UNIVERSIDAD CATÓLICA

DE SANTA FE

FACULTAD DE PSICOLOGÍA

ESTADÍSTICA

CARRERA: LICENCIATURA EN PSICOLOGÍA

EQUIPO DOCENTE:

TORRES, KARINA

CAMMISI, MARÍA EUGENIA

FERNÁNDEZ, YAMILA

DOCENTES COLABORADORES:

MEYER, REGINA

SCALENGHE, JUAN PABLO

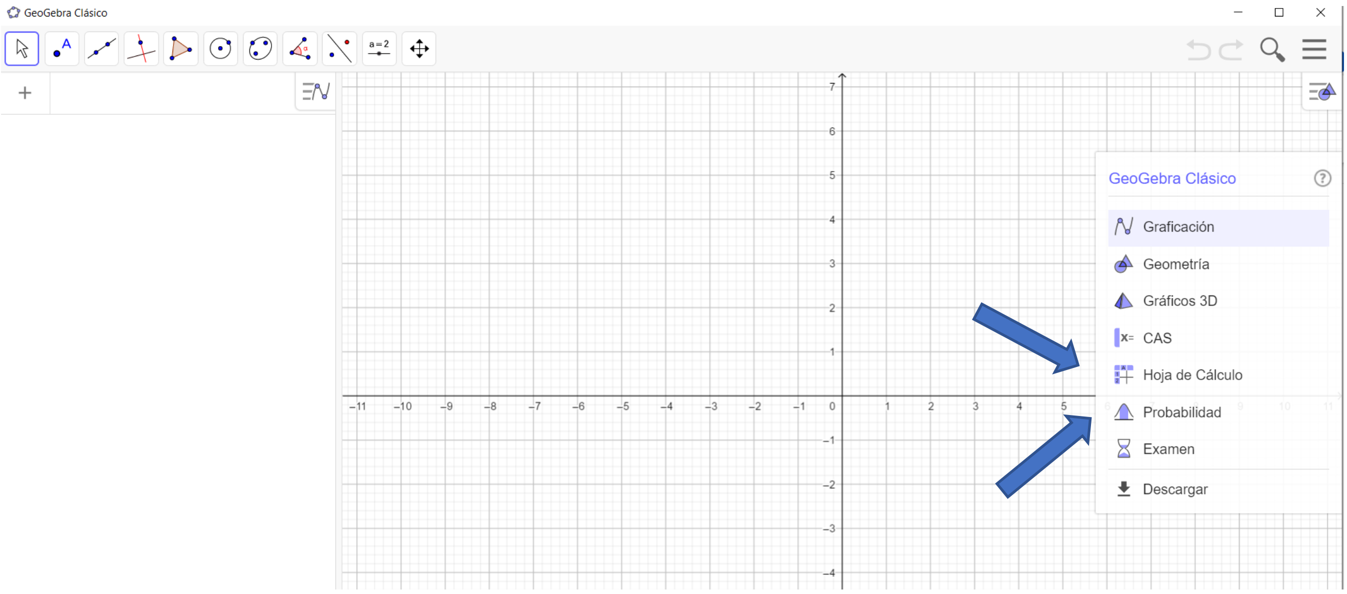
ÍNDICE

PRESENTACIÓN

La asignatura Estadística corresponde al primer año de la carrera Licenciatura en Psicología y es cuatrimestral. Se divide en 5 unidades temáticas:

Unidad 1: CONCEPTOS BÁSICOS.
Unidad 2: ESTUDIO DE DISTRIBUCIONES DE FRECUENCIAS.
Unidad 3: ANÁLISIS EXPLORATORIO DE DATOS.
Unidad 4: DISTRIBUCIONES DE PROBABILIDAD.
Unidad 5: RELACIONES ENTRE VARIABLES ESTADÍSTICAS.

En las distintas temáticas que abordaremos, utilizaremos el software GeoGebra, es gratuito y puede descargarse tato en pc como en celulares. Utilizaremos la vista de hoja de cálculo (para estadística) y probabilidad:



<u>Obtención de la regularidad</u>:
Durante el desarrollo del cuatrimestre se realizará una evaluación parcial, de carácter presencial, que incluirán los contenidos correspondientes a las Unidades 1 y 2. El mismo, debe aprobarse con un mínimo del 60%.
También se evaluará el uso del software Geogebra a través de un trabajo práctico domiciliario que debe estar aprobado con un mínimo de 60%.
Tener un mínimo de 66% de asistencia a clase.
<u>Aprobación de la materia</u>:
En este caso de ser alumno regular, deberá rendir examen final que consistirá en un examen escrito en el que se deberá responder y resolver diversas cuestiones integradoras de los conceptos desarrollados en el cuatrimestre. Este examen se aprobará con un mínimo de 60%.
Caso contrario, se considera al alumno libre de acuerdo con el Artículo 20. De este modo, el alumno deberá rendir un examen escrito teórico-práctico integrador de todos los conceptos contemplados en el programa de la materia. El examen final de esta categoría tendrá mayor extensión que el del alumno regular y se aprobará con un mínimo de 70% sobre el total de las actividades propuestas en dicho examen.

UNIDAD 1: CONCEPTOS BÁSICOS

Antes de comenzar, respondamos las siguientes consignas:

- *¿Qué es para ustedes la Estadística? ¿Para qué piensan que sirve? ¿Por qué aplicarla en Psicología?*
- *Si tuvieran que hacer una encuesta sobre las características de los ingresantes a Psicología… ¿cómo harían la encuesta? ¿a todos los ingresantes o a un grupo? Si lo hacen con un grupo… ¿cómo lo seleccionarían?*

La estadística y sus campos de aplicación

ESTADÍSTICA: Es una disciplina que utiliza recursos matemáticos para organizar y resumir una gran cantidad de *datos* obtenidos de la realidad, e inferir conclusiones respecto a ellos.

Los campos de aplicación de esta disciplina son:

- Organización y análisis de datos no procesados, con el fin de ordenar la información.
- Establecimiento de relaciones de causa y efecto.
- Evaluación de confiabilidad de las conclusiones.
- Supervisión de eventos, tendencias y procesos.
- Predicciones y pronósticos.

¿Para qué sirve la Estadística en Psicología?

Es necesario que los psicólogos en formación desarrollen la competencia para leer, comprender y analizar los resultados presentados en las distintas publicaciones, que utilizan en su gran mayoría los procedimientos estadísticos como método de análisis.

Por otra parte, ayuda a pensar en forma precisa, evaluar la información y aplicar análisis lógico.

Tipos de estadística

Descriptiva: permite analizar características específicas de una población.

Inferencial: proceso mediante el cual se deducen o infieren propiedades de una población a partir del estudio reducido de un grupo, mediante ciertos parámetros.

Conceptos importantes

POBLACIÓN: También llamada Universo, es el conjunto de elementos de referencia sobre el que se realizan observaciones, con el fin de estudiar cierta característica.

MUESTRA: Es un subconjunto de casos o individuos de una población.

Las muestras se obtienen con el fin de inferir propiedades de la totalidad de la población, por lo cual deben escogerse con sumo cuidado. La selección no adecuada de los elementos de la muestra puede provocar errores posteriores a la hora de estimar conclusiones respecto de la población.

Veamos algunos **ejemplos:**

1. Para estudiar cuál es el candidato presidencial por el cual votarán los argentinos en las próximas elecciones, se toma una muestra de 3500 personas de todo el país. La pregunta es la siguiente, ¿por quién votará en las próximas elecciones presidenciales?

En este caso, la **población** es la población electoral del país, es decir, argentinos que pueden votar en las próximas elecciones y la **muestra** es el conjunto de 3500 argentinos que forman parte de esa población.

2. Un estudiante de estadística quiere conocer si los profesores de su universidad, UCSF, prefieren dictar clases con ropa formal o con ropa informal. Para ello, realiza una encuesta a 120 profesores de la UCSF elegidos de forma aleatoria.

Población: Conjunto de todos los profesores de la UCSF.
Muestra: 120 profesores de la UCSF.

TIPOS DE MUESTREO

Para seleccionar la muestra a estudiar existen distintos tipos de muestreos:

Muestreo aleatorio: cuando se selecciónan los elementos al azar y cada miembro tiene igual oportunidad de ser elegido.

Muestreo estratificado: cuando se subdivide en grupos según la variable a investigar. Cada grupo debe corresponder proporcionalmente a la población.

Muestreo sistemático: se establece un patrón o criterio al seleccionar la muestra.

Ejemplo: Supongamos que se desea investigar 4000 estudiantes de una universidad respecto de si creen o no en la vida inteligente en otros planetas.

- **Muestra aleatoria:** se seleccionan 400 estudiantes y se los entrevista al azar.
- **Muestra estratificada:** Si en la población, el 60% de los 4000 estudiantes son mujeres, se debe mantener la misma proporción de mujeres en la muestra. Este tipo de muestreo da más trabajo pero sus resultados son más precisos y confiables.
- **Muestra sistemática:** Se puede elegir de la lista 1 estudiante cada 20 hasta completar los 400.

¿De dónde obtenemos los datos? Existen dos tipos de **FUENTE DE DATOS:**

Primarias: Son las personas o instituciones que han recolectado la información.

Secundarias: Trabajos o publicaciones hechos por personas u organismos que NO recolectaron la información.

Variables y clasificación

Una **variable estadística** es cada una de las características que pueden observarse de un elemento de la muestra.

CLASIFICACIÓN DE LA VARIABLE

La variable se puede clasificar en dos grupos:

a) CUALITATIVAS: Toman valores **NO** numéricos. Dentro de este grupo diferenciamos:

i. NOMINALES: No existe ningún orden entre las categorías de variables. Por ejemplo: el grupo sanguíneo (A, B, AB, 0).

ii. ORDINALES: Existe un cierto orden entre las categorías de las variables. Por ejemplo: El nivel de estudios: Sin estudios, básico, secundario, etc.

b) CUANTITATIVAS: Toman valores numéricos. Dentro de estas se agrupan en:

i. DISCRETAS: Números enteros. Por ejemplo, número de hermanos o de hijos.

ii. CONTINUAS: Puede tomar cualquier valor numérico. Por ejemplo, el peso de un individuo.

Ejemplo 1

El Centro de estudiantes de la Facultad de Psicología de la UBA está interesado en conocer el grado de acuerdo de distintos actores de dicha facultad, respecto de la inclusión de la materia Matemática en el CBC de la carrera. Para ello al finalizar el primer cuatrimestre de 2005, tomó una muestra aleatoria de 1500 estudiantes y 400 docentes de la carrera. A continuación se presentan los resultados (no reales):

Grado de acuerdo	*Alumnos*	*Docentes*
De acuerdo	20%	40%
Indiferente	10%	30%
En desacuerdo	70%	30%
Total	100%	100%

a) Determine quiénes constituyen la población y quiénes la muestra de individuos.
b) Determine la población y la muestra de observaciones
c) Mencione la variable, clasifíquela e indique la escala de medición.

Resolución:

El estudio se lleva a cabo sobre **estudiantes** y **docentes** de la Facultad de Psicología de la UBA, por lo tanto:

a) Hay dos poblaciones de individuos: una está constituida por el conjunto de estudiantes de la carrera de Psicología de la UBA, y la otra por el conjunto de docentes de la carrera de Psicología de la UBA. Hay dos muestras de individuos: una está formada por los 1500 estudiantes seleccionados de la carrera de Psicología de la UBA y la otra por los 400 docentes seleccionados de la misma carrera.

Dicho estudio recoge la opinión de ambos actores respecto del **grado de acuerdo** con la inclusión de la materia Matemática en el CBC de la carrera; en correspondencia con las dos poblaciones y muestras de individuos:

b) Hay dos poblaciones de observaciones: una está formada por el conjunto de valores que representan el grado de acuerdo con la inclusión de Matemática en el CBC para la carrera de Psicología entre los alumnos y la otra por el conjunto de valores que representan el grado de acuerdo con la inclusión de la Matemática en el CBC para la carrera de Psicología entre los docentes. Hay dos muestras de observaciones: una está formada por el conjunto de valores que representan el grado de acuerdo con la inclusión de Matemática en el CBC para la carrera de Psicología entre los 1500 alumnos seleccionados y la otra, por el conjunto de valores que representan el grado de acuerdo con la inclusión de Matemática en el CBC para la carrera de Psicología entre los 400 docentes seleccionados.

c) Variable: Grado de acuerdo con la inclusión de Matemática en el CBC para la carrera de Psicología. Es una variable **Cualitativa**. Escala de medición: **Ordinal** pues sus valores indican un orden o jerarquía.

Ejemplo 2

Clasifique las siguientes variables e indique el nivel de medición de las escalas adoptadas.

a) Calificación de 8 individuos de un grupo según su posición en rendimiento: 1 **al de mayor rendimiento**, 8 **al de menor rendimiento**.

b) Psicodiagnóstico correspondiente a un paciente según los cuadros clínicos (neurosis, psicosis, etc.)

c) Nacionalidad de un sujeto.

d) Cantidad de palabras correctamente leídas por un disléxico en un minuto.

e) Cociente intelectual (Binet-Stern).

f) Tiempo que tarda un alumno de Psicología en concluir su carrera.

g) Temperatura en grados centígrados en Ciudad de Santa Fe a las 00hrs.

h) **Orden de mérito** obtenido por un aspirante en un concurso.

i) Edad de una persona en años cumplidos.

Resolución:

a) Cualitativa. Escala de medición: **Ordinal.** En este caso, los valores asignados representan sólo un orden en el rendimiento de los 8 individuos. De dos individuos a los que se ha asignado números diferentes, puede decirse cuál de ellos representa un mayor rendimiento, pero no cuánto mayor es el mismo.

b) Cualitativa. Escala de medición: **Nominal.** Los valores asignados a esta variable son sólo atributos. Dados los valores correspondientes a dos pacientes, sólo podemos decir si poseen el mismo o diferente psicodiagnóstico.

c) Cualitativa. Escala de medición: **Nominal.** La nacionalidad es un atributo del individuo e, igual que en el punto anterior, de dos sujetos sólo diremos si poseen la misma o diferente nacionalidad.

d) Cuantitativa discreta. Escala de medición: **De razón.** El cero de la variable es absoluto, indica que una persona disléxica no pudo leer correctamente al menos una palabra en un minuto. Es una variable discreta ya que no es posible que la persona lea, por ejemplo, dos palabras y media.

e) Cuantitativa continua. Escala de medición: **De Intervalo.** O **Cualitativa.** Escala de medición: **Ordinal.** Una persona que posee un IQ de 100, tiene una inteligencia más cercana a alguien con un IQ de 110 que otra con un IQ de 60; parece ser una escala de intervalos, pero es difícil establecer que la escala realmente posea intervalos iguales entre las unidades adyacentes. Sin embargo, muchos investigadores consideran tales variables como si fuesen medidas en escalas de intervalos, en particular cuando el instrumento de medición es bastante estándar, cómo es el caso del WAIS.

f) Cuantitativa continua. Escala de medición: **De razón.** El 0 de la escala no tiene por qué ser un valor posible para la variable.

g) Cuantitativa continua. Escala de medición: **De intervalo.** El 0 de la escala no es absoluto, no representa realmente ausencia de temperatura.

h) Cualitativa. Escala de medición: **Ordinal.** Los valores asignados a cada aspirante indican un orden o jerarquía.

i) Cuantitativa discreta. (pues se considera en años cumplidos) Escala de medición: **De intervalo.** El 0 no significa ausencia de variable. Una persona de 3 meses tiene tiempo de vida, pero aún no cumplió años.

Ejemplo 3

Un equipo de psicopedagogos desea estudiar las habilidades adquiridas en el aprendizaje escolar de la lectoescritura mediante una prueba de rendimiento (en la escuela usual de 1 a 10) y el nivel de agresividad (bajo/medio/alto) de niños provenientes de hogares con carencias económicas de Santa Fe.
Determine la población, las variables estadísticas, su clasificación y su escala de medición.

Resolución:

La **población** está formada por los niños escolarizados provenientes de hogares con carencias económicas de Santa Fe.
Las **variables estadísticas** son “calificación en la prueba de habilidades en lectoescritura” y “nivel de agresividad”.
La calificación en una prueba es una **variable cuantitativa discreta**, sus valores son cantidades numéricas y están representados por los números del 1 al 10 (enteros). El “nivel de agresividad” es **cualitativa** pues sus valores son “bajo, medio y alto”.
Por último, en la variable “calificación en la prueba de habilidades de lectoescritura”, la escala de medición se considera **de razón** si la habilidad se representa por el rendimiento entendido como cantidad de respuestas correctas en la prueba. La escala de medición de la “agresividad” es claramente **ordinal** pues sus valores indican un rango o jerarquía.

Ejemplo 4

Una Consultora de Recursos Humanos está interesada en estudiar el clima laboral de una organización. Con este fin se efectúa una evaluación objetiva al contar la cantidad de “quejas” diarias que brindan 30 de sus 150 empleados.
a) Determine la población y muestra de individuos.
b) ¿Cuál es la variable estadística? Clasifíquela.
c) Determine la escala de medición de la variable.

Resolución:

a) La **población** está constituida por el conjunto de empleados de la organización (que en este caso son 150). La **muestra** está constituida por los 30 empleados de la organización seleccionados.
b) La **variable estadística** es “cantidad de quejas diarias de los empleados”. La variable es **cuantitativa discreta** pues la cantidad de quejas es un número entero no negativo.
c) La escala de medición de la variable es **de razón** pues el valor cero indica que el empleado no ha realizado ninguna queja.

UNIDAD 2: ESTUDIO DE DISTRIBUCIONES DE FRECUENCIAS

Tablas de frecuencia para variables cualitativas

Al trabajar con variables cualitativas, sólo podemos organizarlas en categorías o grupos a través de tablas que incluyan al menos un de estas tres columnas con la siguiente información:

1- *Frecuencia Absoluta (fa)*: Es el número de veces que se repite un determinado valor de la variable.
2- *Frecuencia Relativa (fr)*: es la razón entre la frecuencia absoluta y la cantidad total de observaciones. Por lo tanto en una muestra de ***n*** datos: ***fr = fa / n.***
3- *Frecuencia relativa porcentual (fr%)*: es la frecuencia relativa multiplicada por 100, es decir, ***fr% = (fa/n) * 100.***

Llamamos tabla completa cuando incluimos todas las columnas.

Ejemplo 1

Se realizó una encuesta para conocer el nivel de estudios alcanzado por 50 personas de barrio Las Flores de la ciudad de Santa Fe, dividiendolo en 4 categorias: sin estudios, estudios primarios, estudios secundarios y estudios superiores.

Estudios alcanzados	**Cantidad de personas (fa)**	**Cantidad relativas de personas (fr)**	**Cantidad relativa porcentual de personas (fr%)**
Sin Estudios	6	0,12	12
Estudios Primarios	15	0,30	30
Estudios Secundarios	20	0,40	40
Estudios Superiores	9	0,18	18
Totales	50	1	100

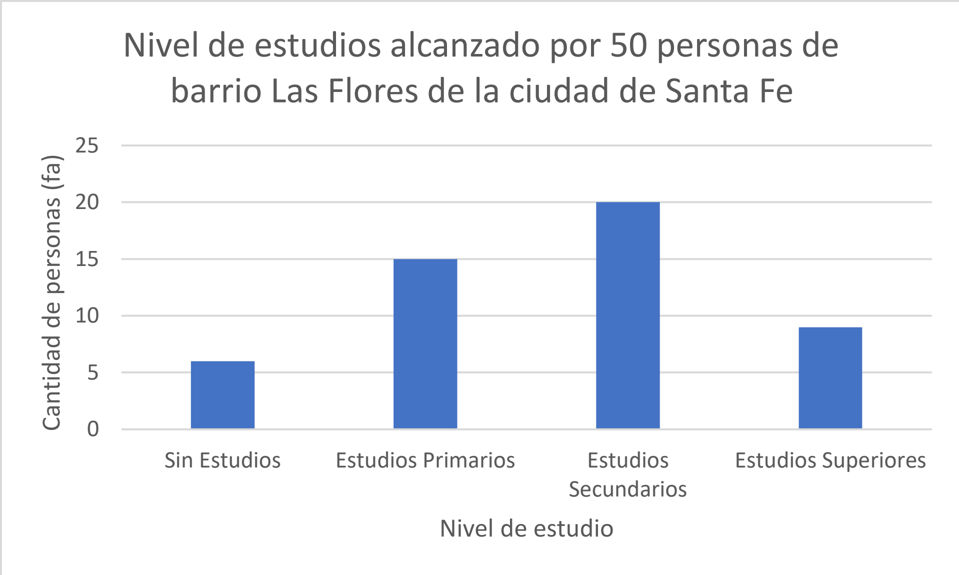
¿Qué información debe contener un gráfico?

• Título General y títulos en los ejes
• Nombre de las categorías y clases
• Ubicación del cero y escalas apropiadas
• Referencias si es necesario
• Fuentes de donde se extrajo la información

Gráficos para variables cualitativas

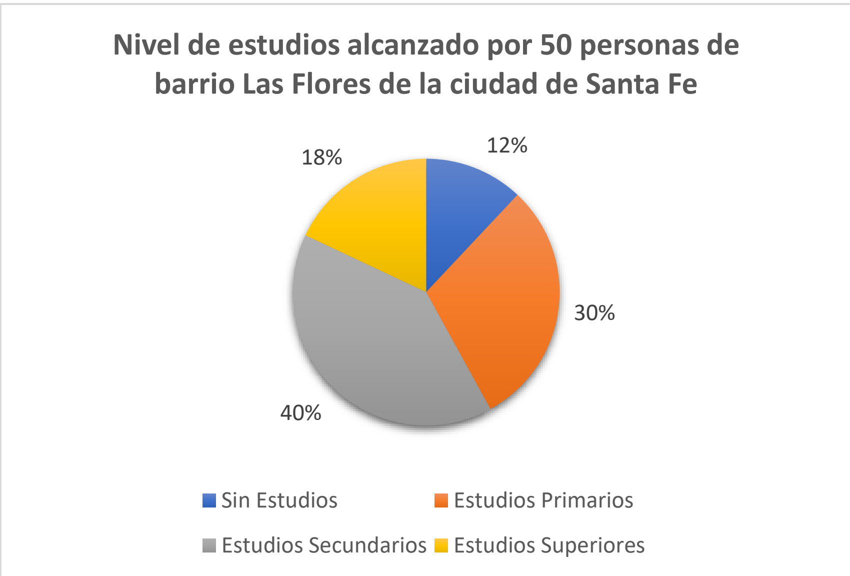
Gráfico de barras verticales

Sobre cada categoría de la variable realizamos una barra cuya altura sea la frecuencia que estemos representado (la escala se indica en el eje vertical).
Las barras se realizan todas de igual ancho y se pintan del mismo color.



- *¿La única frecuencia que puede representarse es la absoluta?*
- *¿Se te ocurren otras formas de realizar un gráfico de barras?*

Gráfico de sectores o torta



Se utiliza para representar la frecuencia relativa porcentual.

Para realizarlo a mano necesitamos transportador.

Para determinar la amplitud del ángulo de cada categoría (con frecuencia absoluta fa) realizamos el siguiente cálculo:

$$amplitud = \frac{fa.360°}{n}$$

Este gráfico NO es recomendable cuando la variable tiene más de cuatro categorías ya que podría verse "deformado".

Ejemplo 2

Se realizó una encuesta para conocer la principal preocupación de los habitantes del país. Se encuestaron 500 personas en el centro de la Ciudad de Santa Fe en Enero de 2020.

Problemas del País	Cantidad de Habitantes (fa)	Cantidad relativa de habitantes (fr)	Cantidad relativa porcentual de habitantes (fr%)
Inflación	185	0,37	37
Desempleo	140	0,28	28

Inseguridad	75	0,15	15
Corrupción	50	0,10	10
Educación	30	0,06	6
Salud	20	0,04	4
Totales	500	1	100

Gráfico de Barras horizontales (similar al de barras, pero invirtiendo los ejes)

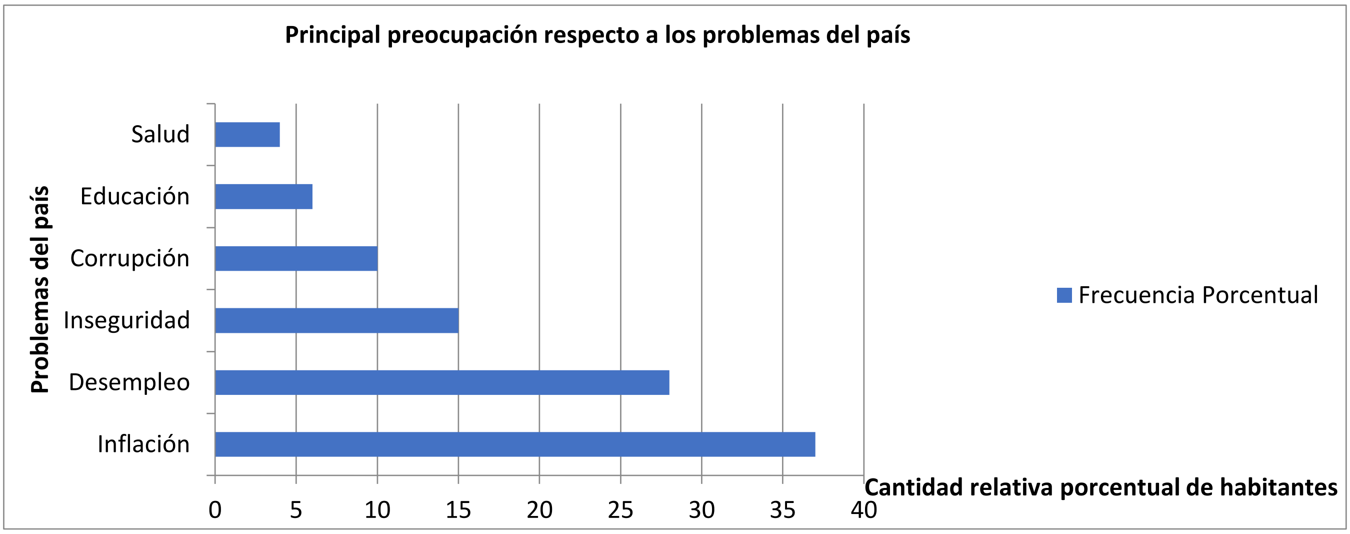
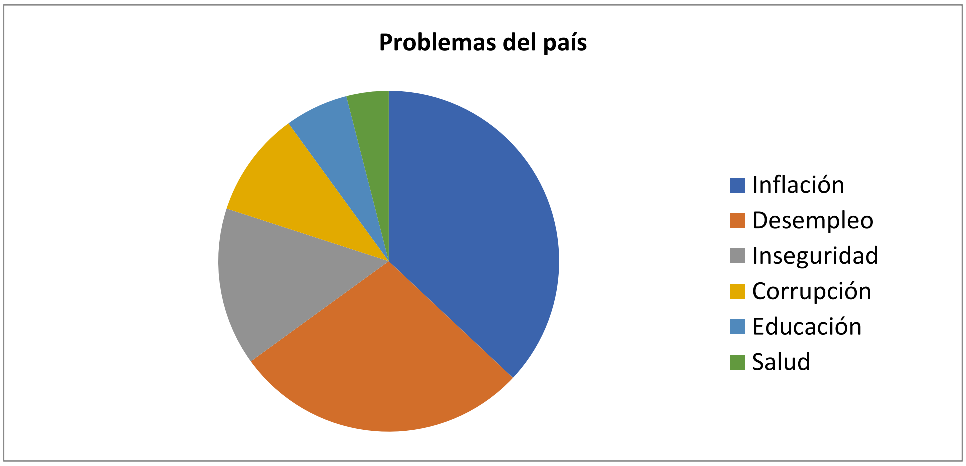


Gráfico de sectores o torta



¿Los gráficos presentados son adecuados? ¿Por qué?

Informes estadísticos

Un Informe estadístico basado sólo en gráficos debe contener:

- Población con la que se va a trabajar
- Muestra: tamaño y tipo
- Variable o Variables que se están estudiando

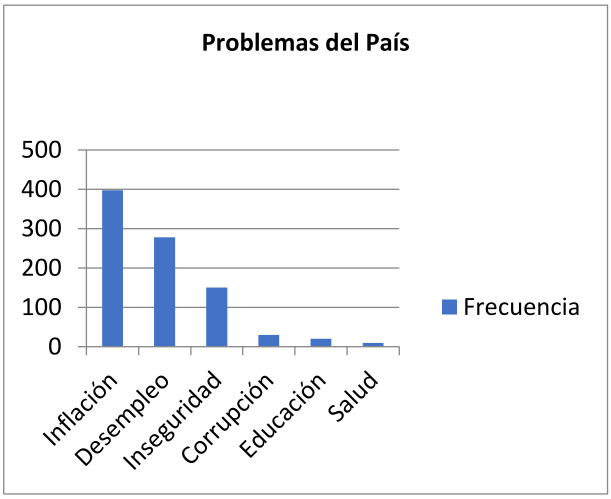
- Mencionar el instrumento de recolección de datos.
- Describir las características principales de la distribución, brindando información sobre porcentajes y medidas estadísticas.

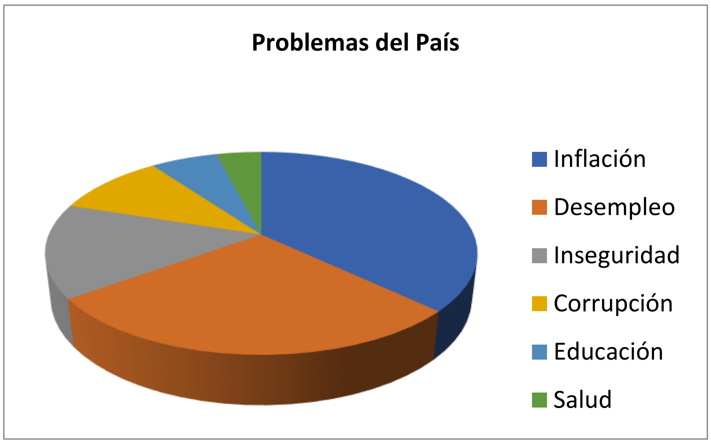
Informe ejemplo 2

Se realizó una encuesta para conocer la principal preocupación de los habitantes del país a principio del nuevo mandato presidencial. A tal efecto, se encuestaron 500 personas en el centro de la Ciudad de Santa Fe en enero de 2020 y se obtuvieron los siguientes resultados:

- El 37% de los encuestados, 185 personas, consideró a la Inflación
- El 28% de los encuestados, 140 personas, consideró al Desempleo
- El 15% de los encuestados, 75 personas, consideró a la Inseguridad
- El 10% de los encuestados, 50 personas, consideró a la Corrupción
- El 6% de los encuestados, 30 personas, consideró a la Educación
- El 4% de los encuestados, 20 personas, consideró a la Salud

Gráficos distorsionados





¿Qué opinas acerca de los gráficos anteriores? ¿Son correctos? ¿qué les cambiarías?

Otros gráficos para variables cualitativas

Comparación de dos o más caracteres

<u>Gráfico de barras bidireccionales</u>: Se utilizan cuando se desea graficar cantidades positivas y negativas, como pérdidas y ganancias, cambios en porcentaje, etc.

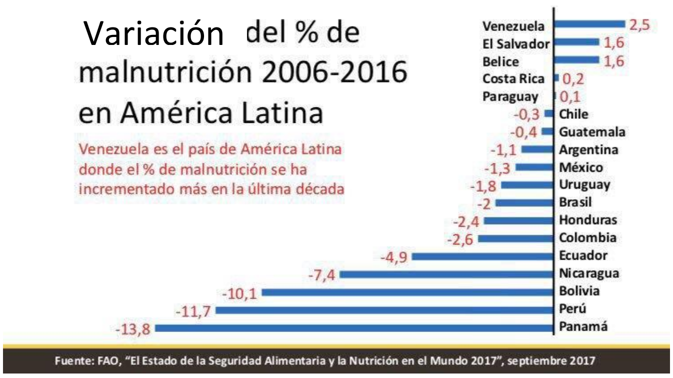
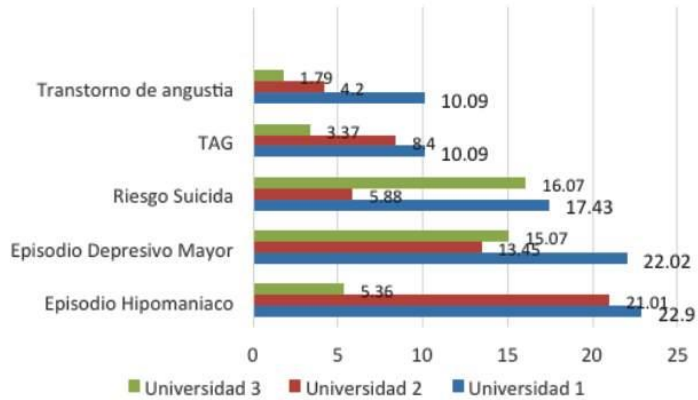


Gráfico de barras adosadas: Son útiles cuando se necesitan realizar comparaciones entre grupos (sexos, instituciones privadas y públicas, países, etc.)

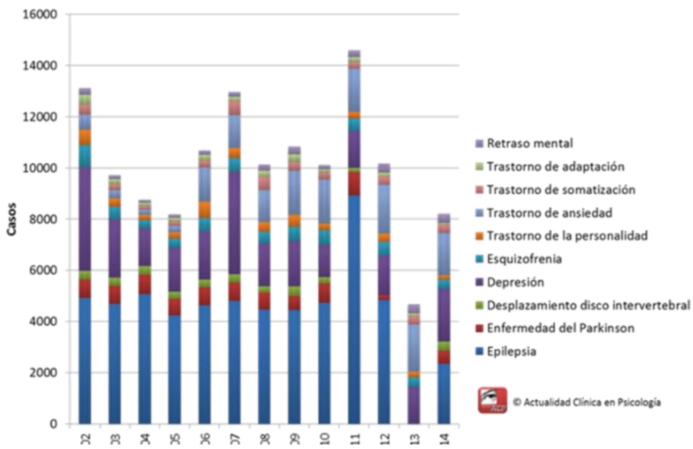


TAG: trastorno de ansiedad generalizada

Gráfico 1. Frecuencias de los principales trastornos mentales de los Estudiantes de Medicina en tres universidades de Lambayeque-Perú 2012.

Gráfico de barras apiladas: Son útiles cuando se quiere mostrar cual es la parte del total representa cada grupo:

PREVALENCIA Y TENDENCIA DE LOS PRINCIPALES TRASTORNOS MENTALES EN LA CIUDAD DE MÉXICO; ANÁLISIS DE DATOS DEL 2002 AL 2014

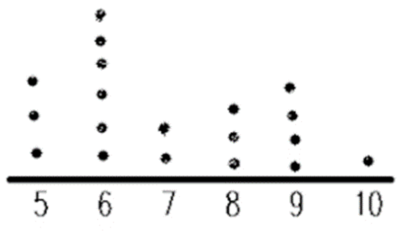


Gráficos para variables cuantitativas

Gráfico de puntos

- Es el gráfico más simple para variables cuantitativas discretas y se usa cuando se tiene un lote de datos pequeño (sin agrupar).
- Se dibuja cada valor observado como un punto ubicado en el valor que corresponda sobre el eje horizontal, ubicando aquellos datos que toman los mismos valores, unos sobre otros.

Ejemplo con el siguiente lote: 5, 7, 6, 7, 8, 5, 6, 6, 10, 9, 9, 5, 8, 9, 6, 8, 6, 9 y 6.



En GeoGebra:

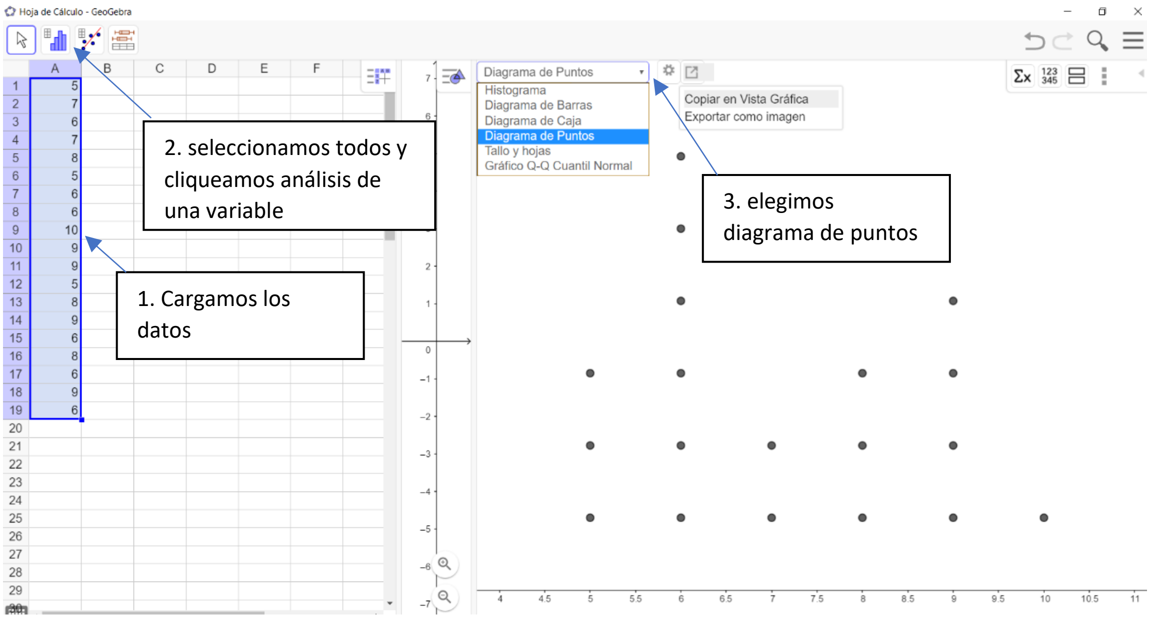
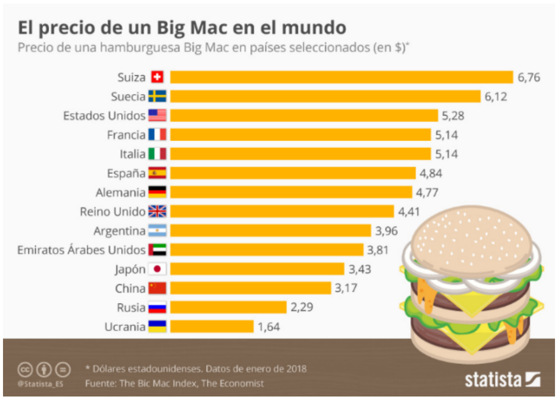


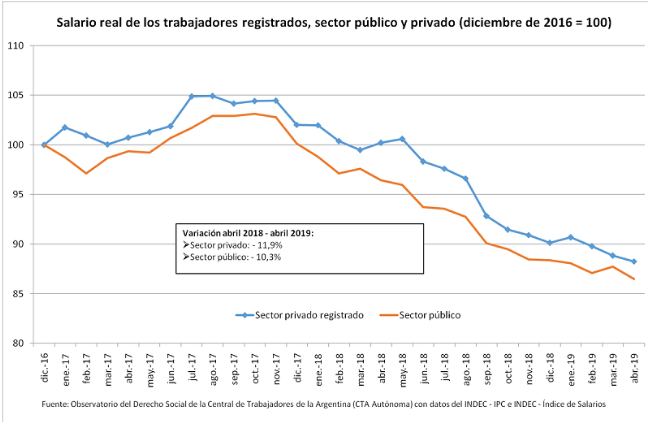
Gráfico de barra (casos puntuales)

Una variable cuantitativa medida para diferentes segmentos de poblaciones o para diferentes categorías se puede graficar usando gráficos de torta o de barras. Este es un caso especial en el que se puede utilizar este tipo de gráficos para estas variables.



<u>Gráfica series de tiempo</u>

Una variable cuantitativa medida a través del tiempo corresponde a lo que se denomina una serie de tiempo. Se puede graficar por medio de un gráfico de líneas o un gráfico de barras (si es un período corto).



Tablas y gráficos para variables cuantitativas agrupadas en frecuencias simples

Al trabajar con variables cuantitativas discretas, cuya dispersión de datos (del valor menor al valor mayor) no sea muy grande, podemos agruparlas en categorías o grupos y ordenarlos de menor a mayor. Al realizar tablas podemos realizar 6 columnas con la siguiente información y sus frecuencias acumuladas:

1- *Frecuencia Absoluta (fa)*: Es el número de veces que se repite un determinado valor de la variable

2- *Frecuencia Relativa (fr)*: es la razón entre la frecuencia absoluta y la cantidad total de observaciones. Por lo tanto, en una muestra de ***n*** datos: ***fr = fa / n***

3- *Frecuencia relativa porcentual (fr%)*: es la frecuencia relativa multiplicada por 100, es decir, ***fr% = (fa/n) * 100***

4- *Frecuencias acumuladas*: las tres columnas anteriores (fa, fr y fr%) pueden acumularse sumando desde el menor valor de la variable hasta la que corresponde al mayor valor de la misma.

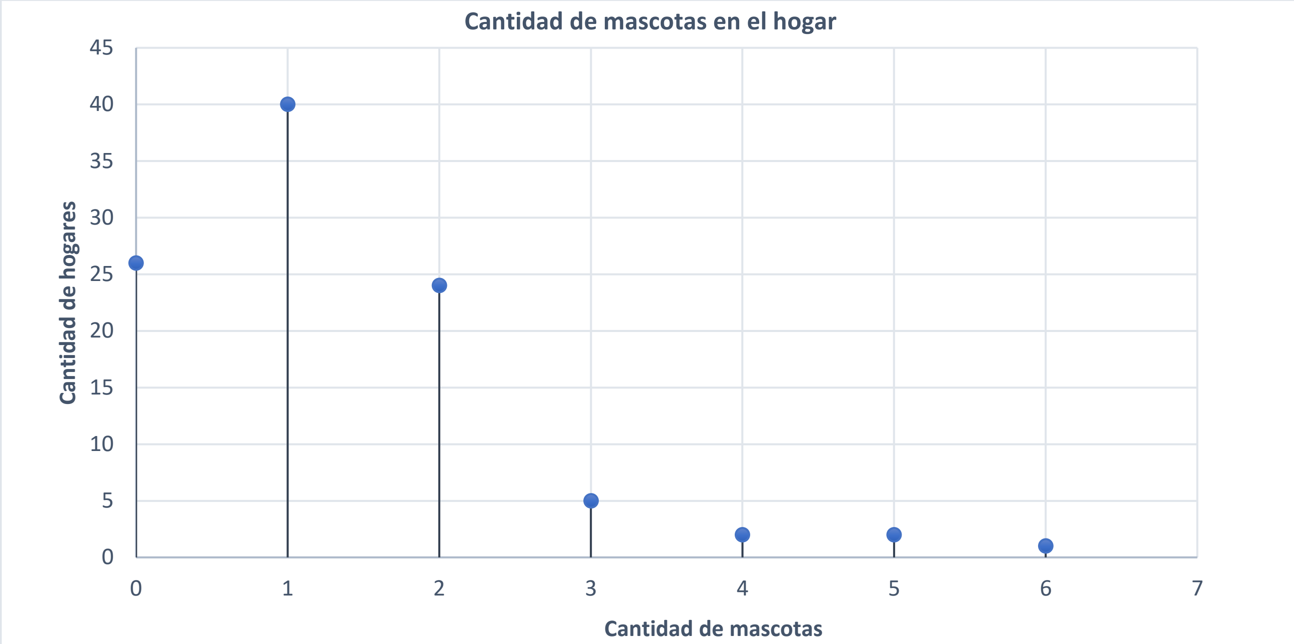
Cantidad de mascotas en el hogar	Cantidad de hogares (fa)	Cantidad de hogares acumulada (Fa)	Cantidad relativa de hogares (fr)	Cantidad relativa de hogares acumulada (Fr)	Cantidad relativa porcentual de hogares (fr%)	Cantidad relativa porcentual de hogares (fr%)
0	26	26	0,26	0,26	26	26
1	40	66	0,4	0,66	40	66
2	24	90	0,24	0,9	24	90
3	5	95	0,05	0,95	5	95
4	2	97	0,02	0,97	2	97
5	2	99	0,02	0,99	2	99
6	1	100	0,01	1	1	100
TOTALES	**100**		**1**		**100**	

n=tamaño de muestra

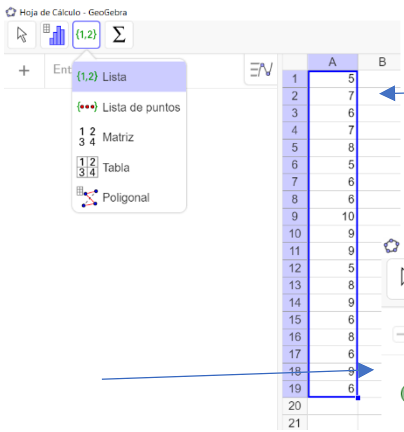
La columna de fr siempre suma 1 (o un valor aproximado si redondeamos)

La columna de fr% siempre suma 100 (o un valor aproximado si redondeamos)

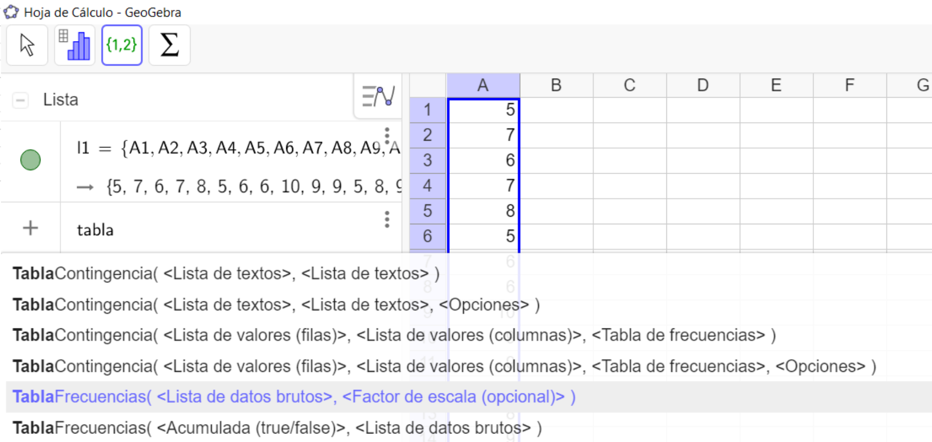
__Gráfico de bastones__: similar al de barras, pero solo hacemos una línea sobre el valor de la variable:



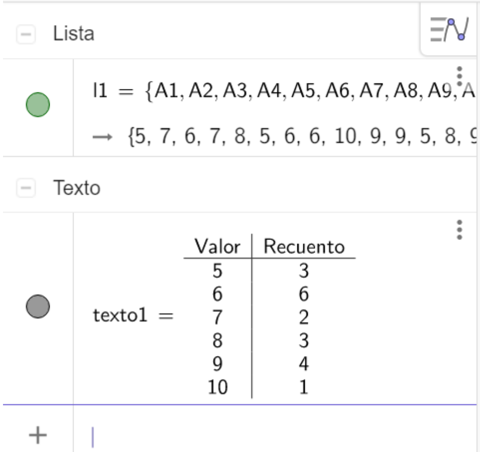
Siguiendo el ejemplo anterior, si tenemos los datos cargados en GeoGebra podemos hacer:



Seleccionamos los datos y creamos una lista (por defecto se llama l1 pero se le puede cambiar el nombre).



En la vista algebraica, en entrada escribimos tabla de frecuencias y seleccionamos la opción que se muestra.



En el lugar de lista de datos brutos ponemos l1 (o el nombre que hayamos dado a la lista) y en factor de escala nada. En consecuencia, obtendremos la tabla con frecuencias absolutas.

Tablas y gráficos para variables cuantitativas en datos agrupados en intervalos

Al trabajar con variables cuantitativas discretas, cuya dispersión de datos (del valor menor al valor mayor) sea muy grande, o trabajemos con variables cuantitativas continuas, debemos crear la tabla de frecuencias, agrupando los datos en intervalos de igual amplitud. Para ello, se deben seguir los siguientes pasos en el orden que se establece:

1. Determinar el tamaño de la muestra o población (n o N)
2. Calcular en nro. de intervalos por medio de la siguiente regla: (se toma **sólo la parte entera** del resultado)

   REGLA DE STURGES= Nro. Int. = K = {1+3,322. Log n}

3. Obtener el ancho o amplitud de los intervalos de la siguiente manera: A = (máx – mín)/ K , donde K es la parte entera, es decir, el número que está delante de la coma, obtenido en el punto 2.
   En este caso se obtiene un valor aproximado que puede redondearse al entero siguiente.

4. Determinar los límites inferior y superior de cada intervalo, considerando que se debe trabajar con intervalos semiabiertos, es decir incluyendo la variable de la izquierda, pero no la de la derecha (por ejemplo, dado el intervalo [4 – 5,5) incluimos variables cuyo valor sea 4 o mayor, pero menores a 5,5).
5. Construir la tabla con la cantidad de intervalos obtenidos y con la amplitud determinada e incluyendo las frecuencias que correspondan según el paso anterior. Las columnas son las mismas que utilizamos para datos agrupados en frecuencias simples.

Ejemplo

Se anotaron los pesos de 30 jugadores del plantel de los pumas y se obtuvo lo siguiente:

94, 70, 78, 90, 119, 103, 109, 88, 78, 98, 114, 79, 93, 88, 97, 101, 117, 107, 109, 93, 86, 89, 104, 90, 93, 98, 106, 88, 92 y 101

1- Calculamos la cantidad de intervalos (n = 30)

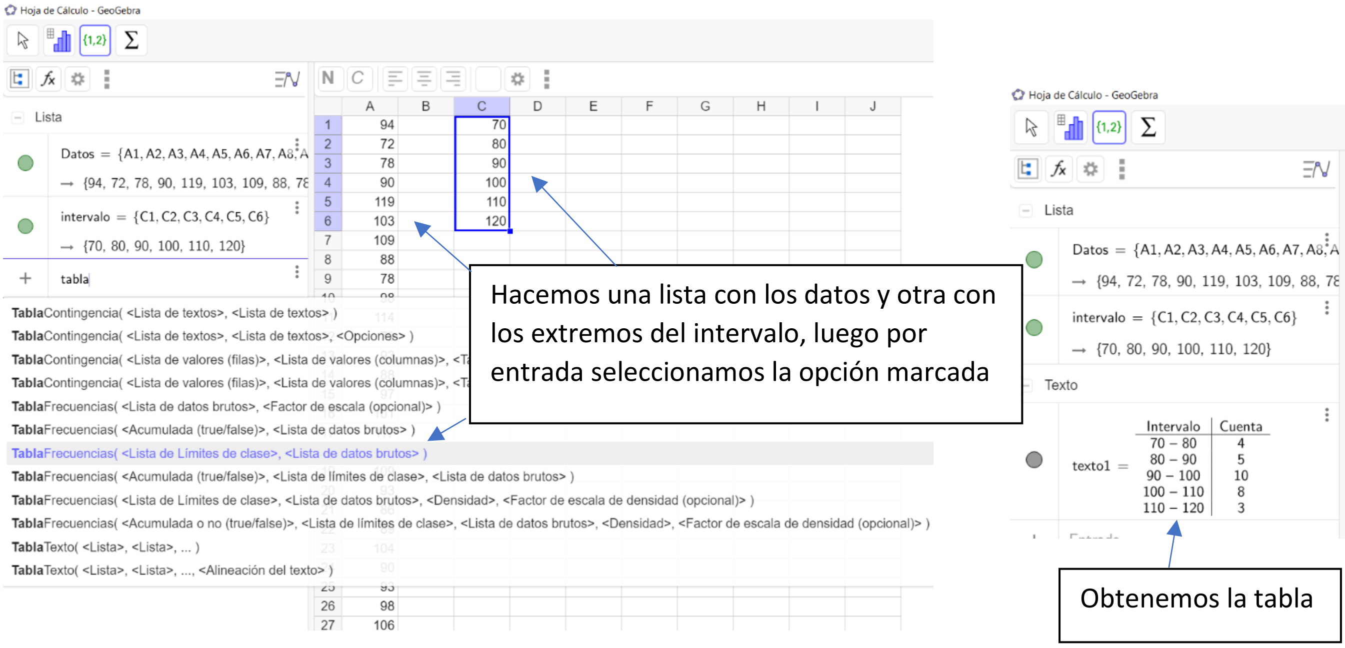
k= (1+ 3,322 . Log 30) = (5,91) Tomamos sólo la parte entera, es decir 5 intervalos

2- Luego, calculamos la amplitud de los intervalos: A= (119-70)/5= 9,8 . En este caso redondeamos a 10

3- A continuación, construimos la tabla con 5 intervalos de amplitud 10.

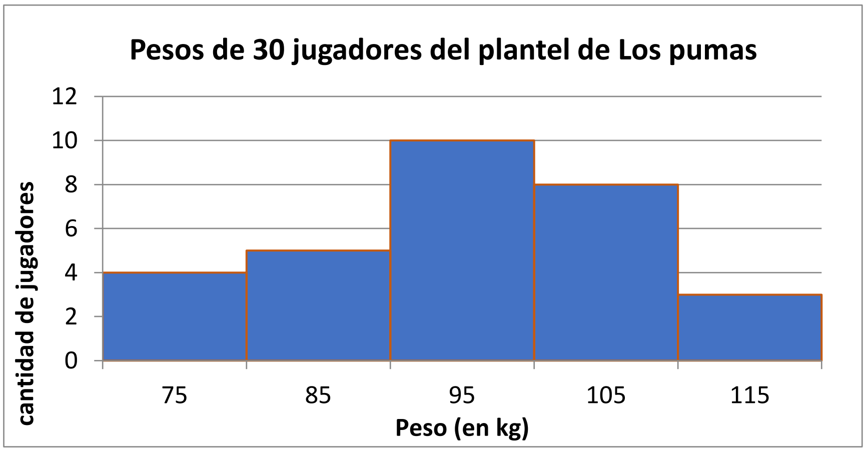
Peso en kg	Frecuencia Abs	Frec. Abs Acumulada	Frecuencia Relativa	Frec Rel Acumulada	Frecuencia %	Frec % Acumulada
Intervalos	fa	Fa	fr	Fr	fr%	Fa %
[70 - 80)	4	4,00	0,13	0,13	13	13
[80 - 90)	5	9,00	0,17	0,30	17	30
[90 - 100)	10	19,00	0,33	0,63	33	63
[100 - 110)	8	27,00	0,27	0,90	27	90
[110 - 120]	3	30,00	0,10	1,00	10	100
Totales	30		1		100	

En GeoGebra:



Histograma:

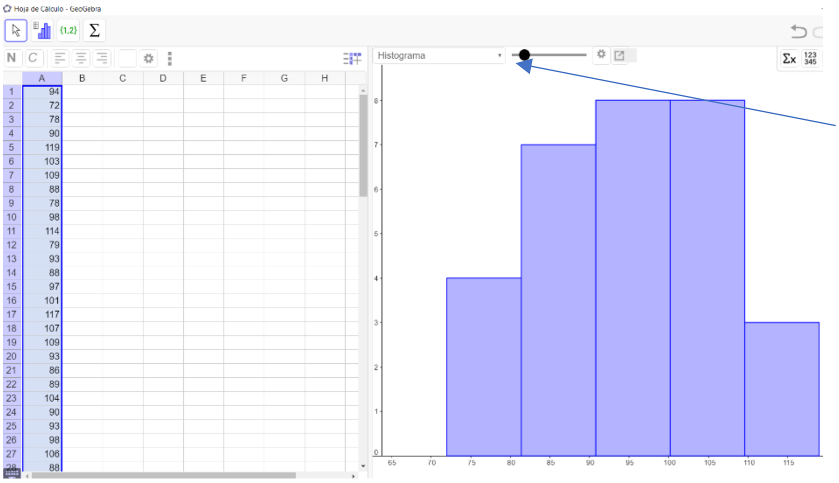
Continuando con el ejemplo de los pesos de los jugadores de los pumas, realizamos el histograma utilizando la frecuencia absoluta: [Notar que el gráfico está marcando el punto medio de cada Intervalo (80-70= 75)]. A continuación, se detalla su construcción.



- Se deben ubicar los extremos de los intervalos, o bien los puntos medios de los mismos sobre el eje horizontal y las frecuencias correspondientes (pueden ser las absolutas, relativas o porcentuales) sobre el eje vertical.
- Tener en cuenta que la continuidad de los intervalos debe quedar representada en el gráfico por la unión de las barras
- El área de cada barra puede representar:

- la proporción de observaciones que caen o se ubican en una clase o intervalo determinado.
- el número de observaciones que se ubican en una clase determinada

En GeoGebra cargamos los datos y al igual que en las variables sin agrupar seleccionamos análisis de una variable:

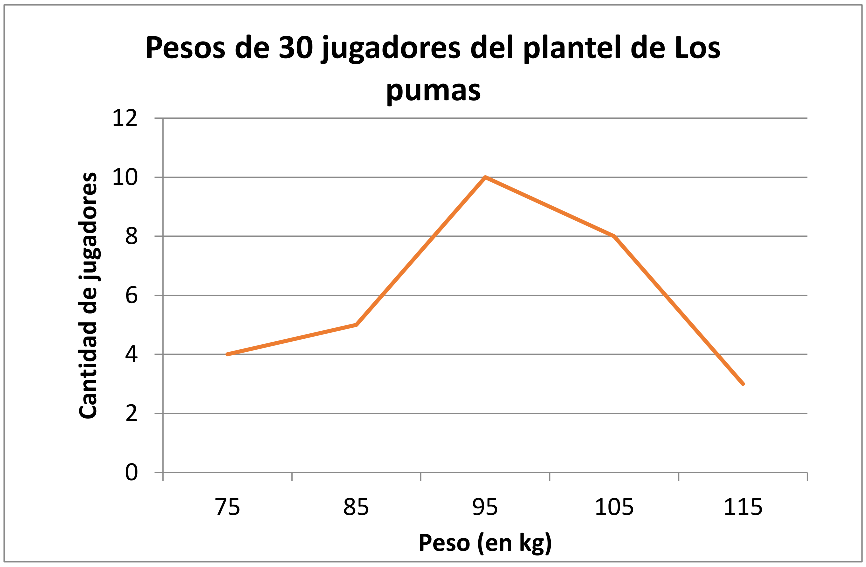


Seleccionamos histograma y en el deslizador podemos elegir cuantos intervalos queremos.

Notar que los dos gráficos tienen diferencias ya que el software no realiza redondeos si no que lo hace de forma exacta. De todas maneras, la forma de la distribución es similar.

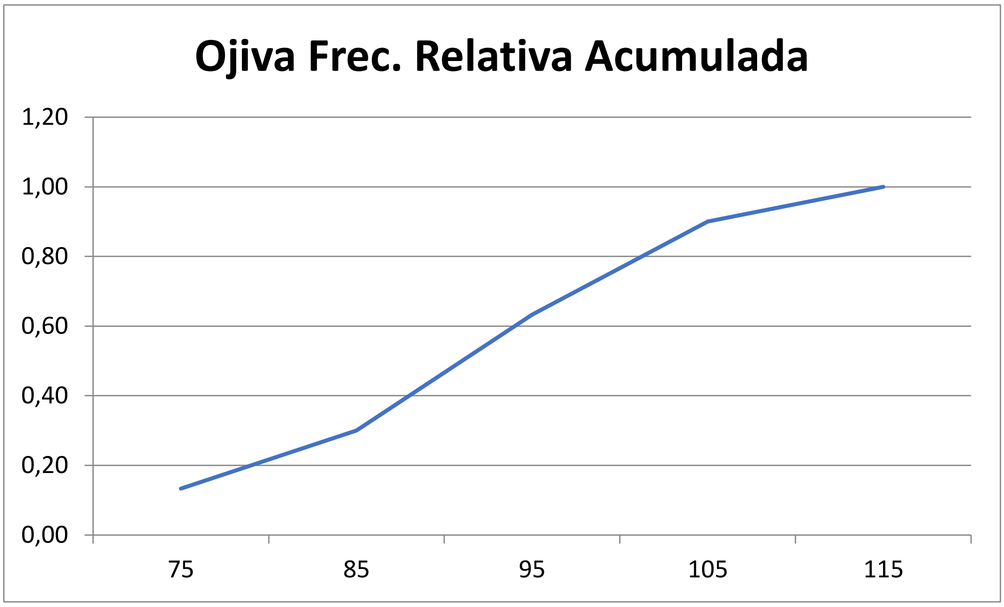
__Polígono de frecuencia__

Es un gráfico de líneas continuas, en dónde se marca el punto medio de cada intervalo a la altura que corresponda según la frecuencia, pudiendo representar cualquiera de ellas, absolutas, relativas o porcentuales. (Este gráfico no se encuentra disponible en GeoGebra)



__Ojiva__

Es un gráfico de líneas continuas que se construye uniendo los puntos cuya abscisa son los extremos derechos cada intervalo, y la ordenada la frecuencia acumulada, absoluta, relativa o porcentual. (No disponible en GeoGebra).



<u>Histograma de frecuencias con anchos de intervalos no homogéneos</u>

- Cuando se tienen intervalos con amplitud distinta, no se puede aplicar la regla de Sturges, por lo que en esa situación se deben agrupar los datos siguiendo algún criterio subjetivo
- Para graficar un histograma a partir de datos agrupados de esta manera, se debe considerar que: el área de cada barra debe ser proporcional a la cantidad de datos que hay en ella.
- En consecuencia, se debe graficar la densidad (en lugar de frecuencias)
- La densidad de cada barra se obtiene así: frecuencia absoluta/ancho del intervalo
- El histograma resultante tendrá barras de distintos anchos y alturas correspondientes a las densidades de cada intervalo

Resumen gráficos

Variables

Cualitativas

Frecuencias absolutas, Relativas y porcentuales

Barras

Sectores o torta

Cuantitativas

Discreta en lotes pequeños

Gráfico de puntos (f.absoluta)

Gráfico de bastones (fa, fr, fr%)

En algunos casos excepcionales puede utilizarse barra (cuando aparece un número pero asociado a otra categoría, por ejemplo país)

Agrupadas

Para fa, fr, y fr%:
-Histograma
-Polígono de frecuencia

Para frecuencias acumuladas:
ojiva

PRÁCTICA 1

En todos los ejercicios que puedas, aplicar GeoGebra para verificar tus respuestas

Actividad 1: Clasificar cada una de las variables y definir la escala de medición.

a) Psicodiagnóstico correspondiente a un paciente según los cuadros clínicos (psicosis, neurosis, etc).
b) Edad (dada en años cumplidos).
c) Nacionalidad de un sujeto.
d) Cantidad de palabras correctamente leídas por un disléxico en un minuto.
e) Presión sanguínea de una persona (en mm de mercurio).
f) Tiempo que tarda un alumno de psicología en concluir su carrera.
g) Temperatura en grados centígrados en la ciudad de Santa Fe a las 0 hs.
h) Fumador (sí o no).
i) Estatura.
j) Talle de ropa. (S, M, L, XL, XXL)
k) Peso.
l) Nivel de estrés.
m) Técnicas de estudio.
n) Número de hermanos.
o) Tipos de enfermedades mentales.
p) Etapas del ciclo vital.
q) Calidad de vida.
r) Lugar de residencia.

Actividad 2: En cada uno de los siguientes estudios, especifica:

i) Las variables que se estudian.
ii) Clasifica ambas variables e identifica la escala de medición en la que se mide cada una.
iii) La población y la muestra.

Estudio 1: Un psicólogo de la salud está interesado en saber si la motivación por miedo es eficaz para reducir la incidencia en el consumo de tabaco. Se eligen 40 adultos fumadores entre los individuos que viven en la ciudad XX donde trabaja el psicólogo. Se pide a veinte que fumen un cigarrillo, después de lo cual presencian una película cruel que trata sobre la forma en que el cigarrillo provoca cáncer. Se muestran imágenes cruentas de pulmones y otros órganos enfermos de fumadores fallecidos, en un esfuerzo por inducir el miedo a fumar en estos sujetos. El otro grupo recibe el mismo tratamiento, excepto que ven una película neutral, no relacionada con el consumo de tabaco. Durante los dos meses posteriores, el experimentador mantiene un registro acerca de la cantidad de cigarros fumados por día por los participantes. Se calcula entonces para cada grupo el promedio de cigarros fumados diariamente, desde la exhibición de las películas, y estos promedios se comparan para determinar si el filme inductor del miedo tuvo el efecto deseado o no.

Estudio 2: Un psicólogo clínico y un psiquiatra están interesados en evaluar tres métodos para tratar la depresión: la medicación, la reestructuración cognoscitiva y el ejercicio. Se incluye una cuarta condición de tratamiento, un grupo que se aliviará solo con espera, para tener un grupo control. Se toma una muestra de 60 personas deprimidas de la ciudad de Santa Fe, asignando al azar 15 individuos a cada método de tratamiento. Los tratamientos se administran durante 6 meses, después de lo cual cada persona recibe un cuestionario diseñado para medir el grado de depresión. El cuestionario tiene una escala de 0 a 100, donde los puntajes mayores indican un mayor grado de depresión.

Actividad 3: Dar ejemplos de variables que pueden definirse y ser estudiadas en las siguientes poblaciones y clasificar según su tipo:

a) Medicamentos que se expenden en una farmacia.
b) Pacientes que concurren a un consultorio psicológico.
c) Alumnos ingresantes a la Facultad de Psicología.

Actividad 4: El centro de estudiantes de la Facultad de Psicología está interesado en conocer el grado de acuerdo de alumnos y docentes de la facultad respecto de la inclusión de la materia Matemática en la carrera. Para ello al finalizar el primer cuatrimestre del año 2013, se tomó una muestra aleatoria de 1500 estudiantes y 400 docentes de la carrera. A continuación, se presentan los resultados:

Grado de acuerdo	**Alumnos**	**Docentes**
De acuerdo	20%	40%
Indiferente	10%	30%
En desacuerdo	70%	30%
Total	100%	100%

a) Determinar la población, muestra, variable (clasificar) y escala de medición.
b) Completar la tabla con las restantes frecuencias.
c) Graficar adecuadamente teniendo en cuenta el tipo de variable estudiada. En el grafico debe apreciarse la comparación entre alumnos y docentes.
d) ¿Qué opina la mayoría de los alumnos? ¿Por qué sabemos que es la mayoría?
e) ¿La diferencia de docentes que están de acuerdo con los que no lo están es significativa como para valerse de eso para tomar una decisión?

Actividad 5: Un grupo de estudiantes realizó una investigación para comparar diferentes cuestiones de interés entre un barrio y las familias de la escuela. Se encuestaron a personas del Barrio los Hornos de la ciudad de Santa Fe y algunos familiares de los alumnos de la escuela a la que pertenecen. Luego de juntar los datos, armaron la siguiente tabla:

ÚLTIMO NIVEL EDUCATIVO COMPLETO	**Porcentaje de familiares de alumnos**	**Porcentaje de personas en el barrio ¨República Los Hornos¨**
Primario completo	**17%**	**53,65%**
Secundario completo	**36%**	**36,58%**
Terciario completo	**47%**	**9.7%**
TOTAL	**100%**	**100%**

a) ¿Está bien que se haga la tabla sólo en frecuencia porcentual? ¿por qué?

b) ¿Cuál es la variable? Clasificar y da su escala de medición.
c) Realizar un gráfico adecuado que se pueda comprar las dos muestras.
d) Realizar un informe que de información significativa.

Actividad 6: En los siguientes planteos responder las siguientes cuestiones

i) ¿Cuál es la población en estudio y cuál la muestra?
ii) ¿Cuál es la variable que se estudia y cómo se clasifica?
iii) ¿Qué escala de medición puede utilizarse?

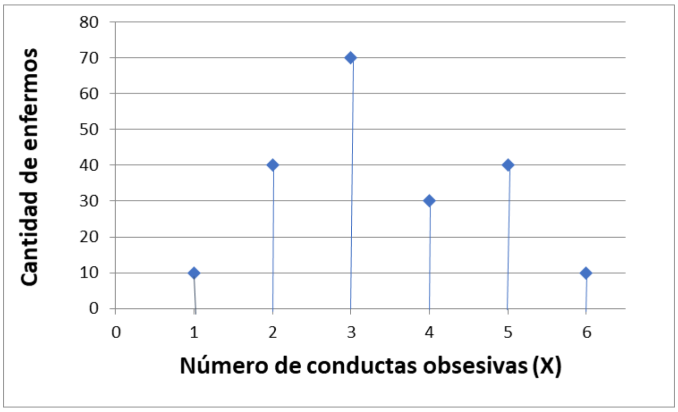
a) El gobierno de un país desea construir un barrio modelo para erradicar las villas miseria en un lugar muy distante de donde éstas se encuentran. A fin de conocer la posibilidad de adaptación al nuevo barrio que tendrían las familias erradicadas se realiza una encuesta a 50 de ellas preguntándoles si se mudarían o no.
b) El dueño de una fábrica de 80 obreros desea conocer el porcentaje de ausentismo entre los obreros de la misma. Para esto, registra los ausentes por día durante un mes.
c) Una firma desea realizar un estudio sobre los ingresos familiares en un área determinada de nuestra ciudad. Con dicho fin se seleccionan aleatoriamente 50 familias dentro de dicha área y se estudian sus ingresos anuales.
d) Una empresa recibe una partida de determinada pieza y desea conocer la calidad de la misma. Extrae una muestra de 100 piezas y mediante ciertos indicadores de prueba de calidad, determina si acepta o rechaza la pieza.

Actividad 7: En una encuesta a 66 hogares de un barrio vulnerado de la Ciudad de Santa Fe en el año 2022, una de las preguntas era: "¿tienen acceso a gas natural?"

Hubo: 21 que respondieron si, 29 que respondieron no y 16 que respondieron no, tenemos garrafa.

a) ¿Cuál es la población y la muestra?
b) Definir la variable en estudio y clasificarla. Explicitar el nivel de medición.
c) Realizar una tabla de frecuencias absolutas y porcentuales.
d) La mayoría de los hogares ¿Cuenta o no con gas natural?
e) Realizar un breve informe.

Actividad 8: El siguiente gráfico muestra el número de conductas obsesivas observadas en un grupo de enfermos mentales de una institución de la ciudad de Bs. As.



a) Responda verdadero o falso y justifique:
   i. La variable es cuantitativa continua.
   ii. La escala es de intervalo.
   iii. El grupo está formado por 200 enfermos.

b) Reconstruya la tabla de frecuencias completa y responda:
   i. ¿Por debajo de que número de conductas obsesivas se encuentra el 75% de los enfermos?
   ii. Se considera que el tratamiento ha sido efectivo en el 25% de enfermos con menor número de conductas obsesivas. ¿Cuántos enfermos son? ¿Cuál es el número máximo de conductas obsesivas que presentaron?

c) Realizar un breve informe estadístico.

Actividad 9: Se quiere estudiar el tiempo que necesitan lo alumnos del nivel polimodal de la Escuela Parroquial “Santa Lucía” para realizar un ejercicio de múltiple opción. Los siguientes datos corresponden al tiempo en minutos que han necesitado 30 alumnos de dicho nivel y escuela para realizar el ejercicio mencionado

4,1	3,1	0,1	6,5	5	2,5	7,4	10	3,3	8	2	0,4	6,4	1,1	9,5
2,8	1,2	1,3	4,1	9,5	4,6	4,3	3,6	5,5	1,4	2,8	1,5	1,6	7,3	7

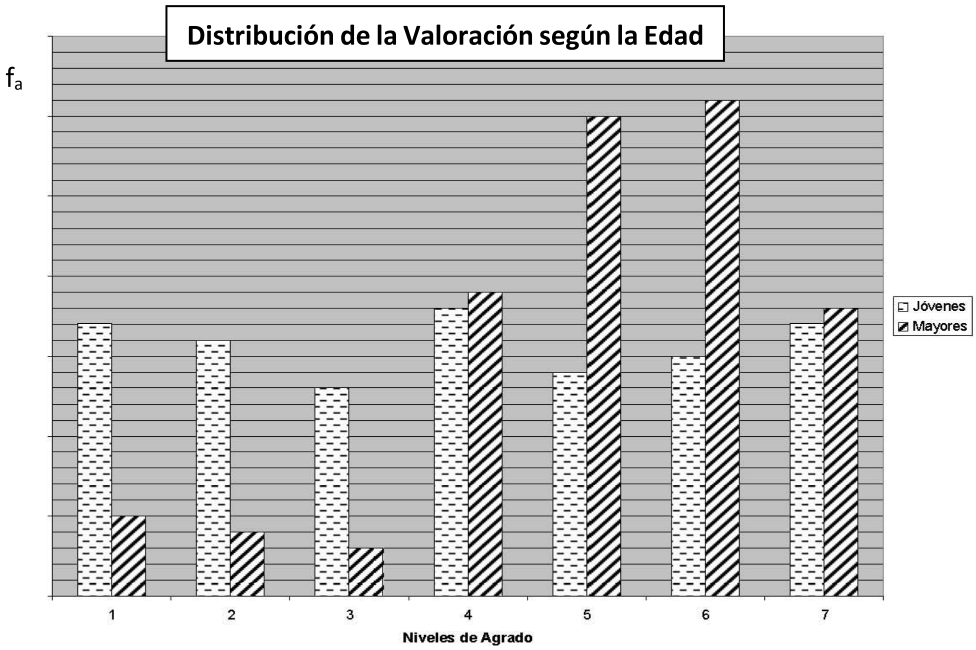
a) Definir la población y la variable en estudio.
b) ¿Cuál es la escala de medición?
c) ¿De qué forma se podrían resumir convenientemente estos datos?
d) Escribir un breve informe estadístico que describa el desempeño general de los alumnos en este múltiple opción.

Actividad 10: La suma total de las notas de los exámenes parciales de 50 estudiantes fueron las siguientes:

11	18	48	45	28	35	31	70	2	82	20	43	40	36	44	3	65	40	17	80	43	69	43	33	15
26	92	59	42	8	1	15	32	12	47	52	60	57	32	62	41	10	51	59	81	11	46	31	22	71

a) Definir la población, la variable y clasificarla. Explicitar la escala de medición.
b) Agrupar los datos en una tabla de frecuencias que corresponda.
c) Describir brevemente las características de la muestra a partir de lo observado en la tabla.
d) Realiza la ojiva y responde cuál es el porcentaje aproximado de notas superiores a 65.

Actividad 11: Para analizar si jóvenes y mayores utilizaban diferencialmente los valores 1 (muy desagradable) a 7 (muy agradable) al juzgar el agrado suscitado por ciertas imágenes, un grupo de investigadores muestran la distribución de frecuencias correspondiente. La escala tomada en el eje de ordenadas es de 20 en 20. Los resultados se exhiben en el siguiente gráfico:



a) Reconstruir las tablas de frecuencias completas para cada grupo.
b) Realizar un informe que permita comparar ambos grupos.

Actividad 12: Se aplicó una batería de pruebas diagnósticas a una muestra de 25 alumnos ingresantes a primer año de la carrera de psicología y se obtuvieron los siguientes puntajes:

4 5 1 6 3 9 1 7 8 7 7 4 10 7 8 10 7 7 8 10 5 8 9 9 8

a) Definir la variable en estudio y clasificarla, dar la escala de medición.
b) Resumir en forma adecuada el conjunto de datos.
c) ¿Cuántos alumnos obtuvieron un puntaje menor a 6?
d) ¿Qué cantidad de alumnos obtuvieron puntajes superiores a 8?
e) ¿Qué porcentaje de alumnos obtuvieron más de 4?
f) ¿Qué porcentaje de alumnos obtuvieron menos de 2?

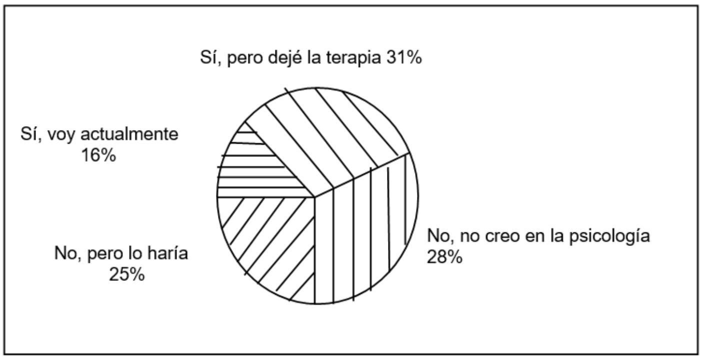
Actividad 13: En una encuesta a 300 jefes y jefas de familia de la ciudad de Santa Fe se les preguntó sobre la cantidad de hijos que tenían y resultaron los siguientes datos:

Cantidad de hijos	0	1	2	3	4	5	6	TOTAL
Cantidad de familias	78	90	32	25	50	15	10	300

Basándote en una tabla de frecuencias adecuada, responde a las siguientes preguntas:

a) ¿Cuántas familias tienen 3 hijos?
b) ¿Qué porcentaje de familias tienen 5 hijos?
c) ¿Cuántas familias tienen 3 hijos o menos?
d) ¿Qué porcentaje de familias tienen 2 hijos o menos?
e) Representa las frecuencias relativas porcentuales y las frecuencias acumuladas porcentuales. A partir del gráfico que corresponda responder los ítems f a i.
f) Considerando el 50% de las familias con menor número de hijos. ¿Cuál es el número máximo de hijos que tiene este grupo de familias?
g) ¿Cuántas familias tienen entre 4 y 6 hijos?
h) ¿Qué porcentaje de familias tienen entre 2 y 4 hijos?
i) ¿Qué número de familias tienen 4 hijos o menos?

Actividad 14: El siguiente diagrama circular muestra los resultados de una encuesta de opinión llevada a cabo sobre 30.646 personas de Ciudad de Buenos Aires y Gran Buenos Aires, publicada en el diario Clarín el 26 de octubre de 2005. La pregunta a la cual respondían es si alguna vez habían concurrido al psicólogo:



a) Construya una tabla de frecuencias a partir del gráfico.
b) Determine cuantas personas no creen en la psicología y cuántas han hecho terapia, pero han abandonado.
c) Realice un breve informe detallando lo observado en el estudio de la muestra.

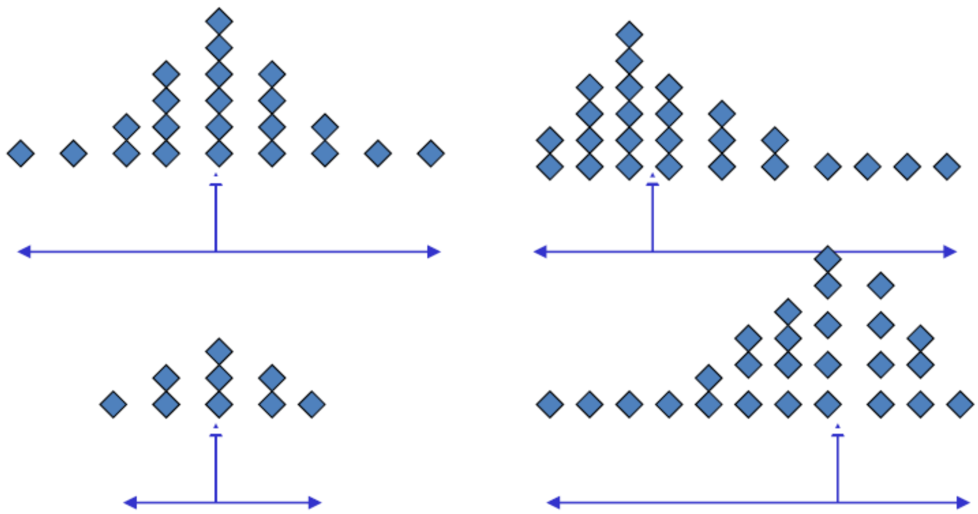
UNIDAD 3: ANÁLISIS EXPLORATORIO DE DATOS

Estructura de una distribución

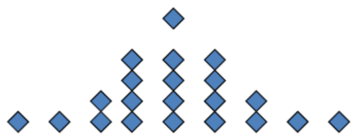
Muchas veces, nos interesa conocer el "comportamiento" de una distribución, si se concentra en algún dato, si es homogénea o heterogénea, etc. Para describir el aspecto general de una distribución, se recomienda:

- Determinar su centro y su dispersión.
- Observar si la distribución tiene una forma simple, que se pueda describir en pocas palabras. Para ello una de las cosas que se deben analizar es la simetría.

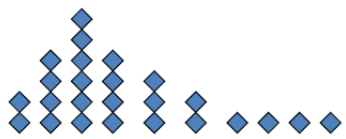
¿Qué podemos analizar a partir de un gráfico? (Centro y dispersión)



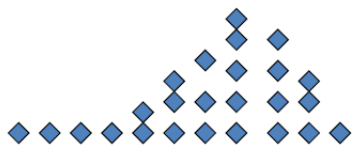
¿Dónde se concentran o se centran los datos sobre el eje horizontal, y cómo se dispersan alrededor del "centro"?



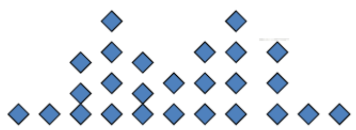
Simétrica y Unimodal: **con un solo "pico"**



Asimétrica a derecha: **pocos datos en los valores mayores de la variable.**



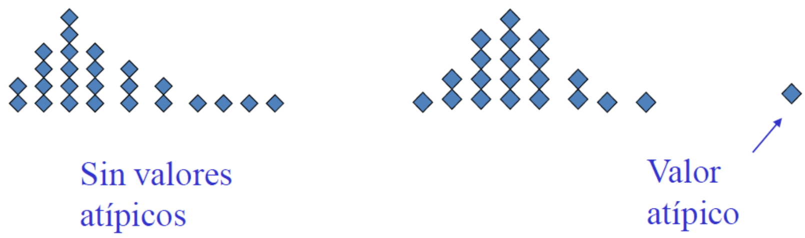
Asimétrica a izquierda: **pocos datos en los valores menores de la variable**



Simétrica y Bimodal: **dos "picos"**

Interpretando gráficos: valores atípicos (“outliers”)

Se denominan valores atípicos a datos que se encuentran alejados de la distribución.

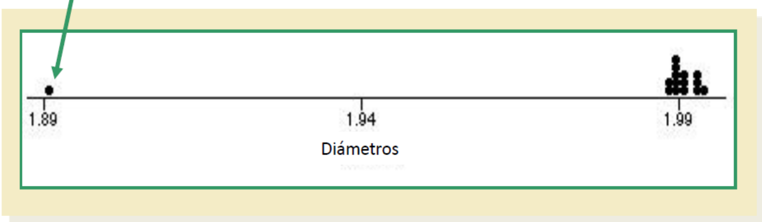


<u>¿Es usual encontrar valores atípicos o alejados?</u>

Puede haber casos en los que se presenten naturalmente y otros en los que se cometa algún error, como veremos en el siguiente **ejemplo**:

En un proceso de control de calidad se miden los diámetros de una pieza hecha por una máquina (en cm). El técnico anota 15 diámetros, pero no se da cuenta de que ha cometido un error de escritura al anotar el segundo dato.

1,991	1,891	1,991	1,988	1,993	1,989	1,990	1,988
1,988	1,993	1,991	1,989	1,989	1,993	1,990	1,994

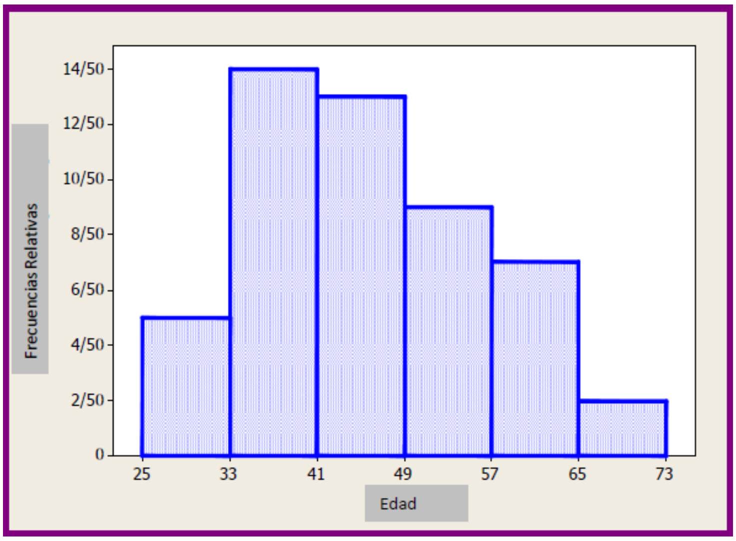


> ¿Qué consecuencias puede traer la NO detección de valores atípicos?

Más adelante, luego de estudiar algunas medidas, veremos la forma de detectarlos.

Describiendo una distribución a partir de un gráfico

Cuando realicemos informes estadísticos, nos va a interesar realizar una descripción lo más exhaustiva posible. Si debemos realizarlo a partir de un gráfico, podemos observar algunas cuestiones como forma, máximo y mínimo, moda, dispersión, etc. Tomemos como ejemplo el gráfico que se muestra a la derecha que muestra edad de empleados de una empresa X:



¿Forma? Asimétrica a la derecha

¿Tiene valores atípicos? NO. Los datos incluso se ven bastante concentrados.

¿Qué proporción de empleados son menores de 41 años? $\frac{5}{50} + \frac{14}{50} = \frac{19}{50} = 0{,}38$

¿Entre qué valores se concentra la mayoría de los datos? Sumamos las frecuencias de los intervalos con mayor cantidad de observaciones, $(14 + 13)/50 = 27/50 = 0{,}54$. De esta manera observamos que supera el 50% por lo tanto nos podemos referir a la mayoría. Vemos entonces, que la mayoría tiene entre 33 y 49 años.

Máximo y mínimo: La mayor edad observada es 73 y la menor 25 (aproximadamente)

Diagrama de tallo y hojas

El diagrama de tallo y hojas consiste en un rápido, simple e informativo modo de organizar los datos (para conjunto pequeño de datos).

Para construirlo, se separa cada observación en un tallo que contenga todos los dígitos a excepción del último de la derecha y la hoja, que será este último dígito.

EJEMPLO: Consideremos las edades de 11 personas dadas en años:

21 – 32 – 34 – 34 – 42 – 44 – 46 – 48 – 52 – 56 – 64

Podemos agrupar los datos de la siguiente manera:

Tallo	Hoja
2	1
3	244
4	2468
5	26
6	4

Unidad: dato.1 ⟶ $\text{Unidad} = \frac{\text{dígito real}}{(\text{tallo} \cup \text{hoja})}$

Pasos a seguir para la construcción del diagrama:

1. Ordenar los datos de menor a mayor.
2. Calcular el número máximo de líneas en las que se puede extender el gráfico, mediante la siguiente expresión:

Nº máx. línea = [10.log(n)] (para muestras entre 20 y 300 datos)

3. Dividir cada dato obtenido en dos partes: tallo y hojas. El tallo puede tener como máximo tres dígitos y se colocan verticalmente en forma creciente. Las hojas pueden tener sólo un dígito y se ubican en el tallo correspondiente, de menor a mayor.
4. Dado que en algunos casos es necesario truncar los datos, siempre se debe especificar la unidad en la que están expresadas los valores del diagrama.

Algunas observaciones:

• Si en un tallo no aparece ninguna hoja, se lo debe escribir igual, dejando vacío el espacio de las hojas.

• Este resumen muestra la forma general, variación y rango de la distribución.

• También muestra todos los valores sin agrupar, los valores atípicos (si se observan 3 o 4 tallos sin hojas entre algún dato y el resto de la distribución), mínimo y máximo.

• Es rápido y fácil de construir.

• Se puede construir un diagrama doble que sirva para comparar grupos. (uno izquierda y otro a derecha)

Ejemplo de construcción de un diagrama de tallo y hoja:

Los siguientes datos son las edades de 50 empleados de la UCSF. Resumir los por medio de un diagrama de tallo y hojas.

34 48 70 63 52 52 35 50 37 43 53 43 52 44
42 31 36 48 43 26 58 62 49 34 48 53 39 45
34 59 34 66 40 59 36 41 35 36 62 34 38 28
43 50 30 43 32 44 58 53

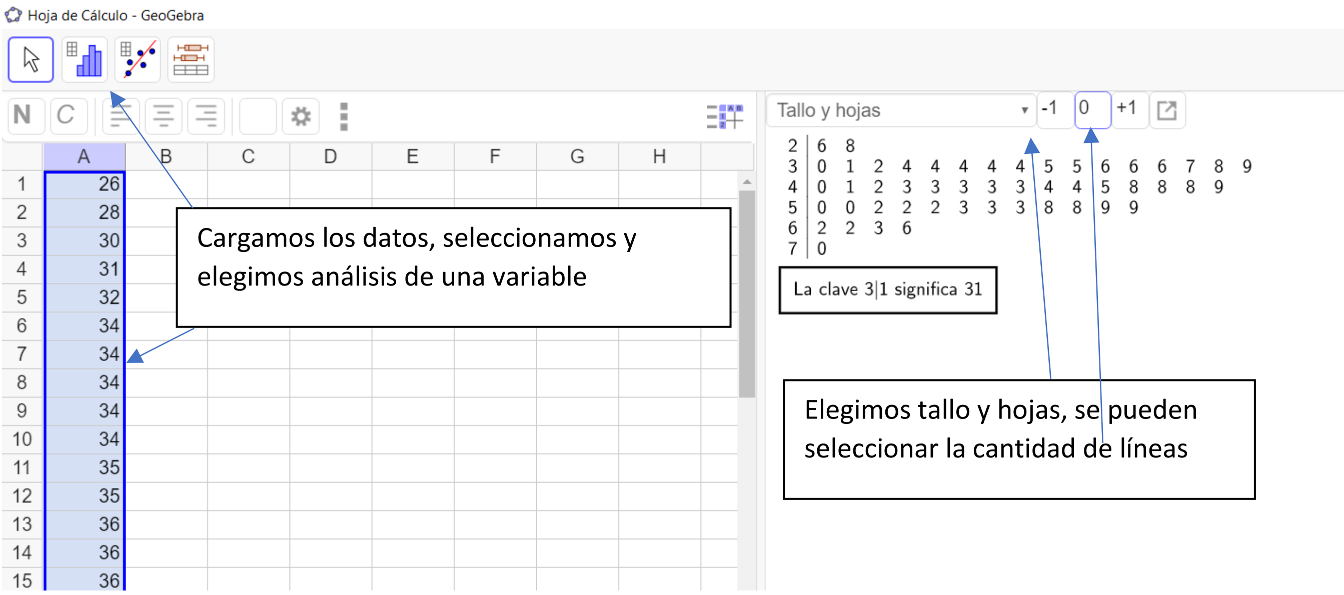
Ordenamos los datos: 26 28 30 31 32 34 34 34 34 34 35 35 36 36 36 37 38 39 40 41 42 43 43 43 43 43 44 44 45 48 48 48 49 50 50 52 52 52 53 53 53 58 58 59 59 62 62 63 66 70

Calculamos el Nº máximo de líneas: [10.log50]=[16,989]=16líneas
Tallo menor: 2 hojas del tallo menor: 6 y 8
Tallo mayor: 7 hojas del tallo mayor: 0
Construimos el gráfico resultando la siguiente representación:

```
2 | 6 8
3 | 0 1 2 4 4 4 4 4 5 5 6 6 6 7 8 9
4 | 0 1 2 3 3 3 3 3 4 4 5 8 8 8 9
5 | 0 0 2 2 2 3 3 3 8 8 9 9
6 | 2 2 3 6
7 | 0
Unidad: dato.1
```

No olvidar de citar la unidad que se usó

En GeoGebra:



Medidas descriptivas

Las medidas descriptivas se obtienen a partir del análisis de una muestra. En casi todos los casos se pueden calcular para variables cuantitativas (con algunas excepciones que ya veremos).
La importancia de éstas radica en que, si sabemos interpretarlas bien, nos brindan información muy valiosa respecto del comportamiento de la muestra.
Estudiaremos:

- **Medidas de tendencia central:** media ($\bar{x}$), mediana (Me) y modo (Mo).
- **Medidas de dispersión:** rango (R), desviación estándar (s), varianza (s^2) y coeficiente de variación (CV).
- **Medidas de posición:** cuartiles (Qi), deciles (Di) y percentiles (Pi).
- **Medidas de asimetría:** coeficiente de Bowley ($C_A B$) y coeficiente de Pearson ($C_A P$).

Medidas de tendencia central

Son medidas estadísticas que se usan para describir como se puede resumir la localización de los datos de la muestra. Ubican e identifican el punto alrededor del cual se centran los datos y nos indican hacia donde se inclinan o se agrupan más los mismos. Las más utilizadas son: la media, la mediana y la moda.

MEDIA O PROMEDIO ARITMETICO: Es la medida de tendencia central más conocida y utilizada, la misma no es otra cosa que el resultado de dividir la sumatoria de todos los valores que toma una variable entre el número de ellos (n), por lo tanto, este parámetro depende en gran medida de cada uno de los valores que forman la serie de datos.

Su fórmula es, en el caso de datos no agrupados: $\bar{x} = \frac{1}{n}.\sum_i x_i$

Hay que señalar, que esta medida de tendencia central se encuentra afectada por valores extremos de la distribución, es decir, aquellos que están situados lejos del centro de la distribución, y por lo tanto no es recomendable como medida representativa, en distribuciones en las cuales los datos se encuentran muy dispersos.

Ejemplo de cálculo de la media para datos sin agrupar:

Sean las calificaciones obtenidas por un grupo de 50 estudiantes en un examen con una escala de 1 a 20 puntos: 6 – 12 – 20 – 17 – 3 – 6 – 8 – 10 – 10 – 10 – 9 – 14 – 15 – 16 – 18 – 17 – 13 – 11 – 11 – 9 – 9 – 4 – 7 – 14 – 13 – 12 – 10 – 9 – 13 – 15 – 19 – 5 – 19 – 20 – 3 - 5 – 7 – 11 – 18 – 17 - 6 - 7 – 8 – 14 – 12 – 18 – 10 – 9 – 14 – 16.

En este caso, si sumamos todos los valores de la variable y dividimos por el total de datos que es 50, obtenemos: $\bar{x} = 11{,}58$.

Esto significa que el promedio de calificaciones del examen es 11, 58 puntos.

En el caso en que los datos estén agrupados en datos simples en una tabla de frecuencia, la fórmula puede reescribirse de la siguiente manera:

$\bar{x} = \frac{1}{n}.\sum_i x_i.f_a$ donde f_a es la frecuencia absoluta.

En este sentido, recordemos que la frecuencia absoluta contabiliza la cantidad de repeticiones de un valor determinado de la variable. Veámoslo en el ejemplo anterior:

xi	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
fa	0	0	2	1	2	3	3	2	5	5	3	3	3	4	2	2	3	3	2	2

En este caso:

$$\bar{x} = \frac{1.0 + 2.0 + 3.2 + 4.1 + 5.2 + 6.3 + 7.3 + 8.2 + 9.5 + 10.5 + 11.3 + 12.3 + 14.3 + 15.2 + 16.2 + 17.3 + 18.3 + 19.2 + 20.2}{50}$$

$$= 11{,}58.$$

Ejemplo de cálculo de la media para datos agrupados en intervalos de clase:

Veamos ahora el caso en que los datos se presenten en una tabla de frecuencias, pero ya dispuestos en intervalos de clase, por ejemplo, la siguiente situación:

Intervalos	fa	Pm – puntos medios
[2,5 – 5,5)	5	4
[5,5 – 8,5)	8	7
[8,5 – 11,5)	13	10
[11,5 – 14,5)	10	13
[14,5 – 17,5)	7	16
[17,5 - 20,5)	7	19

En este caso, como la fórmula de promedio no puede aplicarse sobre intervalos, necesitamos calcular los puntos medios, que se encuentran en la tercera columna y que se obtienen sumando los dos extremos del intervalo y dividiendo por dos.

Podemos calcular el promedio con la fórmula de la diapositiva anterior: $\bar{x} = \frac{1}{n}.\sum_i p_{mi}.f_a$ donde p_{mi} es el punto medio de cada intervalo, es decir, la marca de clase.

Algunas cuestiones a tener en cuenta:

- El promedio no puede calcularse en el caso de variables cualitativas
- En el caso de una muestra con valores extremo, si bien puede calcularse, no resulta una medida confiable a la hora de describir los datos
- En los dos primeros casos, la media se obtuvo de manera exacta en el último caso, si los datos ya vienen dados agrupados en intervalos, el promedio es una aproximación al promedio real ya que no se conoce la muestra original ni tampoco podría reconstruirse a partir de la tabla.

Se puede calcular con la calculadora. A continuación, dejamos el tutorial con las dos calculadoras más comunes:

https://www.youtube.com/watch?v=qguhqq0xvM0&ab_channel=Andr%C3%A9sArcos

https://www.youtube.com/watch?v=sRiMUgfxzwo&ab_channel=Andr%C3%A9sArcos

MEDIANA: La mediana está representada por el valor central de una distribución, y es definida como el valor que divide a la muestra en dos partes iguales. Al igual que el promedio puede obtenerse para casos sin agrupar o agrupados. Para su cálculo, los datos deben estar ordenados de menor a mayor. Es una medida muy utilizada en aquellos casos donde el promedio no resulta confiable por las razones que ya mencionamos anteriormente.

Datos sin agrupar:

Se tienen las siguientes calificaciones: 8- 10- 9- 7- 7- 5- 6- 11- 11- 12.

Las ordenamos de menor a mayor: 5- 6- 7- 7- 8- 9- 10- 11- 11- 12.

Buscamos la posición con la fórmula: $Posición\ mediana = \frac{n+1}{2} = \frac{10+1}{2} = 5{,}5$

Ubicamos dicha posición en la lista de datos:

$$5- \ 6- \ 7- \ 7- \ \underbrace{8-9}_{\substack{entre \\ 8\ y\ 9}}- \ 10- \ 11- \ 11- \ 12$$

Como la posición no es entera, se calcula un promedio entre los dos valores que “rodean” a dicha posición. En este caso $\frac{8+9}{2} = 8{,}5$. Luego Me = 8,5.

¿Cómo interpretamos la mediana?

En el ejemplo anterior podríamos decir que el 50% de los datos de la muestra es menor a 8 5 y el restante 50% es mayor a él. Notar que si la posición hubiera sido entera, directamente la mediana es un valor de la muestra y no es necesario realizar ningún promedio entre dos observaciones.

Datos agrupados en intervalos: para el cálculo procedemos de la siguiente manera:

1) Se divide $n/2$ para obtener la posición de la mediana, este valor se busca en las frecuencias acumuladas (Fa). Si no aparece, ubicamos la primera frecuencia acumulada que contenga al valor. De esa manera identificamos el intervalo donde se encuentra la mediana.

2) Se determina el límite inferior del intervalo, la frecuencia acumulada del intervalo anterior y la frecuencia absoluta del intervalo actual. Todos estos datos los aplicaremos en la siguiente fórmula:

$$M_e = L_i + \frac{\frac{n}{2} - Fa\ (ant)}{fa}.A$$

donde Li es el límite inferior del intervalo que contiene a la mediana, $Fa(ant)$ es la frecuencia acumulada del intervalo anterior, fa es la frecuencia absoluta del intervalo actual y A es el ancho del intervalo.

Calculemos la mediana, retomando el **ejemplo** ya visto:

$$Posición\ mediana = \frac{n}{2} = \frac{50}{2} = 25$$

Intervalos	fa	Fa
[2,5 – 5,5)	5	5
[5,5 – 8,5)	8	13
[8,5 – 11,5)	13	26
[11,5 – 14,5)	10	36
[14,5 – 17,5)	7	43
[17,5 - 20,5)	7	50

El intervalo que contiene a la mediana será [8,5 – 11,5). Aplicamos la fórmula:

$$M_e = 8{,}5 + \frac{25 - 13}{13}.3 = 11{,}26$$

- *Para pensar: ¿Puede la mediana calcularse en el caso en que la variable de estudio sea cualitativa? De ser posible, describir en qué condiciones puede obtenerse.*

MODO O MODA: La moda o modo está representada por el o los valores de una distribución que tienen la mayor frecuencia absoluta. Si los datos no están agrupados se obtiene por simple observación, ya que será el valor más recurrente en la distribución.

Retomando el ejemplo de las calificaciones de los 50 estudiantes:

xi	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
fa	0	0	2	1	2	3	3	2	5	5	3	3	3	4	2	2	3	3	2	2

Observar que hay dos calificaciones, 9 y 10, con la máxima frecuencia absoluta. Por lo tanto, la variable tiene dos modos en este caso.

En el caso en que los datos estén agrupados en una tabla de frecuencias, ya no hablaremos de modo sino de **INTERVALO MODAL**, que será aquel con mayor frecuencia absoluta. Por ejemplo:

Intervalos	fa	Fa
[2,5 – 5,5)	5	5
[5,5 – 8,5)	8	13
[8,5 – 11,5)	$\boxed{13}$	26
[11,5 – 14,5)	10	36
[14,5 – 17,5)	7	43
[17,5 - 20,5)	7	50

En este caso el intervalo modal es [8,5 11,5) que es el que tiene mayor frecuencia igual a 13.

Para pensar: ¿Puede el modo calcularse en el caso en que la variable de estudio sea cualitativa? Fundamentar tu respuesta.

Medidas de dispersión

Las medidas de dispersión, también llamadas medidas de variación nos determinan la variabilidad de una muestra indicando mediante un número, si las diferentes puntuaciones de una variable están muy alejadas de la media o no. Cuanto mayor sea el valor de cualquiera de las medidas de dispersión, mayor será la variabilidad de los datos y por tanto la muestra se denomina heterogénea. En el caso contrario, en que estas medidas sean más pequeñas, la muestra será homogénea y las observaciones estarán cercanas al valor del promedio.

Al calcular el promedio, es común que éste vaya acompañado por alguna medida de dispersión, ya que la media en sí no nos brinda información, por ejemplo, de la existencia de valores extremos, mientras que las medidas de dispersión sí lo hacen. De aquí la importancia de estos parámetros ya que de alguna manera evalúan la confiabilidad o no de las medidas de tendencia central.

Resumiendo, si las medidas de dispersión son pequeñas, las medidas de tendencia central son válidas y representativas de la muestra, en caso contrario no son tan confiables.

RANGO (R): Es la diferencia entre los dos valores extremos de la variable, esto es: $R = máx - mín$.

Es la medida de dispersión más sencilla de calcular y también la que menos información proporciona, pues solo se calcula teniendo en cuenta el valor de dos observaciones de la muestra, lo cual puede provocar un análisis inadecuado. Informa sobre el recorrido de la variable al registrar de alguna manera la distancia entre el mínimo y el máximo de la variable.

Ejemplo

Retomemos nuevamente el ejemplo de las calificaciones de los 50 estudiantes. La muestra es:

6 12 20 17 3 6 8 10 10 10 9 14 15 16 18 17 13 11 11 9 9 4 7 14 13 12 10 9 13 15 19 5 19 20 3 5 7 11 18 17 6 7 8 14 12 18 10 9 14 16.

Claramente, el mínimo es 3 y el máximo es 20. Luego $R = 20 - 3 = 17$.

DESVIACIÓN ESTÁNDAR (s): También llamada desviación típica, siempre será una medida positiva, aunque en algunos casos ya veremos que puede valer 0. Esta medida nos permite averiguar el promedio aritmético de fluctuación de los datos respecto a la media aritmética. Si bien existe una fórmula matemática para calcularla, es una de las medidas que se puede obtener usando el modo estadístico de cualquier calculadora científica. En los enlaces de video de la página 32 encontrarán como realizar en calculadora, además de la media, la desviación.
En el video se habla de desviación poblacional y desviación muestral, en este curso siempre trabajaremos con la desviación muestral.
Luego de ver el video, calculen la desviación estándar para esta pequeña muestra, donde se muestran las edades en años de los niños en la sala de espera de un consultorio pediátrico:

5 10 12 4 3 1 1 8

Deberías obtener $s = 4{,}1057$
La unidad de la desviación estándar coincide con la unidad de la variable estadística que se esté estudiando. Esto es, por ejemplo, en el caso anterior, la desviación estándar es de 4,1057 años.

> *Para pensar: ¿En qué situación, de acuerdo con la definición que vimos de la desviación estándar, consideras que este parámetro puede valer cero? ¿Por qué?*

VARIANZA (s^2): Es otra medida de dispersión, que calcula el promedio de las desviaciones de la media, elevadas al cuadrado. De manera más sencilla, podemos obtenerla elevando al cuadrado la desviación estándar s.
Por ejemplo, si s es 4,1057 años en el ejemplo de la variable que representaba las edades de los niños en un consultorio pediátrico la varianza será en este caso:

$$s^2 = (4{,}1057\ años)^2 = 16{,}856\ años^2$$

Naturalmente, la unidad de medición para la varianza es la unidad de la variable al cuadrado, algo que pueden observar fácilmente a partir del cálculo anterior.
La información que proporcionan la desviación estándar y la varianza es la misma desde un punto de vista práctico, aunque es más sencillo analizar la desviación estándar ya que está expresada en la misma unidad que la variable.

COEFICIENTE DE VARIACIÓN (CV): Es una medida que proporciona información sobre la dispersión relativa del conjunto de datos y se calcula como el cociente entre la desviación estándar y su media aritmética. En la mayoría de los casos este resultado se multiplica por 100 para expresarlo como un porcentaje.

$$CV = \frac{S}{\overline{X}} . 100\%$$

Puesto que tanto la desviación estándar como la media "heredan" las unidades de medición de la variable, el CV es una medida **independiente** de las unidades originales de los datos. Debido a la propiedad anterior, el CV es el parámetro adecuado para comparar la variabilidad de dos muestras que podrían estar medidas en diferentes unidades.
Ejemplo: Supongamos que conocemos la media y la desviación estándar de dos muestras distintas. En ambas se estudia el peso de un grupo de personas que siguen un plan de alimentación determinado.
Muestra 1: Los datos se miden en kg, y se sabe que $\bar{x} = 75{,}5kg \quad y \quad s = 3{,}5kg$.
Luego:

$$CV_1 = \frac{3{,}5\ \cancel{kg}}{75{,}5\ \cancel{kg}} . 100\% = 0{,}046.100\% = 4{,}6\%$$

Muestra 2: Los datos se miden en g, y se sabe que $\bar{x} = 54300\ g \quad y \quad s = 4800g$.

$$CV_1 = \frac{4800\ \cancel{g}}{54300\ \cancel{g}}.100\% = \ 0{,}088.100\% = 8{,}8\%$$

Aquí debemos concluir que la muestra más heterogénea o con mayor variabilidad es la segunda ya que posee un coeficiente de variación más elevado.

Lo mismo sucede si comparamos dos muestras que tienen distintas medias y desviaciones, aun cuando estén en la misma unidad. El CV en estos casos es más confiable a la hora de tomar una decisión.
En la mayoría de los casos, esta medida será menor a 100 aunque los valores por encima de 100 indicarán que estamos en presencia de una muestra con una muy elevada dispersión.

- *Ejercicio para pensar: Dos laboratorios de medicamentos tienen maneras diferentes de pagar a sus visitadores médicos La compañía A lo hace mediante un sueldo fijo mensual y la compañía B mediante un porcentaje sobre las ventas efectuadas La distribución de los salarios (en pesos) es la siguiente:*

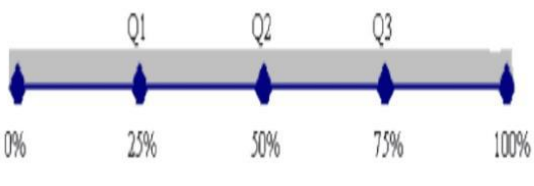
Laboratorio A: 42000 35000 54000 74300 29200 43780 95000 120320 48500
Laboratorio B: 10000 8500 2500 25300 7500 4800 9700 2900 6500
Usando las medidas que creas conveniente, determinar ¿en qué laboratorio el sueldo medio es más equitativo? ¿Por qué?

Medidas de posición

Son valores que permiten dividir a la muestra ordenada en partes con igual número de observaciones o datos en cada una de ellas.
Se usan para describir la posición que tiene un dato especifico en relación con el resto de los datos.
Las medidas de posición más utilizadas son:
Cuartiles (Qi): dividen a la muestra en cuatro partes iguales (con la misma cantidad de datos).
Deciles (Di): dividen a la muestra en diez partes iguales (con la misma cantidad de datos).
Percentiles (Pi): dividen a la muestra en cien partes iguales (con la misma cantidad de datos).
Veremos en detalle cómo podemos calcular cada una de ellas y su interpretación en el contexto en el cual está definida la variable.

Cuartiles (Qi)

Son valores que dividen al lote de datos (ordenado de menor a mayor) en cuatro grupos iguales con la misma cantidad de elementos.
Luego existen tres cuartiles: Q1, Q2 y Q3. Gráficamente:



A partir del gráfico puede observarse que:

a) El primer cuartil Q1 es el valor cuya frecuencia relativa porcentual acumulada es 25%. Esto quiere decir que el 25% de los datos en la muestra son menores a este valor.

b) El segundo cuartil Q2 coincide con la mediana, pues es el valor cuya frecuencia relativa porcentual acumulada es 50%. Por lo tanto, no se considera relevante ya que no aporta información nueva.

c) El tercer cuartil Q3 es el valor cuya frecuencia relativa porcentual acumulada es 75%. Esto es, el 75% de los datos en la muestra es inferior a esta medida.

Cálculo de los cuartiles:

Como ya hemos mencionado, la muestra debe estar ordenada de menor a mayor y sólo será de interés calcular Q1 y Q3.

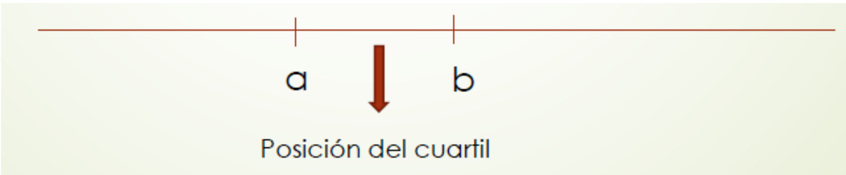
Obtenemos la posición de cada uno de ellos a través de las fórmulas:

$$Pos(Q1) = \frac{n+1}{4} \qquad\qquad Pos(Q3) = \frac{n+1}{4}.3$$

Nuevamente como ocurría en el caso de la mediana, esta posición podría ser entera o no.

En el caso en que sea entera (caso más sencillo), ubicamos dicha posición en la muestra ordenada y ése será el valor de Q1 o Q3.

En el peor de los casos la posición puede ser decimal. Es decir, “caer” entre dos valores de la muestra, que llamaremos **a** y **b**. Gráficamente:



En este caso, aplicamos la siguiente fórmula para obtener Q1 o Q3:

$$Q1, Q3 \;=\; a \;+\; parte\ decimal\ de\ la\ posición\,.\,(b - \;a)$$

Veamos algunos ejemplos

1- Consideremos la muestra: 3- 3- 5- 6- 6- 7- 8. Por simpleza, la muestra ya está ordenada (si no habría que ordenarla). Calculemos las posiciones de ambos cuartiles:

$$Pos(Q1) = \frac{7+1}{4} = 2 \qquad\qquad Pos(Q3) = \frac{7+1}{4}.3 = 6$$

En la muestra, la posición 2 está ocupada por el 3 y la posición 6 corresponde al 7. Luego Q1 = 3 y Q3 = 7. Esto significa que el 25% de los datos son menores a 3 y que el 75% de los datos son menores a 7.

Tomemos ahora la muestra: 3- 3- 5- 6- 6- 7- 8- 9. En este caso:

$$Pos(Q1) = \frac{8+1}{4} = 2{,}25 \qquad\qquad Pos(Q3) = \frac{8+1}{4}.3 = 6{,}75$$

Luego para obtener los cuartiles debemos aplicar la fórmula antes mencionada:

$$Q \;=\; 3 + \;0{,}25.\,(5 \;-\; 3) = \;3{,}5 \qquad\qquad Q3 \;=\; 7 \;+\; 0{,}75.\,(8 \;-\; 7) \;=\; 7{,}75$$

Esto significa que el 25% de los datos será inferior a 3,5 y el 75% de los datos es menor a 7,75.

Por lo general, en el contexto de un problema, no se les pedirá específicamente que calculen los cuartiles, sino que se formulará alguna pregunta respecto de la muestra que pueda responderse a través de estas medidas.

2- El tiempo que transcurre entre la finalización de la presentación de un chiste y el momento en que una persona comienza a reírse se denomina tiempo de reacción. En este contexto, la presentación del chiste es un estímulo y la aparición de la risa, la reacción. Se hizo una experiencia, con un determinado grupo, en el que se midió el tiempo de reacción de sus integrantes ante un chiste y se registraron los siguientes datos en décimas de segundos (ds):

29 34 26 31 38 35 36 32 34 33 30

Responder

1. ¿Cuál es el tiempo de reacción máximo por debajo del cual se encuentra el 25% del grupo?
2. Se decide seleccionar aquellos sujetos cuyos tiempos de reacción se ubican en el 25% de tiempos más altos. ¿Por encima de qué tiempo debo elegir estos individuos? ¿Cuántos son?

Solución: En el ítem (a) se nos pide determinar el tiempo de reacción máximo por debajo del cual se encuentra el 25% de los datos. Claramente se trata de Q1.

- Ordenamos la muestra de menor a mayor:

  26- 29- 30- 31- 32- 33- 34- 34- 35- 36- 38

y calculamos la posición de Q1: $\frac{n+1}{4}.1 = \frac{11+1}{4}.1 = \frac{12}{4}.1 = 3.1 = 3.$

- En la muestra, la posición 3 corresponde al tiempo 30ds. Esto significa que el 25 % de las personas del grupo tuvo un tiempo de reacción inferior a 30 décimas de segundo.
- En el ítem (b) se nos pide determinar el tiempo por encima del cual se deben seleccionar los individuos cuyos tiempos de reacción corresponden al 25% de tiempo más altos, esto nos deja un 75% en la parte inferior, por lo tanto, debemos calcular entonces Q3:

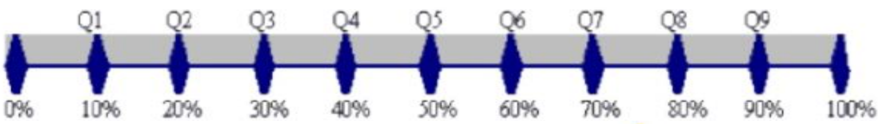
  En la muestra, la posición de Q3: $\frac{n+1}{4}.3 = \frac{11+1}{4}.3 = \frac{12}{4}.3 = 3.3 = 9$ corresponde al tiempo 35 décimas de segundo. Esto significa que se escogerán 3 individuos que son los que tienen tiempos de reacción mayores 35 (o iguales). (35,36 y 38)

IMPORTANTE: Dado que estas medidas son límites entre los datos que quedan por debajo o por arriba del mismo, vamos a considerar siempre incluir este límite en las respuestas. Como el caso del 35 en el ejemplo anterior.

Para pensar y debatir: ¿Para qué tipos de variables es posible calcular los cuartiles? ¿Por qué?

Deciles (Di) y Percentiles (Pi)

- Similarmente al caso de los cuartiles, los deciles y percentiles dividen a la muestra en 10 y 100 partes iguales respectivamente. A continuación, se muestra la representación gráfica de los deciles, que son 9:



- No se muestra una representación para los percentiles debido a que son 99. Pero cada tramo en este caso corresponde al 1% de los datos.
- **Para calcular deciles o percentiles el procedimiento es similar al del cálculo de cuartiles:**
  - La muestra se ordena de menor a mayor.

- Se calcula la posición con la fórmula:

$$Pos\left(Di\right)=\frac{n+1}{10}.\,i \qquad Pos(Pj)=\frac{n+1}{100}.\,j, \qquad i=1,\ldots,9;\ j=1,\ldots,99$$

Los índices i y j de las fórmulas anteriores tomarán los valores que correspondan de acuerdo con la medida que desee obtenerse. Esto es, si quiero calcular el decil 6, luego i = 6 y la posición será:

$$Pos(D6)=\frac{n+1}{10}.\,6$$

- De igual manera para obtener los percentiles.
- Nuevamente, aquí puede ocurrir que la posición sea entera, con lo cual sólo hay que identificar el valor en la muestra que corresponda a dicha posición. O puede pasar que sea una posición decimal, caso en el cual se aplica la misma fórmula que en el caso de los cuartiles.

Por ejemplo, tomemos nuevamente la muestra anterior: 26- 29- 30- 31- 32- 33- 34- 34- 35- 36- 38. Queremos calcular el D4 y P52:

$$Pos(D4)=\frac{11+1}{10}.4=4{,}8 \qquad y \qquad Pos(P52)=\frac{11+1}{100}.52=6{,}24$$

Luego **D4 = 31 + 0,8. (32 – 31) = 31,8** y **P52 = 33 + 0,24. (34 – 33) = 33, 24**

Esto quiere decir que el 40% de los datos será inferior a 31,8 y el 52% de los datos será inferior a 33,24

Para pensar y debatir:

1) ¿Para qué tipos de variables es posible calcular los deciles y percentiles? ¿Por qué?

2) Si de los 50 sujetos que realizan un examen, 40 de ellos obtienen una puntuación superior a 8 quiere decir que P20 = 8.

3) Si mediante una prueba hemos seleccionado a 20 de los 50 aspirantes a un puesto laboral, los cuales han obtenido una puntuación superior a 5, podemos afirmar que P60 = 5.

4) A continuación aparecen las puntuaciones en el factor extraversión – introversión, obtenidas para 20 sujetos:

4- 7- 5- 4- 3- 6- 6- 8- 9 -8 -6- 4- 3- 1- 1- 1-4 -7- 8- 9

Para un estudio posterior se van a seleccionar aquellos individuos cuyas puntuaciones correspondan al 50% central de la muestra. ¿Cuántos individuos se eligen y entre qué rango de puntuación?

Medidas de asimetría

Como vimos, podemos analizar la estructura de las distribuciones viendo su centro, su dispersión y su simetría. ¿Cómo podemos detectar la simetría y la asimetría numéricamente?

COEFICIENTE DE ASIMETRÍA DE PEARSON:

$C_AP=\frac{\bar{x}-M_O}{s}$ donde $\bar{x}$ es la media, s la desviación y $\mathrm{M_O}$ la moda

COEFICIENTE DE ASIMETRÍA DE BOWLEY:

$C_AB=\frac{Q_3+Q_1-2M_e}{Q_3-Q_1}$ donde Q_1 es el cuartil 1, Q_3 el cuartil 3 y $\mathrm{M_e}$ la mediana.

- Si el coeficiente es *cero* la distribución es *Simétrica*
- Si el coeficiente es *positivo (y mayor que 2)*, la distribución es *asimétrica hacia la derecha* (sesgada a la derecha)
- Si el coeficiente es *negativo (y menor que -2)*, la distribución es *asimétrica hacia la izquierda* (sesgada a la izquierda)

Notar que en el caso en que tengamos una distribución bimodal no es conveniente calcular el coeficiente de asimetría de Pearson, pues hay más de un modo en la muestra. En ese caso, se debe calcular el segundo coeficiente.
Cuánto más elevado sea el resultado de cualquiera de los dos coeficientes, la distribución será más asimétrica. Por el contrario, si el valor oscila en el intervalo [-2, 2], diremos que la distribución es aproximadamente simétrica.

Ejemplo
En una encuesta a 300 jefes y jefas de familia de la ciudad de Santa Fe se les preguntó sobre la cantidad de hijos que tenían y resultaron los siguientes datos:

Cantidad de hijos	0	1	2	3	4	5	6	TOTAL
Cantidad de familias	78	90	32	25	50	15	10	300

Coeficiente de asimetría de Pearson: $C_{AP} = \frac{\bar{X}-M_O}{s}$, $\mathrm{M_O}$:moda, $\bar{X}$: media, y s desviación

$$\bar{X} = \frac{0.78 + 1.90 + 2.32 + 3.25 + 4.50 + 5.15 + 6.10}{300} = \frac{0 + 90 + 64 + 75 + 200 + 75 + 60}{300} = \frac{564}{300} = 1{,}88$$

Por otra parte, la frecuencia más grande (n° de familias) es 90, y corresponde al valor de la variable 1, $\mathrm{M_O} = 1$ y $\mathrm{s} = 1{,}7519$

$$C_{AP} = \frac{1{,}88 - 1}{1{,}7519} = 0{,}5023$$

El coeficiente es positivo, pero menor a 2, por lo tanto, la distribución es aproximadamente simétrica o es levemente asimétrica a la derecha.

Coeficiente de asimetría de Bowley: $C_{AB} = \frac{Q_3+Q_1-2M_e}{Q_3-Q_1}$

Calculemos el Q_1, Q_3 y la mediana:

Posición Q_1: $\frac{301}{4} = 75{,}25$

Posición M_e: $2.\frac{301}{4} = 150{,}5$

Posición Q_3: $3.\frac{301}{4} = 225{,}75$

Variable	fa	FAA
0	78	78
1	90	168
2	32	200
3	25	225
4	50	275
5	15	290
6	10	300
TOTAL	300	

Posición Q_1

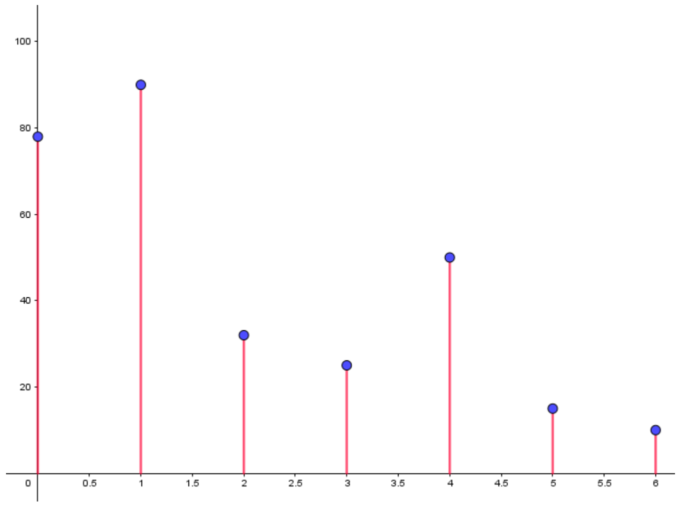
Posición M_e

Posición Q_3

$Q_1 = 0$, $M_e = 1$ y $Q_3 = 3 + 0{,}75.(4-3) = 3{,}75$

$$C_{AB} = \frac{Q_3 + Q_1 - 2M_e}{Q_3 - Q_1} = \frac{3{,}75 + 0 - 2.1}{3{,}75 - 0} = \frac{1{,}75}{3{,}75} = 0{,}46666\ldots$$

Es positivo y menor a 2, por lo tanto es levemente asimétrica a derecha o aproximadamente simétrica.
Veamos esto gráficamente:



Levemente asimétrica a derecha

o aproximadamente simétrica

Diagrama de caja y bigote

Este diagrama fue propuesto por Tukey y se obtiene a partir del ***resumen de los cinco números***: cuartiles 1 y 3, mediana, el valor máximo y mínimo. (y los valores atípicos).

Ya hemos definido los valores atípicos, ahora veremos como determinarlos.

¿Cómo detectamos los valores atípicos?

Lo que debemos hacer, es determinar los límites de las zonas de existencia de valores atípicos, es decir, números que nos determinen que antes o después de ellos, los valores se considerarán atípicos.

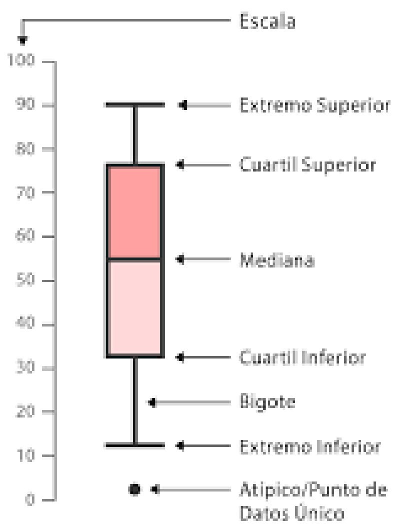
Valores atípicos por exceso

- $Q_3 + 1{,}5.(Q_3 - Q_1)$ Valores alejados (si hay datos por arriba de este valor se consideran atípicos)
- $Q_3 + 3.(Q_3 - Q_1)$ Valores muy alejados

Valores atípicos por defecto

- $Q_1 - 1{,}5.(Q_3 - Q_1)$ Valores alejados (si hay datos por debajo de este valor se consideran atípicos)
- $Q_1 - 3.(Q_3 - Q_1)$ Valores muy alejados

Diagrama de caja y bigote:



(También se puede hacer horizontal).

Cada tramo entre cuartiles contiene el 25% de los datos.

Ejemplo:

Se ha sugerido agua corriente como método de acondicionamiento cardiovascular para atletas lesionados y otros que deseen un programa de aerobics de bajo impacto. Un estudio publicado en la Journal of Sports Medicine investigó la relación entre la cadencia de ejercicio y la frecuencia cardiaca, al medir las frecuencias cardiacas de 20 voluntarios sanos a una cadencia de 48 ciclos por minuto (un ciclo formado por dos pasos). Los datos aparecen a continuación:

87 109 79 80 96 95 90 92 96 98 101 91 78 112 94 98 94 107 81 96

Son 20 datos, calculemos los cuartiles:

Posición $Q_1: \frac{21}{4} = 5{,}25$

Posición $M_e: 2.\frac{21}{4} = 10{,}5$

Posición $Q_3: 3.\frac{21}{4} = 15{,}75$

Ordenemos los datos de menor a mayor:

78 79 80 81 87 90 91 92 94 94 95 96 96 96 98 98 101 107 109 112

Posición Q_1 Posición M_e Posición Q_3

$Q_1 = 87 + 0{,}25.(90 - 87) = 87{,}75$

$M_e = 94 + 0{,}50.(95 - 94) = 94{,}50$

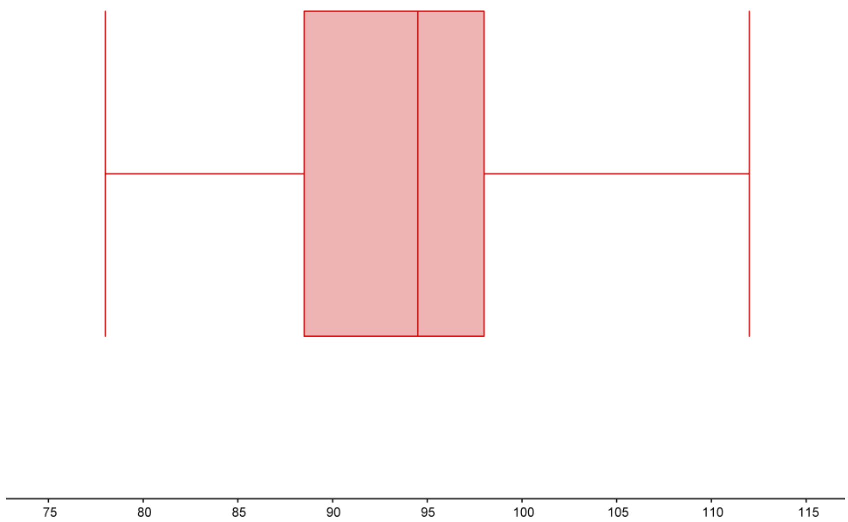
$Q_3 = 98 + 0{,}75.(98 - 98) = 98$

Veamos si existen valores atípicos:

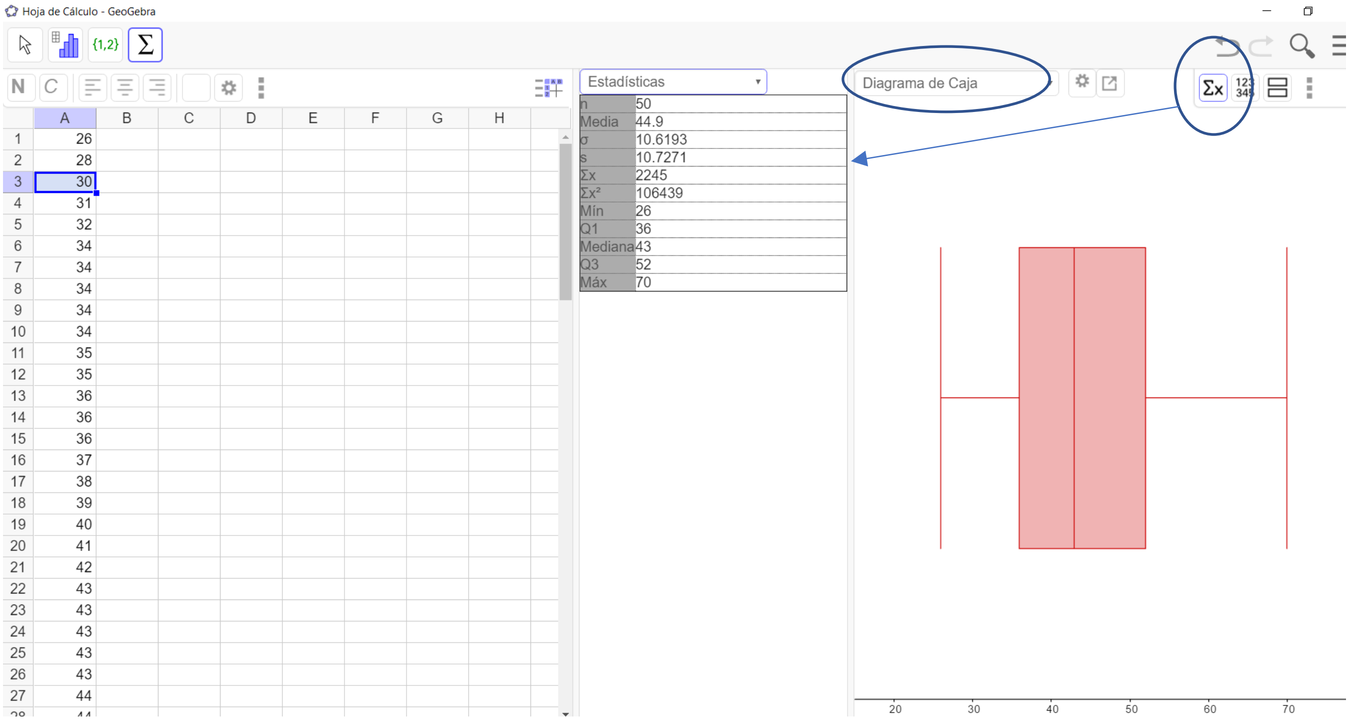
$98 + 1{,}5.(98 - 87{,}75) = 113{,}375$ Este valor es el límite para detectar valores atípicos por exceso, el valor más alto de la distribución es 112, que no supera dicho límite. Veamos que pasa por defecto:

$87{,}75 - 1{,}5.(98 - 87{,}75) = 72{,}375$ Este valor es el límite para detectar valores atípicos por defecto, el valor más bajo de la distribución es 78, que no está por debajo de dicho límite. Conclusión: no existen valores atípicos.

El diagrama de caja y bigote correspondiente es el siguiente:



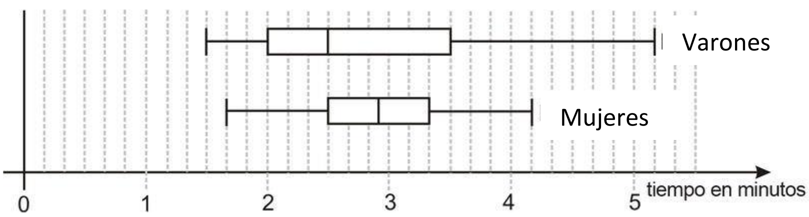
En GeoGebra, podemos obtener medidas antes mencionadas y hacer el diagrama de cajas, veamos para un lote de datos (el que vimos en el diagrama de tallos y hojas):



Comparación entre distribuciones a partir del diagrama de caja y bigote

Este diagrama se puede utilizar para comparar dos distribuciones en una misma escala.

Ejemplo: Los diagramas de caja y bigotes mostrados abajo representan los tiempos hechos por cada alumno de una clase durante una carrera de 150 yardas con obstáculos. Los tiempos se han separado según género, es decir, tanto para el grupo de varones como para el de mujeres. Cada uno de dichos grupos cree que posee los mejores tiempos. Determinar el resumen de cinco números para cada grupo. Proporcionar un argumento convincente para cada grupo.



Solución

La comparación de los dos conjuntos de datos mediante diagramas de caja y bigotes es relativamente sencilla. Por ejemplo, podemos observar, tanto a través del rango como del rango intercuartílico $(Q_3 - Q_1)$, que los datos para los varones están más esparcidos.

El resumen de los cinco números para cada grupo es mostrado en la tabla que sigue.

	Varones	Mujeres
Menor Valor	1:30	1:40

	Varones	Mujeres
Primer Cuartil	2:00	2:30
Mediana	2:30	2:55
Tercer Cuartil	3:30	3:20
Mayor Valor	5:10	4:10

Aunque conviene aclarar que cada juego deportivo necesita tener un conjunto de reglas para evitar confusiones sobre quién gana, cada grupo podría usar los siguientes argumentos a su favor.

Varones

Los varones poseen la marca de tiempo más corto (1 minutos 30 segundos), de donde se concluye que el individuo más rápido fue un varón.

Los varones también tienen la mediana más pequeña (2 min 30 segundos); lo que significa que la mitad de los varones habían terminado la carrera cuando solamente un cuarto de las mujeres lo había hecho (sabemos que únicamente un cuarto de las mujeres había terminado porque su primer cuartil fue también de 2:30).

Mujeres

Los varones tuvieron el tiempo más largo (5 minutos 10 segundos), de modo que para cuando todas las mujeres ya habían terminado, aún quedaba, al menos, un varón sin completar la carrera.

Las mujeres tuvieron el tercer cuartil más pequeño. (3 min 20 segundos); lo que significa que aun sin tomar en cuenta el cuarto de tiempos más largos, las mujeres fueron las más rápidas.

PRÁCTICA 2

En todos los ejercicios que puedas, aplicar GeoGebra para verificar tus respuestas

Actividad 1: Se han medido los niveles de ozono alrededor de Madrid. Las concentraciones de magnitud alrededor de 220 unidades por billón pueden ocasionar quemaduras en los ojos y son peligrosas para la vida animal. Se han obtenido los siguientes datos:

160 176 160 180 167 164 165 163 162 168 173 179 170 196 185 163 162 163 172 162 167 161 169 178 161

a) Diseñar un gráfico de tallos y hojas para estos datos, especificando la unidad correspondiente.
b) Realizar una tabla de distribución de frecuencias y redactar un breve informe a partir de ella y utilizando también el diagrama realizado en el ítem anterior.

Actividad 2: Los datos que figuran a continuación corresponden a los puntos anotados por dos jugadores de baloncesto a lo largo de 30 partidos.

Jugador A: 34, 11, 19, 13, 20, 20, 33, 43, 22, 30, 25, 17, 28, 21, 22, 40, 40, 27, 41, 39, 39, 31, 18, 27, 35, 33, 35, 21, 19, 44.

Jugador B: 37, 1, 10, 2, 12, 13, 34, 54, 16, 31, 21, 6, 28, 15, 16, 48, 48, 25, 51, 47, 47, 32, 8, 25, 38, 35, 38, 13, 11, 56.

a) Realizar el gráfico que creas conveniente para comparar ambos grupos.
b) Realizar un diagrama de tallo y hojas comparativo y destacar las características más sobresalientes de ambos grupos a partir de él.

Actividad 3: Se realiza un estudio para ayudar a comprender el efecto del fumar en los patrones del sueño. La variable aleatoria considerada es X, el tiempo en minutos que se tarda en quedarse dormido. Las muestras de fumadores y no fumadores producen estas observaciones sobre X:

NO FUMADORES	FUMADORES
17 19 18 15 18 17	15 20 17 21 16 24
16 19 19 23 24 20	16 21 18 22 15 25
19 22 20 24 25 21	22 22 19 25 18 25
21 18 22 20 23 20	25 24 15 24 21 16
21 16 23 20 17 21	24 25 15 18 23 17
21 22 21 20 20 20	23 25 16 17 24 19
19 18 19 22 19 17	15 15 19 23 23 25

a) Construir un diagrama de tallos y hojas comparativo para ambos grupos.
b) Es recomendable el diagrama de tallos y hojas cuando los datos tienen una amplia gama de valores?. Justifique.

Actividad 4: El siguiente diagrama de tallo y hoja corresponde al porcentaje de algodón en un material utilizado para la fabricación de camisas para caballeros. La unidad correspondiente es 0.1. Reconstruir la muestra original y realizar el gráfico que te parezca más adecuado para representar los datos.

Tallo	**Hojas**
32	156789
33	114566666688
34	01112235566666777779
35	00111234456789
36	234888
37	13689

Actividad 5: i) Realiza, para cada lote de datos, un gráfico que sea acorde a los datos recolectados.
ii) Sin obtener ninguna medida, ordena a tu criterio las siguientes distribuciones de acuerdo al nivel de variabilidad que representan (la primera es la menos variable y la última es la más variable).

a- 10 8 6 0 8 3 2 5 8 0
b- 1 3 3 5 5 5 7 7 9
c- 20 1 2 5 4 4 4 0
d- 5 5 5 5 5 5 5 5 5

iii) Calcular las medidas de tendencia central y la desviación estándar de cada una y verificar si el orden que diste en el ítem anterior es correcto. Analizar a qué se deben las diferencias encontradas entre las medidas.

Actividad 6: Un profesor desea averiguar el nivel de conocimiento de literatura de sus dos grupos de alumnos A y B para planificar sus clases de acuerdo con dicho nivel. Para ello elige al azar a 15 alumnos de cada grupo y les aplica un test elaborado para tal fin, obteniendo los resultados siguientes:

Grupo A	4	3	7	5	6	4	5	4	5	6	7	7	3	4	5
Grupo B	8	9	1	2	8	8	4	3	2	2	10	7	8	2	1

a) Analizando cada conjunto de datos, responde: En cuál de los dos grupos le será más fácil la planificación de la asignatura? Justifica tu respuesta.
b) Qué medidas utilizarías para responder a la pregunta del ítem a)? Justifica tu elección y luego calcúlalas y verifica si la respuesta dada en a) coincide o no con lo obtenido numéricamente.

Actividad 7: En las siguientes situaciones ¿qué medida de tendencia central crees que es más adecuada para resumir los datos?

a) La municipalidad de un pueblo está estudiando la posibilidad de aplicar un impuesto sobre la renta a sus habitantes. Para ello, quiere conocer el promedio de renta de estos ciudadanos y tomarlo como referencia para el cálculo de la base impositiva.

b) Un psicólogo desea estimar el nivel de estrés promedio de los obreros de la construcción sometidos a situaciones de riesgo. Con dicho objetivo, aplica un test a los obreros de una empresa constructora.

Actividad 8: La edad a la que comenzaron a fumar los individuos de una muestra extraída de una población de fumadores se detalla a continuación en función del sexo:

Edad	Hombres	Mujeres
(10 – 15]	6	8
(15 – 18]	12	21
(18 – 20]	26	35
(20 – 25]	34	6
(25 – 30]	9	4
(30 – 40]	4	2
Total	**91**	**76**

a) Representa las frecuencias relativas de cada sexo en un mismo gráfico que permita realizar comparaciones. Basándote en el gráfico expresa tus conclusiones en relación con las características principales de cada grupo.
b) Calcula el promedio y la desviación estándar en cada uno de los grupos y sacar conclusiones.

Actividad 9: Si la edad media de las mujeres que han contraído matrimonio en un determinado juzgado en los últimos 20 días es 26 con una desviación típica de 0, ¿cuál es la edad de la mujer que se ha casado en último lugar? ¿Cuál es la edad de la más joven?

Actividad 10: Retomando los datos de la actividad 3, responda:
a) ¿Cuál de los dos grupos (fumadores y no fumadores) es más variable? Justifique.
b) Encuentra la media y la desviación estándar de la muestra para cada grupo y ver si tienen relación con la respuesta del ítem anterior.
c) Comenta qué clase de impacto parece tener el fumar sobre el tiempo que se requiere para quedar dormido.

Actividad 11: Las notas en un examen de acceso a la universidad en ciertos estudios con excesiva demanda y poca salida laboral siguen una distribución con una media de 6.3 y rango intercuartílico de 1.6. Si dicha facultad sólo quiere admitir el 25% de las solicitudes con notas más altas y sabiendo que la distribución es simétrica, ¿cuál debe ser la nota mínima para acceder a estos estudios?

Actividad 12: Los datos que se presentan a continuación detallan el porcentaje de familias que vive por debajo del umbral de pobreza en 20 países del Tercer Mundo:

20	46	65	38	52	67	29	36	45	51
53	64	61	60	55	45	39	45	66	23

a) Realiza el diagrama de tallo y hojas y el diagrama de caja correspondiente.

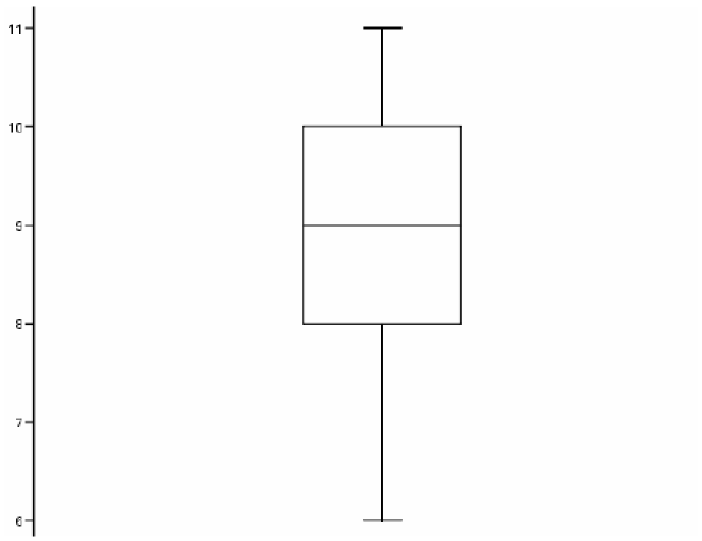
b) ¿Hay valores alejados en esta distribución? ¿Cuáles son?
c) Basándote en los diagramas, ¿se podría decir que la distribución es simétrica?
d) Calcular las medidas de tendencia central y de dispersión adecuadas.
e) Calcular el coeficiente de asimetría e interpretarlo.

Actividad 13: Los siguientes datos representan los tiempos de reacción frente a un determinado estímulo que se aplicó en un experimento a una muestra aleatoria de 14 ratas de laboratorio:

5.4 - 1.3 - 4.1 - 2.9 - 0.7 - 2.6 - 8.5 - 0.6 - 7.1 - 1.9 - 8.2 - 3.9 - 14.8 - 9.8

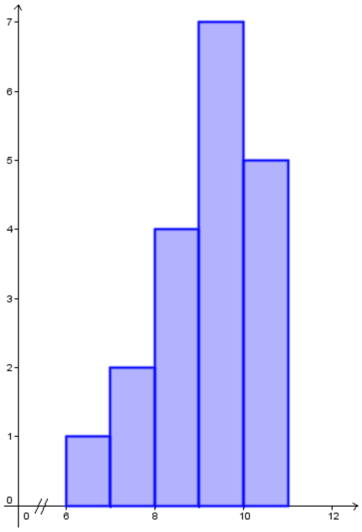
a) Traza un diagrama de tallo y hojas.
b) Realiza el resumen de los cinco números y el diagrama de caja.
c) ¿Es asimétrica esta distribución? Justificar.
d) ¿Cuál de los dos gráficos representa mejor la información en este caso? ¿Por qué?
e) ¿Sería adecuada una tabla de frecuencia con intervalos? ¿Por qué?

Actividad 14: El siguiente diagrama de caja corresponde a una muestra aleatoria de 20 niños con desórdenes de conducta anotando el tiempo necesario en horas que requirió para lograr un plan integral de tratamiento con cada uno de ellos:



a) Realiza un breve informe a partir de lo que observas en el diagrama.
b) ¿Hay valores atípicos en esta distribución?
c) Interpreta los valores de los cuartiles y la mediana en el contexto del problema.
d) ¿Qué se puede decir respecto de la simetría de la distribución?

Actividad 15: i) El tratamiento de los niños con desórdenes de la conducta puede ser complejo. El mismo se puede proveer en una variedad de escenarios dependiendo de la severidad de los comportamientos. Además del reto que ofrece el tratamiento, se encuentran la falta de cooperación del niño/niña y el miedo y la falta de confianza de los adultos. Para poder diseñar un plan integral de tratamiento, los psicólogos de niños y adolescentes puede utilizar la información del niño, la familia, los profesores y de otros especialistas médicos para entender las causas del desorden. Para ello, un psicólogo local ha considerado una muestra aleatoria de 19 niños, anotando el tiempo, medido en horas, necesario que requiere en cada niño para lograr un plan integral del tratamiento, obteniéndose lo siguiente:



Decidir si son verdaderas o falsas las afirmaciones. Justificar en todos los casos.

a) Se puede afirmar que, en promedio, el tiempo requerido, en cada niño, para lograr un plan integral del tratamiento es de 7 horas.

b) El 50% inferior de los tiempos observados es de 9 horas.

c) Podríamos decir que la distribución de los tiempos necesarios que requiere en cada niño para lograr un plan integral del tratamiento es aproximadamente simétrica.

ii) En el colegio X de Buenos Aires, se está realizando un experimento para comprobar si realmente la enseñanza de las matemáticas por computadora supera la enseñanza tradicional, en cuanto a resultados inmediatos de los niños. Dicho experimento se realizó con los estudiantes de cuarto grado de primaria y ocurrió así: se tomó un grupo control formado por 17 de esos niños, elegidos de su curso aleatoriamente, y se les enseñó un cierto tema a la manera tradicional. También se tomó un grupo experimental formado por otros 17 niños, elegidos del mismo curso en forma aleatoria y se les enseñó el mismo tema con asistencia de la computadora. Cuando terminaron de ver ese tema, se les hizo a los 34 alumnos el mismo examen. Las calificaciones obtenidas por los niños se presentan a continuación:

Grupo control 36 35 36 41 41 36 32 29 28 40 33 40 40 38 40 30 41
Grupo Experimental 40 38 38 44 44 40 36 39 38 40 38 44 44 38 40 38 40

a. ¿Cuál es la variable en estudio?
b. ¿Cuál de las medidas de tendencia central podría describir el comportamiento de la variable en cada uno de los dos grupos? ¿Sería representativa? Justifique.
c. Cuál de las dos muestras es más dispersa?
d. ¿Cree usted que las calificaciones obtenidas en los dos grupos corroboran la hipótesis de que la enseñanza por computadora supera en resultados inmediatos a la enseñanza tradicional? Explique su respuesta.

PRÁCTICA COMPLEMENTARIA A LAS PRÁCTICAS 1 Y 2

En todos los ejercicios que puedas, aplicar GeoGebra para verificar tus respuestas

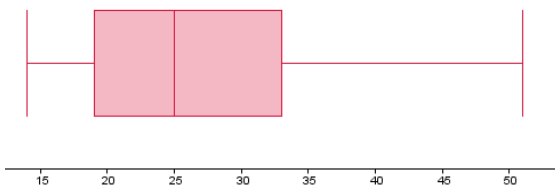
<u>Actividad 1.</u> A continuación se muestran las edades de motociclistas en el momento en que resultaron mortalmente heridos en accidentes de tránsito (según datos del U. S. Department of Transportation).

Edades de motociclistas en el momento en el que resultaron mortalmente heridos en accidentes de tránsito

1	4 6 7 7 8 8 9
2	0 1 3 3 4 5 5 7 8 8 9
3	0 1 3 4 7 8
4	0 2
5	1

La tecla 3|1 significa 31

Edades de motociclistas en el momento en el que resultaron mortalmente heridos en accidentes de tránsito



a) Definir la variable de estudio y clasificarla. ¿A partir de cuál de los diagramas es posible reconstruir la muestra original? Explica.

b) En base a los dos esquemas, realiza un breve informe estadístico que describa las características principales de la muestra.

c) Responde verdadero o falso a las siguientes aseveraciones:

i) Se encuestaron a 30 motociclistas.

ii) La distribución de datos es aproximadamente simétrica.

iii) Las edades correspondientes al 75% superior de los datos son mayores a 33 años.

iv) El 50%central de los datos varía entre las edades de 20 y 35.

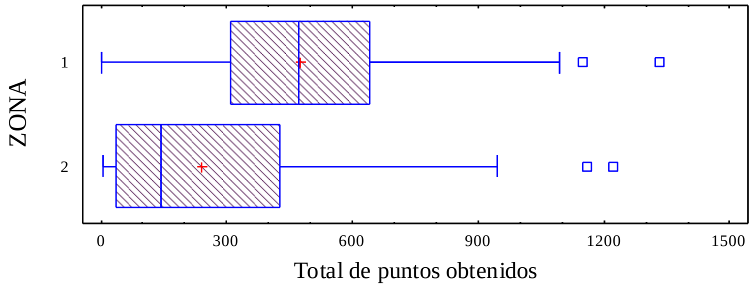
<u>Actividad 2.</u> A continuación se presentan los diagramas de caja y resúmenes numéricos obtenidos a partir de datos reales que han sido tomados de una muestra de alumnos de varias Escuelas de Educación Media de la provincia de Santa Fe. Estas escuelas han sido agrupadas de acuerdo a la zona en la que se encuentran, correspondiendo la **zona 1**, a aquellas <u>escuelas ubicadas en ciudades</u>, y la **zona 2**, a <u>escuelas ubicadas en zonas rurales</u>. A estos alumnos se les ha aplicado un test de que mide diversas aptitudes académicas y la variable representada en los diagramas es el puntaje total obtenido en dicho test. <u>Utilizando las representaciones y resúmenes numéricos que siguen</u>, responde a las siguientes preguntas:

a. ¿En cuál de las dos zonas, los alumnos han obtenido puntuaciones menos variables? Justifica tu respuesta indicando en base a qué medidas realizas tu conclusión.

b. Si tuvieras que elegir uno de estos grupos para representar a la provincia en un encuentro a nivel nacional, ¿Cuál de las dos zonas elegirías? ¿Por qué?

Resumen Estadístico para TOTAL (Zona 1)	**Resumen Estadístico para TOTAL (Zona 2)**
Tamaño de la muestra: 145 alumnos	*Tamaño de la muestra*: 177 alumnos
Mediana = 470	Mediana*:* 142
Media = 474,08	Media = 240,89
Desviación estándar = 263,25	Desviación estándar = 250,94
Cuartil 1 = 310 Cuartil 3 = 642	Cuartil 1 = 35 Cuartil 3 = 427
Mínimo = 0 Máximo = 1333	Mínimo = 6 Máximo = 1221

Diagramas de caja del Total obtenido en función de la procedencia del alumno



Actividad 3. a) Se le ha preguntado a un grupo de personas respecto de la cantidad de hijos que cada uno tenía. La siguiente es la tabla de frecuencias correspondiente a dicha muestra, pero se encuentra incompleta. Completarla y calcular: media, mediana, modo, desviación estándar y coeficiente de variación. Confeccionar el gráfico que consideres apropiado para exponer los datos y realiza un breve informe sobre los resultados.

Xi	fa	Fa	fr
1	4		0.08
2	4		
3		16	0.16
4	7		0.14
5	5	28	
6		38	
7	7	45	0.14
8			
	n =		

b) A tres grupos de niños se les enseña inglés con tres métodos diferentes A, B y C. Al finalizar el curso las puntuaciones fueron las siguientes:

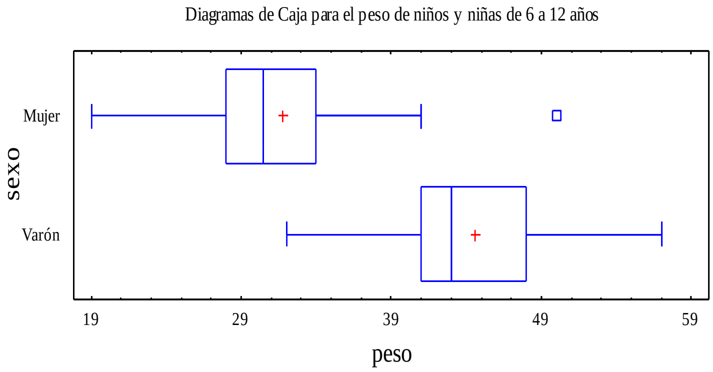
A	1	1	1	3	1	2
B	1	1	2	2	5	1
C	8	6	1	9	5	1

¿Son los tres métodos igualmente eficaces? Justifica. De ser posible escoja el mejor de los métodos.

c) Explicar detalladamente la veracidad o no de las siguientes afirmaciones: “La media está menos afectada que la mediana por la existencia de valores extremos en la muestra”; “La mediana puede calcularse para variables cualitativas en todos los casos”.

Actividad 4. Los siguientes diagramas de caja presentan los datos referentes a los pesos de 30 niñas y 30 niños cuyas edades oscilan entre 6 y 12 años. De acuerdo a las normas dadas por la Organización Mundial de la Salud, el peso normal para las niñas de estas edades, deberían variar entre 17 y 43 kg respectivamente, y para los niños deberían variar entre 18 y 40 kg. Aquellos niños o niñas que tengan pesos inferiores a los definidos como normales se los considera en riesgo de desnutrición y los que tengan pesos superiores a los normales, se considera que tienen sobrepeso y riesgo a ser obesos.

Basándote exclusivamente en la información que puedes extraer de los diagramas de caja presentados, responde a las siguientes preguntas:



a. ¿Se puede afirmar que hay niñas que sufren desnutrición o sobrepeso? ¿Y en el grupo de los niños? Justifica tus respuestas.

b. ¿Cuál es el peso máximo que tiene el 50 % con pesos más bajos en las niñas?

c. ¿Entre qué valores varía el 50 % central de los pesos de los varones?

d. ¿Cuál de los dos grupos podría considerarse que tiene pesos que se acercan más a los normales? Justifica tu elección.

e. ¿Podríamos afirmar que ambas distribuciones son aproximadamente simétricas? ¿Por qué?

<u>Actividad 5.</u> En un grupo de estudiantes se considera el número de ensayos que necesita cada uno para memorizar una lista de seis pares de palabras. Los resultados fueron:

5 8 3 9 6 7 10 6 7 4 6 9 5 6 7 9 4 6 8 7

a) Construir la tabla de frecuencias.

b) Calcular las medidas de tendencia central e interpretar cada una. Razonar si en este caso sería representativo el promedio del grupo de datos.

c) ¿Cuántos ensayos precisa como mínimo el 15% superior de la muestra?

d) Aquellos estudiantes cuya cantidad de ensayos quede "por fuera" del 50% central de los datos, serán sometidos a otros tests. Determinar el rango de ensayos para los cuales esto ocurrirá.

e) Comentar rasgos generales de la muestra en cuanto a simetría y dispersión.

f) Un grupo de 20 actores fue sometido a la misma experiencia que los estudiantes mencionados arriba. Para ellos resultó una media de 4,8 y un desvío de 1,8. En base a los resúmenes estadísticos adecuados señalar cuál es el grupo de mejor desempeño en la experiencia realizada. Justificar su respuesta.

<u>Actividad 6.</u> Una compañía está estudiando el número de veces por día que es utilizado su cajero automático localizado en un supermercado. A continuación, se muestra el número de veces que fue usado durante los últimos 30 días:

83, 63, 95, 64, 80, 36, 84, 84, 78, 76, 73, 61, 84, 68, 59, 54, 52, 84, 75, 65, 95, 59, 90, 47, 70, 52, 87, 61, 77, 60

a) Desarrolla un diagrama de tallo y hojas.

b) En promedio, ¿cuántas veces al día fue usado el cajero?

c) ¿Cuáles son el menor y el mayor número de veces que fue usado? Calcula el número de veces por encima del cual fue usado el 75% de las veces.

d) Caracteriza la simetría.

e) ¿Existen valores atípicos? Justificar.

Actividad 7. Los niños, a diferencia de los adultos, tienden a recordar las películas, cuentos e historias como una sucesión de acciones más que el argumento en forma global y de conjunto. En el relato de una película, por ejemplo, utilizan con frecuencia las palabras "y entonces...". Una psicóloga con suprema paciencia pidió a 50 niños que le contaran una determinada película que ellos habían visto. Consideró la variable: cantidad de "y entonces..." utilizados en el relato y registró los siguientes datos:
8 15 22 19 15 17 18 20 17 12 16 16 17 21 23 18 20 21 20 20 15 18 17 19 20 23 22 10 17 19 19 21 20 18 18 24 11 19 31 16 17 18 19 20 18 18 40 18 19 16
Como parte del mismo estudio la experimentadora obtuvo de 50 adultos el mismo tipo de datos. Estos fueron:
10 12 5 8 13 10 12 8 7 9 11 10 9 9 11 15 12 17 14 10 9 8 15 16 10 14 7 16 9 1 4 11 12 7 9 10 3 11 14 8 12 5 10 9 7 11 14 10 15 9

Para ambas variables:

a) Construya la tabla de frecuencias.

b) Calcule la media, la mediana y la moda.

c) ¿Qué gráficos harías? (No realice)

d) Indique en cuál grupo los integrantes son más parecidos en cuanto a la cantidad de “y entonces…” utilizados en el relato de una película. Justifique su respuesta.

Actividad 8. Los datos siguientes corresponden al número de ordenadores portátiles vendidos trimestralmente por distintas tiendas informáticas de una franquicia:

12	18	35	35	36	36	37	37	38	38	38	38
38	39	39	39	39	40	40	41	41	42	42	47

a) La franquicia quiere premiar al 25% de las tiendas que mas ventas tengan en portátiles. ¿Cuál será el número de portátiles a vender por una tienda para beneficiarse de este premio?

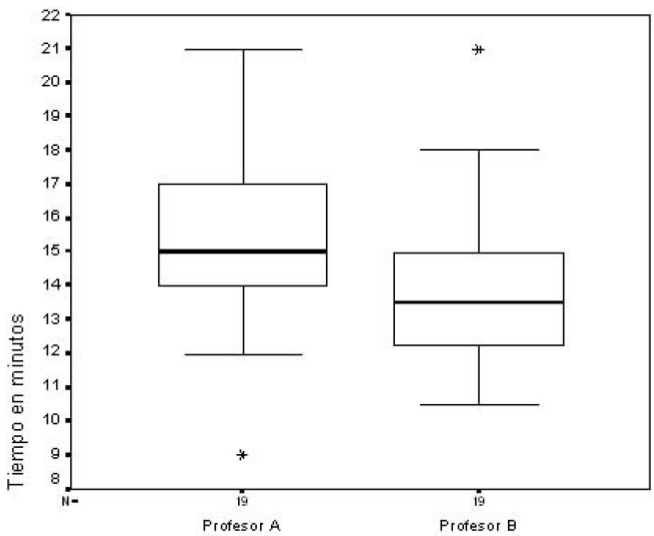
b) Por algún motivo se desea advertir al 25% de las tiendas que menos venden respecto de mejorar esta realidad. ¿Por debajo de que valor de ventas se encuentran las tiendas en cuestión?

c) ¿Es simétrica la distribución? Justifique.

d) Si tuvieras que resumir la información en un gráfico, cual elegirías y por qué.

Actividad 9. Dos profesores (A y B) están interesados en estudiar los hábitos de sueño de los estudiantes en sus clases. Ambos profesores registran el tiempo (en minutos) que demoran en quedarse dormidos sus alumnos desde que empieza la clase. El gráfico muestra los tiempos que demoran en quedarse dormidos los alumnos de los profesores A y B.

Realice un informe completo de la información proporcionada por los diagramas.



Actividad 10. El siguiente diagrama de tallo y hoja corresponde a las edades de actores varones y mujeres de Hollywood, responder verdadero o falso justificando apropiadamente en cada caso:

Edades de actrices	Tallo	Edades de actores
1 3 4 4 4 5 6	2	2 4 6 9
1 8 9	3	1 1 1 3 3
2 3 8 9 9	4	0 0 2 2 7
1 1 1 5 5	5	3 7 7
0 1 1 2 2 9	6	1 1 3 8 8 9
2 2 4 4	7	5 5 6 6 7 9
0 1	8	3 3 5 6

Unidad: 1

a. La muestra presenta las edades de 25 actores y 30 actrices.
b. El diagrama muestra la existencia de valores extremos.
c. El 50% de las edades de las actrices están por debajo de 51 años.
d. El 25% de los actores son mayores a 68 años.
e. La moda para la muestra de actrices es 24 años y en el caso de los actores es 31 años.

UNIDAD 4: DISTRIBUCIONES DE PROBABILIDAD.

Probabilidad

La probabilidad se define como la posibilidad de ocurrencia de cualquier evento o suceso.

EJEMPLO:

1. Probabilidad de obtener un 6 al lanzar un dado.
2. Probabilidad de encontrar una familia que tenga 8 hijos en la ciudad de Santa Fe.
3. Probabilidad de éxito al aplicar un nuevo tratamiento.

Distintos tipos de probabilidad

Hay distintos tipos de probabilidad, según las características del experimento que se esté realizando:

- **PROBABILIDAD CLÁSICA O A PRIORI:**

$$p = \frac{\text{n° casos favorables}}{\text{n° casos posibles}}$$

En este tipo de probabilidad, conocemos previamente que todos los casos tienen la misma probabilidad de aparecer, es decir son equiprobables.

En el primer ejemplo anterior, conocemos de antemano qué probabilidad hay de obtener un 6, sin necesidad de hacer el experimento aleatorio y además todos los casos del dado tienen la misma probabilidad de aparecer. En estos casos la probabilidad se obtiene haciendo:

$$p = \frac{\text{n° casos favorables}}{\text{n° casos posibles}} = \frac{1}{6}$$

- **PROBABILIDAD EMPÍRICA O A POSTERIORI:**

$$f_r = \frac{\text{frecuencia absoluta}}{\text{n}}$$

En el segundo ejemplo, la probabilidad se obtendrá a partir de una muestra de familias de Santa Fe a los que se les preguntará el número de hijos que tienen.

- **PROBABILIDAD SUBJETIVA:**

En el tercer ejemplo, no se dispone de información previa. En este caso, la probabilidad es asignada por una persona y puede variar de un individuo a otro de acuerdo a la información relacionada que tenga sobre el fenómeno en estudio.

Algunos conceptos probabilísticos básicos:

- **EXPERIMENTO ALEATORIO**: Procedimiento por medio del cual se obtiene una observación o una medida y cuyo resultado depende del azar.
  EJEMPLO: Lanzar un dado, lanzar una moneda, registrar la edad de los individuos, etc.
- **EVENTO SIMPLE**: Es el resultado que se observa en una sola repetición de un experimento aleatorio. Uno y sólo un evento simple puede ocurrir cada vez que se realiza un experimento aleatorio. Denotamos a un evento simple con la letra E_i o sencillamente con una letra mayúscula.
  EJEMPLO: Sea el experimento aleatorio “Lanzar un dado”:
  Eventos simples: 1 (E_1), 2 (E_2), 3 (E_3), 4 (E_4), 5 (E_5) y 6 (E_6).

- **ESPACIO MUESTRAL**: Es el conjunto de todos los eventos simples de un experimento aleatorio.

  EJEMPLO: Lanzamiento de un dado:

  $$S = \{1,2,3,4,5,6\}$$

  Lanzamiento de una moneda: $S = \{C, X\}$.
- **EVENTO COMPUESTO**: Es una colección de uno o más eventos simples.

  EJEMPLO: Lanzamiento de un dado.

  A: "Que salga un número impar"

  B: "Que salga un número mayor a 2"

  Entonces, $A = \{E1, E3, E5\}, B = \{E3, E4, E5, E6\}$
- **EVENTO CONJUNTO**: Es aquel que cumple con dos o más características.

  EJEMPLO: Lanzamiento de un dado

  A: "Que salga un número impar"

  B: "Que salga un número mayor que 2"

  C: "Que salga un número impar y mayor que 2", en este caso C es un evento conjunto.
- **EVENTOS MUTUAMENTE EXCLUYENTES**: Dos eventos son **mutuamente excluyentes** si, cuando uno de ellos ocurre, el otro no puede ocurrir.

  EJEMPLO: Experimento: "Lanzar un dado".

  A: "Que salga un número impar"

  B: "Que salga un número mayor que 2"

  Estos dos eventos o sucesos **NO** son mutuamente excluyentes.

  C: "Observar un 6" y D: "Observar un 3"

  Estos dos eventos son **mutuamente excluyentes**.
- **EVENTO IMPOSIBLE**: Si un evento A no puede ocurrir se denomina **evento imposible** y P(A)=0.

  EJEMPLO: Experimento: "Lanzar un dado".

  A: "Que salga 8"
- **EVENTO SEGURO**: Si el evento A ocurre cada vez que se realiza el experimento se denomina evento seguro y P(A)=1.

  EJEMPLO: Experimento: "Lanzar un dado".

  A: "Que salga un número del 1 al 6"

Observaciones:

- La probabilidad de un evento mide "con qué frecuencia" pensamos que el evento A puede ocurrir. Lo simbolizamos como P(A).
- Supongamos que se repite un experimento n veces. La frecuencia relativa para un evento A es:

$$f_r = \frac{\text{frecuencia absoluta}}{n}$$

- A medida que n aumenta, la frecuencia relativa se acerca al valor teórico de probabilidad.
- La suma de las probabilidades de todos los eventos simples de un espacio muestral S es igual a 1.
- La probabilidad de un evento es un valor entre 0 y 1 (ambos inclusive). Por lo tanto, la probabilidad NUNCA nos debe dar un valor mayor a 1.

Ejemplos: cálculo de probabilidades

EJEMPLO 1: Si se lanza una moneda, ¿cuál es la probabilidad de que salga cara?

Definamos el espacio muestral; $S = \{C, X\}$. Tenemos dos casos posibles (cara y cruz)

Sea C= "sale cara", entonces tenemos un caso posible. Luego:

$$P(C) = \frac{1}{2} = 0{,}5$$

Respuesta: La probabilidad de que salga cara es 0,5.

EJEMPLO 2: Si se lanzan dos monedas legales, ¿cuál es la probabilidad de observar al menos una cara?

Definamos el espacio muestral; $S = \{\boldsymbol{CC}, \boldsymbol{CX}, \boldsymbol{XC}, XX\}$. Tenemos 4 casos posibles.

Sea A= "sale al menos una cara", entonces tenemos 3 casos posibles (resaltados en negrita en S). Luego:

$$P(A) = P(CC) + P(CX) + P(XC) = \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}$$

Respuesta: La probabilidad de observar al menos una cara es $\frac{3}{4}$.

EJEMPLO 3: Se ha realizado una encuesta a todos los empleados de una fábrica y se les ha preguntado si son fumadores activos. Los resultados se presentan en la **tabla de contingencia** siguiente:

	Fumador	**No fumador**	**Total**
Mujer	70	50	**120**
Hombre	40	40	**80**
Total	**110**	**90**	**200**

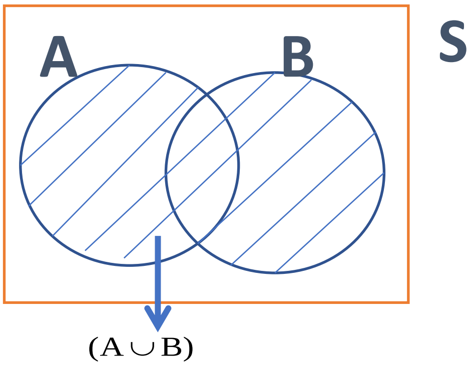
Distribución marginal por renglón (Mujer u hombre)

Distribución marginal por columna (Fumador o No fumador)

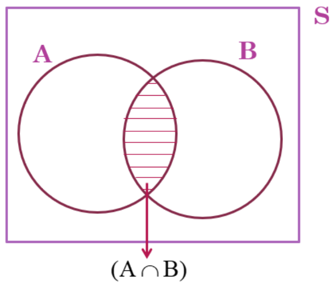
- Un evento simple: A: "La persona es mujer"
- Un evento conjunto: B: "La persona es hombre y fuma".
- El evento "mujer no fumadora" es un evento conjunto.
- Si A: "La persona fuma" entonces el complemento de A se denota como $C(A) = \bar{A} = A^C$ = "La persona no fuma".

Relaciones entre eventos

- **UNIÓN DE DOS EVENTOS**: La unión de dos eventos, A y B, es el evento en el que puede ocurrir A o B o ambos. Cuando se realiza el experimento lo simbolizamos como $(A\ o\ B) o\ (A \cup B)$

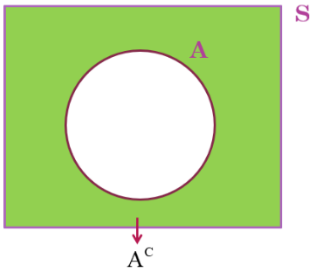


- **INTERSECCIÓN DE DOS EVENTOS**: La intersección de dos eventos, A y B, es el evento en el que A y B ocurren simultáneamente al realizar el experimento. Lo simbolizamos como (A y B) o (A∩B).



- Si los eventos A y B son **mutuamente excluyentes** entonces P(A∩B) =0.

- **COMPLEMENTO DE UN EVENTO**: El complemento de un evento A consiste en todos los resultados del experimento que no pertenecen al evento A. Lo simbolizamos como A^C.



EJEMPLOS:

Se selecciona al azar a un estudiante de una clase y se observa su color de cabello y el género.

Definimos los siguientes eventos:

A: "El estudiante tiene cabellos castaños"

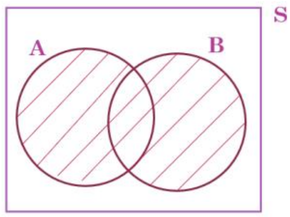
B: "El estudiante es hombre"

C: "El estudiante es mujer"

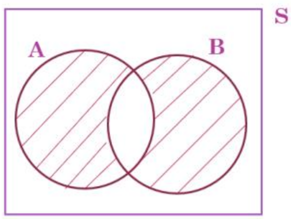
- B y C son mutuamente excluyentes.
- A^C: El estudiante no tiene cabellos castaños.
- B∩C: El estudiante es hombre y mujer. (Evento imposible)
- $B \cup C$: "El estudiante es hombre o mujer" =Todos los estudiantes=S. (Evento seguro)

Reglas de probabilidad

- **REGLA DE LA ADICIÓN:** Para dos eventos cualesquiera, A y B, la probabilidad de que uno de ellos ocurra o que ocurran ambos está dada por: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$



- **REGLA DE LA ADICIÓN:** Para dos eventos cualesquiera, A y B, la probabilidad de que uno de ellos ocurra, pero no ambos, está dada por: $P(A \cup B) = P(A) + P(B) - 2.P(A \cap B)$



EJEMPLO: Supongamos que hay 120 estudiantes en una clase, y que observamos su género y si tienen o no cabello castaño:

	Castaño	**No Castaño**	**Total**
Hombre	20	40	**60**
Mujer	30	30	**60**
Total	**50**	**70**	**120**

Se elige un estudiante al azar y se definen los siguientes eventos:

A: "La persona tiene cabello castaño". Luego, $P(A) = \frac{50}{120}$

B: "La persona es mujer". Entonces, $P(B) = \frac{60}{120}$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{50}{120} + \frac{60}{120} - \frac{30}{120} = \frac{80}{120} = \frac{2}{3}$$

<u>OBSERVACIÓN</u>: Si dos eventos son **mutuamente excluyentes**, $P(A \cap B) = 0$, luego,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = P(A) + P(B)$$

EJEMPLO: A: "la persona es hombre de cabello castaño". $P(A) = \frac{20}{120}$

B: "la persona es mujer de cabello castaño". $P(B) = \frac{30}{120}$

A y B son mutuamente excluyentes entonces $P(A \cap B) = 0$. Por lo tanto:

$$P(A \cup B) = P(A) + P(B) = \frac{20}{120} + \frac{30}{120} = \frac{50}{120} = \frac{5}{12}$$

- **REGLA DEL COMPLEMENTO:** Para cualquier evento A: $P(A \cap A^C) = 0$ dado que A y A^C son **mutuamente excluyentes**.

  $P(A \cup A^C) = 1$, dado que A o A^C pueden ocurrir.

  En efecto,

$$P(A \cup A^C) = P(A) + P(A^C) = 1$$

Entonces, $\boxed{P(A^C) = 1 - P(A)}$

EJEMPLO: Si seleccionamos al azar a un estudiante de la clase,

A: "La persona es hombre": $P(A) = \frac{60}{120} = \frac{1}{2}$

B: "La persona es mujer"

A y B son complementarios, entonces:

$$P(B) = 1 - P(A) = 1 - \frac{60}{120} = \frac{60}{120} = \frac{1}{2}$$

- **INDEPENDENCIA DE DOS EVENTOS:** Dos eventos, A y B, son **independientes** si y sólo si la probabilidad del evento B no es afectado o cambiado por la ocurrencia del evento A, o viceversa.

EJEMPLO DE INDEPENDENCIA DE DOS EVENTOS:
Se lanza un sólo dado dos veces, y se definen los eventos:

A: "Se observa un 2 en el primer lanzamiento."

B: "Se observa un 2 en el segundo lanzamiento"

Si el dado no está cargado, la probabilidad del evento A es $P(A) = \frac{1}{6}$

Consideremos la probabilidad del evento B. Sin importar si el evento A ha ocurrido o no, la probabilidad de observar un 2 en el segundo lanzamiento es aún $\frac{1}{6}$. Se podría escribir:

P(B) dado que A ocurrió $= \frac{1}{6}$

P(B) dado que A no ocurrió = $\frac{1}{6}$

Puesto que la probabilidad del evento B no cambia por la ocurrencia del evento A, se dice que A y B son **eventos independientes**.

Independencia, probabilidad condicional y regla de multiplicación

- **PROBABILIDAD CONDICIONAL:** La probabilidad de un evento A, dado que ha ocurrido el evento B, se llama **probabilidad condicional de A, dado B**, se denota por P(A|B) y está dado por:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{si } P(B) \neq 0$$

La probabilidad condicional del evento B, dado que ha ocurrido el evento A es:

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad \text{si } P(A) \neq 0$$

EJEMPLO:
Supongamos que en la población hay 51% de hombres y 49% de mujeres, y que las proporciones de hombres y mujeres daltónicos se muestran en la tabla de probabilidad siguiente:

	Hombres (B)	Mujeres (B^C)	Total
Daltónico (A)	0,04	0,002	**0,042**
No daltónico (A^C)	0,47	0,488	**0,958**
Total	**0,51**	**0,49**	**1**

a) Si se escoge una persona al azar de esta población y se encuentra que es hombre (evento B), ¿cuál es la probabilidad de que el hombre sea daltónico (evento A)?

b) ¿Cuál es la probabilidad de ser daltónico dado que la persona es mujer?

RESPUESTA:

a) Si se sabe que ha ocurrido el evento B, se debe restringir la atención al 51% de la población que es hombre. La probabilidad de ser daltónico, dado que la persona es un hombre, es:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0{,}04}{0{,}51} = 0{,}078$$

b) Ahora se está restringiendo a que el 49% de la población es mujer, entonces:

$$P(A|B^C)=\frac{P(A\cap B^C)}{P(B^C)}=\frac{0{,}002}{0{,}49}=0{,}004$$

Observe que la probabilidad del evento A cambió, lo que depende de si ocurrió el evento B. Esto indica que son dos eventos dependientes.

*Observar que en la tabla de contingencia se trabaja con proporciones, y anteriormente hemos visto tablas con la cantidad de observaciones. Ambas formas de trabajar son correctas. Lo que NO se recomienda es armar la tabla con porcentajes, ya que en ocasiones debemos multiplicar dos valores de la tabla y al hacer %. % obtenemos por 10000 (lo cual puede equivocarse a la hora de interpretar)

- **REGLA DE LA MULTIPLICACIÓN PARA LA INTERSECCIÓN:**

Para dos eventos, A y B, la probabilidad de que A y B ocurran juntos es: $P(A\cap B)=P(A).P(B|A)$

O bien $P(A\cap B)=P(B).P(A|B)$

Dos eventos, A y B, <u>son independientes</u> si y sólo si $P(A\cap B)=P(A).P(B)$

EJEMPLO: En una población, el 10% de las personas tienen un alto riesgo de sufrir un ataque al corazón. Si se seleccionan 3 personas al azar de esta población. ¿Cuál es la probabilidad de que exactamente una de ellas 3 tenga riesgo de sufrir un ataque?

A: "tener riesgo de ataque", $P(A)=0{,}1$.

N: "No tener riesgo de ataque", $P(N)=0{,}9$.

Sea B: "de las 3 personas seleccionadas, exactamente una tiene riesgo de ataque"

Como la probabilidad de tener un ataque no se relaciona con la probabilidad de que otra tenga un ataque tenemos:

$$P(B)=P(ANN)+P(NAN)+P(NNA)=(0{,}1).(0{,}9).(0{,}9)+(0{,}9).(0{,}1).(0{,}9)+(0{,}9).(0{,}9).(0{,}1)=0{,}243$$

Cuando dos eventos son independientes, es decir, si la probabilidad del evento B es la misma, si el evento A ha ocurrido o no, entonces el evento A no afecta al evento B y

$$P(B|A)=\frac{P(A\cap B)}{P(A)}=\frac{P(A).P(B)}{P(A)}=P(B)$$

<u>COMPROBACIÓN DE INDEPENDENCIA</u>

Dos eventos A y B son independientes si y sólo si $P(A\cap B)=P(A).P(B)$ $o\ si$ $P(B|A)=P(B)$.

De lo contrario, se dicen que son dependientes.

EJEMPLO: Se lanza dos monedas y se observa el resultado.

Se definen los eventos:

A: "Sale cara en la primera moneda."

B: "Sale cruz en la segunda moneda."

¿Son independientes los eventos A y B?

SOLUCIÓN: $S = \{CC, CX, XC, XX\}$

$$P(A) = \frac{1}{2} \quad Y \quad P(B) = \frac{1}{2}$$

Puesto que $P(A).P(B) = \frac{1}{2}.\frac{1}{2} = \frac{1}{4}$ y $P(A \cap B) = \frac{1}{4}$

Entonces, $P(A).P(B) = P(A \cap B)$ y los dos eventos **son independientes.**

Distribuciones de probabilidad

Variable aleatoria

Denominaremos *variable aleatoria* (de aquí en adelante VA) a una función que asigna a cada elemento del espacio muestral un número real. Denotaremos las VA con letras mayúsculas.
Si X es un VA, y A es un elemento de S entonces X(A) es un número real.

Al ser X una función, a todo elemento de S le corresponde uno y solo un elemento de $\mathbb{R}(reales)$.
El conjunto imagen de X, denominado W, está formado por todos los valores que puede tomar la VA y se lo denomina espacio muestral sustitutivo.

Ejemplo: Si un experimento consiste en tomar una pareja al azar de una comunidad, el espacio muestral S está formado por todos los posibles resultados del experimento, en este caso por todas las parejas de la comunidad; téngase en cuenta que, aunque sólo se va a seleccionar una pareja, puede ser elegida cualquiera. Si la variable de interés para el experimento es el número de hijos de la pareja seleccionada, se asigna un número a cada elemento de S, por tanto, se ha definido una función de S en $\mathbb{R}$, ya que cada pareja le corresponde uno y sólo uno, que es el de hijos que tienen.
Supóngase que en la comunidad donde se lleva a cabo el experimento el máximo número de hijos que tiene una pareja es 6, en este caso $W = \{0, 1, 2, 3, 4, 5, 6\}$, es el conjunto de valores que puede tomar la VA X (X= número de hijos). Al elegir una pareja la única característica de interés es el número de hijos y el conjunto W es el espacio muestral sustitutivo.

En un mismo espacio muestral es posible definir varias VA; por ejemplo, para el caso anterior podemos definir el experimento de elegir una pareja al azar y observar la edad promedio. Supongamos que la pareja más joven de la comunidad tiene una edad promedio de 18 años la más longeva 75. Definimos $W' = \{18, \dots, 75\}$.

Para pensar en clase:
1) Consideramos un experimento aleatorio de lanzar una moneda al aire tres veces y anotamos el resultado (cara o cruz). Se define la variable aleatoria X como número de caras aparecidas en los tres lanzamientos. Calcular el espacio muestral y comprobar que es una variable aleatoria.
2) Consideramos un experimento aleatorio de lanzar dos dados. Definir tres VA que pueden ser de interés observar en el mismo.
3) Determinar si las siguientes son o no variables aleatorias. En aquellas que si son, escribir un posible experimento asociado a las mismas:
a) Medida de presión sistólica.
b) Color de ojos.

c) Cantidad de números pares.
d) Número de caras.
e) Tipo de patología.

Distribución de probabilidad de variables aleatorias discretas

Si una variable aleatoria X definida sobre un espacio muestral S toma un número **finito** (una cantidad que puedo contar) o infinito numerable (son infinitos, pero pueden ordenarse como 1, 2, 3, 4…), entonces decimos que X es una variable aleatoria discreta (VAD).

Ejemplos: número de hijos, número de pacientes atendidos en consulta médica, número de intervenciones quirúrgicas en un determinado hospital, etc.

Dada una VAD definiremos una función de W (conjunto de posibles valores de X) en el intervalo [0, 1], de tal manera que a todo elemento de W tenga asociado un número y solo uno en el intervalo [0, 1], si este número es la probabilidad de que la VA tome un determinado valor; dicha función se denomina **función de probabilidad de la variable aleatoria discreta f(xi).**

La distribución de probabilidad de una variable aleatoria X es similar a las distribuciones de frecuencias relativas. Se pueden representar por medio de un gráfico, tabla o fórmula que proporcionan los posibles valores de la variable X y la probabilidad asociada a ese valor.

Una función de probabilidad de una variable aleatoria discreta debe cumplir las siguientes condiciones:

$a)\ f(x_i) = P(X = x_i)$

$b)\ f(x_i) \geq 0$

Esto es que la probabilidad no puede ser negativa

$c)\ \sum f(x_i) = 1$

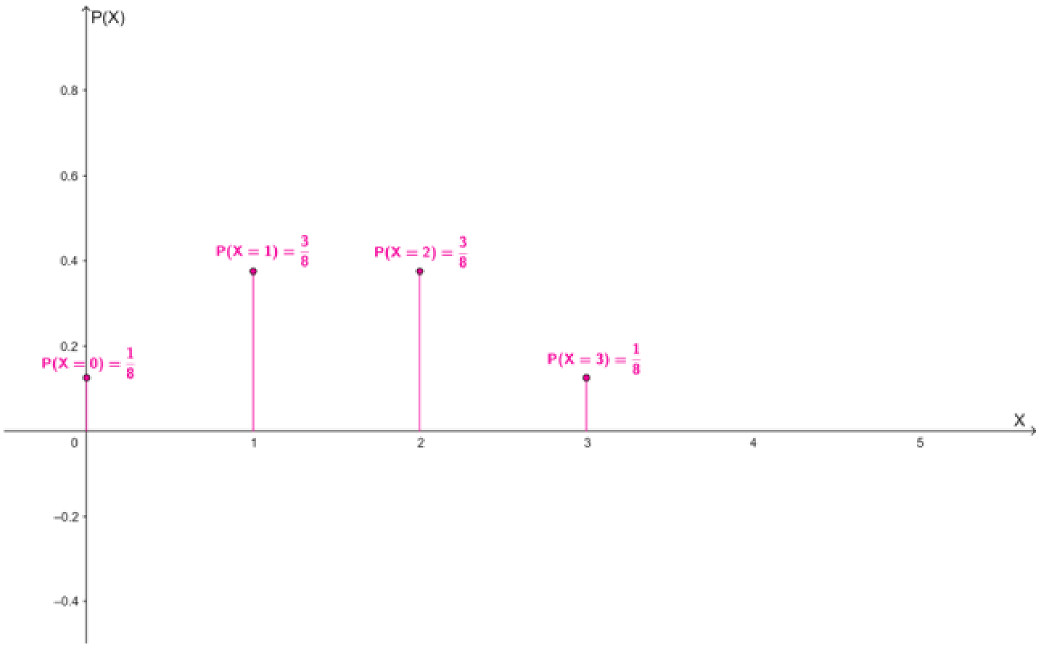
La suma de todas las probabilidades es 1.

EJEMPLO: Lanzamos 3 veces una moneda y definimos la VAD X: "Número de caras que aparecen en los tres lanzamientos". Ya hemos calculado las probabilidades en un ejercicio anterior, estas son:

$$P(X=0)=\frac{1}{8},\ P(X=1)=\frac{3}{8},\ P(X=2)=\frac{3}{8} \text{ y } P(X=3)=\frac{1}{8}$$

Se puede poner estas probabilidades en una tabla o en un gráfico de bastones, como se muestra a continuación:

X	P(X)
0	$\frac{1}{8}$
1	$\frac{3}{8}$
2	$\frac{3}{8}$
3	$\frac{1}{8}$



Las distribuciones de probabilidad se pueden utilizar para describir una población, de la misma manera en que hemos descripto las muestras.

Forma: Simetría, acampanada, unimodal, etc.

Valores alejados: medidas inusuales.

Centro y dispersión: media y desviación estándar. Una media poblacional se denomina con **µ** y la desviación estándar poblacional con **σ**.

La **esperanza o valor esperado E(X)** de una variable aleatoria, es el *valor medio* de un fenómeno aleatorio.

La **varianza Var(X)** de una variable aleatoria es una medida de dispersión definida como la esperanza del cuadrado de la desviación de dicha variable respecto a su media.

Si X es una variable aleatoria discreta con una distribución de probabilidad P(X) entonces la media o esperanza matemática, varianza y desviación estándar están dadas por las siguientes fórmulas:

Media o esperanza: $E(X) = \mu = \sum X.P(X)$

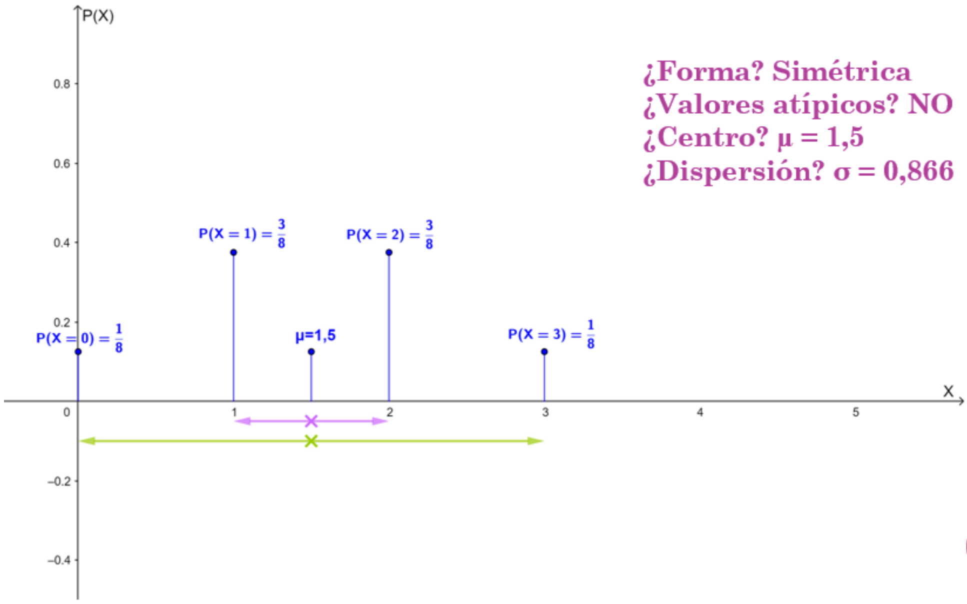
Varianza: $\sigma^2 = \sum(X-\mu)^2.P(X) = \sum X^2.P(X) - E(X)$

Desviación estándar: $\sigma = \sqrt{\sigma^2}$

EJEMPLO: Lanzamos 3 veces una moneda y se observa el número de caras.

X	P(X)	X.P(X)	(X-µ)².P(X)
0	$\frac{1}{8}$	$0.\frac{1}{8}=0$	$(-1,5)^2.\frac{1}{8}$
1	$\frac{3}{8}$	$1.\frac{3}{8}=\frac{3}{8}$	$(-0,5)^2.\frac{3}{8}$
2	$\frac{3}{8}$	$2.\frac{3}{8}=\frac{6}{8}=\frac{3}{4}$	$(0,5)^2.\frac{3}{8}$
3	$\frac{1}{8}$	$3.\frac{1}{8}=\frac{3}{8}$	$(1,5)^2.\frac{1}{8}$

$$\mu = \sum X.P(X) = \frac{12}{8} = 1{,}5$$
$$\sigma^2 = \sum (X-\mu)^2.P(X) = 0{,}28125 + 0{,}09375 + 0{,}09375 + 0{,}28125 = 0{,}75$$
$$\sigma = \sqrt{0{,}75} = 0{,}866$$



Las VAD sólo pueden tomar valores enteros, los cuales provienen de contar el número de individuos que cumple determinada característica. Existen diversos tipos de distribuciones discretas pero las dos que sirven como modelo en un gran número de aplicaciones son la **Binomial** y la de **Poisson**. En esta asignatura estudiaremos la distribución Binomial.

DISTRIBUCIÓN BINOMIAL

1) Un experimento binomial consiste en **n** ensayos idénticos de **Bernoulli**.
2) Los procesos de tipo Bernoulli se caracterizan por la obtención de dos resultados posibles: **éxito** o **fracaso**.
3) La probabilidad de éxito se denota con **p** y es constante para todas las repeticiones del experimento. La probabilidad de fracaso es $q = 1 - p$, pues se trata de eventos complementarios y mutuamente excluyentes.
4) Cada uno de los ensayos Bernoulli proporciona resultados que son independientes.

Algunos ejemplos de ensayos Bernoulli:

1. Sexo de recién nacidos.
2. Artículos defectuosos en una línea de producción.
3. Un consumidor está o no de acuerdo en pagar con tarjeta.

¿CÓMO RECONOCER SI UNA VARIABLE ES BINOMIAL?:

Cada vez que realizamos un ejercicio de probabilidad, debemos SI O SI definir la variable en cuestión y justificar el modelo elegido. Para justificar que una distribución es binomial debemos observar que consiste en varias pruebas idénticas repetidas, teniendo cada prueba dos posibles resultados excluyentes. Un experimento puede considerarse binomial si cumple las siguientes condiciones:

a. Cada experimento consiste en n pruebas idénticas

b. El resultado de cada prueba es dicotómico y puede ser evaluado como A o B.
c. Las n pruebas son idénticas. La probabilidad de que A ocurra es la misma en todas las pruebas del experimento.
d. La variable aleatoria binomial X, es el número de sucesos A, obtenidos en las n pruebas.

e. Cada una de las n pruebas es un ensayo de Bernoulli

EJEMPLO: En un determinado país, el 15% de la población tiene el gen del mal de Alzheimer. Si definimos la variable X: "Cantidad de personas con el gen de Alzheimer", el éxito está dado por el hecho de encontrar un gen en una persona. (Aclaración: el éxito está asociado a la definición de la variable, no a lo que consideramos que sea mejor).

Para la primera persona, P(gen) = 0,15.
Para la segunda persona, P(gen) ≈ 0,15, dado que la primera persona ya ha sido seleccionada.
No obstante, como se consideran poblaciones grandes, consideramos las probabilidades iguales. p es constante.
El hecho de que una persona tenga o no el gen no influye sobre el resultado al observar otra persona. **Los resultados son independientes.**
Hay un éxito que es "encontrar el gen en una persona" y un fracaso que es "no encontrar el gen en una persona"
Entonces se cumplen las condiciones de un experimento Bernoulli por tanto, la variable es **binomial**.
En general, para un experimento **binomial** con **n** ensayos y una probabilidad **p** de éxito para cada ensayo, diremos que la probabilidad de obtener **k** éxitos en **n** ensayos es:

$$P(X=k)=\frac{n!}{k!\,(n-k)!}.p^{k}.q^{n-k},\ k=0,1,2,\ldots.,n$$
$$\text{y } n!=n.(n-1).(n-2).....2.1 \text{ y } 0!=1.$$

Esto equivale a: $P(X=k)=\binom{n}{k}.p^{k}.q^{(n-k)}$ donde $\binom{n}{k}$ son las combinaciones de n en k (lo obtenemos con la calculadora)

***Para pensar*:** Analizar el caso del lanzamiento de una moneda tres veces.

EJEMPLO/EJERCICIO: Los tubos electrónicos producidos en una fábrica se clasifican en defectuosos o no defectuosos. Si los tubos se venden en la caja de 3 elementos y si se sabe que solo el 2% de la producción son defectuosos:

a) ¿Cuál es la probabilidad de que una caja con 3 artículos sea defectuosa?
b) ¿Cuál es la probabilidad de que una caja no tenga ningún tubo defectuoso?
c) ¿Cuál es el número promedio de artículos defectuosos que se espera encontrar en esta línea de producción?

RESPUESTA:
Variable: X: "Números de tubos defectuosos en cajas de 3 elementos."
¿Es binomial?: SI. **(¿POR QUÉ?)** p = 0,02 y q = 0,98.
Espacio muestral: S = {(ND ND ND), (D ND ND), (ND D ND), (ND ND D),(D D ND), (D ND D), (ND D D), (D D D)}

X	P(X)
0	0,98.0,98.0,98=$(0,98)^3$=**0,9412**
1	(0,98.0,98.0,02).3=$(0,98)^2$.3.0,02=**0,0576**
2	(0,98.0,02.0,02).3=$(0,02)^2$.3.0,98=**0,0012**
3	0,02.0,02.0,02=$(0,02)^3$=**0,000008**

a) $P(X=3)=(0,2)^3=0{,}000008$

b) $P(X=0)=0{,}9412$

c) $E(X)=\Sigma X.P(X)=0.0{,}9412+1.0{,}0576+2.0{,}0012\ +3.0{,}000008=\mathbf{0,06}$

LA MEDIA Y LA DESVIACIÓN ESTÁNDAR EN UNA DISTRIBUCIÓN BINOMIAL:

Media: $E(X)=n.p$

Varianza: $Var(X)=\sigma^2=n.p.q$

Desviación: $\sigma=\sqrt{n.p.q}$

EJEMPLO: Una persona que juega a lanzar dardos acierta en el blanco 80% de las veces. Si tira 5 veces al blanco, ¿cuál es la probabilidad de que exactamente 3 veces de en el blanco?

X: "cantidad de veces que da en el blanco" $n=5, p=0{,}8\ y\ q=0{,}2$. **(¿por qué es binomial?)**

$$P(X=3)=\frac{5!}{3!\,2!}.(0{,}8)^3.(0{,}2)^2=0{,}2048$$

¿Cuál es la probabilidad de que dé en el blanco más de 3 veces?

$$P(X>3)=P(X=4)+P(X=5)=$$

$$=\frac{5!}{4!\,1!}.(0{,}8)^4.(0{,}2)^1+\frac{5!}{5!\,0!}.(0{,}8)^5.(0{,}2)^0=0{,}7373$$

EJEMPLO:

1. Un tratamiento cura un 70% de los pacientes afectados de la micosis dérmica pitiriasis versicolor. En un estudio clínico se aplica el tratamiento a 10 pacientes.

Resolver las siguientes cuestiones:

a) Probabilidad de que curen exactamente 4 pacientes.
b) Probabilidad de que curen al menos 3 pacientes.
c) Número de pacientes que se espera que curen.

$n=10$ $p=0{,}7$ $q=0{,}3$ es una binomial porque tenemos dos resultados posibles excluyentes. Además, las 10 pruebas (pacientes) son idénticas y el valor de p es el mismo para todos.

X: "cantidad de pacientes curados"

a) $P(X=4)=\binom{10}{4}.0{,}7^4.0{,}3^6=210.0{,}7^4.0{,}3^6=0{,}0367\ldots$

B) Que se curen al menos tres: que se curen 3, 4, 5, 6...10. (se suman todas las probabilidades)

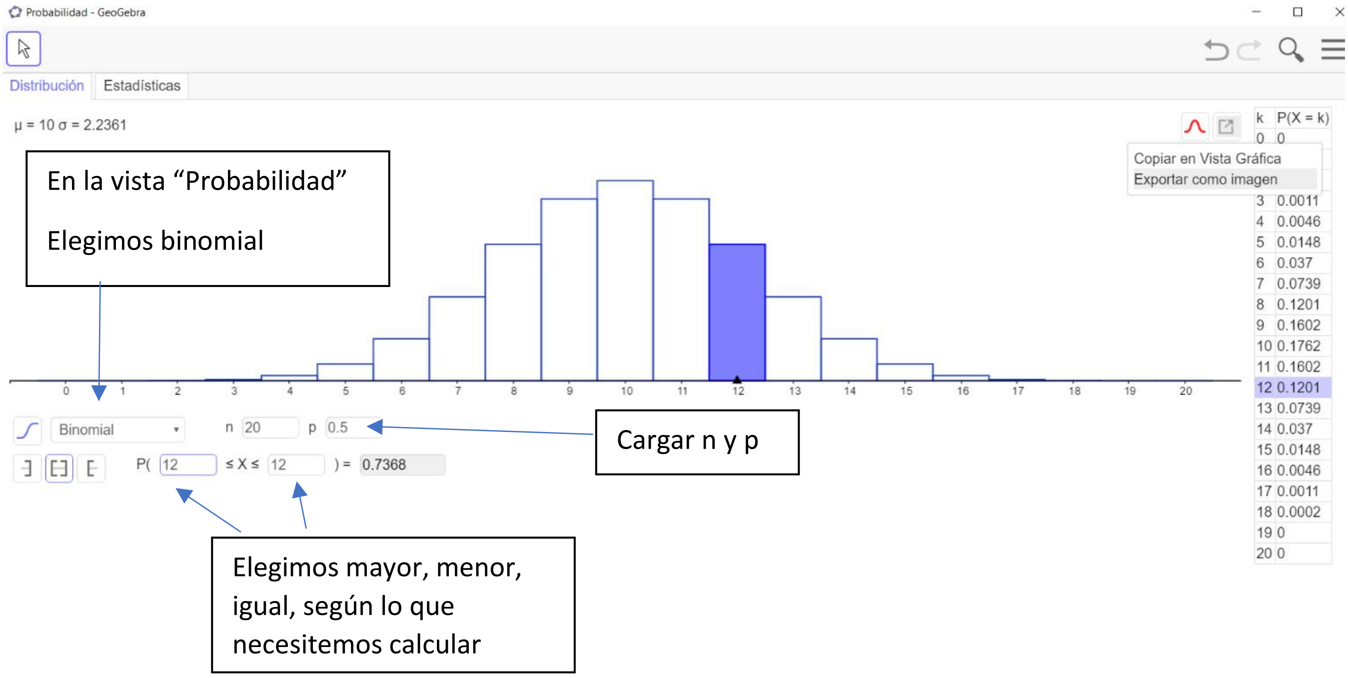
Eso es lo mismo que hacer $1 - P(X=0) - P(X=1) - P(X=2) = 1 - [P(X=0) + P(X=1) + P(X=2)]$

$1 - \binom{10}{0}.0{,}7^0.0{,}3^{10} - \binom{10}{1}.0{,}7^1.0{,}3^9 - \binom{10}{2}.0{,}7^2.0{,}3^8 = 0{,}9984\ldots$ **(¿por qué es lo mismo?)**

c) $E(x) = n.p = 10.0{,}7 = 7$

El valor esperado es que de un total de 10 pacientes se curen 7

En GeoGebra:



Distribución de probabilidad de variables aleatorias continuas

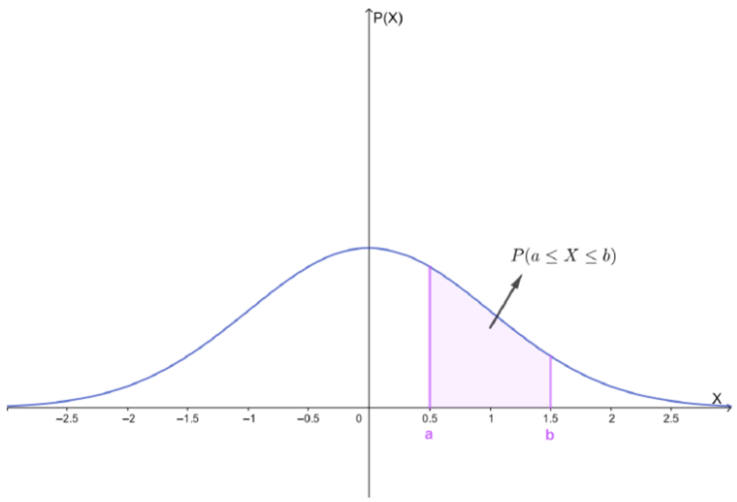
Están asociadas a variables aleatorias continuas. Ejemplo Alturas, pesos, tiempo de vida útil, etc.

Una curva de densidad describe la distribución de probabilidad de una variable continua. Es la aproximación “de un polígono de frecuencia.

La probabilidad se puede describir por una fórmula f(X) o P(X) llamada distribución de probabilidad o función de densidad de la variable aleatoria X.

El área bajo la curva es igual a 1, ya que representa la suma de todas las probabilidades.

$$P\ (a \le X \le b) = \text{á}\boldsymbol{rea\ bajo\ la\ curva\ entre\ a\ y\ b}.$$



La probabilidad para cualquier valor particular a de la variable es $P(X = a) = 0$ (esto no implica imposibilidad). Una de las variables aleatorias continuas más importante es la **variable aleatoria normal**.

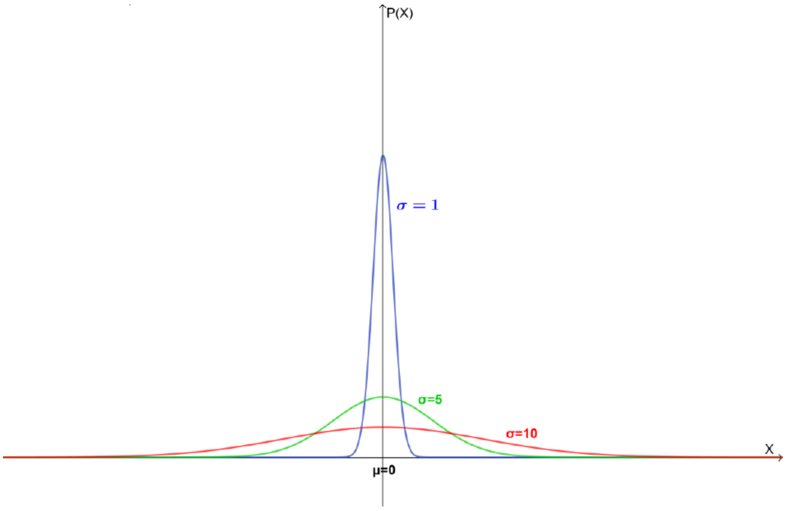
DISTRIBUCIÓN NORMAL

La fórmula (aunque utilizaremos una tabla) que genera la distribución de probabilidad normal es

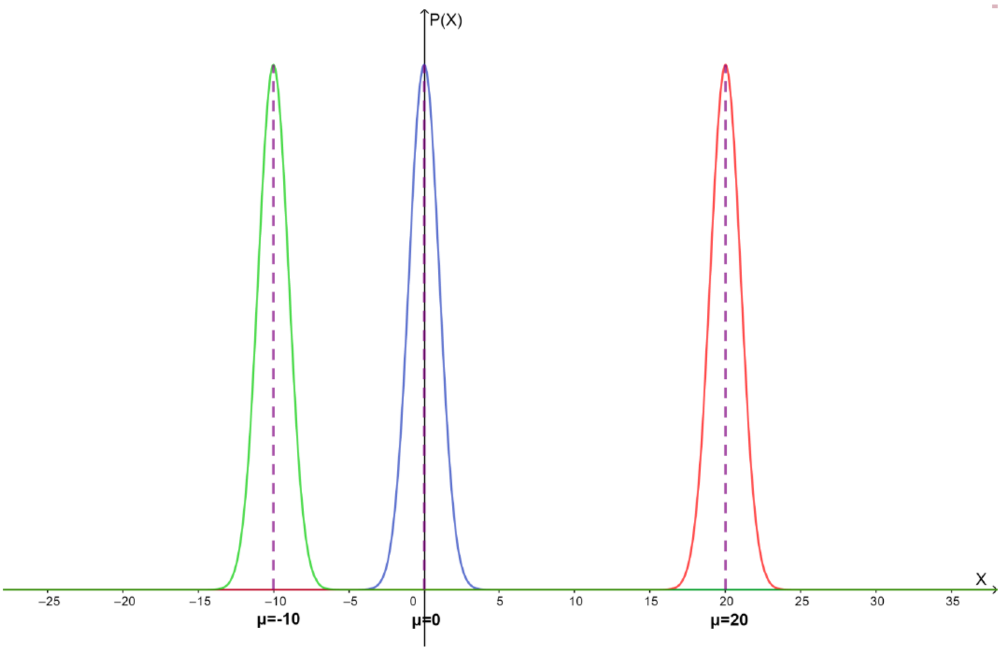
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

μ y σ son la media y desviación poblacional.

La forma y centro de la curva de densidad cambia de acuerdo a los valores de la media y de la desviación estándar. Por ejemplo



Los valores distintos de μ trasladan la gráfica de la distribución a lo largo del eje X y los diferentes valores de σ determinan el grado de aplanamiento o levantamiento de la gráfica.



La media de la distribución puede ser cualquier valor numérico negativo, cero o positivo.

La gráfica muestra 3 curvas normales con la misma desviación, pero con tres medias distintas.

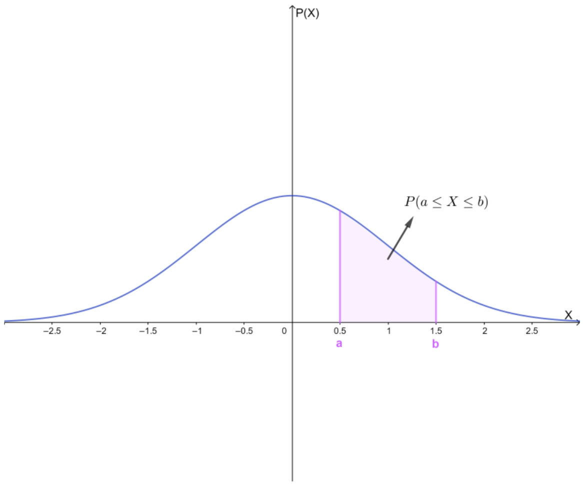
- La curva normal es unimodal y simétrica con respecto a su media μ, por lo tanto, $\bar{X} = Me = Mo$.

- La desviación estándar, representada por σ especifica el grado de dispersión respecto de la media.
- Existe toda una familia de distribuciones normales, cada una con su correspondiente μ y σ.
- En la distribución normal estándar $\mu = 0 \; y \; \sigma = 1$.

Cálculo de probabilidades (a través de tabla):

Habíamos dicho que:

$$P(a \leq X \leq b) \to$$



Este valor debe buscarse en la tabla de la distribución normal, pero para esto se debe estandarizar (ya que la tabla es de distribución normal estándar) cada valor de X a través de la fórmula:

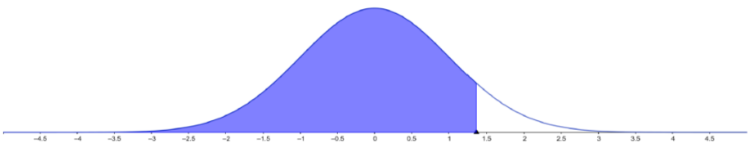
$$Z = \frac{x - \mu}{\sigma}$$

Características de la distribución normal estándar:

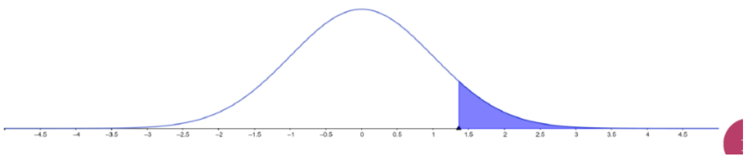
- μ=0 y σ=1.
- Es simétrica alrededor de z=0.
- El área total bajo la curva es 1.

Uso de la tabla (ver tabla en Anexo): En la primera columna encontramos la parte entera y la primera cifra decimal del valor de z, en la primera fila, la segunda cifra decimal. Buscando un valor particular de z, en la intersección de su fila y columna correspondiente encontramos cuatro dígitos (agregar adelante 0,) que proporcionan el área bajo la curva en un determinado dicho valor de z.

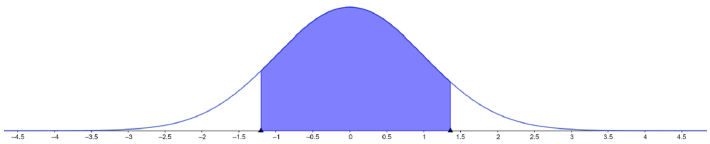
- $P(Z < 1{,}36) = 0{,}9131$



- $P(Z > 1{,}36) = 1 - P(Z \leq 1{,}36) = 1 - 0{,}9131 = 0{,}0869$

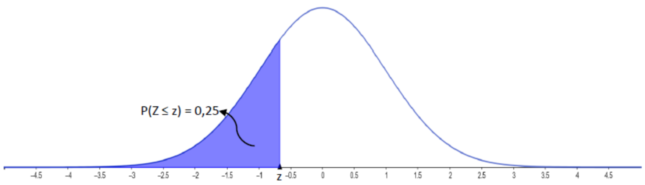


○ $P(-1{,}20 \leq Z \leq 1{,}36) = P(Z \leq 1{,}36) - P(Z \leq -1{,}20)$
$= 0{,}9131 - 0{,}1151 = 0{,}7950.$



- Se sabe que la probabilidad de encontrar un valor menor o igual que Z es de 0,25 ¿Cuánto vale z?

$P(Z \leq z) = 0{,}25$



Primero, deberíamos buscar el valor de la tabla que más se aproxima a 0 25 (dentro de la tabla).
Luego, vemos el valor de z correspondiente.
<u>Solución</u>: $z = 0{,}67$.

Si $\mu \neq 0$ y $\sigma \neq 1$ se debe proceder estandarizando de la siguiente manera:
<u>EJEMPLO</u>: $X \sim N(5{,}2)$, $\mu = 5$ y $\sigma = 2$.

Calcular $P(X > 7)$:

$$P(X > 7) = P\left(\frac{X-\mu}{\sigma} > \frac{7-5}{2}\right) = P\left(Z > \frac{2}{2}\right) = P(Z > 1) = 1 - P(Z \leq 1)$$
$$= 1 - 0{,}8413 = 0{,}1587$$

<u>*Problemas con la distribución normal*</u>:

1. En los estudios dactiloscópicos, una importante característica cuantitativa es el número total de surcos en los dedos de un individuo. Supóngase que los números totales de surcos de los individuos de cierta población están distribuidos en forma normal con μ=140 y σ=50.

Encontrar la probabilidad de que un individuo seleccionado al azar dentro de la población tenga un número de surcos:

a) de 200 o más.
b) menor de 100.
c) entre 100 y 200.
d) entre 200 y 250.

RESPUESTAS: Variable: X: “Número total de surcos.”

a) Planteamos lo siguiente:

$$P(X \geq 200) = P\left(Z \geq \frac{200-140}{50}\right) = P\left(Z \geq \frac{60}{50}\right) = P\left(Z \geq \frac{6}{5}\right) = P(Z \geq 1{,}2) =$$
$$= 1 - P(Z < 1{,}2) = 1 - 0{,}8849 = 0{,}1151$$

b) Tenemos lo siguiente:

$$P(X < 100) = P\left(Z < \frac{100-140}{50}\right) = P(Z < -0{,}8) = 0{,}2119$$

c) Planteamos:

$$P(100 < X < 200) = P(-0{,}8 < Z < 1{,}2) = 0{,}8849 - 0{,}2119 = 0{,}673$$

d) Por último, tenemos que:

$$P(200 < X < 250) = P(1{,}2 < Z < 2{,}2) = 0{,}9861 - 0{,}8849 = 0{,}1012$$

2. El nivel de estrés en personas de 30 o 40 años está normalmente distribuido con μ = 1 punto y σ = 0,10 puntos.

a) ¿Qué probabilidad hay de seleccionar al azar una persona de la población que tenga un nivel de estrés comprendido entre 0,80 y 0,85?

$$P(0{,}80 \leq X \leq 0{,}85) = P(0{,}80 < X < 0{,}85) = P\left(\frac{0{,}80-1}{0{,}10} < Z < \frac{0{,}85-1}{0{,}10}\right) =$$
$$= P(-2 < Z < -1{,}5) = P(Z < -1{,}5) - P(Z < -2) = 0{,}0668 - 0{,}0228 = 0{,}044$$

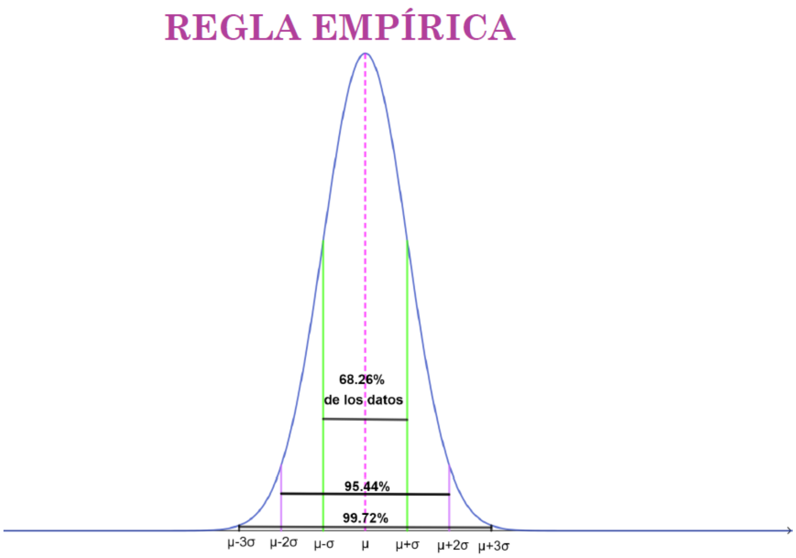
b) ¿Cuál es el nivel mínimo de estrés que puede tener una persona que esté dentro del 1% de individuos con niveles muy altos de estrés?

$$P(X > x) = 0{,}01 \Rightarrow P\left(Z > \frac{x-1}{0{,}1}\right) = 0{,}01 \Rightarrow 1 - P\left(Z \leq \frac{x-1}{0{,}1}\right) = 0{,}01$$
$$\Rightarrow P\left(Z \leq \frac{x-1}{0{,}1}\right) = 0{,}99 \Rightarrow \frac{x-1}{0{,}1} = 2{,}33 \Rightarrow x = 2{,}33.0{,}1 + 1 \Rightarrow x = 1{,}233$$

Por lo tanto, el nivel mínimo de estrés correspondiente al grupo que tiene niveles altos de estrés es 1,233 puntos (dicho valor corresponde al P_{99}).

APROXIMACIÓN DE DISTRIBUCIONES EMPÍRICAS A LA NORMAL

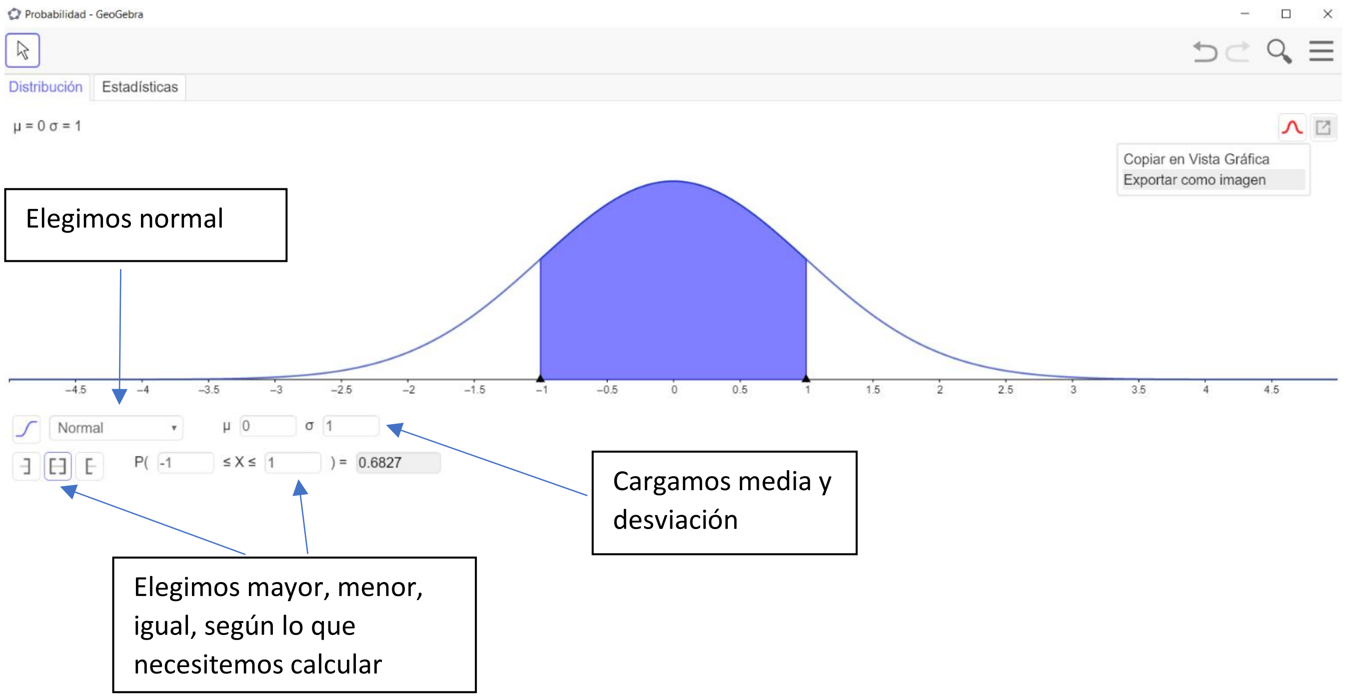
En muchas situaciones, no podemos determinar a priori si la variable que se observa tiene o no una forma normal. En esta situación, podemos buscar relaciones entre las distribuciones normal y la distribución de frecuencias empíricas y observar si la distribución empírica se aproxima a la normal. Una forma de hacer esto (para variables cuantitativas continuas o discretas) es aplicando la regla empírica y observando si la distribución de frecuencia es unimodal y simétrica.



Esta regla establece que para datos con distribución en forma de campana:

1) El 68% de los datos están a menos de una desviación estándar de la media. Este es el valor del área en esa región.
2) El 95% de los datos están a menos de dos desviaciones estándar de la media.
3) Casi todos los elementos están a menos de tres desviaciones estándar de la media.

En GeoGebra, calculamos las probabilidades:



PRÁCTICA 3

En todos los ejercicios que puedas, aplicar GeoGebra para verificar tus respuestas

Actividad 1. Se lanza una moneda tres veces seguidas y se observa si sale cara o cruz.

a. Escribe el espacio muestral, es decir, todos los resultados posibles que puedes obtener.

b. Clasifica la variable: "número de cruces que se obtienen en los tres lanzamientos".

c. ¿Cuál sería la probabilidad de obtener dos cruces?

d. ¿Qué probabilidad hay de obtener por lo menos una cruz?

Actividad 2. Considera el siguiente experimento: Se lanzan dos dados y se observan los valores obtenidos.

a. Escribe el espacio muestral con todos los resultados posibles para este experimento.

b. ¿Cuál es la probabilidad de obtener dos unos?

c. ¿Cuál es la probabilidad de obtener un 2 y un 6?

d. ¿Cuál es la probabilidad de obtener dos números iguales?

Actividad 3. Una baraja de 52 cartas contiene 26 cartas rojas y 26 cartas negras. Se saca una carta del mazo, se observa su color y se la devuelve al mazo.

a. ¿Cuál es la probabilidad de que la carta escogida sea negra?

b. ¿Cuál es la probabilidad de que la carta escogida sea roja?

c. Supongamos que sacaste una carta y resultó ser negra. Ahora escoges una segunda carta sin haber devuelto al mazo la primera carta, ¿cuál es la probabilidad de que la segunda carta sea negra? ¿Y cuál es la probabilidad de que sea roja?

Actividad 4. Se lanza una moneda dos veces seguidas y se anota si sale cara o cruz en cada lanzamiento.

a. Escribe el espacio muestral, es decir, todos los resultados posibles que puedes obtener. ¿Cuál es el total de resultados posibles?

Basándote en el espacio muestral obtenido en el ítem a, responde a las siguientes preguntas:

b. Si nos interesa estudiar la variable X= "Número de caras que se obtiene en dos lanzamientos". ¿Cuál es la probabilidad de éxito? ¿Y la de fracaso?

c. ¿Cuál sería la probabilidad de obtener por lo menos una cara?

Actividad 5. Se tienen dos eventos A y B definidos en un mismo espacio muestral, tales que:

P (A)=0,39 P (B)=0,21 P (A o B)=0,47

Encuentre la probabilidad de que:

a. No se presente ni A ni B.

b. Se presenten tanto A como B.

c. Se presente B dado que A ya se ha presentado.

d. Se presente A dado que B ya se ha presentado.

Actividad 6. Si A y B son eventos mutuamente excluyentes, P (A)=0,3 y P (B)=0,5, encuentre:

a. P $(A \cup B)$

b. P (A^c)

c. P $(A^c \cap B)$

Actividad 7. Un espacio muestral S consta de cinco eventos simples con estas probabilidades:

$$P(E_1) = P(E_2) = 0.15,\ P(E_3) = 0.4,\ P(E_4) = 2P(E_5)$$

Encuentre las probabilidades para los eventos E_4 y E_5.

Actividad 8. Un espacio muestral contiene 10 eventos simples: $E_1, E_2, \ldots, E_{10}$.

Si $P(E_1) = 3.P(E_2) = 0.45$ y los eventos simples restantes son equiprobables, encuentre las probabilidades de los eventos simples restantes.

Actividad 9. El secretario del sindicato que agrupa a los trabajadores de una empresa, redactó una lista de demandas para presentarle al gerente de la empresa. Previamente, indagó acerca del grado de apoyo al paquete de demandas con el que contaba. Para ello, realizó un sondeo aleatorio entre los dos grupos principales de trabajadores: Maquinistas (M) e Inspectores (I). Tomó 30 trabajadores de cada grupo, obteniendo los siguientes resultados:

Grado de apoyo al paquete de demandas	**M**	**I**	**Total**
Apoyo fuerte	9	10	19
Apoyo leve	11	3	14
Indecisos	2	2	4
Levemente opuestos	4	8	12
Fuertemente opuestos	4	7	11
Total	**30**	**30**	**60**

a. Si se selecciona al azar a un trabajador del grupo sondeado,

i. ¿cuál es la probabilidad de que sea un maquinista?

ii. ¿cuál es la probabilidad de que esté indeciso?

iii. ¿qué es más probable que ocurra, que apoye el paquete o que se oponga a él?

b. ¿Cuál es la probabilidad de que un maquinista, seleccionado al azar del grupo sondeado, apoye levemente al paquete? Interprete.

c. ¿Cuál es la probabilidad de que un inspector, seleccionado al azar del grupo sondeado, esté fuertemente opuesto al paquete?

d. Si se selecciona al azar a un trabajador del grupo que apoya levemente el paquete, ¿cuál es la probabilidad de que sea un maquinista?

e. ¿Cambiarían si se tomara una nueva muestra de 30 trabajadores de cada grupo? ¿Por qué?

Actividad 10. Una urna contiene un total de 75 bolitas: 35 son azules y 25 de éstas, son rayadas. Las 40 restantes son rojas, de las cuales 30 también están rayadas. Si se selecciona al azar una bolita de esta urna,

a. ¿Cuál es la probabilidad de que sea azul? Interprete.

b. ¿Cuál es la probabilidad de que sea lisa?

c. ¿Cuál es la probabilidad de que sea azul y rayada?

d. ¿Cuál es la probabilidad de que sea roja y que no esté rayada?

e. ¿Cuál es la probabilidad de que sea roja o rayada?

f. ¿Cuál es la probabilidad de que sea verde?

Actividad 11. Una pequeña presa hidroeléctrica tiene 4 compuertas que fallan y son reparadas de manera independiente una de la otra cuando se produce alguna falla. Por experiencia, se sabe que cada compuerta está fuera de servicio 4% de todo el tiempo.

a. ¿Cuál es la probabilidad de que las cuatro compuertas estén fuera de servicio?

b. ¿Cuál es la probabilidad de que las cuatro compuertas funcionen adecuadamente?

c. Si la compuerta 1 está fuera de servicio, ¿cuál es la probabilidad de que las compuertas 2 y 3 también estén fuera de servicio?

Actividad 12. Diga si la siguiente aseveración es Verdadera o Falsa. Justifique su elección.

Si P (H∪Q)=0,70; P (H|Q)=P (Q|H) y P (H|Q)=0,30 entonces P (H)=0,10 y P (Q)=0,90.

Actividad 13. Un psicólogo escolar dispone de los datos de 10 niños de un colegio:

Niño	Sexo	Escolarización	CI
1	V	No	72
2	V	Sí	95
3	M	No	75
4	V	Sí	100
5	M	Sí	110
6	M	No	78
7	V	Sí	105
8	V	No	65
9	M	Sí	115
10	M	Sí	105

Se definen tres sucesos:

A: "El sujeto extraído es varón"

B: "El sujeto extraído está escolarizado"

C: "El sujeto extraído supera la media de su grupo en CI"

Si se extrae un sujeto al azar:

a. ¿Cuál es la probabilidad de que sea varón?

b. ¿Cuál es la probabilidad de que esté escolarizado?

c. ¿Cuál es la probabilidad de que supere la media de su grupo en CI?

d. ¿Cuál es la probabilidad de que sea varón y esté escolarizado?

e. ¿Cuál es la probabilidad de que sea varón y supere la media de su grupo en CI?

f. ¿Cuál es la probabilidad de que sea mujer y supere la media de su grupo en CI?

g. ¿Cuál es la probabilidad de que sea varón, esté escolarizado y supere la media de su grupo en CI?

h. Elabore la tabla de contingencia entre las variables sexo y escolarización. Responda:

- **i.** Dado que el sujeto extraído es un varón ¿Cuál es la probabilidad de que esté escolarizado?
- **ii.** Dado que el sujeto extraído está escolarizado ¿Cuál es la probabilidad de que sea varón?
- **iii.** Dado que el sujeto extraído es una mujer ¿Cuál es la probabilidad de que no esté escolarizada?

Actividad 14. A un grupo de 100 sujetos se les pasa una prueba sobre "satisfacción con el estilo de vida" y otra sobre "depresión". Los resultados indican que 40 sujetos superaron la media en satisfacción y 65 sujetos eran depresivos. Asimismo, dentro de los depresivos se encontró que sólo 10 obtenían puntuaciones superiores a la media en satisfacción.

Se definen dos sucesos: A: "Tener una puntuación superior a la media en satisfacción con el estilo de vida"

B: "Ser depresivo"

a. ¿Cuál es la probabilidad de que, al extraer un sujeto, sea depresivo?

b. ¿Son los sucesos A y B independientes?

c. Si se ha seleccionado un sujeto que puntúa por encima de la media en satisfacción con el estilo de vida, ¿cuál es la probabilidad de que sea depresivo?

d. ¿Cuál es la probabilidad de que ocurra al menos uno de los sucesos, A o B?

DISTRIBUCIONES DE PROBABILIDAD:

Actividad 15. Entre las mujeres que trabajan, el 25% nunca se ha casado. Se seleccionan al azar a 10 mujeres con empleo.

a. ¿Cuál es la probabilidad de que exactamente 2 mujeres de la muestra nunca se hayan casado?

b. ¿Cuál es la probabilidad de que como máximo 3 mujeres nunca se hayan casado?

Actividad 16. Se ha pasado un test de inteligencia a un grupo de alumnos de 12 años y se ha podido determinar que la distribución de las puntuaciones siguen una distribución normal con media 100 y desviación típica 16.

a. ¿Qué proporción de alumnos tienen puntuaciones de inteligencia más bajos que 84?

b. ¿Qué proporción de alumnos tienen notas mayores que 84 puntos?

c. ¿Qué proporción de alumnos tienen notas entre 84 y 116?

Actividad 17. En una determinada ciudad, la cuota por contribuyente del impuesto municipal de automóviles sigue una distribución normal con media $65 y desviación típica $26.

a. ¿Qué porcentaje de contribuyentes paga una cuota comprendida entre $26 y $52?

b. ¿Cuál es la probabilidad de que un contribuyente pague más de $70 por el impuesto municipal?

Actividad 18. La prueba WAIS es una prueba de inteligencia para adultos. La distribución de los resultados de esta prueba para personas mayores de 16 años es aproximadamente normal con media 100 y desviación típica de 15.

a. ¿Cuál es la probabilidad de que un individuo escogido al azar tenga un resultado de 105 o superior?

b. ¿Qué porcentaje de personas tendrán resultados que varíen entre 90 y 110 puntos?

Actividad 19. La variable Tiempo en Diciembre (TIEMPO_DIC) corresponde a datos reales que han sido tomados de una muestra de 96 alumnos de un Profesorado de Educación Física. Estos alumnos han realizado un determinado tipo de entrenamiento con el fin de mejorar sus tiempos en una carrera de 30 metros, el tiempo está dado en segundos. A continuación, se presentan la tabla de frecuencias, histograma, diagrama de tallos y hojas y resumen estadístico correspondiente a dicha variable.

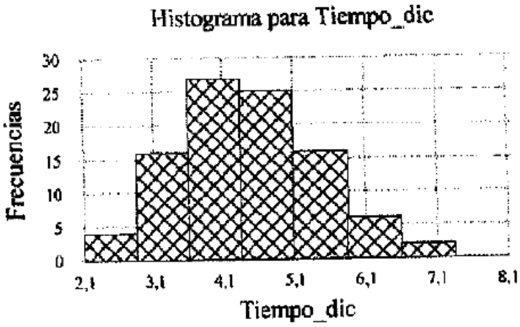
Tabla de Frecuencias para la variable TIEMPO_DIC

Límite Inferior	Límite Superior	Punto Medio	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia Acumulada	Frec. Acum. Relativa
2,1	2,6	2,35	1	0,0104	1	0,0104
2,6	3,1	2,85	8	0,0833	9	0,0937
3,1	3,6	3,35	8	0,0833	17	0,1770
3,6	4,1	3,85	17	0,1771	34	0,3541
4,1	4,6	4,35	21	0,2187	55	0,5728
4,6	5,1	4,85	15	0,1563	70	0,7291
5,1	5,6	5,35	12	0,1250	82	0,8541
5,6	6,1	5,85	8	0,0833	90	0,9374
6,1	6,6	6,35	4	0,0417	94	0,9791
6,6	7,1	6,85	2	0,0208	96	1
TOTAL			96	1		

Diagrama de Tallos y hojas para la variable TIEMPO_DIC

2|4
2|67899
3|0001233344
3|5666788888999999
4|00111222222233334444
4|555566777788899999
5|00112233344444
5|6666779
6|0144
6|699

Histograma para la variable TIÈMPO_DIC



Resumen Estadístico para la variable TIEMPO_DIC:

Tamaño de la muestra: 96 alumnos
Mediana: 4,4 seg.
Media: 4,45
Coeficiente de asimetría: 0,25
Moda: 4,2 seg.
Desviación Típica: 0,99

Utilizando las representaciones anteriores responde a las siguientes preguntas:

a. ¿De qué tipo es la variable con la que se está trabajando?

b. ¿Cuántas personas obtuvieron un tiempo mayor o igual a la moda? Explica.

c. ¿Es aproximadamente simétrica esta distribución? Justifica tu respuesta aplicando la o las propiedades que conozcas.

d. Esta distribución, ¿tiene valores atípicos? ¿En qué te apoyas para justificar tu respuesta?

e. ¿Sería adecuado afirmar que esta distribución es aproximadamente normal? Realiza todos los cálculos pertinentes y utiliza todas las propiedades conocidas para justificar tu afirmación.

f. ¿Cuál es la probabilidad de que un alumno escogido al azar de esta muestra haya hecho un tiempo de 6,43 segundos o mayor?

g. ¿Entre qué tiempos alrededor de la media estaría el 99,7% de los alumnos?

Actividad 20. Se conoce por estudios anteriores que la probabilidad de que un agente de seguros consiga vender una póliza en una entrevista con un cliente es de 0,1.

a. ¿Cuál es la variable en estudio?

b. ¿Cuáles son la probabilidad de éxito y de fracaso?

c. ¿Cuántas pólizas esperaría vender este agente si entrevista a 15 personas? (Esperanza=n.p)

Actividad 21. En un barrio en el que hay muchos bares, se sabe que el 30% de sus vecinos sufren de insomnio. Si se toma una muestra de 10 vecinos, ¿cuál es la probabilidad de que 2 de ellos sufran de insomnio? ¿Cuál es la variable en estudio?

Actividad 22. En un sondeo sobre la actitud hacia la donación de órganos se encuentra que en una determinada población hay un 80% de sujetos que están a favor. Si se extrae una muestra aleatoria de 10 sujetos obtenga lo siguiente:

a. Probabilidad de que 4 personas estén a favor.

b. Probabilidad de que más de 4 personas estén a favor.

c. Probabilidad de que menos de 4 personas estén a favor.

d. Probabilidad de que como máximo 4 personas estén a favor.

e. Probabilidad de que como mínimo haya 7 personas a favor.

f. Probabilidad de que estén en contra 4 o más personas.

Actividad 23. La variable extroversión se distribuye según el modelo normal con media 50 y desviación típica 10. Conteste a las siguientes preguntas:

a. Probabilidad de que los sujetos obtengan como mucho una puntuación de 35.

b. Probabilidad de que los sujetos obtengan una puntuación mayor de 60 en extroversión.

c. Calcular la puntuación en extroversión que deja por debajo de sí al 80% de los sujetos.

d. Probabilidad de observar un valor comprendido entre 42 y 59.

e. Calcular los valores que acotan el 50% central de sujetos.

f. Si se extrae una muestra aleatoria de 25 alumnos, ¿cuál es la probabilidad de que su media aritmética sea mayor de 55?

Actividad 24. Un profesor ha calculado que el tiempo invertido por los estudiantes en hacer ejercicios sigue una distribución normal con media 150 minutos y una desviación estándar de 40 minutos.

a. ¿Cuántos minutos tarda un alumno si el 90% de sus compañeros tardan más que él?

b. ¿Cuántos minutos tarda un alumno si el 80% de sus compañeros tardan menos que él?

Actividad 25. Los puntos de un Test de Aptitud se describen según una normal con media 420 y desviación estándar de 80.

a. Se elige una persona al azar, ¿cuál es la probabilidad de que una puntuación esté entre 400 y 480?

b. ¿Cuál es la puntuación necesaria para estar entre el 10% mejor?

c. Sin hacer cálculos, determinar en cuál de los siguientes intervalos es más probable que esté la puntuación de una persona elegida al azar: [400,440] ó [440,480] ó [480,520] ó [520,560]. ¿Por qué?

Actividad 26. Decir si las siguientes afirmaciones son Verdaderas o Falsas y, <u>en todos los casos</u>, justificar:

a. Seis meses después de divorciarse, cada uno de los ex esposos de una pareja realiza un test para medir su adaptación al divorcio. Luego de realizar el test, la esposa obtiene una puntuación de 63 y el esposo de 59 puntos. Por lo general, la distribución de las puntuaciones de este test es normal con una media para mujeres divorciadas de 60 puntos y su desviación es de 6 puntos, mientras que la distribución para los hombres divorciados que hacen este test también es normal con una media de 55 puntos y una desviación de 4 puntos. Se considera además que, a mayor puntuación obtenida luego de realizar el test, mayor es el nivel de adaptación. Por lo tanto, podemos concluir que al cabo de 6 meses se ha adaptado mejor la mujer.

b. Se encontró que un conjunto de calificaciones de exámenes en un curso de Psicología se distribuye normalmente con una media de 73 y una desviación estándar de 8. En consecuencia:

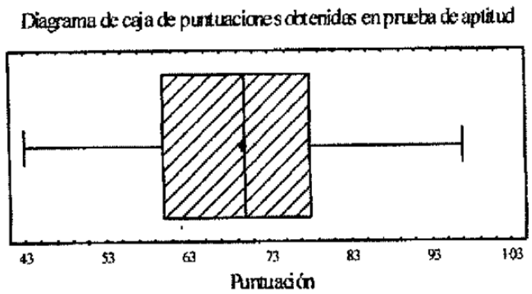
- **i.** La probabilidad de obtener a lo más una calificación de 91 en este examen es de 0,72.
- **ii.** El porcentaje de estudiantes que sacaron una calificación entre 65 y 81 es de 68%.
- **iii.** Si el profesor califica por campana (es decir, otorga la calificación A al 10% superior de la clase sin importar la calificación), es mejor obtener una calificación de 81 en este examen que una calificación de 68 en otro examen con distribución normal, en el que la media es 62 y la desviación estándar es 3.

Actividad 27. Se pasó un test de introversión-extroversión a una población de estudiantes. Por medio de esta prueba, cada alumno obtiene una puntuación (los resultados altos corresponden a las personas más introvertidas) que corresponde a la variable que indica el grado de introversión o de extroversión de cada individuo. A partir de la población observada se determinó que la distribución de la población tiene una media de 82 puntos y una desviación de 12 puntos. Se pasó esta misma prueba a otros tres estudiantes, los cuales obtuvieron las siguientes puntuaciones: Juan obtuvo 88 puntos, Jorge 96 puntos y María 72 puntos. A partir de las puntuaciones estandarizadas, ¿cuál de los tres es el más introvertido y cuál es el más extrovertido?

Actividad 28. En un test de aptitud que se administró a 39 aspirantes a un cargo del Poder Judicial de la provincia de Santa Fe, se obtuvieron los resultados que se muestran a continuación en el diagrama de tallo y hojas y en el diagrama de caja. La escala de este test de aptitud va de 0 a 100 y se considera que a mayor valor dentro de la escala, mayor es la aptitud del aspirante con respecto al puesto que va a desempeñar. Ingresan aquellos aspirantes que obtienen un puntaje de 60 o más y sólo aquellos que obtienen un puntaje mayor a 75 pueden acceder a puestos de mayor jerarquía.

Diagrama de tallo y hojas

4|367
5|003899
6|0124577999
7|00113456788
8|234578
9|027



Media = 69,72 y desviación = 13,43

En base a la información descrita antes y a los diagramas, responde a las siguientes cuestiones:

a. ¿Sería adecuado afirmar que la distribución de las puntuaciones obtenidas por todos los aspirantes al Poder Judicial de la provincia de Santa Fe es aproximadamente normal? Justifica tu respuesta verificando **todas** las propiedades necesarias.

b. En función de la conclusión dada en el ítem anterior, calcula la probabilidad de que un aspirante escogido al azar de la población de la que se tomó esta muestra tenga una puntuación mayor o igual a 80. Interpreta dicho resultado.

c. Para poder ingresar al Poder Judicial de la provincia de Santa Fe, ¿en qué percentil mínimo debería ubicarse un aspirante?

UNIDAD 5: RELACIONES ENTRE VARIABLES ESTADÍSTICAS

Diagramas de dispersión

Si queremos analizar si hay relación entre dos variables cuantitativas, el tipo de relación y si es posible, predecir el valor de una de ellas en función a la otra, es muy útil comenzar con un diagrama de dispersión.

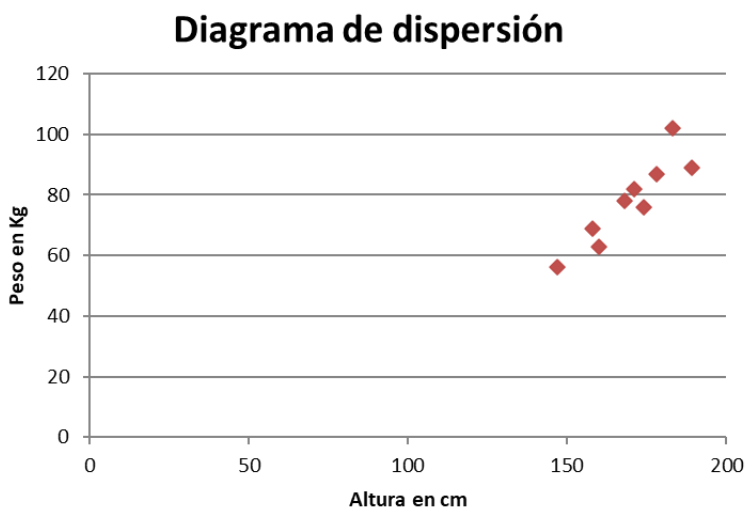
El diagrama muestra una nube de puntos con los valores obtenidos de las variables. Si podemos determinar la variable independiente (es aquella cuyo valor no depende de la otra variable) debemos dibujarla en el eje horizontal x, mientras que la variable dependiente (es aquella cuyos valores dependen de los que tomen otra variable) debemos dibujarla en el eje vertical y.

¿Cómo construimos un Diagrama de Dispersión?

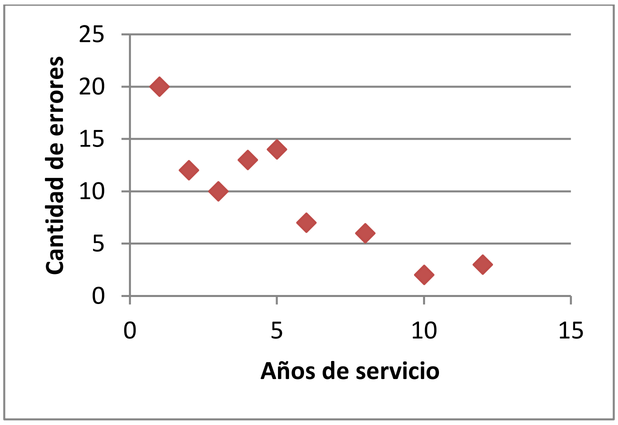
- Recolectamos n parejas de datos de la forma (Xi, Yi), con i = 1, 2, 3,…n donde Xi y Yi representan los valores respectivos de las dos variables. Los datos se suelen representar en una tabla.
- Diseñamos las escalas apropiadas para los ejes X y Y.
- Graficamos las parejas de datos (puntos coordenados).

Veamos un **ejemplo**: Queremos ver si existe y el tipo de relación entre el peso y la altura de 9 personas:

altura (cm)	peso (kg)
171	82
147	56
183	102
168	78
160	63
158	69
174	76
189	89
178	87

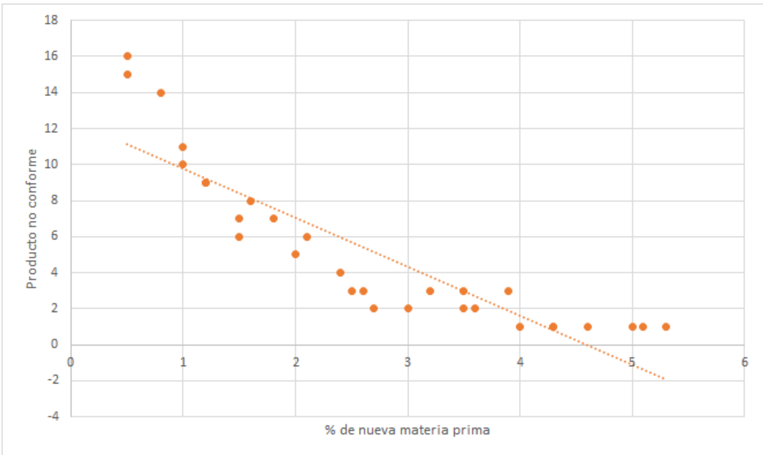


Otros ejemplos:

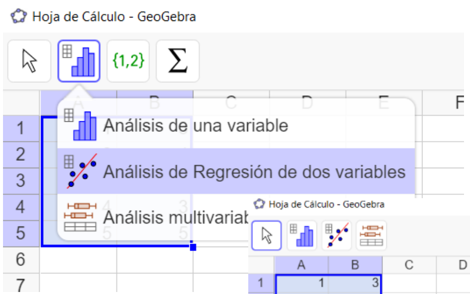


¿Notaron diferencias en los tres diagramas anteriores? ¿Son iguales los comportamientos de las nubes de puntos?

- La lectura del diagrama se hace en base al tipo de relación entre los datos; lo fuerte o débil de la relación, la forma de la relación y la posible presencia de punto anómalos.
- La relación entre dos variables se denomina “**correlación positiva o directa**” cuando a un aumento en el valor de la variable X le acompaña un aumento en la otra variable (1er ejemplo, a mayor altura, mayor peso). La nube de puntos tiene un aspecto creciente.
- La relación entre dos variables se denomina **“correlación negativa o inversa”** cuando a un aumento en el valor de la variable X le acompaña la disminución en la otra variable (en el 2do. ejemplo, a mayor cantidad de años de servicios, menor cantidad de errores). La nube de puntos tiene un aspecto decreciente.
- La relación entre dos variables se denomina **“incorreladas”** cuando el comportamiento de una no puede predecir a la otra, no hay relación aparente. La nube de puntos no da ningún indicio, es muy dispersa (en el 3er ejemplo, no hay clara relación entre la experiencia y el tiempo).

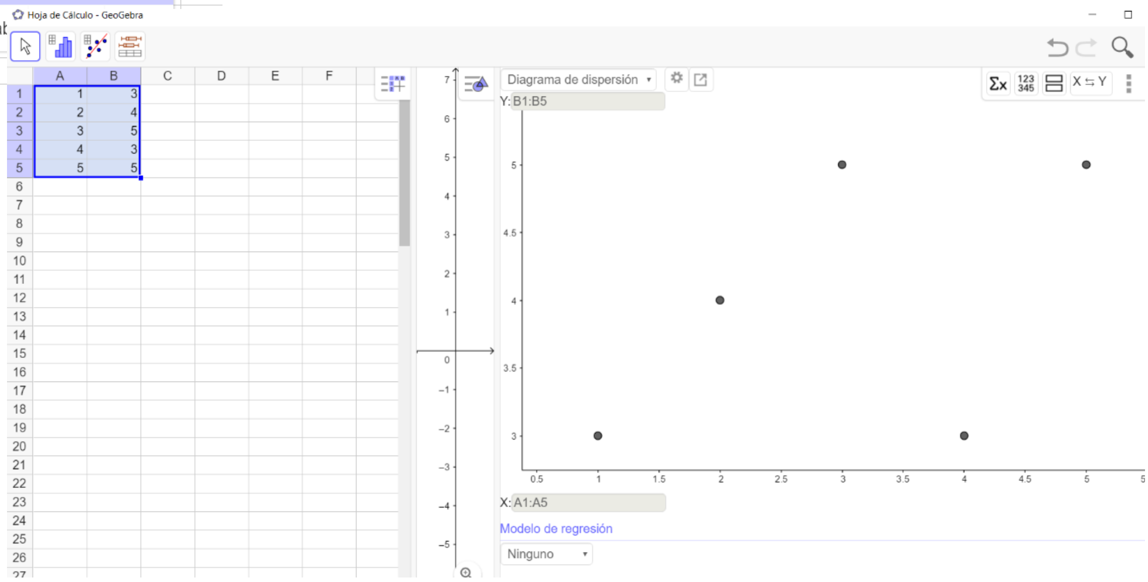


En GeoGebra:



Cargamos los datos, seleccionamos y elegimos la opción análisis de regresión.

Elegimos diagrama de dispersión.



Coeficiente de Correlación de Pearson (r):

- La cuantificación de la fuerza de la relación lineal entre dos variables cuantitativas se estudia por medio del cálculo del coeficiente de correlación de Pearson. Dicho coeficiente oscila entre [–1 y +1]. Un valor de 1 indica una relación lineal positiva o directa perfecta; así como un valor de –1 nos indica una relación negativa o inversa perfecta. Una correlación próxima a cero indica que no hay relación lineal entre las dos variables. El realizar la representación gráfica (diagrama de dispersión) de los datos para demostrar la relación entre el valor del coeficiente de correlación y la forma de la gráfica es fundamental ya que existen relaciones no lineales.
- El coeficiente de correlación lineal sólo detecta relaciones lineales entre dos variables.
- Si el coef. r es mayor a 0,7 o menor a -0,7, hay una relación fuerte y, si r es mayor a 0,3 o menor a -0,3 hay una relación que puede estudiarse.
- Si r está comprendido -0,3 y 0, 3 la relación es casi inexistente, por lo tanto, no se puede estudiar esa relación.
- Se puede hacer con calculadora:
  https://www.youtube.com/watch?v=SrGHclGORX4&ab_channel=Andr%C3%A9sArcos
  https://www.youtube.com/watch?v=4cQe6J7RzAI&ab_channel=CONSTANZAALEJANDRAAHUMADA

El coeficiente de correlación posee las siguientes características:

a. El valor del coeficiente de correlación **es independiente de cualquier unidad usada** para medir las variables.

b. El valor del coeficiente de correlación se altera de forma importante ante la presencia de un valor extremo, como sucede con la desviación típica.

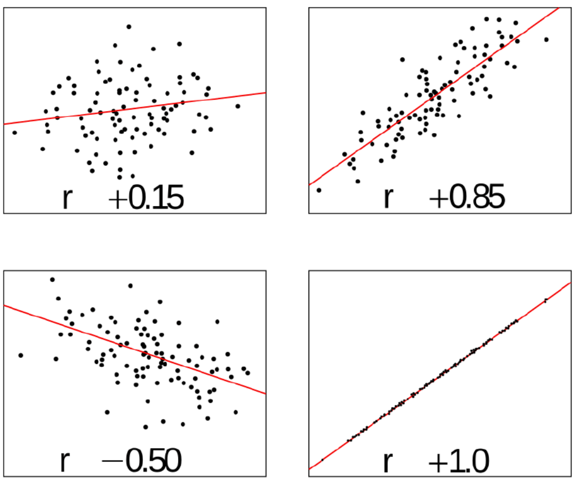
c. El coeficiente de correlación **mide solo la relación con una línea recta**. Dos variables pueden tener una relación curvilínea fuerte, a pesar de que su correlación sea pequeña. Por tanto, cuando analicemos las relaciones entre dos variables debemos representarlas gráficamente y luego calcular el coeficiente de correlación.

d. Mientras más grande sea la muestra, más exacta será la estimación.

e. **La correlación no implica causalidad**. La causalidad es un juicio de valor que requiere más información que un simple valor cuantitativo de un coeficiente de correlación.

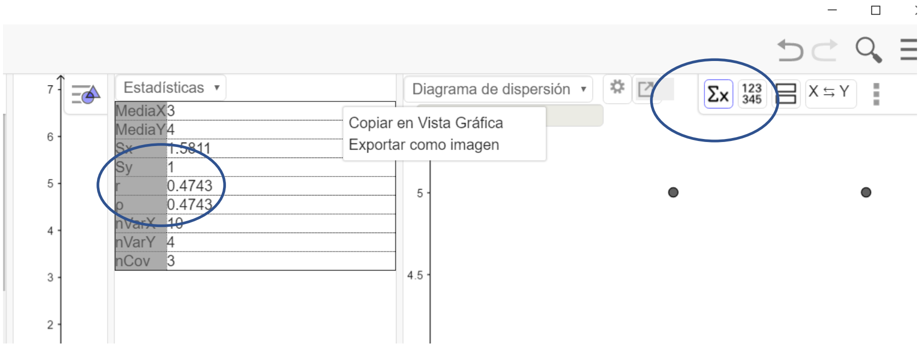
COEFICIENTE DE CORRELACIÓN Y DIAGRAMA DE DISPERSIÓN

En los diagramas de dispersión, podemos intuir el valor del Coeficiente de Pearson: a mayor dispersión de la nube de puntos, el coeficiente es cercano a 0, a mayor concentración, se acercará a valores 1 ó -1 en función de la tendencia de la nube:



-1	-0,7		-0,3	0	0,3		0,7	1
Negativa	Negativa	Negativa	Negativa	Incorreladas	Positiva	Positiva	Positiva	Positiva
perfecta	fuerte	moderada	débil		débil	moderada	fuerte	perfecta

En GeoGebra:



calculamos r.

Coeficiente de determinación (bondad de ajuste)

Al calcular el coeficiente de correlación no podemos distinguir entre la causa y el efecto. Cuando hay una correlación entre dos variables X e Y se podrá considerar cuatro explicaciones posibles:

- La correlación es espuria, depende de la muestra
- X es a causa de Y
- Y es la causa de X
- Una variable latente es la causa de la correlación

Para medir la proporción de variación que es explicada por la variable independiente, podemos calcular el coeficiente de determinación:

Coef. de Determinación: **CD = r ^{2} x 100**

Continuando con el ejemplo de altura y peso: el coeficiente de Pearson es = 0,91, y CD = 83,08%, es decir que el aumento de peso en función a la altura de la persona se explica en un 83,08%. El 16,92% restante esta explicado por otros factores.

Ejemplo: En la siguiente tabla se indica la cantidad de horas de TV de programas violentos por día que miran 10 niños y la conducta agresiva (en una escala de 0 a 10) de los mismos:

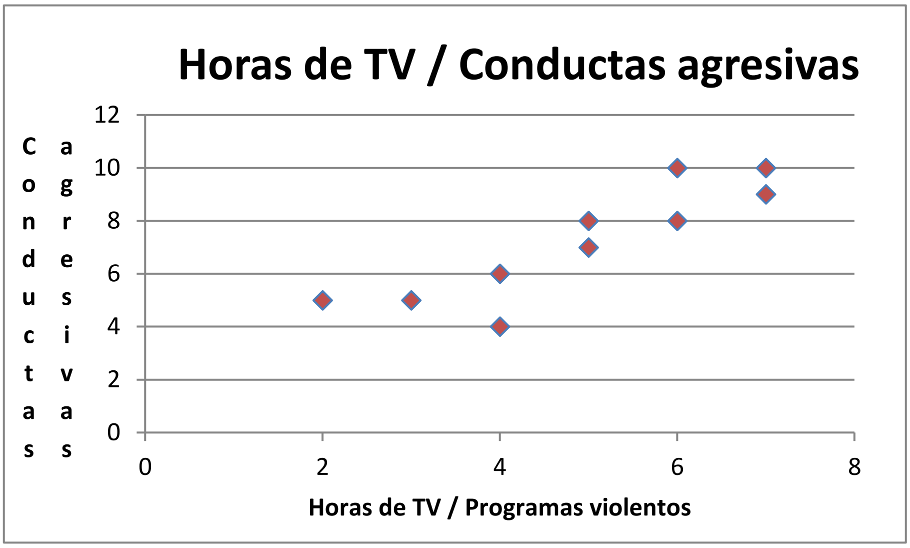
Horas de TV / Prog. Violentos	6	6	7	2	7	4	5	5	4	3
Conductas agresivas	10	8	10	5	9	6	7	8	4	5

Debemos

- Investigar si existe relación lineal entre las variables
- Determinar además la intensidad de la relación e interpretar en el contexto del problema

- ¿En qué proporción pueden justificarse las conductas agresivas de los niños mediante su edad?

Primero comenzamos por confeccionar el diagrama de dispersión: observando el mismo, podemos ver una relación lineal positiva, bastante fuerte. A medida que aumentan la cantidad de horas que están los niños mirando programas violentos aumentan la cantidad de conductas agresivas.



Para determinar la intensidad de la relación, calculamos el coeficiente r con la calculadora: r = 0,87.
El coeficiente de Person confirma lo visto en el gráfico, se trata de una relación fuerte, a mayor cantidad de horas, mayor cantidad de conductas violentas.
Para conocer la proporción que se justifica calculamos CD = 76,77%
El Coeficiente de determinación nos muestra que el aumento de la cantidad de conductas violentas se explica en un 76,77% por la cantidad de horas que los niños miran TV. El resto (23,22%) se debe a otros factores desconocidos.

Errores de interpretación

Primer ejemplo: Un periódico universitario entrevista a un psicólogo sobre las evaluaciones que hacen los estudiantes de sus profesores. El psicólogo afirma: la evidencia demuestra que la correlación entre la capacidad investigadora de los profesores y la evaluación docente que hacen los estudiantes es próxima a cero. El titular del periódico dice: El psicólogo dijo que los buenos investigadores tienden a ser malos profesores y viceversa.

¿Es correcta la afirmación del periódico?

Segundo ejemplo: ¿Qué tipo de error se cometió en cada caso?

- Existe una correlación alta entre el sexo de los trabajadores y sus ingresos
- Hallamos una correlación alta (r=1,09) entre las evaluaciones de los profesores hechas por los estudiantes y las hechas por los profesores
- La correlación hallada entre la densidad de siembra y el rendimiento del maíz fue de r= 0,23 hectolitros.

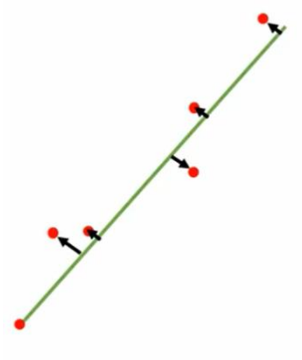
Regresión lineal simple

La regresión lineal es una técnica estadística que analiza si los valores de una variable a la que se denomina *dependiente* (Y), pueden predecirse mediante, un modelo lineal, conociendo los valores de otras variables a la que se denomina *independientes* (X).
El objetivo del análisis de regresión es predecir los valores de la variable respuesta (Y) en función de los valores de las variables independientes.

Si el número de variables independientes es una, la regresión es *simple,* en este caso la dependencia funcional de Y respecto de X, se expresa de la siguiente manera: $Y = f(X)$.
La **relación lineal simple** se expresa en general de la forma:

$Y = A + BX$, B se denomina coeficiente de regresión

El objetivo de estudio de la regresión es estimar A y B con el menor error posible. Dado que el cálculo de los coeficientes de A y B es tedioso, se puede realizar con calculadora. En los siguientes links encontrarán la forma de hacerlo con dos calculadoras distintas:
https://www.youtube.com/watch?v=M8YCXigXMa0
https://www.youtube.com/watch?v=4_WO31Dapv0
En un diagrama de dispersión podremos observar que existen infinitas rectas que pasen entre sus puntos, sin embargo, la recta de regresión es aquella que la suma de las distancias de todos los puntos a ella es la mínima:



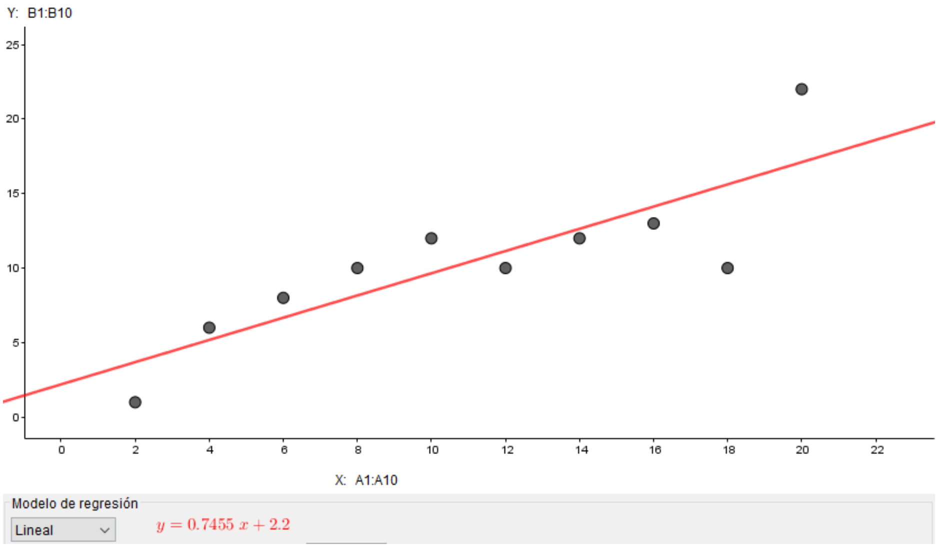
EJEMPLO

Consideremos las variables X e Y que corresponden a puntuaciones obtenidas en dos test aplicados a un grupo de niños, digamos el test 1 (X) y el test 2 (Y):
X: 2 4 6 8 10 12 14 16 18 20
Y: 1 6 8 10 12 10 12 13 10 22

El coeficiente de correlación de Person es $r = 0{,}8391$.
Calculemos el coeficiente de determinación: $r^2 = 0{,}8391^2 = 0{,}70408 \ldots$ *esto es* $70{,}408\%$ esto quiere decir que la puntuación obtenida en el test 2 se explica en un 70,408% por la puntuación obtenida en el test 1. El porcentaje restante se debe a otros factores.
Ahora hallemos A y B: $A = 2{,}2 \quad B = 0{,}7455$
Luego,

$Y = 2{,}2 + 0{,}7455X$

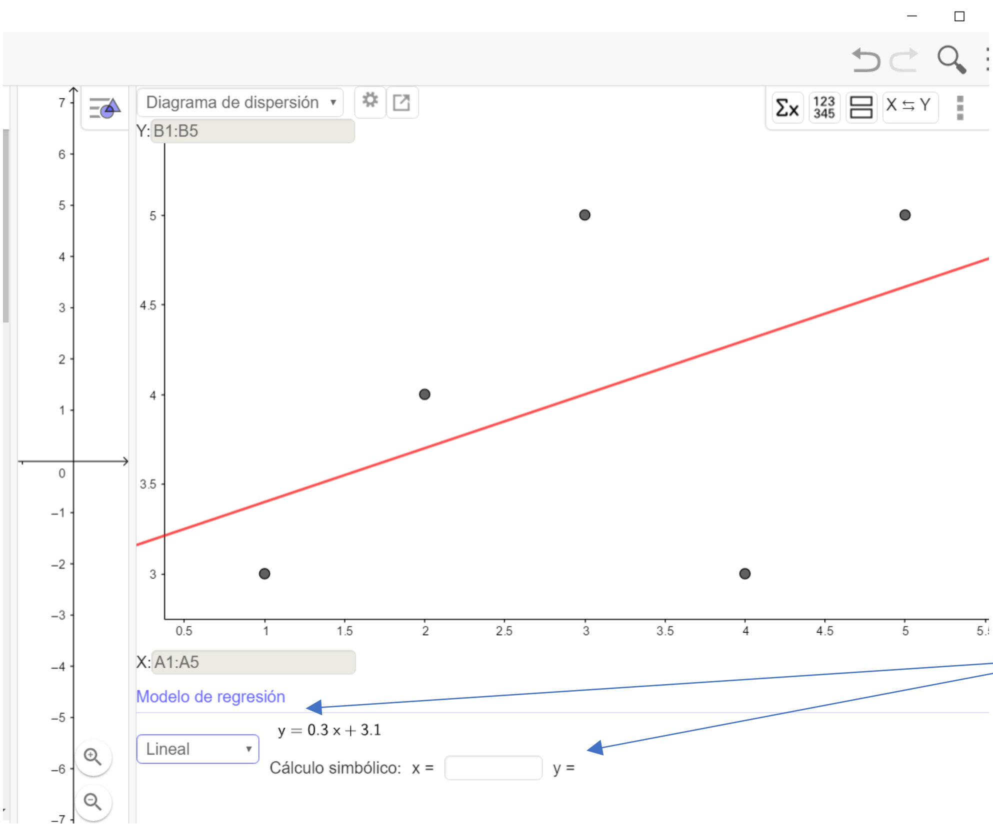
Gráficamente:



A partir de la fórmula podemos predecir (de modo estimativo), ¿qué puntaje se obtendrá en el test 2 si en el test 1 se obtuvo 19 puntos?

$$Y = 2{,}2 + 0{,}7455X = 2{,}2 + 0{,}7455.19 = 16{,}3645$$

En GeoGebra:



Elegimos la opción modelo de regresión lineal, obtenemos el gráfico, A y B. Además, podemos calcular y conociendo el valor de x

PRÁCTICA 4

En todos los ejercicios que puedas, aplicar GeoGebra para verificar tus respuestas

Actividad 1. En la siguiente tabla se muestran las puntuaciones recogidas a partir de una muestra de 27 sujetos en una escala observacional de estrés y en un test orientado a evaluar la utilización de mecanismos de afrontamiento. El rango de puntuaciones en ambas variables puede oscilar entre 0 a 100, significando las puntuaciones más altas un mayor estrés y una mayor capacidad de utilización de mecanismos de afrontamiento, respectivamente.

Caso	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Estrés	61	26	32	22	38	80	17	10	47	15	50	25	50	30	78	10	35	31	4	6	7	17	37	45	50	67	70
Afronta	38	80	40	84	62	18	65	78	22	60	50	58	20	45	19	84	63	43	87	84	83	85	35	15	29	28	35

a) ¿Puedes definir cuál es la variable dependiente y cuál es la variable independiente?

b) Dibuja un diagrama de dispersión. En base a lo que observas en dicho diagrama, ¿crees que existe relación entre las variables estrés y afrontamiento? Si es así, ¿qué tipo de relación existe?

c) Calcula el coeficiente de correlación. ¿Se corrobora la hipótesis planteada en el inciso b)?

Actividad 2. Los siguientes valores corresponden a distintos coeficientes de correlación:

0,55 0,09 -0,77 0,1 -0,12

a) ¿Cuál de estos coeficientes indica la relación más fuerte?

b) ¿Cuál de estos coeficientes indica la relación más débil?

c) Ordena los coeficientes (de menor a mayor) de acuerdo al grado de relación que indican.

Actividad 3. Si calculaste **r** y obtuviste **r = -1,3**, sabrías con certeza que:

a) la relación es extremadamente fuerte.

b) la relación es ciertamente negativa.

c) las dos anteriores.

d) se ha cometido un error de cálculo.

Actividad 4. Las calificaciones obtenidas por 10 alumnos al aplicarles dos tests son las siguientes:

Test 1	2	7	8	9	10	12	14	10	16	12
Test 2	5	8	10	12	12	14	15	16	18	20

El test 1 refleja la capacidad verbal y el 2 es un test manipulativo.

a) Realiza el diagrama de dispersión correspondiente y analiza el tipo de relación que se presenta.

b) ¿Sería adecuado calcular el coeficiente de correlación lineal para este par de variables? En el caso de que tu respuesta sea afirmativa, calcúlalo.

c) Calcula el valor del coeficiente de determinación.

d) Interpreta los resultados de los dos coeficientes en relación con el contexto planteado.

Actividad 5. Los datos dados en la tabla siguiente muestran la relación entre los coeficientes de inteligencia verbales y no verbales de la prueba de inteligencia de Lorge- Thorndike y el rendimiento en lectura y aritmética medido por la prueba de habilidades básicas (ITBS). En cada grado, cada correlación está basada en aproximadamente 2500 alumnos.

	CI verbal			CI no verbal		
Nivel de grado	**3**	**5**	**7**	**3**	**5**	**7**
Lectura	0,68	0,76	0,81	0,53	0,65	0,67
Aritmética	0,66	0,72	0,74	0,61	0,68	0,71

Basado en estos datos, ¿las siguientes afirmaciones son verdaderas o falsas? Justifica.

a) La correlación entre las medidas de inteligencia y el rendimiento parece incrementarse con el grado.

b) Los CI no verbales están correlacionados tan alto con el rendimiento como con los CI verbales.

c) Los CI verbal y no verbal tienden a correlacionarse ligeramente más alto con lectura que con aritmética.

d) La correlación entre ambas medidas de rendimiento y ambas medidas de inteligencia es fuerte en cada uno de los tres grados.

Actividad 6. En un estudio sociológico se postula que la actitud racista viene determinada fundamentalmente por el nivel de desocupación que existe en una sociedad. Para avalar dicha afirmación, el estudio proporciona los datos relativos a una muestra de 10 localidades, cuyos niveles de desocupación (expresado en porcentaje) y racismo (medido a partir de un índice) son los siguientes:

Desocupación (%)	7	13	5	23	33	21	18	30	15	27
Racismo	22	29	15	37	50	35	32	40	30	38

a) ¿Podría afirmarse que existe una relación lineal fuerte entre el nivel de desocupación y el índice de racismo?

b) ¿Se podría concluir que la variación en el nivel de racismo queda bien explicada por la variación en el nivel de desocupación?

Actividad 7. Un editor de un periódico universitario entrevista a un psicólogo y le pregunta sobre las evaluaciones que hacen los estudiantes de sus profesores. El psicólogo afirma: "*la evidencia demuestra que la correlación entre la capacidad investigadora de los profesores y la evaluación docente que hacen los estudiantes es próxima a cero*". El titular del periódico dice: "*el psicólogo Cruz dijo que los buenos investigadores tienden a ser malos profesores y viceversa*". Explica por qué el titular del periódico no refleja el sentido de las palabras del psicólogo Cruz.

Actividad 8. Cada una de las siguientes afirmaciones contiene un error. Explica en cada caso dónde está la incorrección.

a) "Existe una correlación alta entre el sexo de los trabajadores y sus ingresos"

b) "Hallamos una correlación alta (r = 1,09) entre las evaluaciones de los profesores hechas por los estudiantes y las hechas por los profesores".

c) "La correlación hallada entre el nivel de estrés y el tipo de dieta utilizada fue de r = 0,23 kilogramos"

Actividad 9. En una muestra de 10 alumnos de enseñanza secundaria se han medido dos variables: rendimiento en el curso, cuantificado como el promedio de las calificaciones de las asignaturas del curso (Y); y el promedio de horas de estudio semanal durante el curso, obtenido a partir de auto-informe de los propios estudiantes (X). Los datos obtenidos son los que se muestran a continuación:

X	5	12	7	9	15	10	12	8	18	14
Y	3	6	4	5	9	6	6	5	9	7

Tomando en cuenta el rendimiento en el curso y el tiempo de estudio semanal de estos 10 alumnos, ¿podríamos decir que se presenta una asociación fuerte entre ambas variables? Explica tu razonamiento.

Actividad 10. En los siguientes diagramas de dispersión indicar: si se presenta alguna relación, el tipo de relación, su intensidad aproximada y su dirección. ¿Se podría hablar en cada caso de una relación de causa y efecto?

Diagrama de dispersión 1

Peso en función de la altura

Gráfico de dispersión: altura vs. Peso

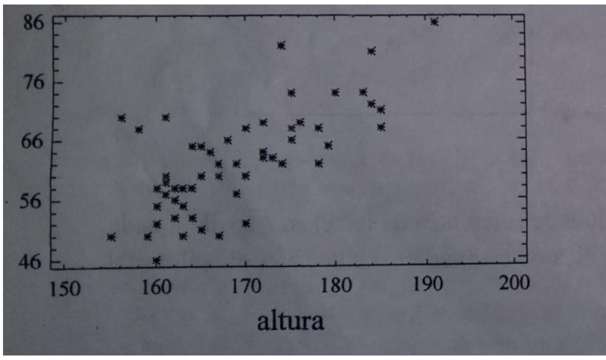


Diagrama de dispersión 2

Peso en función de la Talla

Gráfico de dispersión: Talla vs. Peso

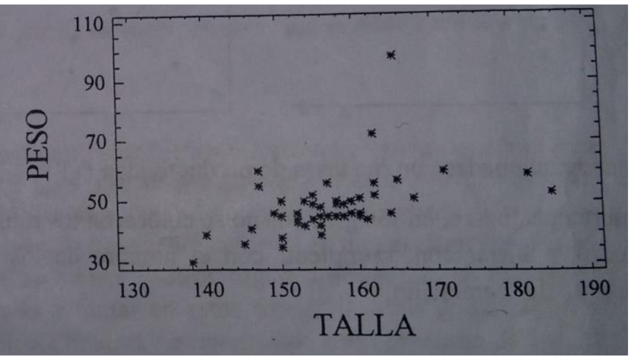
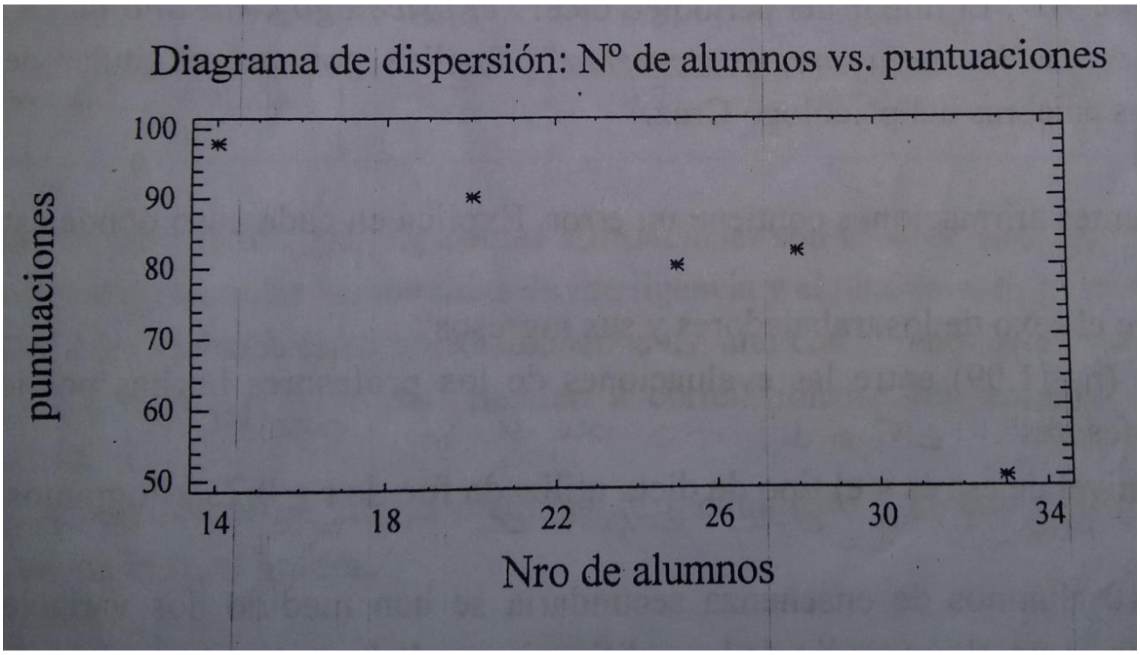


Diagrama de dispersión 3: Rendimiento escolar (expresado en puntuaciones en función de la cantidad de alumnos en la clase)

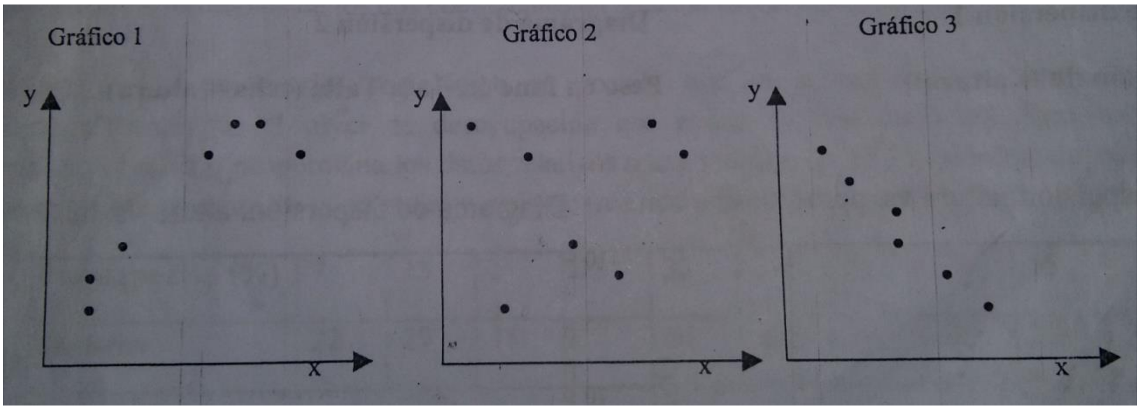


Actividad 11. A continuación se presentan tres afirmaciones referidas a las conclusiones de un estudio realizado acerca de las tasas de nacimiento, suicidio, crecimiento económico y productividad, junto con tres gráficos de dispersión.

Afirmación 1: En países con un desarrollo tecnológico alto tales como Japón, Estados Unidos, Alemania, Inglaterra, Francia, Italia y Canadá, se tienen bajas tasas de nacimiento (TN) asociadas con altas tasas de suicidio (TS).

Afirmación 2: Algunos economistas afirman que independientemente de los países que se estudian, a altas tasas de crecimiento (TC) se asocian altas tasas de productividad (TP).

Afirmación 3: Tanto economistas como demógrafos afirman que las tasas de suicidio (TS) no parecen estar correlacionados con las tasas de productividad (TP).



Desafortunadamente, en los gráficos no se colocaron los rótulos de referencia de los ejes. Asociar a cada gráfica una afirmación, completar con el nombre de los ejes y estimar en cada caso el valor del coeficiente de correlación.

Actividad 12. Se consideran los puntajes de un grupo de niños de 5to grado en una Prueba de Autoconcepto de Piers-Harris (PRUEBA1) y en el Inventario de Autoestima de CooperSmith (PRUEBA2), obteniéndose:

PRUEBA1 8 8 12 12 16 16 20 20 24 24

PRUEBA2 8 6 6 10 8 14 14 12 16 12

a) Confeccione el diagrama de dispersión. b) Determine, mediante un coeficiente adecuado, el grado de ajuste del modelo lineal a los datos. Interprete su resultado. c) Determine, mediante un coeficiente adecuado, el grado y sentido de la asociación lineal entre ambas variables. Interprete su resultadod) Halle la recta de regresión para predecir el puntaje en PRUEBA2, a partir del puntaje obtenido en PRUEBA1. Represente gráficamente. e) Prediga el puntaje en PRUEBA2 para un niño que obtuvo 10 puntos en PRUEBA1.

Actividad 13. Un psicólogo desea relacionar los puntajes en un test de Inteligencia (X), con los puntajes de otro que mide el conformismo (Y). Administró ambos tests a un grupo de sujetos y obtuvo:

X 6 10 5 8 8 7 9 4 7 4

Y 8 1 11 5 4 6 3 11 6 12

a) Confeccione el diagrama de dispersión. b) Determine mediante un coeficiente adecuado el grado y el sentido de la relación lineal entre ambas variables. Interprete el resultado. c) Determine mediante un coeficiente adecuado el grado de ajuste del modelo lineal a los datos. Interprete el resultado. d) Halle la recta de regresión para predecir el puntaje en el test de inteligencia conocido el puntaje obtenido en el test de conformismo. Obtenga la representación gráfica. e) Prediga el puntaje en la prueba de inteligencia de Pablo que obtuvo 2 en la prueba de conformismo.

Actividad 14.

A cincuenta niños de 3er grado, elegidos aleatoriamente dentro de un determinado Distrito Escolar les fueron administradas dos pruebas: una de dibujo y otra de imaginación creadora. Con esos pares de valores se calcularon las rectas de regresión, a saber:

Regresión del puntaje en dibujo sobre el de imaginación creadora: $Y' = -10 + 1,95*X$

Regresión del puntaje en imaginación creadora sobre el de dibujo: $X' = 7,09 + 0,3759*Y$

a) ¿Qué puntaje en dibujo puede esperarse de un niño cuyo puntaje en imaginación creadora fue 7?.

b) ¿Qué porcentaje de la variabilidad del puntaje en dibujo es explicado por el de imaginación creadora?

c) ¿Cuál es el sentido de la relación entre estas variables? Justifique su respuesta.

PRÁCTICA COMPLEMENTARIA A LAS PRÁCTICAS A LAS 3 Y 4

En todos los ejercicios que puedas, aplicar GeoGebra para verificar tus respuestas

Actividad 1. Tras un test de cultura general se observa que las puntuaciones obtenidas siguen una distribución normal con media 65 y desvío 18. Se desea clasificar a los examinados en tres grupos (de baja cultura general, de cultura general aceptable, de excelente cultura general) de modo que hay en el primero un 20% la población, un 65% el segundo y un 15% en el tercero.

a) ¿Cuáles han de ser las puntuaciones que marcan el paso de un grupo al otro? Justificar.

b) Calcular la probabilidad de que una persona seleccionada al azar obtenga una puntuación entre 75 y 85.

c) ¿Es poco probable que una persona obtenga menos de 1 punto en el test? Justificar.

Actividad 2. Un agente de seguros contrata 5 pólizas con personas de la misma edad y de buena salud. Según las tablas en uso, la probabilidad de que una persona de esa edad esté viva dentro de 30 años es 2/3. Hallar la probabilidad de que dentro de 30 años vivan:

a) Las 5 personas.

b) Al menos 4.

c) A lo sumo una.

Actividad 3. Se ha medido la motivación ante el estudio a 12 sujetos, antes y después de participar en un programa de innovación didáctica, obteniéndose los siguientes datos:

Sujeto	1	2	3	4	5	6	7	8	9	10	11	12
Pre-prueba	54	56	66	48	46	60	55	57	51	43	56	52
Post-prueba	68	61	55	56	58	65	68	58	49	66	62	65

a) ¿Cuáles son las variables involucradas?

b) Determine si existe relación lineal entre ellas. Justifique.

c) ¿Cómo explicarías la influencia de una variable sobre la otra?

Actividad 4. Los 200 alumnos de Primer Año de un instituto se distribuyen por sexo y modalidad de la forma siguiente:

Modalidad	Alumnos (H)	Alumnas (M)	Total
Ciencias (C)		50	110
Letras (L)			
Total	88		200

a) Completar la tabla.

b) Hallar la probabilidad de que:

ii) estudie letras

iii) estudie ciencias sabiendo que es mujer

iv) sea hombre y estudie letras

v) sea hombre o estudie ciencias.

Actividad 5. Las temperaturas máximas reales (en grados Fahrenheit) durante septiembre se describen con el siguiente resumen de los cinco números: 62, 72, 76, 80, 85. Utilizando esos valores y suponiendo que los datos corresponden una distribución normal con desviación 2.2, responder:

a. ¿Cuál es la mediana? Justificar.

b. Si se calcula una temperatura máxima de algún día seleccionado al azar en el mes de septiembre, calcule la probabilidad de que oscile entre 72 y 76°F.
c. Si se tiene una temperatura máxima de algún día seleccionado al azar en el mes de septiembre, calcule la probabilidad de que sea menor que 72°F o mayor que 76°F.
d. Si se seleccionan al azar dos días diferentes de septiembre, calcule la probabilidad de que ambos días tengan temperaturas máximas entre 72 y 76°F, suponiendo independencia de los sucesos.

Actividad 6. Un taller sabe que por término medio acuden: por la mañana 3 automóviles con problemas eléctricos, 8 con problemas mecánicos y 3 con problemas de chapa, y por la tarde 2 con problemas eléctricos, 3 con problemas mecánicos y 1 con problemas de chapa. Calcula la probabilidad de que un automóvil:

- **a.** acuda por la tarde.
- **b.** acuda por problemas mecánicos.
- **c.** que tiene problemas eléctricos acuda por la mañana.

Actividad 7. Se realizó un estudio para determinar si existían diferencias significativas entre estudiantes de medicina aceptados a través de programas especiales (como el de acción afirmativa) y estudiantes de medicina aceptados a través de los criterios regulares de admisión. Se encontró que el 94% de los estudiantes de medicina aceptados a través de programas especiales se graduaron (según datos del *Journal of the American Medical Association*).
a. Si se seleccionan al azar 10 de los estudiantes de los programas especiales, calcule la probabilidad de que al menos nueve se gradúen.
b. ¿Sería infrecuente que de 10 estudiantes de los programas especiales, seleccionados al azar, solamente se graduaran siete? ¿Por qué?

Actividad 8. En una población se sabe que el 30% escucha los informativos por la radio; el 60% por la televisión; y el 20% los escucha por los dos medios de comunicación. Si se elige una persona al azar, determina la probabilidad de que:
a) escuche alguno de los medios de comunicación.
b) escuche la radio y que no ve la televisión.
c) escuche sólo uno de los dos medios.

Actividad 9. La tabla siguiente muestra las notas obtenidas por 8 alumnos en un examen, las horas de estudio dedicadas a su preparación y las horas que vieron la televisión los días previos al examen.

Nota	5	6	7	3	5	8	4	9
Horas de estudio	7	10	9	4	8	10	5	14
Horas de TV	7	6	2	11	9	3	9	5

a) Representa gráficamente los diagramas correspondientes a nota-estudio y nota-TV.
b) ¿Se observa correlación entre las variables estudiadas? ¿De qué tipo? ¿En qué caso estimas que es más fuerte? ¿En qué porcentaje influyen las horas de estudio y las horas de TV en las notas obtenidas?
c) ¿Es posible calcular el coeficiente de correlación? Si es así, ¿se ratifica lo respondido en b)?

Actividad 10: El grado de satisfacción de un individuo al finalizar cierta tarea se mide mediante un índice que sigue una distribución normal con media de 40 puntos y desviación de 10 puntos. En este contexto podemos clasificar la población en estudio como insatisfecha si el citado índice es menor que 35, satisfecha si está entre 35 y 55 y, muy satisfecha si el índice es mayor que 55.

- **a.** ¿Qué porcentaje de satisfechos podemos decir que hay en esa población?
- **b.** ¿Cuál es la probabilidad de encontrar una persona, seleccionada al azar, que esté muy satisfecha?

Actividad 11. Un estudio reciente reveló que el 64% de las mujeres mayores de 18 años, consideran a la nutrición la prioridad en su vida. Se seleccionó una muestra de 6 mujeres. Determinar la probabilidad de que:

- **c.** a) 3 o más consideren importante la dieta diaria.
- **d.** b) Más de 3 pero menos de 4 consideren importante el aspecto dietético.
- **e.** c) Exactamente 5 no consideren fundamental la alimentación.

Actividad 12. Una empresa ha encontrado que la duración de sus llamadas telefónicas a larga distancia, tiene aproximadamente una distribución normal, con media de 3 minutos y desviación típica de 3 minutos.

a) ¿En qué proporción las llamadas a larga distancia tienen una duración de más de 2 minutos, pero de menos de 3 y medio minutos?

b) Una secretaria va a hacer una llamada a larga distancia. ¿Cuál es la probabilidad de que dure más de 5 minutos?

Actividad 13. En una ciudad, el 40% de la población tiene cabellos castaños, el 25% tiene ojos castaños y el 15% tiene cabellos y ojos castaños. Se escoge una persona al azar:

a) Si tiene los cabellos castaños, ¿cuál es la probabilidad de que tenga también ojos castaños?

b) Si tiene ojos castaños, ¿cuál es la probabilidad de que no tenga cabellos castaños?

c) ¿Cuál es la probabilidad de que no tenga cabellos ni ojos castaños?

Actividad 14. A dos pacientes de un hospital, un hombre y una mujer, se les ha diagnosticado una fobia severa. Los psicólogos pronostican que con el tratamiento prescripto la probabilidad de cura para el hombre es de 50% y para la mujer del 60%, Si se consideran las dos situaciones completamente independientes calcular la probabilidad:

a. Que los dos se curen

b. Que el hombre no supere la fobia

c. Que la mujer no se cure

d. Que ni el hombre ni la mujer se curen.

Actividad 15 Un investigador estaba interesado en le relación entre el grado de empatía de los psicoterapeutas y el nivel de satisfacción de sus pacientes con la terapia. Como estudio piloto se analizaron cuatro parejas de terapeuta y pacientes. Estos son los resultados:

Nº de pareja	Empatía del terapeuta	Satisfacción del paciente
1	70	4
2	94	5
3	36	2
4	48	1

a. A través de un diagrama de dispersión, determine si existe algún tipo de relación entre las variables.

b. Confirme lo expuesto en a) calculando alguna medida apropiada e interpretándola.
c. ¿En qué proporción la satisfacción del paciente es explicada por la empatía del terapeuta?

Actividad 16 El grado de satisfacción de un individuo al finalizar cierta tarea se mide mediante un índice que sigue una distribución normal con media de 40 puntos y desviación de 10 puntos. En este contexto podemos clasificar la población en estudio como insatisfecha si el citado índice es menor que 35, satisfecha si está entre 35 y 55 y, muy satisfecha si el índice es mayor que 55.

a. ¿Qué porcentaje de satisfechos podemos decir que hay en esa población?
b. ¿Cuál es la probabilidad de encontrar una persona, seleccionada al azar, que esté muy satisfecha?

Actividad 17 Responde si las siguientes afirmaciones son verdaderas o falsas y justifica en todos los casos. No se considerarán válidas aquellas respuestas que sólo consignen si es verdadero o falso y no tengan justificaciones

a) Para poder utilizar una distribución binomial en el cálculo de probabilidades, se debe verificar solamente que la variable sea cuantitativa y que exista una probabilidad de éxito.
b) Para describir el comportamiento de un lote de datos basta con informar sobre alguna de las medidas de tendencia central.
c) La desviación estándar de un lote de datos puede ser un valor negativo.
d) Un diagrama de caja que tenga la mediana a igual distancia de los cuartiles 1 y 3, y con “colas” o “bigotes” de la misma longitud, representa siempre a una distribución asimétrica.
e) Se ha realizado un estudio sobre la cantidad de horas semanales que miran TV un grupo de niños de 6 años con el fin de analizar si el tiempo que miran TV tiene relación con la cantidad de acciones violentas o agresivas hacia sus compañeros de colegio. A partir de los datos obtenidos, se calculó el coeficiente de correlación entre las dos variables estudiadas y se obtuvo r = 0,92. A partir de estos datos responde, para cada sub-item, si es verdadero o falso y justifica:

i. La variable dependiente es: “cantidad de horas semanales que mira TV cada chico”.
ii. La variable independiente es cuantitativa continua con escala de intervalo.
iii. Existe una relación débil entre la cantidad de horas que mira TV cada chico y el nivel de agresión hacia sus compañeros.
iv. El 84,64 % del nivel de agresividad de los niños está explicado por la cantidad de horas de TV que miran por semana.
v. El encargado de brindar información sobre este estudio enunció lo siguiente: *“Se puede decir que la relación entre las variables observadas es lineal e inversa, por lo que cuanto más horas de TV miran los chicos menor es el nivel de agresividad”*

Actividad 18. La siguiente tabla resume los resultados de 985 muertes de peatones, causadas por accidentes.

	Peatón intoxicado	Peatón no intoxicado
Conductor intoxicado	59	79
Conductor no intoxicado	266	581

a) Determina si los sucesos “conductor intoxicado” y “peatón intoxicado” son independientes.
b) Calcula las siguientes probabilidades:
   i) de que ambos no estén intoxicados.
   ii) de que el conductor esté intoxicado.
   iii) de que el peatón no esté intoxicado sabiendo que el conductor sí lo está.
c) Da ejemplos de un suceso simple, un suceso compuesto y un suceso imposible, relacionados con el experimento.

ANEXO

Tabla binomial

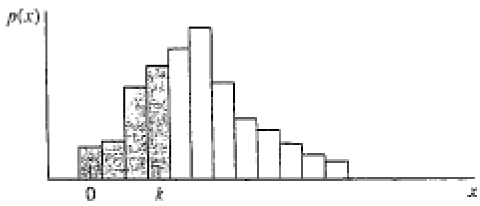


TABLA 1 Probabilidades binomiales acumuladas

Los valores tabulados son $P(x \leq k) = p(0) + p(1) + \cdots + p(k)$

(Los cálculos se redondean en la tercera cifra decimal.)

$n = 2$

k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.980	.902	.810	.640	.490	.360	.250	.160	.090	.040	.010	.002	.000	0
1	1.000	.998	.990	.960	.910	.840	.750	.640	.510	.360	.190	.098	.020	1
2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	2

$n = 3$

k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.970	.857	.729	.512	.343	.216	.125	.064	.027	.008	.001	.000	.000	0
1	1.000	.993	.972	.896	.784	.648	.500	.352	.216	.104	.028	.007	.000	1
2	1.000	1.000	.999	.992	.973	.936	.875	.784	.657	.488	.271	.143	.030	2
3	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	3

$n = 4$

k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.961	.815	.656	.410	.240	.130	.062	.026	.008	.002	.000	.000	.000	0
1	.999	.986	.948	.819	.652	.475	.312	.179	.084	.027	.004	.000	.000	1
2	1.000	1.000	.996	.973	.916	.821	.688	.525	.348	.181	.052	.014	.001	2
3	1.000	1.000	1.000	.998	.992	.974	.938	.870	.760	.590	.344	.185	.039	3
4	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	4

$n = 5$

							p							
k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.951	.774	.590	.328	.168	.078	.031	.010	.002	.000	.000	.000	.000	0
1	.999	.977	.919	.737	.528	.337	.188	.087	.031	.007	.000	.000	.000	1
2	1.000	.999	.991	.942	.837	.683	.500	.317	.163	.058	.009	.001	.000	2
3	1.000	1.000	1.000	.993	.969	.913	.812	.663	.472	.263	.081	.023	.001	3
4	1.000	1.000	1.000	1.000	.998	.990	.969	.922	.832	.672	.410	.226	.049	4
5	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	5

$n = 6$

							p							
k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.941	.735	.531	.262	.118	.047	.016	.004	.001	.000	.000	.000	.000	0
1	.999	.967	.886	.655	.420	.233	.109	.041	.011	.002	.000	.000	.000	1
2	1.000	.998	.984	.901	.744	.544	.344	.179	.070	.017	.001	.000	.000	2
3	1.000	1.000	.999	.983	.930	.821	.656	.456	.256	.099	.016	.002	.000	3
4	1.000	1.000	1.000	.998	.989	.959	.891	.767	.580	.345	.114	.033	.001	4
5	1.000	1.000	1.000	1.000	.999	.996	.984	.953	.882	.738	.469	.265	.059	5
6	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	6

$n = 7$

							p							
k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.932	.698	.478	.210	.082	.028	.008	.002	.000	.000	.000	.000	.000	0
1	.998	.956	.850	.577	.329	.159	.062	.019	.004	.000	.000	.000	.000	1
2	1.000	.996	.974	.852	.647	.420	.227	.096	.029	.005	.000	.000	.000	2
3	1.000	1.000	.997	.967	.874	.710	.500	.290	.126	.033	.003	.000	.000	3
4	1.000	1.000	1.000	.995	.971	.904	.773	.580	.353	.148	.026	.004	.000	4
5	1.000	1.000	1.000	1.000	.996	.981	.938	.841	.671	.423	.150	.044	.002	5
6	1.000	1.000	1.000	1.000	1.000	.998	.992	.972	.918	.790	.522	.302	.068	6
7	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	7

$n = 8$

							p							
k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.923	.663	.430	.168	.058	.017	.004	.001	.000	.000	.000	.000	.000	0
1	.997	.943	.813	.503	.255	.106	.035	.009	.001	.000	.000	.000	.000	1
2	1.000	.994	.962	.797	.552	.315	.145	.050	.011	.001	.000	.000	.000	2
3	1.000	1.000	.995	.944	.806	.594	.363	.174	.058	.010	.000	.000	.000	3
4	1.000	1.000	1.000	.990	.942	.826	.637	.406	.194	.056	.005	.000	.000	4
5	1.000	1.000	1.000	.999	.989	.950	.855	.685	.448	.203	.038	.006	.000	5
6	1.000	1.000	1.000	1.000	.999	.991	.965	.894	.745	.497	.187	.057	.003	6
7	1.000	1.000	1.000	1.000	1.000	.999	.996	.983	.942	.832	.570	.337	.077	7
8	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	8

$n = 9$

k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.914	.630	.387	.134	.040	.010	.002	.000	.000	.000	.000	.000	.000	0
1	.997	.929	.775	.436	.196	.071	.020	.004	.000	.000	.000	.000	.000	1
2	1.000	.992	.947	.738	.463	.232	.090	.025	.004	.000	.000	.000	.000	2
3	1.000	.999	.992	.914	.730	.483	.254	.099	.025	.003	.000	.000	.000	3
4	1.000	1.000	.999	.980	.901	.733	.500	.267	.099	.020	.001	.000	.000	4
5	1.000	1.000	1.000	.997	.975	.901	.746	.517	.270	.086	.008	.001	.000	5
6	1.000	1.000	1.000	1.000	.996	.975	.910	.768	.537	.262	.053	.008	.000	6
7	1.000	1.000	1.000	1.000	1.000	.996	.980	.929	.804	.564	.225	.071	.003	7
8	1.000	1.000	1.000	1.000	1.000	1.000	.998	.990	.960	.866	.613	.370	.086	8
9	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	9

$n = 10$

k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.904	.599	.349	.107	.028	.006	.001	.000	.000	.000	.000	.000	.000	0
1	.996	.914	.736	.376	.149	.046	.011	.002	.000	.000	.000	.000	.000	1
2	1.000	.988	.930	.678	.383	.167	.055	.012	.002	.000	.000	.000	.000	2
3	1.000	.999	.987	.879	.650	.382	.172	.055	.011	.001	.000	.000	.000	3
4	1.000	1.000	.998	.967	.850	.633	.377	.166	.047	.006	.000	.000	.000	4
5	1.000	1.000	1.000	.994	.953	.834	.623	.367	.150	.033	.002	.000	.000	5
6	1.000	1.000	1.000	.999	.989	.945	.828	.618	.350	.121	.013	.001	.000	6
7	1.000	1.000	1.000	1.000	.998	.988	.945	.833	.617	.322	.070	.012	.000	7
8	1.000	1.000	1.000	1.000	1.000	.998	.989	.954	.851	.624	.264	.086	.004	8
9	1.000	1.000	1.000	1.000	1.000	1.000	.999	.994	.972	.893	.651	.401	.096	9
10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	10

$n = 11$

k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.895	.569	.314	.086	.020	.004	.000	.000	.000	.000	.000	.000	.000	0
1	.995	.898	.697	.322	.113	.030	.006	.001	.000	.000	.000	.000	.000	1
2	1.000	.985	.910	.617	.313	.119	.033	.006	.001	.000	.000	.000	.000	2
3	1.000	.998	.981	.839	.570	.296	.113	.029	.004	.000	.000	.000	.000	3
4	1.000	1.000	.997	.950	.790	.533	.274	.099	.022	.002	.000	.000	.000	4
5	1.000	1.000	1.000	.988	.922	.754	.500	.246	.078	.012	.000	.000	.000	5
6	1.000	1.000	1.000	.998	.978	.901	.726	.467	.210	.050	.003	.000	.000	6
7	1.000	1.000	1.000	1.000	.996	.971	.887	.704	.430	.161	.019	.002	.000	7
8	1.000	1.000	1.000	1.000	.999	.994	.967	.881	.687	.383	.090	.015	.000	8
9	1.000	1.000	1.000	1.000	1.000	.999	.994	.970	.887	.678	.303	.102	.005	9
10	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.996	.980	.914	.686	.431	.105	10
11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	11

$n = 12$

k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.886	.540	.282	.069	.014	.002	.000	.000	.000	.000	.000	.000	.000	0
1	.994	.882	.659	.275	.085	.020	.003	.000	.000	.000	.000	.000	.000	1
2	1.000	.980	.889	.558	.253	.083	.019	.003	.000	.000	.000	.000	.000	2
3	1.000	.998	.974	.795	.493	.225	.073	.015	.002	.000	.000	.000	.000	3
4	1.000	1.000	.996	.927	.724	.438	.194	.057	.009	.001	.000	.000	.000	4
5	1.000	1.000	.999	.981	.882	.665	.387	.158	.039	.004	.000	.000	.000	5
6	1.000	1.000	1.000	.996	.961	.842	.613	.335	.118	.019	.001	.000	.000	6
7	1.000	1.000	1.000	.999	.991	.943	.806	.562	.276	.073	.004	.000	.000	7
8	1.000	1.000	1.000	1.000	.998	.985	.927	.775	.507	.205	.026	.002	.000	8
9	1.000	1.000	1.000	1.000	1.000	.997	.981	.917	.747	.442	.111	.020	.000	9
10	1.000	1.000	1.000	1.000	1.000	1.000	.997	.980	.915	.725	.341	.118	.006	10
11	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.986	.931	.718	.460	.114	11
12	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	12

$n = 15$

k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.860	.463	.206	.035	.005	.000	.000	.000	.000	.000	.000	.000	.000	0
1	.990	.829	.549	.167	.035	.005	.000	.000	.000	.000	.000	.000	.000	1
2	1.000	.964	.816	.398	.127	.027	.004	.000	.000	.000	.000	.000	.000	2
3	1.000	.995	.944	.648	.297	.091	.018	.002	.000	.000	.000	.000	.000	3
4	1.000	.999	.987	.836	.515	.217	.059	.009	.001	.000	.000	.000	.000	4
5	1.000	1.000	.998	.939	.722	.403	.151	.034	.004	.000	.000	.000	.000	5
6	1.000	1.000	1.000	.982	.869	.610	.304	.095	.015	.001	.000	.000	.000	6
7	1.000	1.000	1.000	.996	.950	.787	.500	.213	.050	.004	.000	.000	.000	7
8	1.000	1.000	1.000	.999	.985	.905	.696	.390	.131	.018	.000	.000	.000	8
9	1.000	1.000	1.000	1.000	.996	.966	.849	.597	.278	.061	.002	.000	.000	9
10	1.000	1.000	1.000	1.000	.999	.991	.941	.783	.485	.164	.013	.001	.000	10
11	1.000	1.000	1.000	1.000	1.000	.998	.982	.909	.703	.352	.056	.005	.000	11
12	1.000	1.000	1.000	1.000	1.000	1.000	.996	.973	.873	.602	.184	.036	.000	12
13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.995	.965	.833	.451	.171	.010	13
14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.995	.965	.794	.537	.140	14
15	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	15

$n = 20$

k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.818	.358	.122	.012	.001	.000	.000	.000	.000	.000	.000	.000	.000	0
1	.983	.736	.392	.069	.008	.001	.000	.000	.000	.000	.000	.000	.000	1
2	.999	.925	.677	.206	.035	.004	.000	.000	.000	.000	.000	.000	.000	2
3	1.000	.984	.867	.411	.107	.016	.001	.000	.000	.000	.000	.000	.000	3
4	1.000	.997	.957	.630	.238	.051	.006	.000	.000	.000	.000	.000	.000	4
5	1.000	1.000	.989	.804	.416	.126	.021	.002	.000	.000	.000	.000	.000	5
6	1.000	1.000	.998	.913	.608	.250	.058	.006	.000	.000	.000	.000	.000	6
7	1.000	1.000	1.000	.968	.772	.416	.132	.021	.001	.000	.000	.000	.000	7
8	1.000	1.000	1.000	.990	.887	.596	.252	.057	.005	.000	.000	.000	.000	8
9	1.000	1.000	1.000	.997	.952	.755	.412	.128	.017	.001	.000	.000	.000	9
10	1.000	1.000	1.000	.999	.983	.872	.588	.245	.048	.003	.000	.000	.000	10
11	1.000	1.000	1.000	1.000	.995	.943	.748	.404	.113	.010	.000	.000	.000	11
12	1.000	1.000	1.000	1.000	.999	.979	.868	.584	.228	.032	.000	.000	.000	12
13	1.000	1.000	1.000	1.000	1.000	.994	.942	.750	.392	.087	.002	.000	.000	13
14	1.000	1.000	1.000	1.000	1.000	.998	.979	.874	.584	.196	.011	.000	.000	14
15	1.000	1.000	1.000	1.000	1.000	1.000	.994	.949	.762	.370	.043	.003	.000	15
16	1.000	1.000	1.000	1.000	1.000	1.000	.999	.984	.893	.589	.133	.016	.000	16
17	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.996	.965	.794	.323	.075	.001	17
18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.992	.931	.608	.264	.017	18
19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.988	.878	.642	.182	19
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	20

$n = 25$

							p							
k	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99	k
0	.778	.277	.072	.004	.000	.000	.000	.000	.000	.000	.000	.000	.000	0
1	.974	.642	.271	.027	.002	.000	.000	.000	.000	.000	.000	.000	.000	1
2	.998	.873	.537	.098	.009	.000	.000	.000	.000	.000	.000	.000	.000	2
3	1.000	.966	.764	.234	.033	.002	.000	.000	.000	.000	.000	.000	.000	3
4	1.000	.993	.902	.421	.090	.009	.000	.000	.000	.000	.000	.000	.000	4
5	1.000	.999	.967	.617	.193	.029	.002	.000	.000	.000	.000	.000	.000	5
6	1.000	1.000	.991	.780	.341	.074	.007	.000	.000	.000	.000	.000	.000	6
7	1.000	1.000	.998	.891	.512	.154	.022	.001	.000	.000	.000	.000	.000	7
8	1.000	1.000	1.000	.953	.677	.274	.054	.004	.000	.000	.000	.000	.000	8
9	1.000	1.000	1.000	.983	.811	.425	.115	.013	.000	.000	.000	.000	.000	9
10	1.000	1.000	1.000	.994	.902	.586	.212	.034	.002	.000	.000	.000	.000	10
11	1.000	1.000	1.000	.998	.956	.732	.345	.078	.006	.000	.000	.000	.000	11
12	1.000	1.000	1.000	1.000	.983	.846	.500	.154	.017	.000	.000	.000	.000	12
13	1.000	1.000	1.000	1.000	.994	.922	.655	.268	.044	.002	.000	.000	.000	13
14	1.000	1.000	1.000	1.000	.998	.966	.788	.414	.098	.006	.000	.000	.000	14
15	1.000	1.000	1.000	1.000	1.000	.987	.885	.575	.189	.017	.000	.000	.000	15
16	1.000	1.000	1.000	1.000	1.000	.996	.946	.726	.323	.047	.000	.000	.000	16
17	1.000	1.000	1.000	1.000	1.000	.999	.978	.846	.488	.109	.002	.000	.000	17
18	1.000	1.000	1.000	1.000	1.000	1.000	.993	.926	.659	.220	.009	.000	.000	18
19	1.000	1.000	1.000	1.000	1.000	1.000	.998	.971	.807	.383	.033	.001	.000	19
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.991	.910	.579	.098	.007	.000	20
21	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.967	.766	.236	.034	.000	21
22	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.991	.902	.463	.127	.002	22
23	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.973	.729	.358	.026	23
24	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.996	.928	.723	.222	24
25	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	25

Tabla Normal

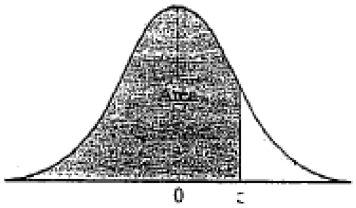


TABLA 3 Áreas bajo la curva normal

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

TABLA 3 *(continuación)*

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Tabla t de student

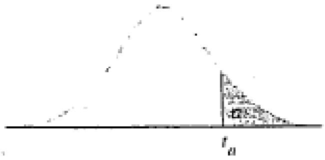


TABLA 4
Valores críticos de t

gl	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	gl
1	3.078	6.314	12.706	31.821	63.657	1
2	1.886	2.920	4.303	6.965	9.925	2
3	1.638	2.353	3.182	4.541	5.841	3
4	1.533	2.132	2.776	3.747	4.604	4
5	1.476	2.015	2.571	3.365	4.032	5
6	1.440	1.943	2.447	3.143	3.707	6
7	1.415	1.895	2.365	2.998	3.499	7
8	1.397	1.860	2.306	2.896	3.355	8
9	1.383	1.833	2.262	2.821	3.250	9
10	1.372	1.812	2.228	2.764	3.169	10
11	1.363	1.796	2.201	2.718	3.106	11
12	1.356	1.782	2.179	2.681	3.055	12
13	1.350	1.771	2.160	2.650	3.012	13
14	1.345	1.761	2.145	2.624	2.977	14
15	1.341	1.753	2.131	2.602	2.947	15
16	1.337	1.746	2.120	2.583	2.921	16
17	1.333	1.740	2.110	2.567	2.898	17
18	1.330	1.734	2.101	2.552	2.878	18
19	1.328	1.729	2.093	2.539	2.861	19
20	1.325	1.725	2.086	2.528	2.845	20
21	1.323	1.721	2.080	2.518	2.831	21
22	1.321	1.717	2.074	2.508	2.819	22
23	1.319	1.714	2.069	2.500	2.807	23
24	1.318	1.711	2.064	2.492	2.797	24
25	1.316	1.708	2.060	2.485	2.787	25
26	1.315	1.706	2.056	2.479	2.779	26
27	1.314	1.703	2.052	2.473	2.771	27
28	1.313	1.701	2.048	2.467	2.763	28
29	1.311	1.699	2.045	2.462	2.756	29
∞	1.282	1.645	1.960	2.326	2.576	∞

FUENTE: de "Table of Percentage Points of the *t*-Distribution", *Biometrika* 32 (1941):300. Reproducida con permiso del consejo de administración de *Biometrika*.

Regla de Sturges:

$k = 1 + 3{,}322 \log(n)$

Siendo k la cantidad de intervalos y n el total de datos en la muestra.

$A = \frac{máx - mín}{k}$.

A es el ancho del intervalo. Máx y mín, máximo y mínimo de la muestra.

N° máximo de líneas en el diagrama de tallo y hoja

$$10.\log(\mathrm{n})$$

Medidas estadísticas:

Promedio o media aritmética ($\bar{x}$):

- Datos sin agrupar: $\bar{x} = \frac{1}{n}.\sum_i x_i$ donde x_i son os valores de la variable.
- Datos agrupados en frecuencias simples: $\bar{x} = \frac{1}{n}.\sum_i x_i.f_a$ f_a es la frecuencia absoluta.

- Datos agrupados en intervalos: $\bar{x} = \frac{1}{n}.\sum_i p_{mi}.f_a$ donde p_{mi} es el punto medio de cada intervalo, es decir, la marca de clase.

Mediana (M_e):

- Datos sin agrupar: $Posición\ mediana = \frac{n+1}{2}$, luego busco en la tabla.
- Datos agrupados en intervalos:

$$Posición\ mediana = \frac{n}{2}$$

$$M_e = l_{i-1} + c.\frac{\frac{n}{2} - Fa_{i-1}}{Fa_i - Fa_{i-1}}$$

Donde:

l_{i-1}: Extremo izquierdo del intervalo que contiene a la mediana.

c: Ancho del intervalo.

Fa_{i-1}: Frecuencia acumulada del intervalo anterior.

Fa_i: Frecuencia acumulada del intervalo de la mediana.

Varianza: $s^2 = \frac{1}{n-1}.\sum_i (x_i - \bar{x})^2$

Desviación estándar: $s = \sqrt{\frac{1}{n-1}.\sum_i (x_i - \bar{x})^2}$

Coeficiente de variación: $CV = \frac{s}{\bar{x}}.100\%$

Cuartiles, deciles y percentiles

Posición $(Q_i) = \frac{n+1}{4}.i$ Q_i: cuartil i, con i=1 y 3

Posición $(D_i) = \frac{n+1}{10}.i$ D_i: decil i, con i=1,..., 10.

Posición $(P_i) = \frac{n+1}{10}.i$ P_i: percentil i, con i=1,..., 100.

$Q_i = a + d.(b - a)$

$D_i = a + d.(b - a)$

$P_i = a + d.(b - a)$

Donde:

- a: valor anterior a la posición.
- b: valor posterior a la posición.
- d: parte decimal de la posición.

Zonas de existencia de valores atípicos en diagrama de caja

- $Q_3 + 1{,}5.(Q_3 - Q_1)$
- $Q_1 - 1{,}5.(Q_3 - Q_1)$
- $Q_1 - 3.(Q_3 - Q_1)$

Asimetría

Coeficiente de asimetría de Pearson

$$C_{AP} = \frac{\bar{X} - M_O}{s}, \quad \mathrm{M_O}\text{:moda}$$

Coeficiente de asimetría de Bowley

$$C_{AB} = \frac{Q_3 + Q_1 - 2M_e}{Q_3 - Q_1}$$

Binomial: $P(X = r) = \binom{n}{r}.p^r.q^{(n-r)}$

$$E(X) = n.p$$

$$Var(X) = n.p.q$$

Normal: $(Tabla)$ $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$

Tipificación de la variable: $Z = \frac{x-\mu}{\sigma}$

Aproximación de una Binomial a una Normal: $si\ p > 0{,}05; q > 0{,}05\ y\ n \geq 30$ Entonces

$B(n,p)$ puede aproximarse a una Normal con $\mu = n.p$ y $\sigma = \sqrt{n.p.q}$