Primero la Secundaria

MATEMÁTICA

Módulo 4

Probabilidad y estadística

Ministerio de Educación, Cultura, Ciencia y Tecnología
Presidencia de la Nación

Primero la Secundaria

MATEMÁTICA

Módulo

4

Probabilidad y estadística

Contenido

Presentación

Antes de comenzar con los temas del cuarto módulo de Matemática (¡y último!), te acercamos algunas orientaciones para organizar el estudio.

Para repasar para el examen, te brindamos material de estudio que te servirá para recordar los temas que ya trabajaste en clase y encontrarás actividades que te ayudarán a concentrarte en los temas más importantes. Te sugerimos que dediques dos horas por día al estudio y la práctica, así en una semana alcanzarás a preparar todo el módulo.

En este módulo repasaremos temas sobre probabilidad: su significado y utilización, y cómo calcularla en casos simples. El objetivo es que puedas aprender los conceptos, puedas deducir la probabilidad de algunos eventos que no sean muy complejos, interpretes esa probabilidad y puedas aplicarla.

Luego, trabajaremos con conceptos de estadística. Haremos representaciones gráficas para poder entender algunos cálculos para caracterizar conjuntos de datos. Este tema es muy importante porque vivimos rodeados de información con datos sobre muchísimas variables: estadísticas de poblaciones, de precios, de salarios, de salud. Este módulo te permitirá comprender las características esenciales de las estadísticas sobre datos. El objetivo es que puedas incorporar herramientas de análisis, tanto para comprender los análisis que realizan otros como para hacer los propios.

Nos centraremos en aspectos prácticos y, en lugar de hacer desarrollos teóricos complejos y abstractos, vamos a ir aprendiendo en base a ejemplos. La guía está dividida en dos bloques, uno sobre probabilidad y otro sobre estadística. Al final de cada bloque encontrarás una serie de ejercicios. Para el aprendizaje del contenido de este módulo es fundamental trabajar los ejercicios, asegurate de hacer todos. Antes de seleccionar la opción que consideres correcta, hacé el ejercicio a conciencia en un cuaderno, calculá el resultado y luego buscalo en las opciones. Leé bien la solución donde encontrarás explicado cómo resolverlo. Leé la forma de llegar al resultado tanto si pudiste encontrar la solución correcta como si no lo lograste, ya que las explicaciones pueden aclarar más lo ejercitado. Luego de unos días, volvé a hacer los ejercicios para tener bien claro si pudiste incorporar los temas.

Probabilidad

La probabilidad indica la fracción de veces que un suceso ocurrirá. Es el cociente (división) entre la cantidad de casos "favorables" (en que el suceso efectivamente ocurre) frente a la cantidad de casos posibles. Los casos favorables no pueden ser más que los casos posibles, por ello la probabilidad es un número menor o igual a uno. Además, las cantidades no pueden ser negativas, por lo cual, la probabilidad tampoco. Entonces, la probabilidad de un suceso será un número entre 0 y 1 inclusive. Muchas veces, la probabilidad también se expresa en porcentaje, de 0 a 100%, que es equivalente.

Veamos un ejemplo cotidiano: las radios, los diarios y la televisión constantemente nos informan sobre el pronóstico del clima. Cuando nos anuncian que hay 85% de probabilidad de lluvia, ¿qué nos están diciendo? Que si hubiese infinitos días como ese, en el 85% de esos días lloverá y los restantes 15% de los días no lloverá. No podemos asegurar qué va a pasar en un día en particular, ese día puede pertenecer al grupo de los días con lluvia o al grupo de los días sin lluvia. Pero sí podemos asegurar que hay mayor probabilidad de que llueva a que no llueva –de acuerdo con factores climáticos como la humedad, la nubosidad, etc– y, además, podemos cuantificar esa probabilidad. Habíamos dicho que la probabilidad es el cociente entre la cantidad de sucesos "favorables" frente a la cantidad de posibles. En este caso, los sucesos favorables son los días con lluvia y los posibles son todos los días (incluyendo los días en los que no llueve y los días con lluvia). La probabilidad "favorable" significa que el suceso en cuestión puede ocurrir, no es algo subjetivo sobre si es conveniente o perjudicial.

Cálculo de probabilidad

Ya sabemos qué significa la probabilidad, pero ¿cómo se obtiene? Hay dos formas:

- En base a "experimentos": haciendo estadística de situaciones. En el caso de la lluvia, sería llevar un gran registro de las variables importantes (supongamos que son: viento, nubosidad, presión, temperatura) junto con el dato de si llovió o no. Luego, si queremos saber la probabilidad de que hoy llueva, nos fijamos en nuestro registro todos los casos donde el viento, la nubosidad, la presión y la temperatura son como hoy. Contamos todos esos casos (eventos "posibles") y también contamos la cantidad de esos casos en los que llovió (eventos "favorables"). Hacemos el cociente favorables/posibles y obtenemos la probabilidad de lluvia en un día con características determinadas. Podemos hacer este análisis experimental si disponemos de un registro con gran cantidad de datos.

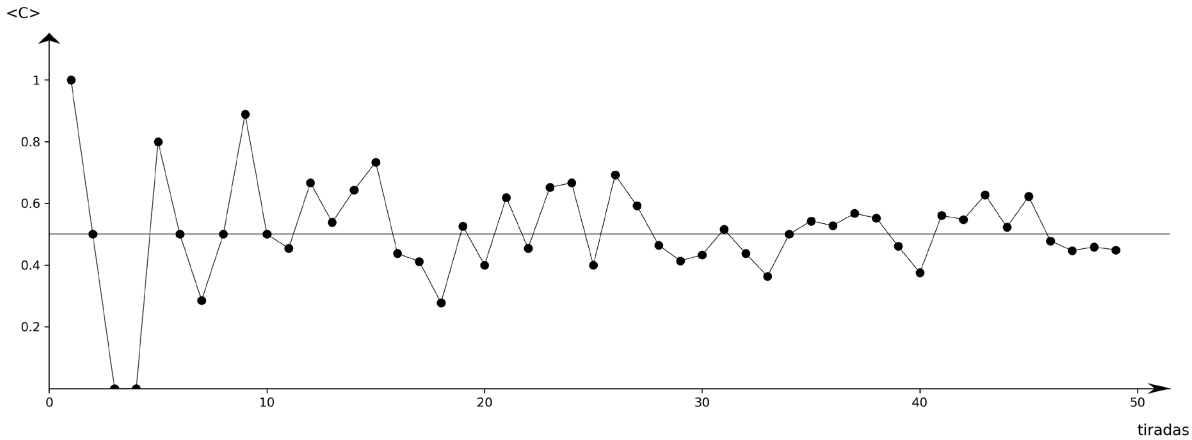
- En base a “modelos matemáticos“: haciendo modelos con ecuaciones que describan el comportamiento del sistema. En el caso de la lluvia sería analizar, por ejemplo, cómo influyen los vientos y las temperaturas en las nubes y cómo estas últimas generan lluvia. Estos modelos pueden requerir grandes cantidades de cálculos y simulaciones.

Cara o seca

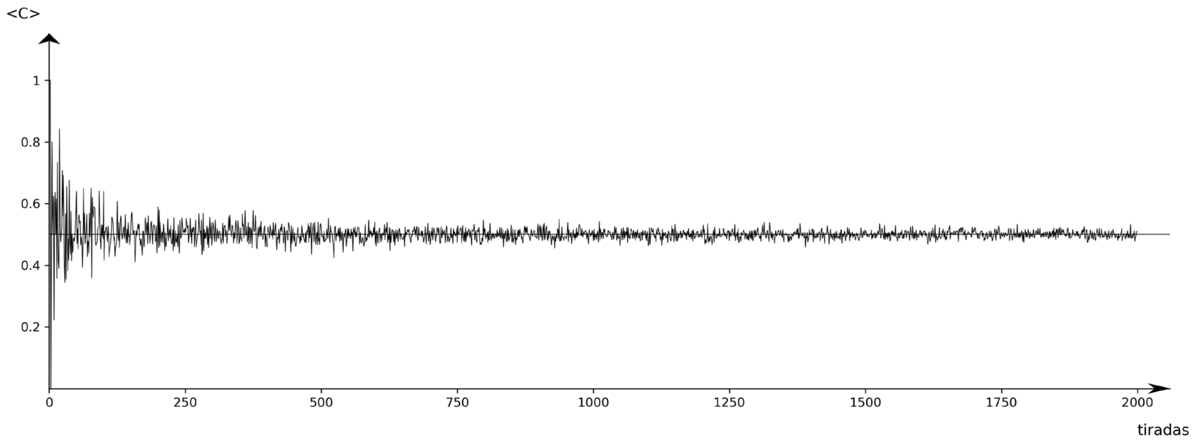
En esta sección vamos a ver un ejemplo de cómo encontrar una probabilidad en base a los dos métodos que nombramos. Comenzaremos con el primer método: en base a un experimento. Las curvas que vamos a ir graficando no son un objetivo en sí mismo de este módulo, las vamos a ir construyendo para poder acercarnos a los conceptos. Para esta parte del módulo vas a necesitar una moneda, para calcular la probabilidad de obtener el lado “cara“ al arrojarla. Hacé este experimento: llená una tabla como la que te mostramos. Por cada fila vas a hacer una “serie” de tantas tiradas como te dice la primera columna, luego anotá en la segunda columna la cantidad de veces que salió cara y en la tercera columna, el cociente (al que llamaremos <C>) que resulta de dividir la cantidad de caras por la cantidad de tiradas. Por ejemplo, si para 5 tiradas obtenemos 3 caras y 2 secas, el valor de <C> será 3/5 = 0,6:

Cantidad de tiradas	Cantidad de “caras”	<C>
1		
2		
3		
5	3	0,6
10		
20		
30		
40		
50		

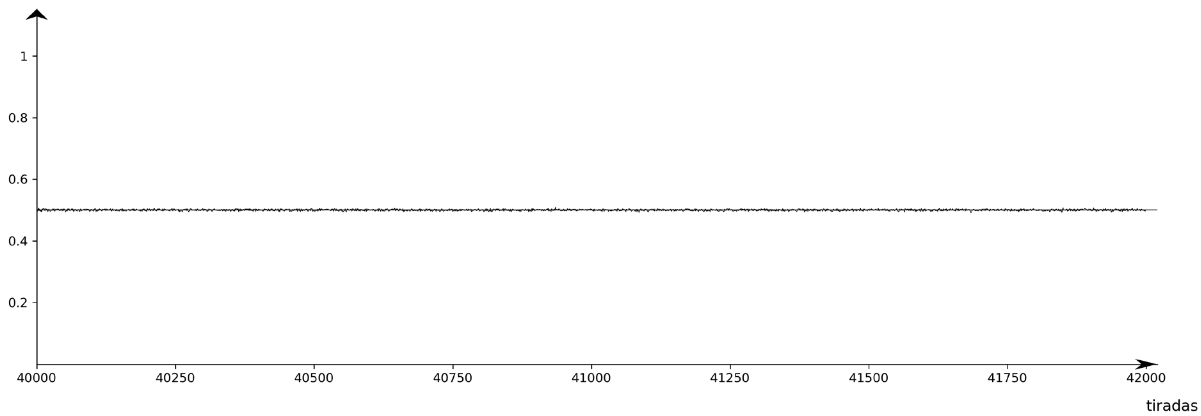
Ahora, graficá <C> en función de la cantidad de tiradas. Hacé tu gráfico y obtendrás uno similar al que te mostramos (aquí lo hicimos con más mediciones):



La figura anterior muestra que <C> tiene mucha fluctuación, es decir, que no es una curva suave sino que tiene muchos saltos y picos que están por arriba y por debajo del valor medio. Tu curva tendrá otra forma porque los eventos de arrojar la moneda no fueron los mismos que los que hicimos nosotros. A medida que aumentamos la cantidad de tiradas, la fluctuación disminuye y <C> se va pareciendo cada vez más a 0,5. Veremos qué ocurre cuando llevamos la cantidad de tiradas a un número mucho mayor, por ejemplo, hasta 2000:



Podemos observar cómo la fluctuación va disminuyendo. Y si vamos a un número de tiradas aún mayor, fijate qué ocurre en el gráfico:



tiradas con las que calculamos <C>. Podemos hacer crecer la cantidad de tiradas todo lo que queramos, pero aunque sea cada vez menor, la fluctuación estará ahí. El cálculo que venimos haciendo es: la cantidad de caras (casos favorables) sobre la cantidad de tiradas (casos posibles: caras y secas). En definitiva, estamos estimando la probabilidad de obtener una cara en una tirada. Tendríamos el valor "verdadero" de la probabilidad si pudiésemos hacer infinitas tiradas, algo imposible. Con una cantidad finita de tiradas obtenemos una estimación de la probabilidad dentro de un margen de error que será menor cuanto más tiradas podamos hacer. El margen de error se puede calcular, pero ese cálculo excede el objetivo de este módulo. Pero sí vamos a mencionar algo muy importante: el margen de error baja a medida que aumentamos la cantidad de veces que arrojamos la moneda. De las dos opciones que nombramos para encontrar la probabilidad, recién lo hicimos en base a **experimentos**. (*Te habrás preguntado si realmente arrojamos la moneda más de 40.000 veces por cada uno de los puntos del último gráfico. La respuesta es que no, lo hicimos en base a arrojar "monedas virtuales" a través de un programa de computación*).

Conclusión: para obtener la probabilidad de un suceso en forma de "experimento", repetimos muchas veces los eventos con resultado al azar. La probabilidad será la cantidad de veces que ocurrió el evento dividida la cantidad de repeticiones de los eventos del experimento. Cuantas más veces repitamos el experimento, mejor será nuestra estimación de la probabilidad.

Ahora encontraremos la probabilidad de obtener cara al arrojar una moneda con el otro método: en base a modelos. Podemos asegurar que la moneda es esencialmente igual de ambas caras y que no se puede tener control sobre cómo caerá. En base a este breve análisis, la probabilidad de que salga cara debe ser la misma de que salga seca, esto implica que debe ser ½=0,5. En el caso de la moneda hay dos sucesos posibles –cara o seca- y ambos tienen la misma probabilidad de suceder (no hay ningún motivo para creer que es más probable que pase uno que el otro), por eso decimos que los sucesos son **equiprobables**. Para el ejemplo de la moneda, el modelo es muy simple, pero en otros casos el modelo puede ser extremadamente complejo (inclusive hay muchos modelos para los cuales ni siquiera es posible realizar los cálculos).

En los casos en los que no hay motivo para que un suceso tenga más probabilidad que otro, la probabilidad de cada uno de ellos debe ser idéntica, y la calcularemos dividiendo 1 por la cantidad de sucesos posibles.

Otros casos

Muchos juegos de azar se basan en sorteos de eventos equiprobables, que quiere decir que la probabilidad de obtener un determinado resultado en un sorteo es la misma que la de obtener cualquier otro resultado. Todos los números de un dado tienen la misma probabilidad de quedar en la cara de arriba. Después de mezclar el mazo de cartas, la probabilidad de que nos toque el as de espadas o el cuatro de copas es la misma. En el bingo, la probabilidad de obtener la bolilla 1 o la 12 del bolillero es la misma. La bolita puede terminar en cualquiera de las 37 posiciones de la ruleta con igual probabilidad. En todos estos casos nos basamos en la simetría, en lo iguales que son las piezas, etc. para determinar que la probabilidad de un suceso en particular se calcula dividiendo 1 por la cantidad de sucesos posibles. Esta estrategia es válida solamente en el caso de que todos los sucesos posibles tengan la misma probabilidad, y esto se denomina "equiprobable". Este camino para calcular la probabilidad de un suceso se conoce como "fórmula de Laplace".

Mirá el siguiente video disponible en este link:

https://www.educ.ar/recursos/102674/estadistica-y-probabilidad

(la explicación está entre el minuto 12:37 y el minuto 13:59).

Este video se encuentra en la plataforma, Página del estudiante: Recursos para el estudio / Matemática 4 / Probabilidad.

Sigamos con otro ejemplo. Para cada problema en particular, habrá que definir bien cuáles serían los casos favorables y cuáles todos los posibles. Para obtener la probabilidad de que ocurra un suceso, el cálculo consiste en contar "casos favorables" y "casos posibles" para hacer el cociente entre ellos: favorables/posibles. Mirá este ejemplo en el mismo video que compartimos recién. Ahora observá del minuto 6:50 al minuto 9:23 En el video, los sucesos posibles son los que corresponden a tomar cualquier caramelo de la bolsa y los sucesos favorables, el haber tomado un caramelo de sabor a fruta.

Recordá que el link es:

https://www.educ.ar/recursos/102674/estadistica-y-probabilidad

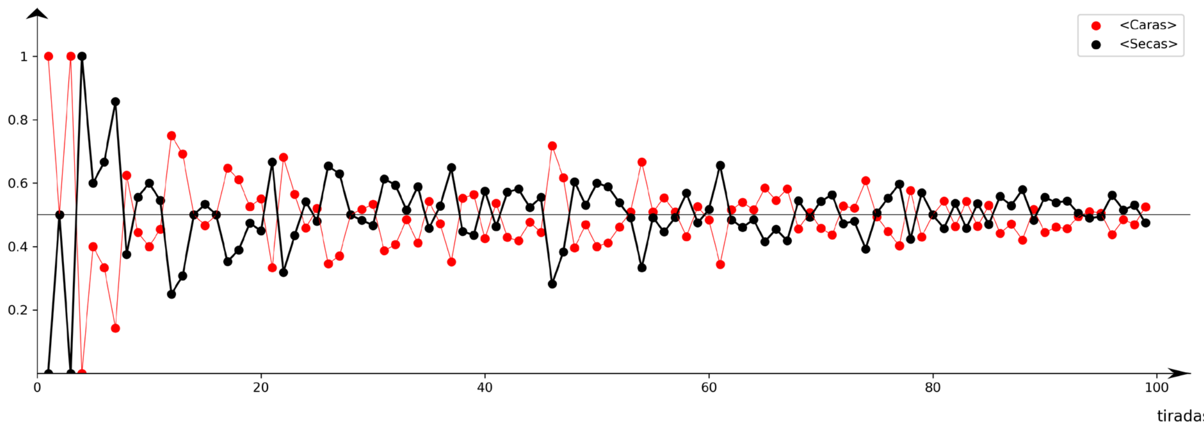
Y que encontrarás el video en la plataforma, Página del estudiante: Recursos para el estudio / Matemática 4 / Probabilidad.

Otro ejemplo parecido a arrojar una moneda es el juego gauchezco de la taba. En la taba se hacen apuestas sobre cual lado de un hueso quedará hacia arriba tras ser arrojado al aire. Ahora bien, a diferencia de la moneda, el hueso tiene dos caras, pero una forma irregular. ¿Cuál es la probabilidad de que tras arrojarlo quede para arriba una de sus "caras"? En principio, uno se tentaría a decir que es 1/2, pero como la taba no tiene una forma regular y simétrica como la moneda, no podemos predecir de entrada la probabilidad. En estos casos, no queda otra alternativa que hacer una tabla como la que hicimos con la moneda y encontrar la probabilidad en forma de experimento.

Independencia de sucesos

Para esta parte vamos a volver a utilizar la moneda. Haremos series de 2 tiradas. Tiramos una vez, supongamos que salió "cara", ¿cuál es la probabilidad de que vuelva a salir cara? Para los casos donde la primera tirada sea cara, vamos a anotar cuantas de las segundas tiradas dieron cara y cuantas seca. Lo repetimos hasta llegar a tener 100 casos para los cuales la primera moneda es cara. ¿Pudiste sacar alguna conclusión? (*No se te pedirá que hagas curvas como éstas en los ejercicios ni en los exámenes, las utilizamos aquí como acercamiento al tema*).

Te mostramos ese mismo experimento a continuación. En lugar de hacer una serie de 100 casos, hicimos muchas series (*también con ayuda de nuestra computadora*) con distinta cantidad de tiradas. Te mostramos cómo fue nuestro resultado para la segunda moneda cuando la primera fue cara:



Si observás bien la figura, tiene el mismo tipo de comportamiento que cuando habíamos mirando solamente las tiradas de una moneda. Se ve una fluctuación muy grande alrededor de ½ al tener pocas tiradas. Esta fluctuación va disminuyendo con la cantidad de tiradas pero siempre alrededor de ½.

Algunas veces habrá más caras que secas y otras veces más secas que caras. Nosotros repetimos el experimento con 10000 tiradas y obtuvimos 5046 caras. Es decir, más allá de la fluctuación, cerca de la mitad de los eventos fueron caras y la otra mitad secas. La conclusión es importante: **los sucesos no guardan memoria, cada evento es independiente del anterior**. Es decir, no importa qué resultado haya dado el primer tiro, el segundo no va a variar en función del primero. Para nuestro ejemplo, que la primera vez haya sido cara o seca no afecta qué va a salir en la segunda tirada.

Contando posibles sucesos

Seguiremos con el ejemplo de la moneda. Si arrojamos 2 veces la moneda, ¿cuál es la posibilidad de obtener una cara en la primera tirada y una cara en la segunda tirada? Armemos una tabla con todas las posibilidades. En cada fila habrá un suceso o caso posible, es decir habrá 4 sucesos posibles:

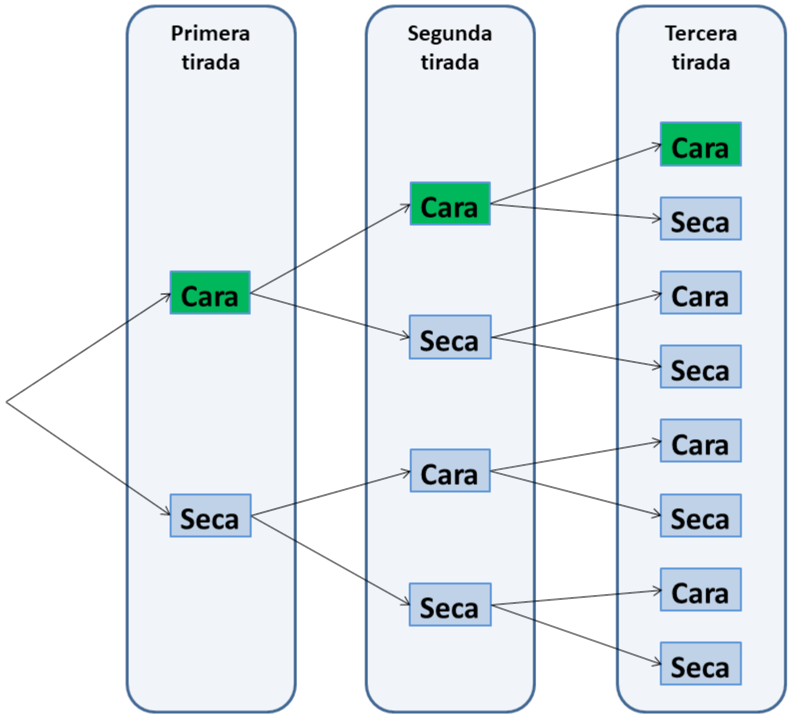
	Tirada 1	Tirada 2
1	Cara	Cara
2	Cara	Seca
3	Seca	Cara
4	Seca	Seca

Vemos que hay 4 casos posibles, de los cuales solo uno es favorable (cara en ambas tiradas), entonces la probabilidad será de ¼ (casos favorables sobre casos posibles).

Si ahora queremos saber la probabilidad de obtener 3 caras al tirar 3 veces la moneda, podríamos hacer una tabla como la anterior, pero con 3 columnas incluyendo todos los casos posibles.

	Primera tirada	Segunda tirada	Tercera tirada
1	Cara	Cara	Cara
2	Cara	Cara	Seca
3	Cara	Seca	Cara
4	Cara	Seca	Seca
5	Seca	Cara	Cara
6	Seca	Cara	Seca
7	Seca	Seca	Cara
8	Seca	Seca	Seca

Analicemos con una figura:



En la figura resumimos todas las combinaciones, que son los sucesos posibles. Siguiendo las líneas de arriba, encontramos el suceso favorable que marcamos con verde. Vamos a analizar la cantidad de posibilidades, es decir, sucesos posibles. Como vimos recién, cada tirada es un evento independiente del anterior. Analicemos entonces cómo se relacionan estos tres eventos (tres tiradas) para comprender qué cantidad de sucesos posibles existen.

Vemos que en la primera tirada se abren dos posibilidades: cara o seca. Para la segunda tirada también la moneda podrá quedar en cara o seca, lo que equivale a que cada una de las dos posibilidades que ya teníamos se abre en dos. Esto se calcula multiplicando la posibilidad de sucesos posibles en la primera tirada por la cantidad de sucesos posibles de la segunda tirada: $2 * 2 = 4$. Estas son las posibles combinaciones entre primera y segunda tirada.

La tercera tirada nuevamente nos puede dar dos opciones, con lo cual cada una de las cuatro posibilidades anteriores se abre en 2 nuevas. Combinando, tenemos $2 * 2 * 2 = 8$ posibles combinaciones en las 3 tiradas.

Entonces la cantidad de sucesos posibles son:

$$S = 2 * 2 * 2 = 2^3 = 8$$

¿Qué hicimos? Multiplicamos la cantidad de sucesos posibles en un evento (tirada) por sí misma tantas veces como la cantidad de eventos (cantidad de tiradas) haya. La cantidad de sucesos posibles en una tirada son 2, y la cantidad de tiradas son 3. Entonces, multiplicamos 3 veces al número 2 por sí mismo.

Como vimos más arriba, nuevamente la probabilidad de obtener (cara, cara, cara) será la cantidad de sucesos favorables F dividido la cantidad de sucesos posibles S:

$$P = \frac{F}{S} = \frac{1}{8}$$

¿Cuál hubiese sido la probabilidad de obtener 1) cara, 2) seca, 3) cara? También 1/8, ya que los posibles sucesos siguen siendo 8 y la cantidad de sucesos favorables es 1.

¿Cuál hubiese sido la probabilidad de obtener 1) cara, 2) cara o seca, 3) seca? En este caso ¼, ya que los posibles sucesos siguen siendo 8, pero la cantidad de sucesos favorables son 2: (cara, cara, seca) y (cara, seca, seca). Entonces sería 2/8 que es igual a ¼.

¿Habrá alguna forma de calcular la cantidad de sucesos sin hacer la tabla? La respuesta es sí. Cuando los eventos son equiprobables e independientes (no importa qué salga en el anterior o los anteriores para definir qué saldrá en los siguientes), definimos j como la cantidad de sucesos posibles en un evento. Por ejemplo, j=2 en el caso de la moneda por haber 2 posibles resultados (cara o seca) por cada tirada. O j=6 en el caso de arrojar un dado ya que tiene 6 posibles resultados. Ahora, la cantidad de posibles sucesos tras n repeticiones de eventos independientes será:

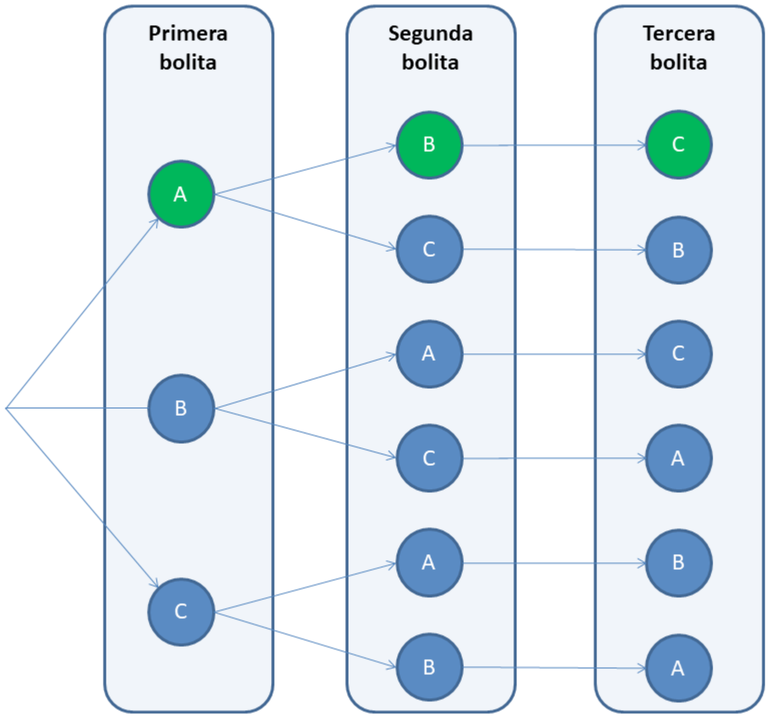
$$S = j^n$$

Conclusión: la cantidad de secuencias posibles al repetir n veces eventos iguales e independientes es igual a la cantidad de sucesos posibles en cada evento elevada a la *n.*

Combinatoria

Ordenando todos los elementos

Analicemos la siguiente situación. Tenemos una bolsa con 3 bolitas, una marcada con la letra “A”, otra con la “B” y otra con la “C”. Vamos sacando de a una, ¿cuál es la probabilidad de obtener la secuencia A-B-C? La secuencia es el orden en el que se van sacando las bolitas, y la probabilidad de una secuencia en particular será: 1 (cantidad de eventos favorables) dividido la cantidad de formas diferentes de ordenarlas (cantidad de eventos posibles). En primer lugar, debemos notar que sacar la segunda bolita **no es un evento independiente** ya que en la bolsa ya no está la bolita que sacamos al principio. El análisis lo haremos con un árbol similar al que utilizamos antes con las monedas:



Para la primera bolita tenemos 3 posibilidades (es decir, al haber tres bolitas, al sacar la primera hay posibilidad de que salga cualquiera de las 3 que hay). Ahora, al sacar la segunda, nos quedan dos posibilidades porque ya habíamos quitado una bolita de la bolsa y solo quedan dos. Para la tercera bolita, una vez que quitamos las dos primeras, nos queda una única posibilidad. Esto hace que la cantidad de sucesos posibles sea:

$$S = 3 * 2 * 1 = 6$$

¿Qué hicimos? Fuimos multiplicando la cantidad de bolitas por la cantidad de bolitas menos 1, luego por la cantidad de bolitas menos 2; y frenamos ahí porque ya no quedan bolitas. Al haber un solo evento favorable (el pintado con verde) la probabilidad es $P = F/S = 1/6$. Si tuviésemos n bolitas diferentes, la cantidad de órdenes posibles en los que las podríamos haber sacado de la bolsa son:

$$S = n * (n-1) * (n-2) * (n-3) *....... * 1$$

Los puntos suspensivos indican que la operación se repite hasta que el paréntesis se hace 1. Esta cuenta es muy usual en matemática y recibe el nombre de factorial. El factorial se simboliza con el signo de admiración “!”:

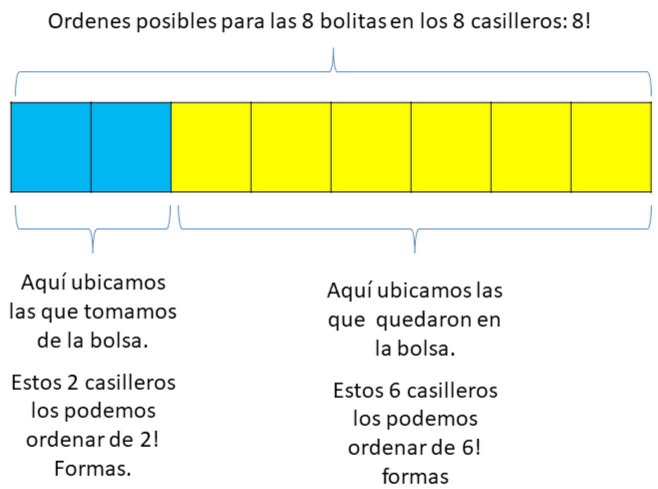
$$S = n!$$

Esta formulación nos sirve para calcular la cantidad de formas posibles en que podrían haber salido las bolitas de la bolsa y también, para calcular las diferentes formas de ordenar elementos. Por ejemplo, si queremos ordenar 4 personas en 4 sillas, podríamos numerar las sillas del 1 al 4, y meter 4 bolillas en una bolsa (correspondientes a los números de las sillas) y, a medida que entraran las personas, ir sacando de la bolsa el número de silla que le corresponde a cada uno. Entonces, la cantidad de formas de ordenar (acomodar) a las personas, es equivalente a la cantidad de secuencias posibles que podemos obtener al ir retirando bolillas de la bolsa.

Conclusión: la cantidad de formas de ordenar n elementos diferentes es *n*!

Combinaciones sin repetición

Supongamos que tenemos 8 bolitas en una bolsa, cada una identificada con una letra: A, B, C, D, E, F, G, H. Si saco 2 bolitas, ¿cuál es la probabilidad de obtener la bolita A y la B sin importar el orden en el cual las obtengo? Veamos como contar los sucesos. Empecemos por ver todos los posibles órdenes si sacamos todas las bolitas de a una y las vamos acomodando en casilleros. Los dos primeros van a ser para las 2 bolitas que tomamos de la bolsa y los siguientes 6 para las bolitas que hubiesen quedado en ella:



Primero nos fijamos cuántos órdenes posibles hay para ubicar las 8 bolitas en los 8 casilleros, por lo que vimos antes hay 8! órdenes posibles. En nuestro caso es indistinto si sacamos primero la bolita A y después la D o si sacamos primero la D y después la A, el orden de los dos primeros casilleros no nos importa, ambos casilleros significan "bolita elegida"; y tampoco nos importa en cuál ponemos la primera y en cuál la segunda. Esto implica que debemos dividir la cantidad de órdenes que teníamos para las 8 bolitas por la cantidad de órdenes posibles de esos dos casilleros celestes (2!). Lo mismo ocurre con las que quedaron en la bolsa, el orden dentro de ellas no importa, como son 6 casilleros, los órdenes dentro de la bolsa son 6! y dividimos los casos que teníamos por este número. De esta forma **la cantidad de sucesos posibles son**:

$$S = \frac{8!}{2! * 6!} = \frac{40320}{2 * 720} = 28$$

Entonces, obtenemos que **la probabilidad de sacar las bolitas A y B es 1/28**, ya que hay un suceso favorable frente a 28 posibles.

Podemos desarrollar los factoriales y luego simplificamos lo que aparezca en el numerador y denominador:

$$S = \frac{8!}{2! * 6!} = \frac{(1 * 2 * 3 * 4 * 5 * 6 * 7 * 8)}{(1 * 2) * (1 * 2 * 3 * 4 * 5 * 6)} = \frac{7 * 8}{1 * 2} = \frac{56}{2} = 28$$

El 6! se debía a que quedaban 6 bolitas en la bolsa. La cantidad de bolitas que quedan en la bolsa, es la cantidad total de bolitas (8) menos las que saco (2), entontes:

$$S = \frac{8!}{2! * (8-2)!}$$

Prestá atención, extenderemos este concepto para el caso general: **si tengo un conjunto de *n* elementos distintos, para encontrar la cantidad de subconjuntos diferentes con r elementos utilizo la siguiente fórmula**:

$$S = \frac{n!}{r! * (n-r)!}$$

En el caso que veníamos analizando, teníamos un conjunto de 8 elementos, por eso n=8, y los subconjuntos son todos los posibles grupos de 2 bolitas, lo que implica que $r = 2$.

En esta sección que trata sobre probabilidad repasaste el significado de ella y algunos modos para calcularla. Tratamos secuencias de eventos independientes y cómo calcular la cantidad de secuencias diferentes que podemos obtener. Luego, calculamos la cantidad de formas en las cuales podemos ordenar n elementos distintos y la cantidad de subgrupos diferentes que pueden extraerse de un conjunto de elementos distintos. Utilizamos esas cantidades de posibles resultados para calcular la probabilidad de algún resultado en particular. Ahora, ¡a practicar!

A. Indicá si el enunciado es Verdadero o Falso.

En la quiniela, el jugador apuesta a un determinado número y, si ese número resulta sorteado, recibirá un premio. El apostador puede elegir el número o pedir al empleado que la computadora elija los números al azar. Entonces:

“no es conveniente que la elección se haga a través de la computadora porque disminuye las probabilidades de ganar ya que, para ganar, el número tuvo que haber salido sorteado 2 veces“.

○ Verdadero ○ Falso

B. Indicá si el enunciado es Verdadero o Falso.

En todas las agencias de lotería aparece una tabla con los números que más salieron en los últimos sorteos, esa información no tiene ningún valor al momento de hacer una nueva apuesta.

○ Verdadero ○ Falso

C. Indicá si el enunciado es Verdadero o Falso.

Al arrojar 6000 veces un dado de seis caras, voy a obtener 1000 veces el “2”

○ Verdadero ○ Falso

D. Indicá si el enunciado es Verdadero o Falso.

La probabilidad de tirar dos veces los dados y obtener primero un 3 y luego un 1 es distinta a la de obtener primero un 3 y luego otro 3.

○ Verdadero ○ Falso

E. Indicá si el enunciado es Verdadero o Falso.

Si hoy hay un 90% de probabilidad de lluvia, la probabilidad de que no llueva es 10% que es menor a la probabilidad de que sí llueva, lo que implica que va a llover.

○ Verdadero ○ Falso

F. Indicá la respuesta correcta.

En una competencia escolar se eligen al azar 2 representantes de un curso de 10 alumnos. Juan y Andrea tienen muchas ganas de resultar seleccionados. ¿Cuál es la probabilidad de que Juan y Andrea resulten representantes del curso?

- 1/45
- 1/20
- 1/19
- 1/10

G. Indicá la respuesta correcta.

Alejo va 9 días seguidos a hacer un trámite a una oficina. Al llegar, el sistema de turnos sortea en cuál de las 10 ventanillas será atendido.

¿Cuál es la probabilidad de que sea atendido los nueve días en la misma ventanilla?

(*Ayuda: tener en cuenta que la ventanilla del primer día puede ser cualquiera, lo importante es que en los siguientes 8 días resulte la misma*)

- $(1/9)^{10}$
- $(1/10)^{9}$
- $(1/10)^{8}$
- 1/10 * 9

H. Indicá la respuesta correcta.

Martín juega a la “generala” con 5 dados. Ya tiró dos veces y tiene separados dos dados en los que salió “4” y dos dados en los que salió “5”. Le queda un tiro y espera obtener un juego de “full” (dos dados que resultaron del mismo número y tres que resultaron de otro mismo número), para ello necesita que el resultado al arrojar el último dado sea un “4” o un “5”. ¿Cuál es la probabilidad P de que obtenga un “full”?

- P = 1/6 * 1/6
- P = 1/6
- P = 1/3
- P = 1/2

I. Indicá la respuesta correcta.

En una caja hay un papel blanco, uno celeste, uno rojo, uno verde y otro azul. Se los mezcla y se sacan dos papeles al azar. ¿Cuál es la probabilidad de que sean los colores celeste y blanco de la bandera de Argentina?

- 1/25
- 1/20
- 1/10
- 1/5

J. Indicá la respuesta correcta.

En un juego de Backgamon se está por tirar dos dados y si se obtiene un doble (dos números iguales) se gana el partido en esa jugada. ¿Cuál es la probabilidad "P" de ganar en esta jugada?

- P=1/36
- P=1/12
- P=1/6
- P=1/3

K. Indicá la respuesta correcta.

Si se lanza una moneda al aire dos veces, ¿cuál es la probabilidad de obtener al menos una vez el lado "cara"?

- 1/4
- 1/3
- 1/2
- 3/4

Estadística

La estadística es una ciencia que estudia datos cuantitativos (aquello que se puede medir o contar). Se encarga de la recolección, análisis e interpretación. El estudio de los datos permite primero, describir un sistema para luego explicar su comportamiento.

Los análisis estadísticos se pueden aplicar para cuestiones científicas y también para la toma de decisiones de gobiernos, empresas, etc. Los datos pueden tener cualquier tipo de origen, y es condición que se pueda asignar algún tipo de evaluación numérica. Algunos datos ya son numéricos como cuánto pesa la gente de una población, el ingreso económico de cada familia, etc. Otros datos no son numéricos, pero tienen asociados la cantidad de veces en que ocurren (que se denomina “ocurrencias”). Te explicamos con un ejemplo: en una estadística sobre si los ciudadanos votan a uno u otro partido político, el partido político no es un número en sí mismo, pero la cantidad de votantes que elige cada partido político sí lo es. Los índices que se utilizan para medir alguna característica de un sistema de datos pueden ser objetivos, es decir que no dependen de opiniones, o no serlo. En el caso de la población de un país, el índice de natalidad es un número objetivo, indica cuantos bebés nacen cada cierta cantidad de habitantes en un año. Otros índices tienen una subjetividad escondida, lo que significa que hay valoraciones y opiniones que influyen en su elección. El índice de pobreza de un país mide la proporción de pobres en una población, para ello se debe definir cuáles condiciones hacen que una persona sea considerada como pobre, y ello dependerá de alguna “ecuación”. En nuestro país se suele calcular la cantidad de pobres en base a quiénes con sus ingresos no llegan a cubrir una canasta básica que incluye, entre otras cosas, comida, vestimenta y transporte. La ONU (Organización de las Naciones Unidas) agrega además en la medición factores relacionados con la salubridad del entorno medioambiental, la expectativa de vida y otros factores de calidad de vida. La elección de cuál es el índice correcto es subjetiva, pero una vez fijado el criterio, ese criterio finalmente arroja valores numéricos que no dependen de quién los evalúe aunque luego su interpretación dependerá de la aceptación que tenga la forma de definir cuándo una persona es pobre. En este módulo nos centraremos en la descripción de los datos una vez ya medidos.

Histogramas

Un histograma es una representación gráfica de cómo se distribuyen ciertos datos. Los histogramas en sí mismos no forman un objetivo de este módulo y no serán parte de los ejercicios ni del examen; pero son necesarios para la comprensión de algunos conceptos y, por eso, los incluimos dentro de las explicaciones. Un histograma es un gráfico de barras donde el área de cada barra es proporcional a la cantidad de elementos con alguna característica. Es una forma visual de representar cómo es la distribución de los datos. Antes de seguir leyendo, fijate que en la página 25 encontrarás el ejemplo del histograma con las barras azules). Como en otros casos, vamos a explicar en base a un ejemplo. Supongamos que tenemos un grupo de 30 personas y queremos saber qué alturas tienen. Tomamos las medidas de todos y obtenemos los siguientes valores:

Alturas = {160cm; 164cm; 170cm; 169cm; 161cm; 153cm; 153cm; 173cm; 166cm; 169cm; 155cm; 170cm; 157cm; 161cm; 165cm; 177cm; 161cm; 160cm; 173cm; 161cm; 153cm; 183cm; 166cm; 158cm; 160cm; 162cm; 156cm; 160cm; 169cm; 163cm}

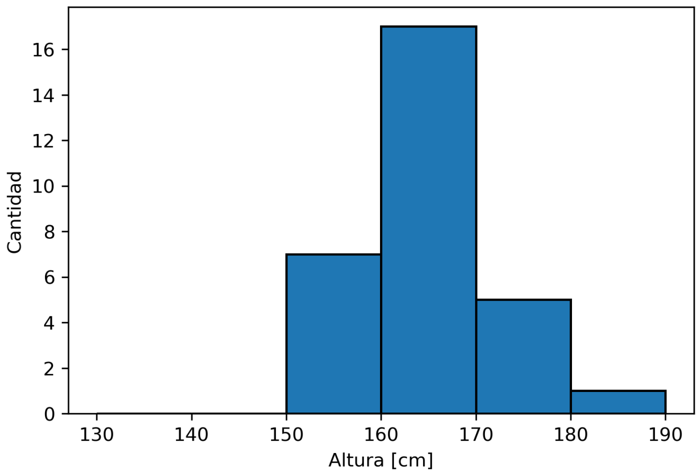
Las alturas que medimos están ordenadas según las fuimos midiendo. Reordenemos de menor a mayor:

Alturas = { 153cm; 153cm; 153cm; 155cm; 156cm; 157cm; 158cm; 160cm; 160cm; 160cm; 160cm; 161cm; 161cm; 161cm; 161cm; 162cm; 163cm; 164cm; 165cm; 166cm; 166cm; 169cm; 169cm; 169cm; 170cm; 170cm; 173cm; 173cm; 177cm; 183cm}

Ahora, vamos a dividir las alturas en pequeños rangos o grupos, y cada rango irá desde una “altura inferior” hasta una “altura superior”. A cada rango lo llamamos “bin”. Contaremos cuántas personas tienen una altura superior o igual a la “altura inferior” y que al mismo tiempo sea menor a la “altura superior”, esa cantidad estará asignada a ese bin. Analicemos los bines de 10cm de ancho que van de 130cm a 200cm. Para el bin que va desde 130cm hasta 140cm no hay ninguna persona, lo mismo ocurre para el bin que va desde 140cm hasta 150cm. Para el bin que va desde 150cm hasta 160cm tenemos las personas que miden: 153cm; 153cm; 153cm; 155cm; 156cm; 157cm; 158cm. Entonces para este bin hay 7 personas. Para el rango que va desde 150cm hasta 160cm hay 17 personas con alturas: 160cm; 160cm; 160cm; 160cm; 161cm; 161cm; 161cm; 161cm; 162cm; 163cm; 164cm; 165cm; 166cm; 166cm; 169cm; 169cm; 169cm. Luego, completamos para el resto de los bines. Anotamos la cantidad de personas en cada bin en una tabla:

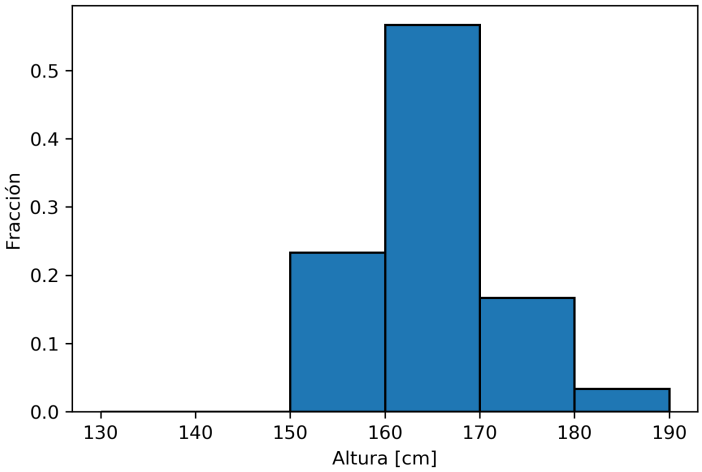
Altura Inferior [cm]	Altura Superior [cm]	Cantidad	Fracción
130	140	0	0
140	150	0	0
150	160	7	0,233
160	170	17	0,567
170	180	5	0,167
180	190	1	0,033
190	200	0	0

Los histogramas son gráficos de barras cuyas alturas corresponden a la cantidad de elementos dentro de cada bin y cuyo ancho se extiende entre los límites inferior y superior:



El histograma nos permite observar de forma muy rápida la distribución de un conjunto de mediciones de una variable, en este caso de alturas, sin tener el detalle de cuánto es el valor exacto de cada medición. En este ejemplo, condensamos en 4 barras la información de 30 mediciones de alturas. ¿Qué información podemos tener observando el histograma? En estas 4 barras podemos ver que la mayoría mide entre 160cm y 170cm, que ninguno tiene una altura menor a 150cm y ninguno, una mayor a 190cm, que hay una sola persona con altura mayor a 180cm. Nos da mucha información procesada y ordenada que no hubiésemos visualizado con los 30 valores sueltos.

En la tabla de la página anterior , fijate que en la última columna, cuyo nombre es “fracción”, se indicó cuánto representa la cantidad de personas dentro de un bin respecto del total de personas, es decir, la división entre estas magnitudes. Otra forma de hacer histogramas es graficando estas fracciones como alturas de la barra. La información que nos brinda este histograma es muy similar, podremos tener una lectura muy rápida del porcentaje de personas con cierto rango de alturas frente al total independientemente de cuántas personas hay en total en el campo muestreado. En principio perdemos información de la cantidad de personas, pero nos permite comparar poblaciones con distinta cantidad de miembros. Si vamos a comparar las alturas de las personas de un país con las personas de otros, vamos a comparar las fracciones en cada bin, ya que el total de personas en un país puede ser muy diferente al del otro. El histograma que venimos analizando, como fracción de personas que entran dentro de cada bin para la misma población, es:



Valores que caracterizan un conjunto de valores

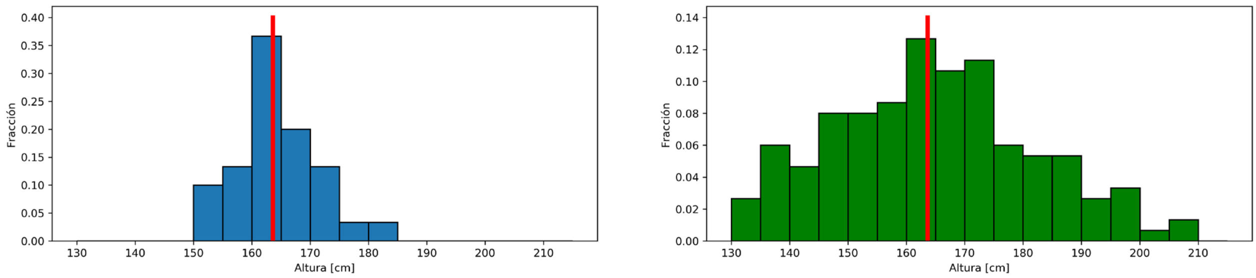
En muchas ocasiones queremos resumir la distribución de los valores de una muestra en pocos números. Tenemos miles de datos y queremos en pocos valores dar idea de cómo son esos datos. Uno de esos números es la llamada "media". La media es el promedio entre todos los datos. Si tenemos un conjunto de "*N*" mediciones tipo "*x*", numero las mediciones con un índice *i* y, entonces, cada medición tiene un valor que llamo "x_i". La media será:

$$\bar{x} = \frac{1}{N} * \sum_{i=1}^{N} x_i$$

La rayita sobre la x: $\bar{x}$ indica que es el valor medio. La letra griega mayúscula Σ (sigma) indica sumatoria, es decir que se sumará desde *i*=1 hasta i=*N* todo lo que está a su derecha, en este caso los valores de x_i La expresión anterior es equivalente a:

$$\bar{x} = \frac{x_1 + x_2 + ... + x_N}{N}$$

La media equivale al "promedio" de las mediciones. En el ejemplo que venimos siguiendo de las alturas, la altura media (hacé como práctica el cálculo promediando las 30 alturas) es 163,6cm. Esta información condensa, en un único número, una característica importante de las alturas de la población. Comparando las alturas medias de dos países podemos saber, por ejemplo, si en promedio en un país son más altos o bajos que en otro. Si bien ese número nos da información de la altura promedio, no nos da información de cómo están distribuidas las alturas. En la siguiente figura, vamos a graficar en azul el histograma de alturas del grupo con el cual veníamos trabajando y en verde, el histograma de otro grupo de personas cuya altura media es idéntica pero cuya distribución es diferente. Como la cantidad de personas en cada grupo también es diferente, vamos a comparar la fracción de personas dentro de cada bin.



En los histogramas marcamos la media con una línea roja, la media es idéntica en ambos casos. La diferencia entre ambas distribuciones es clara: en la distribución mostrada en verde hay mucha más dispersión de alturas, se alcanzan alturas más bajas y más altas que en la mostrada en azul. También podemos observar que la distribución en verde no tiene un pico tan marcado como la distribución mostrada en azul. Entonces, se observa que las alturas se separan en distinta magnitud de la media.

Buscaremos la forma de tener una medida global de las distancias a la media para mostrar cuan dispersos están los datos. Para medir la "distancia" entre una altura cualquiera h_i y la altura media $\bar{h}$ podríamos tomar la diferencia h_i-$\bar{h}$. Las alturas mayores que $\bar{h}$ tendrán signo positivo y las menores tendrán signo negativo. El problema es que cuando sumemos todas las distancias para obtener el promedio como una medida global, se anularán. Las distancias positivas se compensarán con las negativas y la medida global de las distancias será siempre nula. La operación que se realiza para convertir las distancias negativas en positivas y así obtener una medida global de la dispersión de los datos es:

$$\sigma^2 = \frac{1}{N} * \sum_{i=1}^{N} (h_i - \bar{h})^2$$

¿Qué hicimos? sumamos los cuadrados de las diferencias entre cada altura y la altura media y luego, dividimos por la cantidad de muestras. Esto es equivalente a calcular el promedio de los cuadrados de las diferencias entre cada altura con la media. Dentro de nuestro ejemplo de alturas que tienen unidad de cm, esos cuadrados de las diferencias de las distancias tendrán unidad de cm^2 que es unidad de área. Como queremos tener una unidad de alturas, tomamos la raíz cuadrada para volver a tener unidades de distancia:

$$\sigma = \sqrt{\frac{1}{N} * \sum_{i=1}^{N} (h_i - \bar{h})^2}$$

La última expresión es el desvío estándar de las alturas y se lo suele designar con la letra griega sigma minúscula: σ. Así como calculamos esta medida de la dispersión de las alturas, podríamos haber calculado el desvío estándar de cualquier variable numérica de una muestra.

En el caso de la población que graficamos en “azul”, el desvío estándar es σ_{azul}=7,17cm. En el caso de la población “verde” es σ_{verde}=18,21cm. Como σ_{verde} es mayor σ_{azul} podemos concluir que las alturas del conjunto “verde” tienen una mayor dispersión alrededor de la altura media que la población “azul”.

El desvío estándar de un conjunto de *N* datos de tipo *x* es una medida de cuanto se alejan los datos de su media $\bar{x}$ y se calcula como:

$$\sigma = \sqrt{\frac{1}{N} * \sum_{i=1}^{N} (x_i - \bar{x})^2}$$

Una estadística significa resumir un conjunto de datos, que por su cantidad puede abrumar, en un conjunto de valores, gráficos o funciones mucho más simples. Usualmente la estadística implica una reducción de la información, pero también permite hacer un análisis no accesible solamente con la lista de valores. Al analizar un conjunto de datos no sólo hay que observar el valor medio de las variables de interés, también hay que analizar cómo se distribuyen esas variables. Nosotros hemos visto algunos cálculos que arrojan valores que nos permiten describir a una población que conocemos completamente. Tenemos que tener claro que sólo con el valor medio y el desvío estándar no podemos describir completamente cómo se componen todos los datos.

Hay una parte importante de la estadística que trabaja en cómo inferir las características de una población a la que por su naturaleza no podemos acceder en forma completa. Un ejemplo podría ser la cantidad de arsénico en el agua bebible de cada casa, es imposible analizar todas las casas. En estos casos se selecciona un subconjunto del conjunto de todas las casas. Se analiza la variable de la cantidad de arsénico en ese subconjunto de casas y luego, con mucha matemática mediante, se infiere cómo es en el total de la población. Las encuestas son otro ejemplo: se encuesta a una cantidad de personas para que contesten algunas preguntas y se generaliza a toda la población.

A. Indicá si el enunciado es Verdadero o Falso.

La media de una variable en una población indica cuan dispersos se encuentran los datos.

○ Verdadero ○ Falso

B. Indicá si el enunciado es Verdadero o Falso.

La media de la distribución de una variable nos otorga una medida de su promedio pero nada dice sobre cómo es su distribución.

○ Verdadero ○ Falso

C. Indicá si el enunciado es Verdadero o Falso.

Los informes basados en la estadística son objetivos, es decir, no encierran opinión.

○ Verdadero ○ Falso

D. Indicá la respuesta correcta.

Se tiene una serie de mediciones de peso de cuatro personas. Estas mediciones arrojaron los siguientes valores: {65kg, 82kg, 68kg, 57kg}

¿Cuál es la media del peso de esa población?

- 65kg
- 66,5kg
- 68kg
- 69,5kg

E. Indicá la respuesta correcta.

Se tiene una serie de mediciones de peso de cuatro personas. Estas mediciones arrojaron los siguientes valores: { 65kg, 82kg, 68kg, 57kg}

¿Cuál es el desvío estándar del peso de esa población?

- 7,02 kg
- 9,03 kg
- 16,63
- 17,00

Resumen

En esta guía repasaste algunas herramientas de probabilidad y estadística. Estudiamos la forma para distinguir sucesos equiprobables y observamos que, en esos casos, la probabilidad de obtener un resultado se logra dividiendo los eventos favorables por todos los posibles. Abordamos distintas maneras de contar cantidad de sucesos posibles y cómo estos conteos nos sirven para calcular probabilidades. En estadística, estudiamos la media y el desvío estándar de una muestra como valores que nos permite caracterizar a un conjunto de datos.

Claves de corrección

Probabilidad

A. La respuesta correcta es: "Falso". El sorteo que hace la quiniela no tiene información de si la elección del número fue hecha por el apostador o por la computadora de la agencia. El azar no tiene información ni memoria de lo que haya ocurrido antes.

B. La respuesta correcta es: "Verdadero", ya que cada sorteo es un suceso independiente que no está condicionado por lo que haya ocurrido con anterioridad.

C. La respuesta correcta es: "Falso". Si bien la probabilidad de obtener un 2 en cada tirada es 1/6, estamos frente a una situación de eventos al azar. La probabilidad nos dice que en promedio 1 de cada 6 eventos será un "2" pero no se puede asegurar el resultado de un evento o una serie de eventos en particular.

D. La respuesta correcta es: "Falso". La probabilidad de que en el primer sorteo obtengamos un determinado número "a" es 1/6; de la misma forma, la probabilidad de que en el segundo evento obtengamos otro número "b" es 1/6. Al ser eventos independientes, la probabilidad de obtener primero "a" y después "b" es 1/6 * 1/6 = 1/3, y es independiente de cuánto valga "a" o "b" y si "a" y "b" son iguales o diferentes.

E. La respuesta correcta es: "Falso", ya que las probabilidades indican cual suceso tiene más chance de suceder, pero no indican que va a ocurrir seguro. No indican que el evento más probable seguro ocurrirá y el menos probable seguro que no.

F. La respuesta correcta es: "1/45".
Para encontrarla debes utilizar la siguiente expresión:

$$S = \frac{n!}{r! * (n-r)!}$$

que indica la cantidad de posibles grupos de r alumnos que representen al grado. Se aplica con *n*=cantidad de alumnos en el grado, *r*=2 que es la cantidad de alumnos a seleccionar:

$$S = \frac{10!}{2! * (10-2)!} = \frac{10!}{2! * (8)!}$$

$$S = \frac{1 * 2 * 3 * 4 * 5 * 6 * 7 * 8 * 9 * 10}{(1 * 2) * (1 * 2 * 3 * 4 * 5 * 6 * 7 * 8)}$$

$$S = \frac{9 * 10}{1 * 2} = 45$$

Como hay 45 posibles duplas equiprobables y una única dupla que es un evento favorable, su probabilidad es 1/45.

- **G. La respuesta correcta es "$(1/10)^8$"**. Para encontrarla debés hacer el siguiente razonamiento:
  - El primer día puede sortearse cualquier ventanilla, la ventanilla que resulte sorteada determinará que las siguientes deben ser esa y no otra.
  - La probabilidad de que en cualquier día siguiente la ventanilla sorteada sea la del primer día es 1/10.
  - Cada día es un evento independiente, y son 8 días para los cuales la ventanilla tiene que ser la misma que el primer día, entonces:

$$P = (1/10)^8$$

- **H. La respuesta correcta es "P= 1/3"**. Para encontrarla debés hacer el siguiente razonamiento:
  - Hay seis posibles sucesos equiprobables para los números que puede arrojar el dado.
  - Hay dos sucesos favorables: que salga un 4 o que salga un 5.
  - La probabilidad se calcula dividiendo la cantidad de sucesos favorables respecto a los posibles:

$$P = 2/6 = 1/3$$

- **I. La respuesta correcta es: "1/10"**

Para encontrarla debés utilizar la siguiente expresión:

$$S = \frac{n!}{r! * (n-r)!}$$

que indica la cantidad de posibles grupos de *r* colores. Se aplica con *n*=5 (cantidad de papeles en la caja) y *r*=2 (cantidad de papeles que forman un grupo):

$$S = \frac{5!}{2! * (5-2)!} = \frac{5!}{2! * 3!}$$

$$S = \frac{1 * 2 * 3 * 4 * 5}{(1 * 2) * (1 * 2 * 3)}$$

$$S = \frac{4 * 5}{1 * 2} = 10$$

Como hay 10 posibles conjuntos de 10 colores y todos son equiprobables, la probabilidad de cada uno de ellos es 1/10

J. La respuesta correcta es: "P = 1/6".

Para encontrarla debés hacer uno de los siguientes razonamientos (indicamos dos formas de llegar):

- El primer dado puede arrojar cualquier número, el segundo debe copiarlo y la probabilidad de copiarlo es P=1/6
- Hay 6 sucesos posibles para cada dado, y como son independientes, los sucesos posibles para ambos son S=6*6. Los eventos favorables son F=6 (doble "1", doble "2", doble "3", doble "4", doble "5" y doble "6"). Entonces la probabilidad es P = F/S = 6/36 = 1/6

K. La respuesta correcta es: "3/4". Los sucesos posibles son: (cara, cara); (cara, seca); (seca, cara); (seca, seca). De los 4 sucesos posibles hay 3 que son favorables ya que dieron al menos una cara. La probabilidad se obtiene dividiendo la cantidad de sucesos favorables respecto de los posibles: 3/4.

Estadística

A. La respuesta correcta es: "Falso". La media da información sobre el promedio de la variable y se pierde información sobre la dispersión.

B. La respuesta correcta es: "Verdadero", ya que se condensa toda la información en un promedio y se pierde información de cuáles son los valores que llevan a ese promedio y la forma en la que están distribuidos.

C. La respuesta correcta es: "Falso", ya que hay muchos indicadores para hacer una estadística, elegir cuáles indicadores medir y mostrar es una decisión subjetiva.

D. La respuesta correcta es: "68 kg". Para encontrarla se debe hacer la siguiente operación:

$$\bar{p} = \frac{1}{N} \sum_{i=1}^{N} p_i$$

donde llamamos p_i a cada uno de los pesos y N=4 es la cantidad de elementos en la muestra:

$$\bar{p} = \frac{1}{4} \sum_{i=1}^{N} p_i = \frac{65kg + 82kg + 68kg + 57kg}{4} = \frac{272kg}{4} = 68kg$$

E. La respuesta correcta es: "9,03kg". Primero se debe calcular la media del peso p =68kg (es el promedio de los pesos). Luego, se calcula el desvío estándar como:

$$\sigma = \sqrt{\frac{1}{4} * \sum_{i=1}^{N} (p_i - 68kg)^2}$$

$$\sigma = \sqrt{\frac{1}{4} * [(65kg - 68kg)^2 + (82kg - 68kg)^2 + (68kg - 68kg)^2 + (57kg - 68kg)^2]}$$

$$\sigma = \sqrt{\frac{1}{4} * (9 + 196 + 0 + 121)\ kg}$$

$$\sigma = \sqrt{\frac{1}{4} * 326\ kg} = 9{,}03\ kg$$

Ministerio de Educación,
Cultura, Ciencia y Tecnología
Presidencia de la Nación