# Estadística

Alejandro Cáceres (alejandro.caceres.dominguez@upc.edu)

2023-05-04

# Contents

# Chapter 1

# Objetivo

Este es el curso de introducción a la estadística de la EEBE (UPC).

La estadística es un **lenguaje** que permite afrontar problemas nuevos, sobre los que no tenemos solución, y en donde interviene la **aleatoridad**.

En este curso trataremos los **conceptos fundamentales** de estadística.

- 3 horas de **teoría** por semana: Explicaremos los conceptos, haremos ejercicios.
- 6 horas de **estudio individual** por semana: Notas notas de curso y los recursos en ATENEA.
- 2 horas de Solución de problemas con **R**: Sesiones presenciales con ordenador (Prácticas).

Las fechas de exámenes y material de estudio adicional se pueden encontrar en **ATENEA metacurso**:

Objetivos de evaluación:

**Q1** (10%): Prueba en ordenador duración 2h en las fechas indicadas.

a. Dominio de comandos básicos en R (Prácticas)
b. Capacidad de calcular estadísticos descriptivos y gráficos, en situaciones concretas (Teoría/Práctica)
c. Conocimiento sobre la regresión lineal (Prácticas)

**EP1** (25%): Prueba escrita (2-3 problemas)

a. Capacidad de interpretación de enunciados en fórmulas de probabilidad (Teoría).
b. Conocimiento de las herramientas básicas para solucionar problemas de probabilidad conjunta y probabilidad condicional (Teoría).

c. Dominio matemático de funciones de probabilidad para calcular sus propiedades básicas (Teoría).

**Q2** (10%): Prueba en ordenador duración 2h en las fechas indicadas

a. Capacidad de identificación de modelos de probabilidad en problemas concretos (Teoría/Práctica).
b. Uso de funciones de R para calcular probabilidades de modelos probabilísticos (Práctica/Teoría)

**Q3** (10%): Prueba en ordenador duración 2h en las fechas indicadas

a. Capacidad de identificación de un estadístico de muestreo y sus propiedades (Teoría/Práctica)
b. Conocimiento de cómo calcular la probabilidad de los estadísticos de muestreo (Teoría/Práctica)
c. Uso de comandos en R para calcular probabilidades y hacer simulaciones de muestras aleatorias (Prácticas)

**EP2** (10%): Prueba escrita (2-3 problemas)

a. Capacidad matemática para determinar estimadores puntuales de modelos de probabilidad.
b. Conociemiento de las propiedades de los estimadores puntuales.

**CG** (5%): Prueba escrita (2 preguntas sobre un texto)

a. Capacidad de expresión escrita sobre un tema relacionado a la estadística.

**EP3** (30%): Prueba por ordenador presencial (2-3 problemas)

a. Conocimiento de los intervalos de confianza y sus propiedades (Teoría).
b. Capacidad de identificar el tipo de intervalo de confianza en un problema concreto (Teoría).
c. Capacidad de interpretación del tipo de hipótesis a usar en un problema concreto (Teoría).
d. Propiedades de las pruebas de hipótesis.
e. Uso de comandos en R para resolver problemas de intervalos de confianza y prueabas de hipótesis (Práctica).

coordinadores:

- Luis Mujica (luis.eduardo.mujica@upc.edu)
- Pablo Buenestado (pablo.buenestado@upc.edu)

## 1.1 Lectura recomendada

- Las notas de clase se nuestra sección estarán accesibles en ATENEA en pdf y en html.

- Douglas C. Montgomery and George C. Runger. "Applied Statistics and Probability for Engineers" 4th Edition. Wiley 2007.

# Chapter 2

# Descripción de datos

En este capítulo, presentaremos herramientas para describir datos.

Lo haremos utilizando tablas, figuras y estadísticos descriptivos de tendencia central y dispersión.

También presentaremos conceptos clave en estadística como experimentos aleatorios, observaciones, resultados y frecuencias absolutas y relativas.

## 2.1 Método científico

Uno de los objetivos del método científico es proporcionar un marco para resolver los problemas que surgen en el estudio de los fenómenos naturales o en el diseño de nuevas tecnologías.

Los humanos modernos han desarrollado un **método** durante miles de años que todavía está en desarrollo.

El método tiene tres actividades humanas principales:

- *Observación* caracterizada por la adquisición de **datos**
- *Razón* caracterizada por el desarrollo de **modelos** matemáticos
- *Acción* caracterizada por el desarrollo de nuevos **experimentos** (tecnología)

Su compleja interacción y resultados son la base de la *actividad científica*.

## 2.2 Estadística

La estadística se ocupa de la interacción entre *modelos* y *datos* (la parte inferior de la figura).

Las preguntas de tipo estadístico son:

- ¿Cuál es el mejor modelo para mis datos (inferencia)?
- ¿Cuáles son los datos que produciría un determinado modelo (predicción)?

## 2.3 Datos

Los datos se presentan en forma de observaciones.

Una **Observación** o *Realización* es la adquisición de un número o una característica de un experimento.

Por ejemplo, tomemos la serie de números que se producen por la repetición de un experimento (1: éxito, 0: fracaso)

… 1 0 0 1 0 **1** 0 1 1 …

El número en negrita es **una observación** en una repetición del experimento

Un **resultado** es una **posible** observación que es el resultado de un experimento.

**1** es un resultado, **0** es el otro resultado del experimento.

Recuerda que la observación es **concreta** es el número que obtienes un día en el laboratorio. El resultado **abstracto** es una de las características del tipo de experimento que estás realizando.

## 2.4 Tipos de resultado

En estadística nos interesan principalmente dos tipos de resultados.

- **Categóricos**: Si el resultado de un experimento es una cualidad. Pueden ser nominales (binario: sí, no; múltiple: colores) u ordinales cuando las cualidades pueden jerarquizarse (gravedad de una enfermedad).
- **Numéricos**: Si el resultado de un experimento es un número. El número puede ser discreto (número de correos electrónicos recibidos en una hora, número de leucocitos en sangre) o continuo (estado de carga de la batería, temperatura del motor).

## 2.5 Experimentos aleatorios

Se puede decir que el tema de estudio de la estadística son los experimentos aleatorios, el medio por el cual producimos datos.

**Definición:**

Un **experimento aleatorio** es un experimento que da diferentes resultados cuando se repite de la misma manera.

Los experimentos aleatorios son de diferentes tipos, dependiendo de cómo se realicen:

- en el mismo objeto (persona): temperatura, niveles de azúcar.
- sobre objetos diferentes pero de la misma medida: el peso de un animal.
- sobre eventos: el número de huracanes por año.

## 2.6 Frecuencias absolutas

Cuando repetimos un experimento aleatorio con resultados **categóricos**, registramos una lista de resultados.

Resumimos las observaciones contando cuántas veces vimos un resultado particular.

**Frecuencia absoluta**:

$$n_i$$

es el número de veces que observamos el resultado $i$.

**Ejemplo (leucocitos)**

Extraigamos un leucocito de **un** donante y anotemos su tipo. Repitamos el experimento $N = 119$ veces.

`(célula T, célula T, neutrófilo, ..., célula B)`

La segunda **célula T** en negrita es la segunda observación. La última **célula B** es la observación número 119.

Podemos listar los **resultados** (categorías) en una **tabla de frecuencia**:

```
##      outcome ni
## 1     T Cell 34
## 2     B cell 50
## 3   basophil 20
## 4   Monocyte  5
## 5 Neutrophil 10
```

De la tabla, podemos decir que, por ejemplo, $n_1 = 34$ es el número total de células T observadas en la repetición del experimento. También notamos que el número total de repeticiones $N = \sum_i n_i = 119$.

## 2.7 Frecuencias relativas

También podemos resumir las observaciones calculando la **proporción** de cuántas veces vimos un resultado en particular.

$$f_i = n_i/N$$

donde $N$ es el número total de observaciones

En nuestro ejemplo se registraron $n_1 = 34$ células T, por lo que nos preguntamos por la proporción de células T del total de 119. Podemos agregar estas proporciones $f_i$ en la tabla las frecuencias.

```
##      outcome ni         fi
## 1     T Cell 34 0.28571429
## 2     B cell 50 0.42016807
## 3   basophil 20 0.16806723
## 4   Monocyte  5 0.04201681
## 5 Neutrophil 10 0.08403361
```

Las frecuencias relativas son **fundamentales** en estadística. Dan la proporción de un resultado en relación con los otros resultados. Más adelante las entenderemos como las observaciones de las probabilidades.

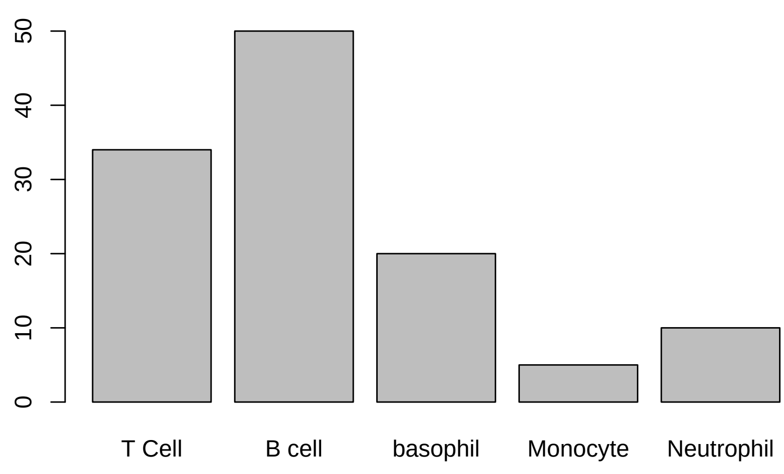
Para las frecuencias absolutas y relativas tenemos las propiedades

- $\sum_{i=1..M} n_i = N$
- $\sum_{i=1..M} f_i = 1$

donde $M$ es el número de resultados.

## 2.8 Diagrama de barras

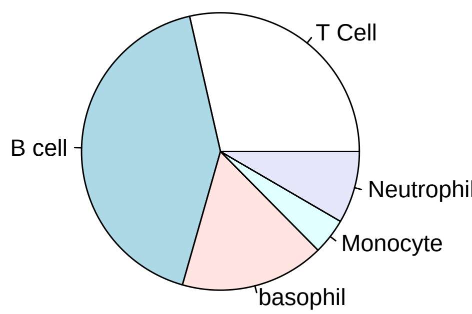
Cuando tenemos muchos resultados y queremos ver cuáles son los más probables, podemos usar un gráfico de barras que es una cifra de $n_i$ Vs los resultados.



## 2.9 Gráfico de sectores (pie)

También podemos visualizar las frecuencias relativas con un gráfico de sectores.

El área del círculo representa el 100% de las observaciones (proporción = 1) y las secciones las frecuencias relativas de cada resultado.




## 2.10 Variables categóricas ordinales

El tipo de leucocito de los ejemplos anteriores es una variable nominal **categórica**. Cada observación pertenece a una categoría (cualidad). Las categorías no siempre tienen un orden determindado.

A veces, las variables **categóricas** se pueden **ordenar** cuando cumplen una clasificación natural. Esto permite introducir **frecuencias acumulativas**.

**Ejemplo (misofonía)**

Este es un estudio clínico en 123 pacientes que fueron examinados por su grado de misofonía. La misofnía es ansiedad/ira descontrolada producida por ciertos sonidos.

Cada paciente fue evaluado con un cuestionario (AMISO) y se clasificaron en 4 grupos diferentes según la gravedad.

Los resultados del estudio son

```
##   [1] 4 2 0 3 0 0 2 3 0 3 0 2 2 0 2 0 0 3 3 0 3 3 2 0 0 0 4 2 2 0 2 0 0 0 3 0 2
```

```
##  [38] 3 2 2 0 2 3 0 0 2 2 3 3 0 0 4 3 3 2 0 2 0 0 0 2 2 0 0 2 3 0 1 3 2 4 3 2 3
##  [75] 0 2 3 2 4 1 2 0 2 0 2 0 2 2 4 3 0 3 0 0 0 2 2 1 3 0 0 3 2 1 3 0 4 4 2 3 3
## [112] 3 0 3 2 1 2 3 3 4 2 3 2
```

Cada observación es el resultado de un experimento aleatorio: medición del nivel de misofonía en un paciente. Esta serie de datos se puede resumir en términos de los resultados en la tabla de frecuencia

```
##   outcome ni         fi
## 1       0 41 0.33333333
## 2       1  5 0.04065041
## 3       2 37 0.30081301
## 4       3 31 0.25203252
## 5       4  9 0.07317073
```

## 2.11 Frecuencias acumuladas absolutas y relativas

La gravedad de la misofonía es **categórica ordinal** porque sus resultados pueden ordenarse en relación con su grado.

Cuando los resultados se pueden ordenar, es útil preguntar cuántas observaciones se obtuvieron hasta un resultado dado. Llamamos a este número la **frecuencia acumulada absoluta** hasta el resultado $i$:

$$N_i = \sum_{k=1..i} n_k$$

También es útil para calcular la **proporción** de las observaciones que se obtuvo hasta un resultado dado

$$F_i = \sum_{k=1..i} f_k$$

Podemos agregar estas frecuencias en la **tabla de frecuencias**

```
##   outcome ni         fi  Ni        Fi
## 0       0 41 0.33333333  41 0.3333333
## 1       1  5 0.04065041  46 0.3739837
## 2       2 37 0.30081301  83 0.6747967
## 3       3 31 0.25203252 114 0.9268293
## 4       4  9 0.07317073 123 1.0000000
```

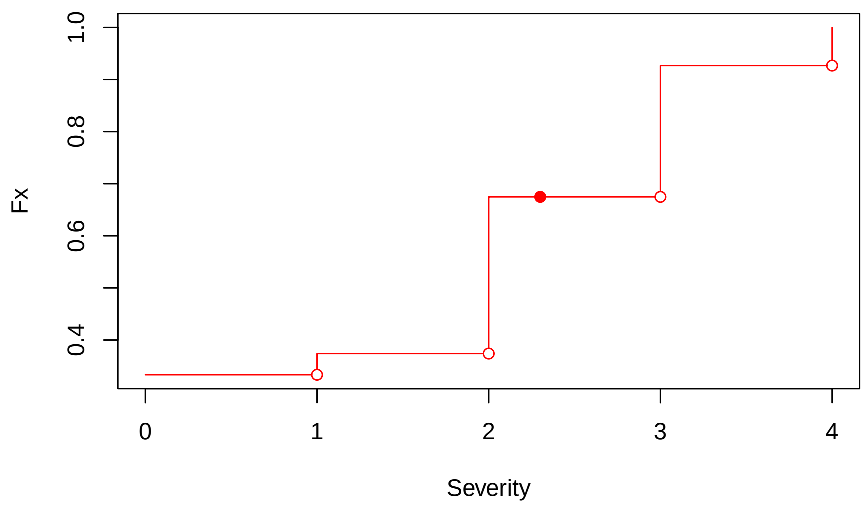
Por lo tanto, el **67 %** de los pacientes tenían misofonía hasta la gravedad **2** y el **37 %** de los pacientes tenían una gravedad inferior o igual a **1**.

## 2.12 Gráfica de frecuencia acumulada

$F_i$ es una cantidad importante porque nos permite definir la acumulación de probabilidades hasta niveles intermedios.

La probabilidad de un nivel intermedio $x$ ($i \leq x < i+1$) es solo la acumulación hasta el nivel inferior $F_x = F_i$.

$F_x$ es por lo tanto una función de rango **continuo**. Podemos dibujarla con respecto a los resultados.




Por lo tanto, podemos decir que el **67 %** de los pacientes tenían misofonía hasta gravedad 2.3, aunque 2.3 no es un resultado observado.

## 2.13 Variables numéricas

El resultado de un experimento aleatorio puede producir un número. Si el número es **discreto**, podemos generar una tabla de frecuencias, con frecuencias absolutas, relativas y acumulativas, e ilustrarlas con gráficos de barras, de sectores y acumulativos.

Cuando el número es **continuo** las frecuencias no son útiles, lo más probable es que observemos o no un número contínuo en particular.

**Ejemplo (misofonía)**

Los investigadores se preguntaron si la convexidad de la mandíbula afectaría la gravedad de la misofonía. La hipótesis científica es que el ángulo de convexidad de la mandíbula puede influir en el oído y su sensibilidad. Estos son los resultados de la convexidad de la mandíbula (grados) para cada paciente:

```
##   [1]  7.97 18.23 12.27  7.81  9.81 13.50 19.30  7.70 12.30  7.90 12.60 19.00
##  [13]  7.27 14.00  5.40  8.00 11.20  7.75  7.94 16.69  7.62  7.02  7.00 19.20
##  [25]  7.96 14.70  7.24  7.80  7.90  4.70  4.40 14.00 14.40 16.00  1.40  9.76
##  [37]  7.90  7.90  7.40  6.30  7.76  7.30  7.00 11.23 16.00  7.90  7.29  6.91
##  [49]  7.10 13.40 11.60 -1.00  6.00  7.82  4.80 11.00  9.00 11.50 16.00 15.00
##  [61]  1.40 16.80  7.70 16.14  7.12 -1.00 17.00  9.26 18.70  3.40 21.30  7.50
##  [73]  6.03  7.50 19.00 19.01  8.10  7.80  6.10 15.26  7.95 18.00  4.60 15.00
##  [85]  7.50  8.00 16.80  8.54  7.00 18.30  7.80 16.00 14.00 12.30 11.40  8.50
##  [97]  7.00  7.96 17.60 10.00  3.50  6.70 17.00 20.26  6.64  1.80  7.02  2.46
## [109] 19.00 17.86  6.10  6.64 12.00  6.60  8.70 14.05  7.20 19.70  7.70  6.02
## [121]  2.50 19.00  6.80
```

## 2.14 Transformando datos continuos

Como los resultados continuos no se pueden contar (de manera informativa), los transformamos en variables categóricas ordenadas.

1) Primero cubrimos el rango de las observaciones en intervalos regulares del mismo tamaño (contenedores)

```
## [1] "[-1.02,3.46]" "(3.46,7.92]"  "(7.92,12.4]"  "(12.4,16.8]"  "(16.8,21.3]"
```

2) Luego mapeamos cada observación a su intervalo: creando una variable categórica **ordenada**; en este caso con 5 resultados posibles

```
##   [1] "(7.92,12.4]"  "(16.8,21.3]"  "(7.92,12.4]"  "(3.46,7.92]"  "(7.92,12.4]" 
##   [6] "(12.4,16.8]"  "(16.8,21.3]"  "(3.46,7.92]"  "(7.92,12.4]"  "(3.46,7.92]" 
##  [11] "(12.4,16.8]"  "(16.8,21.3]"  "(3.46,7.92]"  "(12.4,16.8]"  "(3.46,7.92]" 
##  [16] "(7.92,12.4]"  "(7.92,12.4]"  "(3.46,7.92]"  "(7.92,12.4]"  "(12.4,16.8]" 
##  [21] "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(16.8,21.3]"  "(7.92,12.4]" 
##  [26] "(12.4,16.8]"  "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]" 
##  [31] "(3.46,7.92]"  "(12.4,16.8]"  "(12.4,16.8]"  "(12.4,16.8]"  "[-1.02,3.46]"
##  [36] "(7.92,12.4]"  "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]" 
##  [41] "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(7.92,12.4]"  "(12.4,16.8]" 
##  [46] "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(12.4,16.8]" 
##  [51] "(7.92,12.4]"  "[-1.02,3.46]" "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]" 
##  [56] "(7.92,12.4]"  "(7.92,12.4]"  "(7.92,12.4]"  "(12.4,16.8]"  "(12.4,16.8]" 
##  [61] "[-1.02,3.46]" "(12.4,16.8]"  "(3.46,7.92]"  "(12.4,16.8]"  "(3.46,7.92]" 
##  [66] "[-1.02,3.46]" "(16.8,21.3]"  "(7.92,12.4]"  "(16.8,21.3]"  "[-1.02,3.46]"
##  [71] "(16.8,21.3]"  "(3.46,7.92]"  "(3.46,7.92]"  "(3.46,7.92]"  "(16.8,21.3]" 
##  [76] "(16.8,21.3]"  "(7.92,12.4]"  "(3.46,7.92]"  "(3.46,7.92]"  "(12.4,16.8]" 
##  [81] "(7.92,12.4]"  "(16.8,21.3]"  "(3.46,7.92]"  "(12.4,16.8]"  "(3.46,7.92]" 
##  [86] "(7.92,12.4]"  "(12.4,16.8]"  "(7.92,12.4]"  "(3.46,7.92]"  "(16.8,21.3]" 
```

```
##  [91] "(3.46,7.92]"  "(12.4,16.8]"  "(12.4,16.8]"  "(7.92,12.4]"  "(7.92,12.4]"
##  [96] "(7.92,12.4]"  "(3.46,7.92]"  "(7.92,12.4]"  "(16.8,21.3]"  "(7.92,12.4]"
## [101] "(3.46,7.92]"  "(3.46,7.92]"  "(16.8,21.3]"  "(16.8,21.3]"  "(3.46,7.92]"
## [106] "[-1.02,3.46]" "(3.46,7.92]"  "[-1.02,3.46]" "(16.8,21.3]"  "(16.8,21.3]"
## [111] "(3.46,7.92]"  "(3.46,7.92]"  "(7.92,12.4]"  "(3.46,7.92]"  "(7.92,12.4]"
## [116] "(12.4,16.8]"  "(3.46,7.92]"  "(16.8,21.3]"  "(3.46,7.92]"  "(3.46,7.92]"
## [121] "[-1.02,3.46]" "(16.8,21.3]"  "(3.46,7.92]"
```

Por tanto, en lugar de decir que el primer paciente tenía un ángulo de convexidad de 7.97, decimos que su ángulo estaba entre el intervalo (o **bin**) $(7.92, 12.4]$.

Ningún otro paciente tenía un ángulo de 7.97, pero muchos tenían ángulos entre $(7.92, 12.4]$.

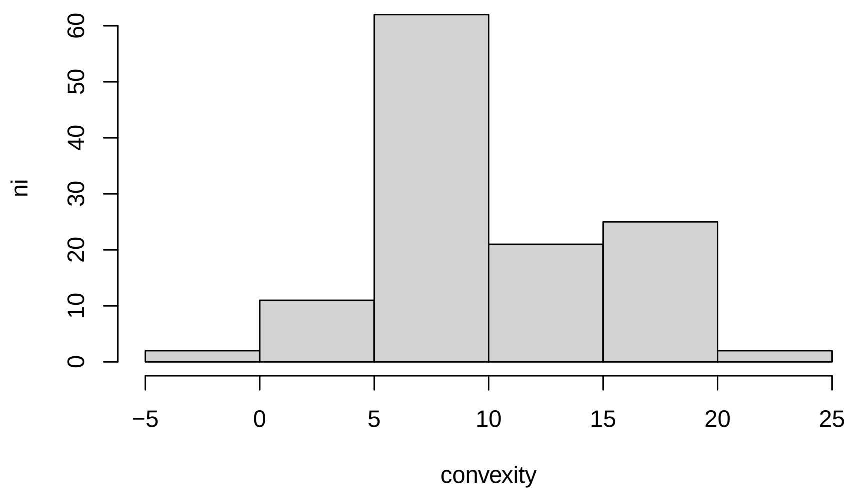
## 2.15 Tabla de frecuencias para una variable continua

Para una partición regular dada del intervalo de resultados en intervalos, podemos producir una tabla de frecuencias como antes

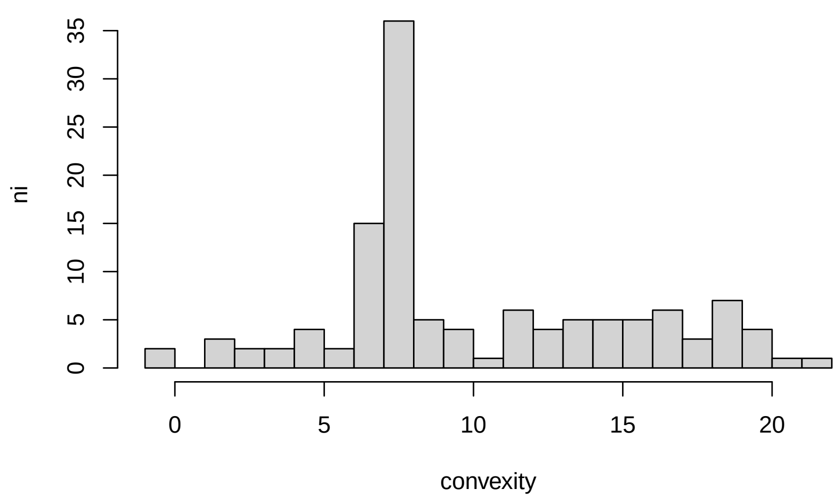
```
##         outcome ni         fi  Ni         Fi
## 1 [-1.02,3.46]  8 0.06504065   8 0.06504065
## 2  (3.46,7.92] 51 0.41463415  59 0.47967480
## 3  (7.92,12.4] 26 0.21138211  85 0.69105691
## 4  (12.4,16.8] 20 0.16260163 105 0.85365854
## 5  (16.8,21.3] 18 0.14634146 123 1.00000000
```

## 2.16 Histograma

El histograma es la gráfica de $n_i$ o $f_i$ Vs los resultados en intervalos (bins). El histograma depende del tamaño de los bins.




Este es un histograma con 20 bins.




Vemos que la mayoría de las personas tienen ángulos dentro de $(7, 8]$

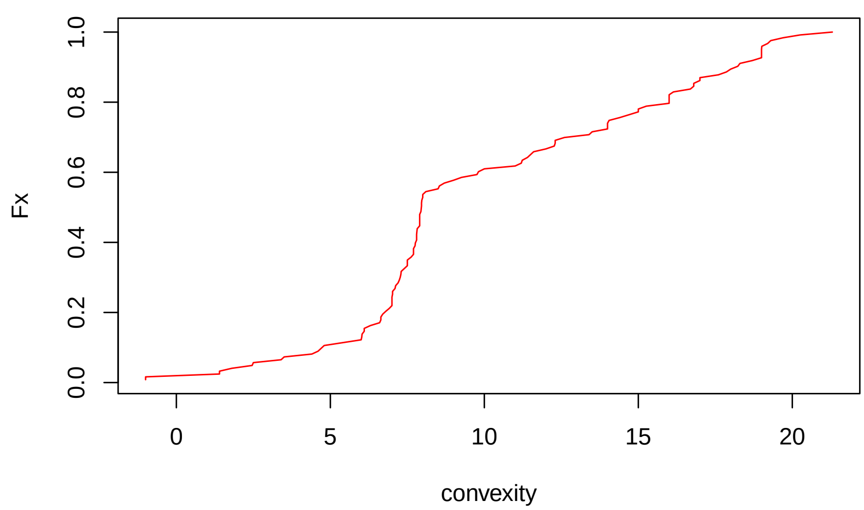
## 2.17 Gráfica de frecuencia acumulada

También podemos graficar $F_x$ contra los resultados. Como $F_x$ es de rango continuo, podemos ordenar las observaciones ($x_1 < ...x_j < x_{j+1} < x_n$) y por lo tanto

$$F_x = \frac{k}{n}$$

para $x_k \leq x < x_{k+1}$.

$F_x$ se conoce como la **distribución** de los datos. $F_x$ no depende del tamaño del bin. Sin embargo, su **resolución** depende de la cantidad de datos.




## 2.18 Estadísticas de resumen

Las estadísticas de resumen son números calculados a partir de los datos que nos dicen características importantes de las variables numéricas (discretas o continuas).

Por ejemplo, tenemos estadísticas que describen los valores extremos:

- **mínimo**: el resultado mínimo observado
- **máximo**: el resultado máximo observado

## 2.19 Promedio (media muestral)

Una estadística importante que describe el valor central de los resultados (dónde esperar la mayoría de las observaciones) es el **promedio**

$$\bar{x} = \frac{1}{N} \sum_{j=1..N} x_j$$

donde $x_j$ es la **observación** $j$ de un total de $N$.

**Ejemplo (Misofonía)**

La convexidad promedio se puede calcular directamente a partir de las **observaciones**

$\bar{x} = \frac{1}{N} \sum_j x_j$

$= \frac{1}{N}(7.97 + 18.23 + 12.27... + 6.80) = 10.19894$

Para variables **categóricamente ordenadas**, podemos usar las frecuencias relativas para calcular el promedio

$\bar{x} = \frac{1}{N} \sum_{i=1...N} x_j = \frac{1}{N} \sum_{i=1...M} x_i * n_i$

$$= \sum_{i=1...M} x_i * f_i$$

donde pasamos de sumar $N$ **observaciones** a sumar $M$ **resultados**.

La forma $\bar{x} = \sum_{i=1...M} x_i f_i$ muestra que el promedio es el **centro de gravedad** de los resultados. Como si cada resultado tuviera una densidad de masa dada por $f_i$.

**Ejemplo (Misofonía)**

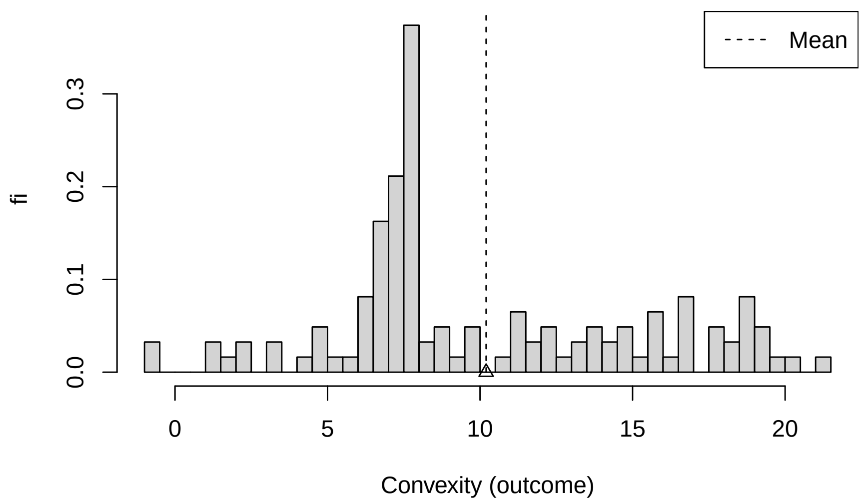
La **severidad** promedio de la misofonía en el estudio se puede calcular a partir de las frecuencias relativas de los **resultados**

```
##   outcome ni         fi
## 1       0 41 0.33333333
## 2       1  5 0.04065041
## 3       2 37 0.30081301
## 4       3 31 0.25203252
## 5       4  9 0.07317073
```

$\bar{x} = 0 * f_0 + 1 * f_1 + 2 * f_2 + 3 * f_3 + 4 * f_4 = 1.691057$

## 2.20 Promedio

El promedio es también el centro de gravedad de las variables continuas. Ese es el punto donde las frecuencias reativas se equilibran.




## 2.21 mediana

Otra medida de centralidad es la mediana. La mediana $x_m$, o $q_{0.5}$, es el valor por debajo del cual encontramos la mitad de las observaciones. Cuando ordenamos las observaciones $x_1 < ...x_j < x_{j+1} < x_N$, las contamos hasta encontrar la mitad de ellas. $x_m$ es tal que

$$\sum_{i \leq m} 1 = \frac{N}{2}$$

**Ejemplo (Misofonía)**

Si ordenamos los ángulos de convexidad, vemos que 62 observaciones (individuos) ($N/2 \sim 123/2$) están por debajo de 7.96. La **convexidad mediana** es por lo tanto $q_{0.5} = x_{62} = 7.96$

```
##  [1] -1.00 -1.00  1.40  1.40  1.80  2.46  2.50  3.40  3.50  4.40  4.60  4.70
## [13]  4.80  5.40  6.00  6.02  6.03  6.10  6.10  6.30  6.60  6.64  6.64  6.70
## [25]  6.80  6.91  7.00  7.00  7.00  7.00  7.02  7.02  7.10  7.12  7.20  7.24
## [37]  7.27  7.29  7.30  7.40  7.50  7.50  7.50  7.62  7.70  7.70  7.70  7.75
## [49]  7.76  7.80  7.80  7.80  7.81  7.82  7.90  7.90  7.90  7.90  7.90  7.94
```

```
## [61]  7.95  7.96
```

```
##  [1]  7.96  7.97  8.00  8.00  8.10  8.50  8.54  8.70  9.00  9.26  9.76  9.81
## [13] 10.00 11.00 11.20 11.23 11.40 11.50 11.60 12.00 12.27 12.30 12.30 12.60
## [25] 13.40 13.50 14.00 14.00 14.00 14.05 14.40 14.70 15.00 15.00 15.26 16.00
## [37] 16.00 16.00 16.00 16.14 16.69 16.80 16.80 17.00 17.00 17.60 17.86 18.00
## [49] 18.23 18.30 18.70 19.00 19.00 19.00 19.00 19.01 19.20 19.30 19.70 20.26
## [61] 21.30
```

```
## [1] 7.96
```

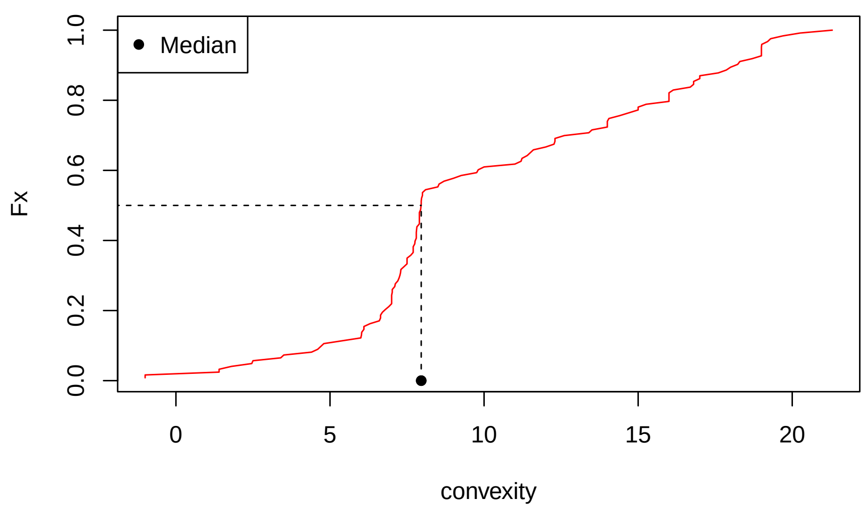
En términos de frecuencias, $q_{0.5}$ hace que la frecuencia acumulada $F_x$ sea igual a 0.5

$$\sum_{i=0,\ldots m} f_i = F_{q_{0.5}} = 0.5$$

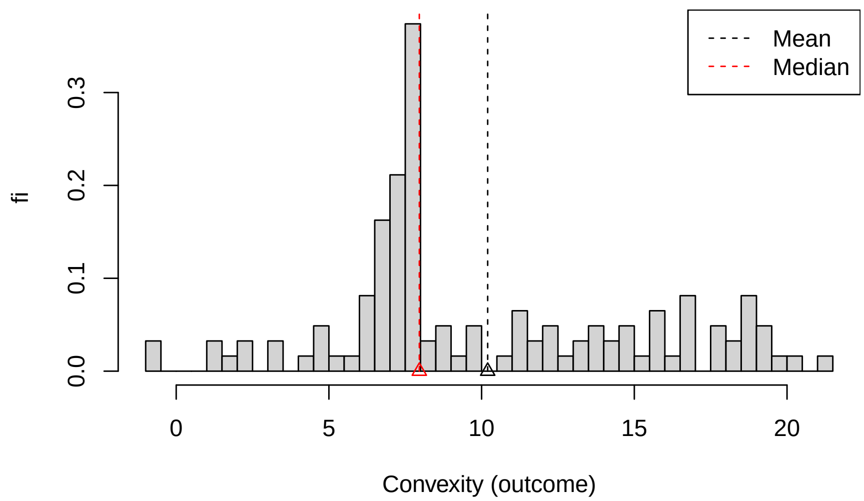
o

$$q_{0.5} = F^{-1}(0.5)$$

En el gráfico de distribución, la mediana es el valor de $x$ en el que se encuentra la mitad del máximo de $F$.




El promedio y la mediana no siempre son iguales.

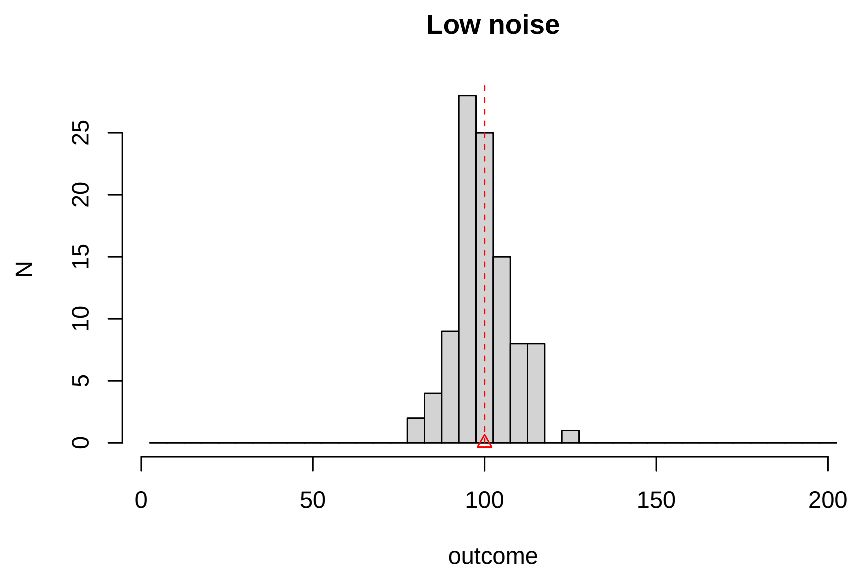


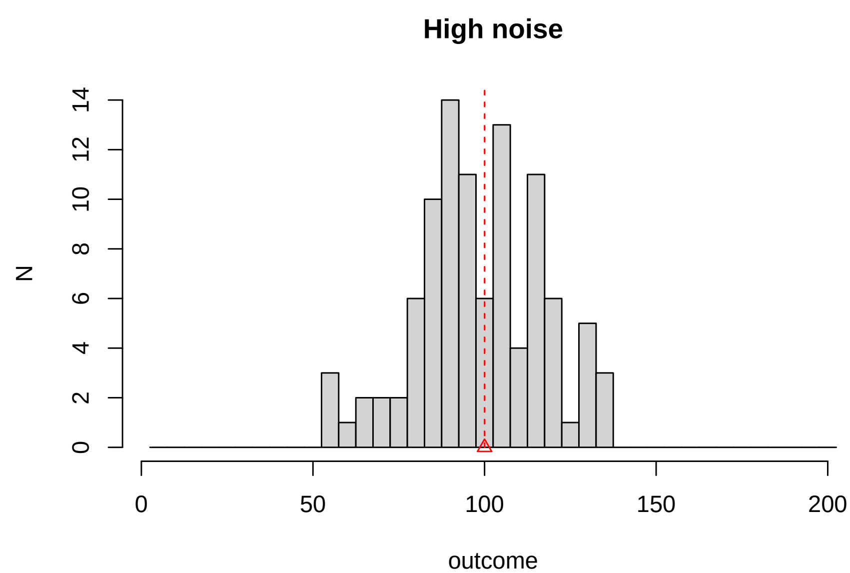

## 2.22 Dispersión

Otras estadísticas de resumen importantes de las observaciones son las de **dispersión**.

Muchos experimentos pueden compartir su media, pero difieren en cuán **dispersos** son los valores.

La dispersión de las observaciones es una medida del **ruido**.




## 2.23 Variación de la muestra

La dispersión sobre la media se mide con la varianza muestral

$$s^2 = \frac{1}{N-1}\sum_{j=1..N}(x_j - \bar{x})^2$$

Este número, mide la distancia cuadrada promedio de las **observaciones** al promedio. La razón de $N-1$ se explicará cuando hablemos de inferencia, cuando estudiemos la dispersión de $\bar{x}$, además de la dispersión de las observaciones.

En términos de las frecuencias de las variables **categóricas y ordenadas**

$$s^2 = \frac{N}{N-1}\sum_{i=1...M}(x_i - \bar{x})^2 f_i$$

$s^2$ se puede considerar como el **momento de inercia** de las observaciones.

La raíz cuadrada de la varianza de la muestra se denomina **desviación estándar** $s$.

**Ejemplo (Misofonía)**

La desviación estándar del ángulo de convexidad es

$s = [\frac{1}{123-1}((7.97 - 10.19894)^2 + (18.23 - 10.19894)^2 + (12.27 - 10.19894)^2 + ...)]^{1/2} = 5.086707$

La convexidad de la mandíbula se desvía de su media en 5.086707.

## 2.24 Rango intercuartílico (IQR)

La dispersión de los datos también se puede medir con respecto a la mediana usando el **rango intercuartílico**:

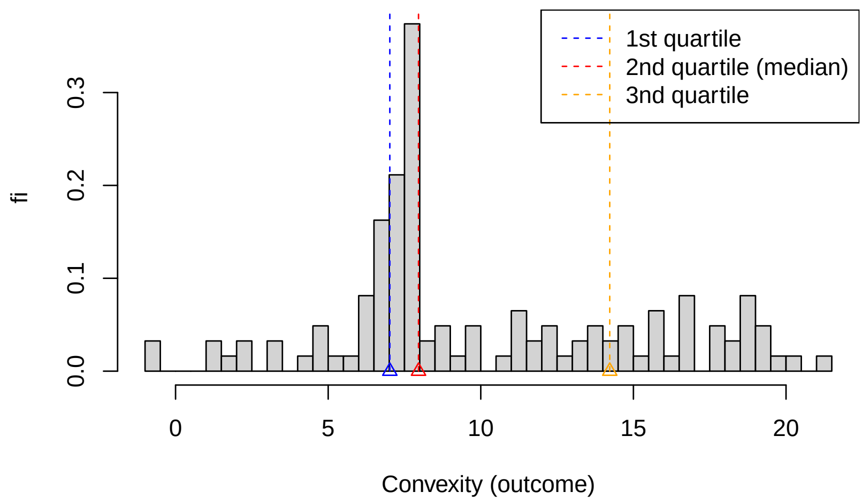
1) Definimos el **primer** cuartil como el valor $x_m$ que hace que la frecuencia acumulada $F_{q_{0.25}}$ sea igual a 0.25 ($x$ donde hemos acumulado una cuarta parte de las observaciones)

$$F_{q_{0.25}} = 0.25$$

1) Definimos el **tercer** cuartil como el valor $x_m$ que hace que la frecuencia acumulada $F_{q_{0.75}}$ sea igual a 0.75 ($x$ donde hemos acumulado tres cuartos de observaciones)

$$F_{q_{0.75}} = 0.75$$

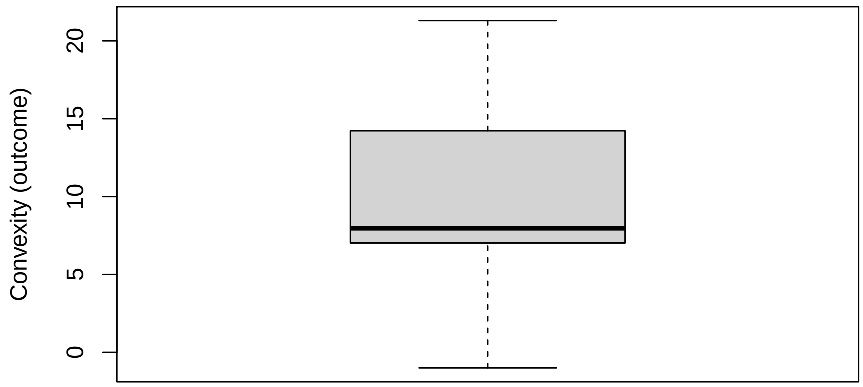
3) El **rango intercuartílico** (IQR) es $IQR = q_{0.75} - q_{0.25}$. Esa es la distancia entre el tercer y el primer cuartil y captura el 50% central de las observaciones




## 2.25 Diagrama de caja

El rango intercuartílico, la mediana y los 5% y 95% de los datos se pueden visualizar en un **diagrama de caja**.

En el diagrama de caja, los valores de los resultados están en el eje y. El IQR es la caja, la mediana es la línea del medio y los bigotes marcan los 5% y 95% de los datos.




## 2.26 Preguntas

**1)** En el siguiente diagrama de caja, el primer cuartil y el segundo cuartil de los datos son:

**a:** $(-1.00, 21.30)$; **b:** $(-1.00, 7.02)$; **c:** $(7.02, 7.96)$; **d:** $(7.02, 14.22)$

**2)** La principal desventaja de un histograma es que:

**a:** Depende del tamaño del bin; **b:** No se puede utilizar para variables categóricas; **c:** No se puede usar cuando el tamaño del bin es pequeño; **d:** Se usa solo para frecuencias relativas;

**3)** Si las frecuencias acumuladas relativas de un experimento aleatorio con resultados $\{1, 2, 3, 4\}$ son: $F(1) = 0.15$, $F(2) = 0.60$, $F(3) = 0.85$, $F(4) = 1$.

Entonces la frecuencia relativa para el resultado 3 es

**a:** 0.15; **b:** 0.85; **c:** 0.45; **d:** 0.25

**4)** En una muestra de tamaño 10 de un experimento aleatorio obtuvimos los siguientes datos:

8, 3, 3, 7, 3, 6, 5, 10, 3, 8.

El primer cuartil de los datos es:

**a:** 3.5; **b:** 4; **c:** 5; **d:** 3

**5)** Imaginemos que recopilamos datos para dos cantidades que no son mutuamente excluyentes, por ejemplo, el sexo y la nacionalidad de los pasajeros de un vuelo. Si queremos hacer un solo gráfico circular para los datos, ¿cuál de estas afirmaciones es verdadera?

**a:** Solo podemos hacer un gráfico circular de nacionalidad porque tiene más de dos resultados posibles; **b:** Podemos hacer un gráfico circular para una variable nueva que marca el sexo **y** la nacionalidad; **c:** Podemos hacer un gráfico circular para la variale sexo o la variable nacionalidad; **d:** Solo podemos elegir si hacemos un gráfico circular para el sexo o un gráfico circular para la nacionalidad.

## 2.27 Ejercicios

#### 2.27.0.1 Ejercicio 1

Hemos realizado un experimento 8 veces con los siguientes resultados

```
## [1]  3  3 10  2  6 11  5  4
```

Responde las siguientes cuestiones:

- Calcula las frecuencias relativas de cada resultado.
- Calcula las frecuencias acumuladas de cada resultado.
- ¿Cuál es el promedio de las observaciones?
- ¿Cuál es la mediana?
- ¿Cuál es el tercer cuartil?
- ¿Cuál es el primer cuartil?

#### 2.27.0.2 Ejercicio 2

Hemos realizado un experimento 10 veces con los siguientes resultados

```
##  [1] 2.875775 7.883051 4.089769 8.830174 9.404673 0.455565 5.281055 8.924190
##  [9] 5.514350 4.566147
```

Considera 10 bins de tamaño 1: [0,1], (1,2]…(9,10).

Responde las siguientes cuestiones:

- Calcula las frecuencias relativas de cada resultado y dibuja el histograma
- Calcula las frecuencias acumulativas de cada resultado y dibuja la gráfica acumulativa.
- Dibuja un diagrama de caja .

# Chapter 3

# Probabilidad

En este capítulo introduciremos el concepto de probabilidad a partir de frecuencias relativas.

Definiremos los eventos como los elementos sobre los que se aplica la probabilidad. Los eventos compuestos se definirán usando álgebra de conjuntos.

Luego discutiremos el concepto de probabilidad condicional derivado de la probabilidad conjunta de dos eventos.

## 3.1 Experimentos aleatorios

Recordemos el objetivo básico de la estadística. La estadística se ocupa de los datos que se presentan en forma de observaciones.

- Una **observación** es la adquisición de un número o una característica de un experimento

Las observaciones son realizaciones de **resultados**.

- Un **resultado** es una posible observación que es el resultado de un experimento.

Al realizar experimentos, a menudo obtenemos resultados diferentes. La descripción de la variabilidad de los resultados es uno de los objetivos de la estadística.

- Un **experimento aleatorio** es un experimento que da diferentes resultados cuando se repite de la misma manera.

La pregunta filosófica detrás es ¿Cómo podemos conocer algo si cada vez que lo miramos cambia?

## 3.2 Probabilidad de medición

Nos gustaría tener una medida para el resultado de un experimento aleatorio que nos diga **cuán seguros** estamos de observar el resultado cuando realicemos un **futuro** experimento aleatorio.

Llamaremos a esta medida la probabilidad del resultado y le asignaremos valores:

- 0, cuando estamos seguros de que la observación **no** ocurrirá.
- 1, cuando estamos seguros de que la observación sucederá.

## 3.3 Probabilidad clásica

**Siempre que** un experimento aleatorio tenga $M$ resultados posibles que son todos **igualmente probables**, la probabilidad de cada resultado $i$ es

$$P_i = \frac{1}{M}$$

.

La probabilidad clásica fue defendida por Laplace (1814).

Dado que cada resultado es **igualmente probable** en este tipo de experimento, declaramos una completa ignorancia y lo mejor que podemos hacer es distribuir equitativamente la misma probabilidad para cada resultado.

- No observamos $P_i$
- Deducimos $P_i$ de nuestra razón y no necesitamos realizar ningún experimento para conocerla.

**Ejemplo (dado):**

¿Cuál es la probabilidad de que obtengamos 2 en el lanzamiento de un dado?

$P_2 = 1/6 = 0.166666$.

## 3.4 Frecuencias relativas

¿Qué sucede con los experimentos aleatorios cuyos posibles resultados **no** son igualmente probables?

¿Cómo podemos entonces definir las probabilidades de los resultados?

**Ejemplo (experimento aleatorio)**

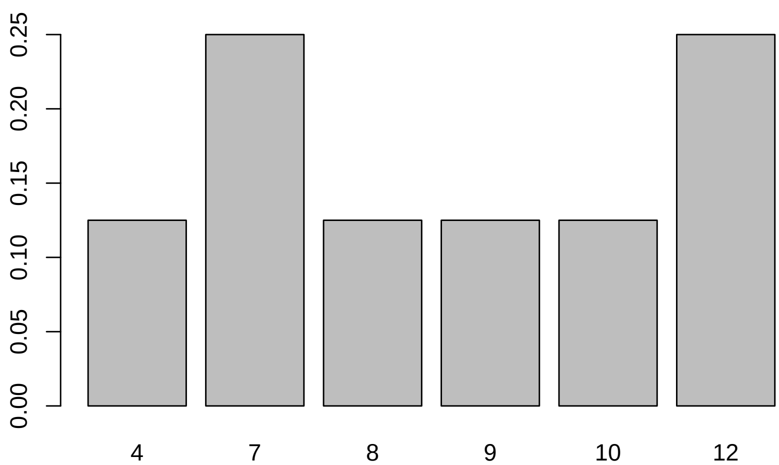
Imaginemos que repetimos un experimento aleatorio 8 veces y obtenemos las siguientes observaciones

8 4 12 7 10 7 9 12

- ¿Qué tan seguro estamos de obtener el resultado 12 en la siguiente observación?

La tabla de frecuencias es

```
##   outcome ni    fi
## 1       4  1 0.125
## 2       7  2 0.250
## 3       8  1 0.125
## 4       9  1 0.125
## 5      10  1 0.125
## 6      12  2 0.250
```




La **frecuencia relativa** $f_i = \frac{n_i}{N}$ parece una medida de probabilidad razonable porque

- es un número entre 0 y 1.
- mide la proporción del total de observaciones que observamos de un resultado particular.

Como $f_{12} = 0.25$ entonces estaríamos un cuarto seguros, una de cada 4 observaciones, de obtener 12.

**Pregunta**: ¿Qué tan bueno es $f_i$ como medida de certeza del resultado $i$?

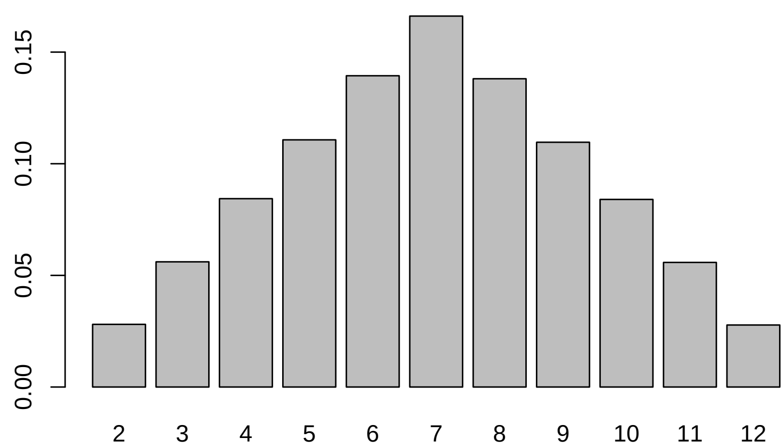
**Ejemplo (experimento aleatorio con mas repeticiones)**

Digamos que repetimos el experimento 100000 veces más:

La tabla de frecuencias es ahora

```
##    outcome    ni      fi
## 1        2  2807 0.02807
## 2        3  5607 0.05607
## 3        4  8435 0.08435
## 4        5 11070 0.11070
## 5        6 13940 0.13940
## 6        7 16613 0.16613
## 7        8 13806 0.13806
## 8        9 10962 0.10962
## 9       10  8402 0.08402
## 10      11  5581 0.05581
## 11      12  2777 0.02777
```

y el gráfico de barras es

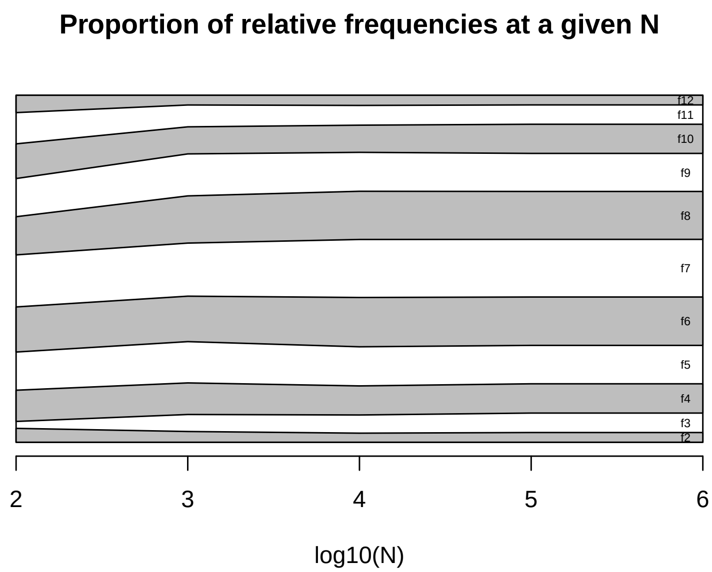



Aparecieron nuevos resultados y $f_{12}$ ahora es solo 0.027, y entonces estamos sólo un $\sim 3\%$ seguros de obtener 12 en el próximo experimento. Las probabilidades medidas por $f_i$ cambian con $N$.

## 3.5 Frecuencias relativas en el infinito

Una observación crucial es que si medimos las probabilidades de $f_i$ en valores crecientes de $N$ **¡convergen**!

En este gráfico cada sección vertical da la frecuencia relativa de cada observación. Vemos que después de $N = 1000$ ($log10(N) = 3$) las proporciones apenas varían con mas $N$.




Encontramos que cada una de las frecuencias relativas $f_i$ converge a un valor constante

$$lim_{N\to\infty} f_i = P_i$$

## 3.6 Probabilidad frecuentista

Llamamos **Probabilidad** $P_i$ al límite cuando $N \to \infty$ de la **frecuencia relativa** de observar el resultado $i$ en un experimento aleatorio.

Defendida por Venn (1876), la definición frecuentista de probabilidad se deriva de datos/experiencia (empírica).

- No observamos $P_i$, observamos $f_i$
- **Estimamos** $P_i$ con $f_i$ (normalmente cuando $N$ es grande), escribimos:

$$\hat{P}_i = f_i$$

Similar a la relación entre **observación** y **resultado**, tenemos la relación entre **frecuencia relativa** y **probabilidad** como un valor concreto de una cantidad abstracta.

## 3.7 Probabilidades clásicas y frecuentistas

Tenemos situaciones en las que se puede usar la probabilidad clásica para encontrar el límite de frecuencias relativas.

- Si los resultados son **igualmente probables**, la probabilidad clásica nos da el límite:

$$P_i = lim_{N\to\infty}\frac{n_i}{N} = \frac{1}{M}$$

- Si los resultados en los que estamos interesados pueden derivarse de otros resultados **igualmente probables**; Veremos más sobre esto cuando estudiemos los modelos de probabilidad.

**Ejemplo (suma de dos dados)**

Nuestro ejemplo anterior se basa en la **suma de dos dados**. Si bien realizamos el experimento muchas veces, anotamos los resultados y calculamos las **frecuencias relativas**, podemos conocer el valor exacto de probabilidad.

Esta probabilidad **se deduce** del hecho de que el resultado de cada dado es **igualmente probable**. A partir de esta suposición, podemos encontrar que (Ejercicio 1)

$$P_i = \begin{cases} \frac{i-1}{36}, & i \in \{2,3,4,5,6,7\} \\ \frac{13-i}{36}, & i \in \{8,9,10,11,12\} \end{cases}$$

La motivación de la definición frecuentista es **empírica** (datos) mientras que la de la definición clásica es **racional** (modelos). A menudo combinamos ambos enfoques (inferencia y deducción) para conocer las probabilidades de nuestro experimento aleatorio.

## 3.8 Definición de probabilidad

Una probabilidad es un número que se asigna a cada resultado posible de un experimento aleatorio y satisface las siguientes propiedades o **axiomas**:

1) cuando los resultados $E_1$ y $E_2$ son mutuamente excluyentes; es decir, solo uno de ellos puede ocurrir, entonces la probabilidad de observar $E_1$ **o** $E_1$, escrito como $E_1 \cup E_2$, es su suma:

$$P(E_1 \cup E_2) = P(E_1) + P(E_2)$$

2) cuando $S$ es el conjunto de todos los resultados posibles, entonces su probabilidad es 1 (al menos se observa algo):

$$P(S) = 1$$

3) La probabilidad de cualquier resultado está entre 0 y 1

$$P(E) \in [0, 1]$$

Propuesto por Kolmogorov's hace menos de 100 años (1933)

## 3.9 Tabla de probabilidades

Las propiedades de Kolmogorov son las reglas básicas para construir una **tabla de probabilidad**, de manera similar a la tabla de frecuencia relativa.

**Ejemplo (Dado)**

La tabla de probabilidad para el lanzamiento de un dado

| resultado | Probabilidad |
|---|---|
| 1 | 1/6 |
| 2 | 1/6 |
| 3 | 1/6 |
| 4 | 1/6 |
| 5 | 1/6 |
| 6 | 1/6 |
| $P(1 \cup 2 \cup ... \cup 6)$ | 1 |

Verifiquemos los axiomas:

1) Donde $1 \cup 2$ es, por ejemplo, el **evento** de lanzar un 1 **o** un 2. Entonces

$$P(1 \cup 2) = P(1) + P(2) = 2/6$$

2) Como $S = \{1, 2, 3, 4, 5, 6\}$ se compone de resultados **mutuamente excluyentes**, entonces

$$P(S) = P(1 \cup 2 \cup ... \cup 6) = P(1) + P(2) + ... + P(n) = 1$$

3) Las probabilidades de cada uno de resultados están entre 0 y 1.

## 3.10 Espacio muestral

El conjunto de todos los resultados posibles de un experimento aleatorio se denomina **espacio muestral** y se denota como $S$.

El espacio muestral puede estar formado por resultados categóricos o numéricos.

*Por ejemplo:*

- temperatura humana: $S = (36, 42)$ grados Celsius.

- niveles de azúcar en humanos: $S = (70 - 80) mg/dL$
- el tamaño de un tornillo de una línea de producción: $S = (70 - 72) mm$
- número de correos electrónicos recibidos en una hora: $S = \{1, ...\infty\}$
- el lanzamiento de un dado: $S = \{1, 2, 3, 4, 5, 6\}$

## 3.11 Eventos

Un **evento** $A$ es un **subconjunto** del espacio muestral. Es una **colección** de resultados.

*Ejemplos de eventos:*

- El evento de una temperatura saludable: $A = 37 - 38$ grados Celsius
- El evento de producir un tornillo con un tamaño: $A = 71.5mm$
- El evento de recibir más de 4 emails en una hora: $A = \{4, \infty\}$
- El evento de obtener un número menor o igual a 3 en la tirada de a dice: $A = \{1, 2, 3\}$

Un evento se refiere a un posible conjunto de **resultados**.

## 3.12 Álgebra de eventos

Para dos eventos $A$ y $B$, podemos construir los siguientes eventos derivados utilizando las operaciones básicas de conjuntos:

- Complemento $A'$: el evento de **no** $A$
- Unión $A \cup B$: el evento de $A$ **o** $B$
- Intersección $A \cap B$: el evento de $A$ **y** $B$

**Ejemplo (dado)**

Lancemos un dado y veamos los eventos (conjunto de resultados):

- un número menor o igual a tres $A : \{1, 2, 3\}$
- un número par $B : \{2, 4, 6\}$

Veamos como podemos construir nuevos eventos con las operaciones de conjuntos:

- un número no menor de tres: $A' : \{4, 5, 6\}$
- un número menor o igual a tres **o** par: $A \cup B : \{1, 2, 3, 4, 6\}$
- un número menor o igual a tres **y** par $A \cap B : \{2\}$

## 3.13 Resultados mutuamente excluyentes

Los resultados como tirar 1 y 2 en un dado son eventos que no pueden ocurrir al mismo tiempo. Decimos que son **mutuamente excluyentes**.

En general, dos eventos denotados como $E_1$ y $E_2$ son mutuamente excluyentes cuando

$$E_1 \cap E_2 = \emptyset$$

*Ejemplos:*

- El resultado de tener una gravedad de misofonía de 1 y una gravedad de 4.
- Los resultados de obtener 12 y 5 al sumar el lanzamiento de dos dados.

De acuerdo con las propiedades de Kolmogorov, solo los resultados **mutuamente excluyentes** se pueden organizar en **tablas de probabilidad**, como en las tablas de frecuencias relativas.

## 3.14 Probabilidades conjuntas

La **probabilidad conjunta** de $A$ y $B$ es la probabilidad de $A$ y $B$. Eso es

$$P(A \cap B)$$

o $P(A, B)$.

Para escribir probabilidades conjuntas de eventos no mutuamente excluyentes ($A \cap B \neq \emptyset$) en una tabla de probabilidad, notamos que siempre podemos descomponer el espacio muestral en conjuntos **mutuamente excluyentes** que involucran las intersecciones:

$S = \{A \cap B, A \cap B', A' \cap B, A' \cap B'\}$

**Consideremos el diagrama de Ven** para el ejemplo donde $A$ es el evento que corresponde a sacar número menor o igual que 3 y $B$ corresponde a un número par:

Las **marginales** de $A$ y $B$ son la probabilidad de $A$ y la probabilidad de $B$, respectivamente:

- $P(A) = P(A \cap B') + P(A \cap B) = 2/6 + 1/6 = 3/6$
- $P(B) = P(A' \cap B) + P(A \cap B) = 2/6 + 1/6 = 3/6$

Podemos ahora escribir la **tabla de probabilidad** para las probabilidades conjuntas

| Resultado | Probabilidad |
|---|---|
| $(A \cap B)$ | $P(A \cap B) = 1/6$ |
| $(A \cap B')$ | $P(A \cap B') = 2/6$ |
| $(A' \cap B)$ | $P(A' \cap B) = 2/6$ |
| $(A' \cap B')$ | $P(A' \cap B') = 1/6$ |

| Resultado | Probabilidad |
|---|---|
| suma | 1 |

Cada resultado tiene *dos* valores (uno para la característica del tipo $A$ y otro para el tipo $B$)

## 3.15 Tabla de contingencia

La tabla de probabilidad conjunta también se puede escribir en una **tabla de contingencia**

| | $B$ | $B'$ | suma |
|---|---|---|---|
| $A$ | $P(A \cap B)$ | $P(A \cap B')$ | $P(A)$ |
| $A'$ | $P(A' \cap B)$ | $P(A' \cap B')$ | $P(A')$ |
| suma | $P(B)$ | $P(B')$ | 1 |

Donde las marginales son las sumas en las márgenes de la tabla, por ejemplo:

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

En nuestro ejemplo, la tabla de contingencia es

| | $B$ | $B'$ | suma |
|---|---|---|---|
| $A$ | $1/6$ | $2/6$ | $3/6$ |
| $A'$ | $2/6$ | $1/6$ | $3/6$ |
| suma | $3/6$ | $3/6$ | 1 |

## 3.16 La regla de la suma:

La regla de la suma nos permite calcular la probabilidad de $A$ o $B$, $P(A \cup B)$, en términos de la probabilidad de $A$ y $B$, $P(A \cap B)$. Podemos hacer esto de tres maneras equivalentes:

1) Usando las marginales y la probabilidad conjunta

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

2) Usando solo probabilidades conjuntas

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

3) Usando el complemento de la probabilidad conjunta

$$P(A \cup B) = 1 - P(A' \cap B')$$

**Ejemplo (dado)**

Tomemos los eventos $A : \{1, 2, 3\}$, sacar un número menor o igual que 3, y $B : \{2, 4, 6\}$, sacar un número par en el lanzamiento de un dado.

Por lo tanto:

1) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 3/6 + 3/6 - 1/6 = 5/6$
2) $P(A \cup B) = P(A \cap B) + P(A \cap B') + P(A' \cap B) = 1/6 + 2/6 + 2/6 = 5/6$
3) $P(A \cup B) = 1 - P(A' \cap B') = 1 - 1/6 = 5/6$

En la tabla de contingencia $P(A \cup B)$ corresponde a las casillas en negrita (o sea todas menos 1/6 de abajo a la derecha)

| | $B$ | $B'$ |
|---|---|---|
| $A$ | **1/6** | **2/6** |
| $A'$ | **2/6** | 1/6 |

## 3.17 Preguntas

Recopilamos la edad y categoría de 100 deportistas en una competición

| | $edad : junior$ | $edad : senior$ |
|---|---|---|
| $categoria : 1ra$ | 14 | 12 |
| $categoria : 2a$ | 21 | 18 |
| $categoria : 3a$ | 22 | 13 |

**1)** ¿Cuál es la probabilidad estimada de que un deportista sea de 2ª categoría y senior?

**a:** 18/100; **b:** 18/43; **c:** 18; **d:** 18/39

**2)** ¿Cuál es la probabilidad estimada de que el atleta no esté en la tercera categoría y sea senior?

**a:** 35/100; **b:** 30/100; **c:** 22/100; **d:** 13/100

**3)** ¿Cuál es la probabilidad marginal de la tercera categoría?

**a:** 13/100; **b:** 35/100; **c:** 22/100; **d:** 13/22

**4)** ¿Cuál es la probabilidad marginal de ser senior?

**a:** 13/100; **b:** 43/100; **c:** 43/57; **d:** 57/100

**5)** ¿Cuál es la probabilidad de ser senior o de tercera categoría?

**a:** 65/100; **b:** 86/100; **c:** 78/100; **d:** 13/100

## 3.18 Ejercicios

#### 3.18.0.1 Probabilidad clásica: Ejercicio 1

- Escribe la tabla de **probabilidad conjunta** para los **resultados** de lanzar dos dados; en las filas escribe los resultados del primer dado y en las columnas los resultados del segundo dado.
- ¿Cuál es la probabilidad de sacar $(3,4)$? (R:1/36)
- ¿Cuál es la probabilidad de tirar 3 y 4 con cualquiera de los dos dados? (R:2/36)
- ¿Cuál es la probabilidad de tirar 3 en el primer dado o 4 en el segundo? (A:11/36)
- ¿Cuál es la probabilidad de tirar 3 o 4 con cualquier dado? (R:20/36)
- Escribe la **tabla de probabilidad** para el resultado de la **suma** de dos dados. Supon que el resultado de cada dado es **igualmente probable**. Verifica que es:

$$P_i = \begin{cases} \frac{i-1}{36}, & i \in \{2,3,4,5,6,7\} \\ \frac{13-i}{36}, & i \in \{8,9,10,11,12\} \end{cases}$$

#### 3.18.0.2 Probabilidad frecuentista: Ejercicio 2

El resultado de un experimento aleatorio es medir la gravedad de la misofonía **y** el estado de depresión de un paciente.

- Gravedad de la misofonía: $S_M : \{M_0, M_1, M_2, M_3, M_4\}$
- Depresión: $S_D : \{D', D\})$

Escribe la tabla de contingencia para las frecuencias absolutas ($n_{M,D}$) para un estudio sobre un total de 123 pacientes en el que se observó

- 100 individuos no tuvieron depresión.
- Ningún individuo con misofonía 4 y sin depresión.
- 5 individuos con misofonía de grado 1 y sin depresión.
- El mismo número que el caso anterior para individuos con depresión y sin misofonía.
- 25 individuos sin depresión y grado 3 de misofonía.
- El número de misofónicos sin depresión para los grados 2 y 0 se repartieron a cantiaddes iguales.

- El número de individuos con depresión y misofonía incrementó progresivamente en múltiplos de tres, empezando en 0 individuos para grado 1.

Reponde las siguientes preguntas:

- ¿Cuantos individuos tuvieron misofonía? (R:83)
- ¿Cuantos individuos tuvieron misofonía de grado 3? (R:31)
- ¿Cuantos individuos tuvieron misofonía de grado 2 sin depresión? (R:35)

Escribe las tabla de consingencia para frecuencias relativas $f_{M,D}$. Supongamos que $N$ es grande y que las frecuencias absolutas **estiman** las probabilidades $f_{M,D} = \hat{P}(M \cap D)$. Responde las siguientes preguntas:

- ¿Cuál es la probabilidad marginal de misofonía de gravedad 2? (R: 0.3)
- ¿Cuál es la probabilidad de no ser misofónico **y** no estar deprimido? (R:0.284)
- ¿Cuál es la probabilidad de ser misofónico **o** estar deprimido? (R: 0.715)
- ¿Cuál es la probabilidad de ser misofónico **y** estar deprimido? (R: 0.146)
- Describir en lenguaje hablado los resultados con probabilidad 0.

#### 3.18.0.3 Ejercicio 3

Hemos realizado un experimento aleatorio 10 veces, que consiste en anotar el sexo y el estado vital de pacientes con algún tipo de cáncer después de 10 años del diagnóstico. Obtuvimos los siguientes resultados

```
##         A     B
## 1    male  dead
## 2    male  dead
## 3    male  dead
## 4  female alive
## 5    male  dead
## 6  female alive
## 7  female  dead
## 8  female alive
## 9    male alive
## 10   male alive
```

- Crea la tabla de contingencia para el número ($n_{i,j}$) de observaciones de cada resultado $(A, B)$
- Crea la tabla de contingencia para la frecuencia relativa ($f_{i,j}$) de los resultados
- ¿Cuál es la frecuencia marginal de ser hombre? (R/0.6)
- ¿Cuál es la frecuencia marginal de estar vivo? (R/0.5)
- ¿Cuál es la frecuencia de estar vivo **o** ser mujer? (R/0.6)

#### 3.18.0.4 Teoría: Ejercicio 4

- De la segunda forma de la regla de la suma, obtener la primera y la tercera forma.
- ¿Cuál es la regla de la suma de la tercera forma para la probabilidad de tres eventos $P(A \cup B \cup C)$?

# Chapter 4

# Probabilidad condicional

En este capítulo, introduciremos la probabilidad condicional.

Usaremos la probabilidad condicional para definir la independencia estadística.

Discutiremos el teorema de Bayes y discutiremos una de sus principales aplicaciones, que es la eficacia de predicción de una herramienta de diagnóstico.

## 4.1 Probabilidad conjunta

Recordemos que la probabilidad conjunta de dos eventos $A$ y $B$ se define como su intersección

$$P(A, B) = P(A \cap B)$$

Ahora, imagina experimentos aleatorios que miden dos tipos diferentes de resultados.

- altura y peso de un individuo: $(h, w)$
- tiempo y posición de una carga eléctrica: $(p, t)$
- el lanzamiento de dos dados: $(n_1,n_2)$
- cruzar dos semáforos en verde: $(\bar{R}_1\ , \bar{R}_2)$

A menudo nos interesa saber si los valores de un resultado **condicionan** los valores del otro.

## 4.2 Independencia estadística

En muchos casos, estamos interesados en saber si dos eventos a menudo tienden a ocurrir juntos. Queremos poder discernir entre dos casos.

- **Independencia** entre eventos. Por ejemplo, sacar un 1 en un dado no hace más probable sacar otro 1 en un segundo dado.
- **Correlación** entre eventos. Por ejemplo, si un hombre es alto, probablemente sea pesado.

**Ejemplo (conductor)**

Realizamos un experimento para averiguar si observar fallas estructurales en un material afecta su conductividad.

Los datos se verían como

| Director de orquesta | Estructura | Conductividad |
|---|---|---|
| $c_1$ | con fallas | defectuosa |
| $c_2$ | sin fallas | sin defectos |
| $c_3$ | con fallas | defectuosa |
| ... | ... | ... |
| $c_i$ | sin fallas | defectuosa* |
| ... | ... | ... |
| ... | ... | ... |
| $c_n$ | con fallas | sin defectos* |

Podemos esperar que la conductividad defectuosa ocurra más a menudo con fallas que sin fallas si las fallas afectan la conductividad.

Imaginemos que a partir de los datos obtenemos la siguiente tabla de contingencia de **probabilidades conjuntas** estimadas

| | con fallas (F) | sin fallas (F') | suma |
|---|---|---|---|
| defectuoso (D) | 0.005 | 0.045 | 0.05 |
| sin defectos (D') | 0.095 | 0.855 | 0.95 |
| suma | 0.1 | 0.9 | 1 |

donde, por ejemplo, la probabilidad conjunta de $F$ y $D$ es

- $P(D, F) = 0.005$

y las probabilidades marginales son

- $P(D) = P(D, F) + P(D, F') = 0.05$
- $P(F) = P(D, F) + P(D', F) = 0.1.$

## 4.3 La probabilidad condicional

La conductividad defectuosa es **independiente** de tener fallas estructurales en el material si la probabilidad de tener conductividad defectuosa ($D$) es la misma **ya sea** que tenga fallas ($F$) o no ($F'$) .

Consideremos primero solamente los materiales que tienen fallas.

Dentro de aquellos materiales que tienen fallas ($F$), ¿cuál es la probabilidad estimada de que sean defectuosos?

$\hat{P}(D|F) = \frac{n_{F,D}}{n_F} = \frac{n_{F,D}/n}{n_F/n} = \frac{f_{F,D}}{f_F}$

$$= \frac{\hat{P}(F,D)}{\hat{P}(F)}$$

Por lo tanto, en el límite cuando $N \to \infty$, tenemos

$$P(D|F) = \frac{P(F,D)}{P(D)} = \frac{P(F \cap D)}{P(D)}$$

**Definición:**

La *probabilidad condicional* de un evento $B$ dado un evento $A$, indicado como $P(A|B)$, es

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

Podemos probar que la probabilidad condicional satisface los axiomas de probabilidad. La probabilidad condicional se puede entender como una probabilidad con un espacio muestral dado por $B$: $S_B$. En nuestro ejemplo, los materiales con fallas.

## 4.4 Tabla de contingencia condicional

Si dividimos las columnas de la tabla de probabilidad conjunta por las probabilidades marginales de los efectos condicionantes ($F$ y $F'$), podemos escribir **una tabla de contingencia condicional**

| | F | F' |
|---|---|---|
| D | P(D \| F) | P(D \| F') |
| D' | P(D' \| F) | P(D' \| F') |
| suma | 1 | 1 |

Donde las probabilidades por columnas suman uno. La primera columna muestra las probabilidades de ser defectuoso o no solo de que los materiales que tienen fallas (primera condición: $F$). La segunda columna muestra las probabilidades solo para los materiales que no tienen fallas (segunda condición: $F'$).

Las probabilidades condicionales son las probabilidades del evento dentro de cada condición. Las leemos como:

- $P(D|F)$: Probabilidad de tener conductividad defectuosa **si** tiene fallas
- $P(D'|F)$: Probabilidad de no tener conductividad defectuosa **si** tiene fallas
- $P(D|F')$: Probabilidad de tener conductividad defectuosa **si** no tiene fallas
- $P(D'|F')$ Probabilidad de no tener conductividad defectuosa **si** no tiene fallas

## 4.5 Independencia estadística

En nuestro ejemplo, la tabla de contingencia condicional es

| | F | F' |
|---|---|---|
| D | P(D\|F) = 0.05 | P(D\|F')=0.05 |
| D' | P(D'\|F)=0.95 | P(D'\|F')=0.95 |
| suma | 1 | 1 |

¡Observamos que las probabilidades marginales y condicionales son las mismas!

- $P(D|F) = P(D|F') = P(D)$
- $P(D'|F) = P(D'|F') = P(D')$

Esto quiere decir que la probabilidad de observar un conductor defectuoso **no** depende tener una falla estructural o no.

Concluimos que la conductividad defectuosa no se ve afectada por tener una falla estructural.

**Definición**

Dos eventos $A$ y $B$ son estadísticamente independientes si ocurre cualquiera de los casos equivalentes

1) $P(A|B) = P(A)$; $A$ es independiente de $B$
2) $P(B|A) = P(B)$; $B$ es independiente de $A$

y por la definición de probabilidad condicional

3) $P(A \cap B) = P(A|B)P(B) = P(A)P(B)$

Esta tercera forma es un enunciado sobre las probabilidades conjuntas. Dice que podemos obtener probabilidades conjuntas por la multiplicación de las marginales.

En nuestra tabla de probabilidad conjunta original

| | F | F' | suma |
|---|---|---|---|
| D | 0.005 | 0.045 | 0.05 |
| D' | 0.095 | 0.855 | 0.95 |
| suma | 0.1 | 0.9 | 1 |

podemos confirmar que todas las entradas de la matriz son el producto de las marginales. Por ejemplo: $P(F)P(D) = P(D \cap F)$ y $P(D')P(F') = P(D' \cap F')$. Por lo tanto, ser defectuoso es independiente de tener un defecto.

**Ejemplo (Monedas)**

Queremos confirmar que los resultados de lanzar dos monedas son independientes. Consideramos que todos los resultados son igualmente probables:

| resultado | Probabilidad |
|---|---|
| $(H, T)$ | 1/4 |
| $(H, H)$ | 1/4 |
| $(T, T)$ | 1/4 |
| $(T, H)$ | 1/4 |
| suma | 1 |

donde $(H, T)$ es, por ejemplo, el evento de cara en la primera moneda y cruz en la segunda moneda. La tabla de contingencia para las probabilidades conjuntas es:

| | H | T | suma |
|---|---|---|---|
| H | 1/4 | 1/4 | 1/2 |
| T | 1/4 | 1/4 | 1/2 |
| suma | 1/2 | 1/2 | 1 |

De esta tabla vemos que la probabilidad de obtener una cara y luego una cruz es el producto de las marginales $P(H, T) = P(H) * P(T) = 1/4$. Por lo tanto, el evento de cara en la primera moneda y cruz en la segunda son independientes.

Si elaboramos la tabla de contingencia condicional sobre el lanzamiento de la primera moneda veremos que obtener cruz en la segunda moneda no está condicionado por haber obtenido cara en la primera moneda: $P(T|H) = P(T) = 1/2$

## 4.6 Dependencia estadística

Un ejemplo importante de dependencia estadística se encuentra en el desempeño de **herramientas de diagnóstico**, donde queremos determinar el estado de un sistema (s) con resultados

- inadecuado (si)
- adecuado (no)

con una prueba (t) con resultados

- positivo
- negativo

Por ejemplo, probamos una batería para saber cuánto tiempo puede durar. Tensamos un cable para saber si resiste llevar cierta carga. Realizamos una PCR para ver si alguien está infectado.

## 4.7 Prueba de diagnóstico

Consideremos diagnosticar una infección con una nueva prueba. Estado de infección:

- si (infectado)
- no (no infectado)

Prueba:

- positivo
- negativo

La **tabla de contingencia condicional** es lo que obtenemos en un ambiente controlado (laboratorio)

| | Infección: si | Infección: No |
|---|---|---|
| Test: positivo | P(positivo \| si) | P(positivo \| no) |
| Test: negativo | P(negativo \| si) | P(negativo \| no) |
| suma | 1 | 1 |

Miremos las entradas de la tabla 1) Tasa de verdaderos positivos (Sensibilidad): La probabilidad de dar positivo **si** tiene la enfermedad $P(positivo|si)$

2) Tasa de verdaderos negativos (Especificidad): La probabilidad de dar negativo **si** no tiene la enfermedad $P(negativo|no)$

3) Tasa de falsos positivos: la probabilidad de dar positivo **si** no tiene la enfermedad $P(positivo|no)$

4) Tasa de falsos negativos: la probabilidad de dar negativo **si** tiene la enfermedad $P(negativo|si)$

Alta correlación (dependencia estadística) entre la prueba y la infección significa valores altos de las probabilidades 1 y 2 **y** valores bajos para las probabilidades 3 y 4.

**Ejemplo (COVID)**

Ahora consideremos una situación real. En los días iniciales de la pandemia de coronavirus no había una medida de la eficacia de las PCR para detectar el virus. Uno de los primeros estudios publicados (https://www.nejm.org/doi/full/10.1056/NEJMp2015897) encontró que

- Las PCR tuvieron una sensibilidad del 70%, en condición de infección.
- Las PCR tuvieron una especificidad del 94%, en condición de no infección.

La tabla de contingencia condicional es

| | Infección: si | Infección: No |
|---|---|---|
| Test: positivo | P(positivo\|si)=0.7 | P(positivo\|no)=0.06 |
| Test: negativo | P(negativo\|si)=0.3 | P(negativo\|no)=0.94 |
| suma | 1 | 1 |

Por lo tanto, los errores en las pruebas de diagnóstico fueron:

- La tasa de falsos positivos es $P(positivo|no) = 0.06$
- La tasa de falsos negativos es $P(negativo|si) = 0.3$

## 4.8 Probabilidades inversas

Nos interesa encontrar la probabilidad de estar infectado si la prueba da positivo:

$$P(si|positivo)$$

Para eso:

1. Recuperamos la tabla de contingencia para probabilidades conjuntas, multiplicando por las marginales

| | Infección: si | Infección: No | suma |
|---|---|---|---|
| Test: positivo | P(positivo \| si)P(si) | P(positivo \| no)P(no) | P(positivo) |
| Test: negativo | P(negativo \| si)P(si) | P(negativo \| no) P(no) | P(negativo) |
| suma | P(si) | P(no) | 1 |

2. Usamos la definición de probabilidades condicionales para filas en lugar de columnas (dividimos por la marginal de los resultados de la prueba)

| | Infección: si | Infección: No | suma |
|---|---|---|---|
| Test: positivo | P(si\|positivo) | P(sin\|positivo) | 1 |
| Test: negativo | P(si\|negativo) | P(sin\|negativo) | 1 |

Por ejemplo:

$$P(si|positivo) = \frac{P(positivo|si)P(si)}{P(positivo)}$$

Para aplicar esta fórmula necesitamos las marginales $P(si)$ (incidencia) y $P(positivo)$.

- Para encontrar $P(si)$, necesitamos un nuevo estudio: el primer estudio de prevalencia en España mostró que durante el confinamiento $P(si) = 0.05$, $P(no) = 0.95$, antes del verano de 2020.
- Para encontrar $P(positivo)$, podemos usar la definición de probabilidad marginal y condicional:

$$P(positivo) = P(positivo \cap si) + P(positivo \cap no)$$

$$= P(positivo|si)P(si) + P(positivo|no)P(no)$$

Esta última relación de las marginales se llama **regla de probabilidad total**.

## 4.9 Teorema de Bayes

Después de sustituir la regla de probabilidad total en $P(si|positivo)$, tenemos

$$P(si|positivo) = \frac{P(positivo|si)P(si)}{P(positivo|si)P(si) + P(positivo|no)P(no)}$$

Esta expresión se conoce como **teorema de Bayes**. Nos permite invertir los condicionales:

$$P(positivo|si) \to P(si|positivo)$$

O **evaluar** una prueba en una condición controlada (infección) y luego usarla para **inferir** la probabilidad de la condición cuando la prueba es positiva.

**Ejemplo (COVID)**:

El rendimiento de la prueba fue:

- Sensibilidad: $P(positivo|si) = 0.70$
- Tasa de falsos positivos: $P(positivo|no) = 1 - P(negativo|no) = 0.06$

El estudio en población española dio:

- $P(si) = 0.05$
- $P(no) = 1 - P(si) = 0.95$.

Por lo tanto, la probabilidad de estar infectado en caso de dar positivo era:

$$P(si|positivo) = 0.38$$

Concluimos que en ese momento las PCR no eran muy buenas para **confirmar** infecciones.

Sin embargo, apliquemos ahora el teorema de Bayes a la probabilidad de no estar infectado si la prueba fue negativa.

$$P(no|negativo) = \frac{P(negativo|no)P(no)}{P(negativo|no)P(no) + P(negativo|si)P(si)}$$

La sustitución de todos los valores da

$$P(no|negativo) = 0.98$$

Por lo tanto, las pruebas eran buenas para **descartar** infecciones y un requisito justo para viajar.

**Teorema de Bayes**

En general, podemos tener más de dos eventos condicionantes. Por lo tanto, el teorema de Baye dice:

Si $E1, E2, ..., Ek$ son $k$ eventos mutuamente excluyentes y exhaustivos y $B$ es cualquier evento, entonces la probabilidad inversa $P(Ei|B)$ es

$$P(Ei|B) = \frac{P(B|Ei)P(Ei)}{P(B|E1)P(E1) + ... + P(B|Ek)P(Ek)}$$

El denominador es la regla de probabilidad total para la marginal $P(B)$, en términos de las marginales $P(E1), P(E2), ...P(Ek)$.

$$P(B) = P(B|E1)P(E1) + ... + P(B|Ek)P(Ek)$$

**Árbol condicional**

La regla de probabilidad total también se puede ilustrar usando un árbol **condicional**.

**Regla de probabilidad total** para la marginal de $B$: ¿De cuántas maneras puedo obtener el resultado $B$?

$P(B) = P(B|A)P(A) + P(B|A')P(A')$

## 4.10 Ejercicios

## 4.11 Preguntas

Recopilamos la edad y categoría de 100 deportistas en una competición

| | *junior* | *senior* |
|---|---|---|
| 1*er* | 14 | 12 |
| 2*do* | 21 | 18 |
| 3*er* | 22 | 13 |

**1)** ¿Cuál es la probabilidad estimada de que el atleta esté en la tercera categoría si el atleta es junior?

**a:** 22; **b:** 22/100; **c:** 22/57; **d:** 22/35;

**2)** ¿Cuál es la probabilidad estimada de que el atleta sea junior y esté en 1ra categoría si el atleta no está en 3ra categoría?

**a:** 14/35; **b:** 14/65; **c:** 14/100; **d:** 14/26

**3)** Una prueba diagnóstica tiene una probabilidad de 8/9 de detectar una enfermedad si los pacientes están enfermos y una probabilidad de 3/9 de detectar la enfermedad si los pacientes están sanos. Si la probabilidad de estar enfermo es 1/9. ¿Cuál es la probabilidad de que un paciente esté enfermo si una prueba detecta la enfermedad?

**a:** $\frac{8/9}{8/9+3/9} * 1/9$; **b:** $\frac{3/9}{8/9+3/9} * 1/9$; **c:** $\frac{3/9*8/9}{8/9*1/9+3/9*8/9}$; **d:** $\frac{8/9*1/9}{8/9*1/9+3/9*8/9}$;

**4)** Como se comenta en las notas, una prueba PCR para coronavirus tenía una sensibilidad del 70 % y una especificidad del 94 % y en España durante el confinamiento hubo una incidencia del 5 %. Con estos datos, ¿cuál era la probabilidad de dar positivo en España ($P(positivo)$)

**a:** 0.035; **b:** 0.092; **c:** 0.908; **d:** 0.95

**5)** Con los mismos datos que en la pregunta 4, dar positivo en la PCR y estar infectado no son eventos independientes porque:

**a:** La sensibilidad es del 70%; **b:** La sensibilidad y la tasa de falsos positivos son diferentes; **c:** La tasa de falsos positivos es del 0.06%; **d:** la especificidad es del 96%

#### 4.11.0.1 Ejercicio 1

Se prueba el rendimiento de una máquina para producir varillas de torneado de alta calidad. Estos son los resultados de las pruebas

| | Redondeado: si | Redondeado: No |
|---|---|---|
| superficie lisa: si | 200 | 1 |
| superficie lisa: no | 4 | 2 |

- ¿Cuál es la probabilidad estimada de que la máquina produzca una varilla que no satisfaga ningún control de calidad? (R: 2/207)
- ¿Cuál es la probabilidad estimada de que la máquina produzca una varilla que no satisfaga al menos un control de calidad? (R: 7/207)
- ¿Cuál es la probabilidad estimada de que la máquina produzca varillas de superficie redondeada y alisada? (R: 200/207)
- ¿Cuál es la probabilidad estimada de que la barra sea redondeada si la barra es lisa? (R: 200/201)
- ¿Cuál es la probabilidad estimada de que la varilla sea lisa si es redondeada? (R: 200/204)
- ¿Cuál es la probabilidad estimada de que la varilla no sea ni lisa ni redondeada si no satisface al menos un control de calidad? (R: 2/7)
- ¿Son eventos independientes la lisa y la redondez? (No)

#### 4.11.0.2 Ejercicio 2

Desarrollamos un test para detectar la presencia de bacterias en un lago. Encontramos que si el lago contiene la bacteria, la prueba es positiva el 70% de las veces. Si no hay bacterias, la prueba es negativa el 60% de las veces. Implementamos la prueba en una región donde sabemos que el 20% de los lagos tienen bacterias.

- ¿Cuál es la probabilidad de que un lago que dé positivo esté contaminado con bacterias? (R: 0.30)

#### 4.11.0.3 Ejercicio 3

Se prueba el rendimiento de dos máquinas para producir varillas de torneado de alta calidad. Estos son los resultados de las pruebas

**Máquina 1**

| | Redondeado: si | Redondeado: No |
|---|---|---|
| superficie lisa: si | 200 | 1 |
| superficie lisa: no | 4 | 2 |

**Máquina 2**

| | Redondeado: si | Redondeado: No |
|---|---|---|
| superficie lisa: si | 145 | 4 |
| superficie lisa: no | 8 | 6 |

- ¿Cuál es la probabilidad de que la barra sea redondeada? (R: 357/370)
- ¿Cuál es la probabilidad de que la varilla haya sido producida por la máquina 1? (R: 207/370)
- ¿Cuál es la probabilidad de que la varilla no sea lisa? (R: 20/370)
- ¿Cuál es la probabilidad de que la varilla sea lisa o redondeada o producida por la máquina 1? (R: 364/370)
- ¿Cuál es la probabilidad de que la varilla quede redondeada si es alisada y de la máquina 1? (R: 200/201)
- ¿Cuál es la probabilidad de que la varilla no esté redondeada si no está alisada y es de la máquina 2? (R: 6/8)
- ¿Cuál es la probabilidad de que la varilla haya salido de la máquina 1 si está alisada y redondeada? (R: 200/345)
- ¿Cuál es la probabilidad de que la varilla haya venido de la máquina 2 si no pasa al menos uno de los controles de calidad? (R:0.72)

#### 4.11.0.4 Ejercicio 4

Queremos cruzar una avenida con dos semáforos. La probabilidad de encontrar el primer semáforo en rojo es 0.6. Si paramos en el primer semáforo, la probabilidad de parar en el segundo es 0.15. Mientras que la probabilidad de detenernos en el segundo si no nos detenemos en el primero es 0.25.

Cuando intentamos cruzar ambos semáforos:

- ¿Cuál es la probabilidad de tener que detenerse en cada semáforo? (R:0.09)
- ¿Cuál es la probabilidad de tener que parar en al menos un semáforo? (R:0.7)
- ¿Cuál es la probabilidad de tener que detenerse en un solo semáforo? (R:0.61)
- Si paré en el segundo semáforo, ¿cuál es la probabilidad de que tuviera que parar en el primero? (R: 0.47)
- Si tuviera que parar en cualquier semáforo, ¿cuál es la probabilidad de que tuviera que hacerlo dos veces? (R: 0.12)
- ¿Parar en el primer semáforo es un evento independiente de detenerse en el segundo semáforo? (No)

Ahora, queremos cruzar una avenida con tres semáforos. La probabilidad de encontrar un semáforo en rojo solo depende de la anterior. En concreto, la probabilidad de encontrar un semáforo en rojo dado que el anterior estaba en rojo es de 0.15. Mientras que la probabilidad de encontrar un tráfico justo en rojo dado que el anterior estaba en verde es de 0.25. Además, la probabilidad de encontrar el primer semáforo en rojo es de 0.6.

- ¿Cuál es la probabilidad de tener que parar en cada semáforo? (R:0.013)
- ¿Cuál es la probabilidad de tener que parar en al menos un semáforo? (R:0.775)
- ¿Cuál es la probabilidad de tener que detenerse en un solo semáforo? (R:0.5425)

consejos:

- Si la probabilidad de que un semáforo esté en rojo depende únicamente del anterior, entonces $P(R_3|R_2, R_1) = P(R_3|R_2, \bar{R}_1) = P(R_3|R_2)$ y $P(R_3|\bar{R}_2, R_1) = P(R_3|\bar{R}_2, \bar{R}_1) = P(R_3|\bar{R}_2)$
- La probabilidad conjunta de encontrar tres semáforos en rojo se puede escribir como: $P(R_1, R_2, R_3) = P(R_3|R_2)P(R_2|R_1)P(R_1)$

#### 4.11.0.5 Ejercicio 5

Una prueba de calidad en un ladrillo aleatorio se define por los eventos:

- Pasar la prueba de calidad: $E$, no pasar la prueba de calidad: $\bar{E}$
- Defectuoso: $D$, no defectuoso: $\bar{D}$

Si la prueba diagnóstica tiene sensibilidad $P(E|\bar{D}) = 0.99$ y especificidad $P(\bar{E}|D) = 0.98$, y la probabilidad de pasar la prueba es $P(E) = 0.893$ entonces

- ¿Cuál es la probabilidad de que un ladrillo elegido al azar sea defectuoso $P(D)$? (R:0.1)
- ¿Cuál es la probabilidad de que un ladrillo que ha pasado la prueba sea realmente defectuoso? (R:0.022)
- La probabilidad de que un ladrillo no sea defectuoso **y** que no pase la prueba (R:0.009)
- ¿Son $D$ y $\bar{E}$ estadísticamente independientes? (No)

# Chapter 5

# Variables aleatorias discretas

## 5.1 Objetivo

En este capítulo definiremos las variables aleatorias y estudiaremos variables aleatorias **discretas**.

Definiremos la función de masa de probabilidad y sus principales propiedades de media y varianza. Siguiendo el proceso de abstracción de las frecuencias relativas en probabilidades, también definimos la distribución de probabilidad como el caso límite de la frecuencia relativa acumulada.

## 5.2 Frecuencias relativas

Las frecuencias relativas de los resultados de un experimento aleatorio son una medida de su propensión. Podemos usarlos como estimadores de sus probabilidades, cuando repetimos el experimento aleatorio muchas veces ($n \to \infty$).

Definimos tendencia central (promedio), dispersión (varianza muestral) y la distribución de frecuencias de los datos ($F_i$).

En términos de probabilidades, ¿cómo se definen estas cantidades?

## 5.3 Variable aleatoria

Definimos las frecuencias relativas sobre las **observaciones** de los experimentos. Ahora definimos las cantidades equivalentes para las probabilidades en términos de los **resultados** de los experimentos. Nos ocuparemos únicamente de resultados de tipo numérico.

Una **variable aleatoria** es un símbolo que representa un **resultado numérico** de un experimento aleatorio. Escribimos la variable aleatoria en **mayúsculas** (es decir, $X$).

**Definición:**

Una **variable aleatoria** es una función que asigna un **número** real a un **evento** del espacio muestral de un experimento aleatorio.

Recuerda que un evento puede ser un resultado o una colección de resultados.

Cuando la variable aleatoria toma un **valor**, indica la realización de un **evento** de un experimento aleatorio.

*Ejemplo:*

Si $X \in \{0, 1\}$, entonces decimos que $X$ es una variable aleatoria que puede tomar los valores 0 o 1.

## 5.4 Eventos de observar una variable aleatoria

Hacemos la distinción entre variables en el espacio modelo con letras mayúsculas, como entidades abstractas, y la realización de un evento o resultado particular. Por ejemplo:

- $X = 1$ es el **evento** de observar la variable aleatoria $X$ con valor 1
- $X = 2$ es el **evento** de observar la variable aleatoria $X$ con valor 2

...

**En general:**

- $X = x$ es el **evento** de observar la variable aleatoria $X$ ($X$ mayúscula) con valor $x$ ($x$ pequeño).

## 5.5 Probabilidad de variables aleatorias

Nos interesa asignar probabilidades a los eventos de observar un valor particular de una variable aleatoria.

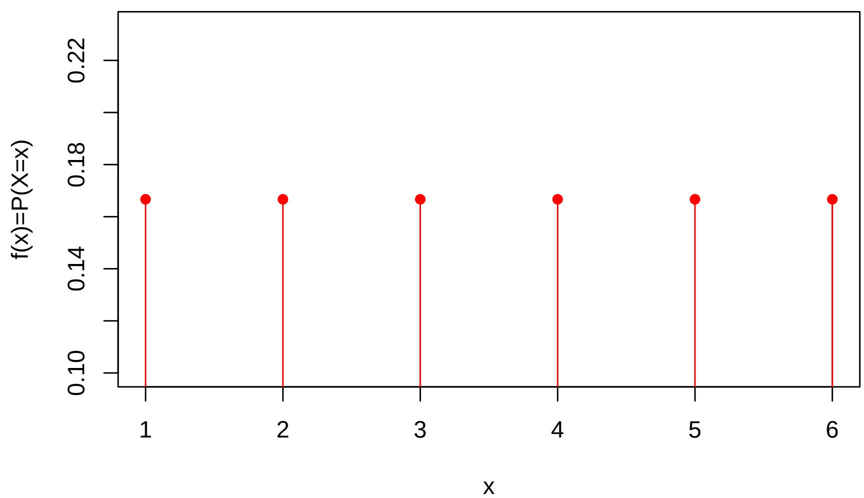
Por ejemplo, para los dados escribiremos la tabla de probabilidad como

| $X$ | Probabilidad |
|---|---|
| 1 | $P(X = 1) = 1/6$ |
| 2 | $P(X = 2) = 1/6$ |
| 3 | $P(X = 3) = 1/6$ |
| 4 | $P(X = 4) = 1/6$ |
| 5 | $P(X = 5) = 1/6$ |
| 6 | $P(X = 6) = 1/6$ |

donde hacemos explícitos los eventos de que la variable toma un resultado dado $X = x$.

## 5.6 Funciones de probabilidad

Debido a que $x$ (minúscula) es una variable numérica, las probabilidades de la variable aleatoria se pueden dibujar




o escrir como una función

$$f(x) = P(X = x) = 1/6$$

## 5.7 Funciones de probabilidad

Podemos **crear** cualquier tipo de función de probabilidad si satisfacemos las reglas de probabilidad de Kolmogorov:

Para una variable aleatoria discreta $X \in \{x_1, x_2, .., x_M\}$, una **función de masa de probabilidad** que se usa para calcular probabilidades

- $f(x_i) = P(X = x_i)$

siempre es positiva

- $f(x_i) \geq 0$

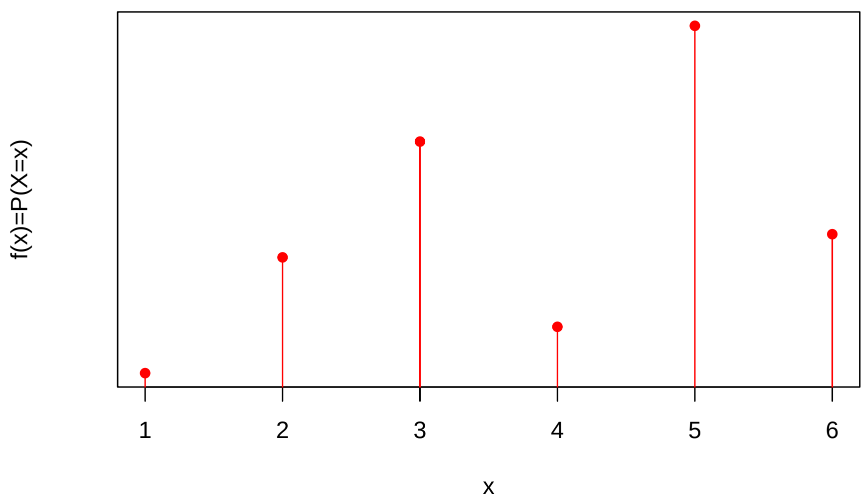
y su suma sobre todos los valores de la variable es 1:

- $\sum_{i=1}^{M} f(x_i) = 1$

Donde $M$ es el número de resultados posibles.

Ten en cuenta que la definición de $X$ y su función de masa de probabilidad es general **sin referencia** a ningún experimento. Las funciones viven en el espacio modelo (abstracto).

Aquí tenemos un ejemplo




$X$ y $f(x)$ son objetos abstractos que pueden corresponder o no a un experimento. Tenemos la libertad de construirlos como queramos siempre que respetemos su definición.

Las funciones de masa de probabilidad tienen algunas **propiedades** que se derivan exclusivamente de su definición.

## 5.8 Probabilidades y frecuencias relativas

### Considera el ejemplo

Haz el siguiente experimento: En una urna pon 8 bolas y:

- marca 1 bola con el número $-2$
- marca 2 bolas con el número $-1$
- marca 2 bolas con el número $0$

- marca 2 bolas con el número 1
- marca 1 bolas con el número 2

Y considere realizar el siguiente **experimento aleatorio:** Tome una bola y lea el número.

A partir de la probabilidad clásica, podemos escribir la tabla de probabilidades, para lo cual no necesitamos realizar ningún experimento

| $X$ | $P(X=x)$ |
|---|---|
| $-2$ | $1/8 = 0.125$ |
| $-1$ | $2/8 = 0.25$ |
| 0 | $2/8 = 0.25$ |
| 1 | $2/8 = 0.25$ |
| 2 | $1/8 = 0.125$ |

Ahora, realicemos el experimento 30 veces y escribamos la tabla de frecuencia

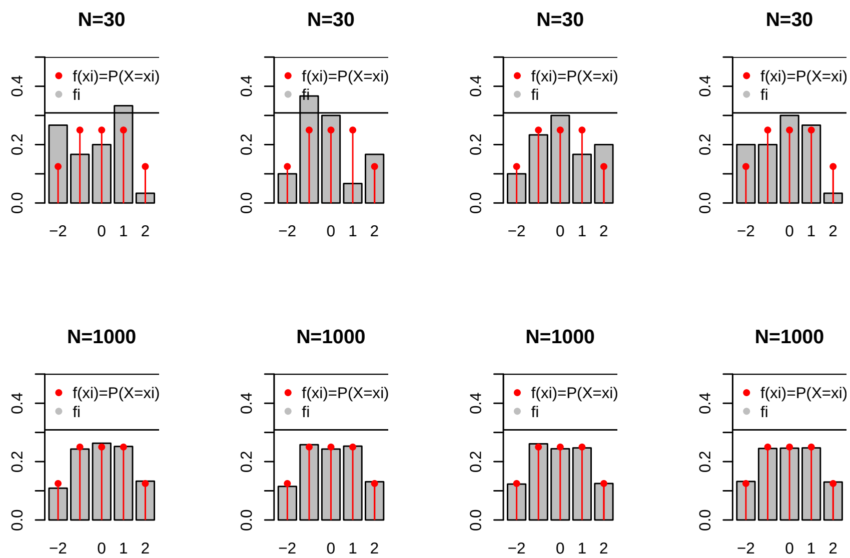
| $X$ | $f_i$ |
|---|---|
| $-2$ | 0.132 |
| $-1$ | 0.262 |
| 0 | 0.240 |
| 1 | 0.248 |
| 2 | 0.118 |

La probabilidad frecuentista nos dice

$$lim_{N\to\infty} f_i = f(x_i) = P(X = x_i)$$

Entonces, si no conocíamos el montaje del experimento (caja negra), lo mejor que podemos hacer es **estimar** las probabilidades con las frecuencias, obtenidas de $N$ repeticiones del experimento aleatorio:

$$f_i = \hat{P}_i$$




Cada vez que estimamos las probabilidades, nuestras estimaciones $\hat{P}_i = f_i$ cambian. Pero $P_i$ es una cantidad abstracta que nunca cambia. A medida que aumenta $N$, nos acercamos más a ella.

## 5.9 La media o el valor esperado

Cuando discutimos las estadísticas de resumen de los datos, definimos el **centro** de las observaciones como un valor alrededor del cual se concentran las frecuencias de los resultados.

Usamos el **promedio** para medir el centro de gravedad de los **datos**. En términos de las frecuencias relativas de los valores de los resultados discretos, escribimos el promedio como

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

$$\sum_{i=1}^{M} x_i f_i$$

**Definición**

La **media** ($\mu$) o valor esperado de una variable aleatoria discreta $X$, $E(X)$, con función de masa $f(x)$ está dada por

$$\mu = E(X) = \sum_{i=1}^{M} x_i f(x_i)$$

Es el centro de gravedad de las **probabilidades**: El punto donde se equilibran las cargas de probabilidad.

De la definición tenemos

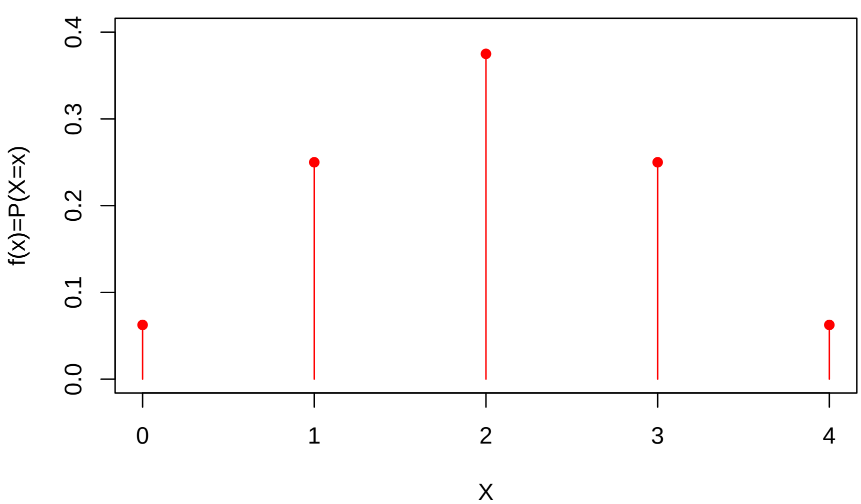
$$\bar{x} \to \mu$$

en el **límite** cuando $N \to \infty$ como la frecuencia tiende a la función de masa de probabilidad $f_i \to f(x_i)$.

**Ejemplo**

¿Cuál es la media de $X$ si su función de masa de probabilidad $f(x)$ está dada por

| $X$ | $f(x) = P(X = x)$ |
|---|---|
| 0 | 1/16 |
| 1 | 4/16 |
| 2 | 6/16 |
| 3 | 4/16 |
| 4 | 1/16 |




$$\mu = E(X) = \sum_{i=1}^{m} x_i f(x_i)$$

$E(X) =$**0** * 1/16 + **1** * 4/16 + **2** * 6/16 + **3** * 4/16 + **4** * 1/16 =2

La media $\mu$ es el centro de gravedad de la función de masa de probabilidad y **no cambia**. Sin embargo, el promedio $\bar{x}$ es el centro de gravedad de las observaciones (frecuencias relativas) **cambia** con diferentes datos.

## 5.10 Varianza

Cuando discutimos los estadísticos de resumen, también definimos la dispersión de las observaciones como una distancia promedio de los datos al promedio.

**Definición**

La varianza, escrita como $\sigma^2$ o $V(X)$, de una variable aleatoria discreta $X$ con función de masa $f(x)$ viene dada por

$$\sigma^2 = V(X) = \sum_{i=1}^{M} (x_i - \mu)^2 f(x_i)$$

$\sigma = \sqrt{V(X)}$ se llama la **desviación estándar** de la variable aleatoria.

La varianza es la dispersión de las **probabilidades** con respecto a la media: El momento de inercia de las probabilidades sobre la media.

**Ejemplo**

¿Cuál es la varianza de $X$ si su función de masa de probabilidad $f(x)$ está dada por

| $X$ | $f(x) = P(X = x)$ |
|---|---|
| 0 | 1/16 |
| 1 | 4/16 |
| 2 | 6/16 |
| 3 | 4/16 |
| 4 | 1/16 |

$$\sigma^2 = V(X) = \sum_{i=1}^{m} (x_i - \mu)^2 f(x_i)$$

$V(X) =$**(0-2)**$^2$* 1/16 + **(1-2)**$^2$* 4/16 + **(2- 2)**$^2$* 6/16 + **(3-2)**$^2$* 4/16 + **(4-2)**$^2$* 1/ 16 = 1

$$V(X) = \sigma^2 = 1$$

$$\sigma = 1$$

## 5.11 Funciones de probabilidad para funciones de $X$

En muchas ocasiones, estaremos interesados en resultados que sean función de las variables aleatorias. Quizás nos interese el cuadrado del número de contagios de gripe, o la raíz cuadrada del número de correos electrónicos en una hora.

**Definición**

Para cualquier función $h$ de una variable aleatoria $X$, con función de masa $f(x)$, su valor esperado viene dado por

$$E[h(X)] = \sum_{i=1}^{M} h(x_i) f(x_i)$$

Esta es una definición importante que nos permite probar tres propiedades de la media y la varianza que se usan con frecuencia:

1) La media de una función lineal es la función lineal de la media:

$$E(a \times X + b) = a \times E(X) + b$$

para $a$ y $b$ escalares (números ).

2) La varianza de una función lineal de $X$ es:

$$V(a \times X + b) = a^2 \times V(X)$$

3) La varianza **con respecto al origen** es la varianza **con respecto a la media** más la media al cuadrado:

$$E(X^2) = V(X) + E(X)^2$$

**Ejemplo**

¿Cuál es la varianza $X$ con respecto al origen, $E(X^2)$, si su función de masa de probabilidad $f(x)$ está dada por

| $X$ | $f(x) = P(X = x)$ |
|---|---|
| 0 | 1/16 |
| 1 | 4/16 |
| 2 | 6/16 |
| 3 | 4/16 |
| 4 | 1/16 |

$$E(X^2) = \sum_{i=1}^{m} x_i^2 f(x_i)$$

$E(X^2)$ =**(0)**$^{2*}$ 1/16 + **(1)**$^{2*}$ 4/16 + **(2)** $^{2*}$ 6/16 + **(3)**$^{2*}$ 4/16 + **(4)**$^{2*}$ 1/16 =5

También podemos verificar:

$$E(X^2) = V(X) + E(X)^2$$

$5 = 1 + 2^2$

## 5.12 Distribución de probabilidad

Cuando discutimos las estadísticas de resumen, también definimos la **distribución** de frecuencias (o la frecuencia acumulada relativa) $F_i$. $F_i$ es una cantidad importante porque es una función continua $F_x$ es por lo tanto una función de rango **continuo**, incluso si los resultados son discretos.

**Definición:**

La función de **distribución de probabilidad** se define como

$$F(x) = P(X \leq x) = \sum_{x_i \leq x} f(x_i)$$

Esa es la probabilidad acumulada hasta un valor dado $x$

$F(x)$ satisface por lo tanto satisface:

1) $0 \leq F(x) \leq 1$
2) Si $x \leq y$, entonces $F(x) \leq F(y)$

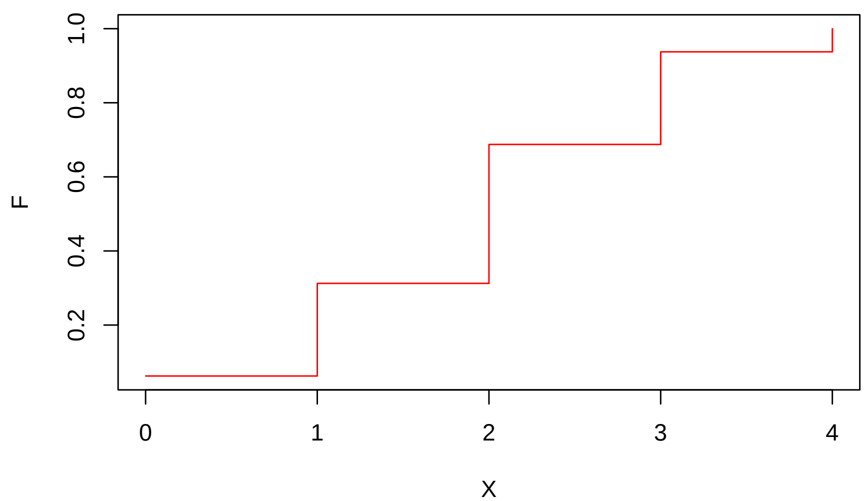
Para la función de masa de probabilidad:

| $X$ | $f(x) = P(X = x)$ |
|---|---|
| 0 | 1/16 |
| 1 | 4/16 |
| 2 | 6/16 |
| 3 | 4/16 |
| 4 | 1/16 |

La distribución de probabilidad es:

$$F(x) = \begin{cases} 1/16, & \text{if } 0 \leq x < 1 \\ 5/16, & 1 \leq x < 2 \\ 11/16, & 2 \leq x < 3 \\ 15/16, & 4 \leq x < 5 \\ 16/16, & x \leq 5 \end{cases}$$

Para $X \in \mathbb{Z}$




## 5.13 Función de probabilidad y distribución de probabilidad

La función de probabilidad y la distribución son equivalentes. Podemos obtener uno del otro y viceversa.

$$f(x_i) = F(x_i) - F(x_{i-1})$$

con

$$f(x_1) = F(x_1)$$

para $X$ tomando valores en $x_1 \leq x_2 \leq ... \leq x_n$

**Ejemplo**

De la distribución de probabilidad:

$$F(x) = \begin{cases} 1/16, & \text{if } 0 \leq x < 1 \\ 5/16, & 1 \leq x < 2 \\ 11/16, & 2 \leq x < 3 \\ 15/16, & 4 \leq x < 5 \\ 16/16, & x \leq 5 \end{cases}$$

Podemos obtener la función masa de probabilidad.

$f(0) = F(0) = 1/16$ $f(1) = F(1) - f(0) = 5/32 - 1/32 = 4/16$ $f(2) = F(2) - f(1) - f(0) = F(2) - F(1) = 6/16$ $f(3) = F(3) - f(2) - f(1) - f(0) = F(3) - F(2) = 4/16$ $f(4) = F(4) - F(3) = 1/16$

## 5.14 Cuantiles

Finalmente, podemos usar la distribución de probabilidad $F(x)$ para definir la mediana y los cuartiles de la variable aleatoria $X$.

En general, definimos el **q-cuantil** como el valor $x_p$ **bajo** el cual hemos acumulado q*100% de la probabilidad

$$q = \sum_{i=1}^{p} f(x_i) = F(x_p)$$

- La **mediana** es valor $x_m$ tal que $q = 0.5$

$$F(x_m) = 0.5$$

- El cuantil 0.05 es el valor $x_r$ tal que $q = 0.05$

$$F(x_r) = 0.05$$

- El cuantil de $0,25$ es el **primer cuartil** el valor $x_s$ tal que $q = 0.25$

$$F(x_s) = 0.25$$

## 5.15 Resumen

| nombres de cantidades | modelo (no observado) | datos (observados) |
|---|---|---|
| función de masa de probabilidad // frecuencia relativa | $f(x_i) = P(X = x_i)$ | $f_i = \frac{n_i}{N}$ |

| nombres de cantidades | modelo (no observado) | datos (observados) |
|---|---|---|
| distribución de probabilidad // frecuencia relativa acumulada | $F(x_i) = P(X \leq x_i)$ | $F_i = \sum_{k \leq i} f_k$ |
| media // promedio | $\mu = E(X) = \sum_{i=1}^{M} x_i f(x_i)$ | $\bar{x} = \sum_{j=1}^{N} x_j / N$ |
| varianza // varianza de la muestra | $\sigma^2 = V(X) = \sum_{i=1}^{M} (x_i - \mu)^2 f(x_i)$ | $s^2 = \sum_{j=1}^{N} (x_j - \bar{x})^2 / (N-1)$ |
| desviación estándar // muestra sd | $\sigma = \sqrt{V(X)}$ | $s$ |
| varianza con respecto al origen // 2º momento muestral | $E(X^2) = \sum_{i=1}^{M} x_i^2 f(x_i)$ | $m_2 = \sum_{j=1}^{N} x_j^2 / n$ |

Ten en cuenta:

- $i = 1...M$ es un **resultado** de la variable aleatoria $X$.
- $j = 1...N$ es una **observación** de la variable aleatoria $X$.

Propiedades:

- $\sum_{i=1...N} f(x_i) = 1$
- $f(x_i) = F(x_i) - F(x_{i-1})$
- $E(a \times X + b) = a \times E(X) + b$; for $a$ and $b$ scalars.
- $V(a \times X + b) = a^2 \times V(X)$
- $E(X^2) = V(X) + E(X)^2$

## 5.16 Preguntas

**1)** Para una función de masa de probabilidad no es cierto que

**a:** la suma de los valores de su imagen es 1; **b:** sus valores pueden interpretarse como probabilidades de eventos; **c:** siempre es positiva; **d:** no puede tomar el valor 1;

**2)** El valor de una variable aleatoria representa

**a:** una observación de un experimento aleatorio; **b:** la frecuencia de un resultado de un experimento aleatorio; **c:** un resultado de un experimento aleatorio; **d:** una probabilidad de un resultado;

**3)** El valor estimado de una probabilidad $\hat{P}_i$ es igual a la probabilidad $P_i$ cuando el número de repeticiones del experimento aleatorio es

**a:** grande; **b:** infinito; **c:** pequeño **d:** cero;

**4)** Si una función de masa de probabilidad es simétrica alrededor de $x = 0$

**a:** La media es menor que la mediana; **b:** La media es mayor que la mediana; **c:** La media y la mediana son iguales; **d:** La media y la mediana son diferentes de 0;

**5)** La media y la varianza

**a:** son inversamente proporcionales; **b:** son valores esperados de funciones de $X$; **c:** de una función lineal son la función lineal de la media y la función lineal de la varianza; **d:** cambia cuando repetimos el experimento aleatorio;

## 5.17 Ejercicios

#### 5.17.0.1 Ejercicio 1

Considera la siguiente variable aleatoria $X$ sobre los resultados

| resultado | $X$ |
|---|---|
| $a$ | 0 |
| $b$ | 0 |
| $c$ | 1.5 |
| $d$ | 1.5 |
| $e$ | 2 |
| $f$ | 3 |

a) Si cada resultado es igualmente probable, ¿cuál es la función de masa de probabilidad de $x$?

b) Encuentra:

- $P(X > 3)$
- $P(X = 0 \cup X = 2)$
- $P(X \leq 2)$

#### 5.17.0.2 Ejercicio 2

Dada la función de masa de probabilidad

| $x$ | $f(x) = P(X = x)$ |
|---|---|
| 10 | 0.1 |
| 12 | 0.3 |
| 14 | 0.25 |
| 15 | 0.15 |
| 17 | ? |
| 20 | 0.15 |

- ¿Cuál es su valor esperado y su desviación estándar? (R: 14,2; 2,95)

#### 5.17.0.3 Ejercicio 3

Dada la distribución de probabilidad para una variable discreta $X$

$$F(x) = \begin{cases} 0, & x < -1 \\ 0.2, & x \in [-1, 0) \\ 0.35, & x \in [0, 1) \\ 0.45, & x \in [1, 2) \\ 1, & x \geq 2 \end{cases}$$

- encuentra $f(x)$
- encuentra $E(X)$ y $V(X)$ (R:1; 1.5)
- cuál es el valor esperado y la varianza de $Y = 2X + 3$ (R:5, 6)
- ¿Cuál es la mediana y el primer y tercer cuartil de $X$? (R:2,0,2)

#### 5.17.0.4 Ejercicio 4

Estamos probando un sistema para transmitir imágenes digitales. Primero consideramos el experimento de enviar 3 píxeles y tener como resultados **posibles** eventos como $(0, 1, 1)$. Este es el evento de recibir el primer píxel sin error, el segundo con error y el tercero con error.

- Enumera en una columna el espacio muestral del experimento aleatorio.
- En la segunda columna asigna la variable aleatoria que cuenta el número de errores transmitidos para cada resultado

Considera que tenemos un canal totalmente ruidoso, es decir, cualquier resultado de tres píxeles es igualmente probable.

- ¿Cuál es la probabilidad de recibir errores de 0, 1, 2 o 3 en la transmisión de 3 píxeles? (R: 1/8; 3/8; 3/8; 1/8)
- Dibuja la función de masa de probabilidad para el número de errores
- ¿Cuál es el valor esperado para el número de errores? (R:1.5)
- ¿Cuál es su varianza? (R: 0,75)
- Dibuja la distribución de probabilidad
- ¿Cuál es la probabilidad de transmitir al menos 1 error? (R:7/8)

# Chapter 6

# Variables aleatorias continuas

## 6.1 Objetivo

En este capítulo estudiaremos variables aleatorias continuas.

Definiremos la función de densidad de probabilidad, su media y varianza. De forma similar a las variables aleatorias discretas, definiremos la función de distribución de probabilidad.

## 6.2 Variables aleatorias continuas

En el capítulo pasado usamos las probabilidades de variables aleatorias discretas para definir la función de masa de probabilidad

$$f(x) = P(X = x)$$

Donde la probabilidad de que la variable aleatoria tome el valor $x$ la entendemos como el valor de su frecuencia relativa, cuando el número de repeticiones del experimiento aleatorio tiende a infinito.

Cuando hablamos de datos continous vimos que teníamos que transformarlos en variables discretas (bins) para producir tablas de frecuencias relativas o histogramas. Veamos cómo definir las probabilidades de las variables continuas teniendo en cuenta estas particiones.

**Ejemplo ( misofonía )**

Reconsideremos el ángulo de convexidad de los pacientes con misofonía (Sección 2.21). El ángulo de convexidas de 123 pacientes fue medido. Entendimos cada medición como el resultado de un experimento aleatorio que repetimos 123 veces y que podíamos describir en una tabla de frecuencias o en un histograma.

Para hacer esto redefinimos los resultados como pequeños intervalos regulares (bins) y calculamos la frecuencia relativa de cada intervalo.

```
##        outcome ni         fi
## 1 [-1.02,3.46]  8 0.06504065
## 2  (3.46,7.92] 51 0.41463415
## 3  (7.92,12.4] 26 0.21138211
## 4  (12.4,16.8] 20 0.16260163
## 5  (16.8,21.3] 18 0.14634146
```

## 6.3 frecuencias relativas

Por lo tanto, definimos la probabilidad de observar un intervalo $i$ como la frecuencia relativa del intervalo cuando $N \rightarrow \infty$

$$f_i = \frac{n_i}{N} \rightarrow P(x_i \leq X \leq x_i + \Delta x)$$

Esta probabilidad depende de la longitud de los bins $\Delta x$.

Si hacemos los bins cada vez más pequeños, las frecuencias se hacen más pequeñas y, por lo tanto,

$$P(x_i \leq X \leq x_i + \Delta x) \rightarrow 0$$

cuando $\Delta x \rightarrow 0$ porque $n_i \rightarrow 0$

Veamos cómo las frecuencias relativas se hacen más pequeñas cuando dividimos el rango de $X$ en 20 bins

```
##          outcome ni         fi
## 1  [-1.02,0.115]  2 0.01626016
## 2   (0.115,1.23]  0 0.00000000
## 3    (1.23,2.34]  3 0.02439024
## 4    (2.34,3.46]  3 0.02439024
## 5    (3.46,4.58]  2 0.01626016
## 6    (4.58,5.69]  4 0.03252033
## 7     (5.69,6.8] 11 0.08943089
## 8     (6.8,7.92] 34 0.27642276
## 9    (7.92,9.04] 12 0.09756098
## 10   (9.04,10.2]  4 0.03252033
## 11   (10.2,11.3]  3 0.02439024
## 12   (11.3,12.4]  7 0.05691057
## 13   (12.4,13.5]  2 0.01626016
## 14   (13.5,14.6]  6 0.04878049
```

```
## 15   (14.6,15.7]  4 0.03252033
## 16   (15.7,16.8]  8 0.06504065
## 17     (16.8,18]  4 0.03252033
## 18     (18,19.1]  9 0.07317073
## 19   (19.1,20.2]  3 0.02439024
## 20   (20.2,21.3]  2 0.01626016
```

## 6.4 función de densidad de probabilidad

Definimos una cantidad en un punto $x$ que es **la probabilidad por unidad de distancia** de que una observación esté en el bin **infinitesimal** entre $x$ y $x+dx$

$$f(x) = \frac{P(x \leq X \leq x+dx)}{dx}$$

$f(x)$ se llama la **función de densidad de probabilidad**.

Por lo tanto, la probabilidad de observar $X$ entre $x$ y $x+dx$ está dada por

$$P(x \leq X \leq x+dx) = f(x)dx$$

**Definición**

Para una variable aleatoria continua $X$, una función de **densidad de probabilidad** es tal que

1) Es positiva:

$$f(x) \geq 0$$

2) La probabilidad de observar **cualquier** valor de $x$ es 1:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

3) La probabilidad de observar un valor dentro de un intervalo es el **área bajo la curva**:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

Las propiedades aseguran que $f(x)dx$ satisfacen las propiedades de probabilidad de Kolmogorov.

La función de densidad de probabilidad es un paso más en la abstracción de probabilidades en la que añadimos el límite continuo

$$dx \to 0$$

Todas las propiedades de las probabilidades se traducen en términos de densidades y por lo tanto cambiamos sumatorios por integrales

$$\sum \to \int$$

Las densidades de probabilidad son cantidades matemáticas que no necesariamente representan experimientos aleatorios.

Un interés fundamental en estadística es describir las densidades que describen nuestro experimento aleatorio concreto.

## 6.5 Área total bajo la curva

**Ejemplo (gotas de lluvia)**

tomemos una **densidad de probabilidad** que podría describir la variable aleatoria que mide dónde cae una gota de lluvia en una canaleta de 100 cm de longitud.

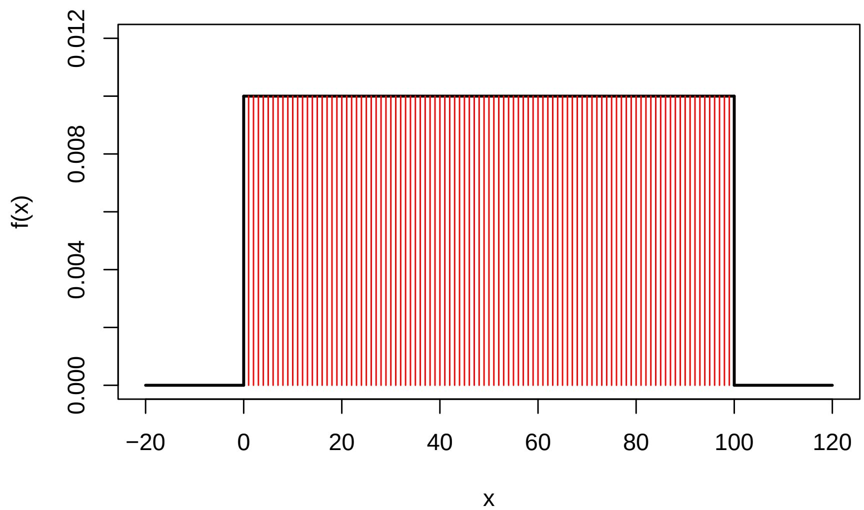
$$f(x) = \begin{cases} \frac{1}{100}, & \text{si } x \in (0, 100) \\ 0, & de\, otra\, forma \end{cases}$$

Verifiquemos que la función satisface las tres propiedades de una densidad de probabilidad.

1) es evidente a partir de la definición que $f(x) \geq 0$

2) La probabilidad de observar **cualquier valor** de $X$ es el **área total bajo la curva**

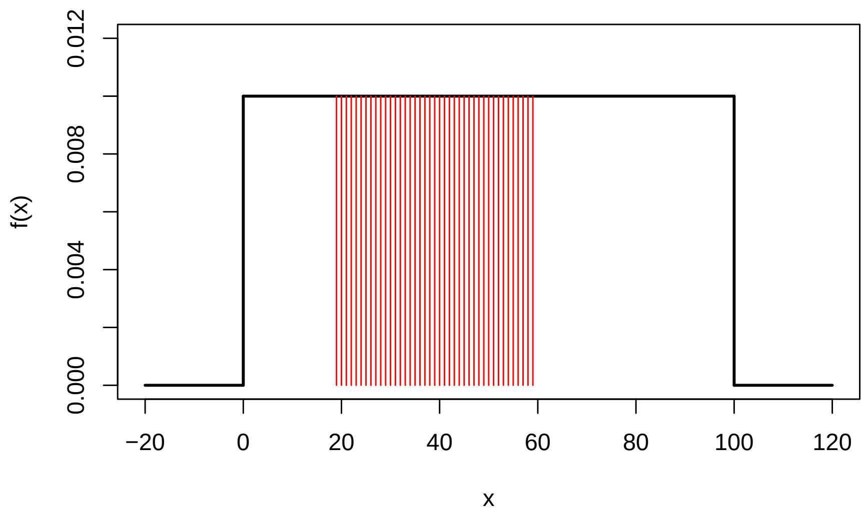
$P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f(x)dx = 100 * 0.01 = 1$




## 6.6 Área bajo la curva

3) La probabilidad de observar $X$ en un intervalo es el **área bajo la curva** dentro del intervalo

- $P(20 \leq X \leq 60) = \int_{20}^{60} f(x)dx = (60 - 20) * 0.01 = 0.4$




## 6.7 Probabilidades de variables continuas

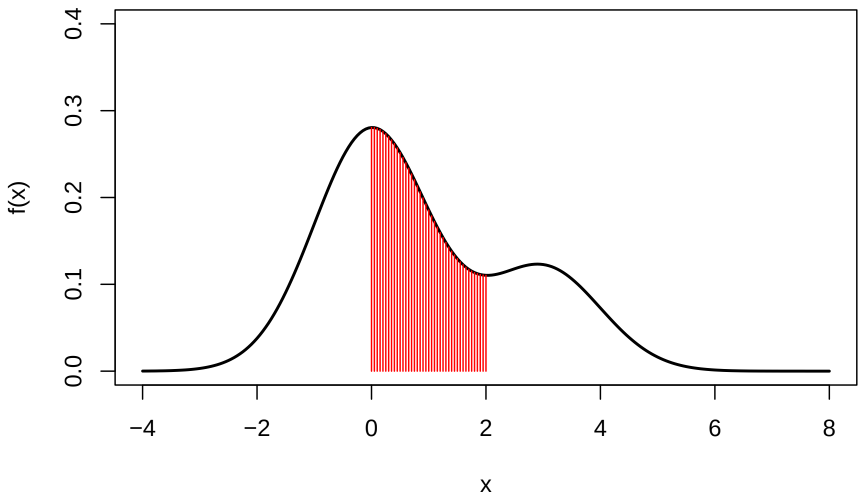
Para variables continuas, calculamos la probabilidad de que la variable esté en un intervalo dado. Eso es

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

Recordemos que para variables continuas, la probabilidad de que el experimento nos dé un número real particular es cero: $P(X = a) = 0$

La probabilidad $P(a \leq X \leq b)$ es el área bajo la curva de $f(x)$ entre $a$ y $b$

- $P(a \leq X \leq b) = \int_a^b f(x)dx$




## 6.8 Distribución de probabilidad

La **distribución de probabilidad** $F(c)$ definida como la acumulación de probabilidad hasta el resultado $C$

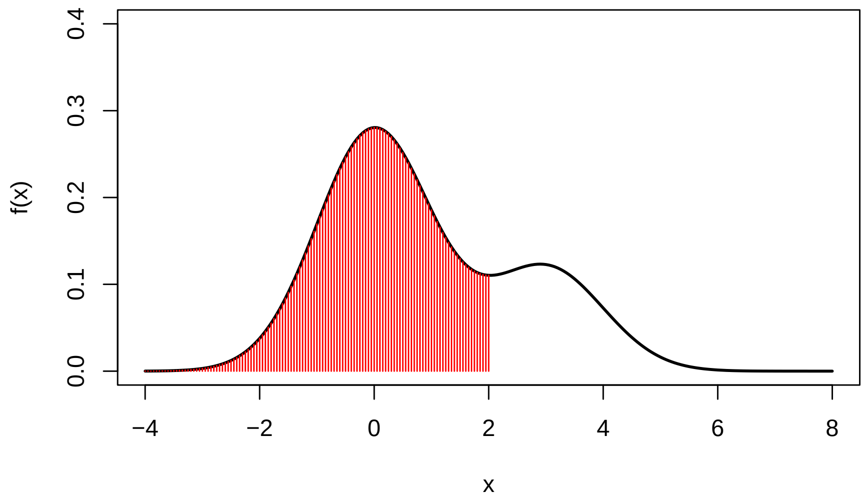
$F(c) = P(X \leq c)$

se puede utilizar para calcular la probabilidad $P(a \leq X \leq b)$.

Si consideramos que:

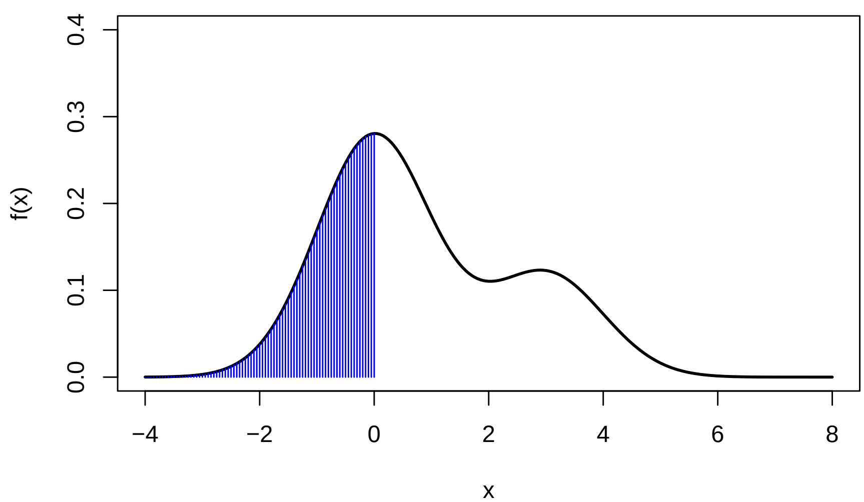
1) la probabilidad acumulada hasta $b$ está dada por

- $F(b) = P(X \leq b) = \int_{-\infty}^{b} f(x)dx$



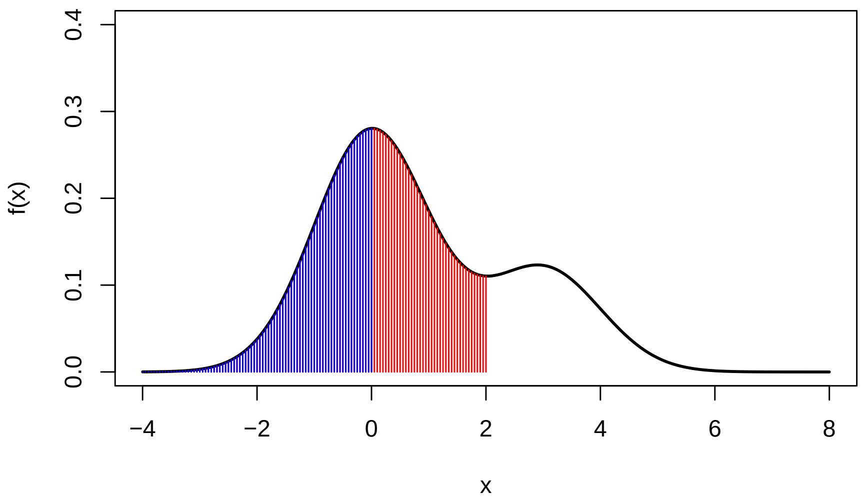

2) la probabilidad acumulada hasta $a$ es

- $F(a) = P(X \leq a)$




Entonces, la probabilidad entre $a$ y $b$ viene dada por la diferencia en el valor de la distribución de probabilidad

- $P(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$




**Definición**

La **distribución de probabilidad** de una variable aleatoria continua se define como

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$$

y tiene las siguientes propiedades:

1) Está entre 0 y 1:

$$F(-\infty) = 0 \; y \; F(\infty) = 1$$

2) Siempre aumenta:

$$F(a) \leq F(b)$$

si $a \leq b$

3) Se puede utilizar para calcular probabilidades:

$$P(a \leq X \leq b) = F(b) - F(a)$$

4) Recupera la densidad de probabilidad:

$$f(x) = \frac{dF(x)}{dx}$$

Usamos **distribuciones de probabilidad** para **calcular probabilidades** de una variable aleatoria dentro de intervalos, y su derivada es la función de densidad de probabilidad.

**Ejemplo (gotas de lluvia)**

Para la función de densidad uniforme:

$$f(x) = \begin{cases} \frac{1}{100}, & \text{si } x \in (0, 100) \\ 0, & de\, otra\, forma \end{cases}$$

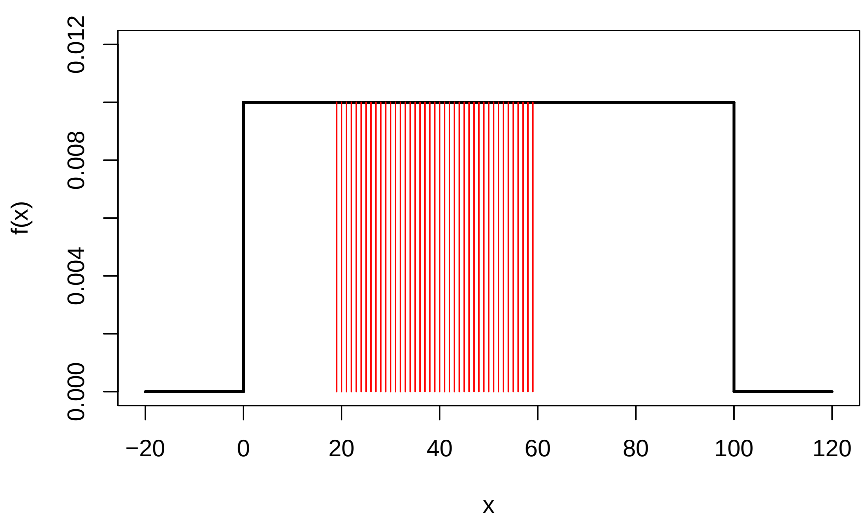
Encontramos que la distribución de probabilidad es

$$F(a) = \begin{cases} 0, & \text{si } a \leq 0 \\ \frac{a}{100}, & \text{si } a \in (0, 100) \\ 1, & \text{si } 100 \leq a \\ \end{cases}$$

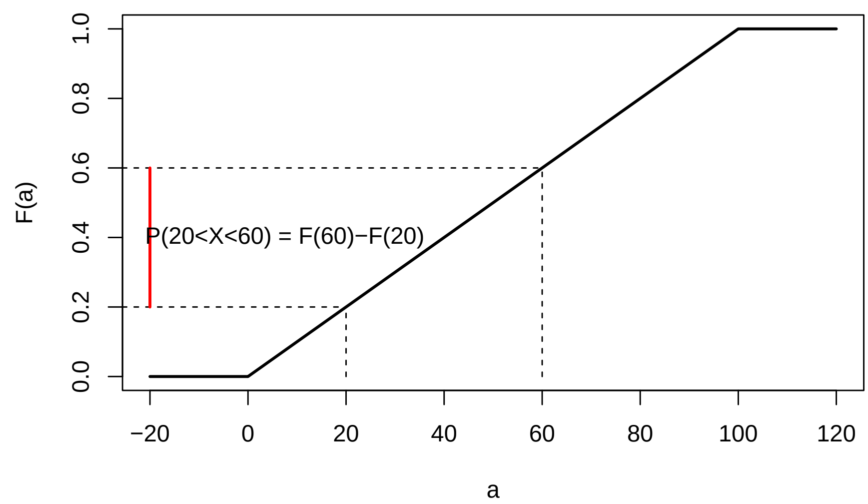
## 6.9 Gráficas de probabilidad

1) Podemos dibujar la probabilidad de una variable aleatoria en un intervalo como el *área* bajo la curva de la **densidad**. Por ejemplo

$$P(20 < X < 60)$$




2) También podemos dibujar la probabilidad $P(20 < X < 60)$ como la *diferencia* en los valores de la **distribución**




## 6.10 Media

Como en el caso discreto, la **media** mide el centro de masa de las probabilidades

**Definición**

Supongamos que $X$ es una variable aleatoria continua con función de probabilidad **densidad** $f(x)$. El valor medio o esperado de $X$, denotado como $\mu$ o $E(X)$, es

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Es la versión continua del centro de gravedad.

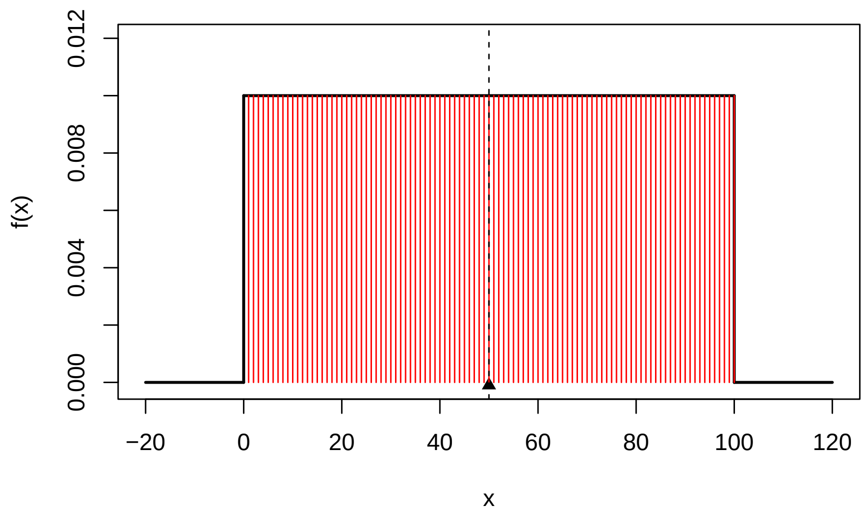
**Ejemplo (gotas de lluvia)**

La variable aleatoria con densidad de probabilidad

$$f(x) = \begin{cases} \frac{1}{100}, & \text{si } x \in (0, 100) \\ 0, & de\, otra\, forma \end{cases}$$

Tiene un valor esperado en

$$E(X) = 50$$




## 6.11 Varianza

Como en el caso discreto, la varianza mide la dispersión de probabilidades sobre la media

**Definición**

Supongamos que $X$ es una variable aleatoria continua con función de densidad de probabilidad $f(x)$. La varianza de $X$, denotada como $\sigma^2$ o $V(X)$, es

$$\sigma^2 = V(X) = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx$$

Es la versión continua del momento de inercia.

## 6.12 Funciones de $X$

En muchas ocasiones, estaremos interesados en resultados que sean función de las variables aleatorias. Tal vez nos interese el cuadrado de la elongación de un muelle, o la raíz cuadrada de la temperatura de un motor.

**Definición**

Para cualquier función $h$ de una variable aleatoria $X$, con función de masa $f(x)$, su valor esperado viene dado por

$$E[h(X)] = \int_{-\infty}^{\infty} h(x)f(x)dx$$

De esta definición recuperamos las mismas propiedades que en el caso discreto

1) La media de una función lineal es la función lineal de la media:

$$E(a \times X + b) = a \times E(X) + b$$

para $a$ y $b$ escalares.

2) La varianza de una función lineal de $X$ es:

$$V(a \times X + b) = a^2 \times V(X)$$

3) La varianza sobre el origen es la varianza sobre la media más la media al cuadrado:

$$E(X^2) = V(X) + E(X)^2$$

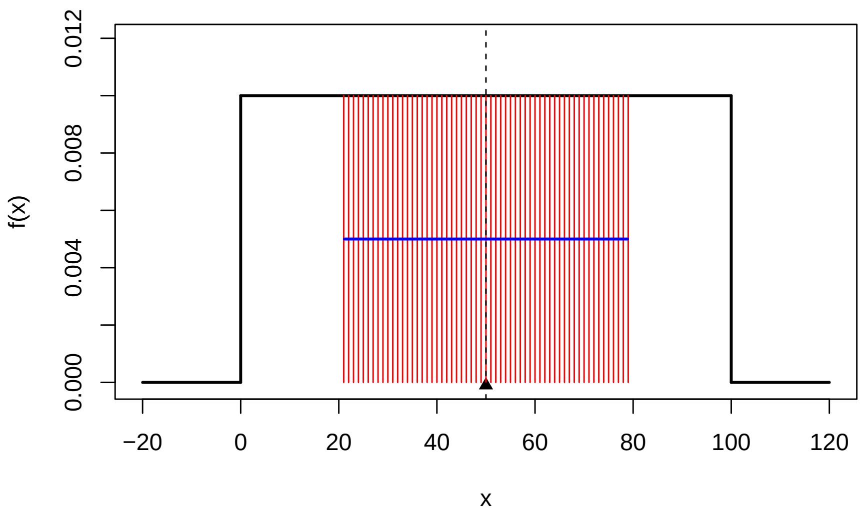
## 6.13 Ejercicios

##### 6.13.0.1 Ejercicio 1

Para la densidad de probabilidad

$$f(x) = \begin{cases} \frac{1}{100}, & \text{si } x \in (0, 100) \\ 0, & de\, otra\, forma \end{cases}$$

- calcular la media (R:50)
- calcular la varianza usando $E(X^2) = V(X) + E(X)^2$ (R:100^2/12)
- calcula $P(\mu - \sigma \leq X \leq \mu + \sigma)$ (R: 0.57)
- ¿Cuáles son el primer y tercer cuartiles? (R: 25; 75)




#### 6.13.0.2 Ejercicio 2

Dado

$$f(x) = \begin{cases} 0, & x < 0 \\ ax, & x \in [0,3] \\ b, & x \in (3,5) \\ \frac{b}{3}(8-x), & x \in [5,8] \\ 0, & x > 8 \end{cases}$$

- ¿Cuáles son los valores de $a$ y $b$ tales que $f(x)$ es una función de densidad de probabilidad continua ? (R: 1/15; 1/5)
- ¿Cuál es la media de $X$? (R:4)

#### 6.13.0.3 Ejercicio 3

Para la densidad de probabilidad

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x \geq 0 \\ 0, & de\, otra\, forma \end{cases}$$

- Confirmar que se trata de una densidad de probabilidad
- Calcular la media (R: $1/\lambda$)
- Calcule el valor esperado de $X^2$ (R: $2/\lambda^2$)

- Calcular la varianza (R: $1/\lambda^2$)
- Hallar la distribución de probabilidad $F(a)$ (R: $1 - exp(-\lambda a)$)
- Encuentra la mediana (R: $\log 2\ /\lambda$)

#### 6.13.0.4 Ejercicio 4

Dada la distribución acumulativa de una variable aleatoria $X$

$$F(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{80}(17 + 16x - x^2), & x \in [-1, 7) \\ 1, & x \geq 7 \end{cases}$$

calcular:

- $P(X > 0)$ (R:63/80)
- $E(X)$ (R:1.93)
- $P(X > 0|X < 2)$ (R:28/45)

# Chapter 7

# Modelos de probabilidad para variables aletorias discretas

## 7.1 Objetivo

En este capítulo veremos algunas funciones de masa de probabilidad que se utilizan para describir experimentos aleatorios comunes.

Introduciremos el concepto de parámetro y por tanto de modelos paramétricos.

En particular, discutiremos las funciones de probabilidad uniforme y de Bernoulli y cómo se usan para derivar las funciones de probabilidad binomial y binomial negativa. También hableramos del modelo hipergeométrico.

## 7.2 Función de probabilidad

Recordemos que una función de masa de probabilidad de una **variable aleatoria discreta** $X$ con valores posibles $x_1, x_2, .., x_M$ es **cualquier función** tal que

1) Nos permite calcular probabilidades para todos los resultados

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

2) Siempre es positiva:

$$f(x_i) \geq 0$$

3) La probabilidad de obtener algo en el experimento aleatorio es 1

$$\sum_{i=1}^{M} f(x_i) = 1$$

Estudiamos dos **propiedades importantes:**

1) La media como medida de tendencia central:

$$E(X) = \sum_{i=1}^{M} x_i f(x_i)$$

2) La varianza como medida de dispersión:

$$V(X) = \sum_{i=1}^{M} (x_i - \mu)^2 f(x_i)$$

## 7.3 Modelo de probabilidad

Un **modelo de probabilidad** es una función de masa de probabilidad que puede representar las probabilidades de un experimento aleatorio.

**Ejemplos:**

1) La función de masa de probabilidad definida por

| $X$ | $f(x)$ |
|---|---|
| $-2$ | $1/8$ |
| $-1$ | $2/8$ |
| $0$ | $2/8$ |
| $1$ | $2/8$ |
| $2$ | $1/8$ |

Representa la probabilidad de sacar **una** bola de una urna donde hay dos bolas con etiquetas: $-1, 0, 1$ y una bola con etiquetas: $-2, 2$.

2) $f(x) = P(X = x) = 1/6$ representa la probabilidad de los resultados de **un** lanzamiento de un dado.

## 7.4 Modelos paramétricos

Cuando tenemos un experimento aleatorio con $M$ resultados posibles, necesitamos encontrar $M$ números para determinar la función de masa de probabilidad.

Como en el ejemplo 1 anterior, necesitábamos 5 valores en la columna $f(x)$ de la tabla de probabilidad.

Sin embargo, **en muchos casos**, podemos formular funciones de probabilidad $f(x)$ que dependen únicamente de **muy pocos** números. Al igual que en el ejemplo 2 anterior, solo necesitábamos saber cuántos resultados posibles puede dar un dado.

**Ejemplo (probabilidad clásica):**

Un experimento aleatorio con $M$ resultados igualmente probables tiene una función de masa de probabilidad:

$$f(x) = P(X = x) = 1/M$$

Sólo necesitamos saber $M$.

Los números que **necesitamos saber** para determinar completamente una función de probabilidad se llaman **parámetros**.

## 7.5 Distribución uniforme (un parámetro)

El ejemplo anterior es la interpretación clásica de la probabilidad y define nuestro primer modelo paramétrico.

**Definición**

Una variable aleatoria $X$ con resultados $\{1, ...M\}$ tiene una **distribución uniforme** discreta si todos sus resultados $M$ tienen la misma probabilidad

$$f(x) = \frac{1}{M}$$

$M$ es el parámetro natural del modelo. Una vez que definimos $M$ para un experimento, elegimos una función de masa de probabilidad particular. La función anterior es realmente una **familia** de funciones que dependen de $M$: $f(x; M)$.

La media y la varianza de una variable que sigue una distribución uniforme son:

$$E(X) = \frac{M + 1}{2}$$

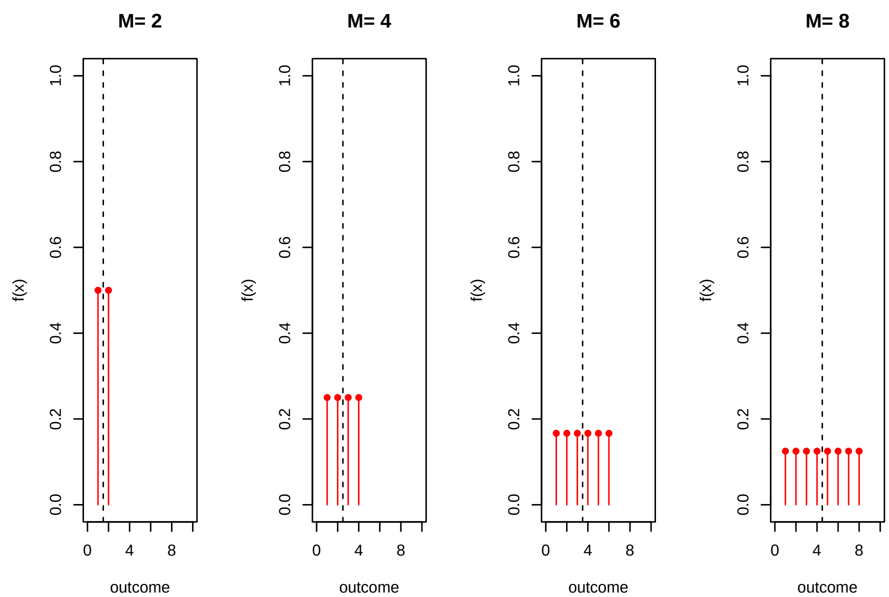
y

$$V(X) = \frac{M^2 - 1}{12}$$

Nota: $E(X)$ y $V(X)$ también son **parámetros**. Si conocemos alguno de ellos, entonces podemos determinar completamente la distribución. Por ejemplo:

$$f(x) = \frac{1}{2E(X) - 1}$$

Veamos algunas funciones de masa de probabilidad en la familia de modelos paramétricos uniformes:




## 7.6 Distribución uniforme (dos parámetros)

Presentemos un nuevo modelo de probabilidad **uniforme** con **dos parámetros**: los resultados mínimo y máximo.

Si la variable aleatoria toma valores en $\{a, a+1, ...b\}$, donde $a$ y $b$ son números enteros y todos los resultados son igualmente probables, entonces

$$f(x) = \frac{1}{b - a + 1}$$

porque $M = b - a + 1$.

Entonces decimos que $X$ se distribuye uniformemente entre $a$ y $b$ y escribimos

$$X \rightarrow Unif(a, b)$$

**Propiedades:**

Si $X$ se distribuye uniformemente entre $a$ y $b$

$$X \rightarrow Unif(a, b)$$

1) Su media es

$$E(X) = \frac{b+a}{2}$$

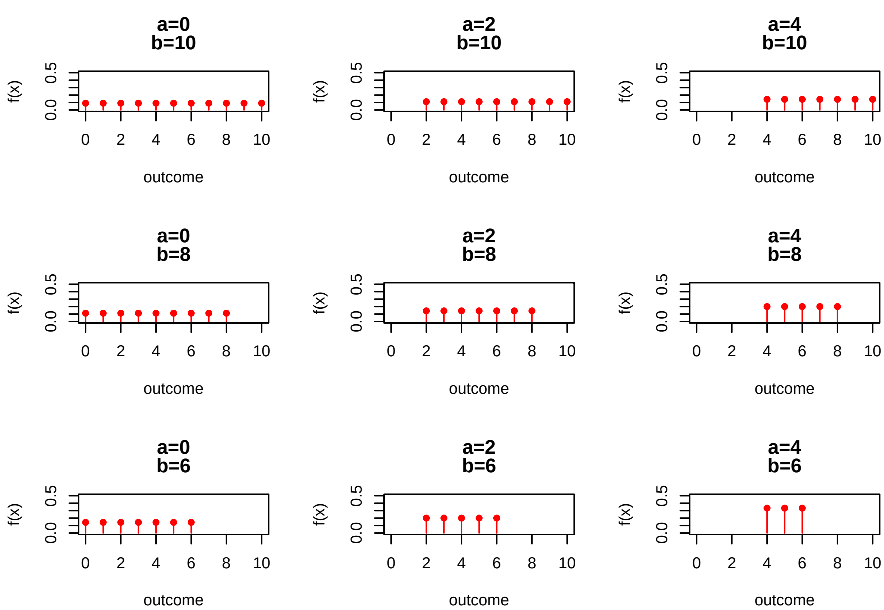
2) Su varianza es

$$V(X) = \frac{(b-a+1)^2 - 1}{12}$$

Para probar esto cambia las variables $X = Y + a - 1$, $y \in \{1, ...M\}$.

**Funciones de masa de probabilidad**

Veamos algunas funciones de masa de probabilidad en la familia de modelos paramétricos uniformes:




**Ejemplo (clases escolares):**

¿Cuál es la probabilidad de observar a un niño de una edad particular en una escuela primaria (si todas las clases tienen la misma cantidad de niños)?

De la configuración del experimento sabemos: $a = 6$ y $b = 11$ entonces

$$X \to Unif(a = 6, b = 11)$$

eso es

$$f(x) = \frac{1}{6}$$

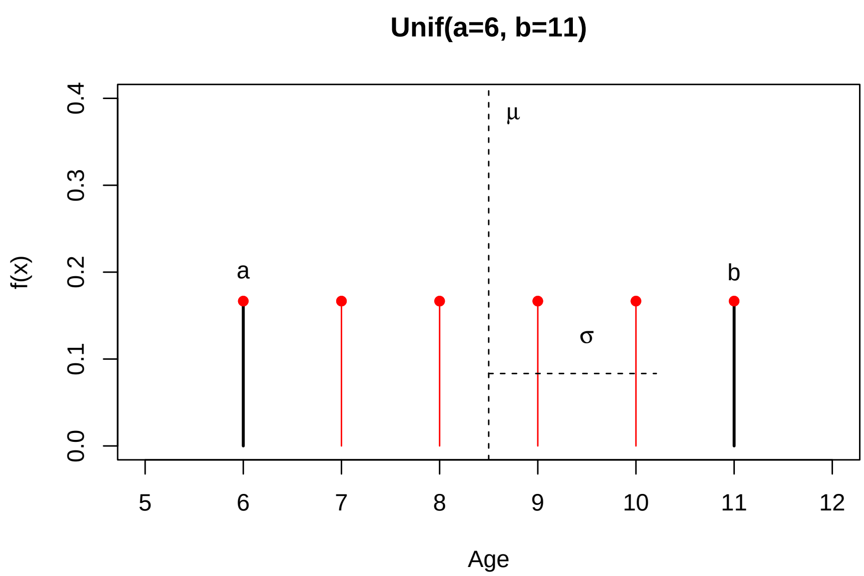
para $x \in \{6, 7, 8, 9, 10, 11\}$, y 0 en caso contrario.

La media y la varianza de esta función de masa de probabilidad es:

- $E(X) = 8.5$
- $V(X) = 2.916667$

Recuerda

- El valor esperado es la **media** $\mu = 8.5$
- La **desviación estándar** $\sigma = 1.707825$ es la distancia promedio desde la media y se calcula a partir de la raíz cuadrada de la varianza.




**Parámetros y Modelos:**

Un **modelo** es una función particular $f(x)$ que **describe** nuestro experimento.

Si el modelo es una función **conocida** que depende de algunos parámetros, al cambiar el valor de los parámetros producimos una **familia de modelos**: $f(x; a, b)$.

El conocimiento de $f(x)$ se reduce al conocimiento del valor de los parámetros $a$, $b$.

Idealmente, el modelo y los parámetros son **interpretables**.

En nuestro ejemplo, $a$ representa la edad mínima en la escuela y $b$ la edad máxima. Pueden considerarse como las **propiedades físicas** del experimento.

## 7.7 ensayo de Bernoulli

Intentemos avanzar desde el caso de probabilidad igual y supongamos un modelo con solo dos resultados posibles ($A$ y $B$) que tienen probabilidades **desiguales**

**Ejemplos:**

- Anotar el sexo de un paciente que acude a urgencias de un hospital ($A : masculino$ y $B : femenino$).
- Registrar si una máquina fabricada es defectuosa o no ($A : defectuosa$ y $B : no\ defectuosa$).
- Dar en el blanco ($A : xito$ y $B : fracaso$).
- Transmitir un píxel correctamente ($A : s$ y $B : no$).

En estos ejemplos, la probabilidad del resultado $A$ suele ser **desconocida**.

**Modelo de probabilidad:**

Introduciremos la probabilidad de un resultado ($A$) como el **parámetro** del modelo. El modelo se puede escribir en diferentes formas.

1) Como una tabla de probabilidad:

| *Resultado* | $P_i$ |
|---|---|
| $A$ | $p$ |
| $B$ | $1-p$ |

- $i \in \{A, B\}$
- resultado $A$ (éxito): tiene probabilidad $p$ (parámetro)
- resultado $B$ (fracaso): tiene una probabilidad $1-p$

2) Como función de masa de probabilidad de la variable aleatoria $K$ tomando valores $\{0, 1\}$ para $B$ y $A$, respectivamente.

$$f(k) = \begin{cases} 1-p, & k = 0\,(event\,B) \\ p, & k = 1\,(event\,A) \end{cases}$$

3) Como una función de $k$

$$f(k;p) = p^k(1-p)^{1-k}$$

para $k = (0, 1)$.

Entonces decimos que $K$ sigue una distribución de Bernoulli con parámetro $p$

$$K \to Bernoulli(p)$$

**Propiedades:**

Si $K$ sigue una distribución de Bernoulli, entonces

1) su media es

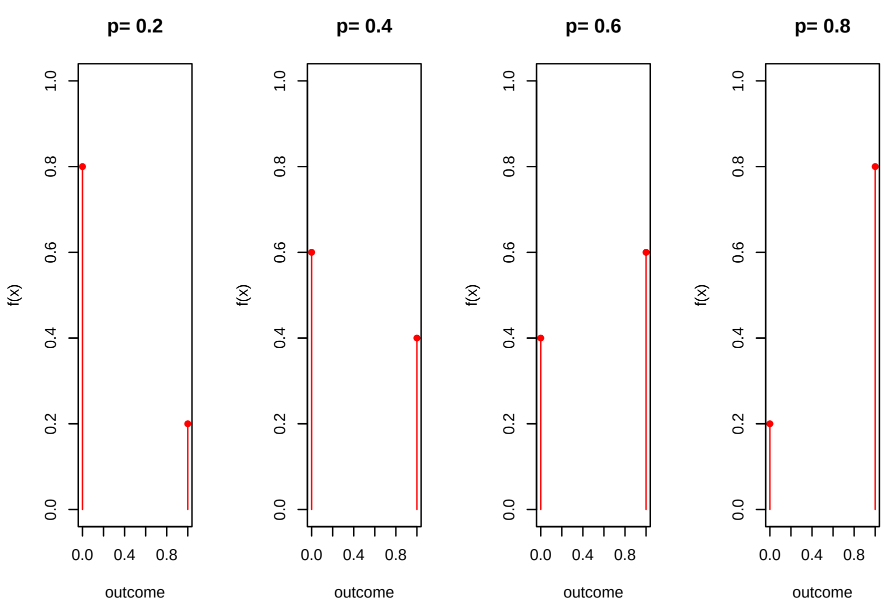
$$E(K) = p$$

2) su varianza es

$$V(K) = (1-p)p$$

Ten en cuenta que la probabilidad del resultado $A$ es el parámetro $p$ que es lo mismo que su valor en $k = 1$: $f(1) = P(k = 1)$. El parámetro determina completamente la función de masa de probabilidad, incluidas su media y varianza.

Veamos algunas funciones de masa de probabilidad en la familia de modelos paramétricos uniformes:



## 7.8 Experimento binomial

Cuando estamos interesados en predecir **frecuencias absolutas** cuando conocemos el parámetro $p$ de un ensayo particular de Bernoulli, entonces

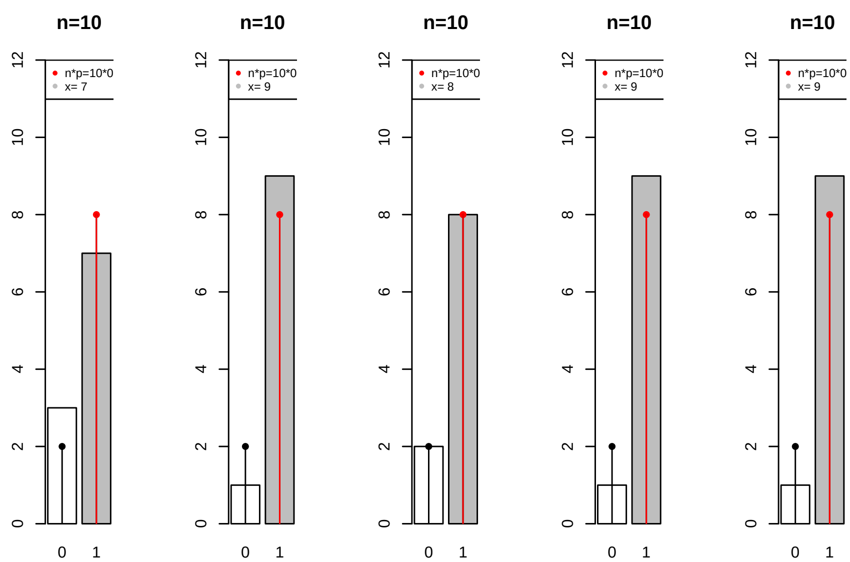
1) **repetimos** el ensayo de Bernoulli $n$ veces y contamos cuantas veces obtuvimos $A$ usando la frecuencia absoluta de $A$: $N_A$.

2) definimos una **variable aleatoria** $X = N_A$ tomando valores $x \in 0, 1, ...n$

Cuando repetimos $n$ veces una ensayo de Bernoulli, obtenemos un valor para $n_A$. Si realiza otras pruebas de Bernoulli de $n$, entonces $n_A$ da otro valor. $N_A = X$ es la variable aleatoria, $n_A = x$ es su observación.

**Ejemplo (experimento binomial):**

Imaginemos una repetición de un **experimento binomial** 5 veces. Cada experimento binomial es la repetición de un **ensayo de Bernoulli** $n = 10$ veces. En cada **Experimento Binomial** obtuvimos un valor diferente de $x$ que mide la frecuencia relativa de $A$. En este experimento queremos **predecir** el valor de $X$, es decir, nos pregutamos cuál es la probabilidad de un número específico de eventos (por ejemplo, $X = 9$).




**Ejemplos (Experimentos binomiales en contexto):**

- Anotamos el sexo de $n = 10$ pacientes que acuden a urgencias de un hospital. ¿Cuál es la probabilidad de que $X = 9$ los pacientes sean hombres cuando $p = 0.8$?

- Intentamos $n = 5$ veces de dar en un blanco ($A : xito$ y $B : fracaso$). ¿Cuál es la probabilidad de que alcancemos el objetivo $X = 5$ veces cuando normalmente lo hacemos el 25% de las veces ($p = 0.25$)?
- Transmitimos $n = 100$ píxeles correctamente ($A : s$ y $B : no$). ¿Cuál es la probabilidad de que $X = 2$ píxeles sean errores, cuando la probabilidad de error es $p = 0.1$?

## 7.9 Función de probabilidad binomial

Supongamos que **sabemos** el valor real del parámetro del ensayo de Bernoulli $p$ ($lim_{n\to\infty} x = np$).

Cuando repetimos un ensayo de Bernoulli y paramos en la repetición número $n$, ¿el valor $x$ que obtenemos es común o raro? ¿cuál es su función de masa de probabilidad $P(X = x) = f(x)$?

**Ejemplo (transmisión de píxeles):**

¿Cuál es la probabilidad de observar errores $X = x$ al transmitir $n = 4$ píxeles, si la probabilidad de error es $p$?

Consideremos que

1) Una variable aleatoria del **experimento de transmisión** es el vector

$$(K_1, K_2, K_3, K_4)$$

donde una observación puede ser $(K_1 = 0, K_2 = 1, K_3 = 0, K_4 = 1)$ o $(0, 1, 0, 1)$.

2) Cada

$$K_i \to Bernoulli(p)$$

$k_i \in \{0, 1\}$

3) $X = N_A$ se puede calcular como la suma

$$X = \sum_{i=1}^{4} K_i$$

$x \in \{0, 1, 2, 3, 4\}$. Por ejemplo $X = 2$ para el resultado $(0, 1, 0, 1)$.

Ahora veamos las probabilidades del número de **errores** y luego las generalizaremos.

1) ¿Cuál es la probabilidad de observar 4 **errores**?

La probabilidad de observar 4 errores es la probabilidad de observar un error en $1^{er}$ **y** $2^o$ **y** $3^o$ **y** $4^o$ píxel:

$$P(X = 4) = P(1, 1, 1, 1) = p * p * p * p = p^4$$

porque $K_i$ son **independientes**.

2) ¿Cuál es la probabilidad de observar 0 **errores**?

La probabilidad de errores 0 es la probabilidad conjunta de observar **ningún error** en **cualquier** transmisión:

$$P(X=0) = P(0,0,0,0) = (1-p)(1-p)(1-p)(1-p) = (1-p)^4$$

3) ¿Cuál es la probabilidad de observar 3 **errores**?

La probabilidad de 3 errores es la **suma** de la probabilidad de observar 3 errores en **eventos diferentes**:

$$P(X=3) = P(0,1,1,1)+P(1,0,1,1)+P(1,1,0,1)+P(1,1,1,0) = 4p^3(1-p)^1$$

porque todos estos eventos son **mutuamente excluyentes**.

4) Por lo tanto, la probabilidad de $x$ **errores** es

$$f(x) = \begin{cases} 1*p^0(1-p)^4, & x=0 \\ 4*p^1(1-p)^3, & x=1 \\ 6*p^2(1-p)^2, & x=2 \\ 4*p^3(1-p)^1, & x=3 \\ 1*p^4(1-p)^0, & x=4 \end{cases}$$

o más en breve

$$f(x) = \binom{4}{x} p^x (1-p)^{4-x}$$

para $x = 0, 1, 2, 3, 4$

donde $\binom{4}{x}$ es el número de **posibles resultados** (transmisiones de 4 píxeles) con $x$ errores.

**Definición:**

La función de **probabilidad binomial** es la función de masa de probabilidad de observar $x$ resultados de tipo $A$ en $n$ ensayos independientes de Bernoulli, donde $A$ tiene la misma probabilidad $p$ en cada ensayo.

La función está dada por

$f(x) = \binom{n}{x} p^x (1-p)^{nx}$, $x = 0, 1, ...n$

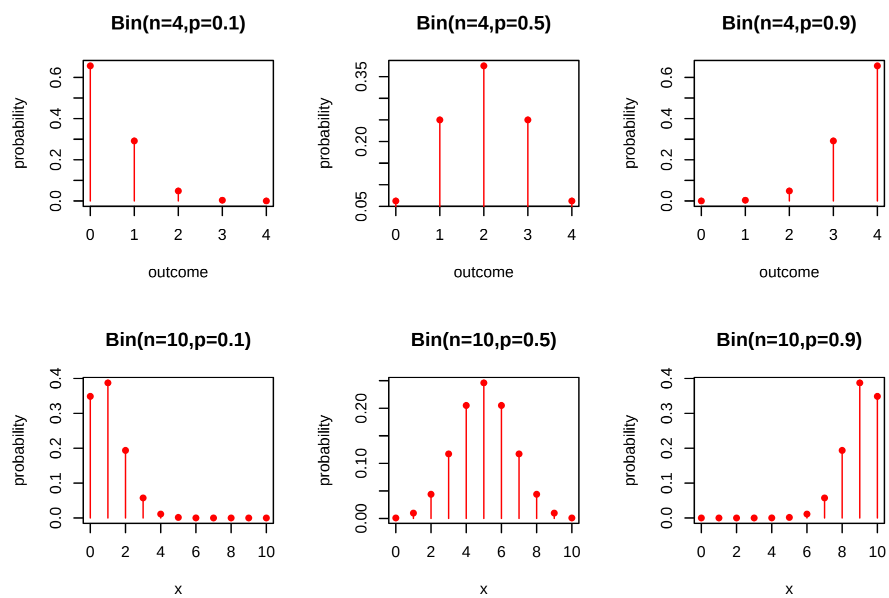
$\binom{n}{x} = \frac{n!}{x!(nx)!}$ se denomina **coeficiente binomial** y da el número de formas en que se pueden obtener $x$ eventos de tipo $A$ en un conjunto de $n$.

Cuando una variable $X$ tiene una función de probabilidad binomial decimos que se distribuye binomialmente y escribimos

$$X \rightarrow Bin(n, p)$$

donde $n$ y $p$ son parámetros.

Veamos algunas funciones de masa de probabilidad en la familia de modelos paramétricos binomiales:




**Propiedades:**

Si una variable aleatoria $X \rightarrow Bin(n, p)$ entonces

1) su media es

$$E(X) = np$$

2) su varianza es

$$V(X) = np(1-p)$$

Estas propiedades se pueden demostrar por el hecho de que $X$ es la suma de $n$ variables de Bernoulli independientes. Por lo tanto,

$E(X) = E(\sum_{i=1}^{n} K_i) = np$

y

$V(X) = V(\sum_{i=1}^{n} K_i) = n(1-p)p$

**Ejemplo (transmisión de píxeles):**

- El valor esperado para el número de errores en la transmisión de 4 píxeles es $np = 4 * 0.1 = 0.4$ cuando la probabilidad de error es 0.1.
- La varianza es $n(1-p)p = 0.36$
- ¿Cuál es la probabilidad de observar 4 errores?

Dado que estamos repitiendo una ensayo de Bernoulli $n = 4$ veces y contando el número de eventos de tipo $A$ (errores), cuando $P(A) = p = 0.1$ entonces

$$X \rightarrow Bin(n = 4, p = 0.1)$$

Eso es

$$f(x) = \binom{4}{x} 0.1^x (1-0.1)^{4-x}$$

$P(X = 4) = f(4) = \binom{4}{4} 0.1^4 0.9^0 = 0.1^4 = 10^{-4}$

En R dbinom(4,4,0.1)

- ¿Cuál es la probabilidad de observar 2 errores?

$P(X = 2) = \binom{4}{2} 0.1^2 0.9^2 = 0.0486$

En R dbinom(2,4,0.1)

**Ejemplo (encuestas de opinión):**

- ¿Cuál es la probabilidad de observar **como máximo** 8 votantes del partido de gobierno en una encuesta electoral de tamaño 10, si la probabilidad de un voto para el partido es de 0.9?

Para este caso

$$X \rightarrow Bin(n = 10, p = 0.9)$$

Eso es

$$f(x) = \binom{10}{x} 0.9^x (0.1)^{4-x}$$

Queremos calcular: $P(X \leq 8) = F(8) = \sum_{i=1..8} f(x_i) = 0.2639011$

en R pbinom(8,10, 0.9)

## 7.10 Función de probabilidad binomial negativa

Ahora imaginemos que estamos interesados en contar los píxeles bien transmitidos antes de que ocurra un **número dado** de errores. Digamos que podemos **tolerar** $r$ errores en la transmisión.

Nuestro experimento aleatorio ahora es: Repetir las pruebas de Bernoulli hasta que observemos que el resultado $A$ aparece $r$ veces.

El resultado del experimento es el número de eventos $n_B = y$

Estamos interesados en encontrar la probabilidad de observar un número particular de eventos $B$, $P(Y = y)$, donde $Y = N_B$ es la variable aleatoria.

**Ejemplo (transmisión de píxeles):**

¿Cuál es la probabilidad de observar $y$ píxeles bien transmitidos ($B$) antes de $r$ errores ($A$)?

Primero encontremos la probabilidad de **un** evento de transmisión **en particular** con $y$ número de píxeles correctos ($B$) y $r$ número de errores ($A$).

$$(0, 0, 1, ., 0, 1, ...0, 1)$$

donde consideramos que hay $y$ ceros y $r$ unos. Por lo tanto, observamos $y$ píxeles correctos en un total de $y + r$ píxeles.

La probabilidad de este evento es:

$$P(0, 0, 1, ., 0, 1, ...0, 1) = p^r(1 - p)^y$$

Recuerda que $p$ es la probabilidad de error ($A$).

¿Cuántos **eventos de transmisión** pueden tener $y$ píxeles correctos (0) antes de $r$ errores (1)?

Ten en cuenta que

1) El último pixel es fijo (marca el final de la transmisión)
2) El número total de formas en que $y$ el número de ceros se puede asignar en $y + r - 1$ píxeles es: $\binom{y+r-1}{y}$

Por lo tanto, la probabilidad de observar $y$ 1 antes de $r$ 0 (cada 1 con probabilidad $p$) es

$$P(Y = y) = f(y) = \binom{y + r - 1}{y} p^r(1 - p)^y$$

para $y = 0, 1, ...$

Entonces decimos que $Y$ sigue una distribución binomial negativa y escribimos

$$Y \to NB(r, p)$$

donde $r$ y $p$ son parámetros que representan la tolerancia y la probabilidad de un solo error (evento $A$).

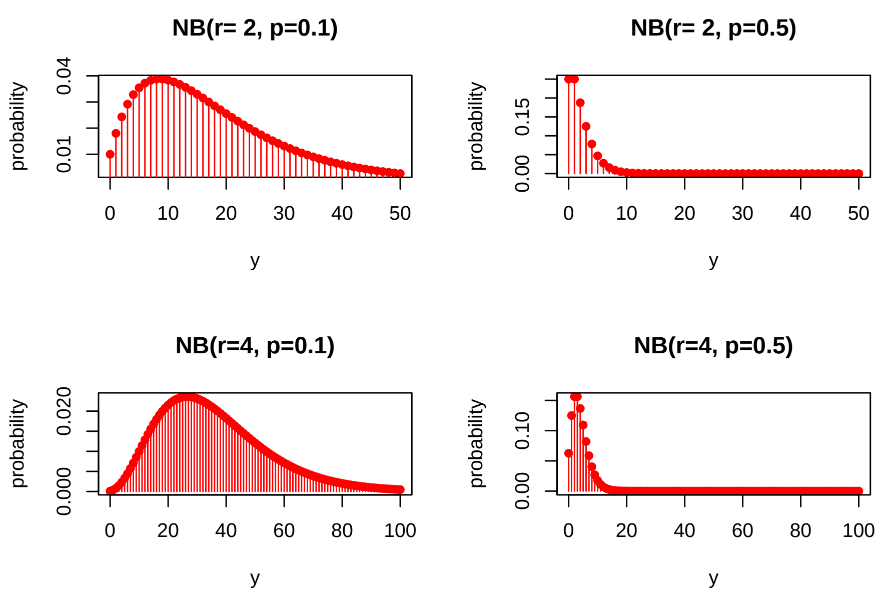
**Propiedades:**

Una variable aleatoria $Y \to NB(r, p)$ tiene

1) media
$$E(Y) = r\frac{1-p}{p}$$

2) y varianza
$$V(Y) = r\frac{1-p}{p^2}$$

Veamos algunas funciones de masa de probabilidad en la familia de modelos paramétricos binomiales negativos:




**Ejemplo (sitio web)**

Un sitio web tiene tres servidores. Un servidor opera a la vez y solo cuando falla una solicitud se usa otro servidor.

Si se sabe que la probabilidad de que falle una solicitud es $p = 0.0005$, entonces

- ¿Cuál es el número esperado de solicitudes exitosas antes de que las tres computadoras fallen?

Ya que estamos repitiendo un ensayo de Bernoulli hasta $r = 3$ se observan eventos de tipo $A$ (fallo) (cada uno con $P(A) = p = 0.0005$) y estamos contando el número de eventos de tipo $B$ (solicitudes exitosas) entonces

$$Y \to NB(r = 3, p = 0.0005)$$

Por lo tanto, el número esperado de solicitudes antes de que el sistema falle es:

$E(Y) = r\frac{1-p}{p} = 3\frac{1-0.0005}{0.0005} = 5997$

Ten en cuenta que en realidad hay pruebas de 6000.

- ¿Cuál es la probabilidad de observar 5 solicitudes exitosas antes de que el sistema falle?

$f(5) = \binom{7}{5}0.0005^3 0.9995^5 = 2.618444 \times 10^{-9}$

- ¿Cuál es la probabilidad de tratar con un máximo de 5 solicitudes exitosas antes de que el sistema falle?

En R esto se calcula con dnbinom(5,3,0.0005)

Por lo tanto, queremos calcular la distribución de probabilidad en 5:

$F(5) = P(Y \leq 5) = \Sigma_{y=0}^{5} f(y)$

$= \sum_{y=0}^{5} \binom{y+2}{y} 0.0005^r 0.9995^y$

$= \binom{2}{0}0.0005^3 0.9995^0 + \binom{3}{1}0.0005^3 0.9995^1$

$+\binom{4}{2}0.0005^3 0.9995^2 + \binom{5}{3}0.0005^3 0.9995^3$

$+\binom{6}{4}0.0005^3 0.9995^4 + \binom{7}{5}0.0005^3 0.9995^5$

$= 6.9 \times 10^{-9}$

En R esto se calcula con pnbinom(5,3,0.0005)

**Ejemplos**

- ¿Cuál es la probabilidad de observar 10 píxeles correctos antes de 2 errores, si la probabilidad de error es 0.1?

$f(10; r = 2, p = 0.1) = 0.03835463$

en R dnbinom(10, 2, 0.1)

- ¿Cuál es la probabilidad de que entren 2 chicas antes que 4 chicos entren a clase si la probabilidad de que entre un chico es de 0.55?

$f(2; r = 4, p = 0.55) = 0.1853$

en R dnbinom(2, 4, 0.55)

## 7.11 Distribución geométrica

Llamamos **distribución geométrica** a la distribución **binomial negativa** con $r = 1$

La probabilidad de observar $B$ eventos antes de observar el **primer** evento de tipo $A$ es

$$P(Y = y) = f(y) = p(1 - p)^y$$

$$Y \to Geom(p)$$

que tiene

1) media
$$E(Y) = \frac{1 - p}{p}$$

2) y varianza
$$V(Y) = \frac{1 - p}{p^2}$$

## 7.12 Modelo hipergeométrico

El **modelo hipergeométrico** surge cuando queremos contar el número de eventos de tipo $A$ que se extraen de una población finita.

El modelo general es considerar $N$ bolas totales en una urna. Marquemos $K$ con la etiqueta $A$ y $NK$ con la etiqueta $B$. Saquemos $n$ bolas una por una sin reemplazo en la urna y luego contemos cuántos $A$ obtuvimos.

El modelo **Binomial** se puede derivar del modelo **Hipergeométrico** cuando consideramos que $N$ es infinito, o que cada vez que sacamos una bola la volvemos a colocar en la urna.

**Ejemplo (varicela):**

Una escuela de $N = 600$ niños tiene una epidemia de varicela. Testamos a $n = 200$ niños y observamos que $x = 17$ dieron positivo. Si supiéramos que un total de $K = 64$ estaban realmente infectados en la escuela, ¿cuál es la probabilidad de nuestra observación?

**Definición:**

La probabilidad de obtener $x$ casos (tipo $A$) en una muestra de $n$ extraída de una población de $N$ donde $K$ son casos (tipo $A$).

$P(X = x) = P(una\, muestra) \times (Nmero\, de\, formas\, de\, obteniendo\, x)$

$$= \frac{1}{\binom{N}{n}} \binom{K}{x} \binom{NK}{nx}$$

donde $k \in \{\max(0, n + KN), ... \min(K, n)\}$

Entonces decimos que $X$ sigue una distribución hipergeométrica y escribimos

$$X \rightarrow Hipergeom(N, K, n)$$

El modelo hipergeométrico tiene tres parámetros.

**Propiedades:**

Si $X \rightarrow Hypergeometric(N, K, n)$ entonces tiene

1) media

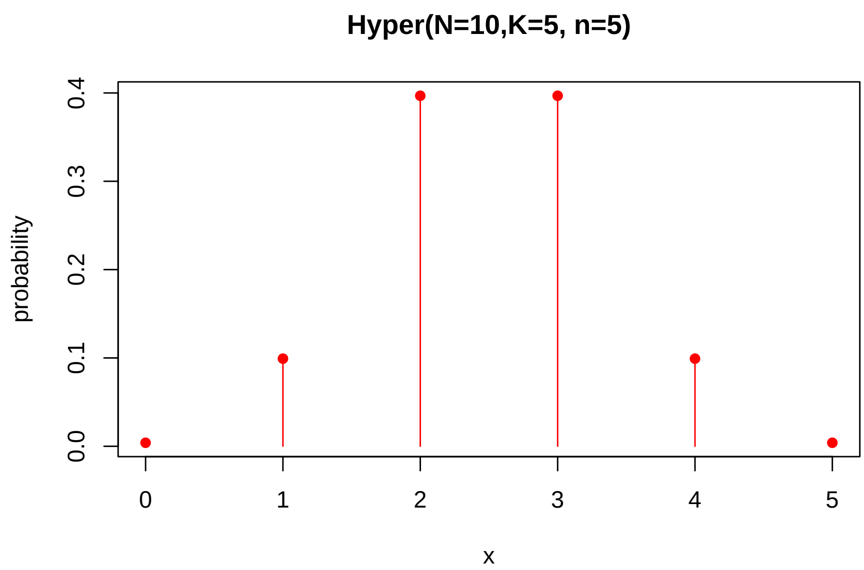
$$E(X) = n\frac{K}{N} = np$$

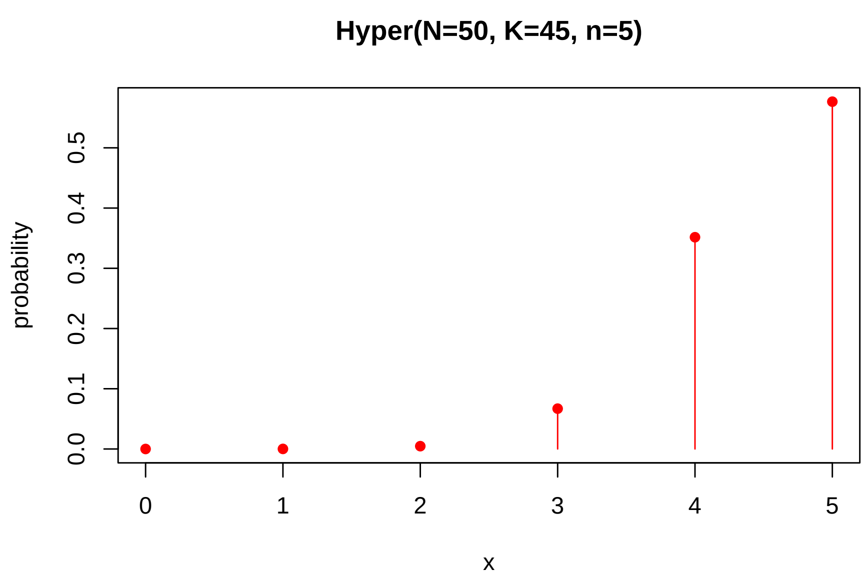
2) y varianza

$$V(X) = np(1-p)\frac{Nn}{N-1}$$

cuando $p = \frac{K}{N}$ es la proporción de casos ($A$) en una población de tamaño $N$. Ten en cuenta que cuando $N \rightarrow \infty$ recuperamos las propiedades binomiales.

Veamos algunas funciones de masa de probabilidad en la familia de modelos paramétricos hipergeométricos:




**Ejemplo (varicela):**

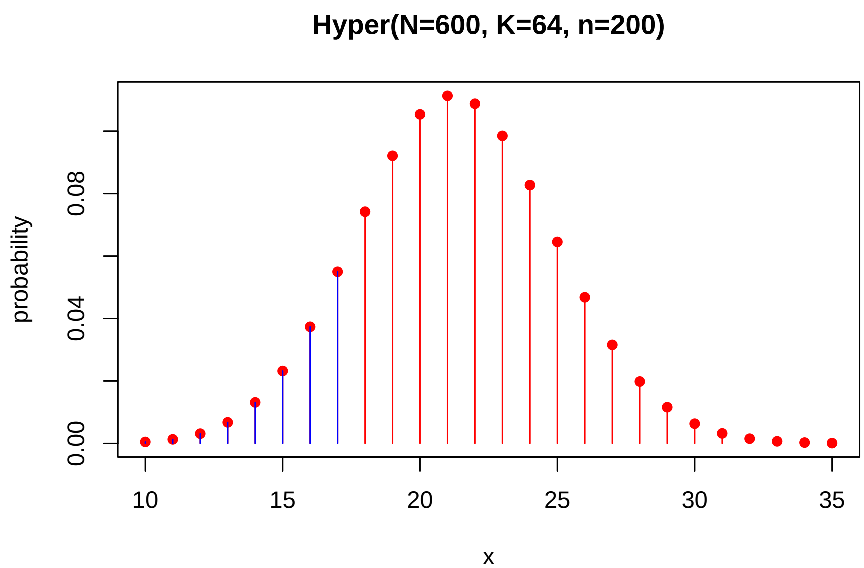
- ¿Cuál es la probabilidad ver como mucho 17 casos de varicela en en una muestra de 200 alumnos de una ecuela de 600 alumnos donde 64 están infectados?

La probabilidad que necesitamos calcular es $P(X \leq 17) = F(17)$

donde $X \to Hypergeometric(N = 600, K = 64, n = 200)$

en R phyper(17, 64, 600-64, 200)=0.140565

La solución es la adición de las agujas azules en el gráfico.




## 7.13 Preguntas

**1)** ¿Cuál es el valor esperado y la varianza del número de fallos en 100 prototipos, cuando la probabilidad de un fallo es de 0.25?

**a:** 0.25, 0.1875; **b:** 25, 0.1875; **c:** 0.25, 18.75; **d:** 25, 18.75

**2)** ¿Qué modelo de probabilidad describe mejor el número de mesas disponibles a la hora de la cena en un restaurante?

**a:** Binomial; **b:** Uniforme; **c:** Binomial negativo; **d:** Hipergeométrico

**3)** El valor esperado de una distribución Binomial no es

**a:** $n$ veces el valor esperado de un Bernoulli; **b:** el valor esperado de un Hipergeométrico, cuando la población es muy grande; **c:** $np$; **d:** el límite de la frecuencia relativa cuando el número de repeticiones es grande

**4)** Las encuestas de opinión para las elecciones de EE. UU. dan una probabilidad de 0.55 de que un votante esté a favor del partido republicano. Si realizamos nuestra propia encuesta y preguntamos a 100 personas al azar en la calle, ¿cómo calcularías la probabilidad de que en nuestra encuesta los demócratas ganen las elecciones?

**a:**pbinom(x=49, n=100, p=0.55)=0.13; **b:**1-pbinom(x=49, n=100, p=0.55)=0.86; **c:**pbinom(x=51, n=100, p=0.45)=0.90; **d:**1-pbinom(x=51, n=100, p=0.45)=0.095

**5)** En un examen un alumno cuando no sabe la respuesta elige al azar una de las cuatro respuestas en una pregunta de selección múltiple. Si no sabe 10 preguntas, ¿cuál es la probabilidad de que al menos 5 preguntas sean correctas?

**a:**dbinom(x=4, n=10, p=0.25); **b:**pbinom(x=4, n=10, p=0.75); **c:**dbinom(x=4, n=10, p=0.75); **d:**1-pbinom(x=4, n=10, p=0.25)

## 7.14 Ejercicios

#### 7.14.0.1 Ejercicio 1

Si el 25% de los tornillo producidos por una máquina son defectuosos, determina la probabilidad de que, de 5 tornillos elegidos al azar

- ningún tornillo sea defectuoso (R:0.2373)
- 1 tornillo sea defectuoso (R:0.3955)
- 2 tornillos sean defectuosos (R:0.2636)
- como máximo 2 tornillos sean defectuosos (R:0.8964)

#### 7.14.0.2 Ejercicio 2

En una población, la probabilidad de que nazca un niño es $p = 0.51$. Considera una familia de 4 hijos.

- ¿Cuál es la probabilidad de que una familia tenga un solo niño? (R: 0.240)
- ¿Cuál es la probabilidad de que una familia tenga una sola niña? (R: 0.259)
- ¿Cuál es la probabilidad de que una familia tenga solo un niño o solo una niña? (R: 0.4999)
- ¿Cuál es la probabilidad de que la familia tenga por mucho dos niños? (R: 0.6723)
- ¿Cuál es la probabilidad de que la familia tenga al menos dos niños? (R: 0.7023)
- ¿Cuál es el mínimo número de hijos que debe tener una familia para que la probabilidad de tener al menos una niña sea mayor a 0.75?(R:$n = 3 \geq \log(0.25)/\log(0.51) = 2.05$)

#### 7.14.0.3 Ejercicio 3

Un motor de búsqueda falla al recuperar información con una probabilidad de 0.1

- Si nuestro sistema recibe 50 solicitudes de búsqueda, ¿cuál es la probabilidad de que el sistema no responda a tres de ellas? (R: 0.1385651)
- ¿Cuál es la probabilidad de que el motor complete con éxito 15 búsquedas antes del primer fallo? (R:0.020)
- Consideramos que un motor de búsqueda funciona suficientemente bien cuando es capaz de encontrar información como mínimo para 10 solicitudes por cada 2 fallos. ¿Cuál es la probabilidad de que en un ensayo de fiabilidad nuestro buscador sea satisfactorio? (R: 0.659)

# Chapter 8

# Modelos de Poisson y Exponencial

## 8.1 Objetivo

En este capítulo veremos dos modelos de probabilidad estrechamente relacionados: los modelos **Poisson** y **exponencial**.

El modelo de Poisson es para variables aleatorias discretas, mientras que la función exponencial es para variables aleatorias **continuas**

## 8.2 Modelos de probabilidad para variables discretas

En el capítulo anterior construimos modelos complejos a partir de modelos simples. En cada etapa, introdujimos algún concepto novedoso:

**Uniforme**: interpretación clásica de la probabilidad ↓ **Bernoulli**: Introducción de un **parámetro** $p$ (familia de modelos) ↓ **Binomial**: Introducción a la **Repetición** de un experimento aleatorio ($n$ ensayos de Bernoulli) ↓ **Poisson**: Repetición de un experimento aleatorio dentro de un intervalo continuo, sin **control** sobre cuándo/dónde ocurre el ensayo de Bernoulli.

El último modelo es para procesos de Poisson que describen la repetición de un experimento aleatorio con la aleatoriedad adicional del momento en que la repeticiones se producen.

## 8.3 Experimento de Poissson

Imagina que estamos observando eventos que **dependen** de **intervalos** de tiempo o distancia.

Por ejemplo:

- coches que llegan a un semáforo
- mensajes que recibimos en el teléfono móvil
- impurezas que ocurren al azar en un alambre de cobre

Supongamos que los eventos son resultados de ensayos de Bernoulli **independientes**, cada uno de los cuales aparece aleatoriamente en un intervalo continuo, y queremos **contarlos**.

¿Cuál es la probabilidad de observar $X$ eventos en una unidad de intervalo (tiempo o distancia)?

**Ejemplo (Impurezas en un alambre):**

Imaginemos que algunas impurezas se depositan al azar a lo largo de un cable de cobre. Quieremos contar el número de impurezas en un centímetro de alambre ($X$).

Considera que sabemos que en promedio hay 10 impurezas por centímetro $\lambda = 10/cm$.

¿Cuál es la probabilidad de observar $X = 5$ impurezas en una muestra de un centímetro en particular?

## 8.4 Función de masa de probabilidad de Poisson

Para calcular la función masa de probabilidad $f(x) = P(X = x)$ del ejemplo anterior dividimos el centímetro en micrómetros ($0.0001cm$).

Los micrómetros son lo suficientemente pequeños como para

1) que hay ao no haya una impureza en cada micrómetro

2) cada micrómetro se pueda considerar como un **ensayo de Bernoulli**

**De la función binomial a la función de probabilidad de Poisson**

La probabilidad de observar $X$ impurezas en $n = 10,000\mu$ (1cm) sigue aproximadamente una distribución binomial

$f(x) \sim \binom{n}{x} p^x (1-p)^{nx}$

donde $p$ es la probabilidad de encontrar una impureza en un micrómetro.

Dado que el valor esperado de una variable Binomial es $E(X) = np$. Este es el número promedio de impurezas por 1 cm o $\lambda = np$. Por lo tanto, sustituimos $p = \lambda/n$

$$f(x) \sim \binom{n}{x}(\frac{\lambda}{n})^x(1-\frac{\lambda}{n})^{nx}$$

Dado que **podría** haber todavía dos impurezas en un micrómetro, necesitamos aumentar la partición del alambre y $n \to \infty$.

Luego en el límite:

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Donde $\lambda$ es constante porque es la densidad de impurezas por centímetro, una **propiedad física** del sistema. $\lambda$ es por lo tanto el **parámetro** del modelo de probabilidad.

**Detalles de la derivación:**

Para $f(x) \sim \binom{n}{x}(\frac{\lambda}{n})^x(1-\frac{\lambda}{n})^{nx}$

en el límite ($n \to \infty$)

1) $\frac{1}{n^x}\binom{n}{x} = \frac{1}{n^x}\frac{n!}{x!(n-x)!} = \frac{(n-x)!(n-x+1)...(n-1)n}{n^x x!(n-x)!} = \frac{n(n-1)..(n-x+1)}{n^x x!} \to \frac{1}{x!}$
2) $(1-\frac{\lambda}{n})^n \to e^{-\lambda}$ (definition of exponential)
3) $(1-\frac{\lambda}{n})^{-x} \to 1$

Poniendo todo junto entonces:

$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$

**Definición**

Dado

1) un intervalo en los números reales
2) hay eventos que ocurren al azar en el intervalo
3) se conoce el número promedio de eventos en el intervalo ($\lambda$)
4) si se puede encontrar una pequeña partición regular del intervalo tal que en cada partición la podamos considerar como un ensayo de Bernoulli.

Entonces, la variable aleatoria $X$ que cuenta eventos a lo largo del intervalo es una variable **Poisson** con función de masa de probabilidad

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \lambda > 0$$

**Propiedades:** Cuando $X \to Poiss(\lambda)$ tiene

1) media

$$E(X) = \lambda$$

2) y varianza
$$V(X) = \lambda$$

**Ejemplos**

1) ¿Cuál es la probabilidad de recibir 4 correos electrónicos en una hora, cuando el promedio de correos electrónicos en una hora es de 1?

Tenemos que la varible es de Poisson: $X \to Poiss(\lambda)$ con $\lambda = 1$ y su función de masa de probabilidad es:

$$f(x) = \frac{e^{-1}1^x}{x!}$$

Por lo tanto la probabilidad de que la variable tome valor 4 es $P(X = 4)$:

$f(4; \lambda = 1) = \frac{e^{-1}1^4}{4!} = 0.01532831$

in R dpois(4,1)

2) ¿Cuál es la probabilidad de recibir 4 correos electrónicos en **tres horas**, cuando el promedio de correos electrónicos en una hora es de 1?

La unidad sobre la cual hacemos los conteos ha cambiado de 1 hora a 2 horas, por lo tanto tenemos que **re-escalar** $\lambda$. Si antes el promedio de correos era $\lambda = 1$ en una hora, el promedio de correos en tres horas es ahora 3: $\lambda_{3h} = 3\lambda_{1h} = 3*1 = 3$

Tenemos que la varible es de Poisson: $X \to Poiss(\lambda_{3h})$ con $\lambda_{3h} = 3$ y su función de masa de probabilidad es:

$$f(x) = \frac{e^{-3}3^x}{x!}$$

Por lo tanto la probabilidad de que la variable tome valor 4 es $P(X = 4)$:

$f(4; \lambda = 3) = \frac{e^{-3}3^4}{4!} = 0.1680314$

in R dpois(4,3)

3) ¿Cuál es la probabilidad de contar **al menos** 10 automóviles que llegan a un peaje en un minuto, cuando el promedio de automóviles que llegan a un peaje en un minuto es de 5;

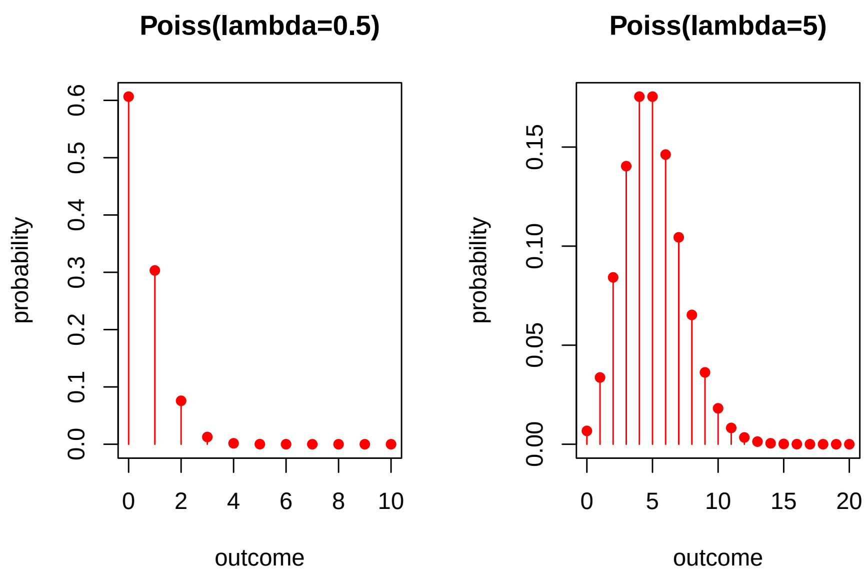
Tenemos que la varible es de Poisson: $X \to Poiss(\lambda)$ con $\lambda = 5$ y su función de masa de probabilidad es:

$$f(x) = \frac{e^{-5}5^x}{x!}$$

$P(X \leq 10) = F(10; \lambda = 5) = \sum_{x=0,\ldots 10} f(x; \lambda = 5) = 0.9863047?$

en R ppois(10,5)

Veamos algunas funciones de masa de probabilidad en la familia de modelos paramétricos de Poisson:




## 8.5 Modelos de probabilidad para variables continuas

Los modelos de probabilidad para variables continuas son **funciones de densidad** de probabilidad $f(x)$ que **creemos** describen experimentos aleatorios reales.

La función de densidad de probabilidad $f(x)$

1) es positiva

$$f(x) \geq 0$$

2) nos permite calcular probabilidades usando el área bajo la curva:

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

3) es tal que la probabilidad de que obtengamos cualquier resultado es 1:

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

## 8.6 Experimento exponencial

Volvamos a un **proceso de Poisson** definido por la probabilidad

$$f(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \lambda > 0$$

para el número de eventos ($k$) en un intervalo.

Consideremos ahora que estamos interesados en la duración/tiempo que debemos esperar hasta que ocurra el **primer** conteo.

Podemos preguntarnos por la probabilidad de que el primer evento ocurra después de la duración/tiempo $X$.

Por lo tanto, dado que $X$ es una variable aleatoria **continua**, buscamos su función de densidad de probabilidad $f(x)$.

## 8.7 Densidad de probabilidad exponencial

La probabilidad de 0 eventos **si** un intervalo tiene unidad $x$ (rescalando como en el ejemplo 2) es

$$f(0|x) = \frac{e^{-x\lambda}(x\lambda)^0}{0!}$$

o

$$f(0|x) = e^{-x\lambda}$$

Podemos tratar esto como la probabilidad condicional de 0 eventos dada una distancia $x$: $f(K = 0|X = x)$ y aplicar el teorema de Bayes para invertirlo:

$$f(x|0) = Cf(0|x) = Ce^{-x\lambda}$$

Esta es la **probabilidad de observar una distancia** $x$ para 0 eventos. Esta es la distancia hasta el primer evento.

**Definición**

En un proceso de Poisson con parámetro $\lambda$ la probabilidad de esperar una distancia/tiempo $X$ entre dos conteos viene dada por la **densidad de probabilidad**

$$f(x) = Ce^{-x\lambda}$$

- $C$ es una constante que asegura: $\int_{-\infty}^{\infty} f(x)dx = 1$

- por integración $C = \lambda$

Por lo tanto:

$$f(x) = \lambda e^{-\lambda x}, x \geq 0$$

$\lambda$ es el parámetro del modelo, también conocido como **tasa de decaimiento**.

**Propiedades:**

Cuando $X \rightarrow Exp(\lambda)$ entonces

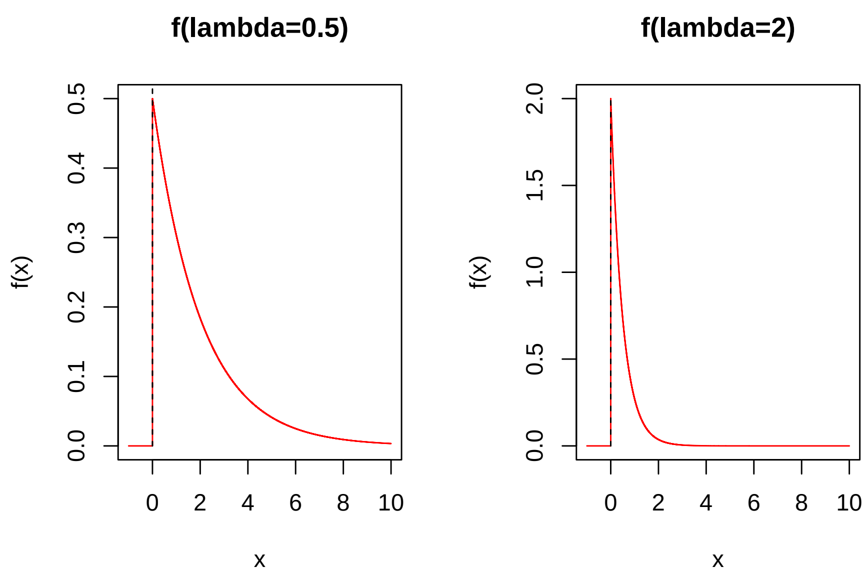
1) tiene media

$$E(X) = \frac{1}{\lambda}$$

2) y varianza

$$V(Y) = \frac{1}{\lambda^2}$$

Veamos un par de densidades de probabilidad en la familia exponencial




## 8.8 Distribución exponencial

Consideremos las siguientes preguntas:

1) En un proceso de Poisson ¿Cuál es la probabilidad de observar un intervalo **menor** que $a$ hasta el primer evento?

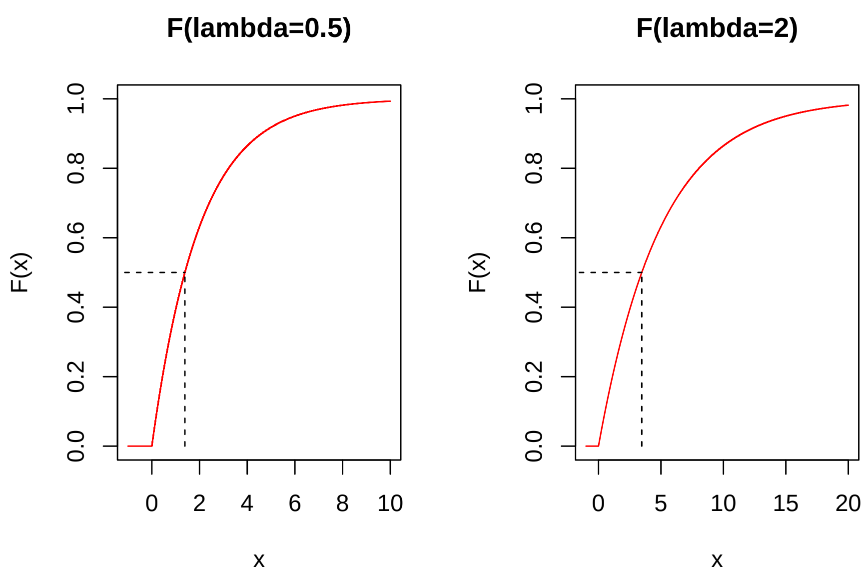
Recuerda que esta probabilidad $F(a) = P(X \leq a)$ es la densidad de probabilidad

$$F(a) = \lambda \int_{\infty}^{a} e^{-x\lambda} dx = 1 - e^{-a\lambda}$$

2) En un proceso de Poisson ¿Cuál es la probabilidad de observar un intervalo **mayor** que $a$ hasta el primer evento?

$$P(X > a) = 1 - P(X \leq a) = 1 - F(a) = e^{-a\lambda}$$

Veamos un par de distribuciones exponenciales de la familia exponencial




La mediana $x_m$ es tal que $F(x_m) = 0.5$. Eso es $x_m = \frac{\log(2)}{\lambda}$

**Ejemplos**

1) ¿Cuál es la probabilidad de que tengamos que esperar un bus por más de 1 hora cuando en promedio hay dos buses por hora?

$$P(X > 1) = 1 - P(X \leq 1) = 1 - F(1, \lambda = 2) = 0.1353353$$

En R 1-pexp(1,2)

2) ¿Cuál es la probabilidad de tener que esperar menos de 2 segundos para detectar una partícula cuando la tasa de desintegración radiactiva es de 2 partículas por segundo; $F(2, \lambda = 2)$

$$P(X \leq 2) = F(2, \lambda = 2) = 0.9816844$$

En R pexp(2,2)

## 8.9 Preguntas

**1)** Durante la Segunda Guerra Mundial, en un día de bombardeo sobre Londres, el valor eperado de que cayera una bomba en $1.5km^2$ era de 0.92. La probabilidad de que en Hyde Park, de área aproximadamente $1.5km^2$, cayeran como mucho dos bombas era de

**a:**1-ppois(x=2, lambda=0.92); **b:**ppois(x=2, lambda=0.92); **c:**1-dpois(x=2, lambda=0.92); **d:**dpois(x=2, lambda=0.92)

**2)** La probabilidad de que un pasajero tenga que esperar menos de 20 minutos hasta que llegue el próximo taxi a su parada está mejor descrita por

**a:** Un modelo de Poisson sobre el número de taxis que pasan cada 20 minutos; **b:** Una distribución exponencial con $\lambda = 1/20$ ; **c:** Un modelo binomial que cuenta el número de taxis cada 20 minutos **d:** Una distribución uniforme entre 0 y 20 minutos;

**3)** A partir de la distribución de probabilidad exponencial de la siguiente figura, ¿cuál es el valor más posible de la mediana?

**a:** 2; **b:** 3; **c:** 4; **d:** 5

## 8.10 Ejercicios

#### 8.10.0.1 Ejercicio 1

El promedio de llamadas telefónicas por hora que ingresan a la centralita de una empresa es de 150. Encuentra la probabilidad de que durante un minuto en particular haya

- 0 llamadas telefónicas (R:0.082)
- 1 llamada telefónica (R:0.205)
- 4 o menos llamadas (R:0.891)
- más de 6 llamadas telefónicas (R:0.0141)

#### 8.10.0.2 Ejercicio 2

La cantidad promedio de partículas radiactivas que golpean un contador Geiger en una planta de energía nuclear bajo control es de 2.3 por minuto.

- ¿Cuál es la probabilidad de contar exactamente 2 partículas en un minuto? (R:0.265)
- ¿Cuál es la probabilidad de detectar exactamente 10 partículas en 5 minutos? (R:0.112)
- ¿Cuál es la probabilidad de al menos un conteo en dos minutos? (R:0.9899)
- ¿Cuál es la probabilidad de tener que esperar menos de 1 segundo para detectar una partícula radiactiva, después de encender el detector? (R:0.037)
- Sospechamos que una planta nuclear tiene una fuga radiactiva si esperamos menos de 1 segundo para detectar una partícula radiactiva, después de encender el detector. ¿Cuál es la probabilidad de que cuando visitemos 5 plantas que están bajo control, sospechemos que al menos una tiene una fuga? (R:0.1744).

# Chapter 9

# Distribución normal

## 9.1 Objetivo

En este capítulo introduciremos la distribución de probabilidad normal.

Hablaremos de su origen y de sus principales propiedades.

## 9.2 Historia

En 1801, Gauss analizó los datos obtenidos sobre la posición de Ceres, un gran asteroide entre Marte y Júpiter.

En ese momento, la gente sospechaba que era un nuevo planeta, ya que se movía día a día contra las estrellas fijas. En enero, se podía ver en el horizonte justo antes del amanecer. Sin embargo, a medida que pasaban los días, Ceres salía cada vez más tarde hasta que ya no se pudo ver más debido a la salida del Sol.

Gauss entendió que las medidas para la posición de Ceres tenían errores.

Por lo tanto, estaba interesado en descubrir cómo se **distribuían** las observaciones para poder encontrar la órbita más **probable**. Con la órbita, podía derivar la masa del objeto y luego decidir si era un planeta o sólo un gran asteroide.

Los datos estaban disponibles sólo para el mes de enero. Después de lo cual Ceres desaparecería. Quería **predecir** hacia dónde deberían apuntar los astrónomos sus telescopios para encontrarlo seis meses después al anochecer, una vez que hubiera pasado por detrás del Sol.

Gauss tuvo que dar cuenta de los errores en la posición de ceres en un día determinado debido a la medición

Gauss supuso que

1) los errores pequeños eran más probables que los errores grandes

2) el error a una distancia $-\epsilon$ del varlor en la posición de Ceres era igualmente probable que una distancia $\epsilon$

3) la pisición más **verosimil** (que nos creemos más) de Ceres en un momento dado en el cielo era el **promedio** de múltiples mediciones de altitud en esa latitud.

Eso fue suficiente para mostrar que las desviaciones de las observaciones $y$ **de la órbita** satisfacian la ecuación

$$\frac{df(y)}{dy} = -Cyf(y)$$

con $C$ una costante positiva. La solución de esta ecuación diferencial es:

$$f(y) = \frac{\sqrt{C}}{\sqrt{2\pi}} e^{-\frac{Cy^2}{2}}$$

*The evolution of the normal distribution, Saul Stahl, Mathemetics Magazine, 2006.

## 9.3 Densidad normal

Densidad de probabilidad de Gauss da la distribución de los errores de medición desde la posición **real** pero **desconocida** de Ceres en el cielo. Hagamos un par de cambios en la función.

1- Escribamos la densidad de errores desde el horizonte usando la variable aleatoria $X$, o sea $y = x - \mu$. $\mu$ es la posición **real** pero **desconocida** de Ceres desde el horizonte. Después de un cambio de variable encontramos la función de densidad de probabilidad:

$$f(x) = \frac{\sqrt{C}}{\sqrt{2\pi}} e^{-C(x-\mu)^2}$$

2) Cambiemos de nombre la variable $C$ por $\frac{1}{\sigma^2}$

Entonces, llegamos a la siguente definición.

## 9.4 Definición

Una variable aleatoria $X$ definida en los números reales tiene una densidad **Normal** si toma la forma

$$f(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x\in\mathbb{R}$$

La variable tiene

1) media

$$E(X)=\mu$$

que para Gauss representaba la posición real de Ceres.

2) y varianza

$$V(X)=\sigma^2$$

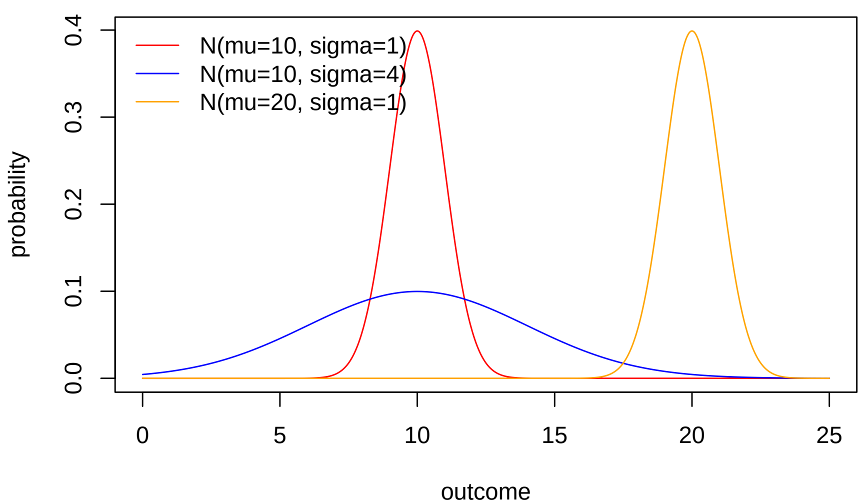
que representaba la dispersión del error en las observaciones, que depende de la calidad del telescopio.

$\mu$ y $\sigma$ son los **dos parámetros** que describen completamente la función de densidad normal y su **interpretación** depende del experimento aleatorio.

Cuando $X$ sigue una densidad Normal, es decir, se distribuye normalmente, escribimos

$$X\rightarrow N(\mu,\sigma^2)$$

Veamos algunas densidades de probabilidad en el modelo paramétrico normal



## 9.5 Distribución de probabilidad

La distribución de probabilidad de la densidad Normal:

$$F(a) = P(Z \leq a)$$

es la función de **error** definida por el área bajo la curva de $-\infty$ a $a$

$$F(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

La función se encuentra en la mayoría de los programas de computadora y no tiene una forma cerrada de funciones conocidas.

**Ejemplo (altura de mujeres)**

1) ¿Cuál es la probabilidad de que una mujer de la población tenga una altura máxima de $150cm$ si las mujeres tienen una altura media de $165cm$ con una desviación estándar de $8cm$?

$P(X \leq 150) = F(150, \mu = 165, \sigma = 8) = 0.03039636$

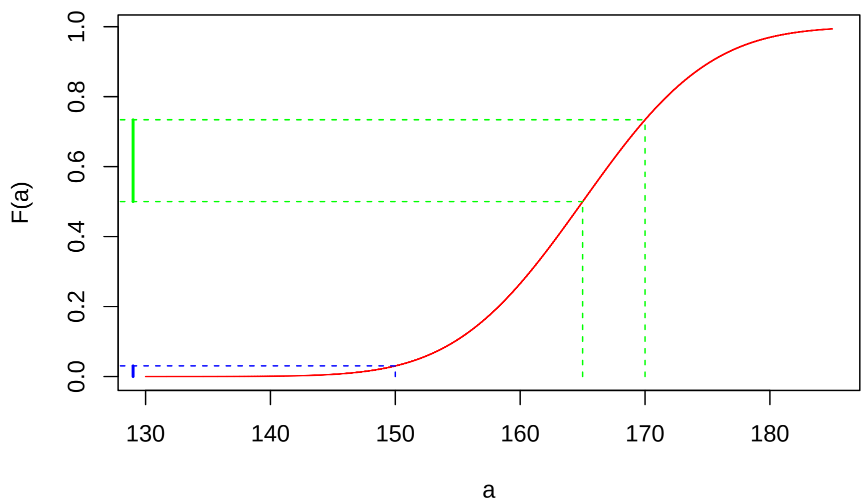
en R pnorm(150, 165, 8)

2) ¿Cuál es la probabilidad de que la altura de una mujer en la población esté entre $165cm$ y $170cm$?

$P(165 \leq X \leq 170) = F(170, \mu = 165, \sigma = 8) - F(165, \mu = 165, \sigma = 8) = 0.2340145$

en R pnorm(170, 165, 8)-pnorm(165, 165, 8)

Veamos la función de distribución de probabilidad




3) ¿Cuál es el primer cuartíl para altura de las mujeres?

El primer cuartíl se define como:

$F(x_{0.25}, \mu = 165, \sigma = 8) = 0.25$

o

$x_{0.25} = F^{-1}(0.25, \mu = 165, \sigma = 8) = 159.6041$

en R qnorm(0.25, 165, 8)

**Propiedades de la distribución Normal**

1) la media $\mu$ es también la mediana ya que divide las medidas en dos
2) Los valores de $x$ que caen más allá de $2\sigma$ se consideran **raros** 5%
3) Los valores de $x$ que caen más allá de $3\sigma$ se consideran **extremadamente raros** 0.2%

**Ejemplo (altura de mujeres)**

Podemos definir los límites de **observaciones comunes** para la distribución de la altura de las mujeres en la población.

1) a una distancia de una desviación estándar de la media, encontramos 68% de la población

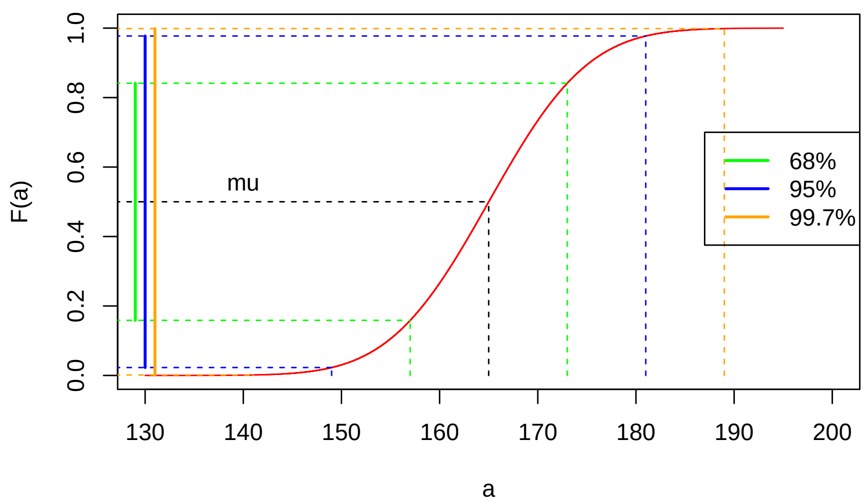
$$P(165 - 8 \leq X \leq 165 + 8) = P(157 \leq X \leq 173) = F(173) - F(157) = 0.68$$

2) a una distancia de dos desviaciones estándar de la media, encontramos 95% de la población

$$P(165 - 2 \times 8 \leq X \leq 165 + 2 \times 8) = F(181) - F(149) = 0.95$$

3) a una distancia de tres desviaciones estándar de la media, encontramos 99.7% de la población

$$P(165 - 3 \times 8 \leq X \leq 165 + 3 \times 8) = F(189) - F(141) = 0.997$$




## 9.6 Densidad normal estándar

La densidad normal estándar es la densidad particular de la familia normal

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

Por lo tanto, es la densidad con

1) media

$$E(X) = \mu = 0$$

2) y varianza

$$V(X) = \sigma^2 = 1$$

Cuando una variable aleatoria sigue una densidad de probabilidad normal, decimos que se distribuye normalmente y escribimos

$$X \to N(0, 1)$$

**Estandarización:**

Todas las variables normales se pueden **estandarizar**. Esto significa que si $X \to N(\mu, \sigma^2)$, entonces podemos transformar la variable a una **variable estandarizada**

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

que tendrá densidad:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Por lo tanto, para cualquier $X \to N(\mu, \sigma^2)$

$$Z = \frac{X - \mu}{\sigma} \to N(0, 1)$$

Puedes demostrar esto reemplazando $x = \sigma z + \mu$ y $dx = \sigma dz$ en la expresión de probabilidad que tenemos

$$P(x \leq X \leq x + dx) = P(z \leq Z \leq z + dz)$$

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

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

## 9.7 Distribución estándar

La distribución estándar es:

$$\phi(a) = F_{N(0,1)}(a) = P(Z \leq a)$$

es la función **error** definida por

$$\phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

Debido a que la distribución estándar es especial y aparecerá con frecuencia, usamos la letra $\phi$ para ello.

Puedes encontrarla en la mayoría de los programas de computadora. En R es pnorm(x) con los parámetros predeterminados, 0 y 1.

Normalmente definimos los límites de las **observaciones más comunes** para la variable estándar

1) El rango intercuartílico

$$P(-0.67 \leq X \leq 0.67) = 0.50$$

en R: c(qnorm(0.25), qnorm(0.75))

2) El rango del 95%

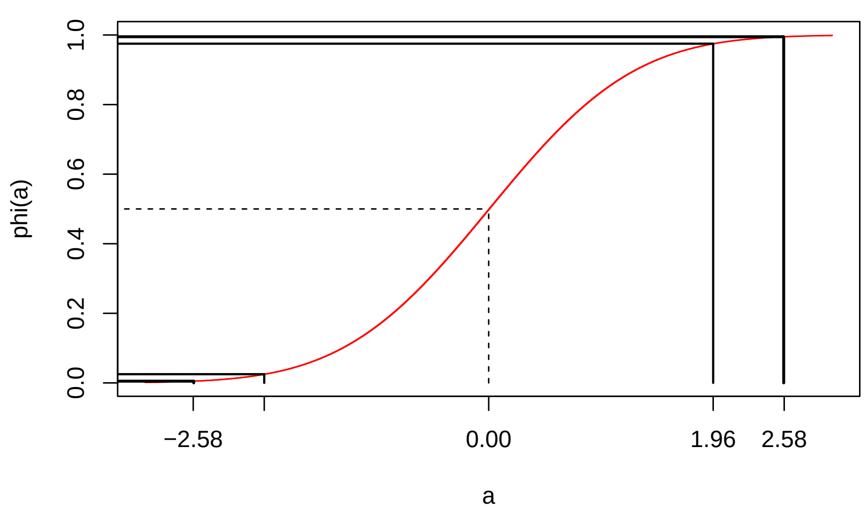
$$P(-1.96 \leq X \leq 1.96) = 0.95$$

en R: c(qnorm(0.025), qnorm(0.975))

3) El rango del 99%

$$P(-2.58 \leq X \leq 2.58) = 0.99$$

en R: c(qnorm(0.005), qnorm(0.995))



La probabilidad de cualquier variable normal $X \to N(\mu, \sigma^2)$ usando la distribución estándar

$F(a) = P(X < a) = P(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma})$

$$= P(Z < \frac{a-\mu}{\sigma}) = \phi(\frac{a-\mu}{\sigma})$$

Para calcular $P(a \leq X \leq b)$, usamos la propiedad de las distribuciones de probabilidad

$F(b) - F(a) = P(X \leq b) - P(X \leq a)$

$$= \phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})$$

## 9.8 Resumen de modelos de probabilidad

| Modelo | X | rango de x | f(x) | E(X) | V(X) |
|---|---|---|---|---|---|
| Uniforme | número entero o real | $[a, b]$ | $\frac{1}{n}$ | $\frac{b+a}{2}$ | $\frac{(b-a+1)^2-1}{12}$ |
| Bernoulli | evento A (1) | 0,1 | $(1-p)^{1-x}p^x$ | $p$ | $p(1-p)$ |
| Binomial | # de eventos A en $n$ repeticiones de ensayos de Bernoulli | 0,1,... | $\binom{n}{x}(1-p)^{nx}p^x$ | $np$ | $np(1-p)$ |
| Binomial negativo para eventos | # de eventos B (0) en repeticiones de Bernoulli antes de $r$ eventos A (1) | 0,1,.. | $\binom{x+r-1}{x}(1-p)^x p^r$ | $\frac{r(1-p)}{p}$ | $\frac{r(1-p)}{p^2}$ |

| Modelo | X | rango de x | f(x) | E(X) | V(X) |
|---|---|---|---|---|---|
| Hipergeométrico | # de eventos A en una muestra $n$ de la población $N$ con $K$ numero de As | $\max(0, n + KN)$, ... $\min(K, n)$ | $\frac{1}{\binom{N}{n}}\binom{K}{x}\binom{N-K}{n-x}$ | $n * \frac{N}{K}$ | $n\frac{N}{K}(1 - \frac{N}{K})\frac{Nn}{N-1}$ |
| Poisson | # de eventos A en un intervalo | 0,1, .. | $\frac{e^{-\lambda}\lambda^x}{x!}$ | $\lambda$ | $\lambda$ |
| Exponencial | Intervalo entre dos eventos A | $[0, \infty)$ | $\lambda e^{-\lambda x}$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |
| Normal | medida con errores simétricos cuyo valor más probable es la media | $(-\infty, \infty)$ | $\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ | $\mu$ | $\sigma^2$ |

## 9.9 Funciones R de modelos de probabilidad

| modelo | R f(x) | R F(x) |
|---|---|---|
| Uniforme (continuo) | dunif(x, a, b) | punif(x, a, b) |
| Binomial | dbimon(x,n,p) | pbimon(x,n,p) |
| Binomial negativo para eventos | dnbinom(x,r,p) | pnbinom(x,r,p) |
| Hipergeométrico | dhyper(x, K, N-K, n) | phyper(x, K, N-K, n) |
| Poisson | dpois(x, lambda) | ppois(x, lambda) |
| Exponencial | dexp(x, lambda) | pexp(x, lambda) |
| normales | dnomr(x, mu, sigma) | pnomr(x, mu, sigma) |

## 9.10 Preguntas

**1)** Para una variable normal estándar

**a:** 50% de sus observaciones están entre $(-0.67, 0.67)$; **b:** 2% de sus observaciones son inferiores a $-2.58$; **c:** 5% de sus observaciones son superiores a 1.96; **d:** 25% de sus observaciones están entre $(-1.96, -0.67)$

**2)** si sabemos que $\phi(-0.8416212) = 0.2$ entonces que es $\phi(0.8416212)$

**a:** 0.1; **b:** 0.2; **c:** 0.8; **d:** 0.9

**3)** el tercer cuartil de una variable normal con media 10 y desviación estándar 2 es

**a:** qnorm(1/3, 10, 2)=9.138545; **b:** qnorm(1-0.75, 10, 2)=8.65102 ; **c:** qnorm(1-1/3, 10, 2)=10.86145 ; **d:** qnorm(0.75, 10, 2)= 11.34898

**4)** la probabilidad de que una variable normal con media 10 y desviación estándar 2 esté en $(-\infty, 10)$ es

**a:** 0.25; **b:** 0.5; **c:** 0.75; **d:** 1:

**5)** No es cierto que para una variable normal estándar

**a:** su media y mediana son iguales; **b:** la distribución de probabilidad estándar se puede utilizar para calcular sus probabilidades; **c:** su rango intercuartílico es el doble de su desviación estándar; **d:** 5% de sus observaciones están a una distancia desde 0 mas grande que su desviación estándar

## 9.11 Ejercicios

#### 9.11.0.1 Ejercicio 1

Encuentra el área bajo la curva normal estándar en los siguientes casos:

- Entre $z = 0.81$ y $z = 1.94$ (R:0.182)
- A la derecha de $z = -1.28$ (R:0.899)
- A la derecha de $z = 2.05$ o a la izquierda de $z = -1.44$ (R:0.0951)

#### 9.11.0.2 Ejercicio 2

- ¿Cuál es la probabilidad de que la altura de un hombre sea al menos 165cm si la media poblacional es 175cm y la desviación estándar es 10cm? (R:0.841)
- ¿Cuál es la probabilidad de que la altura de un hombre esté entre 165cm y 185cm? (R:0.682)
- ¿Cuál es la altura que define el 5% de los hombres más pequeños? (R:158.55)

# Chapter 10

# Distribuciones de muestreo

## 10.1 Objetivo

En este capítulo, vamos a estudiar las estimaciones de la media y la varianza de las distribuciones normales utilizando **muestras aleatorias**.

Introduciremos la **media muestral** y la **varianza muestral** como variables aleatorias que estiman los parámetros de la distribución normal.

La media muestral y la varianza muestral tienen funciones de densidad de probabilidad, estas se denominan **funciones de densidad muestral**.

## 10.2 Muestra aleatoria

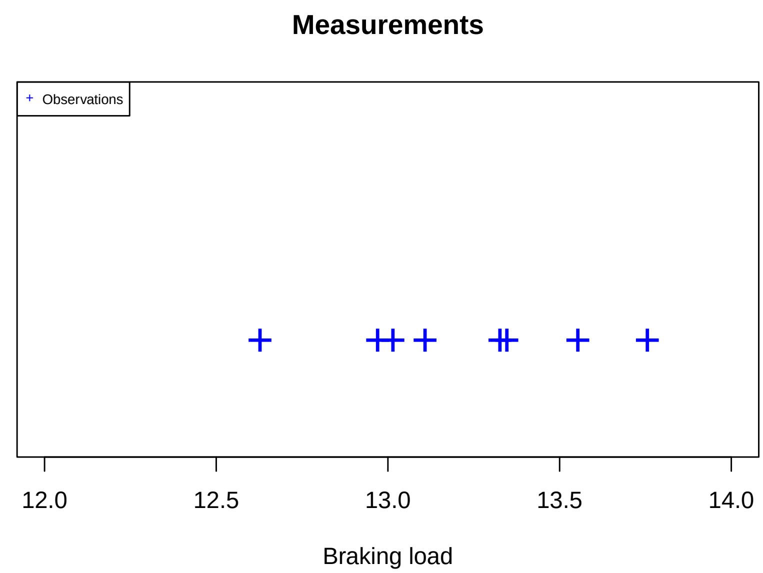
**Ejemplo (Cables)**

Imagina que un cliente le pide a tu empresa metalúrgica que le venda cables a 8 que pueden transportar hasta 96 Toneladas; eso es 12 Toneladas cada uno. Debes garantizar que ninguno de ellos frenará con este peso.

Tienes en **existencia** un conjunto de cables que podrían servir, pero no estás seguro. Por lo que tomas 8 cables aleatoriamente, y los cargas hasta que se rompen.

Decimos que tomas una **muestra aleatoria** de tamaño 8, lo que significa que repites el experimento aleatorio 8 veces. Aquí están los resultados

```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```




**Definición:**

Una **muestra aleatoria** de tamaño $n$ es la **repetición** de un experimento aleatorio $n$ **independientes** veces.

- Una muestra aleatoria es una **variable aleatoria** $n$-dimensional

$$(X_1, X_2, ...X_n)$$

donde $X_i$ es la *i-ésima* repetición del experimento aleatorio con distribución común $f(x;\theta)$ para cualquier $i$

- **Una observación** de una muestra aleatoria es el conjunto de $n$ valores obtenidos de los experimentos

$$(x_1, x_2, ...x_n)$$

Nuestra **observación** de la muestra de tamaño 8 cables fue

```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```

**Ejemplo (Cables)**

En la muestra observada de la carga de rotura de los cables se observó que

1) Ninguno de ellos rompió en 12 Toneladas.
2) Hubo uno que rompió en 12.62747 Toneladas.

¿Te arriesgas y vendes una muestra aleatoria de 8 cables de tu stock? ¿Qué sucede si su empresa es responsable de la rotura de un cable y tiene que pagar una multa elevada?

Para garantizar al cliente que los cables no se romperán a 12 Toneladas, nos gustaría ver que $P(X \leq 12)$ es razonablemente bajo.

## 10.3 Cálculo de probabilidades

Para calcular probabilidades necesitamos:

1. Un modelo de probabilidad (función de probabilidad)
2. Los parámetros del modelo (los valores de la función de probabilidad)

Vamos a **suponer** que la carga de rotura de los cables sigue una función de densidad de probabilidad **normal**.

$$X \rightarrow N(x; \mu, \sigma^2)$$

Para calcular $P(X \leq 12)$, necesitamos los parámetros $\mu$ y $\sigma^2$. ¿Cómo podemos estimar los parámetros de la muestra observada?

## 10.4 Estimación de los parámetros

Para encontrar valores probables para los parámetros usamos datos. Por lo tanto, tomamos una **muestra aleatoria**. Es decir, repetimos el experimento $n$ veces, recolectamos datos y los usamos para estimar los parámetros.

**Estimación de la media y la varianza**

Recordemos que para una variable aleatoria discreta, definimos la media como

$$\mu = \sum_{i}^{m} x_i f(x_i)$$

que es el centro de gravedad de las **probabilidades**, donde $f(x_i)$ es la función de probabilidad. Esta definición fue motivada por el centro de gravedad de las **observaciones**

$$\bar{x} = \frac{1}{n} \sum_{i}^{n} x_i = \sum_{i}^{m} x_i f_i$$

que definimos como **promedio**, y donde $f_i$ son las frecuencias relativas. Recuerda que $n$ es el número de observaciones (puede ser tan grande como queramos) y $m$ es el número de resultados posibles (normalmente fijado por el espacio muestral). Discutimos que cuando $n \rightarrow \infty$ entonces

$$\hat{P}(X = x) = f_i$$

Esto significa que las probabilidades pueden ser **estimadas** (poniéndose un **sombrero**) por las frecuencias relativas cuando $n$ es grande, porque $lim_{n\to\infty} f_i = f(x_i)$. Por lo tanto, también deberíamos tener que la **media** $\mu$ puede ser estimada por el **promedio** $\bar{x}$

$$\hat{\mu} = \bar{x} = \sum_{i}^{m} x_i \hat{P}(X = x)$$

Asi pues, podemos tomar el centro de la función de probabilidad como el centro de gravedad de los datos.

Con la varianza

$$\sigma^2 = \sum_{i}^{m} (x_i - \mu)^2 f(x_i)$$

tenemos una situación similar. En el límite cuando $n \to \infty$

$$\hat{\sigma}^2 = s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

y suponemos que el momento de inercia de los datos es cercano al momento de inercia de las probabilidades.

**Ejemplo (Cables)**

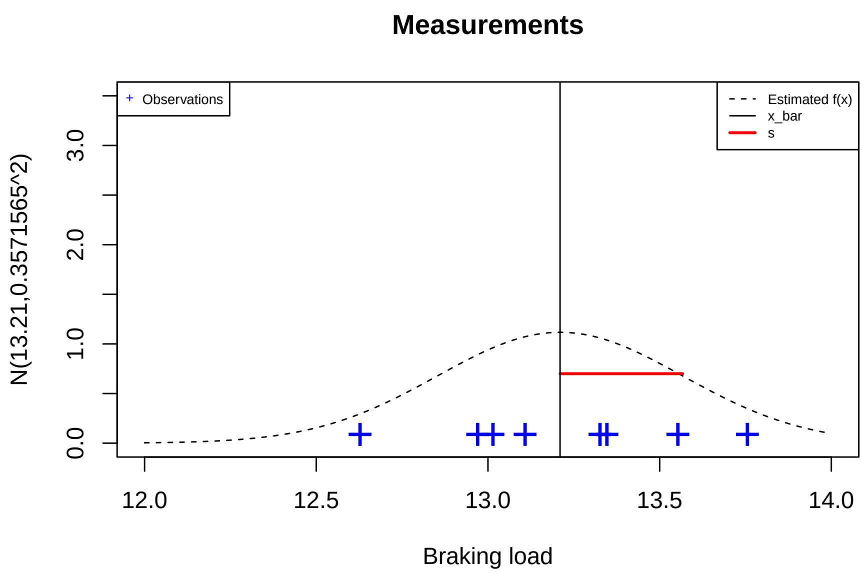
Suponiendo que la carga de rotura de nuestro cable es una variable aleatoria normal

$$X \to N(x; \mu, \sigma^2)$$

**usamos** las estimaciones $\bar{x}_{stock} = 13,21$ (mean(x)) y $s^2 = 0.3571565^2$ (sd( x)^2) como los valores de $\mu$ y $\sigma^2$. De tal forma que el modelo **ajustado** es

$$X \to N(x; \mu = 13.21, \sigma^2 = 0.3571565^2)$$

En este problema **no sabíamos** $\mu$ o $\sigma$ y, por lo tanto, estamos adivinando su valor y el modelo subyacente




¿Cuál es la probabilidad de que el cable se rompa a 12 Toneladas?

Como

$$X \to N(x; \mu = 13.21, \sigma^2 = 0.3571565^2)$$

entonces

$$P(X \leq 12) = F(12; \mu = 13.21, \sigma^2 = 0.1275608)$$

En R pnorm(12,13.21, 0.3571565)= 0.000352188

Dada la muestra **observada**, existe una probabilidad estimada de 0.03% de que un solo cable se rompa en 12 Toneladas. Tenemos un argumento probabilístico para vender los cables.

## 10.5 Margen de error de las estimaciones

Cuando estimamos los parámetros usando datos, como al tomar el valor de

$$\hat{\mu} = \bar{x}$$

por el valor de $\mu$; y el valor de

$$\hat{\sigma}^2 = s^2$$

por el valor de $\sigma^2$, sabemos que estamos **cometiendo un error**. Sabemos que si tomamos otra muestra de tamaño 8 cables **la estimación cambiará**, porque el promedio $\bar{x}$ cambiará.

¿Podemos tener una idea de cuán grande es el error de nuestra estimación?

Lo primero que debemos darnos cuenta es que el valor numérico que obtenemos para

$$\bar{x}$$

es la observación de una **variable aleatoria**

$$\bar{X}$$

**Definición**

La **media muestral** (o promedio) de una muestra aleatoria de tamaño $n$ se define como

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

El promedio es una **variable aleatoria** que en nuestra muestra de tamaño 8 tomó el valor

$$\bar{x}_{stock} = 13.21$$

Si tomamos otra muestra, este número cambiará.

**La media como estimador**

El número $\bar{x}$ se puede usar para **estimar** el parámetro desconocido $\mu$ porque la variable aleatoria $\bar{X}$ satisface estas dos propiedades importantes

1) es **insesgada**:
$$E(\bar{X}) = \mu$$

2) es **consistente**:
$$lim_{n\to\infty} V(\bar{X}) = 0$$

La primera propiedad se tiene porque

$$E(\bar{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = E(X) = \mu$$

La segunda propiedad se tien porque

$$V(\bar{X}) = V(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{V(\sum_{i=1}^{n} X_i)}{n^2} = \frac{V(X)}{n} = \frac{\sigma^2}{n}$$

Que utiliza el hecho de que cada experimento aleatorio en la muestra es independiente y por lo tanto $V(\sum_{i=1}^{n} X_i) = nV(X)$.

*Estimación de $\mu$*

Como consecuencia de las propiedades 1 y 2, entendemos que el valor $\bar{x}$ **se concentra más y más cerca** de $\mu$ a medida que aumenta $n$. Esto significa que el error que cometemos cuando tomamos un valor de $\bar{x}$ como la estimación de $\mu$

$$\bar{x} = \hat{\mu}$$

se vuelve más y más pequeño a medida que la muestra se hace más y más grande.

## 10.6 Inferencia

Sabemos que cuando tomamos muestras grandes, nuestro error es pequeño. Sin embargo, para un valor dado de $n$ queremos tener una **medida del error**. Por lo tanto, nos preguntamos por la **probabilidad de cometer un error** de un tamaño dado cuando estimamos $\mu$ con $\bar{x}$.

Cuando calculamos probabilidades en un estimador, decimos que estamos haciendo una **inferencia**. Los problemas de inferencia suelen surgir cuando nos interesa calcular la probabilidad de cometer un error al estimar $\mu$ con $\bar{x}$.

Para calcular probabilidades necesitamos

1. Un modelo de probabilidad (función de probabilidad)
2. Los parámetros del modelo (los valores de la función de probabilidad)

¿Cuáles son las funciones de probabilidad de $\bar{X}$ y $S^2$ para que podamos calcular sus probabilidades?

Estas funciones de probabilidad se denominan **funciones de probabilidad de muestreo**, porque se derivan de un experimento de muestreo.

**Ejemplo (cables)**

Hagamos una pregunta de inferencia. Imagina que nuestros cables están **certificados** para romperse con una carga promedio de $\mu = 13$ Toneladas con varianza $\sigma^2 = 0.35^2$.

Si tomamos una muestra aleatoria de 8 cables, ¿Cuál es la probabilidad de que la media de la muestra $\bar{X}$ esté dentro de un **margen de error** de 0.25 Toneladas de la media $\mu$?

$$P(-0.25 \leq \bar{X} - \mu \leq 0.25)$$

Para calcular esta probabilidad, necesitamos conocer la función de probabilidad de $\bar{X}$.

## 10.7 Distribución media muestral

**Teorema:** Si $X$ sigue una distribución normal

$$X \to N(\mu, \sigma^2)$$

entonces $\bar{X}$ es normal

$$\bar{X} \to N(\mu, \frac{\sigma^2}{n})$$

y $\bar{X}$ tiene

1) media (insesgado)
$$E(\bar{X}) = \mu$$
y
2) varianza (consistente)
$$V(\bar{X}) = \frac{\sigma^2}{n}$$

Llamamos $se = \sqrt{V(\bar{X})}$ al **error estándar** de la media muestral. El error estándar también se escribe como $\sigma_{\bar{x}}$. Ten en cuenta que este es el error que esperamos cuando usamos $\bar{x}$ como el valor de $\mu$, y es el sesgo que necesitábamos corregir para $S_n^2$.

Entonces, si **sabemos** $\mu$ y $\sigma$, podemos calcular las **probabilidades de** $\bar{X}$ usando la distribución normal.

Recuerda que tenemos **dos funciones de probabilidad**:

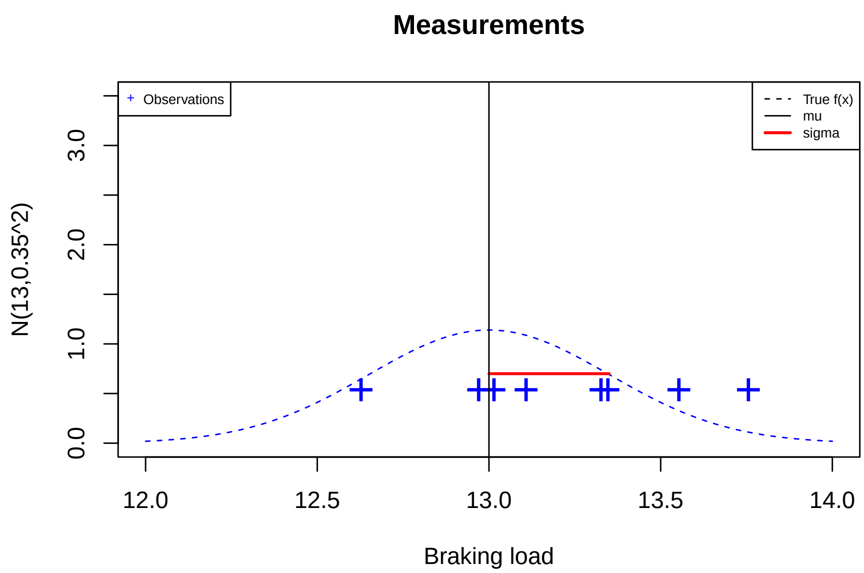
1. La función de probabilidad de $X$ también se conoce como la función de probabilidad de la **población**
2. La función de probabilidad de $\bar{X}$ es una función de probabilidad de la **muestra**.

**Ejemplo (cables)**

*Densidades de probabilidad para $X$ y $\bar{X}$*

En nuestro nuevo problema, ahora **sabemos** $\mu$ y $\sigma$ y la función de probabilidad de la **población**

$$X \to N(\mu = 13, \sigma^2 = 0.35^2)$$

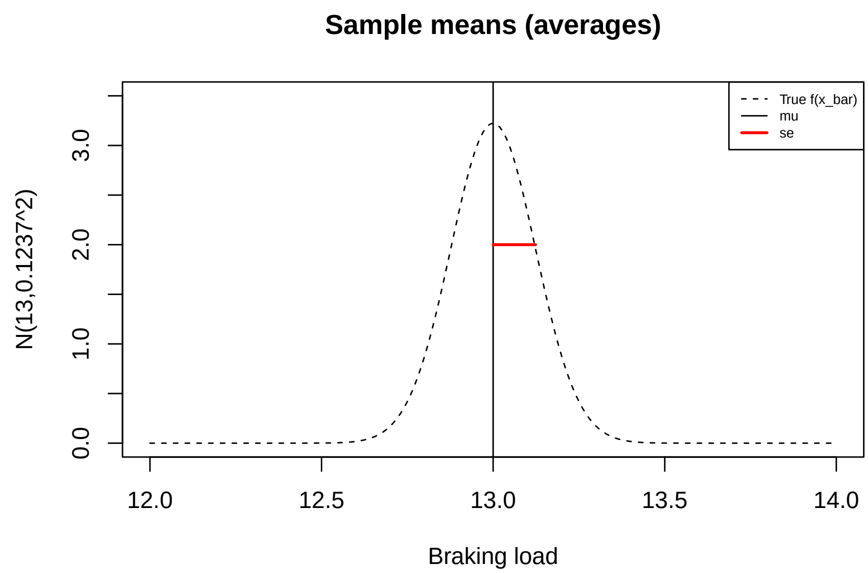



Dado que $X$ es normal, entonces $\bar{X}$ es normal y, por lo tanto, también conocemos la función de probabilidad de la media muestral $\bar{X}$

$$\bar{X} \rightarrow N(13, \frac{0.35^2}{8})$$

que tiene media y varianza

1) $E(\bar{X}) = \mu = 13$
2) $V(\bar{X}) = \frac{\sigma^2}{n} = \frac{0.35^2}{8} = 0.01530169$




Finalmente queremos calcular **la probabilidad** de que nuestra estimación tenga un margen de error de 0.25. Esa es una distancia de 0.25 de la media. Eso es

$$P(-0.25 \leq \bar{X} - 13 \leq 0.25) = P(12.75 \leq \bar{X} \leq 13.25)$$

$= F(13.25; \mu, \sigma^2/n) - F(12.75; \mu, \sigma^2/n)$

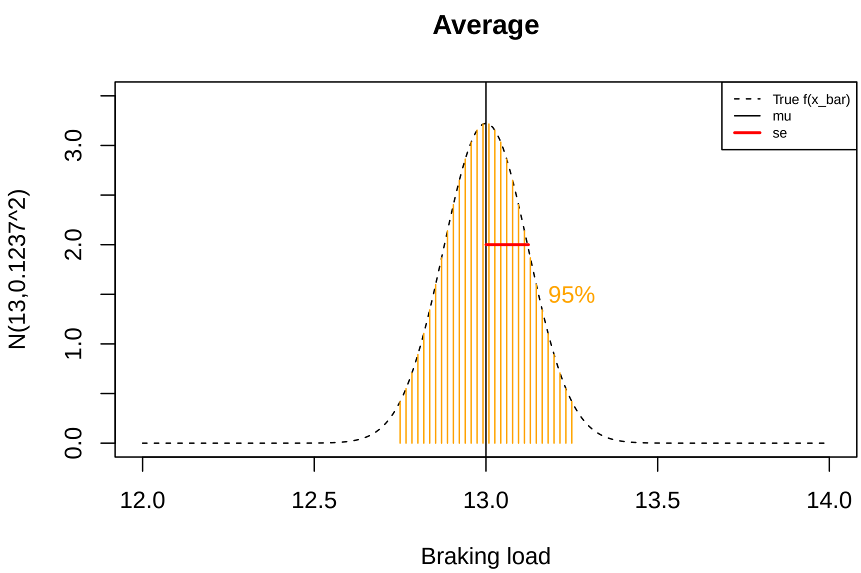
En R podemos calcularlo como:

pnorm(13.25, 13, 0.1237)-pnorm(12.75, 13, 0.1237)=0.956.

Recuerda: $se = \sigma_{\bar{x}} = \sqrt{0.01530169} = 0.1237$

Por lo tanto el 95.6% de los promedios $\bar{X}$ de muestras aleatorias de tamaño 8 están a una distancia de 0.25 de la media $\mu = 13$.

Si vendemos nuestro proceso para construir los cables, podemos decirles a los nuevos fabricantes que cuando sigan nuestras instrucciones, pueden probar el proceso tomando una muestra de tamaño 8 cables. En ese caso, pueden esperar que el promedio de la muestra caiga entre $(12.75, 13.25)$ alrededor de 95% de las veces.



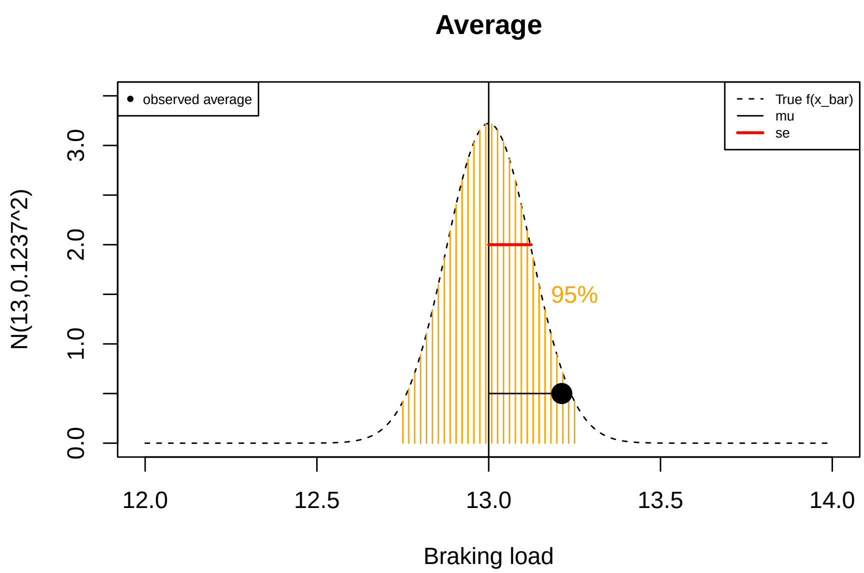

Cuando realizamos el muestreo aleatorio observamos:

```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```

Asumiendo que $\mu = 13$ entonces nuestro **error observado** en la estimación de la media es la diferencia

$$\bar{x}_{stock} - \mu = 13.21 - 13 = 0.21$$

Lo cual está dentro del margen de error de 95.6% y por lo tanto, debemos considerar que el proceso de fabricación está funcionando como se esperaba.




### 10.7.1 Suma muestral

Si estamos interesados en usar todos los 8 cables al mismo tiempo para transportar un total de 96 Toneladas, entonces deberíamos considerar **sumar** sus contribuciones individuales.

La **suma muestral** es la **estadística**

$$Y = n\bar{X} = \sum_{i=1}^{n} X_i$$

Una **estadística** es cualquier función de la muestra aleatoria $(X_1, ...X_n)$.

**Teorema:** si $X$ sigue una distribución normal

$$X \rightarrow N(\mu, \sigma^2)$$

entonces $Y$ es normal

$$Y \rightarrow N(n\mu, n\sigma^2)$$

$Y$ tiene

1) media

$$E(Y) = n\mu$$

2) varianza

$$V(Y) = n\sigma^2$$

**Ejemplo (suma de cables)**

¿Cuál es la probabilidad de que cuando juntamos todos los cables, puedan llevar un peso total entre $102 = 8(13 - 0.25)$ y $106 = 8(13 + 0.25)$ Toneladas?

**Sabemos** que para nuestros cables

$$X \to N(\mu = 13, \sigma^2 = 0.35^2)$$

entonces

$$Y \to N(n\mu = 104, n\sigma^2 = 8 \times 0.35^2)$$

con media y varianza

1) $E(Y) = n\mu = 104$
2) $V(Y) = n\sigma^2 = 8 \times 0.35^2 = 0.98$; $\sqrt{V(Y)} = 0.9899495$

Queremos calcular

$$P(102 \leq Y \leq 106)$$

$= F(102; n\mu, n\sigma^2) - F(106; n\mu, n\sigma^2)$

En R podemos calcularlo como:

pnorm(106, 104, 0.9899495)-pnorm(102, 104, 0.9899495)=0.956.

Por lo tanto 95.6% del peso total que pueden llevar 8 cables están entre 102 y 106 Toneladas, o una distancia de $8 * 0.25 = 2$ Toneladas de la media total $n\mu = 104$.

## 10.8 Variaza muestral

Al estimar la varianza

$$s^2 = \hat{\sigma}$$

También cometemos un error. ¿Cómo podemos estimar el error que cometemos?

**Definición**

La **varianza muestral** $S^2$ de una muestra aleatoria de tamaño $n$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

es la dispersión de las medidas al rededor de $\bar{X}$. En nuestra muestra de tamaño 8, $S^2$ tomó el valor

$$s^2_{stock} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0.1275608$$

$S^2$ es

1) insesgada: $E(S^2) = V(X) = \sigma^2$
2) consistente: $n \to \infty$, $V(S^2) \to 0$

y por lo tanto $S^2$ estima consistentemente $\sigma^2$

Podemos tomar un valor de $s^2$ como estimación para $\sigma^2$ o

$$s^2 = \hat{\sigma}^2$$

De manera similar a $\hat{\mu}$, el error de esta estimación se hace cada vez más pequeño a medida que $n$ se hace cada vez más grande.

**La varianza muestral insesgada (¿por qué dividimos entre n-1?)**

Podríamos proponer estimar $\sigma^2$ dividiendo las diferencias cuadráticas de $\bar{X}$ por $n$

$$S^2_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$S^2_n$ es por lo tanto

1) **sesgada**: $E(S^2_n) = \sigma^2 - \frac{\sigma^2}{n} \neq \sigma^2$
2) pero consistente $V(S^2_n) \to 0$ cuando $n \to \infty$

El término de sesgo $\frac{\sigma^2}{n}$ surge porque $S^2_n$ mide la dispersión al rededor de $\bar{X}$ y no al rededor de $\mu$. Recuerda que el error que cometemos cuando sustituimos $\bar{x}$ por $\mu$ es la varianza de $\bar{X}$: $\sigma^2/n$. Corrijamos el sesgo, escribiendo la ecuación 1 anterior como:

$$E(\frac{n}{n-1}S^2_n) = \sigma^2$$

Podemos definir la **varianza muestral** (corregida)

$$S^2 = \frac{n}{n-1}S^2_n = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

que es un estimador insesgado de $\sigma^2$ porque $E(S^2) = \sigma^2$.

También podemos tener problemas de inferencia cuando estamos interesados en la probabilidad de la **varianza muestral** $S^2$.

Considera un proceso de control de calidad que requiera que los cables se produzcan cerca del valor especificado $\mu$. No queremos cables que se rompan demasiado lejos de la media.

Si una muestra de tamaño 8 cables está muy dispersa ($S^2 > 0.3$), detenemos la producción: el proceso está fuera de control.

¿Cuál es la probabilidad de que la varianza muestral de una muestra de tamaño 8 cables sea mayor que los 0.3 requeridos?

## 10.9 Probabilidades de la varianza muestral

**Teorema:** Si $X$ sigue una distribución normal

$$X \to N(\mu, \sigma^2)$$

La **estadística**:

$$W = \frac{(n-1)S^2}{\sigma^2} \to \chi^2(n-1)$$

tiene una distribución $\chi^2$ (chi-cuadrado) con $df = n-1$ grados de libertad dada por

$$f(w) = C_n w^{\frac{n-3}{2}} e^{-\frac{w}{2}}$$

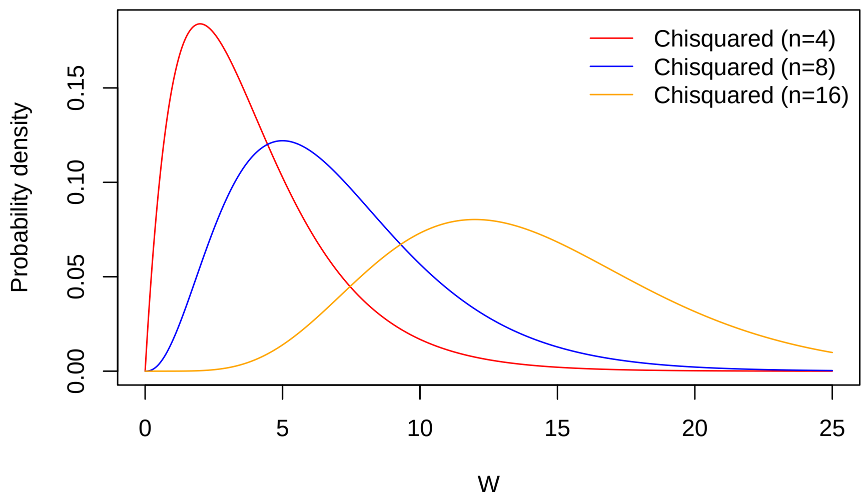
dónde:

1) $C_n = \frac{1}{2^{(n-1)/2}\sqrt{\pi(n-1)}}$ asegura $\int_{-\infty}^{\infty} f(t)dt = 1$
2) $\Gamma(x)$ es el factorial de Euler para números reales

Si **sabemos** el valor de $\sigma$, podemos calcular las probabilidades de $S^2$ usando la distribución $\chi^2$ para $W$.

## 10.10 $\chi^2$-estadística

La densidad de probabilidad $\chi^2$ tiene un parámetro $df = n-1$, llamado grados de libertad. Veamos algunas densidades de probabilidad en la familia de modelos de probabilidad $\chi^2$




**Ejemplo (variaciones en la rotura del cable)**

Si **sabemos** que nuestros cables

$$X \to N(\mu = 13, \sigma^2 = 0.35^2)$$

entonces

$$W = \frac{(n-1)S^2}{\sigma^2} = \frac{7S^2}{0.35^2} \to \chi^2(n-1)$$

podemos calcular

$$P(S^2 > 0.3) = P(\frac{(n-1)S^2}{\sigma^2} > \frac{(n-1)0.3}{\sigma^2})$$

$= P(W > \frac{7*0.3}{0.35^2}) = P(W > 17.14286)$

$= 1 - P(W \leq 17.14286)$

$= 1 - F_{\chi^2, df=7}(17.14286) = 0.016$

En R 1-pchisq(17.14286, df=7)=0.016

Solo hay una probabilidad de 1% de obtener un valor superior a $s^2 = 0.3$. Por lo tanto, $s^2 > 0.3$ parece ser un buen criterio para detener la producción y revisar el proceso.

Si tomamos una muestra aleatoria y obtenemos un valor de $s^2$ que es mayor que 0.3, será una observación rara si todo está bien. Solemos creer que los valores observados son comunes, no raros, por lo que podemos pensar que algo no está bien.

Cuando realizamos el muestreo aleatorio observamos:

```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```

Por lo tanto, nuestro valor observado fue $s^2_{stock} = 0.1275608$

La muestra no está muy dispersa porque $s^2_{stock} < 0.3$ y creemos que todo está bien y la producción está bajo control.

## 10.11 Preguntas

**1)** La media muestral es un estimador insesgado de la media poblacional porque

**a:** El valor esperado de la media muestral es la media poblacional; **b:** El valor esperado de la media poblacional es la media muestral; **c:** El error estándar tiende a cero cuando $n$ tiende a infinito; **d:** La varianza de la media muestral tiende a cero cuando $n$ tiende a infinito;

**2)** ¿Por qué se usa la estadística $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$ en su lugar de $S_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$ para estimar la varianza de una variable aleatoria?

**a:** porque su varianza es 0; **b:** porque es un estimador consistente de $\sigma^2$; **c:** porque es un estimador insesgado de $\sigma^2$; **d:** porque es la distancia cuadrática promedio a la media muestral ($\bar{X}$);

**3)** ¿Cuál es la varianza de la media muestral $\bar{X} = \frac{1}{n}\sum_{i=1}^n X_i$?

**a:**$\sigma$; **b:**$\frac{\sigma}{\sqrt{n}}$; **c:**$\sigma^2$; **d:**$\frac{\sigma^2}{n}$;

**4)** ¿Cuál es la media y la varianza de la suma muestral?

**a:**$\mu$, $n\sigma$; **b:**$n\mu$,$n\sigma$; **c:**$\mu$, $n\sigma^2$; **d:**$n\mu$, $n\sigma^2$;

**5)** Una pregunta de inferencia implica:

**a:** calcular el valor esperado de un estimador; **b:** estimar el valor de un parámetro; **c:** calcular una probabilidad de un estimador; **d:** ajustar un modelo de probabilidad;

## 10.12 Ejercicios

#### 10.12.0.1 Ejercicio 1

Una empresa de electrónica fabrica resistencias que tienen una resistencia media de 100 ohmios y una desviación estándar de 10 ohmios. La distribución de la resistencia es normal.

- ¿Cuál es la media muestral de $n = 25$ resistencias? (R:100)
- ¿Cuál es la varianza de la media muestral de $n = 25$ resistencias? (R:4)
- ¿Cuál es el error estándar de la media muestral de $n = 25$ resistencias? (R:2)
- Encuentra la probabilidad que una muestra aleatoria de $n = 25$ resistencias tenga una resistencia promedio de menos de 95 ohmios (R: 0.0062)

#### 10.12.0.2 Ejercicio 2

Un modelo de batería carga una media de 75% de su capacidad en una hora con una desviación estándar de 15%.

- Si la carga de la batería es una variable normal, ¿cuál es la probabilidad de que la diferencia de carga entre la media muestral de 25 baterías y la carga media sea como mucho de 5%? (R:0.9044)
- Si cargamos 100 baterías, ¿cuál es esa probabilidad? (R:0.9991)
- Si en cambio solo cargamos 9 baterías, ¿qué carga $c$ es superada por la media muestral con probabilidad de 0.015? (R:85.850)

# Chapter 11

# Teorema central del límite

## 11.1 Objetivo

En este capítulo introduciremos el **margen de errores** al estimar la media de la distribución de la población por el promedio.

Discutiremos cómo el **Teorema central del límite** nos permitirá calcular el margen de error para cualquier tipo de distribución si la muestra es grande.

También introducirá la estadística t, para calcular el margen de error cuando la muestra es pequeña pero la distribución de la población es normal.

## 11.2 Margen de error

Al decidir si el error de estimación de $\mu$ por la media muestral $\bar{x}$ es grande o no, generalmente lo comparamos con una tolerancia **predefinida**.

El **margen de error** a nivel de 5% es la distancia $m$ de $\bar{X}$ de $\mu$ que captura 95% de las estimaciones:

$$P(-m \leq \bar{X} - \mu \leq m) = P(\mu - m \leq \bar{X} \leq \mu + m) = 0.95$$

Esto significa que 95% de los posibles resultados de $\bar{X}$ están a una distancia $m$ de $\mu$

## 11.3 Ejemplo (cables)

Si tomamos una muestra de cables de 8 de una población de cables cuya carga de ruptura sigue una distribución normal con parámetros **conocidos** $\mu = 13$ y $\sigma^2 = 0.35^2$,

$$X \rightarrow N(\mu = 13, \sigma^2 = 0.35^2)$$

¿Cuál es el margen de error cuando estimamos $\mu$ por $\bar{x}$?

**Calculando el maging de error de una variable normal**

Queremos saber el número $m$ en la ecuación

$$P(\mu - m \leq \bar{X} \leq \mu + m) = 0.95$$

Para resolver esta ecuación necesitamos dos pasos. **Primero**, necesitamos saber la **distribución** de $\bar{X}$.

1. Cuando $X$ es normal $X \rightarrow N(\mu, \sigma^2)$ entonces

$$\bar{X} \rightarrow N(\mu, \frac{\sigma^2}{n})$$

**Entonces** necesitamos **estandarizar** $\bar{X}$. Recuerda que para estandarizar una variable normal **restamos su media** y la **dividimos por su desviación estándar**.

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \rightarrow N(0, 1)$$

2. Sustituyendo la media de $\bar{X}$ y su desviación estándar en la ecuación del margen de error, tenemos:

$$P(\mu - m \leq \bar{X} \leq \mu + m) = P(-\tfrac{m}{\sigma/\sqrt{n}} \leq \tfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \tfrac{m}{\sigma/\sqrt{n}})$$

$$= P(-\frac{m}{\sigma/\sqrt{n}} \leq Z \leq \frac{m}{\sigma/\sqrt{n}}) = 0.95$$

Compáralo con el gráfico anterior: $\frac{m}{\sigma/\sqrt{n}}$ es la distancia alrededor de 0 que captura 95% de la variable normal estándar de distribución. La distancia deja una probabilidad de 2.5% en cada extremo de la distribución. Para la parte superior de la cola, esto es

$$\frac{m}{\sigma/\sqrt{n}} = \phi^{-1}(0.975) = 1.96$$

donde $\phi^{-1}$ es la inversa de la distribución normal estándar (qnorm(0.975)). Por lo tanto

$$m = 1.96\frac{\sigma}{\sqrt{n}}$$

**Ejemplo (cables)**

La media muestral $\bar{X}$ de una muestra de cables de 8 sigue una distribución normal con:

1. media $E(\bar{X}) = \mu$

y

2. error estándar $se = \sqrt{V(\bar{X})} = \frac{\sigma}{\sqrt{n}} = \frac{0.35}{\sqrt{8}}$

Entonces el margen de error en 5% es:

$$m = 1.96 \frac{0.35}{\sqrt{8}} = 0.24$$

Podemos esperar que 95% de los promedios ($\bar{x}$) para la carga de rotura de los cables de 8 caigan entre $(13 - 0.24, 13 + 0.24) = (12.76, 13.24)$

## 11.4 Teorema central del límite

Pudimos resolver el margen de error porque asumimos que esa variable $X$ era normal. ¿Qué pasa si $X$ sigue cualquier otra distribución de probabilidad?

**Teorema:** Para cualquier variable aleatoria $X$ con cualquier tipo de distribución

$$X \to f(x; \theta)$$

la estadística estandarizada

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}}$$

se aproxima a una distribución estándar

$$Z \to_d N(0, 1)$$

cuando $n \to \infty$

**Consecuencia:** Podemos calcular las probabilidades de $\bar{X}$ si $n$ es grande, usando la distribución normal:

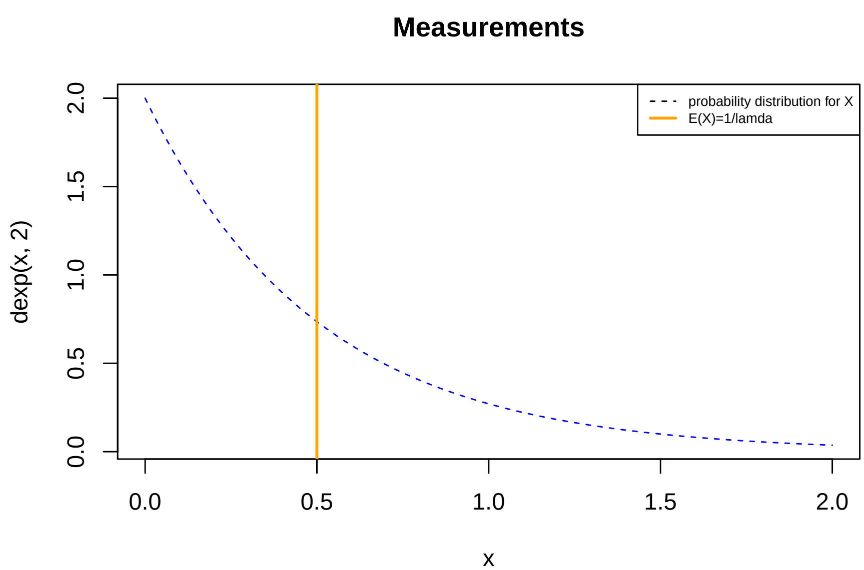
$$\bar{X} \sim_{aprox} N(\mu, \frac{\sigma^2}{n})$$

**Ejemplo (fármaco en concentración sanguínea):**

Considera un experimento en el que queremos medir la concentración en sangre de un fármaco después de 10 horas de administración en 30 pacientes.

Si **sabemos** que los niveles siguen una distribución exponencial

$$X \to exp(\lambda = 2)$$




La media y la varianza son:

- $\mu = \frac{1}{\lambda} = 0.5$
- $\sigma^2 = \frac{1}{\lambda^2} = 0.25$

Por lo tanto, la media y el error estándar de $\bar{X}$ son:

- $E(\bar{X}) = \frac{1}{\lambda} = 0.5$
- $se = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{1}{n\lambda^2}} = 0.091$

Como $n \geq 30$

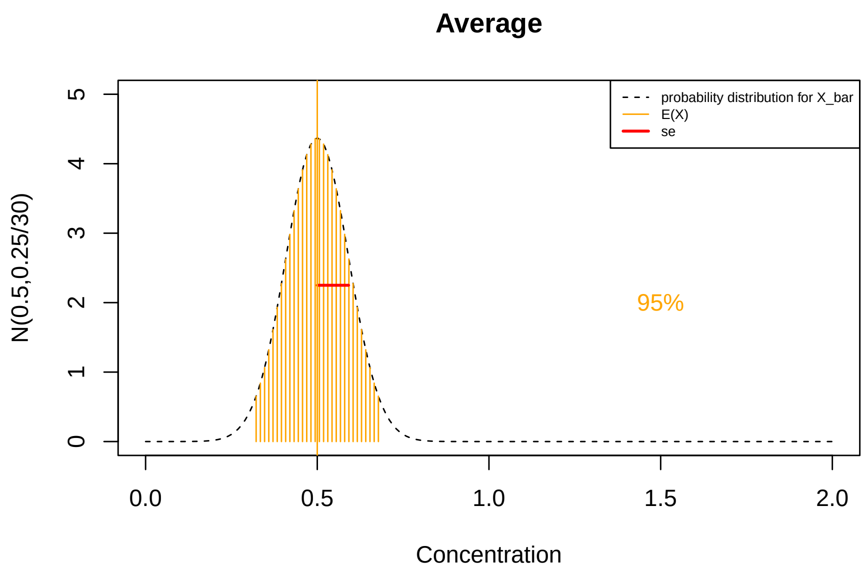
$$Z = \frac{\bar{X} - \lambda}{\sqrt{\frac{1}{n\lambda^2}}}$$

es una variable normal estándar y: $\bar{X} \sim_{aprox} N(\lambda, \frac{1}{n\lambda^2})$

El margen de error al nivel de 5% se puede calcular de nuevo con la distribución estándar

$$m = \phi^{-1}(0.75)\frac{\sigma}{\sqrt{n}} = 1.96\frac{0.25}{\sqrt{30}} = 0.1789227$$

Podemos esperar que 95 de los promedios de muestras de pacientes de 30 caigan entre $(0.5 - 0.178, 0.5 + 0.178) = (0.322, 0.678)$




## 11.5 Suma muestral y CLT

Para cualquier variable aleatoria $X$ con distribución **desconocida** (cualquier tipo de)

$$X \to f(x; \theta)$$

la estadística estandarizada

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{n\bar{X} - n\mu}{\sqrt{n}\sigma}$$

se aproxima a una distribución estándar

$$Z \to_d N(0, 1)$$

cuando $n \to \infty$

**Consecuencia:** Podemos calcular probabilidades para la suma muestral $Y = n\bar{X}$ si $n$ es grande, usando la distribución normal:

$$\bar{Y} \sim_{aprox} N(n\mu, n\sigma)$$

## 11.6 Preguntas

**1)** La importancia del Teorema central del límite es que se aplica a la estandarización de

**a:** Una variable aleatoria; **b:** La media muestral de una variable normal; **c:** La media muestral de una variable aleatoria; **d:** Una variable normal;

**2)** Cuando $n > 30$, la media muestral de una variable aleatoria ($\frac{1}{n}\sum_i Xi$) y su suma muestral ($\sum_i Xi$) se pueden aproximar a

**a:** $N(\frac{\mu}{n}, \frac{\sigma^2}{n})$ y $N(\mu, \sigma^2)$; **b:** $N(\mu, n\sigma^2)$ y $N(\mu, \frac{\sigma^2}{n})$; **c:** $N(\mu, \frac{\sigma^2}{n})$ y $N(\mu, \frac{\sigma^2}{n})$; **d:** $N(\mu, \frac{\sigma^2}{n})$ y $N(n\mu, n\sigma^2)$

**3)** Para una variable normal estándar, si ponemos el número $z_{0.025}$ en la definición del margen de error $m = z_{0.025}\frac{\sigma}{\sqrt{n}}$, entonces se referirá a

**a:** El primer cuartil; **b:** El número en el que la distribución ha acumulado 0.975 de probabilidad; **c:** El número en el que la distribución ha acumulado una probabilidad de 0.025; **d:** El tercer cuartil;

**4)** La probabilidad de que la media muestral de 50 observaciones esté una distancia de un error estandar ($se = \frac{\sigma}{\sqrt{n}}$) de de la media poblacional $\mu$ es:

**a:**2*(1-pnorm(1))= 0.3173105; **c:**2*(1-pnorm(2))=0.04550026; **b:**-1+2*pnorm(1)=0.6826895; **d:**-1+2*pnorm(2)=0.9544997

**5)** Una resonancia magnética del hipocampo del cerebro tiene 100 píxeles. Esperamos que 90% de los píxeles sean blancos (tejido cerebral). Según el Teorema central del límite , ¿cuál es la probabilidad de que el escaneo de un paciente tenga como máximo 85% de píxeles blancos?

**a:**pnorm(0.9, 0.85, sqrt(0.85*0.15)/10); **b:**dnorm(0.85, 0.9, sqrt(0.9*0.1)/10); **c:**pnorm(0.85, 0.9, sqrt(0.9*0.1)/10); **d:**dnorm(0.9, 0.85, sqrt(0.85*0.15)/10)

## 11.7 Ejercicios

#### 11.7.0.1 Ejercicio 1

Se necesita un componente electrónico para el correcto funcionamiento de un telescopio. Necesita ser reemplazado inmediatamente cuando se desgasta.

La vida media del componente ($\mu$) es de 100 horas y su desviación estándar $\sigma$ es de 30 horas.

- ¿Cuál es la probabilidad de que el promedio de la vida media de 50 componentes esté dentro de 1 hora de la vida media de un solo componente? (R:0.1863)
- ¿Cuántos componentes necesitamos para que el telescopio esté operativo 2750 horas consecutivas con al menos 0.95 de probabilidad? (R:31)

#### 11.7.0.2 Ejercicio 2

La probabilidad de que se encuentre una mutación particular en la población es de 0.4. Si testamos 2000 personas por la mutación:

- ¿Cuál es la probabilidad de que el número total de personas con la mutación esté entre 791 y 809? (R:0.31)

sugerencia: usa el CLT con una muestra de ensayos de Bernoulli de 2000. Esto se conoce como la **aproximación normal de la distribución binomial** que es buena cuando $p$ y $1 - p$ son ambis mayor que 5.

#### 11.7.0.3 Ejercicio 3

Una máquina automática llena tubos de ensayo con muestras biológicas con una media de $\mu = 130$mg y una desviación estándar de $\sigma = 5$mg.

- para una muestra aleatoria de tamaño 50. ¿Cuál es la probabilidad de que la media muestral (promedio) esté entre 128mg y 132mg? (R:0.995)
- ¿Cuál debería ser el tamaño de la muestra ($n$) tal que la media muestral $\bar{X}$ sea superior a 131gr con una probabilidad menor o igual a 0.025?(R:97)

#### 11.7.0.4 Ejercicio 4

En el Caribe, parece haber un promedio de huracanes de 6 por año. Teniendo en cuenta que la formación de huracanes es un proceso de Poisson, los meteorólogos planean estimar el tiempo medio entre la formación de dos huracanes. Planean recolectar una muestra de tamaño 36 para los tiempos entre dos huracanes.

- ¿Cuál es la probabilidad de que su promedio muestral esté entre 45 y 60 días? (R:0.39)
- ¿Cuál debe ser el tamaño de la muestra para que tengan una probabilidad de 0.025 de que la media muestral sea mayor a 70 días? (R:169)

# Chapter 12

# Máxima verosimilitud y Método de los Momentos

## 12.1 Objetivo

En este capítulo discutiremos qué es un **estimador** y daremos algunos ejemplos. Luego introduciremos dos métodos para obtener **estimadores** de los parámetros de los modelos de probabilidad.

Estos son la **máxima verosimilitud** y el **método de los momentos**.

## 12.2 Estadística

**Definición**

Una **estadística** es cualquier función de una **muestra aleatoria**

$$T(X_1, X_2, ..., X_n)$$

Por lo general, devuelve un número.

Las estadísticas son **variables aleatorias** y sus **distribuciones de probabilidad** se llaman **distribuciones de muestreo**

Las estadísticas tienen diferentes funciones:

1. **Descripción** de los datos de una muestra

- ubicación: $\bar{X}$
- Mínimo: $\min\{X_i\}$
- Máximo: $\max\{X_i\}$

2. **Estimación** de los **parámetros** de un modelo de probabilidad

- media: $\bar{X}$ para $\mu$
- varianza: $S^2$ para $\sigma^2$

3. **Inferencia** para decir algo sobre los parámetros dados los datos

- media: $Z$, $T$
- varianza: $\chi^2$

Recuerda: Todas son variables aleatorias. Cada vez que tomamos otra muestra cambian su valor.

**Definición de estimadores**

Un **estimador** es un estadístico cuyos valores observados se utilizan para estimar los **parámetros** de la distribución de la población sobre la que se define la muestra.

Si escribimos la distribución de la población como

$$X \to f(x;\theta)$$

entonces $\theta$ es un parámetro y $\Theta$ es una variable aleatoria cuyas observaciones $\hat{\theta}$ tomamos como estimaciones de $\theta$

$$\hat{\theta} \sim \theta$$

Por lo tanto hay tres cantidades diferentes que debemos considerar:

1. $\theta$ es un **parámetro** de la distribución de la población $f(x;\theta)$
2. $\Theta$ es un **estimador** de $\theta$: Una variable aleatoria
3. $\hat{\theta}$ es la **estimación** de $\theta$: un valor realizado de $\Theta$

**Ejemplo (media de la muestra)**

Cuando tenemos una variable aleatoria normal

$$X \to N(\mu, \sigma^2)$$

identificamos las tres cantidades diferentes:

1. $\mu$ es un **parámetro** de la distribución de la **población**: $N(\mu, \sigma^2)$
2. $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ es un **estimador** de $\mu$
3. $\bar{x} = \hat{\mu}$ es la **estimación** de $\mu$

**Ejemplo (varianza de la muestra)**

Cuando tenemos una variable aleatoria normal

$$X \to N(\mu, \sigma^2)$$

1. $\sigma^2$ es un **parámetro** de la distribución de la población
2. $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ es un **estimador** de $\sigma^2$
3. $s^2 = \hat{\sigma}^2$ es la **estimación** de $\sigma^2$

## 12.3 Propiedades

1. Un estimador es **insesgado** si su valor esperado es el parámetro

$$E(\Theta) = \theta$$

Por ejemplo:

- $\bar{X}$ es un estimador **insesgado** de $\mu$ porque $E(\bar{X}) = \mu$
- $S^2$ es un estimador **insesgado** de $\sigma^2$ porque $E(S^2) = \sigma^2$

2. Un estimador es **consistente** cuando sus valores observados están cada vez menos dispersos al rededor de la medida a medida que el tamaño de la muestra crece

$$lim_{n\rightarrow\infty} V(\Theta) = 0$$

Por ejemplo:

- $\bar{X}$ es **consistente** porque $V(\bar{X}) = \frac{\sigma^2}{n} \rightarrow 0$ cuando $n \rightarrow \infty$.

## 12.4 Máxima verosimilitud

¿Cómo se pueden obtener **estimadores** de los parámetros de **cualquier** modelo de probabilidad?

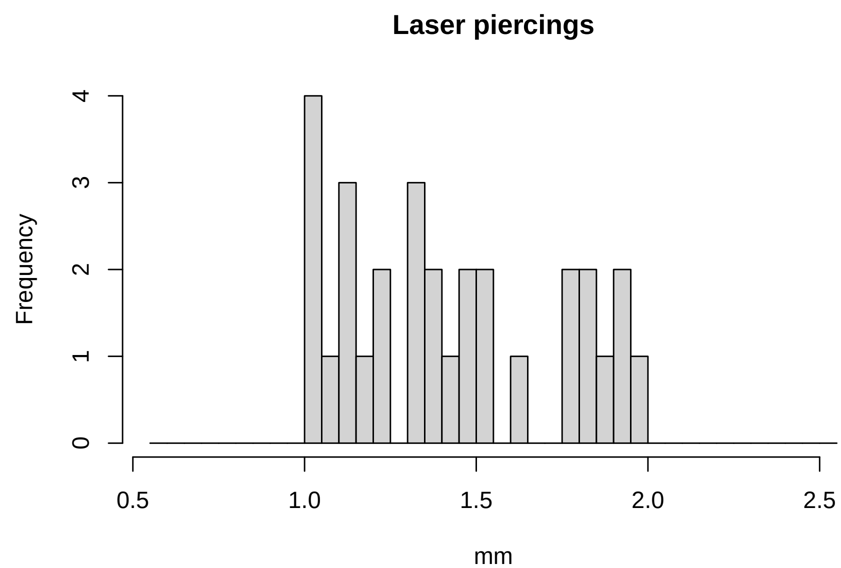
**Ejemplo (Láser)**

Imagina que diseñamos un láser con un diámetro de $1mm$ que queremos usar para aplicaciones clínicas.

Queremos caracterizar el diámetro de un piercing en un tejido realizado con láser y tomar una muestra aleatoria de 30 cortes realizados con láser. Aquí están los resultados

```
##  [1] 1.11 1.64 1.20 1.79 1.89 1.01 1.31 1.81 1.34 1.25 1.92 1.24 1.49 1.36 1.03
## [16] 1.82 1.09 1.01 1.14 1.91 1.80 1.51 1.44 1.98 1.46 1.53 1.33 1.39 1.12 1.04
```

y el histograma




¿Cuál sería una función de probabilidad que podría describir los datos?

Para ello seguimos el siguiente proceso:

1. Proponemos **un modelo** que depende de parámetros
2. Derivamos los **estimadores** para los parámetros, por máxima verosimilitud o el método de momentos.
3. Finalmente usamos el estimador para **estimar los parámetros** con los datos.

*Proponiendo una densidad de probabilidad*

En muchas aplicaciones, podemos proponer la forma de una densidad de probabilidad que depende de algunos parámetros. Proponer un modelo de probabilidad se hace siguiendo **propiedades generales** de las observaciones, o lo que esperamos observar. El modelado requiere experiencia, habilidad y conocimiento de varias funciones matemáticas. Sin embargo, en la mayoría de los casos se suelen aplicar **modelos bien conocidos**.

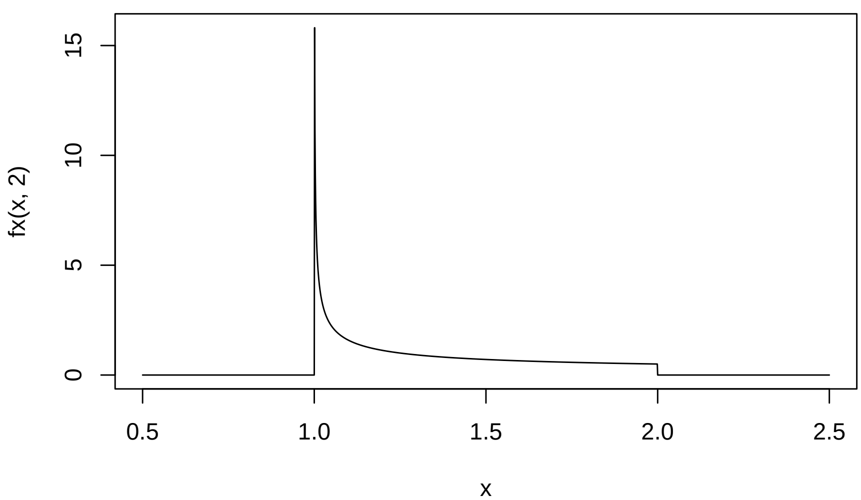
**Ejemplo (Láser)**

En nuestro ejemplo, podemos considerar, por ejemplo, que se debe dar la máxima probabilidad a los diámetros de $x = 1mm$, y que los diámetros deben disminuir como la potencia inversa de algún parámetro **desconocido** $\alpha$, con un límite de $2mm$ más allá del cual la probabilidad es de 0.

Una distribución de densidad de probabilidad adecuada es

$$f(x) = \begin{cases} \frac{1}{\alpha}(x-1)^{\frac{1}{\alpha}-1}, & \text{if } x \in (1,2) \\ 0, & x \notin (1,2) \end{cases}$$

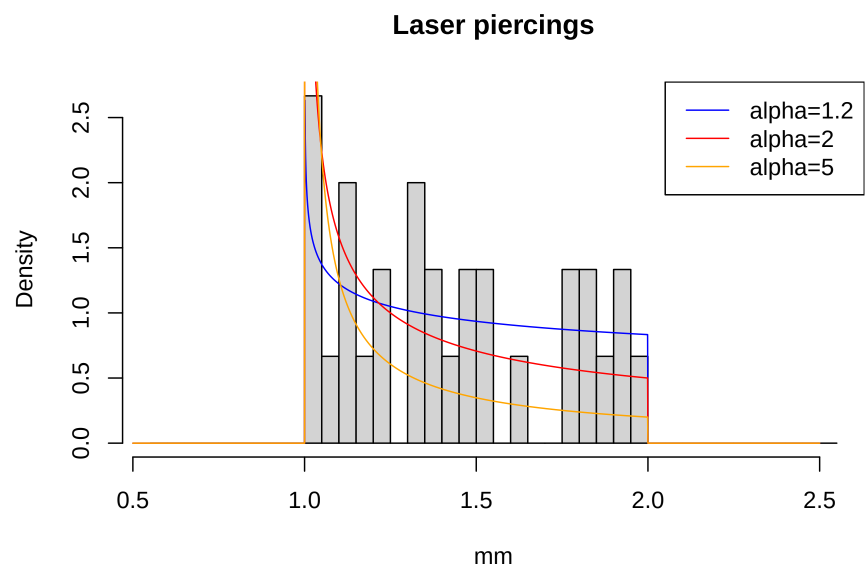
Donde $\alpha$ es un parámetro. Esta es una densidad de probabilidad porque se integra a uno y es positiva. En particular, para $\alpha = 2$ podemos graficarlo




*Derivar los estimadores*

Si realizamos una muestra de tamaño $n$: $(X_1, ...X_n)$, ¿Cómo debemos combinar los datos para obtener el mejor valor de $\alpha$?

Muchos valores de para el parámetro podrían explicar los datos. Nos interesa **un criterio** para elegir un valor en particular.




El método de **máxima verosimilitud** nos da un estimador para $\alpha$

$$\hat{\alpha}_{ml}$$

## 12.5 Máxima verosimilitud

El objetivo es encontrar el valor del parámetro que **creemos** es el que **mejor** representa los datos.

El método de máxima verosimilitud se basa en la búsqueda del valor del parámetro que hace más **probable** la **observación** de la muestra.

**Máxima verosimilitud paso 1**

Calculamos la probabilidad de haber observado la muestra $n$: $x_1, ...x_n$. Es el producto de probabilidades porque las observaciones son independientes entre sí:

$$P(M = x_1, ...x_n) = P(X = x_1)P(X = x_2)...P(X = x_n)$$
$$= f(x_1; \alpha)f(x_2; \alpha)...f(x_n; \alpha)$$

A esta función la llamamos **función de verosimilitud** y consideramos que:

- Una vez observados los datos estos son **fijos**
- La incógnita es $\alpha$

$$L(\alpha) = \Pi_{i=1..n} f(x_i;\alpha)$$

**Ejemplo (Láser)**

Para el experimento con láser la función de verosimilitud es

$L(\alpha; x_1, ..x_n) = \frac{1}{\alpha^n}\Pi_{i=1..n}(x_i - 1)^{\frac{1-\alpha}{\alpha}} = \frac{1}{\alpha^n}\{(x_1 - 1)(x_2 - 1)...(x_n - 1)\}^{\frac{1-\alpha}{\alpha}}$

**Máxima verosimilitud paso 2**

Entonces nos preguntamos: ¿cuál es el valor de $\alpha$ que hace que la muestra observada sea el evento más probable? Por lo tanto, queremos maximizar $L(\alpha)$ con respecto a $\alpha$. Como tenemos la multiplicación de muchos factores es más fácil maximizar el logaritmo de $L(\alpha)$. Esto se llama la función logaritmo de verosimilitud:

$$\ln L(\alpha; x_1, ..x_n)$$

**Ejemplo (Láser)**

En el ejemplo del láser, tomamos el logaritmo y obtenemos la **Logaritmo de verosimilitud**

$$\ln L(\alpha; x_1, ..x_n) = -n\ln(\alpha) + \frac{1-\alpha}{\alpha}\Sigma_{i=1...n}\ln(x_i - 1)$$

**Máxima verosimilitud paso 3**

Finalmente **maximizamos** el logaritmo de verosimilitud con respecto al parámetro. Por lo tanto, diferenciamos el log-verosimilitud con respecto al parámetro $\alpha$, igualamos a cero y resolvemos para el máximo.

$$\frac{d\ln L(\alpha)}{d\alpha}\Big|_{\hat{\alpha}} = 0$$

El valor máximo del parámetro se denomina **estimación de máxima verosimilitud** para el parámetro y se escribe con un sombrero $\hat{\alpha}$.

**Ejemplo (Láser)**

Derivamos la función log-verosimilitud

$$\frac{d\ln L(\alpha)}{d\alpha} = -\frac{n}{\alpha} - \frac{1}{\alpha^2}\Sigma_{i=1...n}\ln(x_i - 1)$$

El máximo es donde la derivada es 0. Este máximo es el valor de nuestro estimador $\hat{\alpha}_{ml}$.

$$\hat{\alpha}_{ml} = -\frac{1}{n}\sum_{i=1}^{n}\ln(x_i - 1)$$

El estimador del parámetro es por lo tanto (nótese las letras mayúsculas)

$$A = -\frac{1}{n}\sum_{i=1}^{n}\ln(X_i - 1)$$

Que es una variable aleatoria, función de la muestra aleatoria

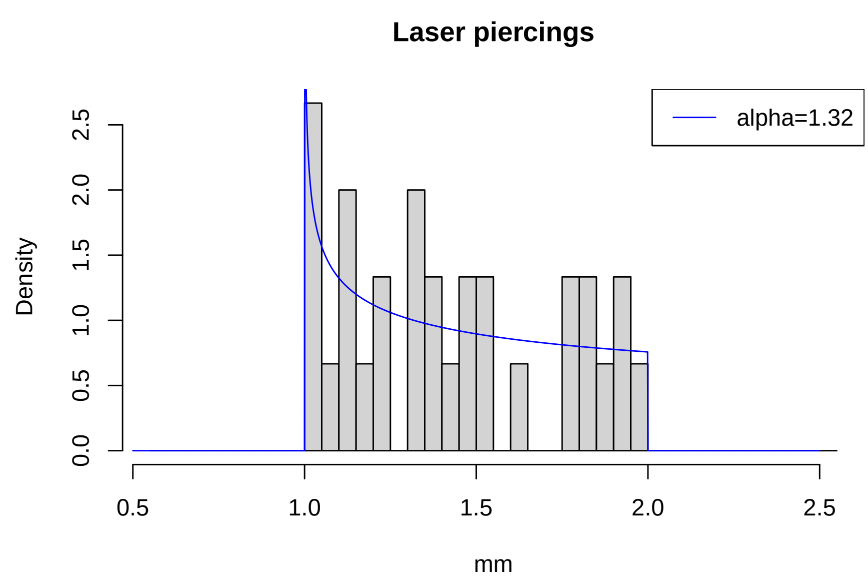
$$(X_1, X_2, ...X_n)$$

*Estimando los parámetros con los datos*

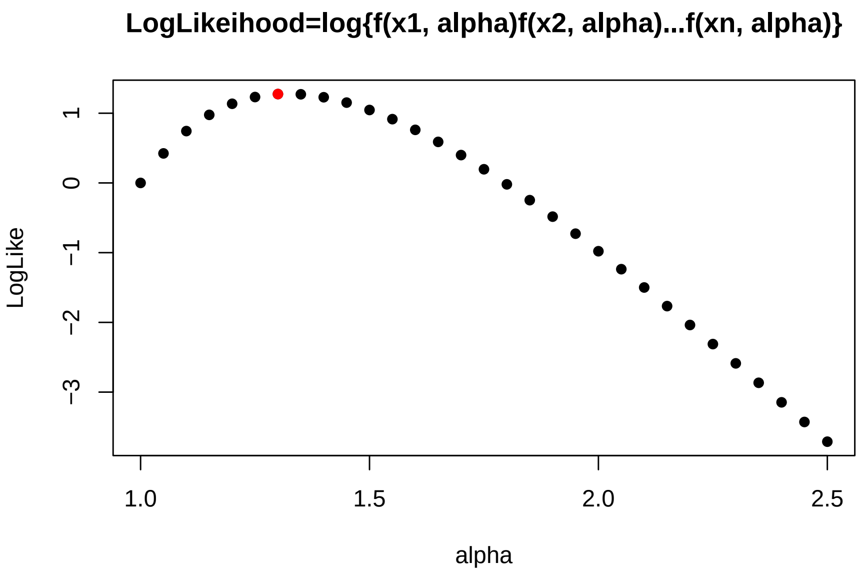
En nuestro ejemplo, tenemos entonces la observación de la muestra aleatoria como un conjunto de 30 números $(x_1, x_2, ...x_{30})$, por lo tanto sustituimos los números en el estimador y esto nos dará su valor observado.

$\hat{\alpha}_{ml} = -\frac{1}{30}\{\ln(1.11 - 1) + \ln(1.64 - 1) + ... \ln(1.04 - 1)\} = 1.320$

Por lo tanto, la estimación de máxima verosimilitud del parámetro es 1.320. Si sustituimos este valor en la función de probabilidad y lo superponemos con el histograma, podemos ver que nos da una descripción adecuada de los datos.



Veamos la función del logaritmo de la verosimilitud para nuestros 30 cortes láser. Recuerda, los datos son fijos para nuestro experimento y $\alpha$ varía. La función tiene un máximo. Sin embargo, si tomamos otra muestra, esta función cambia y también su lo hará su máximo.




**Máxima verosimilitud: Historia**

Para inferir la verdadera posición de Ceres en un momento dado, Gauss derivó la función de error

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

Donde la posición **verdadera** de Ceres era la media $\mu$. ¿Cómo podemos combinar los datos para tener la mejor estimación de la posición de Ceres?

¿Cuál es la estadística que mejor puede describir su posición?

Esta pregunta se puede formular como: ¿Cuál es la estimación de máxima verosimilitud de $\mu$ para una variable aleatoria normal?

**Máxima verosimilitud de la distribución normal**

Para una variable aleatoria normal

$$X\rightarrow N(\mu,\sigma^2)$$

.

¿Cuáles son los estimadores de $\mu$ y $\sigma^2$ que maximizan la probabilidad de los datos observados?

Seguimos el método de máxima verosimilitud:

1. La función de verosimilitud, o la probabilidad de haber observado la muestra $(x_1, ....x_n)$ es

$L(\mu, \sigma^2) = \Pi_{i=1..n} f(x_i; \mu, \sigma)$

$$= (\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2}$$

2. Tomamos el logaritmo de $L$ y calculamos la función **log-verosimilitud**

$$\ln L(\mu, \sigma^2) = -n\ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\Sigma_i(x_i-\mu)^2$$

Las estimaciones de $\mu$ y $\sigma^2$ son donde la probabilidad es máxima. Dan la probabilidad más alta para los datos de la muestra.

3. Diferenciamos con respecto a $\mu$ y $\sigma^2$. Estas dos derivadas nos dan dos ecuaciones, una para cada uno de los parámetros. Para derivar con respecto a $\sigma^2$, es más fácil hacer una sustitución $t = \sigma^2$.

a) $\frac{d\ln L(\mu,\sigma^2)}{d\mu} = \frac{1}{\sigma^2}\sum_i (x_i - \mu)$

b) $\frac{d\ln L(\mu,\sigma^2)}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_i (x_i - \mu)^2$

Las derivadas son 0 en el máximo

a) $\frac{1}{\hat{\sigma}^2}\sum_i (x_i - \hat{\mu}) = 0$
b) $-\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\sum_i (x_i - \hat{\mu})^2 = 0$

resolviendo ambas ecuaciones para los parámetros, encontramos para $\mu$

$$\hat{\mu}_{ml} = \frac{1}{n}\sum_i x_i = \bar{x}$$

y para $\sigma^2$

$$\hat{\sigma}^2_{ml} = \frac{1}{n}\sum_i (x_i - \bar{x})^2$$

Por lo tanto la **media mestral** o promedio $\bar{X}$ es el estimador de máxima verosimilitud de la media $\mu$ de la población. Gauss demostró que la estadística en la que más deberíamos confiar (la que tienen la mayor verosimilitud) para la posición real de Ceres era el **promedio**. Gauss, al resolvier la posición de Ceres, no solo descubrió la distribución normal, sino que también creó el análisis de regresión y mostró la importancia del promedio. Es debido a él que usamos el promedio para muchas cosas, y no otro tipo de estadísticas.

Además, el estimador de máxima verosimilitud de $\sigma^2$ es un estimador **sesgado** porque se puede demostrar que

$$E(\hat{\sigma}^2_{ml}) = \sigma^2 + \frac{\sigma^2}{n} \neq \sigma^2$$

Fue Fisher quien demostró que a pesar de ser sesgado este estimador es importante, ya que lo usó para generalizar el teorema del límite central, donde pierde su sesgo en $n \to \infty$.

## 12.6 Método de los Momentos

El método de máxima verosimilitud tiene como objetivo producir los estimadores de distribuciones de probabilidad a partir de datos. Sin embargo, existe otra forma de producir esos estimadores, que se basa en la idea frecuentista de las probabilidades.

Hbíamos visto que las frecuancias relativas tienden a las probabilidades cuando $n$ es grande $f_i \to f(x_i)$, y como consequencia

$$\bar{x} \to \mu$$

El centro de gravedad de los datos tiende al centro de gravedad de la probabilidad. El método de los momentos dice que podemos tomar el valor **observado** de la media meustral $\bar{X}$ como estimador de $E(X) = \mu$

$$E(X) \sim \bar{x} = \hat{\mu}$$

Es decir que la variable aleatoria que estima la media de la población es el promedio:

$$\bar{X} = \frac{1}{n} \sum_i X_i$$

que también se denomina el primer **momento de muestra**

En general, si $X \to f(x, \theta)$ el estimador del parámetro $\theta$ se obtiene entonces de la ecuación:

$$E(X; \hat{\theta}) = \bar{x}$$

debido a que el valor esperado de la variable aletoria siempre es función del parámetro $\theta$.

**Ejemplo (exponencial)**

Si una variable aleatoria sigue una distribución exponencial

$$X \hookrightarrow exp(\lambda)$$

entonces podemos usar el método de los momentos para estimar $\lambda$. El método consta de tres pasos:

1. Calcular el valor esperado de la variable

$$E(X;\lambda) = \mu$$

2. Escribir la ecuación donde el valor esperado es igual al primer momento muestral

$$\frac{1}{\bar{\lambda}} = \bar{x}$$

3. Resolver para el parámetro

$$\hat{\lambda} = \frac{1}{\bar{x}}$$

En términos de datos, esto es $\hat{\lambda} = (\frac{1}{n}\sum_i x_i)^{-1}$. Es decir que si queremos estimar el parámetro $\lambda$ de una variable exponencial de un experimento aleatorio, debemos tomar una muestra aleatoria, sacar su promedio y tomar su inversa. El resultado es el estimador de parámetro que después, junto al modelo, lo podemos usar para calcular las probabilidades de observaciones futuras.

**Ejemplo (Baterías)**

Supongamos que tenemos varias baterías (nuevas y viejas) que cargamos durante el período de 1 hora. Medimos el estado de carga de la batería, siendo 1 un 100% de carga.

El estado de carga de una batería es una variable aleatoria que puede tener una distribución uniforme, donde no sabemos el valor mínimo que puede tomar $x$, pero sabemos que el máximo es 1 (100% de carga)

$$f(x) = \begin{cases} \frac{1}{1-a}, & \text{if } x \in (a,1) \\ 0, & x \notin (a,1) \end{cases}$$

¿Cuál es el estimador de $a$ (la carga mínima después de una hora)?

Si ejecutamos un experimento y obtenemos $x_1, ...x_n$, nos preguntamos ¿cómo podemos estimar $a$ a partir de los datos?

Seguimos los tres pasos del método de los momentos:

1. Calculamos el valor esperado de la variable aleatoria

$$E(X) = \frac{a+1}{2}$$

2. Obtenemos la ecuación para $\hat{a}$ donde igualamos el valor esperado al primer momento muestral

$$\frac{\hat{a}+1}{2} = \bar{x}$$

3. Resolvemos para el estimador $\hat{a}$

$$\hat{a} = 2\bar{x} - 1$$

Este es el estimador de la carga mínima que podemos observar.

Ten en cuenta si tomaramos el mínimo de las observaciones esto sería claramente subóptimo. El método nos dio una respuesta inteligente que también se puede resumir en los siguientes pasos

a) Podemos calcular $\bar{x}$ con precisión creciente dada por $n$
b) Sabemos que ninguna medida supera $b = 1$
c) Luego calculamos la distancia entre $\bar{x}$ y $b$ que es $1 - \bar{x}$
d) Esta distancia la restamos al promedio $\bar{x}$ para estimar el valor mínimo de carga:

$$\bar{x} - (1 - \bar{x}) = 2\bar{x} - 1$$

Esta debería ser nuestra mejor suposición para $\hat{a}$. Como tal llegamos a la misma estimación dada por el método de los momentos.

## 12.7 Método de Momentos para varios parámetros

El método dice que se puede encontrar un estimador para el parámetro $\theta$ de $f(x;\theta)$ a partir de la ecuación:

$$E(X) = \frac{1}{n}\sum_i x_i$$

Si hay más parámetros, usamos los **momentos de muestra** más altos. Consideremos que el segundo momento muestral es

$$\frac{1}{n}\sum_i X_i^2$$

Por lo tanto, una observación de este momento es cercana a $E(X^2)$

$$E(X^2) = \frac{1}{n}\sum_i x_i^2$$

El método para dos parámetros dice que se puede encontrar una estimación para los parámetros $\theta_1$ y $\theta_2$ de $f(x;\theta_1,\theta_2)$ a partir de las ecuaciones:

a. $E(X) = \frac{1}{n}\sum_i x_i$

b. $E(X^2) = \frac{1}{n}\sum_i x_i^2$

Podemos tener tantas ecuaciones como parámetros necesitemos calcular, incrementando el grado de los momentos, es decir las potencias de $X$.

**Ejemplo (Distribución normal)**

Si $X$ se distribuye normalmente, tenemos dos parámetros para estimar

$$X \to N(\mu, \sigma^2)$$

Seguimos los pasos del método de los momentos para dos parámetros:

1. Calculamos el valor esperado de la variable

$$E(X) = \mu$$

y el valor esperado de $X^2$

$$E(X^2) = \sigma^2 + \mu^2$$

$E(X^2)$ se sigue de la propiedad: $E(X^2) = V(X) + \mu^2$

2. Obtenemos las ecuaciones para los parámetros donde hacemos (a) el valor esperado de la variable igual al primer momento muestral, y (b) el valor esperado del segundo momento igual al segundo momento muestral

a. $E(X)$ se estima por

$$\hat{\mu} = \frac{1}{n}\sum_i x_i$$

b. $E(X^2)$ se estima por

$$\hat{\sigma}^2 - \hat{\mu}^2 = \frac{1}{n}\sum_i x_i^2$$

3. Resolvemos los parámetros

La primera ecuación da directamente el estimador de la media $\mu$.

$$\hat{\mu} = \frac{1}{n}\sum_i x_i$$

Que de nuevo es el promedio. De la segunda ecuación obtenemos

$$\hat{\sigma}^2 = \frac{1}{n}\sum_i x_i^2 - \hat{\mu}^2$$

que también se puede escribir como:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_i (x_i - \hat{\mu})^2$$

Encontramos que el método de los momentos y las estimaciones de máxima verosimilitud para la distribución normal son iguales. Sin embargo, este no siempre es el caso.

**Ejemplo (láser)**

¿Cuál es el estimador del parámetro $\alpha$ para el corte láser dado por el método de los momentos?

$$f(x;\alpha) = \begin{cases} \frac{1}{\alpha}(x-1)^{\frac{1}{\alpha}-1}, & \text{if } x \in (1,2) \\ 0, & x \notin (1,2) \end{cases}$$

Donde $\alpha$ es un parámetro.

El método dice que se puede encontrar un estimador para el parámetro $\alpha$ de $f(x;\alpha)$ a partir de la ecuación:

$$E(X) = \frac{1}{n}\sum_i x_i$$

por $\hat{\alpha}$

1. Calculamos el valor esperado $E(X)$

$$E(X) = \int_{-\infty}^{\infty} x f(x;\alpha) dx$$

Considera un cambio de variables $Z = X - 1$ entonces $E(X) = E(Z) + 1$ y

$E(Z) = \frac{1}{\alpha}\int_0^1 z z^{\frac{1-\alpha}{\alpha}} dz = \frac{1}{\alpha}\int_0^1 z^{1+\frac{1-\alpha}{\alpha}} dz$

$= \frac{1}{\alpha} \frac{z^{2+\frac{1-\alpha}{\alpha}}}{2+\frac{1-\alpha}{\alpha}} |_0^1 = \frac{1}{1+\alpha}$

Por lo tanto,

$$E(X) = E(Z+1) = \frac{1}{1+\alpha} + 1$$

2. Obtenemos la ecuación para $\hat{\alpha}_m$ donde igualamos el valor esperado al primer momento muestral. Sustituyendo $\hat{\alpha}_m$, el método de los momentos nos da la ecuación

$$\frac{1}{1+\hat{\alpha}} + 1 = \bar{x}$$

3. Resolvemos para $\hat{\alpha}$

$$\hat{\alpha}_m = \frac{1}{\bar{x}-1} - 1$$

4. Calculamos el valor de nuestros datos

$$\hat{\alpha}_m = 1.314$$

Tenga en cuenta que este es un ejemplo para el cual las estimaciones por máxima verosimilitud y el método de momentos son **diferentes**.

La estimación de máxima verosimilitud fue:

$$\hat{\alpha}_{ml} = -\frac{1}{n}\sum_{i=1}^{n} \ln(x_i - 1) = 1.320$$

El método de estimación de momentos fue:

$$\hat{\alpha}_m = \frac{1}{\bar{x}-1} - 1 = 1.314$$

Necesitamos estudios de **simulación**, donde **sabemos** el verdadero valor del parámetro $\alpha$, para encontrar cuál de estas estadísticas tiene menos error.

Nota: los datos para perforaciones con láser de 30 se simularon con $\alpha = 2$, por lo tanto, debemos preferir la estimación de máxima verosimilitud. Para obtener mejores estimaciones de $\alpha$ necesitamos aumentar el tamaño de la muestra.

## 12.8 Ejercicios Máxima Verosimilitud

#### 12.8.0.1 Ejercicio 1

Toma una variable aleatoria con la siguiente función de densidad de probabilidad

$$f(x) = \begin{cases} (1+\theta)x^{\theta}, & \text{if } x \in (0,1) \\ 0, & x \notin (0,1) \end{cases}$$

- ¿Cuál es la estimación de máxima verosimilitud para $\theta$?
- Si tomamos una muestra de 5 obsevaciones $x_1 = 0.92$; $x_2 = 0.79$; $x_3 = 0.90$; $x_4 = 0.65$; $x_5 = 0.86$

¿Cuál es el valor estimado del parámetro $\theta$?

- Calcula $E(X) = \mu$ en función de $\theta$. ¿Cuál es la estimación de máxima verosimilitud para $\mu$?

#### 12.8.0.2 Ejercicio 2

Para una variable aleatoria con una función de probabilidad binomial

$$f(x;p) = \binom{n}{x} p^x (1-p)^{n-x}$$

- ¿Cuál es el estimador de máxima verosimilitud de $p$ para una muestra de tamaño 1 de esta variable aleatoria?
- ¿Es el estimador insesgado y/o consistente?
- En **un** examen de 100 estudiantes observamos $x_1 = 68$ estudiantes que aprobaron el examen. ¿Cuál es la estimación de máxima verosimilitud para la probabilidad de pasar el examen?

#### 12.8.0.3 Ejercicio 3

Toma una variable aleatoria con la siguiente función de densidad de probabilidad

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

- ¿Cuál es la estimación de máxima verosimilitud para $\lambda$?
- Si tomamos una muestra de 5 observaciones $x_1 = 0.223$ $x_2 = 0.681$; $x_3 = 0.117$; $x_4 = 0.150$; $x_5 = 0.520$

¿Cuál es el valor estimado del parámetro $\lambda$?

- ¿Cuál es la estimación de máxima verosimilitud de la varianza de la variable exponencial $\sigma^2 = \frac{1}{\lambda^2}$?

- ¿Es $\hat{\sigma}^2$ un estimador insesgado y consistente de $\sigma^2$?

## 12.9 Método de los momentos

#### 12.9.0.1 Ejercicio 1

¿Cuáles son los estimadores de los siguientes modelos paramétricos dados por el método de los momentos?

| Model | f(x) | E(X) |
|---|---|---|
| Bernoulli | $p^x(1-p)^{1-x}$ | $p$ |
| Binomial | $\binom{n}{x}p^x(1-p)^{n-x}$ | $np$ |
| Geometrica | $p(1-p)^{x-1}$ | $\frac{1}{p}$ |
| Binomial Negativa | $\binom{x+r-1}{x}p^r(1-p)^x$ | $r\frac{1-p}{p}$ |
| Poisson | $\frac{e^{-\lambda}\lambda^x}{x!}$ | $\lambda$ |
| Exponencial | $\lambda e^{-\lambda x}$ | $\frac{1}{\lambda}$ |
| Normal | $\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ | $\mu$ |

#### 12.9.0.2 Ejercicio 2

Toma una variable aleatoria con la siguiente función de densidad de probabilidad

$$f(x) = \begin{cases} (1+\theta)x^\theta, & \text{if } x \in (0,1) \\ 0, & x \notin (0,1) \end{cases}$$

- Calcula $E(X)$ como una función de $\theta$
- ¿Cuál es la estimación de $\theta$ utilizando el método de los momentos?
- Si tomamos una muestra de 5 observaciones $x_1 = 0,92; \quad x_2 = 0,79; \quad x_3 = 0,90; \quad x_4 = 0,65; \quad x_5 = 0,86$

¿Cuál es el valor estimado del parámetro $\theta$?

#### 12.9.0.3 Ejercicio 3

Considera una variable aleatoria discreta $X$ que sigue una distribución binomial negativa con función de masa de probabilidad:

$$f(x) = \binom{x+r-1}{x}p^r(1-p)^x$$

Dado que

- $E(X) = \dfrac{r(1-p)}{p}$

- $V(X) = \dfrac{r(1-p)}{p^2}$

calcular:

- Una estimación del parámetro $r$ y una estimación del parámetro $p$ obtenidas a partir de una muestra aleatoria de tamaño $n$ por el método de los momentos.
- Los valores de las estimaciones de $r$ y $p$ para la siguiente muestra aleatoria:

$$x_1 = 27; \qquad x_2 = 8; \qquad x_3 = 22; \qquad x_4 = 29; \qquad x_5 = 19; \qquad x_5 = 32$$

# Chapter 13

# Intervalos de confianza

## 13.1 Objetivo

En este capítulo, presentaremos el concepto de **intervalos de confianza** para medias, proporciones y varianzas.

Derivaremos las fórmulas para los intervalos de confianza bajo diferentes condiciones, como normalidad con varianza conocida y desconocida, y $n$ grande.

## 13.2 Estimación de la media

Hemos visto que siempre que tomamos una muestra aleatoria $(X_1, X_2, ...X_n)$, la media muestral

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

es un estimador de la media $\mu$ de la variable aleatoria $X$. El estimador es **insesgado**

- $E(\bar{X}) = \mu$

y **consistente**

- $V(\bar{X}) = \frac{\sigma^2}{n}$

donde $\sigma^2$ es la varianza de $X$. Llamamos a $se = \frac{\sigma}{\sqrt{n}}$ el **error estándar**.

cuando tomamos los valores de la muestra aleatoria, tomamos el valor de $\bar{x}$ como el valor de la media. Eso es

$$\bar{x} = \hat{\mu}$$

Como $\bar{X}$ es una variable aleatoria, la estimación de la media **cambia** cuando tomamos **otra muestra**.

## 13.3 Margen de error

Al decidir si el **error** en la estimación

$$\bar{X} - \mu$$

es grande o no, generalmente lo comparamos con una tolerancia predefinida. **Si sabemos** que la distribución de $X$ es normal $X \rightarrow N(\mu, \sigma^2)$ y el valor de $\mu$, podemos calcular qué tan lejos cae la estimación de $\bar{x}$ de $\mu$.

Definimos el **margen de error** al nivel de 5% como la distancia $m$ tal que la distribución de $\bar{X}$ captura 95% de las estimaciones:

$$P(-m \leq \bar{X} - \mu \leq m) = P(\mu - m \leq \bar{X} \leq \mu + m) = 0.95$$

Si $\bar{X}$ se distribuye normalmente, entonces el margen de error es

$$m = z_{0.025}\frac{\sigma}{\sqrt{n}} = 1.96 \times se$$

donde $z_{0.025} = \phi^{-1}(0.975) = \text{qnorm}(0.975)$

**Ejemplo (cables):**

Tomamos una muestra aleatoria de tamaño 8: Cargamos un cable hasta que se rompan y registramos la carga de rotura.

**Si sabemos** que la rotura de los cables realmente se distribuyen normalmente

$$X \rightarrow N(\mu = 13, \sigma^2 = 0.35^2)$$

entonces la media muestral es normal

$$\bar{X} \rightarrow N(13, \frac{0.35^2}{8})$$

Con media $E(\bar{X}) = 13$ y error estándar $se = \frac{0.35}{\sqrt{8}} = 0.1237$

Por lo tanto, el margen de error en 95% es

$$m = z_{0,025}\frac{\sigma}{\sqrt{n}} = 1.96 \times se = 1.96\frac{0.35}{\sqrt{8}} = 0.24$$

Ahora, tomamos la muestra aleatoria y encontramos los resultados.

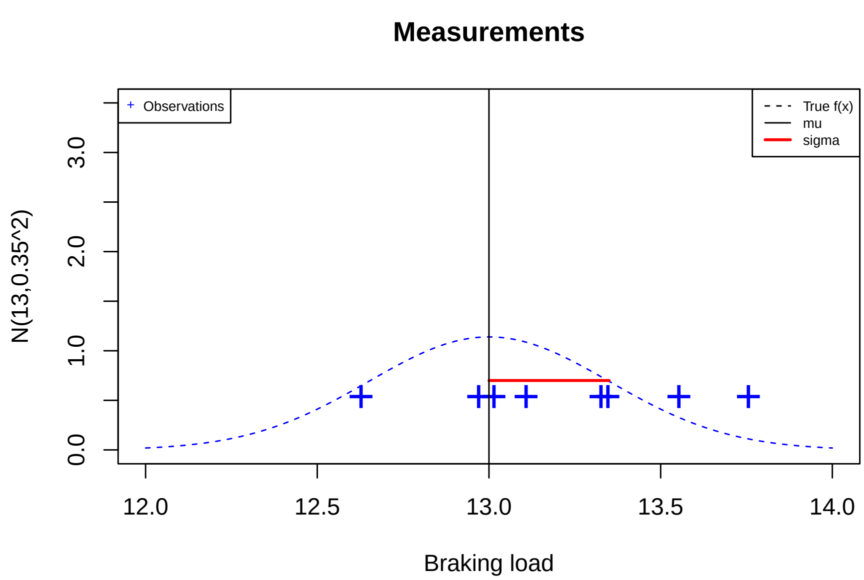
```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```

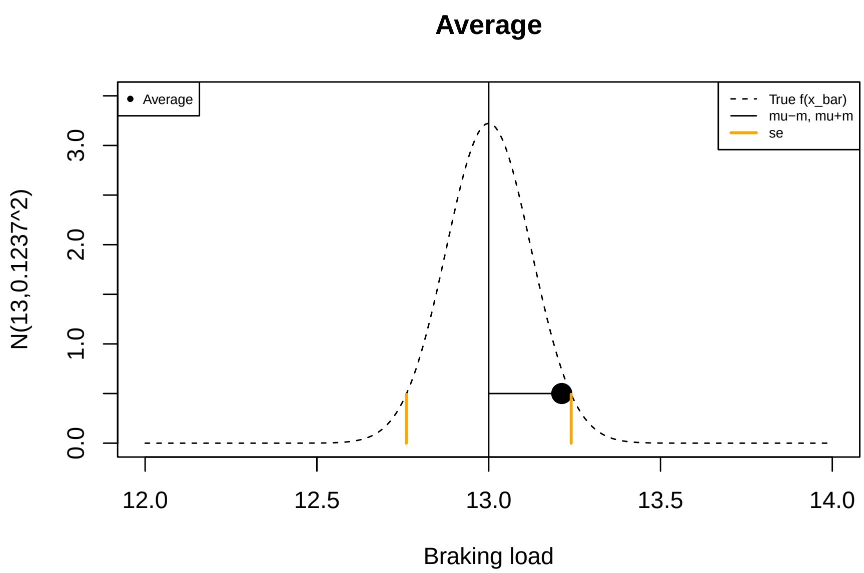
El promedio observado es $\bar{x} = 13.21$, y el error que cometeríamos si reemplazamos $\mu$ por $\bar{x}$, sería

$$\bar{x} - \mu = 13.21 - 13 = 0.21$$

El **error observado** está dentro del margen de error

$$\bar{x} - \mu < m$$



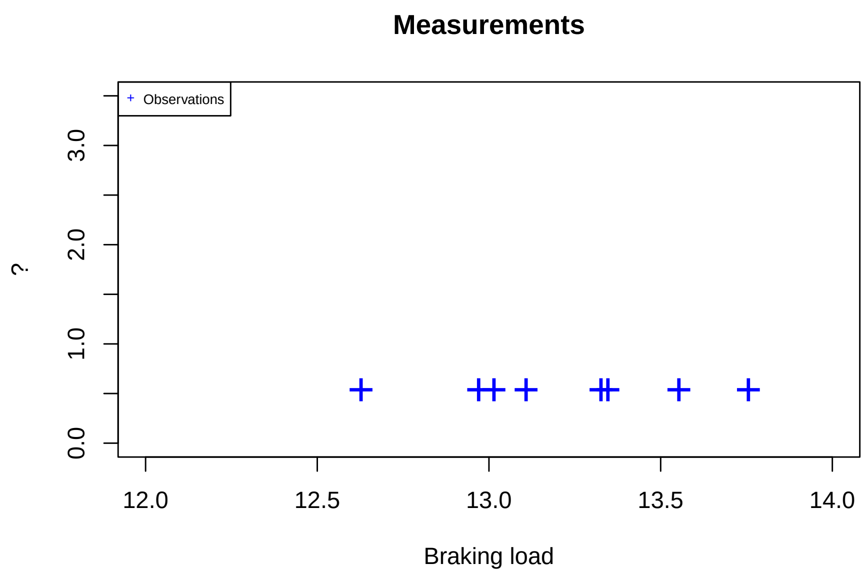



## 13.4 Estimación de intervalo para la media

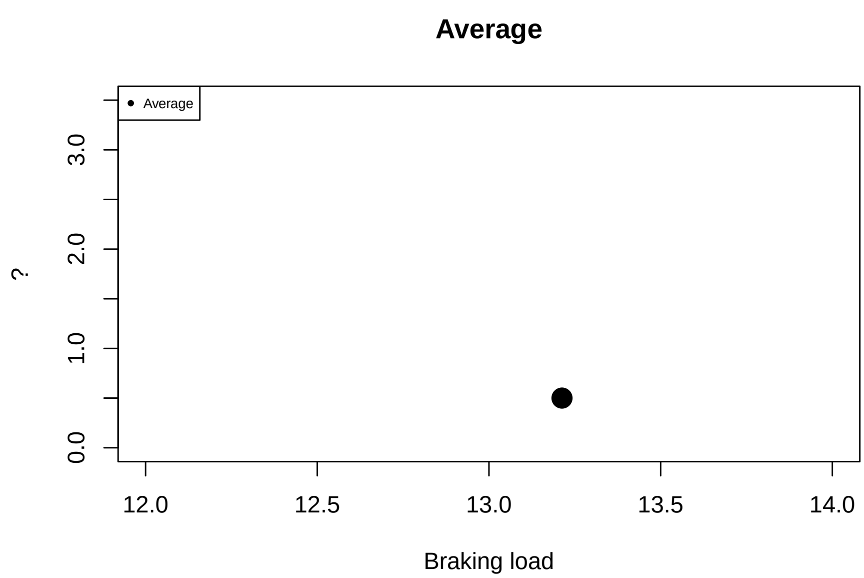
El problema es que en la vida real **no sabemos** los valores reales de $\mu$ o $\sigma$ para

$$X \to N(\mu, \sigma^2)$$

Empezamos tomando la muestra




y luego calcular la media




¿Cuál es el valor de $\mu$?

Nuestros datos sugieren que es $\bar{x} = 13.21$. Reemplazar $\mu$ por $\bar{x}$ se denomina

**estimación puntual** del parámetro. Pero ¿qué tan **seguros** estamos al hacer este remplazo? después de todo, sabemos que estamos cometiendo un error, pero no sabemos qué tan grande es.

Definimos el intervalo de confianza para $\mu$. De la ecuación del margen de error

$$P(-m \leq \bar{X} - \mu \leq m) = 0.95$$

resolvamos para $\mu$ que es **la verdadera incógnita**

$$P(\bar{X} - m \leq \mu \leq \bar{X} + m) = 0.95$$

Los límites izquierdo y derecho de la desigualdad son variables aleatorias que motivan la definición del **intervalo de confianza aleatorio en** 95%**:**

$$(L, U) = (\bar{X} - m, \bar{X} + m)$$

Este intervalo es una nueva **variable aleatoria** y tiene por definición una probabilidad de 0.95 de contener $\mu$.

El **intervalo observado** que obtenemos del experimento es (en minúsculas)

$$(l, u) = (\bar{x} - m, \bar{x} + m)$$

Este intervalo contiene o no el parámetro $\mu$: ¡**nunca lo sabremos**!

Decimos que tenemos una confianza de 95% en que el intervalo $(l, u)$ capturará el verdadero parámetro desconocido $\mu$. Piensa en comprar un billete de lotería del raspa y gana pero que no podemos raspar para ver el premio. El billete tiene o no el premio sólo que no lo sabemos.

### 13.4.1 Caso 1 (varianza conocida)

Los intervalos de confianza se pueden estimar en diferentes casos. El primer caso es cuando

1. $X$ es una variable normal
2. y conocemos el valor de $\sigma$

el intervalo de confianza en 95% es

$$(l, u) = (\bar{x} - m, \bar{x} + m)$$

donde

$$m = z_{0.025} \frac{\sigma}{\sqrt{n}}$$

Eso es:

$$(l, u) = (\bar{x} - z_{0.025}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0.025}\frac{\sigma}{\sqrt{n}})$$

**Ejemplo (cables):**

En nuestro ejemplo, asumimos que $X$ se distribuye normalmente y que sabemos $\sigma^2 = 0.35^2$.

1. Dado que $\bar{X}$ es normal, el margen de error es

$$m = z_{0.025}\frac{\sigma}{\sqrt{n}}$$

2. Como sabemos $\sigma^2 = 0.35^2$, entonces el intervalo de confianza de 95% es

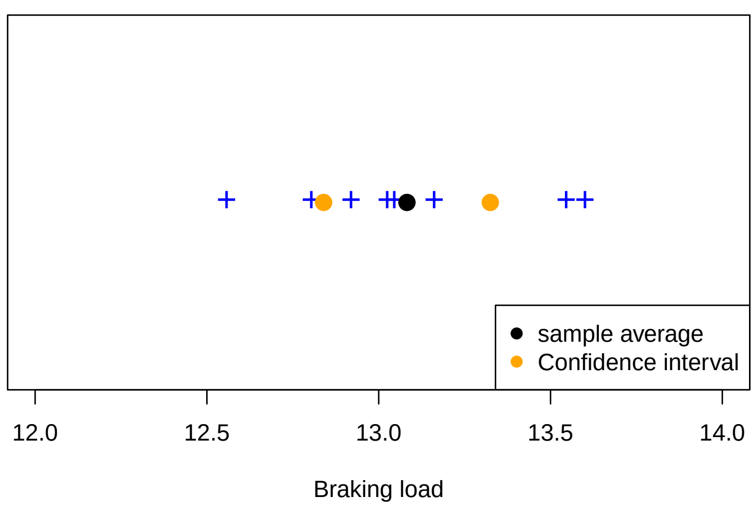
$(l, u) = (\bar{x} - m, \bar{x} + m) =$

$$(\bar{x} - z_{0.025}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0.025}\frac{\sigma}{\sqrt{n}}) = (12.97, 13.45)$$

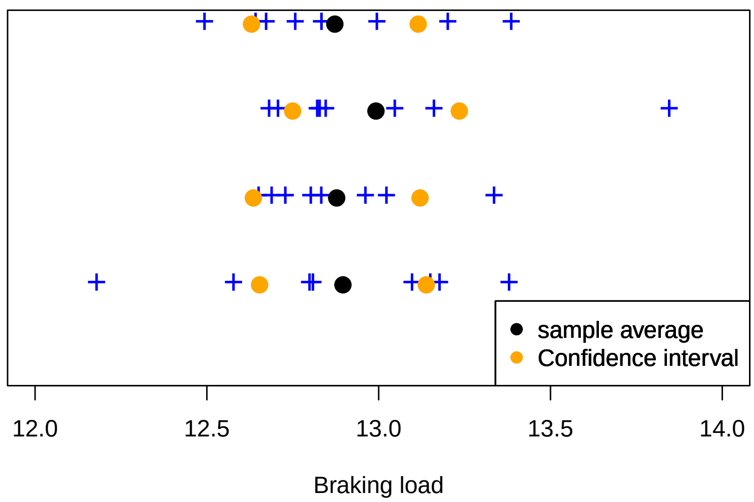
o también podemos escribirlo como

$$\hat{\mu} = \bar{x} \pm m = 13.21 \pm 0.24$$

Esto significa que, al estimar la media por el promedio, confiamos en las unidades pero no tanto en los lugares decimales.



Recuerda que el intervalo de confianza $(l, u)$ es una observación del intervalo de confianza aleatorio $(L, U)$. Por lo tanto, cada vez que obtenemos una nueva muestra entonces $(l, u)$ cambia. Si realizamos muestras de 100 de tamaño $n$ entonces 95 de los intervalos de confianza contendrán $\mu$, ¡pero no sabemos cuál!




### 13.4.2 Nivel de confianza

Podemos cambiar nuestra confianza de 95% a 99%. Cuando calculamos el margen de error en 95%, dejamos por fuera $\alpha = 0.05$ de probabilidad, 0.025 a cada lado.

Ahora, podemos dejar fuera $\alpha = 0.01$ de probabilidad, 0.005 a cada lado. Por lo tanto, el intervalo de confianza de 99% es

$(l, u) = (\bar{x} - z_{0,005}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0,005}\frac{\sigma}{sqrtn})$

$$= (\bar{x} - 2.58\frac{\sigma}{\sqrt{n}}, \bar{x} + 2.58\frac{\sigma}{\sqrt{n}})$$

donde $z_{0.005} = \phi^{-1}(0.995) = \text{qnorm}(0.995)$. También podemos escribirlo como

$$\hat{\mu} = \bar{x} \pm 2.58\frac{\sigma}{\sqrt{n}}$$

Para nuestros cables, el intervalo de confianza de 99% es

$$\hat{\mu} = 13.21 \pm 0.31$$

Si queremos tener más confianza, ¡necesitamos intervalos de confianza más grandes!

**Ejemplo (Energía de impacto):**

Un material metálico se prueba por impacto para medir la energía requerida para cortarlo a una temperatura dada. Se cortaron diez probetas de acero A238 a 60ºC con las siguientes energías de impacto (J):

$64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, 64.3$

Si **suponemos** que la energía del impacto se distribuye normalmente con $\sigma = 1J$, ¿cuál es el intervalo de confianza de 95% para la media de estos datos?

Sabemos

1. $X \to N(\mu, \sigma^2)$
2. $\sigma = 1J$
3. $\alpha = 0.05$ (el límite de confianza)

El intervalo de confianza de 95% es entonces

$(l, u) = (\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}})$

$$= (64, 46 - 1.96\frac{1}{\sqrt{10}}, 64, 46 + 1.96\frac{1}{\sqrt{10}}) = (63, 84, 65, 08)$$

o

$$\hat{\mu} = 64.46 \pm 0.61$$

esto nos dice que podemos estar seguros del primer dígito (6), algo seguros del segundo (4) e inseguros de los decimales (46).

¿Qué pasa si $\sigma^2$ es **desconocido**?

## 13.5 Margen de error para varianza desconocida

Pudimos calcular el intervalo de confianza $(l, u) = (\bar{x} - m, \bar{x} - m)$ porque pudimos encontrar el margen de error

$$m = 1.96\frac{\sigma}{\sqrt{n}}$$

desde que **sabíamos** $\sigma$. $\sigma$ es un parámetro de la distribución que normalmente **no conocemos**, así cmo $\mu$. Para encontrar el margen de error con varianza **desconocida**, necesitamos el siguiente teorema, debido a Gosset

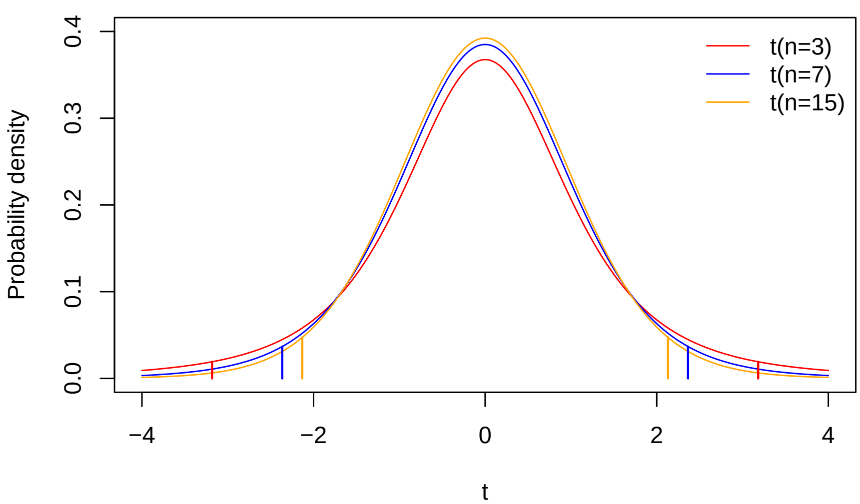
### 13.5.1 Thorem (estadística T)

Cuando $X$ es normal, entonces la estadística estandarizada

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

Sigue una distribución $t$ con $n-1$ grados de libertad, donde $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Por lo tanto, podemos calcular probabilidades para $\bar{X}$, incluso si no conocemos $\sigma$.

Veamos algunas densidades de probabilidad en la familia de las distribuciones $T$.




Ahora necesitamos volver a calcular el margen de error $m$ al nivel de 5% cuando usamos la distribución $t$

$P(\mu - m \leq \bar{X} \leq \mu + m)$

$$= P(-\frac{m}{s/\sqrt{n}} \leq T \leq \frac{m}{s/\sqrt{n}}) = 0,95$$

$$m = t_{0.025,n-1}\frac{s}{\sqrt{n}}$$

donde $t_{0.025,n-1}$ es el valor $T$ que deja 2.5% de probabilidad en el lado derecho de la distribución $t$ con $n-1$ grados de libertad.

### 13.5.2 Caso 2 (varianza desconocida)

El segundo caso para calcular los intervalos de confianza es más realista. Si

1. $X$ es una variable normal

el intervalo de confianza en 95% es

$$(l,u) = (\bar{x} - m, \bar{x} + m)$$

donde

$$m = t_{0.025,n-1}\frac{s}{\sqrt{n}}$$

Eso es:

$$(l,u) = (\bar{x} - t_{0.025,n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{0.025,n-1}\frac{s}{\sqrt{n}})$$

donde $t_{0.025,n-1} = F^{-1}(0.975)$=qt(0.975, n-1)

**Ejemplo (Energía de impacto):**

Un material metálico se prueba por impacto para medir la energía requerida para cortarlo a una temperatura dada. Se cortaron diez probetas de acero A238 a 60ºC con las siguientes energías de impacto (J):

$64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, 64.3$

Si **suponemos** que la energía del impacto se distribuye normalmente pero **no conocemos** la varianza, ¿cuál es el intervalo de confianza de 95% para la media de estos datos?

Sabemos

- $\bar{x} = 64.46$
- $s = 0.227$
- $\alpha = 0.05$
- $t_{0.025,9} = 2.26$ obtenido de $t_{0.025,9} =$ qt(0.975, 9)

El intervalo de confianza es entonces

$(l,u) = (\bar{x} - t_{0.025,9}\frac{s}{\sqrt{n}}, \bar{x} + t_{0.025,9}\frac{s}{\sqrt{n}})$

$$= (64,46 - 2,26\frac{0,227}{\sqrt{10}}, 64,46 + 2,26\frac{0,227}{\sqrt{10}})$$
$$= (64,29, 64,62)$$

Tenga en cuenta que cuando asumimos $\sigma = 1$ el intervalo de confianza $(63.84, 65.08)$ era mayor. Por lo tanto, los datos sugieren que $\sigma < 1$.

En R, podemos calcular el intervalo de confianza con: `t.test(c(64.1,64.7,64.5,64.6,64.5,64.3,64.6,64.8,64.2,64.3))`

## 13.6 Estimación de proporciones

**Ejemplo (vacuna):**

Se seleccionó una muestra aleatoria de 400 pacientes para probar una nueva vacuna contra el virus de la influenza, después de 6 meses de vacunación, 136 estaban enfermos. ¿Cuál es la eficacia esperada de la vacuna?

Dado que cada vacunación $X_i$ es un ensayo de Bernoulli

$$X \rightarrow Bernoulli(p)$$

Con media $\mu = p$ y varianza $\sigma^2 = p(1-p)$.

La muestra es algo como

$$(x_1, x_2, x_3, ...x_n) = (0, 1, 0, ..1, 0)_{400}$$

con 136 unos y en un total de 400 repeticiones. La muestra tiene un promedio de $\bar{x} = \frac{1}{400}\sum_i^{400} x_i = 134/400 = 0.34$. Dado que la media muestral es un estimador insesgado de $\mu$, entonces podemos tener una estimación puntual para $p$

$$\hat{p} = \bar{x} = 134/400 = 0.34$$

Esto tiene sentido porque $\bar{x}$ es la frecuencia relativa observada para el número de **unos** en la muestra ($f_1$). Y como tal, es un estimador de la probabilidad de observar un uno en un ensayo de Bernoulli.

$$f_1 = \hat{P}(X = 1)$$

Esto es consistente con la definición frecuentista de probabilidad que vimos en el capítulo 2. Pero, ¿qué confianza tenemos en esta estimación? Queremos un intervalo de confianza para $p$

### 13.6.1 Caso 3 (proporciones)

1. Cuando $\hat{p}n > 5$ y $(\hat{p} - 1)n > 5$, la **estadística estandarizada** de $\bar{X}$ se puede aproximar mediante un estándar distribución (TCL)

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - p}{\left[\frac{p(1-p)}{n}\right]^{1/2}} \rightarrow N(0, 1)$$

y el intervalo de CI al 95% de $p$ es:

$$CI = (l, u) = (\bar{x} - z_{0.025}\left[\frac{\bar{x}(1-\bar{x})}{n}\right]^{1/2}, \bar{x} + z_{0.025}\left[\frac{\bar{x}(1-\bar{x})}{n}\right]^{1/2})$$

Donde estimamos la varianza de Bernoulli $\sigma = p(1-p)$ por $\hat{\sigma} = \bar{x}(1-\bar{x})$.

**Ejemplo (vacuna):**

En nuestro caso, estamos contando 136 fallos de inmunización de 400 inoculaciones.

sabemos

- $\bar{x} = 134/400 = 0.34$
- $z_{0.025} = 1.96$

Por lo tanto, el intervalo de confianza de 95% para $p$ es

$(l, u) = (\bar{x} - 1.96[\frac{\bar{x}(1-\bar{x})}{n}]^{1/2}, \bar{x} + 1.96[\frac{\bar{x}(1-\bar{x})}{n}]^{1/2})$

$$= (0.29, 0.39)$$

La probabilidad de fracaso de la vacuna es

$$\hat{p} = 0.34 \pm 0.05$$

Nota: Las encuestas de intención de voto (ensayo de Bernoulli) en una muestra de $n$ individuos reportan este tipo de estimación con su **margen de error**. No significa que el **valor verdadero** de $p$ esté dentro de este intervalo con una probabilidad de 95%. Significa que tenemos una confianza del 95% de haber atrapado al $p$ que representa esta muestra en particular.

## 13.7 Estimación de la varianza

Hemos visto que siempre que tomamos una muestra aleatoria $(X_1, X_2, ...X_n)$, la varianza muestral

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

es un estimador de la media $\sigma^2$ de la variable aleatoria $X$. El estimador es **insesgado**

- $E(S^2) = \sigma^2$

y también es **consistente**. Cuando tomamos los valores de la muestra aleatoria, tomamos el valor de $S^2$ como el valor de la varianza; eso es

$$s^2 = \hat{\sigma}^2$$

Dado que $S^2$ es una variable aleatoria, la estimación de la varianza cambia cuando tomamos otra muestra.

**Ejemplo (energía de impacto)**

Un material metálico se prueba por impacto para medir la energía requerida para cortarlo a una temperatura dada. Se cortaron diez probetas de acero A238 a 60ºC con las siguientes energías de impacto (J):

$64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, 64.3$

¿Cuál es la estimación de la varianza de estos datos?

$$s^2 = 0.05155556$$

sd(c(64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, 64.3))^2

¿Cuánta confianza tenemos en los decimales de la estimación?

## 13.8 Intervalo de confianza para la varianza

Para calcular un intervalo de confianza para la varianza, necesitamos una estadística que sea una función de $S^2$ y nos permita calcular probabilidades. Usaremos el siguiente teorema

### 13.8.1 Teorema ($\chi^2$):

Cuando $X$ es normal, entonces la estadística estandarizada

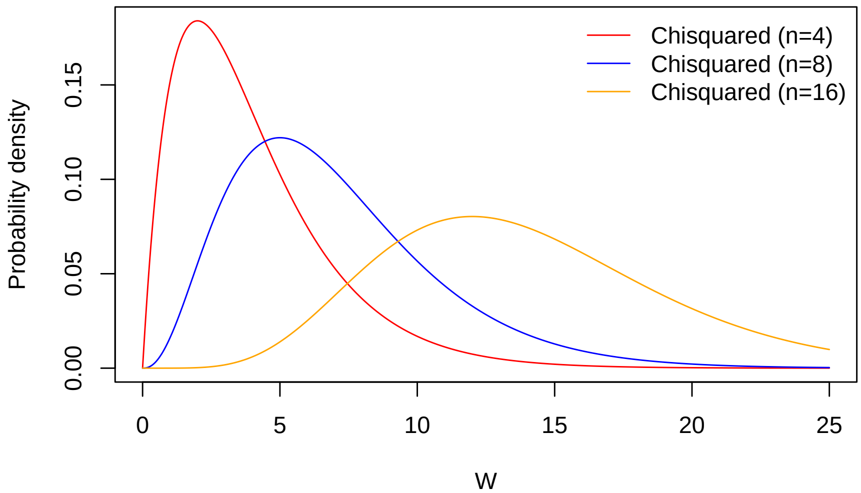
$$W = \frac{S^2(n-1)}{\sigma^2}$$

sigue una distribución $\chi^2$ con $n-1$ grados de libertad

$$\frac{S^2}{\sigma^2}(n-1) \to \chi^2_{n-1}$$

Por lo tanto, podemos calcular probabilidades para $W$.

Veamos algunas densidades de probabilidad en la familia de las distribuciones $\chi^2$.




### 13.8.2 Intervalo de confianza para la varianza

Buscamos el intervalo de confianza de $\sigma^2$ con confianza 95% $(L, U)$ tal que

$$P(L \leq \sigma^2 \leq U) = 0.95$$

Entonces podemos usar el $\chi^2$ para determinar el 95% de la distribución alrededor de $W$. Comencemos definiendo los valores que capturan los 95% de la distribución

$$P(\chi^2_{0.975,n-1} \leq W \leq \chi^2_{0.025,n-1}) = 0.95$$

Reemplazando el valor de $W$

$$P(\chi^2_{0.975,n-1} \leq \frac{S^2}{\sigma^2}(n-1) \leq \chi^2_{0.025,n-1}) = 0.95$$

y resolviendo para $\sigma^2$

$$P(\frac{S^2(n-1)}{\chi^2_{0.025,n-1}} \leq \sigma^2 \leq \frac{S^2(n-1)}{\chi^2_{0.975,n-1}}) = 0.95$$

Encontramos un intervalo aleatorio que captura $\sigma^2$ con 95% de confianza

$$(L,U) = (\frac{S^2(n-1)}{\chi^2_{0.025,n-1}}, \frac{S^2(n-1)}{\chi^2_{0.975,n-1}})$$

### 13.8.3 Caso 4 (varianza)

1. Cuando $X$ es una variable normal

El intervalo de confianza al 95% **observado** (minúsculas) es

$$(l,u) = (\frac{s^2(n-1)}{\chi^2_{0.025,n-1}}, \frac{s^2(n-1)}{\chi^2_{0.975,n-1}})$$

dónde

- $\chi^2_{0.975,n-1} = F^{-1}(0.025) =$ qchisq(0.025, df=n-1)
- $\chi^2_{0.025,n-1} = F^{-1}(0,975)$ =qchisq(0,975, df=n-1) para $n = 10$ o $df = n-1 = 9$

**Ejemplo (energía de impacto)**

Un material metálico se prueba por impacto para medir la energía requerida para cortarlo a una temperatura dada. Se cortaron diez probetas de acero A238 a 60ºC con las siguientes energías de impacto (J):

$64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, 64.3$

¿Cuál es el intervalo de confianza para la varianza de estos datos?

$$(l,u) = (\frac{s^2(n-1)}{\chi^2_{0.025,n-1}}, \frac{s^2(n-1)}{\chi^2_{0.975,n-1}})$$

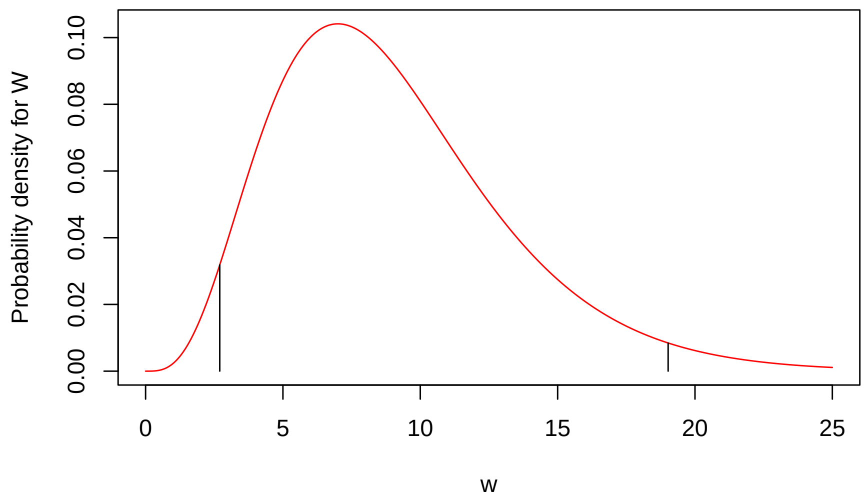
1. La desviación estándar de los datos es $s^2 = 0.05155556$
2. $n = 10$
3. Luego calculamos $\chi^2_{0.025,n-1}$ y $\chi^2_{0.975,n-1}$

```
chi0.975 <- qchisq(0.025, df=9)
chi0.975
```

```
[1] 2.700389
```

```
chi0.025 <- qchisq(0.975, df=9)
chi0.025
```

```
[1] 19.02277
```




Por lo tanto

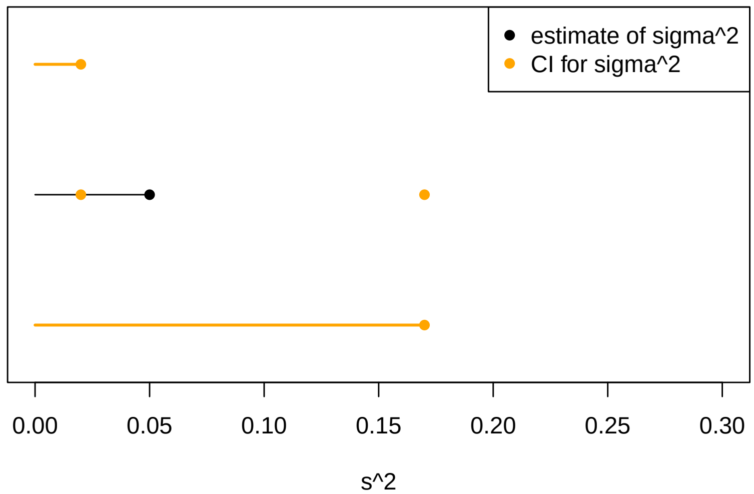
$$(l,u) = (\frac{0.227^2(10-1)}{19.02277}, \frac{0.227^2(10-1)}{2.700389}) = (0.02, 0.17)$$

Recuerda que cuando habíamos calculado el intervalo de confianza para la media, asumimos $\sigma^2 = 1$ (caso 1). Ahora podemos decir que esta elección no fue consistente con los datos porque vemos que el intervalo de confianza no contiene el valor $\sigma^2 = 1$ .

Según los datos, $\sigma^2 \neq 1$ con una confianza de 95%.

en R: biblioteca(Ecfun); confint.var(0.05155556, 9)

El intervalo para la varianza **no es simétrico** y no podemos formularlo como un margen de error con $\pm$.


estimate of sigma^2
CI for sigma^2
0.00
0.05
0.10
0.15
0.20
0.25
0.30
s^2