Estadística

Alejandro Caceres

2022-11-27

Contents

Chapter 1

Objetivo

- Este es el curso de introducción a la estadística de la EEBE (UPC).
- Las fechas de exámenes y material de estudio adicional se pueden encontrar en ATENEA

1.1 Lectura recomendada

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

Chapter 2

Descripción de datos

2.1 Objetivo

- Datos: discretos, continuos
- Resumir datos en tablas y figuras.

2.2 Estadísticas

- Resolver problemas de manera sistemática (ciencia, tecnología e ingeniería)
- ¡Los humanos modernos usamos un **método** general históricamente desarrollado durante miles de años! ... y aún en desarrollo.
- Tiene tres componentes principales: observación, lógica y generación de nuevo conocimiento.

2.3 Metodo científico

2.4 Resultado

Observación o *Realización*

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

… 1 0 0 1 0 **1** 0 1 1 … (el número en negrita es una observación en una repetición del experimento)

Resultado

- Un **resultado** es una de las posibles observaciones de un experimento.

1 es un resultado, **0** es el otro resultado

2.5 Tipos de resultado

- **Categórico**: Si el resultado de un experimento solo puede tomar valores discretos (número de piezas de automóvil producidas por hora, número de leucocitos en sangre)
- **Continuo**: Si el resultado de un experimento solo puede tomar valores continuos (estado de carga de la batería, temperatura del motor).

2.6 Experimentos aleatorios

Definición:

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

Ejemplos:

- en el mismo objeto (persona): temperatura, niveles de azúcar.
- sobre objetos diferentes pero de la misma medida: el peso de un animal.
- sobre eventos: número de correos electrónicos recibidos en una hora.

2.7 Frecuencias absolutas

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

Resumimos las observaciones **categóricas** contando cuántas veces vimos un resultado en particular.

Frecuencia absoluta:

$$n_i$$

es el número de veces que observamos el resultado i

2.8 Ejemplo

Experimento aleatorio: extraiga un leucocito de **un** donante y anote su tipo. Repita el experimento $N = 119$ veces.

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

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

- Por ejemplo: $n_1 = 34$ es el número total de células T
- $N = \sum_i n_i = 119$

2.9 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 registran $n_1 = 34$ células T, por lo que la frecuencia relativa nos da la proporción de células T de un total de 119.

2.10 Ejemplo

```
##      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
```

Tenemos

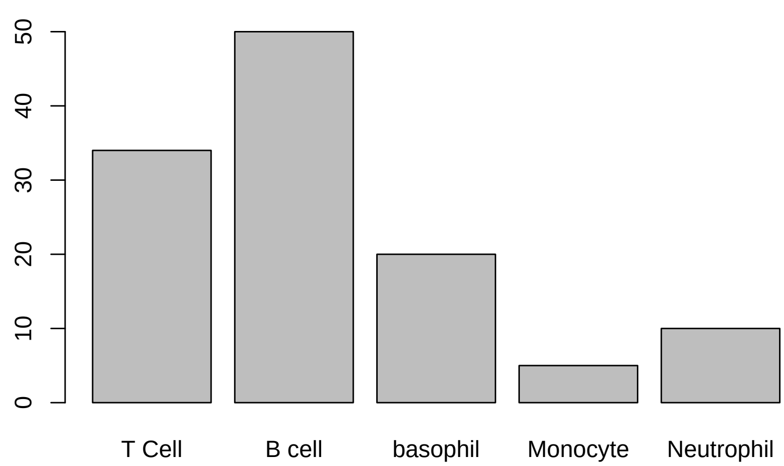
$\sum_{i=1..M} n_i = N$

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

donde M es el número de resultados.

2.11 Diagrama de barras

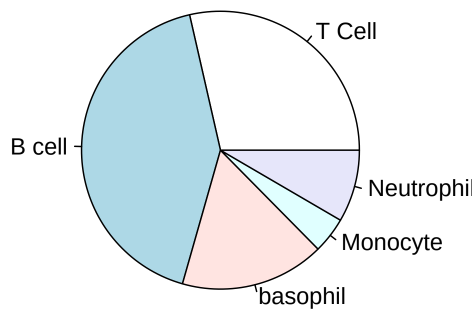
Podemos graficar n_i Vs los resultados, dándonos un gráfico de barras



2.12 Gráfico de sectores

Podemos visualizar las frecuencias relativas con un gráfico de sectores

- Donde el área del círculo representa el 100% de las observaciones (proporción = 1) y las secciones las frecuencias relativas de todos los resultados.



2.13 Variables categóricas y ordenadas

Los tipos de células no están ordenados de manera lógica en relación con los resultados. Sin embargo, a veces las variables **categóricas** se pueden **ordenar**.

Estudio de misofonía:

- 123 pacientes fueron examinados por misofonía: ansiedad/ira producida por ciertos sonidos
- Se clasificaron en 4 grupos diferentes según la gravedad.

2.14 Ejemplo

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
```

y su tabla de frecuencias

```
##   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
```

2.15 Frecuencias acumuladas absolutas y relativas

La gravedad de la misofonía es **categórica** y **ordenada**.

Cuando los resultados se pueden ordenar, entonces es útil preguntarse por el **número** de observaciones que se obtuvieron hasta un resultado dado. Llamamos a este número la frecuencia acumulada absoluta hasta el resultado i:

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

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

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

2.16 Tabla de frecuencia

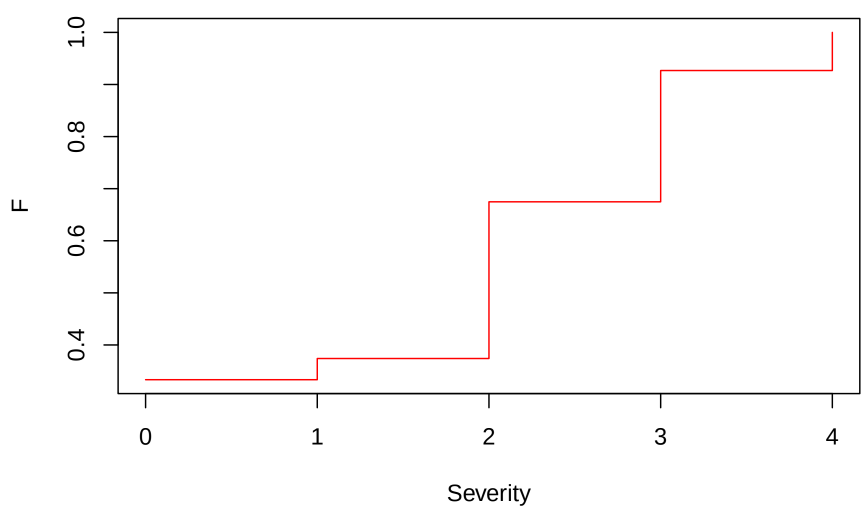
```
##   outcome ni         fi  Ni        Fi
## 0       0 41 0.33333333  41 0.3333333
## 1       1  5 0.04065041  46 0.3739837
```

```
## 2       2 37 0.30081301  83 0.6747967
## 3       3 31 0.25203252 114 0.9268293
## 4       4  9 0.07317073 123 1.0000000
```

- **67%** de los pacientes tenían misofonía hasta la gravedad **2**
- **37%** de los pacientes tienen una gravedad menor o igual a **1**

2.17 Gráfica de frecuencia acumulada

También podemos graficar la frecuencia acumulada Vs los resultados



2.18 Variables continuas

El resultado de un experimento aleatorio también puede dar resultados continuos.

En el estudio de 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 para la convexidad de la mandíbula (grados)

```
##   [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
```

2.19 Contenedores

¡Los resultados continuos no se pueden contar!

Las transformamos en variables categóricas ordenadas

- Cubrimos el rango de las observaciones en intervalos regulares del mismo tamaño (bins)

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

2.20 Crear una variable categórica a partir de una continua

- Asignamos 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]"
```

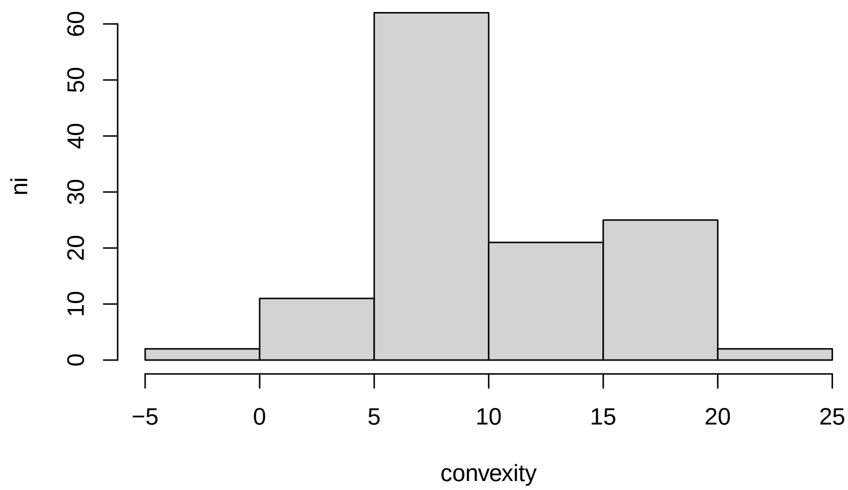
2.21 Tabla de frecuencias para una variable continua

```
##        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
```

2.22 Histograma

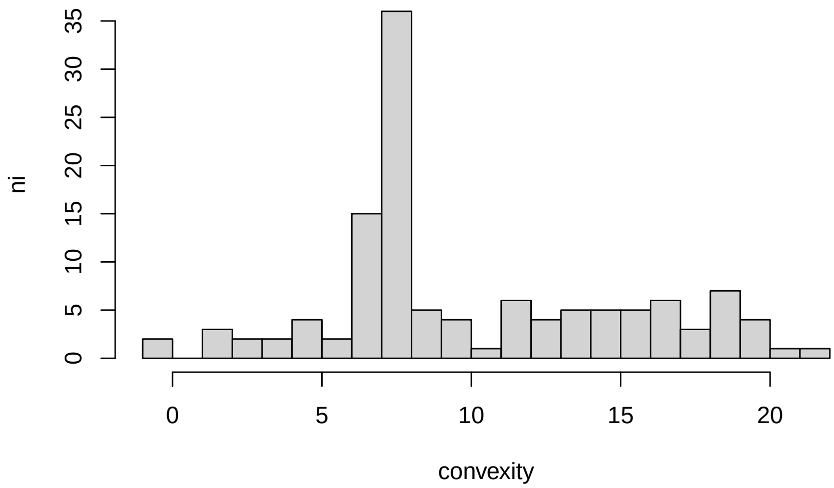
El histograma es la gráfica de n_i o f_i Vs los resultados (bins). El histograma depende del tamaño de los contenedores.

2.23 Tabla de frecuencias para una variable continua



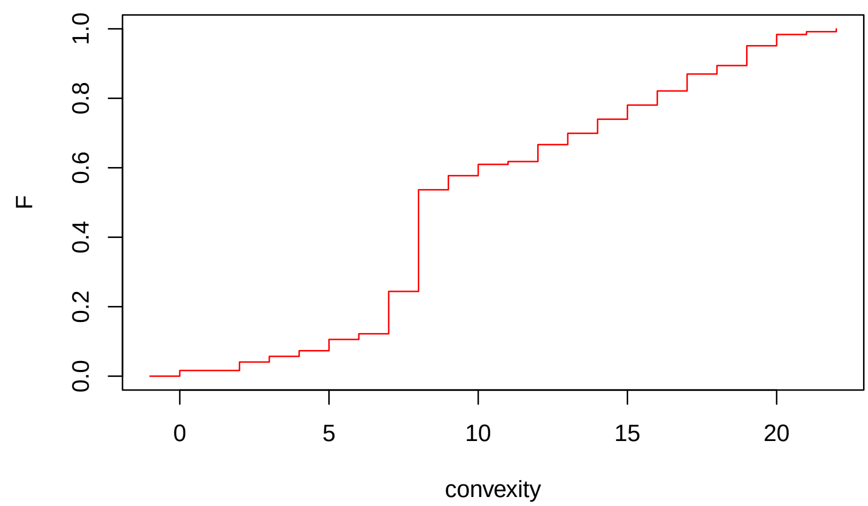
2.24 Histograma

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



2.25 Gráfica de frecuencia acumulada: Variables continuas

También podemos graficar la frecuencia acumulada Vs los resultados



2.26 Resumen estadístico

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

Valores límite:

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

Valor central para los resultados

- El promedio se define como

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

donde x_j es la **observación** j (convexidad) de un total de N.

2.27 Promedio

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

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

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

2.28 Promedio (ordenado categóricamente)

Para las variables **ordenadas categóricamente**, podemos usar la tabla de frecuencias para calcular el promedio

```
##   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
```

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

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

$= 0 * f_0 + 1 * f_1 + 2 * f_2 + 3 * f_3 + 4 * f_4 = 1,691057$

(note el cambio de N a M en la segunda suma)

2.29 Promedio (ordenado categóricamente)

En términos de los **resultados** de las variables ordenadas categóricas, el **promedio** se puede escribir como

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

de un total de M posibles resultados (número de niveles de gravedad).

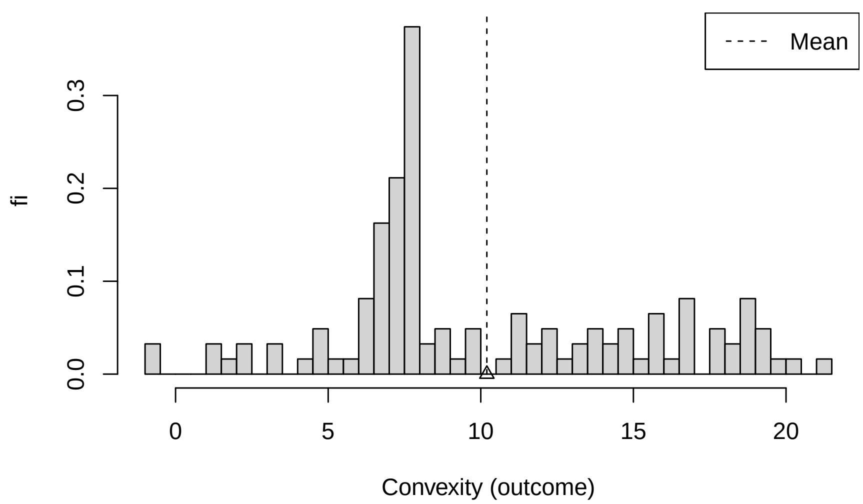
$\bar{x}$ es el **valor central** o centro de gravedad de los resultados. Como si cada resultado tuviera una densidad de masa dada por f_i.

2.30 Promedio

- El promedio no es el resultado de una observación (experimento aleatorio).
- Es el resultado de una serie de observaciones (muestra).
- Describe el número donde se equilibran los valores observados.

Por eso escuchamos, por ejemplo, que un paciente con una infección puede contagiar a una media de 2,5 personas.

2.31 Promedio



2.32 mediana

Otra medida de centralidad es la mediana. La mediana $q_{0.5}$ es el valor x_p

$$mediana(x) = q_{0.5} = x_p$$

debajo del cual encontramos la mitad de las observaciones

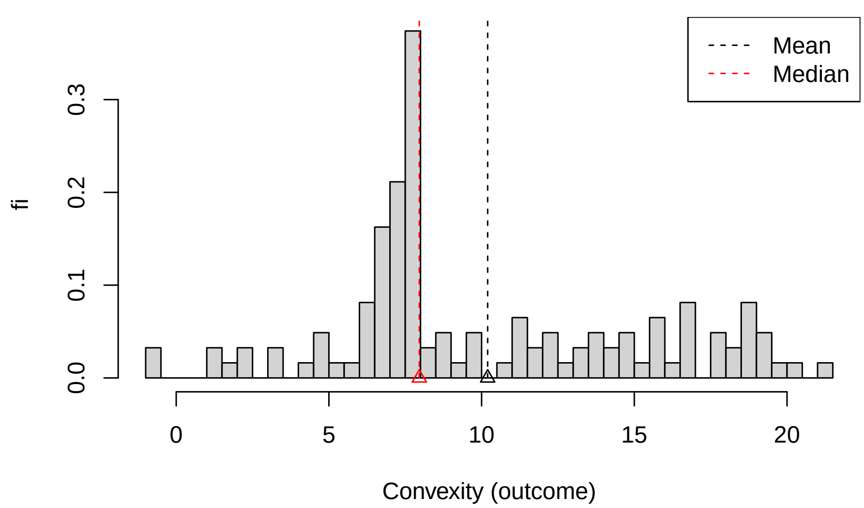
$$\sum_{x \leq x_p} 1 = \frac{N}{2}$$

o en términos de frecuencias, es el valor x_p que hace que la frecuencia acumulada F_p sea igual a 0.5

$$q_{0.5} = \sum_{x \leq x_p} f_x = F_p = 0.5$$

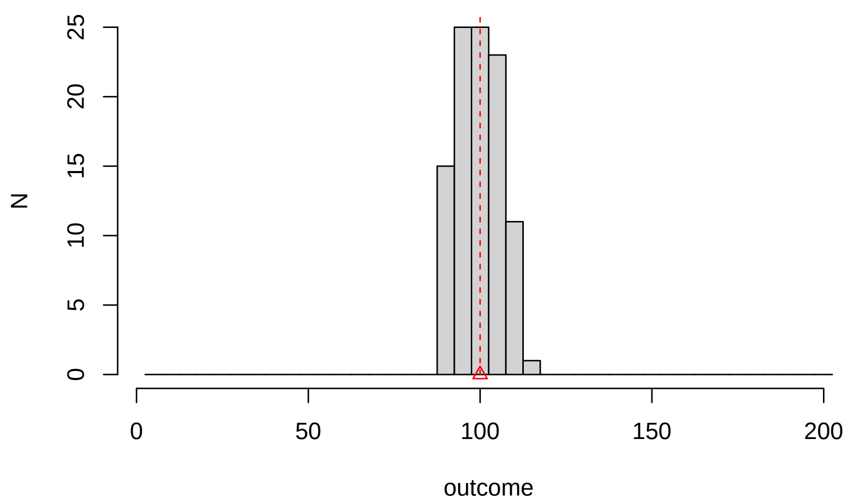
2.33 Mediana Vs Promedio

- Promedio: Centro de masa (compensa valores distantes)
- Mediana: La mitad de la masa

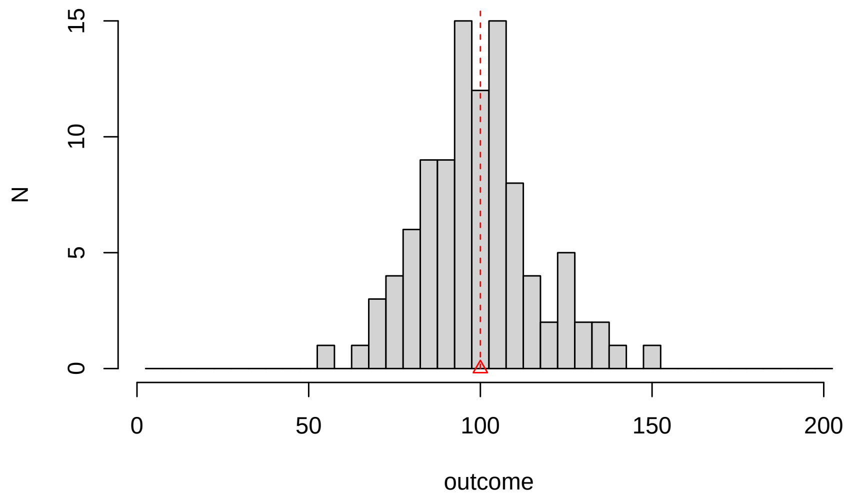


2.34 Dispersión

Una medida importante de los resultados es su **dispersión**. Muchos experimentos pueden compartir su media, pero difieren en la dispersión de los valores.



2.35 Dispersión



2.36 Variación de la muestra

La dispersión con respecto a la media se mide con el

- La varianza muestral:

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

Mide la distancia cuadrada promedio de las **observaciones** al promedio. La razón de $N-1$ se explicará cuando hablemos de inferencia.

2.37 Variación de la muestra

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

$$s^2 = \frac{N}{N-1}\sum_x (x-\bar{x})^2 f_x$$

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

2.38 Desviación Estándar

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

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 $5,086707$.

2.39 RIC

- La dispersión de datos también se puede medir con respecto a la mediana por el **rango intercuartílico**
- Definimos el **primer** cuartil como el valor x_p que hace que la frecuencia acumulada F_p sea igual a $0,25$

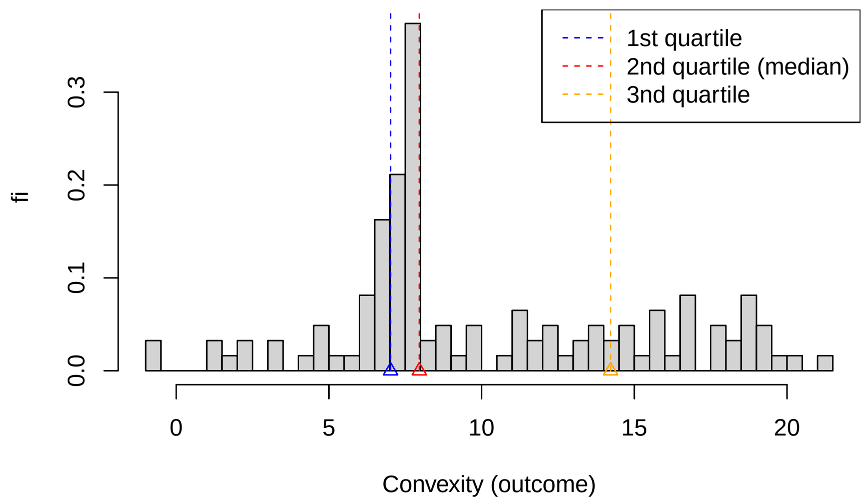
$$q_{0.25} = \sum_{x\le x_p} f_x = F_p = 0.25$$

- También definimos el **tercer** cuartil como el valor x_p que hace que la frecuencia acumulada F_p sea igual a $0,75$

$$q_{0.75} = \sum_{x\le x_p} f_x = F_p = 0.75$$

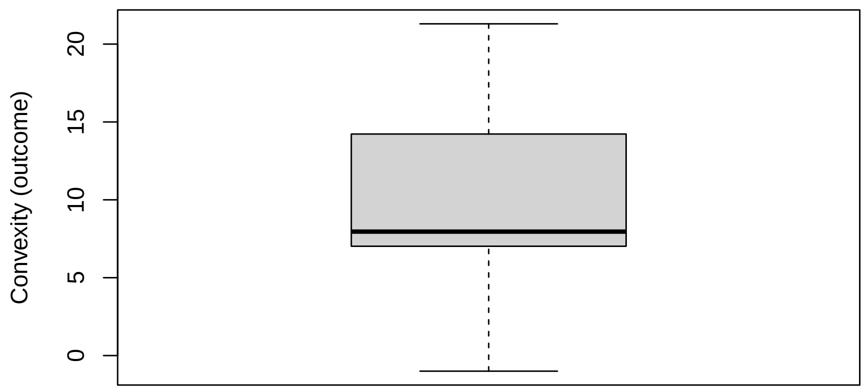
2.40 RIC

La distancia entre el tercer cuartil y el primer cuartil se denomina **rango intercuartílico** (RIC) y captura el 50 % central de las observaciones



2.41 Diagrama de caja

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


Convexity (outcome)
0
5
10
15
20

Chapter 3

Probabilidad

3.1 Objetivo

- Definición de probabilidad
- Álgebra de probabilidad
- Probabilidad conjunta

3.2 Experimentos aleatorios

Observación

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

Salir

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

Experimento aleatorio

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

3.3 Probabilidad

La **probabilidad** de un resultado es una medida de cuán seguros estamos de observar ese resultado al realizar un experimento aleatorio.

- 0: Estamos seguros de que la observación **no** ocurrirá.
- 1: Estamos seguros de que la observación sucederá.

3.4 Ejemplo

- Considere las siguientes observaciones de un experimento aleatorio:

1 5 1 2 2 1 2 2

- ¿Qué tan seguro estamos de obtener 2 en la siguiente observación?

3.5 Ejemplo

La tabla de frecuencias es

```
##   outcome ni    fi
## 1       1  3 0.375
## 2       2  4 0.500
## 3       5  1 0.125
```

La **frecuencia relativa** f_i

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

Como $f_2 = 0.5$ entonces estaríamos 50 seguros de obtener 2 en la siguiente repetición del experimento.

3.6 Frecuencia relativa

¿f_i es una buena medida de certeza?

Digamos que repetimos el experimento 12 veces más:

1 5 1 2 2 1 2 2 **3 1 1 3 3 1 6 3 5 6 4 4**

La tabla de frecuencias es ahora

```
##   outcome ni  fi
## 1       1  6 0.3
## 2       2  4 0.2
## 3       3  4 0.2
## 4       4  2 0.1
## 5       5  2 0.1
## 6       6  2 0.1
```

Aparecieron nuevos resultados y f_2 ahora es 0.2, ahora estamos un 20% seguros de obtener 2 en el próximo experimento… la probabilidad no debería depender de N

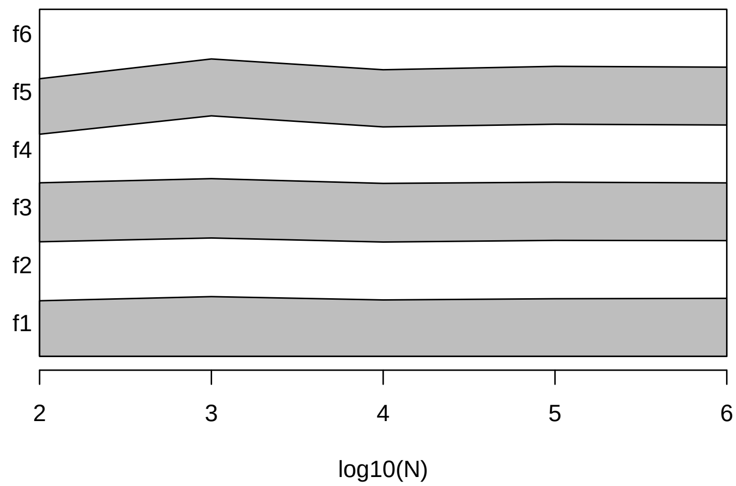
3.7 En el infinito

Digamos que repetimos el experimento 1000 veces:

```
##   outcome  ni    fi
## 1       1 171 0.171
## 2       2 149 0.149
## 3       3 153 0.153
## 4       4 152 0.152
## 5       5 178 0.178
## 6       6 197 0.197
```

Encontramos que f_i está convergiendo a un valor constante

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



3.8 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 interpretación frecuentista de probabilidades se deriva de datos/experiencia (empírica).

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

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

3.9 Probabilidad clásica

Cada vez que un experimento aleatorio tiene M resultados posibles que son todos **igualmente probables**, la probabilidad de cada resultado es $\frac{1}{M}$.

Defendida por Laplace (1814).

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

¿Y si te dijera que nuestro experimento fue tirar un dado? entonces

$P_2 = 1/6 = 0.166666$.

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

3.10 Probabilidades clásicas y frecuentistas

3.11 Probabilidad

La probabilidad es un número entre 0 y 1 que se asigna a cada miembro E de una colección de **eventos** de un **espacio muestral** (S) de un experimento aleatorio.

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

donde $E \in S$

3.12 Espacio muestral

Empezamos razonando cuáles son todos los valores posibles (resultados) que podría dar un experimento aleatorio.

Tenga en cuenta que no tenemos que observarlos en un experimento en particular: estamos usando **razón/lógica** y no observación.

Definición:

- El conjunto de todos los resultados posibles de un experimento aleatorio se denomina **espacio muestral** del experimento

- El espacio muestral se denota como S.

3.13 Ejemplos de espacios muestrales

- temperatura 35 y 42 grados centígrados
- niveles de azúcar: 70-80mg/dL
- el tamaño de un tornillo de una línea de producción: 70 mm-72 mm
- número de correos electrónicos recibidos en una hora: 0-100
- un lanzamiento de dados: 1, 2, 3, 4, 5, 6

3.14 Espacios muestrales discretos y continuos

- Un espacio muestral es discreto si consiste en un conjunto de resultados finito o infinito numerable.

- Un espacio muestral es continuo si contiene un intervalo (ya sea de longitud finita o infinita) de numeros reales.

3.15 Evento

Definición:

Un **evento** es un **subconjunto** del espacio muestral de un experimento aleatorio. Es una **colección** de resultados.

Ejemplos de eventos:

- El evento de una temperatura saludable: temperatura 37-38 grados centígrados
- El evento de producir un tornillo con un tamaño: de 71,5 mm
- El evento de recibir más de 4 correos electrónicos en una hora.
- El evento de obtener un número menor de 3 en el lanzamiento de un dado

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

3.16 Operaciones de eventos

Para dos eventos A y B, podemos construir los siguientes eventos derivados:

- 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

3.17 Ejemplo de operaciones de eventos

Tomar

- Evento $A : \{1, 2, 3\}$ un número menor o igual a tres en el lanzamiento de un dado
- Evento $B : \{2, 4, 6\}$ un número par en el lanzamiento de un dado

Nuevos eventos:

- No menos de tres: $A' : \{4, 5, 6\}$
- Menor o igual a tres **o** par: $A \cup B : \{1, 2, 3, 4, 6\}$
- Menor o igual a tres **y** par $A \cap B : \{2\}$

3.18 Resultados

Los resultados son eventos que son **mutuamente excluyentes**

Definición:

Dos eventos denotados como E_1 y E_2, tales que

$$E_1 \cap E_2 = \emptyset$$

No pueden ocurrir al mismo tiempo.

Ejemplo:

- El resultado de obtener 1 **y** el resultado de obtener 5 en el lanzamiento de un dado son mutuamente excluyentes:

- El evento de obtener 1 y 5 está vacío:

$$\{1\} \cap \{5\} = \emptyset$$

3.19 Definición de probabilidad

Una probabilidad es un número que se asigna a cada evento posible (E) de un espacio muestral (S) de un experimento aleatorio que cumple las siguientes propiedades:

- $P(S) = 1$
- $0 \leq P(E) \leq 1$
- cuando $E_1 \cap E_2 = \emptyset$

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

Propuesto por Kolmogorov (1933)

3.20 Propiedades de probabilidad

Kolmogorov dice que podemos construir una tabla de probabilidad (al igual que la tabla de frecuencia relativa)

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

Como $\{1, 2, 3, 4, 5, 6\}$ son mutuamente excluyentes, entonces

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

3.21 Regla de adición

Cuando A y B no son mutuamente excluyentes, entonces:

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

Donde $P(A)$ y $P(B)$ se denominan **probabilidades marginales**

3.22 Ejemplo de regla de adición

Tomar

- Evento $A : \{1, 2, 3\}$ un número menor o igual a tres en el lanzamiento de un dado
- Evento $B : \{2, 4, 6\}$ un número par en el lanzamiento de un dado

después:

- $P(A) : P(1) + P(2) + P(3) = 3/6$
- $P(G) : P(2) + P(4) + P(6) = 3/6$
- $P(A \cap B) : P(2) = 1/6$

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

Nota: $P(2)$ aparece en $P(A)$ y $P(B)$ por eso lo restamos con la intersección

3.23 Diagrama de Venn

Tenga en cuenta que siempre se puede descomponer el espacio muestral en conjuntos **mutuamente excluyentes** que involucran las intersecciones:

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

Marginales:

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

3.24 Tabla de probabilidades

Veamos la tabla de probabilidades.

resultado	Probabilidad
$A \cap B$	$P(A \cap B)$
$A \cap B'$	$P(A \cap B')$
$A' \cap B$	$P(A' \cap B)$
$A' \cap B'$	$P(A' \cap B')$
suma	1

3.25 Ejemplo de tabla de probabilidades

También escribimos $A \cap B$ como (A, B) y lo llamamos la **probabilidad conjunta** de A y B

En nuestro ejemplo:

resultado	Probabilidad
(A, B)	$P(A, B) = 1/6$
(A, B')	$P(A, B') = 2/6$
(A', B)	$P(A', B) = 2/6$
(A', B')	$P(A', B') = 1/6$
suma	1

Nota: cada resultado tiene *dos* valores (uno para la característica del tipo A y otro para el tipo B)

3.26 Tabla de contingencia

Podemos organizar la probabilidad de **resultados conjuntos** en una **tabla de contingencia**

	B	B'	suma
A	$P(A, B)$	$P(A, B')$	$P(A)$

	B	B'	suma
A'	$P(A', B)$	$P(A', B')$	$P(A')$
suma	$P(B)$	$P(B')$	1

marginales:

- $P(A) = P(A, B') + P(A, B)$
- $P(B) = P(A', B) + P(A, B)$

3.27 Ejemplo de tabla de contingencia

- Evento $A : \{1, 2, 3\}$ un número menor o igual a tres en el lanzamiento de un dado
- Evento $B : \{2, 4, 6\}$ un número par en el lanzamiento de un dado

	B	B'	suma
A	1/6	2/6	3/6
A'	2/6	1/6	3/6
suma	3/6	3/6	1

Tres formas de la **regla de la suma**:

$P(A \cup B)$

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

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

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

3.28 Estudio de misofonía

En el estudio de misofonía, se evaluó a los pacientes según la gravedad de su misofonía **y** si estaban deprimidos.

El resultado de un experimento aleatorio es medir la gravedad de la misofonía **y** el estado de depresión de un paciente. La repetición del experimento aleatorio consistía en realizar las mismas dos mediciones en otro paciente.

```
##     Misofonia.dic depresion.dic
## 1              4             1
## 2              2             0
## 3              0             0
## 4              3             0
## 5              0             0
## 6              0             0
## 7              2             0
## 8              3             0
## 9              0             1
## 10             3             0
## 11             0             0
## 12             2             0
## 13             2             1
## 14             0             0
## 15             2             0
## 16             0             0
## 17             0             0
## 18             3             0
## 19             3             0
## 20             0             0
## 21             3             0
## 22             3             0
## 23             2             0
## 24             0             0
## 25             0             0
## 26             0             0
## 27             4             1
## 28             2             0
## 29             2             0
## 30             0             0
## 31             2             0
## 32             0             0
## 33             0             0
## 34             0             0
## 35             3             0
## 36             0             0
## 37             2             0
## 38             3             1
## 39             2             0
## 40             2             0
## 41             0             0
## 42             2             0
## 43             3             0
## 44             0             0
## 45             0             0
```

```
## 46                2              0
## 47                2              0
## 48                3              0
## 49                3              0
## 50                0              0
## 51                0              0
## 52                4              1
## 53                3              0
## 54                3              1
## 55                2              1
## 56                0              1
## 57                2              0
## 58                0              0
## 59                0              0
## 60                0              0
## 61                2              0
## 62                2              0
## 63                0              0
## 64                0              0
## 65                2              0
## 66                3              1
## 67                0              0
## 68                1              0
## 69                3              0
## 70                2              0
## 71                4              1
## 72                3              0
## 73                2              1
## 74                3              0
## 75                0              1
## 76                2              0
## 77                3              0
## 78                2              0
## 79                4              1
## 80                1              0
## 81                2              0
## 82                0              0
## 83                2              0
## 84                0              0
## 85                2              0
## 86                0              1
## 87                2              0
## 88                2              0
## 89                4              1
## 90                3              0
## 91                0              1
```

```
## 92                3              0
## 93                0              0
## 94                0              0
## 95                0              0
## 96                2              0
## 97                2              0
## 98                1              0
## 99                3              0
## 100               0              0
## 101               0              0
## 102               3              1
## 103               2              0
## 104               1              0
## 105               3              0
## 106               0              0
## 107               4              1
## 108               4              1
## 109               2              0
## 110               3              0
## 111               3              0
## 112               3              1
## 113               0              0
## 114               3              0
## 115               2              0
## 116               1              0
## 117               2              0
## 118               3              1
## 119               3              0
## 120               4              1
## 121               2              0
## 122               3              0
## 123               2              0
```

3.29 Tabla de contingencia para frecuencias

- Para el número de observaciones $n_{i,j}$ de cada resultado (x_i, y_i), misofonía: $x \in \{0, 1, 2, 3, 4\}$ y depresión $y\ en\{0, 1\}$ (no:0, sí:1)

```
##
##                Depression:0 Depression:1
##   Misophonia:4            0            9
##   Misophonia:3           25            6
```

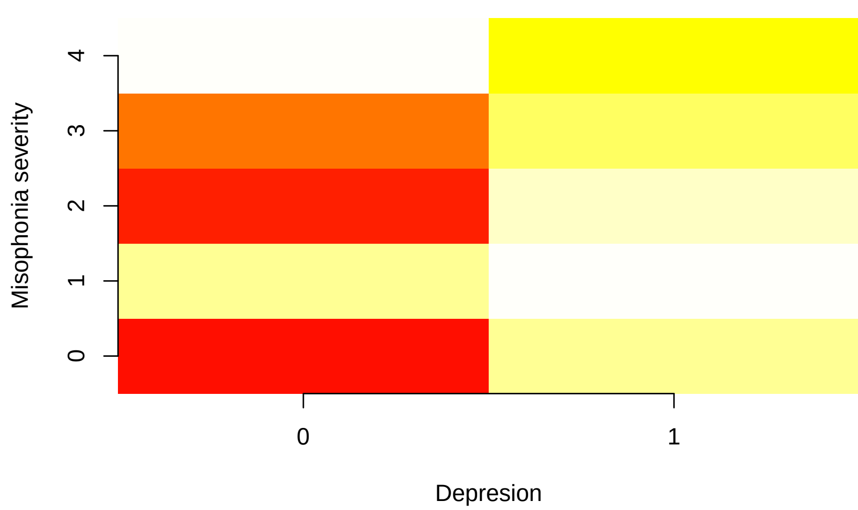
```
##    Misophonia:2           34            3
##    Misophonia:1            5            0
##    Misophonia:0           36            5
```

- para las frecuencias relativas $f_{i,j}$

```
##
##                 Depression:0 Depression:1
##    Misophonia:4   0.00000000   0.07317073
##    Misophonia:3   0.20325203   0.04878049
##    Misophonia:2   0.27642276   0.02439024
##    Misophonia:1   0.04065041   0.00000000
##    Misophonia:0   0.29268293   0.04065041
```

3.30 Mapa de calor

La tabla de contingencia se puede trazar como un **mapa de calor**



3.31 Variables continuas

En el estudio de misofonía también se midió la protrusión mandibular como posible factor cefalométrico de la enfermedad.

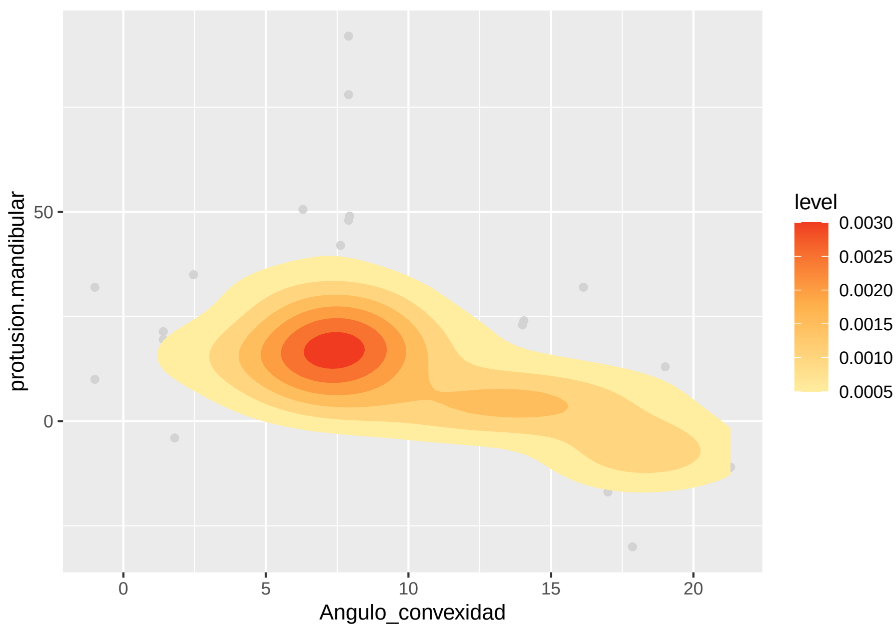
```
##    Angulo_convexidad protusion.mandibular
## 1               7.97                13.00
## 2              18.23                -5.00
## 3              12.27                11.50
## 4               7.81                16.80
## 5               9.81                33.00
## 6              13.50                 2.00
## 7              19.30                -3.90
## 8               7.70                16.80
## 9              12.30                 8.00
## 10              7.90                28.80
## 11             12.60                 3.00
## 12             19.00                -7.90
## 13              7.27                28.30
## 14             14.00                 4.00
## 15              5.40                22.20
## 16              8.00                 0.00
## 17             11.20                15.00
## 18              7.75                17.00
## 19              7.94                49.00
## 20             16.69                 5.00
## 21              7.62                42.00
## 22              7.02                28.00
## 23              7.00                 9.40
## 24             19.20               -13.20
## 25              7.96                23.00
## 26             14.70                 2.30
## 27              7.24                25.00
## 28              7.80                 4.90
## 29              7.90                92.00
## 30              4.70                 6.00
## 31              4.40                17.00
## 32             14.00                 3.30
## 33             14.40                10.30
## 34             16.00                 6.30
## 35              1.40                19.50
## 36              9.76                22.00
## 37              7.90                 5.00
## 38              7.90                78.00
## 39              7.40                 9.30
## 40              6.30                50.60
```

```
## 41                 7.76                 18.00
## 42                 7.30                 18.00
## 43                 7.00                 10.00
## 44                11.23                  4.00
## 45                16.00                 13.30
## 46                 7.90                 48.00
## 47                 7.29                 23.50
## 48                 6.91                 37.60
## 49                 7.10                 15.00
## 50                13.40                  5.10
## 51                11.60                 -2.20
## 52                -1.00                 32.00
## 53                 6.00                 25.00
## 54                 7.82                 24.00
## 55                 4.80                 33.60
## 56                11.00                  3.30
## 57                 9.00                 31.50
## 58                11.50                 12.80
## 59                16.00                  3.00
## 60                15.00                  6.00
## 61                 1.40                 21.40
## 62                16.80                -10.00
## 63                 7.70                 19.00
## 64                16.14                 32.00
## 65                 7.12                 15.00
## 66                -1.00                 10.00
## 67                17.00                -16.90
## 68                 9.26                  2.00
## 69                18.70                -10.10
## 70                 3.40                 12.20
## 71                21.30                -11.00
## 72                 7.50                  5.20
## 73                 6.03                 16.00
## 74                 7.50                  5.80
## 75                19.00                  5.20
## 76                19.01                 13.00
## 77                 8.10                 13.60
## 78                 7.80                 16.10
## 79                 6.10                 33.20
## 80                15.26                  4.00
## 81                 7.95                 12.00
## 82                18.00                 -1.50
## 83                 4.60                 18.30
## 84                15.00                  3.00
## 85                 7.50                 15.80
## 86                 8.00                 27.10
```

```
## 87                 16.80              -10.00
## 88                  8.54               25.00
## 89                  7.00               27.10
## 90                 18.30               -8.00
## 91                  7.80               12.00
## 92                 16.00               -8.00
## 93                 14.00               23.00
## 94                 12.30                5.00
## 95                 11.40                1.00
## 96                  8.50               18.90
## 97                  7.00               15.00
## 98                  7.96               22.00
## 99                 17.60               -3.50
## 100                10.00               20.00
## 101                 3.50               12.20
## 102                 6.70               14.70
## 103                17.00               -5.00
## 104                20.26               -4.15
## 105                 6.64               11.00
## 106                 1.80               -4.00
## 107                 7.02               25.00
## 108                 2.46               35.00
## 109                19.00               -5.00
## 110                17.86              -30.00
## 111                 6.10               12.20
## 112                 6.64               19.00
## 113                12.00                1.60
## 114                 6.60               20.00
## 115                 8.70               17.10
## 116                14.05               24.00
## 117                 7.20                7.10
## 118                19.70              -11.00
## 119                 7.70               21.30
## 120                 6.02                5.00
## 121                 2.50               12.90
## 122                19.00                5.90
## 123                 6.80                5.80
```

3.32 Variables continuas

En el estudio de misofonía también se midió la protrusión mandibular como posible factor cefalométrico de la enfermedad.

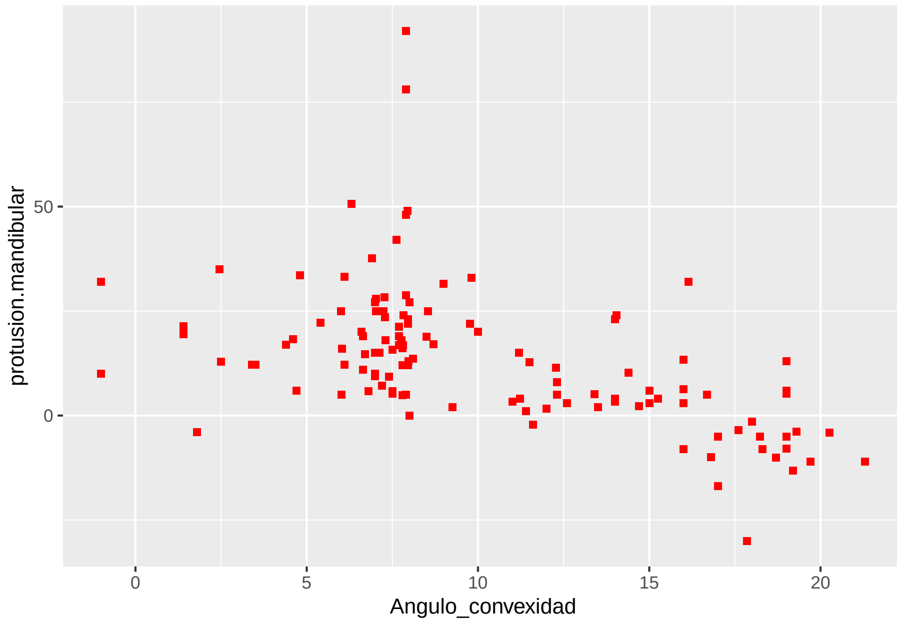


3.33 Gráfico de dispersión

- El **histograma** depende del tamaño del contenedor (píxel).

- Si el píxel es lo suficientemente pequeño como para contener una sola observación, el mapa de calor da como resultado un **diagrama de dispersión**

El diagrama de dispersión es la ilustración de una "tabla de contingencia" para variables continuas cuando el contenedor (píxel) es lo suficientemente pequeño como para contener una sola observación (que consta de un par de valores).



protusion.mandibular
50
0
0
5
10
15
20
Angulo_convexidad

Chapter 4

Probabilidad condicional

4.1 Objetivo

- Probabilidad condicional
- Independencia
- Teorema de Bayes

4.2 Probabilidad conjunta

La probabilidad conjunta de dos eventos A y B es

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

Imaginemos un experimento aleatorio que mide dos tipos diferentes de resultados.

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

En muchos casos, nos interesa saber si los valores de un resultado **condicionan** los valores del otro.

4.3 Diagnósticos

Consideremos una **herramienta de diagnóstico**

Queremos encontrar el estado de un sistema (s):

- inadecuado (sí)
- adecuado (no)

con una prueba (t):

- positivo
- negativo

Probamos una batería para saber cuánto tiempo puede vivir. Tensamos un cable para saber si resiste llevar cierta carga. Realizamos una PCR para ver si alguien está infectado.

4.4 Prueba de diagnóstico

Consideremos diagnosticar una infección con una nueva prueba.

Estado de infección:

- si (infectado)
- no (no infectado)

Prueba:

- positivo
- negativo

4.5 Observaciones

Cada individuo es un experimento aleatorio con dos medidas: (Infección, Prueba)

Asunto	Infección	Prueba
s_1	si	positivo
s_2	no	negativo
s_3	si	positivo
...	...	...

Asunto	Infección	Prueba
s_i	no	positivo*
...	...	...
...	...	...
s_n	si	negativo*

4.6 Tablas de contingencia

- Por el número de observaciones de cada resultado

	Infección: sí	Infección: no	suma
Test: positivo	18	12	30
Test: negativo	30	300	330
suma	48	312	360

- Para las frecuencias relativas, si $N >> 0$ tomaremos $f_{i,j} = \hat{P}(x_i, y_j)$

	Infección: sí	Infección: no	suma
Test: positivo	0.05	0.0333	0.0833
Test: negativo	0.0833	0.833	0.9166
suma	0.133	0.866	1

4.7 La probabilidad condicional

Pensemos primero en términos de aquellos que están **infectados**

Dentro de los que están infectados (**sí**), ¿cuál es la probabilidad de dar positivo?

- Sensibilidad (tasa de verdaderos positivos)

$$\hat{P}(positivo|s) = \frac{n_{positivo,s}}{n_s}$$

$$= \frac{\frac{n_{positivo,s}}{N}}{\frac{n_s}{N}} = \frac{f_{positivo,s}}{f_s}$$

Por lo tanto, en el límite, esperamos tener una probabilidad del tipo

$$P(positivo|s) = \frac{P(positivo, s)}{P(s)} = \frac{P(positivo \cap s)}{P(s)}$$

4.8 La probabilidad condicional

Definición: La probabilidad condicional de un evento B dado un evento A, denotado como $P(A|B)$, es

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

- se puede probar que la probabilidad condicional satisface los axiomas de probabilidad.
- la probabilidad condicional es la probabilidad bajo el espacio muestral dado por B: S_B.

4.9 Tabla de contingencia condicional

	Infección: Sí	Infección: No
Test: positivo	P(positivo | sí)	P(positivo | no)
Test: negativo	P(negativo | sí)	P(negativo | no)
suma	1	1

- Tasa de verdaderos positivos (Sensibilidad): La probabilidad de dar positivo **si** se tiene la enfermedad $P(positivo|s)$
- Tasa de verdaderos negativos (Especificidad): La probabilidad de dar negativo **si** no se tiene la enfermedad $P(negativo|no)$
- Tasa de falsos positivos: La probabilidad de dar positivo **si** no se tiene la enfermedad $P(positivo|no)$
- Tasa de falsos negativos: la probabilidad de dar negativo **si** se tiene la enfermedad $P(negativo|s)$

4.10 Ejemplo de tabla de contingencia condicional

Tomando las frecuencias como estimaciones de las probabilidades, entonces

	Infección: Sí	Infección: No
Test: positivo	18/48 = 0.375	12/312 = 0.038
Test: negativo	30/48 = 0.625	300/312 =0.962
suma	1	1

Nuestra herramienta de diagnóstico tiene baja sensibilidad (0.375) pero alta especificidad (0.962).

4.11 Regla de multiplicación

Ahora imaginemos la situación real, donde queremos obtener la probabilidad **conjunta** de la probabilidad **condicional**

- Se (realizaron) PCR para coronavirus [https://www.nejm.org/doi/full/10.1056/NEJMp2015897] en personas en el hospital que estamos seguros de estar infectadas. Este test tiene una sensibilidad del 70%. También se ha probado en el laboratorio en condiciones sin infección con una especificidad del 96 %.
- Un estudio de prevalencia en España mostró que $P(s) = 0.05$, $P(no) = 0.95$ antes del verano.

Con estos datos, ¿cuál era la probabilidad de que una persona seleccionada al azar de la población diera positivo **y** estuviera infectada: $P(s \cap positivo) = P(s, positivo)$?

4.12 Rendimiento de diagnóstico

Para estudiar el rendimiento de una nueva prueba diagnóstica:

- selecciona muestras que son inadecuadas (enfermedad: **sí**) y aplica la prueba, tratando de encontrar su sensibilidad: $P(positivo|s)$ (0.70 para PCR)

- selecciona muestras que son adecuadas (enfermedad: **no**) y aplica la prueba, tratando de encontrar su especificidad: $P(negativo|no)$ (0.96 para PCR)

	Infección: Sí	Infección: No
Test: positivo	P(positivo|sí)=0.7	P(positivo|no)=0.06
Test: negativo	P(negativo|sí)=0.3	P(negativo|no)=0.94
suma	1	1

De esta matriz, ¿podemos obtener $P(s, positivo)$?

4.13 Regla de multiplicación

¿Cómo se recupera la probabilidad conjunta de la probabilidad condicional?

Para dos eventos A y B tenemos la regla de la multiplicación

$$P(A, B) = P(A|B)P(B)$$

que se sigue de la definición de probabilidad condicional.

4.14 Tabla de contingencia en términos de probabilidades condicionales

	Infección: Sí	Infección: No	suma
Test: positivo	P(positivo | sí)P(sí)	P(positivo | no)P(no)	P(positivo)
Test: negativo	P(negativo | sí)P(sí)	P(negativo | no) P(no)	P(negativo)
suma	P(sí)	P(no)	1

Por ejemplo, la probabilidad de dar *positivo* y estar infectado s:

- $P(positivo, s) = P(positivo \cap s) = P(positivo|s)P(s)$

4.15 Árbol condicional

4.16 Tabla de contingencia en términos de probabilidades condicionales

	Infección: sí	Infección: no	suma
Test: positivo	0.035	0.057	0.092
Test: negativo	0.015	0.893	0.908
suma	0.05	0.95	1

- $P(positivo, si) = 0.035$

Pero también encontramos la probabilidad marginal de ser **positivo**:

- $P(positivo) = 0.092$

4.17 Regla de probabilidad total

	Infección: Sí	Infección: No	suma
Test: positivo	P(positivo | sí)P(sí)	P(positivo | no)P(no)	P(positivo)
Test: negativo	P(negativo | sí)P(sí)	P(negativo | no) P(no)	P(negativo)
suma	P(sí)	P(no)	1

Cuando escribimos las marginales desconocidas en términos de sus probabilidades condicionales, lo llamamos **regla de probabilidad total**

- $P(positivo) = P(positivo|s)P(s) + P(positivo|no)P(no)$
- $P(negativo) = P(negativo|s)P(s) + P(negativo|no)P(no)$

4.18 Árbol condicional

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

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

4.19 Encontrar probabilidades inversas

De la tabla de contingencia condicional

	Infección: Sí	Infección: No
Test: positivo	P(positivo | sí)	P(positivo | no)
Test: negativo	P(negativo | sí)	P(negativo | no)
suma	1	1

¿Cómo podemos calcular la probabilidad de estar infectado si la prueba da positivo: $P(s|positivo)$?

4.20 Recuperar probabilidades conjuntas

1. Recuperamos la tabla de contingencia para probabilidades conjuntas

	Infección: Sí	Infección: No	suma
Test: positivo	P(positivo | sí)P(sí)	P(positivo | no)P(no)	P(positivo)
Test: negativo	P(negativo | sí)P(sí)	P(negativo | no) P(no)	P(negativo)
suma	P(sí)	P(no)	1

4.21 Condicionales inversas

2. Calculamos las probabilidades condicionales para la prueba:

$$P(infeccin|prueba) = \frac{P(prueba|infeccin)P(infeccin)}{P(prueba)}$$

	Infección: Sí	Infección: No	suma
Test: positivo	P(sí|positivo)	P(sin|positivo)	1
Test: negativo	P(sí|negativo)	P(sin|negativo)	1

Por ejemplo:

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

como normalmente no tenemos $P(positivo)$, usamos la regla de **probabilidad total** en el denominador

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

4.22 Teorema de Bayes

La expresion:

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

se llama **teorema de Bayes**

Teorema

Si $E1, E2, ..., Ek$ son k eventos mutuamente excluyentes y exhaustivos y B es cualquier evento,

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

Permite invertir los condicionales:

$$P(B|A) \rightarrow P(A|B)$$

O **diseñe** una prueba B en condición controlada A y luego utilícela para **inferir** la probabilidad de la condición cuando la prueba es positiva.

4.23 Ejemplo: teorema de Bayes

Teorema de Bayes:

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

sabemos:

- $P(positivo|s) = 0.70$
- $P(positivo|no) = 1 - P(negativo|no) = 0.06$
- la probabilidad de infección y no infección en la población: $P(s) = 0.05$ y $P(no) = 1 - P(s) = 0.95$.

Por lo tanto:

$$P(s|positivo) = 0.47$$

Las pruebas no son tan buenas para **confirmar** infecciones.

4.24 Ejemplo: teorema de Bayes

Apliquémoslo ahora a la probabilidad de no estar infectado si la prueba es negativa.

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

La sustitución de todos los valores da

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

Las pruebas son buenas para **descartar** infecciones.

4.25 Independencia estadística

En muchas aplicaciones, queremos saber si el conocimiento de un evento condiciona el resultado de otro evento.

- hay casos en los que queremos saber si los eventos no están condicionados

4.26 Independencia estadística

Considere los conductores para los cuales medimos sus fallas superficiales y si su capacidad de conducción es defectuosa.

Las **probabilidades conjuntas** estimadas son

	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

- $P(D, F) = 0.005$

Las probabilidades marginales son

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

4.27 Independencia estadística

¿Cuál es la **probabilidad condicional** de observar un conductor defectuoso si tiene un defecto?

	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

¡Las probabilidades marginales y condicionales son las mismas!

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

La probabilidad de observar un conductor defectuoso **no** depende de haber observado o no un defecto.

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

4.28 Independencia estadística

Dos eventos A y B son estadísticamente independientes si

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

y por la regla de la multiplicación, su probabilidad conjunta es

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

la multiplicación de sus probabilidades marginales.

4.29 Productos de productos marginales

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

Confirme que todas las entradas de la matriz son el producto de los marginales.

Por ejemplo:

- $P(F)P(D) = P(D \cap F)$
- $P(D')P(F') = P(D' \cap F')$

4.30 Ejemplo

Resultados de lanzar dos monedas: $S = (H, H), (H, T), (T, H), (T, T)$

	H	T	suma
H	1/4	1/4	1/2
T	1/4	1/4	1/2
suma	1/2	1/2	1

- Obtener cara en la primera moneda no condiciona obtener cruz en el resultado de la segunda moneda $P(T|H) = P(T) = 1/2$
- la probabilidad de obtener cara y después cruz es el producto de cada resultado independiente $P(H, T) = P(H) * P(T) = 1/4$

Chapter 5

Variables aleatorias discretas

5.1 Objetivo

- Variables aleatorias
- Función de probabilidad
- Media y varianza
- Distribución de probabilidad

5.2 ¿Cómo asignamos valores de probabilidad a los resultados?

5.3 Variable aleatoria

Definición:

Una **variable aleatoria** es una función que asigna un **número** real a cada **resultado** en el espacio muestral de un experimento aleatorio.

- Por lo general, una variable aleatoria es el valor de la **medida** de interés que se realiza en un experimento aleatorio.

Una variable aleatoria puede ser:

- Discreta (nominal, ordinal)
- Continua (intervalo, relación)

5.4 Variable aleatoria

Un **valor** (o **resultado**) de una variable aleatoria es uno de los números posibles que la variable puede tomar en un experimento aleatorio.

Escribimos la variable aleatoria en **mayúsculas**.

Ejemplo:

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

Observación de una variable aleatoria

- Una observación es la **adquisición** del valor de una variable aleatoria en un experimento aleatorio

Ejemplo:

1 0 0 1 0 **1** 0 1 1

El número en negrita es una observación de X

5.5 Eventos de observar una variable aleatoria

- $X = 1$ 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 con valor x (pequeño x)

- Dos valores cualesquiera de una variable aleatoria definen dos eventos **mutuamente excluyentes**.

5.6 Probabilidad de variables aleatorias

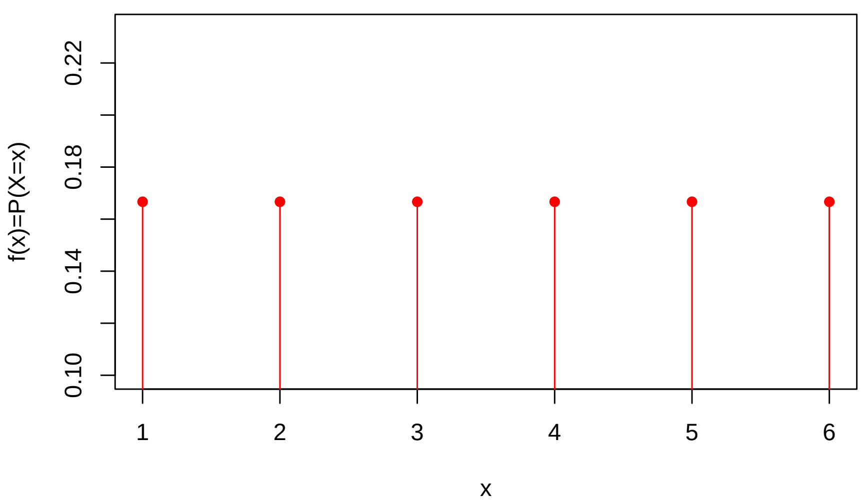
Nos interesa asignar probabilidades a los valores de una variable aleatoria.

Ya hemos hecho esto para los dados: $X \in \{1, 2, 3, 4, 5, 6\}$ (interpretación clásica de probabilidad)

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

5.7 Funciones de probabilidad

- Podemos escribir la tabla de probabilidad
- graficarla

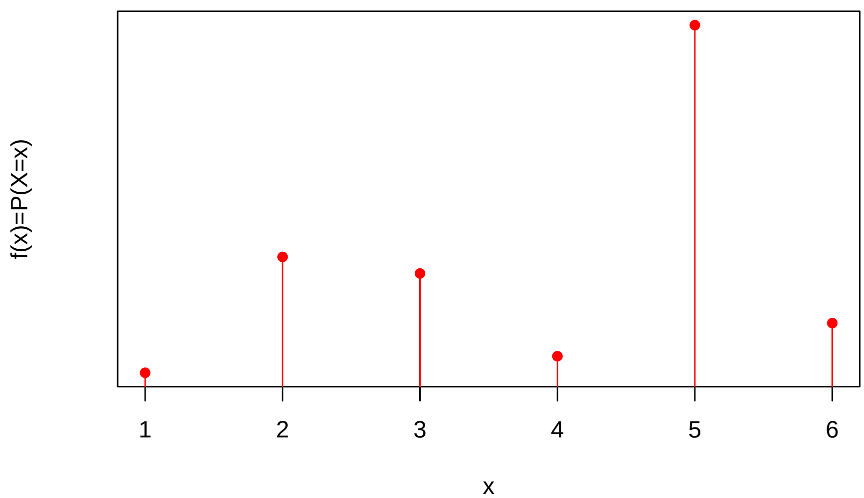


- o escribirla como la función

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

5.8 Funciones de probabilidad

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



5.9 Funciones de probabilidad

Para una variable aleatoria discreta $X \in \{x_1, x_2, .., x_M\}$, una **función de masa de probabilidad**

siempre es positiva

- $f(x_i) \geq 0$

se utiliza para calcular probabilidades

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

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

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

5.10 Funciones de probabilidad

- Tenga 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).
- X y $f(x)$ son objetos abstractos que pueden o no asignarse a un experimento
- Tenemos la libertad de construirlos como queramos siempre que respetemos su definición.
- Tienen algunas **propiedades** que se derivan exclusivamente de su definición.

5.11 Ejemplo: función de masa de probabilidad

Considere la siguiente variable aleatoria X sobre los resultados

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

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

5.12 Tabla de probabilidad para resultados igualmente probables

resultado	Probabilidad (resultado)
a	1/6
b	1/6
c	1/6
d	1/6
e	1/6
f	1/6

5.13 Tabla de probabilidad para X

X	$f(x) = P(X = x)$
0	$P(X = 0) = 2/6$
1.5	$P(X = 1.5) = 2/6$
2	$P(X = 2) = 1/6$
3	$P(X = 3) = 1/6$

Podemos calcular, por ejemplo, las siguientes probabilidades de eventos en los valores de X

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

5.14 Ejemplo

Modelo de probabilidad:

Considere el siguiente experimento: En una urna ponga 8 bolas y:

- marque 1 bola con el número -2
- marque 2 bolas con el número -1
- marque 2 bolas con el número 0

- marque 2 bolas con el número 1
- marque 1 bola con el número 2

experimento: Tome una bola y lea el número.

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	$1/8 = 0.125$

5.15 Ejemplo

Considere otro experimento en el que no sabemos qué hay en la urna anterior. Sacamos una bola 30 veces, escribimos su númeror y la devolvemos a la urna.

- no sabemos cuáles son los eventos primarios con iguales probabilidades.
- y **estimamos** la **función de masa de probabilidad** a partir de las **frecuencias relativas** observadas para cada variable aleatoria

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

5.16 Probabilidades y frecuencias

Para calcular las frecuencias relativas f_i trenemos que

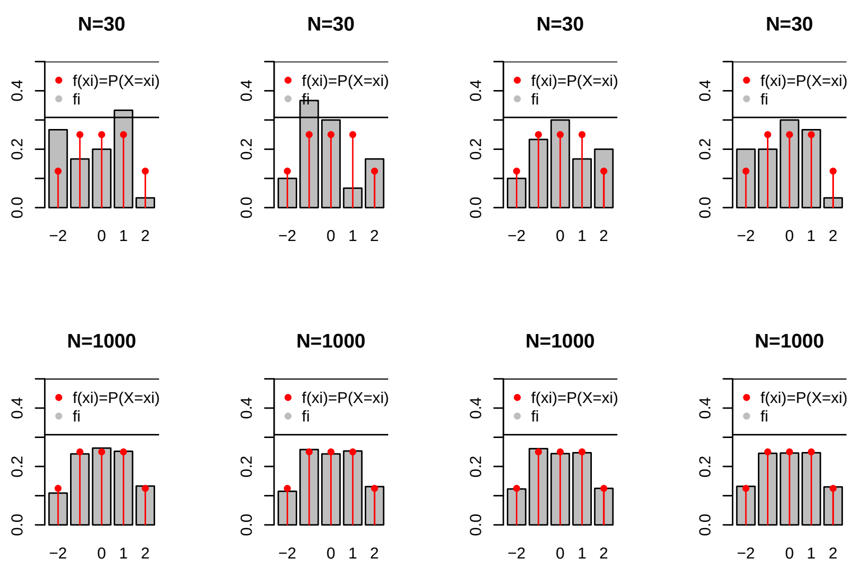
- **repetir** el experimento N veces (tenemos que volver a poner la bola en la urna cada vez) y al final calcular

$$f_i = n_i/N$$

Estamos suponiendo que:

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

5.17 Probabilidades y frecuencias relativas



- En este ejemplo, **sabemos** el **modelo** de probabilidad $f(x) = P(X = x)$ por diseño.
- Nunca observamos $f(x)$
- Podemos usar frecuencias relativas para estimar las probabilidades

$$f_i = \hat{f}(x_i) = \hat{P}(X = x_i)$$

(f_i depende de N)

5.18 Media y Varianza

Las funciones de masa de probabilidad $f(x)$ tienen dos propiedades principales

- su centro
- su dispersión

Podemos preguntar,

- ¿Alrededor de qué valores de X se concentró la probabilidad?
- ¿Qué tan dispersos son los valores de X en relación a sus probabilidades?

5.19 Media y Varianza

5.20 Media

Recuerde que el **promedio** en términos de las frecuencias relativas de los valores de x_i (resultados ordenados categóricos) se puede escribir como

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

Definición

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

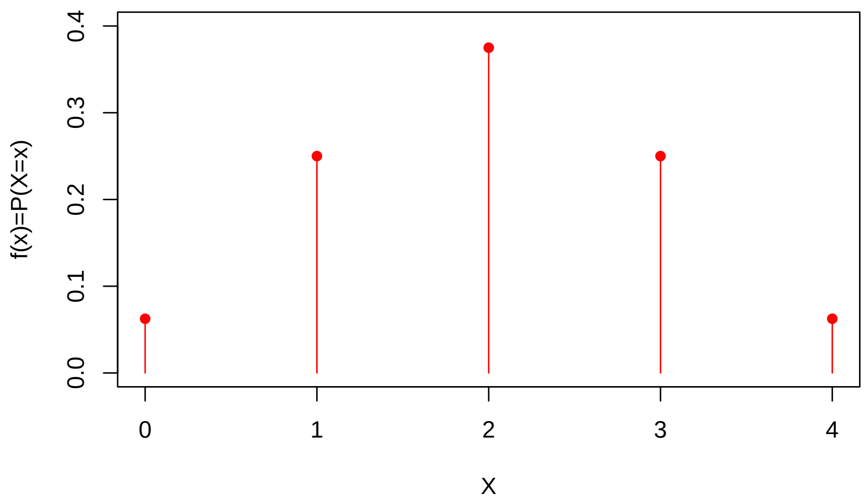
$$\mu = E(X) = \sum_{i=1}^{M} x_i f(x_i)$$

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

5.21 Ejemplo: Media

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

$P(X = 0) = 1/16$ $P(X = 1) = 4/16$ $P(X = 2) = 6/16$ $P(X = 3) = 4/16$ $P(X = 4) = 1/16$



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

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

5.22 Media y Promedio

- La media μ es el centro de gravedad de función de masa de probabilidad y **no cambia**

Por ejemplo de

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

X	$P(X = x)$
2	$1/8 = 0.125$

- El promedio $\bar{x}$ es el centro de gravedad de las observaciones (frequencias relativas) **y cambia** de acuerdo a los datos.

Por ejemplo de

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

5.23 Variación

En términos similares definimos la distancia media al cuadrado de la media:

Definición

La varianza, escrita como σ^2 o $V(X)$, de una variable aleatoria discreta X con función de masa $f(x)$ está 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
- Piense en ello como el momento de inercia de las probabilidades sobre la media.

5.24 Ejemplo: Varianza

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

$P(X = 0) = 1/16$ $P(X = 1) = 4/16$ $P(X = 2) = 6/16$ $P(X = 3) = 4/16$ $P(X = 4) = 1/16$

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

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

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

$$\sigma = 1$$

5.25 Funciones de X

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** importantes de la mediana y la varianza:

- 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) .

- La varianza de una función lineal de X es:

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

- La varianza **sobre el origen** es la varianza **sobre la media** más la media al cuadrado:

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

5.26 Ejemplo: Varianza sobre el origen

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

$P(X = 0) = 1/16$ $P(X = 1) = 4/16$ $P(X = 2) = 6/16$ $P(X = 3) = 4/16$ $P(X = 4) = 1/16$

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

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

También podemos verificar:

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

$5 = 1 + 2^2$

5.27 Distribución de probabilidad

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:

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

5.28 Ejemplo: distribución de probabilidad

Para la función de masa de probabilidad:

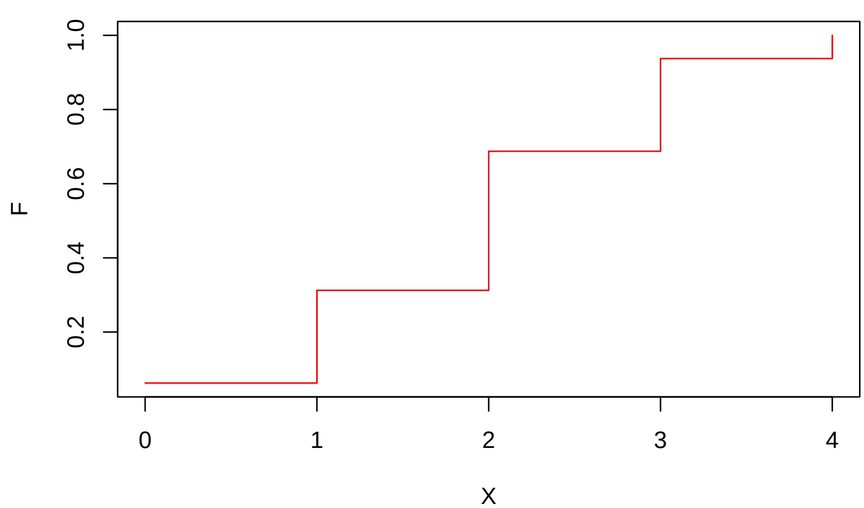
$f(0) = P(X = 0) = 1/16$ $f(1) = P(X = 1) = 4/16$ $f(2) = P(X = 2) = 6/16$ $f(3) = P(X = 3) = 4/16$ $f(4) = P(X = 4) = 1/16$

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 $X \in \mathbb{Z}$

5.29 Probability distribution



5.30 Función de probabilidad y Distribución de probabilidad

Calcule la función de probabilidad de masa de la siguiente distribución de probabilidad:

$F(0) = 1/16$, $F(1) = 5/16$, $F(2) = 11/16$, $F(3) = 15/16$, $F(4) = 16/16$,

Trabajemos al revés.

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

5.31 Función de probabilidad y Distribución de probabilidad

La distribución de probabilidad es otra forma de especificar la probabilidad de una variable aleatoria.

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

con

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

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

5.32 Cuantiles

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 $0,25$ es el primer **cuartil** o sea el valor x_s tal que $q = 0,25$

$$F(x_s) = 0,25$$

5.33 Resumen

nombres de cantidades	modelo (no observado)	datos (observados)
función de masa de probabilidad // frecuencia relativa	$f(x_i) = P(X = x_i)$	$f_i = \frac{n_i}{N}$
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 sobre el origen // 2º momento muestral	$E(X^2) = \sum_{i=1}^{M} x_i^2 f(x_i)$	$m_2 = \sum_{j=1}^{N} x_j^2 / n$

Tenga en cuenta que:

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

Propiedades:

- $\sum_{i=1...N} f(x_i) = 1$
- $f(x_i) = F(x_i) - F(x_{i-1})$
- $E(a \times X + b) = a \times E(X) + b$; para los escalares a y b.
- $V(a \times X + b) = a^2 \times V(X)$
- $E(X^2) = V(X) + E(X)^2$

Chapter 6

Variables aleatorias continuas

6.1 Objetivo

- Función de densidad de probabilidad
- Media y varianza
- Distribución de probabilidad

6.2 Variable aleatoria continua

¿Qué sucede con las variables aleatorias continuas?

Reconsideremos el ángulo de convexidad de los pacientes con misofonía (Sección 2.21).

- Para esta variabler redefinimos los resultados como pequeños intervalos regulares (bins) y calculamos la frecuencia relativa para cada uno de ellos como hicimos en el caso discreto.

```
##         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 Variable aleatoria continua

Consideremos nuevamente que sus frecuencias relativas son las probabilidades cuando $N \rightarrow \infty$

$$f_i = \frac{n_i}{N} \rightarrow f(x_i) = P(X = x_i)$$

La probabilidad depende ahora de la longitud de los bins Δx. Si hacemos los contenedores cada vez más pequeños, las frecuencias se hacen más pequeñas y, por lo tanto,

$P(X = x_i) \rightarrow 0$ cuando $\Delta x \rightarrow 0$, porque $n_i \rightarrow 0$

```
##           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 Variable aleatoria continua

Definimos una cantidad en un punto x que es la cantidad de probabilidad por unidad de distancia que encontraríamos en un contenedor **infinitesimal** dx en x

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

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

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

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

6.5 Variable aleatoria continua

Definición

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

La función es positiva:

- $f(x) \geq 0$

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

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

La probabilidad de observar **cualquier** valor es 1:

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

6.6 Variable aleatoria continua

- La función de densidad de probabilidad es un paso adelante en la abstracción de probabilidades: sumamos el límite continuo ($dx \to 0$).
- Todas las propiedades de las probabilidades se traducen en términos de densidades ($\sum \to \int$).
- La asignación de probabilidades a una variable aleatoria se puede realizar con argumentos de equiprobabilidad (clásicos).
- Las densidades son cantidades matemáticas que algunas asignarán a experimentos y otras no. *¿Qué densidad corresponderá mejor a mi experimento?*

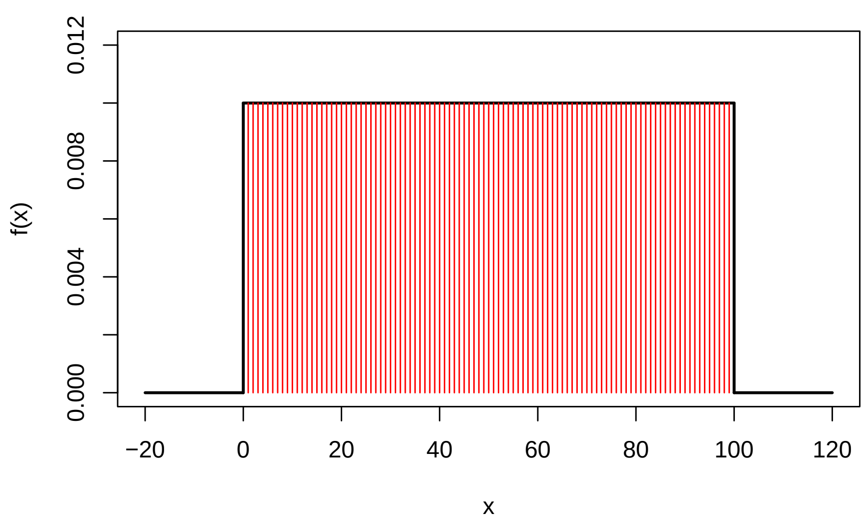
6.7 Área total bajo la curva

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

$$f(x) = \begin{cases} \frac{1}{100}, & \text{si } x \in (0, 100) \\ 0, & si\, no \end{cases}$$

Entonces la probabilidad de **cualquier** observación es el **área total bajo la curva**

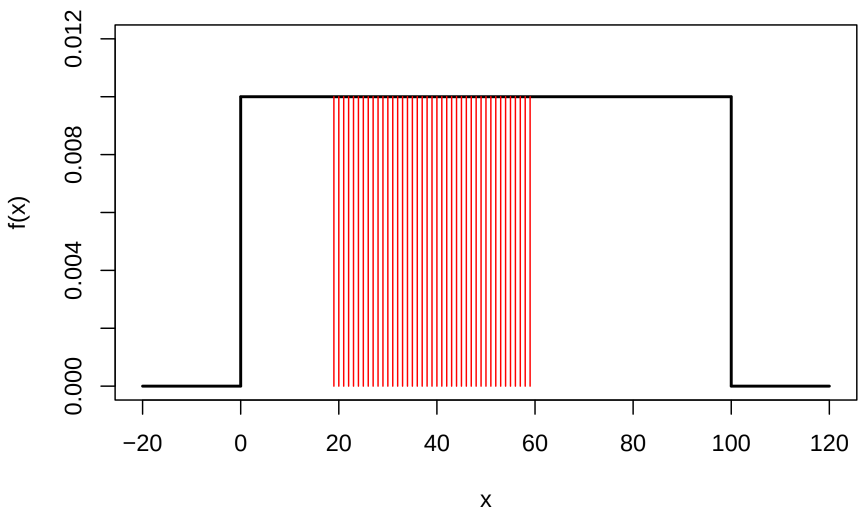
$P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f(x)dx = 100 * 0.01 = 1$



6.8 Área bajo la curva

La probabilidad de observar x en un intervalo es el **área bajo la curva** dentro del intervalo

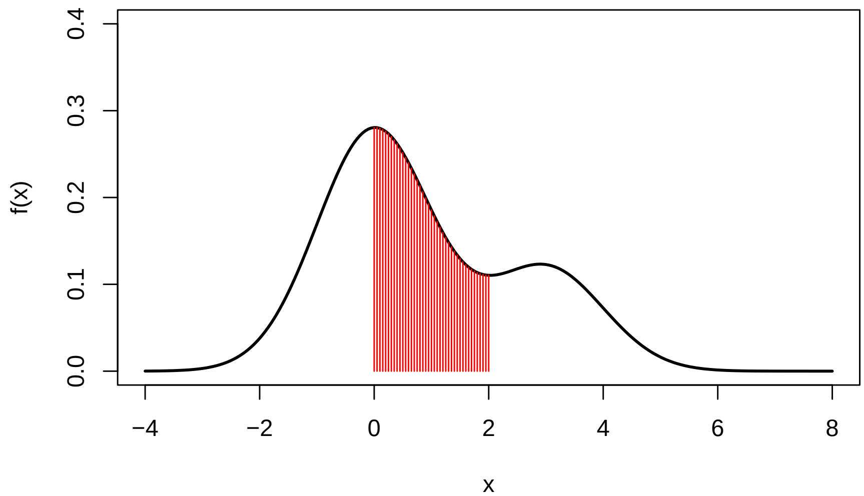
- $P(20 \leq X \leq 60) = \int_{20}^{60} f(x)dx = (60 - 20) * 0.01 = 0.4$



6.9 Área bajo la curva

En general, $f(x)$ debe satisfacer:

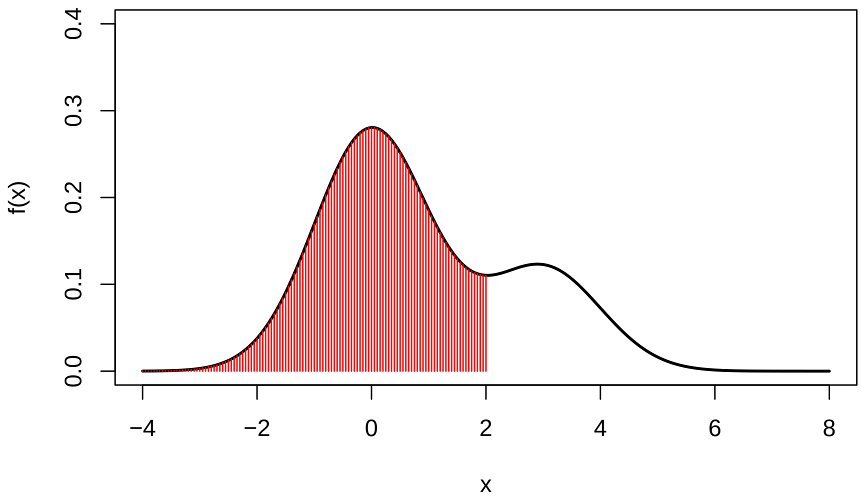
- $0 \leq P(a \leq X \leq b) = \int_{a}^{b} f(x)dx \leq 1$



6.10 Distribución de probabilidad

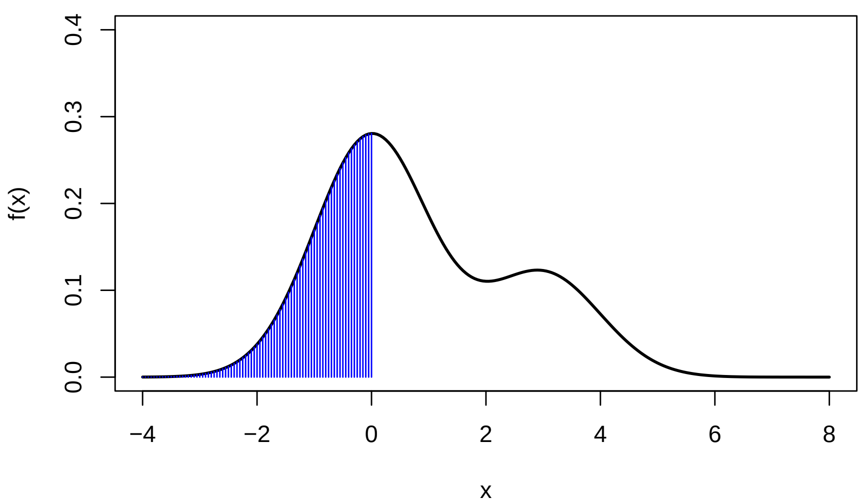
La probabilidad acumulada hasta b está definida por la distribución de probabilidad F

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



La probabilidad acumulada hasta a es

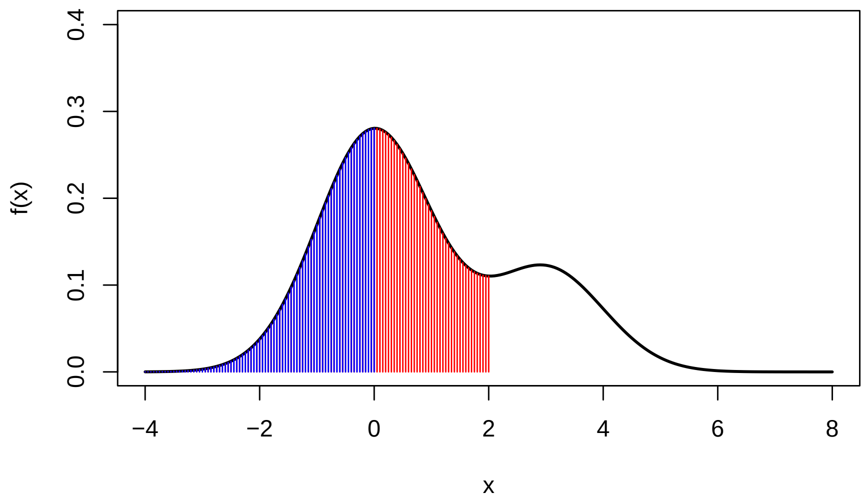
- $F(a) = P(X \leq a)$



6.11 Distribución de probabilidad

La probabilidad entre a y b está definida por la distribución de probabilidad F

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



6.12 Distribución de probabilidad

La distribución de probabilidad de una variable aleatoria continua se define como $F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)dx$

con las propiedades que:

Está entre 0 y 1:

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

Siempre aumenta:

- si $a \leq b$ entonces $F(a) \leq F(b)$

Se puede utilizar para calcular probabilidades:

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

Recupera la densidad de probabilidad:

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

Usamos **distribuciones de probabilidad** para **calcular probabilidades** de una variable aleatoria en intervalos

---

6.13 Distribución de probabilidad

Para la funcion de densidad uniforme:

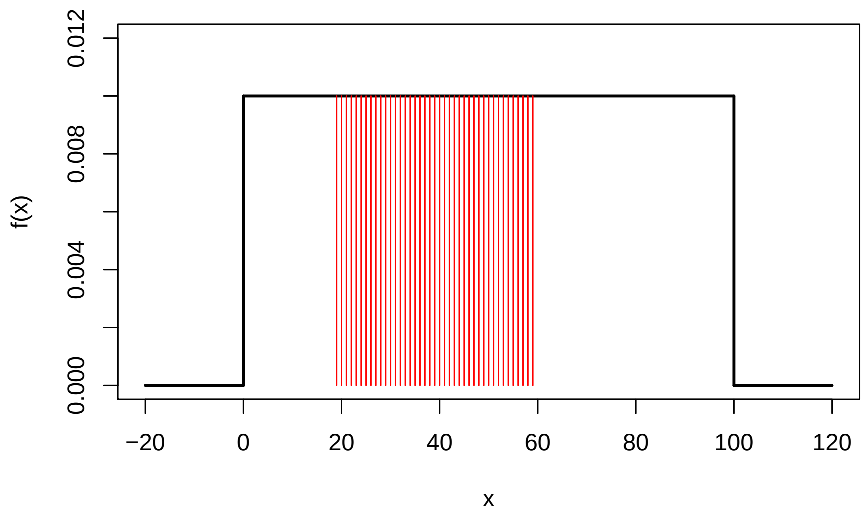
$$f(x) = \begin{cases} \frac{1}{100}, & \text{si } x \in (0, 100) \\ 0, & si\, no \end{cases}$$

La distribución de probabilidad es

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

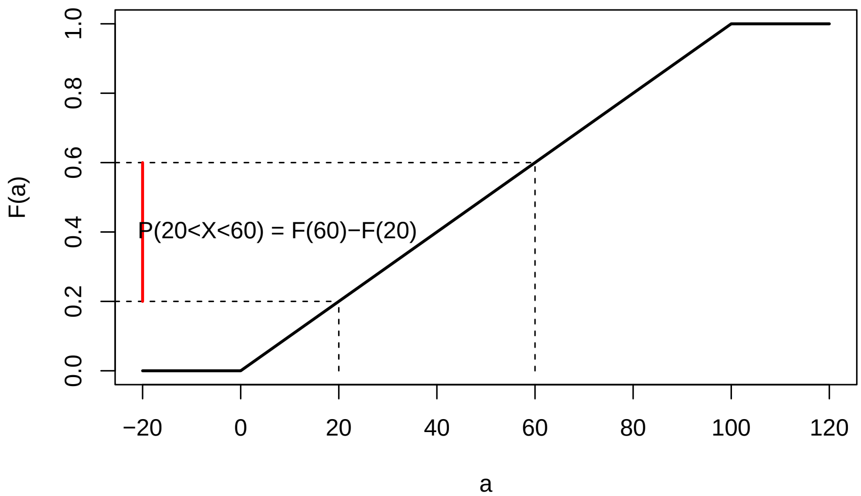
6.14 Gráficos de probabilidad

La probabilidad $P(20 < X < 60)$ es el *área* bajo la curva de **densidad**



6.15 Gráficos de probabilidad

La probabilidad $P(20 < X < 60)$ es la *diferencia* en valores de **distribución**



6.16 Media

Como en el caso discreto, la **media** mide el centro de la distribución

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 μ o $E(X)$, es

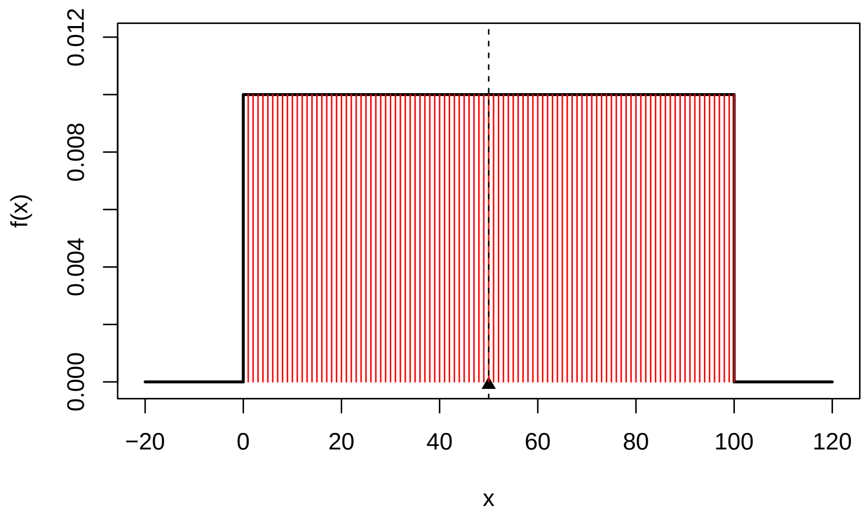
$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Es la versión continua del centro de masa.

6.17 Media

$$f(x) = \begin{cases} \frac{1}{100}, & \text{si } x \in (0, 100) \\ 0, & si\,no \end{cases}$$

$E(X) = 50$



6.18 Varianza

Como en el caso discreto, la varianza mide la dispersión con respecto a 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 σ^2 o $V(X)$, es

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

6.19 Funciones de X

Definición

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

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

Y tenemos las mismas propiedades que en el caso discreto

- 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.

- La varianza de una función lineal de X es:

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

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

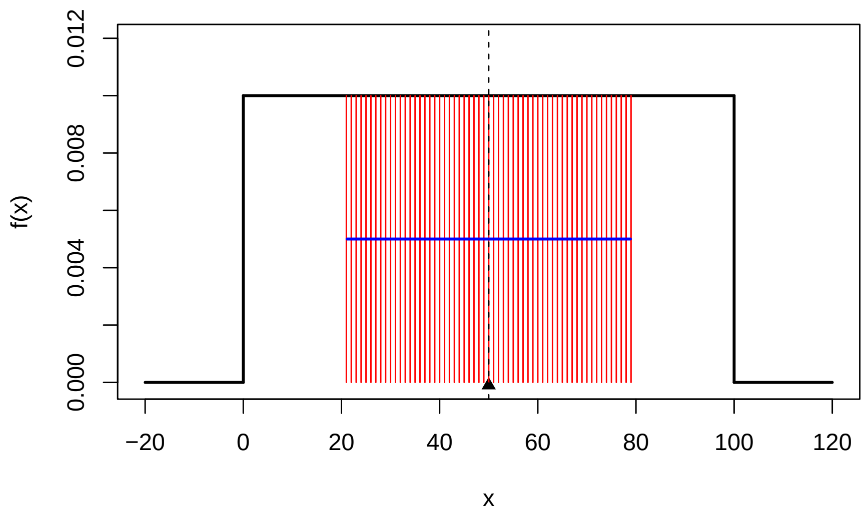
$$E(X^2) = V(X) + E(X)^2$$

6.20 Ejemplo

- para la densidad de probabilidad

$$f(x) = \begin{cases} \frac{1}{100}, & \text{si } x \in (0, 100) \\ 0, & si\, no \end{cases}$$

- calcule la media
- calcule la varianza usando $E(X^2) = V(X) + E(X)^2$
- calcule $P(\mu - \sigma \leq X \leq \mu + \sigma)$
- ¿Cuáles son el primer y tercer cuartiles?


0.012
0.008
0.004
0.000
f(x)
−20
0
20
40
60
80
100
120
x

Chapter 7

Modelos de probabilidad para variables aleatorias discretas

7.1 Objetivo

Modelos probabilidad:

- Funciones de probabilidad uniforme y de Bernoulli
- Funciones de probabilidad binomial y binomial negativa

7.2 Función de probabilidad

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

es positiva:

- $f(x_i) \geq 0$

Nos permite calcular probabilidades:

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

La probabilidad de observar algún resultado es 1

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

Propiedades:

Tendencia central:

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

Dispersión:

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

Son objetos abstractos con propiedades generales que pueden o no **describir** un proceso natural o de ingeniería.

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:

- $f(x) = P(X = x) = 1/6$ representa la probabilidad de los resultados de **una** tirada de dados.
- La función de masa de probabilidad

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 por etiqueta: $-1, 0, 1$ y una bola por etiqueta: $-2, 2$.

7.4 Modelos paramétricos

Cuando realizamos un experimento aleatorio y **no** sabemos las probabilidades de los resultados:

- Siempre podemos formular el modelo dado por las frecuencias relativas: $\hat{P}(X = x_i) = f_i$ (donde $i = 1...M$).

Necesitamos encontrar M números cada uno dependiendo de N.

En muchos casos:

- Podemos formular funciones de probabilidad $f(x)$ que dependen solamente de **muy pocos** números.

Ejemplo:

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

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

Solo necesitamos saber M.

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

7.5 Distribución uniforme (un parámetro)

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

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

Con media y varianza:

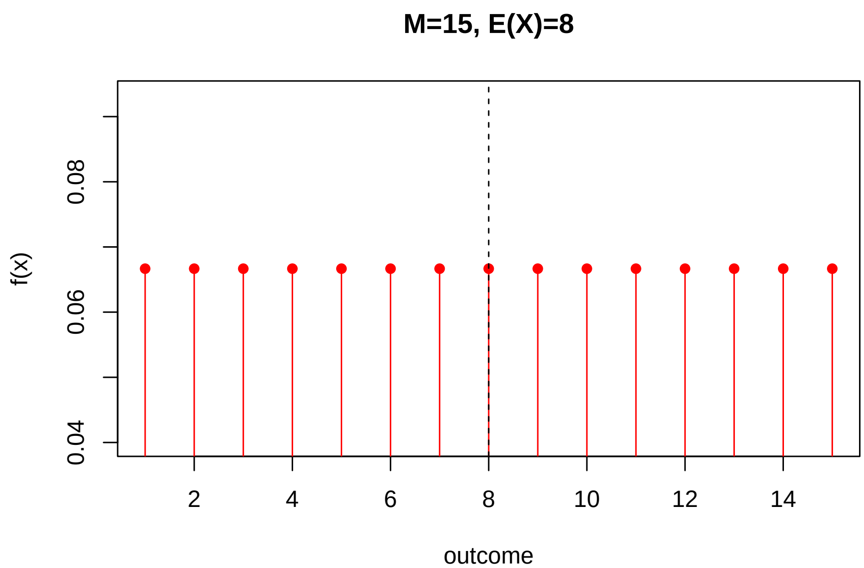
$E(X) = \frac{M+1}{2}$

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

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

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

7.6 Distribución uniforme



7.7 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

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

como $M = b - a + 1$.

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

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

7.8 Distribución uniforme (dos parámetros)

Ejemplo:

¿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)?

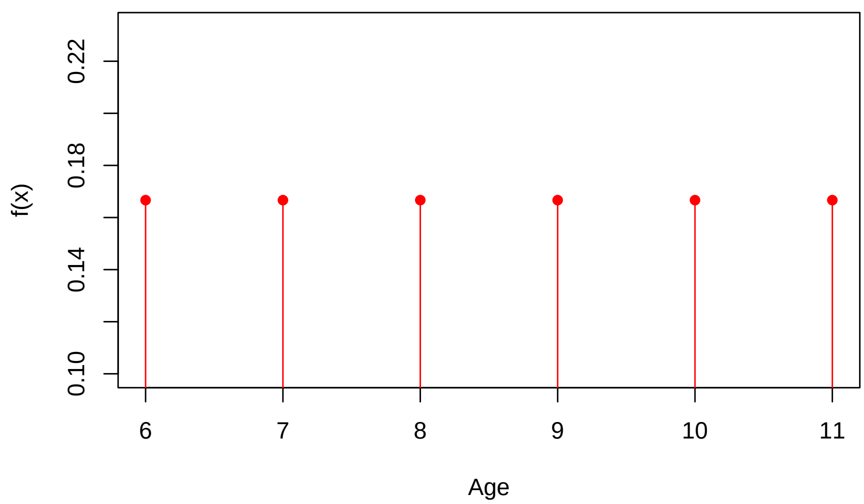
Del experimento sabemos: $a = 6$ y $b = 11$ entonces

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

eso es

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

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



7.9 Distribución uniforme

El modelo de probabilidad de una variable aleatoria X

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

para $x \in \{a, a+1, ...b\}$

tiene media y varianza:

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

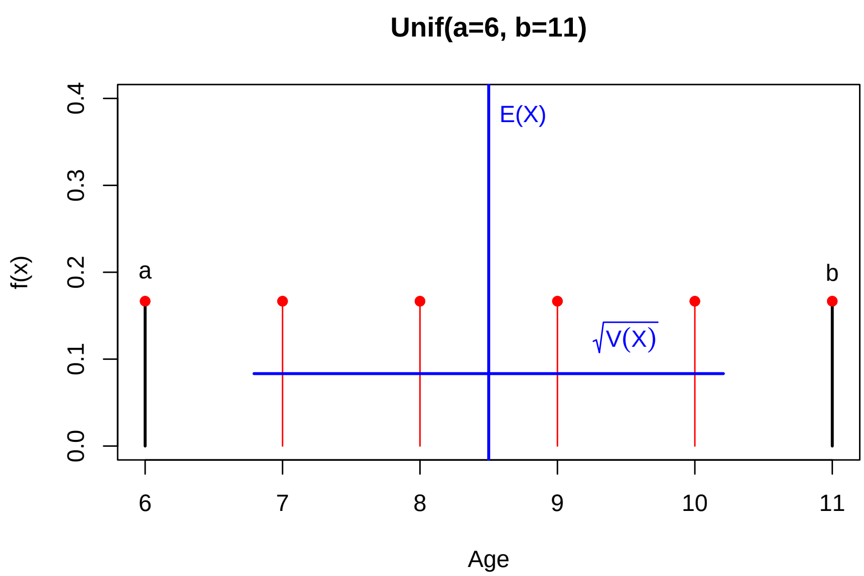
(Cambiar variables $X = Y + a - 1$, $y \in \{1, ...M\}$)

Podemos especificar a y b o $E(X)$ y $V(X)$.

En nuestro ejemplo:

- $E(X) = (11+6)/2 = 8.5$
- $V(X) = (6^2-1)/12 = 2.916667$

7.10 Distribución uniforme (dos parámetros)



7.11 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**
- El conocimiento de $f(x)$ se reduce al conocimiento del valor de los parámetros
- Idealmente, el modelo y los parámetros son **interpretables**

Ejemplo:

Modelo: Los datos de nuestro experimento se producen mediante un proceso aleatorio en el que cada edad tiene la **misma probabilidad** de ser observada.

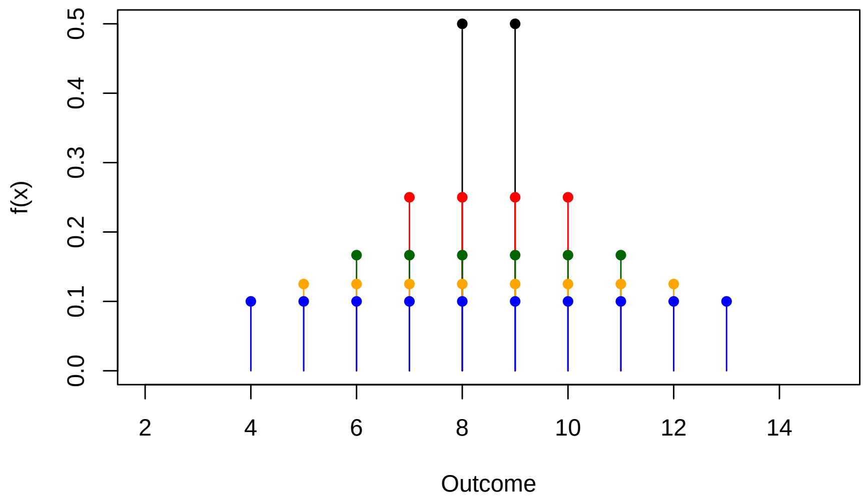
Parámetros: a es la edad mínima, $E(X)$ es la edad esperada… son **propiedades físicas** del experimento.

7.12 Parámetros y Modelos

Ejemplo:

Una **familia** de modelos obtenidos a partir de distribuciones uniformes de dos parámetros cambiando las **varianzas** y manteniendo una media constante ($E(X) = 8.5$). Da como resultado **cambiar** los resultados **mínimo** y **máximo**.

- Nota: solo un modelo tiene sentido para nuestro experimento (solo un modelo puede representar las edades de los niños en una escuela).



- Podemos pensar en **familias** que cambian solo en la **media**, solo el **mínimo** o solo el **máximo**

7.13 Ensayo de Bernoulli

Intentemos avanzar desde el caso de probabilidad igual y supongamos un modelo con dos resultados (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 está defectuosa o no ($A: defectuosa$ y $B: buena$).
- Dar en el blanco ($A: xito$ y $B: fracaso$).
- Transmitiendo un píxel correctamente ($A: s$ y $B: no$).

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

7.14 Ensayo de Bernoulli

Introduciremos la probabilidad de un resultado (A) como el **parámetro** del modelo:

- resultado A (éxito): tiene probabilidad p (parámetro)
- resultado B (fracaso): tiene una probabilidad $1 - p$

O podemos escribir la función de masa de probabilidad de K tomando valores $\{0, 1\}$ para A y B

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

o más en breve

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

para $k = (0, 1)$

Solo necesitamos saber p.

7.15 Ensayo de Bernoulli

Una variable de Bernoulli K con resultados $\{0, 1\}$ tiene una función de masa de probabilidad

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

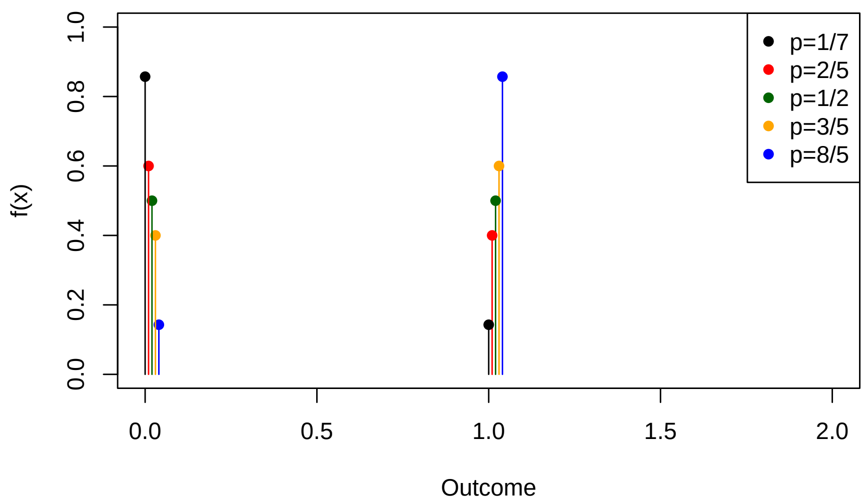
Con media y varianza:

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

Nota:

- La probabilidad del resultado A es el parámetro p que es lo mismo que $f(0) = P(X = 0)$.
- Como p suele ser **desconocido**, normalmente lo estimamos por la frecuencia relativa (más sobre esto en las secciones de inferencia): $\hat{p} = f_A = \frac{n_A}{N}$

7.16 Ensayo de Bernoulli



7.17 Distribución binomial

Cuando estamos interesados en aprender sobre un ensayo de Bernoulli en particular

- Repetimos el ensayo de Bernoulli N veces y contamos cuantas veces obtuvimos A (n_A).
- Definimos una variable aleatoria $X = n_A$ tomando valores $x \in 0, 1, ...N$

Ahora preguntamos por la probabilidad de observar x eventos de tipo A en la repetición de n ensayos independientes de Bernoulli, cuando la probabilidad de observar A es p.

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

7.18 Ejemplos: distribución binomial

- Anotar el sexo de $n = 10$ pacientes que acuden a urgencias de un hospital. ¿Cuál es la probabilidad de que $x = 6$ los pacientes sean hombres cuando $p = 0,9$?
- Intentar $n = 5$ veces para dar en el blanco ($A : xito$ y $B : fracaso$). ¿Cuál es la probabilidad de que alcance el objetivo $x = 5$ veces cuando normalmente lo hago el 20% de las veces ($p = 0,25$)?
- Transmitiendo $n = 100$ píxeles correctamente ($A : s$ y $B : no$). ¿Cuál es la probabilidad de que $x = 2$ píxeles sean errores, cuando la probabilidad de error es $p = 0,1$?

7.19 Distribución binomial

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

Considere las variables aleatorias 4: K_1, K_2, K_3 y K_4 que registran si se ha cometido un error en el 1^o, 2^o, 3^r y 4^o píxel.

Por lo tanto

- k_i toma valores $\{correcto : 0; error : 1\}$
- $X = \sum_{i=1}^{4} K_i$ toma valores $\{0, 1, 2, 3, 4\}$

Entonces la probabilidad de observar 4 errores es:

- $P(X = 4) = P(1, 1, 1, 1) = p * p * p * p = p^4$ porque K_i son independientes.

La probabilidad de observar 0 errores es:

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

La probabilidad de errores de 3 es:

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

7.20 Distribución binomial

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.

7.21 Distribución binomial: 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

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

$\binom{n}{x} = \frac{n!}{x!(n-x)!}$ 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.

7.22 Distribución binomial: Media y Varianza

La media y la varianza de $X \to Bin(n, p)$ son

- $E(X) = np$
- $V(X) = np(1 - p)$
- Dado que X es la suma de n variables independientes de Bernoulli

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

y

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

Ejemplo:

- 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 $n(1 - p)p = 0.36$

Recuerde: Podemos especificar los parámetros n y p, o los parámetros $E(X)$ y $V(X)$

7.23 Ejemplo 1

Ahora respondamos:

- ¿Cuál es la probabilidad de observar 4 errores al transmitir 4 píxeles, si la probabilidad de un error es de 0.1?

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

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

Eso es

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

7.24 Ejemplo 1

- Queremos calcular:

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

En R dbinom(4,4,0.1)

- También podemos calcular:

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

En R dbinom(2,4,0.1)

7.25 Ejemplo 2

- ¿Cuál es la probabilidad de observar a lo mucho 8 votantes del partido de gobierno en una encuesta electoral de tamaño 10, si la probabilidad de un voto positivo es de 0.9?

Para este caso

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

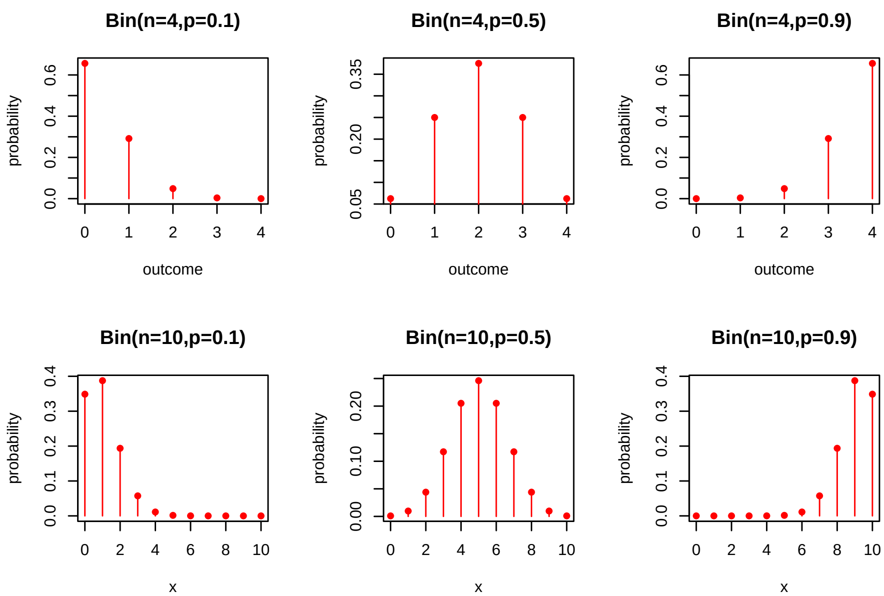
Eso es

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

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

en R pbinom(8,10, 0.9)

7.26 Distribución binomial



7.27 Distribución 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.

- Experimento: Supongamos que realizamos ensayos de Bernoulli hasta que observamos que el resultado A aparece r veces.

- Variable aleatoria: Contamos el número de eventos B

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

7.28 Distribución binomial negativa

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

$(0, 0, 1, ., 0, 1, ...0, 1)$ (hay y ceros y r unos)

Observamos y píxeles correctos en un total de $y + r$ intentos.

Por lo tanto

- $P(0, 0, 1, ., 0, 1, ...0, 1) = (1 - p)^y p^r$ (Recuerda: p es la probabilidad de error)

¿Cuántas transmisiones pueden tener y píxeles correctos antes de r errores?

Nota:

- El último bit es fijo (marca el final de la transmisión)
- El número total de transmisiones con y número de píxeles correctos (B) que podemos obtener en $y + r - 1$ intentos es: $\binom{y+r-1}{y}$

7.29 Distribución binomial negativa

Por lo tanto, la probabilidad de observar y eventos de tipo B antes de r eventos de tipo A (con probabilidad p) es

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

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.

7.30 Media y Varianza

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

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

7.31 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

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

$$Y \to Geom(p)$$

con media

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

7.32 Ejemplo

- Un sitio web tiene tres servidores.
- Un servidor opera a la vez y solo cuando falla una solicitud se utiliza 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 los tres servidores fallen?

7.33 Ejemplo

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

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

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

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

- Tenga en cuenta que para enviar este número de solicitures hemos enviado un total de $6000 = 5997 + 3$

7.34 Ejemplo

¿Cuál es la probabilidad de tratar con éxito como máximo 5 solicitudes antes de que el sistema falle?

Recuerde la función de distribución: $F(y) = P(Y \leq 5)$

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

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

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

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

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

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

En R pnbinom(5,3,0.0005)

7.35 Ejemplos

Con la función de probabilidad binomial negativa:

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

Ahora podemos responder preguntas como:

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

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

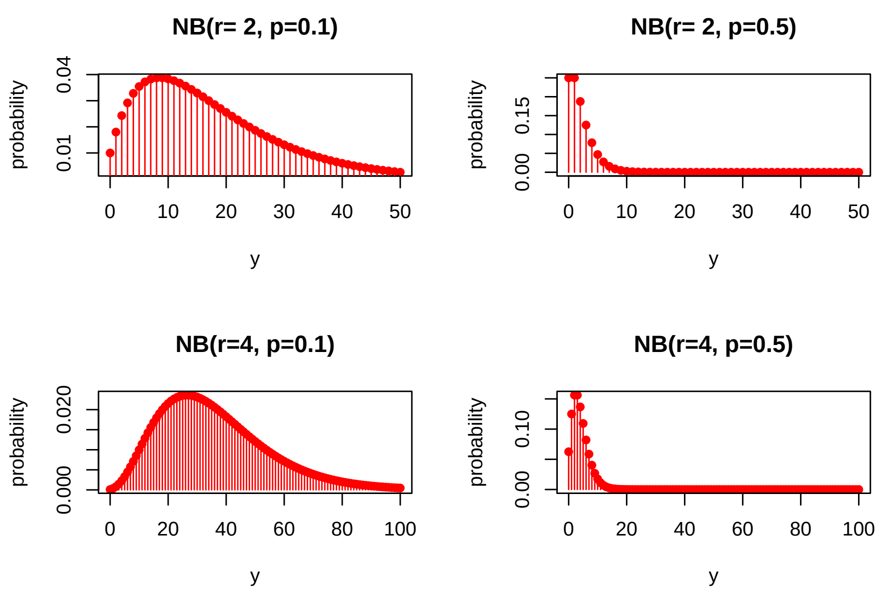
en R dnbinom(10, 2, 0.1)

- ¿Cuál es la probabilidad de que entren 2 chicas antes que 4 chicos si la probabilidad de que entre una chica es de 0.5?

$f(2; r = 4, p = 0.5) = 0.15625$

en R dnbinom(2, 4, 0.5)

7.36 Distribución binomial negativa



7.37 Distribución hipergeométrica

La probabilidad de obtener x casos de hepatitis C en una muestra de n extraída de una población de N donde K tiene hepatitis C es

$P(X = x) = P(one\,sample) \times (Number\,of\,ways\,of\,obtaining\,x)$

$$= \frac{1}{\binom{N}{n}} \binom{K}{x} \binom{N-K}{n-x}$$

donde $k \in \{\max(0, n+K-N), ... \min(K, n)\}$

$$X \to Hypergeometric(N, K, n)$$

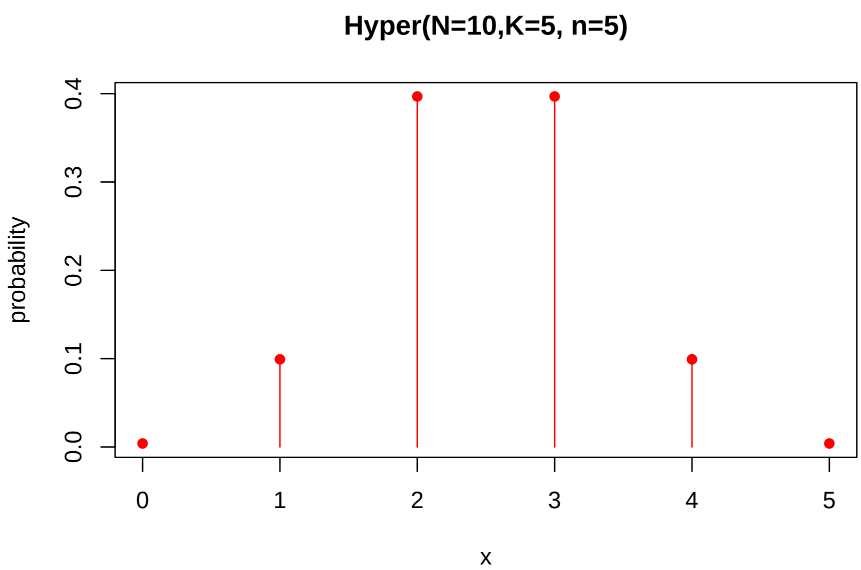
7.38 Distribución hipergeométrica

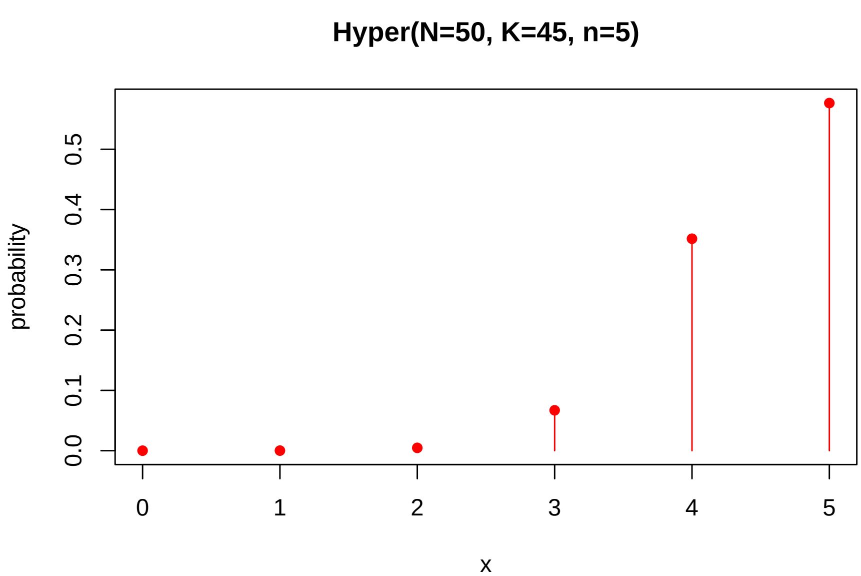
Una variable hipergeométrica tiene

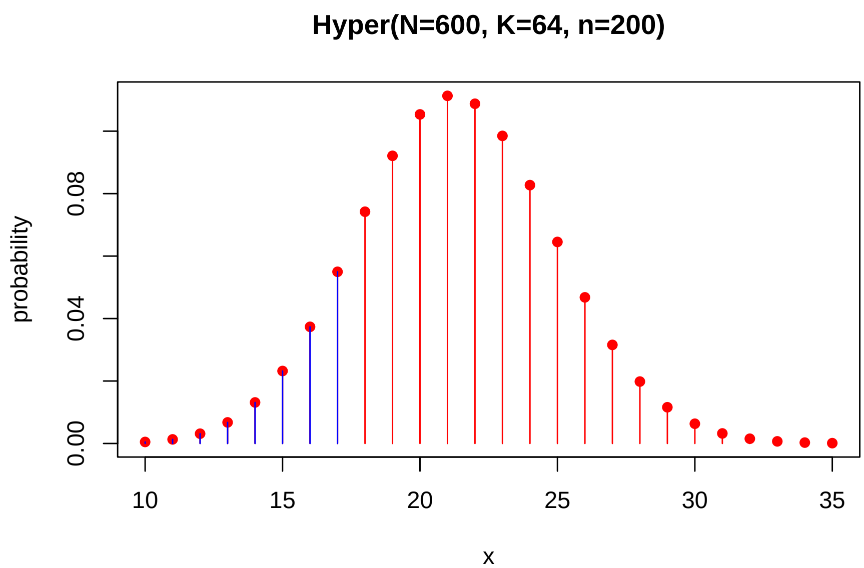
- media: $E(X) = n\frac{K}{N} = np_0$
- varianza: $V(X) = np_0(1-p_0)\frac{N-n}{N-1}$

cuando $p_0 = \frac{K}{N}$ es la proporción de hepatitis C en una población de tamaño N.

7.39 Distribución hipergeométrica






Hyper(N=600, K=64, n=200)
probability
0.00
0.04
0.08
10
15
20
25
30
35
x

Chapter 8

Modelos de Poisson y Exponencial

8.1 Objetivo

Modelo de probabilidad discreta:

- Poisson

Modelo de probabilidad continua:

- Exponencial

8.2 Modelos de probabilidad discreta

Estamos construyendo modelos más complejos a partir de modelos simples:

Uniforme: interpretación clásica de la probabilidad $\downarrow$ **Bernoulli**: Introducción de un **parámetro** p (familia de modelos) $\downarrow$ **Binomial**: **Repetición** de un experimento aleatorio (n-veces ensayos de Bernoulli) $\downarrow$ **Poisson**: Repetición de un experimento aleatorio dentro de un intervalo continuo, **sin control** sobre cuándo/dónde ocurre la ensayo de Bernouilli.

8.3 Contando eventos

Imagine que estamos observando eventos que **dependen** de **intervalos** de tiempo o distancia.

- coches que llegan a un semáforo
- mensajes 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**.

8.4 Contando eventos

¿Cuál es la probabilidad de observar X eventos en una unidad de intervalo (tiempo o distancia)?

Imagine que algunas impurezas en un alambre de cobre se depositan al azar a lo largo de un alambre

- en cada centímetro, contamos un promedio de $\lambda = 10/cm$.
- dividimos el centímetro en micrómetros ($0.0001cm$)

8.5 Distribución de Poisson

Los micrómetros son lo suficientemente pequeños tal que

- hay o no hay una impureza en cada micrómetro
- cada micrómetro puede considerarse un **ensayo de Bernoulli**

8.6 Distribución de Poisson

La probabilidad de observar X impurezas en $n = 10,000\mu$ (1cm) sigue aproximadamente una distribución binomial

$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$

donde p es la probabilidad de encontrar una impureza en un micrómetro.

Recuerdemos que $E(X) = np$ entonces para $\lambda = np$ (número promedio de impurezas por 1 cm), podemos escribir

$$P(X = x) = \binom{n}{x}(\frac{\lambda}{n})^x(1 - \frac{\lambda}{n})^{n-x}$$

- **Podría** todavía haber dos impurezas en un micrómetro, por lo que debemos aumentar la partición del cable y $n \to \infty$.

Entonces en el límite:

$$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

Donde λ es constante porque es la densidad de impurezas por centímetro, una **propiedad física** del sistema.

8.7 Distribución de Poisson: detalles de la derivación

Para $P(X = x) = \binom{n}{x}(\frac{\lambda}{n})^x(1 - \frac{\lambda}{n})^{n-x}$

en el límite ($n \to \infty$)

- $\frac{1}{n^x}\binom{n}{x} = \frac{1}{n^x}\frac{n!}{x!(n-x)!} = \frac{(n-x)!(n-x+1)\ldots(n-1)n}{n^x x!(n-x)!} = \frac{n(n-1)..(n-x+1)}{n^x x!} \to \frac{1}{x!}$
- $(1 - \frac{\lambda}{n})^n \to e^{-\lambda}$ (definición de exponencial)
- $(1 - \frac{\lambda}{n})^{-x} \to 1$

Por lo tanto $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

8.8 Distribución de Poisson

Definición

Dado

- un intervalo en los números reales
- los eventos ocurren al azar en el intervalo
- se conoce el número medio de conteos en el intervalo (λ)
- se puede encontrar una pequeña partición regular del intervalo tal que cada uno de ellos pueda considerarse un ensayo de Bernoulli

Entonces ...

8.9 Distribución de Poisson

Definición

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:

- media $E(X) = \lambda$
- varianza $V(X) = \lambda$

8.10 Distribución de Poisson

Con la función de probabilidad de Poisson:

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

para $x \in \{0, 1, ...\}$

Podemos responder preguntas como:

- ¿Cuál es la probabilidad de recibir 4 correos electrónicos en una hora, cuando el promedio de correos electrónicos en una horas es de 1?

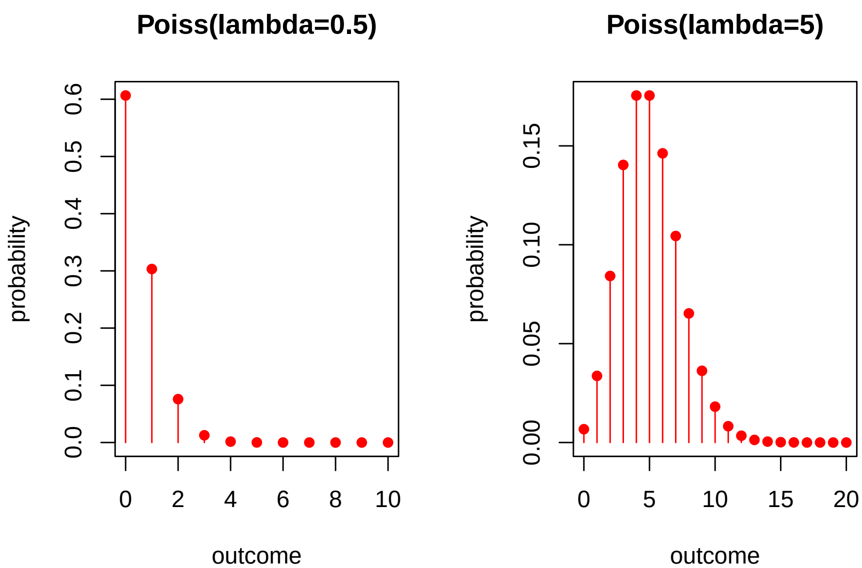
$f(4; \lambda = 1) = 0.18$

in R dpois(2,1)

- ¿Cuál es la probabilidad de contar al menos 10 coches que llegan a un peaje en un minuto, cuando el promedio de autos que llegan a un peaje en un minuto es de 5; $P(X \leq 10) = F(10; \lambda = 5) = 0.98$?

in R ppois(10,5)

8.11 Distribución de Poisson



8.12 Modelos de probabilidad continua

Los modelos de probabilidad continua son funciones de densidad de probabilidad $f(x)$ de variables aleatorias continuas que **creemos** describen experimentos aleatorios reales.

Definición:

Positiva:

- $f(x) \geq 0$

Permite calcular probabilidades usando el área bajo la curva:

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

La probabilidad de observar algún valor es 1:

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

8.13 Densidad exponencial

Volvamos a la probabilidad de Poisson para el número de eventos (k) en un intervalo

$$f(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \lambda > 0$$

- Consideremos ahora solo el primer evento (duración/tiempo)
- la distancia/tiempo que debemos esperar hasta el primer evento es una variable aleatoria **continua**.

Podemos preguntar por la probabilidad de que el primer evento esté a la distancia X.

8.14 Densidad exponencial

La probabilidad de no observar ningún evento **si** un intervalo tiene unidad x 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 en una distancia x: $f(K = 0|X = x)$ y aplicar el teorema de Bayes para invertirlo:

$$f(x|0) = Cf(0|x) = Ce^{-x\lambda}$$

Entonces podemos calcular la **probabilidad de observar una distancia** x con 0 eventos (esta es la distancia hasta el primer evento, o la distancia entre dos eventos).

8.15 Densidad exponencial

En un proceso de Poisson con parámetro λ la probabilidad de esperar una distancia/tiempo X hasta el primer evento tiene una **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

$$f(x) = \lambda e^{-\lambda x}$$

8.16 Densidad exponencial

Una variable aleatoria exponencial X tiene una densidad de probabilidad

$$f(x) = \lambda e^{-\lambda x}, x \geq 0$$

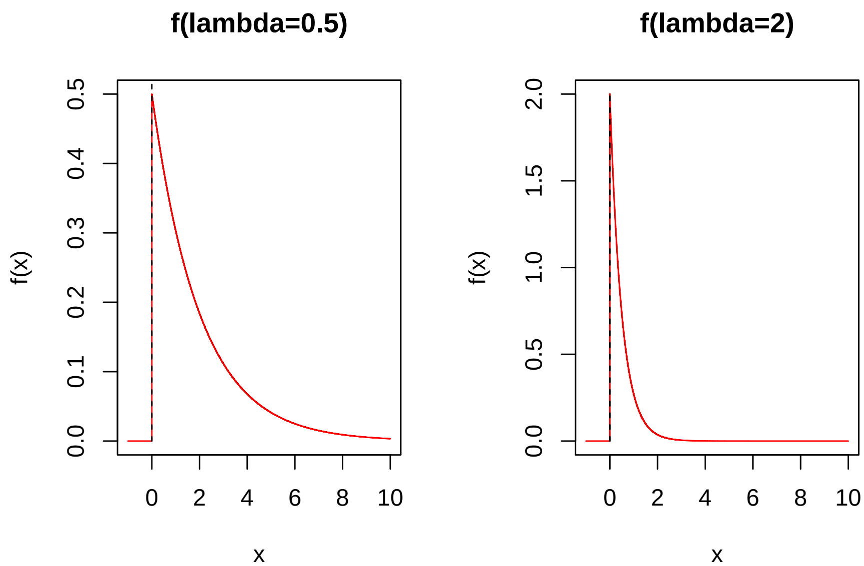
Propiedades:

- Media: $E(X) = \frac{1}{\lambda}$
- Varianza: $V(Y) = \frac{1}{\lambda^2}$

Donde λ es su único parámetro, conocido como **tasa de decaimiento**.

Nota: El modelo exponencial es un modelo general. Puede describir el tiempo/duración hasta la primera cuenta en un proceso de Poisson del tamaño de un huevo hecho por un taladro.

8.17 Densidad exponencial



8.18 Distribución exponencial

En un proceso de Poisson: ¿Cuál es la probabilidad de observar una distancia **menor** que a hasta el primer event?

Recuerde que esta probabilidad $F(a) = P(X \leq a)$ es la densidad de probabilidad

$$F(a) = \lambda \int_{\infty}^{a} e^{-x\lambda} dx = 1 - e^{-a\lambda}$$

- ¿Cuál es la probabilidad de observar una distancia **mayor** que a hasta el primer evento?

$$P(X > a) = 1 - P(X \leq a) = 1 - F(a) = e^{-a\lambda}$$

8.19 Distribución exponencial

Con la función de densidad exponencial:

$$f(x) = \lambda e^{-\lambda x}$$

Podemos responder preguntas como:

- ¿Cuál es la probabilidad de que tengamos que esperar un bus por más de 1 hora cuando en promedio hay dos buses por hora?

$$P(X > 1) = 1 - P(X \leq 1) = 1 - F(1, \lambda = 2) = 0.1353$$

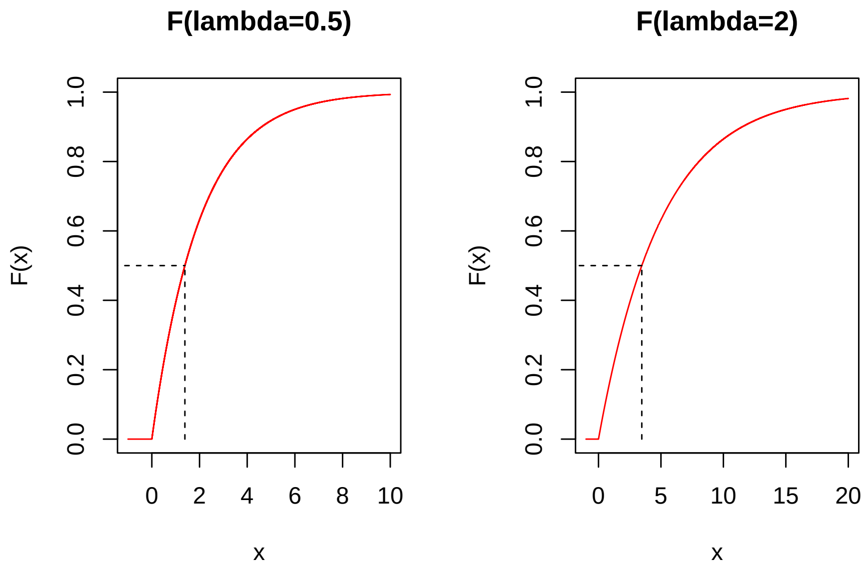
en R 1-pexp(1,2)

- ¿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 cada segundo? $F(2, \lambda = 2)$

$$P(X \leq 2) = F(2, \lambda = 2) = 0.981$$

en R pexp(2,2)

8.20 Distribución exponencial



La mediana x_m es tal que $F(x_m) = 0.5$. Eso es $x_m = \frac{\log(2)}{\lambda}$

Chapter 9

Distribución normal

9.1 Objetivo

Modelo de probabilidad para valiables continuas:

- Distribución normal

9.2 Modelo de probabilidad para valiables continuas

Modelo de probabilidad para valiables continuas son funciones de densidad de probabilidad $f(x)$ de variables aleatorias continuas que **creemos** describen experimentos aleatorios reales.

Definición:

Positiva:

- $f(x) \geq 0$

Permite calcular probabilidades usando el área bajo la curva:

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

La probabilidad de algún valor es 1:

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

9.3 Densidad normal

En 1801 Gauss analizó la órbita de Ceres (gran asteroide entre Marte y Júpiter).

- La gente sospechaba que era un nuevo planeta.
- Las medidas tenían errores.
- Le interesaba saber cómo se distribuían las observaciones para poder encontrar la órbita más probable.
- Quería predecir hacia dónde deberían apuntar los astrónomos sus telescopios para encontrarlo unos meses después de que hubiera pasado por detrás del Sol.

9.4 Densidad normal

Errores debidos a la medición.

9.5 Densidad normal

Él asumió que

- los errores pequeños eran más probables que los errores grandes
- el error a una distancia $-\epsilon$ o ϵ de la medida más probable era igualmente probable
- la altitud más **probable** de Ceres en un momento dado en el cielo era el **promedio** de varias mediciones de altitud en esa latitud.

9.6 Densidad normal

Eso fue suficiente para mostrar que las desviaciones aleatorias y **de la órbita** distribuidas como

$$f(y) = \frac{h}{\sqrt{\pi}} e^{-h^2 y^2}$$

*La evolución de la distribución Normal, Saul Stahl, Revista de Matemáticas, 2006.

9.7 Densidad normal

Escribamos la distribución de errores

$$f(y) = \frac{h}{\sqrt{\pi}} e^{-h^2 y^2}$$

para los errores de medidas desde el horizonte X entonces $y = x - x_0$

$$f(x) = \frac{h}{\sqrt{\pi}} e^{-h^2 (x-x_0)^2}$$

- La **media** de esta densidad de probabilidad es:

$E(X) = \mu = x_0$, que representa la **verdadera** posición de Ceres desde el horizonte (propiedad del sistema físico).

- La **varianza** es:

$V(X) = \sigma^2 = \frac{1}{2h^2}$, que representa la dispersión del error en las observaciones (propiedad del sistema de medida).

9.8 Definición

Una variable aleatoria X definida en los números reales tiene una densidad **Normal** si toma la forma

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

con media y varianza:

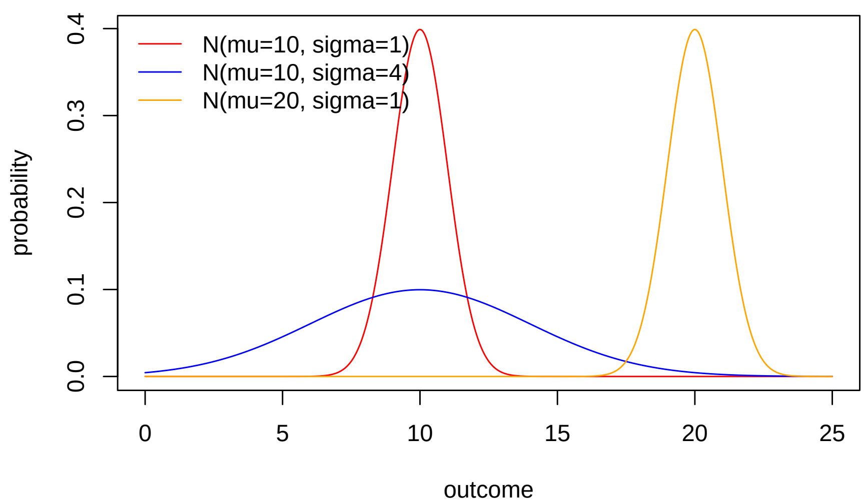
- $E(X) = \mu$
- $V(X) = \sigma^2$

μ y σ 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

$$X \rightarrow N(\mu, \sigma^2)$$

9.9 Densidad de probabilidad normal (gaussiana)



9.10 Distribución normal

La distribución de probabilidad de la densidad Normal:

$$F_{normal}(a) = P(Z \leq a)$$

es la función de **error** definida por el área bajo la curva de $-\infty$ a a

$$F_{normal}(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

La función se encuentra en la mayoría de los programas informáticos.

9.11 Distribución normal

Cuando

$$X \to N(\mu, \sigma^2)$$

Podemos hacer preguntas como:

- ¿Cuál es la probabilidad de que una mujer en la población mida como máximo $150cm$ de altura si las mujeres tienen una altura media de $165cm$ con una desviación estándar de $8cm$?

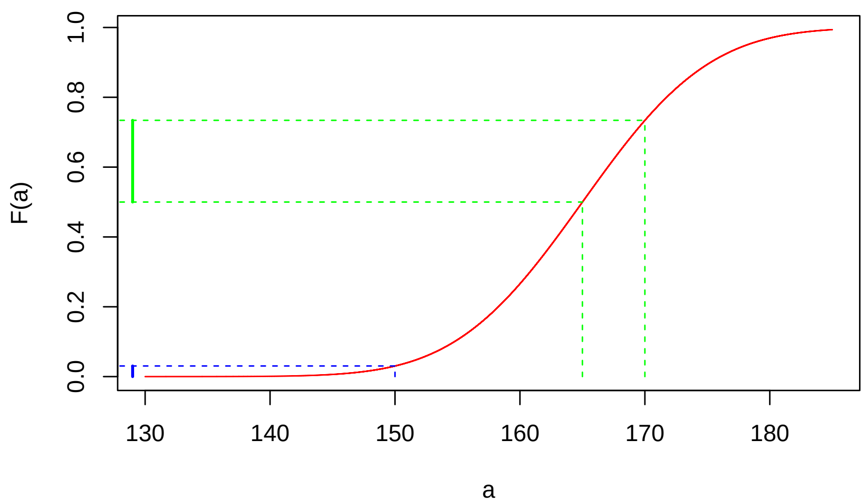
$P(X \leq 150) = F(150, \mu = 165, \sigma = 8) = 0.03039636$

in R pnorm(150, 165, 8) - ¿Cuál es la probabilidad de que la altura de una mujer en la población esté entre $165cm$ y $170cm$?

$P(165 \leq X \leq 170) = F(170, \mu = 165, \sigma = 8) - F(165, \mu = 165, \sigma = 8) = 0.2340145$

in R pnorm(170, 165, 8)-pnorm(165, 165, 8)

9.12 Distribución normal



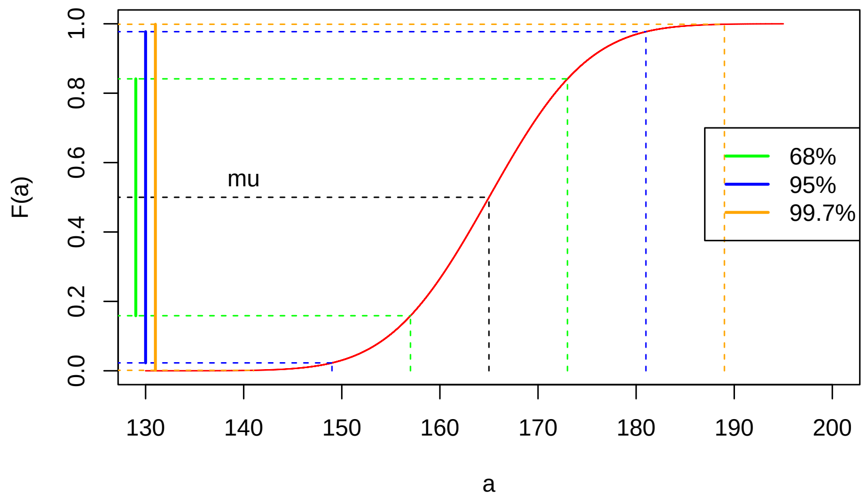
9.13 Distribución normal

- la media μ es también la mediana ya que divide las medidas en dos
- Los valores de x que caen más allá de 2σ se consideran **raros** 5%
- Los valores de x que caen más allá de 3σ se consideran **extremadamente raros** 0.2%

9.14 Distribución normal

Podemos definir los límites de **observaciones comunes** para la distribución de la altura de las mujeres en la población.

- $P(165 - 8 \leq X \leq 165 + 8) = P(157 \leq X \leq 173) = 0.68$
- $P(165 - 2 \times 8 \leq X \leq 165 + 2 \times 8) = P(149 \leq X \leq 181) = 0.95$
- $P(165 - 3 \times 8 \leq X \leq 165 + 3 \times 8) = P(141 \leq X \leq 189) = 0.997$



9.15 Densidad normal estándar

Cambiemos las variables a una **variable estandarizada**

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

en la densidad

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

reemplazando $x = \sigma z + \mu$ y $dx = \sigma dz$ en la expresión de probabilidad que tenemos

$P(x \leq X \leq x + dx) = P(z \leq Z \leq z + dz)$

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

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

obtenemos la forma **estandarizada** de la densidad normal.

9.16 Densidad normal estándar

Definición

Una variable aleatoria Z definida en los números reales tiene una densidad **estándar** si toma la forma

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz, z \in \mathbb{R}$$

con media y varianza

- $E(X) = 0$
- $V(X) = 1$.

9.17 Densidad normal estándar

La densidad estándar:

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz, z \in \mathbb{R}$$

- es la densidad normal $N(\mu = 0, \sigma^2 = 1)$
- cualquier variable distribuida normalmente X puede transformarse en una variable Z

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

que sigue una distribución estándar:

$$Z \to N(0, 1)$$

9.18 Distribución normal

Todas las densidades normales se pueden obtener a partir de la densidad estándar con los valores de μ y σ

9.19 Distribución estándar

La distribución de probabilidad de la densidad estándar:

$$\phi(a) = F_{estndar}(a) = P(Z \leq a)$$

es la función **error** definida por

$$\phi(a) = \int_{-\infty}^{a} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$

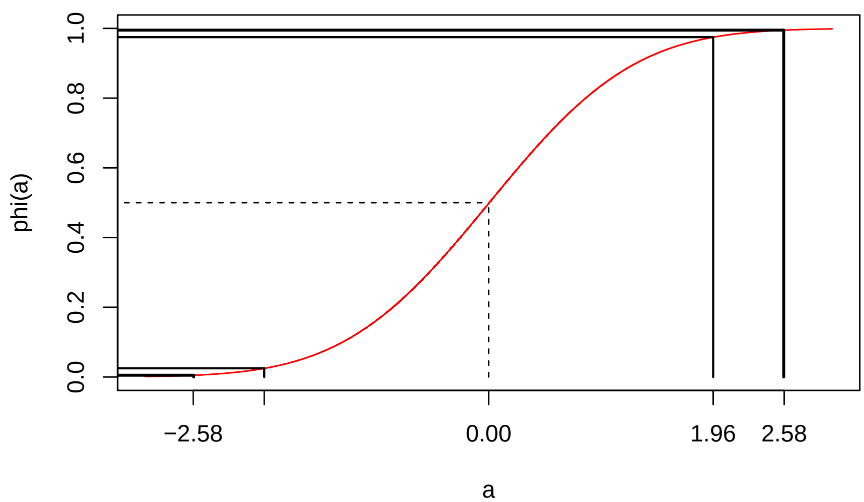
You can find it in most computer programs

9.20 Standard normal density

9.21 Densidad normal estándar

Definimos los límites de las **observaciones más comunes** para la variable estándar

- $P(-0.67 \leq X \leq 0.67) = 0.50$
- $P(-1.96 \leq X \leq 1.96) = 0.95$
- $P(-2.58 \leq X \leq 2.58) = 0.99$



9.22 Distribuciones normal y estándar

Para cualquier variable normalmente distribuida X, tal que

$$X \to N(\mu, \sigma^2)$$

su distribución $F(a) = P(X \leq a)$ se puede calcular a partir de

$$F(a) = \phi(\frac{a-\mu}{\sigma})$$

9.23 Distribución normal

Para calcular $P(a \leq X \leq b)$, usamos la propiedad de las distribuciones de probabilidad

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

estandaricemos
$= P(\frac{X-\mu}{\sigma} \leq \frac{a-\mu}{\sigma}) - P(\frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma})$
$= P(Z \leq \frac{b-\mu}{\sigma}) - P(Z \leq \frac{a-\mu}{\sigma})$
$= \phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})$
Entonces

$$F(b) - F(a) = \phi(\frac{b-\mu}{\sigma}) - \phi(\frac{a-\mu}{\sigma})$$

Las probabilidades de **cualquier variable normal** se pueden obtener de la **distribución estándar**, después de la estandarización (restar la media y dividir por la desviación estándar).

9.24 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	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,...	$\binom{n}{x}(1-p)^{n-x}p^x$	np	$np(1-p)$
Binomial negativo para eventos	# de eventos B en repeticiones de Bernoulli antes de r As	0,1,..	$\binom{x+r-1}{x}(1-p)^x p^r$	$\frac{r(1-p)}{p}$	$\frac{r(1-p)}{p^2}$
Hipergeométrico	#eventos A en una muestra n de la población N con K As	$\max(0, n+K-N)$, ... $\min(K, n)$	$\frac{1}{\binom{N}{n}}\binom{K}{x}\binom{N-K}{n-x}$	$n * \frac{N}{K}$	$n\frac{N}{K}(1-\frac{N}{K})\frac{N-n}{N-1}$

Modelo	X	rango de x	f(x)	E(X)	V(X)
Poisson	# de eventos A en un intervalo	0,1, ..	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ
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	$(-\infty, \infty)$	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2

9.25 Funciones R para modelos de probabilidad

Modelo	R
Uniforme (continuo)	dunif(x, a, b)
binomial	dbimon(x,n,p)
Binomial negativo para eventos	dnbinom(x,r,p)
Hipergeométrico	dhyper(x, K, N-K, n)
Poisson	dpois(x, lambda)
Exponencial	dexp(x, lambda)
Normal	dnomr(x, mu, sigma)

Chapter 10

Distribuciones de muestreo

10.1 Objetivo

Distribuciones para

- Media (promedio) muestral
- suma muestral
- Varianza muestral

10.2 Distribución normal

Cuando tenemos una variable aleatoria normal

$$X \to N(x; \mu, \sigma^2)$$

¿Cómo estimamos μ y σ^2?

- necesitamos tomar una **muestra aleatoria**
- necesitamos **estimar** cada parámetro

10.3 Ejemplo: Cuando no conocemos los parametros

Imaginemos que un cliente que nos pide a nuestra empresa metalúrgica que le vendamos 8 cables que pueden cargar hasta 96 Toneladas; eso es 12 Toneladas cada uno.

- Tenemos en **stock** un conjunto de cables que podrían hacer el trabajo.

¿Podemos utilizar los cables en stock o necesitaríamos producir unos nuevos?

10.4 Ejemplo

Tomamos una muestra de 8 experimentos aleatorios, cada uno de los cuales consiste en cargar un cable hasta que se rompa y anotamos la carga de rotura.

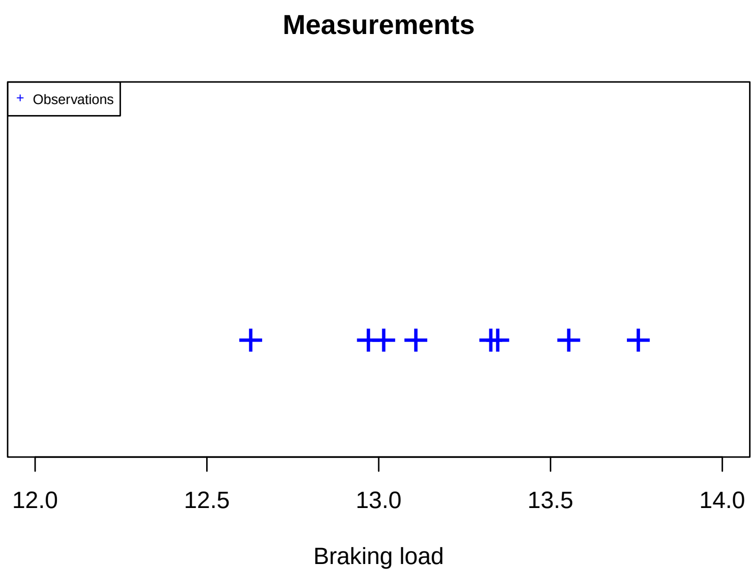
Estos son los resultados: La observación de una **muestra** de tamaño 8

```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```

- Ninguno se rompió a 12 Toneladas.

- Hubo uno que se rompió a 12.62747 Toneladas.

¿Nos arriesgamos y vendemos una muestra aleatoria 8 cables de nuestro inventario?



10.5 Muestra aleatoria

Una **muestra aleatoria** de tamaño n es la **repetición** de un experimento aleatorio n veces de forma **independiente**.

- Una muestra aleatoria es una **variable aleatoria** de 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 8 cables fue

```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```

10.6 Ejemplo

Nos gustaría calcular $P(X \leq 12)$.

Vamos a **suponer** que el punto de rotura se distribuye **normalmente**.

$$X \rightarrow N(x; \mu, \sigma^2)$$

- Para calcular $P(X \leq 12)$ necesitamos los parámetros μ y σ^2.
- ¿Cómo **estimamos** los parámetros usando la muestra observada?

10.7 Promedio o media muestral

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 8 cables tomó el valor

$$\bar{x}_{stock} = 13.21$$

10.8 Promedio como estimador

Este número puede usarse para **estimar** el parámetro desconocido μ porque:

- $E(\bar{X}) = E(X) = \mu$
- $V(\bar{X}) = \frac{V(X)}{n} = \frac{\sigma^2}{n}$

(dado que cada experimento aleatorio en la muestra es independiente)

como

- $n \to \infty$, $V(\bar{X}) \to 0$

entonces

- $\bar{x}$ **se concentra cada vez más cerca** de μ a medida que aumenta n.

Podemos tomar un valor de $\bar{x}$ como **estimación** para μ o

$$\bar{x} = \hat{\mu}$$

10.9 Varianza muestral

Definición

La **varianza de la muestra** 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 8 cables, S^2 tomó el valor

$$s^2_{stock} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0.1275608$$

El valor esperado de S^2 es

- $E(S^2) = V(X) = \sigma^2$ (insesgado)

y por lo tanto S^2 es

- un estimador de $V(X)$
- también se concentra alrededor de σ^2 porque como $n \to \infty$, $V(\bar{S^2}) \to 0$ (consistente)

Podemos tomar un valor de s^2 como estimación para σ^2 o

$$s^2 = \hat{\sigma}^2$$

10.10 Varianza muestral

S^2 tiene como objetivo estimar la dispersión de los resultados al rededor de μ (la varianza)

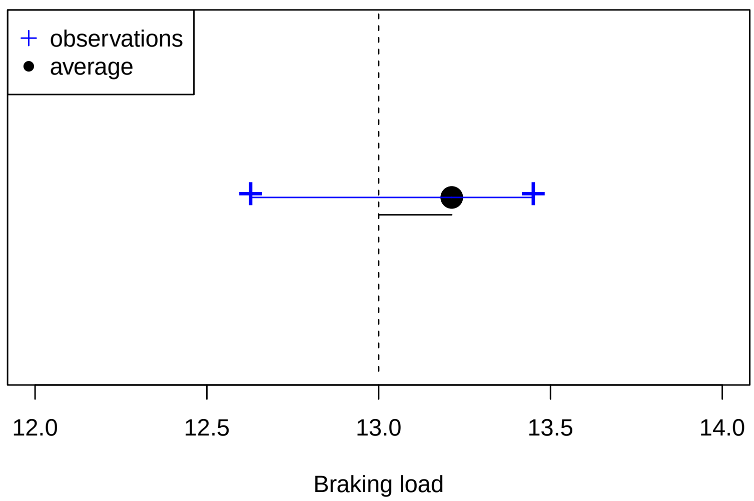
Si usamos $\bar{X}$ como estimador de μ, debemos corregir su dispersión (es decir, el error cuadrático medio de $\bar{X}$).

La corrección se logra dividiendo por $n-1$ y no por n en la definición de S^2

Para:

$S_n^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2$

$E(S_n^2) = \sigma^2 - \frac{\sigma^2}{n} \neq \sigma^2$ (decimos que S_n^2 es en estimador **sesgado** de)



10.11 Ajuste de un modelo

Ajustamos en un modelo cuando

- **estimamos** los parámetros del modelo

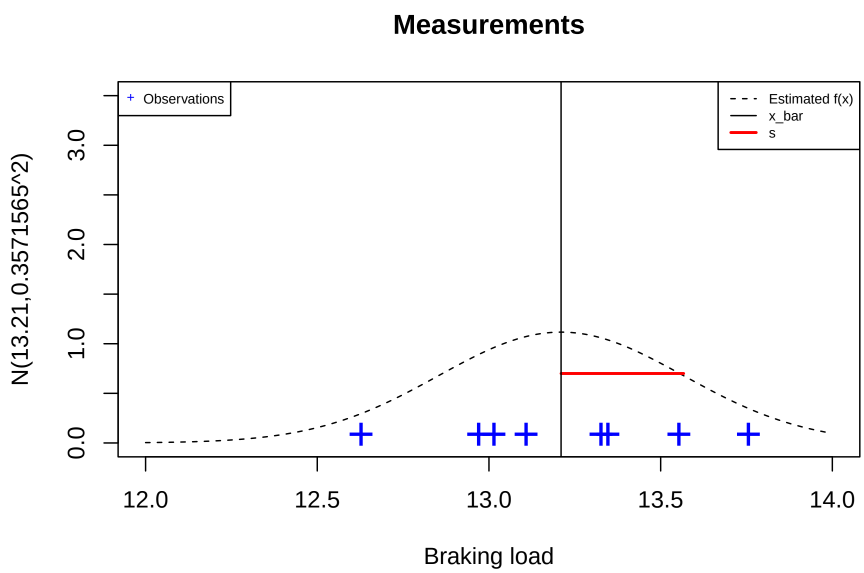
También decimos que **entrenamos** un modelo (aprendizaje automático)

Asumiendo que

$$X \rightarrow N(x; \mu, \sigma^2)$$

Como no conocemos los parámetros, **sustituimos** las estimaciones $\bar{x}$ y s^2 como los valores de μ y σ^2

$$X \rightarrow N(x; \mu = 13.21, \sigma^2 = 0.3571565^2)$$



10.12 Predicción

Predecimos el valor de un **resultado** cuando calculamos su **probabilidad**

¿Cuál es la probabilidad de que el cable se rompa a 12 Toneladas?

Si asumimos la variable aleatoria

$$X \rightarrow N(x; \mu, \sigma^2)$$

Sustituimos las estimaciones $\bar{x}$ y s^2 en la distribución de probabilidad

$$P(X \leq 12) = F_{normal}(12; \mu = 13.21, \sigma^2 = 0.1275608)$$

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 a 12 Toneladas.

10.13 Inferencia

Cuando tenemos una variable aleatoria normal

$$X \to N(x; \mu, \sigma^2)$$

Y conocemos μ y σ^2.

Podemos hacer inferencias sobre $\bar{X}$, es decir calcular probabilidades de la variable aleatoria $\bar{X}$.

Cuando hacemos **inferencias**, generalmente hacemos la pregunta:

¿Qué tan **seguros** estamos de que el valor del estimador **está cerca** del **parámetro verdadero**?

10.14 Ejemplo: Cuando sí conocemos los parametros

Imaginemos que nuestros cables están certificados para romper con una carga media de $\mu = 13$ Toneladas con varianza $\sigma^2 = 0.35^2$.

Tomamos una muestra aleatoria de 8 cables

```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```

¿Podemos afirmar que en realidad producimos cables más resistentes porque obtuvimos $\bar{x} = 13,21$ en esta muestra de 8 cables?

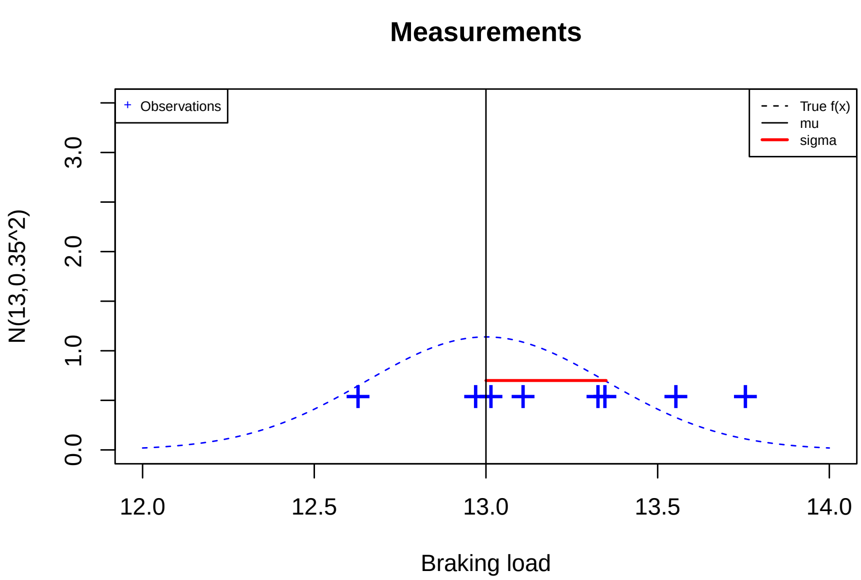
Necesitamos calcular probabilidades de $\bar{X}$.

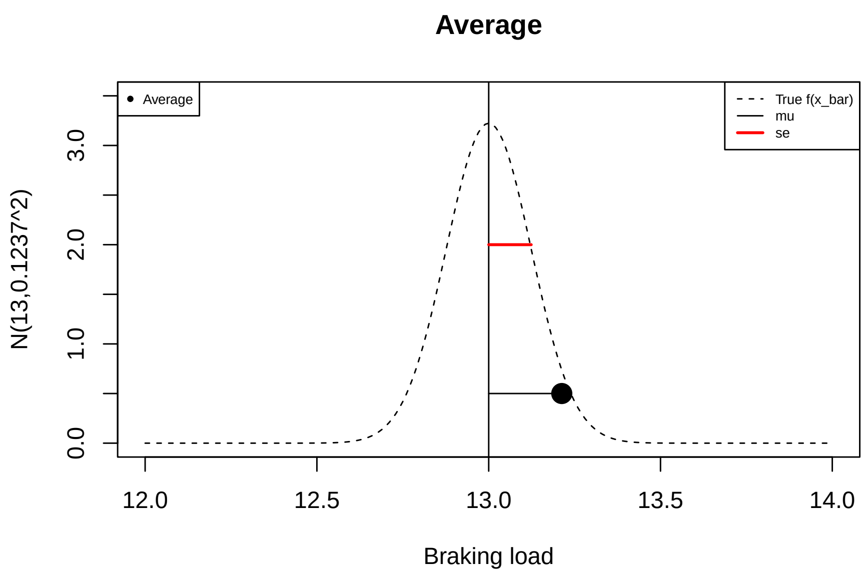
- Cuál es la probabilidad de que la distancia entre $\bar{X}$ y μ sea menos de $\bar{x}_{stock} - \mu = 0.21$?

$$P(-0.21 \leq \bar{X} - \mu \leq 0.21)$$

10.15 Desnidad para X y para $\bar{X}$

Cuando **sabemos** que los parámetros **verdaderos** son $\mu = 13$ y $\sigma = 0.35$ esto es lo que veríamos





10.16 Distribución media muestral

Teorema: Cuando X sigue una distribución normal $X \to N(\mu, \sigma^2)$

$\bar{X}$ es normal:

$$\bar{X} \to N(\mu, \frac{\sigma^2}{n})$$

Entonces, si **sabemos** μ y σ podemos calcular las **probabilidades de** $\bar{X}$ usando la distribución normal.

La media y la varianza de $\bar{X}$ son

- $E(\bar{X}) = \mu$
- $V(\bar{X}) = \frac{\sigma^2}{n}$

10.17 Inferencia del promedio

Ejemplo:

Si **sabemos** que la rotura de nuestros cables realmente se distribuye como

$$X \to N(\mu = 13, \sigma^2 = 0.35^2)$$

entonces

$$\bar{X} \to N(13, \frac{0.35^2}{8})$$

- $E(\bar{X}) = 13$
- $V(\bar{X}) = \frac{0.35^2}{8} = 0.01530169$

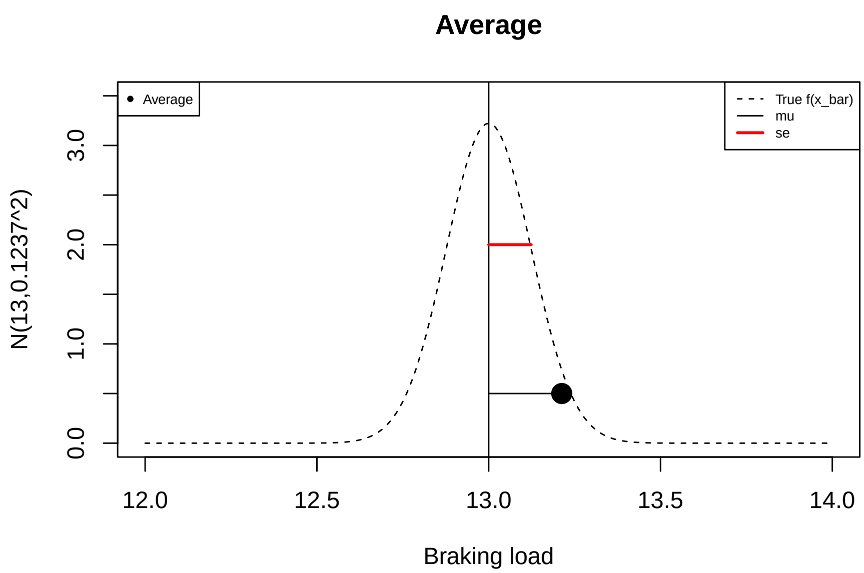
Nuestro **error observado** en la estimación de la media es la diferencia

$$\bar{x}_{stock} - \mu = 13.21 - 13 = 0.21$$

Nos preguntamos: ¿Es este un error **típico**?

10.18 Densidad para $\bar{X}$

Si **supiéramos** que los parámetros **verdaderos** son $\mu = 13$ y $\sigma = 0.35$ este es el error que veríamos



10.18.1 Probabilidades de $\bar{X}$

Si **sabemos** que la carga de rotura de nuestros cables **realmente** se distribuye como

$$\bar{X} \to N(\mu = 13, \frac{\sigma^2}{n} = 0.1237^2)$$

¿Cuál es la probabilidad de observar un **error de estimación** de μ (distancia entre $\bar{X}$ y μ) menor a 0.21?

Queremos calcular

$$P(-0.21 \leq \bar{X} - 13 \leq 0.21) = P(12.79 \leq \bar{X} \leq 13.21)$$

$= F_{normal}(13, 21; \mu, se^2) - F_{normal}(12, 79; \mu, se^2)$

En R podemos calcularlo como:

pnorm(13.21, 13, 0.1237)-pnorm(12.79, 13, 0.1237)=0.9104.

$91, 0\%$ de los errores son inferiores a $0, 21$, por lo que el error **observado** no parece demasiado típico (solo 9% de los errores son superiores).

Tal vez tengamos cables más fuertes de lo que pensábamos.

10.18.2 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 de la muestra** es la **estadística**:

$$Y = n\bar{X} = \sum_{i=1}^{n} X_i$$

Teorema: Si $X \to N(\mu, \sigma^2)$ entonces

$$Y \to N(n\mu, n\sigma^2)$$

Con media y varianza:

- $E(Y) = n\mu$
- $V(Y) = n\sigma^2$

10.18.3 Inferencia sobre la suma muestral

Si **sabemos** que nuestros cables

$$X \to N(\mu = 13, \sigma^2 = 0.35^2)$$

entonces

$$Y \to N(n\mu = 104, n\sigma^2 = 8 \times 0.35^2)$$

- $E(Y) = 104$
- $V(Y) = 8 \times 0.35^2 = 0.98$

Para nuestra muestra de 8 cables, observamos

- $y_{stock} = 105.7014$

y, por tanto, el **error observado** en la estimación de la media de la **verdadera** carga de rotura total ($n\mu$) de 8 cables fue

- $y_{stock} - n\mu = 1.7014$

¿Es este un error **típico**?

10.18.4 Probabilidades de la suma muestral: Propagación del error

¿Cuál es la probabilidad de observar una diferencia $Y - E(Y)$ menor que 1.7014?

Queremos calcular la probabilidad

$$P(-1.7014 \leq \bar{Y} - 104 \leq 1.7014) = P(102.2986 \leq Y \leq 105.7014)$$

$= F_{normal}(105.7014; n\mu, n\sigma^2) - F_{normal}(102.2986; n\mu, n\sigma^2)$

En R podemos calcularlo como:

pnorm(105.7014, 104, sqrt(0.98)) - pnorm(102.2986, 104, sqrt(0.98))=0.914.

$91,4\%$ de las veces tenemos sumas de cargas que son menores que $1,7014$.

Esta proporción es mayor que la de cables individuales.

10.19 Inferencia en la varianza muestral

Consideremos un proceso de control de calidad que requiera que los cables se produzcan cerca del valor especificado μ.

Si una muestra de cables de 8 es 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 8 cables sea mayor que los $0,3$ requeridos?

10.20 Probabilidades de la varianza muestral

Teorema: Cuando X sigue una distribución normal

$$X \rightarrow N(\mu, \sigma^2)$$

La **estadística**:

$$W = \frac{(n-1)S^2}{\sigma^2} \to \chi^2(n-1)$$

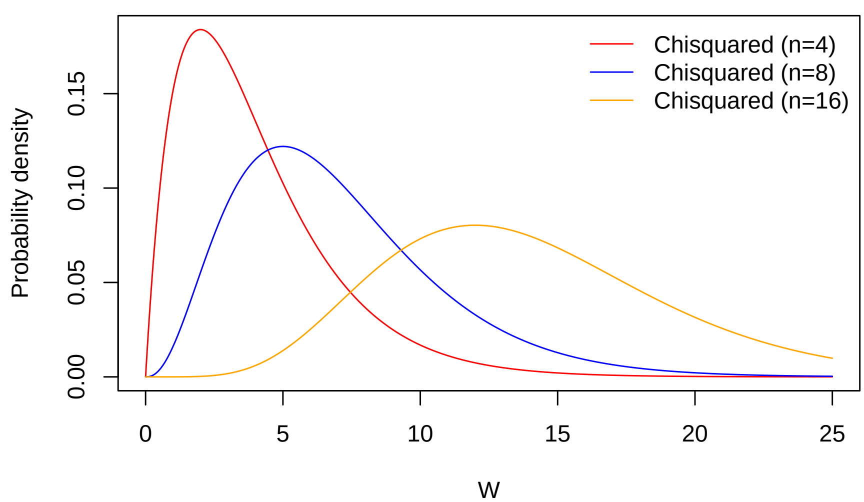
tiene una distribución χ^2 (chi-cuadrado) con $df = n-1$ grados de libertad dada por

$$f(w) = C_n w^{\frac{n-3}{2}} e^{-\frac{w}{2}}$$

dónde:

- $C_n = \frac{1}{2^{(n-1)/2}\sqrt{\pi(n-1)}}$ asegura $\int_{-\infty}^{\infty} f(t)dt = 1$
- $\Gamma(x)$ es el factorial de Euler para números reales
- Si **sabemos** los valores verdaderos de μ y σ podemos calcular las probabilidades de S^2 usando la distribución χ^2 para W.

10.21 χ^2-estadística



10.22 χ^2-estadística

Si **sabemos** que nuestros cables realmente se distribuyen como

$$X \rightarrow N(\mu = 13, \sigma^2 = 0.35^2)$$

entonces podemos calcular

$$P(S^2 > 0.2) = P(\frac{(n-1)S^2}{\sigma^2} > \frac{(n-1)0.3}{\sigma^2})$$

$= P(W > \frac{(n-1)0.3}{\sigma^2})$

$= 1 - P(W \leq \frac{(n-1)0.3}{\sigma^2}) = 1 - P(W \leq \frac{(8-1)0.3}{0.1225})$

$= 1 - F_{\chi^2, df=7}(17.14286) = 0.016$

En R 1-pchisq(17.14286, df=7)=0.016

Solo hay una probabilidad de 1% de obtener un valor superior a $s^2 = 0,3$.

- $s^2 > 0.3$ parece ser un buen criterio para detener la producción y revisar el proceso.
- nuestro valor observado fue $s^2_{stock} = 0.1275608$
- la muestra no está demasiado dispersa y creemos que la producción está bajo control.

Chapter 11

teorema del límite central

11.1 objetivo

- Margen de errores
- Teorema del límite central -t-estadística

11.2 Margen de error

Al decidir si un **error observado** es grande o no, generalmente lo comparamos con una tolerancia **predefinida**.

- El **margen de error** al nivel de 5% es la distancia m tal que la distribución de $\bar{X}$ captura 95% de las estimaciones:

$$P(-m \leq \bar{X} - \mu \leq m) = P(\mu - m \leq \bar{X} \leq \mu + m) = 0.95$$

- o que 95% de los valores de $\bar{X}$ están a una distancia m de μ

11.3 Margen de error

Sigamos con el ejemplo de la carga de rotura.

para la muestra de 8 cables

```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```

el **error observado** es la diferencia

$$\bar{x}_{stock} - \mu = 13.21 - 13 = 0.21$$

¿Está este valor por debajo del margen de error de 5%?

11.4 Estadística Z

Si **sabemos** que nuestros cables realmente distribuyen como

$$X \to N(\mu = 13, \sigma^2 = 0.35^2)$$

entonces,

$$\bar{X} \to N(\mu, \frac{\sigma^2}{n})$$

y el margen de error de 5% para el promedio en nuestra muestra de 8 cables se puede calcular a partir de la **estadística estandarizada**:

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \to N(0, 1)$$

11.5 Estadística Z

para calcular el margen de error m al nivel de 5% estandarizamos (restamos μ y dividimos por $\sigma/\sqrt{n}$)

$$P(\mu - m \leq \bar{X} \leq \mu + m) = P(-\tfrac{m}{\sigma/\sqrt{n}} \leq \tfrac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \leq \tfrac{m}{\sigma/\sqrt{n}})$$

$$= P(-\frac{m}{\sigma/\sqrt{n}} \leq Z \leq \frac{m}{\sigma/\sqrt{n}}) = 0,95$$

(compararlo con el gráfico) tenemos

$$m = z_{0,025}\frac{\sigma}{\sqrt{n}} = 1,96 \times se = 1,96\frac{0,35}{\sqrt{8}} = 0.24$$

donde $z_{0.025} = 1.96$ es el valor Z que deja 2.5% a cada lado de la densidad normal estándar (0.025-cuantil)

Nuestro error observado 0.21

- es menor que el margen de error 0.24 en el nivel 5%.
- y, por tanto, se espera dentro de los 95% de errores.

Si una observación de $\bar{x}$ distancia más de ~ 2 veces el se, decimos que el error es **inusualmente** grande.

11.6 Estadística Z

Definición

Para una variable aleatoria normal X

$$X \to N(\mu, \sigma^2)$$

con **conocido** σ

La estadística Z:

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}}$$

es una variable aleatoria estándar cuyos cuantiles $1 - \alpha/2$ ($z_{1-\alpha/2}$) dan una medida del margen de error de $\bar{X}$ en $1 - \alpha$ nivel

$$m = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Una situación común:

- ¿Qué sucede cuando X no se distribuye normalmente?

11.7 Teorema del límite central

Para cualquier variable aleatoria X con distribución **desconocida** (de cualquier tipo)

$$X \rightarrow f(x;\theta)$$

la estadística estandarizada

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}}$$

se aproxima a una distribución estándar

$$Z \rightarrow_d N(0,1)$$

cuando $n \rightarrow \infty$

Por lo tanto:

- Podemos calcular probabilidades para $\bar{X}$ si n es grande, usando la distribución normal:

$$\bar{X} \sim_{aprox} N(E(X), \frac{V(X)}{n})$$

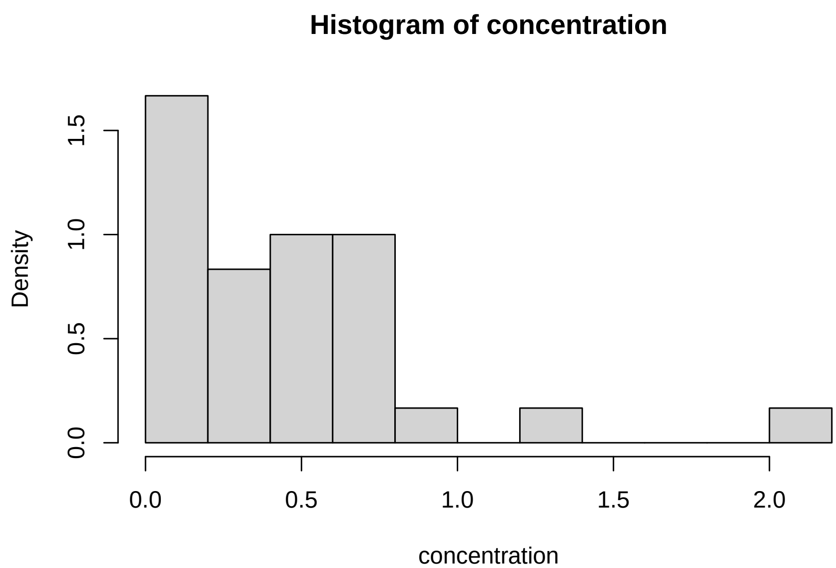
11.8 Teorema del límite central

Ejemplo:

Consideremos un experimento en el que medimos la concentration en sangre de un fármaco después de 10 horas de administración en pacientes de 30. Obtenemos los siguientes resultados:

```
##  [1] 0.42172863 0.28830514 0.66452743 0.01578868 0.02810549 0.15825061
##  [7] 0.15711365 0.07263340 1.36311823 0.01457672 0.50241503 0.24010736
## [13] 0.14050681 0.18855892 0.09414202 0.42489306 0.78160177 0.23938021
## [19] 0.29546742 2.02050586 0.42157487 0.48293561 0.74263790 0.67402224
## [25] 0.58426449 0.80292617 0.74837143 0.78532627 0.01588387 0.29892485
```

- el promedio es $\bar{x} = 0.56$
- el histograma de los resultados es:



11.9 Teorema del límite central

Si **sabemos** que los niveles siguen una distribución exponencial

$$X \to exp(\lambda = 2)$$

La media y la varianza son:

- $E(X) = \frac{1}{\lambda} = 0.5$
- $V(X) = \frac{1}{\lambda^2} = 0.25$

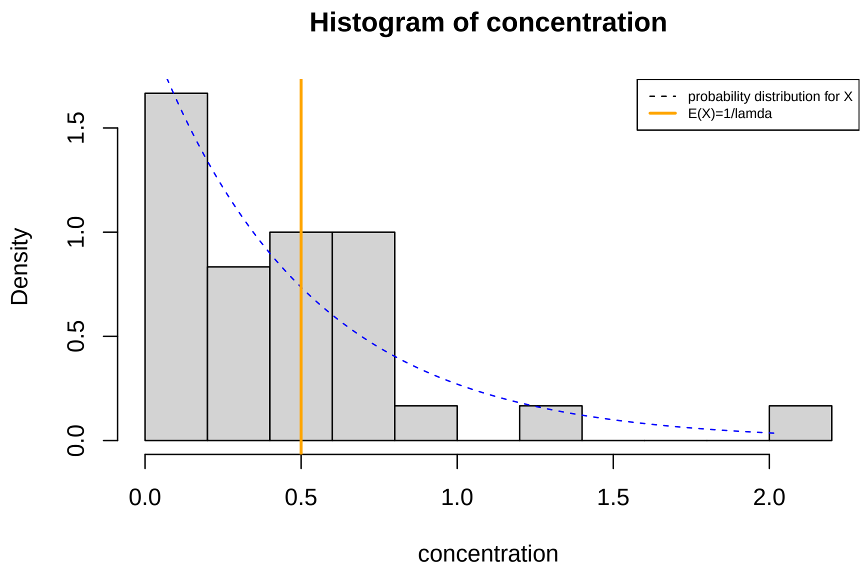
Por lo tanto la media y la varianza de $\bar{X}$ son:

- $E(\bar{X}) = \frac{1}{\lambda} = 0.5$
- $V(\bar{X}) = \frac{V(X)}{n} = \frac{1}{n\lambda^2} = 0.25/30$

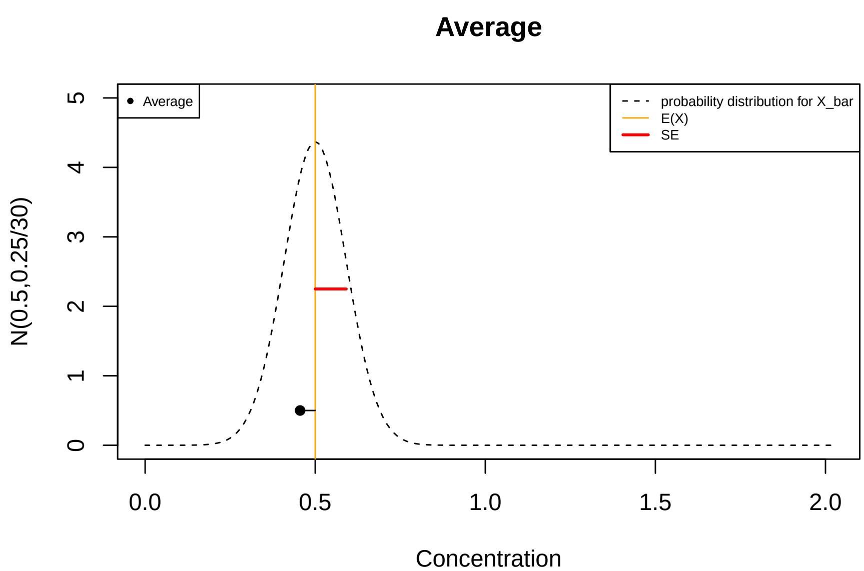
Como $n \geq 30$

$$Z = \frac{\bar{X} - \lambda}{\sqrt{\frac{1}{n\lambda^2}}}$$

es una variable normal estándar y: $\bar{X} \sim_{aprox} N(\lambda, \frac{1}{n\lambda^2})$



Histogram of concentration
probability distribution for X
E(X)=1/lamda
Density
0.0
0.5
1.0
1.5
2.0
concentration



Average
Average
probability distribution for X_bar
E(X)
SE
N(0.5,0.25/30)
0
1
2
3
4
5
0.0
0.5
1.0
1.5
2.0
Concentration

11.10 Margen de error con TCL

Ya que

$$\bar{X} \sim_{aprox} N(E(X), \frac{V(X)}{n})$$

El margen de error en el nivel de 5%

$$P(E(X) - m \leq \bar{X} \leq E(X) + m) = 0.95$$

se puede calcular de nuevo con la distribución estándar

$$m = z_{0.025}\sqrt{\frac{V(X)}{n}} = 1.96\sqrt{\frac{0.25}{30}} = 0.1789227$$

Observamos $\bar{x} = 0.5638725$ por lo tanto el **error observado** en la estimación es

$$\bar{x} - E(X) = 0.5638725 - 0.5 = 0.063$$

que está dentro del margen de error.

El error que observamos es común y dentro del 95% de errores.

11.11 Suma de muestra y TCL

Para cualquier variable aleatoria X con distribución **desconocida** (cualquier tipo de)

$$X \to f(x; \theta)$$

la estadística estandarizada

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{n\bar{X} - nE(\bar{X})}{\sqrt{nV(\bar{X})}}$$

se aproxima a una distribución estándar

$$Z \to_d N(0,1)$$

cuando $n \to \infty$

Por lo tanto:

- Podemos calcular probabilidades para la suma muestral $Y = n\bar{X}$ si n es grande, usando la distribución normal:

$$\bar{Y} \sim_{aprox} N(nE(X), nV(X))$$

11.12 Desconocido σ pero grande n

Para cualquier variable aleatoria X con distribución **desconocida** (cualquier tipo de)

$$X \to f(x; \theta)$$

con varianza **desconocida** $V(X)$, podemos estimar el error estándar ($se = \sqrt{V(X)/n}$) por la desviación estándar de la muestra

$$\hat{se} = \frac{s}{\sqrt{n}}$$

y escribe la estadística estandarizada

$$Z = \frac{\bar{X} - E(\bar{X})}{\frac{s}{\sqrt{n}}}$$

$$Z \to_d N(0, 1)$$

para recuperar el TCL cuando $n \to \infty$ (una buena aproximación es cuando $n > 30$)

11.13 Estadística T

Cuando

- σ es **desconocido**

y

- n es pequeño (no se puede aplicar TCL)

Sin embargo, si X es normal

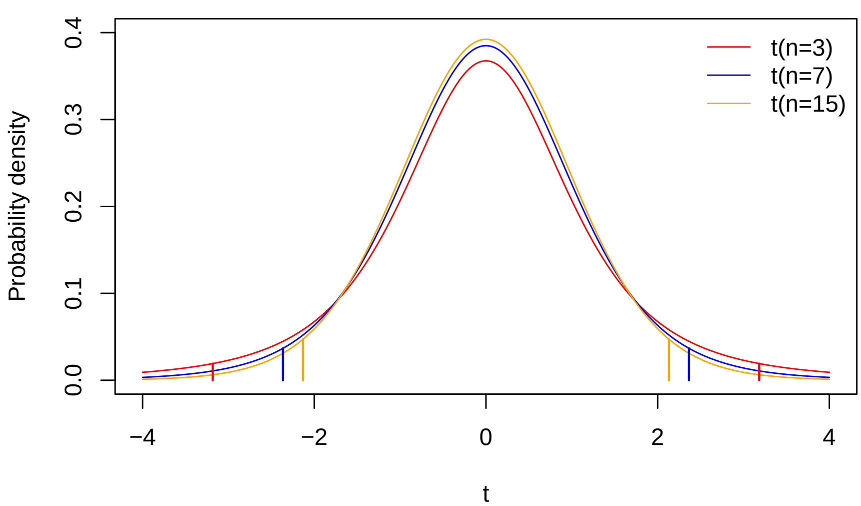
$$X \to N(\mu, \sigma^2)$$

entonces la estadística estandarizada

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

Sigue una distribución t con $n-1$ grados de libertad, y podemos calcular probabilidades en $\bar{X}$.

11.14 Estadística T



11.15 Estadística T

Para calcular el margen de error m a un nivel de 5% cuando n es pequeño, σ desconocido pero X normal

$$P(\mu - m \leq \bar{X} \leq \mu + m) = P(-\tfrac{m}{s/\sqrt{n}} \leq \tfrac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \leq \tfrac{m}{s/\sqrt{n}})$$
$$= P(-\frac{m}{s/\sqrt{n}} \leq T \leq \frac{m}{s/\sqrt{n}}) = 0,95$$

Usamos la distribución t

$$m = t_{0.025,n-1}\frac{s}{\sqrt{n}}$$

donde $t_{0.025,n-1}$ es el valor T que deja 2.5% a cada lado de la distribución t con $n-1$ grados de libertad (0.025-cuantil)

11.16 Ejemplo 1

Volviendo al ejemplo de la carga de rotura, calculamos el margen de error con **conocido** $\sigma^2 = 0.35^2$.

$$m = z_{0,025}\frac{\sigma}{\sqrt{n}} = 1,96 \times se = 1,96\frac{0,35}{\sqrt{8}} = 0,24$$

- En la mayoría de las aplicaciones **no conocemos** los parámetros

Si solo asumiéramos que la carga de rotura es una variable aleatoria normal

$$X \rightarrow N(\mu, \sigma^2)$$

con **desconocido** μ y σ^2 luego de los datos

- $s_{stock} = \sqrt{0.1275608}$

y el margen de error es

$$m = t_{0.025,n-1}\frac{s}{\sqrt{n}} = 2.36 \times \hat{se} = 2.36\frac{0.3571565}{\sqrt{8}} = 0.29$$

donde $t_{0.025,n-1} = 2.36$

en R es qt(1-0.025, 7)

Aumentó a partir del valor que obtuvimos con **conocido** σ

11.17 Ejemplo 2

También podemos preguntar por la probabilidad de observar un error en la estimación de μ (distancia entre $\bar{X}$ y μ) menor que el valor observado 0.21?

Entonces queremos calcular

$$P(-0.21 \leq \bar{X} - \mu \leq 0.21) = P(\frac{-0.21}{s/\sqrt{n}} \leq T \leq \frac{0.21}{s/\sqrt{n}})$$

$= P(\frac{-0.21}{0.3571565/\sqrt{8}} \leq T \leq \frac{0.21}{0.3571565/\sqrt{8}})$

$= F_{t,n-1}(0,21) - F_{t,n-1}(-0,21)$

En R podemos calcularlo como:

pt(1.663052, 7)-pt(-1.663052, 7)=0.859.

$85,9\%$ de los errores son inferiores a $0,21$, por lo que el error **observado** parece más típico que el 91% que obtenemos con $\sigma^2 = 0,35^2$.

Tenga en cuenta que en los cálculos hemos sustituido $\sigma = 0.35$ por una estimación más alta $s = 0.3571565$ **obtenida de los datos**.

Chapter 12

Máxima verosimilitud

12.1 Objetivo

- Máxima verosimilitud
- Método de los Momentos

12.2 Estadística

Definición

Dada una muestra aleatoria $X_1, ...X_n$, una **estadística** es cualquier función de valor real de las variables aleatorias que definen la muestra aleatoria: $f(X_1, ...X_n)$

- $\bar{X} = \frac{1}{N}\sum_{j=1..N} X_j$
- $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
- $\max X_1, X_n$

son estadísticas

12.3 Estimador

Definición

Un **estimador** es un estadístico Θ cuyos valores $\hat{\theta}$ son medidas de un parámetro θ de la distribución poblacional sobre la que se define la muestra: $E(\Theta) \sim \theta$

$X \hookrightarrow f(x;\theta)$

Entonces

- θ es un **parámetro** de la distribución de la población $f(x;\theta)$
- Θ es un **estimador** de θ: Una variable aleatoria
- $\hat{\theta}$ es la **estimación** de θ: Un valor realizado de Θ

12.4 Estimador

12.5 Ejemplos 1: promedio (media de la muestra)

Cuando $X \hookrightarrow N(\mu, \sigma^2)$

Para la media:

- μ es un **parámetro** de la distribución poblacional $N(\mu, \sigma^2)$
- $\bar{X}$ es un **estimador** de μ
- $\bar{x} = \hat{\mu} = 13.21\,Tons$ es el **estimado** de μ

12.6 Ejemplos 2: Variación de la muestra

Cuando $X \hookrightarrow N(\mu, \sigma^2)$

Para la varianza:

- σ^2 es un **parámetro** de la distribución de la población $N(\mu, \sigma^2)$
- S^2 es un **estimador** de σ^2
- $s^2 = \widehat{\sigma^2} = 0.127\,Tons^2$ es la **estimación** de σ^2

12.7 Sesgo

Un estimador es insesgado (no tiene sesgo) si $E(\Theta) = \theta$

- $\bar{X}$ es un estimador **insesgado** de μ porque $E(\bar{X}) = \mu$
- S^2 es un estimador **insesgado** de σ^2 porque $E(S^2) = \sigma^2$

12.8 Consistencia

Un estimador es consistente si $V(\Theta) \to 0$ cuando $n \to \infty$

- $\bar{X}$ es **consistente** porque $V(\bar{X}) = \frac{\sigma}{n} \to 0$ cuando $n \to \infty$.
- S^2 también es **coherente** (no lo mostraremos).

12.9 Máxima verosimilitud

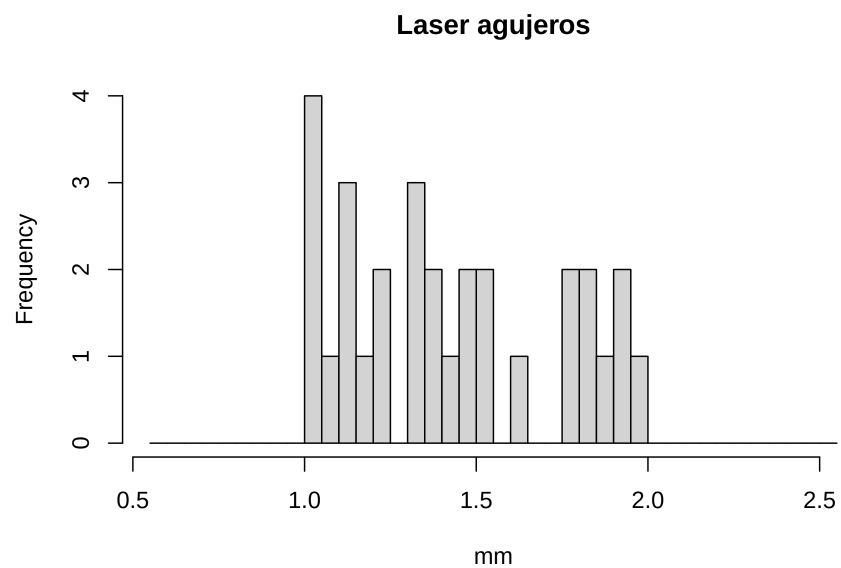
¿Cómo podemos **estimar** el parámetro de **cualquier** modelo paramétrico?

- Imaginemos que diseñamos un láser con un diámetro de 1 mm que queremos usar para aplicaciones clínicas.
- Queremos caracterizar el diámetro de un agujero en un tejido realizado con el láser
- y tomar una muestra aleatoria de 30 cortes realizados con el láser

```
##  [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
```

12.10 Ejemplo

con histograma



12.11 Densidad de probabilidad

Consideramos que

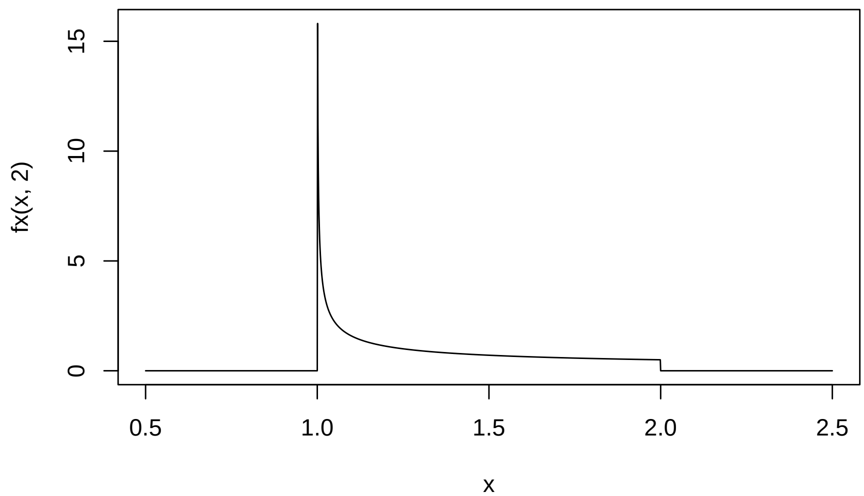
- se debe dar la máxima probabilidad a los diámetros de $x = 1mm$
- los diámetros deben disminuir exponencialmente en probabilidad a medida que aumentan de tamaño con un límite de $2mm$ más allá del cual la probabilidad es de 0.

Una distribución de densidad de probabilidad adecuada es

$$f(x) = \begin{cases} \frac{1}{\alpha}(x-1)^{\frac{1-\alpha}{\alpha}}, & \text{if } x \in (1,2) \\ 0, & x \notin (1,2) \end{cases}$$

Donde α es un parámetro.

12.12 Densidad de probabilidad



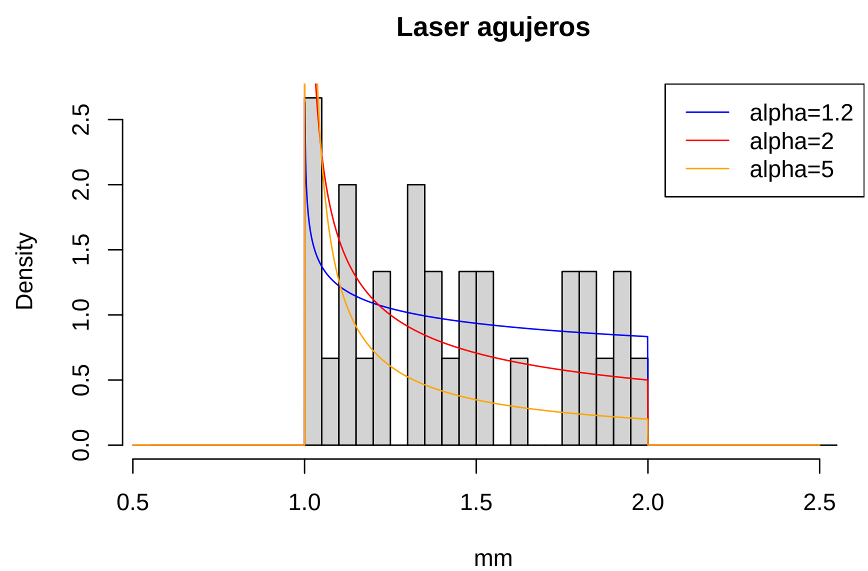
Si realizáramos una muestra de n: $X_1, ...X_n$

- La **máxima verosimilitud** es un método que **nos da el estimador** para α

$$\hat{\alpha}_{ml}$$

¿Cómo debemos combinar los datos para obtener el mejor valor de $\hat{\alpha}_{ml}$?

12.13 Ejemplo: Máxima probabilidad



12.14 Máxima verosimilitud

El objetivo es encontrar el valor del parámetro que **creemos** que puede representar **mejor** los datos.

Buscamos el parámetro que hace más **probable** la **observación** de la muestra.

Recordemos que:

- Las probabilidades se asignan a las observaciones.
- Las probabilidades no se asignan a parámetros (asignamos creencias, probabilidades).

Se supone que los parámetros no deben cambiar, son propiedades del sistema.

12.15 Método paso 1

1. Calculamos la probabilidad de haber observado la muestra n: $x_1, ...x_n$

$P(x_1, ...x_n) = P(X = x_1)P(X = x_2)...P(X = x_n)$

$$= f(x_1; \alpha)f(x_2; \alpha)...f(x_n; \alpha)$$

- Una vez observados los datos se **fijan**.
- La incógnita es α
- Esta probabilidad como función de α la llamamos **función de verosimilitud**

$$L(\alpha) = \Pi_{i=1..n} f(x_i; \alpha)$$

entonces en nuestro caso

$L(\alpha; x_1, ..x_n) = \frac{1}{\alpha^n}\Pi_{i=1..n}(x_i - 1)^{\frac{1-\alpha}{\alpha}} = \frac{1}{\alpha^n}\{(x_1 - 1)(x_2 - 1)...(x_n - 1)\}^{\frac{1-\alpha}{alpha}}$

12.16 Método paso 2

Queremos maximizar $L(\alpha)$ con respecto a α.

Como tenemos la multiplicación de muchos factores es más fácil maximizar el logaritmo de $L(\alpha)$

2. Tomemos el logaritmo para obetener el **Logaritmo de la verosimilitud**

$$\ln L(\alpha; x_1, ..x_n) = -n \ln(\alpha) + \frac{1 - \alpha}{\alpha}\Sigma_{i=1...n} \ln(x_i - 1)$$

12.17 Método paso 3

3. Maximizamos el log de la verosimilitud con respecto al parámetro

Por lo tanto,

- diferenciamos con respecto a α

$\frac{d \ln L(\alpha)}{d\alpha} = -\frac{n}{\alpha} - \frac{1}{\alpha^2}\Sigma_{i=1...n} \ln(x_i)$

- El máximo es donde la derivada es 0. Este máximo es el valor de nuestro estimador $\hat{\alpha}_{ml}$.

$\hat{\alpha}_{ml} = -\frac{1}{n}\Sigma_{i=1...n}\ln(x_i - 1)$

12.18 Método paso 3

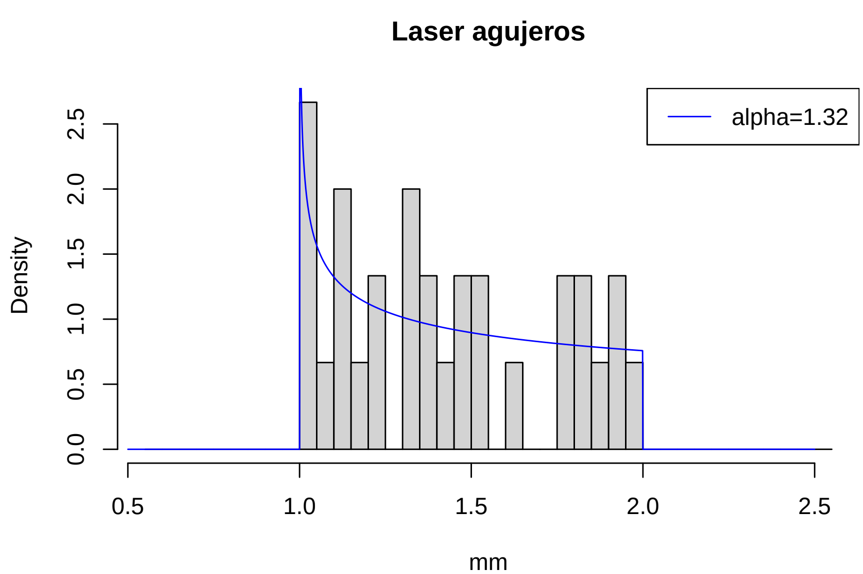
$\hat{\alpha}_{ml} = -\frac{1}{n}\Sigma_{i=1...n}\ln(x_i - 1)$

es la **estadística** que estima el parámetro.

En nuestro ejemplo calculamos:

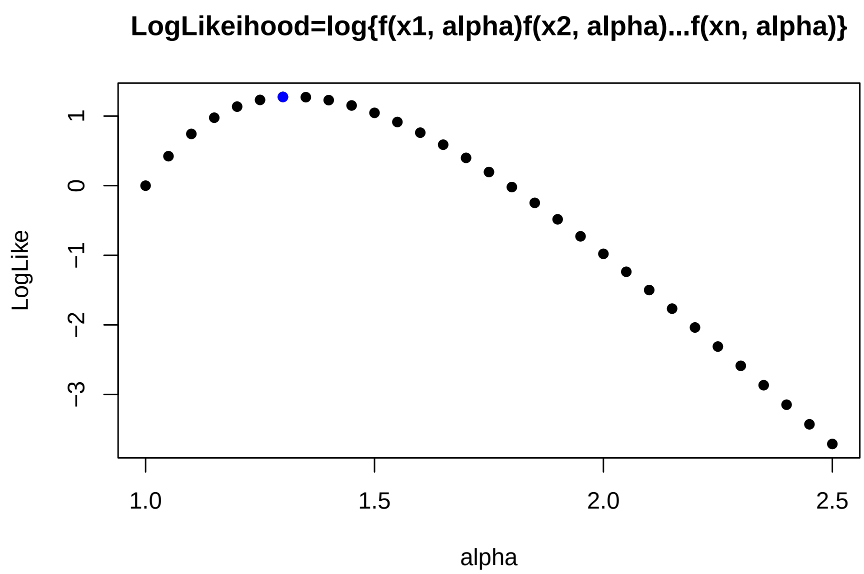
$\hat{\alpha}_{ml} = -\frac{1}{n}\{\ln(1.11 - 1) + \ln(1.64 - 1) + ... \ln(1.04 - 1)\} = 1.320$

12.19 Estimacion



12.20 Estimacion

- Este es el log de la verosimilitud de nuestros 30 cortes con láser.
- Si tomamos otra muestra esta función cambia y también su máximo.



12.21 Distribución normal

Imaginemos que tomamos una muestra de 8 cables para estimar la carga de rotura de cables y

- **supongamos** que
$$X \to N(\mu, \sigma^2)$$
.
- ¿Cuáles son los estimadores de μ y σ^2 de la muestra?

¿Qué valores de los parámetros describen mejor los datos?

12.22 Distribución normal

1. La función de verosimilitud, la probabilidad de haber observado $(x_1, ...x_n)$ es

$L(\mu, \sigma^2) = \Pi_{i=1..n} N(x_i; \mu, \sigma)$

$= (\frac{1}{\sigma\sqrt{2\pi}})^n e^{-\frac{1}{2\sigma^2}\sum_i (x_i-\mu)^2}$

2. Podemos tomar el logaritmo de L y calcular la **el log de la verosimilitud**

$\ln L(\mu, \sigma^2) = -n \ln(\sigma\sqrt{2\pi}) - \frac{1}{2\sigma^2}\Sigma_i (x_i - \mu)^2$

12.23 Distribución normal

Las estimaciones de μ, σ^2 son donde la probabilidad es máxima y dan la máxima probabilidad a para los datos.

3. diferenciamos con respecto a μ y σ^2 (tratándola como una variable por ejemplo $t = \sigma^2$)

- $\frac{d \ln L(\mu,\sigma^2)}{d\mu} = \frac{1}{\sigma^2}\sum_i (x_i - \mu)$
- $\frac{d \ln L(\mu,\sigma^2)}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_i (x_i - \mu)^2$

12.24 Distribución normal

Derivando e igualando a 0 encontramos los máximos

- $\frac{1}{\hat{\sigma}^2}\sum_i (x_i - \hat{\mu}) = 0$
- $-\frac{n}{2\hat{\sigma}^2} + \frac{1}{2\hat{\sigma}^4}\sum_i (x_i - \hat{\mu})^2 = 0$

resolviendo los parámetros encontramos

- $\hat{\mu}_{ml} = \frac{1}{n}\sum_i x_i = \bar{x}$ (el promedio)
- $\widehat{\sigma^2}_{ml} = \frac{1}{n}\sum_i (x_i - \bar{x})^2$ (la varianza de la muestra **sin corregir**)

El estimador de máxima verosimilitud de σ^2 es un estimador **sesgado** como: $E(\hat{\sigma}^2_{ml}) \neq \sigma^2$

12.25 Máxima verosimilitud: Historia

¿Cuál es la estadística que mejor representa la verdadera posición de Ceres?

12.26 Máxima verosimilitud: Historia

Gauss propuso que en un momento **dado**

- la posición **verdadera** de Ceres era la media μ
- las probabilidades alrededor de la media eran simétricas.

12.27 Máxima verosimilitud: Historia

Gauss descubrió que si el promedio ($\bar{x}$) es el valor **más verosimil** para la posición real de Ceres (μ), entonces la densidad de probabilidad de los errores es

$$\frac{h}{\sqrt{\pi}} e^{-h^2(y-\mu_Y)^2}$$

a la que llamamos la Gaussiana y que Pearson (1920) la bautizó como la curva normal.

Nota: Suponemos que la posición **verdadera** de Ceres existe y es μ.

- ¿Podemos decir lo mismo para altura de los hombres μ? (Galton)

12.28 Método de los Momentos

El método de máxima verosimilitud tiene como objetivo producir los estimadores de distribuciones de probabilidad a partir de datos.

- ¿Hay otra forma de producir esos estimadores? ¿serán iguales?

12.29 Método de los Momentos

Reescribamos la estimación $\hat{\mu} = \bar{x}$ para una variable normal en términos de los resultados de X

Por ejemplo:

$$\hat{\mu} = \frac{1}{n}\sum_i x_i = \sum_x x\frac{n_x}{n}$$

y recordemos que en el límite $n \to \infty$ la interpretación frecuentista requiere $\frac{n_x}{n} \to P(X = x)$ y por lo tanto en el límite

$$\hat{\mu} = \frac{1}{n}\sum_i x_i \to E(X) = \mu$$

12.30 Método de los Momentos

El método de los momentos dice que podemos tomar el valor **observado** del promedio $\bar{X}$ como estimador de $E(X) = \mu$

$$E(X) \sim \bar{x}$$

$\bar{X} = \frac{1}{N}\sum_i X_i$ se llama el primer **momento muestral**

Si $X \to f(x, \theta)$, el estimador del parámetro θ se obtiene entonces de la ecuación:

$$E(X) = \bar{x}$$

para el valor de $\hat{\theta}$ observado

Ejemplo: Si

$X \hookrightarrow exp(\lambda)$ entonces

1. Calculamos el valor esperado $E(X) = \frac{1}{\lambda}$
2. Formulamos la ecuación donde igualamos el valor esperado al primer momento muestral $\frac{1}{\lambda} = \bar{x} = \frac{1}{n}\sum_i x_i$
3. Resolvemos para el parámetro

$$\hat{\lambda} = \frac{1}{\bar{x}}$$

12.31 Método de los Momentos

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 a 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 al cabo de una hora)?

- Realizamos un experimento y obtenemos $x_1, ...x_n$ ¿cómo podemos estimar a a partir de los datos?

12.32 Método de los Momentos

1. Calculamos el valor esperado de la variable aleatoria

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

2. Formulamos la ecuación para el parámetro $\hat{a}$ igualando el valor esperado al primer momento muestral

$$\frac{\hat{a}+1}{2} = \bar{x}$$

3. resolvemos para $\hat{a}$

$$\hat{a} = 2\bar{x} - 1$$

Este es el estimador de la carga mínima que podemos observar.

12.33 Método de los Momentos

Ten en cuenta que tomar el mínimo de las medidas es claramente subóptimo.

El método nos dio una respuesta inteligente:

- podemos calcular $\bar{x}$ con precisión creciente dada por n
- Sabemos que ninguna medida supera $b = 1$
- Luego calcula la distancia entre $\bar{x}$ y b: $1 - \bar{x}$
- Restarlo de $\bar{x}$: $\bar{x} - (1 - \bar{x}) = 2\bar{x} - 1$

12.34 Método de los Momentos

El método dice que se puede encontrar una estimación para el parámetro θ de $f(x; \theta)$ a partir de la ecuación:

$$E(X) = \frac{1}{n} \sum_i x_i$$

para $\hat{\theta}$.

Si hay más parámetros, usamos los **momentos muestrales** más altos

- El segundo momento muestral es

$$E(X^2) = \frac{1}{n} \sum_i X_i^2$$

Una observación de este momento es
$\frac{1}{n} \sum_i x_i^2$.

El método dice que se puede encontrar una estimación para los parámetros θ_1 y θ_2 de $f(x; \theta_1, \theta_2)$ a partir de las ecuaciones:

a. $E(X) = \frac{1}{n} \sum_i x_i$

b. $E(X^2) = \frac{1}{n} \sum_i x_i^2$

Para $\hat{\theta}_1$ y $\hat{\theta}_2$

Podemos entonces incrementar el número de ecuaciones tanto como parametros queramos estimar.

12.35 Distribución normal

Si X se distribuye normalmente tenemos dos parámetros a estimar

$$X \to N(\mu, \sigma^2)$$

1. Calculamos la media y el valor esperado del segundo momento $E(X^2)$:

$E(X) = \mu$ y $E(X^2) = \sigma^2 - \mu^2$

2. Obtenemos las ecuaciones para los parámetros donde hacemos que el valor esperado sea igual al primer momento de la muestra, y el segundo momento sea igual al valor esperado del segundo momento

a. $\hat{\mu} = \frac{1}{n}\sum_i x_i$
b. $\hat{\sigma}^2 - \hat{\mu}^2 = \frac{1}{n}\sum_i x_i^2$

12.36 Distribución normal

3. Resolvemos los parámetros

La primera ecuación nos da el estimador de la media μ.

a. $\hat{\mu} = \frac{1}{n}\sum_i x_i$

De la segunda ecuación obtenemos

b. $\hat{\sigma}^2 = \frac{1}{n}\sum_i x_i^2 - \hat{\mu}^2$

que también se puede escribir como:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_i (x_i - \hat{\mu})^2$$

12.37 Método de los Momentos

¿Cuál es el estimador del parámetro α para el corte láser dado por el método de los momentos?

$$f(x;\alpha) = \begin{cases} \frac{1}{\alpha}(x-1)^{\frac{1-\alpha}{\alpha}}, & \text{if } x \in (1,2) \\ 0, & x \notin (1,2) \end{cases}$$

Donde α es un parámetro.

12.38 Método de los Momentos

El método dice que se puede encontrar una estimación para el parámetro α de $f(x;\alpha)$ a partir de la ecuación:

$$E(X) = \frac{1}{n}\sum_i x_i$$

para $\hat{\alpha}$

1. Calculamos el valor esperado $E(X)$

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

12.39 Método de los Momentos

Consideremos 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. Formulamos la ecuación para el parámetro $\hat{\alpha}$ igualando el valor esperado al primer momento muestral

$$\frac{1}{1+\hat{\alpha}} + 1 = \bar{x}$$

3. resolvemos para $\hat{\alpha}$

$$\hat{\alpha}_m = \frac{1}{\bar{x}-1} - 1$$

4. Calculamos el valor para nuestros datos

$\hat{\alpha}_m = 1.314$

12.40 Método de los Momentos

Ten en cuenta que este es un ejemplo para el cual las estimaciones por máxima verosimilitud y el método de momentos son diferentes

- $\hat{\alpha}_{ml} = -\frac{1}{n}\sum_{i=1}^{n} \ln(x_i - 1) = 1.320$
- $\hat{\alpha}_m = \hat{\alpha}_m = \frac{1}{\bar{x}-1} - 1 = 1.314$

Necesitamos estudios de simulación, donde sepamos el verdadero valor del parámetro α, para encontrar cuál de estas estadísticas tiene menos error cuadrático medio.

Nota: los datos de 30 perforaciones con láser se simularon con $\alpha = 2$, por lo que debemos preferir la estimación de máxima verosimilitud.

Para obtener mejores estimaciones de α necesitamos aumentar el tamaño de la muestra.

Chapter 13

Estimación intervalar

13.1 Objetivo

- Estimación intervalar para la media y la proporción
- Estimación intervalar para la varianza

13.2 Promedio o media muestral

Definición

La media muestral (o promedio) de una **muestra aleatoria** de tamaño n es

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Dado que cada experimento aleatorio es independiente, la media y la varianza de $\bar{X}$ son

- $E(\bar{X}) = E(X)$
- $V(\bar{X}) = \frac{V(X)}{n}$
- $se = \sqrt{V(\bar{X})}$ se conoce como el error estandard

$\bar{X}$ es por lo tanto

- un **estimador** de $E(X)$, es decir μ.
- una variable aleatoria

13.3 Inferencia en el promedio

Ejemplo:

Realizamos 8 experimentos aleatorios: cargamos un cable hasta que se rompe y registramos la carga de rotura. Estos son los resultados.

```
## [1] 13.34642 13.32620 13.01459 13.10811 12.96999 13.55309 13.75557 12.62747
```

Si **sabemos** que nuestros cables realmente se distribuyen como

$$X \to N(\mu = 13, \sigma^2 = 0.35^2)$$

entonces

$$\bar{X} \to N(13, \frac{0.35^2}{8})$$

- $E(\bar{X}) = 13$
- $V(\bar{X}) = \frac{0.35^2}{8} = 0.01530169$
- $se = \frac{0.35}{\sqrt{8}} = 0.1237$

entonces el **error observado** en la estimación es la diferencia

$$\bar{x}_{stock} - \mu = 13.21 - 13 = 0.21$$

13.4 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.

- El **margen de error** a nivel de 5% es la distancia m tal que la distribución de $\bar{X}$ captura 95% de las estimaciones:

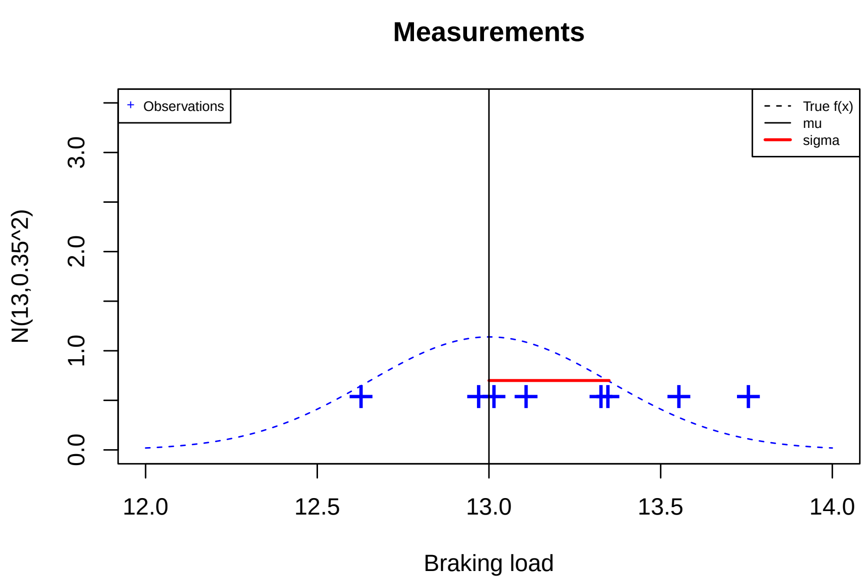
$$P(-m \leq \bar{X} - \mu \leq m) = P(\mu - m \leq \bar{X} \leq \mu + m) = 0.95$$

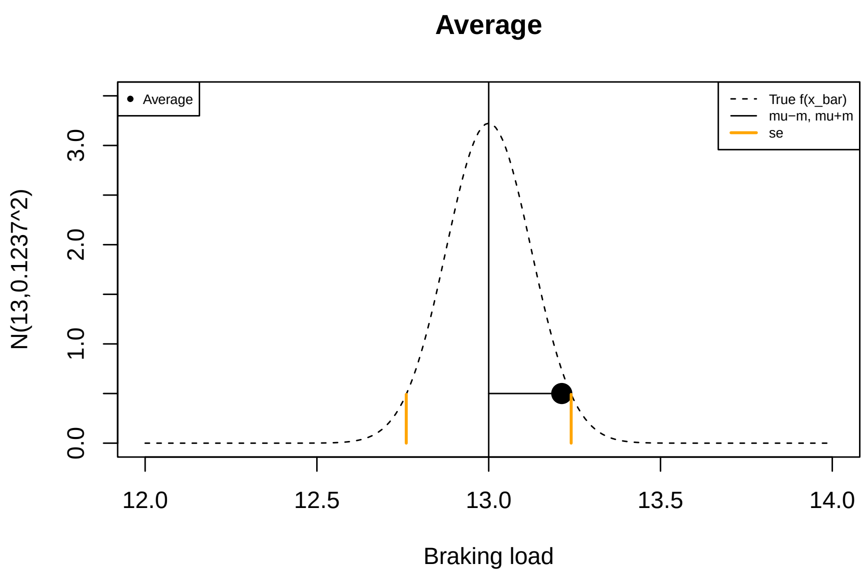
- o que 95% de los valores de $\bar{X}$ están a una distancia m de μ.

En nuestro ejemplo, asumimos que $\bar{X}$ se distribuye normalmente, entonces

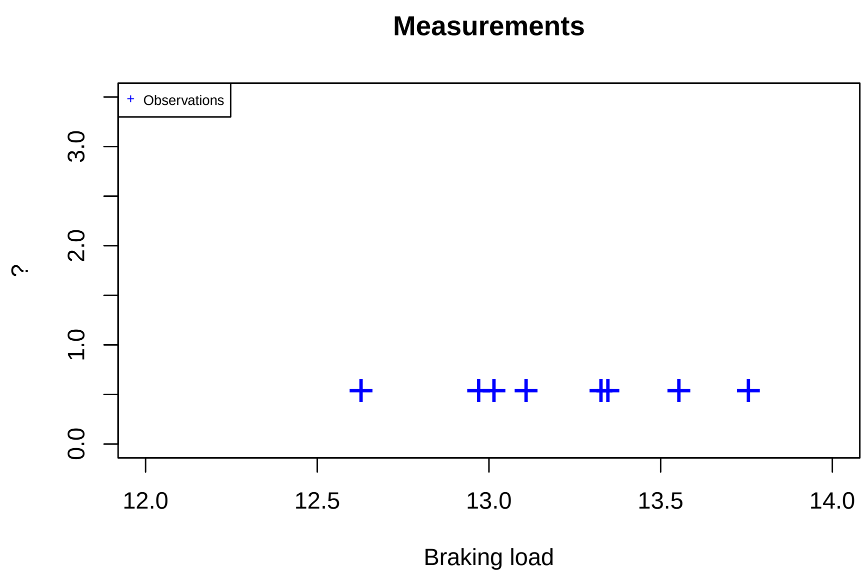
$$m = z_{0,025}\frac{\sigma}{\sqrt{n}} = 1,96 \times se = 1,96\frac{0,35}{\sqrt{8}} = 0,24$$

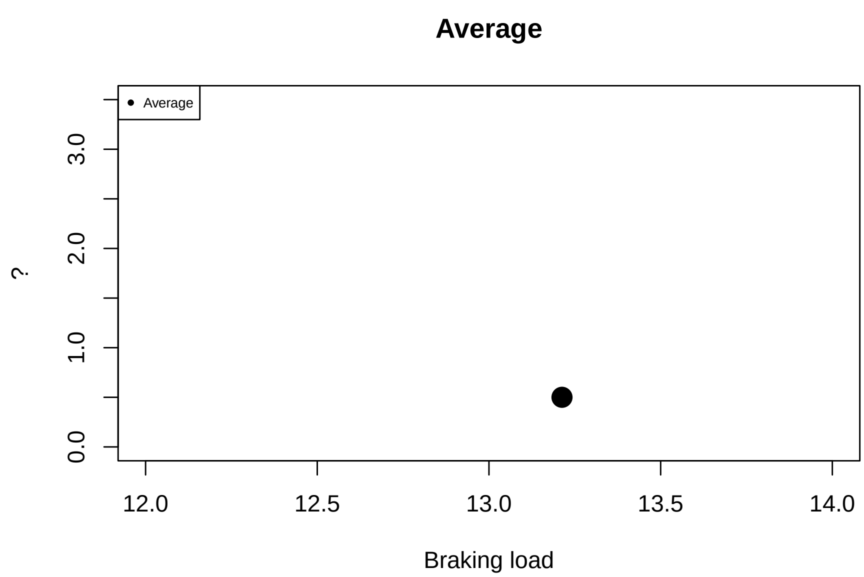
13.5 Densidad de probabilidad de X Vs densidad de probabilidad de $\bar{X}$




Average
Average
True f(x_bar)
mu−m, mu+m
se
N(13,0.1237^2)
0.0
1.0
2.0
3.0
12.0
12.5
13.0
13.5
14.0
Braking load

13.6 Vida real





13.7 Estimación intervalar

De la ecuación del margen de error:

$$P(-m \leq \bar{X} - \mu \leq m) = 0.95$$

resolvamos para μ (la incógnita real)

$$P(\bar{X} - m \leq \mu \leq \bar{X} + m) = 0.95$$

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 **variable aleatoria** y tiene por definición una probabilidad de 0.95 de contener μ.

13.8 Estimación intervalar

Cuando realizamos n-experimentos aleatorios (muestra de tamaño n) podemos calcular m **si**

- X es normal
- conocemos σ^2.

El intervalo que obtenemos del experimento es (en minúsculas)

$$(l, u) = (\bar{x} - m, \bar{x} + m)$$

- este intervalo contiene o no el parámetro μ: **nunca lo sabremos**!
- Decimos que tenemos una confianza de 95% que el intervalo (l, u) capturará el verdadero parámetro desconocido μ. Piensa en comprar un billete de lotería del que no sabes el resultado.

13.9 Estimación intervalar

En nuestro ejemplo, asumimos que $\bar{X}$ se distribuye normalmente, entonces

$$m = z_{0.025}\frac{\sigma}{\sqrt{n}}$$

y el intervalo de confianza de 95% es

$$(l, u) = (\bar{x} - z_{0.025}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0.025}\frac{\sigma}{\sqrt{n}}) = (12.97, 13.45)$$

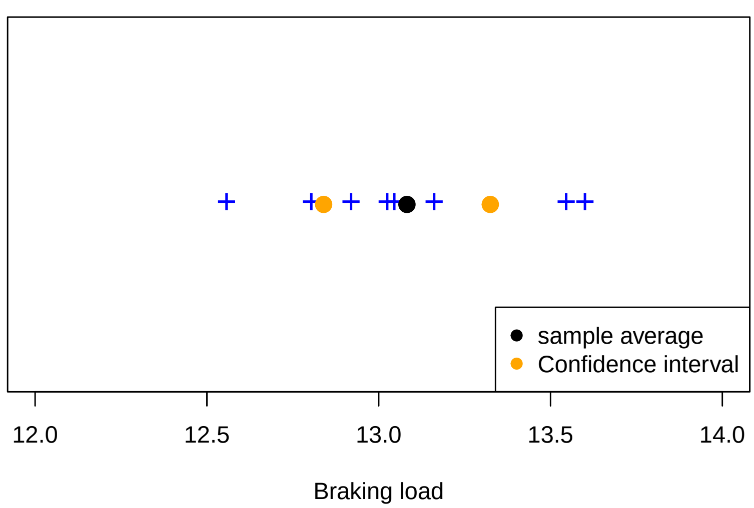
o

$$\hat{\mu} = 13.21 \pm 0.24$$

También significa que, en la estimación, confiamos en las unidades pero no tanto en los lugares decimales.

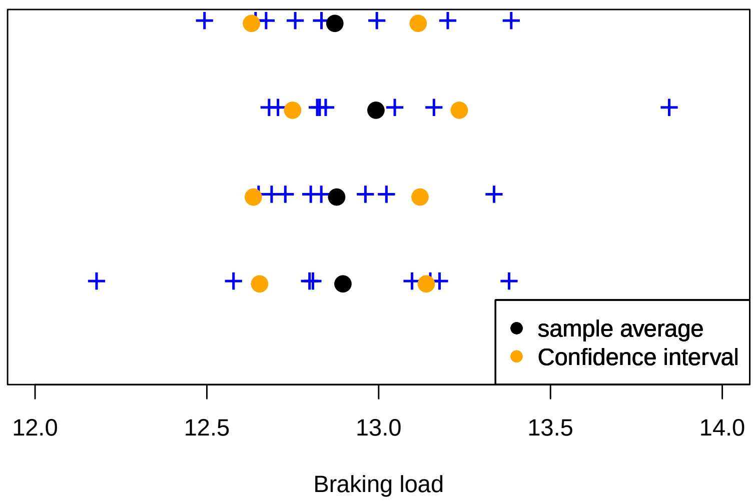
13.10 Estimación intervalar

Para una muestra de 8 observaciones, tenemos una estimación de la media y un intervalo de confianza



13.11 Estimación intervalar

Cada vez que obtenemos una nueva muestra, las estimaciones cambian. Si realizamos 100 muestras de tanaño n entonces 95 de los intervalos de confianza contendrán μ (¡no sabemos cuáles!)



13.12 Estimación de intervalos

Podemos cambiar nuestra confianza de 95% a 99%

- Habíamos dejado fuera $\alpha = 0.05$ de probabilidad, 0.025 en cada lado.
- Ahora, podemos dejar fuera $\alpha = 0.01$ probabilidad, 0.005 en cada lado.

Por lo tanto, el intervalo de confianza de 99% es

$(l, u) = (\bar{x} - z_{0,005}\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{0,005}\frac{\sigma}{\sqrt{n}})$

$$= (\bar{x} - 2,58\frac{\sigma}{\sqrt{n}}, \bar{x} + 2,58\frac{\sigma}{\sqrt{n}})$$

0

$$\hat{\mu} = \bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}}$$

Si queremos tener más confianza, ¡necesitamos intervalos de confianza más grandes!

Para nuestros cables:

$$\hat{\mu} = 13.21 \pm 0.31$$

13.13 Interval estimation

13.14 Estimación intervalar

Un material metálico se impacta para medir la energía requerida para cortarlo a una temperatura dada.

- Se cortaron diez probetas de acero A238 a 60ºC a 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 sabemos que la energía del impacto se distribuye normalmente con $\sigma = 1J$ ¿cuál es el IC de 95% para la media de estos datos?

13.15 Ejemplo

Sabemos

- $x_i = \{64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, 64.3\}$
- $X \to N(\mu, \sigma^2)$
- $\sigma = 1J$
- $\alpha = 0.05$

El intervalo de confianza es entonces

$$CI = (\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{x} + 1.96\frac{\sigma}{\sqrt{n}})$$

$$= (64.46 - 1.96\frac{1}{\sqrt{10}}, 64.46 + 1.96\frac{1}{\sqrt{10}}) = (63.84, 65.08)$$

or

$$\hat{\mu} = 64.46 \pm 0.61$$

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 **no conocemos** σ^2?

13.16 Estadística T

Cuando

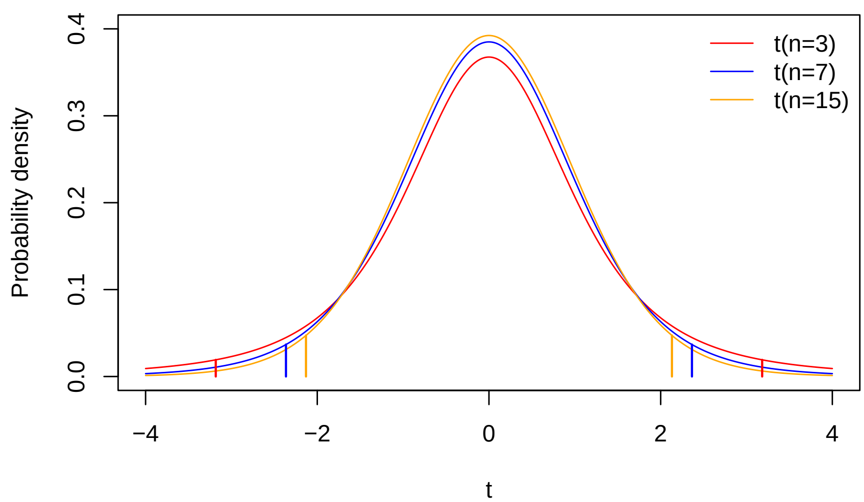
- X es normal, y
- σ es **desconocido**

entonces la estadística estandarizada

$$T = \frac{\bar{X} - \mu}{\frac{S}{\sqrt{n}}}$$

Sigue una distribución t con $n-1$ grados de libertad, y podemos calcular probabilidades para $\bar{X}$.

13.17 Estadística T



13.18 Estadística T

Para calcular el margen de error m al 95% cuando

- X normal y
- σ^2 es desconocido

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 al lado derecho de la distribución t con $n-1$ grados de libertad (0.025-cuantil)

el intervalo de confianza al 95% es entonces

$$(l, u) = (\bar{x} - t_{0.025,n-1}\frac{s}{\sqrt{n}}, \bar{x} + t_{0.025,n-1}\frac{s}{\sqrt{n}})$$

en R: $t_{0.025,n-1}$=qt(1-0.025, n-1)

13.19 Ejemplo

Un material metálico se impacta para medir la energía requerida para cortarlo a una temperatura dada.

- Se cortaron diez probetas de acero A238 a 60ºC a 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 sabemos que la energía del impacto se distribuye normalmente pero **no sabemos** la varianza, ¿cuál es el IC de 95% para la media de estos datos?

13.20 Ejemplo

- $\bar{x} = 64.46$
- $s = 0.227$
- $\alpha = 0.05$
- $t_{0.025,9} = 2.26$ obtained from $P(T \leq t_{0.025,9}) = 0.975$; qt(1-0.025, 9)

El intervalo de confianza es entonces

$CI = (\bar{x} - t_{0.025,9}\frac{s}{\sqrt{n}}, \bar{x} + t_{0.025,9}\frac{s}{\sqrt{n}})$

$$= (64.46 - 2.26\frac{0.227}{\sqrt{10}}, 64.46 + 2.26\frac{0.227}{\sqrt{10}})$$
$$= (64.29, 64.62)$$

habíamos visto un intervalo mas grande $CI = (63.84, 65.08)$ cuando $\sigma = 1$. Los datos entonces sugieren que $\sigma < 1$.

R: t.test(c(64.1,64.7,64.5,64.6,64.5,64.3,64.6,64.8,64.2,64.3))

13.21 CI con CLT

Si

- **no sabemos** cómo se distribuye X, pero
- tomamos una muestra grande de $n \geq 30$

Entonces podemos usar el CLT para encontrar los intervalos de confianza.

El intervalo de confianza a 95% es entonces

$$(l, u) = (\bar{x} - z_{0.025}\frac{s}{\sqrt{n}}, \bar{x} + z_{0.025}\frac{s}{\sqrt{n}})$$

dado que $t_{0.025,n-1} \to z_{0.025}$ para $n \to \infty$, entonces también está bien usar la distribución T en este caso.

Nota: Esta es la razón por la que R solo implementa t.test y no z.test en las funciones base para calcular CI.

Ejemplo:

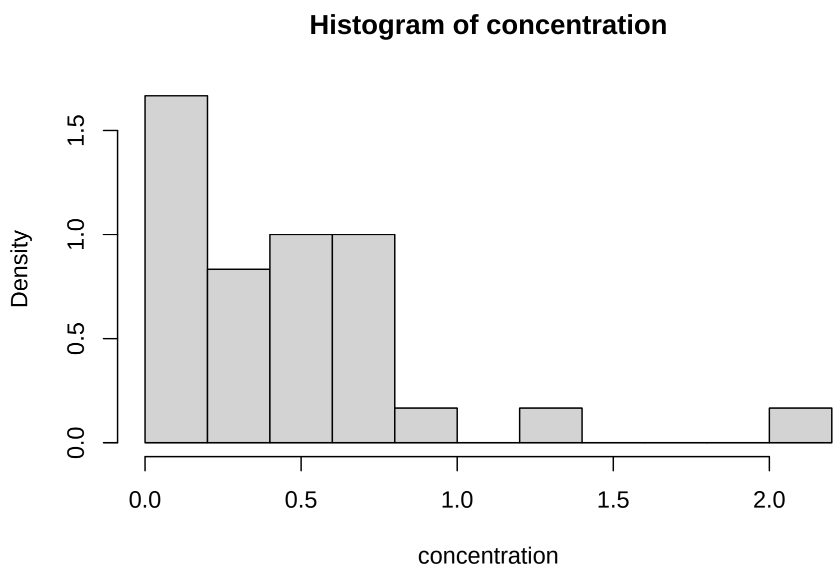
Consideremos un experimento en el que medimos la concentración en sangre de un fármaco después de 10 horas de administración en 30 pacientes. Obtenemos los siguientes resultados:

```
##  [1] 0.42172863 0.28830514 0.66452743 0.01578868 0.02810549 0.15825061
##  [7] 0.15711365 0.07263340 1.36311823 0.01457672 0.50241503 0.24010736
## [13] 0.14050681 0.18855892 0.09414202 0.42489306 0.78160177 0.23938021
## [19] 0.29546742 2.02050586 0.42157487 0.48293561 0.74263790 0.67402224
## [25] 0.58426449 0.80292617 0.74837143 0.78532627 0.01588387 0.29892485
```

- el promedio es $\bar{x} = 0.4556198$

- la desviación estándar es $s = 0.4335571$

- el histograma de los resultados es:



13.22 Teorema del límite central

Asumimos que $X \to exp(\lambda = 2)$

Con media y varianza:

- $E(X) = \frac{1}{\lambda}$
- $V(X) = \frac{1}{\lambda^2}$

¿Cuál es el CI para la media $E(X) = \mu$?

- Usamos un IC del 95% para estimarlo

Como $n \geq 30$ podemos usar el CLT

$$\bar{X} \sim_{aprox} N(\lambda, \frac{1}{n\lambda^2})$$

y el intervalo de confianza de 95% es entonces

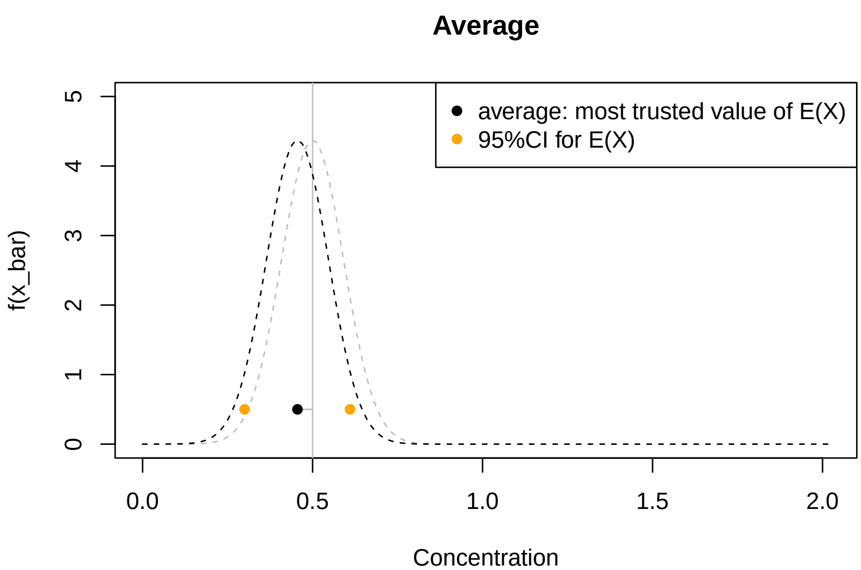
$$(l, u) = (\bar{x} - z_{0.025}\frac{s}{\sqrt{n}}, \bar{x} + z_{0.025}\frac{s}{\sqrt{n}})$$

$(l, u) = (0.4556198 - 1.96\frac{0.4335571}{\sqrt{30}}, 0.4556198 + 1.96\frac{0.4335571}{\sqrt{30}})$

$$= (0.300, 0.610)$$

o

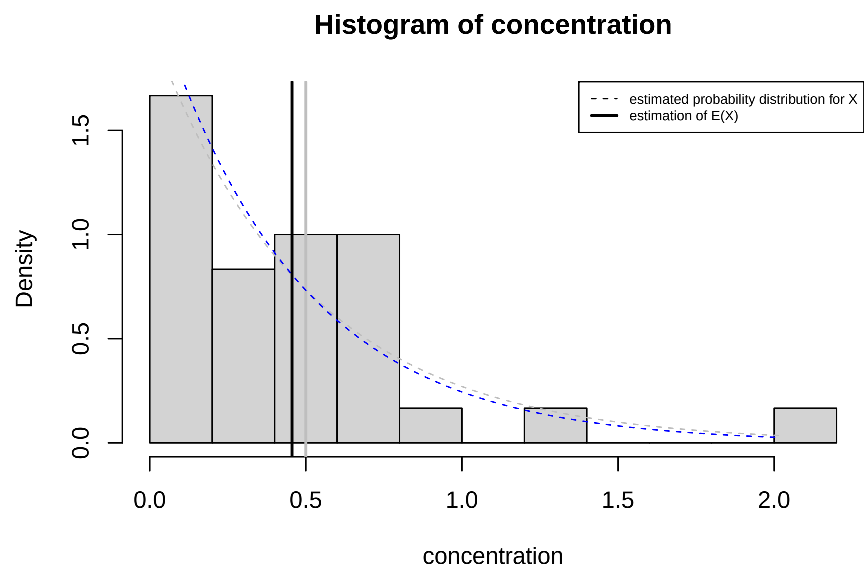
$$\hat{\mu} = 0.45 \pm 0.15$$



13.23 Estimación de parámetros

Como $E(X) = \mu = \frac{1}{\lambda}$ entonces

$$\hat{\lambda} = \frac{1}{\hat{\mu}} = 2.194812$$



o su IC de 95%:

$$\hat{\lambda} = (1.66, 3.33)$$

13.24 Estimación intervalar para proporciones

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?

Tenemos 136 fallas en 400 ensayos, cada ensayo es un juicio de Bernoulli

$$X \to Bernoulli(p)$$

con:

- la probabilidad p de fallo para una persona ($x = 1$)
- media $E(X) = p$
- varianza $V(X) = p(1 - p)$

Queremos tener un IC de 95% para p.

13.25 Estimación intervalar para proporciones

Si la distribución de un experimento aleatorio es

$$X \rightarrow Bernoulli(p)$$

Entonces $\bar{X}$ tiene

- media $E(\bar{X}) = E(X) = p$ (estimador insesgado de p)
- varianza $V(\bar{X}) = \frac{V(X)}{n} = \frac{p(p-1)}{n}$ (estimador consistente de p)

$$\hat{p} = \bar{x}$$

13.26 Estimación intervalar para proporciones

Cuando $\hat{p}n > 5$ y $(\hat{p} - 1)n > 5$

- La **estadística estandarizada** de $\bar{X}$ se puede aproximar mediante una distribución estándar

$$Z = \frac{\bar{X} - E(\bar{X})}{\sqrt{V(\bar{X})}} = \frac{\bar{X} - p}{\left[\frac{p(1-p)}{n}\right]^{1/2}} \rightarrow N(0,1)$$

- El intervalo de CI de 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 $p(1-p)$ por $\bar{x}(1-\bar{x})$.

13.27 Estimación intervalar para proporciones

En nuestro caso, estamos contando fallos en vacunas 136 en 400 ensayos

sabemos

- $\bar{x} = 134/400 = 0.34$
- $z_{0.025} = 1.96$

$CI = (l, u) = (\bar{x} - 1,96\big[\frac{\bar{x}(1-\bar{x})}{n}\big]^{1/2}, \bar{x} + 1,96\big[\frac{\bar{x}(1-\bar{x})}{n}\big]^{1/2})$

$$= (0.29, 0.39)$$

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%.

13.28 Estimación intervalar para la varianza

Un material metálico se impacta para medir la energía requerida para cortarlo a una temperatura dada.

- Se cortaron diez probetas de acero A238 a 60ºC a las siguientes energías de impacto (J)
- $64.1, 64.7, 64.5, 64.6, 64.5, 64.3, 64.6, 64.8, 64.2, 64.3$

Sabemos que la estimación de $s^2 = 0.227^2 = 0.051$, pero ¿cuál es su intervalo de confianza?

13.29 Estimación intervalar para la varianza

Si

- X es **normal**

entonces

$$W = \frac{S^2(n-1)}{\sigma^2}$$

Captura la proporción en el error de σ^2 y sigue una distribución χ^2 con $n-1$ grados de libertad

$$\frac{S^2}{\sigma^2}(n-1) \to \chi^2_{n-1}$$

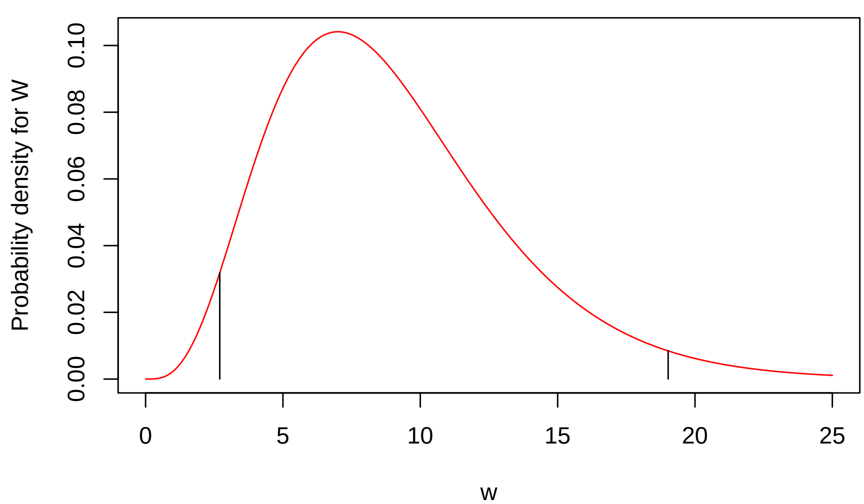
- Buscamos el intervalo de confianza de σ^2 con confianza 95% (L, U) tal que

$$P(L \leq \sigma^2 \leq U) = 0.95$$

Podemos usar el χ^2 para determinar el 95% de la distribución alrededor de W

$$P(\chi^2_{0.975,n-1} \leq W \leq \chi^2_{0.025,n-1}) = 0.95$$

13.30 χ^2-statistic



13.31 Estimación intervalar para la varianza

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 σ^2

$$P(\frac{S^2(n-1)}{\chi^2_{0.025,n-1}} \leq \sigma^2 \leq \frac{S^2(n-1)}{\chi^2_{0.975,n-1}}) = 0.95$$

El intervalo **aleatorio** con una confianza al 95% es

$$(L, U) = (\frac{S^2(n-1)}{\chi^2_{0.025,n-1}}, \frac{S^2(n-1)}{\chi^2_{0.975,n-1}})$$

y el intervalo de confianza **observado** al 95% (minúscula)

$$(l, u) = (\frac{s^2(n-1)}{\chi^2_{0.025,n-1}}, \frac{s^2(n-1)}{\chi^2_{0.975,n-1}})$$

13.32 Estimación intervalar para la varianza

$\chi^2_{0.975,n-1} = F^{-1}(0.025)$ para $n = 10$ o $df = n - 1 = 9$

```
chi0.975 <- qchisq(0.025, df=9)
chi0.975
```

```
[1] 2.700389
```

```
chi0.025 <- qchisq(0.975, df=9)
chi0.025
```

```
[1] 19.02277
```

13.33 Estimación intervalar

En nuestro ejemplo

- $s = 0.227$
- $n = 10$

$\hat{\sigma}^2 = (l, u) = (\frac{s^2(n-1)}{\chi^2_{0.025,n-1}}, \frac{s^2(n-1)}{\chi^2_{0.975,n-1}})$

$$= (\frac{0.227^2(10-1)}{19.02277}, \frac{0.227^2(10-1)}{2.700389}) = (0.02, 0.17)$$

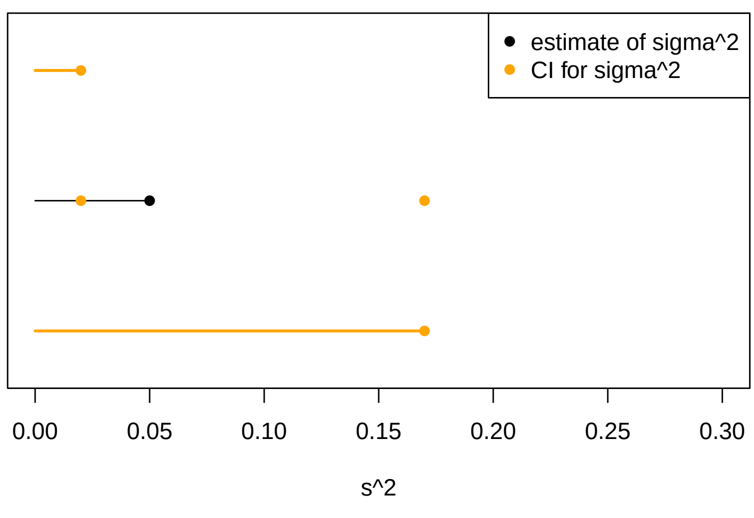
Según los datos, $\sigma^2 \neq 1$ con una confianza de 95%.

- ¿Habíamos cometido un error al considerar $\sigma = 1$ cuando calculamos el primer IC para estos datos?

en R: library(Ecfun); confint.var(0.05, 9)

13.34 Estimación intervalar

El intervalo para la varianza **no es simétrico** y no podemos formularlo como un margen de error estimado de $\pm$.



Chapter 14

Ejercicios

14.1 Descripción de datos

14.1.0.1 Ejercicio 1

Hemos realizado un experimento 12 veces con los siguientes resultados

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

Responde las siguientes preguntas:

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

14.1.0.2 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
```

Considere 10 contenedores de tamaño 1: [0,1], (1,2]…(9,10).

Responde las siguientes preguntas:

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

14.2 Probabilidad

14.2.0.1 Ejercicio 1

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: $x \in \{0, 1, 2, 3, 4\}$
- Depresión: $y \in \{0, 1\}$ (no:0, si:1)

```
##   Misofonia.dic depresion.dic
## 1             4             1
## 2             2             0
## 3             0             0
## 4             3             0
## 5             0             0
## 6             0             0
```

Un estudio en 123 pacientes mostró las frecuencias $n_{x,y}$ dadas en la tabla de contingencia:

```
##
##                Depression:0 Depression:1
##   Misophonia:4            0            9
##   Misophonia:3           25            6
##   Misophonia:2           34            3
##   Misophonia:1            5            0
##   Misophonia:0           36            5
```

Supongamos que $N >> 0$ y que las frecuencias **estiman** las probabilidades $f_{x,y} = \hat{P}(X, Y)$

```
##
##                Depression:0 Depression:1
##   Misophonia:4   0.00000000   0.07317073
##   Misophonia:3   0.20325203   0.04878049
##   Misophonia:2   0.27642276   0.02439024
##   Misophonia:1   0.04065041   0.00000000
##   Misophonia:0   0.29268293   0.04065041
```

- ¿Cuál es la probabilidad marginal de misofonía de gravedad 3?(R/0.3)
- ¿Cuál es la probabilidad de no ser misofónico **y** no estar deprimido?(R/0.293)
- ¿Cuál es la probabilidad de ser misofónico **o** deprimido?(R/0.293)
- ¿Cuál es la probabilidad de ser misofónico **y** deprimido?(R/0.707)
- Describir en palabras los resultados con probabilidad 0.

14.2.0.2 Ejercicio 2

Hemos realizado un experimento 10 veces con 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
```

- Crear la tabla de contingencia para el número ($n_{i,j}$) de observaciones de cada resultado (A, B)
- Crear 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** mujer? (R/0.6)

14.3 La probabilidad condicional

14.3.0.1 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: Sí	Redondeado: No
superficie lisa: sí	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: 201/201)
- ¿Cuál es la probabilidad estimada de que la varilla sea lisa si es redondeada? (R: 201/204)
- ¿Cuál es la probabilidad estimada de que la varilla no sea ni lisa ni redondeada si no cumple al menos un control de calidad? (R: 2/7)

- ¿Son eventos independientes la suavidad y la redondez? (no)

14.3.0.2 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)

14.3.0.3 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: Sí	Redondeado: No
superficie lisa: sí	200	1
superficie lisa: no	4	2

Máquina 2

	Redondeado: Sí	Redondeado: No
superficie lisa: sí	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)

14.3.0.4 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 hubiera tenido que parar en el primero? (R: 0.62)
- 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: $P(R_1, R_2, R_3) = P(R_3|R_2)P(R_2|R_1)P(R_1)$

14.3.0.5 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)

14.4 Variables aleatorias

14.4.0.1 Ejercicio 1

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)

14.4.0.2 Ejercicio 2

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: 6)
- ¿Cuál es la mediana y el primer y tercer cuartil de X? (R:2,0,2)

14.4.0.3 Ejercicio 3

Estamos probando un sistema para transmitir imágenes digitales. Primero consideramos el experimento de enviar 3 píxeles y tener por ejemplo 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.

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

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

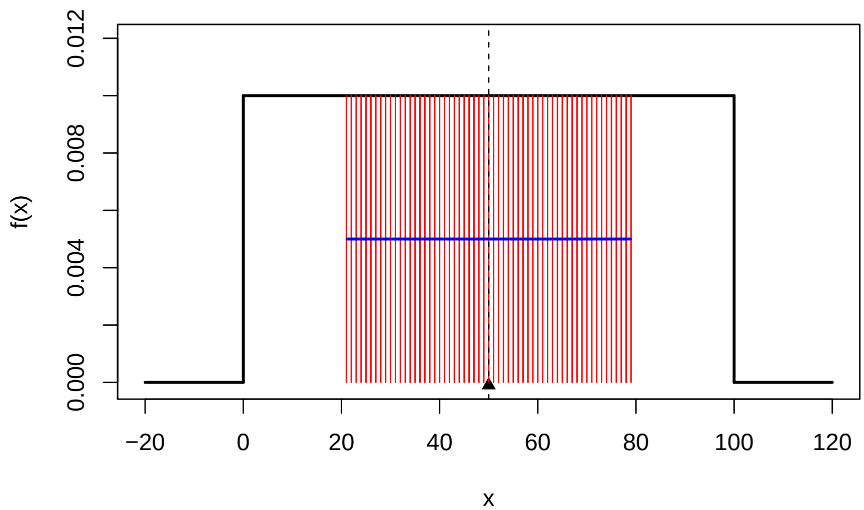
- ¿Cuál es la probabilidad de recibir errores de 0, 1, 2 o 3 en la transmisión de 3 píxeles? (R: 1/8; 3/8; 3/8; 1/8)
- Dibuje la función de masa de probabilidad para el número de errores
- ¿Cuál es el valor esperado para el número de errores? (R:1.5)
- ¿Cuál es su varianza? (R: 0.75)
- Dibujar la distribución de probabilidad
- ¿Cuál es la probabilidad de transmitir al menos 1 error? (R:7/8)

14.4.0.4 Ejercicio 4

Para la densidad de probabilidad

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

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



14.4.0.5 Ejercicio 5

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)

14.4.0.6 Ejercicio 6

Para la densidad de probabilidad

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{if } x \geq 0 \\ 0, & si\,no \end{cases}$$

- 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: $\log 2/\lambda$)

14.4.0.7 Ejercicio 7

Dada la distribución de probabilidad 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)

14.5 Modelos de probabilidad

14.5.0.1 Ejercicio 1

En una población, la probabilidad de que nazca un niño es $p = 0.51$. Considere una familia de 4 hijos.

- ¿Cuál es la probabilidad de que una familia de 4 hijos 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 o solo un niño o solo una niña?(R: 0.4999)
- ¿Cuál es la probabilidad de que la familia tenga al menos dos niños? (R: 0.7023)
- ¿Cuál es el número de hijos que debe tener una familia para que la probabilidad de tener al menos una niña sea superior a 0.75? (R:$n = 3 > \log(0.25)/\log(0.51)$)

14.5.0.2 Ejercicio 2

Un motor de búsqueda falla al recibir una petición 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 buscador funciona suficientemente bien cuando es capaz de encontrar información para mas de 10 solicitudes por cada 2

fallos. ¿Cuál es la probabilidad de que en un ensayo de fiabilidad nuestro motor de búsqueda sea satisfactorio? (R: 0.697)

14.5.0.3 Ejercicio 3

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 observar al menos una partículas en dos minutos? (R:0.953)
- ¿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).

14.5.0.4 Ejercicio 4

- ¿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)

14.6 Muestreo y teorema del límite central

14.6.0.1 Ejercicio 1

Un tipo de batería carga hasta el 75% de su capacidad en una hora, con una desviación estándar de 15%.

- Si cargamos 25 baterias, ¿cuál es la probabilidad de que la diferencia de carga entre el promedio de la muestra y la media sea como máximo de un 5%? (R:0.9044)
- Si cargamos 100 baterías, ¿cuál es esa probabilidad? (R:0.9991)

- Si, en cambio, solo cargamos 9 baterías, ¿cuál valor c que es superado por la media muestral con una probabilidad del 0.015? (R:85.850)

14.6.0.2 Ejercicio 2

Se necesita un componente electrónico para el correcto funcionamiento de un telescopio. Necesita ser reemplazado inmediatamente cuando se desgasta.

La vida media del componente (μ) es de 100 horas y su desviación estándar σ 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 una probabilidad de por lo menos $0,95$? (R:31)

14.6.0.3 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) está entre 128 y 132gr?
- ¿Cuál debe ser el tamaño de la muestra (n) para que la media muestral $\bar{X}$ sea mayor a 131gr con una probabilidad menor o igual a 0.025?

14.6.0.4 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?
- ¿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?

14.6.0.5 Ejercicio 5

La probabilidad de que se encuentre una mutación particular en la población es de 0.4. Si probamos 2000 personas para la mutación:

- ¿Cuál es la probabilidad de que el número total de personas con la mutación esté entre 791 y 809?

sugerencia: use el CLT con una muestra de ensayos de Bernoulli de 2000. Esto se conoce como la aproximación normal de la distribución binomial.

14.7 Estimadores puntuales

14.7.0.1 Ejercicio 1

Considere el modelo de probabilidad

$$f(x) = \begin{cases} 1/2 - a, & \text{si } x = -1 \\ 1/2, & \text{si } x = 0 \\ a, & 1\text{si } x = 1 \end{cases}$$

donde a es un parámetro.

Calcule la media y la varianza de la estadística:

$$T = \frac{\bar{X}}{2} + \frac{1}{4}$$

donde $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$

- ¿T es un estimador sesgado de a?
- ¿Es T consistente? es decir, $V(T) \to 0$ cuando $N \to \infty$

14.7.0.2 Ejercicio 2

- ¿Es $\bar{X}^2 = (\frac{1}{N}\sum_{i=1}^{N} X_i)^2$ un estimador insesgado de $E(X)^2$?

14.8 Máxima verosimilitud

14.8.0.1 Ejercicio 1

Tome 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 θ?
- Si tomamos una muestra de 5 con 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 θ?

- Calcular $E(X) = \mu$ en función de θ. ¿Cuál es la estimación de máxima verosimilitud para μ?

14.8.0.2 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?
- En **un** examen de 100 estudiantes observamos $x_1 = 68$ estudiantes que aprobaron el examen. ¿Cuál es la estimación de p?

14.8.0.3 Ejercicio 3

Tome una variable aleatoria con la siguiente función de densidad de probabilidad

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } 0 \leq x \\ 0, \; de\, lo\, contrario & \end{cases}$$

- ¿Cuál es la estimación de máxima verosimilitud para λ?
- Si tomamos una muestra de 5 con observaciones $x_1 = 0.223$ $\quad x_2 = 0.681;$ $\quad x_3 = 0,117;$ $\quad x_4 = 0,150;$ $\quad x_5 = 0.520$

¿Cuál es el valor estimado del parámetro λ?

- ¿Cuál es la estimación de máxima verosimilitud del parámetro $\alpha = \frac{n}{\lambda}$
- ¿El α un estimador insesgado y consistente de la media de la suma muestral $E(Y)$, donde $Y = \sum_1^n X_i$?

14.9 Método de los momentos

14.9.0.1 Ejercicio 1

¿Cuáles son los estimadores de los siguientes modelos paramétricos dados por el método de los momentos?

modelo	f(x)	E(X)
Bernoulli	$p^x(1-p)^{1-x}$	p
binomial	$\binom{n}{x}p^x(1-p)^{n-x}$	np
Geométrico desplazado	$p(1-p)^{x-1}$	$\frac{1}{p}$
Binomial negativo	$\binom{x+r-1}{x}p^r(1-p)^x$	$r\frac{1-p}{p}$
Poisson	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ
Exponencial	$\lambda e^{-\lambda x}$	$\frac{1}{\lambda}$
normales	$\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ

14.9.0.2 Ejercicio 2

Tome una variable aleatoria con la siguiente función de densidad de probabilidad

$$f(x) = \begin{cases} (1+\theta)x^{\theta}, & \text{if } x \in (0,1) \\ 0, & x \notin (0,1) \end{cases}$$

- Calcule $E(X)$ como una función de θ
- ¿Cuál es la estimación de θ utilizando el método de los momentos?
- Si tomamos una muestra de 5 con 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 θ?

14.9.0.3 Ejercicio 3

Considere 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:

$$x_1 = 27; \quad x_2 = 8; \quad x_3 = 22; \quad x_4 = 29; \quad x_5 = 19; \quad x_5 = 32$$