

Estadística

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

2023-03-22

Contents

1	Objetivo	7
1.1	Lectura recomendada	8
2	Descripción de datos	11
2.1	Método científico	11
2.2	Estadística	11
2.3	Datos	12
2.4	Tipos de resultado	12
2.5	Experimentos aleatorios	12
2.6	Frecuencias absolutas	13
2.7	Frecuencias relativas	13
2.8	Diagrama de barras	14
2.9	Gráfico de sectores (pie)	15
2.10	Variables categóricas ordinales	15
2.11	Frecuencias acumuladas absolutas y relativas	16
2.12	Gráfica de frecuencia acumulada	17
2.13	Variables numéricas	17
2.14	Transformando datos continuos	18
2.15	Tabla de frecuencias para una variable continua	19
2.16	Histograma	19
2.17	Gráfica de frecuencia acumulada	21
2.18	Estadísticas de resumen	21
2.19	Promedio (media muestral)	22
2.20	Promedio	23
2.21	mediana	23
2.22	Dispersión	25
2.23	Variación de la muestra	26
2.24	Rango intercuartílico (IQR)	27
2.25	Diagrama de caja	28
2.26	Preguntas	29
2.27	Ejercicios	30
3	Probabilidad	31

3.1	Experimentos aleatorios	31
3.2	Probabilidad de medición	32
3.3	Probabilidad clásica	32
3.4	Frecuencias relativas	32
3.5	Frecuencias relativas en el infinito	34
3.6	Probabilidad frecuentista	35
3.7	Probabilidades clásicas y frecuentistas	36
3.8	Definición de probabilidad	36
3.9	Tabla de probabilidades	37
3.10	Espacio muestral	37
3.11	Eventos	38
3.12	Álgebra de eventos	38
3.13	Resultados mutuamente excluyentes	38
3.14	Probabilidades conjuntas	39
3.15	Tabla de contingencia	40
3.16	La regla de la suma:	40
3.17	Preguntas	41
3.18	Ejercicios	42
4	Probabilidad condicional	45
4.1	Probabilidad conjunta	45
4.2	Independencia estadística	46
4.3	La probabilidad condicional	47
4.4	Tabla de contingencia condicional	47
4.5	Independencia estadística	48
4.6	Dependencia estadística	50
4.7	Prueba de diagnóstico	50
4.8	Probabilidades inversas	51
4.9	Teorema de Bayes	52
4.10	Ejercicios	54
4.11	Preguntas	54
5	Variables aleatorias discretas	59
5.1	Objetivo	59
5.2	Frecuencias relativas	59
5.3	Variable aleatoria	59
5.4	Eventos de observar una variable aleatoria	60
5.5	Probabilidad de variables aleatorias	60
5.6	Funciones de probabilidad	61
5.7	Funciones de probabilidad	61
5.8	Probabilidades y frecuencias relativas	62
5.9	La media o el valor esperado	64
5.10	Varianza	66
5.11	Funciones de probabilidad para funciones de X	67
5.12	Distribución de probabilidad	68
5.13	Función de probabilidad y distribución de probabilidad	69

5.14	Cuantiles	70
5.15	Resumen	70
5.16	Preguntas	71
5.17	Ejercicios	72
6	Variables aleatorias continuas	75
6.1	Objetivo	75
6.2	Variables aleatorias continuas	75
6.3	frecuencias relativas	76
6.4	función de densidad de probabilidad	77
6.5	Área total bajo la curva	78
6.6	Área bajo la curva	79
6.7	Probabilidades de variables continuas	80
6.8	Distribución de probabilidad	81
6.9	Gráficas de probabilidad	85
6.10	Media	86
6.11	Varianza	87
6.12	Funciones de X	87
6.13	Ejercicios	88
7	Modelos de probabilidad para variables aleatorias discretas	91
7.1	Objetivo	91
7.2	Función de probabilidad	91
7.3	Modelo de probabilidad	92
7.4	Modelos paramétricos	92
7.5	Distribución uniforme (un parámetro)	93
7.6	Distribución uniforme (dos parámetros)	94
7.7	ensayo de Bernoulli	97
7.8	Experimento binomial	99
7.9	Función de probabilidad binomial	100
7.10	Función de probabilidad binomial negativa	104
7.11	Distribución geométrica	107
7.12	Modelo hipergeométrico	107
7.13	Preguntas	110
7.14	Ejercicios	111

Chapter 1

Objetivo

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

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

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

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

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

Objetivos de evaluación:

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

coordinadores:

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

1.1 Lectura recomendada

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

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

Chapter 2

Descripción de datos

En este capítulo, presentaremos herramientas para describir datos.

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

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

2.1 Método científico

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

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

El método tiene tres actividades humanas principales:

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

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

2.2 Estadística

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

Las preguntas de tipo estadístico son:

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

2.3 Datos

Los datos se presentan en forma de observaciones.

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

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

... 1 0 0 1 0 1 0 1 1 ...

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

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

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

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

2.4 Tipos de resultado

En estadística nos interesan principalmente dos tipos de resultados.

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

2.5 Experimentos aleatorios

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

Definición:

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

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

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

2.6 Frecuencias absolutas

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

Resumimos las observaciones contando cuántas veces vimos un resultado particular.

Frecuencia absoluta:

$$n_i$$

es el número de veces que observamos el resultado i .

Ejemplo (leucocitos)

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

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

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

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

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

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

2.7 Frecuencias relativas

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

$$f_i = n_i/N$$

donde N es el número total de observaciones

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

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

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

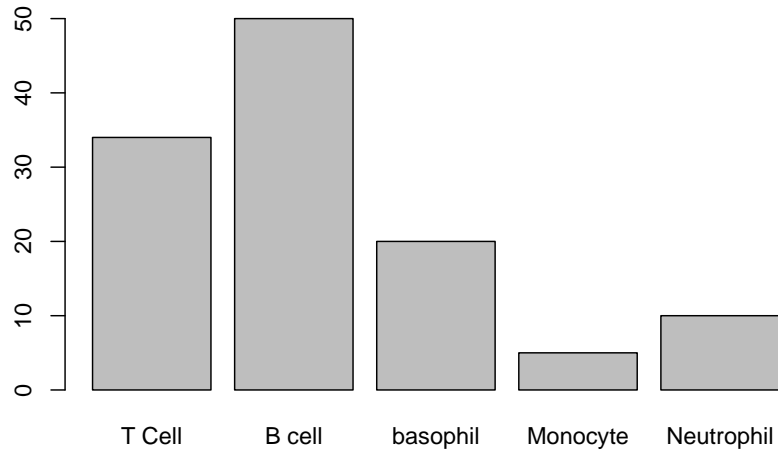
Para las frecuencias absolutas y relativas tenemos las propiedades

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

donde M es el número de resultados.

2.8 Diagrama de barras

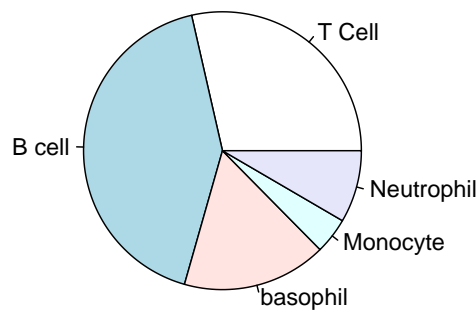
Cuando tenemos muchos resultados y queremos ver cuáles son los más probables, podemos usar un gráfico de barras que es una cifra de n_i Vs los resultados.



2.9 Gráfico de sectores (pie)

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

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



2.10 Variables categóricas ordinales

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

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

Ejemplo (misofonía)

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

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

Los resultados del estudio son

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

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

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

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

2.11 Frecuencias acumuladas absolutas y relativas

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

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

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

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

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

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

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

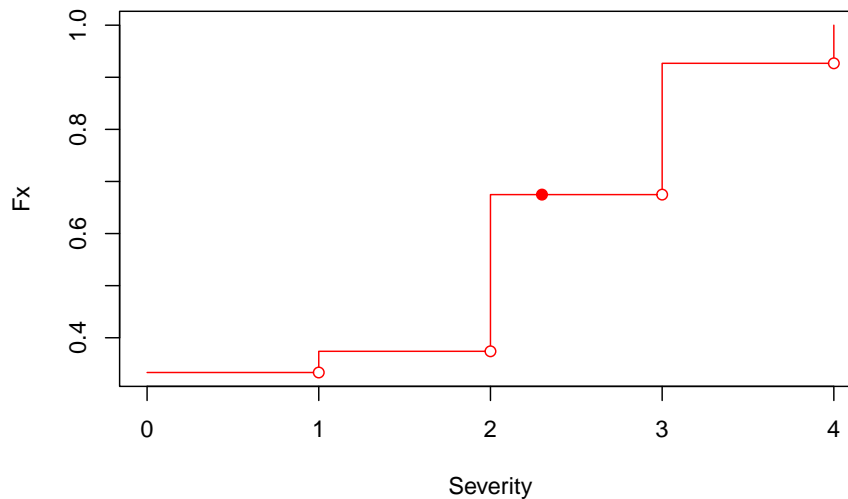
Por lo tanto, el **67 %** de los pacientes tenían misofonía hasta la gravedad **2** y el **37 %** de los pacientes tenían una gravedad inferior o igual a **1**.

2.12 Gráfica de frecuencia acumulada

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

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

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



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

2.13 Variables numéricas

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

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

Ejemplo (misofonía)

Los investigadores se preguntaron si la convexidad de la mandíbula afectaría la

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

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

2.14 Transformando datos continuos

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

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

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

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

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

2.15. TABLA DE FRECUENCIAS PARA UNA VARIABLE CONTINUA 19

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

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

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

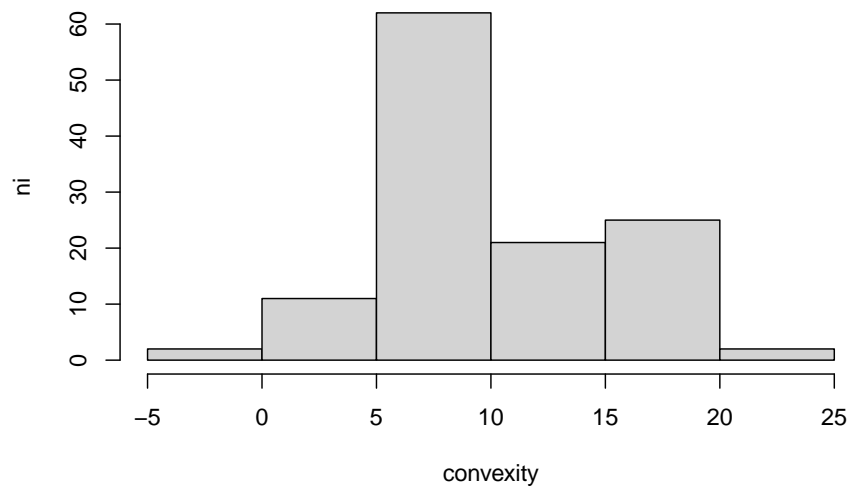
2.15 Tabla de frecuencias para una variable continua

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

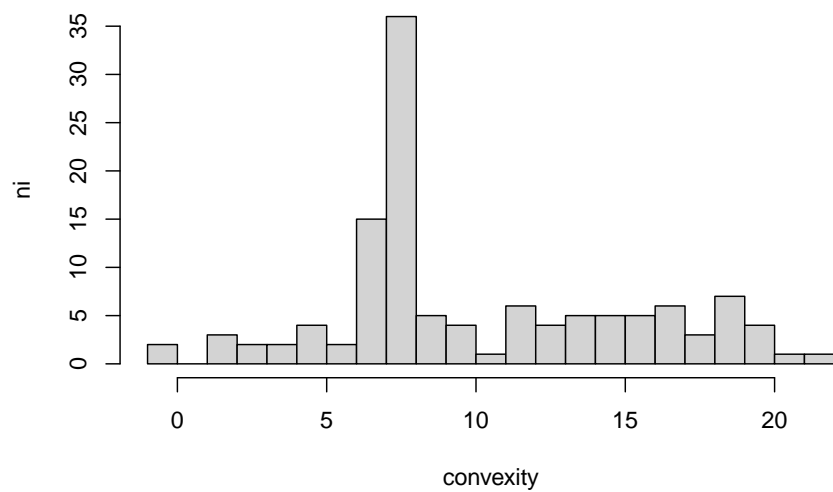
##	outcome	ni	fi	Ni	Fi
## 1	[-1.02,3.46]	8	0.06504065	8	0.06504065
## 2	(3.46,7.92]	51	0.41463415	59	0.47967480
## 3	(7.92,12.4]	26	0.21138211	85	0.69105691
## 4	(12.4,16.8]	20	0.16260163	105	0.85365854
## 5	(16.8,21.3]	18	0.14634146	123	1.00000000

2.16 Histograma

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



Este es un histograma con 20 bins.



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

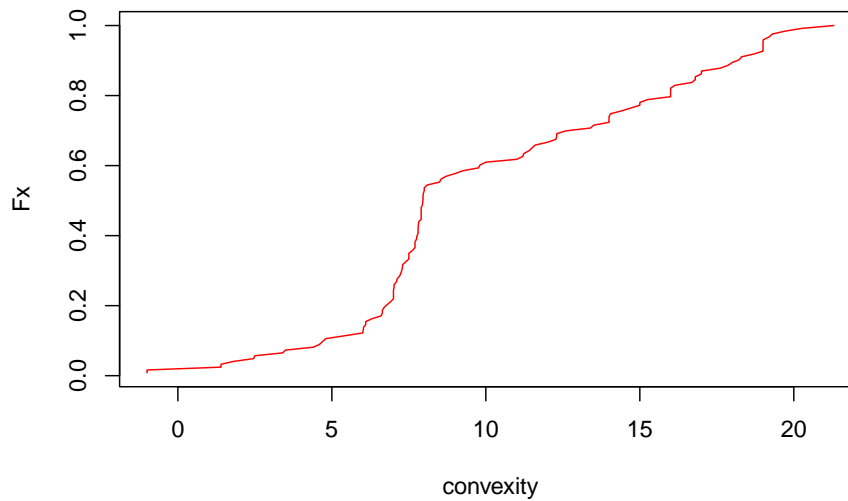
2.17 Gráfica de frecuencia acumulada

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

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

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

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



2.18 Estadísticas de resumen

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

Por ejemplo, tenemos estadísticas que describen los valores extremos:

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

2.19 Promedio (media muestral)

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

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

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

Ejemplo (Misofonía)

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

$$\begin{aligned} \bar{x} &= \frac{1}{N} \sum_j x_j \\ &= \frac{1}{N} (7.97 + 18.23 + 12.27 \dots + 6.80) = 10.19894 \end{aligned}$$

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

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

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

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

Ejemplo (Misofonía)

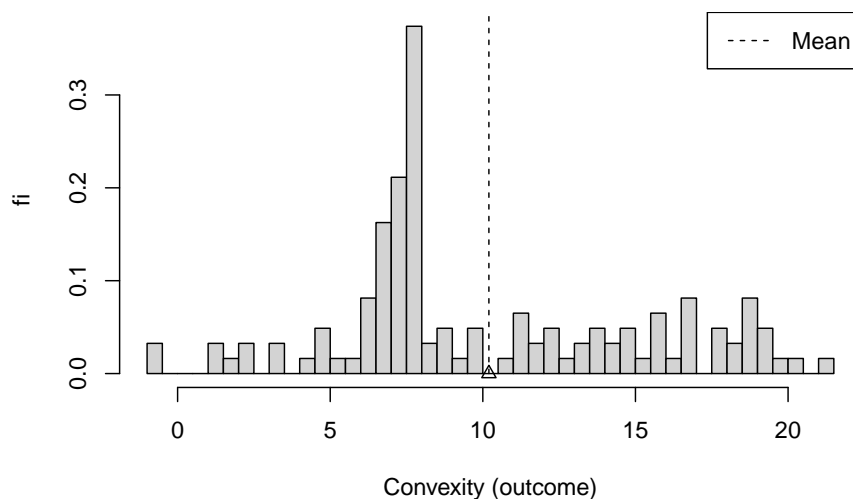
La **severidad** promedio de la misofonía en el estudio se puede calcular a partir de las frecuencias relativas de los **resultados**

```
## outcome ni      fi
## 1      0 41 0.3333333
## 2      1  5 0.04065041
## 3      2 37 0.30081301
## 4      3 31 0.25203252
## 5      4  9 0.07317073
```

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

2.20 Promedio

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



2.21 mediana

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

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

Ejemplo (Misofonía)

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

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

```
## [61] 7.95 7.96

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

## [1] 7.96
```

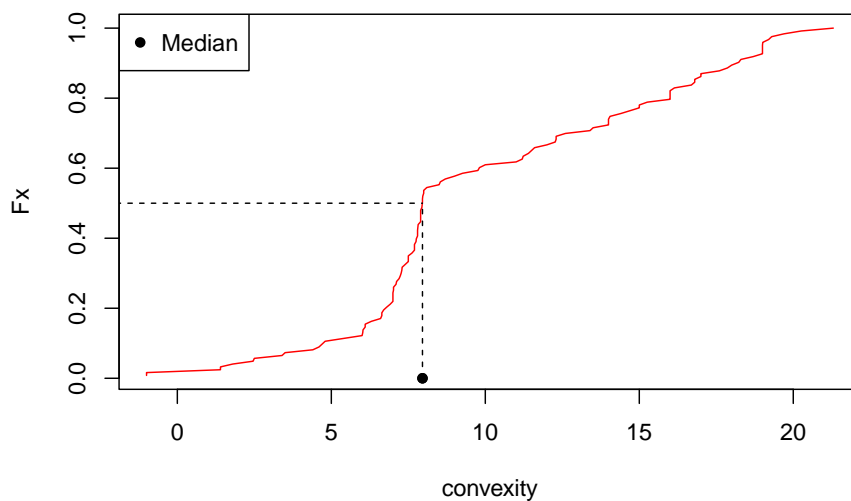
En términos de frecuencias, $q_{0.5}$ hace que la frecuencia acumulada F_x sea igual a 0.5

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

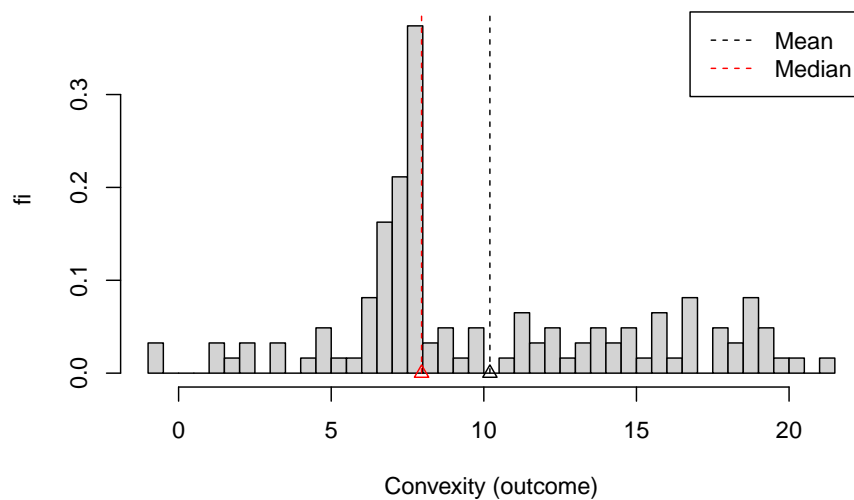
o

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

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



El promedio y la mediana no siempre son iguales.

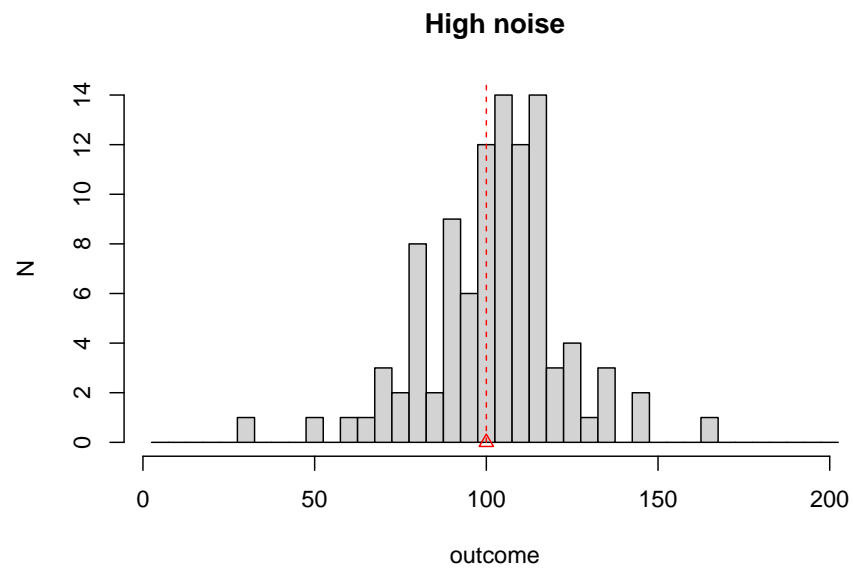
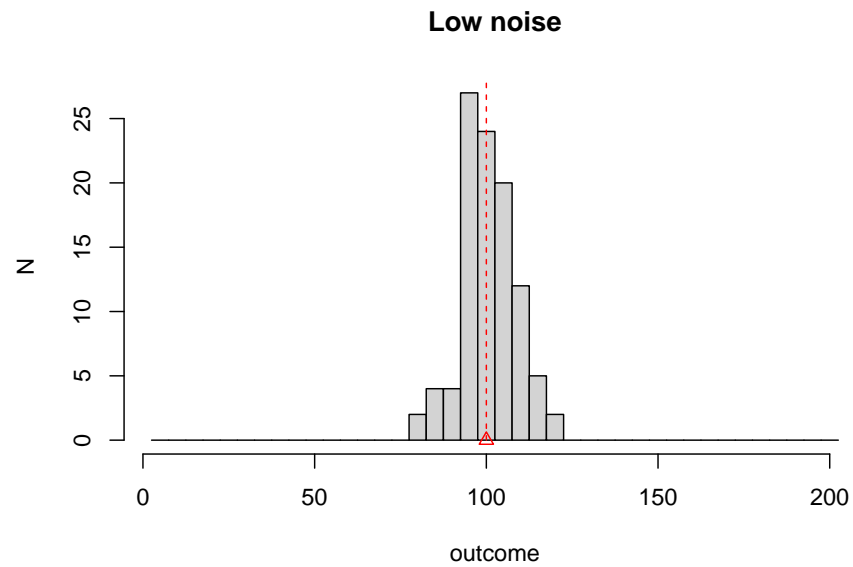


2.22 Dispersión

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

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

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



2.23 Variación de la muestra

La dispersión sobre la media se mide con la varianza muestral

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

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

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

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

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

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

Ejemplo (Misofonía)

La desviación estándar del ángulo de convexidad es

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

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

2.24 Rango intercuartílico (IQR)

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

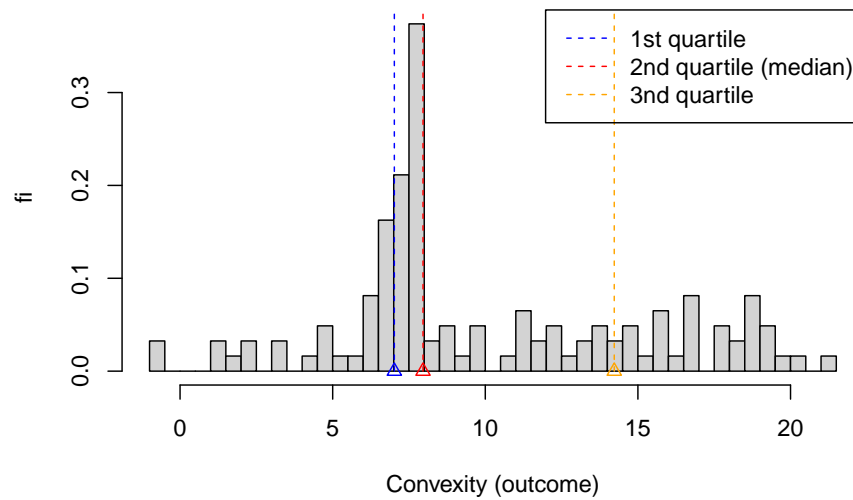
- 1) Definimos el **primer** cuartil como el valor x_m que hace que la frecuencia acumulada $F_{q_{0.25}}$ sea igual a 0.25 (x donde hemos acumulado una cuarta parte de las observaciones)

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

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

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

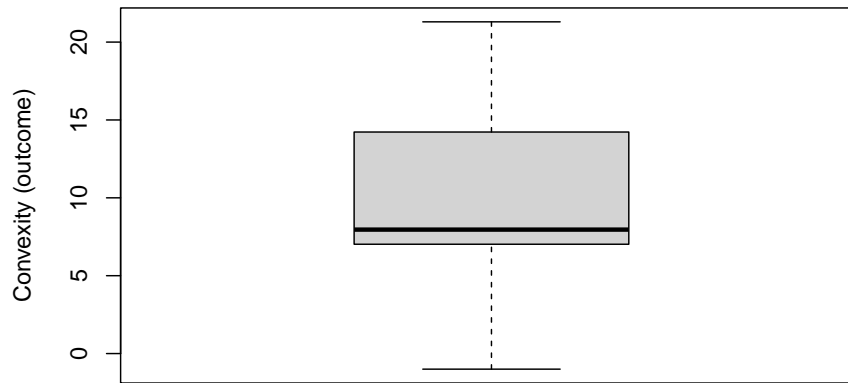
- 3) El **rango intercuartílico** (IQR) es $IQR = q_{0.75} - q_{0.25}$. Esa es la distancia entre el tercer y el primer cuartil y captura el 50% central de las observaciones



2.25 Diagrama de caja

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

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



2.26 Preguntas

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

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

2) La principal desventaja de un histograma es que:

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

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

Entonces la frecuencia relativa para el resultado 3 es

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

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

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

El primer cuartil de los datos es:

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

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

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

2.27 Ejercicios

2.27.0.1 Ejercicio 1

Hemos realizado un experimento 8 veces con los siguientes resultados

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

Responde las siguientes cuestiones:

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

2.27.0.2 Ejercicio 2

Hemos realizado un experimento 10 veces con los siguientes resultados

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

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

Responde las siguientes cuestiones:

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

Chapter 3

Probabilidad

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

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

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

3.1 Experimentos aleatorios

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

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

Las observaciones son realizaciones de **resultados**.

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

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

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

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

3.2 Probabilidad de medición

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

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

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

3.3 Probabilidad clásica

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

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

.

La probabilidad clásica fue defendida por Laplace (1814).

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

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

Ejemplo (dado):

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

$$P_2 = 1/6 = 0.166666.$$

3.4 Frecuencias relativas

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

¿Cómo podemos entonces definir las probabilidades de los resultados?

Ejemplo (experimento aleatorio)

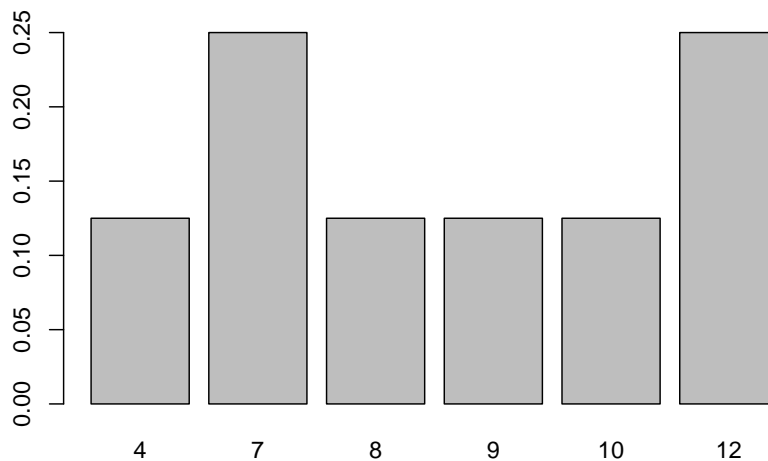
Imaginemos que repetimos un experimento aleatorio 8 veces y obtenemos las siguientes observaciones

8 4 12 7 10 7 9 12

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

La tabla de frecuencias es

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



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

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

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

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

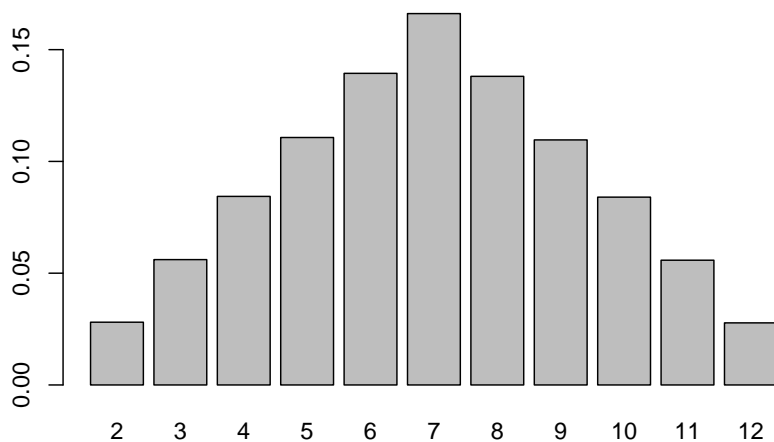
Ejemplo (experimento aleatorio con mas repeticiones)

Digamos que repetimos el experimento 100000 veces más:

La tabla de frecuencias es ahora

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

y el gráfico de barras es



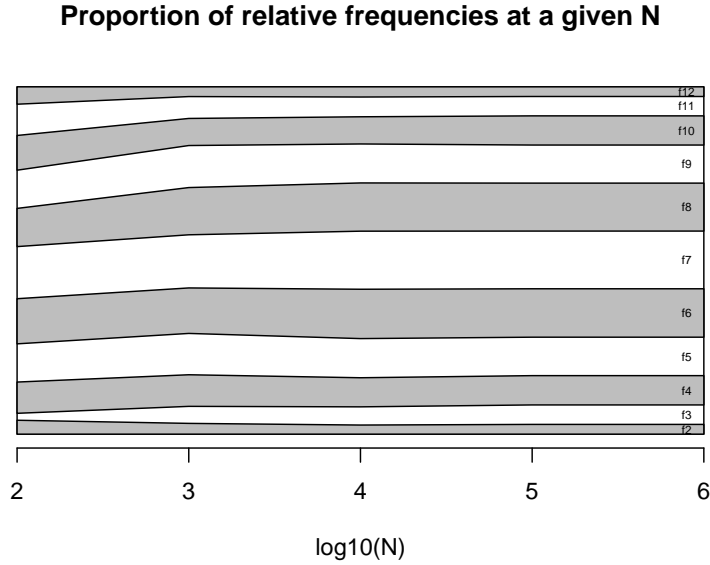
Aparecieron nuevos resultados y f_{12} ahora es solo 0.027, y entonces estamos sólo un $\sim 3\%$ seguros de obtener 12 en el próximo experimento. Las probabilidades medidas por f_i cambian con N .

3.5 Frecuencias relativas en el infinito

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

En este gráfico cada sección vertical da la frecuencia relativa de cada observación. Vemos que después de $N = 1000$ ($\log_{10}(N) = 3$) las proporciones apenas varían

con mas N .



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

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

3.6 Probabilidad frecuentista

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

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

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

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

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

3.7 Probabilidades clásicas y frecuentistas

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

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

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

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

Ejemplo (suma de dos dados)

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

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

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

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

3.8 Definición de probabilidad

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

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

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

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

$$P(S) = 1$$

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

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

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

3.9 Tabla de probabilidades

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

Ejemplo (Dado)

La tabla de probabilidad para el lanzamiento de un dado

resultado	Probabilidad
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6
$P(1 \cup 2 \cup \dots \cup 6)$	1

Verifiquemos los axiomas:

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

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

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

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

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

3.10 Espacio muestral

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

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

Por ejemplo:

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

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

3.11 Eventos

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

Ejemplos de eventos:

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

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

3.12 Álgebra de eventos

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

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

Ejemplo (dado)

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

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

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

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

3.13 Resultados mutuamente excluyentes

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

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

$$E_1 \cap E_2 = \emptyset$$

Ejemplos:

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

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

3.14 Probabilidades conjuntas

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

$$P(A \cap B)$$

o $P(A, B)$.

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

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

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

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

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

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

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

Resultado	Probabilidad
suma	1

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

3.15 Tabla de contingencia

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

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

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

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

En nuestro ejemplo, la tabla de contingencia es

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

3.16 La regla de la suma:

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

- 1) Usando las marginales y la probabilidad conjunta

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

- 2) Usando solo probabilidades conjuntas

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

3) Usando el complemento de la probabilidad conjunta

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

Ejemplo (dado)

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

Por lo tanto:

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

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

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

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

	B	B'
A	1/6	2/6
A'	2/6	1/6

3.17 Preguntas

Recopilamos la edad y categoría de 100 deportistas en una competición

	<i>edad : junior</i>	<i>edad : senior</i>
<i>categoría : 1ra</i>	14	12
<i>categoría : 2a</i>	21	18
<i>categoría : 3a</i>	22	13

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

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

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

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

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

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

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

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

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

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

3.18 Ejercicios

3.18.0.1 Probabilidad clásica: Ejercicio 1

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

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

3.18.0.2 Probabilidad frecuentista: Ejercicio 2

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

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

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

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

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

Reponde las siguientes preguntas:

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

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

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

3.18.0.3 Ejercicio 3

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

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

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

3.18.0.4 Teoría: Ejercicio 4

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

Chapter 4

Probabilidad condicional

En este capítulo, introduciremos la probabilidad condicional.

Usaremos la probabilidad condicional para definir la independencia estadística.

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

4.1 Probabilidad conjunta

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

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

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

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

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

4.2 Independencia estadística

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

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

Ejemplo (conductor)

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

Los datos se verían como

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

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

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

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

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

- $P(D, F) = 0.005$

y las probabilidades marginales son

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

4.3 La probabilidad condicional

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

Consideremos primero solamente los materiales que tienen fallas.

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

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

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

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

Definición:

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

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

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

4.4 Tabla de contingencia condicional

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

	F	F'
D	$P(D F)$	$P(D F')$
D'	$P(D' F)$	$P(D' F')$
suma	1	1

Donde las probabilidades por columnas suman uno. La primera columna muestra las probabilidades de ser defectuoso o no solo de que los materiales que tienen

fallas (primera condición: F). La segunda columna muestra las probabilidades solo para los materiales que no tienen fallas (segunda condición: F').

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

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

4.5 Independencia estadística

En nuestro ejemplo, la tabla de contingencia condicional es

	F	F'
D	$P(D F) = 0.05$	$P(D F')=0.05$
D'	$P(D' F)=0.95$	$P(D' F')=0.95$
suma	1	1

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

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

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

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

Definición

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

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

y por la definición de probabilidad condicional

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

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

En nuestra tabla de probabilidad conjunta original

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

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

Ejemplo (Monedas)

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

resultado	Probabilidad
(H, T)	1/4
(H, H)	1/4
(T, T)	1/4
(T, H)	1/4
suma	1

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

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

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

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

4.6 Dependencia estadística

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

- inadecuado (si)
- adecuado (no)

con una prueba (t) con resultados

- positivo
- negativo

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

4.7 Prueba de diagnóstico

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

- si (infectado)
- no (no infectado)

Prueba:

- positivo
- negativo

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

	Infección: si	Infección: No
Test: positivo	$P(\text{positivo} \text{si})$	$P(\text{positivo} \text{no})$
Test: negativo	$P(\text{negativo} \text{si})$	$P(\text{negativo} \text{no})$
suma	1	1

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

- 2) Tasa de verdaderos negativos (Especificidad): La probabilidad de dar negativo **si** no tiene la enfermedad $P(\text{negativo} | \text{no})$
- 3) Tasa de falsos positivos: la probabilidad de dar positivo **si** no tiene la enfermedad $P(\text{positivo} | \text{no})$

- 4) Tasa de falsos negativos: la probabilidad de dar negativo **si** tiene la enfermedad $P(\text{negativo}|\text{si})$

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

Ejemplo (COVID)

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

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

La tabla de contingencia condicional es

	Infección: si	Infección: No
Test: positivo	$P(\text{positivo} \text{si})=0.7$	$P(\text{positivo} \text{no})=0.06$
Test: negativo	$P(\text{negativo} \text{si})=0.3$	$P(\text{negativo} \text{no})=0.94$
suma	1	1

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

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

4.8 Probabilidades inversas

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

$$P(\text{si}|\text{positivo})$$

Para eso:

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

	Infección: si	Infección: No	suma
Test: positivo	$P(\text{positivo} \text{si})P(\text{si})$	$P(\text{positivo} \text{no})P(\text{no})$	$P(\text{positivo})$
Test: negativo	$P(\text{negativo} \text{si})P(\text{si})$	$P(\text{negativo} \text{no})P(\text{no})$	$P(\text{negativo})$
suma	$P(\text{si})$	$P(\text{no})$	1

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

	Infección: si	Infección: No	suma
Test: positivo	$P(\text{si} \text{positivo})$	$P(\text{sin} \text{positivo})$	1
Test: negativo	$P(\text{si} \text{negativo})$	$P(\text{sin} \text{negativo})$	1

Por ejemplo:

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

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

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

$$\begin{aligned} P(\text{positivo}) &= P(\text{positivo} \cap \text{si}) + P(\text{positivo} \cap \text{no}) \\ &= P(\text{positivo}|\text{si})P(\text{si}) + P(\text{positivo}|\text{no})P(\text{no}) \end{aligned}$$

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

4.9 Teorema de Bayes

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

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

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

$$P(\text{positivo}|\text{si}) \rightarrow P(\text{si}|\text{positivo})$$

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

Ejemplo (COVID):

El rendimiento de la prueba fue:

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

El estudio en población española dio:

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

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

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

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

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

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

La sustitución de todos los valores da

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

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

Teorema de Bayes

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

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

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

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

$$P(B) = P(B|E1)P(E1) + \dots + P(B|Ek)P(Ek)$$

Árbol condicional

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

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

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

4.10 Ejercicios

4.11 Preguntas

Recopilamos la edad y categoría de 100 deportistas en una competición

	<i>junior</i>	<i>senior</i>
<i>1er</i>	14	12
<i>2do</i>	21	18
<i>3er</i>	22	13

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

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

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

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

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

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

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

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

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

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

4.11.0.1 Ejercicio 1

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

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

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

4.11.0.2 Ejercicio 2

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

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

4.11.0.3 Ejercicio 3

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

Máquina 1

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

Máquina 2

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

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

4.11.0.4 Ejercicio 4

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

Cuando intentamos cruzar ambos semáforos:

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

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

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

consejos:

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

4.11.0.5 Ejercicio 5

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

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

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

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

Chapter 5

Variables aleatorias discretas

5.1 Objetivo

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

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

5.2 Frecuencias relativas

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

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

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

5.3 Variable aleatoria

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

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

Definición:

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

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

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

Ejemplo:

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

5.4 Eventos de observar una variable aleatoria

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

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

...

En general:

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

5.5 Probabilidad de variables aleatorias

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

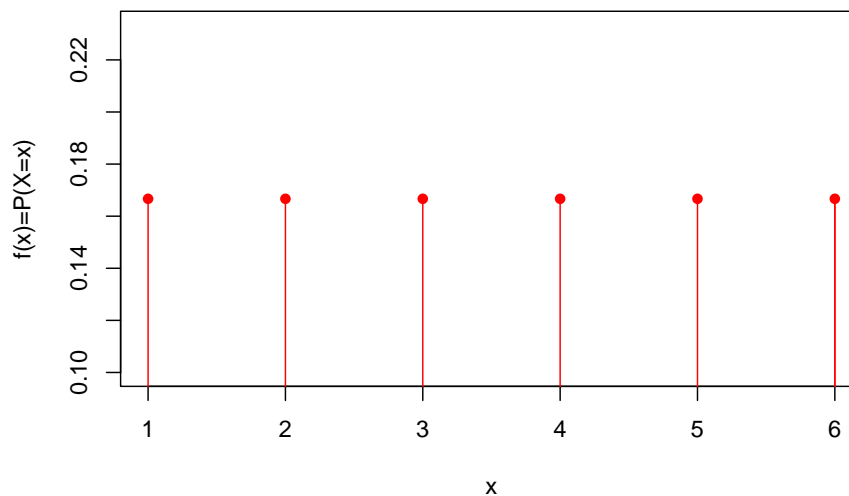
Por ejemplo, para los dados escribiremos la tabla de probabilidad como

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

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

5.6 Funciones de probabilidad

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



o escribir como una función

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

5.7 Funciones de probabilidad

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

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

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

siempre es positiva

- $f(x_i) \geq 0$

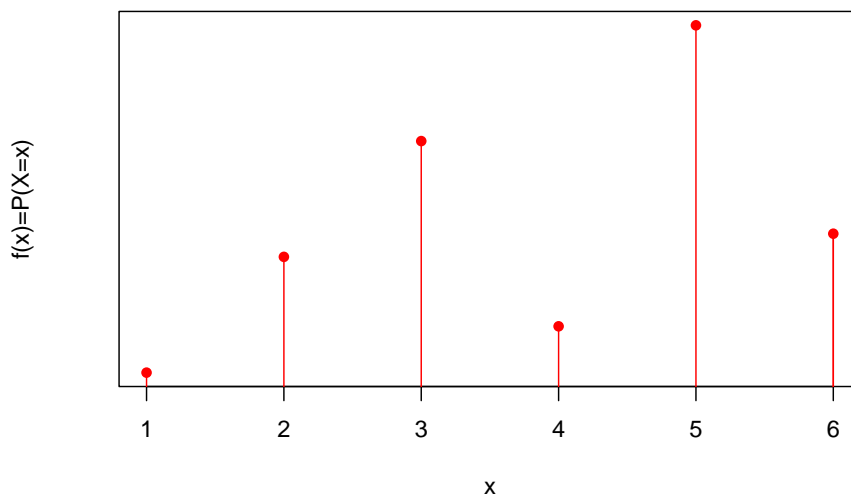
y su suma sobre todos los valores de la variable es 1:

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

Donde M es el número de resultados posibles.

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

Aquí tenemos un ejemplo



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

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

5.8 Probabilidades y frecuencias relativas

Considera el ejemplo

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

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

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

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

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

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

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

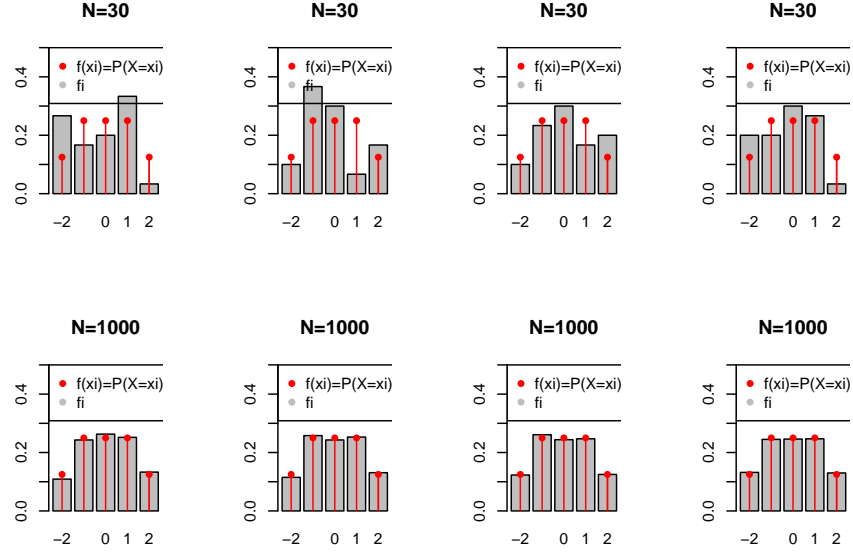
X	f_i
-2	0.132
-1	0.262
0	0.240
1	0.248
2	0.118

La probabilidad frecuentista nos dice

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

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

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



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

5.9 La media o el valor esperado

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

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

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

Definición

La **media** (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.

De la definición tenemos

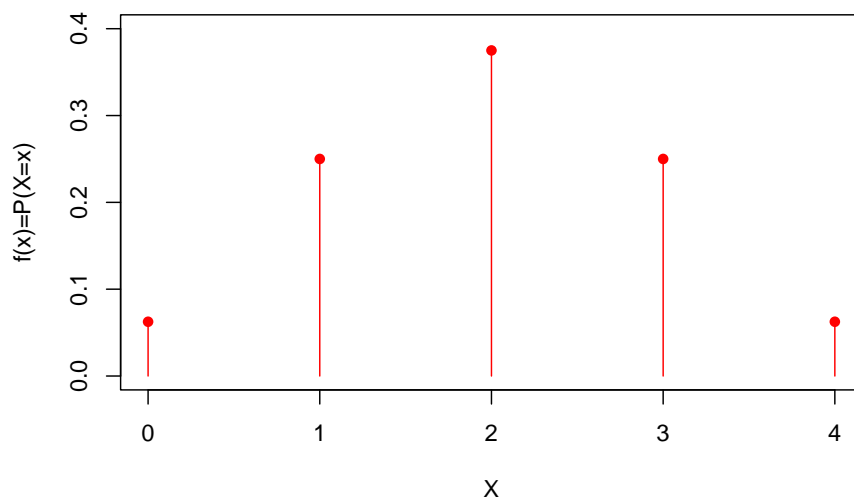
$$\bar{x} \rightarrow \mu$$

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

Ejemplo

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

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



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

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

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

5.10 Varianza

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

Definición

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

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

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

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

Ejemplo

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

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

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

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

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

$$\sigma = 1$$

5.11 Funciones de probabilidad para funciones de X

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

Definición

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

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

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

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

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

para a y b escalares (números).

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

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

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

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

Ejemplo

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

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

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

$$E(X^2) = (0)^2 \cdot 1/16 + (1)^2 \cdot 4/16 + (2)^2 \cdot 6/16 + (3)^2 \cdot 4/16 + (4)^2 \cdot 1/16 = 5$$

También podemos verificar:

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

$$5 = 1 + 2^2$$

5.12 Distribución de probabilidad

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

Definición:

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

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

Esa es la probabilidad acumulada hasta un valor dado x

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

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

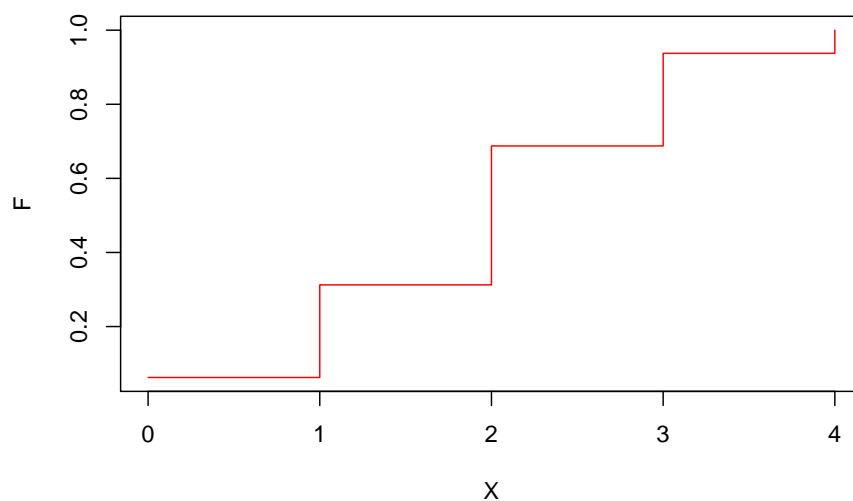
Para la función de masa de probabilidad:

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

La distribución de probabilidad es:

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

Para $X \in \mathbb{Z}$



5.13 Función de probabilidad y distribución de probabilidad

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

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

con

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

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

Ejemplo

De la distribución de probabilidad:

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

Podemos obtener la función masa de probabilidad.

$$\begin{aligned} f(0) &= F(0) = 1/16 & f(1) &= F(1) - f(0) = 5/32 - 1/32 = 4/16 & f(2) &= F(2) - \\ & & f(1) - f(0) &= F(2) - F(1) = 6/16 & f(3) &= F(3) - f(2) - f(1) - f(0) = F(3) - \\ & & & & F(2) &= 4/16 & f(4) &= F(4) - F(3) = 1/16 \end{aligned}$$

5.14 Cuantiles

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

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

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

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

$$F(x_m) = 0.5$$

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

$$F(x_r) = 0.05$$

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

$$F(x_s) = 0.25$$

5.15 Resumen

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

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

Ten en cuenta:

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

Propiedades:

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

5.16 Preguntas

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

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

2) El valor de una variable aleatoria representa

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

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

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

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

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

5) La media y la varianza

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

5.17 Ejercicios

5.17.0.1 Ejercicio 1

Considera la siguiente variable aleatoria X sobre los resultados

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

- a) Si cada resultado es igualmente probable, ¿cuál es la función de masa de probabilidad de x ?
- b) Encuentra:
- $P(X > 3)$
 - $P(X = 0 \cup X = 2)$
 - $P(X \leq 2)$

5.17.0.2 Ejercicio 2

Dada la función de masa de probabilidad

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

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

5.17.0.3 Ejercicio 3

Dada la distribución de probabilidad para una variable discreta X

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

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

5.17.0.4 Ejercicio 4

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

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

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

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

Chapter 6

Variables aleatorias continuas

6.1 Objetivo

En este capítulo estudiaremos variables aleatorias continuas.

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

6.2 Variables aleatorias continuas

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

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

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

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

Ejemplo (misofonía)

Reconsideremos el ángulo de convexidad de los pacientes con misofonía (Sección 2.21). El ángulo de convexidad de 123 pacientes fue medido. Entendimos cada

medición como el resultado de un experimento aleatorio que repetimos 123 veces y que podíamos describir en una tabla de frecuencias o en un histograma.

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

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

6.3 frecuencias relativas

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

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

Esta probabilidad depende de la longitud de los bins Δx .

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

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

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

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

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

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

6.4 función de densidad de probabilidad

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

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

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

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

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

Definición

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

- 1) Es positiva:

$$f(x) \geq 0$$

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

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

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

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

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

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

$$dx \rightarrow 0$$

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

$$\Sigma \rightarrow \int$$

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

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

6.5 Área total bajo la curva

Ejemplo (gotas de lluvia)

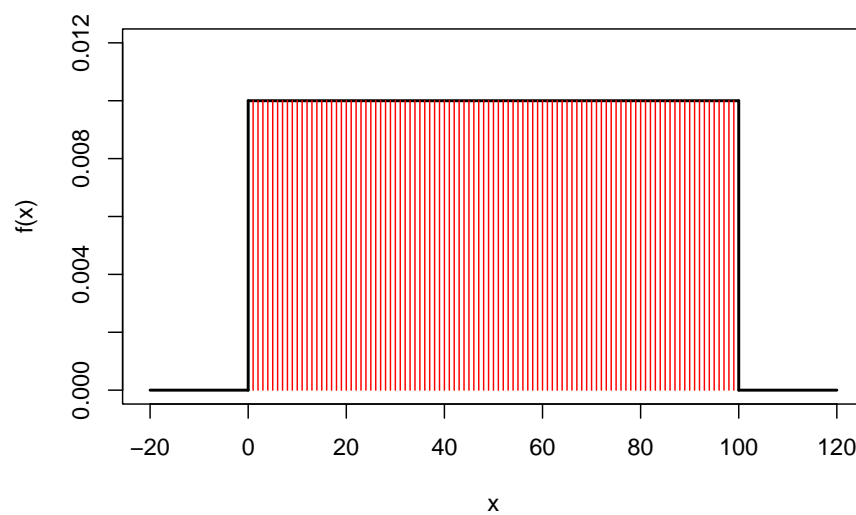
tomemos una **densidad de probabilidad** que podría describir la variable aleatoria que mide dónde cae una gota de lluvia en una canaleta de 100 cm de longitud.

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

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

- 1) es evidente a partir de la definición que $f(x) \geq 0$
- 2) La probabilidad de observar **cualquier valor** de X es el **área total bajo la curva**

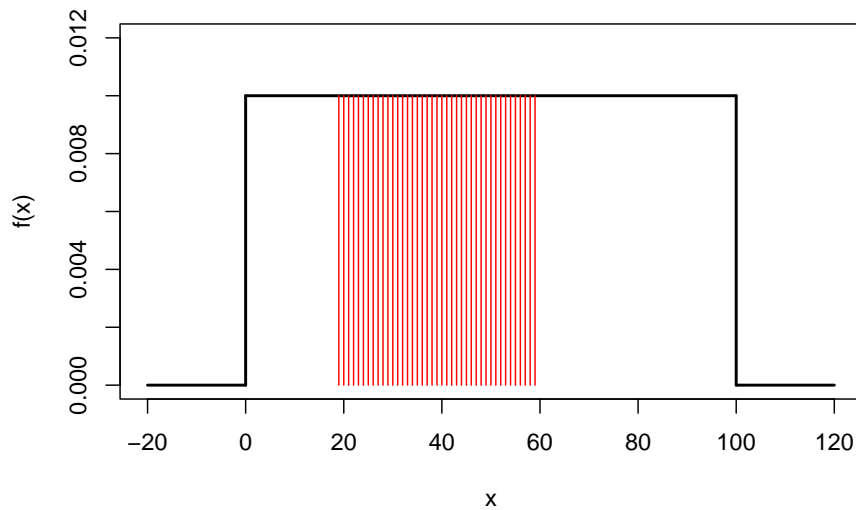
$$P(-\infty \leq X \leq \infty) = \int_{-\infty}^{\infty} f(x)dx = 100 * 0.01 = 1$$



6.6 Área bajo la curva

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

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



6.7 Probabilidades de variables continuas

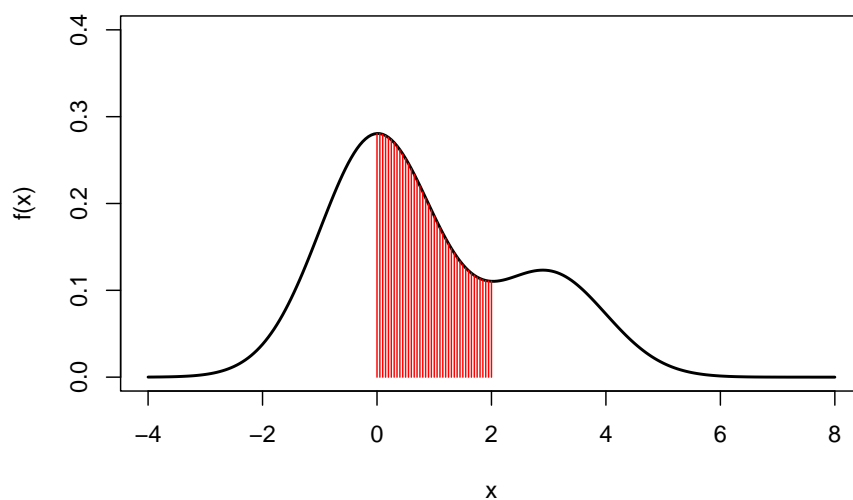
Para variables continuas, calculamos la probabilidad de que la variable esté en un intervalo dado. Eso es

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

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

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

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



6.8 Distribución de probabilidad

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

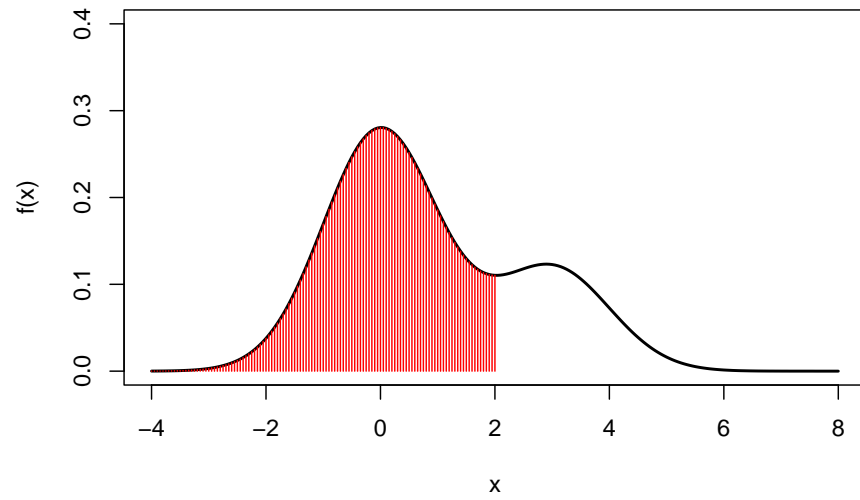
$$F(c) = P(X \leq c)$$

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

Si consideramos que:

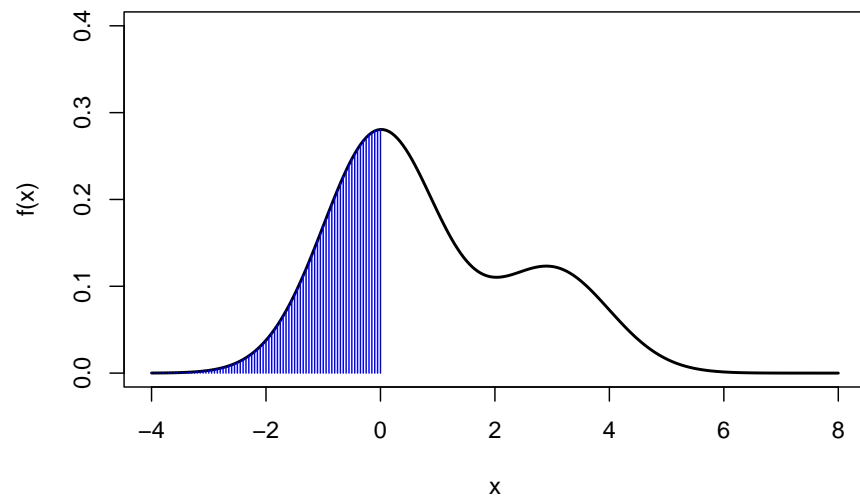
1) la probabilidad acumulada hasta b está dada por

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



2) la probabilidad acumulada hasta a es

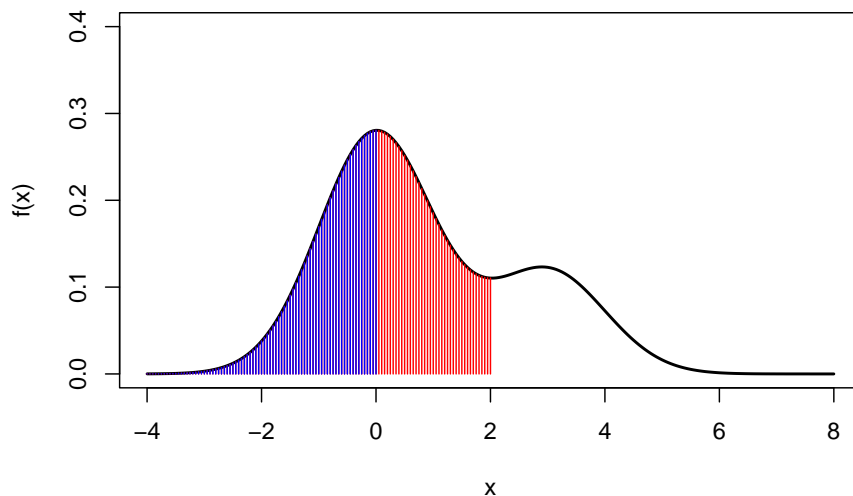
- $F(a) = P(X \leq a)$



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

la distribución de probabilidad

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



Definición

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

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

y tiene las siguientes propiedades:

- 1) Está entre 0 y 1:

$$F(-\infty) = 0 \text{ y } F(\infty) = 1$$

- 2) Siempre aumenta:

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

si $a \leq b$

- 3) Se puede utilizar para calcular probabilidades:

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

4) Recupera la densidad de probabilidad:

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

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

Ejemplo (gotas de lluvia)

Para la función de densidad uniforme:

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

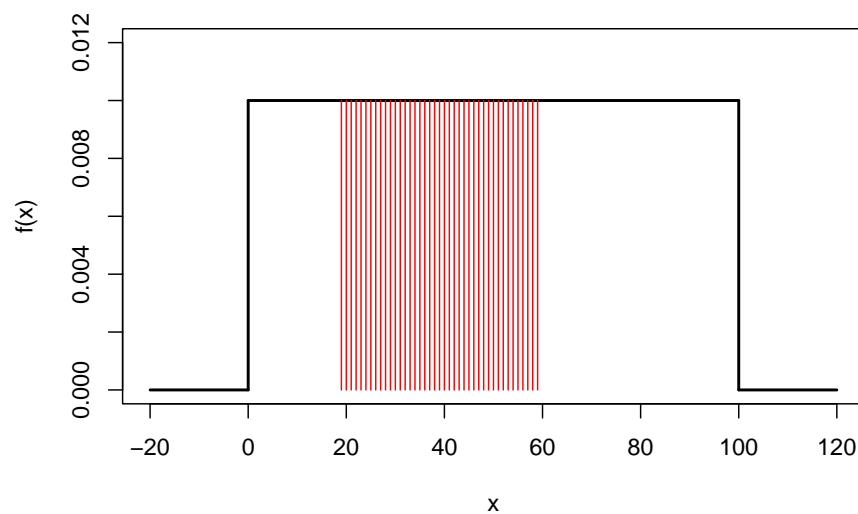
Encontramos que la distribución de probabilidad es

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

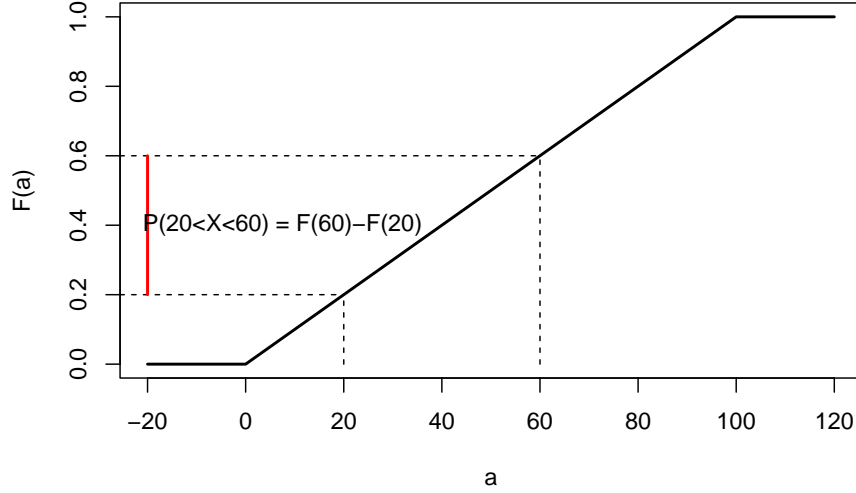
6.9 Gráficas de probabilidad

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

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



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



6.10 Media

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

Definición

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

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

Es la versión continua del centro de gravedad.

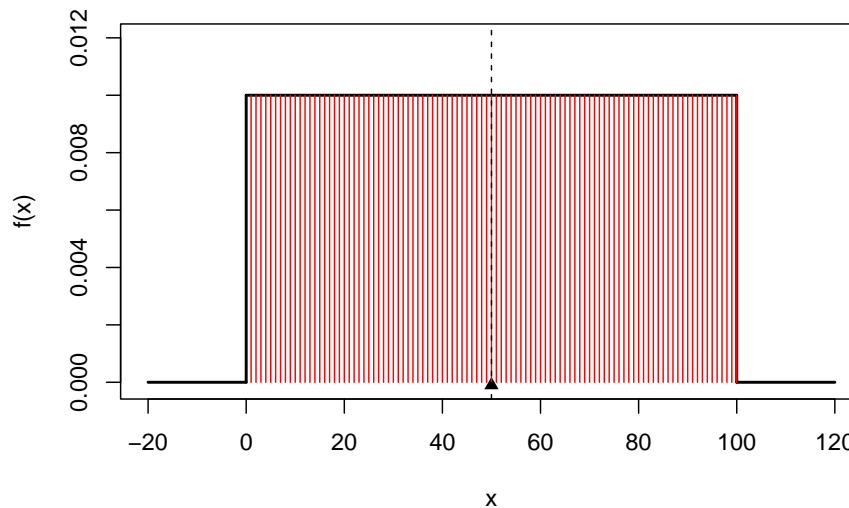
Ejemplo (gotas de lluvia)

La variable aleatoria con densidad de probabilidad

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

Tiene un valor esperado en

$$E(X) = 50$$



6.11 Varianza

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

Definición

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

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

Es la versión continua del momento de inercia.

6.12 Funciones de X

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

Definición

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

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

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

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

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

para a y b escalares.

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

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

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

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

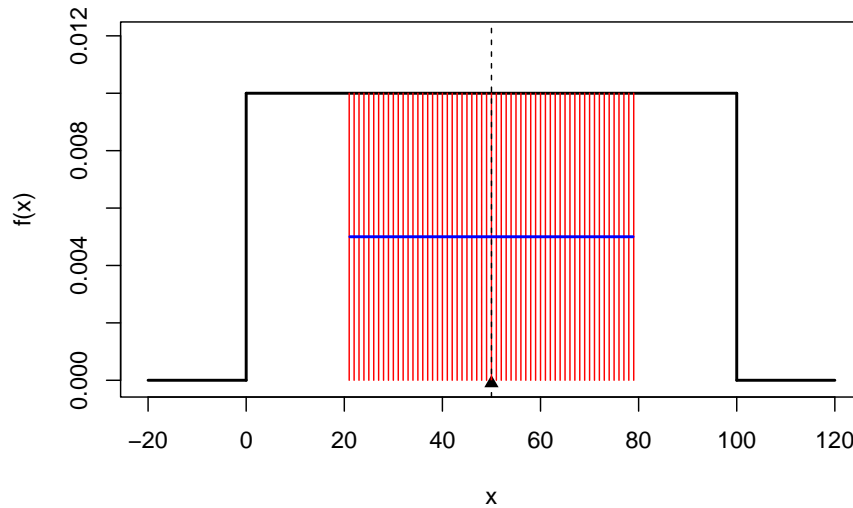
6.13 Ejercicios

6.13.0.1 Ejercicio 1

Para la densidad de probabilidad

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

- calcular la media (R:50)
- calcular la varianza usando $E(X^2) = V(X) + E(X)^2$ (R:100²/12)
- calcula $P(\mu - \sigma \leq X \leq \mu + \sigma)$ (R: 0.57)
- ¿Cuáles son el primer y tercer cuartiles? (R: 25; 75)

**6.13.0.2 Ejercicio 2**

Dado

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

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

6.13.0.3 Ejercicio 3

Para la densidad de probabilidad

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

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

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

6.13.0.4 Ejercicio 4

Dada la distribución acumulativa de una variable aleatoria X

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

calcular:

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

Chapter 7

Modelos de probabilidad para variables aleatorias discretas

7.1 Objetivo

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

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

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

7.2 Función de probabilidad

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

- 1) Nos permite calcular probabilidades para todos los resultados

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

- 2) Siempre es positiva:

$$f(x_i) \geq 0$$

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

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

Estudiamos dos **propiedades importantes**:

- 1) La media como medida de tendencia central:

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

- 2) La varianza como medida de dispersión:

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

7.3 Modelo de probabilidad

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

Ejemplos:

- 1) La función de masa de probabilidad definida por

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

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

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

7.4 Modelos paramétricos

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

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

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

Ejemplo (probabilidad clásica):

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

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

Sólo necesitamos saber M .

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

7.5 Distribución uniforme (un parámetro)

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

Definición

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

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

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

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

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

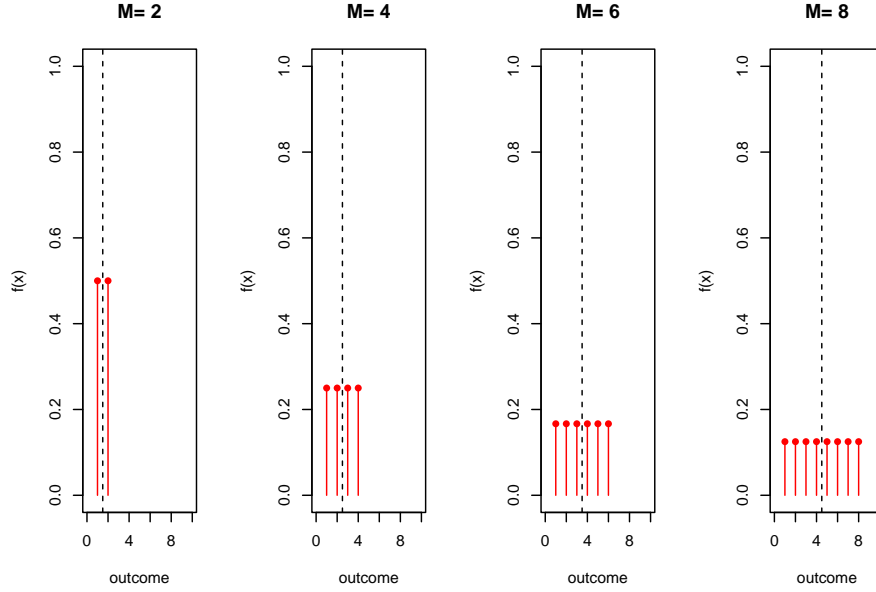
y

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

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

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

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



7.6 Distribución uniforme (dos parámetros)

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

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

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

porque $M = b - a + 1$.

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

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

Propiedades:

Si X se distribuye uniformemente entre a y b

$$X \rightarrow \text{Unif}(a, b)$$

1) Su media es

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

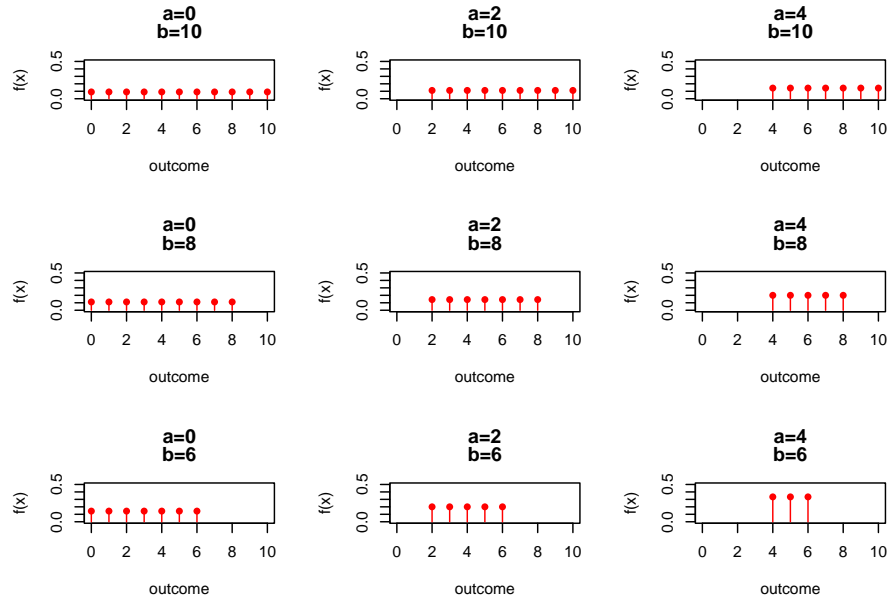
2) Su varianza es

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

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

Funciones de masa de probabilidad

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

**Ejemplo (clases escolares):**

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

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

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

eso es

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

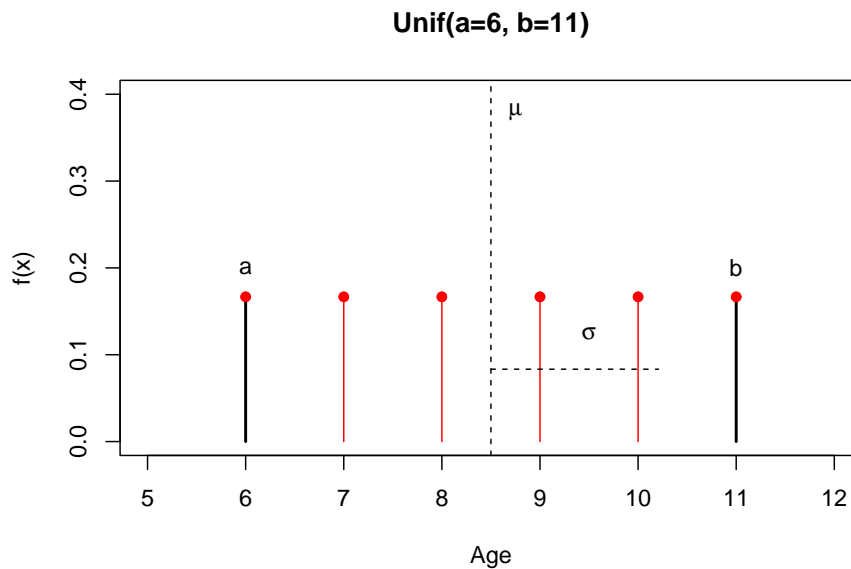
para $x \in \{6, 7, 8, 9, 10, 11\}$, y 0 en caso contrario.

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

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

Recuerda

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



Parámetros y Modelos:

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

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

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

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

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

7.7 ensayo de Bernoulli

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

Ejemplos:

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

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

Modelo de probabilidad:

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

- 1) Como una tabla de probabilidad:

<i>Resultado</i>	P_i
A	p
B	$1 - p$

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

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

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

- 3) Como una función de k

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

para $k = (0, 1)$.

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

$$K \rightarrow \text{Bernoulli}(p)$$

Propiedades:

Si K sigue una distribución de Bernoulli, entonces

- 1) su media es

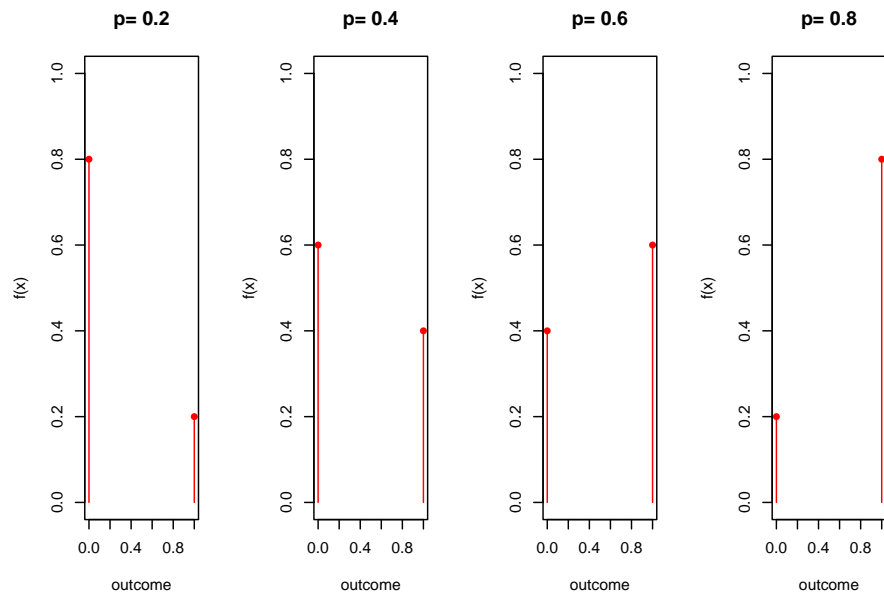
$$E(K) = p$$

- 2) su varianza es

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

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

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



7.8 Experimento binomial

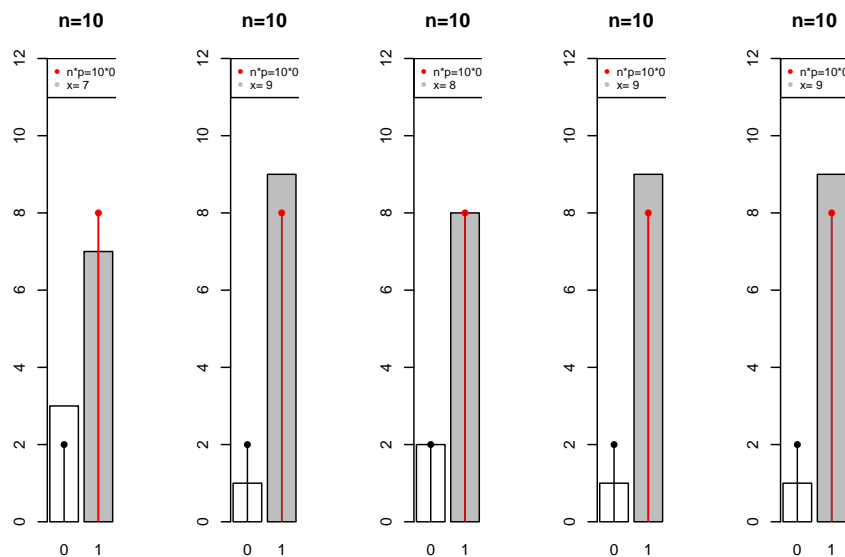
Cuando estamos interesados en predecir **frecuencias absolutas** cuando conocemos el parámetro p de un ensayo particular de Bernoulli, entonces

- 1) **repetimos** el ensayo de Bernoulli n veces y contamos cuantas veces obtuvimos A usando la frecuencia absoluta de A : N_A .
- 2) definimos una **variable aleatoria** $X = N_A$ tomando valores $x \in 0, 1, \dots, n$

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

Ejemplo (experimento binomial):

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



Ejemplos (Experimentos binomiales en contexto):

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

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

7.9 Función de probabilidad binomial

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

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

Ejemplo (transmisión de píxeles):

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

Consideremos que

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

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

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

- 2) Cada

$$K_i \rightarrow \text{Bernoulli}(p)$$

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

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

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

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

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

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

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

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

porque K_i son **independientes**.

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

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

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

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

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

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

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

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

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

o más en breve

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

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

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

Definición:

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

La función está dada por

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

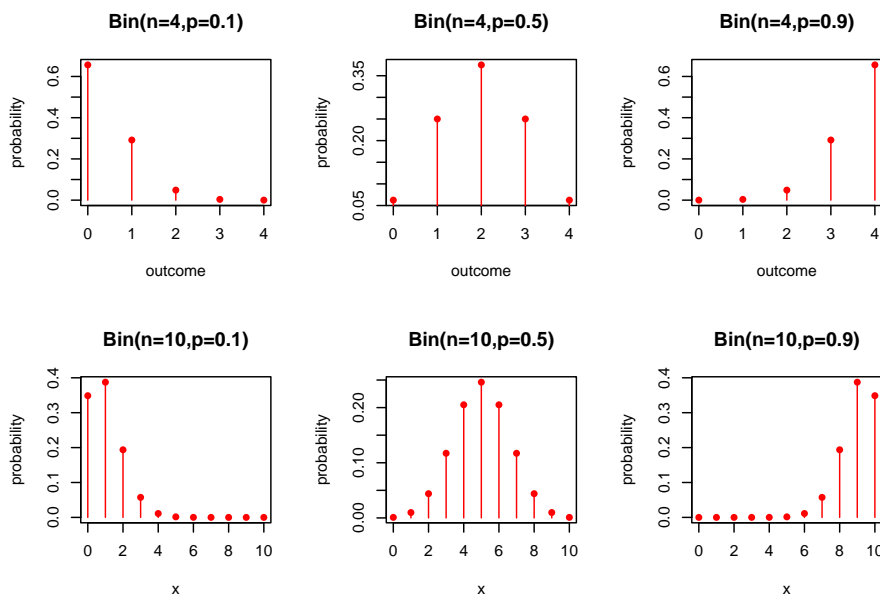
$\binom{n}{x} = \frac{n!}{x!(n-x)!}$ se denomina **coeficiente binomial** y da el número de formas en que se pueden obtener x eventos de tipo A en un conjunto de n .

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

$$X \rightarrow \text{Bin}(n, p)$$

donde n y p son parámetros.

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



Propiedades:

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

- 1) su media es

$$E(X) = np$$

- 2) su varianza es

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

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

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

y

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

Ejemplo (transmisión de píxeles):

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

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

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

Eso es

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

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

En R `dbinom(4,4,0.1)`

- ¿Cuál es la probabilidad de observar 2 errores?

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

En R `dbinom(2,4,0.1)`

Ejemplo (encuestas de opinión):

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

Para este caso

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

Eso es

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

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

en R `pbinom(8,10, 0.9)`

7.10 Función de probabilidad binomial negativa

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

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

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

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

Ejemplo (transmisión de píxeles):

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

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

$$(0, 0, 1, \dots, 0, 1, \dots, 0, 1)$$

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

La probabilidad de este evento es:

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

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

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

Ten en cuenta que

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

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

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

para $y = 0, 1, \dots$

Entonces decimos que Y sigue una distribución binomial negativa y escribimos

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

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

Propiedades:

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

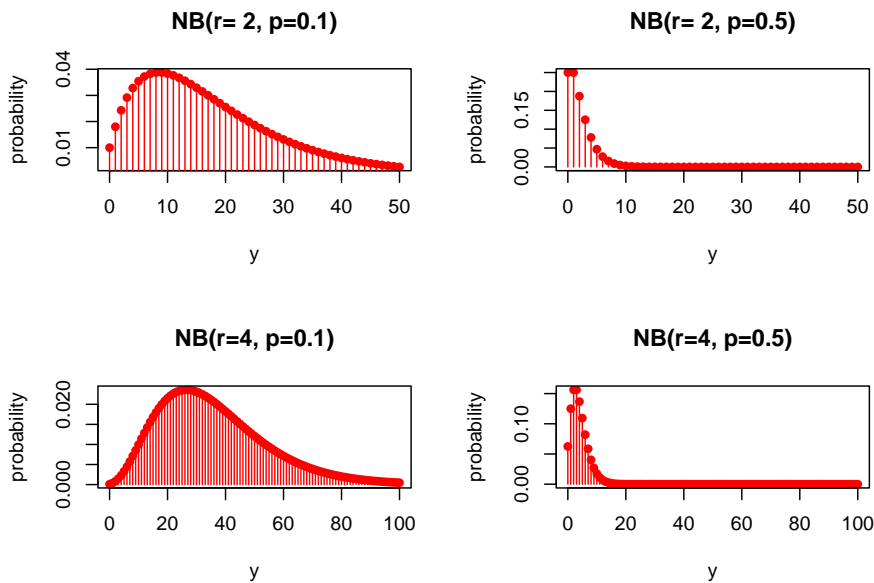
1) media

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

2) y varianza

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

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



Ejemplo (sitio web)

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

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

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

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

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

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

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

Ten en cuenta que en realidad hay pruebas de 6000.

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

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

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

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

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

$$\begin{aligned} F(5) &= P(Y \leq 5) = \sum_{y=0}^5 f(y) \\ &= \sum_{y=0}^5 \binom{y+2}{y} 0.0005^3 0.9995^y \\ &= \binom{2}{0} 0.0005^3 0.9995^0 + \binom{3}{1} 0.0005^3 0.9995^1 \\ &\quad + \binom{4}{2} 0.0005^3 0.9995^2 + \binom{5}{3} 0.0005^3 0.9995^3 \\ &\quad + \binom{6}{4} 0.0005^3 0.9995^4 + \binom{7}{5} 0.0005^3 0.9995^5 \\ &= 6.9 \times 10^{-9} \end{aligned}$$

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

Ejemplos

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

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

en R `dnbinom(10, 2, 0.1)`

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

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

en R `dnbinom(2, 4, 0.55)`

7.11 Distribución geométrica

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

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

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

$$Y \rightarrow \text{Geom}(p)$$

que tiene

1) media

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

2) y varianza

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

7.12 Modelo hipergeométrico

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

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

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

Ejemplo (varicela):

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

Definición:

La probabilidad de obtener x casos (tipo A) en una muestra de n extraída de una población de N donde K son casos (tipo A).

$$P(X = x) = P(\text{una muestra}) \times (\text{Número de formas de obteniendo } x)$$

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

donde $k \in \{\max(0, n + KN), \dots \min(K, n)\}$

Entonces decimos que X sigue una distribución hipergeométrica y escribimos

$$X \rightarrow \text{Hipergeom}(N, K, n)$$

El modelo hipergeométrico tiene tres parámetros.

Propiedades:

Si $X \rightarrow \text{Hypergeometric}(N, K, n)$ entonces tiene

1) media

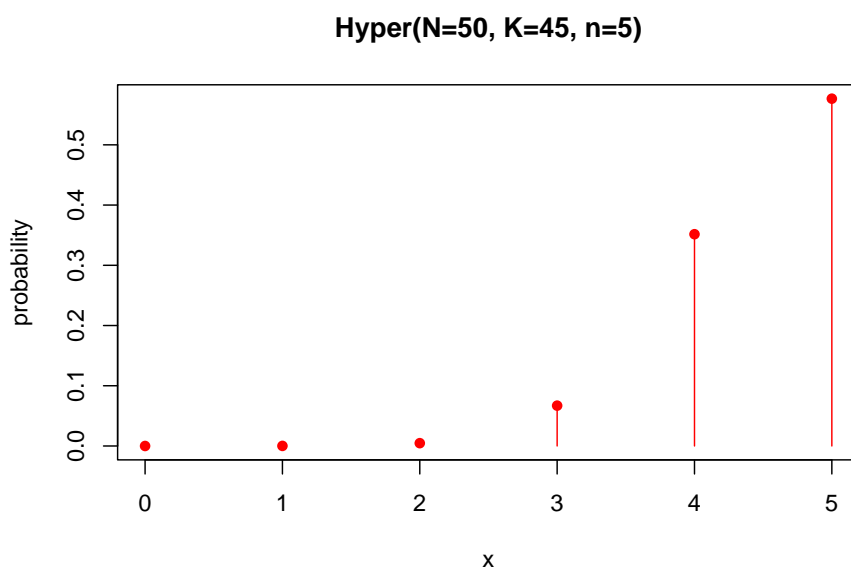
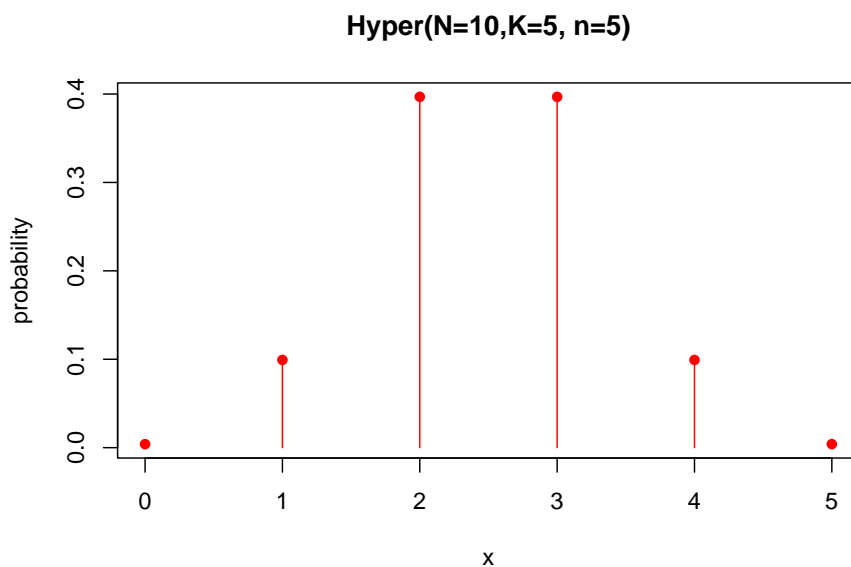
$$E(X) = n \frac{K}{N} = np$$

2) y varianza

$$V(X) = np(1-p) \frac{N-n}{N-1}$$

cuando $p = \frac{K}{N}$ es la proporción de casos (A) en una población de tamaño N . Ten en cuenta que cuando $N \rightarrow \infty$ recuperamos las propiedades binomiales.

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

**Ejemplo (varicela):**

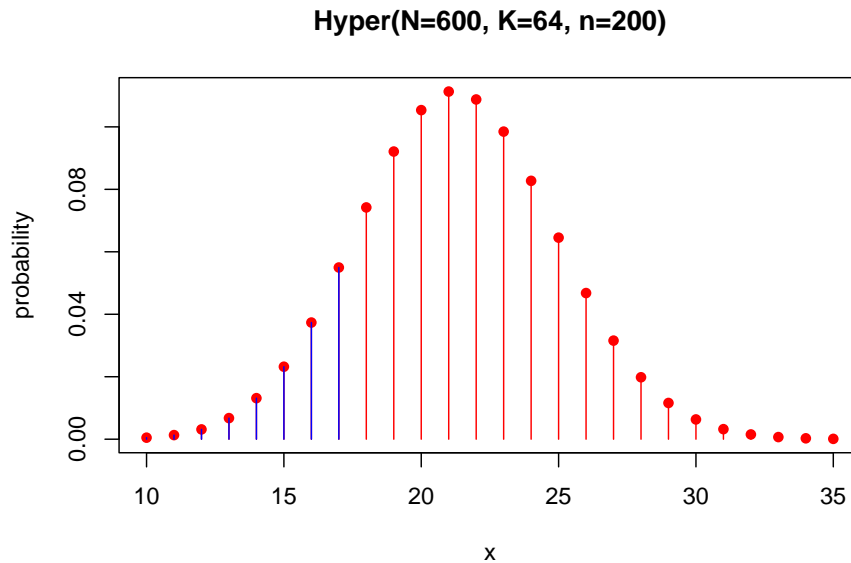
- ¿Cuál es la probabilidad ver como mucho 17 casos de varicela en una muestra de 200 alumnos de una escuela de 600 alumnos donde 64 están infectados?

La probabilidad que necesitamos calcular es $P(X \leq 17) = F(17)$

donde $X \rightarrow \text{Hypergeometric}(N = 600, K = 64, n = 200)$

en R `phyper(17, 64, 600-64, 200)=0.140565`

La solución es la adición de las agujas azules en el gráfico.



7.13 Preguntas

1) ¿Cuál es el valor esperado y la varianza del número de fallos en 100 prototipos, cuando la probabilidad de un fallo es de 0.25?

a: 0.25, 0.1875; **b:** 25, 0.1875; **c:** 0.25, 18.75; **d:** 25, 18.75

2) ¿Qué modelo de probabilidad describe mejor el número de mesas disponibles a la hora de la cena en un restaurante?

a: Binomial; **b:** Uniforme; **c:** Binomial negativo; **d:** Hipergeométrico

3) El valor esperado de una distribución Binomial no es

a: n veces el valor esperado de un Bernoulli; **b:** el valor esperado de un Hipergeométrico, cuando la población es muy grande; **c:** np ; **d:** el límite de la frecuencia relativa cuando el número de repeticiones es grande

4) Las encuestas de opinión para las elecciones de EE. UU. dan una probabilidad de 0.55 de que un votante esté a favor del partido republicano. Si realizamos

nuestra propia encuesta y preguntamos a 100 personas al azar en la calle, ¿cómo calcularías la probabilidad de que en nuestra encuesta los demócratas ganen las elecciones?

a: pbinom(x=49, n=100, p=0,55)=0,13; **b:** 1-pbinom(x=49, n=100, p=0,55)=0,86; **c:** pbinom(x=51, n=100, p=0,45)=0,90; **d:** 1-pbinom(x=51, n=100, p=0,45)=0,095

5) En un examen un alumno elige al azar una de las cuatro respuestas que no sabe. Si no sabe 10 preguntas, ¿cuál es la probabilidad de que al menos 5 preguntas sean correctas?

a: dbinom(x=5, n=10, p=0,25)~ 0,05; **b:** pbinom(x=5, n=10, p=0,75)~ 0,07; **c:** dbinom(x=5, n=10, p=0,75)~ 0,05; **d:** 1-pbinom(x=5, n=10, p=0,25)~ 0,02

7.14 Ejercicios

7.14.0.1 Ejercicio 1

Si el 25% de los tornillos producidos por una máquina son defectuosos, determina la probabilidad de que, de 5 tornillos elegidos al azar

- ningún tornillo sea defectuoso (R:0.2373)
- 1 tornillo sea defectuoso (R:0.3955)
- 2 tornillos sean defectuosos (R:0.2636)
- como máximo 2 tornillos sean defectuosos (R:0.8964)

7.14.0.2 Ejercicio 2

En una población, la probabilidad de que nazca un niño es $p = 0.51$. Considera una familia de 4 hijos.

- ¿Cuál es la probabilidad de que una familia tenga un solo niño? (R: 0.240)
- ¿Cuál es la probabilidad de que una familia tenga una sola niña? (R: 0.259)
- ¿Cuál es la probabilidad de que una familia tenga solo un niño o solo una niña? (R: 0.4999)
- ¿Cuál es la probabilidad de que la familia tenga por mucho dos niños? (R: 0.37470)
- ¿Cuál es la probabilidad de que la familia tenga al menos dos niños? (R: 0.7023)
- ¿Cuál es el mínimo número de hijos que debe tener una familia para que la probabilidad de tener al menos una niña sea mayor a 0.75? (R: $n = 3 \geq \log(0.25)/\log(0.51) = 2.05$)

7.14.0.3 Ejercicio 3

Un motor de búsqueda falla al recuperar información con una probabilidad de 0.1

- Si nuestro sistema recibe 50 solicitudes de búsqueda, ¿cuál es la probabilidad de que el sistema no responda a tres de ellas? (R: 0.1385651)
- ¿Cuál es la probabilidad de que el motor complete con éxito 15 búsquedas antes del primer fallo? (R:0.020)
- Consideramos que un motor de búsqueda funciona suficientemente bien cuando es capaz de encontrar información como mínimo para 10 solicitudes por cada 2 fallos. ¿Cuál es la probabilidad de que en un ensayo de fiabilidad nuestro buscador sea satisfactorio? (R: 0.659)