

Cap 1

Variables

↳ Numericas:

↳ Continuas: Toman cualquier valor en un rango.

↳ Discretas: No toman cualquier valor.

↳ Categoricals: Nivel

↳ Nominales: No hay orden natural.

↳ Ordinales: Hay un orden natural.

CAP 2

• Estadísticas descriptivas para datos Numéricos.

1. Media → promedio

→ Pto de equilibrio de la distribución.

→ \bar{x} media muestral

→ μ_x media poblacional

↳ `mean()`, (`na.rm = True` → para omitir datos faltantes)

2. Mediana

→ valor central

↳ `median()`

3. Moda → valor + frecuente

↳ unimodales, bimodales, multimodales.

↳ `mfvc()` (paquete `modeest`)

4. Medidas de dispersión

↳ varianza: `var()`

↳ Desviación estándar: `sd()` } semejanzas o diferencias entre las observaciones.

↳ Rango: `range()`, `min()`, `max()`

↳ Cuantil: `quantile()`

↳ Rango Intercuartil (IQR): + dispersión en los datos → + IQR

↳ `IQR()`

* Valores atípicos o outliers

↳ mediana y IQR

• Estadísticas descriptivas para datos categóricos.

1. frecuencia

2. Proporción

3. Tabla de contingencia → `table()` y `xtabs()`

↳ `marginsum()` → totales por filas.

↳ `addmargins()` → totales por filas y los agrega a la tabla.

Gráficos → ggpubr()

2

1 variable Numerica

→ Histograma: muestra la distribución de frecuencias (densidad).

↳ `gghistogram()`

→ De caja: - Puntos = datos atípicos

↓
`ggboxplot()`

- Línea horizontal → mediana

- extremos → inferior: primer cuartil

→ superior: tercer cuartil

- bigotes → límite 1.5 veces el IQR.

1 variable Categorica

→ De barras → `ggbarplot()`

→ cada barra es la proporción.

→ De torta → `ggpie()`

2 variables numericas

→ Dispersión: `ggscatter()` } Relación entre variables independientes: No hay tendencia

+ : /

- : \

2 variables categoricas

→ barras apiladas

→ barras agrupadas:

→ barras estandarizadas

↳ `geom-bar(position = "stack")`

↳ `geom-bar(position = "dodge")`

↳ `geom-bar(position = "fill")`

→ De mosaico: `ggplot() + geom-mosaic()`

Una variable numerica y categorica

→ de cajas: `ggboxplot()`

→ de Tiras: `ggstripchart()`

CAP 3

- Variables discretas: paquete → `discreteRV`

↳ `RV()`: Distribución (dato, prob)

↳ `E()`: valor esperado

↳ `V()`: varianza

↳ `SD()`: desviación estandar.

↳ `BofIID(X, n=?)`: crear vector según n de experimentos

1. Distribución normal
- ↳ unimodal
 - ↳ simétrica
- $dnorm()$: densidad
 $pnorm()$: función de distribución acumulada
 $qnorm()$: percentil para probabilidades dadas en p.
 $rnorm()$: genera aleatoriamente n observaciones según mean y sd.

↳ gráfico Q-Q

↳ permite visualizar

la distribución de las observaciones.

↳ $qqnorm()$

2. Distribución Z o Distribución normal estándar

↳ media = 0

↳ Dev. Estándar = 1

3. Distribución chi-cuadrado

↳ caracterizar valores positivos y habitualmente desviados a la derecha.

↳ grados de libertad: $n-1$

1. $dchisq()$

2. $pchisq()$

3. $qchisq()$

4. $rchisq()$

4. Distribución t de Student

↳ muestras pequeñas

↳ ↑ grados de libertad → ↑ cercano a normal.

1. $dt()$

2. $pt()$

3. $qt()$

4. $rt()$

5. Distribución F

↳ Comparar varianzas.

1. $df()$

2. $pf()$

3. $qf()$

4. $rf()$

Distribuciones Discretas

1. Bernoulli

↳ éxito = 1 = p

↳ fracaso = $1-p = q$

↳ \hat{p} (proporción de muestras) = cantidad de p / cantidad de intentos

↳ paquete extraDistr

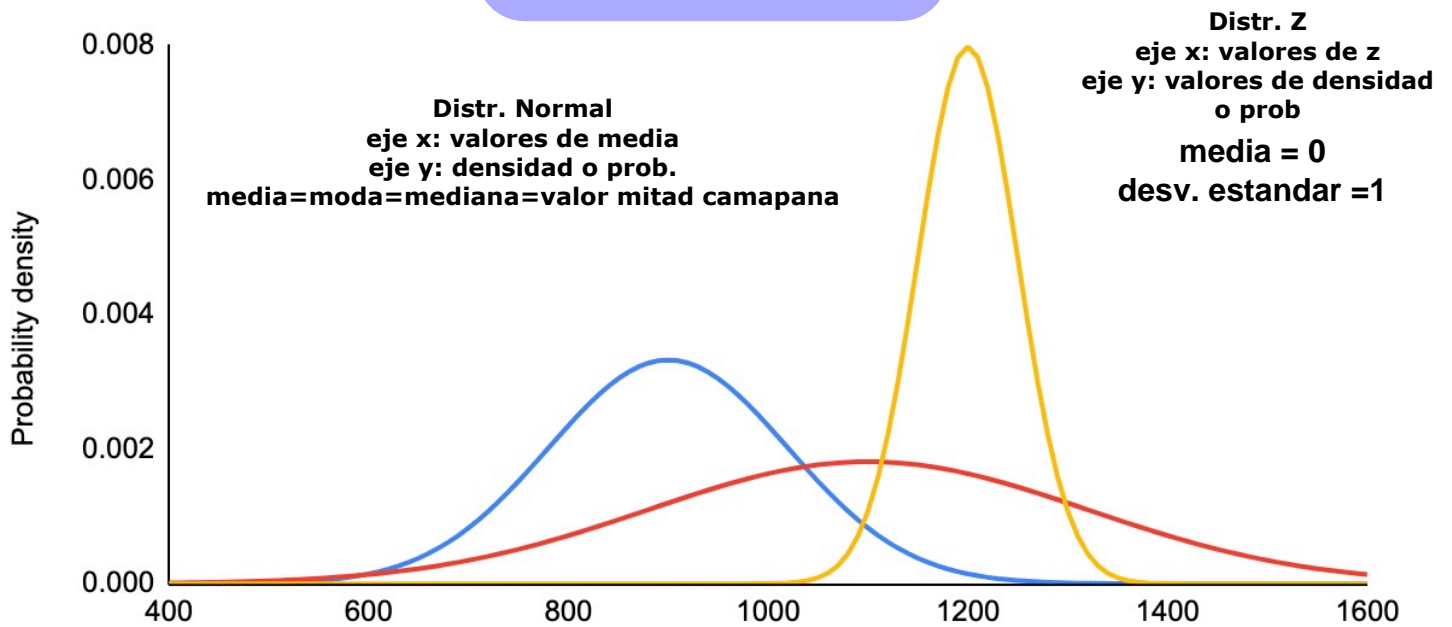
1. $dbern()$

3. $gbern()$

2. $pbern()$

4. $rbern()$

Normal distributions



2 Geométrica \rightarrow # intentos hasta obtener un éxito.

- 1 \rightarrow `dgeom()`
- 2 \rightarrow `pgeom()`
- 3 \rightarrow `qgeom()`
- 4 \rightarrow `rgeom()`

3 Binomial \rightarrow k éxitos en n intentos

- 1 \rightarrow intentos son independientes.
- 2 \rightarrow " son fijos
- 3 \rightarrow cada resultado se puede clasificar como éxito o fracaso.
- 4 \rightarrow p es = para cada intento.

- 1 `dbinom()`
- 2 `pbinom()`
- 3 `qbinom()`
- 4 `rbinom()`

4 Binomial Negativa \rightarrow k -ésimo éxito al n -ésimo intento.

- condiciones: 1 *

- 2 ***
- 3 ****
- 4 último intento es un éxito.

- 1 \rightarrow `dnbinom()`
- 2 \rightarrow `pnbinom()`
- 3 \rightarrow `qnbinom()`
- 4 \rightarrow `rnbinom()`

5 Poisson \rightarrow cantidad de eventos en una población grande en un lapso de tiempo dado.

- 1 \rightarrow `dpois()`
- 2 \rightarrow `ppois()`
- 3 \rightarrow `qpois()`
- 4 \rightarrow `rpois()`

CAP 4

1 Estimadores puntuales \rightarrow parametro \Rightarrow población
 \rightarrow Estadístico \Rightarrow muestra

2 Error Estandar: σ / \sqrt{n}

- \rightarrow obs. son independientes
 - \rightarrow $n \geq 30$
 - \rightarrow No es muy asimétrica
- } modelo normal

3 Pruebas de hipótesis

- \rightarrow Hipótesis nula (H_0)
- \rightarrow Hipótesis alternativa (H_A)

- Error tipo I = rechazar H_0 cuando es verdadera
- Error tipo II = Aceptar H_0 cuando es falsa y H_A verdadera.

4. valor de P : Cuantificar con fuerte es la evidencia en contra H_0 .

$P < \alpha \rightarrow$ se rechaza H_0

CAP 5

* valor p \rightarrow $pt()$

1. Prueba Z: $z.test()$

- $\rightarrow n \geq 30$
- $\rightarrow n < 30$, varianza de la población.
- \rightarrow observaciones son independientes.
- \rightarrow población sigue una distribución normal. \rightarrow grafico Q-Q

Shapiro Wilk
($shapiro.test()$)

2. Prueba T de Student

2.1: 1 muestra

- \rightarrow observaciones indepen.
- \rightarrow Distribución cercana a la normal.

$t.test(x, ...)$

2.2: 2 muestras pareadas

$\rightarrow t.test(x, y, ...)$

2.3: 2 muestras independientes.

Diferencia de medios: $\mu_1 - \mu_2 = d_0$

- \rightarrow cada muestra cumple con la condición
- \rightarrow Son independientes entre si (muestras).

CAP 6

- $\alpha \rightarrow$ error tipo I
- $\beta \rightarrow$ error tipo II

* poder estadístico $= 1 - \beta \rightarrow$ Permite determinar el tam. del muestra para un cierto tamaño de efecto. probabilidad de correctamente rechazar H_0 cuando es falsa

* tamaño del efecto \rightarrow Cuantificación de la diferencia entre 2 grupos

- \uparrow poder si \uparrow tamaño
- \downarrow tamaño, el poder se acerca al α .

* $power.t.test(n, delta, sd, sig.level, type, alternative)$ \rightarrow power

• Tamaño del efecto

\rightarrow d de cohen: $d = \frac{\bar{x} - \mu_0}{s}$

• Poder

\rightarrow $power.t.test(n, d, sig.level, power, type, alternative)$

CAP 7

Metodo de Wald

→ obs. son independientes

→ Condición de éxito-fracaso:

- min 10 obs de éxito
- min 10 obs de fracaso.

→ $np \geq 10$

→ $n(1-p) \geq 10$

Para una proporción

→ intervalo de confianza

$$SE = \sqrt{\frac{p(1-p)}{n}}$$

$$\text{inferior} = p - z.\text{critico} \cdot SE$$

$$\text{superior} = p + z.\text{critico} \cdot SE$$

$$z.\text{critico} =$$

$$\rightarrow qnorm(\alpha/2, \text{lower.tail} = \text{FALSE})$$

\hat{p} = valor nulo
 p = p éxito

→ Hipotesis

$$\text{error} = \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

$$Z = \frac{p - \hat{p}}{\text{error}}$$

$$p = pnorm(z, \text{lower.tail} = \text{FALSE})$$

Para 2 proporciones.

$$\hat{p} = p_1 - p_2$$

→ cada una tiene modelo normal

→ Son independientes.

$$\text{error}p_1 = (p_1 \times (1 - p_1)) / n_1$$

$$\text{error}p_2 = (p_2 \times (1 - p_2)) / n_2$$

$$SE = \sqrt{\text{error}p_1 + \text{error}p_2}$$

$$z.\text{critico} = qnorm(\alpha/2, \text{lower.tail} = \text{FALSE})$$

$$\text{inferior} = \hat{p} - z.\text{critico} \times SE$$

$$\text{superior} = \hat{p} + z.\text{critico} \times SE$$

$$p.\text{Agrup} = (p_1 + p_2) / (n_1 + n_2)$$

$$\text{error}_1 = p.\text{Agrup} (1 - p.\text{Agrup}) / n_1$$

$$\text{error}_2 = p.\text{Agrup} (1 - p.\text{Agrup}) / n_2$$

$$\text{error.E6} = \sqrt{\text{error}_1 + \text{error}_2}$$

$$Z = \hat{p} - \text{valor-nulo} / \text{error.E6}$$

$$p = 2 \times pnorm(Z, \text{lower.tail} = \text{FALSE})$$

2 Metodo de Wilson

↳ `prop.test(x, n, p, alternative, conf.level, ...)`

- x = nº de exitos
- n = tam muestra
- p = valor nulo

3 Poder y Pruebas de proporciones.

↳ `power.prop.test(n, p1, p2, sig.level, power, alternative)`

- n = nº de obs.

↳ `pwr.p.test(h, n, sig.level, power, alternative)` → única proporción

↳ `pwr.2p.test(h, n, sig.level, power, alternative)` → 2 proporciones = tam muestra.

↳ `pwr.2p2n.test(h, n1, n2, sig.level, power, alternative)`

↳ 2 proporciones y tam muestra \neq

→ tam. del efecto: paquete (pwr)

↳ `ES.h(p1, p2)`

→ `bbomsize(p1, p2, fraction, alpha, power)`

↳ paquete: Hmisc

↳ fraction: $n_1 / (n_1 + n_2)$

CAP 8

1 Prueba chi-cuadrado de Pearson

↳ Obs independientes entre si.

↳ Al menos 5 observaciones esperadas en cada grupo.

-1.1: Homogeneidad: 2 poblaciones presentan las mismas proporciones en los diferentes niveles de una variable categorica.

↳ `chisq.test()`

-1.2: Bondad de Ajuste: -permite comprobar si una distribución de frecuencias observada se asemeja a una distribución esperada.
- Comprobar si una muestra es representativa de la población.

2

Prueba exacta de Fisher (muestras pequeñas)

↳ alternativa a independencia.

• H_0 : variables indep.

• H_A : variables relacionadas

↳ `fisher.test(x, conf.level)` : x tabla de contingencia.

3

Prueba de McNemar (muestras pequeñas)

↳ Una misma característica, con respuesta dicotómica, se mide en 2 ocasiones \neq para los = sujetos (muestras pareadas) y queremos determinar si se produce o no un cambio significativo entre ambas mediciones.

- H_0 : no hay cambios significativos en las muestras.

- H_A : Si " " " " " " " " " " " "

↳ `mcnemar.test(tabla)`

↳ obtener p \rightarrow `pchisq()`

4

Prueba Q de Cochran

↳ la variable de respuesta es dicotómica y la variable independiente tiene + de 2 observaciones pareadas (cuando ambas son dicotómicas, es equivalente a McNemar).

- H_0 : proporción de éxitos es la = para todos los grupos

- H_A : " " " " " " " " " " " "

• Condiciones

↳ variable respuesta es dicotómica

↳ " indep. es categórica

↳ obs con indep.

↳ $n \cdot k \geq 24$, k = cantidad de niveles en la variable independiente

↳ `Cochran.q.test(formula, data, alpha)` / paquete: R Videllemire