Capitulo 2: Se tiene la definición de Estimador puntual, distribución de frecuencia.
Capitulo 3:Variable aleatoria, valor esperado.
Capitulo 4: Se tiene la definición de intervalos de confianza,error estándar, nivel de significación, ley de los grandes números, Modelos estadisticos, estadístico de prueba.
Capitulo 6: Se tiene la definición de poder, tamaño de efecto.
------------------------------- ----------------------------------------

Rbind(dataframe, tabla) sirve para añadir a un dataframe o matriz algo.

Si obtengo información de un csv y necesito obtener una muestra, solo debo realizar un sample_n (datos csv, cantidad de la muestra)

**Muestra<- sample_n(datos, n)**

Otro método.

```
set.seed(random)
muestra_N <- sample(datos[["Id"]], N, replace = FALSE, prob = NULL)
x_datos <- t(datos %>% select("x"))
y_datos <- t(datos %>% select("y"))

i = 1
x <- c()
y <- c()

while(i <= N){
  pos <- muestra_N[i]
  x[i] <- x_datos[pos]
  y[i] <- y_datos[pos]
  i=i+1
}
new_datos <- data.frame(muestra_N,x,y)
```

**Resumen formulas por capítulos:**

**Capitulo 1:**

(a) Texto plano delimitado por tabulaciones.
datos2 <- read.delim ( file . choose ())
(b) Valores separados por comas (inglés).
datos3 <- read .csv("C:\\ Inferencia \\ ejemplo1 -csv -eng.csv ")
(c) Valores separados por punto y comas (español).
datos4 <- read . csv2 (" ejemplo1 -csv -esp. csv")

**Capitulo 2:**

```
# Cálculo de varias medidas para la variable Potencia.
medidas_potencia <- datos %>% summarise(Media = mean(Potencia),
                                        Mediana = median(Potencia),
                                        Varianza = var(Potencia),
                                        IQR = IQR(Potencia))
```

**Graficos:**

**Una variable numérica:** histograma pag 24. Pag 25 grafico de cajas, cuantiles, bigotes,etc.
**Una variable categorica:** Grafico de barras pag 26 y grafico de tortas pag 27
**Dos variables numéricas:** Grafico de dispersión pag 28
**Dos variables categóricas:** Grafico de barras apiladas, agrupadas y estandarizadas. Pag 30
**Una variable numérica y otra categorica:** Grafico de cajas ‘pag 33 ografico de tiras.

## Capitulo 3:

**Variables aleatorias**
**-Calcular el valor esperado .**
**esperado <- E(X)**
**cat (" Valor esperado :", esperado , "\n")**
**-Calcular la varianza .**
**varianza <- V(X)**
**cat (" Varianza :", varianza , "\n")**
**-Calcular la desviaci ón está ndar .**
**desviacion <- SD(X)**
**cat (" Desviaci ón está ndar :", desviacion , "\n")**

## Distribucion Normal: Pag: 42:

**Es importante señalar que, por defecto, lower.tail toma el valor verdadero, con lo que pnorm() y qnorm() operan con la cola inferior de la distribución. Si, en cambio, lower.tail = FALSE, dichas funciones operan con la cola superior (es decir, pnorm() nos entrega la probabilidad de que la variable tome valores mayores que un valor dado).**

## Grafico Q-Q: Permite observar si los datos siguen la normal, mientras los puntos sigan a la recta mas parecida sera la distribución a la normal. PAG 43 Y 44.

```
library(ggpubr)

# Cargar datos.
datos <- read.csv2("C:/Inferencia/Mtcars.csv", stringsAsFactors = TRUE,
                   row.names = 1)

# Gráfico Q-Q para la variable Rendimiento.
g <- ggqqplot(datos,
              x = "Rendimiento",
              color = "red")
```

## Distribucion Z: pag 44 y 45

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

Su valor z, que determina cuán por encima o por debajo de la media (en términos de la desviación estándar) se encuentra dicha observación x. Así, observaciones cuyos valores z sean negativos estarán por debajo de la media. Análogamente, un valor Z positivo indica que la observación está por sobre la media. Mientras mayor sea el valor absoluto de su valor z (|z|), más inusual será la observación.

## Distribucion Chi cuadrado: pag 45

**Se usa para caracterizar valores siempre positivos y habitualmente desviados a la derecha.**
**El único parámetro de esta distribución corresponde a los grados de libertad.**
**dchisq (x,df)**
**pchisq (q,df,lower.tail)**
**qchisq (p,df,lower.tail)**
**rchisq (n,df)**

- `x, q` son vectores de cuantiles (enteros no negativos).
- `p` es un vector de probabilidades.
- `n` es la cantidad de observaciones.
- `df` son los grados de libertad.
- `lower.tail` es análogo al de la función `pnorm`.

## Distribucion T de Student: pag 45

Se emplea con muestras pequeñas y se usan los grados de libertar como único parámetro.
Mientras aumenten los grados la distribución mas se asemeja a la normal.

Dt(x,df)
Pt(q,df,lower.tail)
Qt(p´,df,lower.tail)
Rt(n,df)

**Distrigucion F : pag 47**
**Usada para comparar varianzas.**
**Se requiere dos distribuciones X^2 y dos grados de libertad respectivamente.**
Dt(x,df1,df2)
Pt(q, df1,df2,lower.tail)
Qt(p´, df1,df2,lower.tail)
Rt(n, df1,df2)

## Distribuciones discretas: pag 48

## Bernoulli:

Una variable aleatoria de Bernoulli es aquella en que cada intento individual tiene solo dos resultados posibles: “éxito”, que ocurre con una probabilidad p y se representa habitualmente con un 1, y “fracaso”, que ocurre con probabilidad q = 1- p y suele representarse por un 0.
Se requiere que los datos sean independientes.
Paquete extraDistr ofrece 4 funciones para la distribución
Dbern(x,prob)
Pbern(q,prob,lower.tail)
Qbern(p,prob,lower.tail)
Rbern(n,prob)

## Distribucion Geometrica: pag 48

Describe la cantidad de intentos que debemos realizar hasta obtener un éxito para variables de Bernoulli independientes e idénticamente distribuidas, es decir, que no se afectan unas a otras y cada una con igual probabilidad de éxito.
Dgeom(x,prob)

## Distribución binomial: pag 49-50

Describe la probabilidad de tener exactamente k éxitos en n intentos independientes de Bernoulli con probabilidad de éxito p.

ANTES DE DECIDIR UTILIZAR LA DISTRIBUCION BINOMIAL SE DEBEN SABER 4 PUNTOS:
1. Los intentos son independientes.
2. La cantidad de intentos (n) es fija.
3. El resultado de cada intento puede ser clasificado como éxito o fracaso.
4. La probabilidad de éxito (p) es la misma para cada intento.
Dbinom(x,size,prob)

- x es un vector numérico.
- p es un vector de probabilidades.
- n es la cantidad de observaciones.
- size corresponde al número de intentos.
- prob es la probabilidad de éxito de cada intento.

## Distribución binomial Negativa: pag 51

Describe la probabilidad de encontrar el k-ésimo éxito al n-ésimo intento.
En el caso binomial, en general se tiene una cantidad fija de intentos y se considera la cantidad de éxitos. En el caso binomial negativo, se examina cuántos intentos se necesitan para observar una cantidad fija de éxitos y se requiere que la última observación sea un éxito.

1. Los intentos son independientes.
2. El resultado de cada intento puede ser clasificado como éxito o fracaso.
3. La probabilidad de éxito (p) es la misma para cada intento.
4. El último intento debe ser un éxito.
dnbinom(x,size,prob)

Donde:

- x, q son vectores de cuantiles (enteros no negativos).
- p es un vector de probabilidades.
- n es la cantidad de observaciones.
- lambda es un vector no negativo de medias.
- lower.tail es análogo al de la función pnorm.

## Distribucion de Poisson: Pag 52

Útil para estimar la cantidad de eventos en una población grande en un lapso de tiempo dado, por ejemplo, la cantidad de contagios de influenza entre los habitantes de Santiago en una semana.
dpois (x,labmda)

# Capitulo 4:

**Prueba bilateral: pag 60**

H0: El nuevo sistema, en promedio, tarda lo mismo que el antiguo en procesar las transacciones, es decir: muN =muA.

HA: Los sistemas requieren, en promedio, cantidades de tiempo diferentes para procesar las transacciones, es decir: muN =!muA TIENE DOS COLAS MUN<MUA Y MUN>MUA

**Prueba unilateral: pag 60**

H0: El nuevo sistema tarda, en promedio, lo mismo que el antiguo en procesar las transacciones, es decir: muN = muA.

HA: El nuevo sistema tarda, en promedio, menos que el antiguo en procesar las transacciones, es decir: muN < muA TIENJE SOLO UNA COLA.

z* qnorm

El error tipo I corresponde a rechazar H0 cuando en realidad es verdadera, mientras que el error tipo II corresponde a no rechazarla cuando en realidad HA es verdadera.

**Prueba formal de hipotesis con valores P: pag 62**

Cuanto menor sea el valor p, más fuerte será la evidencia en favor de HA por sobre H0. Y aquí la ventaja de usar este método para decidir: el valor p se puede comparar directamente con el nivel de significación _, y si p es menor que el nivel de significación se considera evidencia suficiente para rechazar la hipótesis nula en favor de la hipótesis alternativa. En este ejemplo, p = 0; 040 < _ = 0; 05, por lo que se rechaza H0 en favor de HA. Pero como se dijo cuando usamos intervalos de confianza, el valor p está cerca del valor _ y convendría ser menos tajante en la decisión y evaluar la posibilidad de ampliar la muestra para conseguir evidencia más definitiva.

IMPORTANTE ESTO:

Como p > alfa, se falla en rechazar la hipótesis nula en favor de la hipótesis alternativa, por lo que se concluye con un 95% de confianza. Se acepta H0.

Como p < alfa, se rechaza la hipótesis nula en favor de la hipótesis alternativa, por lo que se concluye con un 95% de confianza. Se acepta Ha.

## El efecto del nivel de significación pag 66

**Si resulta costoso o peligroso cometer un error de este tipo, debemos requerir evidencia más fuerte para rechazar la hipótesis nula (es decir, reducir la probabilidad de que esto ocurra), lo que podemos lograr usando un valor más pequeño para el nivel de significación, por ejemplo, alfa = 0.01. Sin embargo, esto necesariamente aumentará la probabilidad de cometer un error de tipo II. Si, por el contrario, el costo o el peligro de cometer un error de tipo II (no rechazar H0 cuando en realidad HA es verdadera) es mayor, debemos escoger un nivel de significación más elevado (por ejemplo, alfa = 0,10).**

## Estimadores puntuales con distribución cercana a la normal pag 67

Prueba de hipótesis usando el modelo normal:

1. Formular la hipótesis nula (H0) y alternativa (HA) en lenguaje llano y luego en notación matemática.
2. Identificar un estimador puntual (estadístico) adecuado e insesgado para el parámetro de interés.
3. Verificar las condiciones para garantizar que la estimación del error estándar sea razonable y que la distribución muestral del estimador puntual siga aproximadamente una distribución normal.
4. Calcular el error estándar. Luego, graficar la distribución muestral del estadístico bajo el supuesto de que H0 es verdadera y sombrear las áreas que representan el valor p.
5. Usando el gráfico y el modelo normal, calcular el valor p para evaluar las hipótesis y escribir la conclusión en lenguaje llano.

**ESTIMADORES CON OTRAS DISTRIBUCIONES PAG 67**

# CAPITULO 5

**PRUEBA Z: PAG 71: ADECUADA PARA INFERIR ACERCA DE LAS MEDIAS CON UNA O DOS MUESTRAS.**
**Condiciones:**
**• La muestra debe tener al menos 30 observaciones. Si la muestra tiene menos de 30 observaciones, se debe conocer la varianza de la población.**
**• Las observaciones deben ser independientes, es decir que la elección de una observación para la muestra no influye en la selección de las otras.**
**• La población de donde se obtuvo la muestra sigue aproximadamente una distribución normal. Con la prueba Q-Q o shapiro.test(x).**
**DONDE X ES LA MUESTRA. Si p obtenido en shpairo es mayor a nivel de significacancia sigue la normal pero si es menor no sigue.**

$$Z = \frac{\overline{x} - \mu}{\sigma} = \frac{26,066 - 20}{2,32} = 2.6147$$

2 * pnorm(2.6147, lower.tail = FALSE), al ser bilateral se multiplica el valor por dos.
z.test(x,mu, stdev, alternative, conf.level)

**Donde x= vector con observaciones, mu es el valor nulo , stdev desviación estándar de la población, alternative = que tipo de prueba es (depende de la hipotesis) y conf.level es el 1- alfa 0.05**

**desviacion <- SD(X)**
**cat (" Desviaci ón está ndar :", desviacion , "\n")**
**ADECUADA PARA ASEGURAR O DESCARTAR QUE LA MEDIA DE LA POBLACION TIENE UN CIERTO VALOR HIPOTETICO**
**Ejemplo de pregunta:**

*En un estudio acerca de la sintomatología de pacientes con problemas cardíacos, los investigadores estimaron que la edad promedio de los pacientes diagnosticados como enfermos es de 60,5 años.* **y los del ejemplo.**

**Sirve utilizer prueba Z o te de student Comunmente se usa mas la t de student.**
**PRUEBA T DE STUDEN: PAG 74-78:**
La prueba t de Student, basada en la distribución t, es en consecuencia la alternativa más ampliamente empleada para inferir acerca de una o dos medias muestrales.

## Prueba t para una muestra:
## VERIFICAR:
0. No tiene un mínimo de observaciones.
1. Las observaciones son independientes entre sí.
2. Las observaciones provienen de una distribución cercana a la normal. (GRAFICO Q-Q O SHAPIRO.TEST)

t.test(x,alternative, mu, conf.level)
DONDE: X = MUESTRA , ALTERNATIVE TIPO DE HIPOTESIS, MU VALOR NULO Y CONF.LEVEL 1-ALFA

## Prueba t para DOS MUESTRAS PAREADAS pag 78:

Se tienen dos muestras del mismo tamaño,
Se les entregar las dos muestras avisar que estan pareadas, alternative depende de que tipo de hipotesis se tenga, valor nulo es el de la hipotesis.
prueba _2 <- t. test (x = t_A, y = t_B, paired = TRUE , alternative = " two. sided ", mu = valor _nulo , conf . level = 1 - alfa )

## Prueba t para dos muestras independientes: pag 80:

la prueba t se usa para comparar las medias de dos poblaciones en que las observaciones con que se cuenta no tienen relación con ninguna de las otras observaciones, ni influyen en su selección, ni en la misma ni en la otra muestra.

```
prueba <- t. test (x = vacuna _A,
                   y = vacuna _B,
                   paired = FALSE ,
                   alternative = " greater ",
                   mu = 0,
                   conf . level = 1 - alfa )
```

1. Cada muestra cumple las condiciones para usar la distribución t.
2. Las muestras son independientes entre sí.

Van las muestras, paired falso porque no son pareadas alternative ver hipótesis mu =

## CAPITULO 6:

Error tipo I: rechazar H0 en favor de HA cuando H0 es en realidad verdadera.
Error tipo II: no rechazar H0 en favor de HA cuando HA es en realidad verdadera.
Poder estadístico = 1-beta que es la probabilidad de correctamente rechazar H0 cuando es falsa.

• El poder de la prueba aumenta mientras mayor es el tamaño del efecto (en este caso, la distancia entre el valor nulo y la media de la muestra).
• A medida que el tamaño del efecto disminuye (es decir, el estimador se acerca al valor nulo), el poder se aproxima al nivel de significación.
• Usar un valor de alfa más exigente (menor), manteniendo constante el tamaño de la muestra, hace que la curva de poder sea más baja para cualquier tamaño del efecto (lo que verifica la relación entre alfa y beta).
• Usar una muestra más grande aumenta el poder de la prueba para cualquier tamaño del efecto distinto de 0.

**Tamaño del efecto pag 88:**

**D de cohen** $$d = \frac{\overline{x} - \mu_0}{s}$$ **donde**

- $\overline{x}$: media muestral.
- $\mu_0$: media teórica para el contraste (valor nulo).
- $s$: desviación estándar de la muestra con $n-1$ grados de libertad.

**Calculo del poder en R: pag 94:**
**power.t.test(n, delta, sd, sig.level, power, type, alternative)**

- `n`: tamaño de la muestra (por cada grupo, si corresponde).
- `delta`: diferencia observada entre las medias, o entre la media muestral y el valor nulo, no estandarizada.
- `sd`: desviación estándar observada.
- `sig.level`: nivel de significación.
- `power`: poder de la prueba.
- `type`: tipo de prueba t de Student (“`two.sample`” para diferencia de medias, “`one.sample`” para una sola muestra o “`paired`” para dos muestras pareadas).
- `alternative`: tipo de hipótesis alternativa (“`one.sided`” si es unilateral, “`two.sided`” si es bilateral).

para usar la función power.t.test() siempre debemos señalar el tipo de prueba t con el que estamos trabajando y si la hipótesis alternativa es de una o dos colas. Esta función nos permite calcular cualquiera de los demás argumentos (tamaño de la muestra, tamaño del efecto, desviación estándar, nivel de significación o poder estadístico) para la prueba en cuestión a partir de los 4 argumentos restantes. Así, al argumento que queremos calcular se le asigna el valor NULL en la llamada.

Delta aveces puede tomar el valor de d de cohen.
d_cohen_4 <- (media_pregunta2-media)/desviacion_estandar
¿cuántos bidones deberían revisarse para conseguir un poder estadístico de 0,8 y un nivel de significación de 0,05?
Se utiliza d de cohen para utilizar el poder y encontrar las n bidones.

# Capitulo 7: inferencia con proporciones muestrales:

Métodos de Wald y de Wilson para inferir acerca de una y dos proporciones.

**MÉTODO DE WALD pag 99:**

cuando queremos responder preguntas del tipo “¿qué proporción de la ciudadanía apoya al gobierno actual?”

Condiciones:

1. Las observaciones de la muestra son independientes.
2. Se cumple la condición de éxito-fracaso, que establece que se espera observar al menos 10 observaciones correspondientes a éxito y al menos 10, correspondientes a fracasos.
En consecuencia, podemos asumir que la distribución muestral de ^p sigue aproximadamente a la normal.

**Método de Wald para una proporción pag 99 -101**

error _est _hip <- sqrt (( valor _ nulo * (1 - valor _ nulo )) / n)

Z <- (p_ exito - valor _ nulo ) / error _est _hip

p <- pnorm (Z, lower . tail = FALSE )

**Método de Wald para dos proporciones**

Condiciones:

1. Cada proporción, por separado, sigue el modelo normal.

2. Las dos muestras son independientes una de la otra.

p <- pnorm (Z, lower . tail = FALSE )

DOS METODOS:

**CUANDO LA HIPOTESIS NULA SUPONE QUE NO HAY DIFERENCIA ENTRE LAS PROPORCIONES PAG: 102 EJEMPLO:**

**CUANDO LA HIPOTESIS NULA TIENE UN VALOR DISTENTO DE 0 PAG:104 EJEMPLO**

**METODO DE WILSON PAG 105.**

Muy similar a método de wald pero mejor aceptado por tener menos aproximaciones.

SE PUEDE EMPLEAR PARA UNA O DOS PROPORCIONES.

**Al trabajar con dos proporciones no puede establecer un valor nulo diferente de cero, es por este mortivo que si tengo valor nulo diferente de cero se debe utilizar wald pag 104.**

prop . test ( exitos , n = n, p = valor _nulo , alternative = " greater ", conf . level = 1 - alfa )

- `x`: cantidad de éxitos en la muestra.
- `n`: tamaño de la muestra.
- `p`: valor nulo (por defecto, `p=NULL`).
- `alternative`: tipo de hipótesis alternativa, por defecto bilateral (`alternative="two.sided"`), y valores "less" y "greater" para hipótesis unilaterales.
- `conf.level`: nivel de confianza (`conf.level=0.95` por defecto).

Método de Wilson para dos proporciones

prueba <- prop . test ( exitos , n = n, alternative = "two. sided ", conf . level = 1 - alfa )

Comunmente se puede utilizar Wald o Wilson, las preguntas a responder para su uso es:

Estudios previos habían determinado que la proporción de autoras en la especialidad de medicina interna era de 42%. ¿Respaldan estos datos tal estimación?

Igualdad de proporciones o preguntas que digan diferencia en la proporción de autoras en x e de y es de 0,28. ¿A cuántos -- deberíamos monitorear para obtener un intervalo de confianza del 95% y poder estadístico de 80%, si se intenta mantener aproximadamente la misma proporción de gente estudiada en cada caso?

Se asigna la diferencia = 0.28

Necesitamos hacer un ajuste de modo que la diferencia de las proporciones tenga el valor mismo que el del enunciado, es decir, 0.28. Para esto realizamos la siguiente

```
p_xNueva <- (p_x+p_y)/2 - diferencia/2
p_yNueva <- (p_x+p_y)/2 + diferencia/2
diferencia_esperada <- abs(p_x - p_y)

# Usamos la prueba de Wilson para calcular el tamaño de las muestras
resultado <- bsamsize(p_xNueva, p_yNueva, fraccion, alfa, poder)
```

**PODER Y PRUEBAS DE PROPORCIONES**

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

n: número de observaciones por cada grupo.
p1: probabilidad de éxito en un grupo.
p2: probabilidad de éxito en otro grupo.
sig.level: nivel de significación. – corresponde a alfa o 0.05 que s eusa comúnmente.
power: poder de la prueba.
alternative: tipo de hipótesis alternativa (“one.sided” si es unilateral, “two.sided” si es bilateral).

Otro método del poder.

• pwr.p.test(h, n, sig.level, power, alternative): para pruebas con una única proporción.
• pwr.2p.test(h, n, sig.level, power, alternative): para pruebas con dos proporciones donde ambas muestras son de igual tamaño.
• pwr.2p2n.test(h, n1, n2, sig.level, power, alternative): para pruebas con dos proporciones y muestras de diferente tamaño.
El tamaño del efecto puede calcularse como muestra la ecuación 7.6 pag:107, implementada en R en la función ES.h(p1, p2) del paquete pwr. Donde p1 y p2 son las proporciones calculadas en pag 102.

En el caso de una única proporción, los autores del paquete `pwr` sugieren usar $p_2 = 0,5$ (Champely y col., 2020).

Otra función que nos puede ser de ayuda es `bsamsize(p1, p2, fraction, alpha, power)`, del paquete `Hmisc`. En el caso de una prueba de Wilson con dos muestras, calcula los tamaños de cada grupo dados los siguientes argumentos:

- `p1`: probabilidad de la población para el grupo 1.
- `p2`: probabilidad del grupo 2.
- `fraction`: fracción de las observaciones en el grupo 1 ($n1/(n1 + n2)$).
- `alpha`: nivel de significación.
- `power`: poder deseado.

# CAPITULO 8. PAG 109

pruebas para inferir acerca de proporciones **cuyas hipótesis nula y alternativa no mencionan parámetro alguno**. Es más, **ninguna de ellas hace alguna suposición sobre la distribución de la población desde donde proviene la muestra analizada**. Es por esta razón que a estas pruebas (y a otras que se abordan en capítulos posteriores) se les denomina no paramétricas o libres de distribución.

### PRUEBA CHI CUADRADO DE PEARSON PAG 109.

Sirve para inferir cuando se tienen dos variables categóricas y una de ellas es dicotómica(solo dos niveles).
Condiciones:
1. Las observaciones deben ser independientes entre sí.
2. Debe haber a lo menos 5 observaciones esperadas en cada grupo.

## Prueba chi-cuadrado de homogeneidad pag 110.

Esta prueba resulta adecuada si queremos determinar si dos poblaciones (la variable dicotómica) presentan las mismas proporciones en los diferentes niveles de una variable categórica.  Hipotesis a contratar pag 111.

Pag 112 ejemplo R.
prueba <- chisq . test (tabla , correct = FALSE )
pchisq(valor pagina  111, df = 4 (pag 112), lower.tail = FALSE)

### Prueba chi-cuadrado de bondad de ajuste pag 112.

Esta prueba permite comprobar si una distribución de frecuencias observada se asemeja a una distribución esperada. Usualmente se emplea para comprobar si una muestra es representativa de la población. Hipotesis a contratar pag 113.
Pag 113 ejemplo R.

prueba <- chisq . test (tabla , correct = FALSE )

## Prueba chi-cuadrado de independencia pag 114.

Esta prueba permite determinar si dos variables categóricas, de una misma población, son estadísticamente independientes o si, por el contrario, están relacionadas. Hipotesis a contratar pag 114.
Pag 115 ejemplo R.

## PRUEBAS PARA MUESTRAS PEQUEÑAS pag 115.

Hemos visto que la prueba chi2 nos pide que las observaciones esperadas para cada grupo sean a lo menos 5. Sin embargo, hay escenarios donde esta condición no se cumple, por lo que debemos recurrir a alguna alternativa.

### Prueba exacta de Fisher pag 115.

La prueba exacta de Fisher es una alternativa a la prueba chi cuadrado de independencia en el caso de que ambas variables sean dicotómicas. Así, las hipótesis a contrastar son:

H0: las variables son independientes.
HA: las variables están relacionadas.

fisher.test(x, conf.level) conf.level = 1-alfa
ejemplo en R pag 117.

# Prueba de mcNemar pag 118.

**Esta prueba resulta apropiada cuando una misma característica, con respuesta dicotómica, se mide en dos ocasiones diferentes para los mismos sujetos (muestras pareadas) y queremos determinar si se produce o no un cambio significativo entre ambas mediciones.**

### Las hipótesis asociadas a la prueba de mcNemar son:

H0: no hay cambios significativos en las respuestas.
HA: sí hay cambios significativos en las respuestas.

pchisq(0.083 valor tabla 118, 1, lower.tail = FALSE)
prueba <- mcnemar . test ( tabla )

Ejemplo pag 118.
Ejemplo en R pag 119.

## Prueba Q de cochran pag 120.

La prueba Q de Cochran es una extensión de la prueba de mcNemar, adecuada cuando la variable de respuesta es dicotómica y la variable independiente tiene más de dos observaciones pareadas.
Las hipótesis contrastadas por la prueba Q de Cochran son:
H0: la proporción de “éxitos” es la misma para todos los grupos.
HA: la proporción de “éxitos” es distinta para al menos un grupo.

Como ya debemos suponer, esta prueba también requiere que se cumplan algunas condiciones:
1. La variable de respuesta es dicotómica.
2. La variable independiente es categórica.
3. Las observaciones son independientes entre sí.
4. El tamaño de la muestra es lo suficientemente grande. Glen (2016a) sugiere que n * k  mayor o igual a  24, donde n es el tamaño de la muestra (la cantidad de instancias, para el ejemplo) y k, la cantidad de niveles en la variable independiente.

cochran.qtest(formula, data,alpha = 0.05)

- `formula`: fórmula de la forma `respuesta ~ independiente | bloques`.
- `data`: matriz de datos en formato largo.
- `alpha`: nivel de significación.

En este punto, debemos mencionar que la hipótesis nula de la prueba Q de Cochran no es específica, sino que comprueba la igualdad de todas las proporciones. Esta clase de hipótesis nula suele llamarse ómnibus (en ocasiones también colectiva o global). Así, se dice que la prueba Q de Cochran es una prueba ómnibus porque tiene esta clase de hipótesis nula, con la dificultad de que solo detecta si existe al menos bloque con una proporción de “éxito” diferente. Sin embargo, de ser afirmativa la respuesta, no nos dice qué grupos presentan diferencias (Lane, s.f.). Desde luego, existen métodos para responder a esta última pregunta, llamados pruebas post-hoc, o también a posteriori. Reciben este nombre porque se realizan una vez que se ha concluido gracias a la prueba ómnibus que existen diferencias significativas.
pairwiseMcnemar(formula, data, method) method= Bonferroni o Holm