# Estadística I – clase 8

## MISCELÁNEAS

# Comparación de proporciones en dos muestras independientes

# Problema

- Se quiere determinar si existe relación entre tener un nacimiento con bajo peso al nacer y el hecho de que la madre haya fumado durante el embarazo.

# Situación

- Se tienen dos variables dicotómicas medidas en la misma muestra de pacientes
- Low vale 1 si se trata de un nacimiento con bajo peso, 0 en caso contrario
- SMOKE vale 1 si la madre fumó durante el embarazo,0 en caso contrario
- Se puede pensar que se tienen dos muestras,
  - Muestra de mujeres que fumaron durante el embarazo
  - Muestra de mujeres que no fumaron durante el embarazo

# Situación

- Este problema puede pensarse como de comparación del comportamiento de la variable LOW en dos poblaciones similares, que sólo difieren en cuanto a si la madre fumó o no durante el embarazo
- La diferencia con los casos vistos anteriormente es que la variable LOW no es cuantitativa

# Prueba chi cuadrado

- Las hipótesis que se contrastan son:

    $H_0$: no hay diferencia en la proporción de bajo peso al nacer entre los grupos.

    $H_1$: hay diferencia en las proporciones.

# Prueba de chi cuadrado

- El estadístico Chi cuadrado compara las frecuencias obtenidas en la tabla de contingencia con las que debería haber si no hubiera asociación entre las variables

- Las frecuencias esperadas se obtienen de la siguiente forma

$$e_{ij} = \frac{(\text{frecuencia marginal fila } i) \times (\text{frecuencia marginal columna } j)}{\text{Total de observaciones}}$$

# Prueba chi cuadrado

- Frecuencias observadas

```
> x$observed   # observed counts
          LOWBWT$SMOKE
LOWBWT$LOW  0  1
         0 86 44
         1 29 30
```

Si las variables son independientes,

las tablas de frecuencias esperadas y observadas

serán similares.

- Frecuencias esperadas

```
> x$expected   # expected counts under the null
          LOWBWT$SMOKE
LOWBWT$LOW        0        1
         0 79.10053 50.89947
         1 35.89947 23.10053
```

# Prueba chi cuadrado

- Estadístico chi cuadrado

$$\chi^2 = K \sum_i \sum_j \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

Este estadístico tiene distribución asintótica chi cuadrado con r grados de libertad

$$r = (filas - 1) * (columnas - 1)$$

# Prueba chi cuadrado

Limitaciones en el uso de la prueba chi-cuadrado

- Como la distribución del estadístico es aproximada, no se recomienda su uso si existen frecuencias esperadas inferiores a 5.
- Solución: Test exacto de Fisher

# Test exacto de Fisher

- Se basa en el cálculo de la probabilidad de haber obtenido esas frecuencias observadas, condicionados a los marginales fijos.
- A diferencia del test chi-cuadrado la distribución utilizada es exacta

# Comparación de métodos

```
> # comparación de dos proporciones
> xtab <- table(LOWBWT$LOW,LOWBWT$SMOKE)
> dimnames(xtab) <- list(
+   Bajo_peso = c("si", "no"),
+   Fumadora = c("si", "no"))
> xtab
         Fumadora
Bajo_peso si no
       si 86 44
       no 29 30

> prop_test(xtab)
n statistic    df      p p.signif
1   189       4.24     1 0.0396 *
```

```
> fisher_test(xtab)
     n    p p.signif
1    189 0.0362 *
```

# Conclusiones de utilización

Prueba chi cuadrado

- Se utiliza con muestras grandes, dado que su distribución es aproximada
- Se desaconseja su utilización si hay frecuencias esperadas inferiores a 5.

Test exacto de Fisher

- Se utiliza con muestras pequeñas
- Es la alternativa a utilizar cuando el Chi cuadrado no pueda utilizarse

# Test binomial

# Test binomial

- La distribución binomial se utiliza para problemas que involucran a una proporción poblacional.
- Se basa en el conteo de la ocurrencia de un suceso dicotómico (éxito/fracaso)
- Ejemplos
  - Una persona fuma/no fuma
  - Un tratamiento es efectivo/no es efectivo
  - Un paciente sobrevive/no sobrevive a una enfermedad

# Test binomial

- En todos los casos, se supone que a nivel poblacional existe p que es la proporción de individuos que tienen la propiedad que se está midiendo
- La variable binomial cuenta el número de éxitos en una muestra de tamaño n.
- La muestra pertenece a una población cuya propiedad de éxito es p
- Con estos parametros, pueden calcularse con precisión las probabilidades vinculadas al testeo de hipótesis relacionadas a proporciones.

# Test binomial

- Test binomial de una muestra

$H_0 : p = p_0$

$H_1 : p \neq p_0$

$H_1 : p < p_0$

$H_1 : p > p_0$

# Test binomial

- Test binomial de dos muestras

$H_0: p_1 - p_0 = 0$

$H_1: p_1 - p_0 \neq 0$

$H_1: p_1 - p_0 < 0$

$H_1: p_1 - p_0 > 0$

- Cuando la muestra es grande, el estadístico de prueba se aproxima por normal, por el teorema central del límite

# Test binomial

- En un estudio de los efectos del estrés un investigador enseñó a 18 estudiantes dos métodos diferentes para hacer el mismo nudo.
- La mitad de los sujetos (seleccionados aleatoriamente) aprendieron primero el método A y la otra mitad aprendió primero el método B.
- Posteriormente -a medianoche y después de un examen final de cuatro horas de duración-, a cada sujeto se le pidió que hiciera el nudo.
- La predicción fue que el estrés induciría regresión, esto es, que los sujetos regresarían al primer método aprendido para hacer el nudo.
- Cada sujeto fue categorizado conforme a si usó el primer o el segundo método aprendido de hacer nudos cuando se le pedía que hiciera el nudo bajo estrés. 16 de ellos utilizaron el método aprendido primero

# Test binomial

$H_0: p = 0.5$ $H_1: p \neq 0.5$

```
> binom_test(16,18,p = 0.5,
+            alternative = "two.sided",
+            conf.level = 0.95,
+            detailed = T)
      n estimate   statis       p      parameter conf.low conf.high method alternative
1    18    0.889         16 0.00131         18    0.653     0.986 Exact~   two.sided
```

# Comparación de proporciones en dos muestras apareadas

Test de McNemar

# Test de McNemar

- La prueba de McNemar para la significación de los cambios es particularmente aplicable a los problemas tipo "antes-después", en los cuales cada sujeto se utiliza como su propio control y en los que las mediciones se realizan ya sea en escala nominal u ordinal
- Es un test tipo chi cuadrado en el que se comparan frecuencias esperadas y observadas

# Test de McNemar

| | | DESPUES | |
|---|---|---|---|
| | | - | + |
| ANTES | + | A | B |
| | - | C | D |

$$\chi^2 = \frac{(A-D)^2}{A+D}$$

- Este estadístico tiene distribución chi cuadrado con 1 grado de libertad
- Cuando las frecuencias esperadas son bajas, existe una versión exacta que puede calcularse

# Test de McNemar

- Durante las campañas presidenciales de 1980 en Estados Unidos se realizaron debates televisivos entre dos o más candidatos. Un investigador en técnicas de comunicación estaba interesado -tanto como los candidatos- en determinar si los debates entre los candidatos presidenciales en las elecciones de 1980 eran efectivos o no en cuanto a cambiar las preferencias de los televidentes hacia los distintos candidatos.
- Se predijo que si los candidatos (Jimmy Carter y Ronald Reagan) eran igualmente efectivos, habría cambios comparables en las preferencias a cada candidato por parte de los televidentes.

# Test de McNemar

- Por otro lado, si un candidato era más efectivo o persuasivo durante el debate, entonces habría un cambio diferencial de un candidato a otro.
- Para evaluar la efectividad del debate, el investigador seleccionó 70 adultos al azar antes del debate y les pidió que indicaran sus preferencias hacia ambos candidatos.
- Después de la conclusión del debate, les volvió a preguntar acerca de su predilección.
- Así, en cada caso el conocía las preferencias de las personas antes del debate y después del mismo.

# Test de McNemar

```
> # Data: frequencies of smokers before and after interventions
> xtab <- as.table(
+   rbind(c(13, 28), c(27,7)))
> dimnames(xtab) <- list(
+   before = c("Carter", "Reagan"),
+   after = c("Carter", "Reagan")
+ )
> xtab
        after
before   Carter Reagan
  Carter     13     28
  Reagan     27      7
> # Compare the proportion
> mcnemar_test(xtab)
n statistic    df      p p.signif method
1    75          0      1      1 ns       McNemar test
```

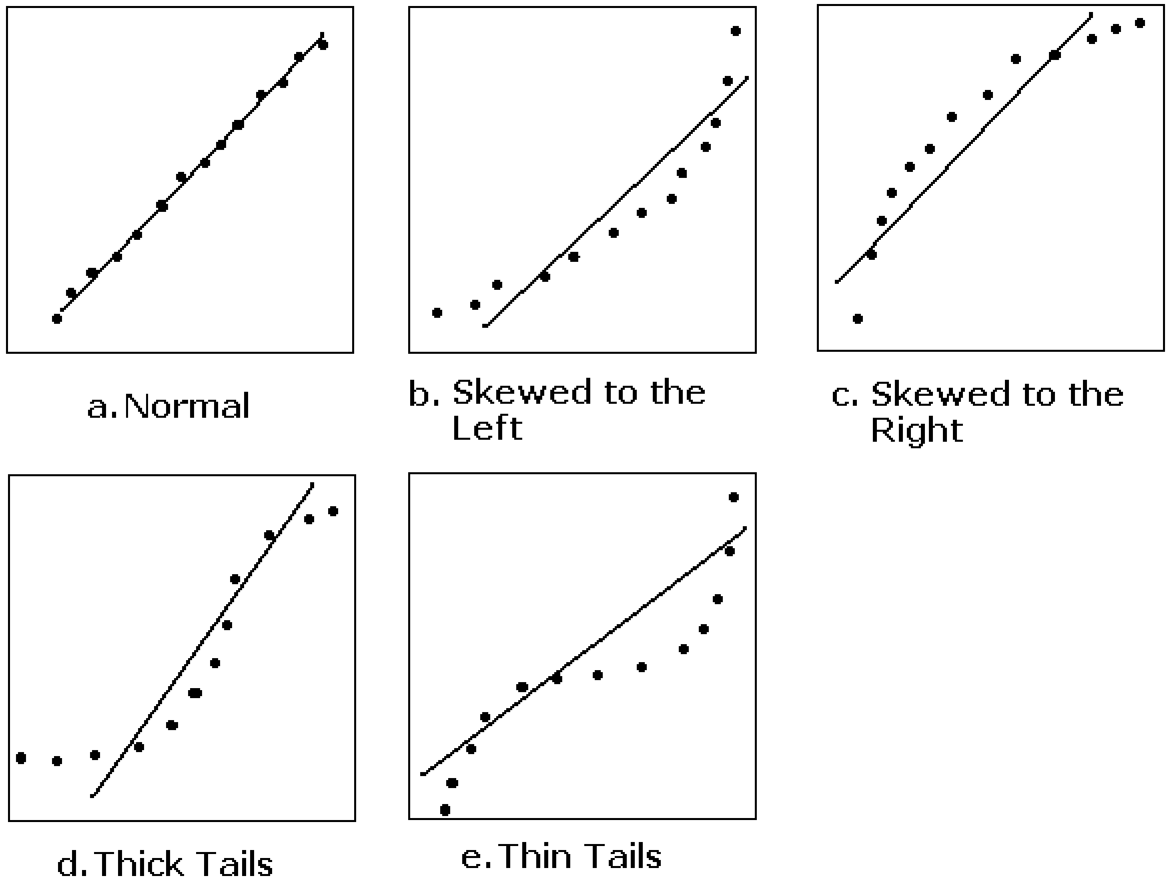
# Detección de normalidad y homocedasticidad

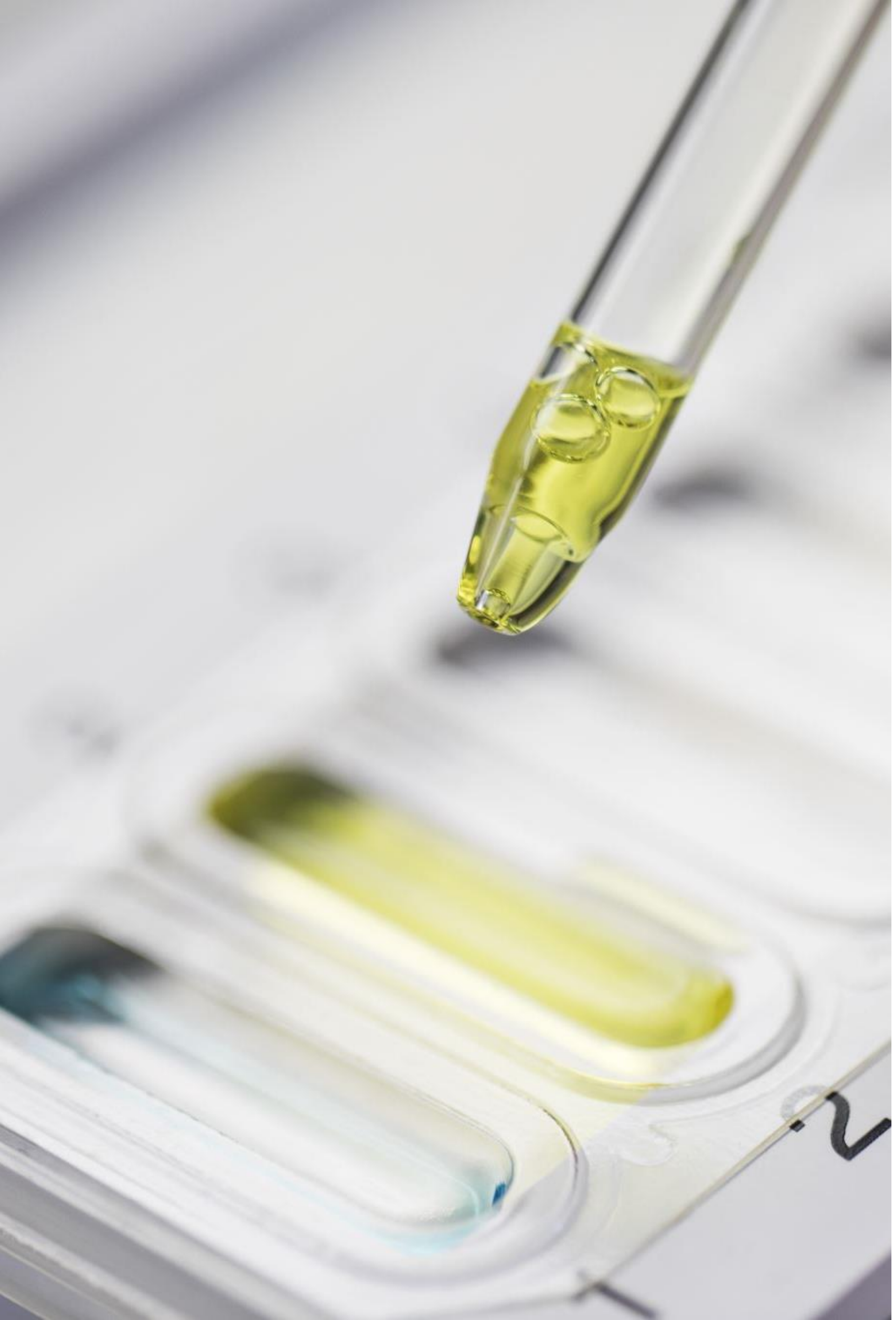
# Detección de normalidad en una muestra

Se utiliza el gráfico Q-Q Plot que grafica los cuantiles de la muestra contra los cuartiles teóricos de una distribución normal.

Si la muestra es normal, los puntos deben caer sobre la bisectriz

# Q-Q Plots


a. Normal
b. Skewed to the Left
c. Skewed to the Right
d. Thick Tails
e. Thin Tails

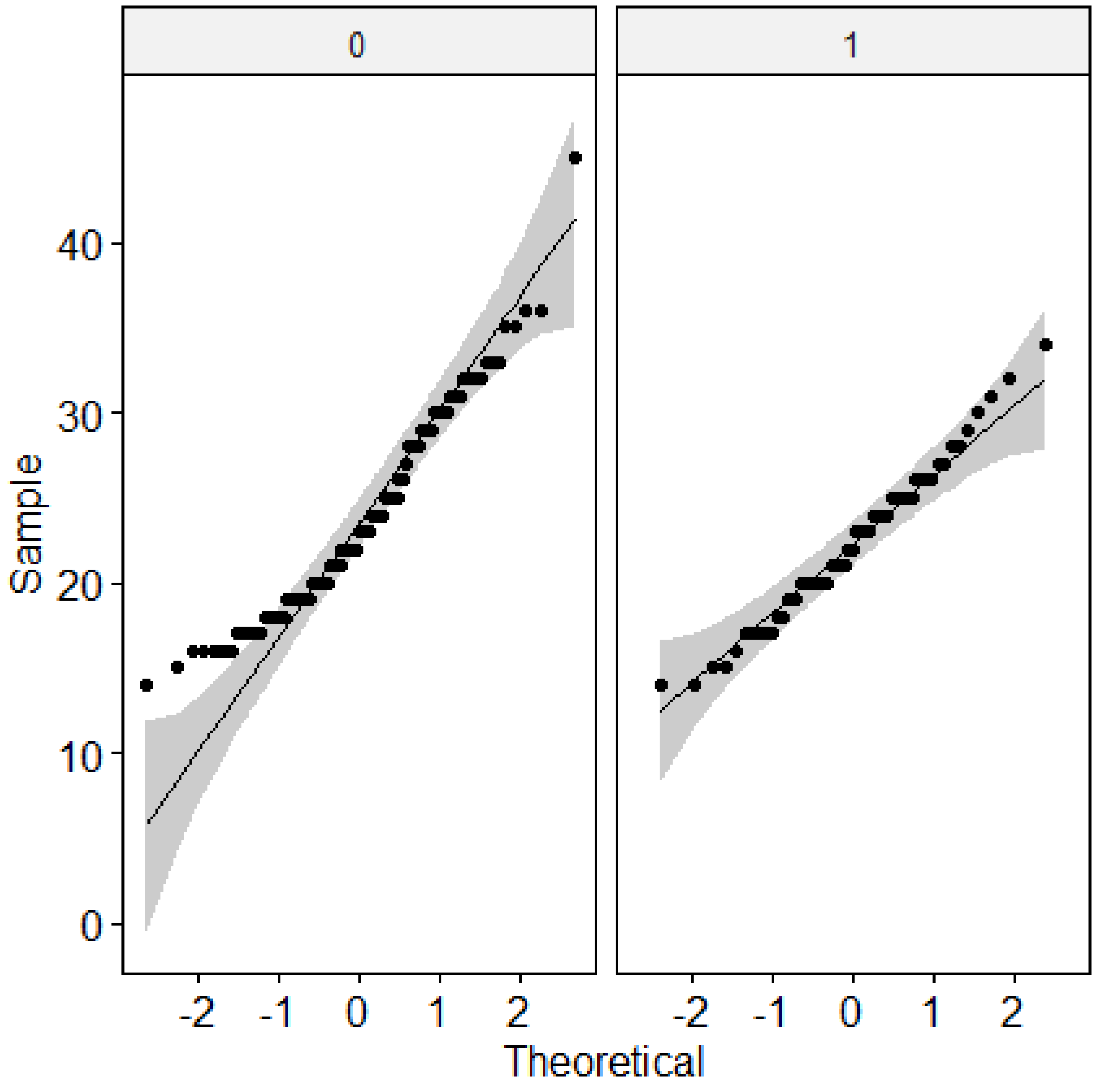


# Test de normalidad

- **Shapiro – Wilks**: para muestras más pequeñas
- **Kolmogorov – Smirnov**: para muestras grandes por ser noparamétrico

- $H_0$: los datos son compatibles con una distribución normal
- $H_1$: los datos no son compatibles con una distribución normal

- Siempre deben complementarse con un diagnostico gráfico ya que los test dependen de los tamaños muestrales




# Testeo de normalidad

```
> LOWBWT %>%
+   shapiro_test(AGE)
# A tibble: 1 × 3
  variable statistic         p
  <chr>        <dbl>     <dbl>
1 AGE          0.960 0.0000319
> LOWBWT %>% levene_test(AGE ~ factor(LOW))
# A tibble: 1 × 4
    df1   df2 statistic     p
  <int> <int>     <dbl> <dbl>
1     1   187      2.70 0.102
```

# Test de igualdad de varianzas

- Test de Levene
- $H_0$: las varianzas de las poblaciones son iguales
- $H_1$: las varianzas de las poblaciones no son iguales

# TALLER DE CALCULO DEL TAMAÑO MUESTRAL Y LA POTENCIA EN R

Dr. Roberto Muiños

# Porqué calcular el tamaño muestral?

- Una muestra pequeña puede impedir detectar un efecto de interés en un experimento
- Una muestra demasiado grande puede generar la utilización innecesaria de recursos difíciles de conseguir
- Objetivo: determinar la cantidad de muestra necesaria para garantizar la detección de los efectos de interés, sin utilizar más recursos de los estrictamente necesarios

# Principales elementos del cálculo del tamaño muestral

- Tamaño del efecto: magnitud del efecto en la hipótesis alternativa.
- Potencia: probabilidad de rechazar correctamente la hipótesis nula cuando es falsa
- Nivel de significación ($\alpha$): probabilidad de rechazar la hipótesis nula cuando es verdadera

# Principales elementos del cálculo del tamaño muestral

- Tamaño del efecto: magnitud del efecto en la hipótesis alternativa.
  - Mientras mayor es el tamaño del efecto, más sencilla es su detección y menor es la cantidad de casos necesarios

# Principales elementos del cálculo del tamaño muestral

- Potencia: probabilidad de rechazar correctamente la hipótesis nula cuando es falsa
  - $Potencia = 1 - \beta$, siendo $\beta$ la probabilidad de error tipo II (falso negativo).
  - Mientras mayor es la potencia, mayor es la probabilidad de detectar un efecto si está presente, y mayor es la cantidad de muestra necesaria
  - Valor habitual utilizado: 0.80

# Principales elementos del cálculo del tamaño muestral

- Nivel de significación ($\alpha$): probabilidad de rechazar la hipótesis nula cuando es verdadera
  - Probabilidad de error de tipo I
  - Valor habitual utilizado: 0.05

# Ejemplo

Los resultados obtenidos por un gabinete psicológico durante los últimos 15 años indican que los pacientes con insomnio duermen un promedio de 4 horas con un desvío de 2 hora. Se toma una muestra de 20 personas con diagnóstico de insomnio para determinar si esta afirmación es válida y se obtiene un promedio de 4.8 horas. Utilizar $\alpha = 0.05$

- Determinar la decisión de la prueba de hipótesis.
- De acuerdo a la decisión, qué tipo de error podría estar cometiéndose?
- Calcular la potencia de acuerdo a la decisión obtenida.
- Calcular el tamaño de muestra necesario para que pueda decidirse con una potencia del 80%

# Ejemplo

Los resultados obtenidos por un gabinete psicológico durante los últimos 15 años indican que los pacientes con insomnio duermen un promedio de 4 horas con un desvío de 2 horas. Se toma una muestra de 20 personas con diagnóstico de insomnio para determinar si esta afirmación es válida y se obtiene un promedio de 4.4 horas.

X= cantidad de horas dormidas $X \sim N(\mu, \sigma)$

$$\mu = cantidad\ promedio\ de\ horas\ dormidas\ por\ la\ población\ de\ insomnes$$

$$\mu_0 = 4\ horas$$

Hipótesis a contrastar:

$$H_0 = los\ insomnes\ duermen\ en\ promedio\ 4\ horas$$

$$H_1 = los\ insomnes\ no\ duermen\ en\ promedio\ 4\ horas$$

$$H_0\colon\ \mu = 4 \qquad\qquad H_1\colon\ \mu \neq 4$$

Test para la media de una normal con sigma conocido

# Test para la media de la normal

Sigma conocido

$$E = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0,1)$$

Sigma desconocido

$$E = \frac{\bar{X} - \mu_0}{S / \sqrt{n}} \sim t_{n-1}$$

# Ejemplo

$$H_0\!: \ \mu = 4 \qquad\qquad H_1\!: \ \mu \neq 4$$

$$E = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}} = \frac{4{,}4-4}{2/\sqrt{20}} = 0{,}894$$

No significativo

Posible error de tipo II

- Cálculo de la potencia para este problema: 0,136
- Por lo tanto, la probabilidad de error de tipo II es 1-0,136=0,864
- Qué cantidad de casos se necesitaría para tener una Probabilidad error de tipo II de 0.2?: 200 casos.

# Principales elementos del cálculo del tamaño muestral

- El tamaño muestral (*n*) en una prueba de hipótesis está relacionado con varios elementos
  - Tipo de variable a comparar (cuantitativa, dicotómica)
  - Tipo de contraste a realizar (bilateral, unilateral, no inferioridad, equivalencia, superioridad)
  - Nivel de significación ($\alpha$)
  - Potencia (1-$\beta$)
  - Tamaño del efecto (*d*)

# Cálculo del tamaño muestral

La potencia y el nivel de significación son requerimientos fijados por los investigadores

El tamaño del efecto está relacionado con los datos

Es, esencialmente, la diferencia estandarizada entre las medias de las hipótesis que se contrastan

# Tamaño del efecto

- Cómo estimar el tamaño del efecto?
  - Utilizar información preliminar las medias y la variación y, a continuación, calcular directamente el tamaño del efecto
  - Utilizar información de otros estudios similares para obtener medias y la variación, y luego calcular directamente el tamaño del efecto
  - Sin información previa, hacer una estimación del tamaño del efecto esperado

# Cálculo del tamaño muestral con R

- Existen muchos paquetes para estas tareas
- Nosotros usaremos dos
  - pwr
  - WebPower

# Principales funciones según la prueba estadística

| # | Name of Test | in R? | Package | Function |
|---|---|---|---|---|
| 1 | One Mean T-test | Yes | pwr | pwr.t.test |
| 2 | Two Means T-test | Yes | pwr | pwr.t.test |
| 3 | Paired T-test | Yes | pwr | pwr.t.test |
| 4 | One-way ANOVA | Yes | pwr | pwr.anova.test |
| 5 | Single Proportion Test | Yes | pwr | pwr.p.test |
| 6 | Two Proportions Test | Yes | pwr | pwr.2p.test |
| 7 | Chi-Squared Test | Yes | pwr | pwr.chisq.test |
| 8 | Simple Linear Regression | Yes | pwr | pwr.f2.test |
| 9 | Multiple Linear Regression | Yes | pwr | pwr.f2.test |
| 10 | Correlation | Yes | pwr | pwr.r.test |
| 11 | One Mean Wilcoxon Test | Yes* | pwr | pwer.t.test + 15% |
| 12 | Mann-Whitney Test | Yes* | pwr | pwer.t.test + 15% |
| 13 | Paired Wilcoxon Test | Yes* | pwr | pwer.t.test + 15% |
| 14 | Kruskal Wallace Test | Yes* | pwr | pwr.anova.test + 15% |
| 15 | Repeated Measures ANOVA | Yes | WebPower | wp.rmanova |
| 16 | Multi-way ANOVA (1 Category of interest) | Yes | WebPower | wp.kanova |
| 17 | Multi-way ANOVA (>1 Category of interest) | Yes | WebPower | wp.kanova |
| 18 | Non-Parametric Regression (Logistic) | Yes | WebPower | wp.logistic |
| 19 | Non-Parametric Regression (Poisson) | Yes | WebPower | wp.poisson |
| 20 | Multilevel modeling: CRT | Yes | WebPower | wp.crt2arm/wp.crt3arm |
| 21 | Multilevel modeling: MRT | Yes | WebPower | wp.mrt2arm/wp.mrt3arm |
| 22 | GLMM | Yes^ | Simr & lme4 | n/a |

*-parametric test with non-parametric correction
^-detailed in future Module

# Cálculo del tamaño muestral para una media poblacional

- La fórmula de cálculo del tamaño muestral está relacionado con el concepto de intervalo de confianza para la media de una normal

# Intervalo de confianza

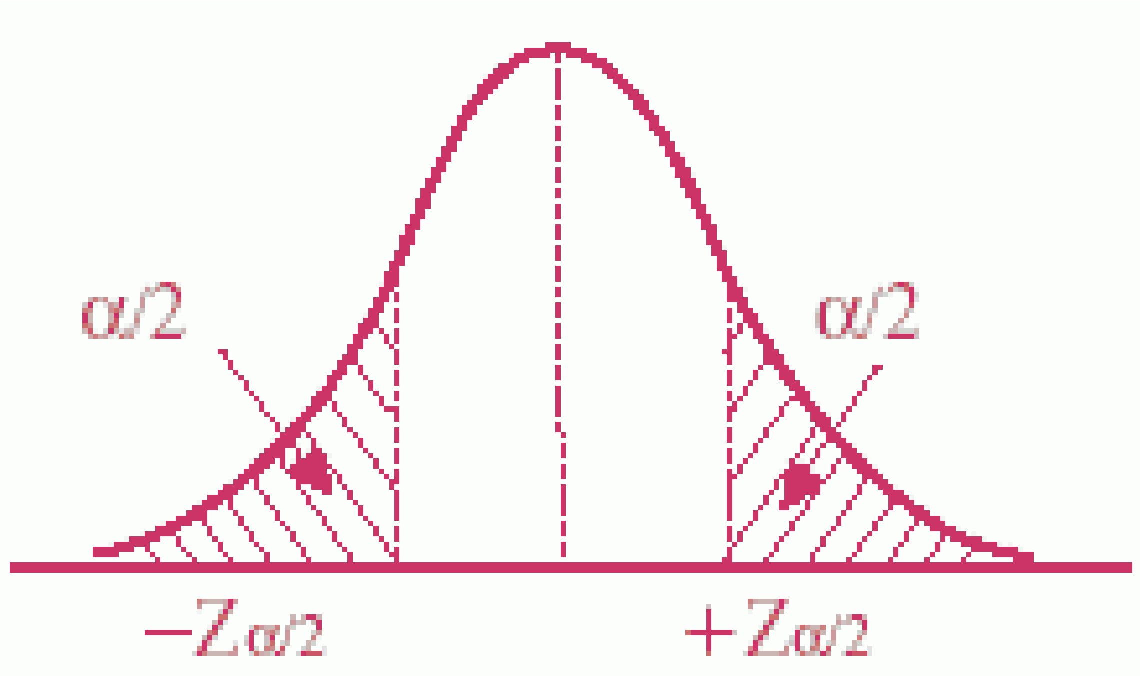
Es una forma alternativa de estimar un parámetro

Consiste en un rango de valores el cual contendrá al valor poblacional con una alta probabilidad

Por ejemplo, el intervalo de confianza para la media de una población se extiende a ambos lados de la media muestral un número fijo de desvíos stándard.

# Cálculo del tamaño muestral para una media poblacional - Intervalo de confianza para la media de una normal

$$\bar{X} \sim N\left(\mu, \sigma/\sqrt{n}\right)$$




$$P\left(-z\alpha/_2 \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq z\alpha/_2\right) = 1-\alpha$$

$$P\left(-z\alpha/_2 * S/\sqrt{n} \leq \bar{X}-\mu \leq z\alpha/_2 * S/\sqrt{n}\right) = 1-\alpha$$

$$|\bar{X}-\mu| \leq z\alpha/_2 * S/\sqrt{n} \leq \delta$$

# Cálculo del tamaño muestral para una media poblacional

$$n = \frac{\left(z_{\alpha/2} + z_{\beta}\right)^2}{\delta^2}\sigma^2$$

Variabilidad

Error tipo II: β

Nivel de significación α
(error tipo I)

Diferencia admitida

¿Cuándo necesito mayor n?

Mayor potencia
Mayor desviación típica
Menor error tipo I
Menor diferencia relevante

# Ejemplo

$$H_0\!: \mu = 4 \qquad\qquad H_1\!: \mu \neq 4$$

$$E = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{4{,}4 - 4}{2/\sqrt{20}} = 0{,}894$$

No significativo

Posible error de tipo II

- Cálculo de la potencia para este problema: 0,136
- Por lo tanto, la probabilidad de error de tipo II es 1-0,136=0,864
- Qué cantidad de casos se necesitaría para tener una Probabilidad error de tipo II de 0.2?: 200 casos.

# Volviendo al ejemplo

**CALCULO DE LA POTENCIA**

```
>  pwr.t.test (d=0.28,
+            sig.level =0.05,
+            power=NULL,
+            n=20,
+            type="two.sample")
```

```
     Two-sample t test power calculation

              n = 20
              d = 0.28
      sig.level = 0.05
          power = 0.1387525
    alternative = two.sided
```

**CALCULO DEL N PARA UNA POTENCIA DE 0.8**

```
>  pwr.t.test (d=0.28,
+            sig.level =0.05,
+            power=0.8,
+            n=NULL,
+            type="two.sample")
```

```
     Two-sample t test power calculation

              n = 201.1909
              d = 0.28
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
```