Práctica 6: Intervalos de confianza y contrastes de hipótesis para una población.

Nota: Para los estudios de proporciones es muy útil tener cargado el plugin "IPSUR". Se carga desde el menú "Herramientas"de Rcommander. Cuando se carga un plugin Rcommander se reinicia (es mejor cargarlo primero antes de empezar la práctica).

1. Distribuciones en el muestreo

En esta sección, consideramos una v.a. X, y se extrae una m.a.s. de la población en la que se observa dicha variable.

La muestra se escribe $(X_1, X_2, \ldots, X_n)$. Con la muestra construimos los estadísticos media y cuasivarianza:

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

Entonces

a) Si X es una v.a cualquiera y n es suficientemente grande, por el Teorema Central del Límite:

$$\bar{X}_n \sim N(\mu_X, \frac{\sigma_X}{\sqrt{n}})$$

b) Si X es normal, para cualquier valor de n:

$$\bar{X}_n \sim N(\mu_X, \frac{\sigma_X}{\sqrt{n}}), \qquad \frac{(n-1)S_n^2}{\sigma_X^2} \sim \chi^2_{n-1}, \qquad \frac{\bar{X}_n - \mu}{\frac{S_n}{\sqrt{n}}} : t_{n-1}$$

Ejercicios:

1. (En clase de TEORIA) El peso por unidad en una variedad de mandarina tiene una distribución normal de media 50 y varianza 5. Para una muestra de tamaño 20,

   a) ¿cuál es la probabilidad de que la media muestral se encuentre entre 45 y 50?

   b) ¿cuál es la probabilidad de que la varianza muestral sea mayor o igual que 7.44?

   c) ¿por encima de qué valor se encontrará la varianza muestral en el 90 % de las muestras ?

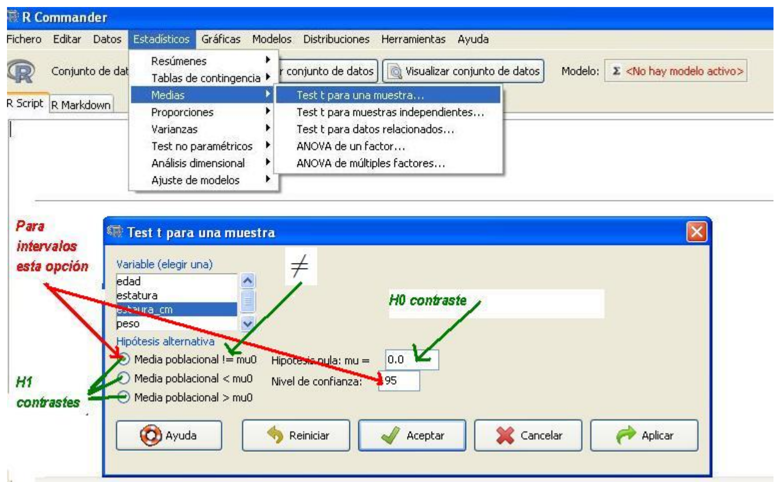
2. Genera 300 muestras de tamaño 5 de una población normal de media 8 y desviación típica 3. Para cada muestra guarda la media y la desviación típica.

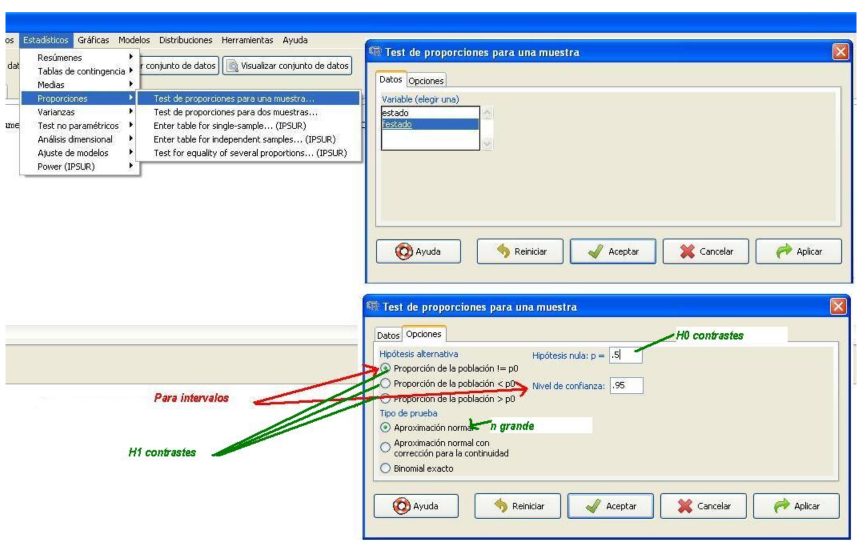
   a) comprueba que el estadístico $\dfrac{\bar{X} - 8}{\sigma_x/\sqrt{5}}$ se comporta como una distribución normal y no como una t-student con 4 grados de libertad.

b) comprueba que el estadístico $\dfrac{\bar{X}-8}{S/\sqrt{5}}$ se comporta como una distribución t-student con 4 grados de libertad y no como una normal.

Nota: Para ello Crea una nueva variable que recoja el valor estadistico propuesto y utiliza el *gráfico comparacion de cuantiles*.

2. Intervalos de confianza y contrastes para la media con varianza desconocida y la proporción.





• Para el estudio de proporciones, hay que tener en cuenta que Rcommander ordena los niveles (modalidades) de un factor por orden alfabético y que asocia el éxito al primer nivel. Si es necesario porque el éxito va asociado al otro nivel hay que reordenar los niveles del factor desde el menú *datos–modificar variables–reordenar niveles* ó bien usar directamente el menu para las proporciones IPSUR.

3. Sobre los datos de Galton (carga los datos galtonR.RData). Supongamos que estos datos son una muestra aleatoria de marineros.

   a) Comprueba que la distribución de los datos sobre la estatura en cm es normal mediante el gráfico de comparación de cuantiles.

   b) Haz un intervalo de confianza al 95 % para la media de la estatura en cm.

```
data:  galton$estaura_cm
t = 495.7403, df = 399, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 172.0475 173.4175
sample estimates:
mean of x
 172.7325
```

   c) Contrasta la hipótesis de que la estatura media no supera 173cm al 1 % de significación.

```
        One Sample t-test

data:  galton$estaura_cm
t = -0.7677, df = 399, p-value = 0.2216
alternative hypothesis: true mean is less than 173
95 percent confidence interval:
    -Inf 173.307
sample estimates:
mean of x
 172.7325
```

   d) Haz un intervalo con un nivel de confianza del 95 % para la proporción de marineros que son solteros.

```
        1-sample proportions test without continuity correction

data:  rbind(.Table), null probability 0.5
X-squared = 225, df = 1, p-value < 2.2e-16
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.8389809 0.9038849
sample estimates:
    p
0.875
```

   e) Contrasta la hipótesis de que la la proporción de marineros que son solteros es menor que el 90%

```
        1-sample proportions test without continuity correction

data:  rbind(.Table), null probability 0.9
X-squared = 2.7778, df = 1, p-value = 0.04779
alternative hypothesis: true p is less than 0.9
95 percent confidence interval:
 0.0000000 0.8997051
sample estimates:
    p
0.875
```

$$\chi^2 = Z^2$$

$$Z = \sqrt{2{,}7778} = \pm 1{,}67$$

Nota: habría sido equivalente realizar el contraste "la proporcion de marineros que son casados es mayor que el 10%

4. Una máquina produce latas de envasado para pimientos. Se selecciona una muestra aleatoria de 15 latas y se mide el diámetro de cada una. Los datos en mm son:

8.24	8.25	8.2	8.23	8.24	8.21	8.26	8.26
8.2	8.25	8.23	8.23	8.19	8.28	8.24	

   a) Comprueba que la distribución de los datos es normal mediante el gráfico de comparación de cuantiles.

   b) Calcula un intervalo de confianza al 95 % para la media de la variable diámetro. (8.2199,8.248)

   c) Contrasta si se puede aceptar que la media de los diámetros es 8.21 con un nivel de significación 0,05.
   tobs=3.674 > 2.145=t(14,0.025) Rechazar Ho
   p_valor=2*P(t(14>3.674)=2*0.0012 = 0.0025 < 0.05 Rechazar Ho

EJERCICIOS PARA ENTREGAR (ejercicios 5, 6 y 7)

HACER LOS EJERCICIOS DE LAS HOJAS DE PROBLEMAS PARA COMPROBAR RESULTADOS!!!!

5. El instituto de meteorología de un país proporciona la media anual de precipitaciones, en mm (espesor en mm de la capa de agua acumulada sobre un suelo horizontal≈ l/m^2), durante 20 años consecutivos:

499.2	555.2	398.8	391.9	453.4	459.8	483.7	417.6	469.2	452.4
499.3	340.6	522.8	469.9	527.2	565.5	584.1	727.3	558.6	338.6

   a) Comprueba si los datos siguen una ley normal.

   b) Proporciona un intervalo de confianza al 95 % para la media anual de precipitaciones.

   c) Si un país se considera árido cuando la media anual de precipitaciones es inferior a 420mm; efectúa un contraste que permita aceptar o rechazar que este país sea árido.

6. Una muestra aleatoria de 500 consumidores de vino participaron en una cata de vinos promovida por de la marca "Bodega". En la cata solo uno de los vimos era de esta marca. Cuando se les preguntó cuál era el vino preferido de todos los catados, 180 de ellos eligieron el de esta marca.

   a) Proporciona un intervalo de confianza al 95 % para que representa la probabilidad de que un consumidor elija este vino.

   b) Efectúa el contraste con hipótesis alternativa de que el porcentaje de consumidores que eligen este vino es mayor del 35 %.

   c) ¿cuántos catadores deberán participar a la prueba para obtener un intervalo de confianza al 95 % con un error máximo de 0.02?

   Nota: Para cargar de forma rápida los datos, debes utilizar el plug-in IPSUR, siguiendo el menú *estadísticos–proporciones–enter table for single sample.*

7. Se estudia la fracción de plantas de maíz de genotipo A. Un modelo genético sugiere que la proporción de plantas con genotipo A es mayor que 1/3. Se hace un análisis de genotipo en una muestra aleatoria de 300 plantas, la cual da como resultado 130 de genotipo A.

   a) Construye un intervalo de confianza al 95 % y establece conclusiones.

   b) ¿Qué tamaño muestral hubiera sido necesario para mantener el margen de error pero aumentar el nivel de confianza al 99 %?

   c) Efectúa el contraste de hipótesis correspondiente planteando como hipótesis alternativa la sospecha del enunciado. Asume $\alpha = 0{,}05$.

   d) También se pide estimar con un intervalo al 99 % de confianza la proporción de plantas que tienen el hongo cuitlacoche (Ustilago maydis). Se sabe que no superará el 20 % pero no se conoce exactamente la prevalencia. En la misma muetra han aparecido 40 plantas con el hongo.

   e) ¿ Qué tamaño muestral hubiera sido necesario si en el apartado anterior hubiéramos exigido un margen de error máximo del 2 %?

3. Inferencia cuando son necesarias las formulas de estadísticos (no hay menús)

Hay que tener en cuenta que los menus resuelven solo algunos de los casos vistos en clase y además que si la muestra no se da dato a dato, sino que se dan los valores muestrales de las v.a. $\overline{X}$ y S ya calculados, no se puede resolver el problema mediante los menus.

Cerramos la práctica resolviendo problemas de inferencia no contemplados en los menús. En tales casos, R se utiliza a modo de calculadora para obtener los valores que nos faltan en el intervalo correspondiente.

A modo de ejemplo mostramos dos intervalos.

1. Intervalo y contraste para la <u>media con varianza conocida</u>.

   Un ingeniero ha diseñado una nueva correa de distribución para el apero de una distribuidora de semillas en de tractor. En su opinión durará 150000km de recorrido de sembrado. Según los históricos la desviación de la variable recorrido es de 1000km. Un proveedor toma una muestra de 80 correas del nuevo diseño, comprobadas en laboratorio, da como resultado una vida media $\bar{x} = 148000$km. Construye un intervalo de confianza y establece conclusiones al 90 %.

   Puesto que el intervalo correspondiente es $\bar{x} \pm z_{\alpha/2}\ \sigma/\sqrt{n}$

   Con R-commander calculamos: $z_{0{,}05} = 1{,}6448$

   El intervalo es, por tanto, $(148000 - 1{,}6448 * 1000/sqrt(80), 148000 + 1{,}6448 * 1000/sqrt(80)) = (147816, 148184)$

2. Intervalo de confianza para la <u>varianza en una población normal</u>.

   Una marca de margarina dietética fue analizada para determinar el nivel de grasas poliinsaturadas (en porcentajes). Un muestra de seis tarrinas proporcionó los siguientes datos: 16.8, 17.2, 17.4, 16.9, 16.5, 17.1.

   Si asumimos normalidad de la variable en estudio, calcula un intervalo al 95 % para la varianza de la variable nivel de grasas.

   Para obtener el intervalo calculamos la desviacion tipica de la muestra con el menu *estadísticos–resumenes numéricos* y los valores de $\chi^2_{n-1;0{,}025}$ y $\chi^2_{n-1;0{,}975}$.

   El intervalo es: $(\frac{s^2(n-1)}{\chi^2_{\alpha/2}}, \frac{s^2(n-1)}{\chi^2_{1-\alpha/2}}) =$ (0.03961,0.61156)