

Inferencia estadística

Sesión 3. Curso: 'Análisis estadístico aplicado con R Commander'

Miguel A. Martinez-Beneito
correo: martinez_mig@gva.es

Inferencia estadística ¿qué es?

En un estudio estadístico siempre tendremos una muestra de una población y nuestro objetivo será obtener conclusiones de la población a partir de nuestra muestra.

Dada una muestra de una población, un modelo estadístico es una conceptualización matemática del proceso que se supone que genera los datos.

Ejemplo

Disponemos de los valores de colesterol en sangre para una muestra de niños (x_1, \dots, x_n) obtenida de varios centros educativos.

Consideramos razonable asumir que dichos valores se distribuyen según una distribución Normal de media μ y varianza σ^2 desconocidas.

Esta última suposición es el modelo estadístico asumido para los datos.

Ejemplo

Disponemos del número observado de defunciones (y_1, \dots, y_m) por cáncer de pulmon en la Comunidad Valenciana para cada uno de los años del periodo 2000-2019.

Asumimos que las observaciones se distribuyen según una distribución de Poisson de parámetro λ desconocido.

La inferencia estadística permite estimar los parámetros (poblacionales) de un modelo estadístico, μ, σ^2 y λ en los ejemplos anteriores, a partir de la información de la muestra disponible.

La inferencia estadística dispone de 3 herramientas fundamentales para llevar a cabo su objetivo:

- Estimación puntual.
- Intervalos de confianza.
- Contrastes de hipótesis.

A continuación describimos cada una de estas 3 herramientas.

Estimación puntual

Su objetivo es obtener estimadores, que son funciones de los valores de la muestra, que aproximen en la medida de lo posible un parámetro (poblacional) del modelo correspondiente.

Para cada uno de los ejemplos anteriores, el objetivo de la estimación puntual será obtener tres funciones: $\hat{\mu} = f_1(x_1, \dots, x_n)$, $\hat{\sigma}^2 = f_2(x_1, \dots, x_n)$ y $\hat{\lambda} = f_3(y_1, \dots, y_m)$ que estimen los valores (desconocidos) de μ , σ^2 y λ , respectivamente.

Muchos de los parámetros de los modelos estadísticos tienen un equivalente muestral que en general serán buenos estimadores del parámetro que queremos conocer.

Ejemplo

Para los datos de colesterol hemos de estimar μ y σ^2 , la media y varianza de la correspondiente población, que podremos estimar con la media y varianza muestral de la muestra que dispongamos:

Ejemplo

$$\hat{\mu} = \bar{x} = \left(\sum_{i=1}^n x_i \right) / n; \quad \hat{\sigma}^2 = s^2 = \left(\sum_{i=1}^n (x_i - \bar{x})^2 \right) / (n - 1)$$

De la misma forma, en los datos de mortalidad, el parámetro λ de la distribución de Poisson es su media poblacional, por tanto podemos estimar dicho valor con la media muestral de los datos:

$$\hat{\lambda} = \bar{y} = \left(\sum_{i=1}^m y_i \right) / m$$

De manera paralela a los equivalentes muestrales de los parámetros de los modelos, existen estimadores alternativos que pueden tener incluso mejores propiedades que los estimadores anteriores.

Ejemplo



Picanúmeros
@Picanumeros

En 2016, un municipio barcelonés tenía una renta bruta media de poco menos de 28.000 € (fuente: Agencia Tributaria). Al año siguiente, la renta media se disparó hasta casi 58.000 €. En el pueblo sospechan que es culpa de una sola persona.

USAD LA MEDIANA



EL PAÍS

Avinyonet del Penedès: el misterio del pueblo que se hizo rico de golpe
Los habitantes de la pequeña localidad barcelonesa atribuyen ser los segundos con más renta a un vecino empresario

elpais.com

La mediana de una muestra es un estimador de μ más robusto (se ve menos influido por observaciones anómalas) que la media muestral.

Estimadores puntuales con R Commander

En la sesión anterior vimos distintos estadísticos que resumían distintas características de muestras y cómo se pueden calcular con R Commander.

Los menús Estadísticos> Resúmenes> Conjunto de datos activo o Resúmenes numéricos o Tabla de Estadísticas son muy útiles.

Ejemplo

El archivo `colesterol.Rdata` contiene los datos de colesterol que ya mencionamos. El menú Resúmenes numéricos nos proporciona tanto la media como la desviación típica muestral, por tanto: $\hat{\mu} = \bar{x} = 171.8$ y $\hat{\sigma}^2 = s^2 = 19.0^2$.

A partir de estos valores podemos asumir que la distribución del colesterol en sangre en la población es: $N(171.8, 19.0^2)$ y de ella podemos conocer, por ejemplo, la probabilidad de que el colesterol de un niño supere las 200 unidades: 0.069 (Menú: Distribuciones> Distribuciones continuas> Distribución Normal> Probabilidades ...).

Intervalos de confianza

Los estimadores puntuales proporcionan una primera estimación de los parámetros del modelo de indudable utilidad. Sin embargo, dichos estimadores ignoran completamente la variabilidad de la estimación.

Los intervalos de confianza suponen la principal herramienta para cuantificar e incorporar la variabilidad a los estimadores puntuales.

Se define el intervalo de confianza (IC) de nivel $100 \cdot (1 - \alpha) \%$ como aquel intervalo, construido a partir de la muestra disponible, con probabilidad del $100 \cdot (1 - \alpha) \%$ de contener el verdadero valor del parámetro que pretende estimar.

La probabilidad de error, α , del IC se dice significación y se suele fijar al 5% (nivel de confianza del 95 %) o al 1 % (nivel de confianza del 99 %).

Intervalo de confianza para la media de una población:

El teorema central del límite nos asegura que la media muestral \bar{x} de cualquier muestra se distribuye según una distribución Normal, de media μ -la media poblacional-, y de varianza σ^2/n , donde σ^2 es la varianza de los datos y n el tamaño de la muestra.

$$\bar{x} \sim N(\mu, \sigma^2/n) \rightarrow \bar{x} - \mu \sim N(0, \sigma^2/n) \rightarrow \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \quad (1)$$

Sabemos que -1.96 y 1.96 son los percentiles al 2.5 % y al 97.5 % de cualquier variable $N(0, 1)$, por tanto $\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ tendrá probabilidad del 95 % de estar en el intervalo $[-1.96, 1.96]$.

Entonces con probabilidad del 95 %:

$$\begin{aligned} \bar{x} - \mu &\in [-1.96 \cdot \sigma/\sqrt{n}, 1.96 \cdot \sigma/\sqrt{n}] \rightarrow \\ -\mu &\in [-\bar{x} - 1.96 \cdot \sigma/\sqrt{n}, -\bar{x} + 1.96 \cdot \sigma/\sqrt{n}] \rightarrow \\ \mu &\in [\bar{x} - 1.96 \cdot \sigma/\sqrt{n}, \bar{x} + 1.96 \cdot \sigma/\sqrt{n}] \end{aligned}$$

Ese último intervalo cumple las propiedades que esperábamos, por tanto es el IC al 95 % para μ que esperábamos.

Intervalos de confianza, el caso general:

El procedimiento general de cálculo de IC es muy similar a lo que hemos visto.

- 1 Determinar una función de los valores de la muestra y los parámetros poblacionales, de la que conozcamos completamente su distribución (ver final de la expresión (1)). Dicha función se conoce como pivote.
- 2 Determinar en la distribución anterior un intervalo que contenga la probabilidad correspondiente al nivel del IC (habitualmente 95 o 99 %).
- 3 Despejar del pivote el parámetro que queramos estimar.

Ejemplo

Queremos calcular un IC al 95 % para el nivel medio de colesterol en niños a partir del banco de datos `colesterol.Rdata`.

Según vimos, para calcular un IC para la media podemos utilizar la relación:

$$\frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

que proporciona el IC que acabamos de ver.

Sin embargo, dicho intervalo depende de σ^2 que no conocemos. Podríamos estimar σ^2 mediante s^2 , en cuyo caso la relación anterior ya no es válida, sino:

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

donde t_{n-1} es la distribución t de Student con $n - 1$ grados de libertad, en nuestro caso $n = 40$.

Ejemplo

Los percentiles al 2.5 y 97.5 % de la distribución t_{39} son, respectivamente -2.02 y 2.02. Así, con probabilidad del 95 %:

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \in [-2.02, 2.02] \rightarrow \dots \rightarrow \mu \in [\bar{x} - 2.02 \cdot s/\sqrt{40}, \bar{x} + 2.02 \cdot s/\sqrt{40}]$$

Teniendo en cuenta que $\bar{x} = 171.8$ $s = 19.0$, entonces $\mu \in [165.8, 177.9]$ con probabilidad del 95 %.

Los IC se suelen utilizar para valorar si existen diferencias 'relevantes' (o en términos estadísticos 'significativas') entre el parámetro poblacional que desconocemos y cierto valor de referencia.

A la vista del ejemplo anterior podemos concluir que μ es significativamente distinto de 150, pero no de 170.

Intervalos de confianza en R Commander

R Commander no dispone de una función genérica que, de forma general, devuelva ICs para cualquier modelo estadístico sino que los ICs suelen ser parte de la salida de un gran número de menús.

Simplemente:

- Qué menú de R Commander corresponde al modelo que queremos ajustar, y de cuyos parámetros queremos obtener sus IC.
- Hemos de saber interpretar de forma correcta un IC cuando nos aparezca como parte de la salida de una función.

Ejemplo (IC para la media de una población Normal)

Para el banco de datos `colestero1` estamos interesados en obtener un IC para el colesterol medio en población infantil.

Ejemplo

El menú Estadísticos> Medias> Test t para una muestra compara dicha media con cierto valor concreto que consideremos de referencia y como parte de la salida calcula el IC para la media.

```
> with(colesterol, (t.test(valores, alternative = "two.sided", mu = 0, conf.level = 0.95)))

One Sample t-test

data:  valores
t = 57.281, df = 39, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 165.7817 177.9183
sample estimates:
mean of x
 171.85
```

Ejemplo (IC para la diferencia de medias en poblaciones Normales)

El banco de datos `colesterol` contiene una segunda variable con el centro educativo de cada niño. Nos gustaría conocer el IC para $\mu_1 - \mu_2$, la diferencia de las medias de ambos centros.

Ejemplo

El menú Estadísticos > Medias > Test t para muestras independientes valora las diferencias de las medias de dos poblaciones Normales, en su salida se muestra el IC para $\mu_1 - \mu_2$.

```
> t.test(valores ~ centro, alternative = "two.sided", conf.level = 0.95, var.equal = FALSE, data = colesterol)

Welch Two Sample t-test

data:  valores by centro
t = 2.5036, df = 37.925, p-value = 0.01672
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.698056 25.501944
sample estimates:
mean in group 1 mean in group 2
      178.9         164.8
```

Dicho intervalo sugiere que existen diferencias significativas entre μ_1 y μ_2 .

Ejemplo (IC para una proporción)

El banco de datos colesterol contiene una tercera variable que señala aquellos niños con colesterol superior a 200 unidades. Nos gustaría conocer el IC para π , la probabilidad de que un niño presente un exceso de colesterol.

Ejemplo

El menú Estadísticos> Proporciones> Test de proporciones para una muestra compara dicha proporción con cierto valor de referencia y como parte de la salida calcula el IC para π .

```
Frequency counts (test is for first level):
exceso
  Exceso No exceso
      2         38

      1-sample proportions test without continuity correction

data:  rbind(.Table), null probability 0.5
X-squared = 32.4, df = 1, p-value = 0.00000001255
alternative hypothesis: true p is not equal to 0.5
95 percent confidence interval:
 0.01382067 0.16503877
sample estimates:
      p
0.05
```


Contrastes de hipótesis

Más allá de la estimación que pudiéramos hacer de un parámetro, puntual o por intervalos, en ocasiones queremos responder ciertas preguntas concretas sobre ellos.

- ¿Es la media de mi población distinta de la de la población general?
- ¿Son las medias de los 2 poblaciones que manejo distintas entre sí?

La herramienta adecuada para responder a dichas preguntas son los contrastes de hipótesis.

Para llevar a cabo un contraste de hipótesis habremos de precisar los siguientes elementos:

- Hipótesis a contrastar.
- Significatividad del contraste.

Hipótesis a contrastar:

Para formular un contraste de hipótesis habremos precisar la hipótesis que queremos contrastar, hipótesis nula que denotaremos H_0 .

Alternativamente a la hipótesis nula tendremos una hipótesis alternativa (denotada como H_1) que daremos por buena si H_0 es rechazada.

$H_0 \cup H_1$ debería considerar todos los valores plausibles del parámetro a contrastar

Ejemplo

- Pretendemos demostrar que la ingesta de un fármaco altera los valores de glucosa en sangre de los pacientes que lo consumen. Conocemos que el valor medio de glucosa en sangre en población general es de 90mg/dl. Si μ es el nivel medio de glucosa de los potenciales consumidores del fármaco:

$$\left. \begin{array}{l} H_0 : \mu = 90 \\ H_1 : \mu \neq 90 \end{array} \right\}$$

Ejemplo

- Pretendemos conocer si un fármaco para la diabetes es efectivo o no para reducir la glucosa en sangre. Si μ_a y μ_d son los niveles de glucosa en sangre en la población antes y después de tomar el fármaco:

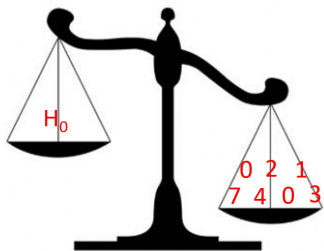
$$\left. \begin{array}{l} H_0 : \mu_a = \mu_d \\ H_1 : \mu_a > \mu_d \end{array} \right\}$$

El primero de los ejemplos se dice contraste bilateral y el segundo unilateral, en función de la forma asumida de la hipótesis alternativa.

Las hipótesis nula y alternativa no juegan un papel equivalente en un contraste de hipótesis.

La hipótesis nula es la que es puesta en tela de juicio y finalmente será rechazada o aceptada.

H_1 simplemente se asume cierta si H_0 es rechazada.



H_0 , como en todo juicio, goza de “presunción de inocencia”, no la rechazaremos a menos que encontremos (mejor dicho, los datos muestren) suficiente evidencia en su contra.

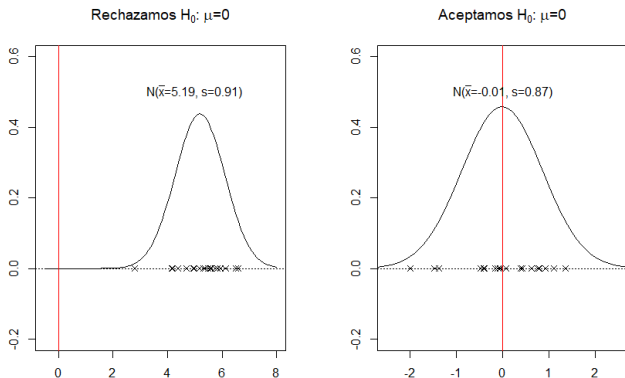
Por el contrario, el aceptar H_0 no quiere decir que sea cierta (sólo que no hemos encontrado evidencia en su contra).



Aquello que queramos probar deberá ser establecido como H_1 en un contraste ya que en caso de rechazar H_0 daremos como probado H_1 . En cambio aceptar H_0 no implica que sea cierta por lo que no nos lleva a nada concluyente.

- H_1 aquello que queremos demostrar.
- H_0 en general representa una igualdad.

Cuantificaremos la evidencia en contra de H_0 en función de la compatibilidad de dicha hipótesis con los datos observados.



Los contrastes de hipótesis son una formalización matemática de esta idea.

Significatividad del contraste

Tras formular las hipótesis del contraste dispondremos de una hipótesis nula, que no sabremos si es cierta o no, y que deberemos rechazar o aceptar en función de los datos.

	H_0 aceptada	H_0 rechazada
H_0 verdadera	Ok (verdadero positivo)	Error tipo I (falso positivo)
H_0 falsa	Error tipo II (falso negativo)	Ok (verdadero negativo)

No sabremos si la decisión tomada sobre H_0 es correcta o no, lo que si podremos hacer es fijar la probabilidad de equivocarnos al rechazar H_0 si ésta fuera cierta, es decir podremos fijar la probabilidad de error de tipo I.

La probabilidad de error de tipo I que estamos dispuestos a asumir en un contraste (denotado como α) se conoce como su significación, que suele ser fijada al 5 % o al 1 %.

La probabilidad de error de tipo II no se fija en un contraste sino que, dada su significación, será función de la potencia estadística de los datos.

Contraste de hipótesis: procedimiento

- Una vez determinadas las hipótesis y significación del contraste buscaremo un pivote (al igual que para construir un IC) que dependa de la muestra, del parámetro de interés y conozcamos su distribución.
- Dados estos elementos, y asumiendo como cierta H_0 , determinaremos la región del pivote con probabilidad α y más favorable a H_1 .
Llamaremos a esa región: *región de rechazo del contraste*.
- Valoraremos si el pivote propuesto cae en la región de rechazo, o no, en cuyo caso rechazaremos la hipótesis nula.

Ejemplo

Consideramos de nuevo los datos de colesterol y queremos ahora valorar si la media poblacional de la variable es distinta de 165, el colesterol medio de los niños a nivel nacional.

Ejemplo

Nos planteamos una significatividad $\alpha = 5\%$. Además, nuestras hipótesis serán:

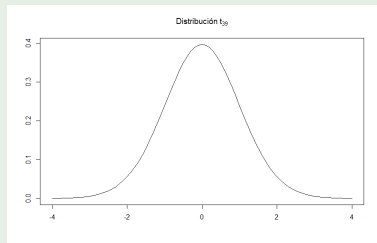
$$\left. \begin{array}{l} H_0 : \mu = 165 \\ H_1 : \mu \neq 165 \end{array} \right\}$$

Para el cálculo del IC para μ empleamos la siguiente variable:

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

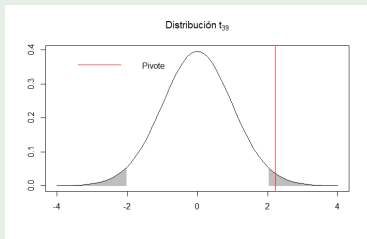
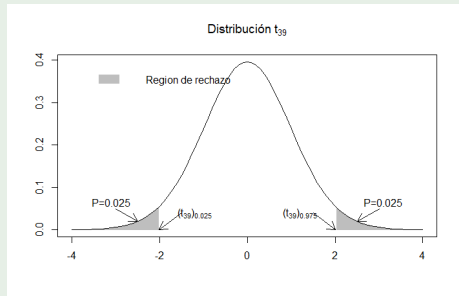
que nos puede servir también ahora como pivote.

La distribución t_{39} tiene esta forma, por lo que sus valores menos probables son los que se hallan alejados del cero.



Ejemplo

La región que corresponde a los valores más compatibles con esta distribución y con probabilidad del 95 % es la delimitada por los percentiles al 2.5 % y al 97.5 % de la distribución.



Por último, habremos de calcular el pivote asumiendo $H_0 : \mu = 165$, $((171.8 - 165)/(19/\sqrt{40}) = 2.23)$, y valorar si cae o no en la región de rechazo.

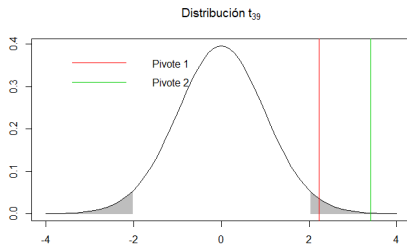
Por tanto rechazamos H_0 y concluimos entonces $H_1 : \mu \neq 165$.

P-valor

Los contrastes de hipótesis simplemente concluyen si podemos aceptar o no la hipótesis nula, dada cierta significación.

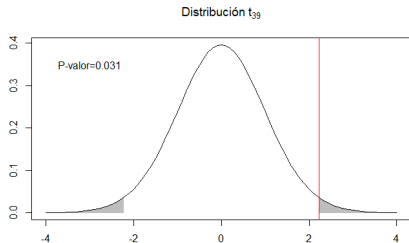
Sin embargo, podemos encontrar niveles muy distintos de evidencia en contra de una hipótesis nula.

Los P-valores cuantifican de forma precisa esos niveles.



Los P-valores se definen como la probabilidad que tendríamos de obtener el pivote disponible, o una más extremo, si fuera cierta H_0 .

Alternativamente, se entiende el P-valor como la probabilidad que asumiríamos de equivocarnos en caso de rechazar H_0 .



Se calculan haciendo coincidir el límite de la región de rechazo con el valor del pivote que tenemos. El tamaño de la región de rechazo correspondiente se define como el P-valor.

En general si el P-valor es menor que la significación asumida, se rechaza H_0 , y si no se aceptará.

En cualquier caso el valor del P-valor es bastante más informativo que el simplemente aceptar o rechazar H_0 .

Ejemplo

Prestamos atención de nuevo a las diferencias del colesterol según centro educativo. Queremos ahora contrastar si hay diferencias significativas ($\alpha = 0.05$) entre el nivel medio de colesterol en ambos centros.

Ejemplo

Nuestras hipótesis serán:

$$\left. \begin{array}{l} H_0 : \mu_1 = \mu_2 \\ H_1 : \mu_1 \neq \mu_2 \end{array} \right\}$$

Aunque en su momento no lo comentamos el pivote que se utilizó para el cálculo del IC fue:

$$\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_1^2/n_1 + s_2^2/n_2}} \sim t_{38}$$

Asumiendo H_0 , dicho pivote vale:

$$\frac{(178.9 - 164.8) - 0}{\sqrt{331.2/20 + 303.1/20}} = 2.504$$

Para calcular el P-valor correspondiente deberíamos calcular el tamaño de la cola de la distribución t_{38} a la derecha de 2.504 y multiplicarlo por 2 (cola izquierda de la distribución).

Ejemplo

La cola de la distribución mencionada tiene un tamaño de 0.0084 (Menú: Distribuciones> Distribuciones continuas> Distribución t> Probabilidades t acumuladas) por lo que el P-valor del contraste será 0.0168.

Como el P-valor es inferior a la significatividad del contraste (0.05) rechazaremos H_0 .

Estos resultados ya los habíamos reproducido al calcular el IC para la diferencia de medias para estos datos

```
> t.test(valores ~ centro, alternative = "two.sided", conf.level = 0.95, var.equal = FALSE, data = colesterol)

Welch Two Sample t-test

data:  valores by centro
t = 2.5036, df = 37.925, p-value = 0.01672
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.698056 25.501944
sample estimates:
mean in group 1 mean in group 2
      178.9         164.8
```

Contrastes de hipótesis en R Commander

R Commander no dispone de una función genérica para realizar contrastes de hipótesis.

Un buen número de menús de R Commander han sido desarrollados exclusivamente para llevar a cabo contrastes de hipótesis o, éstos son parte habitual de la salida de muchos otros menús.

Ya hemos visto cómo la comparación de medias en poblaciones Normales puede ser hecha llamando a cierto menú de R Commander lo que nos ahorra tener que llevar a cabo todo el proceso anterior.

Los ejemplos de comparación de la media del colesterol con un valor de referencia y el porcentaje de exceso de colesterol son a su vez salidas de menú dedicados a hacer contrastos de hipótesis.

Cuando R Commander haga un contraste de hipótesis por nosotros no debemos preocuparnos de buscar un pivote ni nada parecido. Simplemente deberemos conocer la hipótesis que se están contrastando y saber interpretar sus resultados.

Test de comparación de proporciones

R Commander dispone de un test de comparación de proporciones para dos grupos en el menú: Estadísticos> Proporciones> Test de proporciones para dos muestras

Ejemplo

En el banco de datos `colestero1` disponemos de la variable binaria exceso y la variable centro que divide el banco de datos en 2 grupos. Nos planteamos si el porcentaje de excesos de niños con `colestero1` difiere entre ambos grupos:

```
Percentage table:
      exceso
centro Exceso No exceso Total Count
1      10      90      100      20
2       0     100      100      20

2-sample test for equality of proportions without continuity correction

data: .Table
X-squared = 2.1053, df = 1, p-value = 0.1468
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.03147838  0.23147838
sample estimates:
prop 1 prop 2
 0.1   0.0
```

No encontramos diferencias significativas entre ambos grupos.

Test t de comparación de medias para muestras pareadas

La comparación de dos medias o proporciones no es del todo eficiente ya que queremos comparar dos cantidades que no conocemos con detalle.

Sin embargo, si los individuos de ambas muestras están relacionados (por ejemplo: individuos medidos antes y después de cierta intervención, estudios de casos y controles . . .) podremos comparar ambas poblaciones de manera más efectiva.

Compararemos cada una de las observaciones de cada grupo con la respectiva observación del otro grupo en lugar de comparar las medias de ambas poblaciones.

El contraste de hipótesis resultante es en general bastante más potente.

Ejemplo

En el banco de datos `colesterol` disponemos de una variable adicional TrasDieta que contiene el valor del colesterol para cada individuo tras someterse a una dieta específica de 2 semanas.

Ejemplo

Queremos conocer si la dieta ha tenido algún efecto sobre el colesterol medio de los niños.

Hacemos un test t pareado (Estadísticos> Medias> Test t para datos relacionados):

```
> with(colesterol, (t.test(valores, TrasDieta, alternative = "two.sided", conf.level = 0.95, paired = TRUE)))

Paired t-test

data:  valores and TrasDieta
t = 5.9261, df = 39, p-value = 0.0000006559
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.914674 5.935326
sample estimates:
mean of the differences
      4.425
```

Encontramos que antes de la dieta los valores de colesterol son 4.4 unidades superiores que después de la dieta, además la diferencia encontrada entre ambos grupos es muy significativa.

Ejemplo

Si hubiéramos hecho un test t no pareado habríamos obtenido lo siguiente*.

```
> with(cholesterol, (t.test(valores, TrasDieta, alternative = "two.sided", conf.level = 0.95, paired = FALSE)))

Welch Two Sample t-test

data:  valores and TrasDieta
t = 1.0004, df = 77.504, p-value = 0.3202
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.382082 13.232082
sample estimates:
mean of x mean of y
 171.850   167.425
```

Donde las diferencias entre los valores antes y después de la dieta no serían significativas.

* El test no pareado lo hemos hecho editando el comando de R ya que desde los menús de R Commander ese test requiere que todas las observaciones pertenezcan a una misma variable y tengamos una variable que defina los grupos que queremos contrastar.

Test de comparación de varianzas

R Commander también dispone (menú: Estadísticos> Varianzas> Test de Levene) de tests de comparación de varianzas para distintos grupos de observaciones.

Ejemplo

Para los datos de `colesterol` nos planteamos si podemos asumir o no que las varianzas de las observaciones de ambos centros son iguales.

```
> leveneTest(valores ~ centro, data = colesterol, center = "median")
Levene's Test for Homogeneity of Variance (center = "median")
      Df F value Pr(>F)
group  1  0.0581 0.8108
      38
```

No rechazamos que las varianzas sean iguales, por tanto daremos por buena esa hipótesis (que no quiere decir que lo sea).