

CONTRASTE DE HIPÓTE- SIS

Selección de actividades re-
seltas

© Jose Fco. Martínez Boscá, Arnau Mir Torres, Lluís M. Pla
Aragonés, Àngel J. Gil Estallo (Autors) & Àngel A. Juan (Editor)

© FUOC 2009

Introducción

La utilización de los contrastes de hipótesis estadísticos son sumamente frecuentes en la toma de decisiones en diversos campos de las ciencias, aplicadas a múltiples problemas: los económicos, (para estimar, por ejemplo, las ventas futuras de una empresa), los biológicos (probar la efectividad de una droga nueva), los industriales (controles de producción), los médicos e inclusive aquellos vinculados a las Ciencias de la Educación. Los contrastes estadísticos de hipótesis son un tema que pertenece a la rama de la Estadística Inferencial. La comprensión de dichos contrastes es relevante para su utilización dentro de variados campos disciplinares. Según Gardner, puede decirse que la estadística actúa como “puente entre las ciencias naturales y sociales” (1997, 171).

Mediante la construcción de muestras se pretende dar cuenta de cuál es el comportamiento de la población a partir de un conjunto limitado de casos para tomar finalmente decisiones sobre todo el universo. Es así como el contraste estadístico de hipótesis adquiere relevancia, proporcionándonos una herramienta que permite establecer conclusiones acerca de fenómenos poblacionales a partir de los datos disponibles, que habitualmente son de tipo muestrales. Pero para que dichos procedimientos puedan ser llevados a cabo de manera efectiva, es imprescindible conocer la lógica del proceso y los errores que pueden cometerse en la consecuente toma de las decisiones, a fin de realizar una correcta lectura de los resultados a los que se han llegado.

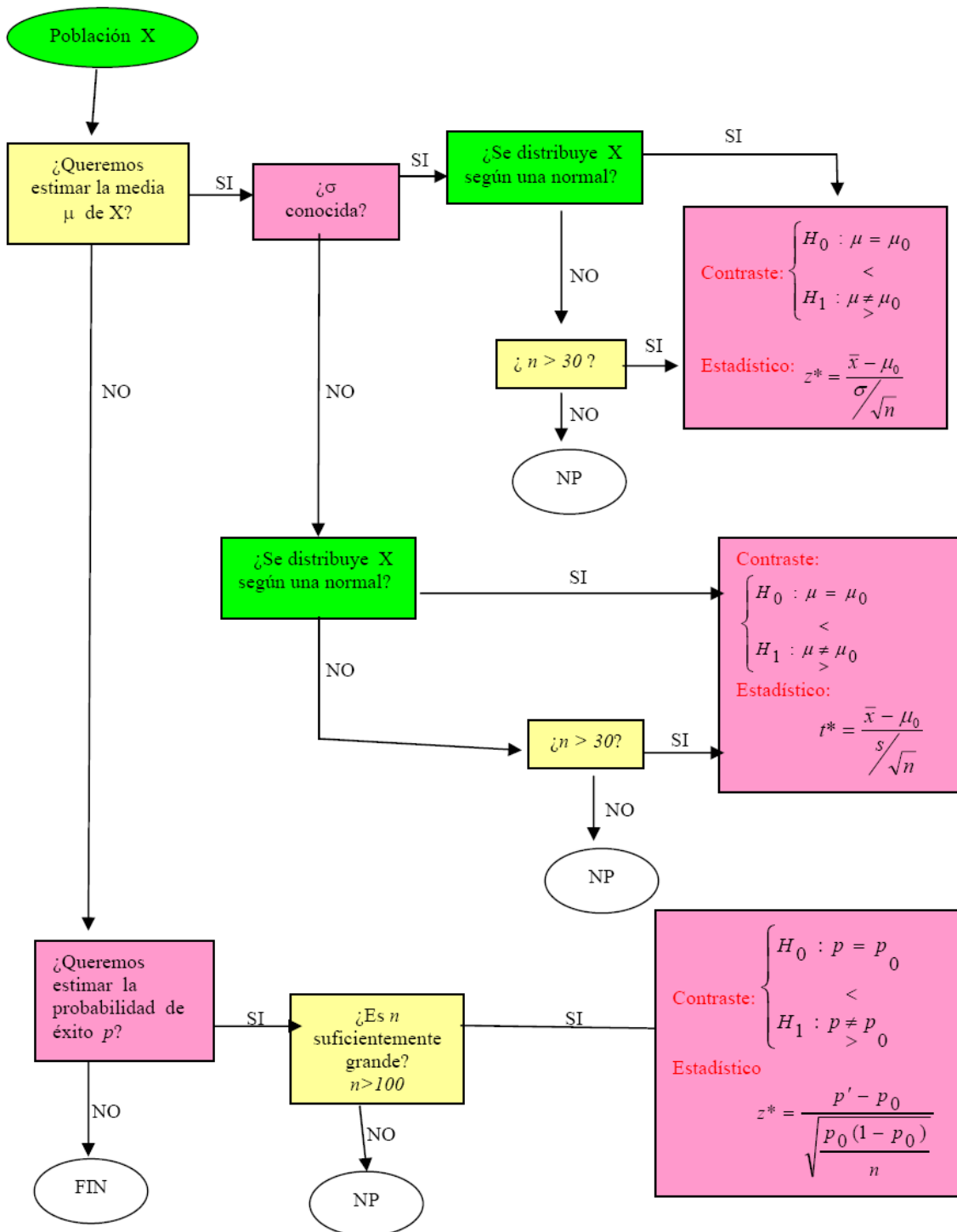
En este módulo se pretende entender qué es y para qué se utiliza un contraste de hipótesis, así como saber calcular e interpretar los estadísticos de contraste y los p-valores a la hora de realizar dichos contrastes para la media poblacional, sea o no conocida la desviación estándar poblacional y para las proporciones.

También haremos contrastes sobre la media de muestras emparejadas. El que dos muestras sean dependientes (emparejadas) o no viene determinado por las fuentes (personas u objetos) que nos aportan las observaciones. Si en la obtención de ambas muestras se han utilizado las mismas fuentes o fuentes asociadas, tendremos dos **muestras dependientes**. Por el contrario, si se han utilizado fuentes completamente distintas hablaremos de muestras independientes.

Supongamos que, al iniciar el semestre, seleccionamos al azar 30 alumnos matriculados en Estadística y les pasamos un test de conocimientos previos. Al final del semestre, seleccionamos otros 30 alumnos al azar y les pasamos un test de conocimientos adquiridos durante el curso. En tal caso, consideraríamos ambas muestras como independientes. Por el contrario, si el test de conocimientos adquiridos se realizase a los mismos 30 alumnos que hicieron el test inicial, entonces hablaríamos de muestras dependientes. Otro ejemplo sería: Se cuestiona si el tiempo de arranque de un ordenador con el antivirus TRONX es más lento que si el ordenador no tiene el antivirus. Con el objetivo de hacer este estudio se ha medido el tiempo de arranque de 10 ordenadores sin tener ningún antivirus instalado. Después, se ha instalado el antivirus en todos estos ordenadores y se ha vuelto a medir el tiempo de arranque. Suponiendo que el tiempo de arranque de un ordenador se distribuye normalmente, decide con $\alpha=0,05$ si la existencia en el ordenador del antivirus hace que los tiempos de arranque sea mayor. Encontrad el p-valor del contraste.

Mapa conceptual

CONTRASTE DE HIPÓTESIS



NP significa que tenemos que hacer servir métodos. No Paramétricos (fuera del contenido del curso)

Actividades

Actividad 1: Estudio sobre la velocidad de un proveedor de Internet.

Contraste de hipótesis paramétricos. Contraste de la media de una población normal con varianza desconocida. Estimación por intervalos: Intervalo de confianza. Software R.

En un test para medir la velocidad de bajada de un proveedor de internet los datos obtenidos de 10 usuarios, en kbps, fueron los siguientes:
150.8, 234, 260, 235.4, 280, 276, 200, 300, 256, 190

Dicho proveedor de Internet publicita que su velocidad de bajada es de 256 kbps. Suponiendo que la velocidad sigue una distribución normal, contrastad la afirmación de la empresa, al nivel de significación del 5%. Calculad un intervalo de confianza para μ al 95% de nivel de confianza.

Para contestar a estas pregunta plantear la hipótesis nula y alternativa, dar la fórmula del estadístico de contraste así como su distribución de probabilidad. Encontrad el p-valor del contraste y el intervalo de confianza correspondiente. Comprobad que las conclusiones son las mismas con el p-valor y con el intervalo.

Solución

$$H_0: \mu = 256$$

$$H_1: \mu \neq 256$$

$$t^i = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{238.2 - 256}{\frac{46.2}{\sqrt{10}}} = \frac{-17.8}{14.61} = -1.22$$

El estadístico del contraste sería :

Debemos calcular ahora el p-valor, es decir la probabilidad de obtener un valor igual o inferior a -1.22 o superior a 1.22 en una distribución t-student con 9 grados de libertad (10-1).

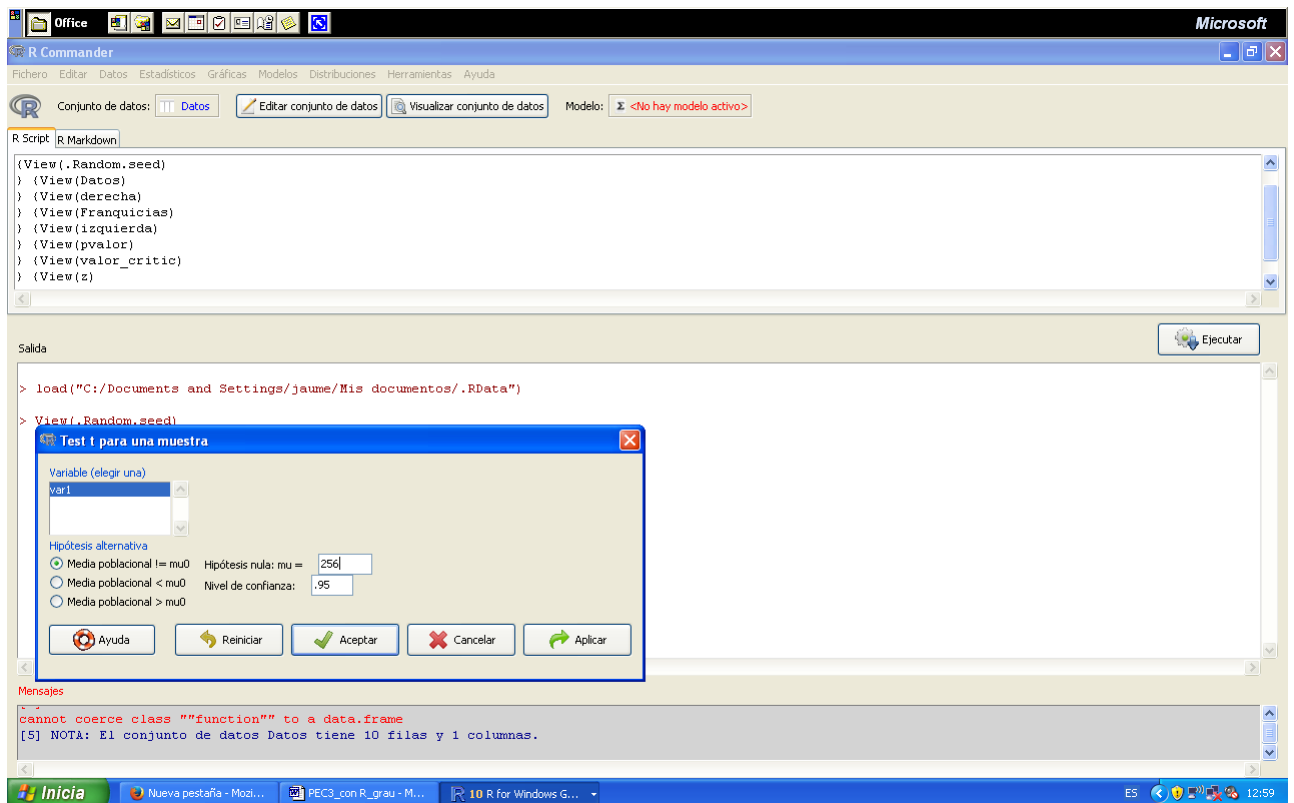
Usando las tablas estadísticas de la distribución t-student con 9 grados de libertad podemos aproximar esa probabilidad por: $p\text{-valor} = 2 * 0.125 = 0.250$

$$\bar{x} \pm t_{9,0.025} \frac{s}{\sqrt{n}}$$

El intervalo de confianza al 95% es :

A continuación lo hacemos con R:

Vamos a R Comander: Estadísticos > Medias y le damos a Test t para una muestra y elegimos la variable: var1, marcamos Media poblacional μ_0 Hipótesis nula: $\mu = 256$ y nivel de confianza: 95



One Sample t-test

```

data: Datos$var1
t = -1.2162, df = 9, p-value = 0.2548
alternative hypothesis: true mean is not equal to 256
95 percent confidence interval:
 205.1497 271.2903
sample estimates:
mean of x
 238.22

```

Como el p-valor (0.25) es más grande que el nivel de significación 0.05 aceptamos que $\mu=256$.

IC al 95% para μ es (205.15 ; 271.29) y como 256 está dentro del intervalo aceptamos también a hipótesis nula. Esto está de acuerdo con la conclusión que hemos sacado con el p-valor.

Actividad 2: Estudio sobre las compras realizadas por Internet en un portal de e-business.

Contraste de hipótesis paramétricos. Contraste para la proporción. Estimación por intervalos: Intervalo de confianza. Software Minitab.

Un portal de e-business sabe que el 60% de todos sus visitantes a la Web están interesados en adquirir sus productos, pero son reacios al comercio electrónico y no realizan finalmente la compra vía Internet.

En cambio, la dirección del portal sabe que en el último año, el porcentaje de gente que está dispuesta a comprar por Internet ha aumentado y esto se debe reflejar en sus resultados empresariales.

a) Se pide contrastar al nivel de significación del 2% si en el último año se ha reducido el porcentaje de gente que no está dispuesta a comprar por Internet. Para esto se cogió una muestra de 500 visitantes, para conocer su opinión, y se observó que el 55% no estaba dispuesto a realizar compras vía on-line.

b) En la empresa anterior se está realizando un estudio para determinar la cantidad mensual de horas de baja laboral, causadas por accidentes en el trabajo. Suponiendo un nivel de significación del 0.05, se ha realizado el siguiente contraste de hipótesis, para comparar si la media de horas es igual o inferior a 18, $H_0: \mu=18$ vs. $H_1: \mu<18$

Obtenemos el siguiente output:

```
One Sample t-test
data: variable$V1
t = 0.33, p-value = 0.63
alternative hypothesis: true mean is less than 18
sample estimates:
mean of x
18.129
```

¿Qué conclusiones sacarías?

Solución

a) Lo que se pide es contrastar lo siguiente:

$$H_0 : p = 0,6$$

$$H_A : p < 0,6$$

Podemos observar que se cumplen los supuestos de normalidad, puesto que la distribución del número de visitantes del portal que no están dispuestos a comprar vía Internet se aproximó a una normal debido a que $n * p_{H_0} \geq 5$ y $n * (1 - p_{H_0}) \geq 5$.

> prop.test(275, 500, alternative='less', p=.6, conf.level=.98, correct=FALSE)

```
1-sample proportions test without continuity correction
data: 275 out of 500, null probability 0.6
X-squared = 5.2083, df = 1, p-value = 0.01124
alternative hypothesis: true p is less than 0.6
98 percent confidence interval:
0.0000000 0.5950852
sample estimates:
p
0.55
```

Del resultado anterior se desprende que el p-valor es $0.01124 < 0.02$, entonces rechazaremos la hipótesis nula a un nivel de significación del 2%

En conclusión, existe evidencia estadística que la proporción de visitantes al portal que están dispuestos a comprar por Internet ha aumentado, es decir, que el porcentaje de visitantes que son reacios a comprar vía on-line ha disminuido.

b) Como el p-valor $0.63 > 0.05$, no podemos rechazar la hipótesis nula, es decir, asumiremos como posible la opción de que el número medio de días de baja laboral sea 18, puesto que no tenemos indicios suficientes para rechazar esta posibilidad.

Actividad 3: Estudio de una empresa informática sobre la eficiencia de un servidor Web.

Contraste de hipótesis paramétricos. Contraste de la media de una población normal con varianza desconocida. Estimación por intervalos: Intervalo de confianza. Software R.

En una empresa informática se desea medir la eficiencia de un servidor Web. Para ello, miden el tiempo de espera del cliente entre la petición que éste hace y la respuesta que le da el servidor.

Los tiempos de espera (en milisegundos) del servidor (TA) para 50 peticiones son:

TA
9,67
9,62
9,50
10,88
8,94
10,59
9,81
9,46
9,26
9,02
8,61
9,42
10,86
10,01
10,55
11,26
10,64
10,23
11,63
8,91
10,27
9,49
8,99
10,09
9,11
9,47
8,08
9,98
10,30
7,05
11,79
9,59
10,88
9,83
10,92
10,98
9,54
10,17
10,32
10,01
9,96

9,28
10,30
11,08
10,05
9,74
11,14
9,44
9,17
10,86

Suponiendo normalidad, contestad a las preguntas siguientes:

- ¿Podemos considerar que el tiempo medio de espera del servidor A es de 9 milisegundos? Razonad la respuesta. Hallad el p-valor del contraste. Tomad $\alpha=0,1$
- Hallar un intervalo de confianza para el tiempo medio de espera del servidor A al 95% de confianza.

Solución

Apartado a)

Para importar los datos a R haremos lo siguiente (si los datos están en un fichero SERVIDORA.txt, dónde indicamos que el separador decimal es la ,):

```
> Data <- as.data.frame(read.table("./servidorA.txt", header=TRUE,
dec = ",", ))
```

Si buscamos con R la media y la desviación típica de la muestra, encontraremos el estadístico de contraste con la instrucción siguiente:

```
> t <- (mean(Data$TA)-9) / (sd(Data$TA)/sqrt(50))
> t
[1] 7.344513
```

Ahora ya podemos encontrar el p-valor:

```
> 2*pt(t, df=49, lower.tail=FALSE)
[1] 1.944584e-09
```

Como el nivel de significación (0,1) es mayor que el p-valor ($1,945 \cdot 10^{-9}$), rechazaremos la hipótesis nula y aceptaremos la alternativa, es decir, no podemos considerar que el tiempo medio de espera del servidor A es de 9 ms.

También se podría haber hecho con otro método, usando directamente la instrucción correspondiente de R:

```
> t.test(Data$TA, mu=9, alternative="two.sided", conf.level=0.9)
One Sample t-test

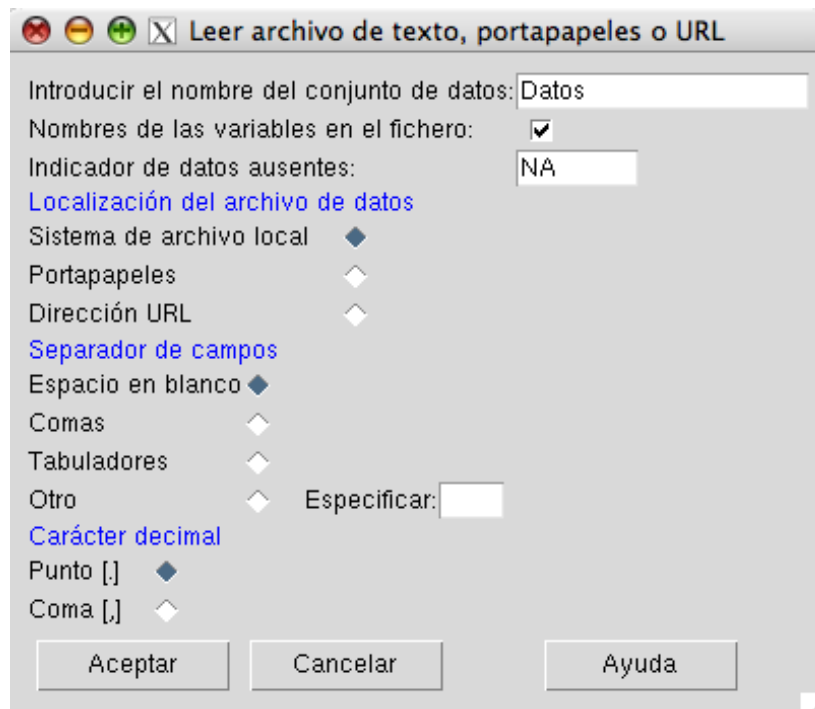
data: Data$TA
t = 7.3445, df = 49, p-value = 1.945e-09
alternative hypothesis: true mean is not equal to 9
90 percent confidence interval:
 9.721565 10.148435
sample estimates:
mean of x
 9.935
```

Como podemos ver, el resultado es el mismo (p-value = $1.945e-09$), por lo tanto llegaríamos a la misma conclusión.

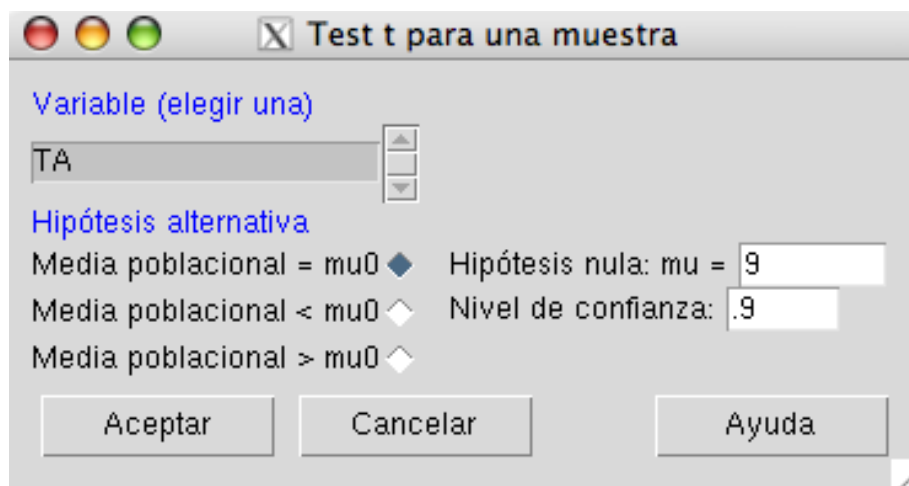
Finalmente se puede hacer usando el denominado Rcommander, que, una vez instalado se llama con la instrucción:

```
> library(Rcmdr)
```


Después hace falta introducir los datos desde *Datos->Importar datos* con lo que nos queda:



y desde *Estadísticos -> medias -> test t* para una muestra hacemos:



con lo que obtenemos:

```

R Commander
Fichero Editar Datos Estadísticos Gráficas Modelos Distribuciones Herramientas Ayuda
Conjunto de datos: Datos Editar conjunto de datos Visualizar conjunto de datos Modelo: <No hay modelo activo>
Ventana de instrucciones
Datos <-
  read.table("/Volumes/NO NAME/2008-09/nou material uoc/r/servidorA.txt",
    header=TRUE, sep=" ", na.strings="NA", dec=".", strip.white=TRUE)
t.test(Datos$TA, alternative='two.sided', mu=9, conf.level=.9)
t.test(Datos$TA, alternative='two.sided', mu=9, conf.level=.9)

Ventana de resultados Ejecutar
> Datos <-
+ read.table("/Volumes/NO NAME/2008-09/nou material uoc/r/servidorA.txt",
+ header=TRUE, sep=" ", na.strings="NA", dec=".", strip.white=TRUE)
> t.test(Datos$TA, alternative='two.sided', mu=9, conf.level=.9)

One Sample t-test

data: Datos$TA
t = 7.3445, df = 49, p-value = 1.945e-09
alternative hypothesis: true mean is not equal to 9
90 percent confidence interval:
 9.721565 10.148435
sample estimates:
mean of x
 9.935

Mensajes

```

Apartado b)

Tenemos que encontrar el intervalo de confianza al 95% para la media de tiempo de espera del servidor A. Tenemos que utilizar la siguiente fórmula:

$$\left(\bar{X}_A - t_{\alpha/2, n-1} \frac{S_A}{\sqrt{n_A}}; \bar{X}_A + t_{\alpha/2, n-1} \frac{S_A}{\sqrt{n_A}} \right)$$

Esto lo podemos calcular, utilizando R, así:

```

> error_std<-sqrt(var(Data$TA))/(sqrt(length(Data$TA)))
> error_std
[1] 0.1273059
> valor_critico<-qt(0.025,df=49,lower.tail=FALSE)
> valor_critico
[1] 2.009575
> ci<-c(mean(Data$TA)-valor_critico*error_std,mean(Data$TA)
+valor_critico*error_std)
> ci
[1] 9.67917 10.19083

```

Por lo tanto, el intervalo de confianza para el tiempo medio de espera de A es (9.67917, 10.19083). Podemos ver que el 9 no está dentro del intervalo, y por lo tanto, se confirma lo que se ha demostrado el apartado a) (aunque con un nivel de significación diferente). También se puede encontrar este intervalo con la instrucción:

```

> t.test(Data$TA,mu=9,alternative="two.sided",conf.level=0.95)

One Sample t-test

```

```

data: Data$TA
t = 7.3445, df = 49, p-value = 1.945e-09
alternative hypothesis: true mean is not equal to 9
95 percent confidence interval:
 9.67917 10.19083
sample estimates:
mean of x
 9.935

```

Actividad 4: Estudio sobre la calidad de un sistema operativo.

Contraste de hipótesis paramétricos. Inferencia sobre datos apareados. Contraste de la media sobre poblaciones normales. Estimación por intervalos: Intervalo de confianza. Software Minitab.

Hemos pedido a 10 informáticos que evalúen, en base a unos criterios preestablecidos, la calidad de un determinado sistema operativo. Las puntuaciones varían entre un mínimo de 0 y un máximo de 15. Pasados tres meses, las mismas 10 personas repiten el proceso de evaluación. Los resultados obtenidos, que introduciremos en las columnas V1 y V2, son los siguientes:

Persona	EV_1	EV_2
1	13,2	14,0
2	8,2	8,8
3	10,9	11,2
4	14,3	14,2
5	10,7	11,8
6	6,6	6,4
7	9,5	9,8
8	10,8	11,3
9	8,8	9,3
10	13,3	13,6

- ❑ Nuestro objetivo es doble: por un lado, pretendemos calcular un intervalo de confianza, a nivel del 95%, para $\mu_A - \mu_B$; por otro, contrastar las hipótesis: $H_0: \mu_A - \mu_B = 0$ vs. $\mu_A - \mu_B \neq 0$.

Debemos hacer un contraste de diferencia de medias apareadas. Introducimos datos.

Vamos a R Comander: Estadísticos > Medias y le damos a Test t para datos relacionados y elegimos las dos variables:

```
with(Dataset, (t.test(V1, V2, alternative='two.sided', conf.level=.95, paired=TRUE)))
```

```

Paired t-test
data: V1 and V2
t = -3.3489, df = 9, p-value = 0.008539
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.6869539 -0.1330461
sample estimates:
mean of the differences
 -0.41

```

Los resultados obtenidos nos dicen que, en base a las observaciones registradas, hay una probabilidad de 0,95 de que $\mu_A - \mu_B$ sea un valor del intervalo (-0,687, -0,133). Además, con un p-valor de 0,009 también podemos afirmar que hay indicios suficientes como para descartar la hipótesis nula. Por tanto, parece sensato pensar que las dos medias poblacionales son distintas.

Notar que esta conclusión es coherente con que el valor 0 no esté incluido en el intervalo de confianza hallado para la diferencia de ambas medias.

Actividad 5: Estudio sobre la media del tiempo de transferencia de ficheros.

Contraste de hipótesis. Contraste de la media con varianza desconocida. Contraste de la media con varianza conocida. Software R.

El responsable de comunicaciones de una empresa informática afirma que la media del tiempo de transferencia por FTP de un fichero de 2Mb es superior a 30 segundos. Para comprobar esta afirmación este tomó una muestra del tiempo de transferencia de 12 ficheros de 2Mb, y se obtuvo que la media y la desviación estándar muestral valen $\bar{x}=30,2$, $s=1,833$ (en segundos).

- Suponiendo que el tiempo de transferencia se distribuye normalmente, a partir de los datos muestrales obtenidos, ¿tenemos las suficientes evidencias para creer la afirmación del responsable? (Tomad $\alpha=0,05$). Encontrad el p-valor del contraste.
- Si además de disponer de estas observaciones nos hubieran dado como información adicional (obtenida de experiencias previas) que la varianza del tiempo de transferencia es de según $\sigma^2=9,2$, ¿hubiésemos llegado a la misma conclusión que en el apartado anterior? Encontrad el p-valor del contraste.

En cada apartado hace falta seguir el siguiente esquema:

- Especificar el tipo de contraste que hace.
- Indicar la hipótesis nula y alternativa.
- Indicar la fórmula del estadístico de contraste así como su distribución de probabilidad.
- Usar el software para encontrar los p-valores correspondientes y los valores críticos.

Solución

Apartado a)

- Debemos hacer un contraste de una media con varianza desconocida.

$$H_0: m=30,$$

- Las hipótesis nula y alternativa son: $H_1: m>30$, donde m representa la media del tiempo de transferencia por FTP de un fichero de 2Mb.

$$t = \frac{\bar{x} - 30}{s / \sqrt{12}},$$

- El estadístico de contraste es: $t = \frac{\bar{x} - 30}{s / \sqrt{12}}$, donde \bar{x} es la media muestral y s es la desviación estándar muestral. La distribución de t es la de t de Student con 11 grados de libertad. La media y la desviación estándar muestral valen respectivamente: $\bar{x}=30,2$, $s=1,833$. El valor del estadístico de contraste vale

```
> t<-(30.2-30)/(1.833/sqrt(12))
> t
[1] 0.3779707
```

- El p-valor

```
> pt(0.3779707, df=11, lower.tail=FALSE)
[1] 0.356322
```

Como que el p-valor vale $0,36 > 0,05$ aceptamos la hipótesis nula, es a decir, la velocidad es 30. El valor crítico valdrá:

```
> qt(0.05, df=11, lower.tail=FALSE)
[1] 1.795885
```

Como $t=0.3779707 < t_{0.05,11}=1.795885$ aceptamos la hipótesis nula y concluimos que la afirmación del responsable no se puede considerar cierta.

Apartado b)

1) Debemos hacer un contraste de una media con varianza conocida.

$$H_0: m=30,$$

2) Las hipótesis nula y alternativa son: $H_1: m > 30$, donde m representa la media del tiempo de transferencia por FTP de un fichero de 2Mb.

$$z = \frac{\bar{x} - 30}{s/\sqrt{12}},$$

3) El estadístico de contraste es: donde \bar{x} es la media muestral y s es la desviación estándar poblacional. La distribución de z es la de una normal $N(0,1)$. La media y la desviación estándar poblacional valen respectivamente: $\bar{x}=30.2, s=\sqrt{9.2} \approx 3.03$. El valor del estadístico de contraste vale:

```
> z <- (30.2 - 30) / (3.03 / sqrt(12))
> z
[1] 0.2286536
```

4) El p-valor vale

```
> pnorm(0.2286536, lower.tail=FALSE)
[1] 0.4095691
```

y el valor crítico

```
> qnorm(0.05, lower.tail=FALSE)
[1] 1.644854
```

En este caso aceptamos la hipótesis nula y concluimos que la afirmación del responsable no es cierta.

Actividad 6: Contraste de la media de la diferencia en el tiempo de arranque de un ordenador con antivirus y sin antivirus.

Contraste de hipótesis. Contraste de la media con varianza desconocida. Muestras apareadas. Programario R.

Se cuestiona si el tiempo (en segundos) de arranque de un ordenador con el antivirus TRONX es más lento que si el ordenador no tiene el antivirus. Con el objetivo de hacer este estudio se ha medido el tiempo de arranque de 10 ordenadores sin tener ningún antivirus instalado, obteniendo los siguientes resultados:

Sin el antivirus	47,8	70,54	59,3	46,0	54,4	70,1	56,9	72,6	62,3	63,3
------------------	------	-------	------	------	------	------	------	------	------	------

Después, se ha instalado el antivirus en todos estos ordenadores y se ha vuelto a medir el tiempo de arranque, obteniendo:

Con el antivirus	46,7	80,8	63,9	43,9	56,7	80,1	60,4	71,2	60,3	69,8
------------------	------	------	------	------	------	------	------	------	------	------

Suponiendo que el tiempo de arranque de un ordenador se distribuye normalmente, decide con $\alpha=0,05$ si la existencia en el ordenador del antivirus hace que los tiempos de arranque sea mayor. Encontrad el p-valor del contraste.

Para resolver el problema, indicad
a) Especificar el tipo de contraste que hace.

- b) Indicar la hipótesis nula y alternativa.
- c) Realizar el test correspondiente.

Solución

a) Debemos hacer un contraste de diferencia de medias con datos apareados.

b) Sea m la diferencia de medias entre el tiempo de arranque de los ordenadores entre los que no tienen antivirus y los que tienen antivirus. O sea,

$$m = m_{\text{no antivirus}} - m_{\text{amb antivirus}}$$

El contraste a hacer es:

$$H_0: m = 0,$$

$$H_1: m < 0.$$

c)

Introducimos los datos

```
> sin_anti<-c(47.8, 70.54, 59.3, 46.0, 54.4, 70.1, 56.9, 72.6,
62.3, 63.3)
> con_anti<-c(46.7, 80.8, 63.9, 43.9, 56.7, 80.1, 60.4, 71.2, 60.3,
69.8)
```

Efectuamos el test de datos emparejados con alternativa "<" o "less":

```
>t.test(sin_anti,con_anti,alternative="less",mu=0,paired=TRUE,conf.
level=0.95)

Paired t-test

data:  sin_anti and con_anti
t = -2.0278, df = 9, p-value = 0.0366
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
 -Inf -0.2934189
sample estimates:
mean of the differences
 -3.056
```

El valor del estadístico de contraste vale $t = -2.0278$ y el p-valor $p=0.0366$. Por lo tanto rechazamos la hipótesis nula y llegamos a la conclusión que el tiempo de arranque es mayor con el antivirus.

Actividad 7: Contraste de la proporción de usuarios de un software.

Contraste de hipótesis. Contraste de la proporción en muestras grandes. Software R.

En una universidad el centro de cálculo afirma que la proporción de terminales que usan el programa estadístico R a primera hora de la mañana es inferior al 50%. Cogida una muestra de 700 terminales a primera hora de la mañana se observó que 410 estaban usando el programa R. Con $\alpha=0,01$, ¿podemos estar de acuerdo con la afirmación del centro de cálculo? Encontrad el p-valor del contraste.

Solución

Debemos hacer un contraste de proporciones. Sea p la proporción de terminales que usan R a

$$H_0: p = 0.5,$$

$$H_1: p < 0.5.$$

primera hora de la mañana. El contraste de hipótesis es:

$$z = \frac{\hat{p} - 0.5}{\sqrt{\frac{0.5 \cdot 0.5}{700}}}$$

El estadístico de contraste vale: dónde \hat{p} es la proporción de terminales que usan R a primera hora de la mañana y vale $410/700=0.586$. El estadístico de contraste sigue aproximadamente la distribución normal $N(0,1)$ si la medida de la muestra es suficientemente grande como es nuestro caso. El valor del estadístico de contraste es

```
> pbar<-410/700
> z<- (pbar-0.5)/sqrt(0.5*0.5/700)
```

```
> z
[1] 4.535574
```

Calculamos el p-valor

```
> #como la alternativa es < haremos
> pnorm(4.535574)
[1] 0.9999971
```

y el valor crítico

```
> qnorm(0.99)
[1] 2.326348
```

Por lo tanto hemos de aceptar la hipótesis nula y concluir que la proporción es el 50% y no inferior al 50%.

Direcciones de interés

<http://oak.cats.ohiou.edu/~wallacd1/shyp.html>

Definición y ejemplos de contraste de hipótesis de una población conocida la media y la desviación estándar de la población.

<http://www.unalmed.edu.co/~estadist/confinterval/intervalconf.htm>

Características y applet del concepto de Intervalo de confianza.

http://e-stadistica.bio.ucm.es/mod_intervalos/intervalos_applet.html

Características y applet del concepto de Intervalo de confianza.

http://e-stadistica.bio.ucm.es/mod_contraste/contraste_applet.html

Applet sobre contraste de hipótesis para muestras independientes.

<http://kitchen.stat.vt.edu/~sundar/java/applets/>

Aplicaciones estadísticas con JAVA.

<http://www.udc.es/dep/mate/recursos.html>

Selección de recursos en Internet para la enseñanza-aprendizaje de la Estadística.

http://www.uoc.edu/in3/e-math/docs/CH_1Pob.pdf

Math-block del proyecto e-math sobre contraste de 1 población con teoría y ejemplos con y sin Minitab.