EJERCICIOS

1. En un colectivo de personas seleccionamos aleatoriamente 30 hombres a los cuales pedimos su altura. La media muestral obtenida es de 1.75 m y la desviación estándar muestral es de 0.09 m. Suponiendo que la altura se distribuye normalmente, se pide:

a) Construye un intervalo de confianza del 90% para la media poblacional de la alzada.

b) Contrasta al 5% de significación la hipótesis que la media poblacional sea igual o inferior a 1.80 m.

Resolución

a) El intervalo será: $1.75 \pm 1.64 \cdot \frac{0.09}{\sqrt{30}}$

1.75 ± 0.0269

es decir, el intervalo de confianza es [1.7231;   1.7769]

b)

$H_0 : \mu = 1.80$
$H_1 : \mu \leq 1.80$

El estadístico de contraste es:

$$EC = \frac{1.75 - 1.80}{0.09 / \sqrt{30}} = -3.04$$

Si suponemos que el tamaño de la muestra es grnde y, puesto que se trata de un contraste unilateral por la izquierda, tenemos que:

EC = (-3.04) < VC (-1.96)

Y, por tanto, se rechaza la hipótesis nula.

2. El servicio de estudios económicos de la UOC quiere realizar un estudio de coyuntura del sector de bibliotecas en Cataluña. Por este motivo, ha seleccionado una muestra representativa del sector para el 1996, y ha destacado que 91 bibliotecas de un total de 144 habían incrementado la suya cifra de usuarios respecto del año anterior.
a) ¿Cuál es la probabilidad de que la proporción de bibliotecas que han incrementado su cifra de usuarios sea superior al 50%?

b) Construye un intervalo de confianza del 95% para la verdadera proporción de bibliotecas que han incrementado la cifra de usuarios.
c) ¿Con qué muestra debería trabajar el IDESCAT si quisiéramos disminuir el margen de error de la apartado b) a la mitad? (trabajando con el mismo nivel de confianza)

Resolución

p = 91 / 144 = 0.63

a) $p(X>0.5) = p\left(Z> \dfrac{0.5-0.63}{\sqrt{\dfrac{0.63*0.37}{144}}}\right) = p(Z > -3.25) = 1 - p(Z < -3.25 = 0.99942$

b) Intervalo = $0.63 \pm 1.96\sqrt{\dfrac{0.63*0.37}{144}} = 0.63 \pm 0.08 = [0.55;0.71]$

c) Si el margen de error es 0.08 y lo queremos reducir a la mitad:

$$0.04 = 1.96\sqrt{\frac{0.63*0.37}{n}} \quad \rightarrow \quad n = 560$$

3. Se cruzaron dos razas de perros y, de cien perros reciente nacidos, 15 son negros y los otros no. De acuerdo con el modelo genético, estos deberán ser en una proporción 3/13 (negros/no negros). Se pide contrastar la hipótesis de consistencia de los datos. Estudiar el modelo por un nivel de significación del 0.05.

Resolución

La hipótesis nula será que nuestra observación deriva del modelo genético, es decir, que no hay diferencia entre la proporción observada y la teórica

La hipótesis alternativa será rechazar la hipótesis nula, es decir, considerar que nuestro modelo no deriva del genético (la diferencia entre proporciones es significativa)

La proporción teórica es de 3 negros/(13+3) total = 3/16 = 0'1875 perros negros
La proporción observada es de 15/100 = 0'15 perros negros.

La diferencia entre la proporción observada y la proporción teórica (que bajo la H_0 debería ser igual a cero) será 0'1875 - 0'15 = 0'0375

Bajo la hipótesis nula la desviación estándar de la proporción será:
$\sqrt{0'1875*(1-0'1875)/100}$ = 0'03903

En definitiva, el estadístico de contraste será: 0'0375/0'03903 = 0'96

Al ser α = 0'05 y tratarse de una prueba bilateral, $\alpha/2$ = 0'025, por lo que $Z_{\alpha/2} = \pm$ 1'96.

En otras palabras, la zona de aceptación de la hipótesis nula es [-1'96, 1'96] y la de rechazo el complementario.

Nuestro estadístico de contraste cae dentro del intervalo de aceptación, por lo que se debe aceptar la hipótesis nula de igualdad entre la proporción observada y la teórica para un nivel de significación de 0'05.

PRÁCTICA 1

En una determinada empresa de nuevas tecnologias, se quiere realizar un estudio para averiguar el número medio de días de baja laboral por empleado, ya que se ha detectado que debido a la presión y responsabilidad del trabajo, el número de días de baja laboral por empleado ha aumentado.

En los últimos años se había estimado que este número medio era de 18 días al año, con una desviación estándar de 2.6.

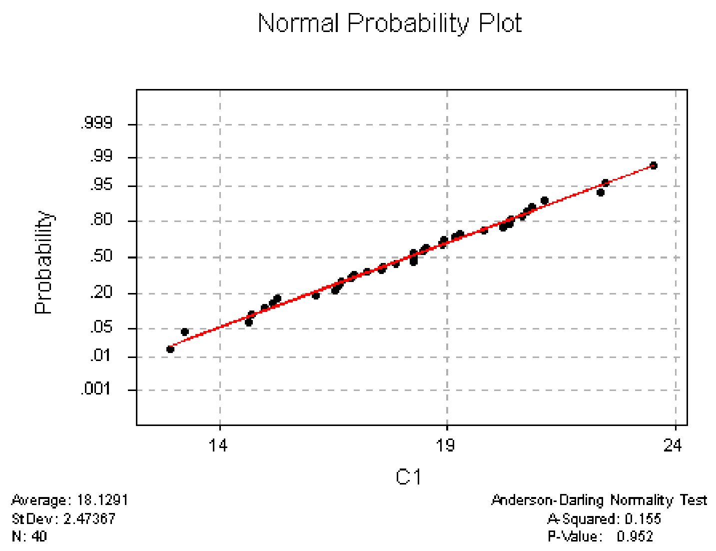
Para realizar este trabajo, se ha hecho un estudio basado en 40 empleados escogidos aleatoriamente, obteniendo los siguientes datos:

18.2630	20.3956	18.4742	18.5457	19.2611
13.2104	18.8842	18.9199	14.7056	16.6708
17.8783	20.6542	18.2433	16.8871	21.1429
15.2542	16.5648	19.1553	16.1106	18.8906
20.7617	22.4938	18.2460	17.2399	17.5592
15.1401	23.5305	20.2367	16.5339	20.8730
19.8144	14.9760	20.3570	17.5765	16.6488
16.9471	22.3734	18.2373	14.6103	12.8947

a) Comprobar que la colección de datos anterior sigue una distribución aproximadamente normal.

Solución

Para comprobar la normalidad de los datos anteriores, seleccionamos *Stat > Basic Statistics > Normality Test.* Así, obtenemos el siguiente gráfico:



Por lo tanto, podemos concluir sin ningún tipo de dudas que estos datos siguen una distribución normal.

Podemos asegurar, además, que como X sigue una distribución normal, $\bar{X}$ también.

b) Como el tiempo medio de dias de baja laboral parece excesivo, se quiere contrastar, con un nivel de significación del 0.05, la hipótesis "oficial" de que el tiempo medio es de 18 dias frente a la hipótesis de que esta media es menor. ¿Qué resultado se obtiene?

Solución

El contraste de hipótesis que estableceremos será H0: $\mu=18$ vs. H1: $\mu<18$

Seleccionamos *Stat* > *Basic Statistics* > *1-Sample* Z:

Z-Test

Test of mu = 18.000 vs mu < 18.000

The assumed sigma = 2.60

Variable	N	Mean	StDev	SE Mean	Z	P
C1	40	18.129	2.474	0.411	0.31	0.62

Dado que el p-valor obtenido 0.62 > 0.05, no descartaremos la hipótesis nula, eso significa que parece razonable considerar que el número medio de días de baja laboral es de 18.

c) Calcula un intervalo de confianza a nivel del 90% para la media poblacional y comenta si el resultado obtenido es coherente con el resultado esperado.

Seleccionamos *Stat* > *Basic Statistics* > *1-Sample* Z, considerando 0.9 como el nivel de confianza:

Z Confidence Intervalos

The assumed sigma = 2.60

Variable	N	Mean	StDev	SE Mean	90.0% CI
C1	40	18.129	2.474	0.411	(17.453, 18.805)

Como observamos, el intervalo de confianza obtenido (17.453, 18.805), es coherente con los resultados esperados ya que contiene al valor medio 18.

d) Finalmente, realiza el mismo contraste que el apartado **b),** pero suponiendo esta vez que no conoces la desviación estándar.

Solución

Análogamente, seleccionamos *Stat > Basic Statistics > 1-Sample t*, obteniendo los siguientes resultados:

```
T-test of the Mean


Test of mu = 18.000 vs mu < 18.000

Variable      N      Mean      StDev     SE Mean     T          P
C1            40     18.129    2.474      0.391      0.33       0.63
```

Por lo tanto, observamos que el p-valor 0.63 > 0.05, eso nos indica que no rechazamos la hipótesis nula, es decir, asumiremos como posible la opción de que el número medio de días de baja laboral sea 18, ya que no tenemos indicios suficientes para rechazar esta posibilidad.

PRACTICA 2

Se han seleccionado 60 estudiantes de la UOC para realizar un estudio acerca del tipo de conexión a internet que poseen los estudiantes. Por ello, estamos interesados en estimar qué proporción del alumnado accede al campus virtual mediante una línea ADSL.

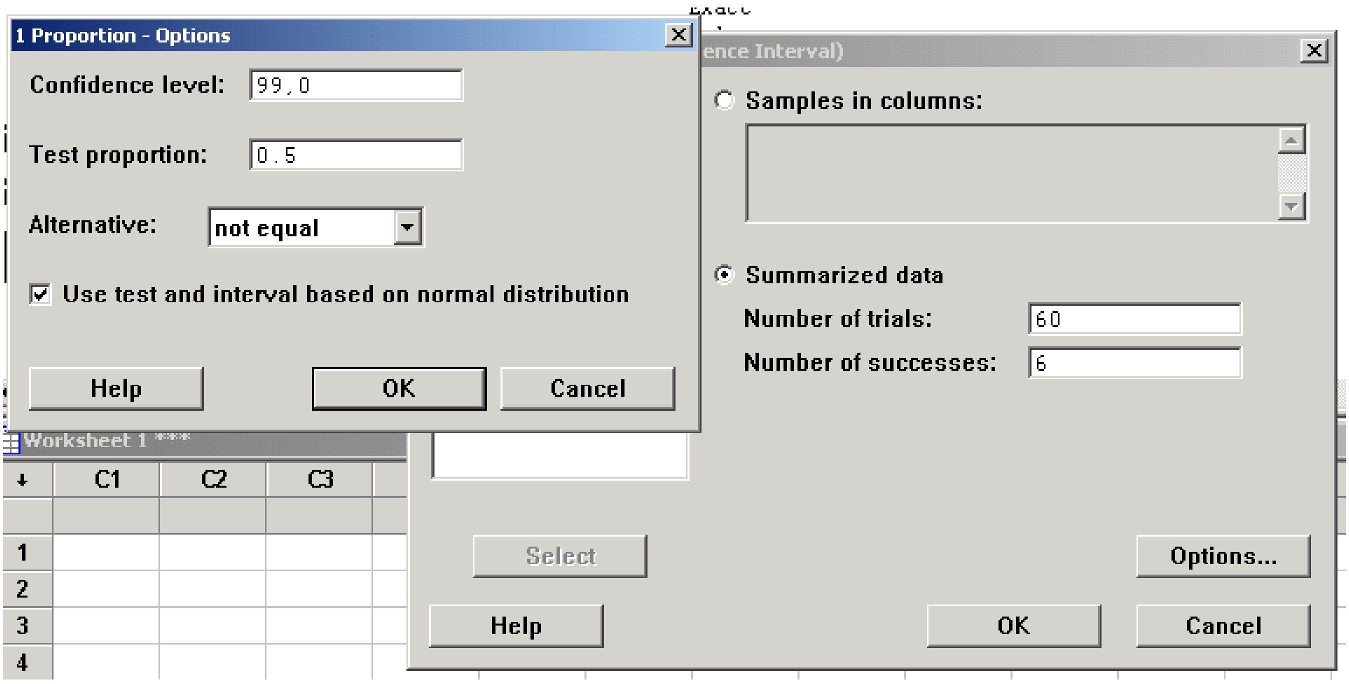
De los estudiantes seleccionados solamente el 10% accede utilizando este medio.

a) Calculad el intervalo de confianza para la proporción de estudiantes que accede al campus utilizando ADSL , a un nivel de confianza del 99%.

Solución

Para calcular el intervalo de confianza para la proporción de estudiantes, seleccionaremos *Stat > Basic Statistics > 1 Proportion*.
Consideraremos el número de pruebas realizadas n=60 estudiantes entrevistados y, de éstos, el número de "éxitos" (estudiantes que poseen ADSL) corresponde al 10%, es decir, 6 estudiantes.

Seleccionamos también "*Options*" y consideramos un constraste de hipótesis, a un nivel de confianza del 99% y podemos suponer un contraste de hipótesis p=0.5 vs p<>0.5 (el estudiante tiene ADSL o no la tiene)



Los resultados obtenidos son:

Test and CI for One Proportion

```
Test of p = 0,5 vs p not = 0,5

Sample      X     N  Sample p         99,0% CI        Z-Value  P-Value
1           6    60  0,100000  (0,000239; 0,199761)     -6,20    0,000
```

Es decir, el intervalo de confianza obtenido con un nivel de confianza del 99% es aprox. (0 , 0.2)

b) Por otra parte, la biblioteca de la UOC ha introducido unos mejoras en el programa informático que se encarga de realizar los préstamos y que permite que estos trámites sean más rápidos. Por ello, quiere evaluar el grado satisfacción del cliente tras algunos meses de funcionamiento.
La biblioteca de la UOC sabe que de los usuarios que acceden a la biblioteca, el 60% NO se muestran satisfechos al realizar un préstamo.
Se pide contrastar, al nivel de significación del 2% si este porcentaje se ha reducido tras haber realizado las mejoras.
Para ello se ha tomado una muestra de 500 usuarios para conocer su opinión y se ha observado que el 55% sigue sin estar satisfecho con los servicios de gestión de préstamos.

Solución:

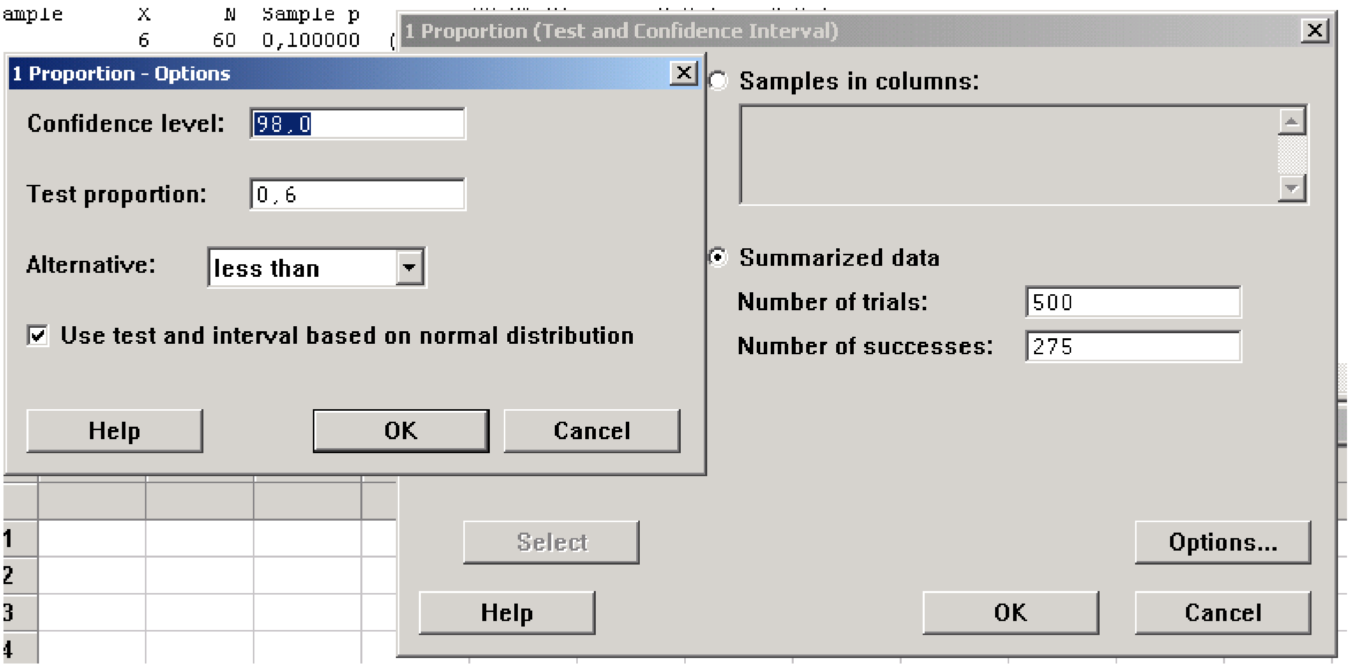
Planteamos el siguiente contraste:

$$H_0 : p = 0,6$$
$$H_{A:} : p < 0,6$$

Sea X="Número de estudiantes insatisfechos"

Consideraremos n=500 estudiantes entrevistados, teniendo en cuenta que el 55% de éstos siguen sin estar satisfechos, es decir, 275 estudiantes.

Seleccionamos *Stat > Basic Statistics > 1 Proportion* y rellenamos como se ve en la imagen:



Los resultados obtenidos son:

Test and CI for One Proportion

```
Test of p = 0,6 vs p < 0,6

Sample      X     N  Sample p  98,0% Upper Bound  Z-Value  P-Value
1         275   500  0,550000           0,595693    -2,28    0,011
```

Por tanto, como el p-valor es menor que el nivel de significación, 0.011 < 0.02, podemos deducir que para este nivel de significación rechazaremos la hipótesis nula, es decir, podemos concluir que las mejoras realizadas en el programa informático han contribuido a que el nivel de satisfacción entre los estudiantes que acceden a realizar el préstamo ha aumentado (o que el nivel de insatisfacción entre los estudiantes ha disminuido).