Comenzado el	martes, 14 de mayo de 2024, 12:05
Estado	Finalizado
Finalizado en	martes, 14 de mayo de 2024, 13:50
Tiempo empleado	1 hora 44 minutos
Calificación	**4,71** de 10,00 (**47,14**%)

Pregunta **1**

Correcta

Se puntúa 1,00 sobre 1,00

Se han recogido datos relativos al número de usuarios de una estación de metro de Madrid durante 33 días y se ha calculado el intervalo de confianza 99% del número medio de usuarios, que resulta: $(750.17, 761.83)$. Se asume normalidad en los datos. Calcula:

a. Estimación puntual de la media: 755.8299 ✔

b. Estimación puntual de la varianza: 169.0545 ✔

c. Si queremos conseguir un intervalo más preciso, reduciendo la amplitud de dicho intervalo a la mitad, cuantos datos necesitaríamos tomar en la nueva muestra? 133 ✔

```
ci = 750.17
cs = 761.83
n = 33
conf = 0.99
alfa = 1 - conf
talfa = qt(1-alfa/2, n-1)
media = (ci+cs)/2
sd = sqrt(n)*(cs-media)/talfa
var = round(sd^2,3)
n2=4*n
```

Pregunta **2**

Parcialmente correcta

Se puntúa 0,20 sobre 1,40

El número de mutaciones genéticas que experimenta un organismo en un laboratorio sigue una distribución de Poisson con una media de 1.2 mutaciones anuales.

a. Hallar la probabilidad de que este organismo sufra alguna mutación en un año. 0.3614331 ✖

b. Calcular la probabilidad de que en 10 meses experimente menos de 2 mutaciones. 0.9933509 ✖

c. Determinar la probabilidad de que sufra al menos 4 mutaciones a lo largo de un año y medio. 0.03640666 ✖

d. Si en el último año este organismo ha experimentado al menos 2 mutaciones, ¿cuál es la probabilidad de que hayan sido exactamente 2? 0.64279 ✔

e. El organismo acaba de sufrir una mutación. ¿Cuál es la probabilidad de que pasen entre 13 y 26 meses hasta que experimente la siguiente? ✖

f. Ha transcurrido ya más de un año desde que el organismo experimentó la última mutación. Calcular la probabilidad de que transcurra al menos otro año hasta que sufra una nueva mutación. ✖

lambda = 1.2

m = 10

n1 = 2

n2 = 4

n3 = 2

m2 = 26

m1 = 13

a. La probabilidad pedida es: p1 = 1-dpois(0,lambda) = 0.6988058
b. La probabilidad pedida es: p2 = ppois(n1-1, lambda*m/12) = 0.7357589
c. La probabilidad pedida es: p3 = 1-ppois(n2-1, lambda*1.5) = 0.1087084
d. La probabilidad pedida es: p4 = dpois(n3, lambda)/(1-ppois(n3-1,lambda)) = 0.64279
e. La probabilidad pedida es: p5 = pexp(m2, lambda/12) - pexp(m1, lambda/12) = 0.1982582
f. La probabilidad pedida es: p6 = 1-pexp(1, lambda) = 0.3011942

Pregunta **3**

Parcialmente correcta

Se puntúa 2,60 sobre 3,60

Nota: Preguntas: **Tipo de Contraste:** Cada respuesta correcta suma 0.1. Cada respuesta erronea resta 0.05. Las respuestas en blanco puntúan 0. Preguntas: **Solución:** Cada respuesta correcta suma 0.3. Cada respuesta erronea resta 0.15. Las respuestas en blanco puntúan 0.

(1) Una empresa de software está analizando la calidad del código que desarrolla. Se seleccionan aleatoriamente 12 códigos y se mide el tiempo que tarda cada uno en ejecutarse, obteniendo los datos:

```
10.55 9.9 10.5 9.32 10.78 9.03 10.59 9.23 9.27 10.67 10.66 9.33
```

Utiliza un nivel de significación del 0.02 para contrastar si el tiempo medio de ejecución es superior a 9 segundos. Se asume que el tiempo de ejecución es normalmente distribuido.

Tipo de contraste: Contraste para la media de una muestra ✔ .

Solución: El contraste es significativo ✔ .

(2) Un equipo de desarrolladores de inteligencia artificial (IA) ha creado dos algoritmos de clasificación de imágenes para detectar la presencia de gatos en fotografías. Después de probar ambos algoritmos en un conjunto de datos de prueba, se obtuvieron los siguientes resultados:

Algoritmo A: Detectó gatos en 85 de 101 imágenes.

Algoritmo B: Detectó gatos en 84 de 102 imágenes.

¿Hay evidencias para concluir que el algoritmo A es mejor que el algoritmo B en cuanto a precisión de detección de gatos en imágenes? Usar un nivel de significación de 0.01.

Tipo de contraste: Contraste para la diferencia de proporciones de dos muestras ✔ .

Solución: El contraste es no-significativo ✔ .

(3) Se está investigando el tiempo medio que tardan los usuarios en completar una tarea en dos versiones diferentes de un software. Se pide a 10 usuarios que completen la misma tarea con cada una de las dos versiones del software y se obtienen los datos de la tabla siguiente:

	versionA	versionB
1	10	10
2	10	10
3	10	10
4	11	11
5	11	12
6	11	13
7	13	13
8	13	13
9	13	13
10	13	13

Usando un nivel de significación del 1%, ¿podemos decir que hay diferencias significativas entre las dos versiones del software, en cuanto al tiempo medio para completar la tarea?

Tipo de contraste: Contraste para la diferencia de medias de dos muestras independientes ✖ .

Solución: El contraste es no-significativo ✔ .

(4) Un equipo de seguridad informática está valorando la efectividad de un software antivirus. Para ello, simula 149 ataques informáticos, de los cuales el software ha sido capaz de detectar 107. ¿Podemos concluir que la proporción de ataques detectados por el software es superior a 70%? Utiliza un nivel de significación del 0.02.

Tipo de contraste: Contraste para la proporción de una muestra ✔ .

Solución: El contraste es no-significativo ✔ .

(5) Un equipo de desarrollo de videojuegos está evaluando la efectividad de un nuevo sistema de detección de trampas en línea. Tras una serie de pruebas, se observa que el sistema ha identificado al 79% de los intentos de trampa. Utiliza un contraste de hipótesis con un nivel de significación de 0.02 para determinar si la probabilidad de que el sistema detecte una trampa es superior al 0.8%.

Tipo de contraste: Contraste para la proporción de una muestra ✔ .

Solución: El contraste es significativo ✖ .

(6) Una empresa informática quiere comprobar si hay diferencias significativas en la estabilidad de dos sistemas operativos en términos de la variabilidad en el tiempo entre fallos. Para ello, toma muestras de los tiempos entre fallos de ambos sistemas operativos, obteniendo los siguientes datos:

Tiempo entre fallos del sistema operativo A:

```
4.71 33.54 6.78 7.5 2.15 0.88 44.26 3.57 4.49 3.62 1.82 8.76 2.53
```

Tiempo entre fallos del sistema operativo B:

```
28.63 12.12 10.76 7.48 9.3 1.12 0.44 2.4 5.33 14.13 18.84 5.04 13.26
```

Utiliza un contraste de hipótesis con un nivel de significación de 0.03 para determinar si hay una diferencia significativa en la variabilidad entre los dos sistemas.

Tipo de contraste: Contraste para el cociente de varianzas de dos muestras ✔ .

Solución: El contraste es no-significativo ✔ .

(7) Se ha comprobado que el número de intervenciones en el equipo de Teams de una asignatura de la URJC sigue una distribución de Poisson. Se eligieron al azar las intervenciones de 24 alumnos en el foro, anotándose el número de veces que habían participado cada uno de ellos, obteniéndose los datos:

```
1 1 4 0 3 4 1 5 5 1 3 3 1 2 5 2 0 2 5 1 6 1 3 4
```

¿Con una significación de 0.1 podemos afirmar que el porcentaje de estudiantes que participan más de 3 veces es inferior al 70%?

Tipo de contraste: Contraste para la proporción de una muestra ✔ .

Solución: El contraste es significativo ✔ .

(8) Se quiere realizar un análisis de los tiempos de acceso a las páginas web de dos bancos A y B. Se realiza una encuesta a 99 y 106 clientes de cada uno de los bancos, respectivamente, obteniendo que los clientes del banco A tardan en promedio 101.19 minutos, con una cuasi-varianza muestral 9.11, mientras que los clientes del banco B tardan en promedio 103.51 minutos, con una cuasi-varianza muestral 9.59. ¿Podemos afirmar que, en media, el tiempo de acceso a la página web del banco B es distinta a la del banco A con una significación del 0.04? Se asume que los tiempos de acceso siguen una distribución Normal.

Tipo de contraste: Contraste para la diferencia de medias de dos muestras independientes ✔ .

Solución: No tenemos suficiente información para resolver el contraste ✖ .

(9) Una empresa está interesada en analizar si una nueva metodología de desarrollo de software produce un código más eficiente en términos de tiempo de ejecución. Toma dos muestras aleatorias de tamaño 13 para cada una de las dos metodologías, cuyos resúmenes númericos son respectivamente:

Tiempo de ejecución de código con la metodología antigua (en segundos):

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
  12.38   12.58   13.00   13.08   13.54   14.05 
```

Tiempo de ejecución de código con la metodología nueva (en segundos):

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max. 
   9.60   10.65   11.14   10.99   11.47   12.22 
```

¿Podemos concluir que la nueva metodología disminuye el tiempo medio de ejecución con respecto a la metodología antigua, usando un nivel de significación del 0.02?

Tipo de contraste: Contraste para la diferencia de medias de dos muestras independientes ✔ .

Solución: El contraste es significativo ✖ .

```
## SOLUTION 1

k1_a = 85

k1_b = 84

n1_a = 101

n1_b = 102

alfa1 = 0.01

test1 <- prop.test(c(k1_a,k1_b), c(n1_a,n1_b))

pvalor1=test1$p.value

signif1 = pvalor1 < alfa1


## SOLUTION 2

n2 = 13

mu2 = 12

y2_a = scan()
```

12.76 14.05 12.38 13 12.46 12.58 12.8 13.2 13.68 13.95 13.54 13.02 12.57

```
y2_b = scan()
```

9.6 11.14 11.47 9.86 10.65 10.29 10.99 11.62 11.33 12.22 10.69 11.2 11.75

```
alfa2 = 0.02

zalfa2 <- qnorm(1-alfa2/2)

pvalor2 = t.test(y2_a,y2_b,alternative="two.sided")$p.value

signif2 = pvalor2 < alfa2

## SOLUTION 3

n3 = 149

k3 = 107

alfa3 = 0.02

p3 = 0.7

test3 = prop.test(k3,n3,p=p3, alternative="less")

pvalor3=test3$p.value

signif3 = pvalor3 < alfa3

## SOLUTION 4

n4 = 10

y4_a = scan()
```

10 10 10 11 11 11 13 13 13 13

```
y4_b = scan()
```

10 10 10 11 12 13 13 13 13 13

```
datos4=data.frame(versionA = y4_a, versionB = y4_b)

alfa4 = 0.01

zalfa4 = qnorm(1-alfa4/2)

dif4=y4_a-y4_b

media4=mean(dif4)

v4 <- var(dif4)

ds4 <- sd(dif4)

ci4 <- media4 - zalfa4*ds4/sqrt(n4)

cs4 <- media4 + zalfa4*ds4/sqrt(n4)

test4 = t.test(y4_a, y4_b, paired = TRUE)

pvalor4=test4$p.value

signif4 = pvalor4 < alfa4

# SOLUTION 5

n5 = 147

k5 = 119

alfa5 = 0.02

p5 = 0.8

test5 = prop.test(k5,n5,p=p5, alternative="less")

pvalor5=test5$p.value

signif5 = pvalor5 < alfa5

## SOLUTION 6

n6_a = 15

n6_b = 12

mu6 = 11

y6_a = scan()
```

4.71 33.54 6.78 7.5 2.15 0.88 44.26 3.57 4.49 3.62 1.82 8.76 2.53

```
y6_b = scan()
```

28.63 12.12 10.76 7.48 9.3 1.12 0.44 2.4 5.33 14.13 18.84 5.04 13.26

```
alfa6 = 0.03

chialfa6_i <- qchisq(1-alfa6/2, n6_a)

chialfa6_s <- qchisq(alfa6/2, n6_a)

test6 = var.test(y6_a,y6_b, alternative="two.sided")

pvalor6 = test6$p.value

signif6 = pvalor6 < alfa6

media6_a <- mean(y6_a)

varianza6_a <- var(y6_a)

desviaciony6_a <- sd(y6_a)

media6_b <- mean(y6_b)

varianza6_b <- var(y6_b)

desviaciony6_b <- sd(y6_b)


## SOLUTION 7

n7 = 24

lambda7 = 3.5

l7 = 3

y7 = scan()
```

1 1 4 0 3 4 1 5 5 1 3 3 1 2 5 2 0 2 5 1 6 1 3 4

```
alfa7 = 0.1

k7=sum(y7>l7)

pr7 = sample(c(0.7, 0.75, 0.8), 1)

test7=prop.test(k7,n7,pr7)

pvalor7=test7$p.value

signif7 = pvalor7 < alfa7

## SOLUTION 8

n8 = 12

mu8 = 9

y8 = scan()
```

10.55 9.9 10.5 9.32 10.78 9.03 10.59 9.23 9.27 10.67 10.66 9.33

```
alfa8 = 0.02

zalfa8 <- qnorm(1-alfa8/2)

test8 = t.test(y8,mu=mu8, alternative="greater")

pvalor8 = test8$p.value

signif8 = pvalor8 < alfa8

media8 <- mean(y8)

v8 <- var(y8)

ds8 <- sd(y8)


ci8 <- media8 - zalfa8*ds8/sqrt(n8)

cs8 <- media8 + zalfa8*ds8/sqrt(n8)


## SOLUTION 9

n9_a = 99

n9_b = 106

m9_a = 101.19

m9_b = 103.51

v9_a  = 9.11

v9_b  = 9.59

alfa9 = 0.04

ci_var<-(v9_a/v9_b)/qf(1-alfa9/2, n9_a-1, n9_b-1)

cs_var<-(v9_a/v9_b)*qf(1-alfa9/2, n9_b-1, n9_a-1)

signif9_var = !((1 < cs_var) && (ci_var < 1))

if (signif9_var){

f9=((v9_a/n9_a+v9_b/n9_b)^2)/(((v9_a/n9_a)^2)/(n9_a-1)+((v9_b/n9_b)^2)/(n9_b-1))

ci9_mu <-(m9_a-m9_b)-qt(1-alfa9/2,f9)*sqrt(v9_a/n9_a+v9_b/n9_b)

cs9_mu <-(m9_a-m9_b)+qt(1-alfa9/2,f9)*sqrt(v9_a/n9_a+v9_b/n9_b)

} else {

f9 = n9_a +n9_b-2

sp9 = sqrt(((n9_a-1)*v9_a+(n9_b-1)*v9_a)/(f9))

ci9_mu = (m9_a-m9_b)-qt(1-alfa9/2,f9)*sp9*sqrt(1/n9_a+1/n9_b)

cs9_mu = (m9_a-m9_b)+qt(1-alfa9/2,f9)*sp9*sqrt(1/n9_a+1/n9_b)

}

signif9_mu = ci9_mu*cs9_mu > 0
```

Pregunta **4**

Sin contestar

Puntúa como 1,40

La latencia de respuesta de un servidor de bases de datos, medida en milisegundos (ms), es una variable aleatoria, X, con función de distribución

$$F(x) = \begin{cases} 0 & \text{si } x < 5, \\ \dfrac{x^2 - 25}{a} & \text{si } 5 \leq x < 10, \\ 1 & \text{si } x \geq 10. \end{cases}$$

El umbral de latencia aceptable (ULA) para el servidor de bases de datos está fijado en 8 ms. Sin embargo, debido a consideraciones de rendimiento, el umbral deseado (UD) es de 7 ms.

a. Calcular la constante a.

b. Hallar la probabilidad de que una respuesta del servidor de bases de datos supere el umbral deseado.

c. Se ha comprobado que una de las respuestas del servidor no supera el umbral de latencia aceptable. ¿Cuál es la probabilidad de que esta respuesta sí supere el umbral deseado?

d. ¿Cual es la latencia media de respuesta del servidor?

e. ¿Cual es la varianza de la latencia de respuesta del servidor?

f. ¿Cual es el percentil 80 de la latencia de respuesta del servidor?

g. El administrador de sistemas necesita obtener una respuesta del servidor de bases de datos que *no* supere el umbral deseado. Realiza solicitudes al servidor, y en caso de que superen el umbral deseado, las descarta y vuelve a enviar la solicitud. ¿Cuál es la probabilidad de que tenga que realizar menos de 2 solicitudes hasta obtener la respuesta que necesita?

h. ¿Cuántas solicitudes se espera que tenga que enviar el administrador de sistemas hasta obtener la respuesta que necesita?

Usando que $F(10) = 1$, obtenemos que la constante a debe valer:

$a = 10^2 - 25$ = 75

Definimos:

$F = function(x)(x^2 - 25)/(a)$

ula = 8

p1 = 1-F(ud)

ud = 7

p2 = (F(ula)-F(ud))/F(ula)

F=function(x){(x$^{2\text{-inf}}$2)/a}

f = function(x){2*x/a}

xf = function(x){2*x^2/a}

x2f = function(x){2*x^3/a}

media = integrate(xf,inf,sup)$value

varianza = integrate(x2f,inf,sup)$value - media^2

per = sample(c(0.6,0.7,0.8),1)

f1 = function(x){integrate(f, inf, x)$value-per}

percentil1 = uniroot(f1,interval=c(0.0001,1000))$root

percentil2 = sqrt(per*a+inf^2)

g = 2

p3 = pgeom(g-2, 1-p1)

mu = 1/(1-p1)

a. La constante a debe valer: 75
b. La probabilidad de que una respuesta del servidor de bases de datos supere el umbral deseado es 0.68
c. La probabilidad de que esta respuesta sí supere el umbral deseado 0.3846154.
d. La latencia media de respuesta del servidor es 7.7777778.
e. La varianza de la latencia de respuesta del servidor es 2.0061728.
f. El percentil 80 de la latencia de respuesta del servidor es 9.2195445.
g. La probabilidad de que tenga que realizar menos de 2 solicitudes hasta obtener una respuesta que no supere el umbral deseado es 0.32.
h. El número esperado de solicitudes es: 3.125.

Pregunta **5**

Parcialmente correcta

Se puntúa 0,91 sobre 1,60

Nota: Tolerancia = 0.005.

Una fábrica de galletas saca al mercado una variedad en cajas de 500 gramos. Según la normativa, se permite una merma en el peso de las cajas de 3%, esto es, el peso de una caja comprada en el mercado debe ser superior a 485 gramos. Después de un amplio estudio, la compañía ha llegado a la conclusión de que el peso real de las galletas que salen al mercado se puede aproximar mediante una variable aleatoria X con distribución normal de media 508 gramos y desviación típica 13.

1. ¿Cuál es la probabilidad de que una caja seleccionada al azar tenga un peso superior al mínimo legal? 0.9615723 ✔ .
2. Un inspector toma una muestra al azar de 10 cajas. ¿Cuál es la probabilidad de que ninguna caja tenga un peso por debajo del mínimo legal? 7.0217e-15 ✖ .
3. La empresa recibe una multa de 1000 euros por cada caja de las seleccionadas que el inspector encuentre con un peso inferior al mínimo legal. ¿Cuál es la multa esperada? 38.42769 ✖ .
4. El inspector se ha dado cuenta de que a la empresa no le importa mucho el importe de la multa y ha decidido endurecer el sistema de sanciones. A partir de ahora, califica la infracción de leve, grave o muy grave, dependiendo de que el peso de una caja esté en $(480, 485]$, $(475, 480]$ o sea inferior a 475 gramos, respectivamente. Si el inspector selecciona una caja al azar, ¿cuál es la probabilidad de que la empresa sea sancionada con una de las siguientes multas?
   - Leve: 0.02280157 ✔ .
   - Grave: 0.01005907 ✔ .
   - Muy grave: 0.005567052 ✔ .
5. Con este nuevo sistema, la multa recibida por cada caja con un peso inferior al legal es de 1000, 5000 y 8000 euros, según la infracción sea leve, grave o muy grave, respectivamente. Se selecciona una caja al azar, ¿cuál es la multa esperada por esa caja? ✖ .

```
mu = 508
sigma=13
(p1 = 1 - pnorm(485, mu, sigma))

n=10
dbinom(10,n, p1)

(multa.esperada = 1000*n*p1)

(p.leve = pnorm(485, mu, sigma) - pnorm(480, mu, sigma))
(p.grave = pnorm(480, mu, sigma) - pnorm(475, mu, sigma))
(p.muygrave = pnorm(475, mu, sigma))

(multa = 0*p1 + 1000*p.leve + 5000* p.grave + 8000* p.muygrave)
```

Pregunta **6**

Incorrecta

Se puntúa 0,00 sobre 1,00

En la consulta de un pediatra, se sabe que la altura de los bebes de un año sigue una distribución Normal y que el percentil 97 es 91.927 centímetros y el percentil 4 es 59.244 centímetros.

a. Si un bebé tiene una altura de 81.07 centímetros, ¿en qué percentil se encuentra? 88 ✖

x1 = 91.927

p1 = 97

x2 = 59.244

p2 = 4

x3 = 81.07

z1 = qnorm(p1/100)

z2 = qnorm(p2/100)

media = x1 - ((x1-x2)/(z1-z2))*z1

sd = ((x1-x2)/(z1-z2))

p3 ← pnorm(x3, media, sd) * 100

a. El percentil correspondiente a una altura de 81.07 es 75.

Actividad previa

Formulario final

Ir a...

Siguiente actividad

Formulas para el primer parcial