Ejercicios resueltos

OCW 2019: *Curso práctico para el análisis e inferencia estadística con Mathematica*

Bloque IV: Inferencia estadística

Equipo docente del curso
Arrospide Zabala, Eneko
Martín Yagüe, Luis
Unzueta Inchaurbe, Aitziber
Soto Merino, Juan Carlos
Durana Apaolaza, Gaizka
Bikandi Irazabal, Iñaki

Departamento de Matemática Aplicada
Escuela de Ingeniería de Bilbao, Edificio II-I

EJERCICIOS DEL BLOQUE 4. INFERENCIA ESTADÍSTICA

Ejercicio nº1

Enunciado

Se está llevando a cabo una campaña de medidas de un nuevo estándar de televisión digital (DVB-T2). Con el objetivo de realizar un estudio de cobertura en diferentes entornos urbanos, se han tomado medidas de la señal recibida en cinco localizaciones de Bilbao y seis localizaciones de Vitoria-Gasteiz. Se sabe que la señal recibida sigue una distribución normal. Para las medidas se ha empleado un receptor estandarizado y calibrado y se han obtenido los siguientes valores del campo eléctrico ($\mu V/m$):

Bilbao	450	415	415	470	450	
Vitoria – Gazteiz	470	480	460	490	430	400

- a) Con un nivel de confianza del 95%, obtenga un intervalo para la media del campo eléctrico que se obtendría en la villa de Bilbao
- b) Con un nivel de significación del 0.05, contraste la hipótesis de que la señal recibida en Bilbao sea de 450 $\mu V/m$
- c) Contraste si la señal recibida en Vitoria-Gasteiz presenta una mayor variabilidad que la recibida en Bilbao con un nivel de significación del 0.05

Resolución

- Introducción de las muestras

```
Bilbao = {450, 415, 415, 470, 450};
Gasteiz = {470, 480, 460, 490, 430, 400};
```

- Cálculo de magnitudes muestrales: tamaño, media y cuasidesviación típica

```
nB = Length[Bilbao]; mB = Mean[Bilbao]; SB = StandardDeviation[Bilbao];

nG = Length[Gasteiz]; mG = Mean[Gasteiz]; SG = StandardDeviation[Gasteiz];
```

- Carga del paquete para estadística inferencial

```
Needs["HypothesisTesting`"]
```

- a) Con un nivel de confianza del 95%, obtenga un intervalo para la media del campo eléctrico que se obtendría en la villa de Bilbao
  - La varianza poblacional de las medidas de señal de Bilbao es desconocida; por tanto, una opción es utilizar la función **StudentTCI**:

```
IC95B = StudentTCI[mB, SB / Sqrt[nB], nB - 1, ConfidenceLevel → 0.95]
{409.904, 470.096}
```

  - El intervalo de confianza pedido es: $\boxed{I_{\mu}^{0.95} = (409.904,\ 470.096)}$

- b) Con un nivel de significación del 0.05, contraste la hipótesis de que la señal recibida en Bilbao sea de 450 $\mu V/m$
  - Resolución con un intervalo de confianza del 95%

$$\boxed{\mu_0 = 450 \in I_\mu^{0.95} = (409.904,\ 470.096)}$$

    - El valor dado pertenece al intervalo obtenido en el apartado anterior por lo que se puede aceptar la hipótesis planteada con un nivel de confianza del 95%
  - Resolución con un contraste de hipótesis (μ_B denota la señal media recibida en Bilbao)

$$\boxed{\begin{array}{l} H_0 : \mu_B = 450 = \mu_0 \\ H_a : \mu_B \neq 450 = \mu_0 \end{array}}$$

    - Cálculo del estadístico de contraste y del p-valor:

```
μ0 = 450;

TTest[Bilbao, μ0, "TestDataTable", AlternativeHypothesis → "Unequal", SignificanceLevel → 0.05]
```

	Statistic	P-Value
T	-0.922531	0.408461

    - El nivel de significación es menor que el p-valor
    - Por lo tanto, el resultado del test no arroja evidencia estadística, con una confianza del 95%, para rechazar la hipótesis nula con lo que se acepta que la señal recibida en Bilbao es de 450 $\mu V/m$

- c) Contraste si la señal recibida en Vitoria-Gasteiz presenta una mayor variabilidad que la recibida en Bilbao con un nivel de significación del 0.05
  - Planteamiento de las hipótesis del contraste: nula y alternativa (σ_B^2 y σ_G^2 denotan las varianzas de las señales recibidas en Bilbao y Gasteiz, respectivamente)

$$\boxed{\begin{array}{l} H_0 : \frac{\sigma_B^2}{\sigma_G^2} \geq 1 \\ H_a : \frac{\sigma_B^2}{\sigma_G^2} < 1 \end{array}}$$

  - Cálculo del estadístico de contraste y del p-valor:

```
FisherRatioTest[{Bilbao, Gasteiz}, 1, "TestDataTable",
 AlternativeHypothesis -> "Less", SignificanceLevel -> 0.05]
```

	Statistic	P-Value
Fisher Ratio	0.51087	0.267473

  - El nivel de significación es menor que el p-valor
  - Por lo tanto, con una confianza del 95%, el resultado del test no arroja evidencia estadística para aceptar que la señal recibida en Gasteiz tiene mayor variabilidad que la recibida en Bilbao
  - Es decir, no hay mayor variabilidad en la señal recibida en Gasteiz

Ejercicio nº2

Enunciado

Una empresa de comida rápida anuncia que el tiempo medio de sus entregas a domicilio es menor o igual a 18 minutos. Se supone que el tiempo de entrega sigue un distribución normal. Los siguientes datos son los tiempos de entrega, en minutos, a 16 clientes:

15.3	19.6	17.3	12.6	21.4	16.3	19.2	20.1
15.7	16.1	17.8	14.9	13.3	20.9	22.5	21.6

- a) Contraste la veracidad del anuncio de la empresa de comida rápida, ¿cuál es el máximo nivel de significación para aceptar la afirmación?
- b) Obtenga un intervalo de confianza para la varianza del tiempo de entrega con un nivel de confianza del 90%

Resolución

- Introducción de la muestra

```
tservicio = {20.4, 19.6, 25.3, 14.6, 21.4,
   16.3, 19.2, 20.1, 16.7, 16.1, 17.8, 14.9, 20.3, 20.9, 22.5, 21.6};
```

- Carga del paquete para estadística inferencial (si es necesario en la sesión)

```
Needs["HypothesisTesting`"]
```

- a) Contraste la veracidad del anuncio de la empresa de comida rápida, ¿cuál es el máximo nivel de significación para aceptar la afirmación?
  - Planteamiento de las hipótesis del contraste: nula y alternativa

$$\boxed{\begin{array}{l} H_0 : \mu \le 18 = \mu_0 \\ H_a : \mu > 18 = \mu_0 \end{array}}$$

  - Cálculo del estadístico de contraste y del p-valor:

```
μ0 = 18;
```

```
TTest[tservicio, μ0, "TestDataTable", AlternativeHypothesis → "Greater"]
```

	Statistic	P-Value
T	1.66607	0.0582198

  - El máximo nivel de significación para aceptar la veracidad del anuncio de la empresa es el p-valor
  - Solución: : $\boxed{\max(\alpha) = p - valor = 0.05822}$

- b) Obtenga un intervalo de confianza para la varianza del tiempo de entrega con un nivel de confianza del 90%

```
ICvar90 = VarianceCI[tservicio, ConfidenceLevel → 0.90]
```

```
{5.24386, 18.052}
```

  - El intervalo de confianza pedido es: $\boxed{I_{\sigma^2}^{0.90} = (5.2439, 18.0520)}$

Ejercicio nº3

Enunciado

El tiempo de uso de un modelo de batería sigue una distribución normal. La empresa que los fabrica cree que el tiempo medio hasta que es necesario cargarla es de 35 horas y la desviación típica 1.7 horas. Se ha seleccionado una muestra aleatoria simple formada por diez baterías y los tiempos resultantes son:

34.3	31.3	36.8	29.8	34.6	31.9	37.1	35.2	34.4	30.7

- a) Obtenga un intervalo de confianza para el tiempo medio de uso de las baterías con un nivel de confianza del 99% y confirme si la empresa está en lo cierto
- b) Contraste si la varianza del tiempo de uso de las baterías es de 1.7 horas con $\alpha = 0.01$

Resolución

- Introducción de la muestra

```
tbateria = {34.3, 31.3, 36.8, 29.8, 34.6, 31.9, 37.1, 35.2, 34.4, 30.7};
```

- Carga del paquete para estadística inferencial (si es necesario en la sesión)

```
Needs["HypothesisTesting`"]
```

- a) Obtenga un intervalo de confianza para el tiempo medio de uso de las baterías con un nivel de confianza del 99% y confirme si la empresa está en lo cierto

```
MeanCI[tbateria, ConfidenceLevel → .99, KnownVariance → None]
```

```
{30.9964, 36.2236}
```

  - El intervalo de confianza pedido es: $\boxed{I_\mu^{0.90} = (30.9964,\ 36.2236)}$
  - Por lo tanto, con una confianza del 99%, puede afirmarse que la empresa está en lo cierto dado que el valor propuesto pertenece al intervalo calculado: $\mu = 35 \in (30.9964,\ 36.2236)$

- b) Contraste si la varianza del tiempo de uso de las baterías es de 1.7 horas con $\alpha = 0.01$
  - Planteamiento de las hipótesis del contraste: nula y alternativa

$$\boxed{\begin{array}{l} H_0 : \sigma^2 = (1.7)^2 = \sigma_0^2 \\ H_a : \sigma^2 \neq (1.7)^2 = \sigma_0^2 \end{array}}$$

  - Cálculo de la región de aceptación y el valor del estadístico del contraste
    - Definición del nivel de significación y tamaño de la muestra

```
alfa = 0.01; n = Length[tbateria];
```

    - Límites de la región de aceptación

```
ji1 = Quantile[ChiSquareDistribution[n - 1], alfa / 2]
```

```
1.73493
```

```
ji2 = Quantile[ChiSquareDistribution[n - 1], 1 - alfa / 2]
```

```
23.5894
```

    - Valor del estadístico del contraste: $\boxed{EC = \dfrac{(n-1)\,S^2}{\sigma_0^2}}$

```
sigma0 = 1.7^2;

ec = (n - 1) * Variance[tbateria] / sigma0
```

20.1415

- El valor del estadístico del contraste pertenece a la región de aceptación del contraste planteado: $ec = 20.1415 \in (1.7349,\ 23.5894)$
- Con una confianza del 99%, la muestra no arroja una evidencia estadística que permita rechazar la hipótesis nula; por tanto, se acepta: $\sigma^2 = (1.7)^2$

- Resolución con la función **VarianceTest**

```
VarianceTest[tbateria, sigma0, "TestDataTable", SignificanceLevel → 0.01]
```

	Statistic	P-Value
Fisher Ratio	20.1415	0.0341176

- Dado que el nivel de significación es menor que el p-valor, se puede aceptar que la desviación típica del tiempo de vida de las baterías es de 1.7 horas con una confianza del 99%
- Para rechazar la hipótesis nula, la significación debería aumentar hasta el valor $\alpha = 0.0341$

Ejercicio nº4

Enunciado

Diez pilotos profesionales han probado dos tipos distintos de neumático hasta que se degradan y no resultan seguros para la conducción. El número de kilómetros recorridos por cada piloto con los dos tipos de neumático se presenta en la siguiente tabla:

Piloto	Neumático A	Neumático B
1	15 100	15 800
2	16 300	16 500
3	17 200	17 000
4	15 700	16 100
5	16 800	16 700
6	17 000	17 500
7	16 700	17 000
8	16 200	16 500
9	15 800	15 500
10	16 300	16 500

Se supone que el número de kilómetros recorridos por ambos tipos de neumático hasta su degradación siguen distribuciones normales. Con un nivel de significación del 10%:

- a) Contraste si los neumáticos del tipo B duran más que los del tipo A
- b) Obtenga un intervalo de confianza para la varianza de los neumáticos del tipo B

Resolución

- Introducción de la muestra

```
neumaticosA = {15 100, 16 300, 17 200, 15 700, 16 800, 17 000, 16 700, 16 200, 15 800, 16 300};

neumaticosB = {15 800, 16 500, 17 000, 16 100, 16 700, 17 500, 17 000, 16 500, 15 500, 16 500};
```

- Cálculo de magnitudes muestrales: tamaño, media y cuasidesviación típica

```
nA = Length[neumaticosA];
mA = Mean[neumaticosA];
SA = StandardDeviation[neumaticosA];
```

```
nB = Length[neumaticosB];
mB = Mean[neumaticosB];
SB = StandardDeviation[neumaticosB];
```

- Carga del paquete para estadística inferencial (si es necesaria en la sesión)

```
Needs["HypothesisTesting`"]
```

- a) Contraste si los neumáticos del tipo B duran más que los del tipo A
  - Planteamiento de las hipótesis del contraste: nula y alternativa

$H_0 : \mu_A \geq \mu_B$	$H_0 : \mu_A - \mu_B \geq 0$
$H_a : \mu_A < \mu_B$	$H_a : \mu_A - \mu_B < 0$

  - Las dos muestras son pareadas, dado que es el mismo piloto el que prueba ambos tipos de neumáticos

```
PairedTTest[{neumaticosA, neumaticosB}, 0, "TestDataTable",
 AlternativeHypothesis → "Less", SignificanceLevel → .10]
```

	Statistic	P-Value
Paired T	−2.	0.0382764

```
PairedTTest[{neumaticosA, neumaticosB}, 0, "TestConclusion",
 AlternativeHypothesis → "Less", SignificanceLevel → .10]
```

```
The null hypothesis that the mean difference is greater than or equal to 0
 is rejected at the 10. percent level based on the Paired T test.
```

  - Se rechaza la hipótesis nula frente a la alternativa con una confianza del 90%
  - La aceptación de la hipótesis nula implicaría reducir la significación hasta el p-valor: $\alpha = 0.0383$
  - Puede plantearse un contraste de hipótesis para la media de una variable
  - Se define una nueva variable D como la diferencia entre los valores observados de los dos tipos de neumáticos:

Piloto	Neumático A	Neumático B	$D = A - B$
1	15 100	15 800	−700
2	16 300	16 500	−200
3	17 200	17 000	200
4	15 700	16 100	−400
5	16 800	16 700	100
6	17 000	17 500	−500
7	16 700	17 000	−300
8	16 200	16 500	−300
9	15 800	15 500	300
10	16 300	16 500	−200

```
neumaticosD = neumaticosA - neumaticosB;
```

  - Las hipótesis del contraste resultan:

$$\boxed{\begin{array}{l} H_0 : \mu_D \geq 0 \\ H_a : \mu_D < 0 \end{array}}$$

- Cálculo de la región de rechazo y el valor del estadístico del contraste
  - Definición del nivel de significación y tamaño de la muestra

```
alfa = 0.10; nD = Length[neumaticosD];
```

  - Cálculo de magnitudes muestrales: media y cuasidesviación típica

```
mD = Mean[neumaticosD]; SD = StandardDeviation[neumaticosD];
```

  - Límite de la región de rechazo

```
t1 = Quantile[StudentTDistribution[nD - 1], alfa / 2]
```

```
-1.83311
```

  - Valor del estadístico del contraste: $\boxed{EC = \dfrac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}}}$

```
ec = mD / (SD / Sqrt[nD])
```

```
-2
```

  - El valor del estadístico del contraste pertenece a la región de rechazo del contraste planteado: $ec = -2 \in (-\infty, -1.83311)$
  - Con una confianza del 90%, la muestra arroja una evidencia estadística que permite rechazar la hipótesis nula; por tanto, se acepta la alternativa: los neumáticos tipo B tienen mayor duración
- Resolución con la función **TTest**

```
TTest[neumaticosD, 0, "TestDataTable", AlternativeHypothesis → "Less"]
```

	Statistic	P-Value
T	-2.	0.0382764

```
TTest[neumaticosD, 0, "TestConclusion",
 AlternativeHypothesis → "Less", SignificanceLevel → 0.10]
```

```
The null hypothesis that the mean of the population is greater than or equal to 0
 is rejected at the 10. percent level based on the T test.
```

- b) Obtenga un intervalo de confianza para la varianza de los neumáticos del tipo B

```
ICvar = VarianceCI[neumaticosB, ConfidenceLevel → 0.90]
```

```
{188487., 959065.}
```

- El intervalo de confianza pedido es: $\boxed{I_{\sigma^2}^{0.90} = (188\,487,\ 959\,065)}$

```
ICsd = Sqrt[ICvar]
```

```
{434.15, 979.319}
```

- El intervalo de confianza para la desviación típica, en km, es: $\boxed{I_{\sigma}^{0.90} = (434.15,\ 979.32)}$