

Ejercicios del primer trabajo de prácticas de Métodos Estadísticos para la Ingeniería

Bloque 1: Análisis de datos

0.1. Control de calidad

Parte de un estudio de control de calidad tuvo como objetivo mejorar una línea de producción. Se midieron los pesos (en onzas) de 50 barras de jabón. Los resultados son los siguientes, ordenados de menor a mayor.

[1] 11.7 12.2 13.2 13.4 14.4 14.4 14.7 14.7 14.8 14.8 15.1 15.2 15.3 15.4 15.5
[16] 15.7 15.7 15.8 15.8 15.9 15.9 15.9 15.9 16.0 16.2 16.4 16.7 16.7 16.8 16.8
[31] 16.8 17.0 17.1 17.2 17.2 17.2 17.4 17.5 17.5 17.5 17.6 17.6 17.8 17.8 17.9
[46] 18.2 18.2 18.7 19.0 19.1

- Halla la media y la varianza de los datos
- Halla la mediana muestral y los cuartiles
- Dibuja un diagrama de cajas con los datos
- ¿Qué proporción de salarios es mayor que la media?
- ¿Qué proporción de los datos está a dos desviaciones típicas de la media? ¿Y a tres desviaciones típicas de la media?
- Dibuja un histograma (representando la probabilidad) y superpón la curva de la distribución Normal que mejor se ajusta a los datos. ¿Dirías que los datos parecen Normales?
- Calcula un intervalo de confianza para la media suponiendo que el valor de la desviación típica de la población es $\sigma = 2$.
- Plantea un contraste para la media de la población suponiendo que nos dicen que su valor es 15. Resuelve el contraste (a un nivel de confianza del 95 %) y calcula el p-valor del estadístico de contraste.

0.2. Conductividad eléctrica

El artículo *Hydrogeochemical Characteristics of Grounwater in a Mid-Western Coastal Aquifer System* (S. Jeen, J. Kim y colaboradores, *Geosciences Journal*, 2001: 339-348) presenta las mediciones de las propiedades de algunas aguas subterráneas poco profundas en cierto sistema de acuíferos de Corea. Las siguientes son las mediciones de la conductividad eléctrica (en microsiemens por centímetro) para 34 muestras de agua:

[1]	2099	528	2030	1350	1018	384	1499	1265	375	424	789	810	522	513	488
[16]	200	215	486	257	557	260	461	500	749	883	711	842	662	846	743
[31]	649	712	673	805	577										

- Halla la media y la varianza de los datos
- Halla la mediana muestral y los cuartiles
- Dibuja un diagrama de cajas con los datos
- ¿Qué proporción de salarios es mayor que la media?
- ¿Qué proporción de los datos está a dos desviaciones típicas de la media?
¿Y a tres desviaciones típicas de la media?
- Dibuja un histograma y superpón la curva de la distribución Normal que mejor se ajusta a los datos. ¿Dirías que los datos parecen Normales?
- Calcula un intervalo de confianza para la media. Usa $\alpha = 0'01$.
- Plantea un contraste para la media de la población suponiendo que nos dicen que su valor es 700. Resuelve el contraste (a un nivel de confianza del 95%) y calcula el p-valor del estadístico de contraste.

0.3. Salario de ingenieros

Un estudio midió el salario (en miles de dólares) de los ingenieros tras diez años de su graduación y estos fueron los resultados:

[1]	30	42	46	50	51	51	54	54	55	56	58	58	60	67	69	89	104	33	49
[20]	53	55	57	57	60	62	62	64	66	68	70	71	73	75	93	35	40	56	60
[39]	62	64	64	67	67	69	70	72	72	78	80	95	101						

- Halla la media y la varianza de los datos
- Halla la mediana muestral y los cuartiles
- Dibuja un diagrama de cajas con los datos
- ¿Qué proporción de salarios es mayor que la media?
- ¿Qué proporción de los datos está a dos desviaciones típicas de la media?
¿Y a tres desviaciones típicas de la media?
- Dibuja un histograma y superpón la curva de la distribución Normal que mejor se ajusta a los datos. ¿Dirías que los datos parecen Normales?
- Calcula un intervalo de confianza para la media. Usa $\alpha = 0'03$.
- Plantea un contraste para el salario medio suponiendo que nos dicen que su valor es 70. Resuelve el contraste (a un nivel de confianza del 95%) y calcula el p-valor del estadístico de contraste.

0.4. Tiempos de espera

Los tiempos de espera se suelen modelizar como una distribución exponencial. Los siguientes valores representan tiempos de espera de una máquina en una línea de producción:

[1]	8.00	2.10	3.80	8.60	7.30	6.10	1.40	2.90	5.50	2.70	4.80	4.60
[13]	1.00	8.70	0.80	0.60	7.20	1.50	1.00	4.00	0.22	0.35	8.90	5.20
[25]	8.80	7.30	0.43	6.00	6.00	10.00	4.80	0.71	4.10	11.00	1.80	

- Halla la media y la varianza de los datos
- Halla la mediana muestral y los cuartiles
- Dibuja un diagrama de cajas con los datos
- ¿Qué proporción de tiempos es menor que la media?
- ¿Qué proporción de los datos está a dos desviaciones típicas de la media? ¿Y a tres desviaciones típicas de la media?
- Dibuja un histograma y superpón la curva de la distribución Exponencial que mejor se ajusta a los datos (es decir, que tiene igual media que los datos)
- Calcula un intervalo de confianza para la media con un coeficiente de confianza de 90%.
- Plantea un contraste para el salario medio suponiendo que nos dicen que su valor es 70. Resuelve el contraste (a un nivel de confianza del 95%) y calcula el p-valor del estadístico de contraste.

0.5. Precios de casas

Los siguientes valores son una muestra de los precios de casas (en miles de dólares) en Estados Unidos:

[1]	184.90	160.00	298.00	269.90	189.90	229.90	175.00	279.50	294.90	324.50
[11]	226.00	329.90	199.90	133.90	324.30	215.30	227.80	208.20	264.00	259.20
[21]	188.00	101.10	202.50	305.20	251.40	177.50	200.20	86.12	244.30	210.50
[31]	147.80	269.40	248.70	326.20						

- Halla la media y la varianza de los datos
- Halla la mediana muestral y los cuartiles
- Dibuja un diagrama de cajas con los datos
- ¿Qué proporción de precios es mayor que la media?
- ¿Qué proporción de los datos está a dos desviaciones típicas de la media? ¿Y a tres desviaciones típicas de la media?
- Dibuja un histograma y superpón la curva de la distribución Normal que mejor se ajusta a los datos. ¿Dirías que los datos parecen Normales?

- Calcula un intervalo de confianza para la media. Utiliza $\alpha = 0'03$.
- Plantea un contraste para el precio medio suponiendo que nos dicen que su valor es 250. Resuelve el contraste (a un nivel de confianza del 95%) y calcula el p-valor del estadístico de contraste.

0.6. Precios de apartamentos en Vitoria

Los siguientes valores representan precios de apartamentos en Vitoria:

[1] 228000.0 409000.0 200000.0 180000.0 443600.0 173000.0 245000.0 210000.0
 [9] 188982.2 193000.0 333000.0 224000.0 560000.0 214400.0 268000.0 230000.0
 [17] 207500.0 254000.0 221000.0 360500.0 250000.0 380000.0 338000.0 372000.0
 [25] 290000.0 270000.0 330000.0 340000.0 270000.0 258000.0 407000.0 178000.0
 [33] 333000.0 403000.0 204000.0 198000.0 330500.0 220000.0 373000.0 230000.0

- Halla la media y la varianza de los datos
- Halla la mediana muestral y los cuartiles
- Dibuja un diagrama de cajas con los datos
- ¿Qué proporción de precios es mayor que la media?
- ¿Qué proporción de los datos está a dos desviaciones típicas de la media?
¿Y a tres desviaciones típicas de la media?
- Dibuja un histograma y superpón la curva de la distribución Normal que mejor se ajusta a los datos. ¿Dirías que los datos parecen Normales?
- Calcula un intervalo de confianza para la media para la media de la población. Usa $\alpha = 0'07$.
- Plantea un contraste para el precio medio suponiendo que nos dicen que su valor es de 250000 euros. Resuelve el contraste (a un nivel de confianza del 95%) y calcula el p-valor del estadístico de contraste.

0.7. Producción de algodón

Los siguientes datos representan la producción de algodón (en miles de kilogramos):

[1] 1.28 1.13 1.63 1.72 1.20 1.09 0.81 1.60 1.39 0.94 1.20 1.54 0.84 1.13 1.62
 [16] 1.17 1.75 1.38 1.13 1.62 0.81 1.20 1.60 1.08 1.36 1.13 1.32 1.13 1.09 1.29
 [31] 1.17 1.63 1.17 1.54 1.52

- Halla la media y la varianza de los datos
- Halla la mediana muestral y los cuartiles
- Dibuja un diagrama de cajas con los datos
- ¿Qué proporción de precios es mayor que la media?
- ¿Qué proporción de los datos está a dos desviaciones típicas de la media?
¿Y a tres desviaciones típicas de la media?

- Dibuja un histograma y superpón la curva de la distribución Normal que mejor se ajusta a los datos. ¿Dirías que los datos parecen Normales?
- Calcula un intervalo de confianza para la media tomando $\alpha = 0'08$.
- Plantea un contraste para la producción media suponiendo que nos dicen que su valor es 1'5. Resuelve el contraste (a un nivel de confianza del 95 %) y calcula el p-valor del estadístico de contraste.

0.8. Tiempos de parada

Los siguientes datos representan los tiempos de parada de un coche que circula a 60 kilómetros por hora con distintos tipos de neumáticos:

[1] 3.94 3.98 4.15 4.47 3.53 4.47 4.17 4.18 4.18 3.74 3.53 3.53 4.34 3.96 4.22
 [16] 4.02 4.25 3.89 3.96 4.22 4.17 3.81 4.18 3.63 3.94 3.96 3.91 3.96 4.09 3.98
 [31] 4.02 4.03 4.02 3.53 3.78 4.34 3.78 4.16

- Halla la media y la varianza de los datos
- Halla la mediana muestral y los cuartiles
- Dibuja un diagrama de cajas con los datos
- ¿Qué proporción de precios es mayor que la media?
- ¿Qué proporción de los datos está a dos desviaciones típicas de la media?
¿Y a tres desviaciones típicas de la media?
- Dibuja un histograma y superpón la curva de la distribución Normal que mejor se ajusta a los datos. ¿Dirías que los datos parecen Normales?
- Calcula un intervalo de confianza para la media para el tiempo medio de parada con $\alpha = 0'02$.
- Plantea un contraste para el tiempo medio de parada suponiendo que nos dicen que su valor es 3'80. Resuelve el contraste (a un nivel de confianza del 95 %) y calcula el p-valor del estadístico de contraste.

0.9. Duración de baterías

Los siguientes datos se midieron durante un experimento en el que se midió la duración de una batería (en días):

[1] 182.33 178.83 183.57 177.33 179.11 181.14 174.22 178.26 179.08 178.89
 [11] 179.96 179.70 178.74 182.65 176.96 179.13 181.94 180.06 179.83 180.78
 [21] 180.47 179.71 181.44 180.74 179.52 177.06 178.81 177.71 175.05 178.77
 [31] 179.57 183.18 183.11 182.22 177.81

- Halla la media y la varianza de los datos
- Halla la mediana muestral y los cuartiles
- Dibuja un diagrama de cajas con los datos

- ¿Qué proporción de precios es mayor que la media?
- ¿Qué proporción de los datos está a dos desviaciones típicas de la media?
¿Y a tres desviaciones típicas de la media?
- Dibuja un histograma y superpón la curva de la distribución Normal que mejor se ajusta a los datos. ¿Dirías que los datos parecen Normales?
- Calcula un intervalo de confianza para la media de la población a un 90 % de confianza.
- Plantea un contraste para la duración media suponiendo que nos dicen que su valor es 175. Resuelve el contraste (a un nivel de confianza del 95 %) y calcula el p-valor del estadístico de contraste.

0.10. Tiempos de parada en boxes

Los siguientes datos son los tiempos de las paradas en boxes de un equipo de Fórmula 1:

```
[1] 5.947 5.664 5.335 5.947 5.981 5.981 5.422 5.257 5.859 5.934 5.335 5.514
[13] 5.292 5.947 5.920 6.245 5.197 5.514 5.920 6.130 4.821 5.243 5.257 5.947
[25] 5.690 5.859 5.422 5.243 5.197 5.422
```

- Halla la media y la varianza de los datos
- Halla la mediana muestral y los cuartiles
- Dibuja un diagrama de cajas con los datos
- ¿Qué proporción de precios es mayor que la media?
- ¿Qué proporción de los datos está a dos desviaciones típicas de la media?
¿Y a tres desviaciones típicas de la media?
- Dibuja un histograma y superpón la curva de la distribución Normal que mejor se ajusta a los datos. ¿Dirías que los datos parecen Normales?
- Calcula un intervalo de confianza para el tiempo medio de parada con $\alpha = 0'09$.
- Plantea un contraste para el tiempo medio de parada suponiendo que nos dicen que su valor es 6 segundos. Resuelve el contraste (a un nivel de confianza del 95 %) y calcula el p-valor del estadístico de contraste.

0.11. Velocidad de saque

Los siguientes datos provienen de una máquina que mide la velocidad en el saque de una pelota de tenis:

```
[1] 66.48 67.49 83.12 66.48 67.49 67.49 74.01 90.24 99.92 83.12
[11] 83.12 56.45 74.84 66.48 105.68 90.24 80.27 56.45 76.50 74.74
[21] 82.93 105.68 90.24 74.01 81.45 99.92 74.01 81.45 80.27 74.01
[31] 97.99 74.01 83.12 90.24 74.84
```

- Halla la media y la varianza de los datos
- Halla la mediana muestral y los cuartiles
- Dibuja un diagrama de cajas con los datos
- ¿Qué proporción de precios es mayor que la media?
- ¿Qué proporción de los datos está a dos desviaciones típicas de la media? ¿Y a tres desviaciones típicas de la media?
- Dibuja un histograma y superpón la curva de la distribución Normal que mejor se ajusta a los datos. ¿Dirías que los datos parecen Normales?
- Calcula un intervalo de confianza para la media. Utiliza $\alpha = 0'025$.
- Plantea un contraste para la velocidad media de saque suponiendo que nos dicen que su valor es 80. Resuelve el contraste (a un nivel de confianza del 95 %) y calcula el p-valor del estadístico de contraste.

Bloque 2: Manipulación de datos en R

0.12. Construcción de matrices

Construye una matriz cuadrada 5×5 con los 25 números naturales comprendidos entre r y $r+24$, siendo r el resto de dividir tu DNI por 9. El resto de dividir x por y se puede obtener en R con $x \% \% y$. Estos números irán introducidos por filas y de izquierda a derecha.

1. Halla la media y la varianza de todos los elementos de la matriz.
2. Halla la media de cada una de las filas y añádela como nueva columna a la derecha de la matriz.
3. Halla la media de cada una de las columnas del apartado anterior y añádelas como nueva fila en la base de la matriz.
4. Halla la suma de los cuadrados de todos los elementos de la matriz.
5. Halla la matriz inversa de la matriz original (si existe).

0.13. Construcción de un vector

Construye un vector x con 100 valores aleatorios correspondientes a una distribución Normal de media tu DNI $\% \% 10$ y varianza tu DNI $\% \% 5 + 1$. Antes de generar los 100 valores pon `set.seed(DNI)`.

1. Muestra los 10 primeros valores de x con una instrucción.
2. Halla la mediana de los valores que superan la media con una instrucción.
3. Halla la media de los valores que están por debajo de la mediana con una instrucción.
4. Halla el valor más próximo a la media con instrucciones (no por observación directa).

Bloque 3: Simulación de datos con R

0.14. Simulación de datos Normales

Tenemos dos variables Normales x y z de parámetros $\mu = 2, \sigma = 3$ y $\mu = 1, \sigma = 4$, respectivamente.

1. Genera 10000 números aleatorios de la variables x y z , tras poner `set.seed(DNI)`
2. Calcula la media y varianza de los datos simulados
3. Muestra los 10 primeros valores generados
4. Calcula la media y varianza aproximadas de la variable $x + z$ utilizando los 10000 valores anteriores de x y z
5. Representa el histograma de esta variable.
6. En vista de lo obtenido, ¿qué puedes decir acerca de las medias y varianzas de las variables x , z y $x + z$? ¿Qué relación tienen?

0.15. Simulación de datos Poisson

Tenemos dos variables de Poisson x e y de medias r =cifra de las unidades de tu DNI (10 si es 0) y s =cifra de las decenas de tu DNI (10 si es 0), respectivamente.

1. Genera 10000 números aleatorios de la variable $x+y$, tras poner `set.seed(DNI)`
2. Muestra los 10 primeros valores generados
3. Calcula la media y varianza aproximadas de la variable $x + y$ mediante los 10000 valores
4. Construye el correspondiente diagrama de barras para $x + y$
5. Genera 10000 números aleatorios (tras poner `set.seed(DNI)`) de una variable de Poisson de media $r + s$
6. Representa el histograma de esta variable.
7. En vista de lo obtenido, ¿qué relación crees que existe entre las variables z , y y $x + y$?

0.16. Simulación del índice de masa corporal

Representa mediante un histograma el índice de masa corporal de un conjunto de 1000 individuos. Para ello simula la obtención de una muestra para el peso con una media de 80 kg y una desviación típica de 10 kg. Haz lo mismo para la estatura, suponiendo una media de 170 cm y una desviación típica de 15 cm. El índice de masa corporal se calcula como el cociente entre el peso y la altura al cuadrado.

0.17. Simulación de una encuesta

Simula una encuesta en las que preguntes 50 veces a 10 personas si fuman o no, siendo la probabilidad de fumar 0'25. Contando el número de veces que obtienes un valor de 5 para la variable “nº de personas que fuman”, calcula con R la probabilidad de obtener un 5 y compara el resultado con el valor teórico. Repite el ejercicio realizando el experimento 100 y 1000 veces.

0.18. Número de motores defectuosos

Considera el problema en el que se quiere ver el número de motores defectuosos de una muestra de 25 motores (supón que la probabilidad de ser defectuosos es $\pi = 0'2$). Obtén el histograma de la media de 30 de estos experimentos. A continuación, superpón una distribución Normal que aproxime la distribución de la media de esos 30 experimentos.

0.19. Simulación de variables Uniformes

Obtén el histograma de la media de 10 variables aleatorias uniformes en $[0,1]$ a partir de 10000 repeticiones de este experimento. Superpón el gráfico de la distribución normal de igual media y desviación típica que las medias muestrales. ¿Se puede afirmar que la media de 10 variables aleatorias uniformes se ajusta bien a una Normal, a pesar de que el tamaño de la muestra es pequeño?

0.20. Simulación de la proporción de caras

Simula la evolución de proporción de caras al tirar una moneda equilibrada. Hazlo con 10000 tiradas. Representa en un gráfico la proporción de caras en función del número de tiradas y la línea hacia la que se estabiliza la proporción.