

Práctica 7: Tests estadísticos

1. La tabla siguiente contiene una muestra de 21 números extraídos de una población con una función de densidad de probabilidad Uniforme en el intervalo $[2,3]$.

2.62	2.10	2.23	2.54	2.31	2.50	2.60
2.67	2.18	2.12	2.05	2.41	2.69	2.37
2.92	2.92	2.78	2.28	2.74	2.82	2.07

- a) Mediante el uso de tests paramétricos, analice si la media y varianza de la muestra tienen diferencias significativas respecto de los parámetros de la función de probabilidad teórica. Utilice un nivel de significancia del 10%.
- b) Mediante un test de bondad de ajuste, analice si la muestra se ajusta a la función de probabilidad teórica con un nivel de significancia del 5%.
2. La tabla siguiente contiene las mediciones de un objeto utilizando dos instrumentos:

Inst. 1	Inst.2
100	97
101	102
103	103
98	96
97	100
100	101
102	100
101	
109	
101	

- a) Calcule un valor representativo (promedio) y su desviación estándar para cada instrumento.
- b) Verifique la validez de la hipótesis: las varianzas poblacionales de ambas muestras son iguales.
- c) Verifique la validez de la hipótesis: las medias poblacionales de ambas muestras son iguales.
- d) Si los resultados de los puntos b) y c) lo habilitan, estime un valor representativo para ambas muestras.
3. Dado el siguiente conjunto de números:

0.201	0.133	0.568	0.496	0.057	0.707	0.645	0.165	0.281
0.823	0.985	0.805	0.743	0.835	0.502	0.094	0.278	0.546
0.277	0.395	0.398	0.624	0.465	0.967	0.788	0.250	0.367

- a) Realice un histograma de paso constante.

- b) Analizando visualmente el histograma, podría sugerir alguna función de densidad de probabilidad que represente lo que muestra el gráfico.
- c) Aplique tests paramétricos para verificar si los parámetros de la función de densidad de probabilidad propuesta en el punto anterior difieren significativamente de los valores teóricos. Aclare cuál es el nivel de significancia utiliza.
- d) Aplique un test de bondad de ajuste para verificar si la función de densidad propuesta ajusta bien a la distribución observada de números. Aclare cuál es el nivel de significancia que utiliza.
4. Un generador de números pseudo – aleatorios es un método iterativo que permiten generar una serie de valores aleatorios (con función de densidad de probabilidad uniforme) dentro de un rango de valores específicos, a partir de un número inicial. Dos de los métodos más utilizados están expresados por las siguientes expresiones: $x_{i+1} = mantiza(997 \cdot x_i)$ $x_{i+1} = mantiza(\pi \cdot x_i)$, donde x_0 es el número semilla y tiene que ser un valor en 0 y 1.
- a) Genere 5 pares de series de números, una serie para cada método. Comience con un número semilla de un solo decimal y luego, vaya agregando un decimal en para cada nueva serie. Genere series de por lo menos 100 números.
- b) Aplique un test de bondad de ajuste para verificar cuál de los generados de números aleatorios analizados es más efectivo.
5. Se toman 300 muestras aire, todas de igual volumen, en una región y se cuenta la cantidad de partículas que cada una contiene. Luego, se analizan los resultados y se confecciona la siguiente tabla que resumen el número de muestras que se encontró con un determinado número de partículas.

Nro. de partículas	0	1	2	3	4	5	6	7
Cant. de muestras	38	75	89	54	20	19	4	1

- a) Calcule las frecuencias observadas.
- b) Encuentre la función de probabilidad de Poisson que mejor representa las frecuencias observadas.
- c) Realice un gráfico ‘Cant. de muestras vs. Nro. de partículas’, donde aparezcan las frecuencias observadas, las frecuencias teóricas (frecuencias esperadas) y la curva que representa la función obtenida en b).

d) Aplique un test de bondad de ajuste con nivel de significancia del 5% para verificar si la función de probabilidad propuesta representa bien las frecuencias observadas.

6. La tabla siguiente muestra la cantidad de causas judiciales que cada juez recibió en los últimos años. La asignación de causas se realiza a través de un sorteo donde cada juez tiene la misma probabilidad de salir sorteado.

Mediante un test estadístico con nivel de significancia del 5%, analice si la asignación de causas muestra alguna desviación del comportamiento esperado. Justifique bien su respuesta.

Juez	1	2-3	4	5	6	7-9	10	11	12
Causas	2712	5005	2367	2453	2680	9013	2713	2864	2925

7. Un experimento consiste en arrojar tres dados a la vez un total de 100 veces y contar la cantidad de seis que salen a en cada tirada. El resultado obtenido es la tabla que se muestra a continuación. Mediante un test de Chi-cuadrado estudie si los dados estaban cargados o no.

Nro de seis	0	1	2	3
Cant. de casos	47	35	14	4

8. Se quiere analizar la hipótesis que la concentración de polvo atmosférico es la misma para todo el planeta. Se acepta que una buena variable para este tipo de estudios es la concentración de partículas que se puede medir en la nieve perenne.

La tabla siguiente muestra dos series de mediciones de concentración de polvo atmosférico en muestras de nieve perenne, en unidades de partes en 1000 millones. La serie S1 corresponde a mediciones realizadas en Groenlandia (Ártico), mientras que la serie S2 corresponde a una base localizada en la Antártica.

S1	3.7	7.8	1.9	2.0	1.1	1.3	1.9	3.7	3.4	1.6	2.4	1.3	2.6	3.7	2.2	1.8	1.2	0.8
S2	3.7	2.0	1.3	3.9	0.2	1.4	4.2	4.9	0.6	1.4	4.4	3.2	1.7	2.1	4.2	3.5		

- Calcule las medias y varianzas muestrales para cada serie.
- Realice un test paramétrico para verificar si existen diferencias significativas entre las varianzas de ambas poblaciones. Especifique claramente la hipótesis que está testeando.
- Realice un teste paramétrico para verificar si existen diferencias significativas entre las medias de ambas poblaciones. Especifique claramente la hipótesis que está testeando.

- d) En base a los resultados de los puntos b) y c), qué conclusión puede sacar respecto de la hipótesis planteada al comienzo del ejercicio.
9. El archivo ‘P7E7 – 2022.txt’ contiene dos columnas de números, las que corresponden a una muestra del par de variables aleatorias (x, y) . A partir de este archivo:
- Realice dos histogramas con paso 0.5, uno para cada variable.
 - Calcule las medias y desviación estándar de la muestra para cada variable. Interprete estos parámetros en relación a los histogramas obtenidos en el punto anterior.
 - Realice un gráfico de dispersión ‘y vs. x’.
 - Analice visualmente si el gráfico anterior sugiere que exista alguna relación lineal entre las variables.
 - Calcule el coeficiente de correlación entre las variables.
 - Aplice el test de significancia de la correlación* para verificar si el coeficiente calculado en a) es significativamente diferente de cero. Explique el significado del resultado del test.
 - Estime mediante Mínimos Cuadrados los coeficientes del modelo $y_i = a x_i + b + v_i$, donde v_i representa el error de medición.
 - Interprete el resultado en g) en relación a la conclusión obtenida en f) usando el test estadístico.
10. Repita los pasos del ejercicio anterior para el archivo ‘P7E8 – 2022.txt’.

*Resumen del Test de significancia de correlación

Hipótesis nula: el coeficiente de correlación de la población es igual a 0, $\rho = 0$.

Hipótesis alternativa: el coeficiente de correlación de la población es diferente de 0, $\rho \neq 0$.

- Se calcula el estadístico $t^* = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ donde r es el coeficiente de correlación de la muestra, n es el tamaño de la muestra.
- Utilizando la tabla de la función de t-student, se obtiene el valor de $t_{1-\alpha, n-2}$ (nivel de significancia α y $n-2$ grados de libertad).

3) Si $t^* > t$ entonces se rechaza la hipótesis nula y se puede decir que ‘existe suficiente evidencia estadística, con significancia α , para concluir que existe una relación lineal entre las dos variables’.

Si $t^* < t$ entonces se dice que ‘no se puede rechazar la hipótesis nula’ y que ‘no existe suficiente evidencia estadística, con significancia α , para concluir que existe una relación lineal entre las dos variables’.