

Inferencia Estadística II, Examen I

Luis Alberto May Custodio

21 de noviembre de 2024

1. De 400 electores incluidos en un estudio, 212 expresaron su apoyo a la política del gobierno, mientras que los otros 188 manifestaron su oposición. Después de esto, el gobierno introdujo un presupuesto y posteriormente los mismos 400 electores fueron entrevistados nuevamente. Ahora se encontró que 196 apoyan la política gubernamental, incluyendo a 17 que anteriormente se habían opuesto.

- a) Explique por qué no sería válido llevar a cabo una prueba de independencia con la siguiente tabla:

	Apoyan	Se oponen
Antes del presupuesto	212	188
Después del presupuesto	196	204

- b) Otra manera de tabular los datos es la siguiente:

	Apoyan después	Se oponen después
Apoyan antes	179	33
Se oponen antes	17	171

Realiza una prueba de independencia en esta tabla y explica cuidadosamente lo que significa el resultado.

- c) Prueba la hipótesis de que la proporción de votantes que apoya la política gubernamental es la misma antes y después de la introducción del presupuesto.
2. Cuando alguien lanzó una moneda 200 veces, se observaron 115 caras y 85 sellos. ¿Es esa evidencia para pensar que la moneda es justa? Prueba la hipótesis de equidad usando un nivel de significancia de (a) 0.05, (b) 0.01. (c) Calcula el valor p de la prueba.
3. En el pasado, los científicos solían confiar en tablas de números aleatorios. Una tabla de números aleatorios de 250 dígitos tenía la distribución de los dígitos 0, 1, 2, ..., 9 mostrada en Cuadro 1.

Cuadro 1: Frecuencias observadas de dígitos										
Dígito	0	1	2	3	4	5	6	7	8	9
Frecuencia	17	31	29	18	14	20	35	30	20	36

- a) ¿Difiere la distribución observada significativamente de la distribución esperada?
- b) ¿Cuál es el valor p de la observación?

Cuadro 2: Distribución de niños y niñas en familias de 5 hijos

Número de niños	5	4	3	2	1	0
Número de familias	18	56	110	88	40	8

4. Una encuesta de 320 familias con 5 hijos cada una reveló la distribución de niños y niñas mostrada en el Cuadro 2.
 - a) ¿Es el resultado consistente con la hipótesis de que los nacimientos de varones y mujeres son igualmente probables?
 - b) ¿Cuál es el valor p de los resultados muestrales?
5. La tienda de alimentos Fresh N Friendly anuncia que sus tiempos de espera en la caja son de cuatro minutos o menos. Un cliente enojado quiere disputar esta afirmación. Toma una muestra aleatoria de compradores en la hora pico y registra sus tiempos de espera en la caja. ¿Puede disputar su afirmación a un nivel de significancia del 10
Tiempos de espera en la caja (en minutos): 3.8, 5.3, 3.5, 4.5, 7.2, 5.1
6. Considera los datos a continuación. ¿Hay evidencia para pensar que la distribución asociada a la Aleación I (F_{AI}) es estocásticamente mayor que la distribución asociada a la Aleación II (F_{AII})? Establece tu hipótesis a probar. De acuerdo con tu conclusión, ¿qué significa eso respecto a la comparación de la resistencia de los cables hechos de la Aleación I y la Aleación II?
 - **Aleación I:** 18.3, 16.4, 22.7, 17.8, 18.9, 25.3, 16.1, 24.2
 - **Aleación II:** 12.6, 14.1, 20.5, 10.7, 15.9, 19.6, 12.9, 15.2, 11.8, 14.7
7. Dos fabricantes de computadoras, Computadoras & Compañía (CC) y Binario (B), compiten por un contrato lucrativo. Por supuesto, cada compañía afirma que sus computadoras son más rápidas, pero ¿cómo debería un cliente elegir entre ellas? Se decidió ejecutar simultáneamente el mismo programa en siete computadoras de cada compañía y ver cuáles terminan primero. Los resultados fueron los siguientes: dos computadoras producidas por CC terminaron primero, seguidas por tres computadoras producidas por B, luego cinco computadoras producidas por CC, y finalmente, cuatro computadoras producidas por B. Los tiempos reales no fueron registrados. Basado en la información, ¿podemos decir que las computadoras producidas por CC son estocásticamente más rápidas que las producidas por B?
8. Los gerentes de un portal de compras en línea sospechan que más clientes participan en compras en línea si se les ofrece alguna recompensa, como un descuento o devolución de efectivo. Para verificar esto, eligieron al azar 12 días y ofrecieron un 5% de descuento en 6 días al azar, pero no ofrecieron incentivos en los otros 6 días. Con el descuento, el portal recibió 1215, 1698, 2629, 1476, 2404 y 2109 visitas. Sin el descuento, se registraron 1433, 889, 1321, 1842, 711 y 1022 visitas. ¿Apoya esto la hipótesis de los gerentes?

9. De acuerdo con la teoría genética, los tipos de sangre MM, MN y NN deberían ocurrir en una población muy grande, con las frecuencias relativas θ^2 , $2\theta(1 - \theta)$ y $(1 - \theta)^2$, donde θ es la frecuencia génica (desconocida).
- a) Las frecuencias observadas en una muestra de tamaño 100 de una población fueron 33, 44 y 23, respectivamente. Pruebe la bondad de ajuste del modelo a estos datos.
- b) Suponga que las frecuencias observadas en una muestra de tamaño 400 fueron cuatro veces las mencionadas en (a). Realice una prueba de bondad de ajuste y explique por qué se obtiene un resultado diferente al obtenido en (a).
10. Un vendedor de semillas afirma que su semilla de chícharo dulce tiene una tasa de germinación de 80 %. Un cliente compró cuatro paquetes de semillas de chícharo dulce, cada paquete de un color diferente (rojo, blanco, azul y amarillo). Posteriormente plantó las 100 semillas de cada paquete y los números de semillas que germinaron en un mes fueron los siguientes:

Cuadro 3: Número de semillas que germinaron y no germinaron por color

	Rojo	Blanco	Azul	Amarillo
Germinaron	75	66	81	74
No germinaron	25	34	19	26

- a) Pruebe la hipótesis de que la tasa de germinación es del 80 % para los cuatro colores.
- b) Pruebe la hipótesis de que la tasa de germinación es la misma para los cuatro colores (con una tasa no necesariamente del 80 %).
11. En la siguiente tabla se clasifican 64 ternas de acuerdo a la edad de la madre al nacer y la distribución de los sexos.

	3 niños	2 niños	2 niñas	3 niñas	Total
Madre por debajo de 30	5	8	9	7	29
Madre por encima de 30	6	10	13	6	35
Total	11	18	22	13	64

- a) ¿Existe alguna evidencia de una asociación entre la distribución del sexo y la edad de la madre?
- b) Suponga que la probabilidad de nacimiento de un varón es 0.50 y que los sexos de las ternas se determinan independientemente. Encuentre la probabilidad de que haya x niños en una terna ($x = 0, 1, 2, 3$) y pruebe si los totales por columna son consistentes con esta distribución.
12. En un estudio se recolectaron algunas características de accidentes de motocicleta. Como parte del estudio, se enviaron cuestionarios a individuos que habían estado involucrados en choques de motocicletas. Una pregunta de interés fue la posible relación entre

Cuadro 4: Relación entre daño en la cabeza y uso de casco

	Sin daño	Daño menor	Daño serio
Usando casco	165	20	33
Sin usar casco	262	53	93

la ocurrencia de daño en la cabeza y el uso de casco. Los siguientes datos reportados corresponden a 626 conductores accidentados que respondieron el cuestionario.

- a) ¿Hay alguna evidencia de una diferencia en las frecuencias relativas de los diferentes tipos de daño entre los dos grupos (uso de casco contra no uso de casco)?
 - b) De aquellos que recibieron un daño en la cabeza, ¿existe alguna evidencia de una diferencia en la frecuencia de daño serio contra daño menor para los dos grupos?
 - c) En estos datos vemos que solo 218 de 626 (35 %) usaron casco. De acuerdo con los datos, ¿existe alguna evidencia de que el uso de casco reduce la posibilidad de un daño en el accidente?
13. Con el fin de investigar si dos poblaciones eran diferentes, se recolectaron observaciones (en escala continua) X_1, X_2, \dots, X_{20} y Y_1, Y_2, \dots, Y_{20} , provenientes de dichas poblaciones, con el fin de aplicar una prueba de suma de rangos para resolver nuestra pregunta. En la siguiente tabla se muestran los valores muestrales observados:

Cuadro 5: Valores muestrales observados

Muestra X					Muestra Y				
-2.604	-2.594	-1.921	-1.756	-1.449	-2.346	-1.113	0.207	0.356	0.873
-0.695	-0.676	-0.624	-0.555	-0.157	0.897	1.074	1.125	1.629	1.796
0.442	0.786	0.932	0.932	1.691	2.015	2.174	2.315	2.630	2.737
1.697	1.777	2.961	2.976	6.297	3.178	3.610	3.714	4.559	4.847

Como las muestras son grandes, podemos utilizar una estadística de prueba:

$$Z_X = \frac{T_X - \mathbb{E}[T_X]}{\sqrt{\mathbb{V}[T_X]}},$$

distribuida aproximadamente como una normal estándar. Comente el resultado de la prueba de suma de rangos aplicada a los datos anteriores.

14. Los siguientes datos muestran la edad al diagnóstico de diabetes tipo II en adultos jóvenes. ¿Es diferente la edad al diagnóstico para hombres y mujeres?
- **Hombres:** 19, 22, 16, 29, 24
 - **Mujeres:** 20, 11, 17, 12