

# Test de hipótesis

Análisis estadístico de datos

2021

1. **Fernet con Coca.** Una vez en un bar un cliente rechazó un vaso de Fernet con Coca argumentando que contenía Pepsi. La barwoman enojada le ofreció cuatro vasos de Fernet con Coca y cinco vasos de Fernet con Pepsi y desafió al cliente a seleccionar de la muestra los cuatro vasos de Fernet con Coca. Identificar para este test la estadística de prueba y la hipótesis nula. Calcular la significancia del test.
2. Considerar un índice para medir la capacidad de una persona para “controlar el enojo”. El índice fue calculado en base a las respuestas a un cuestionario para varones adultos. Se encontró que esta estadística sigue una distribución normal con media  $\mu = 100$  y desviación estándar  $\sigma = 30$ . Unos investigadores hicieron el cuestionario a un grupo de 10 varones presos por crímenes violentos y encontraron que el índice promedio fue 50.8. Plantear una hipótesis nula y otra alternativa para decidir si los presos de crímenes violentos tiene más dificultades para controlar el enojo que las personas de la población general. Determinar cuál es la estadística de prueba. Decidir cuál es la región crítica para rechazar la hipótesis nula con una significancia del 5%. Decidir si en base a las respuestas al cuestionario se puede concluir que los presos por crímenes violentos tienen menos control del enojo que los varones adultos en general (adaptado de [1]).
3. Hacer un histograma con los datos del archivo *chi2.dat*. Graficar el histograma y una función de densidad de probabilidad normal estándar (no ajustar los datos). Tomando como hipótesis nula que los datos siguen esta PDF, calcular el chi-cuadrado del histograma y su pvalor. Comparar el pvalor con el obtenido por el resto de la clase y explicar las diferencias encontradas.
4. Student usó como ejemplo en su artículo de 1908 [2] el rendimiento de la cebada cervecera. Para la prueba se sembraron 11 muestras de cebada de distintas variedades con semillas secadas al aire y otras 11 muestras con semillas secadas en horno. Se busca establecer si el secado en horno mejora el rinde con respecto al secado al aire con una significancia del 5%. En las pruebas se midieron los rendimientos de

Muestra	Aire	Horno
1	1903	2009
2	1935	1915
3	1910	2011
4	2496	2463
5	2108	2180
6	1961	1925
7	2060	2122
8	1444	1482
9	1612	1542
10	1316	1443
11	1511	1535

Table 1: Rendimiento de la cosecha de grano de cebada en libras por acre. Las muestras corresponden a distintas variedades de cebada plantadas en los años 1899 y 1900. .

la tabla 1. Antes de examinar los datos, establecer la estadística de prueba, la hipótesis nula y alternativa y la región crítica. Decidir si los datos indican que las semillas secadas en horno tiene mejor rinde que las secadas al aire.

5. **(Para entregar)** La diabetes se diagnostica en base al nivel de glucosa en sangre. Según la base de datos de indios Pima la glucosa en sangre de personas sanas tiene una media  $\mu_0 = 110.6$  mg/dl y una desviación estándar  $\sigma_0 = 24.8$  mg/dl. Para personas con diabetes la media es  $\mu_1 = 142.3$  mg/dl y la desviación estándar  $\sigma_1 = 29.6$  mg/dl. Asumir que la glucosa en sangre para personas sanas y con diabetes sigue una distribución normal. Siguiendo las recomendaciones de la Organización Mundial de la Salud, a una persona se le diagnostica diabetes si su nivel de glucosa en sangre excede 126 mg/dl. Graficar las distribuciones para la hipótesis nula y la alternativa. Indicar en la figura el valor crítico de la glucosa en sangre. Calcular la precisión del test de glucosa en sangre, la probabilidad que un paciente sano sea diagnosticado como diabético (falso positivo) y la probabilidad que un paciente diabético sea diagnosticado como sano (falso negativo). Si la prevalencia de la diabetes es del 35%, calcular aplicando el teorema de Bayes, la probabilidad que una persona con un test positivo efectivamente tenga diabetes.

## References

- [1] David Weisburd y Chester Britt, *Statistics in Criminal Justice*, Nueva York, 2014
- [2] Student. “The Probable Error of a Mean.” *Biometrika*, vol. 6, no. 1, [Oxford University Press, Biometrika Trust], 1908, pp. 1–25, <https://doi.org/10.2307/2331554>.