Problemas de Bioestadística

Alfredo Sánchez Alberca
asalber@ceu.es
https://aprendeconalf.es

Tabla de contenidos

Prefacio

Colección de problemas de Bioestadística para el Master en Nutrigenómica.

1 Contrastes de hipótesis paramétricos

Ejercicio 1.1. Sabiendo que el año pasado el consumo per cápita de azúcar en España fue de 4.8 kg y que este consumo sigue una distribución normal, hemos seleccionado aleatoriamente a 20 españoles obteniendo una media muestral de 5 kg y una cuasidesviación típica muestral de 0.4 kg. Contrastar la hipótesis de que el consumo de azúcar per cápita de este año en España no ha variado utilizando un nivel de significación del 10%.

Solución

Sea $X \sim N(\mu, \sigma)$ la variable aleatoria que mide el consumo per cápita de azúcar en España.
Número de poblaciones: 1 población. Variable dependiente: Consumo de azúcar (cuantitativa).
Tamaño muestral: $n = 20$.
Contraste de la media de una población normal (prueba t): $H_0 : \mu = 4.8$, $H_1 : \mu \neq 4.8$.
Estadístico del contraste: $t = 2.24$
p-valor: $p = 2P(t(19) > 2.24) = 0.037$.
Como el p-valor es menor que el nivel de significación $\alpha = 0.1$, se rechaza la hipótesis nula y se concluye que el consumo de azúcar per cápita de este año en España ha variado.

Ejercicio 1.2. En una clase de alumnos de primaria se ha comprobado que el 20% del alumnado consume bollería industrial. Para disminuir esta preocupante cifra, el colegio ha implantado un programa de educación nutricional. Después del programa se tomó una muestra aleatoria de 50 alumnos de primaria y se observó que 8 seguían consumiendo bollería industrial. Contrastar con un nivel de significación del 5% si el programa es efectivo.

Solución

Número de poblaciones: 1 población.
Variable dependiente: Consumo de bollería industrial (cualitativa binaria).
Tamaño muestral: $n = 50$.
Contraste para la proporción de una población: $H_0 : p = 0.2$, $H_1 : p < 0.2$.
p-valor: $p = 0.2979$.

Como el p-valor es mayor que el nivel de significación $\alpha = 0.05$, no se rechaza la hipótesis nula y se concluye que no hay evidencias significativas de que el programa sea efectivo.

Ejercicio 1.3. Se sabe que el consumo anual de helado correspondiente a cada español sigue una distribución normal y que el año pasado el consumo medio fue de 20 kg. Queremos contrastar si este año se va a mantener el consumo medio de helado que el año pasado, y para comprobarlo se efectúa una muestra aleatoria de 22 españoles, obteniéndose los resultados recogidos en el fichero consumo-helado.csv. Realizar el contraste con un nivel de significación de 0.10.

Solución

Número de poblaciones: 1 población.
Variable dependiente: Consumo de helado (cuantitativa).
Tamaño muestral: $n = 22$.
Prueba de normalidad de Kolmogorov-Smirnov. p-valor: $p = 0.632$. Se acepta la normalidad.
Contraste de la media de una población normal (prueba t): $H_0 : \mu = 20$, $H_1 : \mu \neq 20$.
p-valor: $p = 0.8033$.
Como el p-valor es mayor que el nivel de significación $\alpha = 0.10$, no se rechaza la hipótesis nula y se concluye que el consumo medio de helado de este año en España se va a mantener.

Ejercicio 1.4. Se ha realizado un estudio para comparar los niveles de hierro en sangre antes y después de un programa de ejercicios. Los datos del estudio están en el fichero niveles-hierro-ejercicio.csv. Realizar el contraste de hipótesis adecuado para ver si el ejercicio aumenta el nivel de hierro con un nivel de significación del 1%.

Solución

Número de poblaciones: 2 poblaciones pareadas (población 1: antes de hacer ejercicio, población 2: después de hacer ejercicio).
Variable dependiente: Nivel de hierro (cuantitativa).
Tamaño muestral: $n = 20$.
Prueba de normalidad de Kolmogorov-Smirnov: p-valor de la diferencia: 0.804. Se acepta la normalidad.
Contraste para la comparación de medias de poblaciones pareadas: $H_0 : \mu_d = 0$, $H_1 : \mu_d < 0$.
p-valor: $p = 0.0011$.
Como el p-valor es menor que el nivel de significación se rechaza la hipótesis nula y

se concluye que el programa de ejercicios ha aumentado significativamente el nivel de hierro en sangre.

Ejercicio 1.5. Se ha realizado un estudio para comparar los niveles de glucosa en sangre de dos grupos: uno sigue una dieta baja en carbohidratos y otro una dieta estándar. Los datos del estudio están en el fichero niveles-glucosa-dietas.csv. Realizar el contraste de hipótesis adecuado para comparar las medias de los dos grupos con un nivel de significación del 5%.

Solución

Número de poblaciones: 2 poblaciones independientes (población 1: dieta baja en carbohidratos, población 2: dieta estándar).
Variable dependiente: Nivel de glucosa (cuantitativa).
Tamaños muestrales: $n_1 = 30$, $n_2 = 30$.
Contraste de comparación de varianzas: $H_0 : \sigma_1^2 = \sigma_2^2$, $H_1 : \sigma_1^2 \neq \sigma_2^2$. p-valor: $p = 0.25$. Se acepta la hipótesis nula de igualdad de varianzas.
Contraste para la comparación de medias de poblaciones independientes con varianzas iguales: $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 < \mu_2$.
p-valor: $p = 0.00009$. Como el p-valor es menor que el nivel de significación se rechaza la hipótesis nula y se concluye que el nivel de glucosa en sangre es significativamente menor en el grupo que sigue una dieta baja en carbohidratos.

Ejercicio 1.6. Se ha realizado un estudio para comparar los niveles de expresión de un gen en función de la dieta (vegetariana u omnívora). Los datos del estudio están en el fichero expresion-genica-dietas.csv. Realizar el contraste de hipótesis adecuado para comparar las medias de los dos grupos.

Solución

Número de poblaciones: 2 poblaciones independientes (población 1: dieta vegetariana, población 2: dieta omnívora).
Variable dependiente: Expresión génica (cuantitativa).
Tamaños muestrales: $n_1 = 20$, $n_2 = 25$.
Prueba de normalidad de Kolmogorov-Smirnov: p-valor primera muestra: 0.4164, p-valor segunda muestra: 0.2767. Se acepta la normalidad en ambos grupos.
Contraste de comparación de varianzas: $H_0 : \sigma_1^2 = \sigma_2^2$, $H_1 : \sigma_1^2 \neq \sigma_2^2$. p-valor: $p = 0.1373$. Se acepta la hipótesis nula de igualdad de varianzas.
Contraste para la comparación de medias de poblaciones independientes con varianzas iguales: $H_0 : \mu_1 = \mu_2$, $H_1 : \mu_1 < \mu_2$.
p-valor: $p = 0.00006$.

Como el p-valor es menor que el nivel de significación, se rechaza la hipótesis nula y se concluye que existen diferencias estadísticamente significativas en los niveles de expresión génica en función de la dieta.

Ejercicio 1.7. Se utiliza un grupo de 150 pacientes para comprobar la teoría de que la vitamina C tiene alguna influencia en el tratamiento del cáncer. Los 150 pacientes fueron divididos en dos grupos de 75. Un grupo recibió 10 gramos de vitamina C y el otro un placebo cada día, además de la medicación habitual. De los que recibieron la vitamina C, 47 presentaban alguna mejoría al cabo de cuatro semanas, mientras que de los que recibieron el placebo, 43 experimentaron mejoría. Contrastar la hipótesis de que la vitamina C tiene influencia en el tratamiento del cáncer con un nivel de significación del 5%.

Solución

Número de poblaciones: 2 poblaciones independientes (población 1: vitamina C, población 2: placebo).
Variable dependiente: Mejoría en el tratamiento del cáncer (cualitativa binaria).
Tamaños muestrales: $n_1 = 75$, $n_2 = 75$.
Contraste para la comparación de proporciones: $H_0 : p_1 = p_2$, $H_1 : p_1 \neq p_2$.
p-valor: $p = 0.6824$.
Como el p-valor es mayor que el nivel de significación, no se rechaza la hipótesis nula y se concluye que no hay evidencias significativas de que la vitamina C tenga influencia en el tratamiento del cáncer.

Ejercicio 1.8. En un estudio sobre el consumo de alcohol entre los jóvenes durante los fines de semana, se preguntó a 100 chicos y a 125 chicas, de los que 63 chicos y 59 chicas contestaron que consumían. En vista de estos datos, ¿existe alguna diferencia significativa entre las proporciones de chicos y chicas que consumen alcohol los fines de semana?

Solución

Número de poblaciones: 2 poblaciones independientes (población 1: chicos, población 2: chicas).
Variable dependiente: Consumo de alcohol (cualitativa binaria).
Tamaños muestrales: $n_1 = 100$, $n_2 = 125$.
Contraste para la comparación de proporciones: $H_0 : p_1 = p_2$, $H_1 : p_1 \neq p_2$.
p-valor: $p = 0.0163$.
Como el p-valor es menor que el nivel de significación, se rechaza la hipótesis nula y se concluye que hay diferencias significativas entre el consumo de alcohol de chicos y chicas.

Ejercicio 1.9. En un estudio de genómica nutricional, se evalúa cómo afecta la suplementación con diferentes tipos de macronutrientes a la expresión de un gen relacionado con el metabolismo energético. Los investigadores dividen a los participantes en tres grupos, cada uno recibiendo una dieta suplementada (1 carbohidratos, 2 proteínas y 3 grasas).Se mide el nivel de expresión génica (ΔCt, relativo a un gen de referencia) en 25 individuos por grupo. Los resultados se recogen en el fichero expresion-genica-suplementacion.csv. Realizar el contraste de hipótesis adecuado para comparar ver si hay diferencias estadísticamente significativas entre las expresiones génicas de los tres grupos.

Solución

Número de poblaciones: 3 poblaciones independientes (población 1: Carbohidratos, población 2: Proteínas, población 3: Grasas).
Variable dependiente: Expresión génica (cuantitativa).
Tamaños muestrales: $n_1 = 75$, $n_2 = 75$ y $n_3 = 75$.
Contraste de normalidad de Kolmogorov-Smirnov: p-valor primera muestra: 0.7571, p-valor segunda muestra: 0.6459, p-valor tercera muestra: 0.9643. Se acepta la normalidad en los tres grupos.
Contraste de homocedasticidad de Levene: p-valor: 0.2126. Se acepta la hipótesis de igualdad de varianzas.
Contraste de ANOVA: $H_0 : \mu_1 = \mu_2 = \mu_3$, H_1 : Al menos dos medias son diferentes.
p-valor: $p = 0.000283$.
Como el p-valor es menor que el nivel de significación, se rechaza la hipótesis nula y se concluye que hay diferencias significativas entre las medias de la expresión génica de los grupos de suplementación.
Contraste post hoc de Tukey:
- Carbohidratos vs Proteínas: $p = 0.0016$. Hay diferencias significativas.
- Carbohidratos vs Grasas: $p = 0.0009$. Hay diferencias significativas.
- Proteínas vs Grasas: $p = 0.9809$. No hay diferencias significativas.

Ejercicio 1.10. En un estudio diseñado para analizar la influencia de un tipo de dieta y de un fármaco en el peso corporal perdido, se ha anotado el número de Kg perdidos en un grupo de personas al cabo de 3 meses de dieta y de tomar el fármaco en el fichero interaccion-dieta-farmaco.csv. Realizar el contraste de hipótesis adecuado para ver si el peso perdido depende de la dieta, del fármaco y de la interacción entre ambos.

Solución

Número de poblaciones: 4 poblaciones independientes (población 1: Dieta No + Fármaco No, población 2: Dieta No + Fármaco Si, población 3: Dieta Si + Fármaco No, población 4: Dieta Si + Fármaco Si).

Variable dependiente: Peso perdido (cuantitativa).
Tamaños muestrales: $n_1 = 5$, $n_2 = 6$, $n_3 = 5$ y $n_4 = 6$.
Contraste de normalidad de Shapiro-Wilk: p-valor primera muestra: 0.4677, p-valor segunda muestra: 0.9672, p-valor tercera muestra: 0.741, p-valor cuarta muestra: 0.719. Se acepta la normalidad en los cuatro grupos.
Contraste de homocedasticidad de Levene: p-valor: 0.845. Se acepta la hipótesis de igualdad de varianzas.
Contraste de ANOVA de dos factores con interacción: H_0 : No hay interacción entre la dieta y el fármaco, H_1 : Hay interacción entre la dieta y el fármaco.

- p-valor dieta: $p = 0.0000$.
- p-valor fármaco: $p = 0.0000$.
- p-valor interacción: $p = 0.0012$.

Como el p-valor es menor que el nivel de significación en los tres casos, se rechaza la hipótesis nula y se concluye que la pérdida de peso depende de la dieta, del fármaco y de la interacción entre ambos.