

Estadística (Q)

Ejercicio para las clases prácticas de la cuarta semana (D), intervalos de confianza y tests de hipótesis

1. Se hacen análisis de sangre a 25 personas elegidas al azar de la población obteniendo los siguientes índices de colesterol:

1.53 1.65 1.72 1.83 1.62 1.75 1.72 1.68 1.65 1.61
1.70 1.60 1.73 1.61 1.52 1.81 1.72 1.50 1.51 1.65
1.58 1.82 1.65 1.72 1.65

- (a) Calcular estimadores insesgados para la esperanza y la varianza poblacional del índice de colesterol.
- (b) Supongamos ahora que la distribución del índice de colesterol de la población es $N(\mu, \sigma^2)$
- Si $\sigma^2 = 0.01$. Hallar un intervalo de confianza de nivel 0.95 para μ . Calcular la longitud del intervalo obtenido. ¿A cuántas personas debería realizarse el estudio si se quiere que la longitud sea menor que 0.05?
 - Si σ^2 es desconocido, hallar un intervalo de confianza para μ de nivel 0.95.
 - Si σ^2 es desconocido, hallar un intervalo de confianza para σ^2 de nivel 0.95.
2. Se quiere aumentar la producción anual de almendras en el noroeste de la provincia de Mendoza, para lo cual se desarrolla un nuevo sistema de riego. Se observó el rendimiento de 140 parcelas elegidas al azar, plantadas con almendros, provistas del nuevo sistema de riego, obteniéndose una media muestral de 525 kilogramos por hectárea y un desvío muestral de 18 kilogramos por hectárea.
- (a) En base a esta muestra, hallar un intervalo de confianza de nivel 0.95 para μ , el rendimiento medio de los campos (del noroeste de Mendoza) con el nuevo sistema de riego. El intervalo hallado, ¿es exacto o asintótico? Justifique su respuesta.
- (b) El rendimiento medio de los campos con el sistema de riego tradicional es de 512 kilogramos por hectárea. ¿Hay razones para sospechar que el nuevo sistema de riego produce un rendimiento medio mayor?
- (c) Decidir si son verdaderas o falsas las siguientes afirmaciones
- Aproximadamente el 95% de las parcelas del noroeste de Mendoza tiene un rendimiento que cae en el intervalo hallado en a), en base a la muestra observada.
 - La probabilidad de que el intervalo anterior contenga a μ es aproximadamente 0.95.
 - Si se extraen muchas muestras de tamaño 140 de manera independiente y para cada una se calcula el intervalo de confianza como en a), aproximadamente el 95% de estos intervalos contendrá a μ .
3. Una muestra de 1000 votantes de determinado país es encuestada respecto a cierta propuesta política. Como resultado se obtiene que 200 están de acuerdo con la propuesta, 600 se oponen y 200 están indecisos.
- (a) Hallar un intervalo de confianza de nivel 90% para la proporción de votantes del país que se oponen a la propuesta. ¿Es exacto o asintótico? Justifique su respuesta.
- (b) ¿Cuántos votantes deberían encuestarse si se quiere que la longitud del intervalo obtenido sea menor o igual a 0.02?
4. El peso medio de calcio en un cemento estándar es de $94g/kg$. Se tomaron 16 muestras de cemento contaminado con plomo obteniéndose, en las 16 determinaciones de calcio, un peso promedio de $87g/kg$. Suponiendo que las mediciones de calcio siguen una distribución normal con desvío estándar $\sigma = 13g/kg$. Se quiere saber si la presencia de plomo afecta el peso medio de calcio en el cemento.
- (a) ¿Qué hipótesis se deben testear para responder a esta pregunta?
- (b) Si se tuviera información adicional respecto del efecto que la presencia del plomo en el cemento tiene sobre el contenido de calcio, en el sentido que este último sólo puede disminuir si hay plomo en el cemento, ¿cuál sería el test correcto para proponer para responder a la pregunta del enunciado, a nivel 5%? ¿Cuánto vale el p-valor en este caso?
- (c) Repita lo realizado en (b) a nivel 1% y 10%.

- (d) Hallar la potencia del test hallado en (b) si el verdadero peso medio del calcio del cemento contaminado es de $90g/kg$. Hallar la función de potencia del test.
- (e) Si se quiere que la potencia del test sea de 0.90 cuando el peso medio del calcio del cemento contaminado es de $90g/kg$, hallar el tamaño de muestra necesario para lograrlo.
- (f) Responder a la pregunta del ítem (a) con un test de hipótesis de nivel del 1%, definiendo claramente las variables aleatorias, los parámetros de interés y las hipótesis en cuestión. Escriba su conclusión en los términos del problema. Repita lo realizado a nivel 5%.
- (g) Calcular el p-valor para el test del ítem anterior.
- (h) Hallar el intervalo de confianza para el peso medio del calcio del cemento contaminado de nivel 99%. ¿Qué relación guarda con lo realizado en (f)? Si quisiéramos testear si el peso medio del calcio del cemento contaminado es de $92g/kg$ a nivel 1%, ¿podríamos sacar una conclusión sin hacer ninguna cuenta más? ¿Qué relación tiene este intervalo con lo realizado en (c)?
5. Una asociación de consumidores, preocupada por la cantidad de grasas contenida en una marca de hamburguesas, envía a un laboratorio independiente una muestra aleatoria de 12 hamburguesas para su análisis. El porcentaje de grasa en cada una de las hamburguesas de la muestra es:

21 18 19 16 18 24 22 19 24 14 18 15

El fabricante afirma que el contenido medio de grasa de este tipo de hamburguesas es menor al 18%. Basándose en la salida de R que figura más abajo, resuelva los siguientes ítems.

- (a) La asociación de consumidores quiere saber si tiene motivos para decir que la afirmación del fabricante es falsa. Asumiendo que el contenido de grasa de cada hamburguesa de esta marca es una v.a con distribución normal y varianza conocida $\sigma^2 = 9$, proponer un test para resolver este problema. Escribir las hipótesis a testear, definir las variables aleatorias y los parámetros involucrados en el test, escribir el estadístico del test y su distribución bajo la hipótesis nula y dar la región de rechazo. ¿Qué le informaría al representante de la asociación de consumidores como conclusión del test?
- (b) Dar un intervalo de confianza de nivel 0.95 para el verdadero contenido medio de grasa de este tipo de hamburguesas.
- (c) Calcular el p-valor para el test del ítem (a). ¿Rechazaría la hipótesis nula a nivel 0.08?
- (d) Si el verdadero contenido de grasa de las hamburguesas fuera del 20% ¿cuál sería la potencia de este test? Si se quiere que esta potencia sea de al menos 0.85, ¿cuántas hamburguesas habría que tomar en la muestra?

```
hamburguesas<-scan()
21 18 19 16 18 24 22 19 24 14 18 15
c(mean(hamburguesas),var(hamburguesas),sd(hamburguesas))
[1] 19.000000 10.545455 3.247377
```