



Facultad de Ingenierías
Inferencia Estadística
Maestría en Inteligencia Artificial y Ciencia de Datos
Prof. Cristian E Garcia.



Desafío 2

Condiciones:

- Subir la tarea en formato pdf en la plataforma UAO-Virtual.
- Es necesario incluir el código de R en formato R. Mostrar los resultados a partir de tablas, gráficos o indicadores que les permita dar respuesta a los planteamientos.
- Deben interpretar los resultados obtenidos en cada situación de acuerdo al contexto.
- Realizar la actividad en grupos máximo de 4 personas.

Situación 1

Un experimento utilizó una muestra de estudiantes universitarios para investigar si el uso del teléfono celular afecta los tiempos de reacción de los conductores. En una máquina que simulaba situaciones de conducción, se encendía de manera irregular una luz roja o verde. Se les indicó a los participantes que presionaran un botón de freno tan pronto como detectaran una luz roja. Bajo la condición de uso del teléfono celular, cada estudiante tenía una conversación con alguien en otra habitación. En la condición de control, los mismos estudiantes escuchaban una transmisión de radio. El archivo de datos **CellPhone** registra los tiempos de respuesta promedio de los estudiantes (en milisegundos) en varias pruebas para cada condición: $\{y_{i1}\}$ para la condición del teléfono celular y $\{y_{i2}\}$ para la condición de control.

[(a)] Las comparaciones de medias o proporciones suponen muestras independientes para los dos grupos. Explica por qué las muestras para estas dos condiciones son **dependientes** en lugar de independientes.

[(b)] Para comparar μ_1 y μ_2 , puedes usar $\{d_i = y_{i1} - y_{i2}, i = 1, \dots, n\}$, aquí con $n = 8$. Especifica el parámetro μ_d y la hipótesis nula H_0 para hacer esto, y explica por qué $\mu_d = \mu_1 - \mu_2$.

[(c)] Indica las suposiciones y la estadística de prueba, y explica por qué sigue una distribución t con $df = n - 1$. Reporta el valor P con una hipótesis alternativa bilateral H_a , e interpreta el resultado. También es posible realizar análisis de pares relacionados

usando intervalos de confianza, al comparar pesos de niñas con anorexia antes y después de un tratamiento analizando la diferencia media de pesos).

Situación 2

Genere 5000 muestras aleatorias de tamaño $n = 10$ de una población normal con media $\mu = 5$ y Varianza 1. Con cada una de ellas construya un intervalo de confianza del 95 % para la media. Cuente que porcentaje de los 5000 intervalos atrapan la verdadera media. Comente el resultado del porcentaje de cobertura y la amplitud promedio.

- (a) Repita el ejercicio para los tamaños de muestra (30, 50, 100). Represente gráficamente el porcentaje de cobertura observado y amplitud promedio e interprete los resultados. Nota: deben comparar el rendimiento de cada métodos utilizando la amplitud promedio y el porcentaje de cobertura y para los métodos bootstrap utilizar $B = 1000$

Situación 3

Una firma decide estudiar una muestra aleatoria de 20 proyectos que envió para ser evaluados, tanto a consultores externos, como a su propio departamento de proyectos. Las variables medidas fueron

X : n^o de días que demoro la evaluación

Y : n^o de variables consideradas en la evaluación.

Z : Consultor al que se le envió el proyecto

$$Z = \begin{cases} -1 & ; \text{Depto. de Evaluación} \\ 0 & ; \text{Robani Consultores} \\ 1 & ; \text{Tanaka Ltda.} \end{cases}$$

W : Costo de la evaluación (en U.F.)

Los resultados de este muestreo son:

Nº	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
X	4	2	8	10	1	3	8	3	2	2	4	4	5	6	7	2	1	3	4	9
Y	3	1	6	8	3	2	6	2	1	1	4	4	4	7	10	3	2	4	5	10
Z	-1	-1	0	0	0	0	1	0	0	1	-1	-1	0	1	1	-1	-1	0	1	-1
w	40	30.5	80.3	68.5	24.7	40.5	90.5	38.5	50.4	50.2	60.1	60.8	70.9	80	90	30	27	40	50	40

- (a) Estime con un 90 % de confianza el costo medio de los proyectos.

(b) Estime con un 90 % de confianza la proporción de proyectos cuyo costo fue inferior a 50 U.F. dado que no involucraron más de 6 variables y que fueron resueltos en un tiempo superior a 2 días.

(c) El Depto. de control afirma que el costo medio de enviar los proyectos a asesores externos es significativamente mayor que el de evaluarlos allí mismo. ¿Qué concluye usando $\alpha = 0,05$? **Nota:** Compruebe primero si las varianzas son iguales o diferentes para poder decidir que test utilizar para la diferencia de medias. **Hint:** Use la distribución F .

(d) Tanaka Ltda. Afirma que la proporción de proyectos que ellos evalúan, que toman más tiempo de más de 4 días, no es superior a la proporción de proyectos que evalúa Robani Consultores, que toman un tiempo de más 4 días, no es superior a la proporción de proyectos que evalúa Robani Consultores, que toman un tiempo más de 4 días. Concluya si la afirmación de Tanaka Ltda. es correcta. (Use $\alpha = 0,01$).

Situación 4 “Investigar un poco”

La siguiente tabla contiene 40 recuentos anuales del número de reclutas y reproductores en una población de salmones. Las unidades están en miles de peces.

R	S	R	S	R	S	R	S
68	56	222	351	311	412	244	265
77	62	205	282	166	176	222	301
299	445	233	310	248	313	195	234
220	279	228	266	161	162	203	229
142	138	188	256	226	368	210	270
287	428	132	144	67	54	275	478
276	319	285	447	201	214	286	419
115	102	188	186	267	429	275	490
64	51	224	389	121	115	304	430
206	289	121	113	301	407	214	235

☼ **Reclutas (R):** peces que ingresan a la población capturable.

☼ **Reproductores (S):** peces que están poniendo huevos. Los reproductores mueren después de poner huevos.

El modelo clásico de Beverton-Holt para la relación entre reproductores y reclutas es:

$$R = \frac{1}{\beta_1 + \beta_2/S}, \quad \beta_1 \geq 0, \quad \beta_2 \geq 0,$$

donde R y S son los números de reclutas y reproductores, respectivamente. Este modelo puede ajustarse mediante regresión lineal con las variables transformadas $1/R$ y $1/S$.

Para mantener una pesca sostenible, la población total solo se estabilizará si $R = S$. La población total disminuirá si se producen menos reclutas de los reproductores que murieron generándolos. Si se producen demasiados reclutas, la población también disminuirá debido a la competencia por los recursos. Por lo tanto, hay un nivel intermedio de reclutas que se puede mantener indefinidamente en una población estable. Este nivel estable es el punto donde la línea de 45° intercepta la curva que relaciona R y S .

Instrucciones

- (a) Ajustar el modelo de Beverton-Holt y encontrar una estimación puntual para el nivel estable de la población donde $R = S$. Usar bootstrap para obtener un intervalo de confianza del 95% y un error estándar, utilizando dos métodos: remuestreo de los residuales y remuestreo de los casos. Representar histogramas para cada distribución bootstrap y comentar sobre las diferencias en los resultados.
- (b) Proporcionar una estimación corregida por sesgo y un error estándar correspondiente para el estimador corregido.
- (c) Usar bootstrap anidado con pivoteo para encontrar un intervalo de confianza del 95% para el punto de estabilización.