



Prueba escrita N° 1

Forma 10

Instrucciones

1. Mientras desarrollen esta prueba, tienen prohibido solicitar ayuda, por cualquier medio y cualquier persona que no pertenezca al equipo informado y autorizado por el profesor. Pueden revisar sus propios apuntes, las lecturas realizadas, el help de R y los scripts publicados en la página del curso en uvirtual.usach.cl. Casos sospechosos de actos deshonestos serán calificados con nota mínima y podrían dar inicio a un proceso sumarial.
2. Su respuesta debe ser un script R, ordenado y muy documentado, que incluya las soluciones a las preguntas planteadas. El nombre del script debe ser Forma-10, que corresponde al número de la forma de su enunciado. No se identifiquen de ninguna otra manera en el script.
3. Uno de los estudiantes del equipo debe enviar la respuesta al correo electrónico del profesor (joseluis.jarra@usach.cl) alrededor de las 20:30. Respuestas recibidas después de las 20:40 van a requerir justificación por el atraso y debe adjuntarse evidencia de que tuvieron problemas de conexión o que no pudieron comenzar a tiempo por un imprevisto.
4. También se recomienda que mantengan actualizada una copia de sus respuestas en el repositorio GITHUB de cada pareja.
5. La prueba tiene 4 preguntas, dos de ellas prácticas, con una puntuación final de 51 puntos. Obtiene nota 4,0 con 50% de la puntuación.
6. Si tienen una pregunta sobre el enunciado, envíe un correo al profesor, quien estará atento a sus mensajes durante el desarrollo de la prueba. Sin embargo, recuerde que se irán respondiendo las preguntas en el orden que llegan, por lo que no esperen respuesta instantánea. Como alternativa, ustedes pueden hacer cualquier suposición que necesiten para responder una pregunta; mientras sea razonable, será considerada en la corrección de la pregunta. Debe estar escrita y justificada explícitamente en su script.
7. Tiene 90 minutos para responder.
8. Se adjunta un archivo CSV con los datos que van a utilizar, que corresponden a un estudio de pacientes con problemas cardíacos. La siguiente tabla resume las columnas presentes:

| Concepto (nombre columna) | Descripción | Medición | Rango de valores |
|---|---|--------------------|--------------------|
| Edad (age) | Edad del paciente | Años | [40,..., 80] |
| Anemia (anaemia) | Si el paciente presenta una reducción de glóbulos rojos o hemoglobina | Booleano | 0,1 |
| Creatina quinasa (creatinine_phosphokinase) | Nivel de la enzima CK en la sangre | mcg / L | [2,..., 1031] |
| Diabetes (diabetes) | Si el paciente tiene diabetes | Booleano | 0,1 |
| Fracción de eyección (ejection_fraction) | Porcentaje de sangre que sale del corazón en cada contracción | Percentage | [19,..., 55] |
| 2da medición de la fracción de eyección (ejection_fraction_2) | Fracción de eyección al tercer día de tratamiento | Percentage | [21,..., 62] |
| Hipertensión (high_blood_pressure) | Si el paciente tiene hipertensión | Booleano | 0,1 |
| Plaquetas (platelets) | Plaquetas en la sangre | Kiloplaquetas / mL | [115.6,..., 381.9] |
| Creatinina sérica (serum_creatinine) | Nivel de creatinina en la sangre | mg / dL | [0.42,..., 2.52] |
| 2da medición Creatinina (serum_creatinine_2) | Nivel de creatinina al tercer día de tratamiento | mg / dL | [0.38,..., 2.58] |
| Sodio sérico (serum_sodium) | Nivel de sodio en la sangre | mEq / L | [128,..., 142] |
| Sexo (sex) | Si el paciente es mujer o hombre | Binario | 0,1 |
| Fumador (smoking) | Si el paciente fuma | Booleano | 0,1 |
| Tiempo (time) | Periodo de control | Días | [5,...,295] |
| Muerte (DEATH_EVENT) | Si el paciente murió durante el periodo de control | Booleano | 0,1 |

mcg / L: microgramos por litro; mL: mililitro; mg / dL: miligramos por decilitro; mEq / L: miliequivalentes por litro.



Pregunta 1

Lea el siguiente enunciado:

Un estudio sobre la mortalidad de pacientes con fallas coronarias, la principal causa de muerte en el mundo, determinó que no había una diferencia significativa en los niveles de creatina quinasa entre los pacientes que presentaron anemia, una carencia de suficientes glóbulos rojos sanos para transportar oxígeno por el cuerpo, durante el periodo de control y los pacientes que no tuvieron esta afección.

(17 puntos) Usando como semilla el valor 127, tome una muestra aleatoria de 25 pacientes que no presentaron anemia durante el periodo de control y 25 pacientes que sí la presentaron y, explicando y justificando paso a paso el procedimiento seguido (hipótesis contrastadas, prueba estadística usada, verificación de condiciones, etc.), realice un análisis inferencial, con un nivel de significación de 0,05, que determine si la afirmación hecha por el estudio es correcta de acuerdo a los datos obtenidos. Además, determine el poder estadístico de la prueba realizada. Además, determine el poder estadístico de la prueba realizada.

Pregunta 2

Lea el siguiente enunciado:

Un estudio sobre la mortalidad de pacientes con fallas coronarias, la principal causa de muerte en el mundo, describió que dos tercios de las personas reclutadas presentaron presión arterial anormalmente alta (hipertensión).

(17 puntos) Usando como semilla el valor 439, tome una muestra aleatoria de 60 pacientes y, explicando y justificando paso a paso el procedimiento seguido (hipótesis contrastadas, prueba estadística usada, verificación de condiciones, etc.), realice un análisis inferencial, con un nivel de significación de 0,01, que determine si los datos obtenidos respaldan la descripción hecha en el estudio. También determine el poder estadístico de la prueba realizada.

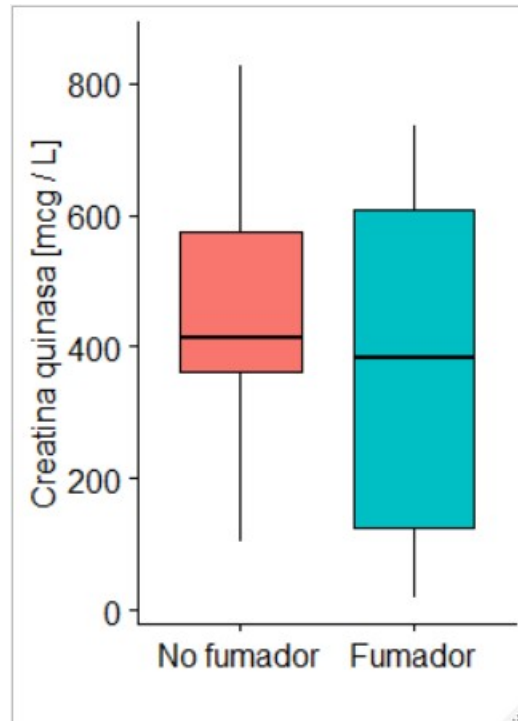
Pregunta 3

(9 puntos) Proponga un ejemplo novedoso (no mencionado en clase ni que aparezca en las lecturas dadas) en donde un estudio o experimento, relacionado con el sentimiento ciudadano en las comunas de Chile que completan cuatro meses de cuarentena, necesite utilizar una Prueba de McNemar. Indique cuáles serían las variables/niveles involucrados en su ejemplo.



Pregunta 4

(8 puntos) Un estudio sobre la mortalidad de pacientes con fallas coronarias, la principal causa de muerte en el mundo, reportó el siguiente gráfico de cajas:



Indique tres características distintas y relevantes de las distribuciones de los datos usados en este ejemplo, indicando claramente cómo los observa en el gráfico.

Buena suerte.