



Prueba escrita Nº 2 Forma 06

INSTRUCCIONES

- Mientras desarrollen esta prueba, tienen prohibido solicitar ayuda, por cualquier medio y a cualquier persona que no pertenezca a la dupla informada y autorizada por su profesor(a). 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.
- 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-XX.R, donde XX corresponde al número de la forma de su enunciado. **No se identifiquen de ninguna otra manera en el script.**
- El enunciado de la prueba estará disponible para ser descargado desde las 19:00 y hasta las 22:00 del viernes 21 de enero en la página del curso en Uvirtual. Su respuesta debe ser subida por este mismo medio antes de las 22:00, cuando se cierra automáticamente el sistema.
- La prueba está diseñada para ser respondida en 90 minutos (de manera presencial en una sala de clases). Entendiendo que la modalidad remota en que se está realizando la evaluación puede significar que algunos de los estudiantes no consigan replicar las condiciones de la modalidad presencial, el tiempo para responder la prueba se extiende a **120 minutos**.
- Por otro lado, se pide que hagan frecuentes actualizaciones del script respuesta (al menos 10) en el repositorio GITHUB de cada equipo. Por este motivo, se otorgan 10 minutos adicionales para responder.
- De esta forma, en caso de dificultades al momento de la entrega, su profesor(a) descargará desde GITHUB la última versión del script respuesta actualizado antes de cumplidos 130 minutos después de que el equipo haya descargado el enunciado o la última versión subida a Uvirtual antes de las 22:00 horas, lo que ocurra primero.
- Si un equipo tiene dificultades durante el desarrollo o la entrega de su respuesta, se sugiere escribir al correo electrónico de su profesor(a) (jose-luis.jara@usach.cl – jacqueline.kohler@usach.cl) y justificar un eventual atraso adjuntando evidencia del imprevisto enfrentado.
- Si tiene una pregunta sobre el enunciado, envíe un correo su profesor(a), 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 una 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.
- La prueba tiene 3 preguntas, dos de ellas prácticas, con una puntuación final de 56 puntos. Obtiene nota 4,0 con 50% de la puntuación.

CRITERIOS DE EVALUACIÓN

Prob.	Categoría	Nivel de logro	P. ideal
1	Datos	Obtienen los datos necesarios para realizar correctamente una prueba de hipótesis pertinente para responder la pregunta planteada	2
	Formulación de hipótesis	Formulan, con claridad y explícitamente, hipótesis nula y alternativa correctas y completas a la prueba de hipótesis pertinente para responder la pregunta	3
	Condiciones	Verifican el cumplimiento de todas las condiciones requeridas por la prueba de hipótesis pertinente para responder la pregunta, usando para ello gráficos o pruebas estadísticas adecuadas	3
	Prueba omnibus	Realizan de forma correcta y completa una prueba omnibus pertinente para responder la pregunta, usando para ello los datos adecuados y los parámetros correspondientes	3
	Prueba post hoc	Realizan de forma correcta y completa una prueba post-hoc pertinente para responder la pregunta, justificando convincentemente su selección, usando para ello los datos adecuados y los parámetros correspondientes	3
	Conclusión	Entregan una conclusión correcta y completa a la pregunta planteada, basándose en el resultado de la prueba realizada (omnibus o post-hoc, según corresponda) y el contexto del problema	3
	Código fuente	Escriben código R -ordenado, bien indentado, sin sentencias espurias y bien comentado- que realiza de forma completa y correcta la prueba seleccionada con los datos adecuados en cada caso	3
	Ortografía y redacción	Escriben con buena ortografía y redacción (<3 errores), usando vocabulario propio de la disciplina y el contexto del problema	3
2	Datos	Obtienen una muestra de datos para poder crear y evaluar modelos de regresión, cumpliendo las restricciones indicadas en el enunciado (semilla, separación entre conjuntos de entrenamiento y prueba, etc.)	3
	Selección de variables	Seleccionan un conjunto de variables relevantes para predecir la variable de salida indicada, utilizando correctamente gráficos y/o utilidades en paquetes de R para explorarlas, desde el conjunto de ocho variables elegidas aleatoriamente como posibles predictores y respetando las otras restricciones indicadas en el enunciado	4
	Modelo	Construyen correctamente un modelo de regresión pertinente para predecir la variable de salida indicada, considerando las variables seleccionadas anteriormente	2
	Confiabilidad del modelo	Escriben comentarios y código en R correcto que verifica las condiciones que garantizan que el modelo de regresión obtenido tiene un buen nivel de ajuste y es generalizable, interpretando explícita y correctamente los resultados obtenidos en cada paso y tomando acciones correctivas apropiadas de ser necesarias o comentando los riesgos asociados	4
	Calidad predictiva del modelo	Escriben código R correcto que evalúa la calidad predictiva del modelo de regresión obtenido, en datos no utilizados para su construcción	2
	Conclusiones	Entrega conclusiones correctas y completas, basadas en las evaluaciones realizadas y el proceso de búsqueda de predictores seguido, que muestran que se ha conseguido un modelo de acuerdo a lo solicitado, escribiendo con buena redacción y ortografía	3
	Código fuente	El script está completo, ordenado y bien indentado, se comenta paso a paso el procedimiento implementado, con buena redacción (basta una lectura para entender) y con buena ortografía (no más de 3 errores), logrando un programa que es fácil de seguir y que no requiere cambios para que funcione	3
	Ortografía y redacción	Escriben con buena ortografía y redacción (<3 errores), usando vocabulario propio de la disciplina y el contexto del problema	3
3	Estudio o experimento	Describe, con claridad y buena ortografía, un estudio o experimento interesante y novedoso (no visto anteriormente en lecturas dadas o en clases), que se da naturalmente en el contexto solicitado, y argumenta sólidamente que, a pesar que existen problemas con la escala de la variable dependiente para ser analizado con una prueba paramétrica, puede ser analizado apropiadamente con la prueba estadística indicada en el enunciado, explicando cómo se organizarían los datos y formulando correctamente las hipótesis nula y alternativa que se necesitaría contrastar	9
Puntaje total			56
Nota			7,0

STAR WARS: LA NACIENTE REBELIÓN EN PROBLEMAS

Hace mucho tiempo, en una galaxia muy, muy lejana...

Tras una serie de exitosos golpes al imperio, la rebelión iniciada secretamente por Obi-Wan Kenobi y el antiguo senador Organa en Tatooine ha comenzado a expandirse y nuevas células han comenzado a surgir en cada rincón de la galaxia.

Decidido a aplastar esta insurgencia, el emperador Palpatine ha enviado a su mano derecha, el Lord Sith Darth Vader, a supervisar el entrenamiento de nuevos soldados imperiales (stormtroopers), quienes han incorporado a reclutas voluntarios y jóvenes secuestrados de sus planetas de origen.

Lord Vader, tras secuestrar a su familia, ha forzado a un joven analista de datos a evaluar el resultado del entrenamiento de nuevos soldados en base a los siguientes datos...

Columna	Descripción	Medición	Rango de valores
id	Código único de identificación del soldado	Numérico	
division	Especialidad del stormtrooper	Categorico	Cavetrooper, Snowtrooper, Lavatrooper, Shoretrooper, Spacetrooper, Sandtrooper, Flametrooper, Recontrooper
eval_instructor	Evaluación general del stormtrooper realizada por su instructor	numérico	[550, ..., 750]
Eval_capitan	Evaluación general del stormtrooper realizada por su capitán	numérico	[550, ..., 750]
eval_comandante	Evaluación general del stormtrooper realizada por su comandante	numérico	[550, ..., 750]
eval_general	Evaluación general del stormtrooper realizada por su general	numérico	[550, ..., 750]
velocidad	Velocidad a la que puede correr el soldado con respecto al óptimo.	porcentaje	
fuerza	Fuerza del soldado con respecto al óptimo.	porcentaje	
resistencia	Resistencia del soldado con respecto al óptimo.	porcentaje	
agilidad	Agilidad del soldado con respecto al óptimo.	porcentaje	
es_clon	El soldado es o no un clon	Categorico	S: sí, N: no
estatura	Estatura del soldado	cm	
peso	Peso del soldado	Kg	
imc	Índice de masa corporal del soldado	Kg/m ²	
armadura	Armadura del soldado con respecto al óptimo.	porcentaje	
Precision	Precisión del soldado con respecto al óptimo.	porcentaje	
eficiencia	Eficiencia del soldado con respecto al óptimo.	porcentaje	

Pregunta 1

(23 puntos) Lord Vader desea saber si los niveles de exigencia con que los distintos oficiales evaluadores (instructor, capitán, comandante y general) califican a los flametroopers son similares, por lo que le ha solicitado estudiar si existen diferencias significativas en el promedio de la evaluación realizada por cada uno de los oficiales. El Lord Sith ha sido muy claro al solicitar un reporte de aquellos oficiales cuyas evaluaciones presenten diferencias.

Pregunta 2

(24 puntos) A fin de determinar si es necesario establecer programas de entrenamiento diferenciados para clones y reclutas, Lord Vader quiere saber si es posible distinguir entre ambas clases de soldados con los datos actuales. Para ello, ha solicitado evaluar un modelo clasificador que contemple entre 2 y 5 variables predictoras. Considere que, para ser aceptable, el modelo:

- Debe lograr una exactitud (*accuracy*) de al menos 0,8 en datos de prueba
- No puede considerar casos con demasiada influencia (considerando la distancia de Cook)
- No debe presentar autocorrelación (usando la prueba de Durbin-Watson para un retardo y un nivel de significación $\alpha = .01$)
- No debe presentar multicolinealidad severa (considerando el factor de inflación de la varianza, con un VIF promedio inferior a 1,03).

Considere la semilla 1030 para obtener una muestra de 400 datos, 80% de los cuales serán empleados para ajustar el modelo y el 20% restante, para evaluarlo.

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 sentir de los santiaguinos ante el aumento de la violencia de la delincuencia, necesite utilizar una prueba de suma de rangos de Wilcoxon (también llamada prueba de Mann–Whitney–Wilcoxon o prueba U de Mann-Whitney) debido a problemas con la escala de la variable dependiente en estudio. Indique cuáles serían las variables involucradas en su ejemplo (con sus respectivos niveles) y las hipótesis nula y alternativa a contrastar.

¡Que la fuerza te acompañe!