

Prueba Escrita N° 3

Forma 06

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-06, que corresponde al número de la forma de su enunciado. No se identifiquen de ninguna otra manera en el script.
3. El enunciado de las pruebas estará disponible para ser descargadas desde las 19:00 y hasta las 22:00 del lunes 16 de agosto en la página del curso en Uvirtual.
4. 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.
5. 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.
6. De esta forma, el/la profesor/a descargará desde GITHUB la versión del script respuesta actualizado 130 minutos después de que el equipo haya descargado el enunciado o la última versión subida antes de las 22:00 horas, lo que ocurra primero.
7. Si un equipo tiene dificultades durante el desarrollo o la entrega de su respuesta, se sugiere escribir al correo electrónico del profesor (joseluis.jara@usach.cl) y justificar un eventual atraso adjuntando evidencia del imprevisto enfrentado.
8. Si tienen una pregunta sobre el enunciado, envíe un correo al profesor de su sección, 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.
9. La prueba tiene 3 preguntas, una de ellas práctica, con una puntuación final de 36 puntos. Obtiene nota 4,0 con 50% de la puntuación.
10. Adjunto a este documento encontrará un conjunto de datos con algunos de los datos relativos a estudiantes de la asignatura de Portugués de dos colegios de Brasil (basado en <https://archive.ics.uci.edu/ml/datasets/Student+Performance>). Dicho conjunto de datos cuenta con las siguientes columnas:
 - a. colegio: 'GP'=Gabriel Pereira, 'MS'=Mousinho da Silveira.
 - b. sexo: 'F'=Femenino, 'M'=Masculino.
 - c. edad [15-22].

- d. residencia: 'rural', 'urbana'.
- e. tam.fam: cantidad de integrantes del grupo familiar. ' ≤ 3 '=3 o menos integrantes, ' >3 '=más de 3 integrantes.
- f. estado.padres: estado de cohabitación de los padres. 'separados'=viven separados, 'juntos'=viven juntos
- g. edu.madre: educación de la madre. 'ninguna', 'primaria' (hasta 4to grado), 'basica', 'media', 'superior'.
- h. edu.padre: educación del padre. 'ninguna', 'primaria' (hasta 4to grado), 'basica', 'media', 'superior'.
- i. trab.madre: trabajo de la madre. 'docencia', 'salud', 'servicios', 'ninguno', 'otro'.
- j. trab.padre: trabajo del padre. 'docencia', 'salud', 'servicios', 'ninguno', 'otro'.
- k. motivo: razón para escoger el colegio. 'distancia' (cerca de la casa), 'reputacion', 'curriculo', 'otro'.
- l. apoderado: 'madre', 'padre', 'otro'.
- m. traslado: tiempo de viaje de la casa al colegio. ' <15 '=menos de 15 min., '15-30'=entre 15 y 30 min., '31-60'=30 min. a 1 h, ' >60 '=más de 1 h.
- n. estudio: tiempo semanal de estudio. ' <2 '=menos de 2 h, '2-5'= 2 a 5 h, '5-10'=5 a 10 h, ' >10 '= más de 10 h.
- o. reprobaciones: número de reprobaciones en años anteriores [1-4].
- p. apoyo.alumno: apoyo académico adicional para el alumno. no, si.
- q. apoyo.familia: apoyo académico para la familia. no, si.
- r. particular: clases particulares de la asignatura. no, si.
- s. actividades: participa en actividades extra-programáticas. no, si.
- t. preescolar: asistió a educación preescolar. no, si.
- u. superior: aspira a cursar estudios superiores. no, si.
- v. internet: cuenta con acceso a Internet en casa. no, si.
- w. relacion: está en una relación romántica. no, si.
- x. relacion.familiar: calidad de la relación familiar. 'muy_mala', 'mala', 'buena', 'muy_buena', 'excelente'.
- y. tiempo.libre: tiempo libre después del colegio. 'muy_poco', 'poco', 'suficiente', 'bastante', 'mucho'.
- z. salidas: sale con amigos. 'muy_poco', 'poco', 'suficiente', 'bastante', 'mucho'.
- aa. alcohol.s: consumo de alcohol durante la semana. 'nada', 'muy_poco', 'poco', 'bastante', 'mucho'.
- bb. alcohol.f: consumo de alcohol durante el fin de semana. 'nada', 'muy_poco', 'poco', 'bastante', 'mucho'.
- cc. salud: estado actual de salud 'muy mala', 'mala', 'buena', 'muy_buena', 'excelente'.
- dd. inasistencias [0-93].
- ee. final: situación final del alumno. 'R'=reprueba, 'A'=aprueba.

Pregunta 1

(21 puntos) Utilizando los datos provistos, construya un modelo de regresión logística para identificar la situación final de los estudiantes (respuesta), que tenga entre 3 y 8 variables predictoras y que consiga una exactitud (*accuracy*), en datos de prueba, superior a 0,70. Considere la semilla 287 para determinar los conjuntos de entrenamiento (80% de las observaciones) y prueba (20% de las observaciones). Verifique que el modelo no presenta multicolinealidad severa (usando el factor de inflación de la varianza).

Pregunta 2

(9 puntos) Proponga un ejemplo novedoso (no mencionado en clase ni que aparezca en las lecturas dadas) en donde dos variables relacionadas con las eliminatorias para el campeonato mundial de fútbol Catar 2022 presenten una correlación con un coeficiente de Pearson de aproximadamente -0,7. Dé valores posibles, que podrían ser reales, y escriba código R para mostrar en un gráfico cómo se ve esta relación.

Pregunta 3

(6 puntos) Proponga un ejemplo novedoso (no mencionado en clase ni que aparezca en las lecturas dadas) de un problema de regresión lineal simple que podría aparecer en relación con las calificaciones finales de estudiantes de primer año de Ingeniería durante la pandemia de Covid-19. Justifique que es plausible obtener los datos para resolver el problema que plantea.

¡Buena suerte!