

# EXAMEN FINAL

ⓘ Esta es una vista previa de la versión publicada del examen

Comenzado: 7 de feb en 10:26

## Instrucciones del examen

El siguiente examen consta de dos partes.

Parte I: 3 Preguntas en base a la teoría (40 puntos).

Parte II: Pregunta de desarrollo en R. (60 puntos). En esta parte deberá adjuntar: Archivo .R con el código de su análisis y respuestas.

Cuentan con 120 minutos. Distribuya bien su tiempo.

Es un examen a libro abierto, pueden utilizar el material de clase, páginas de internet, etc.

Es un examen individual, está prohibida la comunicación con sus compañeros o cualquier tercera persona durante el examen.

Deben mantener las cámaras encendidas durante la duración del examen.

Cualquier acto de deshonestidad será informado a las autoridades competentes.

**Política adicional por virtualidad: El docente en cualquier momento durante el examen puede solicitar aleatoriamente a uno o algunos estudiantes que se unan a una sala de Zoom privada para mostrar el avance de su examen y sustentar oralmente lo que están realizando. Así mismo, el docente puede solicitar sustanciación oral de lo realizado después del examen. Si el estudiante no puede sustentar lo que ha realizado, la nota de dicha pregunta será de cero.**

Al comenzar el examen, el estudiante acepta el siguiente compromiso:

### COMPROMISO DE HONOR

Reconozco que el presente examen está diseñado para ser resuelto de manera individual, y no se permite la ayuda de fuentes no autorizadas ni copiar.

Acepto adicionalmente que el docente puede solicitarme en cualquier momento que sustente oralmente cómo resolví las preguntas del examen. Acepto además que si no puedo sustentar lo que he realizado, la nota de dicha pregunta será de cero.

#### Pregunta 1

20 pts

En un estudio sobre conflictos laborales en industria y derechos laborales en 47 estados de Estados Unidos. Las variables recolectadas son las siguientes:

conflict: % de horas de trabajadores perdidas en paros.

unionization: % de la fuerza laboral en sindicatos.

union\_shop: Variable dicotómica. 1= El estado acepta los acuerdos de seguridad sindical entre empleadores y sindicatos, 0= El estado NO acepta los acuerdos de seguridad sindical entre empleadores y sindicatos.

agricultural: % de trabajadores trabajando en agricultura.

El resumen del modelo estimado se presenta a continuación:

```
##
## Call:
## lm(formula = conflict ~ unionization * union_shop + agricultural,
##     data = workers2)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.1718 -0.4246 -0.2534  0.4082  1.7489
##
## Coefficients:
##              Estimate Std. Error t value
## (Intercept)   -1.28692    0.66595  -1.932
## unionization    0.10928    0.02470   4.425
## union_shop     1.54886    0.75164   2.061
## agricultural    0.00833    0.01727   0.482
## unionization:union_shop -0.07393    0.02818  -2.624
##              Pr(>|t|)
## (Intercept)    0.0601 .
## unionization  6.73e-05 ***
## union_shop    0.0456 *
## agricultural  0.6321
## unionization:union_shop 0.0121 *
## ---
## Signif. codes:
## 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7093 on 42 degrees of freedom
## Multiple R-squared:  0.4124, Adjusted R-squared:  0.3564
## F-statistic: 7.369 on 4 and 42 DF,  p-value: 0.0001367
```

Suponga que un estado actualmente acepta los acuerdos de seguridad sindical entre empleadores y sindicatos y está evaluando NO aceptarlos más.

En base al modelo ¿En cuánto esperaría que aumente o disminuya el % de horas de trabajadores perdidas en paros **si ahora NO aceptara dichos acuerdos** manteniendo los mismos porcentajes de fuerza laboral en sindicatos y de trabajadores en agricultura que actualmente están en 25% y 9%, respectivamente?

Ingrese el valor estimado en el casillero redondeando a 3 decimales. Si es una disminución ingrese el valor con signo negativo.

**Nota:** Para sus cálculos, use los porcentajes proporcionados sin dividir para 100, por ejemplo, un 25% en la base de datos está registrado como 25.

## Pregunta 2

10 pts

En el modelo anterior, interprete el valor estimado del intercepto en el contexto del problema.

[Editor HTML](#)

**B** *I* U A ▾ A ▾  $\bar{I}_x$       $x^2$   $x_2$     
      $\sqrt{x}$    12pt ▾ Párrafo ▾ 

0 palabras

## Pregunta 3

10 pts

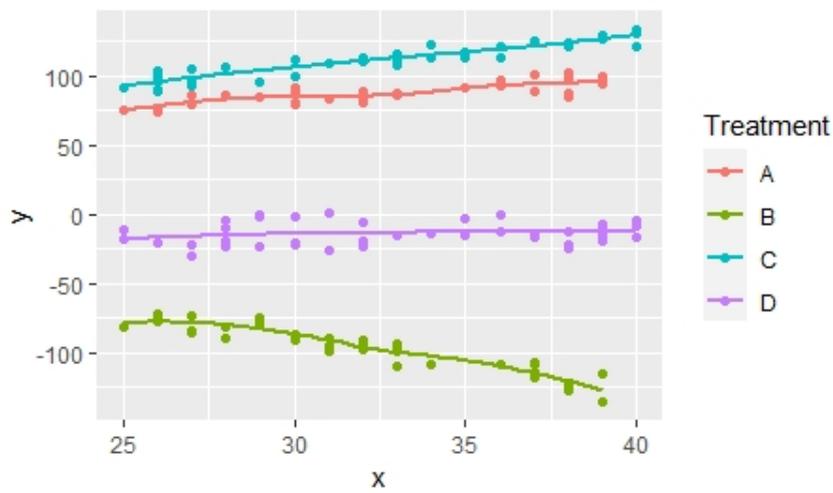
Suponga que se quiere estudiar si existe diferencia de los tratamientos en la variable de respuesta. Se cree además que se puede usar como covariable una variable continua  $x$  si es que ésta ayudara a reducir la variabilidad del error en el modelo.

Suponga que se tienen dos muestras distintas de poblaciones distintas y se quiere usar un modelo para cada muestra.

En el análisis exploratorio de los datos obtenidos se grafica la relación de la variable de respuesta y la covariable diferenciado por cada tratamiento.

Escoja para cada gráfico el tipo de modelo que se ajustaría mejor según el análisis exploratorio y el objetivo del estudio.

**A.**

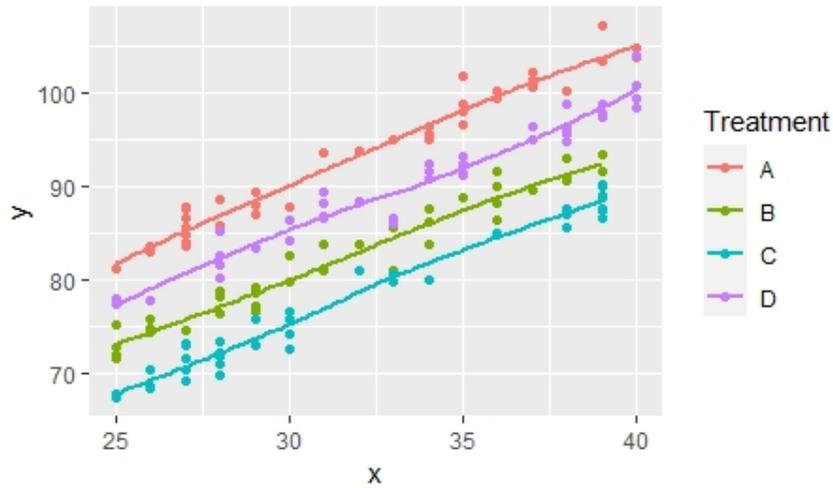


Para el gráfico A el mejor modelo sería

[ Seleccionar ]



**B.**



Para el gráfico B el mejor modelo sería

[ Seleccionar ]



Base de datos: [quality.csv](#)

([https://aulavirtual.espol.edu.ec/assessment\\_questions/6389471/files/4961163/download?verifier=h9G2u0M66DkSq4Ozfmccw891uhc4erEwOHmrx3P](https://aulavirtual.espol.edu.ec/assessment_questions/6389471/files/4961163/download?verifier=h9G2u0M66DkSq4Ozfmccw891uhc4erEwOHmrx3P)).

### PARTE A.

Se desea comparar la calidad del vino (quality: medido en un índice continuo del 1 al 10) dependiendo del tipo de vino (type= white, red).

1. Realice un diagrama de cajas comparando estas dos variables. ¿Hay sugerencia gráfica de que alguno de los dos tipos de vinos tenga una mejor calidad?. (5 pts)
2. Estime un modelo ANOVA. Asuma que los supuestos se cumplen y realice la prueba de hipótesis para responder a la siguiente pregunta: ¿La calidad del vino depende de si es blanco o rojo? Escriba el contraste de hipótesis, el estadístico de prueba y concluya en terminología del problema planteado. (10 pts)
3. Proporcione la ecuación estimada del modelo ANOVA  $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$  para el vino blanco. (10 pts)

### PARTE B

Ahora va a modelar la calidad del vino blanco en base a un grupo de variables físico-químicas por medio de un modelo de regresión lineal.

**Nota:** Solo para el examen, NO compruebe supuestos en ninguno de los pasos a realizar, asuma que los supuestos del modelo se cumplen.

4. Divida apropiadamente su base de datos en tres sets “training”, “validation” y “estimation” tomando en cuenta que tenga suficientes observaciones en cada set. Fije las semillas para la división aleatoria, al menos una semilla debe ser su número de matrícula. Recuerde que solo trabajará con el tipo de vino blanco. (5 pts)
5. En el set de “training” seleccione dos modelos por medio de dos búsquedas automáticas backward, una con el criterio AIC y la otra con el criterio BIC.

Use como modelo completo, un modelo lineal aditivo con todas las variables físico-químicas. El modelo más pequeño será el modelo sólo con el intercepto y la variable volatile.acidity. (10 pts)

6. Compare los modelos seleccionados por AIC y BIC en el set de validación usando el criterio del MSE (Error cuadrático medio de la predicción). Escoja el mejor. (10 pts)
7. Estime el modelo seleccionado en el literal 6 en el set de estimación y proporcione un intervalo del 99% confianza para el coeficiente de UNA de las variables físico químicas que usted considere más importantes para explicar la calidad del vino. Interprete el intervalo en terminología del problema estudiado. (10 pts)

Cargar

Examen guardado en 10:27