

				PUNTAJE	
AÑO:	2022-2023	PERÍODO:	PRIMER TÉRMINO	TEMA 1	
MATERIA:	ESTADÍSTICA II	PROFESORES:	FRANCISCO MOREIRA, PAMELA CROW, ANDREA GARCIA	TEMA 2	
EVALUACIÓN:	SEGUNDA	FECHA:	01/09/2022	TEMA 3	
				TEMA 4	

COMPROMISO DE HONOR

Yo, al firmar este compromiso, reconozco que el presente examen está diseñado para ser resuelto de manera individual, que puedo usar una calculadora ordinaria para cálculos aritméticos, un lápiz o esferográfico; que solo puedo comunicarme con la persona responsable de la recepción del examen; y, cualquier instrumento de comunicación que hubiere traído, debo apagarlo y depositarlo en la parte anterior del aula, junto con algún otro material que se encuentre acompañándolo. No debo además, consultar libros, notas, ni apuntes adicionales a las que se entreguen en esta evaluación. Los temas debo desarrollarlos de manera ordenada.

Firmo al pie del presente compromiso, como constancia de haber leído y aceptar la declaración anterior.

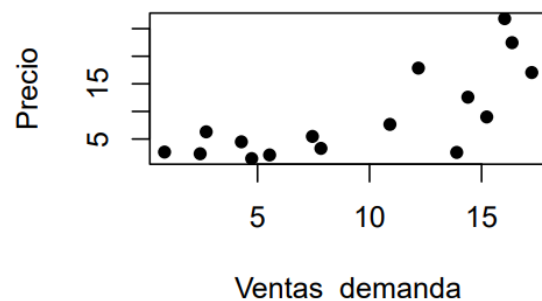
"Como estudiante de ESPOL me comprometo a combatir la mediocridad y actuar con honestidad, por eso no copio ni dejo copiar".

Firma: NÚMERO DE MATRÍCULA: PARALELO:

TEMA 1 (25 PTS)

Las ventas minoristas son una métrica económica que mide la demanda del consumidor de productos terminados. Se desea modelar el precio de televisores en base a las ventas minoristas (demanda) de dicho producto. Para ello se han recogido datos de 16 televisores:

Precio (cientos \$)	Venta_demanda (millones \$)
2.57	13.89
9.00	15.23
22.46	16.36
4.49	4.28
1.45	4.74
6.30	2.71
7.65	10.91
26.81	16.03
12.58	14.39
2.33	2.44
5.47	7.45
2.64	0.85
2.11	5.54
3.30	7.83
17.85	12.18
17.05	17.24



- a) Estime un modelo de regresión lineal simple. Proporcione la ecuación de la recta estimada. (Ayuda: $S_{xx}=478.1446$, $S_{xy}=505.9919$) (5 PTS)

b) Interprete el coeficiente estimado $\hat{\beta}_1$ en el contexto del problema. **(5 PTS)**

c) Proporcione un intervalo del 99% de predicción para el precio de un televisor cuya demanda (ventas minoristas) ascienden a 16.3 millones de \$. Interprete el intervalo en el contexto del problema. **(10 PTS)**

d) El coeficiente de determinación de este modelo asciende a 0.5581. ¿Es un buen modelo para explicar la variabilidad total en el precio? Basado en el gráfico de los datos ¿qué recomendaría para mejorar el modelo? **(5 PTS)**

TEMA 2 (35 PTS)

Un distribuidor de automóviles realizó un estudio para determinar la relación entre el costo de los automóviles (en miles de dólares) con el peso, la potencia y el número de cilindros (4 o 6). Los datos de 20 modelos se corrieron en un paquete estadístico y se obtuvo los siguientes resultados: (utilice un $\alpha = 0.01$)

ANOVA					
Fuentes de Variación	SC	GL	MC	F	p
Regresión	10194.85	3			
Residual	3382.24				
Total	13577.09	19			

Coeficientes				
	Estimador	Desv. Estándar	t	p
(Intercepto)	-50.995			
X1	0.043			
X2	0.165			
X3	8.841			

Intervalos de confianza para β	
Estimador	Error de Estimación
-50.995	-
0.043	0.011
0.165	0.112
8.841	8.342

a) Plantee el modelo estadístico (teórico) a utilizar (ecuación) y sus supuestos (5 PTS)

b) Cuál es el coeficiente de determinación múltiple e interprételo (5 PTS)

- c) Realice la prueba de hipótesis global para determinar si todas las variables explican el precio de los automóviles y calcule el valor p de la prueba. **(10 PTS)**

- d) Realice la prueba de los coeficientes individuales para determinar si alguna de ellas puede eliminarse. **(10 PTS)**

- e) Concluya sobre el modelo. **(5 PTS)**

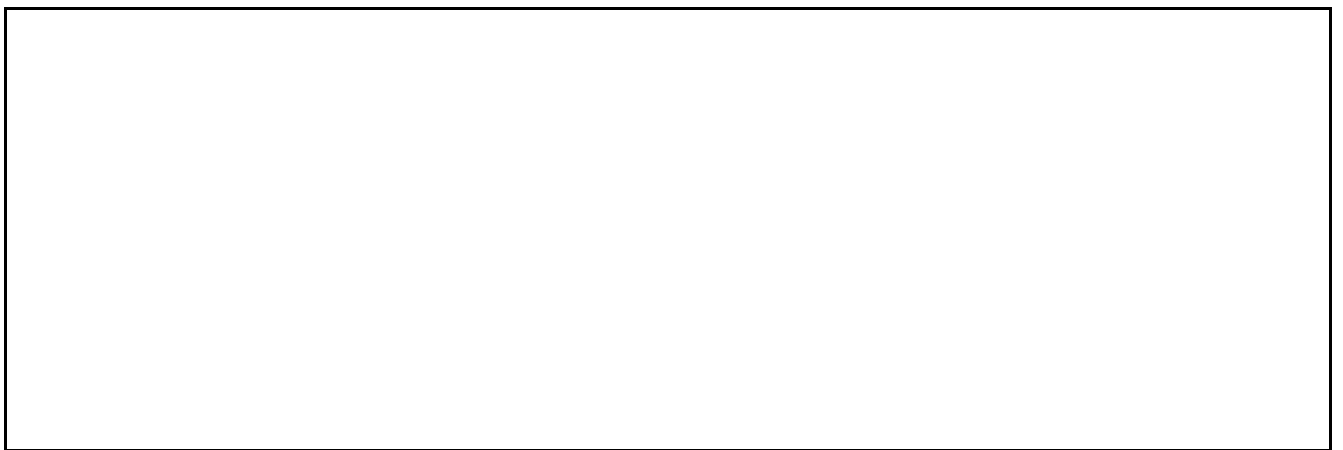
TEMA 3 (20 PTS)

Una empresa de marketing está realizando un estudio de mercado donde desea conocer si la aceptación de un producto difiere según la marca. Se está evaluando 8 marcas distintas. Para ello realiza un experimento con 80 personas, las divide en 8 grupos de 10 personas cada uno y a cada grupo les asigna una marca y les pide que le den un puntaje al producto.

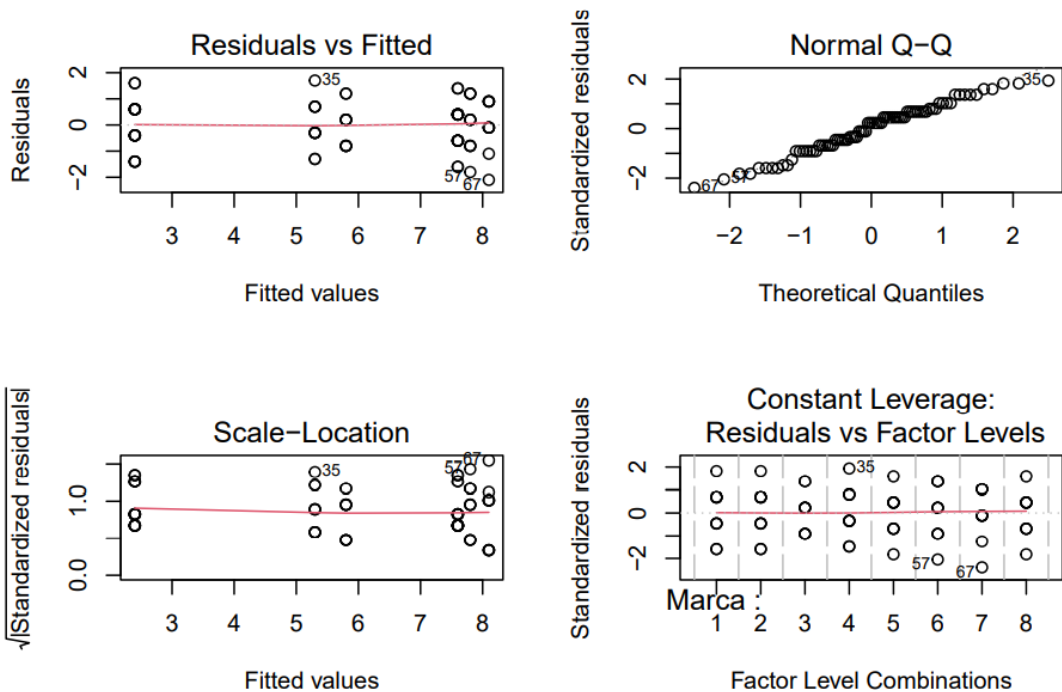
- a) A continuación se presenta los resultados del modelo ANOVA. Escriba el contraste de hipótesis, el estadístico de prueba, el valor p de la prueba y su conclusión al 5% de significancia. La conclusión debe ser redactada con aplicación al problema. **(7 PTS)**

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Marca	7	391.0	55.85	65.07	<2e-16 ***
Residuals	72	61.8	0.86		

 signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



- b) En base a los gráficos de los residuos, ¿se cumplen los supuestos del modelo? Sustente su respuesta: Indique cada supuesto y en qué gráfico se basa para comprobarlo. ¿Es válida su conclusión del literal anterior? **(5 PTS)**



c) A continuación se presenta los resultados de pruebas comparaciones múltiples. A un 5% de significancia, ¿Qué marcas no tienen el mismo nivel de aceptación? ¿Cuál o cuáles marcas tienen mayor aceptación? **(8 PTS)**

```
## Tukey multiple comparisons of means
## 95% family-wise confidence level
##
## Fit: aov(formula = Rating ~ Marca, data = df)
##
## $Marca
##          diff          lwr          upr          p adj
## 2-1 -1.199041e-14 -1.2934489  1.2934489  1.0000000
## 3-1  3.400000e+00  2.1065511  4.6934489  0.0000000
## 4-1  2.900000e+00  1.6065511  4.1934489  0.0000000
## 5-1  5.200000e+00  3.9065511  6.4934489  0.0000000
## 6-1  5.400000e+00  4.1065511  6.6934489  0.0000000
## 7-1  5.700000e+00  4.4065511  6.9934489  0.0000000
## 8-1  5.200000e+00  3.9065511  6.4934489  0.0000000
##
## 3-2  3.400000e+00  2.1065511  4.6934489  0.0000000
## 4-2  2.900000e+00  1.6065511  4.1934489  0.0000000
## 5-2  5.200000e+00  3.9065511  6.4934489  0.0000000
## 6-2  5.400000e+00  4.1065511  6.6934489  0.0000000
## 7-2  5.700000e+00  4.4065511  6.9934489  0.0000000
## 8-2  5.200000e+00  3.9065511  6.4934489  0.0000000
## 4-3 -5.000000e-01 -1.7934489  0.7934489  0.9273590
## 5-3  1.800000e+00  0.5065511  3.0934489  0.0011210
## 6-3  2.000000e+00  0.7065511  3.2934489  0.0001960
## 7-3  2.300000e+00  1.0065511  3.5934489  0.0000120
## 8-3  1.800000e+00  0.5065511  3.0934489  0.0011210
## 5-4  2.300000e+00  1.0065511  3.5934489  0.0000120
## 6-4  2.500000e+00  1.2065511  3.7934489  0.0000017
## 7-4  2.800000e+00  1.5065511  4.0934489  0.0000001
## 8-4  2.300000e+00  1.0065511  3.5934489  0.0000120
## 6-5  2.000000e-01 -1.0934489  1.4934489  0.9997058
## 7-5  5.000000e-01 -0.7934489  1.7934489  0.9273590
## 8-5  5.329071e-15 -1.2934489  1.2934489  1.0000000
## 7-6  3.000000e-01 -0.9934489  1.5934489  0.9959859
## 8-6 -2.000000e-01 -1.4934489  1.0934489  0.9997058
## 8-7 -5.000000e-01 -1.7934489  0.7934489  0.9273590
```

TEMA 4 (20 PTS)

Se desea medir el efecto de un curso de capacitación sobre manejo del tiempo en la productividad de los empleados del área de producción de una empresa. Como el área de producción consta de 3000 empleados, para una primera etapa se seleccionan al azar 10 empleados para que tomen el curso y luego de un mes se mide su productividad y se compara con la productividad del mes anterior antes del curso. Los datos son los siguientes:

empleado	1	2	3	4	5	6	7	8	9	10
antes	30	38	53	79	26	78	81	60	58	15
despues	38	49	64	88	27	75	79	63	70	20

De estudios anteriores se conoce que la diferencia en el índice de productividad no sigue una distribución normal pero sí simétrica. ¿El curso ayuda a aumentar la productividad? Utilice un nivel de significancia del 5%. Utilice la prueba con mayor potencia posible. Escriba el contraste de hipótesis, calcule el estadístico y concluya. (20 PTS)