

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS
DEPARTAMENTO DE MATEMÁTICA
INGENIERÍA EN AUDITORIA Y CONTADURÍA PÚBLICA AUTORIZADA

AÑO LECTIVO: 2019-2020	PERIODO: Segundo Terminó
MATERIA: Técnicas de Muestreo Y Análisis Multivariado	PROFESORES:
EVALUACIÓN: Segunda	Msc. Sandra Gonzalez C.
TIEMPO DE DURACIÓN: 2 horas	FECHA: 30/01/2020

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: 1

Nota 1: Es válido utilizar aproximaciones en caso de ser necesario; use cuatro decimales de aproximación.

Nota 2: Trabaje con el 95% de confianza para sus cálculos

- 1) (20 puntos) La siguiente tabla muestra los recorridos, en tiempos y consumo de combustible de un vehículo en distintas rutas.

Recorrido (Km)	250	126	270	300
Tiempo (horas)	5	2	3	4
Consumo (litros)	27.5	18.9	51.3	48

- a) (2 puntos) Calcule el consumo por cada 100 km y la velocidad media en (*km/hora*) y completar la siguiente tabla de datos.

Consumo a los 100 km	11			16
Velocidad Media		63		75

- b) (5 puntos) Determine los estimadores de mínimos cuadrados ordinarios para un modelo de regresión lineal simple que explique el consumo por cada 100 km en función de la velocidad media y escriba la ecuación de regresión lineal simple que mejor modele los datos.
- c) (5 puntos) Determine la tabla ANOVA, plantee y realice la prueba global del modelo del literal b)
- d) (3 puntos) Determine R^2 del modelo del literal b) utilizando la tabla ANOVA e interprete su respuesta
- e) (5 puntos) Estime el consumo de gasolina a los 100 km para las velocidades medias de $x_1 = 55$ y $x_2 = 95$ y encuentre un intervalo de confianza para cada estimación calculada.

2) (20 puntos) Sea el modelo de regresión lineal múltiple

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon$$

Y los siguientes datos muestrales:

Datos para el modelo

Y	1	3	1	2	3	4	1
X1	7	3	6	3	8	6	9
X2	10	3	8	1	6	3	13

Matrices del modelo:

$$X'X = \begin{pmatrix} 7 & 42 & 44 \\ & 284 & 313 \\ & & 388 \end{pmatrix} \quad X'Y = \begin{pmatrix} 15 \\ 85 \\ 72 \end{pmatrix}$$

$$(X'X)^{-1} = \begin{pmatrix} 1.4992 & -0.3096 & 0.0797 \\ & 0.0957 & -0.0420 \\ & & 0.0275 \end{pmatrix}$$

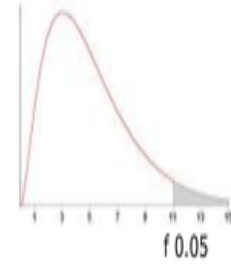
- a) (5 puntos) Determine un modelo de regresión que mejor represente los datos
- b) (2 puntos) Determinar los \hat{y}_i
- c) (5 puntos) Calcule la tabla ANOVA del modelo y realice la prueba global del modelo, utilizando un 95% de confianza.
- d) (3 puntos) ¿Cuáles β_i son significativos en el modelo? Justifique su respuesta
- e) (2 puntos) Realice un intervalo de confianza para cada β_i significativo del modelo planteado
- f) (3 puntos) Determine el coeficiente de determinación R^2 del modelo e interprete su resultado

3) (10 puntos) Califique como verdadera o falsa cada una de las siguientes afirmaciones:

	Verdadera	Falsa
a) En regresión lineal: La prueba de normalidad Shapiro Wilks se prefiere a la prueba de Kolmogorov Smirnov cuando el tamaño de la muestra es menor a 50 (pequeña).		
b) En regresión Lineal: En la prueba de Durbin Watson el estadístico DW toma valores entre 0 y 1.		
c) En regresión Lineal: El término Homocedasticidad se refiere a que la $v(\epsilon_i) = \sigma^2$		
d) Las gráficas P-P y Q-Q se utilizan para verificar la covarianza de los errores		
e) En un modelo de regresión lineal se cumple que $SCT=SCR+SCE$		
f) El coeficiente de determinación R^2 es un valor entre 0 y 1, y mientras más cercano a 1 es mejor el modelo de regresión lineal.		
g) En la prueba global del modelo utilizando la tabla ANOVA, se utiliza el estadístico F para realizar el contraste de hipótesis de la significancia de los β_i ($\beta_1, \beta_2, \dots, \beta_{p-1}$) en el modelo de regresión múltiple.		
h) En regresión lineal múltiple $(X^tX)^{-1}$ existe siempre y cuando el rango de X sea p, es decir, X tenga p columnas linealmente independientes		
i) Un modelo de regresión logística tiene la forma $\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1$		
j) En un modelo de regresión logística $odds = \frac{1-p}{p}$		

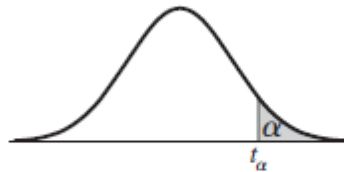
Tabla D.9: VALORES CRÍTICOS DE LA DISTRIBUCIÓN F (0,05)

área a la derecha del valor crítico = 0,05



g.d.l	Grados de libertad del Numerador															g.d.l
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,0	243,9	244,7	245,4	245,9	1
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385	19,396	19,405	19,413	19,419	19,424	19,429	2
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,786	8,763	8,745	8,729	8,715	8,703	3
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,936	5,912	5,891	5,873	5,858	4
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,772	4,735	4,704	4,678	4,655	4,636	4,619	5
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,099	4,060	4,027	4,000	3,976	3,956	3,938	6
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,677	3,637	3,603	3,575	3,550	3,529	3,511	7
8	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,438	3,388	3,347	3,313	3,284	3,259	3,237	3,218	8
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,179	3,137	3,102	3,073	3,048	3,025	3,006	9
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	3,020	2,978	2,943	2,913	2,887	2,865	2,845	10

Tabla 5 Puntos porcentuales de las distribuciones t



$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	gl
3.078	6.314	12.706	31.821	63.657	1
1.886	2.920	4.303	6.965	9.925	2
1.638	2.353	3.182	4.541	5.841	3
1.533	2.132	2.776	3.747	4.604	4
1.476	2.015	2.571	3.365	4.032	5
1.440	1.943	2.447	3.143	3.707	6
1.415	1.895	2.365	2.998	3.499	7
1.397	1.860	2.306	2.896	3.355	8
1.383	1.833	2.262	2.821	3.250	9
1.372	1.812	2.228	2.764	3.169	10
1.363	1.796	2.201	2.718	3.106	11
1.356	1.782	2.179	2.681	3.055	12
1.350	1.771	2.160	2.650	3.012	13
1.345	1.761	2.145	2.624	2.977	14
1.341	1.753	2.131	2.602	2.947	15

Algunas fórmulas: Modelo de regresión lineal simple

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$ $S_{xx} = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{\sum x_i^2}{n} - \bar{x}^2$ $S_{yy} = \frac{\sum (y_i - \bar{y})^2}{n} = \frac{\sum y_i^2}{n} - \bar{y}^2$ $S_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n} = \frac{\sum x_i y_i}{n} - \bar{x} \bar{y}$ $R^2 = \frac{SCR}{SCT}$ <p>Estadístico prueba global del modelo:</p> $F_o = \frac{MCR}{MCE} > F_{(\alpha, p-1, n-p)}$	$SCR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ <p>G de Libertad: p-1</p> $SCE = \sum_{i=1}^n (y_i - \hat{y})^2$ <p>Grados de Libertad: n-p</p> $SCT = \sum_{i=1}^n (y_i - \bar{y})^2$ <p>Grados de Libertad: n-1</p>
<p><u>Estimación modelo de regresión lineal simple</u></p> $S_e = \sqrt{MCE}$ <p>El intervalo de $(1-\alpha)100\%$ de confianza para la estimación individual es:</p> $\hat{y}_o \pm t_{\frac{\alpha}{2}, n-2} S_e \sqrt{\left[1 + \frac{1}{n} + \frac{(x_o - \bar{x})^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \right]}$	