

AÑO: 2019	PERIODO: PRIMER TÉRMINO
MATERIA: ESTADÍSTICA INFERENCIAL	PROFESORES: Mario Solorzano, Jennifer Marcillo, Kenny Escobar, Francisco Moreira
EVALUACIÓN: SEGUNDA	
TIEMPO DE DURACIÓN: 2 HORAS	FECHA: 29 de agosto de 2019

PUNTAJE	
TEMA 1	
TEMA 2	
TEMA 3	

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 (40 PUNTOS)

El conjunto de datos que se analiza a continuación brinda estimaciones del porcentaje de grasa corporal determinado por el peso de una persona bajo el agua y varias medidas de circunferencia corporal para “n” hombres. Es de nuestro interés plantear una ecuación para estimar la grasa corporal, de esta manera la variable respuesta es el porcentaje de grasa corporal (grasacorp), y los predictores de interés son: edad (años), peso (libras), altura (pulgadas) y circunferencias (en cm) del cuello, pecho, abdomen, cadera, muslo, rodilla, tobillo, bíceps, antebrazo y muñeca.

A continuación se detalla la primera propuesta para modelar la variable respuesta

Propuesta 1

call:

lm(formula = grasacorp ~ edad + altura + cuello + pecho + abdomen + cadera + muslo + antebrazo + muñeca, data = data)

Coefficientes:

	Estimador	Desv. Std	valor T	Pr(> t)
Intercepto	6.36691	7.46173	_____	_____
edad	0.07013	_____	2.282	_____
altura	_____	0.08281	-1.744	0.082505
cuello	-0.55744	0.21845	_____	_____
pecho	_____	0.08896	-0.866	_____
abdomen	0.91854	0.08411	_____	_____
cadera	-0.34136	_____	-2.873	_____
muslo	0.24175	_____	1.843	0.066509
antebrazo	_____	0.18891	2.586	_____
muñeca	-1.67707	_____	-3.363	0.000896

Error estándar residual: 4.302 con _____ grados de libertad

R-cuadrado: 0.7452, R-cuadrado ajustado: 0.7357

Estadístico F: 78.63 con 9 y _____ grados de libertad, p-value: < 2.2e-16

a) Complete los espacios en blanco detallando las fórmulas utilizadas. Además, determine el tamaño de la muestra, sabiendo que la Suma Cuadrática del Error es 4478,74



b) ¿Es un buen modelo? En caso de que no lo sea, explique el procedimiento que realizaría para mejorarlo.

c) Se analizaron los supuestos del modelo, encontrándose que no se cumplen dichos supuestos. ¿Cuáles son éstos supuestos?

d) Especifique el modelo teórico y calculado de la propuesta 2

Propuesta 2

Coefficientes:

	Estimador	Desv. Std	valor T	Pr(> t)
Intercepto	-0.51460	0.20000	_____	_____
edad	0.07968	0.02774	_____	_____
cuello	-0.50277	_____	-2.518	0.01244
abdomen	0.87576	0.05451	16.067	< 2e-16
cadera	-0.31089	0.10550	_____	_____
muslo	0.31181	_____	2.501	0.01306
muñeca	_____	0.43574	-4.253	3.01e-05

Error estándar residual: 4.176 con _____ grados de libertad

R-cuadrado: 0.9607, R-cuadrado ajustado: 0.9598

Estadístico F: _____ con _____ y _____ grados de libertad, p-value: < 2.2e-16

Análisis de Varianza (ANOVA)

v. Respuesta: grasacorp

	SC	gl	MC	F value	Pr(>F)
SCR	103425	_____	_____	_____	_____
SCE	_____	_____	_____		
SCT	_____	_____			

e) En la propuesta 2 se detectaron algunas observaciones atípicas que distorsionaban el ajuste del modelo. Complete los espacios vacíos del modelo y la tabla ANOVA ¿Cuántas observaciones fueron eliminadas?



f) Se han validado todos los supuestos del modelo, excepto el de normalidad. Los errores de la propuesta 2 se detallan a continuación. Pruebe que éstos errores cumplen con dicho supuesto (plantee el contraste de hipótesis).

Clases	Frecuencia
$(-9.4, -7.46]$	9
$(-7.46, -5.54]$	12
$(-5.54, -3.63]$	25
$(-3.63, -1.71]$	48
$(-1.71, 0.207]$	42
$(0.207, 2.12]$	34
$(2.12, 4.04]$	33
$(4.04, 5.96]$	24
$(5.96, 7.88]$	18
$(7.88, 9.81]$	4



TEMA 2 (30 PUNTOS)

Se realiza un experimento para examinar el efecto de la edad en el ritmo cardiaco de las personas luego de realizar un ejercicio específico. Se seleccionan aleatoriamente diez personas para cada uno de los siguientes grupos de edad: 10-19, 20-39, 40-59 y 60-69. Cada persona caminó en una cinta de correr durante un periodo de 12 minutos, registrando luego el incremento en el ritmo cardiaco (diferencia en el ritmo cardiaco antes y después del ejercicio). Cálculos preliminares indican una $SCT_{Total}=1002.975$ y una $SCT_{Tratamiento}=67.475$

Edad	Incremento Promedio de Ritmo cardiaco
10-19	30.9
20-39	27.5
40-59	29.5
60-69	28.2

- Plantee el modelo para analizar las diferencias en el incremento del ritmo cardiaco para los grupos de edades, además escriba las funciones de densidad para el incremento de ritmo cardiaco para cada grupo de edad.
- Pruebe que \bar{Y}_1 es un estimador insesgado de μ_1
- ¿Existe evidencia estadística suficiente que indique que hay un efecto del grupo de edad en el incremento del ritmo cardiaco?
- Explique a partir de las distribuciones muestrales de las sumas cuadráticas porqué el estadístico de prueba F sigue una distribución F de Fisher.
- En caso de haber diferencia significativa en el literal b), proponga alguna metodología que permita identificar qué grupo o grupos de edades presentan el mayor incremento promedio de ritmo cardiaco.



TEMA 3 (30 PUNTOS)

Dentro de un gimnasio se han sometido a cierto tipo de ejercicios a varias personas, dándonos dos escenarios posibles:

Escenario 1: Se ha medido el peso (kg.) antes y después a 10 personas para determinar si existen diferencias significativas en cuanto al tipo de ejercicios que realizan durante un mes; dando los siguientes resultados:

Antes	65,2	70,2	80,1	81,4	74,0	64,3	72,5	69,4	68,7	69,5
Después	63,4	68,5	75,4	82,4	70,5	64,3	70,5	72,4	68,7	66,7

Escenario 2: Se ha medido el peso (kg.) en dos días diferentes a dos grupos de 10 personas para determinar si existen diferencias significativas en cuanto a los grupos que se pesaron dependiendo del día; dando los siguientes resultados:

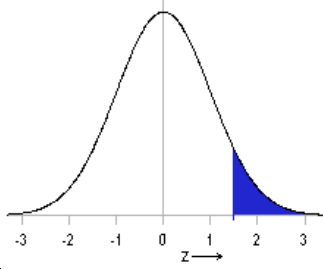
Lunes	75,2	82,4	82,2	91,4	88,7	88,0	74,5	75,9	68,7	58,4
Martes	73,1	85,7	90,0	87,4	85,8	88,0	78,6	74,8	75,4	65,9

- ¿Cuáles son las consideraciones para realizar un contraste no paramétrico para cada escenario?
- Identifique la prueba correspondiente no paramétrica más acertada para cada escenario.
- Realice las pruebas no paramétricas según corresponda para determinar si existen o no diferencias significativas en ambos escenarios.



TABLAS

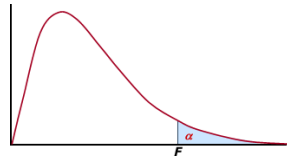
Distribución Normal Estándar



Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,10	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,20	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,30	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,40	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,50	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,60	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,70	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,80	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,90	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,00	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,10	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,20	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,30	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,40	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,50	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,60	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,70	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,80	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,90	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,00	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,10	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,20	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,30	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,40	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
2,50	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
2,60	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
2,70	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
2,80	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
2,90	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
3,00	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010



Distribución F de Fisher



Test de Kolmogorov-Smirnov sobre Bondad de Ajuste

Grados de libertad en el Denominador, ν_2	α	Grados de libertad en el Numerador, ν_1			
		1	2	3	4
1	0,1	39,86	49,50	53,59	55,83
	0,05	161,40	199,50	215,70	224,60
	0,025	647,80	799,50	864,20	899,60
	0,01	4052,00	4999,50	5403,00	5625,00
2	0,1	8,53	9,00	9,16	9,24
	0,05	18,51	19,00	19,16	19,25
	0,025	38,51	39,00	39,17	39,25
	0,01	98,50	99,00	99,17	99,25
...					
30	0,1	2,88	2,49	2,28	2,14
	0,05	4,17	3,32	2,92	2,69
	0,025	5,57	4,18	3,59	3,25
	0,01	7,56	5,39	4,51	4,02
40	0,1	2,84	2,44	2,23	2,09
	0,05	4,08	3,23	2,94	2,61
	0,025	5,42	4,05	3,46	3,13
	0,01	7,31	5,18	4,31	3,83

n	Nivel de significancia α				
	0,1	0,05	0,02	0,01	0,005
7	0.43607	0.48342	0.53844	0.57581	0.60975
8	0.40962	0.45427	0.50654	0.54179	0.57429
9	0.38746	0.43001	0.47960	0.51332	0.54443
10	0.36866	0.40925	0.45562	0.48893	0.51872
11	0.35242	0.39122	0.43670	0.46770	0.49539
...					
47	0.17481	0.19420	0.21715	0.232980	0.24776
48	0.17301	0.19221	0.21493	0.23059	0.24523
49	0.17128	0.19028	0.21281	0.22832	0.24281
50	0.16959	0.18841	0.21068	0.22604	0.24039
>50	$1.22/\sqrt{n}$	$1.36/\sqrt{n}$	$1.52/\sqrt{n}$	$1.63/\sqrt{n}$	$1.73/\sqrt{n}$

Valores críticos para prueba de Rangos con Signos

n	Unilateral $\alpha=0,025$	Unilateral $\alpha=0,05$
	Bilateral $\alpha=0,05$	Bilateral $\alpha=0,1$
5		1
6	1	2
7	2	4
8	4	6
9	6	8
10	8	11
11	11	14
12	14	17
13	17	21
14	21	26
15	25	30
16	30	36
17	35	41
18	40	47
19	46	54
20	52	60

Valores críticos para la prueba de suma de rangos de Wilcoxon Mann-Whitney

Prueba de una cola con $\alpha = 0.025$ o prueba de dos colas con $\alpha = 0.05$

n_1	n_2									
	4	5	6	7	8	9	10	11	12	
1										
2					0	0	0	0	1	
3			0	1	1	2	2	3	3	4
4	0	1	2	3	4	4	5	6	7	
5		2	3	5	6	7	8	9	11	
6			5	6	8	10	11	13	14	
7				8	10	12	14	16	18	
8					13	15	17	19	22	
9						17	20	23	26	
10							23	26	29	
11								30	33	
12									37	

$$u_1 = w_1 - \frac{n_1(n_1 + 1)}{2}$$

$$\mu_{u_1} = \frac{n_1 n_2}{2}$$

$$\sigma_{u_1}^2 = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}$$

