UAEM Universidad Autónoma del Estado de México

CENTRO UNIVERSITARIO UAEM ZUMPANGO

LICENCIATURA EN TURISMO

TEMA: PRUEBAS DE COMPARACION DE MEDIAS Y POBLACIONES

M. EN C. LUIS ENRIQUE KU MOO

FECHA: SEPTIEMBRE DE 2019

UNIDAD DE APRENDIZAJE

"ESTADISTICA"

UNIDAD DE COMPETENCIA III:

"PRUEBAS ESTADÍSTICAS APLICADAS A UN CASO PRÁCTICO EN PARTICULAR RELACIONADO CON EL TURISMO"

3.2. Pruebas de comparación de medias y poblaciones

3.2.1 Pruebas Paramétricas, Prueba T de Student, Análisis de Varianza. (ANOVA) de una sola vía y pruebas de comparación de medias.

3.2.2 Pruebas no Paramétricas, suma de rangos de Wilcoxon y Kruskall –Wallis.

CONTENIDO DE LA UNIDAD DE COMPETENCIA

3.1 Pruebas de asociación

3.1.1 Pruebas paramétricas, Análisis de correlación y Regresión Lineal Simple (RLS).

3.1.2 No Paramétricas, correlación de Spearman, pruebas de independencia y bondad de ajuste de Ji cuadrada

3.2 Pruebas de comparación de medias y poblaciones

3.2.1 Pruebas Paramétricas, Prueba T de Student, Análisis de Varianza (ANOVA) de una sola vía y pruebas de comparación de medias.

3.2.2 Pruebas no Paramétricas, suma de rangos de Wilcoxon y Kruskall-Wallis

3.3 Aplicación a un caso práctico

3.3.1 Diseño y análisis de información

3.3.2 Interpretación y presentación de resultados

OBJETIVO

Objetivos del área curricular o disciplinaria: Analizar y aplicar las diferentes perspectivas teórico-metodológicas de la investigación en ciencias sociales para abordar el estudio del turismo.

Objetivos de la unidad de aprendizaje: Aplicar los métodos y técnicas estadísticas para el análisis e interpretación de datos.

Objetivo de la Unidad de competencia:

Aplicar las pruebas estadísticas que correspondan al planteamiento y solución de problemas o casos relacionados con el Turismo.

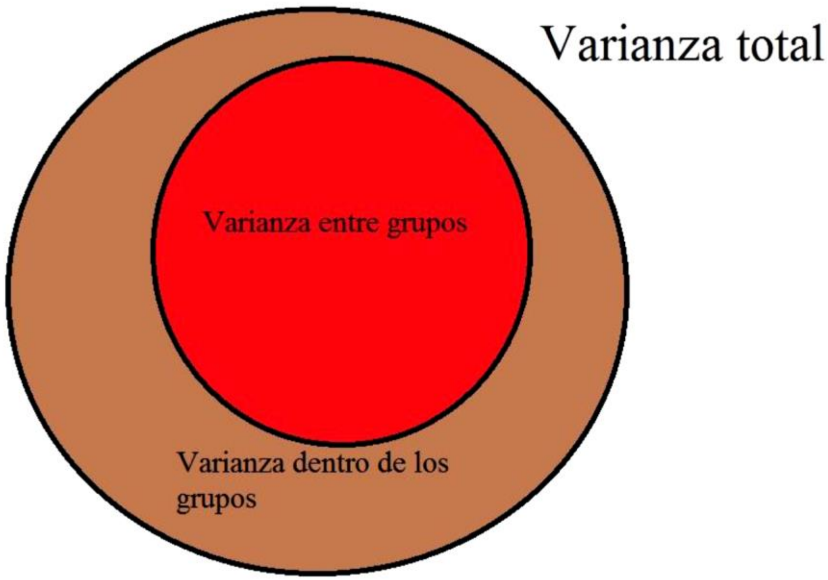
JUSTIFICACIÓN

El presente material sirve de apoyo a la Tercera Unidad de competencia “Pruebas estadísticas aplicadas a un caso práctico en particular relacionado con el Turismo” de la Unidad de aprendizaje Estadística que se imparte en el tercer período de la Licenciatura en Turismo.

Se desarrolla el Tema de “Pruebas de comparación de medias y poblaciones” tanto paramétricas como no paramétricas, para aplicar las pruebas estadísticas que correspondan al planteamiento y solución de problemas o casos relacionados con el Turismo.

ANALISIS DE VARIANZA

Es un método para comparar dos o más medias, que es necesario porque cuando se quiere comparar más de dos medias es incorrecto utilizar repetidamente el contraste basado en la t de Student.



SUPOSICIONES DE ANÁLISIS DE VARIANZA

La distribución F también se usa para probar la igualdad de más de dos medias con una técnica llamada análisis de varianza (Andeva o ANOVA). Condiciones:

- La población tiene una distribución normal, desviaciones estándar iguales y muestras independientes.
- Está determinado por dos parámetros: los grados de libertad (gl) en el numerador y los grados de libertad en el denominador.
- El valor de F no puede ser negativo y es una distribución continua con sesgo positivo.
- Sus valores varían de 0 a ∞ . Conforme $F \to \infty$ la curva se aproxima al eje X.

OBSERVACIONES SOBRE ANOVA

- Si se muestrean k poblaciones, entonces los *gl* (numerador) = $k – 1$
- Si hay un total de n puntos en la muestra, entonces los *gl* (denominador) = $n – k$
- El estadístico de prueba se calcula con: $F = [(\text{SCTr}) /(k - 1)] /[(\text{SCE}) /(N - k)]$.
- SCT es la suma de cuadrados de los tratamientos.
- SCE es la suma de cuadrados del error.
- Sea *TC* el total de la columna, n el número de observaciones en cada columna, y ΣX la suma de todas las observaciones.

ANALISIS DE VARIANZA GRAFICAMENTE

MUESTRA 1	
Unidad 1	X_{11}
Unidad 2	X_{12}
Unidad 3	X_{13}
Unidad 4	X_{14}
Unidad 5	X_{15}
Unidad 6	X_{16}
Unidad 7	X_{17}
Unidad 8	X_{18}
Unidad 9	X_{19}

Promedio 1

$X_{1\bullet}$

MUESTRA 2	
Unidad 1	X_{21}
Unidad 2	X_{22}
Unidad 3	X_{23}
Unidad 4	X_{24}
Unidad 5	X_{25}
Unidad 6	X_{26}
Unidad 7	X_{27}
Unidad 8	X_{28}
Unidad 9	X_{29}

Promedio 2

$X_{2\bullet}$

MUESTRA 3	
Unidad 1	X_{31}
Unidad 2	X_{32}
Unidad 3	X_{33}
Unidad 4	X_{34}
Unidad 5	X_{35}
Unidad 6	X_{36}
Unidad 7	X_{37}
Unidad 8	X_{38}
Unidad 9	X_{39}

Promedio 3

$X_{3\bullet}$

Promedio General

$X_{\bullet\bullet}$

ANALISIS DE VARIANZA

Variación Total. La que toma en cuenta la variación entre todas las unidades tomando en cuenta la diferencia a la gran Media (promedio general)

$$\sum (X_{11} - \chi_{\bullet\bullet})^2 + (X_{12} - \chi_{\bullet\bullet})^2 + \ldots + (X_{39} - \chi_{\bullet\bullet})^2$$

La Varianza ENTRE GRUPOS o tratamiento compara las medias de cada Grupo con la gran Media

$$\sum n_1 (X_{1\bullet} - \chi_{\bullet\bullet})^2 + n_2 (X_{2\bullet} - \chi_{\bullet\bullet})^2 + n_3 (X_{3\bullet} - \chi_{\bullet\bullet})^2$$

La varianza DENTRO GRUPOS (error) considera la variación que hay dentro de cada grupo

Para cada Grupo

$$\sum (X_{11} - \chi_{1\bullet})^2 + (X_{12} - \chi_{1\bullet})^2 + \ldots + (X_{19} - \chi_{1\bullet})^2 +$$

$$\sum (X_{21} - \chi_{2\bullet})^2 + (X_{22} - \chi_{2\bullet})^2 + \ldots + (X_{29} - \chi_{2\bullet})^2 +$$

$$\sum (X_{31} - \chi_{3\bullet})^2 + (X_{32} - \chi_{3\bullet})^2 + \ldots + (X_{39} - \chi_{3\bullet})^2 =$$

PROCEDIMIENTO ANOVA

Para prueba de dos colas, el estadístico de prueba está dado por: $F = \frac{s_1^2}{s_2^2}$

Hipótesis nula: las medias de las poblaciones son iguales. $(\mu_1 = \mu_2 = \cdots = \mu_k)$

Hipótesis alternativa: al menos una de las medias es diferente $(\mu_1 \neq \mu_2 \neq \cdots \neq \mu_k)$

Estadístico de prueba: *F* = (varianza *entre* muestras)/(varianza *dentro* de muestras).

Regla de decisión: para un nivel de significancia α, la hipótesis nula se rechaza si *F* (calculada) es mayor que *F* (en tablas) con grados de libertad en el numerador y en el denominador.

Tabla de ANOVA

- LA TABLA DE ANÁLISIS DE VARIANZA. Resume los valores de las variaciones y los llamados grados de libertad (gl).

Fuente de Variación	gl	Suma de Cuadrados	Cuadrados medios	F
Entre Grupos O tratamientos	K-1	$SCTr=\sum n_i (X_{1\bullet} - \chi_{\bullet\bullet})^2$	CMTr= SCTr/GLTr	CMTr/CME
Intra o Dentro Grupos (Error)	n - K	$SCE=\sum \sum (X_{ij} - \chi_{i\bullet})^2$ $\sum_{j=1}^{k} (n_j - 1) {s_j}^2$	CME= SCE/GLE	
TOTAL	n - 1	$SCT=\sum \sum (X_{ij} - \chi_{\bullet\bullet})^2$		

ANALISIS DE VARIANZA

EJEMPLO.

	Proveedor 1	Proveedor 2	Proveedor 3	Proveedor 4
	18.5	26.3	20.6	25.4
	24.0	25.3	25.2	19.9
	17.2	24.0	20.8	22.6
	19.9	21.2	24.7	17.5
	18.0	24.5	22.9	20.4
Media	19.52	24.26	22.84	21.16
Desv. Est.	2.69	1.92	2.13	2.98

ANALISIS DE VARIANZA

$$\textit{Media General} = \frac{\sum_{j=1}^{c}\sum_{i=1}^{n_j} x_{ij}}{n} = \frac{438.9}{20} = 21.945$$

$SCTr = \sum_{j=1}^{c} n_j\left(\bar{x}_j - \bar{\bar{x}}\right)^2$ =(5)(19.52 - 21.945)^2 + (5)(24.945 – 21.26)^2 + (5)(22.84 -21.945)^2 + (5)(21.16 - 21.945)^2 = 63.2855

SCE = $\sum_{j=1}^{c}\sum_{i=1}^{n_j}\left(x_{ij} - \bar{x}_j\right)^2$ =(18.5 – 19.52)^2 + … + (18 – 19.52)^2 + (26.3 – 24.26)^2 + … + (24.5 – 24.26)^2 + (20.6 – 22.84)^2 + … + (22.9 – 22.84)^2 + (25.4 – 21.16)^2 + … + (20.4 – 21.6)^2 = 97.504

SCE = $\sum_{j=1}^{k}(n_j - 1){s_j}^2 = 4\,(2.69)^2 + 4(1.92)^2 + 4(2.13)^2 + 4(2.98)^2 = 97.36$

$SCT = \sum_{j=1}^{c}\sum_{i=1}^{n_j}\left(x_{ij} - \bar{\bar{x}}\right)^2$= (18.5 – 21.945)^2 + (24 – 21.945)^2 + … + (20.4 – 21.945)^2 = 160.7895

ANALISIS DE VARIANZA

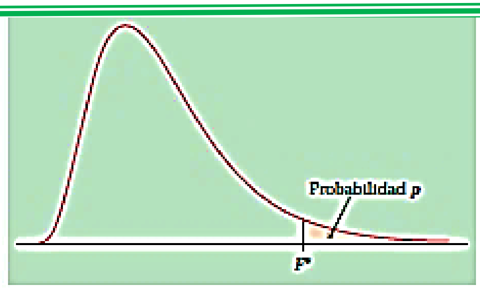
Fuente de Variación	Grados de libertad	Suma de Cuadrados	Cuadrados medios	F
Entre Grupos	3	SCTr = 63.2855	CMTr = 63.2855/3 = 21.095	21.095/6.094 = 3.46
Intra Grupos	16	SCE = 97.504	CME = 97.504/16 = 6.094	
TOTAL	19	SCT = 160.790		

ANALISIS DE VARIANZA: Valor de F

Valores críticos de la distribución F, $\alpha = 0.05$

Grados de libertad del denominador	Grados de libertad del numerador			
	5	6	7	8
1	230	234	237	239
2	19.3	19.3	19.4	19.4
3	9.01	8.94	8.89	8.85
4	6.26	6.16	6.09	6.04
5	5.05	4.95	4.88	4.82
6	4.39	4.28	4.21	4.15
7	3.97	3.87	3.79	3.73
8	3.69	3.58	3.50	3.44
9	3.48	3.37	3.29	3.23
10	3.33	3.22	3.14	3.07

ANALISIS DE VARIANZA Tabla de distribución F

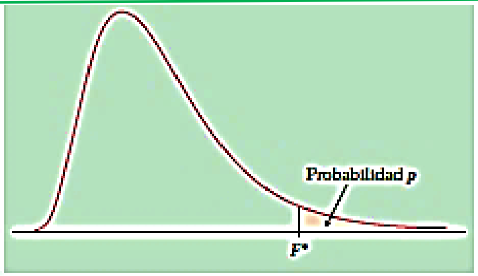


El valor de la tabla para p es el valor crítico F* que deja la probabilidad p a la derecha

TABLA D Valores críticos de la distribución F de Fisher

Grados de libertad en el denominador / Grados de libertad en el numerador

	p	1	2	3	4	5	6	7	8	9
	.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86
	.050	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
1	.025	647.79	799.50	864.16	899.58	921.85	937.11	948.22	956.66	963.28
	.010	4052.2	4999.5	5403.4	5624.6	5763.6	5859.0	5928.4	5981.1	6022.5
	.001	405284	500000	540379	562500	576405	585937	592873	598144	602284
	.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38
	.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38
2	.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39
	.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39
	.001	998.50	999.00	999.17	999.25	999.30	999.33	999.36	999.37	999.39
	.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24
	.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
3	.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47
	.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35
	.001	167.03	148.50	141.11	137.10	134.58	132.85	131.58	130.62	129.86
	.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94
	.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
4	.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90
	.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66
	.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47
	.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32
	.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
5	.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68
	.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16
	.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24
	.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96
	.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
6	.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52
	.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98
	.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	18.69
	.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72
	.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
7	.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82
	.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72
	.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	14.33



El valor de la tabla para p es el valor crítico F* que deja la probabilidad p a la derecha

TABLA D Valores críticos de la distribución F de Fisher (cont.)

Grados de libertad del numerador

10	12	15	20	25	30	40	50	60	120	1000
60.19	60.71	61.22	61.74	62.05	62.26	62.53	62.69	62.79	63.06	63.30
241.88	243.91	245.95	248.01	249.26	250.10	251.14	251.77	252.20	253.25	254.19
968.63	976.71	984.87	993.10	998.08	1001.4	1005.6	1008.1	1009.8	1014.0	1017.7
6055.8	6106.3	6157.3	6208.7	6239.8	6260.6	6286.8	6302.5	6313.0	6339.4	6362.7
605621	610668	615764	620908	624017	626099	628712	630285	631337	633972	636301
9.39	9.41	9.42	9.44	9.45	9.46	9.47	9.47	9.47	9.48	9.49
19.40	19.41	19.43	19.45	19.46	19.46	19.47	19.48	19.48	19.49	19.49
39.40	39.41	39.43	39.45	39.46	39.46	39.47	39.48	39.48	39.49	39.50
99.40	99.42	99.43	99.45	99.46	99.47	99.47	99.48	99.48	99.49	99.50
999.40	999.42	999.43	999.45	999.46	999.47	999.47	999.48	999.48	999.49	999.50
5.23	5.22	5.20	5.18	5.17	5.17	5.16	5.15	5.15	5.14	5.13
8.79	8.74	8.70	8.66	8.63	8.62	8.59	8.58	8.57	8.55	8.53
14.42	14.34	14.25	14.17	14.12	14.08	14.04	14.01	13.99	13.95	13.91
27.23	27.05	26.87	26.69	26.58	26.50	26.41	26.35	26.32	26.22	26.14
129.25	128.32	127.37	126.42	125.84	125.45	124.96	124.66	124.47	123.97	123.53
3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.80	3.79	3.78	3.76
5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.70	5.69	5.66	5.63
8.84	8.75	8.66	8.56	8.50	8.46	8.41	8.38	8.36	8.31	8.26
14.55	14.37	14.20	14.02	13.91	13.84	13.75	13.69	13.65	13.56	13.47
48.05	47.41	46.76	46.10	45.70	45.43	45.09	44.88	44.75	44.40	44.09
3.30	3.27	3.24	3.21	3.19	3.17	3.16	3.15	3.14	3.12	3.11
4.74	4.68	4.62	4.56	4.52	4.50	4.46	4.44	4.43	4.40	4.37
6.62	6.52	6.43	6.33	6.27	6.23	6.18	6.14	6.12	6.07	6.02
10.05	9.89	9.72	9.55	9.45	9.38	9.29	9.24	9.20	9.11	9.03
26.92	26.42	25.91	25.39	25.08	24.87	24.60	24.44	24.33	24.06	23.82
2.94	2.90	2.87	2.84	2.81	2.80	2.78	2.77	2.76	2.74	2.72
4.06	4.00	3.94	3.87	3.83	3.81	3.77	3.75	3.74	3.70	3.67
5.46	5.37	5.27	5.17	5.11	5.07	5.01	4.98	4.96	4.90	4.86
7.87	7.72	7.56	7.40	7.30	7.23	7.14	7.09	7.06	6.97	6.89
18.41	17.99	17.56	17.12	16.85	16.67	16.44	16.31	16.21	15.98	15.77
2.70	2.67	2.63	2.59	2.57	2.56	2.54	2.52	2.51	2.49	2.47
3.64	3.57	3.51	3.44	3.40	3.38	3.34	3.32	3.30	3.27	3.23
4.76	4.67	4.57	4.47	4.40	4.36	4.31	4.28	4.25	4.20	4.15
6.62	6.47	6.31	6.16	6.06	5.99	5.91	5.86	5.82	5.74	5.66
14.08	13.71	13.32	12.93	12.69	12.53	12.33	12.20	12.12	11.91	11.72

ANALISIS DE VARIANZA

TABLA D Valores críticos de la distribución F de Fisher (cont.)

Grados de libertad en el denominador / Grados de libertad en el numerador

Grados de libertad en el denominador	p	1	2	3	4	5	6	7	8	9
8	.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56
	.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
	.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36
	.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91
	.001	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.77
9	.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44
	.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18
	.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03
	.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35
	.001	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11
10	.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35
	.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02
	.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78
	.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94
	.001	21.04	14.91	12.55	11.28	10.48	9.93	9.52	9.20	8.96
11	.100	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27
	.050	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90
	.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59
	.010	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63
	.001	19.69	13.81	11.56	10.35	9.58	9.05	8.66	8.35	8.12
12	.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21
	.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80
	.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44
	.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39
	.001	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48
13	.100	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16
	.050	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71
	.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31
	.010	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19
	.001	17.82	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98
14	.100	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12
	.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65
	.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21
	.010	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03
	.001	17.14	11.78	9.73	8.62	7.92	7.44	7.08	6.80	6.58
15	.100	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09
	.050	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59
	.025	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12
	.010	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89
	.001	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26
16	.100	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06
	.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54
	.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05
	.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78
	.001	16.12	10.97	9.01	7.94	7.27	6.80	6.46	6.19	5.98
17	.100	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03
	.050	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49
	.025	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98
	.010	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68
	.001	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75

TABLA D Valores críticos de la distribución F de Fisher (cont.)

Grados de libertad en el numerador

Grados de libertad en el denominador	p	10	12	15	20	25	30	40	50	60	120	1000
8	.100	2.54	2.50	2.46	2.42	2.40	2.38	2.36	2.35	2.34	2.32	2.30
	.050	3.35	3.28	3.22	3.15	3.11	3.08	3.04	3.02	3.01	2.97	2.93
	.025	4.30	4.20	4.10	4.00	3.94	3.89	3.84	3.81	3.78	3.73	3.68
	.010	5.81	5.67	5.52	5.36	5.26	5.20	5.12	5.07	5.03	4.95	4.87
	.001	11.54	11.19	10.84	10.48	10.26	10.11	9.92	9.80	9.73	9.53	9.36
9	.100	2.42	2.38	2.34	2.30	2.27	2.25	2.23	2.22	2.21	2.18	2.16
	.050	3.14	3.07	3.01	2.94	2.89	2.86	2.83	2.80	2.79	2.75	2.71
	.025	3.96	3.87	3.77	3.67	3.60	3.56	3.51	3.47	3.45	3.39	3.34
	.010	5.26	5.11	4.96	4.81	4.71	4.65	4.57	4.52	4.48	4.40	4.32
	.001	9.89	9.57	9.24	8.90	8.69	8.55	8.37	8.26	8.19	8.00	7.84
10	.100	2.32	2.28	2.24	2.20	2.17	2.16	2.13	2.12	2.11	2.08	2.06
	.050	2.98	2.91	2.85	2.77	2.73	2.70	2.66	2.64	2.62	2.58	2.54
	.025	3.72	3.62	3.52	3.42	3.35	3.31	3.26	3.22	3.20	3.14	3.09
	.010	4.85	4.71	4.56	4.41	4.31	4.25	4.17	4.12	4.08	4.00	3.92
	.001	8.75	8.45	8.13	7.80	7.60	7.47	7.30	7.19	7.12	6.94	6.78
11	.100	2.25	2.21	2.17	2.12	2.10	2.08	2.05	2.04	2.03	2.00	1.98
	.050	2.85	2.79	2.72	2.65	2.60	2.57	2.53	2.51	2.49	2.45	2.41
	.025	3.53	3.43	3.33	3.23	3.16	3.12	3.06	3.03	3.00	2.94	2.89
	.010	4.54	4.40	4.25	4.10	4.01	3.94	3.86	3.81	3.78	3.69	3.61
	.001	7.92	7.63	7.32	7.01	6.81	6.68	6.52	6.42	6.35	6.18	6.02
12	.100	2.19	2.15	2.10	2.06	2.03	2.01	1.99	1.97	1.96	1.93	1.91
	.050	2.75	2.69	2.62	2.54	2.50	2.47	2.43	2.40	2.38	2.34	2.30
	.025	3.37	3.28	3.18	3.07	3.01	2.96	2.91	2.87	2.85	2.79	2.73
	.010	4.30	4.16	4.01	3.86	3.76	3.70	3.62	3.57	3.54	3.45	3.37
	.001	7.29	7.00	6.71	6.40	6.22	6.09	5.93	5.83	5.76	5.59	5.44
13	.100	2.14	2.10	2.05	2.01	1.98	1.96	1.93	1.92	1.90	1.88	1.85
	.050	2.67	2.60	2.53	2.46	2.41	2.38	2.34	2.31	2.30	2.25	2.21
	.025	3.25	3.15	3.05	2.95	2.88	2.84	2.78	2.74	2.72	2.66	2.60
	.010	4.10	3.96	3.82	3.66	3.57	3.51	3.43	3.38	3.34	3.25	3.18
	.001	6.80	6.52	6.23	5.93	5.75	5.63	5.47	5.37	5.30	5.14	4.99
14	.100	2.10	2.05	2.01	1.96	1.93	1.91	1.89	1.87	1.86	1.83	1.80
	.050	2.60	2.53	2.46	2.39	2.34	2.31	2.27	2.24	2.22	2.18	2.14
	.025	3.15	3.05	2.95	2.84	2.78	2.73	2.67	2.64	2.61	2.55	2.50
	.010	3.94	3.80	3.66	3.51	3.41	3.35	3.27	3.22	3.18	3.09	3.02
	.001	6.40	6.13	5.85	5.56	5.38	5.25	5.10	5.00	4.94	4.77	4.62
15	.100	2.06	2.02	1.97	1.92	1.89	1.87	1.85	1.83	1.82	1.79	1.76
	.050	2.54	2.48	2.40	2.33	2.28	2.25	2.20	2.18	2.16	2.11	2.07
	.025	3.06	2.96	2.86	2.76	2.69	2.64	2.59	2.55	2.52	2.46	2.40
	.010	3.80	3.67	3.52	3.37	3.28	3.21	3.13	3.08	3.05	2.96	2.88
	.001	6.08	5.81	5.54	5.25	5.07	4.95	4.80	4.70	4.64	4.47	4.33
16	.100	2.03	1.99	1.94	1.89	1.86	1.84	1.81	1.79	1.78	1.75	1.72
	.050	2.49	2.42	2.35	2.28	2.23	2.19	2.15	2.12	2.11	2.06	2.02
	.025	2.99	2.89	2.79	2.68	2.61	2.57	2.51	2.47	2.45	2.38	2.32
	.010	3.69	3.55	3.41	3.26	3.16	3.10	3.02	2.97	2.93	2.84	2.76
	.001	5.81	5.55	5.27	4.99	4.82	4.70	4.54	4.45	4.39	4.23	4.08
17	.100	2.00	1.96	1.91	1.86	1.83	1.81	1.78	1.76	1.75	1.72	1.69
	.050	2.45	2.38	2.31	2.23	2.18	2.15	2.10	2.08	2.06	2.01	1.97
	.025	2.92	2.82	2.72	2.62	2.55	2.50	2.44	2.41	2.38	2.32	2.26
	.010	3.59	3.46	3.31	3.16	3.07	3.00	2.92	2.87	2.83	2.75	2.68
	.001	5.58	5.32	5.05	4.78	4.60	4.48	4.33	4.24	4.18	4.02	3.87

ANALISIS DE VARIANZA

- Juego de Hipótesis:

Ho: $\mu_1 = \mu_2 = \mu_3 \ldots. = \mu_k$

Ha: $\mu_1 \neq \mu_2 \neq \mu_3 \ \ldots \neq \mu_k$

- Estadístico de prueba: $F_{Calc} = 3.46$
- Regla de decisión:

nivel de significancia .05

la hipótesis nula se rechaza si *F* (calculada) es mayor que *F* (en tablas)

$F_{tab} = 3.24$

Conclusión: Se rechaza La H_o.

ANALISIS DE VARIANZA

Una empresa que renta coches desea saber si existe diferencias significativas en el consumo de gasolina de 3 tipos de coches.

Automóvil A	Automóvil B	Automóvil C
25.1	23.9	26.6
24.7	23.7	25.4
26.0	24.4	25.8
24.3	23.3	24.4
23.9	23.6	24.2
24.2	24.5	25.4

ANALISIS DE VARIANZA

Fuente de Variación	Grados de libertad	Suma de Cuadrados	Cuadrados medios	F
Entre Grupos	3	SCTr =21.5495	CMTr = 10.7748	15.04
Intra Grupos	16	SCE =12.1800	CME = 0.7165	
TOTAL	19	SCT = **33.7295**		

ANALISIS DE VARIANZA

EJEMPLO.

Juego de Hipótesis:

Ho: $\mu_1 = \mu_2 = \mu_3 \ldots. = \mu_k$

Ha: $\mu_1 \neq \mu_2 \neq \mu_3 \ \ldots \neq \mu_k$

Estadístico de prueba: $F_{Calc} = 15.04$

Regla de decisión:

nivel de significancia .05

la hipótesis nula se rechaza si *F* (calculada) es mayor que *F* (en tablas)

$F_{tab} =$

Conclusión:

COMPARACIONES MÚLTIPLES

El análisis de varianza sólo permite concluir que las medias poblacionales no son iguales. En ocasiones se necesita determinar en dónde están las diferencias, específicamente **qué medias son las que difieren.**

Dentro de los intereses más comunes encontramos las pruebas por pares donde:

$H_0: \mu_i = \mu_j$ *para toda* $i \neq j$

$H_a: \mu_i \neq \mu_j$ para toda $i \neq j$

Varios procedimientos: Método de la diferencia mínima de Fisher (DMS), Prueba de Tukey, Prueba del rango múltiple de Duncan, Newman-Keuls, Hsu, etc.

ANALISIS DE VARIANZA

Prueba	Fórmula	Tablas
DMS	$t_{(\alpha/2,\ n-k)}\sqrt{\left[CMe\left(\frac{1}{ni}+\frac{1}{nj}\right)\right]}$	t de tablas de Student
Tukey	$Q_{(\alpha,k,n\text{-}k)}\sqrt{\frac{CMe}{2}\left(\frac{1}{ni}+\frac{1}{nj}\right)}$	Q de tablas de Tukey
Duncan	$= d_{(\alpha,\ p,\ n-k)}\sqrt{\left[CMe\left(\frac{1}{n}\right)\right]}$	Q_d de tablas de Duncan
Scheffé	$= S_{(\alpha,\ p,\ n-k)}\sqrt{\left[CMe\left(\frac{1}{n}\right)\right]}$	S de scheffé

PRUEBA t DE FISHER (DMS)

Juego de Hipótesis:

Ho: $\mu_i = \mu_j$

Ha: $\mu_i \neq \mu_j$

Estadístico de prueba: $\bar{x}_i - \bar{x}_j$

Regla de decisión: para

Rechazar H_0 si $\left|\bar{x}_i - \bar{x}_j\right| \geq DMS$

Con $\mathrm{DMS} = t_{\alpha/2}\sqrt{CME\left[\frac{1}{n_i} + \frac{1}{n_j}\right]}$

donde el valor de $t_{\alpha/2}$ se basa en la distribución t con n_T - k grados de libertad

PRUEBA DMS de Fisher

Ejemplo

	Proveedor 1	Proveedor 2	Proveedor 3
	58 64 55 66 67	58 69 71 64 68	48 57 59 47 49
Media Var Desv. Est.	62 27.5 5.244	66 26.5 5.148	52 31.0 5.568

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	*F*
Tratamientos	520	2	260.00	9.18
Error	340	12	28.33	
Total	860	14		

PRUEBA DMS de Fisher

Ejemplo

Proveedor	n	Media
1	5	62
2	5	66
3	5	52

Medias	Dif.	Signo	DMS
$\lvert\bar{x}_1 - \bar{x}_2\rvert$	4	$<$	7.34
$\lvert\bar{x}_1 - \bar{x}_3\rvert$	10	$>$	7.34
$\lvert\bar{x}_2 - \bar{x}_3\rvert$	14	$>$	7.34

$$\mathrm{DMS} = 2.179\sqrt{28.33\left[\frac{1}{5}+\frac{1}{5}\right]} = 7.34$$

Resultado

1. Se rechazar la H_o de que la media poblacional del número de unidades del proveedor 1 sea igual que la media poblacional del Proveedor 3.

2. Se rechazar la Ho de que la media poblacional del número de unidades del proveedor 2 sea igual que la media poblacional del Proveedor 3.

PRUEBA TUKEY

Juego de Hipótesis.

Ho: $\mu_i = \mu_j$

Ha: $\mu_i \neq \mu_j$

Estadístico de prueba: $Q = Qu\sqrt{\frac{MCE}{2}\left[\frac{1}{n_i} + \frac{1}{n_j}\right]}$

Regla de decisión:

para un nivel de significancia α, la hipótesis nula se rechaza si *la diferencia observada* es mayor que la *Q* (calculada) con grados de libertad en el numerador y en el denominador.

PRUEBA TUKEY

Proveedor	n	Media
1	5	19.52
2	5	24.26
3	5	22.84
4	5	21.16

$$\text{Rango crítico} = \text{Qu}\sqrt{\frac{MCE}{2}\left[\frac{1}{n_i} + \frac{1}{n_j}\right]}$$

$$= 4.05\sqrt{\frac{6.094}{2}\left[\frac{1}{5} + \frac{1}{5}\right]} = 4.471$$

Medias	Dif.	Signo	DHS
$\lvert\bar{x}_1 - \bar{x}_2\rvert$	4.74	>	4.47
$\lvert\bar{x}_1 - \bar{x}_3\rvert$	3.32	<	4.47
$\lvert\bar{x}_1 - \bar{x}_4\rvert$	1.64	<	4.47
$\lvert\bar{x}_2 - \bar{x}_3\rvert$	1.42	<	4.47
$\lvert\bar{x}_2 - \bar{x}_4\rvert$	3.10	<	4.47
$\lvert\bar{x}_3 - \bar{x}_4\rvert$	1.68	<	4.47

Resultado: Existe diferencias significativas entre los proveedores 1 y 2 únicamente.

Conclusión: Los materiales que abastecen los proveedores 1 y 2 son diferentes con un nivel de significancia del 5%.

PRUEBA TUKEY

Fuente de Variación	Grados de libertad	Suma de Cuadrados	Cuadrados medios	F
Entre grupos	3	210.478	70.159	**6.471**
Dentro grupos (Error)	14	151.800	10.843	
TOTAL	17	362.278		

PRUEBA TUKEY

Material	n	Media	s
1	5	68.20	3.114
2	3	62.00	2.646
3	4	60.50	3.697
4	6	68.39	3.406

$$\text{Rango crítico} = Qu\sqrt{\frac{MCE}{2}\left[\frac{1}{n_i}+\frac{1}{n_j}\right]}$$

$$= 4.11\sqrt{\frac{10.843}{2}\left[\frac{1}{n_i}+\frac{1}{n_j}\right]} = ?$$

Medias	**Dif.**	**Signo**	**DHS**
$\lvert\bar{x}_1 - \bar{x}_2\rvert$	6.2	<	6.988
$\lvert\bar{x}_1 - \bar{x}_3\rvert$	7.7	>	6.419
$\lvert\bar{x}_1 - \bar{x}_4\rvert$	.02	<	5.794
$\lvert\bar{x}_2 - \bar{x}_3\rvert$	1.5	<	7.308
$\lvert\bar{x}_2 - \bar{x}_4\rvert$	.6	<	6.766
$\lvert\bar{x}_3 - \bar{x}_4\rvert$	6.5	>	6.176

PRUEBA TUKEY: EJERCICIO

Material	n	Media	s^2
1	8	119	146.86
2	10	107	96.44
3	10	100	173.78

$$Q = Qu\sqrt{\frac{138.4}{2}\left[\frac{1}{n_i} + \frac{1}{n_j}\right]}$$

Medias	Dif.	Signo	DHS
$\lvert\bar{x}_1 - \bar{x}_2\rvert$			
$\lvert\bar{x}_1 - \bar{x}_3\rvert$			
$\lvert\bar{x}_2 - \bar{x}_3\rvert$			

PRUEBA DE LA SUMA DE RANGO DE WILCOXON

Prueba de rangos con signo de Wilcoxon. Es un procedimiento no paramétrico para el análisis de datos de un experimento de muestras pareadas que no requiere la suposición de normalidad.

Se supone que $n_1 \leq n_2$ y se clasifican las observaciones $n_1 + n_2$ en orden de magnitud ascendente asignándoles rangos. Si dos o más observaciones se unen o igualan (idénticas), se emplea la media de los rangos que se habría asignado si las observaciones hubieren diferido. Sea R_1 la suma de los rangos en la muestra más pequeña, y se define $R_2 = n_1(n_1 + n_2 + 1) - R_1$

Cuando tanto n_1 y n_2 son mayores que 8, la distribución de R, puede aproximarse bien mediante la distribución normal con media $\mu_{R_1} = \frac{n_1(n_1+n_2+1)}{2}$ y varianza $\sigma^2_{R_1} = \frac{n_1 n_2(n_1+n_2+1)}{12}$.

En consecuencia, se puede emplear $Z_0 = \frac{R_1-\mu_1}{\sigma_R}$ como una estadística de prueba, y la apropiada región critica de $|Z_0| > Z_\alpha$ para la prueba de hipótesis.

PRUEBA DE LA SUMA DE RANGO DE WILCOXON

Juego de Hipótesis:

H_0: las dos poblaciones son idénticas

H_a: las dos poblaciones no son idénticas

Estadístico de prueba: $Z_0 = \dfrac{R_1 - \frac{n_1(n_1+n_2+1)}{2}}{\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}} = Z_0 = \dfrac{R_1-\mu_1}{\sigma_R}$

Regla de decisión:

a) Rechazar H_0 Si $Z_0 > Z_\alpha$. Prueba en la cola derecha.

b) Rechazar H_0 Si $Z_0 < -Z_\alpha$. Prueba en la cola izquierda.

c) Rechazar H_0 Si $Z_0 > Z_{\alpha/2}$ 0 $Z_0 < -Z_{\alpha/2}$. Prueba de dos colas

Con α nivel de significancia.

PRUEBA DE LA SUMA DE RANGO DE WILCOXON

Ejemplo.

El interés principal es determinar si hay más personas que no se presentan a tomar los vuelos que salen de la ciudad A en comparación con vuelos que salen de la ciudad C. Una muestra de ocho vuelos de A y nueve de C aparece enseguida.

Ciudad A	13	14	10	8	16	9	17	21		$n_1 = 8$
Ciudad C	11	15	10	18	11	20	24	22	25	$n_2 = 9$

Valores	8	9	10	10	11	11	13	14	15	16	17	18	20	21	22	24	25	Suma
Rango A	1	2	3.5				7	8		10	11			14				56.5
Rango C				3.5	5.5	5.5			9			12	13		15	16	17	96.5

$$\text{media} = \mu_{R_1} = \frac{n_1(n_1+n_2+1)}{2} = \frac{8(8+9+1)}{2} = \frac{144}{2} = 72$$

$$\text{Varianza} = \sigma_{R_1}^2 = \frac{n_1 n_2(n_1+n_2+1)}{12} = \frac{(8)(9)(8+9+1)}{12} = 108.$$

PRUEBA DE LA SUMA DE RANGO DE WILCOXON

Ejemplo.

Juego de Hipótesis:

H_0: las dos poblaciones son idénticas

H_a: las dos poblaciones no son idénticas

Estadístico de prueba: $Z_0 = \frac{56-72}{10.39} = \frac{-15.5}{10.39} = -1.492$

Regla de decisión: Rechazar H_0 Si $Z_0 < -Z_\alpha$.

Con $\alpha = .05$, $Z_\alpha = 1.645$.

Por lo que $Z_0 = -1.492 > -Z_\alpha = -1.645$, no se cumple la regla de decisión.

No se rechaza H_0, Las dos poblaciones son idénticas. No hay una diferencia entre los números habituales de personas que no se presentaron en la ciudad A y en la C.

PRUEBA DE LA SUMA DE RANGO DE WILCOXON

Ejercicio. Con los datos de la tabla siguiente, utilice un nivel de significancia de 0.05 para probar la aseveración de que la mediana del IMC de los hombres es igual a la mediana del IMC de las mujeres.

Mediciones del IMC	
Hombres	Mujeres
23.8 (11.5)	19.6 (2.5)
23.2 (9)	23.8 (11.5)
24.6 (14)	19.6 (2.5)
26.2 (17)	29.1 (22)
23.5 (10)	25.2 (15.5)
24.5 (13)	21.4 (5)
21.5 (6)	22.0 (7)
31.4 (24)	27.5 (19)
26.4 (18)	33.5 (25)
22.7 (8)	20.6 (4)
27.8 (20)	29.9 (23)
28.1 (21)	17.7 (1)
25.2 (15.5)	
$n_1 = 13$	$n_2 = 12$
$R_1 = 187$	$R_2 = 138$

$$\text{media} = \frac{n_1(n_1+n_2+1)}{2} = 156$$

$$\text{Var} = \frac{n_1 n_2(n_1+n_2+1)}{12} = 338$$

$$Z_0 = \frac{R_1-\mu_1}{\sigma_R} = -0.979$$

PRUEBA DE KRUSKALL–WALLIS

PRUEBA DE KRUSKALL–WALLIS. Este procedimiento no paramétrico requiere al menos tres muestras independientes, las cuales se seleccionan al azar y no se necesita la suposición de normalidad. La hipótesis nula es que las poblaciones son idénticas.

1. Se clasifican las observaciones n1 + n2 en orden de magnitud ascendente asignándoles rangos.
2. En cada muestra, se calcula la suma de los rangos y el tamaño muestral.
3. Se Calcula el estadístico de prueba. $H = \left[\frac{12}{n_t(n_t+1)}\sum\frac{R_i^2}{n_i}\right] - 3(n_t + 1)$

n_i = número de observaciones en la muestra i
n_t = número total de observaciones en todas las muestras
R_i = suma de los rangos para la muestra i

La prueba es de cola derecha y el estadístico de prueba H puede aproximarse por medio de una distribución chi cuadrada con k - 1 grados de libertad, donde k es el número de muestras diferentes. La regla de decisión es Rechazar H_0 Si $\mathrm{H} \geq \chi^2{}_\alpha$ con gl = k-1 y α nivel de significancia.

PRUEBA DE KRUSKALL–WALLIS

Los empleados de un equipo gerencial que provienen de tres universidades. El director intenta determinar si existen diferencias en el desempeño de los empleados de acuerdo con la universidad de procedencia. Se cuenta con los datos para muestras independientes de 7 gerentes que se graduaron en la universidad A, 6 que provienen de la universidad B y 7 que egresaron de la universidad C.

Universidad		
A	B	C
25	60	50
70	20	70
60	30	60
85	15	80
95	40	90
90	35	70
80		75

Universidad					
A	Rango	B	Rango	C	Rango
25	3	60	9	50	7
70	12	20	2	70	12
60	9	30	4	60	9
85	17	15	1	80	15.5
95	20	40	6	90	18.5
90	18.5	35	5	70	12
80	18.5			75	14
	$\sum = 95$		$\sum = 27$		$\sum = 88$

PRUEBA DE KRUSKALL–WALLIS

Juego de Hipótesis:

H_0: Las dos poblaciones son idénticas

H_a: Las dos poblaciones no son idénticas

Estadístico de prueba: $H = \left[\frac{12}{n_t(n_t+1)} \sum \frac{R_i^2}{n_i}\right] - 3(n_t + 1)$

$$H = \left[\left(\frac{12}{20(21)}\right)\left(\frac{95^2}{7} + \frac{27^2}{6} + \frac{88^2}{7}\right)\right] - 3(21) = 8.92$$

Regla de decisión: Rechazar H_0 Si $\mathrm{H} \geq \chi^2$

Con $\alpha = .05$, $k = 3$, $\chi^2 = 5.991$ por lo que $8.92 \geq 5.991$

Resultado: H_0 es rechazada y se concluye que las tres poblaciones no son iguales.

PRUEBA DE KRUSKALL–WALLIS

Ejercicio. Se recolectaron las calificaciones del examen de aprovechamiento para cuatro diferentes grupos de estudiantes, Realice la prueba usando la prueba *H* de Kruskal-Wallis con α = .05 para probar la hipótesis de que no hay diferencia, en las distribuciones poblacionales de las calificaciones del examen de aprovechamiento.

1	2	3	4
65 (3)	75 (9)	59 (1)	94 (23)
87 (19)	69 (5.5)	78 (11)	89 (21)
73 (8)	83 (17.5)	67 (4)	80 (14)
79 (12.5)	81 (15.5)	62 (2)	88 (20)
81 (15.5)	72 (7)	83 (17.5)	
69 (5.5)	79 (12.5)	76 (10)	
	90 (22)		
$R_1 = 63.5$	$R_2 = 89$	$R_3 = 45.5$	$R_4 = 78$

BIBLIOGRAFIA

1. Anderson, D. R., Sweeney, D. J. y Williams, T. A. Estadística para Administración y Economía. Décima edición. Cengage Editores. México. 2008.

2. Fernández. A. C. Manual de estadística descriptiva aplicada al sector turístico. Ed. Síntesis. España. 2001.

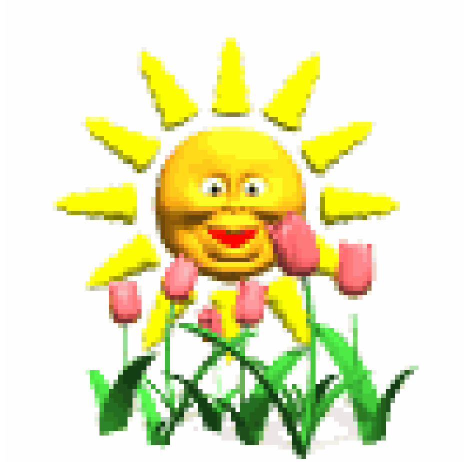
3. Ferran M. SPSS para Windows. Análisis Estadístico. Ed Mc Graw Hill. México 2001

4. Infante, S. G. y Zárate de L. G. Métodos Estadísticos. Ed. Trillas. México. 2000.

5. Levine, D. M., Krehbiel, T. C. y Berenson, M. L. Estadística para administración. Cuarta edición. Pearson. México. 2006

6. Lind Douglas A.,Marchal William G.,Wathen Samuel A. . Estadística aplicada a los negocios la economía. Décimo Tercera edición .Mc Graw Hill 2008.

7. Riquelme P. Tablas y Gráficos en investigaciones. 2004.



FIN DE LA PRESENTACION