UNIVERSIDAD DE SAN PEDRO SULA

2. Análisis de varianza

Introducción

La estadística inferencial no solo realiza estudios con una muestra, también es necesario trabajar con más de una muestra; las que pueden ser dos o más.

Para cada una de las muestras, vistas por separado, es posible calcular la media aritmética y la varianza; por lo que los análisis, se pueden revisar en conjunto para tomar decisiones. Las muestras pueden ser independientes o dependientes.

Distribución F

A menudo se necesita probar si dos poblaciones independientes tienen la misma variabilidad. Para estos casos, la mejor opción es la *Distribución F*.

Para probar la hipótesis de dos varianzas se necesitan dos muestras. En este caso, existe la posibilidad de que las medias aritméticas sean iguales; pero, la variación que hay entre ambas sea diferente.

La distribución F es una familia de distribuciones con las siguientes características:

1. Continua
2. No puede ser negativa
3. Tiene sesgo positivo
4. Es asintótica

El estadístico de prueba para comparar dos varianzas es la Distribución F cuya fórmula se muestra a continuación:

$$F = \frac{s_1^2}{s_2^2}$$

Sin importar si se desea determinar si una población varía más que otra o validar una suposición de una prueba estadística, primero se formula la hipótesis nula asumiendo que ambas son iguales. El formato general de la hipótesis se formula de la siguiente manera:

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_a: \sigma_1^2 \neq \sigma_2^2$$

Para realizar la prueba, se utilizan dos muestras aleatorias con tamaños que pueden ser diferentes. Para la prueba se siguen los 5 pasos enunciados en el capítulo anterior.

UNIVERSIDAD DE SAN PEDRO SULA

Localización de valor crítico en la Distribución F

La distribución F contiene valores diferentes según el nivel de significancia que se elija; los más comunes son 0.05 y 0.01. Localizar el

Los pasos para localizar un valor de F son los siguientes:

1. Determinar el tamaño de cada muestra.
2. Calcular los grados de libertad (n-1)
3. Localizar los grados de libertad (gl) de la primera muestra en la primera línea de la tabla.
4. Localizar los grados de libertad (gl) de la segunda muestra en la primera columna
5. Ubicar la posición en que se cruzan ambos grados de libertad (gl).

Un resumen de una distribución F para un nivel de significancia de 0.05 se muestra a continuación:

Grados de libertad para el denominador	Grados de libertad para el numerador											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**
1	161	200	216	225	230	234	237	239	241	242	244	246
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20

Para un nivel de significancia de 0.01 la distribución F es la siguiente:

Grados de libertad para el denominador	Grados de libertad para el numerador											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157
2	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4
3	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	26.9
4	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.2
5	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.89	9.72
6	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56
7	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31
8	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52
9	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96
10	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.80	3.66
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09

UNIVERSIDAD DE SAN PEDRO SULA

Ejemplo 2.1

1. Encontrar el valor crítico con un nivel de significancia de 0.01 para dos muestras con tamaño 11 y 8 respectivamente.

Desarrollo

OPCIÓN 1	OPCIÓN 2
$n_1 = 11$ $n_2 = 8$ $gl_1 = 11 - 1 = 10$ $gl_2 = 8 - 1 = 7$ $F = 6.62$	$n_1 = 8$ $n_2 = 11$ $gl_1 = 8 - 1 = 7$ $gl_2 = 11 - 1 = 10$ $F = 5.20$

2. Encontrar el valor crítico con un nivel de significancia de 0.05 para dos muestras con tamaño 11 y 8 respectivamente.

Desarrollo

OPCIÓN 1	OPCIÓN 2
$n_1 = 11$ $n_2 = 8$ $gl_1 = 11 - 1 = 10$ $gl_2 = 8 - 1 = 7$ $F = 3.64$	$n_1 = 8$ $n_2 = 11$ $gl_1 = 8 - 1 = 7$ $gl_2 = 11 - 1 = 10$ $F = 3.14$

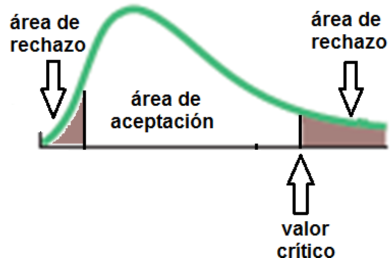
El valor crítico es modificado dependiendo de cuál es la cola que se va analizar. En el caso de la comparación entre dos varianzas, se utiliza la cola de la derecha.

Para determinar la regla de decisión (Paso 4), se evalúa la hipótesis nula; si es igualdad (=) es de dos colas, en caso contrario es de una cola. El proceso estándar para realizar la prueba de hipótesis se enuncia de la siguiente manera:

1. Una vez calculada la varianza se las muestras, la que tiene mayor valor se convierte en la muestra #1 y la de menor valor la muestra #2.
2. Se formula la hipótesis nula y la alternativa (Paso 1)
3. Se selecciona el nivel de significancia como un total (Paso 2)
4. Se evalúa si la hipótesis nula es una igualdad (=) para considerarla de dos colas; en caso contrario, solo es de una cola. (Paso 3)
5. Si el nivel de significancia es a dos colas; éste se divide entre 2
6. Determinar el tamaño de cada muestra
7. Calcular los grados de libertad (tamaño de muestra – 1)
8. Determinar la tabla de Distribución F que se debe utilizar (0.05 ó 0.01)
9. Localizar en la tabla de la Distribución F los grados de libertad de la muestra #1 en la primera fila y los grados de libertad de la muestra #2 en la primera columna y encontrar el valor crítico.
10. Visualizar en la representación gráfica de la distribución F la posición del valor crítico que separa el área de aceptación y el área de rechazo.
11. Calcular el valor de F con los datos de las varianzas de las 2 muestras (La de mayor valor representa la cola de la derecha).

USAP
UNIVERSIDAD DE SAN PEDRO SULA

12. Compara el valor F resultante con el valor crítico y determinar si la hipótesis nula se acepta o se rechaza.



Ejemplo 2.2

1. Suponer que se tiene una muestra de tamaño 8 de una población con distribución normal, con varianza muestral de 56.0; además se tiene otra muestra de tamaño 10 de una población con distribución normal, con varianza muestral de 24. Utilizar el nivel de significancia 0.10 para probar que no hay diferencia en las dos varianzas poblacionales contra la alternativa de que sí existe evidencia de una diferencia significativa en las varianzas poblacionales.

Desarrollo

DATOS INICIALES

Definir cuál es la varianza de la muestra #1 y cuál la #2 (varianza con mayor valor es #1).

$$S_1^2 = 56$$
$$S_2^2 = 24$$

PRUEBA DE HIPÓTESIS

Paso 1: Formular la hipótesis nula y alternativa

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_a: \sigma_1^2 \neq \sigma_2^2$$

Paso 2: Seleccionar el nivel de significancia

$$\alpha = 0.10$$

Paso 3: Determinar el estadístico de prueba

$$F = \frac{S_1^2}{S_2^2}$$

Paso 4: Formular la regla de decisión

- La hipótesis es una igualdad

$$H_0: \sigma_1^2 = \sigma_2^2$$

- El nivel de significancia es de 2 colas

$$\alpha = 0.10/2 = 0.05$$

- Determinar el tamaño de las muestras

$$n_1 = 8$$
$$n_2 = 10$$

- Calcular los grados de libertad

$$gl_1 = 8 - 1 = 7$$
$$gl_2 = 10 - 1 = 9$$

UNIVERSIDAD DE SAN PEDRO SULA

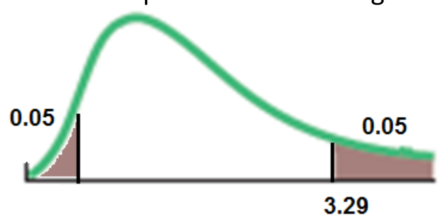
- Determinar el valor crítico en la tabla de la distribución F de 0.05 de nivel de significancia.

	Grados de						
	1	2	3	4	5	6	7
1	161	200	216	225	230	234	237
2	18.5	19.0	19.2	19.2	19.3	19.3	19.4
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29

ador

Valor crítico: F=3.29

- Visualizar la posición 3.29 en la gráfica y definir el área de aceptación y rechazo.

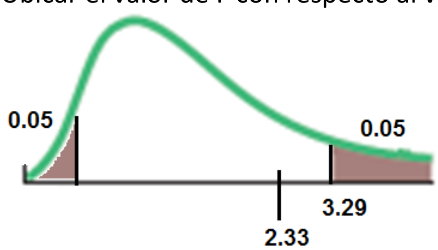


Paso 5: Tomar la decisión

- Calcular el valor de F con las varianzas de las muestras

$$\boldsymbol{F = \frac{s_1^2}{s_2^2} = \frac{56}{24} = 2.33}$$

- Ubicar el valor de F con respecto al valor crítico.



La hipótesis nula se acepta.
No hay evidencia para concluir que exista diferencia en la variación de ambas muestras.

2. Productos Eléctricos Steele, ubicada en el Zip Constantine, ensambla componentes eléctricos para teléfonos celulares. Durante los últimos 10 días el Turno A del departamento de Control de Despachos ha promediado 9 productos rechazados, con una desviación estándar de 2 rechazos por día. El Turno B promedió 8.5 productos rechazados, con una desviación estándar de 1.5 rechazos durante el mismo periodo. Con un nivel de significancia de 0.05, ¿podría concluir que hay más variación en el número de productos rechazados por día en la muestra del Turno A?

<u>DATOS INICIALES</u>

Muestra del turno A: s = 2
Muestra del turno B: s = 1.5

$$s_1^2 = 4$$
$$s_2^2 = 2.25$$

PRUEBA DE HIPÓTESIS

Paso 1: Formular la hipótesis nula y alternativa

$$H_0: \sigma_1^2 \leq \sigma_2^2$$
$$H_a: \sigma_1^2 > \sigma_2^2$$

Paso 2: Seleccionar el nivel de significancia

$$\alpha = 0.05$$

Paso 3: Determinar el estadístico de prueba

$$F = \frac{S_1^2}{S_2^2}$$

Paso 4: Formular la regla de decisión

- La hipótesis es una desigualdad

$$H_0: \sigma_1^2 \leq \sigma_2^2$$
$$1\ cola$$

- El nivel de significancia es de 1 cola

$$\alpha = 0.05$$

- Determinar el tamaño de las muestras

- Calcular los grados de libertad

$$n_1 = 10$$
$$n_2 = 10$$

- Calcular los grados de libertad

$$gl_1 = 10 - 1 = 9$$
$$gl_2 = 10 - 1 = 9$$

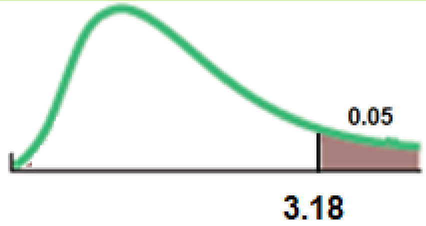
- Determinar el valor crítico en la tabla de la distribución F de 0.05 de nivel de significancia.

		Grados de libertad para el								
		1	2	3	4	5	6	7	8	9
	1	161	200	216	225	230	234	237	239	241
	2	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4
	3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81
	4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00
	5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77
	6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10
	7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68
	8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39
ador	9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18

Valor crítico: F=3.18

- Visualizar la posición 3.18 en la gráfica y definir el área de aceptación y rechazo.

USAP
UNIVERSIDAD DE SAN PEDRO SULA

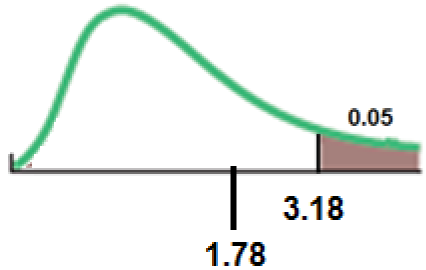


Paso 5: Tomar la decisión

- Calcular el valor de F con las varianzas de las muestras

$$F = \frac{s_1^2}{s_2^2} = \frac{4}{2.25} = 1.78$$

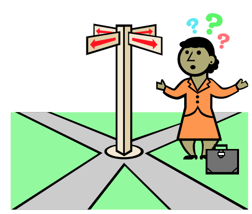
- Ubicar el valor de F con respecto al valor crítico.



H_0 se acepta

No hay suficiente evidencia para determinar que la variación en el turno A es mayor que en el Turno B.

3. En una agencia de servicio de Taxis, del centro de la ciudad al aeropuerto internacional, utilizan dos rutas para llegar, la autopista y la carretera alterna. La distancia que recorre el taxi desde el centro de la ciudad al aeropuerto es mayor que la que se recorre por la carretera alterna; sin embargo, las condiciones de la carretera alterna son deficientes y aunque es más corta, se llega casi al mismo tiempo. El Gerente de la empresa desea estudiar el tiempo que se tarda en conducir por cada una de las rutas y luego comparar los resultados, usando un nivel de significancia de 0.10.



Se recopiló una muestra del tiempo en minutos que tarda un taxi en llegar hasta el aeropuerto de cada una de las rutas y se obtuvieron los siguientes resultados:

Autopista	59	60	61	51	56	63	57	65
Carretera alterna	52	67	56	45	70	54	64	

¿Hay alguna diferencia entre las variaciones de los tiempos del manejo de las dos rutas?

Desarrollo

DATOS INICIALES

Previo a generar los 5 pasos de la prueba de hipótesis, calcular la media aritmética y la varianza de cada una de las muestras (2).

ESTADÍSTICO		autopista	carretera alterna
Media aritmética	$\bar{X} = \frac{\sum X_i}{n}$	$\bar{X} = \frac{472.0}{8} = 59.0$	$\bar{X} = \frac{408.0}{7} = 58.3$
Varianza	$s^2 = \frac{\sum (X - \bar{X}_1)^2}{n-1}$	$s^2 = \frac{134}{8-1} = 19.1$	$s^2 = \frac{485.43}{7-1} = 80.9$
Desviación estándar	$s = \sqrt{s^2}$	$s = \sqrt{9.1}$ =4.38	$s = \sqrt{80.9}$ =8.99
		MUESTRA 2	MUESTRA 1

Las medias aritmética tienen un comportamiento similar, en ambas pistas se tardan un promedio de 58.5 segundos en llegar al aeropuerto; sin embargo, la variabilidad que muestra la desviación estándar es alta en la carretera alterna (la desviación estándar de la autopista es casi la mitad de la de la carretera alterna). Con este resultado, la prueba de la hipótesis se hará con la comparación de las varianzas para comprobar si son similares (aunque su diferencia sea alta).

PRUEBA DE HIPÓTESIS

Paso 1: Formular la hipótesis nula y alternativa

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_a: \sigma_1^2 \neq \sigma_2^2$$

Paso 2: Seleccionar el nivel de significancia

$$\alpha = 0.10$$

Paso 3: Determinar el estadístico de prueba

$$F = \frac{S_1^2}{S_2^2}$$

Paso 4: Formular la regla de decisión

- La hipótesis es una igualdad
$$H_0: \sigma_1^2 = \sigma_2^2$$
- El nivel de significancia es de 2 colas
$$\alpha = 0.10/2 = 0.05$$
- Determinar el tamaño de las muestras
$$n_1 = 7$$
$$n_2 = 8$$
- Calcular los grados de libertad
$$gl_1 = 7 - 1 = 6$$
$$gl_2 = 8 - 1 = 7$$
- Determinar el valor crítico en la tabla de la distribución F con nivel de significancia 0.05

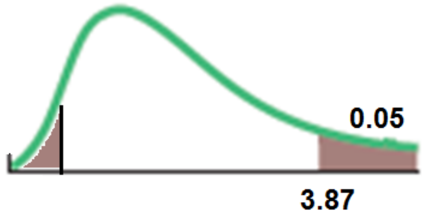
	1	2	3	4	5	6
1	161	200	216	225	230	234
2	18.5	19.0	19.2	19.2	19.3	19.3
3	10.1	9.55	9.28	9.12	9.01	8.94
4	7.71	6.94	6.59	6.39	6.26	6.16
5	6.61	5.79	5.41	5.19	5.05	4.95
6	5.99	5.14	4.76	4.53	4.39	4.28
7	5.59	4.74	4.35	4.12	3.97	3.87

En primera fila se busca gl_1 y en primera columna gl_2.

UNIVERSIDAD DE SAN PEDRO SULA

$$F = 3.87$$

- Visualizar la posición 3.87 en la gráfica de la distribución F.

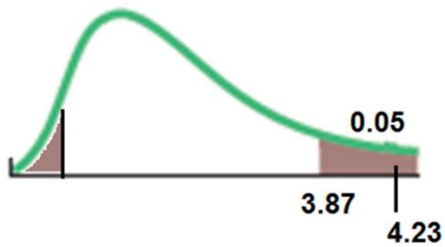


Paso 5: Tomar la decisión

- Calcular el valor de F con las varianzas de las muestras

$$F = \frac{s_1^2}{s_2^2} = \frac{80.9}{19.1} = 4.23$$

- Ubicar el valor de F con respecto al valor crítico.
  El valor 4.23 es mayor que 3.87, el valor F cae en la zona de rechazo.



La hipótesis nula se rechaza.
Se concluye que sí hay una diferencia entre las variaciones de los tiempos de recorrido por las dos rutas.

La prueba ANOVA

Para la prueba de hipótesis con una o dos muestras que están normalmente distribuidas se utilizan las distribuciones Gauss y la t-Student; sin embargo, a menudo se necesitan hacer comparaciones para más de dos medias y para ello se utilizan la metodología del análisis de varianza (ANOVA), que recurre a la distribución F.

Para analizar varias muestras a la vez, se utilizan dos elementos básicos que son:

1. Variación de tratamiento
2. Variación aleatoria

La fuerza de esta metodología radica en que analiza las muestras por separado y en conjunto para obtener un valor promedio que garantice su eficacia.

Aunque la hipótesis es planteada a través de la media aritmética de las muestras, el análisis se realiza a través de la varianza, de allí su nombre.

Los pasos iniciales del proceso consisten en calcular la media aritmética global de las muestran en conjunto y la media aritmética de cada muestra.

USAP
UNIVERSIDAD DE SAN PEDRO SULA

Variación de tratamiento (VT)

Es la diferencia de la media de cada muestra y la media global elevada al cuadrado, el mismo número de veces de cada dato de cada muestra.

$$VT = n_1 \sum (\bar{X}_{m_1} - \bar{X}_g)^2 + n_2 \sum (\bar{X}_{m_2} - \bar{X}_g)^2 + \dots n_k \sum (\bar{X}_{m_k} - \bar{X}_g)^2$$

> "VARIACIÓN DE TRATAMIENTO: Suma de las diferencias entre la media de cada tratamiento y la media global elevada al cuadrado." (Lind |Marchal |Wathen, 2008, p.331).

A cada dato se le coloca la media que le corresponde y luego ese valor se resta con la media global y el resultado se eleva al cuadrado. Esa resta es el mismo número de veces que se tenga una observación en la muestra.

Ejemplo 2.3

1. El gerente de un centro financiero regional desea comparar la productividad, medida por el número de clientes atendidos, de 3 de sus empleados. Selecciona 4 días en forma aleatoria y registra el número de clientes que atendió cada empleado. Los resultados obtenidos fueron:

LOBO	BLANCO	CÓRDOVA
55	66	47
54	76	51
59	67	46
56	71	48

   Calcular la variación de tratamiento.

   Desarrollo

   a. Convertir las 3 muestras como si fuera 1 y calcular la media aritmética de cada muestra y la media aritmética global.

UNIDAD	X_i	$\bar{X}_m$	$\bar{X}_g$
LOBO	55	56	58
	54		
	59		
	56		
BLANCO	66	70	
	76		
	67		
	71		
CÓRDOVA	47	48	
	51		
	46		
	48		

   b. Para cada dato de cada muestra, calcular el cuadrado de la diferencia entre la media de la muestra y la media global.

USAP
UNIVERSIDAD DE SAN PEDRO SULA

UNIDAD	X_i	$\bar{X}_m$	$\bar{X}_g$	Variación Tratamiento	
LOBO	55	56	58	$(56 - 58)^2 =$	4
	54			$(56 - 58)^2 =$	4
	59			$(56 - 58)^2 =$	4
	56			$(56 - 58)^2 =$	4
BLANCO	66	70		$(70 - 58)^2 =$	144
	76			$(70 - 58)^2 =$	144
	67			$(70 - 58)^2 =$	144
	71			$(70 - 58)^2 =$	144
CÓRDOVA	47	48		$(48 - 58)^2 =$	100
	51			$(48 - 58)^2 =$	100
	46			$(48 - 58)^2 =$	100
	48			$(48 - 58)^2 =$	100
			Σ		992

La variación de tratamiento es 992

Variación de aleatoria (VA)

Es la diferencia de cada dato de cada muestra y la media muestral respectiva, elevada al cuadrado.

$$VA = \sum (X_i - \bar{X}_{m_1})^2 + \sum (X_j - \bar{X}_{m_2})^2 + \ldots \sum (X_z - \bar{X}_{m_k})^2$$

"VARIACIÓN ALEATORIA: Suma de las diferencias entre cada observación y su media de tratamiento elevada al cuadrado." (Lind |Marchal |Wathen, 2008, p.331).

Ejemplo 2.4

1. El gerente de un centro financiero regional desea comparar la productividad, medida por el número de clientes atendidos, de 3 de sus empleados. Selecciona 4 días en forma aleatoria y registra el número de clientes que atendió cada empleado. Los resultados obtenidos fueron:

LOBO	BLANCO	CÓRDOVA
55	66	47
54	76	51
59	67	46
56	71	48

Calcular la variación aleatoria.

Desarrollo

El cuadro ya construido, tomar como base la columna del dato y la columna de las medias. Para cada dato de cada muestra, calcular el cuadrado de la diferencia entre el dato y la media de la muestra.

USAP
UNIVERSIDAD DE SAN PEDRO SULA

UNIDAD	X_i	$\bar{X}_m$	Variación Aleatoria	
LOBO	55	56	$(55 - 56)^2 =$	1
	54		$(54 - 56)^2 =$	4
	59		$(59 - 56)^2 =$	9
	56		$(56 - 56)^2 =$	0
BLANCO	66	70	$(66 - 70)^2 =$	16
	76		$(76 - 70)^2 =$	36
	67		$(67 - 70)^2 =$	9
	71		$(71 - 70)^2 =$	1
CÓRDOVA	47	48	$(47 - 48)^2 =$	1
	51		$(51 - 48)^2 =$	9
	46		$(46 - 48)^2 =$	4
	48		$(48 - 48)^2 =$	0
			Σ	90

La variación aleatoria es 90

Estadístico de prueba

El estadístico de prueba para la ANOVA sigue siendo las varianzas; una de ellas es la división entre la variación de tratamiento y los grados de libertad del total de muestras en análisis; la otra es la división entre la variación aleatoria y los grados de libertad con relación a todas las muestras.

$k-1$: Número de muestras menos 1
$n-k$: Total de datos en análisis menos el total de muestras.

$$F = \frac{\begin{array}{c}\textit{Estimación de la varianza poblacional}\\ \textit{basada en las diferencias entre medias muestrales}\end{array}}{\begin{array}{c}\textit{Estimación de la varianza poblacional}\\ \textit{basada en la variación dentro de la muestra}\end{array}} = \frac{\frac{VT}{k-1}}{\frac{VA}{n-k}}$$

Para hacer la labor más dinámica, nada como una tabla que nos permite reducir el tiempo de trabajo utilizando la siguiente tabla:

Resumen de la tabla ANOVA

Variación	Σ^2	n	gl	Estimación Varianza	F
Tratamiento	VT	Muestras	k - 1	$\frac{VT}{k-1}$	$\frac{\frac{VT}{k-1}}{\frac{VA}{n-k}}$
Aleatoria	VA	Total datos	n - k	$\frac{VA}{n-k}$	

USAP
UNIVERSIDAD DE SAN PEDRO SULA

Ejemplo 2.5

1. El gerente de un centro financiero regional desea comparar la productividad, medida por el número de clientes atendidos, de 3 de sus empleados. Selecciona 4 días en forma aleatoria y registra el número de clientes que atendió cada empleado. Los resultados obtenidos fueron:

LOBO	BLANCO	CÓRDOVA
55	66	47
54	76	51
59	67	46
56	71	48

El cálculo de la variación de tratamiento es 992 y la variación aleatoria es 90. Calcular el valor de F.

Desarrollo

Resumen de la tabla ANOVA

Variación	Σ^2	n	gl	Estimación Varianza	F
Tratamiento	992.0	3	2	496.0	49.60
Aleatoria	90.0	12	9	10.0	

$$\boldsymbol{F=\frac{\frac{VT}{k-1}}{\frac{VA}{n-k}}=\frac{\frac{992}{3-1}}{\frac{90}{12-3}}=\frac{\frac{992}{2}}{\frac{90}{9}}=49.6}$$

PRUEBA DE HIPÓTESIS

Para realizar la prueba de hipótesis, se sigue utilizando el mismo concepto. Si se trata de resolver el caso del centro financiero regional, el resumen sería el siguiente:

Ejemplo 2.6

1. El gerente de un centro financiero regional desea comparar la productividad, medida por el número de clientes atendidos, de 3 de sus empleados. Selecciona 4 días en forma aleatoria y registra el número de clientes que atendió cada empleado. Los resultados obtenidos fueron:

LOBO	BLANCO	CÓRDOVA
55	66	47
54	76	51
59	67	46
56	71	48

El cálculo de la variación de tratamiento es 992 y la variación aleatoria es 90. Calcular el valor de F.

Desarrollo

PASO 1: Establecer la hipótesis nula y la alternativa

$$\boldsymbol{H_0: \mu_{lobo}=\mu_{blanco}=\mu_{córdoba}}$$
$$\boldsymbol{H_a: No\ todas\ las\ medias\ son\ iguales}$$

PASO 2: Seleccionar el nivel de significancia

$$\boldsymbol{\alpha=0.10}$$

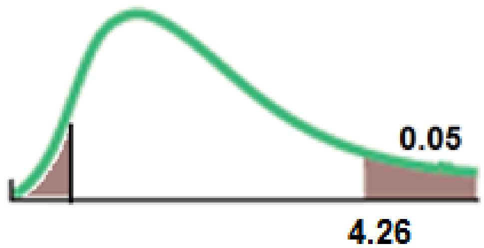
PASO 3: Determinar el estadístico de prueba

$$\boldsymbol{F=\frac{s_1^2}{s_2^2}}$$

PASO 4: Formular la regla de decisión

La hipótesis es una igualdad:	$H_0: \mu_{lobo} = \mu_{blanco} = \mu_{córdoba}$
Total de colas:	2 colas
Nivel se significancia:	$\alpha = \frac{0.10}{2} = 0.05$
Total de muestras:	$k = 3$
Tamaño observaciones:	$n = 12$
Grados de libertad:	$gl_1 = 3 - 1 = 2$
Grados de libertad:	$gl_2 = 12 - 3 = 9$

	1	2
1	161	200
2	18.5	19.0
3	10.1	9.55
4	7.71	6.94
5	6.61	5.79
6	5.99	5.14
7	5.59	4.74
8	5.32	4.46
9	5.12	4.26

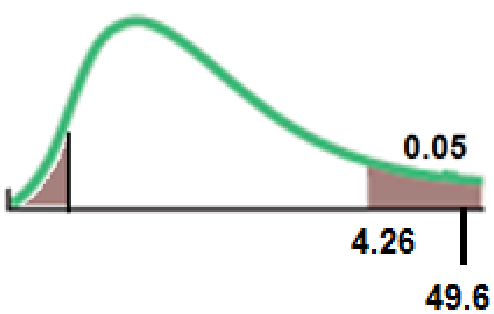


F=4.26

PASO 5: Toma de decisión

- Resumen de la tabla ANOVA

Variación	Σ^2	n	gl	Estimación Varianza	F
Tratamiento	992.0	3	2	496.0	49.60
Aleatoria	90.0	12	9	10.0	



La hipótesis nula se rechaza

Existe evidencia fuerte de que no todas las medias de la población son iguales

UNIVERSIDAD DE SAN PEDRO SULA

2. Desde hace algún tiempo las aerolíneas han reducido sus servicios, como alimentos y bocadillos durante sus vuelos; se ha estado cobrando de manera adicional algunos de los antiguos servicios. La central de aeropuerto desea conocer si este cambio ha producido insatisfacción en los clientes que utilizan sus servicios. Se levantaron 4 muestras sobre este tema en 4 aerolíneas distintas, sobre la satisfacción de los servicios y los resultados que se obtuvieron están mostrados en la siguiente tabla:

American	Taca	Iberia	Spirit
94	75	70	68
90	68	73	70
85	77	76	72
80	83	78	65
	88	80	74
		68	65
		65	

Con un nivel de significancia de 0.01 para 1 cola ¿Se puede concluir que hay alguna diferencia entre los niveles de satisfacción con respecto a las cuatro aerolíneas?

Desarrollo

PASOS INICIALES

Total de muestras: 4

Total de datos: 4+5+7+6= 22

PRUEBA DE HIPÓTESIS

Paso 1: Determinar la hipótesis nula y alternativa

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_a: No\ todas\ las\ medias\ son\ iguales$$

Paso 2: Seleccionar el nivel de significancia

$$\alpha = 0.01\ (1\ cola)$$

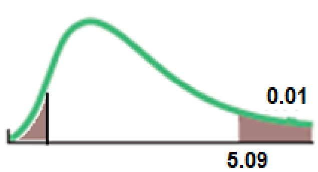
Paso 3: Seleccionar el estadístico de prueba

$$F = \frac{s_1^2}{s_2^2}$$

Paso 4: Formular regla de decisión

$$k = 4$$
$$n = 22$$
$$gl_1 = 4 - 1 = 3$$
$$gl_2 = 22 - 4 = 18$$

	1	2	3
1	4052	5000	5403
2	98.5	99.0	99.2
3	34.1	30.8	29.5
4	21.2	18.0	16.7
5	16.3	13.3	12.1
6	13.7	10.9	9.78
7	12.2	9.55	8.45
8	11.3	8.65	7.59
9	10.6	8.02	6.99
10	10.0	7.56	6.55
11	9.65	7.21	6.22
12	9.33	6.93	5.95
13	9.07	6.70	5.74
14	8.86	6.51	5.56
15	8.68	6.36	5.42
16	8.53	6.23	5.29
17	8.40	6.11	5.18
18	8.29	6.01	5.09



F=5.09

Paso 5: Toma de decisión

- Calcular la media global y las medias de cada muestra.

	American	Taca	Iberia	Spirit
	94	75	70	68
	90	68	73	70
	85	77	76	72
	80	83	78	65
		88	80	74
			68	65
			65	
$\bar{X}_m$	**87.3**	**78.2**	**72.9**	**69**
$\bar{X}_g$	**75.6**			

- Calcular la variación de tratamiento y la variación aleatoria

Aerolínea	X	$\bar{X}_m$	$\bar{X}_g$	Variación tratamiento $(\bar{X}_m - \bar{X}_g)^2$		Variación aleatoria $(X - \bar{X}_m)^2$	
American	94	87.3	75.6	$(87.3-75.6)^2$ =	134.9	$(94 - 87.3)^2$ =	45.6
	90			$(87.3-75.6)^2$ =	134.9	$(90 - 87.3)^2$ =	7.6
	85			$(87.3-75.6)^2$ =	134.9	$(85 - 87.3)^2$ =	5.1
	80			$(87.3-75.6)^2$ =	134.9	$(80 - 87.3)^2$ =	52.6
Taca	75	78.2		$(78.2-75.6)^2$ =	6.6	$(75 - 78.2)^2$ =	10.2
	68			$(78.2-75.6)^2$ =	6.6	$(68 - 78.2)^2$ =	104.0
	77			$(78.2-75.6)^2$ =	6.6	$(77 - 78.2)^2$ =	1.4
	83			$(78.2-75.6)^2$ =	6.6	$(83 - 78.2)^2$ =	23.0
	88			$(78.2-75.6)^2$ =	6.6	$(88 - 78.2)^2$ =	96.0
Iberia	70	72.9		$(72.9-75.6)^2$ =	7.7	$(70 - 72.9)^2$ =	8.2
	73			$(72.9-75.6)^2$ =	7.7	$(73 - 72.9)^2$ =	0.02
	76			$(72.9-75.6)^2$ =	7.7	$(76 - 72.9)^2$ =	9.9
	78			$(72.9-75.6)^2$ =	7.7	$(78 - 72.9)^2$ =	26.4
	80			$(72.9-75.6)^2$ =	7.7	$(80 - 72.9)^2$ =	51.0
	68			$(72.9-75.6)^2$ =	7.7	$(68 - 72.9)^2$ =	23.6
	65			$(72.9-75.6)^2$ =	7.7	$(65 - 72.9)^2$ =	61.7
Spirit	68	69		$(69-75.6)^2$ =	44.0	$(68 - 69)^2$ =	1.0
	70			$(69-75.6)^2$ =	44.0	$(70 - 69)^2$ =	1.0
	72			$(69-75.6)^2$ =	44.0	$(72 - 69)^2$ =	9.0
	65			$(69-75.6)^2$ =	44.0	$(65 - 69)^2$ =	16.0
	74			$(69-75.6)^2$ =	44.0	$(74 - 69)^2$ =	25.0
	65			$(69-75.6)^2$ =	44.0	$(65 - 69)^2$ =	16.0
			Σ		890.7		594.4

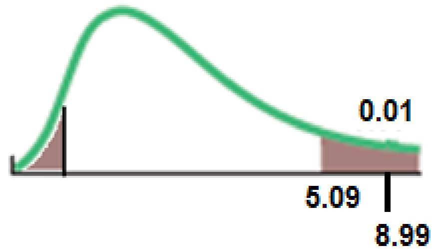
- Resumen de tabla ANOVA

Variación	Σ^2	gl	Estimación Varianza	F
Tratamiento	890.7	3	296.9	8.99
Aleatoria	594.4	18	33.0	

- Calcular el valor de F

$$\boldsymbol{F = 8.99}$$

USAP
UNIVERSIDAD DE SAN PEDRO SULA



La hipótesis nula se rechaza
Existe evidencia que no todas las medias son iguales

Tratamiento e inferencia sobre pares de medias

Al rechazar una hipótesis cuando las medias son más de 2, se concluye que no todas son iguales; pero, no se conoce cuáles son las que difieren. No siempre esta conclusión es satisfactoria, ya que se puede conocer cuáles medias de tratamiento difieren.

La distribución t sirve como base para obtener el factor en que difieren las medias; el cual es conocido como el **error medio cuadrado** (MSE=mean square error), calculado a partir de la variación aleatoria.

$$MSE = \frac{VA}{n-k}$$

Recordar que la ANOVA asume que las muestras vienen de poblaciones normalmente distribuidas.

Intervalo de confianza de la diferencia entre las medidas de tratamiento

El intervalo de confianza de la diferencia entre dos poblaciones se obtiene con la siguiente fórmula:

$$(\bar{X}_1 - \bar{X}_2) \pm t\sqrt{MSE\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

La hipótesis nula es la igualdad que asume que las dos medias muestrales elegidas son iguales. Para concluir que no hay diferencia entre ambas medias, el intervalo debe incluir el 0.

Ejemplo 2.7

1. Siguiendo con el ejemplo de las aerolíneas, que han reducido sus servicios, como alimentos y bocadillos durante sus vuelos; se ha estado cobrando de manera adicional algunos de los antiguos servicios. La central de aeropuerto desea conocer si este cambio ha producido insatisfacción en los clientes que utilizan sus servicios.

   Son 4 muestras de tamaños diferentes, tomar las muestras con la media más alta y la media más baja para evaluar qué tanto difieren entre ambas. Calcular el intervalo del 95% de confianza.

American	Taca	Iberia	Spirit
94	75	70	68
90	68	73	70
85	77	76	72
80	83	78	65
	88	80	74
		68	65
		65	

USAP
UNIVERSIDAD DE SAN PEDRO SULA

	American	Taca	Iberia	Spirit
	94	75	70	68
	90	68	73	70
	85	77	76	72
	80	83	78	65
		88	80	74
			68	65
			65	
$\bar{X}_m$	**87.3**	**78.2**	**72.9**	**69**
$\bar{X}_g$	**75.6**			
n	**4**	**5**	**7**	**6**

Desarrollo

Los datos obtenidos en la prueba de hipótesis son los siguientes:

Variación	Σ^2	gl	Estimación Varianza	F
Tratamiento	890.7	3	296.9	8.99
Aleatoria	594.4	18	33.0	

Media de la muestra de American: $\bar{X}_1 = 87.3$
Media de la muestra de Spirit: $\bar{X}_1 = 69.0$
Error medio al cuadrado: MSE = 33.0
Tamaño de la muestra de American: $n_1 = 4$
Tamaño de la muestra de Spirit: $n_2 = 6$

	Intervalo de confianza, c			
	80%	90%	95%	98%
	Nivel de significancia para una prueb			
gl	0.100	0.050	0.025	0.010
	Nivel de significancia para una prueba			
	0.200	0.10	0.05	0.02
1	3.078	6.314	12.706	31.821
2	1.886	2.920	4.303	6.965
3	1.638	2.353	3.182	4.541
4	1.533	2.132	2.776	3.747
5	1.476	2.015	2.571	3.365
6	1.440	1.943	2.447	3.143
7	1.415	1.895	2.365	2.998
8	1.397	1.860	2.306	2.896
9	1.383	1.833	2.262	2.821
10	1.372	1.812	2.228	2.764
11	1.363	1.796	2.201	2.718
12	1.356	1.782	2.179	2.681
13	1.350	1.771	2.160	2.650
14	1.345	1.761	2.145	2.624
15	1.341	1.753	2.131	2.602
16	1.337	1.746	2.120	2.583
17	1.333	1.740	2.110	2.567
18	1.330	1.734	2.101	2.552

Determinar el valor de t

- Nivel de Aceptación:
  95%
- Grados de libertad variación aleatoria:
  18
- $t = 2.101$

Calcular el Intervalo de confianza

$$(\bar{X}_a - \bar{X}_s) \pm \sqrt{MSE\left(\frac{1}{n_a} + \frac{1}{n_s}\right)} = (87.3 - 69) \pm 2.101\sqrt{33.0\left(\frac{1}{4} + \frac{1}{6}\right)}$$

$$= 18.3 \pm 2.101\sqrt{33.0(0.41667)} = 18.3 \pm 2.101\sqrt{13.8}$$

$$= 18.3 \pm 2.101(3.71) = 18.3 \pm 7.791$$

$$= \begin{cases} 18.3 - 7.791 = 10.5 \\ 18.3 + 7.791 = 26.1 \end{cases}$$

Los dos puntos extremos son positivos
Si hay suficiente evidencia para concluir que estas medias difieren de manera significativa

2. Citrus Clean es un nuevo limpiador multiusos a prueba en el mercado, del cual se han colocado exhibidores en varios supermercados de la ciudad. Una muestra tomada la semana pasada reportó que las cantidades de botellas que se vendieron en cada lugar de los supermercados se muestran en la tabla de la derecha

Cerca del pan	Cerca de la cerveza	Cerca de otros limpiadores
18	12	26
14	18	28
19	10	30
17	16	32

Con un nivel de significancia de 0.05. ¿Hay alguna diferencia entre los promedios de las botellas que se vendieron en los 3 lugares? ¿Qué indica el intervalo de confianza?

PASOS INICIALES

Total de muestras: 3

Total de datos: 4+4+4= 12

PRUEBA DE HIPÓTESIS

Paso 1: Determinar la hipótesis nula y alternativa

$$H_0: \mu_1 = \mu_2 = \mu_3$$

$$H_a: No\ todas\ las\ medias\ son\ iguales$$

Paso 2: Seleccionar el nivel de significancia

$$\alpha = 0.05\ (1\ cola)$$

Paso 3: Seleccionar el estadístico de prueba

$$F = \frac{s_1^2}{s_2^2}$$

Paso 4: Formular regla de decisión

$$k = 3$$

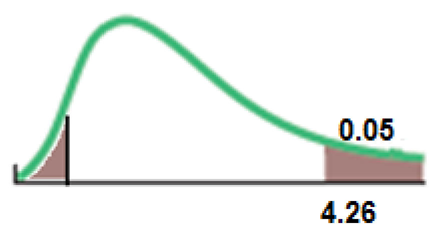
$$n = 12$$

$$gl_1 = 3 - 1 = 2$$

$$gl_2 = 12 - 3 = 9$$

$$F = 4.26$$

	1	2
1	161	200
2	18.5	19.0
3	10.1	9.55
4	7.71	6.94
5	6.61	5.79
6	5.99	5.14
7	5.59	4.74
8	5.32	4.46
9	5.12	4.26



Paso 5: Toma de decisión

- Calcular la media global y las medias de cada muestra.

	Cerca del pan	Cerca de la cerveza	Cerca de otros limpiadores
	18	12	26
	14	18	28
	19	10	30
	17	16	32
$\bar{X}_m$	**17.0**	**14.0**	**29.0**
$\bar{X}_g$	**20.0**		
n	**4**	**4**	**4**

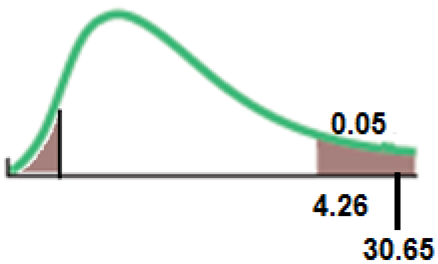
USAP
UNIVERSIDAD DE SAN PEDRO SULA

- Calcular la variación de tratamiento y la variación aleatoria

Ubicación	X	$\bar{X}_m$	$\bar{X}_g$	Variación tratamiento $(\bar{X}_m - \bar{X}_g)^2$	Variación aleatoria $(X - \bar{X}_m)^2$
Cerca del pan	18	17.0	20.0	$(17 - 20)^2$ = 9.0	$(18 - 17)^2$ = 1.0
	14			(17 - 20)2 = 9.0	$(14 - 17)^2$ = 9.0
	19			(17 - 20)2 = 9.0	$(19 - 17)^2$ = 4.0
	17			(17 - 20)2 = 9.0	$(17 - 17)^2$ = 0.0
Cerca de la cerveza	12	14.0		$(14- 20)^2$ = 36.0	$(12 - 14)^2$ = 4.0
	18			(14- 20)2 = 36.0	$(18 - 14)^2$ = 16.0
	10			(14- 20)2 = 36.0	$(10 - 14)^2$ = 16.0
	16			(14- 20)2 = 36.0	$(16 - 14)^2$ = 4.0
Cerca de otros limpiadores	26	29.0		$(29- 20)^2$ = 81.0	$(26 - 29)^2$ = 9.0
	28			(29- 20)2 = 81.0	$(28 - 29)^2$ = 1.0
	30			(29- 20)2 = 81.0	$(30 - 29)^2$ = 1.0
	32			(29- 20)2 = 81.0	$(32 - 29)^2$ = 9.0
			∑	504.0	74.0

- Resumen de la tabla ANOVA

Variación	$\sum^2$	n	gl	Estimación Varianza	F
Tratamiento	504.0	3	2	252.0	30.65
Aleatoria	74.0	12	9	8.2	



La hipótesis nula se rechaza
Hay evidencia de que no todas las medias son iguales

INTERVALO DE CONFIANZA

Determinar el valor de t

- Intervalo de confianza: 95%
- Grados de libertad variación aleatoria: 9

- $\boldsymbol{t = 2.262}$

	Intervalo de confianza, *c*		
	80%	90%	95%
	Nivel de significancia para		
gl	0.100	0.050	0.025
	Nivel de significancia para		
	0.200	0.10	0.05
1	3.078	6.314	12.706
2	1.886	2.920	4.303
3	1.638	2.353	3.182
4	1.533	2.132	2.776
5	1.476	2.015	2.571
6	1.440	1.943	2.447
7	1.415	1.895	2.365
8	1.397	1.860	2.306
9	1.383	1.833	2.262

Calcular el intervalo de confianza

$$(\bar{X}_a - \bar{X}_s) \pm \sqrt{MSE\left(\frac{1}{n_a} + \frac{1}{n_s}\right)}$$

$$= (29 - 17) \pm 2.263\sqrt{8.2\left(\frac{1}{4} + \frac{1}{4}\right)}$$

$$= 12 \pm 2.263\sqrt{4.1}$$

USAP
UNIVERSIDAD DE SAN PEDRO SULA

$$= \mathbf{12 \pm 4.58}$$
$$= \begin{cases} \mathbf{12 - 4.58 = 7.42} \\ \mathbf{12 + 4.58 = 16.58} \end{cases}$$

Los dos puntos extremos son positivos
Si hay suficiente evidencia para concluir que estas medias difieren de manera significativa

Ejercicios

1. Un agente de bienes raíces en el área costera de Georgia desea comparar la variación entre el precio de venta de casas con frente al mar y el de las ubicadas a tres cuadras del mar. Una muestra de 21 casas con frente al mar vendidas el año pasado revelo que la desviación estándar de los precios de venta fue $45,600. Una muestra de 18 casas, también vendidas el año pasado, ubicadas de una a tres cuadradas del mar, revelo que la desviación estándar fue $21,330. Con un nivel de significancia de 0.01, ¿puede concluir que hay más variación en los precios de venta de las casas con frente al mar?

2. En la ciudad hay dos concesionarios Chevrolet, Tiburones y Río Blanco. Las ventas mensuales medias en Tiburones y Rio Blanco son más o menos iguales. Sin embargo, el propietario de Tiburones, considera que sus ventas son más consistentes (variación). A continuación se presenta el número de automóviles nuevos vendidos en Tiburones en los últimos siete meses, y en los últimos ocho meses en Río Blanco. ¿Concuerda con Tiburones? Utilice el nivel de significancia 0.01.

Tiburones	98	78	54	57	68	64	70	
Río Blanco	75	81	81	30	82	46	58	101

3. Se tienen dos muestras de la cantidad de productos con fallas en dos empresas fabricantes de juguetes y se desea probar si existe alguna diferencia en la variación en cuanto a la cantidad de fallas con un nivel de significancia del 10%.

Compañía A	59	45	84	50	60	59	75	
Compañía B	45	56	57	74	53	61	67	79

4. Un inversionista en bienes raíces considera invertir en un centro comercial en los suburbios de Atlanta, Georgia, para lo cual evalúa tres terrenos. El ingreso familiar en el área circundante al centro comercial propuesto tiene una importancia particular. Se selecciona una muestra aleatoria de cuatro familias cerca de cada centro comercial propuesto. A continuación se presentan los resultados de la muestra. Con un nivel de significancia de 0.05, ¿el inversionista puede concluir que hay una diferencia en el ingreso medio? Utilice el procedimiento de prueba de hipótesis habitual de cinco pasos.

Southwyck (miles de dólares)	Franklin Park (miles de dólares)	Old Orchard (miles de dólares)
64	74	75
68	71	80
70	69	76
60	70	78

UNIVERSIDAD DE SAN PEDRO SULA

5. La gerente de una compañía de software desea estudiar el número de horas que los directivos de diversas empresas utilizan sus computadoras de escritorio. El gerente seleccionó una muestra de cinco ejecutivos de cada una de tres industrias. Con un nivel de significancia de 0.05, ¿puede la gerente concluir que hay una diferencia en el número medio de horas por semana utilizando las computadoras en la industria?

Bancaria	Detallista	De seguros
12	8	10
10	8	8
10	6	6
12	8	8
10	10	10

6. Una organización de consumidores desea saber si hay una diferencia en el precio de un juguete en particular en tres tipos de tiendas diferentes. El precio del juguete se investigó en una muestra de cinco tiendas de descuento, cinco tiendas de artículos diversos y cinco tiendas departamentales. Los resultados se muestran a continuación. Utilice el nivel de significancia de 0.05.

Tipo de tienda		
Descuento	**Variedad**	**Departamental**
13	15	19
13	17	17
14	14	16
12	18	20
15	17	19

Si la hipótesis nula se rechaza, elabore un intervalo de confianza de 95% para la diferencia de medias entre las tiendas de descuento y las tiendas por departamentos.

7. La ciudad de Maumee comprende cuatro distritos. Andy North, jefe de la policía, desea determinar si hay una diferencia en el número medio de delitos cometidos en los cuatro distritos. Para esto registra el número de delitos reportados en cada distrito para una muestra de seis días. Con un nivel de significancia de 0.10, ¿el jefe de la policía puede concluir que hay una diferencia en el número medio de delitos? Si la hipótesis nula se rechaza, elabore un intervalo de confianza del 90% para la diferencia de medias entre el Centro y Monclova.

Centro	Island	Monclova	Castaño
13	21	12	16
15	13	14	17
14	18	15	18
15	19	13	15
14	18	12	20
15	19	15	18

USAP
UNIVERSIDAD DE SAN PEDRO SULA

8. Con la siguiente información muestral, comprueba la hipótesis de que las medias de los tratamientos son iguales con un nivel de significancia de 0.10. Si se rechaza H_0, ¿Puede concluir que el tratamiento 2 y 3 difieren? Utilizar un nivel de confianza de 95%

Tratamiento 1	Tratamiento 2	Tratamiento 3
3	9	6
2	6	3
5	5	5
1	6	5
3	8	5
1	5	4
	4	1
	7	5
	6	
	4	

BIBLIOGRAFÍA

- Lind, D.A., Marchal, W.G., Wathen, S.A. (15). (2012). *Estadística Aplicada a los Negocios y la Economía*. México: McGraw-Hill

- David M. Levine, Timothy C. Krehbiel, Mark L. Berenson. 2006. *Estadística para Administración*. (4° edición). Naucalpan de Juárez, México.: Pearson Prentice Hall