



# Cap 12: Análisis de Varianza

Estadística Inferencial

---

Roberto S. Villamarín G

17 de junio de 2025

Pedagogía de las Ciencias Experimentales Matemáticas y Física

# Tabla de contenidos

1. Introducción
2. La distribución F
3. Comparación de dos varianzas poblacionales
4. Suposiciones en el análisis de la varianza (ANOVA)
5. Tratamiento e inferencia sobre pares de medias
6. Análisis de Varianza de dos vías
7. ANOVA de dos vías, con interacción

# Introducción

---

## Para recordar

1. Se analizó el caso en que se seleccionó una muestra de una población.
2. Se utilizó la distribución  $z$  (la distribución normal estándar) o la distribución  $t$  para determinar si era razonable concluir que la media poblacional era igual a un valor específico.
3. Se probó si dos medias poblacionales eran iguales.
4. También se realizaron pruebas de una y dos muestras de las proporciones de las poblaciones, con la distribución normal estándar como la distribución del estadístico de prueba.

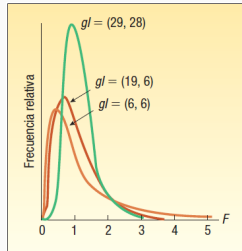
# La distribución F

---

# Características de F

- Con ella se pone a prueba si dos muestras provienen de poblaciones que tienen varianzas iguales, y también se aplica cuando se desea comparar **varias medias poblacionales en forma simultánea**.
- Se denomina análisis de la varianza (ANOVA)
- En las dos situaciones, las poblaciones deben seguir una **distribución normal**, y los datos deben ser al menos de **escala de intervalos**.

- Existe una familia de distribuciones F
- La forma de las curvas cambia cuando varían los grados de libertad.
- La distribución F es continua
- La distribución F no puede ser negativa
- Tiene sesgo positivo
- Es asintótica



**Figura 1:** Distribución Fisher

## **Comparación de dos varianzas poblacionales**

---

## Aspectos generales

- La primera aplicación de la distribución F ocurre cuando se pone a **prueba la hipótesis de que la varianza de una población normal es igual a la varianza de otra población normal.**
- Sirve para probar suposiciones de algunas pruebas estadísticas. La distribución F proporciona un medio para realizar una prueba considerando las varianzas de dos poblaciones normales
- 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.
  - $H_0 : \sigma_1^2 = \sigma_2^2$
  - $H_1 : \sigma_1^2 \neq \sigma_2^2$
- Para realizar la prueba, se selecciona una muestra aleatoria de  $n_1$  observaciones de una población y una muestra aleatoria de  $n_2$  observaciones de la segunda población.

- El estadístico de prueba se define como sigue.

$$F = \frac{s_1^2}{S_2^2} \quad (1)$$

Dónde  $s_1^2$ ;  $S_2^2$  son las varianzas muestrales respectivas.

- Si la hipótesis nula es verdadera, el estadístico de prueba sigue la distribución  $F$  con  $n_1-1$  y  $n_2-1$  grados de libertad
- A fin de reducir el tamaño de la tabla de valores críticos, **la varianza más grande de la muestra se coloca en el numerador**; de aquí, la razón  $F$  que se indica en la tabla siempre es mayor que 1.00.
- Así, el valor crítico de la cola derecha es el único que se requiere.
- El valor crítico de  $F$  de una prueba de dos colas se determina dividiendo el nivel de significancia entre dos  $\alpha/2$  y después se consultan los grados de libertad apropiados en el apéndice  $B \cdot 4$

## Ejemplo

Se ofrece servicio de transporte en limusina del ayuntamiento de Toledo, Ohio, al aeropuerto metropolitano de Detroit por dos rutas.

Una por la carretera 25 y la otra por la autopista I-75. Se requiere estudiar el tiempo que tardaría en conducir al aeropuerto por cada una de las rutas y luego comparar los resultados.

Recopiló los siguientes datos muestrales, reportados en minutos. Usando el nivel de significancia de 0,10, ¿hay alguna diferencia entre las variaciones de los tiempos de manejo por las dos rutas?

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

- Carretera

$$\bar{X} = \frac{\sum X}{n} = \frac{408}{7} = 58,29$$

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}} = \sqrt{\frac{485,43}{7 - 1}} = 8,9947$$

- Autopista

$$\bar{X} = \frac{\sum X}{n} = \frac{472}{8} = 59$$

$$s = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}} = \sqrt{\frac{134}{8 - 1}} = 4,3753$$

- $S_c > S_a$ ; La ruta de la autopista es más larga, pero la carretera tiene mas semáforos.
- Es importante que el servicio que ofrece sea tanto puntual como consistente, por lo que decide realizar una prueba estadística para determinar si en realidad existe una diferencia entre las variaciones de las dos rutas

# Pasos de la prueba de hipótesis

## 1º Planteamiento de las hipótesis

- $H_0 : \sigma_1^2 = \sigma_2^2$
- $H_1 : \sigma_1^2 \neq \sigma_2^2$
- Prueba a dos colas
- Se busca una diferencia entre las variaciones de las dos rutas. No se trata de demostrar que el tiempo que se emplea varía más por una ruta que por la otra.

## 2º $\alpha = 0,10$

## 3º Estadístico de Prueba Distribución $F$

## 4º Valor crítico de $F$

- $\alpha/2 = 0,05$
- $gl_{numerador} = 7 - 1 = 6$ ;  $gl_{denominador} = 8 - 1 = 7$
- Valor crítico  $F_t = 3,87$
- Criterio de Rechazo : Rechazar  $H_0$  si  $F_c > F_t$

| 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 |

**Figura 2:** Valor de F teórico

### 5º Toma de Decisión

$$F = \frac{S_1^2}{S_2^2} = \frac{8,9947^2}{4,3753^2} = 4,23$$

Dado que  $F_c > F_t$  se rechaza la  $H_0$ , y se concluye que *Existen diferencias estadísticamente significativas* entre las variaciones de los tiempos recorridos por las dos rutas.

# NO OLVIDAR!!

1. Poner la mayor de la varianzas en el numerador. Así el cociente siempre será mayor que 1,00
2. ¿Cómo hacer una prueba a una cola?
3. En el ejemplo anterior. Se sospecha que las varianzas por la carretera es mayor que por la autopista, las hipótesis serían:
  - $H_0 : \sigma_1^2 \leq \sigma_2^2$
  - $H_1 : \sigma_1^2 > \sigma_2^2$
4. Se calcula igual  $s_1^2/s_2^2$ , con  $\alpha$  sin cambios. Se puede utilizar la cola superior de la distribución F.
5. Es necesario utilizar software estadístico: *R* o *Excel* por ejemplo.
6. *DISTR.F.INV(probabilidad; gl\_numerador; gl\_denominador)*  
(Versiones anteriores)
7. *INV.F(probabilidad; gl\_numerador; gl\_denominador)* (Versiones actuales)

# Procedimiento en Excel

| arretera | Autopista |                                 |                                         |                  |
|----------|-----------|---------------------------------|-----------------------------------------|------------------|
| 52       | 59        |                                 | Prueba F para varianzas de dos muestras |                  |
| 67       | 60        |                                 |                                         |                  |
| 56       | 61        |                                 | <i>Carretera</i>                        | <i>Autopista</i> |
| 45       | 51        | Media                           | 58,29                                   | 59,00            |
| 70       | 56        | Varianza                        | 80,90                                   | 19,14            |
| 54       | 63        | Observaciones                   | 7,00                                    | 8,00             |
| 64       | 57        | Grados de libertad              | 6,00                                    | 7,00             |
|          | 65        | F                               | 4,23                                    |                  |
|          |           | P(F<=f) una cola                | 0,04                                    |                  |
|          |           | Valor crítico para F (una cola) | 3,87                                    |                  |

**Figura 3:** Prueba F para varianzas de dos muestras en Excel

- Autoevaluación 12-1
- Ejercicios Cap. 12: 1 al 6
- Estadística para la Administración y Negocios 15<sup>º</sup> Edición

## **Suposiciones en el análisis de la varianza (ANOVA)**

---

# ANOVA

Se utiliza para comparar las medias de tres o más poblaciones y determinar si pueden ser iguales, para ello se supone lo siguiente:

1. Las poblaciones siguen una distribución normal
2. Las poblaciones tienen distribuciones estándares ( $\sigma$ ) iguales
3. las poblaciones son independientes

Si cumples estas condiciones,  $F$  se emplea como estadístico de prueba

**¿Por qué usar ANOVA?**

## EVITAR LA ACUMULACIÓN DEL ERROR TIPO I

Si se tienen 4 muestras, se tendría que hacer 6 pruebas (dos a dos), por lo que error acumulado sería  $(0,95)^6 = 0,735$  por lo que  $1 - 0,735 = 0,265$ .

Al usar  $t$  se incrementa el error de  $\alpha = 0,05$  a un nivel intolerable de 0,265. **ANOVA le permite comparar las medias de tratamiento de forma simultánea** y evitar la acumulación del error de tipo I

## Ejemplo

Se desea medir la productividad de tres empleados y los datos son:

| Wolfe | White | Korosa |
|-------|-------|--------|
| 55    | 66    | 47     |
| 54    | 76    | 51     |
| 59    | 67    | 46     |
| 56    | 71    | 48     |

**Cuadro 1:** Datos de tres empleados

- ¿Hay alguna diferencia en el número medio de clientes atendidos?

# Solución- Prueba ANOVA

- Recuerde que se desea probar si *las medias muestrales provienen de una sola población o de poblaciones con medias diferentes*.
- Las medias muestrales se comparan mediante sus varianzas
- Suposición : Las desviaciones estándar de las diversas poblaciones normales, **deben ser las mismas**.

## Estrategia de la ANOVA

Estimar la varianza de la población de dos formas (diferentes) para determinar la razón de dichas estimaciones.

Si la razón es aproximadamente igual a 1, entonces las dos estimaciones son iguales y se concluye que las medias poblacionales **no son iguales**.

**La distribución  $F$  actúa como árbitro para indicar en que instancia la razón de las varianzas muestrales es mucho mayor que 1 para haber ocurrido por casualidad**

## Variación Total

Suma de las diferencias entre cada observación y la media global elevadas al cuadrado

## Variación de tratamiento (población)

Suma de las diferencias entre la media de cada tratamiento (población) y la media total o global elevadas al cuadrado

## Variación aleatoria

Suma de las diferencias entre cada observación y su media de tratamiento elevadas al cuadrado

|             | Wolfe     | White     | Korosa    |
|-------------|-----------|-----------|-----------|
|             | 55        | 66        | 47        |
|             | 54        | 76        | 51        |
|             | 59        | 67        | 46        |
|             | 56        | 71        | 48        |
| $\bar{x}_t$ | <b>56</b> | <b>70</b> | <b>48</b> |
| $\bar{X}_G$ | -         | -         | <b>58</b> |

$$\bar{X}_G = \frac{\sum \bar{x}_t}{n_t} = \frac{\sum_i^n x_i}{n}$$

$$V_{total} = \sum (x_i - \bar{X}_G)^2 = 1082$$

La  $V_{total}$  se divide en dos:

$$V_{Tratamiento} = n \cdot \sum (\bar{x}_t - \bar{X}_G)^2 = 992$$

$$V_{aleatoria} = \sum (x_{it} - \bar{x}_t)^2 = 90$$

# En Excel

|            |    |     |     |                 |     |     |                     |    |   |
|------------|----|-----|-----|-----------------|-----|-----|---------------------|----|---|
|            |    |     |     |                 |     |     |                     |    |   |
|            | W  | W   | K   | Variación total |     |     | Variación aleatoria |    |   |
|            | 55 | 66  | 47  | 9               | 64  | 121 | 1                   | 16 | 1 |
|            | 54 | 76  | 51  | 16              | 324 | 49  | 4                   | 36 | 9 |
|            | 59 | 67  | 46  | 1               | 81  | 144 | 9                   | 9  | 4 |
|            | 56 | 71  | 48  | 4               | 169 | 100 | 0                   | 1  | 0 |
| Media_trat | 56 | 70  | 48  | Var_total       |     |     | 1082                |    |   |
| M_global   | 58 |     |     |                 |     |     |                     |    |   |
|            |    |     |     |                 |     |     |                     |    |   |
|            |    |     |     |                 |     |     |                     |    |   |
|            | 4  | 144 | 100 |                 |     |     |                     |    |   |
|            |    |     | 992 |                 |     |     |                     |    |   |

**Figura 4:** Implementación en Excel

### Observe que:

- Si existe una variación considerable entre las medias de los tratamientos, es lógico que este término sea grande. Es decir la  $V_t$
- Si las medias son similares, este término será un valor bajo.
- El valor más bajo posible es cero. Esto ocurrirá cuando todas las medias de los tratamientos sean iguales.

El estadístico de prueba, que es la razón de las dos estimaciones de la varianza poblacional, se determina a partir de la siguiente ecuación:

$$F = \frac{\text{Estimación de la varianza poblacional basada en las diferencias entre las medias muestrales}}{\text{Estimación de la varianza poblacional basada en la variación dentro de la muestra}}$$

**Figura 5:** Prueba Fisher

$$F = \frac{\sigma_{\text{dif-med-mues}}}{\sigma_{\text{var-dentro-muestra}}}$$

- Número de observaciones (tratamientos) ( $n - 1 = 3 - 1 = 2$ ) para el cálculo de la varianza muestral, hay tres tratamientos
- 12 número total de observaciones (datos) menos el número de tratamientos:  $12 - 3 = 9$

$$F = \frac{\frac{992}{2}}{\frac{90}{9}} = 49,6$$

Dado de  $F \gg 1$  **se concluye que las medias de los tratamientos NO SON IGUALES**

*Hay una diferencia entre los números medios de clientes atendidos por cada uno de los tres empleados*

## Ejemplo 2:

Se ha realizado una encuesta de satisfacción de los clientes a 4 compañías de vuelo. Entre mayor la calificación, mayor el nivel de satisfacción con el servicio. La calificación mayor posible fue 100. (25 preguntas, evaluadas del 1 al 4. Escala: Mas-mejor.)

| Northern | WTA | Pocono | Branson |
|----------|-----|--------|---------|
| 94       | 75  | 70     | 68      |
| 90       | 68  | 73     | 70      |
| 85       | 77  | 76     | 72      |
| 80       | 83  | 78     | 65      |
|          | 88  | 80     | 74      |
|          |     | 68     | 65      |
|          |     | 65     |         |

¿Hay alguna diferencia entre los niveles de satisfacción medios con respecto a las cuatro aerolíneas? Use el nivel de significancia de 0.01.

# Prueba de Hipótesis

- **1º Paso: Hipótesis**

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$

- Las calificaciones son iguales

- $H_1 : \neg(\mu_1 = \mu_2 = \mu_3 = \mu_4)$

- No todas las calificaciones son iguales. Al menos una de ellas es diferente.

- **2º Paso:  $\alpha = 0,01$**

- **3º Paso: Estadístico de prueba:** Distribución  $F$

- **4º Paso: Regla de Decisión**

- Grados de libertad del **numerador**: Numero de tratamientos(aerolíneas) ( $k = 4$ ) menos 1:  $k - 1 = 3$

- Grados de libertad del **denominador**: Numero total de observaciones (datos) menos número de tratamientos (aerolíneas):

- $n - k = 22 - 4 = 18$

- Valor crítico: 5,09 <sup>1</sup>

- **Regla de Decisión:** Rechazar la  $H_0$  si  $F_c > F_t = 5,09$

---

<sup>1</sup>Ver tabla F al 0,01

- 5º Paso: Toma de decisión

| Fuente Variación | Suma Cuadrados       | Grados de libertad | Media Cuadrática          | F                 |
|------------------|----------------------|--------------------|---------------------------|-------------------|
| Tratamientos     | $SST^2$              | $K - 1$            | $MST = \frac{SST}{k - 1}$ | $\frac{MST}{MSE}$ |
| Error            | $SSE^3$              | $n - k$            | $MSE = \frac{SSE}{n - k}$ |                   |
| Total            | $SS \text{ Total}^4$ | $n - 1$            |                           |                   |

**Cuadro 2:** Tabla ANOVA

---

<sup>2</sup>Variación tratamientos

<sup>3</sup>Variación dentro de los tratamientos o Error aleatorio

<sup>4</sup>Variación Total

- Variación Total:  $SS\ TOTAL = \sum (X - \bar{X}_G)^2$
- Variación dentro de los tratamientos o el error aleatorio:  
 $SSE = \sum (X - X_c)^2$
- Variación de los tratamientos:  $SST = SS\ TOTAL - SSE$ :

Empezamos determinando el valor de  $SS_T$

$$SS_T = \sum (x_i - \bar{X}_G)^2 \quad (2)$$

Dónde:

- $x_i$  es cada observación de la muestra
- $\bar{X}_G$  es la media global o total

Ahora determinamos  $SS_E$

$$SS_E = \sum (x_{it} - \bar{X}_t) \quad (3)$$

Dónde:

- $x_{it}$  es cada observación de ese tratamiento
- $\bar{X}_t$  es la media muestral del tratamiento  $t$

$$SST = SS_{total} - SSE \quad (4)$$

|           | Northern | WTA   | Pocono | Branson | Total |
|-----------|----------|-------|--------|---------|-------|
|           | 94       | 75    | 70     | 68      |       |
|           | 90       | 68    | 73     | 70      |       |
|           | 85       | 77    | 76     | 72      |       |
|           | 80       | 83    | 78     | 65      |       |
|           |          | 88    | 80     | 74      |       |
|           |          |       | 68     | 65      |       |
|           |          |       | 65     |         |       |
| Total     | 349      | 391   | 510    | 414     | 1664  |
| n         | 4        | 5     | 7      | 6       | 22    |
| $\bar{x}$ | 87,25    | 78,20 | 72,86  | 69,00   | 75,64 |

$$\bar{X}_G = \frac{1664}{22} = 75,64$$

| (x <sub>i</sub> - x <sub>G</sub> ) |        |         |         | (x <sub>i</sub> - x <sub>G</sub> ) <sup>2</sup> |         |        |         | total  |
|------------------------------------|--------|---------|---------|-------------------------------------------------|---------|--------|---------|--------|
| Northern                           | WTA    | Pocono  | Branson | Northern                                        | WTA     | Pocono | Branson |        |
| 18,364                             | -0,636 | -5,636  | -7,636  | 337,22                                          | 0,40    | 31,77  | 58,31   |        |
| 14,364                             | -7,636 | -2,636  | -5,636  | 206,31                                          | 58,31   | 6,95   | 31,77   |        |
| 9,364                              | 1,364  | 0,364   | -3,636  | 87,68                                           | 1,86    | 0,13   | 13,22   |        |
| 4,364                              | 7,364  | 2,364   | -10,636 | 19,04                                           | 54,22   | 5,59   | 113,13  |        |
|                                    | 12,364 | 4,364   | -1,636  |                                                 | 152,86  | 19,04  | 2,68    |        |
|                                    |        | -7,636  | -10,636 |                                                 |         | 58,31  | 113,13  |        |
|                                    |        | -10,636 |         |                                                 |         | 113,13 |         |        |
|                                    |        |         |         | 650,256                                         | 267,661 | 234,93 | 332,248 | 1485,1 |

**Figura 6:**

| (x_it - x_t) |        |        |         |  | (x_it - x_t)^2 |        |        |         | total  |
|--------------|--------|--------|---------|--|----------------|--------|--------|---------|--------|
| Northern     | WTA    | Pocono | Branson |  | Northern       | WTA    | Pocono | Branson |        |
| 6,75         | -3,20  | -2,86  | -1,00   |  | 45,56          | 10,24  | 8,16   | 1,00    |        |
| 2,75         | -10,20 | 0,14   | 1,00    |  | 7,56           | 104,04 | 0,02   | 1,00    |        |
| -2,25        | -1,20  | 3,14   | 3,00    |  | 5,06           | 1,44   | 9,88   | 9,00    |        |
| -7,25        | 4,80   | 5,14   | -4,00   |  | 52,56          | 23,04  | 26,45  | 16,00   |        |
|              | 9,80   | 7,14   | 5,00    |  | 0,00           | 96,04  | 51,02  | 25,00   |        |
|              |        | -4,86  | -4,00   |  | 0,00           | 0,00   | 23,59  | 16,00   |        |
|              |        | -7,86  |         |  | 0,00           | 0,00   | 61,73  | 0,00    | 594,41 |
| SST= SSt-Sse |        |        |         |  |                |        |        |         |        |
| 890,68       |        |        |         |  |                |        |        |         |        |

Figura 7:

| Fuente de variación | Suma de cuadrado | Grados de libertad | Media Cudrática | F         |
|---------------------|------------------|--------------------|-----------------|-----------|
| Tratamientos        | 890,68           | 3                  | 296,8945887     | 8,9906433 |
| Error               | 594,41           | 18                 | 33,02261905     |           |
| Total               | 1485,09          | 21                 |                 |           |

**Figura 8:**

Dado que  $F_c > F_t$  se **rechaza la  $H_0$** . **La conclusión es que no todas las medias poblacionales son iguales**

**OJO!!!**

En este punto sólo es posible concluir que hay una diferencia entre las medias del tratamiento.

**No se puede determinar cuáles ni cuántos grupos de tratamientos difieren.**

# ANOVA DE UN FACTOR EN EXCEL

|                                   |                          |                           |                                  |                 |                             |                             |
|-----------------------------------|--------------------------|---------------------------|----------------------------------|-----------------|-----------------------------|-----------------------------|
| Análisis de varianza de un factor |                          |                           |                                  |                 |                             |                             |
| RESUMEN                           |                          |                           |                                  |                 |                             |                             |
| <i>Grupos</i>                     | <i>Cuenta</i>            | <i>Suma</i>               | <i>Promedio</i>                  | <i>Varianza</i> |                             |                             |
| Northern                          | 4                        | 349                       | 87,25                            | 36,916667       |                             |                             |
| WTA                               | 5                        | 391                       | 78,2                             | 58,7            |                             |                             |
| Pocono                            | 7                        | 510                       | 72,857143                        | 30,142857       |                             |                             |
| Branson                           | 6                        | 414                       | 69                               | 13,6            |                             |                             |
| ANÁLISIS DE VARIANZA              |                          |                           |                                  |                 |                             |                             |
| <i>Origen de las variaciones</i>  | <i>Suma de cuadrados</i> | <i>Grados de libertad</i> | <i>Promedio de los cuadrados</i> | <i>F</i>        | <i>Probabilidad p_valor</i> | <i>Valor crítico para F</i> |
| Entre grupos                      | 890,68377                | 3                         | 296,89459                        | 8,9906433       | 0,00074277                  | 3,1599076                   |
| Dentro de los grupos              | 594,40714                | 18                        | 33,022619                        |                 |                             |                             |
| Total                             | 1485,0909                | 21                        |                                  |                 |                             |                             |

Figura 9:

- Autoevaluación 12-2
- Ejercicios Cap. 12: 7 al 10
- Estadística para la Administración y Negocios 15o Edición

## Tratamiento e inferencia sobre pares de medias

---

## Recuerde que:

- Al rechazar la  $H_0$  no se puede determinar cuales de las medias son diferentes, **solo permite saber que no todas son iguales**
- Se desea saber: **¿entre qué grupos difieren las medias de tratamiento?**
- Para responder esta pregunta utilizaremos los intervalos de confianza: <sup>5</sup>

$$\bar{X} \pm t \cdot \frac{s}{\sqrt{n}}$$

- En el ejemplo de las aerolíneas, las medias son:
  - Northtern : 87,25
  - Branson: 69,00
  - Pocono: 72857
  - WTA : 78,2

---

<sup>5</sup>Este es uno de los métodos más simples

¿Existe suficiente disparidad para justificar la conclusión de que hay una diferencia significativa entre las calificaciones de satisfacción medias de las aerolíneas

## Recuerde que:

La distribución  $t$ , sirve como base de esta prueba.

Recuerde que una de las suposiciones de ANOVA es que las varianzas poblacionales de todos los tratamientos son las mismas.

Este valor común de la población es el error medio cuadrático, o  $MSE$ , y se determina mediante  $SSE/(n-k)$ .

Un intervalo de confianza de la diferencia entre dos poblaciones se obtiene mediante:

### Intervalo de confianza de la diferencia entre las medias de tratamiento

$$(\bar{X}_1 - \bar{X}_2) \pm t \cdot \sqrt{MSE \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \quad (5)$$

Donde:

- $\bar{X}_1, \bar{X}_2$  es la media de primera y segunda muestra, respectivamente.
- $t$  es el valor de  $t$  que se obtiene de las tablas. Los grados de libertad son  $n - k$
- $MSE$  es el **error medio cuadrático que se obtiene de la tabla ANOVA**  $[SSE/(n - k)]$
- $n_1, n_2$  es el número de observaciones de la primera y segunda muestra, respectivamente.

¿Cómo se concluye que hay diferencias entre los tratamientos?

- Si el punto extremo izquierdo del IC tiene signo negativo y el derecho signo positivo, es decir, incluye al cero, y las dos medias no difieren. Suponga el siguiente caso:

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

$$(\bar{X}_1 - \bar{X}_2) = 5$$

$$t \cdot \sqrt{MSE \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} = 12$$

$$5 \pm 12 \rightarrow [-7; 17]$$

Este intervalo **incluye** el **cero** por lo que se concluye **que no existe diferencia significativa entre las medias de los tratamientos seleccionados**

- Suponga el siguiente escenario:

$$(\bar{X}_1 - \bar{X}_2) = -0,35$$

$$t \cdot \sqrt{MSE \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} = 0,25$$

$$-0,35 \pm 0,25 \rightarrow [-0,60; -0,10]$$

Este intervalo **NO incluye** el **cero** por lo que se concluye **que SI existe diferencia significativa entre las medias de los tratamientos seleccionados**

## Caso aerolíneas

En el ejemplo de las aerolíneas, las medias son:

- Northern :  $\bar{X}_A = 87,25$
- Branson:  $\bar{X}_{US} = 69,00$
- Nivel de confianza 95 %
- $t = 2,101$  con  $n - k = 22 - 4 = 18$  grados de libertad
- $n_E = 4$
- $n_{US} = 6$
- $MSE = 33$  tomado de la tabla ANOVA con  $SSE/(n - k) = 594,4/18$

$$\begin{aligned} & (\bar{X}_1 - \bar{X}_2) \pm t \cdot \sqrt{MSE \left( \frac{1}{n_1} + \frac{1}{n_2} \right)} \\ & (87,25 - 69) \pm 2,101 \sqrt{33 \cdot \left( \frac{1}{4} + \frac{1}{6} \right)} \\ & = 18,25 \pm 7,79 \rightarrow [10,46 \leftrightarrow 26,04] \end{aligned}$$

**El intervalo no incluye al cero, por lo que se puede concluir que las medias de tratamiento difieren de manera significativa**

- Comparar las medias restantes de los otros grupos.
- Autoevaluación 12-3
- Ejercicios Cap. 12: 11 al 14
- Estadística para la Administración y Negocios 15<sup>º</sup> Edición

# **Análisis de Varianza de dos vías**

---

Hasta ahora conocemos que:

- La variación total se dividió en dos categorías:
  - La variación entre los tratamientos
  - La variación dentro de los tratamientos (error o variación aleatoria)

Es decir solo se consideraron las variaciones debidas a los tratamientos y debidas a las diferencias aleatorias.

- Pero, pueden existir otras fuentes o causas de variación (estación del año, el aeropuerto, número de pasajeros, etcétera)
- Al considerar otros factores, se reduce la **varianza del error** (al reducir el denominador de  $F$ , concretamente el término  $SSE$ ), entonces  $F$  será mayor y probablemente ocasionará el rechazo de la hipótesis del tratamiento de medias iguales.
- Si se puede explicar mas la variación, habrá menos ERROR

## Ejemplo

Se considera ampliar el servicio de autobuses, a 4 rutas más. Cada conductor conduce sobre todas las rutas. Los datos (tiempos) registrados son:

| Conductor | Carretera 6 | West End | Hickory St. | Ruta 59 |
|-----------|-------------|----------|-------------|---------|
| Deans     | 18          | 17       | 21          | 22      |
| Snaverly  | 16          | 23       | 23          | 22      |
| Ormson    | 21          | 21       | 26          | 22      |
| Zollaco   | 23          | 22       | 29          | 25      |
| Filbeck   | 25          | 24       | 28          | 28      |

**Cuadro 3:** Tiempo recorrido en minutos

Con un nivel de significancia 0,05. ¿Hay alguna diferencia entre los tiempos de recorrido medios a lo largo de las cuatro rutas?

# Solución

- Empezamos con una **prueba de ANOVA de una vía**, es decir, solo se considera las 4 rutas.
  - $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$
  - $H_1 : \text{No todas las medias de tratamiento son iguales}$
- Cuatro rutas, por lo tanto; grados de libertad del numerador  $k - 1 = 4 - 1 = 3$
- Son 20 observaciones, por lo tanto; grados de libertad del denominador  $n - k = 20 - 4 = 16$
- Valor crítico  $F_{0,05} = 3,24$
- **Regla de Decisión:** Rechazar  $H_0$  si  $F_c > F_t$
- $F_c = 2,482$
- **Conclusión:** Dado que  $F_c > F_t$  se rechaza  $H_0$ , es decir; **no hay una diferencia entre los tiempos de recorrido medios a lo largo de las cuatro rutas. No hay una razón para seleccionar una de las rutas como más rápida que las demás.**

| Análisis de varianza de un factor |                   |                    |                           |           |              |                      |
|-----------------------------------|-------------------|--------------------|---------------------------|-----------|--------------|----------------------|
| RESUMEN                           |                   |                    |                           |           |              |                      |
| Grupos                            | Cuenta            | Suma               | Promedio                  | Varianza  |              |                      |
| Carretera 6                       | 5                 | 103                | 20,6                      | 13,3      |              |                      |
| West End                          | 5                 | 107                | 21,4                      | 7,3       |              |                      |
| Hickory                           | 5                 | 127                | 25,4                      | 11,3      |              |                      |
| Ruta 59                           | 5                 | 119                | 23,8                      | 7,2       |              |                      |
| ANÁLISIS DE VARIANZA              |                   |                    |                           |           |              |                      |
| Origen de las variaciones         | Suma de cuadrados | Grados de libertad | Promedio de los cuadrados | F         | Probabilidad | Valor crítico para F |
| Entre grupos                      | 72,8              | 3                  | 24,266667                 | 2,4825234 | 0,098105     | 3,2388715            |
| Dentro de los grupos              | 156,4             | 16                 | 9,775                     |           |              |                      |
| Total                             | 229,2             | 19                 |                           |           |              |                      |

**Figura 10:** Varianza de una vía en Excel

$$SST = 72,28$$

$$SSE = 156,4$$

$$SS_{total} = 229,2$$

# Variables de bloqueo

- Para el caso de la Varianza de una vía, solo se consideró la rutas (se tomó toda variación restante como aleatoria), **no se consideró a los conductores**
- Si se pudiera considerar el efecto de los diversos conductores, se podría reducir el término  $SSE$ , lo cual generaría un valor mayor de  $F$ .
- A la segunda variable de tratamiento, en este caso los **conductores**, se le conoce como **variable de bloqueo**.

## Variables de bloqueo

Es una segunda variable de tratamiento que, cuando se incluye en el análisis ANOVA, **tendrá el efecto de reducir el término  $SSE$**

## Ejemplo-continuación

- Los conductores son la variable de bloqueo y al eliminar el efecto de los conductores sobre del término  $SSE$  cambiará la razón (la división)  $F$  de la variable de tratamiento.
- En **ANOVA de dos vías**, la suma de los cuadrados debida a los bloqueos se determina mediante

$$SSB = k \cdot \sum (\bar{X}_b - \bar{X}_G)^2 \quad (6)$$

Donde:

- $k$  es el número de tratamientos
- $b$  es el numero de bloqueos
- $\bar{X}_b$  es la media muestral del bloque  $b$
- $\bar{X}_G$  es media global o total.

## Ejemplo-continuación

| Conductor | Carretera 6 | West End | Hickory St. | Ruta 59 | $S_c^6$ | $\bar{X}_c$ |
|-----------|-------------|----------|-------------|---------|---------|-------------|
| Deans     | 18          | 17       | 21          | 22      | 78      | 19,5        |
| Snaverly  | 16          | 23       | 23          | 22      | 84      | 21          |
| Ormson    | 21          | 21       | 16          | 22      | 90      | 22,5        |
| Zollaco   | 23          | 22       | 29          | 25      | 99      | 24,75       |
| Filbeck   | 25          | 24       | 28          | 28      | 105     | 26,25       |

$$\bar{X}_G = 22,8$$

$$SSB = k \cdot \sum (\bar{X}_b - \bar{X}_G)^2$$

$$SSB = 4 \cdot [(19,5 - 22,8)^2 + (21,0 - 22,8)^2 + (22,5 - 22,8)^2 + (24,75 - 22,8)^2 + (26,25 - 22,8)^2] = 119,7$$

---

<sup>6</sup>Suma del conductor

## Ejemplo continuación

### Suma de Errores Cuadráticos de dos vías

$$SSE = SS_{total} - SST - SSB \quad (7)$$

Los valores de los varios componentes de la tabla ANOVA se calculan como sigue.

| Fuente Variación | Suma Cuadrados     | Grados de libertad | Media Cuadrática                   | F                 |
|------------------|--------------------|--------------------|------------------------------------|-------------------|
| Tratamientos     | $SST$              | $k - 1$            | $MST = \frac{SST}{k - 1}$          | $\frac{MST}{MSE}$ |
| Bloqueos         | $SSB$              | $b - 1$            | $MSB = \frac{SSB}{b - 1}$          | $\frac{MSB}{MSE}$ |
| Error            | $SSE$              | $(k - 1)(b - 1)$   | $MSE = \frac{SSE}{(k - 1)(b - 1)}$ |                   |
| Total            | $SS \text{ Total}$ | $n - 1$            |                                    |                   |

**Cuadro 4:** Tabla ANOVA de dos vías

## Ejemplo-continuación

$SSE$  se obtiene con la fórmula 7

$$SSE = SS_{total} - SST - SSB$$

$$SSE = 229,2 - 72,8 - 119,7 = 36,7$$

|                     | (1)                   | (2)                | (3) = (1)/(2)    |
|---------------------|-----------------------|--------------------|------------------|
| Fuente de Variación | Suma de los cuadrados | Grados de libertad | Media cuadrática |
| Tratamientos        | 72,8                  | 3                  | 24,27            |
| Bloqueos            | 119,7                 | 4                  | 29,93            |
| Error               | 36,7                  | 12                 | 3,06             |
| Total               | 229,2                 | 19                 | -                |

## Consideraciones finales

- En este punto hay un desacuerdo. Si el objetivo de la variable de bloqueo (los conductores en este ejemplo) fue sólo reducir la variación del error, no se debe realizar una prueba de hipótesis de las diferencias entre las medias de los bloques.
- Es decir, si el objetivo era reducir el término  $MSE$ , no se debe probar una hipótesis respecto de la variable de bloqueo.
- Por otro lado, quizá se desee dar a los bloques la misma condición que a los tratamientos y realizar una prueba de hipótesis.
- Este último caso, **cuando los bloques son lo bastante importantes para considerarse un segundo factor**, se conoce como **un experimento de dos factores**. *En muchos casos, la decisión no es clara.*

En este ejemplo lo importante es la diferencia entre los tiempos de recorrido de los diversos conductores, por lo que se realizará la prueba de hipótesis. Los dos conjuntos de hipótesis son:

1.
  - $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$   
Las medias de los tratamientos son iguales.
  - $H_1$  Las medias de los tratamientos NO son iguales.
2.
  - $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$   
Las medias de los *bloques* son iguales.
  - $H_1$  Las medias de los *bloques* NO son iguales.

## Caso 1: respecto de las medias de tratamiento (rutas)

Primero se pondrá a prueba la hipótesis respecto de las medias de tratamiento

- Hay  $k-1 = 4-1 = 3$  grados de libertad en el numerador,y;
- $(b-1)(k-1) = (5-1)(4-1) = 12$  grados de libertad en el denominador.
- $F_{0,05} = 3,49$ .
- **Regla de Decisión:** Rechace  $H_0$  si  $F_c > F_t = 3,49$

- 

$$F = \frac{MST}{MSE} = \frac{24,27}{3,06} = 7,93$$

- Dado que  $F_c = 7,93 > F_t = 3,49$

**Se rechaza  $H_0$** , es decir, **se concluye que el tiempo de recorrido medio no es el mismo para todas las rutas.**

## Caso 2: respecto de las medias de los conductores (bloqueos)

Segundo se pondrá a prueba respecto de los bloqueos

- Grados de libertad del numerador  $b - 1 = 5 - 1 = 4$
- Grados de libertad del denominador (los mismos que el caso 1)  
 $(b-1)(k-1) = (5-1)(4-1) = 12$
- $F_{0,05} = 3,26$ .

- **Regla de Decisión:** Rechace  $H_0$  si  $F_c > F_t = 3,26$

- 

$$F = \frac{MSB}{MSE} = \frac{29,93}{3,06} = 9,78$$

- Dado que  $F_c = 9,78 > F_t = 3,26$

**Se rechaza  $H_0$** , es decir, **se concluye que el tiempo medio no es el mismo para los conductores.**

|                        |                                                                     |                      |                       |                                 |           |              |                         |
|------------------------|---------------------------------------------------------------------|----------------------|-----------------------|---------------------------------|-----------|--------------|-------------------------|
|                        | Análisis de varianza de dos factores con una sola muestra por grupo |                      |                       |                                 |           |              |                         |
|                        | RESUMEN                                                             | Cuenta               | Suma                  | Promedio                        | Varianza  |              |                         |
|                        | Deans                                                               | 4                    | 78                    | 19,5                            | 5,6666667 |              |                         |
|                        | Snaverly                                                            | 4                    | 84                    | 21                              | 11,333333 |              |                         |
|                        | Ormsen                                                              | 4                    | 90                    | 22,5                            | 5,6666667 |              |                         |
|                        | Zollaco                                                             | 4                    | 99                    | 24,75                           | 9,5833333 |              |                         |
|                        | Flibeck                                                             | 4                    | 105                   | 26,25                           | 4,25      |              |                         |
|                        | Carretera 6                                                         | 5                    | 103                   | 20,6                            | 13,3      |              |                         |
|                        | West End                                                            | 5                    | 107                   | 21,4                            | 7,3       |              |                         |
|                        | Hickory                                                             | 5                    | 127                   | 25,4                            | 11,3      |              |                         |
|                        | Ruta 59                                                             | 5                    | 119                   | 23,8                            | 7,2       |              |                         |
|                        | ANÁLISIS DE VARIANZA                                                |                      |                       |                                 |           |              |                         |
| bloque<br>(Conductor)  | Origen de las<br>variaciones                                        | Suma de<br>cuadrados | Grados de<br>libertad | Promedio<br>de los<br>cuadrados | F         | Probabilidad | Valor critico<br>para F |
|                        | Filas                                                               | 119,7                | 4                     | 29,925                          | 9,7847411 | 0,000933574  | 3,259166727             |
| Tratamiento<br>(Rutas) | Columnas                                                            | 72,8                 | 3                     | 24,266667                       | 7,9346049 | 0,003507913  | 3,490294819             |
|                        | Error                                                               | 36,7                 | 12                    | 3,0583333                       |           |              |                         |
|                        | Total                                                               | 229,2                | 19                    |                                 |           |              |                         |

bloque  
(Conductor)

Tratamiento  
(Rutas)

**Figura 11:** Anova de 2 vías en Excel

**OJO!!!**

Observe que el valor de p, en los dos casos son menores que  $\alpha$  por lo que se confirma que la decisión de rechazar las  $H_0$  es correcta.

- Comparar las medias restantes de los otros grupos.
- Autoevaluación 12-4
- Ejercicios Cap. 12: 15 al 18
- Estadística para la Administración y Negocios 15<sup>º</sup> Edición

## **ANOVA de dos vías, con interacción**

---

# Overview

- Hasta ahora se estudiaron los **efectos separados o independientes de dos variables**, rutas hacia la ciudad y conductores, respecto a los tiempos de recorrido medios.
- Los resultados muestrales indicaron distintos tiempos medios según las **rutas**, que se puede atribuir a las **distancias** por la rutas elegidas.
- Los resultados también indicaron diferencias entre los **tiempos de conducción medios** de los diversos conductores. que podrían ser por la **velocidades promedios** de cada conductor, sin importar la ruta.
- Existe otro efecto que influye en el tiempo de recorrido. A éste se le denomina **efecto de interacción** entre la ruta y el conductor sobre el tiempo de recorrido
- Es posible que un conductor sea especialmente bueno en una de las rutas (conoce: semáforos, evitar atascos, carriles rápidos, etc. ). En este caso, el **efecto combinado del conductor y la ruta** también explica las diferencias entre los tiempos de recorrido medios.

- **OJO:** Para medir los efectos de interacción es necesario tener al menos dos observaciones en cada celda
- **ANTES:** tratamientos y bloques  
**AHORA:** factores (a las dos variables)
- En el mismo ejemplo anterior:
  - **Factores:** Conductores y rutas
  - **Interacción** entre los dos factores
- **Hay un efecto de las rutas, del conductor, y de la interacción entre ambos factores (conductores y rutas)**
- La interacción tiene lugar si la combinación de dos factores ejerce algún efecto sobre la variable en estudio, además de hacerlo en cada factor por sí mismo.
- A la variable en estudio se le llama **variable de respuesta**.

# Ejemplo de variables con interacción

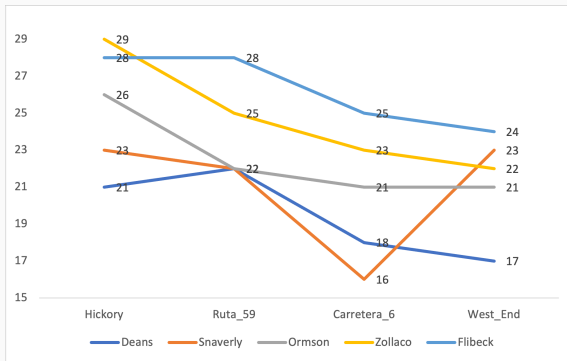
Un ejemplo cotidiano de interacción es el efecto de **dieta y ejercicio** sobre el **peso**.

- En general, se acepta que el **peso de una persona (la variable de respuesta)** se controla mediante dos factores, **dieta y ejercicio**.
- Las investigaciones demuestran que una dieta, por sí sola, afecta al peso de una persona, y también que el solo ejercicio tiene un efecto sobre el peso.
- Sin embargo, el método recomendado para controlar el peso se fundamenta en el **efecto combinado o en la interacción entre dieta y ejercicio**.

## Interacción

El efecto de un factor sobre una variable de respuesta difiere según el valor de otro factor.

# Gráficas de interacción



**Figura 12:** Gráficos de interacción para tiempos de viaje

# Interpretación de las Gráficas de interacción

- Si los segmentos de recta de los conductores son casi paralelos, tal vez no haya interacción.
- Si los segmentos de recta no parecen ser paralelos o se cruzan, esto sugiere una interacción entre los factores
- En la figura 12 se puede ver que:
  - Los segmentos de recta de Zollaco y Filbeck se cruzan entre sí.
  - El segmento de recta de Snaverly de la carretera 6 a West End cruza tres segmentos de recta.

## Conclusión

Estas observaciones *sugieren* una **interacción** entre el **conductor** y la **ruta**.

# Prueba de hipótesis para detectar una interacción

El estudio de los tiempos de recorrido plantea varias preguntas:

- ¿Hay alguna interacción entre rutas y conductores? **Cuestión de mayor interés**
- ¿Los tiempos de recorrido de los conductores son iguales?
- ¿Los tiempos de recorrido de las rutas son iguales
- Estas preguntas se investigan al ampliar el procedimiento ANOVA de dos vías, **agregando otra fuente de variación**, la **INTERACCIÓN**
- Para estimar la suma de ERROR de los cuadrados, es necesario **al menos dos mediciones para cada combinación conductor / ruta**<sup>7</sup>

---

<sup>7</sup>Para Excel es importante que en las columnas se coloque los grupos (tratamientos) y en las filas (bloques) **en ese orden**

## Datos para ANOVA DE INTERACCIÓN

|          | US_6 | West_End | Hickoy St | Route 69 |
|----------|------|----------|-----------|----------|
| Deans    | 18   | 14       | 20        | 19       |
| Deans    | 15   | 17       | 21        | 22       |
| Deans    | 21   | 20       | 22        | 25       |
| Snaverly | 19   | 20       | 24        | 24       |
| Snaverly | 15   | 24       | 23        | 22       |
| Snaverly | 14   | 25       | 22        | 20       |
| Ormson   | 19   | 23       | 25        | 23       |
| Ormson   | 21   | 21       | 19        | 23       |
| Ormson   | 23   | 19       | 24        | 20       |
| Zollaco  | 24   | 20       | 30        | 26       |
| Zollaco  | 20   | 24       | 28        | 25       |
| Zollaco  | 25   | 22       | 29        | 24       |
| Filbeck  | 24   | 24       | 28        | 28       |
| Filbeck  | 25   | 24       | 28        | 30       |
| Filbeck  | 23   | 24       | 28        | 26       |

# Hipótesis de ANOVA de interacción

Ahora ANOVA, tiene 3 conjuntos de hipótesis que se deben probar:

## 1. Hipótesis de interacción

- $H_0$  : No hay interacción entre los conductores y las rutas
- $H_1$  : Si hay interacción entre los conductores y las rutas

## 2. Hipótesis referidas a los tiempos de los conductores

- $H_0$  : Las medias de los conductores son iguales
- $H_1$  : Las medias de los conductores NO son iguales

## 3. Hipótesis referidas a las rutas

- Las medias de las rutas son iguales
- Las medias de las rutas NO son iguales

*Note que se identifica el efecto del conductor como **Factor A** y el efecto de la ruta, como **Factor B***

# Consideraciones para la prueba de Hipótesis

- Cada Hipótesis se prueba con  $F$
- Se puede utilizar  $F_c$  y  $F_t$  o utilizar el  $p\_valor$  para cada prueba
- Se utiliza el mismo valor de  $\alpha = 0,05$  para todos los casos
- Se aplica la misma regla de decisión: *Rechaza  $H_0$  si  $F_c > F_t$  o si  $p\_valor < \alpha$*
- Se calcula la suma cuadrática de los factores y las interacciones (en lugar de la suma cuadrática de los tratamientos y los bloques)

Para calcular la suma debida a una posible interacción utilice la siguiente fórmula:

$$SSI = \frac{n}{bk} \left[ \sum \sum (\bar{X}_{ij} - \bar{X}_i - \bar{X}_j + \bar{X}_G)^2 \right]$$

Donde:

- $i$  subíndice que representa una ruta
- $j$  subíndice que representa a un conductor
- $k$  número de niveles del **factor A** (efecto de la ruta)
- $b$  número de niveles del **factor B** (efecto del conductor)
- $n$  números de observaciones (datos)
- $\bar{X}_{ij}$  Tiempo de recorrido medio en la ruta  $i$  por el conductor  $j$ . Ver figura 12. Gráfica de la interacción de los tiempos de viaje
- $\bar{X}_i$  Tiempo de recorrido medio por la ruta  $i$
- $\bar{X}_j$  Tiempo de recorrido medio por el conductor  $j$
- $\bar{X}_G$  media total

Calculado del  $SSI$ , se procede a calcular el  $SSE$ , de la siguiente manera:

$$SSE = SS_{total} - SS_{factor-A} - SS_{factor-B} - SSI \quad (8)$$

La tabla ANOVA con interacción es:

| Fuente      | Suma cuadrática | gl               | Media cuadrática             | F         |
|-------------|-----------------|------------------|------------------------------|-----------|
| Ruta        | Factor A        | $k - 1$          | $SSA/(k - 1) = MSA$          | $MSA/MSE$ |
| Conductor   | Factor B        | $b - 1$          | $SSB/(b - 1) = MSB$          | $MSB/MSE$ |
| Interacción | SSI             | $(k - 1)(b - 1)$ | $SSI/[(k - 1)(b - 1)] = MSI$ | $MSI/MSE$ |
| Error       | SSE             | $n - kb$         | $SSE/(n - kb) = MSE$         |           |
| Total       | SS total        | $n - 1$          |                              |           |

**Figura 13:** Anova con interacción

# Anova de interacción en Excel

| Análisis de varianza de dos factores con varias muestras por grupo |      |          |         |         |            | ANÁLISIS DE VARIANZA |                           |                   |                    |                           |            |                |                      |
|--------------------------------------------------------------------|------|----------|---------|---------|------------|----------------------|---------------------------|-------------------|--------------------|---------------------------|------------|----------------|----------------------|
|                                                                    |      |          |         |         |            |                      | Origen de las variaciones | Suma de cuadrados | Grados de libertad | Promedio de los cuadrados | F          | Probabilidad d | Valor crítico para F |
| RESUMEN                                                            | US_6 | West_End | Hickory | Ruta_59 | Total      | Conductor            | Muestra                   | 359,1             | 4                  | 89,775                    | 20,877907  | 2,3394E-09     | 2,60597495           |
| Deans                                                              |      |          |         |         |            | Rutas                | Columnas                  | 218,4             | 3                  | 72,8                      | 16,9302326 | 2,9727E-07     | 2,8387454            |
| Cuenta                                                             | 3    | 3        | 3       | 3       | 12         |                      | Interacción               | 110,1             | 12                 | 9,175                     | 2,13372093 | 0,03643096     | 2,0034594            |
|                                                                    |      |          |         |         |            |                      | Dentro del grupo          | 172               | 40                 | 4,3                       |            |                |                      |
| Suma                                                               | 54   | 51       | 63      | 66      | 234        |                      | Total                     | 859,6             | 59                 |                           |            |                |                      |
| Promedio                                                           | 18   | 17       | 21      | 22      | 19,5       |                      |                           |                   |                    |                           |            |                |                      |
| Varianza                                                           | 9    | 9        | 1       | 9       | 9,72727273 |                      |                           |                   |                    |                           |            |                |                      |
|                                                                    |      |          |         |         |            |                      |                           |                   |                    |                           |            |                |                      |
| Snaverly                                                           |      |          |         |         |            |                      |                           |                   |                    |                           |            |                |                      |
| Cuenta                                                             | 3    | 3        | 3       | 3       | 12         |                      |                           |                   |                    |                           |            |                |                      |
| Suma                                                               | 48   | 69       | 69      | 66      | 252        |                      |                           |                   |                    |                           |            |                |                      |
| Promedio                                                           | 16   | 23       | 23      | 22      | 21         |                      |                           |                   |                    |                           |            |                |                      |
| Varianza                                                           | 7    | 7        | 1       | 4       | 12,7272727 |                      |                           |                   |                    |                           |            |                |                      |

Figura 14: Anova con interacción en Excel

Dado que  $p\_valor < \alpha$  se rechaza  $H_0$ , por lo que se concluye que **La combinación de la ruta y el conductor tiene un efecto significativo en la variable respuesta, que es el tiempo de recorrido**

## Considere los siguientes aspectos:

- Los efectos de la interacción proporcionan información acerca de los efectos combinados de las variables.
- Si está presente la interacción, se deberá efectuar una prueba ANOVA de una vía para probar diferencias entre las medias del factor por cada nivel del otro factor
- Este análisis requiere tiempo y esfuerzo, pero los resultados son muy interesantes.
- El análisis continúa con una ANOVA de una vía por cada conductor para probar la hipótesis:  $H_0$ : **Los tiempos de recorrido de las rutas son iguales.**

# Anova de una vía para cada conductor

| Deans: $H_0$ : Los tiempos de recorrido de las rutas son iguales.   |    |        |       |      |       | Snaverly: $H_0$ : Los tiempos de recorrido de las rutas son iguales. |    |        |       |      |       |
|---------------------------------------------------------------------|----|--------|-------|------|-------|----------------------------------------------------------------------|----|--------|-------|------|-------|
| Fuente                                                              | DF | SS     | MS    | F    | P     | Fuente                                                               | DF | SS     | MS    | F    | P     |
| Deans RTE                                                           | 3  | 51.00  | 17.00 | 2.43 | 0.140 | SN RTE                                                               | 3  | 102.00 | 34.00 | 7.16 | 0.012 |
| Error                                                               | 8  | 56.00  | 7.00  |      |       | Error                                                                | 8  | 38.00  | 4.75  |      |       |
| Total                                                               | 11 | 107.00 |       |      |       | Total                                                                | 11 | 140.00 |       |      |       |
| Ormson: $H_0$ : Los tiempos de recorrido de las rutas son iguales.  |    |        |       |      |       | Zollaco: $H_0$ : Los tiempos de recorrido de las rutas son iguales.  |    |        |       |      |       |
| Fuente                                                              | DF | SS     | MS    | F    | P     | Fuente                                                               | DF | SS     | MS    | F    | P     |
| Ormson RTE                                                          | 3  | 51.00  | 17.00 | 3.78 | 0.059 | Z-RTE                                                                | 3  | 86.25  | 28.75 | 8.85 | 0.006 |
| Error                                                               | 8  | 36.00  | 4.50  |      |       | Error                                                                | 8  | 26.00  | 3.25  |      |       |
| Total                                                               | 11 | 87.00  |       |      |       | Total                                                                | 11 | 112.25 |       |      |       |
| Filbeck: $H_0$ : Los tiempos de recorrido de las rutas son iguales. |    |        |       |      |       |                                                                      |    |        |       |      |       |
| Fuente                                                              | DF | SS     | MS    | F    | P     |                                                                      |    |        |       |      |       |
| Filbeck RTE                                                         | 3  | 38.25  | 12.75 | 6.38 | 0.016 |                                                                      |    |        |       |      |       |
| Error                                                               | 8  | 16.00  | 2.00  |      |       |                                                                      |    |        |       |      |       |
| Total                                                               | 11 | 54.25  |       |      |       |                                                                      |    |        |       |      |       |

**Figura 15:** Anova de una vía para cada conductor

## Aspectos finales a considerar

- Recuerde que los resultados de ANOVA de dos vías sin interacción (página 433). En ese análisis, los resultados mostraron en forma clara que el factor “ruta” **tenía un efecto significativo** en el tiempo de recorrido.
- Sin embargo, ahora que se incluye el efecto interacción, los resultados muestran que, **por lo general, la conclusión no es verdadera**
- Los *p\_valor* de Filbeck, Snaverly y Zollaco; son menores que  $\alpha$ , lo que indica que los **tiempos de recorrido medio son distintos**, los de Deans y Ormson, **no difieren de manera significativa**
- Pregunta clave: ¿Por qué existen estas diferencias?  
Respuesta: *Se requiere otra investigación sobre los hábitos de conducción de los cinco conductores*

- **ANOVA de dos vías con interacción, muestra el poder del análisis estadístico.**
- En este análisis se demostró el **efecto combinado del conductor y la ruta** *sobre el tiempo de recorrido*, y **también que los distintos conductores**, en efecto, se comportan de manera diferente cuando recorren sus rutas.
- Conocer los efectos de la interacción es muy importante en muchas aplicaciones, desde áreas científicas, como agricultura y control de calidad, hasta campos gerenciales, como administración de recursos humanos y equidad de género en las tabulaciones salariales y evaluaciones de desempeño.

**¿Preguntas?**

- Autoevaluación 12-5
- Ejercicios Cap. 12: 19 al 22
- Estadística para la Administración y Negocios 15<sup>º</sup> Edición

¿Preguntas?

- Ejercicios Cap. 12: 23 al 48 - Impares
- Estadística para la Administración y Negocios 15<sup>º</sup> Edición