Daena: International Journal of Good Conscience. 5(2) 170-218. Octubre 2010.
ISSN 1870-557X

Representatividad Estadística en las Ciencias Sociales

(Statistical Representativeness in Social Sciences)

Badii, M.H., A. Guillen y J.L. Abreu*

Palabras claves. Métodos estadísticos, pruebas de hipótesis

Resumen. Se presentan de forma simple, breve y por medios de los ejemplos prácticos, los métodos estadísticos que sirven para demostrar la representatividad de una, dos o varias muestras basada en los parámetros (la media y varianza) poblacionales (modelos X^2, t, F, Bartlett y ANOVA). Se indican la manera de reducir el tamaño del error experimental para aclarar el efecto significativo de los grupos o muestras bajo el interés (Diseño Completamente Aleatorio, Diseño de Bloques Aleatorios, Diseño de Cuadros Latinos). Además, se demuestran la forma de eliminar el efecto de las variables auxiliares o colaterales causante del ruido o error en la estadística (Análisis de Covarianza). También se nota el uso del modelo ANOVA Anidada para indicar la forma de segmentar la ANOVA en varias partes y de esta manera poder contestar preguntas más complejas. Finalmente, se describen e explican la forma de optimizar los recursos en la conducción de los experimentos y descubrir el efecto de la interacción entre las variables (Análisis Factorial).

Keywords. Hypothesis testing, statistical methods

Abstract. Statistical methods that serve to represent, based upon population parameters of mean and variance, for one, tow or various samples taken from study populations (X^2, t, F, Bartlett and ANOVA) are discussed. Ways to reduce the amount of experimental error in order to clarify the real effect of the treatments are demonstrated (Completely Randomized Design, Randomized Block Design, and Latin Square Design). The use of Covariance Analysis to show the pure effect of treatment free from the effect of collateral variables is also indicated. Nested ANOVA that allows division of ANOVA in several parts in order to answer complex questions is described and explained. Finally, the utility of Factorial Analysis for optimization of resource use as well as the illustration of interaction effects are noted and analyzed.

Introducción

En nuestro mundo, ocurren cosas, en forma de fenómenos, procesos, objetos, eventos, etc. de manera espacio-temporal. Ahora bien, la ocurrencia de todas estas cosas en el tiempo y en el espacio no es de manera determinísticos con 100% de exactitud. De hecho con la excepción de la muerte y el pago de impuestos, todos los demás cosas tienen una cierta probabilidad (entre cero y 100 por ciento) de ocurrir, por tanto, podemos deducir que las cosas mayoritariamente suceden bajo las leyes de la probabilidad. La estadística es precisamente la ciencia que trata de verificar la validez probabilísticos de todas las cosa, de allí la relevancia de esta rama de las ciencias en las tomas de decisiones prácticamente, en todo lo que concierna a nosotros en la vida (Ostle, 1986; Spiegel, 1991; Snedecor & Cochran; Montgomery, 2001).

En el campo de las ciencias sociales, algunas de las preguntas frecuentes son, por ejemplo:

1. ¿Si la opinión probable de la gente expresada por medio de de uno o varios sondeos reflejará su votación real hacia cierto partido político?
2. ¿Si existe diferencia probabilística entre la eficiencia de un determinado juzgado en comparación con otro juzgado basado en sondeos realizados para este fin?
3. ¿Si las distintas escuelas del derecho conllevan a diferentes niveles de eficacia y/o eficiencia en la hora de emitir dictámenes judiciales por diferentes jueces estimadas basadas en muestreos aleatorios realizados para este fin?
4. ¿Si por medio del muestreo se puede confiar si las diferencias entre los votos son realmente debido a la diferencia de los votos entre las diferentes urnas o estas diferencias están influenciadas por el conteo de los votos durante diferentes horas después de emitir las votas por los votantes?
5. ¿Si los sondeos o muestreos demuestran diferencias entre el número de los divorcios en diferentes meses del año o entre diferentes regiones del país o entre diferentes grupos étnicos?
6. ¿Si en base a los muestreos aleatorios se puede detectar diferencias reales en eficacia entre las secretarias de los juzgados por un lado y los jueces de los juzgados de distintas distritos de un determinado Estado?

Para contestar estas y preguntas similares en término estadístico, es decir para poder verificar la validez estadística de estas preguntas, existen modelos estadísticos cuya selección depende de la habilidad del investigador al emplear el modelo adecuado según la naturaleza de la pregunta. Por tanto, el propósito de esta investigación es presentar por medio de los ejemplos, de manera simple el uso de los métodos estadísticos la noción de la representatividad en las ciencias sociales con particular énfasis en el campo de derecho.

Concepto

El concepto de la representatividad en la estadística indica si la muestra o las muestras (Badii & Castillo, 2009) representan una misma población. Para verificar esta noción se toman en cuenta dos factores. 1) Si se trata de una, dos o más de dos muestras. 2) Si el parámetro poblacional bajo el estudio es la media "*m*" o la varianza poblacional "*v*". Por tanto, en base a la combinación de estas dos factores tendremos seis modelos distintos para poder verificar o constatar las probables preguntas (Tabla A).

En resumen se puede utilizar el siguiente enunciado para poder seleccionar cualquier de los seis modelos de la Tabla "A". ***Verificar si la muestra, 2 muestras,***

o "n" muestras proceden de, o representan la misma población en base a la media "m" o la varianza "v".

Tabla A. Generación de seis modelos para comprobar la representatividad estadística (a: *p* = parámetro poblacional, *v* = varianza, *m* = media).

p^a	Número de muestras		
	1	**2**	**>2**
v	**<u>Modelo</u>: X^2** **Objetivo:** Comparar la varianza de la población (*V*p) y la varianza de la muestra (*V*m). **<u>Ecuación</u>: $X^2 = (Vp/Vm)(gl_m)$** Donde, gl_m = Grado de libertad de la muestra **<u>Juego de hipótesis</u>:** **Ho: $V_m = V_p$** **Ha: $V_m \neq V_p$** Donde, Ho = Hipótesis nula, Ha = Hipótesis alterna	**<u>Modelo</u>: F_{Fisher}** **Objetivo:** Comparar la varianza de la muestra uno (V_1) y la varianza de la muestra dos (V_2). **<u>Ecuación</u>:** **$F_{Fisher} = V_{mayor} / V_{menor}$** **<u>Juego de hipótesis</u>:** **Ho: $V_1 = V_2$** **Ha: $V_1 \neq V_2$**	**<u>Modelo</u>: F_{Max}** **Objetivo:** Comparar las varianzas de más de dos muestras, es decir, V_1, V_2, …, V_n. **<u>Ecuación</u>:** **$F_{Max} = V\text{max} / V\text{min}$** **<u>Juego de hipótesis</u>:** **Ho: $V_1 = V_2 = \ldots = V_n$** **Ha: Al menos una "*V*" es distinta** **<u>Modelo</u>: Bartlett (ver el texto)**
m	**<u>Modelo</u>: t_1** **Objetivo:** Comparar la media de la población (m_p) y la media de la muestra (m_m). **<u>Ecuación</u>: $t_1 = (m_m - m_p) / EE$** Donde, $EE = (V/n)^{1/2}$ *EE*= Error estándar muestreal *V* = Varianza de la muestra *n* = Tamaño de la muestra **<u>Juego de hipótesis</u>:** **Ho: $m_m = m_p$** **Ha: $m_m \neq m_p$**	**<u>Modelo</u>: t_2** **Objetivo:** Comparar la media de la muestra uno (m_1) y la media de la muestra dos (m_2). **<u>Ecuación</u>: $t_2=(m_1–m_2)/EE_p$** Donde, $EE_p=[(V_p/n_1) + (V_p/n_2)]^{1/2}$ $V_p=(SC_1 + SC_2)/(gl_1 + gl_2)$ *EE*=Error pool muestreal *Vp* = Varianza ponderada *SC* = Suma de cuadrados *gl* = Grados de libertad **<u>Juego de hipótesis</u>:** **Ho: $m_1 = m_2$** **Ha: $m_1 \neq m_2$**	**<u>Modelo</u>: ANOVA** **Objetivo:** Comparar las medias de más de dos muestras, es decir, m_1, m_2, …, m_n. **<u>Ecuación</u>:** **$F = V_{tratamiento} / V_{Error}$** Para las ecuaciones, ver la Tabla de ANOVA en la sección de "Análisis de varianza" **<u>Juego de hipótesis</u>:** **Ho: $m_1 = m_2 = \ldots = m_n$** **Ha: Al menos una "m" es distinta.**

Prueba de "X^2"

Definición. Distribución probabilística para determinar si una sola varianza poblacional es igual a cierto valor, mientras que una distribución observada de frecuencias es significativamente diferente de la distribución esperada o teórica, y por último si la clasificación de una variable es independiente de la clasificación de otra. La población que se está muestreando se distribuye normalmente, el estadístico de prueba se distribuye como x^2 con (n-1) grados de libertad. De acuerdo a esto el valor de x^2 puede compararse con el valor crítico de chi cuadrada en la tabla con la finalidad de probar la hipótesis.

Objetivo. Probar la igualdad de varianza de una población normal y la varianza de una muestra extraída de esa población.

Ejemplo. Supongamos el siguiente caso donde se trata de verificar la representatividad de cada muestra en función de la varianza para cada uno de los 3 casos siguientes.
Datos: Datos son resultados de las muestras extraídas de poblaciones en los Estados del *norte* (Tabla 1), *centro* (Tabla 2) y *sur* (Tabla 3) de México y representan en cada Estado los promedios de divorcios ocurridos en 25 matrimonios después de 5 años de estar casados.

Caso A (ejemplo1). Norte de México:
Verificar si la muestra procede de una población cuya varianza es igual a 12 (Tabla 1).

Tabla 1. Promedios de divorcios en 25 matrimonios en Estados del norte de México.

i (Estado)	X (Promedio de divorcios en 25 matrimonios)
Baja California	12.9
Coahuila	14.8
Chihuahua	11.6
Durango	10.4
Nuevo León	16.9
Tamaulipas	13.8

Ecuación:

$X^2 = (Vp/Vm)(gl_m)$

Donde,
$Vm = SC_x / gl$
$SC_x = \Sigma(x^2) - FC$
$FC = (\Sigma x)^2 / n$
SC_x = Suma cuadrados de "x"
FC = Factor de corrección
n = Tamaño de la muestra
gl = Grados de libertad = n -1

Juego de hipótesis:

Ho: Vm = Vp
Ha: Vm ≠ Vp
Donde,
Ho = hipótesis nula
Ha = Hipótesis alterna o de la investigación
V_p = Varianza poblacional o varianza hipotética
V_m = Varianza muestreal

Solución:

$X^2 = (v_{hipotética} / v_{muestra}) * gl_{muestra}$
$v = SC / gl$
$SC = (\sum X^2 - (\sum X)^2 / n$
$V = 5.37$
$X^2 = (v_{hipotética} / v_{muestra}) * gl_{muestra}$
$X^2 = (12 / 5.37) * 5$
$X^2 = 11.17$

$X^2_{Calculada}$ = 11.17. El valor de la $X^2_{Tabulada}$ a nivel de α igual a 0.05 es de 11.07. *Debido a que el valor calculado* (*11.7*) *es mayor que el valor tabulado* (*11.07*)*;* "*Ho*" *se rechaza, es decir, que la variabilidad de la muestra no es igual a la variabilidad de la población. Consecuentemente, la muestra del los Estados del norte de México no es representativa y no procede de la población bajo el estudio.*

Nota: ***Dentro del texto las notaciones de VC y VT*** **(*cuando se utilizan*)** ***indican los valores calculados*** **(en este ejemplo** X^2Calculada **)** ***y tabulados*** **(el el presente**

ejemplo X^2Tabulada)***, respectivamente, para cualquier modelo. Valor numérico de la tabla*** **(*VT*)** ***procede de la tabla correspondiente de "X^2", t o F en el Anexo 1.***

Caso B (Ejemplo 2). Centro de México:
Verificar si la muestra procede de una población cuya varianza es igual a 12 (Tabla 2).

Tabla 2. Promedios de divorcios en 25 matrimonios en Estados del centro de México.	
i (Estado)	**X (Promedio de divorcios en 25 matrimonios)**
Distrito Federal	22.1
Guanajuato	7.4
Hidalgo	8.0
Jalisco	12.4
Estado de México	11.5
Morelos	12.4
Puebla	10.2
Querétaro	12.7

Solución:

$X^2 = (v_{hipotética} / v_{muestra}) * gl_{muestra}$
$v = SC / gl$
$SC = (\sum X^2 – (\sum X)^2 / n$
$V = 20.54$
$X^2 = (v_{hipotética} / v_{muestra}) * gl_{muestra}$
$X^2 = (12 / 20.54) * 7$
$X^2 = 4.09$

$X^2_{Calculada}$ = 4.09. El valor de la $X^2_{Tabulada}$ a nivel de α igual a 0.05 es de 14.06. *Debido a que el valor calculado* (*4.09*) *es menor que el valor tabulado* (*14.06*)*;* "*Ho*" *se acepta, es decir, que la variabilidad de la muestra es igual a la variabilidad de la población. Consecuentemente, la muestra del los Estados del centro de México es representativa y procede de la población bajo el estudio.*

Caso C (ejemplo 3). Sur de México:
Verificar si la muestra procede de una población cuya varianza es igual a 12 (Tabla 3).

Tabla 3. Promedios de divorcios en 25 matrimonios en Estados del sur de México.	
i (Estado)	**X (Promedio de divorcios en 25 matrimonios)**
Campeche	11.3
Chiapas	6.7
Oaxaca	6.5
Quintana roo	12.0
Tabasco	10.4
Yucatán	9.7
Guerrero	8.9

Solución:

$X^2 = (v_{hipotética} / v_{muestra}) * gl_{muestra}$
$v = SC / gl$
$SC = (\sum X^2 – (\sum X)^2 / n$
$V = 4.75$

$X^2 = (v_{hipotética} / v_{muestra}) * gl_{muestra}$
$X^2 = (12 / 4.75) * 6$
$X^2 = 15.76$

$X^2_{Calculada} = 15.76$. El valor de la $X^2_{Tabulada}$ a nivel de α igual a 0.05 es de 12.59. *Debido a que el valor calculado* (*15.76*) *es mayor que el valor tabulado* (*12.59*)*;* "*Ho*" *se rechaza, es decir, que la variabilidad de la muestra no es igual a la variabilidad de la población. Consecuentemente, la muestra del los Estados del sur de México no es representativa y no procede de la población bajo del estudio.*

Prueba de F

Definición. La prueba *f* se utiliza principalmente para probar la igualdad entre dos varianzas. A diferencia de las distribuciones t_1 y chi cuadrada, donde cada una de estas tiene solamente un parámetro, la distribución *f* tiene dos parámetros: los números de grados de libertad (gl) denotados por lo general gl_1 y gl_2. Una distribución específica de *f* o la forma específica de la curva de *f*, está completamente determinada cuando se conoce gl_1 y gl_2.
La prueba de Fisher se utiliza cuando se tienen dos variables nominales . Esta prueba es más precisa que la prueba de Ji-cuadrado o el G-test de independencia cuando los números esperados son pequeños.

Requisitos. El estadístico de prueba *f* toma solamente valores no negativos debido a que tanto el numerador como el denominador de la razón *f* son varianzas, las cuales no pueden ser valores negativos.

Objetivo. Determinar la igualdad entre dos varianzas muestréales.

Ecuación:

$$\mathbf{F_{Fisher} = V_{mayor} / V_{menor}}$$

Donde,
V_{mayor} = Varianza mayor
V_{menor} = Varianza menor

Juego de hipótesis:

Ho: $V_1 = V_2$
Ha: $V_1 \neq V_2$

Donde,
V_1 = varianza de la primea muestra
V_2 = varianza de la segunda muestra

Ejemplo 4. Verificar que las siguientes dos muestras (datos de las prisiones "**A**" y "**B**") proceden de la misma población basada en sus varianzas (Tabla 4).

Tabla 4. Meses de sentencia de los reos en las prisiones "A" y "B" (-: no hay datos para el reo).

i (reo)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
X (años en prisión A)	18	15	16	17	20	23	24	19	25	21
X (años en prisión B)	1.0	1.9	2.1	2.8	3.4	2.5	1.2	1.6	-	-

Juego de hipótesis:

Ho: $V_1=V_2$
Ha: $V_1 \neq V_2$

Solución:

V_1 = 11.7333
V_2 = 0.6626
F_{Fisher} = 11.7333 / 0.6626 = **17.7079**

VC vs VT
ɑ: 0.05
gln: 10-1 =9
gld: 8-1=7

17.7079 > 3.68 :. "Ho" Se rechaza.

Conclusión. Las dos muestras no proceden de la misma población en base a sus varianzas.

Ejemplo 5. Verificar que las dos muestras siguientes (datos de las prisiones "**A**" y "**C**") (Tabla 5) proceden de la misma población basada en sus varianzas.

Tabla 5. Meses de sentencia de los reos en las presiones "A" y "C" (-: no hay datos para el reo).

i (# de reo)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
X (meses en prisión "A")	18	15	16	17	20	23	24	19	25	21
X (meses en prisión "C")	105	147	120	136	141	110	-	-	-	-

Juego de hipótesis:

Ho: $V_1 = V_3$
Ha: $V_1 \neq V_3$
Donde,
V_1 = varianza de la primea muestra
V_3 = varianza de la tercera muestra

Solución:

V_1 = 11.7333, V_3 = 299.5
F_{Fisher} = 299.5/ 11.7333 = **25.5256**

VC vs VT
ɑ: 0.05
gln: 6-1 =5
gld: 10-1=9

25.5256 > 3.48 :. "Ho" Se Rechaza.

Conclusión. Las dos muestras no proceden de la misma población en base a su varianza y por tanto no son representativas de la misma población.

Ejemplo 6. Verificar que las dos muestras siguientes (datos de las prisiones "**B**" y "**C**") proceden de la misma población basada en sus varianzas (Tabla 6).

Tabla 6. Meses de sentencia de los reos en las prisiones "B" y "C" (-: no hay datos para el reo).

i (# de reo)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
X (meses en prisión "B")	1.0	1.9	2.1	2.8	3.4	2.5	1.2	1.6
X (años en prisión "C")	105	147	120	136	141	110	-	-

Juego de hipótesis:

Ho: $V_2 = V_3$
Ha: $V_2 \neq V_3$
Donde,
V_2 = varianza de la segunda muestra
V_3 = varianza de la tercera muestra

Solución:

$V_2 = 0.6626$, $V_3 = 299.5$
$F_{Fisher} =$ 299.5 / 0.6626 = **452.0072**

VC	vs	VT	
452.0072	>	3.97	$\left\{ \begin{array}{l} \alpha\text{: } 0.05 \\ \text{gln: } 6\text{-}1 \\ =5 \end{array} \right.$

:. "Ho" Se Rechaza.

Conclusión. Estas dos muestras no son representativas, es decir, no reflejan la misma población, en otras palabras, no proceden de la misma población en base a sus varianzas.

Modelo de F_{Max}

El modelo de F_{max} es una extensión del modelo de F_{Fisher}, para comparación de más de 2 muestras, con los mismos supuestos de F_{Fisher} y operativamente funciona como este modelo. Acordar que el modelo F_{Fisher} se trata de la división de 2 varianzas, es decir la varianza mayor entre la varianza menor. En caso del modelo de Fmax, se arreglan las varianzas en orden de sus magnitudes, y luego se dividen el valor extremo máximo entre el valor extremo mínimo, por tanto, la ecuación de este modelo sería como siguiente:

$$\mathbf{F_{max} = V_{Max} / V_{Min}}$$

Donde.
$\mathbf{V_{Max}}$ = Varianza máxima
$\mathbf{V_{Max}}$ = varianza mínima

Juego de hipótesis para el modelo de $\mathbf{F_{max}}$:

Ho: $V_{norte} = V_{centro} = V_{sur}$
Ha: No todas las varianzas son iguales (al menos una varianza es distinta de cualquiera otra varianza)

En caso del ejemplo colectivo (Tablas 1, 2 y 3), al compara las varianzas de 3 muestras del norte, centro y sur de México, nos damos cuenta que de las 3 muestras, la varianza del centro tiene el máximo valor (20.54) y la varianza del sur tiene el mínimo valor (4.57). Por tanto:

$$\mathbf{F_{max} = V_{Max} / V_{Min}}$$
$$\mathbf{Fmax = 20.54 / 4.57 = 4.49}$$

Comparando este valor calculado (4.49) con el valor tabulado con 7 y 6 grados de libertad llegamos con el valor de 4.21, lo cual es menor que el valor calculado de la F, y por tanto, “Ho” se rechaza, es decir, las 3 muestras no representan una misma población, sino que al menos una muestra es distinta de cualquier otra muestra, en este caso la muestra del centro es diferente de la muestra del sur del país.

Modelo de Bartlett

Definición. Esta prueba se usa para determinar si las varianzas de *n* poblaciones son homogéneas; es una prueba aproximada de Chi cuadrada, sensible a la normalidad, es decir, no es robusta si las muestras de las poblaciones a las cuales se aplicaron los tratamientos no se distribuyen normalmente.

La prueba de Bartlett parte del supuesto que los datos están asociados con los grupos, cada uno de los cuales tiene un distinto conjunto de observaciones independientes.
Al respecto, se debe considerar que la varianza corresponde a la suma de las diferencias de los valores individuales en relación con el promedio, elevadas al cuadrado y divididas entre los grados de libertad, es decir, son variaciones alrededor de la medida de tendencia central, representativa de la muestra con la cual se estudia un fenómeno, sin embargo, no se puede saber si esas variaciones se deben a errores dados por el fenómeno en sí o a errores del observador o del método para efectuar las mediciones.
Cuando existen tres o más grupos de población que se desea comparar y las varianzas son iguales, se puede considerar que la fuente de error es la misma, en caso contrario, si son desiguales, se tiene la probabilidad de que otra fuente desconocida de error en alguna de las muestras intervenga desfavorablemente en los resultados del análisis estadístico.
En las tareas de la investigación científica es poco probable que al medir las observaciones en tres o más muestras, la variación sea idéntica. Por lo general, esas varianzas tienen magnitudes diferentes y el investigador cae en la incertidumbre de decidir si las fuentes de error fueron las mismas o si intervinieron uno o más agentes de variación.
En resumen este modelo funciona como el modelo de Fmax, sin embargo presente mayor precisión que el Fmax.

Requisitos. La prueba de Bartlett, se basa en una estadística cuya distribución muestreal proporciona valores críticos exactos cuando los tamaños muéstrales son iguales. Estos valores críticos para tamaños iguales de muestras también se pueden utilizar para obtener aproximaciones altamente precisas de los valores críticos para tamaños diferentes de las muestras.

Objetivo. La prueba de Bartlett permite saber, en función de la probabilidad, si la discrepancia entre varianzas fue dada por el azar o por otros factores de error no deseados por el experimentador.

Ecuaciones:

$$\mathbf{B = 2.303\ [(\mathit{log}Vp)(\Sigma \mathit{gl_i}) - (\Sigma \mathit{log}V_i * \mathit{gl_i})]}$$
$$\mathbf{V_p = \Sigma SC_i / \Sigma \mathit{gl_i}}$$
$$\mathbf{SC_i = \Sigma x_i^2 - (\Sigma x_i)^2 / N_T}$$
$$\mathbf{C = 1 + 1/3(\mathit{k}-1)\ [\Sigma(1/\mathit{gl}_i) - 1/\Sigma \mathit{gl}_i]}$$
$$\mathbf{B_C = B / C}$$

Donde,
B = Valor del modelo de Bartlett
V_i = Varianza del grupo "i"
V_p = Varianza ponderada
C = Factor de corrección para el modelo de Bartlett
B_c = Valor corregido del modelo de Bartlett
SC_i = Suma de cuadrados para el grupo "i"
gl_i = Grados de libertad del grupo "i"

Juego de hipótesis:

Ho: Todas las varianzas son iguales
Ha: No es así, es decir, al menos una varianza varía de cualquier otra varianza

Ejemplo 7. Verificar si cuatro Estados (muestras) representan la misma población en base a sus varianzas (Tabla 7). Datos son número de votos electorales (por miles) de cada uno de 5 municipios de cada uno de 4 Estados. Note que el Estado "C" tiene solamente 4 municipios.

Tabla 7. Meses de prisión por cada reo en cada una de cuatro prisiones diferentes.

Repetición (# de municipios)	**Estados**			
	A	**B**	**C**	**D**
1	60.8	68.7	102.6	87.9
2	57.0	67.7	102.1	84.2
3	65.0	74.0	100.2	83.1
3	58.6	66.3	96.5	85.7
5	61.7	69.8	- (a)	90.3
Σx_i	**303.1**	**346.5**	**401.4**	**431.2**
Σx_i^2	**18,411.49**	**24,046.71**	**40,303.46**	**37,220.24**

(a) *: No hay datos, es decir, no hay quinto municipio en el estado "C"*

Juego de hipótesis:

Ho: Todas las varianzas son iguales
Ha: No es así, es decir, al menos una varianza varía de cualquier otra varianza

Solución:

SC_1 = 18,411.49 – $(303.1)^2$ / 5 = 18,411.49 – 91,869.61/5 = 18,411.49-18,373.922 = 37.57

SC_2 = 24,046.71 – $(346.5)^2$ / 5 = 24,046.71 – 120, 062.25/5 = 24,046.71 – 24,012.45 = 34.26
SC_3 = 40,303.46 – $(401.4)^2$ / 4 = 40,303.46 – 161,121.96/4 = 40,303.46-40,280.49 = 22.97
SC_4 = 37,220.24 – $(431.2)^2$ / 5 = 37,220.24 – 185,933.44/5 = 37,220.24- 37,186.68 = 33.56

Varianza	***log*V*gl**
V1= 37.57/ 4= 9.3925	0.9727*4= 3.8911
V2= 34.26/ 4= 8.565	0.9327*4=3.7309
V3= 22.97/ 3= 7.6566	0.8840*3=2.6521
V4= 33.56/ 4= 8.39	0.9237*4=3.6950
	Σ *log*V**gl* = 13.9691

$\mathbf{V_p = \Sigma SC_i / \Sigma \mathit{gl_i}}$ = Vp = (37.57 + 34.267 +22.97 + 33.56) / (4+4+3+4) = 128.36/15 = **8.5573**
$\mathbf{logV_p}$ = log 8.5573 = **0.9323**
B = 2.303 [(*log*Vp)(Σ*gl*$_i$) – (Σ*log*V$_i$gl*$_i$)] =** 2.303 [(0.9323)(15) – (13.9691)] = 2.303 [13.9845-13.9691] = 2.303(0.0154) = **0.0354**
C = 1 + 1/3(*k*-1) [Σ(1/*gl*$_i$) - 1/Σ*gl*$_i$] = 1 + 1/3(4-1) [(1/4 + 1/4 + 1/3 + 1/4) – (1/15)] = **1.1296**
BC = **0.0354/1.1296** = **0.0313**

BC vs VT $\left\{ \begin{array}{l} \alpha\text{: } 0.05 \\ gl\text{: } 4-1= 3 \end{array} \right.$

0.0313 < 7.815 :. "Ho" Se Acepta todas las varianzas son iguales.

Conclusión. Las 4 muestras son representativas y por tanto, proceden de una misma población en base a sus varianzas.

Prueba de t Student

Definición. En la investigación en derecho es común obtener información de campo, resultante de un estudio de muestreo, o derivada de un proceso experimental cuyo interés es comparar dos poblaciones o dos tipos de tratamientos. Así se compara la solución de problemas de índole civil, mercantil, laboral, penal, etc. con sistemas tradicional versus sistemas de manejo de procedimientos nuevos. En este proceso la aleatorización juega un papel fundamental evitando sesgos en las comparaciones, derivados de la asignación aleatoria de los tratamientos a las unidades experimentales da origen a conjuntos independientes de observaciones de la variable aleatoria de interés, cuyo valor es imposible de conocer antes de realizar el experimento.

Requisitos. La prueba t se supone que las observaciones dentro de cada grupo se distribuyen normalmente y las varianzas son iguales en los dos grupos.

Objetivo. El objetivo de la prueba t de student es comprobar si la media de la población es igual a la media de una muestra de datos tomada de la misma población.

Ecuación:

$$\mathbf{t_1 = (m_m - m_p) / EE}$$

Donde,
m_m = Media muestreal
m_p = Media poblacional
EE = Error estándar = $(v/n)^{1/2}$

Juego de hipótesis:

Ho: $m_m = m_p$
Ha: $m_m \neq m_p$

Donde,
m_m = media muestreal
m_p = media poblacional

Ejemplo 8. Verificar si la muestra de juicios familiares (Tabla 8) procede de una población cuya media poblacional es 22 meses.

Juego de hipótesis:

Ho: $m_m = m_p$
Ha: $m_m \neq m_p$

Tabla 8. Duración de cada uno de 10 juicios familiares en meses.										
i (# de juicio)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
X (duración en meses)	18	15	16	17	20	23	24	19	25	21

Solución:

$\Sigma x = 198$
V = SC / gl
V = 11.733
$m_m = 198/10 = 19.8$
$\mathbf{t_1 = (m_m - m_p) / EE} = (19.8 - 22) / (11.733/10)^{1/2} = \mathbf{-2.0310}$

VC vs VT $\begin{cases} \alpha: 0.05 \\ gl: 10\text{-}1=9 \end{cases}$

$|-2.0310| < 2.262$:. “Ho” Se Acepta.
$|\mathbf{-2.0310}|$ = Tomar valor absoluto de 2.0310

Conclusión.- La muestra es representativa, ya que procede de la misma población cuya media es 22 meses.

Ejemplo 9. Verificar si la muestra de juicios laborales (Tabla 9) procede de una población cuya media poblacional es 2.0 años.

Ho: $m_m = m_p$
Ha: $m_m \neq m_p$

Tabla 9. Duración de cada uno de 8 juicios laborales en meses.								
i (# de juicio)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
X (duración en meses)	1.0	1.9	2.1	2.8	3.4	2.5	1.2	1.6

Solución:

$\Sigma x = 16.5$, $m_m = 16.5/8 = 2.0625$
$\mathbf{t_1 = (m_m - m_p) / EE} = (2.0625 - 2.0) / (0.6626 / 8)^{1/2} = 0.0625 / 0.2877 = \mathbf{0.2172}$

VC vs VT $\begin{cases} \alpha: 0.05 \\ gl: 8\text{-}1=7 \end{cases}$

0.2172 < 2.365 :. “Ho” Se Acepta.

Conclusión. La muestra de juicios laborales es una muestra representativa y refleja la misma población cuya media es 2.0 años.

Ejemplo 10. Verificar si la siguiente muestra de juicios penales (Tabla 10) procede de una población cuya media poblacional es 120 semanas.

Ho: $m_m = m_p$
Ha: $m_m \neq m_p$

Tabla 10. Duración de seis juicios penales en semanas.						
i (# de juicio)	**1**	**2**	**3**	**4**	**5**	**6**
X (Duración en meses)	105	147	120	136	141	110

Solución:

$\Sigma x = 759$, $m_m = 759/6 = 126.5$

$\mathbf{t_1 = (m_m - m_p) / EE} = (126.5 - 120) / (299.5 / 6)^{1/2} = 6.5 / 7.0651 = \mathbf{0.92}$

VC vs VT $\begin{cases} \alpha\text{: } 0.05 \\ gl\text{: } 6\text{-}1\text{=}5 \end{cases}$

0.920 < 2.571 :. "Ho" Se Acepta.

Conclusión.- La muestra de juicios penales es una muestra representativa y procede de la misma población cuya media es igual a 120 semanas.

Prueba de t Student para 2 muestras

Definición. En algunas investigaciones en derecho, el investigador estudia los mismos jueces para probar diferentes tipos de tratamientos, solamente desplazando su uso en diferentes tiempos, tal es el caso de aquellos estudios en los que se miden variables de percepción como juicios escritos, orales, etc., es decir, las observaciones sobre un mismo juez se encuentran correlacionadas, ya sea positiva o negativamente.

Requisitos. Para homogeneizar las condiciones en que se comparan los tratamientos, se requiere formar pares de unidades experimentales, de manera que las unidades que integran cada par sean tan semejantes entre sí como sea posible. Entonces, para probar una hipótesis de nulidad, es decir, de igualdad de medias de tratamiento, se requiere considerar la varianza de las diferencias entre los dos tratamientos, en lugar de las varianzas individuales.

Objetivo. El objetivo de la prueba t de student es comprobar la hipótesis que si 2 muestras tienen la misma media. Se puede realizar sobre muestras independientes o aparejadas.

Ejemplo 11. Verificar si dos muestras siguientes proceden de la mis población en base a sus medias (Tabla 11).

Tabla 11. Meses de sentencia de ocho y seis reos en la prisión "I" y "II", respectivamente.

i (# de reo)		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
X (meses en presión)	**Prisión I**	1.0	1.9	2.1	2.8	3.4	2.5	1.2	1.6
	Prisión II	105	147	120	136	141	110	-	-

- *: No hay datos, es decir, no hay dato sobre los reos 7 y 8 de la prisión "II"*

Ecuación: $\mathbf{t_2 = (m_1 - m_2) / EE_p}$

Donde,

$\mathbf{EE_p = [(V_p/n_1) + (V_p/n_2)]^{1/2}}$

$\mathbf{V_p = (SC_1 + SC_2)/(gl_1 + gl_2)}$

EE = Error ponderado muestreal

Vp = Varianza ponderada

SC = Suma de cuadrados

gl = Grados de libertad

Juego de hipótesis:

Ho: $m_1 = m_2$

Ha: $m_1 \neq m_2$

Donde: m_m = Media muestreal, m_p = Media de la población

Solución:

Prisión I: m = 2.06, SC = 1,497.5, gl = 5, n_1 = 8

Prisión II: M = 126.5, SC = 4.6388, gl = 7, n_2 = 6

Ecuación: $t_2 = (m_1 – m_2) / EE_p$

Donde,

$EE_p = [(V_p/n_1)+ (V_p/n_2)]^{1/2}$
$V_p = (SC_1 + SC_2)/(gl_1 + gl_2)$
EE = Error ponderado muestreal
Vp = Varianza ponderada
SC = Suma de cuadrados
gl = Grados de libertad

$V_p = (SC_1 + SC_2)/(gl_1 + gl_2)$ = (1,497.5 + 4.6388) / (5 +7) = **125.1782**
$EE_p = [(V_p/n_1)+ (V_p/n_2)]^{1/2}$ = $[(125.1782/5) + (125.1782/n_2)]^{1/2}$ = **6.0423**
$t_2 = (m_c – m_b) / EE_p$ = (126.5 – 2.06) / 6.0423 = **20.5948**

$$\text{VC} \quad \text{vs} \quad \text{VT} \begin{cases} \alpha\text{: } 0.05 \\ gl\text{: } gl_1+gl_2=5+7=12 \end{cases}$$

20.5948 > 2.179 :. "Ho" Se Rechaza.
Conclusión.- Las 2 muestras no son representativas y por tanto, no proceden de la misma población en base a sus medias.

Análisis de varianza (ANOVA)

Descripción. El análisis de varianza es una técnica estadística que consiste en dividir la variación total de una variable de respuesta, en un conjunto de datos, en fuentes de variación conocidas (tratamientos, bloques, hileras, columnas, etc.) y fuentes de variación desconocidas (error). Mediante este análisis es posible conocer la magnitud de las contribuciones de esas fuentes de variación con respecto a la variación total.

Objetivo. Sus objetivos son:
a) Probar hipótesis acerca de las medias de tratamientos (modelos de efectos fijos)
b) Estimar componentes de varianza (Modelos de efectos aleatorios).

Requisitos. En la estadística paramétrica, el análisis de varianza debe cumplir ciertas suposiciones, ya que de otra manera nos puede llevar a conclusiones erróneas.

Supuestos del modelo:

a) Existe homogeneidad de varianzas entre las distintas poblaciones a las cuales se aplicaron los tratamientos (Homocedasticidad).
b) Las observaciones de cada grupo o tratamiento fueron obtenidas en forma aleatoria de una población normal.
c) Los efectos de tratamientos son aditivos.
d) Las observaciones son independientes (no están correlacionadas).

El análisis de la varianza se puede realizar con tamaños muestréales iguales o distintos, sin embargo, es recomendable tamaños iguales por dos motivos: 1) Si el tamaño es igual, La *f* es insensible a pequeñas variaciones en la asunción de igual varianza. 2) Tamaño igual de muestra minimiza la probabilidad de error tipo II. Hay métodos no paramétricos de ANOVA (Badii et al., 2006).

Ecuaciones del modelo:

Las ecuaciones de ANOVA se indican en forma de la Tabla de ANOVA (Tabla B).

Tabla B. Tabla de ANOVA con las correspondientes columnas, hileras y ecuaciones.

FV	*gl*	SC	Varianza	$F_{Calculada}$
Tratamientos (grupos)	K – 1	$\Sigma[(T_{cg})^2/r_{cg}]$ – FC	SC_t / gl_t	V_T / V_E
Error	gl_T - gl_t	SC_T – SC_t	SC_E / gl_E	
Total	Nt - 1	$\Sigma\ x_i^2$ – FC		

FV = fuente de variación, K = número de grupos, Nt = número total de los elementos, gl_T = grados de libertad total, gl_t = grados de libertad del tratamientos, FC = factor de corrección = $(\Sigma x_i)^2/N_t$, SC_T = suma de cuadrados del total, SC_t = suma de cuadrados del tratamientos, V_T = varianza del tratamientos, V_E = varianza del error

Diseños comunes de ANOVA:

Existen varios tipos de diseños de ANOVA dependiendo de las características de la población bajo del estudio. Por ejemplo, si a parte de la variabilidad debido a los tratamientos no hay otra gradiente de la variabilidad, el diseño de la selección se denomina el Diseño Completamente aleatorio (DCA). Si existe (a parte de los tratamientos) una sola gradiente de la variabilidad, se utiliza el Diseño de Bloques al Azar (DBA). Si hay dos gradientes de variabilidad, se usa el Diseño de Cuadro

Latino (DCL). Este quiere decir, que en cada caso se bloquea el efecto del gradiente extra de variabilidad para poder determinar la diferencia pura entre los tratamientos sin que esta diferencia esté influenciada por cualquier otro factor o variabilidad. Los diseños de uso común en la estadística se indican en la Tabla C (Badii et al., 2004, 2009).

Tabla C. Diseños experimentales de uso común.

Nombre del diseño	**Rasgos**	**Ventajas**	**Eficiencia**
Diseño Completamente al azar *(*Badii et al., 2007b) (UE = Unidad Experimental)	Un factor (el tratamiento) y ningún otro gradiente de variabilidad	**1.** fácil de diseñar **2.** fácil de analizar **3.** diferentes # de repeticiones 4. máximo g.l. para el error	**100%**
Diseño de Bloques al azar (Badii et al., 2007a)	Un factor y un gradiente de variabilidad	**1.** reduce la varianza de error **2.** fácil de analizar **3.** más flexibilidad **4.** más precisión	**167%**
Diseño de Cuadro Latino (Badii et al., 2007a)	Un factor y 2 gradientes de variabilidad	**1.** reduce la varianza de error **2.** fácil de analizar **3.** más flexibilidad **4.** más precisión	**222%**
(a) ***Diseño factorial***: asignación al azar de Unidades Experimentales (UE's) **(b)** ***Parcelas divididas***: sin asignación al azar de las UE´s (Badii et al., 2007a)	Más de un factor y diferentes niveles para cada factor	**1.** más económico **2.** medir las interacciones	**288%**

Diseños multivariados (alto número de variables)	
1. ***Análisis de Componentes Principales***	**1.** Provee ordenación y el perfil jerárquica
2. ***Análisis Factor***	**1.** Reducir el número de las variables para el análisis
3. ***Análisis Discriminante*** (Badii et al., 2007d)	**1.** Agrupar en base a la diferencia **2.** Más riguroso con los supuestos de la normalidad
4. ***Análisis Cluster*** (Badii et al., 2007c)	**1.** Agrupar en base a la similitud **2.** Más robusto con los supuestos de la normalidad
5. ***LISREL*** (**LI**near **S**tructured **REL**ationship) (Rositas et al., 2007), ***EQ, AMOS, PLS Graph, PLS Smart***	**1.** Busca linealizar las interrelaciones entre las variables **2.** Intercambia las variables independientes a las dependientes y viceversa
6. ***Correlación Canónica*** (Badii et al., 2007e)	**1.** Interrelación entre gran número de variables

Diseño completamente al azar (DCA)

Este diseño es un análisis de varianza simple (ANOVA simple) y se emplea cuando a parte de los tratamientos no existe otra gradiente de variabilidad.

Ejemplo 12. Verificar si el promedio de juicios (tratamientos o grupos) durante 4 meses es igual entre 4 jueces procedentes de 4 Estados, es decir, los 4 grupos representan la misma población (Tabla 12.1).

Tabla 12.1. Número de juicios por cada juez durante cada uno de 4 meses.

Repetición (Estado)	Juez			
	A	**B**	**C**	**D**
I	17	14	12	13
II	14	14	12	11
III	13	13	10	11
IV	13	8	9	9
Σx_i	**57**	**49**	**43**	**44**
Σx_i^2	**823**	**625**	**469**	**492**

Ho: $m_1 = m_2 = \ldots = m_n$

Ha: No es así o al menos una media varía de cualquier otra media

Dicho de otra manera: *Ho:* Todas las medias son iguales

Ha: No todas las medias son iguales

El arreglo de los datos de ANOVA se indica en la Tabla 12.2.

Tabla 12.2. Templete de la tabla de ANOVA para DCA.

FV	*gl*	SC	Varianza	$F_{Calculada}$
Tratamientos (grupos)	K – 1	$\Sigma[(T_{cg})^2/r_{cg}]$ – FC	SC_t / gl_t	V_T / V_E
Error	gl_T - gl_t	SC_T – SC_t	SC_E / gl_E	
Total	Nt – 1	Σx_i^2 – FC		

FV = fuente de variación, K = número de grupos, Nt = número total de los elementos, gl_T = grados de libertad total, gl_t = grados de libertad del tratamientos, FC = factor de corrección = $(\Sigma x_i)^2/N_t$, SC_T = suma de cuadrados del total, SC_t = suma de cuadrados del tratamientos, V_T = varianza del tratamientos, V_E = varianza del error

Solución:

Los resultados se indican en la Tabla 12.3.

Tabla 12.3. Tabla de ANOVA para datos de la Tabla 12.1.

Fuente de Variación	*gl*	SC	Varianza	$F_{Calculada}$	F_{Tabla}
Tratamientos	4-1= 3	30.6875	10.2291	2.4427	3.49
Error	19-3= 12	50.25	4.1875		
Total	16-1= 15	80.9375			

$\mathbf{SC_{Total} = \Sigma\, x_i^2 - FC}$
$\mathbf{FC = (\Sigma\, x_i)^2/N_t} = (193)^2/16 = 37{,}249/16 = \mathbf{2{,}328.0625}$
$\mathbf{\Sigma\, x_i^2} = (17)^2 + (14)^2 + \ldots + (9)^2 = \mathbf{2{,}409}$
$\mathbf{SC_{Total}} = 2{,}409 - 2{,}328.0625 = \mathbf{80.9375}$

$\mathbf{SC_{Tratamientos} = \Sigma[(T_{cg})^2/r_{cg}] - FC}$
$\mathbf{\Sigma[(T_{cg})^2/r_{cg}]} = (57)^2/4 + (49)^2/4 + (43)^2/4 + (44)^2/4 = 812.25 + 600.25 + 462.25 + 484 = \mathbf{2{,}358.75}$
$\mathbf{FC = (\Sigma\, x_i)^2/N_t} = (193)^2/16 = 37{,}249/16 = \mathbf{2{,}328.0625}$
$\mathbf{SC_{Tratamientos}} = 2{,}358.75 - 2{,}328.0625 = \mathbf{30.6875}$

$\mathbf{SC_E = SC_{Total} - SC_{tratamientos}}$
$\mathbf{SC_E} = 80.9375 - 30.6875 = \mathbf{50.25}$

$\mathbf{V_t = SC_t / gl_t} = 30.6875 / 3 = \mathbf{10.2291}$
$\mathbf{V_E = SC_E / gl_E} = 50.25 / 12 = \mathbf{4.1875}$
$\mathbf{F_{Calculada} = V_t / V_E} = 10.2291 / 4.1875 = \mathbf{2.4427}$

$$F_{Calculada} \quad \text{vs} \quad F_{Tabulada} \left\{ \begin{array}{l} \alpha\text{: } 0.05 \\ gl\text{: } n=3 \end{array} \right.$$

2.4427 < 3.49 **:.** “Ho” Se Acepta. Todas las medias son iguales.

Conclusión. *Todas las muestras representan la misma población y no hay diferencia entre el promedio de los juicios entre los 4 jueces.*

Diseño de bloques al azar (DBA)

En cualquier experimento, la variabilidad proveniente de un factor de ruido puede afectar los resultados. Un factor de ruido es un factor que probablemente tiene un efecto en la respuesta pero que no nos interesa estudiar. Si el factor de ruido es conocido y controlable, se utilizan bloques para eliminar su efecto en la comparación estadística de los tratamientos.

Este diseño de bloques al azar (DBA) consiste en asignar los tratamientos en forma completamente al azar a un grupo de parcelas, llamados bloques o repeticiones, con la condición de que dicho bloque o conjunto de unidades experimentales sea lo más homogéneo posible. El objetivo de realizar un agrupamiento de unidades experimentales, es escoger aquellas más parecidas de manera de reducir hasta donde sea posible la variabilidad dentro de cada bloque y así evitar que los efectos de tratamiento se vean enmascarados o confundidos por la heterogeneidad de las unidades experimentales a las cuales se asignan los tratamientos de interés. Es por ello que el diseño de bloques completos al azar (DBA) se considera más eficiente que el diseño completamente al azar (DCA) para controlar la variabilidad del material experimental. Generalmente, el número de bloques es igual al número de tratamientos. El adjetivo "completos" se refiere a que todos los tratamientos aparecen representados en cada uno de los bloques del experimento.

Este diseño, ofrece menos grados de libertad para estimar el error, comparado con el DCA, esta reducción se debe a los grados de libertad necesarios para estimar el efecto de bloques. Es por ello que debe existir una razón real de bloqueo para usar este diseño, es decir, entre mayor sea la variabilidad entre bloques más eficiente será el diseño para detectar diferencias entre tratamientos.

Requisitos. Una buena elección del criterio de bloqueo resulta en menor varianza entre las unidades experimentales dentro de los bloques comparada con la varianza entre unidades experimentales de diferentes bloques. Generalmente los criterios de bloqueo son:

1. Proximidad (parcelas vecinas).
2. Características físicas (edad, peso, sexo).
3. Tiempo.
4. Manejo de las unidades experimentales en el experimento.

Objetivo. Determinar si aparte de las diferencias entre las medias de los tratamientos, también existe diferencia significativa entre las medias de los bloques.

Ejemplo 13. Tomando los mismos datos del problema anterior (DCA) (ejemplo 12), verificar si:

1. El promedio de los juicios de 4 grupos (jueces) son iguales, es decir, los 4 grupos representan la misma población.
2. Los diferentes Estados representan la misma población, es decir, la procedencia del juez no impacta sobre el juicio (Tabla 13).

Juego de hipótesis: *Para cada caso: los jueces y los Estados tenemos las siguientes pares de hipótesis:*

Ho: Todas las medias son iguales
Ha: No es así (o al menos una media varía de cualquier otra media)

Tabla 13. Arreglo de datos para el diseño de bloques al azar (DBA).

Estados (Bloques)	**Juez**				**Total de bloque**
	A	**B**	**C**	**D**	
I	17	14	12	13	**56**
II	14	14	12	11	**51**
III	13	13	10	11	**47**
IV	13	8	9	9	**39**
Σ x	**57**	**49**	**43**	**44**	**ΣT** = 193
Σx^2	**823**	**625**	**469**	**492**	**ΣT^2** = 2,409

Solución:

El templete de ANOVA y los resultados de este análisis se indican en la Tabla 13.1 y Tabla 13.2, respectivamente.

Tabla 13.1. Tabla de templete de ANOVA para los datos de la Tabla 13.

FV	***gl***	**SC**	**Varianza**	**$F_{Calculada}$**
Tratamientos	k-1	$\Sigma[(Tcg)^2/r_{cg}) -$ FC	SC_{trata}/gl_{trata}	V_{Trata}/V_{Error}
Bloques	B-1	$\Sigma[(Tcb)^2/r_{cb}) -$ FC	$SC_{Bloques}/gl_{Bloques}$	$V_{Bloques}/V_{Error}$
Error	$gl_{Tot.}-[gl_{trat.}+gl_b]$	$SC_{Total}-[SC_{Trata}+SC_{Bloques}]$	SC_{Error}/gl_{Error}	
Total	n-1	Σx^2 - FC		

gl = Grados de libertad, SC = Suma de cuadrados, K = Número de grupos o tratamientos, B = Número de bloques, n = Número de elementos, $gl_{Tot.}$ = gl total, $gl_{trat.}$= gl de tratamientos, gl_b= gl de bloques, Tcg = Total de cada grupo, r_{cg} = Número de repeticiones de cada grupo, Tcb = Total de cada bloque, r_{cb} = Número de repeticiones de cada bloque, FC = Factor de corrección = $(\Sigma x)^2/N_T$

FC $= (\Sigma\, x_i)^2/N_T = (193)^2/\ 16 = 37{,}249/16 = \mathbf{2{,}328.0625}$
$SC_{Tratamientos} = \Sigma[(Tcg)^2/r_{cg}) -\ FC$
$\Sigma[(Tcg)^2/r_{cg}) = [(57)^2 + (49)^2 + (43)^2 + (44)^2]/4 = \mathbf{2{,}358.75}$
$SC_{Tratamientos} = 2{,}358.75 - 2{,}328.0625 = \mathbf{30.6875}$

$SC_{Bloques} = \Sigma[(Tcb)^2/r_{cb}) -\ FC$
$\Sigma[(Tcb)^2/r_{cb}) = [(56)^2 + (51)^2 + (47)^2 + (39)^2]/4 = \mathbf{2{,}366.75}$
$SC_{Bloques} = 2{,}366.75 - 2{,}328.0625 = \mathbf{38.6875}$

$SC_{Total} = \ \Sigma\, x_i^2 - FC$
$\Sigma\, x_i^2 = (17)^2 + (14)^2 + \ldots + (9)^2 = \mathbf{2{,}409}$
$SC_{Total} = 2{,}409 - 2{,}328.0625 = \mathbf{80.9375}$

$SC_{Error} = SC_{Total}\ [SC_{Tratamientos} + SC_{Bloques}]$
$SC_{Error} = 80.9375 - [30.6875 + 38.6875] = \mathbf{11.5625}$

Tabla 13.2. ANOVA para datos de la Tabla 13 (*: efecto significativo = el rechazo de "Ho").

FV	*gl*	SC	Varianza	$F_{Calculada}$	$F_{Tabulada}$
Tratamientos (grupos)	4-1=3	30.6875	10.2291	7.9621*	3.86
Bloques	4-1=3	38.6875	12.8958	10.0364*	3.86
Error	15-3-3=9	11.5625	1.2849		
Total	16-1=15	80.9375			

FC vs FT $\begin{cases} \alpha: 0.05 \\ gl: n=3 \quad d=9 \end{cases}$

7.9621 > 3.86 :. "Ho" Se Rechaza.

FC vs FT $\begin{cases} \alpha: 0.05 \\ gl: n=3 \quad d=9 \end{cases}$

10.0364 > 3.86 :. "Ho" Se Rechaza.

Conclusiones:

1. Hay diferencia significativa entre los promedios de los juicios entre los jueces, es decir, los Jueces no representan la misma población, sino al contrario reflejan diferentes poblaciones.
2. Hay diferencia significativa entre los promedios de los juicios entre los Estados, por tanto, los 4 Estados representan diferentes poblaciones.

De aquí nos damos cuenta, que los mismos datos tienen diferentes significados cuando se los tratan por 2 modelo distintos, en este caso, el modelo de Diseño

Completamente Aleatorio (DCA) (caso del ejemplo 12), y el modelo actual o Diseño de Bloques al Azar (DBA), como es el ejemplo actual o 13. Hay que recordar que no hubo diferencias significativas entre el promedio de los juicios entre los 4 Jueces con el modelo de DCA, sin embargo, al bloquear el efecto de los Estados (usar el modelo DBA), se encontró diferencias significativas entre los promedios de los juicios tanto por los Jueces como por los Estados. Esta diferencia se debe que la magnitud de la varianza del término de Error era inflada en el caso del modelo de DCA, y al bloquear el efecto de los Estados, se pudo reducir el tamaño de la varianza del término de Error, y por consecuente, la diferencia entre los resultados de estos dos modelos de ANOVA.

Diseño de Cuadro Latino (DCL)

En este diseño los tratamientos son agrupados en dos diferentes formas, por "hileras" y por "columnas", lo cual puede interpretarse como bloques en dos sentidos. La idea básica es que cada tratamiento esté presente en cada "hilera" y en cada "columna", de tal manera que cualquier comparación entre tratamientos no sea afectada por las diferencias entre hileras o columnas. Este arreglo permite remover la variabilidad del material experimental en dos "direcciones" o controlar dos fuentes de variación según sea el caso. En este diseño el número de hileras (h) es igual al número de columnas (c) e igual al número de tratamientos (t). Hay que asignar los tratamientos al azar, de acuerdo a una forma de aleatorización característica del diseño. Letras latinas son usadas para identificar los tratamientos aplicados a las unidades experimentales, los cuales deben ocurrir únicamente una vez en cada hilera y en cada columna.

Objetivo. Determinar si aparte de las diferencias entre las medias de los tratamientos, también existe diferencia significativa entre las medias de los bloques verticales (columnas) y entre las medias de los bloques horizontales (hileras).

Ventajas. Controla las fuentes de variación en las dos direcciones: hileras y columnas, es decir, extrae del error experimental la variación debida a tratamientos, hileras y columnas.

Desventajas. Se pierden grados de libertad en el error experimental, sacrificando la precisión del diseño experimental. Debido a que el número de hileras y columnas debe ser igual al de tratamientos, el número de tratamientos es limitado.

Restricciones:

1. Un tratamiento cualquiera debe estar solamente una vez en una columna.
2. Un tratamiento cualquiera debe estar solamente una vez en una hilera.

Ejemplo 14. Utilizando los datos del problema de DCA (ejemplo 12), vamos a verificar si la diferencia entre los promedios de los juicios se debe: a) a los jueces, b) a los Estados, y c) a los meses de estudio. En otras palabras a parte del bloqueo de los Estados (ejemplo 13 del modelo de DBA), se va a bloquear el efecto del tiempo (como un posible fuente de variación que pudiera afectar los tratamientos.

Ecuaciones:

El arreglo de los datos y las ecuaciones del Diseño de Cuadro Latino se notan en la Tabla 14 y Tabla 14.1 repectivamente.

Las Hipótesis del modelo de DCL:

"Ho"$_{\text{Tratamientos}}$: Todas las medias son iguales o ($m_{t1} = m_{t2} = m_{t3} = m_{t4}$)
"Ha"$_{\text{Tratamientos}}$: Al menos una media es distinta
(t = tratamientos o jueces)

"Ho"$_{BV}$: Todas las medias son iguales o ($m_{BV1} = m_{BV2} = m_{BV3} = m_{BV4}$)
"Ha"$_{BV}$: Al menos una media es distinta
(BV = Bloqueo vertical o Estados)

"Ho"$_{BH}$: Toas las medis son iguales o $m_{BH1} = m_{BH2} = m_{BH3} = m_{BH4}$)
"Ha"$_{BH}$: Al menos una media es distinta
(BH = Bloqueo horizontal o tiempo en meses)

Tabla 14. Arreglo de datos para DCL (letras = número de juicios).

Tiempo (Bloques horizontal = BH)	Estados (Bloques vertical = BV)				Total$_{BH}$
	I	**II**	**III**	**IV**	
1	C 12	C 12	A 13	B 8	**45**
2	B 14	D 11	D 11	A 13	**49**
3	A 17	B 14	C 10	D 9	**50**
4	D 13	A 14	B 13	C 9	**49**
Σx_i (Total$_{BV}$)	**56**	**51**	**47**	**39**	**$\Sigma\Sigma x_i = 193$**
Σx_i^2	**798**	**657**	**559**	**395**	**$\Sigma\Sigma x_i^2 = 2,409$**

Tabla 14.1. Tabla de templete de ANOVA para los datos de la Tabla 14.

FV	*gl*	SC	Varianza	$F_{Calculada}$
Tratamientos	k-1	$\Sigma[(Tcg)^2/r_{cg}) - (\Sigma x)^2/n$	SC_{trata}/gl_{trata}	V_{Trata}/V_{Error}
Estados (BV)	BV-1	$\Sigma[(Tcbv)^2/r_{cbv}) - (\Sigma x)^2/n$	$SC_{Bloques}/gl_{Bloques}$	$V_{Bloques}/V_{Error}$
Tiempo (BH)	BH-1	$\Sigma[(Tch)^2/r_{cbh}) - (\Sigma x)^2/n$	$SC_{Bloques}/gl_{Bloques}$	$V_{Bloques}/V_{Error}$
Error	$gl_{Tot.} - [gl_{trat.} + gl_{bv} + gl_{bh}]$	$SC_{Total} - [SC_{Trata} + SC_{BV} + SC_{BH}]$	SC_{Error}/gl_{Error}	
Total	n-1	$\Sigma x^2 - (\Sigma x)^2/n$		

gl = Grados de libertad, SC = Suma de cuadrados, K = Número de grupos o tratamientos, BV = Número de bloques varticales, BH = Número de bloques horizontales, n = Número de elementos, $gl_{Tot.}$ = gl total, $gl_{trat.}$= gl de tratamientos, gl_{bv}= gl de bloques vericales, gl_{bh}= gl de bloques horizontales, Tcg = Total de cada grupo, r_{cg} = Número de repeticiones de cada grupo, Tcbv = Total de cada bloque vertical, r_{cbv} = Número de repeticiones de cada bloque vertical, Tcbh = Total de cada bloque horizontal, r_{cbh} = Número de repeticiones de cada bloque horizontal, FC = Factor de corrección = $(\Sigma x)^2/n$

Solución:

$FC = (\Sigma x_i)^2/N_T$ $= (193)^2/ 16 = 37{,}249/16=$ **2,328.0625**

$SC_{Tratamientos} = \Sigma[(Tcg)^2/r_{cg}) - FC$

$\Sigma[(Tcg)^2/r_{cg})$ $= [(57)^2 + (49)^2 + (43)^2 + (44)^2]/4 =$ **2,358.75**

$SC_{Tratamientos}$ = 2,358.75 - 2,328.0625 = **30.6875**

$SC_{Bloques\ verticales} = \Sigma[(Tcb)^2/r_{cb}) - FC$

$\Sigma[(Tcb)^2/r_{cb})$ $= [(56)^2 + (51)^2 + (47)^2 + (39)^2]/4 =$ **2,366.75**

$SC_{Bloques\ verticales}$ = 2,366.75 – 2,328.0625 = **38.6875**

$SC_{Bloques\ hortizontales} = \Sigma[(Tcb)^2/r_{cb}) - FC$

$\Sigma[(Tcb)^2/r_{cb})$ $= [(45)^2 + (49)^2 + (50)^2 + (49)^2]/4 =$ **2,331.75**

$SC_{Bloques\ horizontles}$ = 2,331.75 – 2,328.0625 = **3.6875**

$SC_{Total} = \Sigma x_i^2 - FC$

Σx_i^2 $= (17)^2 + (14)^2 + \ldots + (9)^2 =$ **2,409**

SC_{Total} = 2,409 – 2,328.0625 = **80.9375**

$$SC_{Error} = SC_{Total}\ [SC_{Tratamientos} + SC_{Bloques\ horizontales} + SC_{Bloques\ horizontles}\ ]$$
$$SC_{Error} = 80.9375 – [30.6875 + 38.6875 + 3.6875] = \mathbf{7.875}$$

Los resultados se indican en la Tabla 14.2.

Tabla 14.2. Tabla de de ANOVA para los datos de la Tabla 14 (* = efecto significativo).

FV	*Gl*	SC	Varianza	$F_{Calculada}$	$F_{Tabulada}$
Tratamientos (Grupos)	4 – 1 = 3	30.6875	10.2291	7.9621*	4.76
Estados (BV)	4 - 1 = 3	38.6875	12.8958	10.0364*	4.76
Tiempo (BH)	4 – 1 = 3	3.6875	1.2291	0.9364	4.76
Error	15 – (3 + 3 + 3) = 6	11.5625	1.3125		
Total	16 - 1 = 15	80.9375			

Tratamientos

FC vs FT $\begin{cases} \alpha: 0.05 \\ gl: n=3 \quad d=6 \end{cases}$

7.9621 > 4.76 **:.** “Ho” Se Rechaza.

Estados (Bloques verticales)

FC vs FT $\begin{cases} \alpha: 0.05 \\ gl: n=3 \quad d=6 \end{cases}$

10.0364 > 4.76 **:.** “Ho” Se Rechaza.

Tiempo (Bloques horizontales)

FC vs FT $\begin{cases} \alpha: 0.05 \\ gl: n=3 \quad d=6 \end{cases}$

0.9364 < 4.76 **:.** “Ho” Se Acepta.

Conclusión. *La diferencia entre los promedios de los juicios se debe a la diferencia por un lado entre los Jueces, y por otro lado, entre los Estados, más sin embargo, el tiempo no ejerce ningún efecto al respecto. Por tanto, los juicios proceden de diferentes jueces, y diferentes Estados pero reflejan en mismo efecto temporal.*

Análisis de varianza anidada (ANOVA Anidada)

Definición. Se utiliza el análisis de varianza anidado (ANOVA anidada) cuando se tiene una variable de medición y dos o más variables nominales. Las variables nominales se anidan, lo que significa que cada valor de una variable nominal (los subgrupos) se encuentra en combinación con un solo valor de la variable nominal de más alto nivel (los grupos). El nivel nominal de la variable superior puede ser Modelo I o II pero de las variables más bajos deben ser el Modelo II.
El análisis de varianza anidado es una extensión del análisis de varianza de una vía en que se divide cada grupo en subgrupos. En teoría, estos subgrupos se eligen al azar de un conjunto más amplio posible de los subgrupos. Un análisis de varianza anidado tiene una hipótesis nula para cada nivel. En uno de dos niveles ANOVA anidados, una hipótesis nula sería que los subgrupos dentro de cada grupo tienen el mismo medio, la segunda hipótesis nula sería que los grupos tienen los mismos medios.

Objetivo. Comparar varios factores que estén relacionados con una muestra, determinando su igualdad entre medias.
Requisitos. El análisis de varianza anidado, como todos los ANOVAs, supone que las observaciones dentro de cada subgrupo se distribuyen normalmente y tienen varianzas iguales .
Ejemplo 15. Hay 5 distritos con 4 jueces por cada distrito, y analizar el número de juicios por cada juez de cada distrito durante 4 semanas del mes (Tabla 15).

Tabla 15. Número de juicios por juez por semana por distrito[$].

S	Distritos																			
	A				B				C				D				E			
	Jueces				Jueces				Jueces				Jueces				Jueces			
	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
I	6	13	1	7	10	2	4	0	0	10	8	7	11	5	1	0	1	6	3	3
II	2	3	10	4	9	1	1	3	0	11	5	2	0	10	8	8	4	7	0	7
III	0	9	0	7	7	1	7	4	5	6	0	5	6	8	9	6	7	0	2	4
IV	8	8	6	9	12	10	9	1	5	7	7	4	4	3	4	5	9	3	2	0
ΣCJ	16	33	17	27	38	14	21	8	10	34	20	18	21	26	22	19	21	16	7	14
ΣCD	93				81				82				88				58			

$: S = Semana, ΣCJ = Sumatorio para cada juez, ΣCD = Sumatorio para cada distrito

<u>Juegos de hipótesis:</u>

1. La eficacia entre los 5 Distritos:

   Ho: $m_{\text{Dist A}} = m_{\text{Dist B}} = m_{\text{Dist C}} = m_{\text{Dist D}} = m_{\text{Dist E}}$

   Ha: Al menos una media es diferente

2. La eficacia de los 20 Jueces (Suma de jueces o jueces combinados):
Ho: $m_{Juez\ I} = m\ m_{Juez\ II} = m_{t3}\ldots = m\ m_{Juez\ XX}$
Ha: Al menos una media es diferente

3. Para los Jueces de cada Distrito:

3a. Los Jueces del Distrito A:
Ho: $m_{Juez\ I\ Dist\ A} = m_{Juez\ II\ Dist\ A} = m_{Juez\ III\ Dist\ A} = m_{Juez\ IV\ Dist\ A}$
Ha: Al menos una media es diferente

3b. Los Jueces del Distrito B:
Ho: $m_{Juez\ I\ Dist\ B} = m_{Juez\ II\ Dist\ B} = m_{Juez\ III\ Dist\ B} = m_{Juez\ IV\ Dist\ B}$
Ha: Al menos una media es diferente

3c. Los Jueces del Distrito C:
Ho: $m_{Juez\ I\ Dist\ C} = m_{Juez\ II\ Dist\ C} = m_{Juez\ III\ Dist\ C} = m_{Juez\ IV\ Dist\ C}$
Ha: Al menos una media es diferente

3d. Los Jueces del Distrito D:
Ho: $m_{Juez\ I\ Dist\ D} = m_{Juez\ II\ Dist\ D} = m_{Juez\ III\ Dist\ D} = m_{Juez\ IV\ Dist\ D}$
Ha: Al menos una media es diferente

3e. Los Jueces del Distrito E:
Ho: $m_{Juez\ I\ Dist\ E} = m_{Juez\ II\ Dist\ E} = m_{Juez\ III\ Dist\ E} = m_{Juez\ IV\ Dist\ E}$
Ha: Al menos una media es diferente

El templete de ANOVA y las ecuaciones de este diseño se encuentran en la Tabla 15.1.

Solución:

SC Totales

$\mathbf{SC_{Totales} = \Sigma x_i^2 - (\Sigma x_i)^2/N_T}$
$\mathbf{\Sigma x_i} = 402$
$\mathbf{\Sigma\ x_i^2} = 2,990$
$\mathbf{FC = (\Sigma x_i)^2/ N_T} = (402)^2 /80 = 161,604/80 = 2,020.05$
$\mathbf{SC_{Totales}} = 2,990 - 2,020.05 = \mathbf{969.95}$

SC de los 5 distritos

$\mathbf{SC_{Distritos} = \Sigma(Tcg)^2/r_{CG} - (\Sigma x_i)^2/N_T}$

$\boldsymbol{\Sigma(Tcg)^2/r_{CG}}$ = [{$(93)^2 + (81)^2 + (82)^2 + (88)^2 + (58)^2$} / 16] = 2,065.125
$\mathbf{FC_{Distritos} = (\Sigma x_i)^2/ N_T}$ = $(402)^2$ /80 = 161,604/80= 2,020.05
$\mathbf{SC_{Distritos}}$ = 2,065.125 – 2,020.05
$\mathbf{SC_{Distritos}}$ = **45.075**

<u>SC de Jueces del distrito A</u>
$\mathbf{SC_{Dist\ A} = \Sigma(T_{DistA})^2/r_{DistA} - (\Sigma x_{DistA})^2/r_{Corr}}$
$\mathbf{\Sigma(T_{Dist\ A})^2/r_{DistA}}$ = [{$(16)^2 + (33)^2 + (17)^2 + (27)^2$} / 4] = 590.75
$\mathbf{FC_{Dist\ A}}$ = $\mathbf{(\Sigma x_{Dist.A})^2/r_{Dist.A}}$ = $(93)^2/16$ = 540.5625
$\mathbf{SC_{Dist\ A}}$ = 590.75-540.5625
$\mathbf{SC_{Dist\ A}}$ = **50.1875**

<u>SC de Jueces del distrito B</u>
$\mathbf{SC_{Dist\ B} = \Sigma(T_{Dist\ B})^2/r_{DistB} - (\Sigma x_{Dist\ B})^2/r_{Corr}}$
$\mathbf{\Sigma(T_{Dist\ B})^2/r_{Dist\ B}}$ = [{$(38)^2 + (14)^2 + (21)^2 + (8)^2$} / 4] = 536.25
$\mathbf{FC_{Dist\ B}}$ = $\mathbf{(\Sigma x_{Dist.B})^2/r_{Dist.B}}$ = $(81)^2/16$ = 410.0625
$\mathbf{SC_{Dist\ B}}$ = 536.725 - 410.0625
$\mathbf{SC_{Dist\ B}}$ = **126.1875**

<u>SC de Jueces del distrito C</u>
$\mathbf{SC_{Dist\ C} = \Sigma(T_{Dist\ C})^2/r_{Dist\ C} - (\Sigma x_{Dist\ C})^2/r_{Corr}}$
$\mathbf{\Sigma(T_{Dist\ C})^2/r_{Dist\ C}}$ = [{$(34)^2 + (20)^2 + (18)^2 + (21)^2$} / 4] = 495.00
$\mathbf{FC_{Dist\ C}}$ = $\mathbf{(\Sigma x_{Dist.C})^2/r_{Dist.C}}$ = $(82)^2/16$ = 420.25
$\mathbf{SC_{Dist\ C}}$ = 495.00 - 420.25
$\mathbf{SC_{Dist\ C}}$ = **74.75**

<u>SC de Jueces del distrito D</u>
$\mathbf{SC_{Dist\ D} = \Sigma(T_{Dist\ D})^2/r_{Dist\ D} - (\Sigma x_{Dist\ D})^2/r_{Corr}}$
$\mathbf{\Sigma(T_{Dist\ D})^2/r_{Dist\ D}}$ = [{$(21)^2 + (26)^2 + (22)^2 + (19)^2$} / 4] = 490.50
$\mathbf{FC_{Dist\ D}}$ = $\mathbf{(\Sigma x_{Dist.D})^2/r_{Dist.D}}$ = $(88)^2/16$ = 484.00
$\mathbf{SC_{Dist\ D}}$ = 490.50 - 484.00
$\mathbf{SC_{Dist\ D}}$ = **6.50**

<u>SC de Jueces del distrito E</u>
$\mathbf{SC_{Dist\ E} = \Sigma(T_{Dist\ E})^2/r_{Dist\ E} - (\Sigma x_{Dist\ E})^2/r_{Corr}}$
$\mathbf{\Sigma(T_{Dist\ E})^2/r_{Dist\ E}}$ = [{$(21)^2 + (16)^2 + (7)^2 + (14)^2$} / 4] = 235.50
$\mathbf{FC_{Dist\ E}}$ = $\mathbf{(\Sigma x_{Dist.E})^2/r_{Dist.E}}$ = $(58)^2/16$ = 210.25
$\mathbf{SC_{Dist\ E}}$ = 235.50 - 210.20
$\mathbf{SC_{Dist\ E}}$ = **25.25**

SC del Error

SC_E = SC_{Total} – ($\Sigma SC_{Distritos}$) = 969.95 - (45.075 + 50.1875 + 126.1875 + 74.75 + 6.5 + 25.25) = **642**

Tabla 15.1. Templete de la ANOVA para ANOVA anidada$.

FV	*gl*	SC	V	$F_{calculada}$
Distritos (grupos)	k-1	$\Sigma(Tcg)^2/r_{CG} - (\Sigma x_i)^2/N_T$	SC_{Dists}/gl_{Dists}	V_{Dists}/V_{Error}
Σ Jueces combinados	N_T-5	ΣSC_{Traba}	$SC_{\Sigma J}/gl_{\Sigma J}$	$V_{\Sigma J}/V_{Error}$
Jueces del Distrito A	k-1	$\Sigma(T_{DistA})^2/r_{DistA} - (\Sigma x_{DistA})^2/r_{Corr}$	SC_{JDA}/gl_{JDA}	V_{JDA}/V_{Error}
Jueces del Distrito B	k-1	$\Sigma(T_{DistB})^2/r_{DistB} - (\Sigma x_{DistB})^2/r_{Corr}$	SC_{JDb}/gl_{JDB}	V_{JDB}/V_{Error}
Jueces del Distrito C	k-1	$\Sigma(T_{DistC})^2/r_{DistC} - (\Sigma x_{DistC})^2/r_{Corr}$	SC_{JDC}/gl_{JDC}	V_{JDC}/V_{Error}
Jueces del Distrito D	k-1	$\Sigma(T_{DistD})^2/r_{DistD} - (\Sigma x_{DistD})^2/r_{Corr}$	SC_{JDD}/gl_{JDD}	V_{JDD}/V_{Error}
Jueces del Distrito E	k-1	$\Sigma(T_{DistE})^2/r_{DistE} - (\Sigma x_{DistE})^2/r_{Corr}$	SC_{JDE}/gl_{JDE}	V_{JDE}/V_{Error}
Error	SC_T -(SC_{Dists} +SC_{Juecs})	$SC_T - (SC_{Dists} + SC_{Juecs})$		
Total	N_T-1	$\Sigma x_i^2 - (\Sigma x_i)^2/N_T$		
$: ΣJ = Suma de jueces combinados, JDi = Jueces de Distrito “i”, r_{Corr} = 16				

Los resultados se encuentran en la Tabla 15.2.

Conclusiones:

1. El desempeño de los juicios (promedio de juicios) es igual entre los 5 distritos.
2. El promedio de desempeño de todos los jueces (combinados) es igual, es decir, no hay diferencia entre los promedios de desempeño de suma de los 20 jueces.
3. Finalmente, para cada distrito individual, con la excepción del distrito B, los promedios de desempeño de los jueces de los demás distritos (A, C, D y E) son iguales.

Tabla 15.2. Tabla ANOVA para datos de los Distritos.

FV	*gl*	SC	V	$F_{Calculada}$	$F_{Tabulada}$
Distritos	4	45.075	11.2687	1.0531	2.53
Σ Jueces combinados	15	282.875	18.8583	1.7624	1.84
Jueces del Distrito A	3	50.1875	16.7291	1.5634	2.76
Jueces del Distrito B	3	126.1875	42.0625	3.9310*	2.76
Jueces del Distrito C	3	74.75	24.9166	2.3286	2.76
Jueces del Distrito D	3	6.50	2.1666	0.2024	2.76
Jueces del Distrito E	3	25.25	8.4166	0.7865	2.76
Error	60	642.00	10.7		
Total	79	969.95			

Análisis de Covarianza (ANCOVA)

Definición. Es una combinación de dos técnicas: Análisis de la Varianza y Análisis de Regresión. En el Análisis de la Covarianza: La variable respuesta es cuantitativa. Las variables independientes son cualitativas y cuantitativas. Permite eliminar la heterogeneidad causada en la variable de interés (variable dependiente) por la influencia de una o más variables cuantitativas (covariables). Básicamente, el fundamento del ANCOVA es un ANOVA al que en la variable dependiente se le ha eliminado el efecto predicho por una o más covariables por regresión lineal múltiple. La inclusión de covariables puede aumentar la potencia estadística porque a menudo reduce la variabilidad.

Requisitos. El análisis de covarianza es muy estricto en sus requisitos previos; requiere fundamentalmente la *asignación aleatoria de los sujetos a los diversos tratamientos* (o la asignación aleatoria de los grupos a las condiciones experimental y control). Esta necesidad de asignación aleatoria hace que el análisis de covarianza no sea siempre fácilmente posible. Si no hay asignación aleatoria de los sujetos a los diversos grupos o tratamientos, las diferencias en la variable dependiente (en definitiva nuestros resultados) pueden estar relacionadas con diferencias en otras variables, que pueden ser desconocidas, no controladas y que hacen difícil la interpretación de los resultados

Objetivo. Busca comparar los resultados obtenidos en diferentes grupos de una variable cuantitativa, pero "corrigiendo" las posibles diferencias existentes entre los grupos en otras variables que pudieran afectar también el resultado (covariantes).

Ejemplo 16. Se trata de estudiar número de votos multiplicado por 1000 (variable respuesta o “Y”) en 7 Estados (tratamientos o grupos) con 6 repeticiones o bloques por cada Estado. En este estudio también se toman en cuenta una variable auxiliar o covariable “X” que es el número de horas después del inicio de la votación en la cual se realiza el conteo de votos. Queremos ver si el número de votos (Y) está afectado por la hora de contar los votos (X) (Tabla 16).

X = Hora de conteo de los votos después del inicio de la votación
Y = Número de votos por 1000

Tabla 16. Datos para la Análisis de covarianza (ANCOVA) (G = grupos, t = tratamientos).

G = t	Bloques												Σ	
	I		II		III		IV		V		VI			
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
1	41	4.08	24	2.78	31	2.79	46	4.24	32	4.17	38	2.62	**212**	**20.68**
2	40	4.26	36	4.23	44	5.6	48	6.36	47	4.33	47	4.03	**262**	**28.81**
3	37	4.72	32	4.92	38	4.5	41	5.62	42	5.15	40	4.32	**230**	**29.23**
4	32	4.25	38	4.53	40	4.83	40	4.3	51	5.43	46	4.52	**247**	**27.86**
5	37	4	34	5.05	47	5.54	41	4.46	50	6.63	39	5.7	**248**	**31.38**
6	42	6.16	38	5	34	4.61	40	5.41	42	5.13	49	5.86	**245**	**32.17**
7	35	4.59	22	3.63	44	6.2	35	5.47	40	5.16	40	6.07	**216**	**31.12**
Σ	**264**	**32.06**	**224**	**30.1**	**278**	**34.07**	**291**	**35.86**	**304**	**36.0**	**299**	**33.12**	**1660**	**201.25**

Ecuaciones:

Paso 1: Ecuaciones para calcular Suma de Cuadrados de la variable “X”:

$$\mathbf{SC_{Totx} = (\Sigma x_i^2) - FC}$$
$$\mathbf{FC = (\Sigma x_i)^2/N_T}$$
$$\mathbf{SC_t = \Sigma(T_{txi})^2/r_{ti} - FC}$$
$$\mathbf{SC_B = \Sigma(T_{Bxi})^2/r_{Bi} - FC}$$
$$\mathbf{SC_E = SC_{Tot} - (SC_B + SC_t)}$$

Paso 2: Ecuaciones para calcular Suma de Cuadrados de la variable “Y”:
Mismas fórmulas, sólo se cambia la “X” por la “Y”.

Paso 3: Ecuaciones para calcular Suma de Cuadrados del producto “XY”:

$$\mathbf{SC_{Totxy} = \Sigma(x_i y_i) - FC_{xy}}$$
$$\mathbf{FC_{xy} = (\Sigma x_i)(\Sigma y_j) / N_T}$$
$$\mathbf{SC_{txy} = \Sigma[T_{txyi}]^2/r_{txy} - FC_{xy}}$$

$\mathbf{SC_{Bxy} = \Sigma\ [T_{Bxyi}]^2/r_{Bxy}\ - FC_{xy}}$

$\mathbf{SC_E = SC_{Totxy} - (SC_{Bxy} + SC_{txy})}$

Cálculo del coeficiente de la regresión b_{yx}: $b_{yx} = SC_{Exy} / SC_{E(x)}$

Solución:

<u>Paso 1: Cálculo de los datos para "X"</u>

$\mathbf{SC_{Total\ x} = (\Sigma x_i^2) - (\Sigma x_i)^2 / N_T}$
$(\Sigma x_i^2) = (41)^2 + (40)^2 + \ldots + (40)^2 = 67{,}258$
$FC_x = (\Sigma x_i)^2 / N_T = (1660)^2/42 = 2´755{,}600/42= 65{,}609.52$
$SC_{Total\ x} = 67{,}258 - 65{,}609.52= \mathbf{1{,}648.48}$

$\mathbf{SC_{trat.\ x} = \Sigma[(T_{txi})^2/r_{ti}] - FC_X}$
$\mathbf{SC_{trat.\ x} = (212)^2 + (262)^2 + \ldots + (216)^2 / 6] - 65{,}609.52}$
$\mathbf{SC_{trat.\ x} = 65{,}947 - 65{,}609.52 = 337.47}$

$\mathbf{SC_{Bloue\ x} = \Sigma[(T_{Bxi})^2/r_{Bi}] - FC_X}$
$\mathbf{SC_{Bloque\ x} = (264)^2 + (224)^2 + \ldots + (299)^2 / 7] - 65{,}609.52}$
$\mathbf{SC_{Bloque\ x} = 66{,}236.27 - 65{,}609.52 = 626.67}$

$\mathbf{SC_{Ex} = SC_{Total\ x} - (SC_{trat.\ x} + SC_{Bloue\ x})}$
$SC_{Ex} = 1{,}648.48 - (337.47 + 626.76) = \mathbf{684.24}$

<u>Paso 2: Cálculo de los datos para "Y"</u>

$\mathbf{SC_{Total\ y} = (\Sigma y_i^2) - (\Sigma y_i)^2 / N_T}$
$(\Sigma y_i^2) = (4.08)^2 + (4.26)^2 + \ldots + (6.07)^2 = 999.7827$
$FC_y = (\Sigma y_i)^2 / N_T = (201.25)^2/42 = 40{,}501.56/42= 964.32$
$SC_{Total\ y} = 999.7827 - 964.32= \mathbf{35.46}$

$\mathbf{SC_{trat.\ y} = \Sigma[(T_{tyi})^2/r_{ti}] - FC_Y}$
$\mathbf{SC_{trat.\ y} = (20.68)^2 + (28.81)^2 + \ldots + (31.12)^2 / 6] - 964.32}$
$\mathbf{SC_{trat.\ y} = 979.34 - 964.32 = 15.02}$

$\mathbf{SC_{Bloue\ y} = \Sigma[(T_{Byi})^2/r_{Bi}] - FC_Y}$
$\mathbf{SC_{Bloque\ y} = (32.06)^2 + (30.14)^2 + \ldots + (33.17)^2 / 7] - 964.32}$
$\mathbf{SC_{Bloque\ y} = 967.96 - 964.32 = 3.64}$

$\mathbf{SC_{Ey} = SC_{Total\ y} - (SC_{trat.\ y} + SC_{Bloue\ y})}$
$SC_{Ey} = 35.64 - (15.02 + 3.64) = \mathbf{16.8}$

Paso 3: Cálculo de los datos para "XY"

$\mathbf{SC_{Total\ XY} = \Sigma\ x_iy_i - (\Sigma x_i * \Sigma y_i) / N_T}$

$\mathbf{\Sigma\ x_iy_i} = (41 * 4.08) + (40 * 4.26) + \ldots + (40 * 6.07) =$

$\mathbf{FC_{xy}} = \mathbf{(\Sigma x_i * \Sigma y_i)\ / N_T} = (1{,}660 * 203.35) / 42 = 7{,}954.16$

$\mathbf{SC_{Total\ xy}} = 8{,}091.76 - 7{,}954.16 = \mathbf{137.6}$

$\mathbf{SC_{trat.\ xy} = \Sigma[(T_{txyi})/r_{txyi}] - FC_{xy}}$

$\mathbf{SC_{trat.\ xy} = (212 * 20.68) + (262 * 28.81) + \ldots + (216 * 31.12) / 6] - 7{,}954.16}$

$\mathbf{SC_{trat.\ xy} = 7{,}987.08 - 7{,}954.16 = 32.92}$

$\mathbf{SC_{Bloue\ xy} = \Sigma[(T_{Bxyi})^2/r_{Bxyi}] - FC_{xy}}$

$\mathbf{SC_{Bloque\ xy} = (264 * 32.06) + (224 * 30.14) + \ldots + (299 * 33.12) / 7] - 7{,}954.16}$

$\mathbf{SC_{Bloque\ xy} = 7{,}995.54 - 7{,}954.16 = 41.38}$

$\mathbf{SC_{Exy} = SC_{Total\ xy} - (SC_{trat.\ xy} + SC_{Bloue\ xy})}$

$\mathbf{SC_{EY}} = 137.6 - (32.92 + 41.38) = \mathbf{63.3}$

A continuación se indican las Tablas de ANOVA para las variables "X" (Tabla 16.1) y "Y" en la Tabla 16.2.

Tabla 16.1. ANOVA para la variable "X".

FV	gl	SC	V	$F_{Calculada}$	$F_{Tabulada}$
Tratamiento	6	337.47	56.24	2.46*	2.42
Error	30	684.24	22.8		

Tabla 16.2. ANOVA para la variable "Y".

FV	gl	SC	V	$F_{Calculada}$	$F_{Tabulada}$
Tratamiento	6	15.02	2.50	4.46*	2.42
Error	30	16.80	0.56		

De las Tablas arriba se concluye que existen diferencias significativas tanto para la variable "Y" como para la variable "X". El hecho de que la variable auxiliar es significativo, esto indica que debemos verificar la dependencia de la "Y sobre la "X" (por medio de la regresión) y si se encuentra esta dependencia, entonces se debe de ajustar los valores de "Y observadas" o originales y correr un ANOVA sobre los valores ajustados de "Y" para poder descifrar el efecto puro de la *variable repuesta, ya que el efecto significativo de los valores observados de "Y" era influenciado por la variable auxiliar o "X".*

Cálculo de coeficiente de la regresión (b_{yx})
b_{yx} = 63.3/684.24 = **0.0925**

Conclusión.- Por cada aumento unitario de la variable "X" hay un aumento igual a 9.25% en la variable "Y".

Prueba de significancia de la pendiente (b_{yx})

Ho: $b = 0$ No hay dependencia de "Y" sobre la "X"
Ha: $b \neq 0$ Sí hay dependencia de "Y" sobre la "X"

La significancia de la pendiente se indica en la Tabla 16.3.

Tabla 16.3. ANOVA para la regresión.

FV	*gl*	SC	V
Regresión	1	14.10	14.10
Error	40	21.36	0.534
Total	41	35.46	

$SC_x = \Sigma x^2 - (\Sigma x)^2/n = 67{,}258 - 65{,}609.52 = \mathbf{1{,}648.48}$
$SC_r = b_2 {*} SC_x = (0.0925)^2(1{,}648.48) = \mathbf{14.10}$
$SC_T = \Sigma y^2 - (\Sigma y)^2/n = 999.7827 - 964.32 = \mathbf{35.46}$

$$EE_b = \sqrt{\frac{VE}{SCx}} = \sqrt{\frac{0.534}{1{,}648.48}} = \mathbf{0.0179}$$

Donde: EE_b = error estándar de la pendiente y V_E = varianza de error

$$t = \frac{bc - 0}{EEb} = \frac{0.0925 - 0}{0.0179} = \mathbf{5.1675}$$

Donde: b_c = valor calculado de la pendiente

$$\begin{array}{ccc} \text{VC} & \text{vs} & \text{Vt} \\ \downarrow & & \downarrow \\ 5.1675 & > & 2.021 \end{array} \left\{ \begin{array}{l} \alpha\text{: } 0.05 \\ gl\text{: } 42-2=40 \end{array} \right.$$

∴ "Ho" Se rechaza.

Conclusión: Existe dependencia significativa de "Y" sobre la "X".
Ahora se procede con la Tabla de ANCOVA y los tratamientos ajustados (Tabla 16.4).

Tabla 16.4. Tabla completa de Análisis de Covarianza (ANCOVA).

FV	gl	SC_X	SC_{XY}	SC_Y	Valores ajustados			$F_{Calculada}$
					gl	**SC**	**V**	
Total	41	1,645.45	137.60	35.46				
B (Bloques)	5	626.76	41.38	3.64				
t (tratamientos)	6	337.47	32.92	15.02				
E (Error)	30	684.24	63.30	16.8	29	**10.95**	0.37	1.96 / 0.37 = **5.29**
t + E (tratamientos más error)	36	1,021.71	96.22	31.82	35	**22.75**	0.65	
$t_{(Ajust)}$ (tratamientos ajustados)					6	**11.8**	1.96	

$\mathbf{SC_{E(y)(Ajust)} = SC_{E(y)} - [SC_{E(xy)}]^2/SC_{E(x)}} = 16.8 - (63.3)^2/684.24 = \mathbf{10.95}$
$\mathbf{SC_{t+E(Ajust)} = SC_{t+E(y)} - [\{SC_{t+E(xy)}\}^2 / SC_{t+E(x)}]} = 31.82 - (96.22)^2 / 1{,}021.71 = \mathbf{22.75}$
$\mathbf{SC_{Total(Ajust)} = SC_{t+E} - SC_E} = 22.75 - 10.95 = \mathbf{11.8}$
$\mathbf{F_{Calculada} = V_{t(Ajust)} / V_{E(Ajust)}} = 1.96 / 0.37 = \mathbf{5.29}$

VC vs VT $\left\{ \begin{array}{l} \alpha\text{: } 0.05 \\ \text{gl: } 6, 29 \end{array} \right.$

5.29 > 2.42

Conclusión.- Existe un efecto significativo con respecto a la variable respuesta o "Y", es decir, hay diferencias significativas entre las medias ajustadas (los cuales ya son puros e independiente del efecto de la variable auxiliar "X"). Hay que recordar que el ajuste es para liberar el efecto que causa la "X" sobre la "Y".

Ajustes de medias de tratamiento por:

$$\hat{y}_i = \bar{y}_i - b_{yx}(\bar{x}_i - \bar{\bar{x}}_i)$$

$\hat{y}_i$ = Media ajustada de tratamiento "i".
$\bar{y}_i$ = Media sin ajuste de tratamiento "i"
b_{yx} = Coeficiente de regresión = 0.0925
$\bar{x}_i$ = Media de la variable "X" del tratamiento "i"
$\bar{\bar{x}}_i$ = Media general de de la variable "X" por unidad experimental

La confirmación de la validez de las operaciones se indica en la Tabla 16.5.

Tabla 16.5. Confirmación de la validez de los cálculos.

t	$\bar{x}_i$	$(\bar{x}_i - \bar{\bar{x}}_i)$	$b_{yx}(\bar{x}_i - \bar{\bar{x}}_i)$	$\bar{y}_i$	$\hat{\bar{y}}_i$
1	35.33	-4.19	-0.39	3.45	3.84
2	43.66	4.15	0.38	4.81	4.43
3	38.33	-1.19	-0.11	4.87	4.98
4	41.16	1.65	0.15	4.65	4.5
5	41.33	1.81	0.17	5.23	5.06
6	40.83	1.31	0.12	5.36	5.24
7	36	-3.52	-0.33	5.19	5.52
∑	39.52	0.02	-0.01	33.56	33.57
Condiciones necesarias		Esta suma debe ser igual a 0	Esta suma debe ser igual a 0	Estas 2 sumas deben ser iguales	

Conclusiones de checar la validez de la operación (Tabla 16.5)

1. Las sumas de la tercera y la curta columnas (desde la izquierda) son casi iguales a cero, y de hecho la diferencia mínima de estas sumas del cero se debe al factor de redondeo.

2. Las suma de las últimas dos columnas son iguales, es decir, las sumatorias de las medias observadas de la variable respuesta ($\bar{y}_i$) y las medias ajustadas de la misma variable ($\hat{\bar{y}}_i$) son iguales.

3. Por tanto, al reunir estas dos condiciones se concluye que el manejo (la operación) de los datos de ANCOVA fue realizado de forma correcta.

Análisis factorial (ANOVA Factorial)

Descripción. El Análisis Factorial es el nombre genérico que se da a una clase de métodos estadísticos multivariantes (multivariados) cuyo propósito principal es sacar a la luz la estructura subyacente en una matriz de datos. Analiza la estructura de las interrelaciones entre un gran número de variables no exigiendo ninguna distinción entre variables dependientes e independientes. Utilizando esta información se calcula un conjunto de dimensiones latentes, conocidas como

factores, que buscan explicar dichas interrelaciones. Es, por lo tanto, una técnica de reducción de datos dado que si se cumplen sus hipótesis, la información contenida en la matriz de datos puede expresarse, sin mucha distorsión, en un número menor de dimensiones representadas por dichos factores.
El Análisis Factorial es una técnica estadística multivariado (multivariante) cuyo principal propósito es sintetizar las interrelaciones observadas entre un conjunto de variables en una forma concisa y segura como una ayuda a la construcción de nuevos conceptos y teorías. Para ello utiliza un conjunto de variables aleatorias inobservables, que se llaman ***factores comunes***, de forma que todas las covarianzas o correlaciones son explicadas por dichos factores y cualquier porción de la varianza inexplicada por los factores comunes se asigna a términos de error residuales llamados ***factores únicos o específicos***.

Requisitos. Un Análisis Factorial tiene sentido si se cumplen siguientes condiciones:

1. Parsimonia (principio de la navaja de Ockham).
2. Fácil de interpretar**.**
3. El orden de correr el experimento debe ser aleatorio.

Ventajas. Las s de este modelo constan de los siguientes puntos:

1. Optimiza, economiza los dos experimentos en uno.
2. Detecta el efecto de interacción entre los factores.

Objetivo. Es usado para explicar la variabilidad entre las variables observadas en términos de un número menor de variables no observadas llamadas factores.

Ejemplo 17. Determinar el efecto de (a) tres corrientes (3 niveles) de derecho y (b) los años de especialidad (2 niveles, uno bajo y otro alto) sobre la eficacia de culminar juicios durante 2 años (2 repeticiones).

Por tanto aquí tenemos dos factores:

1. Factor **A** lo cual es los corrientes o escuelas de derecho (CD) con 3 niveles o 3 escuelas distintas.
2. Factor **B** lo cual está constituida por los años de especialidad (AE) con 2 niveles.

Queremos medir el efecto de cada factor sobre la eficacia de culminar los juicios y también la probable interacción entre los 2 factores (Tabla 17).

Tabla 17. Datos para el análisis factorial[$].						
Rep	CD1 AEb	CD2 AEb	CD3 AEb	CD1 AEa	CD2 AEa	CD3 AEa
1	48	28	7	62	14	6
2	58	33	15	54	10	9
∑	**106**	**61**	**22**	**116**	**24**	**15**
$: CD: corriente de derecho con 3 niveles (1, 2 y 3), AE: años de especialidad con 2 niveles alto y bajo						

Solución:

Primero, hay que arreglar los datos para un Diseño completamente aleatorio (DCA) (Tabla 17.1).

Tabla 17.1. Templete de la Tabla ANOVA para DCA.				
Fuente de Variación	***gl***	**SC**	**Varianza**	**FC**
Tratamientos (efecto explicado = suma de 2 factores)	k-1	$\Sigma(Tcg)^2/r_{cg}$ – FC	SC_t/gl_t	V_t / V_E
Error	gl_T – gl_t	$SC_T – SC_t$	SC_E/gl_E	
Total	n-1	Σx_i^2 – FC		
$FC = (\Sigma x_i)^2/N_T$				

<u>Cálculo de SC:</u>

$\boldsymbol{FC = (\Sigma x_i)^2/N_T}$
$FC = (344)^2 / 12 = 118.336/12 = \mathbf{9{,}861.33}$
$\boldsymbol{SC_{Total} = \Sigma x_i^2 – FC}$
$SC_{Total} = 14{,}988 – 9{,}861.33 = \mathbf{5{,}126.67}$

$\boldsymbol{SC_{tratamientos} = \Sigma(Tcg)^2/r_{cg} – FC}$
$SC_{tratamientos} = [\{(106)^2 + (61)^2 +…+ (15)^2\} / 2] - 9{,}861.33$
$SC_{tratamientos} = 14{,}849 – 9{,}861.33 = \mathbf{4{,}987.67}$

$\boldsymbol{SC_{Error} = SC_{Total} - SC_{tratamientos}}$
$SC_{Error} = 5{,}126.67 – 4{,}987.67 = \mathbf{139}$

Resultados de Diseño completamente al azar se indican en la Tabla 17.2.

Tabla 17.2. ANOVA (Diseño Completamente al Azar = DCA).					
FV	***gl***	**SC**	**Varianza**	**FC**	**FT**
Tratamientos	5	4,987.67	997.53	43.07*	4.39
Error	6	139	23.16		
Total	11	5,126.67			

Conclusión.- Se demuestra que hay un efecto significativo de los 6 tratamientos. Por consiguiente, hay que cambiar el diseño de DCA al Diseño de Bloques al Azar (DBA) para determinar si hay efecto entre bloques (Tabla 17.3).

Tabla 17.3. Datos para diseño de bloques al azar (DBA).				
Años de especialidad (AE) (Factor "B") (bloques)	**Corrientes de derecho (CD) (Factor "A")**			**∑**
	CD1	**CD2**	**CD3**	
AEb	48 58	28 33	7 15	**189**
AEa	62 54	14 10	6 9	**155**
∑	**222**	**85**	**37**	**GT (Gran total) = 344**

$\boldsymbol{FC} = \mathbf{(GT)^2/\ N_T} = (344)^2/12 = \mathbf{9{,}861.33}$

<u>Cálculo de SC para Factor "A"</u>

$\boldsymbol{SC_A = \Sigma(T_{FA})^2/r_{FA} - FC}$

$SC_A = [(222)^2/4 + (85)^2/4 + (37)^2/4] - 9{,}861.33$

$SC_A = [12{,}321 + 1{,}806.25 + 342.25] - 9{,}861.33$

$\boldsymbol{SC_A = 4{,}608.17}$

<u>Cálculo de SC para Factor "B"</u>

$\boldsymbol{SC_B = \Sigma(T_{FB})^2/r_{FB} - FC}$

$SC_B = [(189)^2/4 + (155)^2/4] - 9{,}861.33$

$SC_B = [5{,}953 + 4{,}004.16] - 9{,}861.33$

$SC_B = \mathbf{96.33}$

<u>Cálculo de SC para Total</u>

$\boldsymbol{SC_T = \Sigma X_i^2 - FC}$
$SC_T = (48^2 + 28^2 + \ldots + 9^2)$ - FC
$SC_T =$ **5,126.67**

Efecto principal = Efecto de Factor A + efecto de factor B
$\mathbf{SC_{Efecto\ principal} = SC_{Efecto\ A} + SC_{Efecto\ B}}$
$\mathbf{SC_{Efecto\ principal} = 4{,}608.17 + 96.33 = 4{,}704.5}$

Efecto explicado = Efecto principal + Efecto de interacción
$\mathbf{SC_{Efecto\ Explicado} = SC_{Efecto\ principal} + SC_{Efecto\ interacción}}$
$\mathbf{4{,}987.67 = 4{,}704.5 + SC_{Efecto\ interacción}}$
$\mathbf{SC_{Efecto\ interacción} = 4{,}987.67 - 4{,}704.5}$
$\mathbf{SC_{Efecto\ interacción} = 283.17}$

Resultados de Diseño de bloques al azar se indican en la Tabla 17.4.

Tabla 17.4. ANOVA para datos de la Tabla 17.3 (arreglado en forma de DBA).

FV	gl	SC	V	FC	FT
Factor A	2	4,608.17	2,304.08	99.48*	5.14
Factor B	1	96.33	96.33	4.15	5.99
A x B	2	283.17	23.16		
Error	6	139			
Total	11	5,126.67			

Conclusión.- No se ve un efecto significativo en bloques (factor B), solo se sigue observando el efecto significativo de los tratamientos (factor A). Ahora se debe realizar una tabla de ANOVA factorial y ver el efecto de los factores por separado con sus interacciones (Tabla 17.5).

<u>De nuevo para refrescar la memoria se notan los siguientes SC´s:</u>

$\mathbf{\mathit{SC_{Principal}} = SC_A + SC_B}$
$\mathbf{\mathit{SC_{Explicado}} = SC_{Principal} + SC_{Interacción} = SC_{tratamientos}}$ **del ANOVA de DCA (**4,987.67**)**
$\mathbf{\mathit{SC_{Interacción}} = SC_{Explicado} - SC_{Principal}}$

$\mathit{SC_{Principal}} = 4{,}680.17 + 96.33 = \mathbf{4{,}704.5}$
$\mathit{SC_{Interacción}} = 4{,}987.67 - 4{,}704.5 = \mathbf{283.17}$

Tabla 17.5. Tabla completa de ANOVA factorial.

FV	gl	SC	V	$F_{Calculada}$	$F_{Tabulada}$
Efecto Principal (A+B)	3	4,704.50	1,568.16	67.70*	4.76
Factor "A"	2	4,608.17	2,304.08	99.48*	5.14
Factor "B"	1	96.33	96.33	4.15	5.99
Interacción ("AB")	2	283.17	141.58	6.11*	5.14
Efecto Explicado (A+B + "AB")	5	4,987.67	997.53	43.07*	4.39
Error (Total - Explicado)	6	139.00	23.16		
Total	11	5,126.67			

Conclusión.- Hay diferencia significativa entre los promedios de los 6 tratamientos. Los tratamientos se dividen en A y B. El factor A sale significativo, pero el factor B no. El término significativo de la interacción nos indica que el comportamiento de un factor no es igual cuando está solo a diferencia de que este en la presencia de otro factor. Hay explicación significativa en la interacción de los factores. Por ende se concluye que:

El factor "A" de corrientes de derecho (CD) no se comporta igual cuando hay nivel bajo de años de especialidad del facto "B" (AEb) en comparación con cuando hay nivel alto (AEa), y este es la esencia de la interacción en la estadística, la cual no se hubiera podido lograr si hubiéramos realizado 2 experimento (uno para cada factor) en lugar de un experimento factorial.
Por tanto, el análisis factorial es: a) económico, ya que ahorra recursos en conducir un experimento en lugar de realizar varios experimentos (dependiendo en el número de los factores involucrados), y b) este análisis detecta el efecto de la interacción entre los factores lo cual es indetectable en los experimentos no-factoriales.

Conclusiones

En muchos campos de estudios una pregunta común es si la muestra o las muestras realmente representan o reflejan la población de interés. Una equivocada representatividad en este sentido puede terminar en tomar decisiones erróneas con subsecuente pérdida, abuso o mal uso de recursos financiero, material, temporal, etc. La selección y la utilización apropiada de los métodos estadísticos evitarán cometer estos tipos de errores. Este trabajo simple precisamente demuestra el uso y el empleo adecuado de las técnicas correctas.

Referencias

Badii, M.H., A.R. Pazhakh, J.L. Abreu & R. Foroughbakhch. 2004. Fundamentos del método científico. InnOvaciOnes de NegOciOs, 1(1): 89–107.

Badii, M.H., J. Castillo & A. Wong. 2006. Diseños de distribución libre. InnOvaciOnes de NegOciOs, 3(1): 141-174.

Badii, M.H., J. Castillo, R. Rositas & G. Ponce. 2007a. Experimental designs. Pp. 335-348. In: M.H. Badii & J. Castillo (eds.). Técnicas Cuantitativas en la Investigación. UANL, Monterrey.

Badii, M.H., J. Castillo, F. Gorjón & R. Foroughbakhch. 2007b. Completely randomized designs. Pp. 307-334. In: M.H. Badii & J. Castillo (eds.). Técnicas Cuantitativas en la Investigación. UANL, Monterrey.

Badii, M.H., J. Castillo, K. Cortes & H. Quiroz. 2007c. Análisis de clusters. Pp. 15-36. In: M.H. Badii & J. Castillo (eds.). Técnicas Cuantitativas en la Investigación. UANL, Monterrey.

Badii, M.H., J. Castillo, J.N. Barragán & A.E. Flores. 2007d. Análisis discriminante. Pp. 119-136. In: M.H. Badii & J. Castillo (eds.). Técnicas Cuantitativas en la Investigación. UANL, Monterrey.

Badii, M.H., J. Castillo, J. Rositas & G. Alarcón. 2007e. Uso de un método de pronóstico en investigación. Pp. 137-155. In: M.H. Badii & J. Castillo (eds.). Técnicas Cuantitativas en la Investigación. UANL, Monterrey.

Badii, M.H. & J. Castillo. 2009. Muestreo Estadístico: Conceptos y Aplicaciones. 225 pp. UANL, Monterrey.

Badii, M.H., J. Castillo, M. Rodríguez & K. Cortez. Papel de la estadística en la investigación científica. Pp. 1-43. En. Badii, M.H. & J. Castillo (eds). Desarrollo Sustentable: Métodos, Aplicaciones y Perspectivas. UANL, Monterrey.

Montgomery, D.C. 2001. Design of Analisis of Experiments. 15th Ed., Wiley. N.Y.

Ostle, B. 1986. Estadística aplicada. Limusa -Wiley S.A., México.

Rositas, J., M.H. Badii, J. Castillo & R. Foroughbakhch. 2007e. Técnicas de investigación basadas en sistemas de modelación estructurada. Pp. 291-304. In: M.H. Badii & J. Castillo (eds.). Técnicas Cuantitativas en la Investigación. UANL, Monterrey.

Snedecor, G. & W. Cochran. 1994. Statistical Methods. 8th Ed. Iowa State University Press. Ames.

Spiegel, M. 1991. Estadística. McGraw-Hill. México.

***** UANL, San Nicolás, N.L., México, mhbadiiz@gmail.com

Anexo 1. Tablas parciales estadísticas.

Tabla parcial de X^2								
gl	**Valores de α**							
	0.50	**0.25**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**	**0.001**
1	0.455	1.325	2.7066	3.841	5.024	6.435	7.879	10.828
2	1.386	2.775	4.605	5.991	7.378	9.210	10.597	13.816
3	2.366	4.108	6.251	7.815	9.348	11.345	12.838	16.266
4	3.357	5.385	7.779	8.488	11.143	13.277	12.860	18.467
5	4.351	6.626	9.236	11.070	12.833	15.086	16.750	20.515
6	5.348	7.841	10.645	12.592	14.449	16.812	18.548	22.458
7	6.346	9.037	2.017	14.067	16.013	18.475	20.276	24.322
8	7.344	10.219	13.362	15.507	17.535	20.090	21.955	26.124
9	8.343	11.389	14.684	16.919	19.023	21.666	23.589	27.877
10	9.342	12.540	15.987	18.307	20.483	23.209	25.188	29.607
…								
140	139.334	150.894	161.827	168.613	174.648	181.840	180.847	197.451

Tabla parcial de distribución F de Fisher a nivel de $\alpha = 0.05$											
Gl	1	2	3	4	5	6	7	8	9	10	…	120
1	161.45	199.5	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88		253.25
2	18.513	19.00	19.164	19.247	19.296	19.329	19.353	19.371	19.385	19.396		19.49
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785		8.549
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964		5.658
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735		4.398
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060		3.705
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637		3.267
8	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347		2.967
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137		2.748
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978		2.580
…												
120	3.920	3.072	2.680	2.447	2.290	2.175	2.087	2.016	1.959	1.910		1.352

Tabla parcial de t

α (1) α (2)	0.50 0.25	0.20 0.10	0.10 0.05	0.05 0.025	0.02 0.01	0.01 0.005	0.005 0.0025	0.002 0.001	0.001 0.0005
1	1.000	3.078	6.314	12.706	31.821	63.657	127.321	318.309	636.619
2	0.816	1.886	2.920	4.303	6.965	9.925	14.089	22.327	31.599
3	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.215	12.924
4	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.727	1.476	2.015	2.571	3.365	4.032	4.775	5.893	6.869
6	0.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
…	…	…	…	…	…	…	…	…	…
∞	0.6745	1.2816	1.5449	1.960	2.3263	2.5758	2.8070	3.0902	3.2905