APUNTES DE DISEÑO DE EXPERIMENTOS.

Úrsula Torres Parejo

Índice

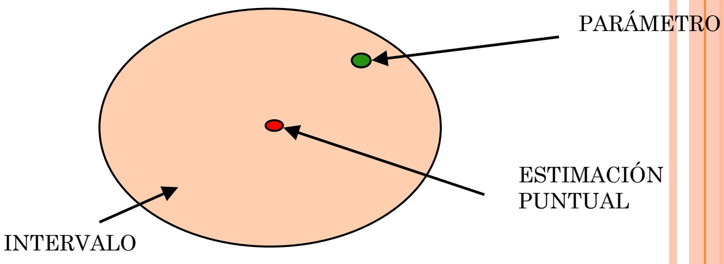
Contraste de hipótesis. Repaso.

INTRODUCCIÓN

La **Inferencia** es el Proceso mediante el cual se utiliza la información de una muestra para extraer conclusiones de la población:

1. **Estimación**: Utiliza la muestra para estimar las características de la población.

    a. **Estimación Puntual**
    b. **Estimación por Intervalos**



2. **Contraste de Hipótesis**: Emite hipótesis sobre las características de la población y comprueba su veracidad.

NOTACIÓN.

El objetivo de la **estimación** de parámetros es proveer de métodos que permitan determinar, con cierta precisión, el valor de los parámetros de un modelo a partir de una muestra extraída de la población.

En la población	Su equivalente	**En la muestra**
Media poblacional: μ Varianza poblacional: σ^2 Proporción poblacional: π	→	Media muestral: $\overline{x}$ Varianza muestral: S^2 Proporción muestral: p

Definiciones.

Una **hipótesis estadística** es una proposición referente a una o varias poblaciones. *Ejemplos: la probabilidad de que salga cara es 0.5, la media poblacional es menor que 9, la varianza poblacional ha aumentado, etc.*

Un **contraste** o **test de hipótesis** es un procedimiento inferencial para decidir cuál de las dos hipótesis debe aceptarse o rechazarse en base a la información obtenida a partir de una muestra.

Hipótesis Nula (H_0): Es la hipótesis de partida: generalmente, la que se viene admitiendo como cierta hasta el momento, la más estable o intuitiva. **Hipótesis que debemos refutar sólo si tenemos una alta evidencia en contra**.

Hipótesis Alternativa (H_1): Es una alternativa a la hipótesis nula que dependerá del contexto.

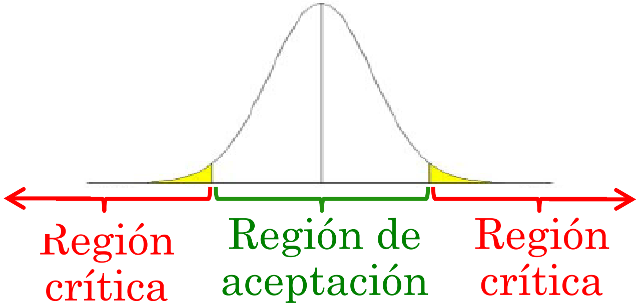
Pasos a seguir en un contraste de hipótesis:

1. **Hipótesis:** Concretar la hipótesis nula (H_0) y formular una hipótesis alternativa (H_1)
2. **Estadístico de Contraste:** Calcular el estadístico de contraste que corresponda a partir de los valores de la muestra.
3. **Regiones de aceptación y de rechazo.** Construir las regiones de aceptación y de rechazo para una nivel de significación establecido.
4. **Decisión:** Si el valor del estadístico pertenece a la región de rechazo, habría que rechazar H_0. Si el estadístico de contraste pertenece a la región de aceptación, habría que aceptar H_0.
5. **Conclusión:** Dar respuesta al problema planteado.

Tipos de contrastes paramétricos.

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

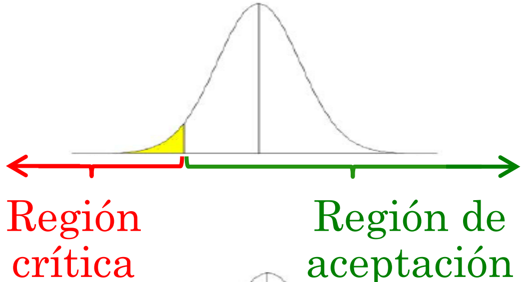
BILATERALES

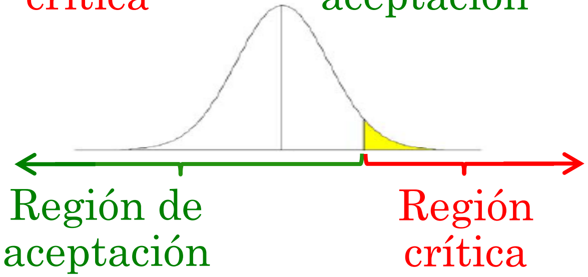


$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases}$$

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

UNILATERALES





EJEMPLO.

Contraste de Hipótesis para la media, con desviación típica conocida.

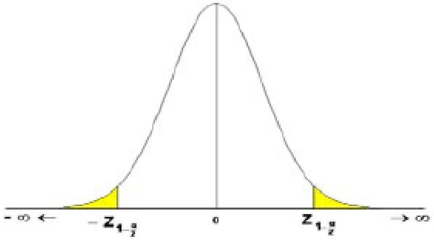
$$H_0 : \mu = \mu_0$$

Estadístico de Contraste $$z_{\exp} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \rightarrow N(0,1)$$

<u>Contrastes Bilaterales:</u>

$$H_1 : \mu \neq \mu_0 \;\Rightarrow\; \text{R.C} = \left\{ |\, z_{\exp} \,| > z_{1-\alpha/2} \right\}$$

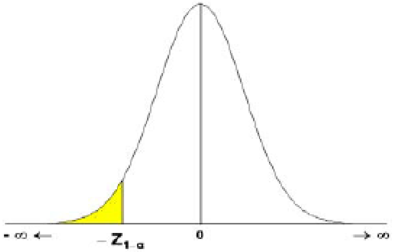
P-valor=2*P[Z>|Z_{exp}|]



<u>Contrastes Unilaterales:</u>

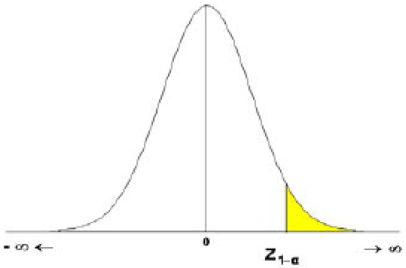
$$H_1 : \mu < \mu_0 \;\Rightarrow\; \text{R.C} = \left\{ z_{\exp} < -z_{1-\alpha} \right\}$$

P-valor=P[Z<Z_{exp}]



$$H_1 : \mu > \mu_0 \;\Rightarrow\; \text{R.C} = \left\{ z_{\exp} > z_{1-\alpha} \right\}$$

P-valor=P[Z>Z_{exp}]



P-VALOR.

Se denomina **p-valor** o **valor p** a la probabilidad de encontrar un valor más extremo que el obtenido para el estadístico de contraste.

Para un nivel de significación α:

Si p-valor $\leq \alpha$, rechazamos H_0

Si p-valor $> \alpha$, aceptamos H_0

ERRORES.

	DECISIÓN	
	Acepto H_0	**Acepto H_1**
Cierto H_0	✔	Error I
Cierto H_1	Error II	✔

Prob(Error I)=P(Rechazar H_0 | H_0 cierta) = α
(α = nivel de significación del test)

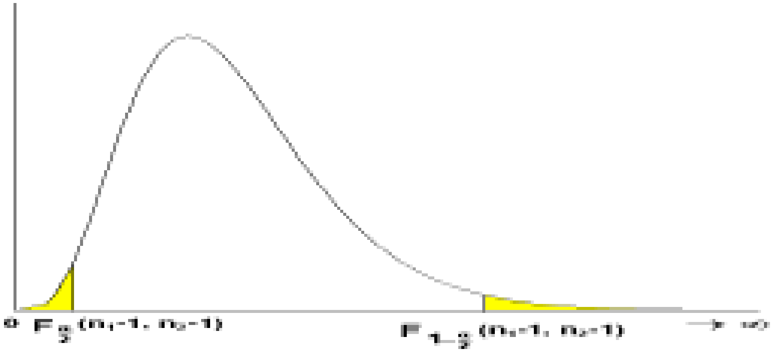
Prob(Error II)=P(Aceptar H_0 | H_0 falsa) = β
($1 - \beta$ = potencia del test)

EJERCICIO PROPUESTO.

En una empresa se sospecha que los salarios de sus empleados no son igualmente homogéneos en el sector administrativo que en el sector del personal de seguridad. Para realizar un estudio histórico sobre la homogeneidad de los salarios, se selecciona al azar una muestra de 5 administrativos y otra de 4 empleados del sector seguridad, siendo la cuasivarianza del salario de los administrativos 5,3 euros^2 y la cuasivarianza del salario del personal de seguridad 5,7 euros^2. A la vista de los datos y admitiendo normalidad en la distribución de salarios, contrastar la hipótesis de igualdad de varianzas para probar si las sospechas son ciertas. Utilizar una significación del 10%.

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \neq \sigma_2^2 \end{cases}$$

$$F_{\exp} = \frac{S_{C1}^2}{Sc_2^2} = \frac{5.3}{5.7} = 0.929$$



$$H_1 : \sigma_1^2 \neq \sigma_2^2 \Rightarrow \text{R.C} = \left\{ F_{\exp} < \text{F}_{n_1-1, n_2-1;\ \alpha/2} \right\} \cup \left\{ F_{\exp} > \text{F}_{n_1-1, n_2-1; 1-\alpha/2} \right\}$$

v \ u	1	2	3	4	5	6	7	8	9	10	11	12	p
1	39,863	49,500	53,593	55,833	57,240	58,204	58,906	59,439	59,858	60,195	60,473	60,705	0,900
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	242,98	243,91	0,950
1	647,79	799,50	864,16	899,58	921,85	937,11	948,22	956,66	963,28	968,63	973,03	976,71	0,975
1	4052,2	4999,5	5403,4	5624,6	5763,6	5859,0	5928,4	5981,1	6022,5	6055,8	6083,3	6106,3	0,990
1	16211	19999	21615	22500	23056	23437	23715	23925	24091	24224	24334	24426	0,995
2	8,526	9,000	9,162	9,243	9,293	9,326	9,349	9,367	9,381	9,392	9,401	9,408	0,900
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385	19,396	19,405	19,413	0,950
2	38,506	39,000	39,165	39,248	39,298	39,331	39,355	39,373	39,387	39,398	39,407	39,415	0,975
2	98,503	99,000	99,166	99,249	99,299	99,333	99,356	99,374	99,388	99,399	99,408	99,416	0,990
2	198,50	199,00	199,17	199,25	199,30	199,33	199,36	199,37	199,39	199,40	199,41	199,42	0,995
3	5,538	5,462	5,391	5,343	5,309	5,285	5,266	5,252	5,240	5,230	5,222	5,216	0,900
3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,812	8,786	8,763	8,745	0,950
3	17,443	16,044	15,439	15,101	14,885	14,735	14,624	14,540	14,473	14,419	14,374	14,337	0,975
3	34,116	30,817	29,457	28,710	28,237	27,911	27,672	27,489	27,345	27,229	27,133	27,052	0,990
3	55,552	49,799	47,467	46,195	45,392	44,838	44,434	44,126	43,882	43,686	43,524	43,387	0,995
4	4,545	4,325	4,191	4,107	4,051	4,010	3,979	3,955	3,936	3,920	3,907	3,896	0,900
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,999	5,964	5,936	5,912	0,950
4	12,218	10,649	9,979	9,605	9,364	9,197	9,074	8,980	8,905	8,844	8,794	8,751	0,975
4	21,198	18,000	16,694	15,977	15,522	15,207	14,976	14,799	14,659	14,546	14,452	14,374	0,990
4	31,333	26,284	24,259	23,155	22,456	21,975	21,622	21,352	21,139	20,967	20,824	20,705	0,995
5	4,060	3,780	3,619	3,520	3,453	3,405	3,368	3,339	3,316	3,297	3,282	3,268	0,900

$$\boxed{F_{u,v;p} = F_{4,3;0.95} = 9.12}$$

$$\boxed{F_{4,3;0.05} = \frac{1}{F_{3,4;0.95}} = \frac{1}{6.59} = 0.15}$$

Región Crítica: $F_{exp} < 0.15$ ó $F_{exp} > 9.12$
Como $0.15 < 0.929 < 9.12$ $\Longrightarrow$ **Acepto H_0**

Conclusión: No existen evidencias estadísticas suficientes al 10% de significación para rechazar la igualdad de varianzas poblacionales, por lo que las sospechas no son ciertas.

TEMA 5.1. INTRODUCCIÓN AL DISEÑO ESTADÍSTICO DE EXPERIMENTOS.

5.1.1. Elementos Intervinientes en un Experimento.
5.1.2. Objetivos Perseguidos.
5.1.3. La Importancia de Planificar la Experimentación.
5.1.4. Etapas de un Experimento.
5.1.5. El Método Tradicional de Experimentación.
5.1.6. Tipos de Variabilidad.
5.1.7. Principios Básicos del Diseño de Experimentos.
5.1.8. Principales Diseños Estadísticos.

En la investigación empírica, cuando se plantea una hipótesis se proyecta un experimento cuyos resultados deben conducir a la aceptación o al rechazo de dicha hipótesis, con el menor riesgo posible de equivocación.

Diseño Estadístico de Experimentos

- Conjunto de Técnicas de Análisis
- Métodos de Construcción de Modelos

Permiten llevar a cabo el proceso

↓

Planificar un experimento

↓

Obtener datos apropiados

↓

Que puedan ser analizados con métodos estadísticos

↓

Para obtener conclusiones válidas y objetivas

5.1.1. Elementos intervinientes en un experimento

1. **Unidades experimentales:** (personas, elementos físicos,…)
2. **Variable respuesta:** variable de interés.
3. **Factores:** variables controladas por el experimentador (niveles de factor o tratamientos).
4. **Error experimental o perturbación:** variables no controladas por el experimentador.
5. **Tamaño del experimento:** número total de observaciones.

Objetivo: Estudiar el efecto que producen los factores sobre la Variable Respuesta.

Ejemplos de sistemas experimentales:

1. Una reacción química, cuyo rendimiento (y) puede ser función, entre otros, del tiempo de reacción (x_1), la temperatura de reacción (x_2), y el tipo de catalizador utilizado (x_3).

   y = variable respuesta

   x_i = factores

   Otras variables que pueden influir son, por ejemplo, la pureza de los reactivos, la limpieza del material, la velocidad de agitación…. (error experimental).

2. Un alimento, producido por mezcla de distintas proporciones de ingredientes (x), lo cual da lugar a distintos olores y sabores (y).
3. Una separación cromatográfica, donde el tiempo de separación (y) depende del pH (x_1) y del porcentaje de modificador orgánico de la fase móvil (x_2)
4. El rendimiento de un determinado tipo de máquina, medido por las unidades producidas (y), que depende del trabajador que la maneja (x_1) y de la marca de la máquina (x_2).

Factores y sus niveles:

Se denomina **factor** a cualquier variable de interés para el experimento. Los **niveles de factor** son los tipos o grados específicos del factor.

Ejemplos de factores cualitativos:

1. Proveedor: distintos proveedores de materia prima.
2. Máquina: diferentes tipos o marcas de máquinas.
3. Trabajador: distintos trabajadores encargados de hacer una tarea.

Ejemplos de factores cuantitativos:

1. Tamaño de memoria: diferentes tamaños de memoria de ordenadores.
2. Temperatura: conjuntos de temperaturas seleccionadas en unos rangos de interés.

Los factores cuantitativos se tratan como cualitativos.

Tratamientos.

Son combinaciones específicas de los niveles de los factores en estudio:

- En un diseño con un único factor, son los distintos niveles del factor.
- En un diseño con varios factores, son las distintas combinaciones de niveles de factores.

5.1.2. Objetivos perseguidos

1. **Obtener un conocimiento inicial del sistema de estudio.** *¿en qué valores de los factores se puede centrar la investigación?*
2. **Determinar la influencia de los factores sobre las respuestas observadas.** *¿Qué factores influyen más? ¿Cómo interaccionan entre ellos?*
3. **Optimizar respuestas.** *¿Qué valores de los factores proporcionan las respuestas de mayor calidad?*
4. **Determinar la robustez del sistema.** *¿Cómo afectan al sistema variaciones no controladas en el valor de los factores?*

5.1.3. La importancia de planificar la experimentación

La experimentación juega un papel fundamental en todos los campos de la investigación y el desarrollo. Nos permite obtener información de calidad que permita **desarrollar nuevos productos** y procesos, **comprender** mejor un sistema y tomar decisiones de cómo **optimizarlo** y mejorar su calidad, comprobar **hipótesis científicas**, etc.

La experimentación se debe planificar principalmente porque:

1. **Normalmente es cara.** La capacidad de experimentar está limitada por el coste en tiempo y en recursos. Por lo tanto, una organización óptima deberá buscar el menor número de experimentos que permita obtener la información buscada.
2. El resultado de un experimento está sujeto a **incertidumbre**. Por tanto, se planificará un diseño que minimice la influencia del error experimental sobre la información buscada.

CONSIDERACIONES AL PLANTEAR EL PROBLEMA:

1. ¿Qué se conoce y qué no se conoce?
   a. ¿Hay zonas de la región experimental donde ya se conoce el resultado?
   b. ¿Qué complejidad se espera en la relación entre los factores y la respuesta? ¿podría ser no lineal?
   c. ¿Podrían existir interacciones?
   d. ¿Cuál es el coste permitido de la experimentación?
   e. ¿Con qué rapidez es necesario proporcionar los resultados?
2. ¿Qué se necesita investigar?
   a. ¿Cuál es el objetivo de la experimentación?
   b. ¿Qué información queremos proporcionar?

5.1.4. Etapas de un experimento

1. **Diseño** del experimento con una estructura lo más adecuada posible a la situación que se desea estudiar y a los medios disponibles:
   a. **Planteamiento** general del problema y de los objetivos que se persiguen.
   b. Selección y definición de la **variable respuesta**.
   c. Elección de los **factores** y niveles que han de intervenir en el experimento.
   d. Determinación del conjunto de **unidades experimentales** incluidas en el estudio.
   e. Determinación del **modelo** que explica la variable respuesta en función de los factores.

2. Realizar la **experimentación** de acuerdo con el plan establecido previamente en el diseño.
3. **Analizar estadísticamente** los resultados obtenidos y comprobar si las hipótesis establecidas y el modelo de diseño elegido se adecuan a la situación estudiada.
4. Realizar las **modificaciones** oportunas para ampliar o modificar el diseño.
5. Obtener las **conclusiones** apropiadas.

El diseño de experimentos tiene como objetivo reducir el error experimental de forma que los posibles efectos de los factores de interés se manifiesten más claramente.

5.1.5. El método tradicional de experimentación.

Consiste en variar un factor cada vez.

1. A partir de unas condiciones iniciales, se realizan experimentos, en los cuales, se mantienen constantes todos los factores excepto uno.
2. De este modo, la variación de la respuesta se puede atribuir a la variación del factor, revelando su efecto.
3. El procedimiento se repite para los otros factores.

Los inconvenientes de este método son:

1. No informa como un factor interactúa con otros factores ni cómo estas interacciones afectan a la respuesta.
2. No proporciona el valor óptimo.
3. Requiere de un gran número de pruebas.
4. Es inviable, en muchos casos, por problemas de tiempo o costo.

¿Qué método utilizar entonces?

Uno que permita estudiar simultáneamente todos los factores.

5.1.6. Tipos de variabilidad

1. **Variabilidad sistemática y planificada:** Debida a diferencias sistemáticas entre las distintas condiciones experimentales impuestas por deseo expreso del experimentador. Esta variabilidad es deseable y se intenta identificar y cuantificar.
2. **Variabilidad típica de la naturaleza del problema y del experimentador:** Debida a la variabilidad no planificada, denominada error de medida. Es impredecible e inevitable, pero es posible estimar su valor y ver cómo se comporta. Es tolerable.

3. **Variabilidad sistemática y no planificada:** Debida a causas desconocidas y no planificadas. Es la causante de conclusiones erróneas y estudios incorrectos. Existen dos estrategias para evitarla:
   a. **La aleatorización.**
   b. **La técnica de bloques.**

5.1.7. Principios básicos del diseño de experimentos.

1. **Aleatorización**: La asignación de las unidades experimentales a los distintos tratamientos y el orden en el que se realizan los ensayos se determina al azar.
2. **Replicación**: El experimento se puede repetir con la finalidad de aumentar su precisión.
3. **Homogeneidad del material experimental:** Se deben dividir o particionar las unidades experimentales en grupos llamados "bloques", de modo que las observaciones realizadas en cada bloque se tomen bajo las condiciones experimentales lo más parecidas posible.

5.1.8. Principales diseños estadísticos.

1. Diseño Completamente Aleatorizado.
2. Diseño en Bloques Aleatorizados.
3. Diseño en Bloques Incompletos Aleatorizado.
4. Diseño en Cuadrados Latinos.
5. Diseño en Cuadrados Grecolatinos.
6. Diseño en Cuadrados de Youden.
7. Diseños Factoriales.

TEMA 5.2. DISEÑO COMPLETAMENTE ALEATORIZADO.

5.2.1. Modelo Estadístico.

5.2.2. Objetivos.

5.2.3. Situaciones.

5.2.4. Estudio de un Experimento. Pasos a Seguir.

5.2.5. Análisis del Modelo.

5.2.6. Diagnosis y Validación del Modelo.

5.2.7. Comparaciones Múltiples.

El experimentador asigna los tratamientos a las unidades experimentales al azar. La única restricción es el número de observaciones que se toman en cada tratamiento.

Este tipo de diseño se utiliza en experimentos en los que no influyen factores bloque.

El modelo matemático tiene la forma:

Respuesta = Constante + Efecto Tratamiento + Error

Ejemplos:

1. Una compañía algodonera que emplea diversos fertilizantes desea comprobar si éstos tienen efectos diferentes sobre el rendimiento de la semilla de algodón.
2. Una industria química, que obtiene un determinado producto, está interesada en comprobar si los cambios de temperatura influyen en la cantidad de producto obtenido.
3. Una profesora de estadística que imparte en grupos experimentales de alumnos, en los que explica la misma materia pero siguiendo distintos métodos de enseñanza, desea comprobar si el método de enseñanza utilizado influye en las calificaciones de los alumnos.

INTERÉS → Un solo factor con varios niveles o tratamientos.

OBJETIVO → Comparar entre sí varios grupos o tratamientos.

TÉCNICA ESTADÍSTICA → Análisis de la varianza de un factor o vía.

MÉTODO → Descomposición de la variabilidad total de un experimento en componentes independientes.

Otros factores que influyen:

1. Pequeñas variaciones en la cantidad de riego, en la pureza de los insecticidas suministrados, etc.
2. La pureza de la materia prima, la habilidad de los operarios, etc.
3. El nivel de preparación del alumno, su grado de atención e interés, etc.

Teóricamente, es posible dividir la variabilidad en dos partes:

-La originada por el factor de interés.

-La producida por los restantes factores (error experimental)

5.2.1. Modelo estadístico.

$$\boxed{y_{ij} = \mu + \tau_i + u_{ij}; \quad i:1,\ldots,I; \;\; j:1,\ldots,n_i}$$

y_{ij} : Variable aleatoria que representa la observación j-ésima del i-ésimo tratamiento.

μ : Efecto constante, común a todos los niveles. Media global.

τ_i : Efecto del tratamiento i-ésimo

u_{ij} : Variables aleatorias que engloban un conjunto de factores (errores experimentales o residuos)

Los residuos deben satisfacer las siguientes propiedades:

- La media es cero: $E[u_{ij}] = 0 \quad \forall\ i, j$
- La varianza es constante: $Var[u_{ij}] = \sigma^2 \quad \forall i, j$
- Son independientes entre sí: $E[u_{ij} u_{rk}] = 0 \ \ i \neq r \ \text{ o } \ j \neq k$
- Siguen una distribución Normal.

5.2.2. Objetivos.

1. Estimar los parámetros del modelo.
2. Contrastar la hipótesis de igualdad de los tratamientos.

$$\boxed{\begin{array}{l} H_0 : \tau_i = 0 \ \ \forall i \ \ \text{(Todos los tratamientos producen el mismo efecto)} \\ H_1 : \tau_i \neq 0 \text{ para algún } i \text{ (Al menos dos tratamientos difieren entre sí)} \end{array}}$$

o equivalentemente

$$\boxed{\begin{array}{l} H_0 : \mu_1 = \mu_2 = \ldots = \mu_I = \mu \quad \text{(Todos los tratamientos tienen la misma media)} \\ H_1 : \mu_i \neq \mu_j \text{ por lo menos para algún par } (i, j) \end{array}}$$

3. Comprobar la idoneidad del modelo.
4. Realizar comparaciones múltiples.

5.2.3. Situaciones.

- Dependiendo de los efectos:
  - **Modelo de efectos fijos**.
  - **Modelo de efectos aleatorios.**
- Dependiendo del tamaño muestral:
  - **Modelo equilibrado o balanceado**: Todas las muestras del mismo tamaño ($n_i = n$).
  - **Modelo no equilibrado o no balanceado**: Todas las muestras no son del mismo tamaño.

5.2.4. Estudio de un experimento. Pasos a seguir.

1. Plantear un modelo que explique los datos.
2. Examinar la adecuación del modelo planteado:

    Detectar graves desviaciones de las hipótesis supuestas en el modelo (**Independencia de los residuos, normalidad de los residuos y homocedasticidad**).

   Si el modelo no es el adecuado tomar medidas correctoras, como transformaciones de los datos o modificar el modelo.
3. Si el modelo es el adecuado se realiza el análisis estadístico de los datos y se evalúa su grado de ajuste.

5.2.5. Análisis del modelo.

Tratamientos		n_i	$y_{i\bullet}$	$\bar{y}_{i\bullet}$
1	$y_{11} \cdots y_{1j} \cdots y_{1n_1}$	n_1	$y_{1\bullet}$	$\bar{y}_{1\bullet}$
.	. . .	.	.	.
i	$y_{i1} \cdots y_{ij} \cdots y_{in_i}$	n_i	$y_{i\bullet}$	$\bar{y}_{i\bullet}$
.	. . .	.	.	.
I	$y_{I1} \cdots y_{Ij} \cdots y_{In_I}$	n_I	$y_{I\bullet}$	$\bar{y}_{I\bullet}$
		N	$y_{\bullet\bullet}$	$\bar{y}_{\bullet\bullet}$

$$y_{i\bullet} = \sum_{j=1}^{n_i} y_{ij} \qquad \bar{y}_{i\bullet} = \frac{y_{i\bullet}}{n_i}$$

$$y_{\bullet\bullet} = \sum_i \sum_j y_{ij} = \sum_{i=1}^{I} y_{i\bullet}$$

$$\bar{y}_{\bullet\bullet} = \frac{y_{\bullet\bullet}}{N} = \frac{1}{N} \sum_i \sum_j y_{ij} = \frac{1}{N} \sum_{i=1}^{I} n_i \bar{y}_{i\bullet}$$

Estimación por máxima verosimilitud:

$$\hat{\mu} = \bar{y}_{\bullet\bullet}$$

$$\hat{\tau}_i = \bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_i \sum_j (y_{ij} - \bar{y}_{i\bullet})^2$$

Descomposición de la variabilidad:

Entre tratamientos: $SCTr = \sum_{i=1}^{I} n_i (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2$

Residual: $SCR = \sum_{i=1}^{I}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\bullet})^2$

Total: $SCT = SCTr + SCR = \sum_{i=1}^{I}\sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{\bullet\bullet})^2$

Tabla ANOVA:

F. V.	S. C.	G. L.	C. M.	Fexp
Entre Tratamientos	$SCTr = \sum_{i=1}^{I} n_i (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2$	I − 1	S_{Tr}^2	S_{Tr}^2 / S_R^2
Residual	$SCR = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\bullet})^2$	N − I	S_R^2	
TOTAL	$SCT = \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{\bullet\bullet})^2$	N − 1	S_T^2	

Los Cuadrados Medios (C.M.) se calculan como la Suma de Cuadrados (S.C.) del Factor de Variabilidad (F.V.) correspondiente entre los Grados de Libertad (G.L.)

Forma práctica de la tabla ANOVA:

F. V.	S. C.	G. L.	C. M.	Fexp
Entre Tratamientos	$SCTr = \sum_{i=1}^{I} \frac{y_{i\bullet}^2}{n_i} - \frac{y_{\bullet\bullet}^2}{N}$	$I - 1$	S_{Tr}^2	S_{Tr}^2 / S_R^2
Residual	$SCR = \sum_{i=1}^{I} \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{I} \frac{y_{i\bullet}^2}{n_i}$	$N - I$	S_R^2	
TOTAL	$SCT = \sum_{i=1}^{I} \sum_{j=1}^{n_i} y_{ij}^2 - \frac{y_{\bullet\bullet}^2}{N}$	$N - 1$	S_T^2	

HIPÓTESIS ESTADÍSTICA:

$$\boxed{H_0 : \mu_1 = \mu_2 = \ldots = \mu_I = \mu} \qquad \boxed{H_0 : \tau_1 = \tau_2 = \ldots = \tau_I = 0}$$

Se demuestra que si la hipótesis nula es cierta, entonces: $\dfrac{S_{Tr}^2}{S_R^2} \rightarrow F_{I-1,J-1}$

Región de Rechazo: $\boxed{F_{\exp} > F_{1-\alpha;I-1,J-1}}$

Región de Aceptación: $\boxed{F_{\exp} \leq F_{1-\alpha;I-1,J-1}}$

Coeficiente de Determinación: $\boxed{R^2 = \dfrac{SCTr}{SCT}}$

R^2 es la proporción de variabilidad total presente en los datos explicada por el modelo **(bondad del ajuste).**

Ejemplo.

Una compañía textil utiliza diversos telares para la producción de telas. Aunque se desea que los telares sean homogéneos con el objeto de producir telas de resistencia uniforme, se supone que puede haber una variación significativa en la resistencia de la tela debido a la utilización de distintos telares. Se tienen 5 telares con los que se realizan determinaciones de la resistencia de la tela. Este experimento se realiza en orden aleatorio y los resultados se muestran en la tabla siguiente:

Telares	Resistencia					
1	51	49	50	49	51	50
2	56	60	56	56	57	
3	48	50	53	44	45	
4	47	48	49	44		
5	43	43	46	47	45	46

En este experimento se han considerado 5 tipos de telares y se han realizado 6, 5, 5, 4 y 6 determinaciones de la resistencia de la tela manufacturada con cada uno.

- La variable de **interés** o la **variable respuesta** es **la resistencia de la tela.**
- El **factor** son **los telares.**
- Se utilizan **5 niveles de factor**.
- El modelo es **unifactorial, de efectos fijos, no-equilibrado.**

Telar.	Observaciones						n_i	$y_{i\bullet}$	$\bar{y}_{i\bullet}$	$\sum y_{ij}^2$	$y_{i\bullet}^2/n_i$
1	51	49	50	49	51	50	6	300	50	15004	15000
2	56	60	56	56	57		5	285	57	16257	16245
3	48	50	53	44	45		5	240	48	11574	11520
4	47	48	49	44			4	188	47	8850	8836
5	43	43	46	~~49~~ 47	45	46	6	270	45	12164	12150
							26	$y_{\bullet\bullet} = 1283$		63849	63751

$$SCT = \sum_{i=1}^{5}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{1}{N} y_{\bullet\bullet}^2 = 63849 - \frac{1283^2}{26} = 537.88$$

$$SCTr = \sum_{i=1}^{5} \frac{y_{i\bullet}^2}{n_i} - \frac{y_{\bullet\bullet}^2}{N} = 63751 - \frac{1283^2}{26} = 439.88$$

$$\mathrm{SCR} = \mathrm{SCT} - \mathrm{SCTr} = 98$$

F. V.	S. C.	G. L.	C. M.	Fexp	C.D.
Trat. = Telares	439.88	4	109.97	23.55	0.8178
Residual	98.00	21	4.67		
TOTAL	537.88	25			

Nivel de Significación: $\alpha = 0.05$			
$F_{0.95;\, 4,21} = 2.84$	Fexp = 23.55	Fexp > F-t	Rech. H_0
Nivel de Significación: $\alpha = 0.01$			
$F_{0.99\;\, 4,21} = 4.37$	Fexp = 23.55	Fexp > F-t	Rech. H_0

Conclusión: Rechazamos la hipótesis de igualdad de medias, por lo que concluimos que hay evidencias para afirmar (con un 5% y un 1% de significación) que la resistencia de la tela se ve afectada por el telar utilizado.

df_2	df_1 1	2	3	4	5	6	7	8	10	12	24	60	120	∞	
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.494	2.425	2.235	2.106	2.059	2.010	0.95
16	6.115	4.687	4.077	3.729	3.502	3.341	3.219	3.125	2.986	2.889	2.625	2.447	2.383	2.316	0.975
16	8.531	6.226	5.292	4.773	4.437	4.202	4.026	3.890	3.691	3.553	3.181	2.933	2.845	2.753	0.99
16	16.12	10.97	9.006	7.944	7.272	6.805	6.460	6.195	5.812	5.547	4.846	4.388	4.226	4.059	0.999
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.412	2.342	2.150	2.017	1.968	1.917	0.95
18	5.978	4.560	3.954	3.608	3.382	3.221	3.100	3.005	2.866	2.769	2.503	2.321	2.256	2.187	0.975
18	8.285	6.013	5.092	4.579	4.248	4.015	3.841	3.705	3.508	3.371	2.999	2.749	2.660	2.566	0.99
18	15.38	10.39	8.487	7.460	6.808	6.355	6.021	5.763	5.390	5.132	4.447	3.996	3.836	3.670	0.999
20	4.351	3.493	3.098	$\boxed{2.866}$	2.711	2.599	2.514	2.447	2.348	2.278	2.082	1.946	1.896	1.843	0.95
20	5.871	4.461	3.859	3.515	3.289	3.128	3.007	2.913	2.774	2.676	2.408	2.223	2.156	2.085	0.975
20	8.096	5.849	4.938	4.431	4.103	3.871	3.699	3.564	3.368	3.231	2.859	2.608	2.517	2.421	0.99
20	14.82	9.953	8.098	7.096	6.461	6.019	5.692	5.440	5.075	4.823	4.149	3.703	3.544	3.378	0.999
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.255	2.183	1.984	1.842	1.790	1.733	0.95
24	5.717	4.319	3.721	3.379	3.155	2.995	2.874	2.779	2.640	2.541	2.269	2.080	2.010	1.935	0.975
24	7.823	5.614	4.718	4.218	3.895	3.667	3.496	3.363	3.168	3.032	2.659	2.403	2.310	2.211	0.99
24	14.03	9.340	7.554	6.589	5.977	5.551	5.235	4.991	4.638	4.393	3.735	3.295	3.136	2.969	0.999
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.165	2.092	1.887	1.740	1.683	1.622	0.95
30	5.568	4.182	3.589	3.250	3.026	2.867	2.746	2.651	2.511	2.412	2.136	1.940	1.866	1.787	0.975
30	7.562	5.390	4.510	4.018	3.699	3.473	3.305	3.173	2.979	2.843	2.469	2.208	2.111	2.006	0.99
30	13.29	8.773	7.054	6.125	5.534	5.122	4.817	4.582	4.239	4.001	3.357	2.920	2.760	2.589	0.999
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.077	2.003	1.793	1.637	1.577	1.509	0.95
40	5.424	4.051	3.463	3.126	2.904	2.744	2.624	2.529	2.388	2.288	2.007	1.803	1.724	1.637	0.975
40	7.314	5.178	4.313	3.828	3.514	3.291	3.124	2.993	2.801	2.665	2.288	2.019	1.917	1.805	0.99
40	12.61	8.251	6.595	5.698	5.128	4.731	4.436	4.207	3.874	3.643	3.011	2.574	2.410	2.233	0.999
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.026	1.952	1.737	1.576	1.511	1.438	0.95
50	5.340	3.975	3.390	3.054	2.833	2.674	2.553	2.458	2.317	2.216	1.931	1.721	1.639	1.545	0.975
50	7.171	5.057	4.199	3.720	3.408	3.186	3.020	2.890	2.698	2.563	2.183	1.909	1.803	1.683	0.99
50	12.22	7.956	6.336	5.459	4.901	4.512	4.222	3.998	3.671	3.443	2.817	2.378	2.211	2.026	0.999

Ejemplo.

En una forja se utilizan varios hornos para calentar muestras de metal. Se supone que todos los hornos operan a la misma temperatura, aunque se sospecha que quizás esto probablemente no sea cierto. Se seleccionan **aleatoriamente** tres hornos y se anotan sus temperaturas en sucesivos calentamientos, obteniéndose las observaciones que se muestran en la tabla:

Hornos	Temperaturas					
1	91.5	98.30	98.10	93.50	93.60	
2	88.50	84.65	79.00	77.35		
3	90.10	84.80	88.25	73.00	71.85	78.65

- La variable de **interés** o la **variable respuesta** es **la temperatura** a la que operan los hornos.
- El **factor** son **los hornos.**
- Se utilizan **3 niveles de factor**.
- El modelo es **unifactorial, de efectos aleatorios, no-equilibrado.**

F. V.	S. C.	G. L.	C. M.	Fexp	C.D.
Trat. = Hornos	594.53	2	297.26	8.62	0.5896
Residual	413.81	12	34.48		
TOTAL	1008.34	14			

Nivel de Significación: $\alpha = 0.05$			
$F_{0.95;\,2,12} = 3.89$	Fexp = 8. 62	Fexp > F-t	Rech. H_0

Conclusión: Rechazamos la hipótesis de igualdad de medias, por lo que concluimos que hay evidencias para afirmar (con un 5% de significación) que todos los hornos no están operando a la misma temperatura.

5.2.6. DIAGNOSIS Y VALIDACIÓN DEL MODELO.

Se trata de comprobar si las hipótesis básicas del modelo están o no en contradicción con los datos observados.

HIPÓTESIS DEL MODELO:

- La media es cero: $E[u_{ij}] = 0 \quad \forall\ i, j$
- La varianza es constante: $Var[u_{ij}] = \sigma^2 \quad \forall i, j$
- Son independientes entre sí: $E[u_{ij} u_{rk}] = 0 \;\; i \neq r \;\; \text{o} \;\; j \neq k$
- Siguen una distribución Normal.

¿CÓMO SE VERIFICAN LAS HIPÓTESIS DE LOS RESIDUOS?

1. Independencia de los residuos:
   a. Gráfico de los residuos en función del tiempo.
2. Normalidad de los residuos:
   a. Histograma: Apariencia de una distribución normal centrada en cero.
   b. Gráfico probabilístico normal (Q-Q-Plot)
3. Homocedasticidad (varianza constante):
   a. Residuos frente a los valores ajustados.
   b. Residuos frente a ciertas variables de interés
   c. Contrastes de igualdad de varianzas.

Independencia de los residuos:

- Gráfico de los residuos en función del tiempo:

  Representación de los residuos frente al orden en el que se recopilaron los datos y buscar rachas de residuos de igual signo, así como cualquier tendencia creciente o decreciente en los mismos, lo que sería un indicio de correlación entre los términos de error y el tiempo, lo cual implicaría que la suposición de independencia se ha violado.

- Test de Durbin-Watson.

NORMALIDAD DE LOS RESIDUOS:

- Histograma:

  Los residuos deben tener la apariencia de una distribución normal centrada en cero.

  Discrepancias: valores muy alejados de los demás que suelen corresponder a datos anómalos.

- Gráfico probabilístico normal (Q-Q-Plot):

  Representación de la función de distribución normal en una escala transformada apropiada de forma que la función quede linealizada, de esta manera, si se dispone de datos extraídos aleatoriamente de esta distribución, la representación en esta escala no debe separarse gráficamente de la línea recta teórica.

HOMOCEDASTICIDAD (VARIANZA CONSTANTE):

- Gráfico de los residuos frente a los valores ajustados:

  Si el gráfico tiene forma de embudo, pone de manifiesto un aumento o disminución de los errores en función de los niveles de factor y el modelo no es el adecuado.

- Gráfico de los residuos en función de ciertas variables de interés:

  Igual que el anterior, pero en función de las variables de interés.

- Contrastes de igualdad de varianzas:
  - Para tamaños de muestras iguales:
    - Test de Cochran
    - Test de Hartley
  - Para tamaños de muestras cualesquiera:
    - Test de Barlett

5.2.7. Comparaciones múltiples.

Son técnicas para identificar qué tratamientos son estadísticamente diferentes entre sí (modelo de efectos fijos).

Objetivo fundamental: Comparar entre sí medias de tratamientos o grupos de ellas.

1. Procedimientos gráficos.
2. Procedimientos analíticos:
    a. Método de la diferencia mínima significativa (LSD).
    b. Método de Bonferroni.
    c. Método de Tukey o Método HSD.
    d. Método de rangos múltiples de Duncan.
    e. Método de Newman-Keuls.
    f. Método de Sheffe.
    g. Método de Games-Howell.

TEMA 5.3. DISEÑO EN BLOQUES.

5.3.1. Introducción.
5.3.2. Modelo Estadístico.
5.3.3. Objetivos.
5.3.4. Diseño en Bloques Aleatorizados Completos.
5.3.5. Diseño en Bloques Incompletos Aleatorizados.
- 5.3.5.1. Análisis del modelo para estudiar el efecto de los tratamientos.
- 5.3.5.2. Análisis del modelo para estudiar el efecto de los bloques.

5.3.1. INTRODUCCIÓN.

En el diseño en bloques las unidades experimentales son heterogéneas entre sí, por eso se agrupan en bloques aquellas unidades más homogéneas, para de esta forma disminuir la variabilidad y en consecuencia el error experimental.

Ejemplo:

Si en la industria algodonera las parcelas de terreno son de la misma calidad y superficie

Homogeneidad entre las unidades experimentales

Diseño Completamente Aleatorizado

Los tratamientos se aplican al azar a las unidades experimentales

Si en la industria algodonera las parcelas de terreno **NO** son de la misma calidad y superficie

Se agrupan las parcelas en **bloques homogéneos** con la misma calidad y superficie y **heterogéneos** entre sí.

Diseño En Bloques Aleatorizado

Los tratamientos se aplican al azar a las unidades experimentales en cada bloque.

1. Se realiza una observación por tratamiento en cada bloque: **N = IJ** observaciones (bloques completos).
2. La asignación de los tratamientos a las unidades experimentales en cada bloque se determina **aleatoriamente**.
3. Los tratamientos y los bloques son **factores de efectos fijos**.
4. **No hay interacción** entre los tratamientos y los bloques: El efecto de un factor no depende del nivel de otro factor.

	Bloques						
		1	2	…	j	…	J
Tratamientos	1	y_{11}	y_{12}	…	y_{1j}	…	y_{1J}
	2	y_{21}	y_{22}	…	y_{2j}	…	y_{2J}
	…	…	…	…	…	…	…
	i	y_{i1}	y_{i2}	…	y_{ij}	…	y_{iJ}
	…	…	…	…	…	…	…
	I	y_{I1}	y_{I2}	…	y_{Ij}	…	y_{IJ}

5.3.2. Modelo estadístico.

$$\boxed{y_{ij} = \mu + \tau_i + \beta_j + u_{ij}; \quad i:1,\ldots,I; \;\; j:1,\ldots,J}$$

y_{ij} : Variable aleatoria que representa la observación del tratamiento i-ésimo en el bloque j-ésimo.

μ : Efecto constante, común a todos los niveles. Media global.

τ_i : Efecto del tratamiento i-ésimo.

β_j : Efecto del bloque j-ésimo.

u_{ij} : Errores experimentales.

Los residuos deben satisfacer las siguientes propiedades:

- La media es cero: $E[u_{ij}] = 0 \quad \forall\, i, j$
- La varianza es constante: $Var[u_{ij}] = \sigma^2 \quad \forall i, j$
- Son independientes entre sí: $E[u_{ij} u_{rk}] = 0 \;\; i \neq r \;\; \text{o} \;\; j \neq k$
- Siguen una distribución Normal.

Dos Factores:

1) **Factor Tratamiento** $\Rightarrow$ Factor Principal.

2) **Factor Bloque** $\Rightarrow$ Factor Secundario (Se introduce únicamente para eliminar su influencia en la variable respuesta)

5.3.3. OBJETIVOS.

1. Estimar los parámetros del modelo.
2. Contrastar la hipótesis de igualdad de los tratamientos.

$$\boxed{\begin{array}{l} H_0 : \tau_i = 0 \quad \forall i \ \text{ (Todos los tratamientos producen el mismo efecto)} \\ H_1 : \tau_i \neq 0 \text{ para algún } i \ \text{ (Al menos dos tratamientos difieren entre sí)} \end{array}}$$

3. Contrastar la hipótesis de igualdad de los bloques.

$$\boxed{\begin{array}{l} H_0 : \beta_j = 0 \ \forall j \ \text{ (Todos los bloques producen el mismo efecto)} \\ H_1 : \beta_j \neq 0 \text{ para algún } j \ \text{ (Al menos dos bloques difieren entre sí)} \end{array}}$$

4. Comprobar la idoneidad del modelo.
5. Realizar comparaciones múltiples.

5.3.4. Diseño en bloques aleatorizados completos.

Análisis del modelo.

Estimación por máxima verosimilitud:

$$\hat{\mu} = \bar{y}_{\bullet\bullet}$$

$$\hat{\tau}_i = \bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}$$

$$\hat{\beta}_j = \bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet}$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_i \sum_j (y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j)^2$$

Descomposición de la variabilidad:

Entre tratamientos: $$SCTr = J\sum_{i=1}^{I}(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2$$

Entre bloques: $$SCBl = I\sum_{j=1}^{J}(\bar{y}_{\bullet j} - \bar{y}_{\bullet\bullet})^2$$

Residual: $$SCR = \sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij} - \bar{y}_{i\bullet} - \bar{y}_{\bullet j} + \bar{y}_{\bullet\bullet})^2$$

Total: $$SCT = SCTr + SCBl + SCR = \sum_{i=1}^{I}\sum_{j=1}^{J}(y_{ij} - \bar{y}_{\bullet\bullet})^2$$

Forma práctica de la tabla anova:

F. V.	S. C.	G. L.	C. M.	Fexp
Tratam.	$SCTr = \frac{1}{J}\sum_{i=1}^{I} y_{i\bullet}^2 - \frac{y_{\bullet\bullet}^2}{IJ}$	I – 1	S_{Tr}^2	S_{Tr}^2 / S_R^2
Bloques	$SCBl = \frac{1}{I}\sum_{j=1}^{J} y_{\bullet j}^2 - \frac{y_{\bullet\bullet}^2}{IJ}$	J – 1	S_{Bl}^2	S_{Bl}^2 / S_R^2
Residual	$SCR = SCT - SCTr - SCBl$	(I – 1) (J – 1)	S_R^2	
TOTAL	$SCT = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}^2 - \frac{y_{\bullet\bullet}^2}{IJ}$	N – 1	S_T^2	

Hipótesis estadística 1:

$$\boxed{H_0 : \mu_1 = \mu_2 = \ldots = \mu_I = \mu} \qquad \boxed{H_0 : \tau_1 = \tau_2 = \ldots = \tau_I = 0}$$

Si la hipótesis nula es cierta, entonces: $\dfrac{S_{Tr}^2}{S_R^2} \to F_{(I-1),(I-1)(J-1)}$

Región de Rechazo: $\boxed{F_{\tau(\exp)} > F_{1-\alpha;(I-1),(I-1)(J-1)}}$

Región de Aceptación: $\boxed{F_{\tau(\exp)} \leq F_{1-\alpha;(I-1),(I-1)(J-1)}}$

Coeficiente de Determinación de los Tratamientos: $\boxed{R_{Tr}^2 = \dfrac{SCTr}{SCT}}$

R^2_{Tr} es la proporción de variabilidad total en la variable respuesta explicada por los tratamientos.

HIPÓTESIS ESTADÍSTICA 2:

$$\boxed{H_0 : \mu_1 = \mu_2 = \ldots = \mu_J = \mu} \qquad \boxed{H_0 : \beta_1 = \beta_2 = \ldots = \beta_J = 0}$$

Si la hipótesis nula es cierta, entonces: $\dfrac{S^2_{Bl}}{S^2_R} \rightarrow F_{(J-1),(I-1)(J-1)}$

Región de Rechazo: $\boxed{F_{\beta(\exp)} > F_{1-\alpha;(J-1),(I-1)(J-1)}}$

Región de Aceptación: $\boxed{F_{\beta(\exp)} \leq F_{1-\alpha;(J-1),(I-1)(J-1)}}$

Coeficiente de Determinación de los Bloques: $\boxed{R^2_{Bl} = \dfrac{SCBl}{SCT}}$

R^2_{Bl} es la proporción de variabilidad total de la variable respuesta explicada por los bloques.

Coeficiente de Determinación del Modelo:

$$\boxed{R^2 = \frac{SCTr + SCBl}{SCT}}$$

R^2 es la proporción de variabilidad total de la variable respuesta explicada por el modelo.

Ejemplo.

Una industria algodonera, interesada en maximizar el rendimiento de la semilla de algodón, quiere comprobar si dicho rendimiento depende del tipo de fertilizante utilizado para tratar la planta. A su disposición tiene 5 tipos de fertilizantes. Como puede haber diferencia entre las parcelas, el experimentador decide realizar un diseño en bloques aleatorizados. Para ello, divide el terreno en 4 bloques y cada bloque en 5 parcelas, fumigando dentro de cada bloque cada una de las parcelas con un fertilizante. Al recoger la cosecha se mide el rendimiento de la semilla, obteniéndose las siguientes observaciones:

	Bloques			
Fertilizantes	A	B	C	D
1	87	86	88	83
2	85	87	95	85
3	90	92	95	90
4	89	97	98	88
5	99	96	91	90

	Bloques						
Fertilizantes	A	B	C	D	$y_{i\bullet}$	$y_{i\bullet}^2$	$\sum y_{ij}^2$
1	87	86	88	83	344	118336	29598
2	85	87	95	85	352	123904	31044
3	90	92	95	90	367	134689	33689
4	89	97	98	88	372	138384	34678
5	99	96	91	90	376	141376	35398
$y_{\bullet j}$	450	458	467	436	1811	656689	164407
$y_{\bullet j}^2$	202500	209764	218089	190096	820449		

$$SCT = \sum_{i=1}^{5}\sum_{j=1}^{4} y_{ij}^2 - \frac{1}{IJ} y_{\bullet\bullet}^2 = 164407 - \frac{1811^2}{20} = 420.95$$

$$SCTr = \frac{1}{J}\sum_{i=1}^{5} y_{i\bullet}^2 - \frac{y_{\bullet\bullet}^2}{IJ} = \frac{656689}{4} - \frac{1811^2}{20} = 186.20$$

$$SCBl = \frac{1}{I}\sum_{i=1}^{4} y_{\bullet j}^2 - \frac{y_{\bullet\bullet}^2}{IJ} = \frac{820449}{5} - \frac{1811^2}{20} = 103.75$$

$$SCR = SCT - SCTr - SCBl = 131$$

F. V.	S. C.	G. L.	C. M.	Fexp
Entre tratamientos	186.20	4	46.5500	4.264
Entre bloques	103.75	3	34.5833	3.168
Residual	131.00	12	10.9166	
TOTAL	420.95	19		

$$R_\tau^2 = \frac{SCT_r}{SCT} = \frac{186.2}{420.95} = 0.4423\ ; \quad R_\beta^2 = \frac{SCBl}{SCT} = \frac{103.75}{420.95} = 0.2464$$

$$R^2 = R_\tau^2 + R_\beta^2 = 0.6887$$

CONTRASTE PARA LOS TRATAMIENTOS:

Nivel de Significación: $\alpha = 0.05$			
$F_{0.95;\,4,12} = 3.26$	$Fexp = 4.264$	$Fexp > F\text{-}t$	Rech. H_0
Nivel de Significación: $\alpha = 0.01$			
$F_{0.01;\,4,12} = 5.41$	$Fexp = 4.264$	$Fexp < F\text{-}t$	Acep. H_0
Nivel Mínimo de Significación		0.0233	

Conclusión:

Para un 5% de significación, rechazamos la hipótesis de igualdad de medias, por lo que concluimos que hay evidencias para afirmar que el rendimiento de la semilla de algodón se ve afectado por el fertilizante utilizado.

Sin embargo, para un 1% de significación, aceptamos la hipótesis nula, por lo que no habría evidencias para afirmar tal hipótesis a este nivel de significación.

CONTRASTE PARA LOS BLOQUES:

Nivel de Significación: $\alpha = 0.05$			
$F_{0.95;\,3,12} = 3.49$	Fexp = 3.168	Fexp < F-t	Acep. H_0
Nivel de Significación: $\alpha = 0.01$			
$F_{0.99;\,3,12} = 5.95$	Fexp = 3.168	Fexp < F-t	Acep. H_0
Nivel Mínimo de Significación		0.0683	

Conclusión:
Tanto para un 1% como para un 5% de significación, aceptamos la hipótesis de igualdad de bloques, por lo que no hay evidencias para decir que el tipo de terreno afecte en el rendimiento de la semilla de algodón.

PRESCINDIENDO DE LOS BLOQUES:

F. V.	S. C.	G. L.	C. M.	Fexp
Entre tratamientos	186.20	4	46.55	2.974
Residual	234.75	15	15.65	
TOTAL	420.95	19		

Nivel de Significación: $\alpha = 0.05$			
$F_{0.05;\,4,15} = 3.05$	Fexp = 2.974	Fexp < F-t	Acep. H_0
Nivel de Significación: $\alpha = 0.01$			
$F_{0.01;\,4,15} = 4.89$	Fexp = 2.974	Fexp < F-t	Acep. H_0
Nivel Mínimo de Significación		0.0561	

Conclusión: Tanto para un 1% como para un 5% de significación, aceptamos la hipótesis de igualdad de tratamientos, por lo que no hay evidencias para decir que el tipo de fertilizante afecte en el rendimiento de la semilla de algodón.

Es mejor el diseño en bloques porque tiene menor nivel mínimo de significación (menor p-valor) y menor varianza residual.

La diagnosis y validación del modelo y la comparación de medias múltiples se realiza como en el diseño completamente aleatorizado.

5.3.5. Diseño en bloques incompletos aleatorizados.

El caso que vamos a estudiar es el del Diseño en Bloques Incompletos **Balanceados:**

1. Cada tratamiento ocurre el mismo número de veces en el diseño.
2. Cada par de tratamientos ocurren juntos el mismo número de veces que cualquier otro par.

PARÁMETROS:

I: Número de tratamientos o niveles del factor principal.
J: Número de bloques.
K: Número de tratamientos por bloque.
R: Número de veces que un tratamiento se presenta en el diseño, es decir, el número de réplicas de un tratamiento dado.
λ: Número de bloques en los que un par de tratamientos ocurren juntos.
N: Número total de observaciones.

Estos parámetros deben verificar las siguientes relaciones:

i) $N = IR = JK$

ii) $\lambda = R\dfrac{K-1}{I-1}$

iii) $J \geq 1$

Ejemplo:

	Bloques				
Tratat.	A	B	C	D	E
1	X	X	X	X	
2	X	X	X		X
3	X	X		X	X
4	X		X	X	X
5		X	X	X	X

I = 5 = J, K = 4 = R, $\lambda = R\dfrac{K-1}{I-1} = 4\dfrac{3}{4} = 3$

5.3.5.1. ANÁLISIS DEL MODELO PARA ESTUDIAR EL EFECTO DE LOS TRATAMIENTOS.

DESCOMPOSICIÓN DE LA VARIABILIDAD:

Entre tratamientos: $SCTr^* = \dfrac{K\sum_{i=1}^{I} T_i^2}{\lambda I}, \qquad T_i = y_{i\bullet} - \dfrac{1}{K}\sum_{j=1}^{J} n_{ij} y_{\bullet j}, \quad i = 1, 2, \ldots, I$

$$n_{ij} = \begin{cases} 1 & \text{si el tratamiento i ocurre en el bloque j} \\ 0 & \text{en otro caso} \end{cases} \qquad \sum_{i=1}^{I} T_i = 0$$

Entre bloques: $SCBl = \dfrac{1}{k}\sum_{j=1}^{J} y_{\bullet j}^2 - \dfrac{y_{\bullet\bullet}^2}{N}$

Total: $SCT = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}^2 - \dfrac{y_{\bullet\bullet}^2}{N}$

Residual: $SCR = SCT - SCTr^* - SCBl$

Forma práctica de la tabla ANOVA:

F. V.	S. C.	G. L.	C. M.	Fexp
Trat-Ajustados	$\dfrac{K\sum_{i=1}^{I} T_i^2}{\lambda I}$	I – 1	$CMTr^*$	$CMTr^* / CMR$
Bloques	$\dfrac{1}{K}\sum_{j=1}^{J} y_{\bullet j}^2 - \dfrac{y_{\bullet\bullet}^2}{N}$	J – 1		
Residual	$SCT - SCTr^* - SCBl$	N – I – J + 1	CMR	
TOTAL	$\sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}^2 - \dfrac{y_{\bullet\bullet}^2}{N}$			

5.3.5.2. ANÁLISIS DEL MODELO PARA ESTUDIAR EL EFECTO DE LOS BLOQUES.

DESCOMPOSICIÓN DE LA VARIABILIDAD:

Entre bloques: $SCBl^{*} = \dfrac{R\sum_{j=1}^{J} B_j^2}{\lambda I}, \qquad B_j = y_{\bullet j} - \dfrac{1}{R}\sum_{i=1}^{I} n_{ij} y_{i\bullet}, \quad j = 1, 2, \ldots, J$

$$n_{ij} = \begin{cases} 1 & \text{si el tratamiento i ocurre en el bloque j} \\ 0 & \text{en otro caso} \end{cases} \qquad \sum_{j=1}^{J} B_j = 0$$

Entre tratamientos: $SCTr = \dfrac{1}{R}\sum_{i=1}^{I} y_{i\bullet}^2 - \dfrac{y_{\bullet\bullet}^2}{N}$

Total: $SCT = \sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}^2 - \dfrac{y_{\bullet\bullet}^2}{N}$

Residual: $SCR = SCT - SCTr - SCBl^{*}$

Forma práctica de la tabla ANOVA:

F. V.	S. C.	G. L.	C. M.	Fexp
Trat-No-Ajustad.	$\frac{1}{R}\sum_{i=1}^{I} y_{i\bullet}^2 - \frac{y_{\bullet\bullet}^2}{N}$	I – 1		
Bloques-Ajustados	$\frac{R\sum_{j=1}^{J} B_j^2}{\lambda J}$	J – 1	$CMBl^*$	$CMBl^* / CMR$
Residual	$SCT - SCTr - SCBl^*$	N – I – J + 1	CMR	
TOTAL	$\sum_{i=1}^{I}\sum_{j=1}^{J} y_{ij}^2 - \frac{y_{\bullet\bullet}^2}{N}$			

Ejemplo.

Se considera de nuevo el caso de la industria algodonera. Supongamos que debido a la extensión de las parcelas de terreno y a la falta de recursos, no se pueden aplicar los fertilizantes en cada bloque, si no que sólo se pueden aplicar 4 de los 5 fertilizantes en cada bloque. Se decide utilizar un diseño en bloques incompletos balanceado.

	Bloques				
Fertilizantes	B_1	B_2	B_3	B_4	B_5
1	94	96	100	92	
2	95	75	76		92
3	76	100		97	98
4	94		102	93	96
5		75	91	86	95

I = 5 = J, K = 4 = R, $\lambda = R\dfrac{K-1}{I-1} = 4\dfrac{3}{4} = 3$

EFECTO DE LOS TRATAMIENTOS:

	Bloques					
Fertilizantes	B_1	B_2	B_3	B_4	B_5	$y_{i\bullet}$
1	94	96	100	92		382
2	95	75	76		92	338
3	76	100		97	98	371
4	94		102	93	96	385
5		75	91	86	95	347
$y_{\bullet j}$	359	346	369	368	381	

$$T_i = y_{i\bullet} - \frac{1}{K}\sum_{j=1}^{5} n_{ij} y_{\bullet j}, \quad i = 1, 2, \ldots, 5$$

$$T_1 = 382 - \frac{1}{4}(359+346+369+368) = 21.5$$

$$T_2 = 338 - \frac{1}{4}(359+346+369+381) = -25.75$$

$$T_3 = 371 - \frac{1}{4}(359+346+368+381) = 7.5$$

$$T_4 = 385 - \frac{1}{4}(359+369+368+381) = 15.75$$

$$T_5 = 347 - \frac{1}{4}(346+369+368+381) = -19$$

$$\sum_{i=1}^{5} T_i = 0$$

	Bloques							
Fertiliz.	B_1	B_2	B_3	B_4	B_5	$y_{i\cdot}$	$y_{i\cdot}^2$	$\sum y_{ij}^2$
1	94	96	100	92		382	145924	36516
2	95	75	76		92	338	114244	28890
3	76	100		97	98	371	137641	34789
4	94		102	93	96	385	148225	37105
5		75	91	86	95	347	120409	30327
$y_{\bullet j}$	359	346	369	368	381	1823	666443	167627
$y_{\bullet j}^2$	128881	119716	136161	135424	145161	665343		

$$SCT = \sum_{i=1}^{5}\sum_{j=1}^{5} y_{ij}^2 - \frac{1}{N} y_{\bullet\bullet}^2 = 167627 - \frac{1823^2}{20} = 1460.55$$

$$SCTr^* = \frac{K\sum_{i=1}^{5} T_i^2}{\lambda I} = \frac{4}{3\times 5} 1790.625 = 477.5$$

$$SCBl = \frac{1}{K}\sum_{j=1}^{5} y_{\bullet j}^2 - \frac{y_{\bullet\bullet}^2}{N} = \frac{665343}{4} - \frac{1823^2}{20} = 169.3$$

$$SCR = SCT - SCTr^* - SCBl = 813.75$$

F. V.	S. C.	G. L.	C. M.	Fexp
Tratamientos-Ajustados	477.5	4	119.375	1.614
Bloques-No-Ajustaos	169.3	4		
Residual	813.75	11	73.97	
TOTAL	1460.55	19		

Nivel de Significación: $\alpha = 0.05$			
$F_{0.95;\, 4,11} = 3.36$	Fexp = 1.614	Fexp < F-t	Acep. H_0
Nivel de Significación: $\alpha = 0.01$			
$F_{0.99\; 4,11} = 5.67$	Fexp = 1.614	Fexp < F-t	Acep. H_0
Nivel Mínimo de Significación		0.243	

Conclusión: Tanto para un 1% como para un 5% de significación, aceptamos la hipótesis de igualdad de tratamientos, por lo que no hay evidencias para decir que el tipo de fertilizante afecte en el rendimiento de la semilla de algodón.

EFECTO DE LOS BLOQUES:

	Bloques					
Fertilizantes	B_1	B_2	B_3	B_4	B_5	$y_{i\bullet}$
1	94	96	100	92		382
2	95	75	76		92	338
3	76	100		97	98	371
4	94		102	93	96	385
5		75	91	86	95	347
$y_{\bullet j}$	359	346	369	368	381	

$$B_j = y_{\bullet j} - \frac{1}{R}\sum_{i=1}^{5} n_{ij} y_{i\bullet}, \quad j = 1,2,\ldots.5$$

$$B_1 = 359 - \tfrac{1}{4}(382+338+371+385) = -10$$

$$B_2 = 346 - \tfrac{1}{4}(382+338+371+347) = -13.5$$

$$B_3 = 369 - \tfrac{1}{4}(382+338+385+347) = 6$$

$$B_4 = 368 - \tfrac{1}{4}(382+371+385+347) = -3.25$$

$$B_5 = 381 - \tfrac{1}{4}(338+371+385+347) = 20.75$$

$$\sum_{j}^{5} B_j = 0$$

$$SCBl^{*} = \frac{R\sum_{j=1}^{J} B_j^2}{\lambda J} = \frac{4}{3\times 5} 759.375 = 202.5$$

$$SCTr = \frac{1}{R}\sum_{i=1}^{5} y_{i.}^2 - \frac{y_{..}^2}{N} = \frac{666443}{4} - \frac{1823^2}{20} = 444.3$$

$SCT = 1460.55$ $\qquad$ $SCR = 813.75$

F. V.	S. C.	G. L.	C. M.	Fexp
Tratam-No-Ajustados	444.3	4		
Bloques-Ajustaos	202.5	4	50.625	0.684
Residual	813.75	11	73.97	
TOTAL	1460.55	19		

Nivel de Significación: $\alpha = 0.05$			
$F_{0\,95;\,4,11} = 3.36$	Fexp = 0.684	Fexp < F-t	Acep. H_0
Nivel de Significación: $\alpha = 0.01$			
$F_{0.99\;4,11} = 5.67$	Fexp = 0.684	Fexp < F-t	Acep. H_0
Nivel Mínimo de Significación		0.626	

Conclusión: Tanto para un 1% como para un 5% de significación, aceptamos la hipótesis de igualdad de bloques, por lo que no hay evidencias para decir que el tipo de parcela afecte en el rendimiento de la semilla de algodón.

PRESCINDIENDO DE LOS BLOQUES:

F. V.	S. C.	G. L.	C. M.	Fexp
Tratamientos	444.3	4	111.075	
Residual	202.5 + 813.75 = 1016.25	15	67.75	1.639
TOTAL	1460.55	19		

Nivel de Significación: $\alpha = 0.05$			
$F_{0\,95;\,4,15} = 3.06$	Fexp = 1.639	Fexp < F-t	Acep. H_0
Nivel de Significación: $\alpha = 0.01$			
$F_{0\,99;\,4,15} = 4.89$	Fexp = 1.639	Fexp < F-t	Acep. H_0
Nivel Mínimo de Significación	0.217		

Conclusión: Tanto para un 1% como para un 5% de significación, aceptamos la hipótesis de igualdad de tratamientos, por lo que no hay evidencias para decir que el tipo de fertilizante afecte en el rendimiento de la semilla de algodón.

Es mejor el diseño sin bloques porque tiene menor nivel mínimo de significación (menor p-valor) y menor varianza residual.

COEFICIENTES DE DETERMINACIÓN:

- MODELO EN BLOQUES

$$R^2 = \frac{SCTr^* + SCBl}{SCT} = \frac{477.5+169.3}{1460.55} = 0.442$$

- MODELO SIN BLOQUES

$$R^2 = \frac{SCTr}{SCT} = \frac{444.3}{1460.55} = 0.304$$

TEMA 5.4. DISEÑO EN CUADRADOS.

5.4.1. Diseño en Cuadrados Latinos.

5.4.2. Diseño en Cuadrados Greco-latinos.

5.4.3. Diseño en Cuadrados de Youden.

5.4.1. Diseño en cuadrados latinos.

- Se controlan 3 fuentes de variabilidad: 1 factor principal y 2 factores de bloque.
- Cada uno de los factores tiene el mismo número de niveles, K.
- Cada nivel del factor principal aparece una vez en cada fila y una vez en cada columna.
- No hay interacción entre los factores.
- El número de observaciones es K^2.

A	B	C
B	C	A
C	A	B

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

A	B	C	D	E
D	E	A	B	C
B	C	D	E	A
E	A	B	C	D
C	D	E	A	B

1. Se elige aleatoriamente un cuadrado latino de los disponibles.
2. Se asigna aleatoriamente el orden de las filas y las columnas.
3. Se asignan aleatoriamente los tres factores a las filas, columnas y letras.

Ejemplo.

Se quiere estudiar el rendimiento de la semilla de trigo. Al plantear este experimento se pensó que podría conseguirse mayor precisión si se controlaba la variabilidad introducida por los tipos de abono e insecticida. El instituto de experimentación agrícola está interesado en estudiar 4 tipos de semillas de trigo (S1, S2, S3, S4) y decide realizar el experimento utilizando un diseño en cuadrado latino. Para ello selecciona 4 niveles para cada una de las variables bloque: Abono (A1, A2, A3, A4) e insecticida (I1, I2, I3, I4).

1. Se elige un cuadrado latino al azar:

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

2. Se asigna al azar el orden de las filas: 2,3,1,4

B	A	D	C
C	D	A	B
A	B	C	D
D	C	B	A

1. Se elige un cuadrado latino al azar:

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

2. Se asigna al azar el orden de las filas: 2,3,1,4

B	A	D	C
C	D	A	B
A	B	C	D
D	C	B	A

3. Se asigna al azar el orden de las columnas: 4,3,1,2

C	D	B	A
B	A	C	D
D	C	A	B
A	B	D	C

4. Se asignan al azar las filas, las columnas y las letras latinas a los tres factores:
Filas → Insecticidas
Columnas → Semillas
Letras Latinas → Abonos

	Semillas			
Insecticidas.	**S1**	**S2**	**S3**	**S4**
I1	A3	A4	A2	A1
I2	A2	A1	A3	A4
I3	A4	A3	A1	A2
I4	A1	A2	A4	A3

5. Por convenio se sitúa el factor principal en las celdillas. Reordenando el diseño anterior se obtiene la siguiente tabla:

	Abonos			
Insecticidas.	**A1**	**A2**	**A3**	**A4**
I1	S4	S3	S1	S2
I2	S2	S1	S3	S4
I3	S3	S4	S2	S1
I4	S1	S2	S4	S3

Modelo estadístico.

$$\boxed{y_{ij(h)} = \mu + \tau_i + \beta_j + \gamma_h + u_{ij(h)}; \quad i:1,\ldots,K; \;\; j:1,\ldots,K; \;\; h:1,\ldots,K}$$

$y_{ij(h)}$: Observación de la i-ésima fila, j-ésima columna y h-ésima letra latina.

μ : Efecto constante, común a todos los niveles. Media global.

τ_i : Efecto del i-ésimo nivel del factor fila.

β_j : Efecto del j-ésimo nivel del factor columna.

γ_h : Efecto del h-ésimo nivel del factor letra latina.

$u_{ij(h)}$: Errores experimentales, v.a. independendientes y normales.

ANÁLISIS DEL MODELO.

	Abonos			
Insecticidas.	**A1**	**A2**	**A3**	**A4**
I1	S4	S3	S1	S2
I2	S2	S1	S3	S4
I3	S3	S4	S2	S1
I4	S1	S2	S4	S3

↓

$y_{11(4)}$	$y_{12(3)}$	$y_{11(4)}$	$y_{14(2)}$
$y_{21(2)}$	$y_{22(1)}$	$y_{23(3)}$	$y_{24(4)}$
$y_{31(3)}$	$y_{32(4)}$	$y_{33(2)}$	$y_{34(1)}$
$y_{41(1)}$	$y_{42(2)}$	$y_{43(4)}$	$y_{44(3)}$

- $N = K^2$ es el número total de observaciones.

- El total y la media general: $y_{\bullet\bullet\bullet} = \sum_{i=1}^{K}\sum_{j=1}^{K} y_{ij(\bullet)} \quad \bar{y}_{\bullet\bullet\bullet} = \frac{y_{\bullet\bullet\bullet}}{K^2}$

- El total y la media por fila: $y_{i\bullet\bullet} = \sum_{j=1}^{K} y_{ij(\bullet)} \quad \bar{y}_{i\bullet\bullet} = \frac{y_{i\bullet\bullet}}{K}$

- El total y la media por columna: $y_{\bullet j\bullet} = \sum_{i=1}^{K} y_{ij(\bullet)} \quad \bar{y}_{\bullet j\bullet} = \frac{y_{\bullet j\bullet}}{K}$

- El total y la media por letra: $y_{\bullet\bullet h} = \sum_{i=1}^{K}\sum_{j=1}^{K} y_{ij(h)} \quad \bar{y}_{\bullet\bullet h} = \frac{y_{\bullet\bullet h}}{K}$

Estimación por máxima verosimilitud:

$$\hat{\mu} = \bar{y}_{\bullet\bullet\bullet}$$

$$\hat{\tau}_i = \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}$$

$$\hat{\beta}_j = \bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}$$

$$\hat{\gamma}_h = \bar{y}_{\bullet\bullet h} - \bar{y}_{\bullet\bullet\bullet}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_i \sum_j (y_{ij(h)} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j - \hat{\gamma}_h)^2$$

Descomposición de la variabilidad:

Entre filas: $$SCF = K\sum_{i=1}^{K}(\overline{y}_{i\bullet\bullet} - \overline{y}_{\bullet\bullet\bullet})^2$$

Entre columnas: $$SCC = K\sum_{j=1}^{K}(\overline{y}_{\bullet j\bullet} - \overline{y}_{\bullet\bullet\bullet})^2$$

Entre letras: $$SCL = K\sum_{h=1}^{K}(\overline{y}_{\bullet\bullet h} - \overline{y}_{\bullet\bullet\bullet})^2$$

Residual: $$SCR = \sum_{i=1}^{K}\sum_{j=1}^{K}(y_{ij(h)} - \overline{y}_{i\bullet\bullet} - \overline{y}_{\bullet j\bullet} - \overline{y}_{\bullet\bullet h} + 2\overline{y}_{\bullet\bullet\bullet})^2 = \sum_{i=1}^{K}\sum_{j=1}^{K}u_{ij(h)}^2$$

Total: $$SCT = SCF + SCC + SCL + SCR = \sum_{i=1}^{K}\sum_{j=1}^{K}(y_{ij(h)} - \overline{y}_{\bullet\bullet\bullet})^2$$

FORMA PRÁCTICA DE LA TABLA ANOVA:

F. V.	S. C.	G. L.	C. M.	Fexp
Filas	$SCF = \frac{1}{K}\sum_{i=1}^{K} y_{i\bullet\bullet}^2 - \frac{y_{\bullet\bullet\bullet}^2}{N}$	K – 1	S_F^2	S_F^2 / S_R^2
Columnas	$SCC = \frac{1}{K}\sum_{j=1}^{K} y_{\bullet j\bullet}^2 - \frac{y_{\bullet\bullet\bullet}^2}{N}$	K – 1	S_C^2	S_C^2 / S_R^2
Letras	$SCL = \frac{1}{K}\sum_{h=1}^{K} y_{\bullet\bullet h}^2 - \frac{y_{\bullet\bullet\bullet}^2}{N}$	K – 1	S_L^2	S_L^2 / S_R^2
Residual	$SCR = SCT - SCF - SCC - SCL$	(K – 1) (K – 2)	S_R^2	
TOTAL	$SCT = \sum_{i=1}^{K}\sum_{j=1}^{K} y_{ij(h)}^2 - \frac{y_{\bullet\bullet\bullet}^2}{N}$	N – 1		

Hipótesis estadísticas:

$$\boxed{H_0 : \tau_1 = \tau_2 = \ldots = \tau_K = 0} \longrightarrow \boxed{\frac{S_F^2}{S_R^2} \rightarrow F_{(K-1),(K-1)(K-2)}}$$

$$\boxed{H_0 : \beta_1 = \beta_2 = \ldots = \beta_K = 0} \longrightarrow \boxed{\frac{S_C^2}{S_R^2} \rightarrow F_{(K-1),(K-1)(K-2)}}$$

$$\boxed{H_0 : \gamma_1 = \gamma_2 = \ldots = \gamma_K = 0} \longrightarrow \boxed{\frac{S_L^2}{S_R^2} \rightarrow F_{(K-1),(K-1)(K-2)}}$$

Coeficientes de Determinación:

$$\boxed{R^2 = \frac{SCF + SCC + SCL}{SCT}}$$

$$\boxed{R_F^2 = \frac{SCF}{SCT}} \qquad \boxed{R_C^2 = \frac{SCC}{SCT}} \qquad \boxed{R_L^2 = \frac{SCL}{SCT}}$$

Ejemplo.

Se quiere estudiar el rendimiento de la semilla de trigo. Al plantear este experimento se pensó que podría conseguirse mayor precisión si se controlaba la variabilidad introducida por los tipos de abono e insecticida. El instituto de experimentación agrícola está interesado en estudiar 4 tipos de semilla de trigo (A, B, C, D) y decide realizar el experimento utilizando un diseño en cuadrado latino. Para ello selecciona 4 niveles para cada una de las variables de bloque: abono (A.1, A.2, A.3, A.4) e insecticida (I.1, I.2, I.3, I.4).

Abonos	Insecticidas			
	I.1	I.2	I.3	I.4
A.1	C 7	D 8	B 4	A 3
A.2	B 15	A 16	C 18	D 23
A.3	D 18	C 12	A 12	B 10
A.4	A 14	B 13	D 16	C 14

Abonos	Insecticidas					
	I.1	I.2	I.3	I.4	$y_{i..}$	$y_{i..}^2$
A.1	C 7	D 8	B 4	A 3	22	484
A.2	B 15	A 16	C 18	D 23	72	5184
A.3	D 18	C 12	A 12	B 10	52	2704
A.4	A 14	B 13	D 16	C 14	57	3249
$y_{.j.}$	54	49	50	50	203	11621
$y_{.j.}^2$	2916	2401	2500	2500	10317	
$\sum y_{ij(.)}^2$	794	633	740	834	3001	

Abonos	Insecticidas			
	I.1	I.2	I.3	I.4
A.1	C 7	D 8	B 4	A 3
A.2	B 15	A 16	C 18	D 23
A.3	D 18	C 12	A 12	B 10
A.4	A 14	B 13	D 16	C 14

Letra Latina	Observaciones				$y_{..h}$	$y^2_{..h}$
A	3	16	12	14	45	2025
B	4	15	10	13	42	1764
C	7	18	12	14	51	2601
D	8	23	18	16	65	4225
					203	10615

$$SCT=\sum_i\sum_j y_{ij(\cdot)}^2-\frac{1}{K^2}y_{\cdots}^2=3001-\frac{1}{4^2}\times 203^2=425.4375$$

$$SCF=\frac{1}{K}\sum_i y_{i\cdot\cdot}^2-\frac{1}{K^2}y_{\cdots}^2=\frac{1}{4}11621-\frac{1}{4^2}\times 203^2=329.6875$$

$$SCC=\frac{1}{K}\sum_j y_{\cdot j\cdot}^2-\frac{1}{K^2}y_{\cdots}^2=\frac{1}{4}10317-\frac{1}{4^2}\times 203^2=3.6875$$

$$SCL=\frac{1}{K}\sum_h y_{\cdot\cdot h}^2-\frac{1}{K^2}y_{\cdots}^2=\frac{1}{4}10615-\frac{1}{4^2}\times 203^2=78.1875$$

$$\mathrm{SCR}=\mathrm{SCT}-\mathrm{SCF}-\mathrm{SCC}-\mathrm{SCL}=13.875$$

F. V.	S. C.	G. L.	C. M.	Fexp	R^2
Filas = Abonos	329.6875	3	109.89583	47.523	0.7749
Columnas = Insecticidas	3.6875	3	1.22917	0.532	0.0087
L. Latina = Semillas	78.1875	3	26.06250	11.270	0.1838
Residual	13. 8750	6	2.31250		
TOTAL	425.4375	15			0.9674

Nivel de Significación: $\alpha = 0.05$				
F. = Abonos	$F_{0.95;\,3,6} = 4.7571$	Fexp = 47.523	Fexp > F-t	Rech. H_0
C. = Insecticidas	$F_{0.95;\,3,6} = 4.7571$	Fexp = 0.532	Fexp < F-t	Acep. H_0
L. L. = Semillas	$F_{0.95;\,3,6} = 4.7571$	Fexp = 11.270	Fexp > F-t	Rech. H_0

PRESCINDIENDO DE LOS INSECTICIDAS:

F. V.	S. C.	G. L.	C. M.	Fexp	R^2
F. = Abonos	329.6875	3	109.89583	56.3192	0.7749
L. L. = Semillas	78.1875	3	26.06250	13.3564	0.1838
Residual	13. 8750 + 3.6875 = 17. 5625	9	1.9513		
TOTAL	425.4375	15			0.9587

Nivel de Significación: $\alpha = 0.05$				
F. = Abonos	$F_{0.95;\,3,9} = 3.86$	Fexp = 56.3192	Fexp > F-t	Rech. H_0
L. L. = Semillas	$F_{0.95;\,3,9} = 3.86$	Fexp = 13.3564	Fexp > F-t	Rech. H_0

Es mejor el modelo que prescinde de los insecticidas.

5.4.2. Diseño en cuadrados greco-latinos.

- Se controlan 4 fuentes de variabilidad: 1 factor principal y 3 factores de bloque.
- Cada uno de los factores tiene el mismo número de niveles, K.
- Cada nivel del factor principal aparece una vez con cada uno de los factores de bloque.
- No hay interacción entre los factores.
- El número de observaciones es K^2.

A	B	C	D
D	C	B	A
B	A	D	C
C	D	A	B

A	B	C	D
C	D	A	B
D	C	B	A
B	A	D	C

α	β	γ	δ
γ	δ	α	β
δ	γ	β	α
β	α	δ	γ

A α	B β	C γ	D δ
D γ	C δ	B α	A β
B δ	A γ	D β	C α
C β	D α	A δ	BΓ

Modelo estadístico.

$$\boxed{y_{ij(hp)} = \mu + \tau_i + \beta_j + \gamma_h + \delta_p + u_{ij(hp)}; \quad i, j, h, p : 1, \ldots, K}$$

$y_{ij(hp)}$: Observación de la i-ésima fila, j-ésima columna y h-ésima letra latina y p-ésima letra griega.

μ : Efecto constante, común a todos los niveles. Media global.

τ_i : Efecto del i-ésimo nivel del factor fila.

β_j : Efecto del j-ésimo nivel del factor columna.

γ_h : Efecto del h-ésimo nivel del factor letra latina.

δ_p : Efecto del p-ésimo nivel del factor letra griega.

$u_{ij(hp)}$: Errores experimentales, v.a. independendientes y normales.

ANÁLISIS DEL MODELO.

- $N = K^2$ es el número total de observaciones.
- El total y la media general: $y_{\bullet\bullet\bullet\bullet} = \sum_{i=1}^{K}\sum_{j=1}^{K} y_{ij(\bullet\bullet)}$ $\quad \bar{y}_{\bullet\bullet\bullet\bullet} = \dfrac{y_{\bullet\bullet\bullet\bullet}}{K^2}$
- El total y la media por fila: $y_{i\bullet\bullet\bullet} = \sum_{j=1}^{K} y_{ij(\bullet\bullet)}$ $\quad \bar{y}_{i\bullet\bullet\bullet} = \dfrac{y_{i\bullet\bullet\bullet}}{K}$
- El total y la media por columna: $y_{\bullet j\bullet\bullet} = \sum_{i=1}^{K} y_{ij(\bullet\bullet)}$ $\quad \bar{y}_{\bullet j\bullet\bullet} = \dfrac{y_{\bullet j\bullet\bullet}}{K}$
- El total y la media por letra latina: $y_{\bullet\bullet h\bullet} = \sum_{i=1}^{K}\sum_{j=1}^{K} y_{ij(h\bullet)}$ $\quad \bar{y}_{\bullet\bullet h\bullet} = \dfrac{y_{\bullet\bullet h\bullet}}{K}$
- El total y la media por letra griega: $y_{\bullet\bullet\bullet p} = \sum_{i=1}^{K}\sum_{j=1}^{K} y_{ij(\bullet p)}$ $\quad \bar{y}_{\bullet\bullet\bullet p} = \dfrac{y_{\bullet\bullet\bullet p}}{K}$

Estimación por máxima verosimilitud:

$$\hat{\mu} = \bar{y}_{\bullet\bullet\bullet\bullet}$$

$$\hat{\tau}_i = \bar{y}_{i\bullet\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet\bullet}$$

$$\hat{\beta}_j = \bar{y}_{\bullet j\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet\bullet}$$

$$\hat{\gamma}_h = \bar{y}_{\bullet\bullet h\bullet} - \bar{y}_{\bullet\bullet\bullet\bullet}$$

$$\hat{\delta}_p = \bar{y}_{\bullet\bullet\bullet p} - \bar{y}_{\bullet\bullet\bullet\bullet}$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_i \sum_j (y_{ij(hp)} - \hat{\mu} - \hat{\tau}_i - \hat{\beta}_j - \hat{\gamma}_h - \hat{\delta}_p)^2$$

Descomposición de la variabilidad:

Entre filas: $SCF = K\sum_{i=1}^{K}(\bar{y}_{i\bullet\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet\bullet})^2$

Entre columnas: $SCC = K\sum_{j=1}^{K}(\bar{y}_{\bullet j\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet\bullet})^2$

Entre letras latinas: $SCL = K\sum_{h=1}^{K}(\bar{y}_{\bullet\bullet h\bullet} - \bar{y}_{\bullet\bullet\bullet\bullet})^2$

Entre letras griegas: $SCG = K\sum_{p=1}^{K}(\bar{y}_{\bullet\bullet\bullet p} - \bar{y}_{\bullet\bullet\bullet\bullet})^2$

Residual: $SCR = \sum_{i=1}^{K}\sum_{j=1}^{K}(y_{ij(hp)} - \bar{y}_{i\bullet\bullet\bullet} - \bar{y}_{\bullet j\bullet\bullet} - \bar{y}_{\bullet\bullet h\bullet} - \bar{y}_{\bullet\bullet\bullet p} + 3\bar{y}_{\bullet\bullet\bullet\bullet})^2$

Total: $SCT = SCF + SCC + SCL + SCG + SCR = \sum_{i=1}^{K}\sum_{j=1}^{K}(y_{ij(hp)} - \bar{y}_{\bullet\bullet\bullet\bullet})^2$

Forma práctica de la tabla anova:

F. V.	S. C.	G. L.	C. M.	Fexp
Filas	$SCF = \frac{1}{K}\sum_{i=1}^{K} y_{i\cdots}^2 - \frac{y_{\cdots\cdot}^2}{N}$	K – 1	S_F^2	S_F^2 / S_R^2
Columnas	$SCC = \frac{1}{K}\sum_{j=1}^{K} y_{\cdot j\cdot\cdot}^2 - \frac{y_{\cdots\cdot}^2}{N}$	K – 1	S_C^2	S_C^2 / S_R^2
L-Latinas	$SCL = \frac{1}{K}\sum_{h=1}^{K} y_{\cdot\cdot h\cdot}^2 - \frac{y_{\cdots\cdot}^2}{N}$	K – 1	S_L^2	S_L^2 / S_R^2
L-Griegas	$SCG = \frac{1}{K}\sum_{p=1}^{K} y_{\cdot\cdot\cdot p}^2 - \frac{y_{\cdots\cdot}^2}{N}$	K – 1	S_G^2	S_G^2 / S_R^2
Residual	$SCR = SCT - SCF - SCC - SCL - SCG$	(K – 1) (K – 3)	S_R^2	
TOTAL	$SCT = \sum_{i=1}^{K}\sum_{j=1}^{K} y_{ij(hp)}^2 - \frac{y_{\cdots\cdot}^2}{N}$	N – 1		

Hipótesis estadísticas:

$$\boxed{H_0 : \tau_1 = \tau_2 = \ldots = \tau_K = 0} \longrightarrow \boxed{\frac{S_F^2}{S_R^2} \rightarrow F_{(K-1),(K-1)(K-3)}}$$

$$\boxed{H_0 : \beta_1 = \beta_2 = \ldots = \beta_K = 0} \longrightarrow \boxed{\frac{S_C^2}{S_R^2} \rightarrow F_{(K-1),(K-1)(K-3)}}$$

$$\boxed{H_0 : \gamma_1 = \gamma_2 = \ldots = \gamma_K = 0} \longrightarrow \boxed{\frac{S_L^2}{S_R^2} \rightarrow F_{(K-1),(K-1)(K-3)}}$$

$$\boxed{H_0 : \delta_1 = \delta_2 = \ldots = \delta_K = 0} \longrightarrow \boxed{\frac{S_G^2}{S_R^2} \rightarrow F_{(K-1),(K-1)(K-3)}}$$

Coeficientes de Determinación:

$$\boxed{R^2 = \frac{SCF + SCC + SCL + SCG}{SCT}}$$

$$\boxed{R_F^2 = \frac{SCF}{SCT}} \qquad \boxed{R_C^2 = \frac{SCC}{SCT}} \qquad \boxed{R_L^2 = \frac{SCL}{SCT}} \qquad \boxed{R_G^2 = \frac{SCG}{SCT}}$$

Ejemplo.

En la obtención de un determinado producto se está interesado en comparar 4 procedimientos. Se supone que en dicha obtención también puede influir la temperatura, presión y tipo de catalizador empleado, decidiéndose realizar un experimento en cuadrado greco-latino. Para ello se consideran 4 niveles de cada uno de estos factores. La tabla adjunta muestra el cuadrado greco-latino que resulta elegido y las cantidades de producto obtenidas. En dicha tabla: las filas representan el factor principal, procedimientos; Las columnas representan el factor temperatura; Las letras latinas representan el factor presión; Las letras griegas representan el factor catalizador.

Procedimientos	T. 1	T. 2	T. 3	T. 4
P. 1	β C 5	α B 12	δ A 13	γ D 13
P. 2	γ B 6	δ C 10	α D 15	β A 11
P. 3	δ D 7	γ A 5	β B 5	α C 7
P. 4	α A 11	β D 10	γ C 8	δ B 9

Procedimientos	T. 1	T. 2	T. 3	T. 4	$y_{i\ldots}$	$y_{i\ldots}^2$
P. 1	β C 5	α B 12	δ A 13	γ D 13	43	1849
P. 2	γ B 6	δ C 10	α D 15	β A 11	42	1764
P. 3	δ D 7	γ A 5	β B 5	α C 7	24	576
P. 4	α A 11	β D 10	γ C 8	δ B 9	38	444
$y_{.j..}$	29	37	41	40	147	5633
$y_{.j..}^2$	841	1369	1681	1600	5491	
$\sum y_{ij(..)}^2$	231	369	483	420	1503	

Letra Latina	Observaciones				$y_{..h.}$	$y^2_{..h.}$
A	11	5	13	11	40	1600
B	6	12	5	9	32	1024
C	5	10	8	7	30	900
D	7	10	15	13	45	2025
					147	5549

Letra Griega	Observaciones				$y_{...p}$	$y^2_{...p}$
α	11	12	15	7	45	2025
β	5	10	5	11	31	961
γ	6	5	8	13	32	1024
δ	7	10	13	9	39	1521
					147	5531

$$SCT=\sum_i\sum_j y_{ij(..)}^2-\frac{1}{K^2}y_{....}^2=1503-\frac{1}{4^2}\times 147^2=152.4375$$

$$SCF=\frac{1}{K}\sum_i y_{i...}^2-\frac{1}{K^2}y_{....}^2=\frac{1}{4}5633-\frac{1}{4^2}\times 147^2=57.6875$$

$$SCC=\frac{1}{K}\sum_j y_{.j..}^2-\frac{1}{K^2}y_{....}^2=\frac{1}{4}5491-\frac{1}{4^2}\times 147^2=22.1875$$

$$SCL=\frac{1}{K}\sum_h y_{..h.}^2-\frac{1}{K^2}y_{....}^2=\frac{1}{4}5549-\frac{1}{4^2}\times 147^2=36.6875$$

$$SCL=\frac{1}{K}\sum_p y_{...p}^2-\frac{1}{K^2}y_{....}^2=\frac{1}{4}5531-\frac{1}{4^2}\times 147^2=32.1875$$

$$SCR = SCT - SCF - SCC - SCLL - SCLG = 3.6875$$

F. V.	S. C.	G. L.	C. M.	Fexp	R^2
Filas = Procedimientos	57.6875	3	19.2291	15.644	0.3784
Columnas = Temperatura	22.1875	3	7.3958	6.017	0.1455
L. Latina = Presión	36.6875	3	12.2291	9.949	0.2406
L. Griega = Catalizador	32.1875	3	10.7291	8.729	0.2111
Residual	3.6875	3	1.2291		
TOTAL	152.4375	15			0.9756

Nivel de Significación: $\alpha = 0.05$				
F. = Procedimientos	$F_{0.95;3,3} = 9.28$	Fexp = 15.644	Fexp > F-t	Rech. H_0
C. = Temperatura	$F_{0.95;3,3} = 9.28$	Fexp = 6.017	Fexp < F-t	Acep. H_0
L. L. = Presión	$F_{0.95;3,3} = 9.28$	Fexp = 9.949	Fexp > F-t	Rech. H_0
L. G. = Catalizador	$F_{0.95;3,3} = 9.28$	Fexp = 8.729	Fexp < F-t	Acep. H_0

PRESCINDIENDO DE LA TEMPERATURA (COLUMNAS):

F. V.	S. C.	G. L.	C. M.	Fexp	R^2
Procedimientos	57.6875	3	19.2291	4.45889	0.3784
Presión	36.6875	3	12.2291	2.8375	0.2406
Catalizador	32.1875	3	10.7291	2.48788	0.2111
Residual	25.875	6	4.3125		
TOTAL	152.4375	15			0.8301

Nivel de Significación: $\alpha = 0.05$				
Procedimientos	$F_{0.95;\,3,6} = 4.76$	Fexp = 4.45889	Fexp < F-t	Acep. H_0
Presión	$F_{0.95;\,3,6} = 4.76$	Fexp = 2.8375	Fexp < F-t	Acep. H_0
Catalizador	$F_{0.95;\,3,6} = 4.76$	Fexp = 2.48788	Fexp < F-t	Acep. H_0

Es mejor el modelo completo.

5.4.3. Diseño en cuadrados de Youden.

- Un diseño en cuadrado de Youden es un cuadrado latino incompleto.
- Cada tratamiento ocurre una vez en cada columna.
- El número de réplicas de un tratamiento es igual al número de tratamientos por bloque.

Ejemplo.

Si asignamos:

- El factor principal a las letras latinas.
- El factor secundario, con el mismo número de niveles que el factor principal, a las filas.
- El factor secundario, con menor número de niveles que el factor principal, a las columnas

Insecticidas	Abonos		
	A1	A2	A3
I1	A 23	B 25	C 16
I2	B 18	C 15	D 17
I3	C 19	D 25	A 18
I4	D 21	A 12	B 20

Modelo estadístico.

$$\boxed{y_{ijk} = \mu + \tau_i + \beta_j + \gamma_k + u_{ijk}; \quad i:1,\ldots,I; \;\; j:1,\ldots,J; \;\; \mathrm{k}:1,\ldots,K; \;\; I = J; \;\; K < I}$$

y_{ijk} : Observación del i-ésimo tratamiento, j-ésimo bloque y k-ésima columna.

μ : Media general.

τ_i : Efecto del tratamiento i-ésimo.

β_j : Efecto del bloque j-ésimo.

γ_k : Efecto del k-ésimo nivel del factor situado en las columnas.

u_{ijk} : Errores experimentales, v.a. independendientes y normales.

ANÁLISIS DEL MODELO PARA ESTUDIAR EL EFECTO DE LOS TRATAMIENTOS.

DESCOMPOSICIÓN DE LA VARIABILIDAD:

Entre tratamientos: $SCTr^* = \dfrac{K\sum_{i=1}^{I} T_i^2}{\lambda I}$, $\lambda = R\dfrac{K-1}{I-1}$, $T_i = y_{i\bullet\bullet} - \dfrac{1}{K}\sum_{j=1}^{J} n_{ij}\, y_{\bullet j\bullet}$, $i = 1, 2, \ldots, I$

$$n_{ij} = \begin{cases} 1 & \text{si el tratamiento i ocurre en el bloque j} \\ 0 & \text{en otro caso} \end{cases} \qquad \sum_{i=1}^{I} T_i = 0$$

Entre bloques: $SCBl = \dfrac{1}{k}\sum_{j=1}^{J} y_{\bullet j\bullet}^2 - \dfrac{y_{\bullet\bullet\bullet}^2}{N}$

Entre columnas: $SCC = \dfrac{1}{I}\sum_{k=1}^{K} y_{\bullet\bullet k}^2 - \dfrac{y_{\bullet\bullet\bullet}^2}{N}$ Total: $SCT = \sum_{j=1}^{J}\sum_{k=1}^{K} y_{(\bullet)jk}^2 - \dfrac{y_{\bullet\bullet\bullet}^2}{N}$

Residual: $SCR = SCT - SCTr^* - SCBl - SCC$

Forma práctica de la tabla ANOVA:

F. V.	S. C.	G. L.	C. M.	Fexp
Trat-Ajustados	$\dfrac{K\sum_{i=1}^{I} T_i^2}{\lambda I}$	I – 1	$CMTr^*$	$CMTr^* / CMR$
Bloques-No-Ajustados	$\dfrac{1}{K}\sum_{j=1}^{J} y_{\bullet j \bullet}^2 - \dfrac{y_{\bullet\bullet\bullet}^2}{N}$	J – 1		
Columnas	$\dfrac{1}{I}\sum_{k=1}^{K} y_{\bullet\bullet k}^2 - \dfrac{y_{\bullet\bullet\bullet}^2}{N}$	K – 1		
Residual	$SCR = SCT - SCTr^* - SCBl - SCC$	N – 2I – K + 2	CMR	
TOTAL	$\sum_{j=1}^{J}\sum_{k=1}^{K} y_{(\bullet)jk}^2 - \dfrac{y_{\bullet\bullet\bullet}^2}{N}$	N – 1		

ANÁLISIS DEL MODELO PARA ESTUDIAR EL EFECTO DE LOS BLOQUES.

DESCOMPOSICIÓN DE LA VARIABILIDAD:

Entre bloques: $SCBl^{*} = \dfrac{R\sum_{j=1}^{J} B_j^2}{\lambda I}, \quad B_j = y_{\bullet j \bullet} - \dfrac{1}{R}\sum_{i=1}^{I} n_{ij} y_{i\bullet\bullet}, \quad j = 1, 2, \ldots, J$

$$n_{ij} = \begin{cases} 1 & \text{si el tratamiento i ocurre en el bloque j} \\ 0 & \text{en otro caso} \end{cases} \qquad \sum_{j=1}^{J} B_j = 0$$

Entre tratamientos: $SCTr = \dfrac{1}{k}\sum_{i=1}^{I} y_{i\bullet\bullet}^2 - \dfrac{y_{\bullet\bullet\bullet}^2}{N}$

Entre columnas: $SCC = \dfrac{1}{I}\sum_{k=1}^{K} y_{\bullet\bullet k}^2 - \dfrac{y_{\bullet\bullet\bullet}^2}{N}$ Total: $SCT = \sum_{j=1}^{J}\sum_{k=1}^{K} y_{(\bullet)jk}^2 - \dfrac{y_{\bullet\bullet\bullet}^2}{N}$

Residual: $SCR = SCT - SCTr^{*} - SCBl - SCC$

FORMA PRÁCTICA DE LA TABLA ANOVA:

F. V.	S. C.	G. L.	C. M.	Fexp
Trat-No-Ajustados	$\frac{1}{K}\sum_{i=1}^{I} y_{i\bullet\bullet}^2 - \frac{y_{\bullet\bullet\bullet}^2}{N}$	I – 1		
Bloques-Ajustados	$SCBl^* = \frac{R\sum_{j=1}^{J} B_j^2}{\lambda J}$	J – 1	$CMBl^*$	$CMBl^* / CMR$
Columnas	$\frac{1}{I}\sum_{k=1}^{K} y_{\bullet\bullet k}^2 - \frac{y_{\bullet\bullet\bullet}^2}{N}$	K – 1		
Residual	$SCR = SCT - SCTr - SCBl^* - SCC$	N – 2I – K + 2	CMR	
TOTAL	$\sum_{j=1}^{J}\sum_{k=1}^{K} y_{(\bullet)jk}^2 - \frac{y_{\bullet\bullet\bullet}^2}{N}$	N – 1		

Ejemplo.

Consideremos el ejemplo de la semilla de trigo. Al plantear este experimento se pensó que podría conseguirse mayor precisión si se controlaba la variabilidad introducida por los tipos de abono e insecticida. El instituto de experimentación agrícola está interesado en estudiar 4 tipos de semilla de trigo (S1, S2, S3, S4), pero sólo se dispone de 3 tipos de abono. Se decide utilizar un cuadrado de Youden.

Insecticidas	Abonos		
	A1	A2	A3
I1	A 23	B 25	C 16
I2	B 18	C 15	D 17
I3	C 19	D 25	A 18
I4	D 21	A 12	B 20

Insecticidas	Abonos					
	A1	A2	A3	$y_{\bullet j \bullet}$	$y^2_{\bullet j \bullet}$	$\sum_k y^2_{(\bullet)jk}$
I1	A 23	B 25	C 16	64	4096	1410
I2	B 18	C 15	D 17	50	2500	838
I3	C 19	D 25	A 18	62	3844	1310
I4	D 21	A 12	B 20	53	2809	985
$y_{\bullet \bullet k}$	81	77	71	229	13249	4543
$y^2_{\bullet \bullet k}$	6561	5929	5041	17531		

Parámetros del modelo:

$$N = IR = JK = 12$$

$$I = 4 = J; \quad K = 3 = R$$

$$\lambda = R\frac{K-1}{I-1} = 3\frac{2}{3} = 2$$

EFECTO DE LOS TRATAMIENTOS:

Letra Latina	Observaciones			$y_{i\bullet\bullet}$	$y_{i\bullet\bullet}^2$
A	23	18	12	53	2809
B	25	18	20	63	3969
C	16	15	19	50	2500
D	17	25	21	63	3969
				229	13247

$$SCT = \sum_{j=1}^{4}\sum_{k=1}^{3} y_{(\bullet)jk}^2 - \frac{y_{\bullet\bullet\bullet}^2}{N} = 4543 - \frac{229^2}{12} = 172.91$$

$$SCBl = \frac{1}{K}\sum_{j=1}^{4} y_{\bullet j\bullet}^2 - \frac{y_{\bullet\bullet\bullet}^2}{N} = \frac{13249}{3} - \frac{229^2}{12} = 46.25$$

$$SCC = \frac{1}{I}\sum_{k=1}^{K} y_{\bullet\bullet k}^2 - \frac{y_{\bullet\bullet\bullet}^2}{N} = \frac{17531}{4} - \frac{229^2}{12} = 12.66$$

Insecticidas	Abonos						
	A1		A2		A3		$y_{\bullet j \bullet}$
I1	A	23	B	25	C	16	64
I2	B	18	C	15	D	17	50
I3	C	19	D	25	A	18	62
I4	D	21	A	12	B	20	53
$y_{\bullet \bullet k}$	81		77		71		229

$$T_i = y_{i\bullet\bullet} - \frac{1}{K}\sum_{j=1}^{5} n_{ij} y_{\bullet j \bullet}, \quad i = 1,2,\ldots.I$$

$$T_1 = 53 - \tfrac{1}{3}(64+62+53) = -6.66 \qquad T_2 = 63 - \tfrac{1}{3}(64+50+53) = 7.33$$

$$T_3 = 50 - \tfrac{1}{3}(64+50+62) = -8.66 \qquad T_4 = 63 - \tfrac{1}{3}(50+62+53) = 8$$

$$SCTr^* = \frac{K\sum_{i=1}^{I} T_i^2}{\lambda I} = \frac{3(237.33)}{2\times 4} = 89$$

$$SCR = SCT - SCTr^* - SCBl - SCC = 172.91 - 89 - 46.25 - 12.66 = 25$$

F. V.	S. C.	G. L.	C. M.	Fexp
Trat-Ajustados	89	3	29.66	3.56
Bloques-No-Ajustados	46.25	3		
Columnas	12.66	2		
Residual	25	3	8.33	
TOTAL	172.91	11		

Nivel de Significación: $\alpha = 0.05$			
$F_{0.95;\,3,3} = 9.28$	Fexp = 3.56	Fexp < F-t	Acep. H_0
Nivel de Significación: $\alpha = 0.01$			
$F_{0.99;\,3,3} = 29.46$	Fexp = 3.56	Fexp < F-t	Acep. H_0

EFECTO DE LOS BLOQUES:

Insecticidas	Abonos						
	A1		A2		A3		
I1	A	23	B	25	C	16	64
I2	B	18	C	15	D	17	50
I3	C	19	D	25	A	18	62
I4	D	21	A	12	B	20	53
	81		77		71		229

$$B_j = y_{\bullet j \bullet} - \frac{1}{R}\sum_{i=1}^{I} n_{ij} y_{i\bullet\bullet}, \quad j = 1,2,\ldots,J$$

$$B_1 = 64 - \tfrac{1}{3}(53+63+50) = 8.66 \qquad B_2 = 50 - \tfrac{1}{3}(63+50+63) = -8.66$$

$$B_3 = 62 - \tfrac{1}{3}(53+50+63) = 6.66 \qquad B_4 = 53 - \tfrac{1}{3}(53+63+63) = -6.66$$

$$SCBl^* = \frac{R\sum_{j=1}^{4} B_j^2}{\lambda J} = \frac{3(239.11)}{2\times 4} = 89.66$$

$$SCTr = \frac{1}{K}\sum_{i=1}^{I} y_{i\bullet\bullet}^2 - \frac{y_{\bullet\bullet\bullet}^2}{N} = \frac{13247}{3} - \frac{229^2}{12} = 45.58$$

F. V.	S. C.	G. L.	C. M.	Fexp
Trat-No-Ajustados	45.58	3		
Bloques-Ajustados	89.66	3	29.98	3.587
Columnas	12.66	2		
Residual	25	3	8.33	
TOTAL	172.91	11		

Nivel de Significación: $\alpha = 0.05$			
$F_{0.95;\,3,3} = 9.28$	Fexp = 3.587	Fexp < F-t	Acep. H_0
Nivel de Significación: $\alpha = 0.01$			
$F_{0.99;\,3,3} = 29.46$	Fexp = 3.587	Fexp < F-t	Acep. H_0

TEMA 5.5. DISEÑOS FACTORIALES.

5.5.1. Diseños factoriales con dos factores.

5.5.2. Diseños factoriales con tres factores.

5.5.1. DISEÑOS FACTORIALES CON DOS FACTORES.

DISEÑO SIN REPLICACIÓN:

Se toma una observación para cada combinación de niveles de los dos factores. Número de observaciones: N=ab

$$\boxed{y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + u_{ij}; \quad i:1,\dots,a; \;\; j:1,\dots,b;}$$

DISEÑO CON REPLICACIÓN:

Se toman r observaciones para cada combinación de niveles de los dos factores. Número de observaciones: N=abr

$$\boxed{y_{ijs} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + u_{ijs}; \quad i:1,\dots,a; \;\; j:1,\dots,b; \; s:1,\dots,r}$$

$$\boxed{y_{ij} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + u_{ij}; \quad i:1,\dots,a; \;\; j:1,\dots,b;}$$

y_{ij} : Observación del nivel (i) del factor A y del nivel (j) del factor B

μ : Media general

τ_i : Efecto del nivel (i) del factor A

β_j : Efecto del nivel (j) del factor B

$(\tau\beta)_{ij}$: Efecto de la interacción entre A y B

u_{ij} : Errores experimentales

Forma práctica de la tabla ANOVA:

F. V.	S. C.	G. L.	C. M.	Fexp
Factor A	$\frac{1}{br}\sum_{i=1}^{a} y_{i\bullet\bullet}^{2} - \frac{y_{\bullet\bullet\bullet}^{2}}{abr}$	a – 1	CMA	CMA/CMR
Factor B	$\frac{1}{ar}\sum_{j=1}^{b} y_{\bullet j\bullet}^{2} - \frac{y_{\bullet\bullet\bullet}^{2}}{abr}$	b – 1	CMB	CMB/CMR
Interacción	$\frac{1}{r}\sum_{ij} y_{ij\bullet}^{2} - \frac{y_{\bullet\bullet\bullet}^{2}}{abr} - SCA - SCB$	(a – 1)(b – 1)	CMAB	CMAB/CMR
Residual	$SCR =$ $SCT - SCA - SCB - SCAB$	ab(r – 1)	CMR	
TOTAL	$\sum_{ijs} y_{ijs}^{2} - \frac{y_{\bullet\bullet\bullet}^{2}}{abr}$	abr – 1		

Hipótesis estadísticas:

$$\boxed{H_0 : \tau_1 = \tau_2 = \ldots = \tau_a = 0} \longrightarrow \boxed{\frac{S_A^2}{S_R^2} \rightarrow F_{(a-1),ab(r-1)}}$$

$$\boxed{H_0 : \beta_1 = \beta_2 = \ldots = \beta_b = 0} \longrightarrow \boxed{\frac{S_B^2}{S_R^2} \rightarrow F_{(b-1),ab(r-1)}}$$

$$\boxed{H_0 : (\tau\beta)_{ij} = 0} \longrightarrow \boxed{\frac{S_{AB}^2}{S_R^2} \rightarrow F_{(a-1)(b-1),ab(r-1)}}$$

Coeficientes de Determinación:

$$\boxed{R^2 = \frac{SCA + SCB + SCAB}{SCT}}$$

$$\boxed{R_A^2 = \frac{SCA}{SCT}} \qquad \boxed{R_B^2 = \frac{SCB}{SCT}} \qquad \boxed{R_{AB}^2 = \frac{SCAB}{SCT}}$$

5.5.2. Diseños factoriales con tres factores.

DISEÑO SIN REPLICACIÓN:

Se toma una observación para cada combinación de niveles de los tres factores. Número de observaciones: N=abc

$$\boxed{\begin{array}{l} y_{ijk} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + u_{ijk}; \\ i:1,\ldots,a;\ \ j:1,\ldots,b;\ \ k:1,\ldots,c \end{array}}$$

DISEÑO CON REPLICACIÓN:

Se toman r observaciones para cada combinación de niveles de los tres factores. Número de observaciones: N=abcr

$$\boxed{\begin{array}{l} y_{ijks} = \mu + \tau_i + \beta_j + \gamma_k + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + u_{ijks}; \\ i:1,\ldots,a;\ \ j:1,\ldots,b;\ \ k:1,\ldots,c;\ \ s:1,\ldots,r \end{array}}$$

Ejemplo.

Se experimenta en un proceso de fabricación en dos condiciones de funcionamiento τ_1 y τ_2 , con tres tipos de materia prima β_1, β_2 y β_3 y dos valores de una variable de control γ_1 y γ_2, obteniéndose las siguientes observaciones, donde la respuesta es una medida de la calidad del producto.

	γ_1			γ_2		
	β_1	β_2	β_3	β_1	β_2	β_3
τ_1	20	30	12	16	33	8
τ_2	36	38	40	40	44	42

Factores	A	B	C
Niveles	a=2	b=3	c=2

Construya la tabla ANOVA para estudiar el efecto de la interacción doble entre los factores y conteste:

a) ¿Qué factores interaccionan entre sí de forma signifativa? Excluya del modelo las interacciones que no afectan en la variable respuesta.
b) Estudie el efecto de los factores y discuta si es mejor utilizar el modelo completo o si sería mejor prescindir de algún factor.
c) Estudiar la idoneidad del mejor modelo y dar el valor del coeficiente de determinación.

TEMA 6. DISEÑOS NO PARAMÉTRICOS PARA EL ANÁLISIS DE LA VARIANZA .

6.1. Contraste de Rangos de Kruskal-Wallis.

6.2. Contraste de la Mediana.

6.3. Contraste de Cochran.

6.4. Contraste de Rangos de Friedman.

6.1. CONTRASTE DE RANGOS DE KRUSKAL-WALLIS.

1. Las observaciones están medidas en al menos la escala ordinal.
2. Las poblaciones de las que son extraídas las muestras son idénticas, pero pueden diferir en la media.
3. Las muestras de los diferentes grupos son independientes entre sí

$$\boxed{\begin{aligned} &H_0 : \mu_1 = \mu_2 = \ldots = \mu_I \\ &H_1 : \mu_i \neq \mu_j \text{ para algún par } (i, j) \end{aligned}}$$

- Se asignan **rangos** a las N observaciones
- A las observaciones ligadas se les asigna la ***media de los rangos***

Estadístico de contraste:

$$\boxed{\chi^2_{KW} = \frac{12}{N(N+1)} \sum_{i=1}^{I} \frac{R_{i\bullet}^2}{n_i} - 3(N+1)}$$

- N es el número total de observaciones
- n_i es el número de observaciones del i-ésimo tratamiento
- R_i es la suma de los rangos de las observaciones del i-ésimo tratamiento

$$\boxed{Si \quad \chi^2_{KW(\exp)} > \chi^2_{1-\alpha;I-1} \quad \Rightarrow \quad \text{Se rechaza } H_0}$$

Corrección para las observaciones ligadas:

$$\chi^2_{KW*} = \frac{\chi^2_{KW}}{C} \qquad C = 1 - \frac{\sum_{i=1}^{s} T_i}{(N^3 - N)} \qquad T_i = (t_i^3 - t_i)$$

- s es el número de observaciones ligadas
- t_i es el número de observaciones ligadas en el grupo i-ésimo

Ejemplo.

Una compañía publicitaria está interesada en realizar un estudio sobre el número de horas que los adolescentes, pertenecientes a distintos niveles de renta familiar, dedican a ver la televisión. Para saber si hay diferencias significativas entre los distintos niveles de renta, se establecen 5 niveles y se realiza una encuesta a un grupo de adolescentes. La tabla adjunta presenta los resultados de dicha encuesta.

Renta	Número de horas					
1	49	51	50	49	51	51
2	60	56	56	56	57	
3	50	48	53	44	45	44
4	48	47	49	44	43	
5	43	43	46	47	45	46

Observaciones	Grupo	Rango	
43	4	1	
43	5	2	2
43	5	3	
44	3	4	
44	3	5	5
44	4	6	
45	3	7	7.5
45	5	8	
46	5	9	9.5
46	5	10	
47	4	11	11.5
47	5	12	
48	3	13	14.5
48	4	14	

Observaciones	Grupo	Rango	
49	1	15	
49	1	16	16
49	4	17	
50	1	18	11.5
50	3	19	
51	1	20	
51	1	21	21
51	1	22	
53	3	23	23
56	2	24	
56	2	25	25
56	2	26	
57	2	27	27
	2		
60		28	28

TABLA POR GRUPOS

Número de horas									
Nivel 1		Nivel 2		Nivel 3		Nivel 4		Nivel 5	
y_{1j}	Rango	y_{2j}	Rango	y_{3j}	Rango	y_{4j}	Rango	y_{5j}	Rango
49	16	60	28	50	18.5	48	13.5	43	2
51	21	56	25	48	13.5	47	11.5	43	2
50	18.5	56	25	53	23	49	16	46	9.5
49	16	56	25	44	5	44	5	47	11.5
51	21	57	27	45	7.5	43	2	45	7.5
51	21			44	5			46	9.5
$R_{i\bullet}$	113.5		130		72.5		48		42
$R_{i\bullet}^2/n_i$	2147.04		3380		876.04		460.8		294

Observac.	43	43	43	44	44	44	45	45	46	46	47	47	48	48		
Rango	2	2	2	5	5	5	7.5	7.5	9.5	9.5	11.5	11.5	13.5	13.5		
Observac.	49	49	49	50	50	50	51	51	51	53	56	56	56	57	60	
Rango	16	16	16	18.5	18.5	18.5	21	21	21	23	25	25	25	27	28	

$$C = 1 - \frac{\sum_{i=1}^{10} T_i}{N^3 - N} = 1 - \frac{\sum_{i=1}^{10}\left(t_i^3 - t_i\right)}{N^3 - N} = 1 - \frac{5\left(2^3 - 2\right) + 5\left(3^3 - 3\right)}{28^3 - 28} = 1 - \frac{150}{21924} = 0.9931$$

$$\chi^2_{KW(\exp)} = \frac{12}{N(N+1)} \sum_{i=1}^{5} \frac{R_{i\bullet}^2}{n_i} - 3(N+1) =$$

$$= \frac{12}{28 \times 29}(2147.04 + 3380 + 876.04 + 460.8 + 294) - 3(29) = 18.7814$$

$$\chi^2_{KW^*(\exp)} = \frac{\chi^2_{KW(\exp)}}{C} = \frac{18.7803}{0.9931} = 18.9108$$

$$\alpha = 0.05, \quad \chi^2_{0.95;4} = 9.488, \quad \chi^2_{KW^*(\exp)} > \chi^2_{0.95;4} \Rightarrow Se\ rechaza\ H_0$$

6.2. Contraste de la mediana.

1. Las observaciones están medidas al menos en la escala ordinal.
2. Las poblaciones de las que son extraídas las muestras son idénticas, pero pueden diferir en la media.
3. Las muestras de los diferentes grupos son independientes entre sí

$$\boxed{\begin{array}{l} H_0 : \mu_1 = \mu_2 = \ldots = \mu_I \\ H_1 : \mu_i \neq \mu_j \text{ para algún par } (i, j) \end{array}}$$

Contrastar la hipótesis de que las poblaciones tienen la misma mediana es equivalente a contrastar que dichas poblaciones son homogéneas respecto a la clasificación de las observaciones.

O_{i1} : Observaciones $\leq$ que la mediana en el tratamiento i-ésimo
O_{i2} : Observaciones $>$ que la mediana en el tratamiento i-ésimo

Estadístico de contraste: $$\boxed{\chi^2_M = \sum_{i=1}^{I}\sum_{j=1}^{2}\frac{(O_{ij}-E_{ij})^2}{E_{ij}}}$$

O_{ij} : Frecuencia observada de la celdilla *ij*
E_{ij} : Frecuencia esperada bajo la hipótesis de homogeneidad

$$\boxed{Si \quad \chi^2_{M\,(\exp)} > \chi^2_{1-\alpha;I-1} \quad \Rightarrow \quad \text{Se rechaza } \mathrm{H}_0}$$

Este contraste requiere que las frecuencias esperadas no sean demasiado pequeñas. Se recomienda que, como máximo el 20% de las celdillas tengan una frecuencia esperada menor que 5 y ninguna menor que 1.

Ejemplo.

Unos investigadores están interesados en saber si cuatro productos (P) producen, por término medio, el mismo grado de satisfacción. Para ello se tomaron cuatro muestras de 17, 15, 13 y 10 personas, pasándoles a cada una un test que medía el grado de satisfacción con el producto en una escala del 0 al 10. Los resultados se muestran en la tabla adjunta.

A	Grados de ansiedad																
1	4	3	7	8	2	4	7	5	3	5	7	4	1	8	2	5	7
2	5	2	4	7	8	2	4	3	6	3	6	10	2	4	9		
3	4	6	7	9	7	3	1	4	1	7	5	7	4				
4	5	7	3	5	4	9	7	5	7	8							

Número de observaciones: 55. **Mediana: 5.**

Tabla de Frecuencias Observadas					
	Grupos				
	1	2	3	4	Total
≤ 5	11	9	7	5	32
> 5	6	6	6	5	23
Total	17	15	13	10	55

Tabla de Frecuencias Esperadas				
	Grupos			
	1	2	3	4
≤ 5	$\frac{32\times17}{55}=9.89$	$\frac{32\times15}{55}=8.72$	$\frac{32\times13}{55}=7.56$	$\frac{32\times10}{55}=5.81$
> 5	$\frac{23\times17}{55}=7.10$	$\frac{23\times15}{55}=6.27$	$\frac{23\times13}{55}=5.43$	$\frac{23\times10}{55}=4.18$

$$\chi^2_{M(\exp)} = \sum_{i=1}^{4}\sum_{j=1}^{2}\frac{\left(O_{ij}-E_{ij}\right)^2}{E_{ij}} =$$

$$\frac{(11-9.89)^2}{9.89}+\frac{(6-7.10)^2}{7.10}+\frac{(9-8.72)^2}{8.72}+\frac{(6-6.27)^2}{6.27}+$$

$$\frac{(7-7.56)^2}{7.56}+\frac{(6-5.43)^2}{5.43}+\frac{(5-5.81)^2}{5.81}+\frac{(5-4.18)^2}{4.18}=0.6934$$

$$\alpha = 0.05, \quad \chi^2_{0.95;3} = 7.81, \quad \chi^2_{M(\exp)} < \chi^2_{0.95;3} \Rightarrow \textit{Se acepta } H_0$$

6.3. Contraste de Cochran.

Método para examinar si las proporciones de éxito de tres o más tratamientos, con el mismo número de observaciones, difieren significativamente entre sí.

1. Experimentos en bloques con respuestas binarias.
2. Los bloques están formados por grupos homogéneos.
3. Los bloques se eligen al azar.
4. Los tratamientos se asignan al azar dentro de los bloques.

	Tratamientos		
Bloques	1	j	k
1	y_{11}	y_{1j}	y_{1k}
.	.	.	.
i	y_{i1}	y_{ij}	y_{ik}
.	.	.	.
n	y_{n1}	y_{nj}	y_{nk}

$$y_{ij} = \begin{cases} 1 \\ 0 \end{cases} \quad i = 1, 2, \ldots n, \; j = 1, 2, \ldots, k$$

$H_0 : \mu_1 = \mu_2 = \ldots = \mu_I$ (todos los tratamientos tienen el mismo efecto)

$H_1 : \mu_i \neq \mu_j$ para algún par (i, j) (al menos dos tratamientos no tienen el mismo efecto)

Estadístico de contraste:

$$\boxed{\chi^2_C = \frac{(k-1)\left(k\sum_{j=1}^{k} y^2_{\bullet j} - y^2_{\bullet\bullet}\right)}{k y_{\bullet\bullet} - \sum_{i=1}^{n} y^2_{i\bullet}}}$$

$$\boxed{Si \quad \chi^2_{C(\exp)} > \chi^2_{1-\alpha; k-1} \quad \Rightarrow \quad \text{Se rechaza H}_0}$$

Ejemplo.

Unos laboratorios están interesados en comprobar cuatro tratamientos distintos para la mejora de un proceso de fabricación. Para ello se aplican los 4 tratamientos a 15 grupos de 4 productos y se codifica la respuesta como 1 si el tratamiento tuvo éxito en la mejora del producto y como 0 si fracasó. Los resultados se muestran en la siguiente tabla:

	Tratamientos			
Bloques	1	2	3	4
1	1	1	1	1
2	1	0	1	1
3	0	1	1	1
4	1	1	1	1
5	0	0	0	0
6	1	0	0	0
7	1	1	1	0
8	1	1	1	1
9	0	1	1	1
10	1	0	0	1
11	1	1	1	1
12	0	0	0	0
13	1	1	0	0
14	1	0	1	0
15	1	0	1	1

	Tratamientos (j)					
Bloques (i)	1	2	3	4	$y_{i\bullet}$	$y_{i\bullet}^2$
1	1	1	1	1	4	16
2	1	0	1	1	3	9
3	0	1	1	1	3	9
4	1	1	1	1	4	16
5	0	0	0	0	0	0
6	1	0	0	0	1	1
7	1	1	1	0	3	9
8	1	1	1	1	4	16
9	0	1	1	1	3	9
10	1	0	0	1	2	4
11	1	1	1	1	4	16
12	0	0	0	0	0	0
13	1	1	0	0	2	4
14	1	0	1	0	2	4
15	1	0	1	1	3	9
$y_{\bullet j}$	11	8	10	9	38	122
$y_{\bullet j}^2$	121	64	100	81	366	

$$\chi_C^2 = \frac{(k-1)\left(k\sum_{j=1}^{k} y_{\bullet j}^2 - y_{\bullet\bullet}^2\right)}{k y_{\bullet\bullet} - \sum_{i=1}^{n} y_{i\bullet}^2} =$$

$$= \frac{3(4x366 - 38^2)}{4x38 - 122} = 2$$

$$\chi_{C(\exp)}^2 < \chi_{0.95;3}^2 = 7.81 \Rightarrow$$

Se acepta H_0

Los cálculos pueden simplificarse si despreciamos los bloques con idéntica respuesta a todos los tratamientos (todos uno o todos cero).

6.4. Contraste de rangos de Friedman.

1. Experimentos en bloques con observaciones medidas en escala ordinal.
2. Los bloques están formados por grupos homogéneos.
3. Los bloques se eligen al azar.
4. Los tratamientos se asignan al azar dentro de los bloques.
5. Se asignan rangos a las observaciones en cada bloque.

	Tratamientos		
Bloques	1	j	k
1	y_{11}	y_{1j}	y_{1k}
.	.	.	.
i	y_{i1}	y_{ij}	y_{ik}
.	.	.	.
n	y_{n1}	y_{nj}	y_{nk}

R_{ij} : Rango asignado a la observación j-ésima del tratamiento i-ésimo.

$H_0 : \mu_1 = \mu_2 = ... = \mu_I$ (todos los tratamientos tienen el mismo efecto)

$H_1 : \mu_i \neq \mu_j$ para algún par (i, j) (al menos dos tratamientos no tienen el mismo efecto)

Estadístico de contraste: $$\chi_F^2 = \frac{12}{nk(k+1)} \sum_{j=1}^{k} R_{\bullet j}^2 - 3n(k+1)$$

$$Si \quad \chi^2_{F\,(\exp)} > \chi^2_{1-\alpha;k-1} \quad \Rightarrow \quad \text{Se rechaza } H_0$$

Corrección para observaciones ligadas:

$$\chi^2_{F*} = \frac{\chi^2_F}{C} \qquad\qquad C = 1 - \frac{\sum (t_i^3 - t_i)}{n(k^3 - k)}$$

donde t_i es el número de observaciones ligadas en el grupo i-ésimo.

Ejemplo.

Un grupo de investigación desea evaluar la calidad de 4 marcas distintas de un mismo producto. Para ello se eligen al azar a 9 profesionales que asignan a los distintos productos valores en una escala de 1 a 5. Se trata de comprobar si hay diferencias de calidad entre las 4 marcas. Los resultados se encuentran en la siguiente tabla:

	Tratamientos (j)			
Bloques (i)	1	2	3	4
1	3 (2.5)	4 (4)	3 (2.5)	2 (1)
2	2 (2)	2 (2)	4 (4)	2 (2)
3	4 (3.5)	4 (3.5)	3 (1.5)	3 (1.5)
4	1 (1.5)	1 (1.5)	3 (4)	2 (3)
5	4 (3)	3 (2)	5 (4)	1 (1)
6	2 (1)	3 (2.5)	3 (2.5)	4 (4)
7	1 (1)	3 (3)	4 (4)	2 (2)
8	1 (1.5)	1 (1.5)	2 (3.5)	2 (3.5)
9	2 (4)	1 (2)	1 (2)	1 (2)

	Tratamientos (j)				
Bloques (i)	1	2	3	4	$\sum(t^3-t)$
1	3 (2.5)	4 (4)	3 (2.5)	2 (1)	6
2	2 (2)	2 (2)	4 (4)	2 (2)	24
3	4 (3.5)	4 (3.5)	3 (1.5)	3 (1.5)	12
4	1 (1.5)	1 (1.5)	3 (4)	2 (3)	6
5	4 (3)	3 (2)	5 (4)	1 (1)	0
6	2 (1)	3 (2.5)	3 (2.5)	4 (4)	6
7	1 (1)	3 (3)	4 (4)	2 (2)	0
8	1 (1.5)	1 (1.5)	2 (3.5)	2 (3.5)	12
9	2 (4)	1 (2)	1 (2)	1 (2)	24
$R_{\bullet j}$	20	22	28	20	
$R_{\bullet j}^2$	400	484	784	400	2068

$$C = 1 - \frac{\sum\left(t^3 - t\right)}{n\left(k^3 - k\right)} = 1 - \frac{90}{9\left(4^3 - 4\right)} = 0.834$$

$$\chi^2_{F(\exp)} = \frac{12}{nk(k+1)} \sum_{j=1}^{k} R^2_{\bullet j} - 3n(k+1) = \frac{12}{9(4)(5)} 2068 - 3(9)(5) = 2.866$$

$$\chi^2_{F^*(\exp)} = \frac{\chi^2_{F(\exp)}}{C} = \frac{2.866}{0.833} = 3.44$$

$$\alpha = 0.05, \quad \chi^2_{0.95;3} = 7.81, \quad \chi^2_{F^*(\exp)} < \chi^2_{0.95;3} \Rightarrow \textit{Se acepta } H_0$$