ANOVA

Análisis de la Varianza en diseño de experimentos

NATURALEZA DEL DISEÑO EXPERIMENTAL

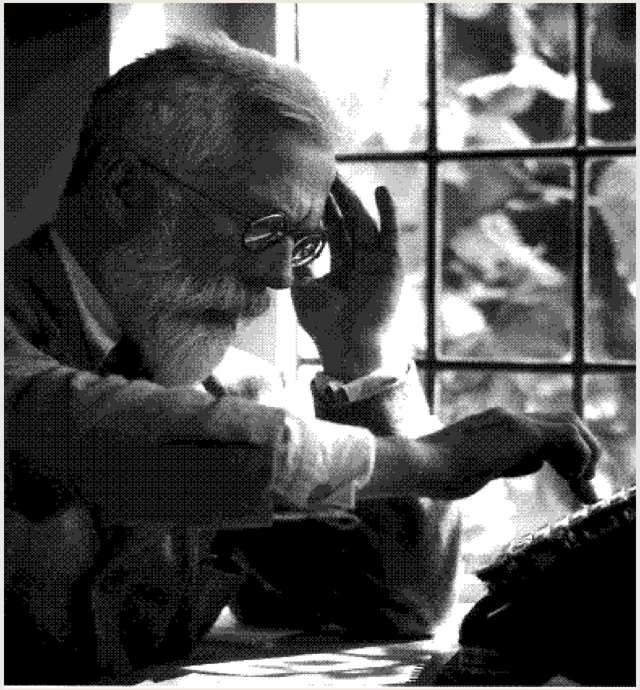
El diseño experimental tiene sus orígenes en los trabajos de Ronald Aylmer Fisher (1890 – 1962) desarrollados en la Estación Agrícola Experimental de Rothamsted, en el Reino Unido, donde introduce los conceptos de aleatorización y el análisis de varianza.

Los cimientos de la materia han llegado a conocerse por el título de su libro The Design of Experiments (1935).

NATURALEZA DEL DISEÑO EXPERIMENTAL

Desde entonces la teoría del diseño experimental ha sido desarrollada y ampliada.

Aplicaciones de esta teoría se encuentran hoy en las investigaciones en ciencias naturales, ingeniería y casi todas las ramas de las ciencias sociales.



Fisher (17/2/1890 – 29/7/1962)

DISEÑOS EXPERIMENTALES

Un diseño estadístico de experimentos (DEE) es el proceso de *planificación* del experimento mediante el cual se *recolectan* datos apropiados que al ser *analizados* por métodos estadísticos estos nos conducen a *conclusiones* metodoló-gicamente válidas.

Es la secuencia completa de pasos a realizar para asegurar que los datos conduzcan a conclusiones válidas con respecto al problema establecido.

DISEÑOS EXPERIMENTALES

SURGEN PREGUNTAS COMO:

¿Cómo se va a medir el efecto?,
¿Cuáles son las características a analizar?
¿Qué otras características afectan a las características que se van a analizar?
¿Cuántas veces deberá ejecutarse el experimento?
¿Cuál será la forma de análisis?

PROPÓSITOS DE UN DEE

- Proporcionar métodos que permitan obtener *la mayor cantidad* de información en forma objetiva, confiable y *al mínimo costo* (dinero, personal, tiempo, etc.).

- *Controlar la variabilidad* que está presente en todo experimento y que afecta a sus resultados.

DEFINICIONES

UNIDAD EXPERIMENTAL

Es la parte más pequeña de material experimental a la que se le aplica el tratamiento y en donde se mide o registra la variable que se investiga.

La unidad experimental por lo general esta conformada por:

-en los experimentos pecuarios *un animal* (vaca, cerdo, pato, etc.)

- en los experimentos forestales, la mayoría de los casos esta conformada por *un árbol*

-en experimentos agrícolos suele ser *una parcela* de tierra en lugar de una planta individual.

TRATAMIENTOS Y FACTORES

- **Tratamiento:** Conjunto de condiciones experimentales o procedimientos que se van a aplicar a una unidad experimental en un diseño elegido y cuyos efectos van a ser registrados (respuesta).

- **Observación:** valor que asume la variable respuesta en una determinada realización.

TRATAMIENTOS Y FACTORES

- **Factor:** Variable controlada por el experimentador o variable independiente. Se estudia su efecto sobre la variable dependiente o respuesta.

- **Nivel del Factor:** es cada una de las categorías, valores o formas específicas del Factor.

Ejemplos

Factor: Tipos de Riego, Dosis del fertilizante, variedades de cultivo, manejo de crianzas, ración alimentaria, profundidad del sembrado, distanciamiento entre plantas, etc. Se acostumbra a simbolizar con A, B, C.

Niveles: Se acostumbra a simbolizar por la letras minúscula del factor y con un subíndice que representan al nivel.

Ejemplo:

A: Tipos de Riego: Goteo Aspersión Secano

Niveles **:** a_0 a_1 a_2

Tipos de Factores

1.-Factores **Cuantitativos.** Son aquellos factores cuyos niveles son cantidades numéricas.

Ejemplo:

Factor A: Dosis de Fertilización**.**

Niveles : 10 kg/Ha ($\mathbf{a_0}$), 20kg/Ha($\mathbf{a_1}$), 30kg/Ha($\mathbf{a_2}$)

2.- Factores **Cualitativos.** Son aquellos factores cuyos niveles son procedimientos o cualidades.

Ejemplo:

Factor A: Variedad de Cultivo

Niveles: Variedad 1($\mathbf{a_0}$), Variedad 2($\mathbf{a_1}$).

Experimentos unifactoriales y multifactoriales

- Unifactorial: un solo factor con sus respectivos niveles. Cada nivel del factor es un tratamiento.
- Multifactorial: dos o más factores.

Un tratamiento es la combinación de un nivel de cada factor.

B

A

	B1	B2	B3
A1	17 20 19 21	15 17 21 18	19 24 25 23
A2	16 19 20 18	15 12 12 14	20 20 24 26
A3	22 20 25 23	16 14 15 15	9 9 12 13

PRINCIPIOS BÁSICOS DEL DISEÑO EXPERIMENTAL

- **Repetición**
- **Aleatorización**
- **Control Local**

REPETICIÓN

Es la *reproducción* o *réplica* del experimento básico (asignación de un tratamiento a una unidad experimental). Son observaciones de un mismo tratamiento en diferentes unidades experimentales.

Las principales razones por las cuales es deseable la repetición son:

- Permite estimar el ***error experimental***
- Aumenta el alcance de la inferencia del experimento por selección y uso apropiado de las u.e.

ALEATORIZACIÓN

Consiste en la *asignación al azar* de los tratamientos a las unidades experimentales con el propósito de asegurar que un determinado tratamiento no presente sesgo.

La aleatorización *hace válidos* los procesos de inferencia estadísticas y los supuestos asegurando conclusiones confiables.

Una de las suposiciones más frecuentes es que las observaciones, o los *errores en ellas están distribuidos independientemente*.

CONTROL LOCAL

(Control del error experimental) consiste en tomar medidas (acciones del experimentador) para hacer el diseño experimental más eficiente, de tal manera que pueda *permitir la reducción del error experimental* y así hacerla más sensible a cualquier prueba de significación.

El error experimental puede reflejar variación del material experimental, errores de medición u observación, errores de experimentación, etc.

CONTROL LOCAL

El error experimental se puede reducir mediante:

♦El uso de un diseño experimental apropiado.

♦ La selección minuciosa del material a usar (lo más homogéneo posible).

♦ El incremento del número de repeticiones en el experimento.

♦ El perfeccionamiento de la técnica experimental y el mayor cuidado al dirigir el experimento.

♦ La utilización de la información proporcionada por variables relacionadas a la variable en estudio (uso de la técnica de covarianza).

TIPOS DE DISEÑOS

- Diseño Completamente Aleatorizado **(DCA)**
- Diseño en Bloques completos aleatorizados **(DBCA)**
- Diseño en Cuadrado Latino **(DCL)**
- Diseño en Parcelas Divididas **(DPD)**
- Otros

Diseño Completamente Aleatorizado (D.C.A)

En este diseño, los tratamientos en estudio se distribuyen al azar en forma ***irrestricta*** sobre todas las unidades experimentales; siendo el número de repeticiones por tratamiento igual ó diferente.

Este diseño se emplea cuando la variabilidad en todo el material experimental es relativamente pequeño y uniformemente distribuido.

Ventajas y Desventajas

- ***Ventajas:***

Fácil de planear y analizar; además es flexible en el empleo del número de tratamientos y repeticiones. Finalmente, permite tener dentro del análisis de varianza el máximo número de grados de libertad para la suma de cuadrados del error.

- ***Desventaja:***

La principal desventaja que presenta este diseño está relacionado a la homogeneidad del material experimental; el cual es difícil de encontrar en experimentos de campo, por lo que su uso se restringe con mucha frecuencia a experimentos de laboratorio, ó donde se pueda tener control de los efectos no considerados en el estudio (ambiente, temperatura, luz, etc.).

Datos

Grupo 1	Grupo 2	Grupo 3	…	Grupo t
Y_{11}	Y_{21}	Y_{31}	…	Y_{t1}
Y_{12}	Y_{22}	Y_{32}	…	Y_{t2}
Y_{13}	Y_{23}	Y_{33}	…	Y_{t3}
:	:	:		:
Y_{1n1}	Y_{2n2}	Y_{3n3}	…	Y_{tnk}

El Grupo 1 tiene n_1 observaciones, el Grupo 2 tiene n_2 observaciones, y así sucesivamente.

Comenzaremos con el caso en que todos los tratamientos tienen igual cantidad de repeticiones. (Diseño balanceado) $n_1 = n_2 = ... = n_t = r$

DCA Fijo o Modelo I

$$Y_{ij} = \mu_i + \varepsilon_{ij} \quad i = 1,\dots,t; \quad j = 1,\dots,r$$

Donde:

Y_{ij}: Es el valor observado de la *j-ésima* observación del *i-ésimo* tratamiento.

μ_i : Es el valor medio poblacional del *i-ésimo* tratamiento.

ε_{ij}: Es el error aleatorio de la *j-ésima* observación correspondiente al i-ésimo tratamiento.

O si en el modelo anterior denotamos a

$$\mu_i = \mu + \tau_i \quad i = 1,\dots,t$$

se obtiene la siguiente forma alternativa del modelo

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad i = 1,\dots,t; \quad j = 1,\dots,r$$

Donde:

Y_{ij}: Es el valor observado de la *j-ésima* observación que recibió el *i-ésimo* tratamiento.

μ : Es el valor medio poblacional general.

τ_i: Es el efecto medio poblacional del *i-ésimo* tratamiento.

ε_{ij}: Es el error aleatorio de la *j-ésima* observación correspondiente al i-ésimo tratamiento.

SUPUESTOS para el modelo $Y_{ij} = \mu_i + \varepsilon_{ij} \quad i = 1,\dots,t; \quad j = 1,\dots,r$

1. Para cada tratamiento la variable aleatoria respuesta Y_{ij} se distribuye normalmente . Para cada T_i

$$Y_{ij} : N(\mu_i ; \sigma_i^2)$$

2. Las Y_{ij} son variables aleatorias independiente entre si

$$\operatorname{cov}(Y_{ij} Y_{kl}) = 0 \quad \forall i \neq k; \quad \forall j \neq l$$

3. Las varianzas de los t tratamientos son iguales (homoscedasticidad)

$$\sigma_1^2 = \sigma_2^2 = \dots = \sigma_t^2 = \sigma^2$$

En consecuencia

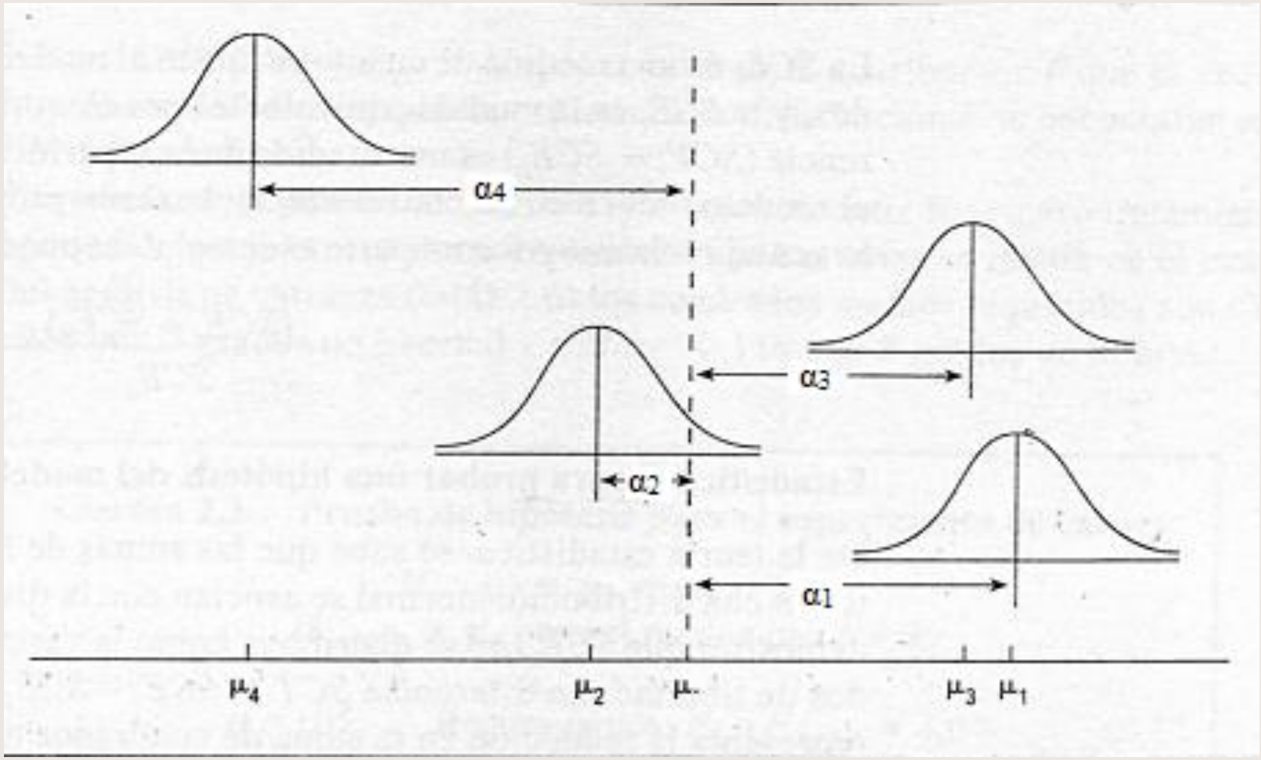
$$\varepsilon_{ij} : N(0; \sigma^2)$$

$$E(y_{ij}) = \mu_i$$

$$Var(Y_{ij}) = \sigma^2$$

DISEÑO COMPLETAMENTE ALEATORIZADO (D.C.A)

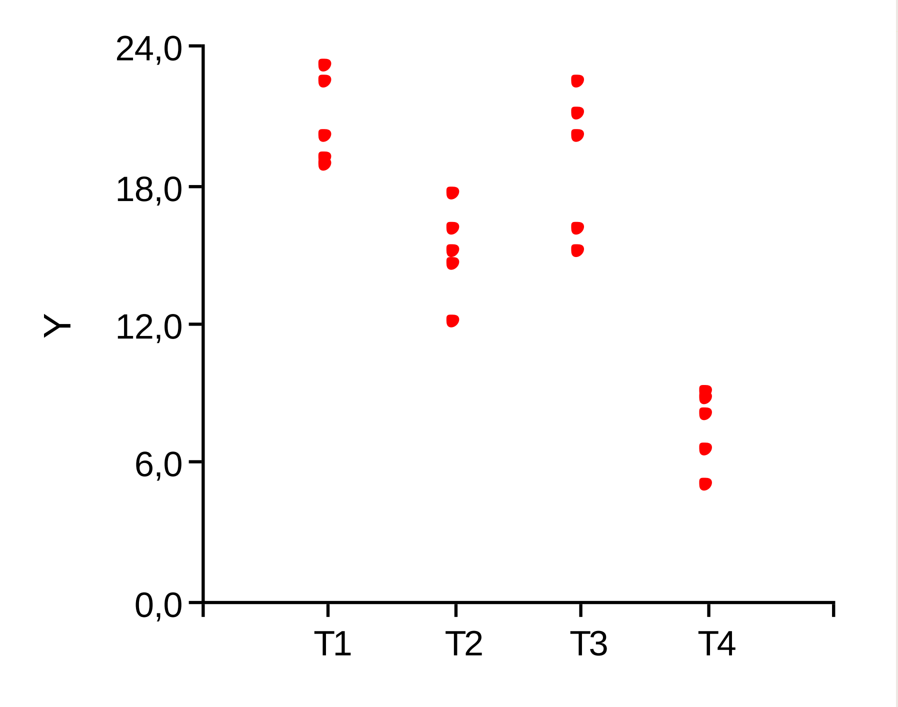
En teoría tenemos las distribuciones correspondientes a los tratamientos para una determinada variable respuesta

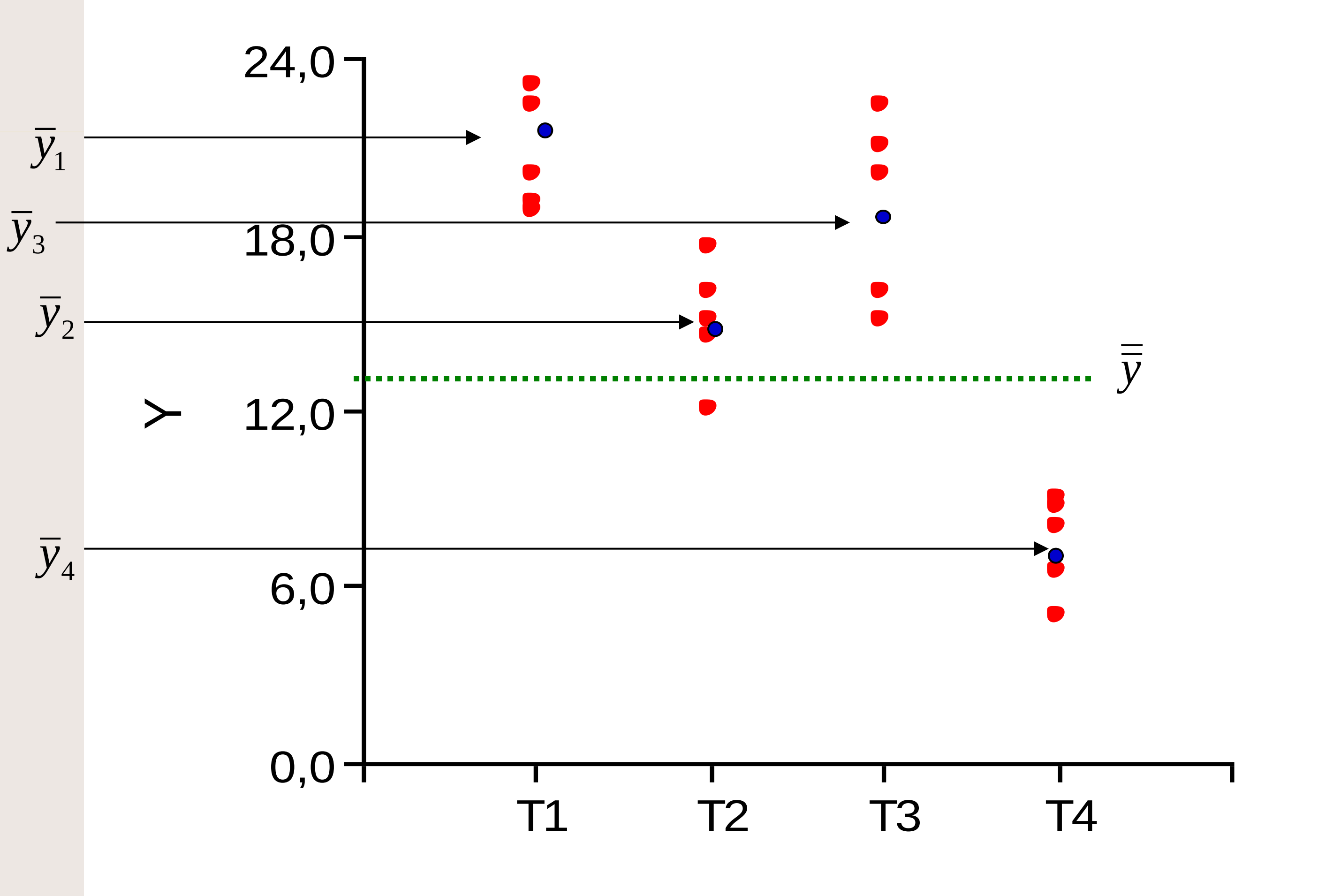


$$\tau_i = \alpha_i = \mu_i - \mu$$

DISEÑO COMPLETAMENTE ALEATORIZADO (D.C.A)

En la práctica sólo tenemos las muestras correspondientes a los tratamientos y sus correspondientes valores respuesta





24,0
18,0
12,0
6,0
0,0
Y
T1
T2
T3
T4
$\bar{y}_1$
$\bar{y}_3$
$\bar{y}_2$
$\bar{y}_4$
$\bar{\bar{y}}$

Partición de la suma de cuadrados

$$\sum_i \sum_j (y_{ij} - \bar{\bar{y}})^2 = \sum_i \sum_j (y_{ij} - \bar{y}_i)^2 + \sum_i \sum_j (\bar{y}_i - \bar{\bar{y}})^2$$

Estas sumas de cuadrados se llaman respectivamente

$$SC_{tot} = SC_{ee} + SC_{Trat}$$

o

$$SC_{tot} = SC_{dentro} + SC_{entre}$$

y tienen distribución Chi cuadrado y son independientes por lo tanto los grados de libertad son:

SC_{tot}: r*t – 1

SC_{ee}: t(r – 1)

SC_{Trat}: (t – 1)

Dividiendo la suma de cuadrados por sus respectivos grados de libertad se obtienen los *cuadrados medios* que son varianzas muestrales.

$$CM_{error} = \frac{SC_{error}}{t(r-1)} = \hat{\sigma}^2 \qquad CM_{trat} = \frac{SC_{trat}}{t-1}$$

Estas varianzas muestrales son estimadores de varianzas poblacionales (parámetros) que están relacionadas con las fuentes de variación correspondientes

$$E(CM_{ee}) = \sigma^2 \qquad E(CM_{trat.}) = \sigma^2 + r\,\frac{\sum_{i=1}^{t} \tau_i^{\,2}}{t-1}$$

El cociente entre éstos tiene distribución F

$$F = \frac{CM_{trat}}{CM_{ee}} \approx F_{(t-1;\,t(r-1))}$$

Tabla ANOVA

Fuentes Variación	Sumas Cuadrados	g.l	Cuadrados Medios	*F*
Entre Trat.	SCTrat.	t-1	SCTrat/g.lTr (I)	*Fo*= I/II
Dentro Trat. (EE)	SCError	t(r-1)	SCError/g.lee (II)	
Total	SCTotal	tr - 1		

Hipótesis en términos de los efectos

$H_0 : \tau_i = 0$ $i = 1,2, ..., t$ (Todos los tratamientos tienen el mismo efecto sobre la variable rta.)
H_1: algún $\tau_i \neq 0$ (No todos los tratamientos tienen el mismo efecto sobre la variable en estudio).

Dócimas en términos de los promedios

H_0: $\mu_1 = \mu_2 = ...= \mu_t = \mu$
H_1: algún $\mu_i \neq \mu$.

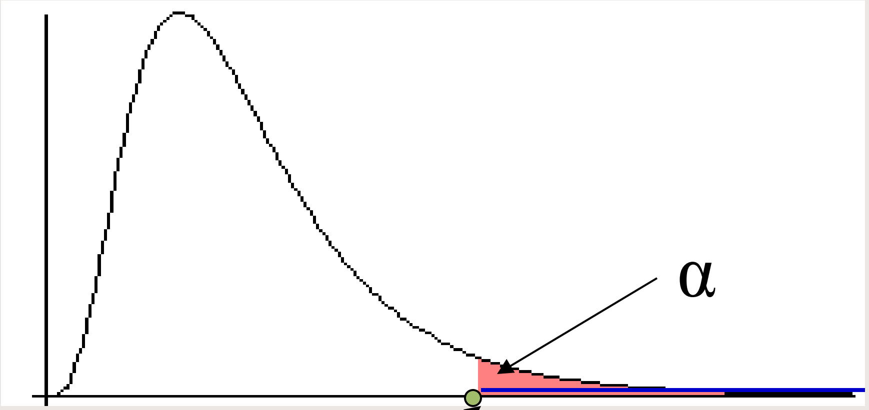
PASOS si se realiza a mano la dócima en el DCA

- **Paso Nº 3 Variable pivotal :**

$$F = \frac{CM_{trat}}{CM_{ee}} \approx F_{(t-1;t(r-1))}$$

- **Paso Nº 4 Región Crítica y regla de decisión:**

La región crítica $F_{(t-1;\ t(r-1))} \geq F_{crit}$



$F_{crit} = F_{(t-1;\ t(r-1));1-\alpha}$

- **Regla de decisión**

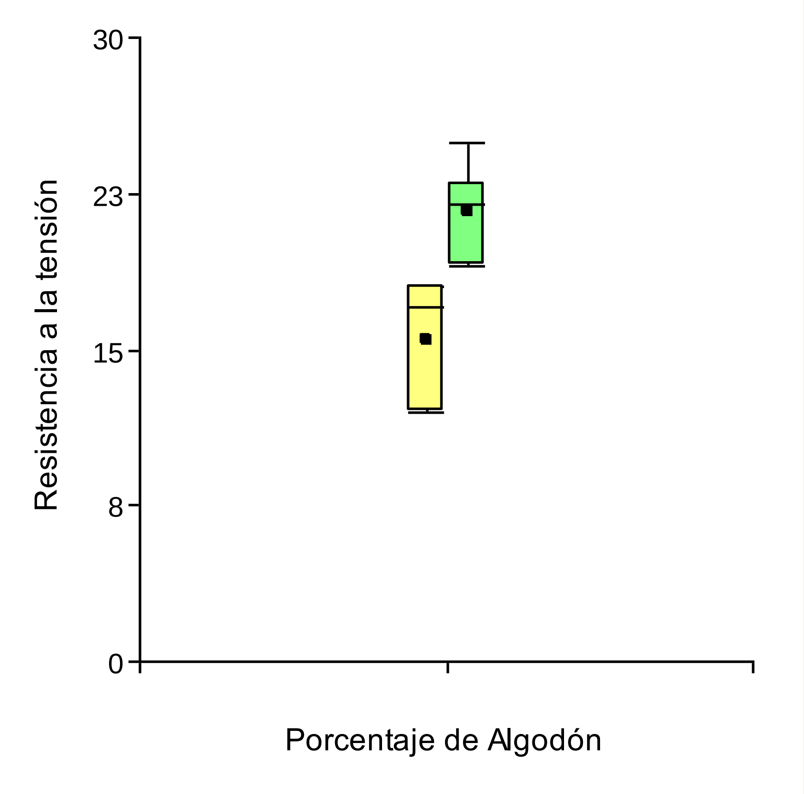
Rechazo H0 si Fcalc $\geq F_{crit;}$ no Rechazo H0 si Fcalc $< F_{crit}$

Un ingeniero de desarrollo de productos está interesado en maximizar la resistencia a la tensión de una nueva fibra sintética que se empleará en la manufactura de tela para camisas de hombre.

El ingeniero sabe por experiencia que la resistencia es influida por el porcentaje de algodón presente en la fibra. Además, él sospecha que elevar el contenido de algodón incrementará la resistencia, al menos inicialmente. También sabe que el contenido de algodón incrementará la resistencia al menos inicialmente. También sabe que el contenido del algodón debe variar aproximadamente entre 20 y 40 % para que la tela resultante tenga otras características de calidad que se desean (como capacidad de recibir un tratamiento de planchado permanente).

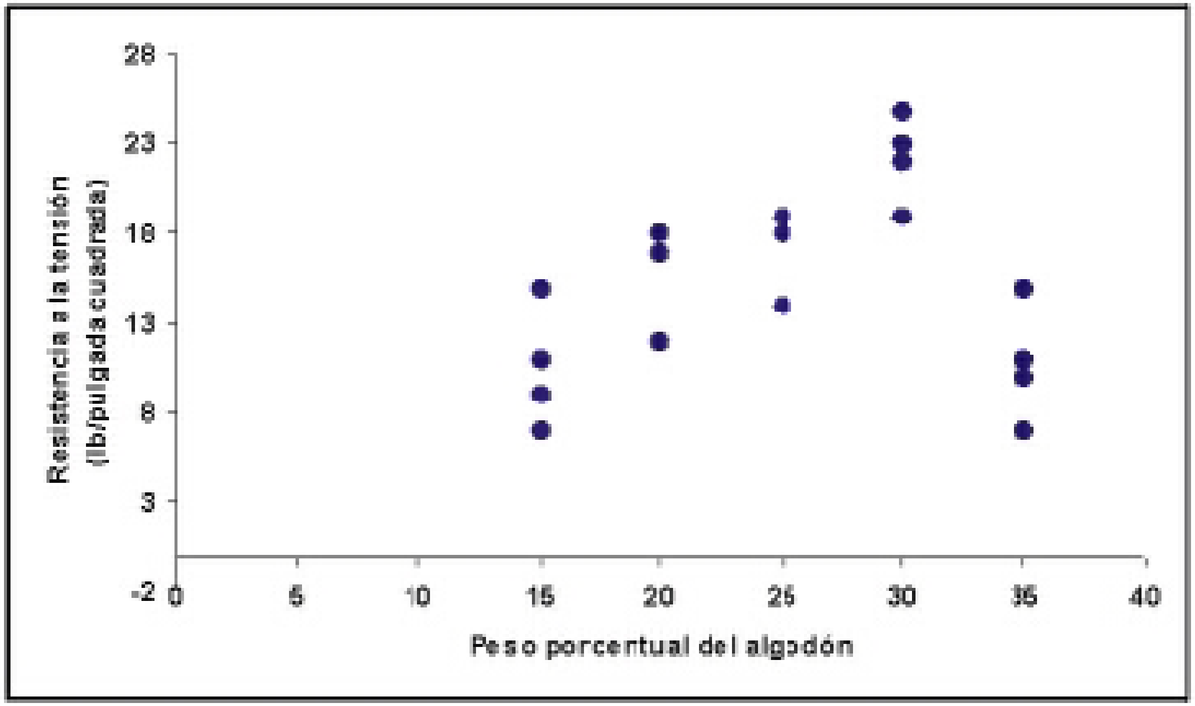
El ingeniero decide probar muestras a dos niveles de porcentaje de algodón: 20% y 30%

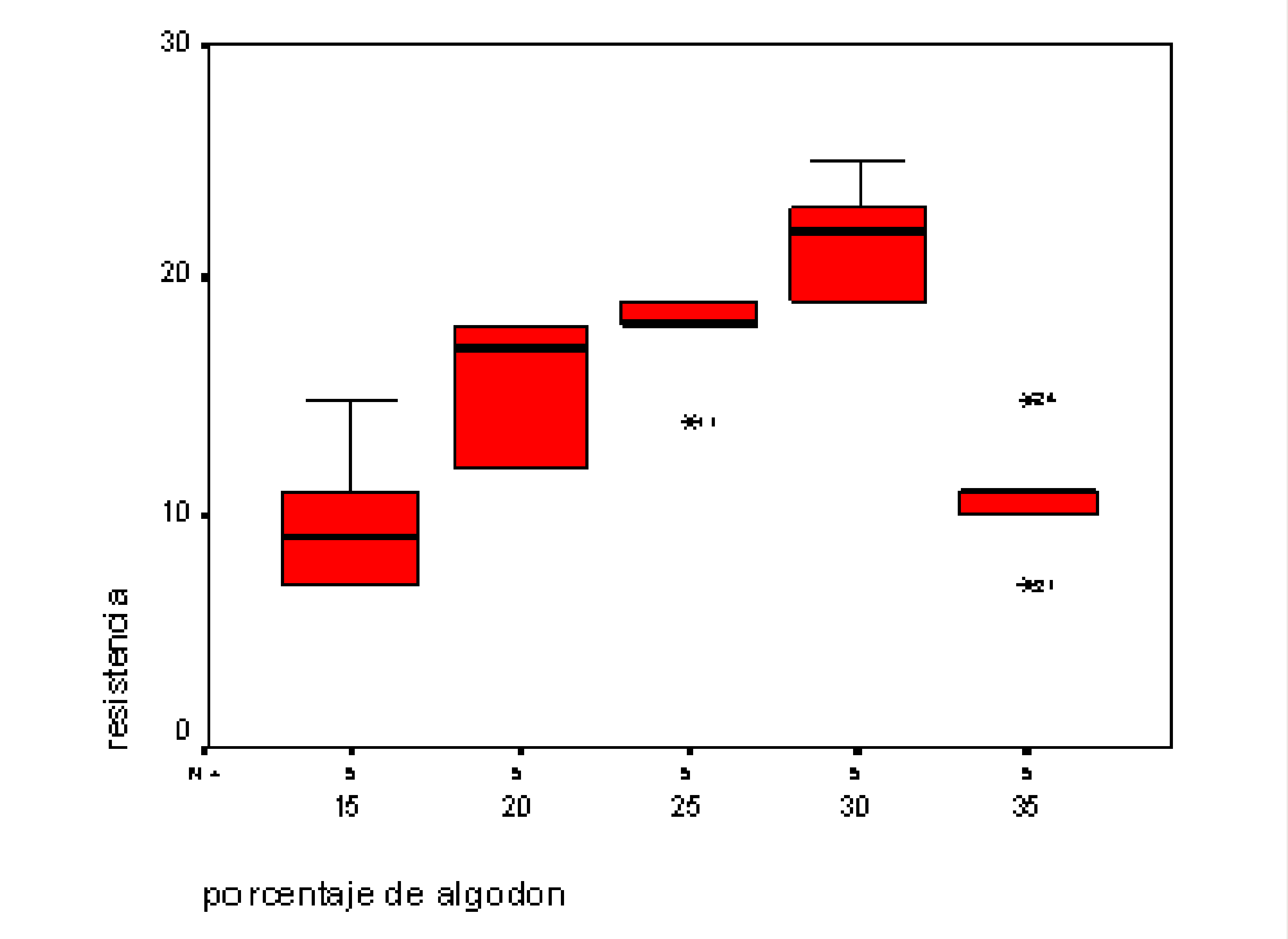
Algodón al 20%	Algodón al 30%
12	19
17	25
12	22
18	19
18	23



El ingeniero decide probar muestras a cinco niveles de porcentaje de algodón: 15, 20, 25, 30 y 35 %. Así mismo, decide ensayar cinco muestras a cada nivel del contenido de algodón.

Porcentaje de algodón	1	2	3	4	5
15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11




30
20
10
0
resistencia
N =
5
5
5
5
5
15
20
25
30
35
porcentaje de algodon