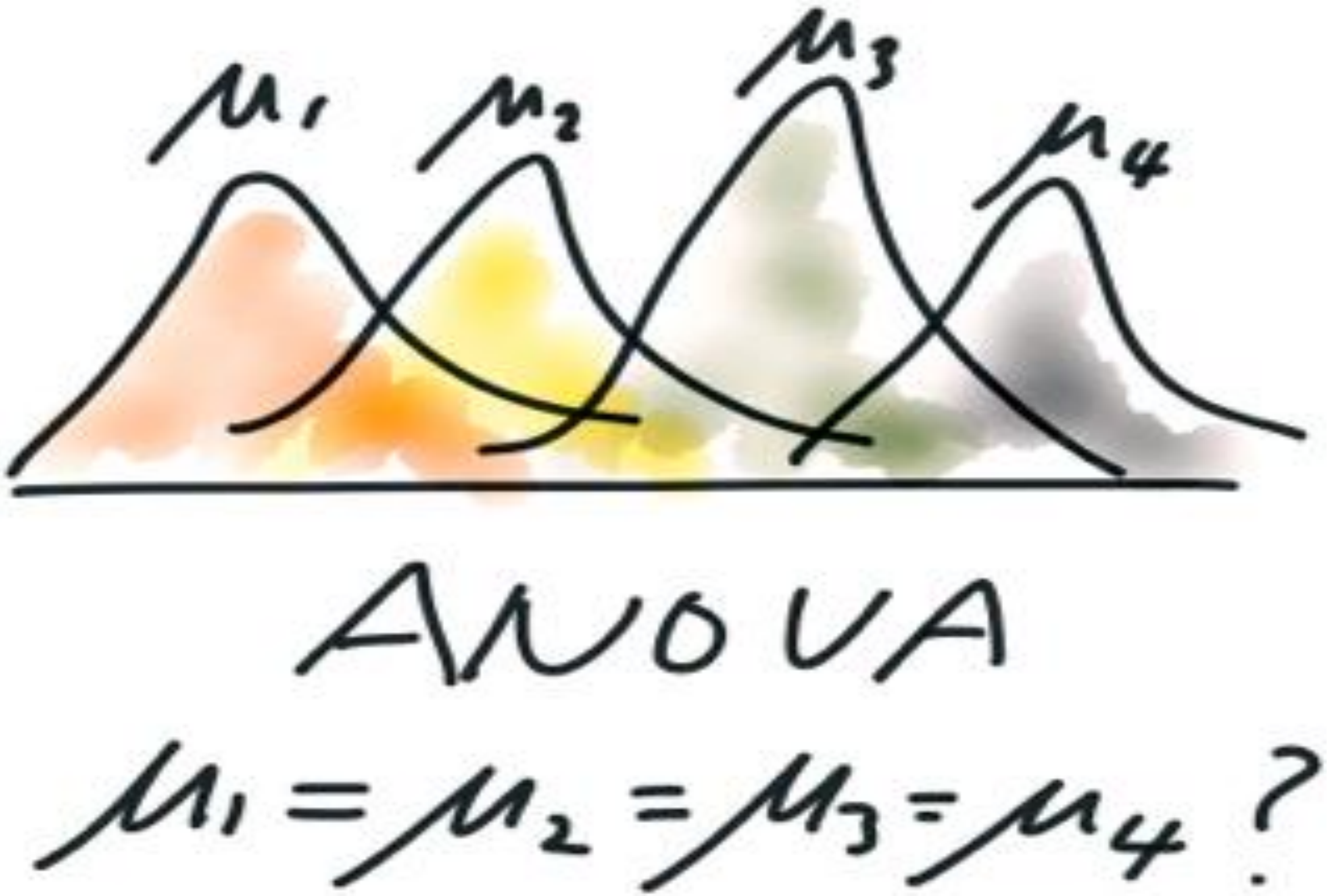
Análisis de variancia

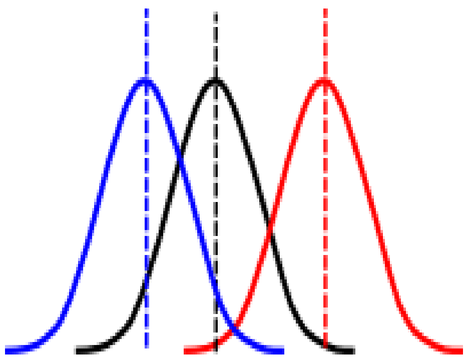


ANOVA

Oscar Centeno Mora

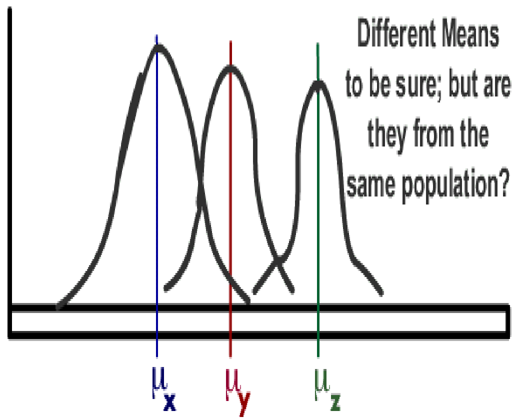
Introducción

- En los temas anteriores se estudiaron los conceptos de prueba de hipótesis y comparación de medias, los cuales son la base para el presente tema.



- Se estudiara el análisis de variancia, el cual se define como

"una técnica en la que la variancia total de un conjunto de datos se divide en dos o más componentes, y cada uno de ellos se asocia con una fuente específica de variación, de manera que durante el análisis es posible encontrar la magnitud con la que contribuye cada una de esas fuentes en la variancia total "



- El desarrollo del análisis de variancia (ANOVA en inglés) se debe sobre todo al trabajo de R.A. Fisher.

Índice

Índice

Fundamentos del ANOVA

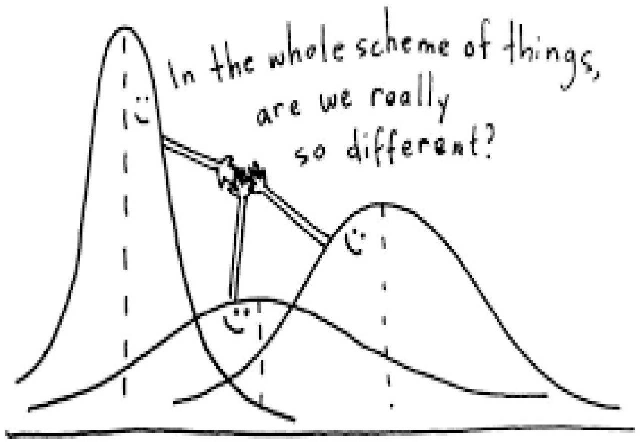
- El Análisis de variancia (ANOVA), se origina en los análisis derivados de experimentos.

- El ANOVA se utiliza para cumplir dos objetivos

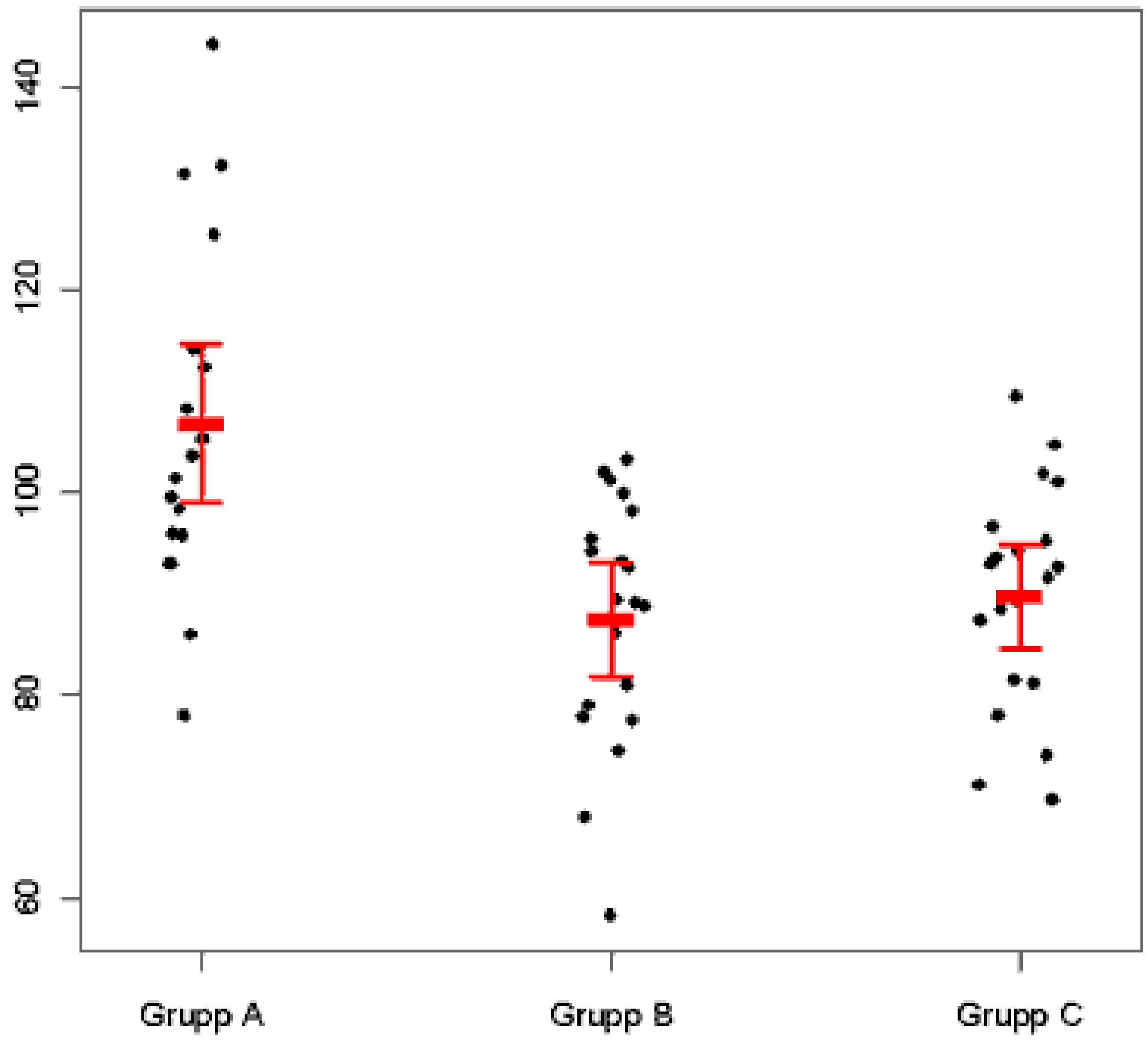
    1. Estimar y probar hipótesis respecto a las variancias
    2. Estimar y probar hipótesis respecto a las medias

- Se estudia únicamente las hipótesis respecto a las medias.

- En los temas anteriores se probó como hacer PH bajo la comparación de 1 ó 2 medias. El ANOVA se aplica para más de dos medias.



Fundamentos del ANOVA

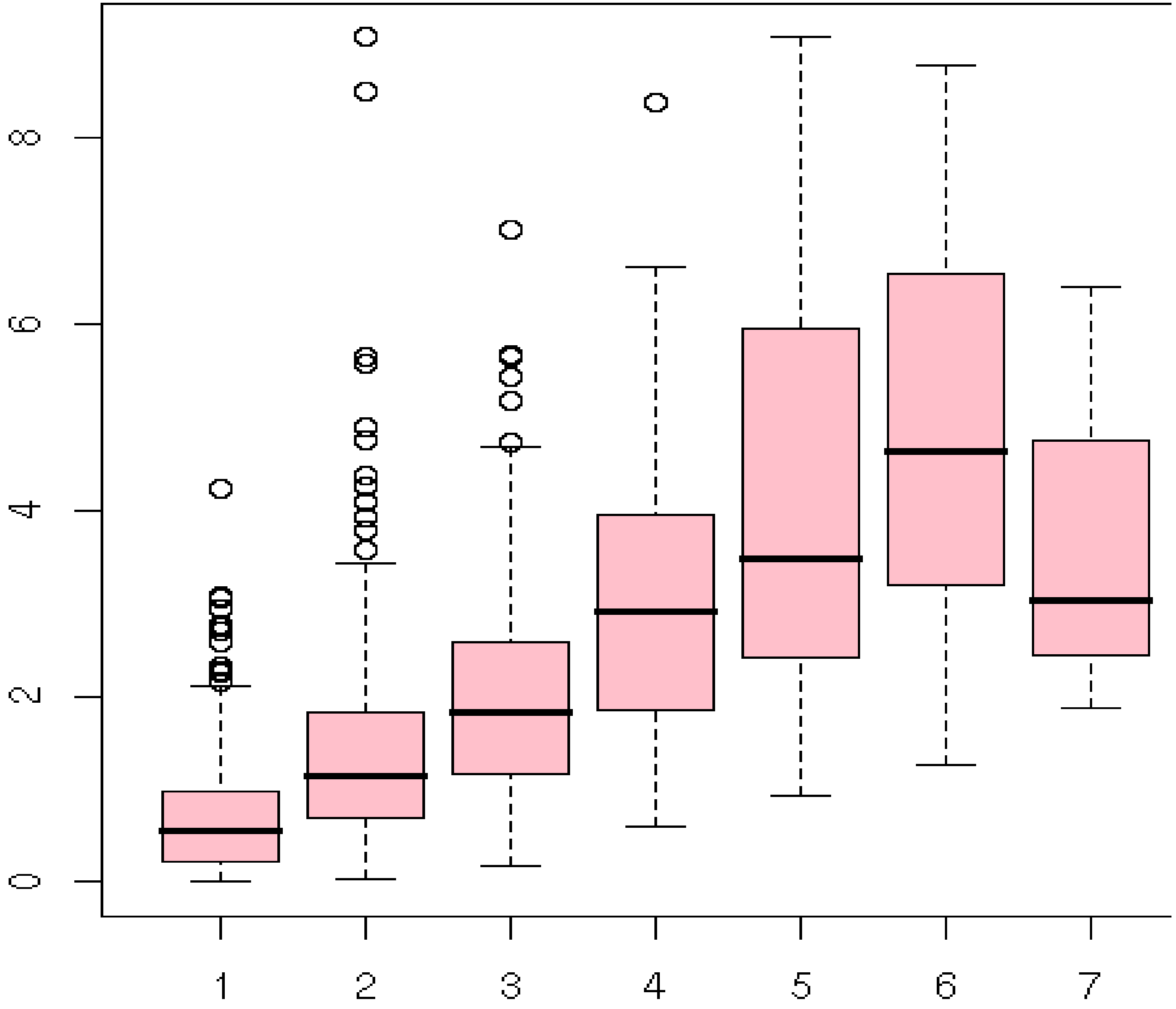


Se tienen 3 poblaciones

Se quiere saber si hay diferencia a nivel de las medias

Se debe contemplar la variabilidad….

Fundamentos del ANOVA



Se tienen 7 poblaciones

Se quiere saber si hay diferencia a nivel de las medias

Se debe contemplar igual la variabilidad….

Índice

Etapas de un ANOVA

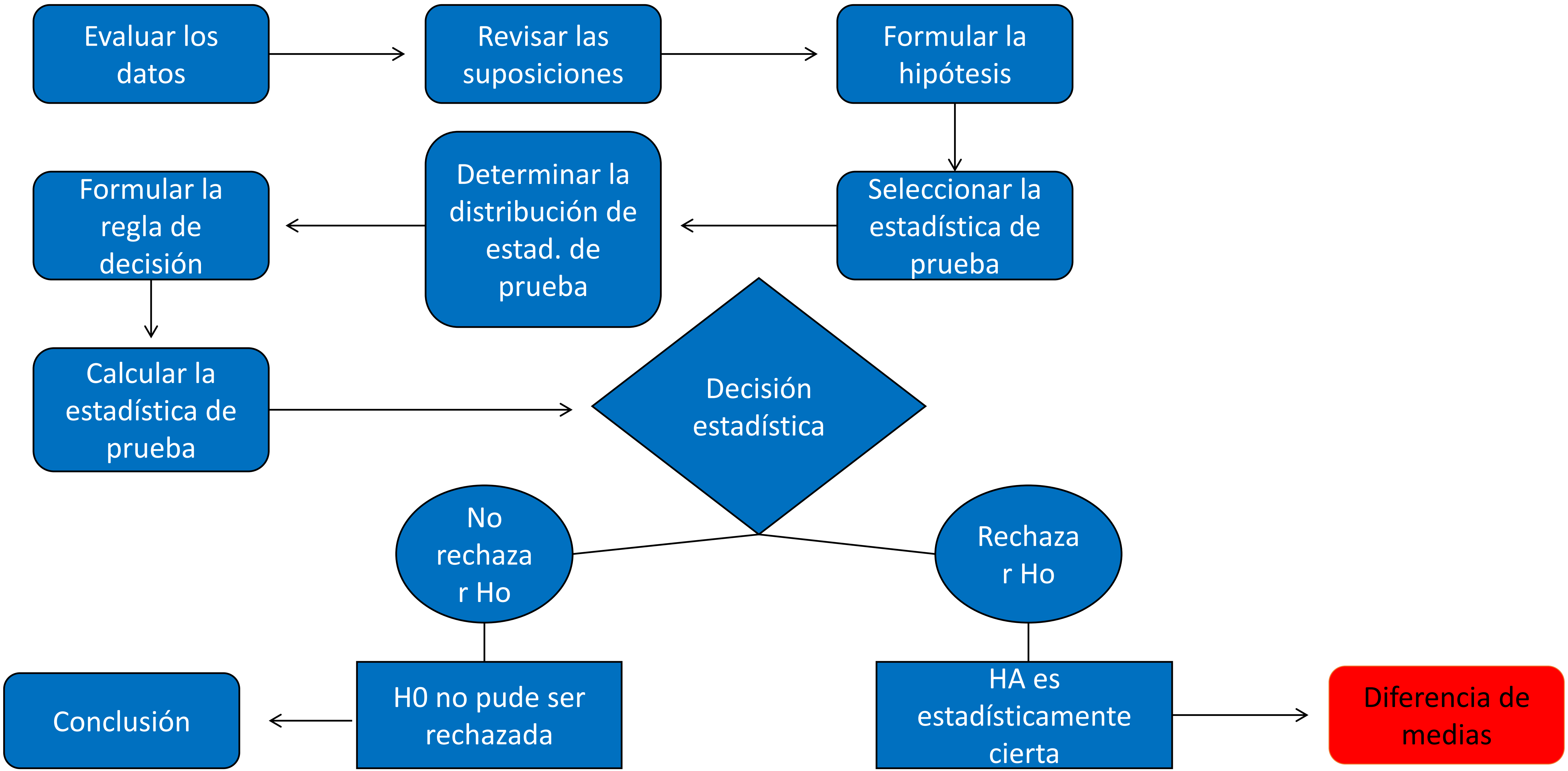
- En la presentación del análisis de variancia , se siguen los mismos pasos vistos para la prueba de hipótesis...

- …más uno adicional si se rechaza la hipótesis de igualdad de medias

- Se vuelve a presentar el esquema para realizar una prueba de hipótesis, dado que debemos entregar nuevos conceptos, necesarios para adaptar el procedimiento al análisis de variancia.

- Las diferencias más importantes se encuentran a nivel de la descripción de los datos, el cálculo de la estadística de prueba, el análisis de las diferencias, y sobre todo, el establecimiento de un modelo lineal…

Recordatorio: Pasos de una prueba de hipótesis: resumen



Índice

Diseño de una vía: fundamentos

- A menudo se desea probar una hipótesis para mas de dos poblaciones.

- Se requiere un método para hallar una diferencias significativa entre varias medias, sin que el nivel de confianza de la prueba se vea afectado.

- El análisis de variancias constituye una solución para lo anterior.

- El análisis de la variancia unilateral, es el más sencillo y en el cuál se estudia una sola fuente de variación o factor.

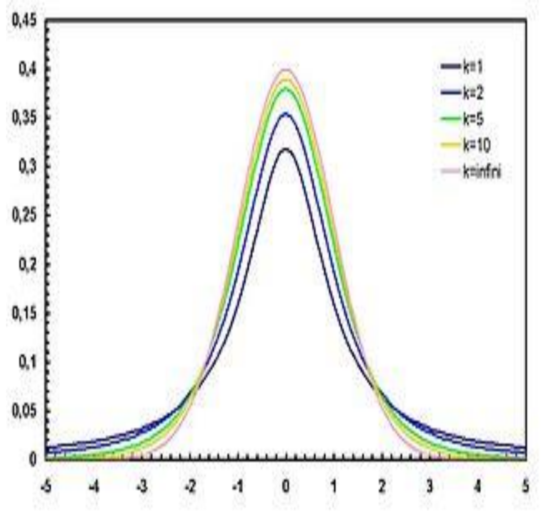
Diseño de una vía: fundamentos

- Es una extensión del procedimiento de la prueba t, para utilizar en dos muestras independientes.

- Utilizar la prueba t con dos muestras independientes es un caso específico del análisis de variancia unilateral.

- Se parte que los individuos pertenecientes a cada uno de los población, corresponde a población independientes.

- Desde un punto de vista experimental, se dice que en vez de poblaciones se poseen tratamientos (factor), y que estos se asignan de forma aleatoria a las personas. Por lo que el tratamiento sería visto como la referencia de las poblaciones.



Diseño de una vía: descripción de los datos

- Las mediciones (u observaciones) que resultan de un análisis de este tipo , junto con las medias y los totales que puedan calcularse a partir de los datos, pueden presentarse de acuerdo a un cuadro donde los factores o tratamientos se presentan en columnas.

- En importante saber determinar que es:
  - La observación x_{ij}
  - El total por factor o tratamiento $T_{.j}$
  - La observación para el j-ésimo factor $\bar{x}_{.j}$
  - Total de todas las observaciones $T_{..}$
  - El promedio general $\bar{x}_{..}$

- Nótese la importancia de conocer la procedencia de la observación de acuerdo al factor.

Diseño de una vía: descripción de los datos

muestra para el diseño complementamente aleatorizado

	Tratamientos					
	1	2	3	...	k	
	x_{11}	x_{12}	x_{13}	...	x_{1k}	
	x_{21}	x_{22}	x_{23}	...	x_{2k}	
	x_{31}	x_{32}	x_{33}	...	x_{3k}	
	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
	$x_{n_1 1}$	$x_{n_2 2}$	$x_{n_3 3}$	...	$x_{n_k k}$	
Total	$T_{.1}$	$T_{.2}$	$T_{.3}$	...	$T_{.k}$	$T_{..}$
Media	$\bar{x}_{.1}$	$\bar{x}_{.2}$	$\bar{x}_{.3}$	...	$\bar{x}_{.k}$	$\bar{x}_{..}$

$$\boldsymbol{x_{ij} = la\ i - ésima\ observación\ a\ partir\ del\ j - ésimo\ factor}$$

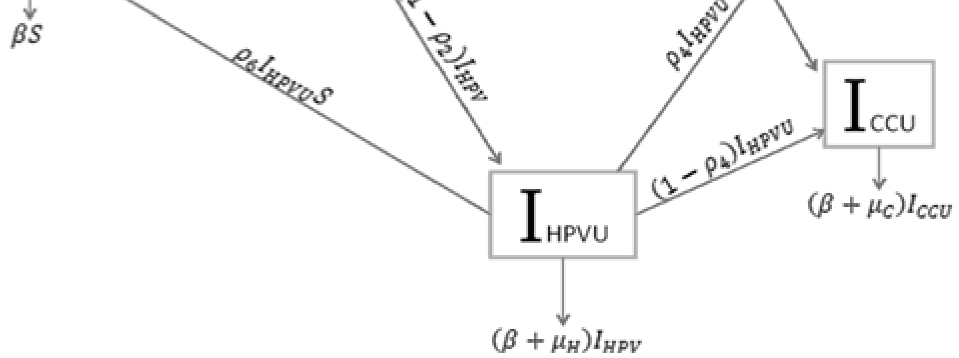
$$T_{.j} = \sum x_{i.} = total\ de\ j - ésimo\ factor$$

$$T_{.j} = \sum x_{i.} = total\ de\ j - ésimo\ factor$$

$$\bar{x}_{.j} = \frac{T_{.j}}{n_j} = medida\ del\ j - ésimo\ factor$$

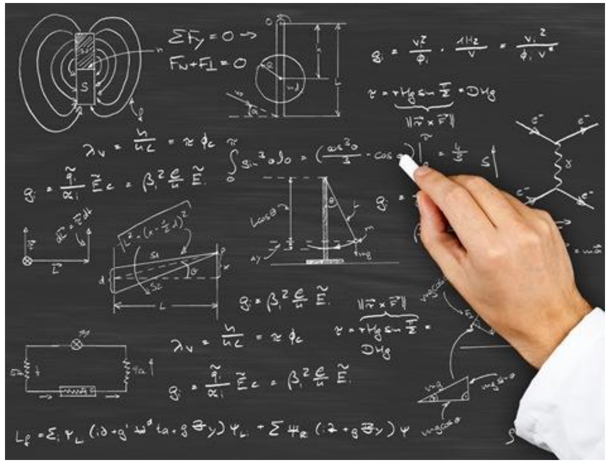
$$T_{..} = \sum T_{.j} = \sum\sum x_{ij} = total\ de\ las\ observaciones$$

$$\bar{x}_{..} = \frac{T_{..}}{N} = promedio\ general$$



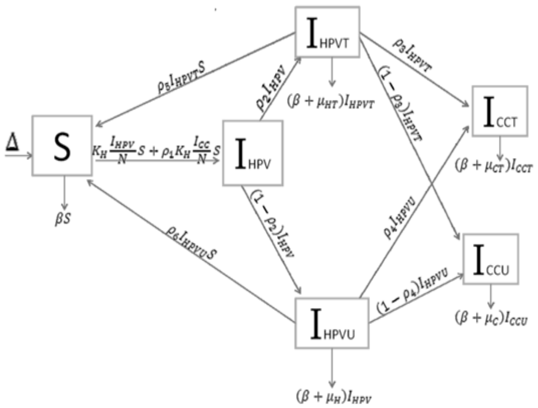
una vía: supuestos

blecen a través de ciertas características del modelo. Estas son:



1. El modelo
2. Componentes del modelo

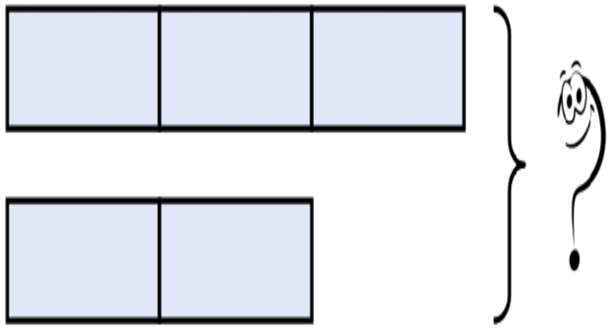
- Antes de pasar a los supuestos del modelo se debe extender e introducir el modelo, además de la significación de cada uno de sus componentes.



- Esto se presente a continuación…

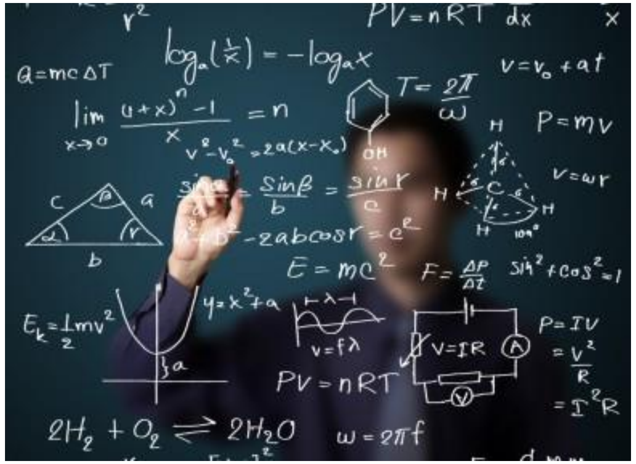
Diseño de una vía: supuestos – el modelo

- En los casos anteriores las pruebas de hipótesis no habían requerido establecer un modelo lineal, aunque estos si lo poseían….



- Un modelo es una representación simbólica de un valor representativo del conjunto de datos.

- Para escribirlo, se debe recordar los valor visto en el cuadro correspondiente a los anteriormente. En todos los casos, el modelo se expresa siempre en base a la observación $\boldsymbol{x}_{\mathbf{ij}}$



- Recordar que el modelo debe contener la expresión individual como la expresión según el factor de interés.

Diseño de una vía: supuestos – el modelo

- El modelo se expresa según la siguiente ecuación:

$$\boldsymbol{x_{ij}} = \boldsymbol{\mu} + \boldsymbol{\tau_i} + \boldsymbol{e_{ij}};$$

$$i = 1,2,\dots,n_j$$
$$j = 1,2,\dots,k$$

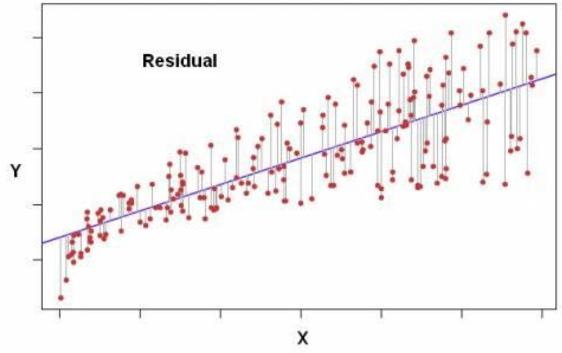
$$\mu = \frac{\sum X}{N}$$

- Los términos en este modelo se definen de la siguiente forma:

1. $\boldsymbol{\mu}$ representa la media general, se le conoce como la media general.

2. $\boldsymbol{\tau_i}$ es la diferencia entre las j-médias, y también se conoce como efecto del tratamiento o del factor.

3. $\boldsymbol{e_{ij}}$ representa la cantidad en que difieren las medidas estimadas y observadas. Se le conoce como término del error.



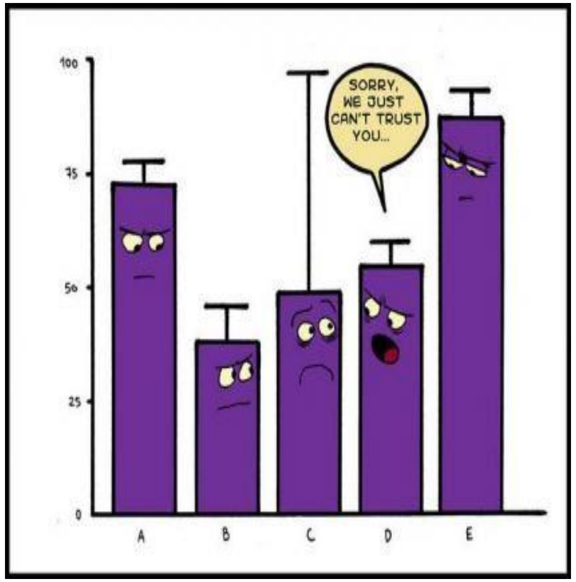
Diseño de una vía: supuestos – componentes

- Al revisar el modelo se aprecia que una observación está compuesta por:

1. La media general
2. El efecto del tratamiento
3. El término de error que representa la desviación de la observación a partir de su media grupal.

- Los ANOVA de una vía, la mayoría de veces interesa verificar el efecto del tratamiento. Cualquier inferencia se aplica a los tratamiento y no se pretende ampliar la inferencia a otro conjunto mayor de tratamientos.

- La restricción anterior se conoce como modelo de efectos fijos, y el presente capítulo se limita a este tipo de modelo.



Diseño de una vía: supuestos

Los supuestos que posee el ANOVA de una vía son los siguientes:

a) Los k tratamientos forman k muestras aleatorias
b) Cada una de las sub muestras poseen una distribución normal
c) Las variancias se suponen iguales entre los grupos
d) Los efectos $\boldsymbol{\tau_i}$ se consideran como constantes desconocidas, cuando sumadas es igual a 0.
e) Los errores $\boldsymbol{e_{ij}}$ tiene un valor esperado de 0.
f) Los errores $\boldsymbol{e_{ij}}$ poseen una variancia constantes
g) Los errores $\boldsymbol{e_{ij}}$ poseen una distribución normal e independiente.

Diseño de una vía: hipótesis

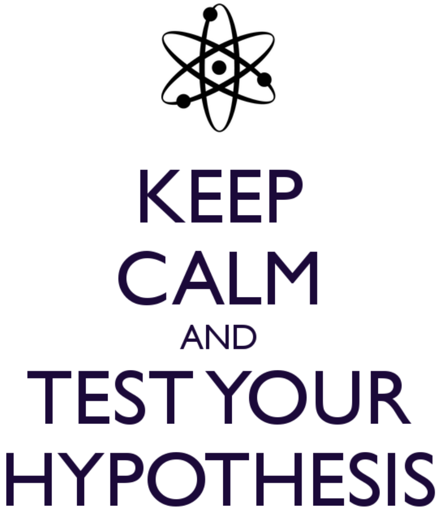
- La hipótesis supone que las poblaciones o tratamientos tienen medias iguales contra la hipótesis alternativa, donde se dice que por lo menos una difiere.

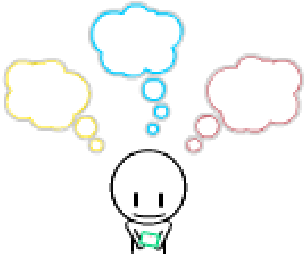
- Esta se expresa como :

$$H_0\colon\ \mu_1 = \mu_2 = \cdots = \mu_k$$
$$H_A\colon \mu_i \neq\ \mu_j$$

- De forma análoga, si las medias de las poblaciones son iguales, y el efecto de cada tratamiento es igual a cero, de tal manera que, alternativamente, las hipótesis pueden escribirse como:

$$H_0\colon\ \tau_j = 0, j = 1{,}2, \ldots k$$
$$H_A\colon no\ todas\ las\ \tau_j = 0$$





Diseño de una vía: estadística de prueba

- La estadística de prueba para el análisis de variancia unilateral es el cálculo de la razón de dos variancias, o la razón de los cuadros medios.

- Para entender esto, se debe analizar el cuadro del análisis ANOVA:

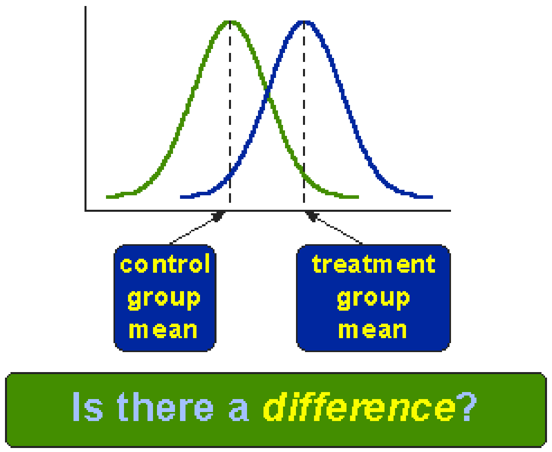
Fuentes de variación	Suma de cuadrados	Grados de libertas	Cuadrado medio	Razón de variación
Entre muestras	SC entre	k-1	CM entre= (SC entre) / (k-1)	R.V= (CM entre) / (CM dentro)
Dentro de los muestras	SC dentro	n-k	CM dentro = (SC dentro) / (n-k)	
Total	SC total	n-1		

Fuentes de variación	Suma de cuadrados	Grados de libertas	Cuadrado medio	Razón de variación
Entre muestras	SC entre	k-1	CM entre= (SC entre) / (k-1)	R.V= (CM entre) / (CM dentro)
Dentro de los muestras	SC dentro	n-k	CM dentro = (SC dentro) / (n-k)	
Total	SC total	n-1		

- Sumas de cuadrados

Diseño de una vía: estadística de prueba – suma de cuadrados

- La suma de cuadrados es la forma en que se descompone las fuentes de variación en un análisis de variancia. Esto es indispensable para la prueba.



- Se posee un total de 3 sumas de cuadrados: suma total de cuadrados, suma de cuadrados dentro de los grupos y suma de cuadrados entre los grupos.

- Se tiene que la suma total de cuadrados es igual a la suma de cuadrados entre grupos o tratamientos más la suma de cuadrados dentro de los grupos o los tratamientos.

$$SS = \sum X^2 - \frac{\left(\sum X\right)^2}{N}$$

- Se expresa como: $SC_{total} = SC_{entre} + SC_{dentro}$

Diseño de una vía: estadística de prueba – suma de cuadrados

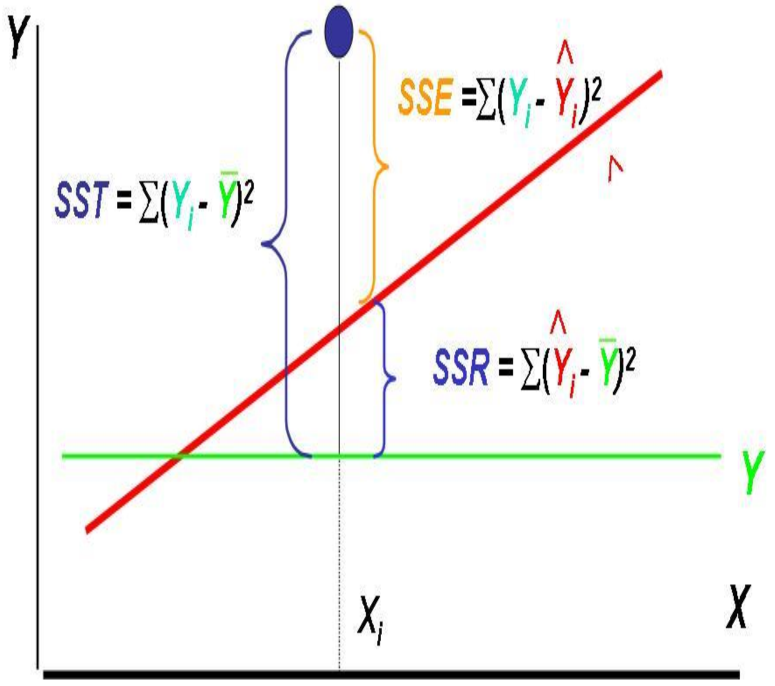
- La suma total de cuadrados, es la suma de los cuadrados de las desviaciones de las observaciones individuales a partir de la media de todas las observaciones tomadas juntas.

- **La suma total de cuadrados** (SC_{total}), se define como:

$$SC_{total} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_{..})^2$$

La $\sum_{j=1}^{k}$ indica la suma de desviciones para los grupos

La $\sum_{i=1}^{n_j}$ indica la suma de los totales por grupos



Diseño de una vía: estadística de prueba – suma de cuadrados

	Tratamiento 1	Tratamiento 2	Tratamiento 3	
	2	12	22	
	4	14	24	
	6	16	26	
	8	18	28	
	10	20	30	
				General
Promedio	**6**	**16**	**26**	**16**
Total	**6**	**6**	**6**	**18**

¿Como se calcula la **suma total de cuadrado ?**

$$SC_{total} = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_{..})^2$$

Diseño de una vía: estadística de prueba – suma de cuadrados

- **La suma de cuadrados dentro de los grupos** implican realizar dentro de cada grupo, la suma de las desviaciones al cuadrado de las observaciones individuales, según la media de interés.

- Después de realizar los cálculos dentro de cada grupo, se obtiene la suma de los resultados individuales del grupo.

- Este componente (suma de cuadrados de los grupos) se designa como SC_{dentro}, y su expresión matemática es:

$$SC_{dentro} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{.j})^2$$

- Esta cantidad se suele denominar como suma de residuales o errores.

Diseño de una vía: estadística de prueba – suma de cuadrados

	Tratamiento 1	Tratamiento 2	Tratamiento 3	
	2	12	22	
	4	14	24	
	6	16	26	
	8	18	28	
	10	20	30	
				General
Promedio	**6**	**16**	**26**	**16**
Total	**6**	**6**	**6**	**18**

¿Como se calcula la **suma dentro de los tratamientos** ?

$$SC_{dentro} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{.j})^2$$

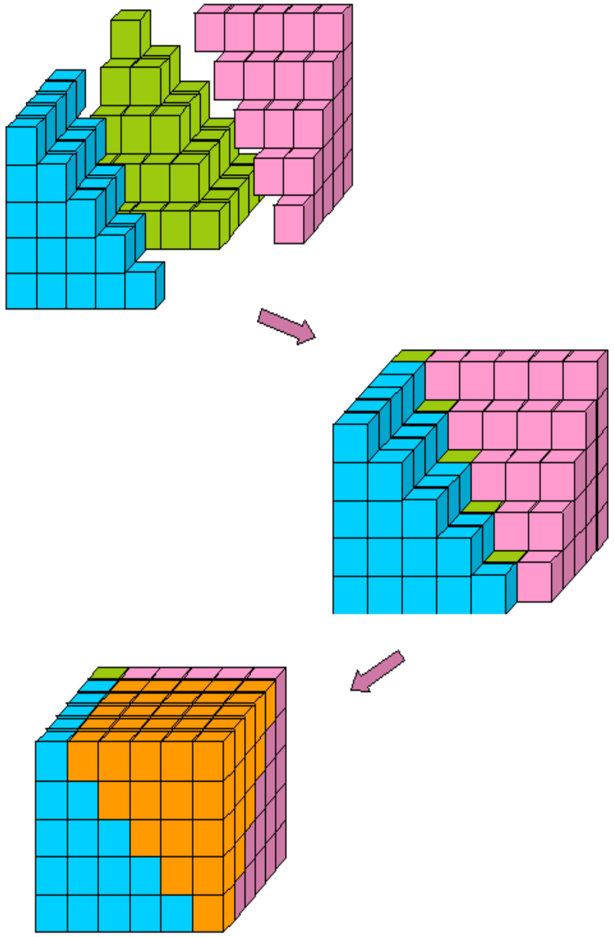
Diseño de una vía: estadística de prueba – suma de cuadrados

- Para obtener el segundo componente de la suma total de cuadrados, se calcula para cada grupo la desviación al cuadrado de la media del grupo a partir de la gran media, y se multiplica el resultado por el tamaño del grupo.

- Finalmente, se suman los resultados de todos los grupos.

- La suma de cuadraros entre los grupos (SC_{entre}) es una medida de la variación entre los grupos y se expresa matemáticamente como:

$$SC_{entre} = \sum_{j=1}^{k} n_j(\bar{x}_{.j} - \bar{x}_{..})^2$$



Diseño de una vía: estadística de prueba – suma de cuadrados

	Tratamiento 1	Tratamiento 2	Tratamiento 3	
	2	12	22	
	4	14	24	
	6	16	26	
	8	18	28	
	10	20	30	
				General
Promedio	**6**	**16**	**26**	**16**
Total	**6**	**6**	**6**	**18**

¿Como se calcula la **suma de cuadrados entre los tratamientos** ?

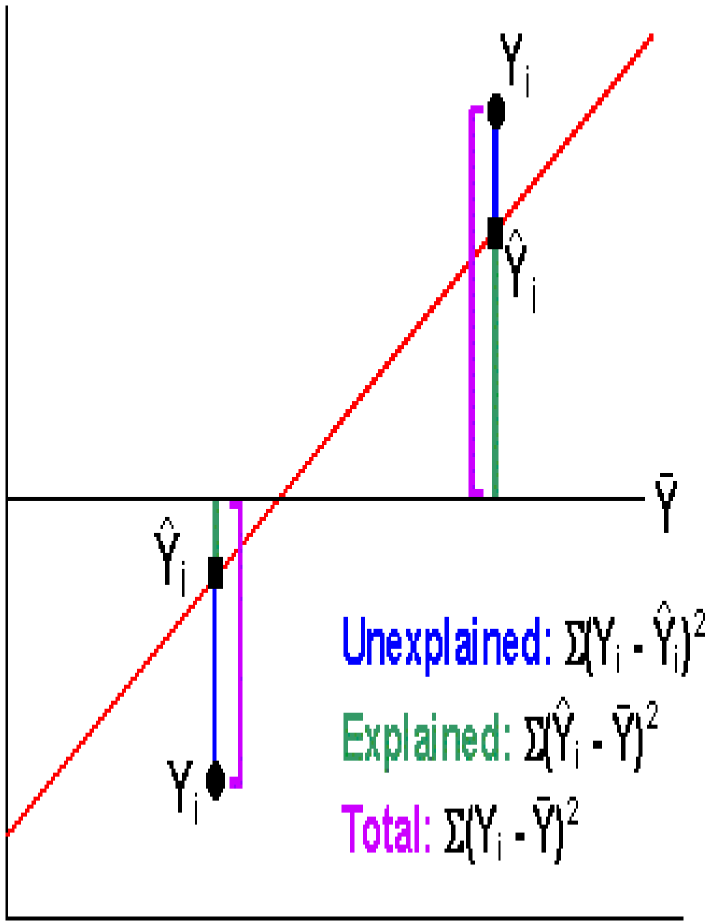
$$SC_{entre} = \sum_{j=1}^{k} n_j(\bar{x}_{.j} - \bar{x}_{..})^2$$

Diseño de una vía: estadística de prueba – suma de cuadrados

- Para resumir, se encuentra que la suma total de cuadrados es igual a la suma de cuadrados entre los grupos mas la suma de los cuadrados dentro de los grupos. Esta relación se expresa como:

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

$$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_{..})^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_{.j})^2 + \sum_{j=1}^{k} n_j(\bar{x}_{.j} - \bar{x}_{..})^2$$



- Con la suma de cuadrados, es posible calcular dos estimaciones de la variancia común de la población, σ^2.

Fuentes de variación	Suma de cuadrados	Grados de libertas	Cuadrado medio	Razón de variación
Entre muestras	SC entre	k-1	CM entre= (SC entre) / (k-1)	R.V= (CM entre) / (CM dentro)
Dentro de los muestras	SC dentro	n-k	CM dentro = (SC dentro) / (n-k)	
Total	SC total	n-1		

- Sumas de cuadrados
- Grados de libertad

Diseño de una vía: estadística de prueba – grados de libertad

- Los grados de libertad de para la fuente de variación entre tratamientos se expresa como:

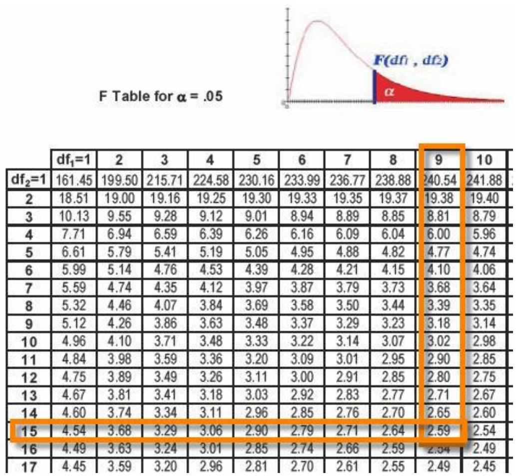
$$GL_{entre} = k - 1$$

- Los grados de libertad para la fuente de variación dentro de los tratamiento se expresa como:

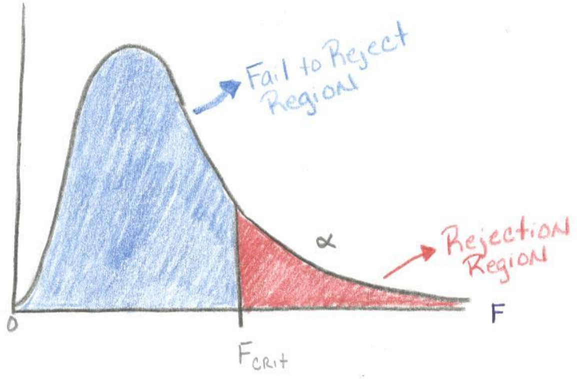
$$GL_{dentro} = n - k$$

- Los grados de libertad totales se expresa como:

$$GL_{total} = n - 1$$



	$df_1=1$	2	3	4	5	6	7	8	9	10
$df_2=1$	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45



¿ Por qué los grados de libertad poseen dichas expresiones?

Fuentes de variación	Suma de cuadrados	Grados de libertas	Cuadrado medio	Razón de variación
Entre muestras	$\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij}-\bar{x}_{.j})^2$	k-1	CM entre= (SC entre) / (k-1)	R.V= (CM entre) / (CM dentro)
Dentro de los muestras	SC dentro	n-k	CM dentro = (SC dentro) / (n-k)	
Total	SC total	n-1		

- Sumas de cuadrados
- Grados de libertad
- Cuadrado medio

Diseño de una vía: cuadrado medio

- Se posee dos tipos de cuadrados medios: entre y dentro. El primero expresa el efecto de las diferencias en los tratamientos, el segundo el grado de error que existe en el modelo…

- El cuadrado medio entre tiene por expresión:

$$CM_{entre} = \frac{\sum_{j=1}^{k} n_j(\bar{x}_{.j} - \bar{x}_{..})^2}{k-1} = \frac{SC_{entre}}{k-1}$$

- El cuadrado medio dentro se expresa como:

$$CM_{dentro} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{ij} - \bar{x}_{.j})^2}{n-k} = \frac{SC_{dentro}}{n-k}$$

¿Qué se puede deducir de los cuadrados medios anteriores?

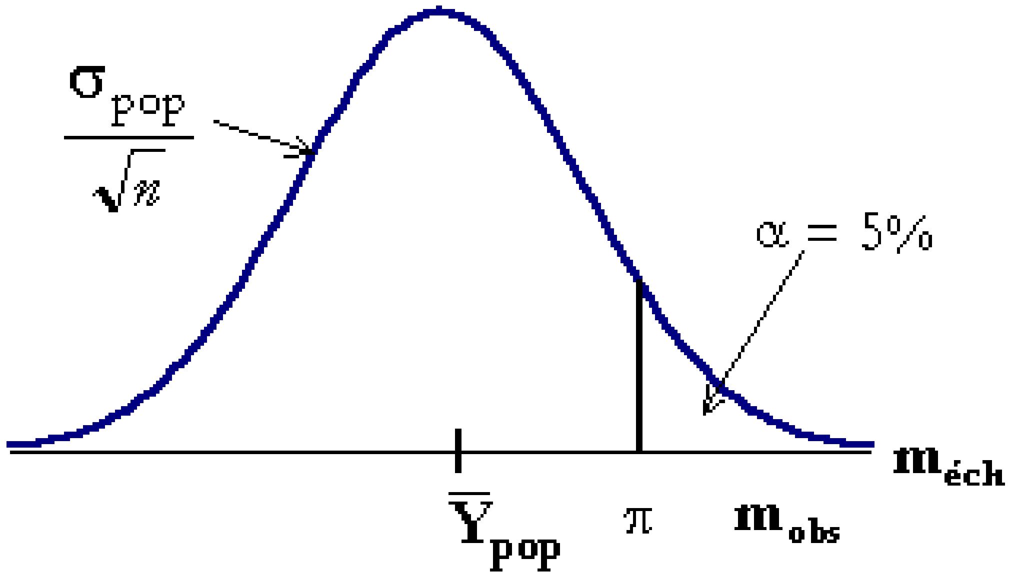
Fuentes de variación	Suma de cuadrados	Grados de libertas	Cuadrado medio	Razón de variación
Entre muestras	SC entre	k-1	CM entre= (SC entre) / (k-1)	R.V= (CM entre) / (CM dentro)
Dentro de los muestras	SC dentro	n-k	CM dentro = (SC dentro) / (n-k)	
Total	SC total	n-1		

- Sumas de cuadrados
- Grados de libertad
- Cuadrado medio
- Estadístico calculado (Razón de variación)

Diseño de una vía: estadística de prueba

- Finalmente se podrá entender el por qué del nombre del análisis de variancia.
- Para contemplar el cuadro del ANOVA, se realiza la razón de variación, esto es, se dividen los cuadrados medios.
- La expresión en la siguiente:

$$R.V. = \frac{CM_{entre}}{CM_{dentro}}$$



Fuentes de variación	Suma de cuadrados	Grados de libertas	Cuadrado medio	Razón de variación
Entre muestras	$\sum_{j=1}^{k} n_j(\bar{x}_{.j} - \bar{x}_{..})^2$	$\boldsymbol{k-1}$	$CM_{entre} = \frac{\sum_{j=1}^{k} n_j(\bar{x}_{.j} - \bar{x}_{..})^2}{k-1}$	$R.V. = \frac{CM_{entre}}{CM_{dentro}}$
Dentro de los muestras	$\sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{.j})^2$	$\boldsymbol{n-k}$	$CM_{dentro} = \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{.j})^2}{n-k}$	
Total	$\sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{..})^2$	$\boldsymbol{n-1}$		

Diseño de una vía: regla de decisión

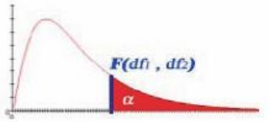
- Para tomar una decisión es necesario comparar la R.V calculada contra el valor critico de *F,* que puede obtenerse de la tabla de la F de Snedecor, con grados de libertad "*k* – 1" en el numerador y "*N* – *k*" en el denominador.

- Si el valor calculado para la R.V. es mayor o igual que el valor crítico *F,* la hipótesis nula es rechazada; pero si es menor que el valor crítico de *F,* no se rechaza la hipótesis nula.

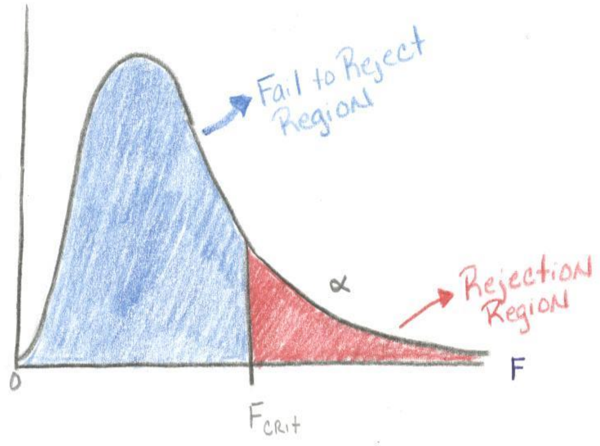
- Nótese que se utiliza por primera vez la F de Snedecor, y que también es la primera vez que se utiliza una distribución que posee dos grados de libertad.

- Nótese que la decisión se hace a través de variancias, y no medias…

F Table for $\alpha = .05$



	$df_1=1$	2	3	4	5	6	7	8	9	10
$df_2=1$	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54	241.88
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45



F - Distribution ($\alpha = 0.05$ in the Right Tail)

df_2 \ df_1 (Denominator Degrees of Freedom \ Numerator Degrees of Freedom)	1	2	3	4	5	6	7	8	9
1	161.45	199.50	215.71	224.58	230.16	233.99	236.77	238.88	240.54
2	18.513	19.000	19.164	19.247	19.296	19.330	19.353	19.371	19.385
3	10.128	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123
4	7.7086	9.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	6.9988
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107
40	4.0847	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240
60	4.0012	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401
120	3.9201	3.0718	2.6802	2.4472	2.2899	2.1750	2.0868	2.0164	1.9588
∞	3.8415	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799

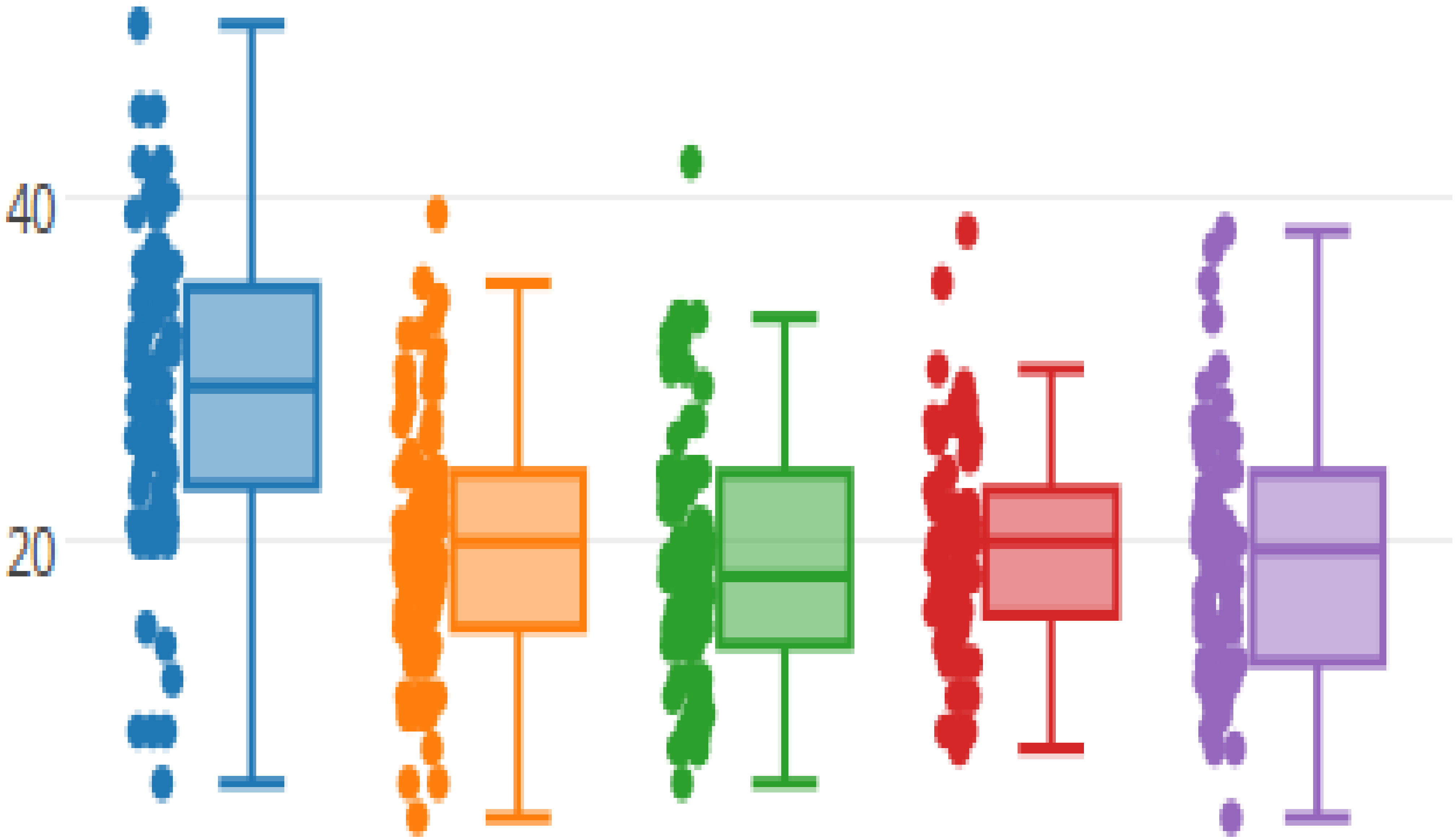
Diseño de una vía: calculo de la estadística de prueba

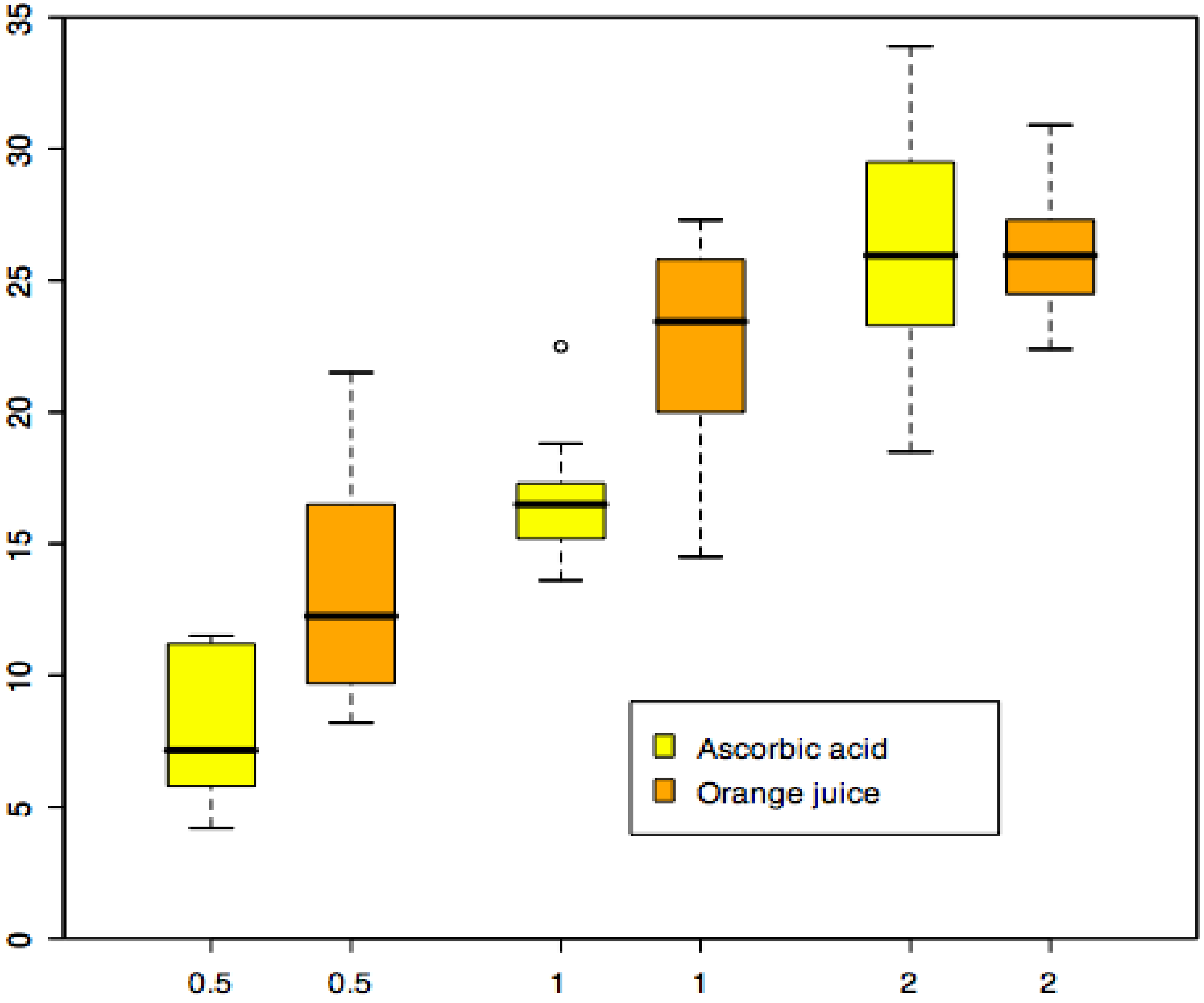
- ¿El rechazo de la hipótesis acerca de las variancias implica un rechazo de la hipótesis de la igualdad de las medias poblacionales? La respuesta es afirmativa…

- Un valor grande de R.V. resultó del hecho de que el cuadrado medio entre los grupos era considerablemente mayor que el cuadrado medio dentro de los grupos.

- Dado que el cuadrado medio entre los grupos se basa en la dispersión de las medias muestrales en torno a su media, esta cantidad será grande cuando exista una gran discrepancia entre los tamaños de las medias muestrales.

- Debido a esto, un valor significativo de la R.V. indica que se rechace la hipótesis nula de la igualdad de las medias poblacionales.


40
20



35
30
25
20
15
10
5
0
0.5
0.5
1
1
2
2
Ascorbic acid
Orange juice

Diseño de una vía: conclusión

- Cuando se rechaza Ho se concluye que no todas las medias son iguales…

- Cuando no se rechaza Ho, se concluye que probablemente todas las medias poblacionales son iguales…

Ejemplo…

- Se desea investigar si el nivel de calcio consumido diariamente en la dieta como tratamiento no farmacológico de la presión sanguínea elevada puede influir benéficamente la función endotelial en la hipertensión experimental mineralo-corticoide-NaCl.

- Se estudio un tipo de mamífero tomando su presión sanguínea luego de 7 semanas, afectados a 4 tipos de dieta, además de su peso.

- Los niveles de tratamiento fueron: sin tratamiento (WKY), dieta rica en calcio (WKY – Ca), dieta con deoxicorticosterona y NaCl (DOC), y dieta rica en calcio y tratamiento (DOC-Ca)

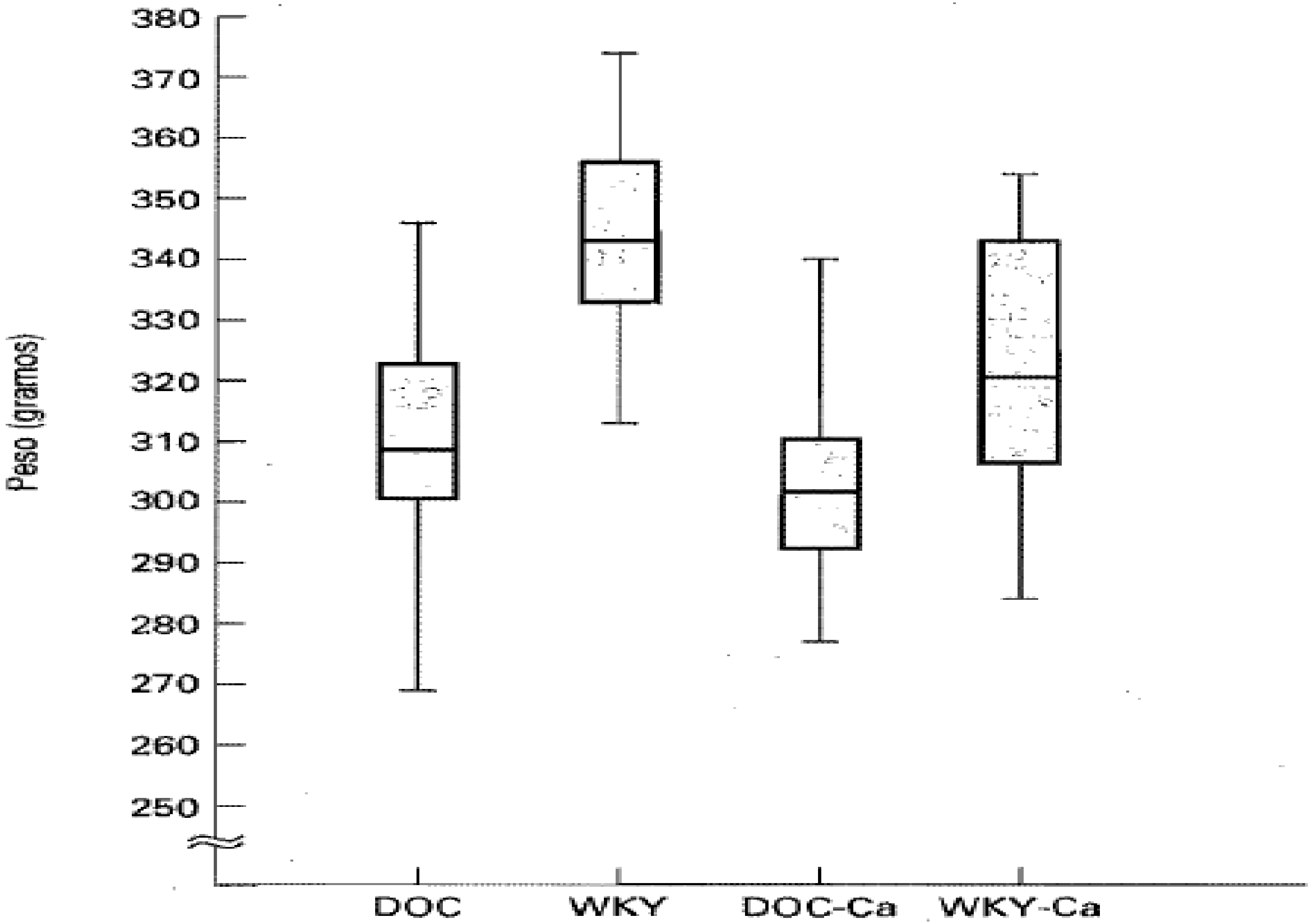
- Se quiere sabe si los 4 tratamientos tienen diferentes efectos en el peso promedio en la especie estudiada.

Ejemplo… 1. Datos

- Los datos corresponden a los pesos de la especie al final del estudio, junto con los totales del tratamiento y las medias (ver siguiente cuadro)

	DOC	WKY	DOC-Ca	WKY-Ca	
	336	328	304	342	
	346	315	292	284	
	269	343	299	334	
	346	368	293	348	
	323	353	277	315	
	309	374	303	313	
	322	356	303	301	
	316	339	320	354	
	300	343	324	346	
	309	343	340	319	
	276	334	299	289	
	306	333	279	322	
	310	313	305	308	
	302	333	290	325	
	269	372	300		
	311		312		**Total**
Promedio	**309,38**	**343,13**	**302,50**	**321,43**	**318,64**
Total	**4950**	**5147**	**4840**	**4500**	**19437**
Muestra	**16**	**15**	**16**	**14**	

Ejemplo… 1. Datos



Ejemplo… seguimos….

2. **Supuestos**: los cuatro conjuntos de datos forman muestras aleatorias simples e independientes, extraídas de cuatro poblaciones que son similares excepto por la condición estudiada. Además las poblaciones de mediciones siguen una distribución normal con variancias iguales.

3. **Hipótesis**:

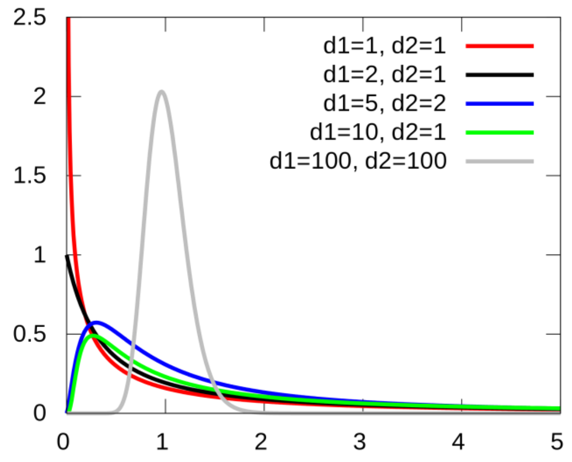
$$H_0:\ \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_A: \mu_i \neq \ \mu_j$$

4. **Estadística de prueba**:

$$R.V. = \frac{CM_{entre}}{CM_{dentro}}$$

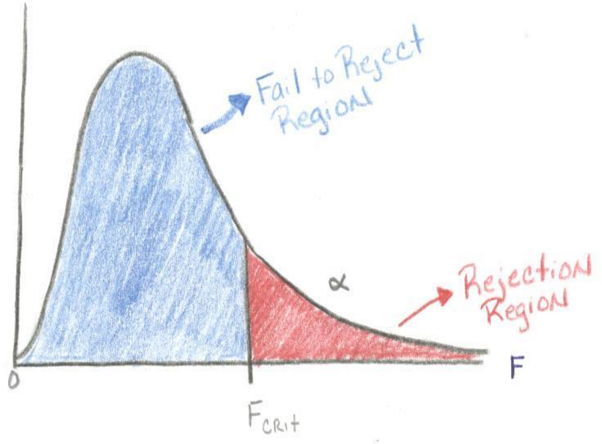
Ejemplo… seguimos….

5. **Distribución de la estadística de prueba**: si Ho es verdadera, y se cumple con las condiciones, entonces R.V. sigue una distribución F con los grados de libertad correspondientes, respectivamente, del numerador (k-1) y denominador (n-k), en este caso 3 (4-1) y 57 (61-4).



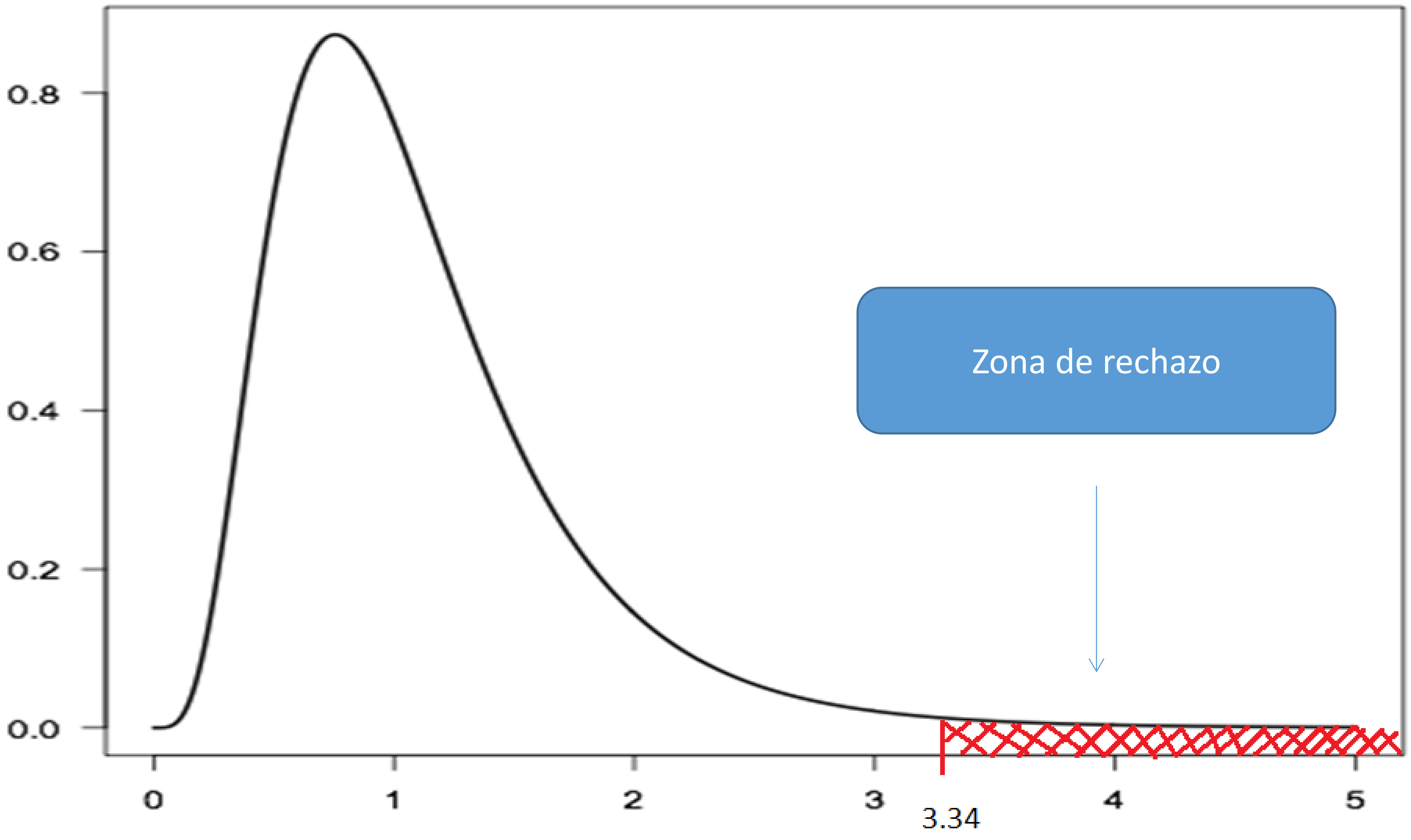
6. **Regla de decisión**: supóngase que $\alpha = 0.05$. El valor crítico para una F con 3 gados en el numerador y 57 en el denominador sería de

De F(3,57)= 3,34.



Por lo tanto, la hipótesis no se rechaza si el valor calculado es menor a 3.34, y se rechaza si es superior a 3.34.

Ejemplo… seguimos….



Ejemplo… seguimos….

7. **Cálculo de la estadística de prueba**: la suma de cuadrados

$SC_{entre} = \sum_{j=1}^{k} n_j \left(\bar{x}_{.j} - \bar{x}_{..}\right)^2 = 16*(309{,}38 - 318{,}64)^2 + \cdots + 14 * (321{,}43 - 318{,}64)^2$

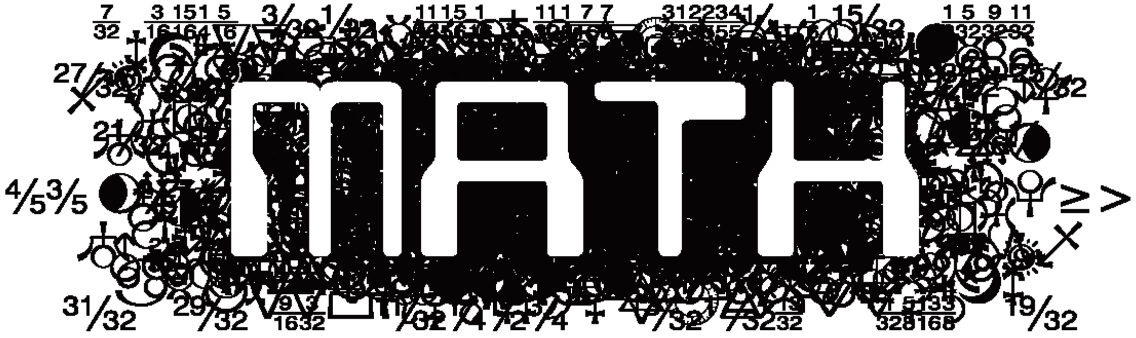
$SC_{entre} = 14649{,}15$

$SC_{dentro} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_{.j})^2 = (336 - 309{,}38)^2 + \cdots + (328 - 343{,}13)^2 + \cdots + (304 - 302{,}50)^2 + \ldots$

$SC_{dentro} = 23210.9023$

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

$SC_{total} = 37860.0547$



Ejemplo… seguimos….

7. **Cálculo de la estadística de prueba**: **grados de libertad**

$$\boldsymbol{GL_{entre} = k - 1 = 3} \qquad \boldsymbol{GL_{dentro} = n - k = 57} \qquad \boldsymbol{GL_{total} = 60}$$

Cuadrado medio

$$CM_{entre} = \frac{SC_{entre}}{k-1} = 4883.0503 \qquad CM_{dentro} = \frac{SC_{dentro}}{n-k} = 407.2088$$

Razón de variancia

$$R.V. = \frac{CM_{entre}}{CM_{dentro}} = 11.99$$

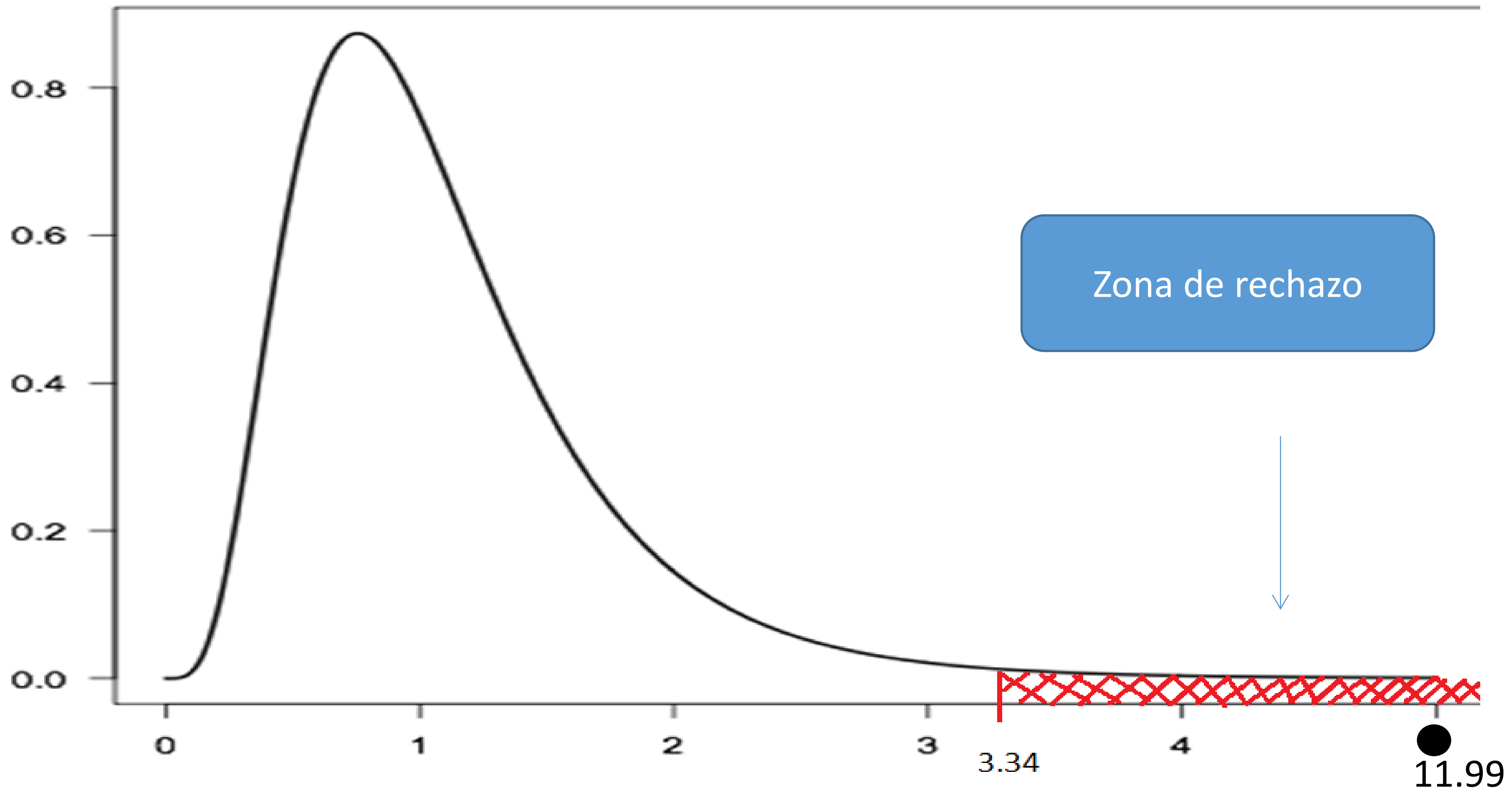
Ejemplo… seguimos….

El cuadro del análisis de variancia es el siguiente

Fuentes de variación	Suma de cuadrados	Grados de libertas	Cuadrado medio	Razón de variación
Entre muestras	14649.1514	3	4883.0503	11.99
Dentro de los muestras	23210.9023	57	407.2088	
Total	37860.0547	60		

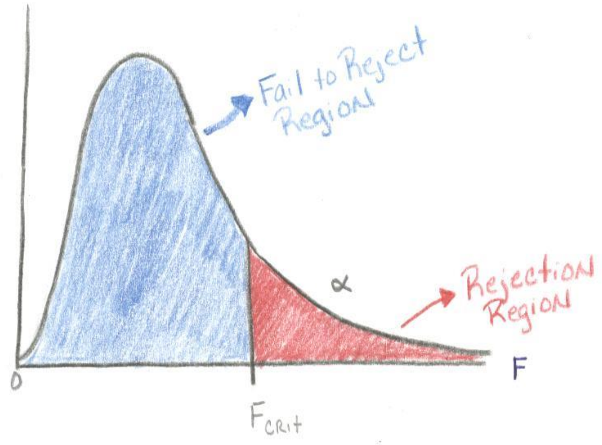
Ejemplo… seguimos….

8. **Decisión estadística**: 11.99 > 3,34. Se rechaza la hipótesis nula…



Ejemplo… y terminamos….

9. **Conclusión**: dado que se rechazó Ho, se concluye que la hipótesis alternativa es verdadera. Es decir, se concluye que los cuatro tratamientos no tienen el mismo efecto en promedio.



10. **Valor de *p***. Puesto que 11.99 > 4.77, por valores de l tabla se sabe que p <0,005 para esta prueba.

Advertencia: el diseño completamente aleatorizado es sencillo. Sin embargo, se debe utilizar cuando las unidades que reciben los tratamientos son homogéneas. Si las unidades experimentales no son homogéneas, ya sea por las características o las condiciones, se de aplicar alguna otra variación.

¿Las pruebas de comparaciones de medias se pudieron haber calculado mediante el método de análisis de variancia?

Índice

1 Fundamentos del ANOVA

2 Etapas de un ANOVA

3 Diseño de una vía o completamente aleatorizado

4 Diferencia de medias

La diferencia de medias

- NO ver con alumnos de audiología 2016.

Índice

Otros diseños

- NO ver con alumnos de audiología 2016.

Índice

Efectos fijos o aleatorios

- NO ver con alumnos de audiología 2016.

Conclusión y discusión

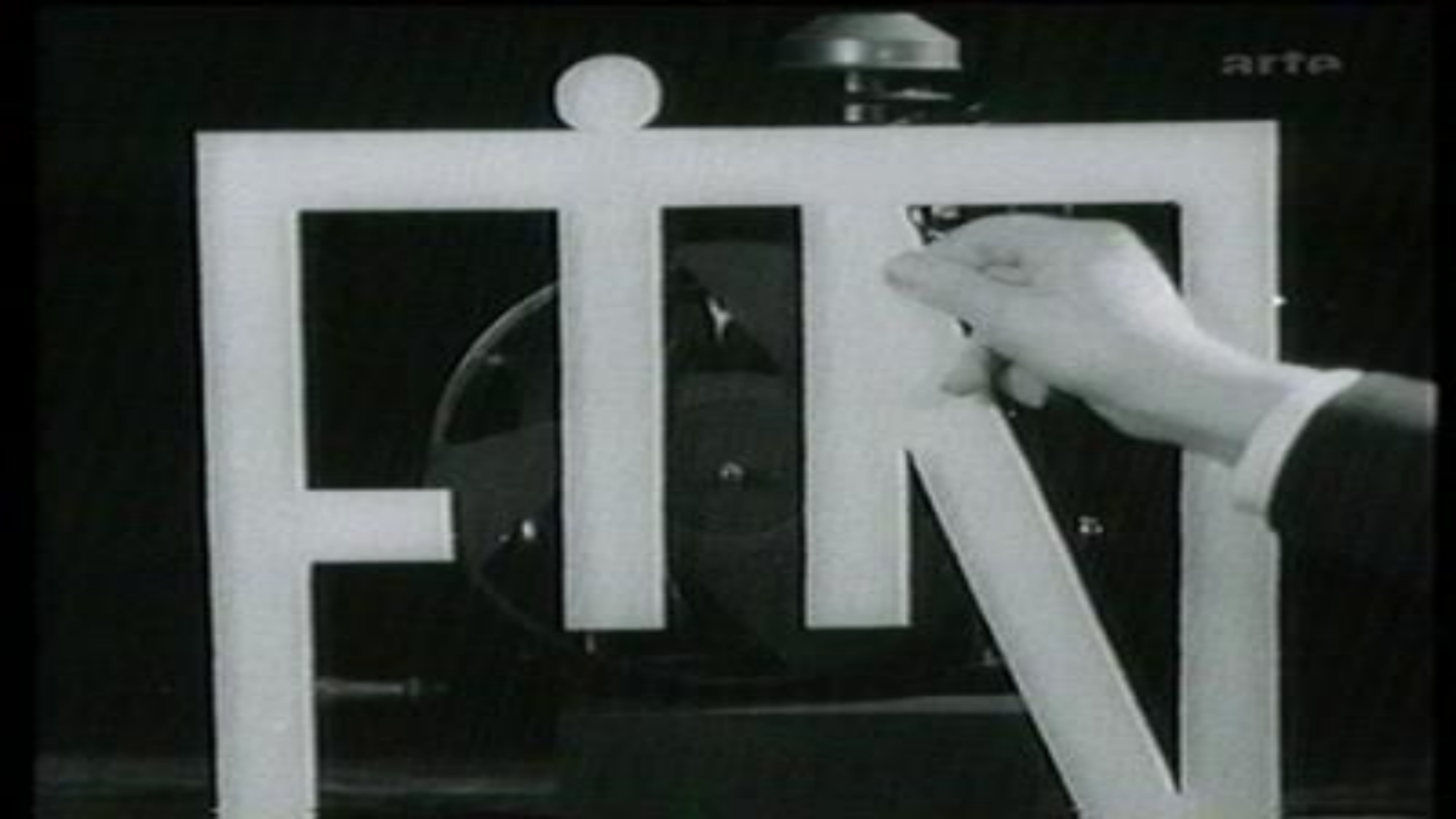
- El análisis de variancia se utiliza cuando se quiere hacer comparaciones para más de 2 medias.

- A diferencia de lo anterior, se debe particionar las fuentes de variación para obtener los resultados.

- Se introdujo la tabla F de Snedecor para poder obtener el resultado del análisis de variancia.

- Al rechazar el análisis de variancia, se debe verificar en grupos es que existen las diferencias.

- Se deben contemplar otras variaciones del ANOVA…


FIN