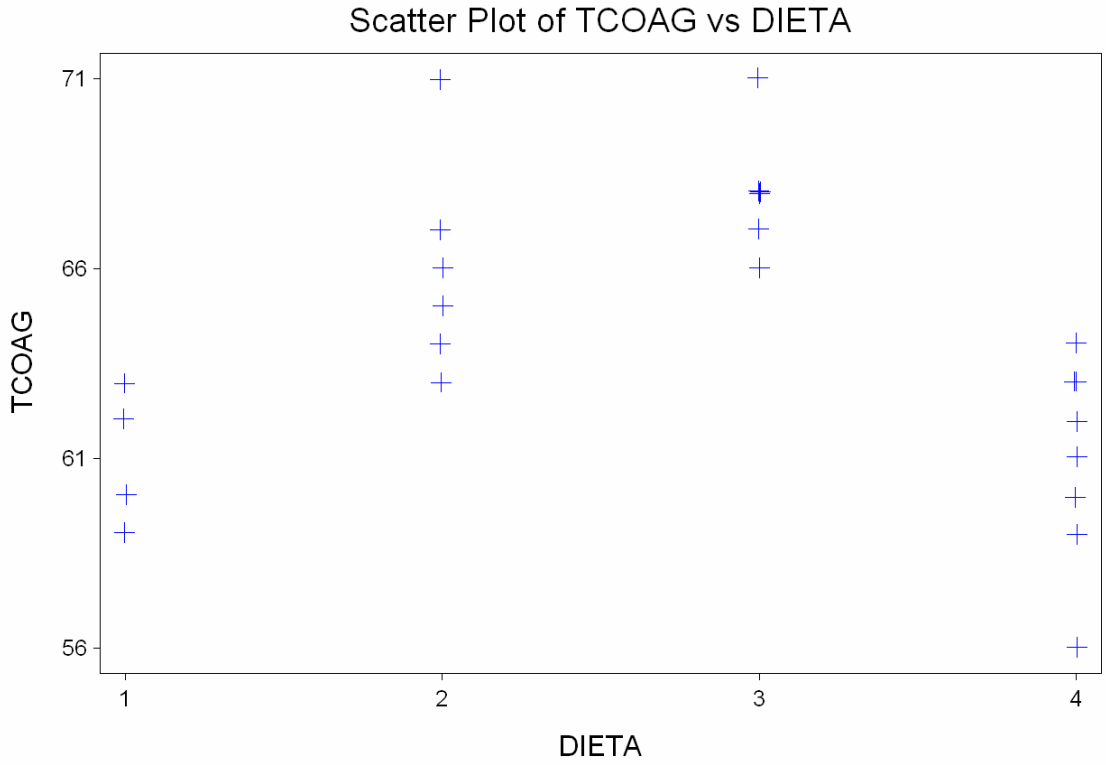
EJEMPLO 1

Se asignaron al azar ratas en condiciones similares a cuatro dietas (A – D).

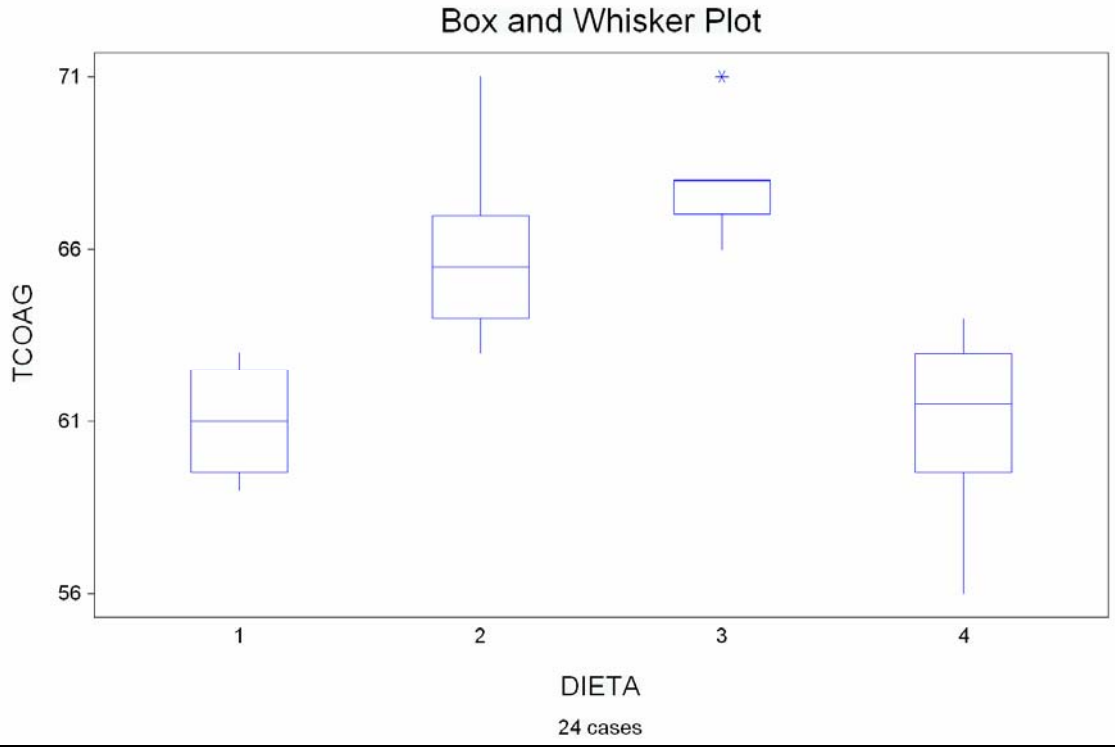
Dos semanas después se midió el tiempo de coagulación.

	DIETA1	DIETA2	DIETA3	DIETA4
	62	63	68	56
	60	67	66	62
	63	71	71	60
	59	64	67	61
		65	68	63
		66	68	64
				63
				59
Media	**61**	**66**	**68**	**61**

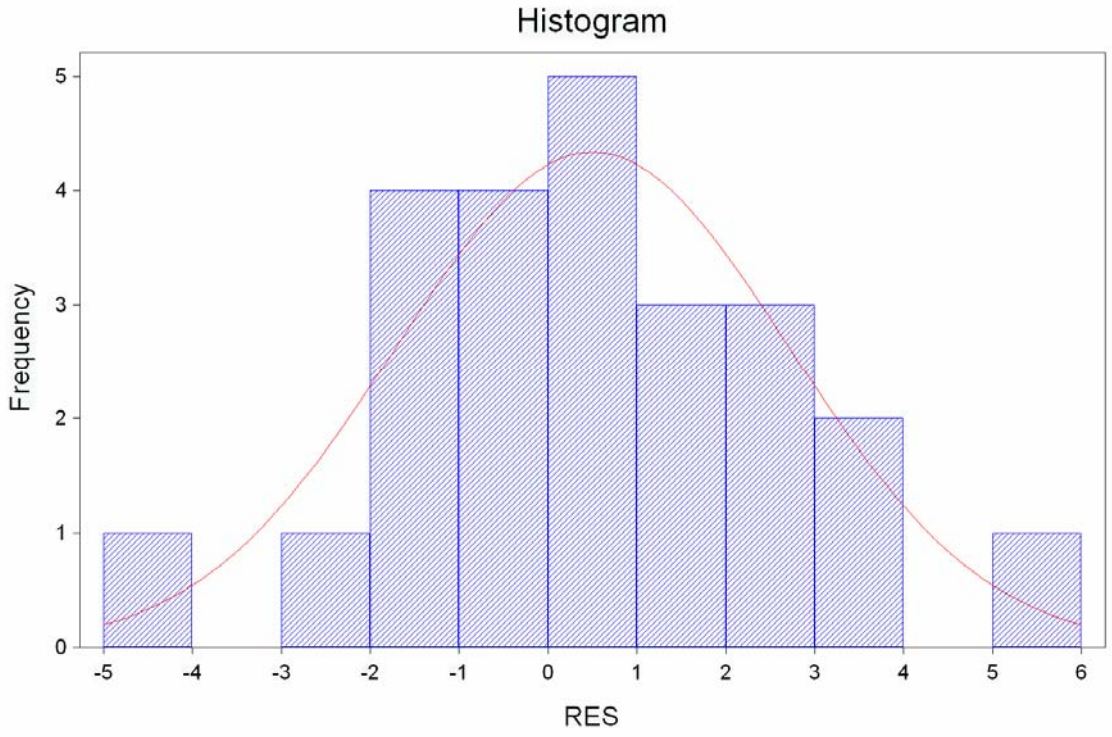
Statistics /Summary/ Scatter plot



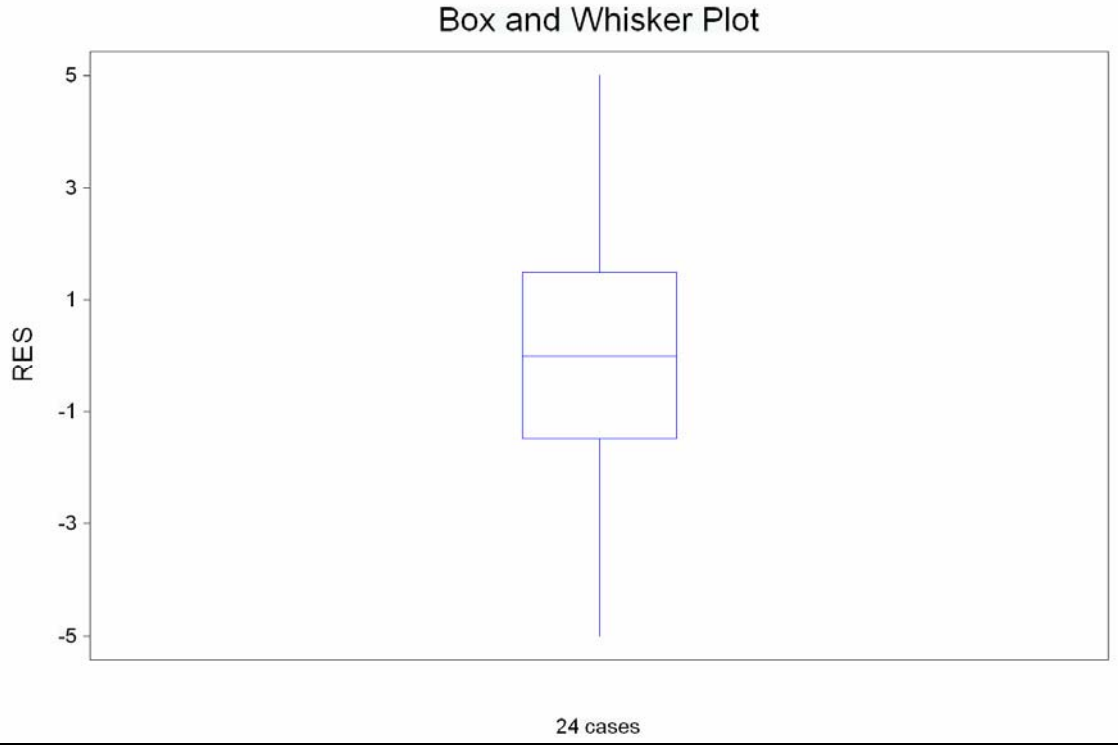
Statistics /Summary/ Box and Whisker plot



Statistics /Summary/ Histogram



Statistics /Summary/ Box and Whisker Plot



ANALISIS DE LA VARIANZA

STATISTIX 7.0

ONE-WAY AOV FOR TCOAG BY DIETA

SOURCE	DF	SS	MS	F	P
BETWEEN	3	228.000	76.0000	13.57	0.0000
WITHIN	20	112.000	5.60000		
TOTAL	23	340.000			

	CHI-SQ	DF	P
BARTLETT'S TEST OF EQUAL VARIANCES	1.67	3	**0.6441**

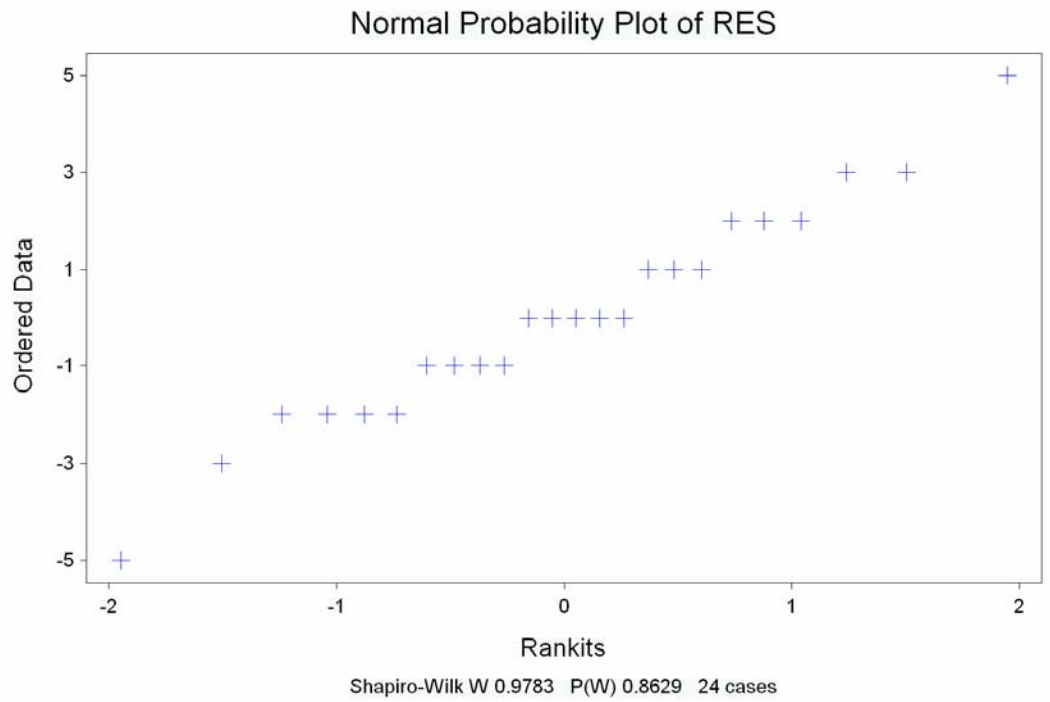
COCHRAN'S Q	0.3811
LARGEST VAR / SMALLEST VAR	2.8571

COMPONENT OF VARIANCE FOR BETWEEN GROUPS	11.9547
EFFECTIVE CELL SIZE	5.9

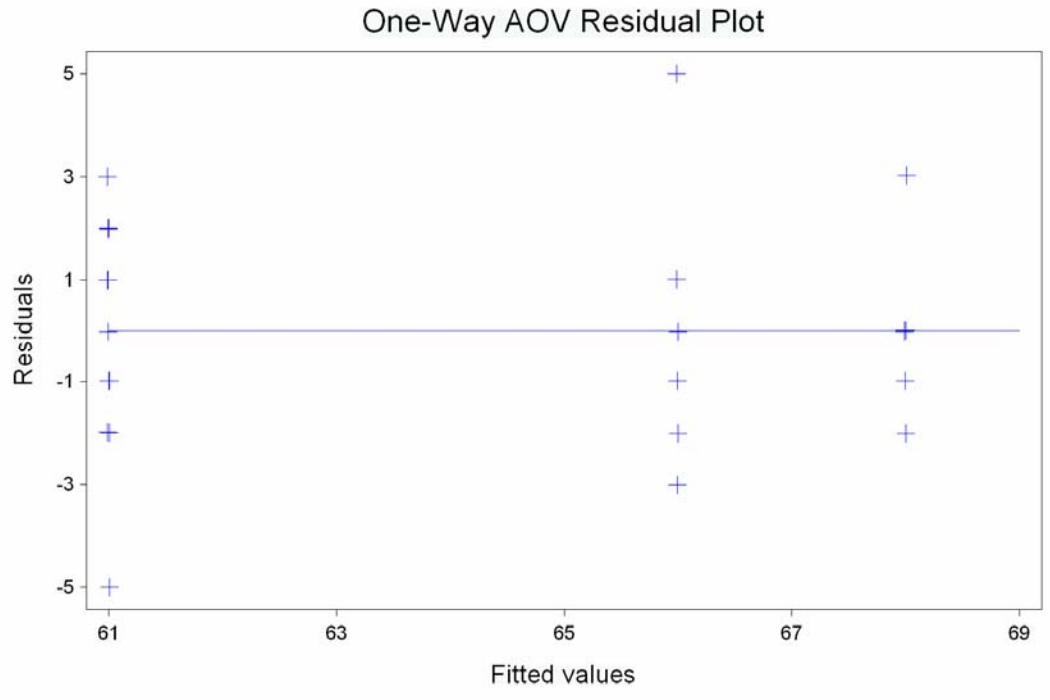
DIETA	MEAN	SAMPLE SIZE	GROUP STD DEV
1	61.000	4	1.8257
2	66.000	6	2.8284
3	68.000	6	1.6733
4	61.000	8	2.6186
TOTAL	64.000	24	2.3664

CASES INCLUDED 24 MISSING CASES 0

(En el menú que aparece con la tabla de ANOVA => Results / Plots / Normal Probability plot)



(En el menú que aparece con la tabla de ANOVA => Results / Plots / Residuals vs fitted values)



Test de Levene

Statistics/ One – Two – Multiple... / One-way anova
DIFABS = |TCOAG – MEDIANA|

```
STATISTIX 7.0

ONE-WAY AOV FOR DIFABS BY DIETA

SOURCE     DF       SS          MS         F        P
-------   ----  ---------   ---------   ------   ------
BETWEEN     3     4.33333     1.44444     0.65   0.5926
WITHIN     20     44.5000     2.22500
TOTAL      23     48.8333
```

COMPARACIONES MULTIPLES

(En el menú que aparece con la tabla de ANOVA => Results / Comparison of means)

BONFERRONI COMPARISON OF MEANS OF TCOAG BY DIETA

DIETA	MEAN	HOMOGENEOUS GROUPS
3	68.000	I
2	66.000	I
1	61.000	.. I
4	61.000	.. I

THERE ARE 2 GROUPS IN WHICH THE MEANS ARE
NOT SIGNIFICANTLY DIFFERENT FROM ONE ANOTHER.

CRITICAL T VALUE 2.927 REJECTION LEVEL 0.050
STANDARD ERRORS AND CRITICAL VALUES OF DIFFERENCES
VARY BETWEEN COMPARISONS BECAUSE OF UNEQUAL SAMPLE SIZES.

TUKEY (HSD) COMPARISON OF MEANS OF TCOAG BY DIETA

DIETA	MEAN	HOMOGENEOUS GROUPS
3	68.000	I
2	66.000	I
1	61.000	.. I
4	61.000	.. I

THERE ARE 2 GROUPS IN WHICH THE MEANS ARE
NOT SIGNIFICANTLY DIFFERENT FROM ONE ANOTHER.

CRITICAL Q VALUE 3.959 REJECTION LEVEL 0.050
STANDARD ERRORS AND CRITICAL VALUES OF DIFFERENCES
VARY BETWEEN COMPARISONS BECAUSE OF UNEQUAL SAMPLE SIZES.

LSD (T) COMPARISON OF MEANS OF TCOAG BY DIETA

DIETA	MEAN	HOMOGENEOUS GROUPS
3	68.000	I
2	66.000	I
1	61.000	.. I
4	61.000	.. I

THERE ARE 2 GROUPS IN WHICH THE MEANS ARE
NOT SIGNIFICANTLY DIFFERENT FROM ONE ANOTHER.

CRITICAL T VALUE 2.086 REJECTION LEVEL 0.050
STANDARD ERRORS AND CRITICAL VALUES OF DIFFERENCES
VARY BETWEEN COMPARISONS BECAUSE OF UNEQUAL SAMPLE SIZES.

SCHEFFE COMPARISON OF MEANS OF TCOAG BY DIETA

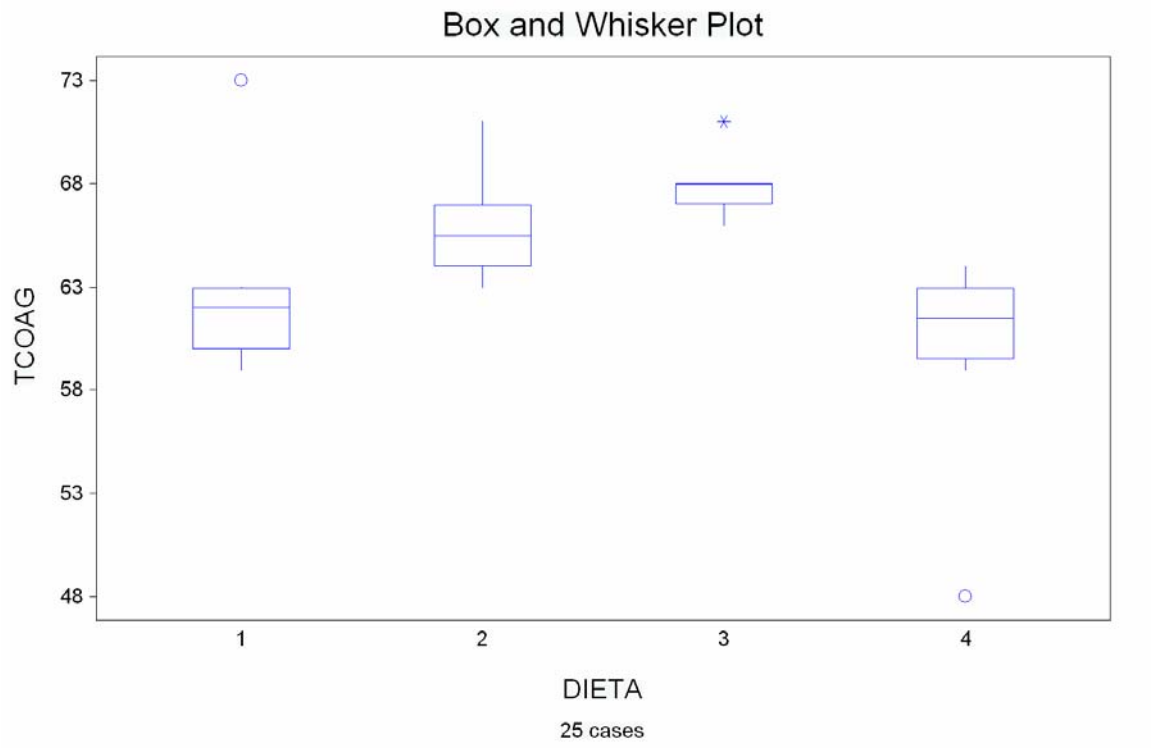
DIETA	MEAN	HOMOGENEOUS GROUPS
3	68.000	I
2	66.000	I
1	61.000	.. I
4	61.000	.. I

THERE ARE 2 GROUPS IN WHICH THE MEANS ARE
NOT SIGNIFICANTLY DIFFERENT FROM ONE ANOTHER.

CRITICAL F VALUE 3.098 REJECTION LEVEL 0.050
STANDARD ERRORS AND CRITICAL VALUES OF DIFFERENCES
VARY BETWEEN COMPARISONS BECAUSE OF UNEQUAL SAMPLE SIZES.

EJEMPLO 2

DIETA1	DIETA2	DIETA3	DIETA4
62	63	68	48
60	67	66	62
63	71	71	60
59	64	67	61
73	65	68	63
	66	68	64
			63
			59



ONE-WAY AOV FOR TCOAG BY DIETA

SOURCE	DF	SS	MS	F	P
BETWEEN	3	249.760	83.2533	4.81	0.0105
WITHIN	21	363.200	17.2952		
TOTAL	24	612.960			

	CHI-SQ	DF	P
BARTLETT'S TEST OF EQUAL VARIANCES	7.07	3	**0.0697**

COCHRAN'S Q 0.4577
LARGEST VAR / SMALLEST VAR 11.179

COMPONENT OF VARIANCE FOR BETWEEN GROUPS 10.6613
EFFECTIVE CELL SIZE 6.2

DIETA	MEAN	SAMPLE SIZE	GROUP STD DEV
1	63.400	5	5.5946
2	66.000	6	2.8284
3	68.000	6	1.6733
4	60.000	8	5.1270
TOTAL	64.040	25	4.1588

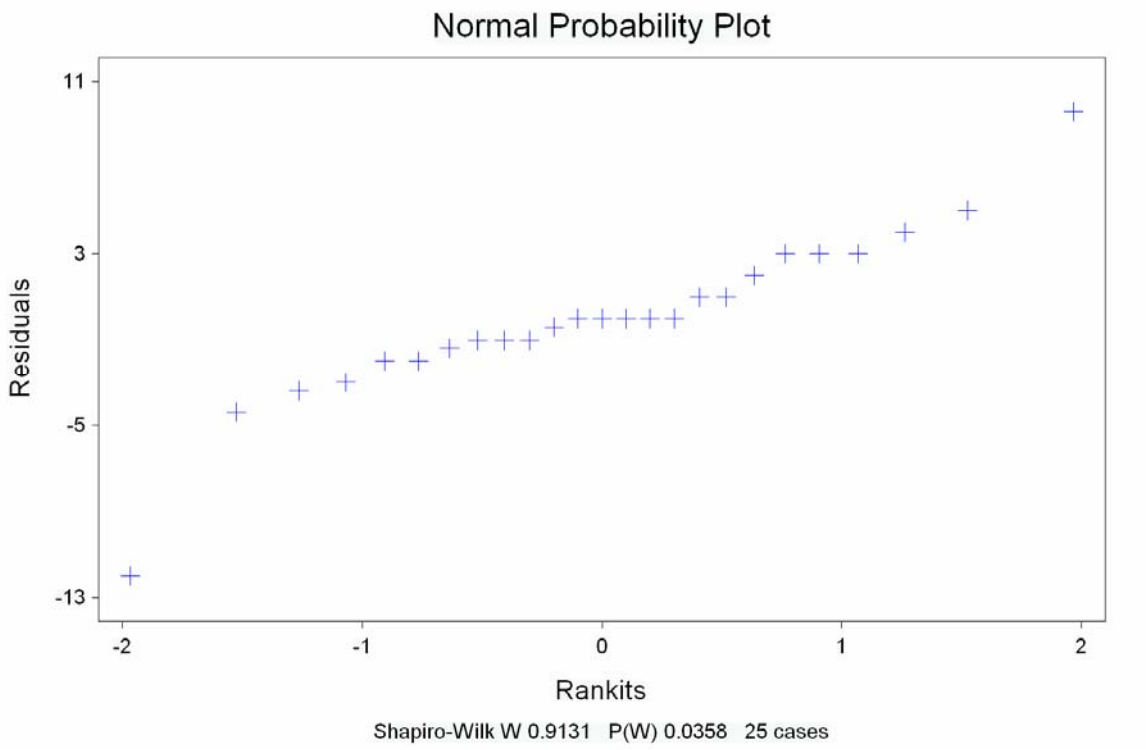
CASES INCLUDED 25 MISSING CASES 0

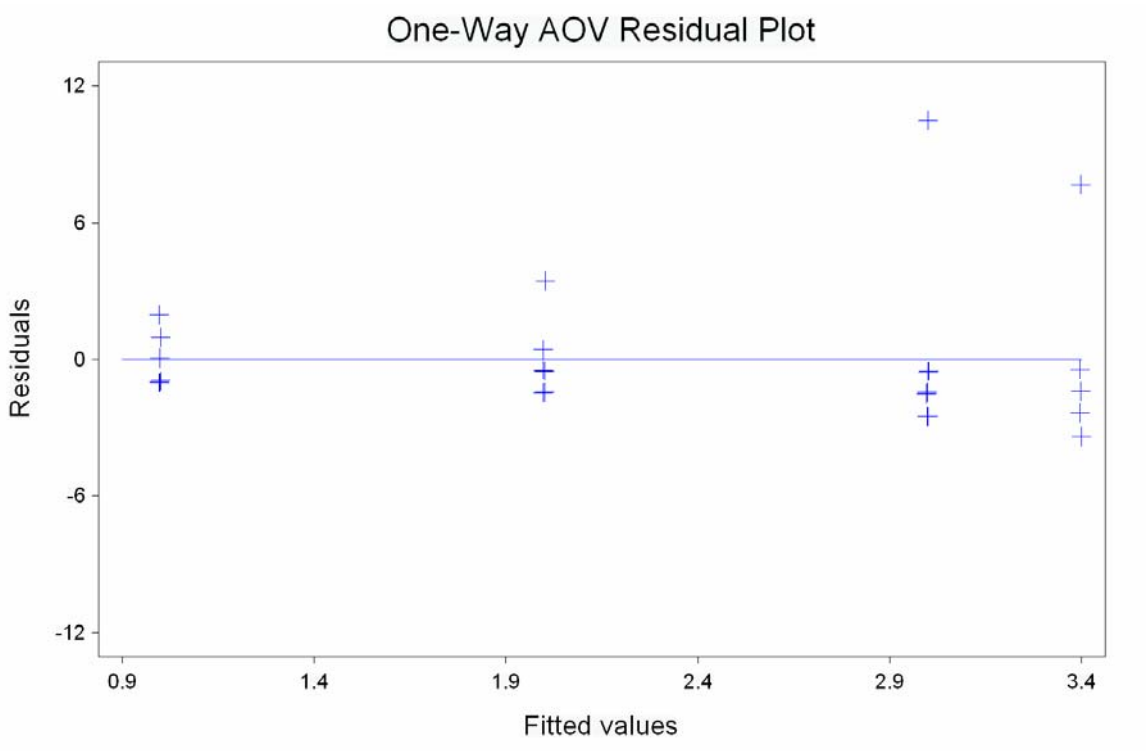
TEST DE LEVENE

ONE-WAY AOV FOR DIFABS BY DIETA

SOURCE	DF	SS	MS	F	P
BETWEEN	3	20.5600	6.85333	0.62	**0.6108**
WITHIN	21	232.700	11.0810		
TOTAL	24	253.260			

	CHI-SQ	DF	P
BARTLETT'S TEST OF EQUAL VARIANCES	8.87	3	0.0311


Normal Probability Plot
Residuals
11
3
-5
-13
-2
-1
0
1
2
Rankits
Shapiro-Wilk W 0.9131 P(W) 0.0358 25 cases


One-Way AOV Residual Plot
Residuals
12
6
0
-6
-12
0.9
1.4
1.9
2.4
2.9
3.4
Fitted values

TEST DE KRUSKAL-WALLIS

	DIETA1	DIETA2	DIETA3	DIETA4
	62 (7.5)	63 (10.5)	68 (21)	48 (1)
	60 (4.5)	67 (18.5)	66 (16.5)	62 (7.5)
	63 (10.5)	71 (23.5)	71 (23.5)	60 (4.5)
	59 (2.5)	64 (13.5)	67 (18.5)	61 (6)
	73 (25)	65 (15)	68 (21)	63 (10.5)
		66 (16.5)	68 (21)	64 (13.5)
				63 (10.5)
				59 (2.5)
Media de los rangos	**10.0**	**16.3**	**20.3**	**7.0**

STATISTICS / ONE – TWO – MULTIP.... / KRUSKAL WALLIS

KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV FOR TCOAG BY DIETA

DIETA	MEAN RANK	SAMPLE SIZE
1	10.0	5
2	16.3	6
3	20.3	6
4	7.0	8
TOTAL	13.0	25

KRUSKAL-WALLIS STATISTIC 13.2470
P-VALUE, USING CHI-SQUARED APPROXIMATION 0.0041

PARAMETRIC AOV APPLIED to RANKS

SOURCE	DF	SS	MS	F	P
BETWEEN	3	711.750	237.250	8.62	**0.0006**
WITHIN	21	577.750	27.5119		
TOTAL	24	1289.50			

TOTAL NUMBER OF VALUES THAT WERE TIED 21
MAX. DIFF. ALLOWED BETWEEN TIES 0.00001

CASES INCLUDED 25 MISSING CASES 0

COMPARACIONES A POSTERIORI DE KRUSKAL-WALLIS

COMPARISONS OF MEAN RANKS OF TCOAG BY DIETA

DIETA	MEAN RANK	HOMOGENEOUS GROUPS
3	20.250	I
2	16.250	I I
1	10.000	I I
4	7.0000	.. I

THERE ARE 2 GROUPS IN WHICH THE MEANS ARE
NOT SIGNIFICANTLY DIFFERENT FROM ONE ANOTHER.

REJECTION LEVEL 0.050
CRITICAL Z VALUE 2.64
CRITICAL VALUES OF DIFFERENCES VARY BETWEEN
COMPARISONS BECAUSE OF UNEQUAL SAMPLE SIZES.

Test no paramétrico para comparer 3 o más muestras: test de Kruskal-Wallis.
Este test es una generalización del test de Wilcoxon- Mann Whitney al caso de más de 2 muestras. Igual que el test de Mann Whitney no requiere que los datos sean normales, y el estadístico de este test no se calcula con los datos originales, sino con los rangos de los datos.

Los supuestos en que se basa el test son:

- Los datos son por lo menos ordinales.

- Además de la independencia entre las observaciones de una misma muestra suponemos independencia entre las observaciones de las distintas muestras.

De cada una de las k poblaciones tenemos una muestra aleatoria de tamaño n_i, es decir:

Muestra de la población i	Tamaño de muestra
$Y_{11}, Y_{12},\ldots\ldots\ldots, Y_{1n1}$	n_1
$Y_{21}, Y_{22},\ldots\ldots\ldots, Y_{2n2}$	n_2
……	…..
……	…..
……	…..
$Y_{k1}, Y_{k2},\ldots\ldots\ldots, Y_{knk}$	n_k
Total de observaciones	$N = n_1 + n_2 + \ldots . n_k$

La hipótesis nula a testear es

H_o : todas las poblaciones tienen la misma distribución

Bajo H_o , todas las observaciones provienen de poblaciones idénticas. Si hacemos un pool con todas las N observaciones Y_{ij} y las ordenamos de menor a mayor, obtendremos los rangos

$$R_{ij}$$

Si H_o es cierta , las observaciones Y_{ij} provienen de una misma distribución y por lo tanto, todas las asignaciones de los rangos a las k muestras tienen la misma chance de ocurrir.

Si H_o es falsa, algunas muestras tenderán a tomar los rangos más pequeños, mientras que otras tenderán a tomar los rangos más grandes.

El test de Kruskal-Wallis mide la discrepancia entre los promedios observados $\overline{R}_{i.}$ para cada tratamiento i y el valor que esperaríamos si Ho fuera cierta.

El SX nos da el valor del estadístico G_{KW}. En este test rechazamos la hipótesis nula de igualdad de medias si G_{KW} es grande.

El estadístico puede calcularse como

$$G_{KW} = \frac{SSB}{\frac{SSTO}{N-1}}$$

donde SSB y SSTO son, respectivamente, la suma de cuadrados between y la suma de cuadrados total para la tabla de análisis de la varianza correspondiente a los rangos de las observaciones.

SX nos da el p-valor usando la aproximación por una distribución: χ^2_{k-1}

Esta aproximación es válida cuando:

k=3 $n_i \geq 6$ para las k muestras o bien k>3 $n_i \geq 5$ para las k muestrasPara el caso en que k=3 y los $n_i \leq 5$ se debe usar la tabla con los percentiles de la distribución exacta.

Como vemos la salida de SX incluye la tabla de ANOVA para los rangos de las observaciones. Esto se basa en que los dos estadísticos están relacionados. Si llamamos F_R al estadístico del test de F aplicado a los rangos tenemos que:

$$F_R = \frac{(N-k)G_{KW}}{(k-1)(N-1-G_{KW})}$$

INTERVALOS DE CONFIANZA SIMULTANEOS

Intervalos de confianza simultáneos (concepto general, no sólo para el análisis de varianza de un factor)

¿Cuál es la **definición** de IC para un parámetro θ?
Recordemos que si $X=(X_1,X_2,...,X_n)$ es la muestra observada, un intervalo $[a(X),b(X)]$ es un IC para θ con nivel $1-\alpha$ si

$$\mathbf{P(\ a(X) \le \theta \le b(X)\ ) = 1-\alpha}$$

Ahora deseamos calcular IC para cada uno de los parámetros θ_j (digamos j=1...,m). Se dice que el intervalo $[a_j(X),b_j(Y)]$ es un IC para θ_j calculado por un método simultáneo si

$$P\left(\bigcap_{j=1}^{m} [a_j(X) \le \theta_j \le b_j(X)] \right) \quad \ge \quad 1-\alpha$$

(27)

o sea que la probabilidad de que todos los IC sean correctos (contengan al verdadero valor del parámetro) es $\ge 1-\alpha$. La probabilidad de que alguno sea incorrecto es $\le \alpha$.

¿Se pueden calcular muchos IC o aplicar muchos tests?

¿Cuál es la crítica que se suele hacer a los IC "usando la distribución t" (de la forma (26)) y a los tests deducidos de estos intervalos?

Si hacemos unos pocos intervalos **elegidos a priori** (antes de observar los datos) la probabilidad de equivocarnos será >5%, pero no será tan alta...

Si por ejemplo tenemos 6 tratamientos y hacemos todas las comparaciones de a pares, el nro. de IC será 15, ¿cuál será la probabilidad de que alguno no contenga al verdadero valor del parámetro? Aunque no la sepamos calcular, es evidente que esta probabilidad es mucho > que 0.05.

Por eso cuando uno planea de antemano hacer uno o muy pocos intervalos o tests puede usar (26), pero en caso contrario conviene utilizar un método de intervalos de confianza simultáneos.

Método de Bonferroni.

Un método muy general (para cualquier modelo) para obtener intervalos de confianza simultáneos es calcular cada uno de ellos con nivel 1-α/m, donde m es el número de IC que se desea calcular. (Ej.: demostrar que de este modo se consigue (27), usando la desigualdad de Bonferroni).

Este método tiene la ventaja de ser muy simple y muy general, pero sólo se usa en la práctica si m es muy pequeño, porque para valores moderados de m da IC de mucha longitud.

Para el caso particular del análisis de la varianza de un factor, basta usar (26), pero reemplazando $t_{n-k,\alpha/2}$ por $t_{n-k,\alpha/2m}$ donde m es el número de IC que se desea calcular:

$$[\overline{X}i - \overline{X}_j - t_{n-k,\alpha/2m}\sqrt{s_p^2\left(\frac{1}{n_i}+\frac{1}{n_j}\right)};\ \overline{X}_i - \overline{X}_j + t_{n-k,\alpha/2m}\sqrt{s_p^2\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}]$$

Método de Tukey.

Los intervalos de Tukey son similares a los dados en (26) pero reemplazando $t_{n-k,\alpha/2}$ por el valor $\mathbf{q_{k,n-k,\alpha}/\sqrt{2}}$

$$\overline{X}_i - \overline{X}_j \pm \frac{1}{\sqrt{2}} q_{k,N-k,\alpha}\sqrt{S_p^2(\frac{1}{n_i}+\frac{1}{n_j})}$$

donde los valores "q" están tabulados y corresponden a la distribución estudiada por Tukey, llamada distribución del "**rango studentizado**" de k variables normales independientes. El $\sqrt{2}$ que aparece se debe simplemente a como se construyó la tabla.

Para el caso originalmente pensado por Tukey en el que los tamaños de muestras son iguales ($n_1=n_2=...=n_I$), este método hace que se cumpla el = en vez del $\geq$ en (27) cuando se realizan todas los comparaciones de a pares. El método de Tukey es óptimo (da IC de la menor longitud posible) cuando se desea calcular IC para todos los pares posibles y los n_j's son iguales.

Para el caso en que los tamaños de muestras no son iguales, se demostró que sigue valiendo (27) pero con ">". En este caso el método se conoce también como "método de Tukey-Kramer".

Tests simultáneos: son los derivados de IC simultáneos. Tienen la propiedad de que la probabilidad de cometer algún error tipo I es menor o igual que α.

Comparación de los métodos considerados

Si se desea calcular un IC o aplicar un test para una sola diferencia de medias elegida **a priori**, evidentemente el método de elección es el basado en la distribución t. Si son unos pocos, elegidos a priori conviene usar Bonferroni. Si se hacen muchas comparaciones de a pares (o algunas elegidas a posteriori, que es "igual que hacer muchas") conviene usar Tukey (da intervalos de menor longitud que Bonferroni).

Para elegir entre Bonferroni y Tukey, no es "trampa" elegir el método que da IC de menor longitud. No se necesita hacer las cuentas del IC para elegir el método: basta comparar quien es menor entre los valores de la tabla de "t" y de la tabla de "q" (entre $t_{n-k,\alpha/2m}$ y $q_{k,n-k,\alpha}/\sqrt{2}$).

Evaluación de los supuestos del ANOVA

Para poder aplicar un análisis de varianza a un conjunto de datos es necesario que éstos satisfagan los supuestos del método, ellos son:

1. Independencia de las observaciones dentro de cada grupo y entre grupos.
2. Distribución normal de las observaciones de cada grupo.
3. Homogeneidad de varianzas (Las varianzas de los k grupos son iguales).

De los dos últimos supuestos, es más grave que no se cumpla la homogeneidad de varianza, especialmente cuando el diseño no es balanceado (igual número de observaciones en los distintos grupos).

¿Cómo chequeamos estos supuestos?

1. *Independencia* $\rightarrow$ Depende del diseño del estudio.
   - En un ensayo equivale a pedir asignación aleatoria de los pacientes a los grupos.
   - En un estudio observacional equivale a pedir que se hayan obtenido muestras aleatorias de cada población.
   - No se satisface si el mismo sujeto es observado en más de una oportunidad.
   - No se satisface en observaciones apareadas o "matched"

2. *Distribución normal de las observaciones*
   - Si cada grupo tiene un buen número de datos, chequear gráficamente (box-plots, histogramas, normal probability plot) y analíticamente (Test de Shapiro-Wilk).
   - Si el número de datos en cada grupo es pequeño, se chequea este supuesto utilizando los RESIDUOS. Definimos el residuo de un dato como la diferencia entre ese dato y el valor estimado para la media de esa población, es decir RESIDUO = (DATO – MEDIA DEL GRUPO).
     Si todas las poblaciones son normales y tienen la misma varianza, los N residuos pueden ser considerados como una muestra de la misma distribución $\rightarrow$ $N(0 , \sigma^2)$.
     Por lo tanto, graficamos los N residuos o hacemos el test de Shapiro-Wilk sobre los N residuos.

3. *Homogeneidad de varianzas*

- Si las observaciones de cada grupo son normales, entonces uno de los tests que se pueden usar para testear la hipótesis de homogeneidad de varianzas es el test de Bartlett. STATA produce automáticamente este test a continuación de la Tabla de Anova cuando se usa el comando *oneway*.
- Supuestos del test de Bartlett:

- las observaciones son independientes,
- los datos tienen distribución normal.

- Hipótesis del Test de Bartlett

  H_o: $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$ versus H_1 : "alguna varianza es diferente".

Método Gráfico

1) Un modo simple y útil para estudiar el supuesto de homogeneidad de varianzas es hacer un GRAFICO DE RESIDUOS vs VALORES PREDICHOS (o fitted values). Este gráfico es sumamente útil cuando se utilizan modelos más complejos de análisis de varianza o cuando se hace un análisis de regresión; casos en los que es necesario además estudiar si el modelo propuesto es adecuado para los datos
2) Grafico de probabilidad normal o q-q plot

¿POR QUÉ NO TRATAR EL PROBLEMA DE COMPARACIÓN DE VARIAS MEDIAS SIMPLEMENTE APLICANDO EL TEST T PARA DOS MUESTRAS TANTAS VECES COMO SEA NECESARIO?

En nuestro ensayo con 4 tratamientos deberíamos hacer 6 comparaciones:

T1 vs T2	T1 vs T3	T1 vs T4
T2 vs T3	T2 vs T4	T3 vs T4

Y si tuviéramos 6 tratamientos? ¡ 15 comparaciones ! En general el número de comparaciones de a pares que se pueden realizar con k grupos es $k\,(k - 1)/2$.

¡¡¡ Para 10 tratamientos, se deben hacer 45 tests t !!! Pero el problema no es sólo el trabajo ....

→ ¿Cuál es el **nivel de significación conjunto para este conjunto de tests t** que estamos haciendo?

La característica más importante de un test de hipótesis es que es un instrumento que nos permite medir o controlar la probabilidad de que ocurra error de tipo I. Recordemos,

α = *nivel de significación del test* = P(Error Tipo I) =

= P(rechazar H_o cuando H_o es verdadera)

Volvamos a nuestro ejemplo con 4 tratamientos y en el que sería necesario hacer 6 tests para comparar todos los grupos de a pares. Si cada test t se hace con un nivel de significación $\alpha = 0.05$ (es decir rechazamos la hipótesis nula cuando el p-valor < 0.05), ¿Cuál es el NIVEL GENERAL de los 6 tests que estamos haciendo?

Trataremos de calcularlo. Para ello supondremos que :

- todas las muestras provienen de poblaciones con la misma media μ.
- que cada test se hace con nivel de significación α=0.05, es decir

  P(rechazar H_o) = 0.05 y por lo tanto P(no rechazar H_o) = 0.95.
- que los 6 tests que estamos haciendo son independientes (este supuesto es falso!!!, pero permite simplificar los cálculos).

Entonces,

P(no rechazar ninguna de las 6 hipótesis nulas) =

P(no rechazar la 1º H_o $\cap$ no rechazar la 2º H_o $\cap$ ... $\cap$ no rechazar la 6º H_o) =

$= (0.95)^6 = 0.735$

El NIVEL GENERAL o SIMULTÁNEO de los 6 tests, es la probabilidad de cometer Error Tipo I en al menos uno de los tests, es decir

α^* = P(rechazar H_o en al menos 1 test | todas las H_o son verdaderas)

= 1 – 0.735 = 0.265 $\Longrightarrow$ ¡ 26 % !!!!!

(En realidad el problema es aún más complejo, porque estamos utilizando el mismo conjunto de datos para realizar todos los tests por lo que estos no pueden suponerse independientes!!!)

El hecho que el nivel general del conjunto de tests aumenta a medida que aumenta el número de comparaciones, es la razón fundamental por la cual NO DEBE realizarse la comparación de las medias de varias poblaciones a través de comparaciones de a pares.

$\rightarrow$ Un segundo inconveniente al hacer todos los test t de a pares, es que en cada test t suponemos homogeneidad de varianzas, por lo tanto estamos suponiendo que TODAS las poblaciones tienen la misma varianza. Sin embargo, en cada test usamos una estimación diferente de la varianza poblacional. Parece más razonable utilizar una estimación de la varianza poblacional basada en todas las muestras, tal como se hace en el ANOVA.

Algunos comentarios sobre los tests de comparaciones múltiples

- Sólo cuando rechazamos H_o en el Análisis de Varianza, tiene sentido hacer tests a posteriori. Si el ANOVA, que es más potente que las comparaciones individuales, fracasa en rechazar H_o, entonces no tiene sentido "bucear" en los datos para rastrear posibles diferencias.
- El número de comparaciones que se pretendan realizar NO DEBE DECIDIRSE DESPUES DE MIRAR LOS DATOS. O se toma *m* = número de total de comparaciones posibles ó las comparaciones de interés están planeadas de antemano.

(En algunos experimentos hay comparaciones entre grupos que son de interés para el investigador, se las denomina comparaciones pre-planeadas o comparaciones *a priori*).

- Tanto el test de Bonferroni como el de Sidak son poco potentes, son muy conservativos, en el sentido de que fallan en detectar diferencias que verdaderamente existen. Existen otros métodos de comparaciones múltiples más potentes, tales como por ejemplo el test de Tukey o el test de Duncan.

Algunos comentarios sobre el ANALISIS de la VARIANZA

- ¿ Qué hacer cuando no se satisfacen los supuestos del ANOVA? Existen alternativas no paramétricas para la comparación simultánea de 3 o más grupos de observaciones. El método más conocido es el Test de Kruskal-Wallis.
- En esta clase hemos considerado el caso más simple de un análisis de varianza, que corresponde a k grupos de observaciones independientes, donde los grupos están definidos en base a un único factor. En nuestro ejemplo, el factor que define los grupos es el tipo de droga recibido. En este caso utilizamos un ANOVA a un criterio de clasificación (one-way) o ANOVA de un factor, ya que hay un único factor bajo estudio. Queremos estudiar la influencia del factor droga sobre la variable respuesta (TAS).
- Consideremos el mismo problema de comparar 4 drogas, pero supongamos que se utiliza un diseño en el que cada paciente recibe los 4 tratamientos. En este caso las observaciones de los distintos grupos no son independientes. El modelo propuesto para realizar el análisis DEBE tener en cuenta que hay observaciones que provienen del mismo sujeto. $\Rightarrow$ Diseño en BLOQUES.
- Puede ser de interés realizar un ensayo en el cual se compare el efecto de las 4 drogas sobre la TAS, y simultáneamente se prueben dos dosis (alta y baja). En este caso hay dos factores bajo estudio DROGA (4 niveles) y DOSIS (2 niveles), en realidad tenemos 8 tratamientos con la siguiente estructura:

	Droga 1	Droga 2	Droga 3	Droga 4
Dosis Baja				
Dosis Alta				

Tenemos un diseño FACTORIAL de los tratamientos, en el cual interesa estimar, no sólo el efecto de cada droga y de cada dosis, sino además si existe interacción entre la droga y la dosis. Es decir, responder a la pregunta si el efecto de la droga es independiente o no de la dosis.

- **Análisis de la varianza** hace referencia a un conjunto de metodologías que se inscribe dentro de un contexto más general, MODELOS LINEALES. La idea básica es que las observaciones pueden ser descriptas a través de un modelo que es una suma de diferentes contribuciones, donde cada contribución depende de un parámetro que mide el efecto de algún o algunos de los factores bajo estudio. La ventaja de esta

metodología es que puede ser utilizada para diseños experimentales muy complejos. El problema general consiste en estimar los parámetros del modelo y construir tests para diferentes hipótesis sobre estos parámetros.