14. ANÁLISIS DE LA VARIANZA DE UN FACTOR

14.1 INTRODUCCIÓN

El análisis de la varianza de un factor es una extensión del test de t para dos muestras independientes, para comparar K muestras.

Hemos visto métodos para decidir si la diferencia entre dos medias muestrales es (ó no es) significativa. Muchas veces interesa comparar más de dos medias. Podríamos pensar en usar múltiples tests de t si se satisfacen los supuestos de Normalidad o las correspondientes alternativas no paramétricas. Pero esta forma no es la más eficiente.

¿Qué tipo de problemas vamos a estudiar?

Muchas veces interesa estudiar el efecto de una o más condiciones sobre una variable. En el contexto del análisis de la varianza es habitual llamar "factor" al tipo de condición que interesa estudiar. Esta puede ser **cuantitativa**: puede tomar valores en la recta real (por ejemplo: presión, temperatura), o **cualitativa** (operador).

Ejemplos: se desea comparar

1. La concentración de proteína en una solución para muestras almacenadas con temperaturas diferentes. La variable es la concentración observada. El **factor** es "la temperatura de almacenamiento". Las diferentes temperaturas se denominan **niveles** o **tratamientos**.
2. La concentración media de un analito obtenido utilizando métodos diferentes. El factor en estudio es el método.
3. La media de los resultados obtenidos por cada uno de tres operadores que utilizan los mismos equipos. El factor en estudio es el operador, con tres niveles uno para cada operador.

En todos estos ejemplos hay dos posibles fuentes de variación:

- Debida al factor en estudio.
- Error aleatorio en la medición. Siempre está presente.

en los valores de la variable en estudio, también llamada **variable** de **respuesta**.

El análisis de la varianza (ANOVA) es una técnica que se utiliza para separar y evaluar diferentes causas de variación. En los ejemplos anteriores se lo podría utilizar para separar la variación debida al error aleatorio de la debida a cambios en los niveles del factor.

¿Cómo comparamos las medias de más de dos poblaciones?

Realizaremos:
El Análisis de la Varianza si queremos saber si las medias muestrales de los grupos difieren significativamente o no.
Comparaciones Múltiples si deseamos identificar los grupos para los cuales las medias difieren.

14.2 ANÁLISIS DE LA VARIANZA

Ejemplo: Se realiza un estudio para evaluar la estabilidad de un reactivo en diferentes condiciones de almacenamiento:
A: recientemente preparado
B: una hora en la oscuridad
C: una hora con luz tenue
D: una hora con luz fuerte

	Señales de Fluorescencia observadas en las diferentes condiciones de almacenamiento			
	YA	YB	YC	YD
	102	101	97	96
	100	101	95	92
	101	104	99	94
	101	102	97	
media	101	102	97	94
media general	98.8			

¿Son las diferencias observadas entre las medias muestrales atribuibles a variabilidad aleatoria o son significativas y explicadas por el factor en estudio?

Modelo $Y_{kj} = \mu_k + \varepsilon_{kj}$ variables aleatorias independientes

Donde Y_{kj} = j-ésima observación del k-ésimo nivel del factor $1 \le k \le K$, $1 \le j \le J_k$

μ_k es la media poblacional de las observaciones del grupo k y ε_{kj} son variables aleatorias independientes con varianza σ^2_k

La cantidad total de niveles del factor o grupos a comparar es K, la cantidad de observaciones por grupo (J_k) depende del grupo (k). La cantidad total de observaciones es $n = \sum_{k=1}^{K} J_k$.

Hipótesis a testear

H_0: $\mu_1 = \mu_2 = \mu_3 = ... = \mu_K$ vs. H_a: existe al menos una μ_k diferente

El análisis de la varianza reemplaza a los tests de *t* múltiples por un sólo test de *F* que se realiza bajo los supuestos de igualdad de varianzas y de normalidad de las observaciones.

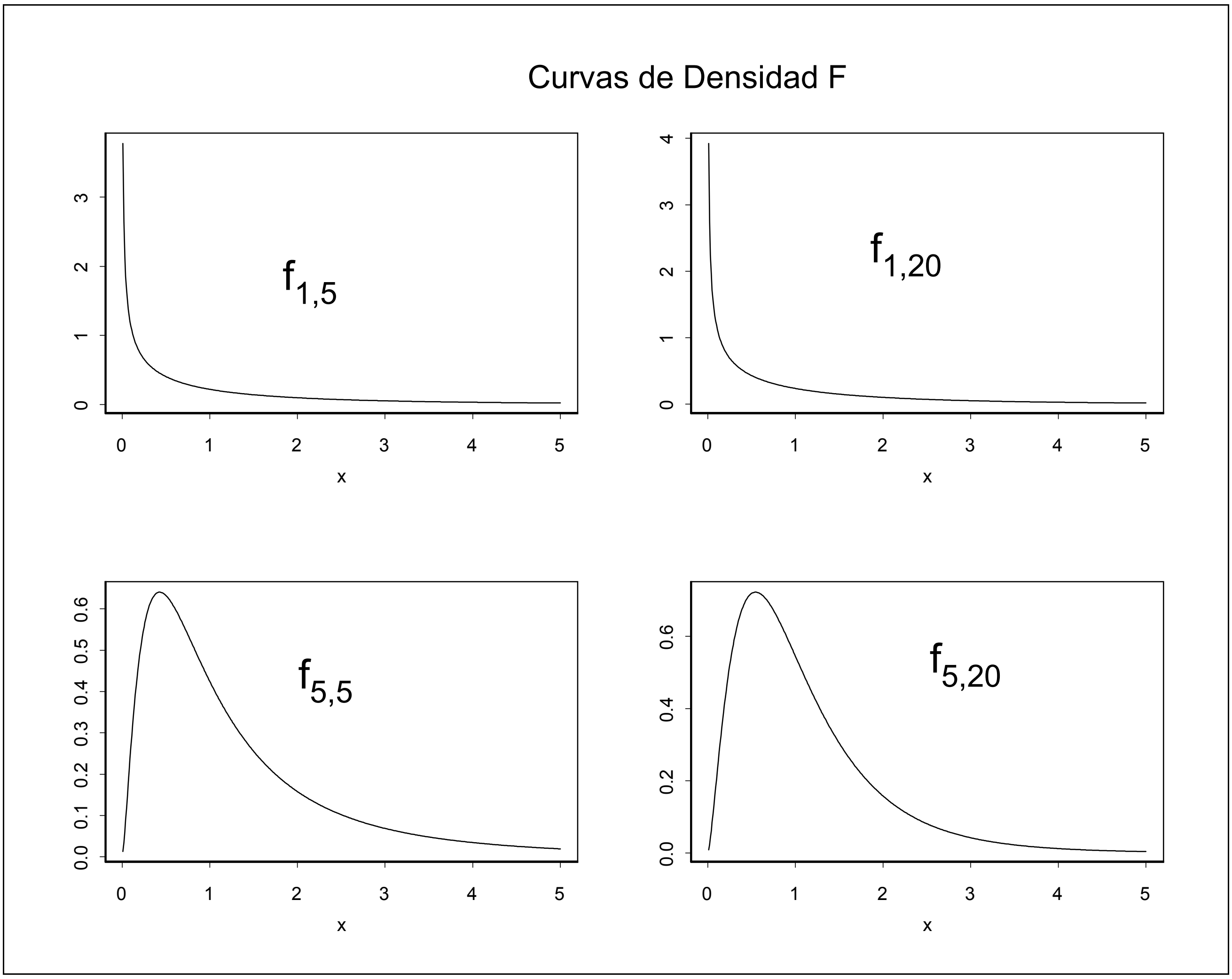
Supuestos

- $\sigma^2_1 = \sigma^2_2 = \sigma^2_3 = \ldots\ldots = \sigma^2_K = \sigma^2$
- Los datos correspondientes a los K niveles del factor son independientes y tienen distribución Normal

$$\left.\begin{array}{llll} & & & \text{media muestral} \\ \text{nivel 1}: Y_{11},\cdots,Y_{1J_1} & \text{i.i.d} & N(\mu_1,\sigma^2) & \overline{Y}_{1.} \\ \text{nivel 2}: Y_{21},\cdots,Y_{2J_2} & \text{i.i.d} & N(\mu_2,\sigma^2) & \overline{Y}_{2.} \\ \text{nivel K}: Y_{K1},\cdots,Y_{KJ_K} & \text{i.i.d} & N(\mu_K,\sigma^2) & \overline{Y}_{K.} \end{array}\right\} \text{independientes}$$

media general $\overline{Y}_{..}$

Estadístico del test $$F = \frac{SS_B/(K-1)}{SS_W/(n-K)} \sim F_{K-1,n-K} \quad \text{bajo } H_0$$



donde

$$SS_B = \sum_{k=1}^{K} J_k (\overline{Y}_{k.} - \overline{Y}_{..})^2, \quad SS_W = \sum_{k=1}^{K} \sum_{j=1}^{J_k} (Y_{kj} - \overline{Y}_{k.})^2 = \sum_{k=1}^{K} (J_k - 1) S_k^2$$

Llamemos $f_{K-1,\, n-K,\, \alpha}$ al valor que deja a la derecha un área α bajo la curva de densidad de la distribución $F_{K-1,\; n-K}$

<u>Región de rechazo de nivel α</u> $F > f_{K-1,n-K,\, \alpha}$

¿De donde surgen los grados de libertad? Se puede demostrar, que si se satisfacen los supuestos del análisis de la varianza resulta:

i) $\dfrac{SS_W}{\sigma^2} \sim \chi^2_{n-K}$ ii) $\dfrac{SS_B}{\sigma^2} \sim \chi^2_{K-1}$ Bajo H_0 y además SS_B y SS_W son independientes

Comentarios

- **Bajo H_0**, todas las observaciones provienen de una misma población Normal con igual media e igual varianza.
- Bajo este supuesto, podríamos medir la variabilidad de todas las observaciones juntas mediante la varianza muestral. Llamamos media general (o gran media) a la media muestral de **todas** las observaciones

$$\overline{Y}_{..}$$

- La varianza muestral usando todas las observaciones es

$$\frac{SST}{n-1} = \frac{(Y_{11} - \overline{\overline{Y}})^2 + ... + (Y_{1J_1} - \overline{\overline{Y}})^2 + ... + (Y_{K1} - \overline{\overline{Y}})^2 + ... + (Y_{KJ_K} - \overline{\overline{Y}})^2}{n-1}$$

  donde **SST** es la **Suma de Cuadrados Total** (Total Sum of Squares).
- Para cada grupo podemos obtener la **varianza muestral** S^2 correspondiente. Así, para el grupo 1 tendríamos:

$$S_1^2 = \frac{(Y_{11} - \overline{Y}_{1.})^2 + (Y_{12} - \overline{Y}_{1.})^2 + ... + (Y_{1J_1} - \overline{Y}_{1.})^2}{J_1 - 1}$$

  donde $\overline{Y}_{1.}$ es la **media muestral** del **grupo 1**. Realizando los mismos cálculos en cada grupo tendremos. $\overline{Y}_1, \overline{Y}_2, ..., \overline{Y}_K$ y S^2_1, S^2_2, ... S^2_K .
- Bajo el supuesto de que <u>todas las varianzas</u> son iguales, podríamos estimar el σ^2 común mediante un promedio pesado de los S^2_k (igual que en el test de t para dos muestras independientes)

$$S_W^2 = \frac{(J_1 - 1)S_1^2 + (J_2 - 1)S_2^2 + ... + (J_K - 1)S_K^2}{J_1 + J_2 + ... + J_K - K} = \frac{SSW}{n-K}$$

  el numerador **SSW** es la **Suma de Cuadrados Dentro** (Sum of Squares Within).
- También podríamos estimar la varianza a partir de los promedios muestrales $\overline{Y}_i$
  SSB la **Suma de Cuadrados Entre** tratamientos (Sum of Squares Between).

$$S_B^2 = \frac{J_1(\overline{Y}_1 - \overline{Y})^2 + J_2(\overline{Y}_2 - \overline{Y})^2 + ... + J_K(\overline{Y}_K - \overline{Y})^2}{K-1} = \frac{SSB}{K-1}$$

- $SST = SSW + SSB$.
  La variación total, medida por la suma de cuadrados total, se separa en las variaciones provenientes de dos fuentes: **entre** grupos y **dentro** de grupos. La suma de cuadrados dentro es también llamada suma de cuadrados del error.
- Si H_0 fuera cierta, $\frac{SSW}{N-k}$ y $\frac{SSB}{k-1}$ deberían ser muy parecidas, por lo tanto el cociente de ambas sumas estaría cercano a 1. Si las medias poblacionales no son todas iguales, generalmente $\mathbf{S^2_B}$ será mayor que $\mathbf{S^2_W}$ y en ese caso el cociente será mayor a 1:

$$F = \frac{S_B^2}{S_W^2} > 1$$

- Bajo H_0 y normalidad, el cociente F sigue una distribución **F** con **K-1** y **n-K** grados de libertad.
- Para determinar si los datos son o no significativos o para calcular el p-valor comparamos con la distribución $F_{K-1,n-K}$.
- si K=2 el estadístico $F = T^2$ (cuadrado del estadístico T del test de t para dos muestras independientes) $T = \frac{(\overline{Y}_{1.} - \overline{Y}_{2.})}{S_p\sqrt{\frac{1}{J_1} + \frac{1}{J_2}}}$

TABLA DE ANÁLISIS DE LA VARIANZA

Source (fuente de variación)	**SS** (suma de cuadrados)	**Df** (grados de libertad)	**MS** (cuadrados medios)	**F**	**p-valor**
Between (entre grupos)	SSB	K-1	SSB / (K-1)	$\frac{SSB / (K-1)}{SSW / (n-K)}$	
Within (dentro de grupos)	SSW	n-K	SSW / (n-K)		
Total	SST	n-1			

¿Con el Statistix? Statistics -> Multi-Sample Tests -> One-Way AOV

Volviendo al ejemplo del estudio para evaluar la estabilidad de un reactivo obtenemos

```
ONE-WAY AOV FOR: XA XB XC XD

SOURCE     DF      SS          MS         F        P
-------   ----  ---------   ---------   ------   ------
BETWEEN     3    142.400     47.4667    21.76   0.0001
WITHIN     11    24.0000     2.18182
TOTAL      14    166.400
```

Rechazamos la hipótesis de igualdad de medias con un p-valor = 0.0001, las medias de las señales de fluorescencia de un reactivo almacenado en distintas condiciones de almacenamiento difieren significativamente. **¿Cuáles son las medias diferentes?**

VARIABLE	MEAN	SAMPLE SIZE	GROUP STD DEV
XA	101.00	4	0.8165
XB	102.00	4	1.4142
XC	97.000	4	1.6330
XD	94.000	3	2.0000
TOTAL	98.800	15	1.4771

Para responder esta pregunta deberemos utilizar métodos de comparaciones múltiples.

14.3 COMPARACIONES MÚLTIPLES

Existen varias alternativas para identificar grupos con distintas medias poblacionales.
En general, lo que hace cada una de ellas es realizar tests de a pares. Veremos algunos de los más usados y difundidos.

14.3.1 Método de Bonferroni

Este método propone la construcción de intervalos de confianza para todos los pares de medias posibles y rechazar su igualdad cuando el correspondiente intervalo no contiene al cero. Si queremos que el **nivel global** del test resultante sea **a lo sumo** α, cada intervalo para $\mu_i-\mu_j$ debería tener un nivel de confianza igual a $1-\alpha^*$, donde

$$\alpha^*=\alpha / m$$

siendo m=K(K-1)/2 cuando realizamos todas las comparaciones posibles.

En nuestro ejemplo como K=4 tenemos que calcular m = 4.3/2=6 intervalos de confianza.

```
BONFERRONI COMPARISON OF MEANS

                            HOMOGENEOUS
VARIABLE        MEAN        GROUPS
---------    ----------    -----------
XB               102.00      I
XA               101.00      I
XC               97.000    .. I
XD               94.000    .. I

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

CRITICAL T VALUE 3.208   REJECTION LEVEL 0.050
STANDARD ERRORS AND CRITICAL VALUES OF DIFFERENCES
VARY BETWEEN COMPARISONS BECAUSE OF UNEQUAL SAMPLE SIZES.
```

El Statistix construye los intervalos de confianza y realiza el test basado en cada uno de los intervalos de a pares. Vemos que las medias de los grupos A y B no difieren significativamente formando un grupo homogéneo, tampoco lo hacen los grupos C y D. Recordemos que A corresponde a las muestras recientemente preparadas y B las almacenadas una hora en la oscuridad. Mientras que los valores de C y D son los obtenidos luego de 1 hora de almacenamiento con luz tenue y fuerte respectivamente. Concluimos que es la exposición a la luz lo que afecta la fluorescencia.

Consideraciones generales

- El procedimiento de comparaciones múltiples que resulta por el método de Bonferroni identifica a μ_i y μ_j como significativamente diferentes si el intervalo para μ_i-μ_j no incluye el 0.
- El Intervalo de Confianza para la diferencia de medias μ_i-μ_j es de la forma

$$\bar{Y}_i - \bar{Y}_j \pm t_{n-K,\alpha^*/2} \sqrt{S_W^2 \left(\frac{1}{J_i} + \frac{1}{J_j}\right)}$$

- Los grados de libertad n-K del intervalo de confianza anterior provienen del estimador de la varianza $\mathbf{S^2_W}$ que se obtiene como parte del análisis de la varianza y utiliza **todos los datos**, no únicamente los de los grupos sobre los que se está realizando la comparación.

14.3.2 Método de Tukey

- Tiene nivel exacto 1-α cuando el tamaño de las muestras es el mismo (J_k = J para todas los grupos). Cuando los tamaños muestrales por grupo no son todos iguales el nivel de confianza es mayor que 1-α. Es decir, el procedimiento de Tukey es conservativo cuando los J_k no son iguales.

- Se basa en el estadístico

$$Q = \max_{i,j} \frac{|\bar{Y}_i - \bar{Y}_j - (\mu_i - \mu_j)|}{\sqrt{S_W^2 \left(\frac{1}{J_i} + \frac{1}{J_j}\right)}}$$

el máximo se toma sobre todos los pares de grupos

- La distribución Q (llamado rango estudentizado), tiene dos parámetros (grados de libertad). Los grados de libertad serán K (cantidad de grupos a comparar) y n-K (grados de libertad de S_W^2). Cuando J_k = J los grados de libertad serán K y n-K=J K - K = K (J-1)

- Cuando J_k = J el intervalo para μ_i -μ_j es de la forma:

$$\bar{Y}_i - \bar{Y}_j \pm Q_{K,n-K,\alpha} \sqrt{S_W^2 / J}$$

Siendo $Q_{K,n-K,\alpha}$ el valor crítico de nivel α de la distribución del rango studentizado Q, es decir que

$$P(Q > Q_{K,n-K,\alpha}) = \alpha$$

- Cuando los J_k son diferentes el intervalo para μ_i -μ_j es de la forma:

$$\overline{Y}_i - \overline{Y}_j \pm \frac{1}{\sqrt{2}} Q_{K,n-K,\alpha} \sqrt{S_W^2 \left(\frac{1}{J_i} + \frac{1}{J_j}\right)}$$

- Cada intervalo que no contiene el 0 da como conclusión que los correspondientes μ_i y μ_j difieren significativamente a nivel α.

Comparación de los métodos

1. Si todos las comparaciones de medias son de interés, el método de Tukey es superior, en tanto da intervalos más cortos.

2. Si sólo algunas comparaciones de a pares son de interés, Bonferroni puede dar mejor.

3. En un problema determinado, para cada método se puede obtener la constante necesaria para construir los intervalos de confianza y luego, elegir la más pequeña. Esto es válido pues no depende de las observaciones, sino que de K y J_i.

14.4 Validación de los supuestos

14.4.1 Homogeneidad de varianzas - homoscedasticidad

La salida de ANOVA del Statistix incluye tres métodos para validar el supuesto de igualdad de varianzas:

- Test de Barlett
- Test de Q de Cochran
- Test de Hartley.

Las hipótesis a testear en este caso son:

Ho: $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2$ vs. Ho: $\sigma_i^2 \neq \sigma_j^2$ algún par de índices $i \neq j$

Test de Barlett. Compara a S^2_W y las S^2_k correspondientes a cada grupo, pero en escala logarítmica. El test se basa en el estadístico $\ln(10)\,\dfrac{q}{c}$

con $$q = (n-K)\log_{10} S_W^2 - \sum_{k=1}^{K}(J_k - 1)\log_{10} S_k^2$$

y $$c = 1 + \frac{1}{3(K-1)}\left(\sum_{k=1}^{K}\frac{1}{J_k - 1} - \frac{1}{n-K}\right),$$

q es grande cuando las varianzas muestrales S^2_k difieren mucho y vale cero cuando las S^2_k son iguales.

La región de rechazo es: $\ln(10)\,\dfrac{q}{c} > \chi^2_{k-1,\ \alpha}$ donde $\chi^2_{k-1,\ \alpha}$ es el valor que deja a la derecha un área α bajo la curva de densidad χ^2 con k-1 grados de libertad.

Este test en muy sensible al alejamiento del supuesto de normalidad.

Test Q de Cochran. Para el caso en que los J_k son todos iguales Cochran propuso el siguiente estadístico:

$$Q = \frac{\max(S_1{}^2, S_2{}^2, ..., S_K{}^2)}{S_1{}^2 + S_2{}^2 + ...S_K{}^2}$$

Se rechaza Ho cuando Q es muy grande, es decir cuando Q es más grande que el percentil 1-α de la distribución de Q cuando las varianzas son iguales (H_0).

Como el anterior es sensible a la falta de normalidad.

Test de Hartley. El estadístico es:

$$F_{\max} = \frac{\max(S_1{}^2, S_2{}^2, ..., S_K{}^2)}{\min(S_1{}^2, S_2{}^2, ..., S_K{}^2)}$$

Región de rechazo: $F_{max} > f_{\max K, \bar{J}-1, \alpha}$

F_{max} tiene una distribución llamada $F_{\max K, \bar{J}-1}$ con dos parámetros K, $\bar{J}$-1
($\bar{J}$ = parte entera (promedio de los J_i)

Este test tiene nivel exacto para igual número de observaciones por muestra y si los tamaños muestrales difieren tiene nivel aproximado.

Este test es sensible a la falta de normalidad.

Para los datos del ejemplo de las condiciones de almacenamiento, no se rechaza el supuesto de homogeneidad de varianzas

```
                    CHI-SQ    DF       P
BARTLETT'S TEST OF  ------  ------  ------
   EQUAL VARIANCES    1.73     3    0.6298

COCHRAN'S Q                     0.4286
LARGEST VAR / SMALLEST VAR      6.0000
```

Test de Levene Modificado. El test de Levene modificado testea la igualdad de varianzas y es resistente a la falta de normalidad. SX no lo calcula directamente, sin embargo se puede calcular fácilmente transformando la variable de respuesta y calculando una nueva Tabla de ANOVA para las variables transformadas. Los pasos a seguir son:

1) Calculamos la mediana de la k-ésimo grupo $\tilde{Y}_k = \underset{j}{med}\ (Y_{kj})$
2) Obtenemos las variables transformadas: $W_{kj} = \left| Y_{kj} - \tilde{Y}_k \right|$
3) Realizamos un ANOVA para W_{kj} .
4) Rechazamos la hipótesis de igualdad de varianzas si el estadístico *F* del paso anterior es grande.

Entre las propuestas para testear homogeneidad de varianzas, este test figura entre los más potentes y resistentes a la violación del supuesto de normalidad.

14.4.2 Normalidad

Hemos visto métodos gráficos (histogramas, box-plots y qq-plots) y analíticos (tests de bondad de ajuste: Shapiro-Wilk) para detectar alejamientos en los datos del supuesto de normalidad. Los alejamientos más importantes a detectar son valores atípicos y asimetría.

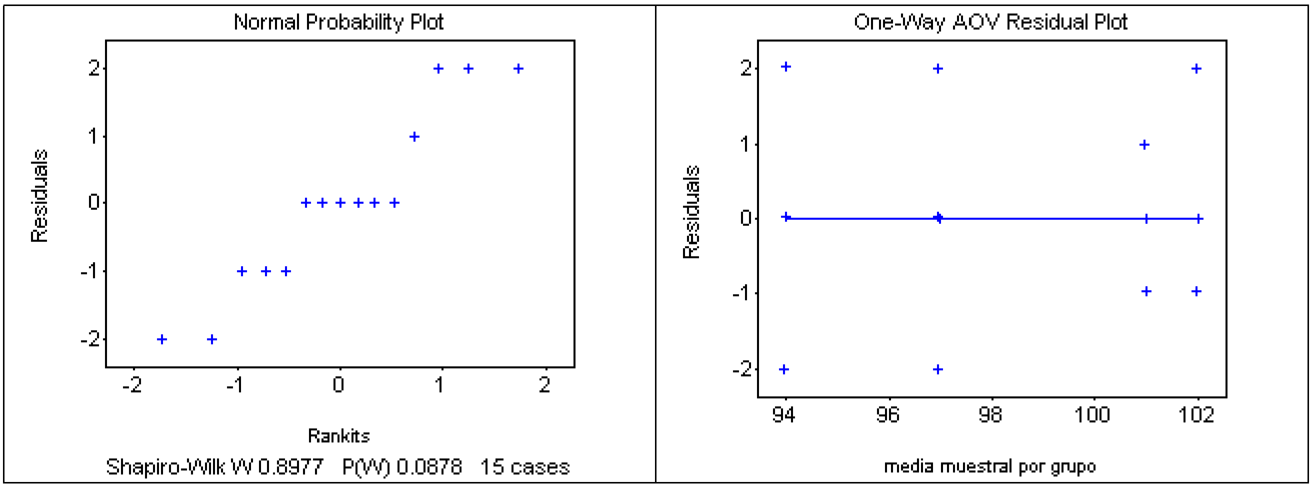
En un análisis de la varianza si se tienen suficientes datos por grupo es posible realizar la validación del supuesto de normalidad en cada uno de ellos. Si no se dispone de suficiente información los controles se realizan utilizando todos los datos juntos. Para ello es necesario restarle a cada dato la media correspondiente al grupo al cual pertenece. Esto es realizar la validación de los supuestos utilizando los **residuos**:

El **residuo** de la observación j es la diferencia entre su valor y la media muestral del grupo k al cual pertenece:

$$r_{kj} = Y_{kj} - \overline{Y}_{k.}$$

El Statistix genera por defecto dos gráficos de los residuos a posteriori del ANOVA
Results -> plots

A continuación se muestran los gráficos para el ejemplo de los 4 métodos de almacenamiento



El estadístico de Shapiro-Wilk (p-valor = 0.0878 > 0.05) y el qq-plot no muestran alejamientos importantes de la normalidad, aunque llaman la atención los valores repetidos y el p-valor que no es demasiado alto. El diagrama de dispersión de los residuos en función de la media muestral por grupo permite una evaluación visual del supuesto de homogeneidad de varianzas. Tampoco aquí se destacan alejamientos del supuesto.

Veamos otro ejemplo

Interesa comparar el volumen espiratorio forzado en 1 segundo (FEV), en pacientes con enfermedad arterial coronaria que provienen de 3 centros médicos. La pregunta planteada es: *¿Hay alguna diferencia entre las medias de FEV de estos tres centros?*

La tabla siguiente muestra el volumen espiratorio forzado en un segundo de pacientes de los tres centros y algunas medidas de resumen

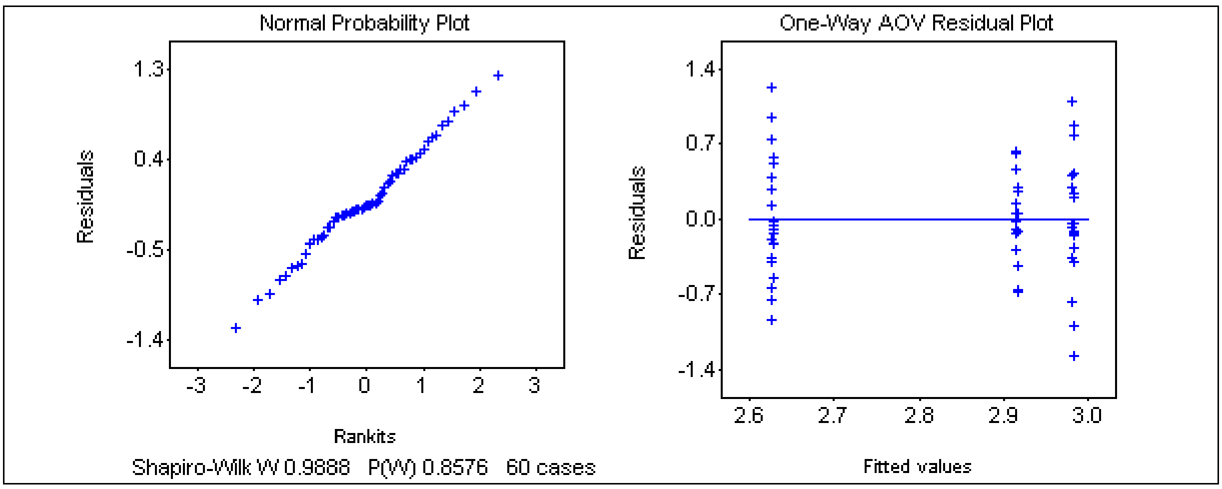
FEV1	FEV2	FEV3
3.13	3.17	3.53
3.57	3.22	2.79
1.86	2.88	3.22
2.27	1.71	2.25
3.01	2.92	2.98
1.69	3.77	2.47
2.40	3.29	2.77
2.51	3.39	2.95
3.86	3.86	3.56
3.36	2.94	2.88
2.56	2.61	2.63
2.55	2.71	3.38
1.98	3.41	3.07
2.57	2.89	2.81
2.08	2.59	3.17
2.47	3.39	2.23
2.47	2.19	M
2.74	4.06	M

	FEV1	FEV2	FEV3
N	23	21	16
MEAN	2.6278	2.9829	2.9181
SD	0.5386	0.5892	0.4031
MINIMUM	1.6900	1.7100	2.2300
1ST QUARTI	2.2700	2.6600	2.6650
MEDIAN	2.5500	2.9200	2.9150
3RD QUARTI	3.0100	3.3900	3.2075
MAXIMUM	3.8600	4.0600	3.5600
MAD	0.3200	0.3300	0.2700

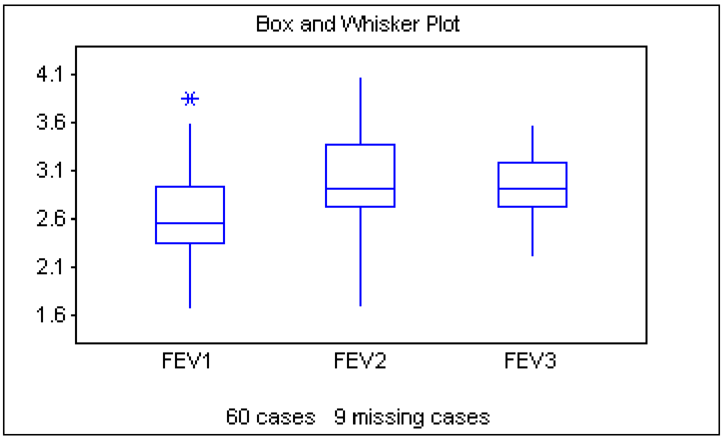
```
2.88       1.98          M
2.60       2.81          M
2.45       2.85          M
2.23          M          M
3.20          M          M
```

```
STATISTIX 7.0                                        FEV, 06/06/05, 01:05:50 p.m.
volumen espiratorio forzado en 1 segundo

ONE-WAY AOV FOR: FEV1 FEV2 FEV3

SOURCE     DF        SS          MS          F        P
-------  ----  ----------  ----------  ------  ------
BETWEEN     2     1.55502     0.77751     2.81  0.0685 ⇐  mayor a 0.05

WITHIN     57     15.7621     0.27653
TOTAL      59     17.3171

                            CHI-SQ    DF        P
BARTLETT'S TEST OF          ------  ------  ------
   EQUAL VARIANCES            2.32       2  0.3138

COCHRAN'S Q                         0.4341
LARGEST VAR / SMALLEST VAR          2.1366

COMPONENT OF VARIANCE FOR BETWEEN GROUPS    0.02532
EFFECTIVE CELL SIZE                            19.8


                               SAMPLE     GROUP
VARIABLE          MEAN          SIZE     STD DEV
---------    ----------    ------    ----------
FEV1             2.6278        23        0.5386
FEV2             2.9829        21        0.5892
FEV3             2.9181        16        0.4031
TOTAL            2.8295        60        0.5259

CASES INCLUDED 60    MISSING CASES 9
```

No se detectan diferencias significativas al 5%, mediante el ANOVA.



No se detectan alejamientos graves de los supuestos de Normalidad y homoscedasticidad. Sin embargo, al realizar los box-plots por centro, encontramos un outlier moderado en el primer centro médico. Si corresponde a un paciente que pertenece a la población en estudio y no hay razones para eliminarlo, ¿cuál es la estrategia adecuada de análisis?



Alternativa no paramétrica: Test de Kruskal-Wallis.

El test de Mann-Witney para dos muestras independientes fue extendido al problema del análisis de K muestras independientes, $K \geq 2$ por Kruskall y Wallis (1952). Interesa testear la <u>hipótesis nula</u> de que la distribución de la variable de interés es la misma en todas las (**K**) poblaciones contra la <u>alternativa</u> de en algunas de las poblaciones la variable tiende a dar valores mayores que en otras.

<u>Supuestos</u>

- Los datos son por lo menos ordinales. Esto significa que las observaciones pueden ordenarse en un orden creciente de acuerdo con alguna propiedad. Las variables continuas satisfacen naturalmente esta condición. Las variables categóricas cuyas categorías pueden ordenarse (por ejemplo por calidad, gravedad ó valor), también.
- Además de la independencia entre las observaciones de una misma muestra suponemos independencia entre las observaciones de las distintas muestras.

Igual que en el anova, de cada una de las K poblaciones obtenemos una muestra aleatoria de tamaño J_k, y observamos la variable de interés en dicha muestra es decir:

Muestra de la población k	Tamaño de muestra
$Y_{11}, Y_{12}, \ldots\ldots\ldots, Y_{1\,J1}$	J_1
$Y_{21}, Y_{22}, \ldots\ldots\ldots, Y_{2\,J2}$	J_2
....	.....
$Y_{k1}, Y_{k2}, \ldots\ldots\ldots, Y_{K\,Jk}$	J_k
Total de observaciones	$n = J_1 + J_2 + \ldots.\ J_k$

<u>Hipótesis nula</u> H_0 : la variable observada tiene la misma distribución en todas las poblaciones.

<u>Hipótesis alternativa</u> H_a: por lo menos en una población, la variable observada tiende a dar valores mayores que en alguna otra población.

Bajo H_0 , todas las observaciones (Y_{kj}) tienen la misma distribución, si las tomamos todas juntas en una única muestra de tamaño n y las ordenamos de menor a mayor, obteniendo los rangos R_{kj} todas las asignaciones de los rangos a las K muestras tienen la misma chance de ocurrir.

Si H_0 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 $\overline{R}_{k.}$ para cada tratamiento k y el valor que esperaríamos si H_0 fuera cierta.

<u>Estadístico del test</u>: $$G_{KW} = \frac{SSB}{\frac{SST}{n-1}}$$

donde SSB y SST son, respectivamente, la *suma de cuadrados between* y la *suma de cuadrados total* para la tabla de análisis de la varianza correspondiente a los <u>rangos de las observaciones</u>.

<u>Región de rechazo</u> G_{KW} es grande.

El SX nos da el valor del estadístico G_{KW}.
SX nos da el p-valor usando la aproximación por una distribución:

$$\chi^2_{K-1}$$

Esta aproximación es válida cuando:

K=3 $J_{ik} \geq 6$ para las k muestras

o bien

K>3 $J_k \geq 5$ para las K muestras

Para el caso en que K=3 y los $J_k \leq 5$ adjuntamos una tabla con los percentiles de la distribución exacta.

Veamos salidas de SX para el ejemplo de FEV

```
volumen espiratorio forzado en 1 segundo

KRUSKAL-WALLIS ONE-WAY NONPARAMETRIC AOV

             MEAN    SAMPLE
VARIABLE     RANK     SIZE
---------   ------   ------
FEV1         23.2      23
FEV2         35.9      21
FEV3         33.9      16
TOTAL        30.5      60

KRUSKAL-WALLIS STATISTIC                      6.5695
P-VALUE, USING CHI-SQUARED APPROXIMATION      0.0374  ⇐  menor a 0.05


PARAMETRIC AOV APPLIED to RANKS

SOURCE    DF       SS         MS        F       P
-------  ----  ---------  ---------  ------  ------
BETWEEN    2     2002.86    1001.43    3.57  0.0346
WITHIN    57     15984.6    280.432
TOTAL     59     17987.5

TOTAL NUMBER OF VALUES THAT WERE TIED   20
MAX. DIFF. ALLOWED BETWEEN TIES    0.00001

CASES INCLUDED 60    MISSING CASES 9
```

Se rechaza la hipótesis nula a favor de la alternativa en algún centro la FEV tiende a dar valores más altos que en algún otro.

La salida de SX incluye también 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:

$$F_R = \frac{(n-K)G_{KW}}{(K-1)(n-1-G_{KW})}$$

Comparaciones Múltiples

Como ocurre con el test de F cuando encontramos diferencias significativas, en general será de interés identificar cuales son las poblaciones que tienen una distribución diferente.

Un método válido para muestras grandes equivalente al procedimiento de Bonferroni para comparaciones múltiples puede emplearse en este caso, siempre que el tamaño de las muestras no sea muy pequeño. Los límites de los intervalos de confianza para todas las m=$\frac{K*(K-1)}{2}$ comparaciones usando los rangos promedios $\overline{R}i.$ pueden calcularse como

$$\overline{R}_{i.} - \overline{R}_{j.} \pm z_{\alpha/2m} \left[\frac{n(n+1)}{12} \left(\frac{1}{J_i} + \frac{1}{J_j} \right) \right]^{1/2}$$

Volviendo al ejemplo, a un nivel del 5% obtenemos:

```
volumen espiratorio forzado en 1 segundo

COMPARISONS OF MEAN RANKS

                MEAN    HOMOGENEOUS
VARIABLE        RANK    GROUPS
---------  ----------  -----------
FEV2          35.881     I
FEV3          33.875     I I
FEV1          23.239    .. I

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

Concluimos al 5%, que los valores de FEV de los pacientes del centro 2 tienden a ser más altos que los de los pacientes del centro 1.