## Diagnóstico en ANOVA

En el caso de Anova 1 Factor computando los residuos $r_{ij}$ una vez calculada la Tabla ANOVA podemos detectar:

- Heterogeneidad de varianzas
- Falta de independencia entre las observaciones
- Presencia de outliers
- Omisión de alguna variable importante
- Falta de Normalidad

Podemos investigar la distribución de los residuos a través de diagramas de tallo-hoja, histogramas, box-plots. Podemos detectar asimetría, presencia de outliers, etc.

Si el tamaño de $n_i$ es razonable, es aconsejable realizarlos para cada nivel del factor. Recordemos que como en regresión, los residuales no son independientes. En general, esta dependencia suele ser despreciable.

Si el tamaño de $n_i$ es razonable, podemos chequear el supuesto de normalidad realizando qq-plots y aplicando el test de Shapiro-Wilk para la observaciones originales en cada nivel. Si no es así, los haremos para todos los residuos juntos.

Si detectásemos residuos grandes o alejados del grueso de los residuos deberíamos estudiar cuidadosamente la situación.

## Otros gráficos

- Diagrama de puntos: se construyen graficando los residuos (o las observaciones originales) de cada nivel del factor en paralelo y nos darán una idea de si el supuesto de homogeneidad de varianzas entre los niveles es razonable o no.
- Valores predichos $\overline{Y}_i$ vs. Residuos: en este gráfico podemos apreciar la bondad del ajuste del modelo y las varianzas de los residuos.
- Gráfico de residuos vs. secuencia temporal: si se tiene registrado el orden en que fueron tomadas las observaciones es aconsejable hacer este gráfico con el fin de detectar alguna tendencia.

- Gráfico de residuales vs. alguna variable de interés: si se midió alguna otra variable (edad, peso, etc) puede ser útil graficar los residuos vs. esta variable. Esto puede contribuir a:

  1. la comprensión del problema
  2. sugerir variables a controlar en una nueva experiencia
  3. ayudar a detectar un factor confundido si no se aleatorizó correctamente.

Para detectar heterogeneidad de varianzas en este modelo existen varios tests específicos cuando la distribución de los datos es normal. Veremos una opción, que es la del Test de Levene, que es válida en un contexto más general.

Supongamos que tenemos un **Anova 1 Factor** en el que comparamos k tratamientos.

Las hipótesis a testear son:

$$\boxed{\mathbf{Ho:}\ \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 \quad \text{vs.} \quad \mathbf{H_1:}\ \sigma_i^2 \neq \sigma_j^2 \text{ para algún par de índices } i \neq j}$$

## Test de Levene Modificado

El test de Levene modificado testea la igualdad de varianzas. Puede calcularse fácilmente transformando la variable de respuesta y calculando una nueva Tabla de ANOVA para las variables transformadas.

Los pasos a seguir son:

0) Computamos la mediana de la i-ésima casilla $\tilde{Y}_i = \underset{j}{med}(Y_{ij})$

1) Calculamos las variables transformadas: $W_{ij} = \left| Y_{ij} - \tilde{Y}_i \right|$

2) Calculamos la Tabla de ANOVA para $W_{ij}$

3) 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.

***Si se rechaza la hipótesis de igualdad de varianzas, tenemos algunas alternativas.***

Si la varianza no es constante, pero se sustenta el supuesto de normalidad, es recomendable usar mínimos cuadrados ponderados o pesados.

Muchas veces la heterogeneidad de varianzas está acompañada por la no normalidad de las observaciones. En este caso, la transformación de la variable de respuesta suele ser un remedio.

Con frecuencia, la misma transformación que estabiliza las varianzas también corrige la falta de normalidad de los datos.

Si esto no se puede lograr, puede combinarse una transformación estabilizadora de varianzas con una alternativa al test de F que sea no paramétrica.

Una posibilidad para encontrar la función transformadora es realizar un gráfico de $\overline{Y}_i$ vs. $S_i$ para visualizar qué tipo de relación tienen.

Veamos un ejemplo.

## Transplante de Corazón

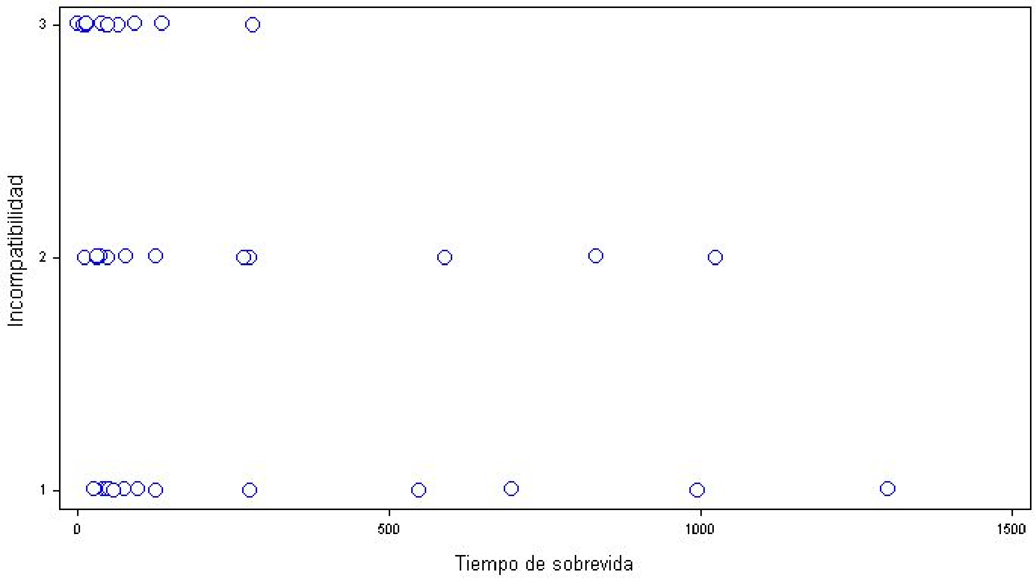
En los transplantes de corazón la similitud entre el tipo de tejido del donante y del receptor es importante, pues grandes diferencias aumentan la probabilidad de rechazo del corazón transplantado. Los datos que analizaremos a continuación corresponden al tiempo de supervivencia de 36 pacientes transplantados. Los datos fueron agrupados en tres categorías de acuerdo con el grado de incompatibilidad entre el tejido del donante y del receptor (baja=1, media=2 y alta=3). Los investigadores quieren determinar si el tiempo medio de sobrevida depende del grado de incompatibilidad.
Las hipótesis a testear son:

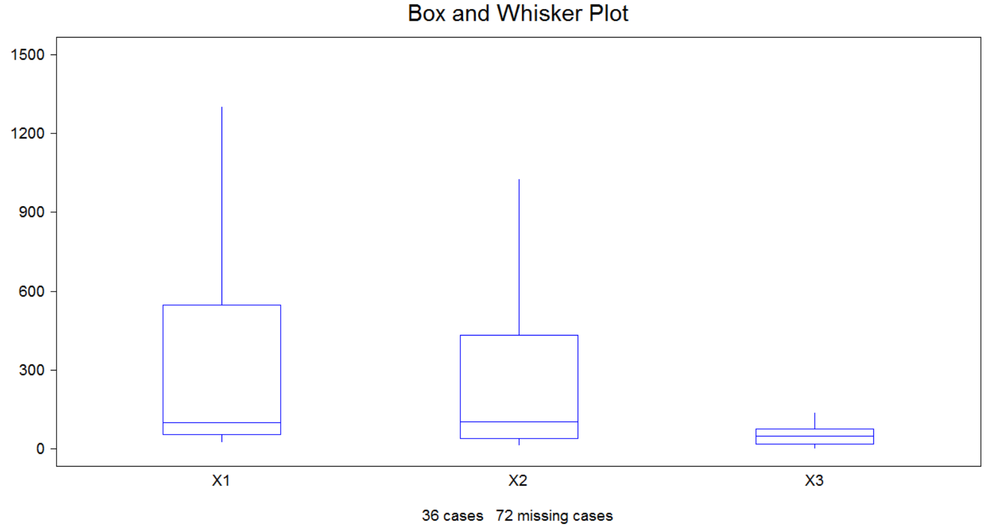
$H_o$: $\mu_1 = \mu_2 = \mu_3$    vs.    $H_1$: no todas las $\mu_i$ son iguales

**Diagrama de Puntos**

El diagrama de puntos sugiere que el tiempo de sobrevida puede disminuir cuando crece la incompatibilidad.

Modelo Lineal  Susana Sombielle FCEyN 2017






Inicialmente se realizó un ANOVA y se calcularon los residuos con fines de diagnóstico. A continuación ofrecemos la salida y algunos gráficos.

| SOURCE | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| BETWEEN | 2 | 455385 | 227693 | 2.13 | 0.1351 |
| WITHIN | 33 | 3530419 | 106982 | | |
| TOTAL | 35 | 398580 | | | |

| Incompat. | MEAN | SAMPLE SIZE | GROUP STD DEV |
|---|---|---|---|
| 1 | 334.92 | 13 | 421.99 |
| 2 | 281.08 | 12 | 347.32 |
| 3 | 69.818 | 11 | 81.607 |
| TOTAL | 235.97 | 36 | 327.08 |

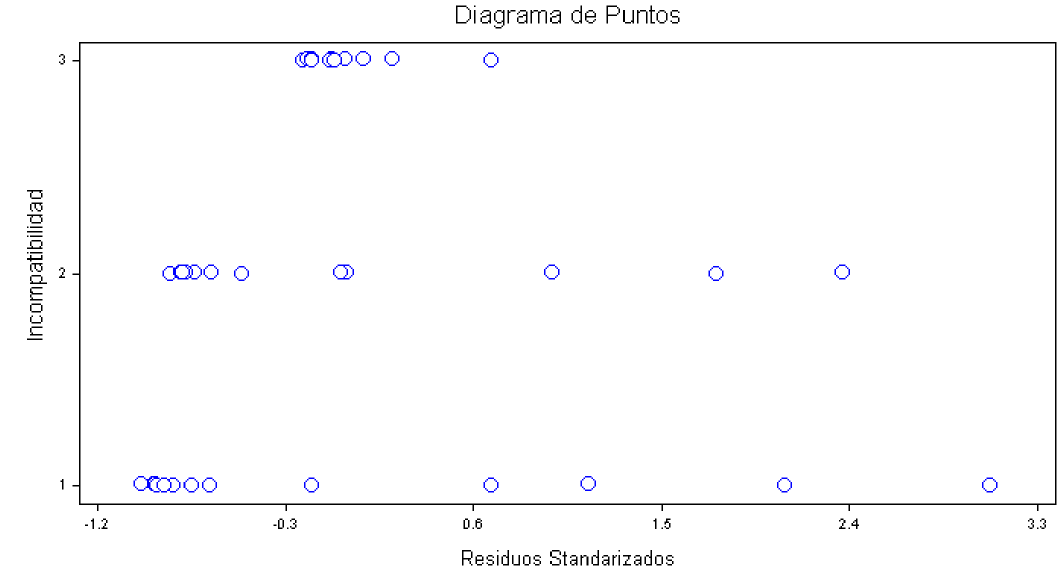
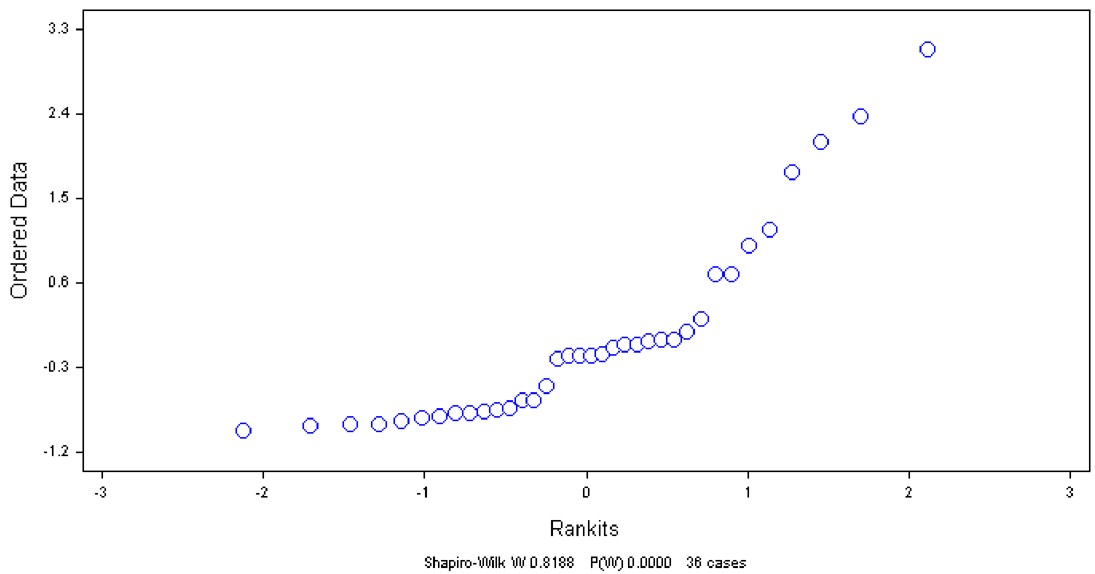


Diagrama de Puntos
Incompatibilidad
3
2
1
-1.2
-0.3
0.6
1.5
2.4
3.3
Residuos Standarizados

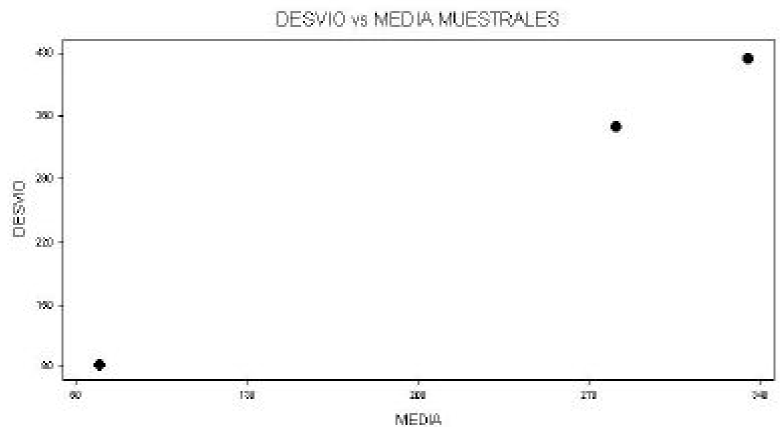

Ordered Data
3.3
2.4
1.5
0.6
-0.3
-1.2
-3
-2
-1
0
1
2
3
Rankits
Shapiro-Wilk W 0.8188    P(W) 0.0000    36 cases

El diagrama de puntos de los residuos standarizados sugiere que la distribución de los residuos es asimétrica a derecha y que la varianza de los residuos podría ser menor cuando hay una alta incompatibilidad.

El test de Levene modificado fue aplicado obteniéndose un p-valor igual a 0.1504.

Por otro lado, el qq-plot de todos los residuos standarizados revela cierta asimetría a derecha y el test de Shapiro-Wilk tiene un p-valor menor que 0.0001.

Si realizamos un scatter plot de $\overline{Y}_i$ vs $s_i$ parece haber una relación lineal entre ambos.



Trabajamos con la nueva variable $Y' = \log(Y)$ y realizamos el análisis de la varianza para ella.

Tabla de ANOVA

| SOURCE | DF | SS | MS | F | P |
|---|---|---|---|---|---|
| BETWEEN | 2 | 12.9734 | 6.48670 | 3.57 | 0.0394 |
| WITHIN | 33 | 59.9250 | 1.81591 | | |
| TOTAL | 35 | 72.8984 | | | |

| IND | MEAN | SAMPLE SIZE | GROUP STD DEV |
|---|---|---|---|
| 1 | 5.0174 | 13 | 1.3338 |
| 2 | 4.8098 | 12 | 1.4213 |
| 3 | 3.6281 | 11 | 1.2790 |
| TOTAL | 4.5237 | 36 | 1.3476 |

El p-valor del test de Levene modificado para la variable transformada es 0.7282. El diagrama de puntos y el qq-plot de los residuos standarizados (p-valor del test de Shapiro -Wilk = 0.1463) también sugieren que la transformación logarítmica es apropiada.

En la tabla de ANOVA vemos  que el estadístico F = 3.57 con un p-valor = 0.0394. Para un nivel $\alpha$=0.05 concluiríamos que la media del logaritmo del tiempo de sobrevida de los transplantados depende del grado de incompatibilidad del tejido entre donante y receptor.

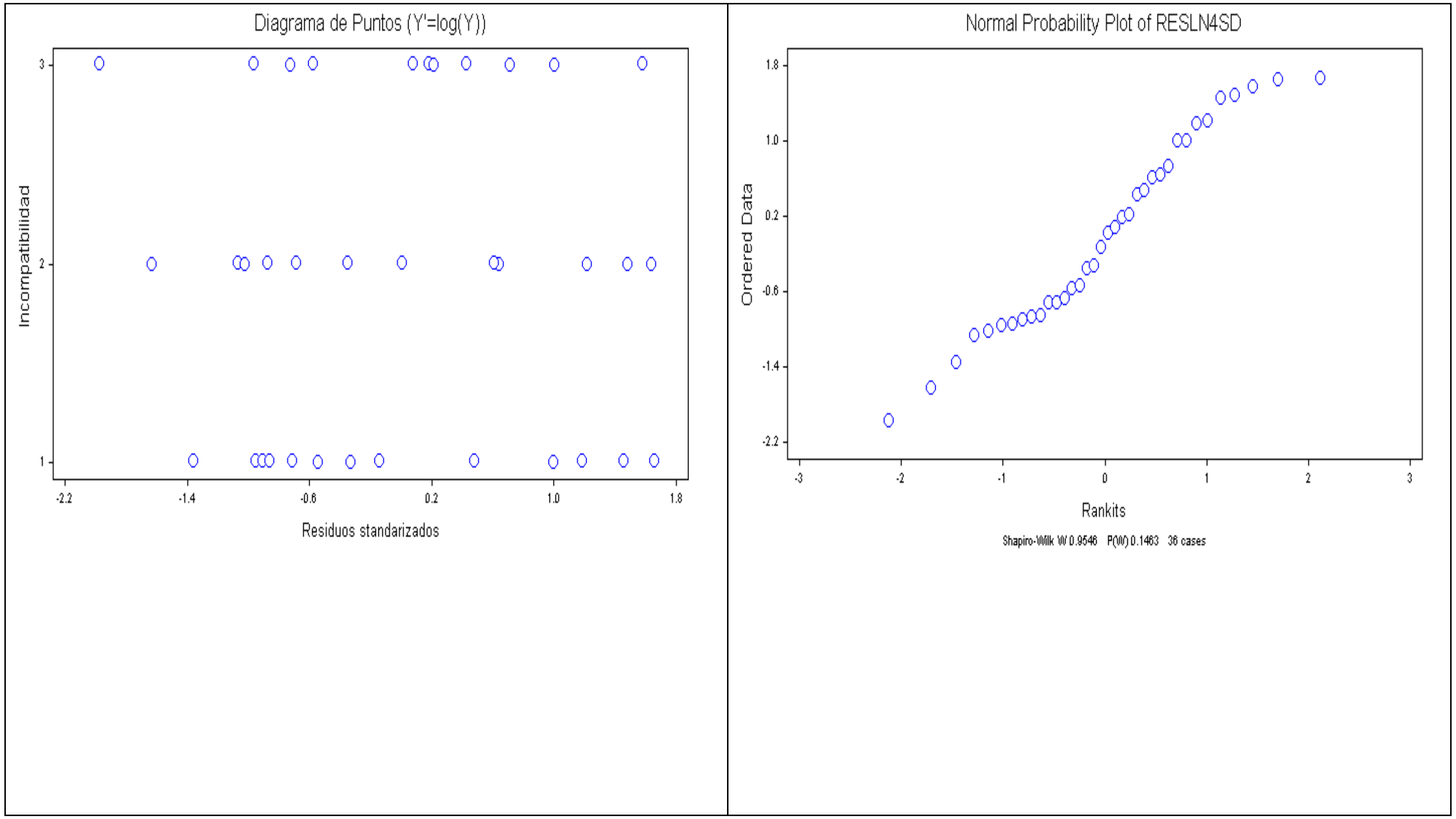


Diagrama de Puntos (Y'=log(Y))
Incompatibilidad
3
2
1
-2.2
-1.4
-0.6
0.2
1.0
1.8
Residuos standarizados
Normal Probability Plot of RESLN4SD
Ordered Data
1.8
1.0
0.2
-0.6
-1.4
-2.2
-3
-2
-1
0
1
2
3
Rankits
Shapiro-Wilk W 0.9546   P(W) 0.1463   36 cases