

Anova unifactorial.

Verificación de las condiciones Anova

Comparación por pares

Grados de Biología y Biología sanitaria

M. Marvá

e-mail: marcos.marva@uah.es

Unidad docente de Matemáticas, Universidad de Alcalá

28 de noviembre de 2017

El problema Anova

Estudiamos la relación

$$C \sim F$$

donde

C variable respuesta cuantitativa

F variable explicativa cualitativa; factor con k niveles

Los individuos de cada nivel del factor se representan por una variable aleatoria X_i que deben verificar

- 1 Independencia de las muestras
- 2 Normalidad: $X_i \sim N(\mu_i, \sigma)$ para cada nivel i del factor
- 3 Homocedasticidad

Verificación de las condiciones Anova

1. Independencia de las muestras

- Depende de un diseño experimental correcto.
- Sabemos poco sobre diseño de experimentos,
supondremos que esa independencia está garantizada

Verificación condiciones Anova: normalidad

2. Normalidad: $X_i \sim N(\mu_i, \sigma)$ para cada nivel i del factor

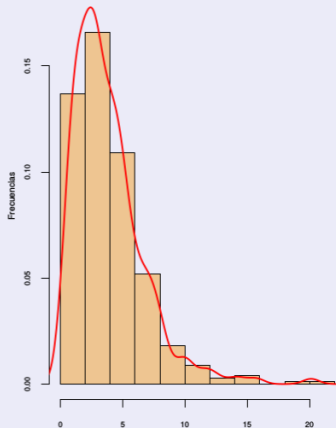
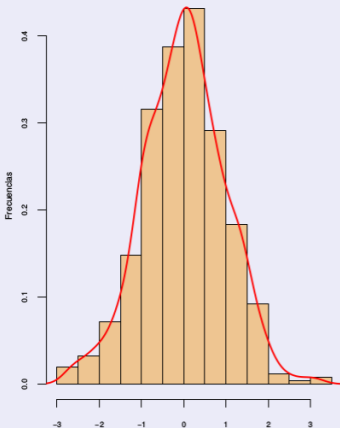
Es la más difícil de verificar (sobre todo con muestras pequeñas)

- 1 Muchas variables se distribuyen normalmente (otras muchas no)
- 2 El contraste Anova de un factor es **robusto** frente a desviaciones moderadas respecto a la normalidad.
- 3 **Análisis exploratorio** (ojo a la escala): histogramas, boxplot y qq-plots de cada grupo, compararlos con los de una población normal
- 4 Para muestras pequeñas, histogramas, boxplots, etc. no ayudan

Verificación condiciones Anova: normalidad

¿Cómo podemos analizar la normalidad de una población?

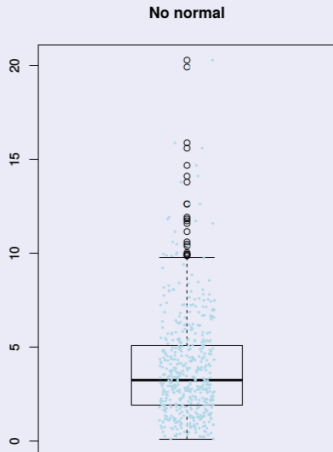
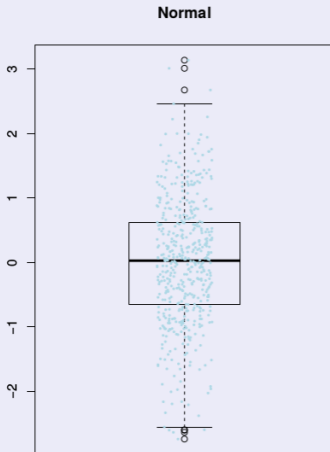
En **muestras grandes** podemos usar *histogramas* y *curvas de densidad*. Las figuras muestran a la izquierda datos normales y a la derecha datos no normales, con $n = 500$ en ambos casos.



Verificación condiciones Anova: normalidad

Boxplots para analizar la simetría.

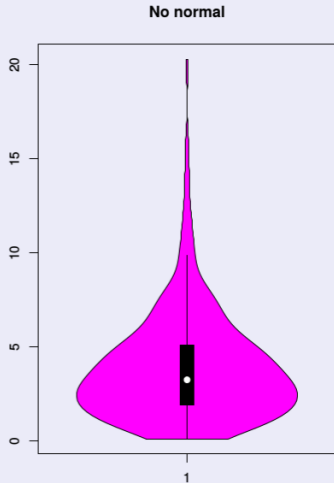
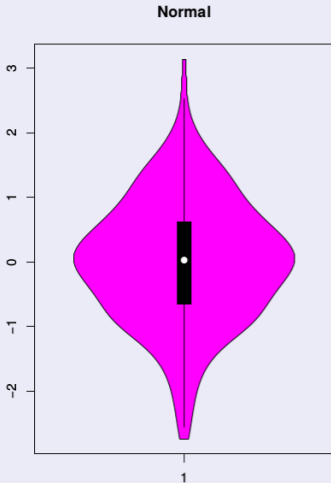
A menudo la simetría es el requisito más importante para que los métodos de la Estadística funcionen. Los boxplots son especialmente útiles para detectar la falta de simetría (¡añade los puntos!).



Verificación condiciones Anova: normalidad

Violinplot.

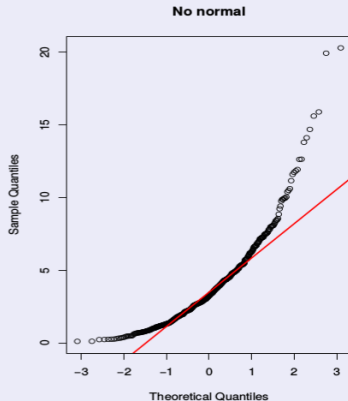
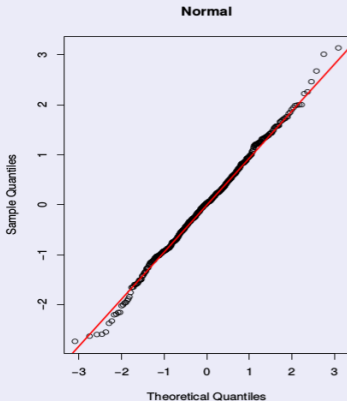
Pero recuerda que este tipo de gráficos son también interesantes y muchas veces más completos que el boxplot.



Verificación condiciones Anova: normalidad

QQplot

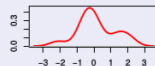
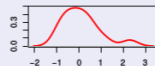
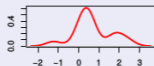
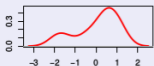
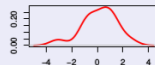
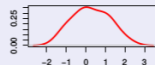
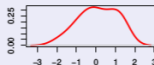
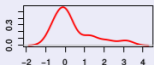
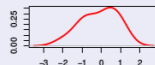
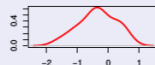
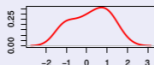
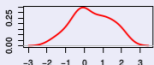
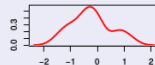
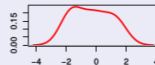
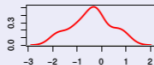
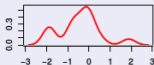
- Son, probablemente, los más utilizados para el análisis de la normalidad.
- Compara los percentiles de la muestra con los de una normal.
- Si la muestra procede de una población normal, los puntos coinciden con la recta.



Verificación condiciones Anova: normalidad

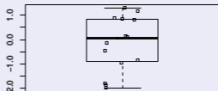
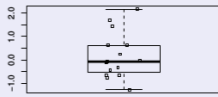
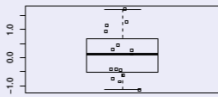
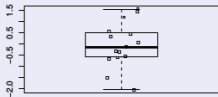
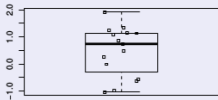
¡¡Precaución con las muestras pequeñas!!

Todas las siguientes gráficas son de muestras de tamaño 15 que **proviene de poblaciones normales**.



Verificación condiciones Anova: normalidad

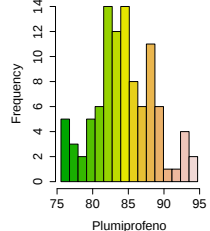
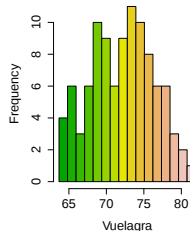
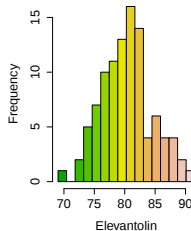
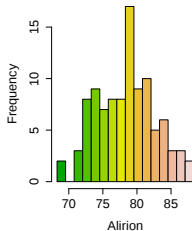
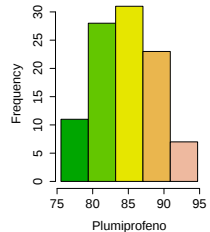
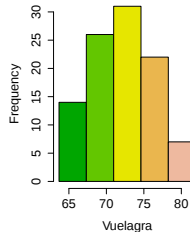
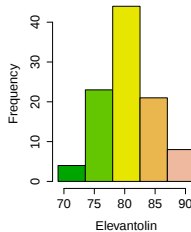
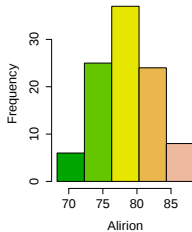
Y con los boxplots sucede algo parecido.



Verificación condiciones Anova: normalidad

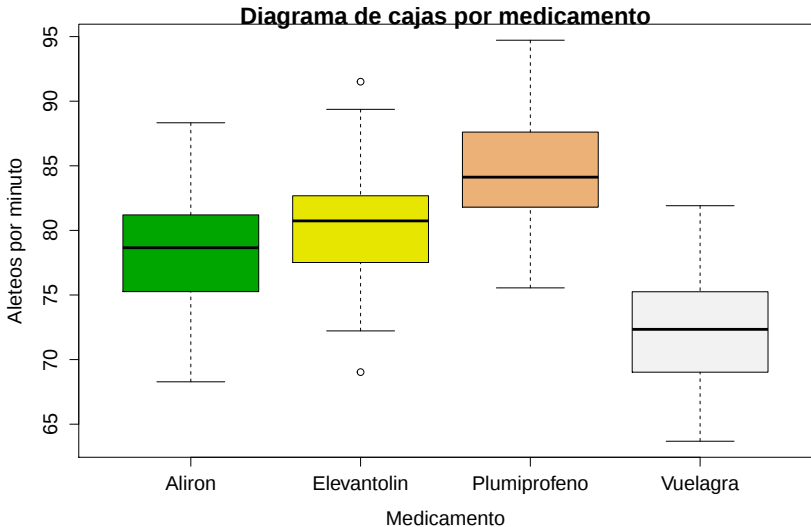
2. Normalidad: $X_i \sim N(\mu_i, \sigma)$ para cada nivel i del factor

Ejemplo: Histogramas para los datos de los frailecillos y con este código



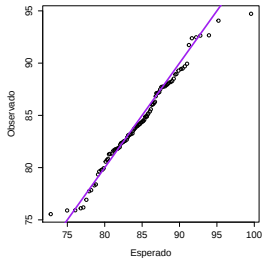
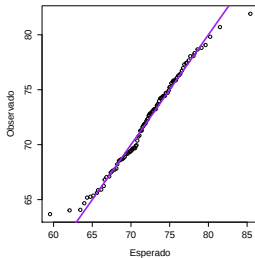
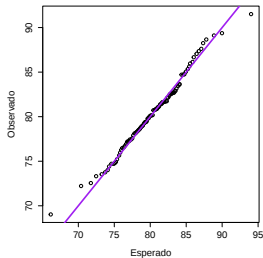
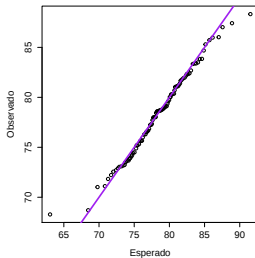
Verificación condiciones Anova: normalidad

Ejemplo: Boxplot para los datos de los frailecillos y con [este código](#)



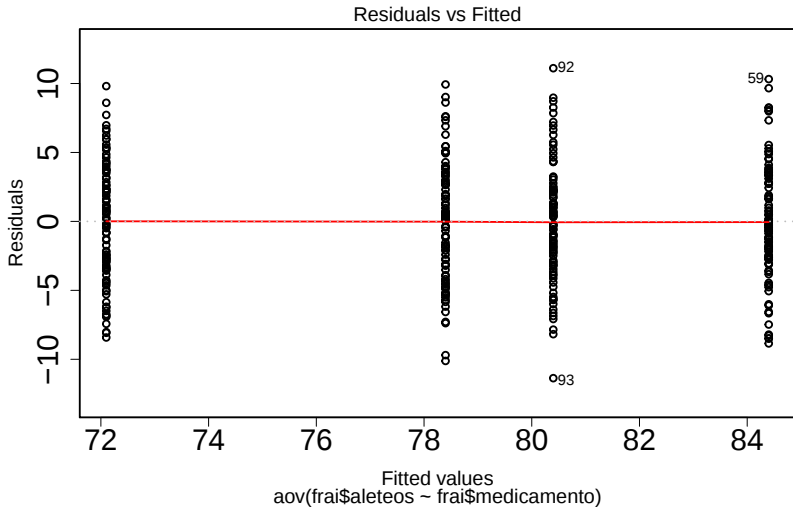
Verificación condiciones Anova: normalidad

Ejemplo: QQ-plots para los datos de los frailecillos y con este código



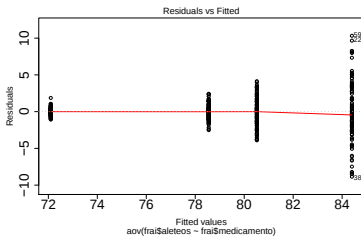
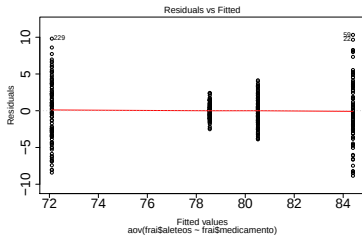
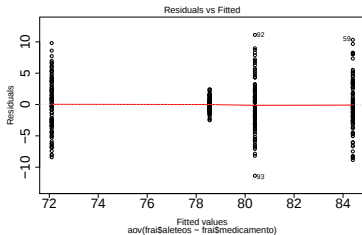
Verificación condiciones Anova: homocedast

Representar residuos por medicamento



Verificación condiciones Anova: homocedast

Tres situaciones DIFERENTES en las que no hay homocedasticidad



Anova significativo: comparación 2 a 2

Comparación ingenua: contraste de diferencia de medias para cada pareja de tratamientos

- La comparación de cada par de niveles implica testar

H_0 : {Dos medias son iguales} H_0 : {Esas medias NO son iguales}

con probabilidad de error de tipo I igual a α .

- La variable

$X = \text{"n"}^0$ de comparaciones en las que se comete error de tipo I"

es una variable binomial de parámetros $n = k$ y $p = \alpha$.

- La probabilidad de equivocarse (rechazar alguna H_0 siendo cierta) es

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (1 - \alpha)^k$$

Ejemplos: para $\alpha = 0.05$

- Factor con 6 niveles, $P(X \geq 1) = 1 - (0.95)^6 \approx 0.27$
- Factor con 14 niveles, $P(X \geq 1) = 1 - (0.95)^{14} \approx 0.51$

Anova significativo: comparación 2 a 2

Ajuste de Bonferroni (detalles en sección 11.6.1 del libro)

Si hay k niveles, **repartir el error de tipo I entre las**

$$\frac{k!}{2!(k-2)!}$$

comparaciones 2 a 2 posibles usando la varianza muestral ponderada

$$s_{\text{pond}}^2 = \frac{SS_{\text{residual}}}{N - k}$$

y el estadístico

$$\Upsilon = \frac{\bar{X}_{\cdot i} - \bar{X}_{\cdot j}}{\sqrt{s_{\text{pond}}^2 \cdot \left(\frac{1}{n_i} + \frac{1}{n_j} \right)}} \sim t_{N-k}$$

para comparar el nivel i con el nivel j

$$\text{p-valor} = \min \left\{ \binom{k}{2} \cdot 2 \cdot P(T_{N-k} > |\Upsilon|), 1 \right\}$$

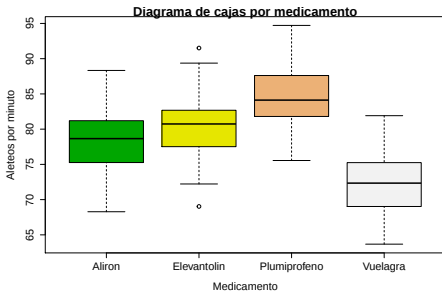
Anova significativo: comparación 2 a 2

Ejemplo Salida R, ajuste de Bonferroni para los datos de los frailecillos

	Aliron	Elevantolin	Plumiprofeno
Elevantolin	0.005	-	-
Plumiprofeno	<2e-16	1e-09	-
Vuelagra	<2e-16	<2e-16	<2e-16

P value adjustment method: bonferroni

Todas las diferencias son significativas; podemos ordenar las medias



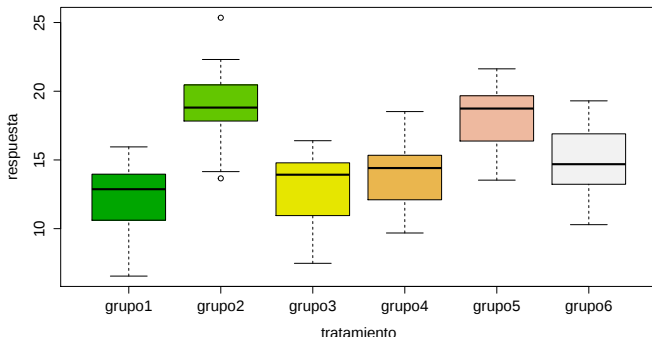
Anova significativo: comparación 2 a 2

Ejemplo Salida R, otras situaciones. Para estos **datos**

	grupo1	grupo2	grupo3	grupo4	grupo5
grupo2	3.83e-10	-	-	-	-
grupo3	1.000000	1.34e-10	-	-	-
grupo4	0.175126	9.59e-09	1.000000	-	-
grupo5	1.00e-06	1.000000	7.93e-06	0.000229	-
grupo6	0.005100	3.50e-06	0.087394	1.000000	0.008866

P value adjustment method: bonferroni

¿Cómo ordenar las medias?



Anova significativo: comparación 2 a 2

