Análisis de la varianza

José Gabriel Palomo Sánchez
gabriel.palomo@upm.es

E.U.A.T.
U.P.M.

Julio de 2011

Índice I

1. Introducción
   1. Comparación de medias
   2. El pricipio de aleatorización
2. El problema de un factor con dos niveles
   1. Obtención de los datos
   2. Hipótesis para la resolución del problema
   3. Consecuencias de las hipótesis
   4. Estimación de los parámetros del modelo
   5. Comparación de dos niveles. Contraste de hipótesis
3. El Análisis de la varianza con un factor
   1. Introducción
   2. Obtención de los datos
   3. Análisis de la varianza con un factor. Hipótesis del modelo
   4. Análisis de la varianza con un factor. Consecuencias de las hipótesis

Índice II

3. El Análisis de la varianza con un factor. (Continuación)
   5. Análisis de la varianza con un factor. Estimación de los parámetros del modelo
   6. Análisis de la varianza con un factor. Planteamiento del contraste de hipótesis
   7. Análisis de la varianza con un factor. Metodología del Análisis de la varianza para la resolución del contraste
   8. Análisis de la varianza con un factor. El test de la F
   9. Análisis de la varianza con un factor. La tabla ADEVA
   10. Análisis de la varianza con un factor. El coeficiente de determinación
   11. Análisis de la varianza con un factor. Diagnosis y validación del modelo
   12. Análisis de la varianza con un factor. Inferencia sobre los parámetros del modelo
   13. Acepciones del Análisis de la varianza. Observaciones

Comparación de medias I

Problema

En ocasiones el investigador desea analizar si el comportamiento medio de una variable **respuesta** depende del valor, **nivel**, al que se encuentre otra, denominada **factor**.

Comparación de medias II. Ejemplo

Un químico desea comprobar si los distintos valores de la presión (1, 2 ó 3 atmósferas) influyen en el rendimiento medio de una reacción química.

- La variable respuesta, en este caso, es el rendimiento de la reacción química, y el factor la presión.

¿Cómo se hace esta comparación?

Comparación de medias III. El principio de aleatorización

- Una vez determinada la variable respuesta, el factor y sus distintos niveles, se procede a la obtención de datos mediante la experimentación.
- Para ello se asignan los distintos niveles del factor a individuos experimentales, elegidos aleatoriamente.
- La elección aleatoria de los individuos a los que se les asignan los niveles del factor tiene por objeto evitar que la influencia de otras variables, no contempladas en el experimento, invaliden las conclusiones del diseño.

Comparación de medias IV. El principio de aleatorización. Ejemplo

Supóngase que se desea analizar si existen diferencias entre la efectividad de dos tratamientos antitabaco: A y B.

- Para comparar los tratamientos se eligen 10 hombres fumadores y se les aplica el tratamiento A.
- También se eligen 10 mujeres y se les aplica el tratamiento B.
- Si se descubren diferencias entre ambos grupos, ¿a qué serán debidas?, ¿a la mayor efectividad de uno de los tratamientos?, ¿o a una disposición diferente entre los dos sexos a la mejoría?

Comparación de medias V. El principio de aleatorización. Ejemplo

La aleatorización supone la elección al azar de las personas a las que se somete a cada uno de los tratamientos. De esta forma habrá personas de los dos sexos en los dos grupos, y no será atribuible al sexo la posible mayor efectividad de un tratamiento sobre el otro.

Comparación de dos niveles I

El caso más sencillo de comparación de medias es aquél en el que existe un único factor con dos niveles, 1 y 2.

- Para analizar si existen diferencias en la respuesta atribuibles a los dos niveles del factor, se eligen aleatoriamente n individuos.

Comparación de dos niveles II

- A n_1 de ellos se les somete al nivel 1 y a los $n_2 = n - n_1$ restantes se les aplica el nivel 2, midiéndose a continuación el valor de la variable respuesta en cada individuo.
  La tabla siguiente resume los datos de la experimentación.

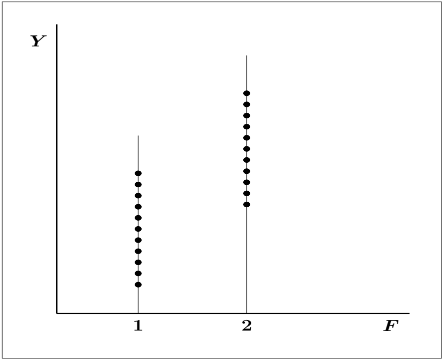
Nivel 1	Nivel 2
y_{11}	y_{21}
y_{12}	y_{22}
$\vdots$	$\vdots$
y_{1n_1}	y_{2n_2}

- **Observación:** No es necesario que $n_1 = n_2$, aunque resulta conveniente que $n_1 \cong n_2$.

Comparación de dos niveles III

Cabe esperar que exista variabilidad entre los individuos tratados con el mismo nivel del factor.

- Por ello, la comparación de los comportamientos de la variable respuesta en los dos niveles del factor se realiza a través de las medias.



Comparación de dos niveles IV. Hipótesis para la resolución del problema

Para el análisis del problema se supondrá que se verifican las siguientes hipótesis:

- La variable respuesta en un individuo se puede descomponer según el modelo:

$$y_{ij} = \mu_i + e_{ij}, \quad \text{con } i = 1, 2 \quad \text{y} \quad j = 1, \ldots, n_i.$$

donde:

  - μ_i es la parte determinista del modelo, y representa el valor medio de la variable respuesta cuando el factor se encuentra en el nivel i.
  - e_{ij} representa el error experimental, y representa la parte aleatoria del modelo.

Comparación de dos niveles V. Hipótesis para la resolución del problema

Además se supondrá que:

- Para todos los valores de i y j,

$$e_{ij} \approx N(0, \sigma).$$

  (El hecho de que el valor de la varianza del error experimental no dependa de los valores de i y de j se conoce con el nombre de **homocedasticidad**.)

- Todos los e_{ij} son independientes entre sí.

Comparación de dos niveles VI. Consecuencias de las hipótesis

Como consecuencia de las hipótesis, se cumple que:

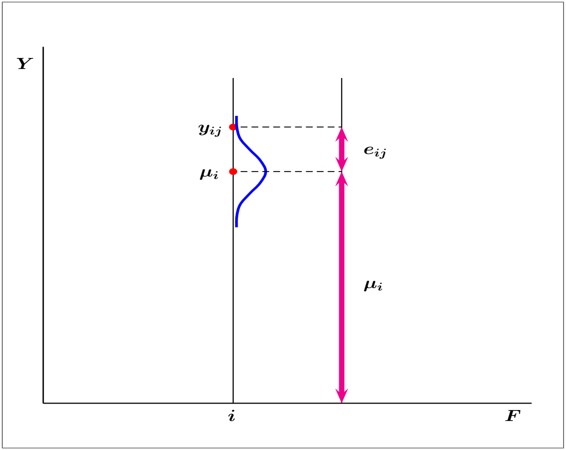
- La variable respuesta en los individuos tratados con el nivel i del factor, F, con $i = 1, 2$, sigue una distribución:

$$(Y|F = i) \approx N(\mu_i, \sigma).$$

- Todos los y_{ij} son independientes entre sí.

Comparación de dos niveles VII. Consecuencias de las hipótesis

Gráficamente,



Comparación de dos niveles VIII. Estimación de los parámetros del modelo

- Para estimar μ_1 y μ_2, se utilizarán las medias muestrales de las observaciones obtenidas en los distintos niveles de los factores:

$$\mu_1 \cong \bar{y}_{1\bullet} \quad \text{y} \quad \mu_2 \cong \bar{y}_{2\bullet}$$

con:

$$\bar{y}_{i\bullet} = \frac{\sum_{j=1}^{n_j} y_{ij}}{n_j}.$$

Comparación de dos niveles IX. Estimación de los parámetros del modelo

- Para estimar σ^2 se utilizará $\hat{s}_R^2$, con

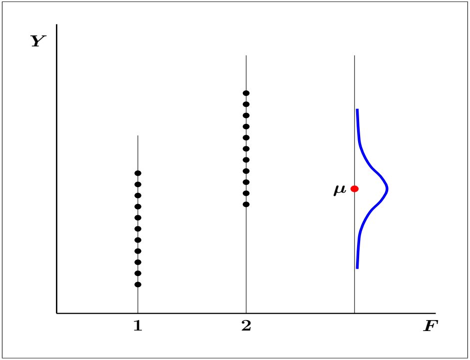
$$\hat{S}_R^2 = \frac{(n_1 - 1)\hat{S}_1^2 + (n_2 - 1)\hat{S}_2^2}{n_1 + n_2 - 2}$$

Observación: $\hat{S}_R^2$ es una aproximación de σ^2 obtenida con los n datos, mediante una ponderación de las varianzas muestrales corregidas, $\hat{S}_1^2$ y $\hat{S}_2^2$ de los dos grupos, que representan, ambas, aproximaciones de σ^2.

Comparación de dos niveles X. Contraste de hipótesis

Se trata de contrastar la hipótesis, H_0, de que tanto las observaciones obtenidas con el nivel 1 del factor, como las obtenidas con el tratamiento 2, provienen de la misma población $N(\mu, \sigma)$.

Gráficamente:

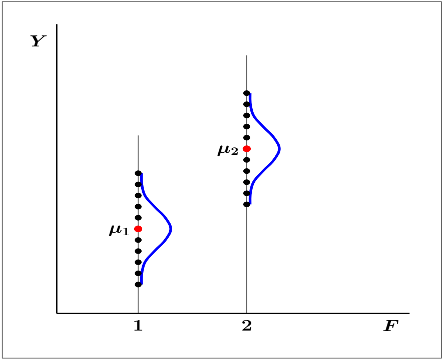


Comparación de dos niveles XI. Contraste de hipótesis

Frente a la hipótesis, H_1, de que las observaciones del nivel 1 son una muestra de una $N(\mu_1, \sigma)$, mientras que las del nivel 2 provienen de otra población $N(\mu_2, \sigma)$. Con

$$\mu_1 \neq \mu_2.$$

Gráficamente:



Comparación de dos niveles XII. Contraste de hipótesis

La realización del contraste:

$$H_0 : \mu_1 = \mu_2,$$

frente a

$$H_1 : \mu_1 \neq \mu_2,$$

se realiza a través del estudio de la diferencia entre las medias de las observaciones de cada uno de los grupos:

$$\bar{y}_{1\bullet} - \bar{y}_{2\bullet}$$

Comparación de dos niveles XIII. Contraste de hipótesis

- De manera efectiva el contraste se lleva a cabo teniendo en cuenta que, si H_0 es cierta y $\mu_1 = \mu_2$, se cumple que

$$\frac{\bar{y}_{1\bullet} - \bar{y}_{2\bullet}}{\hat{S}_R \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \longrightarrow t_{n_1+n_1-2},$$

Comparación de dos niveles XIV. Contraste de hipótesis

- En el caso en que

$$\left|\frac{\bar{y}_{1\bullet}-\bar{y}_{2\bullet}}{\hat{S}_R\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\right| > t_{n_1+n_1-2;\alpha/2},$$

se rechaza la hipótesis nula acerca de la igualdad de medias, aceptándose en caso contrario.

Comparación de dos niveles XV. Contraste de hipótesis

- La hipótesis de igualdad de varianzas se comprueba teniendo en cuenta que, si es cierta la hipótesis nula:

$$H_0 : \sigma_1^2 = \sigma_2^2,$$

se verifica que

$$\frac{\hat{s}_1^2}{\hat{s}_1^2} \longrightarrow F_{n_1-1, n_2-1}.$$

Análisis de la varianza con un factor I

- Supóngase que en el ejemplo anterior referido al rendimiento de una reacción química, se desea comparar K niveles de la presión.
- **Observación:** En este caso se analiza un factor con K niveles.
- Una solución sería comparar los K niveles dos a dos.
- Sin embargo, la realización de los $\binom{K}{2}$ contrastes aumenta considerablemente la probabilidad de encontrar diferencias entre los niveles, aunque no existan.
- Por su mayor eficiencia se emplea el método del **Análisis de la varianza**

ANÁLISIS DE LA VARIANZA CON UN FACTOR II. OBTENCIÓN DE LOS DATOS

1. Para analizar si existen diferencias entre los K niveles, se eligen aleatoriamente n individuos.

2. Se asigna aleatoriamente el nivel 1 a n_1 individuos, el nivel 2 a n_2 individuos,$\ldots$ y el nivel k a n_k individuos.

   En general, se asigna el nivel $i-$ésimo a n_i individuos. Por tanto,

$$\sum_{i=1}^{K} n_i = n.$$

Análisis de la varianza con un factor III. Obtención de los datos

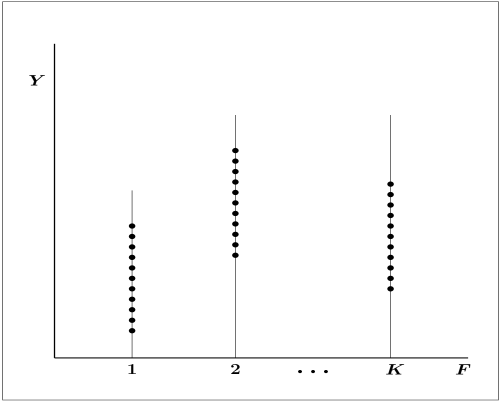
- Se obtendrá una tabla de datos del tipo:

Nivel 1	Nivel 2	$\cdots$	Nivel K
y_{11}	y_{21}	$\cdots$	y_{K1}
y_{12}	y_{22}	$\cdots$	y_{K2}
$\vdots$	$\vdots$		$\vdots$
y_{1n_1}	y_{2n_2}	$\cdots$	y_{Kn_K}

- **Observación:** No es necesario que $n_1 = n_2 \cdots = n_K$, aunque es conveniente que $n_1 \cong n_2 \cdots \cong n_K$.

Análisis de la varianza con un factor IV. Obtención de los datos

Cabe esperar que exista variabilidad entre los individuos obtenidos al mismo nivel del factor. Gráficamente,



Análisis de la varianza con un factor V. Hipótesis del modelo

Como en el caso de la comparación de dos niveles se supondrá que se verifican las siguientes hipótesis:

- La variable respuesta en un individuo se puede descomponer según el modelo del **Análisis de la Varianza con un factor**:

$$y_{ij} = \mu_i + e_{ij}, \quad \text{con } i = 1, \ldots, K \quad \text{y} \quad j = 1, \ldots, n_i.$$

donde:

  - μ_i es la parte determinista del modelo, y representa el valor medio de la variable respuesta cuando el factor se encuentra en el nivel i.
  - e_{ij} representa el error experimental, y representa la parte aleatoria del modelo.

Análisis de la varianza con un factor VI. Hipótesis del modelo

Además se supondrá que:

- Para todos los valores de i y j,

$$e_{ij} \approx N(0, \sigma).$$

  (El hecho de que el valor de la varianza del error experimental no dependa de los valores de i y de j se conoce con el nombre de **homocedasticidad**.)

- Todos los e_{ij} son independientes entre sí.

Análisis de la varianza con un factor VII. Consecuencias de las hipótesis

Como consecuencia de las hipótesis, se cumple que:

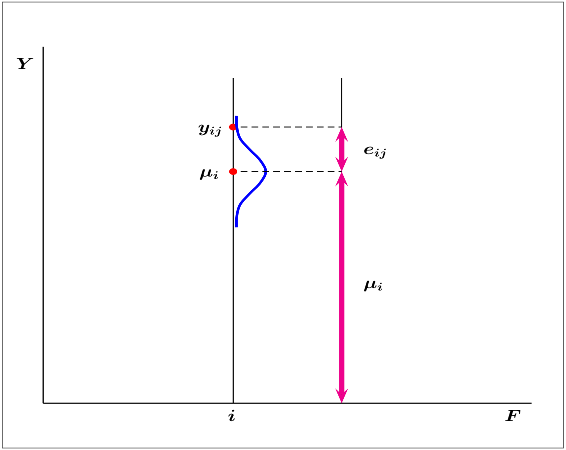
- La variable respuesta en los individuos sometidos al nivel i del factor, F, con $i = 1, 2, \ldots, K$ sigue una distribución:

$$(Y|F = i) \approx N(\mu_i, \sigma).$$

- Todos los y_{ij} son independientes entre sí.

Análisis de la varianza con un factor VIII. Consecuencias de las hipótesis

Gráficamente,



Análisis de la varianza con un factor IX. Estimación de los parámetros del modelo

- Para estimar $\mu_1, \mu_2, \ldots, \mu_K$ se utilizarán las medias muestrales de las observaciones obtenidas en los distintos niveles de los factores:

$$\mu_1 \cong \bar{y}_{1\bullet}, \mu_2 \cong \bar{y}_{2\bullet}, \ldots, \mu_K \cong \bar{y}_{K\bullet}$$

con:

$$\bar{y}_{i\bullet} = \frac{\sum_{j=1}^{n_j} y_{ij}}{n_j}.$$

Análisis de la varianza con un factor X. Estimación de los parámetros del modelo

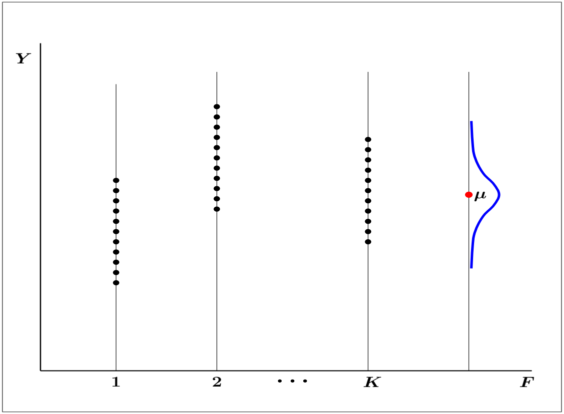
- Para estimar σ^2 se utilizará $\hat{s}_R^2$, con

$$\hat{S}_R^2 = \frac{\sum_{i=1}^{K} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\bullet})^2}{n-K}$$

Observación: $\hat{S}_R^2$ se denomina **varianza residual**, y es una aproximación de σ^2 obtenida con los n datos. Posteriormente se encontrará una interpretación de este estimador de σ^2.

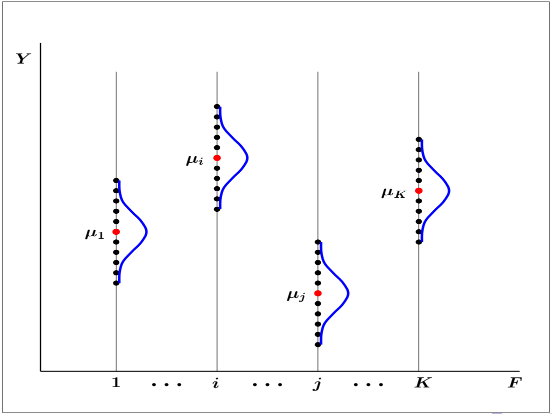
Análisis de la varianza con un factor XI. Contraste de hipótesis

Se trata de contrastar la hipótesis H_0, de que todas las observaciones obtenidas provienen de la misma población $N(\mu, \sigma)$. Gráficamente:



Análisis de la varianza con un factor XII. Contraste de hipótesis

Frente a la hipótesis H_1 de que existen niveles, por lo menos dos, cuyas observaciones provienen de poblaciones normales con la misma varianza, pero con medias distintas. Gráficamente,



Análisis de la varianza con un factor XIII. Metodología del A.V.

La discusión efectiva del contraste:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_K = \mu,$$

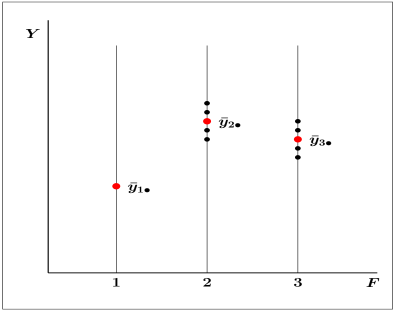
frente a

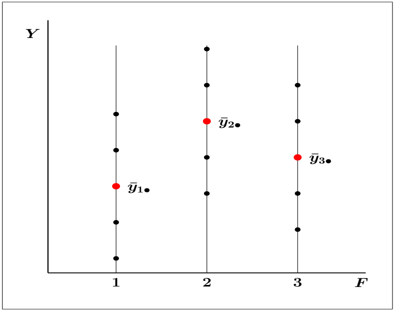
$$H_1 : \text{Existen al menos dos medias, } \mu_i \text{ y } \mu_j\text{, tales que } \ \mu_i \neq \mu_j,$$

se realiza por medio del **Análisis de la varianza.**

Análisis de la varianza con un factor XIV. Metodología del A.V.

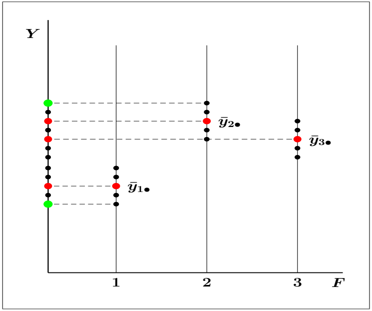
Supóngase que se desea analizar la igualdad de las medias de los distintos niveles en los dos casos siguientes:

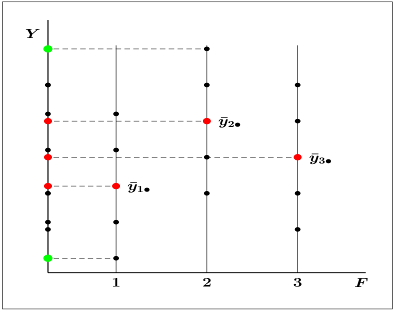




Análisis de la varianza con un factor XV. Metodología del A.V.

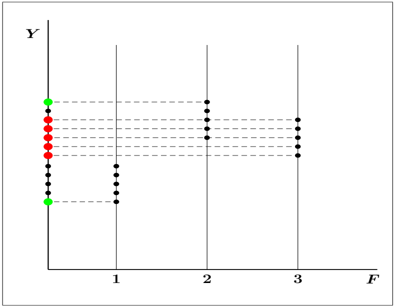
En el primer caso la variabilidad entre las medias de los distintos niveles es más grande, en relación con la variabilidad total, que en el segundo caso. Gráficamente:

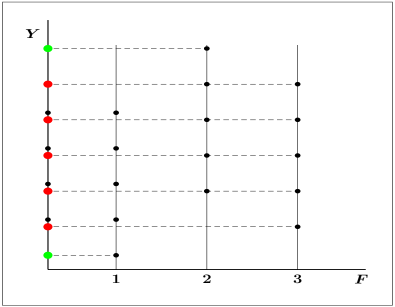




Análisis de la varianza con un factor XVI. Metodología del A.V.

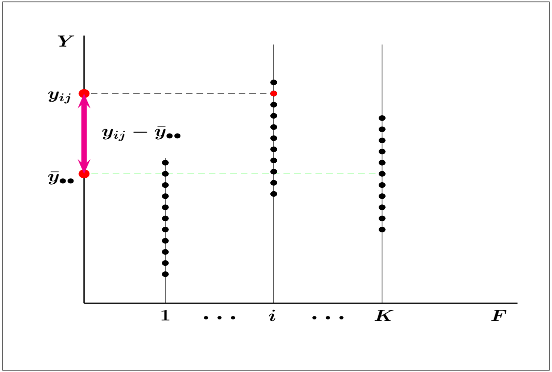
Sin embargo, en el primer caso la variabilidad entre las observaciones correspondientes a un mismo nivel es más pequeña, en relación con la variabilidad total, que en el segundo caso. Gráficamente:





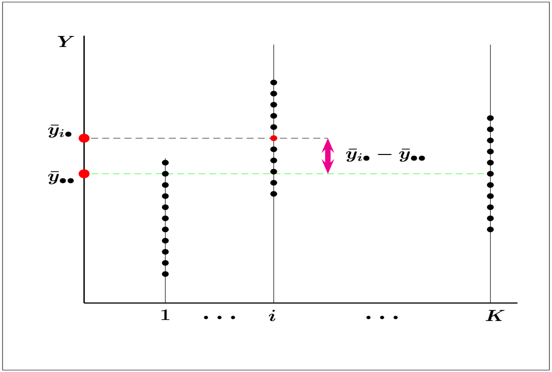
Análisis de la varianza con un factor XVII. Metodología del A.V.

La variabilidad total de todas las observaciones puede medirse a través de las diferencias entre las observaciones, y_{ij}, y la media global, $\bar{y}_{\bullet\bullet}$: $(y_{ij} - \bar{y}_{\bullet\bullet})$. Gráficamente,



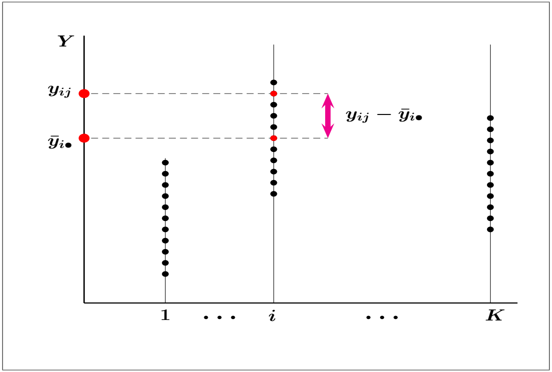
Análisis de la varianza con un factor XVIII. Metodología del A.V.

La variabilidad entre las medias puede medirse a través de las diferencias $\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}$



Análisis de la varianza con un factor XIX. Metodología del A.V.

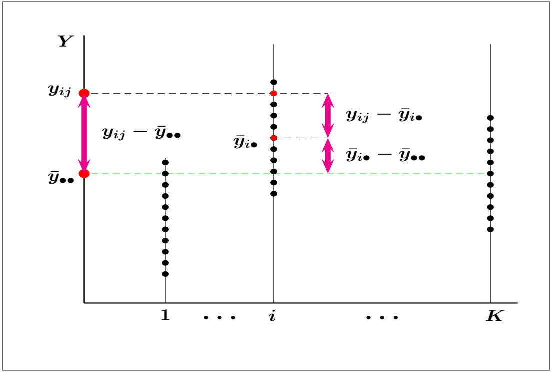
Y la variabilidad dentro de cada nivel puede medirse por las diferencias $y_{ij} - \bar{y}_{i\bullet}$



Análisis de la varianza con un factor XX. Metodología del A.V.

Ahora bien,

$$(y_{ij} - \bar{y}_{\bullet\bullet}) = (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}) + (y_{ij} - \bar{y}_{i\bullet})$$



Análisis de la varianza con un factor XXI. Metodología del A.V.

Como en el caso del cálculo de la varianza de una variable, el análisis de la variabilidad total se realizaría por medio de la suma:

$$VT = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{\bullet\bullet})^2,$$

pero

$$VT = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2 + \sum_{i=1}^{K} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\bullet})^2 + 2 \sum_{i=1}^{K} \sum_{j=1}^{n_i} (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})(y_{ij} - \bar{y}_{i\bullet}).$$

Análisis de la varianza con un factor XXII. Metodología del A.V.

Ahora bien, se puede demostrar que

$$2\sum_{i=1}^{K}\sum_{j=1}^{n_i}(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})(y_{ij}-\bar{y}_{i\bullet})=0.$$

Y resulta que:

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i}(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2=\sum_{i=1}^{K}n_i(\bar{y}_{i\bullet}-\bar{y}_{\bullet\bullet})^2.$$

Análisis de la varianza con un factor XXIII. Metodología del A.V.

Por lo que:

$$VT = \sum_{i=1}^{K} n_i(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2 + \sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\bullet})^2$$

- El sumando $\sum_{i=1}^{K} n_i(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2$ representa la variabilidad entre las medias de los distintos niveles, y se denomina **variabilidad explicada**.

- El sumando $\sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\bullet})^2$ representa la variabilidad interna dentro de los distintos niveles del factor, y se denomina **variabilidad no explicada**.

Análisis de la varianza con un factor XXIV. Metodología del A.V.

En definitiva, se verifica que:

$$VT = VE + VNE$$

- En la medida en que VE sea grande en relación con VNE, habrá evidencia de diferencia de valor entre las medias de los niveles.
- En la medida en que VE sea pequeña en relación con VNE, habrá evidencia de igualdad entre los valores de las medias de los niveles.

Análisis de la varianza con un factor XXV. Metodología del A.V.

- Para discutir la magnitud de la relación entre *VE* y *VNE* es necesario analizar sus distribuciones de probabilidad.

Teorema

1. Si se verifica la hipótesis:

$$\mu_1 = \mu_2 = \cdots = \mu_K = \mu,$$

   la variable VE/σ^2 se distribuye como una χ^2_{K-1}.

2. La variable VNE/σ^2 se distribuye, en cualquier caso, como una χ^2_{n-K} y es independiente con la anterior.

Análisis de la varianza con un factor XXVI. Metodología del A.V.

Consecuencia

Si se verifica la hipótesis:

$$\mu_1 = \mu_2 = \cdots = \mu_K = \mu,$$

la variable

$$\frac{\frac{VE}{\sigma^2(K-1)}}{\frac{VNE}{\sigma^2(n-K)}} \longrightarrow F_{(K-1;n-K)}$$

Análisis de la varianza con un factor XXVII. Metodología del A.V.

- Llamando $\hat{s}_e^2$ al valor de $VE/(\text{K-1})$.
- Y $\hat{s}_R^2$ al valor de $VNE/(\text{n-K})$.

Se tiene que, cuando se cumpla que $\mu_1 = \mu_2 = \cdots = \mu_K = \mu$,

$$\frac{\hat{s}_e^2}{\hat{s}_R^2} \longrightarrow F_{(K-1;n-K)}$$

Análisis de la varianza con un factor XXVIII. Metodología del A.V.

Observaciones:

1. $\hat{s}_e^2$ es una medida de la variabilidad de las medias de las observaciones de los distintos niveles, ponderada por el número de observaciones en cada uno de ellos.
2. $\hat{s}_R^2$ es la varianza residual. Es una estimación centrada de σ^2, que mide la variabilidad dentro de los distintos niveles.
3. $(y_{ij} - \bar{y}_{i\bullet})$ es el residuo de la observación $j-$ésima, diferencia entre la observación y el valor previsto por el modelo para una observación con el tratamiento $i-$ésimo, $\bar{y}_{i\bullet}$.

Análisis de la varianza con un factor XXIX. El test de la F

- Empleando todo lo anterior, para discutir el contraste:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_K = \mu,$$

frente a

$$H_1 : \text{Existen al menos dos medias, } \mu_i \text{ y } \mu_j, \text{ tales que } \; \mu_i \neq \mu_j,$$

basta con analizar el valor del estadístico

$$F = \frac{\hat{s}_e^2}{\hat{s}_R^2}$$

Análisis de la varianza con un factor XXX. El test de la F

- De manera que, si se denomina F_α al valor tal que

  $$P(F_{(K-1;n-K)} > F_\alpha) = \alpha,$$

  cuando

  $$F = \frac{\hat{s}_e^2}{\hat{s}_R^2} < F_\alpha$$

  se aceptará la hipótesis nula, que se rechazará en caso contrario.

Observación: Nótese que el test de la F es un contraste unilateral, en coherencia con la hipótesis que se contrasta.

Análisis de la varianza con un factor XXXI. La tabla ADEVA

Los resultados del test de la F se resumen en la

Tabla ADEVA

Fuentes de variación	**Suma de cuadrados**	**Grados de libertad**	**Varianzas**	F	p-v.
Entre grupos	$\sum n_i(\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2$	$K-1$	$\hat{s}_e^2 = \frac{VE}{K-1}$	$\frac{\hat{s}_e^2}{\hat{s}_R^2}$	p
Residual	$\sum\sum(y_{ij} - \bar{y}_{i\bullet})^2$	$n-K$	$\hat{s}_R^2 = \frac{VNE}{n-K}$		
Total	$\sum\sum(y_{ij} - \bar{y}_{\bullet\bullet})^2$	$n-1$	$\hat{s}_Y^2$		

Análisis de la varianza con un factor XXXII. El coeficiente de determinación

Una medida relativa de la variabilidad total explicada por los niveles del factor es el **coeficente de determinación,** R^2:

$$R^2 = \frac{VE}{VT}$$

Observaciones: En todo caso $0 \leq R^2 \leq 1$.
Además, multiplicado por cien, representa el porcentaje de la variabilidad total de la variable respuesta explicada por el modelo.

Análisis de la varianza con un factor XXXIII. Diagnosis y validación del modelo

- Una vez realizados los cálculos y contrastes anteriores es necesario verificar las hipótesis del modelo.
- Esta verificación se lleva a cabo por medio del análisis de los residuos.
  - La discusión de la normalidad se realiza a través del papel probabilístico normal.
  - La comprobación de la homocedasticidad y de la independencia requiere un gráfico de los residuos frente a los distintos niveles del factor, que debe ser un gráfico sin estructura.

Análisis de la varianza con un factor XXXIV. Diagnosis y validación del modelo

Observación

- En el caso en que el análisis de los residuos no permita validar el modelo, será necesario estudiar transformaciones de los datos que ofrezcan un comportamiento razonable del error experimental.

Análisis de la varianza con un factor XXXV. Inferencia sobre los parámetros del modelo

Realizada la diagnosis del modelo, puede ser necesario hacer inferencia respecto de los parámetros del mismo.

- La inferencia respecto del valor de la media μ_i se puede hacer teniendo en cuenta que:

$$\frac{\bar{y}_{i\bullet} - \mu_i}{\hat{s}_R/\sqrt{n_i}} \longrightarrow t_{n-K}$$

- La comparación de dos medias, μ_i y μ_j se puede estudiar si se tiene en cuenta que:

$$\frac{(\bar{y}_{i\bullet} - \bar{y}_{j\bullet}) - (\mu_i - \mu_j)}{\hat{s}_R\sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \longrightarrow t_{n-K}$$

Análisis de la varianza con un factor XXXVI. Inferencia sobre los parámetros del modelo

- La inferencia respecto de σ^2 se realiza teniendo en cuenta que:

$$\frac{(n-K)\hat{s}_R^2}{\sigma^2} \longrightarrow \chi^2_{n-K}$$

- **Observación**: Existen contrastes para estudiar la igualdad de varianzas de los distintos niveles. Por ejemplo el test de Bartlett o el de Cochran, que suelen estar implementados en los distintos programas informáticos. Su interpretación, a través del p-valor se realiza de la forma habitual.

Acepciones del Análisis de la varianza

Observación

La expresión **Análisis de la varianza** contempla dos acepciones distintas, como se ha descrito en estas notas:

- El problema de comparación de K medias, analizado en este capítulo, se conoce como el «Análisis de la varianza con un factor.»
- Por otra parte, la metodología desarrollada para la discusión del problema anterior, por medio de la comparación de la variabilidad explicada con la no explicada, se conoce por el nombre de la metodología del «Análisis de la varianza.» Esta metodología es muy general, y se emplea en todos los modelos del diseño experimental, así como en los de regresión.