Análisis de Datos

2ºde Biología

ANOVA unifactorial

Universidad Autónoma de Madrid

Departamento de Matemáticas

2012

Tema 1. Análisis de la varianza unifactorial

El problema

Se quiere analizar si una magnitud determinada tiene la misma distribución en varias poblaciones distintas, diferenciadas en el análisis por un solo factor.

Elementos

- Variable respuesta es la variable a explicar Y_i.
- Factor o variable explicativa.
- Niveles (poblaciones, cualidades, grupos, tratamientos, ...) de la variable explicativa (factor). Número de niveles: I.
- Modelo:

  $$Y_i \quad = \mu_i + U \quad = \mu + \alpha_i + U; \quad (i = 1, 2, \ldots, \mathrm{I})$$

Antecedentes

Estadística (primer curso): contrastes «t»

- Comparan las medias de dos grupos
- Ahora compararemos las medias de n grupos

Ejemplo: Ejercicio de la asignatura «Estadística»

6. Se tienen dos métodos, A y B, para determinar el calor latente de fusión del hielo. La siguiente tabla da los resultados obtenidos (en calorías por gramo de masa para pasar de -0.72°C a 0°C) utilizando ambos métodos independientemente:

Método A	79.98 80.04 80.02 80.04 80.03 80.03 80.04 79.97 80.05 80.03 80.02 80.00 80.02
Método B	80.02 79.94 79.98 79.97 79.97 80.03 79.95 79.97

Con un nivel de significación del 10%, ¿existen diferencias significativas entre los resultados medios proporcionados por los dos métodos?

La variable respuesta es «Calor latente de fusión del hielo», el factor «Método de determinación», el número de niveles es $\mathrm{I} = 2$

Primer ejemplo

Proporción de grasa en la leche de varias razas vacunas

REFERENCIA: Sokal, R. R. and Rohlf, F. J. (1981). *Biometry*, 2nd edition, San Fransisco: WH Freeman.

DATOS

Ayrshire	Canadian	Guernsey	Holstein-Fresian	Jersey
3,74	3,92	4,54	3,30	4,80
3,77	4,07	4,59	3,40	5,18
4,08	4,29	4,64	3,55	5,18
4,10	4,38	4,72	3,58	5,18
4,11	4,40	4,83	3,58	5,24
4,25	4,43	4,97	3,59	5,25
4,27	4,46	5,28	3,71	5,41
4,37	4,47	5,30	3,79	5,75
4,41	4,62	5,39	3,83	5,98
4,44	4,85	5,75	4,43	6,55

(...)

En este ejemplo,

- la variable respuesta es «Proporción de grasa en leche»,
- el factor, «Raza de la vaca»,
- el número de niveles, $\mathrm{I} = 5$

Estadísticos descriptivos

Descriptivos

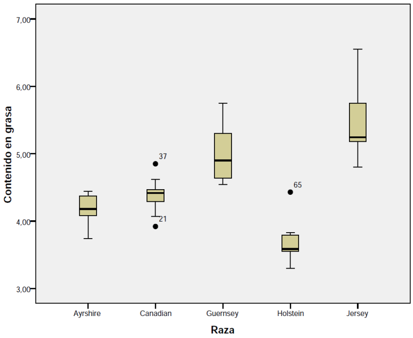
Contenido en grasa

	N	Media	Desviación típica	Error típico
Ayrshire	10	4,1540	,24627	,07788
Canadian	10	4,3890	,26053	,08239
Guernsey	10	5,0010	,40831	,12912
Holstein_Fresian	10	3,6760	,31013	,09807
Jersey	10	5,4520	,50638	,16013
Total	50	4,5344	,72075	,10193

- En la tabla vemos las estimaciones de las medias
- Contrastaremos si estiman o no la misma cantidad
- ¿Podemos afirmar que no todas estiman la misma cantidad?

Comparación visual

- La decisión debe tomarse atendiendo a la variabilidad estimada de los datos
- El diagrama de cajas siguiente nos permite observar la distancia entre las medianas (no las medias[1]) de los grupos en función de las dispersiones



[1] los diagramas de cajas marcan el centro con las medianas no con las medias

Descripción del modelo

$$Y_i = \mu_i + U_i = \mu + \alpha_i + U_i; \quad i = 1, 2, \ldots, \mathrm{I}$$

- Y_i respuesta de la variable en el i-ésimo nivel del factor explicativo
- $\mu_i = E(Y_i)$: valor medio de $Y_i = \mu + \alpha_i; \quad \sum \alpha_i = 0$
- μ_i se descompone como $\mu_i = \mu + \alpha_i; \quad (i = 1, 2, \ldots, \mathrm{I})$ donde α_i representa el efecto del nivel i sobre la media global μ
- U_i es la variación aleatoria de Y_i
- Supondremos que $U_i \sim N(0, \sigma)$, por tanto $Y_i \sim N(\mu_i, \sigma)$
- $\sigma^2 = \text{Var}\, U_i = \text{Var}\, Y_i \quad$ igual para todo i.

Elección de los niveles del factor

Dos formas de elegir los I niveles del factor:

Niveles fijos

Los tratamientos son seleccionados por el experimentador.

EJEMPLO: Efecto sobre la presión arterial de distintos medicamentos: los medicamentos son elegidos por el experimentador.

Niveles aleatorios

Los tratamientos se seleccionan al azar entre todos los posibles.

EJEMPLO: Efecto de un contaminante en las aguas de un lago. Se quiere estudiar si la contaminación es o no uniforme en todo el lago: se seleccionan al azar las I estaciones de muestreo.

NOTA: En las propiedades estadísticas del Análisis de la Varianza unifactorial no hay diferencia entre la selección fija o aleatoria de los niveles.

Muestra aleatoria y datos

Muestra aleatoria

- Y_{ij} es la j-ésima observación dentro del i-ésimo nivel del factor: $i = 1, 2, \ldots, \mathrm{I}$, $j = 1, 2, \ldots, n_i$.
- n_i es el tamaño de la muestra en el nivel i. Si todas las muestras tienen el mismo tamaño el diseño es equilibrado.
- Número total de datos: $n = n_1 + ... + n_{\mathrm{I}}$

Las observaciones se realizarán al azar e independientemente unas de otras.

Datos

- y_{ij} resultado de la j-ésima observación dentro del i-ésimo nivel del factor explicativo

Muestra

Factor	1	2	...	I
	Y_{11}	Y_{21}	...	$Y_{\mathrm{I}1}$
	Y_{12}	Y_{22}	...	$Y_{\mathrm{I}2}$
	$\vdots$	$\vdots$		$\vdots$
	$\vdots$	Y_{2n_2}		$\vdots$
	$\vdots$			$Y_{\mathrm{I}n_{\mathrm{I}}}$
	Y_{1n_1}			
	$\downarrow$	$\downarrow$		$\downarrow$
	$\bar{Y}_{1\cdot}$	$\bar{Y}_{2\cdot}$		$\bar{Y}_{\mathrm{I}\cdot}$
	S_1^2	S_2^2		S_{I}^2

Datos

Factor	1	2	3	Total
	20	15	19	
	18	17	11	
	21	22	18	
	22	24	22	
	19		17	
	25			
n_i	6	4	5	15
$\bar{y}_{i\cdot}$	20,83	19,50	17,40	19,33
s_i^2	6,167	17,67	16,30	12,42

Variables $Y_{ij} \sim N(\mu_i, \sigma)$ independientes; $i=1,\ldots,\mathrm{I}$; $j=1,\ldots,n_i$; $\sum_i n_i = n$.

Estimación de parámetros

Parámetros desconocidos del modelo: $\mu_1, \ldots, \mu_{\mathrm{I}}, \sigma$; en total $\mathrm{I}+1$.

Estimaciones

$$\hat{\mu}_i = \bar{y}_{i\cdot} = \frac{1}{n_i}\sum_j y_{ij}, \quad i = 1, \ldots, \mathrm{I}$$

$$\hat{\sigma}^2 = S_R^2 = \frac{1}{n-\mathrm{I}}\sum_{i=1}^{\mathrm{I}}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_{i\cdot})^2$$

NOTAS:

1. $\sum_j (y_{ij} - \bar{y}_{i\cdot})^2 = (n_i - 1)s_i^2 \quad ; \qquad S_R^2 = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_{\mathrm{I}} - 1)s_{\mathrm{I}}^2}{n - \mathrm{I}}$

2. $\sum_j (y_{ij} - \bar{y}_{i\cdot})^2 = \left(\sum_j y_{ij}^2\right) - n_i \bar{y}_{i\cdot}^2$

Intervalos

Intervalos de confianza

$$\mathrm{IC}_{1-\alpha}(\mu_i) = \left(\bar{y}_{i\cdot} - t_{n-\mathrm{I};\frac{\alpha}{2}} S_R \sqrt{\frac{1}{n_i}}, \quad \bar{y}_{i\cdot} + t_{n-\mathrm{I};\frac{\alpha}{2}} S_R \sqrt{\frac{1}{n_i}} \right)$$

$$\mathrm{IC}_{1-\alpha}(\sigma^2) = \left(\frac{(n-\mathrm{I})S_R^2}{\chi^2_{n-\mathrm{I};\frac{\alpha}{2}}}, \quad \frac{(n-\mathrm{I})S_R^2}{\chi^2_{n-\mathrm{I};1-\frac{\alpha}{2}}} \right)$$

Requisitos de modelo

Requisitos

1. **Normalidad:** En cada nivel del factor la variable es normal.
2. **Homocedasticidad:** En cada nivel del factor la variabilidad es similar
3. **Linealidad:** Los efectos de los factores sobre la variable son aditivos
4. **Independencia:** Las observaciones a realizar son independientes

NOTA: Cualquier desviación importante de estas características del modelo puede conducir a conclusiones erróneas.

Residuos

Según el modelo $Y_{ij} = \mu_i + U_{ij}$, el término U_{ij} es la parte aleatoria de la variable, que se denomina residuo. Independientemente de i, j el residuo U_{ij} tiene distribución $N(0, \sigma)$.

La diagnosis del modelo se realiza por medio del análisis de las estimaciones de los residuos de la muestra:

$$u_{ij} = y_{ij} - \bar{y}_{i\cdot}$$

NOTA: Es una estimación ya que de cada dato y_{ij} se resta el valor estimado de la media $\bar{y}i\cdot$, no la media μ_i.

Los n residuos son datos independientes procedentes de una $N(0, \sigma)$.

Análisis de la varianza

Desarrollado por Ronald Fisher a partir de 1920.

Sir Ronald Aylmer Fisher (1890 - 1962)

Biografía de Fisher en MacTutor:
http://www-history.mcs.st-and.ac.uk/Biographies/Fisher.html

Análisis de la varianza

Análisis de la varianza: ANOVA

Para decidir si las medias de los grupos pueden considerarse no iguales compara la variabilidad en los grupos con la variabilidad entre los grupos.

$$\underset{\text{Desviaciones:}}{(Y_{ij} - \bar{Y}_{\cdot\cdot})} = \underset{\text{intra-grupos}}{(Y_{ij} - \bar{Y}_{i\cdot})} + \underset{\text{entre-grupos}}{(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})}$$

Elevando al cuadrado y sumando resulta:

$$\sum_i \sum_j (y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{i\cdot})^2 + n_i \sum_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

que da una descomposición de la variabilidad total (en términos cuadráticos) como suma de la variabilidad residual más la variabilidad explicada por el modelo.

TEOREMA: Al calcular el cuadrado de la suma desaparecen los dobles productos.

Términos

Términos para estudiar la variabilidad

- **SCE** Suma de cuadrados explicada (variabilidad debida a que hay distintos niveles del factor): $n_i \sum_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$.
- **SCR** Suma de cuadrados residual (variabilidad interna dentro de cada nivel): $\sum_i \sum_j (y_{ij} - \bar{y}_{i\cdot})^2$.
- **SCT** Suma de cuadrados total (variabilidad total de todos los datos): $\sum_i \sum_j (y_{ij} - \bar{y}_{\cdot\cdot})^2$.

$$\text{SCT} = \text{SCE} + \text{SCR}$$

El contraste

Contraste de igualdad de medias

Hipótesis nula: $H_0 \equiv \mu_1 = \mu_2 = \cdots = \mu_I$.
Hipótesis alternativa: $H_1 \equiv \exists i, j : \mu_i \neq \mu_j$.

Se resuelve comparando la SCE y la SCR:

Tabla ANOVA

Fuentes de variación	Sumas	g. l.	Varianza	Test F	p-val
Explicada	SCE	$I-1$	$S_E^2 = \frac{\text{SCE}}{I-1}$	$F = \frac{\text{SCE}\cdot(n-I)}{\text{SCR}\cdot(I-1)}$	p
Residual	SCR	$n-I$	$S_R^2 = \frac{\text{SCR}}{n-I}$		
Total	SCT	$n-1$			

NOTA: ANOVA con $I = 2$ es matemáticamente equivalente al contraste «t» de Student para la igualdad de medias con varianzas iguales.

Coeficiente de determinación R^2

¿Qué proporción de la variabilidad de los datos está explicada por el modelo?
Interpretamos el cociente

$$R^2 = \frac{\mathrm{SCE}}{\mathrm{SCT}}$$

como la fracción de la variabilidad de de los datos explicada por el modelo. El resto a 1, es decir

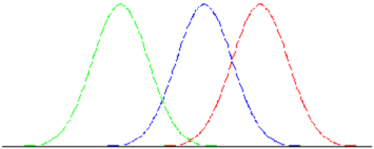
$$\frac{\mathrm{SCR}}{\mathrm{SCT}}$$

será la parte aleatoria de la variabilidad.

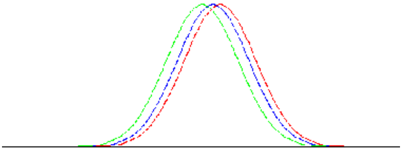
Casuística

La hipótesis alternativa incluye una variedad amplia de resultados. Por ejemplo, para $I = 3$ tendríamos 4 posibilidades:

- Las tres medias son distintas dos a dos.
- Dos medias son iguales y la tercera es distinta (puede ocurrir de tres formas diferentes).



Medias distintas



Medias iguales

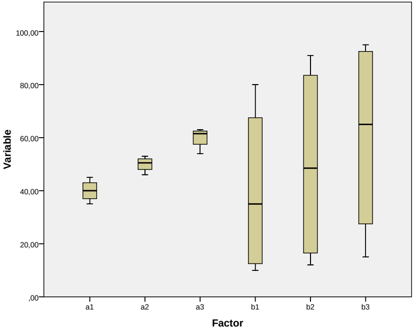
NOTA: Aunque lógicamente estas son las cuatro únicas posibilidades para la hipótesis alternativa cuando $I = 3$, estadísticamente puede ocurrir que rechacemos, digamos $\mu_1 = \mu_3$ pero no rechacemos $\mu_1 = \mu_2$ y $\mu_2 = \mu_3$.

Ejemplo

Consideramos los dos conjuntos de datos A y B siguientes:

(A)				(B)			
	45	53	63		10	12	95
	35	46	62		80	21	15
	41	51	61		15	91	40
	39	50	54		55	76	90
$\bar{Y}_{i\cdot} =$	40	50	60	$\bar{Y}_{i\cdot} =$	40	50	60
$R^2 = 0{,}86$				$R^2 = 0{,}06$			

Diagramas de cajas



Tablas ANOVA

(A)

Origen	*S. C.*	*g. l.*	*Var.*	*F*	*p-val.*	*F-crit.*
Inter grupos	800	2	400	28,13	0,0001	4,26
Intra grupos	128	9	14,22			
Total	928	11				

(B)

Origen	*S. C.*	*g. l.*	*Var.*	*F*	*p-val.*	*F-crit.*
Inter grupos	800	2	400	0,287	0,757	4,26
Intra grupos	12542	9	1394			
Total	13342	11				

Observaciones

El contraste ANOVA equilibrado (con iguales tamaños de las muestras) es bastante fiable (robusto) al rechazar H_0 incluso con desviaciones pequeñas de los requisitos de igualdad de varianzas o normalidad.

Si las varianzas son muy diferentes o se detectan serias desviaciones de la normalidad, se pueden realizar transformaciones de la variable Y que podrían resolver el problema. Por ejemplo tomando el $\log Y$ (si la variabilidad crece con los valores de Y) o alguna potencia de Y.

Otra situación irregular que debe detectarse es la existencia de datos anómalos (outliers). En este caso habría que estudiar más a fondo dichos datos y su posible causa de anomalía.

Análisis posteriores al rechazo de H_0

Al rechazar H_0 tenemos evidencia estadística de que al menos una de las μ_i es diferente de alguna de las otras pero ¿entre cuáles hay diferencia significativa?

Intervalos

Intervalo de confianza para la diferencia de dos medias:

$$IC_{1-\alpha}(\mu_i - \mu_j) = \left[(\bar{y}_{i\cdot} - \bar{y}_{j\cdot}) \pm t_{n-I;\frac{\alpha}{2}} \cdot S_R \sqrt{\frac{1}{n_i} + \frac{1}{n_j}} \quad \right]$$

Contrastes

Contraste de hipótesis sobre la igualad de dos medias al nivel de significación α:

$$H_0 \equiv \mu_i = \mu_j \quad vs. \quad H_1 \equiv \mu_i \neq \mu_j$$

$$\mathcal{R} = \left\{ \left| \frac{\bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}}{S_R \sqrt{\frac{1}{n_i} - \frac{1}{n_j}}} \right| > t_{n-I;\alpha/2} \right\}$$

NOTA: Equivalente a rechazar H_0 cuando el cero está en el intervalo de confianza para la diferencia de medias.

Comparaciones múltiples

Pruebas Post hoc

Si realizamos una comparación (con el mismo nivel α) de todas las posibles parejas de medias la probabilidad de que rechacemos incorrectamente en alguno de los contrastes puede ser muy alta: hasta $1-(1-\alpha)^c$ donde c es el número de contrastes a realizar. Si el factor tiene I niveles, el valor de c será:

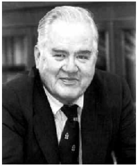
$$\binom{I}{2} = \frac{I \cdot (I-1)}{2}$$

Ejemplo: Si hay cinco niveles del factor, $c = 10$, si hay 10, $c = 45$.

Método de Bonferroni

El contraste múltiple de Bonferroni fija un nivel de significación total α_T y realiza todos los contrastes de parejas con $\alpha = \alpha_T / c$. Es importante señalar que puede ocurrir que rechacemos H_0 en ANOVA y que al usar el método de Bonferroni no encontremos diferencias entre ningún par de medias.

Carlo Emilio Bonferroni (1892 - 1960)



John Wilder Tukey (1915 - 2000)

Ejemplo

Con los datos del último ejemplo, en el que el *p*-valor del contraste ANOVA fue 0,0001, los intervalos de confianza para la diferencia de las medias, con una confianza global del 95 %, serán

$$\mathrm{IC}_{95\,\%}(\mu_1-\mu_2)=\left((40-50)\pm 2{,}96\sqrt{14{,}22(\tfrac{1}{4}+\tfrac{1}{4})}\right)=(-10\pm 7{,}9)$$

$$\mathrm{IC}_{95\,\%}(\mu_1-\mu_3)=\left((40-60)\pm 2{,}96\sqrt{14{,}22(\tfrac{1}{4}+\tfrac{1}{4})}\right)=(-20\pm 7{,}9)$$

$$\mathrm{IC}_{95\,\%}(\mu_2-\mu_3)=\left((50-60)\pm 2{,}96\sqrt{14{,}22(\tfrac{1}{4}+\tfrac{1}{4})}\right)=(-10\pm 7{,}9)$$

Donde $2{,}96 = t_{9,0{,}008}$. En ninguno de los tres casos el intervalo contiene a *cero*, por lo que se puede afirmar, con significación conjunta $\alpha = 0{,}05$, que las tres medias son distintas dos a dos.

Otros contrates

El test de Bonferroni es muy conservador, sobre todo si c es grande.
Por ejemplo, si el factor tiene 5 niveles y fijamos $\alpha_T = 0{,}05$ tendremos que el α para cada contraste entre dos medias es $0{,}005$.

Otros contrates múltiples

Tukey	bueno si el diseño es equilibrado
Scheffé	útil en el caso de tamaños muestrales diferentes; coincide siempre con ANOVA
Dunnett	si hay un *grupo de control*
Duncan	....

Resumen

Modelo: $Y_{ij} \sim N(\mu_i;\sigma^2)$ independientes; $i=1,...,I$; $j=1.,...,n_i$.

$\bar{y}_{i.}=\frac{1}{n_i}\sum_j y_{ij}$; $\sum_i n_i=n$; $\bar{y}_{..}=\frac{1}{n}\sum_i n_i\bar{y}_{i.}$

$\hat{\mu}_i=\bar{y}_{i.}=\frac{1}{n_i}\sum_j y_{ij}$, $i=1,...,I$; $\hat{\sigma}^2=S_R^2=\frac{1}{n-I}\sum_i\sum_j(y_{ij}-\bar{y}_{i.})^2$

$IC_{1-\alpha}(\mu_i)=\left(\bar{y}_{i.}\pm t_{n-I;\alpha/2}\ S_R\sqrt{\frac{1}{n_i}}\right)$; $IC_{1-\alpha}(\sigma^2)=\left(\frac{(n-I)S_R^2}{\chi^2_{n-I;\alpha/2}}\ ;\frac{(n-I)S_R^2}{\chi^2_{n-I;1-\alpha/2}}\right)$

Tabla ANOVA

Suma de cuadrados	g.l.	Varianza	Estadístico
$\text{SCE}=\sum_i n_i(\bar{y}_{i.}-\bar{y}_{..})^2$	$I-1$	$\frac{\text{SCE}}{I-1}$	$F=\frac{\text{SCE}/(I-1)}{\text{SCR}/(n-I)}$
$\text{SCR}=\sum_i\sum_j(y_{ij}-\bar{y}_{i.})^2$	$n-I$	$S_R^2=\frac{\text{SCR}}{n-I}$	
$\text{SCT}=\sum_i\sum_j(y_{ij}-\bar{y}_{..})^2$	$n-1$		

$IC_{1-\alpha}(\mu_i-\mu_j)=\left(\bar{y}_{i.}-\bar{y}_{j.}\pm t_{n-I;\alpha/2}\ S_R\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}\right)$; $S_R^2=\frac{\sum_i(n_i-1)s_i^2}{n-I}$

Ejemplo: Micorriza

Estudio sobre el intercambio de carbono entre árboles de distintas especies a través de la micorriza.
Sobre un determinado número de parejas abedul papirífero (*Betula papyrifera*) — abeto Douglas (*Pseudotsuga menziesii*) se introduce CO_2 marcado con ^{13}C o ^{14}C y se mide la transferencia neta de C entre ellos en tres condiciones distintas para el abeto («Sombra», «Sol y sombra», «Sol») y las mismas para el abedul (pleno sol).
Fuente: Simard et al.; Nature 388 (997) pp. 579—582
(datos están recogidos en
http://www.zoology.ubc.ca/~whitlock/bio300/LectureNotes/)

Modelo

Variable a explicar: transferencia media de C
Factor: Situación del abeto
Niveles del factor: Sombra, Sombra parcial, Sol

Sombra	Sol y sombra	Sol
15,1	4,7	8,9
19,8	12,2	0,1
13,0	15,3	5,0
16,6	8,0	9,5
20,1	7,0	1,4

Descriptivos

Groups	Count	Sum	Average	Variance
Sombra	5	84,6	16,92	9,297
Sol y sombra	5	47,2	9,44	18,113
Sol	5	24,9	4,98	18,107

Tabla ANOVA

F critica al nivel de significación $\alpha = 0{,}05$

Source of Variation	SS	df	MS	F	$p - \text{val}$	Fcrit
Between Groups	364,01	2	182,00	12,00	0,00137	3,89
Within Groups	182,07	12	15,17			
Total	546,08	14				

Al nivel de significación $\alpha = 0{,}05$, se rechaza la hipótesis nula de igualdad de medias.

Conclusión

La media de la variable «transferencia de C» depende de la situación del abeto.

Cabe preguntarse para qué pares puede afirmarse que las medias son diferentes.

Pruebas post-hoc

Error típico en la estimación de la diferencia de las medias:

$$\sqrt{S_R^2(\frac{1}{5}+\frac{1}{5})} = \sqrt{15{,}17\frac{2}{5}} = 3{,}95$$

Valor crítico de t para un nivel de significación conjunto de 0,05:

$t_{12;0{,}008} = 2{,}801$

Diferencias:

	Diferencia	Estadístico t
Sombra — Sol y sombra:	$16{,}92 - 9{,}44 = 7{,}48$	1,89
Sombra — Sol:	$16{,}92 - 4{,}98 = 11{,}94$	3,02
Sol y sombra — Sol:	$9{,}44 - 4{,}98 = 4{,}46$	1,13

Solamente podemos afirmar que las medias de absorción de C son distintas cuando el abeto se encuentra en las dos condiciones extremas: «Sombra» *vs.* «Sol»

Ejemplo

Proporción de grasa en la leche de varias razas vacunas

Sokal, R. R. and Rohlf, F. J. (1981). *Biometry*, 2nd edition, San Fransisco: WH Freeman.

butterfat_short

Ayrshire	Canadian	Guernsey	Holstein-F	Jersey
3.74	3.92	4.54	3.30	4.80
3.77	4.07	4.59	3.40	5.18
4.08	4.29	4.64	3.55	5.18
4.10	4.38	4.72	3.58	5.18
4.11	4.40	4.83	3.58	5.24
4.25	4.43	4.97	3.59	5.25
4.27	4.46	5.28	3.71	5.41
4.37	4.47	5.30	3.79	5.75
4.41	4.62	5.39	3.83	5.98
4.44	4.85	5.75	4.43	6.55
10	10	10	10	10
4.15	4.39	5.00	3.68	5.45
0.25	0.26	0.41	0.31	0.51

Residuos tipificados

Ayrshire	Canadian	Guernsey	Holstein-F	Jersey
−1.68	−1.80	−1.13	−1.21	−1.29
−1.56	−1.22	−1.01	−0.89	−0.54
−0.30	−0.38	−0.88	−0.41	−0.54
−0.22	−0.03	−0.69	−0.31	−0.54
−0.18	0.04	−0.42	−0.31	−0.42
0.39	0.16	−0.08	−0.28	−0.40
0.47	0.27	0.68	0.11	−0.08
0.88	0.31	0.73	0.37	0.59
1.04	0.89	0.95	0.50	1.04
1.16	1.77	1.83	2.43	2.17

ANOVA

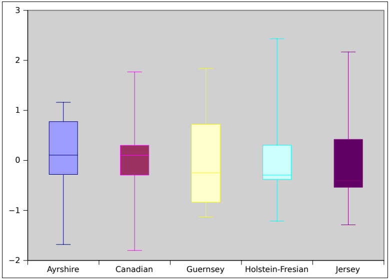
Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Ayrshire	10	41.54	4.154	0.0606
Canadian	10	43.89	4.389	0.0679
Guernsey	10	50.01	5.001	0.1667
Holstein-Fresian	10	36.76	3.676	0.0962
Jersey	10	54.52	5.452	0.2564

ANOVA

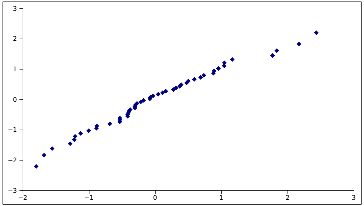
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F critical*
Between Groups	19.62	4	4.906	37.86	0.0000	2.58
Within Groups	5.83	45	0.130			
Total	25.45	49				



histogram

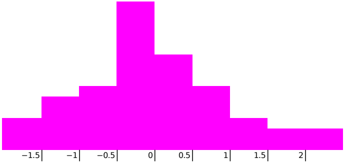
−1.6811	−1.8002	−1.1290	−1.2124	−1.2876
−1.5593	−1.2244	−1.0066	−0.8899	−0.5371
−0.3005	−0.3800	−0.8841	−0.4063	−0.5371
−0.2193	−0.0345	−0.6882	−0.3095	−0.5371
−0.1787	0.04222	−0.4188	−0.3095	−0.4187
0.3898	0.15737	−0.0759	−0.2773	−0.3989
0.4710	0.27252	0.68330	0.10963	−0.0829
0.8771	0.31090	0.73228	0.36758	0.58849
1.0395	0.88665	0.95270	0.49656	1.04270
1.1613	1.76946	1.83437	2.43122	2.16834

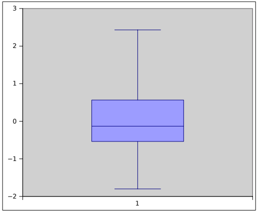
Lilliefors (Kolmogorov-Smirnov) Test Column 1	
Alpha	0.05
p-Value	0.6997
Statistic	0.0744
N	50
Conclusion	Possibly normal



histograma

−1.8002
−1.6811
−1.5593
−1.2876
−1.2244
−1.2124
−1.1290
−1.0066
−0.8899
−0.8841
−0.6882
−0.5371
−0.5371
−0.5371
−0.4188
−0.4187
−0.4063
−0.3989
−0.3800
−0.3095
−0.3095
−0.3005
−0.2773
−0.2193
−0.1787
−0.0829
−0.0759
−0.0345
0.04222
0.10963
0.15737
0.27252
0.31090
0.36758
0.38982
0.47103
0.49656
0.58849
0.68330
0.73228
0.87709
0.88665
0.95270
1.03951
1.04270
1.16133
1.76946
1.83437
2.16834
2.43122





P. Cifuentes

Bonferroni

Bonferroni

Ayrshire	Canadian	Guernsey	Holstein-Fr	Jersey
4.154	4.389	5.001	3.676	5.452

S^2	t(45,0'0025)	error
0.130	2.95	0.1610

Intervalos de confianza conjunta 0'05 (Bonferroni)

			(	)	*sig.*
Ayrshire	Canadian	−0.235	−0.71	0.24	-
	Guernsey	−0.847	−1.32	−0.37	*
	Holstein-Fr	0.478	0.0028	0.95	*
	Jersey	−1.298	−1.77	−0.82	*
Canadian	Guernsey	−0.612	−1.09	−0.14	*
	Holstein-Fr	−0.612	−1.09	−0.14	*
	Jersey	−1.063	−1.54	−0.59	*
Guernsey	Holstein-Fr	1.325	0.85	1.80	*
	Jersey	−0.451	−0.93	0.02	-
Holstein-Fr	Jersey	−1.776	−2.25	−1.30	*