



UNIVERSIDAD NACIONAL AGRARIA LA MOLINA

Diseño y Análisis de Experimentos en Ingeniería y Ciencias Ambientales

Ancova DCA y DBCA

Dr. Christian R. Encina Zelada

cencina@lamolina.edu.pe



Análisis de Covarianza (ANCOVA)

Introducción

- En el análisis de covarianza se combinan los conceptos del **análisis de variancia** para un diseño experimental y para la **regresión**.
- El análisis de covarianza es utilizado en casos en los que la variable respuesta de un diseño experimental **esté relacionada con una o más variables concomitantes**.
- La aplicación del ANCOVA es aplicable a los diseños anteriormente estudiados: DCA, DBCA y otros más.



Finalidad del ANCOVA

Los objetivos más importantes del análisis de covarianza son:

- a. Disminuir el error experimental, con el consiguiente aumento en la precisión del experimento.
- b. Ajustar los promedios de los tratamientos, con la información obtenida de la variable concomitante o independiente.
- c. Hacer una mejor interpretación de los resultados de los experimentos, especialmente en cuanto se relaciona con la naturaleza de los efectos de los tratamientos.



Supuestos del ANCOVA

- Cuando se utiliza el análisis de covarianza es necesario asumir ciertos requisitos que le den validez al análisis. Estas asunciones son:
 - La variable X es fija, medida sin error y no es afectada por los tratamientos.
 - La variable Y deben tener varianzas homogéneas en los tratamientos.
 - La variable Y deben tener distribución normal para cada valor de X.
 - La regresión de X sobre Y, debe ser lineal.
- Los errores se distribuyen independientemente y normal con cero de promedio y con varianza σ^2 .

Modelos aditivos lineales

Los diseños experimentales que van a ser estudiados utilizando ANCOVA son el DCA y DBCA los siguientes:

$$\begin{array}{l} \text{En} \\ \text{DCA:} \end{array} \quad Y_{ij} = \mu_i + \beta(X_{ij} - \bar{X}_{\square\square}) + \varepsilon_{ij} = \mu + \tau_i + \beta(X_{ij} - \bar{X}_{\square\square}) + \varepsilon_{ij}$$

$$\begin{array}{l} \text{En} \\ \text{DBCA:} \end{array} \quad Y_{ij} = \mu_i + \gamma_j + \beta(X_{ij} - \bar{X}_{\cdot\cdot}) + \varepsilon_{ij} = \mu + \tau_i + \gamma_j + \beta(X_{ij} - \bar{X}_{\cdot\cdot}) + \varepsilon_{ij}$$

Modelo aditivo lineal para un DCA

$$Y_{ij} = \mu_i + \beta(X_{ij} - \bar{X}_{\square}) + \varepsilon_{ij} = \mu + \tau_i + \beta(X_{ij} - \bar{X}_{\square}) + \varepsilon_{ij}$$

Donde:

Y_{ij} : es el valor o rendimiento observado en el i-ésimo tratamiento.

μ : es el efecto de la media general.

τ_i : es el efecto del i-ésimo tratamiento.

β : es el coeficiente de regresión lineal del Y sobre X.

X_{ij} : es el valor de la variable independiente en el i-ésimo tratamiento.

\bar{X}_{\square} : es la media de la variable independiente.

ε_{ij} : es el efecto del error experimental en el i-ésimo tratamiento.

t = número de tratamientos.

Procedimiento en la evaluación de los datos

- Evaluación de los datos mediante ANCOVA ($H_0: \beta=0$ vs. $H_1: \beta \neq 0$)
- Prueba hipótesis de la relación de la variable respuesta con la variable concomitante (independiente) (Prueba de la Pendiente)
- Si no se rechaza la hipótesis nula de la prueba de la pendiente, se procede con un ANVA con la variable respuesta y se procede con la prueba hipótesis de tratamientos, sin considerar la variable X.
- Si se rechaza la hipótesis nula de la prueba de la pendiente, se procede con el ANCOVA para obtener las sumas de cuadrados ajustadas ($SC_{aj.}$) y se procede con la prueba hipótesis de tratamientos ajustados.

Ejemplo

Se utilizó un experimento para determinar si tres tipos de dietas producen el mismo peso en gallinas (en cientos de gramos). Por ello se registró el peso inicial (en cientos de gramos) de las gallinas antes del experimento. Los datos obtenidos fueron:

Peso inicial	3	4	7	8	9	8	10	10	11
Peso final	12.0	14.0	16.0	20.2	21.3	18.9	19.0	19.2	20.0
Dieta	A	A	A	B	B	B	C	C	C

- 1) Definir el modelo aditivo lineal o modelo estadístico
- 2) Elaborar ANCOVA y hacer las pruebas respectivas a un nivel de significación del 5%
- 3) Hacer pruebas de comparación

Ejemplo: Modelo aditivo lineal

El modelo aditivo lineal es el siguiente:

$$Y_{ij} = \mu + \tau_i + \beta(X_{ij} - \bar{X}_{\cdot}) + \varepsilon_{ij} \quad i=1, \dots, t$$

Donde:

Y_{ij} : es el peso final de gallinas en cientos de gramos tratadas en el i-ésimo tipo de dieta, de la j-ésima repetición.

μ : es el efecto de la media general de los pesos.

τ_i : es el efecto de la i-ésimo tipo de dieta.

β : es el coeficiente de regresión lineal del Y, el peso final de las gallinas, sobre X, el peso inicial.

X_{ij} : es el peso inicial (en cientos de gramos) de las gallinas tratadas con el i-ésimo tipo de dieta, j-ésima repetición.

\bar{X}_{\cdot} : es el peso medio de las gallinas.

ε_{ij} : es el efecto del error experimental con la i-esima dieta, en la j-ésimo repetición.

$t=3$ (número de tratamientos).

Prueba hipótesis de la relación de Y respect a X (Prueba de la pendiente)

Pruebe si el peso inicial influye sobre el peso final.

Use $\alpha = 0.05$

$H_0: \beta = 0$ (el peso final de las gallinas no depende linealmente del peso inicial)

$H_1: \beta \neq 0$ (el peso final de las gallinas depende linealmente del peso inicial de las gallinas)

$\alpha = 0.05$

Prueba
estadística:

$$F_C = \frac{\left(\frac{E_{XY}^2}{E_{XX}} \right)}{CME_{aj.}} = \frac{\left(\frac{9.77^2}{10} \right)}{0.3789} = 25.19 \sim F_{(1,5)}$$

Criterio de
decisión:

Si $F_C \leq 6.61 = F_{(0.95,1,5)}$ no se rechaza H_0

Si $F_C > 6.61 = F_{(0.95,1,5)}$ se rechaza H_0



Conclusión de la relación de la variable Y respecto a X

Con un nivel de significación del 5% se rechaza la hipótesis nula (H_0). Por lo tanto, se tiene evidencia estadística para afirmar que existen una dependencia lineal del Peso Final respecto al Peso inicial.

Prueba hipótesis para los tratamientos

Probar si al menos una de las dietas produce diferente peso promedio final.

$H_0: \mu_{iajus} = \mu_{iajus}$ para todo $i = 1, 2, 3$.

H_1 : Al menos una μ_{iajus} es diferente.

Use $\alpha = 0.05$

Prueba estadística:
$$F_C = \frac{CM(Trat)_{aj.}}{CME_{aj.}} = \frac{6.79}{0.3789} = 17.92 \sim F_{(2,5)}$$

Criterio de decisión:
$$\begin{aligned} \text{Si } F_C \leq 5.79 &= F_{(0.95, 2, 5)} \text{ no se rechaza } H_0 \\ \text{Si } F_C > 5.79 &= F_{(0.95, 2, 5)} \text{ se rechaza } H_0 \end{aligned}$$



Conclusión respecto comparación delos tratamientos

Con un nivel de significación del 5% se rechaza la hipótesis nula (H_0). Por lo tanto, se tiene evidencia estadística para afirmar que al menos una dieta ocasiona un incremento en el peso distinto al de las demás dietas.

Pruebas de Comparación de medias

Compare los tratamientos utilizando la prueba de Tukey. ¿Qué tratamiento recomendaría?

Use $\alpha = 0.05$

$$\hat{\beta} = \frac{E_{XY}}{E_{XX}} = \frac{9.77}{10.0} = 0.977$$

Prueba de
Tukey

$$ALS(T) = AES(T) * \sqrt{\frac{CME_{ajust}}{2} \left(\frac{1}{r_i} + \frac{1}{r_j} + \frac{(\bar{X}_i - \bar{X}_j)^2}{E_{XX}} \right)}$$

Medias Ajustadas: $\bar{X}_{1\Box} = 4.67$ $\bar{X}_{2\Box} = 8.33$ $\bar{X}_{3\Box} = 10.33$ $\bar{X}_{\Box\Box} = 7.78$

$$\bar{Y}_{i\Boxjust} = \bar{Y}_{i\Box} - \hat{\beta}(\bar{X}_{i\Box} - \bar{X}_{\Box\Box}) \quad \bar{Y}_{1\Box} = 14.0 \quad \bar{Y}_{2\Box} = 20.13 \quad \bar{Y}_{3\Box} = 19.4 \quad \bar{Y}_{\Box\Box} = 17.84$$

$$\bar{Y}_{1\Boxjust} = \bar{Y}_{1\Box} - \hat{\beta}(\bar{X}_{1\Box} - \bar{X}_{\Box\Box}) = 14 - 0.977(4.67 - 7.78) = 17.03$$

$$\bar{Y}_{2\Boxjust} = \bar{Y}_{2\Box} - \hat{\beta}(\bar{X}_{2\Box} - \bar{X}_{\Box\Box}) = 20.13 - 0.977(8.33 - 7.78) = 19.59$$

$$\bar{Y}_{3\Boxjust} = \bar{Y}_{3\Box} - \hat{\beta}(\bar{X}_{3\Box} - \bar{X}_{\Box\Box}) = 19.4 - 0.977(10.33 - 7.78) = 16.91$$

Pruebas de Comparación de medias

$$H_0 : \mu_{1ajust} = \mu_{2ajust} \quad H_0 : \mu_{1ajust} = \mu_{3ajust} \quad H_0 : \mu_{2ajust} = \mu_{3ajust}$$

$$H_1 : \mu_{1ajust} \neq \mu_{2ajust} \quad H_1 : \mu_{1ajust} \neq \mu_{3ajust} \quad H_1 : \mu_{2ajust} \neq \mu_{3ajust}$$

$$\alpha = 0.05$$

$$AES(T) = 4.60$$

$$\underbrace{\begin{matrix} p=3 \\ GLE_{ajust}=5 \\ \alpha=0.05 \end{matrix}}$$

$$Comp A vs. B: ALS(T) = 4.60 \sqrt{\frac{0.3789}{2} \left(\frac{1}{3} + \frac{1}{3} + \frac{(4.67 - 8.33)^2}{10} \right)} = 2.836$$

$$Comp A vs. C: ALS(T) = 4.60 \sqrt{\frac{0.3789}{2} \left(\frac{1}{3} + \frac{1}{3} + \frac{(4.67 - 10.33)^2}{10} \right)} = 3.939$$

$$Comp B vs. C: ALS(T) = 4.60 \sqrt{\frac{0.3789}{2} \left(\frac{1}{3} + \frac{1}{3} + \frac{(8.33 - 10.33)^2}{10} \right)} = 2.068$$

Tabla de comparación de medias

Tratamientos A comparar	$ \bar{Y}_{i\text{ajust}} - \bar{Y}_{j\text{ajust}} $	Sd	ALS (T)	SIG.
A y B	2.562	0.8563	2.836	n.s.
A y C	0.122	0.6175	3.939	n.s.
B y C	2.684	0.4495	2.068	*

Conclusiones.

Con un nivel de significación del 5% se concluye lo siguiente:

No hay diferencias en los pesos promedio finales de las dietas A y B, tampoco hay diferencias entre las dietas A y C. Se puede afirmar que si hay diferencias en los resultados de los pesos finales de las dietas B y C



Análisis de Covarianza en un DBCA

Ejemplo de Aplicación 1.

Se desarrolló un experimento cuyo objetivo era determinar si la exposición en agua calentada artificialmente afectaba el crecimiento de las ostras. Cinco bolsas con diez ostras cada una fueron aleatoriamente asignadas a cinco temperaturas (T_1 , T_2 , T_3 , T_4 , T_5); cada bolsa constituía una unidad experimental. Se utilizaron cinco estanques, cada uno calentado a una de las cinco temperaturas. Las ostras fueron limpiadas y pesadas al comienzo y al final del experimento un mes después. El experimento se repitió cuatro veces para lo cual fueron necesarios 4 meses. Cada repetición constituye un bloque. Los pesos iniciales y finales se presentan en la siguiente tabla:

Análisis de Covarianza en un DBCA

Blo q.	T1		T2		T3		T4		T5		TOTAL	
	X	Y	X	Y	X	Y	X	Y	X	Y	X	Y
I	20.4	24.6	27.2	32.6	26.8	31.7	22.4	29.1	21.8	27.0	118.6	145.0
II	19.6	23.4	32.0	36.6	26.5	30.7	23.2	28.9	24.3	30.5	125.6	150.1
III	25.1	30.3	33.0	37.7	26.8	30.4	28.6	35.2	30.3	36.4	143.8	170.0
IV	18.1	21.8	26.8	31.0	28.6	33.8	24.4	30.2	29.3	35.0	127.2	151.8
Tot al	83.2	100.1	119.0	137.9	108.7	126.6	98.6	123.4	105.7	128.9	515.2	616.9

X: Peso inicial de las
bolsas

Y: Peso final de las
bolsas

Modelo aditivo lineal en un DBCA

$$Y_{ij} = \mu + t_i + \gamma_j + \beta(X_{ij} - \bar{X}_{..}) + \varepsilon_{ij} \quad i=1,\dots,t \quad j=1,\dots,b$$

Donde:

Y_{ij} : es el peso final de una bolsa de ostras tratada con la i-ésima temperatura de agua (tratamiento) en el j-ésimo mes (bloque).

μ : es el efecto de la media general de los pesos.

t_i : es el efecto de la i-esima temperatura del agua.

γ_j : es el efecto del j-esimo bloque.

β : es el coeficiente de regresión lineal del Y, el peso final de las ostras, sobre X, el peso inicial.

X_{ij} : es el peso inicial de una bolsa de ostras tratada con la i-ésima temperatura de agua (tratamiento) en el j-ésimo mes (bloque).

$\bar{X}_{..}$: es el peso medio inicial de las bolsas de ostras.

ε_{ij} : es el efecto del error experimental con la i-esima temperatura de agua, en el j-esimo mes.

$t=5$ (número de tratamientos).

$b=4$ (número de bloques).

Desarrollo del ejemplo de aplicación 1.

Prueba de la pendiente

Prueba de Hipótesis para el Coeficiente de Regresión o de la pendiente

Hipótesis:

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

Estadístico de Prueba:

$$F_C = \frac{\frac{E^2_{xy}}{E_{xx}}}{CME_{aj.}} \sim F_{(1, gl(Error aj.))}$$

Regla de Decisión:

La hipótesis nula se rechaza con un nivel de significación α si el F_C resulta mayor que el valor de tabla $F_{(1-\alpha, 1, gl(Error aj.))}$.

Desarrollo del ejemplo de aplicación 1.

Prueba de la pendiente

$H_0: \beta = 0$ (El peso final de las ostras no depende linealmente del peso inicial)

$H_1: \beta \neq 0$ (El peso final de las ostras sí depende linealmente del peso inicial)

$$\alpha = 0.05$$

$$F_C = \frac{\frac{E^2_{xy}}{E_{xx}}}{\text{CME}_{aj.}} = \frac{\frac{74.50^2}{64.63}}{0.2743} = 313.05 \sim F_{(1,11)}$$

Criterio de
decisión:

$\text{Si } F_C \leq 4.84 = F_{(0.95,1,11)}$ no se rechaza H_0

$\text{Si } F_C > 4.84 = F_{(0.95,1,11)}$ se rechaza H_0

Conclusión:

Con un nivel de significación del 5% se rechaza H_0 y se concluye que existe suficiente evidencia estadística para aceptar que el peso final de las ostras depende linealmente del peso inicial.

Desarrollo del ejemplo de aplicación 1.

Prueba de tratamientos

Prueba de Hipótesis para los efectos de los tratamientos

Hipótesis:

$H_0: \mu_{i\text{ aj.}} = \mu_{\text{ aj.}} \text{ para } i = 1, 2, 3, 4, 5$

$\alpha = 0.05$

$H_1: \mu_{i\text{ aj.}} \neq \mu_{\text{ aj.}} \text{ para al menos algún } i$

ó literalmente:

H_0 : Las cinco temperaturas son igualmente efectivas en el crecimiento de las ostras.

H_1 : Con al menos una de las temperaturas se obtienen resultados diferentes en el crecimiento de ostras.

$$F_c = \frac{CM(\text{Trat aj.})}{CME_{\text{ aj.}}} = \frac{3.1493}{0.2743} = 11.48 \sim F_{(4,11)}$$

Desarrollo del ejemplo de aplicación 1

Criterio de
decisión:

$\text{Si } F_C \leq 3.36 = F_{(0.95, 4, 11)}$ *no se rechaza H_0*

$\text{Si } F_C > 3.36 = F_{(0.95, 4, 11)}$ *se rechaza H_0*

Conclusión:

Con un nivel de significación del 5%, se rechaza H_0 y se concluye que existe suficiente evidencia estadística para aceptar que con al menos en una temperatura se obtiene un peso final diferente para las ostras.

Pruebas de Comparación de Medias de Tratamientos

Pendiente:

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}}$$

Las medias de los tratamientos ajustadas por la regresión, el cual es dado por:

$$\bar{Y}_{j,aj} = \bar{Y}_j - \hat{\beta}(\bar{X}_j - \bar{X}_{..})$$

Las desviaciones estándar para las pruebas son:

Prueba t y DSL

$$s_d = \sqrt{CME_{aj} \left[\frac{1}{r_i} + \frac{1}{r_j} + \frac{(\bar{X}_i - \bar{X}_j)^2}{E_{xx}} \right]}$$

Tukey

$$s_d = \sqrt{\frac{CME_{aj,dy}}{2} \left[\frac{1}{r_i} + \frac{1}{r_j} + \frac{(\bar{X}_i - \bar{X}_j)^2}{E_{xx}} \right]}$$

Dunnett

$$s_d = \sqrt{CME_{aj} \left[\frac{1}{r_T} + \frac{1}{r_i} + \frac{(\bar{X}_T - \bar{X}_i)^2}{E_{xx}} \right]}$$

Desarrollo del ejemplo de aplicación 1.

Comparación de medias (continuación)

Efectué la prueba de Tukey.

Las hipótesis son las siguientes:

$$H_0: \mu_{i \text{ aj.}} = \mu_{j \text{ aj.}} \quad \forall i, j = 1, 2, \dots, 5, \text{ con } i \neq j$$

$$H_1: \mu_{i \text{ aj.}} \neq \mu_{j \text{ aj.}}$$

El coeficiente de regresión estimado es:

$$\hat{\beta} = \frac{E_{xy}}{E_{xx}} = \frac{74.50}{64.63} = 1.1527$$

Las medias de Y ajustadas para cada tratamiento según la formula

$$\bar{Y}_{i \cdot \text{aj.}} = \bar{Y}_{i \cdot} - \hat{\beta}(\bar{X}_{i \cdot} - \bar{X}_{..})$$

Son:

$$\bar{Y}_{1 \cdot \text{aj.}} = 30.74 \quad \bar{Y}_{2 \cdot \text{aj.}} = 29.88 \quad \bar{Y}_{3 \cdot \text{aj.}} = 30.02 \quad \bar{Y}_{4 \cdot \text{aj.}} = 32.13 \quad \bar{Y}_{5 \cdot \text{aj.}} = 31.46$$

Desarrollo del ejemplo de aplicación 1.

Comparación de medias (continuación)

El valor de tabla con $\alpha = 5\%$, $p = 5$ tratamientos y 11 grados de libertad para el error ajustado es $AES(T) = 4.57$. La amplitud limite significativa de Tukey está dada por la siguiente fórmula:

$$ALS(T) = AES(T) \sqrt{\frac{CME_{aj.}}{2} \left[\frac{2}{b} + \frac{(\bar{X}_{i.} - \bar{X}_{j.})^2}{E_{xx}} \right]}$$

Donde $b = 4$, $CME_{aj.} = 0.2743$ y $E_{xx} = 64.63$

ALS: amplitud limite significativa

AES: amplitud estudentizado significativa

Desarrollo del ejemplo de aplicación 1.

Comparación de medias (continuación)

Tratamientos comparados	$ \bar{Y}_{i\cdot} - \bar{Y}_{j\cdot} $	S_d	ALS(T)	Significancia
1 y 2	0.867	0.488	2.232	n.s.
1 y 3	0.724	0.393	1.789	n.s.
1 y 4	1.387	0.316	1.445	n.s.
1 y 5	0.716	0.368	1.684	n.s.
2 y 3	0.143	0.287	1.314	n.s.
2 y 4	2.254	0.352	1.608	*
2 y 5	1.583	0.303	1.386	*
3 y 4	2.111	0.287	1.310	*
3 y 5	1.440	0.264	1.207	*
4 y 5	0.671	0.274	1.254	n.s.

ALS: amplitud limite significativa