<table>
<tr><th colspan="3">RESUMEN COMPARATIVO DE MODELOS ANOVA CON UN FACTOR DE TRATAMIENTOS EN ESTRUCTURAS DE DISEÑO COMPLETAMENTE ALEATORIZADA (DCA)</th></tr>
<tr><th></th><th>Estructura de tratamientos de efectos fijos</th><th>Estructura de tratamientos de efectos aleatorios</th></tr>
<tr><td>Definición de los a niveles del factor de estudio</td><td>Los a niveles de tratamientos observados en el experimento son seleccionados a criterio del investigador (los que le interesa estudiar) y por tanto cualquier inferencia se limita únicamente a los niveles seleccionados.</td><td>Los a niveles de tratamientos observados en el experimento son seleccionados aleatoriamente de una población grande de posibles niveles, sin embargo, la inferencias a realizar se extrapolan a toda la población de niveles.</td></tr>
<tr><td>Objetivo</td><td>Determinar si los niveles considerados tienen efectos sobre la media de la variable respuesta, o equivalentemente, si al menos para un par de niveles hay diferencia en la media de la respuesta.</td><td>Determinar si la variación de los niveles del factor (la aleatoriedad de estos) contribuye a la varianza total de la respuesta, es decir si contribuyen a $Var(Y_{ij})$.</td></tr>
<tr><td>Ejecución del experimento</td><td>Determinado el número de réplicas para cada tratamiento (los tamaños de muestra en cada nivel, n_i, $i = 1, 2, \dots, a$) se seleccionan y asignan aleatoriamente $N = \sum_{i=1}^{a} n_i$ unidades experimentales (U.E) a los tratamientos, las cuales deben ser homogéneas entre sí. Luego, en orden completamente al azar, se mide la variable respuesta a cada U.E según tratamiento asignado.</td><td>Se seleccionan aleatoriamente los a niveles de tratamiento y determinado el número de réplicas para cada tratamiento seleccionado (los tamaños de muestra en cada nivel, n_i, $i = 1, 2, \dots, a$), se procede como en el caso de efectos fijos.</td></tr>
<tr><td>Modelo ANOVA</td><td>Modelo de efectos (fijos) de tratamientos:

$\boldsymbol{Y_{ij} = \mu + \alpha_i + E_{ij}}$, con $\boldsymbol{E_{ij} \sim IID\ N(0, \sigma^2)}$ y sujeto a $\boldsymbol{\sum_{i=1}^{a} n_i \alpha_i = 0}$.

Nota: Si el diseño es balanceado, es decir, si $n_i = n\ \forall i,\ i = 1, 2, \dots, a$, la restricción se reduce a $\sum_{i=1}^{a} \alpha_i = 0$

O bien,

Modelo de medias de tratamientos:

$\boldsymbol{Y_{ij} = \mu_i + E_{ij}}$, con $\boldsymbol{E_{ij} \sim IID\ N(0, \sigma^2)}$.

Donde,
• Y_{ij} es la respuesta en la j – ésima réplica del tratamiento o nivel i del factor;
• μ es la media global de la variable respuesta
• α_i es el efecto fijo del tratamiento o nivel i del factor sobre la media de la variable respuesta. Se dice que son efectos fijos es decir un parámetro y no una variable aleatoria, pues si se repite el experimento siempre estarán presentes los mismos niveles del factor y por tanto los mismos efectos sobre la media de la variable respuesta.
• μ_i es un parámetro que representa la media de la variable respuesta en el tratamiento o nivel i del factor y es tal que $\mu_i = \mu + \alpha_i$;
• E_{ij} es el error aleatorio en la j – ésima réplica del tratamiento o nivel i del factor.</td><td>Modelo de efectos aleatorios:

$\boldsymbol{Y_{ij} = \mu + A_i + E_{ij}}$, con $\boldsymbol{E_{ij} \sim IID\ N(0, \sigma^2)}$, $\boldsymbol{A_i \sim IID\ N(0, \sigma_\alpha^2)}$ y los A_i y los E_{ij} son mutuamente independientes $\forall i, j,\ i = 1, 2, \dots, a$,

Donde,
• Y_{ij} es la respuesta en la j – ésima réplica del tratamiento o nivel i del factor;
• μ es la media global de la variable respuesta
• A_i es el efecto aleatorio del tratamiento o nivel i del factor sobre la media de la variable respuesta. Desde que los niveles observados fueron seleccionados aleatoriamente, estos son una muestra de una variable aleatoria y por los supuestos del modelo, su media es cero y su varianza σ_α^2, de forma que si se repite el experimento no hay por qué esperar que se observen los mismos niveles del factor.
• E_{ij} es el error aleatorio en la j – ésima réplica del tratamiento o nivel i del factor.</td></tr>
<tr><td>Correlación entre las Y_{ij}</td><td>Las Y_{ij} son mutuamente independientes y por tanto no están correlacionadas.</td><td>Los efectos aleatorios inducen correlación (y por tanto dependencia estadística) entre las variables respuesta que tengan el mismo nivel i, y es tal que $Cov(Y_{ij}, Y_{is}) = \sigma_\alpha^2,\ \forall j \neq s$.</td></tr>
<tr><td>Distribución de Y_{ij} y de $\bar{Y}_{i\bullet}$</td><td>$\boldsymbol{Y_{ij} \sim N(\mu_i, \sigma^2) = N(\mu + \alpha_i, \sigma^2)}$ y son mutuamente independientes, aunque no todas son idénticamente distribuidas desde que la media cambia con el nivel del factor. Observe que la varianza total de la variable respuesta es $\boldsymbol{Var(Y_{ij}) = \sigma^2}$.

$\boldsymbol{\hat{\mu}_i = \bar{Y}_{i\bullet} \sim N\left(\mu_i, \frac{\sigma^2}{n_i}\right)}$ y son mutuamente independientes, aunque no son idénticamente distribuidas desde que la media cambia con el nivel del factor y los tamaños de muestra pueden diferir también. $\bar{Y}_{i\bullet}$ es el promedio aritmético de las n_i réplicas de la respuesta en el nivel i.</td><td>$\boldsymbol{Y_{ij} \sim N(\mu, \sigma_\alpha^2 + \sigma^2)}$, pero no todas son mutuamente independientes (solo las respuestas que no comparten el mismo nivel i). Observe que la varianza total de la variable respuesta es $\boldsymbol{Var(Y_{ij}) = \sigma_\alpha^2 + \sigma^2}$.

$\boldsymbol{\bar{Y}_{i\bullet} \sim N\left(\mu, \sigma_\alpha^2 + \frac{\sigma^2}{n_i}\right)}$ y son mutuamente independientes aunque no son idénticamente distribuidas cuando el diseño es desbalanceado, es decir cuando no se tiene mismo número de réplicas por nivel del factor de estudio. $\bar{Y}_{i\bullet}$ es el promedio aritmético de las n_i réplicas de la respuesta en el nivel i.</td></tr>
<tr><td>ANOVA</td><td>
<table>
<tr><th colspan="6">Tabla ANOVA</th></tr>
<tr><th>Fuente</th><th>gl</th><th>SC</th><th>CM</th><th>Esperanza de CM</th><th>F_0</th></tr>
<tr><td>Factor</td><td>$a - 1$</td><td>SSA</td><td>$MSA = SSA/(a-1)$</td><td>$\boldsymbol{E(MSA) = \sigma^2 + \frac{\sum_{i=1}^{a} n_i \alpha_i^2}{(a-1)}}$</td><td>$\frac{MSA}{MSE}$</td></tr>
<tr><td>Error</td><td>$N - a$</td><td>SSE</td><td>$MSE = SSE/(N-a)$</td><td>$E(MSE) = \sigma^2$</td><td></td></tr>
<tr><td>Total</td><td>$N - 1$</td><td>SST</td><td></td><td></td><td></td></tr>
</table>
• $SSA = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{a} n_i \bar{Y}_{i\bullet}^2 - N\bar{Y}_{\bullet\bullet}^2$, Variabilidad explicada por el factor
• $SSE = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\bullet})^2 = \sum_{i=1}^{a} (n_i - 1) S_i^2$Variabilidad debida al error aleatorio
• $SST = SSA + SSE = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{\bullet\bullet})^2$ Variabilidad total observada
• S_i^2 varianza muestral en el nivel i, $\bar{Y}_{\bullet\bullet}$ media muestral de la N observaciones</td><td>
<table>
<tr><th colspan="6">Tabla ANOVA</th></tr>
<tr><th>Fuente</th><th>gl</th><th>SC</th><th>CM</th><th>Esperanza de CM</th><th>F_0</th></tr>
<tr><td>Factor</td><td>$a - 1$</td><td>SSA</td><td>$MSA = SSA/(a-1)$</td><td>$\boldsymbol{E(MSA) = \sigma^2 + c \times \sigma_\alpha^2}$</td><td>$\frac{MSA}{MSE}$</td></tr>
<tr><td>Error</td><td>$N - a$</td><td>SSE</td><td>$MSE = SSE/(N-a)$</td><td>$E(MSE) = \sigma^2$</td><td></td></tr>
<tr><td>Total</td><td>$N - 1$</td><td>SST</td><td></td><td></td><td></td></tr>
</table>
Donde $c = \frac{1}{N(a-1)}[N^2 - \sum_{i=1}^{a} n_i^2]$. Si $n_i = n\ \forall i,\ i = 1, 2, \dots, a$, entonces $c = n$.

• Las sumas de cuadrados se calculan de la misma forma que en modelo con efectos fijos.</td></tr>
</table>

RESUMEN COMPARATIVO DE MODELOS ANOVA CON UN FACTOR DE TRATAMIENTOS EN ESTRUCTURAS DE DISEÑO COMPLETAMENTE ALEATORIZADA (DCA)		
	Estructura de tratamientos de efectos fijos	**Estructura de tratamientos de efectos aleatorios**
Test ANOVA	*Prueba si hay efectos significativos del factor sobre la media de la respuesta*: $\boldsymbol{H_0: \mu_1 = \mu_2 = \cdots, \mu_a}$ **vs.** $\boldsymbol{H_1: \mu_i \neq \mu_j}$ **para al menos un par de niveles de tratamientos,** o equivalentemente, $\boldsymbol{H_0: \alpha_1 = \alpha_2 = \cdots, \alpha_a = 0}$ **vs.** $\boldsymbol{H_1: \alpha_i \neq 0}$ **para al menos un nivel** $\boldsymbol{i}$**.** • Bajo H_0 se dice que el factor no tiene efectos significativos sobre la media de la variable respuesta, en los niveles considerados. • Bajo H_1 se dice que el factor tiene efectos significativos sobre media de la variable respuesta, en los niveles considerados. • Bajo H_0 y los supuestos del modelo, el estadístico de prueba es $F_0 = \frac{MSA}{MSE} \sim f_{a-1,N-a}$ y se rechaza H_0 si $P(f_{a-1,N-a} > F_0)$ es pequeño.	*Prueba si el factor de estudio contribuye significativamente a la varianza total de la respuesta, es decir, si hay componente de varianza debida al factor de estudio:* $\boldsymbol{H_0: \sigma_\alpha^2 = 0}$ **vs.** $\boldsymbol{H_1: \sigma_\alpha^2 > 0}$**.** *(No confundir con test de homogeneidad de varianza de los errores)* • Bajo H_0 se dice que el factor no contribuye significativamente a la varianza total de la variable respuesta, o equivalentemente, no hay componente de varianza debida al factor de estudio. • Bajo H_1 se dice que el factor contribuye significativamente a la varianza total de la variable respuesta, o equivalentemente, hay componente de varianza debida al factor de estudio. • Bajo H_0 y los supuestos del modelo, el estadístico de prueba es $F_0 = \frac{MSA}{MSE} \sim f_{a-1,N-a}$ y se rechaza H_0 si $P(f_{a-1,N-a} > F_0)$ es pequeño.
Inferencias y estimaciones de interés	***Si se rechaza H_0 en el test ANOVA:*** **Estimaciones de (ver en notas de clase distribuciones e I.C):** • Medias de tratamiento: $\hat{\mu}_i = \bar{Y}_{i\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$ • Media global: $\hat{\mu} = \bar{Y}_{\bullet\bullet}$ • Efectos de tratamientos: $\hat{\alpha}_i = \bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}$, o bien $\hat{\alpha}_i = \frac{a-1}{a}\bar{Y}_{i\bullet} - \frac{1}{a}\sum_{j\neq i}\bar{Y}_{j\bullet}$ (esta última fórmula sólo cuando el diseño es balanceado) • Varianza, $\hat{\sigma}^2 = MSE$ **Inferencias:** **Comparaciones de medias de tratamientos** • ***Comparaciones múltiples por pares de medias:*** Se realizan todas las comparaciones por pares de medias $H_0: \mu_i = \mu_j$ vs. $H_1: \mu_i \neq \mu_j$. Luego, se agrupan las medias en grupos tales que todas las medias incluidas en un grupo dado son estadísticamente iguales (*no se aplica igualdad por transitividad*). *Bajo validez de los supuestos sobre los errores del modelo*, la igualdad por pares de medias se realiza así: 1. Método LSD: Usando mínima diferencia significativa (LSD) a un nivel de significancia γ se rechaza H_0 si $\lvert\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}\rvert \geq t_{\gamma/2,N-a}\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$, o usando I.C de $(1-\gamma)100\%$ de confianza para $\mu_i - \mu_j$, se rechaza H_0 si cero no pertenece al intervalo $(\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}) \pm t_{\gamma/2,N-a} \times \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$, donde $t_{\gamma/2,N-a}$ es el percentil crítico de nivel $\gamma/2$ en la distribución t-Student con $N-a$ grados de libertad. 2. Método Tukey: Usando diferencia "honestamente significativa" (HSD) a un nivel de significancia γ se rechaza H_0 si $\lvert\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}\rvert \geq \frac{1}{\sqrt{2}}\, q_\gamma(a, N-a)\sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$, o usando I.C de $(1-\gamma)100\%$ de confianza de Tukey para $\mu_i - \mu_j$, se rechaza H_0 si cero no pertenece al intervalo $(\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}) \pm \frac{1}{\sqrt{2}}\, q_\gamma(a, N-a) \times \sqrt{MSE\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}$, donde $q_\gamma(a, N-a)$ es el rango crítico estudentizado de nivel γ para a medias y con $N-a$ grados de libertad. 3. Otros métodos: Duncan, Dunnet (comparaciones con un nivel de control), ver notas de clase. • ***Comparaciones por contrastes de medias de tratamientos:*** Un contraste de medias de tratamientos es una combinación lineal de medias así $W = \sum_{i=1}^{a} c_i\mu_i$, tal que $\sum_{i=1}^{a} c_i = 0$. Puede definirse un contraste para comparar el promedio de las medias de un grupo de tratamientos contra el promedio de las medias de otro grupo de tratamientos, definiendo los pesos o ponderadores c_i de acuerdo a la conformación de los dos grupos de medias a contrastar. El estimador del contraste es $\widehat{W} = \sum_{i=1}^{a} c_i\bar{Y}_{i\bullet}$. Bajo validez de los supuestos sobre los errores del modelo, $\widehat{W} \sim N\left(W, \sigma^2\sum_{i=1}^{a}\frac{c_i^2}{n_i}\right)$. Las inferencias pueden ser mediante: 1. Tests de hipótesis sobre el contraste (ver notas de clase) 2. I.C para el contraste de $(1-\gamma)100\%$ de confianza $\widehat{W} \pm t_{\gamma/2,N-a}\sqrt{MSE\sum_{i=1}^{a}\frac{c_i^2}{n_i}}$	***Si se rechaza H_0 en el test ANOVA:*** **Estimaciones de (ver en notas de clase las distribuciones e I.C):** • Las componentes de varianza: 1. Debida al error aleatorio: $\hat{\sigma}^2 = MSE$ 2. Debida al factor de estudio: $\hat{\sigma}_\alpha^2 = \frac{MSA-MSE}{c}$ (si el diseño es balanceado $c = n$). 3. Varianza total: $\widehat{Var}(Y_{ij}) = \hat{\sigma}_\alpha^2 + \hat{\sigma}^2$ • La razón de componentes de varianzas: $\hat{\sigma}_\alpha^2/\hat{\sigma}^2$ • Proporción de la varianza total debida al factor de estudio: $\frac{\hat{\sigma}_\alpha^2}{\hat{\sigma}_\alpha^2 + \hat{\sigma}^2}$ • Media global: $\hat{\mu} = \bar{Y}_{\bullet\bullet}$ ***No son válidas las comparaciones de medias de tratamientos por ningún método***, pues los efectos de niveles del factor son aleatorios*!*

Análisis de varianza e inferencias sobre medias de tratamiento en modelo con efectos fijos, bajo no homogeneidad de varianza en errores del modelo (Ver Apéndice A.6 en notas de clase)

- **Para el test ANOVA:** Cuando el diseño es balanceado,

$$F_0 = \frac{MSA}{MSE} \sim f_{\nu_1,\nu_2} \text{ (aproximadamente)},$$

Con $\nu_1 = \frac{a-1}{1+c^2\frac{(a-2)}{a-1}}$, $\nu_2 = \frac{a(n-1)}{1+c^2}$, donde $c^2 = \frac{\sum_{i=1}^{a}\left(S_i^2-\bar{S}^2\right)^2}{a(\bar{S}^2)^2}$, siendo $\bar{S}^2 = \frac{1}{a}\sum_{i=1}^{a} S_i^2$ y los S_i^2 la varianza muestral en los respectivos niveles i del factor de tratamientos.

- **Para I.C de medias de tratamientos, bajo varianza no constante:** Un I.C del $(1-\gamma)100\%$ para μ_i sería simplemente

$$\bar{Y}_{i\bullet} \pm t_{\gamma/2,n_i-1}\frac{S_i}{\sqrt{n_i}}$$

- **En I.C de comparaciones por pares de medias**: Por método LSD tendríamos por la aproximación Sattertwaite (sólo LSD puede usarse porque en los demás métodos no hay correcciones por varianza no constante)

$$\left(\mu_i-\mu_j\right) \in \left(\bar{Y}_{i\bullet}-\bar{Y}_{j\bullet}\right) \pm t_{\gamma/2,\nu} \times \sqrt{\frac{S_i^2}{n_i}+\frac{S_j^2}{n_j}}, \qquad \text{con} \qquad \nu = \frac{\left(\frac{S_i^2}{n_i}+\frac{S_j^2}{n_j}\right)^2}{\frac{\left(\frac{S_i^2}{n_i}\right)^2}{n_i-1}+\frac{\left(\frac{S_j^2}{n_j}\right)^2}{n_j-1}}$$

- **Criterio de decisión en las comparaciones LSD:** Rechazar $H_0\colon \mu_i = \mu_j$, en favor de $H_1\colon \mu_i \neq \mu_j$, si $\left|\bar{Y}_{i\bullet}-\bar{Y}_{j\bullet}\right| \geq t_{\gamma/2,\nu} \times \sqrt{\frac{S_i^2}{n_i}+\frac{S_j^2}{n_j}}$ con $\nu = \frac{\left(\frac{S_i^2}{n_i}+\frac{S_j^2}{n_j}\right)^2}{\frac{\left(\frac{S_i^2}{n_i}\right)^2}{n_i-1}+\frac{\left(\frac{S_j^2}{n_j}\right)^2}{n_j-1}}$

- **Inferencias sobre contrastes de medias:** La aproximación Sattertwaite debe ser usada para hacer inferencias bajo varianza no constante. Sea $W = \sum_{i=1}^{a} c_i\mu_i = \sum_{i=1}^{a} c_i\alpha_i$, desde que $\sum_{i=1}^{a} c_i = 0$, entonces su estimador es $\widehat{W} = \sum_{i=1}^{a} c_i\bar{Y}_{i\bullet}$

  Bajo varianza no constante $\widehat{W} \sim N\left(W, \sum_{i=1}^{a}\frac{c_i^2\sigma_i^2}{n_i}\right)$. La varianza estimada corresponde a $\widehat{\text{Var}}\left(\widehat{W}\right) = \sum_{i=1}^{a}\frac{c_i^2 S_i^2}{n_i}$, y de acuerdo con las ecuaciones (A.11) y (A.12) en notas de clase, tenemos que:

$$T_0 = \frac{\widehat{W}-W}{\sqrt{\sum_{i=1}^{a}\frac{c_i^2S_i^2}{n_i}}} \sim\ t_\nu \text{ (aproximadamente)}, \qquad \text{con} \qquad \nu = \frac{\left(\sum_{i=1}^{a}\frac{c_i^2S_i^2}{n_i}\right)^2}{\sum_{i=1}^{a}\frac{\left(\frac{c_i^2S_i^2}{n_i}\right)^2}{n_i-1}}.$$

  Luego un I.C aproximado del $(1-\gamma)100\%$ para W es

$$\widehat{W} \pm t_{\gamma/2,\nu}\sqrt{\sum_{i=1}^{a}\frac{c_i^2S_i^2}{n_i}}, \qquad \text{con} \qquad \nu = \frac{\left(\sum_{i=1}^{a}\frac{c_i^2S_i^2}{n_i}\right)^2}{\sum_{i=1}^{a}\frac{\left(\frac{c_i^2S_i^2}{n_i}\right)^2}{n_i-1}}$$

<table>
<tr><th colspan="3">RESUMEN COMPARATIVO DE MODELOS ANOVA CON UN FACTOR DE ESTUDIO DE EFECTOS FIJOS EN ESTRUCTURAS DE DISEÑO DCA VS. DBCA</th></tr>
<tr><th></th><th>DCA (Diseño completamente aleatorizado)</th><th>DBCA (Diseño en bloques completos aleatorizados con tamaños de bloques igual a a)</th></tr>
<tr><td>Objetivo</td><td>Determinar si los niveles considerados tienen efectos sobre la media de la variable respuesta, o equivalentemente, si al menos para un par de niveles del factor de tratamientos hay diferencia en la media de la respuesta.</td><td>Comparar las medias de los a tratamientos o niveles del factor de estudio (es decir, determinar si los tratamientos tienen efectos sobre la media de la variable respuesta, o equivalentemente, si al menos para un par de niveles del factor de tratamientos hay diferencia en la media de la respuesta), <u>promediados sobre un rango de condiciones diferentes</u></td></tr>
<tr><td>Ejecución del experimento</td><td>Determinado el número de réplicas para cada tratamiento (los tamaños de muestra en cada nivel, n_i, $i = 1, 2, \dots, a$) <u>se seleccionan y asignan aleatoriamente las $\boldsymbol{N = \sum_{i=1}^{a} n_i}$ unidades experimentales (U.E) a los tratamientos, las cuales deben ser homogéneas entre sí</u>. Luego, en orden completamente al azar, se mide la variable respuesta a cada U.E según tratamiento asignado.</td><td>Determinado el número de bloques b, cada uno de tamaño a (el número de réplicas por tratamiento es igual al número de bloques, es decir, $\boldsymbol{n_{i\bullet} = b, \forall i = 1, 2, \dots, a}$ y el número de U.E dentro de cada bloque es $\boldsymbol{n_{\bullet j} = a, \forall j = 1, 2, \dots, b}$), en cada uno de los bloques se seleccionan y asignan aleatoriamente las U.E, una por tratamiento (<u>las U.E que pertenecen a un mismo bloque son homogéneas entre sí pero las U.E en diferentes bloques son heterogéneas entre sí</u>). Luego, se observan al azar los tratamientos en cada bloque. El total de observaciones es $\boldsymbol{N = \sum_{i=1}^{a} n_{i\bullet} = ab}$.</td></tr>
<tr><td>Modelo ANOVA</td><td>Modelo de efectos fijos de tratamientos:

$\boldsymbol{Y_{ij} = \mu + \alpha_i + E_{ij}}$, con $\boldsymbol{E_{ij} \sim IID\ N(0, \sigma^2)}$ y sujeto a $\boldsymbol{\sum_{i=1}^{a} n_i \alpha_i = 0}$.

Si el diseño es balanceado, es decir, si $n_i = n\ \forall i$, $i = 1, 2, \dots, a$, la restricción se reduce a $\sum_{i=1}^{a} \alpha_i = 0$

Donde,
Y_{ij} es la respuesta en la j – ésima réplica del tratamiento o nivel i del factor de tratamientos;
μ es la media global de la variable respuesta;
α_i es el efecto fijo del tratamiento o nivel i del factor de tratamientos sobre la media de la variable respuesta;
Las medias de tratamientos son $\mu_i = \mu + \alpha_i$; $i = 1, 2, \dots, a$;
E_{ij} es el error aleatorio en la j – ésima réplica del tratamiento o nivel i del factor de tratamientos.</td><td>Modelo de efectos fijos de tratamientos y de bloques: Bajo la presunción de que no hay “interacción” entre factor de tratamientos y de bloques,

$\boldsymbol{Y_{ij} = \mu + \alpha_i + \beta_j + E_{ij}}$, con $\boldsymbol{E_{ij} \sim IID\ N(0, \sigma^2)}$ y sujeto a $\boldsymbol{\sum_{i=1}^{a} \alpha_i = 0}$ y $\boldsymbol{\sum_{j=1}^{b} \beta_j = 0}$.

Donde,
Y_{ij} es la respuesta al tratamiento o nivel i del factor de tratamientos en el j – ésimo bloque;
μ es la media global de la variable respuesta;
α_i es el efecto fijo del tratamiento o nivel i del factor de tratamientos sobre la media de la variable respuesta;
β_j es el efecto fijo del j – ésimo bloque sobre la media de la variable respuesta;
Las medias de tratamiento son $\mu_{i\bullet} = \mu + \alpha_i$, $i = 1, 2, \dots, a$;
E_{ij} es el error aleatorio en la respuesta en el j – ésimo bloque con tratamiento i.</td></tr>
<tr><td>Correlación entre las Y_{ij}</td><td>Las Y_{ij} son mutuamente independientes y por tanto no están correlacionadas.</td><td>Las Y_{ij} son mutuamente independientes y por tanto no están correlacionadas</td></tr>
<tr><td>Distribución de Y_{ij} y de $\bar{Y}_{i\bullet}$</td><td>$\boldsymbol{Y_{ij} \sim N(\mu_i, \sigma^2) = N(\mu + \alpha_i, \sigma^2)}$ y son mutuamente independientes, aunque no todas son idénticamente distribuidas desde que la media cambia con el nivel del factor de tratamientos.
$\boldsymbol{\hat{\mu}_i = \bar{Y}_{i\bullet} \sim N\left(\mu_i, \frac{\sigma^2}{n_i}\right) = N\left(\mu + \alpha_i, \frac{\sigma^2}{n_i}\right)}$ y son mutuamente independientes, aunque no son idénticamente distribuidas desde que la media cambia con el nivel del factor y los tamaños de muestra pueden diferir también. $\bar{Y}_{i\bullet}$ es el promedio aritmético de las n_i réplicas de la respuesta en el nivel o tratamiento i.</td><td>$\boldsymbol{Y_{ij} \sim N(\mu_{i\bullet} + \beta_j, \sigma^2) = N(\mu + \alpha_i + \beta_j, \sigma^2)}$ y son mutuamente independientes, aunque no todas son idénticamente distribuidas desde que la media cambia con el nivel del factor de tratamientos y de bloques.
$\boldsymbol{\hat{\mu}_{i\bullet} = \bar{Y}_{i\bullet} \sim N\left(\mu_{i\bullet}, \frac{\sigma^2}{b}\right) = N\left(\mu + \alpha_i, \frac{\sigma^2}{b}\right)}$ y son mutuamente independientes, aunque no son idénticamente distribuidas desde que la media cambia con el nivel del factor de tratamientos. $\bar{Y}_{i\bullet}$ es el promedio aritmético de las $\boldsymbol{n_{i\bullet} = b}$ réplicas de la respuesta en el nivel o tratamiento i.</td></tr>
<tr><td>ANOVA</td><td><table>
<tr><th colspan="6">Tabla ANOVA</th></tr>
<tr><th>Fuente</th><th>gl</th><th>SC</th><th>CM</th><th>Esperanza de CM</th><th>F_0</th></tr>
<tr><td>Factor</td><td>$a-1$</td><td>SSA</td><td>$MSA = SSA/(a-1)$</td><td>$\boldsymbol{E(MSA) = \sigma^2 + \frac{\sum_{i=1}^{a} n_i \alpha_i^2}{(a-1)}}$</td><td>$\frac{MSA}{MSE}$</td></tr>
<tr><td>Error</td><td>$N-a$</td><td>SSE</td><td>$MSE = SSE/(N-a)$</td><td>$E(MSE) = \sigma^2$</td><td></td></tr>
<tr><td>Total</td><td>$N-1$</td><td>SST</td><td></td><td></td><td></td></tr>
</table>
$SSA = \sum_{i=1}^{a} \sum_{j=1}^{n_i} (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 = \sum_{i=1}^{a} n_i \bar{Y}_{i\bullet}^2 - N\bar{Y}_{\bullet\bullet}^2$, Variabilidad explicada por el factor de tratamientos
$SSE = \sum_{i=1}^{a} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_{i\bullet}\right)^2 = \sum_{i=1}^{a} (n_i - 1) S_i^2$ Variabilidad debida al error aleatorio
$SST = SSA + SSE = \sum_{i=1}^{a} \sum_{j=1}^{n_i} \left(Y_{ij} - \bar{Y}_{\bullet\bullet}\right)^2$ Variabilidad total observada
S_i^2 varianza muestral en el nivel i; $\bar{Y}_{\bullet\bullet}$ media muestral de la N observaciones
Respuesta ajustada $\hat{Y}_{ij} = \bar{Y}_{i\bullet}$.</td><td><table>
<tr><th colspan="6">Tabla ANOVA</th></tr>
<tr><th>Fuente</th><th>gl</th><th>SC</th><th>CM</th><th>Esperanza de CM</th><th>F_0</th></tr>
<tr><td>Factor</td><td>$a-1$</td><td>SSA</td><td>$MSA = SSA/(a-1)$</td><td>$\boldsymbol{E(MSA) = \sigma^2 + \frac{b\sum_{i=1}^{a} \alpha_i^2}{(a-1)}}$</td><td>$\frac{MSA}{MSE}$</td></tr>
<tr><td>Bloques</td><td>$b-1$</td><td>SSB</td><td>$MSB = SSB/(b-1)$</td><td>$\boldsymbol{E(MSB) = \sigma^2 + \frac{a\sum_{j=1}^{b} \beta_j^2}{(b-1)}}$</td><td>$\frac{MSB}{MSE}$</td></tr>
<tr><td>Error</td><td>$(a-1)(b-1)$</td><td>SSE</td><td>$MSE = \frac{SSE}{(a-1)(b-1)}$</td><td>$E(MSE) = \sigma^2$</td><td></td></tr>
<tr><td>Total</td><td>$N-1 = ab-1$</td><td>SST</td><td></td><td></td><td></td></tr>
</table>
$SSA = \sum_{i=1}^{a} \sum_{j=1}^{b} (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2 = b\sum_{i=1}^{a} \bar{Y}_{i\bullet}^2 - ab\bar{Y}_{\bullet\bullet}^2$, Variabilidad explicada por factor de tratamientos
$SSB = \sum_{i=1}^{a} \sum_{j=1}^{b} \left(\bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet}\right)^2 = a\sum_{j=1}^{b} \bar{Y}_{\bullet j}^2 - ab\bar{Y}_{\bullet\bullet}^2$, Variabilidad explicada por los bloques
$SSE = \sum_{i=1}^{a} \sum_{j=1}^{b} \left(Y_{ij} - \bar{Y}_{i\bullet} - \bar{Y}_{\bullet j} + \bar{Y}_{\bullet\bullet}\right)^2 = SST - SSA - SSB$ Variabilidad debida al error aleatorio
$SST = SSA + SSB + SSE = \sum_{i=1}^{a} \sum_{j=1}^{b} \left(Y_{ij} - \bar{Y}_{\bullet\bullet}\right)^2$ Variabilidad total observada
$S_{i\bullet}^2$ varianza muestral en el nivel i de tratamientos; $\bar{Y}_{\bullet\bullet}$ media muestral de la $N = ab$ observaciones; $\bar{Y}_{i\bullet}$ es la media muestral de las $n_{i\bullet} = b$ observaciones en el tratamiento i; $\bar{Y}_{\bullet j}$ es la media muestral de las $n_{\bullet j} = a$ observaciones en el bloque j;
Respuesta ajustada $\hat{Y}_{ij} = \bar{Y}_{i\bullet} + \bar{Y}_{\bullet j} - \bar{Y}_{\bullet\bullet}$</td></tr>
</table>

RESUMEN COMPARATIVO DE MODELOS ANOVA CON UN FACTOR DE ESTUDIO DE EFECTOS FIJOS EN ESTRUCTURAS DE DISEÑO DCA VS. DBCA		
	DCA (Diseño completamente aleatorizado)	**DBCA (Diseño en bloques completos aleatorizados con tamaños de bloques igual a a)**
Test ANOVA	Prueba si hay efectos significativos del factor sobre la media de la respuesta: $H_0: \mu_1 = \mu_2 = \cdots, \mu_a$ **vs.** $H_1: \mu_i \neq \mu_j$ **para al menos un par de niveles de tratamientos,** o equivalentemente, $H_0: \alpha_1 = \alpha_2 = \cdots, \alpha_a = 0$ **vs.** $H_1: \alpha_i \neq 0$ **para al menos un nivel *i*.** • Bajo H_0 se dice que el factor no tiene efectos significativos sobre la media de la variable respuesta, en los niveles considerados. • Bajo H_1 se dice que el factor tiene efectos significativos sobre media de la variable respuesta, en los niveles considerados. Bajo H_0 y los supuestos del modelo, el estadístico de prueba es $F_0 = \frac{MSA}{MSE} \sim f_{a-1,N-a}$ y se rechaza H_0 si $P\left(f_{a-1,N-a} > F_0\right)$ es pequeño.	Prueba si hay efectos significativos del factor sobre la media de la respuesta: $H_0: \mu_{1\bullet} = \mu_{2\bullet} = \cdots, \mu_{a\bullet}$ **vs.** $H_1: \mu_{i\bullet} \neq \mu_{j\bullet}$ **para al menos un par de niveles de tratamientos,** o equivalentemente, $H_0: \alpha_1 = \alpha_2 = \cdots, \alpha_a = 0$ **vs.** $H_1: \alpha_i \neq 0$ **para al menos un nivel *i*.** • Bajo H_0 se dice que el factor de tratamientos no tiene efectos significativos sobre la media de la variable respuesta, en los niveles considerados. • Bajo H_1 se dice que el factor tiene efectos significativos sobre media de la variable respuesta, en los niveles considerados. Bajo H_0 y los supuestos del modelo, el estadístico de prueba es $F_0 = \frac{MSA}{MSE} \sim f_{(a-1),(a-1)(b-1)}$ y se rechaza H_0 si $P\left(f_{(a-1),(a-1)(b-1)} > F_0\right)$ es pequeño. ***No hay conclusión estadística acerca de la igualdad de medias de bloques***; la ***"seudo prueba" F*** sobre los bloques meramente se hace como una forma de establecer la utilidad de haber creado bloques en el experimento. *Con relación a los bloques se evalúa la eficiencia relativa del diseño en bloques vs. un DCA ¿Ayudó el bloqueo a disminuir el tamaño del error experimental al compararlo con un DCA con el mismo número de réplicas de tratamientos? O bien ¿Comparado con un DCA, el diseño en bloques reduce el número de U.E necesarias?* (Ver en notas de clase cálculo de la eficiencia relativa de los DBCA)
Inferencias y estimaciones de interés	***Si se rechaza H_0 en el test ANOVA:*** **Estimaciones de (ver en notas de clase distribuciones e I.C):** • Medias de tratamiento: $\hat{\mu}_i = \bar{Y}_{i\bullet} = \frac{1}{n_i}\sum_{j=1}^{n_i} Y_{ij}$ • Media global: $\hat{\mu} = \bar{Y}_{\bullet\bullet}$ • Efectos de tratamientos: $\hat{\alpha}_i = \bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}$, o bien $\hat{\alpha}_i = \frac{a-1}{a}\bar{Y}_{i\bullet} - \frac{1}{a}\sum_{j\neq i}\bar{Y}_{j\bullet}$ (esta última fórmula sólo cuando el diseño es balanceado) • Varianza, $\hat{\sigma}^2 = MSE$ (tiene $N-a$ grados de libertad) **Inferencias:** **Comparaciones de medias de tratamientos** • ***Comparaciones múltiples por pares de medias:*** Se realizan todas las comparaciones por pares de medias $H_0: \mu_i = \mu_j$ vs. $H_1: \mu_i \neq \mu_j$. Luego, se agrupan las medias en grupos tales que todas las medias incluidas en un grupo dado son estadísticamente iguales (*no se aplica igualdad por transitividad*). *Bajo validez de los supuestos sobre los errores del modelo*, la igualdad por pares de medias se realiza así: 1. Método LSD: Usando mínima diferencia significativa (LSD) a un nivel de significancia γ se rechaza H_0 si $\lvert\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}\rvert \geq t_{\gamma/2,N-a}\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$, o usando I.C de $(1-\gamma)100\%$ de confianza para $\mu_i - \mu_j$, se rechaza H_0 si cero no pertenece al intervalo $\left(\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}\right) \pm t_{\gamma/2,N-a} \times \sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$, donde $t_{\gamma/2,N-a}$ es el percentil crítico de nivel $\gamma/2$ en la distribución t-Student con $N-a$ grados de libertad. 2. Método Tukey: Usando diferencia "honestamente significativa" (HSD) a un nivel de significancia γ se rechaza H_0 si $\lvert\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}\rvert \geq \frac{1}{\sqrt{2}}\, q_\gamma(a, N-a)\sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$, o usando I.C de $(1-\gamma)100\%$ de confianza de Tukey para $\mu_i - \mu_j$, se rechaza H_0 si cero no pertenece al intervalo $\left(\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}\right) \pm \frac{1}{\sqrt{2}}\, q_\gamma(a, N-a) \times \sqrt{MSE\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}$, donde $q_\gamma(a, N-a)$ es el rango crítico estudentizado de nivel γ para a medias y con $N-a$ grados de libertad. 3. Otros métodos: Duncan, Dunnet (comparaciones con un nivel de control), ver notas de clase.	***Si se rechaza H_0 en el test ANOVA:*** **Estimaciones de (ver en notas de clase distribuciones e I.C):** • Medias de tratamiento: $\hat{\mu}_{i\bullet} = \bar{Y}_{i\bullet} = \frac{1}{b}\sum_{j=1}^{b} Y_{ij}$ • Media global: $\hat{\mu} = \bar{Y}_{\bullet\bullet}$ • Efectos de tratamientos: $\hat{\alpha}_i = \bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet}$, o bien $\hat{\alpha}_i = \frac{a-1}{a}\bar{Y}_{i\bullet} - \frac{1}{a}\sum_{j\neq i}\bar{Y}_{j\bullet}$ • Varianza, $\hat{\sigma}^2 = MSE$ (tiene $(a-1)(b-1)$ grados de libertad) **Inferencias:** **Comparaciones de medias de tratamientos (no se aplican a medias de bloques!)** • ***Comparaciones múltiples por pares de medias:*** Se realizan todas las comparaciones por pares de medias $H_0: \mu_{i\bullet} = \mu_{j\bullet}$ vs. $H_1: \mu_{i\bullet} \neq \mu_{j\bullet}$. Luego, se agrupan las medias en grupos tales que todas las medias incluidas en un grupo dado son estadísticamente iguales (*no se aplica igualdad por transitividad*). *Bajo validez de los supuestos sobre los errores del modelo*, la igualdad por pares de medias se realiza así: 1. Método LSD: Usando mínima diferencia significativa (LSD) a un nivel de significancia γ se rechaza H_0 si $\lvert\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}\rvert \geq t_{\gamma/2,(a-1)(b-1)}\sqrt{MSE \times \frac{2}{b}}$, o usando I.C de $(1-\gamma)100\%$ de confianza para $\mu_{i\bullet} - \mu_{j\bullet}$, se rechaza H_0 si cero no pertenece al intervalo $\left(\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}\right) \pm t_{\gamma/2,(a-1)(b-1)}\sqrt{MSE \times \frac{2}{b}}$, donde $t_{\frac{\gamma}{2},(a-1)(b-1)}$ es el percentil crítico de nivel $\gamma/2$ en la distribución t-Student con $(a-1)(b-1)$ grados de libertad. 2. Método Tukey: Usando diferencia "honestamente significativa" (HSD) a un nivel de significancia γ se rechaza H_0 si $\lvert\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}\rvert \geq \frac{1}{\sqrt{2}}\, q_\gamma\left(a,(a-1)(b-1)\right)\sqrt{MSE \times \frac{2}{b}} = q_\gamma\left(a,(a-1)(b-1)\right)\sqrt{\frac{MSE}{b}}$, o usando I.C de $(1-\gamma)100\%$ de confianza de Tukey para $\mu_i - \mu_j$, se rechaza H_0 si cero no pertenece al intervalo $\left(\bar{Y}_{i\bullet} - \bar{Y}_{j\bullet}\right) \pm q_\gamma\left(a,(a-1)(b-1)\right)\sqrt{\frac{MSE}{b}}$, donde $q_\gamma\left(a,(a-1)(b-1)\right)$ es el rango crítico estudentizado de nivel γ para a medias y con $(a-1)(b-1)$ grados de libertad. 3. Otros métodos: Duncan, Dunnet (comparaciones con un nivel de control), se ajustan también los grados de libertad del error a $(a-1)(b-1)$ y los tamaños de muestra por tratamiento iguales a b.

RESUMEN COMPARATIVO DE MODELOS ANOVA CON UN FACTOR DE ESTUDIO DE EFECTOS FIJOS EN ESTRUCTURAS DE DISEÑO DCA VS. DBCA		
	DCA (Diseño completamente aleatorizado)	**DBCA (Diseño en bloques completos aleatorizados con tamaños de bloques igual a a)**
	• ***Comparaciones por contrastes de medias de tratamientos μ_i:*** Un contraste de medias de tratamientos es una combinación lineal de medias así $W=\sum_{i=1}^{a} c_i\mu_i$, tal que $\sum_{i=1}^{a} c_i=0$. Puede definirse un contraste para comparar el promedio de las medias de un grupo de tratamientos contra el promedio de las medias de otro grupo de tratamientos, definiendo los pesos o ponderadores c_i de acuerdo a la conformación de los dos grupos de medias a contrastar. El estimador del contraste es $\widehat{W}=\sum_{i=1}^{a} c_i\bar{Y}_{i\bullet}$. Bajo validez de los supuestos sobre los errores del modelo, $\widehat{W}\sim N\left(W,\sigma^2\sum_{i=1}^{a}\frac{c_i^2}{n_i}\right)$. Las inferencias pueden ser mediante: 1. Tests de hipótesis sobre el contraste (ver notas de clase) 2. I.C para el contraste de $(1-\gamma)100\%$ de confianza $\widehat{W}\pm t_{\gamma/2,N-a}\sqrt{MSE\ \sum_{i=1}^{a}\frac{c_i^2}{n_i}}$	• ***Comparaciones por contrastes de medias de tratamientos $\mu_{i\bullet}$:*** Un contraste de medias de tratamientos es una combinación lineal de medias así $W=\sum_{i=1}^{a} c_i\mu_{i\bullet}$, tal que $\sum_{i=1}^{a} c_i=0$. Puede definirse un contraste para comparar el promedio de las medias de un grupo de tratamientos contra el promedio de las medias de otro grupo de tratamientos, definiendo los pesos o ponderadores c_i de acuerdo a la conformación de los dos grupos de medias a contrastar. El estimador del contraste es $\widehat{W}=\sum_{i=1}^{a} c_i\bar{Y}_{i\bullet}$. Bajo validez de los supuestos sobre los errores del modelo, $\widehat{W}\sim N\left(W,\sigma^2\sum_{i=1}^{a}\frac{c_i^2}{b}\right)$. Las inferencias pueden ser mediante: 1. Tests de hipótesis sobre el contraste (ver notas de clase) 2. I.C para el contraste de $(1-\gamma)100\%$ de confianza $\widehat{W}\pm t_{\gamma/2,(a-1)(b-1)}\sqrt{MSE\ \sum_{i=1}^{a}\frac{c_i^2}{b}}$
Validación de supuestos	Residuos de ajuste corresponden $\hat{E}_{ij}=Y_{ij}-\bar{Y}_{i\bullet}$ • Con su versión estandarizada se analizan gráficos de residuales (residuos vs. niveles del factor y vs. respuesta ajustada ¿media cero, varianza constante y no hay carencia de ajuste?), gráfico de probabilidad normal y tests de probabilidad normal. • *Tests formales de homocedasticidad en los errores del modelo $H_0\colon\sigma_1^2=\sigma_2^2=\cdots=\sigma_a^2$ vs. $H_1\colon\sigma_i^2\neq\sigma_{i'}^2$ para al menos un par de niveles de tratamiento i, i´ , donde $\sigma_i^2=Var(E_{ij})$. Estos tests suponen válido la independencia y la normalidad de los errores de ajuste.* • Si se conoce el orden de corrida, debe evaluarse primero el supuesto de independencia (se evalúan autocorrelaciones con ACF, test Ljung-Box y gráfico de residuos de ajuste vs. orden de corrida) con los residuos de ajuste en el orden de corrida, antes de aplicar tests de normalidad y tests para homocedasticidad.	Residuos de ajuste corresponden $\hat{E}_{ij}=Y_{ij}-\bar{Y}_{i\bullet}-\bar{Y}_{\bullet j}+\bar{Y}_{\bullet\bullet}$ • Con su versión estandarizada se analizan gráficos de residuales (residuos vs. niveles del factor de tratamientos, vs. bloques y vs. respuesta ajustada ¿media cero, varianza constante y no hay carencia de ajuste?), gráfico de probabilidad normal y tests de probabilidad normal. • *No pueden aplicarse tests formales de homocedasticidad pues estos suponen una sola vía de clasificación de los datos experimentales, lo cual no se cumple en un DBCA ni en experimentos factoriales!* • Si se conoce el orden de corrida, debe evaluarse primero el supuesto de independencia (se evalúan autocorrelaciones con ACF, test Ljung-Box y gráfico de residuos de ajuste vs. orden de corrida) con los residuos de ajuste en el orden de corrida, antes de aplicar tests de normalidad.