DISEÑO DE EXPERIMENTOS

Licenciatura en Estadística

Profesores:
Dr. José Alberto Pagura
Lic. Julia Inés Fernández

EXPERIMENTOS CON FACTORES ALEATORIOS

Conceptos introductorios

- En determinados problemas que se tratan mediante diseños experimentales, los niveles de los factores se eligen al azar de un población de un gran número de niveles. Por ejemplo se desea conocer el efecto de los operarios sobre el tiempo en realizar una determinada tarea, o en un proceso en el que un producto es elaborado por varias máquinas y se quiere investigar si hay efecto de las diferentes máquinas. En tales situaciones, los factores estudiados se llaman "aleatorios" y las inferencias se desean realizar sobre la población de niveles representados por los ensayados.

Un factor aleatorio-El modelo

- $y_{ij} = \mu + \tau_i + \varepsilon_{ij}\ ; i = 1 \ldots a, j = 1 \ldots n$
- τ_i y ε_{ij} son variables aleatorias independientes con variancias σ_τ^2 y σ^2 respectivamente.
- La variancia de cualquier observación es $V\left(y_{ij}\right) = \sigma_\tau^2 + \sigma^2$
- σ_τ^2 y σ^2 se llaman componentes de variancia y el modelo presentado se conoce como "de efectos aleatorios" o "de componentes de variancia"

Hipótesis a probar

- Si no hay efecto del factor, $\sigma_\tau^2 = 0$
- Entonces. la hipótesis a probar es:
- $H_0\colon\ \sigma_\tau^2 = 0$ vs. $H_1\colon\ \sigma_\tau^2 > 0$
- El no rechazo de H_0 corresponde a la situación en la que el factor que se estudia no tiene efecto sobre la respuesta.

Supuestos necesarios para la prueba de hipótesis

- Para probar la hipótesis de interés se requiere del cumplimiento de los siguientes supuestos:
- $\varepsilon_{ij} \sim N(0; \sigma^2)$ independientes
- $\tau_i \sim N(0; \sigma_\tau^2)$ independientes
- τ_i y ε_{ij} independientes
- La hipótesis nula se prueba mediante el análisis de la variancia a partir de la partición de la suma de cuadrados total como: $SC_{total} = SC_{factor} + SC_{error}$

El análisis de la variancia

FV	gl	SC	CM	$E(CM)$	
Factor	$a-1$	$SC_{factor}=\sum_{i=1}^{a} n(\bar{y}_{i.}-\bar{y}_{..})^2$	$CM_{factor}=\frac{SC_{factor}}{a-1}$	$E(CM_{factor}) = n\sigma_\tau^2+\sigma^2$	$F_{obs}=\frac{CM_{factor}}{CM_{error}}$
Error	$a(n-1)$	$SC_{error}=\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{i.})^2$	$CM_{error}=\frac{SC_{error}}{a(n-1)}$	$E(CM_{error})=\sigma^2$	
Total	$an-1$	$SC_{total}=\sum_{i=1}^{a}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{..})^2$			

Se tiene que $\frac{SC_{error}}{\sigma^2} \sim \chi^2_{a(n-1)}$

Si H_0 es cierta: $\frac{SC_{factor}}{\sigma^2} \sim \chi^2_{a-1}$; entonces, si H_0 es cierta:

$\frac{CM_{factor}}{CM_{error}} \sim F_{a-1,a(n-1)}$

Estimación de componentes de variancia

- En estos problemas, además de probar si el factor tienen algún efecto sobre la respuesta, interesará la estimación de las componentes de variancia.
- Teniendo en cuenta el análisis de la variancia:
- $\hat{\sigma}^2 = CM_{error}$
- $\hat{\sigma}_\tau^2 = \frac{CM_{factor} - CM_{error}}{n}$
- Si el experimento no es balanceado, llamando con n_i al número de réplicas de cada tratamiento, en lugar de n, se utilizará:
- $n_0 = \frac{1}{a-1}\left(\sum_{i=1}^{n} n_i - \frac{\sum_{i=1}^{a} n_i^2}{\sum_{i=1}^{a} n_i}\right)$

Estimación de los componentes de variancia

- Ocasionalmente puede resultar $CM_{factor} - CM_{error} < 0$, es decir $\hat{\sigma}_{\tau}^{2} < 0$. Por definición nunca una componente de variancia puede ser negativa. Una interpretación de este resultado es que $\hat{\sigma}_{\tau}^{2} = 0$.
- Existen también alternativas que proporcionan resultados siempre mayores que 0 (no serán tratadas en el curso)

Intervalos de confianza para los componentes de variancia

- Si la variable en estudio sigue una distribución normal resulta:

$$\frac{(N-a)CM_{error}}{\sigma^2} \sim \chi^2_{N-a}$$

- Entonces $P\left[\chi^2_{\alpha/2;N-a} \leq \frac{(N-a)CM_{error}}{\sigma^2} \leq \chi^2_{1-\alpha/2;N-a}\right] = 1-\alpha$
- Un intervalo de confianza de $100(1-\alpha)$ para σ^2 se puede escribir como:

$$\frac{(N-a)CM_{error}}{\chi^2_{1-\alpha/2;N-a}} \leq \sigma^2 \leq \frac{(N-a)CM_{error}}{\chi^2_{\alpha/2;N-a}}$$

Intervalos de confianza para los componentes de variancia

- El estimador $\hat{\sigma}_{\tau}^{2} = \frac{CM_{factor} - CM_{error}}{n}$ es una combinación lineal de variables aleatorias que siguen una distribución χ^2 y en consecuencia, no se puede obtener una expresión para su distribución.
- $\hat{\sigma}_{\tau}^{2}$ se construye a partir de:
- $\frac{(a-1)CM_{factor}}{n\sigma_{\tau}^{2}+\sigma^{2}} \sim \chi_{a-1}^{2}$ y $\frac{(N-a)CM_{error}}{\chi_{N-a}^{2}} \sim \chi_{N-a}^{2}$
- Con coeficientes $\frac{n\sigma_{\tau}^{2}+\sigma^{2}}{n(a-1)}$ y $\frac{\sigma^{2}}{n(N-a)}$

Intervalos de confianza para los componentes de variancia

- Si, en cambio, puede encontrarse un intervalo de confianza para $\frac{\sigma_\tau^2}{\sigma_\tau^2+\sigma^2}$ que representa la proporción de la variancia debida a la variabilidad introducida por el factor.
- Teniendo en cuenta que CM_{factor} y CM_{error} son variables aleatorias independientes y que:
- $\frac{CM_{factor}/\left(n\sigma_\tau^2+\sigma^2\right)}{CM_{error}/\sigma^2} \sim F_{a-1;N-a}$

Intervalos de confianza para los componentes de variancia

- $P\left(F_{\alpha/2;a-1;N-a} \leq \frac{CM_{factor}}{CM_{error}} \frac{\sigma^2}{(n\sigma_\tau^2+\sigma^2)} \leq F_{1-\alpha/2;a-1;N-a}\right) = 1-\alpha$
- Trabajando la expresión anterior se puede escribir:
- $P\left(L \leq \frac{\sigma_\tau^2}{\sigma^2} \leq U\right) = 1-\alpha.$ Donde:
- $L = \frac{1}{n}\left(\frac{CM_{factor}}{CM_{error}} \frac{1}{F_{1-\alpha/2;a-1;N-a}} - 1\right)$ y $U = \frac{1}{n}\left(\frac{CM_{factor}}{CM_{error}} \frac{1}{F_{\alpha/2;a-1;N-a}} - 1\right)$
- Mediante operaciones matemáticas sencillas pueden deducirse los límites inferior y superior del $100(1-\alpha)$ de confianza:
- $\frac{L}{1+L} \leq \frac{\sigma_\tau^2}{\sigma+\sigma_\tau^2} \leq \frac{U}{1+U}$

Ejemplo 1

- Una compañía textil fabrica un tejido en un gran número de telares. Se desea que el producto de los telares fuera homogéneo, pero se sospecha que además de la variación natural de la resistencia de la tela del tejido del mismo telar, hay variaciones entre la resistencia de un telar y otro. Para estudiar si es ese el comportamiento de la variable resistencia de la tela, se extraen al azar 4 telares y se realizan 4 ensayos sobre muestras de tela producidas por cada telar.

Resultados

Telar	Observaciones				$y_{i.}$
	1	2	3	4	
1	98	97	99	96	390
2	91	90	93	92	366
3	96	95	97	95	383
4	95	96	99	98	388

Ejemplo 2

- Se seleccionan 8 coladas de barras de acero para la construcción de un determinado diámetro y se somete a una prueba de tensión-deformación a 6 barras de cada colada observando una medida de la resistencia. Se desea determinar si hay efecto de la colada y estimar la variancia de la variable que se estudia. En estas pruebas algunas de esas barras fallan y por lo tanto, no se tienen 6 datos para cada colada.
- La información obtenida se encuentra en la diapositiva siguiente:

Resultados del experimento

COLADA							
1	2	3	4	5	6	7	8
698	671	686	715	679	686	667	686
693	680	692	700	694	699	660	684
700	683	693	*	681	693	666	693
690	666	692	*	689	703	*	675
703	*	696	*	*	702	*	*
708	*	*	*	*	709	*	*

Dos factores aleatorios

- Se tienen dos factores A y B con un gran número de niveles (se suponen ∞) y se ensayarán a niveles de A, b niveles de B en un diseño factorial equilibrado, realizando n réplicas.

Dos factores aleatorios - El modelo estadístico

- El modelo estadístico que se plantea, es:
- $y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}\ ; i = 1 \ldots a, j = 1 \ldots b, k = 1 \ldots n$
- $\tau_i \sim N(0; \sigma_\tau^2)$, independientes ; $\beta_j \sim N\left(0; \sigma_\beta^2\right)$, independientes
- $\tau\beta_{ij} \sim N\left(0; \sigma_{\tau\beta}^2\right)$, independientes ; $\varepsilon_{ijk} \sim N(0; \sigma^2)$, independientes
- E independientes entre ellos

- La varianza de una observación será:

- $V\left(y_{ijk}\right) = \sigma_\tau^2 + \sigma_\beta^2 + \sigma_{\tau\beta}^2 + \sigma^2$

Dos factores aleatorios - El análisis de la varianza

- La SC_{total} se particiona como se ha visto en el modelo a efectos fijos, pero las esperanzas de estos cuadrados medios resultan:
- $E(CM_A) = \sigma^2 + n\sigma_{\tau\beta}^2 + bn\sigma_{\tau}^2$
- $E(CM_B) = \sigma^2 + n\sigma_{\tau\beta}^2 + an\sigma_{\beta}^2$
- $E(CM_{AB}) = \sigma^2 + n\sigma_{\tau\beta}^2$
- $E(CM_{error}) = \sigma^2$

Dos factores aleatorio- El análisis de la variancia

- Las estadísticas de prueba puede deducirse, asumiendo que las distribuciones de las sumas de cuadrados sobre sus esperanzas siguen distribuciones Chi-2 y que son independientes.
- Los siguientes cocientes siguen distribuciones F con los correspondientes grados de libertad:

$$F_0 = \frac{CM_{AB}}{CM_{error}}; F_0 = \frac{CM_A}{CM_{AB}}; F_0 = \frac{CM_B}{CM_{AB}}$$

y permiten evaluar la significación estadística de los efectos

Dos factores aleatorio- El análisis de la variancia

FV	gl	SC	E(CM)	F
Factor A	$a-1$	SC_A	$\sigma^2 + n\sigma_{\tau\beta}^2 + bn\sigma_\tau^2$	CM_A/CM_{AB}
Factor B	$b-1$	SC_B	$\sigma^2 + n\sigma_{\tau\beta}^2 + an\sigma_\beta^2$	CM_B/CM_{AB}
Interacción AB	$(a-1)(b-1)$	SC_{AB}	$\sigma^2 + n\sigma_{\tau\beta}^2$	$CM_{AB}/CM_{residual}$
Residual	$ab(n-1)$	$SC_{residual}$	σ^2	
Total	$abn-1$	SC_{total}		

Dos factores aleatorios – Estimación de las componentes de variancia

- Observando las esperanzas de los cuadrados medios pueden plantearse los siguientes estimadores de las componentes de variancia:
- $\hat{\sigma}^2 = CM_{error}$
- $\hat{\sigma}^2_{\tau\beta} = {(CM_{AB} - CM_{error})}/{n}$
- $\hat{\sigma}^2_{\beta} = {(CM_B - CM_{AB})}/{an}$
- $\hat{\sigma}^2_{\tau} = {(CM_A - CM_{AB})}/{bn}$

Modelo mixto con dos factores

- Se considera un factor A, fijo y un B, aleatorio.
- El modelo estadístico:
- $y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$
- $i = 1 \dots a, j = 1 \dots b, k = 1 \dots n$
- τ_i es un efecto fijo, $\sum_{i=1}^{a} \tau_i = 0$
- $\beta_j \sim N(0; \sigma_\tau^2)$

Modelo mixto con dos factores

- $(\tau\beta)_{ij}$ es una variable aleatoria normal con media 0 pero su variancia es $\frac{a-1}{a}\sigma^2_{\tau\beta}$
- Además: $\sum_{i=1}^{a}(\tau\beta)_{ij} = 0\,, j = 1\ldots b$
- $\mathrm{Cov}\left((\tau\beta)_{ij};(\tau\beta)_{i'j'}\right) = \begin{cases} -\frac{1}{a}\sigma^2_{\tau\beta}\; si\; i \neq i', j = j' \\ 0\; si\; j \neq j', i = i' \end{cases}$
- $\varepsilon_{ijk} \sim NID(0;\sigma^2)$

Modelo mixto con dos factores

- La partición de la suma de cuadrados total así como los grados de libertad son iguales al caso de efectos fijos. Las esperanzas de los cuadrados medios son:
- $E(CM_A) = \sigma^2 + n\sigma_{\tau\beta}^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$
- $E(CM_B) = \sigma^2 + an\sigma_{\beta}^2$
- $E(CM_{AB}) = \sigma^2 + n\sigma_{\tau\beta}^2$
- $E(CM_{Error}) = \sigma^2$

Modelo mixto con dos factores

- Las estadísticas para probar las hipótesis se construyen observando las esperanzas de los cuadrados medios, resultando:
- H_0: $\tau_i = 0$ se prueba con $F_0 = \frac{CM_A}{CM_{AB}} \sim F_{a-1;(a-1)(b-1)}$
- H_0: $\sigma_\beta^2 = 0$ se prueba con $F_0 = \frac{CM_B}{CM_{Error}} \sim F_{b-1;ab(n-1)}$
- H_0: $\sigma_{\tau\beta}^2 = 0$ se prueba con $F_0 = \frac{CM_{AB}}{CM_{Error}} \sim F_{(a-1)(b-1);ab(n-1)}$

Dos factores aleatorio- El análisis de la variancia

Siendo A el factor fijo y B el factor aleatoria, el ANOVA resulta:

FV	gl	SC	E(CM)	F
Factor A	$a-1$	SC_A	$\sigma^2 + n\sigma_{\tau\beta}^2 + \frac{bn\sum_{i=1}^{a}\tau_i^2}{a-1}$	CM_A/CM_{AB}
Factor B	$b-1$	SC_B	$\sigma^2 + an\sigma_{\beta}^2$	$CM_B/CM_{residual}$
Interacción AB	$(a-1)(b-1)$	SC_{AB}	$\sigma^2 + n\sigma_{\tau\beta}^2$	$CM_{AB}/CM_{residual}$
Residual	$ab(n-1)$	$SC_{residual}$	σ^2	
Total	$abn-1$	SC_{total}		

Modelo mixto con dos factores

- Los efectos fijos se estiman de la forma ya vista. El estimador del error a emplear en la estimación por intervalo es el cuadrado medio utilizado en el denominador de la prueba de hipótesis.
- Las componentes de variancia se estiman a partir de considerar las esperanzas de los cuadrados medios

Reglas para los cuadrados medios esperados

- Estas reglas se aplican a los diseños factoriales balanceados, cruzados o anidados
- Se ejemplifican para el modelo mixto con dos factores:

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$$

- A es fijo con a niveles, $i = 1 \dots a$
- B es aleatorio con b niveles, $j = 1 \dots b$
- Número de réplicas n, $k = 1 \dots n$

Reglas para los cuadrados medios esperados

1. El término del error se escribe como $\varepsilon_{(ij\ldots)m}$, en el ejemplo es $\varepsilon_{(ij)k}$
2. Además de la media general y el término de error, el modelo contiene todos los efectos que el experimentador supone que existen. Si en un término, el subíndice asociado a algún efecto aparece entre paréntesis, no hay interacción entre ese factor y los otros que aparecen en el término

Reglas para los cuadrados medios esperados

3. Para cada término del modelo, los subíndices se dividen en tres clases:

   - activos: aquellos que están presentes en el término y no están entre paréntesis

   - inactivos: están entre paréntesis en el término considerado

   - ausentes: están en el modelo pero no en el término considerado

Reglas para los cuadrados medios esperados

4. El número de grados de libertad para cualquier término del modelo es el producto de:

   -el número de niveles menos 1, de los índices activos.

   - el número de niveles de los índices inactivos.

Reglas para los cuadrados medios esperados

5. Cada término del modelo tiene asociado una componente de variancia (efecto aleatorio) o bien un factor fijo (efecto fijo).

En el ejemplo, el término que corresponde a A es: $\frac{\sum_{i=1}^{a} \tau_i^2}{a-1}$

Mientras que al efecto de B le corresponde: σ_β^2

Al efecto AB le corresponde $\sigma_{\tau\beta}^2$ ya que si una interacción tiene al menos una componente aleatoria, el efecto de la interacción es aleatorio

Reglas para los cuadrados medios esperados

6. Se elabora una tabla donde cada fila corresponde a un término y cada columna a un subíndice. Para cada subíndice se indica el número de niveles y la naturaleza del factor (fijo (F) o aleatorio (R))

   a) En cada renglón se escribe 1 si uno de los subíndices inactivos en el componente del renglón coincide con el subíndice de la columna

Reglas para los cuadrados medios esperados

b) En cada renglón, si cualquiera de los subíndices del componente coincide con el subíndice de la columna, se escribe 0 si se trata de un factor fijo, y 1 si el factor es aleatorio

c) En las posiciones que quedan vacías se escribe el número de niveles correspondiente al factor

Reglas para los cuadrados medios esperados

d) Los cuadrados medios esperados para cada término se obtienen de la siguiente forma:

- se ocultan las columnas que corresponden a los subíndices activos en el término
- se observa cada renglón que contiene al menos los subíndices del término en consideración.
- se toma el producto de los números visibles

Reglas para los cuadrados medios esperados

- se multiplica dicho producto por la componente aleatoria o fija según corresponda.
- La suma de estas cantidades es el cuadrado medio esperado.