ANOVA: Interpretación de los resultados

Jhon Jairo Padilla Aguilar, PhD.

Interpretación de los resultados

- Después de realizar el experimento, debe hacerse un análisis estadístico y sacar conclusiones prácticas acerca del problema bajo estudio.
- En muchas ocasiones, esto puede hacerse mediante el análisis de los gráficos (diagramas de caja, diagramas de dispersión)
- Pero en algunos casos es necesario aplicar técnicas más formales

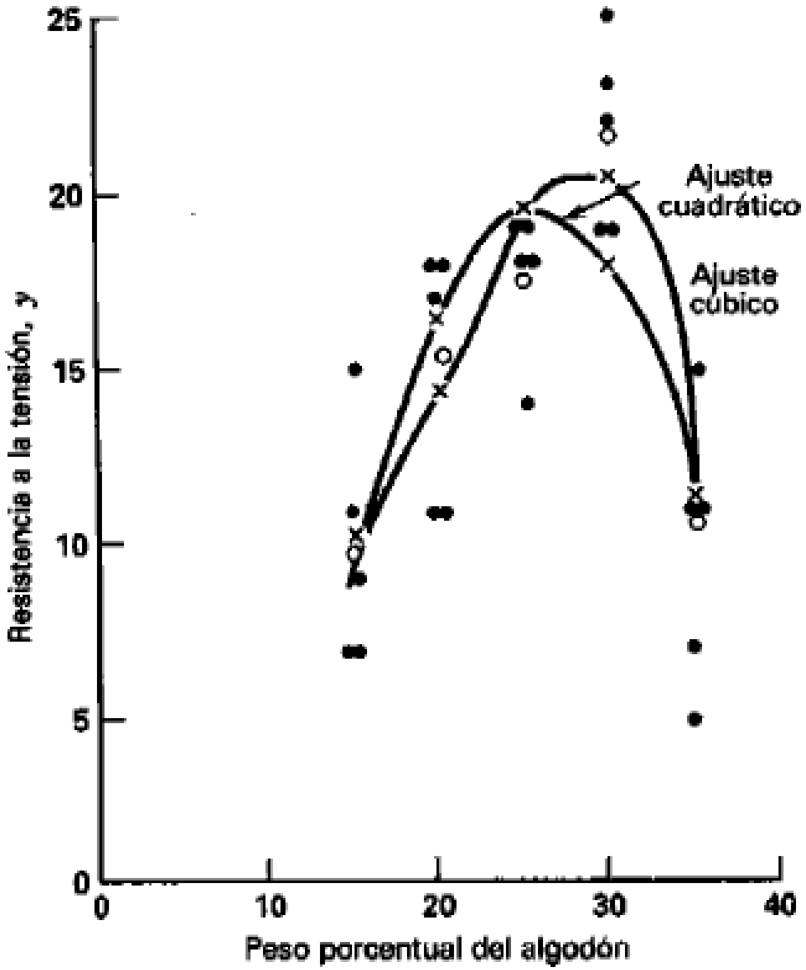
Tipos de factores

- Los factores que intervienen en un experimento pueden ser:
  - Cualitativos: Los niveles no pueden ordenarse por magnitud (Ej: operadores, cambios de turno, lotes de materia prima, etc)
  - Cuantitivos: Aquellos cuyos niveles pueden asociarse con puntos en una escala numérica (Ej: Temperatura, presión, tiempo, etc)
- Ambos tipos de factores se tratan de manera idéntica en el Diseño de Experimentos.

1. Análisis de Regresión

- Con factores cualitativos no hay valores intermedios. Son valores discretos y sólo se pueden sacar conclusiones sobre estos valores en particular.
- Con factores cualitativos, podrían darse valores numéricos intermedios entre dos niveles estudiados (Ej: Se han estudiado los niveles 2.0 y 3.0, pero se quisiera tratar de predecir qué ocurriría con un nivel 2.5)
- En este caso se puede usar una interpolación entre puntos consecutivos. Esto es un modelo empírico del proceso estudiado.
- Al enfoque general para ajustar modelos empíricos se le llama **análisis de regresión**.

1. Análisis de regresión



$$\hat{y} = -39.9886 + 4.596x - 0.0886x^2$$

$$\hat{y} = 62.6114 - 9.0114x + 0.4814x^2 - 0.0076x^3$$

1. Análisis de regresión- Recomendaciones

- En general, sería preferible hacer el ajuste con el polinomio de orden menor que describa adecuadamente el sistema o proceso. **Esto simplifica el modelo!!!!**
- Para el ejemplo, el ajuste del modelo cúbico es mejor que el del modelo cuadrático, por lo que se justifica su uso.
- Otro uso del modelo empírico es para la **optimización del proceso**, es decir, para encontrar los niveles de las variables del diseño que dan como resultado los mejores valores de la respuesta.

2.Comparaciones gráficas de las medias

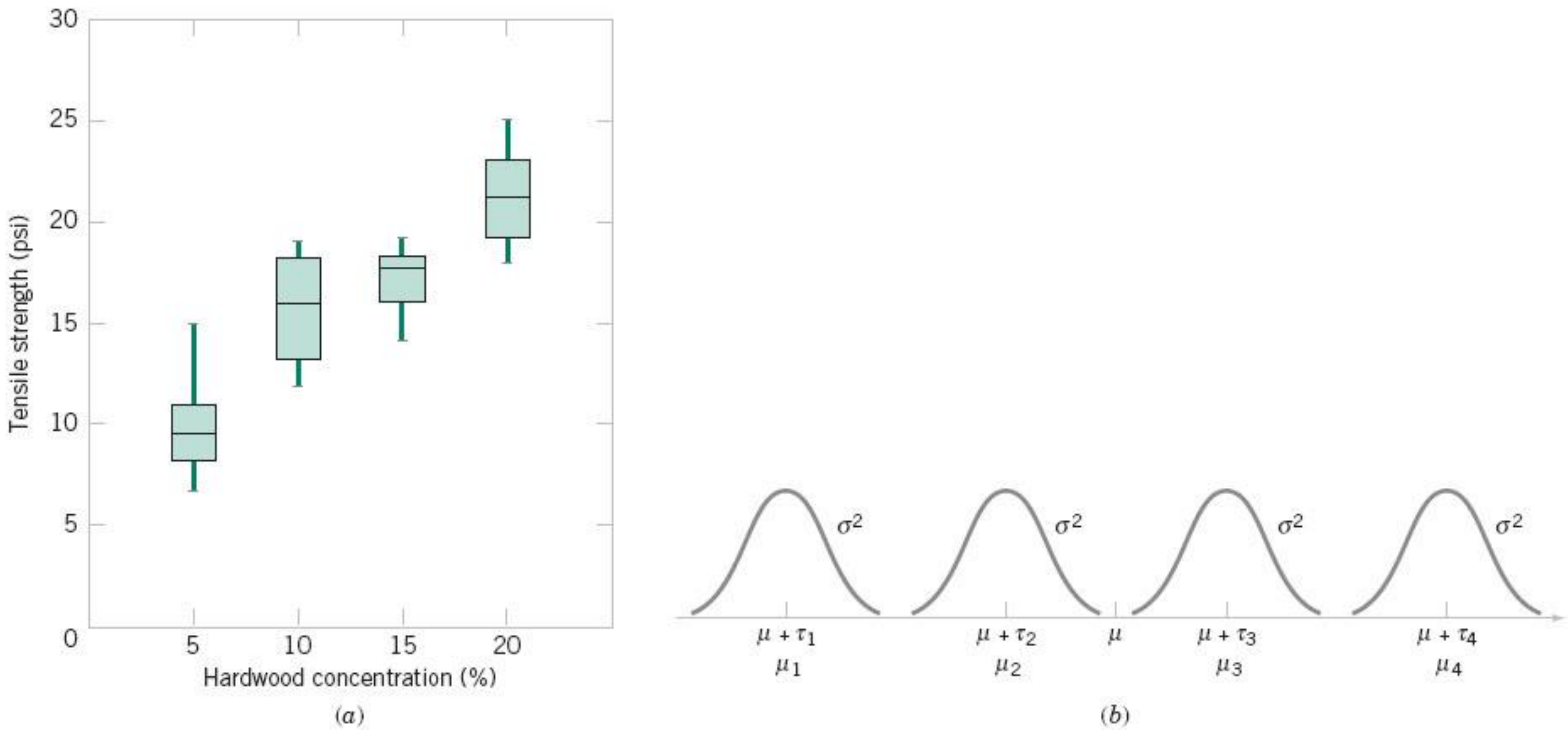


Figure 13-1 (a) Box plots of hardwood concentration data. (b) Display of the model in Equation 13-1 for the completely randomized single-factor experiment.

2. Comparaciones gráficas de medias

- Suponga que todos los promedios de los tratamientos tienen desviación estándar $\sigma/\sqrt{n}$
- Si todos los promedios de los tratamientos son iguales, las medias observadas, se comportarían como si fueran un conjunto de observaciones tomadas al azar de una distribución con media μ y varianza $\sigma/\sqrt{n}$

Comparación gráfica de las medias

- Los valores que no parecen haber sido sacados de la distribución (puntos 1 y 4) se asocian con tratamientos que producen respuestas medias diferentes

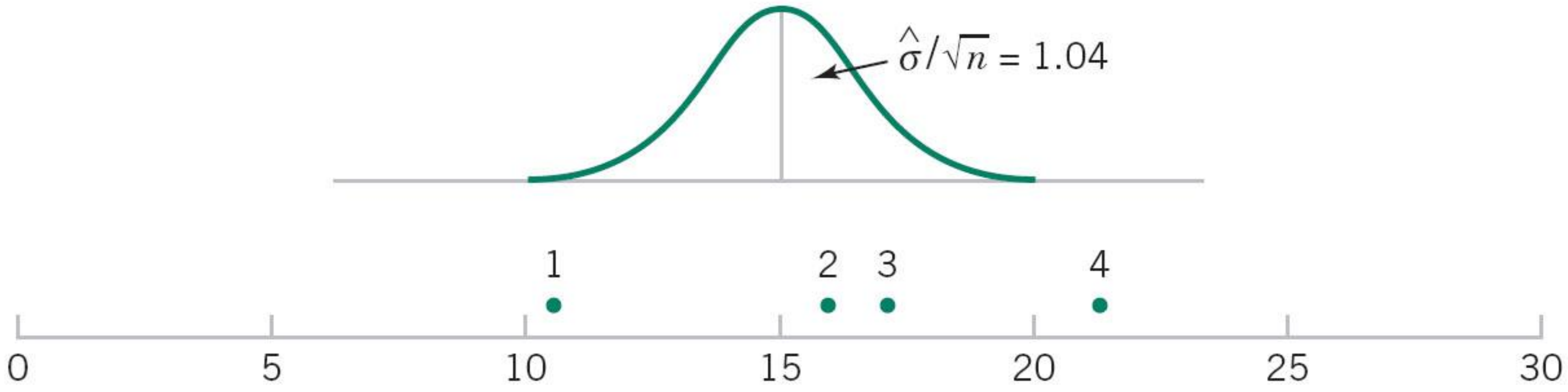


Figure 13-3 Tensile strength averages from the hardwood concentration experiment in relation to a normal distribution with standard deviation $\sqrt{MS_E/n} = \sqrt{6.51/6} = 1.04$.

Comparación gráfica de las medias

- El problema con la prueba anterior es que la varianza no se conoce.
- La varianza se estima con el MSE
- Entonces debe usarse una distribución t en lugar de la normal para hacer la gráfica.
- Los valores de 20 y 25 son los únicos que se adaptan a la distribución t. Los demás no podrían tomarse como medias generadas por esa distribución

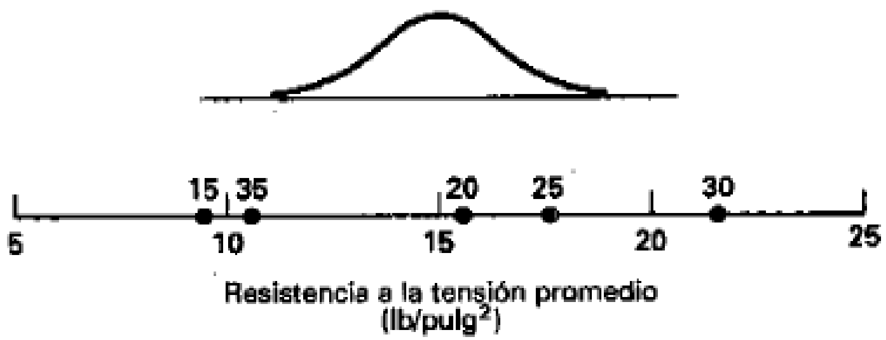


Figura 3-11 Promedio de la resistencia a la tensión del experimento del peso porcentual del algodón en relación con una distribución t con un factor de escalación $\sqrt{MS_E / n} = \sqrt{8.06 / 5} = 1.27$.

Comparación con Contrastes

- En el ejemplo de la fibra de algodón, la hipótesis nula de que las medias eran iguales fue rechazada
- Se sabe que algunas medias son diferentes de la mayoría, pero ¿Cuáles son las que causan esa diferencia?
- En principio, por las gráficas, los puntos 4 y 5 son los que producen medias similares.
- Por tanto, debería probarse la hipótesis

$$H_0: \mu_4 = \mu_5$$
$$H_1: \mu_4 \neq \mu_5$$

Comparación con contrastes

- La hipótesis nula podría probarse usando también la combinación lineal

$$H_0: \mu_4 - \mu_5 = 0$$
$$H_1: \mu_4 - \mu_5 \neq 0$$

- Si por ejemplo, se hubiese sospechado que los puntos 1 y 2 no diferían de los puntos 4 y 5 se podría probar la hipótesis con la combinación lineal

$$H_0: \mu_1 + \mu_2 = \mu_4 + \mu_5$$
$$H_1: \mu_1 + \mu_2 \neq \mu_4 + \mu_5$$

- ó

$$H_0: \mu_1 + \mu_2 - \mu_4 - \mu_5 = 0$$
$$H_1: \mu_1 + \mu_2 - \mu_4 - \mu_5 = 0$$

Comparación con contrastes

- En general, la comparación de las medias de los tratamientos de interés, implicará una combinación lineal de los totales de los tratamientos, denominada **contraste**, como

$$\Gamma = \sum_{i=1}^{a} c_i \mu_i$$

- Donde las constantes de los contrastes suman cero, es decir,

$$\Sigma_{i=1}^{a} c_i = 0.$$

Comparación con contrastes

- Las hipótesis anteriores pueden re-escribirse en términos de contrastes:

$$H_0: \sum_{i=1}^{a} c_i \mu_i = 0$$

$$H_1: \sum_{i=1}^{a} c_i \mu_i \neq 0$$

Pruebas de hipótesis con Contrastes

- Pueden hacerse de dos maneras:
  - Con la prueba t
  - Con la prueba F

Prueba t para los contrastes

- El contraste se escribe en términos de los totales de los tratamientos

$$C = \sum_{i=1}^{a} c_i y_{i.}$$

- Y la varianza de C es $V(C) = n\sigma^2 \sum_{i=1}^{a} c_i^2$

- Cuando los tamaños de las muestras son iguales, el coeficiente

$$\frac{\sum_{i=1}^{a} c_i y_{i.}}{\sqrt{n\sigma^2 \sum_{i=1}^{a} c_i^2}}$$

- Tendrá una distribución normal estándar

Prueba t para los contrastes

- Sustituyendo la varianza desconocida por su estimación (MSE), se utilizará el estadístico

$$t_0 = \frac{\sum_{i=1}^{a} c_i y_{i.}}{\sqrt{n MS_E \sum_{i=1}^{a} c_i^2}}$$

- La hipótesis nula se rechazaría si el valor absoluto de t_0 es mayor que $t_{\alpha/2,\, N-a}$

Prueba F para los contrastes

- Se calcula el índice F_o como

$$F_0 = t_0^2 = \frac{\left(\sum_{i=1}^{a} c_i y_{i.}\right)^2}{nMS_E \sum_{i=1}^{a} c_i^2}$$

- La hipótesis nula se rechazaría si $F_0 > F_{\alpha,1,N-a}$

Prueba F para los contrastes

- El índice F_0 se puede calcular también como

$$F_0 = \frac{MS_C}{MS_E} = \frac{SS_C\ /1}{MS_E}$$

- donde

$$SS_C = \frac{\left(\sum_{i=1}^{a} c_i\, y_{i.}\right)^2}{n\sum_{i=1}^{a} c_i^2}$$

Intervalo de confianza para un contraste

- En lugar de probar hipótesis acerca de un contraste, puede que sea más útil en algunos casos construir un intervalo de confianza.
- Suponga que el contraste de interés es

$$\Gamma = \sum_{i=1}^{a} c_i \mu_i$$

- Al sustituir las medias de los tratamientos con los promedios de los tratamientos se obtiene para tamaños de muestras iguales

$$C = \sum_{i=1}^{a} c_i \bar{y}_{i.} \qquad V(C) = \frac{\sigma^2}{n} \sum_{i=1}^{a} c_i^2$$

Intervalo de confianza para un contraste

- Si se usa MSE para estimar la varianza, el intervalo de confianza será

$$\sum_{i=1}^{a} c_i \bar{y}_{i.} - t_{\alpha/2,N-a}\sqrt{\frac{MS_E}{n}\sum_{i=1}^{a} c_i^2} \leq \sum_{i=1}^{a} c_i \mu_i \leq \sum_{i=1}^{a} c_i \bar{y}_{i.} + t_{\alpha/2,N-a}\sqrt{\frac{MS_E}{n}\sum_{i=1}^{a} c_i^2}$$

Contrastes ortogonales

- Dos contrastes con coeficientes c_i y d_i son ortogonales si para un diseño balanceado:

$$\sum_{i=1}^{a} c_i d_i = 0$$

- Y para un diseño desbalanceado

$$\sum_{i=1}^{a} n_i c_i d_i = 0$$

Utilidad de los contrastes

- Con *a* tratamientos, el conjunto de *a*-1 contrastes ortogonales compone la partición de la suma de cuadrados debida a los tratamientos en *a*-1 componentes independientes con un solo grado de libertad
- Por tanto, **las pruebas que se realizan en los contrastes ortogonales son independientes**.

Utilidad de los contrastes

- En general, el método de contrastes (o de contrastes ortogonales) es útil para lo que se llama **comparaciones pre-planeadas**.
- Los contrastes se especifican antes de llevar a cabo el experimento y de examinar los datos
- Esto se hace debido a que si las comparaciones se seleccionan después de examinar los datos, la mayoría de los experimentadores tendrían un sesgo y tenderían a construir pruebas que correspondan con las diferencias grandes observadas en las medias.

Utilidad de los contrastes

- Las diferencias grandes podrían ser el resultado de la presencia de efectos reales o podrían ser el resultado del error aleatorio.
- Si los experimentadores se inclinan consistentemente a escoger las diferencias más grandes para hacer las comparaciones, inflarán el error tipo I de la prueba, porque es probable que, en un porcentaje inusualmente elevado de las comparaciones seleccionadas, las diferencias observadas serán el resultado del error.
- Al examen de los datos para seleccionar las comparaciones de interés potencial se le llama **curioseo** o **sondeo de los datos**.

Cómo elegir los valores de los coeficientes?

- Existen varias maneras
- Algún elemento en la naturaleza del experimento deberá sugerir las comparaciones que son de interés
- Ejemplo:
  - a=3 tratamientos: tratamiento 1 es el control y los tratamientos 2 y 3 poseen niveles que son de interés para el experimentador.

Ejemplo (continuación)

- Los contrastes ortogonales apropiados podrían ser los siguientes:

Tratamiento	Coeficientes de los contrastes ortogonales	
1 (control)	–2	0
2 (nivel 1)	1	–1
3 (nivel 2)	1	1

- El contraste con los coeficientes -2, 1, 1 compara el efecto promedio del factor con el control
- El contraste con los coeficientes 0, -1, 1 compara los dos niveles del factor de interés

Ejemplo 3-6

- Considere los datos del ejemplo 3-1. Hay cinco medias de los tratamientos y cuatro grados de libertad entre estos tratamientos. Suponga que antes de correr el experimento se especificó la siguiente serie de comparaciones entre las medias de los tratamientos (y sus contrastes asociados):

Hipótesis	Contraste
$H_0{:}\mu_4 = \mu_5$	$C_1 = -y_{4.} + y_{5.}$
$H_0{:}\mu_1 + \mu_3 = \mu_4 + \mu_5$	$C_2 = y_{1.} + y_{3.} - y_{4.} - y_{5.}$
$H_0{:}\mu_1 = \mu_3$	$C_3 = y_{1.} - y_{3.}$
$H_0{:}4\mu_2 = \mu_1 + \mu_3 + \mu_4 + \mu_5$	$C_4 = -y_{1.} + 4y_{2.} - y_{3.} - y_{4.} - y_{5.}$

Resultados obtenidos

- Suponiendo que se siguió el orden aleatorio descrito anteriormente, se muestran las observaciones que se realizaron para la resistencia a la tensión en la siguiente tabla

Peso porcentual del algodón	Resistencia a la tensión observada ($lb/pulg^2$)					Totales y_i	Promedios $\bar{y}_i$
	1	2	3	4	5		
15	7	7	15	11	9	49	9.8
20	12	17	12	18	18	77	15.4
25	14	18	18	19	19	88	17.6
30	19	25	22	19	23	108	21.6
35	7	10	11	15	11	54	10.8
						$y_{..} = 376$	$\bar{y}_{..} = 15.04$

Ejemplo (continuación):

- Observe que los coeficientes de los contrastes son ortogonales.
- Utilizando los datos de la tabla, se encuentra que los valores numéricos de los contrastes y las sumas de los cuadrados son los siguientes:

$$C_1 = \qquad -1(108)+1(54) = -54 \qquad SS_{C_1} = \frac{(-54)^2}{5(2)} = 291.60$$

$$C_2 = +1(49) \qquad +1(88)-1(108)-1(54) = -25 \qquad SS_{C_2} = \frac{(-25)^2}{5(4)} = 31.25$$

$$C_3 = +1(49) \qquad -1(88) \qquad = -39 \qquad SS_{C_3} = \frac{(-39)^2}{5(2)} = 152.10$$

$$C_4 = -1(49)+4(77)-1(88)-1(108)-1(54) = 9 \qquad SS_{C_4} = \frac{(9)^2}{5(20)} = 0.81$$

Ejemplo (continuación):

Tabla 3-11 Análisis de varianza de los datos de la resistencia a la tensión

Fuente de variación	Suma de cuadrados	Grados de libertad	Cuadrado medio	F_0	Valor P
Peso porcentual del algodón	475.76	4	118.94	14.76	<0.001
contrastes ortogonales					
$C_1: \mu_4 = \mu_5$	(291.60)	1	291.60	36.18	<0.001
$C_2: \mu_1 + \mu_3 = \mu_4 + \mu_5$	(31.25)	1	31.25	3.88	0.06
$C_3: \mu_1 = \mu_3$	(152.10)	1	152.10	18.87	<0.001
$C_4: 4\mu_2 = \mu_1 + \mu_3 + \mu_4 + \mu_5$	(0.81)	1	0.81	0.10	0.76
Error	161.20	20	8.06		
Total	636.96	24			

- Observe que la suma de los cuadrados medios de los C_i es igual a la suma de los cuadrados de los tratamientos
- Los valores P muestran que hay diferencias significativas entre los niveles 4 y 5 y los niveles 1 y 3 del peso porcentual del algodón
- El promedio de los niveles 1 y 3 no difiere del promedio de los niveles 4 y 5 con un α=0,05

Método de Scheffé

- Se usa para comparar todos los contrastes
- Hay situaciones en que los experimentadores no pueden conocer de antemano cuáles son los contrastes que quieren comparar, o pueden tener interés en más de *a*-1 posibles comparaciones.
- Este método compara todos y cada uno de los contrastes posibles entre las medias de los tratamientos
- En este método el error de tipo I es a lo sumo α para cualquiera de las comparaciones posibles

Método de Scheffé

- Suponga que se ha determinado un conjunto de m contrastes

$$\Gamma_u = c_{1u}\mu_1 + c_{2u}\mu_2 + \cdots + c_{au}\mu_a \qquad u = 1, 2, \ldots, m$$

- El contraste correspondiente de los promedios de los tratamientos es

$$C_u = c_{1u}\bar{y}_{1.} + c_{2u}\bar{y}_{2.} + \cdots + c_{au}\bar{y}_{a.} \qquad u = 1, 2, \ldots, m$$

- Y el error estándar del contraste es

$$S_{C_u} = \sqrt{MS_E \sum_{i=1}^{a} (c_{iu}^2 / n_i)}$$

- Donde n_i es el número de observaciones en el tratamiento i-ésimo.
- Puede demostrarse que el valor crítico contra el que deberá compararse C_u es:

$$S_{\alpha,u} = S_{C_u}\sqrt{(a-1)F_{\alpha,a-1,N-a}}$$

El método de Scheffé

- Para probar la hipótesis de que el contraste Γ_u difiere de manera significativa de cero, se compara C_u con el valor crítico. Si el valor absoluto de C_u es mayor que el índice S, se rechaza la hipótesis de que el contraste Γ_u es igual a cero.
- El procedimiento de Scheffé puede usarse también para formar intervalos de confianza para todos los contrastes posibles entre las medias de los tratamientos.
- Los intervalos resultantes son intervalos de confianza simultáneos, por lo que la probabilidad de que todos ellos sean verdaderos simultáneamente es al menos $1-\alpha$.

Ejemplo

- Suponga los datos del ejemplo 3-1.
- Suponga que los contrastes de interés son:

$$\Gamma_1 = \mu_1 + \mu_3 - \mu_4 - \mu_5$$

$$\Gamma_2 = \mu_1 - \mu_4$$

- Los valores numéricos de estos contrastes son:

$$\begin{aligned} C_1 &= \bar{y}_{1.} + \bar{y}_{3.} - \bar{y}_{4.} - \bar{y}_{5.} \\ &= 9.80 + 17.60 - 21.60 - 10.80 \\ &= 5.00 \end{aligned}$$

$$\begin{aligned} C_2 &= \bar{y}_{1.} - \bar{y}_{4.} \\ &= 9.80 - 21.60 \\ &= -11.80 \end{aligned}$$

Ejemplo

- Los errores estándar son

$$S_{C_1} = \sqrt{MS_E \sum_{i=1}^{5} (c_{i1}^2 / n_i)} = \sqrt{8.06(1+1+1+1)/5} = 2.54$$

$$S_{C_2} = \sqrt{MS_E \sum_{i=1}^{5} (c_{i2}^2 / n_i)} = \sqrt{8.06(1+1)/5} = 1.80$$

- Los valores críticos son:

$$S_{0.01,1} = S_{C_1} \sqrt{(a-1)F_{0.01,a-1,N-a}} = 2.54\sqrt{4(4.43)} = 10.69$$

$$S_{0.01,2} = S_{C_2} \sqrt{(a-1)F_{0.01,a-1,N-a}} = 1.80\sqrt{4(4.43)} = 7.58$$

Ejemplo:

- Puesto que $|C_1| < S_{0.01,1}$, se concluye que el contraste Γ_1 es igual a cero; es decir, no existe evidencia sólida para concluir que los tratamientos 1 y 3 como grupo difieren de las medias de los tratamientos 4 y 5 como grupo.
- Como $|C_2| > S_{0.01,2}$, se concluye que el contraste Γ_2 no es igual a cero, es decir, las resistencias medias de los tratamientos 1 y 4 difieren significativamente.

Comparación de pares de medias de los tratamientos

- En muchas ocasiones, querrán compararse sólo pares de medias.
- Es posible determinar cuáles son las medias que difieren probando las diferencias entre todos los pares de medias de los tratamientos.
- El método de Scheffé puede aplicarse fácilmente a este problema, pero no es el procedimiento más sensible para tales comparaciones.
- Existen otros procedimientos para comparar medias por pares:
  - Prueba de Tukey
  - Método de la diferencia significativa mínima de Fisher
  - Prueba del rango múltiple de Duncan
  - Prueba de Newman-Keuls

Qué método de comparación por pares usar?

- No hay una respuesta precisa a esa pregunta
- Hay desacuerdos entre los especialistas en estadística
- Método de Diferencia significativa mínima: es eficaz para detectar diferencias reales en las medias si se aplica sólo después de que la prueba F en el análisis de varianza sea significativa en 5 %
- La prueba de rango múltiple de Duncan reporta buen desempeño. Es uno de los métodos más poderosos
- El método de Tukey efectúa un control sobre el índice de error global, por lo que es muy usado
- La potencia de la prueba de Newman-Keuls casi siempre es menor que la de la prueba del rango múltiple de Duncan.

Determinación del tamaño de la muestra

- Es un aspecto crítico del diseño de un experimento
- Se refiere a determinar el número de réplicas que deben correrse
- En general:
  - Para detectar efectos pequeños se requiere un alto número de réplicas
  - Para detectar efectos grandes se requieren menos réplicas
- Los métodos que se estudiarán con un solo factor sirven también para cuando se tienen varios factores

Método 1: Curvas de Operación característica

- La curva de operación característica es una gráfica de la probabilidad de error de tipo II de una prueba estadística.
- La elección del tamaño de la muestra y la probabilidad β de error de tipo II guardan una estrecha relación.
- Suponga que se están probando las hipótesis

$$H_0: \mu_1 = \mu_2$$
$$H_1: \mu_1 \neq \mu_2$$

- Y que las medias no son iguales, por lo que $\delta = \mu_1 - \mu_2$
- Puesto que H_o no es verdadera, la preocupación principal es cometer la equivocación de no rechazar H_o.

Curvas de operación característica

- La probabilidad de error de tipo II depende de la verdadera diferencia de las medias δ.
- A una gráfica de β contra δ para un tamaño particular de la muestra se le llama la **curva de operación característica** o **curva OC**.
- El error β también es una función del tamaño de la muestra.
- En general, para un error dado de δ, el error β se reduce cuando el tamaño de la muestra se incrementa. Por tanto, es más fácil detectar una diferencia especificada en las medias para tamaños grandes de la muestra que para tamaños pequeños.

Curvas de operación característica

- Curva de operación característica para las hipótesis $H_0: \mu_1 = \mu_2$
  $H_1: \mu_1 \neq \mu_2$
- Para el caso de que las dos varianzas poblacionales σ^2_1 y σ^2_2 son desconocidas pero iguales a σ^2 y para un nivel de significación de α=0,05
- También para iguales tamaños de muestra (n_1=n_2=n)
- El parámetro del eje horizontal es d:

$$d = \frac{|\mu_1 - \mu_2|}{2\sigma} = \frac{|\delta|}{2\sigma}$$

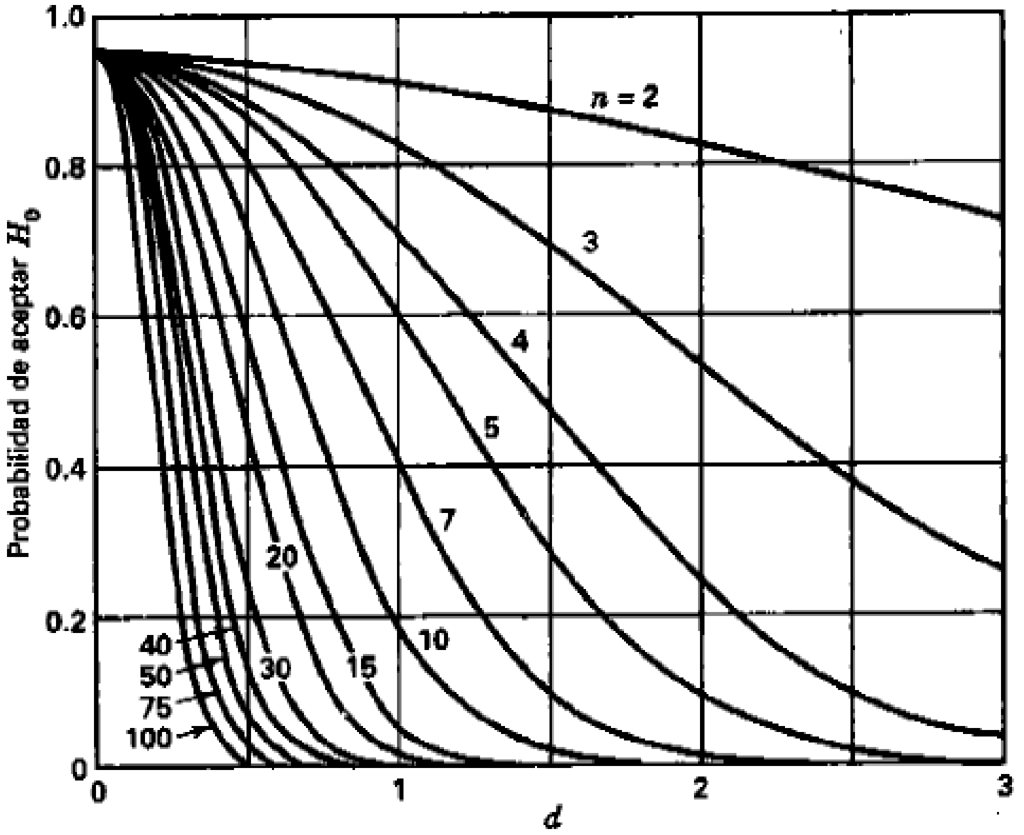


Figura 2-12 Curvas de operación característica para la prueba t de dos colas con α = 0.05. (Reproducida con permiso de "Operating Characteristics Curves for the Common Statistical Tests of Significance", C.L. Ferris, F.E. Grubbs y C.L. Weaver, *Annals of Mathematical Statistics*.)

Curvas de operación característica

- La división de d por 2σ permite al experimentador usar el mismo juego de curvas independientemente del valor de la varianza (la diferencia en las medias se expresa en unidades de desviación estándar)
- El tamaño de la muestra usado para construir las curvas es en realidad n*=2n-1

Curvas de operación característica

Observamos que:

- Entre más grande sea la diferencia en las medias, menor será la probabilidad del error de tipo II para un tamaño de muestra y un valor de α dados
- Para un tamaño de muestra y un α especificados, la prueba detectará con mayor facilidad las diferencias grandes que las pequeñas

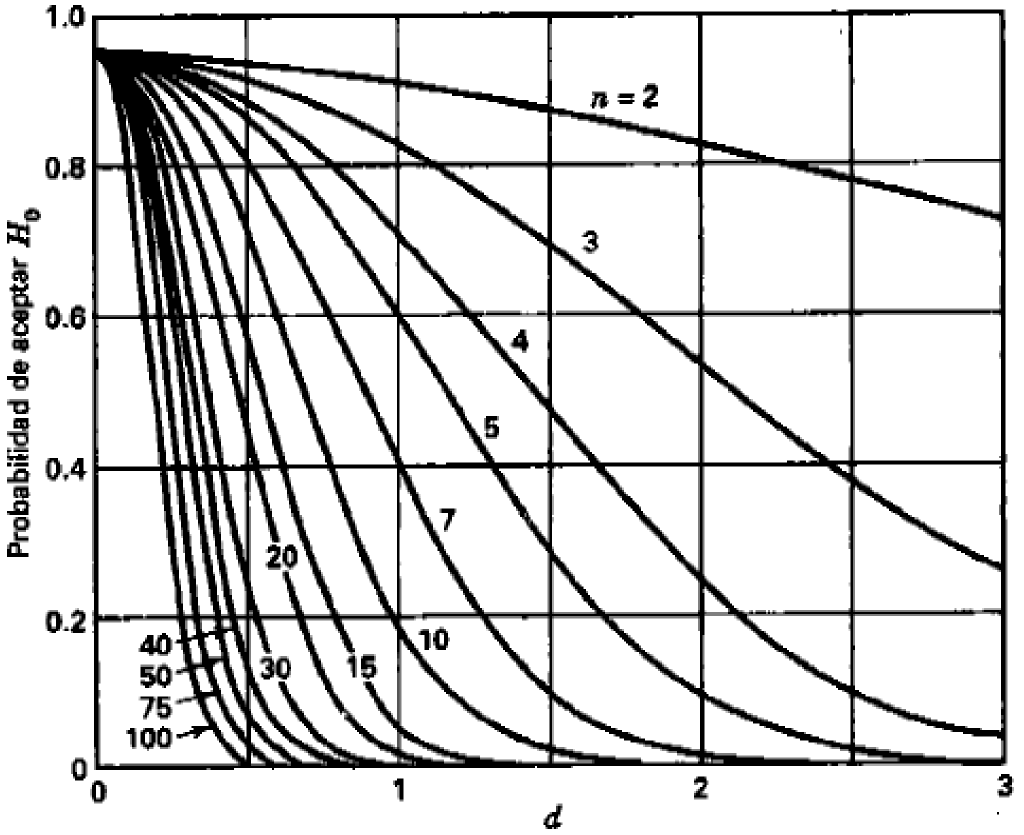


Figura 2-12 Curvas de operación característica para la prueba t de dos colas con $\alpha = 0.05$. (Reproducida con permiso de "Operating Characteristics Curves for the Common Statistical Tests of Significance", C.L. Ferris, F.E. Grubbs y C.L. Weaver, *Annals of Mathematical Statistics*.)

Curvas de operación característica

Observamos que:

- Cuando el tamaño de la muestra se hace más grande, la probabilidad de error de tipo II se hace más pequeña para una diferencia en las medias y un valor de a dados.
- Para detectar una diferencia δ especificada, puede aumentarse la potencia de la prueba incrementando el tamaño de la muestra.

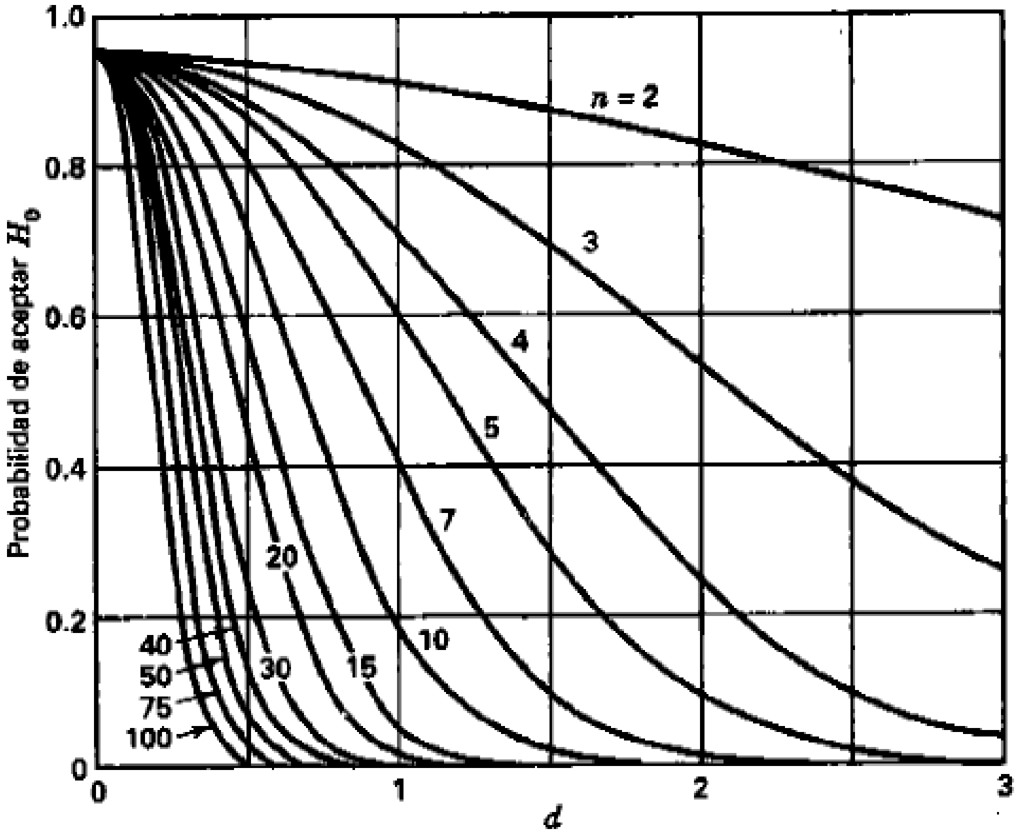


Figura 2-12 Curvas de operación característica para la prueba t de dos colas con $\alpha = 0.05$. (Reproducida con permiso de "Operating Characteristics Curves for the Common Statistical Tests of Significance", C.L. Ferris, F.E. Grubbs y C.L. Weaver, *Annals of Mathematical Statistics*.)

Cómo seleccionar el tamaño de la muestra con la OC?

- Considere el problema del cemento portland comentado antes
- Suponga que si las dos formulaciones difieren en la fuerza promedio hasta en 0.5 Kgf/cm^2, sería deseable detectarlo con una probabilidad alta
- Puesto que este valor es la diferencia crítica en las medias que quiere detectarse, se encuentra que d es

$$d = \frac{|\mu_1 - \mu_2|}{2\sigma} = \frac{0.5}{2\sigma} = \frac{0.25}{\sigma}$$

Cómo seleccionar el tamaño de la muestra con la OC?

- Suponga que con base en la experiencia previa, es altamente probable que la desviación estándar exceda 0.25 kgf/cm^2.
- Por tanto, en este caso d=1
- Si quiere rechazarse la hipótesis nula 95% de las veces cuando la diferencia de las medias es 0.5, entonces β=0.05
- En la figura, con β=0.05 y d=1 se obtiene n*=16 aproximadamente
- Por tanto, el tamaño de la muestra requerido es

$$n = \frac{n^*+1}{2} = \frac{16+1}{2} = 8.5 \simeq 9$$

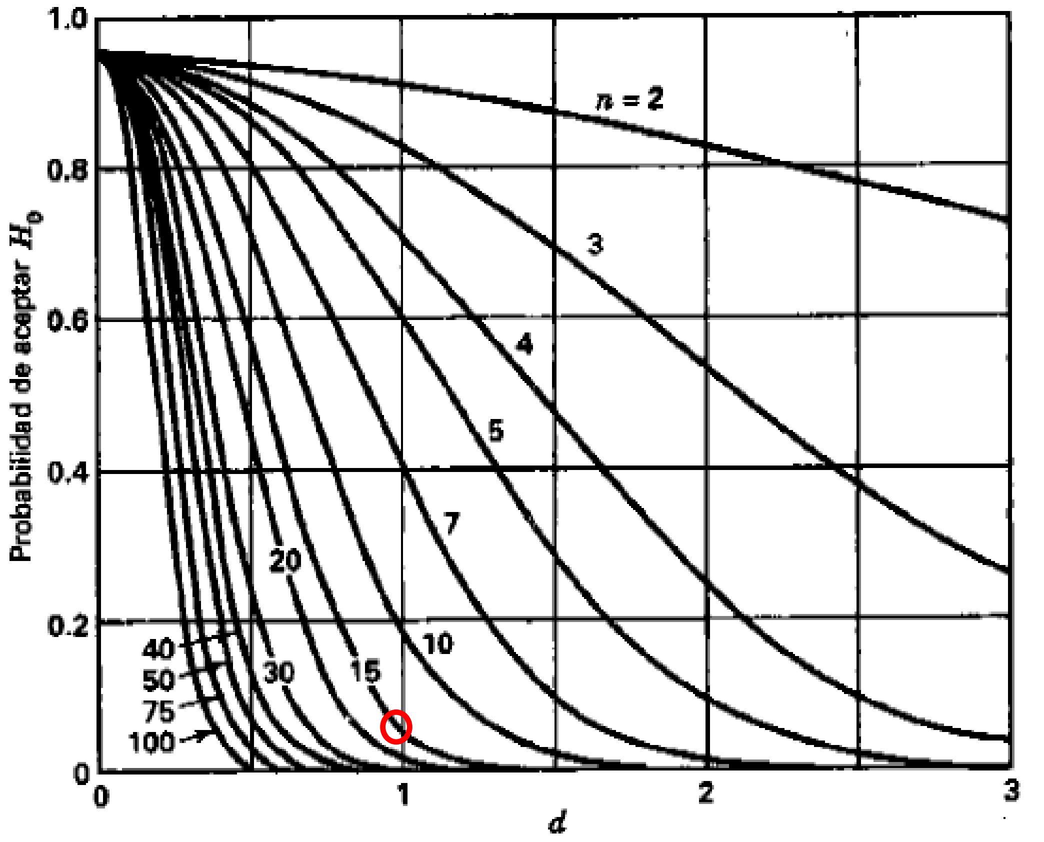


Figura 2-12 Curvas de operación característica para la prueba t de dos colas con α = 0.05. (Reproducida con permiso de "Operating Characteristics Curves for the Common Statistical Tests of Significance", C.L. Ferris, F.E. Grubbs y C.L. Weaver, *Annals of Mathematical Statistics*.)

Tamaño de la muestra para ANOVA

- Considere la probabilidad de error de tipo II del modelo con efectos fijos con el mismo tamaño de la muestra en cada tratamiento:

$$\begin{aligned}\beta &= 1 - P\{\text{Rechazar } H_0 \mid H_0 \text{ es falsa}\} \\ &= 1 - P\{F_0 > F_{\alpha,a-1,N-a} \mid H_0 \text{ es falsa}\}\end{aligned}$$

- Para hacer la evaluación de esta expresión se requiere conocer la distribución del estadístico de prueba F_0 si la hipótesis nula es falsa
- Puede demostrarse que, si H_0 es falsa el estadístico $F_0 = MS_{Tratamientos}/MS_E$ se distribuye como una variable aleatoria F no central con a-1 y N-a grados de libertad y parámetro de no centralidad δ. Si $\delta=0$, la distribución F no central se convierte en la distribución F central común.

Tamaño de la muestra para ANOVA

- Las curvas de probabilidad de error tipo II suelen representarse con respecto a un parámetro Φ
- Este parámetro puede calcularse con base en las medias con la expresión:

$$\Phi^2 = \frac{n\sum_{i=1}^{a} \tau_i^2}{a\sigma^2}$$

- Donde:

$$\tau_i = \mu_i - \bar{\mu} \qquad \bar{\mu} = (1/a)\Sigma_{i=1}^{a} \mu_i$$

- Se requiere también una estimación de la varianza (por experiencia previa) o determinar un posible rango de varianzas y estudiar el efecto de la varianza sobre el tamaño de la muestra requerido antes de seleccionar el valor final

Ejemplo

- Considere el experimento de la resistencia a la tensión descrito en el ejemplo 3-1. Suponga que el experimentador está interesado en rechazar la hipótesis nula con una probabilidad de al menos 0.9 si las medias de los cinco tratamientos son:

$$\mu_1=11 \qquad \mu_2=12 \qquad \mu_3=15 \qquad \mu_4=18 \qquad \text{y} \qquad \mu_5=19$$

- Suponga que se usa un α=0.01

Ejemplo

$$\bar{\mu} = (1/5)75 = 15$$

$$\left.\begin{aligned}
\tau_1 &= \mu_1 - \bar{\mu} = 11 - 15 = -4\\
\tau_2 &= \mu_2 - \bar{\mu} = 12 - 15 = -3\\
\tau_3 &= \mu_3 - \bar{\mu} = 15 - 15 = 0\\
\tau_4 &= \mu_4 - \bar{\mu} = 18 - 15 = 3\\
\tau_5 &= \mu_5 - \bar{\mu} = 19 - 15 = 4
\end{aligned}\right\} \quad \Sigma_{i=1}^{5}\tau_i^2 = 50$$

- Con un σ=3

$$\Phi^2 = \frac{n\sum_{i=1}^{5}\tau_i^2}{a\sigma^2} = \frac{n(50)}{5(3)^2} = 1.11n$$

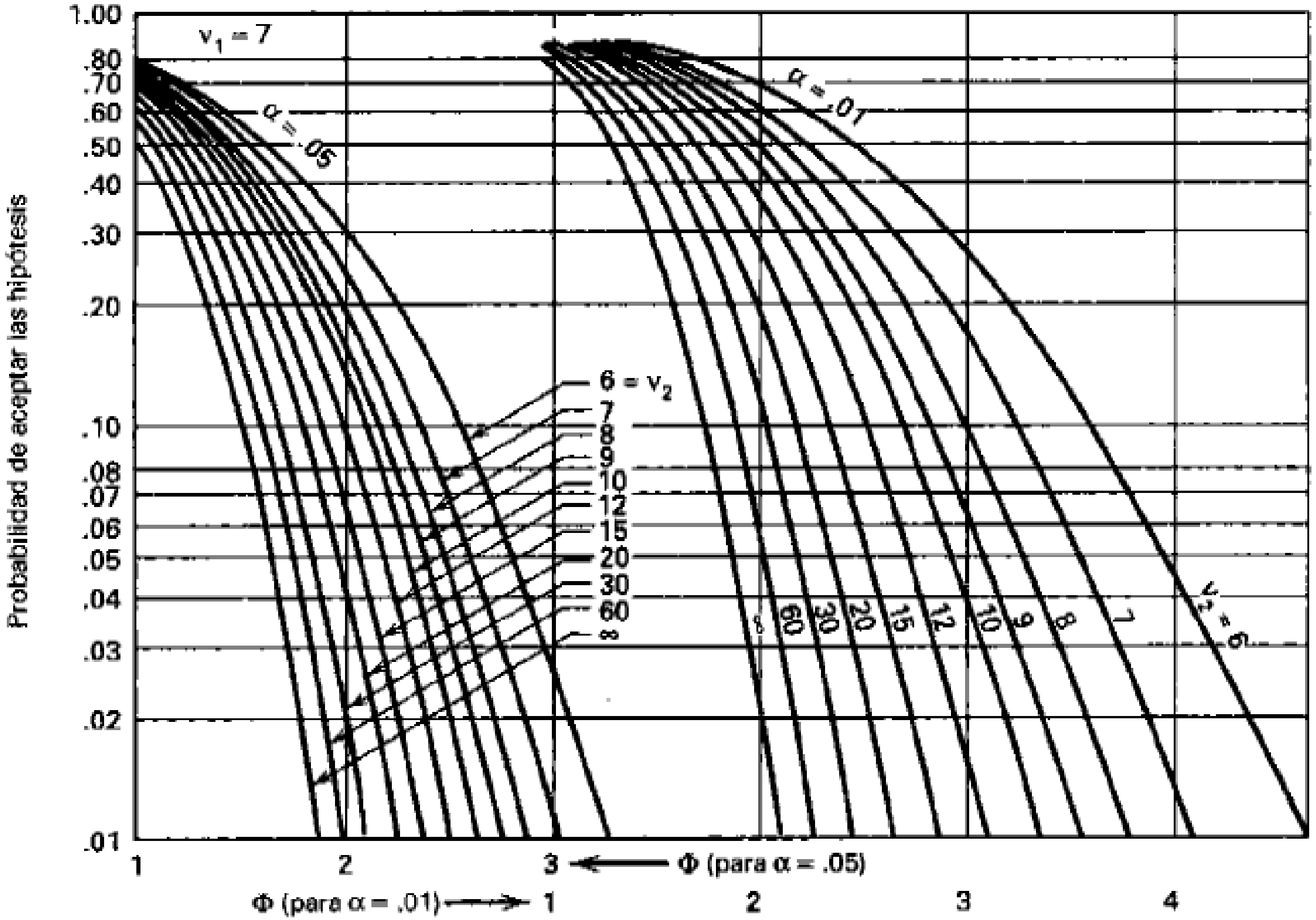
Ejemplo

- Se varía el número de réplicas hasta obtener la potencia requerida:

n	Φ^2	Φ	$a(n-1)$	β	Potencia $(1-\beta)$
4	4.44	2.11	15	0.30	0.70
5	5.55	2.36	20	0.15	0.85
6	6.66	2.58	25	0.04	0.96

- Por tanto, se escoge un n=6 para obtener la sensibilidad dada con α=0.01

Curvas OC para el modelo de efectos fijos



Método 2:Número de réplicas con relación a los intervalos de confianza

- El intervalo de confianza con la varianza estimada con el MSE será:

$$\pm t_{\alpha/2,N-a}\sqrt{\frac{2MS_E}{n}}$$

- Suponiendo un máximo intervalo de confianza de 5 (para un 95% de confianza)
- Con una varianza de 9 y al variar *n* :

$$\pm 2.086\sqrt{\frac{2(9)}{5}} = \pm 3.96 \qquad \pm 2.132\sqrt{\frac{2(9)}{4}} = \pm 4.52 \qquad \pm 2.228\sqrt{\frac{2(9)}{3}} = \pm 5.46$$

- n=4 será el valor que arroja el nivel de confianza requerido