

Capítulo 3

Distribuciones de Referencia

3.1 Error Experimental y Distribuciones de Referencia

Cuando una operación se repite bajo condiciones aproximadamente iguales, los resultados observados nunca son idénticos. Las fluctuaciones que ocurren entre repeticiones son llamadas *ruido*, *variación experimental*, *error experimental* o simplemente *error*. En este contexto, el término se refiere a variaciones inevitables, y no implica ninguna culpa.

El error experimental puede dificultar la identificación de componentes sistemáticas del fenómeno en estudio, y por tanto es necesario desarrollar métodos que tomen en cuenta este factor, como veremos en el siguiente ejemplo. Dicho ejemplo ha sido tomado de Box, Hunter y Hunter (1978)

3.1.1 Un Experimento Industrial

En una fábrica se lleva a cabo un experimento para determinar si un método modificado de producción (B) es mejor que el método que se ha usado tradicionalmente (A) en la producción de un determinado producto químico . El experimento consiste en llevar a cabo una secuencia de 10 lotes del producto con el método A, seguido de una secuencia de 10 lotes usando el método B. Llamaremos *corrida experimental* a cada uno de los puntos del experimento, es decir, a cada lote producido. En cada corrida se registra la cantidad del químico obtenida. Los resultados del experimento se presentan en la tabla 3.1. ¿Proporcionan los datos evidencia suficiente para afirmar que el método B es mejor que el método A?

Método	Número de corrida									
	1	2	3	4	5	6	7	8	9	10
A	89.7	81.4	84.5	84.8	87.3	79.7	85.1	81.7	83.7	84.5
B	84.7	86.1	83.2	91.9	86.3	79.3	82.6	89.1	83.7	88.5

Tabla 3.1: Resultados del experimento en la fábrica

Al calcular los promedios, obtenemos que $\bar{y}_B - \bar{y}_A = 85.54 - 84.24 = 1.3$. Es decir, el promedio del método modificado (B) es 1.3 unidades mayor que el promedio del método standard (A). Sin embargo, observamos en la figura 3.1 que las variaciones de los datos no permiten apreciar si esta diferencia indica o no una superioridad real del método B. Necesitamos, entonces, algún marco de referencia que nos permita decidir si 1.3 es una diferencia “suficientemente grande” como para afirmar que el método modificado es mejor que el método standard, es decir, si la diferencia es *significativa*.

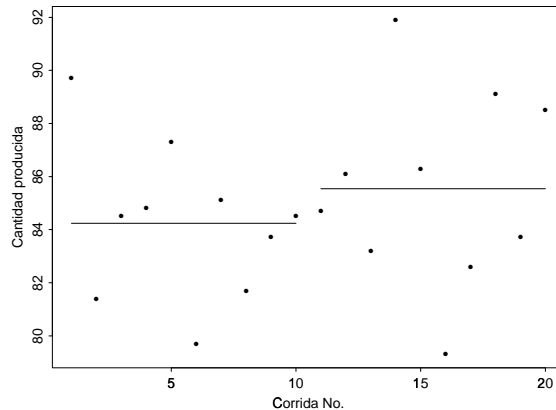


Figura 3.1: Cantidades producidas graficadas en orden cronológico.

Una manera de resolver este problema es buscar alguna *referencia* contra la cual comparar los resultados obtenidos. Por ejemplo, pudiéramos preguntarnos si una diferencia de 1.3 entre los promedios de dos grupos de 10 corridas consecutivas es normal para el método A. Si es así, no tendríamos ninguna razón para afirmar que el método B es mejor que el

método A y no incurriríamos en el gasto que significa modificar la planta para sustituir el método de producción. Sin embargo, si una diferencia de este tipo es “rara” para corridas del método A, parece razonable atribuirla al cambio de método de producción, con lo cual preferiremos el método B al método A.

Por supuesto, esta forma de proceder implica un riesgo, pues aún cuando la diferencia que estemos observando se presente en poquísimas oportunidades, no tenemos garantía de que no haya sido producto del azar. Por lo tanto el significado de “rara” dependerá del riesgo que estamos dispuestos a correr. Por ejemplo, si la modificación de la planta para cambiar al método B de producción implica grandes costos para la empresa, estaremos muy reacios a reconocer que el método B es mejor. Sin embargo, si la reconversión de la planta resulta muy económica, será mucho más fácil tomar una decisión que implique un cambio de método.

En base a este razonamiento, usaremos los registros históricos de 210 corridas consecutivas con el método A, para las cuales calcularemos todos los posibles promedios de 10 observaciones consecutivas. En la tabla 3.2 se muestran tanto las observaciones como los promedios, en orden cronológico por columnas.

En primer lugar, podríamos preguntarnos si las observaciones varían con el tiempo. Una variación en el tiempo indicaría que los promedios no son semejantes, y no sería posible emplearlos como referencia. Afortunadamente, en la figura 3.2 no observamos tendencia alguna que nos impida continuar con nuestro procedimiento.

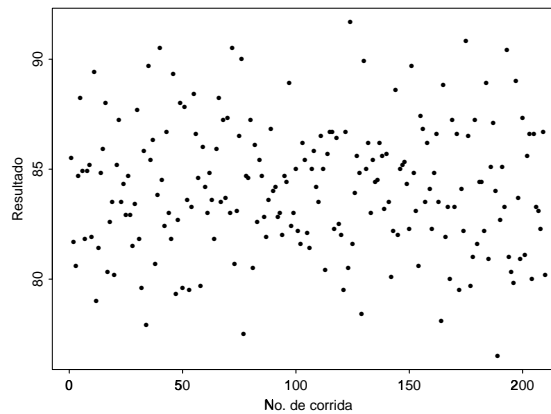


Figura 3.2: Gráfico de datos históricos en orden cronológico.

Obs.	Prom.	Obs.	Prom.	Obs.	Prom.	Obs.	Prom.	Obs.	Prom.	Obs.	Prom.
85.5		85.4	83.57	83.0	84.90	81.4	83.82	83.5	84.77	86.5	84.35
81.7		86.3	83.91	90.5	85.47	85.0	83.43	80.1	84.16	79.7	83.99
80.6		80.7	83.83	80.7	85.18	85.8	83.77	82.2	84.08	81.0	84.09
84.7		83.8	83.87	83.1	85.31	84.2	83.89	88.6	84.40	87.2	84.09
88.2		90.5	84.15	86.5	85.37	83.5	83.74	82.0	84.16	81.6	83.92
84.9		84.5	84.42	90.0	85.55	86.5	84.17	85.0	84.21	84.4	83.70
81.8		82.4	84.70	77.5	84.95	85.0	84.51	85.2	84.11	84.4	84.19
84.9		86.7	84.79	84.7	84.70	80.4	83.93	85.3	84.08	82.2	84.00
85.2		83.0	85.30	84.6	84.79	85.7	83.96	84.3	84.19	88.9	84.67
81.9	83.94	81.8	84.51	87.2	84.78	86.7	84.42	82.3	83.85	80.9	83.68
89.4	84.33	89.3	84.90	80.5	84.53	86.7	84.95	89.7	84.47	85.1	83.54
79.0	84.06	79.3	84.20	86.1	84.09	82.3	84.68	84.8	84.94	87.1	84.28
81.4	84.14	82.7	84.40	82.6	84.28	86.4	84.74	83.1	85.03	84.0	84.58
84.8	84.15	88.0	84.82	85.4	84.51	82.5	84.57	80.6	84.23	76.5	83.51
85.9	83.92	79.6	83.73	84.7	84.33	82.0	84.42	87.4	84.77	82.7	83.62
88.0	84.23	87.8	84.06	82.8	83.61	79.5	83.72	86.8	84.95	85.1	83.69
80.3	84.08	83.6	84.18	81.9	84.05	86.7	83.89	83.5	84.78	83.3	83.58
82.6	83.85	79.5	83.46	83.6	83.94	80.5	83.90	86.2	84.87	90.4	84.40
83.5	83.68	83.3	83.49	86.8	84.16	91.7	84.50	84.1	84.85	81.0	83.61
80.2	83.51	88.4	84.15	84.0	83.84	81.6	83.99	82.3	84.85	80.3	83.55
85.2	83.09	86.6	83.88	84.2	84.21	83.9	83.71	84.8	84.36	79.8	83.02
87.2	83.91	84.6	84.41	82.8	83.88	85.6	84.04	86.6	84.54	89.0	83.21
83.5	84.12	79.7	84.11	83.0	83.92	84.8	83.88	83.5	84.58	83.7	83.18
84.3	84.07	86.0	83.91	82.0	83.58	78.4	83.47	78.1	84.33	80.9	83.62
82.9	83.77	84.2	84.37	84.7	83.58	89.9	84.26	88.8	84.47	87.3	84.08
84.7	83.44	83.0	83.89	84.4	83.74	85.0	84.81	81.9	83.98	81.1	83.68
82.9	83.70	84.8	84.01	88.9	84.44	86.2	84.76	83.3	83.96	85.6	83.91
81.5	83.59	83.6	84.42	82.4	84.32	83.0	85.01	80.0	83.34	86.6	83.53
83.4	83.58	81.8	84.27	83.0	83.94	85.4	84.38	87.2	83.65	80.0	83.43
87.7	84.33	85.9	84.02	85.0	84.04	84.4	84.66	83.3	83.75	86.6	84.06
81.8	83.99	88.2	84.18	82.2	83.84	84.5	84.72	86.6	83.93	83.3	84.41
79.6	83.23	83.5	84.07	81.6	83.72	86.2	84.78	79.5	83.22	83.1	83.82
85.8	83.46	87.2	84.82	86.2	84.04	85.6	84.86	84.1	83.28	82.3	83.68
77.9	82.82	83.7	84.59	85.4	84.38	83.2	85.34	82.2	83.69	86.7	84.26
89.7	83.50	87.3	84.90	82.1	84.12	85.7	84.92	90.8	83.89	80.2	83.55

Tabla 3.2: Registros históricos y promedios de 10 observaciones para el experimento industrial.

Finalmente, en la figura 3.3 se presenta un diagrama de puntos para las 191 diferencias. En este diagrama puede observarse que sólo en 9 de los 191 casos (4.71%) la diferencia es mayor o igual que 1.3. Esto nos lleva a pensar que la diferencia de los resultados obtenidos entre ambos métodos no debe ser producto del azar, sino de una diferencia real entre ambos métodos de producción.

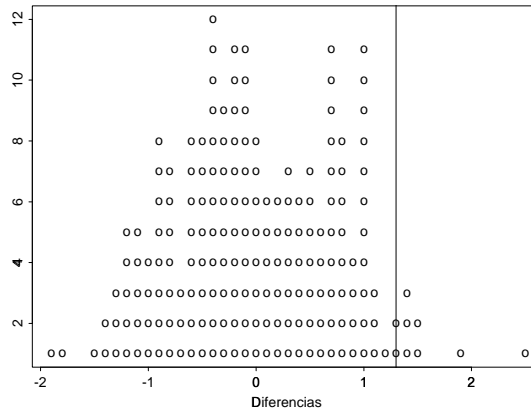


Figura 3.3: Diagrama de puntos de las diferencias.

3.1.2 Pruebas de significancia

Tratemos de resumir el procedimiento que hemos empleado para responder a la pregunta de si el método B es mejor que el método A. En primer lugar, actuamos como abogados del diablo e introducimos una *hipótesis nula*, la cual plantea que los dos métodos son iguales. A la hipótesis de que el método B es mejor que el método A la llamaremos *hipótesis alternativa*. Como es costoso cambiar el método de producción, preferiremos el método B al método A únicamente si los datos proporcionan una evidencia abrumadora a favor del método B. Podríamos decir, entonces, que “la hipótesis nula es inocente mientras no se demuestre lo contrario”.

A fin de expresar estas hipótesis, observemos que en realidad estamos hablando de dos poblaciones hipotéticas: una población de resultados del método A y otra de resultados del método B. Llamemos μ_A a la producción media con el método A, y μ_B a la del método B. Entonces, las hipótesis nula y alternativa se pueden escribir como:

$$H_0 : \mu_A = \mu_B \quad \text{vs} \quad H_1 : \mu_B > \mu_A$$

Para decidir si los datos desacreditan la hipótesis nula lo suficiente como para preferir la hipótesis alternativa, comparemos 1.3 con todas las diferencias de promedios de 10 observaciones. Es decir, el problema involucra los siguientes elementos:

Población	Muestra
Un conjunto muy grande (quizás infinito) de observaciones, de las cuales podría proceder la muestra. (Registros históricos)	Un pequeño conjunto de observaciones ya obtenidas. (Datos del experimento)
Parámetro Media Poblacional μ	Estadística $\bar{Y} = \frac{\sum Y_i}{n}$

En base a estos elementos, deseamos decidir entre $H_0 : \mu_B - \mu_A = 0$ y $H_1 : \mu_B - \mu_A > 0$, que involucran a los parámetros de la población, en base a la información proporcionada por la muestra, la cual resumimos en la estadística $\bar{y}_B - \bar{y}_A$. El problema es, entonces, cómo decidir si $\bar{y}_B - \bar{y}_A$ es lo suficientemente grande como para desacreditar H_0 y convencernos de que vale la pena hacer el gasto involucrado. Para tener un criterio, suponemos que la población tiene una cierta distribución y ésta será nuestra distribución de referencia. Si con nuestra referencia $\bar{y}_B - \bar{y}_A = 1.3$ es poco probable, decimos que la diferencia es *estadísticamente significativa*. Por esta razón, diremos que este procedimiento corresponde a una *prueba de significancia*.

En general, no disponemos de registros históricos. En ese caso, recurrimos a modelos probabilísticos, y suponemos que los datos siguen una cierta distribución de probabilidad. Esta distribución actuará como nuestra referencia, permitiéndonos determinar si el resultado que hemos obtenido es lo suficientemente poco probable como para desacreditar la hipótesis nula.

3.1.3 Errores de tipo I y II

Cuando decidimos probar una hipótesis usando el procedimiento arriba descrito, corremos sin duda el riesgo de cometer errores. Estos errores pueden ser de dos tipos:

- Podemos rechazar la hipótesis nula cuando es cierta. A esta clase de error la llamaremos *Error de Tipo I*.

- Otra posibilidad es no rechazar la hipótesis nula cuando es falsa. Este error es el *Error de tipo II*.

A la probabilidad de cometer un error de tipo I se le suele llamar *Nivel del test* o *Nivel de la prueba* y se suele representar como α .

$$P(\text{Rechazar } H_0 | H_0) = \alpha$$

A la probabilidad de cometer un error de tipo II suele llamársele β .

$$P(\text{No rechazar } H_0 | H_1) = \beta$$

A la probabilidad de rechazar H_0 cuando es falsa, es decir, $1 - \beta$, la llamaremos *Potencia de la prueba*. Es importante destacar que $1 - \beta \neq \alpha$.

Claramente, si hacemos α muy pequeño β aumentará, de la misma forma en que si exigimos muchas pruebas para condenar a un acusado en un juicio corremos un mayor riesgo de no condenar a un culpable. El caso contrario también es cierto: si queremos obtener una potencia grande (β pequeño) α aumentará. El equivalente en el juicio nos indicaría que si queremos condenar a los culpables a toda costa, corremos un riesgo alto de condenar también inocentes. Debemos tomar entonces alguna decisión acerca de cuál error nos produce mayor pérdida y qué riesgo estamos dispuestos a correr. En general, como ya se comentó, elegimos como H_0 la hipótesis más conservadora, es decir, aquella que “será inocente mientras no se demuestre lo contrario”. Como no deseamos condenar a un inocente, fijaremos el nivel α de la prueba como una cantidad cónsona con el riesgo que estamos dispuestos a correr.

La elección del valor de α tiene que ver también con la calidad y la variabilidad de los datos obtenidos. Por ejemplo, si estamos analizando una situación en la cual están involucrados animales, plantas o seres humanos, debemos tomar en cuenta que existe una enorme variabilidad entre individuos. Fijar un valor muy bajo de α podría hacernos incapaces de notar señal alguna, pues la perderíamos dentro del ruido. Si los datos provienen de un experimento controlado, podemos ser algo más exigentes. Algunos valores tradicionales para α son 0.1, 0.05 ó 0.01.

3.1.4 La naturaleza esencial de una prueba de significancia:

Resumiendo toda la discusión anterior, podemos describir así la naturaleza esencial de una prueba de significancia.

1. Obtenemos de los datos algún criterio relevante (o estadístico) apropiado para probar alguna hipótesis particular contra la hipótesis alternativa.
2. Referimos nuestro criterio a alguna distribución de referencia adecuada que mostraría cómo se distribuiría el criterio bajo H_0 .
3. Así podemos calcular la probabilidad de que una discrepancia por lo menos tan grande como lo observado pueda ocurrir.
4. Esta probabilidad es llamada *nivel de significancia*. Si es suficientemente pequeña rechazamos H_0 y decimos que hay una diferencia estadísticamente significativa.

3.2 Distribuciones de Vectores Normales

Antes de pasar a la definición de la Normal multivariada, recordemos que la esperanza de una variable aleatoria n-dimensional $\mathbf{X} = (X_1, X_2, \dots, X_n)'$ se define como

$$E(\mathbf{X}) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{pmatrix}.$$

Así mismo, se define la matrix de varianza de \mathbf{X} como la matriz que contiene todas las varianzas y covarianzas de los elementos de \mathbf{X} , es decir:

$$\begin{aligned} \text{Var}(\mathbf{X}) &= \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \dots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_1, X_2) & \text{Var}(X_2) & \dots & \text{Cov}(X_2, X_n) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(X_1, X_n) & \text{Cov}(X_2, X_n) & \dots & \text{Var}(X_n) \end{pmatrix} \\ &= E \{(\mathbf{X} - E(\mathbf{X}))(\mathbf{X} - E(\mathbf{X}))'\} \end{aligned}$$

En forma equivalente, se define la covarianza entre dos vectores aleatorios \mathbf{X} y \mathbf{Y} como:

$$\begin{aligned} \text{Cov}(X, Y) &= E \{(\mathbf{X} - E(\mathbf{X}))(\mathbf{Y} - E(\mathbf{Y}))'\} \\ &= \begin{pmatrix} \text{Cov}(X_1, Y_1) & \text{Cov}(X_1, Y_2) & \dots & \text{Cov}(X_1, Y_n) \\ \text{Cov}(X_2, Y_1) & \text{Cov}(X_2, Y_2) & \dots & \text{Cov}(X_2, Y_n) \\ \dots & \dots & \dots & \dots \\ \text{Cov}(X_n, Y_1) & \text{Cov}(X_n, Y_2) & \dots & \text{Cov}(X_n, Y_n) \end{pmatrix} \end{aligned}$$

No es difícil (aunque sí laborioso) probar que:

$$\begin{aligned} E(\mathbf{AX}) &= A E(\mathbf{X}) \\ \text{Var}(\mathbf{AX}) &= A \text{Var}(\mathbf{X}) A' \end{aligned}$$

3.2.1 Distribución Normal multivariada

Decimos que el vector aleatorio \mathbf{Y} tiene distribución Normal n -variada con vector de medias $\boldsymbol{\mu}$ y matriz de covarianza Σ si su función de densidad conjunta tiene la forma

$$f(\mathbf{y}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) \right\}$$

En la figura 3.4 se presenta la forma de una normal bivariada, así como sus curvas de nivel. Observemos que cuando se corta la superficie con planos perpendiculares al plano (x, y) se obtienen curvas normales. Esto nos permite ver que todas las distribuciones marginales y condicionales también son normales. Esta propiedad puede extenderse al caso de la normal n -variada, y es posible probar que cualquier subconjunto de k variables tiene distribución conjunta normal k -variada ($k \leq n$).

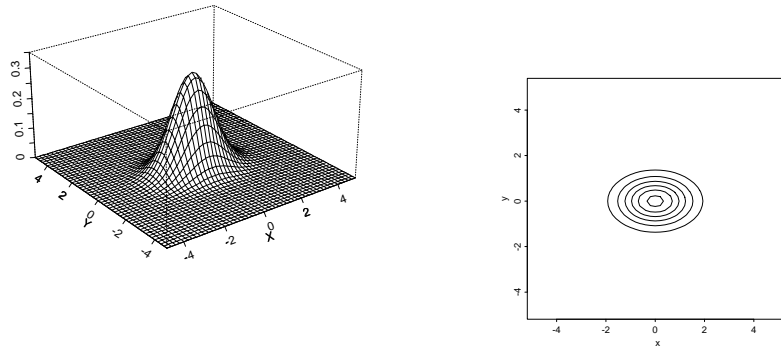


Figura 3.4: Gráfico y curvas de nivel para una normal bivariada

Cuando la matriz de varianzas Σ es diagonal,

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}$$

(es decir, las covarianzas entre las variables son cero), la forma cuadrática que aparece en la exponencial se convierte en:

$$(\mathbf{y} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{y} - \boldsymbol{\mu}) = \sum_{i=1}^n \left(\frac{y_i - \mu_i}{\sigma_i} \right)^2$$

Esto permite escribir la densidad conjunta como el producto de las densidades marginales, e indica que las componentes del vector \mathbf{Y} , Y_1, \dots, Y_n , son independientes. Es decir, *dos variables normales son independientes si y sólo si su covarianza es cero*. (Recordemos que, en general, no es cierto que variables no correlacionadas sean independientes)

Otra propiedad importante tiene que ver con la distribución de combinaciones lineales de variables normales. Puede probarse que al definir

$$\mathbf{X} = A\mathbf{Y},$$

donde A es cualquier matriz $m \times n$, el vector \mathbf{X} tendrá distribución normal m -variada con vector de medias $A\boldsymbol{\mu}$ y matriz de varianza $A\Sigma A'$.

Diremos que \mathbf{Z} tiene distribución normal standard n -variada si $E(\mathbf{Z}) = \mathbf{0}$ y $Var(\mathbf{Z}) = I$. En ese caso:

$$f(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n z_i^2 \right\}$$

Cualquier vector \mathbf{Y} normal n -dimensional con matriz de varianza Σ no singular puede llevarse a un vector \mathbf{Z} normal standard por medio de una transformación lineal, haciendo $\mathbf{Z} = A^{-1}(\mathbf{Y} - \boldsymbol{\mu})$, donde A es una matriz cuadrada tal que

$$\Sigma = AA'$$

(puede garantizarse la existencia de tal A por ser Σ simétrica y positiva definida). La demostración queda como ejercicio para el lector.

A continuación estudiaremos algunas distribuciones derivadas de la normal.

3.2.2 Distribución χ^2 de Pearson

Dadas Z_1, \dots, Z_n independientes e idénticamente distribuidas, $Z_i \sim N(0, 1)$, (es decir, $\mathbf{Z} \sim N(\mathbf{0}, I)$) la norma al cuadrado del vector $\mathbf{Z} = (Z_1, \dots, Z_n)$ sigue una distribución que llamaremos *distribución χ^2 con n grados de libertad*. Es decir:

$$X^2 = Z_1^2 + \dots + Z_n^2 \sim \chi_n^2$$

Puede probarse que una χ^2 es una Gamma con parámetros $\frac{n}{2}$ y $\frac{1}{2}$. Esto permite calcular su esperanza y su varianza, obteniendo para $X^2 \sim \chi_n^2$:

$$E(X^2) = n$$

$$\text{Var}(X^2) = 2n$$

En la figura 3.5 se muestra la forma de la densidad para algunos valores de n . Vemos que la curva se suaviza a medida que los grados de libertad aumentan.

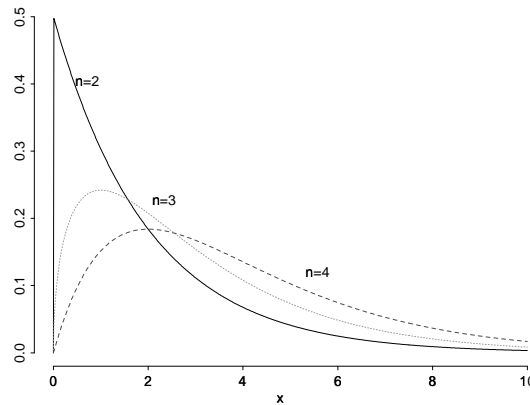


Figura 3.5: Algunas densidades χ^2 .

Los cuantiles correspondientes a ciertos valores comúnmente usados de la probabilidad se encuentran tabulados con respecto a los grados de libertad.

Usando las propiedades de la distribución Gamma, puede verse también que si $X_1^2 \sim \chi_{n_1}^2$ y $X_2^2 \sim \chi_{n_2}^2$ son independientes entonces su suma es también una χ^2 con grados de libertad iguales a la suma de los grados de libertad de X_1^2 y X_2^2 , es decir

$$X_1^2 + X_2^2 \sim \chi_{n_1+n_2}^2.$$

3.2.3 Distribución F de Fisher

Supongamos que \mathbf{X} e \mathbf{Y} son vectores de dimensión n y m , respectivamente, tales que $\mathbf{X} \sim N_n(\mathbf{0}, I)$ e $\mathbf{Y} \sim N_m(\mathbf{0}, I)$. Los cuadrados de sus normas siguen, entonces, distribu-

ciones χ_n^2 y χ_m^2 respectivamente. Decimos que el cociente normalizado:

$$F = \frac{\frac{\chi_n^2}{n}}{\frac{\chi_m^2}{m}}$$

se distribuye como una F con n grados de libertad en el numerador y m grados de libertad en el denominador ($F \sim F_{n,m}$).

Nótese además que esta definición implica que

$$F_{n,m} = F_{m,n}^{-1} = \frac{1}{F_{m,n}}$$

La esperanza y la varianza de esta distribución están dadas por:

$$E(F) = \frac{m}{m-2}, \quad m > 2$$

$$Var(F) = \frac{2m^2(n+m-2)}{n(m-2)^2(m-4)^2}, \quad m > 4$$

Ciertos cuantiles destacados de esta distribución están tabulados en función de n y m , siendo necesaria en este caso una tabla por cada cuantil.

En la figura 3.6 se observa la forma de la densidad correspondiente a esta distribución para algunos valores de n y m .

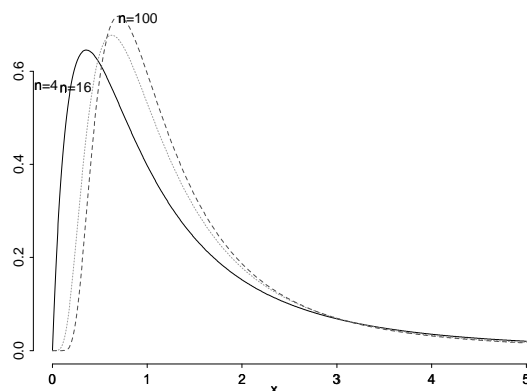


Figura 3.6: Forma de la densidad F , $m = 5$.

3.2.4 Distribución t de Student

Si $Z \sim N(0, 1)$ y $X^2 \sim \chi_n^2$, la variable aleatoria

$$T = \frac{Z}{\sqrt{\frac{X^2}{n}}}$$

sigue una distribución que llamaremos t de Student con n grados de libertad.

La t es una distribución simétrica que tiene colas más pesadas que las de la normal standard, pero se acerca rápidamente a ésta cuando aumentan los grados de libertad, como puede verse en la figura 3.7.

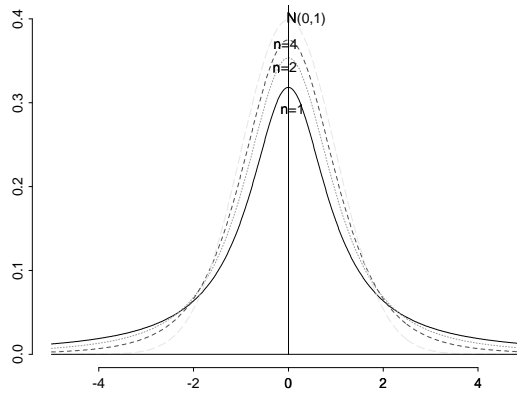


Figura 3.7: Densidad de la distribución t .

Notemos además que si $T \sim t_n$, entonces $T^2 \sim F_{1,n}$.

La variable $T \sim t_n$ tiene esperanza 0 y varianza

$$\text{Var}(T) = \frac{n}{n-2}, \quad n > 2$$

3.3 Algunos teoremas sobre vectores normales

Teorema 3.1 Supongamos que \mathbf{Y} es un vector n -dimensional cuyas componentes son $N(\mu_i, \sigma^2)$ e independientes (es decir, $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2 I)$), y que X_1 y X_2 se obtienen como

funciones lineales de \mathbf{Y} , es decir

$$X_1 = \mathbf{a}'\mathbf{Y} = \sum_{i=1}^n a_i y_i$$

$$X_2 = \mathbf{b}'\mathbf{Y} = \sum_{i=1}^n b_i y_i$$

con \mathbf{a} y \mathbf{b} vectores n -dimensionales fijos. Entonces, X_1 y X_2 serán independientes si y sólo si \mathbf{a} y \mathbf{b} son ortogonales, es decir, si y sólo si $\mathbf{a}'\mathbf{b} = 0$

Demostración: Como X_1 y X_2 son normales por ser combinaciones lineales de un vector normal, basta ver que su covarianza es nula.

$$\begin{aligned} \text{Cov}(X_1, X_2) &= \text{Cov}(\mathbf{a}'\mathbf{Y}, \mathbf{b}'\mathbf{Y}) \\ &= E\{(\mathbf{a}'\mathbf{Y} - \mathbf{a}'\boldsymbol{\mu})(\mathbf{b}'\mathbf{Y} - \mathbf{b}'\boldsymbol{\mu})'\} \\ &= E\{\mathbf{a}'(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})'\mathbf{b}\} \\ &= \mathbf{a}'E\{(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})'\}\mathbf{b} \\ &= \mathbf{a}'\text{Var}(\mathbf{Y})\mathbf{b} \\ &= \mathbf{a}'\sigma^2 I\mathbf{b} \\ &= \sigma^2 \mathbf{a}'\mathbf{b} \end{aligned}$$

Por lo tanto X_1 y X_2 serán independientes si y sólo si $\mathbf{a}'\mathbf{b} = 0$, es decir, si y sólo si es decir \mathbf{a} y \mathbf{b} son ortogonales.

Teorema 3.2 Sean $\mathbf{X}_1 = A_1\mathbf{Y}$ y $\mathbf{X}_2 = A_2\mathbf{Y}$, donde A_1 y A_2 son matrices $k_1 \times n$ y $k_2 \times n$ respectivamente y $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2 I)$. Si $A_1 A_2' = 0_{k_1 \times k_2}$ entonces \mathbf{X}_1 y \mathbf{X}_2 son independientes.

Demostración: La demostración se hace en forma similar al teorema anterior.

Teorema 3.3 Sean $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2 I)$ y $\mathbf{X} = C\mathbf{Y}$, donde $C_{n \times n}$ es una matriz ortogonal ($C^{-1} = C'$). Entonces $\mathbf{X} \sim N_n(C\boldsymbol{\mu}, \sigma^2 I)$.

Demostración: Por las propiedades de la normal multivariada, sabemos que \mathbf{X} sigue una distribución normal, de manera que es suficiente determinar su esperanza y su matriz de varianza y covarianza. Es inmediato que $E(\mathbf{X}) = E(C\mathbf{Y}) = C\boldsymbol{\mu}$. Con respecto a la matriz de varianza,

$$\begin{aligned} \text{Var}(\mathbf{X}) &= E\{(C\mathbf{Y} - C\boldsymbol{\mu})(C\mathbf{Y} - C\boldsymbol{\mu})'\} \\ &= E\{C(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})'C'\} \\ &= CE\{(\mathbf{Y} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\mu})'\}C' \\ &= C\text{Var}(\mathbf{Y})C' \\ &= C\sigma^2IC' \\ &= \sigma^2CC' \end{aligned}$$

Por lo tanto, $\text{Var}(\mathbf{X}) = \sigma^2I$. Esto demuestra el teorema.

3.4 Formas Cuadráticas de Vectores Normales

En la siguiente sección nos ocuparemos de las distribuciones de expresiones de la forma $\mathbf{Y}'A\mathbf{Y}$, donde A es una matriz simétrica y positiva definida, es decir, de formas cuadráticas. En lo que sigue supondremos que $\mathbf{Y} \sim N_n(\mathbf{0}, I)$.

Teorema 3.4 *La forma cuadrática $\mathbf{Y}'A\mathbf{Y}$ se distribuye χ_r^2 si y sólo si A es una matriz idempotente de rango r .*

Demostración: Veremos sólo que la idempotencia de A es condición suficiente. Pueden encontrarse en diversos libros demostraciones de que es necesaria.

Como A es simétrica, puede diagonalizarse usando una matriz ortogonal, es decir, existe una matriz ortogonal C tal que $CAC' = D$, donde D es una matriz diagonal tal que los elementos en la diagonal son los autovalores de la matriz A . Observemos además que como A es idempotente de rango r , sus autovalores deben ser r unos y $n - r$ ceros, de manera que D es una matriz que contiene r unos en la diagonal y el resto son ceros. Podemos suponer, sin pérdida de generalidad, que los r unos se encuentran en las posiciones $d_{11}, d_{22} \dots d_{rr}$.

Definamos $\mathbf{X} = C\mathbf{Y}$. Aplicando el teorema anterior, vemos que $\mathbf{X} \sim N_n(\mathbf{0}, I)$. Como C es una matriz ortogonal, podemos escribir $\mathbf{Y} = C'\mathbf{X}$. Por lo tanto:

$$\begin{aligned} \mathbf{Y}'A\mathbf{Y} &= \mathbf{X}'CAC'\mathbf{X} \\ &= \mathbf{X}'D\mathbf{X} \\ &= \sum_{i=1}^r X_i^2 \end{aligned}$$

Como cada $X_i \sim N(0, 1)$, $\mathbf{Y}'\mathbf{A}\mathbf{Y} = \sum_{i=1}^r X_i^2 \sim \chi_r^2$

Corolario 3.1 Si $\mathbf{Y} \sim N(\mathbf{0}, I)$, el cuadrado de la norma de su proyección sobre un subespacio de dimensión r se distribuye χ_r^2 .

Demostración: Sea A la matriz de la proyección sobre el subespacio de dimensión r . Por tanto, ésta es una matriz simétrica, idempotente y de rango r .

Calculando la norma de la proyección de \mathbf{Y} , obtenemos

$$\|\mathbf{A}\mathbf{Y}\|^2 = \mathbf{Y}'\mathbf{A}'\mathbf{A}\mathbf{Y}$$

Como A es simétrica, $A' = A$ y $A'A = A^2$. Adicionalmente, como A es idempotente, $A^2 = A$. Por lo tanto

$$\|\mathbf{A}\mathbf{Y}\|^2 = \mathbf{Y}'\mathbf{A}\mathbf{Y}$$

Es decir, la norma de $\mathbf{A}\mathbf{Y}$ es una forma cuadrática de un vector $N_n(\mathbf{0}, I)$ donde la matriz de la forma cuadrática es idempotente. Por el teorema anterior

$$\|\mathbf{A}\mathbf{Y}\|^2 \sim \chi_r^2$$

Teorema 3.5 Las formas cuadráticas $\mathbf{Y}'\mathbf{A}\mathbf{Y}$ y $\mathbf{Y}'\mathbf{B}\mathbf{Y}$, con $\mathbf{Y} \sim N(\mathbf{0}, I)$ y A, B simétricas e idempotentes son independientes si y sólo si $AB = 0$.

Demostración: Nuevamente, sólo probaremos que la condición es suficiente.

En primer lugar, observemos que al ser A y B idempotentes, $\mathbf{Y}'\mathbf{A}\mathbf{Y} = \|\mathbf{A}\mathbf{Y}\|^2$ y $\mathbf{Y}'\mathbf{B}\mathbf{Y} = \|\mathbf{B}\mathbf{Y}\|^2$. Por lo tanto, bastaría probar que los vectores $\mathbf{A}\mathbf{Y}$ y $\mathbf{B}\mathbf{Y}$ son independientes. Pero el Teorema 3.2 nos dice que si $AB' = 0$, entonces $\mathbf{A}\mathbf{Y}$ y $\mathbf{B}\mathbf{Y}$ son independientes. Como $AB = 0$ y B es una matriz simétrica, la condición se cumple y la independencia queda garantizada.

Corolario 3.2 Si descomponemos \mathbf{Y} en su proyección ortogonal sobre E_p subespacio de dimensión p y su complemento ortogonal E_{n-p} , los cuadrados de sus normas siguen distribuciones χ_p^2 y χ_{n-p}^2 independientes.

Demostración: Sea A_p la matriz de proyección sobre E_p ; por tanto $(I - A_p)$, también simétrica e idempotente, es la matriz de proyección sobre E_{n-p} . El vector \mathbf{Y} puede descomponerse como la suma de las proyecciones sobre E_p y E_{n-p} , es decir.

$$\mathbf{Y} = A_p \mathbf{Y} + (I - A_p) \mathbf{Y}$$

El corolario 3.1 nos indica que las distribuciones de $\|A_p \mathbf{Y}\|^2$ y $\|(I - A_p) \mathbf{Y}\|^2$ son χ_p^2 y χ_{n-p}^2 respectivamente, y el teorema 3.5 nos permite afirmar que ambas normas son independientes, lo cual finaliza la demostración.

Este último teorema puede extenderse a más de dos espacios ortogonales entre sí, y se demuestra en forma semejante.

Teorema 3.6 *Supongamos que $\mathbf{Y} \sim N_n(\mathbf{0}, I)$ se descompone como*

$$\mathbf{Y} = \mathbf{Y}_1 + \mathbf{Y}_2 + \dots + \mathbf{Y}_k$$

donde las componentes son las proyecciones de \mathbf{Y} sobre subespacios ortogonales E_1, E_2, \dots, E_k , de modo que $\mathbf{Y}_i' \mathbf{Y}_j = 0$ si $i \neq j$. Sea $r_i = \dim(E_i)$ Entonces

$$\|\mathbf{Y}_i\|^2 \sim \chi_{r_i}^2$$

y estas distribuciones son independientes

3.4.1 Algunas aplicaciones

Para aplicar los resultados obtenidos en esta sección, comencemos por probar el siguiente teorema

Teorema 3.7 *Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas como una $N(\mu, \sigma^2)$, y definamos S^2 como*

$$S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1},$$

entonces $\frac{(n-1)S^2}{\sigma^2}$ se distribuye como una χ_{n-1}^2 , y S^2 es independiente de \bar{X} .

Demostración: Observemos que $\bar{X} = \frac{1}{n}\mathbf{1}'\mathbf{X}$, donde $\mathbf{1} = (1, 1, \dots, 1)'$ y $\mathbf{X} = (X_1, \dots, X_n)$.

Definimos $Z_i = \frac{X_i - \mu}{\sigma}$. Claramente $Z_i \sim N(0, 1)$ y se cumple que $\bar{Z} = \frac{\bar{X} - \mu}{\sigma}$

Consideremos la matriz $(I - \frac{1}{n}\mathbf{1}\mathbf{1}')$. Es claro que se trata de una matriz simétrica. Veamos que además es idempotente:

$$\begin{aligned} (I - \frac{1}{n}\mathbf{1}\mathbf{1}')(I - \frac{1}{n}\mathbf{1}\mathbf{1}') &= I - \frac{1}{n}\mathbf{1}\mathbf{1}' - \frac{1}{n}\mathbf{1}\mathbf{1}' + \frac{1}{n}\mathbf{2}\mathbf{1}\mathbf{1}'\mathbf{1}' \\ &= I - \frac{1}{n}\mathbf{2}\mathbf{1}\mathbf{1}' \\ &= I - \frac{1}{n}\mathbf{1}\mathbf{1}' \end{aligned}$$

Tenemos entonces una matriz simétrica, idempotente y de rango $n - 1$ (Corresponde a la proyección sobre el complemento ortogonal del espacio generado por el vector $\mathbf{1}$). Esto implica que $\|(I - \frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{Z}\|^2 \sim \chi_{n-1}^2$. Además

$$\begin{aligned} (I - \frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{Z} &= \mathbf{Z} - \frac{1}{n}\sum_{i=1}^n n\mathbf{Z}_i\mathbf{1} \\ &= \mathbf{Z} - \bar{Z}\mathbf{1} \\ &= \begin{pmatrix} Z_1 - \bar{Z} \\ \vdots \\ Z_n - \bar{Z} \end{pmatrix} \end{aligned}$$

Por lo tanto

$$\begin{aligned} \|(I - \frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{Z}\|^2 &= \sum_{i=1}^n (Z_i - \bar{Z})^2 \\ &= \sum_{i=1}^n \left(\frac{X_i - \mu}{\sigma} - \frac{\bar{X} - \mu}{\sigma} \right)^2 \\ &= \sum_{i=1}^n \frac{1}{\sigma^2} (X_i - \bar{X})^2 \\ &= \frac{(n-1)S^2}{\sigma^2} \end{aligned}$$

Esto prueba que $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$

Para probar la independencia, veamos que \mathbf{Z} puede descomponerse como suma de dos vectores ortogonales haciendo

$$\mathbf{Z} = \frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{Z} + \left(I - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{Z},$$

donde ambos vectores son independientes por los teoremas antes demostrados.

Pero

$$\frac{1}{n}\mathbf{1}\mathbf{1}'\mathbf{Z} = \bar{Z}\mathbf{1} = \frac{\bar{X} - \mu}{\sigma}\mathbf{1}.$$

Por consiguiente, \bar{X} y S^2 son independientes.

Este resultado nos permite además afirmar que

$$\frac{\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{S^2}{\sigma^2}}} = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t_{n-1}$$

Veamos como usar esta última propiedad para construir una prueba de significancia.

Ejemplo 3.1 *El tiempo de vida de una bebida gaseosa embotellada es la variable de interés. Se seleccionan 10 botellas al azar y se prueban, obteniéndose los resultados que se muestran en la tabla 3.3. ¿Permiten los datos afirmar que el tiempo de duración de la bebida es mayor de 125 días?*

Días	
108	124
124	106
115	138
163	159
134	139

Tabla 3.3: Tiempos de vida de diez botellas de bebida gaseosa (Ejemplo 3.1)

Llamemos X a la variable aleatoria que representa el tiempo de vida en botella de la bebida gaseosa, y sea μ la media de X . En este caso la hipótesis nula que deseamos desacreditar es $H_0 : \mu = 125$ contra la alternativa $H_1 : \mu > 125$.

Observemos que $\bar{X} = 131$, de modo que $\bar{X} - 125 = 6$ ¿Es esta diferencia lo suficientemente grande como para rechazar H_0 ?

Decidimos usar como referencia una normal $X \sim N(\mu, \sigma^2)$. Si H_0 fuera cierta, $X \sim N(125, \sigma^2)$, y por lo tanto $\bar{X} \sim N(125, \sigma^2/n)$ y $\sqrt{n}(\bar{X} - 125)/\sigma \sim N(0, 1)$.

Si conociéramos σ , podríamos calcular la probabilidad de que $\sqrt{n}(\bar{X} - 125)/\sigma \geq 6\sqrt{n}/\sigma$ y rechazar H_0 si esta probabilidad es menor que el nivel preestablecido α . Como no conocemos σ , debemos buscar algún criterio alternativo que no dependa de esta cantidad. Pero como se comentó anteriormente,

$$t = \sqrt{n}(\bar{X} - 125)/S \sim t_{n-1}$$

independientemente del valor de σ . Esto nos permite obtener un criterio independiente de σ : rechazaremos H_0 si $t \geq t_{9, \alpha}$

Finalmente, calculando $S^2 = 382$, obtenemos $t = 6\sqrt{10}/382 = 0.77$. Si decidimos usar $\alpha = 0.05$, $t_{9, 0.05} = 1.833$, $t < t_{9, \alpha}$ y no puedo rechazar la hipótesis nula.