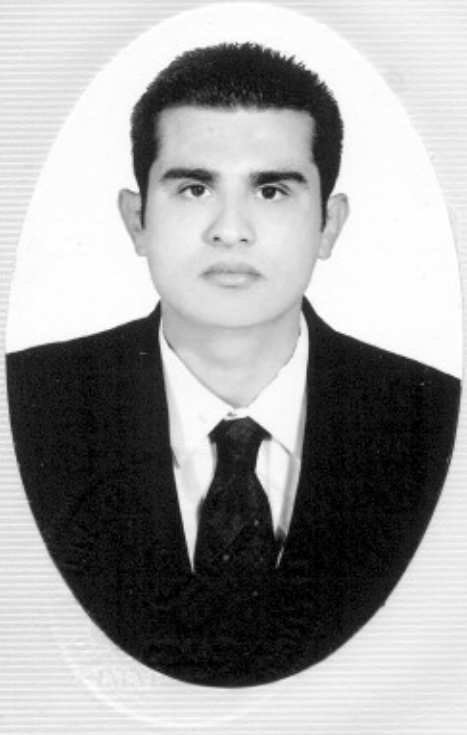
Centro de Investigación en Matemáticas, A.C.

Acta de Examen de Grado

Acta No.: 013

Libro No.: 002

Foja No.: 013



En la Ciudad de Guanajuato, Gto., siendo las 9:30 horas del día 31 de octubre del año 2008, se reunieron los miembros del jurado integrado por los señores:

DR. JOAQUÍN ORTEGA SÁNCHEZ (CIMAT)
DR. JOSÉ MIGUEL PONCIANO CASTELLANOS (CIMAT)
DRA. ELOÍSA DÍAZ-FRANCÉS MURGUÍA (CIMAT)

bajo la presidencia del primero y con carácter de secretario el segundo, para proceder a efectuar el examen que para obtener el grado de

MAESTRO EN CIENCIAS
CON ESPECIALIDAD EN PROBABILIDAD Y ESTADÍSTICA

sustenta

FRANCISCO JAVIER RUBIO ÁLVAREZ

en cumplimiento con lo establecido en los reglamentos y lineamientos de estudios de posgrado del Centro de Investigación en Matemáticas, A.C., mediante la presentación de la tesis

"MODELACIÓN ESTADÍSTICA DEL COCIENTE DE MEDIAS β DE DOS VARIABLES ALEATORIAS NORMALES".

Los miembros del jurado examinaron alternadamente al (la) sustentante y después de deliberar entre sí resolvieron declararlo (a):

Aprobado

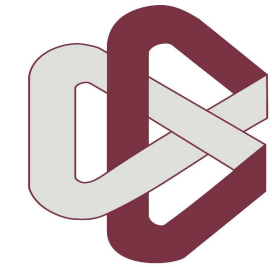
DR. JOAQUÍN ORTEGA SÁNCHEZ
Presidente

DR. JOSÉ MIGUEL PONCIANO CASTELLANOS
Secretario

DRA. ELOÍSA DÍAZ-FRANCÉS MURGUÍA
Vocal

CIMAT
DIRECCION GENERAL

DR. OSCAR ADOLFO SÁNCHEZ VALENZUELA
Director General



CIMAT

Centro de Investigación en Matemáticas A.C.

Modelación estadística del cociente de medias β de dos variables aleatorias normales

T E S I S

Que para obtener el título de:

Maestría en Ciencias con orientación en Probabilidad y Estadística

P r e s e n t a :

Francisco Javier Rubio Alvarez

Directora:

Dra. Eloísa Díaz-Francés Murguía

Guanajuato, Guanajuato, Octubre de 2008

Contenido

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología por el apoyo económico recibido durante dos años para la realización de mis estudios de maestría.

Al Centro de Investigación en Matemáticas A. C. por el financiamiento para la conclusión de esta tesis.

A Eloísa, por sus valiosos comentarios, el tiempo dedicado y su paciencia en la dirección y revisión de este trabajo.

A D. A. Sprott por sus comentarios sobre esta tesis y por inculcar en todos sus cursos la buena costumbre de cuestionar, discutir y proponer.

A los profesores del área de Probabilidad y Estadística del Centro de Investigación en Matemáticas por su enseñanza.

A mi familia por su apoyo incondicional a lo largo de mi carrera.

Prefacio

En diversas áreas de aplicaciones, pero de manera especial en la Citometría de Flujo para estimar el ADN nuclear de una planta, resulta de interés estimar la razón de medias β de dos variables aleatorias normales, X y Y, cuando se cuenta con observaciones pareadas de dichas variables. También interesa comparar si k grupos o poblaciones distintas comparten un mismo valor de β, y de hecho es importante poder cuantificar las diferencias que haya.

Debido seguramente a una confusión matemática que aún prevalece, en la literatura de Citometría se procede a analizar la razón $Z = X/Y$ como si también fuese distribuida como una variable aleatoria normal cuya media es precisamente el parámetro de interés β con el cual se estima el ADN de la planta de interés, Dolězel (1998), Palomino (1999). Seguramente piensan que se trata de una manera válida de simplificar la dimensión del problema, pues ya solamente hay que tratar con una variable aleatoria en vez de dos. Así, para comparar el parámetro β entre k grupos distintos de interés (por ejemplo plantas jóvenes y adultas ó varias posiciones en una planta dada), proceden a aplicar un Análisis de Varianzas (ANOVA) para evaluar la hipótesis nula de igualdad de estos parámetros β en todos los grupos. En el caso de no rechazar la hipótesis nula, entonces frecuentemente proceden a juntar todas las observaciones de los k grupos y las tratan como una muestra única de tamaño n, $z_1, ..., z_n$ (corresponde un valor z_i para cada par de observaciones x_i, y_i para $i = 1, ..., n$). Entonces proceden a estimar β de manera puntual con el estimador habitual $\overline{z} = \sum_{j=1}^{n} z_j/n$. En contraste, bajo el modelo normal conjunto asociado para (X, Y), en el caso de independencia, el estimador de

máxima verosimilitud es $\widehat{\beta} = \overline{x}/\overline{y}$, el cual no es necesariamente igual a $\overline{z}$.

Este error de aplicación en un área científica importante fue lo que motivó el presente trabajo. Si bien es cierto que la densidad de Z tiene como un parámetro a β y que su forma puede ser acampanada y simétrica bajo ciertas condiciones en los otros dos parámetros que la rigen, también es cierto que esta densidad no es normal y que en ocasiones puede ser marcadamente asimétrica e incluso bimodal. Es un hecho que Z no tiene ningún momento finito y sólo por eso parecería absurdo aproximar su distribución con la normal que sí tiene todos sus momentos finitos. Por tanto, es importante valorar el error matemático que se comete al suponer una distribución normal para Z con el fin de estimar β.

A partir del error matemático, asumir una distribución normal para Z también supone un error en las inferencias sobre β. En este trabajo se intenta cuantificar el error en las inferencias de β que han cometido los científicos en la literatura de Citometría de Flujo y otras áreas científicas que analizan las razones z_i de datos pareados, asignándoles una distribución normal. También se desea aquí presentar una forma alternativa eficiente de estimar β a través de su verosimilitud perfil a partir del modelo normal conjunto asociado a las variables X y Y. En contraste con realizar un ANOVA para Z, las gráficas de las perfiles de β para cada grupo de los k de interés permiten comparar la información contenida en los datos sobre estos parámetros $\beta_i,\ i = 1, ..., k$, posiblemente diferentes para cada grupo, de manera mucho más clara y útil. No es necesario que se cumpla el supuesto de normalidad para Z, sólo es necesario que sea razonable este supuesto para X y Y. Tampoco se necesita que la varianza sea la misma en todos los k grupos como requieren los supuestos del ANOVA. Más aún, las muestras pueden ser muy pequeñas, incluso de dos parejas de observaciones por grupo como se mostrará en un ejemplo en el Capítulo 3.

De manera adicional, se caracteriza en este trabajo la relación entre la forma de la densidad de Z y los valores de sus parámetros. Así, se muestra

bajo qué condiciones en los parámetros es que una cierta distribución normal puede aproximar razonablemente bien a la distribución de Z en un intervalo centrado en la moda de la densidad. También se muestra bajo cuales condiciones es que las inferencias sobre β hechas a partir de la densidad de Z serán similares a las hechas con el modelo conjunto de X y Y, que serían las correctas.

Al utilizar la densidad de Z como medio para estimar β, se tiene una asignación implícita de los roles que juegan las variables X y Y, en términos de cuál de ellas es la del numerador y cuál es la del denominador del cociente. En este trabajo se analiza también el efecto de intercambiar los roles de X y Y en las inferencias sobre β al utilizar la densidad de Z, la aproximación normal a ella y también bajo el modelo conjunto de X y Y. Por consideraciones lógicas, lo que se infiera sobre β debería ser equivalente a lo que se infiera sobre $1/\beta$ ya que es una reparametrización uno a uno de β. Esta comparación tiene como objetivo analizar la invarianza de las inferencias sobre β ante reparametrizaciones de estos modelos.

En el Capítulo 1 de esta tesis se muestra una reseña histórica de las expresiones y caracterizaciones que se han hecho para la densidad de Z. También en este capítulo se muestra cuándo la distribución de Z bajo ciertas condiciones en los parámetros de las densidades de X y Y, específicamente en sus coeficientes de variación, puede aproximarse bien con una distribución normal.

En el Capítulo 2 se presenta cómo estimar los parámetros de los distintos modelos considerados en el Capítulo 1 utilizando un enfoque de verosimilitud. Además se comparan para varios ejemplos, las dos aproximaciones normales contempladas para la distribución de Z.

Se presenta un análisis comparativo de las inferencias sobre β obtenidas a partir de las verosimilitudes perfil de β bajo los distintos modelos considerados.

En el Capítulo 3 se presenta la noción de la invarianza de las inferencias

de β ante el cambio de roles de X y Y.

Un resultado central de esta tesis es que se mostró que una aproximación normal a la distribución de Z solamente será razonable si los coeficientes de variación de X y Y son ambos menores a 0.1. Sin embargo, aún en ese caso los intervalos de estimación para β con la aproximación normal para Z que usan en diversas áreas pueden ser en exceso angostos si la muestra de observaciones pareadas (x_i, y_i) $i = 1, ..., n$ es pequeña.

Se recomienda utilizar la verosimilitud perfil de β bajo el modelo adecuado para hacer inferencia en términos de intervalos de verosimilitud-confianza. En el caso de contar con muestras provenientes de k poblaciones, se recomienda graficar juntas las verosimilitudes perfiles de β para los diversos k grupos de interés y no solamente dar estimadores puntuales de β.

En el caso particular de Citometría de Flujo, la perfil de β debe calcularse como se sugiere en Díaz-Francés y Sprott (2001) debido a la intervención física del citometrista en la obtención de los datos que ocasiona que haya muchos más parámetros de estorbo que en los modelos considerados en esta tesis.

Desgraciadamente, los científicos que trabajan en Citometría no están al tanto de que el análisis estadístico que hacen simplifica los supuestos que deben considerarse y por tanto puede ser incorrecto. Consecuentemente, no les ha interesado cuantificar el error que pueden cometer. Finalmente, se desea resaltar aquí, que las gráficas de la verosimilitud perfil de β (bajo el modelo adecuado) son mucho más informativas para comparar grupos que las estimaciones puntuales de β por sí solas u otros métodos utilizados.

Todos los cálculos hechos en esta tesis fueron realizados en una computadora personal con procesador Pentium IV. Para los cálculos y la graficación se utilizó Mathematica V. 5.0, Matlab V. 7.0 y R.

Capítulo 1

Modelos para la inferencia estadística del cociente de medias β de dos variables aleatorias normales X y Y

1.1 Introducción

En este capítulo se describen tres modelos estadísticos que se han considerado para hacer inferencias sobre el cociente de medias $\beta = E[X]/E[Y]$ de dos variables aleatorias normales e independientes X y Y cuando se tienen datos pareados. Las inferencias se harán bajo un enfoque de verosimilitud. Además se presentan aproximaciones a la distribución de $Z = X/Y$.

En la Sección 1.2 se presentan los antecedentes históricos sobre la densidad del cociente $Z = X/Y$. En la Sección 1.3, se presenta el modelo conjunto asociado a (Z, Y). En la Sección 1.4 se presenta un modelo marginal de Z que surge de condicionar el modelo conjunto de la Sección 1.3. En la Sección 1.5 se considera el caso particular de contar con datos pareados por razones experimentales. En particular es de interés comparar el valor de β en k grupos poblacionales distintos; se valora aquí lo adecuado de un Análisis de Varianzas para tal fin. Adicionalmente se presentan aquí las aproximaciones al modelo

marginal Z. La primera es una aproximación normal descrita en la literatura de Citometría de Flujo, la segunda es una aproximación propuesta por David Hinkley (1969) y la tercera es una aproximación normal que se propone en esta tesis por primera vez.

1.2 Antecedentes históricos

A continuación se dará una reseña histórica sobre el análisis de la densidad de $Z = X/Y$, el cociente de dos variables aleatorias normales y también sobre la inferencia del cociente $\beta = E[X]/E[Y]$ de sus medias.

Karl Pearson (1910) publicó fórmulas para la aproximación numérica de la densidad de la variable Z, para el caso en que los coeficientes de variación[1] de X y Y sean pequeños. Si X y Y están correlacionadas, Pearson mencionó que estas fórmulas de aproximación no son manejables en la práctica. Esta era una técnica importante en la época debido a la ausencia de computadoras ya que facilitaba los cálculos y la graficación de densidades.

A. S. Merrill (1928) presentó una fórmula para aproximar numéricamente la densidad de Z e indicó que existen casos en los que esta densidad tiene forma similar a una normal.

C. C. Craig (1929) publicó fórmulas para aproximar la densidad de la variable Z basadas en métodos diferentes a los usados por Pearson (1910) y además realizó un estudio comparativo con estas fórmulas.

R. C. Geary (1930) mostró una transformación para normalizar a la variable Z. Además indicó que esta transformación se puede utilizar en el caso en que la media de la variable Y sea al menos tres veces mayor que la desviación estándar de Y, lo cual es equivalente a pedir que el coeficiente de variación de Y fuera menor que $1/3$.

E. C. Fieller (1932) calculó la densidad de Z mediante la marginalización

[1]El coeficiente de variación de una variable aleatoria X se define como el cociente de la desviación estándar entre la media de esa variable $\sqrt{Var(X)}/E(X)$.

de la densidad conjunta de Z y Y. La expresión final contenía una integral definida. Esta misma densidad conjunta de Z y Y será considerada en la Sección 1.3 de esta tesis.

J. H. Curtiss (1941) demostró la existencia de la función de densidad de la variable Z mostrando primero que la función de distribución de Z es absolutamente continua. También calculó la densidad de Z en el caso en que $E[X] = 0$ y $E[Y] = 0$. Además remarcó que es aparentemente imposible evaluar la densidad de Z en forma cerrada para el caso general.

C. C. Craig (1942) calculó la densidad de Z para el caso general en que X y Y estén correlacionadas. La densidad calculada no es una expresión cerrada ya que está en términos de una integral definida. La expresión obtenida por Craig es más compacta que la expresión obtenida por Fieller (1932), pero de igual manera, está expresada en términos de una integral definida.

G. Marsaglia (1965) realizó un análisis descriptivo de la densidad de Z, mostrando gráficas de los casos en que la densidad de Z es simétrica, asimétrica, con colas pesadas e incluso bimodal.

D. V. Hinkley (1969) calculó la función de distribución y la densidad de Z para el caso general y propuso una aproximación a la densidad de Z. La aproximación propuesta es el límite uniforme de la densidad de Z cuando el coeficiente de variación de la variable Y converge a cero. En la Sección 1.5.3 se describirá con mayor detalle esta aproximación.

J. Hayya et. al. (1975) propusieron una aproximación normal a una transformación $T(Z)$ de la variable Z; actualmente se le conoce a $T(Z)$ como la transformación de Geary-Hinkley. La idea era utilizar una aproximación de segundo orden mediante una serie de Taylor de la media y la varianza de Z y aproximar a $T(Z)$ con una normal con esta media y varianza. Concluyeron con la recomendación, basados en simulaciones de Monte Carlo, de utilizar esta aproximación a la transformación $T(Z)$ cuando el coeficiente de variación de Y sea menor a 0.39 .

S. Shanmugalingam (1982) realizó un estudio de simulación en el cual con-

cluyó que la transformación logarítmica parece normalizar los datos provenientes de una variable Z asimétrica.

A. Cedilnik et. al. (2004) presentaron la densidad del cociente de dos normales como un producto de dos factores, el primero era una densidad Cauchy y el segundo, una función "*complicada*", según sus palabras. El principal objetivo de esta publicación era describir la forma de la densidad basados en un parámetro de forma definido en ese mismo artículo. Además presentaron una aplicación de esta densidad a un problema de regresión.

T. Pham-Gia et. al. (2006) dieron la expresión de la densidad de Z en términos de la función de Hermite y de la función Hipergeométrica Confluente.

El problema sobre la inferencia del cociente de medias β de dos variables aleatorias normales ha sido analizado recientemente desde distintos paradigmas estadísticos.

E. C. Fieller (1940) presentó un método para hacer inferencias sobre el cociente de medias de dos poblaciones normales mediante una cantidad pivotal la cual es conocida actualmente con el nombre de Pivotal de Fieller.

M. Mendoza y E. Gutiérrez-Peña (1999) publicaron un trabajo sobre la inferencia de β desde el punto de vista Bayesiano e hicieron un análisis comparativo con la inferencia pivotal que se deriva del pivotal de Fieller. Una conclusión de este artículo es que los intervalos de probabilidad obtenidos con el método Bayesiano son más cortos que los intervalos de confianza obtenidos con el pivotal de Fieller.

D. A. Sprott (2000) mostró cómo realizar inferencias sobre el cociente de dos parámetros de localización para el caso de datos pareados provenientes de un modelo de localización y escala. Este es un caso más general que el aquí considerado.

Kuethe, et. al. (2000) presentaron la densidad marginal de la variable $Z = X/Y$ como una opción para hacer inferencias sobre el cociente de medias β en aplicaciones a fisiología.

E. Díaz-Francés y D. A. Sprott (2003), para uno de los modelos conside-

rados por Mendoza y Gutiérrez-Peña, el modelo de Cox, presentaron cómo hacer inferencia sobre β con un enfoque de verosimilitud. En este artículo mostraron que el pivotal de Fieller no contiene necesariamente toda la información sobre el cociente de medias y que constituye tan sólo un factor de la función de verosimilitud correspondiente bajo este modelo. Por ello, para hacer inferencias sobre β hay que considerar a la verosimilitud completa y al hacerlo, éstas pueden ser tan eficientes como alternativas Bayesianas adecuadas.

C. G. Qiao et. al (2006) señalaron que la esperanza del cociente $Z = X/Y$ de dos variables normales independientes no existe. Sin embargo, mencionaron que la media muestral de Z puede usarse para estimar β cuando el coeficiente de variación de Y sea menor que 0.2. Realizaron un estudio comparativo de dos estimadores $\overline{X}/\overline{Y}$ y $\overline{Z}$, del cual concluyen que es preferible utilizar $\overline{X}/\overline{Y}$ para estimar β si X y Y son independientes. Además recomendaron utilizar el estimador $\overline{Z}$ sólo cuando los datos disponibles sean únicamente los cocientes individuales $z_i = x_i/y_i$ y que se desconozcan los valores (x_i, y_i) para $i = 1, ..., n$.

F. J. Rubio (2008) en un reporte técnico mostró que en el artículo de Mendoza y Gutiérrez-Peña (1999) había un error numérico cuya consecuencia es que el intervalo de probabilidad con el método Bayesiano es más largo que el intervalo de confianza obtenido con el pivotal de Fieller, contrario a la conclusión presentada en el artículo de Mendoza y Gutiérrez-Peña.

En la literatura estadística de Citometría de Flujo, hay quienes utilizan la variable aleatoria $Z = X/Y$ para estimar β, ya que la densidad de esta variable depende también del parámetro de interés β. Kuethe, et. al. (2000) también notaron este hecho en otro contexto. En algunos trabajos como Dolezel et. al. (1998), Lysák (1998), Palomino et. al. (1999), se ha hecho el supuesto de que esta variable Z tiene distribución normal. Es bien conocido que si marginalmente X y Y son normales, el cociente tendrá una distribución G_Z distinta a la normal y ésta se dará más adelante en la Sección 1.4 de manera explícita. Sin embargo, debido al amplio uso de Z para hacer inferencia

sobre β en campos como la Citometría de Flujo y otros como el Análisis de Varianzas, es importante cuantificar el tipo de errores que se puede cometer en la inferencia de β al usar una aproximación normal a la distribución de Z. En esta tesis también se desea mostrar cuál es la manera más eficiente de estimar β.

1.3 Modelo conjunto para X y Y

Sean X y Y variables aleatorias independientes con distribución normal, $X \backsim N(\mu_x, \sigma_x)$ y $Y \backsim N(\mu_y, \sigma_y)$, donde $\mu_x > 0$, $\mu_y > 0$, $\sigma_x > 0$, $\sigma_y > 0$. Nótese que se considera sólo el caso de medias positivas, el cual es razonable en muchas situaciones en la naturaleza.

La densidad conjunta es:

$$\begin{aligned} f_{X,Y}(x, y; \mu_x, \mu_y, \sigma_x, \sigma_y) &= f_X(x; \mu_x, \sigma_x) f_Y(y; \mu_y, \sigma_y) \\ &= \frac{1}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left[\frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right]\right\}. \end{aligned} \tag{1.1}$$

El parámetro de interés es la razón de medias $\beta = \mu_x/\mu_y$. Este modelo se puede expresar en términos de $Z = X/Y$ y Y, haciendo un cambio de variable como se muestra enseguida. Para cuantificar el error cometido al utilizar la variable Z para estimar β conviene hacer el siguiente cambio de variable de (X, Y) a (Z, Y^*) para obtener la distribución conjunta de las variables Z y Y^*, donde la relación entre las variables es:

$$\begin{aligned} Z &= X/Y, \\ Y^* &= Y. \end{aligned}$$

El Jacobiano o valor absoluto del determinante de la matriz de primeras derivadas para el cambio de variables es:

$$|\jmath| = \begin{vmatrix} \frac{\partial x}{\partial y^*} & \frac{\partial x}{\partial z} \\ \frac{\partial y^*}{\partial y^*} & \frac{\partial y^*}{\partial z} \end{vmatrix} = \begin{vmatrix} z & y \\ 1 & 0 \end{vmatrix} = |-y| = |y|.$$

Como $Y^*= Y$, de aquí en adelante se hará mención a la densidad conjunta de (Z, Y) en vez de la de (Z, Y^*) por sencillez. Esta densidad aplicando el Teorema de Cambio de Variable se calcula de la siguiente manera:

$$\begin{aligned} g_{Z,Y}(z, y; \mu_x, \sigma_x, \mu_y, \sigma_y) &= f_X(yz; \mu_x, \sigma_x) f_Y(y; \mu_y, \sigma_y)|y| \\ &= \frac{|y|}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{(zy-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right)\right]. \end{aligned}$$

Expandiendo los binomios y sumando:

$$g_{Z,Y}(z, y; \mu_x, \sigma_x, \mu_y, \sigma_y) = \frac{|y|}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left[\frac{\sigma_y^2(z^2y^2 + \mu_x^2 - 2zy\mu_x) + \sigma_x^2(y^2 + \mu_y^2 - 2y\mu_y)}{\sigma_x^2\sigma_y^2}\right]\right\};$$

factorizando y y y^2 :

$$\begin{aligned} g_{Z,Y}(z, y; \mu_x, \sigma_x, \mu_y, \sigma_y) =& \frac{|y|}{2\pi\sigma_x\sigma_y} \times \\ \times \exp&\left\{-\frac{1}{2}\left[\frac{y^2(z^2\sigma_y^2 + \sigma_x^2) - y(2z\mu_x\sigma_y^2 + 2\mu_y\sigma_x^2) + (\mu_x^2\sigma_y^2 + \mu_y^2\sigma_x^2)}{\sigma_x^2\sigma_y^2}\right]\right\}. \end{aligned} \tag{1.2}$$

Se definirán los siguientes parámetros $\beta, \rho, \delta_y, \delta_x$, que tienen un significado importante: $\beta = \frac{\mu_x}{\mu_y}$ es el cociente de medias, el cual es el parámetro de interés; $\delta_y = \frac{\sigma_y}{\mu_y}$ es el coeficiente de variación de la variable Y; $\rho = \frac{\sigma_y}{\sigma_x}$ es el cociente de la desviación estándar de Y entre la desviación estándar de X; finalmente se definirá el coeficiente de variación de X como $\delta_x = \frac{\sigma_x}{\mu_x}$. Como veremos en la sección (1.3.3) δ_y servirá para determinar si la aproximación normal a la variable Z es razonable. Conviene realizar la siguiente reparametrización de la densidad $g(z, y)$:

$$(\mu_x, \sigma_x, \mu_y, \sigma_y) \rightleftarrows (\beta, \delta_y, \rho, \sigma_x^*).$$

Esta reparametrización es uno a uno y su transformación inversa es:

$$\begin{aligned}
\sigma_x &= \sigma_x^*, \\
\sigma_y &= \rho\sigma_x^*, \\
\mu_y &= \frac{\rho}{\delta_y}\sigma_x^*, \\
\mu_x &= \frac{\beta\rho}{\delta_y}\sigma_x^*.
\end{aligned}$$

Para simplificar notación se utilizará σ_x en lugar de σ_x^*. El objetivo de esta reparametrización es utilizar la densidad $g(z, y)$ para hacer inferencias sobre el parámetro de interés β.

Multiplicando el cociente del argumento de la exponencial por $(1/\sigma_x^2)/(1/\sigma_x^2)$ y utilizando esta reparametrización se tiene de (1.2) que:

$$\begin{aligned}
g_{Z,Y}(z, y; \beta, \delta_y, \rho, \sigma_x) &= \frac{|y|}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left[\frac{y^2(z^2\rho^2+1) - y(2z\mu_x\rho^2 + 2\mu_y) + (\mu_x^2\rho^2 + \mu_y^2)}{\sigma_y^2}\right]\right\} \\
&= \frac{|y|}{2\pi\sigma_x\sigma_y} \exp\left\{-\frac{1}{2}\left[\frac{(\mu_x^2\rho^2 + \mu_y^2)}{\sigma_y^2}\right]\right\} \times \\
&\quad \times \exp\left\{-\frac{1}{2}\left[\frac{y^2(z^2\rho^2+1) - y(2z\mu_x\rho^2 + 2\mu_y)}{\sigma_y^2}\right]\right\}.
\end{aligned}$$

Multiplicando por $(1/\mu_y^2)/(1/\mu_y^2)$ en el argumento de la primera exponencial y por $(1/\mu_y)/(1/\mu_y)$ el segundo término de la segunda exponencial se tiene que:

$$\begin{aligned}
&g_{Z,Y}(z, y; \beta, \delta_y, \rho, \sigma_x) = \\
&= \frac{|y|}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{(\beta^2\rho^2+1)}{2\delta_y^2}\right] \exp\left\{-\frac{1}{2}\left[\frac{y^2}{\sigma_y^2}(z^2\rho^2+1) - \frac{2y}{\sigma_y}\left(\frac{z\beta\rho^2+1}{\delta_y}\right)\right]\right\}.
\end{aligned}$$

La simplificación final se obtiene multiplicando σ_y^2 por σ_x^2/σ_x^2 y σ_y por σ_x/σ_x :

$$\begin{aligned}
&g_{Z,Y}(z, y; \beta, \delta_y, \rho, \sigma_x) = \\
&= \frac{|y|}{2\pi\sigma_x^2\rho} \exp\left[-\frac{(\beta^2\rho^2+1)}{2\delta_y^2}\right] \exp\left\{-\frac{1}{2}\left[\frac{y^2}{\sigma_x^2\rho^2}(z^2\rho^2+1) - \frac{2y}{\sigma_x\rho}\left(\frac{z\beta\rho^2+1}{\delta_y}\right)\right]\right\}. \quad (1.3)
\end{aligned}$$

Nótese que (1.3) es una función de los parámetros $(\beta, \delta_y, \rho, \sigma_x)$; a esta densidad se le llamará aquí la densidad del "Modelo Conjunto de Y y Z".

1.4 Modelo marginal para $\mathbf{Z = X/Y}$

La densidad de la variable $Z = X/Y$ puede obtenerse marginalizando la densidad conjunta $g(z, y)$ dada en (1.3) :

$$g_Z(z) = \int_{-\infty}^{\infty} g_{Z,Y}(z, y)\, dy.$$

Consecuentemente; la densidad conjunta (1.3) puede factorizarse como:

$$g_{Z,Y}(z, y; \beta, \rho, \delta_y, \sigma_x) = g_Z(z; \beta, \delta_y, \rho) g_{Y|Z}(y|z; \beta, \rho, \sigma_x, \delta_y). \tag{1.4}$$

En este trabajo se desea explorar la validez del uso de la densidad marginal $g_Z(z)$ para hacer inferencia sobre β. La factorización (1.4) puede interpretarse como una separación de la información contenida en la muestra en dos partes: la información dada por Z y la información de Y condicional en Z. Nótese que la distribución condicional $g_{Y|Z}(y|z; \beta, \rho, \sigma_x, \delta_y)$ depende de los cuatro parámetros y en especial puede contener información relevante sobre el parámetro de interés β. Debido a esto, no sería óptimo hacer inferencia sobre β sólo con g_Z sin cuantificar la información sobre este parámetro que esté contenida en $g_{Y|Z}$. Por otro lado, la densidad g_Z del cociente $Z = X/Y$ depende solamente de tres parámetros (β, δ_y, ρ).

Integrando en el caso que $\mu_x > 0$ y $\mu_y > 0$ se tiene que (ver apéndice A1):

$$\begin{aligned} g_Z(z; \beta, \delta_y, \rho) &= \exp\left(-\frac{1+\beta^2\rho^2}{2\delta_y^2}\right) \frac{\rho}{\pi(1+\rho^2 z^2)} \\ &\quad + \frac{\rho(1+\beta\rho^2 z)}{\sqrt{2\pi}\delta_y(1+\rho^2 z^2)^{3/2}} \exp\left[-\frac{\rho^2(z-\beta)^2}{2\delta_y^2(1+\rho^2 z^2)}\right] \operatorname{erf}\left[\frac{1+\beta\rho^2 z}{\delta_y\sqrt{2(1+\rho^2 z^2)}}\right], \end{aligned} \tag{1.5}$$

donde la función erf es la función de error:

$$erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\, dt.$$

La función erf es continua en $\mathbb{R}$ y su imagen es el intervalo $(-1, 1)$. La gráfica de la función erf es la siguiente:

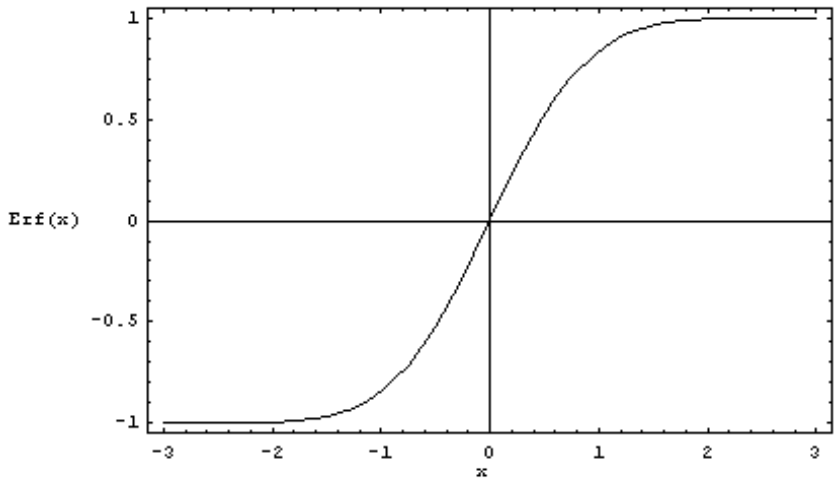


Figura 1.1: Gráfica de la función erf.

En el apéndice A2 se muestran varias propiedades de la función erf, además de una aproximación útil para su cálculo numérico.

Dependiendo de los valores de los parámetros, principalmente el coeficiente de variación δ_y, la densidad de Z dada en (1.5) puede ser asimétrica o incluso bimodal. Se resalta que la variable aleatoria Z no tiene momentos finitos. En el caso en que $\mu_x = \mu_y = 0$ y $\sigma_x = \sigma_y = 1$, la densidad de Z es la densidad de una variable con distribución Cauchy estándar.

En la siguiente figura se muestran varias gráficas de la densidad de Z para diferentes valores de los parámetros:

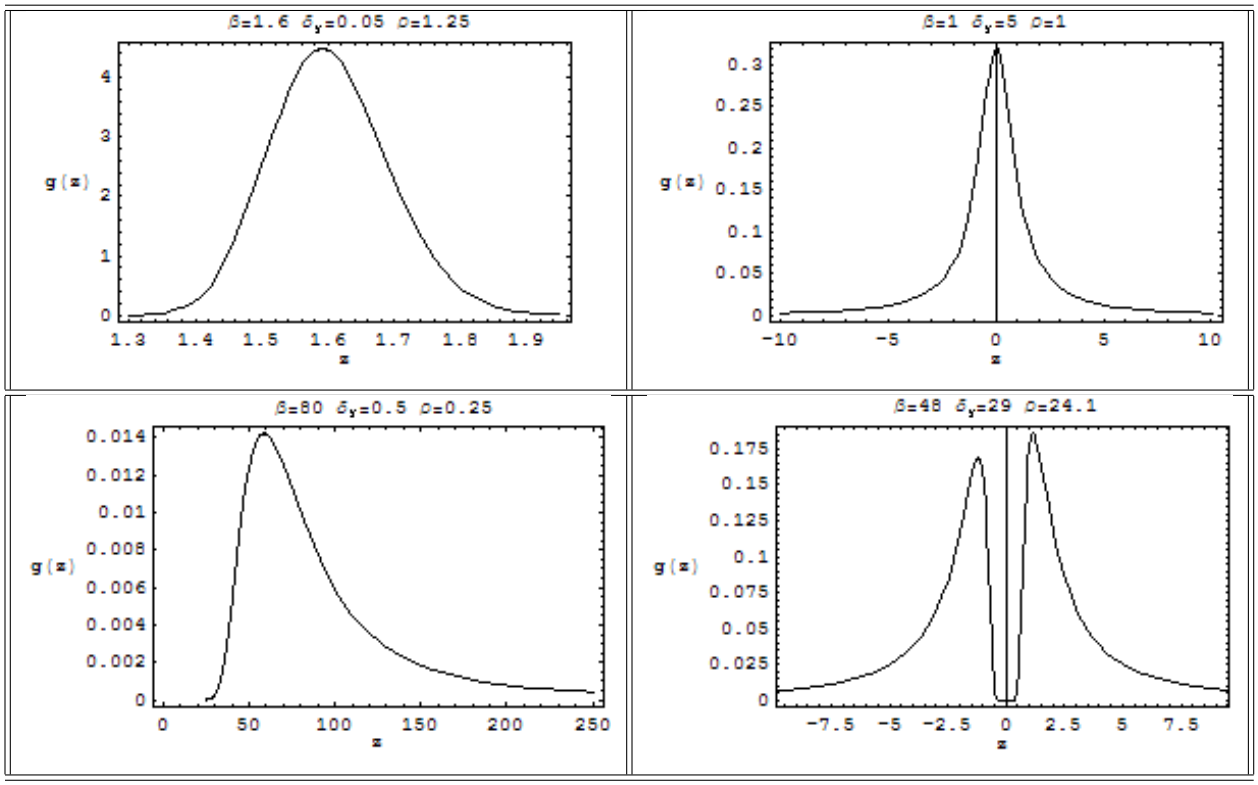


Figura 1.2: Densidad $g_Z(z)$ para diferentes valores de los parámetros.

Como se observa, la forma de g_Z puede ser simétrica, asimétrica e incluso bimodal, pero en ningún caso la variable Z tiene momentos finitos.

1.5 El caso especial de datos pareados

En esta sección se presentan tres aproximaciones a la densidad marginal g_Z, para el caso particular de tener datos pareados. La primera es una aproximación normal estimada propuesta en la literatura de Citometría de Flujo; la segunda es una aproximación publicada por David Hinkley (1969) y la tercera es una aproximación normal que se propone por primera vez en esta tesis. Además se describe el ANOVA de una vía debido a que en Citometría de Flujo lo usan para comparar el ADN de plantas, que es equivalente al parámetro β, entre k grupos.

Sea $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ una muestra independiente de variables aleatorias normales, pareadas e independientes. Así para cada pareja (x_i, y_i) existe

un valor $z_i = x_i/y_i$ para $i = 1, ..., n$. El hecho de que sean pareados se debe a razones físicas y experimentales. Esto ocasiona que se tengan n parámetros μ_i adicionales puesto que se supone entonces que $x_i \backsim N(\beta\mu_i, \sigma_x)$ y $y_i \backsim N(\mu_i, \sigma_y)$ para $i = 1, ..., n$. Es decir, las n parejas comparten una razón de medias común a todas ellas.

1.5.1 Comparación de β en varios grupos a través del Análisis de Varianzas

A continuación se realizará una pequeña descripción del modelo para el Análisis de Varianzas (ANOVA) de una vía.

El modelo lineal que se considera para la variable de interés Z_{ij} cuando se tienen k grupos o poblaciones es :

$$Z_{ij} = \mu + \alpha_i + \epsilon_{ij}, \qquad \epsilon_{ij} \overset{iid}{\backsim} N(0, \sigma),$$

donde $i = 1, ..., k$ indica el grupo específico y el índice j indica la observación en la muestra correspondiente $j = 1, ..., n_i$. Es decir, el valor que toma Z depende de un valor medio μ que describe a todos los grupos, más un parámetro propio de localización de cada grupo específico α_i, más una perturbación normal con media cero y varianza constante σ^2. Así la media de Z en el grupo i toma el valor de $\mu + \alpha_i$.

El objetivo principal del ANOVA es verificar si las observaciones independientes de Z en los k grupos muestran evidencia en contra de la hipótesis nula:

$$H_0 : \alpha_1 = \alpha_2 = ... = \alpha_k = 0.$$

Es decir, se desea probar la hipótesis nula de igualdad de medias para los k grupos.

Por lo tanto los supuestos para el ANOVA son que Z_{ij} es normal en los k grupos y que tiene varianza igual a σ^2 en todos ellos. Ahora las observaciones

que se recopilan bajo este modelo son n_i observaciones para cada uno de los k grupos, Z_{ij}, donde $j = 1, ..., n_i$ y $i = 1, ..., k$. Es decir, sea z_{ij} la observación $j - ésima$ del grupo i. Se pueden descomponer las observaciones como la suma de tres términos:

$$z_{ij} = \overline{z}_{..} + (\overline{z}_{i.} - \overline{z}_{..}) + (z_{ij} - \overline{z}_{i.}), \tag{1.6}$$

donde:

$$\overline{z}_{i.} = \sum_{j=1}^{n_i} z_{ij}, \qquad \overline{z}_{..} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} z_{ij}.$$

Nótese que $\overline{z}_{..}$ es un estimador natural de la media grupal y que $(\overline{z}_{i.} - \overline{z}_{..})$ los es del parámetro α_i.

El término $(\overline{z}_{i.} - \overline{z}_{..})$ representa la estimación de la diferencia α_i de la media grupal $(\mu + \alpha_i)$ con respecto a la media total μ. El término $(z_{ij} - \overline{z}_{i.})$ representa la estimación de la diferencia de la observación ij-ésima con respecto a la media grupal, $(\mu + \alpha_i)$.

Ahora, consideremos la suma de los cuadrados de las discrepancias de las observaciones Z_{ij} con las medias grupales y la global (1.6), las cuales representan la variación muestral entre los grupos (SCE) y la variación muestral dentro de cada grupo (SCD):

$$\begin{aligned} SCT &= \sum_{i=1}^{k} \sum_{j=1}^{n_i} (z_{ij} - \overline{z}_{..})^2 . \\ SCE &= \sum_{i=1}^{k} n_i (\overline{z}_{i.} - \overline{z}_{..})^2 , \\ SCD &= \sum_{i=1}^{k} \sum_{j=1}^{n_i} (z_{ij} - \overline{z}_{i.})^2 . \end{aligned}$$

Se puede demostrar que la variación muestral total está dada por la suma de SCE y SCD, esto es:

$$SCT = SCE + SCD = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \left(z_{ij} - \overline{z}_{..}\right)^2 .$$

R. A. Fisher en 1925 demostró que el estadístico $F = SCE/\,SCD$, tiene distribución F de Fisher-Snedecor con $(n-k, k-1)$ grados de libertad si los k grupos tienen medias iguales, donde $n = \sum_{i=1}^{k} n_i$. Por lo tanto puede ser utilizado para probar la hipótesis de igualdad de medias. Fisher propuso este método como un algoritmo aritmético que permitía hacer los cálculos fácilmente. En esa época no se contaba con las capacidades gráficas de cómputo actuales y por tanto era difícil obtener gráficas de verosimilitud.

1.5.2 Aproximaciones a la distribución de Z

Primera aproximación normal a $g_Z(z)$

En esta sección se presenta la aproximación normal implícita en la literatura de Citometría de Flujo que se describe a continuación.

La Citometría de Flujo es una técnica de análisis celular que implica medir las características de dispersión de luz y fluorescencia que poseen las células de plantas y animales que se han teñido, conforme se las hace pasar a través de un rayo LASER. En particular, se obtienen datos pareados (x_j, y_j) de dos normales independientes, que corresponden a las mediciones del ADN de dos plantas, una de interés y otra de control que se trituran juntas, se preparan en solución, se tiñen y se pasan a través de un equipo electrónico llamado Citómetro de Flujo. El citometrista además controla la escala en la que se obtienen las mediciones para cada pareja, lo que da lugar a que sea razonable que las medias $\mu_i = E[X_i], i = 1, ..., n$ sean distintas. Estas son las razones físicas que da origen a que los datos sean pareados pero en citometría no se toman estos hechos en cuenta en los análisis estadísticos que se efectúan. En Doležel (1998), Lysák (1998) y Palomino (1999) se realiza un ANOVA a los cocientes x_j/y_j para diferentes especies de plantas, con lo cual están suponiendo implícitamente que el cociente de dos variables normales independientes sigue

una distribución normal. Luego usan a la media:

$$\overline{z} = \frac{1}{n}\sum_{j=1}^{n}\frac{x_j}{y_j},$$

como un estimador del cociente de medias β común a todos los grupos si no se rechazara la hipótesis nula, que β sea igual en todos los grupos.

Esto conlleva un error lógico importante: no rechazar la hipótesis nula no es equivalente a aceptarla. Es decir, a pesar de que no se rechace H_0, bien puede ser que las medias para Z en cada grupo discrepen ligeramente. De igual manera es importante valorar si el supuesto de varianzas iguales en los k grupos es razonable, así como el supuesto de la distribución normal de Z.

En contraste, se verá en el siguiente capítulo que al graficar la verosimilitud perfil de β para cada grupo en una misma gráfica, se pueden dar intervalos de estimación para β en cada grupo y así evaluar la magnitud de las discrepancias entre grupos si las hubiera.

Aquí se explorará la propuesta de usar una distribución normal como una aproximación para la densidad del cociente de dos normales independientes, con el fin de cuantificar el error que se comete al considerar que los cocientes $z_j = x_j/y_j$ son normales. El único supuesto que se hará es que $x_i \backsim N(\mu_x, \sigma_x)$ y $y_i \backsim N(\mu_y, \sigma_y)$ para $i = 1, ..., n$. La aproximación normal estimada $N(\overline{z}, s_z)$ que se usa en la literatura está asociada a una variable aleatoria normal con media $\overline{z}$ y varianza s_z^2 donde:

$$\begin{aligned} \overline{z} &= \frac{1}{n}\sum_{j=1}^{n}\frac{x_j}{y_j}, \\ s_z^2 &= \frac{1}{n}\sum_{j=1}^{n}\left(\frac{x_j}{y_j} - \overline{z}\right)^2, \end{aligned}$$

y n es el número de datos pareados (x_j, y_j). Aquí $\overline{z}$ y s_z^2 son los estimadores habituales de máxima verosimilitud para la media y la varianza respectivamente de una variable normal. La función de distribución y de densidad de esta variable normal son respectivamente:

$$F_N(w) = \Phi\left(\frac{w-\overline{z}}{s_z}\right), \tag{1.7}$$
$$f_N(w) = \frac{1}{s_z}\phi\left(\frac{w-\overline{z}}{s_z}\right),$$

donde Φ es la distribución de una normal estándar y ϕ la densidad de una normal estándar.

Aproximación de David Hinkley a g_z

David Hinkley (1969) propuso la siguiente aproximación a la función de distribución de Z :

$$F^*(w) = \Phi\left(\frac{w\mu_y - \mu_x}{\sigma_x\sigma_y\sqrt{\frac{1}{\sigma_y^2} + \frac{w^2}{\sigma_x^2}}}\right). \tag{1.8}$$

Esta aproximación tiene algunas propiedades de convergencia como se especifica en el siguiente resultado.

Teorema *1 (David Hinkley, 1969).* *Sean X y Y variables aleatorias independientes con distribución normal, $X \backsim N(\mu_x, \sigma_x)$ y $Y \backsim N(\mu_y, \sigma_y)$, donde $\mu_x \neq 0$, $\mu_y \neq 0$, $\sigma_x > 0$, $\sigma_y > 0$, sea F_Z la función de distribución de la variable $Z = X/Y$. Para $w \in \mathbb{R}$ fijo se define:*

$$F^*(w) : = P[X - wY \leq 0] = \Phi\left(\frac{w\mu_y - \mu_x}{\sigma_x\sigma_y\sqrt{\frac{1}{\sigma_y^2} + \frac{w^2}{\sigma_x^2}}}\right),$$

entonces $F_Z \xrightarrow{u} F^*$ cuando $\delta_y = \sigma_y/\mu_y \to 0$.

Es decir, cuando el coeficiente de variación δ_y tiende a cero, la distribución de Z converge uniformemente a la F^* propuesta en (1.8). Nótese que F^* no es una distribución normal, ni necesariamente simétrica. En el Apéndice A3 se da la demostración detallada de este teorema, completando y explicando los pasos que Hinkley presenta en su publicación.

Aproximación normal propuesta a g_z

La idea de esta propuesta surge de querer averiguar si existe una distribución normal que aproxime bien a la distribución de Z bajo condiciones específicas en los parámetros de las distribuciones de X y Y. Concretamente, se va a evaluar el caso en que δ_x y δ_y sean pequeños, tal que $\delta_y < 0.1$, $\sqrt{\delta_x^2 + \delta_y^2} < 1/3$ y que la densidad de Z sea simétrica con respecto a su moda. Para esto se utiliza la aproximación mediante la expansión en una serie de Taylor de la media y la varianza de la variable Z dada por Mood et. al. (1974) (ver el Apéndice A4) :

$$\begin{aligned} E[Z] &\approx \frac{\mu_X}{\mu_Y}, \qquad \text{y} \\ Var[Z] &\approx \left(\frac{\mu_X}{\mu_Y}\right)^2 \left(\frac{\sigma_x^2}{\mu_X^2} + \frac{\sigma_y^2}{\mu_Y^2}\right). \end{aligned}$$

Utilizando estas expresiones se propone en esta tesis la siguiente aproximación normal a la función de distribución de Z:

$$F(z) = \Phi\left[\frac{z - \mu_x/\mu_y}{\sqrt{\left(\sigma_x/\mu_y\right)^2 + \left(\mu_x/\mu_y\right)^2 \left(\sigma_y/\mu_y\right)^2}}\right]. \tag{1.9}$$

Si los parámetros se reemplazan por sus estimadores de máxima verosimilitud, se tendrá una versión estimada de esta aproximación normal.

Como se verá en varios ejemplos, esta aproximación teórica sigue más de cerca a la densidad de Z que la propuesta en Citometría pues toma en cuenta la dispersión de X y de Y por separado, en vez de considerar solamente la varianza de los cocientes Z.

Finalmente, se presenta un teorema que indica bajo cuáles condiciones en los parámetros es posible que exista una distribución normal que aproxime bien a la distribución del cociente $Z = X/Y$ para una variable aleatoria normal X fija. Por lo menos debe ocurrir que $\delta_x \in (0, 1/3)$ para que pueda existir tal variable Y, que a su vez deberá cumplir ciertas condiciones.

Teorema 2. *Sea X una variable aleatoria con distribución normal con media $\mu_x > 0$, varianza σ_x^2 y coeficiente de variación $\delta_x = \sigma_x / \mu_x$ tal que $0 < \delta_x < \lambda < 1$. Dado $\varepsilon \in (0,1)$, existen un valor $\gamma(\varepsilon) \in (0,\lambda)$ y una variable aleatoria normal Y, independiente de X con media $\mu_y > 0$, varianza σ_y^2 y coeficiente de variación $\delta_y = \sigma_y / \mu_y$, que cumple con:*

$$0 < \delta_y < \gamma(\varepsilon) < \sqrt{\lambda^2 - \delta_x^2} < \lambda,$$

y es tal que si $Z = X/Y$, entonces para z en el intervalo:

$$I = \left\{ \mu_x/\mu_y \pm \frac{1}{\lambda} \sqrt{\left(\sigma_x/\mu_y\right)^2 + \left(\mu_x/\mu_y\right)^2 \left(\sigma_y/\mu_y\right)^2} \right\} = \left\{ \beta \pm \frac{\beta}{\lambda} \sqrt{\delta_x^2 + \delta_y^2} \right\}, \tag{1.10}$$

se cumple que:

$$|F(z) - F_Z(z)| < \varepsilon, \tag{1.11}$$

donde F es es la función de distribución de una variable aleatoria normal con media β y desviación estándar $\beta\sqrt{\delta_x^2 + \delta_y^2}$,

$$\begin{aligned} F(z) &= \Phi\left[\frac{z - \mu_x/\mu_y}{\sqrt{\left(\sigma_x/\mu_y\right)^2 + \left(\mu_x/\mu_y\right)^2 \left(\sigma_y/\mu_y\right)^2}} \right] \\ &= \Phi\left[\frac{z - \beta}{\beta\sqrt{\delta_x^2 + \delta_y^2}} \right], \end{aligned} \tag{1.12}$$

y F_Z es la distribución de Z.

Es decir, este teorema muestra que dada una variable aleatoria X con coeficiente de variación menor a λ es posible encontrar otra variable aleatoria normal Y tal que $Z = X/Y$ se pueda aproximar bien por una distribución normal en un intervalo centrado en β cuando se cumplen las condiciones en los parámetros mencionadas. En el apéndice A5 se presenta la demostración de este teorema.

La utilidad de este resultado es demostrar formalmente que existirá una aproximación normal buena siempre que δ_y sea pequeño y que δ_x cumpla con las restricciones mencionadas, es decir que también sea pequeño. Cabe

resaltar que las justificaciones que se han dado en la literatura sobre la aproximación normal a la densidad de Z han sido empíricas o se han hecho mediante simulaciones de Monte Carlo. En este trabajo se muestra empíricamente a través de ejemplos que la aproximación normal a $g_Z(z)$ es razonablemente buena cuando ambos δ_x y δ_y sean menores a 0.1.

Capítulo 2

Inferencia sobre los parámetros de los modelos considerados para datos pareados

2.1 Introducción

En este capítulo se presentan las funciones de verosimilitud construidas a partir de los tres modelos estadísticos $g(z, y; \beta, \rho, \delta_y, \sigma_x)$, $g_Z(z; \beta, \rho, \delta_y)$, $f_N(z; \beta, \sigma_z)$ descritos en el capítulo anterior. También se presenta el cálculo de los estimadores de máxima verosimilitud a partir del modelo conjunto $g(z, y; \beta, \rho, \delta_y, \sigma_x)$ para varios ejemplos simulados y reales. Además, para los ejemplos simulados se compara la densidad $g_Z(z; \beta, \rho, \delta_y)$ con las aproximaciones normales $f_N(z; \overline{z}, s_z)$ y la densidad teórica propuesta a partir del Teorema 2 $f_{NP}(z; \beta, \delta_x, \delta_y)$.

2.2 Funciones de verosimilitud bajo los modelos considerados

Para cada uno de los tres modelos estadísticos mencionados en el Capítulo 1 $g(z, y; \beta, \rho, \delta_y, \sigma_x)$, $g_Z(z; \beta, \rho, \delta_y)$, $f_N(z; \beta, \sigma_z)$ se puede obtener la verosimilitud perfil

de β para hacer inferencias sobre este parámetro bajo cada modelo. Estas verosimilitudes permiten hacer inferencia por separado para β cuando se desconocen los parámetros restantes del modelo. Se compararán las inferencias que se hagan sobre β con estos tres modelos para varios ejemplos y se valorarán cuándo es razonable la aproximación normal que se ha usado en la literatura a la luz que se sabe que la perfil de β bajo el modelo conjunto es el patrón de referencia y también porque se conoce el valor teórico de β para los ejemplos simulados. A continuación se definen las verosimilitudes relevantes utilizando los diferentes modelos.

A partir del modelo $g(z, y)$ la verosimilitud para los cuatro parámetros $(\beta, \rho, \delta_y, \sigma_x)$ para una muestra dada $\underline{z}=(z_1, ..., z_n)$ y $\underline{y}=(y_1, ..., y_n)$ es:

$$L_{ZY}(\beta, \rho, \delta_y, \sigma_x; \underline{z}, \underline{y}) \propto \prod_{j=1}^{n} g(z_j, y_j; \beta, \rho, \delta_y, \sigma_x). \tag{2.1}$$

A partir del modelo $g_Z(z)$ la verosimilitud para los parámetros (β, ρ, δ_y) para una muestra dada $\underline{z}$ es:

$$L_Z(\beta, \rho, \delta_y; \underline{z}) \propto \prod_{j=1}^{n} g_Z(z_j, y_j; \beta, \rho, \delta_y). \tag{2.2}$$

Para la maximización numérica de esta función resulta útil la aproximación a la función erf mostrada en el Apéndice $A2$, sobre todo cuando el tamaño de muestra n es grande. Los estimadores para β, ρ y δ_y utilizando este modelo, serán diferentes en general a los estimadores obtenidos con el modelo conjunto (Z, Y), debido al factor adicional $g(Y|Z)$ en (1.4).

La función de verosimilitud correspondiente al modelo normal $f_N(z; \beta, \sigma)$ para los parámetros (β, σ) es:

$$L_N(\beta, \sigma_z; \underline{z}) \propto \prod_{j=1}^{n} f_N(z_j; \beta, \sigma_z) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_z} \exp\left[-\frac{(z_j - \beta)^2}{2\sigma_z^2}\right]. \tag{2.3}$$

Para los datos pareados (x_i, y_i) $i = 1, ..., n$, se verá que es preferible utilizar la verosimilitud conjunta (2.1) debido a que contiene toda la información sobre β

que está en la muestra, mientras que (2.2), (2.3) y sólo contienen la información sobre β contenida en la parte marginal de Z. Sin embargo se desea valorar el error que se comete al usar el modelo marginal $g_Z(z)$ para hacer inferencia sobre β puesto que este modelo se ha utilizado en diversas aplicaciones, así como valorar también el error que se comete al utilizar la aproximación normal (2.3) que también ha sido usada en la literatura de Citometría de Flujo.

Si tenemos una muestra de variables aleatorias independientes (x_1, y_1) ... (x_n, y_n) distribuidas como (1.1), entonces la función de verosimilitud para el modelo de X y Y conjunto es:

$$\begin{aligned} L(\mu_x, \sigma_x, \mu_y, \sigma_y; \underline{x}, \underline{y}) &\propto \frac{1}{\sigma_x^n \sigma_y^n} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n} \left[\frac{(x_j - \mu_x)^2}{\sigma_x^2} + \frac{(y_j - \mu_y)^2}{\sigma_y^2} \right] \right\} \\ &= \frac{1}{\sigma_x^n \sigma_y^n} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n} \left[\left(\frac{(x_j - \overline{x})^2}{\sigma_x^2} + \frac{(y_j - \overline{y})^2}{\sigma_y^2} \right) \right.\right. \\ &\quad \left.\left. + \frac{n(\overline{x} - \mu_x)^2}{\sigma_x^2} + \frac{n(\overline{y} - \mu_y)^2}{\sigma_y^2} \right] \right\} \\ &= \frac{1}{\sigma_x^n} \exp\left\{ -\frac{1}{2} \left[\sum_{j=1}^{n} \frac{(x_j - \overline{x})^2}{\sigma_x^2} + \frac{n(\overline{x} - \mu_x)^2}{\sigma_x^2} \right] \right\} \\ &\quad \times \frac{1}{\sigma_y^n} \exp\left\{ -\frac{1}{2} \left[\sum_{j=1}^{n} \frac{(y_j - \overline{y})^2}{\sigma_y^2} + \frac{n(\overline{y} - \mu_y)^2}{\sigma_y^2} \right] \right\}. \end{aligned}$$

Debido a esta última factorización se tiene que en este caso la estimación de parámetros es equivalente al caso univariado donde se tienen dos muestras separadas $X = (x_1, ..., x_n)$ y $Y = (y_1, ..., y_n)$. Para la variable X se tiene que la log-verosimilitud marginal es:

$$l(\mu_x, \sigma_x; \underline{x}) = -n \log(\sigma_x) - \frac{1}{2} \left[\sum_{j=1}^{n} \frac{(x_j - \overline{x})^2}{\sigma_x^2} + \frac{n(\overline{x} - \mu_x)^2}{\sigma_x^2} \right].$$

El resultado es análogo para la variable Y. Las entradas del vector de primeras derivadas son:

$$
\begin{aligned}
Sc(\mu_x;\underline{x}) &= \frac{\partial l(\mu_x,\sigma_x)}{\partial \mu_x} = -\frac{2n(\overline{x}-\mu_x)}{\sigma_x^2}, \\
Sc(\sigma_x;\underline{x}) &= \frac{\partial l(\mu_x,\sigma_x)}{\partial \sigma_x} = -\frac{n}{\sigma_x} + \frac{1}{\sigma_x^3}\left[\sum_{j=1}^{n}(x_j-\overline{x})^2 + \frac{n(\overline{x}-\mu_x)^2}{\sigma_x^2}\right].
\end{aligned}
$$

Igualando éstas a cero, se obtienen los estimadores de máxima verosimilitud para estos cuatro parámetros:

$$
\begin{aligned}
\widehat{\mu_x} &= \overline{x}, \\
\widehat{\mu_y} &= \overline{y}, \\
\widehat{\sigma_x^2} &= \frac{1}{n}\sum_{j=1}^{n}(x_j-\overline{x})^2, \\
\widehat{\sigma_y^2} &= \frac{1}{n}\sum_{j=1}^{n}(y_j-\overline{y})^2.
\end{aligned}
\tag{2.4}
$$

Entonces, por la propiedad de invarianza de la función de verosimilitud (Kalbfleisch, 1985) se tiene que los estimadores de máxima verosimilitud de los parámetros de interés son:

$$
\widehat{\beta} = \frac{\widehat{\mu_x}}{\widehat{\mu_y}}, \tag{2.5}
$$

$$
\begin{aligned}
\widehat{\delta_y} &= \frac{\sqrt{\widehat{\sigma_y^2}}}{\widehat{\mu_y}}, \\
\widehat{\delta_x} &= \frac{\sqrt{\widehat{\sigma_x^2}}}{\widehat{\mu_x}}, \\
\widehat{\rho} &= \sqrt{\frac{\widehat{\sigma_y^2}}{\widehat{\sigma_x^2}}}.
\end{aligned}
\tag{2.6}
$$

Finalmente, para el modelo normal $f_N(z;\beta,\sigma)$ los estimadores de máxima verosimilitud correspondientes son:

$$\begin{aligned} \widehat{\beta}_N &= \overline{z}, \\ \widehat{\sigma}_z^2 &= \frac{1}{n}\sum_{j=1}^{n}(z_j - \overline{z})^2. \end{aligned} \tag{2.7}$$

Para la verosimilitud dada en (2.2) los estimadores de máxima verosimilitud se tienen que encontrar mediante métodos numéricos, pues no hay una solución analítica cerrada.

2.3 Ejemplos de aproximaciones normales a la distribución de Z

En esta sección se presentarán seis ejemplos donde se analizan y se comparan la aproximaciones normales $f_N(z;\overline{z}, s_z)$ y $f_{NP}(z;\beta,\delta_x,\delta_y)$ a la densidad teórica $g_Z(z;\beta,\rho,\delta_y)$ para diferentes valores del coeficiente de variación δ_y. Los estimadores $\overline{z}$ y s_z serán calculados a partir de datos simulados (X,Y) con los valores teóricos de los parámetros $(\mu_x,\mu_y,\sigma_x,\sigma_y)$. La aproximación normal propuesta f_{NP} se calcula y grafica con los valores de los parámetros que se usaron para simular datos. No se sugiere usarla para estimar β, sino que meramente se presenta con fines comparativos de mostrar si es razonable o no aproximar a g_Z con una normal.

La elección de estos seis ejemplos se hizo con la finalidad de ilustrar también lo que sucede con las inferencias con los tres modelos $g(z,y)$, $g_Z(z)$ y $f_N(z)$ en diferentes escenarios. Esto es, se desea ilustrar casos cuando la densidad g_Z sea simétrica, asimétrica, con colas pesadas, bimodal y ante diferentes tamaños de muestra. La forma de g_Z depende fuertemente de la magnitud de δ_y. Por tanto también δ_y será crucial para determinar si es razonable aproximar g_Z con una densidad normal.

A continuación se muestra una tabla con las características de cada ejemplo:

Ejemplo	δ_y	Datos simulados	Forma de g_Z
1	0.025	30	Simétrica
2	0.025	5	Simétrica
3	0.17	30	Ligeramente Asimétrica
4	0.5	30	Asimétrica
5	5	30	Colas pesadas
6	29	30	Bimodal

Tabla 2.1: Características de los ejemplos.

Para los Ejemplos 1, 3, 4, 5 y 6 se simularon 30 datos de normales independientes $X \sim N(\mu_x, \sigma_x)$y $Y \sim N(\mu_y, \sigma_y)$ los cuales se consideraron pareados de acuerdo a su orden de aparición; luego con estas muestras se calcularon las razones $z_i = x_i/y_i,\ i = 1, ..., n$ y se estimaron los parámetros de las densidades con las expresiones dadas en (2.5). El Ejemplo 2 es similar pero sólo con 5 parejas de datos simulados (x_i, y_i).

Ejemplo 1

Este es un caso donde la aproximación normal a $g_Z(z;\beta,\rho,\delta_y)$ es razonable porque ambos δ_x y δ_y son pequeños (y nótese que menores a 0.1). Se simularon 30 datos $x_i \sim N(50, 2.5)$, $y_i \sim N(80, 2)$. En la Figura 2.1 se muestra la gráfica de la densidad teórica $g_Z(z;\beta,\rho,\delta_y)$, la aproximación normal estimada $N(z;\overline{z},\widehat{\sigma})$ a partir de los datos simulados y la aproximación normal propuesta teórica $N\left(z;\beta,\beta\sqrt{\delta_x^2+\delta_y^2}\right)$, donde $\beta = 0.625$ y $\beta\sqrt{\delta_x^2+\delta_y^2} = 0.0349$, además de la tabla con los parámetros estimados bajo estos modelos:

$\overline{z}$	$\overline{x}/\overline{y}$	s_z	$\sqrt{\widehat{\delta}_x^2+\widehat{\delta}_y^2}$	$\widehat{\delta}_x$	$\widehat{\delta}_y$
0.6310	0.6307	0.0356	0.0588	0.0536	0.0243

Tabla 2.2: Estimadores para los datos del Ejemplo 1.

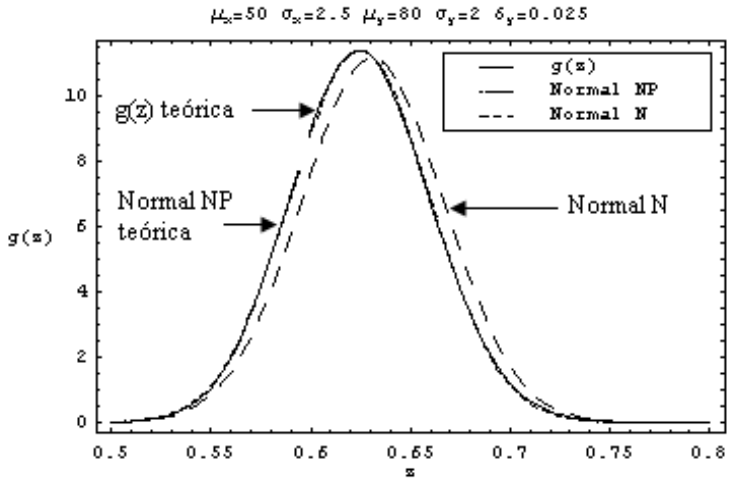


Figura 2.1: Densidad de Z y aproximaciones normales.

Nótese que $0 < \delta_x < 1/3$, y $0 < \delta_y < 1/3$. Se observa que en este caso las aproximaciones normales son razonables y que la aproximación normal propuesta teórica se aproxima muy bien a la densidad teórica de Z.

Ejemplo 2

En este caso se simularon con los mismos parámetros que en el Ejemplo 1 pero aquí ahora con una muestra mucho mas pequeña, de tamaño $n = 5$. Si bien el ajuste es razonablemente bueno, no lo es tanto como en el ejemplo

anterior debido a que la estimación de parámetros se realiza con una muestra muy pequeña.

Se simularon 5 datos con los mismos modelos que en el Ejemplo 1 $x_i \sim N(50, 2.5)$, $y_i \sim N(80, 2)$. A continuación se muestra la gráfica de las densidades teóricas $g_Z(z;\beta,\rho,\delta_y)$, la aproximación normal propuesta teórica $N\left(z;\beta,\beta\sqrt{\delta_x^2+\delta_y^2}\right)$, donde $\beta = 0.625$ y $\beta\sqrt{\delta_x^2+\delta_y^2} = 0.0349$, la aproximación normal estimada $N(z;\overline{z},\widehat{\sigma})$ y la tabla de los parámetros estimados correspondientes:

$\overline{z}$	$\overline{x}/\overline{y}$	s_z	$\sqrt{\widehat{\delta}_x^2+\widehat{\delta}_y^2}$	$\widehat{\delta}_x$	$\widehat{\delta}_y$
0.6221	0.6221	0.0322	0.0680	0.0614	0.0292

Tabla 2.3: Estimadores para los datos del Ejemplo 2.

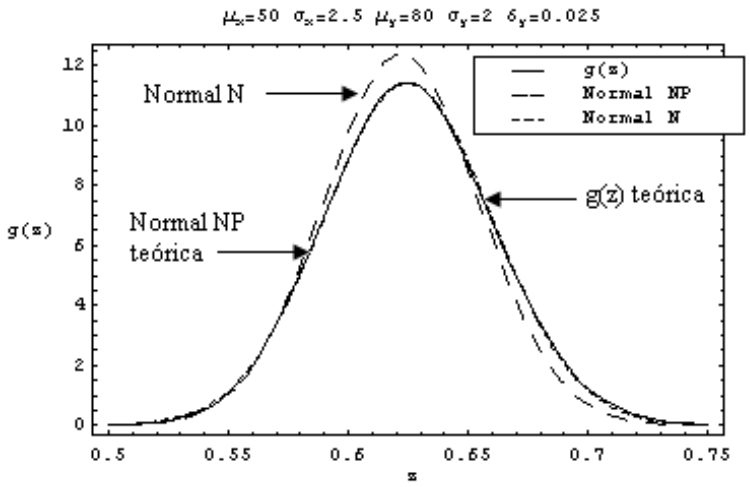


Figura 2.2 : Densidad de Z y aproximaciones normales.

Nuevamente se observa que tanto δ_x como δ_y teóricos y estimados son menores que 0.1 y que una aproximación normal es razonable. Más no necesariamente es ésta la f_N estimada con estos datos. En este ejemplo se observa un mejor ajuste a la densidad de Z de la aproximación normal propuesta teórica que con f_N estimada.

Ejemplo 3

En este ejemplo se muestra un caso cuando el coeficiente de variación de Y es tal que $0.1 < \delta_y < 0.2$, en donde la aproximación normal no es muy buena a pesar de estar en el rango recomendado por Qiao, et. al. (2006),

para aproximar a $\overline{x}/\overline{y}$ con $\overline{z}$. Esto se debe a que la densidad $g_Z(z;\beta,\rho,\delta_y)$ es ligeramente asimétrica.

Se simularon 30 datos $x_i \sim N(1, 0.17)$, $y_i \sim N(1, 0.17)$. A continuación se muestra la gráfica de las densidades $g_Z(z;\beta,\rho,\delta_y)$, la aproximación normal estimada $N(z;\overline{z},\widehat{\sigma})$ y la aproximación normal propuesta teórica $N\left(z;\beta,\beta\sqrt{\delta_x^2+\delta_y^2}\right)$, donde $\beta = 1$ y $\beta\sqrt{\delta_x^2+\delta_y^2} = 0.2404$. La tabla con los estimadores es la siguiente:

$\overline{z}$	$\overline{x}/\overline{y}$	s_z	$\sqrt{\widehat{\delta}_x^2+\widehat{\delta}_y^2}$	$\widehat{\delta}_x$	$\widehat{\delta}_y$
1.0360	0.9902	0.3089	0.2387	0.1512	0.1846

Tabla 2.4: Estimadores para los datos del Ejemplo 3.

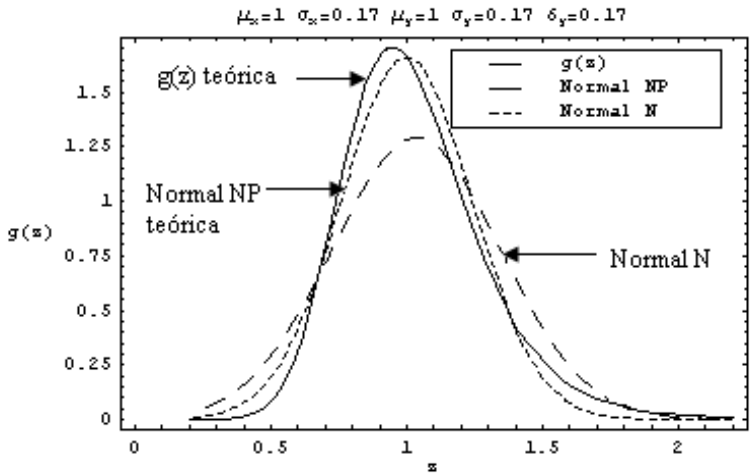


Figura 2.3: Densidad de Z y aproximaciones normales.

En este caso como δ_y es mayor a 0.1, entonces ni la aproximación normal teórica ni la estimada son buenas. Cuando el estimador de δ_y es grande se puede detectar esta situación. Nótese que la densidad de Z es ligeramente asimétrica, por lo tanto no es razonable utilizar una aproximación normal.

Ejemplo 4

Aquí se muestra el caso en que la densidad teórica $g_Z(z;\beta,\rho,\delta_y)$ es fuertemente asimétrica ya que $\delta_y = 0.5$ es muy grande y por lo tanto no es razonable hacer una aproximación normal, por lo que la aproximación normal teórica no se presenta.

Se simularon 30 datos $x_i \sim N(80, 2)$, $y_i \sim N(1, 0.5)$. A continuación se muestra la gráfica de las densidades $g_Z(z;\beta,\rho,\delta_y)$ teórica y la normal estimada $N(z;\overline{z},\widehat{\sigma})$.

La tabla con los estimadores correspondientes es:

$\overline{z}$	$\overline{x}/\overline{y}$	s_z	$\sqrt{\widehat{\delta}_x^2+\widehat{\delta}_y^2}$	$\widehat{\delta}_x$	$\widehat{\delta}_y$
115.408	82.9048	77.8053	0.4716	0.0264	0.4708

Tabla 2.5: Estimadores para los datos del Ejemplo 4.

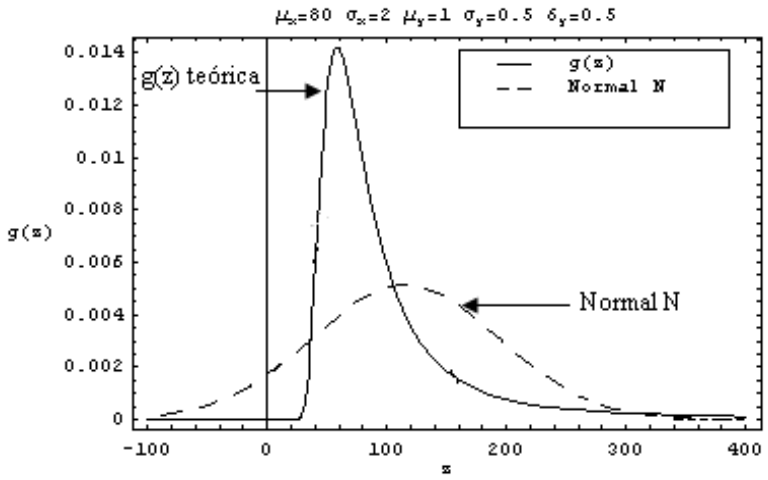


Figura 2.4: Densidad de Z y aproximaciones normales.

En este caso la densidad de Z es asimétrica, por lo tanto no es razonable utilizar una aproximación normal. Además $\sqrt{\widehat{\delta}_x^2+\widehat{\delta}_y^2} > 1/3$ y $\widehat{\delta}_y > 1/3$ lo cual indica que una aproximación normal no será razonable.

Ejemplo 5

Caso con el coeficiente de variación $\delta_y = 5$ grande en el que no hay una buena aproximación normal a la densidad $g_Z(z;\beta,\rho,\delta_y)$, debido a que esta densidad tiene colas pesadas ($\delta_x = 5$ también es grande).

Se simularon 30 datos $x_i \sim N(1,5)$, $y_i \sim N(1,5)$. A continuación se muestra la gráfica de la densidad teórica $g_Z(z;\beta,\rho,\delta_y)$ y la aproximación normal estimada $N(z;\overline{z},\widehat{\sigma})$. La tabla con los parámetros estimados es:

$\overline{z}$	$\overline{x}/\overline{y}$	s_z	$\sqrt{\widehat{\delta}_x^2+\widehat{\delta}_y^2}$	$\widehat{\delta}_x$	$\widehat{\delta}_y$
-1.6706	-5.9651	3.8974	18.0074	2.6156	-17.8164

Tabla 2.6: Estimadores para los datos del Ejemplo 5.

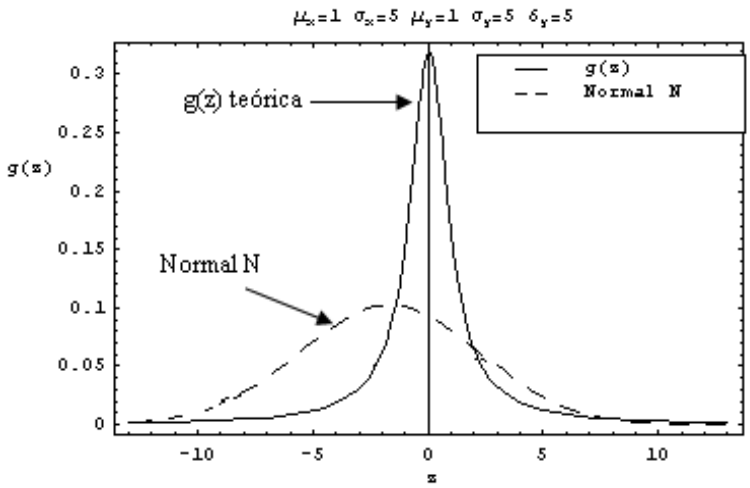


Figura 2.5: Densidad de Z y aproximaciones normales.

En este ejemplo el estimador $\widehat{\delta}_y$ es muy grande en valor absoluto y esto indica claramente que una aproximación normal no será buena.

Ejemplo 6

En este ejemplo se muestra el caso en que la densidad $g_Z(z;\beta,\rho,\delta_y)$ es bimodal. La bimodalidad de esta densidad suele presentarse cuando el coeficiente de variación δ_y es grande, usualmente mayor a 10.

Se simularon 30 datos $x_i \sim N(80,2)$, $y_i \sim N(1.6\underline{6}, 48.3\underline{3})$, donde $\delta_y = 29$. A continuación se muestra la gráfica de la densidad teórica $g_Z(z;\beta,\rho,\delta_y)$ y la aproximación normal estimada $N(z;\overline{z},\widehat{\sigma})$. La tabla de los parámetros estimados es:

$\overline{z}$	$\overline{x}/\overline{y}$	s_z	$\sqrt{\widehat{\delta}_x^2+\widehat{\delta}_y^2}$	$\widehat{\delta}_x$	$\widehat{\delta}_y$
1.6359	8.2980	5.1780	5.9964	0.0208	5.9964

Tabla 2.7: Estimadores para los datos del Ejemplo 6.

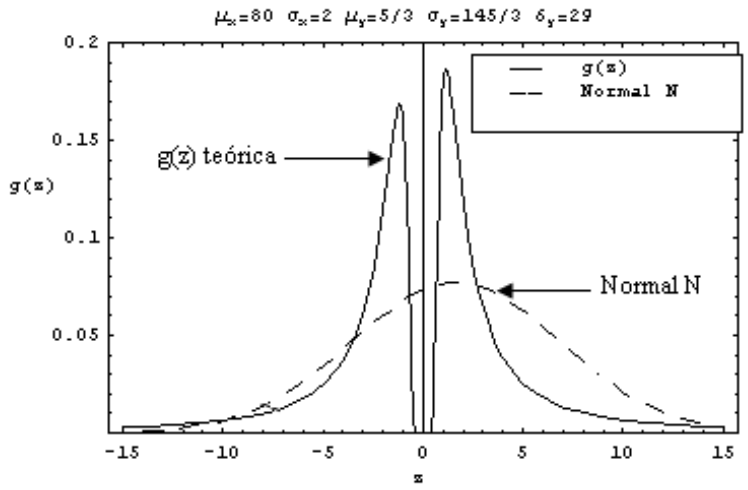


Figura 2.6: Densidad de Z y aproximaciones normales.

En este caso la aproximación no es razonable debido a la bimodalidad. Este es un caso extremo para ejemplificar que existen situaciones en las que es absurdo dar una aproximación normal a la densidad de Z.

Empíricamente se ha visto en estos 6 ejemplos que para que una aproximación normal sea buena, debe ocurrir que $\widehat{\delta}_x$ y $\widehat{\delta}_y$ sean menores a 0.1.

2.4 Verosimilitudes perfiles de β bajo los modelos considerados

En esta sección se presentan las verosimilitudes perfiles de β construidas a partir de los tres modelos estadísticos $g(z,y;\beta,\rho,\delta_y,\sigma_x)$, $g_Z(z;\beta,\rho,\delta_y)$ y $f_N(z;\beta,\sigma)$ para los datos de los seis ejemplos descritos. La idea es poder comparar las consecuencias de adoptar cada modelo en cuanto a las inferencias que se harán sobre β.

La verosimilitud perfil relativa (véase Kalbfleisch, 1985, Sección 10) de β a partir de los tres modelos estadísticos $g(z,y;\beta,\rho,\delta_y,\sigma_x)$, $g_Z(z;\beta,\rho,\delta_y)$ y $f_N(z;\beta,\sigma)$ se definen respectivamente como:

$$R_{\beta}^{ZY} = R_{P}^{(z,y)}(\beta; \underline{z}, \underline{y}) \propto \max_{\rho,\delta_y,\sigma_x|\beta} L_{ZY}(\beta, \rho, \delta_y, \sigma_x; \underline{z}, \underline{y}) \Big/ L_{ZY}(\widehat{\beta}, \widehat{\rho}, \widehat{\delta}_y, \widehat{\sigma}_x; \underline{z}, \underline{y}) \,, \tag{2.8}$$

$$R_{\beta}^{Z} = R_{P}^{Z}(\beta; \underline{z}) \propto \max_{\rho,\delta_y|\beta} L_{Z}(\beta, \rho, \delta_y; \underline{z}) \Big/ L_{Z}(\widehat{\beta}, \widehat{\rho}, \widehat{\delta}_y; \underline{z}) \,, \tag{2.9}$$

$$y \quad R_{\beta}^{N} = R_{P}^{N}(\beta; \underline{z}) \propto \max_{\sigma|\beta} L_{N}(\beta, \sigma; \underline{z}) \Big/ L_{N}(\widehat{\beta}, \widehat{\sigma}; \underline{z}) \,. \tag{2.10}$$

Es de interés cuantificar el error cometido al utilizar el modelo (2.9) para hacer inferencias sobre el parámetro β ya que también es utilizado, como en Kuethe et. al. (2000).

2.5 Ejemplos de inferencias sobre β

A continuación se retoman los Ejemplos 1, 3 y 6 donde g_Z era simétrica, ligeramente asimétrica y bimodal, mostrados en la Tabla 2.1 de la Sección 2.2 y se realizarán inferencias sobre el parámetro β con las tres verosimilitudes perfiles descritas en la Sección 2.4. Con estos ejemplos se ilustra que pueden existir casos en los cuales puede ser razonable utilizar una aproximación normal como en el Ejemplo 1, o en los que es absurdo utilizar una aproximación normal, como se muestra en los Ejemplos 3 y 6; el Ejemplo 3 tiene como objetivo mostrar que las inferencias por intervalo pueden ser diferentes para los modelos considerados, incluso para valores de los parámetros estimados dentro del rango $\widehat{\delta}_x$ y $\widehat{\delta}_y < 0.2$ recomendado por Qiao et. al. (2006).

Se muestran además dos ejemplos adicionales de datos reales que se denotarán como Ejemplo 7 y Ejemplo 8. En el Ejemplo 7 se utilizan mediciones de Citometría de Flujo realizadas en plantas de agave *Tequilana Weber* variedad azul y de maíz *Zea*. En el Ejemplo 8 se utilizan datos tomados en pacientes diabéticos y normales, Miller (1997, pp. 258).

Ejemplo 1

A continuación se muestran las verosimilitudes perfil para $\beta : R_{ZY}, R_Z$ y R_N, utilizando los datos simulados utilizados en la Sección 2.2.

Recuérdese que en este caso se simularon 30 datos $x_i \sim N(50, 2.5)$, $y_i \sim N(80, 2)$ donde los valores teóricos de los parámetros son $\beta = 50/80 = 0.625$, $\delta_x = 0.05$ y $\delta_y = 0.025$. Es decir, este es un caso donde la aproximación normal queda bien y además la magnitud de los parámetros teóricos asemeja la situación de datos reales como los del Ejemplo 7. Los parámetros estimados son:

$\widehat{\beta}_{ZY}$	$\widehat{\beta}_Z$	$\widehat{\beta}_N$	$\widehat{\delta}_x$	$\widehat{\delta}_y$	$\sqrt{\widehat{\delta}_x^2 + \widehat{\delta}_y^2}$
0.6307	0.6306	0.6310	0.0315	0.0220	0.0588

Tabla 2.8: Estimadores de β, $\widehat{\delta}_x$ y $\widehat{\delta}_y$ para los datos del Ejemplo 1.

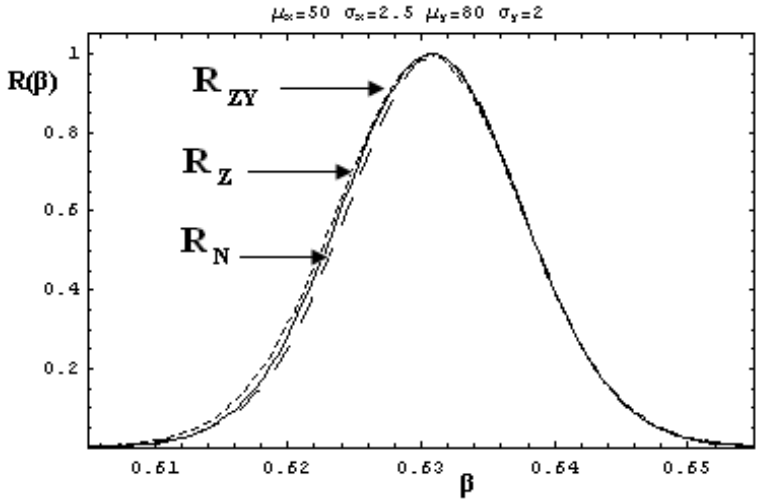


Figura 2.7: Verosimilitudes perfiles de β para los datos del Ejemplo 1.

Nótese que ambos estimadores $\widehat{\delta}_x$ y $\widehat{\delta}_y$ son menores que 0.1. Aquí se observa que los estimadores de máxima verosimilitud del parámetro de interés β son muy parecidos y cercanos al valor teórico con que se simularon los datos. Además se obtienen verosimilitudes perfiles para β muy similares con los tres modelos estadísticos. Esto nos indica que la distribución condicional $Y|Z$ no contiene información relevante sobre el parámetro β, lo cual se ve reflejado en la semejanza de la verosimilitud perfil de β del modelo conjunto R_{ZY} y la verosimilitud perfil de β del modelo marginal. Debido a que los coeficientes de variación estimados $\widehat{\delta}_x$ y $\widehat{\delta}_y$ son pequeños, la aproximación normal es buena

y se llega a las mismas inferencias sobre β sin importar cuál modelo se use.

Ejemplo 3

En este caso se utilizarán los datos del Ejemplo 3 de la Sección 2.2 donde se simularon 30 datos de las variables $x_i \sim N(1, 0.17)$, $y_i \sim N(1, 0.17)$,donde los valores teóricos de los parámetros son $\beta = 1$ y $\delta_x = \delta_y = 0.17$. Los parámetros estimados son:.

$\widehat{\beta}_{ZY}$	$\widehat{\beta}_Z$	$\widehat{\beta}_N$	$\widehat{\delta}_x$	$\widehat{\delta}_y$	$\sqrt{\widehat{\delta}_x^2 + \widehat{\delta}_y^2}$
0.9902	0.9802	1.0360	0.1512	0.1846	0.2387

Tabla 2.9: Estimadores de β, $\widehat{\delta}_x$ y $\widehat{\delta}_y$ para los datos del Ejemplo 3.

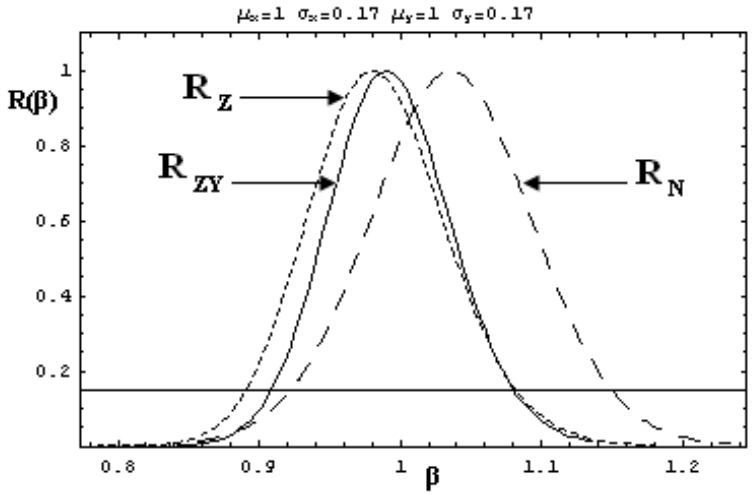


Figura 2.8: Verosimilitudes perfiles de β para los datos del Ejemplo 2.

Nótese que ambos estimadores $\widehat{\delta}_x$ y $\widehat{\delta}_y$ son mayores a 0.1. En este ejemplo se pueden observar varios puntos. En Qiao et. al. (2006) se recomienda usar $\overline{z}$ como aproximación a $\overline{x}/\overline{y}$ si $\widehat{\delta}_y < 0.2$. Podemos observar en la tabla de estimadores que, en efecto, estas estimaciones puntuales de β son parecidas, aunque el estimador obtenido con la aproximación normal estimada es ligeramente mayor. En la Figura 2.8 se marcó la línea horizontal en el nivel del 15% de verosimilitud para los tres modelos. Los intervalos de verosimilitud son diferentes para los tres modelos estadísticos; sin embargo los intervalos de verosimilitud de R_Z tienen una mayor intersección con los intervalos de

verosimilitud de R_{ZY} y ambos difieren bastante del intervalo de la aproximación normal R_N. Esto se debe a que en este caso no es razonable utilizar una aproximación normal a la densidad marginal g_Z como podemos ver en la Figura 2.3 puesto que, en este caso, se estaría sobre-estimando a β. Lo recomendable aquí es adoptar el modelo conjunto y por tanto R_{ZY}. Este es un caso donde el modelo condicional $g_{Y|Z}$ sí contiene un poco de información sobre β.

Ejemplo 6

En este ejemplo se utilizaron los datos del Ejemplo 6 donde se simularon 30 datos $x_i \sim N(80, 2)$, $y_i \sim N(1.6\underline{6}, 48.3\underline{3})$, donde los valores teóricos de los parámetros son $\beta = 48$, $\delta_x = 0.025$ y $\delta_y = 29$. Recuérdese que como δ_y es muy grande, la densidad de Z es bimodal. Los parámetros estimados son:

$\widehat{\beta}_{ZY}$	$\widehat{\beta}_Z$	$\widehat{\beta}_N$	$\widehat{\delta}_x$	$\widehat{\delta}_y$	$\sqrt{\widehat{\delta}_x^2 + \widehat{\delta}_y^2}$
8.2980	9.4755	1.7204	0.0208	5.9964	5.9964

Tabla 2.10: Estimadores de β, $\widehat{\delta}_x$ y $\widehat{\delta}_y$ para los datos del Ejemplo 6.

Nótese que $\widehat{\delta}_y$ es gigantesco a pesar de estar muy alejado del valor teórico $\delta_y = 29$. Los valores estimados están muy alejados de los valores teóricos como se muestra en la Tabla 2.7, esto es debido a que los valores simulados de Y tienen una varianza grande. Equivalentemente, para tal valor de δ_y el tamaño de muestra de 30 datos resulta muy pequeño. A continuación se muestran las verosimilitudes perfiles para β con los tres modelos:

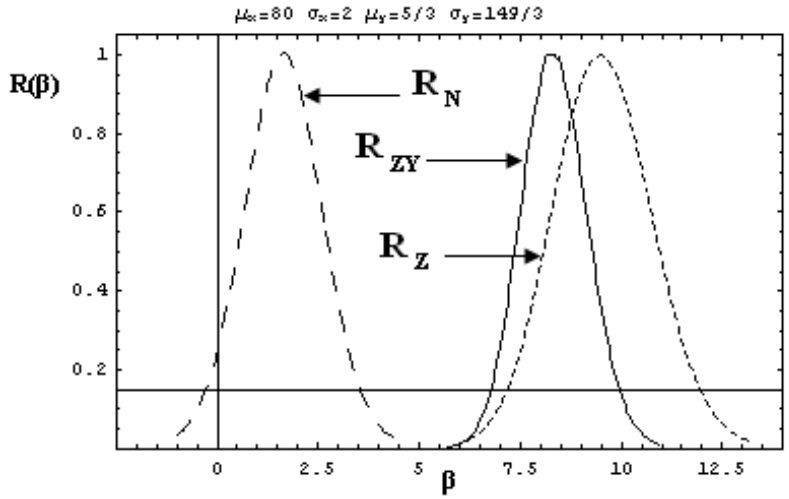


Figura 2.9: Verosimilitudes perfiles de β para los datos del Ejemplo 5.

Este es un caso extremo en el que claramente una aproximación normal no es razonable. Además se tiene que la verosimilitud perfil de β utilizando el modelo estadístico $g_Z(z;\beta,\rho,\delta_y)$ es bimodal, como se ilustra en la Figura 2.10, donde se amplió el intervalo de graficación para resaltar este hecho, mientras que la verosimilitud del modelo conjunto R_{ZY} siempre es unimodal. Así, la Figura 2.9 sólo muestra una parte de esta perfil de β, R_Z.

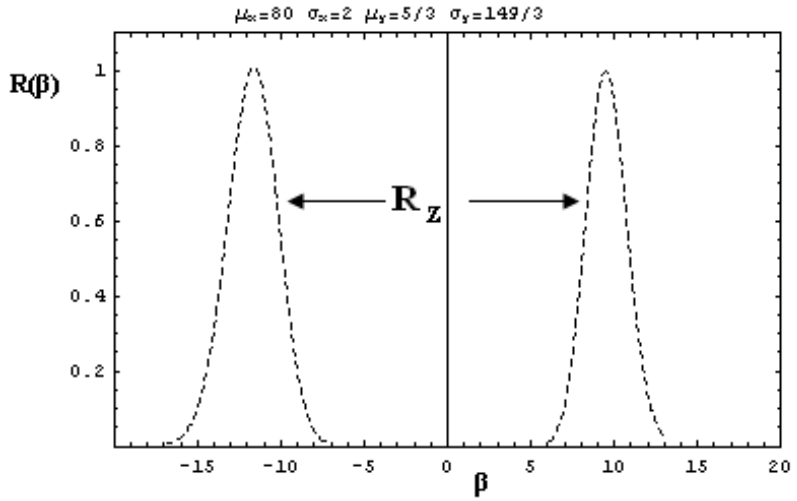


Figura 2.10: Bimodadlidad de la verosimilitud perfil de β para el modelo marginal.

La verosimilitud perfil de β del modelo conjunto (Z, Y) es unimodal, por lo tanto la bimodalidad de R_Z es un indicador de la pérdida de información relevante sobre el parámetro β contenida en la parte condicional $Y|Z$.

Ejemplo 7

En Citometría de Flujo se desea estimar la cantidad del ADN en el núcleo de una célula de una planta de interés. La cantidad de ADN de una célula se mide a través de un rayo láser que emite el citómetro de flujo y del cual se capta el reflejo que da la hélice de ADN debido a la tintura que se le agregó al preparar las hojas de las plantas en el laboratorio. Debido al proceso de medición del citómetro de flujo, existen fuentes de variabilidad y es por ello que a pesar de que todas las células de agave que no están a punto de reproducirse tienen la misma cantidad de ADN, las mediciones pueden ser distintas y mostrar mucha variación. Este error se manifiesta en la desviación de los datos alrededor del verdadero valor del contenido de ADN en la célula. Para el proceso de estimación del ADN de una planta de interés se requiere contar con una planta de control con ADN conocido que sea cercano en magnitud al de la planta de interés. Se machacan juntos con una navaja fina un trozo de la planta de interés y otro de la planta de control en una solución con una tinta fluorescente que penetra en el núcleo de las células y se aloja entre la hélice del ADN. Esta solución se pasa posteriormente a través del citómetro. Este procedimiento produce un pareamiento natural de las mediciones, porque para cada solución hecha a partir de trozos de la planta de interés y control, el citómetro arroja un histograma como el siguiente para una solución analizada de aproximadamente 32 mil células de agave azul *Tequilana Weber* y maíz *Zea* (tomado de Díaz-Francés y Sprott, 2001):

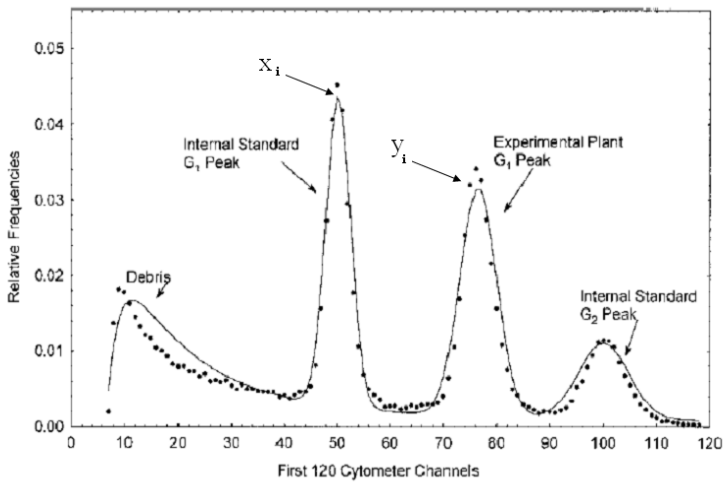


Figura 2.11: Ejemplo de una salida del citómetro para la muestra i-ésima.

Aquí x_i es el pico G1 de la planta de control y y_i es el pico G1 de la planta de interés, produciendo así el par (x_i,y_i). El pico G1 denota la media del ADN nuclear de células que no han iniciado todavía su proceso de reproducción celular (antes de reproducirse en dos células hijas, se duplica el ADN nuclear). Es decir, (x_i,y_i) son estimadores de las primeras dos de las tres medias normales de esta mezcla de lognormal con tres normales. Esta medición se repite n veces y se obtienen así n pares (x_i,y_i) cuyo histograma se muestra en la Figura 2.11. En adición al proceso descrito, para cada pareja (x_i,y_i) el citometrista controla la escala cuando comienzan a observar las mediciones ya que mueve al eje horizontal que se exhibe en la Figura 2.11 para que el primer pico (donde cae X_i) esté cercano al canal 50. Esto es porque las observaciones entre los canales 40 y 200 tienen mayor calidad. Así que por la génesis de los datos, es razonable suponer que $x_i \sim N(\beta\mu_i, \sigma_x)$ y $y_i \sim N(\mu_i, \sigma_y)$ para $i = 1, ..., n$. Estos supuestos se consideran en Díaz-Francés y Sprott (2003). Sin embargo para poder cuantificar el error en la literatura de citometría de manera más sencilla se simplificarán estos supuestos y se considerará que todas las μ_i para $i = 1, ..., n$ son iguales.

La fórmula para δ el ADN contenido en el núcleo de una célula de la planta

de interés está dada por:

$$\delta = \frac{\mu}{\eta}\kappa = \left(\frac{\text{Media del ADN de la planta de interés}}{\text{Media del ADN de la planta de control}}\right)\kappa = \beta\kappa,$$

donde κ es el ADN conocido, contenido en el núcleo de una célula de la planta de control.

Así se supondrá que los estimadores (x_i, y_i) se distribuyen como normales independientes con medias positivas μ, η y además se contará con una muestra de tamaño n de datos naturalmente pareados.

A continuación se muestran los datos presentados por Díaz-Francés y Sprott (2001), donde para cuatro plantas de agave, dos jóvenes y dos adultas, se tomaron mediciones en tres diferentes posiciones de la planta: posición A ó interna basal, posición B ó interna apical y posición C ó externa basal, las cuales se muestran en la Figura 2.12. Simplificando, se supondrá que $x_i \backsim N(\mu_x, \sigma_x)$ y $y_i \backsim N(\mu_y, \sigma_y)$ con $i = 1, 2$:

Posición A	Réplica 1		Réplica 2	
Edad	x_i	y_i	x_i	y_i
Joven	50.804	79.303	53.243	81.075
Adulta	53.217	81.070	48.918	76.784
Posición B	Réplica 1		Réplica 2	
Edad	x_i	y_i	x_i	y_i
Joven	51.257	79.258	51.888	81.194
Adulta	51.478	78.764	49.068	76.574
Posición C	Réplica 1		Réplica 2	
Edad	x_i	y_i	x_i	y_i
Joven	48.852	76.174	50.071	78.125
Adulta	52.161	79.882	50.229	76.536

Tabla 2.11: Datos de Citometría, Díaz-Francés y Sprott (2001).

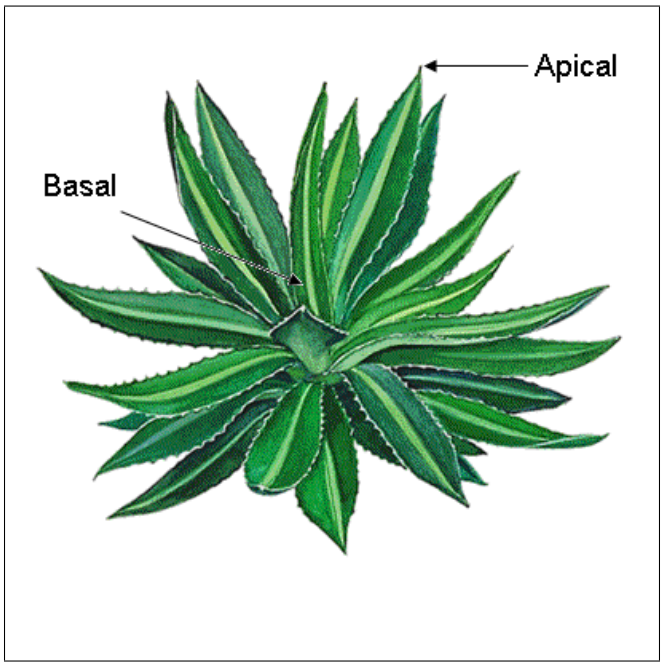


Figura 2.12: Posiciones del agave.

En la Figura 2.12 se muestran las partes apical y basal de una hoja central de agave, la cual tiene una parte interna y una externa que es más dura. En este caso es razonable el supuesto de independencia de x_i con y_i ya que las mediciones de ADN en la planta de interés son independientes de las mediciones de ADN en la planta de control (maíz).

A continuación se muestran las verosimilitudes perfiles para β con el modelo conjunto (Z, Y), que simplifica los supuestos que hay que considerar, utilizando los datos de plantas jóvenes y adultas en cada posición.

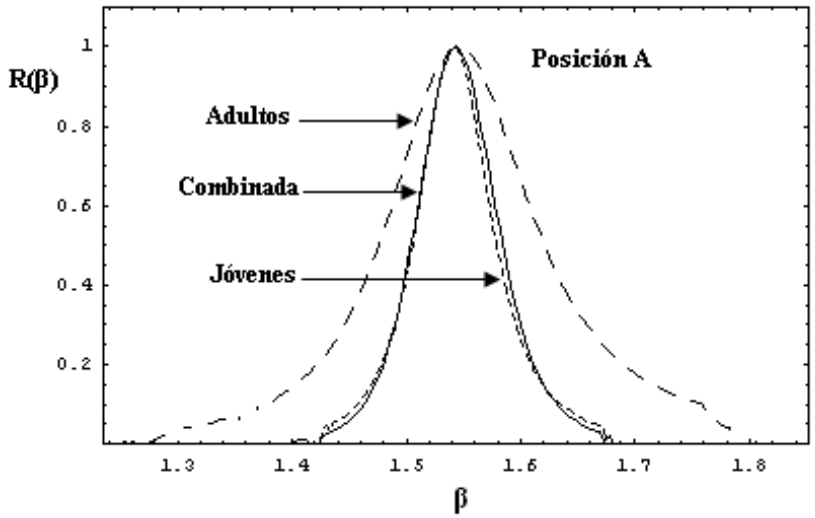


Figura 2.13: Pefiles de β con los datos de adultos, jóvenes y combinados para la posición A.

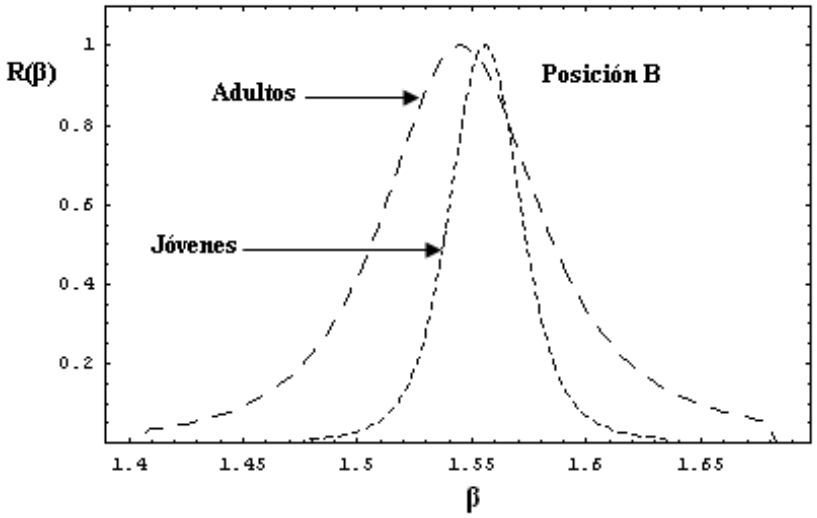


Figura 2.14: Pefiles de β con los datos de adultos y jóvenes para la posición B.

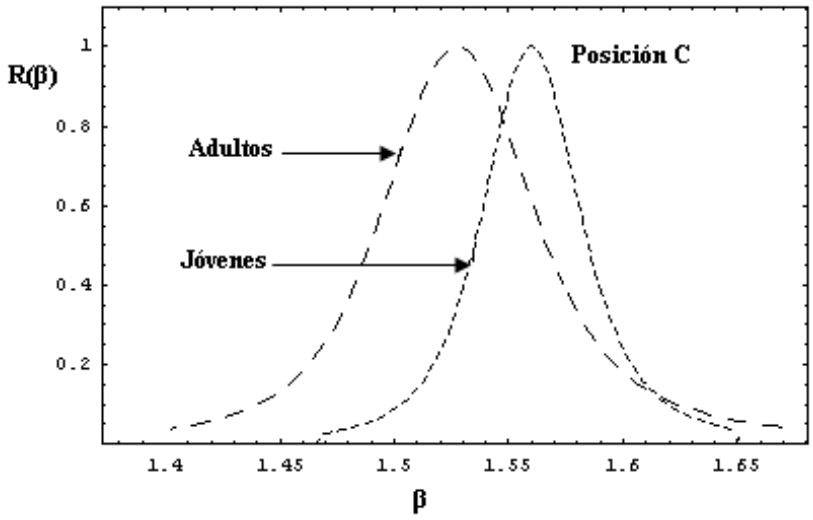


Figura 2.15: Pefiles de β con los datos de adultos y jóvenes para la posición C.

La teoría biológica indica que la cantidad del ADN contenido en una célula no cambia con la edad del individuo. Sin embargo, este hecho no necesariamente se observa al contar con mediciones.

En la posición A la inferencia sobre el ADN de plantas jóvenes y adultas es muy parecida, ya que las verosimilitudes del parámetro β son similares. Las diferencias en la estimación de β que se observan en las posiciones B y C, ambas externas, se deben seguramente a que en las plantas adultas la parte externa de la hoja tiene mucha más fibra que en las plantas jóvenes, lo que impide que se absorba bien la tinta en la hélice de ADN.

Por lo tanto la mejor posición para hacer inferencias sobre el parámetro de interés resulta ser la posición A ó interna basal, donde las plantas jóvenes y adultas contienen información homogénea sobre β que se puede combinar. Por ello es razonable juntar los cuatro datos obtenidos de plantas jóvenes y adultas a través del producto de las verosimilitudes correspondientes para obtener así la verosimilitud combinada que se muestra en la Figura 2.13.

Utilizando los cuatro datos de la posición A, los parámetros estimados con este conjunto de datos son:

$\widehat{\beta}_{ZY}$	$\widehat{\beta}_{Z}$	$\widehat{\beta}_{N}$	$\widehat{\delta}_{x}$	$\widehat{\delta}_{y}$	$\sqrt{\widehat{\delta}_{x}^{2}+\widehat{\delta}_{y}^{2}}$
1.5434	1.5439	1.5441	0.0220	0.0351	0.0414

Tabla 2.12: Estimadores de β, $\widehat{\delta}_x$ y $\widehat{\delta}_y$ para los datos de la posición A.

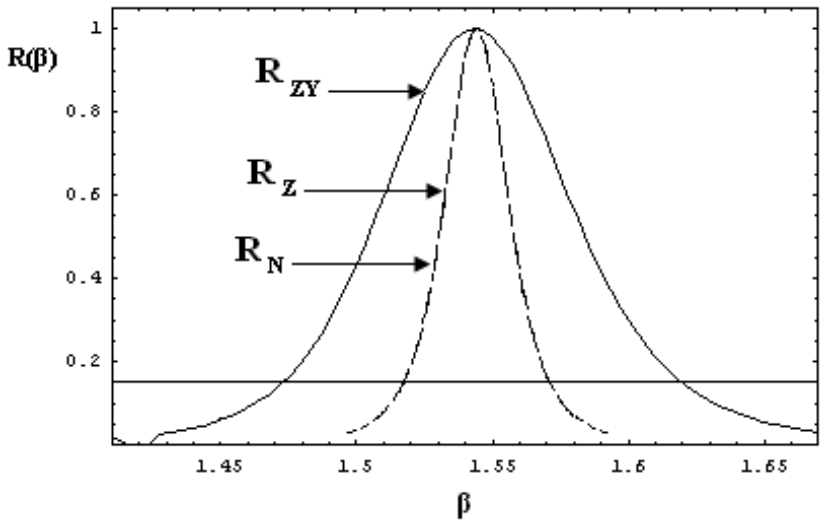


Figura 2.16: Verosimilitudes perfiles de β para los datos presentados por Díaz-Francés y Sprott [17].

Los estimadores de máxima verosimilitud de β son muy parecidos para los tres modelos estadísticos. Debido a que los coeficientes de variación estimados de X y Y son pequeños y menores a 0.1, es de esperarse que la aproximación normal a la densidad de Z sea buena. Sin embargo, este ejemplo nos sirve para detectar que cuando el tamaño de muestra es pequeño, la parte condicional $Y|Z$ sí puede contener información relevante sobre β. Esto no necesariamente se ve reflejado en la estimación puntual, sino en la dispersión de las tres curvas y por tanto en la longitud de los intervalos de verosimilitud; es decir, en la precisión para estimar β. Los modelos R_Z y R_N producen una mayor precisión ya que se obtienen intervalos de verosimilitud de menor longitud que con el modelo R_{ZY}. En este caso la aproximación normal a g_Z es buena y las verosimilitudes perfiles de β, R_Z, R_N son muy parecidas; sin embargo, son más angostas que R_{ZY}. Esto se debe a que la densidad condicional de $Y|Z$ sí contiene información sobre β.

Díaz-Francés y Sprott (2001) propusieron la siguiente verosimilitud para hacer inferencias sobre β en el caso de Citometría de Flujo:

$$L(\beta) \propto \left[\frac{\sum (x_i - \beta y_i)^2}{\lambda^2 + \beta^2}\right]^{-n/2}, \tag{2.11}$$

donde λ es tomado comúnmente como 1. Este modelo toma en cuenta el procedimiento del citometrista al recabar los datos, en el que tiene control manual sobre la escala en la que se obtienen las mediciones. Por ello se involucra a un parámetro adicional por cada pareja de observaciones: $x_i \backsim N(\mu_i, \sigma_x)$, $y_i \backsim N(\beta\mu_i, \sigma_y)$. Para este conjunto de datos, la verosimilitud perfil de β de (2.11) coincide con la aproximación normal R_N. Esto se debe a que es un modelo particular que incluye más información. Por lo tanto se tiene que la aproximación normal utilizada en citometría, a pesar de todos los supuestos simplificadores que toman, arroja resultados similares en este ejemplo para las inferencias sobre el parámetro β para este conjunto de datos de agave y maíz donde los coeficientes de variación δ_x y δ_y asociados son pequeños.

Lo importante es verificar que los coeficientes de variación estimados δ_x y δ_y sean ambos pequeños, menores a 0.1, así como lo fueron en este caso. En este caso resulta que la aproximación normal R_N coincide con la verosimilitud (2.11) como se muestra en la Figura 2.17. Por lo tanto en este caso particular al utilizar la verosimilitud perfil R_N para hacer inferencias sobre β con este conjunto de datos no se cometen errores grandes. Sin embargo, en un caso general, es preferible usar la verosimilitud (2.11) puesto que el modelo que la sustenta describe mejor el mecanismo de obtención de datos y también a la luz de la sencillez de graficar (2.11).

Se resalta también que al usar (2.11), las discrepancias entre las perfiles de β para las plantas jóvenes y adultas en las tres posiciones son mucho más marcadas y esto conlleva a una selección inmediata de la Posición A como la mejor para extender el experimento y así estimar adecuadamente el ADN del agave.

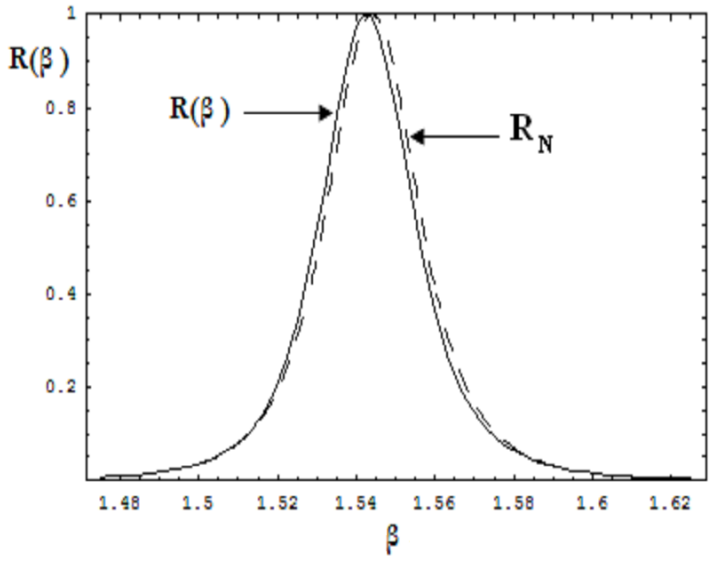


Figura 2.17: Verosimilitudes perfiles de β de Díaz-Francés y Sprott [17] $R(\beta)$ y aproximación normal R_N.

Ejemplo 8

A continuación se presentan unos datos del libro de Miller (1997, pp. 258) sobre pacientes con diabetes. En un estudio de diabetes, 21 pacientes, se identificaron siete como *normales*, siete como *diabéticos moderados* y siete como *diabéticos graves* mediante una prueba de glucosa previa y de un examen médico exploratorio. A los pacientes se les dio una infusión constante de glucosa durante un periodo de tiempo. La siguiente Tabla 3 muestra las concentraciones de glucosa e insulina antes y después de la prueba así como su diferencia:

Pacientes con tolerancia normal a la glucosa					
Glucosa			Insulina		
Antes	Después	Diferencia de Glucosa x_i	Antes	Después	Diferencia de Insulina y_i
86	150	64	26	53	27
80	174	94	11	46	35
73	137	64	16	72	56
81	166	85	28	57	29
84	153	69	12	48	36
82	170	88	24	210	186
82	164	82	24	51	27
Pacientes con diabetes moderada					
Glucosa			Insulina		
Antes	Después	Diferencia de Glucosa x_i	Antes	Después	Diferencia de Insulina y_i
88	198	110	21	72	51
131	300	169	76	264	188
105	238	133	32	102	70
93	193	100	18	64	46
93	220	127	29	163	134
94	187	93	34	138	104
98	217	119	30	75	45

Pacientes con diabetes grave					
Glucosa			Insulina		
Antes	Después	Diferencia de Glucosa x_i	Antes	Después	Diferencia de Insulina y_i
154	330	176	26	42	16
164	324	160	49	111	62
185	379	194	26	44	18
254	426	172	44	61	17
175	370	195	44	90	46
157	286	129	32	83	51
320	486	166	22	46	24

Tabla 2.13. Datos de glucosa e insulina en pacientes, Miller (1197).

Es de interés calcular el índice insulinogénico $\Delta I/\Delta G$ y compararlo entre los tres grupos, donde ΔI es la media teórica de la distribución de las y_i para las que es razonable suponer un modelo normal. De igual manera ΔG es la media de las variables aleatoria x_i. El índice insulinogénico estima la respuesta pancreática temprana a la carga oral de 75 gr. de glucosa, expresada por la proporción del incremento de insulina respecto al incremento de glucosa 30 minutos después de que se dio la carga oral de glucosa. Nótese que el índice insulinogénico equivale así a la razón de medias normales β, puesto que es razonable suponer que las diferencias de glucosa x_i observadas son una muestra de una normal y lo mismo para las diferencias de insulina observadas y_i. La pregunta de interés es si β es la misma en los tres grupos de pacientes considerados.

Nuevamente, a manera de simplificación se considerará que las diferencias de glucosa de todos los pacientes dentro de un mismo grupo se distribuyen normales con la misma media $\beta\mu_y$. De igual manera para las diferencias de insulina y_i, que se supondrán normales con media μ_y y varianza σ_y^2.

A continuación se muestran las gráficas de las verosimilitudes perfiles de β con los tres modelos para los tres grupos de pacientes:

Pacientes normales:

$\widehat{\beta}_{ZY}$	$\widehat{\beta}_{Z}$	$\widehat{\beta}_{N}$	$\widehat{\delta}_{x}$	$\widehat{\delta}_{y}$	$\sqrt{\widehat{\delta}_{x}^{2}+\widehat{\delta}_{y}^{2}}$
0.725	0.481	0.710	0.948	0.920	1.321

Tabla 2.14: Estimadores de β, $\widehat{\delta}_x$ y $\widehat{\delta}_y$ para los datos de pacientes normales.

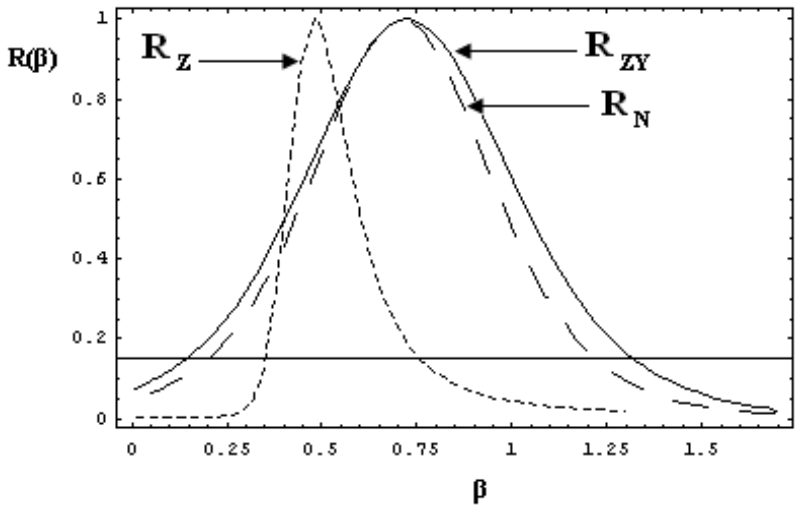


Figura 2.18: Verosimilitudes perfiles de β para los datos de pacientes normales.

Pacientes con diabetes moderada:

$\widehat{\beta}_{ZY}$	$\widehat{\beta}_{Z}$	$\widehat{\beta}_{N}$	$\widehat{\delta}_{x}$	$\widehat{\delta}_{y}$	$\sqrt{\widehat{\delta}_{x}^{2}+\widehat{\delta}_{y}^{2}}$
0.749	0.602	0.730	0.550	0.192	0.582

Tabla 2.15: Estimadores de β, $\widehat{\delta}_x$ y $\widehat{\delta}_y$ para los datos de pacientes con diabetes moderada.

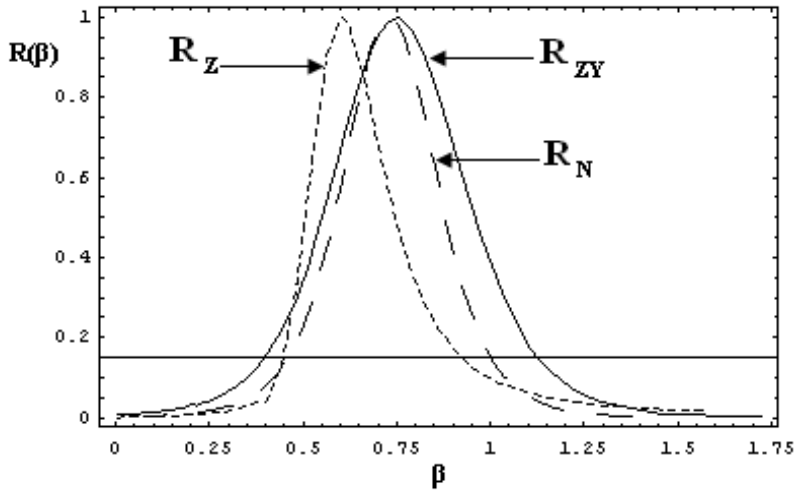


Figura 2.19: Verosimilitudes perfiles de β para los datos de pacientes con diabetes moderada.

Pacientes con diabetes grave:

$\widehat{\beta}_{ZY}$	$\widehat{\beta}_{Z}$	$\widehat{\beta}_{N}$	$\widehat{\delta}_{x}$	$\widehat{\delta}_{y}$	$\sqrt{\widehat{\delta}_{x}^{2}+\widehat{\delta}_{y}^{2}}$
0.196	0.145	0.206	0.528	0.122	0.541

Tabla 2.16: Estimadores de β, $\widehat{\delta}_x$ y $\widehat{\delta}_y$ para los datos de pacientes con diabetes grave.

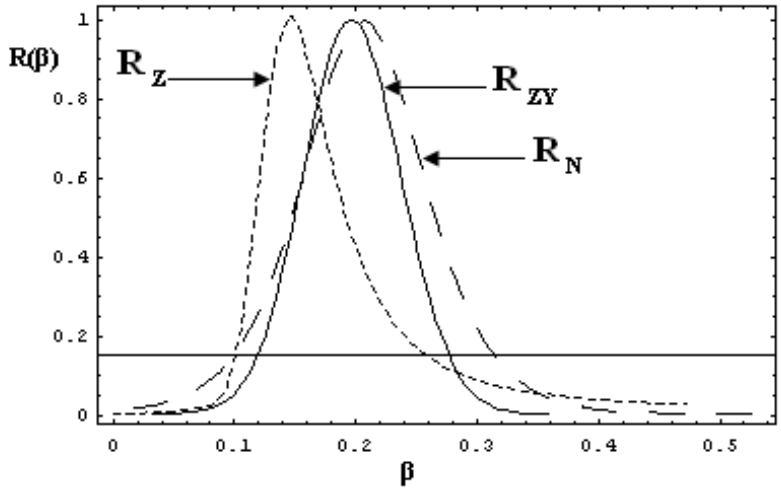


Figura 2.20: Verosimilitudes perfiles de β para los datos de pacientes con diabetes grave.

En los tres grupos se observa que se pierde información al utilizar el modelo marginal debido a que los valores de $\widehat{\delta}_y$ son mayores a 0.1; especialmente esto es más notorio en los pacientes normales. La estimación puntual de β con g_Z es marcadamente menor que con el modelo conjunto (Z, Y). La aproximación normal a g_Z no es razonable porque g_Z es asimétrica. Por tanto el modelo a usar, de los que se consideran aquí, es el conjunto (Z, Y) en los tres grupos de pacientes.

Los intervalos de verosimilitud del 15% para β son diferentes para los tres modelos estadísticos y lo son más para los pacientes normales. Además en este caso no es razonable utilizar la aproximación normal a g_Z.

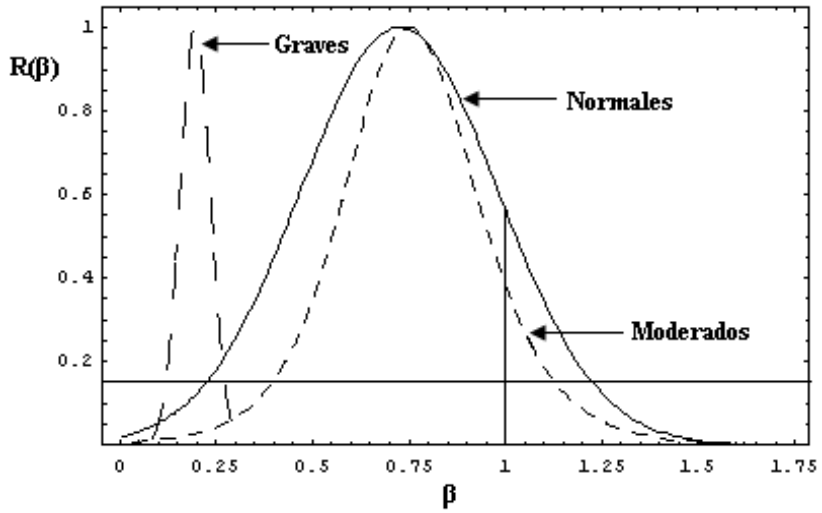


Figura 2.21: Verosimilitudes pefiles de β con el modelo conjunto para los tres grupos de pacientes.

La Figura 2.21 muestra las tres perfiles para β bajo el modelo conjunto, para los tres grupos de pacientes. Se observa que hay un traslape de las verosimilitudes perfiles de β de los pacientes normales y de los pacientes con diabetes moderada. Por lo tanto se observa que existe una respuesta pancreática temprana similar para el grupo pacientes normales y el grupo de pacientes con diabetes moderada, en cambio hay una diferencia relevante con el grupo de pacientes con diabetes grave. Se resalta el hecho que $\beta = 1$ es moderadamente plausible para los pacientes normales y moderados. Esto quiere decir que es plausible que $\Delta I = \Delta G$ en estos grupos de pacientes. Sin embargo, el valor de β es claramente menor en el grupo de pacientes con diabetes grave. Esto indica que para estos pacientes ΔI es menos de la cuarta parte de ΔG.

La Figura 2.21, muestra cómo las gráficas de las perfiles de β dan mucho mayor información que los resultados de un ANOVA, donde solamente se hubiese concluido que hay evidencia fuerte de que el índice insulinogénico no es igual en los tres grupos de pacientes. No se hubiera podido cuantificar tan sólo con el ANOVA la magnitud de las discrepancias ni cuál o cuáles de los grupos diferían. En contraste, las perfiles de β dan mucho mayor información como se mostró aquí. El grupo de pacientes con diabetes grave sí presenta una marcada diferencia en el índice insulinogénico con respecto a los otros

dos grupos.

2.6 Conclusiones

En todos los ejemplos simulados mostrados se observa que se obtienen mejores estimaciones sobre β con el modelo conjunto R_{ZY}. En el caso de contar con datos pareados experimentalmente como los ejemplos 7 y 8 es posible que otro modelo más estructurado, como el de Díaz-Francés y Sprott (2001) de una mejor respuesta porque describe mejor la génesis de los datos. Se resalta que si el tamaño de muestra aumentara, ambos modelos, el conjunto g_{ZY} y el (2.11) resultarían entonces equivalentes las perfiles de β asociadas.

Al utilizar el modelo marginal para estimar β se puede incurrir en una pérdida de información e incluso subestimar o sobre-estimar fuertemente el parámetro de interés como se mostró en el Ejemplo 8.

En el Ejemplo 8 se habría rechazado la igualdad de medias β con un ANOVA. En cambio, con las perfiles de β del modelo conjunto se puede decir mucho más sobre este parámetro para los tres grupos.

El interés relevante en biología, medicina y en otras áreas es realmente la estimación de β. Al realizar un ANOVA no se da respuesta a ello. En cambio al utilizar la verosimilitud perfil de β se pueden dar inferencias sobre este parámetro en términos de intervalos de verosimilitud-confianza bajo el modelo que resulte adecuado y además se pueden comparar las gráficas obtenidas para cada grupo.

Capítulo 3

Invarianza de las estimaciones de β ante reparametrizaciones y cambio de roles de las variables

3.1 Introducción

En esta sección se explorará la invarianza de la estimación de β y de $1/\beta$ con los tres diferentes modelos presentados anteriormente al cambiar de roles las variables X y Y. Hay que tener en cuenta que por consideraciones lógicas, las inferencias que se hagan sobre β deberían coincidir con las de $1/\beta$ pues guardan una relación uno a uno. Esta comparación se basa en la idea expresada por R. A. Fisher (1973, pp. 146):

"A primary, and really very obvious consideration is that if an unknown parameter θ is being estimated, any one-valued function of θ is necessarily being estimated by the same operation. The criteria used in the theory must, for this reason, be invariant for all such functional transformations of the parameters".

Por la propiedad de invarianza de la función de verosimilitud, para el modelo conjunto (Z, Y) las estimaciones puntuales de β y de $1/\beta$ son invariantes frente a reparametrizaciones. Además también lo serán ante el cambio de

roles de las variables X y Y. No ocurrirá lo mismo necesariamente con las estimaciones de β y de $1/\beta$ que se hagan con g_Z al cambiar el papel de las variables X y Y puesto que para este modelo es importante el rol que juegan. Aquí será importante la información sobre β que tenga el factor condicional $Y|Z$ en (1.4). Mientras menor sea la información, mayor será la invarianza en las inferencias sobre β. Tampoco será invariante la aproximación normal a g_Z. Aún cuando la aproximación normal a g_Z fuera buena, dependería de que g_Z contuviera casi toda la información sobre β para que pueda ser invariante ante el cambio de roles de X y Y.

Así, si se desea estimar $1/\beta = \mu_y/\mu_x$ mediante verosimilitud, podemos utilizar dos métodos:

1. El primero consiste en estimar $\beta = \mu_x/\mu_y$ y utilizar la propiedad de invarianza de la función de verosimilitud y de los EMV, aplicando la reparametrización uno a uno $\varphi(\beta) = 1/\beta$, para obtener así una estimación de μ_y/μ_x. Esto equivale a reescalar el eje horizontal de las verosimilitudes perfiles de β mediante la función φ, para así obtener la gráfica de la verosimilitud perfil de $1/\beta$. Se denotarán las verosimilitudes perfiles de $1/\beta$ obtenidas con este primer método como: $R^{-1}_{ZY}(\beta)$ para el modelo (1.3), R^{-1}_Z para el modelo marginal (1.5) y R^{-1}_N para la aproximación normal (1.7).

2. El segundo método consiste en intercambiar los papeles de las variables X y Y, esto es, utilizar $Z^*= Y/X$ y $Y^*= X$, con lo cual se obtienen modelos en términos de $1/\beta = \mu_y/\mu_x$. Se denotarán las verosimilitudes perfil correspondientes como: $R_{ZX^*}(\beta)$ para el modelo (Z^*, X), $R_{Z^*}(\beta)$ para el modelo $g_Z(z^*)$ y $R_{N^*}(\beta)$ para la aproximación normal a Z^* (1.7).

Si se llega a inferencias semejantes sobre $1/\beta$ con ambos caminos, se dirá que hay invarianza en la estimación de β. A continuación se retomarán cuatro ejemplos del capítulo anterior en los que se analizará la invarianza de

las inferencias sobre β utilizando los tres modelos estadísticos con la notación descrita en el párrafo anterior.

3.2 Ejemplos sobre la invarianza de las inferencias de β y de $1/\beta$

Ejemplo 1

Este es un caso donde $g_Z(z;\beta,\rho,\delta_y)$ es simétrica y la aproximación normal $f_N(z;\overline{z},\widehat{\sigma})$ es buena puesto que δ_y es pequeño. Recuérdese que estos datos fueron simulados como $x_i \sim N(50, 2.5)$, $y_i \sim N(80, 2)$. Este es también un caso donde existe invarianza en la estimación de β al cambiar los roles de las variables X y Y bajo todos los modelos. Los parámetros estimados con ambos métodos son similares:

$\widehat{\beta}_{ZY}^{-1}$	$\widehat{\beta}_{Z}^{-1}$	$\widehat{\beta}_{N}^{-1}$
1.5851	1.5857	1.5846
$\widehat{\beta}_{ZX^*}$	$\widehat{\beta}_{Z^*}$	$\widehat{\beta}_{N^*}$
1.5853	1.5857	1.5897

Tabla 3.17: Estimadores de β para los datos del Ejemplo 1.

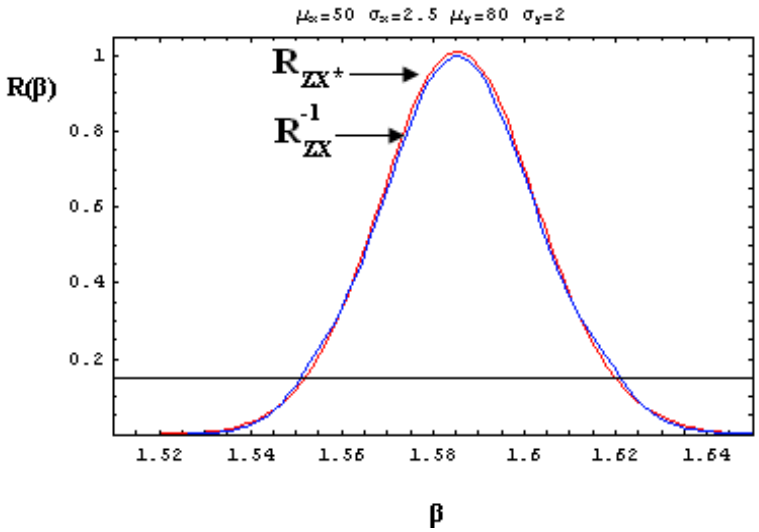


Figura 3.1: Verosimilitudes perfiles de $1/\beta$ con el modelo conjunto.

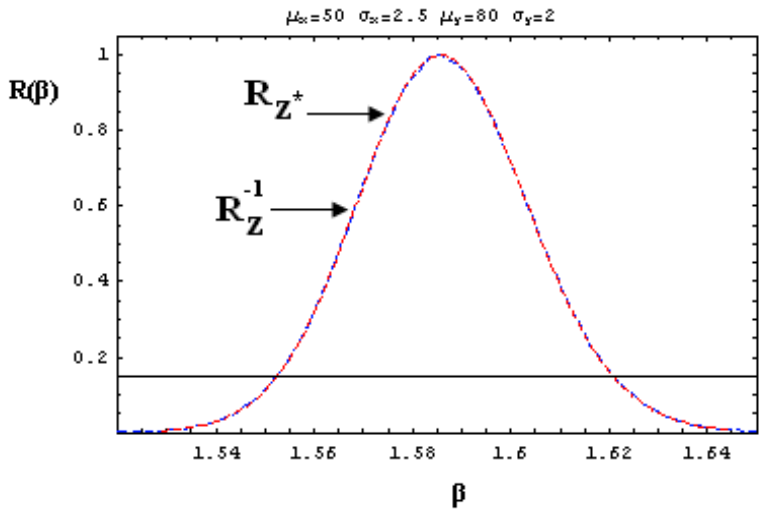


Figura 3.2: Verosimilitudes perfiles de $1/\beta$ con el modelo marginal.

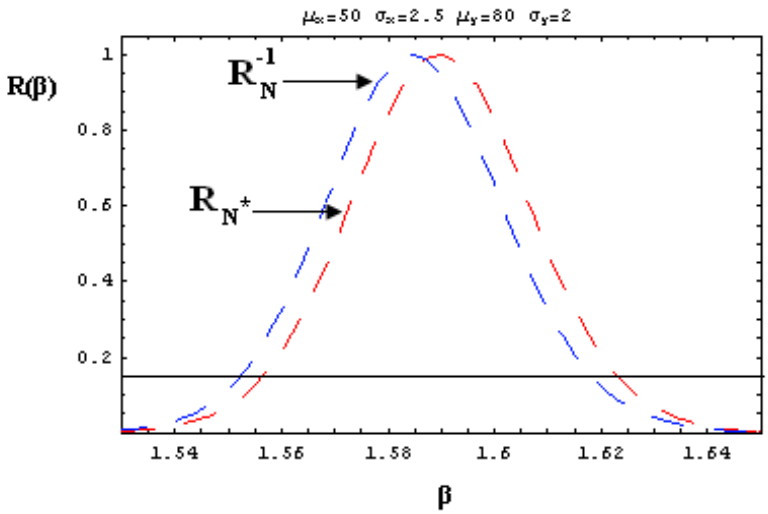


Figura 3.3: Verosimilitudes perfiles de $1/\beta$ con la aproximación normal.

Aquí los estimadores de máxima verosimilitud de β son prácticamente el mismo para los tres modelos estadísticos pero la aproximación normal a la densidad de Z no es del todo invariante en la inferencia sobre β, lo cual se ve reflejado en las verosimilitudes, que están ligeramente trasladadas. En contraste la estimación con el modelo conjunto y con el marginal sí son invariantes en las inferencias sobre β en términos de estimación puntual y también de intervalos de verosimilitud confianza.

Ejemplo 3

Se retoman los datos del Ejemplo 3 del capítulo anterior. Este es un caso donde g_Z es asimétrica y no es razonable usar las aproximaciones normales. En este caso no existe invarianza en la inferencia de β cuando se usa la aproximación normal a la densidad de Z. Sin embargo, los parámetros estimados están en el rango recomendado por Qiao et. al. (2006) para utilizar $\overline{z}$ como estimador de $\overline{x}/\overline{y}$ y se observa que $\widehat{\beta}_{ZX^*}$ es cercano a $\widehat{\beta}_{Z^*}$. Los parámetros estimados con ambos métodos son:

$\widehat{\beta}_{ZY}^{-1}$	$\widehat{\beta}_{Z}^{-1}$	$\widehat{\beta}_{N}^{-1}$
1.009	1.0201	0.9652
$\widehat{\beta}_{ZX^*}$	$\widehat{\beta}_{Z^*}$	$\widehat{\beta}_{N^*}$
1.009	1.0202	1.0359

Tabla 3.18: Estimadores de β para los datos del Ejemplo 2.

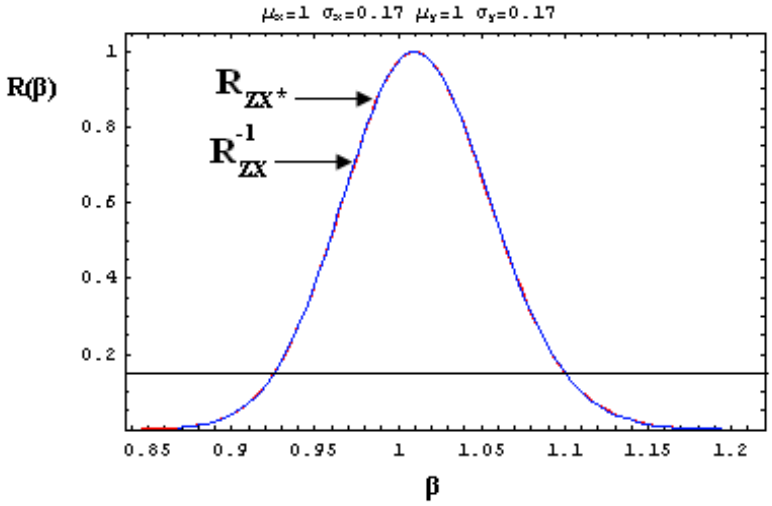


Figura 3.4: Verosimilitudes perfiles de $1/\beta$ con el modelo conjunto .

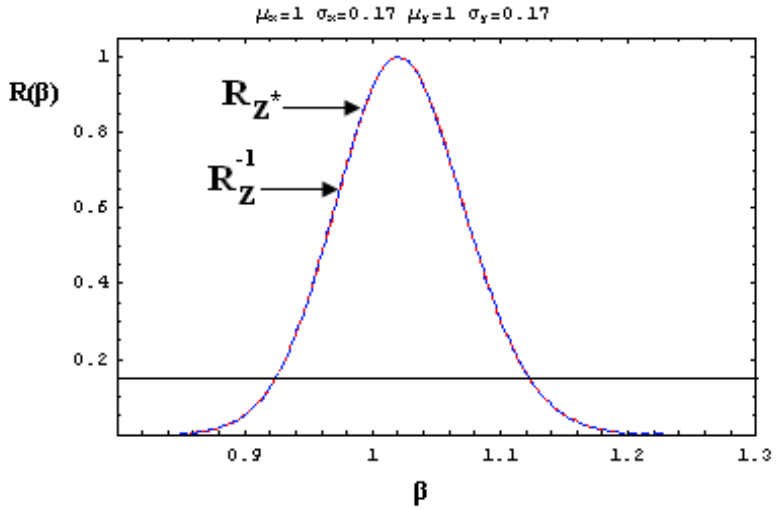


Figura 3.5: Verosimilitudes perfiles de $1/\beta$ con el modelo marginal .

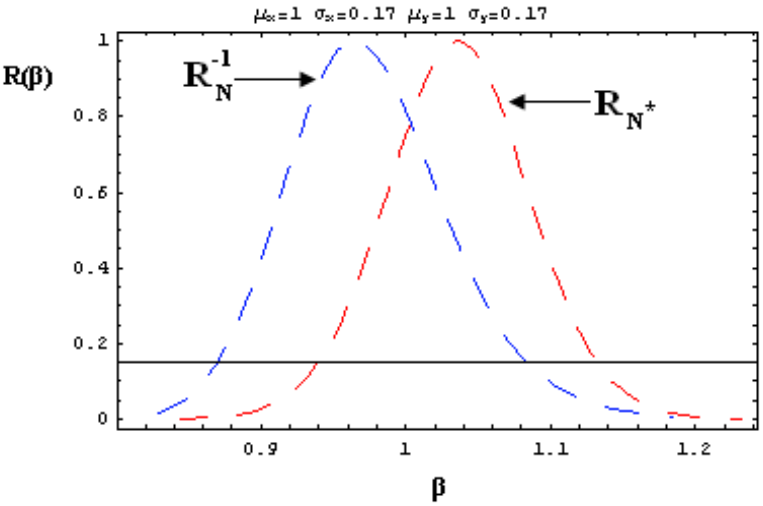


Figura 3.6: Verosimilitudes perfilesde $1/\beta$ con la aproximación normal.

Aquí se observa que la recomendación de Qiao de utilizar $\overline{z}$ para aproximar $\overline{x}/\overline{y}$ si $\widehat{\delta}_y < 0.2$ debe tomarse con precaución si lo que se desea es hacer inferencia con intervalos sobre β. Aquí se cumple esta restricción y sin embargo no existe invarianza en la inferencia por intervalos sobre β para la aproximación normal como se muestra en la Figura 3.6, a pesar de que los estimadores $\widehat{\beta}_{ZX^*}$ y $\widehat{\beta}_{Z^*}$ sean cercanos. Nuevamente aquí se recomienda usar el modelo conjunto.

Ejemplo 6

En este ejemplo se utilizarán los datos del Ejemplo 6 del capítulo anterior donde se simularon datos $x_i \sim N(80, 2)$, $y_i \sim N(1.66, 48.33)$. El valor teórico del parámetro de interés es $\beta = 48$ y el coeficiente de variación teórico de Y es $\delta_y = 29$. En este caso la densidad de $Z = X/Y$ es bimodal, por lo tanto ninguna aproximación normal es razonable. Los parámetros estimados con ambos métodos son:

$\widehat{\beta}_{ZY}^{-1}$	$\widehat{\beta}_{Z}^{-1}$	$\widehat{\beta}_{N}^{-1}$
0.1205	0.1122	0.6112
$\widehat{\beta}_{ZX^*}$	$\widehat{\beta}_{Z^*}$	$\widehat{\beta}_{N^*}$
0.1205	0.1205	0.1374

Tabla 3.19: Estimadores de β para los datos del Ejemplo 6.

El modelo conjunto presenta invarianza en las inferencias puntuales como era de esperarse pero todas ellas son sumamente malas y distantes del valor verdadero de $\beta = 48$. Debido a que la variable Y tiene una varianza muy grande en contraste a la de X, se tiene que el tamaño de muestra es muy pequeño para tener una buena estimación de los parámetros. En Qiao et. al. (2006) mencionan que es raro encontrar en aplicaciones una variable aleatoria normal con coeficiente de variación mayor a 5. Aquí se muestra que en los casos en que hay bimodalidad se tiene el problema adicional de tener varianzas grandes y requerir de un mayor tamaño de muestra para obtener inferencias razonables, incluso bajo el modelo conjunto que es el que se recomienda considerar para hacer inferencias sobre β. Este es un caso donde las inferencias sobre β son pésimas incluso con el modelo conjunto. Por tanto, en aquellos casos que el estimador $\widehat{\delta}_y$ sea grande, es muy importante contar con un buen tamaño de muestra y no usar a g_Z para hacer inferencia sobre β sino el modelo conjunto (Z, Y).

Ejemplo 7

Se utilizaron los datos descritos en el Ejemplo 7 de la Sección 3.2 solamente para la posición Interna Basal de las plantas de agave, a manera de ejemplificar si las inferencias sobre β basadas en las aproximaciones normales eran invariantes frente al cambio de roles de las variables. Aquí se está ignorando el hecho de que el citometrista controla la escala en la que se miden las observaciones, a diferencia del modelo de Díaz-Francés y Sprott (2001).

En este caso existe invarianza en la estimación de β para los tres modelos considerados porque $\widehat{\delta}_x$ y $\widehat{\delta}_y$ son muy pequeños y $\sqrt{\widehat{\delta}_x^2 + \widehat{\delta}_y^2} < 1/3$. Sin embargo, tanto g_Z como la aproximación normal producen sobre-precisión en los intervalos de verosimilitud que son mucho más angostos que los intervalos que se obtienen a partir del modelo conjunto. Los intervalos de 15% de verosimilitud se marcan en la Figura 3.7 donde se muestran las seis verosimilitudes perfiles de $1/\beta$ sobrepuestas, utilizando ambos métodos para los tres modelos estadísticos.

Esto indica que además de checar que los valores de $\widehat{\delta}_x$ y $\widehat{\delta}_y$ estén en el rango recomendado, también es necesario revisar que el tamaño de muestra no sea muy pequeño si se va a usar una aproximación normal. Los parámetros estimados con ambos métodos son:

$\widehat{\beta}_{ZY}^{-1}$	$\widehat{\beta}_{Z}^{-1}$	$\widehat{\beta}_{N}^{-1}$
1.5434	1.5439	1.5438
$\widehat{\beta}_{ZX^*}$	$\widehat{\beta}_{Z^*}$	$\widehat{\beta}_{N^*}$
1.5434	1.5438	1.5441

Tabla 3.20: Estimadores de β para los datos del Ejemplo 7.

Obsérvese que los estimadores puntuales de β son prácticamente invariantes, pero, nótese que las curvas perfiles del modelo conjunto discrepan en la apertura y precisión que dan sobre β con respecto a las del modelo marginal de Z y su aproximación normal. En otros ejemplos simulados se observó que al aumentar los tamaños de muestras en situaciones similares, las curvas que aquí discrepan, se acercaban.

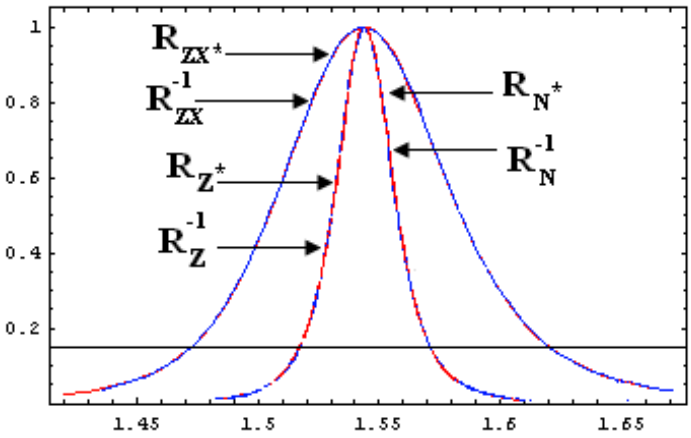


Figura 3.7: Verosimilitudes perfiles de $1/\beta$ con los tres modelos .

Ejemplo 8.

Se retomarán los datos del Ejemplo 8.

Se utilizará este conjunto de datos para evaluar la invarianza en las inferencias sobre $\Delta I/\Delta G$ para solamente uno de los grupos de pacientes, los pacientes con diabetes moderada. Para esto se cambiarán los roles de las variables X y Y, y entonces se aplicarán los dos métodos descritos en la Sección 4.1.

A continuación se muestran las gráficas de las verosimilitudes perfiles de β para los tres modelos presentados en el Capítulo 1 utilizando los dos métodos descritos en la Sección 4.1 para el conjunto de datos de pacientes con diabetes moderada:

$\widehat{\beta}^{-1}_{ZY}$	$\widehat{\beta}^{-1}_{Z}$	$\widehat{\beta}^{-1}_{N}$
0.7497	0.6026	0.6026
$\widehat{\beta}_{ZX^*}$	$\widehat{\beta}_{Z^*}$	$\widehat{\beta}_{N^*}$
0.7497	0.6026	0.7305

Tabla 21: Estimadores de β para los datos de pacientes con diabetes moderada.

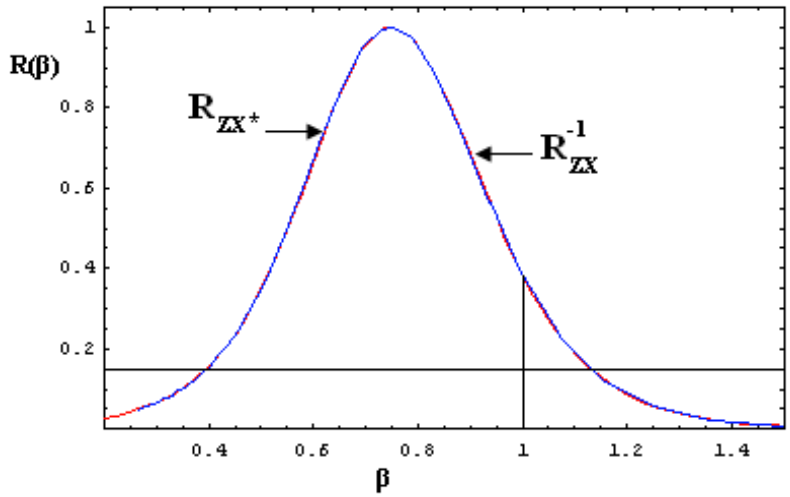


Figura 3.8: Verosimilitudes perfiles de β con el modelo conjunto.

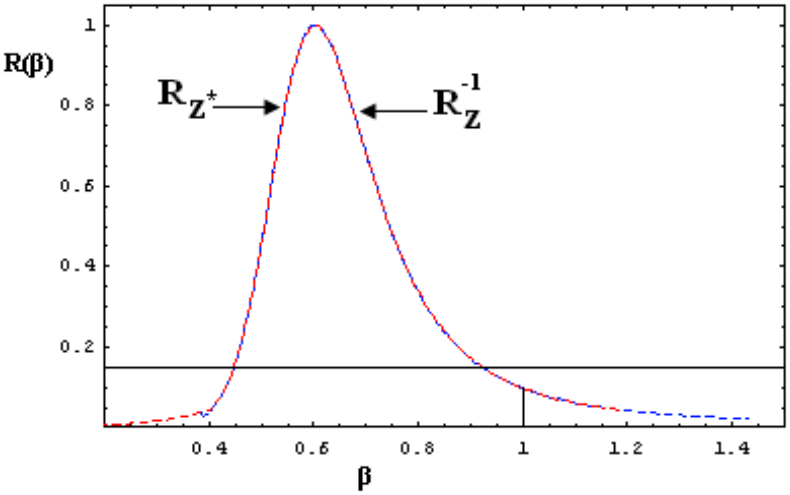


Figura 3.9: Verosimilitudes perfiles de β con el modelo marginal.

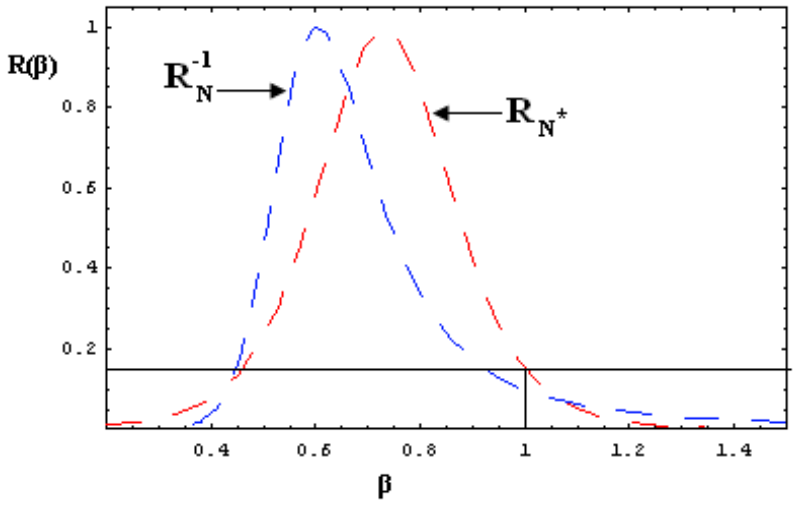


Figura 3.10: Verosimilitudes perfiles de β con la aproximación normal.

Para este conjunto de datos no existe invarianza en la inferencia de $\Delta I/\Delta G$ utilizando la aproximación normal. Además la densidad de la variable Z es asimétrica, por lo tanto no es razonable de entrada utilizar una aproximación normal. En cambio, sí hay invarianza como era de esperarse para los modelos (Z, Y) y el marginal de Z.

Se observa que con el modelo marginal g_Z y también consecuentemente con la aproximación normal se tiende a subestimar el índice insulinogénico. De manera especial este modelo asocia una plausibilidad bajísima al valor $\beta = 1$ en comparación a lo que se tiene con el modelo conjunto (Z, Y).

3.3 Conclusiones

Cuando se usa el modelo (Z, Y) para hacer inferencias sobre β, a través de la verosimilitud perfil y de intervalos de verosimilitud-confianza, dichas inferencias serán invariantes frente a reparametrizaciones uno a uno y frente al cambio de roles de las variables X y Y. Solamente cuando las aproximaciones normales sean razonables para g_Z será que todas las inferencias sobre β bajo los distintos modelos considerados coincidirán. Cuando la densidad g_Z es asimétrica o bimodal, las inferencias sobre β hechas con este modelo no son invariantes para estimar $1/\beta$.

Capítulo 4

Conclusiones generales

En esta tesis se caracterizó la forma de la densidad del cociente $Z = X/Y$ de dos variables aleatorias normales independientes X y Y en términos de los parámetros, especialmente en términos de los coeficientes de variación δ_x y δ_y. Dicha variable Z no tiene momentos finitos y puede ser marcadamente asimétrica, simétrica, con colas pesadas e incluso bimodal. Se demostró analíticamente bajo cuáles condiciones en δ_x y δ_y es posible aproximar dicha densidad con la de una normal, en un intervalo centrado en el parámetro de interés $\beta = E(X)/E(Y)$.

En la literatura de Citometría de Flujo, tradicionalmente se ha supuesto que Z sigue una distribución normal, se estima con esta densidad el cociente medias β de manera puntual y se utilizan Análisis de Varianzas para comparar los valores de β en varios grupos poblacionales. En este trabajo se cuantificó el error de tales procedimientos, dependiendo de condiciones en δ_x y δ_y. De manera más específica, se mostró que al considerar variables aleatorias normales independientes X y Y, suponer que el cociente $Z = X/Y$ es normal será razonable solamente cuando δ_x y δ_y sean menores a 0.1. En otros casos, la densidad de Z podría ser asimétrica o incluso bimodal.

Se mostró también cómo las gráficas de la verosimilitud perfil de β obtenidas a partir de un modelo adecuado, son sumamente informativas cuando el objetivo es realizar inferencias sobre este parámetro de interés y comparar los

valores de β en varios grupos poblacionales. En contraste, los resultados de un ANOVA y la estimación puntual de β con un modelo normal, como tradicionalmente se efectúa en áreas como la Citometría de Flujo y otras, son mucho menos informativas, pueden no ser aplicables y conducir a conclusiones equivocadas sobre el parámetro de interés.

Con el ANOVA sólo se puede concluir si hay evidencia en contra de la igualdad de medias. Cuando se rechaza la hipótesis nula de que las medias sean iguales no se sabe cuál o cuales grupos son los que discrepan en β, ni en qué magnitud. Por otra parte, cuando no se rechaza la hipótesis nula, no se sabe si verdaderamente β es común a todos los grupos o no; lo único cierto es que los datos no contienen evidencia suficientemente fuerte en contra de dicha hipótesis. En contraste, al utilizar la verosimilitud perfil de β se pueden dar estimaciones puntuales e intervalos de verosimilitud-confianza para β en cada grupo, lo cual es mucho más informativo ya que muestra la evidencia a favor de valores de β que está sustentada por los datos.

Se mostró además que el modelo marginal $g_Z(Z)$ puede no contener toda la información sobre β ya que pierde la información que contiene el modelo condicional $g(Y|Z)$ sobre β. Este modelo $g_Z(Z)$ tiene la ventaja de tener menos parámetros (tres) identificables que el modelo conjunto $g(Z, Y)$ que tiene cuatro parámetros; sin embargo g_Z presenta una mayor complejidad para maximizarse numéricamente y así poder hacer inferencia sobre β con enfoques de verosimilitud.

Para datos pareados se recomienda utilizar la verosimilitud perfil de β del modelo presentado en Díaz-Francés y Sprott (2001). En el caso de datos citométricos, este modelo considera todos los aspectos del manejo del citómetro para la obtención de los datos que involucra un parámetro adicional por cada pareja observada (x_i, y_i). En Citometría de Flujo, se ha usado la aproximación normal a g_Z para estimar β indiscriminadamente y sin tomar precaución alguna. Lo mínimo que se debe hacer antes de considerar esta aproximación es revisar las magnitudes de los coeficientes de variación estimados $\widehat{\delta}_x$ y $\widehat{\delta}_y$.

En el ejemplo que se mostró de agave, la aproximación normal fue razonable porque $\widehat{\delta}_x$ y $\widehat{\delta}_y$ fueron pequeños, pero en otros casos de Citometría bien podría no serlo por lo que hay que estar siempre al pendiente.

En la literatura estadística se ha mencionado que la densidad de $Z = X/Y$ es aproximadamente normal cuando el coeficiente de variación δ_y es pequeño. Esto se ha mostrado solamente de manera empírica o mediante simulaciones. En esta tesis se demuestra analíticamente que existe una aproximación normal razonable a g_Z en un intervalo centrado en β cuando los coeficientes de variación δ_x y δ_y son pequeños.

Al utilizar el modelo marginal $g_Z(Z)$ se tiene una asignación implícita de roles (numerador y denominador) a las variables X y Y. En esta tesis se muestra que esta asignación es crucial debido a que la variable en el denominador determina de manera importante, dependiendo de la magnitud de δ_y, la forma de la densidad de la variable Z y por tanto si una aproximación normal será razonable o no. Una aproximación normal sería adecuada sólo si ambos δ_x y δ_y son menores que 0.1, para que además las inferencias sobre β sean invariantes frente a un cambio de roles de las variables X y Y.

En general, se recomienda analizar cuidadosamente el contexto del fenómeno de interés y entender claramente el mecanismo aleatorio que generó a los datos para así poder elegir al mejor modelo para hacer inferencia sobre β, a través de la verosimilitud perfil correspondiente. En el caso de k poblaciones distintas, se recomienda graficar juntas las verosimilitudes perfiles de β de todos los grupos, bajo el modelo elegido, con fines comparativos y dar las inferencias sobre β en términos de intervalos de verosimilitud-confianza para cada grupo. Esto es mucho más informativo y pertinente que contar solamente con estimaciones puntuales o con los resultados de un ANOVA.

Apéndices

A1. Densidad de Z bajo independencia de X y Y

Se calculará la densidad del cociente $Z = X/Y$ de dos normales independientes X y Y marginalizando la densidad conjunta de Z y Y dada en (1.3) :

$$g_Z(z) = \int_{-\infty}^{\infty} g(z, y)\, dy,$$

Primeramente se separa la integral como sigue:

$$g_Z(z) = \int_{-\infty}^{0} g_{ZY}(z, y)\, dy + \int_{0}^{\infty} g_{ZY}(z, y)\, dy = I_1 + I_2.$$

Para simplificar notación se definirá:

$$A = \frac{1}{2\pi\sigma_x^2\rho} \exp\left[-\frac{(\beta^2\rho^2 + 1)}{2\delta_y^2}\right].$$

Entonces podemos reescribir I_1 como:

$$I_1 = \int_{-\infty}^{0} g_{ZY}(z, y)\, dy = -A \int_{-\infty}^{0} y \exp\left\{-\frac{1}{2}\left[\frac{y^2}{\sigma_x^2\rho^2}(z^2\rho^2 + 1) - \frac{2y}{\sigma_x\rho}\left(\frac{z\beta\rho^2 + 1}{\delta_y}\right)\right]\right\} dy.$$

Se utilizará el siguiente cambio de variable:

$$\begin{aligned} u &= \frac{y^2}{\sigma_x^2\rho^2}(z^2\rho^2 + 1) - \frac{2y}{\sigma_x\rho}\left(\frac{z\beta\rho^2 + 1}{\delta_y}\right), \\ du &= 2\left[y\left(\frac{z^2\rho^2 + 1}{\sigma_x^2\rho^2}\right) - \frac{z\beta\rho^2 + 1}{\sigma_x\rho\delta_y}\right]. \end{aligned}$$

Entonces se tiene que:

$$\begin{aligned} I_1 &= \int_{-\infty}^{0} g(z,y)\,dy = -\frac{A}{2\left(\frac{z^2\rho^2+1}{\sigma_x^2\rho^2}\right)}\int_{-\infty}^{0} 2\left[y\left(\frac{z^2\rho^2+1}{\sigma_x^2\rho^2}\right)-\frac{z\beta\rho^2+1}{\sigma_x\rho\delta_y}\right] \\ &\quad \times \exp\left\{-\frac{1}{2}\left[\frac{y^2}{\sigma_x^2\rho^2}(z^2\rho^2+1)-\frac{2y}{\sigma_x\rho}\left(\frac{z\beta\rho^2+1}{\delta_y}\right)\right]\right\}dy \\ &\quad -\frac{A}{2\left(\frac{z^2\rho^2+1}{\sigma_x^2\rho^2}\right)}\int_{-\infty}^{0} 2\left(\frac{z\beta\rho^2+1}{\sigma_x\rho\delta_y}\right) \\ &\quad \times \exp\left\{-\frac{1}{2}\left[\frac{y^2}{\sigma_x^2\rho^2}(z^2\rho^2+1)-\frac{2y}{\sigma_x\rho}\left(\frac{z\beta\rho^2+1}{\delta_y}\right)\right]\right\}dy \\ &= I_{11}+I, \end{aligned}$$

donde:

$$\begin{aligned} I_{11} &= -\frac{A}{2\left(\frac{z^2\rho^2+1}{\sigma_x^2\rho^2}\right)}\int_{\infty}^{0}\exp\left(-\frac{u}{2}\right)du \\ &= \frac{1}{2\pi\sigma_x^2\rho}\exp\left[-\frac{(\beta^2\rho^2+1)}{2\delta_y^2}\right]\left(\frac{\sigma_x^2\rho^2}{z^2\rho^2+1}\right) \\ &= \exp\left(-\frac{1+\beta^2\rho^2}{2\delta_y^2}\right)\frac{\rho}{2\pi(1+\rho^2z^2)}, \end{aligned}$$

y

$$\begin{aligned} I_{12} &= -\frac{A}{2\left(\frac{z^2\rho^2+1}{\sigma_x^2\rho^2}\right)}\int_{-\infty}^{0} 2\left(\frac{z\beta\rho^2+1}{\sigma_x\rho\delta_y}\right) \\ &\quad \times \exp\left\{-\frac{1}{2}\left[\frac{y^2}{\sigma_x^2\rho^2}(z^2\rho^2+1)-\frac{2y}{\sigma_x\rho}\left(\frac{z\beta\rho^2+1}{\delta_y}\right)\right]\right\}dy. \end{aligned}$$

Conviene definir las siguientes variables para simplificar notación:

$$\begin{aligned} \theta &= \frac{1}{2}\left(\frac{z^2\rho^2+1}{\sigma_x^2\rho^2}\right), \\ \lambda &= \left(\frac{z\beta\rho^2+1}{\sigma_x\rho\delta_y}\right), \\ M &= A\left(\frac{\sigma_x^2\rho^2}{z^2\rho^2+1}\right)\left(\frac{z\beta\rho^2+1}{\delta_y}\right). \end{aligned}$$

Entonces:

$$I_{12}= -M\int_{-\infty}^{0}\exp\left(-\theta y^2+\lambda y\right)dy = -M\exp\left(\frac{\lambda^2}{4\theta}\right)\int_{-\infty}^{0}\exp\left[-\theta\left(y-\frac{2\lambda}{\theta}\right)^2\right]dy.$$

Realizando el cambio de variable $u=\sqrt{\theta}\left(y-\frac{2\lambda}{\theta}\right)$, se tiene que:

$$\begin{aligned} I_{12} &= -\frac{M}{\sqrt{\theta}}\exp\left(\frac{\lambda^2}{4\theta}\right)\int_{-\infty}^{-\frac{\lambda}{2\sqrt{\theta}}} e^{-u^2}du \\ &= -\frac{M}{\sqrt{\theta}}\exp\left(\frac{\lambda^2}{4\theta}\right)\int_{\frac{\lambda}{2\sqrt{\theta}}}^{\infty} e^{-u^2}du \\ &= -\frac{M}{\sqrt{\theta}}\exp\left(\frac{\lambda^2}{4\theta}\right)\frac{\sqrt{\pi}}{2}\operatorname{erfc}\left(\frac{\lambda}{2\sqrt{\theta}}\right), \end{aligned}$$

donde $\operatorname{erfc}(x)=1-\operatorname{erf}(x)$. Sustituyendo los valores de θ,λ y M se tiene finalmente que:

$$I_{12}= -\frac{1}{2}\frac{\rho(1+\beta\rho^2 z)}{\sqrt{2\pi}\delta_y(1+\rho^2z^2)^{3/2}}\exp\left[-\frac{\rho^2(z-\beta)^2}{2\delta_y^2(1+\rho^2z^2)}\right]\operatorname{erfc}\left[\frac{1+\beta\rho^2z}{\delta_y\sqrt{2(1+\rho^2z^2)}}\right].$$

Es conveniente separar la integral I_2 como:

$$I_2= \int_0^{\infty} g(z,y)\,dy = \int_0^{x} g(z,y)\,dy + \int_x^{\infty} g(z,y)\,dy = I_{21}+I_{22},$$

donde:

$$x = \frac{\sigma_x\rho\left(z\beta\rho^2+1\right)}{\delta_y\left(z^2\rho^2+1\right)}.$$

Usaremos la siguiente relación:

$$\frac{y^2}{\sigma_x^2\rho^2}(z^2\rho^2+1)-\frac{2y}{\sigma_x\rho}\left(\frac{z\beta\rho^2+1}{\delta_y}\right)=\left(\frac{y\sqrt{z^2\rho^2+1}}{\sigma_x\rho}-\frac{z\beta\rho^2+1}{\delta_y\sqrt{z^2\rho^2+1}}\right)^2-\left(\frac{z\beta\rho^2+1}{\delta_y\sqrt{z^2\rho^2+1}}\right)^2.$$

Entonces para la integral I_{21} se tiene que:

$$I_{21}= \int_0^{x} g(z,y)\,dy = A\exp\left[\frac{1}{2}\frac{\left(z\beta\rho^2+1\right)^2}{\delta_y^2\left(z^2\rho^2+1\right)}\right]\int_0^{x}\exp\left[-\frac{1}{2}\left(\frac{y\sqrt{z^2\rho^2+1}}{\sigma_x\rho}-\frac{z\beta\rho^2+1}{\delta_y\sqrt{z^2\rho^2+1}}\right)^2\right].$$

Haciendo el cambio de variable:

$$V = \frac{y\sqrt{z^2\rho^2+1}}{\sigma_x\rho},$$

se tiene que:

$$I_{21} = A\exp\left[\frac{1}{2}\frac{(z\beta\rho^2+1)^2}{\delta_y^2(z^2\rho^2+1)}\right]\int_0^{x^*}\frac{\sigma_x\rho}{\sqrt{z^2\rho^2+1}}V\exp\left[-\frac{1}{2}\left(V-\frac{z\beta\rho^2+1}{\delta_y\sqrt{z^2\rho^2+1}}\right)^2\right]\frac{\sigma_x\rho}{\sqrt{z^2\rho^2+1}}dV,$$

donde:

$$x^* = \frac{\sqrt{z^2\rho^2+1}}{\sigma_x\rho}x.$$

Por simplificación se definirá:

$$\begin{aligned} B &= A\exp\left[\frac{1}{2}\frac{(z\beta\rho^2+1)^2}{\delta_y^2(z^2\rho^2+1)}\right] \\ &= \frac{1}{2\pi\sigma_x^2\rho}\exp\left[-\frac{\rho^2(z-\beta)^2}{2\delta_y^2(1+\rho^2z^2)}\right]. \end{aligned}$$

Haciendo ahora el cambio de variable:

$$W = V - \frac{z\beta\rho^2+1}{\delta_y\sqrt{z^2\rho^2+1}},$$

se tiene que:

$$I_{21} = B\frac{\sigma_x^2\rho^2}{(z^2\rho^2+1)}\int_{-x^{**}}^{0}(W+x^{**})\exp\left(-\frac{1}{2}W^2\right)dW,$$

donde:

$$x^{**} = \frac{z\beta\rho^2+1}{\delta_y\sqrt{z^2\rho^2+1}}.$$

Hacemos el cambio de variable $T = -W$ y se tiene que:

$$
\begin{aligned}
I_{21} &= B\frac{\sigma_x^2\rho^2}{(z^2\rho^2+1)}\int_0^{x^{**}}(x^{**}-T)\exp\left(-\frac{1}{2}T^2\right)dT \\
&= B\frac{\sigma_x^2\rho^2 x^{**}}{(z^2\rho^2+1)}\int_0^{x^{**}}\exp\left(-\frac{1}{2}T^2\right)dT - B\frac{\sigma_x^2\rho^2}{(z^2\rho^2+1)}\int_0^{x^{**}}T\exp\left(-\frac{1}{2}T^2\right)dT.
\end{aligned}
$$

Haciendo el cambio de variable $t=\frac{T}{\sqrt{2}}$, se tiene que para la primera integral:

$$
\begin{aligned}
&B\frac{\sigma_x^2\rho^2 x^{**}}{(z^2\rho^2+1)}\int_0^{x^{**}}\exp\left(-\frac{1}{2}T^2\right)dT = \\
&B\frac{\sigma_x^2\rho^2 x^{**}}{(z^2\rho^2+1)}\sqrt{\frac{\pi}{2}}\,\mathrm{erf}\left(\frac{x^{**}}{\sqrt{2}}\right) = \\
&B\frac{\sigma_x^2\rho^2}{(z^2\rho^2+1)}\frac{z\beta\rho^2+1}{\delta_y\sqrt{z^2\rho^2+1}}\sqrt{\frac{\pi}{2}}\,\mathrm{erf}\left(\frac{z\beta\rho^2+1}{\delta_y\sqrt{2\left(z^2\rho^2+1\right)}}\right) = \\
&\frac{1}{2}\frac{\rho(1+\beta\rho^2 z)}{\sqrt{2\pi}\delta_y(1+\rho^2z^2)^{3/2}}\exp\left[-\frac{\rho^2(z-\beta)^2}{2\delta_y^2(1+\rho^2z^2)}\right]\mathrm{erf}\left[\frac{1+\beta\rho^2 z}{\delta_y\sqrt{2(1+\rho^2z^2)}}\right].
\end{aligned}
$$

Para la segunda integral se tiene que:

$$
\begin{aligned}
&B\frac{\sigma_x^2\rho^2}{(z^2\rho^2+1)}\int_0^{x^{**}}T\exp\left(-\frac{1}{2}T^2\right)dT = \\
&B\frac{\sigma_x^2\rho^2}{(z^2\rho^2+1)}\left[1-\exp\left(x^{**}\right)\right] = \\
&B\frac{\sigma_x^2\rho^2}{(z^2\rho^2+1)}\left\{1-\exp\left[-\frac{1}{2}\frac{\left(z\beta\rho^2+1\right)^2}{\delta_y^2\left(z^2\rho^2+1\right)}\right]\right\} = \\
&\frac{1}{2\pi\sigma_x^2\rho}\exp\left[-\frac{\rho^2(z-\beta)^2}{2\delta_y^2(1+\rho^2z^2)}\right]\frac{\sigma_x^2\rho^2}{(z^2\rho^2+1)}\left\{1-\exp\left[-\frac{1}{2}\frac{\left(z\beta\rho^2+1\right)^2}{\delta_y^2\left(z^2\rho^2+1\right)}\right]\right\} = \\
&\frac{1}{2}\frac{\rho}{\pi(1+\rho^2z^2)}\left\{\exp\left[-\frac{\rho^2(z-\beta)^2}{2\delta_y^2(1+\rho^2z^2)}\right]-\exp\left(\frac{1+\beta^2\rho^2}{2\delta_y^2}\right)\right\}.
\end{aligned}
$$

Por lo tanto:

$$
\begin{aligned}
I_{22} &= \int_x^\infty g(z,y)\,dy \\
&= \frac{1}{2\pi\sigma_x^2\rho}\exp\left[-\frac{(\beta^2\rho^2+1)}{2\delta_y^2}\right]\int_x^\infty y\exp\left\{-\frac{1}{2}\left[\frac{y^2}{\sigma_x^2\rho^2}(z^2\rho^2+1)\frac{2y}{\sigma_x\rho}\left(\frac{z\beta\rho^2+1}{\delta_y}\right)\right]\right\}dy \\
&= A\int_x^\infty y\exp\left\{-\frac{1}{2}\left[\frac{y^2}{\sigma_x^2\rho^2}(z^2\rho^2+1)-\frac{2y}{\sigma_x\rho}\left(\frac{z\beta\rho^2+1}{\delta_y}\right)\right]\right\}dy.
\end{aligned}
$$

Ahora, realizando el cambio de variable:

$$V = \frac{y\sqrt{z^2\rho^2+1}}{\sigma_x\rho},$$

se tiene que:

$$I_{22} = A\exp\left[\frac{1}{2}\frac{(z\beta\rho^2+1)^2}{\delta_y^2(z^2\rho^2+1)}\right]\int_{x^*}^{\infty}\frac{\sigma_x\rho}{\sqrt{z^2\rho^2+1}}V\exp\left[-\frac{1}{2}\left(V-\frac{z\beta\rho^2+1}{\delta_y\sqrt{z^2\rho^2+1}}\right)^2\right]\frac{\sigma_x\rho}{\sqrt{z^2\rho^2+1}}dV,$$

donde:

$$x^* = \frac{\sqrt{z^2\rho^2+1}}{\sigma_x\rho}x.$$

Sea:

$$\begin{aligned} B &= A\exp\left[\frac{1}{2}\frac{(z\beta\rho^2+1)^2}{\delta_y^2(z^2\rho^2+1)}\right] \\ &= \frac{1}{2\pi\sigma_x^2\rho}\exp\left[-\frac{\rho^2(z-\beta)^2}{2\delta_y^2(1+\rho^2z^2)}\right]. \end{aligned}$$

Haciendo ahora el cambio de variable:

$$W = V - \frac{z\beta\rho^2+1}{\delta_y\sqrt{z^2\rho^2+1}},$$

se tiene que:

$$I_{22} = B\frac{\sigma_x^2\rho^2}{(z^2\rho^2+1)}\int_{\widetilde{x}}^{\infty}(x^{**}+W)\exp\left(-\frac{1}{2}W^2\right)dW,$$

donde:

$$\begin{aligned} x^{**} &= \frac{z\beta\rho^2+1}{\delta_y\sqrt{z^2\rho^2+1}}, \\ \widetilde{x} &= x^* - x^{**} = 0. \end{aligned}$$

Haciendo el cambio de variable:

$$t = \frac{T}{\sqrt{2}},$$

se tiene para la primera integral:

$$\begin{aligned}
& B\frac{\sigma_x^2\rho^2 x^{**}}{(z^2\rho^2+1)}\int_0^\infty \exp\left(-\frac{1}{2}T^2\right)dT \\
= \;& B\frac{\sigma_x^2\rho^2 x^{**}}{(z^2\rho^2+1)}\sqrt{\frac{\pi}{2}}\,\mathrm{erf}\,c\,(0) \\
= \;& B\frac{\sigma_x^2\rho^2}{(z^2\rho^2+1)}\frac{z\beta\rho^2+1}{\delta_y\sqrt{z^2\rho^2+1}}\sqrt{\frac{\pi}{2}} \\
= \;& \frac{1}{2}\frac{\rho(1+\beta\rho^2 z)}{\sqrt{2\pi}\delta_y(1+\rho^2z^2)^{3/2}}\exp\left[-\frac{\rho^2(z-\beta)^2}{2\delta_y^2(1+\rho^2z^2)}\right].
\end{aligned}$$

Para la segunda integral:

$$\begin{aligned}
& B\frac{\sigma_x^2\rho^2}{(z^2\rho^2+1)}\int_0^\infty W\exp\left(-\frac{1}{2}W^2\right)dW \\
= \;& B\frac{\sigma_x^2\rho^2}{(z^2\rho^2+1)} \\
= \;& \frac{\rho}{2\pi(1+\rho^2z^2)}\exp\left[-\frac{\rho^2(z-\beta)^2}{2\delta_y^2(1+\rho^2z^2)}\right].
\end{aligned}$$

Así, finalmente retomando todas las expresiones se obtiene:

$$
\begin{aligned}
g_Z(z) &= I_{11} + I_{12} + I_{21} + I_{22} \\
&= \exp\left(-\frac{1+\beta^2\rho^2}{2\delta_y^2}\right)\frac{\rho}{2\pi(1+\rho^2z^2)} \\
&\quad -\frac{1}{2}\frac{\rho(1+\beta\rho^2 z)}{\sqrt{2\pi}\delta_y(1+\rho^2z^2)^{3/2}}\exp\left[-\frac{\rho^2(z-\beta)^2}{2\delta_y^2(1+\rho^2z^2)}\right]\operatorname{erfc}\left[\frac{1+\beta\rho^2 z}{\delta_y\sqrt{2(1+\rho^2z^2)}}\right] \\
&\quad +\frac{1}{2}\frac{\rho(1+\beta\rho^2 z)}{\sqrt{2\pi}\delta_y(1+\rho^2z^2)^{3/2}}\exp\left[-\frac{\rho^2(z-\beta)^2}{2\delta_y^2(1+\rho^2z^2)}\right]\operatorname{erf}\left[\frac{1+\beta\rho^2 z}{\delta_y\sqrt{2(1+\rho^2z^2)}}\right] \\
&\quad -\frac{1}{2}\frac{\rho}{\pi(1+\rho^2z^2)}\left\{\exp\left[-\frac{\rho^2(z-\beta)^2}{2\delta_y^2(1+\rho^2z^2)}\right]-\exp\left(\frac{1+\beta^2\rho^2}{2\delta_y^2}\right)\right\} \\
&\quad +\frac{1}{2}\frac{\rho(1+\beta\rho^2 z)}{\sqrt{2\pi}\delta_y(1+\rho^2z^2)^{3/2}}\exp\left[-\frac{\rho^2(z-\beta)^2}{2\delta_y^2(1+\rho^2z^2)}\right] \\
&\quad +\frac{\rho}{2\pi(1+\rho^2z^2)}\exp\left[-\frac{\rho^2(z-\beta)^2}{2\delta_y^2(1+\rho^2z^2)}\right] \\
&= \exp\left(-\frac{1+\beta^2\rho^2}{2\delta_y^2}\right)\frac{\rho}{\pi(1+\rho^2z^2)} \\
&\quad +\frac{\rho(1+\beta\rho^2 z)}{\sqrt{2\pi}\delta_y(1+\rho^2z^2)^{3/2}}\exp\left[-\frac{\rho^2(z-\beta)^2}{2\delta_y^2(1+\rho^2z^2)}\right]\operatorname{erf}\left[\frac{1+\beta\rho^2 z}{\delta_y\sqrt{2(1+\rho^2z^2)}}\right]. \qquad (4.1)
\end{aligned}
$$

Reparametrización Hinkley.

David Hinkley (1969) obtuvo la densidad g_Z como se muestra a continuación. Hay una correspondencia uno a uno entre la expresión anterior (4.1) y la densidad obtenida por Hinkley (4.7). Sean

$$
\begin{aligned}
a(z) &= \frac{1}{\sigma_y}(1+\rho^2z^2)^{\frac{1}{2}}, && (4.2)\\
b(z) &= \frac{1}{\sigma_x}\left(\frac{z}{\delta_x}+\frac{1}{\rho\delta_y}\right), && (4.3)\\
c &= \frac{1}{\delta_x^2}+\frac{1}{\delta_y^2}, && (4.4)\\
d(z) &= \exp\left[\frac{b^2(z)-ca^2(z)}{2a^2(z)}\right] \quad y && (4.5)\\
\operatorname{erf}\left(\frac{z}{\sqrt{2}}\right) &= \Phi(z)-\Phi(-z). && (4.6)
\end{aligned}
$$

Entonces la densidad $g_Z(z)$ se puede reescribir como:

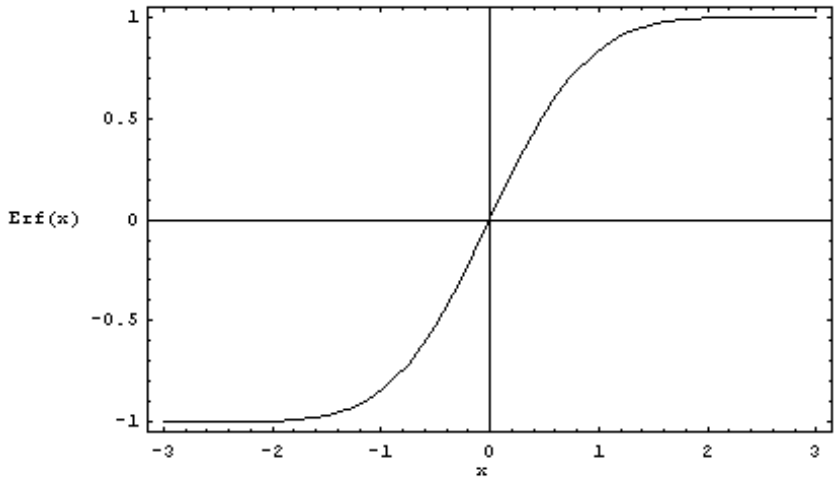
$$g_Z(z) = \frac{b(z)d(z)}{\sqrt{2\pi}\sigma_x\sigma_y a^3(z)} \left\{ \Phi\left[\frac{b(z)}{a(z)}\right] - \Phi\left[-\frac{b(z)}{a(z)}\right] \right\} + \frac{1}{\pi\sigma_x\sigma_y a^2(z)} \exp\left(-\frac{c}{2}\right). \tag{4.7}$$

A2. Función de error

Las integrales de error ó función de error se definen como:

$$\begin{aligned} \operatorname{erf}(x) &= \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt, \\ \operatorname{erfc}(x) &= 1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt. \end{aligned}$$

La gráfica de la función erf es la siguiente:



A continuación se enuncian cuatro propiedades importantes de esta función:

Propiedad 1.

$$\operatorname{erf}(x) - \operatorname{erf}(-x) = 2\operatorname{erf}(x).$$

Propiedad 2.

$$\begin{aligned} \lim_{x \longrightarrow -\infty} \operatorname{erf}(x) &= -1, \\ \lim_{x \longrightarrow \infty} \operatorname{erf}(x) &= 1. \end{aligned}$$

Propiedad 3.

$$\begin{aligned} \Phi(x) &= \frac{1}{2}\left[1+\operatorname{erf}\left(\frac{x}{\sqrt{2}}\right)\right], \qquad y \\ \operatorname{erf}(x) &= 2\Phi(\sqrt{2}x)-1, \end{aligned}$$

donde:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt.$$

Propiedad 4.

$$\operatorname{erf}(x) = \pi^{-1/2}\gamma\left(\frac{1}{2}, x^2\right),$$

donde:

$$\gamma(a,x) = \int_0^x t^{a-1}e^{-t}dt.$$

Una aproximación útil a la función erf es la que presenta Sergei Winitzki (2008) :

$$\operatorname{erf}(x) \approx \left[1-\exp\left(-x^2\frac{\frac{4}{\pi}+ax^2}{1+ax^2}\right)\right]^{1/2},$$

donde $a = \frac{8}{3\pi}\frac{\pi-3}{4-\pi}$.

A3. Demostración del Teorema 1

Teorema *1 (David Hinkley, 1969).* *Sean X y Y variables aleatorias independientes con distribución normal, $X \backsim N(\mu_x, \sigma_x)$ y $Y \backsim N(\mu_y, \sigma_y)$, donde $\mu_x \neq 0$, $\mu_y \neq 0$, $\sigma_x > 0$, $\sigma_y > 0$, sea F_Z la función de distribución de la variable $Z = X/Y$. Para $w \in \mathbb{R}$ fijo se define:*

$$F^*(w) : = P[X - wY \leq 0] = \Phi\left(\frac{w\mu_y - \mu_x}{\sigma_x \sigma_y \sqrt{\frac{1}{\sigma_y^2} + \frac{w^2}{\sigma_x^2}}}\right), \tag{4.8}$$

entonces $F_Z \xrightarrow{u} F^*$ cuando $\delta_y = \sigma_y/\mu_y \to 0$.

A continuación se da la demostración de este teorema. La idea básica de la demostración fue tomada del artículo de Hinkley (1969), pero aquí se presentan los pasos con mayor detalle.

Demostración. Por definición:

$$\begin{aligned} F_Z(w) &= P[Z \leq w] = P[X/Y \leq w] \\ &= P[X/Y \leq w, Y > 0] + P[X/Y \geq w, Y < 0] \\ &= P[X/Y \leq w \,|Y > 0]\, P[Y > 0] + P[X/Y \geq w \,|Y < 0]\, P[Y < 0] \\ &= P[X - wY \leq 0 \,|Y > 0]\, P[Y > 0] + P[X - wY \geq 0 \,|Y < 0]\, P[Y < 0]. \end{aligned}$$

Dado que $Y \backsim N(\mu_y, \sigma_y)$ se tiene que:

$$\begin{aligned} P[Y < 0] &= \Phi\left(-\frac{\mu_y}{\sigma_y}\right), \\ P[Y > 0] &= \Phi\left(\frac{\mu_y}{\sigma_y}\right). \end{aligned}$$

Notemos que:

$$\begin{aligned} P[Y < 0] &\to 0, \quad \text{cuando } \delta_y \to 0, \\ P[Y > 0] &\to 1, \quad \text{cuando } \delta_y \to 0. \end{aligned}$$

Además, para w fijo se tiene que $X - wY \backsim N\left(\mu_x - w\mu_y, \sqrt{\sigma_x^2 + w^2\sigma_y^2}\right)$, entonces:

$$P\left[X - wY \leq 0\right] = \Phi\left(\frac{w\mu_y - \mu_x}{\sigma_x\sigma_y\sqrt{\frac{1}{\sigma_y^2} + \frac{w^2}{\sigma_x^2}}}\right) = F^*\left(w\right).$$

Utilizando las relaciones $P\left(A \cup B\right) = P\left(A\right) + P\left(B\right) - P\left(A \cap B\right)$ y $P\left(A^c\right) = 1 - P\left(A\right)$ se tiene que:

$$\begin{aligned} P\left[X - wY \leq 0, Y > 0\right] &= P\left[X - wY \leq 0\right] + P\left[Y > 0\right] - P\left[\left\{X - wY \leq 0\right\} \cup \left\{Y > 0\right\}\right] \\ &= F^*\left(w\right) + 1 - P\left[Y < 0\right] - \left(1 - P\left[X - wY > 0, Y < 0\right]\right) \\ &= F^*\left(w\right) - P\left[Y < 0\right] + P\left[X - wY > 0, Y < 0\right]. \end{aligned}$$

Por lo tanto:

$$\begin{aligned} F_Z\left(w\right) &= F^*\left(w\right) - P\left[Y < 0\right] + 2P\left[X - wY > 0, Y < 0\right] \\ &= F^*\left(w\right) - P\left[Y < 0\right]\left(1 - 2P\left[X - wY > 0\,|Y < 0\right]\right). \end{aligned}$$

Ahora $\left(1 - 2P\left[X - wY > 0\,|Y < 0\right]\right) \in \left[-1, 1\right]$, entonces:

$$\begin{aligned} &\left|F_Z\left(w\right) - F^*\left(w\right)\right| = \\ &\left|P\left[Y < 0\right]\left(1 - 2P\left[X - wY > 0\,|Y < 0\right]\right)\right| \leq P\left[Y < 0\right] = \Phi\left(-\frac{\mu_y}{\sigma_y}\right). \end{aligned}$$

Dado $0 < \varepsilon < 1$ existen μ_y y σ_y tales que $\Phi\left(-\frac{\mu_y}{\sigma_y}\right) = \varepsilon$, por lo tanto $F_Z\left(w\right) \xrightarrow{u} F^*\left(w\right)$ cuando $\delta_y = \frac{\mu_y}{\sigma_y} \to 0$. ■

En el caso límite la aproximación a la densidad g_z tendrá la forma:

$$f^*\left(w\right) = \frac{b\left(w\right)d\left(w\right)}{\sqrt{2\pi}\sigma_x\sigma_y a^3\left(w\right)}.$$

Las definiciones de las funciones $a\left(\cdot\right)$, $b\left(\cdot\right)$, $d\left(\cdot\right)$ se dan en el apéndice A1.

A4. Aproximación de la media y la varianza de Z

Resultado de Mood et. al. (1974 pp. 181). Sean X e Y variables aleatorias con medias $E[X]=\mu_X,\ E[Y]=\mu_Y,$ varianzas $Var[X]=\sigma_X^2,\ Var[Y]=\sigma_Y^2$ y $Z=X/Y$; entonces, si la esperanza y la varianza de Z existen, se pueden aproximar así:

$$\begin{aligned} E[Z] &\approx \frac{\mu_X}{\mu_Y}-\frac{1}{\mu_Y^2}Cov[X,Y]+\frac{\mu_X}{\mu_Y^3}\sigma_Y^2, \qquad \text{y} \\ Var[Z] &\approx \left(\frac{\mu_X}{\mu_Y}\right)^2\left(\frac{\sigma_X^2}{\mu_X^2}+\frac{\sigma_Y^2}{\mu_Y^2}-\frac{2Cov[X,Y]}{\mu_X\mu_Y}\right). \end{aligned}$$

Demostración. Sea:

$$Z=h(X,Y)=X/Y.$$

La expansión en una serie de Taylor de esta función alrededor del punto (μ_X,μ_Y) hasta el término de segundo orden es:

$$\begin{aligned} Z &= h(X,Y)\approx h(\mu_X,\mu_Y)+(X-\mu_X)\left.\frac{\partial h(X,Y)}{\partial X}\right|_{(X,Y)=(\mu_X,\mu_Y)} \\ &\quad +(Y-\mu_Y)\left.\frac{\partial h(X,Y)}{\partial Y}\right|_{(X,Y)=(\mu_X,\mu_Y)} \\ &\quad +\frac{(X-\mu_X)^2}{2}\left.\frac{\partial^2 h(X,Y)}{\partial X^2}\right|_{(X,Y)=(\mu_X,\mu_Y)} \\ &\quad +\frac{(Y-\mu_Y)^2}{2}\left.\frac{\partial^2 h(X,Y)}{\partial Y^2}\right|_{(X,Y)=(\mu_X,\mu_Y)} \\ &\quad +(X-\mu_X)(Y-\mu_Y)\left.\frac{\partial^2 h(X,Y)}{\partial X\partial Y}\right|_{(X,Y)=(\mu_X,\mu_Y)}. \end{aligned}$$

Tomando la esperanza a esta aproximación:

$$\begin{aligned} E(Z) &\approx h(\mu_X,\mu_Y)+\frac{1}{2}Var[X]\left.\frac{\partial^2 h(X,Y)}{\partial X^2}\right|_{(X,Y)=(\mu_X,\mu_Y)} \\ &\quad +\frac{1}{2}Var[Y]\left.\frac{\partial^2 h(X,Y)}{\partial Y^2}\right|_{(X,Y)=(\mu_X,\mu_Y)} \\ &\quad +Cov[X,Y]\left.\frac{\partial^2 h(X,Y)}{\partial X\partial Y}\right|_{(X,Y)=(\mu_X,\mu_Y)}. \end{aligned}$$

A continuación se calculan cada uno de los términos de esta aproximación:

$$\begin{aligned}
\left.\frac{\partial h(X,Y)}{\partial X}\right|_{(X,Y)=(\mu_X,\mu_Y)} &= \frac{1}{\mu_Y}, \\
\left.\frac{\partial h(X,Y)}{\partial Y}\right|_{(X,Y)=(\mu_X,\mu_Y)} &= -\frac{\mu_X}{\mu_Y^2}, \\
\left.\frac{\partial^2 h(X,Y)}{\partial X^2}\right|_{(X,Y)=(\mu_X,\mu_Y)} &= 0, \\
\left.\frac{\partial^2 h(X,Y)}{\partial Y^2}\right|_{(X,Y)=(\mu_X,\mu_Y)} &= \frac{\mu_X}{\mu_Y^3}, \\
\left.\frac{\partial^2 h(X,Y)}{\partial X \partial Y}\right|_{(X,Y)=(\mu_X,\mu_Y)} &= -\frac{1}{\mu_Y^2}.
\end{aligned}$$

Por lo tanto:

$$E\left[Z\right] \approx \frac{\mu_X}{\mu_Y} + \frac{\mu_X}{\mu_Y^3} Var[Y] - \frac{1}{\mu_Y^2} Cov[X,Y].$$

Utilizando la aproximación de primer orden en la serie de Taylor y en el caso de independencia, la varianza puede aproximarse por:

$$Var[Z] \approx \left(\frac{\sigma_X}{\mu_Y}\right)^2 + \left(\frac{\sigma_Y}{\mu_Y}\right)^2 \left(\frac{\mu_X}{\mu_Y}\right)^2.$$

■

A5. Demostración del Teorema 2

Teorema 2. *Sea X una variable aleatoria con distribución normal con media $\mu_x > 0$, varianza σ_x^2 y coeficiente de variación $\delta_x = \sigma_x / \mu_x$ tal que $0 < \delta_x < \lambda < 1$. Dado $\varepsilon \in (0,1)$, existen un valor $\gamma(\varepsilon) \in (0,\lambda)$ y una variable aleatoria normal Y, independiente de X con media $\mu_y > 0$, varianza σ_y^2 y coeficiente de variación $\delta_y = \sigma_y / \mu_y$, que cumple con:*

$$0 < \delta_y < \gamma(\varepsilon) < \sqrt{\lambda^2 - \delta_x^2} < \lambda,$$

y es tal que si $Z = X/Y$, entonces para z en el intervalo:

$$I = \left\{ \mu_x/\mu_y \pm \frac{1}{\lambda} \sqrt{\left(\sigma_x/\mu_y\right)^2 + \left(\mu_x/\mu_y\right)^2 \left(\sigma_y/\mu_y\right)^2} \right\} = \left\{ \beta \pm \frac{\beta}{\lambda} \sqrt{\delta_x^2 + \delta_y^2} \right\}, \tag{4.9}$$

se cumple que:

$$|F(z) - F_Z(z)| < \varepsilon, \tag{4.10}$$

donde F es es la función de distribución de una variable aleatoria normal con media β y desviación estándar $\beta\sqrt{\delta_x^2 + \delta_y^2}$,

$$\begin{aligned} F(z) &= \Phi\left[\frac{z - \mu_x/\mu_y}{\sqrt{\left(\sigma_x/\mu_y\right)^2 + \left(\mu_x/\mu_y\right)^2 \left(\sigma_y/\mu_y\right)^2}} \right] \\ &= \Phi\left[\frac{z - \beta}{\beta\sqrt{\delta_x^2 + \delta_y^2}} \right], \end{aligned} \tag{4.11}$$

y F_Z es la distribución de Z.

Es decir, bajo las condiciones en los parámetros mencionadas, la distribución normal propuesta (4.11) aproximará muy de cerca a la distribución F_Z en el intervalo descrito en (4.9).

Demostración. Defínase:

$$h_1(z) = \frac{z\mu_y - \mu_x}{\sigma_x \sigma_y \sqrt{\frac{1}{\sigma_y^2} + \frac{z^2}{\sigma_x^2}}} = \frac{z\mu_y - \mu_x}{\sqrt{\sigma_x^2 + z^2\sigma_y^2}}, \qquad z > 0,$$

y

$$h_2(z) = \frac{z - \mu_x/\mu_y}{\sqrt{\left(\sigma_x/\mu_y\right)^2 + \left(\mu_x/\mu_y\right)^2 \left(\sigma_y/\mu_y\right)^2}} = \frac{\mu_y z - \mu_x}{\sqrt{\sigma_x^2 + \mu_x^2 \left(\sigma_y/\mu_y\right)^2}}, \qquad z > 0.$$

Nótese que $h_1(z)$ es equivalente a la expresión del argumento de la aproximación que da Hinkley en (4.8). Luego, considérense a las siguientes distribuciones:

$$\begin{aligned} F^*(w) &= \Phi[h_1(w)], \\ F(w) &= \Phi[h_2(w)], \end{aligned}$$

donde F^* es la aproximación de Hinkley a la distribución de Z y F es la aproximación normal propuesta en (4.11). Se demostrará primero que dado $0 < \varepsilon' < 1$ existe un valor positivo $\eta(\varepsilon')$, tal que si $\delta_y < \eta$ entonces para todo $z \in I$ se cumple que:

$$|h_1(z) - h_2(z)| < \frac{\varepsilon'}{\lambda}. \tag{4.12}$$

Esto con el fin de mostrar que si el coeficiente de variación δ_y es suficientemente pequeño, entonces los argumentos de la distribución de la normal propuesta y de la aproximación de Hinkley son cercanos. Dada la definición de h_1 y h_2 y que $z > 0$ se tiene la siguiente relación:

$$\begin{aligned} h_1(z) &= h_2(z), \qquad \text{si } z = \mu_x/\mu_y = \beta, \\ h_1(z) &< h_2(z), \qquad \text{si } z > \beta, \\ h_1(z) &< h_2(z), \qquad \text{si } z < \beta. \end{aligned}$$

Por lo tanto $h_1(z) \leq h_2(z)$ para $z \in I$. Ahora:

$$\frac{d}{dz}(h_2(z) - h_1(z)) = \frac{\mu_y}{\sqrt{\sigma_x^2 + \mu_x^2\left(\sigma_y/\mu_y\right)^2}} - \frac{\mu_y\sigma_x^2 + \mu_x\sigma_y^2 z}{\left(\sigma_x^2 + z^2\sigma_y^2\right)^{3/2}}.$$

A continuación analizaremos el comportamiento de esta derivada para demostrar que las diferencias máximas entre h_2 y h_1 se encuentran en los extremos del intervalo I ya que $h_2(z) - h_1(z)$ siempre es no negativa y alcanza el valor 0 en β, que es el punto medio del intervalo I.

Si $z > \mu_x/\mu_y$, entonces:

$$\begin{aligned}
\frac{\mu_y}{\sqrt{\sigma_x^2+\mu_x^2\left(\sigma_y/\mu_y\right)^2}}-\frac{\mu_y\sigma_x^2+\mu_x\sigma_y^2 z}{\left(\sigma_x^2+z^2\sigma_y^2\right)^{3/2}} &= \mu_y\left[\frac{1}{\sqrt{\sigma_x^2+\mu_x^2\left(\sigma_y/\mu_y\right)^2}}-\frac{\sigma_x^2+\left(\mu_x/\mu_y\right)\sigma_y^2 z}{\left(\sigma_x^2+z^2\sigma_y^2\right)^{3/2}}\right] \\
&> \mu_y\left[\frac{1}{\sqrt{\sigma_x^2+\mu_x^2\left(\sigma_y/\mu_y\right)^2}}-\frac{\sigma_x^2+\left(\mu_x/\mu_y\right)\sigma_y^2 z}{\left(\sigma_x^2+\left(\mu_x/\mu_y\right)\sigma_y^2 z\right)^{3/2}}\right] \\
&= \mu_y\left[\frac{1}{\sqrt{\sigma_x^2+\mu_x^2\left(\sigma_y/\mu_y\right)^2}}-\frac{1}{\sqrt{\sigma_x^2+\left(\mu_x/\mu_y\right)\sigma_y^2 z}}\right]>0.
\end{aligned}$$

Si $z < \mu_x/\mu_y$, entonces:

$$\begin{aligned}
\frac{\mu_y}{\sqrt{\sigma_x^2+\mu_x^2\left(\sigma_y/\mu_y\right)^2}}-\frac{\mu_y\sigma_x^2+\mu_x\sigma_y^2 z}{\left(\sigma_x^2+z^2\sigma_y^2\right)^{3/2}} &= \mu_y\left[\frac{1}{\sqrt{\sigma_x^2+\mu_x^2\left(\sigma_y/\mu_y\right)^2}}-\frac{\sigma_x^2+\left(\mu_x/\mu_y\right)\sigma_y^2 z}{\left(\sigma_x^2+z^2\sigma_y^2\right)^{3/2}}\right] \\
&< \mu_y\left[\frac{1}{\sqrt{\sigma_x^2+\mu_x^2\left(\sigma_y/\mu_y\right)^2}}-\frac{\sigma_x^2+z^2\sigma_y^2}{\left(\sigma_x^2+z^2\sigma_y^2\right)^{3/2}}\right] \\
&= \mu_y\left[\frac{1}{\sqrt{\sigma_x^2+\mu_x^2\left(\sigma_y/\mu_y\right)^2}}-\frac{1}{\sqrt{\sigma_x^2+z^2\sigma_y^2}}\right]<0.
\end{aligned}$$

Por lo tanto $[h_2(z)-h_1(z)]$ es decreciente para $z < \beta$ y es creciente para $z > \beta$.

Debido a este comportamiento se tiene que la diferencia $(h_2 - h_1)$ toma sus valores máximos en los extremos del intervalo I, por lo tanto para que se cumpla la desigualdad (4.12) en todo I es suficiente si esta desigualdad se cumple en ambos extremos del intervalo I.

Si la desigualdad (4.12) se cumple en el extremo izquierdo de I se tiene que:

$$\begin{aligned}
A &= h_2\left[\mu_x/\mu_y-\frac{1}{\lambda}\sqrt{\left(\sigma_x/\mu_y\right)^2+\left(\mu_x/\mu_y\right)^2\left(\sigma_y/\mu_y\right)^2}\right] \\
&\quad -h_1\left[\mu_x/\mu_y-\frac{1}{\lambda}\sqrt{\left(\sigma_x/\mu_y\right)^2+\left(\mu_x/\mu_y\right)^2\left(\sigma_y/\mu_y\right)^2}\right] \\
&\quad -\frac{1}{\lambda}+\frac{1}{\lambda}\frac{\sqrt{\sigma_x^2+\mu_x^2\left(\sigma_y/\mu_y\right)^2}}{\sqrt{\sigma_x^2+\left(\mu_x\left(\sigma_y/\mu_y\right)-\frac{1}{\lambda}\sqrt{\sigma_x^2\left(\sigma_y/\mu_y\right)^2+\mu_x^2\left(\sigma_y/\mu_y\right)^4}\right)^2}}. \qquad (4.13)
\end{aligned}$$

Nótese que $A < \varepsilon'/\lambda$ es equivalente a:

$$\frac{\sqrt{\sigma_x^2 + \mu_x^2 \left(\sigma_y/\mu_y\right)^2}}{\sqrt{\sigma_x^2 + \left(\mu_x \left(\sigma_y/\mu_y\right) - \frac{1}{\lambda}\sqrt{\sigma_x^2 \left(\sigma_y/\mu_y\right)^2 + \mu_x^2 \left(\sigma_y/\mu_y\right)^4}\right)^2}} < 1 + \varepsilon'.$$

Luego:

$$\begin{aligned}
\sqrt{\sigma_x^2 + \mu_x^2\delta_y^2} &< (1+\varepsilon')\sqrt{\sigma_x^2 + \left(\mu_x\delta_y - \frac{1}{\lambda}\sqrt{\sigma_x^2\delta_y^2 + \mu_x^2\delta_y^4}\right)^2}, \\
\sigma_x^2 + \mu_x^2\delta_y^2 &< (1+\varepsilon')^2\left[\sigma_x^2 + \left(\mu_x\delta_y - \frac{1}{\lambda}\sqrt{\sigma_x^2\delta_y^2 + \mu_x^2\delta_y^4}\right)^2\right], \\
\sigma_x^2\left[1 - (1+\varepsilon')^2\right] &< \left(\mu_x\delta_y - \frac{1}{\lambda}\sqrt{\sigma_x^2\delta_y^2 + \mu_x^2\delta_y^4}\right)^2 (1+\varepsilon')^2 - \mu_x^2\delta_y^2 \\
&= \mu_x^2\delta_y^2\left[\left(1 - \frac{1}{\lambda}\sqrt{\delta_x^2 + \delta_y^2}\right)^2 (1+\varepsilon')^2 - 1\right].
\end{aligned}$$

Por lo tanto $A < \varepsilon'/\lambda$ es equivalente a la siguiente expresión:

$$\delta_x^2\left[(1+\varepsilon')^2 - 1\right] > \delta_y^2\left[1 - \left(1 - \frac{1}{\lambda}\sqrt{\delta_x^2 + \delta_y^2}\right)^2 (1+\varepsilon')^2\right]. \tag{4.14}$$

Para δ_x, λ y ε' fijos se cumple que:

$$\lim_{\delta_y \to 0} \delta_y^2\left[1 - \left(1 - \frac{1}{\lambda}\sqrt{\delta_x^2 + \delta_y^2}\right)^2 (1-\varepsilon')^2\right] = 0.$$

Dependiendo del signo de $\left[1 - \left(1 - \frac{1}{\lambda}\sqrt{\delta_x^2 + \delta_y^2}\right)^2 (1+\varepsilon')^2\right]$, este límite será por la derecha o por la izquierda.

Entonces existe $\eta_1 = \eta_1(\varepsilon', \delta_x, \lambda) > 0$ tal que si $\delta_y < \eta_1$ se cumple la desigualdad $A < \varepsilon'/\lambda$.

Análogamente, para el extremo derecho del intervalo I se tiene:

$$\begin{aligned}
B &= h_2\left[\mu_x/\mu_y + \frac{1}{\lambda}\sqrt{\left(\sigma_x/\mu_y\right)^2 + \left(\mu_x/\mu_y\right)^2\left(\sigma_y/\mu_y\right)^2}\right] \\
&\quad - h_1\left[\mu_x/\mu_y + \frac{1}{\lambda}\sqrt{\left(\sigma_x/\mu_y\right)^2 + \left(\mu_x/\mu_y\right)^2\left(\sigma_y/\mu_y\right)^2}\right] \\
&= \frac{1}{\lambda} - \frac{1}{\lambda}\frac{\sqrt{\sigma_x^2 + \mu_x^2\left(\sigma_y/\mu_y\right)^2}}{\sqrt{\sigma_x^2 + \left[\mu_x\left(\sigma_y/\mu_y\right) + \frac{1}{\lambda}\sqrt{\sigma_x^2\left(\sigma_y/\mu_y\right)^2 + \mu_x^2\left(\sigma_y/\mu_y\right)^4}\right]^2}}.
\end{aligned}$$

$B < 3\varepsilon'$ es equivalente a:

$$\delta_x^2 \left[1 - (1-\varepsilon')^2\right] > \delta_y^2 \left[\left(1 + \frac{1}{\lambda}\sqrt{\delta_x^2 + \delta_y^2}\right)^2 (1-\varepsilon')^2 - 1\right]. \tag{4.15}$$

Para δ_x, λ y ε' fijos se cumple que:

$$\lim_{\delta_y \to 0} \delta_y^2 \left[\left(1 + \frac{1}{\lambda}\sqrt{\delta_x^2 + \delta_y^2}\right)^2 (1-\varepsilon')^2 - 1\right] = 0.$$

Dependiendo del signo de $\left[\left(1 + \frac{1}{\lambda}\sqrt{\delta_x^2 + \delta_y^2}\right)^2 (1-\varepsilon')^2 - 1\right]$, este límite será por la derecha o por la izquierda.

Entonces existe $\eta_2 = \eta_2(\varepsilon', \delta_x, \lambda) > 0$ tal que si $\delta_y < \eta_2$ se cumple la desigualdad $B < \varepsilon'/\lambda$. Eligiendo $\eta \leq \min\{\eta_1, \eta_2\}$, se tiene que $|h_1(z) - h_2(z)| < \varepsilon'/\lambda$, para todo $z \in I$.

Equivalentemente para $z \in I$ se tiene que:

$$h_2(z) - \varepsilon'/\lambda < h_1(z) < h_2(z) + \varepsilon'/\lambda.$$

Luego para $z \in I$:

$$\begin{aligned} |F(z) - F^*(z)| &= |\Phi[h_2(z)] - \Phi[h_1(z)]| = \Phi[h_2(z)] - \Phi[h_1(z)] \\ &< \Phi[h_2(z)] - \Phi[h_2(z) - \varepsilon'/\lambda]. \end{aligned}$$

Utilizando el Teorema del valor medio para integrales y el hecho de que la densidad de una normal estándar ϕ tiene un máximo de altura $M = 1/\sqrt{2\pi}$ que se alcanza en cero y por tanto se tiene que existe $\xi_1 \in (h_2(z) - \varepsilon'/\lambda, h_2(z))$ tal que:

$$\Phi[h_2(z)] - \Phi[h_2(z) - \varepsilon'/\lambda] = \int_{h_2(z)-\varepsilon'/\lambda}^{h_2(z)} \phi(t)\, dt = \phi(\xi_1)\, \varepsilon'/\lambda \leq M\varepsilon'/\lambda,$$

Por lo tanto si $\delta_y < \eta$ se cumple que $|F(z) - F^*(z)| < M\varepsilon'/\lambda$ y esta cota superior no depende de z. Eligiendo $\varepsilon' < \varepsilon\lambda/2M$ se tiene finalmente que:

$$|F(z) - F^*(z)| < \frac{\varepsilon}{2}.$$

Ahora, debido a que F_Z converge uniformemente a F^* cuando $\delta_y \to 0$ (Hinkley, 1969) se tiene que existe $\eta_3 = \eta_3(\varepsilon)$ tal que para $\delta_y < \eta_3$ se cumple que:

$$|F^*(z) - F_Z(z)| < \frac{\varepsilon}{2}.$$

Por lo tanto dado $0 < \varepsilon < 1$ si $0 < \delta_y < \min\left(\eta(\varepsilon, \delta_x, \lambda), \eta_3(\varepsilon), \sqrt{\lambda^2 - \delta_x^2}\right) = \gamma$ y para todo $z \in I$, por desigualdad del triángulo:

$$|F(z) - F_Z(z)| \leq |F(z) - F^*(z)| + |F^*(z) - F_Z(z)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

■

Bibliografía

[1] Cedilnik, A., Košmelj, K., and Blejec, A. (2004). The Distribution of the Ratio of Jointly Normal Variables. *Metodološki zvezki*, Vol. 1, No. 1, pp. 99-108.

[2] Craig, C. C. (1929). The Frequency of Function of y/x. *The Annals of Mathematics*, 2nd Ser., Vol. 30, No. 1/4, pp. 471-486.

[3] Curtiss, J. H. (1941). On the Distribution of the Quotient of Two Chance Variables. *The Annals of Mathematical Statistics*, Vol. 12, No. 4, pp. 409-421.

[4] Dalgard, P. (2002). *Introductory Statistics with R*. New York: Springer-Verlag, pp. 111-115.

[5] Díaz-Francés, E. y Sprott, D. A. (2003). Inference for the Ratio of Two Normal Means with Unspecified Variances. *Biometrical Journal*, Vol. 46, No. 1, 83–89.

[6] Díaz-Francés, E. y Sprott, D. A. (2001). Statistical Analysis of Nuclear Genome Size of Plants With Flow Cytometer Data. *Cytometry*, Vol. 45, pp. 244–249.

[7] Doležel, J., Greilhuber, J., Lucretti, S., Meister, A., Lysák, M. A., Nardi y L., Obermayer, R. (1998). Plant Genome Size Estimation by Flow Citometry: Inter-laboratory Comparison. *Annals of Botany*, Vol. 82, pp. 17-26.

[8] Fieller, E. C. (1932). The Distribution of the Index in a Normal Bivariate Population. *Biometrika*, Vol. 24, No. 3/4, pp. 428-440.

[9] Fieller, E. C. (1940). The Biological Standardization of Insulin. *Supplement to the Journal of the Royal Statistical Society*, Vol. 7, No. 1, pp. 1-64.

[10] Fisher, R.A. (1973). *Statistical Methods and Scientific Inference*. Macmillan, New York, third edition. Reprinted in Fisher (1990), pp. 146.

[11] Geary, R. C. (1930).The Frequency Distribution of the Quotient of Two Normal Variates. *Journal of the Royal Statistical Society*, Vol. 93, No. 3, pp. 442-446.

[12] Hayya, J., Armstrong, D. y Gressis, N. (1975). A Note on the Ratio of Two Normally Distributed Variables. *Management Science*, Vol. 21, No. 11, pp. 1338-1341.

[13] Hinkley, D. V. (1969). On the Ratio of Two Correlated Normal Variables. *Biometrika*, Vol. 56, No. 3, pp. 635-639.

[14] Kalbfleisch, J.G. (1985). *Probability and Statistical Inference Volume 2: Statistical Inference*. New York: Springer-Verlag.

[15] Kamerud, D. B. (1978). The Random Variable X/Y, X, Y, normal. *The American Mathematical Monthly*, Vol. 85 pp. 206–207.

[16] Kuethe, D. O., Caprihan, A., Gach, H. M., Lowe, I. J. y Fukushima, E. (2000). Imaging Obstructed Ventilation with NMR Using Inert Fluorinated Gases. *Journal of Applied Physiology*, Vol. 88 pp. 2279-2286.

[17] Lisák, M. A. y Doležel, J. (1998). Estimation of nuclear DNA content in Sesleria (Poaceae). *Caryologia*, Vol. 52, No. 2, pp. 123-132.

[18] Marsaglia, G. (1965). Ratios of Normal Variables and Ratios of Sums of Uniform Variables. *Journal of the American Statistical Association*, Vol. 60, No. 309, pp. 193-204.

[19] Mendoza, M. y Gutiérrez-Peña, E. (1999). Bayesian Inference for the Ratio of the Means of Two Normal Populations with Unequal Variances. *Biometrical Journal*, Vol. 41 No. 2, pp. 133-147.

[20] Merrill, A. S. (1928). Frequency Distribution of an Index When Both the Components Follow the Normal Law. *Biometrika*, Vol. 20A, No. 1/2, pp. 53-63.

[21] Miller, R. G. (1997). *Beyond ANOVA*. Texts in Statistical Science.

[22] Mood, A., Graybill, F. y Boes, D. (1974), 3a. Edición. *Introduction to the Theory of Statistics*. Nueva York: McGrawHill, pp. 181.

[23] Palomino, G., Doležel, J., Cid, R., Brunner, I., Méndez, I. y Rubluo, A. (1999). Nuclear Genome Stability of *Mammilaria san-angelensis* (Cactaceae) Regenerants Induced by Auxins in Long-Term in Vitro Culture. *Plant Science*, Vol. 141, pp. 191-200.

[24] Pearson, K. (1910). On the Constants of Index-Distributions as Deduced From the Like Constants for the Components of the Ratio, With Special Reference to the Opsonic Index. *Biometrika*, Vol. 7, No. 4, pp. 531-541.

[25] Pham-Gia, T. y Turkkan, N. (2006). Density of the Ratio of Two Normal Random Variables. *Communications in Statistics*.

[26] Qiao C. G.,Wood, G. R., Lai, C. D. y Luo, D. W. (2006). Comparison of Two Common Estimator of the Ratio of the Means of Independent Normal Variables in Agricultural Research. *Journal of Applied Mathematics and Decision Sciences*, Vol. 2006, pp. 1–14.

[27] Rubio, F. J. (2008). Inferencia sobre el cociente de dos medias normales. *Comunicación Técnica CIMAT*, No I-08-02/17-01-2008.

[28] Shanmugalingam, S. (1982). On the Analysis of the Ratio of Two Correlated Normal Variables . *The Statistician*, Vol. 31, No. 3, pp. 251-258.

[29] Sprott D. A. (2000), The Estimation of Ratios From Paired Data. En el libro de Ahmed, S. E. y Reid, N. (2000). *Empirical Bayes and Likelihood Inference*. Lecture Notes in Statistics, New York: Springer-Verlag. Capítulo 10.

[30] Sprott, D. A. (2000). *Statistical Inference in Science*. New York: Springer Series in Statistics, pp. 7-15.

[31] Winitzki, S. (2008). A Handy Approximation for the Error Function and Its Inverse. Notas electrónicas disponibles en http://homepages.physik.uni-muenchen.de/~Winitzki/erf-approx.pdf.

[32] Winitzki, S. (2003). Uniform Approximations for Transcendental Functions. Notas electrónicas disponibles en http://homepages.physik.uni-muenchen.de.