ANÁLISIS ESTADÍSTICO MULTIVARIADO

Análisis Multivariado

1. Inferencia Estadística
   - Distribución normal
   - Distribución de Wishart
   - Distribución T^2 de Hotelling
2. Estimación MV
3. Máxima verosimilitud y test de hipótesis
4. Test para la media de una población normal
5. Tests sobre la matriz de Variancias y Covariancias
6. Test de normalidad
   - Asimetría y kurtosis multivariada
7. ANEXO: relación entre el test de razón de verosimilitudes y la T^2 de Hotelling

Introducción

- Una variable aleatoria vectorial es el resultado de observar p características en un elemento de la población.
- Una variable aleatoria vectorial p-dimensional es continua (discreta) cuando lo es cada una de las p variables escalares que la componen.
- Se ha definido la distribución conjunta de una variable aleatoria vectorial cuando se especifique:
  - El espacio muestral o conjunto de sus valores posibles. El espacio muestral es en general un subconjunto de R^p.
  - Las probabilidades de cada resultado posible (subconjunto de puntos) del espacio muestral.
  - La función de distribución conjunta de una variable aleatoria vectorial $F(x)$ se define en un punto $x_0 = (x_1^0, x_2^0, \ldots, x_p^0)$:

$$F(x^0) = P(x \leq x^0) = P(x_1 \leq x_1^0, \ldots, x_p \leq x_p^0)$$

La distribución normal

- Si generalizamos la expresión de la distribución normal univariada, diremos que $x \sim N_p(\mu, \Sigma)$ si su función de densidad es:

$$f(x) = \frac{1}{\sqrt{|\Sigma|}(2\pi)^{p/2}} e^{-1/2(x-\mu)'\Sigma^{-1}(x-\mu)}$$

- La distribución es simétrica alrededor de μ
- La simetría se comprueba sustituyendo en la densidad por $\mu \pm a$ y observando que $f(\mu + a) = f(\mu - a)$.
- La distribución tiene un único máximo en μ. Como Σ es definida positiva $\Longrightarrow (x-\mu)'\Sigma^{-1}(x-\mu) > 0$ y la densidad $f(x)$ será máxima cuando dicho término sea igual a cero, en $x = \mu$.

La distribución normal

- La media del vector aleatorio normal es μ y su matriz de variancias y covariancias es Σ.
- Si p variables aleatorias tienen distribución conjunta normal y no están correlacionadas entonces son independientes.
- Las distribuciones marginales son normales.
- Cualquier conjunto de $h < p$ variables es normal h-dimensional.
- Si y es $k \times 1$ y $k \leq p$, el vector $y = Ax$, donde A es una matriz $k \times p$, es normal k-dimensional. En particular, cualquier variable escalar $y = a'x$ (siendo a' un vector $1 \times p$ no nulo) tiene distribución normal.

La distribución normal

- Cualquier vector x normal p-dimensional con matriz Σ no singular puede convertirse mediante una transformación lineal en un vector z normal p-dimensional con vector de medias 0 y matriz de variancias y covariancias igual a la matriz identidad I. Llamaremos normal p-dimensional estándar a la densidad de z que viene dada por:

$$f(x) = \frac{1}{(2\pi)^{p/2}} e^{-\frac{1}{2}z'z} = \prod_{i=1}^{p} \frac{1}{(2\pi)^{1/2}} e^{-\frac{1}{2}z_i^2}$$

- Al ser Σ definida positiva, existe A cuadrada y simétrica que verifica $\Sigma = AA$. Definimos una nueva variable $z = A^{-1}(x - \mu)$, entonces $x = \mu + Az$. La función de densidad de z es $f_z(z) = f_x(\mu + Az)|A|$.

La distribución normal

- Al cortar con hiperplanos paralelos al definido por las p variables que forman la variable vectorial x, se obtienen las curvas de nivel cuya ecuación es $(x-\mu)'\Sigma^{-1}(x-\mu) = \text{constante}$.
- Las curvas de nivel son elipsoides y definen una medida de la distancia de un punto al centro de la distribución.
- La distancia de Mahalanobis D^2 se distribuye como una χ^2 con p grados de libertad.
- Si $z = A^{-1}(x-\mu)$, y como $\Sigma^{-1} = A^{-1}A^{-1}$ se obtiene que $D^2 = z'z = \sum_{i=1}^{p} z_i^2$, donde cada z_i es $N(0,1)$. Por lo tanto $D^2 \sim \chi^2_p$.

La distribución normal

- Consideremos una partición del vector de variables aleatorias x:

$$x = (x_1, x_2)'$$

$$\mu = (\mu_1, \mu_2)'$$

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

La expresión de la distribución condicional es:

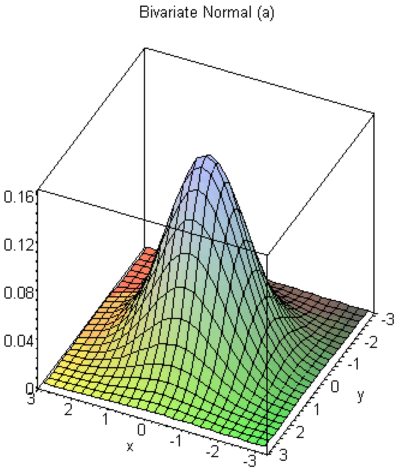
$$f(x_1/x_2) = \frac{f(x_1, x_2)}{f(x_2)}$$

Se puede demostrar que

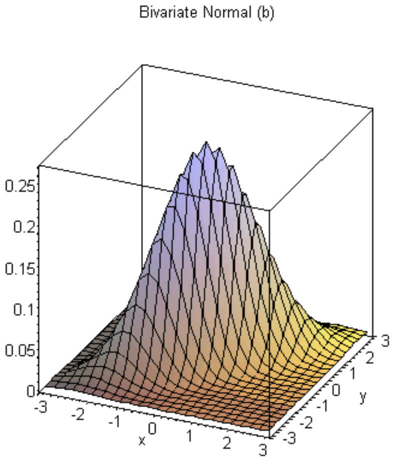
$$E[x_1/x_2] = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$$

$$Var[x_1/x_2] = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$$

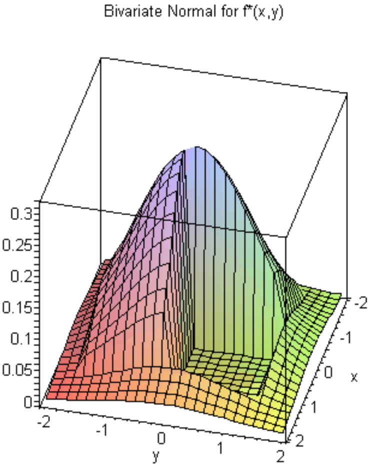
Normal bivariada con $\mu_1 = \mu_2 = 0$,$\sigma_1 = \sigma_2 = 1$,$\rho = 0$



Normal bivariada con $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1, \rho = 0.8$



Normal bivariada



La distribución de Wishart

- Se utiliza para representar la distribución muestral de las matrices de variancias y covariancias en muestras de variables aleatorias normales multivariadas.
- En el caso unidimensional, se utiliza la χ^2 de Pearson y la distribución de Wishart estándar puede considerarse como una generalización multivariante de esta distribución.
- Consideremos un conjunto de n vectores aleatorios $(x_1, x_2, \dots, x_n)$ donde cada $x_i \sim_{iid} N_p(0, I)$. La estimación de su matriz de variancias y covariancias se obtendrá como:

$$S = \frac{1}{n} \sum_{i=1}^{n} x_i x_i' = \frac{W}{n}$$

siendo $W_{(p \times p)}$ una matriz simétrica y definida positiva que se distribuye como una Wishart con n grados de libertad.

La distribución de Wishart

- La distribución conjunta de los $p(p+1)/2$ elementos distintos de la matriz W es:

$$f(w_{11}, w_{12}, \ldots, w_{pp}) = c|W|^{\frac{(n-p-1)}{2}} e^{-\frac{1}{2}tr(W)}$$

  donde c es una constante para que la función satisfaga la condición de cierre (integral=1).
- Para $p = 1$ se obtiene la expresión de la χ^2 univariada.
- En general expresaremos $W \sim W_p(n)$ donde p indica que se trata de la distribución de los elementos de una matriz simétrica de orden p, y n indica la cantidad de grados de libertad.
- Todas las combinaciones de elementos de la matriz W que conduzcan a los mismos valores de traza y determinante tienen la misma probabilidad.

La distribución de Wishart

- Consideremos ahora una muestra de n vectores $(x_1, x_2, \ldots, x_n)$ donde cada $x_i \sim_{iid} N_p(0, \Sigma)$, siendo $\Sigma_{(p \times p)}$ es simétrica, no singular, definida positiva y constante.

$$f(w_{11}, w_{12}, \ldots, w_{pp}) = c|\Sigma|^{-\frac{n}{2}}|W|^{\frac{(n-p-1)}{2}} e^{-\frac{1}{2}tr(\Sigma^{-1}W)}$$

  Diremos que W se comporta de acuerdo a la distribución de Wishart con n grados de libertad y matriz de parámetros Σ y lo expresaremos como $W \sim W_p(n, \Sigma)$.

- La esperanza de la distribución es $E[W] = n\Sigma$, por lo tanto $E[W/n] = \Sigma$.

La distribución de Wishart

- La suma de dos distribuciones χ^2 independientes también tiene distribución χ^2. Del mismo modo, si $W_1 \sim W_p(n_1, \Sigma)$ y $W_2 \sim W_p(n_2, \Sigma)$ son independientes, entonces $W_1 + W_2 \sim W_p(n_1 + n_2, \Sigma)$.
- Si A es una matriz $(h \times p)$ de constantes y $W \sim W_p(n, \Sigma)$, entonces $AWA' \sim W_h(n, A\Sigma A')$.
- Si S es la matriz de variancias y covariancias muestral $S = X'PX/n$ donde $P = (I - 11'/n)$ es idempotente, entonces $nS \sim W_p(n-1, \Sigma)$.
- El estimador

$$\widehat{S} = \frac{X'PX}{n-1} = \frac{nS}{n-1}$$

será insesgado para la matriz de variancias y covariancias Σ.

La distribución T^2 de Hotelling

- Si x es un vector aleatorio $N_p(\mu, \Sigma)$:

$$(x-\mu)'\Sigma^{-1}(x-\mu) \sim \chi^2_p$$

  Reemplazando Σ por su estimación $\widehat{S}$, la distribución que se obtiene se denomina T^2 de Hotelling.

- En general, si $x \sim N_p(\mu, \Sigma)$ y $(n-1)\widehat{S} \sim W_p(n-1, \Sigma)$:

$$T^2 = (x-\mu)'\widehat{S}^{-1}(x-\mu) \sim T^2_{Hotelling}(p, n-1)$$

- Asintóticamente, $\widehat{S} \to \Sigma$, $T^2 \to D^2_M$ (la distancia de Mahalanobis) y $T^2_{Hotelling}(p, n-1) \to \chi^2_p$.

La distribución T^2 de Hotelling

- Para n grande, la distribución de Hotelling es muy similar a una χ^2_p pero para tamaños de muestra mas pequeños presenta mayor variabilidad debido a la utilización de la matriz de variancias y covariancias estimada en lugar de la matriz de variancias y covariancias poblacional.
- Si $\bar{x} \sim N_p(\mu, n^{-1}\Sigma)$, la distribución del estadístico $n(\bar{x}-\mu)'\widehat{S}^{-1}(\bar{x}-\mu)$ es también una T^2 de Hotelling.
- Si $p = 1$ la T^2 es simplemente un estadístico t^2 (t-Student al cuadrado), por lo tanto $T^2_{1,m} = t^2_m$.
- La distribución T^2 no está tabulada ya que se puede demostrar la siguiente relación:

$$F_{p,(n-p)} = \frac{(n-p)}{p(n-1)} T^2_{p,n-1}$$

Estimación MV

- Sea x_1, x_2,..., x_n una muestra aleatoria simple donde $x_i \sim N_p(\mu, \Sigma)$.
- Hallaremos los estimadores máximo verosímiles de los parámetros desconocidos μ y Σ. Lo primero que debemos hacer es construir la función de densidad conjunta de las observaciones:

$$f(X/\mu, \Sigma) = \prod_{i=1}^{n} |\Sigma|^{-1/2}(2\pi)^{-p/2} exp\{-(1/2)(x_i - \mu)'\Sigma^{-1}(x_i - \mu)\}$$

y la función soporte ignorando las constantes será:

$$L(\mu, \Sigma/X) = -\frac{n}{2} log|\Sigma| - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)'\Sigma^{-1}(x_i - \mu)\}$$

Estimación MV

- Esta función es siempre negativa ya que tanto el determinante de Σ como la forma cuadrática son positivas por ser Σ definida positiva
- Expresaremos esta función de manera mas conveniente, llamando $\bar{x}$ al vector de medias muestral y desarrollando la forma cuadrática:

$$\sum_{i=1}^{n}(x_i-\mu)'\Sigma^{-1}(x_i-\mu)=\sum_{i=1}^{n}(x_i-\bar{x})'\Sigma^{-1}(x_i-\bar{x})+n(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)$$

Consideremos el primer término de esta última expresión:

$$tr(\sum_{i=1}^{n}(x_i-\bar{x})'\Sigma^{-1}(x_i-\bar{x}))=\sum_{i=1}^{n}tr[(x_i-\bar{x})'\Sigma^{-1}(x_i-\bar{x})]$$

Estimación MV

$$tr(\sum_{i=1}^{n}(x_i-\bar{x})'\Sigma^{-1}(x_i-\bar{x})) = \sum_{i=1}^{n} tr[(x_i-\bar{x})'\Sigma^{-1}(x_i-\bar{x})]$$

$$= \sum_{i=1}^{n} tr[\Sigma^{-1}(x_i-\bar{x})(x_i-\bar{x})'] = tr[\Sigma^{-1}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})']$$

y llamando

$$S = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})'$$

a la matriz de variancias y covariancias muestral podemos reescribir la función soporte como:

$$L(\mu,\Sigma/X) = -\frac{n}{2}log|\Sigma| - \frac{n}{2}tr(\Sigma^{-1}S) - \frac{n}{2}(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)$$

- Esta función solo depende de la muestra a través de los valores de $\bar{x}$ y S que serán por lo tanto los estimadores de μ y Σ.

Estimación MV

- Todas las muestras que proporcionen los mismos valores de $\bar{x}$ y S darán lugar a las mismas inferencias respecto a los parámetros.
- Para obtener el estimador del vector de medias en la población, tenemos en cuenta que Σ^{-1} es semidefinida positiva por lo tanto la forma cuadrática $(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu) \geq 0$.
- Como este término aparece con signo negativo el valor de μ que maximiza la función soporte es aquel que hace este término lo mas chico posible, y se hará igual a cero cuando $\widehat{\mu} = \bar{x}$ por lo que concluimos que $\bar{x}$ es el estimador máximo verosímil de μ.
- Sustituyendo este estimador en la función soporte este t'ermino desaparece. Para obtener el máximo de la función soporte con respecto a Σ sumaremos la constante $n/2log|S|$. De este modo escribimos la funcion soporte como:

$$L(\mu, \Sigma/X) = \frac{n}{2} log|\Sigma^{-1}S| - \frac{n}{2} tr(\Sigma^{-1}S)$$

Estimación MV

$$L(\mu, \Sigma/X) = \frac{n}{2} log|\Sigma^{-1}S| - \frac{n}{2} tr(\Sigma^{-1}S)$$

- Esta expresión es útil ya que el valor de la verosimilitud escrita de esta forma no depende de las unidades de medida de las variables y también es fácil comprobar que el valor de la similitud es invariante ante transformaciones lineales no singulares de las variables. Llamaremos λ_i a los autovalores de la matriz $\Sigma^{-1}S$ entonces:

$$L(\mu, \Sigma/X) = \frac{n}{2} \sum_{i=1}^{n} log(\lambda_i) - \frac{n}{2} \sum_{i=1}^{n} \lambda_i = \frac{n}{2} \sum_{i=1}^{n} (log\lambda_i - \lambda_i)$$

- Esta expresión indica que la verosimilitud es una suma de funciones del tipo $logx - x$. Derivando con respecto a x es inmediato que una función de este tipo tiene un máximo para $x = 1$. Por lo tanto, $L(\Sigma/X)$ será máxima si todos los autovalores de $\Sigma^{-1}S$ son iguales a 1, lo cual si logra si $\Sigma^{-1}S = I$. Esto se logra tomando como estimador de máxima verosimilitud de Σ a $\widehat{\Sigma} = S$.

Estimación MV

- Los estimadores MV de μ y Σ son entonces $\bar{x}$ y S.
- Finalmente, como en el caso unidimensional se puede demostrar que $\bar{x} \sim N_p(\mu, \Sigma/n)$.
- Además, nS se distribuye como una Wishart, $W_p(n-1, \Sigma)$.
- El estimador S es sesgado pero podemos calcular $\frac{n}{n-1}S$ que es insesgado para Σ.
- Estos estimadores gozan de todas las buenas propiedades de los estimadores máximo verosímiles: consistencia, eficiencia y normalidad asintótica.

Máxima Verosimilitud

- El método de máxima verosimilitud elige como estimador de los parámetros desconocidos aquellos valores que hacen máxima la probabilidad de que el modelo a estimar genere la muestra observada.
- En condiciones muy generales, el método de máxima verosimilitud permite obtener estimadores con las siguientes propiedades:
    - Son asintóticamente insesgados
    - La distribución asintótica de los estimadores es normal
    - Se obtienen estimadores asintóticamente eficientes (mínima variancia)
    - Si existe un estadístico suficiente para el parámetro de interés, es el estimador máximo verosímil
    - Se obtienen estimadores invariantes: si $\hat{\theta}$ es el estimador MV de θ y $g(\cdot)$ es una función cualquiera del vector de parámetros, entonces, bajo condiciones generales $g(\hat{\theta})$ es el estimador MV de $g(\theta)$.

Inferencia

- Para realizar tests para parámetros vectoriales podemos aplicar la teoría del test de razón de verosimilitudes que nos permite plantear pruebas estadísticas que tienen propiedades óptimas para tamaños de muestra grande.
- Dado un vector de parámetros θ, s-dimensional que toma valores en el espacio paramétrico Ω (donde Ω es un subconjunto de R^s) suponemos que se desea contrastar la siguiente hipótesis:

$$H_0)\theta \in \Omega_0$$

  frente a la hipótesis alternativa:

$$H_1)\theta \in \Omega - \Omega_0$$

- El contraste se basa en comparar la verosimilitud de la muestra bajo ambas hipótesis.

Inferencia

- La máxima probabilidad de obtener la muestra observada bajo H_0 se obtiene de la siguiente manera:
  - Si Ω_0 determina un valor único para los parámetros $\theta = \theta_0$ entonces se calcula la probabilidad de los datos suponiendo θ_0.
  - Si Ω_0 permite muchos valores, elegiremos entre ellos el valor del parámetro que haga máxima la probabilidad de obtener la muestra.
- Como la probabilidad de la muestra observada es proporcional a la distribución conjunta de las observaciones, sustituyendo en esta función los datos disponibles resulta la función de verosimilitud.
- Calculando el máximo de esta función en Ω_0 se obtiene el máximo valor de la verosimilitud compatible con H_0, que representaremos como $f(H_0)$.

Inferencia

- La máxima probabilidad de obtener la muestra observada bajo H_1 se calcula obteniendo el máximo absoluto de la función sobre todo el espacio paramétrico.
- Estrictamente debería calcularse en el conjunto $\Omega - \Omega_0$, pero es más simple hacerlo sobre todo el espacio ya que en general se obtiene el mismo resultado. La razón es que habitualmente H_0 impone restricciones en el espacio paramétrico mientras que H_1 supone que estas restricciones no existen.
- Representaremos el máximo de la función de verosimilitud como $f(H_1)$, que corresponde al estimador MV de los parámetros.
- Luego compararemos $f(H_0)$ y $f(H_1)$, construyendo el cociente que denominamos razón de verosimilitudes:

$$RV = \frac{f(H_0)}{f(H_1)}$$

Inferencia

- Por construcción, $RV \leq 1$ y rechazaremos H_0 cuando RV sea suficientemente pequeño. La región de rechazo de H_0 vendrá definida por $RV \leq a$ donde a se determinará de acuerdo al nivel de significación elegido para el test.
- Cuando el tamaño de la muestra es grande, se puede demostrar:

$$\lambda = -2log(RV) = 2(logf(H_1) - logf(H_0)) \sim_{H_0} \chi^2_{gl}$$

- donde los grados de libertad se determinan a partir de la diferencia de dimensión entre los espacios Ω y Ω_0.
- Genéricamente Ω es de dimensión s y la dimensión de Ω_0 es $s-r$, siendo r el número de restricciones lineales sobre el vector de parámetros. En ese caso, la cantidad de grados de libertad del estadístico es igual a r.

Test para la media de una población normal

- Consideremos una muestra $(x_1, x_2, \ldots, x_n)$ donde $x_i \sim_{iid} N_p(\mu, \Sigma)$, $i = 1, 2, \ldots, n$ y se desea probar:

  $$H_0)\mu = \mu_0,\ \Sigma = \text{cualquiera} \qquad H_1)\mu \neq \mu_0,\ \Sigma = \text{cualquiera}$$

  Bajo H_1, $\hat{\mu} = \bar{x}$ y $\widehat{\Sigma} = S$, por lo tanto:

  $$\boxed{L(H_1) = -\frac{n}{2} log|S| - \frac{np}{2}}$$

  Bajo H_0, $\hat{\mu} = \mu_0$:

  $$\boxed{L(H_0) = -\frac{n}{2} log|S_0| - \frac{np}{2}}$$

  Por lo tanto,

  $$\boxed{\lambda = 2(L(H_1) - L(H_0)) = n log \frac{|S_0|}{|S|}}$$

Test para la media de una población normal

- Se rechaza H_0 cuando $L(H_1)$ sea significativamente mayor que $L(H_0)$. Esta condición equivale a que la variancia generalizada bajo H_0, $|S_0|$ sea significativamente mayor que bajo H_1.
- Asintóticamente:
$$\lambda = nlog \frac{|S_0|}{|S|} \sim_{H_0} \chi^2_p$$
- La dimensión del espacio paramétrico bajo H_0 es $p + p(p-1)/2 = p(p+1)/2$, la cantidad de términos en Σ; mientras que la dimensión del espacio paramétrico bajo H_1 es $p + p(p+1)/2$.
- Fijado un nivel de significación α, si $\lambda^{obs} \geq \chi^2_{p,\alpha}$ entonces se rechaza H_0.

Test para la media de una población normal

- En este caso podemos obtener la distribución exacta del ratio de verosimilitudes. Es posible demostrar que

$$\frac{|S_0|}{|S|} = 1 + \frac{T^2}{n-1}$$

  donde el estadístico $T^2 = (n-1)(\bar{x} - \mu_0)' S^{-1} (\bar{x} - \mu_0)$ sigue la distribución T^2 de Hotelling con p y $n-1$ grados de libertad.
- Utilizando la relación entre el estadístico T^2 y F, podemos calcular los percentiles de T^2.
- Como la diferencia de soportes es una función monótona de T^2 podemos utilizar directamente este estadístico en lugar de la razón de verosimilitudes y rechazaremos H_0 cuando T^2 sea significativamente grande.

Tests sobre la matriz de Variancias y Covariancias

- Veremos 3 alternativas para el test sobre la matriz de variancias y covariancias de variables normales:
  1. Hipótesis nula: la matriz de variancias y covariancias toma un valor fijo dado
  2. Hipótesis nula: la matriz de variancias y covariancias es diagonal y por lo tanto las variables no están correlacionadas
  3. Hipótesis nula: la matriz de variancias y covariancias es diagonal, por lo tanto las variables no están correlacionadas, y además las variables tienen todas la misma variancia (test de esfericidad).

Tests sobre la matriz de Variancias y Covariancias

- Supongamos que se desea realizar un test para las hipótesis:

$$H_0)\Sigma = \Sigma_0,\ \mu\text{=cualquiera} \qquad H_1)\Sigma \text{ y } \mu\text{=cualquiera}$$

- Se calcula el máximo de la función de verosimilitud bajo H_0 y H_1.

$$L(\mu, \Sigma/x) = -\frac{n}{2} log|\Sigma| - \frac{n}{2} tr(\Sigma^{-1}S) - \frac{n}{2}(\bar{x}-\mu)'\Sigma^{-1}(\bar{x}-\mu)$$

Bajo H_0 el valor de $\widehat{\Sigma} = \Sigma_0$ y $\hat{\mu} = \bar{x}$, por lo tanto

$$\boxed{L(H_0) = -\frac{n}{2} log|\Sigma_0| - \frac{n}{2} tr(\Sigma_0^{-1}S)}$$

Mientras que bajo H_1 resulta $\hat{\mu} = \bar{x}$ y $\widehat{\Sigma} = S$

$$\boxed{L(H_1) = -\frac{n}{2} log|S| - \frac{np}{2}}$$

Tests sobre la matriz de Variancias y Covariancias

- El estadístico λ será:

$$\boxed{\lambda = 2(L(H_1) - L(H_0)) = n \cdot log\frac{|\Sigma_0|}{|S|} + n \cdot tr(\Sigma_0^{-1}S) - np}$$

- El test consiste en comparar el valor teórico Σ_0 con respecto al estimado S en términos del determinante y la traza.
- Asintóticamente

$$\lambda \sim \chi^2_{p(p+1)/2}$$

- Este test sirve para constrastar si $\Sigma_0 = I$ y el estadístico resulta:

$$\boxed{\lambda = 2(L(H_1) - L(H_0)) = -nlog|S| + n \cdot tr(S) - np}$$

y tiene la misma distribución que antes.

Tests sobre la matriz de Variancias y Covariancias

- Test de independencia:

  $H_0) \Sigma = \Sigma_0$ diagonal, μ=cualquiera $\qquad H_1) \Sigma$ y μ=cualquiera

- Bajo H_0, $\widehat{\Sigma}_0 = Diag(S)$, y el estadístico del test es el siguiente:

$$\lambda = n \cdot log \frac{\prod_{i=1}^{p} s_{ii}}{|S|} + n \cdot tr(\widehat{\Sigma_0}^{-1} S) - np$$

y como $tr(\widehat{\Sigma_0}^{-1} S) = tr(\widehat{\Sigma_0}^{-1/2} S \widehat{\Sigma_0}^{-1/2}) = tr(R) = p$ y además $|R| = |\widehat{\Sigma_0}^{-1}||S|$ el test estadístico se reduce a

$$\boxed{\lambda = -n \cdot log|R|}$$

y su distribución asintótica será χ^2 con $p(p-1)/2$ grados de libertad.

Tests sobre la matriz de Variancias y Covariancias

- Test de esfericidad:
- Se plantean las siguientes hipótesis:

  $$H_0)\Sigma = \sigma^2 I, \ \mu\text{=cualquiera} \qquad H_1)\ \Sigma \text{ y } \mu\text{=cualquiera}$$

  Haciendo las sustituciones necesarias, se obtiene bajo H0:

  $$\boxed{L(H_0) = -\frac{np}{2} log(\sigma^2) - \frac{n}{2\sigma^2} tr(S)}$$

  y derivando con respecto a σ^2 es fácil comprobar que el estimador máximo verosímil de σ^2 es $tr(S)/p$, la variancia media estimada.
- La función soporte bajo la hipótesis alternativa es la misma que en el test anterior.

Tests sobre la matriz de Variancias y Covariancias

- El estadístico λ resulta:

$$\lambda = n \cdot log \frac{\hat{\sigma}^{2p}}{|S|} + \frac{n \cdot tr(S)}{\hat{\sigma}^2} - np$$

y sustituyendo $\hat{\sigma}^2 = tr(S)/p$:

$$\boxed{\lambda = np \cdot log(\hat{\sigma}^2) - nlog(|S|)}$$

Su distribución asintótica sera χ^2 con $p(p+1)/2-1 = (p+2)(p-1)/2$ grados de libertad.

Ajustes en la distribución

- La aproximación de la distribución del estadístico λ a la χ^2 cuando el tamaño de muestra no es muy grande puede mejorarse introduciendo factores de ajuste o corrección.
- Box (1949) y Bartlett (1954) han demostrado que las aproximaciones mejoran si sustituimos en los estadísticos anteriores n por n_c, donde n_c es menor que n y depende de p y del test.
- Por ejemplo Box demostró que el test de independencia mejora si sustituimos n por $n_c = n - (2p + 11)/2$.
- Estas correcciones pueden ser importantes si el tamaño muestral es pequeño, p es grande y el estadístico obtenido está cerca del valor crítico, pero no van a ser importantes si p/n es pequeño y el valor observado del estadístico es claramente concluyente en cualquiera de las dos direcciones (rechazo o no rechazo de la H_0).

Test de igualdad de medias

- Supongamos que hemos observado una muestra de tamaño n de una variable p dimensional que puede estratificarse en G grupos, de manera que existen n_1 observaciones que pertenecen al grupo 1, n_2 observaciones que pertenecen al grupo 2, etc.
- Un problema interesante es comprobar si las medias de los G grupos son iguales.
- Las hipótesis a contrastar son:

$$H_0) \mu_1 = \mu_2 = \cdots = \mu$$

$$H_1) \text{ no todas las medias } \mu_i \text{ son iguales}$$

  y además Σ es definida positiva e idéntica en todos los grupos.
- La función de verosimilitud bajo H_0 de una muestra normal homogénea ya se ha calculado:

$$\boxed{L(H_0) = -\frac{n}{2} log|S| - \frac{np}{2}}$$

Test de igualdad de medias

- Bajo H_1 los n vectores de observaciones se subdividen en G grupos diferentes. La función de verosimilitud bajo H_1 sera:

$$f(\mu_1, \ldots, \mu_G, \Sigma / x) =$$

$$= \frac{1}{|\Sigma|^{\frac{n}{2}} (2\pi)^{\frac{np}{2}}} e^{-\frac{1}{2}\left(\sum_{g=1}^{G} \sum_{j=1}^{n_g} (x_{jg} - \mu_g)' \Sigma^{-1} (x_{jg} - \mu_g)\right)}$$

  donde x_{jg} es el vector j de variables del grupo g y μ_g es su media.
- La maximización de esta función en el espacio paramétrico definido por H_1 implica utilizar la media muestral de cada grupo como estimador de μ_g, y la estimación de la matriz de variancias y covariancias común será S_w.

Test de igualdad de medias

$$\sum_{g=1}^{G}\sum_{j=1}^{n_g}(x_{jg}-\bar{x}_g)'\Sigma^{-1}(x_{jg}-\bar{x}_g) =$$

$$tr\left(\sum_{g=1}^{G}\sum_{j=1}^{n_g}(x_{jg}-\bar{x}_g)'\Sigma^{-1}(x_{jg}-\bar{x}_g)\right) =$$

$$tr\left(\sum_{g=1}^{G}\sum_{j=1}^{n_g}\Sigma^{-1}(x_{jg}-\bar{x}_g)(x_{jg}-\bar{x}_g)'\right) = tr(\Sigma^{-1}W)$$

Donde la matriz W es la matriz de la suma de cuadrados dentro de los grupos:

$$W = \sum_{g=1}^{G}\sum_{j=1}^{n_g}(x_{jg}-\bar{x}_g)(x_{jg}-\bar{x}_g)'$$

Test de igualdad de medias

- Sustituyendo en la función de verosimilitud y tomando logaritmos se obtiene:
$$L(\Sigma/x) = \frac{n}{2}log|\Sigma^{-1}| - \frac{n}{2}tr(\Sigma^{-1}W/n)$$
- La variancia común a los grupos cuando estos tienen distintas medias se estima como :
$$\widehat{\Sigma} = S_w = \frac{W}{n}$$
- Sustituyendo estas expresiones en la función soporte, obtenemos:
$$\boxed{L(H_1) = -\frac{n}{2}log|S_w| - \frac{np}{2}}$$
- El estadístico del test resulta:
$$\boxed{\lambda = nlog\frac{|S|}{|S_w|}}$$

Test de igualdad de medias

- La distribución asintótica de λ es χ^2 con $p(G-1)$ grados de libertad:
- bajo la hipótesis nula hay que estimar los p elementos del vector de medias, común a todos los grupos y la matriz de variancias y covariancias, sumando en total $p + p(p-1)/2$ parámetros.
- bajo la hipótesis alternativa, hay que estimar G vectores de medias y la matriz de variancias y covariancias, lo que supone estimar $Gp + p(p-1)/2$ parámetros.
- Rechazaremos H_0 cuando λ sea suficientemente grande, reflejando que la variabilidad suponiendo que H_0 es cierta y medida por S sea mucho mayor a la variabilidad cuando permitimos que las medias de los grupos sean distintas, medidas por S_w.
- La aproximación a la distribución χ^2 del cociente de verosimilitudes puede mejorarse para tamaños de muestra pequeños reemplazando n por $m = (n-1) - (p-G)/2$.

Test de igualdad de medias

- Un enfoque alternativo: llamemos variabilidad total de los datos a la suma que mide los desvíos con respecto a la media general:

$$T = \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})'$$

- Vamos a descomponer la matriz $T = B + W$.
- La matriz W es la matriz de desvíos con respecto a la media de cada grupo que definimos antes. La matriz B medirá la variabilidad explicada por las diferencias entre las medias.

$$T = \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})' = \sum_{g=1}^{G} \sum_{j=1}^{n_g} (x_{jg} - \bar{x}_g + \bar{x}_g - \bar{x})(x_{jg} - \bar{x}_g + \bar{x}_g - \bar{x})'$$

$$T = \sum_{g=1}^{G} \sum_{j=1}^{n_g} (x_{jg} - \bar{x}_g)(x_{jg} - \bar{x}_g)' + \sum_{g=1}^{G} \sum_{j=1}^{n_g} (\bar{x}_g - \bar{x})(\bar{x}_g - \bar{x})'$$

Test de igualdad de medias

- La matriz de variabilidad explicada o sumas de cuadrados entre grupos resulta:

$$B = \sum_{g=1}^{G} n_g(\bar{x}_g - \bar{x})(\bar{x}_g - \bar{x})'$$

- Para plantear el test de igualdad de medias podemos comparar el tamaño de las matrices W y B, calculando el cociente $|B + W|/|W|$ cuya distribución se puede aproximar por la distribución F. Para tamaños de muestra moderados el test es similar al de razón de verosimilitudes que puede escribirse como:

$$\boxed{\lambda' = mlog\frac{|T|}{|W|} = mlog\frac{|B + W|}{|W|} = mlog|I + W^{-1}B| = m\sum_i log(1 + \lambda_i)}$$

- Expresando con λ_i a los autovalores de la matriz $W^{-1}B$.

Asimetría y Kurtosis

- La generalización de estos coeficientes para el caso multivariado no es inmediata. A continuación vemos una de las propuestas más utilizadas y se debe a Mardia (1970).
- Se propuso calcular las distancias de Mahalanobis para cada par de elementos muestrales (i, j):

$$d_{ij} = [(x_i - \bar{x})' S^{-1} (x_j - \bar{x})]$$

Coeficiente de asimetría multivariante:

$$A_p = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}^3$$

Coeficiente de kurtosis multivariante:

$$K_p = \frac{1}{n} \sum_{i=1}^{n} d_{ii}^2$$

Test de normalidad

- La normalidad de las distribuciones univariantes se puede comprobar a partir de muchos tests. En términos de los coeficientes de asimetría y kurtosis:

$$A = \frac{m_3}{m_2^{3/2}} \qquad K = \frac{m_4}{m_2^2}$$

- Donde llamamos con m_h a los momentos muestrales de orden h:

$$m_h = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^h$$

- Se puede demostrar que si los datos son normales, asintóticamente:

$$A \sim N(0, 6/n) \qquad K \sim N(3, 24/n)$$

Por lo tanto:

$$X^2 = \frac{nA^2}{6} + \frac{n(K-3)^2}{24} \sim_{H_0} \chi_2^2$$

Asimetría y Kurtosis

- La generalización de estos coeficientes para el caso multivariado no es inmediata. A continuación vemos una de las propuestas más utilizadas y se debe a Mardia (1970).
- Se propuso calcular las distancias de Mahalanobis para cada par de elementos muestrales (i, j):

$$d_{ij} = [(x_i - \bar{x})' S^{-1} (x_j - \bar{x})]$$

Coeficiente de asimetría multivariante:

$$A_p = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}^3$$

Coeficiente de kurtosis multivariante:

$$K_p = \frac{1}{n} \sum_{i=1}^{n} d_{ii}^2$$

Asimetría y Kurtosis

Propiedades:

- Para variables escalares, $A_p = A^2$:

$$A_p = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} d_{ij}^3 = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left[\frac{(x_i - \bar{x})(x_j - \bar{x})}{s^2} \right]^3 =$$

$$A_p = \frac{1}{n^2 s^6} \left[\sum_{i=1}^{n} (x_i - \bar{x})^3 \right]^2 = A^2$$

- El coeficiente de asimetría es no negativo, y solo será igual a cero si los datos son simétricos.

Asimetría y Kurtosis

Propiedades:

- Para variables escalares $K = K_p$.

$$d_{ii} = \left[\frac{(x_i - \bar{x})(x_i - \bar{x})}{s^2} \right]^2 = \frac{(x_i - \bar{x})^4}{s^4}$$

- Los coeficientes son invariantes ante transformaciones lineales de los datos.
- Si $y = ax + b$, los coeficientes de asimetría y kurtosis de x e y son idénticos.

Test de normalidad multivariada

- El supuesto de normalidad multivariada implica la normalidad de las distribuciones marginales, pero no siempre ocurre lo contrario. Veamos la generalización de los tests a partir de los coeficientes de asimetría y kurtosis:

$$A_p = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} d_{ij}^3 \qquad K_p = \frac{1}{n}\sum_{i=1}^{n} d_{ii}^2$$

  donde d_{ii} es la distancia de mahalanobis de cada observación con respecto a la media muestral.

- Asintóticamente, si los datos provienen de una población normal multivariada:

$$nA_p/6 \sim \chi^2_{\frac{p(p+1)(p+2)}{6}} \qquad K_p \sim N(p(p+2), 8p(p+2)/n)$$

- La potencia de estos tests no es muy alta a menos que se disponga de una muestra muy grande.

Test de normalidad multivariada

En la práctica se rechaza la hipótesis de normalidad debido a...:

- Las distribuciones marginales son aproximadamente simétricas y las relaciones entre las variables son lineales, pero existen observaciones atípicas que no pueden explicarse por la hipótesis de normalidad. En este caso, si eliminamos los valores atípicos, la hipótesis de normalidad conjunta no se rechaza y los métodos basados en el supuesto de normalidad suelen dar buenos resultados.
- Algunas (o todas) las distribuciones marginales son simétricas y existen relaciones no lineales entre las variables. Una solución simple que suele funcionar bastante bien en este caso es transformar las variables para obtener simetría en las distribuciones marginales y relaciones lineales entre las variables.

Anexo: Test de razón y T^2 de Hotelling

- Para demostrar la relación entre la razón de verosimilitudes y la T^2 de Hotelling utilizaremos el siguiente resultado:
- Lema 1: Si A es una matriz no singular y b un vector,

$$|I + Abb'| = 1 + b'Ab$$

- En efecto, la matriz bb' tiene rango 1 y también tendrá rango 1 el producto Abb'. Por lo tanto, Abb' tiene un único autovalor no nulo.
- Llamando λ a ese autovalor y v al autovector asociado, como $Abb'v = \lambda v$, multiplicando por b' se obtiene que $\lambda = b'Ab$. De este modo la matriz $I + Abb'$ tendrá un autovalor igual a $1 + \lambda$ y los restantes serán iguales a 1.
- Como el determinante es igual al producto de los autovalores de la matriz, el lema queda demostrado.

Anexo: Test de razón y T^2 de Hotelling

- Ahora consideremos la siguiente expresión:

$$nS_0 = \sum_{i=1}^{n} (x_i - \bar{x} + \bar{x} - \mu_0)(x_i - \bar{x} + \bar{x} - \mu_0)'$$

y desarrollando la expresión anterior asociando los términos $(x_i - \bar{x})$ y $(\bar{x} - \mu_0)$ resulta

$$nS_0 = nS + n(\bar{x} - \mu_0)(\bar{x} - \mu_0)'$$

por lo tanto, podemos escribir:

$$\frac{|S_0|}{|S|} = \frac{|S + (\bar{x} - \mu_0)(\bar{x} - \mu_0)'|}{|S|}$$

Anexo: Test de razón y T^2 de Hotelling

$$\frac{|S_0|}{|S|} = \frac{|S + (\bar{x} - \mu_0)(\bar{x} - \mu_0)'|}{|S|}$$

$$\frac{|S_0|}{|S|} = |S^{-1}||S + (\bar{x} - \mu_0)(\bar{x} - \mu_0)'| = |I + S^{-1}(\bar{x} - \mu_0)(\bar{x} - \mu_0)|$$

y aplicando el lema anterior

$$|I + S^{-1}(\bar{x} - \mu_0)(\bar{x} - \mu_0)| = 1 + (\bar{x} - \mu_0)' S^{-1}(\bar{x} - \mu_0)$$

con lo que finalmente obtenemos que

$$\frac{|S_0|}{|S|} = 1 + (\bar{x} - \mu_0)' S^{-1}(\bar{x} - \mu_0) = 1 + \frac{T^2}{n-1}$$