

# Propiedades de los estimadores de Máxima verosimilitud

Se puede probar que los estimadores de máxima verosimilitud poseen una propiedad de invarianza : Si  $\hat{\theta}$  es el MLE de  $\theta$  entonces el MLE de  $h(\theta)$  es  $h(\hat{\theta})$

**Por ejemplo**

$$h(\theta) = h(\mu, \Sigma) = \mu' \Sigma^{-1} \mu$$

El MLE de  $h(\theta)$  está dado por

$$h(\hat{\theta}) = h(\hat{\mu}, \hat{\Sigma}) = \hat{\mu}' \hat{\Sigma}^{-1} \hat{\mu} = \bar{x}' \left( \left( \frac{n-1}{n} \right) S \right)^{-1} \bar{x}$$

El MLE de  $h(\sigma_{ii}) = \sqrt{\sigma_{ii}}$  está dado por  $\sqrt{\hat{\sigma}_{ii}}$

$$\hat{\sigma}_{ii} = \frac{1}{n} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2$$

es el MLE de  $\sigma_{ii} = \text{Var}(x_i)$

# Propiedades de los estimadores de Máxima verosimilitud

La densidad conjunta sobre la muestra completa de observaciones  $\mathbf{x}_1, \dots, \mathbf{x}_n$ :

$$\begin{aligned} f(\mathbf{x}_1, \dots, \mathbf{x}_n) &= \frac{1}{(2\pi)^{np/2} |\mathbf{\Sigma}|^{n/2}} e^{-\frac{1}{2} \left[ \text{tr} \left[ \mathbf{\Sigma}^{-1} \left( \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' \right) \right] + n(\bar{\mathbf{x}} - \mu)\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}} - \mu)' \right]} \\ &= (2\pi)^{-np/2} |\mathbf{\Sigma}|^{-n/2} \exp \left\{ -\frac{1}{2} \left[ \text{tr} \left[ \mathbf{\Sigma}^{-1} \left( \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' \right) \right] + n(\bar{\mathbf{x}} - \mu)\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}} - \mu)' \right] \right\} \end{aligned}$$

depende unicamente de  $\bar{\mathbf{x}}$  y  $(n-1)\mathbf{S}$

Toda la información sobre los parámetros de la población normal multivariada,  $\mu$  y  $\mathbf{\Sigma}$ , está contenida en  $\bar{\mathbf{x}}$  y  $(n-1)\mathbf{S}$  obtenida a traves de la matriz de datos  $\mathbf{X}$ . Entonces decimos que  $\bar{\mathbf{x}}$  y  $(n-1)\mathbf{S}$  son estadísticos suficientes para la densidad conjunta normal multivariada.

# Distribuciones muestrales de $\bar{\mathbf{x}}$ y $\mathbf{S}$

- La suposición de que  $\mathbf{x}_1, \dots, \mathbf{x}_n$  constituye una muestra aleatoria de una población normal multivariada con media  $\boldsymbol{\mu}$  y covarianza  $\boldsymbol{\Sigma}$  determina completamente las distribuciones muestrales de  $\bar{\mathbf{x}}$  y  $\mathbf{S}$ .
- En el caso univariado ( $p = 1$ ) sabemos que  $\bar{x}$  es normal con

$$\text{media} = \mu \quad \text{y} \quad \text{varianza} = \frac{1}{n} \sigma^2 = \frac{\text{varianza poblacional}}{\text{tamaño de la muestra}}$$

- Análogamente para el caso multivariado ( $p \geq 2$ ),  $\bar{\mathbf{x}}$  es normal multivariada con

$$\text{media} = \boldsymbol{\mu} \quad \text{y} \quad \text{matriz de covarianzas} = \frac{1}{n} \boldsymbol{\Sigma}$$

# Distribuciones muestrales de $\bar{x}$ y $S$

Similarmente, para una muestra aleatoria  $x_1, \dots, x_n$  de una distribución normal univariada con media  $\mu$  y varianza  $\sigma^2$

$$(n-1)s^2 = \sum_{j=1}^n (x_j - \bar{x})^2 \sim \sigma^2 \chi_{n-1}^2 \sim \sigma^2 \sum_{j=1}^{n-1} z_j^2 = \sum_{j=1}^{n-1} (\sigma z_j)^2$$

donde

$$\sigma z_j \sim N(0, \sigma^2), j = 1, \dots, n-1 \text{ e independientes}$$

Análogamente, para el caso multivariado ( $p \geq 2$ ) (i.e.,  $\mathbf{x}$  es normal con media  $\mu$  y covarianza  $\Sigma$ ),  $\mathbf{S}$  sigue una distribución Wishart, denotada por  $W_m(\bullet|\Sigma)$ , donde

$W_m(\bullet|\Sigma)$  = distribución Wishart con  $m$  grados de libertad

$$= \text{distribución de } \sum_{j=1}^m \mathbf{z}_j \mathbf{z}_j', \text{ donde } \mathbf{z}_j \sim N_p(0, \Sigma) \text{ y } \mathbf{z}_j \text{ independientes}$$

# Propiedades importantes de la distribución Wishart

- La distribución Wishart existe únicamente si  $n > p$
- Si

$$\mathbf{A}_1 \sim W_{m_1}(\mathbf{A}_1|\Sigma)$$

independientemente de

$$\mathbf{A}_2 \sim W_{m_2}(\mathbf{A}_2|\Sigma)$$

entonces

$$\mathbf{A}_1 + \mathbf{A}_2 \sim W_{m_1+m_2}(\mathbf{A}_1 + \mathbf{A}_2|\Sigma)$$

donde  $\Sigma$  es una matriz de covarianza común.

- Si  $\mathbf{A} \sim W_m(\mathbf{A}|\Sigma)$  entonces

$$\mathbf{CAC}' \sim W_m(\mathbf{CAC}'|\mathbf{C}\Sigma\mathbf{C}')$$

donde  $\mathbf{C}$  es una matriz de constantes de orden  $h \times p$

# Propiedades importantes de la distribución Wishart

- Estas propiedades se obtienen directamente de su definición como suma de los productos independientes  $\mathbf{z}_j \mathbf{z}_j'$
- Cuando existe, la distribución Wishart tiene una densidad de la forma

$$W_{n-1}(\mathbf{A}|\mathbf{\Sigma}) = \frac{|\mathbf{A}|^{(n-p-2)/2} e^{-tr(\mathbf{A}\mathbf{\Sigma}^{-1})/2}}{2^{p(n-1)/2} \pi^{p(p-1)/4} |\mathbf{\Sigma}|^{(n-1)/2} \prod_{i=1}^p \Gamma\left(\frac{1}{2}(n-i)\right)}$$

para una matriz simétrica definida positiva  $\mathbf{A}$  y donde  $\Gamma$  es la función gamma.

# Resumen de las distribuciones muestrales de $\bar{x}$ y $S$

Las distribuciones de  $\bar{x}$  y  $S$  se resumen en el siguiente resultado:

**Resultado:** Sea  $x_1, \dots, x_n$  una muestra aleatoria de tamaño  $n$  obtenida de una distribución normal  $p$ -variada, con media  $\mu$  y matriz de covarianza  $\Sigma$ . Entonces

- 1  $\bar{x} \sim N_p(\mu, \frac{\Sigma}{n})$
- 2  $(n-1)S \sim W_{n-1}(\bullet | \Sigma)$  con  $n-1$  grados de libertad
- 3  $\bar{x}$  y  $S$  son independientes

# Comportamiento de $\bar{x}$ y $S$ para muestras grandes

- El teorema central del límite (Univariado) – supone que

$$x = \sum_{i=1}^n V_i$$

donde las  $V_i$  son independientes y tienen una variabilidad aproximadamente igual.

- Entonces la distribución de  $x$  llega a ser aproximadamente normal conforme el tamaño de la muestra  $n$  aumente.
- Esto se cumple, *no importando la distribución de la población de donde se obtiene la muestra*
- Como consecuencia del TCL la distribución muestral de  $\bar{x}$  para un tamaño de muestra grande, se aproxima a una normal sin importar la forma de la distribución de la población raíz.



# Comportamiento de $\bar{\mathbf{x}}$ y $\mathbf{S}$ para muestras grandes

- Para el caso multivariado, estadísticos como  $\bar{\mathbf{x}}$  y  $\mathbf{S}$  tienen propiedades análogas para muestras grandes.
- En particular para  $\bar{\mathbf{x}}$  se puede enunciar la versión multivariada del teorema central del límite.

## Teorema central del límite (versión multivariada)

Sean  $\mathbf{x}_1, \dots, \mathbf{x}_n$  observaciones independientes de cualquier población con media  $\boldsymbol{\mu}$  y covarianza finita  $\boldsymbol{\Sigma}$  (no singular).

Entonces:

$$\bar{\mathbf{x}} \sim N_p \left[ \boldsymbol{\mu}, \frac{1}{n} \boldsymbol{\Sigma} \right]$$

para  $n$  grande y mayor a  $p$ .

Esto es equivalente a

$$\sqrt{n}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \overset{\sim}{\sim} N_p[\mathbf{0}, \boldsymbol{\Sigma}]$$

para  $n$  grande con respecto a  $p$ .

# Comportamiento de $\bar{\mathbf{x}}$ y $\mathbf{S}$ para muestras grandes

- Cuando  $n$  es grande, la matriz de covarianza muestral  $\mathbf{S}$  se aproxima a la covarianza poblacional  $\mathbf{\Sigma}$  con alta probabilidad.
- Frecuentemente se sustituye la covarianza muestral  $\mathbf{S}$  en la covarianza poblacional  $\mathbf{\Sigma}$  lo cual tiene efectos insignificantes en los cálculos de probabilidades posteriores.
- Por tanto se tiene que

$$\bar{\mathbf{x}} \sim N_p \left[ \mu, \frac{1}{n} \mathbf{S} \right]$$

para  $n$  grande y mayor a  $p$ .

Esto es equivalente a

$$\sqrt{n}(\bar{\mathbf{x}} - \mu) \sim N_p[0, \mathbf{S}]$$

de nuevo para  $n$  grande con respecto a  $p$ .

# Comportamiento de $\bar{\mathbf{x}}$ y $\mathbf{S}$ para muestras grandes

Otro resultado importante como consecuencia del TCL es:

$$n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu}) \overset{\circ}{\sim} \chi_p^2$$

para  $n$  grande y mayor a  $p$ .

Este resultado se obtiene de acuerdo a una de las propiedades de la normal multivariada y por el hecho de que

$$\bar{\mathbf{x}} \sim \mathbf{N}_p \left[ \boldsymbol{\mu}, \frac{1}{n} \mathbf{S} \right]$$

para  $n$  grande y mayor a  $p$ .

Entonces sustituimos  $\mathbf{x}$  por  $\bar{\mathbf{x}}$  y  $\boldsymbol{\Sigma}$  por  $\mathbf{S}/n$  en

$$(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})$$

# La $T^2$ de Hotelling

- Si  $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , la variable  $(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$  sigue una distribución  $\chi_p^2$
- Si sustituimos  $\boldsymbol{\Sigma}$  por su estimación  $\hat{\boldsymbol{\Sigma}} = \mathbf{S}$ , la distribución que se obtiene se denomina  $T^2$  de Hotelling.
- En general, si  $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  y  $(n-1)\mathbf{S} \sim W_p(\mathbf{S}|\boldsymbol{\Sigma})$ , la distribución de la variable escalar

$$T^2 = (\mathbf{x} - \boldsymbol{\mu})' \mathbf{S}^{-1}(\mathbf{x} - \boldsymbol{\mu})$$

se denomina distribución  $T^2$  de Hotelling con  $p$  y  $n-1$  grados de libertad. Diremos que  $T^2 \sim T^2(p, n-1)$

- Como  $\mathbf{S}$  converge a  $\boldsymbol{\Sigma}$  cuando  $n$  es grande, entonces  $T^2$  converge a la distancia Mahalanobis y por tanto la distribución de Hotelling converge a la distribución  $\chi_p^2$

# La $T^2$ de Hotelling

- Sea  $\bar{\mathbf{x}}$  la media muestral, sabemos que  $\bar{\mathbf{x}} \sim N_p(\mu, \frac{\Sigma}{n})$ , entonces la distribución de

$$(\bar{\mathbf{x}} - \mu)' (\mathbf{S}^{-1}/n) (\bar{\mathbf{x}} - \mu) = n(\bar{\mathbf{x}} - \mu)' \mathbf{S}^{-1} (\bar{\mathbf{x}} - \mu)$$

es también una  $T^2$  de Hotelling.

- Observemos que si  $p = 1$ , la  $T^2$  se reduce a

$$T^2 = \frac{n(\bar{x} - \mu)^2}{s^2} = t^2$$

y coincide con el estadístico  $t$  de *Student*. Por tanto

$$T^2(1, m) = t_m^2$$

# La $T^2$ de Hotelling

- La  $T^2$  de Hotelling está también relacionada con la distribución  $F$  del análisis de varianza de Fisher. Se puede probar que mediante una simple transformación,  $T^2$  se reduce a la distribución  $F$ . Así, esta transformación esta dada por

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

- Esto permite calcular la distribución de  $T^2$  en función de las tablas de la distribución  $F$ .
- Este resultado es consistente con la convergencia de  $T^2$  a la  $\chi_p^2$ , cuando  $n$  es grande, ya que implica que  $pF_{p,n-p}$  tiende a una distribución  $\chi_p^2$

# Evaluando la suposición de normalidad

- Muchas de las técnicas estadísticas multivariadas, asumen que los vectores  $\mathbf{x}_i$  de observaciones provienen de una distribución normal multivariada.
- Sin embargo sabemos también que si la muestra de observaciones es grande y las técnicas dependen únicamente de  $\bar{\mathbf{x}}$  o de distancias que implican  $\bar{\mathbf{x}}$  de la forma  $n(\bar{\mathbf{x}} - \boldsymbol{\mu})' \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})$ , entonces la suposición de normalidad para las observaciones individuales  $\mathbf{x}_i$  no es tan crucial (esto por el TCL).
- En cualquiera de estas situaciones, la calidad de las inferencias que se harán dependerá de que tan cercana esté la población real raíz a la forma normal multivariada supuesta.
- Por tanto es necesario tener procedimientos para detectar los casos en los que los datos se desvían de forma moderada a extrema de una distribución normal multivariada

# Evaluando la suposición de normalidad

**De acuerdo a las propiedades de la distribución normal multivariada, sabemos que**

- Todas las combinaciones lineales de normales son normales
- Los contornos de la densidad normal multivariada son elipsoides concéntricos

**Estos hechos sugieren la investigación de las siguientes preguntas:**

- 1 Las distribuciones marginales de los elementos de  $\mathbf{x}$  son normales? Que se puede decir acerca de algunas pocas combinaciones lineales de los componentes  $x_i$ ?
- 2 Los scatterplots bivariados tendrán forma de elipses, como se esperaría de poblaciones normales multivariadas?
- 3 Existen algunas observaciones inusuales (outliers) que deben ser revisadas?



# Evaluando la suposición de normalidad

- Nos enfocaremos en evaluar la normalidad de las observaciones, en una o dos dimensiones, por ejemplo, distribuciones marginales y scatter plots de pares de variables.
- Es difícil construir una prueba global confiable de normalidad conjunta en más de dos dimensiones, debido a que existen un gran número de factores que no se pueden controlar, causando confusión, derivando en pruebas poco confiables.
- Sin embargo debemos pagar un precio por concentrarnos únicamente en examinar los casos univariados y bivariados de normalidad.

# Evaluando la suposición de normalidad

- Nunca podremos estar seguros que no hemos perdido alguna característica que es relevante únicamente en dimensiones mas grandes. Por ejemplo, es posible construir distribuciones bivariadas no-normales con marginales normales.
- Sin embargo muchos tipos de no-normalidad en los datos multivariados son frecuentemente reflejados en sus distribuciones marginales y scatter plots.
- En el trabajo práctico, la investigación de normalidad en una o dos dimensiones es basicamente suficiente.
- Afortunadamente no es muy frecuente encontrar conjuntos de datos que sean normales en una representación en baja dimensión, pero no-normales en dimensiones superiores.

# Evaluando la suposición de normalidad univariada

**Las herramientas usadas frecuentemente para determinar normalidad univariada incluyen**

- La regla empírica,

$$P(\mu - 1\sigma \leq x \leq \mu + 1\sigma) = 0.68$$

$$P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = 0.95$$

$$P(\mu - 3\sigma \leq x \leq \mu + 3\sigma) = 0.997$$

- Diagramas de puntos (para muestras pequeñas) e histogramas (para muestras grandes)
- Pruebas de bondad de ajuste tales como la prueba Chi-cuadrada y la prueba de Kolmogorov-Smirnov.
- La prueba desarrollada por Shapiro y Wilk [1965] llamada la prueba de Shapiro-Wilk.
- Q-Q plots (que grafica los cuantiles muestrales contra los cuantiles esperados si los datos siguen una distribución normal)

# Ejemplo: Evaluando la suposición de normalidad univariada

Supongamos que tenemos las siguientes 15 observaciones muestrales(ordenadas) sobre alguna variable aleatoria  $x$ :

Observaciones ordenadas $x_{(j)}$
1.43
1.62
2.46
2.48
2.97
4.03
4.47
5.76
6.61
6.68
6.79
7.46
7.88
8.92
9.42

Apoyan estos datos la suposición de que fueron elegidos de una población normal?

# Ejemplo: Evaluando la suposición de normalidad univariada

Para determinar normalidad por la regla empírica, estandarizamos los datos a una variable aleatoria normal estándar. Para nuestros datos tenemos

$$\bar{x} = 5.26, \sqrt{s} = 2.669$$

*así los valores correspondientes  
de la normal estandar para  
nuestros datos son*

*Nueve de las observaciones (o  
60%) están en una desviación  
estandar de la media, y las 15  
observaciones están en dos des-  
viaciones estandar de la media.  
Esto apoya la suposición de que  
fueron elegidas de una pobla-  
ción normal?*

Ordered Observations $x_{(j)}$	Standard Normal Variable $z_{(j)}$
1.43	-1.436
1.62	-1.367
2.46	-1.050
2.48	-1.045
2.97	-0.860
4.03	-0.461
4.47	-0.299
5.76	0.185
6.61	0.504
6.68	0.530
6.79	0.570
7.46	0.822
7.88	0.981
8.92	1.371
9.42	1.556

# Ejemplo: Evaluando la suposición de normalidad univariada

Se puede utilizar R o cualquier otro paquete para calcular la prueba de Shapiro-Wilk y el p-valor correspondiente:

```
#R#  
x <-c(1.43,1.62,2.46,2.48,2.97,4.03,  
4.47,5.76,6.61,6.68,6.79,7.46,7.88,8.92,9.42)  
shapiro.test(X)  
  
DATA stuff;  
INPUT x;  
LABEL x='Observed Values of X';  
CARDS;  
1.43  
1.62  
2.46  
2.48  
2.97  
4.03  
4.47  
5.76  
6.61  
6.68  
6.79  
7.46  
7.88  
8.92  
9.42  
;  
PROC UNIVARIATE DATA=stuff NORMAL;  
TITLE4 'Using PROC UNIVARIATE for tests of univariate normality';  
VAR x;  
RUN;
```

# Ejemplo: Evaluando la suposición de normalidad univariada

```
Tests for Normality
```

Test	--Statistic--	-----p Value-----
Shapiro-Wilk	W 0.935851	Pr < W 0.3331
Kolmogorov-Smirnov	D 0.159493	Pr > D >0.1500
Cramer-von Mises	W-Sq 0.058767	Pr > W-Sq >0.2500
Anderson-Darling	A-Sq 0.362615	Pr > A-Sq >0.2500

- Una herramienta gráfica también muy útil para evaluar la suposición de normalidad son los Q-Q plot.
- Los Q-Q plot grafican los cuantiles muestrales vs los cuantiles esperados si las observaciones fueran normalmente distribuidas.
- Si los puntos están muy cercanos a una línea recta de 45 grados, la suposición de normalidad es coherente.
- Si los puntos están alejados de esta línea, entonces la normalidad de los datos está en entredicho



# Construcción de un Q-Q plot

- 1 Colocamos los valores observados en orden creciente,  $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ . Los  $x_{(j)}$  son los cuantiles muestrales.
- 2 Se calcula el nivel corregido de continuidad de la probabilidad acumulada  $(j - 0.5)/n$  para los datos de la muestra.
- 3 Se encuentran los cuantiles normales estandar (valores de la distribución  $N(0,1)$ ) que tienen un nivel de probabilidad acumulada de  $(j - 0.5)/n$  – llamando a estos los  $q_{(j)}$ . Es decir, se encuentra un valor de  $z$  tal que

$$p[z \leq q_{(j)}] = \int_{-\infty}^{q_{(j)}} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \frac{j - \frac{1}{2}}{n}$$

- 4 Se grafican los pares  $(q_{(j)}, x_{(j)})$ . Si los puntos están sobre o cerca de una línea recta, las observaciones apoyan la suposición de que los *datos pudieron ser elegidos de una población normal*.

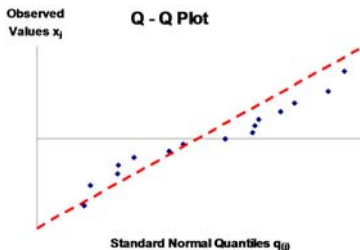
# Ejemplo de como obtener un Q-Q Plot

Obtener el Q-Q plot para la muestra de 15 datos:

Ordered Observations $x_{(j)}$	Adjusted Probability Level $(j-0.5)/n$	Standard Normal Quantiles $q_{(j)}$
1.43	0.033	-1.834
1.62	0.100	-1.282
2.46	0.167	-0.967
2.48	0.233	-0.728
2.97	0.300	-0.524
4.03	0.367	-0.341
4.47	0.433	-0.168
5.76	0.500	0.000
6.61	0.567	0.168
6.68	0.633	0.341
6.79	0.700	0.524
7.46	0.767	0.728
7.88	0.833	0.967
8.92	0.900	1.282
9.42	0.967	1.834

# Ejemplo de como obtener un Q-Q Plot

...y el Q-Q plot resultante tiene la forma



- No aparecen grandes desviaciones de la línea de 45° dibujada a través de los puntos, pero el ajuste tampoco es muy bueno. . .
- Los QQ-plots son no son particularmente informativos a menos que la muestra sea suficientemente grande
- Puede existir un poco de variabilidad en la rectitud del QQ-plot para muestras pequeñas aun cuando los datos provienen de una población normal

# Prueba de normalidad basada en la rectitud del QQ plot

- La rectitud o linealidad de los puntos en el QQ plot se puede medir mediante el coeficiente de correlación de los puntos  $q_{(j)}$  y  $x_{(j)}$
- El coeficiente de correlación entre  $q_{(j)}$  y  $x_{(j)}$  mide la relación lineal entre ellos, entonces entre mayor sea la correlación, mayor será la relación lineal entre  $q_{(j)}$  y  $x_{(j)}$ .
- Por tanto mayor la rectitud del Q-Q plot y en consecuencia se concluirá que los datos son normales.
- Esta prueba es una versión modificada de la prueba de Shapiro-Wilks. Para muestras grandes, las dos pruebas son aproximadamente la misma.

# Prueba de normalidad basada en la rectitud del QQ plot

- Esta prueba fue construida por Looney & Gulledge [1985]. Ellos sugieren calcular el coeficiente de correlación de Pearson entre  $q_{(j)}$  y  $x_{(j)}$  mediante la expresión

$$r_Q = \frac{\sum_{j=1}^n (x_{(j)} - \bar{x})(q_{(j)} - \bar{q})}{\sqrt{\sum_{j=1}^n (x_{(j)} - \bar{x})^2} \sqrt{\sum_{j=1}^n (q_{(j)} - \bar{q})^2}}$$

- Se rechaza la hipótesis de normalidad en un nivel de significancia  $\alpha$  si  $r_Q$  está por debajo de los puntos críticos dados.

# Tabla de los valores $\alpha$ de la prueba para $r_Q$

Critical Points for Q-Q Plot Correlation Coefficient Test for Normality\*

Sample Size n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
5	0.8299	0.8788	0.9032
10	0.8801	0.9198	0.9351
15	0.9126	0.9389	0.9503
20	0.9269	0.9508	0.9604
25	0.9410	0.9591	0.9665
30	0.9479	0.9652	0.9715
35	0.9538	0.9682	0.9740
40	0.9599	0.9726	0.9771
45	0.9632	0.9749	0.9792
50	0.9671	0.9768	0.9809
55	0.9695	0.9787	0.9822
60	0.9720	0.9801	0.9836
75	0.9771	0.9838	0.9866
100	0.9822	0.9873	0.9895
150	0.9879	0.9913	0.9928
200	0.9905	0.9931	0.9942
300	0.9935	0.9953	0.9960

\*Table 4.2 p. 182 from Johnson, R.A. & Wichern, D.W. *Applied Multivariate Statistical Analysis* (5<sup>th</sup> Ed.)

# Ejemplo: Prueba de normalidad basada en la rectitud del QQ plot

Para el ejemplo, los cálculos intermedios para obtener  $r_Q$  están dados en la siguiente tabla:

$x_{(j)} - \bar{x}$	$(x_{(j)} - \bar{x})^2$	$q_{(j)} - \bar{q}$	$(q_{(j)} - \bar{q})^2$	$(x_{(j)} - \bar{x})(q_{(j)} - \bar{q})$
-3.83	14.697	-1.834	3.363	7.031
-3.65	13.314	-1.282	1.642	4.676
-2.80	7.855	-0.967	0.936	2.711
-2.79	7.777	-0.728	0.530	2.030
-2.30	5.270	-0.524	0.275	1.204
-1.23	1.514	-0.341	0.116	0.419
-0.80	0.637	-0.168	0.028	0.134
0.49	0.244	0.000	0.000	0.000
1.35	1.810	0.168	0.028	0.226
1.41	2.001	0.341	0.116	0.482
1.52	2.315	0.524	0.275	0.798
2.19	4.808	0.728	0.530	1.596
2.62	6.860	0.967	0.936	2.534
3.66	13.387	1.282	1.642	4.689
4.15	17.235	1.834	3.363	7.614
0.00	99.724	0.000	13.781	36.143

## Ejemplo: Prueba de normalidad basada en la rectitud del QQ plot

Se evalúa el coeficiente de correlación de Pearson entre  $q_{(j)}$  y  $x_{(j)}$ , obteniendo:

$$\begin{aligned} r_Q &= \frac{\sum_{j=1}^n (x_{(j)} - \bar{x})(q_{(j)} - \bar{q})}{\sqrt{\sum_{j=1}^n (x_{(j)} - \bar{x})^2} \sqrt{\sum_{j=1}^n (q_{(j)} - \bar{q})^2}} \\ &= \frac{36.143}{\sqrt{99.724} \sqrt{13.781}} \\ &= 0.9749513 \end{aligned}$$

El tamaño de muestra es  $n = 15$ , así los puntos críticos para la prueba de normalidad son 0.9503 en  $\alpha = 0.10$ , 0.9389 en  $\alpha = 0.05$ , y 0.9126 en  $\alpha = 0.01$ . Por tanto no rechazamos la hipótesis de normalidad en cualquier  $\alpha$  mas grande que 0.01.



# Evalando normalidad Bivariada

- Al abordar el tema de la normalidad multivariada, las herramientas anteriores ayudan en la determinación de normalidad para las distribuciones marginales univariadas.
- Sin embargo, también debemos evaluar las distribuciones bivariadas (que deben ser normales si la distribución conjunta global es normal multivariada).
- Los métodos comúnmente usados para evaluar la normalidad bivariada son:
  - Diagramas de dispersión
  - Diagramas Chi-Cuadrado

# Ejemplo: Evaluando normalidad Bivariada

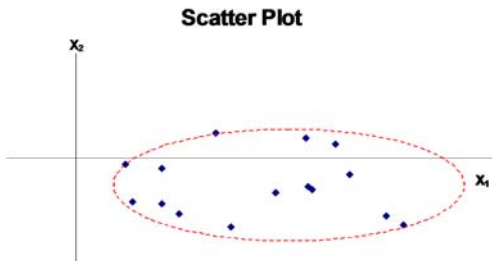
Consideremos las siguientes quince observaciones muestrales (ordenadas) tomadas de las variables aleatorias  $x_1$  y  $x_2$ :

$x_{j1}$	$x_{j2}$
1.43	-0.69
1.62	-5.00
2.46	-1.13
2.48	-5.20
2.97	-6.39
4.03	2.87
4.47	-7.88
5.76	-3.97
6.61	2.32
6.68	-3.24
6.79	-3.56
7.46	1.61
7.88	-1.87
8.92	-6.60
9.42	-7.64

¿Estos datos apoyan la suposición que fueron extraídos de una población padre normal bivariada?

# Ejemplo: Evaluando normalidad Bivariada

El diagrama de dispersión de los pares  $(x_1, x_2)$  tiene una forma parecida a un elipse, por tanto esto apoya la suposición que estos datos fueron extraídos de una distribución normal bivariada (y que tienen poco o nada de correlación).



# Evaluación por resultado de normal bivariada

- Otra forma de evaluar la normalidad bivariada de los datos, se obtiene a partir de un resultado de la normal multivariada: El conjunto de resultados bivariados  $\mathbf{x}$  tal que

$$(\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \leq \chi_2^2(0.5)$$

tiene probabilidad .5.

- Por tanto, se debería esperar que aproximadamente el mismo porcentaje, 50 %, de las observaciones muestrales estén en la elipse dada por

$$\{\forall \mathbf{x} : (\mathbf{x} - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \leq \chi_2^2(0.5)\}$$

donde  $\boldsymbol{\mu}$  se reemplaza por  $\bar{\mathbf{x}}$  y  $\boldsymbol{\Sigma}^{-1}$  por  $\mathbf{S}^{-1}$ . De lo contrario, la suposición de normalidad sería sospechosa.

# Ejemplo: Evaluación por resultado de normal bivariada

Consideremos la misma muestra de datos bivariados:

$x_{j1}$	$x_{j2}$
1.43	-0.69
1.62	-5.00
2.46	-1.13
2.48	-5.20
2.97	-6.39
4.03	2.87
4.47	-7.88
5.76	-3.97
6.61	2.32
6.68	-3.24
6.79	-3.56
7.46	1.61
7.88	-1.87
8.92	-6.60
9.42	-7.64

$$\bar{\mathbf{x}} = \begin{bmatrix} 5.26 \\ -3.09 \end{bmatrix}, \mathbf{S} = \begin{bmatrix} 7.12 & -0.67 \\ -0.67 & 12.43 \end{bmatrix}$$
$$\rightarrow \mathbf{S}^{-1} = \begin{bmatrix} 0.1411 & 0.0076 \\ 0.0076 & 0.0808 \end{bmatrix}$$
$$\chi^2_2(0.5) = 1.39$$

## Ejemplo: Evaluación por resultado de normal bivariada

Así, cualquier observación  $\mathbf{x} = (x_1, x_2)$  que cumple

$$\begin{pmatrix} x_1 - 5.26, & x_2 + 3.09 \end{pmatrix} \begin{pmatrix} 0.1411 & 0.0076 \\ 0.0076 & 0.0808 \end{pmatrix} \begin{pmatrix} x_1 - 5.26 \\ x_2 + 3.09 \end{pmatrix} \leq 1.39$$

cae dentro o sobre el contorno estimado del 50 %. De lo contrario, la observación está fuera de este contorno.

El primer par de observaciones es  $\mathbf{x} = (1.43, -0.69)$ , entonces

$$\begin{aligned} & \begin{pmatrix} 1.43 - 5.26, & -0.69 + 3.09 \end{pmatrix} \begin{pmatrix} 0.1411 & 0.0076 \\ 0.0076 & 0.0808 \end{pmatrix} \begin{pmatrix} 1.43 - 5.26 \\ -0.69 + 3.09 \end{pmatrix} \\ & \quad \begin{pmatrix} -3.83 & 2.4 \end{pmatrix} \begin{pmatrix} 0.1411 & 0.0076 \\ 0.0076 & 0.0808 \end{pmatrix} \begin{pmatrix} -3.83 \\ 2.4 \end{pmatrix} \\ & \quad = 2.4 > 1.39 \end{aligned}$$

Por tanto este punto cae fuera del contorno del 50 %

## Ejemplo:

En la tabla se presentan las distancias cuadradas generalizadas de cada dato al centroide  $\bar{x}$

$x_{j1}$	$x_{j2}$	$d^2_j$
1.43	-0.69	2.400
1.62	-5.00	2.279
2.46	-1.13	1.336
2.48	-5.20	1.548
2.97	-6.39	1.739
4.03	2.87	2.976
4.47	-7.88	2.005
5.76	-3.97	0.090
6.61	2.32	2.737
6.68	-3.24	0.281
6.79	-3.56	0.333
7.46	1.61	2.622
7.88	-1.87	1.138
8.92	-6.60	2.686
9.42	-7.64	3.819

5 distancias son menores a 1.39, es decir, 33 % de los datos caen dentro del contorno del 50 %. Esta diferencia en proporciones podría proveer evidencia para rechazar el supuesto de normalidad bivariada.

Sin embargo, el tamaño de la muestra (15) es muy pequeña para llegar a esta conclusión.

# Diagrama Chi cuadrado

- Existe un método más formal para evaluar la normalidad conjunta para un conjunto de datos, basado en la distancia cuadrada generalizada

$$d_j^2 = (\mathbf{x}_j - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}}), j = 1, 2, \dots, n$$

donde  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  son muestras de observaciones.

- Cuando la población raíz es normal multivariada y  $n$  y  $n-p$  son mas grandes que 25 o 30, cada una de las distancias cuadradas  $d_1^2, d_2^2, \dots, d_n^2$  deben comportarse como una chi-cuadrada.
- La idea es graficar los cuantiles de las distancias de la muestra vs los cuantiles esperados de una distribución chi cuadrada.
- Si el diagrama se parece a una linea recta de 45 grados que pasa por el origen, esto sugiere que los datos son normales.



# Pasos para construir un diagrama Chi-cuadrado:

- 1 Se ordenan las distancias cuadradas de la mas pequeña a la mas grande como  $d_{(1)}^2 \leq d_{(2)}^2 \leq \dots \leq d_{(n)}^2$
- 2 Se encuentran los cuantiles correspondientes,  $q_{c,p}(\frac{j-0.5}{n})$ , de la distribución chi cuadrada con  $p$  grados de libertad.
- 3 Se grafican los pares de cuantiles  $(q_{c,p}(\frac{j-0.5}{n}), d_j^2)$ . Si los puntos están sobre o cerca de una línea recta de 45 grados, esto apoya la suposición de que las observaciones son elegidas de una población normal.

# Ejemplo: construcción del diagrama chi-cuadrado para la muestra de datos bivariados

- Para nuestros datos bivariados tenemos

$$\bar{\mathbf{x}} = \begin{bmatrix} 5.26 \\ -3.09 \end{bmatrix}$$
$$\mathbf{S} = \begin{bmatrix} 7.12 & -0.67 \\ -0.67 & 12.43 \end{bmatrix} \rightarrow \mathbf{S}^{-1} = \begin{bmatrix} 0.1411 & 0.0076 \\ 0.0076 & 0.0808 \end{bmatrix}$$

# Ejemplo: construcción del diagrama chi-cuadrado para la muestra de datos bivariados

- Las distancias cuadradas generalizadas de cada punto al centroide  $\bar{x}$  están dadas en la tabla de la izquierda

$x_{j1}$	$x_{j2}$	$d_j^2$
1.43	-0.69	2.400
1.62	-5.00	2.279
2.46	-1.13	1.336
2.48	-5.20	1.548
2.97	-6.39	1.739
4.03	2.87	2.976
4.47	-7.88	2.005
5.76	-3.97	0.090
6.61	2.32	2.737
6.68	-3.24	0.281
6.79	-3.56	0.333
7.46	1.61	2.622
7.88	-1.87	1.138
8.92	-6.60	2.686
9.42	-7.64	3.819

$x_{j1}$	$x_{j2}$	$d_j^2$
5.76	-3.97	0.090
6.68	-3.24	0.281
6.79	-3.56	0.333
7.88	-1.87	1.138
2.46	-1.13	1.336
2.48	-5.20	1.548
2.97	-6.39	1.739
4.47	-7.88	2.005
1.62	-5.00	2.279
1.43	-0.69	2.400
7.46	1.61	2.622
8.92	-6.60	2.686
6.61	2.32	2.737
4.03	2.87	2.976
9.42	-7.64	3.819

Ordenamos  
las distancias  
cuadradas  
generalizadas  
(de menor a  
mayor)

# Ejemplo: construcción del diagrama chi-cuadrado para la muestra de datos bivariados

Entonces se encuentran los correspondientes percentiles

$100 \left( \frac{j-0.5}{n} \right)^{th}$  de la distribución Chi-cuadrada con  $p$  grados de libertad

$x_{j1}$	$x_{j2}$	$d_j^2$	$(j-0.5)/n$	$q_{c,2}[(j-0.5)/n]$
5.76	-3.97	0.090	0.033	0.068
6.68	-3.24	0.281	0.100	0.211
6.79	-3.56	0.333	0.167	0.365
7.88	-1.87	1.138	0.233	0.531
2.46	-1.13	1.336	0.300	0.713
2.48	-5.20	1.548	0.367	0.914
2.97	-6.39	1.739	0.433	1.136
4.47	-7.88	2.005	0.500	1.386
1.62	-5.00	2.279	0.567	1.672
1.43	-0.69	2.400	0.633	2.007
7.46	1.61	2.622	0.700	2.408
8.92	-6.60	2.686	0.767	2.911
6.61	2.32	2.737	0.833	3.584
4.03	2.87	2.976	0.900	4.605
9.42	-7.64	3.819	0.967	6.802

*Construimos ahora el scatter plot de los pares*

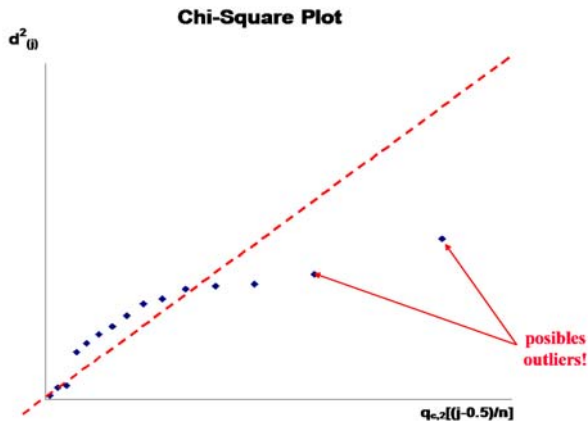
$$\left( q_{c,p}\left(\frac{j-0.5}{n}\right), d_j^2 \right)$$

*Si estos puntos están sobre una línea*

*recta, los datos apoyan la suposición*

# Ejemplo: construcción del diagrama chi-cuadrado para la muestra de datos bivariado

La gráfica no parece apoyar la suposición de que las observaciones fueron extraídos de una población normal bivariada



# Diagrama chi-cuadrado

- Este procedimiento no está limitado al caso bivariado; puede usarse para evaluar normalidad en dimensiones mayores a 2 ( $p > 2$ ).
- Algunos investigadores también sugieren calcular la correlación entre  $d_{(j)}$  y  $q_{c,p}[(j - .5)/n]$  y usar la prueba de Looney & Gullledge para evaluar la rectitud del diagrama chi-cuadrado.
- Para el ejemplo la correlación es  $r_Q = 0.8952$ .
- De la tabla para la prueba de normalidad basada en el coeficiente de correlación, y considerando un tamaño de muestra  $n = 15$ , los puntos críticos son 0.9503 en  $\alpha = 0.10$ , 0.9389 en  $\alpha = 0.05$ , y 0.9126 en  $\alpha = 0.01$ .
- Por tanto rechazamos la hipótesis de normalidad en cualquier  $\alpha$  mas grande que 0.01

# Resumen de las técnicas para evaluar la suposición de normalidad multivariada.

Se basan principalmente en el cálculo de

$d_j^2 = (\mathbf{x}_j - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})$ ,  $j = 1, 2, \dots, n$ , y éstas se comparan con los cuantiles de la  $\chi^2$ . La normalidad  $p$ -variada es adecuada si

- 1 Aproximadamente la mitad de las  $d_j^2$  son menores o iguales al cuantil del 50%,  $q_{c,p}(0.5)$ , de la distribución chi-cuadrada con  $p$  grados de libertad
- 2 O bien, si un diagrama de las distancias ordenadas

$$d_{(1)}^2 \leq d_{(2)}^2 \leq \dots \leq d_{(n)}^2 \quad \text{vs} \quad q_{c,p}\left(\frac{1-0.5}{n}\right), q_{c,p}\left(\frac{2-0.5}{n}\right), \dots, q_{c,p}\left(\frac{n-0.5}{n}\right)$$

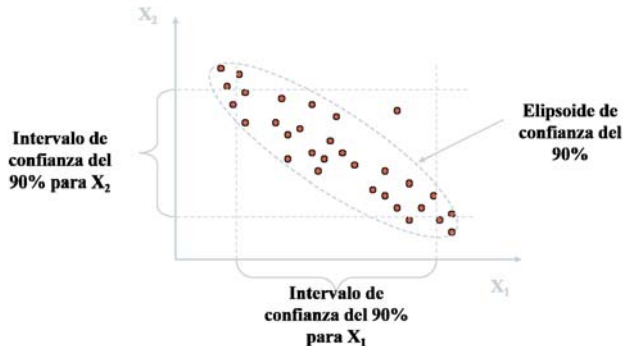
respectivos, se aproxima a una línea recta de 45 grados que pasa por el origen.





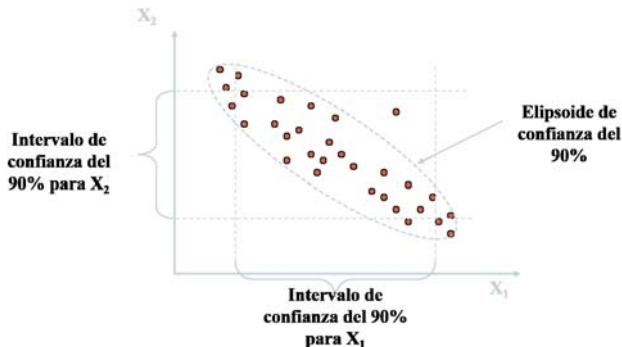
# Detección de Outlier

El punto está alejado del patrón elíptico que muestran los demás datos. Sin embargo de forma separada, cada uno de sus componentes tiene un comportamiento típico. Cada componente cae dentro de un intervalo de confianza del 90 %.



# Detección de Outliers

- Por tanto, este outlier bivariado no puede ser detectado al inspeccionar gráficamente las distribuciones marginales de  $x_1, x_2$ .
- En general, en altas dimensiones pueden existir outliers que no pueden detectarse mediante diagramas univariados e incluso de los scatter plots.



# Una estrategia para la detección de outlier multivariados:

- Buscar los outliers univariados
  - Se calculan los valores estandarizados  $z_{jk} = (x_{jk} - \bar{x}_k) / \sqrt{s_{kk}}$  para  $j = 1, \dots, n$ , y cada columna  $k = 1, \dots, p$ . Se examinan estos valores para grandes o pequeños valores de  $n$  y  $p$
  - Dot plots, histogramas, diagramas de tallo y hoja, para cada variable
  - Q-Q Plots
- Buscar los outliers bivariados
  - Distancias cuadradas generalizadas de cada par de variables a su centroide, su gráfica podría identificar valores outliers
  - Scatter plots para cada par de variables (matriz de scatter plots)
  - Diagramas Chi-Cuadrado para cada par de variables
- Buscar los outliers p-dimensionales
  - Distancias cuadradas generalizadas
  - Diagramas Chi- cuadrados

Observe que NINGUNA ESTRATEGIA garantiza la detección de outliers!

# Ejemplo de detección de outliers

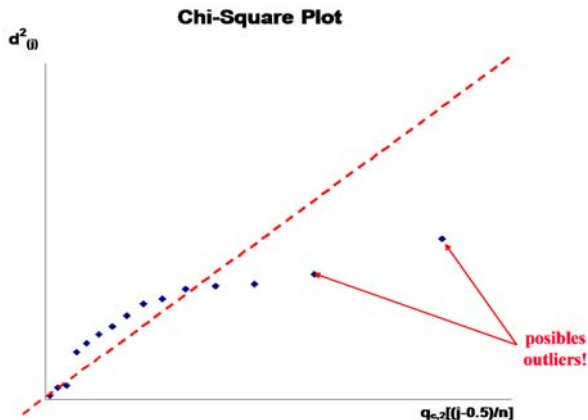
Aquí están calculados los valores estandarizados  $z_{jk}$  y las distancias cuadradas generalizadas  $d_j^2$  para los datos anteriores:

$x_{j1}$	$z_{j1}$	$x_{j2}$	$z_{j2}$	$d_j^2$
5.76	0.185	-3.97	-0.250	0.090
6.68	0.530	-3.24	-0.043	0.281
6.79	0.570	-3.56	-0.131	0.333
7.88	0.981	-1.87	0.347	1.138
2.46	-1.050	-1.13	0.556	1.336
2.48	-1.045	-5.20	-0.599	1.548
2.97	-0.860	-6.39	-0.936	1.739
4.47	-0.299	-7.88	-1.359	2.005
1.62	-1.367	-5.00	-0.541	2.279
1.43	-1.436	-0.69	0.681	2.400
7.46	0.822	1.61	1.333	2.622
8.92	1.371	-6.60	-0.994	2.686
6.61	0.504	2.32	1.536	2.737
4.03	-0.461	2.87	1.691	2.976
9.42	1.556	-7.64	-1.291	3.819

Esto se ve un poco inusual en  
el espacio de  $p = 2$

# Ejemplo de detección de outliers

Del diagrama Chi-cuadrada se comprueba que este dato está muy alejado de los restantes y del origen, por tanto podría ser un dato outlier



# Comentarios finales sobre los outliers

- Si se identifican outliers en un conjunto de datos multivariados, se debe verificar primero su contenido o veracidad.
- Dependiendo de la naturaleza de los outliers y los objetivos de la investigación, éstos pueden eliminarse o ponderarse de una forma apropiada en análisis posteriores.
- Aún cuando muchas técnicas estadísticas asumen poblaciones normales, aquellas que se basan en los vectores de medias muestrales usualmente no serán afectados por un número pequeño de outliers.

# Transformaciones para aproximar a normalidad

Que pasa cuando la suposición de normalidad no se cumple en un conjunto de datos?

En este caso existen dos alternativas:

- 1 Ignorar la comprobación de normalidad y proceder como si los datos fueran normalmente distribuidos. Esta práctica no es recomendable, ya que en muchas situaciones esto podría conducir a conclusiones erróneas.
- 2 Convertir los datos no normales a datos que sean aproximadamente normales, mediante transformaciones. Entonces el análisis de la teoría normal se puede realizar sobre los datos transformados.

# Transformaciones para aproximar a normalidad

- Una transformación es una reexpresión de los datos en diferentes unidades
- Por ejemplo, cuando un histograma de observaciones presenta una cola larga a la derecha, al transformar las observaciones aplicando logaritmos o raíces cuadradas, frecuentemente se mejora la simetría alrededor de la media y por tanto la aproximación a una distribución normal
- En muchas situaciones, las transformaciones de los datos, proveen expresiones mas naturales de las características que se están estudiando



# Transformaciones para aproximar a normalidad

- Transformaciones para convertir datos no normales a datos aproximadamente normales se sugieren usualmente por:
  - 1 Consideraciones teóricas
  - 2 Los datos brutos así lo requieren
- Transformaciones útiles para aproximar a la normalidad

Escala Original	Escala transformada
Conteos, $y$	$\sqrt{y}$
Proporciones, $\hat{p}$	$\text{logit}(\hat{p}) = \frac{1}{2} \log \left( \frac{\hat{p}}{1-\hat{p}} \right)$
Correlaciones, $r$	Transformación $z$ de Fisher, $z(r) = \frac{1}{2} \log \left( \frac{1+r}{1-r} \right)$

# Transformaciones de potencia

- En muchas situaciones, la elección de la transformación no es obvia. En este caso, es mejor dejar que los datos sugieran una transformación adecuada.
- Una familia de transformaciones útiles para este propósito son las transformaciones de potencia.
- Para seleccionar una transformación de potencia, el investigador debe analizar en la marginal los diagramas de puntos o los histogramas y decidir si valores grandes deben ser considerados o descartados para mejorar la simetría alrededor de la media.
- Es común realizar cálculos de prueba y error con algunos valores de las potencias en las transformaciones, hasta obtener una mejor aproximación a la normal.
- La transformación final elegida para los datos debe ser examinada mediante Q-Q plots o cualquier otro criterio, para ver si la suposición de normalidad se cumple.

# Transformaciones de potencia para variables continuas

- Box y Cox [1964] sugieren un método para encontrar una transformación apropiada a partir de la familia de transformaciones de potencia dada por

$$x^{(\lambda)} = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(x) & \lambda = 0 \end{cases}$$

La cual es continua en  $\lambda$  para  $x > 0$ .

- Las transformaciones de potencias están definidas únicamente para variables positivas, aunque esto no es muy restrictivo como parece.
- Se le puede sumar una constante  $m$  a cada observación en el conjunto de datos, de tal forma que si algunos de los valores son negativos, entonces  $x + m > 0$
- Cuando  $\lambda > 1$ , la transformación produce una mayor separación o dispersión de los valores grandes de  $x$ .
- Cuando  $\lambda < 1$ , el efecto es contrario, valores grandes de  $x$  tienden a concentrarse y los valores pequeños a dispersarse

# Transformaciones de potencia para variables continuas

Dada las observaciones  $x_1, \dots, x_n$ , la solución de Box-Cox para elegir una potencia apropiada  $\lambda$ , es la solución que maximiza la expresión

$$t(\lambda) = -\frac{n}{2} \ln \left[ \frac{1}{n} \sum_{j=1}^n (x_j^{(\lambda)} - \bar{x}^{(\lambda)})^2 \right] + (\lambda - 1) \sum_{j=1}^n \ln(x_j)$$

donde

$$x_j^{(\lambda)} = \begin{cases} \frac{x_j^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(x_j) & \lambda = 0 \end{cases}$$

$$\bar{x}^{(\lambda)} = \frac{1}{n} \sum_{j=1}^n x_j^{(\lambda)} = \frac{1}{n} \sum_{j=1}^n \left( \frac{x_j^\lambda - 1}{\lambda} \right)$$

# Transformación de Box-Cox

- $\lambda$  es el parámetro de la transformación que se estima a partir de los datos
- El primer término de  $\iota(\lambda)$ , aparte de una constante, es el logaritmo de una función de verosimilitud normal, después de maximizarla con respecto a los parámetros poblacionales  $\mu$  y  $\sigma$ .
- Al encontrar un valor de  $\lambda$  que maximice la función de verosimilitud normal, estamos encontrando una transformación de los datos, que será mas aproximada a la distribución normal.



# Transformaciones de Box-Cox de observaciones multivariadas

- Cuando tenemos un conjunto de observaciones multivariadas  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ , se selecciona una transformación de potencia para cada variable. Sean  $\lambda_1, \dots, \lambda_p$  las transformaciones de potencias para las  $p$  variables. Cada  $\lambda_k$  seleccionada maximiza

$$t_k(\lambda) = -\frac{n}{2} \ln \left[ \frac{1}{n} \sum_{j=1}^n (x_{jk}^{(\lambda_k)} - \bar{x}_k^{(\lambda_k)})^2 \right] + (\lambda_k - 1) \sum_{j=1}^n \ln(x_{jk})$$

- Este procedimiento es equivalente transformar cada distribución marginal a una distribución aproximadamente normal.
- Aunque el hecho de que las marginales sean normales no es suficiente para asegurar que la distribución conjunta sea normal.
- Sin embargo en aplicaciones prácticas esto podría ser suficientemente para asegurar normalidad conjunta.

# Transformaciones Box-Cox de observaciones multivariadas

- Para obtener normalidad conjunta, podemos empezar con los valores de las transformaciones de las distribuciones marginales,  $\hat{\lambda}_1, \dots, \hat{\lambda}_p$  obtenidas previamente.
- Entonces mediante un proceso iterativo obtenemos un conjunto de valores  $\lambda = (\lambda_1, \dots, \lambda_p)$  el cual colectivamente maximiza

$$l_k(\lambda_1, \dots, \lambda_p) = -\frac{n}{2} \ln |S(\lambda)| + \sum_{i=1}^p \left[ (\lambda_i - 1) \sum_{j=1}^n \ln(x_{ji}) \right]$$

donde  $\mathbf{S}(\lambda)$  es la matriz de covarianza muestral

$$\mathbf{S}(\lambda) = \frac{1}{n} \sum_{j=1}^n (\mathbf{x}_j^{(\lambda)} - \bar{\mathbf{x}}^{(\lambda)})(\mathbf{x}_j^{(\lambda)} - \bar{\mathbf{x}}^{(\lambda)})', \quad \bar{\mathbf{x}}^{(\lambda)} = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j^{(\lambda)}$$



# Transformaciones de observaciones multivariadas

- La obtención del parámetro vectorial  $\lambda$ , mediante la maximización de  $\iota(\lambda_1, \dots, \lambda_p)$  es mas complicada y además no suele aportar mejoras importantes respecto a transformar inividualmente cada variable para que las marginales sean normales.
- Por ejemplo en el caso bivariado, transformar cada distribución marginal a una normal es más o menos equivalente a transformar directamente la distribución bivariada a una distribución normal bivariada.
- En general es más fácil seleccionar apropiadas transformaciones para las distribuciones marginales que para las distribuciones conjuntas.
- **Moraleja:** la transformación de potencias de Box-Cox sobre los vectores de observaciones, es muy complicado y es mejor transformar a normales cada uno de sus componentes