Displays for Statistics 5401/8401

Lecture 6

September 19, 2005

Christopher Bingham, Instructor

612-625-1024, kb@umn.edu
372 Ford Hall

Class Web Page

http://www.stat.umn.edu/~kb/classes/5401

Multivariate population mean vector

Suppose $\mathbf{X} = [X_1, X_2, \ldots X_p]'$ is a random vector ($X_1, \ldots, X_p$ are p jointly distributed random variables).

The ***population mean vector*** (true mean, expectation) of $\mathbf{X}$ is

$$\mu_X = E[\mathbf{X}] = [\mu_1,\mu_2,\ldots,\mu_p]',$$
$$\mu_j = E[X_j],\ j = 1,\ldots,p$$

μ_X has the same dimensions as $\mathbf{X}$ ($p \times 1$).

Note:

$\overline{\mathbf{x}}$ is the vector of univariate *sample* means

μ_X is the vector of the univariate *population* means.

The population mean or expectation of a random matrix $\mathbf{Y} = [y_{ij}]$ is the matrix

$$\mu_Y = [\mu_{y_{ij}}] = [\ E[y_{ij}]\ ].$$

It can be useful to equate a n by p *matrix*, say $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p]$, to the np by 1 vector

$$\mathrm{vec}(\mathbf{X}) \equiv \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \cdots \\ \mathbf{X}_p \end{bmatrix},\ np \times 1$$

Clearly

$$\mu_{\mathrm{vec}(\mathbf{X})} = \mathrm{vec}(\mu_{\mathbf{X}})$$

that is the matrix is "unravelled" or unrolled column by column. This is what `vector(x)` creates when x is a matrix.

Population variance matrix

Vocabulary

The ***population covariance*** between X_j and X_k, $j \neq k$, is

- $\sigma_{jk} \equiv E[(X_j-\mu_j)(X_k-\mu_k)]$.

Properties

- $\sigma_{jk} = \sigma_{kj}$ (symmetry)
- $\sigma_{jk} > 0 \Leftrightarrow$ positive association
- $\sigma_{jk} < 0 \Leftrightarrow$ negative association

The ***population variance*** is

- $\sigma_j^2 = \sigma_{jj} = E[(X_j-\mu_j)^2] \geq 0$
- $|\sigma_{jk}| \leq \sqrt{\{\sigma_{jj}\sigma_{kk}\}}$, $\sigma_{jk}^2 \leq \sigma_{jj}^2\sigma_{kk}^2$
- $-1 \leq \rho_{jk} \leq 1$, $\rho_{jk} = \sigma_{jk}/\sqrt{\{\sigma_{jj}\sigma_{kk}\}}$
  $= \mathrm{cov}[X_j,X_k]/\{SD[X_j]SD[X_k]\}$, $SD[X_j] = \sqrt{\sigma_{jj}}$

ρ_{jk} is the population (true) correlation between X_j and X_k.

The population ***covariance matrix*** or ***variance matrix*** of **X** = $[X_1,\ldots,X_p]'$ is the p×p matrix

$$V[\mathbf{x}] = \Sigma = \Sigma_x = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \ldots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} & \ldots & \sigma_{2p} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} & \ldots & \sigma_{3p} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \sigma_{1p} & \sigma_{2p} & \sigma_{3p} & \ldots & \sigma_{pp} \end{bmatrix}$$

- $\boldsymbol{\Sigma}$ is symmetric ($\boldsymbol{\Sigma}' = \boldsymbol{\Sigma}$)
- The *diagonal elements* σ_{jj} of $\boldsymbol{\Sigma}$ are *variances* (σ_j^2)
- The *off-diagonal* σ_{jk}, $j \neq k$ are *covariances*.
- When $\mathbf{a} = [a_1,a_2,\ldots,a_p]'$ is constant so $\mathbf{a}'\mathbf{x}$ is a linear combination, with variance $V[\mathbf{a}'\mathbf{x}] = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} \geq 0$. Hence $\boldsymbol{\Sigma}$ is positive semi-definite

I hope I don't need to say this:

It is important to distinguish these ***population*** variances and covariances from the ***sample*** variances and covariances

- $s_{jj} = \sum_{1\leq i\leq n}(X_{ij} - \overline{X}_j)^2/(n-1)$ is not the same as σ_{jj}
- $s_{jk} = \sum_{1\leq i\leq n}(X_{ij} - \overline{X}_j)(X_{ik} - \overline{X}_k)/(n-1)$ is not the same as σ_{jk}

and to distinguish the *population* variance matrix $\boldsymbol{\Sigma}$ from the *sample* variance matrix **S**.

You never test a null hypothesis about the value of $\overline{\mathbf{x}}$ or **S**.

To state that a test statistic tests the null hypothesis H_0: $\overline{\mathbf{x}} = \mathbf{0}$ is *nonsense*. Probably what is meant is H_0: $\boldsymbol{\mu} = \mathbf{0}$.

Population correlation matrix

The *population* (Pearson) *correlation* between X_j and X_k is

$$\rho_{jk} \equiv \sigma_{jk}/\sqrt{(\sigma_{jj}\sigma_{kk})} = \rho_{kj} = \mathrm{cov}[Z_j,Z_k].$$

Standardized X_j is $Z_j = (X_j-\mu_j)/\sqrt{\sigma_{jj}}$.

Note: $\rho_{jk} = 0 \Leftrightarrow \sigma_{jk} = 0$

The symmetric p by p matrix

$$\mathbf{R} = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \ldots & \rho_{1p} \\ \rho_{12} & 1 & \rho_{23} & \ldots & \rho_{2p} \\ \rho_{13} & \rho_{23} & 1 & \ldots & \rho_{3p} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \rho_{1p} & \rho_{2p} & \rho_{3p} & \ldots & 1 \end{bmatrix} = \mathbf{D}^{-1}\boldsymbol{\Sigma}\mathbf{D}^{-1}$$

is the *population correlation matrix*

- $\mathbf{D} = \mathrm{diag}[\sqrt{\sigma_{11}},\sqrt{\sigma_{22}},\ldots,\sqrt{\sigma_{pp}}]$
- $\mathbf{D}^{-1} = \mathrm{diag}[1/\sqrt{\sigma_{11}},1/\sqrt{\sigma_{22}},\ldots,1/\sqrt{\sigma_{pp}}]$.

When $\boldsymbol{\Sigma} = \mathrm{diag}[\sigma_{11},\sigma_{22},\ldots,\sigma_{pp}]$ is diagonal,

- $\rho_{ij} = 0$, $i \neq j$
- $\mathbf{R} = I_p = \mathrm{diag}[1, 1,\ldots, 1]$.

Univariate normal Distribution

X is $N(\mu,\sigma^2)$ means X is normal with mean μ and variance σ^2. Its density is

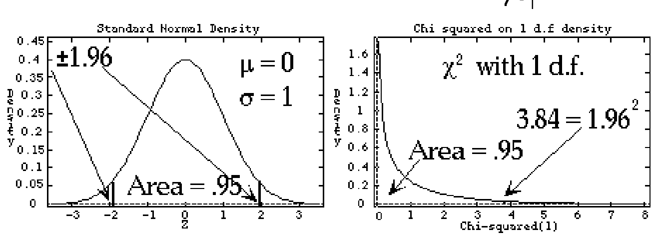
$$f(x,\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left[\frac{(x-\mu)^2}{\sigma^2}\right]} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

where $z = (x-\mu)/\sigma$ (*standardized* x).

Facts:

- z is <u>standard normal</u> $N(0,1^2)$
- $z^2 = \dfrac{(x-\mu)^2}{\sigma^2}$ is χ_1^2 (chi-squared, 1 DF)

Densities of Z and χ_1^2



The <u>total area</u> under each curve is 1.

The area under the χ_1^2 curve to the left of $3.84 = 1.96^2$ is

$$P(\chi_1^2 \leq 3.84) = P(|z| \leq 1.96) = .95$$

Multivariate normal distribution

Notation

$\mathbf{x} = [x_1, x_2, \ldots, x_p]'$ is $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$,

- $\boldsymbol{\mu}$ p×1 vector
- $\boldsymbol{\Sigma}$ p×p *positive definite* symmetric matrix.

Density of $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ random vector $\mathbf{x}$:

$$f(\mathbf{x},\Sigma,\mu) = \frac{e^{-\frac{1}{2}\left[(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right]}}{(2\pi)^{p/2}\sqrt{\det(\Sigma)}} = \frac{e^{-\frac{1}{2}Q(x-\mu,\Sigma)}}{(2\pi)^{p/2}\sqrt{\det(\Sigma)}}$$

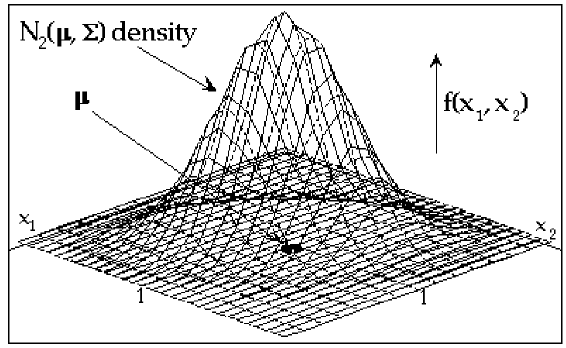
Multivariate Normal Facts

- $E[\mathbf{x}] = \boldsymbol{\mu}$ p parameters
- $V[\mathbf{x}] = \boldsymbol{\Sigma}$ p(p+1)/2 parameters, p variances, (p-1)/2 covariances
- $Q(\mathbf{x} - \boldsymbol{\mu},\boldsymbol{\Sigma}) \equiv (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$
  $= \text{trace}(\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})') \sim \chi_p^2$

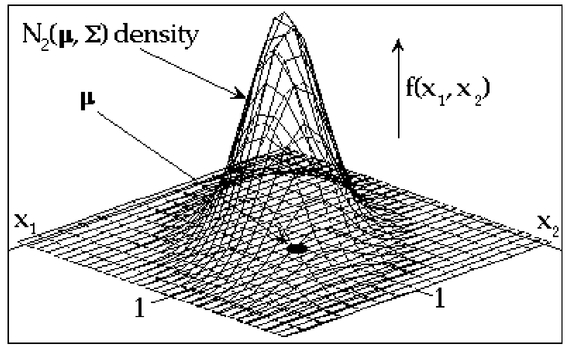
When $p = 1$ ($N(\mu, \sigma^2)$)

- $\boldsymbol{\mu} = \mu$, $\boldsymbol{\Sigma} = \sigma^2$
- $(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = ((x - \mu)/\sigma)^2 = \chi_1^2$

Shape of bivariate density



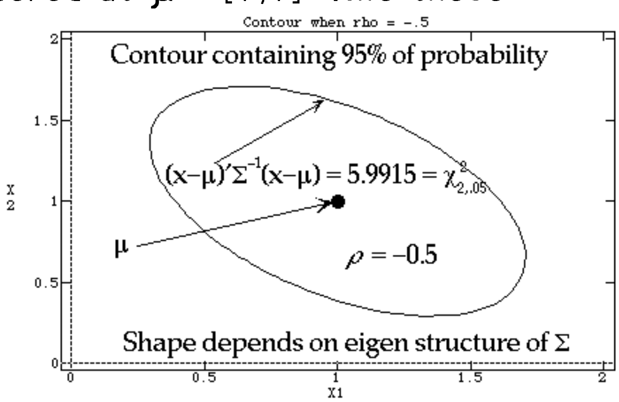
$\boldsymbol{\mu} = [1,1]'$
$\sqrt{\sigma_{11}} = .3$
$\sqrt{\sigma_{22}} = .3$,
$\rho_{12} = -0.5$
Centered at $\boldsymbol{\mu}$

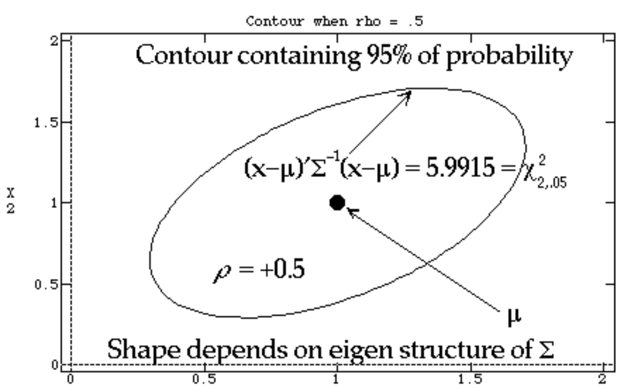


$\boldsymbol{\mu} = [1,1]'$
$\sqrt{\sigma_{11}} = .3$
$\sqrt{\sigma_{22}} = .3$,
$\rho_{12} = +0.5$
Centered at $\boldsymbol{\mu}$

- The mode (maximum) is at $\boldsymbol{\mu}$
- Every contour (level curve) is an ellipse centered at $\boldsymbol{\mu}$

All contours (level curves) are ellipses centered at $\boldsymbol{\mu} = [1,1]'$ like these:





Standard Multivariate Normal

When $\mathbf{z}$ is $N_p(\mathbf{0}, \mathbf{I}_p)$, its density is

$$f(\mathbf{z}) = \frac{e^{-\frac{1}{2}\mathbf{z}'\mathbf{z}}}{(2\pi)^{p/2}} = \frac{e^{-\frac{1}{2}\sum_{i=1}^{p} z_i^2}}{(2\pi)^{p/2}} = \frac{e^{-\frac{z_1^2}{2}}}{\sqrt{2\pi}}\frac{e^{-\frac{z_2^2}{2}}}{\sqrt{2\pi}}\cdots\frac{e^{-\frac{z_p^2}{2}}}{\sqrt{2\pi}},$$

That is

- $\mu_1 = \mu_2 = \ldots = \mu_p = 0$
- $\sigma_{11} = \sigma_{22} = \ldots = \sigma_{pp} = 1$
- $\sigma_{ij} = 0$, $i \neq j \Rightarrow \rho_{ij} = 0$
- The z_i's are independent $N_1(0, 1^2)$ because
  $$f(\mathbf{z}) = f(z_1)\times f(z_2)\times \ldots \times f(z_p)$$
- $Q(\mathbf{z} - \boldsymbol{\mu}_z,\boldsymbol{\Sigma}_z) = (\mathbf{z} - \mathbf{0})'\mathbf{I}_p^{-1}(\mathbf{z} - \mathbf{0}) = \sum_{1\le i\le p} z_i^2$ is distributed as χ_p^2 by a standard univariate result.

Properties of Multivariate Normal

The following properties are basically mathematical theorems. They are important in part because they provide a basis assessing or testing multivariate normality. If a sample of data appears not to satisfy one of these properties, it is evidence the sample is not from a multivariate normal population.

1. All *marginal* distributions are normal.

- *Each* x_i, <u>ignoring</u> other x's, is $N_1(\mu_i,\sigma_{ii})$
- *Any* subset of variables, <u>ignoring</u> other x's, is *multivariate* normal.

Application: If a univariate sample consisting of the values of x_j does not appear to come from a normal population, then the multivariate data is probably not multivariate normal.

Specifically, when $\mathbf{x}' = [\mathbf{x}_1', \mathbf{x}_2']$, where $\mathbf{x}_i$ is a p_1 by 1 vector, $\mathbf{x}_2$ is a p_2 by 1 vector than you can partition $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$:

$$\boldsymbol{\Sigma} = \begin{array}{c} \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21}=\boldsymbol{\Sigma}_{12}' & \boldsymbol{\Sigma}_{22} \end{bmatrix} \\ \begin{array}{cc} p_1 \qquad\qquad & p_2 \end{array} \end{array}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \begin{array}{c} p_1 \\ p_2 \end{array}$$

where

- $\boldsymbol{\mu}_1$ is $p_1\times 1$ and $\boldsymbol{\mu}_2$ is $p_2\times 1$
- $\boldsymbol{\Sigma}_{11}$ is $p_1\times p_1$ and $\boldsymbol{\Sigma}_{22}$ is $p_2\times p_2$
- $\boldsymbol{\Sigma}_{12}$ is $p_1\times p_2$ and $\boldsymbol{\Sigma}_{21} = \boldsymbol{\Sigma}_{12}'$ is $p_2\times p_1$

This property states

$$\mathbf{x}_1 \text{ is } N_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}),\ \mathbf{x}_2 \text{ is } N_{p_2}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22})$$

2. All *conditional* distributions are normal.

- Distribution of $p_2\times 1$ $\mathbf{x}_2$ given $p_1\times 1$ $\mathbf{x}_1$ is
$$N_{p_2}(\boldsymbol{\mu}_2 + \boldsymbol{\beta}_{2\cdot 1}'(\mathbf{x}_1 - \boldsymbol{\mu}_1), \boldsymbol{\Sigma}_{22\cdot 1}),$$

That is,

$E[\mathbf{x}_2|\mathbf{x}_1] = \boldsymbol{\mu}_2+\boldsymbol{\beta}_{2\cdot 1}'(\mathbf{x}_1-\boldsymbol{\mu}_1)$, p_2 by 1

$V[\mathbf{x}_2|\mathbf{x}_1] = \boldsymbol{\Sigma}_{22\cdot 1}$, p_2 by p_2

- This is a *linear* regression of $\mathbf{x}_2$ on $\mathbf{x}_1$

Application: if the dependence of one variable, say x_j, on another, say x_k, is not linear, then $\mathbf{X}$ is probably not multivariate normal.

- $\boldsymbol{\beta}_{2\cdot 1} = \boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $p_1 \times p_2$ is a *matrix* of *population regression coefficients*.

- $V[\mathbf{x}_2|\mathbf{x}_1] = \boldsymbol{\Sigma}_{22\cdot 1}$ does not depend on $\mathbf{x}_1$.
- $\boldsymbol{\Sigma}_{22\cdot 1} \equiv \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\beta}_{2\cdot 1}'\boldsymbol{\Sigma}_{11}\boldsymbol{\beta}_{2\cdot 1}$
$= V[\mathbf{x}_2 - E[\mathbf{x}_2|\mathbf{x}_1]]$
$= V[\mathbf{x}_2 - \boldsymbol{\mu}_2 - \boldsymbol{\beta}_{2\cdot 1}'(\mathbf{x}_1-\boldsymbol{\mu}_1)]$
That is,
$$\boldsymbol{\Sigma}_{22\cdot 1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$$
is the variance matrix of the
residuals $= \mathbf{x}_2 - E[\mathbf{x}_2|\mathbf{x}_1]$.

Application: If regression diagnostics in a linear regression of one variable on the others indicate non-constant variance, that indicates the data do not come from a multivariate normal population.

Bivariate case

When $p = 2$ and $p_1 = p_2 = 1$, these are

- $\beta_{2.1} = \sigma_{12}/\sigma_{11}$ (simple linear regression)
- $\sigma_{22.1} = \sigma_{22} - \sigma_{12}^2/\sigma_{11} = \sigma_{22} - \beta_{2.1}^2\sigma_{11}$
$= (1 - \rho_{12}^2)\sigma_{22}$

3. *Linear combinations are normal*

- $\mathbf{a}'\mathbf{x} = \sum_{1\leq i\leq p} a_i x_i$ is $N_1(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$
- $\mathbf{A}'\mathbf{x} = [\mathbf{a}_1'\mathbf{x},\ldots,\mathbf{a}_q'\mathbf{x}]'$ is $N_q(\mathbf{A}'\boldsymbol{\mu}, \mathbf{A}'\boldsymbol{\Sigma}\mathbf{A})$ when $\mathbf{A} = [\mathbf{a}_1,\ldots,\mathbf{a}_q]$ is p by q whose columns define linear combinations.

Example: If $\{d_i\}$, $d_i = x_{i2} - x_{i1}$, $i = 1,\ldots,n$ does not appear to be normal, then $\mathbf{x}$ is probably not multivariate normal.

4. The distribution of

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

is χ_p^2.

Except for a factor of 1/2, $Q(\mathbf{x}-\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the exponent in the density.

5. *Zero covariance or correlation implies independence*

When $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$,

that is, $\boldsymbol{\Sigma}_{12} = \mathrm{Cov}[\mathbf{x}_1, \mathbf{x}_2] = 0$, then $\mathbf{x}_1$ and $\mathbf{x}_2$ are *independent*.

In particular, when $\sigma_{ij} = 0$, x_i and x_j are independent

Since $\rho_{ij} = 0 \Leftrightarrow \sigma_{ij} = 0$, uncorrelated x_i and x_j are independent.

Standardization

When Y is a *univariate* random variable, an important re-expression of Y is as a *standardized random variable* (**z-score**)

$$Z = (Y - \mu_Y)/\sigma_Y$$

Z has mean 0 and standard deviation 1:

$$\mu_z = 0,\ \sigma_z = 1$$

Example:

When Y = <u>test statistic</u> or <u>estimator</u> (e.g. $\overline{x}$) with *hypothesized* mean $\mu_Y = \mu_0$, often 0, and *standard error* σ_Y.

Then $Z = (Y - \mu_0)/\sigma_Y$ is a Z-statistic for testing H_0: $\mu_Y = \mu_0$

You can often replace an <u>unknown</u> σ_Y by an *estimator* $\hat{\sigma}_Y$ and standardize to get the test statistic

$$t = (Y - \mu_0)/\hat{\sigma}_Y$$

- When Y is (approximately) $N(\mu_Y, \sigma_Y^2)$, Z is (approximately) $N(0,1^2)$ and $Z^2 = (Y - \mu_Y)^2/\sigma_Y^2$ or $t^2 = (Y - \mu_Y)^2/\hat{\sigma}_Y^2$ is (approximately) χ_1^2 (χ^2 on 1 d.f.).
- When Y is exactly $N(\mu_Y,\sigma_Y^2)$ and $\hat{\sigma}_Y^2$ is an <u>independent</u> estimate of σ_Y^2 such that $\hat{\sigma}_Y^2/\sigma_Y^2 = \chi_{f_e}^2/f_e$, then t is distributed as t_{f_e} = **Student's t** on f_e degrees of freedom and

  $t^2 = (Y - \mu_Y)^2/\hat{\sigma}_Y^2$ is distributed as F_{1,f_e}.

Note: $E[\chi_f^2] = f$, so $E[\chi_f^2/f] = 1$

Notation: I consistently use the *notation*

f_e = **error degrees of freedom**

Later I will use the notation

f_h = **hypothesis degrees of freedom**

These notations are not used in the text.

Familiar example

$Y = \overline{X}$ and you are testing H_0: $\mu = \mu_0$.

Then $\sigma_Y = \sigma_{\overline{X}} = \sigma_X/\sqrt{n}$ and $\hat{\sigma}_Y = \hat{\sigma}_{\overline{X}} = s/\sqrt{n}$ and

$$Z = (\overline{X} - \mu_0)/(\sigma/\sqrt{n})$$
$$t = (\overline{X} - \mu_0)/(s/\sqrt{n})$$

When $\overline{x}$ is computed from a random sample:

- In large samples, $Z \sim N(0, 1^2)$ and $Z^2 \sim \chi_1^{\,2}$
- When X is normal $t \sim t_{n-1}$ and $t^2 \sim F_{1,n-1}^{\,2}$

Multivariate standardization

A multivariate vector **Z** is *standardized* when

- $\boldsymbol{\mu}_Z = \mathbf{0}$
- $V[\mathbf{Z}] = \mathbf{I}_p = \text{diag}[1,1,\ldots,1]$.

When **Y** is an multivariate random vector, then for *any* p×p matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_p]$,

$$E[\mathbf{A}'(\mathbf{Y} - \boldsymbol{\mu}_Y)] = \mathbf{A}'E[\mathbf{Y} - \boldsymbol{\mu}_Y] = \mathbf{A}'\mathbf{0} = \mathbf{0}.$$

So it's easy to transform **Y** to a form with mean **0**.

It's harder to find a matrix **A** such that

$$\mathbf{Z} = \mathbf{A}'(\mathbf{Y} - \boldsymbol{\mu}_Y)$$

has variance matrix $\mathbf{I}_p$.

However, when you have such an **A**,

$$\mathbf{Z} = \mathbf{A}'(\mathbf{Y} - \boldsymbol{\mu}_Y)$$

is a standardized version of **Y**.

Or, if $\mathbf{C} = \mathbf{A}^{-1}$,

$$\mathbf{Z} = (\mathbf{C}')^{-1}(\mathbf{Y} - \boldsymbol{\mu}_Y)$$

is a standardized version of **Y**.

Matrix Square Roots

Vocabulary:

Let **B** be a *positive semi-definite* p×p *symmetric* matrix. Then, when the p×p matrix **C** satisfies $\mathbf{C}'\mathbf{C} = \mathbf{B}$, we say **C** is a ***matrix square root*** of **B**.

- **C** is not unique; you can choose **C** to be *symmetric*, *triangular* $\begin{bmatrix} x &&&&& \\ x & x &&&& \\ x & x & x &&& \\ x & x & x & x && \\ x & x & x & x & x & \\ x & x & x & x & x & x \end{bmatrix}$ or $\begin{bmatrix} x & x & x & x & x & x \\ & x & x & x & x & x \\ && x & x & x & x \\ &&& x & x & x \\ &&&& x & x \\ &&&&& x \end{bmatrix}$, or none of these.

```
Cmd> b # previously entered matrix
(1,1)          16          12
(2,1)          12          10

Cmd> upper <- cholesky(b); upper # Upper triangular
(1,1)           4           3
(2,1)           0           1

Cmd> upper' %*% upper # Check:  upper' upper = b
(1,1)          16          12
(2,1)          12          10

Cmd> matsqrt(b) # does same as cholesky()
(1,1)           4           3
(2,1)           0           1

Cmd> lower <- matsqrt(b,lower:T); lower # lower triangular sqrt
(1,1)      1.2649           0      4/sqrt(10)            0
(2,1)      3.7947      3.1623     12/sqrt(10)  10/sqrt(10)

Cmd> lower' %*% lower # Check:  lower' lower = b
(1,1)          16          12
(2,1)          12          10
```

```
Cmd> sym <- matsqrt(b,symmetric:T); sym #symmetric sqrt
(1,1)        3.43       2.058
(2,1)       2.058       2.401

Cmd> sym' %*% sym # or sym %*% sym because sym is symmetric
(1,1)          16          12
(2,1)          12          10
```

You can get still other square roots by swapping the rows:

```
Cmd> asym <- sym[vector(2,1),]; asym #not symmetric, triangular
(1,1)       2.058       2.401
(2,1)        3.43       2.058

Cmd> asym' %*% asym
(1,1)          16          12
(2,1)          12          10
```

Vocabulary

When **C** is *upper triangular*, **C'C** = **B** is the ***Cholesky Decomposition*** of **B**.

MacAnova

When **b** is a symmetric matrix,

```
Cmd> c <- cholesky(b)
```

computes the upper triangular square root of **b**.