Displays for Statistics 5401/8401

Lecture 10

September 28, 2005

Christopher Bingham, Instructor

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

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

Ellipsoids

When $\mathbf{Q} = [q_{ij}]$ is p×p positive definite with inverse $\mathbf{Q}^{-1} = [q^{jk}]$, then

$$\sum_{1\le i\le p}\sum_{1\le j\le p} q^{ij}(x_i - x_{i0})(x_j - x_{j0}) = (\mathbf{x}-\mathbf{x}_0)'\mathbf{Q}^{-1}(\mathbf{x}-\mathbf{x}_0) = K^2$$

defines a p-dimensional ***ellipsoid*** with center at $\mathbf{x}_0$ (ellipse when p = 2).

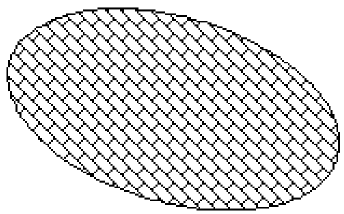
The surface or boundary of the ellipsoid consists of all $\mathbf{x}$ such that this equation is satisfied:

$$\{\mathbf{x} \mid (\mathbf{x}-\mathbf{x}_0)'\mathbf{Q}^{-1}(\mathbf{x}-\mathbf{x}_0) = K^2\}.$$

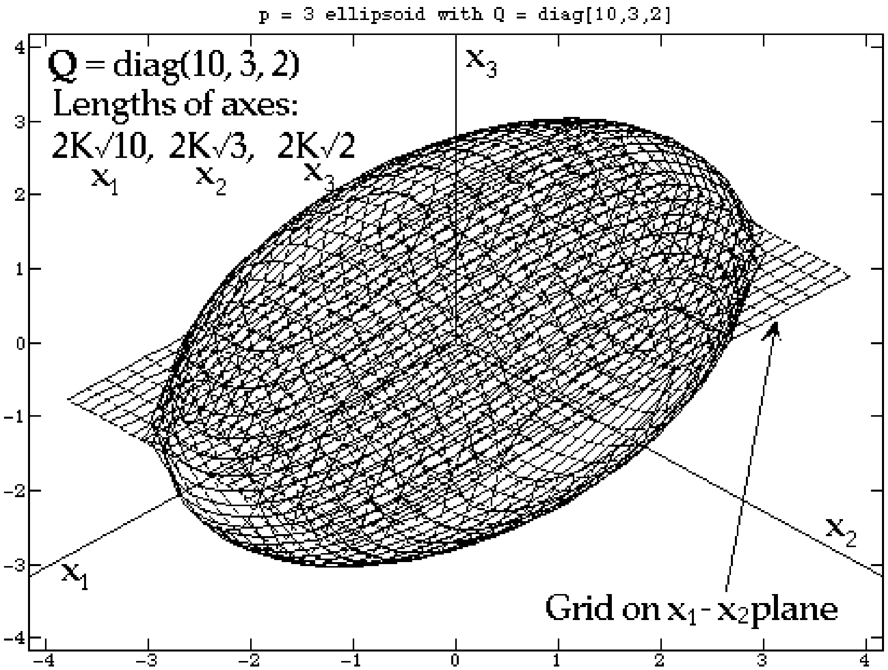
The surface together with the interior of the ellipsoid consists is

$$\{\mathbf{x} \mid (\mathbf{x}-\mathbf{x}_0)'\mathbf{Q}^{-1}(\mathbf{x}-\mathbf{x}_0) \le K^2\}$$

Ellipse together with its interior



Here is an ellipsoid with $p = 3$, centered at $\mathbf{x}_0 = [0, 0, 0,]'$ with $\mathbf{Q} = \text{diag}(10, 3, 2)$.



- When $\mathbf{Q}$ is <u>diagonal</u> (as here), $q_{ii} > 0$ and the equation for the surface is

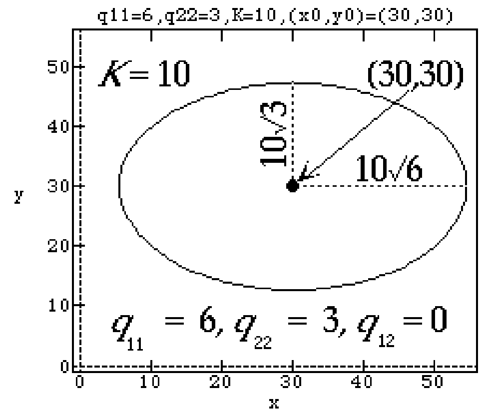
$$(\mathbf{x} - \mathbf{x}_0)'\mathbf{Q}^{-1}(\mathbf{x} - \mathbf{x}_0) = \sum_{1 \leq i \leq p} (x_i - x_{0i})^2/q_{ii} = K^2$$

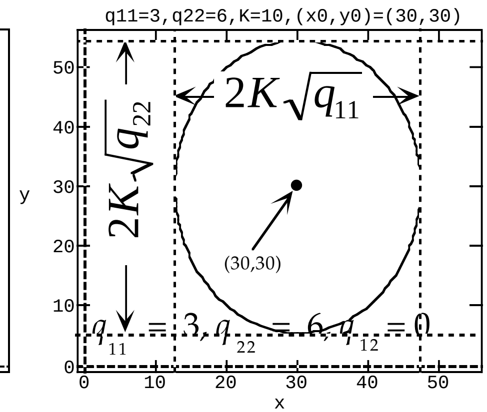
For $p = 2$, this is

$$(x_1 - x_{01})^2/q_{11} + (x_2 - x_{02})^2/q_{22} = K^2$$

Statistics 5401 Lecture 10 September 28, 2005

- Ellipsoids with <u>diagonal</u> $\mathbf{Q}$ have principal axes parallel the coordinate axes





- When $q_{11} = q_{22}$, $q_{12} = 0$ the ellipse is a circle with radius $K\sqrt{q_{11}}$ and diameter $2K\sqrt{q_{11}}$.
- When $q_{11} \neq q_{22}$, $q_{12} = 0$ lengths (diameters) in the x- and y- directions are $2K\sqrt{q_{11}}$ and $2K\sqrt{q_{22}}$.

A diagonal **Q**:

- Has <u>eigenvalues</u> $q_{11}, q_{22}, \ldots, q_{pp}$, rearranged in decreasing order. That is, if $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ are the eigenvalues, each $\lambda_j = q_{\ell\ell}$ for some ℓ.
- Has <u>eigenvectors</u> consisting of one 1 and p-1 0's like

$$\begin{bmatrix} 0 \\ 0 \\ \ldots \\ 1 \\ \ldots \\ 0 \end{bmatrix} \quad \text{A column of } \mathbf{I}_p$$

They are parallel the coordinate axes in p-dimensional space.

For p = 2, the eigenvectors are

$$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Statistics 5401 Lecture 10 September 28, 2005

When **Q** is ***not* diagonal**,
The *center* of the ellipsoid is at $\mathbf{x}_0$

Its *shape* is determined by the <u>eigenvalues</u> $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p$ of **Q**

Its *orientation* is determined by the <u>eigenvectors</u> $\mathbf{u}_1,\ldots,\mathbf{u}_p$, $\|\mathbf{u}_i\| = 1$ of **Q**.

Its <u>size</u> depends on $K\sqrt{\lambda_j}$, i = 1,...,p.

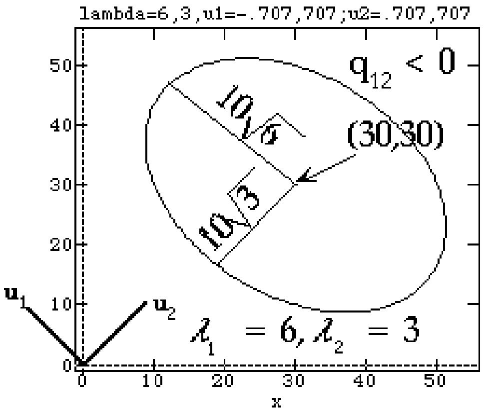
- The <u>longest axis</u> of the ellipsoid is parallel to $\mathbf{u}_1$ and has length $2\sqrt{\lambda_1}K$.
- The <u>2nd longest axis</u> perpendicular to $\mathbf{u}_1$ is parallel $\mathbf{u}_2$ and has length $2\sqrt{\lambda_2}K$.

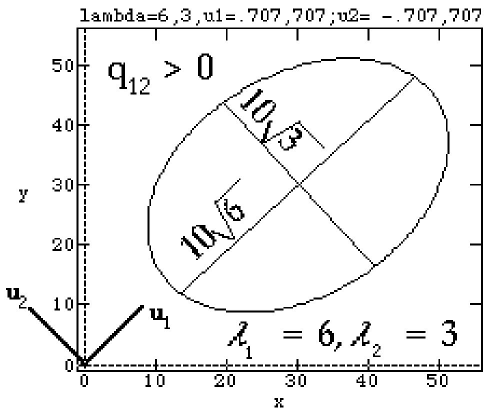
  . . .
- The <u>shortest axis</u> is parallel $\mathbf{u}_p$ and has length $2\sqrt{\lambda_p}K$. It is perpendicular to $\mathbf{u}_1,\ldots, \mathbf{u}_{p-1}$.

Reminder: $\mathbf{u}_1,\ldots, \mathbf{u}_p$ are *orthonormal*, that is, $\mathbf{u}_j'\mathbf{u}_k = 0$ ($\mathbf{u}_j$ <u>ortho</u>gonal to $\mathbf{u}_k$), $j \neq k$, and $\|\mathbf{u}_j\| = 1$ ($\mathbf{u}_j$ <u>normal</u>ized).

Two ellipses with same eigenvalues, and eigenvectors in different order

$\lambda_1 = 6$, $\lambda_2 = 3$, K = 10





$$Q_{left} = \begin{bmatrix} 4.5 & -1.5 \\ -1.5 & 4.5 \end{bmatrix} \quad Q_{right} = \begin{bmatrix} 4.5 & 1.5 \\ 1.5 & 4.5 \end{bmatrix}$$

Both **Q**'s have

$$\lambda_1 = 6,\ \lambda_2 = 3,\ \lambda_1\lambda_2 = 18,\ \lambda_1/\lambda_2 = 2.$$

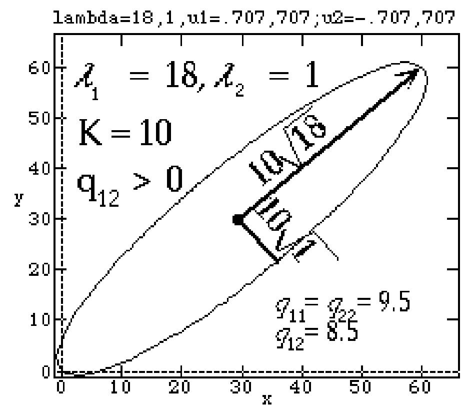
Eigenvectors

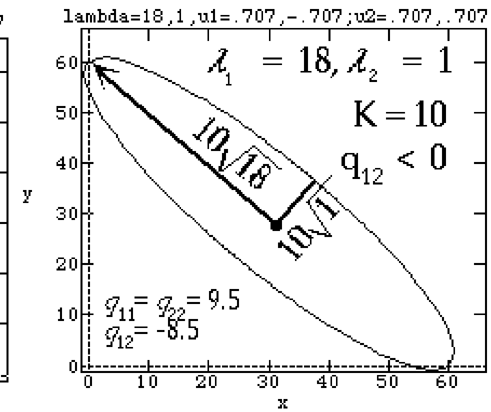
- Left Plot $\mathbf{u}_1$ $\mathbf{u}_2$
- Right Plot $\mathbf{u}_2$ $\mathbf{u}_1$

$$\begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \quad \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

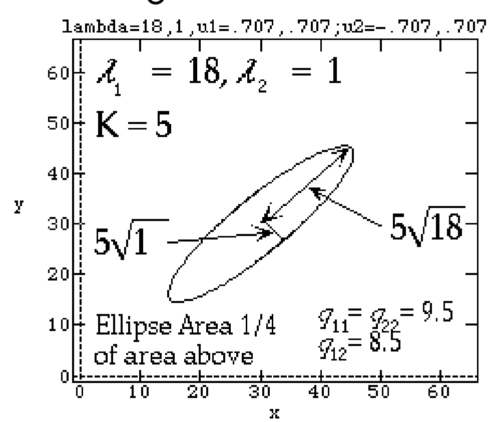
The larger λ_1/λ_2 the "thinner" the ellipse.

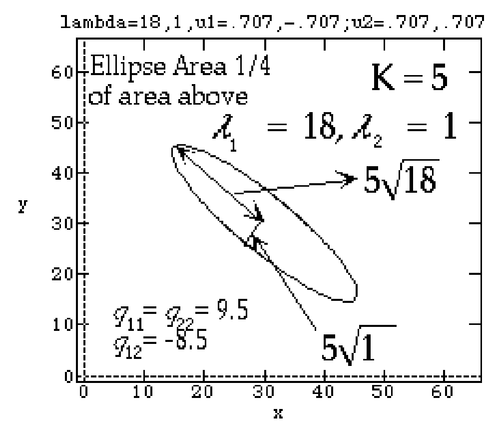
$\lambda_1 = 18$, $\lambda_2 = 1$, $\lambda_1\lambda_2 = 18$, $\lambda_1/\lambda_2 = 18$





Linear dimensions with K = 5 are half those with K = 10 but area is 1/4 (in higher dimensions volume would be $(1/2)^p$ as large.





Contours

Vocabulary: Let $h(\mathbf{x})$ be a function of a p-dimensional vector $\mathbf{x}$.

For a constant c, a *contour* of $h(\mathbf{x})$ is $\{\mathbf{x} \mid h(\mathbf{x}) = c\}$, the set of $\mathbf{x}$ with $h(\mathbf{x}) = c$.

When p = 2, a contour is a *level curve* on the surface whose height at $\mathbf{x}$ is $h(\mathbf{x})$.

A multivariate normal density is $h(\mathbf{x})$ =

$$f(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-(p/2)}\det(\boldsymbol{\Sigma})^{-1/2}e^{-(1/2)(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}$$

It has <u>ellipsoidal contours</u>:

Each *contour* is

$$\{\mathbf{x} \mid f(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = c > 0\}$$

which is equivalent to

$$\{\mathbf{x} \mid (\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = K^2\}$$

where

$$K^2 = -2\log(c) + \log(\det \boldsymbol{\Sigma})/2 + (p/2)\log\pi$$

Note: The maximum (mode) of the density is at $\mathbf{x} = \boldsymbol{\mu}$ so for the contour to exist, $c \leq f(\boldsymbol{\mu}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-(p/2)}\det(\boldsymbol{\Sigma})^{-1/2}$.

The contour

$$\{\mathbf{x} \mid (\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu}) = K^2\}$$

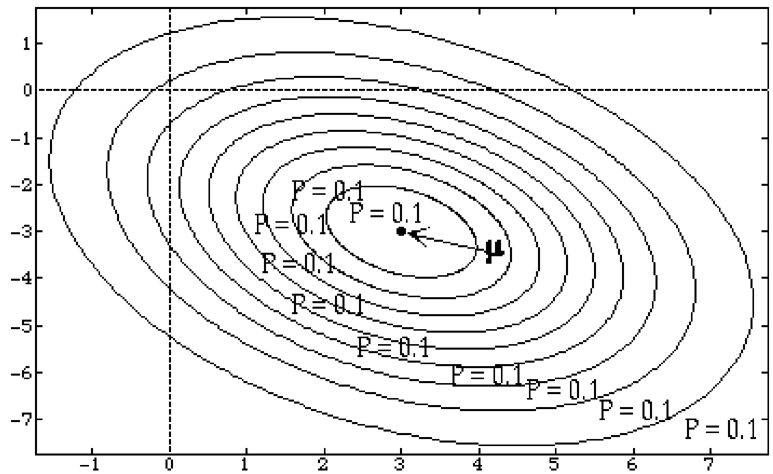
is an ellipsoid

- centered at $\mathbf{x}_0 = \boldsymbol{\mu}$
- with $\mathbf{Q} = \boldsymbol{\Sigma}$

For any fixed K, the larger the eigenvalues of $\boldsymbol{\Sigma}$ are

- the larger the ellipsoid (contour) is.
- the more scattered data will tend to be.

Contour plot of bivariate normal.



The contours plotted are the ellipses

$$\{\mathbf{x} \mid (\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \chi_2^{\,2}(i/10)\},$$

$i = 1, 2, \ldots, 9$, where $\{\chi_2^{\,2}(i/10)\}$ are $\chi_2^{\,2}$ probability points computed by

```
invchi(run(9)/10,2, upper:T).
```

Because $(\mathbf{x} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \chi_2^{\,2}$, there is probability $1/10 = 0.1$

- in the central ellipse
- between any two "adjacent" ellipses
- beyond the outer ellipse.

Univariate situations when a sum of squares has a χ^2 distribution

- $x_1,\ldots,x_n$ is $N_1(\mu, \sigma^2)$ random sample. Then $f_e s^2/\sigma^2 = \sum_i (x_i - \overline{x})^2/\sigma^2 = RSS/\sigma^2$ is exactly $\chi_{f_e}^{\,2}$, where $f_e = n-1$.
- Multiple regression or ANOVA model with independent $N(0,\sigma^2)$ errors ε_i with mean square error $s^2 = RSS/f_e$. Then $f_e s^2/\sigma^2 = RSS/\sigma^2$ is exactly $\chi_{f_e}^{\,2}$, where f_e = error degrees of freedom = $n - k$, where k = number of parameters including the intercept, if any.

You can express both these facts as

$$f_e s^2 = RSS = \sigma^2 \chi_{f_e}^{\,2}$$

Additional (monastic) fact:

s^2 is *independent* of $\overline{x}$ or of the estimated regression or ANOVA coefficients.

Wishart distribution

The Wishart distribution is a generalization of $\sigma^2\chi_f^2$ to *random matrices*.

A ***Wishart* random matrix W** is a random p×p positive definite symmetric matrix with a specific distribution $W_p(f;\mathbf{\Sigma})$.

- $W_p(f;\mathbf{\Sigma})$ depends on degrees of freedom f (as $\sigma^2\chi_f^2$ does).
- $W_p(f;\mathbf{\Sigma})$ depends on a positive definite variance matrix $\mathbf{\Sigma}$ ($\sigma^2 > 0$ for $\sigma^2\chi_f^2$).

When p = 1, $W_1(f;\sigma^2) = \sigma^2\chi_f^2$.

See Rao or Anderson for full details.

Facts

- $E[\mathbf{W}] = f\times\mathbf{\Sigma}$ so $E[(1/f)\mathbf{W}] = \mathbf{\Sigma}$
- When $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are a random sample from $N_p(\boldsymbol{\mu},\mathbf{\Sigma})$,

  $f_e\mathbf{S} = (n-1)\mathbf{S} = \sum_{1\leq i\leq n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})'$ is $W_p(f_e;\mathbf{\Sigma})$, $f_e \equiv n - 1$ (= univariate f_e)

  Thus $E[f_e\mathbf{S}] = f_e\mathbf{\Sigma} \Rightarrow E[\mathbf{S}] = \mathbf{\Sigma}$, so $\hat{\mathbf{\Sigma}} = \mathbf{S}$ is *unbiased* for $\mathbf{\Sigma}$

  $\overline{\mathbf{x}}$ is independent of $\mathbf{S}$
- In multivariate linear regression and MANOVA with errors (true residuals) $\boldsymbol{\varepsilon}_i$ that are independent $N_p(\mathbf{0},\mathbf{\Sigma})$, the p×p matrix of ***residual sums of squares and products***

  $\text{RCP} = \sum_i(\mathbf{x}_i - \hat{\mathbf{x}})(\mathbf{x}_i - \hat{\mathbf{x}})' = W_p(f_e;\mathbf{\Sigma})$

  where f_e = error degrees of freedom, (same as error df in univariate regression or ANOVA).

Example - One-way ANOVA and MANOVA

One-way *analysis of variance* is a way to analyze

- independent
- (univariate) normal
- random samples of sizes $n_1, n_2, \ldots, n_g$

from g groups or populations, all with variance σ^2.

A one-way ANOVA with g groups has

- **Error d.f.** $f_e = N - g$, $N = n_1 + \ldots + n_g$.
- **Error SS** = SSW =
  $$RSS = \sum_{1\le j\le g}\sum_{1\le i\le n_j}(x_{ij}-\overline{x}_{.j})^2 = \sigma^2\chi_{f_e}^2 = \sigma^2\chi_{N-g}^2$$
- **Hypothesis d.f.** $f_h = g - 1$
- **Among groups SS** = SS_{groups} =
  $$SSB = \sum_{1\le j\le g} n_j(\overline{x}_{.j}-\overline{x}_{..})^2$$
  When H_0: $\mu_1 = \mu = \ldots = \mu_g$ is true, SS_{groups} is $\sigma^2\chi_{f_h}^2 = \sigma^2\chi_{g-1}^2$.

You need normality for exactness.

Statistics 5401 Lecture 10 September 28, 2005

A one-way *multivariate ANOVA* (MANOVA) based on

- independent
- $N_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma})$, $j = 1, \ldots, g$
- random samples of sizes $n_1, n_2, \ldots, n_g$

from g groups or populations, all with variance matrix $\boldsymbol{\Sigma}$ has

- Error d.f. $f_e = N - g$, $N = n_1 + \ldots + n_g$
- $\hat{\mathbf{x}}_{ij} = \overline{\mathbf{x}}_{.j}$,
- **Error** SSCP matrix = $\mathbf{E}$ = RCP =
  $$\sum_{1\le j\le g}\sum_{1\le i\le n_j}(\mathbf{x}_{ij}-\overline{\mathbf{x}}_{.j})(\mathbf{x}_{ij}-\overline{\mathbf{x}}_{.j})' = W_p(f_e;\boldsymbol{\Sigma})$$
  $$= W_p(N - g;\boldsymbol{\Sigma})$$
- **Hypothesis d.f.** $f_h = g - 1$
- **Among groups** $SSCP_{groups}$ =
  $$\mathbf{H} = \sum_{1\le j\le g} n_j(\overline{\mathbf{x}}_{.j} - \overline{\mathbf{x}}_{..})(\overline{\mathbf{x}}_{.j} - \overline{\mathbf{x}}_{..})'$$
  When H_0: $\boldsymbol{\mu}_1 = \boldsymbol{\mu} = \ldots = \boldsymbol{\mu}_g$ is true, $\mathbf{H}$ is $W_p(f_h;\boldsymbol{\Sigma}) = W_p(g-1;\boldsymbol{\Sigma})$, independent of $\mathbf{E}$.

You need normality for exactness.

Here is situation which often occurs in MANOVA and multivariate regression with N_p errors.

You have an estimator $\hat{\boldsymbol{\theta}}$ of a vector $\boldsymbol{\theta}$ of parameters such that

- $\hat{\boldsymbol{\theta}}$ is $N_p(\boldsymbol{\theta}, V[\hat{\boldsymbol{\theta}}])$
- $V[\hat{\boldsymbol{\theta}}] = K\,\boldsymbol{\Sigma}$, where K is a <u>known</u> constant but $\boldsymbol{\Sigma}$ is unknown and must be estimated

  Example: for $N_p(\boldsymbol{\mu},\boldsymbol{\Sigma})$, $\boldsymbol{\theta} = \boldsymbol{\mu}$, $\hat{\boldsymbol{\theta}} = \overline{\mathbf{x}}$, $\overline{\mathbf{x}}$ is $N_p(\boldsymbol{\mu}, K\,\boldsymbol{\Sigma})$, with $K = 1/n$

- $\mathbf{S}$ is an unbiased estimate of $\boldsymbol{\Sigma}$, <u>independent</u> of $\hat{\boldsymbol{\theta}}$, with $f_e\mathbf{S} = W_p(f_e,\boldsymbol{\Sigma})$.

  Example: $\mathbf{S} = (1/f_e)\sum_{1\leq i\leq n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})'$
  $f_e = n - 1$ and $f_e\mathbf{S} = W_p(n-1,\boldsymbol{\Sigma})$

Statistics 5401 Lecture 10 September 28, 2005

Facts: Under these conditions (normal estimator, independent Wishart estimate of $\boldsymbol{\Sigma}$

- $\hat{V}[\hat{\boldsymbol{\theta}}] = K\mathbf{S}$ is an <u>unbiased</u> estimator of $V[\hat{\boldsymbol{\theta}}] = K\boldsymbol{\Sigma}$.
- $T^2 \equiv (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\{\hat{V}[\hat{\boldsymbol{\theta}}]\}^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$
  $= (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\{K\mathbf{S}\}^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$
  $= ((f_e p)/(f_e - p + 1))F_{p,f_e-p+1}$

  Equivalently
  $\{(f_e - p + 1)/(pf_e)\}T^2 = F_{p,f_e-p+1}$

Particular cases

- *Single Sample* Hotelling's T^2, based on a *random sample* $\mathbf{x}_1, \ldots, \mathbf{x}_n$ from a multivariate population with mean $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{\theta} = \boldsymbol{\mu},\ \hat{\boldsymbol{\theta}} = \overline{\mathbf{x}}$$
$$V[\overline{\mathbf{x}}] = (1/n)\boldsymbol{\Sigma} \text{ so } K = 1/n$$
$$\hat{V}[\overline{\mathbf{x}}] = (1/n)\mathbf{S}, \text{ with } f_e = n - 1$$
$$T^2 = (\overline{\mathbf{x}} - \boldsymbol{\mu})'\hat{V}[\overline{\mathbf{x}}]^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu})$$
$$= n(\overline{\mathbf{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu})$$

When $\mathbf{x}$ is $N_p(\boldsymbol{\mu}\ \boldsymbol{\Sigma})$,

$$T^2 = ((f_e p)/(f_e - p + 1))F_{p,f_e-p+1}$$
$$= \{(p(n-1))/(n-p)\}F_{p,n-pn}.$$

You use $T^2 = (\overline{\mathbf{x}} - \boldsymbol{\mu}_0)'\hat{V}[\overline{\mathbf{x}}]^{-1}(\overline{\mathbf{x}} - \boldsymbol{\mu}_0)$ as a test statistic to test H_0: $\boldsymbol{\mu} = \boldsymbol{\mu}_0$, computing P-value or critical value using the F-distribution.

Two sample comparison of means

Suppose you have two
- *independent*
- *random samples*

$$\mathbf{x}_{1,1},\mathbf{x}_{2,1},\ldots,\mathbf{x}_{n_1,1} \text{ and } \mathbf{x}_{1,2},\mathbf{x}_{2,2},\ldots,\mathbf{x}_{n_2,2}$$

from two populations

Population 1: mean = $E[\mathbf{x}_1] = \boldsymbol{\mu}_1$, $V[\mathbf{x}_1] = \boldsymbol{\Sigma}_1$

Population 2: mean = $E[\mathbf{x}_2] = \boldsymbol{\mu}_2$, $V[\mathbf{x}_2] = \boldsymbol{\Sigma}_2$

Suppose your interest is in $\boldsymbol{\theta} \equiv \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$.

Then
- $\hat{\boldsymbol{\theta}} \equiv \overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2$
- $V[\hat{\boldsymbol{\theta}}] = V[\overline{\mathbf{x}}_1] + V[\overline{\mathbf{x}}_2]$

$$= (1/n_1)\boldsymbol{\Sigma}_1 + (1/n_2)\boldsymbol{\Sigma}_2.$$

Unpooled two-sample T^2

- *Unpooled* estimate of $V[\hat{\boldsymbol{\theta}}]$ is
  $\hat{V}[\hat{\boldsymbol{\theta}}] = \hat{V}[\overline{\mathbf{x}_1}] + \hat{V}[\overline{\mathbf{x}_2}] = (1/n_1)\mathbf{S}_1 + (1/n_2)\mathbf{S}_2$
  where $\mathbf{S}_1$ and $\mathbf{S}_2$ are (unbiased) sample variance matrices.
  $\hat{V}[\hat{\boldsymbol{\theta}}]$ is an unbiased estimate of $V[\hat{\boldsymbol{\theta}}]$
- $T^2 = (\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2})'\hat{V}[\hat{\boldsymbol{\theta}}]^{-1}(\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2})$
  $= (\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2})'\left(n_1^{-1}\mathbf{S}_1 + n_2^{-1}\mathbf{S}_2\right)^{-1}(\overline{\mathbf{x}_1} - \overline{\mathbf{x}_2})$
  tests H_0: $\boldsymbol{\theta} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \mathbf{0}$
  When H_0: $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ is true and n_1 and n_2 are <u>large</u> , $T^2 \mathrel{\widetilde{=}} \chi_p^2$
- Even with normal $\mathbf{x}_1$ and $\mathbf{x}_2$, when $n_1 \neq n_2$, unpooled T^2 is *not* $((pf_e)/(f_e - p + 1))F(p, f_e - p + 1)$
- Unpooled $T^2 \neq$ "classical" <u>pooled</u> two-sample T^2 except when $n_1 = n_2$.