Displays for Statistics 5401/8401

Lecture 12

October 3, 2005

Christopher Bingham, Instructor

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

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

Bonferronized t vs Hotelling's T^2

Both prescribe rules describing when you reject H_0.

Two-sample case with H_0: $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$.

Hotelling's T^2 test rule is:

Reject H_0 when
$$T^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\hat{V}[\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2]^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \geq T^2(\alpha)$$

Bonferronized t test rule is:

Reject H_0 when any
$$|t_j| > t((\alpha/p)/2)$$
with $t_j = (\bar{x}_{j1} - \bar{x}_{j2})/\sqrt{\{\hat{v}_{jj}[\bar{x}_{j1} - \bar{x}_{j2}]\}}$

The null hypotheses are identical.

$\hat{v}_{jj}[\bar{x}_{j1} - \bar{x}_{j2}] = (1/n_1 + 1/n_2)s_{jj}$ is a diagonal element of $\hat{V}[\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2] = (1/n_1 + 1/n_2)\mathbf{S}_{pooled}$

Advantages of Bonferronized t

- Easy to compute, well understood
- When you reject H_0 (because at least one $|t_j|$ is large), you know *how* H_0 appears to be violated because you know which t-statistics are significant.
- *Weaker* assumptions: <u>univariate</u> normality and equality of σ_{jj}'s not ρ_{jk}'s.
- When you have no good reason to believe the variances are equal, you can use the "unpooled" t-statistics
$$t_j = (\bar{x}_{j1} - \bar{x}_{j2})/\sqrt{\{\hat{V}[\bar{x}_{j1}]+\hat{V}[\bar{x}_{j1}]\}}, \text{ with}$$
$$df \cong \frac{2\{\hat{V}[\bar{x}_{j1}] + \hat{V}[\bar{x}_{j2}]\}^2}{\hat{V}[\bar{x}_{j1}]^2/(n_1-1) + \hat{V}[\bar{x}_{j2}]^2/(n_2-1)}$$
$$\hat{V}[\bar{x}_{jk}] = s_{jj}^{(k)}/n_k,\ k = 1, 2$$
There is no easy "fix" for T^2 when you can't assume $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$.

Disadvantages of Bonferronized t

- Bonferronized t can give different result from T^2
- It can give different results when you replace the original p variables by p linear combinations $y_j = \mathbf{a}_j'\mathbf{x}$, j = 1,...,p.

That is $\mathbf{x} \rightarrow \mathbf{y} = \mathbf{A}\mathbf{x} = [\mathbf{a}_1'\mathbf{x}, \ldots, \mathbf{a}_p'\mathbf{x}]$ where $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_p]'$ is $p \times p$ and <u>invertible</u>.

Example: $\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 1 & \ldots & 1 \\ 1 & -1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & -2 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & 1 & 1 & \ldots & & -(p-1) \end{bmatrix}$

$y_1 = \mathbf{a}_1'\mathbf{x} = x_1 + x_2 + \ldots + x_p$,
$y_2 = \mathbf{a}_2'\mathbf{x} = x_1 - x_2$,
$y_3 = \mathbf{a}_3'\mathbf{x} = 2\{(x_1 + x_2)/2 - x_3\}$, ...
$y_p = \mathbf{a}_p'\mathbf{x} = (p-1)\{(x_1+\ldots+x_{p-1})/(p-1) - x_p\}$

This might be useful with repeated measures data.

Two sample case:

$\mu_{y_1} = A\mu_1$, $\mu_{y_2} = A\mu_2$,

$\mu_1 = A^{-1}\mu_{y_1}$, $\mu_2 = A^{-1}\mu_{y_2}$.

Thus H_0: $\mu_1 = \mu_2$ and H_0: $\mu_{y_1} = \mu_{y_2}$ are essentially the same.

For Hotelling's T^2

- (T^2 based on **x**) = (T^2 based on **y** = **Ax**).
- Conclusions about H_0: $\mu_1 = \mu_2$ will be the same whether you analyze **x** or **y**.

But for Bonferronized t

- The t's computed from $\{y_j = \mathbf{a}_j'\mathbf{x}\}$ will almost certainly not be the same as the t's computed from $\{x_j\}$.
- The conclusions from **y** can differ from the conclusions from **x**.

Power issues

Depending on

- Σ

and

- the actual mean vector or vectors,

T^2 may have higher or lower power than Bonferronized t.

- When some correlations ρ_{ij} are large, T^2 may be much more powerful (more likely to reject H_0) than Bonferronized t
- For some alternative hypothesis H_a, particularly when one element of μ or $\mu_1 - \mu_2$ is large and the others are small, Bonferronized t may have larger power than T^2.

Simultaneous estimation

Suppose you are estimating q unknown parameters $\theta_1,\ldots,\theta_q$, say q means, simultaneously, that is, *in parallel*.

Let $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_q]'$.

Example: 1 sample multivariate mean μ

- q = p = dimension of **x**
- $\boldsymbol{\theta} = \boldsymbol{\mu}$ with $\theta_j = \mu_j$, j = 1, ... p

A **point estimate** of $\boldsymbol{\theta}$ is a vector $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{X})$ computed from data **X** which is a "good (?) guess" for the value of $\boldsymbol{\theta}$.

In some cases the best *point estimate* of $\boldsymbol{\theta}$ is the vector of the best univariate estimates of each θ_i.

Example (continued):

Estimate $\boldsymbol{\mu}$ by $\hat{\boldsymbol{\mu}} = \overline{\mathbf{x}} = [\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_p]'$.

Here $\hat{\theta}_i = \hat{\mu}_i = \overline{x}_i$.

Even in this simple case, Stein showed that $\overline{\mathbf{x}}$ was not an "admissible" estimator with p > 2, in the sense that with respect to one criterion, you can always find a "better" estimator thatn $\overline{\mathbf{x}}$.

Nonetheless, we will ignore the problem.

A point estimate $\hat{\boldsymbol{\theta}}$ is not enough. Tiy aksi want to know which values of $\boldsymbol{\theta}$ are plausible in light of $\mathbf{X}$, that is which $(\theta_1, \ldots, \theta_q)$ are not inconsistent with the data.

For q = 1, you usually use *interval estimation*. In frequentist statistics this is a confidence interval (θ_L, θ_U), where $\theta_L = \theta_L(\mathbf{X})$ and $\theta_U = \theta_U(\mathbf{X})$ are random variables computed from data $\mathbf{X}$. The interval is random.

Generally (θ_L, θ_U) ***encloses*** $\hat{\theta}$, that is

$$\theta_L \leq \hat{\theta} \leq \theta_U.$$

Often $\theta_L = \hat{\theta} - K\hat{\sigma}_{\hat{\theta}}$, $\theta_U = \hat{\theta} + K\hat{\sigma}_{\hat{\theta}}$, K=t or z where K is a *critical value*. $K\hat{\sigma}_{\hat{\theta}}$ is the *margin of error*.

When estimating a scale parameters, $\theta_L = \hat{\theta}/K_1$, $\theta_H = \hat{\theta}/K_2$ with $K_1 > 1$ and $K_2 < 1$.

The **defining property** for $(\theta_L, \theta_U) = (\theta_L(\mathbf{X}), \theta_U(\mathbf{X}))$ to be a confidence interval with *confidence level* $1 - \alpha$ is

$P((\theta_L, \theta_U)$ encloses *true* $\theta) \geq 1 - \alpha$
where "encloses" means $\theta_L \leq \theta \leq \theta_U$

This has meaning because (θ_L, θ_U) is a *random interval*.

Often, "$\geq 1-\alpha$" is actually "$= 1-\alpha$".

When P(encloses) > $1 - \alpha$, the interval is *conservative*.

A confidence interval gives a reasonably clear idea of how far $\hat{\theta}$ *might* be from θ.

It also gives an idea as to how far θ might be from $\hat{\theta}$.

Note the switch of θ and $\hat{\theta}$.

With a *vector* $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_q]'$ of q parameters, you could work with each θ_i *separately*, and compute individual $1 - \alpha$ confidence intervals $(\theta_{iL}, \theta_{iU})$ for each θ_i:

$$\theta_{iL} = \theta_{iL}(\mathbf{X}),\ \theta_{iU} = \theta_{iU}(\mathbf{X}),\ i = 1,\ldots,q.$$

With q intervals, *any* of which might *not* contain the true θ_j, there is an *increased* probability that at least one of the intervals does *not* contain the true θ_j.

That means

P(*all* intervals contain their θ_i) =
$P(\theta_{1L} \leq \theta_1 \leq \theta_{1U}, \ldots, \theta_{qL} \leq \theta_q \leq \theta_{qU}) \approx 1 - q\alpha << 1 - \alpha$

so you would not have high *simultaneous* confidence.

This is the *multiple confidence interval problem*.

If the $\{\theta_{iL}\}$ and $\{\theta_{iU}\}$ are computed so that

$P(\theta_{1L} \leq \theta_1 \leq \theta_{1U}$ and ... and $\theta_{qL} \leq \theta_q \leq \theta_{qU}) \geq 1 - \alpha$

the intervals $(\theta_{iL}, \theta_{iU})$, i = 1,...,q are called *simultaneous confidence intervals* with confidence level $\geq 1 - \alpha$.

You can accomplish this is by *Bonferronizing* univariate confidence intervals, that is using $\alpha' = \alpha/q$ to compute each interval (individual confidence level $1 - \alpha/q$). Then

$1 - P(\theta_{1L} \leq \theta_1 \leq \theta_{1U}, \ldots, \theta_{qL} \leq \theta_q \leq \theta_{qU}) =$
$P((\theta_{iL}, \theta_{iU})$ does *not* contain θ_i for some i)
$\leq q\alpha' = \alpha$,

so

$$P(\theta_{1L} \leq \theta_1 \leq \theta_{1U}, \ldots, \theta_{qL} \leq \theta_q \leq \theta_{qU}) \geq 1 - \alpha$$

so the simultaneous confidence level is at least $1 - \alpha$.

Confidence regions

- Suppose $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_q]'$ is a vector of q unknown parameters.
- Let $R(\mathbf{X})$ be a q dimensional region (area, volume, etc.) that depends on a data matrix $\mathbf{X}$.

For any $\boldsymbol{\theta}$, when you know $\mathbf{X}$, you can determine whether or not $\boldsymbol{\theta}$ is in $R(\mathbf{X})$.

Since $\mathbf{X}$ is random
- $R(\mathbf{X})$ is random
- For each $\boldsymbol{\theta}$, whether $R(\mathbf{X})$ contains $\boldsymbol{\theta}$ is random event with some probability.

When p = 1, $R(\mathbf{X})$ is usually the interval:

$$R(\mathbf{X}) = \{\theta \mid \theta_L(\mathbf{X}) \leq \theta \leq \theta_U(\mathbf{X})\}$$

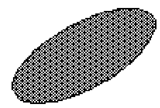
between *confidence limits* θ_L and θ_U.

In higher dimensions, $R(\mathbf{X})$ can have a variety of shapes, including (hyper) rectangular or ellipsoidal.

Definition

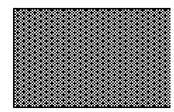
$R(\mathbf{X})$ is a $1 - \alpha$ *confidence region* for $\boldsymbol{\theta}$ if $P(R(\mathbf{X})$ contains $\boldsymbol{\theta}) \geq 1 - \alpha$, for all $\boldsymbol{\theta}$

When q > 1, useful confidence regions are usually
- *bounded*
- *connected*
- have *no holes*.



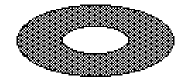
Ellipse

or



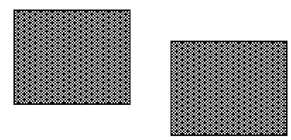
Box

but not



Has hole

and not

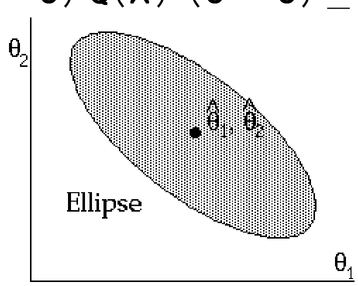


Disconnected

Shapes

- *Ellipses* (p = 2) or *ellipsoids* (p > 2)

$$\{\boldsymbol{\theta} \mid (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\mathbf{Q}(\mathbf{X})^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq K^2\}$$



$\mathbf{Q}(\mathbf{X})$ is an invertible matrix which depends on sample size usually on the data $\mathbf{X}$.

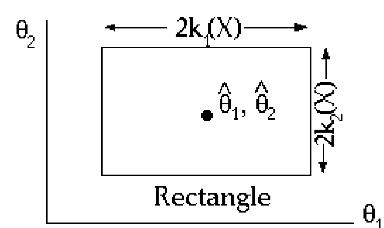
The region $\mathbf{R}(\mathbf{X})$ consists of the interior and boundary of the ellipse.

$\mathbf{Q}(\mathbf{X})$ will usually be $\hat{V}[\hat{\boldsymbol{\theta}}]$.

For example, $\mathbf{Q}(\mathbf{X}) = \mathbf{S}/n$ when $\boldsymbol{\theta} = \boldsymbol{\mu}$ and $\hat{\boldsymbol{\theta}} = \overline{\mathbf{x}}$.

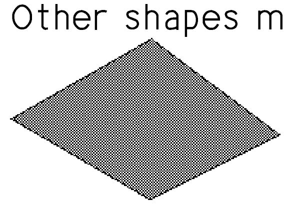
- "Boxes": rectangles when p = 2, rectangular boxes when p=3, or their generalization to p > 3:

$$\{\boldsymbol{\theta} \mid \ |\theta_i - \hat{\theta}_i| \leq k_i(\mathbf{X}),\ i = 1,\ldots,q\}$$

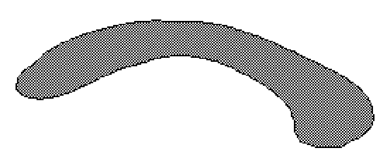


where $k_i(\mathbf{X})$ are numbers computed from the data and critical values, for instance $k_i(\mathbf{X}) = t_{n-1}((\alpha/p)/2)\sqrt{\{s_{ii}/n)\}}$

Other shapes might be



or



Shapes like the one on the right arise in non-linear problems.

Relationship of confidence regions and tests

Suppose for *every* value $\boldsymbol{\theta}^*$ of $\boldsymbol{\theta}$ there is a test statistic $T(\mathbf{X},\boldsymbol{\theta}^*)$ and critical value designed to test $H_0(\boldsymbol{\theta}^*)$: $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ at significance level α. Then

$$R(\mathbf{X}) = \{\boldsymbol{\theta} \mid H_0(\boldsymbol{\theta}) \text{ \underline{not rejected} by } T(\mathbf{X},\boldsymbol{\theta})\}$$

is a $1 - \alpha$ confidence region for θ.

Why?

The true value $\boldsymbol{\theta}$ is contained in $R(\mathbf{X})$ when and only when $H_0(\boldsymbol{\theta})$ is *not* rejected and this occurs with probability $1 - \alpha$.

Conversely, when $R(\mathbf{X})$ is a $1 - \alpha$ confidence region it defines a significance level α test of H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$:

"Reject H_0 if $\boldsymbol{\theta}_0$ is not in $R(\mathbf{X})$"

When H_0 is true, $P(\text{reject } H_0) \leq \alpha$

Shape of a confidence region based on Hotelling's T^2 test

Suppose a test of H_0: $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ is

Reject H_0 when $T^2 = T^2(\boldsymbol{\theta}_0) = T^2(\boldsymbol{\theta}_0,\mathbf{X}) > K_\alpha^{\,2}$ where $T^2(\boldsymbol{\theta}_0,\mathbf{X}) = (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)'\{\hat{V}[\hat{\boldsymbol{\theta}}]\}^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ with

$$K_\alpha^{\,2} = \chi_q^{\,2}(\alpha) \text{ (\underline{large} n)}$$

or

$$K_\alpha^{\,2} = \{(f_e q)/(f_e-q+1)\}F_{q,f_e-q+1}(\alpha) \text{ (\underline{small} n)}$$

$f_e = n-1$ or n_1+n_2-2, or whatever is appropriate.

When do you use small as opposed to large sample critical values?

In my opinion, you use the small sample critical value whenever f_e is defined and you can determine $F_{q,f_e-q+1}(\alpha)$.

That is, essentially always.

The confidence region $R(\mathbf{X})$ corresponding to T^2 consists of all $\boldsymbol{\theta}$ you would *not* reject ($T^2(\boldsymbol{\theta}) \leq K_\alpha^{\,2}$)

$$\begin{aligned} R(\mathbf{X}) &= \{\boldsymbol{\theta} \mid T^2(\boldsymbol{\theta}) \leq K_\alpha^{\,2}\} \\ &= \{\boldsymbol{\theta} \mid (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\{\hat{V}[\hat{\boldsymbol{\theta}}]\}^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \leq K_\alpha^{\,2}\} \\ &= \{\boldsymbol{\theta} \mid (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\{\hat{V}[\hat{\boldsymbol{\theta}}]\}^{-1}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) \leq K_\alpha^{\,2}\} \end{aligned}$$

Note the swap of $\boldsymbol{\theta}$ in $\hat{\boldsymbol{\theta}}$ in this last.

- $R(\mathbf{X})$ is an *ellipsoid*
- $R(\mathbf{X})$ is *centered* at $\hat{\boldsymbol{\theta}}$
- $R(\mathbf{X})$ has shape and orientation determined by the eigenvalues and vectors of $\hat{V}[\hat{\boldsymbol{\theta}}]$
- $R(\mathbf{X})$ has size proportional to K_α.

Note: In the formula for an ellipsoid

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

$\mathbf{x} = \boldsymbol{\theta}$, $\mathbf{x}_0 = \hat{\boldsymbol{\theta}}$ and $\mathbf{Q} = \hat{V}[\hat{\boldsymbol{\theta}}]$.

When $\hat{V}[\hat{\boldsymbol{\theta}}] = c\times\mathbf{S}$, with $c = 1/n$ (1 sample) or $c = 1/n_1+1/n_2$ (2 sample), say,

- $\hat{V}[\hat{\boldsymbol{\theta}}]$ has the same eigenvectors as $\mathbf{S}$
- The eigenvalues of $\hat{V}[\hat{\boldsymbol{\theta}}]$ are $c\times\lambda_i$, where λ_i are eigenvalues of $\mathbf{S}$
- Lengths of axes are $2K_\alpha\sqrt{\{c\times\lambda_i\}}$.

⇒ *orientation* and *shape* of $R(\mathbf{X})$ is determined by the eigenvectors and eigenvalues λ_i of $\mathbf{S}$.

- The "cloud of points" in a scatter plot of the data will have the same shape orientation as a confidence ellipse for the mean.
- The cloud of points will be larger than the confidence ellipse.

Shape of a confidence region corresponding to Bonferronized z- or t-tests

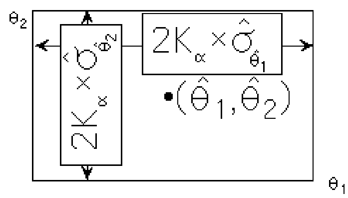
A confidence region R(**X**) related to q Bonferronized t- or z-tests of H_0: **θ** = **θ**$_0$ $((\hat\theta_j - \theta_{0j})/\hat\sigma_{\hat\theta_j})$ j = 1, ..., q, is

R(**X**) =
$\{\boldsymbol\theta \mid \hat\theta_j - K_\alpha\times\hat\sigma_{\hat\theta_j} \le \theta_j \le \hat\theta_j + K_\alpha\times\hat\sigma_{\hat\theta_j}, j=1,\ldots,q\}$,

with
$K_\alpha = t_{f_e}(\alpha'/2)$ or $K_\alpha = z(\alpha'/2)$, $\alpha' = \alpha/q$

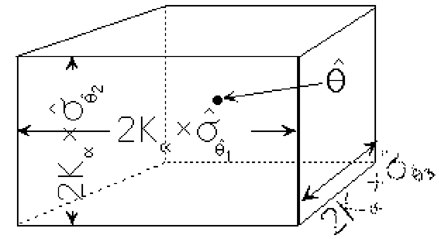
R(**X**) is rectangular or "<u>box shaped</u>".

When q = 2, R(**X**) is



a *rectangle* centered at $\hat{\boldsymbol\theta}$ with sides having lengths $k_j(\mathbf{X}) = 2K_\alpha\times\hat\sigma_{\hat\theta_j}$:

When q = 3, R(**X**) is



a *box* with sides $k_j(\mathbf{X}) = 2\times K_\alpha\times\hat\sigma_{\hat\theta_j}$..

The length $2K_\alpha\times\hat\sigma_{\hat\theta_j}$ of dimension j of the box is proportional to the *estimated standard error* $\hat\sigma_{\hat\theta_j}$ of $\hat\theta_j$.

- $\hat\sigma_{\hat\theta_j}^2$ is a diagonal element of $\hat V[\hat{\boldsymbol\theta}]$.
- Smaller α ⇒ larger K_α ⇒ larger box
- The area or volume of the box is proportional to $(K_\alpha)^q$.

For box-shaped confidence regions,

$$P(R(\mathbf{X}) \text{ contains } \boldsymbol\theta) \ge 1 - \alpha$$

with *inequality* (> 1 - α) when p > 1

Example:
When $\theta_j = \mu_j$ and $\hat{\boldsymbol\theta} = \overline{\mathbf{x}}$, so q = p, the dimensions of the box are proportional to the standard errors

$$\hat\sigma_{\overline{x}_j} = \sqrt{(s_{jj}/n)} = \sqrt{(s_j^2/n)} = s_j/\sqrt{n},$$

$s_{jj} = s_j^2$ = the sample variance of x_j.

Here
$k_j(\mathbf{X}) = 2K_\alpha\hat\sigma_{\overline{x}_j} = 2K_\alpha\sqrt{(s_{jj}/n)}$, j = 1,...,p
with

- $K_\alpha = t_{f_e}(\alpha'/2)$ <u>small sample</u>

or

- $K_\alpha = z(\alpha'/2)$ <u>large sample</u>

where Bonferronized α' = α/p

Summary

- **Box shaped** confidence regions based on Bonferronized z-tests or -t-tests

R(**X**) =
$\{\boldsymbol\theta \mid \hat\theta_j - \tilde K_\alpha\times\hat\sigma_{\hat\theta_j} \le \theta_j \le \hat\theta_j + \tilde K_\alpha\times\hat\sigma_{\hat\theta_j}, j=1,\ldots,q\}$,

with
$\tilde K_\alpha = t_{f_e}(\alpha'/2)$ or $\tilde K_\alpha = z(\alpha'/2)$, $\alpha' = \alpha/q$

$\hat\sigma_{\theta_j}$ is the estimated standard error of $\hat\theta_j$.

The shape is determined by the values of $\hat\sigma_{\theta_j}^2$, the *diagonal* elements of $\hat V[\hat{\boldsymbol\theta}]$.

- **Ellipsoidal** confidence regions based on Hotelling's T^2 test:

  $R(\mathbf{X}) = \{\boldsymbol\theta \mid (\boldsymbol\theta - \hat{\boldsymbol\theta})'\{\hat V[\hat{\boldsymbol\theta}]\}^{-1}(\boldsymbol\theta - \hat{\boldsymbol\theta}) \le K_\alpha^2\}$,
  $K_\alpha^2 = \chi_q^2(\alpha)$ or $\{(f_e q)/(f_e - q + 1)\}F_{q,f_e-q+1}(\alpha)$.

The *shape* is determined by *eigenvalues* of $\hat V[\hat{\boldsymbol\theta}]$. The *orientation* is determined by the *eigenvectors* of $\hat V[\hat{\boldsymbol\theta}]$.