Inferences about a Mean Vector
Edps/Soc 584, Psych 594

Carolyn J. Anderson

Department of Educational Psychology
ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

© Board of Trustees, University of Illinois

Spring 2017
Overview

- Goal
- Univariate Case
- Multivariate Case
  - Hotelling $T^2$
  - Likelihood Ratio test
  - Comparison/relationship
- IF Reject $H_0$...
  - Confidence regions
  - Simultaneous comparisons (univariate/one-at-a-time)
  - $T^2$–intervals
  - Bonferroni intervals
  - Comparison

- Large sample inferences about a population mean vector.

Reading: Johnson & Wichern pages 210–260
Goal

Inference: To make a valid conclusion about the means of a population based on a sample (information about the population).

When we have \( p \) correlated variables, they must be analyzed **jointly**.

Simultaneous analysis yields stronger tests, with better error control.

The tests covered in this set of notes are all of the form:

\[
H_0 : \mu = \mu_o
\]

where \( \mu_{p \times 1} \) vector of populations means and \( \mu_{o,p \times 1} \) is the some specified values under the null hypothesis.
Univariate Case

We’re interested in the mean of a population and we have a random sample of $n$ observations from the population,

$$X_1, X_2, \ldots, X_n$$

where (i.e., Assumptions):

- Observations are independent (i.e., $X_j$ is independent from $X_{j'}$ for $j \neq j'$).

- Observations are from the same population; that is,

$$E(X_j) = \mu \text{ for all } j$$

- If the sample size is “small”, we’ll also assume that

$$X_j \sim \mathcal{N}(\mu, \sigma^2)$$
Hypothesis & Test

- **Hypothesis:**
  \[ H_o : \mu = \mu_o \quad \text{versus} \quad H_1 : \mu \neq \mu_o \]
  where \( \mu_o \) is some specified value. In this case, \( H_1 \) is 2–sided alternative.

- **Test Statistic:**
  \[ t = \frac{\bar{X} - \mu_o}{s/\sqrt{n}} \]
  where \( \bar{X} = (1/n) \sum_{j=1}^{n} X_j \) and
  \[ s = \sqrt{(1/(n-1)) \sum_{j=1}^{n} (X_j - \bar{X})^2} \]

- **Sampling Distribution:** If \( H_o \) and assumptions are true, then the sampling distribution of \( t \) is Student’s - t distribution with \( df = n - 1 \).

- **Decision:** Reject \( H_o \) when \( t \) is “large” (i.e., small \( p \)-value).
Picture of Decision

Each green area = $\alpha/2 = 0.025\ldots$

Students $t$-distribution with $df=10$
Confidence Interval

Confidence Interval: A region or range of plausible μ’s (given observations/data). The set of all μ’s such that

\[
\left| \bar{x} - \mu_o \right| \leq t_{n-1,(\alpha/2)} \frac{s}{\sqrt{n}}
\]

where \( t_{n-1,(\alpha/2)} \) is the upper \((\alpha/2)100\%\) percentile of Student’s t-distribution with \( df = n - 1 \). . . . OR

\[
\left\{ \mu_o \text{ such that } \bar{x} - t_{n-1,(\alpha/2)} \frac{s}{\sqrt{n}} \leq \mu_o \leq \bar{x} + t_{n-1,(\alpha/2)} \frac{s}{\sqrt{n}} \right\}
\]

A 100(1 − α)\(th\) confidence interval or region for \(\mu\) is

\[
\left( \bar{x} - t_{n-1,(\alpha/2)} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1,(\alpha/2)} \frac{s}{\sqrt{n}} \right)
\]

Before for sample is selected, the ends of the interval depend on random variables \(\bar{X}\)’s and \(s\); this is a random interval. 100(1 − α)\(th\) percent of the time such intervals with contain the “true” mean \(\mu\).
Prepare for Jump to \( p \) Dimensions

Square the test statistic \( t \):

\[
 t^2 = \frac{(\bar{x} - \mu_o)^2}{s^2/n} = n(\bar{x} - \mu_o)(s^2)^{-1}(\bar{x} - \mu_o)
\]

So \( t^2 \) is a squared statistical distance between the sample mean \( \bar{x} \) and the hypothesized value \( \mu_o \).

Remember that \( t_{df}^2 = \mathcal{F}_{1,df} \)?

That is, the sampling distribution of

\[
 t^2 = n(\bar{x} - \mu_o)(s^2)^{-1}(\bar{x} - \mu_o) \sim \mathcal{F}_{1,n-1}.
\]
Multivariate Case: Hotelling’s $T^2$

For the extension from the univariate to multivariate case, replace scalars with vectors and matrices:

$$T^2 = n(\bar{X} - \mu_o)'S^{-1}(\bar{X} - \mu_o)$$

$\bar{X}_{p \times 1} = (1/n) \sum_{j=1}^{n} X_j$

$\mu_o, (p \times 1) = (\mu_{1o}, \mu_{2o}, \ldots, \mu_{po})$

$S_{p \times p} = \frac{1}{n-1} \sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})'$

$T^2$ is “Hotelling’s $T^2$”

The sample distribution of $T^2$

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

We can use this to test $H_0: \mu = \mu_o$...assuming that observations are a random sample from $\mathcal{N}_p(\mu, \Sigma)$ i.i.d.
Hotelling's $T^2$

Since

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

We can compute $T^2$ and compare it to

$$\frac{(n - 1)p}{n - p} F_{p, (n - p)} (\alpha)$$

OR use the fact that

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

Compute $T^2$ as

$$T^2 = n(\bar{x} - \mu_o)S^{-1}(\bar{x} - \mu_o)'$$

and the

$$p\text{-value} = \text{Prob} \left\{ F_{p, (n - p)} \geq \frac{(n - p)}{(n - 1)p} T^2 \right\}$$

Reject $H_o$ when $p\text{-value}$ is small (i.e., when $T^2$ is large).
A Really Little Example

\( n = 3 \) and \( p = 2 \)

Data: \( \mathbf{X} = \begin{pmatrix} 6 & 9 \\ 10 & 6 \\ 8 & 3 \end{pmatrix} \)

\( \mathcal{H}_0 : \mu = \begin{pmatrix} 9 \\ 5 \end{pmatrix} \)

\( \mathcal{H}_0 : \mu' = (9, 5) \)

Assuming data come from a multivariate normal distribution and independent observations,

\[
\bar{\mathbf{x}} = \begin{pmatrix} 8 \\ 6 \end{pmatrix} \quad \mathbf{S} = \begin{pmatrix} 4 & -3 \\ -3 & 9 \end{pmatrix}
\]

\[
\mathbf{S}^{-1} = \frac{1}{4(9) - (-3)(-3)} \begin{pmatrix} 9 & 3 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{pmatrix}
\]
Simple Example continued

\[ T^2 = n(\bar{x} - \mu_o)'S^{-1}(\bar{x} - \mu_o) \]
\[ = 3 \left( (8 - 9), (6 - 5) \right) \left( \begin{array}{cc} 1/3 & 1/9 \\ 1/9 & 4/27 \end{array} \right) \left( \begin{array}{c} (8 - 9) \\ (6 - 5) \end{array} \right) \]
\[ = 3(-1, 1) \left( \begin{array}{cc} 1/3 & 1/9 \\ 1/9 & 4/27 \end{array} \right) \left( \begin{array}{c} -1 \\ 1 \end{array} \right) \]
\[ = 3(7/27) = 7/9 \]

Value we need for \( \alpha = .05 \) is \( F_{2,1}(.05) = 199.51 \).

\[ \frac{(3 - 1)^2}{3 - 2} \times 199.51 = 4(199.51) = 798.04. \]

Since \( T^2 \sim \frac{(n-1)p}{(n-p)} F_{p,n-p} \), we can compare our \( T^2 \) to 798.04.

Alternatively, we could compute \( p \)-value: compare \( .25(7/9) = 0.194 \) to \( F_{2,1} \) and we get \( p \)-value = .85.

Do not reject \( H_o \). (\( \bar{x} \) and \( \mu \) are “close” in the figure).
Example: WAIS and \( n = 101 \) elderly subjects

From Morrison (1990), *Multivariate Statistical Methods*, pp 136–137:

There are two variables, verbal and performance scores for \( n = 101 \) elderly subjects aged 60–64 on the Wechsler Adult Intelligence test (WAIS).

Assume that the data are from a bivariate normal distribution with unknown mean vector \( \mu \) and unknown covariance matrix \( \Sigma \).

\[
H_0 : \mu = \begin{pmatrix} 60 \\ 50 \end{pmatrix} \quad \text{versus} \quad H_0 : \mu \neq \begin{pmatrix} 60 \\ 50 \end{pmatrix}
\]

Sample mean vector and covariance matrix:

\[
\bar{x} = \begin{pmatrix} 55.24 \\ 34.97 \end{pmatrix} \quad \text{and} \quad S = \begin{pmatrix} 210.54 & 126.99 \\ 126.99 & 119.68 \end{pmatrix}
\]
**$T^2$ for WAIS example**

We need

$$S^{-1} = \begin{pmatrix} 0.01319 & -0.0140 \\ -0.0140 & 0.02321 \end{pmatrix}$$

Compute test statistic:

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

$$= 101 ((55.24 - 60), (34.97 - 50)) \begin{pmatrix} 0.01319 & -0.0140 \\ -0.0140 & 0.02321 \end{pmatrix} \begin{pmatrix} 55.24 - 60 \\ 34.97 - 50 \end{pmatrix}$$

$$= 357.43$$

So to test the hypothesis, compute

$$\frac{(n-p) T^2}{(n-1)p} = \frac{(101 - 2) 357.43}{(101 - 1)2} = 176.93$$

Under the null hypothesis, this is distributed as $F_{p,(n-p)}$. Since $F_{2,99}(\alpha = .05) = 3.11$, we reject the null hypothesis.

**Big question:** was the null hypothesis rejected because of the verbal score, performance score, or both?
Back to the Univariate Case

Recall that for the univariate case

\[ t = \frac{\bar{X} - \mu_o}{s/\sqrt{n}} \quad \text{or} \quad t^2 = \frac{(\bar{X} - \mu_o)^2}{s^2/n} = n(\bar{X} - \mu_o)(s^2)^{-1}(\bar{X} - \mu_o) \]

Since \( \bar{X} \sim \mathcal{N}(\mu, (1/n)\sigma^2) \),

\[ \sqrt{n}(\bar{X} - \mu_o) \sim \mathcal{N}(\sqrt{n}(\mu - \mu_o), \sigma^2) \]

This is is a linear function of \( \bar{X} \), which is a random variable.

We also know that

\[ (n - 1)s^2 = \sum_{j=1}^{n}(X_j - \bar{X})^2 \sim \sigma^2 \chi^2_{(n-1)} \]

because

\[ \frac{\sum_{j=1}^{n}(X_j - \bar{X})^2}{\sigma^2} = \sum_{j=1}^{n} \frac{Z_j^2}{\sigma^2} \sim \chi^2_{(n-1)} \]
Back to the Univariate Case continued

So

\[ s^2 = \frac{\sum_{j=1}^{n}(X_j - \bar{X})^2}{n - 1} = \text{chi-square random variable} \]

Putting this all together, we find

\[ t^2 = \left( \begin{array}{c} \text{normal} \\ \text{random} \\ \text{variable} \end{array} \right) \left( \begin{array}{c} \text{chi-square random variable} \\ \text{degrees of freedom} \end{array} \right)^{-1} \left( \begin{array}{c} \text{normal} \\ \text{random} \\ \text{variable} \end{array} \right) \]

Now we’ll go through the same thing but with the multivariate case...
The Multivariate Case

\[ T^2 = \sqrt{n}(\bar{X} - \mu_o)'(S)^{-1}\sqrt{n}(\bar{X} - \mu_o) \]

Since \( \bar{X} \sim \mathcal{N}_p(\mu, (1/n)\Sigma) \) and \( \sqrt{n}(\bar{X} - \mu_o) \) is a linear combination of \( \bar{X} \),

\[ \sqrt{n}(\bar{X} - \mu_o) \sim \mathcal{N}_p(\sqrt{n}(\mu - \mu_o), \Sigma) \]

Also

\[ S = \frac{\sum_{j=1}^{n}(X_j - \bar{X})(X_j - \bar{X})'}{(n-1)} \]

\[ = \frac{\sum_{j=1}^{n}Z_jZ_j'}{(n-1)} \]

\[ = \begin{pmatrix} \text{Wishart random matrix with df = } n - 1 \end{pmatrix} \]

\[ \text{degrees of freedom} \]

where \( Z_j \sim \mathcal{N}_p(0, \Sigma) \) i.i.d. . . . if \( H_o \) is true.
The Multivariate Case continued

Recall that a Wishart distribution is a matrix generalization of the chi-square distribution.

The sampling distribution of \((n - 1)S\) is Wishart where

\[
W_m(\cdot|\Sigma) = \sum_{j=1}^{m} Z_j Z_j'
\]

where \(Z_j \sim \mathcal{N}_p(0, \Sigma) \ i.i.d..\)

So,

\[
T^2 = \begin{pmatrix}
\text{multiavirate normal random vector} \\
\text{ Wishart random matrix } \\
\text{ degrees of freedom }
\end{pmatrix}^{-1}
\begin{pmatrix}
\text{multiavirate normal random vector}
\end{pmatrix}
\]
Invariance of $T^2$

$T^2$ is invariant with respect to change of location (i.e., mean) or scale (i.e. covariance matrix); that is, a $T^2$ is invariant by linear transformation.

Rather than $X_{p \times 1}$, we may want to consider

$$Y_{p \times 1} = C_{p \times p} X_{p \times 1} + d_{p \times 1}$$

where $C$ is non-singular (or equivalently $|C| > 0$, or $C$ has $p$ linearly independent rows (columns), or $C^{-1}$ exists).

$$\nu \mu_y = C \mu_x + d \quad \text{and} \quad \Sigma_y = C \Sigma_x C'$$

The $T^2$ for the $Y$–data is exactly the same as the $T^2$ for the $X$–data (see text for proof).

This result it true for the univariate $t$-test.
Another approach to testing null hypothesis about mean vector $\mu$ (as well as other multivariate tests in general).

- It’s equivalent to Hotelling’s $T^2$ for $H_0 : \mu = \mu_o$ or $H_0 : \mu_1 = \mu_2$.
- It’s more general than $T^2$ in that it can be used to test other hypotheses (e.g., those regarding $\Sigma$) and in different circumstances.
- Foreshadow: When testing more than 1 or 2 mean vectors, there are lots of different test statistics (about 5 common ones).
- $T^2$ and likelihood ratio tests are based on different underlying principles.
Underlying Principles

$T^2$ is based on the union-intersection principle, which takes a multivariate hypothesis and turns it into a univariate problem by considering linear combinations of variables. i.e.,

$$T^2 = a'(\bar{X} - \mu_o)$$

is a linear combination.

We select the combination vector $a$ that lead to the largest possible value of $T^2$. (We’ll talk more about this later). The emphasis is on the “direction of maximal difference”.

The likelihood ratio test the emphasis is on overall difference.

**Plan:** First talk about the basic idea behind Likelihood ratio tests and then we’ll apply it to the specific problem of testing $\mu = \mu_o$. 
Basic idea of Likelihood Ratio Tests

- $\Theta_o = \text{a set of unknown parameters under } H_o \text{ (e.g., } \Sigma)\text{.}$
- $\Theta = \text{the set of unknown parameters under the alternative hypothesis (model), which is more general (e.g., } \mu \text{ and } \Sigma)\text{.}$
- $L(\cdot)$ is the likelihood function. It is a function of parameters that indicates “how likely $\Theta$ (or $\Theta_o$) is given the data”.
- $L(\Theta) \geq L(\Theta_o)$.
  - The more general model/hypothesis is always more (or equally) likely than the more restrictive model/hypothesis.

The Likelihood Ratio Statistic is

$$\Lambda = \frac{\max L(\Theta_o)}{\max L(\Theta)} \rightarrow \bar{X} = \hat{\mu} \quad \text{MLE of mean}$$
$$S_n = \hat{\Sigma} \quad \text{MLE of covariance matrix}$$

If $\Lambda$ is “small”, then the data are not likely to have occurred under $H_o \rightarrow \text{Reject } H_o$. 

If $\Lambda$ is “large”, then the data are likely to have occurred under $H_o \rightarrow \text{Retain } H_o$. 
Likelihood Ratio Test for Mean Vector

Let $X_j \sim \mathcal{N}_p(\mu, \Sigma)$ and i.i.d.

$$\Lambda = \frac{\max_{\Sigma} [\mathcal{L}(\mu_o, \Sigma)]}{\max_{\mu, \Sigma} [\mathcal{L}(\mu, \Sigma)]}$$

where

- $\max_{\Sigma} = \text{the maximum of } \mathcal{L}(\cdot) \text{ over all possible } \Sigma \text{'s.}$
- $\max_{\mu, \Sigma} = \text{the maximum of } \mathcal{L}(\cdot) \text{ over all possible } \mu \text{'s & } \Sigma \text{'s.}$

$$\Lambda = \left( \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_o|} \right)^{n/2}$$

where

- $\hat{\Sigma} = \text{MLE of } \Sigma = (1/n) \sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})'$
- $\hat{\Sigma}_o = \text{MLE of } \Sigma \text{ assuming that } \mu = \mu_o$
  $$= (1/n) \sum_{j=1}^{n} (X_j - \mu_o)(X_j - \mu_o)'$$
Likelihood Ratio Test for Mean Vector

\[ \Lambda = \left( \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_o|} \right)^{n/2} \]

\[ \Lambda = \text{(ratio of two generalized sample variances)}^{n/2} \]

- If \( \mu_o \) is really “far” from \( \mu \), then \( |\hat{\Sigma}_o| \) will be much larger than \( |\hat{\Sigma}| \), which uses a “good” estimator of \( \mu \) (i.e., \( \bar{X} \)).
- The likelihood ratio statistic \( \Lambda \) is called “Wilk’s Lambda” for the special case of testing hypotheses about mean vectors.
- For large samples (i.e., large \( n \)),

\[ -2 \ln(\Lambda) \sim \chi^2_p, \]

which can be used to test \( H_0 : \mu = \mu_o \)
Degrees of Freedom for LR Test

We need to consider the number of parameter estimates under each hypothesis:

The alternative hypothesis ("full model") ,
\[ \Theta = \{ \mu, \Sigma \} \rightarrow p \text{ means} + \frac{p(p-1)}{2} \text{ covariances} \]

The null hypothesis,
\[ \Theta_0 = \{ \Sigma \} \rightarrow \frac{p(p-1)}{2} \text{ covariances} \]

degrees of freedom = \( df = \text{difference between number of parameters estimated under each hypothesis} = p \)

If the \( H_0 \) is true and all assumptions valid, then for large samples,
\[ -2 \ln(\Lambda) \sim \chi^2_p. \]
Example: 4 Psychological Tests

\( n = 64, \ p = 4, \ \bar{x}' = (14.15, 14.91, 21.92, 22.34), \)

\[
S = \begin{pmatrix}
10.388 & 7.793 & 15.298 & 5.374 \\
15.298 & 13.707 & 57.058 & 15.932 \\
5.374 & 6.176 & 15.932 & 22.134 \\
\end{pmatrix}
& \quad \det(S) = 61952.085
\]

Test: \( H_0 : \mu' = (20, 20, 20, 20) \) versus \( H_o : \mu' \neq (20, 20, 20, 20) \)

\[
\Sigma_o = \frac{1}{n}(X-1\mu_o)'(X-1\mu_o) = \begin{pmatrix}
44.375 & 37.438 & 3.828 & -8.406 \\
37.438 & 42.344 & 3.703 & -5.859 \\
3.828 & 3.703 & 59.859 & 20.187 \\
-8.406 & -5.859 & 20.187 & 27.281 \\
\end{pmatrix}
\]

\( \det(\Sigma_o) = 518123.8. \)

Wilk's Lambda is \( \Lambda = (61952.085/518123.8)^{64/2} = 3.047E - 30, \) and

Comparing \(-2 \ln(\Lambda) = 135.92659\) to a \( \chi^2_4 \) gives \( p-\text{value} < .01. \)
Comparison of $T^2$ & Likelihood Ratio

Hotelling’s $T^2$ and Wilk’s Lambda are functionally related. Let $X_1, X_2, \ldots, X_n$ be a random sample from a $\mathcal{N}_p(\mu, \Sigma)$ population, then the test of $H_0: \mu = \mu_o$ versus $H_A: \mu \neq \mu_o$ based on $T^2$ is equivalent to the test based on $\Lambda$. The relationship is given by

$$(\Lambda)^{2/n} = \left(1 + \frac{T^2}{(n-1)}\right)^{-1}$$

So,

$$\Lambda = \left(1 + \frac{T^2}{(n-1)}\right)^{-n/2} \quad \text{and} \quad T^2 = (n-1)\Lambda^{-2/n} - (n-1)$$

Since they are inversely related,

- We reject $H_0$ for “large” $T^2$
- We reject $H_0$ for “small” $\Lambda$. 
Example: Comparison of $T^2$ & Likelihood Ratio

Using our 4 psychological test data, we found that

$$(\Lambda) = 3.047E - 30$$

If we compute Hotelling’s $T^2$ for these data we’d find that

$$T^2 = 463.88783$$

$$\Lambda = \left(1 + \frac{463.88783}{(64 - 1)}\right)^{-64/2} = 3.047E - 30$$

and

$$T^2 = (64 - 1)(3.047E - 30)^{-2/64} - (64 - 1)$$

Note: I did this in SAS. The SAS/IML code is on the web-site if you want to check this for yourself.
After Rejection: Confidence Regions

Our goal is to make inferences about populations from samples.

In univariate statistics, we form confidence intervals; we’ll generalize this to multivariate confidence region.

General definition: A **confidence region** is a region of likely values of parameters $\theta$ which is determined by data:

$$R(\mathbf{X}) = \text{confidence region}$$

where

- $\mathbf{X}' = (X_1, X_2, \ldots, X_n)$; that is, data.
- $R(\mathbf{X})$ is a $100(1 - \alpha)\%$ confidence region if before the sample was selected

$$\text{Prob}[R(\mathbf{X}) \text{ contains the true } \theta] = 1 - \alpha$$
Confidence Region for $\mu$

For $\mu_{p \times 1}$ of a $p$-dimensional multivariate normal distribution,

$$\text{Prob} \left[ n(\bar{X} - \mu)'S^{-1}(\bar{X} - \mu) \leq \frac{(n-1)p}{n-p} F_{p, n-p}(\alpha) \right] = 1 - \alpha$$

...before we have data (observations).

i.e., $\bar{X}$ is within $\sqrt{\frac{(n-1)p}{n-p} F_{p, n-p}(\alpha)}$ of $\mu$ with probability $1 - \alpha$

(where distance is measured or defined in terms of $nS^{-1}$).

For a typical sample,

1. Calculate $\bar{x}$ and $S$.
2. Find $(n-1)p/(n-p)F_{p, n-p}(\alpha)$.
3. Consider all $\mu$'s that satisfy the equation

$$n(\bar{X} - \mu)'S^{-1}(\bar{X} - \mu) \leq \frac{(n-1)p}{n-p} F_{p, n-p}(\alpha)$$

This is the confidence region, which is an equation of an ellipsoid.
Confidence Region for $\mu$ continued

To determine whether a particular $\mu^*$ falls within in a confidence region, compute the squared statistical distance of $\bar{X}$ to $\mu^*$ and see if it’s less than or greater than $(n-1)p \frac{n-1}{n-p} F_{p,n-p}(\alpha)$.

The confidence region consists of all vectors $\mu_o$ that lead to retaining the $H_0: \mu = \mu_o$ using Hotelling’s $T^2$ (or equivalently Wilk’s lambda).

These regions are ellipsoids where their shapes are determined by $S$ (the eigenvalues and eigenvectors of $S$).

We’ll continue our WAIS example of $n = 101$ elderly and the verbal and performance sub-tests of WAIS ($p = 2$).

Recall that $H_0 : \mu' = (60, 50)$

But first a closer look at the ellipsoid...
The Shape of the Ellipsoid

- The ellipsoid is centered at $\bar{x}$.
- The direction of the axes are given by the eigenvectors $e_i$ of $S$.
- The (half) length of the axes equal

$$\sqrt{\lambda_i} \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)} = \frac{\sqrt{\lambda_i}}{\sqrt{n}} c$$

So, from the center, which is at $\bar{x}$, the axes are

$$\bar{x} \pm \sqrt{\lambda_i} \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p}(\alpha)} e_i$$

where $Se_i = \lambda_i e_i$ for $i = 1, 2, \ldots, p$. 
WAIS Example

Equation for the \((1 - \alpha)100\%\) confidence region:

\[
n(\bar{x} - \mu)' S^{-1}(\bar{x} - \mu) \leq \frac{(n - 1)p}{(n - p)} F_{p, n-p}(\alpha)
\]

or

\[
T^2 \leq \frac{(n - 1)p}{(n - p)} F_{p, n-p}(\alpha)
\]

The confidence region is an ellipse (ellipsoid for \(p > 2\)) centered at \(\bar{x}\) with axes

\[
\bar{x} \pm \sqrt{\lambda_i} \sqrt{\frac{p(n - 1)}{n(n - p)} F_{p, n-p}(\alpha)} \ e_i
\]

where \(\lambda_i\) and \(e_i\) are the eigenvalues and eigenvectors, respectively, of \(S\) (\(\lambda_i\) is not Wilk's lambda).

For the WAIS data,

\[
\lambda_1 = 299.982, \quad e'_1 = (.818, .576)
\]

\[
\lambda_2 = 30.238, \quad e'_2 = (-.576, .818)
\]
WAIS Example: Finding Major and Minor

\[ \bar{x} \pm \sqrt{\lambda_i} \sqrt{\frac{p(n - 1)}{n(n - p)}} F_{p, n-p}(\alpha) \ e_i \]

The major axis:

\[
\begin{pmatrix}
55.24 \\
34.97
\end{pmatrix}
\pm \sqrt{299.982} \sqrt{\frac{2(101 - 1)}{101(101 - 2)}} 3.11 \begin{pmatrix}
.818 \\
.576
\end{pmatrix}
\]

which gives us \((51.71, 32.48)\) and \((58.77, 37.46)\).

The minor axis:

\[
\begin{pmatrix}
55.24 \\
34.97
\end{pmatrix}
\pm \sqrt{30.238} \sqrt{\frac{2(101 - 1)}{101(101 - 2)}} 3.11 \begin{pmatrix}
-.576 \\
.818
\end{pmatrix}
\]

which gives us \((56.03, 33.85)\) and \((54.45, 36.09)\).
Graph of 95% Confidence Region

Length of major = 8.64 (half-length = 4.32)
Length of minor = 2.74 (half-length = 1.37)

Performance

Verbal

(56.03, 33.85)
(58.77, 37.46)
(56.03, 33.85)
(54.45, 36.09)
(54.45, 36.09)
(51.71, 32.48)
(51.71, 32.48)

$\bar{x}' = (55.24, 34.97)$
Example continued

We note that \( \mu' = (60, 50) \) is not in the confidence region. Using the equation for the ellipse, we find

\[
T^2 = 357.43 > (100(2)/99)(3.11) = 6.283,
\]
so \((60, 50)\) is not in the 95% confidence region.

What about \( \mu' = (60, 40) \)?

\[
T^2 = 101 \left( (55.24 - 60), (34.97 - 40) \right)
\times \left( \begin{array}{cc}
.01319 & -.0140 \\
-.0140 & .02321
\end{array} \right)
\left( \begin{array}{c}
55.24 - 60 \\
34.97 - 40
\end{array} \right)
= 21.80
\]

Since 21.80 is greater than 6.28, \((60, 40)\) also in not in 95% confidence region.
Alternatives to Confidence Regions

The confidence regions consider all the components of $\mu$ jointly. We often desire a confidence statement (i.e., confidence interval) about individual components of $\mu$ or a linear combination of the $\mu_i$'s. We want all such statements to hold simultaneously with some specified large probability; that is, want to make sure that the probability that any one of the confidence statements is incorrect is small.

Three ways of forming simultaneous confidence intervals considered:

- “one-at-a-time” intervals
- $T^2$ intervals
- Bonferroni
“One-at-a-Time” Intervals

(they’re related to the confidence region).

Let \( \mathbf{X} \sim \mathcal{N}_p(\mu, \Sigma) \), where \( \mathbf{X}' = (X_1, X_2, \ldots, X_p) \) and consider the linear combination

\[
Z = a_1X_1 + a_2X_2 + \cdots + a_pX_p = a'\mathbf{X}
\]

From what we know about linear combinations of random vectors and multivariate normal distribution, we know

\[
E(Z) = \mu_z = a'\mu
\]

\[
\text{var}(Z) = \sigma_Z^2 = a'\Sigma a
\]

\( Z \sim \mathcal{N}_1(a'\mu, a'\Sigma a) \)

Estimate \( \mu_Z \) by \( a'\bar{X} \) and estimate \( \text{var}(Z) = a'\Sigma a \) by \( a'Sa \).
Univariate Intervals

A Simultaneous 100(1 − α)% confidence interval for $\mu_Z$ where $Z = a'X$ with unknown $\Sigma$ (but known $a$) is

$$\bar{z} \pm t_{n-1,(\alpha/2)} \sqrt{\frac{a'Sa}{n}}$$

where $t_{n-1,(\alpha/2)}$ is the upper 100(α/2) percentile of Student's $t$-distribution with $df = n - 1$

Can put intervals around any element of $\mu$ by choice of $a$'s:

$$a = (0, 0, \ldots, 1, 0, \ldots 0)_{i^{th} element}$$

So $a'\mu = \mu_i$ \hspace{1cm} $a'\bar{x} = \bar{x}_i$ \hspace{1cm} and \hspace{1cm} $a'Sa = s_{ii}$

and the “one-at-a-time” interval for $\mu_i$ is

$$\bar{x}_i \pm t_{n-1,(\alpha/2)} \sqrt{\frac{s_{ii}}{n}}$$
**WAIS Example: One-at-a-time Intervals**

**Univariate Confidence Intervals**

\[
\bar{x}_i \pm t_{n-1,(\alpha/2)} \sqrt{\frac{s_{ii}}{n}}
\]

We’ll let \(\alpha = .05\) (for a 95% confidence interval), so \(t_{100,.025} = 1.99\).

For verbal score:

\[
55.24 \pm 1.99\sqrt{210.54/101}
\]

\[
55.24 \pm 2.87 \rightarrow (52.37, 58.11)
\]

For performance score:

\[
34.97 \pm 1.99\sqrt{119.68/101} = 2.17
\]

\[
34.97 \pm 2.17 \rightarrow (32.80, 37.14)
\]

For our hypothesized values \(\mu_{o1} = 60\) and \(\mu_{o2} = 50\), neither are in the respective intervals.
Graph of one-at-a time intervals

Multivariate versus Univariate:
95% Confidence region (ellipse)
95% one-at-a-time intervals
Problem with Univariate Intervals

Problem with the Global coverage rate: If the rate is \( 100(1 - \alpha)\% \) for one interval, then the overall experimentwise coverage rate could be much less that \( 100(1 - \alpha)\% \).

If you want the overall coverage rate to be \( 100(1 - \alpha)\% \), then we have to consider simultaneously all possible choices for the vector \( \mathbf{a} \) such that the coverage rate over all of them is \( 100(1 - \alpha)\% \).

How?

What \( \mathbf{a} \) gives the maximum possible test-statistic? Using this \( \mathbf{a} \), consider the distribution for the maximum.

If we achieve \( (1 - \alpha) \) for the maximum, then the remainder (all others) have > \( (1 - \alpha) \).

We use the distribution of the maximum for our “fudge-factor.”

The largest value is proportional to \( \mathbf{S}^{-1} (\bar{\mathbf{x}} - \mu_o) \)
Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathcal{N}_p(\mu, \Sigma)$ population with $\det(\Sigma) > 0$, then simultaneously for all $a$, the interval

$$a' \bar{x} \pm \sqrt{\frac{p(n-1)}{(n-p)}} F_{p,n-p}(\alpha) \sqrt{\frac{a'Sa}{n}}$$

will contain $a'\mu$ with coverage rate $100(1-\alpha)\%$.

These are called "$T^2$–intervals" because the "fudge-factor" $(p(n-1)/(n-p)) F_{p,n-p}$ is the distribution of Hotelling's $T^2$. Set $a'_i = (0, 0, \ldots, 1, 0, \ldots, 0)$ for $i = 1, \ldots, p$. & compute

$$a'_i \bar{x} \pm \sqrt{\frac{p(n-1)}{(n-p)}} F_{p,n-p}(\alpha) \sqrt{\frac{a'Sa}{n}}, \quad i = 1, \ldots, p.$$
**$T^2$ Intervals**

$$a_i' \bar{x} \pm \sqrt{\frac{p(n-1)}{(n-p)}} F_{p,n-p}(\alpha) \sqrt{\frac{a'Sa}{n}}, \quad i = 1, \ldots, p,$$

are Component $T^2$ Intervals and are useful for “data snooping” because the coverage rate remains fixed at $100(1 - \alpha)\%$ regardless of

- The number of intervals you construct
- Whether or not the $a$’s are chosen *a priori*
WAIS Example

For the verbal score:

$$55.24 \pm \sqrt{\frac{100(2)}{99}} (3.11) \sqrt{210.54/101} = 55.24 \pm 3.62 \rightarrow (51.62, 58.86)$$

For the performance score:

$$34.97 \pm \sqrt{\frac{100(2)}{99}} (3.11) \sqrt{119.68/101} = 34.97 \pm 2.73 \rightarrow (32.24, 37.70)$$
WAIS: Comparison

Performance

Verbal

ellipse Confidence region

--- $T^2$ intervals

--- --- Univariate intervals
## Summary of Comparison

<table>
<thead>
<tr>
<th>One-at-a-Time</th>
<th>$T^2$ Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrower (more precise)</td>
<td>Wider (less precise)</td>
</tr>
<tr>
<td>More powerful</td>
<td>Less powerful</td>
</tr>
<tr>
<td>Liberal</td>
<td>Conservative</td>
</tr>
<tr>
<td>Coverage rate $&lt; 100(1 - \alpha)$</td>
<td>Coverage rate $\geq 100(1 - \alpha)$</td>
</tr>
<tr>
<td>Coverage rate depends on number of intervals and $S$.</td>
<td>Coverage rate does not depend on number of intervals.</td>
</tr>
<tr>
<td>Accuracy may be OK provided if reject $H_0 : \mu = \mu_o$.</td>
<td>Good if do a lot of intervals (e.g., $&gt; p$)</td>
</tr>
</tbody>
</table>

Compromise: Bonferroni
Bonferroni Intervals

This method will

- Give narrower (more precise) intervals than $T^2$, but not as narrow are the univariate ones.
- Good if
  - The intervals that you construct are decided upon *a priori*.
  - You only construct $\leq p$ intervals.
- Suppose that we want to make $m$ confidence statements about $m$ linear combinations
  \[ a_1^\prime \mu, \ a_2^\prime \mu, \ \ldots, \ a_m^\prime \mu \]
- It uses a form of the Bonferroni inequality.
Bonferroni Inequality

\[
\text{Prob\{all intervals are valid\}} = 1 - \text{Prob\{at least 1 false\}} \\
\geq 1 - \sum_{i=1}^{m} \text{Prob\{i}^{th}\text{ interval is false\}} \\
= 1 - \sum_{i=1}^{m} 1 - \text{Prob\{i}^{th}\text{ interval is true\}} \\
= 1 - \sum_{i=1}^{m} \alpha_i
\]

This is a form of the Bonferroni inequality:

\[
\text{Prob\{all intervals are true\}} \geq 1 - (\alpha_1 + \alpha_2 + \cdots + \alpha_m)
\]

We set \(\alpha_i = \alpha/m\) using a pre-determined \(\alpha\)-level, then

\[
\text{Prob\{all intervals are true\}} \geq 1 - \left(\frac{\alpha}{m} + \frac{\alpha}{m} + \cdots + \frac{\alpha}{m}\right) = 1 - \alpha
\]

\(m\) of these
Bonferroni Confidence Statements

Use $\alpha/m$ for each of the $m$ intervals (both $\alpha$ and specific intervals pre-determined)

$$a'\bar{x} \pm \left[ t_{n-1,(\alpha/2m)} \sqrt{\frac{a'Sa}{n}} \right]$$

We just replace the "fudge-factor"

**WAIS example:** We’ll only consider $a'_1 = (1,0)$ and $a_2 = (0,1)$ (i.e., the component means).

$$df = n - 1 = 101 - 1 = 100$$

$$\alpha = .05 \longrightarrow \alpha/2 = .025$$

$$t_{100, (.025/2)} = 2.2757$$

You can get t’s from the “pvalue.exe” program on course web-site (under handy programs and links), or from SAS using, for example...
WAIS & Bonferroni Intervals

data tvalue;
  df= 100;
p = 1 − .05/(2 * 2);  * ←− α/(p × m);
t= quantile('t',p,100);
proc print;
run;

Verbal Scores:

55.25 ± 2.2757 \sqrt{210.54/101}
± 2.2757(1.4438)
± 3.2856 ↦ (51.95, 58.53)

Performance Scores:

34.97 ± 2.2757 \sqrt{119.68/101}
± 2.2757(1.08855)
± 2.477 ↦ (32.49, 37.45)
WAIS: All four Confidence Methods

- Ellipse Confidence region
- \( T^2 \) intervals
- Bonferroni
- Univariate intervals

Performance

Verbal

C.J. Anderson (Illinois)
Interval Methods Comparisons

Points are some possible places for $\mu_o$.
Few last statements on Confidence Statements

- Hypothesis testing of \( H_0 : \mu = \mu_o \) may lead to some seemingly inconsistent results. For example,
  - The multivariate tests may reject \( H_0 \), but the component means are within their respective confidence intervals for them (regardless of how intervals are computed, e.g., the red dot).
  - Separate \( t \)-tests for component means may not be rejected, but you do reject for multivariate (e.g., orange dot).

- The confidence region, which contains all values of \( \mu_o \) for which the null hypothesis would not be rejected, is the only one that takes into consideration the covariances, as well as variances.

- Multivariate approach is most powerful.

- In higher dimensions, we can’t “see” what’s going on, but concepts are same.
In the Face of Inconsistencies

or to get a better idea of what’s going on...

Recall that $T^2$ is based on the “union intersection” principle:

$$T^2 = na'(\bar{X} - \mu_o)$$

where $a$ is the one that gives the largest value for $T^2$ among all possible vectors $a$. This vector is

$$a = (\bar{X} - \mu_o)'S^{-1}$$

Examining $a$ can lead to insight into why $H_o: \mu = \mu_o$ was rejected.

For the WAIS example when $H_o: \mu' = (60, 50)$,

$$(\bar{X} - \mu_o)'S^{-1} = (0.15, -0.28)$$

Note: $(\bar{X} - \mu_o)' = (-4.76, -15.03)$
Large-Sample Inferences

about a population mean vector $\mu$

So far, we’ve assumed that $X_j \sim \mathcal{N}_p(\mu, \Sigma)$. But what if the data are not multivariate normal?

We can still make inferences (hypothesis testing & make confidence statements) about population means IF we have Large samples relative to $p$ (i.e., $n - p$ is large).

Let $X_1, X_2, \ldots, X_n$ be a random sample from a population with $\mu$ and $\Sigma$ ($\Sigma$ is positive definite)

$$T^2 = n(\bar{x} - \mu_o)'S^{-1}(\bar{x} - \mu_o) \approx \chi^2_p$$

- $\approx$ means “approximately”.
- $\text{Prob}(n(\bar{x} - \mu_o)'S^{-1}(\bar{x} - \mu_o)) \leq \chi^2_p(\alpha) \approx 1 - \alpha$.
- As $n$ gets large, $\mathcal{F}_{p,n-p}$ and $\chi^2_p(\alpha)$ become closer in value:

$$\text{As } n \to \infty, \quad \frac{(n-1)p}{n-p} \mathcal{F}_{p,n-p} \to \chi^2_p$$

(Show this)
Large-Sample Inferences continued

For large $n - p$,

- Hypothesis test:
  \[ H_0 : \mu = \mu_0 \]

  Reject $H_0$ if $T^2 > \chi^2_p(\alpha)$ where $\chi^2_p(\alpha)$ is the upper $\alpha^{th}$ percentile of the chi-square distribution with $df = p$.

- Simultaneous $T^2$ intervals:
  \[ a'\bar{x} \pm \sqrt{\chi^2_p(\alpha)}\sqrt{\frac{a'Sa}{n}} \]

- Confidence region for $\mu$:
  \[ (\bar{x} - \mu)'S^{-1}(\bar{x} - \mu) \leq \frac{\chi^2_p(\alpha)}{n} \]
WAIS: Large-Sample

- WAIS example with $n = 101$,

\[ F_{p, n-p}(\alpha) = F_{2, 99}(0.05) = 3.11 \]

\[ \frac{(n - 1)p}{n - p} F_{p, n-p} = \frac{100(2)}{99} (3.11) = 6.28 \]

\[ \chi^2_2(0.05) = 5.99 \]

The value 6.28 is fairly close to 5.99.

- It’s generally true that the more you assume, the more powerful your test (more precise estimates).

- The larger $n \to$, the more power. . . This is generally true.
Show How to Do Tests, etc. . . .

- SAS PROC IML and tests
- Use Psychological test scores (on course web-site)