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_o...
    - 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:

$$H_o{:}\boldsymbol{\mu} = \boldsymbol{\mu}_o$$

where $\boldsymbol{\mu}_{p\times 1}$ vector of populations means and $\boldsymbol{\mu}_{o,p\times 1}$ 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:

$$H_o : \mu = \mu_o \qquad \text{versus} \qquad H_1 : \mu \neq \mu_o$$

where μ_o is some specified value. In this case, H_1 is 2–sided alternative.

- Test Statistic:

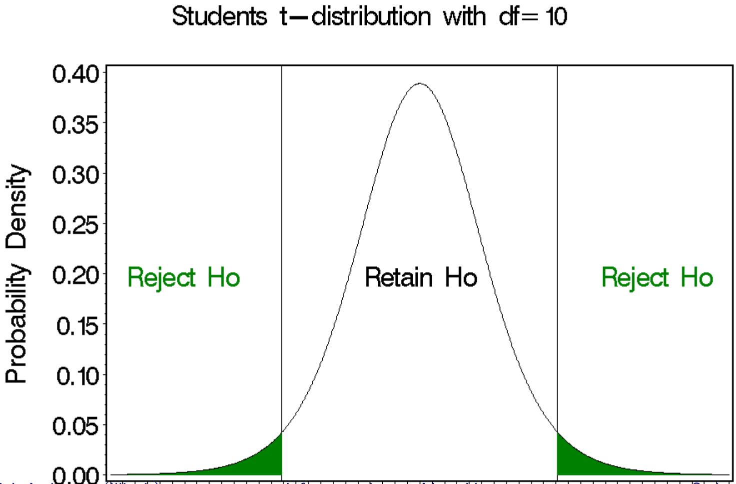
$$t = \frac{\bar{X} - \mu_o}{s/\sqrt{n}}$$

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 $= \alpha/2 = .025\ldots$



Confidence Interval

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

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

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 $100(1-\alpha)^{th}$ confidence interval or region for μ is

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

Before for sample is selected, the ends of the interval depend on random variables $\bar{X}$'s and s; this is a random interval. $100(1-\alpha)^{th}$ percent of the time such intervals with contain the "true" mean μ.

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 μ_o.

Remember that $t^2_{df} = \mathcal{F}_{1,df}$?

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:

$$T^2 = n(\bar{\mathbf{X}} - \boldsymbol{\mu}_o)'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}_o)$$

- $\bar{\mathbf{X}}_{p\times 1} = (1/n)\sum_{j=1}^{n} \mathbf{X}_j$
- $\boldsymbol{\mu}_{o,(p\times 1)} = (\mu_{1o}, \mu_{2o}, \ldots, \mu_{po})$
- $S_{p\times p} = \frac{1}{n-1}\sum_{j=1}^{n}(\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})'$

T^2 is "Hotelling's T^2"

The sample distribution of T^2

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

We can use this to test $H_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$. . . assuming that observations are a random sample from $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ $i.i.d.$

Hotelling's T^2

Since

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

We can compute T^2 and compare it to

$$\frac{(n-1)p}{n-p}\mathcal{F}_{p,(n-p)}(\alpha)$$

OR use the fact that

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

Compute T^2 as

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

and the

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

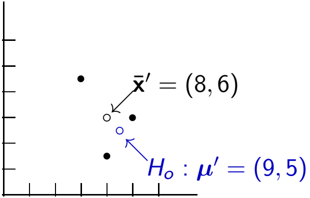
Reject H_o when p-value is small (i.e., when T^2 is large).

A Really Little Example

$n = 3$ and $p = 2$

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

$$H_o : \boldsymbol{\mu} = \begin{pmatrix} 9 \\ 5 \end{pmatrix}$$



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

$$\bar{\mathbf{x}} = \begin{pmatrix} 8 \\ 6 \end{pmatrix} \qquad \mathbf{S} = \begin{pmatrix} 4 & -3 \\ -3 & 9 \end{pmatrix}$$

$$\mathbf{S}^{-1} = \frac{1}{4(9) - (-3)(-3)} \begin{pmatrix} 9 & 3 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{pmatrix}$$

Simple Example continued

$$\begin{aligned} T^2 &= n(\bar{\mathbf{x}}-\boldsymbol{\mu}_o)'\mathbf{S}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}_o) \\ &= 3\left((8-9),(6-5)\right)\begin{pmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{pmatrix}\begin{pmatrix} (8-9) \\ (6-5) \end{pmatrix} \\ &= 3(-1,1)\begin{pmatrix} 1/3 & 1/9 \\ 1/9 & 4/27 \end{pmatrix}\begin{pmatrix} -1 \\ 1 \end{pmatrix} \\ &= 3(7/27) = 7/9 \end{aligned}$$

Value we need for $\alpha = .05$ is $\mathcal{F}_{2,1}(.05) = 199.51$.

$$\frac{(3-1)2}{3-2}199.51 = 4(199.51) = 798.04.$$

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

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

Do not reject H_o. ($\bar{x}$ and $\boldsymbol{\mu}$ 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 $\boldsymbol{\mu}$ and unknown covariance matrix $\boldsymbol{\Sigma}$.

$$H_o : \boldsymbol{\mu} = \begin{pmatrix} 60 \\ 50 \end{pmatrix} \qquad \text{versus} \qquad H_o : \boldsymbol{\mu} \neq \begin{pmatrix} 60 \\ 50 \end{pmatrix}$$

Sample mean vector and covariance matrix:

$$\bar{\mathbf{x}} = \begin{pmatrix} 55.24 \\ 34.97 \end{pmatrix} \qquad \text{and} \qquad \mathbf{S} = \begin{pmatrix} 210.54 & 126.99 \\ 126.99 & 119.68 \end{pmatrix}$$

T^2 for WAIS example

We need

$$\mathbf{S}^{-1} = \begin{pmatrix} .01319 & -.0140 \\ -.0140 & .02321 \end{pmatrix}$$

Compute test statistic:

$$\begin{aligned} T^2 &= n(\bar{\mathbf{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \\ &= 101\left((55.24 - 60), (34.97 - 50)\right)\begin{pmatrix} .01319 & -.0140 \\ -.0140 & .02321 \end{pmatrix}\begin{pmatrix} 55.24 - 60 \\ 34.97 - 50 \end{pmatrix} \\ &= 357.43 \end{aligned}$$

So to test the hypothesis, compute

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

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

$$\frac{\sum_{j=1}^{n}(X_j - \bar{X})^2}{\sigma^2} = \sum_{j=1}^{n} Z_j^2 \sim \chi^2_{(n-1)}$$

Back to the Univariate Case continued

So

$$s^2 = \frac{\sum_{j=1}^{n}(X_j - \bar{X})^2}{n-1} = \frac{\text{chi-square random variable}}{\text{degrees of freedom}}$$

Putting this all together, we find

$$t^2 = \begin{pmatrix} \text{normal} \\ \text{random} \\ \text{variable} \end{pmatrix} \left(\frac{\text{chi-square random varible}}{\text{degress of freedom}} \right)^{-1} \begin{pmatrix} \text{normal} \\ \text{random} \\ \text{variable} \end{pmatrix}$$

Now we'll go through the same thing but with the multivariate case. . .

The Multivariate Case

$$T^2 = \sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu}_o)'(\mathbf{S})^{-1}\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu}_o)$$

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

$$\sqrt{n}(\bar{\mathbf{X}} - \boldsymbol{\mu}_o) \sim \mathcal{N}_p(\sqrt{n}(\boldsymbol{\mu} - \boldsymbol{\mu}_o), \boldsymbol{\Sigma})$$

Also

$$\begin{aligned}
\mathbf{S} &= \frac{\sum_{j=1}^{n}(\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})'}{(n-1)} \\
&= \frac{\sum_{j=1}^{n}\mathbf{Z}_j\mathbf{Z}_j'}{(n-1)} \\
&= \left(\frac{\text{Wishart random matrix with df} = n-1}{\text{degrees of freedom}}\right)
\end{aligned}$$

where $\mathbf{Z}_j \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma})$ $i.i.d.\ldots$ 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 $(n-1)\mathbf{S}$ is Wishart where

$$\mathbf{W}_m(\cdot|\mathbf{\Sigma}) = \sum_{j=1}^{m} \mathbf{Z}_j \mathbf{Z}_j'$$

where $\mathbf{Z}_j \sim \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma})$ $i.i.d.$.
So,

$$T^2 = \begin{pmatrix} \text{multiavirate} \\ \text{normal} \\ \text{random} \\ \text{vector} \end{pmatrix} \left(\frac{\text{Wishart random matrix}}{\text{degress of freedom}} \right)^{-1} \begin{pmatrix} \text{multiavirate} \\ \text{normal} \\ \text{random} \\ \text{vector} \end{pmatrix}$$

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 $\mathbf{X}_{p\times 1}$, we may want to consider

$$\mathbf{Y}_{p\times 1} = \underbrace{\mathbf{C}_{p\times p}}_{scale}\mathbf{X}_{p\times 1} + \underbrace{\mathbf{d}_{p\times 1}}_{location}$$

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

$$v\mu_y = \mathbf{C}\boldsymbol{\mu}_x + \mathbf{d} \qquad \text{and} \qquad \boldsymbol{\Sigma}_y = \mathbf{C}\boldsymbol{\Sigma}_x\mathbf{C}'$$

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.

Likelihood Ratio

- Another approach to testing null hypothesis about mean vector $\boldsymbol{\mu}$ (as well as other multivariate tests in general).
- It's equivalent to Hotelling's T^2 for $H_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$ or $H_o : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$.
- It's more general than T^2 in that it can be used to test other hypotheses (e.g., those regarding $\boldsymbol{\Sigma}$) 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.,

$$T^2 = \mathbf{a}'(\bar{\mathbf{X}} - \boldsymbol{\mu}_o)$$

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 $\boldsymbol{\mu} = \boldsymbol{\mu}_o$.

Basic idea of Likelihood Ratio Tests

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

The Likelihood Ratio Statistic is

$$\Lambda = \frac{\max \mathcal{L}(\Theta_o)}{\max \mathcal{L}(\Theta)} \quad \rightarrow \quad \begin{array}{lll} \bar{\mathbf{X}} = \hat{\boldsymbol{\mu}} & \text{MLE of mean} \\ \mathbf{S}_n = \hat{\mathbf{\Sigma}} & \text{MLE of covariance matrix} \end{array}$$

If Λ is "small", then the data are not likely to have occurred under $H_o \longrightarrow$ Reject H_o.

If Λ is "large", then the data are likely to have occurred under $H_o \longrightarrow$ Retain H_o.

Likelihood Ratio Test for Mean Vector

Let $\mathbf{X}_j \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and *i.i.d*.

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

where

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

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

where

- $\hat{\boldsymbol{\Sigma}}$ = MLE of $\boldsymbol{\Sigma} = (1/n)\sum_{j=1}^{n}(\mathbf{X}_j - \bar{\mathbf{X}})(\mathbf{X}_j - \bar{\mathbf{X}})' = \mathbf{S}_n$
- $\hat{\boldsymbol{\Sigma}}_o$ = MLE of $\boldsymbol{\Sigma}$ assuming that $\boldsymbol{\mu} = \boldsymbol{\mu}_o$
  $= (1/n)\sum_{j=1}^{n}(\mathbf{X}_j - \boldsymbol{\mu}_o)(\mathbf{X}_j - \boldsymbol{\mu}_o)'$

Likelihood Ratio Test for Mean Vector

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

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

- If $\boldsymbol{\mu}_o$ is really "far" from $\boldsymbol{\mu}$, then $|\hat{\mathbf{\Sigma}}_o|$ will be much larger than $|\hat{\mathbf{\Sigma}}|$, which uses a "good" estimator of $\boldsymbol{\mu}$ (i.e., $\bar{\mathbf{X}}$).
- The likelihood ratio statistic Λ 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 $H_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$

Degrees of Freedom for LR Test

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

The alternative hypothesis ("full model") ,

$$\Theta = \{\boldsymbol{\mu}, \boldsymbol{\Sigma}\} \longrightarrow p \text{ means} \quad + \quad \frac{p(p-1)}{2} \text{ covariances}$$

The null hypothesis,

$$\Theta_o = \{\boldsymbol{\Sigma}\} \longrightarrow \frac{p(p-1)}{2} \text{ covariances}$$

$$\begin{aligned} \text{degrees of freedom} = df \quad &= \quad \text{difference between number of parameters} \\ &\quad\quad \text{estimated under each hypothesis} \\ &= \quad p \end{aligned}$$

If the H_o is true and all assumptions valid, then for large samples, $-2\ln(\Lambda) \sim \chi^2_p$.

Example: 4 Psychological Tests

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

$$\mathbf{S} = \begin{pmatrix} 10.388 & 7.793 & 15.298 & 5.3740 \\ 7.793 & 16.658 & 13.707 & 6.1756 \\ 15.298 & 13.707 & 57.058 & 15.932 \\ 5.374 & 6.176 & 15.932 & 22.134 \end{pmatrix} \quad \& \ \det(\mathbf{S}) = 61952.085$$

Test: $H_o : \boldsymbol{\mu}' = (20, 20, 20, 20)$ versus $H_o : \boldsymbol{\mu}' \neq (20, 20, 20, 20)$

$$\boldsymbol{\Sigma}_o = \frac{1}{n}(\mathbf{X} - \mathbf{1}\boldsymbol{\mu}_o')'(\mathbf{X} - \mathbf{1}\boldsymbol{\mu}_o') = \begin{pmatrix} 44.375 & 37.438 & 3.828 & -8.406 \\ 37.438 & 42.344 & 3.703 & -5.859 \\ 3.828 & 3.703 & 59.859 & 20.187 \\ -8.406 & -5.859 & 20.187 & 27.281 \end{pmatrix}$$

$\det(\boldsymbol{\Sigma}_o) = 518123.8$.

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

Comparing $-2\ln(\Lambda) = 135.92659$ to a χ^2_4 gives p–value $<< .01$.

Comparison of T^2 & Likelihood Ratio

Hotelling's T^2 and Wilk's Lambda are functionally related.

Let $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from a $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population, then the test of $H_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$ versus $H_A : \boldsymbol{\mu} \neq \boldsymbol{\mu}_o$ based on T^2 is equivalent to the test based on Λ.
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_o for "large" T^2
- We reject H_o for "small" Λ.

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'ld 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 generalization this to multivariate confidence region.

General definition: A confidence region is a region of likely values of parameters θ which is determined by data:

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

where

- $\mathbf{X}' = (\mathbf{X}_1, \mathbf{X}_2, \ldots \mathbf{X}_n)$; 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 μ

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

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

. . . before we have data (observations).

i.e., $\bar{\mathbf{X}}$ is within $\sqrt{\frac{(n-1)p}{n-p}\mathcal{F}_{p,n-p}(\alpha)}$ of $\boldsymbol{\mu}$ with probability $1-\alpha$ (where distance is measured or defined in terms of $n\mathbf{S}^{-1}$).

For a typical sample,

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

$$n(\bar{\mathbf{X}}-\boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{X}}-\boldsymbol{\mu}) \leq \frac{(n-1)p}{n-p}\mathcal{F}_{p,n-p}(\alpha)$$

This is the confidence region, which is an equation of an ellipsoid.

Confidence Region for μ continued

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

The confidence region consists of all vectors $\boldsymbol{\mu}_o$ that lead to retaining the $H_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$ using Hotelling's T^2 (or equivalently Wilk's lambda).

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

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

Recall that $H_o : \boldsymbol{\mu}' = (60, 50)$

But first a closer look at the ellipsoid...

The Shape of the Ellipsoid

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

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

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

$$\bar{\mathbf{x}} \pm \sqrt{\lambda_i}\sqrt{\frac{p(n-1)}{n(n-p)}\mathcal{F}_{p,n-p}(\alpha)} \quad \mathbf{e}_i$$

where $\mathbf{S}\mathbf{e}_i = \lambda_i \mathbf{e}_i$ for $i = 1, 2, \ldots, p$.

WAIS Example

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

$$n(\bar{\mathbf{x}}-\boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}) \le \frac{(n-1)p}{(n-p)}\mathcal{F}_{p,n-p}(\alpha)$$

$$\text{or} \qquad T^2 \le \frac{(n-1)p}{(n-p)}\mathcal{F}_{p,n-p}(\alpha)$$

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

$$\bar{\mathbf{x}} \pm \sqrt{\lambda_i}\sqrt{\frac{p(n-1)}{n(n-p)}\mathcal{F}_{p,n-p}(\alpha)} \quad \mathbf{e}_i$$

where λ_i and $\mathbf{e}_i$ are the eigenvalues and eigenvectors, respectively, of $\mathbf{S}$ (λ_i is not Wilk's lambda).

For the WAIS data,

$$\lambda_1 = 299.982, \quad \mathbf{e}_1' = (.818, .576)$$

$$\lambda_2 = 30.238, \quad \mathbf{e}_2' = (-.576, .818)$$

WAIS Example: Finding Major and Minor

$$\bar{\mathbf{x}} \pm \sqrt{\lambda_i}\sqrt{\frac{p(n-1)}{n(n-p)}\mathcal{F}_{p,n-p}(\alpha)} \quad \mathbf{e}_i$$

The major axis:

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

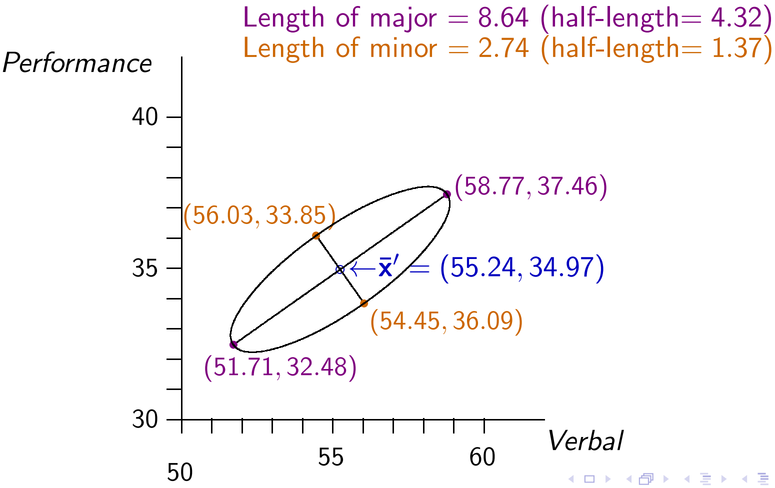
which gives us $(51.71, 32.48)$ and $(58.77, 37.46)$.

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



Example continued

We note that $\boldsymbol{\mu}_o' = (60, 50)$ 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 $\boldsymbol{\mu}' = (60, 40)$?

$$\begin{aligned} T^2 &= 101\left((55.24-60),(34.97-40)\right) \\ &\quad \times \begin{pmatrix} .01319 & -.0140 \\ -.0140 & .02321 \end{pmatrix} \begin{pmatrix} 55.24-60 \\ 34.97-40 \end{pmatrix} \\ &= 21.80 \end{aligned}$$

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 $\boldsymbol{\mu}$ jointly.
We often desire a confidence statement (i.e, confidence interval) about individual components of $\boldsymbol{\mu}$ or a linear combination of the μ_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 $\mathbf{X} \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\mathbf{X}' = (X_1, X_2, \ldots, X_p)$ and consider the linear combination

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

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

$$\begin{aligned} E(Z) &= \mu_z = \mathbf{a}'\boldsymbol{\mu} \\ \text{var}(Z) &= \sigma_Z^2 = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a} \\ Z &\sim \mathcal{N}_1(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}) \end{aligned}$$

Estimate $\boldsymbol{\mu}_Z$ by $\mathbf{a}'\bar{\mathbf{X}}$ and estimate $\text{var}(Z) = \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}$ by $\mathbf{a}'\mathbf{S}\mathbf{a}$.

Univariate Intervals

A Simultaneous $100(1-\alpha)\%$ confidence interval for μ_Z where $Z = \mathbf{a}'\mathbf{X}$ with unknown $\mathbf{\Sigma}$ (but known $\mathbf{a}$) is

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

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

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

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

$$\text{So } \mathbf{a}'\boldsymbol{\mu} = \boldsymbol{\mu}_i \qquad \mathbf{a}'\bar{\mathbf{x}} = \bar{x}_i \qquad \text{and} \qquad \mathbf{a}'\mathbf{S}\mathbf{a} = s_{ii}$$

and the "one-at-a-time" interval for μ_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

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

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

For verbal score:

$$55.24 \pm 1.99\sqrt{210.54/101}$$

$$55.24 \pm 2.87 \qquad \longrightarrow \qquad (52.37, 58.11)$$

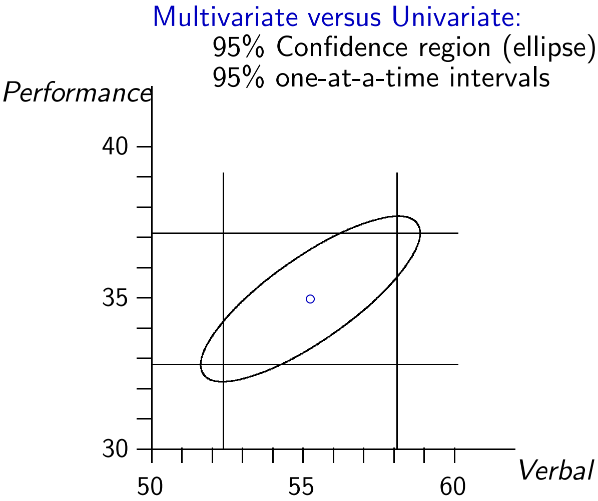
For performance score:

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

$$34.97 \pm 2.17 \qquad \longrightarrow \qquad (32.80, 37.14)$$

For our hypothesized values $\mu_{o1} = 60$ and $\mu_{o2} = 50$, neither are in the respective intervals.

Graph of one-at-a time intervals



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 **a** such that the coverage rate over all of them is $100(1-\alpha)\%$

How?

What **a** gives the maximum possible test-statistic? Using this **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 $\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_o)$

T^2 Intervals

Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ population with $\det(\boldsymbol{\Sigma}) > 0$, then simultaneously for all $\mathbf{a}$, the interval

$$\mathbf{a}'\bar{\mathbf{x}} \pm \sqrt{\frac{p(n-1)}{(n-p)}\mathcal{F}_{p,n-p}(\alpha)}\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}$$

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

These are called "T^2–intervals" because the "fudge-factor" $(p(n-1)/(n-p))\mathcal{F}_{p,n-p}$ is the distribution of Hotelling's T^2.

Set $\mathbf{a}_i' = (0, 0, \ldots, \underbrace{1}_{i^{th} element}, 0, \ldots 0) \qquad i = 1, \ldots, p$. & compute

$$\underbrace{\mathbf{a}_i'\bar{\mathbf{x}}}_{\bar{x}_i} \pm \sqrt{\frac{p(n-1)}{(n-p)}\mathcal{F}_{p,n-p}(\alpha)}\underbrace{\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}}_{s_{ii}/n}, \qquad i = 1, \ldots, p.$$

T^2 Intervals

$$\mathbf{a}_i'\bar{\mathbf{x}} \pm \sqrt{\frac{p(n-1)}{(n-p)}\mathcal{F}_{p,n-p}(\alpha)}\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}, \qquad i = 1, \ldots, p.$$

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

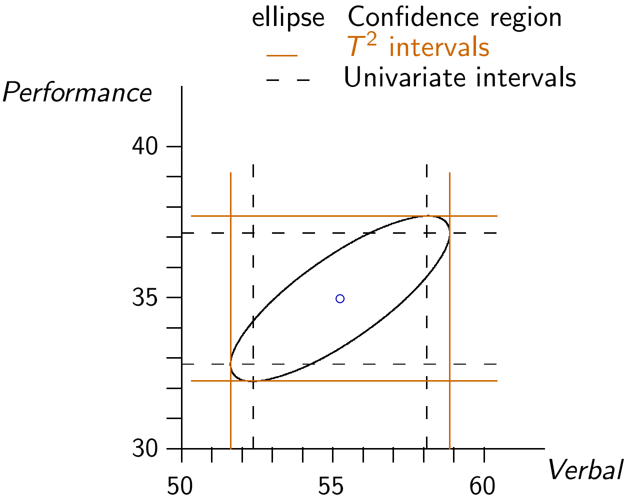
For the verbal score:

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

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



Summary of Comparison

One-at-a-Time	T^2 Intervals
Narrower (more precise)	Wider (less precise)
More powerful	Less powerful
Liberal	Conservative
Coverage rate $< 100(1-\alpha)$	Coverage rate $\geq 100(1-\alpha)$
Coverage rate depends on number of intervals and $\mathbf{S}$.	Coverage rate does not depend on number of intervals.
Accuracy may be OK provided if reject $H_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$.	Good if do a lot of intervals (e.g., $> p$)

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

$$\mathbf{a}_1'\boldsymbol{\mu}, \quad \mathbf{a}_2'\boldsymbol{\mu}, \quad \ldots, \quad \mathbf{a}_m'\boldsymbol{\mu}$$

- It uses a form of the Bonferroni inequality.

Bonferroni Inequality

$$
\begin{aligned}
\text{Prob}\{\text{all intervals are valid}\} &= 1 - \text{Prob}\{\text{at least 1 false}\} \\
&\geq 1 - \sum_{i=1}^{m} \text{Prob}\{i^{th} \text{ interval is false}\} \\
&= 1 - \sum_{i=1}^{m} 1 - \text{Prob}\{i^{th} \text{ interval is true}\} \\
&= 1 - \sum_{i=1}^{m} \alpha_i
\end{aligned}
$$

This is a form of the Bonferroni inequality:

$$\text{Prob}\{\text{all intervals are ture}\} \geq 1 - (\alpha_1 + \alpha_2 + \cdots + \alpha_m)$$

We set $\alpha_i = \alpha/m$ using a pre-determined α–level, then

$$\text{Prob}\{\text{all intervals are ture}\} \geq 1 - \underbrace{(\alpha/m + \alpha/m + \cdots + \alpha/m)}_{m \text{ of these}} = 1 - \alpha$$

Bonferroni Confidence Statements

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

$$\mathbf{a}'\bar{\mathbf{x}} \pm \underbrace{t_{n-1,(\alpha/2m)}} \sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}$$

We just replace the "fudge-factor"

WAIS example: We'll only consider $\mathbf{a}_1' = (1,0)$ and $\mathbf{a}_2 = (0,1)$ (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;
```

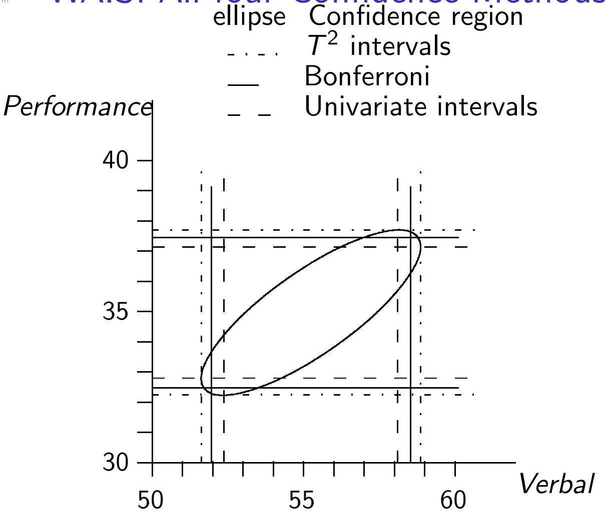
Verbal Scores:

$$\begin{aligned} 55.25 &\pm 2.2757\sqrt{210.54/101} \\ &\pm 2.2757(1.4438) \\ &\pm 3.2856 \longrightarrow (51.95, 58.53) \end{aligned}$$

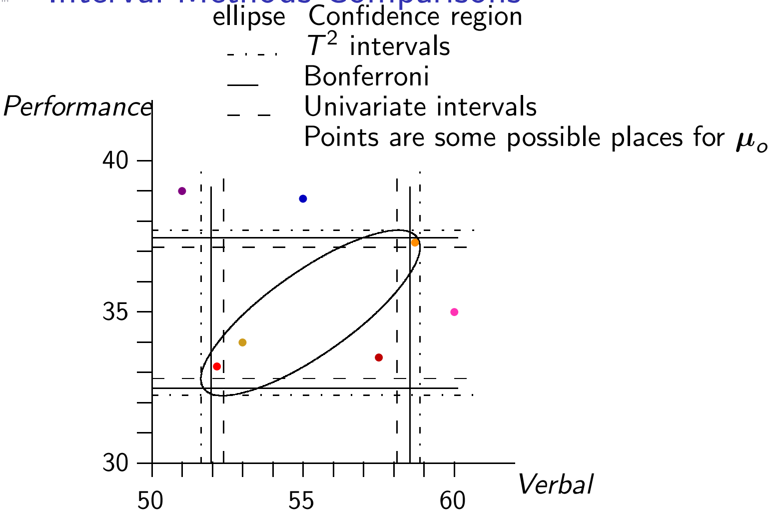
Performance Scores:

$$\begin{aligned} 34.97 &\pm 2.2757\sqrt{119.68/101} \\ &\pm 2.2757(1.08855) \\ &\pm 2.477 \longrightarrow (32.49, 37.45) \end{aligned}$$

WAIS: All four Confidence Methods



Interval Methods Comparisons



Few last statements on Confidence Statements

- Hypothesis testing of $H_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$ may lead to some seemingly inconsistent results. For example,
  - The multivariate tests may reject H_o, 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 μ_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:

$$T^2 = n\mathbf{a}'(\bar{\mathbf{X}} - \boldsymbol{\mu}_o)$$

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

$$\mathbf{a} = (\bar{\mathbf{X}} - \boldsymbol{\mu}_o)'\mathbf{S}^{-1}$$

Examining **a** can lead to insight into why $H_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$ was rejected.

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

$$(\bar{\mathbf{X}} - \boldsymbol{\mu}_o)'\mathbf{S}^{-1} = \begin{pmatrix} 0.15 & -0.28 \end{pmatrix}$$

Note: $(\bar{\mathbf{X}} - \boldsymbol{\mu}_o)' = (-4.76, -15.03)$

Large-Sample Inferences

about a population mean vector $\boldsymbol{\mu}$

So far, we've assumed that $\mathbf{X}_j \sim \mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. 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 $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n$ be a random sample from a population with $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ ($\boldsymbol{\Sigma}$ is positive definite)

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

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

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

(Show this)

Large-Sample Inferences continued

For large $n - p$,

- Hypothesis test:

$$H_o : \boldsymbol{\mu} = \boldsymbol{\mu}_o$$

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

- Simultaneous T^2 intervals:

$$\mathbf{a}'\bar{\mathbf{x}} \pm \sqrt{\chi^2_p(\alpha)}\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}$$

- Confidence region for $\boldsymbol{\mu}$:

$$(\bar{\mathbf{x}} - \boldsymbol{\mu})'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}) \leq \frac{\chi^2_p(\alpha)}{n}$$

WAIS: Large-Sample

- WAIS example with $n = 101$,

$$\mathcal{F}_{p,n-p}(\alpha) = \mathcal{F}_{2,99}(.05) = 3.11$$

$$\frac{(n-1)p}{n-p}\mathcal{F}_{p,n-p} = \frac{100(2)}{99}(3.11) = 6.28$$

$$\chi^2_2(.05) = 5.99$$

  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 $n \rightarrow$, 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)