

Tests for Multivariate Means

Max Turgeon

STAT 7200–Multivariate Statistics

Objectives

- Construct tests for a single multivariate mean
- Discuss and compare confidence regions and confidence intervals
- Describe connection with Likelihood Ratio Test
- Construct tests for two multivariate means
- Present robust alternatives to these tests

Test for a multivariate mean: Σ known

- Let $\mathbf{Y}_1, \dots, \mathbf{Y}_n \sim N_p(\mu, \Sigma)$ be independent.
- We saw in a previous lecture that

$$\bar{\mathbf{Y}} \sim N_p\left(\mu, \frac{1}{n}\Sigma\right).$$

- This means that

$$n(\bar{\mathbf{Y}} - \mu)^T \Sigma^{-1} (\bar{\mathbf{Y}} - \mu) \sim \chi^2(p).$$

- In particular, if we want to test $H_0 : \mu = \mu_0$ at level α , then we reject the null hypothesis if

$$n(\bar{\mathbf{Y}} - \mu_0)^T \Sigma^{-1} (\bar{\mathbf{Y}} - \mu_0) > \chi_{\alpha}^2(p).$$

Example i

```
library(dslabs)
library(tidyverse)

dataset <- filter(gapminder, year == 2012,
                  !is.na(infant_mortality))

dataset <- dataset[,c("infant_mortality",
                      "life_expectancy",
                      "fertility")]

dataset <- as.matrix(dataset)
```

Example ii

```
dim(dataset)
```

```
## [1] 178    3
```

```
# Assume we know Sigma
```

```
Sigma <- matrix(c(555, -170, 30, -170, 65, -10,  
                  30, -10, 2), ncol = 3)
```

```
mu_hat <- colMeans(dataset)
```

```
mu_hat
```

Example iii

```
## infant_mortality  life_expectancy      fertility
##           25.824157           71.308427      2.868933

# Test mu = mu_0
mu_0 <- c(25, 50, 3)
test_statistic <- nrow(dataset) * t(mu_hat - mu_0) %*%
  solve(Sigma) %*% (mu_hat - mu_0)

c(drop(test_statistic), qchisq(0.95, df = 3))

## [1] 7153.275387      7.814728
```

Example iv

```
drop(test_statistic) > qchisq(0.95, df = 3)
```

```
## [1] TRUE
```

Test for a multivariate mean: Σ unknown

- Of course, we rarely (if ever) know Σ , and so we use its MLE

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})^T$$

or the sample covariance S_n .

- Therefore, to test $H_0 : \mu = \mu_0$ at level α , then we reject the null hypothesis if

$$T^2 = n(\bar{\mathbf{Y}} - \mu_0)^T S_n^{-1} (\bar{\mathbf{Y}} - \mu_0) > c,$$

for a suitably chosen constant c that depends on α .

- **Note:** The test statistic T^2 is known as *Hotelling's T^2* .

Test for a multivariate mean: Σ unknown ii

- We will show that (under H_0) T^2 has a simple distribution:

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

- In other words, we reject the null hypothesis at level α if

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

Example (revisited) i

```
n <- nrow(dataset); p <- ncol(dataset)

# Test mu = mu_0
mu_0 <- c(25, 50, 3)
test_statistic <- n * t(mu_hat - mu_0) %*%
  solve(cov(dataset)) %*% (mu_hat - mu_0)

critical_val <- (n - 1)*p*qf(0.95, df1 = p,
                           df2 = n - p)/(n-p)
```

Example (revisited) ii

```
c(drop(test_statistic), critical_val)
```

```
## [1] 5121.461370      8.059773
```

```
drop(test_statistic) > critical_val
```

```
## [1] TRUE
```

Distribution of T^2

We will prove a more general result that we will also be useful for more than one multivariate mean.

Theorem

Let $\mathbf{Y} \sim N_p(0, \Sigma)$, let $m\mathbf{W} \sim W_p(m, \Sigma)$, and assume \mathbf{Y}, \mathbf{W} are independent. Define

$$T^2 = m\mathbf{Y}^T \mathbf{W}^{-1} \mathbf{Y}.$$

Then

$$\frac{m - p + 1}{mp} T^2 \sim F(p, m - p + 1),$$

where $F(\alpha, \beta)$ denotes the non-central F -distribution with α, β degrees of freedom.

- First, if we write $\Sigma = LL^T$, we can replace \mathbf{Y} by $L^{-1}\mathbf{Y}$ and W with L^TWL without changing T^2 .
 - In other words, without loss of generality, we can assume $\Sigma = I_p$.
- Now, note that since \mathbf{Y} and W are independent, the conditional distribution of mW given \mathbf{Y} is also $W_p(m, I_p)$.
- Consider \mathbf{Y} a fixed quantity, and let H be an orthogonal matrix whose first column is $\mathbf{Y}(\mathbf{Y}^T\mathbf{Y})^{-1/2}$.
 - The other columns can be chosen by finding a basis for the orthogonal complement of \mathbf{Y} and applying Gram-Schmidt to obtain an orthonormal basis.

Proof ii

- Define $V = H^T W H$. Conditional on \mathbf{Y} , this is still distributed as $\frac{1}{m} W_p(m, I_p)$.
 - This distribution does not depend on \mathbf{Y} , and therefore V and \mathbf{Y} are independent.
- Decompose V as such:

$$\begin{pmatrix} v_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix},$$

where v_{11} is a (random) scalar.

Proof iii

- By result A.2.4g of MKB (see supplementary materials), the $(1, 1)$ element of V^{-1} is given by

$$v_{11|2}^{-1} = (v_{11} - V_{12}V_{22}^{-1}V_{21})^{-1}.$$

- Moreover, note that $v_{11|2} \sim \chi^2(m - p + 1)$.
- We now have

$$\begin{aligned}\frac{1}{m}T^2 &= \mathbf{Y}^T W^{-1} \mathbf{Y} \\ &= (H\mathbf{Y})^T (H^T W H)^{-1} (H^T \mathbf{Y}) \\ &= (H\mathbf{Y})^T (V)^{-1} (H^T \mathbf{Y}) \\ &= (\mathbf{Y}^T \mathbf{Y})^{1/2} v_{11|2}^{-1} (\mathbf{Y}^T \mathbf{Y})^{1/2} \\ &= (\mathbf{Y}^T \mathbf{Y}) / v_{11|2}.\end{aligned}$$

- In other words, we have expressed $\frac{1}{m}T^2$ as a ratio of chi-squares.
- Therefore, we have

$$\begin{aligned}\frac{m-p+1}{mp}T^2 &= \left((\mathbf{Y}^T \mathbf{Y})/p \right) / \left(v_{11|2}/(m-p+1) \right) \\ &\sim F(p, m-p+1).\end{aligned}$$

□

Confidence region for μ

- Analogously to the univariate setting, it may be more informative to look at a *confidence region*:
 - The set of values $\mu_0 \in \mathbb{R}^p$ that are supported by the data, i.e. whose corresponding null hypothesis $H_0 : \mu = \mu_0$ would be rejected at level α .
- Let $c^2 = \frac{(n-1)p}{(n-p)} F_\alpha(p, n-p)$. A $100(1 - \alpha)\%$ confidence region for μ is given by the ellipsoid around $\bar{\mathbf{Y}}$ such that

$$n(\bar{\mathbf{Y}} - \mu)^T S_n^{-1} (\bar{\mathbf{Y}} - \mu) < c^2, \quad \mu \in \mathbb{R}^p.$$

Confidence region for μ ii

- We can describe the confidence region in terms of the eigendecomposition of S_n : let $\lambda_1 \geq \dots \geq \lambda_p$ be its eigenvalues, and let v_1, \dots, v_p be corresponding eigenvectors of unit length.
- The confidence region is the ellipsoid centered around $\bar{\mathbf{Y}}$ with axes

$$\pm c\sqrt{\lambda_i}v_i.$$

Visualizing confidence regions when $p > 2$

- When $p > 2$ we cannot easily plot the confidence regions.
 - Therefore, we first need to project onto an axis or onto the plane.
- **Theorem:** Let $c > 0$ be a constant and A a $p \times p$ positive definite matrix. For a given vector $\mathbf{u} \neq 0$, the projection of the ellipse $\{\mathbf{y}^T A^{-1} \mathbf{y} \leq c^2\}$ onto \mathbf{u} is given by

$$c \frac{\sqrt{\mathbf{u}^T A \mathbf{u}}}{\mathbf{u}^T \mathbf{u}} \mathbf{u}.$$

Visualizing confidence regions when $p > 2$ ii

- If we take \mathbf{u} to be the standard unit vectors, we get confidence *intervals* for each component of μ :

$$LB = \bar{\mathbf{Y}}_j - \sqrt{\frac{(n-1)p}{(n-p)} F_\alpha(p, n-p) (s_{jj}^2/n)}$$
$$UB = \bar{\mathbf{Y}}_j + \sqrt{\frac{(n-1)p}{(n-p)} F_\alpha(p, n-p) (s_{jj}^2/n)}.$$

Example

```
n <- nrow(dataset); p <- ncol(dataset)

# Test mu = mu_0
mu_0 <- c(25, 50, 3)
test_statistic <- n * t(mu_hat - mu_0) %*%
  solve(cov(dataset)) %*% (mu_hat - mu_0)

critical_val <- (n - 1)*p*qf(0.95, df1 = p,
                           df2 = n - p)/(n-p)
sample_cov <- diag(cov(dataset))

cbind(mu_hat - sqrt(critical_val*
                    sample_cov/n),
      mu_hat + sqrt(critical_val*
```

Visualizing confidence regions when $p > 2$ (cont'd) i

- **Theorem:** Let $c > 0$ be a constant and A a $p \times p$ positive definite matrix. For a given pair of perpendicular unit vectors $\mathbf{u}_1, \mathbf{u}_2$, the projection of the ellipse $\{\mathbf{y}^T A^{-1} \mathbf{y} \leq c^2\}$ onto the plane defined by $\mathbf{u}_1, \mathbf{u}_2$ is given by

$$\left\{ (U^T \mathbf{y})^T (U^T A U)^{-1} (U^T \mathbf{y}) \leq c^2 \right\},$$

where $U = (\mathbf{u}_1, \mathbf{u}_2)$.

Example (cont'd) i

```
U <- matrix(c(1, 0, 0,  
              0, 1, 0),  
            ncol = 2)  
R <- n*solve(t(U) %*% cov(dataset) %*% U)  
transf <- chol(R)
```

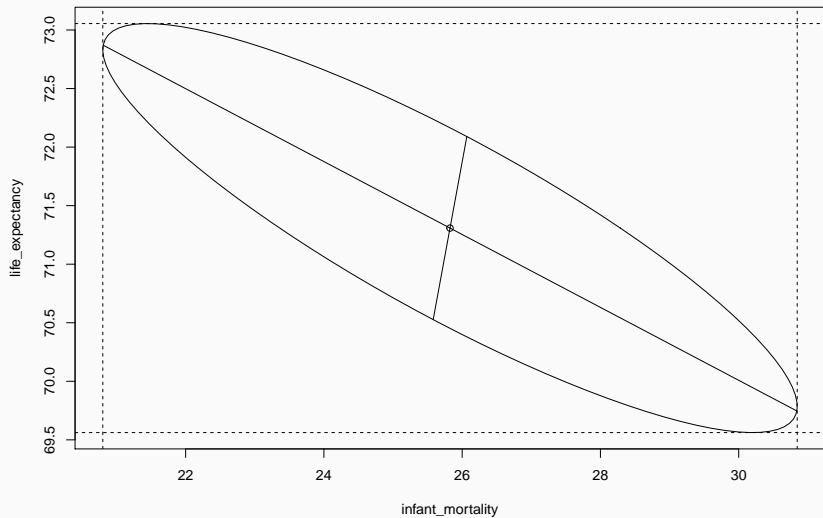
Example (cont'd) ii

```
# First create a circle of radius c
theta_vect <- seq(0, 2*pi, length.out = 100)
circle <- sqrt(critical_val) * cbind(cos(theta_vect), sin(theta_vect))
# Then turn into ellipse
ellipse <- circle %%% t(solve(transf)) +
  matrix(mu_hat[1:2], ncol = 2,
        nrow = nrow(circle),
        byrow = TRUE)
```


Example (cont'd) iii

```
# Eigendecomposition
decomp <- eigen(t(U) %*% cov(dataset) %*% U)
first <- sqrt(decomp$values[1]) *
  decomp$vectors[,1] * sqrt(critical_val)
second <- sqrt(decomp$values[2]) *
  decomp$vectors[,2] * sqrt(critical_val)
```

Example (cont'd) iv



Simultaneous Confidence Statements i

- Let $w \in \mathbb{R}^p$. We are interested in constructing confidence intervals for $w^T \mu$ that are simultaneously valid (i.e. right coverage probability) for all w .
- Note that $w^T \bar{\mathbf{Y}}$ and $w^T S_n w$ are both scalars.
- If we were only interested in a particular w , we could use the following confidence interval:

$$\left(w^T \bar{\mathbf{Y}} \pm t_{\alpha/2, n-1} \sqrt{w^T S_n w / n} \right).$$

Simultaneous Confidence Statements ii

- Or equivalently, the confidence interval contains the set of values $w^T \mu$ for which

$$t^2(w) = \frac{n(w^T \bar{\mathbf{Y}} - w^T \mu)^2}{w^T S_n w} = \frac{n(w^T (\bar{\mathbf{Y}} - \mu))^2}{w^T S_n w} \leq F_\alpha(1, n-1).$$

- **Strategy:** Maximise over all w :

$$\max_w t^2(w) = \max_w \frac{n(w^T (\bar{\mathbf{Y}} - \mu))^2}{w^T S_n w}.$$

- Using the Cauchy-Schwarz Inequality:

$$\begin{aligned}(w^T(\bar{\mathbf{Y}} - \mu))^2 &= (w^T S^{1/2} S^{-1/2}(\bar{\mathbf{Y}} - \mu))^2 \\&= ((S^{1/2} w)^T (S^{-1/2}(\bar{\mathbf{Y}} - \mu)))^2 \\&\leq (w^T S_n w)((\bar{\mathbf{Y}} - \mu)^T S_n^{-1}(\bar{\mathbf{Y}} - \mu)).\end{aligned}$$

- Dividing both sides by $w^T S_n w/n$, we get

$$t^2(w) \leq n(\bar{\mathbf{Y}} - \mu)^T S_n^{-1}(\bar{\mathbf{Y}} - \mu).$$

- Since the Cauchy-Schwarz inequality also implies that the inequality is an *equality* if and only if w is proportional to $S_n^{-1}(\bar{\mathbf{Y}} - \mu)$, it means the upper bound is attained and therefore

$$\max_w t^2(w) = n(\bar{\mathbf{Y}} - \mu)^T S_n^{-1}(\bar{\mathbf{Y}} - \mu).$$

- The right-hand side is Hotteling's T^2 , and therefore we know that

$$\max_w t^2(w) \sim \frac{(n-1)p}{(n-p)} F(p, n-p).$$

Simultaneous Confidence Statements v

- **Theorem:** Simultaneously for all $w \in \mathbb{R}^p$, the interval

$$\left(w^T \bar{\mathbf{Y}} \pm \sqrt{\frac{(n-1)p}{n(n-p)} F_\alpha(p, n-p) w^T S_n w} \right).$$

will contain $w^T \mu$ with probability $1 - \alpha$.

- **Corollary:** If we take w to be the standard basis vectors, we recover the projection results from earlier.

Further comments

- If we take $w = (0, \dots, 0, 1, 0, \dots, 0, -1, 0, \dots, 0)$, we can also derive confidence statements about mean differences $\mu_i - \mu_k$.
- In general, simultaneous confidence statements are good for exploratory analyses, i.e. when we test many different contrasts.
- However, this much generality comes at a cost: the resulting confidence intervals are quite large.
 - Since we typically only care about a finite number of hypotheses, there are more efficient ways to account for the exploratory nature of the tests.

Bonferroni correction i

- Assume that we are interested in m null hypotheses
 $H_{0i} : w_i^T \mu = \mu_{0i}$, at confidence level α_i , for $i = 1, \dots, m$.
- We can show that

$$\begin{aligned} P(\text{none of } H_{0i} \text{ are rejected}) &= 1 - P(\text{some } H_{0i} \text{ is rejected}) \\ &\geq 1 - \sum_{i=1}^m P(H_{0i} \text{ is rejected}) \\ &= 1 - \sum_{i=1}^m \alpha_i. \end{aligned}$$

Bonferroni correction ii

- Therefore, if we want to control the overall error rate at α , we can take

$$\alpha_i = \alpha/m, \quad \text{for all } i = 1, \dots, m.$$

- If we take w_i to be the i -th standard basis vector, we get simultaneous confidence intervals for all p components of μ :

$$\left(\bar{\mathbf{Y}}_i \pm t_{\alpha/2p, n-1}(\sqrt{s_{ii}^2/n}) \right).$$

Example i

```
# Let's focus on only two variables
dataset <- dataset[,c("infant_mortality",
                      "life_expectancy")]

n <- nrow(dataset); p <- ncol(dataset)
```

Example ii

```
alpha <- 0.05
mu_hat <- colMeans(dataset)
sample_cov <- diag(cov(dataset))

# Simultaneous CIs
critical_val <- (n - 1)*p*qf(1-0.5*alpha, df1 = p,
                           df2 = n - p)/(n-p)

simul_ci <- cbind(mu_hat - sqrt(critical_val*
                                sample_cov/n),
                  mu_hat + sqrt(critical_val*
                                sample_cov/n))
```

Example iii

```
# Univariate without correction
```

```
univ_ci <- cbind(mu_hat - qt(1-0.5*alpha, n - 1) *  
                  sqrt(sample_cov/n),  
                  mu_hat + qt(1-0.5*alpha, n - 1) *  
                  sqrt(sample_cov/n))
```

```
# Bonferroni adjustment
```

```
bonf_ci <- cbind(mu_hat - qt(1-0.5*alpha/p, n - 1) *  
                  sqrt(sample_cov/n),  
                  mu_hat + qt(1-0.5*alpha/p, n - 1) *  
                  sqrt(sample_cov/n))
```

```
simul_ci
```

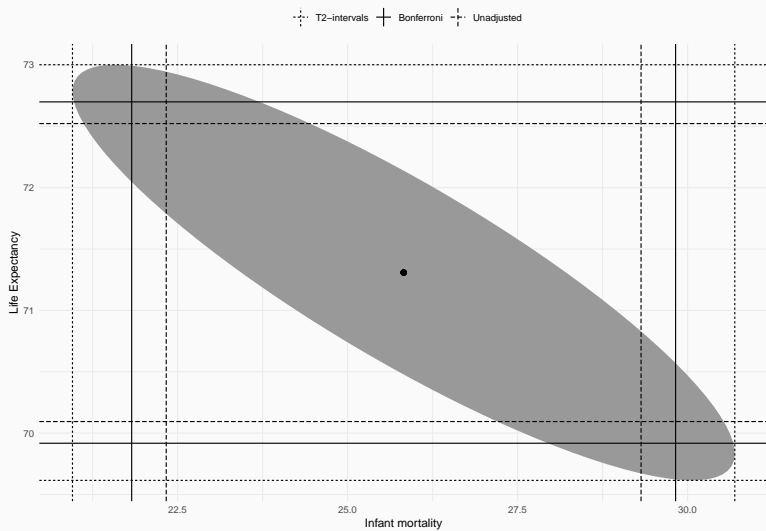
```
##                [,1]      [,2]  
## infant_mortality 20.95439 30.69392  
## life_expectancy  69.61504 73.00181
```

```
univ_ci
```

```
##                [,1]      [,2]  
## infant_mortality 22.33295 29.31537  
## life_expectancy  70.09441 72.52244
```

```
bonf_ci
```

```
##                [,1]      [,2]  
## infant_mortality 21.82491 29.8234  
## life_expectancy  69.91775 72.6991
```



Summary of confidence statements

- *So which one should you use?*
 - Use the confidence region when you're interested in a single multivariate hypothesis test.
 - Use the simultaneous (i.e. T^2) intervals when testing a large number of contrasts.
 - Use the Bonferroni correction when testing a small number of contrasts (e.g. each component of μ).
 - (Almost) **never** use the unadjusted intervals.
- We can check the coverage probabilities of each approach using a simulation study:
 - https://www.maxturgeon.ca/f19-stat4690/simulation_coverage_probability.R

Likelihood Ratio Test i

- There is another important approach to performing hypothesis testing:
 - **Likelihood Ratio Test**
- General strategy:
 - i. Maximise likelihood under the null hypothesis: L_0
 - ii. Maximise likelihood over the whole parameter space: L_1
 - iii. Since the value of the parameters under the null hypothesis is in the parameter space, we have $L_1 \geq L_0$.
 - iv. Reject the null hypothesis if the ratio $\Lambda = L_0/L_1$ is small.

Likelihood Ratio Test ii

- In our setting, recall that the likelihood is given by

$$L(\mu, \Sigma) = \prod_{i=1}^n \left(\frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp \left(-\frac{1}{2} (\mathbf{y}_i - \mu)^T \Sigma^{-1} (\mathbf{y}_i - \mu) \right) \right).$$

- Over the whole parameter space, it is maximised at

$$\hat{\mu} = \bar{\mathbf{Y}}, \quad \hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\mathbf{Y}_i - \bar{\mathbf{Y}})(\mathbf{Y}_i - \bar{\mathbf{Y}})^T.$$

- Under the null hypothesis $H_0 : \mu = \mu_0$, the only free parameter is Σ , and $L(\mu_0, \Sigma)$ is maximised at

$$\hat{\Sigma}_0 = \frac{1}{n} \sum_{i=1}^n (\mathbf{Y}_i - \mu_0)(\mathbf{Y}_i - \mu_0)^T.$$

Likelihood Ratio Test iii

- With some linear algebra, you can check that

$$L(\hat{\mu}, \hat{\Sigma}) = \frac{\exp(-np/2)}{(2\pi)^{np/2} |\hat{\Sigma}|^{n/2}}$$
$$L(\mu_0, \hat{\Sigma}_0) = \frac{\exp(-np/2)}{(2\pi)^{np/2} |\hat{\Sigma}_0|^{n/2}}.$$

- Therefore, the likelihood ratio is given by

$$\Lambda = \frac{L(\mu_0, \hat{\Sigma}_0)}{L(\hat{\mu}, \hat{\Sigma})} = \left(\frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} \right)^{n/2}.$$

- The equivalent statistic $\Lambda^{2/n} = |\hat{\Sigma}|/|\hat{\Sigma}_0|$ is called *Wilks' lambda*.

Distribution of Wilk's Lambda

- What is the sampling distribution of Λ under the null hypothesis? It turns out that

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

where T^2 is Hotelling's statistic.

- Therefore the two tests are equivalent.
- But note that $\Lambda^{2/n}$ involves computing two determinants, whereas T^2 involves inverting a matrix.

Comparing two multivariate means

Equal covariance case i

- Now let's assume we have two independent multivariate samples of (potentially) different sizes:
 - $\mathbf{Y}_{11}, \dots, \mathbf{Y}_{1n_1} \sim N_p(\mu_1, \Sigma)$
 - $\mathbf{Y}_{21}, \dots, \mathbf{Y}_{2n_2} \sim N_p(\mu_2, \Sigma)$
- We are interested in testing $\mu_1 = \mu_2$.
 - Note that we assume *equal covariance* for the time being.
- Let $\bar{\mathbf{Y}}_1, \bar{\mathbf{Y}}_2$ be their respective sample means, and let S_1, S_2 , their respective sample covariances.
- First, note that

$$\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2 \sim N_p \left(\mu_1 - \mu_2, \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \Sigma \right).$$

Equal covariance case ii

- Second, we also have that $(n_i - 1)S_i$ is an estimator for $(n_i - 1)\Sigma$, for $i = 1, 2$.
 - Therefore, we can *pool* both $(n_1 - 1)S_1$ and $(n_2 - 1)S_2$ into a single estimator for Σ :

$$S_{pool} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2}{n_1 + n_2 - 2}.$$

- Putting these two observations together, we get a test statistic for $H_0 : \mu_1 = \mu_2$:

$$T^2 = (\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2)^T \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) S_{pool} \right]^{-1} (\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2).$$

- Under the null hypothesis, we get

$$T^2 \sim \frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)} F(p, n_1 + n_2 - p - 1).$$

Example i

```
dataset1 <- gapminder %>%  
  filter(year == 2012,  
         continent == "Africa",  
         !is.na(infant_mortality)) %>%  
  select(life_expectancy, infant_mortality) %>%  
  as.matrix()  
dim(dataset1)
```

```
## [1] 51  2
```

Example ii

```
dataset2 <- gapminder %>%  
  filter(year == 2012,  
         continent == "Asia",  
         !is.na(infant_mortality)) %>%  
  select(life_expectancy, infant_mortality) %>%  
  as.matrix()  
dim(dataset2)
```

```
## [1] 45  2
```

Example iii

```
n1 <- nrow(dataset1); n2 <- nrow(dataset2)
p <- ncol(dataset1)
```

```
(mu_hat1 <- colMeans(dataset1))
```

```
##   life_expectancy infant_mortality
##           62.14314           52.32745
```

```
(mu_hat2 <- colMeans(dataset2))
```

Example iv

```
## life_expectancy infant_mortality
##           73.76667           20.84000
```

```
(S1 <- cov(dataset1))
```

```
##           life_expectancy infant_mortality
## life_expectancy      48.7241      -107.1926
## infant_mortality    -107.1926      504.2972
```

```
(S2 <- cov(dataset2))
```

Example v

```
##                life_expectancy infant_mortality
## life_expectancy      26.08727      -65.19568
## infant_mortality    -65.19568      256.40655
```

```
# Even though it doesn't look reasonable
# We will assume equal covariance for now
```

Example vi

```
mu_hat_diff <- mu_hat1 - mu_hat2

S_pool <- ((n1 - 1)*S1 + (n2 - 1)*S2)/(n1+n2-2)

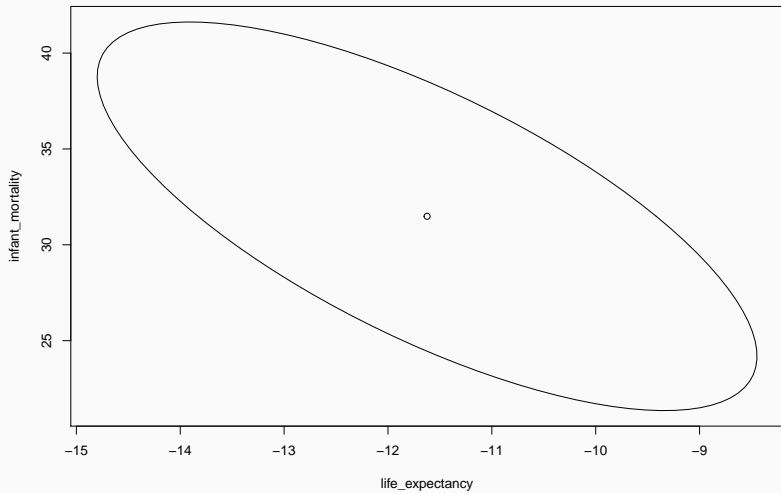
test_statistic <- t(mu_hat_diff) %*%
  solve((n1^-1 + n2^-1)*S_pool) %*% mu_hat_diff

const <- (n1 + n2 - 2)*p/(n1 + n2 - p - 2)
critical_val <- const * qf(0.95, df1 = p,
                          df2 = n1 + n2 - p - 2)

drop(test_statistic) > critical_val

## [1] TRUE
```

Comparing Africa vs. Asia



Unequal covariance case i

- Now let's turn our attention to the case where the covariance matrices are **not** equal:
 - $\mathbf{Y}_{11}, \dots, \mathbf{Y}_{1n_1} \sim N_p(\mu_1, \Sigma_1)$
 - $\mathbf{Y}_{21}, \dots, \mathbf{Y}_{2n_2} \sim N_p(\mu_2, \Sigma_2)$
- Recall that in the univariate case, the test statistic that is typically used is called *Welch's t-statistic*.
 - The general idea is to adjust the degrees of freedom of the *t*-distribution.
 - **Note:** This is actually the default test used by `t.test`!
- Unfortunately, there is no single best approximation in the multivariate case.

Unequal covariance case ii

- First, observe that we have

$$\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2 \sim N_p \left(\mu_1 - \mu_2, \frac{1}{n_1} \Sigma_1 + \frac{1}{n_2} \Sigma_2 \right).$$

- Therefore, under $H_0 : \mu_1 = \mu_2$, we have

$$(\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2)^T \left(\frac{1}{n_1} \Sigma_1 + \frac{1}{n_2} \Sigma_2 \right)^{-1} (\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2) \sim \chi^2(p).$$

- Since S_i converges to Σ_i as $n_i \rightarrow \infty$, we can use Slutsky's theorem to argue that if both $n_1 - p$ and $n_2 - p$ are “large”, then

$$T^2 = (\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2)^T \left(\frac{1}{n_1} S_1 + \frac{1}{n_2} S_2 \right)^{-1} (\bar{\mathbf{Y}}_1 - \bar{\mathbf{Y}}_2) \approx \chi^2(p).$$

Unequal covariance case iii

- Unfortunately, the definition of “large” in this case depends on how different Σ_1 and Σ_2 are.
- Alternatives:
 - Use one of the many approximations to the null distribution of T^2 (e.g. see Timm (2002), Section 3.9; Rencher (1998), Section 3.9.2).
 - Use a resampling technique (e.g. bootstrap or permutation test).
 - Use Welch’s t-statistic for each component of $\mu_1 - \mu_2$ with a Bonferroni correction for the significance level.

Nel & van der Merwe Approximation

- First, define

$$W_i = \frac{1}{n_i} S_i \left(\frac{1}{n_1} S_1 + \frac{1}{n_2} S_2 \right)^{-1}.$$

- Then let

$$\nu = \frac{p + p^2}{\sum_{i=1}^2 \frac{1}{n_i} (\text{tr}(W_i^2) + \text{tr}(W_i)^2)}.$$

- One can show that $\min(n_1, n_2) \leq \nu \leq n_1 + n_2$.
- Under the null hypothesis, we approximately have

$$T^2 \approx \frac{\nu p}{\nu - p + 1} F(p, \nu - p + 1).$$

Example (cont'd) i

```
test_statistic <- t(mu_hat_diff) %*%  
  solve(n1^-1*S1 + n2^-1*S2) %*% mu_hat_diff  
  
critical_val <- qchisq(0.95, df = p)  
  
drop(test_statistic) > critical_val  
  
## [1] TRUE
```

Example (cont'd) ii

```
W1 <- S1 %*% solve(n1^-1*S1 + n2^-1*S2)/n1
W2 <- S2 %*% solve(n1^-1*S1 + n2^-1*S2)/n2

trace_square <- sum(diag(W1%*%W1))/n1 +
  sum(diag(W2%*%W2))/n2
square_trace <- sum(diag(W1))^2/n1 +
  sum(diag(W2))^2/n2

(nu <- (p + p^2)/(trace_square + square_trace))

## [1] 88.85241
```

Example (cont'd) iii

```
const <- nu*p/(nu - p - 1)
critical_val <- const * qf(0.95, df1 = p,
                           df2 = nu - p - 1)

drop(test_statistic) > critical_val

## [1] TRUE
```

Comparing Africa vs. Asia

