Tests for Multivariate Means II

Max Turgeon

STAT 4690–Applied Multivariate Analysis

Repeated Mesures Design

Contrast matrices

- A *contrast* is a linear combination θ of variables such that its coefficients sum to zero.
  - E.g. $\theta = (1, -1, 0)$ or $\theta = (2, -1, -1)$.
- A *contrast matrix* is a matrix C whose rows are contrasts (so the row-sums are zero) and are linearly independent.
  - E.g. $C = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}$.
- As their name suggests, contrasts and contrast matrices are used to contrast (or compare) different combinations of variables.

Testing Structural Relations

- Let C be a $q \times p$ contrast matrix, and let $\bar{\mathbf{Y}}$ be the (p-dimensional) sample mean and S_n, the $(p \times p)$ sample covariance.
- We can test the null hypothesis $H_0 : C\mu = 0$ using Hotelling's T^2:

$$T^2 = n(C\bar{\mathbf{Y}})^T(CS_nC^T)^{-1}(C\bar{\mathbf{Y}}).$$

- **What is the sampling distribution?** $C\bar{\mathbf{Y}}$ is q-dimensional and CS_nC^T is $q \times q$, therefore

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

Repeated Measurements i

- Suppose that our random sample $\mathbf{Y}_1, \ldots, \mathbf{Y}_n \sim N_p(\mu, \Sigma)$ be such that each component of $\mathbf{Y}_i$ represent a repeated measurement on the same experimental unit.
    - E.g. Grades on different tests, blood pressure measurements at different doctor visits.
- **Question**: Is there any evidence that the means differ between the measurements?
    - Or in other words: are all components of μ equal?

Repeated Measurements ii

- Consider the following $(p-1) \times p$ contrast matrix:

$$C = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}.$$

- We thus have

$$C\mu = \begin{pmatrix} \mu_1 - \mu_2 \\ \mu_1 - \mu_3 \\ \vdots \\ \mu_1 - \mu_p \end{pmatrix}.$$

Repeated Measurements iii

- To test the null hypothesis $H_0 : C\mu = 0$, we use T^2 as above:

$$T^2 = n(C\bar{\mathbf{Y}})^T(CS_nC^T)^{-1}(C\bar{\mathbf{Y}}),$$

where

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

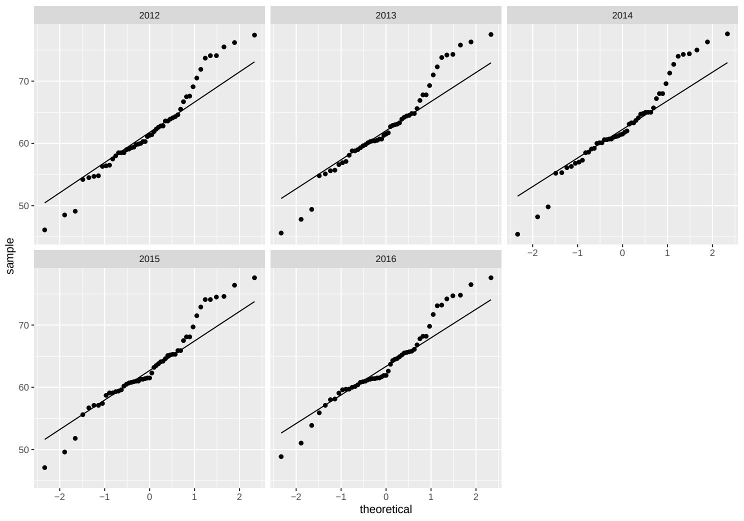
Example i

```r
library(tidyverse)
library(dslabs)

dataset <- gapminder %>%
  filter(year %in% 2012:2016,
         continent == "Africa") %>%
  select(year, country, life_expectancy)
```

Example ii

```r
# QQ-plots to assess normality
dataset %>%
  ggplot(aes(sample = life_expectancy)) +
  stat_qq() + stat_qq_line() +
  facet_wrap(~year)
```

Example iii



Example iv

```r
C <- matrix(c(1, -1, 0, 0, 0,
              1, 0, -1, 0, 0,
              1, 0, 0, -1, 0,
              1, 0, 0, 0, -1),
            ncol = 5, byrow = TRUE)
C
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    1   -1    0    0    0
## [2,]    1    0   -1    0    0
## [3,]    1    0    0   -1    0
```

Example v

```
## [4,]    1    0    0    0   -1
```

```r
# Transform data into wide format
dataset <- dataset %>%
  spread(year, life_expectancy)

head(dataset)
```

Example vi

```
##        country 2012 2013 2014 2015  2016
## 1      Algeria 76.2 76.3 76.3 76.4 76.50
## 2       Angola 58.5 58.8 59.2 59.6 60.00
## 3        Benin 61.4 61.7 62.0 62.3 62.60
## 4     Botswana 56.5 56.9 57.3 58.7 60.13
## 5 Burkina Faso 59.9 60.3 60.6 60.9 61.20
## 6      Burundi 61.1 61.3 61.4 61.4 61.40
```

Example vii

```r
# Compute test statistic
dataset <- dataset %>%
  select(-country) %>%
  as.matrix()
n <- nrow(dataset); p <- ncol(dataset)

mu_hat <- colMeans(dataset)
mu_hat
```

```
##     2012     2013     2014     2015     2016
## 62.14314 62.54510 62.77843 63.27843 63.78843
```

Example viii

```r
Sn <- cov(dataset)
test_statistic <- n * t(C %*% mu_hat) %*%
  solve(C %*% Sn %*% t(C)) %*% (C %*% mu_hat)

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

drop(test_statistic) > critical_val
```

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

Other contrast matrices i

- What about other contrast matrices of the same size? For example:

$$\tilde{C} = \begin{pmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}.$$

- Do we get the same inference results? **YES**
- Let $C, \tilde{C}$ be two $(p-1) \times p$ contrast matrices.

Other contrast matrices ii

- Since their rows are independent, there exists an invertible $(p-1) \times (p-1)$ matrix B such that $\tilde{C} = BC$.

$$
\begin{aligned}
(\tilde{C}\bar{\mathbf{Y}})^T(\tilde{C}S_n\tilde{C}^T)^{-1}(\tilde{C}\bar{\mathbf{Y}}) &= (BC\bar{\mathbf{Y}})^T(BCS_nC^TB^T)^{-1}(BC\bar{\mathbf{Y}}) \\
&= (C\bar{\mathbf{Y}})^TB^T(BCS_nC^TB^T)^{-1}B(C\bar{\mathbf{Y}}) \\
&= (C\bar{\mathbf{Y}})^T(CS_nC^T)^{-1}(C\bar{\mathbf{Y}})
\end{aligned}
$$

- In other words, we get the same test statistic whether we use C or $\tilde{C}$.

Confidence regions and Confidence Intervals

- As discussed earlier, we can use T^2 to create a confidence region around $C\bar{\mathbf{Y}}$:

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

- We can also construct T^2 intervals for any contrast θ:

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

- Or we can construct Bonferroni-adjusted confidence intervals for each row c_i of C:

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

Example (cont'd) i

```
alpha <- 0.05
mu_contr <- C %*% mu_hat
sample_cov <- diag(C %*% Sn %*% t(C))

mu_contr
```

```
##             [,1]
## [1,] -0.4019608
## [2,] -0.6352941
## [3,] -1.1352941
## [4,] -1.6452941
```

Example (cont'd) ii

```
# Simultaneous CIs
simul_ci <- cbind(mu_contr - sqrt(critical_val*
                                    sample_cov/n),
                  mu_contr + sqrt(critical_val*
                                    sample_cov/n))
```

Example (cont'd) iii

```r
# Bonferroni adjustment
bonf_ci <- cbind(mu_contr - qt(1-0.5*alpha/(p-1),
                               n - 1) *
                   sqrt(sample_cov/n),
                 mu_contr + qt(1-0.5*alpha/(p-1),
                               n - 1) *
                   sqrt(sample_cov/n))
```

```r
simul_ci
```

```
##               [,1]       [,2]
## [1,] -0.5902699 -0.2136517
## [2,] -0.9641199 -0.3064684
## [3,] -1.5762989 -0.6942893
## [4,] -2.3083908 -0.9821975
```

```r
bonf_ci
```

```
##               [,1]       [,2]
## [1,] -0.5495288 -0.2543928
## [2,] -0.8929777 -0.3776105
## [3,] -1.4808865 -0.7897017
## [4,] -2.1649283 -1.1256599
```

Comments

- The test above is best used when we cannot make any assumptions about the covariance structure Σ.
- When we assume Σ has a special structure, it is possible to build more powerful tests.
  - E.g. If the repeated measurements are taken over time, it may be reasonable to assume an autoregressive structure.
- Similarly, if we are interested in a specific relationship between the components of μ, it is possible to build more powerful tests.
  - E.g. Linear relationship between the components when measurements are taken over time.

Comparing two multivariate means

Equal covariance case i

- Now let's assume we have *two* independent multivariate samples of (potentially) different sizes:
  - $\mathbf{Y}_{11}, \ldots, \mathbf{Y}_{1n_1} \sim N_p(\mu_1, \Sigma)$
  - $\mathbf{Y}_{21}, \ldots, \mathbf{Y}_{2n_1} \sim N_p(\mu_2, \Sigma)$
- 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.

Equal covariance case ii

- First, note that

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

- 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:

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

Equal covariance case iii

- 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

```r
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

```r
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

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

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

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

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

Example iv

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

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

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

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

Example v

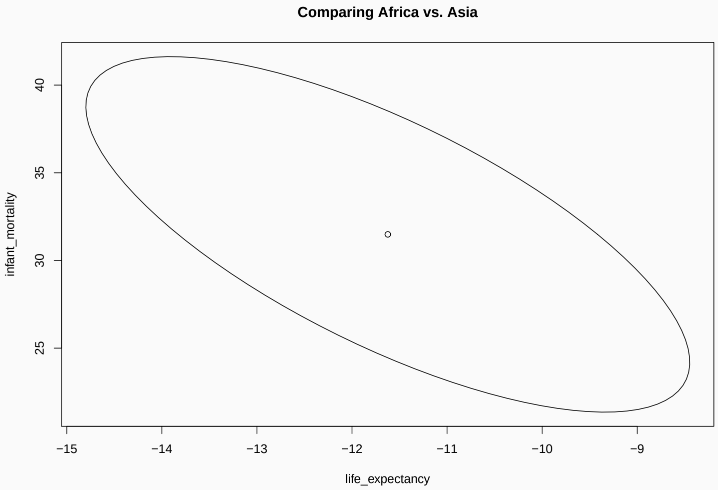
```
##                  life_expectancy infant_mortality
## life_expectancy         26.08727        -65.19568
## infant_mortality       -65.19568        256.40655
```

```
# 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
40
35
30
25
infant_mortality
−15
−14
−13
−12
−11
−10
−9
life_expectancy

Unequal covariance case i

- Now let's turn our attention to the case where the covariance matrices are **not** equal:
    - $\mathbf{Y}_{11}, \ldots, \mathbf{Y}_{1n_1} \sim N_p(\mu_1, \Sigma_1)$
    - $\mathbf{Y}_{21}, \ldots, \mathbf{Y}_{2n_1} \sim N_p(\mu_2, \Sigma_2)$
- 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 $n_i \to \infty$, 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

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

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

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

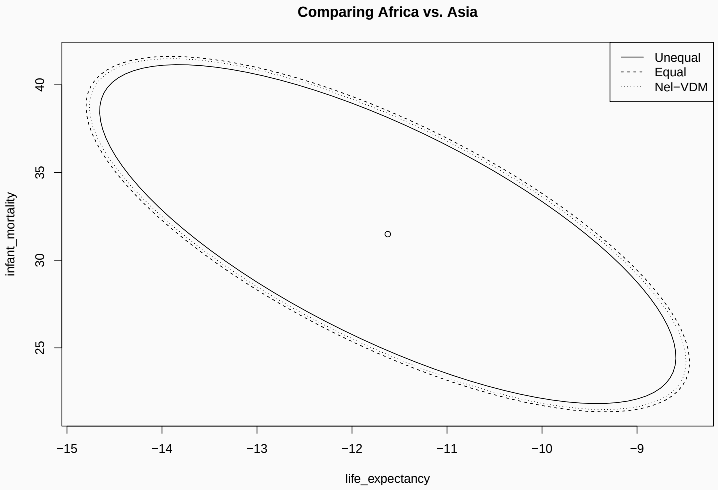
Example (cont'd) i

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

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

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

drop(test_statistic) > critical_val
```


Comparing Africa vs. Asia
Unequal
Equal
Nel−VDM
infant_mortality
40
35
30
25
−15
−14
−13
−12
−11
−10
−9
life_expectancy

Robustness

- To perform the tests on means, we made two main assumptions (listed in order of **importance**):
    1. Independence of the observations;
    2. Normality of the observations.
- Independence is the most important assumption:
    - Departure from independence can introduce significant bias and will impact the coverage probability.
- Normality is not as important:
    - Both tests for one or two means are relatively robust to heavy tail distributions.
    - Test for one mean can be sensitive to skewed distributions; test for two means is more robust.