Tests for Multivariate Means

Max Turgeon

STAT 4690–Applied Multivariate Analysis

Tests for one multivariate mean

Review of univariate tests i

- Let $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ be independently distributed, and let $\bar{X}$ and s^2 be the sample mean and variance, respectively.
- **When σ^2 is known**
  - $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, or equivalently $\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi^2(1)$.
  - $100(1-\alpha)\%$ confidence interval: $(\bar{X} - z_{\alpha/2}(\sigma/\sqrt{n}), \bar{X} + z_{\alpha/2}(\sigma/\sqrt{n}))$.
- **When σ^2 is unknown**
  - $\frac{\bar{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$, or equivalently $\left(\frac{\bar{X}-\mu}{s/\sqrt{n}}\right)^2 \sim F(1, n-1)$.
  - $100(1-\alpha)\%$ confidence interval: $(\bar{X} - t_{\alpha/2,n-1}(s/\sqrt{n}), \bar{X} + t_{\alpha/2,n-1}(s/\sqrt{n}))$.

Review of univariate tests ii

- In particular, if we want to test $H_0 : \mu = \mu_0$ when σ^2 is unknown, then we reject the null hypothesis if

$$\left|\frac{\bar{X} - \mu_0}{s/\sqrt{n}}\right| > t_{\alpha/2,n-1}, \text{ or } \left(\frac{\bar{X} - \mu_0}{s/\sqrt{n}}\right)^2 > F_\alpha(1, n-1).$$

The multivariate tests for a single mean vector have direct analogues.

Test for a multivariate mean: Σ known

- Let $\mathbf{Y}_1, \ldots, \mathbf{Y}_n \sim N_p(\mu, \Sigma)$ be independent.
- We saw in the 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

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

Example i

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

dataset <- gapminder %>%
  filter(year == 2012,
         !is.na(infant_mortality)) %>%
  select(infant_mortality,
         life_expectancy,
         fertility) %>%
  as.matrix()
```

Example ii

```r
# Assume we know Sigma
Sigma <- matrix(c(555, -170, 30, -170, 65, -10,
                  30, -10, 2), ncol = 3)

mu_hat <- colMeans(dataset)
mu_hat
```

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

Example iii

```r
# 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)

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

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

Test for a multivariate mean: Σ unknown i

- 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

- It turns out 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)

```r
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)

drop(test_statistic) > critical_val
```

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

Confidence region for μ i

- 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 $\lambda_1 \geq \cdots \geq \lambda_p$ be its eigenvalues, and let $v_1, \ldots, v_p$ 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 $p > 2$ i

- 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

```r
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
$$\left\{(U^T\mathbf{y})^T(U^TAU)^{-1}(U^T\mathbf{y}) \leq c^2\right\},$$
where $U = (\mathbf{u}_1, \mathbf{u}_2)$.

Example (cont'd) i

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

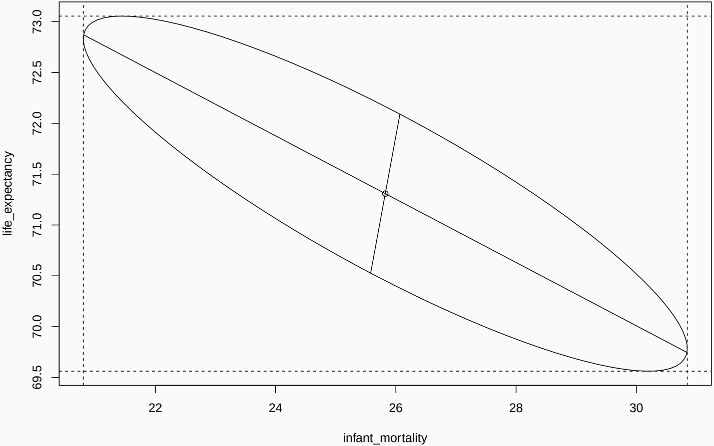
Example (cont'd) ii

```
# First create a circle of radius c
theta_vect <- seq(0, 2*pi, length.out = 100)
circle <- sqrt(critical_val) * cbind(cos(theta_vect),
# 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^TS_nw are both scalars.
- If we were only interested in a particular w, we could use the following confidence interval:

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

Simultaneous Confidence Statements ii

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

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

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

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

Simultaneous Confidence Statements iii

- 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

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

Simultaneous Confidence Statements iv

- 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 $1 - \alpha$.
- **Corrolary**: If we take w to be the standard basis vectors, we recover the projection results from earlier.

Further comments

- If we take $w = (0, \ldots, 0, 1, 0, \ldots, 0, -1, 0, \ldots, 0)$, 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

$$\begin{aligned} P(\text{all } H_{0i} \text{ are true}) &= 1 - P(\text{at least one } H_{0i} \text{ is false}) \\ &\geq 1 - \sum_{i=1}^{m} P(H_{0i} \text{ is false}) \\ &= 1 - \sum_{i=1}^{m} \alpha_i. \end{aligned}$$

Bonferroni correction ii

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

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

- 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 <- gapminder %>%
  filter(year == 2012,
         !is.na(infant_mortality)) %>%
  select(infant_mortality,
         life_expectancy) %>%
  as.matrix()

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

Example ii

```r
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

```r
# 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))
```

```r
# 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))
```

```r
simul_ci
```

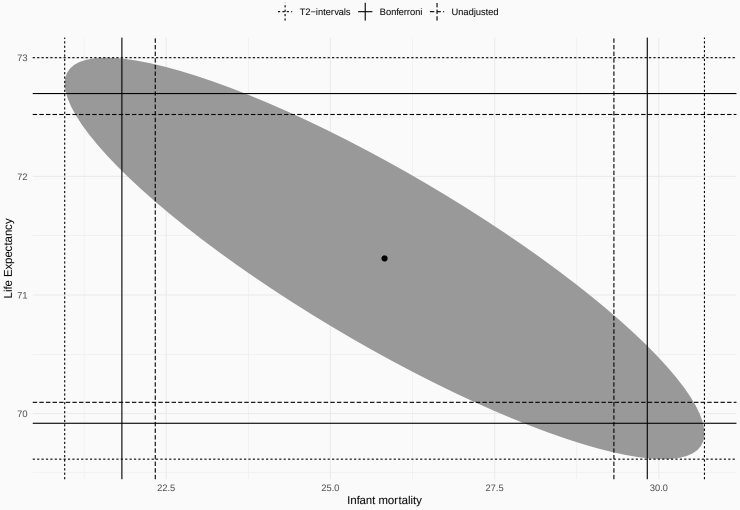
```
##                         [,1]     [,2]
## infant_mortality 20.95439 30.69392
## life_expectancy  69.61504 73.00181
```

```r
univ_ci
```

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

```r
bonf_ci
```

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


T2−intervals
Bonferroni
Unadjusted
73
72
71
70
Life Expectancy
22.5
25.0
27.5
30.0
Infant mortality

Summary

- *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:
  1. Maximise likelihood under the null hypothesis: L_0
  2. Maximise likelihood over the whole parameter space: L_1
  3. Since the value of the parameters under the null hypothesis is in the parameter space, we have $L_1 \geq L_0$.
  4. 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 algbera, you can check that

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

- Therefore, the likelihood ratio is given by

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

Likelihood Ratio Test iv

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