Multivariate Analysis of Variance

Max Turgeon

STAT 7200–Multivariate Statistics

Objectives

- Introduce MANOVA as a generalization of Hotelling's T^2
- Present the four classical test statistics
- Discuss approximations for their null distribution

Quick Overview

What do we mean by Analysis of Variance?

- ANOVA is a collection of statistical models that aim to analyze and understand the differences in means between different subgroups of the data.
    - As such, it can be seen as a generalisation of the t-test (or of Hotelling's T^2).
    - Note that there could be multiple, overlapping ways of defining the subgroups (e.g multiway ANOVA)
- It also provides a framework for hypothesis testing.
    - Which can be recovered from a suitable regression model.
- **Most importantly**, ANOVA provides a framework for understanding and comparing the various sources of variation in the data.

Review of univariate ANOVA i

- Assume the data comes from g populations:

$$\begin{matrix} X_{11}, & \ldots, & X_{1n_1} \\ \vdots & \ddots & \vdots \\ X_{g1}, & \ldots, & X_{gn_g} \end{matrix}$$

- Assume that $X_{\ell 1}, \ldots, X_{\ell n_\ell}$ is a random sample from $N(\mu_\ell, \sigma^2)$, for $\ell = 1, \ldots, g$.
    - **Homoscedasticity**
- We are interested in testing the hypothesis that $\mu_1 = \ldots = \mu_g$.

Review of univariate ANOVA ii

- *Reparametrisation*: We will write the mean $\mu_\ell = \mu + \tau_\ell$ as a sum of an overall component μ (i.e. shared by all populations) and a population-specific component τ_ℓ.
    - Our hypothesis can now be rewritten as $\tau_\ell = 0$, for all ℓ.
    - We can write our observations as

      $$X_{\ell i} = \mu + \tau_\ell + \varepsilon_{\ell i},$$

      where $\varepsilon_{\ell i} \sim N(0, \sigma^2)$.
    - **Identifiability**: We need to assume $\sum_{\ell=1}^g \tau_\ell = 0$, otherwise there are infinitely many models that lead to the same data-generating mechanism.
- *Sample statistics*: Set $n = \sum_{\ell=1}^g n_\ell$.

Review of univariate ANOVA iii

- Overall sample mean: $\bar{X} = \frac{1}{n}\sum_{\ell=1}^{g}\sum_{i=1}^{n_\ell} X_{\ell i}$.
- Population-specific sample mean: $\bar{X}_\ell = \frac{1}{n_\ell}\sum_{i=1}^{n_\ell} X_{\ell i}$.

- We get the following decomposition:

$$\left(X_{\ell i} - \bar{X}\right) = \left(\bar{X}_\ell - \bar{X}\right) + \left(X_{\ell i} - \bar{X}_\ell\right).$$

- Squaring the left-hand side and summing over both ℓ and i, we get

$$\sum_{\ell=1}^{g}\sum_{i=1}^{n_\ell}\left(X_{\ell i} - \bar{X}\right)^2 = \sum_{\ell=1}^{g} n_\ell \left(\bar{X}_\ell - \bar{X}\right)^2 + \sum_{\ell=1}^{g}\sum_{i=1}^{n_\ell}\left(X_{\ell i} - \bar{X}_\ell\right)^2.$$

- This is typically summarised as $SS_T = SS_M + SS_R$:
  - The **total sum of squares**: $SS_T = \sum_{\ell=1}^{g}\sum_{i=1}^{n_\ell}\left(X_{\ell i} - \bar{X}\right)^2$

Review of univariate ANOVA iv

- The **model** (or treatment) **sum of squares**: $SS_M = \sum_{\ell=1}^g n_\ell \left(\bar{X}_\ell - \bar{X}\right)^2$
- The **residual sum of squares**: $SS_R = \sum_{\ell=1}^g \sum_{i=1}^{n_\ell} \left(X_{\ell i} - \bar{X}_\ell\right)^2$

- Yet another representation is the *ANOVA table*:

Source of Variation	Sum of Squares	Degrees of freedom
Model	SS_M	$g-1$
Residual	SS_R	$n-g$
Total	SS_T	$n-1$

Review of univariate ANOVA v

- The usual test statistic used for testing $\tau_\ell = 0$ for all ℓ is

$$F = \frac{SS_M/(g-1)}{SS_R/(n-g)} \sim F(g-1, n-g).$$

- We could also instead reject the null hypothesis for *small* values of

$$\frac{SS_R}{SS_R + SS_M} = \frac{SS_R}{SS_T}.$$

This is the test statistic that we will generalize to the multivariate setting.

Multivariate ANOVA i

- The setting is similar: Assume the data comes from g populations:

$$\begin{array}{ccc} \mathbf{Y}_{11}, & \dots, & \mathbf{Y}_{1n_1} \\ \vdots & \ddots & \vdots \\ \mathbf{Y}_{g1}, & \dots, & \mathbf{Y}_{gn_g} \end{array}$$

- Assume that $\mathbf{Y}_{\ell 1}, \dots, \mathbf{Y}_{\ell n_\ell}$ is a random sample from $N_p(\mu_\ell, \Sigma)$, for $\ell = 1, \dots, g$.
    - **Homoscedasticity** is key here again.
- We are again interested in testing the hypothesis that $\mu_1 = \ldots = \mu_g$.
- *Reparametrisation*: We will write the mean as $\mu_\ell = \mu + \tau_\ell$

Multivariate ANOVA ii

- $\mathbf{Y}_{\ell i} = \mu + \tau_\ell + \mathbf{E}_{\ell i}$, where $\mathbf{E}_{\ell i} \sim N_p(0, \Sigma)$.
- **Identifiability**: We need to assume $\sum_{\ell=1}^{g} \tau_\ell = 0$.
- Instead of a decomposition of the sum of squares, we get a decomposition of the outer product:

$$(\mathbf{Y}_{\ell i} - \bar{\mathbf{Y}})(\mathbf{Y}_{\ell i} - \bar{\mathbf{Y}})^T.$$

- The decomposition is given as

$$\sum_{\ell=1}^{g}\sum_{i=1}^{n_\ell}(\mathbf{Y}_{\ell i} - \bar{\mathbf{Y}})(\mathbf{Y}_{\ell i} - \bar{\mathbf{Y}})^T = \sum_{\ell=1}^{g} n_\ell(\bar{\mathbf{Y}}_\ell - \bar{\mathbf{Y}})(\bar{\mathbf{Y}}_\ell - \bar{\mathbf{Y}})^T + \sum_{\ell=1}^{g}\sum_{i=1}^{n_\ell}(\mathbf{Y}_{\ell i} - \bar{\mathbf{Y}}_\ell)(\mathbf{Y}_{\ell i} - \bar{\mathbf{Y}}_\ell)^T.$$

Multivariate ANOVA iii

- Between sum of squares and cross products matrix: $B = \sum_{\ell=1}^{g} n_\ell(\bar{\mathbf{Y}}_\ell - \bar{\mathbf{Y}})(\bar{\mathbf{Y}}_\ell - \bar{\mathbf{Y}})^T$.
- Within sum of squares and cross products matrix: $W = \sum_{\ell=1}^{g} \sum_{i=1}^{n_\ell}(\mathbf{Y}_{\ell i} - \bar{\mathbf{Y}}_\ell)(\mathbf{Y}_{\ell i} - \bar{\mathbf{Y}}_\ell)^T$.
- Note that $W = \sum_{\ell=1}^{g}(n_\ell - 1)S_\ell$, and therefore $W_p(n-g, \Sigma)$.
- Moreover, using Cochran's theorem, we can show that W and B are independent, and that under the null hypothesis that $\tau_\ell = 0$ for all $\ell = 1, \ldots, g$, we also have

$$B \sim W_p(g-1, \Sigma).$$

Multivariate ANOVA iv

- Similarly as above, we have a *MANOVA table*:

Source of Variation	Sum of Squares	Degrees of freedom
Model	B	$g-1$
Residual	W	$n-g$
Total	$B+W$	$n-1$

Likelihood Ratio Test i

- To test the null hypothesis $H_0 : \tau_\ell = 0$ for all $\ell = 1, \ldots, g$, we will use *Wilk's lambda* as our test statistic:

$$\Lambda^{2/n} = \frac{|W|}{|B + W|}.$$

- As the notation suggests, this is the *likelihood ratio test statistic*.
- Under the unrestricted model (i.e. no constraint on the means), each mean parameter is maximised independently, and the maximum likelihood estimator for the covariance matrix is the pooled covariance:

$$\hat{\mu}_\ell = \bar{\mathbf{Y}}_\ell, \qquad \hat{\Sigma} = \frac{1}{n}W.$$

Likelihood Ratio Test ii

- Under the null model (i.e. all means are equal), all observations $\mathbf{Y}_{\ell i}$ come from a unique distribution $N_p(\mu, \Sigma)$, and so the maximum likelihood estimators are

$$\hat{\mu} = \bar{\mathbf{Y}}, \qquad \hat{\Sigma} = \frac{1}{n}(B + W).$$

- Putting this together, we get

$$\begin{aligned}
\Lambda &= \frac{(2\pi)^{-np/2}\exp(-np/2)|\frac{1}{n}(B+W)|^{-n/2}}{(2\pi)^{-np/2}\exp(-np/2)|\frac{1}{n}W|^{-n/2}} \\
&= \frac{|\frac{1}{n}(B+W)|^{-n/2}}{|\frac{1}{n}W|^{-n/2}} \\
&= \left(\frac{|W|}{|B+W|}\right)^{n/2}.
\end{aligned}$$

Likelihood Ratio Test iii

- From the general asymptotic theory, we now that

$$-2\log\Lambda \approx \chi^2((g-1)p).$$

- Using Bartlett's approximation, we can get an unbiased test:

$$-\left(n-1-\frac{1}{2}(p+g)\right)\log\Lambda \approx \chi^2((g-1)p).$$

- In particular, if we let $c = \chi^2_\alpha((n-1)p)$ be the critical value, we reject the null hypothesis if

$$\Lambda \leq \exp\left(\frac{-c}{n-1-0.5(p+g)}\right).$$

Example i

```r
## Example on producing plastic film
## from Krzanowski (1998, p. 381)
tear <- c(6.5, 6.2, 5.8, 6.5, 6.5, 6.9, 7.2,
          6.9, 6.1, 6.3, 6.7, 6.6, 7.2, 7.1,
          6.8, 7.1, 7.0, 7.2, 7.5, 7.6)
gloss <- c(9.5, 9.9, 9.6, 9.6, 9.2, 9.1, 10.0,
           9.9, 9.5, 9.4, 9.1, 9.3, 8.3, 8.4,
           8.5, 9.2, 8.8, 9.7, 10.1, 9.2)
opacity <- c(4.4, 6.4, 3.0, 4.1, 0.8, 5.7, 2.0,
             3.9, 1.9, 5.7, 2.8, 4.1, 3.8, 1.6,
             3.4, 8.4, 5.2, 6.9, 2.7, 1.9)
```

Example ii

```
Y <- cbind(tear, gloss, opacity)
Y_low <- Y[1:10,]
Y_high <- Y[11:20,]
n <- nrow(Y); p <- ncol(Y); g <- 2

W <- (nrow(Y_low) - 1)*cov(Y_low) +
  (nrow(Y_high) - 1)*cov(Y_high)
B <- (n-1)*cov(Y) - W
(Lambda <- det(W)/det(W+B))
```

```
## [1] 0.4136192
```

Example iii

```r
transf_lambda <- -(n - 1 - 0.5*(p + g))*log(Lambda)
transf_lambda > qchisq(0.95, p*(g-1))
```

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

```r
# Or if you want a p-value
pchisq(transf_lambda, p*(g-1), lower.tail = FALSE)
```

```
## [1] 0.002227356
```

Example iv

```
# R has a function for MANOVA
# But first, create factor variable
rate <- gl(g, 10, labels = c("Low", "High"))

fit <- manova(Y ~ rate)
summary_tbl <- broom::tidy(fit, test = "Wilks")
# Or you can use the summary function
```

```
knitr::kable(summary_tbl, digits = 3)
```

Example v

term	df	wilks	statistic	num.df	den.df	p.value
rate	1	0.414	7.561	3	16	0.002
Residuals	18	-	-	-	-	-

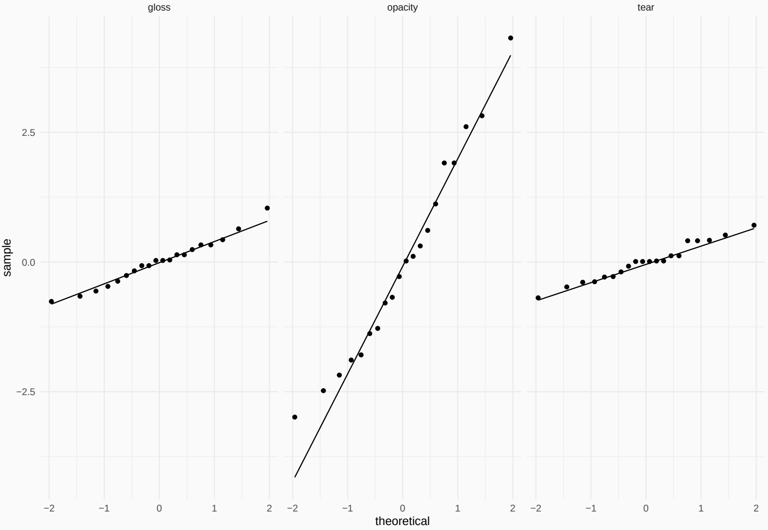
Example vi

```r
# Check residuals for evidence of normality
library(tidyverse)
resids <- residuals(fit)

data_plot <- gather(as.data.frame(resids),
                    variable, residual)

ggplot(data_plot, aes(sample = residual)) +
  stat_qq() + stat_qq_line() +
  facet_grid(. ~ variable) +
  theme_minimal()
```

Example vii

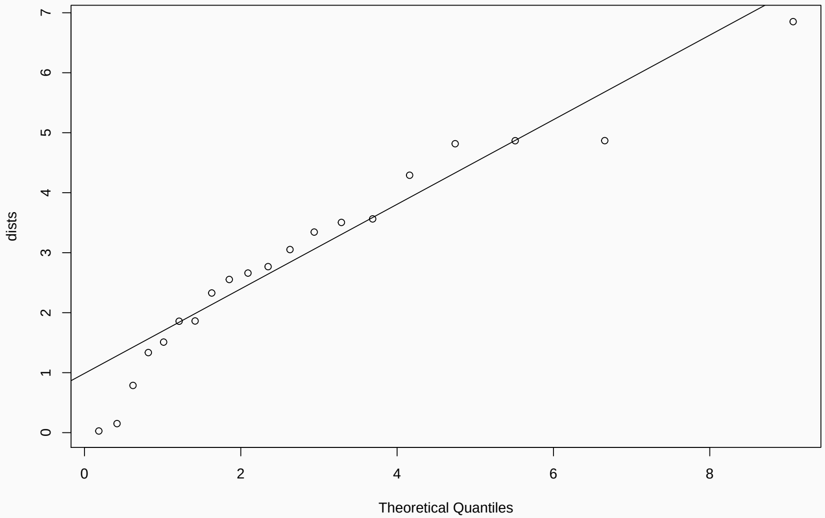


Example viii

```r
# Next: Chi-squared plot
Sn <- cov(resids)
dists <- mahalanobis(resids, colMeans(resids), Sn)
df <- mean(dists)

qqplot(qchisq(ppoints(dists), df = df),
       dists, xlab = "Theoretical Quantiles")
qqline(dists, distribution = function(p) {
  qchisq(p, df = df)
  })
```

Example ix



Comments i

- The output from R shows a different approximation to the Wilk's lambda distribution, due to Rao.
- There are actually 4 tests available in R:
  - Wilk's lambda;
  - Pillai-Bartlett;
  - Hotelling-Lawley;
  - Roy's Largest Root.

- Since we only had two groups in the above example, we were only comparing two means.
    - Wilk's lambda was therefore equivalent to Hotelling's T^2.
    - But of course MANOVA is much more general.
- We can assess the normality assumption by looking at the residuals $\mathbf{E}_{\ell i} = \mathbf{Y}_{\ell i} - \bar{\mathbf{Y}}_{\ell}$.

- The Wilks' lambda statistic can be expressed in terms of the eigenvalues $\lambda_1, \ldots, \lambda_s$ of the matrix $W^{-1}B$, where $s = \min(p, g-1)$:

$$\Lambda^{2/n} = \prod_{i=1}^{s} \frac{1}{1+\lambda_i}.$$

Other MANOVA Test Statistics ii

- The four classical multivariate test statistics are:

$$\begin{aligned} \text{Wilks' lambda} &: \prod_{i=1}^{s} \frac{1}{1+\lambda_i} = \frac{|W|}{|B+W|} \\ \text{Pillai's trace} &: \sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i} = \operatorname{tr}\left(B(B+W)^{-1}\right) \\ \text{Hotelling-Lawley trace} &: \sum_{i=1}^{s} \lambda_i = \operatorname{tr}\left(W^{-1}B\right) \\ \text{Roy's largest root} &: \frac{\lambda_1}{1+\lambda_1}. \end{aligned}$$

Other MANOVA Test Statistics iii

- Under the null hypothesis, all four statistics can be approximated using the F distribution.
    - For one-way MANOVA with $g = 2$ groups, these tests are actually all equivalent.
- In general, as the sample size increases, all four tests give similar results. For finite sample size, Roy's largest root has good power only if the leading eigenvalue λ_1 is significantly larger than the other ones.

Example i

```
knitr::kable(broom::tidy(fit), digits = 3)
```

term	df	pillai	statistic	num.df	den.df	p.value
rate	1	0.586	7.561	3	16	0.002
Residuals	18	-	-	-	-	-

Example ii

```r
knitr::kable(broom::tidy(fit, test = "Hotelling-Lawley"),
             digits = 3)
```

term	df	hl	statistic	num.df	den.df	p.value
rate	1	1.418	7.561	3	16	0.002
Residuals	18	-	-	-	-	-

Strategy for Multivariate Comparison of Treatments

1. Try to identify outliers.
    - This should be done graphically at first.
    - Once the model is fitted, you can also look at influence measures.
2. Perform a multivariate test of hypothesis.
3. If there is evidence of a multivariate difference, calculate Bonferroni confidence intervals and investigate component-wise differences.
    - The projection of the confidence region onto each variable generally leads to confidence intervals that are too large.