

Topic Six: Inferences About Multiple Mean Vectors

Preview

- Motivation:
 - The one-sample inferential techniques we just learned can be easily extended to the two-sample setting.
 - For comparing 3 or more means, we can similarly generalize ANOVA to the multivariate setting.
- Goals:
 - Learn how to perform inference in models with two or more comparison groups.

Paired Comparisons and Repeated Measures Designs

Review: A sensible design for comparing two conditions is to administer both conditions to a sample of individuals and take paired differences. For example, to compare the efficacy of two moisturizing creams, we might administer cream one to the right hand and cream two to the left hand of a sampled individual; we would want to randomize which hand gets which treatment for each individual. Let (X_{j1}, X_{j2}) be the pair of observations on individual j and $D_j = X_{j1} - X_{j2}$ be their difference, $j = 1, 2, \dots, n$. Then, assuming that the D_j are i.i.d. $N(\delta, \sigma_d^2)$, we can test $H_0 : \mu_1 - \mu_2 = \delta_0$ vs. $H_1 : \mu_1 - \mu_2 \neq \delta_0$ with a one-sample t test on the D_j , using the test statistic

$$t = \frac{\bar{D} - \delta_0}{s_d / \sqrt{n}}$$

where

$$\bar{D} = \frac{1}{n} \sum_{j=1}^n D_j \quad \text{and} \quad s_d^2 = \frac{1}{n-1} \sum_{j=1}^n (D_j - \bar{D})^2$$

Under H_0 , $t \sim t_{n-1}$, so rejecting H_0 if $|t| > t_{n-1}(\alpha/2)$ corresponds to testing at significance level α . Similarly, a $100(1 - \alpha)\%$ confidence interval for $\delta = \mu_1 - \mu_2$ is

$$\bar{D} \pm t_{n-1}(\alpha/2) \frac{s_d}{\sqrt{n}}$$

By making comparisons within individuals, paired designs are able to eliminate much of the individual-to-individual variability from the problem.

Paired Multivariate Data: Let X_{kji} be the observation for variable i on individual j under condition k , $k = 1, 2$, $j = 1, 2, \dots, n$, $i = 1, 2, \dots, p$. Let $D_{ji} = X_{1ji} - X_{2ji}$ be the paired difference for variable i on individual j and $\mathbf{D}'_j = [D_{j1}, D_{j2}, \dots, D_{jp}]$. Assuming that the \mathbf{D}_j are i.i.d. $N_p(\boldsymbol{\delta}, \Sigma_d)$,

$$T^2 = n (\bar{\mathbf{D}} - \boldsymbol{\delta})' \mathbf{S}_d^{-1} (\bar{\mathbf{D}} - \boldsymbol{\delta})$$

is distributed as an $[(n-1)p/(n-p)]F_{p,n-p}$ random variable, where

$$\bar{\mathbf{D}} = \frac{1}{n} \sum_{j=1}^n \mathbf{D}_j \quad \text{and} \quad \mathbf{S}_d = \frac{1}{n-1} \sum_{j=1}^n (\mathbf{D}_j - \bar{\mathbf{D}}) (\mathbf{D}_j - \bar{\mathbf{D}})'$$

Thus, a $100(1-\alpha)\%$ confidence region for $\boldsymbol{\delta}$ consists of all $\boldsymbol{\delta}$ such that

$$n (\bar{\mathbf{D}} - \boldsymbol{\delta})' \mathbf{S}_d^{-1} (\bar{\mathbf{D}} - \boldsymbol{\delta}) \leq \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$$

Simultaneous T^2 $100(1-\alpha)\%$ confidence intervals for the individual mean differences δ_i are given by

$$\bar{d}_i \pm \sqrt{\frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)} \sqrt{\frac{s_{d_i}^2}{n}}$$

And simultaneous Bonferroni $100(1-\alpha)\%$ confidence intervals for the δ_i are given by

$$\bar{d}_i \pm t_{n-1} \left(\frac{\alpha}{2p} \right) \sqrt{\frac{s_{d_i}^2}{n}}$$

If n and $n-p$ are large, T^2 is approximately distributed as a χ_p^2 random variable, regardless of the form of the underlying population of differences, and the above regions / intervals can be adjusted accordingly.

Example: Effluent Data Eleven samples of water discharge from a municipal wastewater treatment plant were obtained, and each was analyzed by two laboratories. The interest was in whether the two laboratories gave results that were consistent with one another. The two variables measured on each sample were biochemical oxygen demand (BOD) and suspended solids (SS).

| Sample | Lab 1 | | Lab 2 | | Difference | |
|--------|-----------|-----------|-----------|-----------|---------------|---------------|
| | x_{1j1} | x_{1j2} | x_{2j1} | x_{2j2} | δ_{j1} | δ_{j2} |
| 1 | 6 | 27 | 25 | 15 | -19 | 12 |
| 2 | 6 | 23 | 28 | 13 | -22 | 10 |
| 3 | 18 | 64 | 36 | 22 | -18 | 42 |
| 4 | 8 | 44 | 35 | 29 | -27 | 15 |
| 5 | 11 | 30 | 15 | 31 | -4 | -1 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| 11 | 20 | 14 | 39 | 21 | -19 | -7 |

We have

$$\bar{\mathbf{d}} = \begin{bmatrix} -9.36 \\ 13.27 \end{bmatrix} \quad \text{and} \quad \mathbf{S}_d = \begin{bmatrix} 199.26 & 88.38 \\ 88.38 & 418.61 \end{bmatrix}$$

Example Continued: The value of T^2 is 13.6, and

$$[2(10)/9]F_{2,9}(0.05) = 9.47$$

Thus, T^2 falls outside of the 95% confidence region for $\boldsymbol{\delta}$. Testing at $\alpha = 0.05$, we reject $H_0 : \boldsymbol{\delta} = \mathbf{0}$ and conclude that there is a nonzero mean difference between the two laboratories' measurements. The Bonferroni 95% confidence intervals for the mean differences are

$$\delta_1 : (-20.57, 1.85) \quad \text{and} \quad \delta_2 : (-2.97, 29.52)$$

The intervals both barely include zero, disagreeing slightly with the conclusion of the confidence region. Overall, I would say there is *modest* evidence of differences between the laboratories. It appears that lab 1 tends to give smaller BOD measurements and bigger SS measurements, relative to lab 2. Note that the normality assumption is suspect for the SS variable.

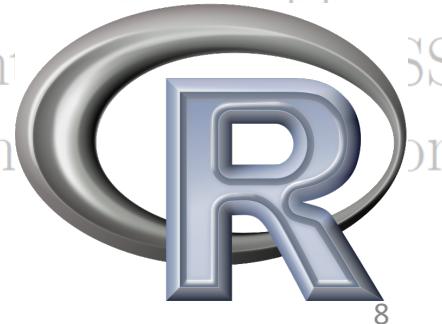
Example Continued: The value of T^2 is 13.6, and

$$[2(10)/9]F_{2,9}(0.05) = 9.47$$

Thus, T^2 falls outside of the 95% confidence region for $\boldsymbol{\delta}$. Testing at $\alpha = 0.05$, we reject $H_0 : \boldsymbol{\delta} = \mathbf{0}$ and conclude that there is a nonzero mean difference between the two laboratories' measurements. The Bonferroni 95% confidence intervals for the mean differences are

$$\delta_1 : (-20.57, 1.85) \quad \text{and} \quad \delta_2 : (-2.97, 29.52)$$

The intervals both barely include zero, disagreeing slightly with the conclusion of the confidence region. Overall, I would say there is *modest* evidence of differences between the laboratories. It appears that lab 1 tends to give smaller BOD measurements than lab 2, relative to lab 2. Note that the normality assumption is suspect for the SS variable.



Contrast Matrices: Let $\bar{\mathbf{x}}$ be the $2p \times 1$ vector of sample averages for the p variables on the two treatments, ordered as

$$\bar{\mathbf{x}}' = [\bar{x}_{11}, \bar{x}_{12}, \dots, \bar{x}_{1p}, \bar{x}_{21}, \bar{x}_{22}, \dots, \bar{x}_{2p}]$$

Similarly, let \mathbf{S} be the $2p \times 2p$ matrix of sample variances and covariances, arranged as

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix}_{(p \times p) \quad (p \times p)}$$

where \mathbf{S}_{11} contains the sample variances and covariances for the p variables on treatment 1, \mathbf{S}_{22} is for treatment 2, and $\mathbf{S}_{12} = \mathbf{S}_{21}$ contains the sample covariances between the two treatments. Now define

$$\mathbf{C}_{(p \times 2p)} = \begin{bmatrix} 1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & -1 \end{bmatrix}$$

Then $\mathbf{d}_j = \mathbf{C}\mathbf{x}_j$, $j = 1, 2, \dots, n$, $\bar{\mathbf{d}} = \mathbf{C}\bar{\mathbf{x}}$, and $\mathbf{S}_d = \mathbf{C}\mathbf{S}\mathbf{C}'$, so that

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

Thus, it is not actually necessary to compute the paired differences as we did originally. Each row \mathbf{c}_i' of \mathbf{C} is called a *contrast vector*, and \mathbf{C} , because its rows are linearly independent contrast vectors, is called a *contrast matrix*.

Repeated Measures Designs: Suppose we have a *single* response variable that we wish to compare across $q \geq 2$ treatments. A generalization of the paired design would administer all q treatments to each sampled individual, giving rise to observations $\mathbf{X}'_j = [X_{j1}, X_{j2}, \dots, X_{jq}]$, $j = 1, 2, \dots, n$. Such a design is called *repeated measures*, because all treatments are administered to each unit. To do inference on treatment differences, we can select an appropriate contrast matrix. For example, to compare treatments $2, 3, \dots, q$ to treatment 1:

$$\mathbf{C}_1\boldsymbol{\mu} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{bmatrix} = \begin{bmatrix} \mu_1 - \mu_2 \\ \mu_1 - \mu_3 \\ \vdots \\ \mu_1 - \mu_q \end{bmatrix}$$

Or, to compare treatment i to treatment $i - 1$, $i = 2, 3, \dots, q$:

$$\mathbf{C}_2\boldsymbol{\mu} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_q \end{bmatrix} = \begin{bmatrix} \mu_2 - \mu_1 \\ \mu_3 - \mu_2 \\ \vdots \\ \mu_q - \mu_{q-1} \end{bmatrix}$$

Inference for Repeated Measures Designs: With the assumption that the \mathbf{X}_j are i.i.d. $N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, and for any contrast matrix \mathbf{C} ,

$$n (\mathbf{C}\bar{\mathbf{X}} - \mathbf{C}\boldsymbol{\mu})' (\mathbf{C}\mathbf{S}\mathbf{C}')^{-1} (\mathbf{C}\bar{\mathbf{X}} - \mathbf{C}\boldsymbol{\mu}) \sim \frac{(n-1)(q-1)}{(n-q+1)} F_{q-1, n-q+1}$$

Thus, to test $H_0 : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$ vs. $H_1 : \mathbf{C}\boldsymbol{\mu} \neq \mathbf{0}$ at significance level α , we reject H_0 if

$$T^2 = n (\mathbf{C}\bar{\mathbf{x}})' (\mathbf{C}\mathbf{S}\mathbf{C}')^{-1} \mathbf{C}\bar{\mathbf{x}} > \frac{(n-1)(q-1)}{(n-q+1)} F_{q-1, n-q+1}(\alpha)$$

Similarly, a $100(1-\alpha)\%$ confidence region for $\mathbf{C}\boldsymbol{\mu}$ is determined by the set of all $\mathbf{C}\boldsymbol{\mu}$ such that

$$n (\mathbf{C}\bar{\mathbf{x}} - \mathbf{C}\boldsymbol{\mu})' (\mathbf{C}\mathbf{S}\mathbf{C}')^{-1} (\mathbf{C}\bar{\mathbf{x}} - \mathbf{C}\boldsymbol{\mu}) \leq \frac{(n-1)(q-1)}{(n-q+1)} F_{q-1, n-q+1}(\alpha)$$

And simultaneous $100(1-\alpha)\%$ confidence intervals for the $q-1$ contrasts in \mathbf{C} are given by:

$$\begin{aligned} T^2 : \mathbf{c}'\bar{\mathbf{x}} &\pm \sqrt{\frac{(n-1)(q-1)}{(n-q+1)} F_{q-1, n-q+1}(\alpha)} \sqrt{\frac{\mathbf{c}'\mathbf{S}\mathbf{c}}{n}} \\ \text{Bonferroni} : \mathbf{c}'\bar{\mathbf{x}} &\pm t_{n-1} \left(\frac{\alpha}{2(q-1)} \right) \sqrt{\frac{\mathbf{c}'\mathbf{S}\mathbf{c}}{n}} \end{aligned}$$

Example: Dog Anesthetics Each of $n = 19$ dogs was given the anesthetic drug pentobarbitol. Each dog was then given carbon dioxide (CO_2) at each of two pressure levels. Halothane (H) was then added, and the administration of CO_2 was repeated. Under each of the $q = 4$ treatment conditions, the time (in milliseconds) between heartbeats was measured. Because each dog was measured under each treatment condition, this is a repeated measures study. Define treatment 1 as high CO_2 pressure without H (with mean response μ_1), treatment 2 as low CO_2 pressure without H (μ_2), treatment 3 as high CO_2 with H (μ_3), and treatment 4 as low CO_2 with H (μ_4). We have the following summary statistics:

$$\bar{\mathbf{x}} = \begin{bmatrix} 368.21 \\ 404.63 \\ 479.26 \\ 502.89 \end{bmatrix} \quad \text{and} \quad \mathbf{S} = \begin{bmatrix} 2819.29 & 3568.42 & 2943.49 & 2295.35 \\ 3568.42 & 7963.14 & 5303.98 & 4065.44 \\ 2943.49 & 5303.98 & 6851.32 & 4499.63 \\ 2295.35 & 4065.44 & 4499.63 & 4878.99 \end{bmatrix}$$

Data are from Table 6.2 in the book.

Example Continued: Consider the three mean contrasts:

$$H \text{ "main effect"} : (\mu_3 - \mu_1) + (\mu_4 - \mu_2)$$

$$CO_2 \text{ "main effect"} : (\mu_1 - \mu_2) + (\mu_3 - \mu_4)$$

$$H-CO_2 \text{ "interaction"} : (\mu_1 - \mu_2) - (\mu_3 - \mu_4)$$

We can represent these contrasts with the contrast matrix

$$C = \begin{bmatrix} -1 & -1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

The value of T^2 is 116, and since this is greater than $[18(3)/16]F_{3,16}(0.05) = 10.94$, we reject $H_0 : C\mu = \mathbf{0}$ (no treatment effects) at significance level $\alpha = 0.05$. Bonferroni 95% confidence intervals for the three contrasts are

$$H \text{ "main effect"} : (150.51, 268.12)$$

$$CO_2 \text{ "main effect"} : (-103.70, -16.41)$$

$$H-CO_2 \text{ "interaction"} : (-65.42, 39.85)$$

Both presence of halothane and *low* CO₂ pressure are associated with longer times between heartbeats. There is not a significant interaction.

Example Continued: Consider the three mean contrasts:

$$H \text{ "main effect"} : (\mu_3 - \mu_1) + (\mu_4 - \mu_2)$$

$$CO_2 \text{ "main effect"} : (\mu_1 - \mu_2) + (\mu_3 - \mu_4)$$

$$H-CO_2 \text{ "interaction"} : (\mu_1 - \mu_2) - (\mu_3 - \mu_4)$$

We can represent these contrasts with the contrast matrix

The value of T^2 is
10.94, we reject H_0
 $\alpha = 0.05$. Bonferroni

Note that, because the measurements with halothane followed those without, there is the possibility that the H effect is just due to time. In other words, H is *confounded with time*. Ideally, the time order of all treatments would be randomized to avoid such issues.

H-CO₂ interaction : (-05.42, 59.85)

Both presence of halothane and *low* CO₂ pressure are associated with longer times between heartbeats. There is not a significant interaction.

Example Continued: Consider the three mean contrasts:

Note also that we have not assumed any special structure for the covariance matrix Σ . If, for example, correlation is expected to decay over time, such structure can be incorporated into inference, resulting in greater power. See the references in the textbook for more.

The F statistic is 10.94, which is significant at $\alpha = 0.05$.

we reject H_0 to time. In other words, H_0 is *confounded* with time. Ideally, the time order of all treatments would be randomized to avoid such issues.

$H_0 \cup H_1$ "interaction" : (-65.42, 59.85)

Both presence of halothane and *low* CO₂ pressure are associated with longer times between heartbeats. There is not a significant interaction.

Example Continued: Consider the three mean contrasts:

We note also that we have not assumed any special structure for the covariance matrix Σ . If, for example, correlation is expected to decay over time, such structure can be incorporated into inference, resulting in greater power. See the references in the textbook for more.

The p-value for the interaction term is 0.0001094, which is less than 0.05. Therefore, we reject H_0 at the $\alpha = 0.05$ level. Bonferroni

to time. In other words, H_0 is *confounded* with time. Ideally, the time order of all treatments would be randomized to avoid such issues.

H_0 : CO₂ "interaction" : (-65.42, 59.85)

Both presence of halothane and low CO₂ pressure are associated with longer times between heartbeats. There is not a significant interaction.



Comparing Mean Vectors from Two Populations

Review: Let $X_{11}, X_{21}, \dots, X_{n_1 1}$ be i.i.d. $N(\mu_1, \sigma_1^2)$ and $X_{12}, X_{22}, \dots, X_{n_2 2}$ be i.i.d. $N(\mu_2, \sigma_2^2)$, where the samples from the two populations are independent of one another (no repeated measures). To test $H_0 : \mu_1 - \mu_2 = \delta_0$:

If (1) either n_1 or n_2 are small, and (2) it is reasonable to assume that $\sigma_1^2 = \sigma_2^2$, then compute

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{s_{\text{pooled}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$$s_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Reject H_0 if $|t| > t_{n_1+n_2-2}(\alpha/2)$.

Review: Let $X_{11}, X_{21}, \dots, X_{n_1 1}$ be i.i.d. $N(\mu_1, \sigma_1^2)$ and $X_{12}, X_{22}, \dots, X_{n_2 2}$ be i.i.d. $N(\mu_2, \sigma_2^2)$, where the samples from the two populations are independent of one another (no repeated measures). To test $H_0 : \mu_1 - \mu_2 = \delta_0$:

If (1) either n_1 or n_2 are small, and (2) it is not reasonable to assume that $\sigma_1^2 = \sigma_2^2$, then compute

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Reject H_0 if $|t| > t_\nu(\alpha/2)$, where

$$\nu = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}$$

Review: Let $X_{11}, X_{21}, \dots, X_{n_1 1}$ be i.i.d. $\mathcal{N}(\mu_1, \sigma_1^2)$ and $X_{12}, X_{22}, \dots, X_{n_2 2}$ be i.i.d. $\mathcal{N}(\mu_2, \sigma_2^2)$, where the samples from the two populations are independent of one another (no repeated measures). To test $H_0 : \mu_1 - \mu_2 = \delta_0$:

If both n_1 and n_2 are large, we do not have to assume that the population distributions are normal. Compute

$$z = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Reject H_0 if $|z| > z(\alpha/2)$.

Multivariate Generalization: Let $\mathbf{X}_{11}, \mathbf{X}_{12}, \dots, \mathbf{X}_{1n_1}$ be i.i.d. $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ and $\mathbf{X}_{21}, \mathbf{X}_{22}, \dots, \mathbf{X}_{2n_2}$ be i.i.d. $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2)$, where the samples from the two populations are independent of one another (no repeated measures). If $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2$, we can compute a pooled sample covariance matrix

$$\begin{aligned}\mathbf{S}_{\text{pooled}} &= \frac{\sum_{j=1}^{n_1} (\mathbf{X}_{1j} - \bar{\mathbf{X}}_1) (\mathbf{X}_{1j} - \bar{\mathbf{X}}_1)' + \sum_{j=1}^{n_2} (\mathbf{X}_{2j} - \bar{\mathbf{X}}_2) (\mathbf{X}_{2j} - \bar{\mathbf{X}}_2)'}{n_1 + n_2 - 2} \\ &= \frac{n_1 - 1}{n_1 + n_2 - 2} \mathbf{S}_1 + \frac{n_2 - 1}{n_1 + n_2 - 2} \mathbf{S}_2\end{aligned}$$

and the T^2 statistic

$$\begin{aligned}T^2 &= (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{S}_{\text{pooled}} \right]^{-1} (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)) \\ &= \begin{pmatrix} \text{multivariate normal} \\ \text{random variable} \end{pmatrix} \left(\frac{\text{(scaled) Wishart}}{\text{random matrix}} \right)^{-1} \left(\begin{array}{c} \text{multivariate normal} \\ \text{random variable} \end{array} \right) \\ &= N_p(\mathbf{0}, \boldsymbol{\Sigma}) \left(\frac{W_{n_1+n_2-2}(\boldsymbol{\Sigma})}{n_1 + n_2 - 2} \right)^{-1} N_p(\mathbf{0}, \boldsymbol{\Sigma})\end{aligned}$$

which is distributed as

$$\frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)} F_{p, n_1 + n_2 - p - 1}$$

To test $H_0 : \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\delta}_0$, compute

$$T^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - \boldsymbol{\delta}_0)' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{S}_{\text{pooled}} \right]^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - \boldsymbol{\delta}_0)$$

and reject H_0 if $T^2 > c^2$, where

$$c^2 = \frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - p - 1)} F_{p, n_1 + n_2 - p - 1}(\alpha)$$

Similarly, a $100(1 - \alpha)\%$ confidence region for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ consists of all $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ within squared statistical distance c^2 of $\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2$. Simultaneous T^2 $100(1 - \alpha)\%$ confidence intervals for linear combinations of $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ are given by

$$\mathbf{a}' (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \pm c \sqrt{\mathbf{a}' \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{S}_{\text{pooled}} \mathbf{a}}$$

And simultaneous Bonferroni $100(1 - \alpha)\%$ confidence intervals for the p component mean differences are given by

$$\bar{x}_{1i} - \bar{x}_{2i} \pm t_{n_1 + n_2 - 2} \left(\frac{\alpha}{2p} \right) \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right) s_{ii, \text{pooled}}}$$

Unequal Covariance Matrices: If $\Sigma_1 \neq \Sigma_2$, we can no longer pool covariance estimates, and we can only approximate the sampling distribution of T^2 statistics. Now,

$$T^2 = (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))' \left[\frac{1}{n_1} \mathbf{S}_1 + \frac{1}{n_2} \mathbf{S}_2 \right]^{-1} (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))$$

is *approximately* distributed as

$$\frac{\nu p}{\nu - p + 1} F_{p, \nu - p + 1}$$

where

$$\nu = \frac{p + p^2}{\sum_{i=1}^2 \frac{1}{n_i} \left\{ \text{tr} \left[\left(\frac{1}{n_i} \mathbf{S}_i \left(\frac{1}{n_1} \mathbf{S}_1 + \frac{1}{n_2} \mathbf{S}_2 \right)^{-1} \right)^2 \right] + \left(\text{tr} \left[\frac{1}{n_i} \mathbf{S}_i \left(\frac{1}{n_1} \mathbf{S}_1 + \frac{1}{n_2} \mathbf{S}_2 \right)^{-1} \right] \right)^2 \right\}}$$

Hypothesis tests and confidence regions / intervals proceed in the usual way. Bonferroni intervals use critical values from the t_ν distribution.

Large Sample Sizes: If both $n_1 - p$ and $n_2 - p$ are large, we need not worry about whether the population distributions are normal or whether the covariances are equal, because

$$T^2 = (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))' \left[\frac{1}{n_1} \mathbf{S}_1 + \frac{1}{n_2} \mathbf{S}_2 \right]^{-1} (\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2))$$

is approximately distributed as χ_p^2 , regardless. Hypothesis tests and confidence regions / intervals proceed in the usual way. Bonferroni intervals use critical values from the standard normal distribution.

Example: Container Data Let us compare the mean concentrations of aluminum and iron between the two container types. We will assume normality and equal covariances. We have $n_1 = n_2 = 10$ and

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 0.8802 \\ 0.1237 \end{bmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 0.7292 \\ 0.0296 \end{bmatrix}, \quad \mathbf{S}_{\text{pooled}} = \begin{bmatrix} 0.00187 & 0.00014 \\ 0.00014 & 0.00014 \end{bmatrix}$$

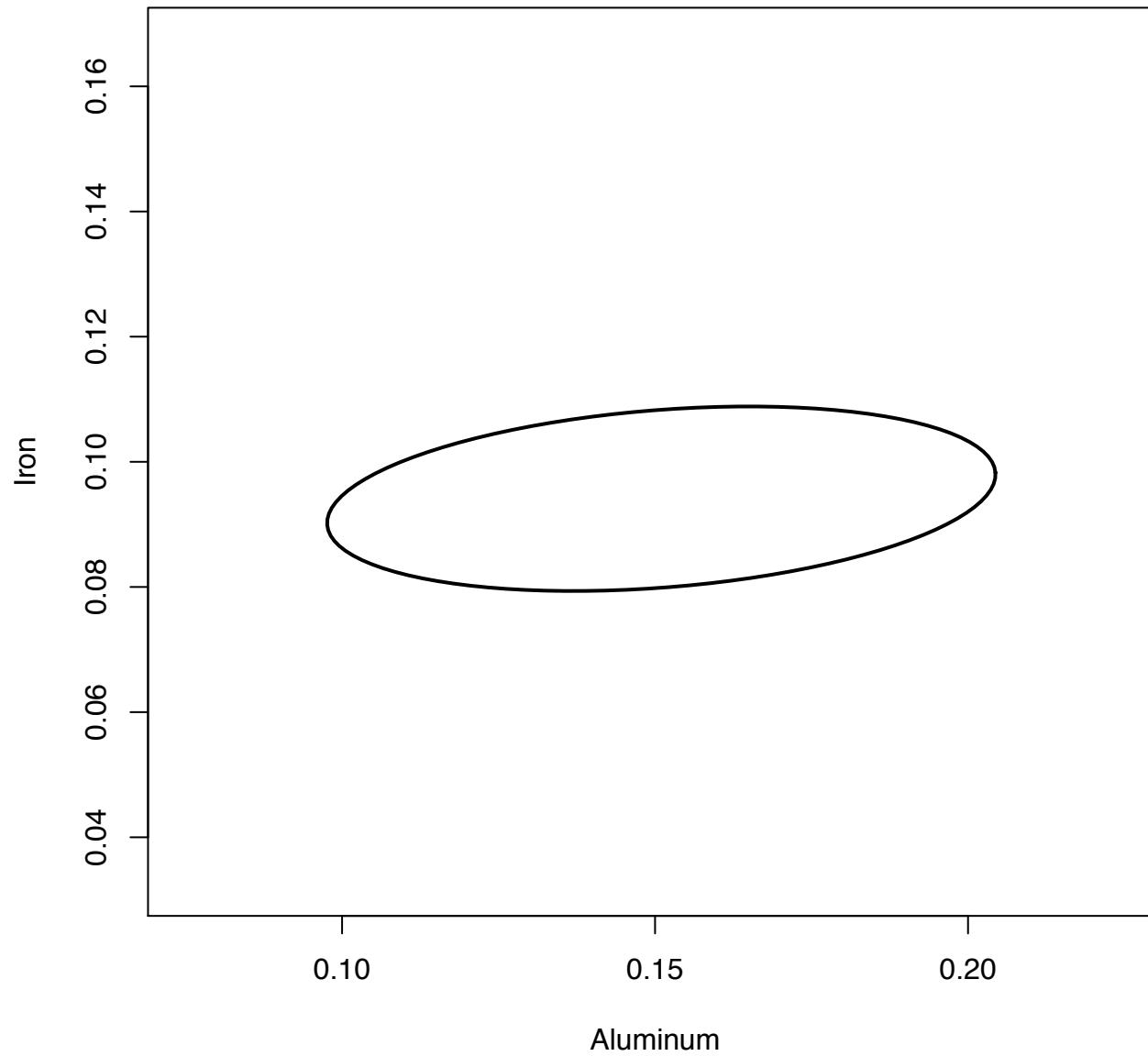
The value of T^2 is 320.55, and, since this is greater than $[18(2)/17]F_{2,17}(0.05) = 7.61$, we reject $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ at $\alpha = 0.05$. A 95% confidence ellipsoid is shown on the next slide and clearly does not contain $\mathbf{0}$. Bonferroni 95% confidence intervals for the component mean differences are

Aluminum: $(0.1037, 0.1983)$

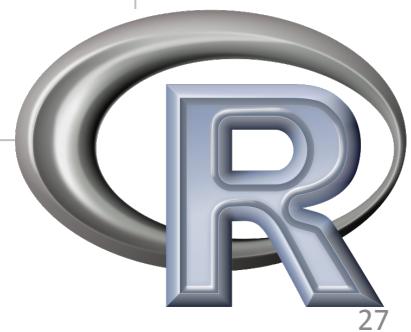
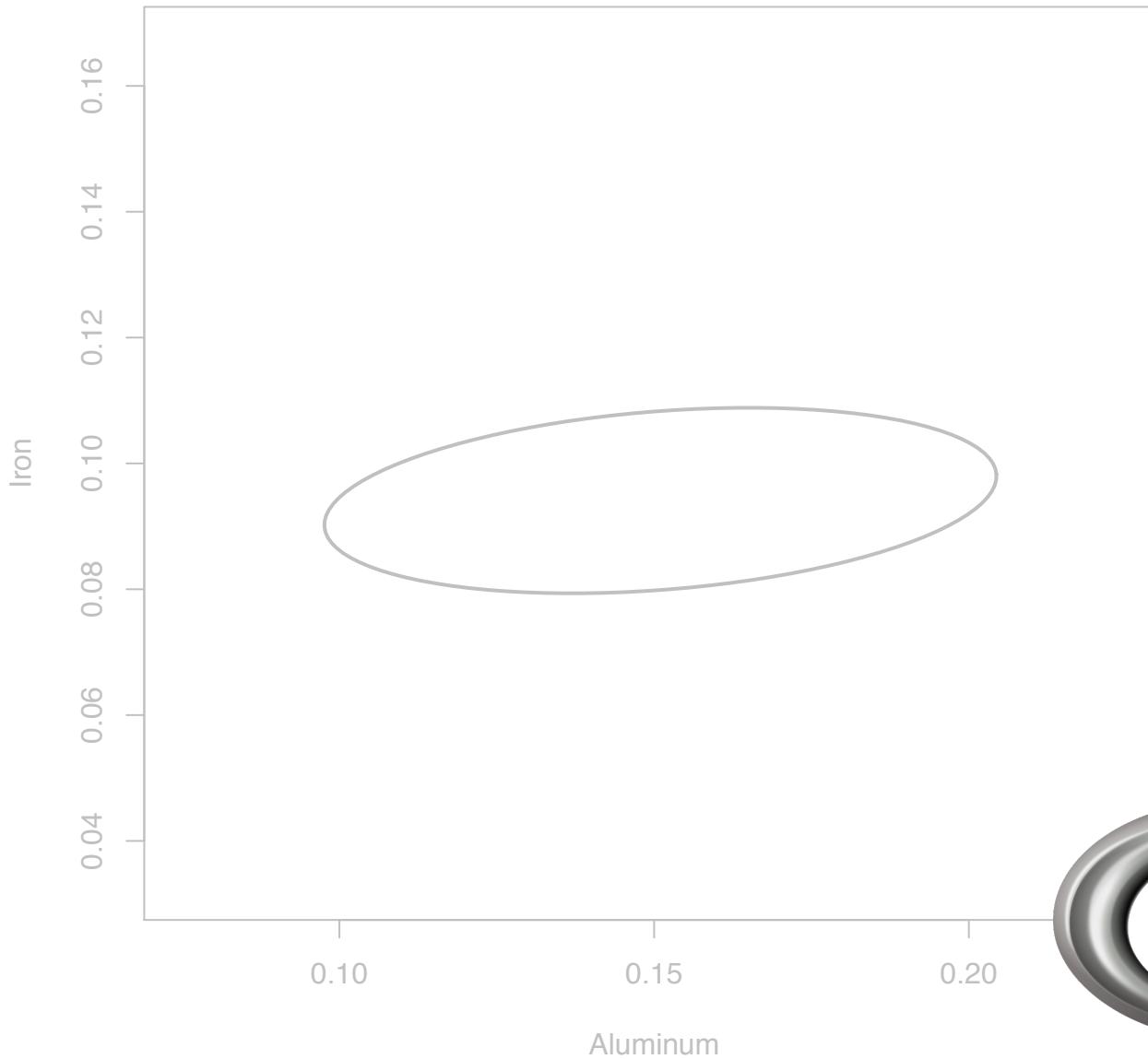
Iron: $(0.0810, 0.1072)$

There is evidence of differences in concentration for both elements. The analysis that allows for unequal covariances comes to the same conclusions.

95% Confidence Region for $\mu_1 - \mu_2$



95% Confidence Region for $\mu_1 - \mu_2$



Comparing Mean Vectors from More Than Two Populations (MANOVA)

ANOVA Review: Let $X_{l1}, X_{l2}, \dots, X_{ln_l}$ be a random sample from a $N(\mu_l, \sigma^2)$ population, $l = 1, 2, \dots, g$, and assume that the g random samples are independent of one another. The analysis of variance (ANOVA) model parameterizes the means as

$$\begin{array}{c} \mu_l \\ \left(\begin{array}{c} l\text{th population} \\ \text{mean} \end{array} \right) \end{array} = \begin{array}{c} \mu \\ \left(\begin{array}{c} \text{overall} \\ \text{mean} \end{array} \right) \end{array} + \begin{array}{c} \tau_l \\ \left(\begin{array}{c} l\text{th population} \\ (\text{treatment}) \text{ effect} \end{array} \right) \end{array}$$

The constraint that $\sum_{l=1}^g n_l \tau_l = 0$ is employed to make the model identifiable. The null hypothesis for testing equality of means is $H_0 : \tau_1 = \tau_2 = \dots = \tau_g = 0$, which is equivalent to testing $\mu_1 = \mu_2 = \dots = \mu_g$. A natural model for the response X_{lj} is now

$$\begin{array}{c} X_{lj} \\ \text{(response)} \end{array} = \begin{array}{c} \mu \\ \text{(overall mean)} \end{array} + \begin{array}{c} \tau_l \\ \left(\begin{array}{c} \text{treatment} \\ \text{effect} \end{array} \right) \end{array} + \begin{array}{c} e_{lj} \\ \left(\begin{array}{c} \text{random} \\ \text{error} \end{array} \right) \end{array}$$

where the e_{lj} are i.i.d. $N(0, \sigma^2)$ random variables. Given data, the analogous version of the above model is

$$\begin{array}{c} x_{lj} \\ \text{(observation)} \end{array} = \begin{array}{c} \bar{x} \\ \left(\begin{array}{c} \text{overall} \\ \text{sample mean} \end{array} \right) \end{array} + \begin{array}{c} (\bar{x}_l - \bar{x}) \\ \left(\begin{array}{c} \text{estimated} \\ \text{treatment effect} \end{array} \right) \end{array} + \begin{array}{c} (x_{lj} - \bar{x}_l) \\ \text{(residual)}_9 \end{array}$$

Subtracting \bar{x} from both sides and squaring gives

$$(x_{lj} - \bar{x})^2 = (\bar{x}_l - \bar{x})^2 + (x_{lj} - \bar{x}_l)^2 + (\bar{x}_l - \bar{x})(x_{lj} - \bar{x}_l)$$

Since $\sum_{j=1}^{n_l} (x_{lj} - \bar{x}_l) = 0$, summing over j gives

$$\sum_{j=1}^{n_l} (x_{lj} - \bar{x})^2 = n_l (\bar{x}_l - \bar{x})^2 + \sum_{j=1}^{n_l} (x_{lj} - \bar{x}_l)^2$$

Finally, summing over l , we have

$$\begin{aligned} \sum_{l=1}^g \sum_{j=1}^{n_l} (x_{lj} - \bar{x})^2 &= \sum_{l=1}^g n_l (\bar{x}_l - \bar{x})^2 + \sum_{l=1}^g \sum_{j=1}^{n_l} (x_{lj} - \bar{x}_l)^2 \\ \left(\begin{array}{c} \text{SS}_{\text{cor}} \\ \text{total (corrected) SS} \end{array} \right) &\quad \left(\begin{array}{c} \text{SS}_{\text{tr}} \\ \text{between (treatment) SS} \end{array} \right) \quad \left(\begin{array}{c} \text{SS}_{\text{res}} \\ \text{within (residual) SS} \end{array} \right) \end{aligned}$$

or, with $n = n_1 + n_2 + \dots + n_g$,

$$\begin{aligned} \sum_{l=1}^g \sum_{j=1}^{n_l} x_{lj}^2 &= n \bar{x}^2 + \sum_{l=1}^g n_l (\bar{x}_l - \bar{x})^2 + \sum_{l=1}^g \sum_{j=1}^{n_l} (x_{lj} - \bar{x}_l)^2 \\ (\text{SS}_{\text{obs}}) &\quad (\text{SS}_{\text{mean}}) \quad (\text{SS}_{\text{tr}}) \quad (\text{SS}_{\text{res}}) \end{aligned}$$

Subtracting \bar{x} from both sides and squaring gives

$$\text{Since } \sum_{j=1}^{n_l} (x_{lj})$$

Having decomposed the total variability, we can ask whether the variability between populations is significantly greater than that within populations. Under H_0 , they should be the same. The ANOVA test statistic is the ratio of the between-group variance to the within-group variance. In order to compute these variances, we first need the appropriate degrees of freedom for dividing the SS terms.

Finally, summing over l , we have

$$\begin{aligned} \sum_{l=1}^g \sum_{j=1}^{n_l} (x_{lj} - \bar{x})^2 &= \sum_{l=1}^g n_l (\bar{x}_l - \bar{x})^2 + \sum_{l=1}^g \sum_{j=1}^{n_l} (x_{lj} - \bar{x}_l)^2 \\ \left(\begin{array}{c} \text{SS}_{\text{cor}} \\ \text{total (corrected) SS} \end{array} \right) &\quad \left(\begin{array}{c} \text{SS}_{\text{tr}} \\ \text{between (treatment) SS} \end{array} \right) \quad \left(\begin{array}{c} \text{SS}_{\text{res}} \\ \text{within (residual) SS} \end{array} \right) \end{aligned}$$

or, with $n = n_1 + n_2 + \dots + n_g$,

$$\begin{aligned} \sum_{l=1}^g \sum_{j=1}^{n_l} x_{lj}^2 &= n\bar{x}^2 + \sum_{l=1}^g n_l (\bar{x}_l - \bar{x})^2 + \sum_{l=1}^g \sum_{j=1}^{n_l} (x_{lj} - \bar{x}_l)^2 \\ (\text{SS}_{\text{obs}}) &\quad (\text{SS}_{\text{mean}}) \quad (\text{SS}_{\text{tr}}) \quad (\text{SS}_{\text{res}}) \end{aligned}$$

Let $\mathbf{y}' = [x_{11}, \dots, x_{1n_1}, x_{21}, \dots, x_{2n_2}, \dots, x_{gn_g}]$ be the vector of all observations. This observation vector has the “freedom” to lie anywhere in $n = \sum_l n_l$ dimensions. The overall mean vector $\bar{x}\mathbf{1} = [\bar{x}, \bar{x}, \dots, \bar{x}]'$ must lie along the equiangular (one-dimensional) line of $\mathbf{1}$. We can write the treatment vector as

$$(\bar{x}_1 - \bar{x}) \mathbf{u}_1 + (\bar{x}_2 - \bar{x}) \mathbf{u}_2 + \cdots + (\bar{x}_g - \bar{x}) \mathbf{u}_g$$

where, with $N_1 = 0$ and $N_l = \sum_{k=1}^{l-1} n_k$, $l = 2, 3, \dots, g$, \mathbf{u}_l has ones in elements $N_l + 1, N_l + 2, \dots, N_l + n_l$ and zeros everywhere else. In other words, the treatment vector lies in the (g -dimensional) hyperplane of linear combinations of the $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_g$. Since $\mathbf{1} = \mathbf{u}_1 + \mathbf{u}_2 + \cdots + \mathbf{u}_g$, the overall mean vector lies in this same hyperplane, and it will always be perpendicular to the treatment vector. Hence, the treatment vector has the freedom to lie in any of the $g - 1$ dimensions not occupied by the overall mean vector. The residual vector is perpendicular to both the overall mean vector and treatment vector and therefore has the freedom to lie in $n - (g - 1) - 1 = n - g$ dimensions.

The sums of squares and associated degrees of freedom are summarized by an ANOVA table:

| Source of variation | Sum of squares (SS) | Degrees of freedom (d.f.) |
|------------------------|---|------------------------------|
| Treatments | $\text{SS}_{\text{tr}} = \sum_{l=1}^g n_l (\bar{x}_l - \bar{x})^2$ | $g - 1$ |
| Residual | $\text{SS}_{\text{res}} = \sum_{l=1}^g \sum_{j=1}^{n_l} (x_{lj} - \bar{x}_l)^2$ | $n - g$ |
| Total | $\text{SS}_{\text{cor}} = \sum_{l=1}^g \sum_{j=1}^{n_l} (x_{lj} - \bar{x})^2$ | $n - 1$ |

We reject $H_0 : \tau_1 = \tau_2 = \dots = \tau_g = 0$ at level α if

$$F = \frac{\text{SS}_{\text{tr}}/(g-1)}{\text{SS}_{\text{res}}/(n-g)} > F_{g-1, n-g}(\alpha)$$

Multivariate Analysis of Variance (MANOVA): The MANOVA model is the vector generalization of the ANOVA model:

$$\mathbf{X}_{lj} = \boldsymbol{\mu} + \boldsymbol{\tau}_l + \mathbf{e}_{lj}, \quad j = 1, 2, \dots, n_l \quad \text{and} \quad l = 1, 2, \dots, g$$

where the \mathbf{e}_{lj} are i.i.d. $N_p(\mathbf{0}, \Sigma)$ random variables and $\sum_{l=1}^g n_l \boldsymbol{\tau}_l = \mathbf{0}$. As with ANOVA, we decompose total variation into between- and within-population components. First:

$$\begin{array}{ccccccccc} \mathbf{x}_{lj} & = & \bar{\mathbf{x}} & + & (\bar{\mathbf{x}}_l - \bar{\mathbf{x}}) & + & (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) \\ & & & & & & & \\ (\text{observation}) & & \left(\begin{array}{c} \text{overall} \\ \text{sample mean} \end{array} \right) & & \left(\begin{array}{c} \text{estimated} \\ \text{treatment effect} \end{array} \right) & & & (\text{residual}) \end{array}$$

Now note that we can write

$$\begin{aligned} (\mathbf{x}_{lj} - \bar{\mathbf{x}}) (\mathbf{x}_{lj} - \bar{\mathbf{x}})' &= [(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})] [(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})]' \\ &= (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)' + (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})' \\ &\quad + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}}) (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)' + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})' \end{aligned}$$

The sum over j of the middle two expressions is the zero matrix.

Multivariate Analysis of Variance (MANOVA): The MANOVA model is the vector

$$\sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}}) (\mathbf{x}_{lj} - \bar{\mathbf{x}})' = \sum_{l=1}^g n_l (\bar{\mathbf{x}}_l - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})' + \sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)'$$

$\begin{pmatrix} \mathbf{B} + \mathbf{W} \\ \text{total (corrected) SSP} \end{pmatrix} \quad \begin{pmatrix} \mathbf{B} \\ \text{between (treatment) SSP} \end{pmatrix} \quad \begin{pmatrix} \mathbf{W} \\ \text{within (residual) SSP} \end{pmatrix}$

decompose total variation into between and within population components. This.

$$\mathbf{x}_{lj} = \bar{\mathbf{x}} + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}}) + (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)$$

(observation) (overall sample mean) (estimated treatment effect) (residual)

Now note that we can write

$$\begin{aligned} (\mathbf{x}_{lj} - \bar{\mathbf{x}}) (\mathbf{x}_{lj} - \bar{\mathbf{x}})' &= [(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})] [(\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})]' \\ &= (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)' + (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})' \\ &\quad + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}}) (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)' + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})' \end{aligned}$$

The sum over j of the middle two expressions is the zero matrix.

Multivariate Analysis of Variance (MANOVA):

The MANOVA model is the vector

$$\sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}}) (\mathbf{x}_{lj} - \bar{\mathbf{x}})' = \sum_{l=1}^g n_l (\bar{\mathbf{x}}_l - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})' + \sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)'$$

| | | |
|--|---|---------------------------------------|
| $\mathbf{B} + \mathbf{W}$ total (corrected) SSP | \mathbf{B} between (treatment) SSP | \mathbf{W} within (residual) SSP |
|--|---|---------------------------------------|

decompose total variation into between and within population components. First,

$$\mathbf{x}_{lj} = \bar{\mathbf{x}} + (\bar{\mathbf{x}}_l - \bar{\mathbf{x}}) + (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)$$

$$(\text{observation}) \quad \begin{pmatrix} \text{overall} \\ \text{sample mean} \end{pmatrix} \quad \begin{pmatrix} \text{estimated} \\ \text{treatment effect} \end{pmatrix} \quad (\text{residual})$$

Note that

$$\begin{aligned} (\mathbf{x}_{lj} - \bar{\mathbf{x}}) (\mathbf{x}_{lj} - \bar{\mathbf{x}})' &= \sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)' \\ &= (n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2 + \cdots + (n_g - 1)\mathbf{S}_g \end{aligned}$$

The sum over j of the sample covariance matrices \mathbf{S}_l is the sample covariance matrix of the l th sample. This is analogous to the numerator of the $\mathbf{S}_{\text{pooled}}$ sample covariance matrix we used in the two-sample case.

The sums of squares (and cross-products) and associated degrees of freedom are summarized by a MANOVA table:

| Source of variation | Sum of squares and cross-products | Degrees of freedom (d.f.) |
|------------------------|--|------------------------------|
| Treatments | $\mathbf{B} = \sum_{l=1}^g n_l (\bar{\mathbf{x}}_l - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_l - \bar{\mathbf{x}})'$ | $g - 1$ |
| Residual | $\mathbf{W} = \sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l) (\mathbf{x}_{lj} - \bar{\mathbf{x}}_l)'$ | $n - g$ |
| Total | $\mathbf{B} + \mathbf{W} = \sum_{l=1}^g \sum_{j=1}^{n_l} (\mathbf{x}_{lj} - \bar{\mathbf{x}}) (\mathbf{x}_{lj} - \bar{\mathbf{x}})'$ | $n - 1$ |

The most commonly used test of $H_0 : \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_g = \mathbf{0}$ rejects if

$$\Lambda^* = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|}$$

is too small. The statistic Λ^* is called *Wilks's lambda*. Several other statistics exist, including Pillai's trace, the Lawley-Hotelling statistic, and Roy's largest root statistic. All are based on likelihood ratio theory and have similar operating characteristics.

The sampling distribution of Λ^* under H_0 is known exactly for a handful of conditions and can be approximated generally with large sample sizes.

| No. of variables | No. of groups | Sampling distribution under H_0 |
|------------------|---------------|---|
| $p = 1$ | $g \geq 2$ | $\left(\frac{n-g}{g-1}\right) \left(\frac{1-\Lambda^*}{\Lambda^*}\right) \sim F_{g-1, n-g}$ |
| $p = 2$ | $g \geq 2$ | $\left(\frac{n-g-1}{g-1}\right) \left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2(g-1), 2(n-g-1)}$ |
| $p \geq 1$ | $g = 2$ | $\left(\frac{n-p-1}{p}\right) \left(\frac{1-\Lambda^*}{\Lambda^*}\right) \sim F_{p, n-p-1}$ |
| $p \geq 1$ | $g = 3$ | $\left(\frac{n-p-2}{p}\right) \left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2p, 2(n-p-2)}$ |
| $p \geq 1$ | $g \geq 2$ | $-(n - 1 - \frac{p+g}{2}) \ln \Lambda^* \stackrel{d}{\sim} \chi^2_{p(g-1)}$ |

The sampling distribution of Λ^* under H_0 is known exactly for a handful of conditions and can be approximated generally with large sample sizes.

| No. of variables | No. of groups | Sampling distribution under H_0 |
|------------------|---------------|---|
| $p = 1$ | $g \geq 2$ | $\left(\frac{n-g}{g-1}\right) \left(\frac{1-\Lambda^*}{\Lambda^*}\right) \sim F_{g-1, n-g}$ |
| $p = 2$ | $g \geq 2$ | $\left(\frac{n-g-1}{g-1}\right) \left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2(g-1), 2(n-g-1)}$ |
| $p \geq 1$ | $g = 2$ | $\left(\frac{n-p-1}{p}\right) \left(\frac{1-\Lambda^*}{\Lambda^*}\right) \sim F_{p, n-p-1}$ |

This is a large-sample approximation.

$$\left(\frac{n-p-1}{p}\right) \left(\frac{1-\sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2p, 2(n-p-2)}$$

$$p \geq 1 \quad g \geq 2 \quad -\left(n - 1 - \frac{p+g}{2}\right) \ln \Lambda^* \stackrel{d}{\sim} \chi^2_{p(g-1)}$$

Simultaneous Confidence Intervals: We use

$$\hat{\tau}_{ki} - \hat{\tau}_{li} = (\bar{x}_{ki} - \bar{x}_i) - (\bar{x}_{li} - \bar{x}_i) = \bar{x}_{ki} - \bar{x}_{li}$$

to estimate $\tau_{ki} - \tau_{li}$, the mean difference between populations k and l on component i . We have

$$\text{Var}(\bar{X}_{ki} - \bar{X}_{li}) = \left(\frac{1}{n_k} + \frac{1}{n_l} \right) \sigma_{ii}$$

where σ_{ii} is the i th diagonal element of Σ . We can estimate Σ by dividing \mathbf{W} by its degrees of freedom:

$$\hat{\Sigma} = \frac{1}{n-g} \mathbf{W}$$

and hence estimate $\text{Var}(\bar{X}_{ki} - \bar{X}_{li})$ as

$$\widehat{\text{Var}}(\bar{X}_{ki} - \bar{X}_{li}) = \left(\frac{1}{n_k} + \frac{1}{n_l} \right) \frac{w_{ii}}{n-g}$$

where w_{ii} is the i th diagonal element of \mathbf{W} . Bonferroni $100(1-\alpha)\%$ confidence intervals for all $g(g-1)/2$ pairwise comparisons on the p variables can then be constructed as

$$\bar{x}_{ki} - \bar{x}_{li} \pm t_{n-g} \left(\frac{\alpha}{pg(g-1)} \right) \sqrt{\frac{w_{ii}}{n-g} \left(\frac{1}{n_k} + \frac{1}{n_l} \right)}$$

for all components $i = 1, 2, \dots, p$ and all differences $l < k = 1, 2, \dots, g$.

Example: Nursing Homes For each of $n_1 = 271$ privately-owned, $n_2 = 138$ nonprofit, and $n_3 = 107$ government-owned nursing homes ($g = 3$), we have observations on $p = 4$ variables: X_1 = cost of nursing labor, X_2 = cost of dietary labor, X_3 = cost of plant operation and maintenance labor, and X_4 = cost of housekeeping and laundry labor. The sample mean vectors and covariance matrices are

$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 2.066 \\ 0.480 \\ 0.082 \\ 0.360 \end{bmatrix} \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 2.167 \\ 0.596 \\ 0.124 \\ 0.418 \end{bmatrix} \quad \bar{\mathbf{x}}_3 = \begin{bmatrix} 2.273 \\ 0.521 \\ 0.125 \\ 0.383 \end{bmatrix}$$

and

$$\mathbf{S}_1 = \begin{bmatrix} 0.291 & -0.001 & 0.002 & 0.010 \\ -0.001 & 0.011 & 0.000 & 0.003 \\ 0.002 & 0.000 & 0.001 & 0.000 \\ 0.010 & 0.003 & 0.000 & 0.010 \end{bmatrix} \quad \mathbf{S}_2 = \begin{bmatrix} 0.561 & 0.011 & 0.001 & 0.037 \\ 0.011 & 0.025 & 0.004 & 0.007 \\ 0.001 & 0.004 & 0.005 & 0.002 \\ 0.037 & 0.007 & 0.002 & 0.019 \end{bmatrix}$$

$$\mathbf{S}_3 = \begin{bmatrix} 0.261 & 0.030 & 0.003 & 0.018 \\ 0.030 & 0.017 & 0.000 & 0.006 \\ 0.003 & 0.000 & 0.004 & 0.001 \\ 0.018 & 0.006 & 0.001 & 0.013 \end{bmatrix}$$

Example Continued: We have

$$\mathbf{B} = \begin{bmatrix} 3.469 & 1.099 & 0.811 & 0.586 \\ 1.099 & 1.231 & 0.450 & 0.616 \\ 0.811 & 0.450 & 0.232 & 0.231 \\ 0.586 & 0.616 & 0.231 & 0.309 \end{bmatrix} \quad \text{and} \quad \mathbf{W} = \begin{bmatrix} 183.093 & 4.417 & 0.995 & 9.677 \\ 4.417 & 8.197 & 0.548 & 2.405 \\ 0.995 & 0.548 & 1.379 & 0.380 \\ 9.677 & 2.405 & 0.380 & 6.681 \end{bmatrix}$$

so that $\Lambda^* = |\mathbf{W}|/|\mathbf{B} + \mathbf{W}| = 0.763$. Since the sample sizes are large, we can use

$$-\left(n - 1 - \frac{p + g}{2}\right) \ln \Lambda^* = -\left(516 - 1 - \frac{4 + 3}{2}\right) \ln 0.763 = 138.359$$

as our test statistic and compute a p-value using the χ^2_8 distribution. The p-value is nearly zero, so we reject H_0 : no differences in mean cost vectors between the nursing home types. Allowing for all $pg(g-1)/2 = 4(3)(2)/2 = 12$ pairwise comparisons, we can compute Bonferroni 95% confidence intervals. For example, a 95% confidence interval for the mean difference in plant operation and maintenance costs between private- and government-owned homes is

$$\begin{aligned} \bar{x}_{13} - \bar{x}_{33} &\pm t_{n-g} \left(\frac{0.05}{pg(g-1)} \right) \sqrt{\frac{w_{33}}{n-g} \left(\frac{1}{n_1} + \frac{1}{n_3} \right)} \\ &= 0.082 - 0.125 \pm 2.878 \sqrt{\frac{1.379}{516-3} \left(\frac{1}{271} + \frac{1}{107} \right)} \end{aligned}$$

or the interval from -0.060 to -0.026. The government homes tend to spend more on plant operation and maintenance than the private homes.

Example Continued: We have

$$\mathbf{B} = \begin{bmatrix} 3.469 & 1.099 & 0.811 & 0.586 \\ 1.099 & 1.231 & 0.450 & 0.616 \\ 0.811 & 0.450 & 0.232 & 0.231 \\ 0.586 & 0.616 & 0.231 & 0.309 \end{bmatrix} \quad \text{and} \quad \mathbf{W} = \begin{bmatrix} 183.093 & 4.417 & 0.995 & 9.677 \\ 4.417 & 8.197 & 0.548 & 2.405 \\ 0.995 & 0.548 & 1.379 & 0.380 \\ 9.677 & 2.405 & 0.380 & 6.681 \end{bmatrix}$$

so that $\Lambda^* = |\mathbf{W}|/|\mathbf{B} + \mathbf{W}| = 0.763$. Since the sample sizes are large, we can use

$$-\left(n - 1 - \frac{p + g}{2}\right) \ln \Lambda^* = -\left(516 - 1 - \frac{4 + 3}{2}\right) \ln 0.763 = 138.359$$

as our test statistic and compute a p-value using the χ_8^2 distribution. The p-value is nearly zero, so we reject H_0 : no differences in mean cost vectors between the nursing home types. Allowing for all $pg(g-1)/2 = 4(3)(2)/2 = 12$ pairwise comparisons, we can compute Bonferroni 95% confidence intervals. For example, a 95% confidence interval for the mean difference in plant operation and maintenance costs between private- and government-owned homes is

$$\begin{aligned} \bar{x}_{13} - \bar{x}_{33} &\pm t_{n-g} \left(\frac{0.05}{pg(g-1)} \right) \sqrt{\frac{w_{33}}{n-g} \left(\frac{1}{n_1} + \frac{1}{n_3} \right)} \\ &= 0.082 - 0.125 \pm 2.878 \sqrt{\frac{1.379}{516-3} \left(\frac{1}{27} + \frac{1}{27} \right)} \end{aligned}$$



or the interval from -0.060 to -0.026. The government homes tend to have lower plant operation and maintenance than the private homes.

Testing for Equality of Covariance Matrices

Testing for Equality of Covariance Matrices: The assumption we have been making of equal covariance matrices is a strong one. To determine whether it is reasonable, we could test $H_0 : \Sigma_1 = \Sigma_2 = \cdots = \Sigma_g$. Assuming multivariate normal populations, the likelihood ratio statistic for testing H_0 is

$$\Lambda = \prod_{l=1}^g \left(\frac{|\mathbf{S}_l|}{|\mathbf{S}_{\text{pooled}}|} \right)^{(n_l-1)/2}$$

where $\mathbf{S}_{\text{pooled}} = \mathbf{W}/(n - g)$. Box's test is based on the large-sample approximation of

$$-2 \ln \Lambda = (n - g) \ln |\mathbf{S}_{\text{pooled}}| - \sum_{l=1}^g [(n_l - 1) \ln |\mathbf{S}_l|]$$

With

$$u = \left[\sum_{l=1}^g \frac{1}{(n_l - 1)} - \frac{1}{n - g} \right] \left[\frac{2p^2 + 3p - 1}{6(p + 1)(g - 1)} \right]$$

we have

$$-2(1 - u) \ln \Lambda \stackrel{\cdot}{\sim} \chi_{\nu}^2$$

where $\nu = p(p + 1)(g - 1)/2$. This approximation works well if each $n_l > 20$, $p \leq 5$, and $g \leq 5$.

Nursing Home Data: To test $H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3$, we have $-2 \ln \Lambda = 244.146$ and $u = 0.013$, so $-2(1 - u) \ln \Lambda = 240.972$. Comparing this to the χ^2_{20} distribution, we obtain a p-value that is very nearly zero. We therefore conclude that the covariances are not equal, even though they appear qualitatively equal to the eye.

Nursing Home Data: To test $H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3$, we have $-2 \ln \Lambda = 244.146$ and $u = 0.013$, so $-2(1 - u) \ln \Lambda = 240.972$. Comparing this to the χ^2_{20} distribution, we obtain a p-value that is very nearly zero. We therefore conclude that the covariances are not equal, even though they appear qualitatively equal to the eye.

With reasonably large sample sizes, MANOVA is robust to modest differences between the covariance matrices. Also, when the sample sizes are equal, the effect of unequal covariances may be small. Thus, depending on the situation, we need not necessarily abandon MANOVA when the equal covariance assumption is rejected by Box's test.

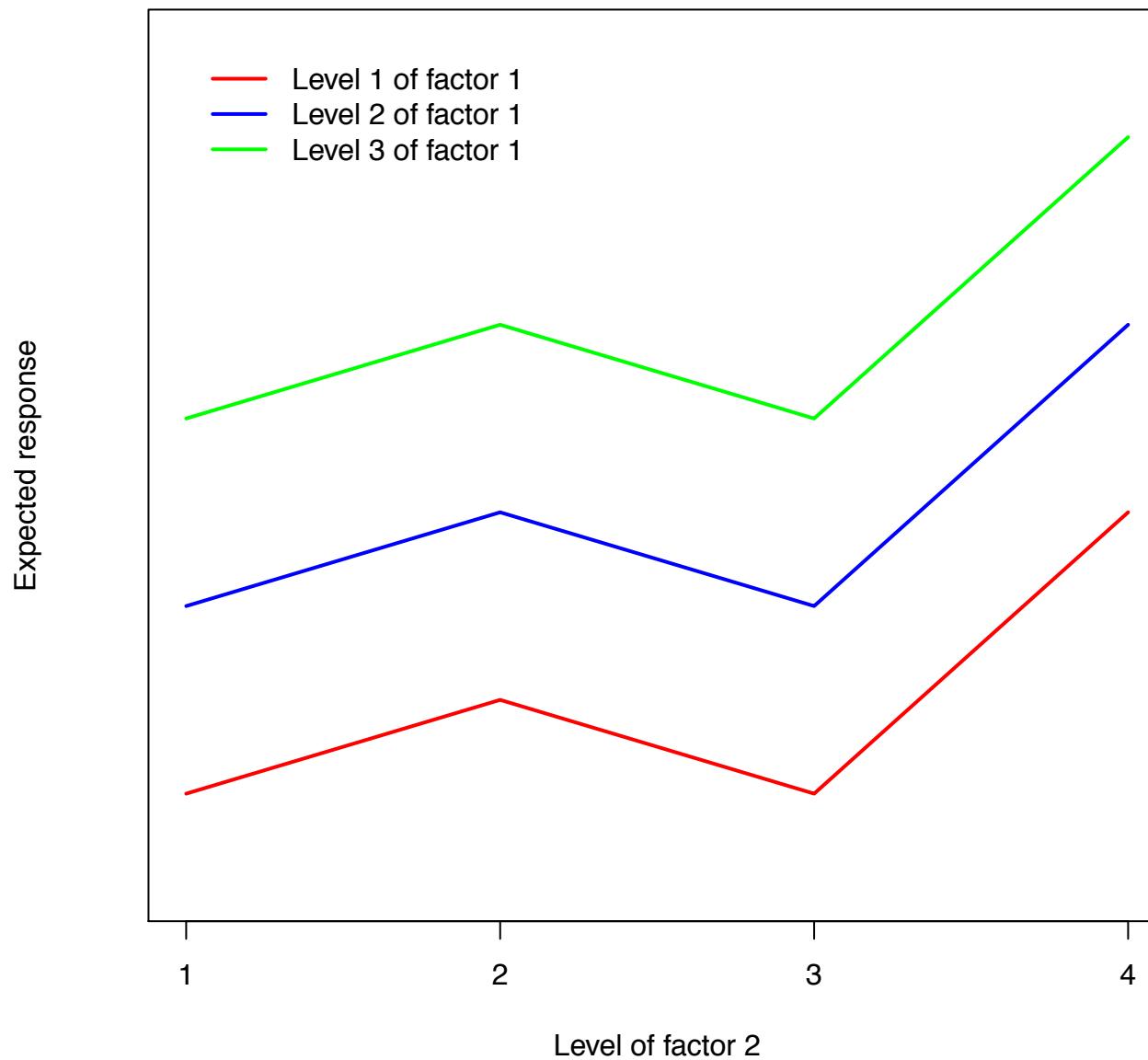
Two-Way MANOVA

Two-Way ANOVA Review: Now let there be *two* sets of experimental conditions, called factor 1 and factor 2. Suppose there are g levels of factor 1 and b levels of factor 2, and that n independent observations are obtained in each of the gb combinations of levels. With X_{lkr} the r th observation at level l of factor 1 and level k of factor 2, we can write

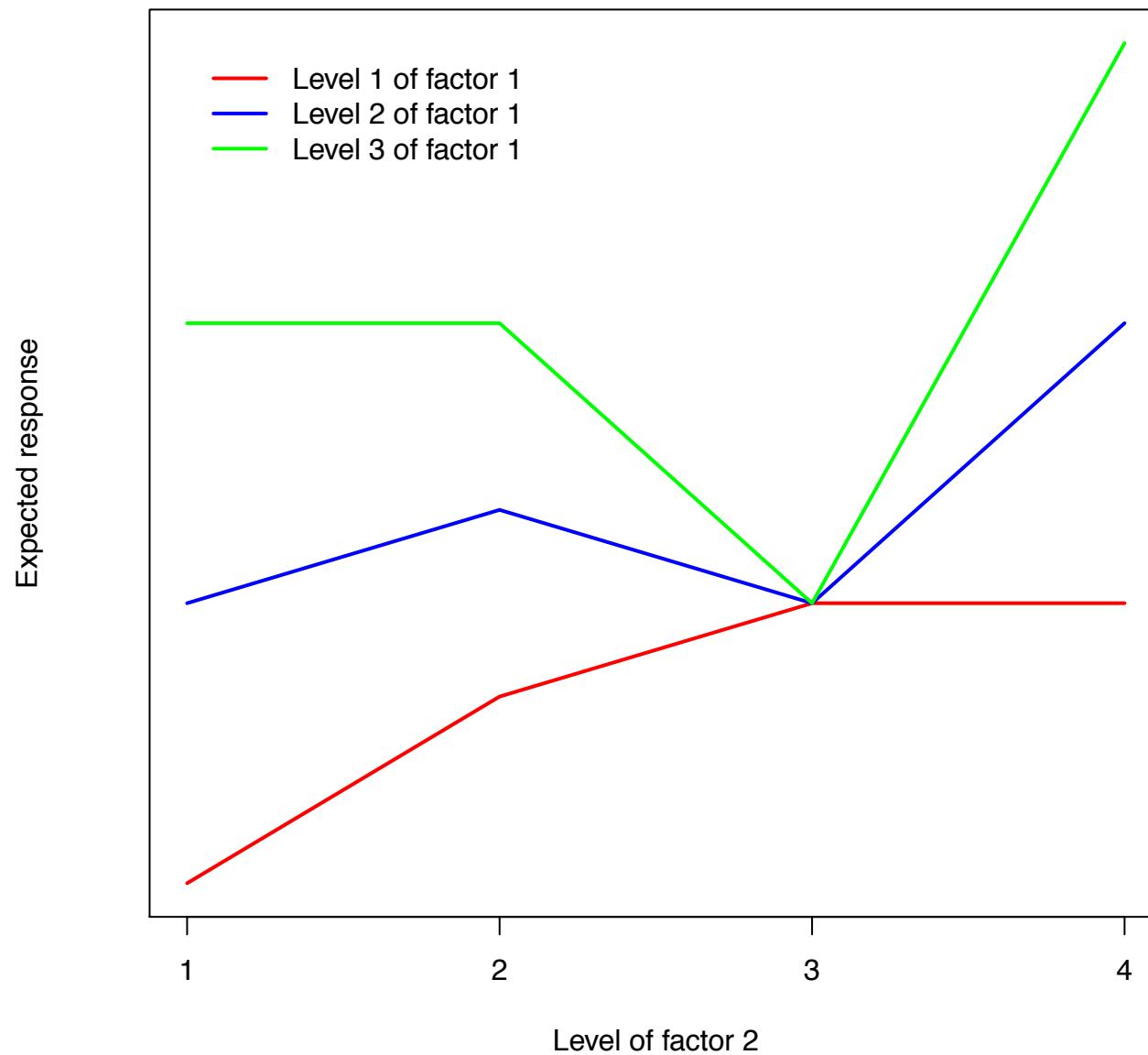
$$X_{lkr} = \mu + \tau_l + \beta_k + \gamma_{lk} + e_{lkr}$$

where $\sum_{l=1}^g \tau_l = \sum_{k=1}^b \beta_k = \sum_{l=1}^g \gamma_{lk} = \sum_{k=1}^b \gamma_{lk} = 0$ and the e_{lkr} are i.i.d. $N(0, \sigma^2)$ random variables. We interpret μ as the overall mean, the τ_l as the factor 1 main effects, the β_k as the factor 2 main effects, and the γ_{lk} as the interaction effects between factors 1 and 2. The interaction terms allow for non-additive effects of factors 1 and 2. As we will see in a moment, interaction terms complicate the model's interpretation.

Without Interaction



With Interaction



We can decompose each observation as

$$x_{lkr} = \bar{x} + (\bar{x}_{l\cdot} - \bar{x}) + (\bar{x}_{\cdot k} - \bar{x}) + (\bar{x}_{lk} - \bar{x}_{l\cdot} - \bar{x}_{\cdot k} + \bar{x}) + (x_{lkr} - \bar{x}_{lk})$$

where \bar{x} is the overall average, $\bar{x}_{l\cdot}$ is the average for the l th level of factor 1, $\bar{x}_{\cdot k}$ is the average for the k th level of factor 2, and \bar{x}_{lk} is the average for the l th level of factor 1 and k th level of factor 2. Subtracting \bar{x} from both sides and summing the squares gives

$$\begin{aligned} \sum_{l=1}^g \sum_{k=1}^b \sum_{r=1}^n (x_{lkr} - \bar{x})^2 &= \sum_{l=1}^g bn (\bar{x}_{l\cdot} - \bar{x})^2 + \sum_{k=1}^b gn (\bar{x}_{\cdot k} - \bar{x})^2 \\ &\quad + \sum_{l=1}^g \sum_{k=1}^b n (\bar{x}_{lk} - \bar{x}_{l\cdot} - \bar{x}_{\cdot k} + \bar{x})^2 \\ &\quad + \sum_{l=1}^g \sum_{k=1}^b \sum_{r=1}^n (x_{lkr} - \bar{x}_{lk})^2 \end{aligned}$$

or

$$SS_{\text{cor}} = SS_{\text{fac 1}} + SS_{\text{fac 2}} + SS_{\text{int}} + SS_{\text{res}}$$

| Source of variation | Sum of squares (SS) | Degrees of freedom (d.f.) |
|------------------------|---|------------------------------|
| Factor 1 | $SS_{\text{fac } 1} = \sum_{l=1}^g bn (\bar{x}_{l\cdot} - \bar{x})^2$ | $g - 1$ |
| Factor 2 | $SS_{\text{fac } 2} = \sum_{k=1}^b gn (\bar{x}_{\cdot k} - \bar{x})^2$ | $b - 1$ |
| Interaction | $SS_{\text{int}} = \sum_{l=1}^g \sum_{k=1}^b n (\bar{x}_{lk} - \bar{x}_{l\cdot} - \bar{x}_{\cdot k} + \bar{x})^2$ | $(g - 1)(b - 1)$ |
| Residual | $SS_{\text{res}} = \sum_{l=1}^g \sum_{k=1}^b \sum_{r=1}^n (x_{lkr} - \bar{x}_{lk})^2$ | $gb(n - 1)$ |
| Total | $SS_{\text{cor}} = \sum_{l=1}^g \sum_{k=1}^b \sum_{r=1}^n (x_{lkr} - \bar{x})^2$ | $gbn - 1$ |

Inference for Two-Way ANOVA: The following F statistics can be used to individually test $H_{01} : \tau_1 = \tau_2 = \cdots = \tau_g = 0$ (no factor 1 main effects), $H_{02} : \beta_1 = \beta_2 = \cdots = \beta_b = 0$ (no factor 2 main effects), and $H_{03} : \gamma_{11} = \gamma_{12} = \cdots = \gamma_{gb} = 0$ (no interaction), respectively:

$$F_1 = \frac{\text{SS}_{\text{fac } 1}/(g - 1)}{\text{SS}_{\text{res}}/(gb(n - 1))} \quad F_2 = \frac{\text{SS}_{\text{fac } 2}/(b - 1)}{\text{SS}_{\text{res}}/(gb(n - 1))}$$

$$F_3 = \frac{\text{SS}_{\text{int}}/((g - 1)(b - 1))}{\text{SS}_{\text{res}}/(gb(n - 1))}$$

Note, however, that the interpretation of main effects is complicated by the inclusion of an interaction. Interactions allow for factor 1 effects to vary across the levels of factor 2, so that the τ_l represent factor 1 effects that have been averaged over the levels of factor 2. It is therefore possible for there to be interaction effects but no main effects. For example, consider the following means when $g = 3$ and $b = 2$:

| | $l = 1$ | $l = 2$ | $l = 3$ |
|---------|---------|---------|---------|
| $k = 1$ | 6.00 | 4.75 | 4.25 |
| $k = 2$ | 4.00 | 5.25 | 5.75 |

We have $\mu = 5.00$, $\tau_1 = \tau_2 = \tau_3 = 0$, $\beta_1 = \beta_2 = 0$, $\gamma_{11} = 1.00$, $\gamma_{12} = -1.00$, $\gamma_{21} = -0.25$, $\gamma_{22} = 0.25$, $\gamma_{31} = -0.75$, and $\gamma_{32} = 0.75$. While the main effects are zero, the factors *should* be considered “interesting” overall since their means do differ across their levels (it’s just that their effects average to zero across the levels of the other factor). Note that we could specifically test for an overall “factor 1 effect” with the null hypothesis $H_0 : \tau_1 = \tau_2 = \tau_3 = \gamma_{11} = \gamma_{12} = \gamma_{21} = \gamma_{22} = \gamma_{31} = \gamma_{32} = 0$; this could be done with, e.g., a likelihood-ratio statistic. See the `anova` function in R for testing nested models.

Two-Way MANOVA: With \mathbf{X}_{lkr} the p -dimensional response vector for the r th observation of level l of factor 1 and level k of factor 2, the MANOVA model is

$$\mathbf{X}_{lkr} = \boldsymbol{\mu} + \boldsymbol{\tau}_l + \boldsymbol{\beta}_k + \boldsymbol{\gamma}_{lk} + \mathbf{e}_{lkr}$$

where $\sum_{l=1}^g \boldsymbol{\tau}_l = \sum_{k=1}^b \boldsymbol{\beta}_k = \sum_{l=1}^g \boldsymbol{\gamma}_{lk} = \sum_{k=1}^b \boldsymbol{\gamma}_{lk} = \mathbf{0}$ and the \mathbf{e}_{lkr} are i.i.d. $N_p(\mathbf{0}, \boldsymbol{\Sigma})$, $r = 1, 2, \dots, n$.

We can similarly decompose \mathbf{x}_{lkr} as

$$\mathbf{x}_{lkr} = \bar{\mathbf{x}} + (\bar{\mathbf{x}}_{l\cdot} - \bar{\mathbf{x}}) + (\bar{\mathbf{x}}_{\cdot k} - \bar{\mathbf{x}}) + (\bar{\mathbf{x}}_{lk} - \bar{\mathbf{x}}_{l\cdot} - \bar{\mathbf{x}}_{\cdot k} + \bar{\mathbf{x}}) + (\mathbf{x}_{lkr} - \bar{\mathbf{x}}_{lk})$$

Generalizing the sum of squares expression from two-way ANOVA, we have

$$\begin{aligned} \sum_{l=1}^g \sum_{k=1}^b \sum_{r=1}^n (\mathbf{x}_{lkr} - \bar{\mathbf{x}}) (\mathbf{x}_{lkr} - \bar{\mathbf{x}})' &= \sum_{l=1}^g bn (\bar{\mathbf{x}}_{l\cdot} - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_{l\cdot} - \bar{\mathbf{x}})' \\ &\quad + \sum_{k=1}^b gn (\bar{\mathbf{x}}_{\cdot k} - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_{\cdot k} - \bar{\mathbf{x}})' \\ &\quad + \sum_{l=1}^g \sum_{k=1}^b n (\bar{\mathbf{x}}_{lk} - \bar{\mathbf{x}}_{l\cdot} - \bar{\mathbf{x}}_{\cdot k} + \bar{\mathbf{x}}) (\bar{\mathbf{x}}_{lk} - \bar{\mathbf{x}}_{l\cdot} - \bar{\mathbf{x}}_{\cdot k} + \bar{\mathbf{x}})' \\ &\quad + \sum_{l=1}^g \sum_{k=1}^b \sum_{r=1}^n (\mathbf{x}_{lkr} - \bar{\mathbf{x}}_{lk}) (\mathbf{x}_{lkr} - \bar{\mathbf{x}}_{lk})' \end{aligned}$$

or

$$\text{SSP}_{\text{cor}} = \text{SSP}_{\text{fac 1}} + \text{SSP}_{\text{fac 2}} + \text{SSP}_{\text{int}} + \text{SSP}_{\text{res}}$$

| Source of variation | Sum of squares and crossproducts (SSP) | Degrees of freedom (d.f.) |
|------------------------|--|------------------------------|
| Factor 1 | $\text{SSP}_{\text{fac } 1} = \sum_{l=1}^g bn (\bar{\mathbf{x}}_{l\cdot} - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_{l\cdot} - \bar{\mathbf{x}})'$ | $g - 1$ |
| Factor 2 | $\text{SSP}_{\text{fac } 2} = \sum_{k=1}^b gn (\bar{\mathbf{x}}_{\cdot k} - \bar{\mathbf{x}}) (\bar{\mathbf{x}}_{\cdot k} - \bar{\mathbf{x}})'$ | $b - 1$ |
| Interaction | $\text{SSP}_{\text{int}} = \sum_{l=1}^g \sum_{k=1}^b n (\bar{\mathbf{x}}_{lk} - \bar{\mathbf{x}}_{l\cdot} - \bar{\mathbf{x}}_{\cdot k} + \bar{\mathbf{x}}) (\bar{\mathbf{x}}_{lk} - \bar{\mathbf{x}}_{l\cdot} - \bar{\mathbf{x}}_{\cdot k} + \bar{\mathbf{x}})'$ | $(g - 1)(b - 1)$ |
| Residual | $\text{SSP}_{\text{res}} = \sum_{l=1}^g \sum_{k=1}^b \sum_{r=1}^n (\mathbf{x}_{lkr} - \bar{\mathbf{x}}_{lk}) (\mathbf{x}_{lkr} - \bar{\mathbf{x}}_{lk})'$ | $gb(n - 1)$ |
| Total | $\text{SSP}_{\text{cor}} = \sum_{l=1}^g \sum_{k=1}^b \sum_{r=1}^n (\mathbf{x}_{lkr} - \bar{\mathbf{x}}) (\mathbf{x}_{lkr} - \bar{\mathbf{x}})'$ | $gbn - 1$ |

Inference for Two-Way MANOVA: As with two-way ANOVA, we can separately test for main effects of the two factors as well as for an interaction.

To test $H_0 : \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_g = \mathbf{0}$ (no factor 1 main effects), compute

$$\Lambda^* = \frac{|\text{SSP}_{\text{res}}|}{|\text{SSP}_{\text{fac } 1} + \text{SSP}_{\text{res}}|}$$

Then, if H_0 is true,

$$\left(\frac{(gb(n-1) - p + 1)/2}{(|(g-1) - p| + 1)/2} \right) \left(\frac{1 - \Lambda^*}{\Lambda^*} \right) \sim F_{\nu_1, \nu_2}$$

where $\nu_1 = |(g-1) - p| + 1$ and $\nu_2 = gb(n-1) - p + 1$.

Inference for Two-Way MANOVA: As with two-way ANOVA, we can separately test for main effects of the two factors as well as for an interaction.

To test $H_0 : \beta_1 = \beta_2 = \dots = \beta_b = \mathbf{0}$ (no factor 2 main effects), compute

$$\Lambda^* = \frac{|\text{SSP}_{\text{res}}|}{|\text{SSP}_{\text{fac } 2} + \text{SSP}_{\text{res}}|}$$

Then, if H_0 is true,

$$\left(\frac{(gb(n-1) - p + 1)/2}{(|(b-1) - p| + 1)/2} \right) \left(\frac{1 - \Lambda^*}{\Lambda^*} \right) \sim F_{\nu_1, \nu_2}$$

where $\nu_1 = |(b-1) - p| + 1$ and $\nu_2 = gb(n-1) - p + 1$.

Inference for Two-Way MANOVA: As with two-way ANOVA, we can separately test for main effects of the two factors as well as for an interaction.

To test $H_0 : \boldsymbol{\gamma}_{11} = \boldsymbol{\gamma}_{12} = \cdots = \boldsymbol{\gamma}_{gb} = \mathbf{0}$ (no interaction), compute

$$\Lambda^* = \frac{|\text{SSP}_{\text{res}}|}{|\text{SSP}_{\text{int}} + \text{SSP}_{\text{res}}|}$$

Then, if H_0 is true,

$$\left(\frac{(gb(n-1) - p + 1)/2}{(|((g-1)(b-1) - p| + 1)/2} \right) \left(\frac{1 - \Lambda^*}{\Lambda^*} \right) \sim F_{\nu_1, \nu_2}$$

where $\nu_1 = |(g-1)(b-1) - p| + 1$ and $\nu_2 = gb(n-1) - p + 1$.

Recall the previous caution about the interpretation of main effects in the presence of interaction. If there *is* an interaction, it arguably does not make much sense to test the main effects. Thus, in practice, the interaction is often tested first, and if it is statistically insignificant, the main effects are then tested. Note that there are large-sample approximate tests that can be used in place of the above exact tests; see the textbook for details.

Inference for Two-Way MANOVA: Bonferroni confidence intervals for contrasts can be constructed in the usual way. For example, $100(1 - \alpha)\%$ confidence intervals for $\tau_{li} - \tau_{mi}$ are

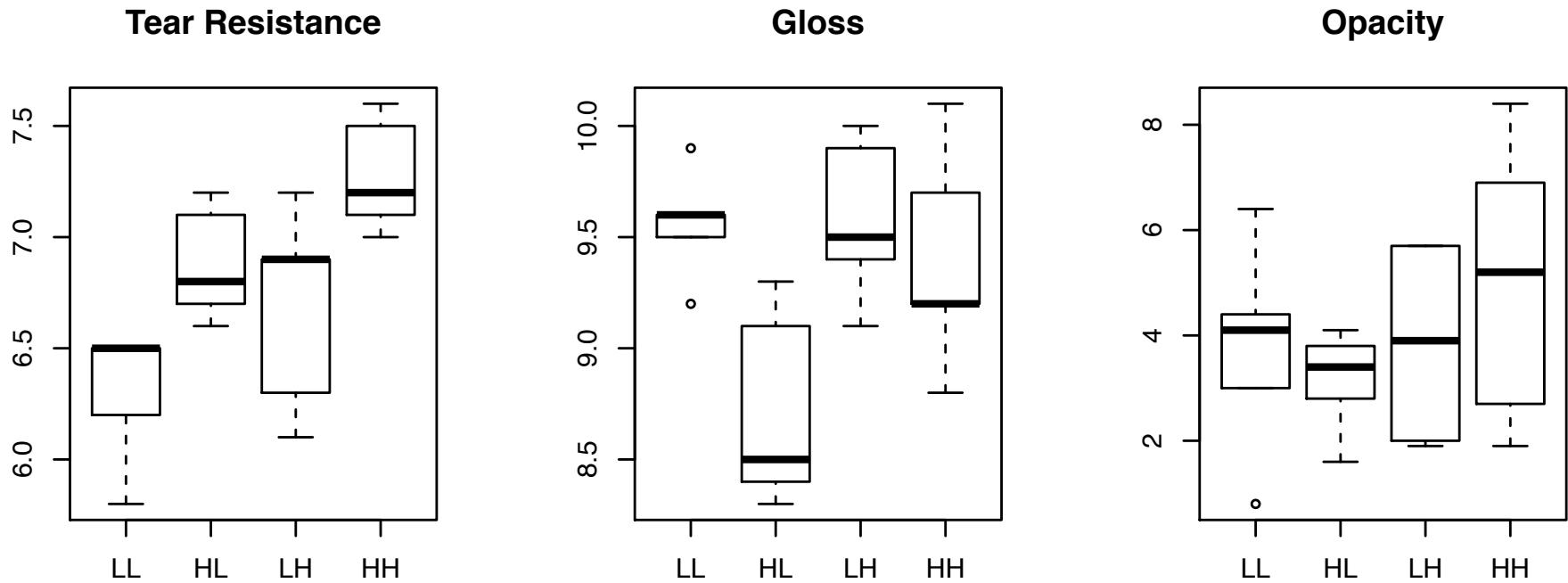
$$\bar{x}_{l.i} - \bar{x}_{m.i} \pm t_\nu \left(\frac{\alpha}{pg(g-1)} \right) \sqrt{\frac{E_{ii}}{\nu} \frac{2}{bn}}$$

where $\nu = gb(n-1)$, E_{ii} is the i th diagonal element of $\mathbf{E} = \text{SSP}_{\text{res}}$, and $\bar{x}_{l.i} - \bar{x}_{m.i}$ is the i th component of $\bar{\mathbf{x}}_l - \bar{\mathbf{x}}_m$. Similarly, $100(1 - \alpha)\%$ confidence intervals for $\beta_{ki} - \beta_{qi}$ are

$$\bar{x}_{.ki} - \bar{x}_{.qi} \pm t_\nu \left(\frac{\alpha}{pb(b-1)} \right) \sqrt{\frac{E_{ii}}{\nu} \frac{2}{gn}}$$

where ν and E_{ii} are as above and $\bar{x}_{.ki} - \bar{x}_{.qi}$ is the i th component of $\bar{\mathbf{x}}_{.k} - \bar{\mathbf{x}}_{.q}$. Intervals for the interaction effects can be constructed in the same way, as can intervals for the effects of one factor within a specified level of the other; for example, we could construct an interval for $(\tau_{li} + \gamma_{lk}) - (\tau_{mi} + \gamma_{mk})$ (the mean difference for component i , comparing levels l and m of factor 1, *within* level k of factor 2).

Example: Plastic Film: Data were collected on samples of plastic film during extrusion through a die. There are two factors, rate of extrusion and amount of an additive (both with two levels - low and high, so $g = b = 2$). Within each of the four factor level combinations, $n = 5$ plastic samples were observed, for which $p = 3$ variables were measured: X_1 = tear resistance, X_2 = gloss, and X_3 = opacity.

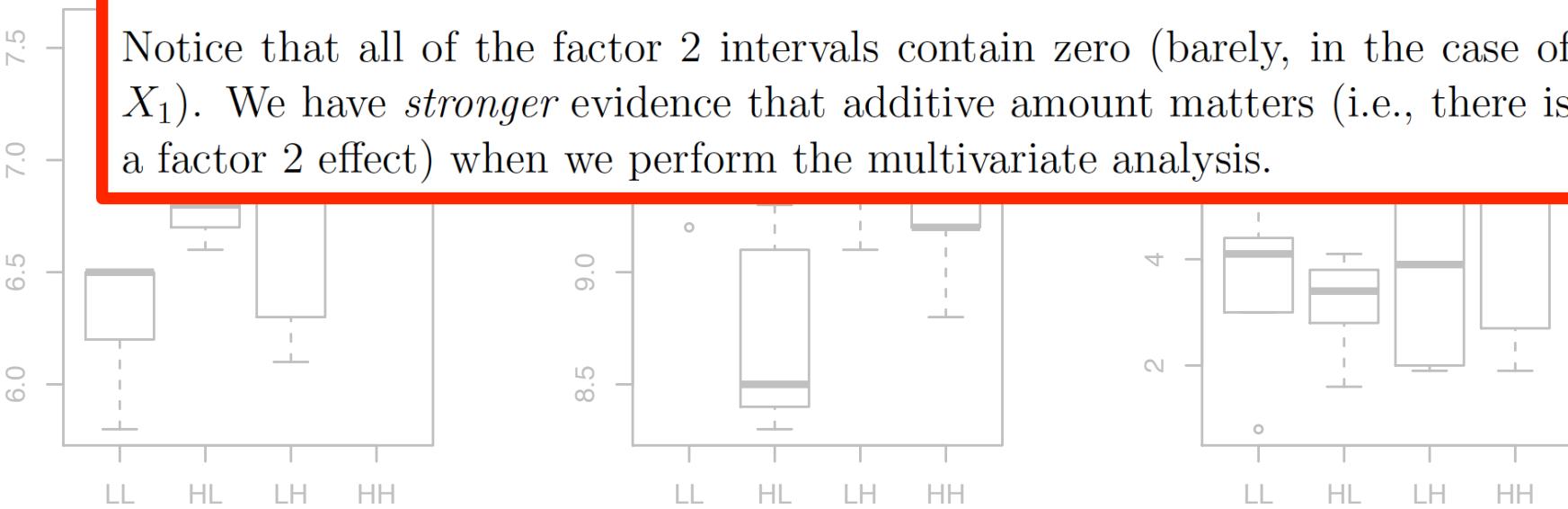


Data are from Table 6.4 in the book.

Example: Plastic Film: Data were collected on samples of plastic film during extrusion through a die. There are two factors, rate of extrusion and amount of additive. Within each factor there are two levels, low (LL) and high (HL). The interaction is not statistically significant (p -value = 0.302). The main effects for both factor 1 (p -value = 0.003) and factor 2 (p -value = 0.025) are. Bonferroni 95% confidence intervals are shown below.

| | X_1 | X_2 | X_3 |
|----------|------------------|-----------------|-----------------|
| Factor 1 | (-0.987, -0.193) | (0.026, 0.994) | (-2.698, 2.118) |
| Factor 2 | (-0.787, 0.007) | (-0.834, 0.134) | (-3.398, 1.418) |

Notice that all of the factor 2 intervals contain zero (barely, in the case of X_1). We have *stronger* evidence that additive amount matters (i.e., there is a factor 2 effect) when we perform the multivariate analysis.

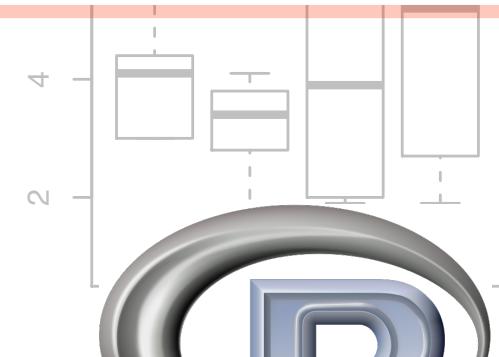
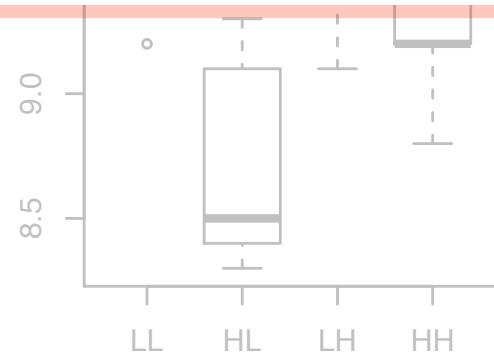
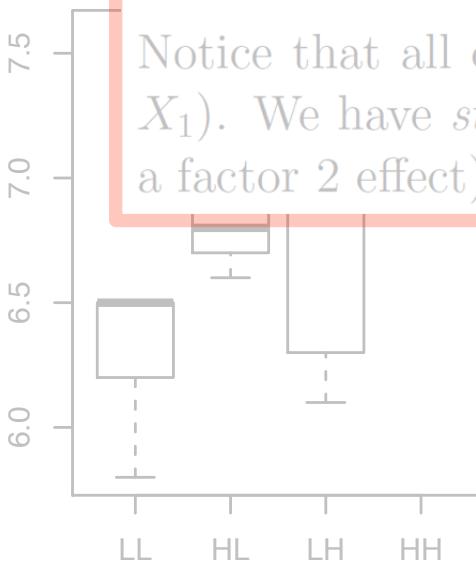


Example: Plastic Film: Data were collected on samples of plastic film during extrusion through a die. There are two factors, rate of extrusion and amount of additive. The interaction is not statistically significant (p -value = 0.302). The main effects for both factor 1 (p -value = 0.003) and factor 2 (p -value = 0.025) are significant. Within effects for both factor 1 (p -value = 0.003) and factor 2 (p -value = 0.025) are significant. The Bonferroni 95% confidence intervals are shown below.

= gloss

| | X_1 | X_2 | X_3 |
|----------|------------------|-----------------|-----------------|
| Factor 1 | (-0.987, -0.193) | (0.026, 0.994) | (-2.698, 2.118) |
| Factor 2 | (-0.787, 0.007) | (-0.834, 0.134) | (-3.398, 1.418) |

Notice that all of the factor 2 intervals contain zero (barely, in the case of X_1). We have *stronger* evidence that additive amount matters (i.e., there is a factor 2 effect) when we perform the multivariate analysis.



Data are from Table 6.4 in the book.

Profile Analysis

Profile Analysis: The question of equality of mean vectors can be broken into a series of hypotheses regarding their shape or *profile*. In the case of two populations with means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$:

1. Hypothesis 1: The profiles are parallel. This can be formulated as $H_{01} : \mu_{1i} - \mu_{1i-1} = \mu_{2i} - \mu_{2i-1}, i = 1, 2, \dots, p.$
2. Hypothesis 2: Assuming that the profiles are parallel, are they coincident (do they lie on top of each other)? This would mean $H_{02} : \mu_{1i} = \mu_{2i}, i = 1, 2, \dots, p.$
3. Hypothesis 3: Assuming that the profiles are coincident, are they level (are all means equal to the same constant)? This would mean $H_{03} : \mu_{11} = \mu_{12} = \dots = \mu_{2p}.$

Let $\mathbf{X}_{11}, \mathbf{X}_{12}, \dots, \mathbf{X}_{1n_1} \stackrel{\text{i.i.d.}}{\sim} N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$ and $\mathbf{X}_{21}, \mathbf{X}_{22}, \dots, \mathbf{X}_{2n_2} \stackrel{\text{i.i.d.}}{\sim} N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$ be independent samples from the two populations.

Test for Parallel Profiles: We can write the conditions of H_{01} as $\mathbf{C}\boldsymbol{\mu}_1 = \mathbf{C}\boldsymbol{\mu}_2$, where

$$\mathbf{C}_{((p-1) \times p)} = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

Using \mathbf{C} to transform the random vectors, we have that the $\mathbf{C}\mathbf{X}_{1j}$ are i.i.d. $N_{p-1}(\mathbf{C}\boldsymbol{\mu}_1, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}')$ and the $\mathbf{C}\mathbf{X}_{2j}$ are i.i.d. $N_{p-1}(\mathbf{C}\boldsymbol{\mu}_2, \mathbf{C}\boldsymbol{\Sigma}\mathbf{C}')$ (and the two samples are still independent of one another). To test H_{01} , we can now simply apply the usual two-sample T^2 test with equal covariances:

$$T^2 = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{C}' \left[\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{C} \mathbf{S}_{\text{pooled}} \mathbf{C}' \right]^{-1} \mathbf{C} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

Under H_{01} , T^2 is distributed as

$$\frac{(n_1 + n_2 - 2)(p - 1)}{n_1 + n_2 - p} F_{p-1, n_1 + n_2 - p}$$

Test for Coincident Profiles, Given That They Are Parallel: If the profiles are parallel, then $\mu_{1i} - \mu_{2i} = c$ for some constant c , $i = 1, 2, \dots, p$. Equivalently, we would have $\mathbf{1}'\boldsymbol{\mu}_1 - \mathbf{1}'\boldsymbol{\mu}_2 = pc$. Coincident profiles would mean that $c = 0$, so we can consider $H_{02} : \mathbf{1}'\boldsymbol{\mu}_1 = \mathbf{1}'\boldsymbol{\mu}_2$. Since the $\mathbf{1}'\mathbf{X}_{1j}$ are i.i.d. $N(\mathbf{1}'\boldsymbol{\mu}_1, \mathbf{1}'\boldsymbol{\Sigma}\mathbf{1})$ and the $\mathbf{1}'\mathbf{X}_{2j}$ are i.i.d. $N(\mathbf{1}'\boldsymbol{\mu}_2, \mathbf{1}'\boldsymbol{\Sigma}\mathbf{1})$, we can test H_{02} with the usual *univariate* two-sample t test with equal variances:

$$T^2 = \left(\frac{\mathbf{1}' (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)}{\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2} \right) \mathbf{1}' \mathbf{S}_{\text{pooled}} \mathbf{1}}} \right)^2$$

Under H_{02} , T^2 is distributed as $F_{1, n_1 + n_2 - 2}$.

Test for Level Profiles, Given That They Are Coincident: If the profiles are coincident, then $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \boldsymbol{\mu}$, and we can consider the \mathbf{X}_{1j} and \mathbf{X}_{2j} as constituting one random sample of size $n_1 + n_2$ from the $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution. The profiles are level if $\mu_1 = \mu_2 = \dots = \mu_p$, which we can formulate as $H_{03} : \mathbf{C}\boldsymbol{\mu} = \mathbf{0}$, where, e.g., \mathbf{C} is the contrast matrix defined before. With $\bar{\mathbf{x}}$ and \mathbf{S} the sample mean and covariance based on all $n_1 + n_2$ observations, we can test H_{03} by applying the usual one-sample T^2 test:

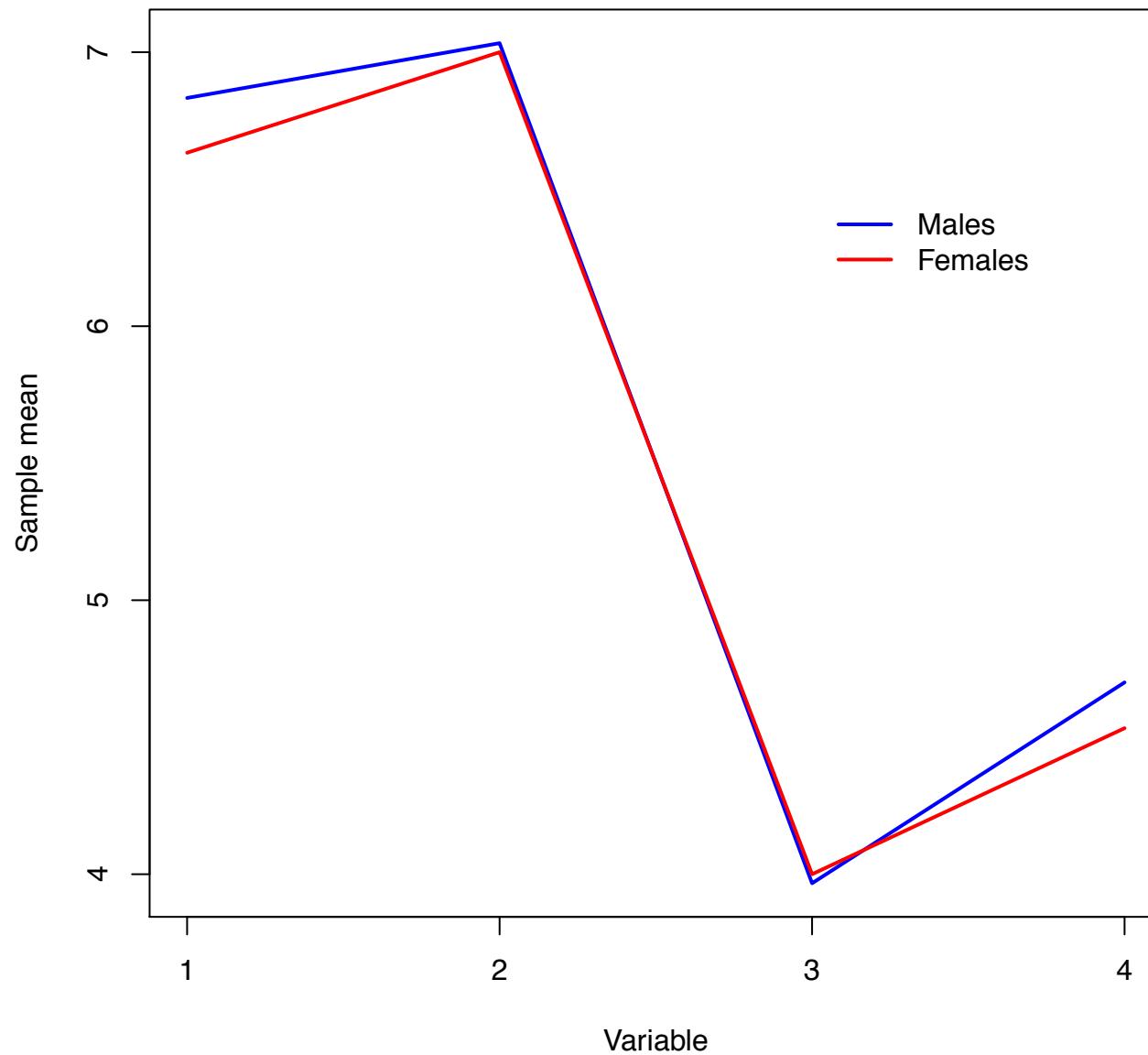
$$(n_1 + n_2) \bar{\mathbf{x}}' \mathbf{C}' [\mathbf{C} \mathbf{S} \mathbf{C}']^{-1} \mathbf{C} \bar{\mathbf{x}}$$

Under H_{03} , T^2 is distributed as

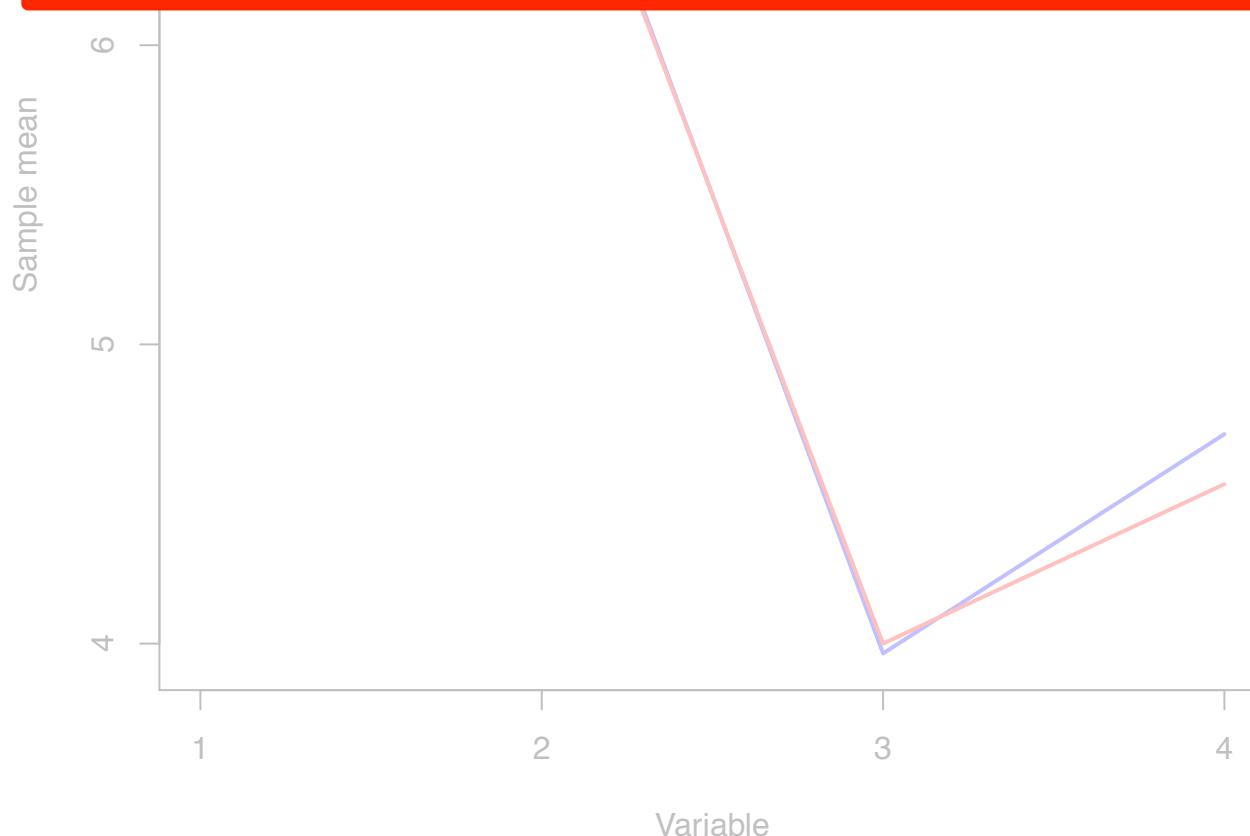
$$\frac{(n_1 + n_2 - 1)(p - 1)}{n_1 + n_2 - p + 1} F_{p-1, n_1+n_2-p+1}$$

Example: Love and Marriage Each of $n_1 = 30$ recently-married males and $n_2 = 30$ recently-married females were administered a survey with $p = 4$ questions: X_1 measures a subject's perceived *contributions* to the marriage on an 8-point scale, X_2 measures a subject's perceived *outcomes* from the marriage on an 8-point scale, X_3 measures a subject's level of *passionate love* for their partner on a 5-point scale, and X_4 measures a subject's level of *companionate love* for their partner on a 5-point scale. We have

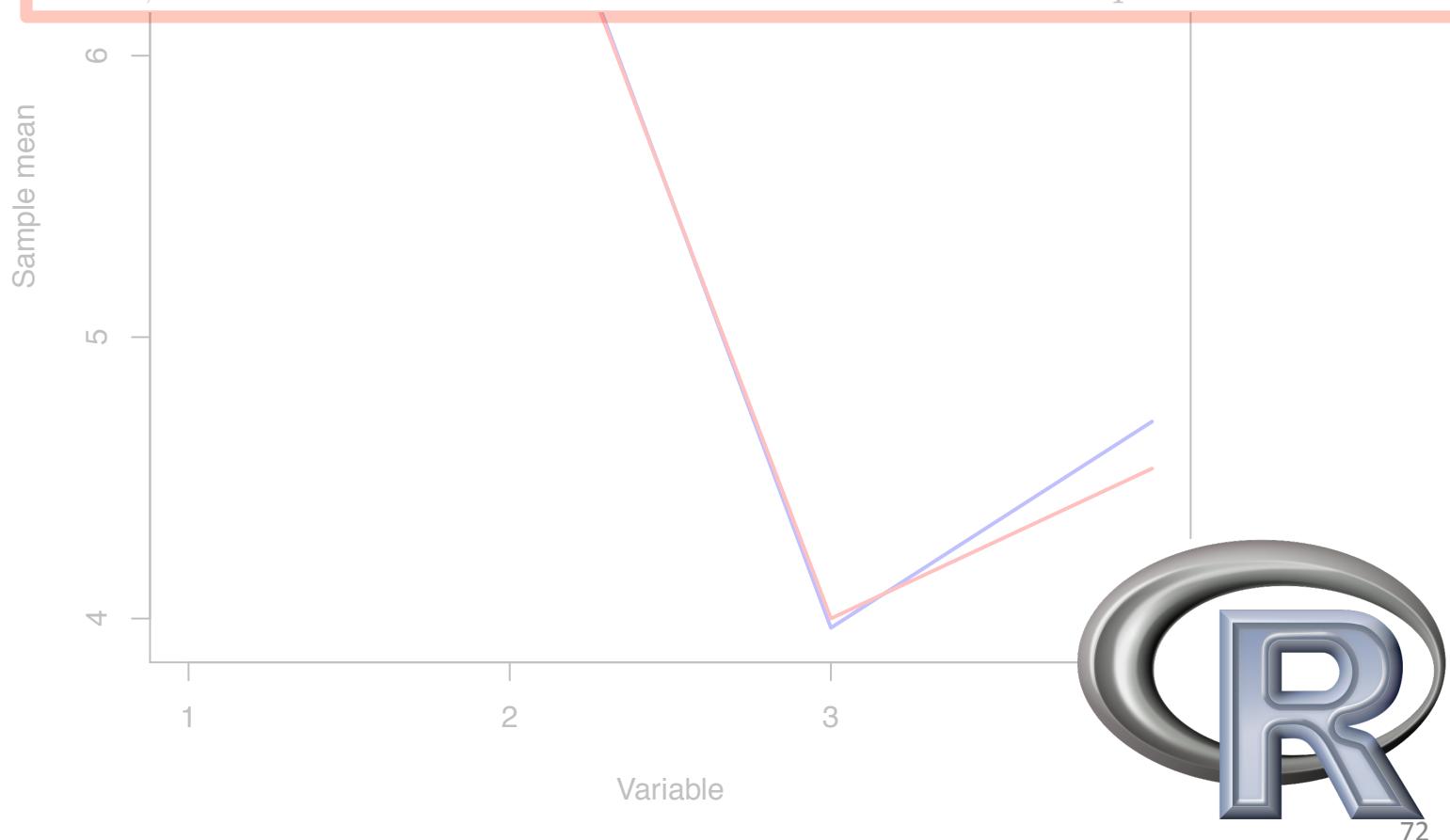
$$\bar{\mathbf{x}}_1 = \begin{bmatrix} 6.833 \\ 7.033 \\ 3.967 \\ 4.700 \end{bmatrix}, \quad \bar{\mathbf{x}}_2 = \begin{bmatrix} 6.633 \\ 7.000 \\ 4.000 \\ 4.533 \end{bmatrix}, \quad \mathbf{S}_{\text{pooled}} = \begin{bmatrix} 0.606 & 0.262 & 0.066 & 0.161 \\ 0.262 & 0.637 & 0.173 & 0.143 \\ 0.066 & 0.173 & 0.810 & 0.029 \\ 0.161 & 0.143 & 0.029 & 0.306 \end{bmatrix}$$



We do not reject H_{01} (p -value = 0.81), concluding that it is reasonable to view the profiles as parallel. We then also do not reject H_{02} (p -value = 0.48), concluding that it is reasonable to view the profiles as coincident. Since the variables were not all measured on the same scale, it does not make sense to test for whether the profiles are level.



We do not reject H_{01} (p -value = 0.81), concluding that it is reasonable to view the profiles as parallel. We then also do not reject H_{02} (p -value = 0.48), concluding that it is reasonable to view the profiles as coincident. Since the variables were not all measured on the same scale, it does not make sense to test for whether the profiles are level.



Repeated Measures and Growth Curves

Growth Curves: Recall that in a repeated-measures design, the same characteristic is observed at multiple times or locations on the same subject. When measurements are taken over time, we refer to the mean pattern over time as a *growth curve*.

Let $\mathbf{X}_{l1}, \mathbf{X}_{l2}, \dots, \mathbf{X}_{ln_l} \stackrel{\text{i.i.d.}}{\sim} N_p(\boldsymbol{\mu}_l, \boldsymbol{\Sigma})$ be a random sample from the l th population, $l = 1, 2, \dots, g$, where the elements of \mathbf{X}_{lj} consist of measurements of the same characteristic at times t_1, t_2, \dots, t_p . We could test the relationships between the g growth curves using profile analysis. Alternatively, in order to allow for a smooth relationship between the means over time, we could specify, e.g., a q th-order *polynomial* growth curve model:

$$\begin{aligned} E(\mathbf{X}_{lj}) &= \begin{bmatrix} \beta_{l0} + \beta_{l1}t_1 + \beta_{l2}t_1^2 + \cdots + \beta_{lq}t_1^q \\ \beta_{l0} + \beta_{l1}t_2 + \beta_{l2}t_2^2 + \cdots + \beta_{lq}t_2^q \\ \vdots \quad \vdots \quad \vdots \quad \ddots \quad \vdots \\ \beta_{l0} + \beta_{l1}t_p + \beta_{l2}t_p^2 + \cdots + \beta_{lq}t_p^q \end{bmatrix} \\ &= \begin{bmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^q \\ 1 & t_2 & t_2^2 & \cdots & t_2^q \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & t_p & t_p^2 & \cdots & t_p^q \end{bmatrix} \begin{bmatrix} \beta_{l0} \\ \beta_{l1} \\ \vdots \\ \beta_{lq} \end{bmatrix} = {}_{(p \times (q+1))((q+1) \times 1)}^{\mathbf{B}} \boldsymbol{\beta}_l \end{aligned}$$

One natural null hypothesis to test is then $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_g$, that all of the comparison groups have the same growth curve.

Estimation and Inference: The maximum likelihood estimators are

$$\hat{\beta}_l = (\mathbf{B}' \mathbf{S}_{\text{pooled}}^{-1} \mathbf{B})^{-1} \mathbf{B}' \mathbf{S}_{\text{pooled}}^{-1} \bar{\mathbf{X}}_l, \quad l = 1, 2, \dots, g$$

where, with $n = n_1 + n_2 + \dots + n_g$,

$$\mathbf{S}_{\text{pooled}} = \frac{1}{n-g} ((n_1-1)\mathbf{S}_1 + (n_2-1)\mathbf{S}_2 + \dots + (n_g-1)\mathbf{S}_g)$$

The covariance of the m.l.e.s is

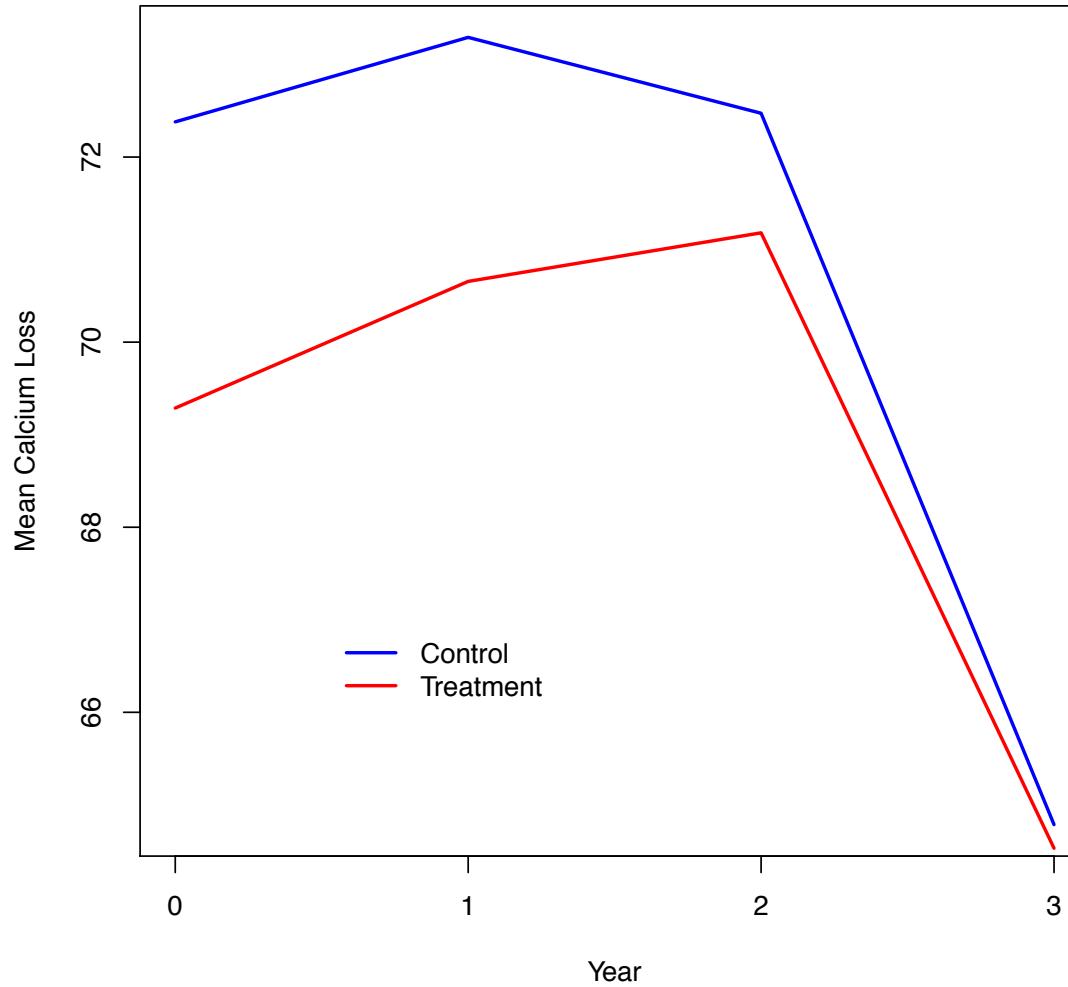
$$\text{Cov}(\hat{\beta}_l) = \frac{k}{n_l} (\mathbf{B}' \boldsymbol{\Sigma}^{-1} \mathbf{B})^{-1}$$

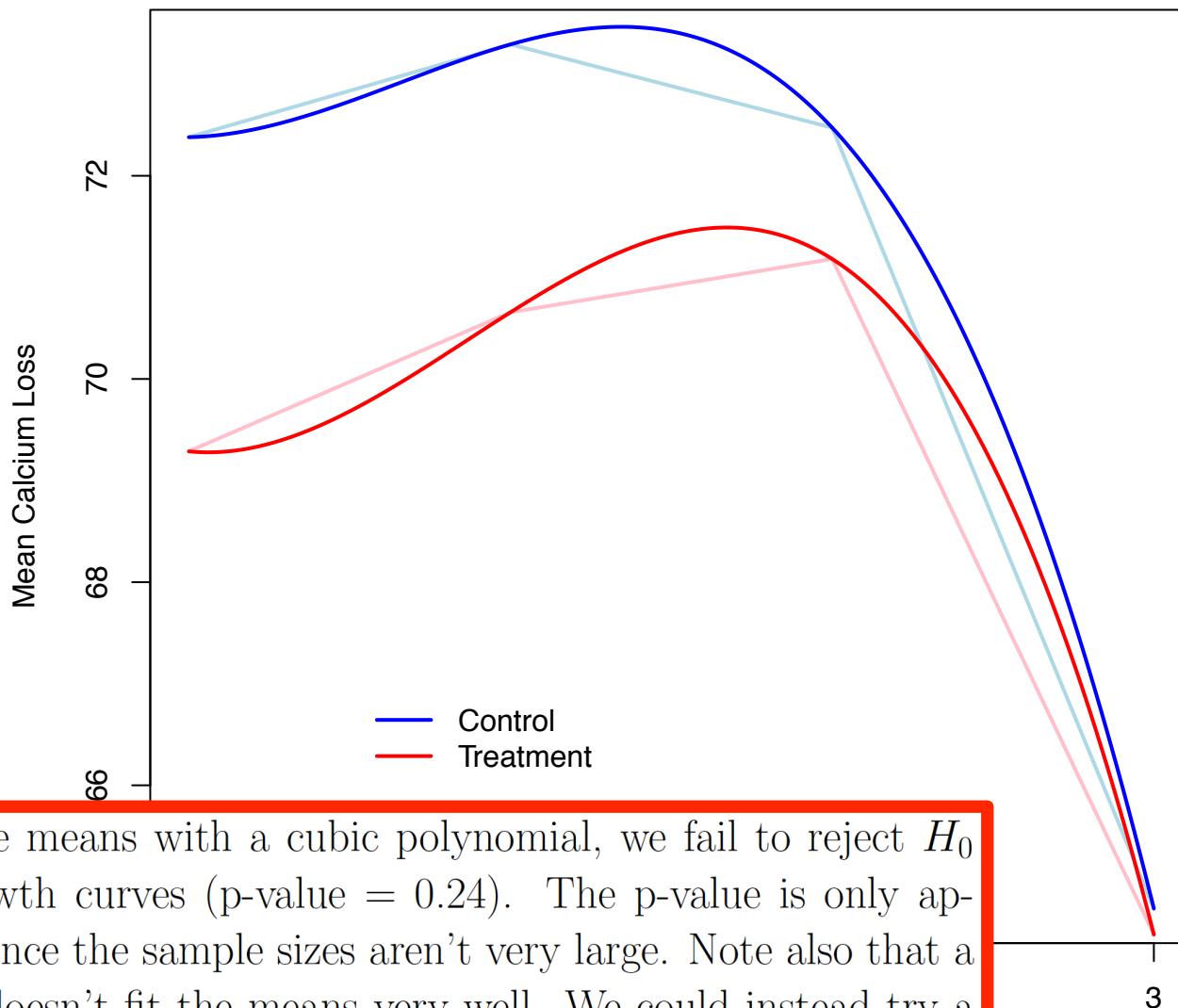
where $k = ((n-1)(n-g-1))/((n-g-p+q)(n-g-p+q+1))$, and we can estimate $\text{Cov}(\hat{\beta}_l)$ by plugging in $\mathbf{S}_{\text{pooled}}$ for $\boldsymbol{\Sigma}$. The m.l.e.s are themselves multivariate normal random variables and are independent of one another. Thus, when $g = 2$, we can test $H_0 : \beta_1 = \beta_2$ with a two-sample T^2 statistic:

$$T^2 = (\hat{\beta}_1 - \hat{\beta}_2)' \left[k \left(\frac{1}{n_1} + \frac{1}{n_2} \right) (\mathbf{B}' \mathbf{S}_{\text{pooled}}^{-1} \mathbf{B})^{-1} \right]^{-1} (\hat{\beta}_1 - \hat{\beta}_2)$$

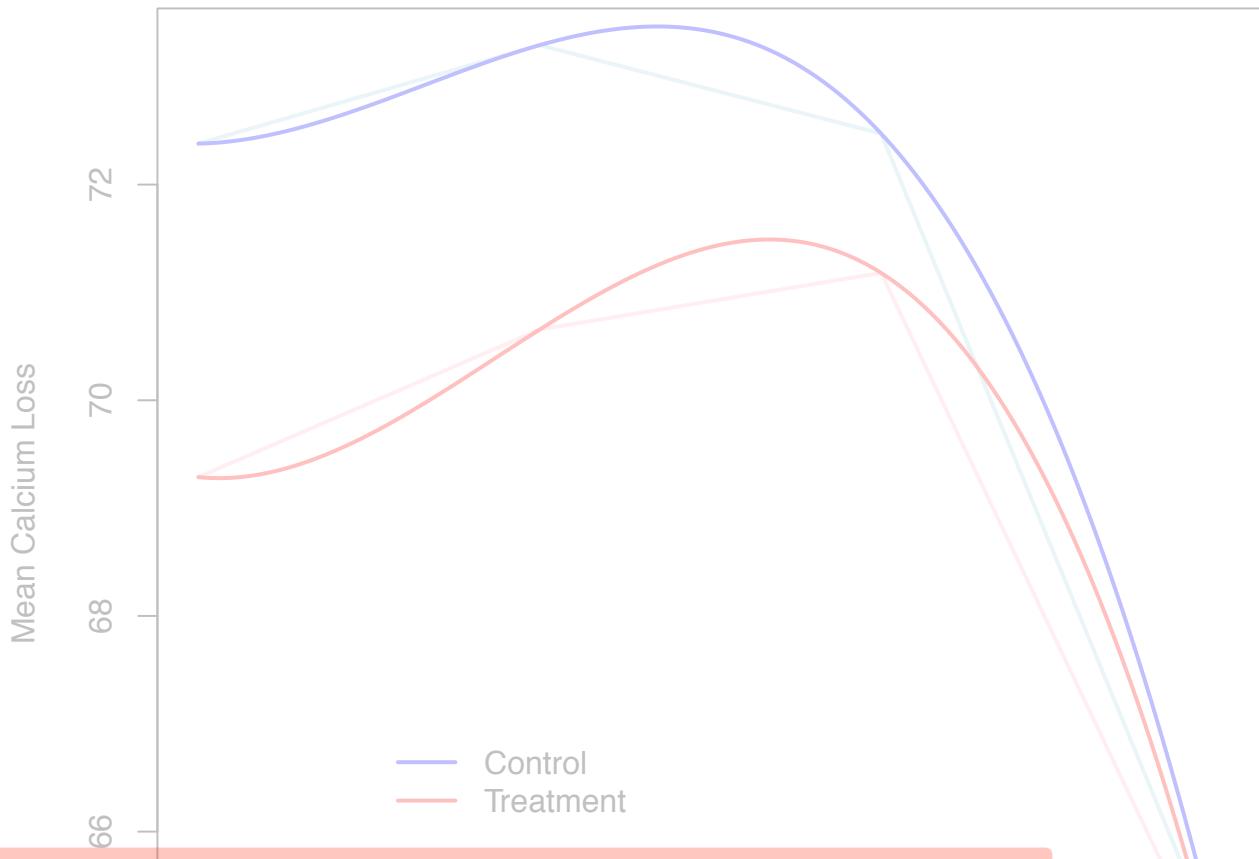
With n_1 and n_2 large, T^2 has an approximate χ_{q+1}^2 distribution under H_0 .

Example: Calcium Loss For $n_1 = 15$ subjects in a control group and $n_2 = 16$ subjects in a treatment group, we have measurements of calcium density loss in the dominant ulna bone at $p = 4$ time points (baseline, year 1, year 2, and year 3).





Modeling the means with a cubic polynomial, we fail to reject H_0 of equal growth curves ($p\text{-value} = 0.24$). The p -value is only approximate, since the sample sizes aren't very large. Note also that a polynomial doesn't fit the means very well. We could instead try a more flexible, non-parametric smoothing method like splines.



Modeling the means with a cubic polynomial, we fail to reject H_0 of equal growth curves ($p\text{-value} = 0.24$). The p -value is only approximate, since the sample sizes aren't very large. Note also that a polynomial doesn't fit the means very well. We could instead try a more flexible, non-parametric smoothing method like splines.



Multivariate Linear Regression

Linear Regression Review: With Y_j a univariate response on subject j , for which covariates $z_{j1}, z_{j2}, \dots, z_{jr}$ have been measured, $j = 1, 2, \dots, n$, the linear regression model is

$$Y_j = \beta_0 + \beta_1 z_{j1} + \beta_2 z_{j2} + \cdots + \beta_r z_{jr} + \epsilon_j$$

where the ϵ_j are i.i.d. random error terms with mean 0 and variance σ^2 ; often, the ϵ_j are assumed to be normally distributed, but this is not necessary with large n . In matrix form, we can write

$$\underset{(n \times 1)}{\mathbf{Y}} = \underset{(n \times (r+1))}{\mathbf{Z}} \underset{((r+1) \times 1)}{\boldsymbol{\beta}} + \underset{(n \times 1)}{\boldsymbol{\epsilon}}$$

The covariates are treated as known, so the unknown parameters of the model are $\boldsymbol{\beta}$ and σ^2 . Estimation is carried out by choosing the value of $\hat{\boldsymbol{\beta}}$ that minimizes the sum of squared differences between the y_j and their fitted values:

$$\hat{\boldsymbol{\beta}} = \underset{\mathbf{b}}{\operatorname{argmin}} (\mathbf{y} - \mathbf{Z}\mathbf{b})' (\mathbf{y} - \mathbf{Z}\mathbf{b}) = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{y}$$

We have $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and

$$\operatorname{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{Z}'\mathbf{Z})^{-1}$$

Furthermore, with large n (or, assuming the population distribution is normal), $\hat{\boldsymbol{\beta}}$ can be assumed to be a multivariate normal random vector. We can estimate σ^2 as the (appropriately-scaled) sample variance of the residuals:

$$\hat{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\beta}})' (\mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\beta}})}{n - (r + 1)}$$

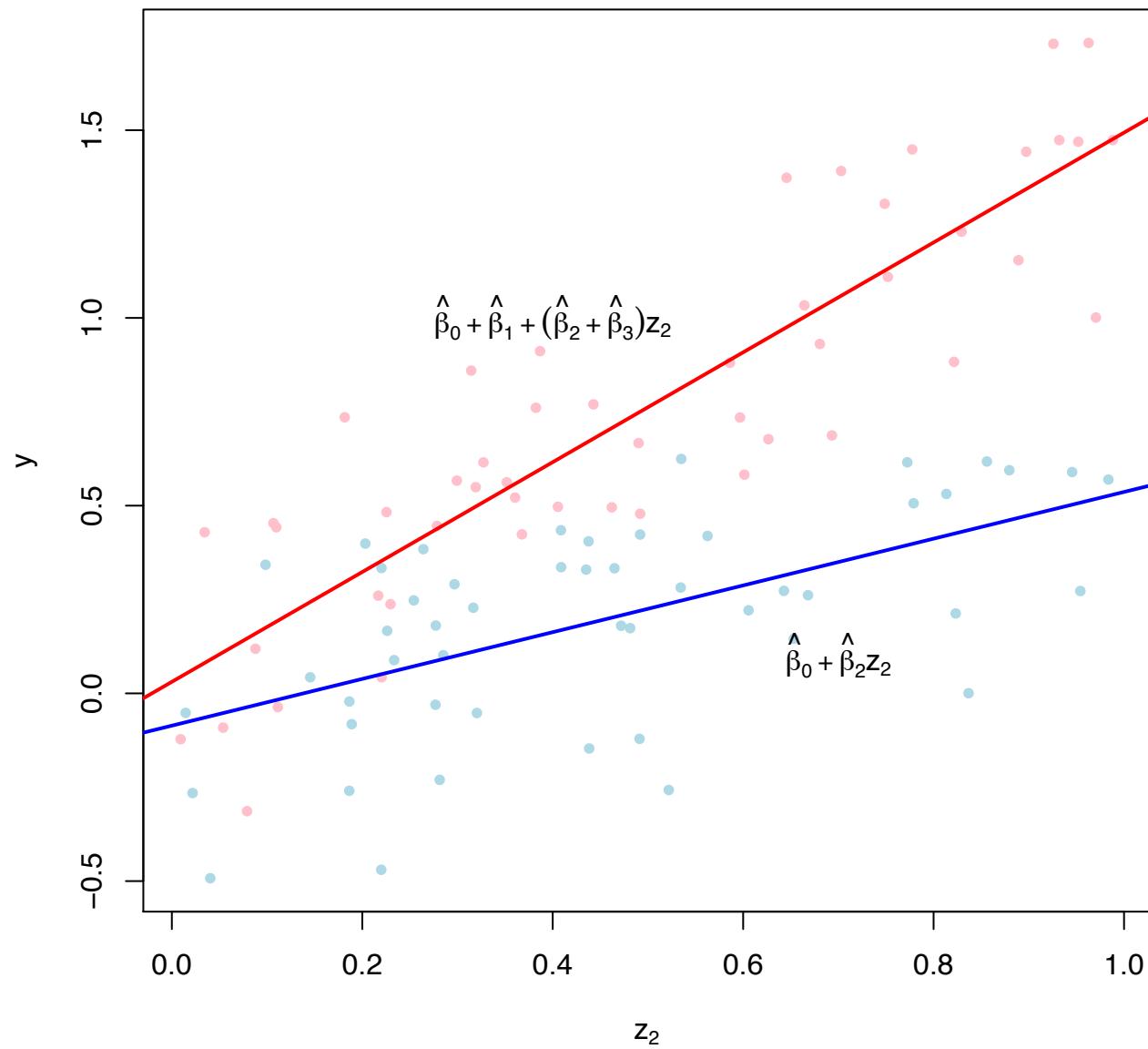
Example: Suppose we have $n = 100$ subjects, 50 each in two comparison groups, for which a univariate response and a single continuous covariate were observed. Let \mathbf{z}_1 be the vector of group indicators (equal to 1 if a subject is in group 2, 0 otherwise). And let \mathbf{z}_2 be the vector containing the covariate value for each subject. Consider the linear regression model

$$Y_j = \beta_0 + \beta_1 z_{1j} + \beta_2 z_{2j} + \beta_3 (z_{1j} \times z_{2j}) + \epsilon_j$$

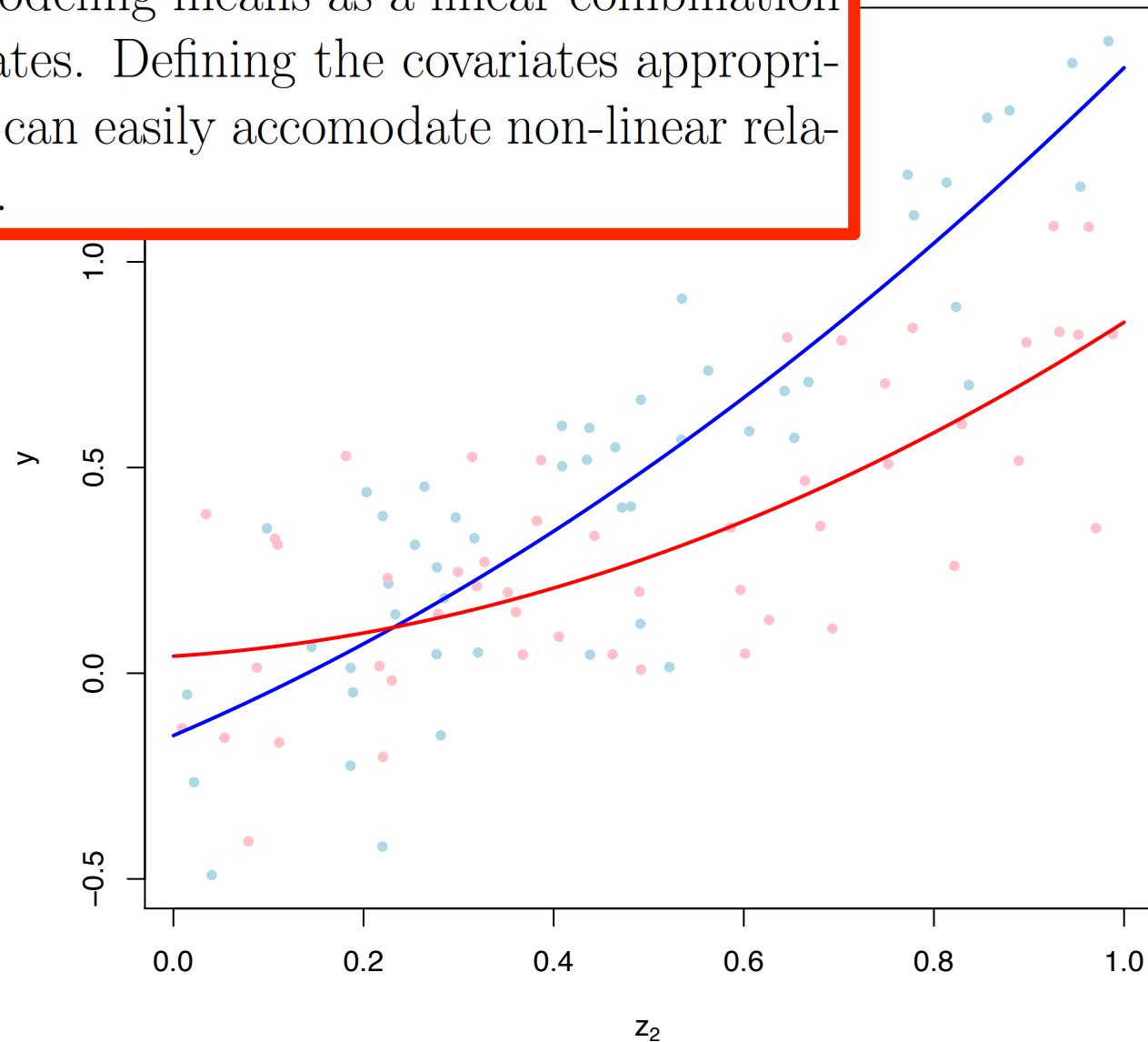
We have the following coefficient interpretations:

- β_0 : Mean response when $z_1 = 0$ and $z_2 = 0$ (the “ y -intercept” of the line in group one).
- $\beta_0 + \beta_1$: Mean response when $z_1 = 1$ and $z_2 = 0$ (the “ y -intercept” of the line in group two).
- β_2 : Mean change in response associated with a one-unit increase in z_2 among subjects in group one (the slope of the line in group one).
- $\beta_2 + \beta_3$: Mean change in response associated with a one-unit increase in z_2 among subjects in group two (the slope of the line in group two).

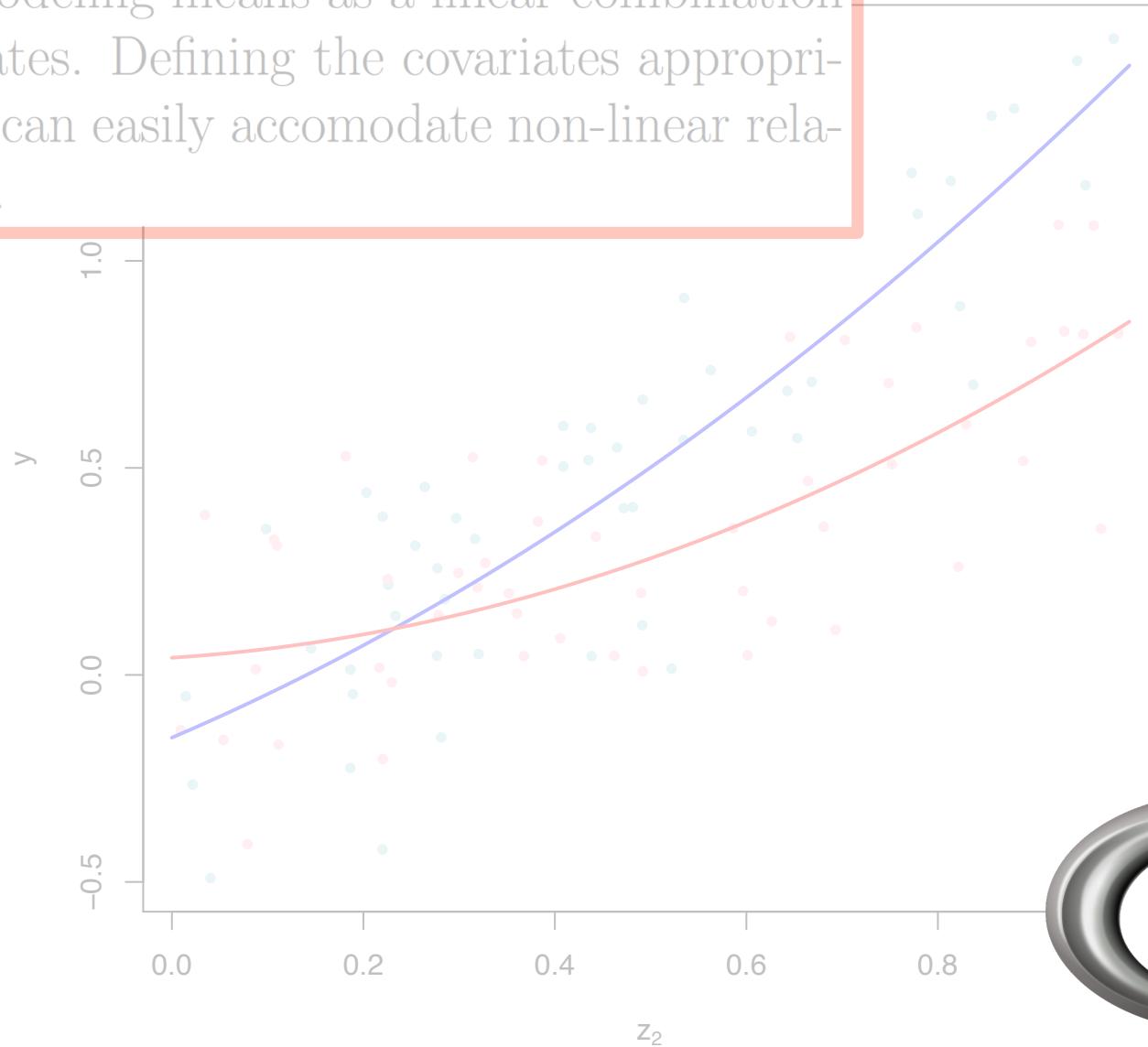
Example inferential applications: (1) test $H_0 : \beta_3 = 0$ (is there an interaction?); (2) test $H_0 : \beta_2 = \beta_3 = 0$ (is there an overall linear relationship between mean response and z_2 ?); (3) construct a confidence interval for $\beta_0 + \beta_1 + 0.5(\beta_2 + \beta_3)$ (mean response of a subject in group 2 when $z_2 = 0.5$).



The “linear” in linear regression just means that we are modeling means as a linear combination of covariates. Defining the covariates appropriately, we can easily accomodate non-linear relationships.



The “linear” in linear regression just means that we are modeling means as a linear combination of covariates. Defining the covariates appropriately, we can easily accomodate non-linear relationships.



ANOVA as Linear Regression: Recall the one-way ANOVA model

$$X_{lj} = \mu + \tau_l + e_{lj}, \quad l = 1, 2, \dots, g \text{ and } j = 1, 2, \dots, n_l$$

with $\sum_{l=1}^g \tau_l = 0$ and the e_{lj} assumed to be i.i.d. $N(0, \sigma^2)$ random errors.

Define $\mathbf{Y}' = [X_{11}, X_{12}, \dots, X_{1n_1}, X_{21}, \dots, X_{gn_g}]$ as the vector of all $n = \sum_{l=1}^g n_l$ observations. Also define the “dummy variables” z_{lj} to be indicators for whether subject j belongs to population l (equal to 1 if yes, 0 otherwise). Now consider the linear regression model

$$Y_j = \beta_0 + \beta_1 z_{1j} + \beta_2 z_{2j} + \cdots + \beta_g z_{gj} + \epsilon_j$$

We have $\beta = \mu$ and $\beta_l = \tau_l$, $l = 1, 2, \dots, g$, with the constraint that $\sum_{l=1}^g \beta_l = 0$. Thus, ANOVA can be formulated as a special case of linear regression.

Multivariate Linear Regression: In the multivariate case, the response is a random vector. With m variables, r covariates, and a sample size of n , the model becomes

$$\mathbf{Y}_j = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 z_{j1} + \boldsymbol{\beta}_2 z_{j2} + \cdots + \boldsymbol{\beta}_r z_{jr} + \boldsymbol{\epsilon}_j$$

where the $\boldsymbol{\epsilon}_j$ are independent with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}$. Define $\mathbf{Y}_{(i)}$ to be the n -vector of observations on variable i , $i = 1, 2, \dots, m$. Similarly, $\boldsymbol{\beta}_{(i)}$ is the vector of regression coefficients for this variable, and $\boldsymbol{\epsilon}_{(i)}$ is the vector of errors. Now, with

$$\begin{aligned}\mathbf{Y}_{(n \times m)} &= [\mathbf{Y}_{(1)} \mid \mathbf{Y}_{(2)} \mid \cdots \mid \mathbf{Y}_{(m)}] & \boldsymbol{\beta}_{((r+1) \times m)} &= [\boldsymbol{\beta}_{(1)} \mid \boldsymbol{\beta}_{(2)} \mid \cdots \mid \boldsymbol{\beta}_{(m)}] \\ \boldsymbol{\epsilon}_{(n \times m)} &= [\boldsymbol{\epsilon}_{(1)} \mid \boldsymbol{\epsilon}_{(2)} \mid \cdots \mid \boldsymbol{\epsilon}_{(m)}]\end{aligned}$$

we can write the model as

$$\mathbf{Y}_{(n \times m)} = \mathbf{Z}_{(n \times (r+1))((r+1) \times m)} \boldsymbol{\beta}_{((r+1) \times m)} + \boldsymbol{\epsilon}_{(n \times m)}$$

where $E(\boldsymbol{\epsilon}_{(i)}) = \mathbf{0}$ and $\text{Cov}(\boldsymbol{\epsilon}_{(i)}, \boldsymbol{\epsilon}_{(k)}) = \sigma_{ik} \mathbf{I}$, $i, k = 1, 2, \dots, m$. The m observations on subject j have covariance matrix $\boldsymbol{\Sigma} = \{\sigma_{ik}\}$, but observations from different subjects are uncorrelated.