Chapter 7: Grouped Multivariate Data

- In some situations, we have multivariate data that are known to come from two or more populations.
- For example, we may have test score data for Americans and similar test score data for Europeans.
- In Fisher's iris data set, measurements were made on irises from three different species.
- It may be of interest to compare the mean vectors between two populations or across several populations.
- With univariate data (one variable per individual), when the data are normally distributed, the *t-test* is used to compare two means, and the *analysis of variance* (ANOVA) F-test is used to compare three or more means.
- With multivariate data (two or more variables per individual), we have analogous tests, the Hotelling's T^2 test and the *multivariate analysis of variance* (MANOVA).

Hotelling's T^2 Test

- Hotelling's T^2 Test is a formal test of whether the $q \times 1$ mean vectors of two populations are equal, i.e.:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 \quad vs. \quad H_a : \boldsymbol{\mu}_1 \neq \boldsymbol{\mu}_2$$

- Under the null hypothesis, the expected values of each variable $X_1, X_2, \ldots, X_q$ are equal for the two populations.

- If one or more components of the mean vectors differ from population 1 to population 2, then H_a is true.

Assumptions of Hotelling's T^2 Test

- We assume we have two random samples, one of n_1 individuals (from the first population) and the other of n_2 individuals (from the second population).
- We further assume the data from each population are multivariate normal and the two populations have the same covariance matrix.
- If the sample sizes n_1 and n_2 are large, the Hotelling test is *approximately correct* even if the populations are not multivariate normal.
- If the covariance matrices are not equal for the two populations, a large-sample procedure based on the χ^2 distribution is available (see Johnson and Wichern, 2002, p. 291).

Relationship to the Two-Sample t-test for Univariate Data

- When we take independent univariate samples from two populations (having a common variance) and test $H_0 : \mu_1 = \mu_2$, we use the test statistic

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

where S_p^2 is the pooled sample variance

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

- The form of the test statistic T^2 in the multivariate situation is somewhat similar.

Hotelling's T^2 Test Statistic

- The Hotelling's test statistic is

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)' \mathbf{S}^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)$$

  where $\mathbf{S}$ is a "pooled" estimate of the common covariance matrix, i.e.:

$$\mathbf{S} = \frac{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}{n_1 + n_2 - 2}$$

- If the assumptions of the test are met, then under H_0, a rescaled version of the test statistic

$$F = \frac{(n_1 + n_2 - q - 1)T^2}{(n_1 + n_2 - 2)q}$$

  has an F-distribution with q numerator degrees of freedom and $n_1 + n_2 - q - 1$ denominator d.f.

- Hence the P-value of the test is the area to the right of F in that F-distribution.

Inference about a Single Multivariate Mean Vector

- In univariate statistics, we have a one-sample t-test or CI about a single population mean.
- With multivariate data, in addition to comparing two mean vectors, we could perform a hypothesis test about a single mean vector $\boldsymbol{\mu}$.
- Another Hotelling T^2-type statistic can be used to test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ vs. $H_a : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, where $\boldsymbol{\mu}_0$ is some *pre-specified* vector (some *hypothesized* mean vector).
- We could also generate a q-dimensional *confidence region* about the population mean vector $\boldsymbol{\mu}$.
- These inferences assume that the data come from a multivariate normal population.
- The inferences will be *approximately correct* when the data are not multivariate normal, but the sample size n is large.

Hotelling's Test about a Single Multivariate Mean Vector

- The one-sample version of Hotelling's T^2 statistic is

$$T^2 = n(\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)' \mathbf{S}^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{\mu}_0)$$

- We reject $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ in favor of $H_a : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$ if $\frac{(n-q)}{(n-1)q} T^2 > F_{q,n-q,\alpha}$.
- The P-value is the area to the right of $\frac{(n-q)}{(n-1)q} T^2$ in the $F_{q,n-q}$ distribution.
- Furthermore, a q-dimensional *confidence region* about the population mean vector $\boldsymbol{\mu}$ is defined by all $\boldsymbol{\mu}$ such that

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

- If $q = 2$, this is an ellipse in the 2-dimensional plane, centered at $\bar{\boldsymbol{x}}$.
- If $q > 2$, this is an *ellipsoid* (hyper-ellipse).

Relationship between Univariate and Multivariate Tests

- Does the multivariate Hotelling's T^2 test always yield the same conclusions as doing a series of individual univariate t-tests for each variable or each comparison?
- No. Sometimes each univariate t-test is not significant, yet the multivariate T^2 test is significant.
- In other cases, the multivariate T^2 test is not significant, yet one or more of the univariate t-tests are significant.
- Even if we use Bonferroni corrections on the series of individual t-tests, this discrepancy still occurs.
- This is because the individual t-tests do not account for the correlations between the variables, but the multivariate T^2 test does.
- See graphical example with SAT scores.

Chapter 7, continued: MANOVA

- The Multivariate Analysis of Variance (MANOVA) technique extends Hotelling T^2 test that compares two mean vectors to the setting in which there are $m \geq 2$ groups.
- We wish to compare the mean vectors across all m groups.
- For example, recall that in Fisher's iris data set, four measurements were made on irises from three different species.
- We could test whether the 4×1 mean vectors were equal for the three species (setosa, versicolor, virginica).
- Another example: Comparing the mean vectors (for several skull measurements) across the five epochs from which the skulls were found.
- MANOVA extends the ordinary ANOVA to the case in which $q \geq 2$ variables are measured on each observation.

The MANOVA Model

- If X_{ijk} represents the j-th observation on variable k in the i-th group ($k = 1, \ldots, q,\ j = 1, \ldots, n_i,\ i = 1, \ldots, m$), then the ANOVA model decomposes X_{ijk} as:

$$X_{ijk} = \mu_k + \alpha_{ik} + \epsilon_{ijk}$$

- Here, μ_k is a general effect associated with the k-th variable, α_{ik} is the effect of group i on variable k, and ϵ_{ijk} is a random error term.

- In vector form, this is

$$\mathbf{X}_{ij} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_{ij}, j = 1, \ldots, n_i,\ i = 1, \ldots, m$$

  where these are each $q \times 1$ vectors.

- Similarly to the Hotelling test, we assume the vector $\boldsymbol{\epsilon}_{ij} = \left(\epsilon_{ij1}, \ldots, \epsilon_{ijq}\right)'$ is multivariate normal with mean vector $\mathbf{0}$ and common covariance matrix $\boldsymbol{\Sigma}$ across all m populations.

- In practice, the normality assumption can be checked via a chi-squared plot of the residual vectors $\mathbf{x}_{ij} - \bar{\boldsymbol{x}}_i$.

The MANOVA Null Hypothesis

- The mean vector for group i can be written as:

$$\boldsymbol{\mu}_i = (\mu_1 + \alpha_{i1}, \ldots, \mu_q + \alpha_{iq})'$$

- Therefore, the MANOVA null hypothesis of equal mean vectors across the groups, i.e.,

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_m$$

  can also be written as

$$H_0 : \alpha_{ik} = 0, i = 1, \ldots, m,\ k = 1, \ldots, q.$$

- The alternative hypothesis is that at least two of the groups' mean vectors differ, i.e., $\boldsymbol{\mu}_i \neq \boldsymbol{\mu}_{i'}$ for some $i \neq i'$.

Between-Groups and Within-Groups Matrices

- The MANOVA test statistic is based on two matrices, $\mathbf{H}$ and $\mathbf{E}$:

$$\mathbf{H} = \sum_{i=1}^{m} n_i(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_i - \bar{\boldsymbol{x}})', \quad \mathbf{E} = \sum_{i=1}^{m}\sum_{j=1}^{n_i}(\mathbf{x}_{ij} - \bar{\boldsymbol{x}}_i)(\mathbf{x}_{ij} - \bar{\boldsymbol{x}}_i)'$$

- The matrix $\mathbf{H}$ contains the between-groups sum of squares for each variable along its diagonal, and the between-groups sum of cross-products for the variables in its off-diagonal elements.
- The matrix $\mathbf{E}$ contains the within-groups sum of squares for each variable along its diagonal, and the within-groups sum of cross-products for the variables in its off-diagonal elements.
- $\mathbf{H}$ is the generalization of the treatment sum of squares (SSTR) in ordinary ANOVA, and $\mathbf{E}$ is the generalization of the error sum of squares (SSE) in ordinary ANOVA.
- The matrix $\mathbf{H} + \mathbf{E}$ contains the total (corrected) sum of squares and cross products for the entire data set.

Test Statistic

- Unlike in ordinary ANOVA, where $F = MSTR/MSE$ gives the most powerful test of the ANOVA hypothesis, *no one test statistic* is uniformly most powerful in testing the MANOVA null hypothesis.
- The most common test statistic is Wilks' Lambda, a ratio of determinants:

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{H} + \mathbf{E}|}$$

- Wilks' Lambda is related to the likelihood ratio criterion for this situation.
- A statistic that is a decreasing function of Λ has an F-distribution under H_0.
- It corresponds to the F-statistic in the univariate situation and can be used via an F-test (or, for large samples, an approximate χ^2 test).

Other Test Statistics

- Other proposed test statistics include Roy's greatest root (the largest eigenvalue of $\mathbf{E}^{-1}\mathbf{H}$), the Lawley-Hotelling trace $tr(\mathbf{E}^{-1}\mathbf{H})$, or the Pillai trace $tr(\mathbf{H}(\mathbf{H}+\mathbf{E})^{-1})$.
- These also have approximate F-distributions under the null hypothesis.
- With large samples, these four test statistics are nearly equal and will give essentially equivalent conclusions.
- With two groups, these four test statistics are equal and are equivalent to the F-value in Hotelling's T^2 test.
- For moderate sample sizes, the Wilks, Lawley-Hotelling, and Pillai methods have similar power.
- Roy's method appears to be best only when there is a difference in just one component of the mean vector, and only one of the groups is different from the others.
- Pillai's trace may be more robust to non-normality, but transforming the data is suggested if the model residuals show departures from normality.

Multiple Comparisons in MANOVA

- If the F-test rejects the null that all m groups have equal mean vectors, we can use multiple comparisons to determine which specific pairs of groups have differing mean vectors (and which components of those mean vectors differ).

- This is most simply done via Bonferroni confidence intervals for all the differences $\alpha_{ik} - \alpha_{i'k}$ for all $i \neq i'$, where $i, i' \in \{1, \ldots, m\}$.

- With family confidence level $1 - \alpha$, the set of CIs defined by:

$$(\bar{x}_{ik} - \bar{x}_{i'k}) \pm t_{n-m, \frac{\alpha}{qm(m-1)}} \sqrt{\frac{E_{kk}}{n-m}\left(\frac{1}{n_i} + \frac{1}{n_{i'}}\right)}$$

  contain the $\alpha_{ik} - \alpha_{i'k}$.

- Any such interval that does not contain zero would indicate a significant difference in that component of the mean vector between that pair of groups.

Two-Way MANOVA

- We could have q response variables measured at various levels (categories) of two factors.
- Note that here, the factors are simply grouping variables, not the type of factors that we encountered in factor analysis!
- The one-way MANOVA model can be extended to a two-way model such as:

$$\mathbf{X}_{hij} = \boldsymbol{\mu} + \boldsymbol{\alpha}_h + \boldsymbol{\beta}_i + \boldsymbol{\gamma}_{hi} + \boldsymbol{\epsilon}_{hij},$$

  where these are each $q \times 1$ vectors.
- Here, h indicates the level for the first factor, i indicates the level for the second factor, and j indicates the observation within the factor level combination.
- The vector $\boldsymbol{\gamma}_{hi}$ contains the effects for interaction between the two factors.

Tests in the Two-Way MANOVA

- When the random error vector is multivariate normal, tests about the interaction effects and factor effects can be carried out via F-tests.
- If the interaction effects are significant, then the effect of one factor is dependent on the level of the other factor.
- In this case, the F-tests for the factor effects are not applicable.
- Separate interaction plots for each variable can be done to investigate the nature of the interaction.
- If the interaction effects are not significant, then we can test whether the mean vectors differ across the levels of factor 1, and whether the mean vectors differ across the levels of factor 2.