Multivariate hypothesis testing

Contents

1. **Multivariate two-sample test**
2. Multivariate analysis of variance (MANOVA)
3. PERMANOVA (by Eduard Szöcs)

Learning targets

- Explaining and applying the Hotelling‘s T^2-Test.
- Explaining and applying MANOVA and PERMANOVA.

Learning targets and study questions

- Explaining and applying the Hotelling‘s T^2-Test.
  - Explain the similarities between the univariate and multivariate two sample test.
  - What are the assumptions of the test (and of MANOVA)?
- Explaining and applying MANOVA and PERMANOVA.
  - List the four MANOVA test statistics and outline their conditions of use
  - How does PERMANOVA differ from MANOVA? When should it be used?
  - What are the assumptions of PERMANOVA?

Multivariate comparison of two groups

Comparison of the central tendency of two groups concerning one response variable

Cu content of soil

	Cu mg/kg			
Soil a	4.2	3.2	4.2	5.5
Soil b	8.9	8.4	7.8	9.5

H_0: $\mu_a = \mu_b$

t-test

Comparison of the central tendency of two groups concerning *k* responses

Cu and Fe content of soil

	Fe mg/kg			Cu mg/kg		
Soil a	94.0	88.3	78.9	4.2	3.2	4.2
Soil b	55.9	64.3	44.9	8.9	8.4	7.8

H_0: $\boldsymbol{\mu_a} = \boldsymbol{\mu_b}$

Hotelling's T^2-Test

Multivariate comparison of two groups

Formula for *t*-test (equal variance)

$$t = \frac{\overline{x_1} - \overline{x_2}}{s_{x_1 x_2}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad \text{with} \quad s_{x_1 x_2} = \sqrt{\frac{(n_1 - 1)s_{x_1}^2 + (n_2 - 1)s_{x_2}^2}{n_1 + n_2 - 2}}$$

„Difference of means divided by standard error of difference“

Formula for Hotelling‘s T^2-Test

Difference of mean vectors

$$T^2 = \frac{n_1 n_2}{n_1 + n_2} \boxed{(\overline{\boldsymbol{x}_1} - \overline{\boldsymbol{x}_2})^t} \boldsymbol{S}^{-1} \boxed{(\overline{\boldsymbol{x}_1} - \overline{\boldsymbol{x}_2})}$$

Inverse of Covariance matrix

Assumptions of the Hotelling's T^2 test and of MANOVA

- Independence of observations
- Multivariate normal distribution of dependent variables within each group
  → Visual checking
- Homogeneity of covariance matrices (i.e. homoscedasticity) within each group
  → hypothesis test: multivariate generalization of Levene’s test of homogeneity of variances
  → Visual checking (informs on collinearity)

Multivariate hypothesis testing

Contents

Multivariate comparisons: MANOVA

Comparison of the central tendency of $n \geq 2$ groups with respect to one variable

Cu content of soil

	Cu mg/kg			
Soil a	4.2	3.2	4.2	5.5
Soil b	..	..	..	..
Soil c				
Soil d	..	..	..	..
Soil e	8.9	8.4	7.8	9.5

H_0: $\mu_a = \mu_b = \ldots = \mu_e$

`aov(), anova() ...`

ANOVA

Comparison of the central tendency of $n \geq 2$ groups with respect to $k \geq 2$ variables

Cu and Fe content of soil

	Fe mg/kg				Cu mg/kg			
Soil a	94.0	88.3	78.9	85.9	4.2	3.2	4.2	5.5
Soil b	..	..	..	..	..	..	..	..
Soil c	..	..	..	..	..	..	..	..
Soil d	..	..	..	..	..	..	..	..
Soil e	55.9	64.3	44.9	63.4	8.9	8.4	7.8	9.5

H_0: $\boldsymbol{\mu}_{\boldsymbol{a}} = \boldsymbol{\mu}_{\boldsymbol{b}} = \ldots = \boldsymbol{\mu}_{\boldsymbol{e}}$

`manova(), rda() ...`

MANOVA

Multivariate comparisons: MANOVA

- **Example:** How do nutrients, light and a herbivore affect benthic stream algae?

Table 1 Summary of (M)ANOVAs (*P*-values) comparing diatom community variables among nutrient, grazer and light treatments. MANOVA *P*-values are for the Pillai's Trace statistic. Data were arcsine square-root transformed prior to analysis. *Post hoc* test rankings are shown for significant effects of nutrients and grazing, and significant light effects are summarised in a similar way. Treatments were abbreviated as follows. Nutrients: ambient (A), medium (M), high (H) and very high (V); grazers: no grazers (N), low (L), medium (M) and high (H); light: ambient (A) and reduced (R)

Dependent variable	Nutrients	Ranking	Grazer	Ranking	Light	Ranking	Nutrients × Grazer	Nutrients × Light	Grazer × Light	Nutrients × Grazer × Light
Community composition (MANOVA; 13 taxa)	**<0.0001**		**0.01**		**<0.0001**		0.05	**<0.0001**	0.26	0.65
Navicula cryptotenella	**0.004**	(V=H=M)>A	0.33		0.53		0.32	0.17	0.94	0.62
Diatoma hiemale	**0.04**	H>M	0.10		0.32		0.63	0.79	0.27	**0.04**
Nitzschia palea	**<0.0001**	(V=H)>M>A	0.33		**<0.0001**	A>R	0.69	0.06	0.23	0.74
Nitzschia dissipata	**<0.0001**	(V=H)>M>A	0.51		**<0.0001**	A>R	0.49	0.09	0.52	0.89
Fragilaria vaucheriae	**0.01**	(V=H)>M	0.34		**<0.0001**	A>R	0.70	0.96	**0.04**	0.71
Encyonema minuta	**0.001**	H>(V=M=A)	0.16		**<0.0001**	A>R	0.93	0.47	0.13	0.56
Cymbella kappii	0.29		0.87		**<0.0001**	A>R	0.21	0.40	0.11	0.90
Synedra spp.	0.17		0.75		**<0.0001**	A>R	0.63	0.34	0.76	0.16
Gomphonema minutum	**0.01**	(A=M)>H	**<0.0001**	H>(M=L=N)	**0.004**	R>A	0.26	0.27	0.80	0.98
Achnanthidium minutissimum	**<0.0001**	A>M>H; A>V	0.71		**<0.0001**	R>A	0.45	0.93	0.18	0.37
Rossithidium petersenii	**<0.0001**	A>M>(H=V)	0.97		**<0.0001**	R>A	0.91	**0.01**	0.79	0.88
Cocconeis placentula	**0.001**	M>(A=H)	0.30		**<0.0001**	R>A	0.83	**0.003**	**0.04**	0.33
Planothidium lanceolatum	**<0.0001**	V>H>M>A	0.16		**<0.0001**	R>A	0.16	**<0.0001**	0.06	0.98

P values < 0.05 are printed in bold.

Lange et al. 2011

MANOVA test statistics

- MANOVA:
  - Wilks lambda
    most popular, but lower power in most cases
  - Pillais trace
    relatively robust to deviations from test assumptions for balanced designs; standard in R; highest power except for collinear case
  - Roys largest root
    most powerful in case of high collinearity between the variables (all groups means differ in one direction) if assumptions met; strongest influenced by heterogeneity of covariance matrices (α-error)
  - Hotelling-Lawley trace
    can be converted to Hotelling T^2, most efficient for two groups

→ Converge for large sample sizes

Assumptions of MANOVA

How serious is violation of assumptions?

- MANOVA relatively robust to deviations from assumptions
- Heterogeneity of covariance matrices especially problematic for unequal group sample sizes
- For ecological data assumptions usually not met

What to do if assumptions are not met?

- Non-parametric and robust MANOVA
- RDA can be used for MANOVA based on permutations, more robust to deviation from multivariate normality
- Multivariate GLMs for non-normally distributed data (e.g. poisson or binomial data)

Multivariate hypothesis testing

Contents

Learning targets

- Explaining and applying the Hotelling‘s T^2-Test.
- Explaining and applying MANOVA and PERMANOVA.

Learning targets and study questions

- Explaining and applying the Hotelling‘s T^2-Test.
  - Explain the similarities between the univariate and multivariate two sample test.
  - What are the assumptions of the test (and of MANOVA)?
- Explaining and applying MANOVA and PERMANOVA.
  - List the four MANOVA test statistics and outline their conditions of use
  - How does PERMANOVA differ from MANOVA? When should it be used?
  - What are the assumptions of PERMANOVA?

Multivariate comparison of two groups

Comparison of the central tendency of two groups concerning one response variable

Comparison of the central tendency of two groups concerning *k* responses

Cu content of soil

	Cu mg/kg			
Soil a	4.2	3.2	4.2	5.5
Soil b	8.9	8.4	7.8	9.5

Cu and Fe content of soil

	Fe mg/kg			Cu mg/kg		
Soil a	94.0	88.3	78.9	4.2	3.2	4.2
Soil b	55.9	64.3	44.9	8.9	8.4	7.8

H_0: $\mu_a = \mu_b$

t-test

H_0: $\boldsymbol{\mu_a} = \boldsymbol{\mu_b}$

Hotelling's T^2-Test

4

The univariate and multivariate comparison of the central tendency follow a similar concept. Note that the results for a series of *t*-tests are not necessarily identical to the result of the Hotelling T^2-Test (and that for a series of *t*-tests the alpha error would increase to $1-0.95^k$ where *k* is the number of *t*-tests).

μ indicates the true mean of a vector *x*.

Multivariate comparison of two groups

Formula for *t*-test (equal variance)

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_{x_1 x_2}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \quad \text{with} \quad s_{x_1 x_2} = \sqrt{\frac{(n_1-1)s_{x_1}^2 + (n_2-1)s_{x_2}^2}{n_1+n_2-2}}$$

„Difference of means divided by standard error of difference"

Formula for Hotelling's T^2-Test

Difference of mean vectors

$$T^2 = \frac{n_1 n_2}{n_1+n_2}\boxed{(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)^t}\,\boldsymbol{S}^{-1}\boxed{(\overline{\boldsymbol{x}}_1 - \overline{\boldsymbol{x}}_2)}$$

Inverse of Covariance matrix

5

Squaring of the *t*-statistic of the univariate *t*-test yields a test statistic that is similar to T^2.

Assumptions of the Hotelling's T^2 test and of MANOVA

- Independence of observations
- Multivariate normal distribution of dependent variables within each group
  → Visual checking
- Homogeneity of covariance matrices (i.e. homoscedasticity) within each group
  → hypothesis test: multivariate generalization of Levene's test of homogeneity of variances
  → Visual checking (informs on collinearity)

The assumptions for the Hotelling test are similar to the assumptions for the univariate *t*-test (e.g. normal distribution, homogeneity of variance).

The visual checking of homogeneity of variances is rather cumbersome, especially in the case of multiple groups in MANOVA (next slides). However, it provides information on potential collinearity between variables. Anderson (2006) introduced a hypothesis test that is relatively robust to deviations from multivariate normality and part of the demonstration in R.

Anderson MJ. 2006. Distance-based tests for homogeneity of multivariate dispersions. Biometrics. 62:245–253.

Multivariate hypothesis testing

Contents

1. Multivariate two-sample test
2. **Multivariate analysis of variance (MANOVA)**
3. PERMANOVA (by Eduard Szöcs)

7

Multivariate comparisons: MANOVA

Comparison of the central tendency of $n \geq 2$ groups with respect to one variable

Comparison of the central tendency of $n \geq 2$ groups with respect to $k \geq 2$ variables

Cu content of soil

	Cu mg/kg			
Soil a	4.2	3.2	4.2	5.5
Soil b	..	..	..	..
Soil c				
Soil d	..	..	..	..
Soil e	8.9	8.4	7.8	9.5

H_0: $\mu_a = \mu_b = ... = \mu_e$

`aov(), anova() ...`

ANOVA

Cu and Fe content of soil

	Fe mg/kg				Cu mg/kg			
Soil a	94.0	88.3	78.9	85.9	4.2	3.2	4.2	5.5
Soil b	..	..	..	..	..	..	..	..
Soil c	..	..	..	..	..	..	..	..
Soil d	..	..	..	..	..	..	..	..
Soil e	55.9	64.3	44.9	63.4	8.9	8.4	7.8	9.5

H_0: $\boldsymbol{\mu_a} = \boldsymbol{\mu_b} = \boldsymbol{...} = \boldsymbol{\mu_e}$

`manova(), rda() ...`

MANOVA

8

The computation of ANOVA and MANOVA is very similar (see Zar 2013: Chapter 16). Note that the results of a MANOVA and of several univariate ANOVAs can be different. For example, the MANOVA could indicate significant differences between the group means, whereas the ANOVA would detect none, and conversely. The reason is that variables can be correlated and this influences the MANOVA. In the case of any correlations, the results of the MANOVA equal those obtained from multiple ANOVAs (Tabachnick & Fidell 2007. Using multivariate statistics: 268). But note the alpha error inflation, when conducting multiple ANOVAs! In case of correlations, the relationship between power and correlation between dependent variables is complex. Cole et al. (1994) state: “[...] power increases as correlations between dependent variables with large consistent effect sizes (that are in the same direction) move from near 1.0 toward -1.0 , (b) power increases as correlations become more positive or more negative between dependent variables that have very different effect sizes (i.e., one large and one negligible), and (c) power increases as correlations between dependent variables with negligible effect sizes shift from positive to negative.”

The two-group MANOVA is equivalent to the Hotelling T^2 Test, the results are the same in terms of the F-test.

Cole DA, Maxwell SE, Arvey R, Salas E. 1994. How the power of MANOVA can both increase and decrease as a function of the intercorrelations among the dependent variables. Psychological Bulletin. 115:465.

Multivariate comparisons: MANOVA

- **Example:** How do nutrients, light and a herbivore affect benthic stream algae?

Table 1 Summary of (M)ANOVAS (*P*-values) comparing diatom community variables among nutrient, grazer and light treatments. MANOVA *P*-values are for the Pillai's Trace statistic. Data were arcsine square-root transformed prior to analysis. *Post hoc* test rankings are shown for significant effects of nutrients and grazing, and significant light effects are summarised in a similar way. Treatments were abbreviated as follows. Nutrients: ambient (A), medium (M), high (H) and very high (V); grazers: no grazers (N), low (L), medium (M) and high (H); light: ambient (A) and reduced (R)

Dependent variable	Nutrients	Ranking	Grazer	Ranking	Light	Ranking	Nutrients × Grazer	Nutrients × Light	Grazer × Light	Nutrients × Grazer × Light
Community composition (MANOVA; 13 taxa)	**<0.0001**		**0.01**		**<0.0001**		0.05	**<0.0001**	0.26	0.65
Navicula cryptotenella	**0.004**	(V=H=M)>A	0.33		0.53		0.32	0.17	0.94	0.62
Diatoma hiemale	**0.04**	H>M	0.10		0.32		0.63	0.79	0.27	**0.04**
Nitzschia palea	**<0.0001**	(V=H)>M>A	0.33		**<0.0001**	A>R	0.69	0.06	0.23	0.74
Nitzschia dissipata	**<0.0001**	(V=H)>M>A	0.51		**<0.0001**	A>R	0.49	0.09	0.52	0.89
Fragilaria vaucheriae	**0.01**	(V=H)>M	0.34		**<0.0001**	A>R	0.70	0.96	**0.04**	0.71
Encyonema minuta	**0.001**	H>(V=M=A)	0.16		**<0.0001**	A>R	0.93	0.47	0.13	0.56
Cymbella kappii	0.29		0.87		**<0.0001**	A>R	0.21	0.40	0.11	0.90
Synedra spp.	0.17		0.75		**<0.0001**	A>R	0.63	0.34	0.76	0.16
Gomphonema minutum	**0.01**	(A=M)>H	**<0.0001**	H>(M=L=N)	**0.004**	R>A	0.26	0.27	0.80	0.98
Achnanthidium minutissimum	**<0.0001**	A>M>H; A>V	0.71		**<0.0001**	R>A	0.45	0.93	0.18	0.37
Rossithidium petersenii	**<0.0001**	A>M>(H=V)	0.97		**<0.0001**	R>A	0.91	**0.01**	0.79	0.88
Cocconeis placentula	**0.001**	M>(A=H)	0.30		**<0.0001**	R>A	0.83	**0.003**	**0.04**	0.33
Planothidium lanceolatum	**<0.0001**	V>H>M>A	0.16		**<0.0001**	R>A	0.16	**<0.0001**	0.06	0.98

P values < 0.05 are printed in bold.

Lange et al. 2011

Lange K., Liess A., Piggott J.J., Townsend C.R. & Matthaei C.D. (2011) Light, nutrients and grazing interact to determine stream diatom community composition and functional group structure. Freshwater Biology 56, 264–278.

MANOVA test statistics

- MANOVA:
  - Wilks lambda
    most popular, but lower power in most cases
  - Pillais trace
    relatively robust to deviations from test assumptions for balanced designs; standard in R; highest power except for collinear case
  - Roys largest root
    most powerful in case of high collinearity between the variables (all groups means differ in one direction) if assumptions met; strongest influenced by heterogeneity of covariance matrices (α-error)
  - Hotelling-Lawley trace
    can be converted to Hotelling T^2, most efficient for two groups

Converge for large sample sizes

10

In practice, most statistical software computes all four test statistics. In those cases where they differ regarding the acceptance or rejection of H_0, one should carefully examine the data regarding deviations from assumptions and interpret the test statistics in this light (see Rencher 2012: 189f for details). In addition, visualisation of group means and observations with respect to up to 3 variables should be considered (perhaps in conjunction with methods such as ordination that may help to reduce the number of variables).

All four test statistics rely on eigenvalues, and approximations of the F-statistic of the univariate ANOVA exist (see Rencher 2012: 180ff).

Assumptions of MANOVA

How serious is violation of assumptions?

- MANOVA relatively robust to deviations from assumptions
- Heterogeneity of covariance matrices especially problematic for unequal group sample sizes
- For ecological data assumptions usually not met

What to do if assumptions are not met?

- Non-parametric and robust MANOVA
- RDA can be used for MANOVA based on permutations, more robust to deviation from multivariate normality
- Multivariate GLMs for non-normally distributed data (e.g. poisson or binomial data)

MANOVA is generally robust to deviations from test assumptions (especially when employing Pillais trace as test statistic, see previous slide). An examination of the effects of deviation from the assumptions concluded that in most situations the parametric MANOVA is the most powerful test (i.e. minimizes the beta error: acceptance of H_0 although it is wrong) and yields valid results (Finch 2005). If the MANOVA follows a balanced design i.e. approximately equal sample sizes of the groups (or largest $n \leq 1.5$ * smallest n) then heterogeneity of the covariance matrices has only minor influence on Pillais trace (Finch 2005, Zar 2013 Ch.6). If the sample sizes are unequal, then heterogeneity of the covariance matrices may affect the α error as follows: "If the larger variances and covariances are associated with the larger samples, the true α-level is reduced and the tests become conservative. On the other hand, if the larger variances and covariances come from the smaller samples, α is inflated, and the tests become liberal "(Rencher 2002: Methods of multivariate Analysis: 177). In both cases the nominal α is different from the real α of the test.

For more information on multivariate GLMs see:

Warton DI, Wright ST, Wang Y. 2011. Distance-based multivariate analyses confound location and dispersion effects. Methods in Ecology and Evolution. 3:89–101.

The use of RDA for (M)ANOVA is detailed in Borcard et al. (2011): 198f

Finch 2005: Comparison of the Performance of Nonparametric and Parametric MANOVA Test Statistics when assumptions are violated. Methodology 1: 27-38