

## Index of Lecture 6

Page	Title
1	Practical information
2	1-way ANOVA – Model
3	ANOVA versus regression
4	1-way ANOVA – Analysis
5	How to proceed after the ANOVA?
6	More about contrasts
7	Multiple comparisons: Overview
8	Bonferroni method
9	Holm method
10	Example: Multiple comparisons
11	Scheffè's method
12	Contrasts for 1-way ANOVA with quantitative groups
13	Beyond 1-way ANOVA
14	Stata do-file (part)
15	Software notes

## PRACTICAL INFORMATION

Today's lecture: 1-way ANOVA review + new topics:  
*contrasts + multiple comparisons + ANOVA vs. regression.*

Guidelines for textbook reading:

- Chapter 3 on 1-Way ANOVA: mostly well-known,<sup>1</sup>
- Chapter 4 on contrasts: short chapter, all relevant,
- Chapter 5 on multiple comparisons: much more detailed than our ambition level, don't focus on mathematical details and read cursorily from Section 5.4.2 onwards,

News/Schedule:

- Schedule for rest of semester (except holidays, 18+21/3):
  - \* lectures: Thursdays 9-11am (278N),
  - \* labs: Mondays 1-4pm (small computer lab),
    - note: lab session tomorrow 1-4pm!
- Moodle system used essentially for
  - \* communication (News and Discussion Forum),
  - \* submission of home assignments,
- Course project:
  - \* time to start thinking about data for your project,
  - \* project outline due March 17.

---

<sup>1</sup> Skip Sections 3.9 and 3.11, and discussion of  $P(p)$  on p. 49.

## 1-WAY ANOVA – MODEL

Rat data example (GO Exercise 3.1, p. 60):

- rat liver weights in percent of body weight following 4 diets (labelled 1-4) randomly allocated to rats,
- notation:  
 $y_{ij}$  = rat liver weight for  $j$ th rat in diet group  $i$ ,  
 $i = 1, \dots, g$  ( $g=4$ ),  
 $j = 1, \dots, n_i$  ( $n_1=7, n_2=n_4=8, n_3=6$ ),
- purpose: assess impact of diets on liver weight.

Statistical model:

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, \dots, g; \quad j = 1, \dots, n_i,$$

where the  $\varepsilon_{ij}$  are i.i.d. and  $\sim N(0, \sigma^2)$ .

Model parameters:

- group (population) means  $\mu_1, \dots, \mu_4$ ,
- common group (population) standard deviation  $\sigma$ .

Alternative model formulations<sup>2</sup>:

$$y_i = \mu_{\text{diet}(i)} + \varepsilon_i, \quad i = 1, \dots, 29 \text{ } (\sim \text{obs. no.}),$$

$$\begin{aligned} y_i &= \beta_0 + \beta_1 1_{\text{diet}2(i)} + \beta_2 1_{\text{diet}3(i)} + \beta_4 1_{\text{diet}4(i)} + \varepsilon_i, \quad \text{or} \\ &= \mu + \alpha_{\text{diet}(i)} + \varepsilon_i, \quad \text{where } \alpha_1 = 0, \end{aligned}$$

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \quad \text{with restrictions on } \alpha_i \text{'s.}$$

---

<sup>2</sup> Restrictions on  $(\alpha_i)$ : either  $\alpha_1 = 0$  (Stata; Minitab **General Regression**; R),  $\alpha_4 = 0$  (SAS), or  $\alpha_1 + \dots + \alpha_4 = 0$  (Minitab **General Linear Model** and **General Regression**).

## ANOVA VERSUS REGRESSION

= two different frameworks for analyzing the *same model* and presenting the results.<sup>3</sup>

Advantages of ANOVA framework:

- no reliance on an, often artificial, reference category,<sup>4</sup>
- extra tools for exploring multiple samples and/or multiple factors, in particular for balanced data.

Advantages of regression framework:

- easier to include continuous predictors (the equivalent of “analysis of covariance” (ANCOVA), which no longer plays any prominent role in ANOVA methods),
- full range of model checking and diagnostic tools (although some of these are of questionable value for categorical predictors; e.g. VIF and leverage).

Minitab vs Stata:

- more complete regression and ANOVA facilities in Stata,<sup>5</sup>
- more easily accessible facilities in Minitab.

---

<sup>3</sup> From Oehlert (p. 44): “Strictly speaking, ANOVA is an arithmetic procedure for partitioning the variability in a data set [...], however [...] we sometimes speak of testing via ANOVA although the test is not really part of the ANOVA.” Other authors (e.g., Christensen 1996, p. 132) use ANOVA “as a name for the entire package of techniques used to compare more than two samples”.

<sup>4</sup> A common mistake within the regression framework is to explore only comparisons with the reference category, cf. L4b–10.

<sup>5</sup> ANOVA facilities in Stata are substantially improved from version 11.

1-WAY ANOVA – ANALYSIS
------------------------

Estimation: ( $g$  groups,  $N$  observations)

- $\hat{\mu}_i = \bar{y}_i$ . ( $\text{Var}(\hat{\mu}_i) = \sigma^2/n_i$ ,  $\text{SE}(\hat{\mu}_i) = s/\sqrt{n_i}$ ),
- $\hat{\sigma}^2 = s^2 = \sum_i \frac{n_i - 1}{N - g} s_i^2 = \frac{\sum_{ij} (y_{ij} - \bar{y}_i)^2}{N - g} = \frac{\text{SS}_E}{\text{DF}_E} = \text{MS}_E$ ,  
 – weighted average of the group sample variances  $s_i^2$ ,
- confidence intervals and tests: “4-step procedure” (L1a–5).

ANOVA table: ( $g$  groups,  $N$  observations)

Source of variation	Degrees of freedom	Sum of squares	Mean square	$F$
Groups/ Treatm.	$\text{DF}_{\text{Trt}} = g - 1$	$\text{SS}_{\text{Trt}} = \sum_i n_i (\bar{y}_i - \bar{y}_{..})^2$	$\text{MS}_{\text{Trt}} = \text{SS}_{\text{Trt}} / \text{DF}_{\text{Trt}}$	$\frac{\text{MS}_{\text{Trt}}}{\text{MS}_E}$
Error	$\text{DF}_E = N - g$	$\text{SS}_E = \sum_{ij} (y_{ij} - \bar{y}_i)^2$	$\text{MS}_E = \text{SS}_E / \text{DF}_E$	
Total	$\text{DF}_T = N - 1$	$\text{SS}_T = \sum_{ij} (y_{ij} - \bar{y}_{..})^2$	$\text{MS}_T = \text{SS}_T / \text{DF}_T$	

- $F$ -test in table is for hypothesis  $H_0$ :  $\mu_1 = \dots = \mu_g$  (all groups equal, homogeneity between groups) against alternative hypothesis  $H_a$ : some  $\mu$ 's differ,
- $P$ -value (for  $F$ -test) =  $\Pr(F \geq F_{\text{obs}})$ ;  $F \sim F(\text{DF}_{\text{Trt}}, \text{DF}_E)$ ,
- $E(\text{MS}_E) = \sigma^2$  and  $E(\text{MS}_{\text{Trt}}) = \sigma^2 + \sum_i n_i (\mu_i - \bar{\mu})^2 / (g - 1)$ ,
- (technical) the ANOVA decomposition is based on the equation

$$(y_{ij} - \bar{y}_{..}) = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{y}_{..}).$$

## HOW TO PROCEED AFTER THE ANOVA?

Q.:  $\left\{ \begin{array}{l} \text{have rejected } H_0 : \mu_1 = \dots = \mu_g, \\ \text{but what relations between } \mu_i \text{'s (which differ)?} \end{array} \right.$

- estimation of parameters:  $\hat{\mu}_1 = \bar{y}_{1.}, \dots, \hat{\mu}_g = \bar{y}_{g.}$ ,  
also of derived parameters such as contrasts,

$$w(\{\mu_i\}) = \sum_{i=1}^g w_i \mu_i = \sum_{i=1}^g w_i \alpha_i, \quad \text{where } \sum_{i=1}^g w_i = 0,$$

examples (for  $g=3$ , or 3 groups):

- \*  $w(\{\mu_i\}) = \mu_1 - \mu_2$  (i.e.,  $w_1=1, w_2=-1, w_3=0$ ),
- \*  $w(\{\mu_i\}) = \frac{1}{2}(\mu_2 + \mu_3) - \mu_1$  (i.e.,  $w = (-1, \frac{1}{2}, \frac{1}{2})$ ),

- confidence intervals or tests for interesting parameters,  
most common examples:

$$\mu_i : \bar{y}_{i.} \pm t(.025, DF_E) \sqrt{MS_E} \sqrt{(1/n_i)},$$

$$\mu_i - \mu_{i'} : \bar{y}_{i.} - \bar{y}_{i'}. \pm t(.025, DF_E) \sqrt{MS_E} \sqrt{(1/n_i) + (1/n_{i'})},$$

- diagram, e.g.  $\hat{\mu}_i$ 's with error bars (SE or conf. interval),
- problems with choice of contrasts/pairwise comparisons:
  - \* many hypotheses; if each test has error of 5%, then  
total error is  $\gg 5\%$ ,
  - \* above methods apply only to preplanned hypotheses,  
not to hypotheses suggested by the data.

## MORE ABOUT CONTRASTS

Formulae to estimate contrasts and SE's, for  $w = w(\{\mu_i\})$ :

$$\hat{w} = \sum_i w_i \bar{y}_i. \quad \text{and} \quad \text{SE}(\hat{w}) = \sqrt{\text{MS}_E} \sqrt{\sum_i w_i^2 / n_i}.$$

Explanation of variation by contrasts:

- idea: every contrast accounts for part of the variation explained by the grouping/model ( $\text{SS}_{\text{Trt}}$ ),
- formula:  $\text{SS}(w) = \hat{w}^2 / (\sum_i w_i^2 / n_i) = \text{MS}_E \times t_w^2$ , where  $t_w$  is the  $t$ -statistic for testing  $w = 0$ ,
- orthogonal contrasts:
  - \* idea: contrasts that explain *different* parts of the variation, to allow *independent* interpretation,
  - \* definition:  $w = \sum_i w_i \mu_i$  and  $w^* = \sum_i w_i^* \mu_i$  are orthogonal if:  $\sum_i w_i w_i^* / n_i = 0$ ,
  - \* there exist at most  $(g - 1)$  pairwise orthogonal contrasts among  $g$  groups,
  - \* example: (3 groups, equal  $n_i$ 's)
 
$$w = \mu_1 - \frac{1}{2}(\mu_2 + \mu_3), \quad \text{and} \quad w^* = \mu_2 - \mu_3,$$
  - \* main result: for *orthogonal* contrasts  $w^{(1)}, \dots, w^{(g-1)}$ , it holds that
 
$$\text{SS}_{\text{Trt}} = \text{SS}(w^{(1)}) + \text{SS}(w^{(2)}) + \dots + \text{SS}(w^{(g-1)}),$$
 — splitting (decomposing)  $\text{SS}_{\text{Trt}}$  into contrast parts,
  - \* problem: not always easy to find useful orthogonal contrasts.

## MULTIPLE COMPARISONS: OVERVIEW

Some terminology and basic facts:

- type I (error) probability: prob. of rejecting  $H_0$ , if  $H_0$  is true,
- per comparison or individual error rate: type I prob. for each test,
- simultaneous or experimentwise or familywise error rate: type I probability for *all* tests, i.e., for rejection of any test in a set (“family”) of tests carried out,
- strong familywise error rate: prob. of rejecting any true null hypotheses (but no impact of false null hypotheses/true rejections),
- all multiple comparisons procedures reduce the type I prob. and increase the type II prob. – a trade-off,
- conservative procedure: too high  $P$ -value(s)  $\sim$  too few hypotheses rejected; opposite of liberal procedure  $\sim$  false significance(s),
- classical strategy (now considered problematic): if overall  $F$ -test is non-sign. and no preplanned hypotheses: *done*, no further “data snooping”! <sup>6</sup>

Simple methods (in this course):

- Bonferroni & Holm corrections for preplanned or all comparisons,
- Scheffé’s method for contrasts suggested by the data.

Other methods exist (in abundance):

- many require balanced data (Tukey, Duncan),
- some are for special cases (Dunnett for comparison with control),
- some assume independent tests (Benjamini & Hochberg’s false discovery rate (i.e., proportion of false rejections) method).

---

<sup>6</sup> Referred to as “protected LSD” method, when combined with LSD comparisons.

## BONFERRONI METHOD

Idea: If  $A$  and  $B$  are events, it always holds that

$$\Pr(A \text{ or } B) \leq \Pr(A) + \Pr(B).$$

In particular, in the context of performing several tests,

$$\Pr(\text{error in one or more tests}) \leq \text{sum of error prob.}$$

Therefore, if we make  $K$  tests/comparisons, we can achieve the *simultaneous* type I probability *for all tests* to be  $\leq \epsilon$ , by taking the type I probability for each test equal to  $\epsilon/K$ .

Adjustment of pairwise  $t$ -tests for  $K$  preplanned tests:

- use  $t$ -distribution percentiles  $\epsilon/(2K)$ , or
- multiply uncorrected  $P$ -values by  $K$ .

Adjustment of pairwise  $t$ -tests for unplanned comparisons:

(suggested by the data, e.g. involving “best” treatment)

- take  $K = \text{total no. of comparisons} = g(g-1)/2$ ,
- use above procedure with that value of  $K$ .

Notes for Bonferroni method:

- gives also simultaneous confidence intervals,<sup>7</sup>
- is *conservative* for controlling strong familywise error rate,
- is available for ANOVA in Minitab only via  
Stat-ANOVA-General Linear Model-Comparisons
- is flexible: applies to a wide range of settings/models.

<sup>7</sup> The prob. of all CIs to simultaneously cover their true value is  $\geq 1 - \epsilon$ .

## HOLM METHOD

Steps of the *sequential* (also called step-down) procedure:

- 1) sort all the  $K$  unadjusted  $P$ -values as:  $P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(K)}$ ,
- 2) for the test corresponding to the  $i$ th ordered  $P$ -value, compute the adjusted  $P$ -value  $P_{(i)}^H = P_{(i)} \times (K - i + 1)$ ,  $i = 1, \dots, K$ ,
- 3) rules for significance:
  - (i) if  $P_{(i)}^H > \epsilon \Rightarrow$  non-significant (at  $\epsilon$ ),
  - (ii) if  $P_{(i)}^H \leq \epsilon$  and also all  $P_{(j)}^H \leq \epsilon$  for all  $j = 1, \dots, i$ ,  $\Rightarrow$  significant (at  $\epsilon$ ).

Notes for Holm method:

- controls the strong familywise error rate, and is less conservative for this than the Bonferroni method,
- does not provide simultaneous confidence intervals,
- is not available in Minitab, but can be carried out manually (by the recipe above),
- adjusted  $P$ -values ( $P_{(i)}^H$  above) are available in Stata, but the sequential rule (ii) must be checked manually,
- is also flexible: applies to a wide range of settings/models.

## MULTIPLE COMPARISONS: EXAMPLE

Rat data:

- assume no preplanned hypotheses or treatment (diet) structure of interest,
- a total of  $K = 4 \cdot (4 - 1)/2 = 6$  multiple comparisons.

<i>P</i> -value		Multiple comparison method		
pair	order	unadjusted	Bonferroni	Holm
2 vs 4	1	.0025	.015	.015
3 vs 4	2	.0068	.041	.034
1 vs 4	3	.106	.633	.422
1 vs 2	4	.128	.768	.384
1 vs 3	5	.205	1	.409
2 vs 3	6	.869	1	.869

\* same conclusions:  
only 4 vs 2,3 signif.

\* Holm  $P <$  Bonf.  $P$

Significance letter coding (groups with same letter *not* sign. different; available in Minitab **General Linear Model**):

- order group means from highest to lowest,
- designate letter *a* to highest group + all groups not significantly different from it,
- designate letter *b* to next group in the same way (but drop if same pattern as for *a*),
- continue through all groups,
- Rat data coding:  $4^a 1^{ab} 3^b 2^b$ .

## SCHEFFÉ'S METHOD

- corrects for examining non-preplanned contrasts,<sup>8</sup>
- “allows” to test contrasts suggested by the data.

Idea: use same procedure as with preplanned contrasts, but correct the *reference distribution*,

not  $\frac{\hat{w} - w}{\text{SE}(\hat{w})} \sim t(\text{DF}_E)$ , but  $[\frac{\hat{w} - w}{\text{SE}(\hat{w})}]^2 / (g - 1) \sim F(g - 1, \text{DF}_E)$ ,

for example,

- test of  $H_0: w = 0$  by  $F = [\frac{\hat{w}}{\text{SE}(\hat{w})}]^2 / (g - 1) \sim F(g - 1, \text{DF}_E)$ ,
- 95% CI for  $w$ :  $\hat{w} \pm \sqrt{(g - 1)F(.05, g - 1, \text{DF}_E)} \text{SE}(\hat{w})$ .

Properties:

- mathematically,

$$[\frac{\hat{w} - w}{\text{SE}(\hat{w})}]^2 / (g - 1) \leq \frac{\text{MS}_{\text{Trt}}}{\text{MS}_E} \sim F(g - 1, \text{DF}_E),$$

- method can never give stronger result than the overall ANOVA  $F$ -test for  $H_0: \mu_1 = \dots = \mu_g$ ,
- there always exists a contrast to give exactly same result as overall  $F$ -test (but it is usually not interesting),
- method is conservative.

---

<sup>8</sup> Method should *not* be used for pairwise comparisons, because too conservative.

## CONTRASTS FOR 1-WAY ANOVA WITH QUANTITATIVE GROUPS

Resin data example: log failure time  $y_i$  of unit  $i$  subjected to temperature  $x_i$ , where  $x_i \in \{175, 194, 213, 231, 250^\circ\text{C}\}$  and  $i = 1, \dots, 37$ .

Orthogonal polynomial contrasts for 1-way ANOVA model:

- $\sim$  model reductions between polynomial regression models,
- split  $SS_{\text{Trt}}$  from 1-way ANOVA into interpretable terms,
- coefficients ( $w_i$ ) listed in textbook Appendix Table D.6.<sup>9</sup>

Illustration: polynomial regression model hierarchy:

Model for $y_i$	Inter- pretation	Contrast estim. (SE)	SS	Model SS	DF	Error SS
$\mu_{\text{temp}(i)} + \varepsilon_i$	ANOVA			3.538	5	0.294
$\downarrow$		(same model!)				
$\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \beta_4 x_i^4 + \varepsilon_i$	4 <sup>th</sup> order regression			3.538	5	0.294
$\downarrow$		-0.038 (.289)	0.000			
$\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \beta_3 x_i^3 + \varepsilon_i$	cubic regression			3.538	4	0.294
$\downarrow$		-0.007 (.112)	0.000			
$\beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i$	quadratic regression			3.538	3	0.294
$\downarrow$		0.400 (.133)	0.083			
$\beta_0 + \beta_1 x_i + \varepsilon_i$	linear regression			3.459	2	0.372
$\downarrow$		-2.19 (.115)	3.332			
$\beta_0 + \varepsilon_i$	i.i.d. sample			0	1	3.831

Conclusion: linear and quadratic contrasts significant, others  $\approx 0$

$\Rightarrow$  quadratic regression model gives best model fit.

<sup>9</sup> The coefficients are valid for equidistant  $x$ 's and equal group sizes.

## BEYOND 1-WAY ANOVA

Methods reviewed for 1-way ANOVA are generalisable to varying extent:

- in multiple regression models, each categorical predictor can/should be assessed separately unless part of interactions (some methods also important for interaction terms  $\Rightarrow$  later lectures),
- construction and assessment of contrasts works for “all” regression models; except that proportion of variation explained is limited to linear models where the factor in question is “unaffected by” (orthogonal to) other effects,
- multiple comparisons are relevant in “all” regression models, but not all methods apply:
  - \* Bonferroni and Holm methods generally applicable,
  - \* many methods limited to balanced ANOVAs, and only few methods extend to GLMs,
  - \* Scheffè method can be applied for Wald-type  $z$ -statistics by comparing  $z^2$  to a  $\chi^2(g-1)$  distribution, where  $g$  = number of groups,
  - \* principles/ideas behind adjustment (e.g. distinction between different error rates) for multiple comparisons are general.

## STATA DO-FILE (PART)

```
insheet using ch03ex1.csv, clear
oneway liverwgt diet, tabulate
anova liverwgt diet /* allows postestimation commands */
regress /* estimates corresponding to anova model */
regress liverwgt i.diet /* totally identical */
* contrasts
anova liverwgt diet
lincom (1.diet+2.diet+3.diet)/3-4.diet
lincom (2.diet+3.diet)/2-1.diet
lincom 2.diet-3.diet
* pairwise comparisons
oneway liverwgt diet, bonferroni /* bonferroni comparisons */
* general method for anova/regression
anova liverwgt diet
pwcompare diet, mcomp(noadjust) /* no adjustment - the default */
pwcompare diet, mcomp(bon) /* Bonferroni method */
* general approach to testing
anova liverwgt diet
test, showorder
matrix input mycon=(1,-1,0,0,0\1,0,-1,0,0\1,0,0,-1,0\
0,1,-1,0,0\0,1,0,-1,0\0,0,1,-1,0)
test, test(mycon) mtest
test, test(mycon) mtest(bon) /* Bonferroni method */
test, test(mycon) mtest(holm) /* Holm method */

insheet using ch03ta1.csv, clear
regress logtime temp
anova logtime temp
scalar F=(.3721-.2937)/(35-32)/.009178 /* lack of fit test */
display F _newline Ftail(3,32,F) /* display F and P-value */
* polynomial contrasts
lincom -2*175.temp-1*194.temp+0*213.temp+1*231.temp+2*250.temp
lincom 2*175.temp-1*194.temp-2*213.temp-1*231.temp+2*250.temp
lincom -1*175.temp+2*194.temp+0*213.temp-2*231.temp+1*250.temp
lincom 1*175.temp-4*194.temp+6*213.temp-4*231.temp+1*250.temp
```

## SOFTWARE NOTES

SAS analysis of 1-way ANOVA and beyond:

- `proc ANOVA`: 1-way and multiple ANOVA,
  - \* limited to balanced designs,
  - \* includes multiple comparison methods (`means` statement),
- `proc glm`: linear models without any restrictions,
  - \* includes multiple comparison methods (`lsmeans` statement),
  - \* includes contrasts (`contrast` and `estimate` statements),
- `proc logistic` (logistic regression) and `proc genmod` (generalized linear models),
  - \* include contrasts (`contrast` statement), but no multiple comparisons,
- `proc multtest`: general multiple testing procedure, for linear models and import of set of unadjusted  $P$ -values.

R analysis of 1-way ANOVA and beyond:

- `oneway.test()` and `pairwise.t.test()` for 1-way ANOVA with multiple comparisons,
- `lm()` and `glm()` functions for fitting linear and generalized linear models (incl. logistic regression), respectively,
- `coef()` and `vcov()` functions extract estimates and the variance-covariance matrix, respectively; further manipulation requires vector/matrix programming (e.g. using `se.contrast()` function) or pre-developed package interface,
- `multcomp` package offers wide variety of multiple comparison procedures, see documentation for use with `lm` and `glm` model fits.