

Index of Lecture 6: Multifactorial design and analysis

Page	Title
1	Practical information
2	One-way ANOVA model and parametrization
3	ANOVA versus regression
4	How to proceed after the ANOVA test?
5	Multiple comparisons: Overview
6	Bonferroni method
7	Holm method
8	Multiple comparisons: Rat data example
9	More about contrasts
10	Scheffè's method
11	Contrasts: Rat data example
12	Beyond one-way ANOVA
13	Multi-factorial designs (recap)
14	Interaction and additivity (recap)
15	Balanced ANOVA example: Bacteria in cheese data
16	More interaction plot examples
17	Textbook examples overview
18	Three-way ANOVA with replication
19	Amylase data: Summary of results
20	Analyse data: Further analyses
21	Additional software notes

PRACTICAL INFORMATION

Today's lecture: expanded review of ANOVA (and regression) with some new topics: contrasts, multiple comparisons and multifactorial designs:

- two-way and three-way ANOVAs: balanced (yes/no) & replication (yes/no).

Notes on for textbook reading (i.e., the GO text):

- Chapter 3 on one-way ANOVA: mostly well-known,¹
- Chapter 4 on contrasts: short chapter, skip Section 4.4,²
- Chapter 5 on multiple comparisons: more detailed than our ambition level, don't focus on mathematical details and read cursorily from Section 5.4.2 onwards,²
- Chapters 8–10 on multifactorial analysis: skip too technical parts³ while focusing on the models and the ANOVA tables,
 - * we are **not** computing SS-values in ANOVA tables manually (only DF-values),
 - * you are **not** expected to compute contrasts involving more than one factor; furthermore, polynomial contrasts can be replaced by regression modelling.

Note: now switching completely to **VHM 802** course schedule and organization; note that home assignment 2 and project information has been posted.

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

² See also article on use of contrasts and multiple comparisons at media page.

³ Technical parts: 177₈–179¹⁵; 179₉–180₅; 181₁₁–181₄; 183₁₀–184¹¹; 184₁₆–185⁴; 192₇–194₁; 205₅–208₆; Figure 8.6; Sections 9.2.3, 9.2.4 and 9.3.

ONE-WAY ANOVA MODEL AND PARAMETRIZATION

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

- rat liver weights in percent of body weight following four diets (labelled 1-4) randomly allocated to rats,

$$y_{ij} = \text{rat liver weight for } j\text{th rat in diet group } i, \\ i = 1, \dots, g \ (g=4); \ j = 1, \dots, n_i \ (n_1=7, n_2=n_4=8, n_3=6),$$

- purpose: assess impact of diets on liver weight.

Statistical model:

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

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

Model parameters (referring to underlying population):

- group means μ_1, \dots, μ_4 , and common within-group/error standard deviation σ .

Alternative formulations of same model:

$$y_i = \mu_{\text{diet}(i)} + \varepsilon_i, \quad i = 1, \dots, 29 \ (\sim \text{observation number}), \\ y_i = \beta_0 + \beta_1 \mathbf{1}_{\text{diet}2(i)} + \beta_2 \mathbf{1}_{\text{diet}3(i)} + \beta_3 \mathbf{1}_{\text{diet}4(i)} + \varepsilon_i, \\ y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad (\text{or, } y_i = \mu + \alpha_{\text{diet}(i)} + \varepsilon_i), \quad \text{with restrictions on } \alpha\text{'s.}^4$$

⁴ Restrictions on (α_i) : either $\alpha_1 = 0$ (Stata; Minitab Regression (default); R), $\alpha_4 = 0$ (SAS), or $\alpha_1 + \dots + \alpha_4 = 0$ (Minitab General Linear Model (default)).

ANOVA VERSUS REGRESSION

= two different frameworks for analyzing the **same model** and presenting the results.⁵

Advantages of ANOVA framework:

- no reliance on an, often artificial, reference category,⁶
- 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: (see 6L–21 for notes on R and SAS)

- more complete regression and ANOVA facilities in Stata,
- more easily accessible facilities in Minitab.

⁵ 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”.

⁶ A common mistake within the regression framework is to explore only comparisons with the reference category, cf. 4bL–10.

HOW TO PROCEED AFTER THE ANOVA TEST?

We rejected overall $H_0: \mu_1 = \dots = \mu_g$, but what relations exist between μ_i 's (which differ)?

- **estimation of parameters:** $\hat{\mu}_i = \bar{y}_i$. and of derived parameters such as **contrasts**:

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

— **examples** (for $g=3$, i.e. 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/tests** for interesting parameters, e.g., using $t^* = t(1 - \frac{\alpha}{2}, DF_E)$:

$$\mu_i : \bar{y}_i. \pm t^* \sqrt{MS_E} \sqrt{(1/n_i)},$$

$$\mu_i - \mu_{i'} : \bar{y}_i. - \bar{y}_{i'}. \pm t^* \sqrt{MS_E} \sqrt{(1/n_i) + (1/n_{i'})},^7$$

- **diagram**, e.g. $\hat{\mu}_i$'s with error bars (interval or margins plot),

- **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,

— therefore **always** an advantage to have any other hypotheses (in addition to the overall H_0) defined **prior to analysis**.

⁷ Note: the margin of error equals the LSD (least significant difference) for unadjusted comparisons between groups.

MULTIPLE COMPARISONS: OVERVIEW

Some terminology and basic facts:

- **type I (error) probability**: probability of rejecting H_0 , when H_0 is actually true,⁷
- **per comparison** or **individual** error rate: type I probability 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; **larger** than individual error rate(s),
- **strong familywise** error rate: probability of rejecting any true null hypotheses (but no impact of false null hypotheses/true rejections),
- **multiple comparison** procedures reduce the type I prob. and increase type II prob. – a trade-off,
- doing (very) many **pairwise *t*-tests** (LSD or Fisher method) is (very) liberal, i.e. likely to have false significances.

(Relatively) **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), and can be used to control specific error rates (GO Display 5.2):

- many require balanced data (Tukey, Duncan) for exact inference,
- some are for special cases (e.g., Dunnett for comparison with control),
- some assume independent tests (e.g., Benjamini & Hochberg’s FDR⁸ method).

⁷ Conversely, the type II (error) probability is for **not** rejecting H_0 , when it is false.

⁸ The false discovery rate (FDR) is the proportion of false rejections out of all rejections, and thus includes the true rejections.

BONFERRONI METHOD

Idea: if we make K comparisons (tests), 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 ϵ/K ,⁹

- either by changing the significance level (to ϵ/K), or by multiplying P -values for individual tests by K (while keeping the significance level unchanged),
- K may be the number of preplanned comparisons, or for unplanned comparisons¹⁰:

$$K = \text{total number of comparisons} = \binom{g}{2} = g(g-1)/2.$$

Notes for Bonferroni method:

- gives also **simultaneous confidence intervals**,¹¹
- is **conservative** (i.e., may lead to too few hypotheses rejected) for controlling the **strong familywise error rate** for unplanned comparisons,
- is available for ANOVA in Minitab only via General Linear Model followed by Comparisons; in Stata, available by `pwcompare` and `test` commands,
- is **flexible**: applies to a wide range of settings/models.

⁹ The (mathematical) justification for the Bonferroni method was explained in VHM 801 (Lecture 9).

¹⁰ Unplanned comparisons include comparisons suggested by the data, e.g. involving the lowest/highest groups.

¹¹ The probability that all CIs simultaneously cover their true value is $\geq 1 - \epsilon$.

HOLM METHOD

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

1) sort 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: $P_{(i)}^H = P_{(i)} \times (K - i + 1)$, for $i = 1, \dots, K$,

3) rules for significance:

(i) if $P_{(i)}^H > \epsilon \Rightarrow$ non-significant (at ϵ),

(ii) if $P_{(i)}^H \leq \epsilon$ and also all $P_{(j)}^H \leq \epsilon$ for all $j = 1, \dots, i$, \Rightarrow significant (at ϵ).

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 (i.e., the $P_{(i)}^H$ above) are available in Stata (`test` command), but the sequential rule (ii) must be checked manually,
- is also **flexible**: applies to the same wide range of settings/models.

MULTIPLE COMPARISONS: RAT DATA EXAMPLE

- group means: $\hat{\mu}_1 = 3.75$, $\hat{\mu}_2 = 3.58$, $\hat{\mu}_3 = 3.60$, $\hat{\mu}_4 = 3.92$,
- assume no preplanned hypotheses or treatment (diet) structure of interest,
- a total of $K = 4 \cdot (4 - 1) / 2 = 6$ pairwise comparisons, shown in the table:

<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 by all methods:
only 4 vs 2,3 significant

* Holm $P <$ Bonferroni P
(except for first P)

Significance letter coding (groups with same letter **not** significantly different; available in Minitab General Linear Model and Stata pwcompare):

- **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: $4^a 1^{ab} 3^b 2^b$.

MORE ABOUT CONTRASTS

- **constructed** to reflect specific (ideally, pre-defined) hypotheses,
- **inference** by 4-step approach from formulas for estimates and SEs, for $w = w(\{\mu_i\})$:

$$\hat{w} = \sum_i w_i \bar{y}_i. \quad \text{and} \quad \text{SE}(\hat{w}) = \sqrt{\text{MS}_E} \sqrt{\sum_i w_i^2 / n_i}.$$
- **variation explained** by contrast (out of SS_{Trt} for the grouping/model):

$$\text{SS}(w) = \hat{w}^2 / (\sum_i w_i^2 / n_i) = \text{MS}_E \times t_w^2, \quad \text{where } t_w = \text{the } t\text{-statistic for testing } w = 0,$$

- **orthogonal contrasts**:

- **idea**: contrasts that explain **different** parts of the variation, to allow **independent** interpretation,¹²
- **definition**: $w = \sum_i w_i \mu_i$ and $w^* = \sum_i w_i^* \mu_i$ are **orthogonal** if: $\sum_i w_i w_i^* / n_i = 0$,
- **fact**: there exist at most $(g - 1)$ pairwise orthogonal contrasts among g groups; these are not unique,
- **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) SS_{Trt} into contrast parts.

¹² In practice, it is not always easy to find useful orthogonal contrasts, in particular in unbalanced designs.

SCHEFFÉ'S METHOD

- corrects for examining **non-preplanned contrasts**¹³, and therefore “allows” to test **contrasts suggested by the data**,
- not available in Minitab/Stata \Rightarrow manual calculation.

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

$$\text{not } \frac{\hat{w} - w}{\text{SE}(\hat{w})} \sim t(\text{DF}_E), \quad \text{but } \left[\frac{\hat{w} - w}{\text{SE}(\hat{w})} \right]^2 / (g - 1) \sim F(g - 1, \text{DF}_E),^{14}$$

for example,

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

Properties:

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

¹³ Method should **not** be used for pairwise comparisons, because in this situation it will be very conservative.

¹⁴ Based on the mathematical relation: $[(\hat{w} - w) / \text{SE}(\hat{w})]^2 / (g - 1) \leq \text{MS}_{\text{Trt}} / \text{MS}_E$.

CONTRASTS: RAT DATA EXAMPLE

- **group means:** $\hat{\mu}_1 = 3.746$, $\hat{\mu}_2 = 3.580$, $\hat{\mu}_3 = 3.598$, $\hat{\mu}_4 = 3.923$,
- **choice of contrasts** guided by group means (in absence of biological hypotheses),
- table of estimates and statistics:

Contrast definition	Weights				Estim.	Stand. err.	Variation	Tests	
	w_1	w_2	w_3	w_4	\hat{w}	SE(\hat{w})	SS(\hat{w}) (%)	Wald t	Scheffé F
$(\mu_1 + \mu_2 + \mu_3)/3 - \mu_4$	1/3	1/3	1/3	-1	-0.281	0.085	0.456 (78.9%)	-3.32	3.67
$(\mu_2 + \mu_3)/2 - \mu_1$	-1	1/2	1/2	0	-0.157	0.094	0.114 (19.6%)	-1.66	0.92
$\mu_2 - \mu_3$	0	1	-1	0	-0.018	0.110	0.001 (0.2%)	-0.17	0.01

- t -tests are assessed in $t(\text{DF}_E) = t(25)$ with $t(0.975, 25) = 2.06$;
 F -tests are assessed in $F(g-1, \text{DF}_E) = F(3, 25)$ with $F(0.95, 3, 25) = 2.99$.

Interpretations:

- contrast between last and first three groups $\sim 80\%$ of variation, and is significant both with Wald test (\sim pre-planned) and Scheffé test,
- other contrasts are far from significant (in particular with F -test),
- due to the unequal group sizes, the contrasts are not orthogonal (seen for example by their proportions out of SS_{Trt} not adding up to 100%).

BEYOND ONE-WAY ANOVA

Methods reviewed for one-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 slides),
- construction and assessment of **contrasts** work 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,
 - * Scheffé’s method can be applied for Wald-type z -statistics in GLMs by comparing z^2 to a $\chi^2(g-1)$ distribution, where g = number of groups,
 - * principles/ideas behind adjustment (e.g. the distinction between different error rates) for multiple comparisons are general.

MULTI-FACTORIAL DESIGNS (RECAP)

Several factors in the same design? — Yes!, in good designs it is possible to separate effects of different factors from each other \Rightarrow

- possible to study **combined effect** of several factors (the presence of interaction),
- if interaction is absent: **cheaper** (less experimental units) than in several one-at-a-time experiments,¹⁵
- **increased scope** of the experiment/study,

and **analysing multi-factorial data by each factor separately**: is generally **wrong** and only gives valid results if at most one factor is of importance.

Design terminology and issues:

- **balancedness**: all (combined) groups are equally large, otherwise unbalanced,
- **completeness**: all (combined) groups are present (no empty cells), otherwise incomplete (should be **avoided**, or at least analyzed **with care**),
- **replication**: some of (combined) groups have $n > 1$, otherwise no replication (all $n = 1$),
- factorial structure can be combined with blocking structures (next lecture).

¹⁵ Simple example: effects of alcohol (A: 0/1) and sleeping pills (B; 0/1). Two “one-at-a-time” studies (effect of A at fixed level of B; effect of B at fixed level of A), each with 20 subjects, give (in the absence of interaction) same precision as a combined study with 5 subjects per alcohol×pill combination.

INTERACTION AND ADDITIVITY (RECAP)

Interaction (synergism/antagonism, covariation):

- the **combined effect of two factors**¹⁶ is not predictable from isolated effects of each of them examined separately,
- the effect of one factor **depends on the level** of the other factor,
- **non-parallel lines in interaction plot** (i.e., plot of combined means versus one factor), where parallel lines \sim **additive effects** (and hence no interaction).
- **note**: interaction is dependent on scale (of outcome), so affected by transformation.

Dealing with interaction between two factors:

- **decomposition of SS** for combined factor $A \times B$:

$$SS_{A \times B} = SS_A + SS_B + SS_{A*B},$$

and each SS corresponds to treatment contrast(s), as illustrated on next page,

- DF formula: $DF_{A*B} = DF_A \cdot DF_B$ (if $A \times B$ complete),
- in presence of an **“important” interaction** (significant and strong), the main effects are of no direct interest¹⁷, but contrasts in combined factor may be of interest.

¹⁶ Interaction between 3 factors: the interaction between two of the factors depends on the level of the third factor.

¹⁷ In GO terminology (Section 8.11), the hierarchy is retained by not removing main effects involved in an interaction, so they may not need to be tested at all.

BALANCED ANOVA EXAMPLE: BACTERIA IN CHEESE DATA

Example 8.6 (p. 178): Nonstarter bacteria in cheddar cheese, a 2×2 -factorial for the bacterial strains R50 and R21:

- SS_{Trt} computed from the four treatments means (1.709, 1.952, 2.153, 2.444),

$$SS_{\text{Trt}} = 3[(1.709 - 2.065)^2 + (1.952 - 2.065)^2 + (2.153 - 2.065)^2 + (2.444 - 2.065)^2] = 0.872,$$

- **decomposition** of SS_{Trt} by **orthogonal contrasts**:

Contrast $w(\{\mu_i\})$	w_1	w_2	w_3	w_4	\hat{w}	$SS(\hat{w})$	$F = t_w^2$
main R50	1	1	-1	-1	-0.935	0.656	7.23
main R21	1	-1	1	-1	-0.535	0.214	2.36
interaction	1	-1	-1	1	0.048	0.002	0.02

SS_{Trt}

- **ANOVA table**:

Source	DF	SS	MS	F	P -value
R50	1	0.656	0.656	7.23	0.028
R21	1	0.214	0.214	2.36	0.16
Interaction	1	0.002	0.002	0.02	0.89
Error	8	0.726	0.091		
Total	11	1.598			

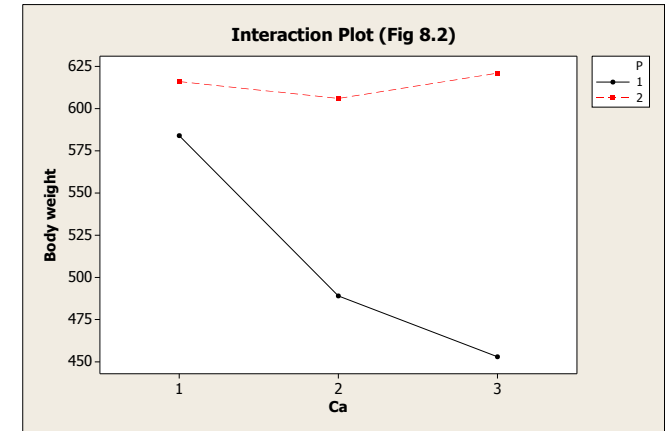
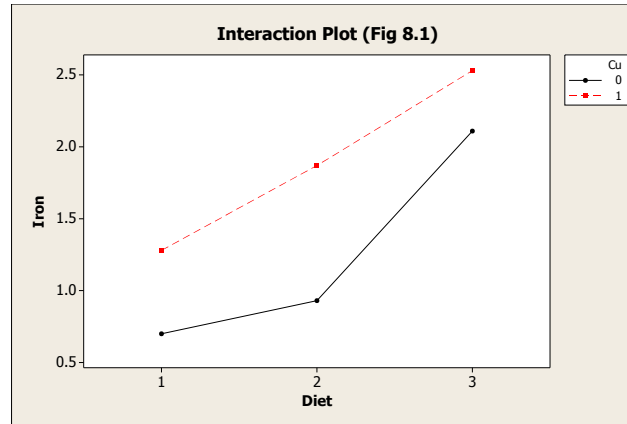
- **conclusions**:

- * absolutely **non-significant interaction** between effects of R50 and R21,
- * positive and weakly significant **main effect of R50**: adding R50 increases TFAA.
- * positive, but non-significant **main effect of R21** \Rightarrow no demonstrated R21 effect,
- the three tests are independent of each other (due to orthogonality of contrasts).

MORE INTERACTION PLOT EXAMPLES

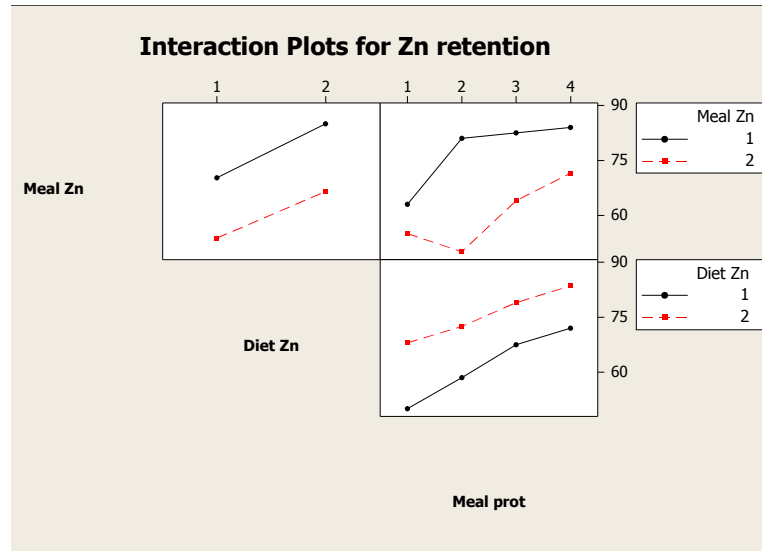
Rat data (Example 8.2):

- **outcome:** iron levels in liver tissue,
- **factors:** milk diet (3), copper deficiency (2),
- **interpretation:** close to additive effects.



Chick data (Example 8.4):

- **outcome:** body weights,
- **factors:** Ca supplement (3), P supplement (2),
- **interpretation:** interaction, with Ca effect for low P (P = 1) only.



Rat data II (Example 8.5):

- **outcome:** Zn retention,
- **factors:** Diet Zn (2), final Meal Zn (2), final Meal protein (4),
- **interpretation:** interaction Meal Zn * Meal protein; other effects additive.

TEXTBOOK EXAMPLES OVERVIEW

ANOVA examples of Chapters 8, 9.1-2 and 10.1-2:

- 8.6: non-starter bacteria in cheese: 2×2 factorial with 3 replicates,
- 8.8: page faults (#) in CPU experiment (program algorithms and settings): $2 \times 3 \times 3 \times 3$ factorial with no replication (**models without replication**¹⁸),
- 8.10, 9.3: amylase activity in maize (analysis and growth temp., variety): $2 \times 2 \times 8$ factorial with 3 replicates, both temperatures quantitative (**polynomial contrasts**¹⁹),
- 9.2: unspecified outcome and factors: 2^4 ($2 \times 2 \times 2 \times 2$) factorial with 2 replicates (**one-cell interaction**²⁰),
- 9.4: seed viability (storage conditions): 3×7 factorial with 3 replicates, both factors quantitative (**polynomial contrasts**¹⁹),
- 10.1-2: amylase activity (8.10) with one observation omitted (**unbalanced data**²¹),
- 10.3: unspecified outcome and factors: 2×2 factorial with highly unequal replication (**unbalanced data**²¹).

¹⁸ Most common method is to omit highest order interaction(s) and thus obtain estimate of residual error: conservative approach if these interactions are non-zero.

¹⁹ For a balanced factor with quantitative (equidistant) values, contrasts \sim **polynomial terms**; not in course syllabus.

²⁰ An interaction may be due to a single mean value deviating from additivity.

²¹ Unbalanced data no longer have orthogonal factorial contrasts \Rightarrow effects depend on other terms in model (as in general regression); two types of sum of squares:

- **partial/adjusted SS**: remove term while keep all others in;
- **sequential SS**: remove terms sequentially (\uparrow) so only terms above are kept in.

THREE-WAY ANOVA WITH REPLICATION

Amylase activity example (GO Example 8.10, p. 195):

- amylase specific activity of sprouted maize under 32 treatment conditions:

y_{ijkl} = activity for maize plant batch l of type (i, j, k)

$i = 1, \dots, 8 \sim$ analysis temperature (10,13,15,20,25,30,35,40°C),

$j = 1, 2 \sim$ growth temperature (13,25°C),

$k = 1, 2 \sim$ variety (B73,O43), and $l = 1, 2, 3 \sim$ replicate,

- completely randomized design (if full randomization),
- **full statistical model** = one-way ANOVA with 32 groups: $y_{ijkl} = \mu_{ijk} + \varepsilon_{ijkl}$, where the errors $\varepsilon_{1111}, \dots, \varepsilon_{8223}$ are i.i.d. and $\sim N(0, \sigma^2)$.
- **decomposition** of combined factor levels into main effects and interactions (first order and second order),

$$\mu_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk},$$

* α_i = main effect of atemp (level i),

* $(\alpha\beta)_{ij}$ = interaction atemp*gtemp (levels i and j),

* $(\alpha\beta\gamma)_{ijk}$ = interaction atemp*gtemp*var (levels i, j and k).

- **model checks** based on full model \Rightarrow same approach as in one-way ANOVA: demonstrating a log-transform to be appropriate.

AMYLASE DATA: SUMMARY OF RESULTS

ANOVA table
(natural log scale):

Source	DF	Seq SS	Adj SS	Adj MS	F	P
atemp	7	3.01613	3.01613	0.43088	78.86	0.000
gtemp	1	0.00438	0.00438	0.00438	0.80	0.374
var	1	0.58957	0.58957	0.58957	107.91	0.000
atemp*gtemp	7	0.08106	0.08106	0.01158	2.12	0.054
atemp*var	7	0.02758	0.02758	0.00394	0.72	0.654
gtemp*var	1	0.08599	0.08599	0.08599	15.74	0.000
atemp*gtemp*var	7	0.04764	0.04764	0.00681	1.25	0.292
Error	64	0.34967	0.34967	0.00546		
Total	95	4.20202				

o **high $R^2 = 1 - 0.3497/4.2020 = 91.7\%$,**

o **non-significant terms:**

atemp*gtemp*var and atemp*var, and atemp*gtemp is close to significant,²²

o **final model:** atemp*gtemp gtemp*var,

but no need to refit model (pool variance terms) because DF_E is large,

o **additive effects** of atemp and var for given gtemp,

o **interaction plots:** strong, parabolic-type effect of atemp (no obvious interpretation of gtemp interaction), and different effects of gtemp for the varieties,

o **presentation** of gtemp*var results by means with SE:

var=B73		var=O43		SE
gtemp=13	gtemp=25	gtemp=13	gtemp=25	
5.85	5.92	5.75	5.70	0.015

o **atemp effects:** multiple comparisons or polynomial modelling.

²² Note that gtemp should not be labelled as non-significant, because it is involved in a significant interaction.

AMYLASE DATA: FURTHER ANALYSES

Polynomial modelling (GO Example 9.3):

- atemp polynomial contrasts not attractive due to non-equidistant temperatures,
- for simplicity, refit with (clearly) non-significant atemp terms omitted before polynomial modelling,
- **quadratic model** in atemp $F = [(0.52148 - 0.42489)/(88 - 78)]/0.00545 = 1.71$
has lack-of-fit test: $\sim F(10, 78)$ under H_0 , $P = 0.094$,
 - no formal evidence against quadratic model, but higher order terms may improve fit,
- **cubic model** in atemp $F = [(0.45037 - 0.42489)/(86 - 78)]/0.00545 = 0.58$
has lack-of-fit test: $\sim F(8, 78)$ under H_0 , $P > 0.5$,
 - cubic polynomial model seems appropriate,
- **interpretation** of fitted model(s): plots of predicted curves.

Illustration: effect of unbalancedness (GO Examples 10.1-2): dropping one observation, the first row in dataset:

- sequential and partial/adjusted (SAS type I and III) sum of squares do no longer coincide: SS values for gtemp range within 0.00140–0.00330 across different models,
- model building must be sequential (as in regression),
- simple means are no longer same as fitted means \Rightarrow care needed with margins.

ADDITIONAL SOFTWARE NOTES

R analysis of one-way ANOVA and beyond:

- `oneway.test()` and `pairwise.t.test()` for one-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.

SAS analysis of one-way ANOVA and beyond:

- `proc ANOVA`: one-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.