Biostats VHM 801 Course
Fall 2016, Atlantic Veterinary College, PEI
Henrik Stryhn

Index of 13-L

Page	Title
1	Practical information
2	Data examples for 2-way ANOVA
3	Factorial designs
4	Notation for 2-way ANOVA
5	1-way ANOVA for 2-way factorial
6	Decomposing a 2-way table of means I
7	Decomposing a 2-way table of means II
8	Additivity and interaction
9	2-way ANOVA models
10	2-way ANOVA analysis
11	Supplementary exercises 13.3 and 13.4
12	Summary of 2-way ANOVA for tomato data
13	Model checking
14	2-way ANOVA without replication
15	Friedman's test
16	P-values
16	Summary notes

Practical Information

Last regular lecture!!

Major news:

- course evaluation will take place in the next lecture,
- course syllabus for final exam added to homepage,
- last home assignment due today. . . ,
- midterm mark: by Dec. 5th, each of you must tell me (in writing/per e-mail) whether you want to use it.

Today's lecture:

- summary worksheet (last!) on regression,
- main topic: 2-way ANOVA (analysis of variance),
  - ∗ non-parametric Friedman's test (extra topic),
- textbook chapters[1] are very (too) brief — some more details in lecture and in home assignment 4 of 2003 (and even more details in VHM 802 and 812),
- tips for comparing and presenting groups (11L–2) apply here as well.

Next (absolutely last!) lecture:

- details about exam,
- brief course review (call for requested topics!).

[1] PSLS 3e Supplement: Sections 26.4-6 (recommended!); S: not covered; IPS 7e: Chapter 13.

Data Examples for 2-Way ANOVA

Energy expenditures in Burkina Faso:[2]

- mean energy expenditures in farming families, divided into man/woman and dry/wet season,
- summarized data (no raw data available),
-

Energy expenditure (calories per day)		Gender	
		men	women
Season	dry	2310	2320
	wet	3460	2890

Phosphorus levels in tomato plants: (PSLS 3e, Ex. 26.12)

- 3 levels of nitrogen fertilizer (0/28/160 kg/ha) applied to two genotypes of tomato plants (wild/mutant, wild being susceptible to Mycorrhizal fungi),
- genotypes $\sim$ blocks, 6 replicates per treatment$\times$block.

Lymphocytes (1000's per μl blood) in mice:

- 4 treatments (A-D), of which D is placebo,
- 4 litters of mice, and 4 mice used from each litter,
- block design.

Treatment	Litter 1	2	3	4
A	7.1	6.1	6.9	5.6
B	6.7	5.1	5.9	5.1
C	7.1	5.8	6.2	5.0
D	6.7	5.4	5.7	5.2

[2] Based on Payne: Nutrition adaptation in man: social adjustments and their nutritional implications, in Blaxter & Waterlow (eds.): *Nutrition Adaptation in Man.*, Libbey, London, 1985.

Factorial Designs

Factor (categorical variable):

- grouping of observations into categories/levels, either by symbols (e.g. letters, roman numbers) or numbers,
- *explanatory variable*, e.g. treatment/control, litter ...
- usually, it does not matter if factors are coded by numbers or symbols: use most natural coding,
- if factors are coded numerically, check DF to ensure their modelling as a grouping.[3]

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

- cheaper (less exp. units) than in separate experiments,[4]
- possible to study combined effect of several factors,
- increased scope of the experiment,

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

Two types of randomization for factorial experiments:

- completely randomized design,
- (randomized) block design.

[3] The software may misunderstand the factor as continuous and estimate a slope.
[4] The advantage arises e.g. if assessment of nitrogen effects can be done on wild and mutant tomato plants combined, in *additive* models introduced later.

NOTATION FOR 2-WAY ANOVA

Data layout and notation:

obs. X_{ijk}		column (C) factor $\sim j$: 1	$\dots$	j	$\dots$	J
	1	$X_{111},\dots,X_{11n_{11}}$	$\dots$	$X_{1j1},\dots,X_{1jn_{1j}}$	$\dots$	$X_{1J1},\dots,X_{1Jn_{1J}}$
row (R)	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
factor	i	$X_{i11},\dots,X_{i1n_{i1}}$	$\dots$	$X_{ij1},\dots,X_{ijn_{ij}}$	$\dots$	$X_{iJ1},\dots,X_{iJn_{iJ}}$
$\sim i$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$
	I	$X_{I11},\dots,X_{I1n_{I1}}$	$\dots$	$X_{Ij1},\dots,X_{Ijn_{Ij}}$	$\dots$	$X_{IJ1},\dots,X_{IJn_{IJ}}$

- X_{ijk} = kth observation in the group defined by row factor R$=i$ and column factor C$=j$,[5]
  - $i = 1,\dots,I$, and I=number of levels of R/rows,
  - $i = j,\dots,J$, and J=no. of levels of C/columns,
  - $k = 1,\dots,n_{ij}$, and n_{ij}=no. of obs. in (i,j)th group.
- denote also by $N = \sum_{ij} n_{ij}$ the total no. of observations, and by $\bar{X} = \sum_{ijk} X_{ijk}/N$ the overall mean,
- terminology: the dataset/design
  - is balanced, if all groups equally large, $(n_{11}=\dots=n_{IJ})$, otherwise unbalanced,
  - is complete, if all IJ groups present, otherwise incomplete,
  - has replication, if some of the n_{ij}'s > 1, otherwise no replication (all $n_{ij} = 1$).

[5] IPS uses the notation: A=row factor, B=column factor.

1-Way ANOVA for 2-Way Factorial

In a 2-way design *with replication*, focusing only on the grouping from the row and column factors (IJ groups) and otherwise forgetting about row and column factors
$\Rightarrow$ 1-way ANOVA for combined factor with IJ levels:

- Model:
  $$X_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \quad \text{for } \varepsilon_{ijk}\text{'s i.i.d. and } \sim \mathrm{N}(0,\sigma),$$
  and where μ_{ij}'s are group (population) means,
- Estimation:
  $$\hat{\mu}_{ij} = \bar{X}_{ij\cdot} \text{ (combined group means)},$$
  $$\hat{\sigma}^2 = s_p^2 = \sum_{ij} \frac{n_{ij}-1}{N-IJ} s_{ij}^2 = \text{MSE}, \text{ and DFE} = N - IJ,$$
  where s_{ij}=sample standard deviation in group (i,j),
- ANOVA table:

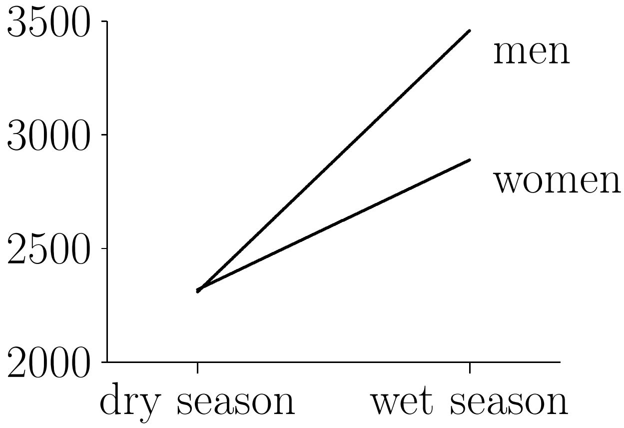
Source	DF	SS	MS	F
Groups	$IJ-1$	$\sum_{ij} n_{ij}(\bar{X}_{ij}-\bar{X})^2$	SSG/DFG	MSG/MSE
Error	$N-IJ$	$\sum_{ijk}(X_{ijk}-\bar{X}_{ij})^2$	SSE/DFE	
Total	$N-1$	$\sum_{ijk}(X_{ijk}-\bar{X})^2$		

- Problem: analysis does not directly give information about row and column factors separately $\Rightarrow$
  need to decompose (split up) model's group terms.

DECOMPOSING A 2-WAY TABLE OF MEANS I

Example: Energy expenditures in Burkina Faso:

Energy exp. (calories)		Gender men	Gender women	Mean
Season	dry	2310	2320	2315
	wet	3460	2890	3175
Mean		2885	2605	2745



Different ways to look at the data:

(i) four separate groups,

(ii) two gender groups for each season,

(iii) two season groups for each gender,

(iv) (overall level), two season groups, two gender groups, association between gender and season.

Decomposition of means corresponding to (iv):

$\bar{X}$ overall mean

2745	2745
2745	2745

$\bar{X}_{\cdot j} - \bar{X}$ gender effect

140	–140
140	–140

$\bar{X}_{i\cdot} - \bar{X}$ season effect

–430	–430
430	430

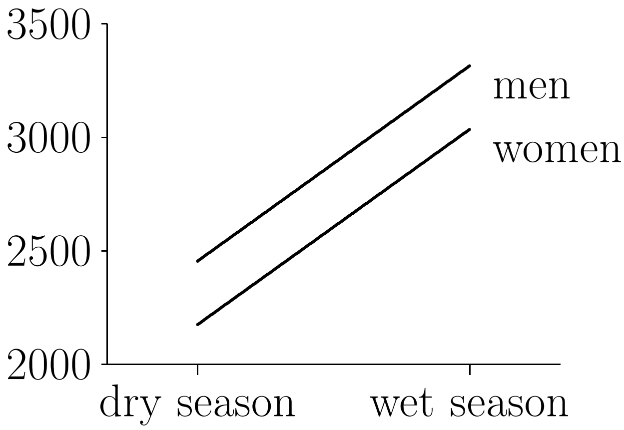
$\bar{X}_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}$

–145	145
145	–145

Decomposing a 2-Way Table of Means II

Modified energy expenditures in Burkina Faso:

Energy exp. (calories)		Gender: men	Gender: women	Mean
Season	dry	2455	2175	2315
	wet	3315	3035	3175
Mean		2885	2605	2745



Decomposition of means corresponding to (iv):

$$\underset{\text{overall mean}}{\bar{X}} \quad \begin{array}{|cc|}\hline 2745 & 2745 \\ 2745 & 2745 \\ \hline\end{array} \qquad \begin{array}{|cc|}\hline 140 & -140 \\ 140 & -140 \\ \hline\end{array} \quad \underset{\text{gender effect}}{\bar{X}_{\cdot j} - \bar{X}}$$

$$\underset{\text{season effect}}{\bar{X}_{i\cdot} - \bar{X}} \quad \begin{array}{|cc|}\hline -430 & -430 \\ 430 & 430 \\ \hline\end{array} \qquad \begin{array}{|cc|}\hline 0 & 0 \\ 0 & 0 \\ \hline\end{array} \quad \begin{array}{c}\bar{X}_{ij} - \bar{X}_{i\cdot} \\ -\bar{X}_{\cdot j} + \bar{X}\end{array}$$

Comparison of two variants of Burkina Faso data:

- same overall level, same overall (average) effects of gender and season,
- modified data: parallel lines ⇒ additive effects, (same effect of one factor at all levels of other factor(s)),
- original data: non-parallel lines ⇒ non-additive effects, or interaction between the factors gender and season.

Interaction and Additivity

Interaction — some other words:

- synergism,
- epistasy (genetics),
- covariation,
- association.

Interaction between two factors:

- the main effects provide an incomplete description, i.e.: the combined effect of two factors is not predictable from the isolated effect of each of them when examined separately,
- the effect of the first factor depends on the level of the second factor (or vice versa) — "it depends…"
- no additivity between factors,
- non-parallel lines.

Always remember!
— interaction is the opposite of additivity, or
— additivity means no interaction!

2-Way ANOVA Models

Basic model (in two equivalent formulations):

$$X_{ijk} = \mu_{ij} + \varepsilon_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk},$$

where the random terms ("errors") ε_{ijk} are i.i.d. and $\sim$ $N(0, \sigma)$, and $i = 1, \ldots, I$; $j = 1, \ldots, J$; $k = 1, \ldots, n_{ij}$.

Parameters and interpretations:

- μ_{ij} = mean of (i, j)th group,
- μ = overall mean,
- α_i = "main effect" of ith row group,
- β_j = "main effect" of jth column group,
- γ_{ij} = "interaction effect" of (i, j)th group,
- σ = population standard deviation (for X's and ε's),

technical note: it is necessary to put some restrictions on α's, β's and γ's (otherwise too many parameters).

Overview of models:[6] (R $\sim$ row factor, C $\sim$ column factor):

EX_{ij}	Model formula	Interpretation	Corresponding model
μ_{ij}	(μ+)R+C+R*C	interaction R∗C	1-way ANOVA for R×C
$\mu + \alpha_i + \beta_j$	(μ+)R+C	additivity	*new* model
$\mu + \alpha_i$	(μ+)R	no effect of C	1-way ANOVA for R
$\mu + \beta_j$	(μ+)C	no effect of R	1-way ANOVA for C
μ	(μ)	no effects	1-sample analysis

note: often the overall mean μ is not included in the model formula.

[6] Final models, i.e. after disregarding non-significant terms

2-Way ANOVA Analysis

Very similar to 1-way ANOVA

- same steps: estimation, model check, ANOVA table with F-tests, contrasts and/or graphical presentation,
- more rows in ANOVA table and more tests,
- σ estimated as $\sqrt{\text{MSE}}$ with DFE degrees of freedom.

General 2-way ANOVA table:

Source of variation	Degrees of freedom	Sum of squares	Mean square	Hypothesis/ F-statistic
Row factor R	$I-1$	$\sum_{ij} n_{ij}(\bar{X}_{i\cdot}-\bar{X})^2$	SSR/DFR	H_0: no row eff. $F=$MSR/MSE
Column factor C	$J-1$	$\sum_{ij} n_{ij}(\bar{X}_{\cdot j}-\bar{X})^2$	SSC/DFC	H_0: no column eff. $F=$MSC/MSE
Interaction R×C	$(I-1)(J-1)$	SSG−SSR−SSC	SSRC/DFRC	H_0: no interaction $F=$MSRC/MSE
Error	$N-IJ$	$\sum_{ijk}(X_{ijk}-\bar{X}_{ij})^2$	SSE/DFE	
Total	$N-1$	$\sum_{ijk}(X_{ijk}-\bar{X})^2$		

Some hints for analysis:

- test for interaction first: if significant, base conclusions on $\hat{\mu}_{ij}$'s ($\bar{X}_{ij\cdot}$'s) alone (contrasts, pairwise comparisons),
- tests for main effects not meaningful in presence of strong interaction! ("read ANOVA table from bottom to top"),
- estimation and interpretation for additive model: separately for row and column factors (based on row and column means).

Supplementary Exercises 13.3 and 13.4

Suppl. ex. 13.3: (response, factors, no. of repl., I, J, N)

(a) response = number of hours of sleep "on a typical night", factors: smoking categories ($I=3$), gender ($J=2$), $n_{ij}=120$, and $N=720$,

(b) response = strength of concrete specimens, factors: mixtures ($I=4$), cycles of freezing and thawing ($J=3$), $n_{ij}=2$, and $N=24$,

(c) response = score on final exam, factors: teaching methods ($I=3$), student's subject of study ($J=2$), $n_{ij}=7$ and $N=42$.

Suppl. ex. 13.4: (sources and degrees of freedom)

(a) smoking categories (DF = 2), gender (DF = 1), interaction (DF = 2), error (DF = 714) and total (DF = 719),

(b) mixtures (DF = 3), cycles (DF = 2), interaction (DF = 6), error (DF = 12) and total (DF = 23),

(c) teaching methods (DF = 2), study subject (DF = 1), interaction (DF = 2), error (DF = 36) and total (DF = 41).

Summary of 2-Way ANOVA for Tomato Data

Statistical model:

$$X_{ijk} = \mu_{ij} + \varepsilon_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk},$$

where $i = 1, 2, 3$ (nitrogen: 0, 28, 160), $j = 1, 2$ (mutant, wild), and $k = 1, \ldots, 6$, or as a model formula:

$$\ln(\text{Phosphorus}) = \text{N} + \text{Type} + \text{N*Type} + \text{Error}.$$

ANOVA table:

Source	DF	SS	MS	F	P-value
Nitrogen	2	0.9171	0.4586	28.4	<0.0005
Genotype	1	3.9654	3.9654	246	<0.0005
Interaction N∗G	2	0.0536	0.0268	1.66	0.21
Error	30	0.4843	0.0161		
Total	35	5.4204	$s = \sqrt{\text{MSE}} = 0.127$		

Hypothesis H_0: all γ_{ij}'s$=0$ (no interaction), H_a: not H_0,

- ∗ Test of H_0: $F_{\text{obs}} = 1.66$, DF$=(2, 30)$, $P=0.21$,
- ∗ Conclusion: no evidence of interaction (on log-scale!),

Main effect hypotheses: strong significance for both N and type (both $P < 0.0005$) ⇒ clear evidence of (some) differences among N groups and between the two genotypes.

Presentation (on log-scale): (using $t^* = t_{.975}(30) = 2.042$)

	Nitrogen			Genotype	
statistic	0	28	160	Mutant	Wild
$\bar{X}_{i\cdot}$ or $\bar{X}_{\cdot j}$	−1.040	−1.223	−1.431	−1.563	−0.900
95% CI	$\pm\, t^*\, s/\sqrt{12} = \pm\, 0.075$			$\pm\, t^*\, s/\sqrt{18} = \pm\, 0.061$	
$\text{LSD}_{0.95}$	$t^*\, s\sqrt{2/12} = 0.106$			N/A	

Conclusion: all N groups clearly statistically different.

Model Checking

Assumptions for 2-way ANOVA model:
same as for 1-way ANOVA (independence, normality, variance homogeneity, same means in row$\times$column groups).

ANOVA models with (1), 2 or more factors:

- often few replications $\Rightarrow$ difficult to check model assumptions separately for each group,
- use instead residuals ("observed $-$ expected") to check model assumptions, similar to linear regression:
  - variance homogeneity:
    plot stand. residuals against model's fitted/expected values and look for unequal variances across range of fittes values,[7]
  - normal distribution:
    normal probability plot of standardized residuals,
  - outliers:
    look for extreme stand. residuals (with same rules as for linear regression),
  - other model violations:
    plot stand. residuals against data order (if applicable) or other variables.

[7] In a two-way ANOVA with replication, the one-way tools: i) max/min rule, ii) variance test, still apply, when groups are defined by combinations of both factors.

2-Way ANOVA without Replication

No replication:

- only one obs. per row×column cell (all $n_{ij} = 1$),
- usually the case in *block designs*,
- data example: lymphocyte data for mice.

Special considerations for ANOVAs without replication:

- cannot estimate interaction, must use *additive model*:
  $X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}$, for ε_{ij}'s i.i.d. and $\sim \mathrm{N}(0, \sigma)$,
  model formula: Lymph = Treatm + Litter + Error,
- the model standard deviation σ *includes* any interaction between the two factors ⇒ model most useful when interaction believed to be absent or weak.

ANOVA table for lymphocyte data:

Source	DF	SS	MS	F	P-value
Treatments	3	1.3250	0.4417	8.83	0.005
Litters	3	6.3950	2.1317	42.6	<0.0005
Error	9	0.4500	0.0500		
Total	15	8.1700			

Conclusions:

- clear treatment effect: only treatment A significantly (5% level) different from placebo (D): $\mathrm{LSD}_{0.95} = 0.36$,
- very clear litter effect, and litter 1 has highest values.

Friedman's Test

= nonparametric (rank) test for treatment effect in a randomized (unreplicated)[8] block design,

- two factors: treatment (`tx`) and block, and one observation per `tx` in each block,
- hypothesis H_0 : no difference between `tx` in the outcome, against a two-sided alternative ($\sim$ ANOVA model),
- no test for (or assumptions about) block effects,[9]
- the tests works by ranking observations within blocks, summing ranks across across blocks, and comparing rank sums in a similar way as for Kruskal-Wallis test (with an approximate χ^2-distribution, df $= I - 1$) $\Rightarrow$ computed by software (Minitab).

Lymphocyte data example — observations and ranks:

Treatment	Litter 1	2	3	4	Rank sum
A	7.1 (3.5)	6.1 (4)	6.9 (4)	5.6 (4)	15.5
B	6.7 (1.5)	5.1 (1)	5.9 (2)	5.1 (2)	6.5
C	7.1 (3.5)	5.8 (3)	6.2 (3)	5.0 (1)	10.5
D	6.7 (1.5)	5.4 (2)	5.7 (1)	5.2 (3)	7.5

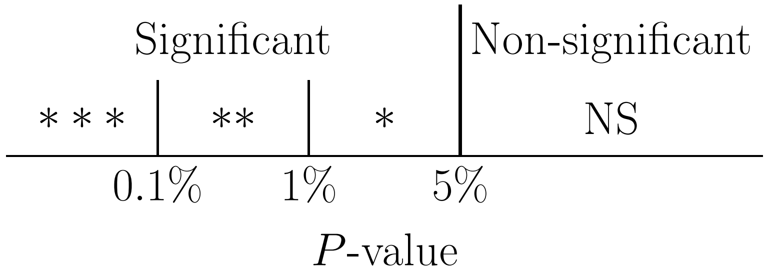
Test statistic: $S = 7.74$, P-value (adj. for ties): $P = 0.052$ $\Rightarrow$ close to significant (here, higher P than for ANOVA).

[8] A less commonly used extension to designs with replication and interaction exists, termed the Scheirer-Rare-Hare test: *Biometrics* **32**, 429–434.

[9] It is possible to get a test for block effects by switching the roles of treatment and block, and recomputing the test.

P-VALUES

Significance as determined from P-value[10]:



Interpretation and suggested wording (personal preferences):

- P-values above 0.05 are usually denoted *non-significant* and interpreted as indicating that the result (observed value of the test statistic) could have occurred by coincidence if H_0 is true, so that H_0 cannot be rejected,
- P-values just around 0.05 (no matter if just above or below) are *borderline significant*, and give *indication* that H_0 may not be true,
- formally, a P-value below 0.05 is *evidence against* H_0,
- P-values at 0.01 or below are *clearly significant* and give clear indication that H_0 is false, so that H_0 should probably be rejected,
- P-values at 0.001 or below are *strongly significant* and show that the data are incompatible with H_0, so that H_0 without doubt should be rejected.

[10] Warning ! — do not confuse P-values and values for p's:

- P-values are values (certain probabilities) computed to interpret test results,
- p's are (probability) parameters in particular models, e.g. $B(n, p)$.

Summary Notes

Key words and concepts for 2-way ANOVA:

- multifactorial designs:
  - advantages over single factor designs: larger scope, allows assessment of combined effects, potentially more efficient,
  - characterizations: factors, factorial notation (e.g. $2\times$ 2-design), balancedness, completeness, replication,
- modelling: 1-way ANOVA for combined factor, decomposition of combined means into main effects and interaction, parametrization and parameter restrictions (technical),
- interaction: non-parallel curves, non-additive effects, effect of one factor depends on another factor,
- analysis: ANOVA table, F-tests for different hypotheses: testing interaction first, interaction plot, model checking by residuals, post-ANOVA analysis using LSD-values and pairwise comparisons (possibly with Bonferroni adjustments),
- nonparametric Friedman's test.