

Index of 10-L

Page	Title
1	Practical information
2	Data example(s): Reading scores
3	Notation for 1-way ANOVA
4	Statistical model
5	Estimation
6	Model checking
7	Hypothesis and test
8	F -distribution
9	ANOVA table
10	Exercises 12.1, 12.9, 12.25
11	Contrasts
12	Pairwise comparisons
13	Bonferroni method for multiple tests
14	Summary of 1-way ANOVA for reading scores
15	Sample size for 1-way ANOVA
16	Kruskal-Wallis test
17	Notes on home assignment II
18	Notes on mid-term exam
19	Summary notes

PRACTICAL INFORMATION

Major news:

- mid-term returned today (solution at course homepage!):
 - * mid-term review possible in today's lab review,
 - * skip mark? — it's your (individual) decision; you'll need to tell me at the end of the course (Dec 2nd),
- home assignment II: also returned today (w/ solution),
- home assignment III at web page: worth 10% of course mark, deadline Thursday 10/11,
- optional project proposal: also due Thursday 10/11,
- now jumping ahead in textbooks:
we'll come back to regression and correlation.

Today's lecture — another new method/analysis. . .

- main topic: 1-way ANOVA (analysis of variance),¹
- contrasts are not part of the course curriculum!²
- jump over: references to residuals and R^2 ,
(we'll talk about these later)
- additional topic: Kruskal-Wallis test (PSLS Chapter 27) — non-parametric 1-way ANOVA.

¹ PSLs 3e: Chapters 24 & 26; S: Section 11.3 (too briefly); IPS 7e: Chapter 12.

² Contrasts are useful!: Yossa & Verdegem (2015), *Aquaculture* **437**, 344–350.

DATA EXAMPLE(S): READING SCORES

Teaching reading comprehension³:

- 66 students, randomly assigned to one of 3 teaching groups, with 22 students in each group,
 - * **Basal**: traditional method,
 - * **DRTA** and **Strat**: innovative methods,
- 2 pre-tests (before teaching) and 3 post-tests (after teaching),
- questions of interest: compare the 3 groups, in particular the new ones versus **Basal**, and also **DRTA** versus **Strat**,
- in the lecture, we look at **pre1** and **post3**.

Outline of data set:

variable	group	observations (22 in each row)							mean	sd
Basal	pre1	4	6	9	12	16	...	9	10.5	2.97
	post3	41	41	43	46	46	...	32	41.0	5.64
DRTA	pre1	7	7	12	10	16	...	10	9.7	2.69
	post3	31	40	48	30	42	...	49	46.7	7.39
Strat	pre1	11	7	4	7	7	...	8	9.1	3.34
	post3	53	47	41	49	43	...	42	44.3	5.77

³ IPS textbook dataset, from a study conducted by Jim Baumann and Leah Jones at the Purdue University School of Education.

NOTATION FOR 1-WAY ANOVA

Data layout and notation:

group	observations (j)	number	mean	std.dev.
1	$X_{11} \quad X_{12} \quad \dots \quad X_{1n_1}$	n_1	\bar{X}_1	s_1
row 2	$X_{21} \quad X_{22} \quad \dots \quad \dots \quad X_{2n_2}$	n_2	\bar{X}_2	s_2
(i)	$\vdots \quad \vdots \quad \dots \quad \dots \quad \vdots$	\vdots	\vdots	\vdots
I	$X_{I1} \quad X_{I2} \quad \dots \quad X_{In_I}$	n_I	\bar{X}_I	s_I

- X_{ij} = j th observation in i th group (row), where
 - * $i = 1, \dots, I$, and I = number of groups (rows),
 - * $j = 1, \dots, n_I$, and n_i = number of obs. in i th group.
- denote also by $N = n_1 + \dots + n_I$ the total number of observations, and by $\bar{X} = \sum_{ij} X_{ij}/N$ the overall mean,
- the dataset/design is
 - * balanced, if all groups equally large, ($n_1 = \dots = n_I$),
 - * unbalanced, otherwise (some groups of different size),balanced designs are nice and simpler, but when using a computer unbalancedness is not a problem.

The natural model would seem to be (assuming normals),

$$\begin{array}{ccc}
 X_{11}, \dots, X_{1n_1} & \text{i.i.d. } N(\mu_1, \sigma_1), & \\
 \vdots & & \vdots \\
 X_{I1}, \dots, X_{In_I} & \text{i.i.d. } N(\mu_I, \sigma_I), &
 \end{array}$$

but we make the additional assumption: $\sigma_1 = \dots = \sigma_I$.

STATISTICAL MODEL

Model:

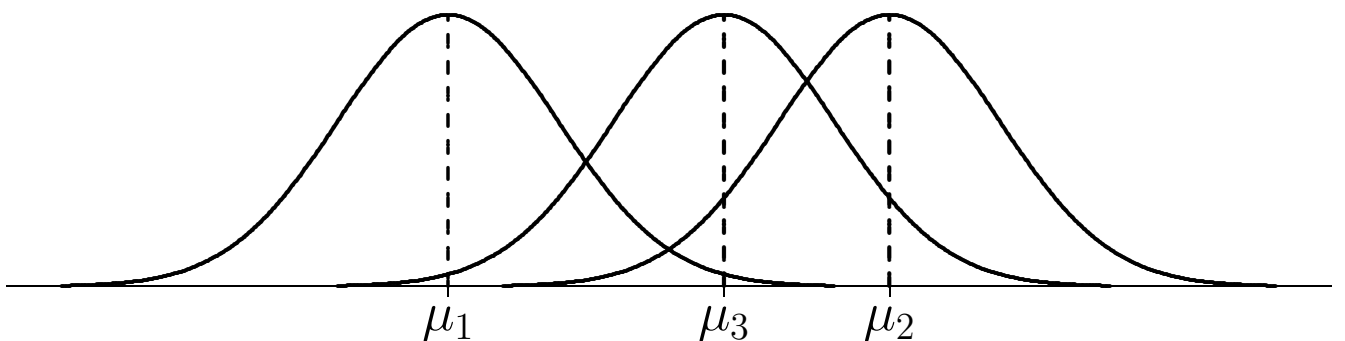
$$X_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, \dots, I; \quad j = 1, \dots, n_i,$$

where ε_{ij} 's are i.i.d. and $\sim N(0, \sigma)$.

i	observations			
1	$X_{11} = \mu_1 + \varepsilon_{11}$	$X_{12} = \mu_1 + \varepsilon_{12}$...	$X_{1n_1} = \mu_1 + \varepsilon_{1n_1}$
2	$X_{21} = \mu_2 + \varepsilon_{21}$	$X_{22} = \mu_2 + \varepsilon_{22}$...	$X_{2n_2} = \mu_2 + \varepsilon_{2n_2}$
\vdots	\vdots	\vdots	...	\vdots
I	$X_{I1} = \mu_I + \varepsilon_{I1}$	$X_{I2} = \mu_I + \varepsilon_{I2}$...	$X_{In_I} = \mu_I + \varepsilon_{In_I}$

- parameters: μ_1, \dots, μ_I (group means) and σ (common standard deviation of all X 's and ε 's),
- $\varepsilon_{ij} = X_{ij} - \mu_i$, i.e. the deviation of X_{ij} from its mean \Rightarrow ε 's interpreted as random errors / perturbations / noise,
- same model as on previous slide: $X_{ij} \sim N(\mu_i, \sigma)$.

Normal distributions for 3 groups:



ESTIMATION

Rules and formulae for estimation of model parameters:

- group sample means as estimates for μ 's:

$$\hat{\mu}_i = \bar{X}_i \sim N(\mu_i, \sigma/\sqrt{n_i}), \quad i = 1, \dots, I,$$

$$\text{SE}(\hat{\mu}_i) = s_p/\sqrt{n_i},$$

- pooled sample variance as estimate for σ^2 (a weighted average of the group variances s_i^2):

$$\hat{\sigma}^2 = s_p^2 = \sum_i \frac{n_i - 1}{N - I} s_i^2 = \sum_{ij} \frac{(X_{ij} - \bar{X}_i)^2}{N - I} = \text{SSE/DFE},$$

$$\hat{\sigma} = s_p = \sqrt{s_p^2},$$

where

* SSE = $\sum_{ij} (X_{ij} - \bar{X}_i)^2$, the within-group sum of squares,

* DFE = $N - I = (n_1 - 1) + \dots + (n_I - 1)$, the (within-group) degrees of freedom for s_p^2 .

Confidence intervals for μ_i the “usual way”, e.g.

$$(1 - \alpha) \text{ CI for } \mu_i : \hat{\mu}_i \pm t^* \text{SE}(\hat{\mu}_i), \quad t^* = t_{1-\alpha/2}(\text{DFE}),$$

note: using the pooled standard deviation s_p and its DF for the confidence intervals.

MODEL CHECKING

Summary of model assumptions:

- (1) all observations are independent,
- (2) all observations are normally distributed,
- (3) all observations have the same standard deviation (often called variance homogeneity or homoscedasticity),
- (4) all observations within a group have the same mean.

Useful graphical displays:

- boxplots/stemplots/dotplots for all groups in same diagram (overview of data, assumptions (2) and (3)),
- normal probability plots for each of the groups (2).

Useful statistics:

- standard descriptive statistics for each group (3),
- normality tests for each of the groups (2).

Practical considerations:

- textbook (PSLS/IPS) guideline for when to “accept” assumption about equal standard deviations: ratio of largest to smallest standard deviation less than 2,
- tests of equal stand. deviations (in software): more sensitive than the 1-way ANOVA itself, but fine if non-significant,
- if all groups show the same non-normal pattern (e.g., skewness), transformation may be a solution,
- if higher group means are associated with higher standard deviation, transformation (log or $\sqrt{\cdot}$) may be a solution.

HYPOTHESIS AND TEST

Consider the overall group hypothesis: $H_0: \mu_1 = \dots = \mu_I$:
(all groups equal, homogeneity between groups)

- alternative hypothesis H_a : some μ 's differ (“two-sided”),⁴
- test statistic calculated in several steps,
 - * define group sum of squares: $SSG = \sum_i n_i (\bar{X}_i - \bar{X})^2$
— a weighted sum of squared deviations between group means and the overall mean
 - * define group degrees of freedom: $DFG = I - 1$,
 - * introduce group mean square (ratio between sum of squares and DF): $MSG = SSG/DFG$,
 - * finally, test statistic is: $F = MSG/s_p^2$,
- some “motivations”:
 - * F compares *variation between groups* with *variation within groups*,
 - * nominator and denominator have similar forms:
 $F = MSG/MSE$, (because $s_p^2 = SSE/DFE = MSE$),
- under H_0 : F -statistic $\sim F(DFG, DFE)$,
and large values are critical for H_0 (one-sided test),
- P -value calculated as: $P = P(F(DFG, DFE) > F_{\text{obs}})$.

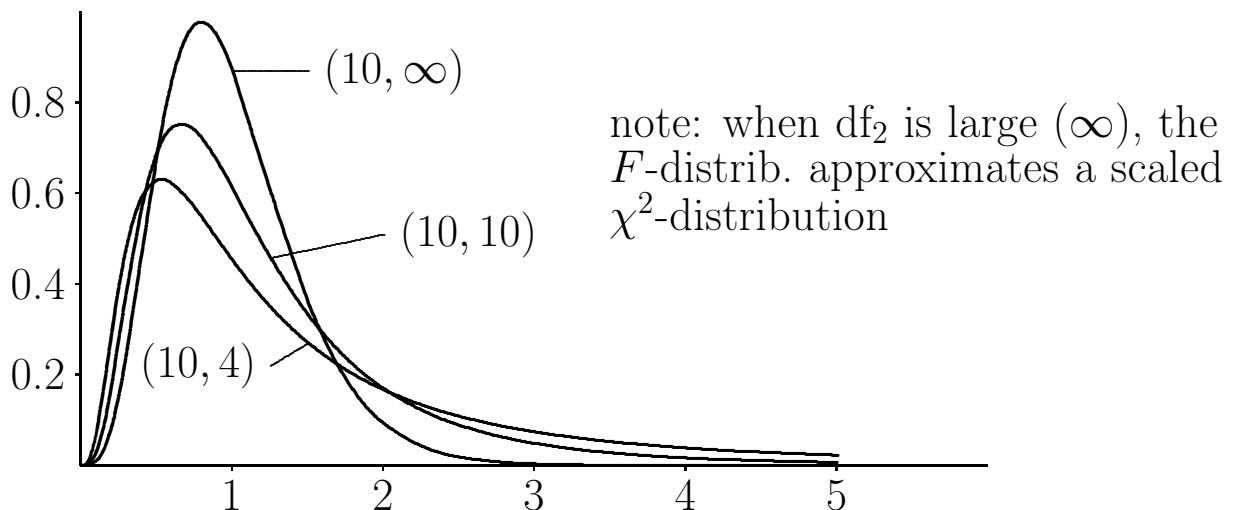
⁴ It is a common misunderstanding that H_a implies all the μ 's to be different, but the opposite of H_0 is just that the means are not all the same.

F-DISTRIBUTION

Another distribution — to be used for tests in normal models (that are more complex than one or two samples):

- F -distributions have two parameters, and we write $F(df_1, df_2)$ to indicate them:
 - * df_1, df_2 are numbers in $\{1, 2, 3, \dots\}$
 - * called *numerator* (df_1) and *denominator* (df_2) “degrees of freedom”, because F variables are usually ratios,
 - * given from the data, and not to be estimated,
 - * the order of df_1 and df_2 is important, because
$$F(df_1, df_2) \neq F(df_2, df_1),$$
- distributions on $(0, \infty)$ — only positive values,
- right skewed; decreasing mean and standard deviation with increasing df_2 ; percentiles in Table F/E of PSLS/IPS.

Some F distributions:



ANOVA TABLE

Analysis of variance table

= convenient layout for summarizing the analysis:

- idea: split variation in the data into parts:
total variation = variation between groups (“Groups”) + variation within groups (“Error”),
- collect quantities related to each source of variation on separate line,
- provides good overview and eases computations,
- generalizes to models with more variables than one.

General ANOVA table:

Source	Degrees of freedom	Sum of squares	Mean square	F	P
Groups	$DFG = I - 1$	$SSG = \sum_i n_i (\bar{X}_i - \bar{X})^2$	$MSG = SSG / DFG$	MSG / MSE	$P(F \geq F_{obs})$
Error	$DFE = N - I$	$SSE = \sum_{ij} (X_{ij} - \bar{X}_i)^2$	$MSE = SSE / DFE$		
Total	$DFT = N - 1$	$SST = \sum_{ij} (X_{ij} - \bar{X})^2$	$(MST = SST / DFT)$	$F \sim F(DFG, DFE)$	

Notes:

- MST often omitted from table (\neq MSG + MSE),
- always remember: $\hat{\sigma} = s_p = \sqrt{MSE}$.

EXERCISES 12.1, 12.9, 12.25

Exercise 12.1: (response, populations, I , n_i 's and N)

- (a) response = tomato yield, $I = 4$ varieties, all $n_i = 12$, and $N = 48$,
- (b) response = rate of attractiveness, $I = 5$ types of packaging, all $n_i = 40$, and $N = 200$,
- (c) response = weight loss, $I = 3$ programs, all $n_i = 20$, and $N = 60$.

Exercise 12.9: (degrees of freedom and hypotheses) For all settings, the hypotheses are:

$H_0: \mu_1 = \mu_2 = \dots = \mu_I$, H_a : some μ 's different, where $I=4, 5$ and 3 , respectively.

- (a) varieties (DF = 3), error (DF = 44), total (DF = 47), $F(3, 44)$,
- (b) packagings (DF = 4), error (DF = 195), total (DF = 199), $F(4, 195)$,
- (c) programs (DF = 2), error (DF = 57), total (DF = 59), $F(2, 57)$.

Exercise 12.25:

	Source	Degrees of freedom	Sum of squares	Mean square	F
(a)	Groups	3	104,855.87		
	Error	32	70,500.59		
	Total				

- (b) $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$, H_a : some μ 's different,
- (c) $F \sim F(3, 32)$ under H_0 , $P = P(F(3, 32) > 15.9) \ll 0.001 \Rightarrow$ reject H_0 : there are some differences between groups,
- (d) $s_p^2 = \text{MSE} = 2203$, $s_p = \sqrt{2203} = 46.9$.

CONTRASTS

Another type of null hypotheses:

- more specific than overall homogeneity of groups,
 - (1) involves two particular groups, e.g. $H_0: \mu_D = \mu_S$, or
 - (2) involves a linear combination of several group means, for example (Reading data): $\mu_B = (\mu_D + \mu_S)/2$,
- typically considered *after* test of overall H_0 ,
- decided prior to data collection and analysis!

Definition:

A *contrast* is a linear combination of group mean parameters of the form (in PLS notation),

$$L = \sum_i c_i \mu_i,$$

where the c_i 's are known constants with sum 0 ($\sum_i c_i = 0$).

Examples from above: (1) $c_D = 1$, $c_S = -1$, $c_B = 0$, and (2) $c_B = 1$, $c_D = -0.5$ and $c_S = -0.5$.

Statistical inference about contrasts:

- estimate: $\hat{L} = \sum_i c_i \bar{X}_i$ (sample contrast),
- standard error: $SE_{\hat{L}} = s_p \sqrt{\sum_i c_i^2 / n_i}$,
- (1- α) CI for L : $\hat{L} \pm t^* SE_{\hat{L}}$, $t^* = t_{1-\alpha/2}(\text{DFE})$,
- test of $H_0: L = 0$: $t = \hat{L} / SE_{\hat{L}} \sim t(\text{DFE})$ under H_0 .

PAIRWISE COMPARISONS

How to proceed after ANOVA and preplanned contrasts?

- if the overall H_0 is non-significant:
no further analysis needed (relevant), report $\hat{\mu}$ and P ,
- if the overall H_0 is significant:
 - * for illustration: plot of $\hat{\mu}_i$'s with error bars (e.g., Minitab),
 - * for informal comparison of group levels — LSD:
 - least significant **d**ifference(s),
 - margin of error of CI's for pairwise differences $\mu_i - \mu_j$ based on $\bar{X}_i - \bar{X}_j$:

$$\text{LSD}_{1-\alpha} = t^* s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}, \quad t^* = t_{1-\alpha/2}(\text{DFE}),$$
 - most useful for *balanced* data (all n_i 's equal), because one LSD-value applies to all comparisons,
 - example (reading data, `post3`):

$$\text{LSD}_{0.95} = t_{0.975}(63) \times 6.314 \times \sqrt{2/22} = 3.81,$$
 - interpretation: smallest distance between \bar{X}_i and \bar{X}_j so that CI for $\mu_i - \mu_j$ does not contain 0
($\Rightarrow H_0 : \mu_i = \mu_j$ signif. at level α , if preplanned),
 - * for formal comparison of group levels we must take into account an increase in overall (simultaneous) error level when doing many comparisons.

BONFERRONI METHOD FOR MULTIPLE TESTS

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

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

In particular, in the context of performing several tests,

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

Therefore, if we make k tests/comparisons, we can achieve the simultaneous error probability for all tests to be $\leq \alpha$, by taking the error probability for each test equal to α/k .

Adjustment of LSD method for k preplanned tests:

$$\text{LSD}_{1-\alpha/k} = t_{1-\alpha/(2k)}(\text{DFE}) s_p \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}.$$

Adjustment of LSD method for unplanned comparisons:

(suggested by the data)

- take $k = \text{total no. of comparisons} = \binom{I}{2} = I(I-1)/2$,
- use above LSD-formula with that value of k .

Notes for Bonferroni method:

- is *conservative* (larger CI's and higher P -values),
- is flexible and applicable to many situations (not only 1-way ANOVA).

Tukey method, described in Chapter 26 of PSLS:

- acceptable method, less conservative than Bonf.,
- method difficult to explain, and less flexible than B.

SUMMARY OF 1-WAY ANOVA FOR READING SCORES

Statistical models (for `pre1` and `post3` variables):

$$X_{ij} = \mu_i + \varepsilon_{ij},$$

where $i = 1, 2, 3$ (Basal, DRTA, Strat), $j = 1, \dots, 22$ (students), and ε_{ij} 's i.i.d. and $\sim N(0, \sigma)$.

ANOVA tables:

	pre1					post3			
Source	DF	SS	MS	F	P	SS	MS	F	P
Groups	2	20.58	10.29	1.13	0.33	357.3	178.7	4.48	0.015
Error	63	572.45	9.09			2511.7	39.9		
Total	65	593.03				2869.0			

Hypothesis $H_0: \mu_1 = \mu_2 = \mu_3$ (no differences), H_a : not H_0 ,

Test of H_0 : $\rightarrow F$ -tests in table:

not significant for `pre1` but significant for `pre3`,

Estimation/Presentation (using $t^* = t_{0.975}(63) = 2.00$):

statistic	pre1	post3		
	overall	Basal	DRTA	Strat
mean	9.79	41.05	46.73	44.27
SE(mean)	$s_p/\sqrt{66} = 0.37$	$s_p/\sqrt{22} = 1.35$		
95% CI	± 0.74	± 2.69		
LSD _{0.95}	—	3.81		

Conclusions: (`pre1`): no differences before teaching,
 (`post3`): no differences D/S or B/S; difference D/B.

SAMPLE SIZE FOR 1-WAY ANOVA

Based on desired precision (e.g., size of CI):

- CI could be for group mean, a difference between group means or a more general contrast,
- requires known (guessed) σ and desired margin of error (m),
- general approach based on approximate 95% CI for a mean (or contrast) parameter μ , based on its estimate $\hat{\mu}$:

$$\hat{\mu} \pm 2 \text{SE}(\hat{\mu})$$

— determine $\text{SE}(\hat{\mu})$ as a function of n , and solve with respect to n the equation: $m \geq 2 \text{SE}(\hat{\mu})$.

Based on power for F -test:

- requires known (guessed) σ and group means μ_i , plus significance level,
- calculation available in Minitab/Stata/R, and on web-sites.⁵

⁵ The recommended website for sample size calculations is <http://homepage.stat.uiowa.edu/~rlenth/Power/>.

KRUSKAL-WALLIS TEST

- Model: I independent samples from different distributions:

$$\begin{array}{ccc} X_{11}, \dots, X_{1n_1} & \text{i.i.d.} & \text{with distribution } \text{Dist}_1, \\ & \vdots & \\ X_{I1}, \dots, X_{In_I} & \text{i.i.d.} & \text{with distribution } \text{Dist}_I, \end{array}$$

- Hypotheses — two possibilities:
 - * H_0 : $\text{Dist}_1 = \dots = \text{Dist}_I$ (same distrib.), H_a : not H_0 ,⁶
 - * assuming “ $\text{Dist}_i = \text{Dist}_0 + \Delta_i$ ” (distributions all of the same shape, and differ only in positions Δ_i),
 H_0 : Δ_i 's=0 (corresponding to same medians) versus two-sided alternative H_a ,
- Test procedure:
 - * rank all observations as if a single sample, and compute rank averages \bar{R}_i for each group i ,
 - * test statistic: sum of squares for ranks,

$$H = \text{const} \times \sum_i n_i (\bar{R}_i - \bar{R})^2, \quad \bar{R} = (N+1)/2,$$
 \sim SSG in 1-way ANOVA for ranks,
 - * under H_0 : distribution of Y has no easy form, and software use different approximations for the P -value (based on the $\chi^2(I-1)$ -distribution).

⁶ More specific wording of H_a : for some (i, i') , P_i is systematically larger than $P_{i'}$; see Chapter 27 of PSLS.

NOTES ON HOME ASSIGNMENT II

Overall impression: good (all eight questions), but...
it is useful to answer all questions asked!

How much detail about models and methods is needed?

- it must be clear exactly which methods were used (without looking at the details of computer listings),
- it must be clear how assumptions were validated and any concerns must be discussed (it's not a valid option to not validate assumptions...),
- standard information to present include: relevant means/proportions, standard dev. (if relevant) and/or standard errors and/or CIs, test statistics from standard distributions and their DFs (if relevant), and P -values.

Some specifics:

- (Q1) the population consisted of volunteers \sim likely selection bias,
- (Q3) a blinded study is not necessarily randomized... ,
- (Q5) probability question, once a normal distribution is assumed; we don't need data, and if anything those for the caffeine group should be considered... ,
- (Q6) the probability computed is really the P -value for testing $H_0 : p = 0.025$ vs. $H_a : p > 0.025$ from observed $Y = 5$ when $Y \sim \text{Bin}(11, p)$; not needed for full score.

NOTES ON MID-TERM EXAM

- marks generally lower than for home assignments (this is “normal”): mean/median $\approx 10/15$, but with some very low marks,
- generally largest problem — choice of method/model:
 - * involves a judgement of what assumptions may be acceptable and how objectives can be met:
 - choice in **b)** between inference for mean (normal distrib.) or median (non-param.), both potentially valid,
 - * the two samples are clearly *independent*, and you should never use any methods involving differences (Minitab listings),
 - * choice between probability calc. and statistical analysis,
- some instances of ineffective exam technique:
 - * “forgetting” to provide what was asked for or to explain the model/method,
 - * recomputing statistics already in the Minitab print (**b,c**); unnecessary: the information is there to be used),
 - * not doing enough questions (running out of time),⁷
 - * writing loong text for the descriptive statistics when these were not asked for,

⇒ what can be learned:

 - * read very carefully, perhaps underline questions asked,
 - * distribute your time more equally on the questions,
- final exam: lots of time/opportunity to prepare yourself.

⁷ An adjustment in scores was made for the exam being quite lengthy for 1 hour.

SUMMARY NOTES

Key words and concepts for 1-way ANOVA:

- comparison of multiple samples (SRS or i.i.d.), assumed normally distributed with separate means but same variance,
- estimation: sample means, pooled variance/standard deviation,
- model checking: normality in each sample, equal standard deviation rule,
- hypothesis and test: F -statistic and F -distribution, ANOVA table: sum of squares (SS), degrees of freedom (DF), mean square (MS),
- after ANOVA: significant test is followed by pairwise comparisons (or contrasts, not in course syllabus), Bonferroni adjustment for multiple comparisons,
- sample size calculations for 1-way ANOVA,
- nonparametric 1-way ANOVA: Kruskal-Wallis test.