

## Index of 9-L: One-way analysis of variance (slide 17 added)

Page	Title
1	Practical information
2	Data example(s): Reading scores
3	Notation for one-way ANOVA
4	Statistical model
5	Estimation
6	Model checking
7	More on model assumptions
8	Hypothesis and test
9	<i>F</i> -distributions
10	ANOVA table
11	Exercises 12.1, 12.9, 12.25
12	Comparing and presenting groups
13	Pairwise comparisons
14	Bonferroni method for multiple tests
15	Summary of one-way ANOVAs for reading scores
16	Method for presenting group comparisons
17	More thoughts about multiple testing
18	Kruskal-Wallis test
19	Summary notes
20	Appendix: Contrasts

## PRACTICAL INFORMATION

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

- **main topic:** one-way ANOVA (analysis of variance),<sup>1</sup>
  - \* the step from two to multiple ( $\geq 2$ ) independent samples for quantitative data,
  - \* some similarity with 2-sample analysis, but also many new features,
- **contrasts** are not part of the course curriculum<sup>2</sup>,
- **additional topic:** Kruskal-Wallis test (PSLS Chapter 27)  
— a non-parametric one-way ANOVA.

Schedule news:

- **home assignment 2** returned on Monday (with questions. . . ),
- **home assignment 3** posted, due Thursday 19/3 (together with *optional project proposal*),
- we need to set a two-hour time slot for the oral exams (April 17?),
- now **jumping ahead** in textbooks:
  - \* we will come back to regression and correlation,
  - \* **skip over** (for now) textbook references to residuals and  $R^2$ .

---

<sup>1</sup> PSLs 4e: Chapters 24 & 26; S: Section 11.3 (too briefly); IPS 7e: Chapter 12.

<sup>2</sup> Contrasts are useful!: Yossa & Verdegem (2015), *Aquaculture* 437, 344–350; see media page.

## DATA EXAMPLE(S): READING SCORES

### Teaching reading comprehension<sup>3</sup>:

- 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:

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

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

## NOTATION FOR ONE-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	$X_{21} \quad X_{22} \quad \dots \quad \dots \quad X_{2n_2}$	$n_2$	$\bar{X}_2$	$s_2$
( $i$ )	$\vdots$	$\vdots \quad \vdots \quad \dots \quad \dots \quad \vdots$	$\vdots$	$\vdots$	$\vdots$
	$I$	$X_{I1} \quad X_{I2} \quad \dots \quad X_{In_I}$	$n_I$	$\bar{X}_I$	$s_I$

- $X_{ij}$  =  $j$ 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 observations 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 are equally large (i.e.,  $n_1 = \dots = n_I$ ), otherwise **unbalanced** (some groups of different size) — **balanced designs** are “nice”.<sup>4</sup>

The **natural model** would seem to be (assuming normal distributions),

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

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

<sup>4</sup> Making some calculations and formulas simpler, but when using a computer unbalancedness is not a problem.

## STATISTICAL MODEL

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

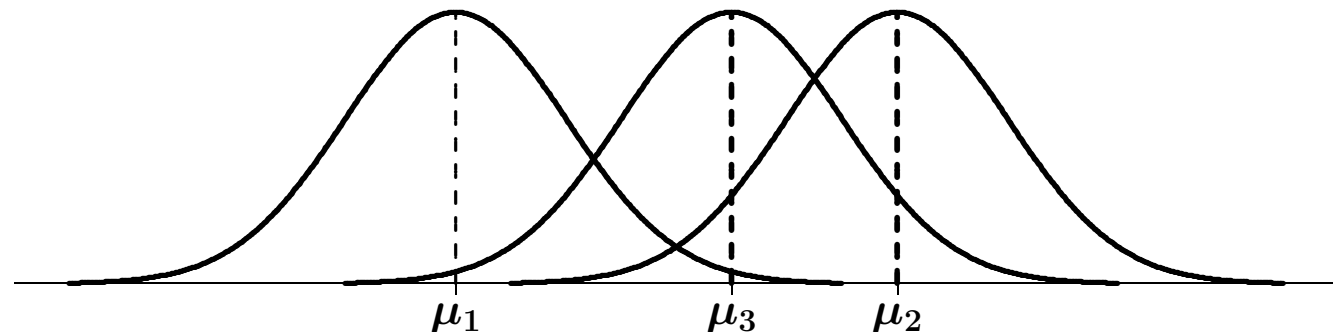
where  $\varepsilon_{ij}$ 's are i.i.d. and  $\sim N(0, \sigma)$ .

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

- **parameters:**  $\mu_1, \dots, \mu_I$  (group means) and  $\sigma$  (common standard deviation of all  $X$ 's and  $\varepsilon$ 's),
- $\varepsilon_{ij} = X_{ij} - \mu_i$ , i.e. **the deviation of  $X_{ij}$  from its mean**  $\Rightarrow \varepsilon$ 's are 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 formulas for estimation of model parameters:

- group sample means as estimates for  $\mu$ '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 estimate for  $\sigma^2$  (a weighted average of group variances  $s_i^2$ ):

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

where we have introduced the **new notation**:

- \* 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 degrees of freedom for  $s_p^2$ .

Confidence intervals for  $\mu_i$  computed 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), **not overall**.

**Extra tool** for model checking: **residuals** — to be introduced in the context of regression (next lecture), but **equally applicable** to ANOVA.

## MORE ON MODEL ASSUMPTIONS

Assumption of **equal standard deviations**:

- textbook (PSLS/IPS) **guideline** based on group standard deviations  $s_i$ :
  - \* okay, if **ratio of largest to smallest standard deviation** less than 2,
  - \* may be too prescriptive for small group sizes ( $n_i$ ) where the  $s_i$  are quite noisy,
- possible to **test** for equal standard deviations (in software):
  - \* **not** (necessarily) a good idea, because tests are more sensitive to unequal standard deviations and non-normality than the ANOVA itself  $\Rightarrow$  risk of stating a problem where there really is none. . . ,
  - \* however fine if such tests (Levene's test is preferable) are non-significant,

Dealing with **violations of assumptions**:

- if all groups show the **same non-normal pattern** (e.g., skewness), transformation may be a solution (e.g., if higher group means are associated with higher standard deviation, transformation with  $\log$  or  $\sqrt{\cdot}$  may work well),
- one-way ANOVA **without equal variances assumption**; Minitab offers:
  - \* adjusted overall  $F$ -test (from 9L–8),
  - \* another pairwise comparison procedure similar to Tukey method (9L–14),
- non-parametric analysis analysis (9L–17).

## HYPOTHESIS AND TEST

Overall group equality hypothesis:

$H_0: \mu_1 = \dots = \mu_I$  (all groups equal, homogeneity between groups),

- alternative hypothesis  $H_a$ : some  $\mu$ 's differ (“two-sided”),<sup>5</sup>
- 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<sup>6</sup>:  $MSG = SSG/DFG$ ,
  - \* finally, the 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:  $F = MSG/MSE$ ,<sup>7</sup>
- under  $H_0$ :  $F$ -statistic  $\sim F(DFG, DFE)$ , and large values are critical for  $H_0 \Rightarrow$   $P$ -value calculated as:  $P = P(F(DFG, DFE) > F_{obs})$ .

<sup>5</sup> It is a common misunderstanding that  $H_a$  implies all the  $\mu$ 's to be different, but the opposite of  $H_0$  is just that the means are not all the same.

<sup>6</sup> Generally, a mean square is a sum of squares divided by the corresponding degrees of freedom.

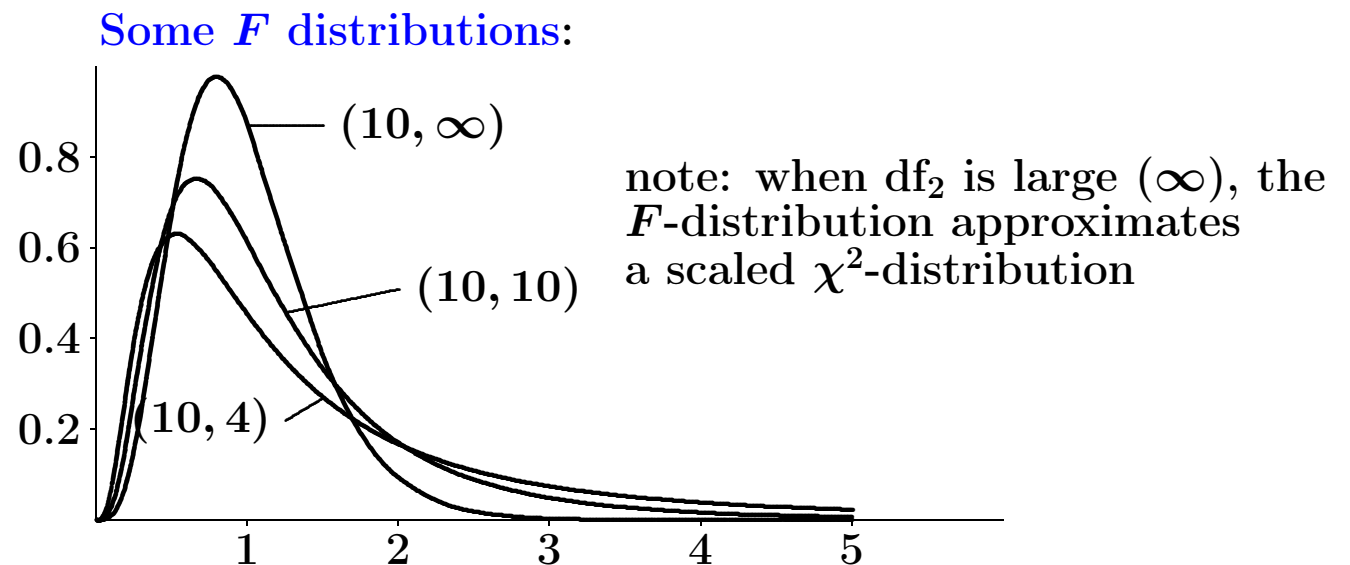
<sup>7</sup> We previously (slide 9L-5) introduced notation to make  $s_p^2 = SSE/DFE = MSE$ .

## $F$ -DISTRIBUTIONS

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:  $F(df_1, df_2) \neq F(df_2, df_1)$ ,
- distributions on  $(0, \infty)$   
— only positive values,

- right skewed;  
decreasing mean  
and standard dev.  
with increasing  $df_2$ ;  
Table F/E of  
of PLS/IPS  
has percentiles.



## ANOVA TABLE

**Analysis of variance table** = convenient layout for summarizing the analysis:

- idea: split variation in the data into parts:

$$\begin{aligned} \text{total variation} &= \text{variation between groups ("Groups")} \\ &+ \text{variation within groups ("Error"),} \end{aligned}$$

- collect quantities related to each source of variation on a separate line,
- provides **good overview** and eases computations (if done by hand),
- generalizes** to models with more variables than one.

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

**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:

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ ,  $H_a : \text{some } \mu\text{'s different}$ ,  
 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				

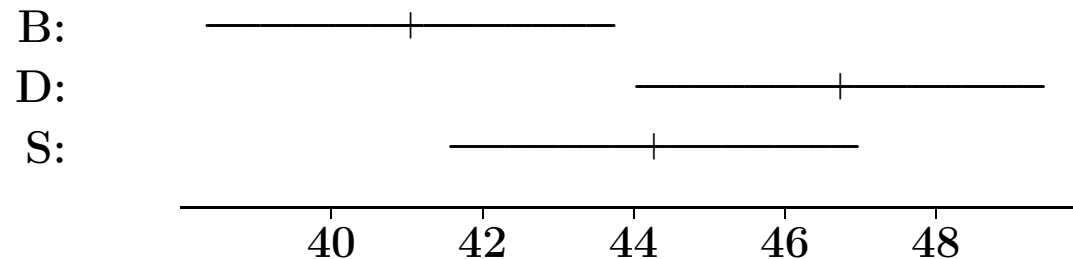
- (b)  $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ ,  $H_a : \text{some } \mu\text{'s different}$ ,
- (c)  $F \sim F(3, 32)$  under  $H_0$ ,  $P = P(F(3, 32) > 15.9) \ll 0.001 \Rightarrow \text{reject } H_0$ ;  
**conclusion:** there are some differences between groups,
- (d)  $s_p^2 = \text{MSE} = 2203$ ,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

## COMPARING AND PRESENTING GROUPS

Group comparisons based on confidence intervals for group means  $\mu_i$ :

- principal method uses test/CI for difference between parameters (next slide),
- but conclusions available from group CIs in 2/3 cases (see figure):

**example:** reading data — 95% CIs for post3:



- \* **B vs. D:** disjoint (non-overlapping) CIs  $\Rightarrow$  significance ( $P < 0.05$ ),
  - \* **D vs. S:** estimate inside another CI  $\Rightarrow$  no significance ( $P > 0.05$ ),
  - \* **B vs. S:** need CI for difference ( $\mu_B - \mu_S$ ) to assess significance,
- method assumes **independent estimates**, and is unadjusted for multiple testing (following slides),
  - method is **applicable** to many other settings than one-way ANOVA.

## PAIRWISE COMPARISONS

Generally, **how to proceed** after ANOVA? (and any pre-planned contrasts (appendix))

If **overall  $H_0$**  is **non-significant**: no further analysis needed (relevant), report  $\hat{\mu}$  and  $P$ .

If **overall  $H_0$**  is **significant**:

- for **illustration**: plot of  $\hat{\mu}_i$ 's with error bars (previous page),
- for **informal comparisons** of group levels — LSD (least significant difference):

- \* margin of error of CIs for pairwise difference  $\mu_i - \mu_j$  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  $n_i$  are equal), because one LSD-value applies to comparisons between any two groups,
- \* **example** (reading data, post3):

$$\text{LSD}_{0.95} = t_{0.975}(63) \times 6.314 \times \sqrt{2/22} = 3.80,$$

- \* **interpretation**: smallest distance between  $\bar{X}_i$  and  $\bar{X}_j$  so that CI for  $\mu_i - \mu_j$  does not contain 0 ( $\Rightarrow H_0 : \mu_i = \mu_j$  significant at level  $\alpha$ , if pre-planned<sup>8</sup>),
- \* same as pairwise 2-sample  $t$ -tests based on  $s_p$  (also called **Fisher** method),
- for **formal comparisons** of group levels we must take into account an increase in overall (simultaneous) error level for multiple unplanned<sup>8</sup> comparisons.

---

<sup>8</sup> The theory behind statistical hypothesis testing is based on pre-determined hypotheses, and **does not apply** to hypotheses suggested by the data!

## BONFERRONI METHOD FOR MULTIPLE TESTS

**Basis:** 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,

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

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 prob. **for each test** equal to  $\alpha/k$ .

**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  $k = \text{total number of comparisons} = \text{“}I \text{ choose } 2\text{”} = I(I-1)/2$ ,
- use above LSD-formula with that value of  $k$ .

**Notes for Bonferroni method:**

- is **conservative** (wider CIs and higher  $P$ -values),
- is flexible and applicable to many situations (not only one-way ANOVA).

Alternative **Tukey method** for multiple comparisons, described in Chapter 26 of PSLS:

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

## SUMMARY OF ONE-WAY ANOVAS FOR READING SCORES

Statistical models (for pre1 and post3 variables):

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

where  $i = 1, 2, 3$  (B, D, S),  $j = 1, \dots, 22$  (students), and  $\varepsilon_{ij}$ 's i.i.d. and  $\sim N(0, \sigma)$ .

ANOVA tables:

		pre1				post3			
Source	DF	SS	MS	<i>F</i>	<i>P</i>	SS	MS	<i>F</i>	<i>P</i>
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  $H_0 \rightarrow$  table *F*-tests: 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
std.dev. $s_p$	$\sqrt{9.09} = 3.01$	$\sqrt{39.9} = 6.31$		
SE(mean)	$s_p/\sqrt{66} = 0.37$	$s_p/\sqrt{22} = 1.35$		
95% CI	$\pm 0.74$	$\pm 2.69$		
LSD <sub>0.95</sub>	—	3.80		

**Conclusions:** (pre1): no differences before teaching,

(post3): no differences D/S or B/S; difference D/B.

## METHOD FOR PRESENTING GROUP COMPARISONS

**Issue:** pairwise comparison results are not easy to present compactly, because the number of comparisons is larger than number of groups ( $I$ ), and growing fast with  $I$  (slide 9L–14).

**Idea:** indicate significance between groups by codes displayed in list of groups (e.g., table of group estimates or CIs).

Most **common system**<sup>9</sup> uses letter codes  $a, b, c \dots$ , so that:

- groups with the same letter are **not** significantly different,
- for **manual construction** of codes, proceed as follows:
  - \* **order** group means from lowest to highest,
  - \* **designate**  $a$  to highest group + all groups not significantly different from it,
  - \* **designate**  $b$  to next group in the same way (but drop if same pattern as for  $a$ ),
  - \* **continue** through all groups,
- **example:** reading data with uncorrected 5% error coding:

$$B^b \quad S^{ab} \quad D^a ,$$

⇒ only significant difference is between B and D (as previously shown).

---

<sup>9</sup> Available in both Minitab and Stata.

## MORE THOUGHTS ABOUT MULTIPLE TESTING

Our approach to multiple testing<sup>10</sup> should reflect the **objectives** of the study and the testing, e.g. confirmatory<sup>11</sup> or exploratory<sup>12</sup>.

Any control for multiple testing to reduce the overall risk of false significances (**type I error**) will also increase the risk of not detecting any truly significant differences (**type II error**) — a trade-off.

**No adjustment** for multiple testing:

- higher overall significance level → conclusions are less certain/strict (which should be reflected in the wording used), e.g. ~ exploratory testing,
- can be argued, that it is “unfair” that each test gets lower power (i.e., higher type II error risk) just because the study involved many tests.

**Full adjustment** (e.g., all pairwise comparisons) for multiple testing:

- simplest approach for interpretations, but risks lowering the type I error level (and raising the type II error level) too much when many tests are involved,
- Tukey method is acceptable, Bonferroni method as well (but conservative ⇒ lower type I error).

**Partial adjustment** for multiple testing, according to prioritization of tests:

- requires explicit statement of pre-study priorities, e.g. ~ confirmatory testing, and careful statement of results/conclusions,
- adjustment for less tests ⇒ lower type II error risk,
- Bonferroni method can be used to adjust for selected tests.

---

<sup>10</sup> See also decision diagram at VHM 801’s media page.

<sup>11</sup> A confirmatory test seeks to prove or disprove (in terms of the statistical evidence obtained) a pre-defined hypothesis.

<sup>12</sup> An exploratory test is used to quantify the strength of a pattern observed in the data, whether pre-defined or not, without aiming for a decision at a controlled significance level.

## KRUSKAL-WALLIS TEST

- **Model:**  $I$  independent samples from different distributions:

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

- **Hypotheses** — two possibilities:

(1)  $H_0$ :  $\text{Dist}_1 = \dots = \text{Dist}_I$  (same distribution),  $H_a$ : not  $H_0$ ,<sup>13</sup>

(2) **assuming** “ $\text{Dist}_i = \text{Dist}_0 + \Delta_i$ ” (distributions differ only in positions  $\Delta_i$ )<sup>14</sup>,

$H_0$ :  $\Delta_i$ '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,

$$H = \text{constant} \times \sum_i n_i (\bar{R}_i - \bar{R})^2, \quad \bar{R} = (N+1)/2,$$

corresponding to SSG in a one-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).

<sup>13</sup> More specific wording of  $H_a$ : for some  $(i, i')$ ,  $P_i$  is systematically larger than  $P_{i'}$ ; see Chapter 27 of PSLS.

<sup>14</sup> The same assumption as for the Mann-Whitney-Wilcoxon test, implying all distributions to have the **same shape**.

## SUMMARY NOTES

### Key words and concepts for one-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, calculations organized in ANOVA table with:
  - \* 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),
  - \* pre-planned or unplanned comparisons,
  - \* Bonferroni adjustment for multiple comparisons,
- **nonparametric** rank-based one-way ANOVA: Kruskal-Wallis test.

## APPENDIX: CONTRASTS

Another type of null hypothesis:

- more specific than overall homogeneity of groups (illustrated by the reading data):
  - (1) involves two particular groups, e.g.  $H_0: \mu_D = \mu_S$ , or
  - (2) involves a linear combination of several group means, e.g.:  $\mu_B = (\mu_D + \mu_S)/2$ ,
- often (not necessarily) considered **after** test of overall  $H_0$ ,
- ideally decided prior to data collection/analysis (otherwise different methods needed).

**Definition:** A **contrast** is a linear combination of group mean parameters of the form,

$$L = \sum_i c_i \mu_i, \quad (\text{PSLS notation})$$

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):  $c_B = 1, c_D = -0.5, c_S = -0.5$ .

**Statistical inference** for contrasts:

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