Chapter 5: Comparisons among several samples, One-way ANOVA

Class 8: 2/25/09 W

Slide 1 Chapter 5: Comparisons among several samples, One-way ANOVA

NOTES:

HW 6 due Monday 3/1/09 9:50

Submit as Myname-HW6.doc (or *.rtf)

- Read Chapter 5 **Comparisons among several samples**
- Comment on Chapter 5 conceptual problems in Blackboard Vista4
- Computation Problem 6
  - Problem 4.30 Sunlight protection factor

EEOS611

Slide 2 HW 6 due Monday 3/1/09 9:50

NOTES:

HW 7 due Thursday 3/4/09 Noon

Submit as Myname-HW7.doc (or *.rtf)

- Read Chapter 6 **Comparisons among several samples**
- Comment on Chapter 6 conceptual problems in Blackboard Vista4
- Computation Problem 7
  - Problem 5.25 Duodenal ulcers

EEOS611

Slide 3 HW 7 due Thursday 3/4/09 Noon

NOTES:

Student Presentations

Starting at 10:50 (8 minutes each)

- Keith Cialino for HW 2
  - Ex 1.21
- Seth Sheldon for HW 3
  - 2.21 Bumpus's data: weights of Bumpus's birds
- Barry Fradkin for HW 4.
  - **3.28 Pollen removal**

EEOS611

Slide 4 Student Presentations

NOTES:

Chapter 4 (End)

Rank-based, nonparametric analogues to the paired t test: Signed rank & Sign tests

Wilcoxon's sign rank test
Frank Wilcoxon of American Cyanamide [See Salsburg, 2001, The Lady Tasting Tea, for biographical sketch]

Asymptotic relative efficiency >0.864, 95.5% for normally distributed data

Fisher's sign test
[See Salsburg, 2001 for biographical sketch]

Asymptotic relative efficiency 63.7% for normally distributed data

EEOS611

Slide 5 Chapter 4 (End)

NOTES:

4.4 Alternatives to the paired t test

Wilcoxon sign-rank and Fisher Sign tests

Slide 6 4.4 Alternatives to the paired t test

NOTES:

Anatomical abnormalities & schizophrenia

Case 2.2 (Sleuth p 30): 15 pairs of twins, paired *t* test

Display 2.2

Differences in volumes (cm^3) of left hippocampus in fifteen sets of monozygotic twins where one twin is affected by schizophrenia

Pair #	Unaffected	Affected	Difference
1	1.94	1.27	0.67
2	1.44	1.63	-0.19
3	1.56	1.47	0.09
4	1.58	1.39	0.19
5	2.06	1.93	0.13
6	1.66	1.26	0.40
7	1.75	1.71	0.04
8	1.77	1.67	0.10
9	1.78	1.28	0.50
10	1.92	1.85	0.07
11	1.25	1.02	0.23
12	1.93	1.34	0.59
13	2.04	2.02	0.02
14	1.62	1.59	0.03
15	2.08	1.97	0.11

Differences

-2	
-1	9
-0	
0	23479
1	0139
2	3
3	
4	0
5	09
6	7
7	

Average: 0.199
Sample SD: 0.238
n: 15

Legend: | 6 | 7 represents 0.67 cm^3

Slide 7 Anatomical abnormalities & schizophrenia

NOTES:

Case 2.2 Statistical Summary

Sleuth, p. 31

There is substantial evidence that the mean difference in the left hippocampus volumes between schizophrenic individuals and their nonschizophrenic twins is nonzero (two-sided *p-value = 0.006, from a paired t test). It is estimated that the mean volume is 0.20 cm^3 smaller for those with schizophrenia (about 11% smaller). A 95% confidence interval for the difference is from 0.07 to 0.33 cm^3.*

Slide 8 Case 2.2 Statistical Summary

NOTES:

Wilcoxon signed rank test

Display 4.12

Signed-rank test statistic computations; schizophrenia study

Pair	Unaffected	Affected	Difference	Ordered Magnitude	Order	Rank	+Ranks	-Ranks
1	1.94	1.27	.67	.02 (+)	1	1	1	
2	1.44	1.63	-.19	.03 (+)	2	2	2	
3	1.56	1.47	.09	.04 (+)	3	3	3	
4	1.58	1.39	.19	.07 (+)	4	4	4	
5	2.06	1.93	.13	.09 (+)	5	5	5	
6	1.66	1.26	.40	.10 (+)	6	6	6	
7	1.75	1.71	.04	.11 (+)	7	7	7	
8	1.77	1.67	.10	.13 (+)	8	8	8	
9	1.78	1.28	.50	.19 (+)	9	9.5	9.5	
10	1.92	1.85	.07	.19 (-)	10	9.5		9.5
11	1.25	1.02	.23	.23 (+)	11	11	11	
12	1.93	1.34	.59	.40 (+)	12	12	12	
13	2.04	2.02	.02	.50 (+)	13	13	13	
14	1.62	1.59	.03	.59 (+)	14	14	14	
15	2.08	1.97	.11	.67 (+)	15	15	15	
							= 110.5	

SPSS d scards pa rs w th equal values*

③ Correct the standard deviation SD(T), based on the pattern of ties [2, 5]

① Order the absolute differences and assign ranks to them

② Signed rank statistics = sum of ranks for positive differences

Slide 9 Wilcoxon signed rank test

NOTES:

Ties in signed rank tests

Two sorts of ties in the signed rank test, 1 of 2

1) If you have identical values in in both pairs, Wilcoxon recommended that those paired observations be dropped. That is still the standard recommendation, and SPSS uses this recommendation.

Hollander and Wolfe's Nonparametric statistics, 2nd ed (p. 46) covers the problem of dropping ties of the first sort. If there are many ties, H & W recommend using another test. They also state that you could leave the tied samples in, and use a random number generator to randomly assign positive or negative signs for the zeros. If you want a more conservative 1-sided test, assign all of the tied differences to the group that would make it less likely to reject the null. For example, if you are testing lipitor's effects on cholesterol and a patient had identical cholesterol levels before and after, then assign that difference as if the lipitor blood sample had the higher cholesterol. If you still reject the null, your conservative test would be less likely to result in a Type I error, but of course the probability of Type II error (failing to reject a false null). Pratt (1959), cited in both Lehmann and Hollander & Wolfe, provides a more thorough review of ties. Lehmann cites more recent papers on dealing with the 1st sort of ties in signed rank tests.

Slide 10 Ties in signed rank tests

NOTES:

Dealing with tied pairs

Two sorts of ties in the signed rank test, 2 of 2

The second sort of ties occurs after the absolute values of the differences are ranked

2) Ties may result after the absolute values of the differences between paired observations are ranked. Two or more differences may have the same absolute value. Those ties are not discarded, and the variance formula is adjusted to take into account the number of tied groups [See next slide]

Slide 11 Dealing with tied pairs

NOTES:

SPSS algorithms, signed rank test

There are exact tests, if no tied ranks

$$Z = \frac{\min(S_p, S_n) - (n(n+1)/4)}{\sqrt{n(n+1)(2n+1)/24 - \sum_{j=1}^{l} (t_j^3 - t_j)/48}}$$

L=tied groups
t_j = items in each tied group

Asymptotic relative efficiency>0.864, 95.5% for normally distributed data

where

n	Number of cases with non-zero differences
l	Number of ties
t_j	Number of elements in the j-th tie, $j = 1,\ldots,l$

EEOS611

Slide 12 SPSS algorithms, signed rank test

NOTES:

Fisher's sign test

Straightforward application of the 1-sample binomial test

- Given that the probability of a + sign = probability of a minus sign = 0.5,
- What is the probability of observing exactly k positive signs in *n* Bernoulli (binomial) trials?
  - $P(X=k) = n \text{ Choose } k * p^k (1-p)^{n-k}$
    - X has a binomial distribution
    - Must sum probability for observed value of k, and all more extreme values of k.
- Statistical sleuth provides only the normal approximation to the binomial, but SPSS will provide the exact test for n<30.

Frequencies

		N
AFFECTED - UNAFFECT	Negative Differences[a]	14
	Positive Differences[b]	1
	Ties[c]	0
	Total	15

a. AFFECTED < UNAFFECT
b. AFFECTED > UNAFFECT
c. AFFECTED = UNAFFECT

Test Statistics[b]

	AFFECTED - UNAFFECT
Exact Sig. (2-tailed)	.0010[a]

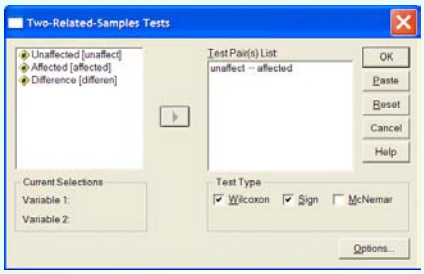
a. Binomial distribution used.
b. Sign Test

EEOS611

Slide 13 Fisher's sign test

NOTES:

Sign test in SPSS



EEOS611

Slide 14 Sign test in SPSS

NOTES:

Schizophrenia data

Sleuth (Ch 4.4.1 p. 99) Sign test: Sleuth p. 99
Sleuth presents the large sample approximation; exact test possible with binomial distribution (used by SPSS)

Display 2.2 p. 29

Differences in volumes (cm^3) of left hippocampus in fifteen sets of monozygotic twins where one twin is affected by schizophrenia

Pair #	Unaffected	Affected	Difference
1	1.94	1.27	0.67
2	1.44	1.63	-0.19
3	1.56	1.47	0.09
4	1.58	1.39	0.19
5	2.06	1.93	0.13
6	1.66	1.26	0.40
7	1.75	1.71	0.04
8	1.77	1.67	0.10
9	1.78	1.28	0.50
10	1.92	1.85	0.07
11	1.25	1.02	0.23
12	1.93	1.34	0.59
13	2.04	2.02	0.02
14	1.62	1.59	0.03
15	2.08	1.97	0.11

Differences: Average: 0.199, Sample SD: 0.238, n: 15

-2 | 1
-1 | 9
-0 |
0 | 23479
1 | 0139
2 | 3
3 |
4 | 0
5 | 09
6 | 7
7 |

Legend: | 6 | 7 represents 0.67 cm^3

Frequencies

		N
AFFECTED - UNAFFECT	Negative Differences[a]	14
	Positive Differences[b]	1
	Ties[c]	0
	Total	15

a. AFFECTED < UNAFFECT
b. AFFECTED > UNAFFECT
c. AFFECTED = UNAFFECT

Test Statistics[b]

	AFFECTED - UNAFFECT
Exact Sig. (2-tailed)	.0010[a]

a. Binomial distribution used.
b. Sign Test

Paired *t* test for equal difference, two-tailed p=0.006 [See sleuth p. 31]

Slide 15 Schizophrenia data

NOTES:

Assumptions of the Wilcoxon signed rank test

- Underlying distribution of the differences are continuous and symmetric about zero [Hollander & Wolfe p. 43]
- Differences within & between pairs on an ordinal scale
- Siegel: use large sample approximation if number of pairs exceeds 25, otherwise use tabulated values of the Wilcoxon signed rank statistic
  - Note that there are exact tests for Wilcoxon signed rank tests if there are ties (described in Hollander & Wolfe p 46-47)

Slide 16 Assumptions of the Wilcoxon signed rank test

NOTES:

Conclusions (1 of 2)

Chapter 4 Alternatives to the t tools

- Consider using alternatives to the t tools if
  - The assumptions are grossly violated or
  - The sample sizes are too small to test distributional assumptions
- Wilcoxon rank sum test
  - Appropriate for small sample sizes, but use the exact tests not the normal approximation
  - Appropriate in the presence of outliers
  - Ties are not a problem if the ties-correction used
  - Not appropriate for samples with unequal variances (try Fligner-Policello only if the sample sizes are large)

EEOS611

Slide 17 Conclusions (1 of 2)

NOTES:

Conclusions (2 of 2)

Chapter 4 Alternatives to the t tools

- Permutation test
  - Appropriate for small sample sizes, when the Student's *t* distribution might not be appropriate
  - Does not protect against the problem of unequal variances (the Fisher-Behrens problem)
    - Note that the solution to Case Study 4.1 is based on the equal variance t test.
- Paired data: tests based on ranks
  - Wilcoxon signed rank test: high power efficiency
  - Sign test, simple application of the 1-sample binomial test

EEOS611

Slide 18 Conclusions (2 of 2)

NOTES:

Chapter 5: Comparisons among several samples

Slide 19 Chapter 5: Comparisons among several samples

NOTES:

ANOVA, Analysis of Variance, the foundation of experimental design

- Most experimental design is based on an ANOVA framework
- One can't really appreciate the need for proper replication without considering the implications for testing treatment effects with ANOVA
  - Hurlbert's (1984) monograph criticizing statistics in ecological papers is largely a criticism of inappropriate ANOVA design
  - Hurlbert's pseudoreplication is Underwood's 'model misspecification' and both are largely based on using an inappropriate ANOVA model
- While ANOVA is a proper subset of the general linear model (GLM) and regression, as we'll see, the concepts involving design and partitioning degrees of freedom are more evident in ANOVA models

R.A. Fisher, inventor of ANOVA

Slide 20 ANOVA, Analysis of Variance, the foundation of experimental design

NOTES:

Case 5.1 Diet restriction & longevity

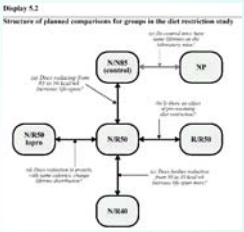
Slide 21 Case 5.1 Diet restriction & longevity

NOTES:

Planned comparisons

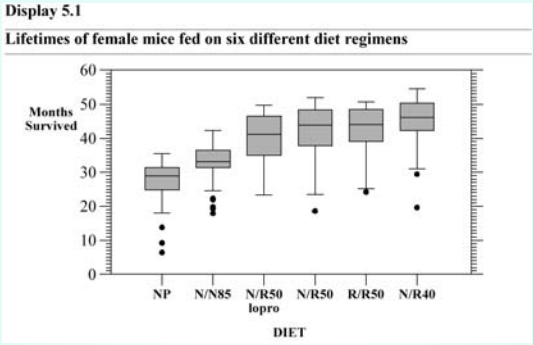
These are *a priori* contrasts, not *a posteriori*

- If hypotheses are set in advance, then you can test at a pre-set alpha level, without *a posteriori* (or post hoc, multiple comparison) adjustment
  - Recall that alpha = P(Type I error)
  - See Cook & Farewell (1996, J. Roy. Stat. Assoc. A). In dose-response studies, no need to adjust for number of dose treatments.
- One large design allows the use of a more precise estimate of the error variance
  - Separate control *vs.* treatment t tests are not powerful
  - If interaction effects are evident, separate tests can be misleading. They can miss interaction effects.



Slide 22 Planned comparisons

NOTES:



Slide 23

NOTES:

Detonator plots

Loughin Kansas State Dept. Statistics

EEOS611

Slide 24 Detonator plots

NOTES:

Summary statistics for lifetimes of mice

Display 5.2, Sleuth 2nd edition page 115

Summary statistics for lifetimes of mice on six different diet regimens

Group	n	Range (mo)	Average	SD	95% CI for Mean
NP	49	6.4 - 35.5	27.4	6.1	25.6 - 29.2
N/N85	57	17.9 - 42.3	32.7	5.1	31.3 - 34.1
N/R50	71	18.6 - 51.9	42.3	7.8	40.5 - 44.1
R/R50	56	24.2 - 50.7	42.9	6.7	41.1 - 44.7
N/R50 lopro	56	23.4 - 49.7	39.7	7.0	37.8 - 41.6
N/R40	60	19.6 - 54.6	45.1	6.7	43.4 - 46.8

EEOS611

Slide 25 Summary statistics for lifetimes of mice

NOTES:

Summary of statistical findings

Case Study 5.1: mouse longevity, 1 of 2

- There is overwhelming evidence that mean lifetimes in the six groups are different (p-value < 0.001); analysis of variance F-test).
- Analysis of the 5 particular questions are
  - **(1)** There is convincing evidence that lifetime increases as a result of restricting the diet from 85 kcal/wk to 50 kcal/wk (**1-sided p-value** < 0.0001; *t* test)
  - **(2)** There is no evidence that reducing the calories before weaning increased lifetime, when the calorie intake after weaning is 50 kcal/wk (1-sided p value = 0.32, *t* test). A 95% CI for the amount by which the lifetime under the R/R50 diet exceeds the lifetime under the N/R50 diet is -1.7 to 2.9 months.

EEOS611

Slide 26 Summary of statistical findings

NOTES:

Summary of statistical findings

Case Study 5.1: mouse longevity, 2 of 2

- Analysis of the 5 particular questions (continued)
  - **(3)** Further restriction of the diet from 50 to 40 kcal/wk increases lifetime by an estimated 2.8 months (95% CI: 0.5 to 5.1 months). The evidence that this effect is greater than zero is moderate (p=0.017, t test)
  - **(4)** There was moderate evidence that lifetime was decreased by the lowering of protein in addition to the 50 kcal/wk diet (2-sided p value =0.024; t-test)
  - **(5)**There is convincing evidence that the control mice live longer than the mice on the non-purified diet (1-sided p-value <0.0001)

EEOS611

Slide 27 Summary of statistical findings

NOTES:

Case study 5.2: The Spock trial

Slide 28 Case study 5.2: The Spock trial

NOTES:

Case 5.2 The Spock trial

Sleuth, page 117: Dr. Spock's *venire* contained only 1 woman, who was released by the prosecution

Display 5.4

Percents of women in 30-juror venires for Boston area U.S. District Court trials, grouped according to the judge presiding

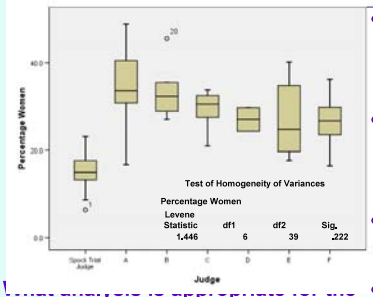
	Spock Trial Judge	Other Boston Area U.S. District Court Judges A	B	C	D	E	F
0	64,87						
1	33,36,50,52,77,86	68				77,97	65
2	31		70,89	10,34,75,75	43,97	15,79	07,35,64,67,95,98
3		08,36	20,27,55	05,19,25,38,38		48	19,62
4		05,89	56			02	

Legend: 4|89 represents a venire with 48.9% women

1) Is there evidence that women were underrepresented on the Spock judge's venires, and 2) Is there evide nce that there are differences in women's representation on the other juries?

Slide 29 Case 5.2 The Spock trial

NOTES:



- The percentage of women on the Spock judge's *venires* were substantially lower than the other judges (t test of Spock judge vs. 'Other judges')
- There is little evidence to reject the null hypothesis of no difference in female representation among the other judges p=0.32 (1-way ANOVA)
- The percentage of women is 15% less on the Spock judge's venires (95% CI: 10% to 20%)
- Gallagher note: *this pooling of judges could be called pseudoreplication, but can be justified as a fixed-effect NESTED ANOVA*

judge effect? There are no true replicates for the Spock-judge effect.

Slide 30

NOTES:

5.2 Comparing any two of several means

5.2.1 An ideal model for several-sample comparisons

Gallagher note: Comparisons among means in ANOVA can be analyzed using t statistics, with a new, more precise estimate of pooled error. It is that pooling, with higher df, that makes ANOVA a more powerful method than multiple t tests.

Slide 31 5.2 Comparing any two of several means

NOTES:

5.2.2 The pooled estimate of the standard deviation, s_p

Display 5.6, Sleuth page 120

Pooled estimate of standard deviation; diet restriction data

Group	n	Sample SD
NP	49	6.1
N/N85	57	5.1
N/R50	71	7.8
R/R50	56	6.7
N/R50 lopro	56	7.0
N/R40	60	6.7

s_p assumes equal variances among groups

1 Calculate the pooled estimate of variance, s_p^2

$$s_p^2 = \frac{(49-1)(6.1)^2+(57-1)(5.1)^2+(71-1)(7.8)^2+(56-1)(6.7)^2+(56-1)(7.0)^2+(60-1)(6.7)^2}{(49-1)+(57-1)+(71-1)+(56-1)+(56-1)+(60-1)}$$

$$= \frac{15,313.90}{343} = 44.647; \quad s_p = \sqrt{44.647} = 6.68$$

2 s_p is the square root

df = 343

3 df is the denominator

Slide 32 5.2.2 The pooled estimate of the standard deviation, sp

NOTES:

Pooled sd (s_p) in *t*-tests & ANOVA

s_p In *t* tests covered In Chapter 3

$$s_p = \sqrt{\frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{(n_1+n_2-2)}}, \qquad \text{d.f.} = n_1+n_2-2.$$

New s_p equation is just an extension of the t-test formula

$$SE(\bar{Y}_2-\bar{Y}_1) = s_p\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}.$$

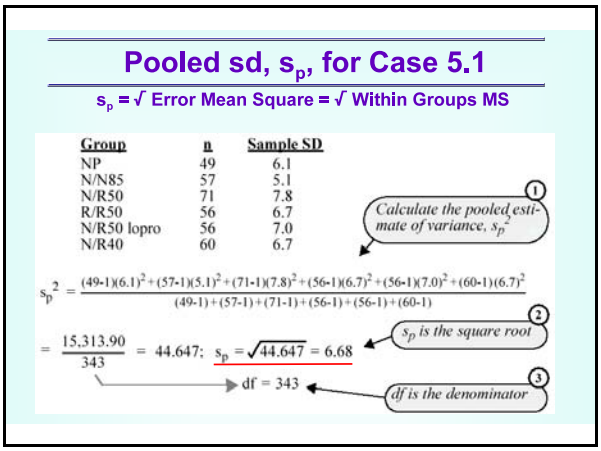
100(1− α)% Confidence Limits for the Difference Between Means

$$(\bar{Y}_2-\bar{Y}_1) \pm t_{df}(1-\alpha/2)SE(\bar{Y}_2-\bar{Y}_1).$$

Using ANOVA, obtain a more precise s_p, √'within groups MS', with more *df* for p values & CI's

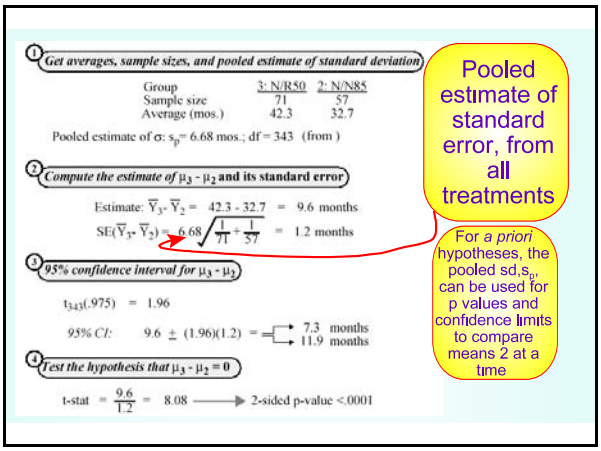
Slide 33 Pooled sd (sp) in t-tests & ANOVA

NOTES:



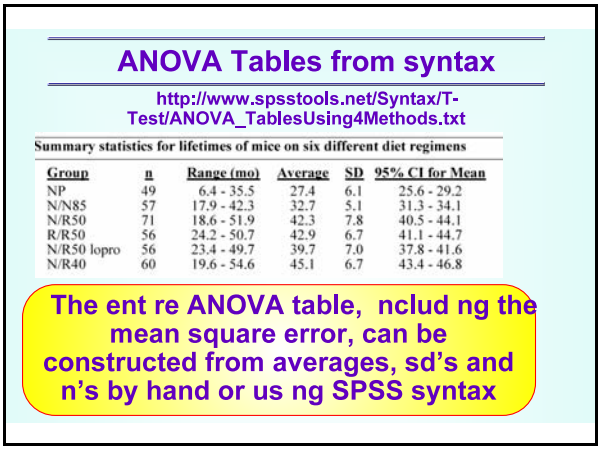
Slide 34 Pooled sd, sp, for Case 5.1

NOTES:



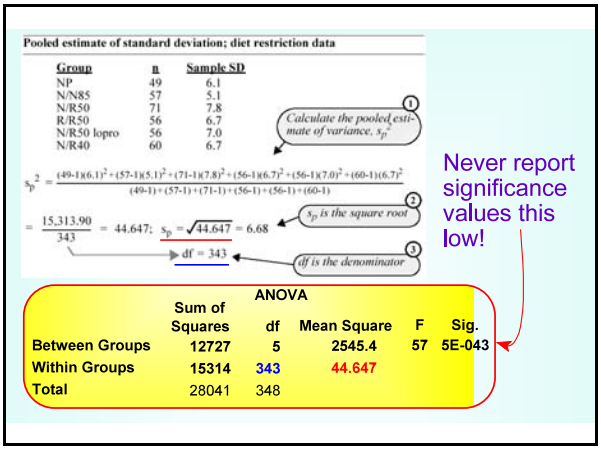
Slide 35

NOTES:



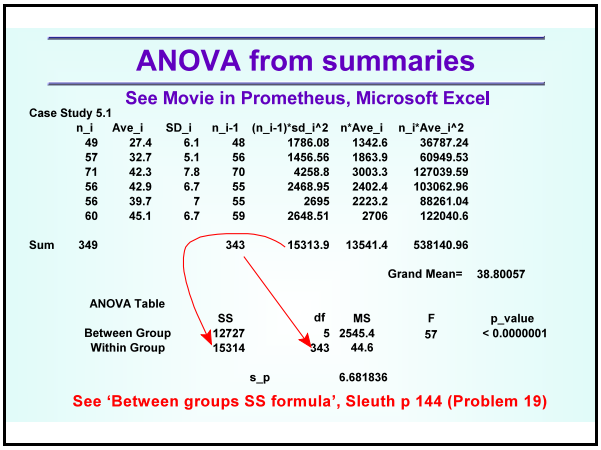
Slide 36 ANOVA Tables from syntax

NOTES:



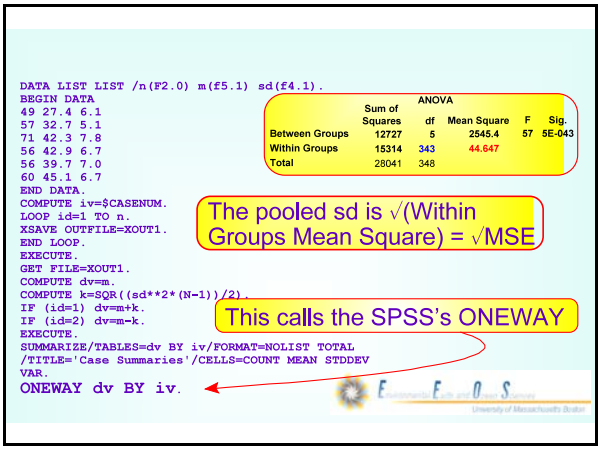
Slide 37

NOTES:



Slide 38 ANOVA from summaries

NOTES:



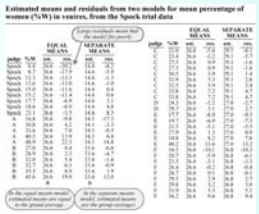
Slide 39

NOTES:

5.3 The One-Way Analysis of Variance *F*-test (Spock data)

5.3.1 Extra-Sum-of Squares principle & equal means model

- Display 5.8, Sleuth page 124
- The extra sum of squares is the single number that summarizes the difference in the sizes of residuals from the full and reduced models, p. 124
- This sum of squares, when divided by the appropriate *df*, estimates a variance
  - The F statistic assess the p-value of the equality of two variance estimates



EEOS611

Slide 40 5.3 The One-Way Analysis of Variance F-test (Spock data)

NOTES:

'Extra sum of squares' *F* statistic

Sleuth Section 5.3.1

Extra sum of squares =
Residual sum of squares (reduced model) -
Residual sum of squares (full model)

F statistic =
{(Extra sum of squares)/(Extra degrees of freedom)}

$\sigma^2_{\text{full model}}$ ← This variance is often the 'within groups' mean square

Tested with $F_{\{\text{Extra } df, \text{ Error } df \text{ full model}\}}$

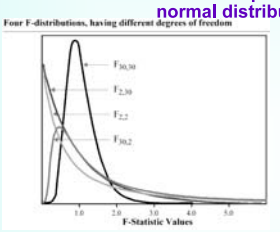
Slide 41 'Extra sum of squares' F statistic

NOTES:

F distribution

Snedecor's named the *F* distribution to honor Ronald Fisher

The F distribution can be regarded as the expected ratios of variances from samples drawn from the same normal distribution



These are probability density functions, with area 1.0. Table A.4 (p. 720) provides the area to the left of the F statistic for df1 & df2

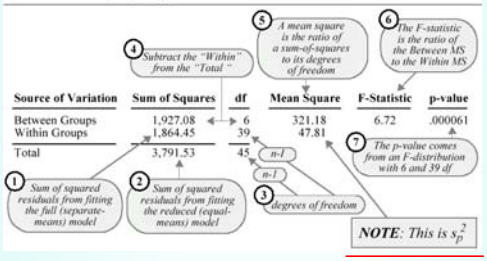
Slide 42 F distribution

NOTES:

ANOVA Table for Spock data

Partition the sum of squares

Analysis of variance table: a test for equal mean percents of women in venires of seven judges; Spock data



Slide 43 ANOVA Table for Spock data

NOTES:

SPSS ANOVA Table

Case 5.2 Spock trial

ANOVA

Percentage Women

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1927	6	321	6.7	.00006
Within Groups	1864	39	48		
Total	3792	45			

- Three ways to do 1-way ANOVA's in SPSS
  - Analyze\compare means\One-way ANOVA
  - Analyze\General Linear Model\Univariate
  - Analyze\Regression\Linear
- Each method has its strengths. All produce identical p values. ANOVA the simplest but least flexible

Slide 44 SPSS ANOVA Table

NOTES:

Case 5.2 ANOVA table, p. 130

Do as separate one-way ANOVAs, *t* tests with appropriate s_p for p-values or as linear contrast (next chapter)

Complete analysis of variance table for three tests involving the mean percents of women in venires of seven judges

Source of Variation	Sum of Squares	df	Mean Square	F-Statistic	p-value
Between Groups	1,927.08	6	321.18	6.72	0.000061
Spock v. Others	1,600.63	1	1,600.63	32.14	0.000001
Among Others	326.45	5	65.29	1.37	0.26
Within Groups	1,864.45	39	47.81		
Total	3,791.53	45			

Warning This ANOVA table only appropriate if 'judges' is regarded as a fixed effect, producing a fixed effect hierarchic (nested) ANOVA (Chapter 16 & Neter *et al.*) The Spock judge effect is nested within the judge effect (Between Groups)

book, corrected on the Sleuth errata web site

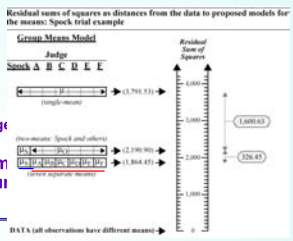
Slide 45 Case 5.2 ANOVA table, p. 130

NOTES:

'Spock judge' vs. other judges

Display 5.11, page 129

- Calculate the mean for the other 6 judges (A-F)
- Find and sum the squared residuals from that new 'other- judge' mean
  - This pooling may NOT be appropriate if there is large judge-to-judge variability
- Then, test that residual sum of squares with an extra sum of squares F test



Slide 46 'Spock judge' vs. other judges

NOTES:

ANOVA: robustness to assumptions

- Normality is not critical. Extremely long-tailed distributions or skewed distributions, coupled with different sample sizes present the only serious distributional problems
- The assumptions of independence within and across groups is critical
- The assumption of equal standard deviations in the populations is crucial. Also called the equal variance assumption, homoscedasticity assumption (vs. Heterorscedasticity)
- The tools are **not** resistant to severely outlying observations.

Slide 47 ANOVA: robustness to assumptions

NOTES:

Assumptions of ANOVA

Not robust to heteroscedasticity! (But Winer et al. argue that p values are robust if sample sizes equal — Sleuth appears to have a counterargument)

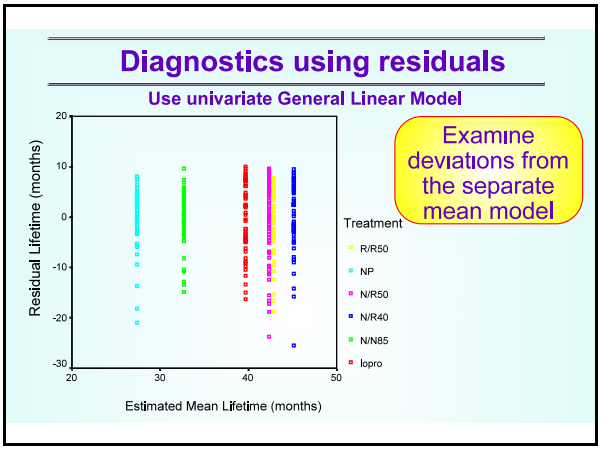
Display 5.13, page 131

Success rates for 95% confidence intervals for $\mu_1-\mu_2$ from samples simulated from normal populations with possibly different SDs

			$\sigma_2=\sigma_1$			$\sigma_2=2\sigma_1$		
n_1	n_2	n_3	$\sigma_3=\sigma_1$	$\sigma_3=2\sigma_1$	$\sigma_3=4\sigma_1$	$\sigma_3=\sigma_1$	$\sigma_3=2\sigma_1$	$\sigma_3=4\sigma_1$
10	10	10	95.4	98.9	99.9	91.9	96.8	99.6
20	10	10	95.5	98.7	99.8	84.8	91.7	98.9
10	20	10	94.1	98.7	99.9	97.0	98.8	99.8
10	10	20	95.6	99.6	99.9	90.4	97.5	99.9

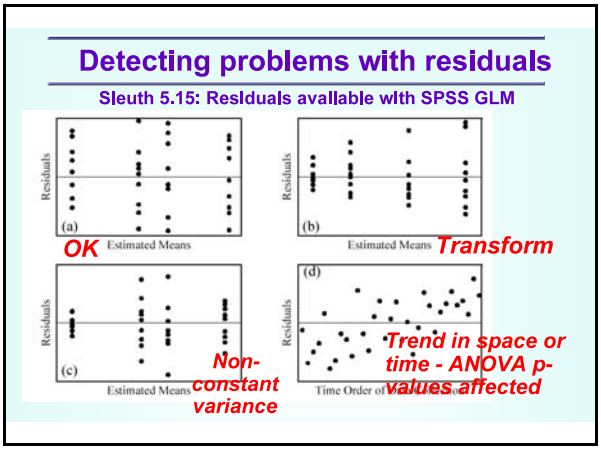
Slide 48 Assumptions of ANOVA

NOTES:



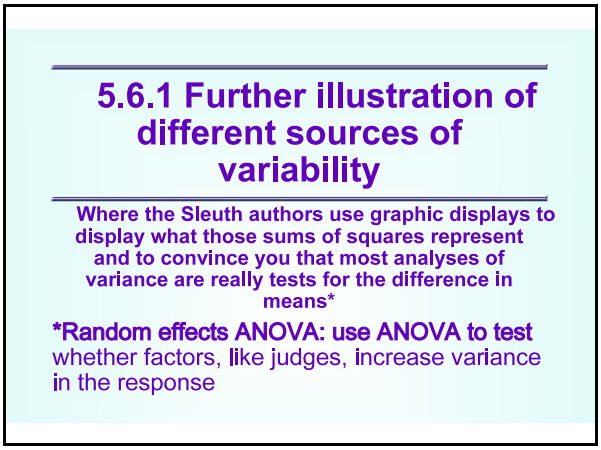
Slide 49 Diagnostics using residuals

NOTES:



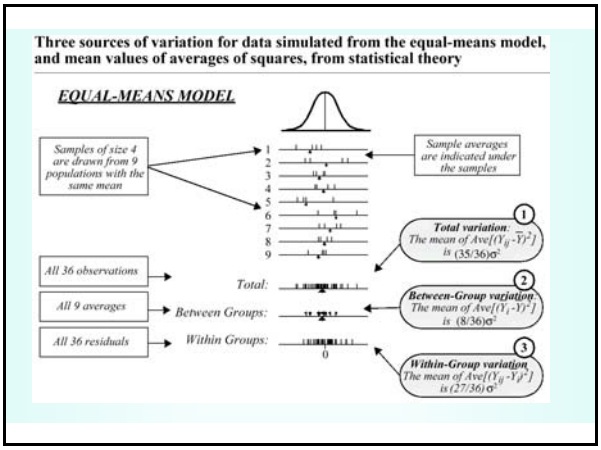
Slide 50 Detecting problems with residuals

NOTES:



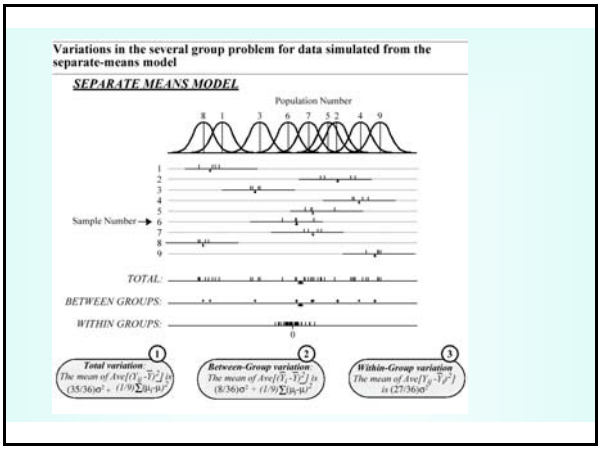
Slide 51 5.6.1 Further illustration of different sources of variability

NOTES:



Slide 52

NOTES:



Slide 53

NOTES:

5.6.2 Kruskal-Wallis Nonparametric ANOVA

Available In SPSS Non-parametric tests
Reslstant to outllers – but susceptlble to unequal varlance

Spock trial data, rank-transformed

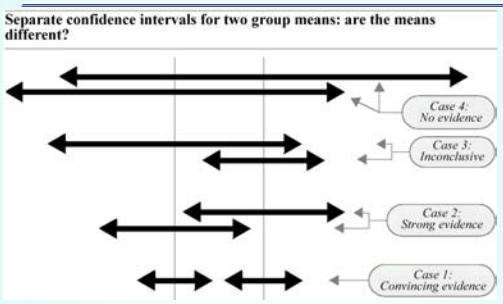
Judge	Rank of venire from smallest (1) to largest (46) percent women								
Spock's	1	2	3	4	5	6	9.5	11	16
A	8	31	37	44	46				
B	22	26	34	36	41	45			
C	14	17	23.5	23.5	30	32.5	35	38.5	38.5
D	19	28							
E	9.5	12	15	25	40	43			
F	7	13	18	20	21	27	29	32.5	42

K W ANOVA does not permit analysis of any but the simplest designs

Slide 54 5.6.2 Kruskal-Wallis Nonparametric ANOVA

NOTES:

Confidence limits & significant differences



Slide 55 Confidence limits & significant differences

NOTES:

Statistical *vs.* Scientific significance

Always report the effect size (don't just report 'significant' or NS)

- Deming: report effect sizes for tests
- Many **statistically significant** results are trivial ecologically (or chemically or socially).
  - All null hypotheses are wrong: $\mu_1 = \mu_2$ and the p-value is often dependent on the sample size
    - A p value of 0.00001 may not be ecologically meaningful if there is only a 1% difference in effects and at least a 5% difference causes changes in the ecosystem
- Tests with large p values may be consistent with important ecological effects
  - What is the probability of Type II error?

Slide 56 Statistical vs. Scientific significance

NOTES:

When is an effect 'random'?

See Sleuth Page 136-138: 'The Random Effects model'

- The differences among subgroup means is NOT of intrinsic interest.
  - You may be interested in whether the effect changes from day to day – i.e, estimating day-to-day or 'among day' variance – but you are not interested specifically in the differences on any pair of days
- If the number of levels of a factor is small relative to the total possible levels of a factor (not the case with district Judges since ALL were sampled)
- Are the subgroups a representative or random sample of some larger group?

EEOS611

Slide 57 When is an effect 'random'?

NOTES:

Quinn & Keough (2002, p. 176) on Fixed vs. Random Factors

Random effects models allow inferences to a larger population

- Investigators use only a random subset of the possibly causal levels of a factor (or factors) and wish to make inferences to all possible levels of the factor
  - e.g., EPA selects a random subsample of zinc-contaminated streams and analyzes the date with a random-effects model
- Q & K: random or at least haphazard selection of experimental or observational units is essential

EEOS611

Slide 58 Quinn & Keough (2002, p. 176) on Fixed vs. Random Factors

NOTES:

Comparing Spock with the other judges: Fixed or random effects?

Is the judge effect fixed or random?

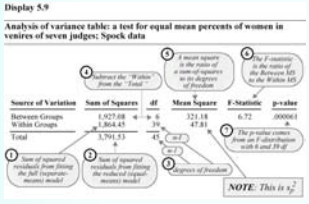
- Type I ANOVA: Fixed effects ANOVA: test for differences in the averages among groups
- Type II ANOVA: Random effects ANOVA: test differences in variances due to the group classification
- Mixed model: Fixed & random factors
- Note
  - The calculations are often identical for random and fixed-effects ANOVA, but the interpretations are different
  - Factorial ANOVA (>1 factor), the F statistics differ among models, with a different denominator mean square for random factors
  - The inference allowed differs among models

Slide 59 Comparing Spock with the other judges: Fixed or random effects?

NOTES:

5.17 Reproduce Display 5.9

Solution as a one-way ANOVA problem
Type I: There is at least 1 difference in the average percentage women jurors that is greater than expected by chance
Type II: There is more judge-to-judge variability in % female voters than expected by chance


Display 5.9

Analysis of variance table: a test for equal mean percents of women in venires of seven judges; Spock data

Source of Variation	Sum of Squares	df	Mean Square	F-Statistic	p-value
Between Groups	1,927.08	6	321.18	6.72	.000061
Within Groups	1,864.45	39	47.81		
Total	3,791.53	45			

Slide 60 5.17 Reproduce Display 5.9

NOTES:

District judges: Random?

If the judge effect is a random factor, this design is pseudoreplicated and invalid. But, the judges are NOT a random subset of a larger class of judges. These 7 judges represent all of the judges. The model is a fixed effect design

Complete analysis of variance table for three tests involving the mean percents of women in venires of seven judges

Source of Variation	Sum of Squares	df	Mean Square	F-Statistic	p-value
Between Groups	1,927.08	6	321.18	6.72	0.000061
Spock v. Others	1,600.63	1	1,600.63	32.14	0.000001
Among Others	326.45	5	65.29	1.37	0.26
Within Groups	1,864.45	39	47.81		
Total	3,791.53	45			

ANOVA uses an inappropriate denominator mean square for the Spock judge effect

Slide 61 District judges: Random?

NOTES:

Fixed *vs.* Random effects

Underwood (1997): Fixed effects 1-way ANOVA

$X_{ij} = \mu + A_i + e_{ij}$

where X_{ij} is jth replicate in ith treatment (ith level of factor A; $i = 1 \ldots a$), A_i is difference between ith level of factor A and overall mean of all levels (μ), e_{ij} is the deviation of replicate j in ith sample from the mean of that population.

Fixed factor:
By definition:

$$\sum_{i=1}^{a} A_i = 0$$

(see Section 7.6).

ıre estimates

Among treatments	$\sigma_\varepsilon^2 + \frac{n\sum_{i=1}^{a}(A_i - \bar{A})^2}{(a-1)}$ or $\sigma_\varepsilon^2 + nk_A^2$
Within treatments	σ_ε^2

where k_A^2 indicates fixed differences, all sampled in the experiment.

Slide 62 Fixed vs. Random effects

NOTES:

Fixed vs. Random factors

Underwood (1997): Random factor (Model II) 1-way

Random factor:

$$E\left(\sum_{i=1}^{a} A_i\right) = 0$$

Meaning you *expect* $\sum_{i=1}^{a} A_i = 0$ on average, over many experiments, but in a single experiment, A_i values as sampled may not sum to zero.

Analysis of variance	Mean square estimates
Among treatments	$\sigma_\varepsilon^2 + n\sigma_A^2$
Within treatments	σ_ε^2

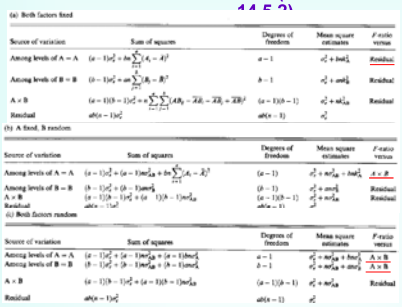
where σ_A^2 is the variance of the population of A_i values sampled in your experiment.

Slide 63 Fixed vs. Random factors

NOTES:

Factorial ANOVA

Tables from Underwood (will be covered in Sleuth 14.5.2)



EEOS611

Slide 64 Factorial ANOVA

NOTES:

Mixed Model Nested ANOVA

A 1 in 67 chance of observing such a difference by chance

Tests of Between-Subjects Effects

Dependent Variable: Percentage Women

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	20003	1	20003	293	
	Error	273	4.003	68[a]		
SPOCK	Hypothesis	1537	1	1537	21.6	.015
	Error	236	3.311	71[b]		
CODE(SPOCK)	Hypothesis	326	5	65	1.4	.258
	Error	1864	39	48[c]		

a. 1.167 MS(CODE(SPOCK)) - .167 MS(Error)
b. 1.337 MS(CODE(SPOCK)) - .337 MS(Error)
c. MS(Error)

This model is not appropriate because the judges are not a random subset of judges

```
UNIANOVA
  percent  BY spock code
  /METHOD = SSTYPE(3)
  /INTERCEPT = INCLUDE
  /CRITERIA = ALPHA(.05)
  /random code
  /DESIGN = spock code(spock) .
```

Slide 65 Mixed Model Nested ANOVA

NOTES:

Counterfactual conditionals

Modus tollens
Hypothesis:
If A then B
Observe 'Not B'
Then conclude: 'Not A' [Reject null]

Counterfactual conditional
If A and C then B
Not C
No inference possible about the truth or falsity of A can be inferred from observing either 'B' or 'Not B'

Modus tollens
If the Spock judge's venire were due to chance, then F=1.0
Observe F=22 (p=0.015, $F_{1,3.3}$)

Counterfactual conditional
If the juries were chosen by chance and the judges were a random subset of judges AND we know that the judges were NOT a random subset of judges. Then F=1
Observing F>1 doesn't allow us to conclude anything about the fairness of the jury selection

Slide 66 Counterfactual conditionals

NOTES:

Conclusions
(1 of 3)

- ANOVA tables can be created from summary statistics
- Assumptions:
  - Homoscedasticity
    - Levene's test a rough guide
    - Boxplots or residual plots are the standard tools for assessing homoscedasticity (equal variance among groups)
    - Spread vs. Level plots
  - Independence of errors among groups a key ANOVA assumption
  - Normally distributed errors (not underlying data) not crucial

EEOS611

Slide 67 Conclusions

NOTES:

Conclusions
(2 of 3)

- An ANOVA is more efficient & powerful than multiple, separate *t* tests
  - The ANOVA error MS (=within groups MS) provides a more **precise** estimate of the population standard deviation [It is not a smaller estimate of error {it is an unbiased estimator}]
- Kruskal-Wallis ANOVA is the rank-based analogue of 1-way ANOVA and is resistant to outliers but not unequal spread
  - Ties correction must be used
  - Effect sizes, hierarchic structure, and covariates difficult to handle

Slide 68 Conclusions

NOTES:

Conclusions
(3 of 3)

- ANOVA tests for difference in means (fixed effect) or whether $\sigma_i^2 = 0$ (random effect) or both (mixed model)
- Fixed vs. random effects
  - The choice of fixed vs. random effects is often crucial and depends on whether the factor levels (judges in the Spock example) represent a random or representative sample from some larger statistical population
  - The *F* statistics and interpretation of the results sometimes change depending on whether fixed or random effects are chosen

EEOS611

Slide 69 Conclusions

NOTES: