Ch 5 (end)
Ch 6: Linear combinations and multiple comparisons of means

Class 9: 3/4/09 W [3/2/09 M was a snow day]

EEOS611

Slide 1 Ch 5 (end)

Ch 6: Linear combinations and multiple comparisons of means

NOTES:

HW 7 due Friday 3/6/09 Noon

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

- Note new due date!
- New WIMBA sesssion tonight at 9 pm to go over analyses
- Read Chapter 6 **Comparisons among several samples**
- Comment on Chapter 6 conceptual problems in Blackboard Vista4
- Computation Problem 7
  - Problem 5.25 Duodenal ulcers
  - Hints: Use boxplots to analyze the equal variance assumption and to check for outliers
  - Use the advice from Display 3.6 (p 66) to evaluate outlier effects
  - Assume that the hypotheses are *a priori*
    - This allows the use of the LSD (Ch 6, p. 162), the approach used in case 5.1
    - You can use ONEWAY or GLM/Univariate, both have an option for post hoc/LSD tests

EEOS611

Slide 2 HW 7 due Friday 3/6/09 Noon

NOTES:

HW 8 due Monday 3/9/09 10 am

Submit as Myname-HW8.doc (or *.rtf)

- Read Chapter **7 Comparisons among several samples**
- Comment on Chapter 7 conceptual problems in Blackboard Vista4
- Computation Problem 8
  - Problem 6.22 A biological basis for homosexuality
  - You must use linear contrasts to solve the problem
  - You can assume that the contrasts were specified *a priori*

EEOS611

Slide 3 HW 8 due Monday 3/9/09 10 am

NOTES:

Student Presentations

Starting at 10:50 (8 minutes each)

- 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 5: Comparisons among several samples

EEOS611

Slide 5 Chapter 5: Comparisons among several samples

NOTES:

ANOVA, Analysis of Variance, the foundation of experimental design

- Most experimental & survey 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 6 ANOVA, Analysis of Variance, the foundation of experimental design

NOTES:

Case 5.1 Diet restriction & longevity

EEOS611

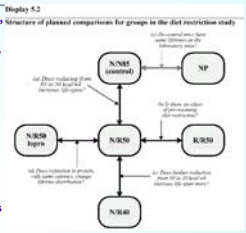
Slide 7 Case 5.1 Diet restriction & longevity

NOTES:

Planned comparisons

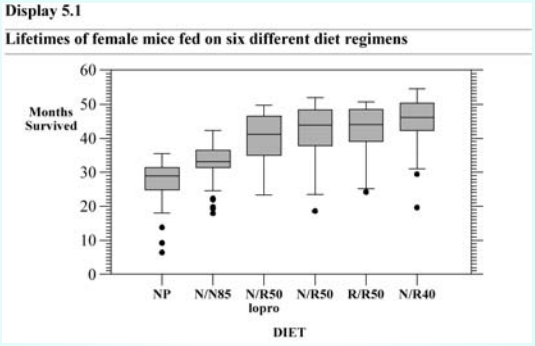
These are *a priori* contrasts, not *a posteriori*

- If hypotheses are specified 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 8 Planned comparisons

NOTES:

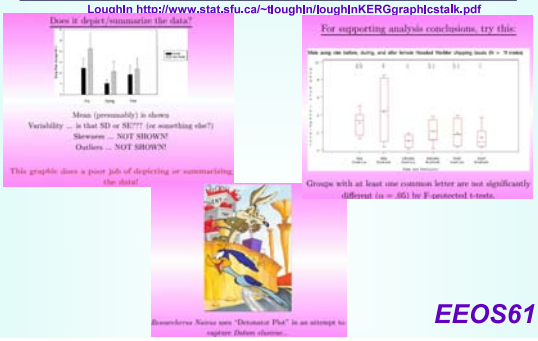


Slide 9

NOTES:

Loughin on detonator plots

Loughlin http://www.stat.sfu.ca/~tloughlin/loughlinKERGgraphicstalk.pdf



EEOS611

Slide 10 Loughin on 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 11 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 12 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)

Note that all 5 of these hypotheses can be tested as Tukey LSD tests (or linear contrasts) see Ch 6

EEOS611

Slide 13 Summary of statistical findings

NOTES:

Case study 5.2: The Spock trial

EEOS611

Slide 14 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

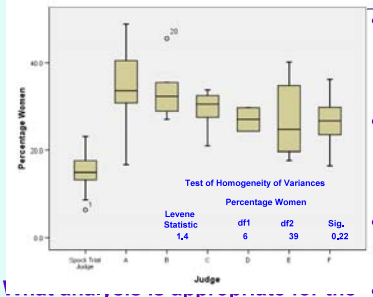
	Spock Trial Judge	Other Boston Area U.S. District Court Judges A	B	C	D	E	F
0	64,87						
1	33,36,50,52,77,86	68				77,97	65
2	31		70,89	10,34,75,75	43,97	15,79	07,35,64,67,95,98
3		08,36	20,27,55	05,19,25,38,38		48	19,62
4		05,89	56			02	

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 15 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 16

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.

EEOS611

Slide 17 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 18 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

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

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

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

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

Note that the s_p is available as theTukey LSD tests (or linear contrasts) see Ch 6

Slide 19 Pooled sd (sp) in t-tests & ANOVA

NOTES:

Pooled sd, s_p, for Case 5.1

s_p = √ Error Mean Square = √ Within Groups MS

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

① Calculate the pooled estimate of variance, s_p^2

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

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

② s_p is the square root

df = 343 ③ df is the denominator

Slide 20 Pooled sd, sp, for Case 5.1

NOTES:

---

① Get averages, sample sizes, and pooled estimate of standard deviation

Group	3: N/R50	2: N/N85
Sample size	71	57
Average (mos.)	42.3	32.7

Pooled estimate of σ: s_p = 6.68 mos.; df = 343 (from)

② Compute the estimate of $\mu_3 - \mu_2$ and its standard error

Estimate: $\bar{Y}_3 - \bar{Y}_2 = 42.3 - 32.7 = 9.6$ months

$SE(\bar{Y}_3 - \bar{Y}_2) = 6.68\sqrt{\frac{1}{71}+\frac{1}{57}} = 1.2$ months

③ 95% confidence interval for $\mu_3 - \mu_2$

$t_{343}(.975) = 1.96$

95% CI: $9.6 \pm (1.96)(1.2) =$ 7.3 months, 11.9 months

④ Test the hypothesis that $\mu_3 - \mu_2 = 0$

t-stat = $\frac{9.6}{1.2} = 8.08$ → 2-sided p-value <.0001

Pooled est mate of standard error, from all treatments

For *a priori* hypotheses, the pooled sd, s_p, can be used for p values and conf dence l m ts to compare means 2 at a t me

Slide 21

NOTES:

ANOVA Tables from syntax

http://www.spsstools.net/Syntax/T-Test/ANOVA_TablesUsing4Methods.txt

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

These CI's are ava lable as the Tukey LSD contrast The ent re ANOVA table can be constructed from averages, sd's and n's by hand or us ng SPSS syntax

Slide 22 ANOVA Tables from syntax

NOTES:

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

① Calculate the pooled estimate of variance, s_p^2

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

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

② s_p is the square root

df = 343 ③ df is the denominator

Never report significance values this low!

ANOVA

	Sum of Squares	df	Mean Square	F	S g
Between Groups	12727	5	2545 4	57	5E 0
W th n Groups	15314	343	44 647		
Total	28041	348			

Slide 23

NOTES:

ANOVA from summaries

See Movie in Prometheus, Microsoft Excel

Case Study 5.1

	n_i	Ave_i	SD_i	n_i-1	(n_i-1)*sd_i^2	n*Ave_i	n_i*Ave_i^2
	49	27.4	6.1	48	1786.08	1342.6	36787.24
	57	32.7	5.1	56	1456.56	1863.9	60949.53
	71	42.3	7.8	70	4258.8	3003.3	127039.59
	56	42.9	6.7	55	2468.95	2402.4	103062.96
	56	39.7	7	55	2695	2223.2	88261.04
	60	45.1	6.7	59	2648.51	2706	122040.6
Sum	349			343	15313.9	13541.4	538140.96

Grand Mean= 38.80057

ANOVA Table

	SS	df	MS	F	p_value
Between Group	12727	5	2545.4	57	< 0.0000001
Within Group	15314	343	44.6		
		s_p	6.681836		

See 'Between groups SS formula', Sleuth p 144 (Problem 19)

Slide 24 ANOVA from summaries

NOTES:

```
DATA LIST LIST /n(F2.0) m(f5.1) sd(f4.1).
BEGIN DATA
49 27.4 6.1
57 32.7 5.1
71 42.3 7.8
56 42.9 6.7
56 39.7 7.0
60 45.1 6.7
END DATA.
COMPUTE iv=$CASENUM.
LOOP id=1 TO n.
XSAVE OUTFILE=XOUT1.
END LOOP.
EXECUTE.
GET FILE=XOUT1.
COMPUTE dv=m.
COMPUTE k=SQR((sd**2*(N-1))/2).
IF (id=1) dv=m+k.
IF (id=2) dv=m-k.
EXECUTE.
SUMMARIZE/TABLES=dv BY iv/FORMAT=NOLIST TOTAL
/TITLE='Case Summaries'/CELLS=COUNT MEAN STDDEV
VAR.
ONEWAY dv BY iv.
```

ANOVA

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	12727	5	2545.4	57	5E-043
Within Groups	15314	343	44.647		
Total	28041	348			

The pooled sd is √(Within Groups Mean Square) = √MSE

This calls the SPSS's ONEWAY

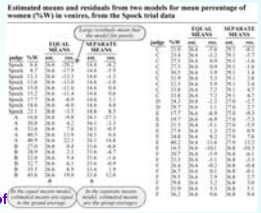
Slide 25

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 26 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

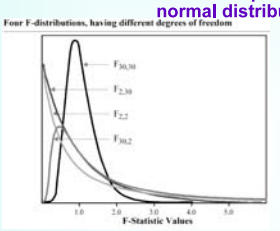
Tested with $F_{\{\text{Extra } df,\ \text{Error } df \text{ full model}\}}$

Slide 27 ‘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 28 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

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			

1. Sum of squared residuals from fitting the full (separate-means) model
2. Sum of squared residuals from fitting the reduced (equal-means) model
3. degrees of freedom
4. Subtract the “Within” from the “Total”
5. A mean square is the ratio of a sum-of-squares to its degrees of freedom
6. The F-statistic is the ratio of the Between MS to the Within MS
7. The p-value comes from an F-distribution with 6 and 39 df

n-1

NOTE: This is s_p^2

Slide 29 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			

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

Slide 30 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

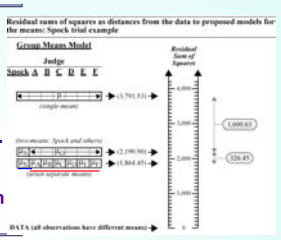
Slide 31 Case 5.2 ANOVA table, p. 130

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 32 '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 33 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. See Display 5.13)

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

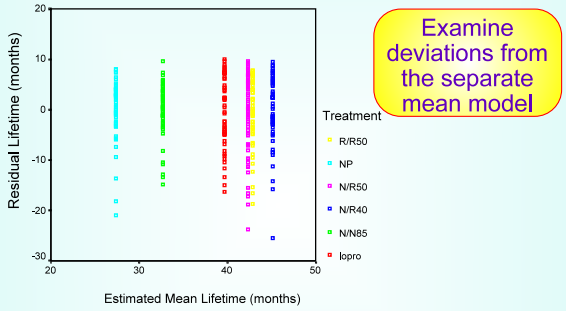
			$\sigma_2=\sigma_1$			$\sigma_2=2\sigma_1$		
n_1	n_2	n_3	$\sigma_3=\sigma_1$	$\sigma_3=2\sigma_1$	$\sigma_3=4\sigma_1$	$\sigma_3=\sigma_1$	$\sigma_3=2\sigma_1$	$\sigma_3=4\sigma_1$
10	10	10	95.4	98.9	99.9	91.9	96.8	99.6
20	10	10	95.5	98.7	99.8	84.8	91.7	98.9
10	20	10	94.1	98.7	99.9	97.0	98.8	99.8
10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

Slide 34 Assumptions of ANOVA

NOTES:

Diagnostics using residuals

Use univariate General Linear Model



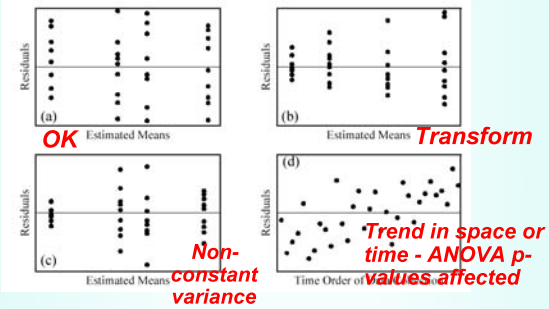
Examine deviations from the separate mean model

Slide 35 Diagnostics using residuals

NOTES:

Detecting problems with residuals

Sleuth 5.15: Residuals available with SPSS GLM



Slide 36 Detecting problems with residuals

NOTES:

5.6.1 Further illustration of different sources of variability

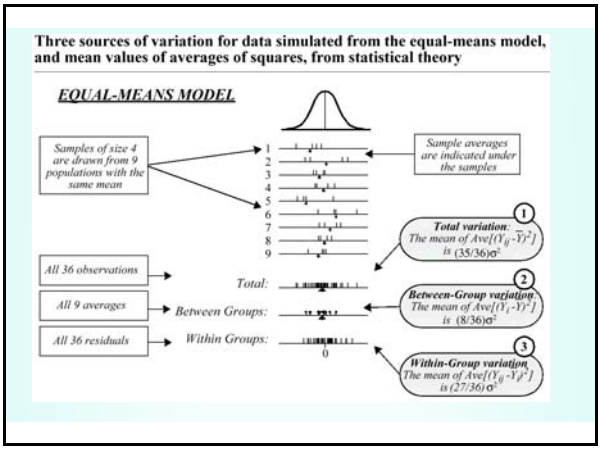
Where the Sleuth authors use graphic displays to display what those sums of squares represent and to convince you that most analyses of variance are really tests for the difference in means*

*Random effects ANOVA: use ANOVA to test whether factors, like judges, increase variance in the response

EEOS611

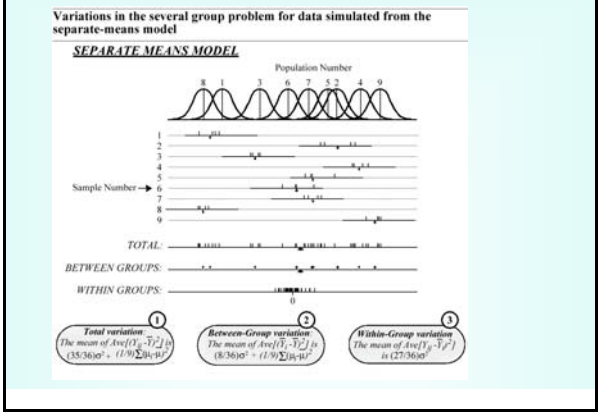
Slide 37 5.6.1 Further illustration of different sources of variability

NOTES:



Slide 38

NOTES:



Slide 39

NOTES:

Slide 40 5.6.2 Kruskal-Wallis Nonparametric ANOVA

5.6.2 Kruskal-Wallis Nonparametric ANOVA

Available in SPSS Non-parametric tests
Resistant to outliers – but susceptible to unequal variance

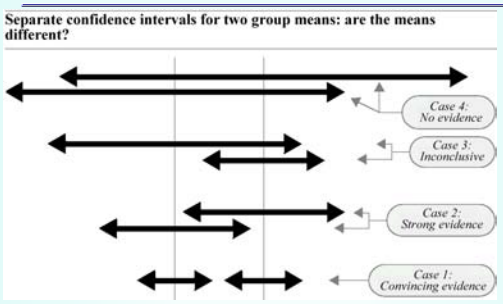
Spock trial data, rank-transformed

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

NOTES:

Confidence limits & significant differences



Slide 41 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).
  - Most null hypotheses ($\mu_1 = \mu_2$) are false and the p-value is often dependent on the sample size
  - e.g., a p value of 0.00001 may not be ecologically meaningful if there is only a minor difference in effects and a much larger difference causes meaningful changes in the ecosystem
- Test statistics with large p values (>0.1) but with broad 95% confidence intervals may be consistent with important ecological effects
  - What is the probability of Type II error?
  - What are the ecological consequences of failing to reject a false null hypothesis

Slide 42 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 43 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 44 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?

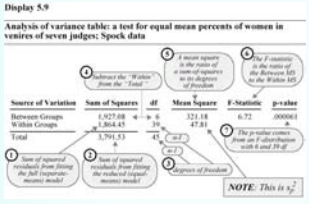
- Type I ANOVA: Fixed effects ANOVA: test for differences in the averages among groups
- Type II ANOVA: Random effects ANOVA: test differences in variances due to the group classification
- Mixed model: Fixed & random factors
- Note
  - The calculations are often identical for random and fixed-effects 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 45 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	1,927.68	6	321.18	6.72	.000061
Within Groups	1,864.45	39	47.81		
Total	3,791.53	45			

Slide 46 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 thus a fixed effect design, and the F statistic is appropriate.

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 47 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).

...re estimates

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

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

Slide 48 Fixed vs. Random effects

NOTES:

Fixed vs. Random factors

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

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	$\sigma_e^2 + n\sigma_A^2$
Within treatments	σ_e^2

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

Slide 49 Fixed vs. Random factors

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 50 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 the expected F is1.0
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
BUT we know that the judges were NOT a random subset of judges (Not C)
Observing F>1 doesn't allow us to conclude anything about the fairness of the jury selection

Slide 51 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

Slide 52 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 53 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

Slide 54 Conclusions

NOTES:

Ch 6: Linear combinations and multiple comparisons of means

EEOS611

Slide 55 Ch 6: Linear combinations and multiple comparisons of means

NOTES:

Case Study 6.1.1

Discrimination against the handicapped

- U.S. Vocational Rehabilitation Act of 1973
- 5 Videotaped job inteviews
  - Applicant appeared with different handicaps
  - Wheelchair
  - Crutches
  - Hearing impaired
  - Amputated
  - No handicap
- 70 undergraduates randomly assigned to view tapes, 14 to each tape.
- Rated on a 1 to 10 applicant qualification scale

EEOS611

Slide 56 Case Study 6.1.1

NOTES:

Display 6.1

Applicant qualification scores; Control group is in the middle of the distribution of scores

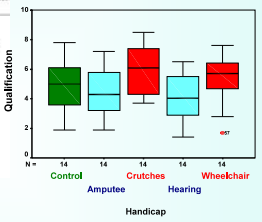
Display 6.1

Stem-and-leaf diagrams of applicant qualification scores given to applicants simulating five different handicap conditions

Applicant's Handicap

	None	Amputee	Crutches	Hearing	Wheelchair
0		9		4	7
1	9	56		149	8
2	5	268	7	479	5
3	06	06	033	237	78
4	129	3589	18	589	03
5	149	1	0234	5	1124
6	17	2	445		246
7	48		5		
8					
9					

Legend: 7| 4 represents a score of 7.4 on the Applicant Qualification Scale



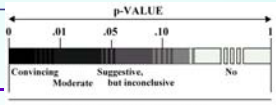
Slide 57 Display 6.1

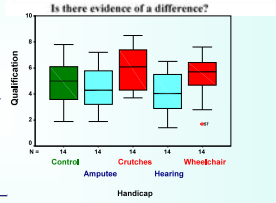
NOTES:

Case 6.01

Summary of Statistical Findings

- Strong but not convincing evidence p = 0.03 that applicants' ratings influenced by handicaps
- Ave 'Crutches' score much higher (1.87 ± 0.73) than hearing, using Tukey-Kramer multiple comparison test (difference in score ± ½ 95% CI)
- The strongest evidence is for a difference between "wheelchair & crutches" vs. "Amputee & hearing" (t-statistic = 3.19 for linear contrasts), with a 1.4±0.9 higher average score for the former group
- None of the feigned handicaps different from control! (The protected least significant differences all have 2-sided p > 0.05)





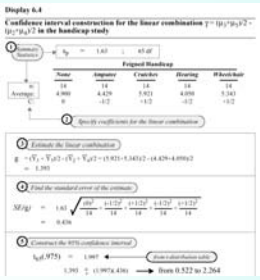
Slide 58 Case 6.01

NOTES:

Display 6.4

Wheelchair + Crutches vs. Amputee + hearing

- Linear contrasts can be solved
  - By hand
  - With SPSS Oneway
  - SPSS GLM (analyze/general linear model/univariate)
  - the appropriate 95% CI's for the average difference can be calculated
- SPSS routine Oneway will calculate the appropriate p value using a linear contrast, but it will not present the appropriate difference in means



Slide 59 Display 6.4

NOTES:

SPSS syntax for Linear contrasts

Display 6.4, p. 149 (1st ed), p. 155 (2nd ed)
Report results as 1.4 ± 0.9

- * DATA order Control Amputee Crutches Hearing Wheelchair.
- ONEWAY
- score BY code
- /CONTRAST = 0 -1 1 -1 1.
- * This call to GLM does it all.
- UNIANOVA
- score BY code
- /METHOD = SSTYPE(3)
- /INTERCEPT = INCLUDE
- /LMATRIX = "Avg A H vs Avg C W" code 0 -1/2 1/2
- /POSTHOC = code (TUKEY SCHEFFE LSD BONFERRONI)
- /CRITERIA = ALPHA(.05)
- /DESIGN = code .

Contrast Results (K Matrix)[a]

	Dependent Variable: Qualification
Contrast Estimate	1.4
Hypothesized Value	0
Difference (Estimate - Hypothesized)	1.4
Std. Error	.44
Sig.	.002
95% Confidence Interval for Difference — Lower Bound	.5
95% Confidence Interval for Difference — Upper Bound	2.3

a. Based on the user-specified contrast coefficients (L') matrix: Avg A H vs Avg C W

Construct the 95% confidence interval

$t_{95}(.975) = 1.997$ ← from t-distribution table

$1.393 \pm (1.997)(.436) \rightarrow$ from 0.522 to 2.264

EEOS611

Slide 60 SPSS syntax for Linear contrasts

NOTES:

5 Construct the 95% confidence interval

$t_{65}(.975) = 1.997$ ← from t-distribution table

ANOVA $1.393 \pm (1.997)(.436)$ → from 0.522 to 2.264

Qualification

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	30.5	4	7.6	2.9	.030
Within Groups	173.3	65	2.7		
Total	203.8	69			

Contrast Coefficients

	Handicap				
Contrast	Control	Amputee	Crutches	Hearing	Wheelchair
1	0	-1	1	-1	1

SPSS Oneway doesn't allow fractional contrast coefficients; the estimates are 2x too large, but the p values are ok

Contrast Tests

Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
2.79	.87	3.19	65	.002

Slide 61

NOTES:

Case 6.1.1

Scope of inference, Questions

- Scope of inference
  - Differences exist, but the situation is complicated by having the control having an average in the middle of the group of 5 treatments
  - How should one compare groups?
- Questions:
  - How does one perform linear contrasts in SPSS?
    - Use Oneway with contrasts
    - Use UNIANOVA (GLM) with /Lmatrix
  - What is the Tukey-Kramer procedure?
  - What is "the protected least significant difference"?
  - When should the Bonferonni & Scheffé procedures be used?

Slide 62 Case 6.1.1

NOTES:

Case 6.1.2

Preexisting preferences of fish – a randomized experiment

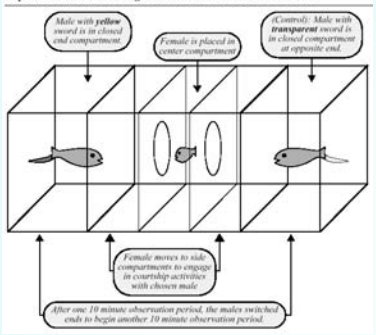
- Sexual selection by females
  - A. L. Basolo
- Southern platyfish: males don't produce the brightly colored sword tail
- Experiment
  - 6 pairs of males surgically given artificial plastic sword tails.
    - 1 individual of each pair received a yellow sword
    - the other a transparent sword.
  - Female fish placed in a compartment
  - Amount of 20 minute periods spent courting with the yellow-sword male recorded.

Slide 63 Case 6.1.2

NOTES:

Display 6.2 p. 145

Experimental tank allowing female fish to choose between males



EEOS611

Slide 64 Display 6.2

NOTES:

Display 6.3

Percent of courtship time spent by 84 females with the yellow-sword male; body sizes of the males are shown in parentheses

Pair 1 (35 mm)	Pair 2 (31 mm)	Pair 3 (33 mm)	Pair 4 (34 mm)	Pair 5 (28 mm)	Pair 6 (34 mm)
43.7	52.5	91.0	72.2	78.3	33.4
54.0	65.6	62.0	58.5	66.0	42.2
49.8	68.5	10.0	51.0	47.7	35.6
65.5	45.9	83.8	56.8	77.5	79.9
53.1	80.2	91.3	92.4	58.3	59.0
53.0	67.0	56.3	55.3	61.1	58.1
62.3	73.0	83.6	59.3	65.1	64.2
49.4	71.7	53.3	42.0	62.9	82.8
45.7	55.0	36.5	68.5	61.0	75.7
56.6	70.0	65.4	78.4		66.3
59.0	63.2	48.1	69.6		56.3
67.8	39.6	50.6	89.2		84.5
73.3	41.0	40.4	67.3		61.1
43.8	59.2	90.6	77.5		87.6
67.4		74.9			
58.1		56.0			
		67.5			
Average: 56.41	60.89	62.43	67.00	64.21	63.34
SD: 9.02	12.48	22.29	14.33	9.41	17.68
n: 16	14	17	14	9	14

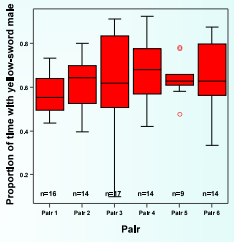
Slide 65 Display 6.3

NOTES:

Sexual preference

Case Study 6.2

- Test for preference for yellow-sword male (expected proportion = ½)
- Test for differences among pairs
- Test for the covariate of male fish weight using a linear contrast



Slide 66 Sexual preference

NOTES:

6.1.2

SUMMARY OF FINDINGS

- No evidence that the mean percentage of time with the yellow-sword male differed from one male pair to another [$P(F_{5,78} \geq 0.79)) = 0.56$]
- No evidence for linear relationship with male body size, from a linear contrast
  - Contrast available with one-way or general linear model
- Mean proportion (± 99.9% CI) with yellow sword is 62.4 (± 5.9) %
- This study provide convincing evidence that the mean percentage of time with the yellow tail exceeds the lack of preference value (50%)

EEOS611

Slide 67 6.1.2

NOTES:

Test for any difference among pairs

Page 158, Sleuth 2nd edition

Display 6.5

Analysis of the pre-existing preference example: F-test for differences in mean percent of time with yellow-tailed male and t-test for linear effect of male body size

ANOVA F-test

Source of Variation	Sum of Squares	df	Mean Square	F-Statistic	p-value
Between Male Groups	938.75	5	187.75	0.786	0.56
Within Groups	18,636.68	78	238.93		
Total	19,575.43	83			

Conclusion: *There is no evidence that the group means are different for different pairs of males (p-value = 0.56, from ANOVA F-statistic).*

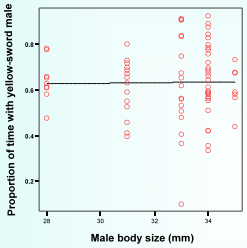
EEOS611

Slide 68 Test for any difference among pairs

NOTES:

Testing a linear contrast

Another way of handling some types of Analysis of Covariance: ANCOVA



Is there a non zero slope between proport on of t me spent w th yellow ta led sword and male pa r body s ze?

EEOS611

Slide 69 Testing a linear contrast

NOTES:

Testing for the male fish weight effect

t-test for Linear Effect of Body Size

Group	n	Average (%)	Standard Deviation	Male Body Size (mm)	Coefficient
Pair 1	16	56.41	9.02	35	5
Pair 2	14	60.89	12.48	31	-3
Pair 3	17	62.43	22.29	33	1 ← C's
Pair 4	14	67.00	14.33	34	3
Pair 5	9	64.21	9.41	28	-9
Pair 6	14	63.34	17.68	34	3
Pooled	84	62.13	15.46	Average =	32.5

① *Calculate the coefficients for the linear combination*

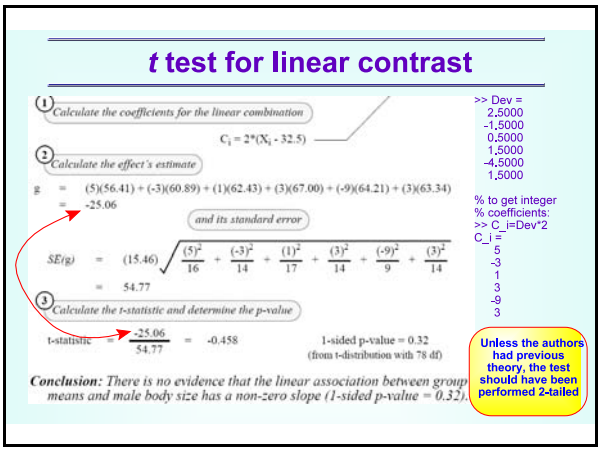
$C_i = 2*(X_i - 32.5)$

Gallagher Matlab code:

```
% Case0602 m
BodySize  [35 31 33 34 28 34];
mn=mean(BodySize);
% Subtract the mean from each body length
Dev  BodySize repmat(mn,size(BodySize))
% Solution:
>> Dev =
    2.5000
   -1.5000
    0.5000
    1.5000
   -4.5000
    1.5000
```

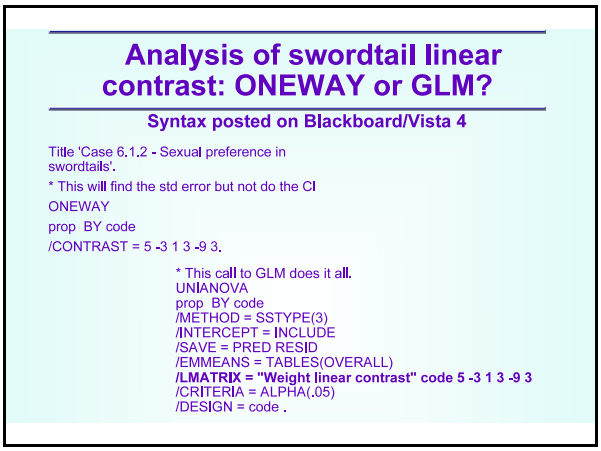
Slide 70 Testing for the male fish weight effect

NOTES:



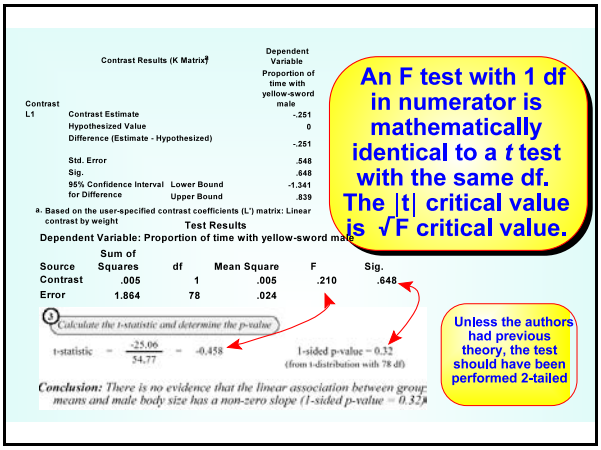
Slide 71 t test for linear contrast

NOTES:



Slide 72 Analysis of swordtail linear contrast: ONEWAY or GLM?

NOTES:



Slide 73

NOTES:

Matlab, Statbox orthpoly.m

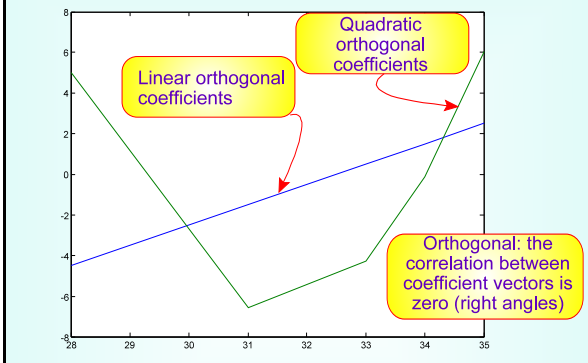
Gordon Smyth's <*Free*> Statbox 4.2
Also includes all of the major probability distributions and includes a nice routine for Poisson regression.
http://www.statsci.org/matlab/statbox.

```
html>> help orthpoly
ORTHPOLY  ORTHPOLY(X,N) calculates the orthogonal polynomials up
to order N corresponding to vector X.
BodySize=[35 31 33 34 28 34]';
format rat;orthpoly(BodySize,2)
ans =
1          5/2        1214/201
1         -3/2       -1318/201
1          1/2        -856/201
1          3/2         -22/201
1         -9/2        1004/201
1          3/2         -22/201
```

Order 2 polynomial. Is there a quadratic or curved effect?

Slide 74 Matlab, Statbox orthpoly.m

NOTES:

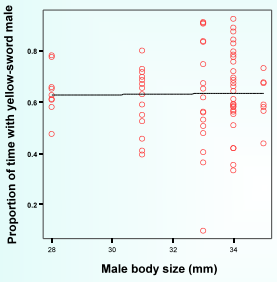


Slide 75

NOTES:

Testing a quadratic contrast

Quadratic implies a unimodal (humped) pattern



Quadratic contrast coefficients have mean zero, are proportional to X^2, and, if orthogonal, are uncorrelated with other contrasts.

EEOS611

Slide 76 Testing a quadratic contrast

NOTES:

Results for a quadratic contrast

Is there any unimodal pattern in fish length *vs.* preference?

```
UNIANOVA
 prop  BY code
 /METHOD = SSTYPE(3)
 /INTERCEPT = INCLUDE
 /EMMEANS =
TABLES(OVERALL)
 /LMATRIX = "Linear contrast by
weight" code 5 -3 1 3 -9 3
 /LMATRIX = "Quadratic contrast
by weight" code 1214 -1318 -856 -
22 1004 -22
 /CRITERIA = ALPHA(.05)
 /DESIGN = code .
```

Contrast Results (K Matrix)[a]

Contrast		Dependent Variable: Proportion of time with yellow-sword male
L1	Contrast Estimate	-36.092
	Hypothesized Value	0
	Difference (Estimate - Hypothesized)	-36.092
	Std. Error	94.197
	Sig.	.703
	95% Confidence Interval for Difference — Lower Bound	-223.624
	95% Confidence Interval for Difference — Upper Bound	151.440

a. Based on the user-specified contrast coefficients (L') matrix: Quadratic contrast by weight

Test Results

Dependent Variable: Proportion of time with yellow-sword male

Source	Sum of Squares	df	Mean Square	F	Sig.
Contrast	.004	1	.004	.147	.703
Error	1.864	78	.024		

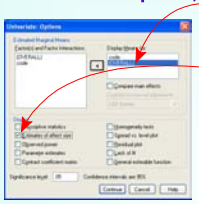
There is little if any evidence indicating a unimodal (quadratic) pattern between female tail preference and male body length (p=0.7, quadratic contrast test)

Slide 77 Results for a quadratic contrast

NOTES:

Estimates of effect size

Sleuth 2e p 152, The mean percentage is 62.4%...



Estimates of effect size, available in GLM/Univariate Needed to estimate effect across fish pairs in Case 6.2 (62.4 ± 3.4% preferred yellow tails)

2. Grand Mean

Dependent Variable: Proportion of time with yellow-sword male

		95% Confidence Interval	
Mean	Std. Error	Lower Bound	Upper Bound
.624	.017	.590	.658

Slide 78 Estimates of effect size

NOTES:

More on Simultaneous Inferences

Confidence limits

- **Individual** (pairwise) confidence level is the frequency with which a single interval captures its parameter.
- **Overall** (familywise or experiment-wise) confidence level is the frequency with which all intervals simultaneously capture their parameters.
- Planned *vs.* Unplanned comparisons

EEOS611

Slide 79 More on Simultaneous Inferences

NOTES:

Multiple comparisons (1 of 2)

Interval half width = Multiplier x Standard error

- LSD (Least Significant Difference): Student's *t* with pooled standard error — **no protection against multiple hypothesis testing**
- F-protected Inference
  - Fisher's protected Least Significant Difference
  - Don't claim a difference if the overall F statistic is not significant
- **Tukey-Kramer,** Studentized range Table A.5
  - Generalization of Tukey's HSD (Honestly Significant Difference) for unequal sample sizes)
  - Games-Howell more robust to unequal variance

EEOS611

Slide 80 Multiple comparisons (1 of 2)

NOTES:

Multiple comparisons (2 of 2)

Interval half width = Multiplier x Standard error

- **Bonferroni**, based on the number of comparisons (α/possible tests)
  - A conservative test (most often applied a posteriori test in drug trials for unplanned comparisons)
  - Test α = Experiment-wise α/k, where k is the number of tests
    - This approximation provides a remarkably accurate estimate of
    - Experiment-wise alpha: $\alpha_{exp} \approx 1-(1 - \alpha_{test})^k$, whre k is the number of tests
    - For example, 20 groups being tested 2 at a time
      - 20 Choose 2 tests = 190
      - Experiment-wise α = 1-(1- **0.05**)^{190}
      - Experiment-wise α = 0.99994
      - But 0.04877683466514 = 1-(1- **0.05/190**)^{190}
- **Scheffé**, based on the number of linear contrasts: most conservative of the widely used multiple comparison tests
- **Others** Sokal & Rohlf's Biometry, Quinn & Keough and Toothacker provide comprehensive listing
  - Newman-Keuls, SNK, *Student-Newman-Keuls*; based on studentized range, more powerful (less conservative) than Tukey-Kramer
  - Duncan's multiple range
  - **Dunnet's, where there is a control group**
  - Dunn's for non-parametric *a posteriori* contrasts

EEOS611

Slide 81 Multiple comparisons (2 of 2)

NOTES:

Display 6.6

Summary of 95% confidence interval procedures for differences between treatment means in the handicap study

Group	Average	Difference with ... hearing	amputee	control	wheelchair
crutches	5.921	1.871	1.492	1.021	0.578
wheelchair	5.343	1.293	0.914	0.443	
control	4.900	0.850	0.471		
amputee	4.429	0.379			
hearing	4.050				

Procedure	95% Interval Halfwidth
LSD	1.233
Tukey-Kramer	1.735
Bonferroni	1.794
Scheffé	1.957

A confidence interval is centered at a difference with halfwidth given by one of the procedures

Slide 82 Display 6.6

NOTES:

SPSS output from GLM

```
UNIANOVA
score  BY code
/METHOD = SSTYPE(3)
/INTERCEPT = INCLUDE
/LMATRIX = "Avg A H vs Avg C W" code 0 -1/2 1/2 -1/2 1/2
/POSTHOC = code ( TUKEY SCHEFFE LSD BONFERRONI )
/CRITERIA = ALPHA(.05)
/DESIGN = code .
```

Qualification

	Handicap	N	Subset 1	Subset 2
Tukey HSD[a,b]	Hearing	14	4.1	
	Amputee	14	4.4	4.4
	Control	14	4.9	4.9
	Wheelchair	14	5.3	5.3
	Crutches	14		5.9
	Sig.		.23	.12
Scheffe[a,b]	Hearing	14	4.1	
	Amputee	14	4.4	
	Control	14	4.9	
	Wheelchair	14	5.3	
	Crutches	14	5.9	
	Sig.		.07	

Means for groups in homogeneous subsets are displayed.
Based on Type III Sum of Squares
The error term is Mean Square(Error) = 2.666.
a. Uses Harmonic Mean Sample Size = 14.000.
b. Alpha = .05.

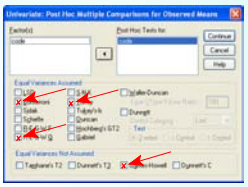
Slide 83 SPSS output from GLM

NOTES:

Multiple Comparisons available in SPSS

See Garson web site

- **Bonferroni: a conservative test (beware Type II error)**
- **Tukey-Kramer**
  - **In SPSS, if you ask for the Tukey test and sample sizes are unequal, you will get the Tukey-Kramer test, using the harmonic mean.**
- **Games-Howell, a modified Tukey-Kramer appropriate when the homogeneity of variances assumption is violated, controls for unequal sample sizes**
- **Ryan test (REGWQ): modified Newman-Keuls test**
  - **Toothaker (1993: 56) calls Ryan the "best choice" among tests supported by major statistical packages because maintains good alpha control (ex., better than Newman-Keuls) while having at least 75% of the power of the most powerful tests (ex., better than Tukey HSD).**



Slide 84 Multiple Comparisons available in SPSS

NOTES:

Quinn & Keough review of multiple comparison tests

- Use planned (*a priori*) contrasts whenever possible for testing specific differences among groups
- "If unplanned comparisons must be used, Ryan's REGW or Tukey's tests are recommended, the latter if simultaneous confidence intervals are required." (P. 207)
  - REGW: Ryan, Einot, Gabriel & Welch procedure.

EEOS611

Slide 85 Quinn & Keough review of multiple comparison tests

NOTES:

Ryan's test: REGW

From SPSS algorithms

Ryan, Einot, Gabriel, and Welsch (R-E-G-W) developed two multiple step-down range tests. Multiple step-down procedures first test whether all means are equal. If all means are not equal, subsets of means are tested for equality. R-E-G-W F is based on an F test and R-E-G-W Q is based on the Studentized range. These tests are more powerful than Duncan's multiple range test and **Student-Newman-Keuls** (which are also multiple step-down procedures), **but they are not recommended for unequal cell sizes.** <*emphasis added by Gallagher*>

Slide 86 Ryan's test: REGW

NOTES:

Display 6.7

2,436 mononucleotides along a DNA molecule. All 40 occurrences of the trinucleotide TGG appear in bold face. Eleven breaks occurred in the string, at the positions indicated by dashes.

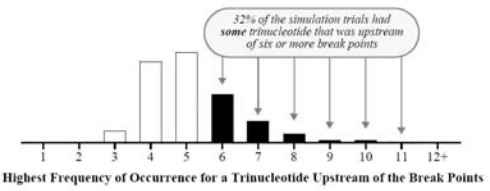
TGG before break in line 7; 6 of the 11 breaks occured 'downstream' of TGG, p= 0.000243

Slide 87 Display 6.7

NOTES:

Display 6.8

Simulated estimate of the distribution of the highest frequency of occurrence of a trinucleotide upstream of eleven randomly-selected breaks



Highest Frequency of Occurrence for a Trinucleotide Upstream of the Break Points

P=0.32 from a Monte Carlo simulation, but p=0.000243 from a test that didn't take into account the number of possible tests

Slide 88 Display 6.8

NOTES:

Conclusions to Chapter 6

1 of 4

- ANOVA is a subset of regression and both are subsets of general linear models
  - SPSS UNIANOVA is the standard GLM package in SPSS
  - GLM/UNIANOVA & regression have the greatest flexibility
- Linear contrasts: can be called through GLM or ANOVA
  - SPSS's Oneway only allows integer contrasts
    - With integer contrasts, p values are identical for any contrast vector multiplied by a scalar (effect sizes and standard errors increase proportionately)
  - With fractional contrasts in GLM/univariate, effect sizes & standard errors don't need to be rescaled
  - Matlab's orthpoly.m (statbox toolbox) solves orthogonal contrasts for any vector of explanatory variables
  - Can be used as an accepted alternative to regression when there is 'lack of fit' due to cluster effects
    - E.g., Boston Harbor regression of biodiversity vs. Year

EEOS611

Slide 89 Conclusions to Chapter 6

NOTES:

Conclusions to Chapter 6

2 of 4

- Planned and unplanned comparisons
  - Always try to specify hypotheses *a priori*, and use the LSD test (or equivalent linear contrast test) at a predetermined experiment-wise α level (usually 0.05)
  - Use the √(Within groups MS) as the pooled estimate of population *s* for these tests
- Unplanned comparisons
  - Also called: *ad hoc*, *a posteriori*, multiple comparison tests
  - Experiment-wise (family-wise) error levels usually set at 0.05

EEOS611

Slide 90 Conclusions to Chapter 6

NOTES:

Conclusions to Chapter 6

3 of 4

- Linear contrasts
  - E.g., Avg (A,B) vs. Avg (C,D,E)
  - Only Scheffé procedure should be used
- Bonferroni
  - A conservative test
  - Test α = Experiment-wise α/ k, where k is the number of tests
  - Experiment-wise $\alpha \approx 1-(1-\text{Test } \alpha)^k$
  - For example, 20 groups being tested 2 at a time
    - 20 Choose 2 tests = 190
    - Experiment-wise $\alpha = 1-(1-.05)^{190}$
    - Experiment-wise α = 0.99994
    - But $0.04877683466514 = 1-(1-\ 0.05/190\ )^{190}$

EEOS611

Slide 91 Conclusions to Chapter 6

NOTES:

Conclusions to Chapter 6

4 of 4

- Tukey-Kramer
  - Tukey's HSD ('Honestly significant difference') with adjustments for unequal sample sizes
  - Assumes equal variance (Games-Howell protects for unequal variance)
- Treatment vs. Control: use Dunnet's test (only n-1 comparisons, not ${}_nC_2$)
- More powerful tests
  - SNK: recommended by Underwood
  - Ryan's test (REGWF), recommended by Quinn & Keough (in addition to Tukey-Kramer)
    - **Ryan's test not suitable for unequal group sizes (SPSS)**

Slide 92 Conclusions to Chapter 6

NOTES: