Tutorial Problems

Question 1

The following data give the arc sine transformations (Y_{ij}) of the percentage survival of the beetle Tribolium castaneum at 4 densities (X = number of eggs per gram of flour medium).

	X (densities)			
	5/g	20/g	50/g	100/g
	61.68	68.21	58.69	53.13
	58.37	66.72	58.37	49.89
	69.30	63.44	58.37	49.82
	61.68	60.84		
	69.30			
Totals	320.33	259.21	175.43	152.84

The total sum of squares for these data is 562.3883.

(a) Find the treatment sum of squares and hence test for differences in the percentage of beetles surviving. (Lecture 14)

(b) Find an appropriate contrast to measure the linear effect of X. Hence find the linear component of the treatment sum of squares and test the hypothesis that Y can be adequately modelled by a linear regression on x. (Lecture 17)

```r
qf(0.95, 3, 11)
```

```
## [1] 3.587434
```

```r
qt(c(0.9, 0.95, 0.975), 15)
```

```
## [1] 1.340606 1.753050 2.131450
```

```r
qt(c(0.9, 0.95, 0.975), 14)
```

```
## [1] 1.345030 1.761310 2.144787
```

```r
qt(c(0.9, 0.95, 0.975), 13)
```

```
## [1] 1.350171 1.770933 2.160369
```

```r
qt(c(0.9, 0.95, 0.975), 12)
```

```
## [1] 1.356217 1.782288 2.178813
```



```r
qt(c(0.9, 0.95, 0.975), 11)
```

```
## [1] 1.363430 1.795885 2.200985
```

Question 2

In order to compare four different brands of golf balls, five balls from each brand are placed in a machine that exerts the force produced by a 200 metre drive. The number of simulated drives needed to crack or chip each ball is recorded below.

A	B	C	D
281	270	218	364
220	334	244	302
274	307	225	325
242	290	273	337
251	331	249	355

An analysis of variance for the model $Y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \epsilon_{ij} \sim N(0, \sigma^2)$, where Y_{ij} denotes the j-th observation on brand i has ANOVA table:

Source of Variation	df	SS	Mean Square	F
Brands	3	29860.4	9953.467	16.42
Residual	16	9698.4	606.15	
Total	19	39558.8		

(a) Calculate 90% Tukey, Bonferroni, and Scheffe simultaneous confidence intervals for the maximal brand effect, adjusting for all pairwise comparisons and compare to the 90% CI for the difference in treatment effect of C and D by measuring how much percentage wider the adjusted CIs are. (Lecture 16)

(b) Decompose the brands sum of squares into three orthogonal components including one which measures the difference between brands A and C, and one which measures the difference between brands B and D. Specify the orthogonal contrasts that correspond to the three 1 d.f. components. (Lecture 15)

```r
qtukey(c(0.9, 0.95, 0.975), 4, 16)
```

```
## [1] 3.520076 4.046093 4.547630
```

```r
qt(c(0.9, 0.95, 0.975), 16)
```

```
## [1] 1.336757 1.745884 2.119905
```

```r
qf(c(0.9, 0.95, 0.975), 3, 16)
```

```
## [1] 2.461811 3.238872 4.076823
```

Computer Problems

Question 1

This question is based on the data collected from the 78 bluegills captured from Lake Mary, Minnesota in 1981. The data is found in Weisberg (1986) A linear model approach to backcalculation of fish length, *Journal of the American Statistical Association*, 81(196) 922-929 or in `bluegills.txt`. The data frame contains two columns: the length (in mm) and the age (in years) of the fish at capture. You will be investigating how the length of a bluegill fish is related to its age.

(a) Plot a scatter plot of age as predictor and length as response with a straight line corresponding to the best line of fit using least squares estimates.

(b) Does the simple linear regression seem appropriate? Explain your conclusion.

(c) Take age as a factor and construct a one-way ANOVA table.

(d) Fit a polynomial regression model that is equivalent to the above ANOVA model.

(e) Test if the polynomial regression model can be simplified. What model have the most appropriate fit? Explain your choice.

