STAT2912

Semester 2	Statistical Tests (Advanced)	2008

Assignment 2

Due Wednesday 15 October, 2008

1. The following are the lifetimes, in hours, of random samples of two kinds of light bulbs in continuous use.

Brand 1:	407	453	378	434	396	441	373	393	386	418
Brand 2:	403	424	383	445	439	417	412	462	432	433

   Calculate an approximate 99% confidence interval, based on the Mann-Whitney statistic, for the average difference in the lifetimes of the two kinds of light bulbs. (Assume that a shift model is appropriate.)

2. Berger and Walker (1972) monitored the heart rates for six tree shrews (*Tupaia glis*) during three different stages of sleep: LSWS (light slow-wave sleep); DSWS (deep slow-wave sleep) and REM (rapid eye movement sleep).

Tree Shrew	LSWS	DSWS	REM
1	14.1	11.7	15.7
2	26.0	21.1	21.5
3	20.9	19.7	18.3
4	19.0	18.2	17.0
5	26.1	23.2	22.5
6	20.5	20.7	18.9

   (a) Construct the two way analysis of variance table based on these data.

   (b) Test for equality of the average heart rates during the three phases of sleep. State your underlying model.

   (c) Calculate a 95% confidence interval to compare the average REM heart rate against the average of LSWS heart rate.

   (d) Test the hypothesis in (b) using the Friedman test statistic. Does this test produce the same conclusions as the one in (b)?

3. In an experiment designed to measure the effects of different dosages of the drug Stilbestrol on the uterine weights of immature rats, 35 rats were divided into 5 groups with 7 rats in each group. The groups were treated with different doses of the drug. The logarithms of the weights are used in the analysis since it was found that this transformation has the effect of producing data with approximately equal variance in each class corresponding to a different dose level. Further the logarithm of dosage is given since it is found that the relationship of dose to uterine weight is approximately linear when both variables are measured on a logarithmic scale. The data are from Lee et al. (1942). Let $y = \log(\text{weight}) - 1.4$ and $x = \log(\text{dose})$. The data are stored in the statistics data library on R as **ratweights**.

(a) Does it matter whether logarithms to base 10 or base e are used?

(b) Treating the different dosages as different treatments use an analysis of variance to test for a dose effect.

(c) Estimate the error variance using the ANOVA model.

(d) This problem can also be analysed via a regression model. Fit the model

$$Y = \alpha + \beta x + \varepsilon,$$

where $\varepsilon \sim N(0, \sigma^2)$.

(e) Produce appropriate residual plots to check the regression model assumptions.

(f) Which model, ANOVA or linear regression produces the smaller error variance estimate?

(g) Use the results from (b) to calculate an approximate 90% confidence interval for the mean uterine weight (not log weight) of rats receiving a log dosage of 0.45.

4. The following output is part of the analysis of data from 'Evidence for and rate of denitrification in the Arabian Sea' (*Deep Sea Research* (1978), pp 431-435). The variables are x, which denotes salinity level (%), and y, which denotes nitrate level ($\mu M/L$).

Source of Variation	df	Sum of Squares
Regression	1	64.49622
Residual	6	6.11253

The fitted regression line is

$$Y = 326.976 - 8.404x.$$

(a) What was the sample size?

(b) Use the above information to calculate the coefficient of correlation between salinity and nitrate level.

(c) Test the significance of the regression.

(d) Calculate a 95% confidence interval for the slope of the regression line.

(e) Calculate a 90% confidence interval for the standard deviation of nitrate level.

5. Birth weight (in kg) and chest size (in cm) at birth were recorded for 9 randomly chosen infants. The correlation coefficient for the sample was $r = 0.73$. Assuming that birth weight and chest size follow a bivariate normal distribution calculate a 95% confidence interval for the population coefficient of correlation, ρ.

6. Consider the two way analysis of variance F-test statistic when we have r blocks but only 2 treatments in each block. Let the observations be X_{ij}, $i = 1, 2, ..r$, $\quad j = 1, 2$.

   (a) Show that the Treatment Sum of Squares is $r(\bar{X}_{\cdot 1} - \bar{X}_{\cdot 2})^2/2$.

   (b) In each block let $D_i = X_{i1} - X_{i2}$. Prove that the Residual Sum of Squares is $(r-1)s_D^2/2$, where s_D^2 is the sample variance of the r differences.

   (c) Hence deduce that the paired t test is equivalent to the 2 way analysis of variance F-test for testing the null hypothesis of no treatment effect against a general two sided alternative.