Semester 2	Statistical Tests (Advanced)	2008

Assignment 2 Solutions

1. The sample sizes are $n_1 = n_2 = 10$. We need to find k such that

$$P(k \leq U \leq 100 - k) \simeq 0.99$$

Using the normal approximation to the distribution of U we get

$$k = 100/2 - 2.576 \times \sqrt{\frac{10 \times 10 \times 21}{12}} = 15.92.$$

Thus if $D_{(i)}$ denotes the ith ordered difference $(x_i - y_i)$, an appoximate 99% CI is $[D_{(16)}, D_{(85)}]$.

By evaluating the differences we get the CI is $[-53,\ 20]$.

2. (a)
```
> sleep.aov=aov(data~sleep+shrew)
> summary(sleep.aov)
            Df  Sum Sq Mean Sq F value   Pr(>F)
sleep        2  16.988   8.494  4.1288  0.04929 *
shrew        5 195.436  39.087 19.0000 8.12e-05 ***
Residuals   10  20.572   2.057
```

(b) The p-value for testing for equal heart rates in the 3 different sleep phases is $p(F_{2,10} \geq 4.129) = 0.049$ Thus there is evidence of a difference in the means.

Let X_{ij} denote the heart rate of the ith shrew in the jth stage of sleep. Model:

$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij},$$

where $\varepsilon_{ij} \sim NID(0, \sigma^2)$.

(c) A 95% CI is

$$(113.9/6 - 126.6/6) \pm 2.228 \times \sqrt{2.057 \times 2/6} = -2.1167 \pm 1.845.$$

(d)
```
> friedman.test(data,sleep,shrew)

Friedman rank sum test

data:  data, sleep and shrew
Friedman chi-squared = 4.3333, df = 2, p-value = 0.1146
```

The Friedman test would not reject the equal mean hypothesis so the conclusion is not the same as in (b).

3. (a) It does not matter which log base is used. Changing from base 10 to base e is only a matter of scale.

(b)
```
> attach(ratweights)
> x1=x
> x1=factor(x1)
> rat.aov=aov(y~x1)
> summary(rat.aov)
            Df   Sum Sq  Mean Sq F value    Pr(>F)
x1           4 0.300182 0.075045  22.034 1.436e-08 ***
Residuals   30 0.102179 0.003406
---
```

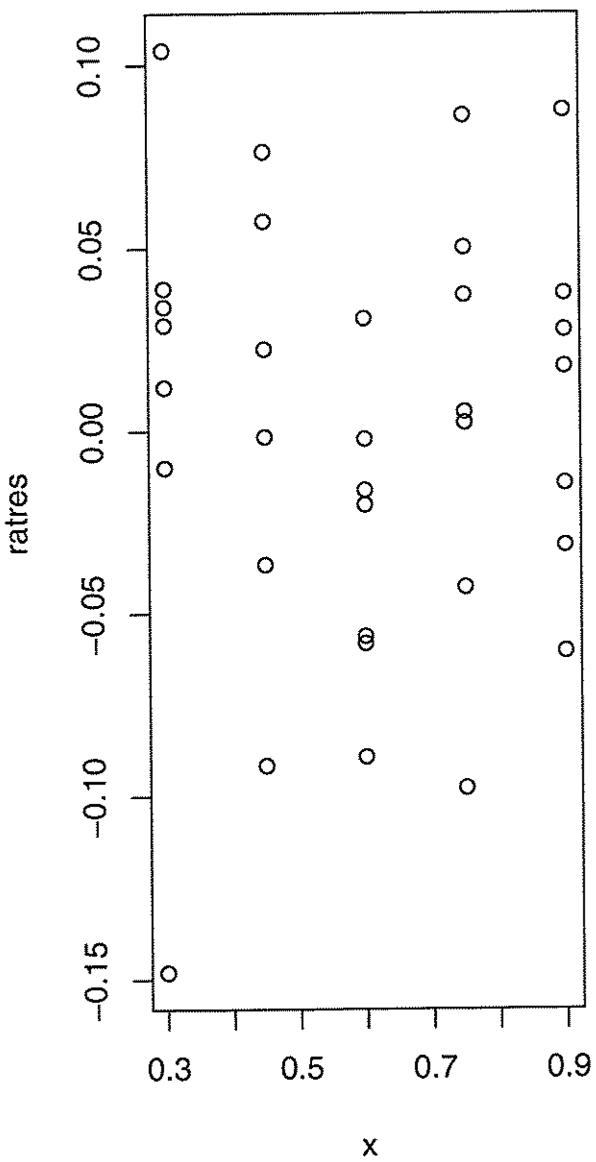
Since the p-value is so small we have evidence to reject the equal mean weight claim.

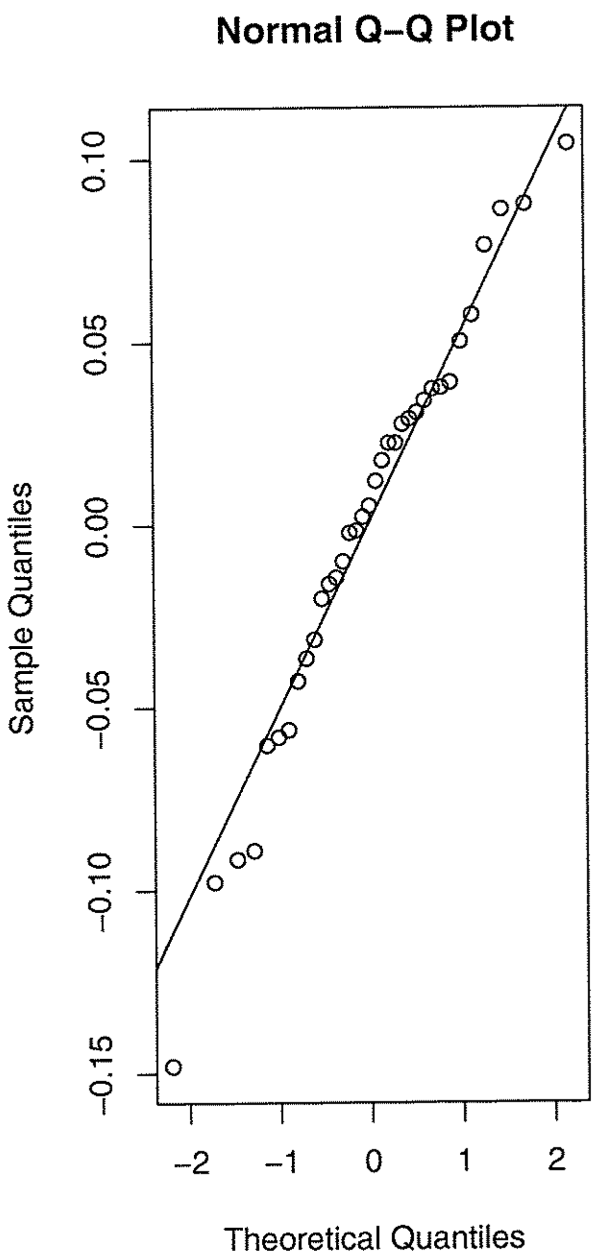
(c) The error variance is estimated to be 0.003406.

(d)
```
> rat.lm=lm(y~x)
> summary(rat.lm)

Call:
lm(formula = y ~ x)

Residuals:
      Min        1Q    Median        3Q       Max
-0.148057 -0.034057  0.005143  0.035543  0.103943

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.01914    0.02931  -0.653    0.518
x            0.43067    0.04605   9.351 8.45e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0578 on 33 degrees of freedom
> ratres=rat.ls$resid
> par(mfrow=c(1,2))
> plot(x,ratres)
> qqnorm(ratres)
> qqline(ratres)
> sum(ratres^2)/33
 [1] 0.003340590
```

(e) The residuals ploted against x do not show any pattern and the residual normal QQ plot is roughly linear supporting the normal error assumption.

(f) The error variance estimate from the linear regression model fit is 0.00334, slightly


ratres
0.10
0.05
0.00
−0.05
−0.10
−0.15
0.3
0.5
0.7
0.9
x


Normal Q–Q Plot
Sample Quantiles
0.10
0.05
0.00
−0.05
−0.10
−0.15
−2
−1
0
1
2
Theoretical Quantiles

smaller than from the ANOVA model.

(g)
```
> t=qt(0.95,30)
> t
[1] 1.697261
> s=sqrt(deviance(rat.aov)/30)
> s
[1] 0.05836062
> m=mean(y[8:14])
> m
[1] 0.1815714
> b=c(m-t*s/sqrt(7), m+t*s/sqrt(7))
> b
[1] 0.1441328 0.2190100
> c(exp(b[1]+1.4), exp(b[2]+1.4))
[1] 4.683908 5.048090
```
The CI is (4.684, 5.048).

4. (a) The sample size is 8.

(b) If the correlation coefficient is r then

$$r^2 = \frac{64.49622}{64.49622 + 6.11253} = 0.91343.$$

Since r has the same sign as the slope, $r = -0.9557$.

(c) Model: $Y_i = \alpha + \beta x_i + \varepsilon_i, \quad \varepsilon_i \sim NID(0, \sigma^2)$. A test for the significance of the regression is a test of $H_0 : \beta = 0$. The F-test statistic is $f = 64.49622/(6.11253/6) = 63.3$. The p-value is $P(F_{1,6} \geq 63.3)$ which is very small. Thus the slope coefficient cannot be set to 0.

(d) A 95% CI for β is

$$-8.404 \pm t_6(0.975) \times \frac{\hat{\sigma}}{\sqrt{S_{xx}}} = -8.4004 \pm 2.447 \times \sqrt{\frac{6.11253/6}{0.91319}}$$

as the regression SS is $\hat{\beta}^2 S_{xx}$, so the CI is

$$-8.404 \pm 2.585.$$

(e) A 90% CI for σ:

$$0.90 = P(1.635 < \chi^2_6 < 12.592)$$

so the CI is

$$(\sqrt{6.11253/12.592}, \sqrt{6.11253/1.635}) = (0.6967,\ 1.9335).$$

5. The sample size is $n = 9$ and $r = 0.73$. A 95% CI for $\tanh^{-1}(\rho)$ is

$$\tanh^{-1}(0.73) \pm 1.96 \times \frac{1}{\sqrt{6}} = (0.12855,\ 1.72889).$$

A 95% CI for ρ is obtained by transforming the endpoints with the tanh function. The CI is (0.1278, 0.9389).

6. (a) The Treatment SS is $r\sum_{j=1}^{2}(\bar{X}_{\cdot j} - \bar{X}_{\cdot\cdot})^2$. Now the overall mean is $\bar{X}_{\cdot\cdot} = (\bar{X}_{\cdot 1} + \bar{X}_{\cdot 2})/2$ so

$$\bar{X}_{\cdot 1} - \bar{X}_{\cdot\cdot} = (\bar{X}_{\cdot 1} - \bar{X}_{\cdot 2})/2$$

and

$$\bar{X}_{\cdot 2} - \bar{X}_{\cdot\cdot} = (\bar{X}_{\cdot 2} - \bar{X}_{\cdot 1})/2$$

Thus the Treatment SS is

$$r\left(\frac{(\bar{X}_{\cdot 1} - \bar{X}_{\cdot 2})^2}{4} + \frac{(\bar{X}_{\cdot 2} - \bar{X}_{\cdot 1})^2}{4}\right) = \frac{r}{2}(\bar{X}_{\cdot 1} - \bar{X}_{\cdot 2})^2.$$

(b) The residual SS is

$$\sum_{i=1}^{r}\sum_{j=1}^{2}(X_{ij} - \bar{X}_{i\cdot} - \bar{X}_{\cdot j} + \bar{X}_{\cdot\cdot})^2.$$

Note

$$\begin{aligned}(X_{i1} - \bar{X}_{i\cdot} - \bar{X}_{\cdot 1} + \bar{X}_{\cdot\cdot})^2 &= \left(X_{i1} - \frac{X_{i1} + X_{i2}}{2} - \bar{X}_{\cdot 1} + \frac{\bar{X}_{\cdot 1} + \bar{X}_{\cdot 2}}{2}\right)^2 \\ &= \left(\frac{X_{i1} - X_{i2}}{2} - \bar{X}_{\cdot 1} - \bar{X}_{\cdot 2}\right)^2\end{aligned}$$

and a similar expression holds for $j = 2$. Thus the Residual SS is

$$\frac{1}{2}\sum_{i=1}^{r}(D_i - \bar{D})^2 = \frac{(r-1)s_D^2}{2}.$$

(c) The 2 way ANOVA F-test statistic is

$$\begin{aligned} f &= \frac{\text{Treatment SS}/(2-1)}{\text{Residual SS}/((2-1)(r-1))} \\ &= \frac{r(\bar{X}_{\cdot 1} - \bar{X}_{\cdot 2})^2/2}{(r-1)s_D^2/(2(r-1))} \\ &= \frac{r\bar{D}^2}{s_D^2} \\ &= t^2, \end{aligned}$$

where t is the paired t-test statistic. Thus the F statistic is just the square of the paired t-test statistic when $s = 2$.