Homework #3 6381, due 6 Feb 06

1. The complete set of the data for Welsh's Example 5 (corrosion of steel plates) is below:

```
y <- scan()
68 76 81 96 92 91 115 133 124 90 209 387 100 106 764 148 394 130 76 77 83
88 96 157 265 123 175 100 95 84 125 122 144 222 195 159 413 86 100 134 174 146

x1time <- scan()
4 4 4 4 4 4 4 4 4
6 6 6 6 6 6 6 6 6
8 8 8 8 8 8 8 8 8
10 10 10 10 10 10 10 10 10
12 12 12 12 12 12

x2temp <- scan()
160 160 160 180 180 180 200 200 200
160 160 160 180 180 180 200 200 200
160 160 160 180 180 180 200 200 200
160 160 160 180 180 180 200 200 200
160 160 160 180 180 180

> tapply(y,x1time,mean)
       4         6         8        10        12
   97.33   258.67    126.67    138.44    175.50

> tapply(y,x2temp,mean)
     160      180      200
  135.00   162.33   181.92

> round( tapply(y,list(x1time,x2temp),mean), 2)
```

```
        160    180    200
  4   75.00  93.00 124.00
  6  228.67 323.33 224.00
  8   78.67 113.67 187.67
  10  93.00 130.33 192.00
  12 199.67 151.33     NA
```

```
> round( sqrt(tapply(y,list(x1time,x2temp),var)), 2)
      160    180    200
4    6.56   2.65   9.00
6  149.47 381.64 147.50
8    3.79  37.74  71.84
10   8.19  11.93  31.61
12 184.88  20.53     NA
```

These within-SDs are varying a LOT (382/2.65=144). Consider instead log(y):

```
yy <- log(y)
```

```
> round( tapply(yy,x1time,mean), 2)
     4    6    8   10   12
  4.56 5.28 4.75 4.88 5.02
```

```
> round( tapply(yy,x2temp,mean), 2)
   160  180  200
  4.70 4.88 5.13
```

```
> round( tapply(yy,list(x1time,x2temp),mean), 2)
       160  180  200
   4  4.31 4.53 4.82
```

```
 6  5.27 5.30 5.28
 8  4.36 4.70 5.19
 10 4.53 4.87 5.25
 12 5.03 5.01   NA
```

```
> round( sqrt(tapply(yy,list(x1time,x2temp),var)), 2)

     160  180  200
 4   0.09 0.03 0.07
 6   0.73 1.16 0.61
 8   0.05 0.31 0.38
 10  0.09 0.09 0.17
 12  0.87 0.13   NA
```

These within-SDs still vary (1.16/.03 = 39) but not as much; typical SD   0.5. Use logs. Fill in values for the NA cell using rnorm(3, 5, sd=.5): 5.295, 4.730, 5.4057. These values correspond to y of approximately 199, 113, 223.

```
newy <- c(y,119,113,223)
newyy <- log(newy)

newx1 <- scan()
4 4 4 4 4 4 4 4 4
6 6 6 6 6 6 6 6 6
8 8 8 8 8 8 8 8 8
10 10 10 10 10 10 10 10 10
12 12 12 12 12 12 12 12 12

newx2 <- scan()
```

```
160 160 160 180 180 180 200 200 200
160 160 160 180 180 180 200 200 200
160 160 160 180 180 180 200 200 200
160 160 160 180 180 180 200 200 200
160 160 160 180 180 180 200 200 200
```

Decompose the sums of squares for the variable "time" (5 levels) and "temp" (3 levels) into their linear, quadratic, ..., components. The coefficients can be found from R using contr.poly(3) and contr.poly(5).

2. What is an 95% upper confidence limit for p, the proportion of "failures" in n bernouilli trials, when the observed number of failures is zero? Do the same when the observed number of failures is 1. Can you find an approximation to this upper 95% bound as a function of n, both when X=0 and X=1? Hint: do not use CI with allowance based on normal approximation $(\hat{p}(1-\hat{p})/\sqrt{n})$; it is unreliable when X is so small. Check how unreliable it will be for various values on n!

3. Prove: If $f_T(t;\theta) = g(Z(t;\theta)) \cdot |\partial/\partial t Z(t;\theta)|$ for some function g and some monotone function Z (monotone in t for every θ, then $Z(t;\theta)$ is pivotal; i.e., its distribution is independent of θ.

4. Suppose you have a Gaussian sample of size n and you want a confidence interval of σ^2 based on the usual chi-squared interval and the sample variance s^2. How large a sample size do you need to be confident of the value of σ^2 to one decimal point? (Same question, for the value of σ to one decimal point.) Answer for 95% and 99%, when s^2 is 1.0, 2.0, ..., 9.0, and when s is 1.0, 2.0, ..., 9.0.