# Assignment 6 - Goodness of Fit and Analysis of Variance

Math 363 - December, 2009

1. Below are the smoking habits of 5375 high school children in Tucson in 1967.

| | student smokes | student does not smoke | total |
|---|---|---|---|
| 2 parents smoke | 400 | 1380 | 1780 |
| 1 parent smokes | 416 | 1823 | 2239 |
| 0 parents smoke | 188 | 1168 | 1356 |
| total | 1004 | 4371 | 5375 |

   Perform a chi square test on this contingency, both by hand and using software and report the $p$-values and the cells that contribute the most to chi square statistic.

2. Ladislaus Bortkiewicz collected data on death of soldiers in the Prussian army from kicks by horses and mules. The studied 10 army corps, each observed over 20 years. In 109 corps-years, no deaths occurred; 65 corps-years had one death, etc.

| deaths/corp/year | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| occurances | 109 | 65 | 22 | 3 | 1 |

   (a) Find the mean for these data.

   (b) Give the expected number of occurances of 0, 1, 2, 3, and 4 or more deaths with a Poisson distribution having the mean above.

   (c) Perform a chi-square goodness of fit test and report your findings.

3. Parents are frequently concerned when their child seems slow to begin walking. In 1972, Science reported on an experiment in which the effects of several different treatments on the age at which a childs first walks were compared. Children in the first group were given special walking exercises for 12 minutes daily beginning at the age 1 week and lasting 7 weeks. The second group of children received daily exercises, but not the walking exercises administered to the first group. The third and forth groups received no special treatment and differed only in that the third groups progress was checked weekly and the forth was checked only at the end of the study

| 1 | 2 | 3 | 4 |
|---|---|---|---|
| 9 | 11 | 11.5 | 13.25 |
| 9.5 | 10 | 12 | 11.5 |
| 9.75 | 10 | 9 | 12 |
| 10 | 11.75 | 11.5 | 13.5 |
| 13 | 10.5 | 13.25 | 11.5 |
| 9.5 | 15 | 13 | |

(a) Enter the data into R

```
> treatment<-c(rep(1,6),rep(2,6),rep(3,6),rep(4,5))
>age<-
c(9,9.5,9.75,10,13,9.5,11,10,10,11.75,10.5,15,11.5,12,9,11.5,13.25,13,13.25,11.5,12,13.5,11.5)
```

(b) Make side by side boxplots. (Check the `formula` option in the R help for `boxplot`.) and describe any differences you see.

(c) Perform a one way analysis of variance. Note that you will have to turn the treatment variable into a factor variable by using `ftreatment<-factor(treatment,c(1:4))`

(d) Find a 90% confidence interval for the mean difference between treatments 1 and 4.