Statistics Midterm Spring 2008

1. Consider a study of the savings rates of various countries. The first five members of the sample are listed below:

   > Save.data <- read.table("savings.data",header=TRUE)
   > Save.data[1:5,]

	Pop15	Pop75	DispInc	Growth	Savings
Australia	29.35	2.87	2329.68	2.87	11.43
Austria	23.32	4.41	1507.99	3.93	12.07
Belgium	23.80	4.43	2108.47	3.82	13.17
Bolivia	41.89	1.67	189.13	0.22	5.75
Brazil	42.19	0.83	728.47	4.56	12.88

   Here *Savings* is the population savings rate, *Pop15* is the percent population under the age of 15, *Pop75* is the percent population over the age of 75, *DispInc* is disposable income, and *Growth* is the population growth rate.

   > attach(Save.data)
   > mod.1 <- lm(Savings ~ Pop15 + Pop75 + DispInc + Growth)
   > summary(mod.1)

   Coefficients:

	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	28.5666100	7.3544986	3.884	0.000334
Pop15	-0.4612050	0.1446425	-3.189	0.002602
Pop75	-1.6915757	1.0835862	-1.561	0.125508
DispInc	-0.0003368	0.0009311	-0.362	0.719296
Growth	0.4096998	0.1961961	2.088	0.042468

   (a) What are the null and alternative hypotheses associated to the p-value .042468?

   (b) What are the models implicitly being compared by the test above?

   (c) Estimate a 95% confidence interval for the Pop15 slope.

(d) Based on this model (mod.1) what would be the effect of a one point increase in population growth on the savings rate (assume Pop15, Pop75, and DispInc are unchanged)?

(e) The next lines of output are:

Residual standard error: 3.803 on yyy degrees of freedom
Multiple R-Squared: 0.3385, Adjusted R-Squared: 0.2797
F-statistic: 5.756 on 4 and yyy DF, p-value: 0.0007902

if the data set had 50 elements, what is yyy and how do you know?

(f) What are the null and alternative hypotheses associated with the p-value 0.0007902?

(g) What are the models implicitly being compared by the test above?

(h) What percentage of the variability in Savings is explained by the model?

(i) Would any confidence intervals produced by the command

> confint(mod.1)

contain 0? If so, which ones and why? If none, why not?

2. Consider a study of the relationship between age and balancing ability.

```
> Balance <- read.table("balance.data",header=TRUE)
> attach(Balance)
> Balance[1:5,]
  movement age.group
1       19   elderly
2       30   elderly
3       20   elderly
4       19   elderly
5       29   elderly
```

Here movement is a measure of the "wobbling" effect when the subject is thrown off balance, so lower numbers indicate better balance. The levels for age.group are "elderly" and "young".

```
> t.test(movement[age.group=="elderly"], movement[age.group=="young"])

        Welch Two Sample t-test

data: movement[age.group == "elderly"] and movement[age.group == "young"]
t = 2.3035, df = 10.971, p-value = yyyyyyy
alternative hypothesis: xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx
95 percent confidence interval:
0.3627401 16.0539266
sample estimates:
mean of x mean of y
26.33333 18.12500
```

(a) What are the null and alternative hypotheses associated with this test?

(b) What can you say about yyyyyyy, the p-value for this test

(c) Had this p-value above (yyyyyyy) been .046 (it wasn't), what conclusion would you draw from this output?

(d) Had the p-value above (yyyyyyy) been .046 (it wasn't), what p-value would be associated with the following test?

> t.test(movement[age.group==“elderly”],movement[age.group==“young”], alternative='greater')