Professor's Notes About Linear Regression Homework 2

- When discussing whether there is a relationship between the response and the explanatory variable you must explicitly note that you are referring to the slope p-value. You cannot just refer the reader to the "p-value" in the table from `summary()` because there are three p-values in that table. Be precise with your language!!
- Remember to use CI when describing rates of change (i.e., slopes) or predictions; don't just use the best estimate.

Male-Female Birth Ratio

1. Yes there is evidence for a significant statistical change in the proportion of male births over the study period because the p-value for the slope (and the equivalent overall F p-value) is very small ($p < 0.0005$; Table 1).

   Table 1. Summary of the linear regression of proportion of males on year.

   ```
                 Estimate Std. Error t value Pr(>|t|)
   (Intercept)  6.201e-01  1.860e-02  33.340  < 2e-16
   year        -5.429e-05  9.393e-06  -5.779 1.44e-05
   ---
   Residual standard error: 0.0002607 on 19 degrees of freedom
   Multiple R-squared: 0.6374,Adjusted R-squared: 0.6183
   F-statistic:  33.4 on 1 and 19 DF,  p-value: 1.439e-05
   ```

2. The proportion of males **declined** between 0.000035 and 0.000074 per year, on average (Table 2).

   Table 2. Confidence intervals for coefficients of the linear regression of proportion of males on year.

   ```
                       2.5 %        97.5 %
   (Intercept)  5.811580e-01  6.590134e-01
   year        -7.394606e-05 -3.462537e-05
   ```

3. The very small slope coefficient is statistically different from zero because the SE for the slope coefficient is very small (0.000009; Table 1) and the overall scale of the measurements is very small.

Ashland Climate

1. Yes, there is a signficant relationship between average precipitation and average high temperature for Ashland, WI ($p < 0.0005$; Table 3). Specifically, as the average high temperature increases by 1°C the mean average precipitation increases between 2.4 and 3.3 mm (Table 4).

   Table 3. Summary of the linear regression results of average monthly preciptation on average high temperature for Ashland, WI.

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  32.0887     3.1782   10.10 1.46e-06
ahi           2.8676     0.2053   13.97 6.91e-08
---
Residual standard error: 7.842 on 10 degrees of freedom
Multiple R-squared: 0.9513,Adjusted R-squared: 0.9464
F-statistic: 195.2 on 1 and 10 DF,  p-value: 6.913e-08
```

   Table 4. Confidence intervals for coefficients of average monthly preciptation on average high temperature for Ashland, WI.

```
                2.5 %    97.5 %
(Intercept) 25.007285 39.170170
ahi          2.410234  3.324969
```

2. The y-intercept value suggests that the mean average monthly preciptation for all months with an average high temperature of 0°C is between 25.0 and 39.2 mm.
3. The mean average preciptation for all months with an average high of 10°C is between 55.7 and 65.8 mm.
4. The mean average preciptation for a month with an average high of 10°C is between 42.6 and 79.0 mm.
5. The prediction interval for the individual (question 4) is wider than the confidence interval for the mean (question 3) because there is more variability in predicting an individual as compared to a mean. Variability for predicting an individual includes both sampling and natural variability, whereas variability for the mean includes only sampling variability.

R Appendix

```r
library(NCStats)
setwd("c:/biometry/")
# First Question
d1 <- read.csv("MaleBirths.csv")
lm1 <- lm(propmale~year,data=d1)
summary(lm1)
confint(lm1)
# Second Question
d2 <- read.csv("Ashland.csv")
lm2 <- lm(aprecip~ahi,data=d2)
summary(lm2)
confint(lm2)
predict(lm2,data.frame(ahi=10),interval="confidence")
predict(lm2,data.frame(ahi=10),interval="prediction")
```