Professor Notes About the "Bivariate EDA - Quant" Homework

- You must provide labeled tables and figures to support your results and refer to these tables in your sentences.
- Do not use the word "correlation" unless you are specifically referring to "r." For example, you would NOT say "the correlation between longevity and gestation is ..." In this case, it is better to replace the word "correlation" with "relationship."
- You must explicitly state where the outliers are located. In this case, it is adequate to note that they are "in the upper-left" corner of the plot. Alternatively, you could note the approximate coordinates of the points.
- Note the sentences that explicitly state whether the correlation coefficient (r) could be used to assess strength or not.
- It is correct to not calculate or report the correlation coefficient because of the presence of outliers. However, you still need to comment on the strength of the relationship. Your comment will be more subjective based on your interpretation of the clustering of the points but it still needs to be made.
- Make sure to use `xlab=` and `ylab=` to provide better labels for the x- and y-axes, respectively, on your scatterplots.
- Note that your scatterplots (and all plots) should be square or only slightly wider than tall (no more than a ratio of 4:3 width:height).

Animal Fat and Breast Cancer

The relationship between age-adjusted death rate and animal fat intake is positive, linear, absent of outliers, and very strong (r=0.949; Figure on homework handout). It was valid to assess strength with the correlation coefficient because of the linear form and lack of outliers.

Animal Longevity and Gestation

The relationship between longevity and gestation for captive anaimls is mostly positive, nonlinear, weak, with one outlier present at an approximate longevity of 40 years and gestation period of 600 days (Figure 1). I did not compute a correlation coefficient because of the nonlinear form and presence of an outlier.
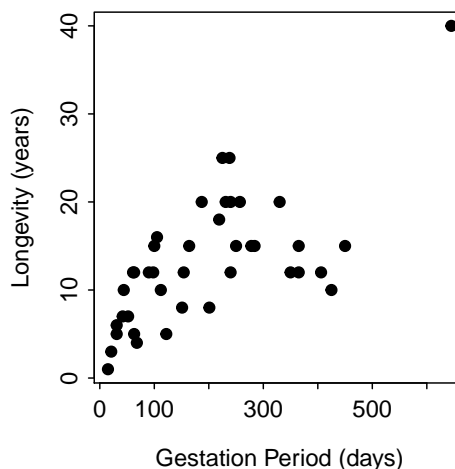

Figure 1. Scatterplot of longevity versus gestation period for a sample of 40 animals.

R Appendix

```
library(NCStats)
setwd('C:/aaaWork/Books/IntroStats/HW/')
```

```
d <- read.csv("animals.csv")
plot(longevity~gestation,data=d,pch=19,ylab="Longevity (years)",
     xlab="Gestation Period (days)")
```