Professor Notes about "Linear Regression" Homework

- The equation of the line should not contain a "Y" or an "X". Rather it should contain "Distance" and "Age".
- Note the use of "on average" in the "interpret slope" and "interpret intercept" questions.
- The "extrapolation" question was realized by looking at the scale on the fitted line plot in (Figure 1). Thus, you must include (and refer to) the fitted-line plot.
- The "concerns," if any, must related to linearity and homoscedasticity. Thus, you must include (and refer to) the fitted-line plot as evidence for your judgment of these two assumptions.
- The correlation coefficient is negative in this case because the association is negative. You must remember to put a negative sign on the square root of the r^2 value if the association is negative.

Sign Legibility and Age

1. The response variable is maximum distance to see the sign. [*Note that distance to see the sign depends on age and distance to see the sign is what is predicted or explained in the ensuing questions.*]
2. The explanatory variable is age.
3. The best-fit line is $distance = -3.01 * age + 576.7$.
4. The slope indicates that for every increase of one year in age the maximum distance to see the sign will decrease by 3.01 ft, on average.
5. The intercept indicates that if the age of the driver was zero, then the maximum distance to see the sign would be 576.7 ft, on average.
6. The predicted maximum distance to see the sign for a 40-year-old driver is 456.4 ft.
7. This prediction should not be made as an age of 90 is outside the observed results for this variable (Figure 1) and is, thus, an extrapolation.
8. The residual for an individual that is 50-years-old and had a maximum distance to see the sign of 410 ft is -16.3 ft. Thus, this individual would have a shorter maximum distance to see the sign than an average 50-year-old.
9. The correlation coefficient between age of the driver and maximum distance to see the sign is -0.80.
10. The proportion of variability in maximum distance to see the sign that is explained by knowing the age of the driver is 0.64.
11. If the age of the driver increased by ten years then the predicted maximum distance to see the sign would decrease by ten slopes or 30.07 ft.
12. I don't have any strong concerns, though the data look very slightly nonlinear and very slightly heteroscedastic (Figure 1).

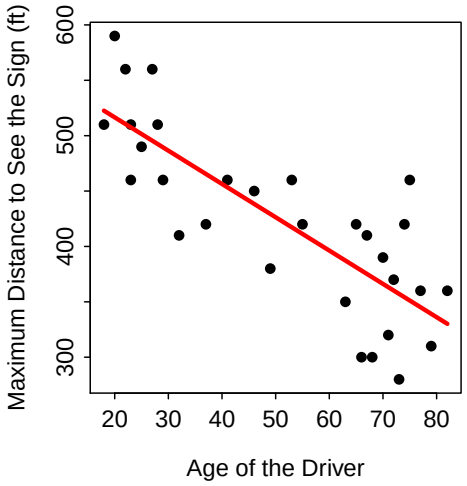

Figure 1. Fitted line plot for the relationship between maximum distance to see a sign and the age of a driver.

Appendix – R Commands

```
library(NCStats)
setwd('C:/aaaWork/Books/IntroStats/HW/')
d <- read.csv("vision.csv")
( lm1 <- lm(distance~age,data=d) )
predict(lm1,data.frame(age=40))
410-predict(lm1,data.frame(age=50))
rSquared(lm1)
corr(~distance+age,data=d)
fitPlot(lm1,ylab="Maximum Distance to See the Sign (ft)",xlab="Age of the Driver")
```