

Professor Notes about “Linear Regression” Homework

- 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).
- The correlation coefficient is positive in this case because the association is positive. You may have to remember to put a negative sign on the square root of the r^2 value if the association was negative.
- When calculating the correlation coefficient, make sure that you compute it with JUST the weekday data and not the combined weekday and weekend (i.e., the original) data. If you mistakenly computed it on the original data, then you would have found a $r = 0.47$

Estimating Fishing Pressure

1. The response variable is fishing pressure.
2. The explanatory variable is index count.
3. The best-fit line is $pressure = 764.3 * index + 2337.3$.
4. The slope indicates that for every increase of one vehicle in the index count the pressure will increase by 764.3 angler-hours, on average.
5. The intercept indicates that if the index count was zero, then the pressure would be 2337.3 angler-hours, on average.
6. This prediction should not be made as an index count of 20 is outside the observed results for this variable (Figure 1) and is, thus, an extrapolation.
7. The predicted amount of fishing pressure for an index count of five vehicles is 6158.6 angler-hours.
8. The residual for an individual with a fishing pressure of 5100 and an index count of eight vehicles is -3351.3 angler-hours.
9. The correlation between fishing pressure and index count is 0.72.
10. The proportion of variability in fishing pressure that is explained by knowing index count is 0.51.
11. If the index count is under-estimated by two vehicles, then the pressure value will be under-predicted by two slopes or 1528.5 angler-hours
12. The data appears to be roughly linear but this is difficult to tell based on the wide scatter of points. My concern, though, is that the data appear to be somewhat heteroscedastic (Figure 1).

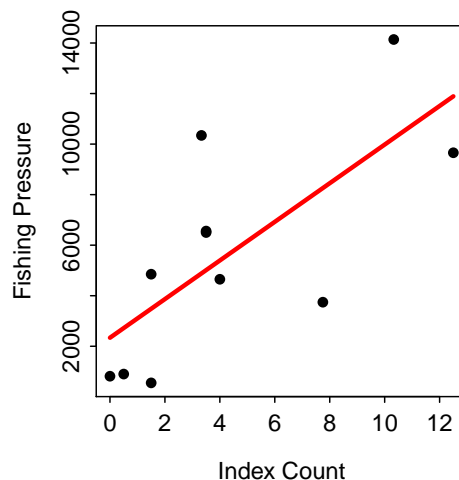


Figure 1. Fitted line plot for the relationship between fishing pressure and number of counted cars on Wyoming reservoirs.

Appendix – R Commands

```
library(NCStats)
setwd('C:/aaaWork/Books/IntroStats/HW/')
d <- read.csv("AlcovaRes.csv")
wd <- filterD(d, day.type=="wday")
( lm1 <- lm(pressure~index, data=wd) )
predict(lm1, data.frame(index=5))
5100-predict(lm1, data.frame(index=8))
rSquared(lm1)
corr(~pressure+index, data=wd)
fitPlot(lm1, ylab="Fishing Pressure", xlab="Index Count")
```