STAT 217: Class handout 11/12

The following is a class exercise that will help you to understand more about interpretation in the context of MLR and some issues that can arise. You will not turn this in, but I expect you to write down answers to the following questions so that you have it in your notes. The questions labeled **Group Discussion Questions** I'd like you to discuss with your group, but you do not need to write down your answer.

```
c.lm <- lm(happiness ~ o.caff)
summary(c.lm)

##
## Call:
## lm(formula = happiness ~ o.caff)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -23.30  -5.87   1.47   6.40  20.16
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -53.297     54.286   -0.98     0.34
## o.caff         1.781      0.271    6.58  3.5e-06
##
## Residual standard error: 11.2 on 18 degrees of freedom
## Multiple R-squared:  0.706,Adjusted R-squared:  0.69
## F-statistic: 43.2 on 1 and 18 DF,  p-value: 3.55e-06

i.lm <- lm(happiness ~ income)
summary(i.lm)

##
## Call:
## lm(formula = happiness ~ income)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -18.18  -6.52   2.07   7.47  13.31
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -5.82e+03   7.17e+02   -8.11  2.0e-07
## income       3.82e-01   4.48e-02    8.54  9.6e-08
##
## Residual standard error: 9.23 on 18 degrees of freedom
## Multiple R-squared:  0.802,Adjusted R-squared:  0.791
## F-statistic: 72.9 on 1 and 18 DF,  p-value: 9.59e-08
```

1. Write out the estimated regression line for the linear model of happiness on caffeine.

2. Write out the estimated regression line for the linear model of happiness on income.

3. Interpret the estimate for the coefficient of caffeine (b_1) in the linear model of happiness on caffeine.

```
ic.lm <- lm(happiness ~ o.caff+income)
summary(ic.lm)
```

```
##
## Call:
## lm(formula = happiness ~ o.caff + income)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -14.292  -5.980  -0.611   5.644  13.513
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.08e+04   3.05e+03   -3.53   0.0026
## o.caff      -1.68e+00   1.01e+00   -1.66   0.1143
## income       7.12e-01   2.03e-01    3.51   0.0027
##
## Residual standard error: 8.81 on 17 degrees of freedom
## Multiple R-squared:  0.83,Adjusted R-squared:  0.81
## F-statistic: 41.4 on 2 and 17 DF,  p-value: 2.92e-07
```

4. Write out the estimated regression line for the linear model of happiness on income and caffeine.

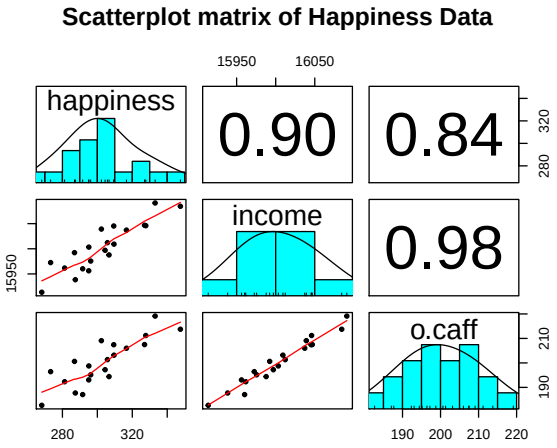
5. Interpret the estimate for the coefficient of caffeine (b_1) in the linear model of happiness on income and caffeine.

6. **Group Discussion**: Explain why the estimate for the coefficient of caffeine changes from positive to negative when we add income into the model.

Take home message: b_1 has a different meaning (and may change drastically) depending on what other terms are in the model!

7. Let's look at some plots to try to make sense of this. **Group Discussion**: Describe the relationship between income and caffeine.

```
require(mosaic)
require(psych)
pairs.panels(~happiness + income + o.caff, pch = 20, ellipse = F, main = "Scatterplot matrix of Happiness
```



We will take some time here to write down some notes.

8. What was the standard error for the coefficient on income when only caffeine was in the model?

9. What is the standard error for the coefficient on income when both caffeine and income are in the model?

10. Based off the two previous questions, calculate $\sqrt{VIF}$ for the coefficient on income. Then, find VIF_{inc}.

11. Below is a table of VIFs. Does this match up, roughly, with what you found? There may be some rounding error.

```
require(car)
vif(ic.lm)


## o.caff income
##   22.5   22.5
```

12. List three ways to detect multicollinearity in your model.

13. **Group Discussion**: We have looked at three models in this handout:

    - Happiness on caffeine
    - Happiness on income
    - Happiness on caffeine and income

    Suppose you were told a person's caffeine intake and income and asked to predict the person's happiness. Which model would you use for prediction and why?