MLDS 401 Homework 3
Due: Monday, October 16, 3:00
Professor Malthouse

Work in your assigned teams. Submit a single assignment with all names.

1. Use the auto data set from JWHT problem 3.9 on page 122.

   (a) Now regress `mpg` on `cylinders`, `displacement`, `weight`, and `year`. Comment on the signs of the estimated coefficients and note which are significantly different from 0. What is value of R^2?

```
Call: lm(formula = mpg ~ cylinders + displacement + weight + year,
    data = auto)

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -14.076941   4.055159  -3.471 0.000575 ***
cylinders     -0.289589   0.329225  -0.880 0.379611
displacement   0.004973   0.006701   0.742 0.458425
weight        -0.006702   0.000572 -11.717  < 2e-16 ***
year           0.764751   0.050684  15.089  < 2e-16 ***

Residual standard error: 3.436 on 392 degrees of freedom
Multiple R-squared:  0.8091,Adjusted R-squared:  0.8072
F-statistic: 415.5 on 4 and 392 DF,  p-value: < 2.2e-16
```

   Answer: It is odd that displacement has a positive sign, although it is not significant. Only weight and year are significant. $R^2 = .8091$.

   (b) Compute the variance inflation factors. What do they tell you?

```
 cylinders displacement       weight         year
 10.524432    16.406259     7.888061     1.173000
```

   Answer: The VIF values for all variables but year are large, indicating multicollinearity: cars that weigh more tend to have larger engine displacement and more cylinders.

   (c) Drop `weight` from the model. What happens to the parameter estimates and R^2?

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  -18.199719   4.688296  -3.882 0.000122 ***
cylinders     -0.620910   0.380657  -1.631 0.103658
displacement  -0.041545   0.006265  -6.632  1.1e-10 ***
year           0.699324   0.058461  11.962  < 2e-16 ***

Residual standard error: 3.988 on 393 degrees of freedom
Multiple R-squared:  0.7423
```

Answer: The coefficients change drastically. Cylinders goes from $-.29$ to $-.62$. Displacement goes from 0.00497 to -0.0415 and now has a positive, significant slope. Year changes less. R^2 decreases a little bit to 0.699. The other variables do the much of the explaining previously done by weight.

(d) Drop `weight` and `displacement` from the model. What happens to the parameter estimates and R^2?

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -17.30285    4.93534  -3.506 0.000507 ***
cylinders    -3.00405    0.13223 -22.718  < 2e-16 ***
year          0.75289    0.06098  12.347  < 2e-16 ***

Multiple R-squared:  0.7135
```

Answer: Cylinders is now significant, and R^2 reduces only slightly to .7135.

(e) For class discussion: Draw a DAG for weight, horsepower, displacement, cylinders and mpg.

2. In class I mentioned that you should control for a variable if it is a fork, but not if it is a pipe or collider. The next three problems will ask you to generate some data to see why. Let $n = 500$ be the number of observations. Generate $w \sim \mathcal{U}[0, 5]$. Then let $x = w + \delta$ where $\delta \sim \mathcal{N}(0, 1)$. Then let $y = 4 + 2x - 3w + \epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$. **For all three problems, our interest is in understanding the effect of x on y, and w is another variable that you are considering including.**

```
library(car)   # for vif function
library(psych) # for describe
n = 500
set.seed(123456)
dat = data.frame(w = 5*runif(n))
dat$x = dat$w + rnorm(n)
dat$y = 4+2*dat$x -3*dat$w + rnorm(n)
plot(dat, pch=".")
```

(a) Is this a pipe, fork or collider? *Answer: Fork*

(b) Generate a correlation matrix and basic descriptive statistics (min, max, mean, sd).

```
> describe(dat)
  vars   n mean   sd median trimmed  mad   min  max range  skew
w    1 500 2.47 1.42   2.44    2.47 1.78  0.01 4.99  4.99  0.03
x    2 500 2.52 1.76   2.56    2.50 2.00 -2.79 7.07  9.85  0.04
```

```
y    3 500 1.60 2.55    1.67     1.61 2.45 -5.64 7.95 13.59 -0.07
> cor(dat)
           w           x            y
w  1.0000000 0.825194559 -0.510735210
x  0.8251946 1.000000000  0.008261618
y -0.5107352 0.008261618  1.000000000
```

(c) Regress y on x. Is the coefficient of x significant at the .05 level? Does a 95% CI cover the true slope for x, namely 2?

Answer: As we see below, the coefficient of x is not significant and $2 \notin [-0.115, 0.139]$. The scatterplot shows no relationship between x and y.

```
> fit = lm(y ~ x, dat)
> summary(fit)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.57175    0.19888   7.903 1.76e-14 ***
x            0.01194    0.06477   0.184    0.854
> confint(fit)
                 2.5 %    97.5 %
(Intercept)  1.1809995 1.9625026
x           -0.1153232 0.1392085
```

(d) Now regress y on both x and w. Is the coefficient of x significant at the .05 level? Does a 95% CI cover the true slope for x, namely 2? *Answer: β_1 is now significantly different from 0 and $2 \in [1.857, 2.037]$.*

```
> fit2 = lm(y~x+w, dat)
> summary(fit2)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.92410    0.09211   42.60   <2e-16 ***
x            1.94695    0.04597   42.35   <2e-16 ***
w           -2.91956    0.05724  -51.01   <2e-16 ***
> confint(fit2)
                2.5 %    97.5 %
(Intercept)  3.743131  4.105073
x            1.856626  2.037279
w           -3.032024 -2.807100
```

(e) What are the values of VIF from this second regression? *Answer: Use **vif** to find 3.134 for each of the two variables, indicating some issues with multicollinearity. If we fit the correct model (with w) we have issues with inflated variance, but we get the right estimate $\beta_1 \approx 2$.*

3. Let $n = 500$ be the number of observations. Generate $x \sim \mathcal{U}[0, 5]$. Then let $y = x + \delta$ where $\delta \sim \mathcal{N}(0, 1)$. Then let $w = 4 + 2x + 3y + \epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$.

```
set.seed(12345)
dat = data.frame(x = 5*runif(n))
dat$y = dat$x + rnorm(n)
dat$w = 2*dat$x +3*dat$y + rnorm(n)
plot(dat, pch=".")
```

(a) Is this a pipe, fork or collider? *Answer: collider*

(b) Generate a correlation matrix and basic descriptive statistics (min, max, mean, sd).

Answer: Strong correations indicating a lot of multicollinearity

```
> describe(dat)
  vars   n  mean   sd median trimmed  mad   min   max range  skew
x    1 500  2.65 1.45   2.71    2.68 1.84  0.01  4.98  4.98 -0.15
y    2 500  2.69 1.78   2.73    2.71 1.86 -2.70  7.71 10.41 -0.14
w    3 500 13.32 7.94  13.66   13.48 9.03 -9.73 31.52 41.25 -0.17
> cor(dat)
          x         y         w
x 1.0000000 0.8278634 0.9192686
y 0.8278634 1.0000000 0.9710719
w 0.9192686 0.9710719 1.0000000
```

(c) Regress y on x. Is the coefficient of x significant at the .05 level? Does a 95% CI cover the true slope for x, namely 1? *Answer:* β_1 *is significant and* $1 \in [0.96, 1.1]$, *and so we get the right answer.*

```
> summary(fit)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.003949   0.093043  -0.042    0.966
x            1.016699   0.030869  32.936   <2e-16 ***
Residual standard error: 0.9968 on 498 degrees of freedom
Multiple R-squared:  0.6854,Adjusted R-squared:  0.6847

> confint(fit)
                 2.5 %    97.5 %
(Intercept) -0.1867546 0.1788566
x            0.9560490 1.0773495
```

(d) Now regress y on both x and w. Is the coefficient of x significant at the .05 level? Does a 95% CI cover the true slope for x, namely 1? *Answer: Still significant, but we have a sign flip and $1 \notin [-0.56, -0.47]$. Introducing w as a control leads to the wronge estimate of β_1!*

```
> fit2 = lm(y~x+w, dat)
> summary(fit2)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.009422   0.028716   0.328    0.743
x           -0.513705   0.024203 -21.225   <2e-16 ***
w            0.303005   0.004405  68.786   <2e-16 ***
Residual standard error: 0.3076 on 497 degrees of freedom
Multiple R-squared:  0.9701,Adjusted R-squared:   0.97

> confint(fit2)
                  2.5 %      97.5 %
(Intercept) -0.04699765  0.06584143
x           -0.56125800 -0.46615296
w            0.29435047  0.31166018
```

(e) What are the values of VIF from this second regression? *Answer: vif = 6.45*

(f) Compare the values of R^2 (or S_e) between the two models. Which model is better according or R^2 (or S_e). Is this the right model? *Answer: R^2 is much better in the model with w (0.97) versus 0.69. This is an example where of why you should not make your model specification decisions using R^2—it could lead you to the wrong model!*

4. Let $n = 500$ be the number of observations. Generate $x \sim \mathcal{U}[0, 5]$. Then let $w = x + \delta$ where $\delta \sim \mathcal{N}(0, 1)$. Then let $y = 2w + \epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$.

(a) Is this a pipe, fork or collider? *Answer: pipe*

(b) Generate a correlation matrix and basic descriptive statistics (min, max, mean, sd).

```
> describe(dat)
  vars   n mean   sd median trimmed  mad   min   max range  skew
x    1 500 2.65 1.45   2.71    2.68 1.84  0.01  4.98  4.98 -0.15
w    2 500 2.69 1.78   2.73    2.71 1.86 -2.70  7.71 10.41 -0.14
y    3 500 5.34 3.67   5.38    5.39 3.94 -7.20 15.04 22.24 -0.15
> cor(dat)
          x         w         y
x 1.0000000 0.8278634 0.8006114
w 0.8278634 1.0000000 0.9649392
y 0.8006114 0.9649392 1.0000000
```

(c) Regress y on x. Is the coefficient of x significant at the .05 level? *Answer: There is a fairly strong positive (and very highly significant) relationship between x and y.*

```
> fit = lm(y ~ x, dat)
> summary(fit)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.04018    0.20561  -0.195    0.845
x            2.03405    0.06822  29.818   <2e-16 ***

Residual standard error: 2.203 on 498 degrees of freedom
Multiple R-squared:  0.641,Adjusted R-squared:  0.6403
```

(d) Now regress y on both x and w. Is the coefficient of x significant at the .05 level? *Answer: The relationship between x and y is no longer significant after we control for w, but it is wrong to conclude that x has no relationship with y. It (x) affects y through w: $x \to w \to y$. If we want to how how x is related to y we should not control for w because it blocks the pipe. These three exercises should make clear that the decision of whether to include "control" variable w depends on understanding the causal relationships between the variables. In the case of a fork we need to include w, but in pipe and collider instances we should not include w.*

```
> fit2 = lm(y~x+w, dat)
> summary(fit2)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.03233    0.09015  -0.359    0.720
x            0.01432    0.05332   0.269    0.788
w            1.98656    0.04342  45.754   <2e-16 ***

Residual standard error: 0.9658 on 497 degrees of freedom
Multiple R-squared:  0.9311,Adjusted R-squared:  0.9308
```

(e) Which model has a better R^2? (You could also compare adjusted R^2, although we have not defined it yet.) *Answer: The model with x and w has a better R^2 (and adjusted R^2). Again, you should be careful in using model fit to decide the model.*

5. (12 points) JWHT problem 3.14a–f on page 125.

(a) Generate the data as shown below, but for those working in Python, I will translate the data generation process. Generate $n = 100$ points where $x_1 \sim \mathcal{U}[0,1]$ (`runif(100)`) and `rnorm(100)` produces 100 $\mathcal{N}(0,1)$ variables.

```
> set.seed(1)
> x1 = runif(100) # part a
> x2 = 0.5*x1 + rnorm(100)/10
> y = 2 + 2*x1 + .3*x2 + rnorm(100)
```

Write out the form of the linear model. What are the true regression coefficents and standard deviation of errors. *Answer:* $\beta_0 = \beta_1 = 2$, $\beta_2 = 0.3$, $\sigma_\epsilon = 1$, $y = 2 + 2x_1 + 0.3x_2 + e$

(b) What is the correlation between x_1 and x_2? *Answer:* ***cor(x1,x2)*** *= 0.83512.*

(c) Using this data, regress y on x_1 and x_2. Describe the results. What are the parameter estimates and how do they relate to the true parameters? Which coefficients are significantly different from 0? Are the true parameters "covered" by 95% CIs? *Answer: Barely reject* $H_0 : \beta_1 = 0$ *(*$P = .0487$*), but not* $H_0 : \beta_2 = 0$ *(*$P = .3754$*). All parameters are covered by the CIs (not required, but the standard error are large giving wide CIs).*

```
> fit = lm(y ~ x1+x2)    # part c
> summary(fit)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.1305     0.2319   9.188 7.61e-15 ***
x1            1.4396     0.7212   1.996   0.0487 *
x2            1.0097     1.1337   0.891   0.3754

> confint(fit)
                   2.5 %   97.5 %
(Intercept)  1.670278673 2.590721
x1           0.008213776 2.870897
x2          -1.240451256 3.259800
```

(d) Regress y on x_1 alone. Is β_1 significantly different from 0? Is the true β_1 covered by by a 95% CI?

Answer: β_1 *is very highly significant (*$P < .0001$*) and the true parameter is covered* $2 \in [1.19, 2.76]$

```
> fit = lm(y~x1)   # Part d
> summary(fit)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.1124     0.2307   9.155 8.27e-15 ***
```

```
x1              1.9759     0.3963   4.986 2.66e-06 ***
> confint(fit)
                2.5 %   97.5 %
(Intercept) 1.654488 2.570299
x1          1.189529 2.762329
```

(e) Regress y on x_2 alone. Is β_2 significantly different from 0? Is the true β_2 covered by by a 95% CI? Are the true parameters "covered" by the 95% confidence intervals? *Answer: β_2 is very highly significant ($P < .0001$) but the true value is not covered, $0.3 \notin [1.64, 4.16]$*

```
> fit = lm(y~x2)   # Part e
> summary(fit)
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.3899     0.1949   12.26  < 2e-16 ***
x2            2.8996     0.6330    4.58 1.37e-05 ***
> confint(fit)
                2.5 %   97.5 %
(Intercept) 2.003116 2.776783
x2          1.643324 4.155846
```

(f) Do the results in (c)–(e) contradict each other? Explain. *Answer: This illustrates the omitted variable bias. The predictors are highly correlated and can serve as proxies for one another. The β_2 coefficient is not significant in part c because of high standard errors, but is significant in part e because it explains some of the variation due to x_1.*

Answer: $\beta_1 = 2$ is in both intervals, but $\beta_2 = .3$ is only the interval for the first one. This is because of the omitted variable bias. The estimate of β_1 is also biased, but not enough to cause the interval not to cover the parameter.

6. (10 points) ACT 2.4: Omitted variables. Suppose that there are two predictor variables, x_1 and x_2, but we fit the straight line model $y = \beta_0 + \beta_1 x_1 + \epsilon$ omitting x_2. If, in fact, the true model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon$, show that

$$\mathbb{E}(\hat{\beta}_1) = \beta_1 + \beta_2 \sum_{i=1}^{n} c_i x_{i2} = \beta_1 + \frac{\beta_2}{S_{11}} \sum_{i=1}^{n} (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) = \beta_1 + \beta_2 r \frac{s_2}{s_1},$$

where $c_i = (x_{i1} - \bar{x}_1)/S_{11}$, $S_{11} = \sum_{i=1}^{n}(x_{i1} - \bar{x}_1)^2$, r is the sample correlation coefficient between x_1 and x_2, and $s_1 > 0$, $s_2 > 0$ are the sample SD's of x_1, x_2, respectively. Thus $\hat{\beta}_1$ is biased with the bias given by $\beta_2 r s_2/s_1$. Under what condition is this bias zero? Discuss how this result applies to JWHT 3.14 that you worked in problem 5.

Answer: We need to establish each of the three equalities in:

$$E(b_1) = \beta_1 + \beta_2 \sum_{i=1}^{n} c_i x_{i2} = \beta_1 + \frac{\beta_2}{S_{11}} \sum_{i=1}^{n} (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) = \beta_1 + \beta_2 r \frac{s_2}{s_1},$$

To establish the first, note that $\sum c_i = 0$, $\sum x_{i1} c_i = 1$, *and* $\mathbb{E}(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$*; the first two require a little bit of algebra, and the third follows from the assumption that* $\mathbb{E}(e) = 0$. *Also note that*

$$b_1 = \frac{\sum (x_{i1} - \bar{x}_1)(y_i - \bar{y})}{\sum (x_{i1} - \bar{x}_1)^2} = \sum c_i (y_i - \bar{y}) = \sum c_i y_i - \bar{y} \underbrace{\sum c_i}_{0} = \sum c_i y_i$$

Then

$$\begin{aligned}
\mathbb{E}(b_1) &= \mathbb{E}\left(\sum c_i y_i\right) = \sum c_i (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}) \\
&= \beta_0 \underbrace{\sum c_i}_{0} + \beta_1 \underbrace{\sum x_{i1} c_i}_{1} + \beta_2 \sum c_i x_{i2} = \beta_1 + \beta_2 \sum c_i x_{i2}
\end{aligned}$$

and we have the first equality. To get the second equality it is slightly easier to go from the third term to the second term, although we can start with either term:

$$\begin{aligned}
\sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2) &= \sum [x_{i1} x_{i2} - \bar{x}_1 x_{i2} - \bar{x}_2 x_{i1} + \bar{x}_1 \bar{x}_2)] \\
&= \sum x_{i1} x_{i2} - \bar{x}_1 \sum x_{i2} - \bar{x}_2 \sum x_{i1} + n \bar{x}_1 \bar{x}_2 \\
&= \sum x_{i1} x_{i2} - n \bar{x}_1 \bar{x}_2 - n \bar{x}_1 \bar{x}_2 + n \bar{x}_1 \bar{x}_2] \\
&= \sum x_{i1} x_{i2} - n \bar{x}_1 \bar{x}_2 = \sum x_{i1} x_{i2} - \bar{x}_1 \sum x_{i2} \\
&= \sum x_{i2} (\sum x_{i1} - \bar{x}_1).
\end{aligned}$$

The third equality follows easily from the definition of $r = \sum (x_{i1} - \bar{x}_1)(x_{i2} - \bar{x}_2)/[(n-1)s_1 s_2]$. *The bias will be 0 when* $r = 0$, *i.e., an orthogonal design (or when* $\beta_2 = 0$ *or* $s_2 = 0$, *although these cases are not interesting). In problem 3.14* $r = .835 > 0$, $\beta_1 = 2 > 0$, *and* $\beta_2 = .3 > 0$. *A simple linear regression with either* x *variable will produce estimates with an upward bias.*

Here is a different solution:

$$\begin{aligned}
\mathbb{E}(\widehat{\beta}_1) &= \frac{1}{S_{11}} \mathbb{E}\left[\sum_{i=1}^{n} (x_{i1} - \overline{x}_1)(y_i - \overline{y})\right] \\
&= \frac{1}{S_{11}} \mathbb{E}\left[\sum_{i=1}^{n} (x_{i1} - \overline{x}_1) y_i\right] \\
&= \frac{1}{S_{11}} \left[\sum_{i=1}^{n} (x_{i1} - \overline{x}_1)(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2})\right].
\end{aligned}$$

Now note that

$$\sum_{i=1}^{n}(x_{i1}-\overline{x}_1)\beta_0 = \beta_0\sum_{i=1}^{n}(x_{i1}-\overline{x}_1) = 0,$$

$$\frac{1}{S_{11}}\sum_{i=1}^{n}\beta_1(x_{i1}-\overline{x}_1)x_{i1} = \frac{\beta_1}{S_{11}}\sum_{i=1}^{n}(x_{i1}-\overline{x}_1)^2 = \beta_1,$$

and

$$\begin{aligned}
&\frac{1}{S_{11}}\left[\beta_2\sum_{i=1}^{n}(x_{i1}-\overline{x}_1)x_{i2}\right]\\
=\ &\frac{1}{S_{11}}\left[\beta_2\sum_{i=1}^{n}(x_{i1}-\overline{x}_1)(x_{i2}-\overline{x}_2)\right]\\
=\ &\beta_2\frac{\sum_{i=1}^{n}(x_{i1}-\overline{x}_1)(x_{i2}-\overline{x}_2)}{\sqrt{S_{11}S_{22}}}\sqrt{\frac{S_{22}}{S_{11}}}\\
=\ &\beta_2 r_{12}\frac{s_2}{s_1},
\end{aligned}$$

where

$$s_1^2 = \frac{\sum_{i=1}^{n}(x_{i1}-\overline{x}_1)^2}{n-1} = \frac{S_{11}}{n-1} \quad \textit{and} \quad s_2^2 = \frac{\sum_{i=1}^{n}(x_{i2}-\overline{x}_2)^2}{n-1} = \frac{S_{22}}{n-1}.$$

For the bias to be zero, the second term involving β_2 must be zero; hence r_{12} must be zero.

7. In class I mentioned that one of the effects of multicollinearity is to increase the variance of the slope estimates. This problem will show you why this is the case. Suppose that $y = \beta_1 z_1 + \beta_2 z_2 + e$, where all variables have been standardized prior to estimation making the intercept unnecessary. You have observed (z_{i1}, z_{i2}, y_i), for $i = 1, \ldots, n$, where the sample means are 0 and the sample variances are 1, i.e.,

$$\bar{z}_j = \frac{1}{n}\sum_{i=1}^{n} z_{ij} = 0, \quad S_j^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_{ij}-0)^2 = 1, \quad \text{and} \quad r = \frac{1}{n-1}\sum_{i=1}^{n} z_{i1}z_{i2},$$

where r is sample correlation between z_1 and z_2. How does VIF $= 1/(1-r^2)$ affect the variance of the slope estimates? Hint: recall that $V(\mathbf{b}) = \sigma^2(\mathbf{Z}^\mathsf{T}\mathbf{Z})^{-1}$ and

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

Answer:

$$\mathbf{Z}^\mathsf{T}\mathbf{Z} = \begin{pmatrix} \sum z_{i1}^2 & \sum z_{i1}z_{i2} \\ \sum z_{i1}z_{i2} & \sum z_{i2}^2 \end{pmatrix} \cdot \frac{n-1}{n-1} = (n-1)\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$$

Note that $(c\mathbf{A})^{-1} = \mathbf{A}^{-1}/c$ *for constant* c*, so*

$$(\mathbf{Z}^\mathsf{T}\mathbf{Z})^{-1} = \frac{1}{(n-1)(1-r^2)} \begin{pmatrix} 1 & -r \\ -r & 1 \end{pmatrix} = \text{VIF} \frac{1}{(n-1)} \begin{pmatrix} 1 & -r \\ -r & 1 \end{pmatrix}$$

The variances of the slope estimates are given by the diagonal, multiplied by the estimate of σ^2*. If we hold* σ^2 *and* n *constant, the diagonal term depends on* $1/(1-r^2)$*, which is minimized when* $r = 0$ *(orthogonal design), and approaches infinity as* r^2 *increases to 1. The interpretation of VIF becomes clear!*

8. Start analyzing the bike data. There is nothing to turn in.