Kyle Nabors
ECN 140
Assignment 3

1 a. minimise $\sum_{i=1}^{n} (y_i - b_1 x_{1i} - b_2 x_{2i})^2$

$$\min_{b_1, b_2} f(x, y, b_1, b_2) \sum_{i=1}^{n} (y_i - b_1 x_{1i} - b_2 x_{2i})^2$$

b. $\frac{\partial f}{\partial b_1} = -2 \sum_{i=1}^{n} (y_i - b_1 x_{1i} - b_2 x_{2i}) x_{1i} = 0$

$$\frac{\partial f}{\partial b_2} = -2 \sum_{i=1}^{n} (y_i - b_1 x_{1i} - b_2 x_{2i}) x_{2i} = 0$$

c. $-2 \sum_{i=1}^{n} (y_i - b_1 x_{1i} - b_2 x_{2i}) x_{1i} = 0$

$$-\sum y_i x_{1i} + b_1 \sum x_{1i}^2 + b_2 \sum x_{1i} x_{2i} = 0$$

$$b_1 \sum x_{1i}^2 = \sum y_i x_{1i} - b_2 \sum x_{1i} x_{2i}$$

$$b_1 = \frac{\sum y_i x_{1i} - b_2 \sum x_{1i} x_{2i}}{\sum x_{1i}^2}$$

$$b_1 = \hat{\beta}_1 - \frac{b_2 \sum x_{1i} x_{2i}}{\sum x_{1i}^2} \qquad \sum x_{1i} x_{2i} = 0$$

$$b_1 = \hat{\beta}_1 - \cancel{\frac{b_2 \cdot 0}{\sum x_{1i}^2}}$$

$$b_1 = \hat{\beta}_1 = \frac{\sum y_i x_{1i}}{\sum x_{1i}^2}$$

d. $\sum x_{1i}x_{2i} \neq 0$

$$\hat{\beta}_1 = \frac{\sum y_i x_{1i} - \hat{\beta}_2 \sum x_{1i} x_{2i}}{\sum x_{1i}^2}$$

$$\hat{\beta}_2 = \frac{\sum y_i x_{2i} - \hat{\beta}_1 \sum x_{1i} x_{2i}}{\sum x_{2i}^2}$$

$$\hat{\beta}_1 = \frac{\sum y_i x_{1i}}{\sum x_{1i}^2} - \hat{\beta}_2 \frac{\sum x_{1i} x_{2i}}{\sum x_{1i}^2}$$

$$\hat{\beta}_1 = \frac{\sum y_i x_{1i}}{\sum x_{1i}^2} - \left(\frac{\sum y_i x_{2i} - \hat{\beta}_1 \sum x_{1i} x_{2i}}{\sum x_{2i}^2} \right) \frac{\sum x_{1i} x_{2i}}{\sum x_{1i}^2}$$

$$\left(1 - \frac{(\sum x_{1i} x_{2i})^2}{\sum x_{2i}^2 \sum x_{1i}^2}\right) \hat{\beta}_1 = \frac{\sum y_i x_{1i}}{\sum x_{1i}^2} - \frac{\sum y_i x_{2i} \sum x_{2i}}{\sum x_{1i} \sum x_{2i}^2}$$

$$\hat{\beta}_1 = \frac{\dfrac{\sum y_i x_{1i}}{\sum x_{1i}^2} - \dfrac{\sum y_i x_{2i} \sum x_{2i}}{\sum x_{1i} \sum x_{2i}^2}}{\left(1 - \dfrac{(\sum x_{1i} x_{2i})^2}{\sum x_{2i}^2 \sum x_{1i}^2}\right)}$$

e. $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + u_i$

$$f = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{1i} - b_2 x_{2i})^2$$

$$\frac{df}{db_0} = -2 \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})$$

$$0 = \sum (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i})$$

$$\sum \hat{\beta}_0 = \sum y_i - \hat{\beta}_1 x_{1i} - \hat{\beta}_2 x_{2i}$$

$$\frac{n}{n} \hat{\beta}_0 = \frac{1}{n} \sum y_i - \hat{\beta}_1 \frac{1}{n} \sum x_{1i} - \hat{\beta}_2 \frac{1}{n} \sum x_{2i}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2$$

$$f = \sum\left(y_i - (\bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2) - \beta_1 \bar{x}_{1i} - \hat{\beta}_2 x_{2i}\right)^2$$

$$= \sum\left((y_i - \bar{y}) - \hat{\beta}_1 (x_{1i} - \bar{x}_1) - \hat{\beta}_2 (x_{2i} - \bar{x}_2)\right)^2$$

$$= \sum\left(\tilde{y}_i - \hat{\beta}_1 \tilde{x}_{1i} - \hat{\beta}_2 \tilde{x}_{2i}\right)^2$$

$$\hat{\beta}_1 = \frac{\sum \tilde{y}_i \tilde{x}_{1i}}{\sum \tilde{x}_{1i}^2} = \frac{\sum (y_i - \bar{y})(x_{1i} - \bar{x}_1)}{\sum (x_{1i} - \bar{x}_1)^2}$$

1. Empirical Exercise
    a. 2.1
        i. After running a regression, we find that the estimated intercept for the regression of average hourly earnings on age is 4.62605 and the slope is .5118171.
        ii. Using our regression, due to Bobs age being 26 his predicted earnings are $17.89/hr. Alexis whose age is 30 has projected earnings of $19.93/hr.
        iii. Using our regression output we can identify our R squared value which indicates that age only accounts for ~1.85% of earnings so no age does not account for a large amount of predicted earnings.
    b. 2.2
        i. After running a regression of average hourly earnings on age we find that the intercept is 4.62605 and the slope is .5118171.
        ii. After running a regression of average hourly earnings on age, gender, and education, we find that the estimated effect of age on earnings is ~1.80%. We also find that the 95% confidence interval for age is (.4327747 - .5877973)
        iii. Comparing our results from 2.1a and 2.2a it does not appear to suffer from omitted variable bias as age is very weakly correlated with female and bachelor.
        iv. Using our regression, we find that bob a 26-year-old male with a high school diploma has estimated earnings of $15.13. We also find that Alexis who is a 30-year-old female with a bachelor's degree has estimated earnings of $21.67.
2. Research Question
    a. In my project I am setting out to predict the correlation between a person's math skills and their willingness to take risks.
    b. If I was to run a randomized control trial for this data, I would run a questionnaire on a variety of subjects. I would have subjects that range in math skills from high school dropout to PhD in mathematics. I would then ask them their choice on a series of lotteries to gauge their willingness to take risks.
    c. The zero conditional mean assumption is unlikely to be true to do being unable to account for all factors. We cannot control for or include all variables in our model and there is likely to be an omitted variable that has some level of correlation with X no matter how small.
    d. Three control variables I would want to include would be economic background, history of gambling, and income level. Due to economists generally thinking about risk differently than the average person I believe it would be a good variable to include. A history of gambling could lead to people having a distorted perception of risk. Due to diminishing marginal utility of wealth it would be good to know the persons income due to it potentially effecting their answer.
    e. Using the esttab command in Stata we can create tables of control variables as seen in figure 1 and figure 2.

f. Marc Oliver Rieger, Mei Wang, Thorsten Hens(2014) Risk Preferences Around the World. Management Science 61(3):637-648. https://doi.org/10.1287/mnsc.2013.1869

g. Using survey results from 53 countries, this paper set out to measure risk preferences from around the world. The data was collected with undergraduate students at over 60 universities. The survey totalled 6,912 results with students from economics or business departments with an average age of 21.5 and a split of 53% male 47% female. The variables that are captured in this paper include both economic conditions such as GDP per capita. The paper also uses cultural factors which are measured by the Hofstede dimensions individualism and uncertainty avoidance. They also compare other factors such as male versus female risk preferences and attitudes towards risk levels based on wealth. They took this data and used various regressions to determine risk levels of various groups and tried to break people into categories and determine signiant factors in a person's risk preferences. With these results being only based on economic oriented students they do not represent a general public but do give insight into the next generation of economists.

Appendix

Figure 1:

	(1) risktaking
patience	**0.198*** **
	(55.69)
posrecip	**-0.00779***
	(-2.04)
altruism	**0.0821*** **
	(20.63)
_cons	**0.00432**
	(1.25)
N	**78917**

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Figure 2:

	(1) risktaking
trust	**0.0533*** **
	(14.40)
age	**-0.0138*** **
	(-69.78)
gender	**-0.219*** **
	(-31.57)
_cons	**0.700*** **
	(73.03)
N	**78039**

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001