Problem Set 1
ECON 30331
(Due at the start of class, Wednesday, January 24, 2018)

Bill Evans
Spring 2018

1. Suppose the daily high temperature (x) for South Bend in the summer months is a uniform distribution between 50 and 100 degrees. The PDF for this distribution is f(x)=1/50.
   a. Show that this is a proper PDF, that is $\int_{50}^{100} \frac{dx}{50} = 1$.
   b. John loves to golf but he will not golf when the temperature is below 65 or above 90. What is the fraction of days John can expect to NOT golf during summer months in South Bend?

2. The exponential distribution is sometimes used to model the "time to failure" of mechanical products. Suppose a light bulb is expected to "fail" after x hours of use and the PDF of this failure time is given by the exponential distribution $f(x) = \beta e^{-\beta x}$ for x>0 and β>0. One useful value to calculate is the "time to failure" which is the probability that the lightbulb will fail sometime through a hours of use. This is defined mathematically as the CDF, or $\Pr(x \le a) = \int_0^a f(x)dx = \int_0^a \beta e^{-\beta x} dx.$
   a. Using the rules of calculus, what is $\Pr(x \le a)$?
   b. Using the results from a), suppose $\beta = 0.001.$ What is the chance the lightbulb will fail within 500 hours of use, or, what is $\Pr(x \le 500)$?
   c. Using the results from a),), suppose $\beta = 0.001.$What is the chance the lightbulb will last longer than 1500 hours of use, or, what is $\Pr(x > 1500)$

3. Suppose that the random variable (z) is distributed as a standard normal distribution. Using a standard normal table, calculate the following probabilities:
   a. Prob($Z \le 0.35$)
   b. Prob($Z > -1.5$)
   c. Prob($-1.22 < z \le 0.57$)

4. Suppose that annual inches of snowfall in South Bend is normally distributed with a mean of 64 inches (μ) and a standard deviation (σ) of 12 inches.
   a. What is the chance that South Bend 12 or fewer inches of annual snowfall?
   b. What is the chance that South Bend will get more than 80 inches of snowfall?

5. Listed below are 10 values from two series. Without calculating values, what is the correlation coefficient for x and y (hint: graph the points).

X	2	4	6	8	10	12	14	16	18	20
Y	-20	-18	-16	-14	-12	-10	-8	-6	-4	-2

6. Many law schools use an 'admission index,' which is a weighted average of college GPA and the Law School Admission Test (LSAT) score, to screen applicants. Applicants whose admission index is above a certain threshold are considered for admission and those with scores below a threshold are rejected immediately. For example, the University of Oklahoma uses an admission index of I = 15*GPA + LSAT. Among the pool of applications for this law school in 2002, the average GPA was 3.2 with a standard deviation of 1.2 the average LSAT score was 150 with a standard deviation of 5, and the **correlation coefficient** between the LSAT and GPA is 0.50. What is the expected value and variance for the admission index at the Oklahoma Law School for their pool of applicants?

7. In this problem, we examine whether wearing a helmet reduces the chance of mortality in a motor cycle crash. The data below is taken from all motorcycle crashes nationwide from 1988-2005 that had a) two people on the bike, and b) the passenger on the bike was not wearing a helmet and died which means the accident was severe enough to generate a fatality. Below is a 2 x 2 table that lists the probabilities of whether the driver on the bike in the accident was wearing a helmet and died in the crash. For example, 28.3% of drivers were wearing a helmet and did not die in a crash, Pr(Wearing a helmet ∩ didn't die) = 0.283

Did the driver die?

		No	Yes	Total
Was driver wearing a helmet?	No	0.161	0.343	0.504
	Yes	0.283	0.213	0.496
	Total	0.444	0.556	1.000

Using the table above, answer the following questions:

a) What is the Pr(Died?)
b) What is the Pr(Was wearing a helmet)?
c) What is the Pr(Died | Was wearing a helmet)?
d) What is the Pr(Died | Was not wearing a helmet)?
e) By how much does the probability of dying in a serious crash go down if the driver is wearing a helmet?

To check your answer to part e) please see the abstract on page 5 of the following report http://www-nrd.nhtsa.dot.gov/Pubs/809715.PDF

8. The *Tornado Fuel Saver* is a device that can be installed in the air-intake of your vehicle's engine that is advertised to improve gas mileage and horsepower. (You may have seen an infomercial for the Tornado Fuel Saver on late-night TV! This product is real.). The manufacturer advertises that the Tornado Fuel Saver can boost you cars' miles per gallon by 21 percent. A car magazine recently conducted a test of the Tornado Fuel Saver and installed the device on 25 cars. The magazine found that miles per gallon improved by an average of 18 percent (x) in this sample. The standard deviation (s) of this estimate is an 8 percent improvement. From this sample of 25, construct a 95% confidence interval for the expected percent improvement in miles per gallon (μ) generated by installing the fuel saver. Can you reject or not reject the null that fuel saver boosts MPG by 21%?

9. A popular weight loss program these days is the 'Atkins Diet' that stresses a high protein/low carbohydrate diet. The Atkins diet is in stark contrast to many diets that stress low fats and /high carbohydrates. To examine which diet produces greater weight loss, a group of researchers randomly assigned 40 overweight patients to either an Atkins type diet (n_A=20) or a Low Fat/High Carb diet (n_L=20). Participants were given detailed guidelines about the diets and a phone number for a 'diet hot line' to call with any questions about the program they were assigned. The participants were weighed at the beginning of the diet and again in 6 months. After 6 months, dieters in the Atkins plan lost an average of 16 pounds whereas the participants in the Low Fat/High Carb diet lost only an average of 7 pounds. A summary of results from the experiment are listed below.

	Atkins Diet	Low Fat/High Carb Diet
n (sample sizes)	20	20
x (average change in weight, in pounds)	-16	-7
s (standard deviation)	12	8

a. Construct a **95% confidence interval** for d =u_A - u_L-- the difference in weight loss generated by the Atkins and Low Fat diets. What are the appropriate degrees of freedom and t-value you should use in this context?

b. Using the confidence interval from a), test the null hypothesis that H_o: d=0 -- there is no difference in the weight loss generated by the Atkins versus the Low Fat diets. Can you reject or not reject the null hypothesis?

c. Using a **t-test** and a **99%** confidence level, test the null hypothesis that H_o: d=0 -- there is no difference in the weight loss generated by the Atkins versus the Low Fat diets. What are the appropriate degrees of freedom of the t-statistic you use in this context? Can you reject or not reject the null hypothesis? Explain your answer.

10. Suppose Y_1 and Y_2 are independent, var(Y_1) = Var(Y_2) = σ^2_y and Z_1 =Y_1 +Y_2. What is Var(Z)?

11. Problem 11 continued -- generalize the results of problem 11. Suppose Y_1,Y_2 .... Y_n are independent, var(Y_1) = Var(Y_2) = ...Var(Y_n)= σ^2_y and Z_2 =Y_1 +Y_2 + Y_3 + ...Y_n. What is Var(Z_2)?