Stat 104: Quantitative Methods for Economics
Homework 5: Due Monday, October 16

Homework policy: ***Homework is due by 8:00 am (EST) on the due date***. Homework is to be handed in via the course website in pdf format. You do not need to type the homework; there are many ways (scanner in the library or phone apps) to convert written homework into a pdf file. Ask the teaching staff if you need assistance.

Late homework will not be accepted. You are encouraged to discuss homework problems with other students (and with the instructor and TFs, of course), but you must write your final answer in your own words. Solutions prepared "in committee" or by copying someone else's paper are not acceptable.

- Please submit your homework in pdf format; this can be done in Word, or OpenOffice or via cellphone apps that will scan and turn into pdf.
- Please make your homework solutions legible by **bolding** or using circles to identify your solution.
- Since we are not printing out anything, use lots of s p a c e for your solutions, and put each answer on a different page if it makes the solution easier to read.
- Please make sure your submitted solutions are in numerical order [problem 1, problem 2 and so on].
- Please keep your computer output to a minimum and focus on the required answer. The easiest way to put your computer output into your homework is to cut and paste it into a Word file and use the font "courier new".
- Please keep in mind the course rules on Academic Honesty and Collaboration

1) The joint probability mass function of random variables X and Y is given by

$$P(X=1 \; and \; Y=1)=1/4 \quad P(X=1 \; and \; Y=2)=1/2$$
$$P(X=2 \; and \; Y=1)=1/8 \quad P(X=2 \; and \; Y=2)=1/8$$

a) Are X and Y independent? Justify your answer
b) Compute P(XY<3)
c) Compute P(X+Y>2)

2) A large insurance agency services a number of customers who have purchased both a homeowner's policy and an automobile policy from the agency. For each type of policy, a deductible amount must be specified. For an automobile policy, the choices are $100 and $200, whereas for a homeowner's policy, the choices are 0, $100, and $300. [editorial note-this problem is a pain-it is a lot of algebra. It is good to do these calculations once by hand. Feel free to curse while you work on this problem.]

Suppose an individual with both types of policy is selected at random from the agency's files. Let

X = the deductible amount on the auto policy, and
Y = the deductible amount on the homeowner's policy.

Suppose the following table represents the joint distribution of X and Y:

	P(x,y)	0	100	300
X	100	0.05	0.2	0.23
	200	0.3	0.2	0.02

a) What are the mean and standard deviation of X?
b) What are the mean and standard deviation of Y?
c) What is the covariance between X and Y?
d) What is the correlation between X and Y?
e) Calculate E(X+Y).
f) Calculate Var(X+Y).
g) Are X and Y independent? Justify your answer.
h) What is the expected value of Y given X= 250?

3) Suppose X and Y are independent random variables, and suppose X is binomial with n = 10 and p = .4; while Y is binomial with n = 12 and p = .2. Find the variance of 2X + 3Y .

4) The annual cost of owning a dog is a normal random variable with mean \$695 and standard deviation \$45. The annual cost of owning a cat is a normal random variable with mean \$705 and standard deviation \$35. What is the probability that the total annual cost of owning one dog and two cats exceeds \$2000? [assume dog and cat ownership is independent].

5) An investor plans to divide \$ 200,000 between two investments. The first yields an expected profit of 8% with a standard deviation of 5% whereas the second yields a profit with expected value 18% and standard deviation 6%. If the investor divides the money equally between these two investments, find the mean and standard deviation of the total profit. Assume the correlation between the two investments is -0.2.

6) When you load the following file of commands into R, you will create four new variables consisting of approximately three years of monthly return data.

```
source("http://people.fas.harvard.edu/~mparzen/stat104/hw5stockcode.txt")
```

The variables created are

- oilret – returns for the OIL etf
- goldret – returns for the GLD etf
- korsret – returns for stock KORS
- spyret – returns for the index (the SPY etf)

a) What company does each symbol represent? Go to finance.yahoo.com to find out.

b) What is the average monthly return for each of the stocks ? What is the standard deviation for the returns of the stocks ? What is the correlation between all the stocks?

c) Find the Beta for each stock. That is run a regression of each stock return as the Y variable and SPY returns as the X variable. Beta is the slope from this regression. Rank the stocks based on their Beta values (smallest to larges). Is the order the same as if you ranked them on their standard deviations from smallest to largest?

d) Create a side by side boxplot for these three stocks. How do they compare? Which looks the riskiest, which the safest?

e) Give the expected return and standard deviation of all the possible two stock portfolios (OIL,KORS), (OIL,GLD), (KORS,GLD) with equal amounts invested in each stock (weights of .5 for each stock).

f) Rank the three portfolios based on their standard deviation. How do they compare with holding one of the individual stocks ?

7) Jill's bowling scores are normally distributed with mean 170 and standard deviation 20, whereas Jack's scores are normally distributed with mean 160 and standard deviation 15. If Jack and Jill each bowl one game, find the probability that

    a) Jack's score is higher (assume independence between Jack and Jill).
    b) the total of their scores is above 350.

8) As we did in class, use http://onlinestatbook.com/stat_sim/sampling_dist/index.html to create a weird looking parent population distribution in the top graph that is NOT normal (i.e., heavily skewed, uniform, bimodal). Using the same population distribution for each, construct the distribution of sample means for n=2,n=10 and n=25. Take 10,000 samples.

    a) Include a screen shot of your parent population and your three distributions of sample means here.
    b) How are your three distributions of sample means similar? How are they different?
    c) Describe how your results relate to the Central Limit Theorem.

9) The Central Limit Theorem is important in statistics because _____.

    a) for any size sample, it says the sampling distribution of the sample mean is approximately normal
    b) for any population, it says the sampling distribution of the sample mean is approximately normal, regardless of the sample size
    c) for a large n, it says the sampling distribution of the sample mean is approximately normal, regardless of the population
    d) for a large n, it says the population is approximately normal

10) The weight of an adult swan is normally distributed with a mean of 26 pounds and a standard deviation of 7.2 pounds. A farmer randomly selected 36 swans and loaded them into his truck. What is the probability that this *flock* of swans weighs > 1000 pounds?

11) The reading speed of second grade students is approximately normal, with a mean of 88 words per minute (wpm) and a standard deviation of 12 wpm.

    a) What is the probability a randomly selected student will read more than 95 words per minute?
    b) What is the probability that a random sample of 12 second grade students results in a mean reading rate of more than 95 words per minute?
    c) What is the probability that a random sample of 24 second grade students results in a mean reading rate of more than 95 words per minute?
    d) What effect does increasing the sample size have on the probability? Provide an explanation for this result.

12) The mean selling price of senior condominiums in Green Valley over a year was \$215,000. The population standard deviation was \$ 25,000. A random sample of 100 new unit sales was obtained.

a) What is the probability that the sample mean selling price was more than \$210,000?
b) What is the probability that the sample mean selling price was between \$ 213,000 and \$ 217,000?
c) Suppose that, after you had done these calculations, a friend asserted that the population distribution of selling prices of senior condominiums in Green Valley was almost certainly not normal. How would you respond?
d) Why can't you answer questions about the probability an indivdual condominium sells for more than \$210,000?