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

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

Not independent. May justify it several ways i.e.:

Marginal Probability P(X=1) = P(X=1 and Y=1) + P(X=1 and Y=2) = ¼ + ½ = 3/4 . P(Y=1) = P(X=1 and Y=1) + P(X=2 and Y=1) = ¼ + 1/8 = 3/8.

We must test if P(X=1 and Y=1) = P(X=1) * P(Y=1). P(X=1 and Y=1) = 1/4 . P(X=1) * P(Y=1) = 3/4 * 3/8 = 0.28 which does not equal ¼ so they are not independent.

b) Compute P(XY<3)

1-1/8 = 7/8

c) Compute P(X+Y>2)

1-1/4 = ¾

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 $250, whereas for a homeowner's policy, the choices are 0, $100, and $200. [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?
Mean(X)= 100 (0.48) + 200 (0.52) = 152.
Standard deviation: (100-152)^2 (0.48) + (200-152)^2 (0.52) = 2496 → square root (2496) = 49.96

b) What are the mean and standard deviation of Y?
Mean of Y = 0* 0.35 + 100*.4 + 300*0.25 = 115
Standard deviation: (0-115)^2 * 0.35 + (100-115)^2 * 0.4 + (300-115)^2 * 0.25 = 13275→ sqrt(13275) = 115.22

c) What is the covariance between X and Y?
E(XY) – E(x) * E(y) = 0 * 0.05 + 0*.3 + 10,000 * 0.2 + 20,000*0.2 + 30,000*0.23 + 60,000 * 0.02 - (152*115) = -3380

d) What is the correlation between X and Y?
-3380/(49.96*115.22) = -0.587

e) Calculate E(X+Y). = 152+115 = 267

f) Calculate Var(X+Y). = 2496+13275+2*(-3380) = 9011

g) Are X and Y independent? Justify your answer. = No, as correlation is nonzero.

h) What is the expected value of Y given X= 200? 0*(0.3/.52) + 100*(0.2/.52)+300*(0.02/0.52) =50

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.

Var(2X + 3Y) = 4Var(X) + 9Var(Y) = 4·10·.4·.6 + 9·12·.2·.8 = 9.6 + 17.28 = 26.88

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].

Let X be the annual cost of owning a dog and Y the annual cost of owning a cat.
The annual cost of owning one dog and two cats is W = X + 2Y.

This is a normal random variable with E(W) = E(X) + 2 E(Y) = 695 + 2 (705) = 2105.
The variance of W is Var(W) = Var(X) + 4Var(Y) = 45^2 + 4(35^2) = 6925.

Thus, sW = √ 6925 = 83.21658.

We have P(W > 2000) = P ((W – 2105)/83.21658 > (2000 – 2105)/83.21658) = P(Z > −1.261768) = P(Z < 1.261768) = 0.8964839.

5) An investor plans to divide $ 200,000 between two investments. The first yields a certain 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.

Mean: 0.5*8+0.5*18 = 13% → 13%*200,000 = 26,000
SD: 0.25*25 + 0.25*36 +(2*0.5*0.5*-0.2*5*6)= 12.25

= 12.25% → Square root of 12.25% = 3.5% → 3.5%*200,000 = 7,000

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.

OIL: **iPath S&P GSCI Crude Oil TR ETN (OIL)**
GLD: **SPDR Gold Shares (GLD)**
KORS: **Michael Kors Holdings Limited (KORS)**
SPY: **SPDR S&P 500 ETF (SPY)**

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?

```
> mean(oilret)
[1] -0.03184388
> sd(oilret)
[1] 0.1067068

> mean(gldret)
[1] 0.001117114
> sd(gldret)
[1] 0.04451872

> mean(korsret)
[1] -0.007756866
> sd(korsret)
[1] 0.1143641
```

```
> mean(spyret)
[1] 0.008435564
> sd(spyret)
[1] 0.02854669
```

Correlation Matrix:

	oilret	gldret	korsret	spyret
oilret	1.00000000	-0.06358263	-0.2174105	0.2509213
gldret	-0.06358263	1.00000000	0.3308233	-0.2396120
korsret	-0.21741046	0.33082326	1.0000000	-0.1220680
spyret	0.25092128	-0.23961199	-0.1220680	1.0000000

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?

OIL:

```
> lm(formula = oilret ~ spyret)

Call:
lm(formula = oilret ~ spyret)

Coefficients:
(Intercept)       spyret
   -0.03976      0.93794
```

GLD:

```
> lm(formula = gldret ~ spyret)

Call:
lm(formula = gldret ~ spyret)

Coefficients:
(Intercept)       spyret
   0.004269    -0.373676
```

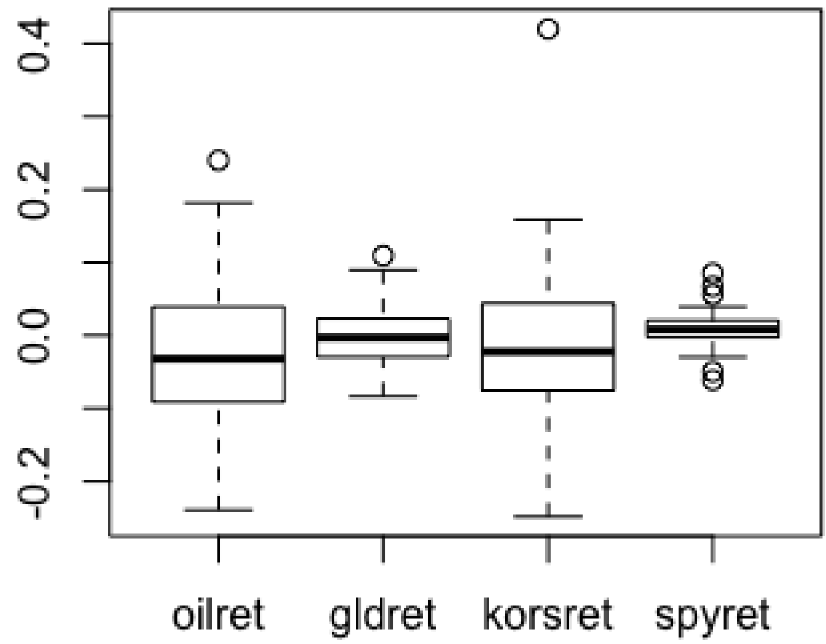
KORS:

```
> lm(formula = korsret ~ spyret)

Call:
lm(formula = korsret ~ spyret)

Coefficients:
(Intercept)       spyret
```

-0.003632   -0.489030

Smallest to Largest: KORS, GLD, OIL
This is the same as ranking by Standard Deviation.

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



GLD looks the safest due to the least amount of variance while OIL and KORS looks equally risky.

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).

> .5*mean(oilret) + .5*mean(korsret)

```
[1] -0.01980037
> sqrt(.5^2*var(oilret) + .5^2*var(korsret) + 2*.5*.5*cov(oilret, korsret))
                monthly.returns
monthly.returns       0.06920832

> .5*mean(oilret) + .5*mean(gldret)
[1] -0.01536338
> sqrt(.5^2*var(oilret) + .5^2*var(gldret) + 2*.5*.5*cov(oilret, gldret))
                monthly.returns
monthly.returns       0.05648928

> .5*mean(korsret) + .5*mean(gldret)
[1] -0.003319876
> sqrt(.5^2*var(korsret) + .5^2*var(gldret) + 2*.5*.5*cov(korsret, gldret))
                monthly.returns
monthly.returns       0.06787805
```

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

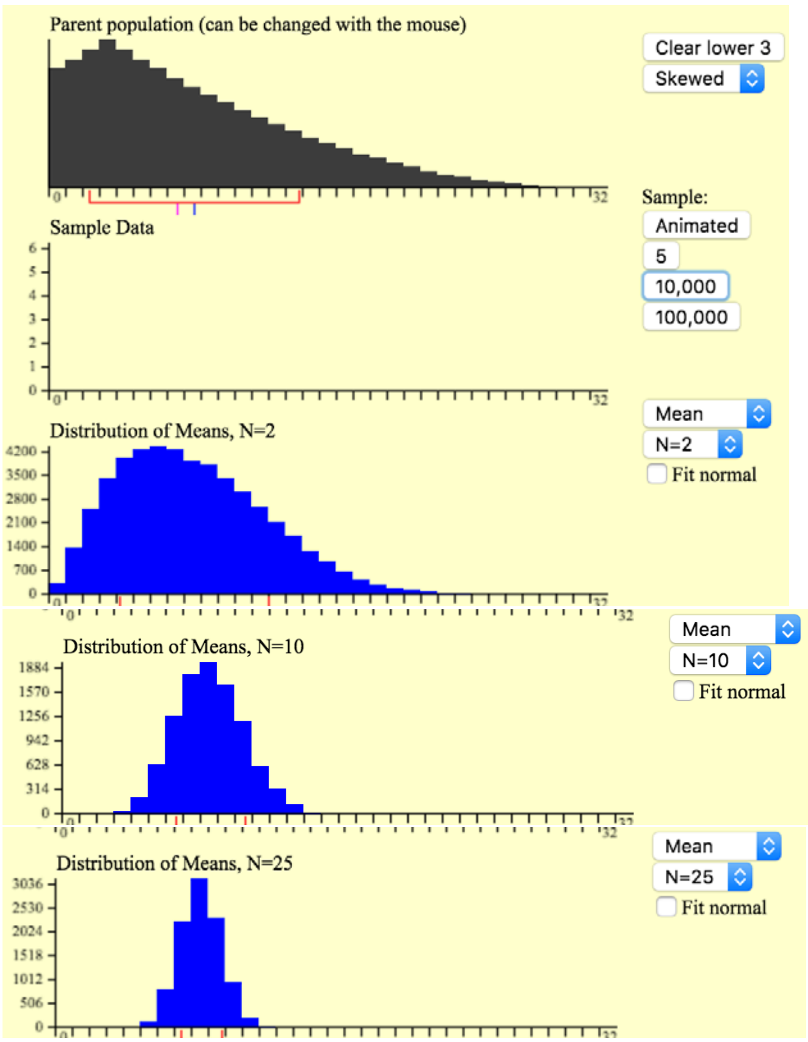
Smallest to Largest: OIL, GLD -> KORS, GLD -> OIL, KORS
Holding GLD individually looks to be the least risky but both OIL and KORS individually are more volatile than diversifying and having a portfolio.

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
N(160-170, 20^2+15^2) = N(-10,625)
P(D>0) = P(Z> (0- -10)/sqrt(625))= P(Z> 0.4) = 1-.6554 = .3446

b) the total of their scores is above 350.
N(160+170, 20^2+15^2) = N(330,625)
P(D>350) = P(Z> (350-330)/sqrt(625))= P(Z> 0.8) = 1-.7881 = .2119

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.



b) They're all relatively centered around the population mean. However, we can see that as n increases the range is much narrower and we get closer to the true population mean.

c) Even though the original population isn't normal, if we take a sample of the population and calculate the distribution of the sample mean we get a normal distribution with mean equal to the population mean. We can see that as n approaches 30 the results get better and better.

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?

To get a flock weight of more than 1000 pounds, we would need an average of 27.778 pounds for a swan.

X = weight of a swan

$$\bar{X} \sim N(26, 1.44)$$

$$\text{P}(z > \frac{27.778 - 26}{\sqrt{1.44}}) = P\left(z > \frac{1.778}{1.2}\right) = P(z > 1.48) = 0.0694$$

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?

Z = (95-88)/12 = .5833
P(X > 95) = P(Z > .5833) = .2798

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?

Z = (95-88)/(12/sqrt(12)) = 2.0207
P(X > 95) = P(Z > 2.0207) = .0217

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?

Z = (95-88)/(12/sqrt(24)) = 2.8577
P(X > 95) = P(Z > 2.8577) = .0021

d) What effect does increasing the sample size have on the probability? Provide an explanation for this result.

As the sample size increases, the probability for that sample mean being greater than 95 decrease. This is because as you take more samples, the closer you expect the sample mean to be close to the population mean. We know this through the Central Limit Theorem.

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?

$Z = (210000-215000)/(25000/\text{sqrt}(100)) = -2$
$P(X > 210000) = P(Z > -2) = .97725$

b) What is the probability that the sample mean selling price was between $ 213,000 and $ 217,000?

$Z = (213000-215000)/(25000/\text{sqrt}(100)) = -.8$
$Z = (217000-215000)/(25000/\text{sqrt}(100)) = .8$

$P(213000 < X < 217000) = P(-.8 < Z < .8) = .57629$

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?

Based on the Central Limit Theorem the sample mean would still approximate the population mean.

d) Why can’t you answer questions about the probability an individual condominium sells for more than $210,000?

The population of condominium’s is not normally distributed.