Homework 5
STAT 104 - Introduction to Quantitative Methods for Economics

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

$P(X = 1 \text{ and } Y = 1) = 1/4 \qquad P(X = 1 \text{ and } Y = 2) = 1/2$
$P(X = 2 \text{ and } Y = 1) = 1/8 \qquad P(X = 2 \text{ and } Y = 2) = 1/8$

$E[X] = 1*0.75 + 2 * 0.25 = 1.25$
$E[X^2] = 1*0.75 + 4 * 0.25 = 1.75$
$E[Y] = 1*0.375 + 2*0.125 = 0.625$
$E[Y^2] = 1*0.375 + 4*0.125 = 0.875$
$Var[X] = 1.75 - (1.25)^2 = 0.1875$
$Var[Y] = 0.875 - (0.625)^2 = 0.4843$

a) Are X and Y independent? Justify your answer

→ X and Y are not independent

Marginal probability $P(Y \mid X = 1) = 0.25 + 0.5 = 0.75$
Marginal probability $P(X \mid Y = 2) = 0.5 + 0.125 = 0.625$

For X and Y to be independent, $P(X = 1 \text{ and } Y = 2) = 0.75 * 0.625$
But that's not the case.

b) Compute P(XY<3)

c) Compute P(X+Y>2)

→ Say $S = X+Y \sim N(1.25 + 0.625, 0.1875 + 0.4843) = N(1.875, 0.82^2)$

Standardization:

$$P(S < 2) = P\left(\frac{S - 1.875}{0.82} < \frac{2 - 1.875}{0.82}\right)$$
$$= P(Z < 0.1524)$$
$$= 0.5596$$

$\therefore P(S > 2) = 1 - 0.5596 =$ **0.4404**

Homework 5
STAT 104 - Introduction to Quantitative Methods for Economics

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	0.48
	200	0.3	0.2	0.02	0.52
		0.35	0.4	0.25	1

a) What are the mean and standard deviation of X?

→ Mean of X = 100 * 0.48 + 200 * 0.52 = 48 + 104 = 152
Variance of X = $(100 – 152)^2$ * 0.48 + $(200 – 152)^2$ * 0.52
= 1297.92 + 1198.08 = 2496
Standard deviation of X = $(2496)^{1/2}$ = 49.95

b) What are the mean and standard deviation of Y?

→ Mean of Y = 100 * 0.4 + 300 * 0.25 = 40 + 75 = 115
Variance of Y = $(0 - 115)^2$ * 0.35 + $(100 – 115)^2$ * 0.4 + $(300 – 115)^2$ * 0.25
= 4628.75 + 90 + 8556.25 = 13275
Standard deviation of Y = $(13275)^{1/2}$ = 115.217

c) What is the covariance between X and Y?

→ (100 - 152) (0 - 115) * 0.05 + (200 - 152) (0 - 115) * 0.3 + (100 - 152) (100 - 115) * 0.2 + (200 - 152) (100 - 115) * 0.2 + (100 - 152) (300 - 115) * 0.23 + (200 - 152) (300 - 115) * 0 .02
= (-52 * -115 * 0.05) + (48 * -115 * 0.3) + (-52 * 15 * 0.2) + (48 * 15 * 0.2) + (-52 * 185 * 0.23) + (48 * 185 * 0.02)
= -299 -1656 – 156 + 144 – 2212.6 + 177.6 = - 4002

d) What is the correlation between X and Y?

→ Cov(X,Y) / Std Dev X * Std Dev Y
= -4002 / 115.217 * 49.95 = -0.6953

e) Calculate E(X+Y).

→ E(X+Y) = E(X) + E(Y) = 152 + 115 = 267

f) Calculate Var(X+Y).

→ Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)
= 13275 + 2496 + 2 * - 4002 = 7767

g) Are X and Y independent? Justify your answer.

→ Given value of P(X = 100 and Y =0) = 0.05
Expected value of P(X = 100 and Y =0) = 0.48 * 0.35 = 0.168

Joint probability of X =100 and Y =0 is not equal to expected value based on marginal probability and hence X and Y are not independent.

h) What is the expected value of Y given X= 200?

→ E(Y | X = 200) = ∑ y P(Y = y | X = 200) for all values of Y
= 0 * P(Y=0|X=200) + 100 * P(Y=100|X=200) + 300 * P(Y=300|X=200)
= 0 + 100*0.5 + 300*0.08
= 74

Homework 5

STAT 104 - Introduction to Quantitative Methods for Economics

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 .

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

→

$Var(X) = npq = 10 * .4 * .6 = 2.4$

$Var(Y) = npq = 12 * .2 * .8 = 1.92$

$VAR(2X + 3Y) = 4(Var(X)) + 9(Var(Y)) = 4 * 2.4 + 9 * 1.92 = 26.88$

Homework 5
STAT 104 - Introduction to Quantitative Methods for Economics

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].

→

$D \sim N(695, 45^2)$
$C \sim N(705, 35^2)$

$P(D + 2C > 2000) = ?$

Say, $D+2C = X$
$\overline{x} \sim N\,(695 + 2 * 705,\ 45^2 + 4 * 35^2) = N(2105,\ 83.22^2)$

$$P(\overline{x} > 2000) = P\left(\frac{x - 2105}{83.22} > \frac{2000 - 2105}{83.22}\right)$$
$$= P(Z > -1.26)$$
$$= 1 - P(Z < -1.26)$$
$$= 1 - 0.1038$$
$$= \mathbf{0.8962}$$

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.

Return on first investment = \$8,000

Return on second investment = \$18,000

Mean of the total Profit = .5 * 8000 + .5 * 18000 = \$13,000

Correlation of total profit = $(.5)^2\,(.05)^2 + (.5)^2\,(0.6)^2 + 2\,(.5)\,(.5)\,(-0.2)\,(0.05)\,(0.06)$

$= (.25)\,(.0025) + (.25)\,(.0036) - .0003$

$= .000625 + .0009 - .0003 = 0.001225$

Standard deviation of total profit = $(0.001225)^{1/2} = 0.035$

Homework 5
STAT 104 - Introduction to Quantitative Methods for Economics

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

GLD - SPDR Gold Shares

KORS - Michael Kors Holdings Limited

SPY - SPDR S&P 500 ETF

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?

→
```
> monthlyReturnSPY = monthlyReturn(SPY)
> monthlyReturnKORS = monthlyReturn(KORS)
> monthlyReturnGLD = monthlyReturn(GLD)
> monthlyReturnOIL = monthlyReturn(OIL)

> mean(monthlyReturnSPY )
[1] 0.006687744

> mean(monthlyReturnKORS  )
[1] -0.007527873

> mean(monthlyReturnGLD)
[1] 0.001358516

> mean(monthlyReturnOIL)
[1] -0.03203832
```

Standard deviation for the returns of the stocks

```
> sd(monthlyReturnGLD)
[1] 0.04467074

> sd(monthlyReturnSPY)
```

```
[1] 0.02905614

> sd(monthlyReturnOIL)
[1] 0.1067084

> sd(monthlyReturnKORS)
[1] 0.1142621
```

Correlation between all the stocks

```
> cor(monthlyReturnSPY, monthlyReturnKORS)
                monthly.returns
monthly.returns      -0.1213283

> cor(monthlyReturnSPY, monthlyReturnGLD)
                monthly.returns
monthly.returns       -0.220175


> cor(monthlyReturnSPY, monthlyReturnOIL)
                     monthly.returns
monthly.returns       0.2522446

> cor(monthlyReturnKORS, monthlyReturnGLD)
                monthly.returns
monthly.returns       0.3307882

> cor(monthlyReturnKORS, monthlyReturnOIL )
                monthly.returns
monthly.returns      -0.2133613

> cor(monthlyReturnGLD, monthlyReturnOIL )
                monthly.returns
monthly.returns     -0.05703014
```

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?

→

Homework 5
STAT 104 - Introduction to Quantitative Methods for Economics

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).

→
E(OIL,KORS) = 0.5 * -0.03203832 + 0.5 * -0.007527873
= -0.016019 – 0.003764 = -0.019783
E(OIL, GLD) = 0.5 * -0.03203832 + 0.5 * 0.001358516
= -0.016019 + 0.000679 = -0.015340
E(KORS, GLD) = 0.5 * -0.007527873 + 0.5 * 0.001358516
= – 0.003764 + 0.000679 = -0.003085

StandardDev(OIL,KORS) = $(0.5^2 * (0.1067084)^2 + 0.5^2 * (0.1142621)^2 + 2 * 0.5 * 0.5 * 0.1067084 * 0.1142621 * -0.2133613)^{1/2}$
$= (.25 * 0.011387 + .25 * 0.013056 – 0.0001301)^{1/2}$
$= (0.0002847 + 0.003264 – 0.0001301)^{1/2}$
= 0.058469
StandardDev(OIL, GLD) = $(0.5^2 * (0.1067084)^2 + 0.5^2 * (0.04467074)^2 + 2 * 0.5 * 0.5 * 0.1067084 * 0.04467074 * -0.05703014)^{1/2}$
$= (.25 * 0.011387 + .25 * 0.001995 – 0.000068)^{1/2}$
$= (0.0002847 + 0.0000499 – 0.000068)^{1/2}$
= 0.016328
StandardDev(KORS, GLD) = $(0.5^2 * (0.1142621)^2 + 0.5^2 * (0.04467074)^2 + 2 * 0.5 * 0.5 * 0. 1142621* 0.04467074 * 0.3307882)^{1/2}$
$= (0.003264 + 0.0000499 + 0.000422)^{1/2}$
= 0.061122

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

→
Portfolio (`OIL, GLD`)

Portfolio (`OIL, KORS`)

Portfolio (`KORS, GLD`)

Homework 5
STAT 104 - Introduction to Quantitative Methods for Economics

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

Jack ~ $N(170, 20^2)$
Jill ~ $N(165, 15^2)$

a) Jack's score is higher (assume independence between Jack and Jill).

→ P(Jack > Jill)
Let D = (Jill-Jack) be the difference in Jack and Jill's scores
So we want to find $P(D < 0)$

From rule of combining normals, we know
D = Jill - Jack ~ $N(165\text{-}170, 15^2 + 20^2) = N(-5, 625)$

Standardization:

$$P(D < 0) = P\left(\frac{D + 5}{\sqrt{(625)}} < \frac{0 + 5}{\sqrt{(625)}}\right)$$
$$= P(Z < 0.2)$$
$$= \mathbf{0.5793}$$

b) the total of their scores is above 350.

→ P(Jack + Jill > 350)
Let S = (Jack+Jill) be the sum of Jack and Jill's scores
So we want to find $P(S > 350)$ i.e. $1 - P(S < 350)$

From rule of combining normals, we know
S = Jill + Jack ~ $N(165+170, 15^2 + 20^2) = N(335, 625)$

Standardization:

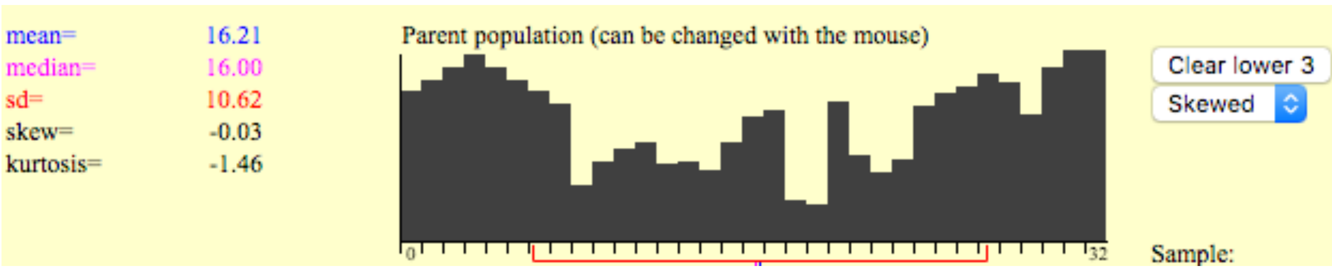
$$P(S < 350) = P\left(\frac{S - 335}{\sqrt{(625)}} < \frac{350 - 335}{\sqrt{(625)}}\right)$$
$$= P(Z < 0.6)$$
$$= 0.7257$$

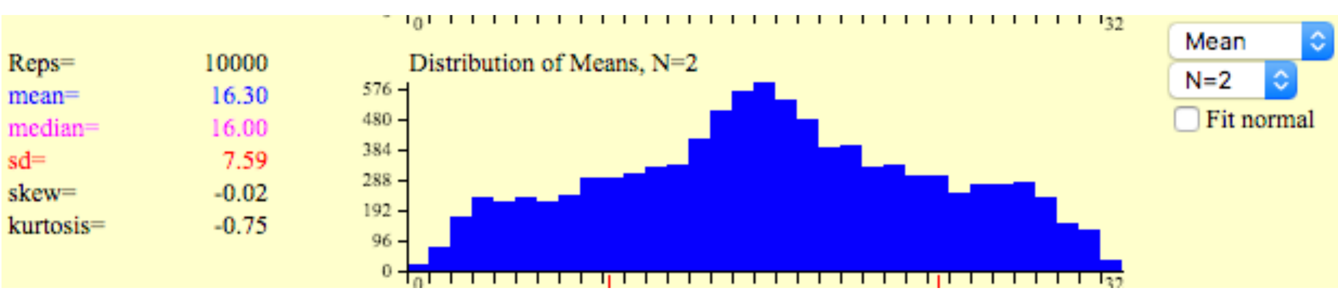
$\therefore P(S > 350) = 1 - 0.7257 = 0.2743$

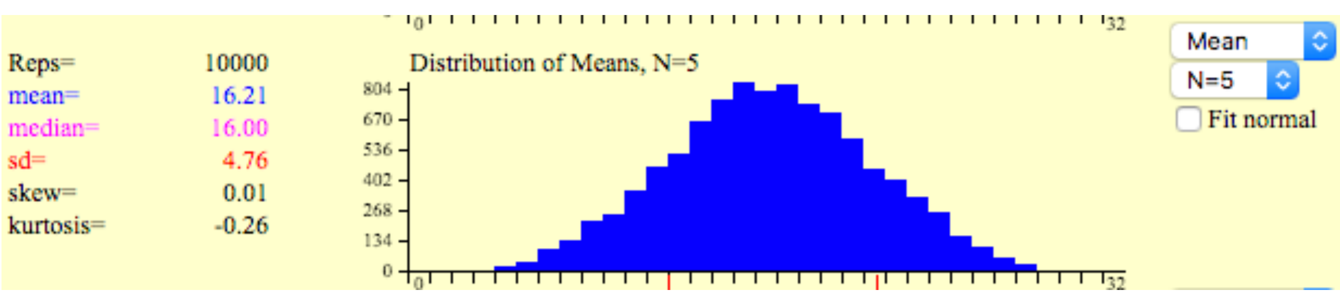
Homework 5
STAT 104 - Introduction to Quantitative Methods for Economics

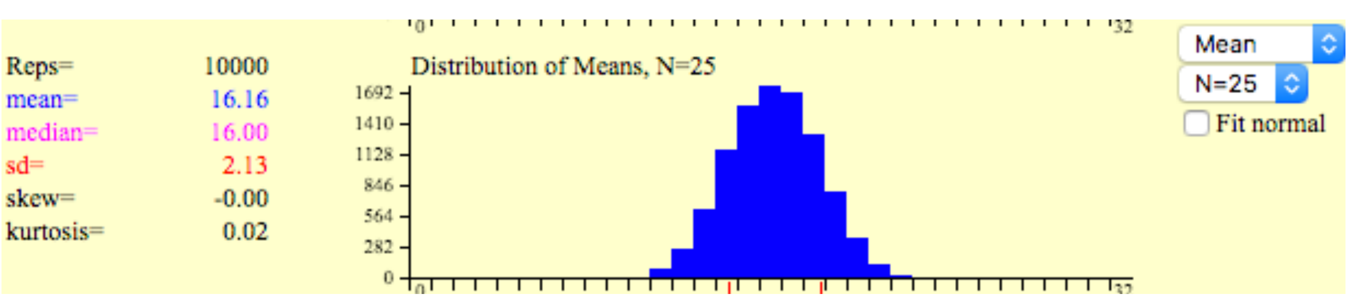
8) As we did in class, use http://onlinestatbook.com/stat_sim/sampling_dist/index.html to create a weird looking parent population distribution in the top graph that is NOT normal (i.e., heavily skewed, uniform, bimodal). Using the same population distribution for each, construct the distribution of 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?

→ Similarity: They sort of resemble a normal distribution
Difference: As N increases, height of distribution increases and width, decreases.

c) Describe how your results relate to the Central Limit Theorem.

→ Even though nothing can be said about the population, once we take samples, the means follow a normal distribution. As the number of samples increases, the normal distribution becomes thinner and taller. Each of the normal distributions is centered around the population mean and the variance decreases as the number of samples increases i.e. we get closer to the population mean.

Homework 5
STAT 104 - Introduction to Quantitative Methods for Economics

9) The Central Limit Theorem is important in statistics because for a large n, it says the sampling distribution of the sample mean is approximately normal, regardless of the population.

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

Homework 5
STAT 104 - Introduction to Quantitative Methods for Economics

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?

→
$X \sim N(26, 7.2^2)$
$N = 36$
$P(\sum X > 1000) = ?$

Dividing both sides by sample size, 36
$P(\sum X > 1000) = P(\sum X/36 > 1000/36)$
$= P(\overline{x} > 27.78)$
$= 1 - P(\overline{x} < 27.78)$

From the Central Limit Theorem, we know
$\overline{x} \sim N(26, \frac{7.2^2}{36}) = N(26, 1.2^2)$

Standardization:
$P(\overline{x} < 27.78) = P(\frac{x - 26}{1.2}) < P(\frac{27.78 - 26}{1.2})$
$= P(Z < 1.48)$
$= 0.9306$

$\therefore$ $P(\sum X > 1000) = 1 - 0.9306 =$ **0.0694**

There is a 6% chance that this flock of swans weigh more than 1000 pounds.

Homework 5
STAT 104 - Introduction to Quantitative Methods for Economics

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.
$X \sim N(88, 12^2)$

a) What is the probability a randomly selected student will read more than 95 words per minute?

$\rightarrow P(x > 95) = ?$
Standardization:
$$P(x > 95) = P\left(\frac{x - 88}{12} > \frac{95 - 88}{12}\right)$$

$= P(Z > 0.583)$
$= 1 - P(Z < 0.583)$
$= 1 - 0.7190$
$= \mathbf{0.281}$

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?

$\rightarrow$ If $n = 12$, $P(\bar{x} > 95) = ?$
$\bar{x} \sim N(88, 12^2/12) = N(88, 3.46^2)$

$$P(\bar{x} > 95) = P\left(\frac{x - 88}{3.46} > \frac{95 - 88}{3.46}\right)$$
$= P(Z > 2.023)$
$= 1 - P(Z < 2.023)$
$= 1 - 0.9783$
$= \mathbf{0.0217}$

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?

$\rightarrow$ If $n = 24$, $P(\bar{x} > 95) = ?$
$\bar{x} \sim N(88, 12^2/24) = N(88, 2.45^2)$

$$P(\bar{x} > 95) = P\left(\frac{x - 88}{2.45} > \frac{95 - 88}{2.45}\right)$$
$= P(Z > 2.86)$
$= 1 - P(Z < 2.86)$
$= 1 - 0.9979$
$= \mathbf{0.0021}$

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

→ Increasing the sample size, reduces the probability.
Explanation:

As the sample size increases, the sample mean gets closer to the population mean, hence reducing the variance. This causes the probability of values further away from the mean to reduce.

Homework 5
STAT 104 - Introduction to Quantitative Methods for Economics

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.

$X \sim N(215000, 25000^2)$
$n = 100$

$\bar{x} \sim N(215000, 2500^2)$

a) What is the probability that the sample mean selling price was more than $210,000?

$P(\bar{x} > 210,000) = ?$

Standardization:

$$P(\bar{x} > 210,000) = P(\frac{x - 215000}{2500} > \frac{210000 - 215000}{2500})$$
$$= P(Z > -2)$$
$$= 1 - P(Z < -2)$$
$$= 1 - 0.0228$$
$$= \mathbf{0.9772}$$

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

$$P(213000 < \bar{x} < 217000) = P(\frac{213000 - 215000}{2500} < \frac{x - 215000}{2500} < \frac{217000 - 215000}{2500})$$
$$= P(-0.8 < Z < 0.8)$$
$$= P(Z < 0.8) - P(Z < -0.8)$$
$$= 0.7881 - 0.2119$$
$$= \mathbf{0.5762}$$

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?

→ That does not impact my calculations since the sample size is large enough for Central Limit Theorem i.e. the distribution of the population does not matter.

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

→ We have no information about the distribution of the population.