Homework 3

Erik Lee

Sun Jul 15 18:23:21 2018

1) The random variable X has a binomial distribution with E(X)=18 and Var(X)=7.2. Find n and p for this distribution [use the formulas for mean and variance].

E(X) = 18 = np Var(X) = 7.2 = (np)q

Solve: 7.2 = (18)q q = 7.2/18 = 0.4 p = (1-q) = 1-0.4 = 0.6 n = 18/p = 18/0.6 = 30

Answer: p = 0.6 n = 30

Checks: E(X) = np = 30*0.6 = 18 Var(X) = npq = 30*0.6*0.4 = 7.2

---

2) Suppose X is a binomial random variable with n=15 and p=0.3. Feel free to use a computer to answer the following:

a) P(X=0)

P(X=0) = P(all failiures) = (1-P)^n = (1-0.3)^15 = 0.7^15 = 0.004747562

b) P(X=2)

```r
dbinom(2,15,0.3)
```

```
## [1] 0.09156
```

P(X=2) = 0.1268

Calculation check: P(X=2) = (n!/(x!(n-x)!)) * (p^x) * ((1-p)^n-x), where n = sample size = 15 and x = sucesses = 2, p = 0.3 and q = 1-p = 0.7

P(X=2) = factorial(15)/(factorial(2)*factorial(15-2))* (0.3^2) * ((1-0.3)^(15-2)) = 105 * 0.09 * 0.009689 = 0.09156 (checks out with the dbinom() r code above)

c) P(X<2)

If P(X<2) then it includes probabilities X=0 and X=1, but not X=2, and no negative X values; that says P(X<2) = P(X<=1) for the discrete random variables X=0 and X=1

```r
# use pbinom to find the cumulative discrete probabilities for everything less than or equal to 1 P(X<=
pbinom(1,15,0.3)
```

```
## [1] 0.03527
```

P(X<2) = P(X<=1) = 0.03527

Check using the sum of P(X=0) and P(X=1)):

```r
dbinom(0,15,0.3) + dbinom(1,15,0.3) # add the dbinoms
```

```
## [1] 0.03527
```

```r
sum(dbinom(0:1,15,0.3)) # sum the dbinom from 0:1
```

```
## [1] 0.03527
```

d) P(X>8)

P(X>8) = 1-P(X<=8) = 0.01524

Using the complement of P(X<=8) and the cumulative distribution for a random binomial variable (pbinom() function), the P(X>8) can be found for all values where X is greater than 8.

```
# The complement of the cumulative distribution from X<=8 will give the cumulative distribution of all
1-pbinom(8,15,0.3)
```

```
## [1] 0.01524
```

```
# Summing the distributions of values greater than 8 will also give the same probability (9<=X<=15)
sum(dbinom(9:15,15,0.3))
```

```
## [1] 0.01524
```

e) E(X)

E(X) = np = 15*0.3 = 4.5

f) Var(X)

Var(X) = npq = np(1-p) = 15*0.3*(0.7) = 3.15

---

3) A school newspaper reporter decides to randomly survey 12 students to see if they will attend YardFest festivities this year. Based on past years, she knows that 22 percent of students attend YardFest festivities. We are interested in the number of students who will attend the festivities.

a) How many of the 12 students do we expect to attend the festivities?

E(X) = n*p = 12*0.22 = 2.64

b) Find the probability that at most 4 students will attend.

P(X<=4) = 0.8979

Solve this by finding the cumulative distributive function of P(X<=4). This includes the range from 0 students up to 4 students (0<=X<=4). The R code below uses pbinom() to calculate the cdf up to 4:

```
pbinom(4,12,0.22)
```

```
## [1] 0.8979
```

```
# summing the individual probabilities from 0:4
sum(dbinom(0:4,12,0.22))
```

```
## [1] 0.8979
```

c) Find the probability that more than 2 students will attend.

P(X>2) = 0.5114

This can be found by using the complement of P(X<=2).

```
# Complement of cdf at P(X<=2) will yield all the sum of the probabilites greater than X=2
1-pbinom(2,12,0.22)
```

```
## [1] 0.5114
```

```
# Same answer using the complement of the sum of the dbinom() from 0<=X<=2
1-sum(dbinom(0:2,12,0.22))
```

```
## [1] 0.5114
```

---

4) Suppose that about 90% of graduating students attend their graduation. A group of 22 graduating students is randomly chosen.

a) How many are expected to attend their graduation?

E(X) = n*p* = *22*0.9 = 19.8

b) Find the probability that 17 or 18 attend.

P(17 or 18) = P(X=17) + P(X=18) =0.1537

This can be computed by the sum of the dbinom() function for X=17 and X=18

```
dbinom(17,22,0.9) + dbinom(18,22,0.9)
```

```
## [1] 0.1537
```

```
# or
sum(dbinom(17:18,22,0.9))
```

```
## [1] 0.1537
```

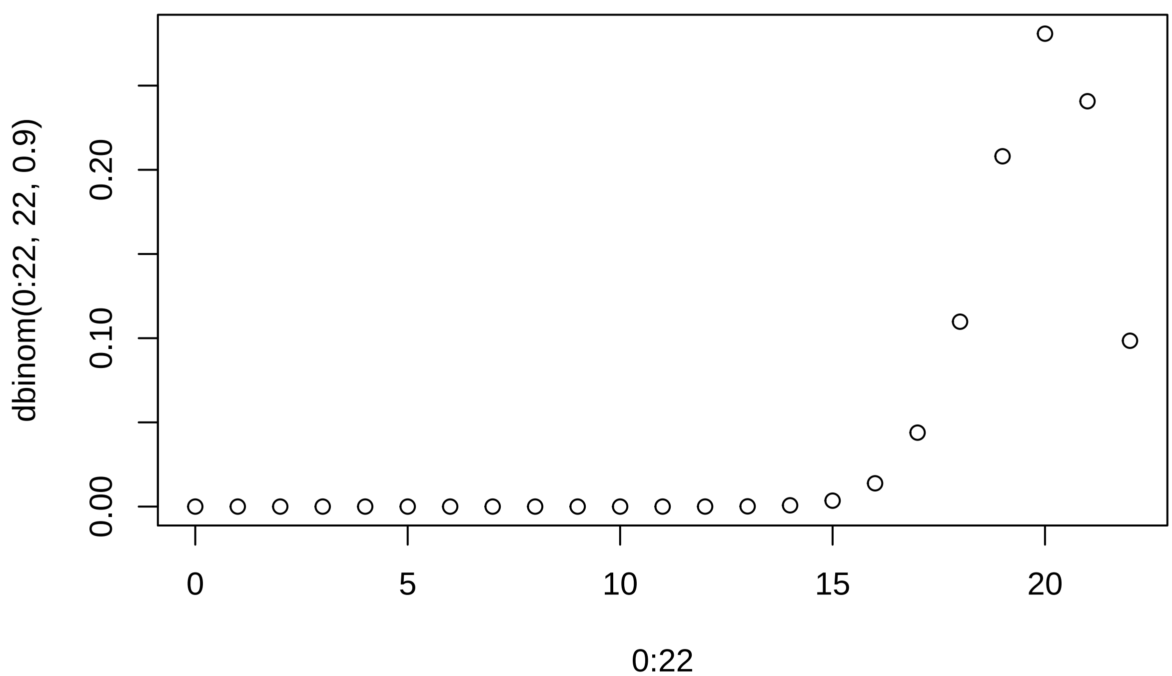
c) Based on numerical values, would you be surprised if all 22 attended graduation? Justify your answer numerically.

Yes, I would be suprised if all 22 students attended graduation. This statement essentially asks for P(X=22), where X represents the random variable of all 22 students in attendence. If this parameter is plugged into dbinom() along with a population of 22 and probability of attendence of 0.9, P(X=22) = 0.09848 ~ 9.8%. This result is a low percentage indicating unlikelihood of all 22 students attending.

```
# Calculate probability of 22 of 22 students attending with 0.9 chance a given student attends
dbinom(22,22,0.9)
```

```
## [1] 0.09848
```

```
# If this distribution is graphed from 0 students to 22 students, we see that at X=22, the probability
plot(0:22,dbinom(0:22,22,0.9))
```



A 9.8% probability for 22 students sounds odd for such a high probability of a given student to attend (90%).

However, since the trials are independent, the probabilites of multiple students attending are a product of their probabilites. This product can be equated to taking the given probability as an exponent (0.9)^X, where X represents the number of students attending. Taking the exponent of a fraction causes the number to decrease as the exponent increases, so (0.9)^22 = 0.09848, quite a small number in comparison to the original probability.

This is an example of an endpoint probability. This is the probability of all successes where all students attend, P(all success) = P(X=22) = p^n = 0.9^22 = 0.09848.

---

5) According to flightstats.com, American Airlines flights from Dallas to Chicago are on time 75% of the time. Suppose 15 flights are randomly selected, and the number of on-time flights is recorded.

a) Explain why this is a binomial experiment.

This scenario fits several key components of a bionmial experiment. The probability (p) of success, an on-time flight in this case, is fixed at 75%. There are a fixed number (n) of randomly selected samples, 15 flights. The observations (flights) are independent, where the timeliness of one flight does not affect another flights timeliness. And there are two possible results for success or faliure, the flight on time or the flight not on time respectively. From these, the scenario fits a binomial experiment.

b) Find and interpret the probability that exactly 10 flights are on time.

P(X=10) = 0.1651

```r
# dbinom() for probability of 10 sucesses, n=15, p=0.75
dbinom(10,15,0.75)
```

```
## [1] 0.1651
```

The condition that 10 flights out of the 15 randomly selected flights arrive on time has a probability of 0.1651 or 16.51%. This accounts for 10/15 being on time and 5/15 not being on time.

c) Find and interpret the probability that between 8 and 10 flights, inclusive, are on time.

P(8<=X<=10) = 0.2962

```r
# dbinom() for X in interval [8,10] inclusive, n=15, p=0.75
sum(dbinom(8:10,15,0.75))
```

```
## [1] 0.2962
```

This condition states that the probability that 8, 9, or 10 flights successfully land on time, out of the 15 randomly selected flight, is 0.2962 or 29.62%. Put another way, there is a 29.62% chance that between 8 and 10 flights land on time, out of the 15 randomly selected flights on record.

d) Find the mean and variance of the number of on time flights

Mean(X) = np = n*p* = *15*0.75 = 11.25

Var(X) = npq = n*p*(1-p) = 15*0.75*(0.25) = 2.812

---

6) Let X be a uniformly distributed random variable on the interval 0 to 1

a) Calculate P(X = 0.25)

P(X=0.25) = 0.25

b) Calculate P(0.7 < X < 1)

P(0.7<X<1) = 1-P(X<0.7) = 1-0.7 = 0.3

c) Calculate the expected value of X

E(X) = (a+b)/2 = (0+1)/2 = 0.5

d) Calculate the variance of X

Var(X) = (b-a)^2 / 2 = (1-0)^2 / 2 = 1/2 = 0.5

---

7) For the standard normal random variable Z, compute the following:

a) $P(0 \leq Z \leq 0.73)$

P(0<=Z<=0.73) = P(Z<0.73) - P(Z<0) = 0.7673 - 0.5000 = 0.2673

b) $P(-1.5 \leq Z \leq 0)$

P(-1.5<=Z<=0) = P(Z<0) - P(Z<-1.5) = 0.5000 - 0.0668 = 0.4332

c) $P(Z > 0.44)$

P(Z>0.44) = 1-P(Z<0.44) = 1 - 0.6700 = 0.33

d) $P(-1.5 \leq Z \leq 0.4)$

P(-1.5<=Z<=0.4) = P(Z<0.4) - P(Z<-1.5) = 0.6554 - 0.0668 = 0.5886

e) $P(Z < 5.23)$

P(Z<5.23) = 1

f) $E(4 - 3Z)$

Z=X a+bX, a=4, b=-3 E(X) = 4+(-3)(0) = 4

g) $Var(4 - 3Z)$

Z=X Var(a+bX) = (b^2)(X) Var(X) = ((-3)^2)(1) = 9

---

8) The time needed to hand stitch a Swoosh soccer ball is normally distributed with mean 43 minutes and standard deviation 3.1 minutes. If 3 workers start at the same time, what is the probability that at least one of them will complete their soccer ball in under 44 minutes?

P(X<44) = 0.6265

This is solved with the function pnorm(), where X = 44, mean = 43, s.d. = 3.1

```
pnorm(44,43,3.1)
```

```
## [1] 0.6265
```

The probabilitiy 1 of the workers completes a soccer ball under 44 minutes is 0.6265 or 62.65%.

---

9) The weight of reports produced in a certain department has a Normal distribution with mean 60g and standard deviation 12g. What is the probability that the next report will weigh less than 45g?

P(X<45) = 0.1056

This is solved with the function pnorm(), where X = 45, mean = 60, s.d. = 12

```r
pnorm(45,60,12)
```

```
## [1] 0.1056
```

The probability of the next report weighing less than 45g is 0.1056 or 10.56%.

---

10) The times taken to complete an introduction to business statistics exam have a normal distribution with a mean of 65 minutes and standard deviation of 7 minutes. There are 150 students who took the exam and students are allowed a total of 75 minutes to take the exam.

a) What is the chance that Mike finished his exam in 63 to 72 minutes?

P(63<X<72) = 0.4538

This can be solved using the pnorm() function to solve for when X is between 63 and 72 minutes:

```r
# Since pnorm() calculates the probability distribution of a normal curve at values equal to or less th

# The concept above translate to the calculation P(63<=X<=72) = P(X<72) - P(X<63). These values can be

pnorm(72,65,7) - pnorm(63,65,7)
```

```
## [1] 0.4538
```

The probability Mike finished his exam in 63 to 72 minutes is 0.4538 or 45.38%.

b) What is the expected number of students who finished in less than 75 minutes?

P(X<75) = 0.9234

E(X=75) = 0.9234*150

This uses the pnorm() function with x = 75, mean = 65, and s.d. = 7

```r
# This is the probabilty a randomly selected student will finish under 75 minutes
pnorm(75,65,7)
```

```
## [1] 0.9234
```

```r
# Expected number of students to finish in under 75 minutes; this is the probability of a student finis
pnorm(75,65,7)*150 # n*p where n=150, p=0.9234
```

```
## [1] 138.5
```

The expected number of students who finish in less than 75 minutes is 138.5. This comes from the product of the probability a given student of said distribution finishes in less than 75 minutes (p = 0.9234 or 92.34%) and the total number of students (n = 150).

It makes sense that such a large percentage of students are expected to finish in time since the teacher/professor has set a time limit of 75 minutes to complete the exam.

c) As some students were not able to finish the exam in time, the instructor allowed 6 more minutes. Given he already spent 75 minutes on the exam, what is the chance that Chris finished his exam in extended time, that is between 75 and 81 minutes

P(75<X<81) = 0.06543

This can be solved using the same method as part a. Using the pnorm() function for the interval between 75 and 81 minutes, mean of 65, and standard deviation of 7:

```r
# To calculate the probabilities for the interval between 75 and 81 for this distribution, find the dif
pnorm(81,65,7) - pnorm(75,65,7)
```

```
## [1] 0.06543
```

```r
pnorm(81,65,7)
```

```
## [1] 0.9889
```

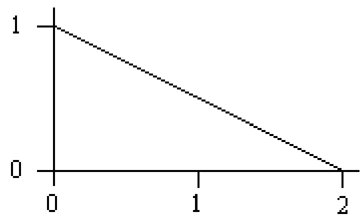
```r
pnorm(75,65,7)
```

```
## [1] 0.9234
```

The probability that Chris finishes his exam in the time between 75 and 81 minutes is 0.06543 or 6.543%. This is a low percentage so the odds are not in Chris's favor. The instructor was generous giving 6 more minutes to finish the exam.

11) Suppose X is a continuous random variable taking values between 0 and 2 and having the probability density function below. Calculate $P(1 \leq X \leq 2)$

P(1<=X<=2) = (0.5)bh = (0.5)(2-1)(0.5) = 0.25



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

P(X=1 and Y=1) = 1/4

P(X=1 and Y=2)=1/2

P(X=2 and Y=1) = 1/8

P(X=2 and Y=2)=1/8

a) Are X and Y independent? Justify your answer

No, X and Y are not independent. This can be checked by the condition that, if X and Y are independent, then P(X=1 and Y=1) = P(X=1)*P(Y=1). Using the values above and a 2x2 table, P(X=1 and Y=1) = 1/4, P(X=1) = 1/4 + 1/2 = 6/8, and P(Y=1) = 1/4 + 1/8 = 3/8. The product P(X=1)*P(Y=1) = (6/8)*(3/8) = 18/64 = 9/32. Since the product 9/32 does not equal 1/4 so P(X=1)*P(X=2) does not equal P(X=1 and Y=1), it follows that X and Y are not independent.

b) Compute P(XY<3)

P(XY<3) = P(X=1 and Y=1) + P(X=1 and Y=2) + P(X=2 and Y=1) = 1/4 + 1/2 + 1/8 = 7/8

c) Compute P(X+Y>2)

P(X+Y>2) = P(X=1 and Y=2) + P(X=2 and Y=1) + P(X=2 and Y=2) = 1/2 + 1/8 + 1/8 = 6/8 = 3/4

---

13) 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:

			Y	
	P(x,y)	0	100	300
X	100	0.1	0.15	0.22
	200	0.3	0.2	0.03

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

E(X) = 153 S.D.(X) = 49.91

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

E(Y) = 128 S.D(X) = 119.3

c) What is the covariance between X and Y?

Cov(X and Y) = E(XY) - E(X)E(Y) = 15400 - (153)(128) = 15400 - 19584 = -4184

d) What is the correlation between X and Y?

Cor(X and Y) = Cov(X and Y)/(SD(X)*SD(Y)) = -4184/(49.91*119.3) = -4184/5954.26 = -0.7027

e) Calculate E(X+Y).

E(X+Y) = E(X) + E(Y) = 153 + 128 = 281

f) Calculate Var(X+Y).

Var(X+Y) = Var(X) + Var(Y) + 2Cov(X and Y) = S.D(X)^2 + S.D.(Y)^2 2Cov(X and Y) = (49.94^2) + (119.3^2) + 2*(-4184) = 2491.0036 + 14232.49 - 8368 = 8358

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

No, X and Y are independent. To test this, the following condition must be true for every joint probability found within the table (prob19.png): the probability P(X=x and Y=y) = P(X=x)*P(Y=y). This states that the joint probability of random variables X and Y at values x and y, respectively, must be equal to the product of the marginal probabilties of X at value x and Y at value y, for every value of x and y possible.

By testing the first cell P(X=100 and Y=0), we find this condition to be false: P(X=100 and Y=0) = 0.1 P(X=100) = 0.47 P(Y=0) = 0.4 P(X=100)*P(Y=0) = 0.4*70.4 = 0.188

Since 0.1 and 0.188 are not equal, the joint and marginal probabilites for X and Y are not equal. This indicates that X and Y are not independent.

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

E(Y|X=200) = (0.3)(0) + (0.2)(100) + (0.03)(300) = 29

---

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

Var(X) = npq = (12)(0.4)(1-0.4) = 2.88 Var(Y) = npq = (10)(0.2)(1-0.2) = 1.6

Var(2X) = (2^2)(2.88) = 11.52 Var(3Y) = (3^2)(1.6) = 14.4

2X + 3Y = 11.52 + 14.4 = 25.92

---

15) 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

Answer is C

---

16) 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?

We want to find the probability that the 36 randomly selected swans have a total combined weight of 1000 pounds. This can be written as the probability that the sum of the weights > 1000. If both sides are divided by 36, P(X_bar>27.78). The averages for weights of the randomly selected swans has to be greater than or equal to 27.78 pounds. Since we have a large sample size of 36, we can apply the Central Limit Theorem to solve for the probability. This can be done by Z scoring the value at 27.78.

Note that for the Central Limit Theorem, the variance is divided by n. This gives a standard of deviation of sqrt((7.2^2)/36) = 1.44

P(X_bar>27.78) = 1 - P(X_bar<27.78) = 1 - P((X_bar-26)/1.44) < ((27.78-26)/1.44)) = 1 - P(Z<1.24) = 1 - 0.8925 = 0.1075 or 10.75%.

Based on the Central Limit Theorem, the probability that this flock of 36 randomly selected swans have a weight of more than 1000 pounds is 0.1075 or 10.75%.

---

17) 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?

P(X>95) = 1 - P(X<95) = 1 - P(Z<(95-88/12)) = 1 - P(Z<0.41) = 1 - 0.6591 = 0.3409 or 34.09%.

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?

P(X_bar>95) = 1 - P(X_bar<95) = 1 - P(Z<((95-88)/(12/sqrt(12)))) = 1 - P(Z<2.02) = 1 - 0.9783 = 0.0217 or 2.17%

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?

P(X_bar>95) = 1 - P(X_bar<95) = 1 - P(Z<((95-88)/(12/sqrt(24)))) = 1 - P(Z<2.86) = 1 - 0.9979 = 0.0021 or 0.21%

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

Increasing the sample size decreases the probability that the condition will come true. Algebraically for the Central Limit Theorem, the standard of deviation is being divided by the square root of the sample size, n. As n gets larger, the value for standard of deviation decreases. This increases the Z scores as it is equal to the difference between the random variable and mean all divided by the standard of deviation and square of n. A higher Z score leads to higher probability. But since we need the complement in this case, it leads to lower probabilities.

This make some sense logically as the random sampling of second grade students get larger and larger, the likelihood that their mean reading rates are more than 95 words per minute gets lower. As more students are measured, the likelihood that they still meet the condition of more than 95 words per minute does down.

---

18) 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?

P(X_bar>210000) = 1 - P(X_bar<210000) = 1 - P(Z< (210000-215000)/(25000/sqrt(100))) = 1 - P(Z<-2.00) = 1 - 0.9772 = 0.0228 or 2.28%

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

P(213000<X_bar<217000) = P(X_bar<217000) - P(X_bar<213000)

Z score for 213000 (Z1) and 217000 (Z2) for mean=215000 and s.d.=25000

Z1 = (213000-215000)/(25000/sqrt(100)) = -2000/2500 = -0.8 Z2 = (217000-215000)/(25000/sqrt(100)) = 2000/2500 = 0.8

Substitute the Z values in for X_bar P(Z<0.8) - P(Z<-0.8) = 0.7881 - 0.2119 = 0.5762 or 57.62%

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?

I would tell this friend that it does not matter that the distribution for the condos in Green Valley is not normal. Since the 100 condos are independently and randomley selected, it follows that the mean of this random sample of condos can be grouped with the means of many trials from different random samples. This postulated/experimental distribution of means of random samples forms a normal distribution, when done over a large set of trials. This is the basis of the Central Limit Theorem and allows for us to prediction of probability for distributions that are not normal.

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

The shape/type of distribution for the population of condiminiums is not specified. As such, the probability for a specific value of X=$210,000 can not be calculated.