Terms: Fall, 2013

ECO 480A Econometrics

Homework 1 Solution

Due Date: 09/20/2013

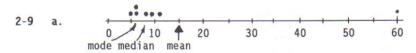
Chap 2

2-7 total = 200 x mean

Formal justification:

$$\bar{X} \equiv \frac{1}{n} \; \Sigma X \label{eq:continuous}$$
 i.e.,
$$\bar{X} \equiv \frac{1}{n} \; (\text{total})$$

Solve: $total = n\overline{X}$



b. Total = 6 + 8 + ... + 60 = 105 thousand

Mean =
$$\frac{\Sigma X}{n}$$
 = $\frac{\text{total}}{n}$ = $\frac{105}{7}$ = 15 thousand

To find the median and mode, we use the observations in the graph above. The median is the middle observation, that is, 8 thousand. The mode is the most frequent value, 6 thousand.

Note that the order is what we also found in the skewed distribution of Figure 2-4b:

c. If production averages 7.8 per country, and if there are 10 countries, then the total production must be $10 \times 7.8 = 78$. (Information about the median or mode was irrelevant, of course.)

Here is the formal justification:

$$\bar{X} \equiv \frac{1}{\bar{n}} \; \Sigma X$$
 i.e.,
$$\bar{X} \equiv \frac{1}{\bar{n}} \; (\text{total})$$

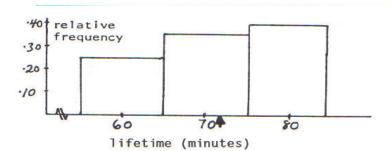
Solve: total = $n\overline{X}$ = (10)(7.8) = 78

2-11

			(a)	(b)
Cell Boundaries	Cell Mid- point X	Freq.	Relative Freq. f/n	Xf
55.00-65.00	60	5	. 25	300
65.00-75.00	70	7	.35	490
75.00-85.00	80	8	.40	640
		20 🗸	1.00 /	1430 20

 $\overline{X} = 71.5$

a.



(a)

(i)
$$\bar{x} = \frac{100*30+110*70}{30+70} = 107$$

(ii)
$$\bar{x} = \frac{100*80+110*20}{20+80} = 102$$

(iii)
$$\bar{x} = \frac{100*50+110*50}{50+50} = 105$$

(iv)
$$\bar{x} = \frac{100*15+110*15}{15+15} = 105$$

(b) True

Since
$$X_1 = n_1 \overline{x_1}, X_2 = n_2 \overline{x_2}, n = n_1 + n_2, X = X_1 + X_2$$

$$X = \frac{\overline{x}}{n} = \frac{n_1 \overline{x}_1 + n_2 \overline{x}_2}{n_1 + n_2}$$

Terms: Fall, 2013

2-17 a. From the box plot in Problem 2-3, we can read off the range from the ends of the whiskers:

Range =
$$16 - 8 = 8$$

We can also read off the IQR from the ends of the box:

$$IQR = 12 - 9 = 3$$

b.

X	X - X	X - X	$(X-\overline{X})^2$
8	-3	3	9
9	-2	2	4
10	- 1	1	1
12	1	1	1
16	5	5	25
X = 11	0 1	$MAD = \frac{12}{5}$	$MSD = \frac{40}{5}$
		= 2.4	= 8.0

variance
$$s^2 = \frac{40}{4} = 10.0$$

standard deviation $s = \sqrt{10.0} = 3.2$

Similarly, for the next 50 measurements, \overline{X} = 71.3, s = 4.14. And for the last 50, X = 70.3, s = 2.75. This shows the same two phenomena as noted in part a: X is decreasing, and so is s.

C.

X	g	iven	f	total	f	xf	x-X	(x-X) ²	(x-X)2f
50	0	1	0	1		60	-12	144	144
55	2	7	5	14		910	- 7	49	686
60 65 70	15	22	38	75		5250	- 2	4	300
75	19	18	6	43		3225	3	9	387
80	11	2	1			1120	8	64	896
75 80 85	3	0	0	14		255	13	169	507
				150	χ̄ =	10,820 150		s ² =	2920 149
					=	72.1 =	72	S	4.43

Alternatively, we could have used the three sums already calculated in part b to easily get the overall sum for the numerator of X:

$$\bar{\chi} = \frac{3,740 + 3,565 + 3,515}{150} = \frac{10,820}{150} = 72.1$$

Or, we could get a better understanding by writing it as:

$$\ddot{X} = \frac{n_1 \ddot{X}_1 + n_2 \ddot{X}_2 + n_3 \ddot{X}_3}{n_1 + \ddot{n}_2 + n_3}$$

$$= \frac{50 (74.8) + 50 (71.3) + 50 (70.3)}{50 + 50 + 50}$$

$$\ddot{X} = \frac{74.8 + 71.3 + 70.3}{3} = 72.1$$

This shows that when all n_i are equal, the overall mean is just the average of the component means.

For the standard deviations, there is no such simple relation. We note that the overall s of 4.43 is <u>greater</u> than the average of the components:

$$4.43 > \frac{4.84 + 4.14 + 2.75}{3}$$

The reason is intuitively clear: the overall sample is spread out not only because of the three individual spreads, but also because the three individual means are spread out too (which makes the overall s larger).

Xi	fi	- Xi- <i>x</i>	$ Xi-x ^2$	$-\frac{1}{ Xi-x ^2}$ fi
5.9	2	0.18	0.0324	0.0648
6.0	16	0.08	0.0064	0.1024
6.1	22	0.02	0.0004	0.0088
6.2	10	0.12	0.0144	0.144
total	50			0.32

$$\overline{X} = \frac{1}{50} (5.9 \times 2 + 6 \times 16 + 6.1 \times 22 + 6.2 \times 10) = 6.08$$

$$S^2 = \frac{1}{49} \, 0.32 = 0.0065 \quad S = 0.0808$$

$$CV = (\frac{S}{\overline{X}})*100\% = \frac{0.0808}{6.08}*100\% = 0.0133$$

(b)

$$\overline{X} = 6.08ounce = 0.38 pound$$

$$S = 0.0808 * \frac{1}{16} = 0.00505$$

$$CV = (\frac{S}{X})*100\% = \frac{0.00505}{0.38}*100\% = 0.0133$$

(c)

Set Y=0.42X+0.24

Then
$$\overline{Y} = 0.24 + 0.42\overline{X} = 2.79$$

$$S_y = 0.42 * 0.0808 = 0.0339$$

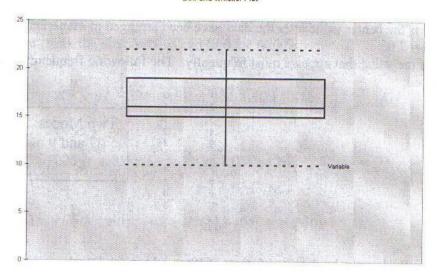
x	f/n	x f/n	x - X	$(x-\overline{X})^2$	$(x-\overline{X})^2f/n$
10	.28	2.8	-6	36	10.08
15	.40	6.0	-1	1	.40
20	.20	4.0	4	16	3.20
25	.08	2.0	9	81	6.48
20 25 30	.04	1.2	14	196	7.84
		Y = 16 0		М	SD = 28 00

$$s^2 = (\frac{n}{n-1}) MSD = \frac{25}{24} 28.00 = 29.2$$

 $s = \sqrt{29.2} = 5.4$

a. Box and Whisker Plot done using PHStat





The lower limit is computed as Q1 - 1.5(Q3 - Q1) = 15 - 1.5(19-15) = 9

The upper limit is Q3 + 1.5(Q3 - Q1) = 19 + 1.5(19-15) = 25.0

Since no value is less than 9 nor greater than 25, there are no outliers in these data.

b. The 60th percentile is found by using $i = \frac{p}{100}(n+1) = \frac{60}{100}(45+1) = 27.60$

Thus, the 60th percentile is somewhere between the 27th and 28th value from the top of the data sorted from low to high. This value is 17.

a.

Method 1						
	X		X - x	$(X-x)^2$		
		14	2	4		
		11	-1	1		
		19	7	49		
		6	-6	36		
15		10	-2	4		
28		60		94		

	Method 2	1
$(X - x^{-})^{2}$	X - x	X
196	14	26
49	-7	5
9	-3	9
36	-6	6
4	2	14
294		60

Method 1:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 60/5 = 12$$

$$S^2 = \frac{\sum_{i=1}^{n} (x - \bar{x})^2}{n - 1} = 94/(5 - 1) = 23.5$$

$$S = \sqrt{S^2} = \sqrt{23.5} = 4.848$$

Method 2:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = 60/5 = 12$$

$$S^2 = \frac{\sum_{i=1}^{n} (x - \bar{x})^2}{n - 1} = 294/(5 - 1) = 73.5$$

$$S = \sqrt{S^2} = \sqrt{73.5} = 8.573$$

b. Method 1:

$$CV = \frac{S}{x}(100) = (4.848/12)(100) = 40.4\%$$

Method 2:

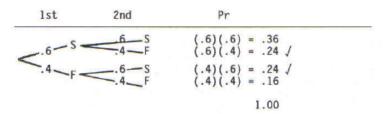
$$CV = \frac{S}{x}(100) = (8.573/12)(100) = 71.4\%$$

Method 1 provides less relative variability in the distances traveled.

c. In this case it would be acceptable to compare standard deviations since the means are the same.

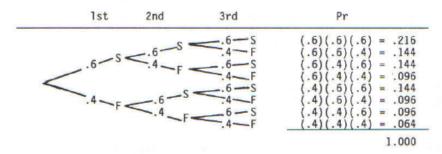
Terms: Fall, 2013

3-5 a. Let F = failure, S = success. Since Pr(F) = .40, Pr(S) = .60. Then for 2 attempts, we have the following tree:



Pr(exactly one failure) = .24 + .24 = .48

For 3 attempts, we have a three-fold tree: b.



Pr (exactly one failure) = .144 + .144 + .144 = .432

3-7 a. Let A denote an acceptable clock U denote an unacceptable clock

It makes no difference to the answer, whether the 3 clocks are all drawn

at once, or one by one. The simplest way to lay out the calculations, however, is to imagine them drawn one by one:

The chance of the first clock being acceptable is 8/12. Then, in the carton there would remain only 7 acceptable clocks in a total of 11, so that the chance of the second clock being acceptable would be 7/11. Similarly, the chance of the third clock then being acceptable would be 6/10. Thus the probability tree branches as follows:

	FIRST	SECOND CLOCK	THIRD	OUTCOME	PROBABILITY
	- 1	6/	10A	Accept lot	8 7 6 12 11 10
	_A </td <td>II A</td> <td>U</td> <td>Reject</td> <td>= .255</td>	II A	U	Reject	= .255
8/12		U		Reject	
	<u> </u>			Reject	

Thus Pr(accepting shipment) = .255 = 25%

b.
$$\frac{6}{12}$$
 $\frac{5}{11}$ $\frac{4}{10}$ = .091 \(\text{ = 9\%} \)

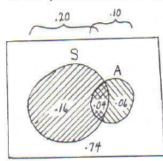
3-9 a. <u>Intuitive Solution</u>: Just imagine what is actually going on. "At least one will work" means that there will be some help given. How often does this happen?

Well, we are told that both fail only 4% of the time. So the remaining

96% of the time we get help.

<u>Venn Diagram Solution</u>: A Venn Diagram is a simple routine procedure that helps in many problems like this.

Let S denote the sprinkler failing and A the alarm failing. We first fill in the diagram with the central lens probability of .04. Then we work gradually outwards using the other given probabilities listed on the top, to eventually fill in the probabilities in all four sections:



"At least one will work" is everything but the lens at the center, so its probability is 1.00 - .04 = .96.

b. "Both will work" is just the clear area of the Venn Diagram, and its probability has already been found:

$$Pr = 1 - (.16 + .04 + .06) = 1 - .26 = .74$$

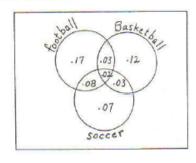
3-11 a.
$$Pr(G_1 \text{ and } G_2 \text{ and } ..) = \frac{1}{2} \times \frac{1}{2} \times ... = (\frac{1}{2})^{10} \approx .001$$

b. From part (a), we found Pr (all girls) = $(1/2)^{10}$. Similarly, Pr (all boys) = $(1/2)^{10}$. The complement is what we want:

Pr (boys and girls both occurring) =
$$1 - (1/2)^{10} - (1/2)^{10}$$

= .998

3-13 The Venn diagram is filled in first at the center: Pr (all 3 sports) = .02. Then we work our way outwards to fill in the rest of the diagram. When this is done we can answer the questions, as follows:



a. sum all the probabilities:
$$.17 + .03 + ... = .52$$

- b. .17
- c. .40
- d. $\frac{.17}{.52} = .3$
- e. $\frac{.40}{.52} = .77$

Terms: Fall, 2013

To answer (d) and (e), it might help to imagine 100 actual students, 52 of whom are athletes, 17 of whom play football only, etc. Then we would get the same answers as before:

d.
$$\frac{17}{52} = .33$$

e.
$$\frac{40}{52} = .77$$

3-15 a. We read the unemployment rate directly from the right-hand (totals) column:

$$Pr(U) = 7.2\%/100\% = 7.2\%$$

b. Pr(U|M) is the male unemployment rate, read from the first (male) column:

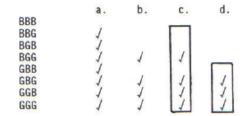
$$Pr(U|M) = 3.9\%/55.8\% = .0699 \approx 7.0\%$$

c. Similarly, Pr(U|F) is the female rate, read from the second (female) column:

$$Pr(U|F) = 3.3\%/44.2\% = .0747 \approx 7.5\%$$

<u>Remarks</u>: Note that the <u>overall</u> unemployment rate (7.2%) is the average of the male rate (7.0%) and female rate (7.5%), -- with the male rate weighted more heavily, since most of the workers are male.

3-17



$$a. 7/8 = .875$$

b.
$$4/8 = .50$$

c.
$$4/7$$
 or $.50/.875 = .57$

d.
$$3/4$$
 or $.375/.50 = .75$

3-20

Event E: First die is $5 = \{(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)\}$

Pr(E)=6/36=1/6

Event F: Total is $7 = \{(1,6),(6,1),(2,5),(5,2),(3,4),(4,3)\}$

Pr(F)=6/36=1/6

Event G: Total is $10 = \{(4,6),(6,4),(5,5)\}$

Pr(G)=3/36=1/12

(a)

 $Pr(F|E)=Pr(F\cap E)/Pr(E)=(1/36)/(1/6)=1/6$

Pr(F)=1/6

(b)

 $Pr(G \mid E) = Pr(G \cap E) / Pr(E) = (1/36) / (1/6) = 1/6$

Pr(G)=1/12

(c)

It is correct

Because of Pr(GIE)>Pr(G), to peek at the first die whether it is a 5 is helpful to bet on whether the dice show 10.

However, because of Pr(FIE)=Pr(F), to peek at the first die whether it is a 5 is not helpful to bet on whether the dice show 7.

3-21		F	F	totals
	E	7.8 %	22.2 %	30.0 %
	E	18.2 %	51.8 %	70.0 %
	totals	26.0 %	74.0 %	100.0 %

a. From the total in the F column,

$$Pr(F) = 26.0 \%$$

b. From the total in the E row, Pr(E) = 30.0 % and hence

$$Pr(F|E) = \frac{Pr(F \text{ and } E)}{Pr(E)}$$

= $\frac{7.8 \%}{30.0 \%} = 26.0 \%$

c. Because Pr(F|E) = Pr(F), F is indeed independent of E, according to (3-20).

3-23

(a)

Pr(U)=8.3/115.5=7.2%

(b)

$$P(U|Y) = Pr(U \cap Y)/Pr(Y) = \frac{3.2/115.5}{23.6/115.5} = 13.6\%$$

(c)

 $Pr(U|Y) \neq Pr(U)$, it is not independent.

Instructor: Fa-Hsiang Chang Terms: Fall, 2013

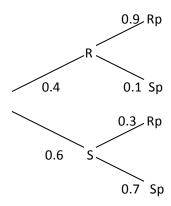
3-26

R=Rainy day

S=Not a rainy day

Rp=Barometer predicted "rain"

Sp=Barometer predicted "no rain"



 $Pr(R \mid Rp) = Pr(R \cap Rp) / Pr(Rp)$

$$= \frac{\Pr(Rp \mid R) \Pr(R)}{\Pr(Rp \mid R) \Pr(R) + \Pr(Rp \mid S) \Pr(S)}$$

$$= \frac{0.9*0.4}{0.9*0.4+0.3*0.6} = 66.7\%$$

Q1.

4.9.

Origination	Early	On-Time	Late	Total
San Francisco	25	50	100	175
Los Angeles	50	100	75	225
Total	75	150	175	400

- a. P(Early) = 75/400 = 0.1875.
- b. P(Los Angeles) = 225/400 = 0.5625.

P(Early Given Los Angeles) = 50/225 = 0.2222.
 P(Early) = 75/400 = 0.1875.

Terms: Fall, 2013

d	Let F = Farly	O= On-Time, and	I. = Late	The events	are as follows:
4.8	Let E - Edity.	O- OH-THIE, and	The Date.	THE CALCULA	THE CALL POPLE OF THE

Elementary Event	Flight 1	Flight 2	Flight 3
Event 1	E	Е	E
Event 2	E	E	0
Event 3	E	E	L
Event 4	E	0	E
Event 5	E	0	0
Event 6	E	0	L
Event 7	Е	L	E
Event 8	E	L	0
Event 9	E	L	L
Event 10	0	E	E
Event 11	0	E	0
Event 12	0	E	L
Event 13	0	0	E
Event 14	0	0	O
Event 15	0	0	L
Event 16	0	L	E
Event 17	0	L	0
Event 18	0	L	L
Event 19	L	E	E
Event 20	L	E	0
Event 21	L	E	L
Event 22	L	0	E
Event 23	L	0	0
Event 24	L	0	L
Event 25	L	L	E
Event 26	L	L	0
Event 27	L	L	L

The sample space consists of the 27 elementary events listed in the table. Sample Space = {Event 1, Event 2, Event 3, ..., Event 27}

Q2

4.18.

a. There are 27 possible elementary events during the 3- 30 minute time slots determined as follows:

Slot 1 Slot 2 Slots 3

3 stations x 3 stations x 3 stations = 27

In order to have all three stations represented, there are 3 possibilities during the first time slot. Once the station for that slot is selected, there are two choices for the second slot. Finally, there will be one choice for the final slot. This, the number of ways this event can happen is $3 \times 2 \times 1 = 6$. Thus, the probability of all three stations being represented is 6/27 = 0.2222

b. The outcome that at least 2 slots are the same channel equals 21, which is total possible outcome 27 minus t 6 that is he or she select all channels(from the answer of question a). The elements of sample space as follow:

222 226 227 262 266 272 277 622 626 662 666 667 676 677 722 727 766 767 772 776 777

So the answer of this question is condition probability p (all the same under at least 2 same channel).

$$\frac{3/27}{21/27} = \frac{1}{7}$$

Q3

Instructor: Fa-Hsiang Chang Terms: Fall, 2013

Q4

```
P(Line 1) = 0.4
P(Line 2) = 0.35
P(Line 3) = 0.25

P(Defective|Line 1) = 0.05
P(Defective|Line 2) = 0.10
P(Defective|Line 3) = 0.07

You need to calculate the probability of each line given you know the cases are defective. Use Bayes' Rule to calculate this.

P(Defective) = P(Defective|Line1)P(Line1) + P(Defective|Line2)P(Line2) + P(Defective|Line3)P(Line3) = (0.05)(0.4) + (0.1)(0.35) + (0.07)(0.25) = 0.0725

P(Line1|Defective) = (0.05)(0.4)/0.0725 = 0.2759
P(Line2|Defective) = (0.10)(0.35)/0.0725 = 0.4828
P(Line3|Defective) = (0.07)(0.25)/0.0725 = 0.2413

The unscaled cans probably came from Line 2
```

Chap 4

a.	×	p(x)	xp(x)	x-μ	$(x-\mu)^2$	$(x-\mu)^2p(x)$
	40	.05	2.00	-22	484	24.20
	50	.15	7.50	-12	144	21.60
	60	.41	24.60	- 2	4	1.64
	70	.34	23.80	- 8	64	21.76
	80	.04	3.20	18	324	12.96
	90	.01	.90	28	784	7.84
	$\mu = 62.00$					$\sigma^2 = 90.00$
					σ =	$\sigma = \sqrt{90} = 9.49$

b. All 60 will be sold if demand is 60 or more. We find this from the last 4 probabilities in the table of p(x) above:

$$Pr(X \ge 60) = .41 + .34 + .04 + .01 = .80$$

There will be some left over if demand is less than 60, which we similarly find:

$$Pr(X < 60) = .05 + .15 = .20$$

As a check, since these two events are complementary, their probabilities should sum to 1.00:

c. To be almost sure of having enough bicycles, we need to order a lot. Let's try ordering 80. Then the chance of running out would be only:

$$Pr(X > 80) = .01$$

But it was specified this chance can be as large as 1-.95 = .05. So we need not shoot quite so high; let's try ordering 70:

Pr(X > 70) = .04 + .01 = .05 $\sqrt{ }$ Thus Pr(X \leq 70) = .95 as required.

(a)

$$Pr(X=0) = 0.015625$$

$$Pr(X=1) = 0.09375$$

$$Pr(X = 2) = 0.234375$$

$$Pr(X = 3) = 0.3125$$

$$Pr(X = 4) = 0.234375$$

$$Pr(X = 5) = 0.09375$$

$$Pr(X = 6) = 0.015625$$

$$\mu = \sum X P(X) = 3.00$$

$$\sigma^2 = \sum (X - \mu)^2 P(X) = 1.50$$

$$\sigma = 1.50^{\circ}0.5 = 1.22$$

i.
$$Pr(X = 3) = 31.25\%$$

ii.
$$Pr(X = 3) + Pr(X = 2) + Pr(X = 4) = 78.125\%$$

iii.
$$Pr(X \ge 3) = Pr(X = 3) + Pr(X = 4) + Pr(X = 5) + Pr(X = 6) = 65.625\%$$

4-11

The number of successful wells S is binomial, with n = 12 and π = .20. From Table III(c),

$$Pr(S \ge 3) = .442 \approx 44\%$$

a.
$$n = 6$$
 and $\pi = .50$. Thus (4-13) gives

$$\mu = n\pi = 6(.50) = 3.00$$

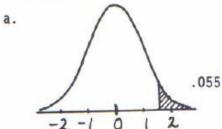
$$\sigma = \sqrt{n\pi(1-\pi)} = \sqrt{6(.50)(.50)} = \sqrt{1.50} = 1.22$$

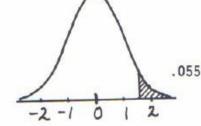
b.
$$n = 12$$
 and $\pi = .20$. Thus

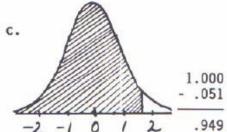
$$\mu = n\pi = 12 (.20) = 2.40$$

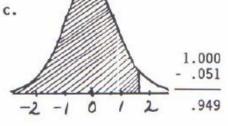
$$\sigma = \sqrt{n\pi(1-\pi)} = \sqrt{12(.20)(.80)} = \sqrt{1.92} = 1.39$$

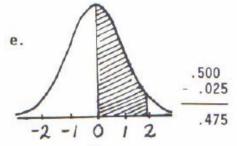
Although we could write out these solutions algebraically, it is easier to work graphically. We look up tail areas in Table IV, and so find;

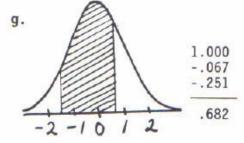


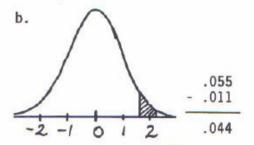


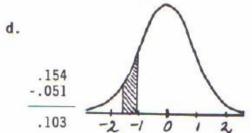


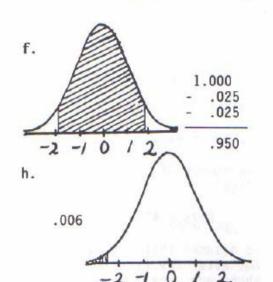












(a)
$$\Pr(X > 20) = \Pr(\frac{X - \mu}{\sigma} > \frac{20 - 16}{5}) = \Pr(Z > 0.8) = 0.212$$

(b)
$$\Pr(20 < X < 25) = \Pr(\frac{20 - 16}{5} < \frac{X - \mu}{\sigma} > \frac{25 - 16}{5}) = \Pr(0.8 < Z < 1.8)$$

= $\Pr(Z > 0.8) - \Pr(Z > 1.8) = 0.176$

(c)
$$Pr(X < 10) = Pr(\frac{X - \mu}{\sigma} > \frac{10 - 16}{5}) = Pr(Z < -1.2) = Pr(Z > 1.2) = 0.115$$

(d)
$$\Pr(12 < X < 24) = \Pr(\frac{12 - 16}{5} < \frac{X - \mu}{\sigma} > \frac{24 - 16}{5}) = \Pr(-0.8 < Z < 1.6)$$

= $1 - \Pr(Z > 0.8) - \Pr(Z > 1.6) = 0.733$

4-22

(a)
$$\mu = 110, \sigma = 20$$

 $\Pr(X < 120) = \Pr(\frac{X - \mu}{\sigma} > \frac{120 - 110}{20}) = \Pr(Z > 0.5) = 0.691$

(b) Suppose the time is *a*

$$Pr(Z < \frac{a-110}{20}) = 0.9$$
 $\frac{a-110}{20} = 1.28$ $a = 135.6$

4-26

(a)
$$\mu = np = 6*0.4 = 2.4$$

(b)
$$R = 200X - 300$$
 $\mu = 200 * 2.4 - 300 = 180$

(c) $\{X=0,1\}$ lose a profit, $\{X=2,3,4,5,6\}$ make a profit The probability of making a profit $\Pr(X \ge 2) = 1 - \Pr(X = 0) - \Pr(X = 1) = 0.76672$ The probability of losing a profit $\Pr(X \le 1) = \Pr(X = 0) + \Pr(X = 1) = 0.23328$

Q1

(a)
$$Pr(X \ge 2) = (0.9)^3 + 3*0.1*(0.9)^2 = 0.972$$

(b)
$$Pr(X = 0) = 0.1^3 = 0.001$$

(c)
$$Pr(X = 1) = 3*0.9*0.1^2 = 0.027$$