CX2100 Prob & Stat

Tutorial #1

rainbow color = Ordinal

Descriptive statistics

1. Categorize the following variables as being qualitative or quantitative and specify the level of measurement scale.

   (a) hair colour quali, Nominal
   (b) number of television sets in a private home quanti, Ratio
   (c) possible responses to questionnaire items: strongly agree, agree, etc. quali, Ordinal
   (d) Intelligence scale of 0 to 100 quanti, interval
   (e) the country where you were born in quali, nominal

2. The following data were obtained for a measurement of certain fuel droplet size.

2.1	2.2	2.2	2.3	2.3	2.4	2.5	2.5	2.5	2.8
2.9	2.9	2.9	3.0	3.1	3.1	3.2	3.3	3.3	3.3
3.4	3.5	3.6	3.6	3.6	3.7	3.7	4.0	4.2	4.5
4.9	5.1	5.2	5.3	5.7	6.0	6.1	7.1	7.8	7.9
8.9									

   (a) Group the droplet sizes and obtain a frequency table using class interval of 1 unit. Construct the frequency histogram and comment on the shape of the distribution.
   (b) Construct the Stem-and-leaf diagram.
   (c) Compute the mean, the 25th, 50th and 75th percentile.
   (d) Construct the box plot for the droplet size data.

3. A frequency distribution of the length of telephone calls monitored at the switchboard of an office is given below. Obtain the mean and the mode of the call duration.

Length of Calls (minutes)	Number of Calls	mid
0 and under 2	10	1
2 and under 4	25	3
4 and under 6	20	5
6 and under 8	40	7
8 and under 10	5	9
Total	100	

4. Find the sample standard deviation for a set of data for which $n = 10$, $\sum x = 50$ and $\sum x^2 = 500$. Explain why it is impossible to have $n = 10$, $\sum x = 50$ and $\sum x^2 = 100$ for a given set of data.

5. Given the following data, determine the Pearson's correlation between variable X and variable Y. Comment on your result.

X	2.5	3.4	5.6	6.7	7.9
Y	10.2	11.8	13.7	19.7	20.6

2) a) Freq table

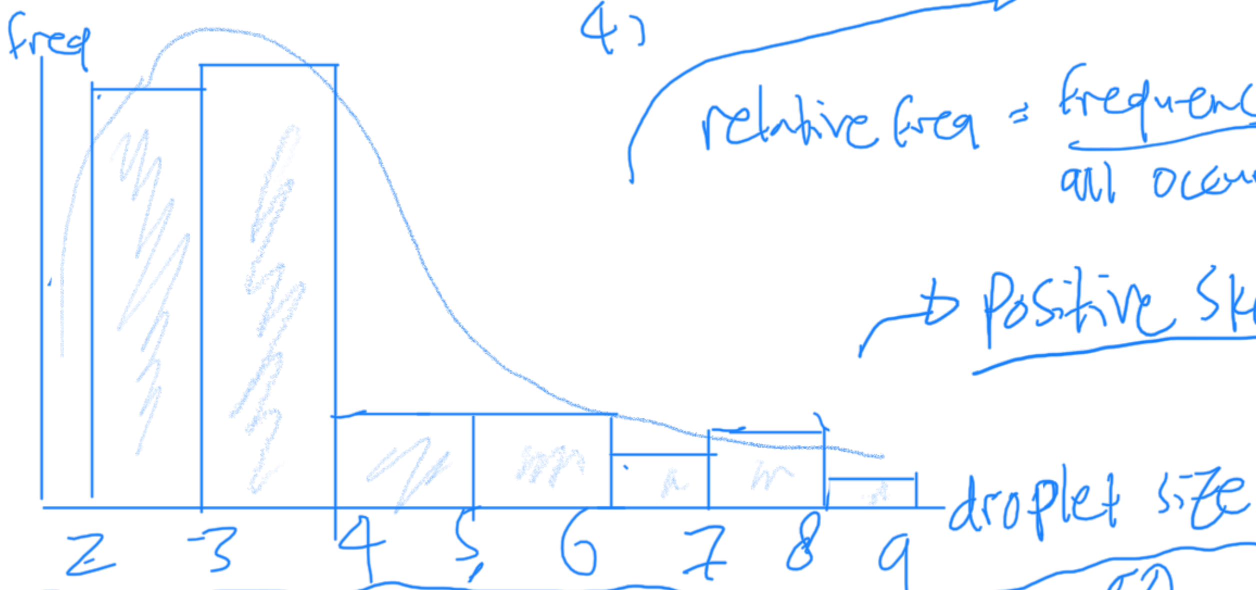
Freq range	Class freq
2 - 3 → class interval	13
3 - 4	14
4 - 5	4
5 - 6	4
6 - 7	2
7 - 8	3
8 - 9	1

b)

8	9
7	1 8 9
6	0 1
5	1 2 3 7
4	0 2 5 9
3	0 1 1 2 3 3 3 4 5 6 6 6 7 7
2	1 2 2 3 3 4 5 5 5 8 9 9 9

Vs → stem n leaf give more info

4)



relative freq = $\frac{\text{frequency}}{\text{all occurrences}}$

→ Positive skew

c) mean = $\frac{162.6}{41} = 3.97$

Range 25th = $\frac{25}{100} \times 42 = 10.5$

50th $\frac{50}{100} \times 42 = 21$

3.4 + 0 = 3.4

25th percentile = 2.8 + (0.1 × 0.5)
= 2.85

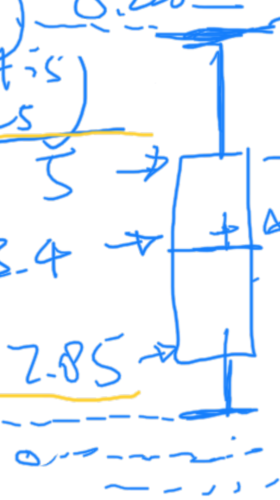
75th $\frac{75}{100} \times 42 = 31.5$

4.9 + (0.2×0.5)
5

d) upper adjecent 7.9 (highest value that is lower than 8.225

8.225

• → 8.9



1 step 3.225

75th percentile 5 →

median 3.4 →

← mean 3.97

25th percentile 2.85 →

lower adjecent value 2.1 =

1 step 3.225

IQR = 2.15

1 step = 1.5 × 2.15
= 3.225

3) mean $= \frac{(0\times1)+(25\times3)+\dots+(5\times9)}{100} = 5.1$

mode $= 7$ (largest freq)

4) i)

$\Sigma x = 50$

$\Sigma x^2 = 500$

$\mu = \frac{\Sigma x}{n} = \frac{50}{10} = 5$

$\sigma = \frac{500 - \frac{(50)^2}{10}}{9}$

$\sigma^2 = \frac{500 - 250}{9}$

$s = \sqrt{27.78}$

$s = 5.27$

$\Sigma x = 50$

$\Sigma x^2 = 100$

$\mu = \frac{\Sigma x}{n} = \frac{50}{10} = 5$

$\sigma^2 = \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1} = \frac{100 - 250}{9}$

$\sigma^2 = -16.67$

Variance cannot be negative

5)

$$\mu_x = \frac{26.1}{5} = 5.22$$

$$\mu_y = \frac{76}{5} = 15.2$$

$$\sigma_x^2 = \frac{\Sigma(x-\mu_x)^2}{N} = \frac{7.3984 + 3.3124 + 0.1444 + 2.1904 + 7.1824}{5} = \frac{20.228}{5}$$

$$= 4.0456$$

$$\sigma_y^2 = \frac{25 + 11.56 + 22.5 + 20.25 + 29.16}{5} = \frac{88.22}{5} = 17.644$$

$$COV(x,y) = \frac{\Sigma(x-\mu_x)(y-\mu_y)}{5}$$

$$= \frac{13.6 + 6.188 - 0.57 + 6.66 + 14.472}{5}$$

$$= 40.35/5 = 8.07$$

$$\rho = \frac{8.07}{\sigma_x \sigma_y} = \frac{8.07}{8.449} = 0.9552$$

6) $N_a = 10$

$\mu_a = 2.4$ hours

standart deviation $= \sqrt{\sigma^2}$

$\sigma_a = 0.8$ hours

Variance $= \sigma_a^2 = 0.64$

$N_b = 5$

$\mu_b = 2$ hours

$\sigma_b = 1.2$ hours

Variance $= \sigma_b^2 = 1.44$

mean $= \frac{\Sigma x}{N}$

$\Sigma x_a = 2.4 \times 10 = 24$

$\Sigma x_b = 2 \times 5 = 10$

$\Sigma x_{a+b} = \frac{34}{15} = 2.267$

$s_a^2 = 0.64 = \frac{1}{9}\left[\Sigma x_a^2 - \frac{24^2}{10}\right]$

$5.76 = \Sigma x_a^2 - 57.6$

$\Sigma x_a^2 = 63.36$

$s_b^2 = 1.44 = \frac{1}{4}\left[\Sigma x_b^2 - \frac{10^2}{5}\right]$

$5.76 = \Sigma x_b^2 - 20$

$\Sigma x_b^2 = 25.76$

$s_{a+b}^2 = \frac{1}{14}\left[(\Sigma x_a + \Sigma x_b) - \frac{(\Sigma x + \Sigma y)^2}{15}\right]$

$= \frac{1}{14}(89.12 - 77.06)$

$s_{a+b} = \sqrt{0.86} = 0.93$

Cannot use Variance Sum law II cause they are same dataset

6. In an attempt to find the mean number of hours his tutorial classmates spent per day preparing for tutorials, John collected data from 10 of his friends in the tutorial group and found that the mean is 2.4 hours with a standard deviation of 0.8 hours. However, a day later he felt that the sample size is too small. So he collected data from another 5 of his friends and found that the mean is 2.0 hours with a standard deviation of 1.2 hours. Find the mean and standard deviation when these 2 sets of data are combined.

Answers

1. (a) qualitative, nominal
   (b) quantitative, ratio
   (c) qualitative, ordinal
   (d) quantitative, interval
   (e) qualitative, nominal

2. (a)

Class	[2,3)	[3,4)	[4,5)	[5,6)	[6,7)	[7,8)	[8,9)
Frequency	13	14	4	4	2	3	1

Histogram and comment on the shape (skewed to the right).

(b) Stem-and-leaf diagram:

(c) mean, 25th, 50th and 75th percentile = 3.97, 2.85, 3.4 and 5.0 respectively.

(d) Box plot:

3. $\bar{x}$ = 5.1 mins, $mode$ = 7 mins (middle of the class interval)

4. s = 5.27, Not possible to have variance < 0.

5. ρ = 0.955. X and Y are highly positive correlated.

6. $\bar{x} = 2.267$, $s = 0.93$

CX2100 Prob & Stat

Tutorial #2

Probability

1. A jar contains four coins: a nickel (5¢), a dime (10¢), a quarter (25¢), and a half-dollar (50¢). Three coins are randomly selected without replacement from the jar. order doesn't matter

   N D Q H

   (a) List all the possible outcomes in sample space S. {NDQ}, {NQH}, {NDH}, {DQH}

   (b) What is the probability that the total amount drawn will equal 60¢ or more?

   40 80 65 85

   >60 $\frac{3}{4}$

2. A mother prepares nine popsicles of different flavours: three of orange, three of cherry and three of grape, for a party of four children. If every child is allowed to choose a popsicle of his/her favourite flavour, what is the probability that all of them will get their choices?

   3 O
   3 C
   3 G

   $1 - \left(\frac{3}{81}\right) \Rightarrow \frac{26}{27}$

3. When two events are mutually exclusive, they cannot both happen when the experiment is performed. Once event B has occurred, event A cannot occur, i.e. P(A|B) = 0 or P(A∩B) = 0, and vice versa. The occurrence of event B certainly affects the probability of occurrence of event A. Therefore, mutually exclusive events must be dependent.

   When two events are independent, the occurrence of event B does not affect the probability of occurrence of event A, i.e. P(A|B) = P(A) or P(A∩B) = P(A)P(B), and vice versa. Event A may still occur even if event B has occurred. Therefore, independent events cannot be mutually exclusive.

   Use the relationships above to fill in the table below:

P(A)	P(B)	Conditions	P(A|B)	P(A∩B)	P(AUB)
0.3	0.4	mutually exclusive	0	0	0.3 + 0.4 = 0.7
0.3	0.4	independent	0.3	0.12	0.7 − 0.12 = 0.58
0.1	0.5	mutually exclusive	0	0	0.6
0.2	0.5	independent	0.2	0.1	0.7 − 0.1 = 0.6

4. A blood disease is found in 2% of the persons in a certain population. A new blood test will correctly identify 96% of the persons with the disease and 94% of the persons without the disease.

   (a) What is the probability that a person who is called positive by the blood test actually has the disease?

   (b) What is the probability that a person who is called negative by the blood test actually does not have the disease?

   (c) Comment on the results obtained in part (a) & (b).

5. (a) A magician has in his pocket a fair coin and a doctored coin where both sides are heads. If he randomly picks a coin to flip, and obtains a head, what is the probability that he picks the fair coin?

   (b) If he flips the same coin the second time and obtains a head again, what is the probability that it is a fair coin?

2) Total # of ways for 4 kids to choose the 3 flavor

$3 \times 3 \times 3 \times 3 = 3^4 = 81$

# of ways for anyone of them not getting his flavour
= # of ways for all to choose same flavour
= 3

P(anyone of them not getting his flavour) $= \frac{3}{81} = \frac{1}{27}$

P(all of them getting their flavours) $= 1 - \frac{1}{27} = \frac{26}{27}$

4) $P(B) = 0.02$ $\quad P(B') = 0.98$ $\quad$ B = event one hv disease

$P(T|B) = 0.96$ $\quad P(T'|B) = 0.04$ $\quad$ T = event one get "+" test

$P(T'|B') = 0.94$ $\quad P(T|B') = 0.06$

a) $P(B|T) = \frac{P(T|B)\,P(B)}{P(T)} = \frac{0.96 \cdot 0.02}{P(T \cap B) + P(T \cap B')}$

$= \frac{0.96 \cdot 0.02}{P(T|B)\,P(B) + P(T|B')\,P(B')} = \frac{0.0192}{0.0192 + 0.0588} = 0.246$

b) $P(B'|T') = \frac{P(T'|B')\,P(B')}{P(T'|B')\,P(B') + P(T'|B)\,P(B)} = 0.999$

c) negative result (B) give almost surety that the person doesn't have disease while positive result (A) might not be accurate

5) F = Fair Coin Flip

H = Head outcome

a) $$P(F|H) = \frac{P(H|F)\,P(F)}{P(H)} = \frac{P(H|F)\,P(F)}{P(H|F)\,P(F) + P(H|F')\,P(F')}$$

$$= \frac{0.5 \times 0.5}{0.5 \times 0.5 + 1 \times 0.5} = \frac{0.25}{0.75} = \frac{1}{3}$$

independent

b) $$P(F|HH) = \frac{P(HH|F)\,P(F)}{P(HH)}$$

$$= \frac{P(HH|F)\,P(F)}{P(HH|F)\,P(F) + P(HH|F')\,P(F')}$$

$$= \frac{1/4 \cdot 1/2}{1/4 \cdot 1/2 + 1 \cdot 1/2} = \frac{1/8}{1/8 + 1/2} = \frac{1/8}{10/16}$$

$$= \frac{16}{80} = \frac{1}{5}$$

Answers

1. (b) $\frac{3}{4}$

2. $\frac{26}{27}$

3.

P(A)	P(B)	Conditions	P(A|B)	P(A∩B)	P(AUB)
0.3	0.4	mutually exclusive	0	0	0.7
0.3	0.4	independent	0.3	0.12	0.58
0.1	0.5	mutually exclusive	0	0	0.6
0.2	0.5	independent	0.2	0.1	0.6

4. (a) 0.246 (b) 0.999

5. (a) $\frac{1}{3}$ (b) $\frac{1}{5}$

CX2100 Prob & Stat

Tutorial #3

Discrete Probability Distribution

1. Suppose that the probabilities are 0.2, 0.4, 0.3 and 0.1 that the number of wills filed on any day at Kusu Island will be 0, 1, 2, or 3.

   $P(x \geq 2) = P(x=2) + P(x=3)$
   $= 0.3 + 0.1 = 0.4$

   (a) What is the probability of having at least 2 wills filed per day?
   (b) Find the expected number of wills filed per day. $0 + 0.4 + 0.6 + 0.3 = 1.3$
   (c) Find the variance of the number of wills filed per day.

2. Given that $f(x) = k/2^x$, is a discrete probability function for a r.v. that can take on the values $x = 0, 1, 2, 3$ and 4. Find k and tabulate the cumulative probability $P(X \leq x)$.

3. A biased die is rolled 50 times and the number of twos appeared is 10. If the die is rolled for another 10 times, determine the following:

   (a) the probability that we get a two exactly 3 times.
   (b) the expected number of twos.
   (c) the variance of the number of twos.

4. The number of calls coming per minute into a hotel reservation center is Poisson random variable with mean 3.

   (a) Find the probability that no calls come in a given 1minute period.
   (b) Assume that the number of calls arriving in two different minutes are independent. Find the probability that at least two calls will arrive in a given two minutes period.

5. The probability that a student fails Subject A exam is 0.05. If the student failed the subject, he will have to re-take it the following semester. Let X be the number of times he attempted to pass the subject.

   (a) Determine and name the probability distribution of X.
   (b) Find the probability that a student will pass the subject with no more than 2 attempts.
   (c) Find the average number of attempts to pass the subject.

Answers

1. (a) 0.4 (b) 1.3 (c) 0.81

2. 16/31

x	0	1	2	3	4
$F(x)= \text{Prob}(X \leq x)$	16/31	24/31	28/31	30/31	1

3. (a) 0.20133 (b) 2 (c) 1.6

4. (a) e^{-3} (b) $1-7e^{-6}$

5. (a) $P(X=k) = 0.05^{k-1}\, 0.95$ (Geometric dist) (b) 0.9975 (c) 1.0526

1) c) $\sigma^2 = \sum x^2 P(x) - \mu^2$

$$(0^2 \cdot 0.2 + 1^2 \cdot 0.4 + 2^2 \cdot 0.3 + 3^2 \cdot 0.1) - 1.3^2$$

$$(0 + 0.4 + 1.2 + 0.9) - 1.69 = 0.81$$

2)

X	0	1	2	3	4
Probability mass function $f(x) = k/2^x$	k	$k/2$	$k/4$	$k/8$	$k/16$
	$16/31$	$8/31$	$4/31$	$2/31$	$1/31$
cumulative distribution function	$16/31$	$24/31$	$28/31$	$30/31$	1

$\rightarrow \sum f(x) = 1$

$$\frac{16k + 8k + 4k + 2k + k}{16} = 1$$

$$\frac{31k}{16} = 1 \qquad k = \frac{16}{31}$$

3) a) Probability of getting 2 $= \frac{10}{50} = 0.2$ Probability of not getting 2 $= 1 - \frac{10}{50} = \frac{40}{50} = 0.8$

let x = no of getting 2

n = 10 trials

$$P(x=3) = {}^{n}C_3 \cdot (0.2)^3 (1-0.2)^{n-3}$$

$$= {}^{10}C_3 \quad 0.008 \quad 0.2097152$$

$$= \frac{10^{5} \cdot 9 \cdot 8^{4} \cdot 7!}{7! \, 3 \cdot 2} \, 0.008 \quad 0.2097152 = 120 \cdot 0.00167\ldots$$

$$= 0.20133$$

b) $E[x] = np(2) = 10 \cdot 0.2$

$= 2$

c) $Var[x] = npq \quad q = 1-p$

$= n\,P(2)(1-P(2))$

$= 10 \cdot 0.2 \quad 0.8$

$= 1.6$

4) a) $X \sim Pois(3)$ per minute (# of call / minute)

$$P(\text{no call}) = \frac{e^{-3} 3^0}{0!} = 0.95$$

b) let x_1 & x_2 = # calls in 1^{st} min & 2^{nd} min.

$P(x_1 + x_2 \geq 2) = 1 - P(x_1 + x_2 < 2)$

$P[x_1 + x_2 = 0] + P[x_1 + x_2 = 1]$

$= P[x_1 = 0] P[x_2 = 0] + P[x_1 = 0] P[x_2 = 1] + P[x_1 = 1] P[x_2 = 0]$

$= 7e^{-6} \rightarrow 1 - 7e^{-6}$

5) $P(\text{fail} | A) = 0.05$ $P(\text{pass} | A) = 0.95$

a) geometric distribution

$$P(X = x) = 0.05^{(x-1)} 0.95$$

b) $P(X <= 2) = P(x=1) + P(x=2)$

$= 0.05^0 0.95 + 0.05^1 0.95$

$= 0.95 + 0.0475$

$= 0.9975$

c) $E[X] = 1/p = 1/0.95$

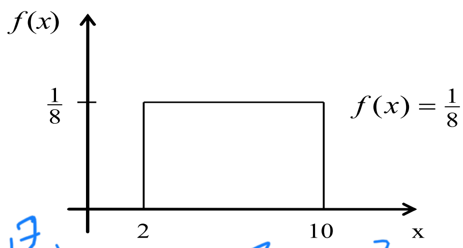
$= 1.0526$

CX2100 Prob & Stat

Tutorial #4

Continuous Probability Distribution

1. The figure below shows the graph of the uniform continuous distribution of a random variable that takes on values on the interval from 2 to 10. Find:



(a) P(X < 7)
(b) E(X)
(c) Var(X)

a) $\int_2^7 \frac{1}{8}x = \frac{7}{8} - \frac{2}{8} = \frac{5}{8} = 0.625$

b) $E[x] = \int_2^{10} \cancel{\frac{1}{8}} x = \left.\frac{x^2}{16}\right|_2^{10} = \frac{96}{16} = 6$ or eye power the mid is 6

2. The waiting time for one to be served in a queueing system is a random variable having an exponential distribution with an average of 4 minutes.
   (a) Determine the variance of the waiting time.
   (b) What is the probability that one has to wait for at least 10 minutes before being served?

3. The cumulative distribution function of the r.v. X is given below:

$$F(x) = \begin{cases} 0, & x < 1 \\ 1 - x^{-3}, & x \geq 1 \end{cases}$$

   (a) Determine the probability density function of X.
   (b) Calculate E[X] and var[X].

4. Given a r.v. having the normal distribution with $\mu = 16.2$ and $\sigma^2 = 1.5625$, find the probabilities that it will take on a value (use the standard normal distribution table)
   (a) greater than 16.8
   (b) between 13.6 and 18.8

5. Studies have shown that 22% of all patients taking a certain antibiotic will get a headache. Use the normal approximation to the binomial distribution to find the probability that among 50 patients taking this antibiotic
   (a) at least 10 will get a headache
   (b) at most 15 will get a headache

Answers

1.	(a) 0.625	(b) 6	(c) $\frac{16}{3}$
2.	(a) 16	(b) $e^{-2.5}$	
3.	(a) $f(x) = \begin{cases} 0, & x < 1 \\ 3x^{-4}, & x \geq 1 \end{cases}$	(b) 3/2	(c) 3/4
4.	(a) 0.3156	(b) 0.9624	
5.	(a) 0.6950	(b) 0.9382	

1 c) $Var[X] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx$

or

$\int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2$ $= 16/3$

2 a) $Var = \mu^2 = 4^2 = 16$

b) $P(X=x) = \frac{1}{\mu} e^{-\frac{x}{\mu}}$

$P(X > 10) = \int_{10}^{\infty} \frac{1}{4} e^{-\frac{x}{4}}\, dx$

3) a) p.d.f. $f(x) = \frac{d F(x)}{dx} = \begin{cases} 0, & x<1 \\ 3x^{-4}, & x \geq 1 \end{cases}$

b) $E[X] = \int_1^{\infty} x \cdot 3x^{-4}\, dx = \int_1^{\infty} 3x^{-3}\, dx = 3/2$

$=$

c) $Var[X] = E[X^2] - E[X]^2$

$= \int^{\infty} x^2 \cdot 3x^{-4}\, dx - (3/2)^2$

$= 3 - \frac{9}{4} = \frac{12-9}{4} = 3/4$

4) a) $Z = \frac{X - \mu}{\sigma}$

$P(X > 16.8) = 1 - P(X \le 16.8)$

$Z = \frac{16.8 - 16.2}{5/4} = 0.6 \frac{4}{5} = \frac{12}{25} = 0.48$

$= 1 - P(Z \le 0.48)$

$= 1 - 0.6844 = 0.3156$

b) $P(13.6 < X < 18.8) = P(-2.08 \le Z \le 2.08)$

$Z = \frac{18.8 - 16.2}{5/4} = 2.08$ $\quad 1 - 2P(Z \le -2.08)$

$1 - 2(1 - P(Z < 2.08))$

$Z = \frac{13.6 - 16.2}{5/4} = -2.08$ $\quad 1 - 2 + 2(0.9812)$

0.9812 $\quad = +0.9624$

5) 22% 50 $= \frac{11}{50}$ $\quad E(X) = 50 \cdot 0.22 = 11$

$Var(X) = 11 \times 0.78 = 8.58$

a) $P(X \ge 10) = P\left(Z > \frac{9.5 - 11}{\sqrt{8.58}}\right)$

$= P(Z \ge -0.512)$

$= P(Z \le 0.51)$

$= 0.6950$

b) $P(X \le 15) = P\left(Z < \frac{15.5 - 11}{\sqrt{8.58}}\right)$

$= P(Z < 1.54)$

$= 0.9382$