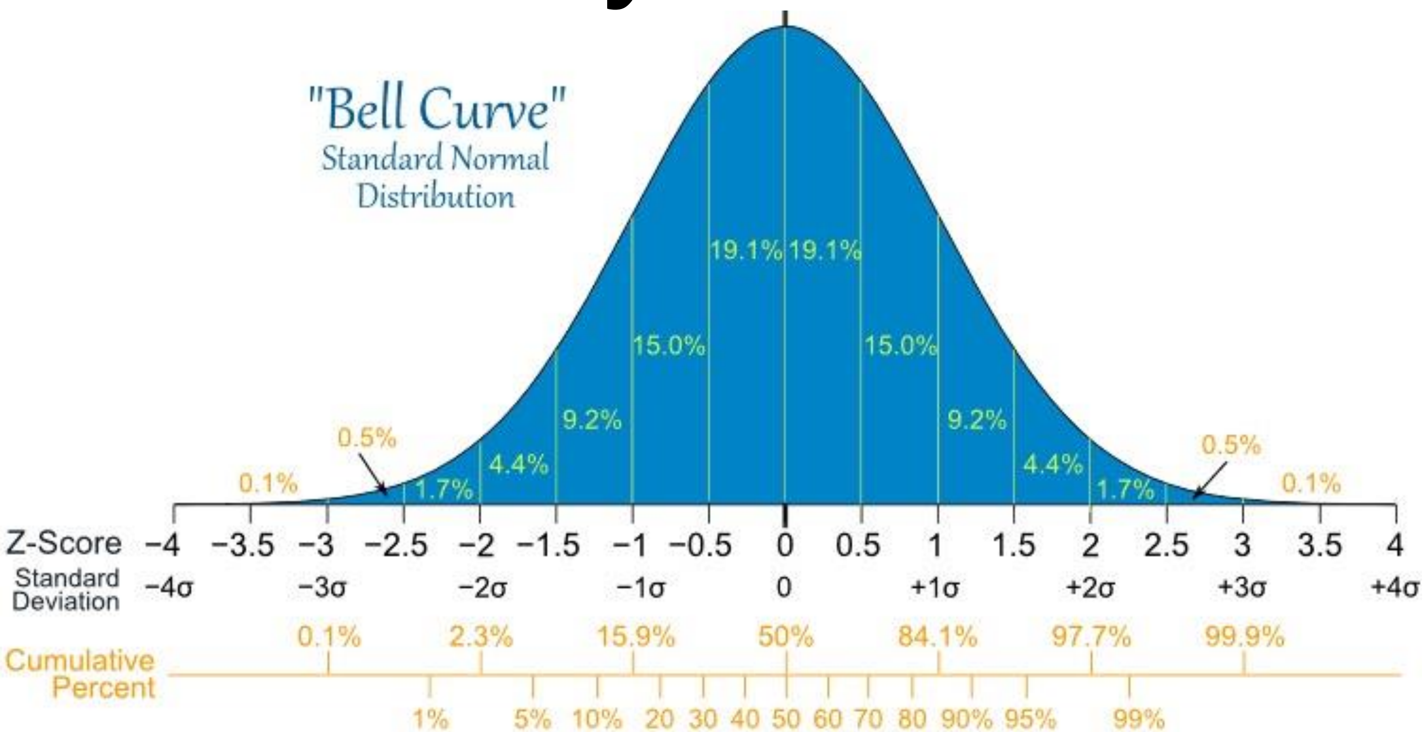
Unit 5
Probability Distribution



Unit Overview:

- **Probability Distribution**
- **Expected Number**
- **Expectation/Fair Game**
- **Probability of Binomial Distribution**
- **Probability of Hypergeometric Distribution**
- **Continuous Probability Distribution for Normal Distribution**
- **Normal Approximation for Discrete Data**
- **Normal Approximation for Binomial Distribution**

Lesson: Introduction to Probability Distribution

Definitions:

________________ ________________
- "X" (Captial letter X)
- a single value for each outcome in an experiment

________________ ________________
- values separate from each other
- finite number of outcomes
- data is counted

________________ ________________
- possible values of x are elements of real numbers
- infinite number of outcomes
- data is measured

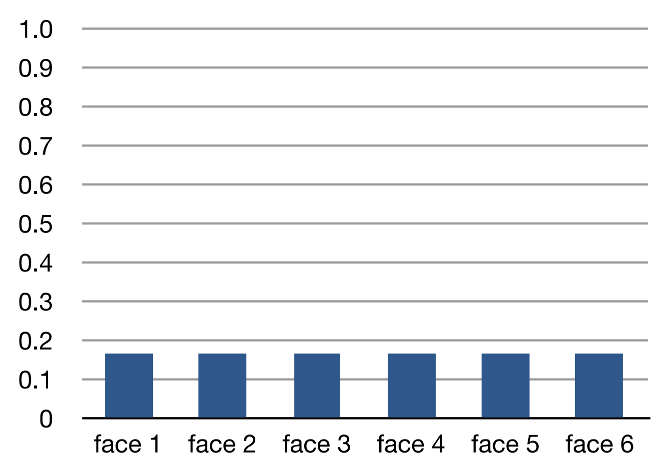
Examples:
Classify each of the following random variables as discrete or continuous:

Random Variables	Discrete OR Continuous
1) the number of phone calls made by a person	
2) the length of time the salesperson spent on the telephone	
3) a company's annual sales	
4) the distance from earth to the sun	
5) the number of widgets sold by a company	

________________ ________________ ________________outcomes of the distribution are equally likely in any single trial
- the sum of the probability is one

Example:

Random Variable, X	Probability, P(x)
1	$\frac{1}{6}$
2	$\frac{1}{6}$
3	$\frac{1}{6}$
4	$\frac{1}{6}$
5	$\frac{1}{6}$
6	$\frac{1}{6}$



- $E(X)$
- predicted average of all possible outcomes of a probability experiment

To calculate probability distribution

$$P(x) = \frac{1}{n}$$

$$E(X) = \sum_{i=1}^{n} x_i P(x_i)$$

Examples:
Given the following probability distribution, determine the expected values:

(a)

Random Variable, X	Probability, P(x)
5	0.3
10	0.25
15	0.45
$E(X) =$	

(b)

Random Variable, X	Probability, P(x)
1 000	0.25
100 000	0.25
1 000 000	0.25
10 000 000	0.25
$E(X) =$	

(c)

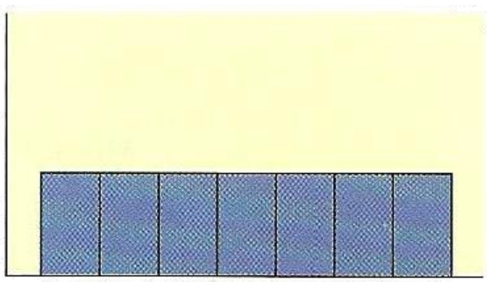
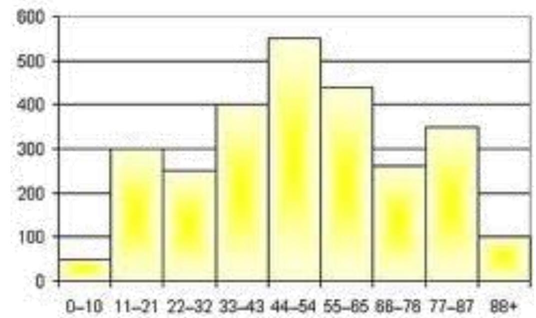
Random Variable, X	Probability, P(x)
1	$\frac{1}{6}$
2	$\frac{1}{5}$
3	$\frac{1}{4}$
4	$\frac{1}{3}$
5	$\frac{1}{20}$
$E(X) =$	

Lesson: Expectation / Fair Game

Definitions:

Probability in a Discrete Uniform Distribution:

$P(x) = \frac{1}{n}$, where n is the number of possible outcomes in the experiment

Uniform Distribution	Non-Uniform Distribution
	

____

- The expected outcome in a fair game is ZERO
- A game in which each player is not more likely to win than another

Fair Game	Not a Fair Game
$E(X) = 0$	$E(X) \neq 0$

Question Page 374 #2:

Explain whether each of the following experiments has a uniform probability distribution:

Experiment	Uniform/Non-Uniform Distribution
a) Selecting the winning number for a lottery	
b) Selecting three people to attend a conference	
c) Flipping a coin	
d) Generating a random number between 1 and 20 with a calculator	
e) Guessing a person's age	
f) Cutting a card from a well-shuffled deck	
g) Rolling a number (sum of the two dice) with two dice	

Example "GAMES"

Ex: You are playing a game with a deck of cards. You randomly chose a card. The following are the results. If it is a face card, you win \$10. If it is an even card, you lose \$3. If it is an odd card, you lose \$5. What is the expectation for this game? (Assume Ace = 1 = odd)

Event	Face	Odd	Even

Example "FAIR GAMES"

Ex: A die game cost \$3 to play. If you roll an even number, you win double the value. If you roll an odd number, you lose the value. Is this a fair game?

(Remember: If a game costs money upfront, then to be fair, the expectation has to have the same amount as the cost of the game.)

Roll a die	Random Variable, X	Probability, P(x)
1		
2		
3		
4		
5		
6		

Lesson: Probability of Binomial Distribution

Definitions:
Probability in a Binomial Distribution:

- ____________________ probability distribution of the number of successes in a sequence of n independent experiment
- 2 possible outcomes only: "____________________" or "____________________"
- "Success" $= p$
- "Failure" $= q$
- $p + q = 1$
- probability of each outcome remains constant throughout the trials (they are __________________)

Probability in a Binomial Distribution:

$$P(x) = \binom{n}{x} p^x q^{n-x}$$

Expectation for a Binomial Distribution:

$$E(X) = np$$

Examples from Textbook Questions Page 385 #1:

Situations	Binomial or not Binomial
a) a child rolls a die ten times and counts the number of 3s	
b) the first player in a free-throw basketball competition has a free-throw success rate of 88.4%. A second player takes over when the first player misses the basket.	
c) A farmer gives 12 of the 200 cattle in a herd an antibiotic. The farmer then selects 10 cattle at random to test for infections to see if the antibiotic was effective.	
d) A factory producing electric motors has a 0.2% defect rate. A quality-control inspector needs to determine the expected number of motors that would fail in a day's production.	

Textbook Example Page 379 Example #1:
A manufacturer of electronics components produces precision resistors designed to have a tolerance of $\pm 1\%$. From qualilty-control testing, the manufacturer knows that about one resistor in six is actually within just 0.3% of its normal value. A customer needs three of these precise resistors. What is the probability of finding exactly three such resistors among the first five tested?

Textbook Example Page 380 Example #2:

Tan's family moves to an area with a different telephone exchange, so they have to get a new telephone number. Telephone numbers in the new exchange start with 446, and all combinations for the four remaining digits are equally likely. Tan's favourite numbers are the prime numbers 2, 3, 5, and 7.

a) Calculate the probability distribution for the number of these prime digits in Tan's new telephone number.

b) What is the expected number of these prime digits in the new telephone number?

Textbook Example Page 384 Example 3:

The Choco-Latie Candies company makes candy-coated chocolates, 40% of which are red. The production line mixes the candies randomly and packages ten per box.

a) What is the probability that at least three candies in a given box are red?

b) What is the expected number of red candies in a box?

Lesson: Probability of Hypergeometric Distribution

Definitions:

Hypergeometric Distribution:

- involves a series of ____________________ trials, each with success or failure as the only possible outcomes
- outcomes doesn't involve success or failure ____________________
- probability of success changes as each trial is made
- random variable is the number of successful trials in an experiment
- probability involves ____

$$P(x) = \frac{\binom{a}{x}\binom{n-a}{r-x}}{\binom{n}{r}} \qquad E(X) = \frac{ra}{n}$$

$$a = number\ of\ successful\ outcomes$$
$$x = number\ of\ successes\ needed$$
$$n = sample\ space$$
$$r = number\ of\ dependent\ trials\ needed$$

Example from Textbook Page 404 #1:

Which of these random variables have a hypergeometric distribution?

Random Variables	Hypergeometric
a) the number of clubs dealt from a deck	
b) the number of attempts before rolling a six with a die	
c) the number of 3s produced by a random number generator	
d) the number of defective screws in a random sample of 20 taken from a production line that has a 2% defect rate	
e) the number of male names on a page selected at random from a telephone book	
f) the number of left-handed people in a group selected from the generate population	
g) the number of left-handed people selected from a group comprised equally of left-handed and right-handed people	

Example:

6 Jurors are needed from a pool of 8 men and 10 women.

a) Determine the probability distribution for the number of women on a civil-count jury selected.
b) What is the expected number of women on the jury?

Example from Textbook Page 401 Example 3:

A box contain seven yellow, three green, five purple, and six red candies jumbled together.

a) what is the expected number of red candies among five candies poured from the box?

Example from Textbook Page 402 Example 4:

In the spring, the Ministry of the Environment caught and tagged 500 raccoons in a widerness area. The raccoons were released after being vaccinated against rabies. To estimate the raccoon population in the area, the ministry caught 40 raccoons during the summer. Of these 15 had tags.
Estimate the raccoon population in the widerness area.

Example:

A hat contains 20 names, 12 of which are female. If five names are drawn from the hat,

a) what is the probability that there is exactly one female name is drawn?

b) What is the expected number of female names?

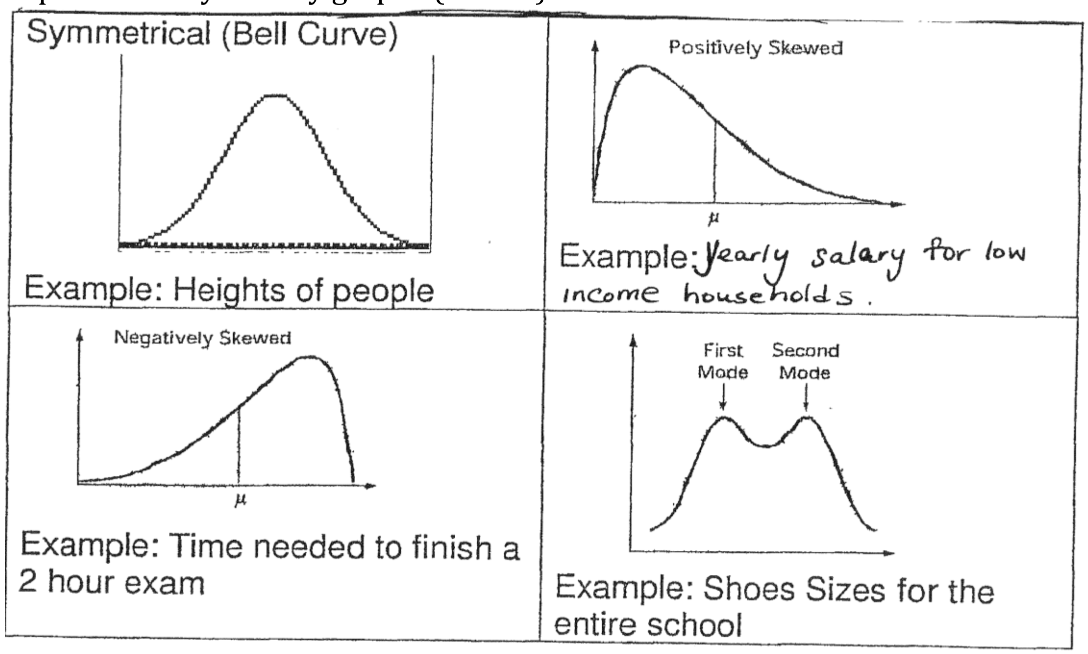
Example:

What is the probability that a card game of 13 cards in hand continas six spades, four hearts, two diamonds, and one club?

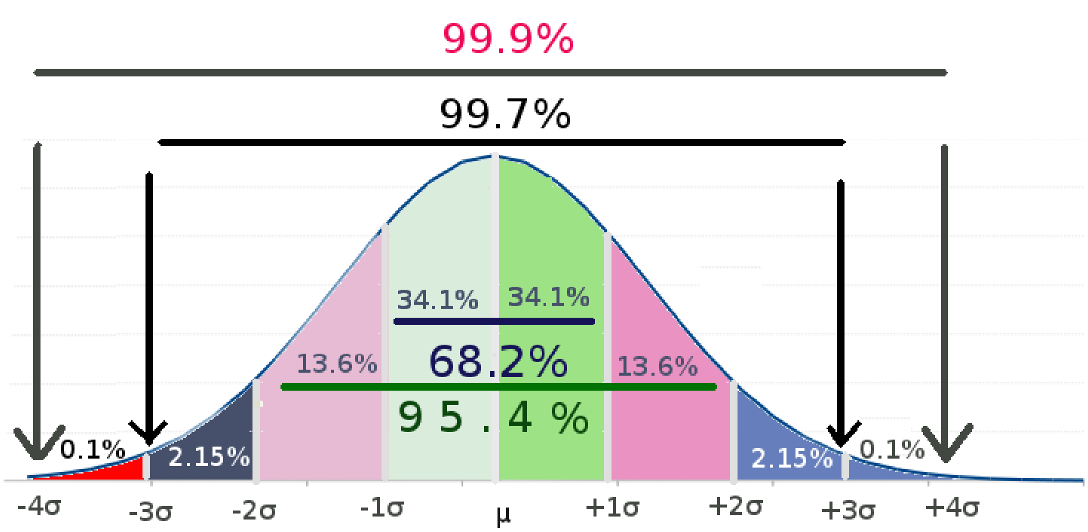
Lesson: Introduction to Continuous Probability Distribution

__________________ probability distributions:

- possible values of the random variable are any ____________________________
- often represented by density graphs (curves)



Characteristics of the Normal Distribution



- It is __________________ about the mean (median and mode are the same value).
- It is bell-shaped.
- The total area under the density curve equals __________________
- When considering continuous distributions, we are most often interested in determining the probability that a variable falls within a particular range of values
- The probability may be determined using the __________________ under the curve

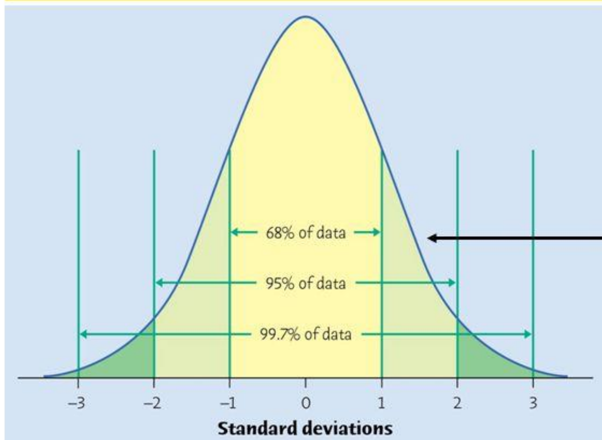
In a normal distribution, the proportions of the data set found within one, two and three standard deviations of the mean always were _________ (within one standard deviation of the mean), ____________ (within two standard deviations of the mean), and ___________ (within three standard deviations of the mean). That is the 68-95-99.75 rule.

Even though the mean and standard deviation affect the position and shape of the normal curve, changing their values does not distort the areas underneath the normal density function and as a result does not distort the proportions of observations found within these regions nor the probabilities determined for these regions. It is the relative position of the boundaries of these regions, measured in standard deviations from the mean, that are important and not the actual values themselves.

THE 68–95–99.7 RULE (aka – The Empirical rule)

In the Normal distribution with mean μ and standard deviation σ:

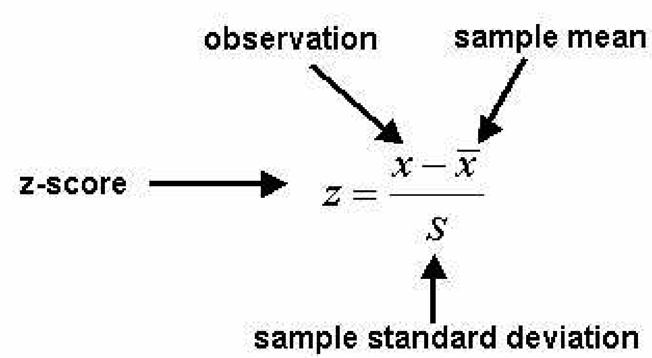
- Approximately **68%** of the observations fall within σ of the mean μ.
- Approximately **95%** of the observations fall within 2σ of μ.
- Approximately **99.7%** of the observations fall within 3σ of μ.



Normal distributions are abbreviated as ; N(μ,σ)

The Normal distribution with mean of 0 and standard deviation of 1 is called the standard Normal curve; N(0,1)

By calculating a _________________, which "standardizes" our observations to the standard normal distribution, we are determining the number of standard deviations that our observation is from the mean and we can use a z-score table to measure the area under the curve.



Strictly speaking, the z-score is given by the following equation.

$$z = \frac{x - \mu}{\sigma}$$

where μ is the true mean and σ is the true standard deviation. Since we usually do not know μ and σ, we estimate these values using $\bar{x}$ and s

Example: Assuming that a normal distribution occurs in a situation, the mean is 14 and standard deviation is 4.5. Determine the probability for the random variable is less than 13.01

Step #1: Convert a particular normal distribution to the standard normal distribution by calculating the z-score

Step #2: Identify the area under the bell curve that is needed for the probability

Step #3: Refer to the Normal Distribution Table for the probability of a z-score that is ____________

Example: Assuming that a normal distribution occurs in a situation, the mean is 14 and standard deviation is 4.5. Determine the probability for the random variable is greater than 17.

Step #1: Convert a particular normal distribution to the standard normal distribution by calculating the z-score

Step #2: Identify the area under the bell curve that is needed for the probability

Step #3: Refer to the Normal Distribution Table for the probability of a z-score that is ______________

Step #4: Determine the probability

Example: Assuming that a normal distribution occurs in a situation, the mean is 14 and standard deviation is 4.5. Determine the probability for the random variables between 10 to 17 inclusively.

Lesson: Normal Approximation and Normal Approximation for Binomial Distribution
Normal Approximation:

- use when model normal distribution for ______________________ data
  (data set is ______________________)
  (data falls into a symmetric unimodal bell shape)
  (discrete data)
- use ______________________ ______________________ [treat discrete data as they are continuous i.e. "60 students" will be treated as $P(59.5 < X < 60.5)$

Example

A factory that makes chocolate covered peanuts packages them in a box. The number of peanuts in the box is assumed to be normally distributed. They found that the boxes have a mean of 200 peanuts with a standard deviation of 12. If a box has fewer than 190 peanuts it will be rejected by quality control. Also, a box with more than 215 peanuts will result in excess costs to the company.

a) What is the probability that a box chosen at random will have exactly 200 peanuts in it?

b) What percent of the production would you expect to lie within acceptable values?

c) If your factory produces 200 000 boxes per shift, how many boxes would deemed rejects?

d) Comment on the quality control of your packaging process.

Lesson: Normal Approximation for Binomial Distribution

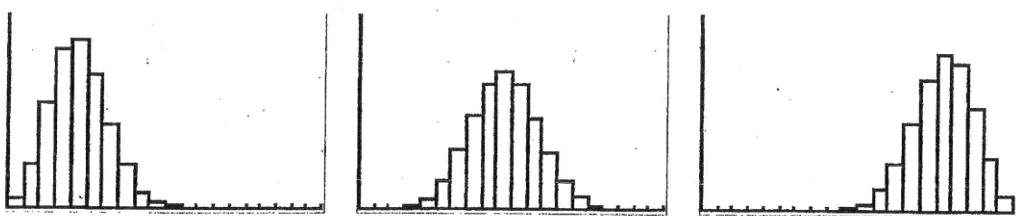
- The Normal distribution can be used to approximate Binomial probabilities when n is large and p is close to ________________. In answer to the question
- the approximation should only be used when both ______________and ______________.
- $(mean) \quad \mu \; or \; = np$
- $(standard\ deviation) \quad \sigma = \sqrt{npq}$

Example #1:

Bill is preparing for a test with 20 questions. He wants to determine his chances of passing if the certainty of answering each question is:

i) 20% ii) 45% iii) 75%

The distribution for the number of questions Bill would get correct is binomial. For all three situations, n = 20 but the values of p and q change. The following graphs are the probability distributions for each case:



p=0.2 & q=0.8 p=0.45 & q=0.55 p=0.75 & q=0.25

a) Which graph appears to have a normal distribution?

b) Which case meets the conditions required for normal distribution?

c) Calculate the mean and the standard deviation for normal approximation.

d) Determine the probability that the person passes the test using normal approximation.

Example #2:

The bottle of Cepsi know that they have 42% of the market. At a booth, 80 people take the Cepsi Challenge.

a) What type of distribution is the number of people who choose Cepsi?

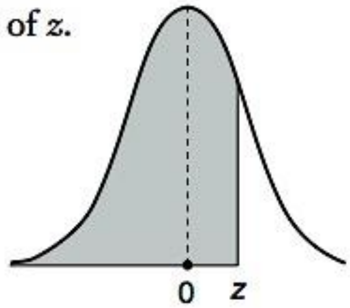
b) Can a normal approximation be made?

c) What is the mean and standard deviation of the approximate normal distribution?

d) What is the probability that the number of people that will choose Cepsi is between 25 and 40?

Areas Under the Normal Distribution Curve

The table lists the shaded area for different values of z.
The area under the entire curve is 1.



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
−2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
−2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
−2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
−2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
−2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
−2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
−2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
−2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
−2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
−1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
−1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
−1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
−1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
−1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
−1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
−1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
−1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
−1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
−1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
−0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
−0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
−0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
−0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
−0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
−0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
−0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
−0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
−0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986