# Statistics Workshop (STA-2023)

## Normal Probability Distribution and Sampling Distributions

**Properties of the Normal distribution**

A probability density function (pdf) is an equation used to compute probabilities of continuous random variables. For us to use the probability density function it must satisfy two properties:

1) the total area under the graph, or curve, of the equation over all possible values of the random variable must be equal to 1

2) the height of the graph, or curve, must be greater than or equal to 0 for all possible values of the random variable.

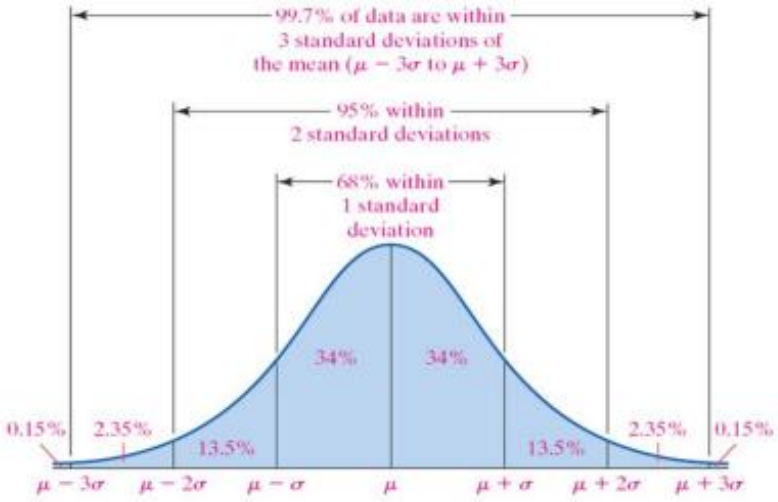
When we say area under the graph over an interval, or area under the curve, that represents the probability of observing a value of the random variable (x) in that interval. A continuous random variable is normally distributed, or has a normal probability distribution, if its relative frequency histogram has the shape of a normal curve (also called bell-shaped).

The properties of the Normal Density Curve are:

1. The normal curve is symmetric about its mean ($\mu$)
2. Because the mean = median = mode, the normal curve has a single peak and the highest point occurs at $x = \mu$
3. The normal curve has inflection points at $\mu - \sigma$ and $\mu + \sigma$, inflection points are the points on the curve where the curvature changes (in this case it is affected by $\sigma$)
4. The area under the normal curve is 1
5. The area under the normal curve to the right of $\mu$ equals the area under the curve to the left of $\mu$, which is equal to $\frac{1}{2}$
6. As x increases without bound, the graph approaches, but never touches, the horizontal axis. As x decreases without bound the graph approaches, but never touches, the horizontal axis.

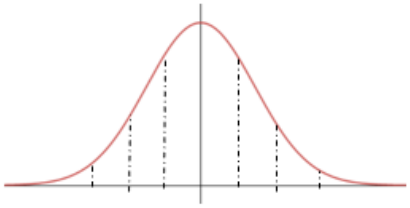
The Empirical Rule states that for data with a symmetric bell-shaped distribution:

1. About 68% of the data lies between $\mu - \sigma$ and $\mu + \sigma$
2. About 95% of the data lies between $\mu - 2\sigma$ and $\mu + 2\sigma$
3. About 99.7% of that data lies between $\mu - 3\sigma$ and $\mu + 3\sigma$

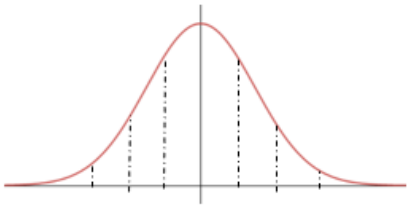



1) Suppose the birth weights of full-term babies are normally distributed with mean $\mu = 3200$ grams and standard deviation $\sigma = 430$ grams.

   a. Draw a normal curve with the parameters labeled.



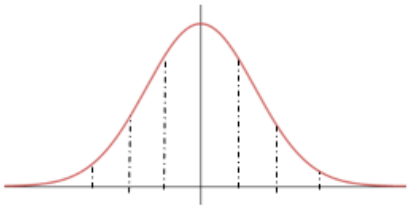
   b. Shade the region that represents the proportion of full-term babies who weigh more than 4320 grams



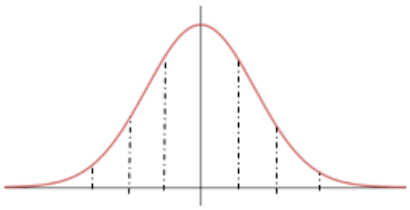
   c. Suppose the area under the normal curve to the right of $X = 4320$ is 0.0228. Provide an interpretation of this result.

   The probability is _____ that the birth weight of a randomly chosen full-term baby in this population is [Less than/More than] _____ grams.

2) The heights of 10-year-old males are normally distributed with mean of $\mu = 55.9$ inches and $\sigma = 5.7$ inches.
   a. Draw a normal curve with the parameters labeled.



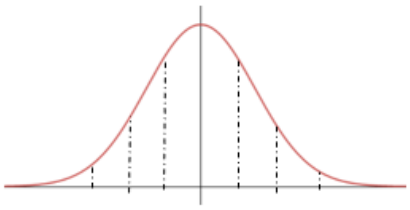
   b. Shade the region that represents the proportion of 10-year-old males who are less than 44.5 inches tall.



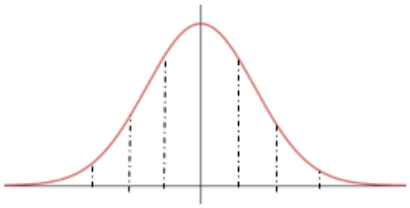
   c. What is the area, or probability, of this shaded region?

3) The lives of refrigerators are normally distributed with mean $\mu = 17$ years and standard deviation $\sigma = 2.5$ years.

   a. Draw a normal curve with parameters labeled.



   b. Shade the region that represents the proportion of refrigerators that last for more than 19.5 years.



   c. What is the area, or probability, of this shaded region?

**Applications of the Normal Distribution**

We use z-scores to help find the area under a normal curve by hand. Suppose that the random variable X is normally distributed with mean $\mu$ and standard deviation $\sigma$. Then the random variable

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

is normally distributed with mean $\mu = 0$ and standard deviation $\sigma = 1$. We are transforming X into a standard normal random variable Z. We use the standard normal table to find the area corresponding to the z-score.

For example, a z-score of -1.17 would have an area to the left of 0.1210.

Figure 15

| z | .00 | .01 | .02 | .03 | .04 | .05 | .06 | .07 | .08 |
|---|---|---|---|---|---|---|---|---|---|
| −3.4 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 | 0.0003 |
| −3.3 | 0.0005 | 0.0005 | 0.0005 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 | 0.0004 |
| −3.2 | 0.0007 | 0.0007 | 0.0006 | 0.0006 | 0.0006 | 0.0006 | 0.0006 | 0.0005 | 0.0005 |
| −1.4 | 0.0808 | 0.0793 | 0.0778 | 0.0764 | 0.0749 | 0.0735 | 0.0721 | 0.0708 | 0.0694 |
| −1.3 | 0.0968 | 0.0951 | 0.0934 | 0.0918 | 0.0901 | 0.0885 | 0.0869 | 0.0853 | 0.0838 |
| −1.2 | 0.1151 | 0.1131 | 0.1112 | 0.1093 | 0.1075 | 0.1056 | 0.1038 | 0.1020 | 0.1003 |
| −1.1 | 0.1357 | 0.1335 | 0.1314 | 0.1292 | 0.1271 | 0.1251 | 0.1230 | 0.1210 | 0.1190 |
| −1.0 | 0.1587 | 0.1562 | 0.1539 | 0.1515 | 0.1492 | 0.1469 | 0.1446 | 0.1423 | 0.1401 |

*Finding the probability of a normal random variable*

Finding the area to the left of X: convert the value of X to a z-score. Then use the standard normal table to find the area to the left of the X value. Or use technology to find the area $P(x < X)$.

Finding the area to the right of X: convert the value of X to a z-score. Then use the standard normal table to find the area to the left of the X value. Then use the complement rule to find the area to the right of the X value. Or use technology to find the area $P(x > X)$.

*Finding the value of a normal random variable*

Often, we want to find the value of a normal random variable that corresponds to a certain probability. In that case, we can find the value applying the method above in a reverse way. Draw a normal curve and shade the desired area with given probability, use the z-table to find the z-score that corresponds to the shaded area, and then obtain the value of $x$ using the formula, $= \frac{x-\mu}{\sigma}$. Or use technology to find the value.

4) In a certain city, the average 20- to 29-year-old man is 71.3 inches tall, with a standard deviation of 3.5 inches, while the average 20- to 29-year-old woman is 63.7 inches tall, with a standard deviation of 3.9 inches. Who is relatively taller a 75-inch man or a 70-inch woman?

   a. The z-score for a 75-inch man is _____

b. The z-score for a 70-inch woman is _____

c. Therefore a _____ is relatively taller than a _____ because the z-score _____ is larger than _____.

5) Determine the area under the standard normal curve that lies to the right of

a. $Z = -1.32$

b. $Z = -.96$

c. $Z = 0.35$

d. $Z = 0.80$

6) Suppose that the lifetimes of light bulbs are approximately normally distributed, with a mean of 56 hours and a standard deviation of 3.5 hours. With this information answer the following questions.

a. What proportion of light bulbs last more than 60 hours?

b. What proportion of light bulbs 50 hours or less?

c. What proportion of light bulbs will last between 57 and 61 hours?

7) The time required for an automotive center to complete an oil change service on an automobile approximately follows a normal distribution, with a mean of 19 minutes and a standard deviation of 4 minutes.

a. The automotive center guarantees customers that the service will take no longer than 20 minutes. If it does take longer, the customer will receive the service for half-price. What percent of customers will receive the service for half-price?

b. If the automotive center does not want to give the discount to more than 4% of its customers, how long should it make the guaranteed time limit?

**The Normal Approximation to the Binomial Probability Distribution**

Review the criteria for the binomial probability experiment:

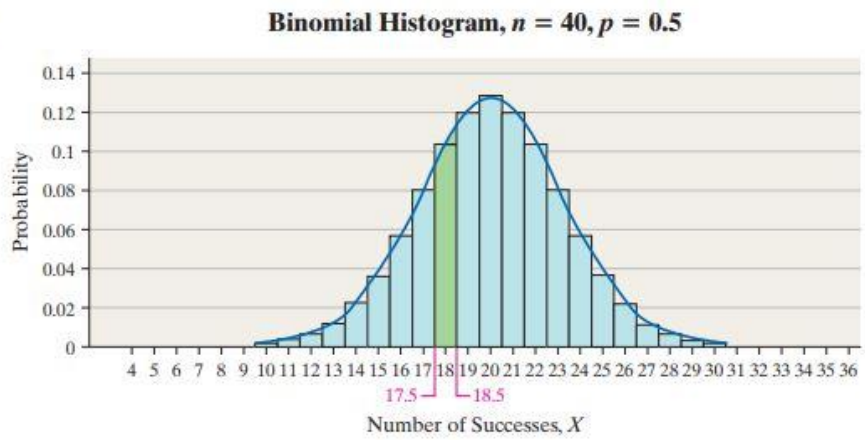
1. The experiment is preformed n independent times, each repetition is called a trial. Independence means that the outcome of one trial will not affect the outcome of the other trials.
2. For each trial there are two mutually exclusive outcomes: success or failure
3. The probability of success, p, is the same for each trial of the experiment.

If $np(1-p) \geq 10$, the binomial random variable X is approximately normally distributed, with mean $\mu_X = np$ and standard deviation $\sigma_X = \sqrt{np(1-p)}$.

Because we are using a continuous density function to approximate the probability of a discrete random variable, we add and/or subtract 0.5 as a correction for continuity.

For example, if we want to approximate $P(X = 18)$, we compute $P(17.5 \leq x \leq 18.5)$.

Figure 36




Some general rules for approximating the probabilities:

- $P(x = a)$ can be calculated by $P(x = a) = P(a - 0.5 \leq x \leq a + 0.5)$
- $P(x \leq a)$ can be calculated by $P(x \leq a) = P(x \leq a + 0.5)$
- $P(x \geq a)$ can be calculated by $P(x \geq a) = P(x \geq a - 0.5)$
- $P(a \leq x \leq b)$ can be calculated by $P(a - 0.5 \leq x \leq b + 0.5)$

8) Whether a randomly selected individual has blood-type O-negative is a binomial random variable. Assume the probability will be approximated using the normal distribution.
   a. Describe the area under the normal curve that will be computed in order to determine the probability that the number of people with blood-type O-negative is between 14 and 25, inclusive.

9) Compute $P(X)$ using the binomial probability formula. Then determine whether the normal distribution can be used to estimate this probability. If so, approximate $P(X)$ using the normal distribution and compare the result to the exact. $n = 62, p = 0.51, \text{and } X = 23$

   a. $P(X) =$ _____ (round to four decimal places)

   b. Can the normal distribution be used to approximate this probability? If yes, Approximate the $P(X)$ using the normal distribution.

   c. By how much do the exact and approximated probabilities differ?

10) In studies for a medication, 7 percent of patients gained weight as a side effect. Suppose 526 patients are randomly selected. Use the normal approximation to the binomial to approximate the probability that

   a. exactly 37 patients will gain weight as a side effect.

   b. no more than 37 patients will gain weight as a side effect.

   c. at least 48 patients will gain weight as a side effect. What does this result suggest?

   d. Since is (More/Fewer) than 5% of the patients, this suggests that the proportion of patients that gain weight as a side effect is (Greater than/Equal to/Less than) 0.05

11) According to a study, 68% of all males between the ages of 18 and 24 live at home. (Unmarried college students living in a dorm are counted as living at home.) Suppose that a survey is administered and 175 of 265 respondents indicated that they live at home.

   a. Use the normal approximation to the binomial to approximate the probability that at least 175 respondents live at home.

   b. Do the results from part (a) contradict the study?

**Distribution of the Sample Mean**

The sampling distribution of a statistic is a probability distribution for all possible values of the statistic computed from a sample of size $n$.

The sampling distribution of the sample mean $\bar{x}$ is the probability distribution of all possible values of the random variable $\bar{x}$ computed from a sample of size $n$ from a population with mean $\mu$ and standard deviation $\sigma$.

The idea behind obtaining the sampling distribution of the sample mean is:

1. Obtain a simple random sample of size $n$.
2. Compute the sample mean.
3. Assuming that we are sampling from a finite population, repeat steps 1 and 2 until all distinct simple random samples of size $n$ have been obtained.

Suppose that a simple random sample of size $n$ is drawn from a population with mean $\mu$ and standard deviation $\sigma$. The sampling distribution of $\bar{x}$ has a mean of $\mu_{\bar{x}} = \mu$ and a standard deviation of $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$. The standard deviation of the sampling distribution of $\bar{x}$ is called the standard error of the mean.

If a random variable X is normally distributed, the sampling distribution of the sample mean, $\bar{x}$, is normally distributed.

The central limit theorem states: regardless of the shape of the underlying population, the sampling distribution of $\bar{x}$ becomes approximately normal as the sample size, $n$, increases.

If the distribution of the population is unknown or not normal, then the distribution of the sample mean is approximately normal provided that the sample size is greater than or equal to 30.

12) Determine $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ from the given parameters of the population and sample size. $\mu = 78, \sigma = 32, n = 86$.

$\mu_{\bar{x}} =$ _____ and $\sigma_{\bar{x}} =$ _____

13) A simple random sample of size n = 91 is obtained from a population with μ = 94 and σ = 20.

a. Describe the sampling distribution of $\bar{x}$

The distribution is (Skewed right/Approximately normal/Uniform/Unknown)

$\mu_{\bar{x}} =$ _____ and $\sigma_{\bar{x}} =$ _____

b. What is $P(\bar{x} > 96)$?

c. What is $P(\bar{x} > 89.3)$?

d. What is $P(91.2 < \bar{x} < 97.4)$?

14) Suppose the lengths of the pregnancies of a certain animal are approximately normally distributed with mean μ = 297 and standard deviation σ = 30.

a. What is the probability that a randomly selected pregnancy lasts less than 279 days?

$P(X < 279) =$

If 100 pregnant individuals were selected independently from this population, we would expect _____ pregnancies to last (More than/Less than/Exactly) 279 days. (round to the nearest integer).

b. Suppose a random sample of $n = 20$ pregnancies is obtained. Describe the sampling distribution of the sample mean length of pregnancies.

The sampling distribution of $\bar{x}$ is (Normal/Skewed right/Skewed left)

$\mu_{\bar{x}} =$ _____ and $\sigma_{\bar{x}} =$ _____

c. What is the probability that a random sample of 20 pregnancies has a mean gestation period of 279 days or less?

$P(\bar{x} < 279) =$

If 100 independent random samples of size $n = 20$ were obtained from this population, we would expect _____ pregnancies to last (More than/Less than/Exactly) 279 days. (round to the nearest integer).

d. What is the probability that a random sample of 38 pregnancies has a mean gestation period of 279 days or less?

$\mu_{\bar{x}} =$ _____ and $\sigma_{\bar{x}} =$ _____

$P(\bar{x} < 279) =$

If 100 independent random samples of size $n = 38$ were obtained from this population, we would expect _____ pregnancies to last (More than/Less than/Exactly) 279 days. (round to the nearest integer).

e. What might you conclude if a random sample of 38 pregnancies resulted in a mean gestation period of 279 days or less?

This result would be (Unusual/Expected) so the sample likely came from a population whose mean gestation period is (Greater than/Less than/Equal to) 279 days.

**Distribution of the Sample Proportion**

Suppose that a random sample of size $n$ is obtained from a population in which each individual either does or does not have a certain characteristic. The sample proportion, denoted with $\hat{p}$ (p-hat), is given by

$$\hat{p} = \frac{x}{n}$$

where $x$ is the number of individuals in the sample with the specified characteristic.

The sample proportion, $\hat{p}$, is a statistic that estimates the population proportion, $p$.

For a simple random sample of size $n$ with a population proportion, $p$,

- The shape of the sampling distribution of $\hat{p}$ is approximately normal provided $np(1-p) \geq 10$
- The mean of the sampling distribution of $\hat{p}$ is $\mu_{\hat{p}} = p$
- The standard deviation of the sampling distribution of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$
- The sample size **cannot** be more than 5% of the population size.

15) Describe the sampling distribution of $\hat{p}$. Assume the size of the population is 30,000. $n = 800, p = 0.64$.

a. The shape of the sampling distribution is (Approximately normal/Not normal) because $n \leq 0.05N$ and ($np(1-p) \geq 10$ or $np(1-p) \leq 10$).

b. Determine the mean and standard deviation of the sample proportion.

$\mu_{\hat{p}} =$ _____ and $\sigma_{\hat{p}} =$ _____

16) Suppose a simple random sample of size $n = 200$ is obtained from a population whose size is and whose population proportion with a specified characteristic is $p = 0.58$.

a. Describe the sampling distribution of $\hat{p}$.

The shape of the sampling distribution is (Approximately normal/Not normal) because $n \leq 0.05N$

and ($np(1-p) \geq 10$ or $np(1-p) \leq 10$).

b. Determine the mean and standard deviation of the sampling distribution of $\hat{p}$.

$\mu_{\hat{p}} =$ _____ and $\sigma_{\hat{p}} =$ _____

c. What is the probability of obtaining $X = 125$ or more individuals with the characteristic? That is,

what is $P(\hat{p} \geq 0.63)$?

$P(\hat{p} \geq 0.63)$

d. What is the probability of obtaining $X = 116$ or fewer individuals with the characteristic? That is,

what is $P(\hat{p} \geq 0.58)$?

$P(\hat{p} \geq 0.58)$

17) According to a survey in a country, 30% of adults do not own a credit card. Suppose a simple random sample of 800 adults is obtained.

a. Describe the sampling distribution of $\hat{p}$. Determine the mean and standard deviation of the sampling

distribution of $\hat{p}$.

The shape of the sampling distribution is (Approximately normal/Not normal) because $n \leq 0.05N$

and ($np(1-p) \geq 10$ or $np(1-p) \leq 10$).

$\mu_{\hat{p}} =$ _____ and $\sigma_{\hat{p}} =$ _____

b. What is the probability that in a random sample of 800 adults, more than 32% do not own a credit card?

$P(\hat{p} > 0.32) =$

If 100 different random samples of 800 adults were obtained, one would expect _____ to result in more than 32% not owning a credit card.

c. What is the probability that in a random sample of 800 adults, between 28% and 34% do not own a credit card?

$P(0.28 < \hat{p} < 0.34) =$

If 100 different random samples of 800 adults were obtained, one would expect _____ to result in between 28% and 34% not owning a credit card.

d. Would it be unusual for a random sample of 800 adults to result in 273 or fewer who do not own a credit card? Why?

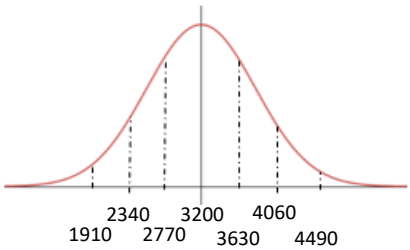
$P(\hat{p} \leq 0.34) =$

The results is (Unusual/Not unusual) because the probability that $\hat{p}$ is less than or equal to the sample proportion is _____, which is (Greater than/Less than) 5%.

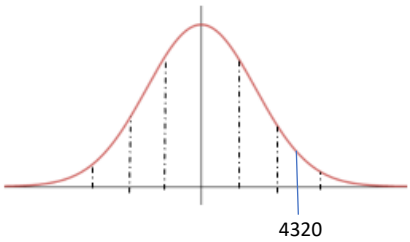
**Answer Key**

1.

(a)



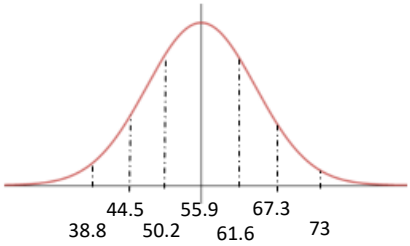

(b)



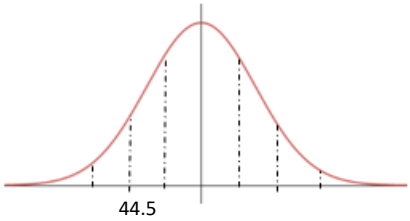

(c) 0.0228, More Than, x = 4320

2.

(a)



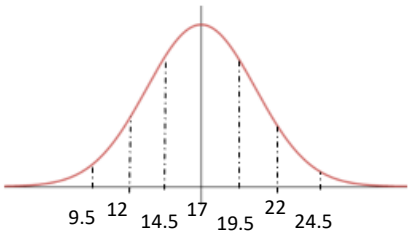

(b)



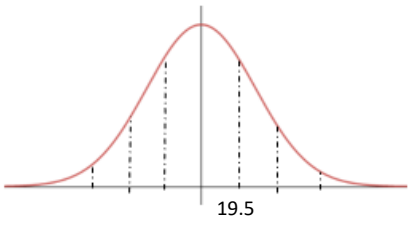

(c) 0.025, hint: use empirical rule

3.

(a)




(b)




(c) 0.16, hint: use empirical rule

4. (a) 1.06 (b) 1.62 (c) 70-inch woman, 75-inch man, 1.62, 1,06

5. (a) 0.9066 (b) 0.8315 (c) 0.3632 (d) 0.2119

6. (a) 0.1265 (b) 0.0432 (c) 0.3110

7. (a) 40.13% (b) 26

8. between 13.5 and 25.5

9. (a) $P(23) = 0.0093$ (b) $\mu_X = 31.32, \sigma_X = 3.936$, $P(22.5 \leq X \leq 23.5) = 0.0093$ (c) 0

10. $\mu_X = 36.82, \sigma_X = 5.852$

(a) $P(36.5 \leq X \leq 37.5) = 0.0681$ (b) $P(x \leq 37.5) = 0.5463$

(c) $P(x \geq 47.5) = 0.0340$ (d) 48, More, Greater than

11. $\mu_X = 180.2, \sigma_X = 7.594$

(a) $P(x \geq 174.5) = 0.7736$ (b) No, because $P(x \geq 175)$ is greater than 0.05

12. $\mu_{\bar{x}} = 78, \sigma_{\bar{x}} = 3.451$

13. (a) Approximately normal, $\mu_{\bar{x}} = 94, \sigma_{\bar{x}} = 2.097$

(b) 0.1701 (c) 0.9875 (d) 0.8567

14. (a) 0.2743, 27, Less than (b) Normal, $\mu_{\bar{x}} = 297, \sigma_{\bar{x}} = 6.708$

(c) 0.0036, 0, Less than (d) $\mu_{\bar{x}} = 297, \sigma_{\bar{x}} = 4.867$, 0.0001, 0, Less than

(e) Unusual, Less than

15. (a) Approximately Normal, $np(1-p) \geq 10$ (b) $\mu_{\hat{p}} = 0.64, \sigma_{\hat{p}} = 0.017$

16. (a) Approximately normal, $np(1-p) \geq 10$ (b) $\mu_{\hat{p}} = 0.58, \sigma_{\hat{p}} = 0.035$

(c) 0.0766 (d) 0.1265

17. (a) Approximately normal, $np(1-p) \geq 10$, $\mu_{\hat{p}} = 0.30, \sigma_{\hat{p}} = 0.016$

(b) 0.1056, 11 (c) 0.8881, 89 (d) 0.9938, Not Unusual, 0.34, Greater than