HW 1-1

1. Below are the ages of the Oscar winners for best actress from 2004 to 2017: 29, 61, 32, 33, 45, 29, 62, 22, 44, 54, 26, 28, 60. Find the mean (rounded to the nearest tenth) and median of these data.

2. Using the boxplot below:

![Boxplot](image)

   a. Identify the values of the 5-number summary.
   b. What is the value of the IQR?

3. Below are the final scores for a week’s N.F.L. games:
   Home team scores: 36, 37, 16, 20, 17, 10, 33, 19, 34, 30, 6, 17, 14, 31, 35, 18
   Visiting team scores: 16, 34, 23, 23, 6, 26, 20, 14, 17, 27, 34, 28, 20, 24, 14, 17
   Find the mean and median of each data set.

4. The data set: 28, 33, 38, 45, 45, 45, 48, 50, 55, 75 represents the ages of randomly selected teachers from a small elementary school. Which value is greater, the mean or median?

5. Given the data set: x, 20, 26, 34, 48, find a value of x which would make the mean and median equal. (Hint: there is more than one possible answer).

6. Given the five-number summary, 10, 16, 18, 24, 32; sketch a boxplot below:

![Boxplot](image)
7. A researcher takes a random sample of 100 cars parked at the Palisades Mall at noon on Friday. Briefly describe the population that the researcher is attempting to analyze.

HW 1-2

Below are the ages of the gold medalists from the 2015 World Athletics Championships in Beijing, China:

Male: 29, 23, 26, 26, 32, 19, 24, 23, 33, 27, 32, 25, 21, 28, 25, 26, 32, 26, 26, 27

1. Find the five number summary for each data set. Create box-and-whisker plots grid below.

2. Write a few sentences comparing the two data sets.

3. Given the boxplot below, write three different (approximate) intervals which contain 50% of the data.
4. Find the value of the median in the graph below:

Ages of Faculty members at Forks High School

2 2
3 12445679
4 002234567
5 025
6
7 1

Key: 4|7 = 47 years of age

5. Find the value of the median in the graph below:

6. The box-and-whisker plot below represents the weights of a random sample of a certain breed of dog.

Describe the center, shape and spread of these data.
7. In the graph below, identify a likely value of the median. Describe the shape of the distribution and make a statement about the mean based on the shape.

HW 1-3

1. In the following data set, which is greater, the mean or the median?
   22, 26, 28, 32, 38, 44, 45, 45, 48, 50, 82.
   What does this comparison tell us about the shape of the distribution?

2. Steve calculated the statistics for a data set of 20 observations and found the mean to be 74.2 and the median to be 65.5. He later found that he had incorrectly entered one of the values, entering 181 instead of 81. Upon correcting his mistake, how would the mean and median change?

3. Given the boxplot below represents the number of eggs laid per day at a particular farm:

The following day, nine eggs are laid. How does this change the boxplot?
4. The back-to-back stem-and-leaf plots below represent the weights (in pounds) of a random sample of students in a high school.

```
<table>
<thead>
<tr>
<th>Girls</th>
<th>Boys</th>
</tr>
</thead>
<tbody>
<tr>
<td>9, 2</td>
<td>9</td>
</tr>
<tr>
<td>6, 1, 0, 0</td>
<td>10</td>
</tr>
<tr>
<td>8, 7</td>
<td>11</td>
</tr>
<tr>
<td>6, 6, 5, 5, 5, 4, 2</td>
<td>12</td>
</tr>
<tr>
<td>7, 1, 0</td>
<td>13</td>
</tr>
<tr>
<td>9, 8</td>
<td>14</td>
</tr>
<tr>
<td>8, 6, 2, 0, 0</td>
<td>15</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
</tr>
<tr>
<td>8, 0</td>
<td>18</td>
</tr>
<tr>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
</tr>
</tbody>
</table>
```

a. Why would a student with this response not get full credit for the AP exam?
b. Describe the shapes of the two distributions, in context.

5. Create a dotplot for the following data, the ages of players on the New York Knicks:

6. Create a boxplot from the stem-and-leaf plot below:

```
Race Running Times in Seconds

<table>
<thead>
<tr>
<th>Stem</th>
<th>Leaves</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>2 6</td>
</tr>
<tr>
<td>13</td>
<td>0 2 5</td>
</tr>
<tr>
<td>14</td>
<td>1 2 4 6</td>
</tr>
<tr>
<td>15</td>
<td>2 3 7 8</td>
</tr>
<tr>
<td>16</td>
<td>1 2 4 6 8</td>
</tr>
<tr>
<td>17</td>
<td>5 7 8</td>
</tr>
<tr>
<td>18</td>
<td>1 3</td>
</tr>
</tbody>
</table>

Key: 14 2 = 14.2 seconds
```
7. The data below is a stem-and-leaf diagram representing the ages of Academy Award winners for best actor and actress. Compare the two distributions.

<table>
<thead>
<tr>
<th>Best Actress</th>
<th>Best Actor</th>
</tr>
</thead>
<tbody>
<tr>
<td>44221 2</td>
<td>9</td>
</tr>
<tr>
<td>9999999988888877776666655555 3</td>
<td>0012223444</td>
</tr>
<tr>
<td>44433333322221111000 3</td>
<td>5566667777788888999999</td>
</tr>
<tr>
<td>998888766655555 3</td>
<td>000111111222223333344</td>
</tr>
<tr>
<td>22111111 4</td>
<td>5555567777888899999</td>
</tr>
<tr>
<td>99855 4</td>
<td>001122334</td>
</tr>
<tr>
<td>4 5      001122334</td>
<td></td>
</tr>
<tr>
<td>5 5       56779</td>
<td></td>
</tr>
<tr>
<td>3211100 6</td>
<td>0000222</td>
</tr>
<tr>
<td>6 6</td>
<td></td>
</tr>
<tr>
<td>4 7</td>
<td></td>
</tr>
<tr>
<td>7 6</td>
<td></td>
</tr>
<tr>
<td>0 8</td>
<td></td>
</tr>
</tbody>
</table>

Key: 4|3 = 43 years old

HW 1-4

1. Create a reasonable box-and-whisker plot from the histogram below:

![Histogram of Salaries](image)
2. Find the approximate value of the 70th percentile from the graph below:

![Fuel Economy for a Random Sample of 2015 Model Year Vehicles](image)

3. A data set has the following statistics:
   Minimum = 17   Q1 = 23   Median = 32   Q3 = 35   Maximum = 53
   Do these data have any outliers? Justify your answer.

4. Find the approximate value of the median in the following graph:

![Heights of Black Cherry Trees](image)

5. In the following graph:

![Histogram](image)

Identify the values of the median and the 80th percentile.
6. Given the boxplot below, write three different (approximate) intervals which contain 50% of the data.

![Boxplot Image]

**WRITTEN RESPONSE – TO BE HANDED IN**

7. A polling firm wants to estimate the percentage of individuals who will vote for the republican candidate for mayor. They take a sample of 250 likely voters in Spring Valley. Describe the population of interest in this example.

**HW 1-5**

1. A botanist measures the height, in feet, of a certain type of plants and finds the following: mean = 2.73 ft, median = 2.4 ft, IQR = 1.2 ft, st dev = 0.54 ft. He then wishes to convert his measurements to inches, find the new mean, median, IQR and standard deviation.

2. Students in a school took the Geometry Regents exam and found that mean was 62.3, IQR = 11.5 and the standard deviation was 6.4. It was found that the state made several errors on the exam, so it was decided to add five points to every student’s score. Find the new values of mean, IQR and standard deviation.

3. It was found that the daily high temperature in Spring Valley in September had a mean of 21.5°C, a median of 20.5°C, and IQR of 9.75°C and a standard deviation of 1.85°C. When temperature is converted from Celsius to Fahrenheit the formula $F = \frac{9}{5}C + 32$. Find the mean, median, IQR and standard deviation in degrees Fahrenheit.

4. A physics teacher gives a chapter test in which the scores have the following information:
   Mean = 60, Median = 66, Standard Deviation = 12, IQR = 34.
   Determine the values of each of above statistics if the teacher curves the scores by adding 10% and then adding 10 points to each score.
5. In the following boxplot,

Determine the approximate values of mean, median and IQR. Each score is multiplied by 3 and 10 points added to arrive at a score out of 100, now determine the transformed mean, median and IQR.

6. In the graph below, estimate the value of the 90\textsuperscript{th} percentile.

The following dot plot shows the daily high temperature in Kats, Colorado in April. Each dot represents a different day.

WRITTEN RESPONSE – TO BE HANDED IN

7. The graph below represents the ages of 36 teachers in an elementary school. Ages of the teachers ranged from 23 to 68 years.
a. Describe the shape of this distribution.
b. Which summary statistic, the mean or the median, should be used to report that the overall ages of the teachers in younger? Explain.
c. The midrange is defined as \( \frac{\text{maximum} + \text{minimum}}{2} \). Compute the value of the midrange for the data on the preceding page. Is midrange considered a measure of center or a measure of spread? Explain.

**HW 1-6**

1. The following table shows the distribution of number of children in a household for a random selection of 60 homes:

<table>
<thead>
<tr>
<th># children</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>frequency</td>
<td>6</td>
<td>11</td>
<td>18</td>
<td>12</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Describe the shape of the distribution

2. The cumulative frequency histogram below represents the number of cell phones owned per household.

Find the values of the median and the IQR.
3. In the dot plot below:

![Dot Plot]

Estimate the values of the 40th percentile and the 90th percentile.

4. A data set had a mean of 45, a median of 52 and a standard deviation of 12. Each individual observation was changed using the following formula:
   \[ x_{new} = 1.4x - 2.8 \]
   Find the new values of the mean, median and standard deviation.

5. The following boxplot represents the scores on a social studies quiz.

![Boxplot]

The teacher converts the scores using the formula \[ x_{new} = 1.8x + 10 \]. Approximate the values for median, IQR and range for the transformed scores. Do you think the mean will be greater than, less than, or equal to the median? Justify your answer.

6. A researcher takes a sample 100 male college graduates in Pennsylvania to determine their average annual salary. What is the population?
In a research study of third-grade students, a test of reading ability the Degree of Reading Power (DRP) was administered to 44 students. The following stem-and-leaf displays these data:

<table>
<thead>
<tr>
<th></th>
<th>1 445899</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>255667789   Key: 3</td>
</tr>
<tr>
<td>3</td>
<td>1334555589</td>
</tr>
<tr>
<td>4</td>
<td>00112345667789</td>
</tr>
<tr>
<td>5</td>
<td>1224</td>
</tr>
</tbody>
</table>

Describe the distribution of data, in context.
1. The following graph represents a density curve:

   a. Determine the total area under the curve
   b. Find \( P(x \leq 3) \).
   c. Find \( P(x \geq 4) \).
   d. Find \( P(2 \leq x \leq 5) \).
   e. Find \( P(x \leq 6) + P(x \geq 6) \)

2. Below is a table for the number of wins for teams in the N.F.L. after 3 weeks, where \( x \) represents the number of wins and \( y \) represents the number of teams with \( x \) wins. (ex. There are 7 teams with 1 win and 15 teams with 2 wins.)

<table>
<thead>
<tr>
<th># of wins (( x ))</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td># of teams (( y ))</td>
<td>2</td>
<td>7</td>
<td>15</td>
<td>8</td>
</tr>
</tbody>
</table>

   a. How many teams are in the league?
   b. What is the probability that a randomly selected team has one win? \( P(x = 1) \)
   c. What is the probability that a randomly selected team has three wins? \( P(x = 3) \)
   d. Find \( P(x \leq 2) \)
   e. Find \( P(x > 1) \)
3. A distribution has a mean of 12 and a standard deviation of 5.8. Four points are added to each score in the data set. What are the new values of mean and standard deviation?

4. A distribution has a median of 64 and an IQR of 25. Each score is multiplied by three. What are the new values of median and IQR?

5. Find the value of the median in the histogram below:

![Histogram](image)

6. A sample is obtained of 100 students at Ramapo High School in order to determine the mean number of hours worked at an after school job per week. Describe the population in this example.

WRITTEN RESPONSE – TO BE HANDED IN

7. According to the United States Golf Association, the ages of frequent golfers are shown in the cumulative frequency graph below:

![Cumulative Frequency Graph](image)

   a. In the context of this question, explain what information is conveyed by the circled point.
   b. Estimate the percentage of golfers between the ages of 59 and 79.

HW 2-2

1. Find the probability that a z-score is (nearest ten-thousandth):
   a. greater than 1.31
   b. between -2.59 and 1.31
   c. less than -0.32
2. Find:
   a. \( P(z < 1.30) \)
   b. \( P(z > -1.98) \)
   c. \( P(-1.28 < z < 1.76) \)

3. Robert and Elizabeth each have a different professor for chemistry. On a midterm exam, Robert scores an 84 while Elizabeth scores a 78. If the students in Robert’s class have a mean score of 72 with a standard deviation of 4, and the students in Elizabeth’s class have a mean score of 62 with a standard deviation of 5, who did better relative to the rest of their class?

4. Find the \( z \)-score that corresponds to a probability of:
   a. 0.0594
   b. 0.6026
   c. 0.4247
   d. 0.9147

5. What \( z \)-scores correspond to the middle 40% of a normal distribution?

6. Find the \( z \)-score such that 80% of the distribution is below this value.

WRITTEN RESPONSE – TO BE HANDED IN

7. Given the following summary statistics, determine if the data contains any outliers. Justify your answer.
   \( n = 25 \)
   \( \bar{x} = 27.3 \)
   \( median = 24.5 \)
   \( min = 16 \)
   \( Q1 = 20 \)
   \( Q3 = 34 \)
   \( max = 55 \)

HW 2-3

1. Assuming a normal distribution with a mean of 20 and a standard deviation of 5, what score associated with the 90th percentile? (round to the nearest tenth)

2. Given a normal distribution with a mean of 60 and a standard deviation of 12, find (round to the nearest ten-thousandth):
   a. \( P(x \geq 65) \)
   b. \( P(45 < x < 70) \)
3. Given a normal distribution with a mean of 80 and a standard deviation of 15, find:
   a. \( P(x < 90) \)
   b. \( P(55 < x < 75) \)

4. In a school of 380 freshmen, the scores on a new English test are approximately normally distributed with a mean of 75 and a standard deviation of 8. Find the approximate number of students who are expected to have a test score above 85.

5. Assume that college women’s heights follow a normal distribution with \( \mu = 65 \) inches and \( \sigma = 2.7 \) inches (round all answers to the nearest ten-thousandth).
   c. What proportion of college women are 62 inches or shorter?
   d. What proportion of college women are between 66 and 70 inches tall?
   e. To the nearest inch, what is the height of a woman in the 75th percentile?

6. A teacher gave a quiz where the mean raw score was 19.3 and the standard deviation was 3.8. When the teacher converted the scores to a 100 point scale, the new mean was 87.2 and the new standard deviation was 15.2. Explain the transformation that occurred for each score.

WRITTEN RESPONSE – TO BE HANDED IN

7. A local real estate company compiles data on the available homes in one particular town. The figure below displays the cumulative relative frequency plot of the homes available for sale (in hundreds of thousands of dollars).
a) In the context of this question, explain what information is conveyed by the circled point.
b) Approximately, what proportion of homes are listed for sale between $400,000 and $600,000?
c) For the values between 7 and 8 on the horizontal axis, the cumulative relative frequency plot is flat. In the context of the question, explain what this means.
d) The owner of the real estate agency wants to personally oversee the top 20 percent most expensive homes. What is the approximate minimum home price that would qualify to be overseen by the owner?

HW 2-4

1. The weights of packages shipped by a delivery company are measured and it is found that the mean weight is 5.8 pounds with a standard deviation of 4.2 pounds. Is there evidence that the population of packages from this company is normally distributed? Justify your answer.

2. Below are the scores (out of 30) for a class of Algebra students:
   25, 17, 23, 14, 20, 16, 20, 8, 12, 18, 19, 16, 13, 20, 19, 15, 14, 4, 25
   Is there evidence that these data are normally distributed? Explain.

3. Henry took a test in English and received a score of 88. The class mean was 80, with a standard deviation of 6.4. If his z-score on his Government test was the same as his English test, what was his grade on Government if the mean for Government was 83 with a standard deviation of 5.6?

4. For a normally distributed population with a mean of 75 and a standard deviation of 8, what is the best estimate of the range of the data given this limited information?

5. Here is a list of exam scores for Mr. Williams’s calculus class:
   60, 61, 61, 65, 72, 75, 75, 78, 81, 81, 85, 89, 91, 98
   What is the percentile of the person whose score was 85?
6. Consider the following cumulative relative frequency graph of the scores of students in an introductory statistics course:

![Cumulative relative frequency graph](image)

A grade of C or C+ is assigned to a student who scores between 55 and 70. Find the approximate percentage of students who obtained a grade of C or C+.

WRITTEN RESPONSE – TO BE HANDED IN

7. For the data: 15, 18, 19, 20, 23, 24, 24, 27, 29, 30, 32, 38, 41, 50, 62
   a. Create a box-and-whisker plot.
   b. Create a stem-and-leaf plot.
   c. Name one characteristic that can be seen in the stem plot that cannot be seen in the boxplot

HW 2-5

1. Students in the senior class of James Woods High School received their GPAs. The GPAs for the entire senior class were approximately normally distributed with a mean of 84.8 and a standard deviation of 6.4. Nick wants to know if he is in the top 10% of his class with a GPA of 93.4. Should Nick’s GPA put him in the top 10%?

2. The following data represents the ages of members of the New York Giants:
   Are these data approximately normally distributed? Justify your answer.
3. The Better Business Council of a large city has concluded that students in the city’s schools are not learning enough about economics to function in the modern world. These findings were based on test results from a random sample of 20 twelfth-grade students who completed a 46-question multiple-choice test on basic economic concepts. The data set below shows the number of questions that each of the 20 students in the sample answered correctly.
12, 16, 18, 17, 18, 33, 41, 44, 38, 35, 19, 36, 19, 13, 43, 8, 16, 14, 10, 9
Are these data approximately normally distributed? Justify your answer.

4. A certain state’s education commissioner released a new report card for all public schools in the state. This report card provides a new tool for comparing schools across the state. One of the key measures that can be computed from the report card is the student-to-teacher ratio, which is the number of students enrolled in a given school divided by the number of teachers at that school. The data below give the student-to-teacher ratio at the 10 schools with the highest proportion of students meeting the state reading standards in the third grade and at the 10 schools with the lowest proportion of students meeting the state reading standards in the third grade.
Ratio at schools with highest proportion: 7, 21, 18, 22, 9, 16, 12, 17, 17, 16
Ratio at schools with lowest proportion: 14, 16, 18, 20, 12, 14, 16, 12, 20, 19
Are the data in each data set approximately normally distributed? Justify your answer.

5. The dot plot below represents the number of children per class in a nursery school of twenty classes:

![Dot plot](image)

a) What is the value of the 30th percentile?
b) Which is greater, the mean or the median?

6. If 30 is added to every number on a list, the only one of the following that is not changed is
A. the mean.
B. the mode.
C. the 75th percentile.
D. the median.
E. the standard deviation.
WRITTEN RESPONSE – TO BE HANDED IN

7. Below are the Regents scores for students in a class. Create box-and-whisker plots of each variable (separately) on the same axes and write a few sentences comparing the two tests:

Integrated Algebra Regents: 74, 84, 91, 86, 93, 85, 81, 77, 82, 80, 82, 81, 83, 88, 84, 93

Algebra 2 & Trig Regents: 69, 64, 69, 80, 80, 91, 64, 98, 61, 91, 65, 48, 80, 64, 71, 80, 76, 81, 71, 79

HW 2-6

1. The following data represent the number of games won by the Patriots and the Jets over the last twelve seasons. Create a boxplot for each team on the same set of axes and write a few sentences comparing the two distributions.
Patriots: 12, 16, 11, 10, 14, 13, 12, 12, 12, 14, 13
Jets: 10, 4, 9, 9, 11, 8, 6, 8, 4, 10, 5, 5

![Boxplot](image)

2. Given a normally distributed population with a mean of 40 and a standard deviation of 6, use the formula for z-score, \( z = \frac{x - \mu}{\sigma} \), determine the score \( x \) that is associated with a z-score of 1.
3. Using the histogram above, create a box-and-whisker plot which would represent the same data.

4. A town’s January high temperatures average 36 degrees F with a standard deviation of 10 degrees, while in July the mean high temperature is 74 degrees with a standard deviation of 8 degrees. In which month is it more unusual to have a day with a high temperature of 55 degrees?

5. IQ tests follow an approximate normal distribution with a mean of 100 and a standard deviation of 15. Use this information to answer each of the following:
   a. What proportion of the population will have an IQ score greater than 135?
   b. What proportion of the population will have an IQ score between 105 and 125?
   c. What IQ score equates to the 70th percentile?
   d. Out of a random sample of 400 individuals, how many will have an IQ score greater than 140?

6. Companies that design furniture for elementary school classrooms produce a variety of sizes for kids of different ages. Suppose the heights of kindergarten children can be described by a normal model with a mean of 38.2 inches and a standard deviation of 1.8 inches.
   a. What proportion of kids should the company expect to be less than three feet tall?
   b. In what height interval should the company expect to find the middle 80% of kindergartners?
   c. At least how tall are the biggest 10% of kindergartners?
   d. In a school with 124 kindergartners, approximately how many would be taller than 41 inches?
7. A college professor polled 25 students and asked how many movies they have watched in the last year. The histogram below shows the distribution of the number of movies watched per student.

a. Write a few sentences to describe the distribution of the number of movies watched by the students.

b. One student had watched 22 movies. If the student had actually watched 32 movies, how would the change affect the mean and median? Justify your answers.
1. A random sample of 1,000 likely voters indicated their preference for the Republican candidate, the Democratic candidate with some voters undecided. The results are in the table below:

<table>
<thead>
<tr>
<th></th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Republican</td>
<td>237</td>
<td>205</td>
<td>442</td>
</tr>
<tr>
<td>Democrat</td>
<td>188</td>
<td>274</td>
<td>462</td>
</tr>
<tr>
<td>Undecided</td>
<td>58</td>
<td>38</td>
<td>96</td>
</tr>
<tr>
<td>Totals</td>
<td>483</td>
<td>517</td>
<td>1,000</td>
</tr>
</tbody>
</table>

a. Is it more likely for a randomly selected voter who is male intending to vote Republican or a female who intends to vote Democrat?
b. Which gender has a higher proportion of undecided voters?
c. Is it more likely for a randomly selected Republican voter to be a female or a Democrat voter to be male?

2. The students in a high school were asked to name their favorite subject among math, English, social studies and science. The results are in the table below:

<table>
<thead>
<tr>
<th></th>
<th>Math</th>
<th>Science</th>
<th>Social Studies</th>
<th>English</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshmen</td>
<td>83</td>
<td>137</td>
<td>96</td>
<td>102</td>
<td>418</td>
</tr>
<tr>
<td>Sophomores</td>
<td>114</td>
<td>91</td>
<td>78</td>
<td>95</td>
<td>378</td>
</tr>
<tr>
<td>Juniors</td>
<td>60</td>
<td>88</td>
<td>102</td>
<td>105</td>
<td>355</td>
</tr>
<tr>
<td>Seniors</td>
<td>117</td>
<td>98</td>
<td>84</td>
<td>50</td>
<td>349</td>
</tr>
<tr>
<td>Totals</td>
<td>374</td>
<td>414</td>
<td>360</td>
<td>352</td>
<td>1,500</td>
</tr>
</tbody>
</table>

a. Is it more likely for a freshman to prefer Math or social studies?
b. Which class has the highest proportion of students who prefer English?
c. Is it more likely for a randomly selected student to be a junior who prefers science or a senior who prefers social studies?
d. Is it more likely for a student who prefers science to be a sophomore or a student who prefers math to be a freshman?

d. What is the probability we select a Human Ecology student?
e. What is the probability that we select a first-born student?
f. What is the probability that the person is first-born and a Human Ecology student?
g. What is the probability that the person is first-born or a Human Ecology student?
4. Put the dot plots in order from smallest standard deviation to largest standard deviation (all three are on the same horizontal scale):

A)  

B)  

C)  

5. Estimate the value of the IQR from the graph below:

6. The following boxplots display the grades in a class for the year 2014 and 2015:

2014 Class

2015 Class

a. Which class had a greater median?
b. Which class had a greater range?
c. Which class had a greater IQR?
d. Which class had more observations?
e. Complete: 75% of the students in the 2015 class had higher grades than ____% of the 2014 class.

WRITTEN RESPONSE – TO BE HANDED IN

7. Create two histograms.

1999 #4
HW 3-2

1. The scatter plot below displays the relationship between daily high temperature (in degrees Celsius) and the amount of ice cream sold at a beach resort. Comment on the scatter plot.

2. Create a scatter plot based on the data below:

<table>
<thead>
<tr>
<th>x</th>
<th>2</th>
<th>4</th>
<th>5</th>
<th>7</th>
<th>8</th>
<th>10</th>
<th>12</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>25</td>
<td>23</td>
<td>21</td>
<td>17</td>
<td>13</td>
<td>11</td>
<td>9</td>
<td>8</td>
</tr>
</tbody>
</table>
3. The scatter plot below shows the relationship and the average number of hours of sleep.

![Scatter plot](image)

Describe the nature of the scatterplot, in context.
How much sleep, on average, does an 18 year old need?

4. Mr. Hale measures the heights of all boys in the sixth grade in his school. He finds that Russell’s height has a z-score of 1.23. Mr. Denver, the science teacher, decides that the heights should have been measured in meters instead of inches. When he converts each boy’s height to meters, what is the new z-score of Russell’s height?

5. A male California sea lion has a mean weight of 300kg with a standard deviation of 35kg; a female has a mean weight of 100kg. A male is chosen at random and found to weigh 363kg. If a randomly chosen female is the same relative weight as the male and weighs 142kg, what is the standard deviation for the female sea lion?

6. Kim wants to determine the grade she needs on a chemistry test to reach a certain percentile. Assume that the scores on the chemistry test are normally distributed with a mean of 82.1 and standard deviation of 5.8. She uses the equation

\[ .84 = \frac{x - 82.1}{5.8} \]

What percentile is she hoping to achieve and what grade will she need on the test?

WRITTEN RESPONSE – TO BE HANDED IN

7. 2001 #1
HW 3-3

1. What is the most likely value of the correlation coefficient for the data graphed in the scatter plot below?

   ![Scatter Plot Image]

   A) \( r = 1 \)  
   B) \( r = 0.8 \)  
   C) \( r = 0 \)  
   D) \( r = -0.8 \)  
   E) \( r = -1 \)

2. A group of randomly selected students in a P.E. class were measured on their performances on a physical fitness test for push-ups and sit-ups. The results are in the table below.

   | push-ups | 27  | 22  | 15  | 35  | 30  | 52  | 35  | 55  | 40  | 40  |
   | sit-ups  | 30  | 26  | 25  | 42  | 38  | 40  | 32  | 54  | 50  | 43  |

   a. Construct a scatter plot for these data.

   ![Scatter Plot Image]

   b. Describe the nature of the relationship between push-ups and sit-ups.
c. What is the value of the correlation coefficient, to the nearest thousandth, between push-ups and sit-ups?

d. What is the value of the correlation coefficient if we use sit-ups and the independent variable and push-ups as the dependent variable?

3. Mr. Mattingly found that the average number of absences for the students in his class so far this year was 2.24 days, with a standard deviation of 1.75 days. Does this evidence suggest that the number of days absent for Mr. Mattingly’s students is approximately normal? Justify your answer.

4. In Olympic bobsled, the maximum weight of a sled (including the driver and brake woman) in the two-woman event is 340 kilograms. A comparison is done of the weight of the two athletes and the weight of the empty sled. Which of the following would be the best approximation of the correlation coefficient between the two variables?
   
   (A) 1  (B) 0.9  (C) 0.5  (D) -0.5  (E) -0.9

5. A car dealership has been running one minute television ads on local TV and recording the number of cars sold during that week. After 10 weeks, the linear correlation between number of ads that week (x) and the number of cars sold (y) resulted in \( r = 0.920 \). Interpret the coefficient of determination in the context of the problem.

6. Boys in the second grade have a mean height of 50.3 inches. If a randomly selected boy has a height that is 1.62 standard deviations above the mean, then what percentage of second grade boys would be taller than the boy selected?
HW 3-4

1. The data below measures, for women aged 30-39, height in inches vs. weight in pounds.

<table>
<thead>
<tr>
<th>Height</th>
<th>58</th>
<th>59</th>
<th>60</th>
<th>61</th>
<th>62</th>
<th>63</th>
<th>64</th>
<th>65</th>
<th>66</th>
<th>67</th>
<th>68</th>
<th>69</th>
<th>70</th>
<th>71</th>
<th>72</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>113</td>
<td>115</td>
<td>118</td>
<td>121</td>
<td>124</td>
<td>128</td>
<td>131</td>
<td>134</td>
<td>137</td>
<td>141</td>
<td>145</td>
<td>150</td>
<td>153</td>
<td>159</td>
<td>164</td>
</tr>
</tbody>
</table>

a. Find the correlation coefficient and coefficient of determination. Explain the meaning of the coefficient of determination.
b. Find the equation of the line of best fit. Explain the meaning of the slope in context.
c. Predict the weight of a woman 58 inches tall.

2. The following scatterplot shows ratings of viewers of a new family movie on a scale of 1 to 10, 10 being best. Each film was rated by a child ($x$) and his or her parent ($y$).

Comment on the relationship between the child’s rating and the adult’s rating.

3. The runners on a men’s college track team were timed for the 60 meter dash and the results displayed in the boxplot below.
a. Comment on the boxplot above.
b. What is the value of the median?
c. Estimate a likely value for the mean.

4. A physics teacher gave a test in two parts. Part I was worth 60 points and part II was worth 40 points. On part I scores were normally distributed with a mean score of 34.3 with a standard deviation of 7.12. Part II scores were normally distributed with a mean of 18.3 and a standard deviation of 4.68.
   a. What is the mean of the combined score for both parts (out of 100)?
   b. What is the standard deviation of the combined score for both parts?
   c. If a student from the class is chosen at random, what is the probability that the student scored over 65 or greater?

5. A study was conducted to investigate the relationship between resale price (in thousands of dollars) and the age (in years) of midsize luxury automobiles. The equation of the line of best fit was \( y = 185.7 - 21.52x \).
   a. Identify and interpret the value of the slope.
   b. Identify and interpret the value of the intercept.

6. The city’s transportation department is interested in studying the relationship between the temperature and the number of passengers that ride the main bus line in order to better serve their customers. The manager recorded the temperature at the beginning of the hour, and then had a bus driver record the number of passengers that boarded the bus throughout the hour. Their findings are listed below. The equation of the line of best fit was \( y = 4.43 + 3.953x \).
   a. Identify and interpret the value of the slope.
   b. Identify and interpret the value of the intercept.

WRITTEN RESPONSE – TO BE HANDED IN
2006 #1
The May/June 1989 issue of *Public Health Reports* published an article titled “A Multistate Analysis of Active Life Expectancy”. The study compared current age and expected years of life remaining.

<table>
<thead>
<tr>
<th>Age</th>
<th>Expected Years Remaining</th>
<th>Predicted Years Remaining</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td>65</td>
<td>16.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>67</td>
<td>15.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>69</td>
<td>13.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>12.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>73</td>
<td>11.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>10.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>77</td>
<td>9.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>79</td>
<td>8.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>81</td>
<td>7.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>83</td>
<td>6.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1. a. Find the equation of the line of best fit, rounding both slope and intercept to the nearest thousandth.
   b. Interpret the value of the slope in the context of the problem.

2. a. Using the equation of the line of best fit, calculate the predicted years of life remaining for each age given.
   b. Calculate the residual for each age given.

3. Given the linear regression equation $\hat{y} = 14.9 + 0.66x$, find the value of the residual for the point (11,20).

4. The relationship between height and weight of a sample of college age women is given by the regression equation $\hat{y} = 4.17x - 186.5$ where $x$ represents the height in inches and $y$ represents the weight in pounds. Interpret the value of the slope in the context of this problem.

6. The five number summary of a students’ test scores is 44, 72, 78, 84, 100. Do these data include any outliers? Justify your answer.

HW 3-6

1. A sample of men agreed to participate in a study to determine the relationship between several variables including height, weight, waist size and percent body fat. A regression was performed for waist size (in inches) and percent body fat. The computer output is below:

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>se of coeff</th>
<th>t-ratio</th>
<th>prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-42.734</td>
<td>2.717</td>
<td>15.7</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Waist</td>
<td>1.700</td>
<td>0.0743</td>
<td>22.9</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

   a. Write the equation of the line of best fit, rounding all values to the nearest thousandth.
   b. What is the value of the correlation coefficient?
   c. Interpret the value of the slope in the context of this problem.

2. A study a random sample of service call records for a computer repair operation were examined and the length of each call (in minutes) and the number of components repaired or replaced were recorded. The computer output of the regression analysis is below:

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>F-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>274195.5</td>
<td>1</td>
<td>274195.5</td>
<td>943</td>
</tr>
<tr>
<td>Residual</td>
<td>346.040</td>
<td>12</td>
<td>29.0707</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>s.e. of Coeff</th>
<th>t-ratio</th>
<th>prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>4.16165</td>
<td>3.355</td>
<td>1.24</td>
<td>0.2385</td>
</tr>
<tr>
<td>Units</td>
<td>15.5088</td>
<td>0.585</td>
<td>38.7</td>
<td>&lt; 0.0001</td>
</tr>
</tbody>
</table>

   a. Write the equation of the line of best fit, rounding all values to the nearest thousandth.
b. What is the value of the correlation coefficient?
c. Interpret the value of the slope in the context of this problem.

3. The population of farmers in the United States by year is given in the table below:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Farmers</td>
<td>32.1</td>
<td>30.5</td>
<td>25.1</td>
<td>9.7</td>
<td>3.0</td>
<td>0.3</td>
</tr>
</tbody>
</table>

a. Find the equation of the line of best fit.
b. Create and sketch a residual plot for these data.
c. Comment on the meaning of the residual plot.

4. In the scatterplot below:

![Scatterplot]

Explain how the removal of the circled point would change the
a. Correlation coefficient
b. Slope
c. Intercept

5. Mr. Collins has three classes. The data for each class in summarized in the table below:

<table>
<thead>
<tr>
<th></th>
<th>Number of observations</th>
<th>Mean</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period 3</td>
<td>20</td>
<td>74.8</td>
<td>77</td>
</tr>
<tr>
<td>Period 4</td>
<td>25</td>
<td>71.6</td>
<td>78.5</td>
</tr>
<tr>
<td>Period 8</td>
<td>22</td>
<td>86.0</td>
<td>91</td>
</tr>
</tbody>
</table>

a. What is the mean of the students in all his classes?
b. What is the median of the students in all his classes?

6. Which histogram has the largest proportion of observations greater than the mean?
CHAPTER 4 – NON-LINEAR REGRESSION

1. The following data represents temperature readings, in degrees Fahrenheit, at different setting on a thermostat for an experimental cooling container.

<table>
<thead>
<tr>
<th>Setting</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp.</td>
<td>88</td>
<td>85</td>
<td>83</td>
<td>81</td>
<td>77</td>
<td>74</td>
<td>71</td>
<td>65</td>
<td>55</td>
<td>44</td>
<td>36</td>
<td>28</td>
<td>14</td>
<td>5</td>
</tr>
</tbody>
</table>

   a. Calculate the equation of the line of best fit.
   b. Sketch the residual plot.

2. Sketch a residual plot with 10 points that would indicate that the linear model is appropriate for the data.

3. Sketch a residual plot with 10 points that would indicate that the linear model is *NOT* appropriate for the data.

4. The population of a small town since 1950 is shown below:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Population (1000s)</td>
<td>50</td>
<td>67</td>
<td>91</td>
<td>122</td>
<td>165</td>
<td>222</td>
<td>300</td>
</tr>
</tbody>
</table>

   a. Perform the linear regression and use the residual plot to determine if the linear model is appropriate.
   b. Transform the data using an exponential model.
   c. Find the line of best fit based on the transformed data.
   d. From the residual plot of the transformed data, is this model appropriate?
   e. Use the transformed model to predict the population in 2014.
5. Run Differential Per Game

The scatterplot above represents the average run differential (runs score minus runs allowed) versus the number of games won for each Major League Baseball team in 2015. The equation of the line of best fit is \( \hat{y} = 81.019 + 15.622x \), where \( x \) represents the average run differential and \( y \) represents the number of games won. Draw the line of best fit in the graph above.

6. Which of the following boxplots is most clearly skewed to the right?

A) 

B)
1. The table below shows the temperature, in degrees Fahrenheit, of an instrument measured as its distance, in centimeters, from a heat source is varied.

<table>
<thead>
<tr>
<th>Distance, cm</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temp, F</td>
<td>130</td>
<td>105</td>
<td>95</td>
<td>87</td>
<td>83</td>
<td>80</td>
<td>78</td>
<td>77</td>
</tr>
</tbody>
</table>

a. Which variable is the response variable?
b. Find the equation of the line of best fit and the linear correlation coefficient?
c. Determine if the linear model is appropriate for these data. Support with statistical evidence.

2. Complete a regression analysis for the following age and income data as indicated.

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>20</th>
<th>25</th>
<th>30</th>
<th>35</th>
<th>40</th>
<th>45</th>
<th>50</th>
<th>55</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Income ($)</td>
<td>18500</td>
<td>23600</td>
<td>29800</td>
<td>38500</td>
<td>49000</td>
<td>64100</td>
<td>78500</td>
<td>102000</td>
<td>130800</td>
</tr>
</tbody>
</table>

Determine whether an exponential or power regression model would be more appropriate for these data.

3. Given the regression equation \( \ln y = 0.1759x + 3.909 \), find the value \( y \) when \( x = 6 \). Is this equation an exponential or a power model? What is the rate of growth and the initial amount?

4. Given the regression equation \( \ln y = 0.5439(\ln x) + 1.992 \), find the value \( y \) when \( x = 4 \). Is this equation an exponential or a power model?
5. Marcia’s doctor told her that the standardized score (z-score) for her cholesterol level, compared to the cholesterol level of other women her age, is 1.72. Interpret the value of her standardized score.

6. Mr. Brady gave a quiz (out of 100 points) where the average score was 88 with a standard deviation of 6.4. Do you believe these scores to be normally distributed? Justify your answer.

WRITTEN RESPONSE – TO BE HANDED IN
1999 #1

HW 4-3

1. The best male long jumpers for State College since 1973 have averaged a jump of 263.0 inches with a standard deviation of 14.0 inches. The best female long jumpers have averaged 201.2 inches with a standard deviation of 7.7 inches. This year Joey jumped 275 inches and his sister, Carla, jumped 207 inches. Both are State College students. Within their groups, which athlete had the more impressive performance? Explain briefly.

2. Measure the number of TVs per person \( x \) and the average life expectancy \( y \) for the world’s nations. There is a high positive correlation: nations with many TVs have higher life expectancies. Could we lengthen the lives of people in Rwanda by shipping them TVs? Explain.

3. People who use artificial sweeteners in place of sugar tend to be heavier than people who use real sugar. Does this mean that artificial sweeteners cause weight gain? Give a more plausible explanation for this association.

4. A study showed that women who work in the production of computer chips have a high number of miscarriages. The union claimed exposure to chemicals causes the miscarriages. Another possible explanation is that the workers spend their entire shift standing up. Can we conclude that exposure to chemicals causes more miscarriages?

5. A study once found that people with two cars live longer than people who own only one car. Owning three cars is even better and so on. There is substantial positive correlation between the number of cars \( x \) and the length of life \( y \). What lurking variables might explain the association?
6. The equation of the line of best fit for the data in the scatterplot below is 
\[ \hat{y} = -1.444x + 30.184 \]. Draw the line of best fit on the graph.
HW 4-4

1. Following World War II, the Allies studied the accuracy of bomber pilots. The study found that there was a large, positive correlation between the number of enemy planes the bomber encountered and the accuracy of bombers. Would you recommend that bombers be sent on missions when there are a large amount of enemy planes in the air?

2. A study was conducted comparing two variables, $x$ and $y$. A linear regression was performed and the correlation coefficient was found to be $r = 0.921$. The scatter plot and residual plot are shown below.

   Below is the scatter plot and residual plot for the relationship between $\ln x$ and $\ln y$ with a correlation coefficient of $r = 0.986$.

   Comment on the nature of the relationship between the two variables.

3. A data set has a variance of 5. The value of each observation in the data set is multiplied by four. What is the variance of the new data?
4. A group of second grade girls is found to have an average height of 51 inches. Tanya is 49 inches tall and is 1.32 standard deviations below the mean. What proportion of the population is taller than 49 inches?

5. Given the regression equation \( \hat{y} = 1.732(\ln x) + 3.842 \), find the value \( y \) when \( x = 8 \).

6. Data set A has a standard deviation of 5.4, data set B has a standard deviation of 2.8. What is the standard deviation of A-B?

**CHAPTER 5 – DATA COLLECTION**

1. Which of the following is true?
   I. In an experiment, some treatment is intentionally forced on one group to note the response.
   II. In an observational study information is gathered on an already existing situation.
   III. Sample surveys are observational studies, not experiments.
   (A) I and II
   (B) I and III
   (C) II and III
   (D) All of the above are true.
   (E) None of the above are true.

2. In one study on the effect of niacin on cholesterol level, 100 subjects who acknowledged being long-time niacin takers had their cholesterol levels compared with those of 100 people who had never taken niacin and 50 were chosen to receive a placebo.
   (A) The first study was a controlled experiment, while the second was an observational study.
   (B) The first study was an observational study, while the second was a controlled experiment.
   (C) Both studies were controlled experiments.
   (D) Both studies were observational studies.
   (E) Each study was part controlled experiment and part observational study.

3. In a 1992 London study, 12 out of 20 migraine sufferers were given chocolate whose flavor was masked by peppermint, while the remaining eight sufferers received a similar-looking, similar tasting tablet that had no chocolate. Within 1 day, five of those receiving chocolate complained of migraines, while no complaints were made by any of those who did not receive chocolate. Which of the following is a true statement?
The study was an observational study of 20 migraine sufferers in which it was noted how many came down with migraines after eating chocolate.

This study was a sample survey in which 12 out of 20 migraine sufferers were picked to receive the peppermint-flavored chocolate.

A census of 20 migraine sufferers was taking, noting how many were given chocolate and how many developed migraines.

A study was performed using chocolate as a placebo to study one cause of migraines.

An experiment was performed comparing a treatment group that was given chocolate to a control group that was not.

4. Suppose you wish to compare the average class size of mathematics classes to the average size of English classes in your high school. Which is the most appropriate technique for gathering the needed data?

(A) Census

(B) Sample survey

(C) Experiment

(D) Observational study

(E) None of these methods is appropriate.

5. Which of the following are true statements?

I. If bias is present in a sampling procedure, it can be overcome by dramatically increasing the sample size.

II. There is no such thing as a “bad sample.”

III. Sampling techniques that use probability techniques effectively eliminate bias.

(A) I only

(B) II only

(C) III only

(D) None are true.

(E) All are true.

6. A girl’s height is measured each year from age 2 to 11. The model for age in years \(x\) versus height in inches \(y\) results in the linear regression equation \(\hat{y} = 2.89x + 28.7\). If she is 47 inches tall at age 6, what is the value of the residual at age 6?

WRITTEN RESPONSE – TO BE HANDED IN

7. 2002 #4
HW 5-2

1. A bank surveyed all of its 60 employees to determine the proportion who participate in volunteer activities?
   a. Is this an experiment or an observational study?
   b. Since the employees were not randomly selected, can the bank use the results to draw a conclusion?

2. Explain a method that could reduce response bias?

3. The school principal wants investigate student opinion about the food served in the school cafeteria. She selects 50 freshmen, 50 sophomores, 50 juniors and 50 seniors. What type of sampling plan has been used?

4. A regression analysis was performed on the number of cell phone subscribers (in thousands) in the United States between 1990 and 1999. A linear regression was performed on the data and the line of best fit was found to be 
   \[ \hat{y} = 9083.349x - 18081823.7 \]
   The residual plot is below:

   a. Is the linear model appropriate for these data?
   b. A transformation was performed by taking the natural logarithm of the number of subscribers and a new regression equation was produced:
      \[ \ln y = 0.309x - 606.048 \]  . Using this model, what would have been the predicted value of cell phone users in 2001?
5. An educator wants to compare the effectiveness of computer software that teaches reading with that of a standard reading curriculum. He tests the reading ability of each student in his class for fourth graders, then divides them into two groups. One group uses the computer regularly, while the other studies a standard curriculum. At the end of the year, he retests all the students and compares the reading ability of the two groups?
   a. Is this an experiment? Justify your answer.
   b. What are the predictor and response variables?

6. What is specifically wrong with this experiment? A scientist wants to figure out how effective SPF 15 sunscreen is so he randomly selects 50 people of various skin types to apply the SPF 15 lotion and sit out in the sun for 4 hours. He then records the results.

WRITTEN RESPONSE – TO BE HANDED IN
7. 2005 #3
HW 5-4

1. An advertisement for an upcoming TV show asked: “Should handgun control be tougher? You call the shots in a special call-in poll tonight. If yes, call 1-900-720-6181. If no, call 1-900-720-6182. Charge is 50 cents for the first minute.” Over 90% of people who call in said “yes.” Explain why this opinion poll is almost certainly biased.

2. A person conducting interviews at the mall. Explain why a large sample of mall shoppers would not provide a trustworthy estimate of the unemployment rate.

3. On the west side of Rocky Mountain National Park, many mature pine trees are dying due to infestation by pine beetles. Scientists would like to use sampling to estimate the proportion of all pine trees in the area that haven’t been infested. Explain why it wouldn’t be practical for scientists to obtain a simple random sample in this setting. Provide an alternate sampling method.

4. A student is interested in estimating the typical amount of sleep high school students get per night. The student surveyed the first 100 students to arrive at school on a particular morning. These students reported an average of 7.2 hours of sleep on the previous night. What type of sample was used? Is this method biased? Explain.

5. Petal lengths of a particular flower are measured to the nearest tenth of a centimeter. Based on the histogram below, explain how to estimate the median and find that value.
6. Given the equation \( \hat{y} = 13.5x + 12.6 \), find the value of the residual for the point (10,30)

7. WRITTEN RESPONSE – TO BE HANDED IN
   2011 #5

HW 5-5

1. What is specifically wrong with this experiment? Gerber wants to compare their best sweet potato baby food to the best Beech-Nut sweet potato baby food they randomly select 1000 customers who had previously purchased Gerber products. They send them both Gerber and Beech Nut brand sweet potato baby food and then give them a rating scale on how their child responded to each brand.

2. What are the four principles of a statistical design of experiments?

3. A school committee member is lobbying for an increase in the gasoline tax to support the county school system. The local newspaper conducted a survey of county residents to assess their support for such an increase. What is the population of interest here?

4. A manufacturer of ready-bake cake mixes is interested in designing a study to test the effects of 4 different temperature levels (300, 325, 350 and 375 degrees F), 2 different types of pans (glass and metal), and 3 different types of ovens (gas, electric and microwave) on the texture of its cakes. How many different treatment combinations are to be used in this study?

5. Juan and Sonia want to play Monopoly, but cannot find any dice. However, Sonia has a random digits table from her AP Stats class. The decide to use the table to simulate the roll of a pair of dice. Describe a method of using random
digits to simulate the roll of two dice and starting with line 121, simulate the first ten rolls of the dice.

6. A regression equation is found comparing the temperature (in degrees Fahrenheit) and sales of soda (in dollars) at a refreshment stand in the park during summer. The equation is $\hat{y} = 52.48x - 3347.6$, $65 < x < 105$. By how much money would the sales of soda be expected to increase if the temperature increases by six degrees?

7. WRITTEN RESPONSE – TO BE HANDED IN
   2015 #5

---

HW 5-6

1. The ability to grow in shade may help pines found in the dry forests of Arizona to resist drought. How well do these pines grow in shade? Investigators planted pine seedlings in a greenhouse in either full light or light reduced to 5% of normal by shade cloth. At the end of the study, they dried the young trees and weighed them. Identify:
   a. The experimental units or subjects
   b. The factors
   c. The treatments
   d. The response variables

2. A large study used records from Canada’s national health care system to compare the effectiveness of two ways to treat prostate disease. The two treatments are traditional surgery and a new method that does not require surgery. The records described many patients whose doctors had chosen each method. The study found that patients treated by the new method were significantly more likely to die within 8 years.
   a. Further study showed that this conclusion was wrong. The extra deaths among patients who got the new method could be explained by lurking variables. What lurking variables might be confounded with a doctor’s choice of surgical or non-surgical treatment?
   b. You have 300 prostate patients who are willing to serve as subjects in an experiment to compare the two methods. Use a diagram to outline the
design of an experiment, be sure to indicate the size of the treatment
groups and the response variable.

3. A pharmaceutical company has designed two possible medications to treat a
disease. One medication is taken orally, the other is an injection. Explain how an
experiment to see which of the two medications is more effective could be
administered as a double-blind experiment.

4. Juan and Sonia want to play Monopoly, but cannot find any dice. However,
Sonia has a random digits table from her AP Stats class. They decide to use the
table to simulate the roll of a pair of dice. Describe a method of using random
digits to simulate the roll of two dice and starting with line 121, simulate the first
ten rolls of the dice.

5. A physical education teacher found there to be a relationship between the number
of sit-ups and the number of push-ups his students could complete for a physical
fitness test. He found the correlation coefficient to be $r=0.894$. Find the value of
the coefficient of determination and interpret the value of the coefficient of
determination.

TO BE HANDED IN

6. Ms. Morgenstern has a total of 83 Algebra students. She has 53 freshmen, 18
sophomores, 7 juniors and 5 seniors. She wants to randomly select four students
for a special pi day project. She wants to simulate the chances that all four of the
selected students are freshmen. Starting with line 81 on the random digits table,
fully explain a method of simulating this situation. Repeat the simulation five
times and determine how many of the five simulations result in a group of four
freshmen.
HW 5-7

1. In 1993, researchers proclaimed that listening to Mozart could make you smarter. Dubbed the Mozart effect, this conclusion was based on a study that showed college students temporarily gained up to 9 IQ points after listening to a Mozart piano sonata. This research has since been criticized by a number of researchers who have been unable to confirm the result in similar studies. Suppose that you wanted to see if there is a Mozart effect for students at DHS. Describe how you might design an experiment for this purpose. Use a diagram to support your answer.

2. A medical researcher is interested in testing a new medication for poison ivy. He decides to conduct a clinical trial on 250 volunteers who are allergic to poison ivy. He purposefully rubs poison ivy on their calf, then after the rash appears, he gives half of the volunteers calamine lotion, and the other half he gives his new medication. Draw a proper and complete design (diagram) for this experiment.

Researchers at U of Penn suggest that a nasal spray derived from pheromones (chemicals emitted by animals when they are trying to attract a mate) may be beneficial in relieving symptoms of PMS. Early tests indicate the spray, called PH80, eases irritability and also reduces some physical symptoms (Los Angeles Times, January 17, 2003)
3. Describe how you might design an experiment using 100 female volunteers who suffer from PMS to determine whether PH80 reduces PMS symptoms.

4. Does your design include a placebo treatment? Why or why not?


6. A soccer team has four games remaining. In each game they have a 54% chance of winning, 18% chance of a tie and 28% chance of a loss. Use the table of random digits, starting with line 126 to simulate the team’s next four games. Perform five simulations and determine how many times the team won three games.

TO BE HANDED IN

7. 1999#3

HW 5-5
HW 5-6

8. Twenty females between the ages of 40 and 55 volunteer for a study of a new cholesterol lowering medication. None of the women were on any cholesterol medication. Each woman had her cholesterol measured before the experiment began. The results, in milligrams per deciliter, were:

Annie 267
Beatrice 244
Cora 226
Deidre 243
Emily 262
Francesca 230
Glynis 251
Heidi 208
Ingrid 293
Jackie 268
Kim 253
Luisa 280
Marie 245
Natalie 299
Ophelia 229
Pippa 252
Roberta 213
Sally 212
Tilda 288
Ulrike 310

9. The grades in Mr. Lloyd’s science class are approximately normally distributed with a mean of 76.9 and a standard deviation of 8.63. Marty claims that his grade puts him in the 90th percentile. What was Marty’s grade?

10. The value of the coefficient of determination was $r = 0.642$. What does this information tell us about the correlation between the two variables?

---

**CHAPTER 6 - PROBABILITY**

6-1

1. Two contestants are on a quiz show, Lorie and Esmeralda. Each is in a soundproof booth and cannot hear what the other is saying. They are each asked to select a number from 1 to 3. They each win $10,000 if they select the same number. What is the probability they win the prize?

2. A family that has three children is randomly selected. What is the probability that 2 of the 3 children are boys?

3. Which of the following histograms has the largest standard deviation? Explain.

   ![Histograms A, B, and C]
4. Describe of the shape of the histogram:

5. A doctor measured the height Radames, of one of his patients, in inches, and compared the height to national averages. The patient had a z-score of -1.13. The new computer system caused the doctor to convert each patient’s height to centimeters. What is the current z-score of Radames’ height?

6. Given: $\mu_x = 33.7$, $\sigma_x = 4.8$, $\mu_y = 20.5$, $\sigma_y = 6.8$, find the mean and standard deviation of $4x - 5y$.

TO BE HANDED IN
7. 1998 #3

6-2

1. The probability that it will rain on Tuesday is 0.3. The probability that Edwin does not show up to work on Tuesday is 0.08. If Edwin’s attendance at work is independent from the weather, what is the probability that it rains and Edwin does not show up to work?

2. Mr. Medler gets off the highway on his commute to work and the probability that the first light is red is 0.42. The probability that the second light is red is 0.35. The probability that both lights are red is 0.231. Are the behavior of the two traffic lights independent? Justify your answer.

3. The current member of the United States Senate are distributed by party and gender below:

<table>
<thead>
<tr>
<th>Party</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democratic/Independent</td>
<td>32</td>
<td>16</td>
</tr>
<tr>
<td>Republican</td>
<td>47</td>
<td>5</td>
</tr>
</tbody>
</table>

Are gender and political party affiliation independent? Justify your answer.
4. The probability that a student will get a 3 or better on the AP Biology exam is 0.78; the probability that a student will get a 3 or better on the AP World exam is 0.56; the probability that the student will get a 3 or better on either the Biology or the World or both is 0.43.
   a. Find the probability that a randomly selected student will get at least 3 on both exams.
   b. Given the student got at least a 3 on World, what is the probability that he or she got at least a 3 on Biology?
   c. Are these two events independent? Explain.
   d. Are these two events mutually exclusive? Explain.

5. Illegal music downloading has become a big problem: 29% of internet users download music files, and 67% of downloaders say they don’t care if the music is copyrighted. What percent of internet users download music and don’t care if it is copyrighted? What assumption must you make in order to calculate the probability in this problem?

6. The weights of certain variety of eggplant are approximately normally distributed with a mean of 2.25 pounds and a standard deviation of 0.14 pounds. How heavy would an eggplant in the 75th percentile weight? How many standard deviations above the mean would this weight lie?

TO BE HANDED IN

7. 2000 #5

6-3

1. Smallville High School has 874 students, 236 freshmen, 224 sophomores, 212 juniors and 202 seniors. Six students are to be randomly selected to have a breakfast one day during school. Describe how a simulation would be performed using a random digits table, including a reference to when the simulation will end, and carry out the simulation beginning with line 120. Repeat six times and determine how many of the six simulations include three seniors.

2. Carla is playing a game using a pair of dice. The dice will be rolled ten times and the number of 7s recorded. Describe how the simulation will be performed, including a stopping rule, and use the table of random digits beginning at line 112, repeat the simulation six times.

3. The figure below represents a plot of land divided into four quadrants with each quadrant divided into four sections. Treat the four quadrants as blocks and randomly assign treatments to the sections within the blocks. Farmer Jones is trying a new type of fertilizer, so the two treatments are old fertilizer and new fertilizer. Describe how the random assignment will take place label each section.
4. A pharmaceutical company is developing a new cholesterol lowering drug and wants to compare the new drug to the drug currently on the market. A group of 24 women have agreed to participate in an experiment where 12 will receive the new drug and 12 will use the current drug. The experimenter wishes to conduct a matched pairs design. All women are ages 40-49 and in generally good health. The women are listed below with their cholesterol levels in milligrams per deciliter. Pair the women and describe how the drugs will be randomly assigned. Perform the random assignment and identify which women receive the new drug and which receive the current drug. If using the table of random digits, start with line 141.

| Ellen – 293  | Fran – 313  | Georgina – 369 | Helen – 379 |
| Rosa – 262   | Sally – 347   | Tanya – 242  | Ursula – 336 |

5. Given: \( \mu_x = 15.2 \), \( \sigma_x = 5.7 \), \( \mu_y = 36.3 \), \( \sigma_y = 17.6 \), find:
   a. \( \mu_{x+y} \)
   b. \( \sigma_{x+y} \)
   c. \( \mu_{12x+10y} \)
   d. \( \sigma_{12x+10y} \)

6. The probability that a teacher will be absent from school is 0.004. What is the probability that he will be absent at least twice in a 180-day school year?
1. Given \( P(A) = 0.4 \), \( P(B) = 0.5 \) and \( P(A \text{ and } B) = 0.15 \), find \( P(A|B) \)

2. Given \( P(C) = 0.72 \), \( P(D) = 0.68 \) and \( P(C \text{ or } D) = 0.86 \), find \( P(D|C) \)

3. The probability that Nickson goes to a party is 0.3, the probability that Shannon goes to the same party is 0.74, the probability that they both go to the party is 0.26. What is the probability that Nickson is at the party given that Shannon is at the party?

4. There is a 0.75 probability that you will go to a movie this weekend. There is a 0.32 probability that you will go out to eat this weekend. There is a 0.17 probability that you will do neither.
   a. What is the probability that you will go out to eat and see a movie?
   b. Given that you saw a movie, what is the probability that you went out to eat?
   c. Given that you went out to eat, what is the probability that you saw a movie?
5. Ramon has applied to both Princeton and Stanford. He thinks the probability that Princeton will admit him is 0.42, the probability that Stanford will admit him is 0.51, and the probability that both will admit him is 0.24.
   1. Make a diagram with the probabilities given marked.
   2. What is the probability that neither university admits Ramon?
   3. What is the probability that he gets into Stanford but not Princeton?
   4. What is the probability that he gets into Princeton, given that he got into Stanford?

6. Two students are trying to conduct a survey of the students in the school concerning their opinions on the new Common Core standards. They collect a sample of 60 students, but are not sure that the results unbiased. Sara proposes to increase the sample size to 120. Nicole proposes to randomly select 15 students from each grade. Which plan will be more successful in reducing bias and why?

TO BE HANDED IN
7. 2001 #3

6-5

1. Electrical connectors from three suppliers. The company prefers Acme Company because only 1% of those connectors prove to be defective, but Acme can deliver only 70% of the connectors needed. The company must also purchase connectors from two other suppliers, 20% from Burns Multinational and the rest from Cola Industries. The rates of defective connectors from Burns and Cola are 2% and 4% respectively. You buy one and find that it is defective.
   2. What is the probability that it came from Acme?
   3. What is the probability that it came from Burns?
   4. What is the probability that it came from Cola?

2. When the male students at RHS were asked, 50% said they do not date someone from RHS. When the female students were asked, 40% said they do not date someone from RHS. The male students make up 52% of the student population. Draw a tree diagram to represent this situation. What is the probability that a randomly selected:
   a. Student does not date someone from RHS?
   b. Student is female?
c. Student is female and does not date someone from RHS?
d. Student who dates someone from RHS is male?

3. A cancer clinic gives free cancer test. It is known that 2% of the people that come into the clinic have cancer. It is known the test comes up positive in 98% of people with cancer. It is known the test comes up positive in 3% of people without cancer. Create the tree diagram:
   a. What is the probability that someone tests positive given that they have cancer?
   b. What is the probability that someone tests negative given that they don’t have cancer?
   c. What is the probability that someone tests positive? Negative?
   d. What is the probability that someone has cancer given that they test positive? (This is called the accuracy of the test)
   e. What is the probability that someone doesn’t have cancer given that they test positive? (this is called a false positive)
   f. What is the probability that someone has cancer given that they test negative?

4. A recent Maryland highway safety study found that in 77% of all accidents the driver was wearing a seatbelt. Accident reports indicated that 92% of those drivers escaped serious injury (defined as hospitalization or death), but only 63% of the non-belted drivers were so fortunate. What’s the probability that a driver who was seriously injured wasn’t wearing a seatbelt?

5. You take a multiple choice test. On any given question, you are certain of the answer 78% of the time, unsure 14% of the time and have no idea on 8% of the questions. If you are sure, you get the correct answer. If you are unsure, you are correct 60% of the time and if you have no idea, you guess and are correct 25% of the time. Below is a tree diagram to illustrate these probabilities.
a. Calculate each of the end probabilities.
b. What is the probability that you got the question correct?
c. If you got the question correct, what was the probability that you were sure of the answer?
d. If you got the question correct, what was the probability that you had no idea?

6. Eric is playing a video game. He must complete two tasks out of three in order to advance to the next level. The probability of completing the first task is 0.8. If he completes the first task, the probability of completing the second task is 0.7; if he fails at the first task the probability of completing the second task is 0.2. If he completes the first two tasks, he advances. If he completes the first, but not the second, the probability of completing the third task is 0.3. If he completes the second, but not the first, the probability of completing the third task is 0.1. Create a tree diagram to answer the following questions:
a. What is the probability of advancing to the next level?
b. Given that he has advanced to the next level, what is the probability that he successfully completed the first task?
c. Given that he has advanced to the next level, what is the probability that he successfully completed the second task?

TO BE HANDED IN
7. 2001 #4

6-6

1. A plumbing contractor obtains 60% of her boiler circulators from company A whose defect rate is 0.005, and the rest from company B whose defect rate is 0.010. If a circulator is defective, what is the probability that it came from company A?

2. Two senior students both apply to the same four colleges. The probabilities of attending each school are below (assume the college one attends is independent of the other):

<table>
<thead>
<tr>
<th></th>
<th>State College</th>
<th>Ivy Univers.</th>
<th>County Community College</th>
<th>Capital University</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anika</td>
<td>0.15</td>
<td>0.05</td>
<td>0.35</td>
<td>0.45</td>
</tr>
<tr>
<td>Karlisha</td>
<td>0.35</td>
<td>0.39</td>
<td>0.01</td>
<td>0.25</td>
</tr>
</tbody>
</table>

a. Create either a tree diagram or a contingency table to illustrate the example.
b. What is the probability that they both will attend Capital University?
c. What is the probability that they both will attend the same college?
d. What is the probability that they will attend different colleges?

3. Solve the following example by creating a contingency table. The probability that Joe will attend Saturday’s party is $\frac{3}{4}$, the probability that Angelique will attend the party is $\frac{2}{3}$, the probability that neither of the will attend the party is $\frac{1}{4}$. Find:
   a. The probability that they will both attend the party.
   b. The probability that Joe will attend the party given that Angelique attends.
   c. The probability that Angelique will attend the party given that Joe attends.

4. A soccer team has a chance to win the league championship. If they win their last game, they clinch the league title. They have a 0.7 probability of winning their last game, a 0.2 chance of losing and a 0.1 chance of a tie. If they lose the game, they must play a playoff game and if they tie, the league championship will be decided by shoot-out. The team has a 0.6 probability of winning the playoff game, a 0.25 chance of losing and a 0.15 chance of a tie. In either game their probability of winning a shoot-out is 0.5.
   a. What is the probability that the team will win the league title?
   b. If they win the league title, what is the probability that they won their last regular game?

5. A bridal magazine wants information regarding college students’ attitudes about wedding costs. Students from a large college campus are randomly selected to take part in a survey. Students who respond yes to the question, “Do you plan to marry someday?” are then asked how much they plan to spend on their wedding. The results of the survey are summarized in the table below:

<table>
<thead>
<tr>
<th></th>
<th>Cost &lt; $10,000</th>
<th>Cost between $10,000 and $20,000</th>
<th>Cost &gt;$20,000</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>147</td>
<td>23</td>
<td>5</td>
<td>175</td>
</tr>
<tr>
<td>Female</td>
<td>5</td>
<td>120</td>
<td>72</td>
<td>197</td>
</tr>
<tr>
<td></td>
<td>152</td>
<td>143</td>
<td>77</td>
<td>372</td>
</tr>
</tbody>
</table>

   a) What is the probability that the cost is less than $10,000 given the respondent is female?
   b) What is the probability that the respondent is male given that the planned cost is between $10,000 and $20,000?
   c) What is the probability that the respondent is female given that the planned cost is over $20,000?
6. Which of the following residual plots provides the best evidence that the linear model is appropriate for the data?

![Residual Plots]

TO BE HANDED IN
7. 2002 #2

7-1

1. Complete the following probability function:

<table>
<thead>
<tr>
<th>$x$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(x)$</td>
<td>0.3</td>
<td>0.2</td>
<td>0.25</td>
<td></td>
<td>0.05</td>
</tr>
</tbody>
</table>

2. Complete the following probability function:

$$P(x) = \frac{x^2 + 3}{70}, \text{ for } x = 1, 2, 3, 4, 5.$$ 

3. The advisor of the Key Club in a school gives a handout to students in randomly selected homerooms to gather information about a possible school dance that would be used as a fundraiser. The teacher distributes 400 surveys to 15 randomly selected homerooms. Three of the teachers never hand out the surveys and among the other twelve homerooms, only 127 total responses are collected.
   a. What type of bias is the major concern about the responses collected and why?
   b. What is the population that the Key Club advisor is trying to generalize to?

4. The number of students in each class in an elementary school is shown below. Describe the distribution.
5. Due to a new rotating schedule, the probability that Mr. O’Connor’s AP World History class meets on any given day is 0.8. On days in which that class meets, the probability that he will be absent is 0.01; on days in which the AP World class does not meet, the probability that he will be absent is 0.05.
   a. Find the probability that Mr. O’Connor is absent on any given day.
   b. Given that Mr. O’Connor is absent, find the probability that his AP World class meets.

6. On a given Sunday, the probability that the Giants win is 0.15 and the probability that the Jets win is 0.35. The probability that both teams win is 0.05.
   a. Find the probability that neither the Giants not the Jets win.
   b. Given that the Giants lost, what is the probability that the Jets won.

7-2

1. Find the expected value:

<table>
<thead>
<tr>
<th>X</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>85</th>
<th>90</th>
<th>95</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(x)</td>
<td>0.14</td>
<td>0.03</td>
<td>0.11</td>
<td>0.09</td>
<td>0.17</td>
<td>0.21</td>
<td>0.17</td>
<td>0.08</td>
</tr>
</tbody>
</table>

2. Find the expected value of a game with the following payouts and probabilities:

<table>
<thead>
<tr>
<th>Payout</th>
<th>$0</th>
<th>$1</th>
<th>$5</th>
<th>$10</th>
<th>$100</th>
<th>$500</th>
<th>$1,000</th>
<th>$10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(x)</td>
<td>0.5</td>
<td>0.17</td>
<td>0.12</td>
<td>0.08</td>
<td>0.06</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
</tr>
</tbody>
</table>

3. The probability that a randomly selected athlete is using performance enhancing drugs is 0.08. A new drug test will cause an athlete to test positive with a probability of .75 and the same test will cause an athlete who has not used drugs to test positive with a probability of 0.03. If the first test is positive, then a second test must be administered. If the second test is positive, then the athlete is suspended. Construct a tree diagram and find:
   a. Find the probability of being suspended.
   b. Find the probability of not being positive.
   c. Find the probability that given an athlete was suspended, they used drugs.
d. Find the probability that given an athlete was not suspended, they did not use drugs.

4. For the following probability distribution function:

<table>
<thead>
<tr>
<th>$x$</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(x)$</td>
<td>.12</td>
<td>.19</td>
<td>.3</td>
<td>$a$</td>
<td>.14</td>
</tr>
</tbody>
</table>

a. Find the value of $a$ which makes the table a valid pdf.

b. Find $\mu_x$

c. Find $\sigma_x$

5. A student takes a twenty-question multiple choice test where each question has five choices. Assume that there is a one in five chance of obtaining the correct answer by guessing. On ten of the questions, she is 90% sure that she answered the question correctly (the other 10% could be right by guessing). On six of the questions, she is 50% sure that she was correct; and on four of the questions, she just guessed. Find the expected value of her grade out of 100 points. [Hint: you will have to compute three separate expected values].

6. The body weight of 8 year old boys is normally distributed with a mean of 56.5 pounds. A randomly selected 65 pound boy is 1.57 standard deviations above the mean. Approximately what percent of 8 year old boys weight more than 65 pounds?

TO BE HANDED IN

7. 2004 #2

7-3

1. The following table shows the relative frequency distribution, $X$, of children (under the age of 18) living in homes in Rockland County. For instance, 16% of households have two children:

<table>
<thead>
<tr>
<th>Number of children</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative frequency</td>
<td>0.35</td>
<td>0.18</td>
<td>0.16</td>
<td>0.11</td>
<td>0.09</td>
<td>0.05</td>
<td>0.03</td>
<td>0.02</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Find:

a. Mean

b. Standard deviation

$P(X \leq 3)$

c. $P(X \geq 6)$

d. Median.

2. The distribution of the number of adults per household in a town are given as:

<table>
<thead>
<tr>
<th>Number of adults $(x)$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(x)$</td>
<td>0.46</td>
<td>0.38</td>
<td>0.09</td>
<td>0.07</td>
</tr>
</tbody>
</table>
The distribution of the number of children (under age 18) per household is given as:

<table>
<thead>
<tr>
<th>Number of children (y)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P(y)$</td>
<td>0.12</td>
<td>0.37</td>
<td>0.22</td>
<td>0.13</td>
<td>0.10</td>
<td>0.06</td>
</tr>
</tbody>
</table>

a. Find the mean and standard deviation of both $x$ and $y$.
b. Find the mean and standard deviation of the number of people (adults plus children) in each household.

3. The distribution of the heights of 12 year old girls is approximately normal with a standard deviation of 1.8 inches. How would the height of a girl in the $83^{rd}$ percentile compare with the mean?

4. The probability that a student will get a 3 or better on the AP Chemistry exam is 0.58; the probability that a student will get a 3 or better on the AP US History exam is 0.37; the probability that the student will get a 3 or better on either the Biology or the World or both is 0.72.
   e. Find the probability that a randomly selected student will get at least on 3 on both exams.
   f. Given the student got at least a 3 on US, what is the probability that he or she got at least a 3 on Chem?
   g. Are these two events independent? Explain.
   h. Are these two events mutually exclusive? Explain.

5. A farm produces chicken eggs with a mean weight of 60 grams and a standard deviation of 3 grams. A carton includes one dozen eggs. The carton itself has a mean weight of 10 grams with a standard deviation of 2 grams. If the weights of the eggs are independent of each other, what is the mean and the standard deviation of the dozen eggs, including the carton.

6. The heights of second grade boys are approximately normal with a mean of 4.2 feet and a standard deviation of 0.3 feet. What would be the mean and standard deviation of the heights, measured in inches?
1. A large ferry can accommodate cars and buses. The toll for cars is $3 and the toll for buses is $10. Let X and Y denote the number of cars and buses respectively, carried on a single trip. Cars and buses are accommodated on separate levels of the ferry, so the X and Y are independent of each other.

<table>
<thead>
<tr>
<th>X</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(X)</td>
<td>.05</td>
<td>.10</td>
<td>.25</td>
<td>.30</td>
<td>.20</td>
<td>.10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Y</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(Y)</td>
<td>.50</td>
<td>.30</td>
<td>.20</td>
</tr>
</tbody>
</table>

a) Compute the mean and st. dev. of X
b) Compute the mean and st. dev. of Y
c) Compute the mean and st. dev. of the total amount of money collected in tolls from cars.
d) Compute the mean and st. dev. of the total amount of money collected in tolls from buses.
2. Given \( P(x) = \frac{x}{10} \), for \( x=1,2,3,4 \) and \( P(y) = \frac{y^2 + 1}{60} \)
   
   a) Complete a probability distribution table for each.
   
   b) Find the mean and standard deviation of \( x \).
   
   c) Find the mean and standard deviation of \( y \).
   
   d) Find the mean and standard deviation of \( x+y \).
   
   e) Find the mean and standard deviation of \( x-y \).

3. Given the following tables

<table>
<thead>
<tr>
<th>A</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(A)</td>
<td>0.4</td>
<td>0.5</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(A)</td>
<td>0.3</td>
<td>0.4</td>
<td>0.1</td>
<td>0.2</td>
</tr>
</tbody>
</table>

Complete the table below by multiplying the probabilities, where the random variable is \( A+B \)

<table>
<thead>
<tr>
<th></th>
<th>( P(A=1) = 0.4 )</th>
<th>( P(A=2) = 0.5 )</th>
<th>( P(A=3) = 0.1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( P(B=1) = 0.3 )</td>
<td>( P(A+B=2) = )</td>
<td>( P(A+B=3) = )</td>
<td>( P(A+B=4) = )</td>
</tr>
<tr>
<td>( P(B=2) = 0.4 )</td>
<td>( P(A+B=3) = )</td>
<td>( P(A+B=4) = )</td>
<td>( P(A+B=5) = )</td>
</tr>
<tr>
<td>( P(B=3) = 0.1 )</td>
<td>( P(A+B=4) = )</td>
<td>( P(A+B=5) = )</td>
<td>( P(A+B=6) = )</td>
</tr>
<tr>
<td>( P(B=4) = 0.2 )</td>
<td>( P(A+B=5) = )</td>
<td>( P(A+B=6) = )</td>
<td>( P(A+B=7) = )</td>
</tr>
</tbody>
</table>

Complete the table below for \( A+B \)

<table>
<thead>
<tr>
<th>A+B</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(A+B)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Calculate the mean and standard deviation for the random variable \( A+B \).

4. An automobile manufacturer buys computer chips from a supplier. The supplier sends a shipment containing 5% defective chips. Each chip chosen from this shipment has a 0.05 probability of being defective and each automobile uses 12 chips selected independently. What is the probability that all 12 chips in a car will work properly?

5. A string of Christmas lights contains 20 lights. The lights are wired in series, so that if any light fails, the whole string will go dark. Each light has probability of 0.02 of failing during a 3-year period. The lights fail independently of each other. What is the probability that the string of lights will remain bright for 3 years?
6. In a certain city 6% of teenagers are married, 25% of married teenagers have children, and 15% of unmarried teenagers have children. If a teenager has a child, what is the probability that the teenager is not married?

7. A plumbing contractor obtains 60% of her boiler circulators from company A whose defect rate is 0.005, and the rest from company B whose defect rate is 0.010. If a circulator is defective, what is the probability that it came from company A?

---

**CHAPTER 8 – BINOMIAL DISTRIBUTION**

1. Given a binomial random variable with a probability of .35. What is the probability of exactly 11 out of 16 successes?

2. The boy’s basketball team at some fictional, hypothetical high school has a probability of winning each game of 0.22. Find the probability that in a 19 game season, the team wins no more than 4 games.

3. A student is taking a midterm exam with 20 multiple choice questions, each with four choices. They have no idea of any of the answers, so they guess on each question. Find the probability that the student gets at least 10 questions correct.

1. Which of the following is a valid binomial probability?
2. An NFL team has a probability of winning each game in their 16 game season of 0.6. What is the probability that they will win less than 9 games?

3. The Student Advisory Council questioned each student in the school as to whether or not they planned to go to the homecoming dance. The results are as follows:

<table>
<thead>
<tr>
<th>Class</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman</td>
<td>63</td>
<td>337</td>
</tr>
<tr>
<td>Sophomore</td>
<td>90</td>
<td>270</td>
</tr>
<tr>
<td>Junior</td>
<td>105</td>
<td>195</td>
</tr>
<tr>
<td>Senior</td>
<td>108</td>
<td>132</td>
</tr>
</tbody>
</table>

Create an appropriate graphical display for these data.

TO BE HANDED IN

4. 1999 #5

8-2

1. Given \( B(20,.7) \), find the value of \( x \) for which \( P(X = x) \) is the greatest and the least.

2. The probability that a randomly selected student was born in December is 0.083. In a class of 24 students, what is the most likely number of students born in December and what is the least likely number of students born in December?

3. The probability that a randomly selected present from the holiday grab bag is candy is 0.3. Write the binomial equation (do not evaluate) with correct substitutions to determine the probability that exactly 3 out of 4 presents from the grab bag will be candy.
4. A fair coin is flipped ten times, each time the result is “tails.” What is the probability of “tails” on the next flip (11th) of the coin?

5. The histogram below shows the body mass of a group of individuals:

![Histogram of Body Mass](image)

Describe the distribution.

6. Events A and B are independent, with P(A)=0.7 and P(A and B)=0.28. Find:
   a. P(B)
   b. P(A or B)

TO BE HANDED IN

7. 1999 #1

---

8-3

1. A certain long jumper will have distances which are approximately normally distributed with a mean of 6.72 meters and a standard deviation of 0.12 meters.
   a. Find the probability that a random jump will be longer than 6.80 meters.
   b. Find the probability that at least 4 out of 6 jumps will be longer than 6.80 meters.

2. Which of the following is true?
   I. The histogram of a binomial distribution with $p=0.5$ is always symmetric no matter what $n$, the number of trials is.
   II. The histogram of a binomial distribution with $p=0.9$ is skewed to the right.
   III. The histogram of a binomial distribution with $p=0.9$ is almost symmetric if $n$ is very large.
3. Bags of Doritos have weights which are normally distributed with a mean of 15 oz. and a standard deviation of 0.21 oz.
   a. What is the probability that a randomly selected bag of Doritos will weigh more than 15.25 oz?
   b. Dan buys 12 bags of Doritos for a party. What is the probability that at least three of the bags will weigh more than 15.25 oz?

4. Suppose that 60% of students who take the AP Statistics exam score 4 or 5, 25% score 3, and the rest score 1 or 2. Suppose further that 95% of those scoring 4 or 5 receive college credit, 50% of those scoring 3 receive such credit, and 4% of those scoring 1 or 2 receive credit. If a student who is chosen at random from among those taking the exam receives college credit, what is the probability that she received a 3 on the exam?

5. A plumbing contractor obtains 60% of her boiler circulators from a company whose defect rate is 0.005, and the rest from a company whose defect rate is 0.010.
   c. What is the proportion of the circulators can be expected to be defective?
   d. If a circulator is defective, what is the probability that it came from the first company?

6. The box plot below displays the results of a quiz for Mr. Bogart’s 80 students.
   Grades of Quiz 3
   
   a. How many students scored better than 85?
   b. How many students scored between 65 and 85?
   c. Did more students score less than 65 or greater than 85?

7. 2000 #3
1. The population \{3,5,8,11\} has a mean \( \mu = 6.75 \) and a standard deviation \( \sigma = 3.031 \). When sampling with replacement, there are 16 different possible ordered samples of size 2 than can be selected from the population. The mean of each of these 16 samples is computed. For example, 1 of the 16 samples is (3,8), which has a mean of 5.5. The distribution of the 16 samples mean has its own mean \( \mu_x \) and its own standard deviation \( \sigma_x \).

a. Is the mean, \( \mu_x \), less than, equal to or greater than \( \mu = 6.75 \)

b. Is the standard deviation, \( \sigma_x \), less than, equal to or greater than \( \sigma = 3.031 \)
2. Let the random variable $X$ represent the number of children in elementary school for families in a certain school district. The probability distribution of $X$ is shown in the table below.

<table>
<thead>
<tr>
<th>$x$</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p(x)$</td>
<td>0.38</td>
<td>0.27</td>
<td>0.19</td>
<td>0.09</td>
<td>0.07</td>
</tr>
</tbody>
</table>

a. Calculate the expected value of $X$.

b. Using prior records of families in the district, a random sample of 20 families found an average of 1.55 elementary school children per family. An assistant superintendent wants to select another random sample of 100 families and compute the average number of elementary age children. How do you expect the average from the new sample to compare to that of the first sample? Justify your response.

c. The median of a random variable is defined as any value of $x$ such that $P(X \leq x) \geq 0.5$ and $P(X \geq x) \geq 0.5$. For the probability distribution shown in the table above, determine the median of $X$.

d. In a sentence or two, comment on the relationship between the mean and the median relative to the shape of this distribution.

3. In the following table, what value for $n$ results in a table showing perfect independence?

<table>
<thead>
<tr>
<th>40</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>$n$</td>
</tr>
</tbody>
</table>

4. Given the probabilities $P(A)=0.3$ and $P(B)=0.2$, what is the probability of $P(A \cup B)$ if:
   a. A and B are mutually exclusive.
   b. A and B are independent.
   c. If B is a subset of A.

5. Grades in an English class are normally distributed with a mean of 75 and a standard deviation of 8.2.
   a) Find the probability that a randomly selected student will score at least 85.
   b) In a class of 24 students, find the expected number of students that will score greater than 85.
   c) In a class of 24 students, find the probability that at least 5 will score greater than 85.

6. A pollster took a sample of 200 men, aged 18-45, living in Rockland County regarding their views on home ownership. What is the population to which the results of the survey can be generalized?
1. In a random sample of 60 adult males, the standard deviation of the sample probability distribution is $\sigma_p = 0.0632$. What sample size should be chosen if we want the standard deviation to be half of the current standard deviation.

2. Which of the following is true regarding the variation of a sampling distribution of a sample proportion?
   (A) Variation depends on the population size as well as sample size.
   (B) The variance of a sampling distribution of a sample proportion for all samples of size 1 is 0.
   (C) As the size of the sample increases, the variation of the sampling distribution approaches the variation of the population.
(D) For a given sample size, the maximum variation in the sampling distribution of a sample proportion occurs when the sample proportion is 0.5.

(E) None of these is true.

3. The weights of female Jersey cattle follow an unknown distribution with a mean of 450kg and a standard deviation of 35kg. Random samples of 20 cows at a time are taken. Describe the sampling distribution.

4. A sampling distribution with \( n=20 \) has a standard deviation of 1.042. What is the standard deviation of the population distribution? (round to the nearest hundredth).

5. The following dotplot displays the number of AP classes taken by students applying to a certain university.

![Dotplot](image)

Estimate the value of the 85\textsuperscript{th} percentile.

6. In Mr. Stewart’s class, his students fall into the following groups:
   - 15% are freshmen and 80% of freshmen will pass the class.
   - 20% are sophomores and 90% of sophomores will pass the class.
   - 30% are juniors and 85% of juniors will pass the class.
   - 35% are seniors and 95% of seniors will pass the class.

   Based on Mr. Stewart’s information, what is the probability that a student is an underclassman (not a senior) and will pass the class?

TO BE HANDED IN
7. 2001 #2
1. In a symmetric distribution with a mean of 12 and a standard deviation of 5, describe the sampling distribution of $\bar{x}$ when the sample size is 25.

2. Suppose that tomatoes weight an average of 10 ounces with a standard deviation of 3 ounces and a store sells boxes containing 12 tomatoes each. If customers determine the average weight of the tomatoes in each box they buy, what will be the mean and standard deviation of these averages?

3. Suppose that the distribution for total amounts spent by students vacationing for a week in Florida is normally distributed with a mean of $650 and a standard
deviation of $120. What is the probability that an SRS of 10 students will spend an average of between $600 and $700?

4. Suppose that the average outstanding credit card balance for young couples is $650 with a standard deviation of $20. In an SRS of 100 couples, what is the probability that the mean outstanding credit balance exceeds $700?

5. The strength of paper coming from a manufacturing plant is known to be 25 pounds per square inch with a standard deviation of 2.3. In a simple random sample of 40 pieces of paper, what is the probability that the mean strength is between 24.5 and 25.5 pounds per square inch?

6. The average outstanding bill for delinquent customer accounts for a national department store chain is $187.50 with a standard deviation of $54.50. In a simple random sample of 50 delinquent accounts, what is the probability that the mean outstanding bill is over $200?

TO BE HANDED IN
7. 2002 #3

CHAPTER 10 – INFERENCE

1. Find the value of $z$ necessary to construct a 95% confidence interval.

2. Find the value of $z$ necessary to construct a 99% confidence interval.

3. A bottling machine is operating with a standard deviation on 0.12 ounces. Suppose that in an SRS of 36 bottles the machine inserted an average of 16.1 ounces into each bottle. Construct and interpret a 95% confidence interval for the mean amount of ounces in the bottles.
4. If 64% of an SRS on 550 people leaving a shopping mall claim to have spent over $25, construct and interpret a 99% confidence interval to estimate the true proportion of shopping mall customers who spend over $25.

5. A service station offers state vehicle inspections. Let the random variable $X$ represent the number of inspections that the station will perform on a randomly selected day. The table below shows the probability distribution of $X$.

<table>
<thead>
<tr>
<th>Number of inspections ($X$)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
<td>0.2</td>
<td>0.15</td>
<td>0.25</td>
<td>0.3</td>
<td>0.1</td>
</tr>
</tbody>
</table>

a) What is the probability that the station will perform at least one inspection?
b) What is the expected value of the number of inspections that the station will perform?
c) What is the probability that the station will perform three inspections, given that at least one inspection was performed?
d) Given that the station performed at least one inspection, would the expected value of the number of inspections performed be less than, equal to or greater than the value found in part (b)? Explain.

6. A new poll concerning the Democratic primary in Iowa is to be conducted. The polling organization sends out 1,500 emails to registered Democratic voters in Iowa.

a) What is the population of interest?
b) Discuss a concern about bias in this situation.

TO BE HANDED IN

7. The owner of a cellphone store has in his inventory 50 cases of iPhones, each with 20 new phones in the case. Describe how a cluster sample would be conducted to obtain a sample size of 80 to estimate the battery strength of the iPhones. Clearly explain how to obtain the sample of size 80.

8. 2003 #3

10-2

1. At a certain plant, batteries are being produced with a life expectancy that has a variance of 5.76 months squared. Suppose the mean life expectancy in an SRS of 64 batteries is 12.35 months. Construct and interpret a 90% confidence interval estimate of life expectancy for all batteries produced at this plant.

2. A new drug results in lowering the heart rate by varying amounts with a standard deviation of 2.49 beats per minute. Construct and interpret a 95% confidence interval estimate for the mean lowering of the heart rate in all patients if a 50-person SRS averages a drop of 5.32 beats per minute.
3. A certain adjustment to a machine will change the length of the parts it is marking but will not affect the standard deviation. The length of the parts is normally distributed and the standard deviation is 0.5 mm. After an adjustment is made, a random sample is taken to determine the mean length of parts now being produced. The resulting lengths are:

| 75.3 | 76.0 | 75.0 | 77.0 | 75.4 | 76.3 | 77.0 | 74.9 | 76.5 | 75.8 |

a) What is the parameter of interest?
b) Find the point estimate for the mean length of all parts now being produced.
c) Find the 99% confidence interval for \( \mu \).

4. A sample of 60 night school students’ ages at a particular college is obtained in order to estimate the mean age of night school students, \( \bar{x} = 25.3 \) years. The population variance is 16.
a) State and verify the assumptions necessary to construct a confidence interval.
b) What is the point estimate for \( \mu \)?
c) Construct and interpret a 95% confidence interval.

5. According to Census data, 12.3% of the United States population is African American. If 20 Americans are selected at random:
a) What is the probability that none of the 20 people in the sample will be African-American?
b) What is the probability that two consecutive samples of 20 people will not include any African-Americans?

6. People in a remote location are contracting a certain disease. A new test to determine if someone has the disease has been developed. If a person has the disease, he or she will test positive 92% of the time. If a person does not have the disease, he or she will test positive 4% of the time. Suppose that 7% of the population has the disease.
a. Create a tree diagram to assign probabilities.
b. Calculate the probability that a randomly selected person will test positive.
c. Given that the person has tested positive, what is the probability that he or she has the disease?

TO BE HANDED IN

7. Mr. Hamilton gave the same midterm exam to two of his classes, the data is shown in the histogram below.
Describe the distribution of midterm test grades.

8. 2003 #4
1. A paint company wants to add a new accelerator to decrease the drying time of latex paint. Several test samples were conducted with the following decrease in drying time:

$$5.2 \ 6.4 \ 3.8 \ 6.3 \ 4.1 \ 2.8 \ 3.2 \ 4.7$$

Construct and interpret a 98% confidence interval for these data. Assume $\sigma = 1.4$ (be sure to: check assumptions, show calculations and interpret the interval in context).

2. The Channel Tunnel train that connect England with France carries up to 650 passengers, and peak speeds of over 190 mph are occasionally obtained. Assume then standard deviation is 19 mph in the course of all journeys back and forth and that the train’s speed is approximately normally distributed. Over the next 20 trips, the mean speed is 184 mph. Construct and interpret a 90% confidence interval for the true mean train speed.

3. A bank randomly selected 250 check-account customers and found that 100 of them also had saving account at this same bank. Construct and interpret a 95% confidence interval for the true proportion of checking-account customer who also have a savings account.

4. Novak Djokovic wins 88.1% of his tennis matches. Assuming that he has the same chance of beating each opponent, what is:
   a. The probability that he will win exactly 10 of his next 12 matches?
   b. The probability that he will lose at least 3 of his next 12 matches?

5. Given a normal distribution with a mean of 22 and a standard deviation of 4. The $P(X \leq 17) = P(X \geq a)$. What is the value of $a$?

6. A chemistry teacher was calculating student grades and found the correlation between the student’s grade on the midterm and their grade for first semester to be equal to $r = 0.873$. The teacher then realized that he had neglected to include weight for the course grade, so each semester grade was increased by 10%. Will the most likely value of the correlation coefficient now be smaller, larger or the same. Explain.

**TO BE HANDED IN**

7. Coach Carruthers thinks there is a difference in the accuracy of two different basketballs, ball A and ball B. Describe how a matched pairs experiment could be conducted using the 15 members of the basketball team to determine if ball A and ball B are more accurate for shooting free throws.

8. 2004 #4
1. It is believed that about 5.2% of adults ages 18-49 watched “Empire” last week.  If we want to conduct a survey to determine the sample size necessary to have a margin of error of ±3% for a 95% confidence interval.  Find the minimum sample size needed.

2. Honda wants to estimate the average weight of brand new 2015 Civics coming off the assembly line.  It is estimated that the standard deviation of weights is 45 pounds.  What is the minimum sample size needed to have a margin of error of no more than 12 pounds?

3. In studying his campaign plans, Mr. Morris wishes to estimate the difference between men’s and women’s views regarding his appeal as a candidate.  He asks his campaign manager to take two random independent samples and find the 99% confidence interval for the difference.  A sample of 1000 voters was taken from each population, with 388 men and 459 women favoring Mr. Morris.

4. Voters in Gotham City will be voting for mayor in November.  The local newspaper wants to conduct a poll to see how many votes are in favor of the incumbent, Mayor Linseed.  They want to results to have a margin of error of no more than ±4% with 90% confidence.  Find the minimum sample size needed.

5. A researcher estimates that typical human body temperature has an estimated standard deviation of 0.72 degrees Fahrenheit.  The formula to convert temperature from Fahrenheit to Celsius is \( C = \frac{5}{9}(F - 32) \).  Find the standard deviation in degrees Celsius.

6. Given:

<table>
<thead>
<tr>
<th>x</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(x)</td>
<td>0.3</td>
<td>0.25</td>
<td>0.15</td>
<td>0.1</td>
<td>.05</td>
<td>a</td>
</tr>
</tbody>
</table>

a) Find \( P(x = 5) \)

b) Find \( P(x \geq 3) \)

c) Find \( E(x) \)

TO BE HANDED IN

7. A statistician working with a small town is trying to project population growth for the town.  He develops a model \( \hat{y} = 3682 + 12.8x \), where \( x \) represents the number of years since 1980 and \( y \) represents the population of the town.  Identify and interpret the value of the slope for this example in context.

8. 2005 #1
1. The gestation period of gray squirrels measured in captivity is listed as 44 days as estimated by the author of *Walker’s Mammals of the World* (1991). It is recognized that the potential life span of animals is rarely attained in nature, but the gestation period could be either shorter or longer. Suppose the gestation period of a sample of 81 squirrels living in the wild is measured using the latest technologies available, and the mean length of time is found to be 42.5 days. Test the hypothesis that squirrels living in the wild have the same gestation period as those in captivity at the 0.05 level of significance. Assume that \( \sigma = 5 \) days.

2. A major car manufacturer wants to test a new engine to determine whether it meets new air-pollution standards. The mean emission of all engines of this type must be less than 20 parts per million of carbon. Ten engines are manufactured for testing purposes, and the emission level of each is determined. The data (in parts per million) are listed below:
   
   15.6  16.2  22.5  20.5  16.4  19.4  16.6  17.9  12.7  13.9

   Do the data supply sufficient evidence to allow the manufacturer to conclude that this type of engine meets the pollution standard at the \( \alpha = 0.01 \) level of significance? Assume that \( \sigma = 3 \).

3. The manager at Air Express feels that the weights of packages shipped recently are less than in the past. Records show that in the past packages have had a mean weight of 36.7 lb with a standard deviation of 14.2 lb. A random sample of last month’s shipping records yielded a mean weight of 32.1 lb for 64 packages. Is this sufficient evidence to reject the null hypothesis in favor of the manager’s claim? Use \( \alpha = 0.01 \).

4. A fire insurance company felt that the mean distance from a home to the nearest fire department in a suburb of Chicago was at least 4.7 mi. Is set its fire insurance rates accordingly. Members of the community set out to show that the mean distance was less than 4.7 mi. This, they felt, would convince the insurance company to lower its rates. They randomly identified 64 homes and measured the distance to the nearest fire department for each. The resulting sample mean was 4.4. If \( \sigma = 2.4 \) mi, does the sample show sufficient evidence to support the community’s claim at the \( \alpha = 0.05 \) level of significance?

5. In an advertisement, a pizza shop claims that its mean delivery time is less than 30 minutes. A random sample of 36 delivery times has a sample mean of 28.5 minutes. Assuming that \( \sigma = 3.5 \), is there enough evidence to support the claim at \( \alpha = 0.01 \)?

6. Employees in a large accounting firm claim that the mean salary of the firm’s accountants is less than that of its competitors, which is $45,000. A random
sample of 30 of the firm’s accountants has a mean salary of $43,500. Assuming that $\sigma = 5,200$, test the employee’s claim at $\alpha = 0.05$. 
1. An article titled “Comparisons of Mathematical Competencies and Attitudes of Elementary Education Majors with Established Norms of a General College Population” (*School Science and Mathematics* Vol 3, No 93, March 1993) reported the mean score on a test of mathematical competency for a random sample of 165 elementary education majors to be 32.63. If the general college population has a mean of 35.7 and a standard deviation of 6.73, is there sufficient evidence to show that elementary education majors score, on average, lower than then general college population? Test at the 1% significance level.

2. Suppose that the average accuracy of all watches being sold today is within 19.8 seconds per month with a standard deviation of 9.1 seconds. A watch company claims that their watches are more accurate. A consumer group tests a random sample of 36 watches from this company and reveals a mean error of 22.7 seconds per month. Is there sufficient evidence for the consumer group to claim that the company tested has less accuracy than other watches at the 2% significance level?

3. An English teacher claims that his students scored better on the Regents exam than the rest of the school. If it is known that all students in the school had a mean score of 71.5 with a standard deviation of 15.6. Does the random sample of the teacher’s student scores below show sufficient evidence that his students do better than the rest of the school at the 5% significance level?

   64, 65, 65, 74, 76, 79, 82, 82, 85, 88, 89, 93

4. The following probability distribution function shows the distribution of scores on a 4th grade spelling test.

<table>
<thead>
<tr>
<th>Score - x</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability - p(x)</td>
<td>0.19</td>
<td>0.08</td>
<td>0.09</td>
<td>0.11</td>
<td>0.24</td>
<td>0.29</td>
</tr>
</tbody>
</table>

   a. Find the mean (show all work!)
   b. Find the standard deviation (show all work!)
   c. Find the median (show all work!)

5. What is the best manner to collect data in order to show a cause and effect relationship?

6. The weights of male German Shepherds is approximately normally distributed with a population standard deviation of 3 kilograms. A breeder takes a random sample of 20 adult male German Shepherds which results in a sample mean of 37.02kg. The breeder claims that the average weight is 35kg. Calculate the z-score for this sample of 20 dogs.

---

TO BE HANDED IN
7. An educational researcher is looking to test the effectiveness of three new math programs for second grade students. A group of 60 students and their parents agree to participate in an after school program in which the math programs will be implemented. Describe an appropriate method for randomly assigning the 60 participants to three groups so that each group has exactly 20 participants.

8. 2005 #2
1. The U.S. Department of Agriculture reports that the mean cost of raising a child from birth to age 2 in a rural area is $8,390. You believe this value is incorrect, so you select a random sample of 900 children (age 2) and find that the mean cost is $8,275. At $\alpha = 0.05$, is there enough evidence to conclude that the mean cost is different from $8,390$? (Assume: $\sigma = 1,540$)

2. The number of years it took a random sample of 16 former smokers to quit permanently is listed. At $\alpha = 0.05$, test the claim that the mean time it takes smokers to quit early is 15 years. (Assume $\sigma = 4.5$)

<table>
<thead>
<tr>
<th>Years</th>
<th>15.7</th>
<th>13.2</th>
<th>22.6</th>
<th>13.0</th>
<th>10.7</th>
<th>18.1</th>
<th>14.7</th>
<th>7.0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>17.3</td>
<td>7.5</td>
<td>21.8</td>
<td>12.3</td>
<td>19.8</td>
<td>13.8</td>
<td>16.0</td>
<td>15.5</td>
</tr>
</tbody>
</table>

3. A company that makes cola drinks states that the mean caffeine content per one 12-ounce bottle of cola is 40 milligrams. Suppose you work as a quality control manager and are asked to verify this claim. During your tests, you find that a random sample of thirty 12-ounce bottles of cola has a mean caffeine content of 39.2 milligrams. At $\alpha = 0.01$, can you reject the company’s claim? (Assume $\sigma = 7.5$)

4. A scientist estimates that the mean nitrogen dioxide level in West London in greater than 28 parts per billion. You want to test this estimate. To do so, you determine the nitrogen oxide levels for 18 randomly selected days. The results (in parts per billion) are listed below. At $\alpha = 0.05$, can you support the scientist’s estimate? (Assume $\sigma = 22$)

<table>
<thead>
<tr>
<th>Parts</th>
<th>27</th>
<th>29</th>
<th>53</th>
<th>31</th>
<th>16</th>
<th>47</th>
<th>22</th>
<th>17</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>46</td>
<td>99</td>
<td>15</td>
<td>20</td>
<td>17</td>
<td>28</td>
<td>10</td>
<td>14</td>
<td>9</td>
</tr>
</tbody>
</table>

5. An auto analyst is conducting a satisfaction survey, sampling from a list of 10,000 new car buyers. The list includes 2,500 Ford buyers, 2,500 GM buyers, 2,500 Honda buyers, and 2,500 Toyota buyers. The analyst selects a sample of 400 car buyers, by randomly sampling 100 buyers of each brand. Is this an example of a simple random sample?

6. A national achievement test is administered annually to 3rd graders. The test has a mean score of 100 and a standard deviation of 15. If Jane's z-score is 1.20, what was her score on the test?

TO BE HANDED IN
7. The junior varsity volleyball team is playing against Midtown High School. The probability that East Ramapo wins the first set is 0.6. If East Ramapo wins the first set, the probability of winning the second set is 0.7; if they lose the first set they probability of winning the second set is 0.4. If either team wins the first two sets, they win the match. If they each win one set, then the probability of either team winning the third set is equally likely. Construct a tree diagram and determine the probability that East Ramapo wins the match.

8. 2005 #5
1. In Illinois, a random sample of 85 eighth-grade students has a mean score of 265 on a national mathematics assessment test. The state school administration claims that the mean score for all students is greater 260. Assuming that the population standard deviation is 52, is there sufficient evidence to support the administration’s claim at the 5% level of significance?

2. A tea drinker’s society estimates that the mean consumption of tea by a person in the U.S. is more than 7 gallons per year. In a sample of 100 people, it is found that the mean consumption is 7.8 gallons. If \( \sigma = 2.7 \), is there sufficient evidence to support the society’s claim at \( \alpha = .05 \)?

3. A coffee shop claims that its fresh-brewed coffee has a mean caffeine content of 80 milligrams per five ounces. A city health agency is asked to test the claim. A random sample of 42 five-ounce coffee serving has a mean caffeine content of 87.1 milligrams. Assuming a population standard deviation of 16.8 milligrams, is there sufficient evidence to show that the mean caffeine content is not 80 milligrams per five ounces at the 2% significance level?

4. A new weight loss program claims that participants will lose at least 10 pounds after the first month. Assuming the amount of weight lost has a population standard deviation of 7 pounds, the data for twelve randomly selected participants is:
   6.6, 8.2, 10.2, 12.0, 13.8, 11.2, 9.3, 7.1, 8.7, 10.5, 12.7, 6.7
Is there sufficient evidence to support the program’s claim at the 5% level of significance?

5. A light bulb manufacturer guarantees that the mean life of a certain type of light bulb is at least 750 hours. A random sample of 36 light bulbs has a mean life of 767.2 hours. Assuming that \( \sigma = 60 \), is there sufficient evidence to show that the manufacturer’s claim is correct at the 5% significance level?

6. A brand of cereal claims that the amount of sodium per serving is not more than 230 milligrams. A random sample of 52 cereal serving has a mean sodium content of 232 milligrams. With a population standard deviation of 10 milligrams, can we show evidence that the amount of sodium is greater than 230 milligrams at the 5% level of significance?
7. The weight of Olympic male speed skaters is approximately normally distributed with a mean of 180 pounds and a standard deviation of 9 pounds. Find, to the nearest percent, the percentage of skaters who weigh between 185 and 195 pounds.

8. 2006 #3
1. The Census Bureau has found that the distribution of ages of heads of households in a particular city is normal with a mean of 41.3 years. The City Council believes that the heads of households are currently older than 41.3 years. A random sample of 20 heads of households indicated a mean of 44.1 years with a standard deviation of 9.6 years. Is there sufficient evidence to support the City Council’s assertion at the 5% significance level?

2. Calcium is a vital nutrient for healthy bones and teeth. The National Institutes of Health (NIH) recommends a calcium intake of 1300 mg per day for teenagers. The NIH is concerned that teenagers aren’t getting enough calcium. A random sample of 20 teens recorded their food and drink consumption for one day. The researchers then computed the calcium intake of each student. Data analysis reveals that the mean was 1198 milligrams and the standard deviation was 411 milligrams. Construct and interpret a 95% confidence interval for the true mean amount of calcium intake for teenagers. Do these data provide sufficient evidence that the true mean calcium intake is less than 1300 mg?

3. A company plans its expansion based on the belief that the mean monthly sales revenue of its 1,000 offices is $0.73 million. A random sample of the sales revenue figures for 40 of its offices for last month indicated a mean revenue of $0.68 million with a standard deviation of $0.02 million. The Chief Operating Officer believes the actual sales revenue is less than $0.73 million. Conduct a hypothesis to test the C.O.O.’s claim.

4. A concerned parents’ group publishes the results of a study that claims the average amount of time that high school students watch television per week is 15.6 hours. The Student Council at a large high school conducts a simple random sample of 15 of its students and find that the mean of the sample is 14.3 hours with a standard deviation of 2.5 hours. Assuming that the population of hours watched is normal, test the claim of the Student Council president that the number of television watched per week is less than 15.6 hours.

5. The equation of the line of best fit is \( \hat{y} = 13.6x - 40.5 \). What is the residual for the point \((10,100)\)?

6. A certain restaurant claims that its hamburgers have no more than 10 grams of fat. A nutritional health agency is asked to disprove the restaurant’s claim. Explain type I and type II error in this example.

TO BE HANDED IN

7. A large university says the mean number of classroom hours per week for full-time faculty is no more than nine hours per week. A statistics professor at the university took a random sample of faculty members in an attempt to disprove the university’s claim. Explain type I and type II error in this example.

8. 2006 #3

HW 11-2
1. A company has developed a new deluxe AAA battery that is supposed to last longer than its regular AAA battery. However, these new batteries are more expensive to produce, so the company would like to be convinced that they really do last longer. Based on years of experience, the company knows that its regular AAA batteries last for 30 hours on continuous use, on average. The company selects a random sample of 15 new batteries and uses them continuously until they are completely drained. The same mean lifetime is 33.9 hours with a standard deviation of 9.8 hours. Assume the population of batteries have a lifetime than is approximately normal. Does this sample provide sufficient evidence that the mean life of the new batteries is more than 30 hours at the 5% significance level?

2. Mr. Grant believes the mean time that it takes his student to complete their homework is less than 20 minutes. He takes a random sample of 32 students and asks them how long they spent on their homework. The sample resulted in a mean of 18.8 minutes with a standard deviation of 3.4 minutes. Do the data provide sufficient evidence to support Mr. Grant’s claim?

3. On a popular self-image test, which results in normally distributed scores, the mean score for public-assistance recipients is expected to be 65. A random sample of 28 public-assistance recipients in Emerson County is given the test. They achieve a mean score of 62.1, and their scores have a standard deviation of 5.83. Construct and interpret a 95% confidence interval of the mean.

4. The weights of the drained fruit found in 21 randomly selected cans of peaches packed by Sunny Fruit Cannery were (in ounces):

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>11.0</td>
<td>11.6</td>
<td>10.9</td>
<td>12.0</td>
<td>11.5</td>
<td>12.0</td>
</tr>
<tr>
<td>10.5</td>
<td>12.2</td>
<td>11.8</td>
<td>12.1</td>
<td>11.6</td>
<td>11.7</td>
</tr>
<tr>
<td>11.2</td>
<td>12.0</td>
<td>11.4</td>
<td>10.8</td>
<td>11.8</td>
<td>10.9</td>
</tr>
</tbody>
</table>

a. Assume normality and construct and interpret a 98% confidence interval for the estimate of the mean weight of drained peaches per can.

b. Does these data provide sufficient evidence that the weight of drained fruit in the population of cans is not equal to 12 ounces?

5. Homes in a nearby college town have a mean value of $88,950. It is assumed that homes in the vicinity of the college have a higher value. To test this theory, a random sample of 12 homes is chosen from the college area. Their mean valuation is $92,460 and the standard deviation if $5,200. Assume prices are normally distributed. Construct and interpret a 90% confidence interval of the mean.

6. It is claimed that the students at a certain university will score an average of 35 on a given test. Is the claim reasonable if a random sample of test scores from this
university yields 33, 42, 38, 37, 30, 42? Assume test results are normally distributed. Construct and interpret a 95% confidence interval.

7.  A fair coin is to be flipped 6 times. The first 5 flips land "heads" up. What is the probability of "heads" on the next (4th) flip of this coin?

8.  The stemplot below shows the yearly earnings per share of stock for two different companies over a sixteen-year period.

<table>
<thead>
<tr>
<th></th>
<th>Company A</th>
<th>Company B</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>58, 75, 96, 98</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>01, 10, 17, 21, 43, 43, 53, 65, 73</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>09, 27, 29</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>73, 27, 05, 02</td>
<td></td>
</tr>
</tbody>
</table>

Identify the median, upper and lower quartiles for each company.

9.  Let X represent a random variable whose distribution is normal, with a mean of 80 and a standard deviation of 12. Write a probability statement equivalent to $P(X > 95)$?

TO BE HANDED IN

10. 2006 #5
1. The staff at a high school decides to start a “biggest loser” weight loss contest. A random sample of 12 participants is listed in the table below.

<table>
<thead>
<tr>
<th>Participant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>240</td>
<td>175</td>
<td>232</td>
<td>198</td>
<td>154</td>
<td>206</td>
<td>155</td>
<td>138</td>
<td>187</td>
<td>209</td>
<td>215</td>
<td>256</td>
</tr>
<tr>
<td>After</td>
<td>220</td>
<td>180</td>
<td>219</td>
<td>203</td>
<td>148</td>
<td>203</td>
<td>163</td>
<td>124</td>
<td>175</td>
<td>202</td>
<td>209</td>
<td>267</td>
</tr>
</tbody>
</table>

Do these data display sufficient evidence that the staff members, on average, lost weight during the contest at the 5% significance level? [Find the t-score, p-value and df only, do not write hypothesis nor verify assumptions.]

2. A teacher wants to compare the results on the Physics Regents between North High School and South High School. A random sample of 35 students at South High School produces a mean of 66.54 and a standard deviation of 15.36. A random sample of 41 students at North High School produces a mean of 56.9 with a standard deviation of 14.56. Construct and interpret a 95% confidence interval for the difference between the mean scores at the two schools. Does this interval demonstrate that the mean scores at South High School are better than the mean scores at North High School?

3. An experiment was designed to estimate the mean difference in weight gain for pigs fed ration A as compared to those fed ration B. Eight pairs of pigs were used. The pigs within each pair were littermates. The rations were assigned at random to the two animals within each pair. The gains (in pounds) after 45 days are shown in the following table.

<table>
<thead>
<tr>
<th>Litter</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ration A</td>
<td>65</td>
<td>37</td>
<td>40</td>
<td>47</td>
<td>49</td>
<td>65</td>
<td>53</td>
<td>59</td>
</tr>
<tr>
<td>Ration B</td>
<td>58</td>
<td>39</td>
<td>31</td>
<td>45</td>
<td>47</td>
<td>55</td>
<td>59</td>
<td>51</td>
</tr>
</tbody>
</table>

Assuming weight gain is normal, find the 95% confidence interval estimate for the mean of the differences between ration A and ration B.

4. To test the effect of a physical fitness course on one’s physical ability, the number of situps that a person could do in one minute, both before and after the course, was recorded. Ten randomly selected participants scored as shown in the following table. Can you conclude that a significant amount of improvement took place? Use \( \alpha = 0.01 \).

<table>
<thead>
<tr>
<th>Before</th>
<th>29</th>
<th>22</th>
<th>25</th>
<th>29</th>
<th>26</th>
<th>24</th>
<th>31</th>
<th>46</th>
<th>34</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>After</td>
<td>30</td>
<td>26</td>
<td>25</td>
<td>35</td>
<td>33</td>
<td>36</td>
<td>32</td>
<td>54</td>
<td>50</td>
<td>43</td>
</tr>
</tbody>
</table>

[Find the t-score, p-value and df, the H and A sections are not needed.]

5. The July 28, 1990 issue of *Science News* reported that smoking boosts death risk for diabetics. Suppose as a follow-up study we investigated the smoking rates for male and female diabetics and obtained the following data.

<table>
<thead>
<tr>
<th>Gender</th>
<th>n</th>
<th>Number Who Smoke</th>
</tr>
</thead>
</table>

a. Test the research hypothesis that the smoking rate (proportion of smokers) is higher for males than for females. Calculate the $p$-value.

b. What decision and conclusion would be reached at the 0.05 level of significance?

[only mechanics and conclusion needed]

6. Quality control tests yielded the following rating on two processes:
   Process A: 1.5, 2.5, 3.5, 2.5
   Process B: 2.5, 3.0, 3.0, 4.0, 3.5, 2.0

   Assume that the samples are from normal populations and are independent.
   a. Calculate the mean and standard deviation of each sample. Find the difference of the sample means.
   b. What is the probability that two samples of these sizes will produce data that indicate that B’s rating exceeds A’s rating by this difference or more, if the processes are in reality of equal effectiveness?

   [only mechanics and conclusion needed]

HW 12-1

1. The Central City Picture News conducts a poll to determine the number of Centry City residents who believe that The Flash exists. Iris West believes that at least
40% of the city residents believe that The Flash exists. In a random sample of 1,560 residents, 664 responded that The Flash is real.
   a. Construct and interpret a 95% confidence interval of the proportion of city residents who believe The Flash is real.
   b. Do these sample results supply evidence that Iris West claim is correct?

2. A pollster wants to determine the proportion of Americans who are in favor of a certain new legislation. The poll wants to construct a 95% confidence interval with a margin of error of no more than 3%.
   a. If the population proportion is unknown, what is the minimum sample size necessary?
   b. If the population proportion is estimated to be 20%, what is the minimum sample size necessary?

3. In a random sample of 345 senior citizens in the state of New York, 152 indicated that they would support a tax cut for individuals who pay school property tax, but no longer have school aged children. Construct (using formula) and interpret a 90% confidence interval for the proportion of New York state senior citizens who would support the tax cut.

4. A teacher wishes to estimate the proportion of students in the school who have a computer in their home. She takes a random sample of 80 students and constructs a 95% confidence interval for the true proportion of students who have a computer in their home. Name two ways that the teacher could make the margin of error smaller.

5. Nickson took a random sample of 450 and Shannan took a random sample of 50. They both got the same sample proportion and each constructed a 95% confidence interval. How would the width of Shannan’s interval differ from the width of Nickson’s interval?

6. The probability that David will go to prom is 0.90, the probability that Stephany will go to prom is 0.84. The probability that neither of them will go is 0.08. What is the probability that they will both go to prom? What is the probability that Stephany will go to prom, given David is going to prom?

TO BE HANDED IN

7. A potato chip producer and its main supplier agree that each shipment of potatoes must meet certain quality standards. If the producer determines that more than 8% of the potatoes in the shipment have “blemishes,” the truck will be sent away to get another load of potatoes from the supplier. Otherwise, the entire truckload will be used to make potato chips. To make the decision, a supervisor will inspect a random sample of potatoes from the shipment. The producer will then perform a hypothesis test. Explain Type I and Type II error in this situation.
1. In 2003, Major League Baseball decided to use the All-Star Game to determine home field advantage in the World Series. In the 80 World Series games played
from 2003 to 2016, the home team won 42 games. Do these data provide sufficient evidence at the 5% significance level to show that the home team has an advantage in World Series games?

2. The senior class faculty advisor agrees that if a vote of two-thirds of the senior class agrees, they will change the theme of prom. Two diligent AP statistics students conduct a random survey in which the results are those in favor of the changing the prom theme is 62% plus or minus a 3% margin of error. Is there evidence that the entire senior class will vote to change the theme? Explain.

3. A government official in China took a random sample of 200 births and found that 111 of those births were males. Is there sufficient evidence to show that the proportion of male births in China is more than 50%? (Conduct a hypothesis test and use formula)

4. A proposition on the next ballot in a certain county will ask if residents want to increase the local sales tax. A random sample of likely voters established a 95% confidence interval with a sample proportion of 0.53 with a margin of error of 0.035. Is there evidence that the proposition will pass? Explain.

5. Voters in Gotham City will be voting for mayor in November. The local newspaper wants to conduct a poll to see how many votes are in favor of the incumbent, Mayor Linseed. They want to results to have a margin of error of no more than ±4% with 90% confidence. Find the minimum sample size needed.

6. A company wants to see if the average time to answer a telephone call is greater than or less than one minute. They collect a random sample of data and find that the mean call time is 53.2 seconds and the p-value is 0.027. What would the null and alternative hypotheses have been and what conclusion would the researchers come to?

TO BE HANDED IN

7. For the following probability distribution function, find the value of the median and justify how you determined this value:

<table>
<thead>
<tr>
<th>x</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>y</td>
<td>0.01</td>
<td>0.03</td>
<td>0.04</td>
<td>0.07</td>
<td>0.11</td>
<td>0.18</td>
<td>0.29</td>
<td>0.27</td>
</tr>
</tbody>
</table>

8. 2007 #2

HW 12-3

1. A school principal wants to compare the passing rates of two math teachers on the Integrated Algebra Regents. Mr. Rogers had 60 of his 75 students pass the test
and Mr. Hooper had 31 of his 48 students pass the test. Do these data indicate to
the principal that Mr. Rogers had better student performance (passing rate) than
Mr. Hooper.

2. A salesman for a new manufacturer of cellular phones claims not only that they
cost less, but also that the percentage of defective cellular phones found among
his products will be no greater than the percentage of defectives found in a
competitor’s line. The retailer took a sample of 150 of his phones found 15 to be
defective, and a sample of 140 of his competitors phones found 6 to be defective.
Do the data support the salesman’s claim?

3. As part of the Pew Internet and American Life Project, researchers conducted two
surveys in 2012. The first survey asked a random sample of 799 U.S. teens about
their use of social media and the Internet. A second survey posed similar
questions to a random sample of 2253 U.S. adults. In these two studies, 80% of
teens and 69% of adults used social-networking sites. Do these data provide
evidence that the population proportion of teens using social networking was
greater than the proportion of adults?

4. Researchers designed a survey to compare the proportions of children who come
to school without eating breakfast in two low-income elementary schools. An
SRS of 80 students from School 1 found that 19 had not eaten breakfast. At
School 2, an SRS of 150 students included 26 who had not eaten breakfast. More
than 1500 students attend each school. Do these data give convincing evidence of
a difference in the population proportions?

5. The $p$-value for a one-sided $t$-test is 0.11. If the test had been two-sided, what
would the $p$-value have been?

6. The probability that Lou fails math is 0.12, the probability that he fails English is
0.07, the probability that he fails PhysEd is 0.18. What is the probability that he
passes all three classes?

TO BE HANDED IN

7. The principal of an elementary school has 40 students volunteer to participate in
an after school program designed to improve reading scores. Each of the students
is given a pretest and will be assigned to either the new reading program or a
program the school has been using for the last six years. Describe how this could
be conducted as a matched pairs design.

8. 2007 #3

HW 12-4

1. Mr. Carpenter feels that he has a higher percentage of male students than his
colleague Miss Withers does. Mr. Carpenter has 70 of his 115 students are males
and Miss Withers has 57 of her 118 students are males. Construct and interpret a 95% confidence interval of these data.

2. A random sample of students in a large high school found that 32 out of 45 girls and 30 out of 48 boys said they were intending to attend the spring musical. A 95% confidence interval of the difference in the proportion of girls and the proportion of boys who intend to attend the musical was calculated. What is the standard error of the difference?

3. Many new products introduced into the market are targeted toward children. The choice of behavior of children with regard to new products is of particular interest to companies that design marketing strategies for these products. As part of one study, randomly selected children in different age groups were compared on their ability to sort new products into the correct product category (milk or juice). Here are the data:

<table>
<thead>
<tr>
<th>Age Group</th>
<th>n</th>
<th>Number who sorted correctly</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-5 year olds</td>
<td>50</td>
<td>10</td>
</tr>
<tr>
<td>6-7 year olds</td>
<td>53</td>
<td>28</td>
</tr>
</tbody>
</table>

Construct and interpret a 98% confidence interval for the difference in the two proportions.

4. Aspirin prevents blood from clotting and so helps prevent strokes. The Second European Stroke Prevention Study asked whether adding another anticlotting drug, named dipyridamole, would be more effective for patients who had already had a stroke. Here are the data on strokes during the two year study:

<table>
<thead>
<tr>
<th></th>
<th>Number of Patients</th>
<th>Number of Strokes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspirin alone</td>
<td>1649</td>
<td>206</td>
</tr>
<tr>
<td>Aspirin + dipyridamole</td>
<td>1650</td>
<td>157</td>
</tr>
</tbody>
</table>

Construct and interpret a 90% confidence interval for the difference in the two proportions.

5. A 98% confidence interval for the proportion of students who have seen “Black Panther” is (.3488,.3912). What is the point estimate for the proportion of students who have seen “Black Panther.”

6. A club is organizing a raffle with several cash prizes. 1,000 raffle tickets will be sold for $5 each and a total of $2,300 in prizes will be awarded. The expected value for the net gain on each ticket is -$0.54. What is the meaning of the expected value in this context?

TO BE HANDED IN

7. A high school wishes to take a sample of 100 students. Explain how this could be conducted as a stratified random sample.
CHAPTER 13 - CHI SQUARE

1. A sociologist claims that the age distribution for the residents of a certain city is the same as it was 10 years ago. The distribution of ages 10 years ago is shown in
the table below. A random sample of 400 residents is taken and the ages recorded. At the 5% significance level is there evidence that the distribution of ages has changed?

<table>
<thead>
<tr>
<th>Ages</th>
<th>Claimed distribution</th>
<th>Survey results</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-9</td>
<td>16%</td>
<td>76</td>
</tr>
<tr>
<td>10-19</td>
<td>20%</td>
<td>84</td>
</tr>
<tr>
<td>20-29</td>
<td>8%</td>
<td>30</td>
</tr>
<tr>
<td>30-39</td>
<td>14%</td>
<td>60</td>
</tr>
<tr>
<td>40-49</td>
<td>15%</td>
<td>54</td>
</tr>
<tr>
<td>50-59</td>
<td>12%</td>
<td>40</td>
</tr>
<tr>
<td>60-69</td>
<td>10%</td>
<td>42</td>
</tr>
<tr>
<td>70+</td>
<td>5%</td>
<td>14</td>
</tr>
</tbody>
</table>

2. A bicycle safety organization claims that fatal bicycle accidents are uniformly distributed throughout the week. The following table lists the day of the week for which 910 randomly selected fatal bicycle accidents occurred. At $\alpha = 0.05$, is the distribution uniform? Assume that the probability of a fatal bicycle accident is the same for each day of the week.

<table>
<thead>
<tr>
<th>Day</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunday</td>
<td>118</td>
</tr>
<tr>
<td>Monday</td>
<td>119</td>
</tr>
<tr>
<td>Tuesday</td>
<td>127</td>
</tr>
<tr>
<td>Wednesday</td>
<td>137</td>
</tr>
<tr>
<td>Thursday</td>
<td>129</td>
</tr>
<tr>
<td>Friday</td>
<td>146</td>
</tr>
<tr>
<td>Saturday</td>
<td>134</td>
</tr>
</tbody>
</table>

3. For a 95% confidence interval for a proportion, what is the minimum sample size needed to ensure that the margin of error will be less than 3.8%?

4. Christian performs a Chi-square test for goodness of fit with seven different categories. His calculated test statistic is $\chi^2 = 13.2$. Will the null hypothesis be rejected at the 5% significance level? Justify your answer.

5. Bryan buys a bunch of bananas. Each banana that Bryan has to select from has a mean weight of 185 grams with a standard deviation of 8 grams. If Bryan randomly selects six bananas, what is the mean and standard deviation of the weight of all of his bananas?

6. A 95% confidence interval is constructed for the mean number of cars in a service station at noon on any given day is (3.7, 8.8). Interpret this confidence interval.

HW 13-2

1. The marketing consultant for a travel agency wants to determine whether certain travel concerns are related to travel purpose. A random sample of 300 travelers is selected and the results are classified as shown in the following table.
a. Assuming the variables are independent, what is the expected number of individuals who travel for business and are concerned about leg room on a plane.

b. What is the number of degrees of freedom in the Chi-square test?

2. A recording company claims that the distribution of musical preferences is as follows:
   - Classical: 4%
   - Country: 31%
   - Rap: 19%
   - Oldies: 2%
   - Pop: 18%
   - Rock: 26%

   A random sample of 150 high school students found the following results:
   - Classical: 2
   - Country: 11
   - Rap: 48
   - Oldies: 7
   - Pop: 47
   - Rock: 35

   Is there sufficient evidence that the musical preferences of students differs from the expected distributions of the record company?

3. A three-year study with 72 chronic cocaine users compared an antidepressant drug called desipramine with lithium (a standard drug to treat cocaine addiction) and a placebo. One-third of the subjects were randomly assigned to receive each treatment. Here are the results:

<table>
<thead>
<tr>
<th></th>
<th>Desipramine</th>
<th>Lithium</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relapse</td>
<td>10</td>
<td>18</td>
<td>20</td>
</tr>
<tr>
<td>No relapse</td>
<td>14</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

   Is there sufficient evidence of an association between treatment and relapse status?

4. A researcher conducts a Chi-square test and calculates $\chi^2 = 10$. For what number of degrees of freedom will the researcher reject the null hypothesis?

5. Of 200 individuals surveyed, 32 said their favorite sport to watch was basketball. The survey included 90 men and 110 women. If gender and favorite sport to watch are independent, how many women would be expected to prefer basketball?
6. A random sample of 100 registered Democrats found that 44 intended to vote for Donald Trump in the primary election. What is the standard error of the statistic?

TO BE HANDED IN

7. A study was conducted on a certain brand of car and measured the relationship between the miles the car had been driven and the resale price of the car. The equation was $\hat{y} = 38257 - 0.16292x$ where $x$ represents the miles drive and $y$ represents the resale price of the car. Identify and interpret the value of the slope in this equation.

8. 2008 #2

HW 14-1

1. A physics class performed an experiment in which the students constructed paper helicopters and randomly assigned 14 helicopters to each of five drop heights: 152cm, 203cm, 254cm, 307cm, and 442cm. The flight time of each helicopter
was measured in seconds. The residuals showed no pattern appeared to show a symmetric, mound-shaped pattern. The computer output was as follows:

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Coef</th>
<th>SE Coef</th>
<th>T</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-0.03761</td>
<td>0.05838</td>
<td>-0.64</td>
<td>0.522</td>
</tr>
<tr>
<td>Drop height (cm)</td>
<td>0.0057244</td>
<td>0.0002018</td>
<td>28.37</td>
<td>0.000</td>
</tr>
</tbody>
</table>

S = 0.168161 \quad R-Sq = 92.2\% \quad R-Sq(adj) = 92.1\%

Write the equation of the line of best fit for these data and find the value of the correlation coefficient.

2. For the computer output above, what is the standard error of the slope, $SE_b$?

3. For the computer output above:
   a. What is the $t$-statistic for the slope of the regression line, and what does this value indicate about the likelihood of rejecting the null hypothesis that $\beta = 0$?
   b. What is the $p$-value for the slope of the regression line, and what does this value indicate about the likelihood of rejecting the null hypothesis that $\beta = 0$?
   c. What is the number of degrees of freedom in this example?

4. A supplier of highway materials claims he can supply an asphalt mixture that will make roads paved with his materials less slippery when wet. A general contractor who builds roads wishes to test the supplier’s claim. The null hypothesis is “roads paced with this asphalt mixture are no less slippery than roads paved with other asphalt.” The alternative hypothesis is “roads paved with this asphalt mixture are less slippery than roads paved with other asphalt.”
   a. Describe type I error in this situation.
   b. Describe type II error in this situation.

5. The probability that Shannen is on time to class on any given day is 84%.
   a. What is the probability that she will be on time to class exactly eight of the next ten days?
   b. What is the probability that she will be late at least three times in the next ten days?

6. The Town of Pawnee wants to estimate the percentage of citizens in favor of constructing a new Senior Citizens Center. The conduct a poll by sending emails to town residents. The results showed only 38% of town residents in favor of the new center. Can these results be trusted to be representative of the entire population of the town? Explain.

TO BE HANDED IN:
7. Angelique and Giselle both perform a hypothesis test of the same sample data. They both have a $t$-statistic with eleven degrees of freedom and a $t=2.45$. Angelique rejects her null hypothesis, whereas Giselle does not. Explain how both girls could be correct.

8. 2003 #5
1.) A company that owns and services a fleet of cars for its sales force has found that the service lifetime of disc brake pads varies from car to car according to a normal distribution with mean $\mu = 55,000$ miles and standard deviation $\sigma = 4500$ miles. The company installs a new brand of brake pads on 8 cars.

a. If the new brand has the same lifetime distribution as the previous type, what is the distribution of the sample mean lifetime for the 8 cars?

b. The average life on the pads on these 8 cars turns out to be $\bar{x} = 51,800$ miles. What is the probability that the sample mean lifetime is 51,800 miles or less if the lifetime distribution is unchanged? (The company takes this probability as evidence that the average lifetime of the new brand of pads is less than 55,000 miles.)

2.) One way of checking the effect of undercoverage, nonresponse and other sources of error in a sample survey is to compare the sample with known facts about the population. About 12% of American adults are black. The proportion $\hat{p}$ of blacks in an SRS of 1500 adults should therefore be close to 12%. It is unlikely to be exactly 12% because of sampling variability.

a.) Can we use the Normal approximation for the sampling distribution of $\hat{p}$ for $n = 1500$ and $p = .12$?

b.) Can we use $\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ as the standard deviation of $\hat{p}$?

c.) If a national sample only contains 9.2% blacks, should we suspect the sampling procedure is somehow under representing blacks?

7. Use the normal approximation of the binomial, given $B(1550,.4)$, find:

e. $P(X \leq 750)$ using the binomial.

f. $P(X \leq 750)$ using the normal approximation.

g. $P(X \geq 600)$ using the binomial.

h. $P(X \geq 600)$ using the normal approximation.
3.) The weight of the eggs produced by a certain breed of hen is normally distributed with mean 65 grams and a standard deviation 5 grams. Think of a carton of eggs as an SRS of size 12 from the population of all eggs. What is the probability that the weight of a carton falls between 750 and 825 grams?

4.) Voter registration records show that 68% of all voters in Indianapolis are registered as Republicans. To test whether the numbers dialed by a random digit dialing service device really are random. You use the device to call 150 randomly chosen residential telephones in Indianapolis. Of the registered voters contacted, 73% are registered Republicans.
   a.) Is 68% a parameter or a statistic? Give the appropriate notation
   b.) Is 73% a parameter or a statistic? Give the appropriate notation
   c.) What are the mean and standard deviation of the sample proportion of registered Republicans in samples of size 150 from Indianapolis?
   d.) Find the probability of obtaining an SRS of size 150 from the population of Indianapolis voters in which 73% or more are registered Republicans. How well is your random digit device working?

9. It is believed that about 5.2% of adults ages 18-49 watched “Empire” last week. If we want to conduct a survey to determine the sample size necessary to have a margin of error of ±3% for a 95% confidence interval. Find the minimum sample size needed.

10. Honda wants to estimate the average weight of brand new 2015 Civics coming off the assembly line. It is estimated that the standard deviation of weights is 45 pounds. What is the minimum sample size needed to have a margin of error of no more than 12 pounds for a 90% confidence interval?

11. The data below shows the ages of winners of the Olympic men’s marathon.
   i. Construct an appropriate graphical display for the ages of men’s Olympic marathon winners.
   ii. Using the graphical display in part (a) describing the distribution of the ages.
   iii. Suppose the next winner of the marathon was 50 years old. How would the mean and median of the new data set with the 50 year old included compare to the mean and median of the original data?

12. Given:

<table>
<thead>
<tr>
<th>x</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>P(x)</td>
<td>0.3</td>
<td>0.25</td>
<td>0.15</td>
<td>0.1</td>
<td>.05</td>
<td>a</td>
</tr>
</tbody>
</table>

d) Find $P(x = 5)$
e) Find $P(x \geq 3)$
f) Find $E(x)$

13. The probability of a manufacturing part being defective is 0.023. A quality control inspector examines 50 parts.
   b. What is the probability that exactly one part is defective?
   c. What is the probability that at least two parts are defective?

Write your work on a separate sheet of paper.

7. The staff at a high school decides to start a “biggest loser” weight loss contest. A random sample of 12 participants is listed in the table below.

<table>
<thead>
<tr>
<th>Participant</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Before</td>
<td>240</td>
<td>175</td>
<td>232</td>
<td>198</td>
<td>154</td>
<td>206</td>
<td>155</td>
<td>138</td>
<td>187</td>
<td>209</td>
<td>215</td>
<td>256</td>
</tr>
<tr>
<td>After</td>
<td>220</td>
<td>180</td>
<td>219</td>
<td>203</td>
<td>148</td>
<td>203</td>
<td>163</td>
<td>124</td>
<td>175</td>
<td>202</td>
<td>209</td>
<td>267</td>
</tr>
</tbody>
</table>

Do these data display sufficient evidence that the staff members, on average, lost weight during the contest at the 5% significance level?

8. A teacher wants to compare the results on the Physics Regents between North High School and South High School. A random sample of 35 students at South High School produces a mean of 66.54 and a standard deviation of 15.36. A random sample of 41 students at North High School produces a mean of 56.9 with a standard deviation of 14.56. Construct and interpret a 95% confidence interval for the difference between the mean scores at the two schools. Does this interval demonstrate that the mean scores at South High School are better than the mean scores at North High School?

9. An experiment was designed to estimate the mean difference in weight gain for pigs fed ration A as compared to those fed ration B. Eight pairs of pigs were used. The pigs within each pair were littermates. The rations were assigned at random to the two animals within each pair. The gains (in pounds) after 45 days are shown in the following table.

<table>
<thead>
<tr>
<th>Litter</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ration A</td>
<td>65</td>
<td>37</td>
<td>40</td>
<td>47</td>
<td>49</td>
<td>65</td>
<td>53</td>
<td>59</td>
</tr>
<tr>
<td>Ration B</td>
<td>58</td>
<td>39</td>
<td>31</td>
<td>45</td>
<td>47</td>
<td>55</td>
<td>59</td>
<td>51</td>
</tr>
</tbody>
</table>

Assuming weight gain is normal, find the 95% confidence interval estimate for the mean of the differences between ration A and ration B.
10. To test the effect of a physical fitness course on one’s physical ability, the number of situps that a person could do in one minute, both before and after the course, was recorded. Ten randomly selected participants scored as shown in the following table. Can you conclude that a significant amount of improvement took place? Use $\alpha = 0.01$.

<table>
<thead>
<tr>
<th>Before</th>
<th>29</th>
<th>22</th>
<th>25</th>
<th>29</th>
<th>26</th>
<th>24</th>
<th>31</th>
<th>46</th>
<th>34</th>
<th>28</th>
</tr>
</thead>
<tbody>
<tr>
<td>After</td>
<td>30</td>
<td>26</td>
<td>25</td>
<td>35</td>
<td>33</td>
<td>36</td>
<td>32</td>
<td>54</td>
<td>50</td>
<td>43</td>
</tr>
</tbody>
</table>

11. The July 28, 1990 issue of *Science News* reported that smoking boosts death risk for diabetics. Suppose as a follow-up study we investigated the smoking rates for male and female diabetics and obtained the following data.

<table>
<thead>
<tr>
<th>Gender</th>
<th>$n$</th>
<th>Number Who Smoke</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>500</td>
<td>215</td>
</tr>
<tr>
<td>Female</td>
<td>500</td>
<td>170</td>
</tr>
</tbody>
</table>

c. Test the research hypothesis that the smoking rate (proportion of smokers) is higher for males than for females. Calculate the $p$-value.
d. What decision and conclusion would be reached at the 0.05 level of significance?

12. Quality control tests yielded the following rating on two processes:

- Process A: 1.5, 2.5, 3.5, 2.5
- Process B: 2.5, 3.0, 3.0, 4.0, 3.5, 2.0

Assume that the samples are from normal populations and are independent.

c. Calculate the mean and standard deviation of each sample. Find the difference of the sample means.
d. What is the probability that two samples of these sizes will produce data that indicate that B’s rating exceeds A’s rating by this difference or more, if the processes are in reality of equal effectiveness?

14. A nutritionist wants to compare the mean protein content of grilled chicken sandwich from Arby’s and McDonald’s. To do so, they randomly select several grilled chick sandwiches from each restaurant and measure the protein content of each. The results are:

- Arby’s: $\bar{x} = 23$ grams, $s = 2.1$ grams, $n = 15$
- McDonald’s: $\bar{x} = 27$ grams, $s = 1.8$ grams, $n = 12$

Assuming that both populations are normal, is there sufficient evidence to show that Arby’s grilled chicken sandwiches have a higher protein content than McDonalds?

15. A study is conducted to determine the food consumption habits of teenage males. A random sample of 20 teenage males was taken and they were asked how many 12-ounce servings of soda they drink per day. The results are:
A nutritionist claims that teenage males drink less than 3 servings of soda per day. Is there sufficient evidence to support his claim?

16. A physician claims that an experimental medication increases an individual’s heart rate. Twelve test subjects are randomly selected and the heart rate of each is measured. The subjects are injected with the medication and, after one hour, the heart rate is measured again. The results are:

<table>
<thead>
<tr>
<th>Before</th>
<th>72</th>
<th>81</th>
<th>76</th>
<th>74</th>
<th>75</th>
<th>80</th>
<th>68</th>
<th>75</th>
<th>78</th>
<th>76</th>
<th>74</th>
<th>77</th>
</tr>
</thead>
<tbody>
<tr>
<td>After</td>
<td>73</td>
<td>80</td>
<td>79</td>
<td>76</td>
<td>76</td>
<td>80</td>
<td>74</td>
<td>77</td>
<td>75</td>
<td>74</td>
<td>76</td>
<td>78</td>
</tr>
</tbody>
</table>

Is there evidence to support the physician’s claim at \( \alpha = 0.05 \)?

17. A company that makes cola drinks states that the mean caffeine content per one 12-ounce bottle of cola is 40 milligrams. Suppose you work as a quality control manager and are asked to verify this claim. During your tests, you find that a random sample of thirty 12-ounce bottles of cola has a mean caffeine content of 39.2 milligrams with a standard deviation of 7.5 milligrams. At \( \alpha = 0.01 \), can you reject the company’s claim?

18. You receive a brochure from a large university. The brochure indicates that the mean class size for full-time faculty is less than 32. You want to test this claim. Eighteen classes taught by full-time faculty are randomly selected and the class size for each is recorded:

35, 28, 29, 33, 32, 40, 26, 25, 29, 28, 30, 36, 33, 28, 27, 30, 28, 25

Is there sufficient evidence to support the claim in the brochure?

19. A guidance counselor claims high school students in a college prep program have higher ACT scores than those in a general program. The mean ACT score for 49 high school students who are in a college prep program is 22.1 and the standard deviation is 4.8. The mean ACT score for 44 high school students who are in a general program is 19.6 and the standard deviation if 5.4. At \( \alpha = 0.05 \), can you support the guidance counselor’s claim?

20. A team of heart surgeons at Saint Ann’s Hospital knows that many patients who undergo corrective heart surgery have a dangerous buildup of anxiety before their scheduled operations. The staff psychiatrist at the hospital has started a new counseling program intended to reduce this anxiety. A test of anxiety is given to patients who know they must undergo heart surgery. The each patient participates in a series of counseling sessions with the staff psychiatrist. At the end of the counseling sessions, each patient is retested to determine anxiety level. The table below indicates the results for a random sample of nine patients. Higher scores mean higher levels of anxiety. From the data, can we conclude that the counseling sessions reduce anxiety at the 1% level of significance?
Patient | Jan | Tom | Diane | Barbara | Mike | Bill | Frank | Carol | Alice  
--- | --- | --- | --- | --- | --- | --- | --- | --- | ---  
Score Before Counseling | 121 | 93 | 105 | 115 | 130 | 98 | 142 | 118 | 125  
Score After Counseling | 76 | 93 | 64 | 117 | 82 | 80 | 79 | 67 | 89

21. A study of fox rabies in southern Germany gave the following information about different regions and the occurrence of rabies in each region. Based on the information, a random sample of 16 locations in region I gave the following information about the number of cases of fox rabies near that location:

| 1 | 8 | 8 | 8 | 7 | 8 | 8 | 1 | 3 | 3 | 3 | 2 | 5 | 1 | 4 | 6 |

A second random sample of 15 locations in region II gave the following information about the number of cases of fox rabies near that location:

| 1 | 1 | 3 | 1 | 4 | 8 | 5 | 4 | 4 | 2 | 2 | 5 | 6 | 9 |

Does this information indicate that there is a difference in the mean number of cases of fox rabies between the two regions at the 5% level of significance?

22. A researcher wants to estimate the proportion of people who will respond “yes” on a survey question. He wants to estimate with a margin of error of at most 3% at a level of 95 percent confidence. What is the smallest sample size that will satisfy these requirements?

23. A comic book store owner wanted to estimate the average age of his customers. He took a random sample of forty customers and constructed a 95% confidence interval that was (32.56, 45.72). Interpret this confidence interval.

24. For problem #5, interpret the confidence level.

25. The probability that Henry will go to prom is 0.90, the probability that Ciara will go to prom is 0.84. The probability that neither of them will go is 0.08. What is the probability that they will both go to prom? What is the probability that Ciara will go to prom, given Henry is going to prom?
26. When a 95% confidence interval for a proportion is constructed, using a random sample of 100, which of the following is true?
   a. The population proportion will be included in the interval.
   b. There is a 95% chance that the sample proportion will be included in the interval.
   c. There is a 95% chance that the population proportion will be included in the interval.
   d. There is a 95% chance that the population proportion will equal the sample proportion.

27. A polling institute conducted a poll to determine the percentage of teenagers who have consumed 12 or more alcoholic beverages in the last year. The poll took a random sample of 160 teenagers and found that 43% had consumed 12 or more alcoholic beverages in the last year. The true population proportion is 47%. What are the mean and standard deviation (using formula) of the sampling proportion of teenagers who have consumed 12 or more alcoholic beverages in the last year?

28.

29. Daniel rolls a die 60 times and the result is 12 “5s”. What is most likely explanation that the result is 12 rather than 10?

1. A sociologist claims that the age distribution for the residents of a certain city is the same as it was 10 years ago. The distribution of ages 10 years ago is shown in the table below. A random sample of 400 residents is taken and the ages recorded. At the 5% significance level is there evidence that the distribution of ages has changed?

<table>
<thead>
<tr>
<th>Ages</th>
<th>Claimed distribution</th>
<th>Survey results</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-9</td>
<td>16%</td>
<td>76</td>
</tr>
<tr>
<td>10-19</td>
<td>20%</td>
<td>84</td>
</tr>
<tr>
<td>20-29</td>
<td>8%</td>
<td>30</td>
</tr>
<tr>
<td>30-39</td>
<td>14%</td>
<td>60</td>
</tr>
</tbody>
</table>
2. A bicycle safety organization claims that fatal bicycle accidents are uniformly distributed throughout the week. The following table lists the day of the week for which 911 randomly selected fatal bicycle accidents occurred. At $\alpha = 0.05$, is the distribution uniform?

<table>
<thead>
<tr>
<th>Day</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunday</td>
<td>118</td>
</tr>
<tr>
<td>Monday</td>
<td>119</td>
</tr>
<tr>
<td>Tuesday</td>
<td>127</td>
</tr>
<tr>
<td>Wednesday</td>
<td>137</td>
</tr>
<tr>
<td>Thursday</td>
<td>129</td>
</tr>
<tr>
<td>Friday</td>
<td>146</td>
</tr>
<tr>
<td>Saturday</td>
<td>135</td>
</tr>
</tbody>
</table>

3. For a 95% confidence interval for a proportion, what is the minimum sample size needed to ensure that the margin of error will be less than 3.8%?

4. Alex performs a Chi-square test for goodness of fit with seven different categories. His calculated test statistic is $\chi^2 = 13.2$. Will the null hypothesis be rejected at the 5% significance level? Justify your answer.

5. The probability that Nalani is on time to school is 0.98, the probability that Erikah is on time to school is 0.76 and the probability that neither is on time is 0.01. Given that Erikah is late to school, what is the probability that Nalani is late?

6. A 95% confidence interval is constructed for the mean number of cars in a service station at noon on any given day is (3.7, 8.8). Interpret this confidence interval.