Text Exercise Set 15

15-1 The weight of Sandor oranges is normally distributed with a mean of 8.5 ounces and a standard deviation of 0.6 ounces. The weight of Merret oranges is normally distributed with a mean of 7.8 ounces and a standard deviation of 0.8 ounces.

(a) About what percentage of Sandor oranges weigh between 7.3 and 9.7 ounces?

(b) In what range can we be certain that the weight for almost 70% of all Merret oranges will lie?

(c) About what percentage of Sandor oranges weigh less than 7.9 ounces?

(d) About what percentage of Merret oranges weigh more than 9.4 ounces?

(e) What can you say about the percentage of Merret oranges which weigh more than 5.2 ounces?

(f) What can you say about the percentage of Merret oranges which weigh less than 5.2 ounces?

(g) How does a Sandor orange weighing 9.0 ounces compare with a Merret orange weighing 9.0 ounces relative to the type of orange?

(h) How does a Sandor orange weighing 8.0 ounces compare with a Merret orange weighing 8.0 ounces relative to the type of orange?

(i) How much would a Merret orange have to weigh to be comparable to a Sandor orange weighing 9.5 ounces?

(j) How much would a Sandor orange have to weigh to be comparable to a Merret orange weighing 7.5 ounces?

15-2 A survey consisting of several questions is designed to measure anxiety levels. When a subject completes all the questions, a score between 20 and 80 can be computed, with 20 representing no anxiety and 80 representing extremely high anxiety. While performing a certain task, the anxiety levels of males are normally distributed with a mean of 55.1 and a standard deviation of 4.8, the anxiety levels of females are normally distributed with a mean of 58.9 and a standard deviation of 3.2.

(a) About what percentage of females have anxiety scores between 55.7 and 62.1 while performing the task?
15-2 - continued

(b) In what range can we be certain that the anxiety scores for about 95% of all males will lie while performing the task?

(c) About what percentage of males have anxiety scores less than 59.9 while performing the task?

(d) About what percentage of females have anxiety scores more than 65.3 while performing the task?

(e) What can you say about the percentage of males with anxiety scores more than 35 while performing the task?

(f) What can you say about the percentage of females with anxiety scores more than 75 while performing the task?

(g) How does a male whose anxiety score while performing the task is 62 compare with a female whose anxiety score is 62 while performing the task, relative to the sexes?

(h) How does a male whose anxiety score while performing the task is 50.5 compare with a female whose anxiety score is 54.2 while performing the task, relative to the sexes?

(i) What anxiety score would a male have to have while performing the task to be comparable to a female with an anxiety score of 64 while performing the task?

(j) What anxiety score would a male have to have while performing the task to be comparable to a female with an anxiety score of 54 while performing the task?

15-3 Figures 15-1a to 15-1f (reproduced here) illustrate various types of density curves that are possible for a population. Indicate which one or ones of Figures 15-1a to 15-1f could potentially be the density curve a population where

(a) the distribution is symmetric;

(b) the mean is larger than the median;

(c) more than half of the items are above the mean;
15-3 - continued

(d) obtaining a value in the middle third of the range, obtaining a value in the lower third of the range, and obtaining a value in the upper third of the range are equally likely;

(e) obtaining a value in the middle third of the range is considerably more likely than obtaining a value in the lower third of the range or obtaining a value in the upper third of the range.

15-4 Figures 15-1a to 15-1f (reproduced here) illustrate various types of density curves that are possible for a population. Indicate which one or ones of Figures 15-1a to 15-1f could potentially be the density curve a population where

(a) the distribution is skewed;

(b) the mean is smaller than the median;

(c) more than half of the items are below the mean;

(d) the mean and median are equal;

(e) obtaining a value in the lower third or upper third of the range is considerably less likely than obtaining a value in the middle third of the range.

15-5 Suppose a random sample of size \( n \) is to be selected from a parent population having a distribution with a density curve similar to the one displayed in Figure 15-1f (reproduced here).

(a) As \( n \) increases, how would the mean of the sampling distribution of \( \bar{x} \) change?

(b) As \( n \) increases, how would the standard deviation of the sampling distribution of \( \bar{x} \) change?
15-5 - continued
(c) As \( n \) increases, how would the shape of the sampling distribution of \( \bar{x} \) change?

(d) Inside the box on the right, make a rough sketch of what the density curve for the sampling distribution of \( \bar{x} \) would look like with a very large sample size \( n \).

15-6 Suppose a random sample of size \( n \) is to be selected from a parent population where is more likely to obtain a value far above or below the mean \( \mu \) of the population than it is to obtain a value close to the mean \( \mu \).

(a) Inside the box on the right, make a rough sketch of what the density curve for the parent population might look like.

(b) As \( n \) increases, how would the mean of the sampling distribution of \( \bar{x} \) change?

(c) As \( n \) increases, how would the standard deviation of the sampling distribution of \( \bar{x} \) change?

(d) As \( n \) increases, how would the shape of the sampling distribution of \( \bar{x} \) change?

(e) Inside the box on the right, make a rough sketch of what the density curve for the sampling distribution of \( \bar{x} \) would look like with a very large sample size \( n \).

15-7 Male and female college students are surveyed in order to record information about their choice of major. Table 15-1 is constructed from the data.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Mathematical Sciences</th>
<th>Natural Sciences</th>
<th>Social Sciences</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>181</td>
<td>185</td>
<td>27</td>
<td>7</td>
</tr>
<tr>
<td>Female</td>
<td>143</td>
<td>136</td>
<td>92</td>
<td>29</td>
</tr>
</tbody>
</table>

Table 15-1

Contingency Table for Exercise 15-7

Type of Major
15-7 - continued

(a) Add the row totals, column totals, and grand total to the contingency table displayed as Table 15-1. Then, explain why it makes no sense to say we want to find the correlation \( r \) between sex of the college student and choice of major.

(b) Is it proper to say that we want to study the differences in choice of major between male and female college students? Why or why not?

(c) Is it proper to say that we want to study the relationship between the different choice of majors? Why or why not?

(d) What proportion of college students chose a social science major?

(e) What proportion of college students chose a natural science major?

(f) What proportion of females chose a social science major?

(g) What proportion of natural science majors are male?

(h) What proportion of females are not math majors?

(i) What proportion of males are either math majors or natural science majors?

(j) Complete the construction of the contingency table of relative frequencies to compare the distribution of choice of major between the two sexes, then complete the construction of the corresponding stacked bar chart, and scale each bar to a height representing 100%.

<table>
<thead>
<tr>
<th>Sex</th>
<th>Type of Major</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mathematical Sciences</td>
</tr>
<tr>
<td>Male</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
</tr>
</tbody>
</table>

(k) From the stacked bar chart in part (j), do you think a relationship might exist, or do you think sex of the student and choice of major are independent?
15-7 - continued

(i) If you think a relationship exists, describe it from the stacked bar chart in part (j), if you think sex of the student and choice of major are independent, say why.

(m) Complete the construction of the contingency table of relative frequencies to compare the distribution of choice of sexes among the different choices of major, then complete the construction of the corresponding stacked bar chart, and scale each bar to a height representing 100%.

<table>
<thead>
<tr>
<th>Type of Major</th>
<th>Mathematical</th>
<th>Natural</th>
<th>Social</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>Sciences</td>
<td>Sciences</td>
<td>Sciences</td>
</tr>
<tr>
<td>Male</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(n) The stacked bar charts constructed in parts (j) and (m) are two different ways of displaying the same data. Do you prefer one over the other? If yes, why? If no, why not?

(o) Would it be appropriate to construct a scatter plot or contiguous box plots with this data? Why or why not?

(p) Could either of the two variables in this data be treated as qualitative-ordinal?

15-8 Table 15-2 displays data gathered from several subjects on three possible treatments of a particular ailment. The three treatments are labeled “Standard,” “New A,” and “New B.”

| Table 15-2  
| Contingency Table for Exercise 15-8 |
| Treatment |         |         |         |
| Complete  | Standard| New A  | New B  |
|           | 52      | 12     | 49     |
| Not Complete | 48  | 63     | 26     |
15-8 - continued

(a) Add the row totals, column totals, and grand total to the contingency table displayed as Table 15-2. Then, explain why it makes no sense to say we want to find the correlation \( r \) between cure and treatment.

(b) Is it proper to say that we want to study the difference between treatment and cures? Why or why not?

(c) Is it proper to say that we want to study the relationship between whether or not a cure is complete and the type of treatment? Why or why not?

(d) What proportion of subjects were completely cured?

(e) What proportion of subjects received the “Standard” treatment?

(f) What proportion of completely cured subjects received the “Standard” treatment?

(g) What proportion of completely cured subjects received the “New A” treatment?

(h) What proportion of subjects who received the “Standard” treatment were completely cured?

(i) What proportion of subjects who received the “New A” treatment were completely cured?

(j) What proportion of completely cured subjects did not receive the “New B” treatment?

(k) What proportion of completely cured subjects received either the “New A” treatment or the “New B” treatment?

(l) Complete the construction of the contingency table of relative frequencies to compare the distribution of complete and incomplete cures among the different treatments, then complete the construction of the corresponding stacked bar chart, and scale each bar to a height representing 100%.

(m) From the stacked bar chart in part (l), do you think a relationship might exist, or do you think the variables “cure” and “treatment” are independent?
(n) If you think a relationship exists, describe it from the stacked bar chart in part (l), if you think the variables "cure" and "treatment" are independent, say why.

(o) Complete the construction of the contingency table of relative frequencies to compare the distribution of treatments between subjects who were completely cured and subjects who were not completely cured; then complete the construction of the corresponding stacked bar chart, and scale each bar to a height representing 100%.

(p) The stacked bar charts constructed in parts (l) and (o) are two different ways of displaying the same data. Do you prefer one over the other? If yes, why? If no, why not?

(q) Would it be appropriate to construct a scatter plot or contiguous box plots with this data? Why or why not?

(r) Could either of the two variables in this data be treated as qualitative-ordinal?

15-9 The variable "Job Satisfaction Score" in the SURVEY DATA, displayed as Data Set 1-1 at the end of Unit 1, is recorded for each of three areas of residence and displayed on the right.

(a) Complete the construction of the dot plot for each area of residence.
(b) For each area of residence, obtain the five-number summary, and decide whether or not there are any outliers.
15-9 - continued

(c) Complete the construction of the box plot for each area of residence.

(d) For which area(s) of residence, if any, does the center of the distribution of "Job Satisfaction Score" appear to be greatest, and for which area(s) of residence does the center of the distribution appear to be smallest?

(e) For which area(s) of residence, if any, does the dispersion of the distribution of "Job Satisfaction Score" appear to be greatest, and for which area(s) of residence does the dispersion of the distribution appear to be smallest?

(f) How does the shape of the distribution of "Job Satisfaction Score" appear to differ, if at all, between the three areas of residence?

15-10 The variable "Yearly Income" ($1000s) in the SURVEY DATA, displayed as Data Set 1-1 at the end of Unit 1, is recorded for each of three areas of residence and displayed below.

(a) Complete the construction of the dot plot for each area of residence.

(b) For each area of residence, obtain the five-number summary, and decide whether or not there are any outliers.
15-10 - continued

(c) Complete the construction of the box plot for each area of residence.
(d) For which area(s) of residence, if any, does the center of the distribution of “Yearly Income” appear to be greatest, and for which area(s) of residence does the center of the distribution appear to be smallest?

(e) For which area(s) of residence, if any, does the dispersion of the distribution of “Yearly Income” appear to be greatest, and for which area(s) of residence does the dispersion of the distribution appear to be smallest?

**Dot Plots for Yearly Income**

- **Rural Area Yearly Income ($1000s)**
- **Suburban Area Yearly Income ($1000s)**
- **Urban Area Yearly Income ($1000s)**

(f) How does the shape of the distribution of “Yearly Income” appear to differ, if at all, between the three areas of residence?

**Box Plots for Yearly Income**

- Rural
- Suburban
- Urban

- **Yearly Income ($1000s)**
  - 20 30 40 50 60 70 80