Answers to Study Questions for Exam Three

Disclaimer: these solutions were created by a human. If you find any errors, or even suspect an error, please contact your instructor to ask about it. Thank you!

1. 
   
   (a) \( \bar{x} = \frac{15+5+3+2+33+2}{6} = \frac{60}{6} = 10 \)

   (b) Be sure to put the data in increasing order first: 2, 2, 3, 5, 15, 33. Since there are 6 values in the set, the median is the average of the 3\(^{rd}\) and 4\(^{th}\) values: \( \frac{3+5}{2} = 4 \).

   (c) The mode is 2, since it occurs more than any other value.

   (d) The range is max – min = 33 – 2 = 31.

   (e) Use a table to find \( \sum (x - \bar{x})^2 \).

<table>
<thead>
<tr>
<th>Data Value</th>
<th>( x )</th>
<th>( x - \bar{x} )</th>
<th>( (x - \bar{x})^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>2 – 10 = -8</td>
<td>64</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2 – 10 = -8</td>
<td>64</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>3 – 10 = -7</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>5 – 10 = -5</td>
<td>25</td>
</tr>
<tr>
<td>15</td>
<td>15</td>
<td>15 – 10 = 5</td>
<td>25</td>
</tr>
<tr>
<td>33</td>
<td>33</td>
<td>33 – 10 = 23</td>
<td>529</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>756</td>
</tr>
</tbody>
</table>

   So \( s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}} = \sqrt{\frac{756}{5}} = \sqrt{151.2} = 12.296. \)

   (f) \( Q_1 = 2. Q_2 = 4. Q_3 = 15. \)

   (g) 

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td></td>
</tr>
</tbody>
</table>

   (h) The middle 50% of the data falls between the two quartiles, so between 2 and 15.
   (i) The upper 50% of the data falls between the median and the highest value, so between 4 and 33.

2. The mean of the data is 31.71, and the standard deviation is 10.75.
3.
(a) **Understand the problem.** I need to devise a set of data with the given characteristics: the average must be 6, the value with the highest frequency must be 4 and the middle value when the set is in ascending order needs to be 5.

**Devise a plan.** I will build my set a step at a time by filling in blanks with values that will create the correct median first (since that’s a location measure) and then make sure I have the correct mode. Once that is done, I’ll have a little more leeway to pick numbers for the rest of the set so that it has the correct mean.

**Carry out the plan.** I’ll try to keep the set small, but I want more than a couple of values so there is room to work in. I’m going to try for six data values. If my set has six values, and the average must be 6, the numbers will need to add up to 6 x 6 = 36.

Space for six values: ___, ___, ___, ___, ___, ___

Since the median must be 5, I’ll try 4 and 6 on either side of the “middle.”

___, ___, 4, 6, ___, ___

Now I will need to be careful to keep smaller values than 4 to the left and larger values than 6 to the right, since my set should be in increasing order to maintain the correct median. Next I’ll make sure I have lots of 4s since that needs to be the mode. I only need two of them if I don’t repeat any other values, but maybe I can use three of them and finish off the left side of the set:

4, 4, 4, 6, ___, ___

Now I just need to make sure that my values sum to 36. I’ve used 4 + 4 + 4 + 6 = 18 so far, so I need 36 – 18 = 18 more. The easiest way to do this is to use 8 and 10, and since these are both larger than 6, they will fit perfectly into the last two spots.

4, 4, 4, 6, 8, 10

**Answer: one such set is 4, 4, 4, 6, 8, 10.**

**Look back and check.** I have checked my calculations: since \( \bar{x} = \frac{3+4+4+6+7+12}{6} = \frac{36}{6} = 6 \), I have the correct mean. The data is in increasing order so the median is \( \frac{4+6}{2} = 5 \). The value of 4 occurs with the highest frequency, so it is the mode. There are many other solutions to this, but this one certainly works.

(b) Since the mean is 7, the total of the scores on the six quizzes is 6*7 = 42. The sum of the recorded quizzes is 8 + 9 + 7 + 5 + 6 = 35. The remaining quiz score must be 42 - 35 = 7.
4. (a) If a set has a standard deviation of zero, **all the data values must be identical**. Example: 5, 5, 5. The mean is \( \bar{x} = \frac{5+5+5}{3} = \frac{15}{3} = 5 \), and standard deviation is zero:

<table>
<thead>
<tr>
<th>Data Value</th>
<th>( x )</th>
<th>( x - \bar{x} )</th>
<th>( (x - \bar{x})^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>5</td>
<td>5 - 5 = 0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>5 - 5 = 0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>5 - 5 = 0</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

\[
s = \sqrt{\frac{\sum (x - \bar{x})^2}{n-1}} = \sqrt{\frac{0}{3}} = \sqrt{0} = 0.
\]

(b) If all the data in a set is very close to the mean, **the standard deviation must be very small since the data is not spread out**. Example: 8.9, 9, 9, 9.1. The mean is 9, and the data is all quite close to this value. The standard deviation is 0.082.

5. **Set A** would have the highest standard deviation, since the data values are more spread out. In set B, the data values are all very close to the mean so the standard deviation would be lower.

6. (a) The two sets cannot be compared without more information. The data in Set A might have very small values (say, short quiz scores, ranging from 0 to 3) whereas the data in Set B might be larger (exam scores ranging from 0 to 100).

(b) If we know that both sets are scores from the same exam, then we can compare the sets using standard deviation. We know that the data in Set B is a lot more spread out than the data in Set A because the standard deviation is larger for Set B.

7. One of the two sets might be much more spread out than the other. Consider set A and set B as follows:

Set A:  1, 100, 100, 199
Set B:  99, 100, 100, 101

Both sets have mean 100 and mode 100, but set A is much more spread out than set B.

8. The mean takes the numerical value of every data value into account, but the median does not. The median is not affected by extreme values but the mean is very sensitive to extreme values. This gives the median an advantage over the mean for data sets with extreme values.
9. The standard deviation will give an indication of how widely the scores vary.

10. The mean will increase by \( k \) and the standard deviation will be unaffected.

11. 

(a) 

<table>
<thead>
<tr>
<th>4</th>
<th>5 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>6 8</td>
</tr>
<tr>
<td>6</td>
<td>1 2 5</td>
</tr>
<tr>
<td>7</td>
<td>1 1 1 7 9</td>
</tr>
<tr>
<td>8</td>
<td>0 3 9 9</td>
</tr>
<tr>
<td>9</td>
<td>4 5 8 8</td>
</tr>
</tbody>
</table>

(b) Since there are 20 values in the set, the median is the average of the 10\(^{th}\) and 11\(^{th}\) values in the set. \( \frac{71 + 77}{2} = 74 \).

(c) The mode is 71, since it occurs with the highest frequency.

(d) \( Q_1 = (61 + 62)/2 = 61.5 \)

\( Q_2 = \text{Median} = 74 \)

\( Q_3 = (89 + 89)/2 = 89 \)

(e) 

![Boxplot](image)
Exam 3 Study Guide Solutions

(f)

<table>
<thead>
<tr>
<th>Score</th>
<th>Frequency f</th>
</tr>
</thead>
<tbody>
<tr>
<td>40 - 49</td>
<td>2</td>
</tr>
<tr>
<td>50 - 59</td>
<td>2</td>
</tr>
<tr>
<td>60 - 69</td>
<td>3</td>
</tr>
<tr>
<td>70 - 79</td>
<td>5</td>
</tr>
<tr>
<td>80 - 89</td>
<td>4</td>
</tr>
<tr>
<td>90 - 99</td>
<td>4</td>
</tr>
</tbody>
</table>

(g)

(h)

(i)

12. (a) There are eleven values in the set. In increasing order they are:

Q₁: 0.25(11) = 2.75 so Q₁ has 2.75 values below it – or, Q₁ is the third value. Q₁ = 23.

Q₂: 0.5(11) = 5.5 so Q₂ has 5.5 values below it. Q₂ = $51.$

Q₃: 0.75(11) = 8.25 so Q₃ has 8.25 values below it – or, Q₃ is the ninth value. Q₃ = 58.
13. (a) Mean = \[
\frac{(0)(4) + (1)(6) + (2)(5) + (3)(3) + (4)(2)}{20}
\]
= \[
\frac{0 + 6 + 10 + 9 + 8}{20}
\] = \[
\frac{33}{20}
\] = 1.65 cars.

(b) Since there are 20 values, the median is between the 10th and 11th values. The 10th value is 1 and the 11th value is 2, so the median is 1.5 cars.

(c) The value that occurred the most often is 1 (with a frequency of 6) so the mode is 1 car.

14. (a) The sum of all the values in the set is 542 so the mean is \[
\frac{542}{20} = 27.1
\] students.

(b) Since there are 20 values in the set, the median is the average of the 10th and 11th values. So the median is \[
\frac{27 + 27}{2} = 27
\] students.

(c) There are two values with the same highest frequency, so both of those are modes: 26 students and 34 students.

15.

16.

17. A pie chart would be more useful in this case. A pie chart makes it easier to compare the sizes of each different category with the rest of the data.

18. A bar graph would be more useful in this case. A bar graph makes it easier to see which data value has the highest frequency at a glance.

19. Remember that relative frequency is the same thing as empirical probability. So we add the relative frequencies for the families of size 2, 3, and 4:

\[
P(\text{family size less than 5}) = 0.416 + 0.238 + 0.198 = 0.852.
\]
21.

(a) 

![Normal Distribution Curve]

(b) 26 and 50 years
(c) 95%
(d) 2 and 74 years

22.

a. Looking up $z = 0.7$ in the table, we get 0.2580 so 25.8%.
b. Looking up $z = 1.91$ in the table, we get 0.4719 so 47.19%.
c. $P(0.7 \leq z \leq 1.91) = 0.2139$
d. $P(0.7 \leq z) = 0.2420$

23.(a) Solve for $x$:

$$z = \frac{x - \bar{x}}{s}$$

$$1.8 = \frac{x - 2.76}{0.64}$$

$$1.152 = x - 2.76$$

$$x = 1.152 + 2.76 = 3.91.$$  

(b) Find the $z$-score using $z = \frac{x - \bar{x}}{s} = \frac{3.77 - 3.24}{0.21} = \frac{0.53}{0.21} = 2.5.$

(c) Even though Jay has a higher GPA (3.91) than Rick (3.77), Jay’s z-score of 1.8 is lower than Rick’s z-score of 2.5. So Rick has the better GPA relative to his college.
24.

a. 

\[ P(70 \leq x) = 0.5 \]

b. Since 70 is the median, 50% of the students scored 70 or higher, so 40 students. Or, use 
\[ P(70 \leq x) = 0.5 \] so 
\[ (0.5)(80) = 40 \text{ students}. \]

c. \[ z = 0 \]

d. \[ z = 1 \]

e. \[ z = -1.6 \]

f. \[ 0.34 = 34\% \] (Or, using the table, 0.341 = 34.1%)

g. Using the table for \[ z = -1.6 \], we see that the percentage between 62 and 70 is .445 = 44.5%. Adding that to 34% we get 78.5%. (Or, using the table, 34.1% + 44.5% = 78.6%.)

h. Since \[ P(62 \leq x \leq 75) = 78.5\% \] (see part h), 78.5% of the students scored between 62 and 75. Thus 
\[ (0.785)(80) = 62.8 \text{ or about 63 students scored between 62 and 75}. \]