Review Exercises

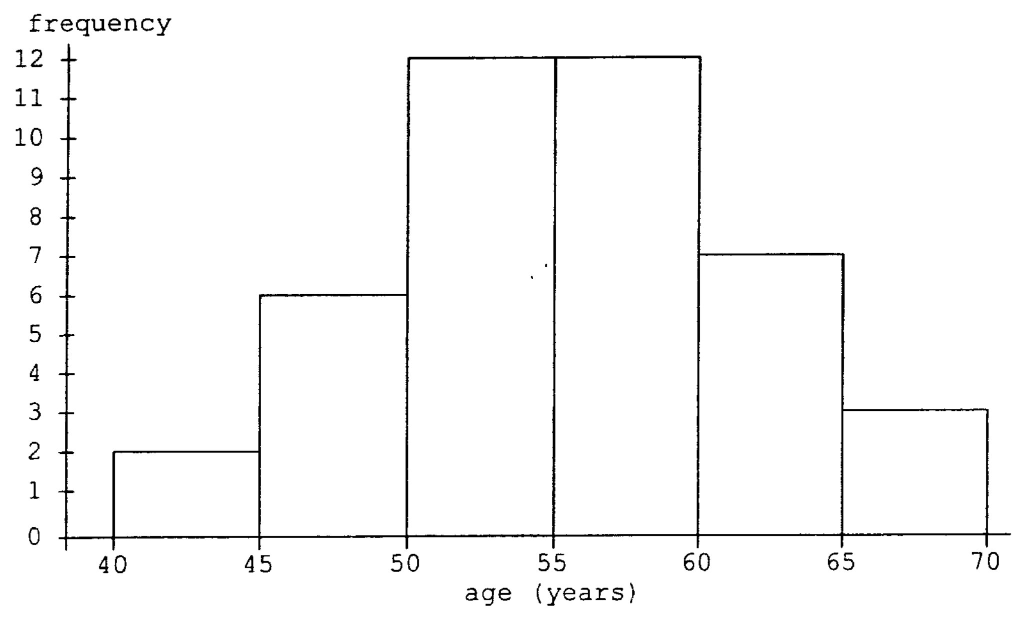
1. The scores arranged in order are:
   42 43 46 46 47 48 49 49 50 51 51 51 51 51 52 52 54 54 54 54 55
   55 55 55 56 56 56 57 57 57 57 58 60 61 61 61 62 64 64 65 68 69
   preliminary values: n = 42, $\Sigma x = 2304$, $\Sigma x^2 = 128{,}014$
   a. $\bar{x} = (\Sigma x)/n = (2304)/42 = 54.9$
   b. $\tilde{x} = (55 + 55)/2 = 55.0$
   c. M = 51
   d. m.r. = (42 + 69)/2 = 55.5
   e. R = 69 - 42 = 27
   f. s = 6.3 (from part g)
   g. $s^2 = [n(\Sigma x^2) - (\Sigma x)^2]/[n(n-1)]$
   $= [42(128{,}014) - (2304)^2]/[42(41)] = (68{,}172)/1722 = 39.6$
   h. For $Q_1 = P_{25}$, $L = (25/100)\cdot 42 = 10.5$ rounded up to 11. And so $Q_1 = x_{11} = 51$.
   i. For P_{30}, $L = (30/100)\cdot 42 = 12.6$ rounded up to 13. And so $P_{30} = x_{13} = 51$.
   j. For $D_7 = P_{70}$, $L = (70/100)\cdot 42 = 29.4$ rounded up to 30. And so $D_7 = x_{30} = 57$.

2. a. $z = (x - \bar{x})/s$
   $z_{42} = (42 - 54.857)/6.292 = -2.04$
   b. Yes; Teddy Roosevelt's inaugural age is unusual since -2.04 < -2.00.
   c. According to the Range Rule of Thumb, the usual values are within 2s of $\bar{x}$.
   usual minimum: $\bar{x} - 2s = 54.9 - 2(6.3) = 42.3$
   usual maximum: $\bar{x} + 2s = 54.9 + 2(6.3) = 67.5$
   In addition to the 42 in part (b), the ages 68 an 69 are also unusual.

3.

age	frequency
40 - 44	2
45 - 49	6
50 - 54	12
55 - 59	12
60 - 64	7
65 - 69	3
	42

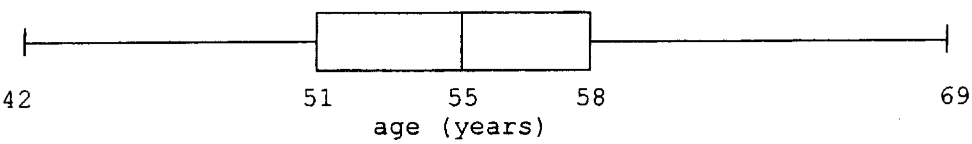
4.



NOTE: Unlike other data, age is usually not reported to the nearest unit. The bars in a histogram extend from class boundary to class boundary. Because of the way that ages are reported, the boundaries here are 40, 45, 50, etc -- i.e., someone 44.9 years old still reports an age of 44 and crosses into the next class only upon turning 45, not upon turning 44.5.

5. The scores are given in order in exercise #1.
   For $Q_1 = P_{25}$, $L = (25/100)\cdot 42 = 10.5$, rounded up to 11.
   For $\tilde{x} = Q_2 = P_{50}$, $L = (50/100)\cdot 42 = 21$ -- an integer, use 21.5.
   For $Q_3 = P_{75}$, $L = (75/100)\cdot 42 = 31.5$, rounded up to 32.

   The 5-number summary is:
   $\min = x_1 = 42$
   $Q_1 = x_{11} = 51$
   $Q_2 = x_{21.5} = (55 + 55)/2 = 55$
   $Q_3 = x_{32} = 58$
   $\max = x_{42} = 69$



6. a. Since scores 48.6 to 61.2 are within 1·s of $\bar{x}$, the Empirical Rule for Data with a Bell-Shaped Distribution states that about 68% of such persons fall within those limits.
   b. Since scores 42.3 to 67.5 are within 2·s of $\bar{x}$, the Empirical Rule for Data with a Bell-Shaped Distribution states that about 95% of such persons fall within those limits.

7. $z = (x - \bar{x})/s$
   management, $z_{72} = (72 - 80)/12 = -0.67$
   production, $z_{19} = (19 - 20)/5 = -0.20$

   The score on the test for production employees has the better relative position since $-0.20 > -0.67$.

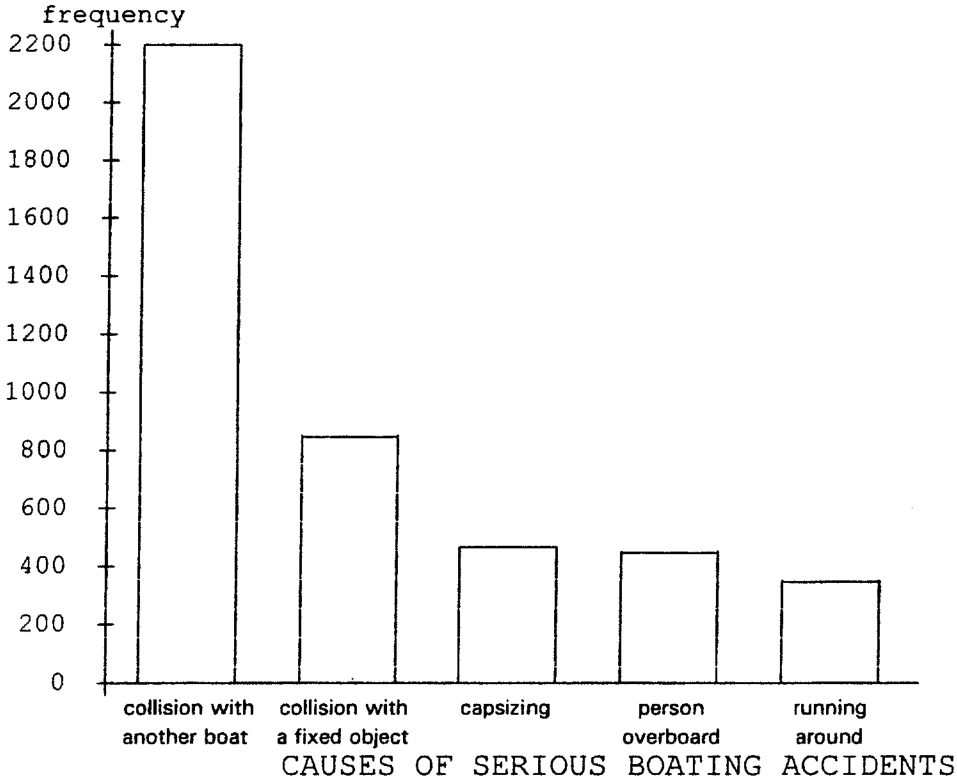
8. In general, the Range Rule of Thumb states that the range typically covers about 4 standard deviations -- with the lowest and highest scores being about 2 standard deviations below and above the mean respectively. The following answers assume the ordered textbook ages range from $x_1 = 0$ years (i.e., copyrighted during the current year) to $x_n = 12$ years.
   a. The estimated mean age is $(x_1 + x_n)/2 = (0 + 12)/2 = 6$ yrs.
   b. The estimated standard deviation of the ages is $R/4 = (x_n - x_1)/4 = (12 - 0)/4 = 3$ yrs.

9. In general, adding the same value to every score will move the all scores up the number line by the same amount. This changes the location of the scores on the number line, but it does not affect the spread of the scores or the shape of the distribution -- i.e., each measure of center will change by the amount added, but the measures of variation will not be affected.
   a. Since adding 5.0 minutes to every score moves all the scores 5 units up on the number line, the mean will increase by 5.0 to $12.3 + 5.0 = 17.3$ minutes.
   b. Since adding 5.0 minutes to every score does not affect the spread of the scores, the standard deviation will not change and remain 4.0 minutes.

c. Since adding 5.0 minutes to every score does not affect the spread of the scores, the variance will not change and will remain $(4.0)^2 = 16.0$.

NOTE: The general principle involved here is discussed for measures of center in exercise #21a of section 2-4 and for measures of variation in exercise #30a of section 2-5.

10. Arranging the categories in decreasing order by frequency produces the following figure.



Cumulative Review Exercises

1. The scores arranged in order are: 0 1 4 7 7 7 11 12 12 12 13 18 19 23 25 30
   preliminary values: $n = 16$, $\Sigma x = 201$, $\Sigma x^2 = 3625$

   a. $\bar{x} = (\Sigma x)/n = (201)/16 = 12.6$
   $\tilde{x} = (12 + 12)/2 = 12.0$
   $M = 7, 12$ (bi-modal)
   m.r. $= (0 + 30)/2 = 15.0$

   b. $R = 30 - 0 = 30$
   $s^2 = [n(\Sigma x^2) - (\Sigma x)^2]/[n(n-1)]$
   $= [16(3625) - (201)^2]/[16(15)] = (17599)/240 = 73.3$
   $s = 8.6$

   c. continuous. Even though the values were reported rounded to whole years, age can be any non-negative value on a continuum. NOTE: While age is continuous, it could be that the population from which the sample was selected was the ages rounded to whole years -- and so it may be argued that "the population from which the given years were drawn" was discrete.

   d. ratio. Differences are consistent and there is a meaningful zero; 8 years is twice as much time as 4 years.

2. a. mode. The median requires at least ordinal level data, and the mean and the midrange require at least interval level data.
   b. convenience. The group was not selected by any process other than the fact that they happened to be the first names on the list.
   c. cluster. The population was divided into units (election precincts), some of which were selected at random to be examined in their entirety.

3. No; using the mean of the state values counts each state equally, while states with more people will have a greater affect on the per capita consumption for the population in all 50 states combined. Use the state populations as weights to find the weighted mean.