**Chapter 2 Exercises**

**Concepts**

1. What does skewness imply about a distribution? How would a positive or negative skew impact the mean of a distribution? The standard deviation?
2. What are contexts in which it might be useful to know the relative standard distance rather than the standard distance? Vice versa?
3. The weighted mean center minimizes the sum of squared distances, while the median center minimizes the sum of distances. If a facility is to be built in an area where one settlement is very far away from the rest of the settlements, would the citizens of that settlement prefer that mean or median center be used to determine the location of the facility? What would the other settlements prefer?

**Exercises**

1. Using the SPSS Housing dataset, answer the following questions:
	1. By hand, visualize the price, number of bedrooms, date built, and floor area of the first 15 records. For each variable, produce a histogram, box plot, and stem and leaf plot.
	2. Use a software package to visualise the price, number of bedrooms, date built, and floor area of all 499 records. Produce a histogram and a box plot for each.
2. Household incomes for 5 houses in a community of 6 total houses are: $90,000, $110,000, $110,000, $132,000, and $125,000. Assuming there is no way to determine the income of the 5th house, estimate the:
	1. The mean and median household income of the community.
	2. Which value would you use as your estimate of the 6th house price and why?
3. A department committee asks 18 random geography majors to rate a class using a scale from 0 to 10. The student responses are included in the table below.
	1. Compute the mean, median, mode, range, variance, standard deviation and interquartile range individually for females and males.
	2. Assume that a female student scores a 9.5 in the same test – what would her *z*-score be? How would this change if the student was a male?
	3. What is the score from (b) above in terms of percentile, for both the cases?

|  |  |  |
| --- | --- | --- |
| Frequency | Score | Gender |
|  |  |  |
| 1 | 5 | Male |
| 2 | 7 | Male |
| 1 | 10 | Male |
| 3 | 5 | Female |
| 4 | 6 | Female |
| 2 | 7 | Female |
| 3 | 8 | Female |
| 2 | 9 | Female |
| Total |  |  |

1. Using the Milwaukee home sales dataset, answer the following questions:
	1. For the finished square feet, number of bedrooms, number of bathrooms, and sales price, give the mean, median, range, and standard deviation.
	2. Using the first 10 records, convert the finished square feet, number of bedrooms, and sales price into *z*-scores.
	3. What percentage of houses have a garage? Air conditioning? An attic?
2. Assume that a community is divided into the following age/sex demographic categories. Find the grouped mean and the grouped variance for the ages of males and females respectively.

|  |  |  |
| --- | --- | --- |
| **Age** | **Male** | **Female** |
| <15 years | 12 | 10 |
| 15‒24 years | 10 | 9 |
| 25‒54 years | 33 | 31 |
| 55‒64 years | 8 | 8 |
| 65+ years | 6 | 8 |

Note that, as the first and last categories are open-ended, an assumption must be made regarding their endpoints. In this example, use 8 and 70 for the first and last groups.

1. Using the SPSS Housing dataset, determine the coefficients of variation for price and floor area.
2. Observations from a percentage distribution have a mean of 52, median of 52.1 and a standard deviation of 7. 8.1% of the observations are greater than 66; 7.9% of the observations are below 38. Based on this information discuss the skew and kurtosis of this distribution when compared to a normal distribution.
3. Draw frequency distributions which are a) positively skewed b) negatively skewed c) leptokurtic (high kurtosis) and not skewed d) platykurtic (low kurtosis) and not skewed.
4. A square grid is imposed over a city map. Give the Euclidean and Manhattan distances between the points at (3,1) and (1,8).
5. Given the set of settlements shown below, find the following:

|  |  |  |  |
| --- | --- | --- | --- |
| **Settlement** | **X** | **Y** | **Population** |
| A | 8.9 | 1.2 | 1,000 |
| B | 7.8 | 2.3 | 50,000 |
| C | 2.3 | 9.1 | 2,500 |
| D | 9.4 | 3.3 | 3,000 |
| E | 8.8 | 3.7 | 8,000 |

* 1. Find the unweighted mean center.
	2. Find the weighted mean center, and describe the differences between the measures.
	3. Find the relative standard distance to the weighted mean center, assuming that the study area ranges from coordinates (0,0) in the southwest to (10,10) in the northeast. Repeat assuming the northeast is defined by the point (12,12).
1. Repeat part (c) from Exercise 8 with the assumption that the study area is defined as a circular area centered at (5,5). Compare the relative distance of the settlements if the radius is 5 versus 10.
2. Find the angular mean and circular variance of the following sample of 10 measurements, and create a rose diagram to visualise them: 3°, 52°, 79°, 90°, 122°, 238°, 241°, 248°, 250°, 324°. Also, interpret the variance value for whether the angular data is dispersed or clustered around a particular direction.

**Chapter 2 solutions**

**Price**







**Bedrooms**







**Year Built**







**Floor Area**







**Price**







**Bedrooms**





**Year Built**





**Floor Area**





* 1. Mean = $113,400, Median = $110,000.
	2. A case could be made to use either value as the estimate of the home. In this case, there does not appear to be any outlier values that would substantially influence the mean. Therefore, we would use the mean.
	3. Female: Mean = (3\*5)+(4\*6)+(2\*7)+(3\*8)/14 = 6.7857; Median = (5,5,5,6,6,6,6,7,7,8,8,8,9,9) = (6+7)/2 = 6.5; Mode = 6; Range = 4;
	Variance = 1.883, Standard Deviation = 1.37, Interquartile range: (8-6) = 2

|  |  |  |
| --- | --- | --- |
| **Y** | **(Y-M)** | **(Y-M)2** |
| 5 | -1.786 | 3.190 |
| 5 | -1.786 | 3.190 |
| 5 | -1.786 | 3.190 |
| 6 | -0.786 | .618 |
| 6 | -0.786 | .618 |
| 6 | -0.786 | .618 |
| 6 | -0.786 | .618 |
| 7 | 0.21 | .044 |
| 7 | 0.21 | .044 |
| 8 | 1.214 | 1.474 |
| 8 | 1.214 | 1.474 |
| 8 | 1.214 | 1.474 |
| 9 | 2.214 | 4.902 |
| 9 | 2.214 | 4.902 |

 Male: Mean = (1\*5) + (2\*7) + (10\*1)/4 = 7.25; Median = (5,7,7,10) = 7; Mode = 7; Range = 5; Variance = 3.1875, Standard Deviation = 1.785, Interquartile range: (8.5-6) = 2.5

|  |  |  |
| --- | --- | --- |
| **Y** | **(Y-M)** | **(Y-M)2** |
| 5 | -2.25 | 5.0625 |
| 7 | -0.25 | .0625 |
| 7 | -0.25 | .0625 |
| 10 | 2.75 | 7.5625 |

b. *Z-*score for 9.5 score female: (9.5-6.7857)/1.37 = 1.98124; Male: (9.5-7.25)/1.785 = 1.26

c. Percentile rank = (number of values below x / n)\*100

Females: 14/15\*100 = 93.33%; Males: 3/5\*100 = 60%



|  |  |  |  |
| --- | --- | --- | --- |
| **Record** | **Finished Square Feet** | **Bedrooms** | **Sales Price** |
| 1 | 0.0688 | 0.66339 | -0.83833 |
| 2 | 0.0335 | -0.37805 | -0.5453 |
| 3 | -0.25369 | -0.37805 | -0.52647 |
| 4 | -0.28578 | -0.37805 | -0.49823 |
| 5 | -0.48793 | -0.37805 | -0.79714 |
| 6 | -0.49916 | -0.37805 | -1.1443 |
| 7 | -0.7302 | -0.37805 | -0.6771 |
| 8 | -0.77673 | -0.37805 | -0.72065 |
| 9 | 0.23566 | -0.37805 | -0.05575 |
| 10 | -0.44943 | 0.66339 | -0.69679 |

* 1. Attic: 36.6% Air Conditioning: 68.3% Garage: 89.9%
1. Females: grouped mean = 38.5, grouped variance = 11.058. Males: grouped mean= 36.5, grouped variance = 11.043
2. Price: 0.511 Floor area: 0.339
3. The data has high kurtosis because: the median is slightly larger than the mean; the right tail has more data than the left tail; the tails contain more than 7.5% as many observations, which is more than 3 times 2.5% ‒ which would be the tail probabilities of a normal distribution.
4. d









1. Euclidean: 7.280. Manhattan: 9
2. A
	1. x = 1.91, y = 2.14
	2. Euclidean distance = 1.1, Manhattan distance = 1.23
	3. Standard distance = 1.17, Relative standard distance = 0.7
	4. 7.44, 3.92
	5. 7.80, 2.77. The center has moved to be closer to the largest settlement.
	6. For 10 by 10: 1.587. For 12 by 12: 1.322.
3. Radius 5: 5.624. Radius 10: 2.812.
4. Angular mean: -55.181°, 55.181° Circular variance: 0.938. The data appears to be clustered toward the 180° direction.

