Quiz 2 practice – Numerical Descriptive Statistics

|  |
| --- |
| Thickness (") |
| 0.754 |
| 0.735 |
| 0.754 |
| 0.748 |
| 0.740 |
| 0.752 |
| 0.747 |
| 0.740 |
| 0.751 |
| 0.741 |
| 0.740 |
| 0.742 |
| 0.748 |
| 0.732 |
| 0.750 |
| 0.747 |
| 0.750 |
| 0.752 |
| 18 total |

 1) To test new equipment for making ¾" plywood,

measurements of thickness were obtained.

Calculate the range and standard deviation of the data.

2) Calculate the standard deviation of a sample of 350 steel rods (use your calculator’s built in functions):

|  |  |
| --- | --- |
| Diameter (mm) | Frequency |
| 10.00 | 40 |
| 10.01 | 75 |
| 10.02 | 100 |
| 10.03 | 90 |
| 10.04 | 45 |

3) The following set of data was taken from repeated measurements of wind velocity at Vancouver Airport over a year and represents the maximum velocity each month. (The data is not ordered by month as given.)

a.) Be able to briefly discuss the monthly maximum wind velocity using a variety of summary statistics. my note: look for standard deviation, range, and max value (anything you can make some assumptions on).

Imagine you are considering the possible wind effects on a new multi-storey hotel to be built in

|  |
| --- |
| Wind Velocity(km/h) |
| 39.16 |
| 115.22 |
| 21.11 |
| 43.04 |
| 37.01 |
| 30.1 |
| 19.33 |
| 70.54 |
| 76.11 |
| 24.35 |
| 109.58 |
| 50.27 |

Richmond. There is a fairly low probability that maximum Wind Velocity will be greater than 3 standard deviations above the mean.

b.) What wind velocity is 3 sample standard deviations above the sample mean?

c.) If we design the hotel so that it can withstand the wind velocity found in part b.), what is the probability that this wind velocity will be exceeded if we assume that extreme wind velocities follow a normal distribution (Empirical Rule)?

|  |  |
| --- | --- |
| **Site A** | **Site B** |
| 2.99 | 1.02 |
| 4.75 | 3.56 |
| 8.79 | 3.5 |
| 5.59 | 3.45 |
| 2.32 | 4.5 |
| 1.9 | 13.6 |
|  | 4.5 |
|  | 2.3 |
|  | 3.5 |
|  | 2.6 |
|  | 3.31 |
|  | 3.1 |

4) The following two sets of sample data are measurements of the drainage rate for a square metre of the soil in two different building sites (liters per m2 per s).

1. Consider using statistics you calculated and other appropriate numbers, to discuss the difference between these two sites. Which of the sites has the most “relatively-variable” drainage? Conclude which site you would be more likely to recommend constructing a building on.
2. For the data set for site B only, assume that the drainage rates should be randomly distributed in a bell shaped curve. Can either or both of the max and min values could be discounted as being too extreme?
3. If you were to exclude the minimum data point from the data set for site A what would be the effect on the mean, median and standard deviation?

|  |  |
| --- | --- |
| Before  Compacting  (g/cm3) | After Compacting  (g/cm3) |
|  |
| 0.1 | 15.46 |
| 1.03 | 17.54 |
| 2.69 | 18.66 |
| 3.54 | 19.03 |
| 4.33 | 19.31 |
| 13.41 | 20.08 |
| 14.04 | 21.2 |
| 14.04 | 21.29 |
| 14.83 | 21.4 |
| 53.26 | 24.68 |

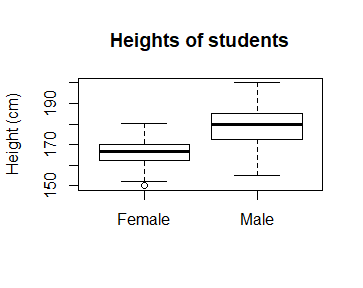
5) The following soil density measurements were made on a strip mall development site in Richmond before and after

compacting the sand.

Make a graphical comparison and also calculate summary statistics to describe the differences before and after. Imagine you are trying to describe how the compacting of the soil has improved the site for building purposes to the client. Use the values which are most suitable for making comparisons. **Hint:** a decent minimal description will use 3 or 4 of the values. Don't waste a lot of time trying to use all of the values.

1. Hank and Pete work in a fabrication facility. Records indicate that Hank completes an average of 62 items per shift, with a standard deviation of 4.2 items. Pete completes an average of 59 items per shift, with a standard deviation of 4.1 items. Which of the two would you consider to be the more consistent worker? Explain.
2. Suppose a lumber mill ships 4x4’s with a mean moisture content of μ = 18% with σ = 0.5%. A histogram reveals that the data is bell-shaped.
3. Use Chebyshev's Rule to determine the minimum percentage of scores that lie between 17% and 19%.
4. Repeat question a.) using the Empirical Rule.
5. Use the Empirical Rule to estimate the percentage of values greater than 19.5%.
6. Use the Empirical Rule to estimate the percentage of values that are between 18% and 18.5%.
7. Building specifications require that the moisture content be under 19% to be closed in. Use the Empirical Rule to estimate the percentage of 4x4’s that is suitable to be closed in.
8. Human temperatures have a mean of 98.20oF and a standard deviation of 0.62oF. Determine whether each of the following temperatures is usual or unusual.
9. 101.00oF
10. 96.90oF
11. 96.98oF

1. The heights of 237 statistics students are represented in the boxplot below. Make one observation about the relative heights of male and female students. Your observation must contain both a reference to the numbers in the boxplot, as well as a proportion or percentage (eg, “x% of female students are shorter than…”).



Answers

1.

|  |  |
| --- | --- |
| Range | 0.022 |
| Standard Dev | 0.006560179 |

2.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Stdev |  | | 0.0120078 | | mm | |
|  |  |  | |  | |

3a.) We see that the mean maximum monthly wind velocity is is 52.99 km/h with a standard deviation of 32.98 km/h. There is considerable variability seen in the standard deviation. Also, the range of values is 95.89 km/h. As a result, it is difficult to predict the maximum wind velocity during a month.

b.) mean + 3s = 52.98 + 3(32.98) = 151.92 km/h

c.) It is a bit of a leap to apply the Empirical Rule since our data does not reflect a normal distribution. If we assume that the population is normal anyway, the probability of being more than 3 standard deviation above the mean is:

(100 - 99.7)/2 = 0.15%

4 a)

Site B has a larger coefficient of variation, larger standard deviation, and larger range so it is relatively more variable than Site A. I would recommend Site A as it's mean drainger is higher and it is more consistent.

b.)

xbar - 3sigma = 4.07 - 3(3.14) = -5.3

xbar - 2sigma = 4.07 - 2(3.14) = - 2.2

xbar + 2sigma = 4.07 + 2(3.14) = 10.3

 xbar +3sigma = 4.07 + 3(3.14) = 13.5 L/m2/s

The maximum value is more than 3 standard deviations above the mean and thus is a very rare value.

c.)

If we remove the minimum from Site A the mean and median should increase and the standard deviation should decrease.

5.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *Before Compaction (g/cm3)* |  | *After Compaction (g/cm3)* |  |  |
|  |  |  |  |  |  |
|  | Mean | 12.127 | Mean | 19.865 |  |
|  | Median | 8.87 | Median | 19.695 |  |
|  | Mode | 14.04 | Mode | #N/A |  |
|  | Standard Deviation | 15.63321 | Standard Deviation | 2.50923738 |  |
|  | Skewness | 2.347438 | Skewness | 0.17569909 |  |
|  | Range | 53.16 | Range | 9.22 |  |
|  | Minimum | 0.1 | Minimum | 15.46 |  |
|  | Maximum | 53.26 | Maximum | 24.68 |  |
|  | Sum | 121.27 | Sum | 198.65 |  |
|  | Count | 10 | Count | 10 |  |

The mean and median soil density have increased after compaction, so overall compacting the soil increases the density. More importantly, by comparing the Ranges and the Coefficients of Variation, we can see that there is much less variability in the soil density after compaction.

6.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Mean | St. Dev. | CV |  |
| Hank | 62 | 4.2 | 6.774194 | % |
| Pete | 59 | 4.1 | 6.949153 | % |
|  |  |  |  |  |

Comparing the CV's we see that Hank is a more consistent worker even though his standard deviation is slightly higher than Pete's.

7a.)

17% is 2σ below the mean and 19 is 2σ above the mean. Applying Chebyshev with k = 2, we have at least 75% of values between 17% and 19%

b.) 95%

c.) 19.5% is 3σ above the mean. 99.7% are within 3σ of the mean when the data is normally distributed. 0.3% of the data is outside the 3σ range. We are only interested in the upper tail of the distribution so 0.3%/2 = 0.15%

d.) We know that 68% of the values are between -1σ and 1σ. 34% of the values should then be located between the mean and 1σ.

e.) 19% is 2σ above the mean. We want the entire bell except the part that is more than 2σ. 5%/2 = 2.5%

8a.) z=4.52; unusual

b.) z=-2.10; unusual

c.) z=-1.97; usual

9) (Possible answer) 50% of the male students are taller than the tallest female student, who is 180 cm tall.