

Guideline for Marking:

See below for how each (sub-) question can be marked according to the points given.

- 2 points: give 1 point if some calculation done; 2 points if the (student's) answer is in general correct except with minor numerical errors; 0 point otherwise.
- 3 points: give 1 point if some calculation done; 2 points if the reasoning or logic behind the answer is mostly correct (more than 70%); 3 points if the answer is in general correct except with minor numerical errors; 0 point otherwise.
- 4 points: give 1 point if some calculation done; give 2 points if the answer is correct for 50% of the solution; give 3 points if the reasoning or logic behind the answer is mostly correct; give 4 points if the answer is in general correct except with minor numerical errors; 0 point otherwise.
- 5 points: give 1 point if some calculation done; give 2 points if the answer is mostly correct for 50% of the solution; give 3 points if the answer is mostly correct for 70% of the solution; give 4 points if the reasoning or logic behind the answer is mostly correct; give 5 points if the answer is in general correct except with minor numerical errors; 0 point otherwise.

Question 1 – Exponential Distribution

The distance between major cracks in a highway follows an exponential distribution with a mean of 5 miles.

- a) What is the probability that there are no major cracks in a 10-mile stretch of the highway? (3 points)

Answer:

$$l = \frac{1 \text{ crack}}{5 \text{ mile}}$$

$$P(X > 10) = e^{-al} = e^{-10 \cdot \frac{1}{5}} = 0.1353$$

- b) What is the probability that there are two major cracks in 10-mile distance? (3 points)

Answer:

$$P(y = 2) = \frac{e^{-\frac{1}{5} * 10} \left(\frac{1}{5} * 10\right)^2}{2!} = 0.2707$$

- c) What is the standard deviation of the distance between major cracks? (3 points)

Answer:

$$\text{var}(x) = \frac{1}{7} \rightarrow \text{Std}(x) = \sqrt{\frac{1}{7}} = \sqrt{5}$$

d) What is the probability that the first major crack occurs between 15 and 20 miles of the start of inspection? (4 points)

Answer:

$$P(15 < x < 20) = P(X < 20) - P(X < 15) = e^{-15 \cdot \frac{1}{5}} - e^{-20 \cdot \frac{1}{5}} = 0.031$$

Question 2 – Normal Calculation

Over the past few years, concerns have grown over increasing traffic in the Berlin. One way the Berlin Department of Transportation measures traffic congestion is by looking at travel time. To help people plan their travel, Berlin Department of Transportation estimates that average trip from central train station to Berlin wall at 5:40 pm (at peak) takes 17 minutes and, 95% the time, the trip will not exceed 30 minutes. They also believe this travel time approximates a normal distribution

a) What is the standard deviation of peak travel time for the trip from central train station to the Berlin wall? (4 points)

From the question, we know that 95 times in 100, the trip will take 30 minutes or less. So, the entire area to the left 30 is 95%. To use the normal table to determine z, subtract .5 from the probability, and you are left with 45%, corresponding to a z-score of 1.645.

$$z = \frac{x - m}{s} \rightarrow s = \frac{30 - 17}{1.645} = 7.903$$

So, the standard deviation is about 8 (7.9) minutes.

b) Using your answer in part a, what proportion of drive times fall between 20 and 30 minutes? (4 points)

We are looking for $P(20 < x < 30)$.

$P(20 < x < 30) = P(x < 30) - P(x < 20) = .95 - .6480 = .302$ or 30% of drive times fall between 20 and 30 minutes.

- c) Write a paragraph explaining your results to a non-technical policy audience. (3 points)

The average trip at peak travel time, 5:40 pm, takes about 17 minutes and most of the time (19/20 times) it takes less than 30 minutes. However, 68% of the time, it will be between about 9 (9.1) and 25 (24.9) minutes. In 3 out of 10 trips, the trip will take between 20 and 30 minutes.

Question 3 – Descriptive Statistics

Forty randomly selected students were asked the number of pairs of sneakers they owned. Let X = the number of pairs of sneakers owned. The results are as follows:

X	Frequency	Relative Frequency	Cumulative Relative Frequency
1	2		
2	5		
3	8		
4	12		
5	12		
7	1		

- Find the sample mean \bar{x} (2pts)
- Find the sample standard deviation, s (2 pts)
- Sample coefficient of variation (2 pts)
- Construct a histogram of the data (4pts)
- Complete the columns of the chart (4 pts)

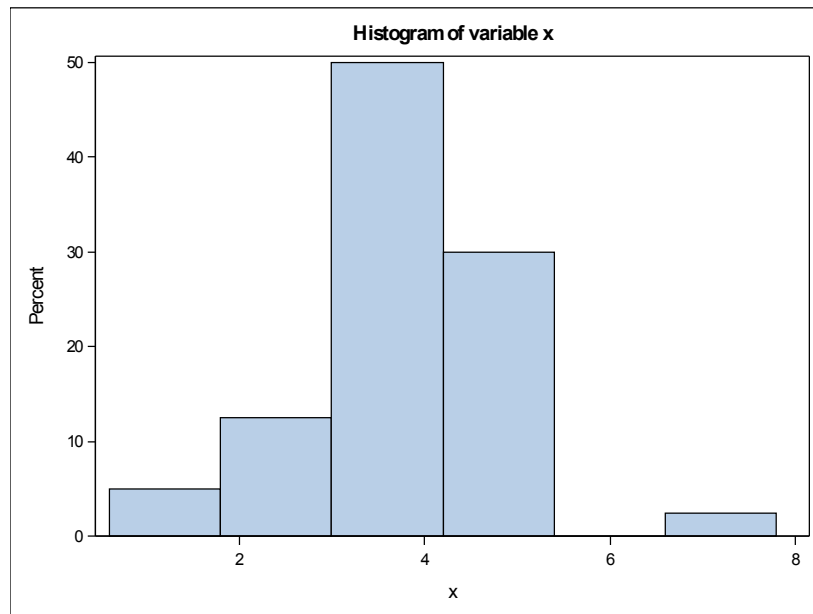
- f. Find the first quartile (2pts)
- g. Find the median. (2 pts)
- h. Find the third quartile (2pts)
- i. Construct a box plot of the data (4pts)
- j. What percent of the students owned at least five pairs? (2pts)
- k. Find the 40th percentile. (2pts)
- l. Find the 90th percentile (2pts)
- m. Construct a stem plot of the data (4pts)

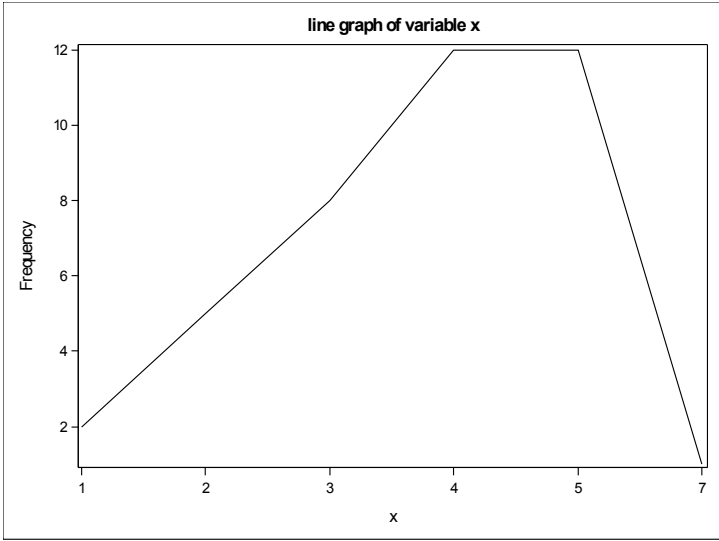
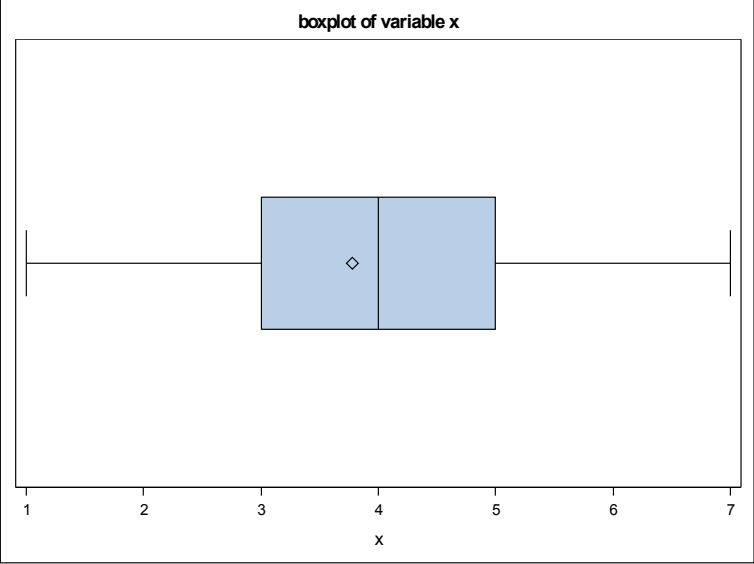
Answer:

Descriptive statistics of variable x

The MEANS Procedure

Analysis Variable : x							
Mean	Std Dev	Coeff of Variation	Lower Quartile	Median	Upper Quartile	40th Pctl	90th Pctl
3.78	1.29	34.19	3.00	4.00	5.00	4.00	5.00





Mini Case Study: Land Use Requirements for Solar Power

Note: You should follow the plan-do-report format to present your solution.

NREL, the National Renewable Energy Laboratory, has analysed data on the land use requirements for 3 different types of solar power installation: Small PV, Large PV and CSP, and published the results in:

<http://www.nrel.gov/docs/fy13osti/56290.pdf>

You are not required to read the whole report, but just the parts referred to below. This assignment does not require an understanding of the difference between PV (photovoltaic) and CSP (concentrating solar power), but if you are interested you could look it up in Wikipedia.

In total 166 solar power projects were analyzed as summarized in Table 2 and Figure 2. NREL distinguishes between direct and total land use requirements, as explained in Section 2 and Figure 1.

The land use requirements per MegaWatt (MW) of generating capacity are given in Section 4.1 and Figure 3, which uses a different type of box & whiskers plot from the one used on ADM2303. Some of the data used in Figure 3 are as follows:

	Small PV		Large PV		CSP	
	Total	Direct	Total	Direct	Total	Direct
Max	17.3	12.2	12.5	10	17.5	13.9
75 %ile	10	7.2	10	7.8	14.8	9.1
25 %ile	6	4.8	7.5	6	7	4.5
Min	2.5	2.2	4.2	4.1	4.9	2

This assignment does not require graphical answers. Give your answer in descriptive text, with reasons.

- (a) (1) If NREL had used the type of plot used on ADM2303, which part(s) of the plot would be removed?
- (b) (2) If NREL had used the type of plot used on ADM2303, which part(s) of the plot might be reduced in size?
- (c) (6) If NREL had used the type of plot used on ADM2303, which of the 6 box and whisker plots would show outliers?
- (d) (3) Suppose you wanted to use the NREL data to estimate total land use requirements per MW of Small PV and CSP. Which of the estimates would be more accurate, small PV or CSP?
- (e) (2) Describe the difference between the total and direct land use requirements per MW of Large PV.

Solar Power NREL (total 10 points)

Plan – comparing with the standard plot covered in 2303 and figuring out how different graphs convey different information (2 points)

Do – Describe the analysis based on given data (as stated below).

Report – Make conclusion (as stated below).

SOLUTION

- (a) the capacity weighted average, since the 2303 box plot just shows the median (1pts)
- (b) the whiskers, since the 2303 box plot whiskers extend at most to $1.5 \cdot \text{IQR}$ which is not necessarily the full range of the data (1pts)
- (c) Small PV total and small PV direct, since the upper whiskers in the 2303 boxplot extend only to 16 and 10.8 acres/MW respectively. Students should do the calculation below or draw 2303-style box/whiskers plots in order to substantiate their answer. (2pts)
- (d) Small PV. Although the range is slightly wider, the IQR is much smaller. Also the small PV plot is based on more data than the CSP plot. (2pts)
- (e) Total land use requirements per MW are more variable than direct requirements, since the range is wider and the IQR is wider. (2pts)

	Small PV		Large PV		CSP	
	Total	Direct	Total	Direct	Total	Direct
Max	17.3	12.2	12.5	10	17.5	13.9
75 %ile	10	7.2	10	7.8	14.8	9.1
25 %ile	6	4.8	7.5	6	7	4.5
Min	2.5	2.2	4.2	4.1	4.9	2
IQR	4	2.4	2.5	1.8	7.8	4.6
Upper whisker	16	10.8	13.75	10.5	26.5	16
Lower whisker	0	1.2	3.75	3.3	-4.7	-2.4
Upper outlier	Yes	Yes	0	0	0	0
Lower outlier	0	0	0	0	0	0