PSET 1 - Due: Sunday, June 23

1. A small survey was conducted in which each respondent was asked how many times, in the previous two-week period, they had eaten at a fast food restaurant. The data appear below

0, 2, 1, 5, 2, 2, 3, 4, 1, 2, 7, 1, 3, 4, 1, 0, 1, 4, 2, 1, 3, 3, 2, 1, 9, 1

(a) Construct a frequency histogram. The histogram should be neat, accurate, and well labeled.

(b) How would you describe the shape the distribution?

(c) Find the proportion of the respondents described by each of the following.

(i) Ate at a fast food restaurant at least four times

(ii) Ate at a fast food restaurant fewer than two times

Given the above data set, we can first find the frequencies of each data point. We can then construct a frequency distribution to get a better look at the data

Number of days	Frequency
0	2
1	8
2	6
3	4
4	3
5	1
6	0
7	1
8	0
9	1

Table 1: *Number of times $n = 26$ people ate a fast food restaurant in the previous two weeks organized as a frequency distribution*

From this, we create the histogram

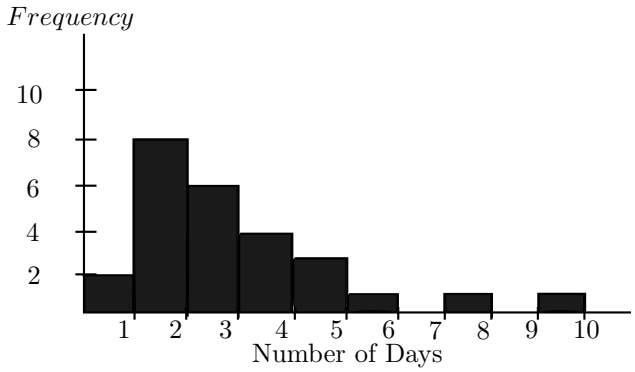


Figure 1: *Number of times $n = 26$ people ate a fast food restaurant in the previous two weeks organized as a histogram for discrete data*

b.) From the histogram, we can see that the shape of the distribution is skewed positively (right)

c.) Furthermore, we can find the relative frequency (proportion) of respondents who ate at a fast food restaurant at least four times by dividing the summation of the frequencies $x \in \{4, 5, 6, ..., 9\}$ by $n = 26$ (the total number of respondents). We find

$$\frac{3 + 1 + 0 + 1 + 0 + 1}{26} = 0.2308 \approx 23\%.$$

We preform a similar calculation to find the number of respondents who ate at a fast food restaurant fewer than two times, in this case we get

$$\frac{2 + 8}{26} = 0.3846 \approx 38\%.$$

2. The blood glucose levels (in milligrams per deciliter) for 25 patients at a medical facility appear below.

63 65 66 68 69 71 73 74 75 75 76 76 77
79 79 81 81 81 83 84 86 87 90 91 95

(a) Construct a relative frequency histogram. Use nine class intervals starting with $60 \leqslant \text{Glucose} < 64$.

(b) How would you describe the shape of the distribution?

(c) What proportion of the patients had a blood glucose level described by each of the following?

(i) At most 71

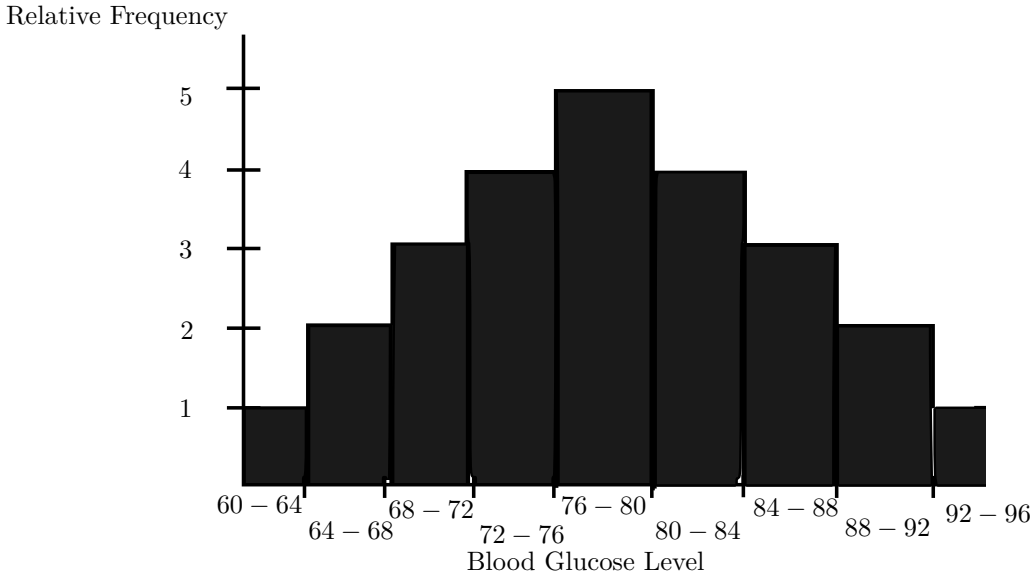
(ii) Between 72 and 91 (inclusive of the endpoints)

We again begin by organizing the data via a frequency distribution

Blood glucose levels (mg/dL)	Frequecy	Relative frequency
60-64	1	0.04
64-68	2	0.08
68-72	3	0.12
72-76	4	0.16
76-80	5	0.2
80-84	4	0.16
84-88	3	0.12
88-92	2	0.08
92-96	1	0.04

Table 2: *Blood glucose levels for $n = 25$ patiens at a medical facility organized as a frequency distribution*

Using the relative frequencies found above, we construct a simple continuous data, equal class width histogram



b.) The shape of this distribution is approximately normal (symmetric)

3. Refer to Problem 1.

(a) Calculate the sample mean.

(b) Calculate the sample median.

(c) How do the answers to (a) and (b) compare?

(i) Why should we have been able to anticipate this?

4. A new type of outdoor paint was tested by painting six homes in the same geographic area. The number of months the paint lasted before fading was recorded and yielded the values below.

10, 60, 50, 30, 40, 20

(a) Calculate the sample range.

(b) Calculate the sample mean.

(c) Calculate the sample variance using the definition formula (i.e. by using the squared

(d) deviations).

(e) Calculate the sample variance using the short-cut formula.

(f) Calculate the sample standard deviation.

(g) Give the units of measurement for (b); for (c); and for (e).

5. Refer to Problem 4.

(a) Add 10 to each data value in Problem 4.

(i) Re-calculate the sample mean.

(ii) How does the answer in (i) compare to that in 4(b)? Be specific.

(iii) Re-calculate the sample variance. Note – using the definition formula might be more illuminating.

(iv) How does the answer in (iii) compare to that in 4(c)? Be specific.

(b) Divide each data value in Problem 4 by 10 (i.e. multiply each data value by 1/10).

(i) Re-calculate the sample mean.

(ii) How does the answer in (i) compare to that in 4(b)? Be specific.

(iii) Re-calculate the sample variance. Note – using the definition formula might be more illuminating.

(iv) How does the answer in (iii) compare to that in 4(c)? Be specific.

6. Suppose that $x_1, x_2, ..., x_n$ is a set of data with $\bar{x} = 40$, $\tilde{x} = 60$, $s_x^2 = 25$ and $s_x = 25$ Suppose that each x_i is transformed to a new data value $y_i = 2x_i - 1$

(a) Compare $\bar{x}$ to $\tilde{x}$. What do the values potentially tell us about the shape of the distribution?

(b) Find each of the following

(i) $\bar{y}$

(ii) $\tilde{y}$

(iii) s_y^2

(iv) s_y