Exam 2 Learning Objectives

Chapter 5 – Exploring Data Distributions

- Know the difference between an individual and a variable.
- Determine the distribution of a variable by finding the values a variable can have and how often they occur.
- Construct a histogram for a data set. ✗ class size (width)
- Describe types of distribution for a histogram by looking at the shape and spread of the distribution.
- Construct a stemplot and/or dotplot for a data set.
- Calculate the mean, median, mode, quartiles and range of a set of data.
- Construct the diagram of a boxplot from the data set's five-number summary.
- Determine if a data set has an outlier.
- Calculate the standard deviation of a small data set.
- Given the mean and standard deviation of a normally distributed data set, compute the first and third quartiles, apply the 68–95–99.7 rule and sketch the normal curve.

Chapter 6 – Exploring Data Relationships

- In a data set determine the response variable and the explanatory variable.
- Draw a scatterplot for a small data set consisting of pairs of numbers and determine how the variables are associated (positively or negatively) and how strongly they are associated (the correlation)
- From a scatterplot, draw an estimated regression line and determine if there are outliers.
- Use the regression line to estimate values of the variables.
- Understand the difference between interpolation and extrapolation and know when extrapolation may not be valid
- Understand correlation and regression describe relationships that need further interpretation because association does not imply causation and outliers have an effect on these relationships.

Chapter 7 – Data For Decisions

- Identify the population and the sample in a given sampling or experimental situation.
- Know the different types of bias that can occur when sampling a population and analyze a sampling example to detect sources of bias.
- Determine a numbering scheme and use it with a table of random digits to select a random sample of a given from a population.
- Know the difference between an observational study and an experiment.
- Recognize the confounding on the effects of two variables in an experiment.
- Know the difference between the experimental group and the control group in an experiment.
- Understand what is meant by statistically significant.
- Describe the placebo effect and why double blindness is desirable in an experiment.
- Define statistical inference and the difference between a parameter and a statistic.
- Compute the sample proportion when both the sample size and number of favorable responses are given.
- Using an appropriate formula, calculate the standard deviation of a given statistic. $SD = \sqrt{\frac{p(1-p)}{n}}$ n is the sample size
- Explain the difference between the population mean and the sample mean. m of error = 2*SD p $\hat{p}$
- Given a sample proportion and sample size, list the range for a 95% and 99.7% confidence interval for the population proportion.
- Calculate differing margins of error for increasing sample sizes or how the sample size needs to increase to achieve a certain margin of error.

95% $\Rightarrow$ $p \pm 2*SD$

99.7% $\Rightarrow$ $p \pm 3*SD$

Exam 2 Practice Problems

1. Below are exam scores for 24 students in Math 167 class.

100	99	96	92	92	91	91	89	87	85
82	80	80	80	79	78	77	76	71	67
62	59	56	53						

$M = \frac{80+80}{2} = 80$

$Q_3 = 91 = \frac{91+91}{2}$

$Q_1 = \frac{71+76}{2} = 73.5$

(a) Show this data in a stem plot.

(b) Show this data in a histogram with class width of 10.

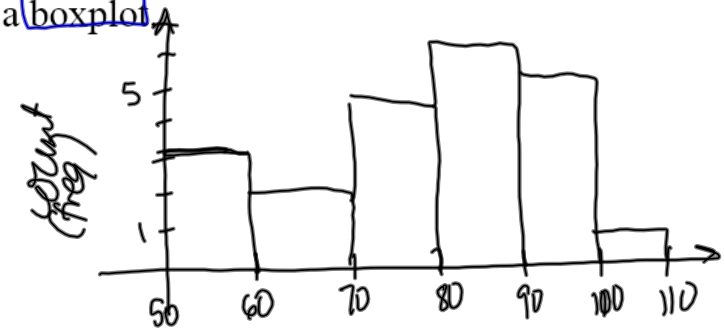
(c) Find the mean and mode (if any) for this data

(d) Display the data in a boxplot.

1a)

5	369
6	27
7	16789
8	00002579
9	112269
10	0

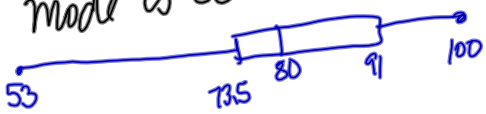
1b)



1c) $\frac{53+56+\ldots+100}{24} = \frac{1922}{24} = 80.0833$

mode is 80

1d)



2. A data set has a minimum value of 25, Q1 = 45, M=50, Q3=57, and a maximum value of 70. Are there outliers?

$IQR = Q_3 - Q_1 = 57 - 45 = 12$

low? $Q_1 - 1.5 \times IQR = 45 - 1.5 \times 12 = 27 \Rightarrow$ at least one low outlier.

high? $Q_3 + 1.5 \times IQR = 57 + 1.5 \times 12 = 75 \Rightarrow$ no high outliers

3. A data set has seven values, 12, 14, 14, 16, 16, 16, 17

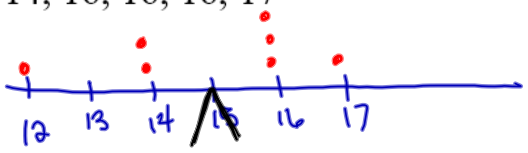
(a) Show this data in a dot plot.

(b) What is the range of values?

(c) Where is the balance point?

(d) What is the standard deviation ?

a)



b) $17-12=5$

c) balance @ mean $\Rightarrow$ $\frac{12+\cdots+17}{7}=15$

d)

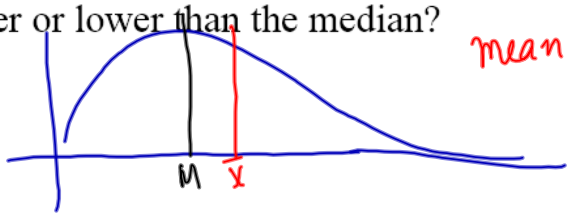
x	$x-\bar{x}$	$(x-\bar{x})^2$
12	$12-15=-3$	9
14	-1	1
14	-1	1
16	1	1
16	1	1
16	2	4
101	0	18

$$s=\sqrt{\frac{\sum(x-\bar{x})^2}{n-1}}=\sqrt{\frac{18}{7-1}}=\sqrt{\frac{18}{6}}=\sqrt{3}$$

mean is pulled towards the tail

4. If a graph is skewed to the right, would you expect the mean to be higher or lower than the median?

mean is higher than median

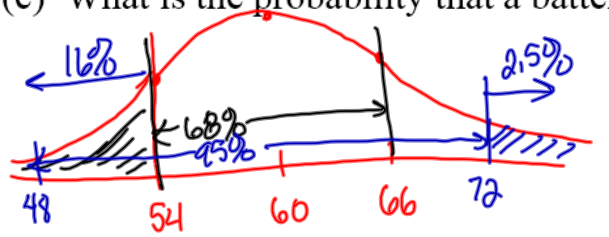


5. The length of time that a particular type of battery lasts is normally distributed with a mean of 60 hours and a standard deviation of 6 hours.

(a) Sketch the appropriate normal curve for this distribution and locate the quartiles.

(b) What is the probability that a battery lasts less than 54 hours? 16%

(c) What is the probability that a battery lasts more than 72 hours? 2.5%



$Q_1=60-.67(6)=55.98$ hours

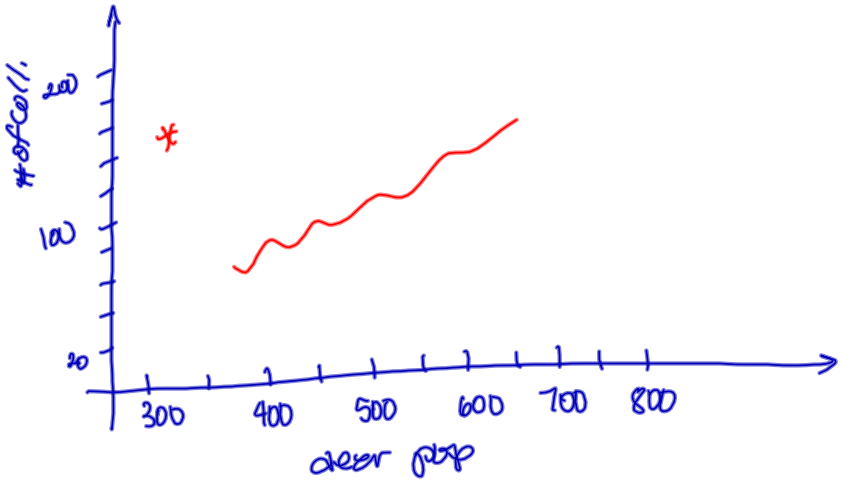
$Q_3=60+.67(6)=64.02$ hours

6. The table below has the deer population in a county in New York and the number of deer-vehicle collisions in the county.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=376529

outliner

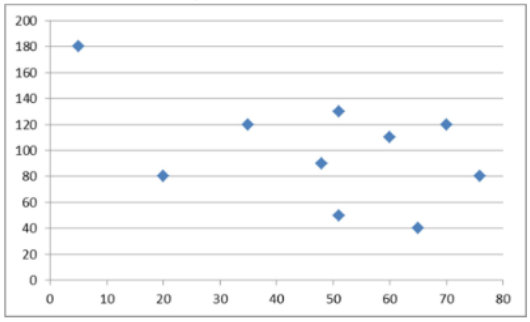
deer population	340	350	480	510	515	600	650	700	760
collisions	150	60	90	130	120	110	140	120	210

(a) Show this data in a scatterplot
(b) Draw a regression line and use the line to estimate the number of collisions when the deer population is 800. ≈ 190
(c) Are there possible outliers? w/all points r = .58, remove outliner, r ≈ .84

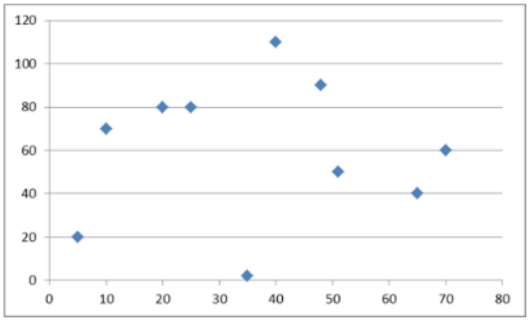


7. Estimate the correlation coefficient for the graphs below choosing from the values $r = 0, \pm 0.5, \pm 0.9$

r = -.5



r = 0



8. A biologist draws a sample of 200 fish from a lake to test for mercury levels. She finds that 10 have levels above limits set as healthy.
(a) What is the population? – all the fish in the lake
(b) What is the sample proportion?

$\hat{p} = 10/200 \rightarrow p$

not everyone is in the phone book.

9. To estimate the mean income of all residents in a town, a sample of people chosen from the telephone directory is surveyed and the mean is found to be \$43,000. The actual mean income in the town is \$40,000. This difference is most likely an example of sampling bias or sampling variability?

10. There is a population of 500 people. We need to choose a simple random sample of 5 for a survey. Decide how to assign numbers these people. Use the table of random digits below starting at line 103 to choose the sample. 001, 002, ... 500 // 500, 200, 089, 008, 192

101	01033	08705	42934	79257	89138	21506	26797
102	49105	00755	39242	50772	44036	54518	56865
103	61589	35486	59500	20060	89769	54870	75586
104	08900	87788	73717	19287	69954	45917	80026
105	75029	51052	25648	02523	84300	83093	39852
106	91276	88988	12439	73741	30492	19280	41255

11. A poll asked 956 licensed drivers whether they supported a nationwide lowering of the drunk driving limit to 0.08% BAC (blood alcohol content), and 72% said they did. Estimate a 95% confidence interval for actual percentage among all licensed drivers.

$\hat{p} = .72 \rightarrow p = .72, n = 956$

$SD = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(.72)(1-.72)}{956}} = 0.0145$

$.72 \pm 2(.0145) \Rightarrow .691$ to $.749$ or 69.1% to 74.9%

Clicker
Feel OK about the exam?
(A) Yes (B) Sort of (C) NO