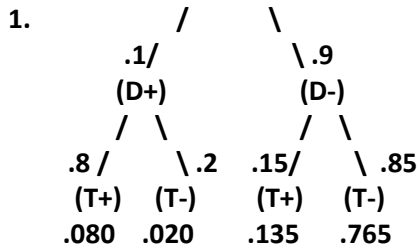


Practice Exam 1 - Answers



1b) $.080 + .135 = .215$

1c) $.080 / .215 = .372$

2. The data shown in the scatter plot is the distance traveled and the airfare for 12 flights on Delta Airlines:
 a. Which of the following is a reasonable estimate of the correlation coefficient? (Circle one answer)

- 1.0 **(0.8)** 0.0 -0.5

b. What does this graph tell us about distance and airfare

As distance goes up price goes up.

c. What is the type and level of distance traveled?

- Type (Circle One) Categorical Discrete **(Continuous)**
 Level (Circle One) Nominal Ordinal Interval **(Ratio)**

3. You have a 70% chance of being on time to class today and a 80% chance of being on time to class tomorrow. Assume these two days are independent events.

- a. Find the probability of being on time to class both today **and** tomorrow.
 $0.7 \times 0.8 = 0.56$
- b. Find the probability of being on time to class at least once today **or** tomorrow.
 $0.7 + 0.8 - 0.56 = 0.94$

4. The following data represent the daily births at a hospital for 20 days

14	15	15	17	17	19	19	21	23	25
26	27	31	35	36	47	48	59	70	99

a) Construct a stem and leaf diagram of the data

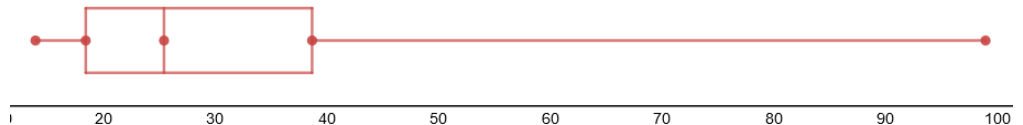
```

1 | 4 5 5 7 7 9 9
2 | 1 3 5 6 7
3 | 1 5 6
4 | 7 8
5 | 9
6 |
7 | 0
8 |
9 | 9
    
```

b) Calculate the interquartile range for this data set.
 $Q1 = 18, Q3 = 41.5 \quad IQR = 41.5 - 18 = 23.5$

c) Calculate the median for this data set.
Median = 25.5

d) Make a box plot for the data

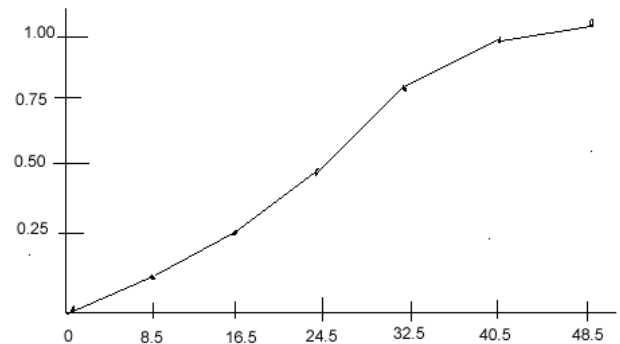
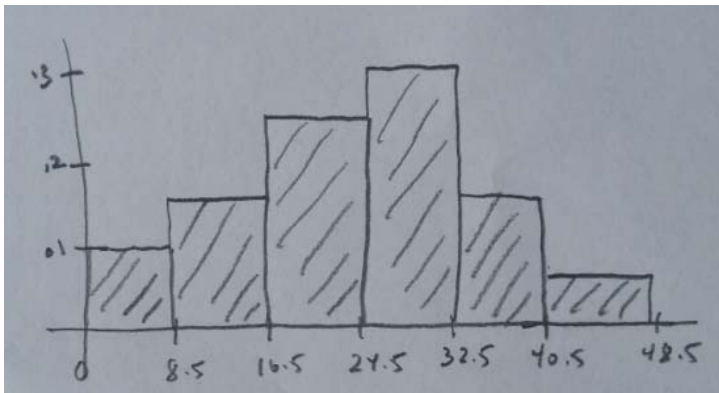


- e) **Right fence = $41.5 + 1.5(23.5) = 76.75$ so 99 is an outlier**
- f) **Without calculating**, what can you say about the mean births for this Hospital.(check one answer below)?
- The mean is greater than the median. Data is skewed right**
 - The mean is less than the median.
 - The mean is about the same as the median.
 - None of the above — no way to know without calculating.

5. The following data represents the hours per week worked outside of school by 200 randomly selected night students at a community college:

Hours	Frequency	Relative Freq	C.R.Freq
0-8	20	0.10	0.10
9-16	30	0.15	0.25
17-24	50	0.25	0.50
25-32	60	0.30	0.80
33-40	30	0.15	0.95
41-48	10	0.05	1.00
Total	200	1.00	

- a) In the space above, determine the relative frequencies and cumulative relative frequencies.
- b) Sketch a relative frequency **histogram**, showing **all horizontal and vertical labels**.
- c) Sketch a cumulative relative frequency **ogive**, showing **all horizontal and vertical labels**.



- d) Estimate the median from the graph.
median = 24.5
- e) What percentage of the night students work 32 hours per week or less?
0.80 (crf)

Without calculating but explaining your reasoning, which of the following is a reasonable estimate for the standard deviation? a) 0.5 b) 1 c) **10** d) 50

10 is the only answer that makes sense since the range is 48 and the range is between 4s and 6s.

6. Determine if each of the following data are categorical, continuous or discrete (circle one for each)

- | | | | |
|--|------------------------|-----------------------|---------------------|
| a. Number of fatalities from a tsunami: | categorical | continuous | discrete |
| b. Time spent in traffic: | categorical | continuous | discrete |
| c. Number of Songs on your I-pod: | categorical | continuous | discrete |
| d. Your student number | categorical | continuous | discrete |
| e. Names of cities in California with a Walmart: | categorical | continuous | discrete |
| f. Price per gallon of gasoline: | categorical | continuous | discrete |
| g. Number of Courses taken in a year. | categorical | continuous | discrete |
| h. Tons of steel used by a manufacturer: | categorical | continuous | discrete |

7. 1000 students (500 morning, 300 afternoon, 200 night) were asked how often they use the campus library. The results are summarized in the table below:

	<i>Never uses library</i>	<i>Sometimes uses library</i>	<i>Frequently uses library</i>	Total
<i>Morning</i>	200	250	150	600
<i>Afternoon</i>	80	145	75	300
<i>Night</i>	80	10	10	100
Total	360	405	235	1000

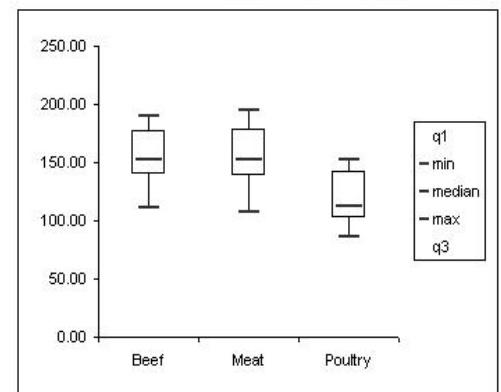
- a. Find the following probabilities:
- i) A randomly selected student never uses the library.
360/1000=.360
 - ii) A randomly selected student is a night student and frequently uses the library.
10/1000=.010
 - iii) **Given** the student is an afternoon student, the student never uses the library.
80/300=.267
- b. Are “Afternoon Student” and “Never uses library” Independent Events? **Justify and explain your answer.**
No. $P(\text{Never}) \neq P(\text{Never}|\text{Afternoon})$ ($0.360 \neq 0.267$)
- c. Would the probabilities generated from this data be classical, empirical or subjective probability?
Empirical – based on data

8. These descriptive statistics and boxplots were generated from data representing **calories per serving** for three types of hotdogs: All **Beef**, Mixed **Meat** and **Poultry**.

Descriptive statistics

	<i>Beef</i>	<i>Meat</i>	<i>Poultry</i>
count	20	17	17
mean	156.85	158.71	118.76
sample variance	512.66	636.85	508.57
sample standard deviation	22.64	25.24	22.55
minimum	111	107	86
maximum	190	195	152
range	79	88	66
median	152.5	153	113
mode	149	#N/A	102

- a. Compare the mean to the median calories for the **Meat** group. Is the result consistent with the shape of the box plot? Explain your answer.
Mean > median, but it is difficult to read the skewness from graph so its unclear.
- b. If the data is approximately bell shaped, between what two values of calories would you expect to find about 95% of the **Beef** data?
(111.57, 202.13)
- c. Which of the three groups has the most variability in calories per serving? Explain your answer.
Meat – highest Standard deviation
- d. Hebrew National All Beef Hotdogs had 190 calories per serving. Calculate and interpret the z-score for Hebrew National Hotdogs using the **Beef** Category data.
Z=1.46 Hebrew national calories are above average.
- e. Determine the probability a randomly selected **Poultry** Hot Dog exceeds 113 calories.
50% (half the data is above the median)
- f. Compare the three groups and draw at least two conclusions from the results.
Chicken dogs are lower in calories. Meat and Beef are about the same. (other answers ok).



9. From samples of a total of 2100 young (18-24 year old) White, Black and Latino men taken in January 2010 in the U.S., the unemployment rate of each sample was determined as given in the following table. (2013, *Urban Institute, The Labor Market Performance of Young Black Men in the Great Recession*). The study used stratified sampling. The Urban Institute concluded that young black men have a higher unemployment during the recession than their white and Latino peers.

Race/Ethnicity	Unemployment Rate
White	15.6%
Black	30.0%
Hispanic	26.9%

- a. What is the population and what is the sample?

Sample: 2100 young (18-24) year old men in the U.S.

Population: All young (18-24) year old men in the U.S.

- b. Identify the steps of the statistical process:

- Ask a question that can be answered with sample data.

Is there a difference in unemployment rates of young men due to race/ethnicity?

- Determine the information needed.

Employment Status and Race/ethnicity

- Collect sample data that is representative of the population.

Stratified Sampling will produce a representative sample

- Summarize, interpret and analyze the sample data.

The tabled data shows young male unemployment rate for Black at 30%, Hispanic at 26.9% and White at 15.6%

- State the results and conclusion of the study.

Young black men have a higher unemployment during the recession than their white and Latino peers.

10. A study was conducted to examine the effects of active recovery (AR), massage (MR), and cold water immersion (CR) on performance of repeated bouts of high-intensity cycling separated by 24 hours. A sample of physically active men aged 18–30 were randomly assigned to one of four groups. Each group performed an intense 18-minute cycling workout after which each underwent a 15-minute recovery period. In the 15 minutes, the first group (AR) continued to cycle at a low level, the second group (MR) received leg massage, the third group (CR) immersed their legs in a bath of cold water. The last group simply sat and rested. The next day the subjects did the same intense 18-minute cycling workout. Each exercise was done on a cycle ergometer so that the work level (measure in kilojoules) was calculated for each. The researchers found that on the second day, that there was no difference in the performance level of the subjects in the AR, MR and CR, but that the subjects who just sat in a chair to rest did less work than the other groups. (*Journal of Strength and Conditioning Research (2004; 18 [4], 855-60*).

- a. What is the explanatory variable?

The recovery methods: AR, MR, CR or rest

- b. What is the response variable?

The difference in work done by the cyclists between the first and second days.

- c. Which groups are the treatment groups?

AR, MR, CR

- d. Is there a control group? If so, which one?

Yes, rest only group.

- e. Is there blinding in this experiment? Explain your answer.

Not possible, since participants know what recovery method they are receiving.

11. 70% of students at a large New York University receive some financial aid. (use binomial table $n=4$, $p=.7$)

- If 4 students are randomly selected, determine the probability that **exactly 2** students in the sample receive some financial aid. $P(X=2) = .265$
- If 4 students are randomly selected, determine the probability that **less than 2** students in the sample receive some financial aid. $P(X<2) = P(0) + P(1) = .084$

12. The random variable X follows the probability distribution function as shown to the right:

x	P(x)
0	0.1
1	0.3
2	0.2
3	
4	0.1

- Determine $P(X=3)$ $P(3)=0.3$
- Determine the population mean. $\mu=2$
- Determine the population variance $\sigma^2=1.4$

13. 40% of students at a college use the cafeteria.

- If 9 students are randomly sampled, determine the probability that less than 3 use the cafeteria.
0.231
- If 9 students are randomly sampled and X represents the number of students in the sample who use the cafeteria, find the **mean** and **standard deviation** of X.
 $\mu=3.6$ students $\sigma=1.469$

14. Deleted

15. Find the 30th percentile for the cooking time for oatmeal which follows a Normal Distribution with a mean of 4 and a standard deviation of 3

30th percentile for Z is -0.52

30th percentile for X is $4 - 0.52(3) = 2.44$ minutes

16. Students' exam scores for a course follow a Normal Distribution with $\mu=70$ and $\sigma=10$.

- Find the probability a randomly selected student scores a **75 or more**.
 $P(X > 75) = P[Z > (75-70)/10] = P(Z > 0.5) = 0.3085$
- Find the exam score which is the **25th percentile** of this distribution.: **$70 - 0.67(10) = 63.3$**
- You take a random sample of 40 students. Find the probability the **sample mean** is between 68 and 72.
 $P(68 < \bar{X} < 72) = P\left(\frac{68-70}{10/\sqrt{40}} < Z < \frac{72-70}{10/\sqrt{40}}\right) = P(-1.26 < Z < 1.26) = 0.7924$
- Would your answer for part c be different if the probability distribution of "exam scores" did not follow a Normal distribution? **Explain your answer.**
NO, because of the CLT.

17. The age of a grove of walnut trees follow a Normal Distribution with $\mu=50$ years and $\sigma=15$ years.

- Find the probability that the age of a randomly selected tree is between 40 and 70 years.
 $P(40 < X < 70) = P\left(\frac{40-50}{15} < Z < \frac{70-50}{15}\right) = P(-0.67 < Z < 1.33) = 0.6568$
- Find the probability of a randomly selected tree has lived exactly 45.231789 years. **0, makes no sense**
- Find the 30th percentile of this distribution. **$50 - 0.52(15) = 42.2$ years**

18. 35% of students at De Anza College plan to transfer to San Jose State. 200 students are randomly selected and the sample proportion \hat{p} will be calculated.

- a. Determine the expected value and standard deviation of the sample proportion.

$$\mu_{\hat{p}} = p = 0.35 \quad \sigma_{\hat{p}} = \sqrt{\frac{(0.35)(1-0.35)}{200}} = 0.0337$$

- b. Determine that the condition for normality is satisfied.

$$np = .35(200) = 70 \quad n(1-p) = .65(200) = 130$$

Both values are at least 10, so condition for normality is satisfied.

- c. Determine the probability the sample proportion exceeds 0.40.

$$\begin{aligned} P(\hat{p} > 0.40) &= P\left(Z > \frac{0.40 - .035}{\sqrt{\frac{(0.35)(1-0.35)}{200}}}\right) \\ &= P(Z > 1.48) = 1 - 0.9306 = 0.0694 \end{aligned}$$