

STP226 Test1 Review Problems, Chapters 1-3, 5-6

1. Use the following fragment of Random Numbers Table to select a Simple Random Sample of 5 individuals out of 172 (numbered 1 to 172). Start at the beginning, go right. List numbers of the individuals selected.

709148039248002215019708142451

Sample is: _____

2. The following table gives information that was collected about a group of infants born June 1st 2010 in certain hospital in Phoenix. Identify all the variables in each column of the following table as categorical (qualitative) or quantitative, if variable is quantitative, decide if it is discrete or continuous.

Baby:	weight (kg)	Gender	Blood type	Eye color	Number of children in the family	Length of pregnancy (weeks)
John	3.6	M	A	blue	0	41
Kasey	4.1	F	O	blue	3	40.5
Loren	4.8	F	B	green	2	42
Kathy	3.2	F	B	brown	1	39.1
Joseph	5.0	M	AB	black	2	42

Qualitative: _____

Quantitative Discrete: _____

Quantitative Continuous: _____

3. The relative frequency table below represents the data on cholesterol levels of a group of patients. Use the table to answer the following questions:
 a. What percentage of the patients have levels of cholesterol of 210 or higher?
 b. Given that the number of patients is 80, how many have levels of cholesterol under 215?

Cholesterol Level	Relative Frequency
195- under 200	.05
200- under 205	.10
205- under 210	.25
210- under 215	.20
215- under 220	.25
220- under 225	.15

a) _____

b) _____

4. Use the following stem & leaf diagram to answer the following questions

```

0 | 5 9
1 | 3 8 8 7 7
2 | 1 3 4 7 8 9 9
3 | 2 3 3 3 4 4 5 6 7 7
4 | 2 3 4 4 5 5 6
5 | 1 2 5 7
6 | 0 3
7 |
8 | 0
9 | 3
10 | 5
    
```

Data : Hair Mercury Content of
Seychelles Fishermen

Stems: tens

leaves: ones

a. Which of the following best describes the rough shape of the distribution?

- i) right skewed (ii) bell shaped (iii) bimodal (iv) uniform (v) left skewed

b. Give Minimum and Maximum values of these data.

5. A car salesperson keeps track of number of cars she sells per week. The number of cars she sold per week in the last 20 weeks of 2009 are as follows:

```

5    0    2    1    2    5    2    4    2    3
1    1    1    1    2    1    3    0    3    4
    
```

Construct a grouped-data table for number of sales per week. Use classes based on single value. Compute frequency and relative frequency for each class. Use 2 decimal places for relative frequency

6. Suppose 11 people participated in a weight loss program and their weight loss was (in pounds)

15, 15, 17, 17, 17, 18, 18, 19, 19, 21, 22

Compute sample standard deviation, use definition and show all work, give units of SD.

7. Suppose a sample of dogs of certain breed has mean of weight of 70 pounds and standard deviation of 5 pounds. Is it correct that using Three Standard Deviations Rule we can say that relatively few dogs in that sample are expected to have a weight less than 55 pounds or more than 85 pounds? Circle one answer below.

Correct

Incorrect

Not enough information

8. Consider the stem-and-leaf diagram given below representing ages for a sample of people participating in a rally against U.S military intervention in Syria.

```

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

stems: tens

leaves: ones

Give observation #4, median and range for that data.

9. List the possible samples (without replacement) of two persons from the population:
Rachel Jason Mark Adam Anna

Describe a method for obtaining a S.R.S of two persons. What are the chances that any particular sample of two persons will be the one selected?

10. Make a stem-and-leaf diagram for the following TV viewing hours data set.

25 41 27 32 43 66 35 31 15 5
34 26 32 38 16 30 38 30 20 21

11. A researcher wants to use a simple random sampling to select 100 ASU students for the survey to measure the life satisfaction level of the entire college student body as ASU. Which one is correct?

- a. To save a time, he can go to a big classroom and distribute 100 survey sheets.
- b. Simple random sampling is convenient even though it causes a bias in sample selection.
- c. Only 100 students cannot reflect as closely as possible the characteristics of the entire students.
- d. Instead of a random-number table, he uses a random-number generated in the software package.

12. Following table gives relative frequency distribution for cholesterol levels of a group of patients.

Cholesterol Level	Relative Frequency
195-under 200	.05
200- under 205	.10
205- under 210	.25
210- under 215	.20
215- under 220	.25
220- under 225	.15

For a randomly selected patient compute following probabilities:

A)P(cholesterol level is at least 205)

B)P(cholesterol level is 200- under 220)

C)P(cholesterol level is under 205 or greater than 215)

13. In questions A-D use $N(0, 1)$ curve, include appropriate sketch for every question. Use tables, show all work

- A). Find area between -1 and 1.
- B). Find area right of 2
- C). Find $Z_{0.2}$.
- D). Find third quartile of $N(0,1)$

14. To describe a data set, which plot or chart is most effective?

Table. Relative-frequency distribution for the political party affiliation in group A

Party	Relative frequency
Democratic	0.34
Republican	0.31
Independent	0.29
Others	0.06
	1

- a. Pie chart b. Bar chart c. Histogram d. Stem-and-Leaf plot

15. X is a variable of height in a sample of group B. Z-score of observation x_1 is -0.5 and z-score of observation x_2 is -1.5. Which one is correct?

- a. x_1 and x_2 has different sample mean.
b. x_1 and x_2 has different sample variance.
c. x_1 and x_2 have values less than the sample mean.
d. x_1 and x_2 are located to the right side of the sample mean.

16. For the experiment of rolling a dice, ANNA defines A as an event the number showing is seven. Which one is correct?

- a. $P(A) = 0$
b. $P(A) = 1$
c. We cannot obtain $P(A)$ since “the number showing is seven” is not an event.

17. In the National Household Survey that was run in Canada in 2011, several questions were asked. Here is a sampling of six questions from the survey:

Q1: What number of hours did you work in the week of May 1 - 7, 2011?

Q2: What is your marital status?

Q3: How much money did you spend on childcare during 2010?

Q4: What is your phone number?

Q5: Do you have any difficulty hearing, seeing or communicating?

Q6: How many minutes does it take you to get to work?

Which questions would have responses that are numeric/quantitative variables?

(A) Q1 and Q3.

(B) Q1, Q3 and Q6.

(C) Q2, Q4 and Q5.

(D) All the responses are numeric/quantitative variables.

18. Using the following random numbers, select a simple random sample of 3 out of 25 subjects.
6094252662172618477433

19. A random sample of cruise ships leaving from the Port of New York showed 5 different destinations: Bahamas, Bermuda, Southampton, Mediterranean, and Caribbean. The frequency of the data is given as follows:

Class	Frequency
Bahamas	2
Bermuda	4
Caribbean	6
Mediterranean	3
Southampton	10

A) Make a bar chart for this data set.

B) In making a pie chart, approximately how many degrees “slice” will correspond to Southampton?

20. Use the values of the sample mean and the sample median to determine whether the distribution is most likely symmetric, skewed to the left, or skewed to the right

- a) Mean=64, Median=63.8
- b) Mean=30, Median=70
- c) Mean=40, Median=20

21. Which of the following measures are robust. Select all that applies.

- b. Mean
- c. Median
- d. IQR
- e. Standard deviation

22. The following data are the incomes (in thousands of dollars) for a sample of 12 households.

35, 29, 44, 72, 34, 64, 41, 50, 54, 104, 39, 58

a) Find the five-number summary. b) Check if there are any outliers. c) sketch a box plot.

23. Below are the midterm scores of a population of 4 students. Find the population mean and standard deviation. Use proper symbols. 82, 95, 67, 92

24. Most college career counselors agree that starting salary is associated with academic major. Even if a person’s first job is not directly related to their course of study, their salary might still be related to their academic major. A recent survey of academic major, and starting salary of graduates showed the following information.

Major	Mean	Standard Deviation
English	\$32,300	\$1,175
Computer Science	\$42,500	\$2,375

A Computer Science major who responded to the survey received a starting salary of \$45,000, and an English major received an offer of \$35,000. Which salary is better, in terms of statistics?

29) Table below gives number of licenced drivers in the USA in 2006

Age	Licenced Drivers (millions)	Relative frequency
16-19	9.2	0.051
20-29	33.6	0.187
30-39	40.8	0.228
40-49	37.0	0.206
50-59	24.2	0.135
60-69	17.5	0.098
70-79	12.7	0.071
80-89	4.3	0.024
totals	179.3	

For randomly selected licenced driver compute following probabilities:

P(driver is below 60)

P(driver is below 60 **and** over 49)

P(driver is over 69 **or** less than 30)

P(driver is **not** below 20)

30) Following table gives distribution of blood type for a random sample of 200 Americans.

Blood Type

	O	A	B	AB	totals
RH positive	78	70	16	8	172
RH negative	12	10	4	2	28
totals	90	80	20	10	200

For randomly selected American compute following probabilities:

P(B and RH negative)

P(A)

P(RH positive)

P(not AB)

P(O or RH negative)

P(A or B)

31) Following table gives distribution of degrees earned by gender for Americans in 2001-2002

	Bachelor's	Master's	Professional	Doctorate	totals
Female	0.390	0.137	0.019	0.011	0.557
Male	0.305	0.098	0.024	0.016	0.443
totals	0.695	0.235	0.043	0.027	1

For randomly selected degree compute following probabilities:

P(degree was awarded to a Female)

P(that was a Bachelor's degree)

P(that was a Bachelor's degree **and** was awarded to a Male)

P(that was a Doctorate **or** was awarded to a Female)

P(degree was Master's **or** Doctorate)

32. Survey of Tempe High School students established following probabilities for number of times randomly selected student eats his/her lunch outside of the school during a week.

Number of times eating out	0	1	2	3	4	5
probability	.05	.10	.25	.30	.20	.15

Could that be a valid probability assignment? Explain why or why not.

33. A frequency distribution for the number of cars owned by each of the **6400** families in a small city is shown below.

Cars owned	0	1	2	3	4	5
Number of families=f	30	1425	2865	1700	324	56

For a family selected at random, let

A= event that the family owns at most 4 cars and B= event that the family owns at least 3 cars

- Compute the probability of the event (AandB).
- Are events A, B mutually exclusive? Explain why or why not.
- Compute the probability of the event (not A)

34 Family has 3 children

- Write down the sample space (use G=girl, B=boy) (Hint: there are 8 possible outcomes, for example one of them is G B G)
- Let A= Girl is first. Compute the probability of the event notA

35. Suppose we roll a balanced die two times, compute the probability that sum of both rolls is not 9.

36. Use $N(0, 1)$ curve, include appropriate sketch for every question.
Use tables, show all work

- Find area between -1.42 and 2.15 .
- Find area right of -1.47
- Find area between 2.25 and 3.15
- Find z-score with 38% area to the left of it
- Find $z_{0.32}$
- Find first quartile of $N(0,1)$
- Find fourth decile of $N(0,1)$
- Find 85th percentile of $N(0,1)$

37. A Classify each of the following attributes as either categorical (qualitative) or numerical (quantitative). For those that are numerical, determine whether they are discrete or continuous.

B. What type of a graphical display is appropriate for these types of data. Select from the following:

- histogram
 - stemplot
 - bar chart
 - box plot
 - pie chart
- Brand of a computer purchased by the customer
 - State of birth for someone born in the United States
 - Number of students in a class of 35 that turned in a term paper before the due date
 - Concentration of a contaminant (micrograms per cubic centimeter) in a water sample
 - Zip code (Think carefully about this one)
 - Actual weight of coffee in a 1-pound can.
 - Price of a textbook.

38. Researcher s analyzed standardized test results and television viewing habits of 1700 children. They found that children who averaged more than 2 hours of TV viewing per day tended to score lower on measures of reading ability and short term memory.

A. Is the study observational or an experiment?

B. Is it reasonable to conclude that watching more than two hours of TV per day is the cause of lower reading scores?

39. A sample consisting of four pieces of luggage was selected from among those checked at an airline counter, yielding the following data on x =weight (in pounds):

$$x_1=33.5 \quad x_2=27.3 \quad x_3=36.7 \quad x_4=30.5$$

Suppose that one more piece is selected, denote its weight by x_5 . Find a value of x_5 such that \bar{x} =median

40. Consider babies born in the “normal” range of 37-43 weeks gestational age. Extensive data supports the assumption that for such babies born in the United States, birth weight is normally distributed with mean 3432 g and standard deviation 482 g.

a) What is the probability that the birth weight of a randomly selected baby of this type exceeds 4000 g?

b) What is the probability that the birth weight of a randomly selected baby of this type exceeds is between 3000 and 4000 g?

c) a) What is the probability that the birth weight of a randomly selected baby of this type is either less than 2000 g or greater than 5000 g?

d) What is the weight of a baby of this type that is the 90th percentile of the distribution?

Key

1. sample is : 148, 39, 2, 19, 142

2. Qualitative: Gender, Blood type, Eye color

Quantitative Discrete: Number of children in the family

Quantitative Continuous: weight , Length of pregnancy

3. a) 60% b) 48

4. a) right skewed b) min=5, max=105

5.

#	f	f/n
0	2	.10
1	6	.30
2	5	.25
3	3	.15
4	2	.10
5	2	.10

20 1.0

6. $\bar{x} = \frac{198}{11} = 18$ $s = \sqrt{\frac{48}{10}} = 2.19 lb$ $(x - \bar{x})$: -3 -3 -1 -1 -1 0 0 1 1 3 4

$(x - \bar{x})^2$: 9 9 1 1 1 0 0 1 1 9 16 sum=48

23. $\mu=84$ $\sigma=10.93$
24. English: $z=2.30$ Computer Science: $z=1.05$, English has higher salary compared to the peers.
25. a) 0.9544 b) 0.0027 c) 4.17 min d) 0.0968 e) 2.71 min
26. Observational study, researcher does not use experimental and placebo groups to which subjects have to be randomly assigned.
27. A) 17, 20, 21, 22, 24 B) T, F, F, T
28. $.34+.32+18$, $1-.18$, $.32$, $0.7+.09+.18$, $1-.34$
29. $.051+.187+.228+.206+.135$, $.135$, $.071+.024+.051+.187$, $1-.051$
30. $4/200$, $80/200$, $172/200$, $1-10/200$, $(90+28-12)/200$, $(80+20)/200$
31. $.557$, $.695$, $.305$, $.027+.557-.011$, $.235+.027$
32. Not Valid, sum of probabilities >1 , it should be $=1$
33. a) $P(A \text{ and } B) = (1700+324)/6400$
 b) Not mutually exclusive since $P(A \cap B) \neq 0$
 c) $P(\text{not } A) = 1 - (324+56)/6400$
34. a) $S = \{BBB, GGG, BGG, GBG, GGB, GBB, BGB, BBG\}$
 b) $\text{not } A = \text{event that first child is not a girl} = \{BBB, BGG, BGB, BBG\}$
 c) $P(\text{not } A) = 3/8$
35. $A = \text{event that sum is } 9 = \{(3,6), (4,5), (5,4), (6,3)\}$ $P(A) = 4/36$
36. Reffer to the class notes for appropriate sketches
- A) $.9842 - .0778 = .9064$, sketch bell shaped curve, 0 in the center, z-scores are on opposite sides of 0, you want area between
- B) $1 - .0708 = .9292$, sketch bell shaped curve, 0 in the center, z-score is left of 0, you want area left of it
- C) $.9992 - .9878 = .0114$, sketch bell shaped curve, 0 in the center, z-scores are on the right side of 0, you want area between
- D) $z = -.31$ (area = .3783), sketch bell shaped curve, 0 in the center, z-score are on the left side of 0, area left of it is .38, you want the z score
- E) $z = .47$ (area = .6808), sketch bell shaped curve, 0 in the center, z-score are on the right side of 0, area left of it is .68, you want the z score
- F) $z = -.67$ (area = .2514), sketch bell shaped curve, 0 in the center, z-score are on the left side of 0, area left of it is .25, you want the z score
- G) $z = -.25$ (area = .4013), sketch bell shaped curve, 0 in the center, z-score are on the left side of 0, area left of it is .40, you want the z score
- H) $z = 1.04$ (area = .8508), sketch bell shaped curve, 0 in the center, z-score are on the right side of 0, area left of it is .85, you want the z score

37.

A Qualitative: a,b,e B) (pie chart, bar chart)

Quantitative discrete: c,g B) (histogram, stemplot, box plot)

Quantitative continuous: d,f B) (histogram, stemplot, box plot)

38.

A) observational study

B) No, cause-and-effect conclusion can be made only based on the experiment, not observational study.

39. $x_5=32$

40.

a) 0.1193

b) 0.6956

c) 0.0021

d) 4049.71 g