60080079 Introduction to Statistical Methods
Semester 2 2023-2024
Homework Assignment 1

21 CST H3Art

1. Here are the first lines of a professor's data set at the end of a statistics course:

NAME	MAJOR	POINTS	GRADE
ADVANI, SURA	COMM	397	B
BARTON, DAVID	HIST	323	C
BROWN, ANNETTE	BIOL	446	A
CHIU, SUN	PSYC	405	B
AORTEZ, MARIA	PSYC	461	A

1.1 In this data set, __ are the individuals and __are the variables. Choose the appropriate answer for each blank from the following options.

1) The students
2) COMM, HIST, BIOL, and PSYC
3) Major, Points, and Grade

1.2 Among the variables, __ and __ are categorical, and __ is quantitative. Choose the appropriate answer for each blank from the following options.

1) Major
2) Points
3) Grade

Write your answer as a five-digit number.

Answer: 13132

2. The number of deaths among persons aged 15 to 24 years in the United Stets in 1999 due to the eight leading causes of death for this age group were: accidents, 13602; homicide, 4989; suicide, 3885; cancer, 1724; heart disease, 1048; congenital defect, 430; respiratory disease, 208; AIDS, 197.

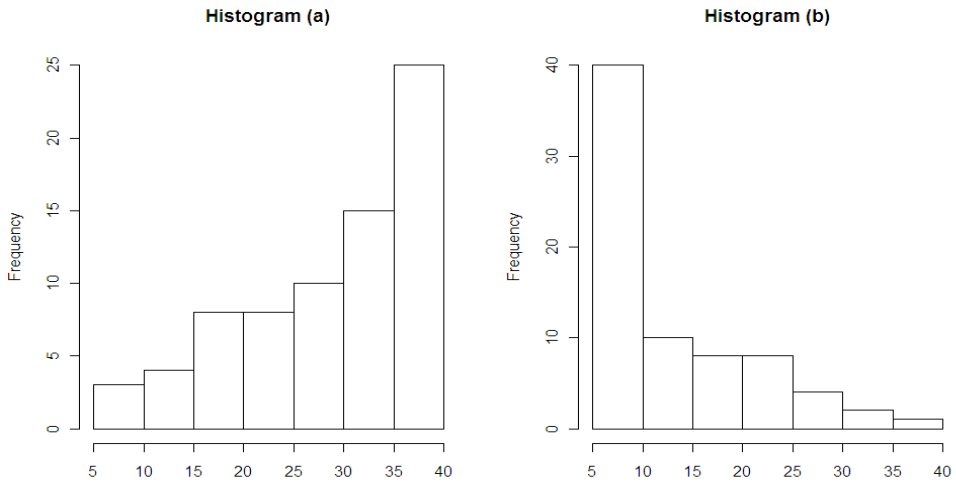
What additional information do you need to make a pie chart?

1) Total of the ages of the people died by the above causes.
2) The deaths from suicide and cancer.
3) No additional information is required.
4) Total deaths or deaths from all other causes.

Write your answer as a single-digit number.

Answer: 4

3. The follows are the histograms for two skewed data sets.



Histogram (a) is skewed to the __, and Histogram (b) is skewed to the __. Choose the appropriate answer for each blank from the following options.

1) right
2) left

Write your answer as a two-digit number.

Answer: 21

4. There is some evidence that increasing the amount of calcium in the diet can lower blood pressure. In a medical experiment one group of men was given a daily calcium supplement, while a control group received a placebo (a dummy pill). The seated systolic blood pressure of all the men was measured before the treatments began and again after 12 weeks. The blood pressure distributions in the two groups should have been similar at the beginning of the experiment. Here are the initial blood pressure readings for the two groups:

Calcium group

107	110	123	129	112	111	107	112	136	102

Placebo group

123	109	112	102	98	114	119	112	110	117	130

Make a stemplot for each group, and use your stemplots to answer the following questions.

4.1 Does your plot show any major differences of the two blood pressure distributions close together?

1) Yes
2) No

4.2 The center of the calcium group is between __, and the center of the placebo group is between __.

1) 90 and 100
2) 100 and 110
3) 110 and 115
4) 120 and 130

Write your answer as a three-digit number.

CALCIUM 茎叶图	PLACEBO 茎叶图
频率 Stem & 叶 3.00 10 . 277 4.00 11 . 0122 2.00 12 . 39 1.00 13 . 6 主干宽度： 10 每个叶： 1 个案	频率 Stem & 叶 1.00 9 . 8 2.00 10 . 29 6.00 11 . 022479 1.00 12 . 3 1.00 13 . 0 主干宽度： 10 每个叶： 1 个案

Answer: 233

5. The follows are data on the total oil recovered from 30 wells.

7.1	21.3	34.6	46.4	64.9	97.7
10.3	24.9	35.1	50.4	69.5	103.1
12.1	29.1	37.7	53.2	79.5	118.2
17.6	31.4	42.7	58.8	81.1	156.5
18.5	32.9	44.5	61.4	92.2	181.2

5.1 The mean of the amounts recovered is 5_.13.

1) 5
2) 6
3) 7
4) 8
5) 9

5.2 The five-number summary of the data set is (7.1, 2_.1, 45.45, 79.5, 1_1.2).

1) 5
2) 6
3) 7
4) 8
5) 9

5.3 Based on the answers from the previous questions, explain how the relationship between the mean and the median reflects the shape of the distribution.

1) Since the mean is smaller, the distribution is skewed to the right.
2) Since the mean is larger, the distribution is skewed to the right.
3) Since the mean is larger, the distribution is skewed to the left.
4) Since the mean is smaller, the distribution is skewed to the left.

Write your answer as a four-digit number.

Answer: 3542

6. Here are the scores on the Survey of Study Habits and attitudes (SSHA) for 18 first-year college women:

154	109	137	115	152	140	154	178	101
103	126	126	137	165	165	129	200	148

The data contain **one** high outlier.

6.1 WITH the outlier, the mean of the data is 141.06, and the median is 1_8.5.
1) 1
2) 2
3) 3
4) 4

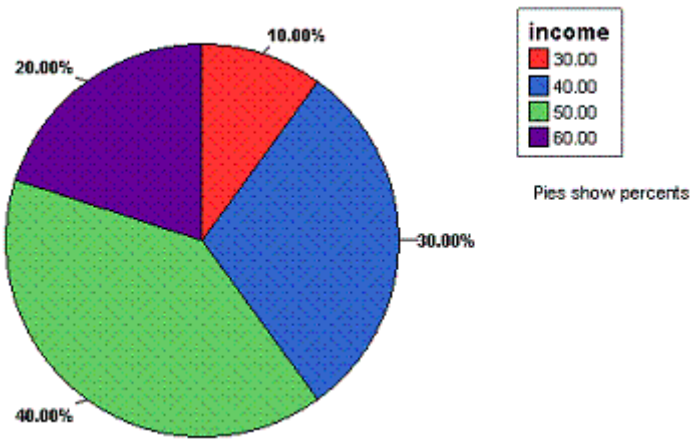
6.2 WITHOUT the outlier, the mean of the data is 1_7.59, and the median is 137.
1) 1
2) 2
3) 3
4) 4

6.3 The __ changed by a larger amount as a result of removing the outlier. This implies that __ is a more resistant measure of center.
1) mean
2) median

Write your answer as a four-digit number.

Answer: 3312

7. The following is the pie chart for income (in thousand).



7.1 How many people are under the category of 30 thousand dollars?

1) 10
2) 30
3) 50
4) There is not enough information.

7.2 If the sample size of the data is 50, how many people are there who make 50 thousand dollars or above per year?

1) 10
2) 30
3) 50
4) There is not enough information.

Write your answer as a two-digit number.

Answer: 42

8. The following is a stemplot displaying the percent of residents aged 25 to 34 in each of the 50 states, ranging from 10.8% to 15.9%.

10	8
11	0
12	1344677889
13	0012455566789999
14	11222344445789
15	24478999

8.1The five-number summary of this distribution is (10.8, 13, 13._, 1_.4, 15.9)

1) 3
2) 4
3) 7
4) 9

8.2 The IQR of this data set is

1) 1.2
2) 1.3
3) 1.4
4) 1.5

8.3 The lower and upper limits of the 1.5×IQR rule for determining outliers are 10.9 and 1_.5, respectively.

1) 6
2) 7
3) 8
4) 9

8.4 Base on the previous question, __ is the outlier of this data set.

1) 10.8

2) 15.9

Write your answer as a five-digit number.

Answer: 42311

9. The following table is the summary of 29 measurements of the density of the earth.

5.5	5.61	4.88	5.07	5.26	5.55	5.36	5.29	5.58	5.65
5.57	5.53	5.62	5.29	5.44	5.34	5.79	5.10	5.27	5.39
5.42	5.47	5.63	5.34	5.46	5.30	5.75	5.68	5.85	

The mean and SD of the data are 5.4_8 and 0.2_1, respectively.
1) 2
2) 3
3) 4
4) 5

Write your answer as a two-digit number.

Answer: 31

10. Using either the *Z* table or software, find the following answers.

10.1 $P(Z < -1.66)$ = __ and $P(Z > 1.66)$ = __.
1) 0.0548
2) 0.9515
3) 0.0485
4) 0.5000

10.2 Based on the answers of the above two questions, we can conclude that
1) $P(Z < -1.66) = P(Z > -1.66)$
2) $P(Z < -1.66) = P(Z > 1.66)$
3) $P(Z < -1.66) = 1 - P(Z > 1.66)$

Write your answer as a three-digit number.

Answer: 332

11. Suppose $Z \sim N(0, 1)$. What is the probability that Z is greater than 0.57?
1) 0.2843
2) 0.7157
3) 0.6485
4) 0.4253

Write your answer as a single-digit number.

Answer: 1

12. Suppose $Z \sim N(0, 1)$. Find the following probabilities.

12.1 $P(Z < 2.85)$ =
1) 0.9978
2) 0.0022
3) 0.7268
4) 1.0205

12.2 $P(Z > 2.85)$ =
1) 0.9978
2) 0.0022
3) 0.7268
4) -0.2252

12.3 $P(-1.66 < Z < 1.66)$ =
1) 0.0970
2) 0.9515
3) 0.0485
4) 0.9030

12.4 $P(-1.66 < Z < 2.85)$ =
1) 0.0507
2) 0.9978
3) 0.9493
4) 0.0970

Write your answer as a four-digit number.

Answer: 1243