14
17
10
14
55

STAT:2010/4200, Statistical Methods and Computing
Spring 2019, Instructor: Cowles
Midterm 1

Show your work on any problems that involve calculations.

Name: Solutions

1. A study examined how long aircraft air conditioning units operated after being repaired. Here are the operating times in hours for one unit:

min Q1

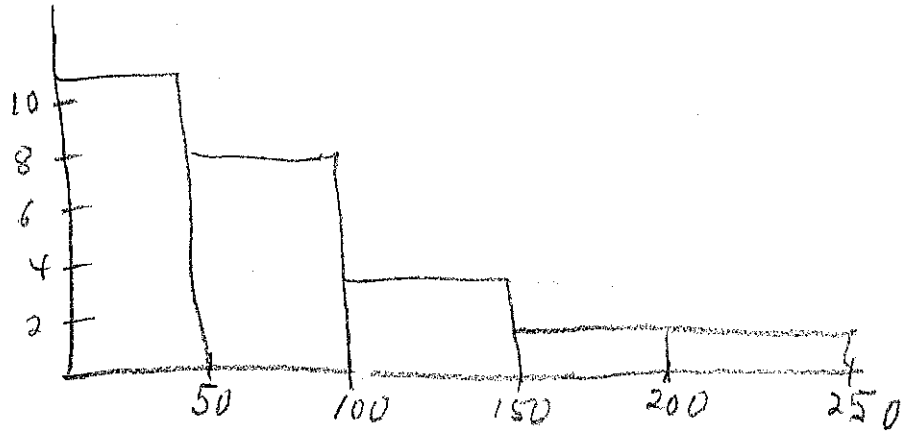
1 4 11 16 18 18 18 24 31 39 46 | 51 54 63 68 77 80 82 97 |
106 111 141 142 | 163 191 | 206 216

median

Q3 max

(a) Draw a histogram of these data, using bins that are 50 hours wide, starting with $0 \leq \text{time} < 50$, $50 \leq \text{time} < 100$, ...

5



(b) The overall shape of the distribution is (circle one):

i. roughly symmetric

2 ii. skewed right

iii. skewed left

iv. none of the above

(c) An appropriate numeric summary for these data is: (circle one)

i. mean and standard deviation

2 ii. five-number summary

(d) Produce the numeric summary that you selected in the previous question. (Numeric answer; show your work.)

5

max 216

Q3 111

med 63

Q1 18

min 1

median of observations > median

median of observations < median

1

2. The polling website Five Thirty-Eight makes the datasets used in some of their articles publicly available. I downloaded a dataset called `women-stem.csv` from this url: `https://github.com/fivethirtyeight/data/tree/master/college-majors`.

   For each major in the STEM disciplines, the dataset provides counts of recent graduates in the major as well as economic variables describing workers in the field. Two of the variables are:

`Median`	`Median earnings of full-time, year-round workers`
`ShareWomen`	`Percent of women among recent graduates with this major`

   Refer to the attached SAS output in answering the following questions.

   (a) In the scatterplot, the response variable is (circle one):
   i. ShareWomen
   ii. Median
   iii. impossible to determine

   (b) The relationship shown in the scatterplot is (circle one):
   i. positive
   ii. negative
   iii. impossible to determine

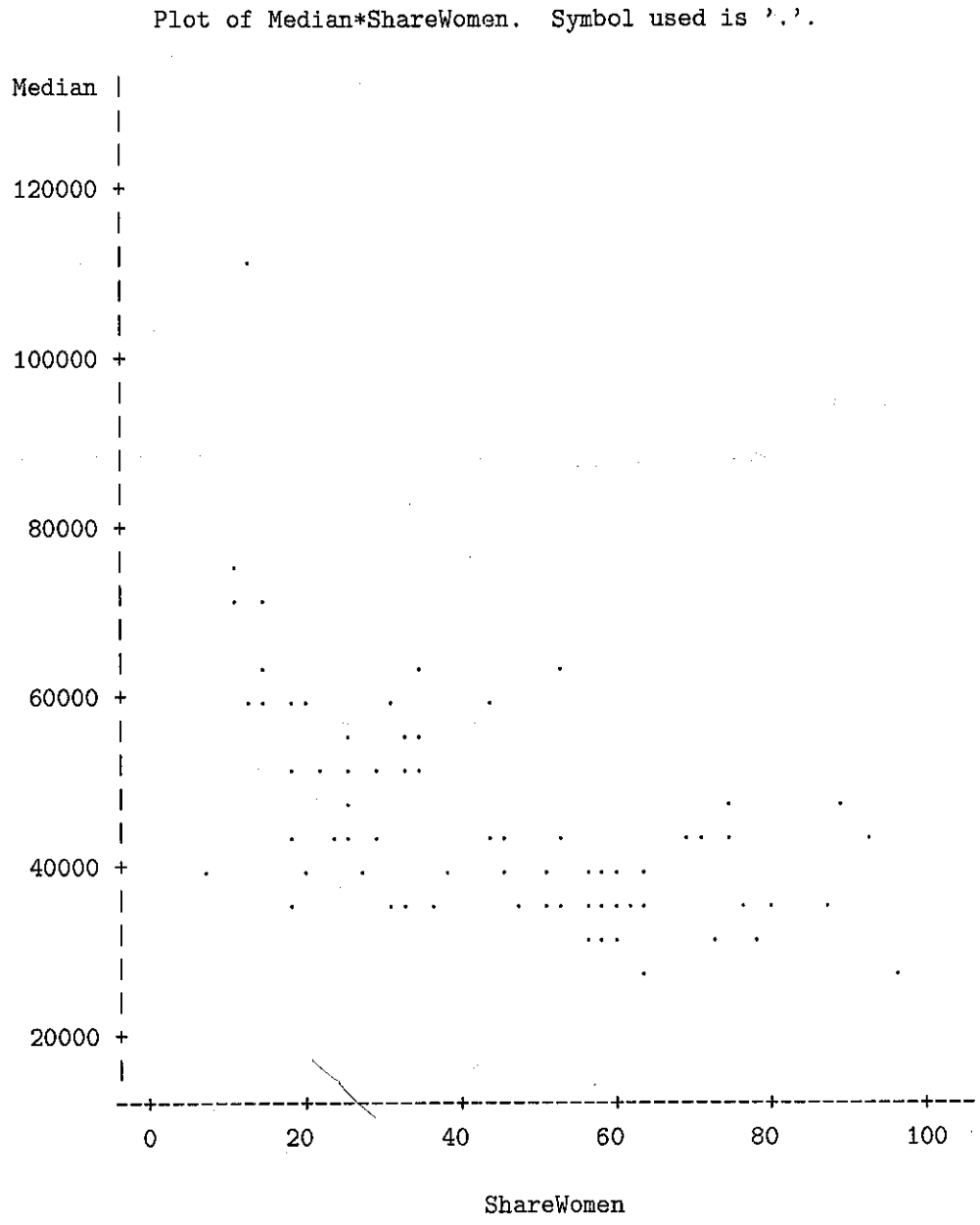
   (c) The relationship shown in the scatterplot is (circle one):
   i. roughly linear
   ii. not linear at all
   iii. impossible to determine

   (d) The relationship shown in the scatterplot is (circle one):
   i. strong
   ii. weak
   iii. there is no relationship

   (e) There is one outlier in the scatterplot. Is it likely to be an influential observation? Briefly explain why or why not.
   Yes. It is near the low end of values in the horizontal direction as well as being an outlier in vertical direction. Removing it would change the regression results a lot.

   (f) Write the regression equation (get numbers from the SAS output).
   $\widehat{\text{Median}} = 60576 - 330.9 \times \text{ShareWomen}$

   (g) Explain in words and numbers what the slope tells us about the relationship between the percent of women in a major and the median income in that field.
   For each 1 percentage point increase in the share of Women in a major, we expect on average $330.90 lower median income.



Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	60576	2651.38968	22.85	<.0001
ShareWomen	1	-330.90121	53.66443	-6.17	<.0001

3. A study in Cameroon found that the wing length of male finches of the species *Pyrenese ostrinus* follows a normal distribution with mean 61.2 mm and standard deviation 1.8 mm.

$\mu = 61.2$ $\sigma = 1.8$ mm

(a) What proportion of male finches have wings longer than 65mm? (Numeric answer; show your work.)

5

$$z = \frac{65 - 61.2}{1.8} = 2.11$$

4/5 if got .9826

$$1 - 0.9826 = 0.0174$$

(b) What is the wing length that only 2% of male finches exceed? (Numeric answer; show your work.)

5

$z = 2.06$

$$X = 61.2 + 2.06(1.8) = 64.9 \text{ mm}$$

4/5 if got $z = -2.06 \approx 2.88$

10

4. The muscle strength of elderly adults may be roughly measured by how strongly they are able to grip the hand of the evaluator. The values of the resulting grip-strength variable are "very strong," "strong," "average," "weak," "very weak."

   (a) The data type of this grip-strength variable is: (circle one)

   i. binary
   ii. nominal
   iii. ordinal
   iv. discrete quantitative
   v. continuous quantitative

   (b) Circle **all** of the numeric summaries that would be appropriate for this variable:

   i. frequency table
   ii. mean
   iii. median
   iv. mode
   v. standard deviation
   vi. IQR
   vii. range
   viii. none of the above

   (c) Circle **all** of the graphical representations that would be appropriate for this variable:

   i. histogram
   ii. bar chart
   iii. pie chart
   iv. stem plot
   v. line plot
   vi. none of the above