

STAT:2010/4200, Statistical Methods and Computing
 Spring 2017, Instructor: Cowles
 Midterm 1

Show your work on any problems that involve calculations.

Name: Shubhan
 Course no. (STAT:2010 or STAT:4200) _____

1. Here is a stem plot of the life expectancies at birth of 30 countries in Europe in 1962.

52 | 1
 53 |
 54 |
 55 |
 56 |
 57 |
 58 |
 59 |
 60 | 9
 61 |
 62 | 7
 63 |
 64 | 458
 65 |
 66 | 8
 67 | 16
 68 | 08
 69 | 22 3 9
 70 | 333358
 71 | 3
 72 | 4
 73 | 2457

(a) What is the median of these values? (Numeric answer)

(b) The shape of the distribution of these values is (circle the best answer):

- i. symmetric
- ii. skewed left
- iii. skewed right

(c) Which numeric summary would best describe the distribution of these data? (circle the best answer)

- i. mean and standard deviation
- ii. IQR and range
- iii. mode and variance
- iv. five number summary

3

1

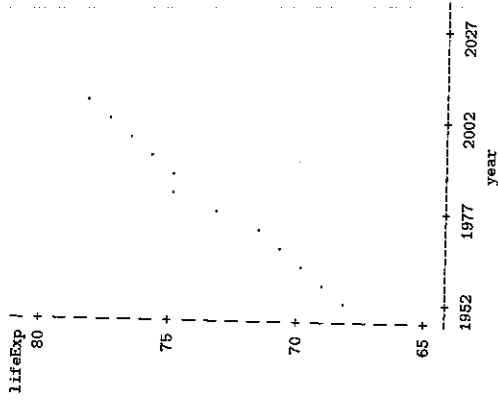
2

1

7

2. Here is a scatterplot of life expectancy at birth in the U.S. for the years 1952 - 2007.

Plot of lifeExp*year. Symbol used is ' . '



(a) The relationship between life expectancy and year is (circle one):

- i. roughly linear
- ii. non-linear
- iii. no relationship
- iv. impossible to tell

(b) The direction of association between life expectancy and year is (circle one):

- i. positive
- ii. negative
- iii. no association
- iv. impossible to tell

(c) The strength of the association between life expectancy and year is (circle one):

- i. very weak
- ii. fairly weak
- iii. no association
- iv. fairly strong
- v. very strong
- vi. impossible to tell

2

3. For the same data plotted in the previous question, here is output from fitting a linear regression model.

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-291.05421	13.78220	-21.12	<.0001
year	1	0.18415	0.00696	26.45	<.0001

(a) Which variable is the response variable in this regression? (circle one)

- i. life expectancy
- ii. year

(b) Write the regression equation that describes these data.

3 life Exp_i = -291.05 + 0.184 year

(c) Use the regression equation to predict U.S. life expectancy in the year 2000. (Numeric answer; show your work.)

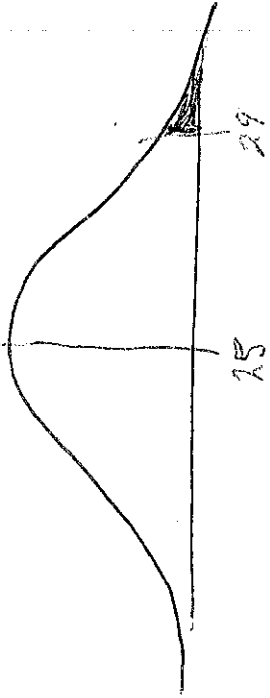
3 -291.05 + 0.184(2000) = 76.95

(d) The correct interpretation of the slope in this model is (circle the one best answer):

- i. For each one-year increase in year, we expect a 291-year decrease in life expectancy.
- ii. We expect life expectancy to increase by about 0.184 years each year.
- iii. We expect life expectancy to increase by about 13.8 years each year.
- iv. There is no linear relationship between life expectancy and year.
- v. None of the above.

(e) The coefficient of determination, R^2 , is likely to be closest to (circle one):

- i. -1.0
- ii. -0.5
- iii. 0.0
- iv. 0.5
- v. 1.0



4. The circumferences of American elm trees follow an approximately normal distribution with mean 25 feet and standard deviation 2.5 feet. What proportion of American elm trees have circumferences greater than 29 feet? (Numeric answer; show your work).

5 $Z = \frac{29 - 25}{2.5} = 1.6$

$Pr(Z \leq 1.6)$ from Table A = 0.9452

$1 - 0.9452 = 0.0548$

5. Briefly explain what is wrong with each of the following statements:

(a) There is a high correlation between Americans' incomes and whether or not they graduated from high school.

2 To use correlation, both variables must be quantitative. Whether graduated is binary.

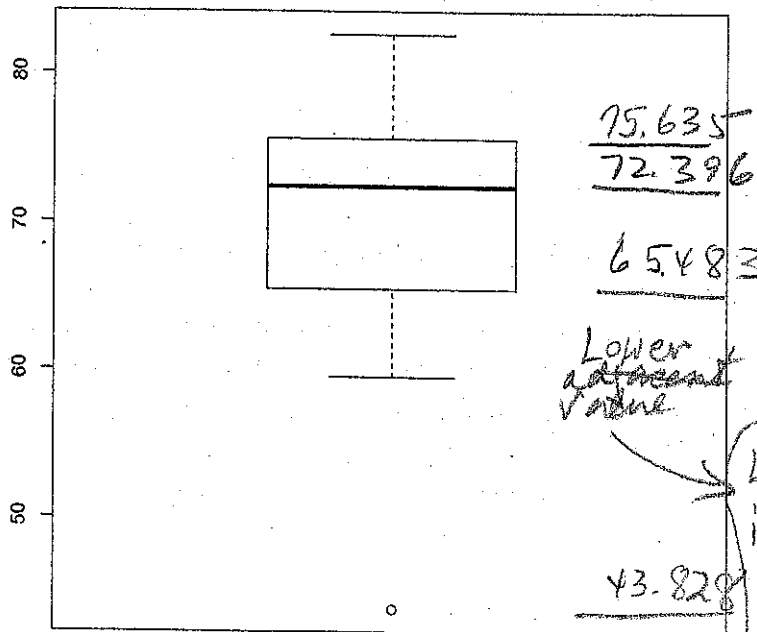
(b) The correlation between the scores on midterm 1 and the scores on midterm 2 was 0.79 points.

2 The sample correlation coefficient has no units.

4. Below is SAS output describing life expectancies for countries in Asia in 2007:

Level	Quantile
100% Max	82.603
99%	82.603
95%	82.208
90%	79.972
75% Q3	75.635
50% Median	72.396
25% Q1	65.483
10%	62.069
5%	59.545
1%	43.828
0% Min	43.828

Here is a boxplot of the same data:



$$IQR = 75.635 - 65.483 = 10.152$$

Lower adjacent value is the lowest value in the data that isn't lower than

$$Q1 - 1.5 \times IQR = 65.483 - 1.5 \times 10.152 = 50.255$$

On each horizontal line beside the plot, indicate the numeric value represented by the corresponding location on the plot. Show your work for any needed calculations.

