

AP STATS
CHAPTER 1 PRACTICE TEST

Name Key
Date _____ Hour _____

Section I: Multiple Choice *Select the best answer for each question.*

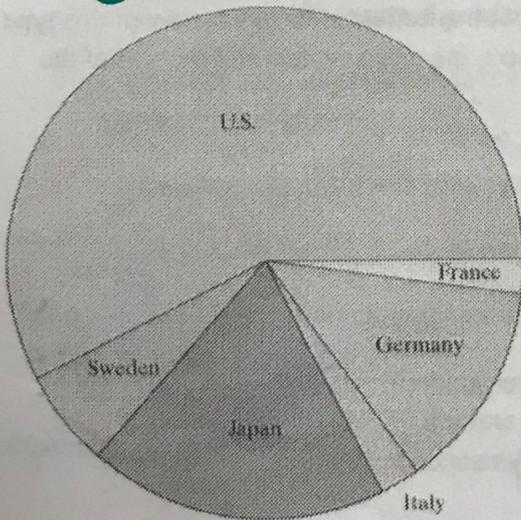
T1.1. You record the age, marital status, and earned income of a sample of 1463 women. The number and type of variables you have recorded is

- (a) 3 quantitative, 0 categorical
- (b) 4 quantitative, 0 categorical
- (c) 3 quantitative, 1 categorical
- **(d) 2 quantitative, 1 categorical**
- (e) 2 quantitative, 2 categorical

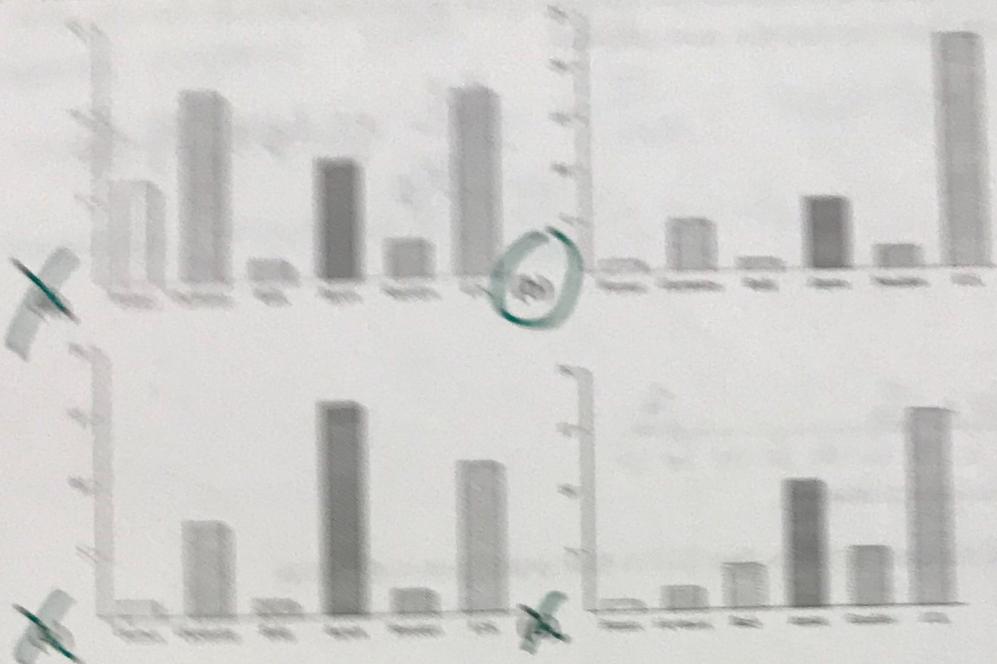
Categorical	Quantitative
- marital status ①	- age ② - income

T1.2. Consumers Union measured the gas mileage in miles per gallon of 38 vehicles from the same model year on a special test track. The pie chart provides information about the country of manufacture of the model cars tested by Consumers Union. Based on the pie chart, we conclude

- (a) Japanese cars get significantly lower gas mileage than cars from other countries. **no**
- (b) U.S. cars get significantly higher gas mileage than cars from other countries. **no**
- (c) Swedish cars get gas mileages that are between those of Japanese and U.S. cars.
- (d) Mercedes, Audi, Porsche, and BMW represent approximately a quarter of the cars tested. **no**
- **(e) More than half of the cars in the study were from the United States.**



12.1 Which of the following bar graphs is equivalent to the pie chart in Question 12.2?



(B) None of these.

12.2 Earthquake intensities are measured using a device called a seismograph, which is designed to be most sensitive to earthquakes with intensities between 4.0 and 5.0 on the Richter scale. Measurements of nine earthquakes gave the following readings:

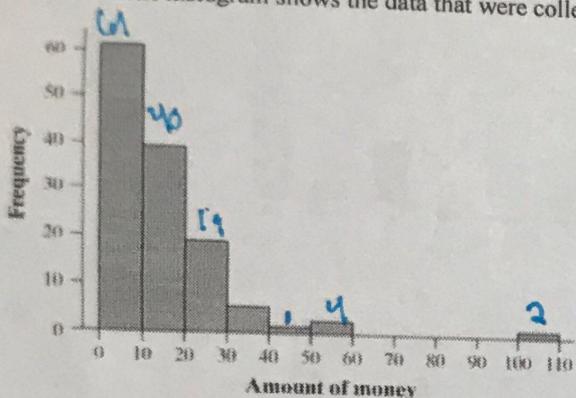
~~4.5~~ ~~4.8~~ ~~4.9~~ 4.4 4.7 4.9 6.0 4.1 5.7

where 4 indicates that the earthquake had an intensity below 4.0 and an 8 indicates that the earthquake had an intensity above 8.0. The median earthquake intensity of the sample is

- (A) 4.75
- (B) 4.9
- (C) 4.47
- (D) 4.70
- (E) Cannot be determined.

L 4.5 5.2 5.5 6.0 6.7 6.9 8.8
 7
 n

Questions T1.5 and T1.6 refer to the following setting. In a statistics class with 136 students, the professor records how much money (in dollars) each student has in his or her possession during the first class of the semester. The histogram shows the data that were collected.



$$\frac{61}{136} \approx 44.9\%$$

T1.5. The percentage of students with less than \$10 in their possession is closest to

- (a) 30%.
- (b) 35%.
- **(c) 50%.**
- (d) 60%.
- (e) 70%.

T1.6. Which of the following statements about this distribution is *not* correct? (false)

- (a) The histogram is right-skewed. *true*
- **(b) The median is less than \$20. *true***
- **(c) The IQR is \$35. *false***
- (d) The mean is greater than the median. *true*
- (e) The histogram is unimodal. *true*

$\frac{136}{2} = 68$
 med = between 68th + 69th data value
 Q_1 is between 34th + 35th data value
 Q_3 is between 102 + 103 data value

Q_1 is in 1st bar \int_0^{10}
 Q_3 is in 3rd bar \int_{10}^{20}
 IQR is \$10

T1.7. Forty students took given in the following set

0	28
1	2245
2	01333358889
3	001356679
4	2244446678
5	000

The third quartile of the

- (a) 45 (b) **44** (c)

T1.8. The mean salary of must be true about the m

- (a) It must be \$3
- (b) It must be la
- **(c) It could be ar**
- (d) It must be la
- (e) It cannot be

Questions T1.9 and T1 operations vary accord were sent to 200 random are returned, researche business. The data are

Size	Respo
Small	125
Medium	81
Large	40

T1.9. Given 2 What percent of

- (a) 12.5% (b)

T1.7. Forty students took a statistics examination having a maximum of 50 points. The score distribution is given in the following stem-and-leaf plot:

0	28
1	2245
2	01333358889
3	001356679
4	22444466788
5	000

$$\frac{40}{2} = 20$$

median between 20th + 21st data value

$$\frac{20}{2} = 10$$

Q3 is between 30th + 31st data value

The third quartile of the score distribution is equal to

- (a) 45 (b) 44 (c) 43 (d) 32 (e) 23

T1.8. The mean salary of all female workers is \$35,000. The mean salary of all male workers is \$41,000. What must be true about the mean salary of all workers?

- (a) It must be \$38,000.
- (b) It must be larger than the median salary.
- (c) It could be any number between \$35,000 and \$41,000.
- (d) It must be larger than \$38,000.
- (e) It cannot be larger than \$40,000.

Questions T1.9 and T1.10 refer to the following setting. A survey was designed to study how business operations vary according to their size. Companies were classified as small, medium, or large. Questionnaires were sent to 200 randomly selected businesses of each size. Since not all questionnaires in a survey of this type are returned, researchers decided to investigate the relationship between the response rate and the size of the business. The data are given in the following two-way table:

Size	Response	No Response	Total
Small	125	75	200
Medium	81	119	200
Large	40	160	200

Given 200

554

T1.9. What percent of all small companies receiving questionnaires responded?

- (a) 12.5% (b) 20.8% (c) 33.3% (d) 50.8% (e) 62.5%

Small and respond

Total small companies

$$\frac{125}{200} = 62.5\%$$

T1.10. Which of the following conclusions seems to be supported by the data?

- (a) There are more small companies than large companies in the survey.
- (b) Small companies appear to have a higher response rate than medium or big companies. *yes*
- (c) Exactly the same number of companies responded as didn't respond.
- (d) Small companies dislike larger companies.
- (e) If we combined the medium and large companies, then their response rate would be equal to that of the small companies.

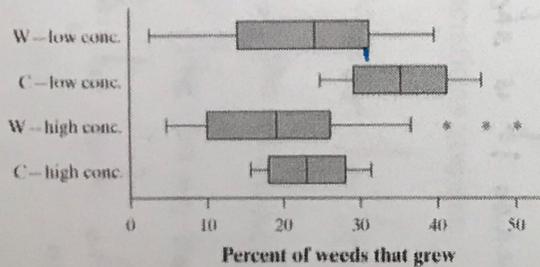
$$\text{small companies response rate} = \frac{125}{200} = 62.5\%$$

$$\text{medium companies response rate} = \frac{81}{200} = 40.5\%$$

$$\text{large companies response rate} = \frac{46}{200} = 23\%$$

T1.11. An experiment was conducted to investigate the effect of a new weed killer to prevent weed growth in onion crops. Two chemicals were used: the standard weed killer (C) and the new chemical (W). Both chemicals were tested at high and low concentrations on a total of 50 test plots. The percent of weeds that grew in each plot was recorded. Here are some boxplots of the results. Which of the following is not a correct statement about the results of this experiment?

false



- (a) At both high and low concentrations, the new chemical (W) gives better weed control than the standard weed killer (C). *True*
- (b) Fewer weeds grew at higher concentrations of both chemicals. *True*
- (c) The results for the standard weed killer are less variable than those for the new chemical. *True*
- (d) High and low concentrations of either chemical have approximately the same effects on weed growth. *False*
- (e) Some of the results for the low concentration of weed killer W show fewer weeds growing than some of the results for the high concentration of W. *True*

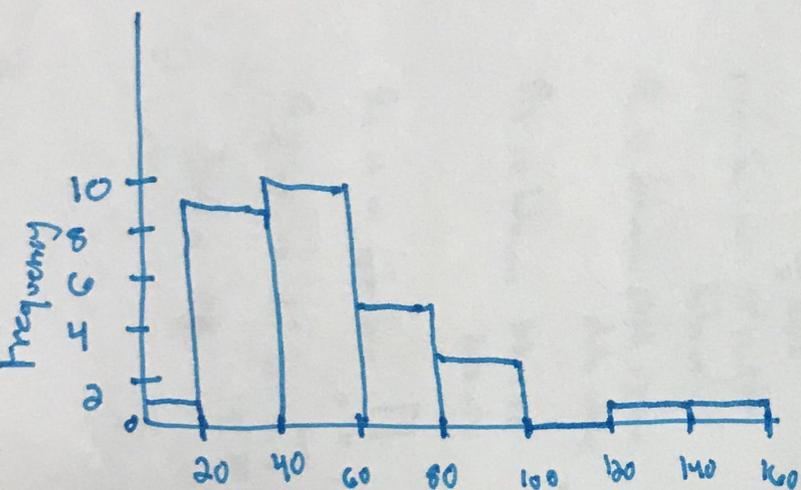
Section II: Free Response Show all your work. Indicate clearly the methods you use, because you will be graded on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

T1.12. You are interested in how much time students spend on the Internet each day. Here are data on the time spent on the Internet (in minutes) for a particular day reported by a random sample of 30 students at a large school:

7	20	24	25	25	28	28	30	32	35
42	43	44	45	46	47	48	48	50	51
72	75	77	78	79	83	87	88	135	151

$\bar{x} = 54.7\bar{6}$ $IQR = 77 - 30 = 47$
 $min = 7$
 $Q_1 = 30$
 $med = 46.5$
 $Q_3 = 77$
 $max = 151$

- (a) Construct a histogram of these data.



Time spent on Internet in Minutes

- (b) Are there any outliers? Justify your answer. $IQR(1.5) = 47(1.5) = 70.5$
 $Q_1 - IQR(1.5) = 30 - 70.5 = -40.5$ Yes, outliers are below -40.5 or above 147.5
 $Q_3 + IQR(1.5) = 77 + 70.5 = 147.5$ so 151 is an outlier.
- (c) Would it be better to use the mean and standard deviation or the median and IQR to describe the center and spread of this distribution? Why?

It would be better to use the median and IQR to describe the center and spread of this distribution because it is skewed to the right.

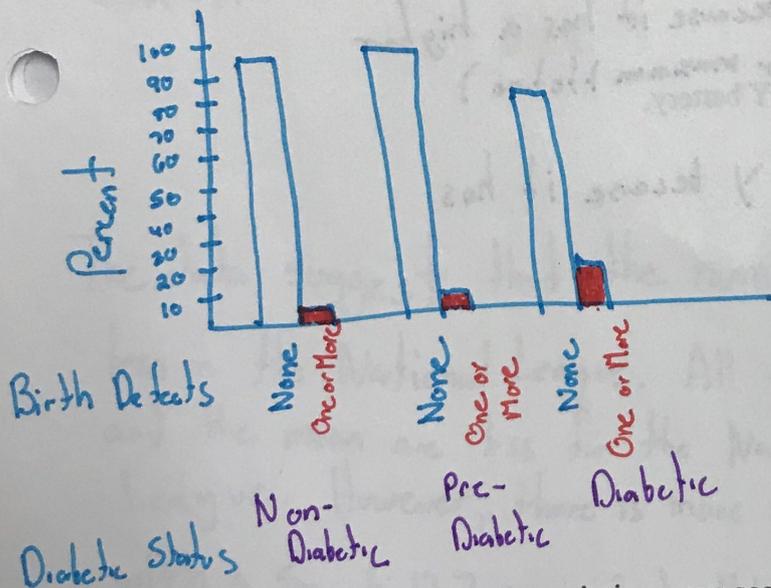
T1.13. A study among the Pima Indians of Arizona investigated the relationship between a mother's diabetic status and the appearance of birth defects in her children. The results appear in the two-way table below.

Birth Defects	Diabetic Status			Total
	Nondiabetic	Prediabetic	Diabetic	
None	754	362	38	1,154
One or more	31	13	9	53
Total	785	375	47	1,207

- ✓ (a) Fill in the row and column totals in the margins of the table.
- (b) Compute (in percents) the conditional distributions of birth defects for each diabetic status.

Birth Defect	Diabetic Status			diabetic status total
	Non-Diabetic	Pre-Diabetic	Diabetic	
None	$754/785 = 96.1\%$	$362/375 = 96.5\%$	$38/47 = 80.9\%$	
One or More	$31/785 = 3.9\%$	$13/375 = 3.5\%$	$9/47 = 19.1\%$	

- (c) Display the conditional distributions in a graph. Don't forget to label your graph completely.



- (d) Comment on any clear associations you see.

There does appear to be an association.

Nondiabetics and pre-diabetics appear to have babies with birth defects at about the same rate. However, those with diabetes have a much higher rate of babies with birth defects.

T1.14. The back-to-back stemplot shows the lifetimes of several Brand X and Brand Y batteries.

Brand X		Brand Y
	1	
	1	7
	2	2
	2	6
2 1 1 0	3	
9 7 7 5	3	
3 2 2 1	4	2 2 3 3 4
	4	5 6 8 8 9
	5	0
	5	

median 430 - 439 hours

median =
390 - 399
hours

Key: 4|2 represents
420-429 hours.

- (a) What is the longest that any battery lasted?

The longest that any battery lasted was between 550 and 559 hours.

- (b) Give a reason someone might prefer a Brand X battery.

Someone might prefer Brand X because it has a higher minimum lifetime. (or higher maximum lifetime)

- (c) Give a reason someone might prefer a Brand Y battery.

Someone might prefer Brand Y because it has a higher median lifetime.

T1.15. During the early part of the 1994 baseball season, many fans and players noticed that the number of home runs being hit seemed unusually large. Here are the data on the number of home runs hit by American League and National League teams in the early part of the 1994 season:

American League: 35 40 43 49 51 54 57 58 58 64 68 68 75 77

National League: 29 31 42 46 47 48 48 53 55 55 55 63 63 67

Compare the distributions of home runs for the two leagues graphically and numerically. Write a few sentences summarizing your findings.

L1 American League: $n = 14$

$$\bar{x} = 56.9$$

$$s_x = 12.7$$

$$\min = 35$$

$$Q_1 = 49$$

$$\text{med} = 57.5$$

$$Q_3 = 68$$

$$\max = 77$$

$$\text{IGR} = 19$$

L2 National League: $n = 14$

$$\bar{x} = 50.1$$

$$s_x = 11.1$$

$$\min = 29$$

$$Q_1 = 46$$

$$\text{med} = 50.5$$

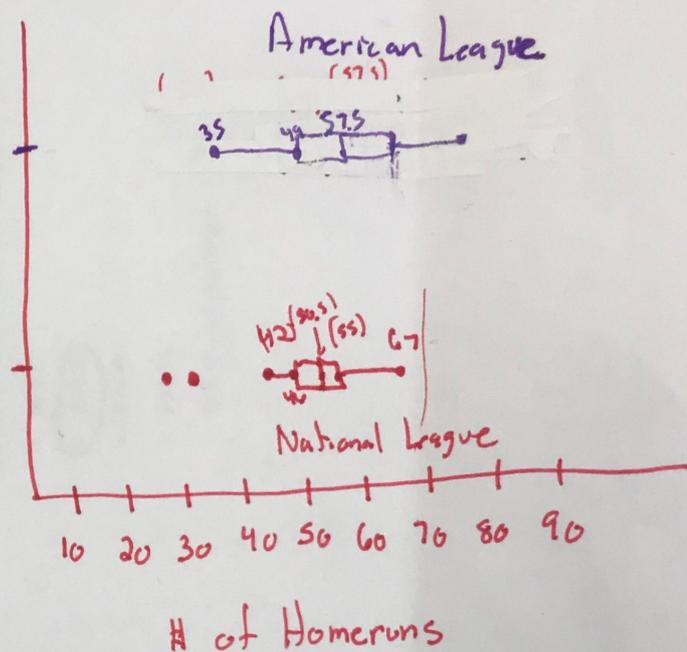
$$Q_3 = 55$$

$$\max = 67$$

$$\text{IGR} = 9$$

American League

National League



The data suggest that the number of home runs is somewhat less in the National League. All 5 #'s in the five # summary and the mean are less for the National League than the American League. However, there is more variability among American League teams, with a s_x of 12.7 compared to 11.1 for the National League. Both distributions are reasonably symmetric. The American League has no outliers, but the teams that hit 29 and 31 home runs are outliers in the National League.