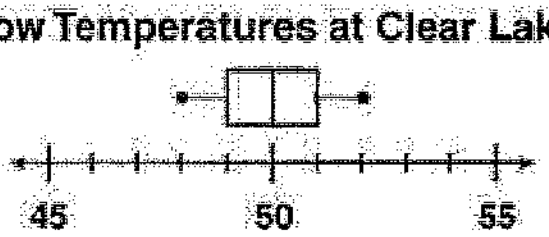
Coordinate Algebra
SID: Statistics Review

Name: _____
Date: _____

1. Judy recorded the low temperatures at Clear Lake for ten days in May. She made the box and whisker plot to display her data.



Which value defines the upper quartile of the temperatures? 55

2. The number of runs scored by the Lions for six games is shown below. (Multiple Choice)

5, 9, 1, 5, 2, 7

If the Lions scored 13 runs in their seventh game, which statement is true?

A. The median and the mean both remain the same.
B. The mean and the median both increase.
C. The median increases and the mean remains the same.
(D.) The mean increases and the median remains the same.

3. The number of runs scored by the Mustangs for six games is shown below.

5, 8, 2, 5, 2, 6

If the Mustangs scored 13 runs in their seventh game, what would happen to the median and mean?

2, 2, 5, 5, 6, 8 13 — 4.67 → 5.83

median stays the same — mean increases

4. Any time a set of data has a large MAD, it means that the data must be skewed to then left or the right.
True or (False)

5. Read the following and answer the questions.
The Caldwell Family and the Greaves Family both have 8 family members.
The ages of the Caldwell Family are ~~29, 2, 27, 41, 9, 54, 13~~, and ~~5~~ 2, 5, 9, 13, 27, 29, 41, 54
The ages of the Greaves Family are ~~22, 21, 1, 7, 27, 48, 39~~, and ~~15~~ 1, 7, 15, 21, 22, 27, 39, 48

a. What are the means for the Caldwell and Greaves Families? What are the medians?

22.5 22.5 C - 20 G - 21.5

b. Which is a better measure of center for each family?

median

c. What is the MAD for the Caldwell Family?

15.25

d. What is the MAD for the Greaves Family?

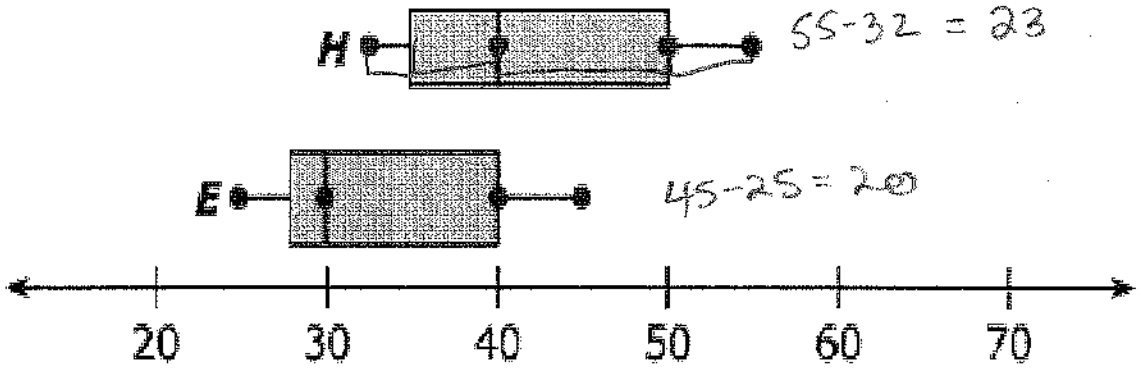
11.625

e. What do these 2 MADs tell you about their families? How do they compare?

The spread of the ages in their Family

Caldwell's Family's ages are more spread than Greaves'

Use the Box and Whisker Plot below to answer #6 & 7.



6. Tell whether the following is true or false for the box and whisker plot above.
   a. Over half of all the parents of elementary school students are 40 or younger. (T) or F
   b. The range of the ages of the parents of elementary students is 20 years. (T) or F
   c. The ages of parents of elementary school students fall within a smaller range than the ages of parents of high school students. (T) or F
   d. The ranges for both elementary and high school parents are the same. T or (F)
   e. The majority of the parents of high school students are 35 or older. (T) or F
   f. A majority of all the parents are 30 years or older. (T) or F

7. What percentage of high school parents are older than 50? Younger than 35?
   25% 25%

Use the table to answer questions #8 – 15.

FREE-TIME ACTIVITY

Activity	Girls	Boys	Total
Listen to music	32	34	66
Play a sport	38	24	62
Read	26	32	58
Watch TV	40	44	84
Total	136	134	270

8. How many girls play sports? 38
9. How many boys read? 32
10. How many students Watch TV overall? 84
11. What is the marginal frequency of students who listen to music? 66
12. What **percentage** of students play sports? 62/270 23%
13. What **percentage** of girls watch TV? 40/136 29%
14. What **fraction** represents the number of boys who listen to music? $34/134 = \frac{17}{67}$
15. What **fraction** represents the number of students who watch TV? $\frac{84}{270} = \frac{14}{45}$
16. What does mean measure? What does MAD measure? What does the median measure? What does range measure?
    center spread center spread
17. Write a statement that describes the correlation coefficient.
    a. -0.8 strong negative correlation
    b. 0.7 strong positive correlation
    c. -1 straight negative line
    d. -0.1 weak negative correlation
    e. 0 no correlation
    f. 0.9 strong positive correlation

18. Tracey and Jason played 5 games on a handheld video game and recorded their scores in the table. Find the mean absolute deviation for their scores and identify which statement below is true when comparing their variability.

Game	Tracey (avg. 20)		Jason (avg. 20)	
Game 1	30	10	15	5
Game 2	20	0	25	5
Game 3	5	15	20	0
Game 4	25	5	25	5
Game 5	20	0	15	5

$6 = 30/5$ $4 = 20/5$

~~A.~~ Tracey's MAD is 26 and she has less variability.

(B.) Tracey's MAD is 6 and she has more variability.

~~C.~~ Jason's MAD is 15.8 and he has less variability

D. Jason's MAD is 4 and he has more variability.

more variability: data is more spread: MAD is bigger

less variability: data is close together: MAD is smaller

19. List the joint frequencies and the marginal frequencies for the following table.

	Curfew: Yes	Curfew: No	Total
Chores: Yes	13	5	18
Chores: No	12	3	15
Total	25	8	33

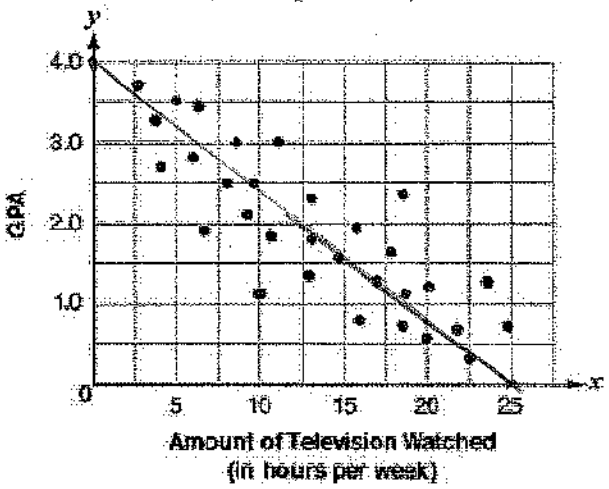
Joint: 13, 5, 12, 3

Marginal: 25, 8, 18, 15

20. When do you use the mean to best represent the measure of center for a set of data and when do you use the median? Mean if the data is symmetric

median if the data is skewed

21. The scatter plot below shows data that were collected to compare the amount of television a student watched (hours per week) and his or her GPA.



a) Describe the relationship between the data. negative

b) Guess the correlation coefficient. -0.6 to -0.8

c) Find the line of best fit. (guess) $y = -\frac{4}{25}x + 4$

d) What is the slope of the line? Describe it in the context of the problem. $-\frac{4}{25}$ GPA decreases as the amount of TV increases

e) What is the y-intercept of the problem? Describe it in the context of the problem. 4.0 when a student watches no TV they have the highest GPA

22. A scatter plot shows a strong negative correlation between watching TV after school and course grade. Based on this information, state whether the following statements could be made and why:

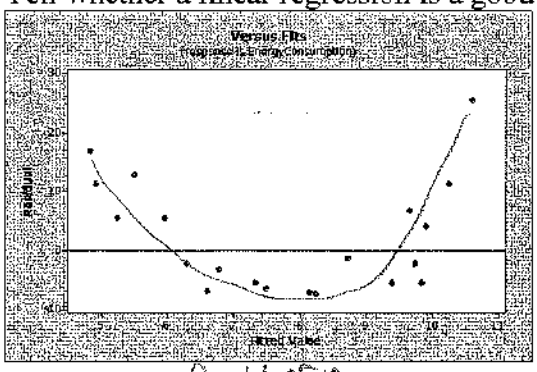
a) Not watching TV causes higher grades no, Causation

b) There is a negative correlation between watching TV and course grades yes, correlation

c) If you watch TV, you have a bad grade. no, Causation

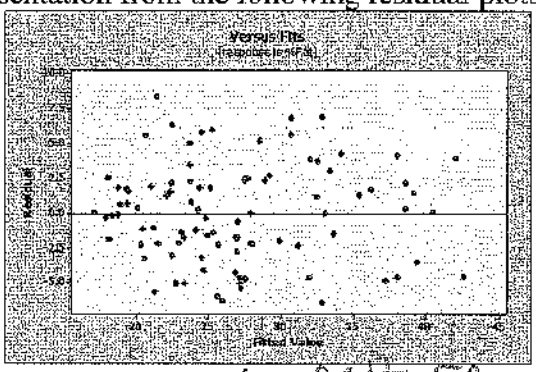
Correlation does not imply Causation

23. Tell whether a linear regression is a good representation from the following residual plots.

a) 

NO

pattern

b) 

yes

no pattern

24. The table shows the percent of the United States population who did not receive needed dental care services due to cost.

Year	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Percent	7.9	8.1	8.7	8.6	9.2	10.7	10.7	10.8	10.5	12.6	13.3

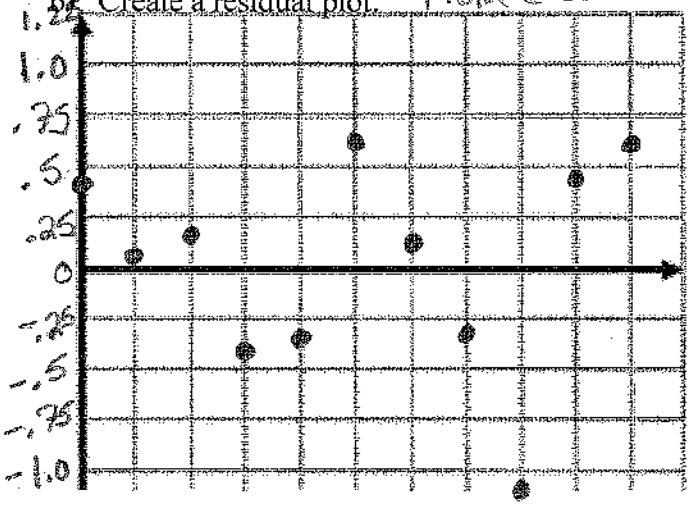
a) If the linear regression equation is $y = 0.512x + 7.541$, find the residuals using the table below:

# of years since 1999	Observed Percent	Predicted Percent	Residual = observed - predicted
0	7.9	7.541	.359
1	8.1	8.053	.047
2	8.7	8.565	.135
3	8.6	9.077	-.477
4	9.2	9.589	-.389
5	10.7	10.101	.599
6	10.7	10.613	.087
7	10.8	11.125	-.325
8	10.5	11.637	-1.137
9	12.6	12.149	.451
10	13.3	12.661	.639

From chart

plug # of yrs into equation

b) Create a residual plot.



c) Does the residual plot suggest a linear relationship? Why or why not?

yes, plot has no pattern
or
plot is random