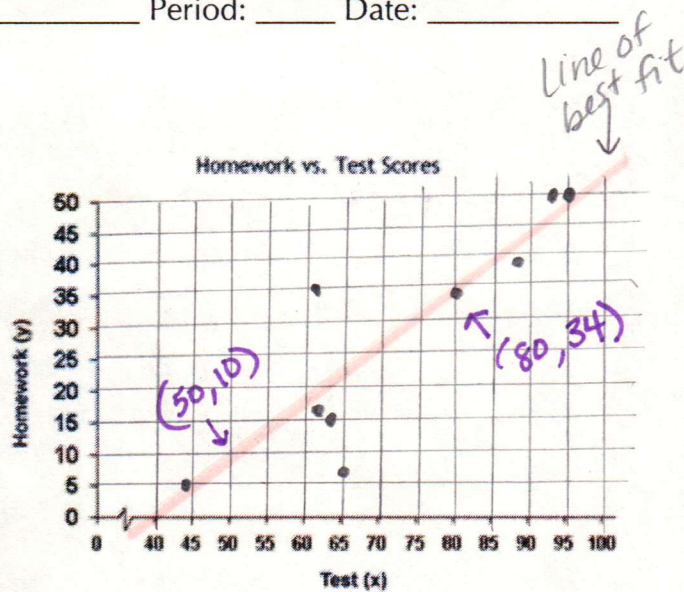


Notes 4.6 – Statistics

Warmup – Create a scatterplot of the given data.

Add grid lines if you believe that would be helpful.

Test Mark (x)	Homework Mark (y)
61	35
95	50
44	5
93	50
63	15
80	34
62	16
95	50
65	7
88	38



Look at the scatter plot, then describe the correlation coefficient.

Strength: *strong*

Direction: *positive*

$$r = .8952$$

Draw a line of best fit through the data.

Find two points on the line of best fit and write an equation for the line.

$$m = \frac{34 - 10}{80 - 50} = \frac{24}{30} = \frac{4}{5}$$

$$y = \frac{4}{5}(x - 50) + 10$$

exact

$$y = .89x - 36.7$$

$$y = \frac{4}{5}x - 40 + 10$$

$$y = \frac{4}{5}x - 30$$

Lesson – Linear Regression

Word	Meaning/Notation	Example
Line of Best Fit	A line that goes through the data, modeling the trend and relationship of the data. Linear or curved	see warmup

The US Census Bureau collects information about the United States. The data in the tables gives the median income based on gender. Dollar amounts have been adjusted for inflation.

The data for men has already been plotted on the graph below.

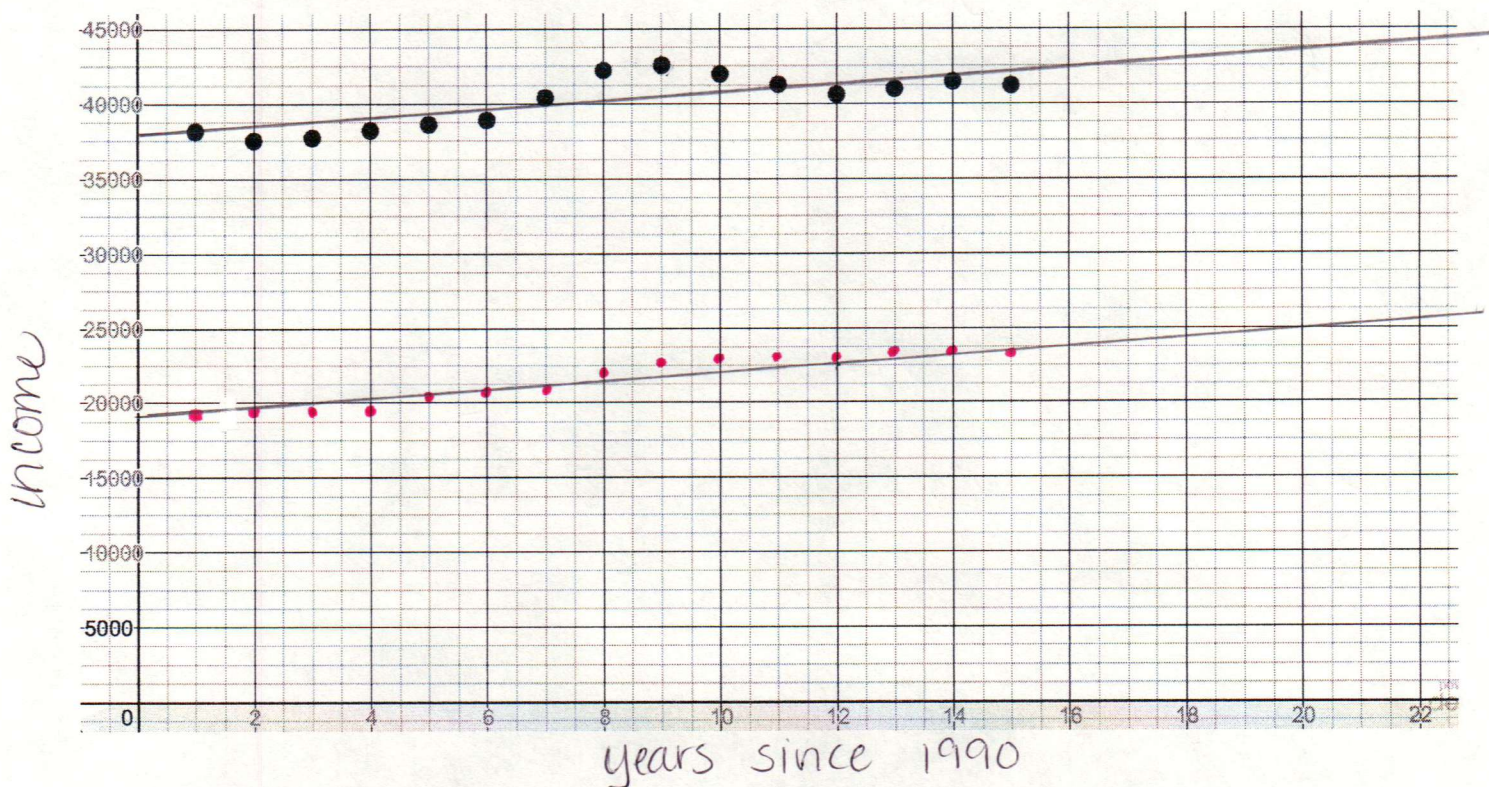
On the same graph, plot the data for women.

Draw the line of best fit for both.

1991 is $x=1$

Year	Median Income for All Men
2005	41196
2004	41464
2003	40987
2002	40595
2001	41280
2000	41996
1999	42580
1998	42240
1997	40406
1996	38894
1995	38607
1994	38215
1993	37712
1992	37528
1991	38145

Year	Median Income for All Women
15 2005	23970
14 2004	23989
13 2003	24065
12 2002	23710
11 2001	23564
10 2000	23551
9 1999	22977
8 1998	22403
7 1997	21759
6 1996	20957
5 1995	20253
4 1994	19158
3 1993	18751
2 1992	18725
1 1991	18649



Data for Men $r = .7928$

Correlation Direction: positive

Correlation Strength: strong

Equation for line of best fit:

$$y = 311.66x + 37,629$$

\uparrow \uparrow
 increase per median
 year in 1990

Data for Women $r = .9636$

Correlation Direction: positive

Correlation Strength: strong

Equation for line of best fit:

$$y = 464.1x + 18052.6$$

\uparrow \uparrow
 increase median in
 per year 1990

Now let's compare your correlation info and equation with the actual correlation coefficient and equation.

Men:

$$r = .7928$$

$$y = 311.66x + 37,629$$

Women

$$r = .9636$$

$$y = 464.1x + 18,052.6$$

- a. How accurate was your information compared to the actual information?
- b. What does the correlation coefficient tell you about the relationship between years and income for men?

moderately strong - close to .8

- c. What does the correlation coefficient tell you about the relationship between years and income for women?

very strong - close to 1

- d. How do the two data sets compare?

the stronger correlation for women creates a more accurate model

- e. What does the slope of the line mean in terms of the data?

how much more you will make each year than the year before

- f. What do the lines tell you about the income of men vs. women from 1991-2005?

income is increasing for both

- g. Estimate what the median income will be for men and women in 2015.

$$\text{men: } y = 311.66(25) + 37629.7 = \$45,421.20$$

$$\text{women: } y = 464.1(25) + 18052.6 = \$29,655.10$$

h. Add the following data to the graph for both men and women.

Year	Median Income for All Men
2011	37653
2010	38014
2009	38588
2008	39134
2007	41033
2006	41103

Year	Median Income for All Women
2011	23395
2010	23657
2009	24284
2008	23967
2007	25005
2006	24429

i. How well does the new data fit with the lines of best fit that you drew?

not well, new data has a negative trend

j. Do you think the estimate of median income you got in (g) is accurate? Why or why not?

k. What reason could there have been for the change in data? In the very long run (decades), do you think the model is still fairly accurate? Why?