Assignment: Two-variable Statistics

Andy Yan

September 2022

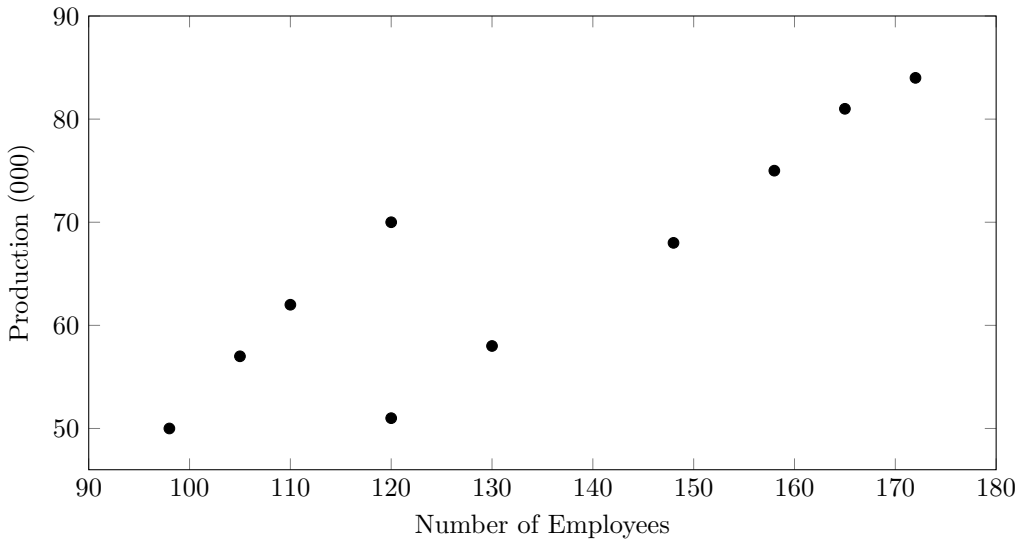
Pg 210 Q8

1 Data

A manufacturing company keeps records of its overall annual production and its number of employees. Data for a ten-year period are shown below.

Year	Number of Employees	Production (000)
1992	158	75
1993	165	81
1994	172	84
1995	148	68
1996	130	58
1997	120	51
1998	98	50
1999	105	57
2000	110	62
2001	120	70

2 Scatter Plot



a) To determine the behaviour of the graph, we must find the correlation coefficient (Pearson's r).

$$r = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sqrt{\sum(x-\overline{x})^2}\sqrt{\sum(y-\overline{y})^2}} \approx \frac{278.49}{26.41 \cdot 11.97} \approx 0.88$$

Since r belongs in the range $\frac{2}{3} \leq r < 1$, this scatter plot graph has a **strong positive linear correlation**.

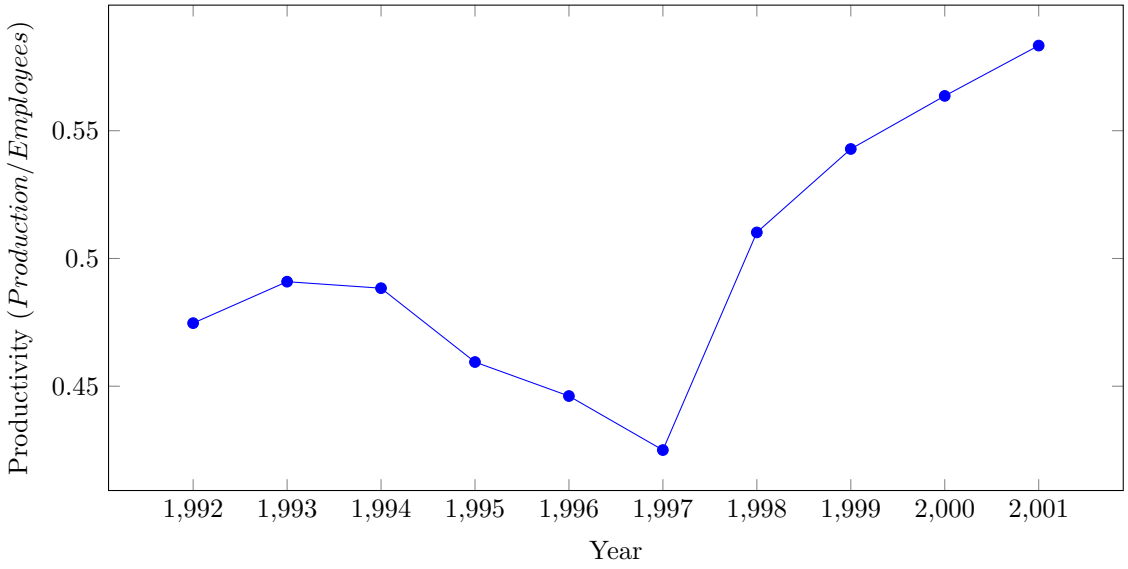
b) **1994-1995**

c) The scatter plot does suggest that there is a hidden variable in play. We can see this when the company has 120 employees for two different, yet, the production differs. This could be due to worker efficiency and

company management. We can assume that the company improved these factors since 1997 and that's why the production value is higher in 2001. The layoffs could suggest the company went under changes which shows us the temporary decline following it. However, we can see that the production starts improving as the company starts to regain employees.

3 Time-Series Graph (Productivity vs Year)

d)



e)

$$y = ax + b$$

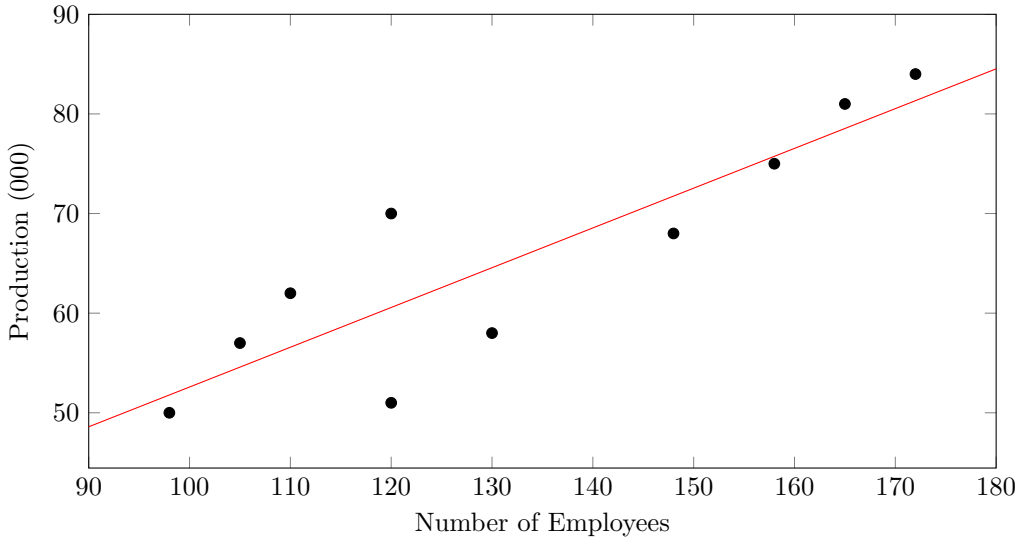
$$a = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2} = \frac{10(89492) - (1326)(656)}{10(182106) - (1326)^2} = \frac{3133}{7848} \approx 0.399$$

$$b = \overline{y} - a\overline{x}$$

$$b = \frac{656}{10} - \frac{3133}{7848} \cdot \frac{1326}{10} \approx 12.665$$

After solving for a and b, we can write the equation for the line of best fit as:

$$y = 0.399x + 12.665$$



f) This system was implemented in 1997 as we can see in our time-series graph that it switches from decreasing to increase after 1997.