

1. The countries of Europe report that 46% of the labor force is female. The United Nations wonders if the percentage of women in the labor force is the same in the United States. Representatives from the United States Department of Labor plan to check a random sample of over 10,000 employment records on file to estimate the percentage of people who identify as female in the United States labor force.

a. The representatives from the Department of Labor want to estimate the percentage of people who identify as women in the United States labor force to within $\pm 5\%$, with 90% confidence. How many employment records should they sample?

$$z^* = 1.6449$$

$$\begin{aligned} n &= \hat{p} \hat{q} \left(\frac{z^*}{ME} \right)^2 \\ &= .46(.54) \left(\frac{1.6449}{.05} \right)^2 \\ &= 268.8 \end{aligned}$$

we need to sample 269 records to have a margin of error of 5%.

b. They actually select a random sample of 525 employment records, and find that 229 of the people identify as women. Create the confidence interval.

$$\hat{p} = \frac{229}{525} = 0.4362$$

$$\begin{aligned} \hat{q} &= 1 - .4362 \\ &= .5638 \end{aligned}$$

$$\begin{aligned} \hat{p} \pm z^* \sqrt{\frac{\hat{p}\hat{q}}{n}} & \quad (.4006, .4718) \\ .4362 \pm 1.6449 \sqrt{\frac{.4362(.5638)}{525}} & \quad (40\%, 47\%) \\ .4362 \pm .0356 & \end{aligned}$$

c. Interpret the confidence interval in this context.

we are 90% confident that the true proportion of people who identify as female in the U.S. labor force is between 40% and 47%.

d. Should the representatives from the Department of Labor conclude that the percentage of women in their labor force is lower than Europe's rate of 46%? Explain.

No, they cannot conclude that because even though 46% is on the high side of the interval, it is within the interval.

2. A company claims to have invented a hand-held sensor that can detect the presence of explosives inside a closed container. Law enforcement and security agencies are very interested in purchasing several of the devices if they are shown to perform effectively. An independent laboratory arranged a preliminary test. If the device can detect explosives at a rate greater than chance would predict, a more rigorous test will be performed.

They placed four empty boxes in the corners of an otherwise empty room. For each trial they put a small quantity of an explosive in one of the boxes selected at random. The company's technician then entered the room and used the sensor to try to determine which of the four boxes contained the explosive. The experiment consisted of 50 trials, and the technician was successful in finding the explosive 16 times. Does this indicate that the device is effective in sensing the presence of explosives?

a. Check the conditions, test an appropriate hypothesis and state your conclusion.

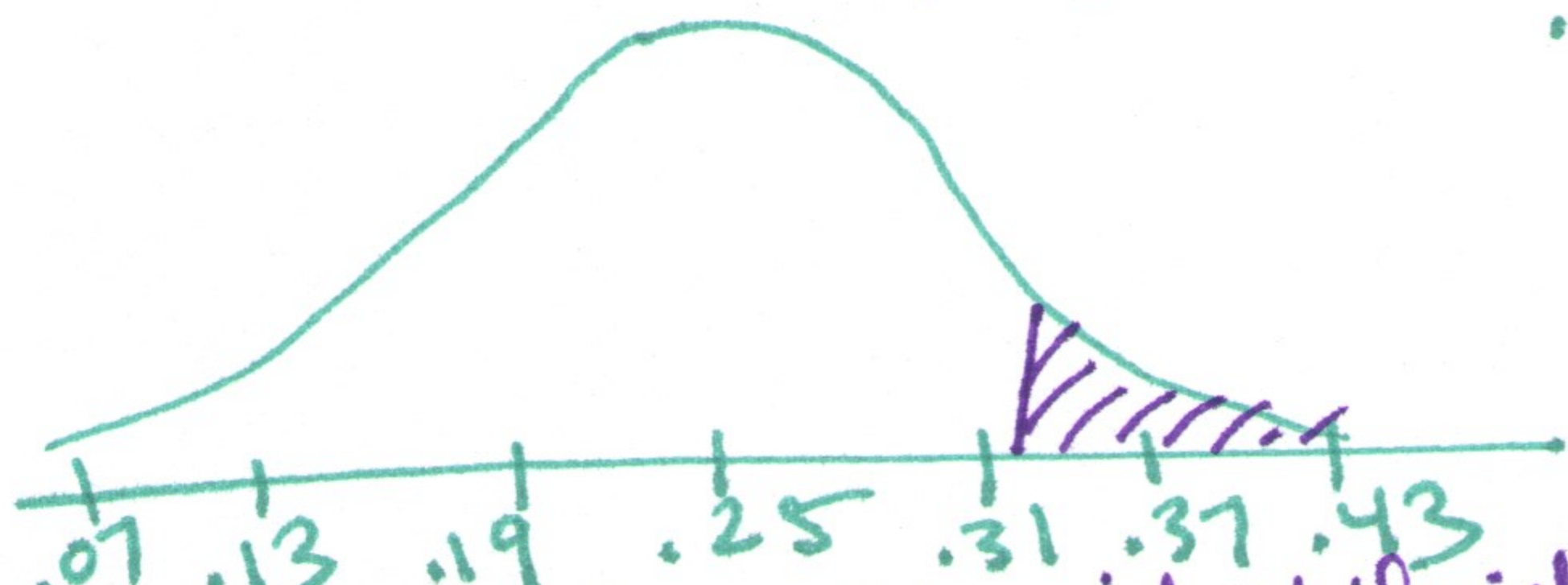
$H_0: p = .25$ no better than chance 1/4
 $H_A: p > .25$ correct more than 25% of the time chance

Conditions

- ① indep - Trials are independent if the explosive is randomly placed
- ② random ✓
- ③ < 10% Trials are independent so this does not matter we could do as many trials as we want. (with replacement)
- ④ $np = 50(.25) = 12.5$ ✓ $nq = 50(.75) = 37.5$ ✓

if H_0 is true:

$$\mu = .25 \quad \sigma = \sqrt{\frac{.25(.75)}{50}} \approx .0612$$



$$\hat{p} = \frac{16}{50} = .32$$

$$p\text{-value} = P(\hat{p} > .32) = .1264$$

The p-value is larger than .05 so we fail to reject the null.

b. Create a 95% confidence interval for the number of sensors that were successful in finding the explosives. Explain whether the confidence interval agrees with the hypothesis test.

$$\hat{p} = .32$$

$$.32 \pm \sqrt{\frac{.32(.68)}{50}}$$

$$.32 \pm .1293$$

$$(.1907, .4493)$$

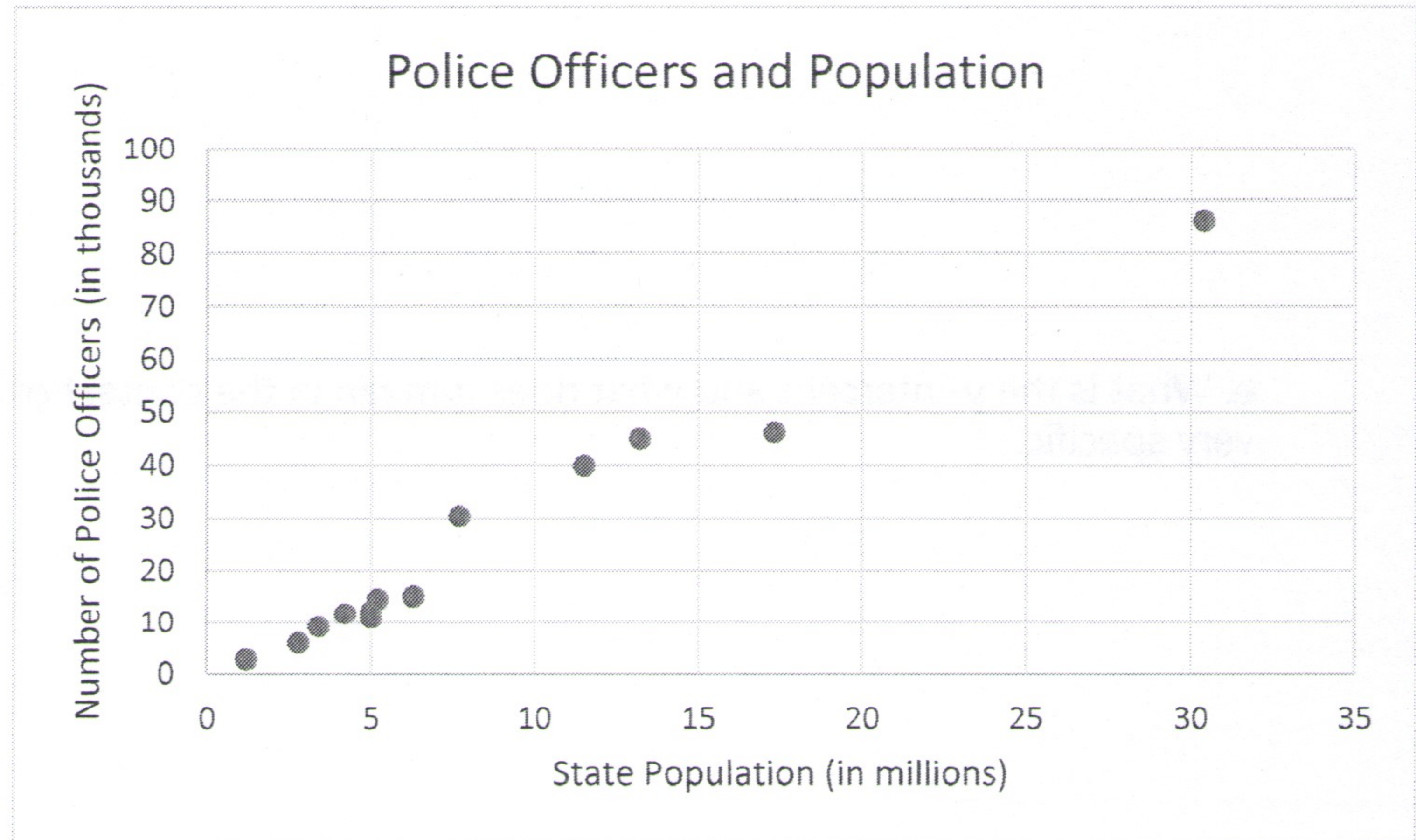
$$(19\%, 45\%)$$

The confidence interval agrees with the hypothesis test because 25% is within the interval.

We do not have enough evidence to suggest the machine is better than chance.

3. The data below shows the population (in millions) of several states and the number of police officers they have (in thousands).

State	Pop	Police
CA	30.4	86.2
CO	3.4	9.2
FL	13.2	45.0
IL	11.5	39.9
IA	2.8	6.0
LA	4.2	11.8
ME	1.2	2.9
MS	5.2	14.6
NJ	7.7	30.5
TN	5.0	12.3
TX	17.3	46.2
VA	6.3	15.2
WA	5.0	10.9



a. Describe the direction, form and strength of the association and indicate whether or not there are any unusual points. Write complete sentences and include the correlation value.

There is a strong, positive, linear (maybe curved) relationship between the state population and the number of police officers. The correlation is very high at 0.98. California has a much larger population so there is a gap but it seems to fit the pattern.

b. Calculate the least-squares regression line using Excel. Round the values to 4 decimal places.

$$\hat{y} = 2.9215x - 0.0008$$

c. Using your regression line, predict the number of police officers that a state with 20 million people would employ.

$$\begin{aligned} \hat{y} &= 2.9215(20) - .0008 \\ &= 58.4292 \text{ thousand police officers} \end{aligned}$$

A state with 20 million people would employ 58,429 police officers according to the model

d. What is the slope and what does it mean in the context of this problem? Include units and be very specific.

$$\text{Slope} = \frac{2.9215 \text{ thousand officers}}{1 \text{ million people}}$$

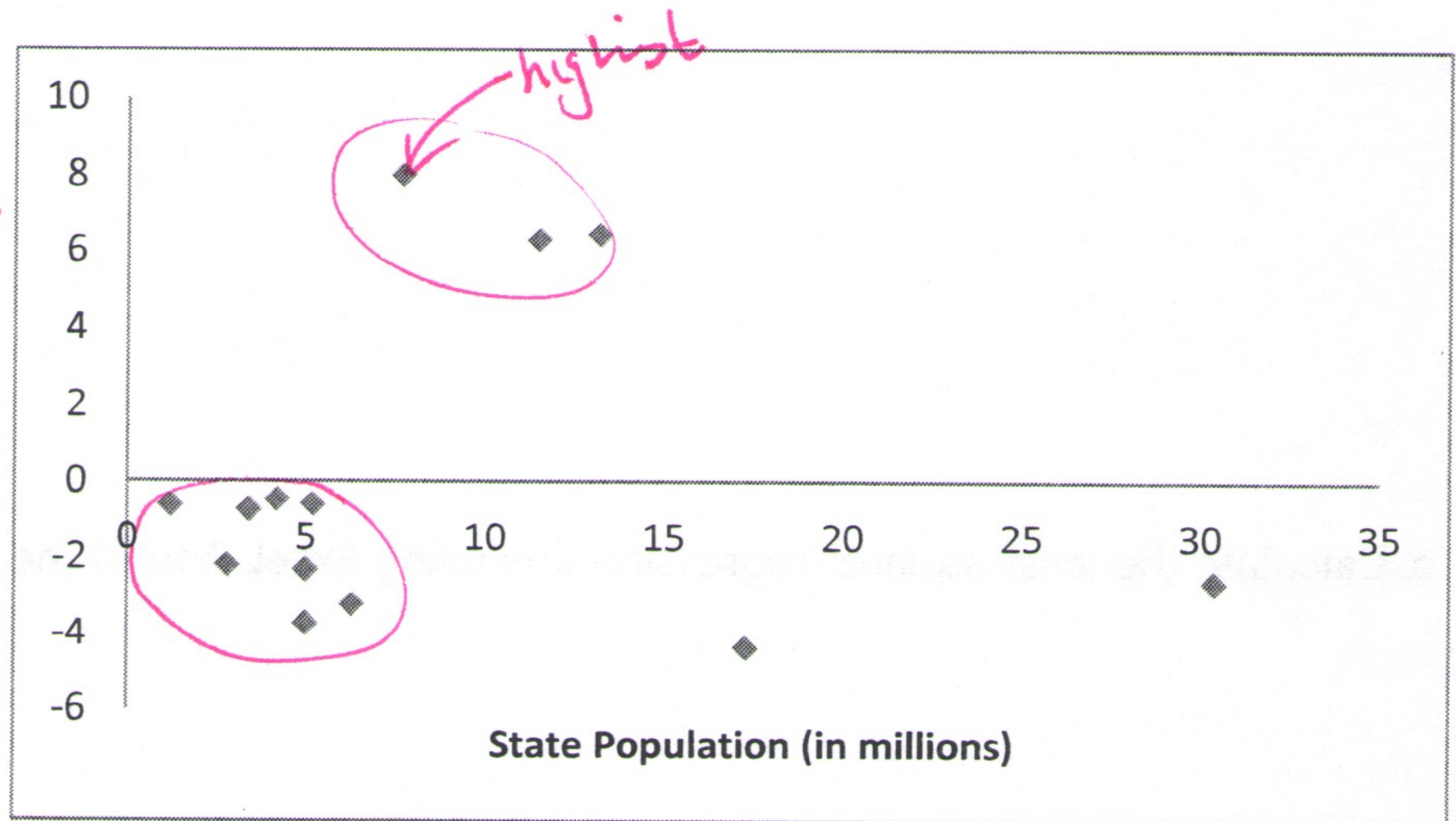
For every increase of one million in population, states employ 2,922 more police officers.

e. What is the y-intercept and what does it mean in the context of the problem? Include units and be very specific.

The y-intercept is -0.0008 thousand officers. That would mean that a state with zero population had -0.8 officers. It doesn't make sense to have a negative number of officers. This is close to the point $(0,0)$, which would make sense.

4. This is a residual plot for the above data. Based on the residual plot, do you think the linear model is appropriate? Explain.

I don't think the model is appropriate because the first grouping is below the line and the next grouping is above. A curved model may fit this data better.



b. Which state has the largest residual? Calculate the residual and explain what it means.

New Jersey has the largest residual.

$$\text{Residual} = \text{observed} - \text{predicted}$$

$$= y - \hat{y}$$

$$= 30.5 - 22.4948$$

$$= 8.0052$$

$$\hat{y} = 2.9215(7.7) - 0.0008$$

$$= 22.4948$$

Cara Lee The model underestimates the number of police officers in New Jersey by 8005.