PIERIAN DATA

EXERCISE #4 - Statistics – ANSWER KEY

1. What is the distinction between a population and a sample?

A population is every member of a group we want to study.
A sample is a smaller, randomly selected set of members of the population.

2. What is the difference between a parameter and a statistic?

A parameter is a characteristic of a population, while a statistic is a characteristic of a sample.

3. A hospital conducts a survey of patients who were given an experimental, lifesaving treatment. Hospital administrators call patients at home and ask them to participate in the survey. What types of sampling bias might be involved?

There are several possible answers here. They include:

1. Undercoverage bias – only those patients who are home at the time of the call and have access to a phone will be included in the survey
2. Self-selection bias – only those patients who agree to participate will be included in the survey
3. Survivorship bias – only those patients who survived the experimental treatment will be included in the survey.

4. What does the Central Limit Theorem say about populations and samples?

The Central Limit Theorem states that the mean values from a group of samples will be normally distributed about the population mean, even if the population itself is not normally distributed.

5. If a population has a mean of 600 and a standard deviation of 50, what is the Standard Error of the Mean for a sample size of 100? What does this value indicate?

$$SE_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{100}} = 5$$

68% of means taken from 100-item samples are expected to fall between 595 and 605.

6. In Hypothesis Testing, what is the difference between a one-tailed and two-tailed test?

If the alternative hypothesis considers only values that are greater than **or** less than the population parameter, then a one-tailed test is used.
If the alternative hypothesis considers any values that are **not equal to** the population parameter, then a two-tailed test is used.

7. A company wants to determine if two different sales departments had statistically the same number of sales per week over the last nine weeks. Perform a Student's t-Test on the following results. We recommend using a spreadsheet!

	Sales per Week		Dept A	Dept B	Dept A	Dept B
	Dept A	Dept B	$(x_1-\bar{x}_1)$	$(x_2-\bar{x}_2)$	$(x_1-\bar{x}_1)^2$	$(x_2-\bar{x}_2)^2$
1	40	43	2	2	4	4
2	36	41	-2	0	4	0
3	42	44	4	3	16	9
4	36	39	-2	-2	4	4
5	35	37	-3	-4	9	16
6	35	35	-3	-6	9	36
7	41	44	3	3	9	9
8	43	46	5	5	25	25
9	34	40	-4	-1	16	1
Sum:	342	369		Sum:	96	104
	$\bar{x}$= 38	$\bar{x}$= 41		s^2=	12	13

$$H_0: \bar{x}_A \geq \bar{x}_B$$

$$H_1: \bar{x}_A < \bar{x}_B$$

$$s^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$$

$$t = \frac{|\overline{x_1} - \overline{x_2}|}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}} = \frac{|38-41|}{\sqrt{\frac{12}{9} + \frac{13}{9}}} = \frac{3}{\sqrt{\frac{25}{9}}} = \frac{3}{\frac{5}{3}} = \frac{9}{5} = \mathbf{1.8}$$

If t-Critical for a one-tailed test with 95% confidence and 16 degrees of freedom is 1.74, what can we conclude about these departments?

We would reject the null hypothesis that mean sales for Dept A is greater than or equal to that of Dept B. Instead we support the alternative hypothesis that Dept B has a greater mean value than Dept A.