

Statistics I

Quiz 1

SOLUTIONS

Problem 1: (30 percent)

Researchers have devised a new test for strep throat. The test has a false positive rate of 10% (meaning that the probability that the test indicates that you have strep throat given that you do not is 0.10). It also has a false negative rate of 2.5% (meaning that the probability that it says that you do not have strep throat given that you do is 0.025).

- a) If 20% of the patients tested in the population actually have strep throat, complete the following table of joint probabilities.

	Person does NOT have strep throat	Person DOES have strep throat
Test says patient does NOT have strep throat	0.72	0.005
Test says patient DOES have strep throat	0.08	0.195

- b) What is the probability that the test will indicate that a patient has strep throat?

$$P(\text{test positive}) = 0.08 + 0.195 = 0.275$$

- c) Given that the test says that someone has strep throat, what is the probability that they do, in fact, have strep throat?

$$P(\text{have strep} | \text{test positive}) = \frac{P(\text{have strep AND test positive})}{P(\text{test positive})} = \frac{0.195}{0.275} = 0.709091$$

- d) Given that the test indicates that someone does NOT have strep throat, what is the probability that they actually do not have the disease?

$$\begin{aligned} P(\text{do not have strep} | \text{test negative}) &= \frac{P(\text{do not have strep AND test negative})}{P(\text{test negative})} \\ &= \frac{0.72}{0.725} = 0.993103 \end{aligned}$$

Please place your name on
Each page now

NAME: _____

- e) Of all the patients given the test at a particular clinic during a one month period, 200 tested negative for the disease (i.e., the test indicated that they did not have the disease). What is the probability that NONE of these 200 patients actually had the disease?

This is just a Binomial distribution:

$$\begin{aligned} P(0 \text{ have disease}) &= P(\text{all 200 who test negative are disease free}) \\ &= (0.993103)^{200} = 0.250552 \end{aligned}$$

Problem 2: (30 Percent)

Ten samples of 5 mm thick steel are tested to determine the actual thickness. The following data are recorded:

5.03	5.07	5.02	5.04
5.05	4.97	5.03	4.99
5.00	4.98		

Rearrange the data in ascending order to facilitate the rest of this problem.

4.97	4.98	4.99	5.00
5.02	5.03	5.03	5.04
5.05	5.07		

- a) Find the sample average

$$\bar{X} = \frac{\sum_{i=1}^{10} x_i}{10} = \frac{50.18}{10} = 5.018$$

- b) Find the sample variance

$$S^2 = \frac{\sum_{i=1}^{10} x_i^2 - 10(\bar{X})^2}{9} = \frac{251.8126 - 10(5.018)^2}{9} = 0.00104$$

Please place your name on
Each page now

NAME: _____

- c) Find the sample median

$$\tilde{X} = \frac{1}{2} \{x_{(5)} + x_{(6)}\} = \frac{1}{2} \{5.02 + 5.03\} = 5.025$$

- d) Find the first quartile (i.e., the value of Q_1) as well as the value of the third quartile (i.e., the value of Q_3).

$$X_{(2.75)} = X_{(2)} + (2.75 - 2) \{X_{(3)} - X_{(2)}\} = 4.98 + 0.75(4.99 - 4.98) = 4.9875$$

$$X_{(8.25)} = X_{(8)} + (8.25 - 8) \{X_{(9)} - X_{(8)}\} = 5.04 + 0.25(5.05 - 5.04) = 5.0425$$

- e) Find the range of the data

$$\text{Range} = 5.07 - 4.97 = 0.10$$

- f) Find the interquartile range.

$$\text{IQR} = Q_3 - Q_1 = 5.0425 - 4.9875 = 0.055$$

Problem 3: (40 Percent)

A tire replacement store records the mileage of tires that are replaced by the firm. They only record data from those customers who are having all four tires replaced for the first time and who have not needed to replace one or more tires due to tire damage. Thus, the data represent the ability of the tires to withstand normal wear and tear in driving conditions. The data are summarized below:

$$\text{Brand 1: } \sum_{i=1}^{100} m_i = 2,916,577$$

$$\sum_{i=1}^{100} m_i^2 = 114,949,427,970$$

$$\text{Brand 2: } \sum_{i=1}^{100} m_i = 3,138,778$$

$$\sum_{i=1}^{100} m_i^2 = 113,075,065,816$$

where m_i is the number of miles that tire set i lasts until it is replaced. Note that the store owner clearly only recorded one mileage number for any set of 4 tires. Thus, there were 100 cars that had brand 1 tires that were replaced and 100 cars with brand 2 tires that were replaced.

- a) Find the sample average tire life for brand.

Please place your name on
Each page now

NAME: _____

$$\text{Brand 1: } \bar{X} = \frac{\sum_{i=1}^{100} m_i}{100} = \frac{2,916,577}{100} = 29,165.77$$

$$\text{Brand 2: } \bar{X} = \frac{\sum_{i=1}^{100} m_i}{100} = \frac{3,138,778}{100} = 31,387.78$$

b) Find the sample variance of the tire life for each brand.

$$\text{Brand 1: } S^2 = \frac{\sum_{i=1}^{100} X_i^2 - \frac{\left[\sum_{i=1}^{100} X_i \right]^2}{100}}{99} = \frac{114949427970 - \frac{2916577^2}{100}}{99} = 301,870,848.49$$

$$\text{Brand 2: } S^2 = \frac{\sum_{i=1}^{100} X_i^2 - \frac{\left[\sum_{i=1}^{100} X_i \right]^2}{100}}{99} = \frac{113075065816 - \frac{3138778^2}{100}}{99} = 147,028,206.9$$

c) Find the sample standard deviation of the tire life for each brand.

$$\text{Brand 1: } S = \sqrt{S^2} = \sqrt{301,870,848.49} = 14,374.43$$

$$\text{Brand 2: } S = \sqrt{S^2} = \sqrt{147,028,206.9} = 12,125.52$$

d) Find the sample mean, variance and standard deviation for all 200 tires taken together.

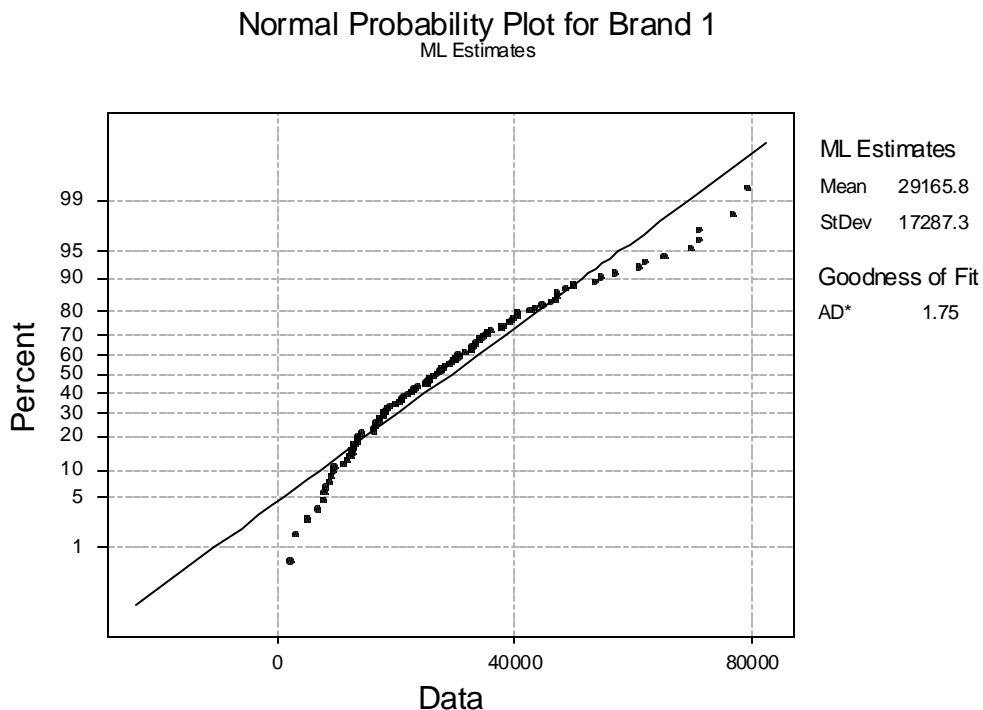
The key here is that the sum of the 200 observations is just the total for brand 1 plus the total for brand 2. Similarly, the sum of the squared values is just the sum of the squared values for brand 1 plus the sum of the squared values for brand 2. Thus,

Please place your name on
Each page now

NAME: _____

$$\begin{aligned}\sum_{i=1}^{200} m_i &= 2,916,577 + 3,138,778 = 6,055,355 \\ \sum_{i=1}^{200} m_i^2 &= 114,949,427,970 + 113,075,165,816 = 228,024,493,786 \\ \bar{X} &= \frac{\sum_{i=1}^{200} m_i}{200} = \frac{6055355}{200} = 30,276.775 \\ S^2 &= \frac{\sum_{i=1}^{200} m_i^2 - \frac{\left[\sum_{i=1}^{200} m_i \right]^2}{200}}{199} = \frac{228024493786 - \frac{6055355^2}{200}}{199} = 224,562,175 \\ S &= \sqrt{S^2} = \sqrt{224,562,175} = 14,985.4\end{aligned}$$

e) The figure below shows the Normal Probability Plot for Brand 1 Tires



Do the data appear to be Normally distributed? If not, what sort of transformation might make the transformed data Normally distributed?

The data do not look normally distributed. I would use a log transform.

Please place your name on
Each page now

NAME: _____

- f) Find the value of the number of miles that a set of tires must last for the z-scores for the two brands to be equal.

$$\frac{m - 29165.77}{17374.43} = \frac{m - 31387.78}{12125.52}$$
$$m = 36520.85$$