

NAME : MALIK FAIZAN AHMED

ID : 14313

Subject : Statistical Inference

Submitted to : Mam wajiha

SUMMERS 2020 FINAL ASSIGNMENT

QUESTION : 1

Q1(a) Suppose a sample of 50 obese patients on a low fat diet lost a mean of 5.5 pounds with a variance of 9 pounds, while another sample of 28 patients on low carb. diet lost a mean of 6.7 pounds with a variance of 16 pounds. Construct 80 % confidence interval for the mean difference of patients on two different diets.

ANSWER:

$$n_1 = 50, \quad n_2 = 28$$

$$X_1 = 5.5, \quad X_2 = 6.7$$

$$S_1 = 9, \quad S_2 = 16$$

$$C.I = 80\% \quad \alpha = 20\% = 0.2$$

$$\alpha/2 = 0.1$$

$$t_{\alpha/2} = 1.2928$$

$$S.P =$$

$$S.P =$$

$$S.P = 3.38$$

$$(X_1 - X_2) - t_{\alpha/2} S.p \leq \mu_1 - \mu_2 \leq (5.5 - 6.7) + (1.2928)$$

$$(-1.2) - (1.03) \leq \mu_1 - \mu_2 \leq (-1.2) + (1.03)$$

$$-2.23 \leq \mu_1 - \mu_2 \leq -0.17$$

(b) Differentiate between z test and t test.

ANSWER:

Z-Tests and T-Tests

Z-tests and t-tests are statistical methods involving data analysis that have applications in business, science, and many other disciplines. Let's explore some of their differences and similarities as well as situations where one of these methods should be used over the other.

Z-tests are statistical calculations that can be used to compare population means to a sample's. The z-score tells you how far, in standard deviations, a data point is from

the mean or average of a data set. A z-test compares a sample to a defined population and is typically used for dealing with problems relating to large samples ($n > 30$). Z-tests can also be helpful when we want to test a hypothesis. Generally, they are most useful when the standard deviation is known.

t-tests are calculations used to test a hypothesis, but they are most useful when we need to determine if there is a statistically significant difference between two independent sample groups. In other words, a t-test asks whether a difference between the means of two groups is unlikely to have occurred because of random chance. Usually, t-tests are most appropriate when dealing with problems with a limited sample size ($n < 30$).

QUESTION : 2

Q2:(a) A survey of 250 students indicated that 107 preferred coffee to tea. Determine 90% confidence interval for the proportion of students who preferred coffee.

ANSWER:

$n=250$, $x=107$. C.I=90% , $\alpha = 10\% = 0.1$, $\alpha/2 = 0.05$

$1-0.05 = 0.95$

$$P = x/n = 107/250 = 0.428$$

$P_{Z_{\alpha/2}}$

$$(0.428) \pm (1.65) \cdot \sqrt{0.428(0.572)/250} \approx 0.428 \pm 1.65$$

$$0.428 \pm (1.65) (0.0312) \approx 0.428 \pm (1.65) (0.0312)$$

$$0.376 \approx 0.479 \quad \text{ANS}$$

(b) Briefly discuss point estimate and interval estimate.

ANSWER:

Point Estimation:

A point estimation is a type of estimation that uses a single value, a sample statistic, to infer information about the population. Point estimation can be a sample statistic. The sample mean of age for the sample, 32, can be used as a point estimation. Point estimation is a single value that can be inferred as a population parameter. Let's discuss populations, parameters, and their relationship to point and interval estimations.

Interval Estimation:

Interval estimation, in statistics, the evaluation of a parameter—for example, the mean (average)—of a population by computing an interval, or range of values, within which the parameter is most likely to be located. Intervals are commonly chosen such that the parameter falls within with a 95 or 99 percent probability, called the confidence coefficient. Hence, the intervals are called confidence intervals, the end points of such an interval are called upper and lower confidence limits.

The interval containing a population parameter is established by calculating that statistic from values measured on a random sample taken from the population and by applying the knowledge (derived from probability theory) of the fidelity with which the properties of a sample represent those of the entire population.

The probability tells what percentage of the time the assignment of the interval will be correct but not what the chances are that it is true for any given sample. Of the intervals computed from many samples, a certain percentage will contain the true value of the parameter being sought.

QUESTION : 3

Q3(a) Given $\sigma = 1000$, $n = 100$ sample mean $(\bar{x}) = 870,000$, determine the confidence interval for 90% and 98% for mean.

ANSWER:

$$\begin{aligned} \sigma &= 1000, n=100, \bar{x}=870,000 \text{ C.I} = 90\% \text{ C.I}=98\% \\ 1-0.05 &= 0.95 \\ Z_{\alpha/2} &= 1.6 + 0.05 = 1.65 \\ X - Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} &\leq \mu \leq X + Z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \\ 870,000 - 1.65 (1000/10) &\leq \mu \leq 870,000 + 1.65 (1000/10) \\ 869,835 &\leq \mu \leq 870,165 \\ 870,000 - 2.33 (100) &\leq \mu \leq 870,000 + 2.33 (100) \\ 869,767 &\leq \mu \leq 870,233 \text{ answer} \end{aligned}$$

(b) State what happens to the size of confidence interval as level of confidence increases?

ANSWER:

Increasing the level of self-confidence increases the risk of error, making the confidence period wider. Lowering self-esteem reduces the risk of error, which in turn reduces confidence.

QUESTION : 4

Q4:(a) The management of a company is trying to determine annual family medical expenses of its employees. The company wishes to be 95% confident that the mean expenses to be correct within +\$50. A previous study indicates the standard

deviation of \$400. How large the sample size is required for the study?

ANSWER:

Explaining confidence interval in few lines ;

Confidence interval is a range of value we are sure our true values lie in .confidence interval, in statistics, refers to the probability that a population parameter will fall between two set values for a certain proportion of times. Confidence intervals measure the degree of uncertainty or certainty in a sampling method. A confidence interval can take any number of probabilities, with the most common being a 95% or 99% confidence level.

Confidence level refers to the percentage of probability, or certainty, that the confidence interval would contain the true population parameter when you draw a random sample many times. Or, in the vernacular, "We are 99% certain (confidence level) that most of these datasets (confidence intervals) contain the true population parameter.

QUESTION : 5

Q5: Fill in the blanks.

ANSWER:

(5.1) 0 and 1

(5.2) decreases

(5.3)hypothesis test

(5.4)A Confidence interval is a range of values that likely would contain an unknown population parameter. Confidence level refers to the percentage of probability, or certainty, that the confidence interval would contain the true population parameter when you draw a random sample many times.

(5) (a) 2.2

(B) 0.2

(C) 20

(6)(a) 8

(B) 4

(C) 81

(D) 80

(5.7) D-Point estimate

(5.8) standard error

(5.9) true

(5.10) 95%