/40

STAT 2509 A

Assignment#1

SOLUTION

[4] **1.** a) What are the important differences between a *parameter* and a *statistic*?

- **Parameter: is a descriptive measure of a population. It is a fixed constant.** (1)
- **Statistic: is a descriptive measure of a sample. It varies from sample to sample.** (1)

b) Explain the difference between a *population* and a *sample*.

- **Population: is a collection of all items of interest** (1)
- **Sample: is a subset of the units in a population** (1)

[1] **2.** Which of the following are measures of Central Tendency?

a) mean & the interquartile range IQR b) median & mode
c) standard deviation & 25th percentile d) mean, median and mode

d) mean, median and mode (1)

[7] **3.** Identify the following variables as : "*purely categorical" (or qualitative)*, "*categorical and ranked*", "*quantitative and discrete*" or "*quantitative and continuous*".

a) the number of students who get a final grade greater than 80% **quantitative and discrete** (1/2) (1/2)

b) dress size: 3, 5, 7, 9, 11, 13, 15, 17 **categorical (qualitative) and ranked** (1/2) (1/2)

c) political affiliation (Liberals, Conservatives, NDP, Green, Independent) **purely categorical (or qualitative)** (1)

d) taste ranking (excellent, good, fair, poor) **categorical (or qualitative) and ranked** (1/2) (1/2)

e) the door chosen by a rat in a maze experiment (A, B or C): **qualitative (or qualitative & ranked)** (1)

f) the winning time for a horse running at the racetrack **quantitative and continuous** (1/2) (1/2)

g) number of children in grade 5 who are reading at or above grade level **quantitative and discrete** (1/2) (1/2)

[6] **4.** Classify each of the following quantities as either a *parameter* or a *statistic*:

(i) $\bar{x}$ - **statistics** (1) (ii) σ^2 - **parameter** (1) (iii) μ - **parameter** (1)

(iv) s^2 - **statistics** (1) (v) β_1 - **parameter** (1) (vi) $\hat{\beta}_0$ - **statistics** (1)

[6] **5.** Find the following values from the tables:

a) $z_{0.025}$ = **<u>1.96</u>** (1)

b) $z_{0.975} = -\ z_{0.025} = -$ **<u>1.96</u>** (1)

c) $z_{0.03}$ = **<u>1.88</u>** (1)

d) $t_{9;0.05}$ = **<u>1.833</u>** (1)

e) $-t_{9;0.05}$ = $-$**<u>1.833</u>** (1)

f) $t_{9;0.95} = -t_{9;0.05} = -$**<u>1.833</u>** (1)

[6] **6.** Consider a normal population distribution with the value of σ known.

a) What is the confidence level for the interval

(i) $\bar{x} \pm 1.96\sigma/\sqrt{n}$ $\Rightarrow z_{\alpha/2}$ **= 1.96** $\Rightarrow \alpha/2 =$ **0.025** $\Rightarrow \alpha =$ **0.05** $\Rightarrow 1-\alpha =$ **0.95**

(1) $\therefore$ **<u>95% C.I.</u> for** μ

(ii) $\bar{x} \pm 1.15\sigma/\sqrt{n}$ $\Rightarrow z_{\alpha/2}$ **= 1.15** $\Rightarrow \alpha/2 =$ **0.125** $\Rightarrow \alpha =$ **0.25** $\Rightarrow 1-\alpha =$ **0.75**

(1) $\therefore$ **<u>75% C.I.</u> for** μ

(iii) $\bar{x} \pm 3.09\sigma/\sqrt{n}$ $\Rightarrow z_{\alpha/2}$ **= 3.09** $\Rightarrow \alpha/2 =$ **0.0010** $\Rightarrow \alpha =$ **0.002** $\Rightarrow 1-\alpha =$ **0.998**

(1) $\therefore$ **<u>99.8% C.I.</u> for** μ

b) What value of z in the confidence interval formula

$$\left(\bar{x} - z_{\alpha/2}\,\sigma/\sqrt{n}, \bar{x} + z_{\alpha/2}\,\sigma/\sqrt{n}\right)$$

results in a confidence level of

(i) 89.68% $\Rightarrow 1-\alpha =$ **0.8968** $\Rightarrow \alpha =$ **0.1032** $\Rightarrow \alpha/2 =$ **0.0516** $\Rightarrow z_{\alpha/2}$ **= <u>1.63</u>** (1)

(ii) 99.20% $\Rightarrow 1-\alpha =$ **0.9920** $\Rightarrow \alpha =$ **0.0080** $\Rightarrow \alpha/2 =$ **0.0040** $\Rightarrow z_{\alpha/2}$ **= <u>2.65</u>** (1)

(iii) 78.88% $\Rightarrow 1-\alpha =$ **0.7888** $\Rightarrow \alpha =$ **0.2112** $\Rightarrow \alpha/2 =$ **0.1056** $\Rightarrow z_{\alpha/2}$ **= <u>1.25</u>** (1)

[6] **7.** For any hypothesis test:

a) Define 2-sided and 1-sided hypotheses and give the steps involved in their testing.

- **2-sided hypothesis**: is a 2-tailed test testing that parameter $\neq$ value, e.g. (1)

$$H_0 : \mu = 0$$
$$H_a : \mu \neq 0$$

- **1-sided hypothesis**: is a 1-tailed test testing that parameter <value or >value, e.g.

$$H_0 : \mu \geq 0 \quad \textbf{or} \quad H_0 : \mu \leq 0$$
$$H_a : \mu < 0 \qquad H_a : \mu > 0$$

(1)

- **Steps involved**: 1) state the null and alternative hypotheses (1/2)
  2) test-statistics (1/2)
  3) rejection (or critical) region (1/2)
  4) conclusion (1/2)

b) What are the two types of error that may be made?

- **Type I error** = error we make when we reject H_0 when it is true.
  P[Type I error] = α (1)

- **Type II error** = error we make when we do not reject H_0 when it is false.
  P[Type II error] = β (1)

[4] **8.** Given that the population variance is defined by

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$$

where: N is the population total (i.e. total number of the observations in the population) and μ the population mean.

Show that
$$\sigma^2 = \frac{1}{N}\left(\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}\right)$$

Solution:

(1)

$$\sum_{i=1}^{N}(x_i - \mu)^2 = \sum_{i=1}^{N}\left(x_i^2 + \mu^2 - 2\mu x_i\right) = \sum_{i=1}^{N} x_i^2 + \sum_{i=1}^{N}\mu^2 - 2\mu\sum_{i=1}^{N} x_i =$$

$$\sum_{i=1}^{N} x_i^2 + N\mu^2 - 2\mu\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} x_i^2 + N\frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N^2} - 2\,\frac{\sum_{i=1}^{N} x_i}{N}\sum_{i=1}^{N} x_i =$$

2

$$\sum_{i=1}^{N} x_i^2 + \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N} - 2\,\frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N} = \sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}$$

1

i.e. $$\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2 = \frac{1}{N}\left(\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N}\right)$$ **Q.E.D.**