/47

STAT 2509 B

Assignment#1

SOLUTION

1. [1] The process of using information from a sample to draw conclusions about the entire population is called

(a) sampling (b) the scientific method
(c) statistical inference (d) descriptive statistics

(c) statistical inference 1

2. [1] A numerical measure computed to describe a characteristic of a population is called a

(a) parameter (b) statistic (c) sample (d) population

(a) parameter 1

3. [7] 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 1/2 **categorical (qualitative) and ranked** 1/2 1/2

c) weight of a newborn in kg **quantitative and continuous** 1/2 1/2

d) rating of a professor as: excellent, good, fair, poor 1/2 **categorical and ranked** 1/2

e) mark out of 100 obtained on a statistics test 1/2 **quantitative and continuous** 1/2

f) letter grade obtained on a statistics test **purely categorical (or categorical and ranked)** 1

g) number of chairs with green upholstery in a conference room 1/2 **quantitative and discrete** 1/2

4. [6] 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

5. Find the following values from the tables:

[6]

a) $z_{0.3015}$ = **0.52** 1

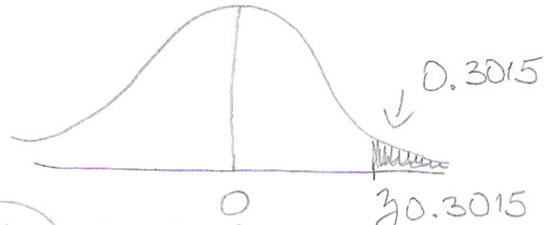
b) $z_{0.6985}$ = - $z_{0.3015}$ = **- 0.52** 1

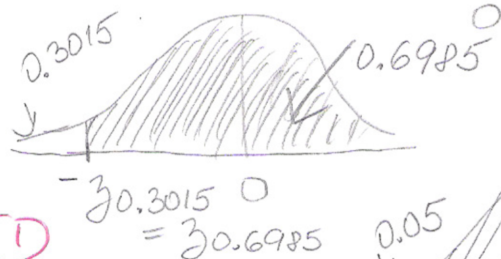
c) $z_{0.002}$ = **2.88** 1

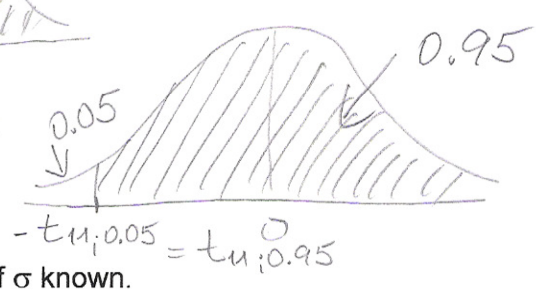
d) $t_{11;0.05}$ = **1.796** 1

e) $-t_{11;0.05}$ = **- 1.796** 1

f) $t_{11;0.95} = -t_{11;0.05}$ = **- 1.796** 1







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

[7]

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**

$\therefore$ **95% C.I. for** μ 1

(ii) $\bar{x} \pm 2.24\sigma/\sqrt{n}$ $\Rightarrow z_{\alpha/2}$ **= 2.24** $\Rightarrow \alpha/2 =$ **0.0125** $\Rightarrow \alpha =$ **0.025** $\Rightarrow 1-\alpha =$ **0.975**

$\therefore$ **97.5% C.I. for** μ 1

(iii) $\bar{x} \pm 3.09\sigma/\sqrt{n}$ $\Rightarrow z_{\alpha/2}$ **= 3.09** $\Rightarrow \alpha/2 =$ **0.0010** $\Rightarrow \alpha =$ **0.0020** $\Rightarrow 1-\alpha =$ **0.998**

$\therefore$ **99.8% C.I. for** μ 1

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** $\Rightarrow z_{\alpha/2}$ **= 1.63** 1

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

(iii) 75.40% $\Rightarrow 1-\alpha =$ **0.7540** $\Rightarrow \alpha =$ **0.2460** $\Rightarrow \alpha/2 =$ **0.1230** $\Rightarrow z_{\alpha/2}$ **= 1.16** 1

c) Would a 90% C.I. be narrower or wider than the 99.20% C.I. in b)?

1

90% C.I. would be narrower than 99.20% C.I. since 90% would have shorter span (i.e. it covers smaller interval of values)

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

[8]

- **2-sided hypothesis: is a 2-tailed test for testing parameter value $\neq 0$** (1)
  (e.g. $H_0 : \mu = 0$ vs. $H_a : \mu \neq 0$)

- **1-sided hypothesis: is a 1-tailed test for testing parameter value < 0** (1/2) **or > 0** (1/2)
  (e.g. $H_0 : \mu \leq 0$ vs. $H_a : \mu > 0$,
  or $H_0 : \mu \geq 0$ vs. $H_a : \mu < 0$)

- **steps involved: 1) state H_0 and H_a** (1)
  2) test-statistics (1)
  3) rejection (critical) region (1)
  4) conclusion (1)

b) For any hypothesis test, 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. (1)
P[Type I error] = α

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

8. If k is a constant and X and Y are random variables, then

[7]

a) (i) E(k)= **k**, (1) (ii) E(kX)= **$kE(X)$**, (1) (iii) E($X \pm Y$)= **$E(X) \pm E(Y)$** (1)

b) (i) V(k)= **0**, (1) (ii) V(kX)= **$k^2V(X)$**, (1) (iii) V($X \pm Y$)= **$V(X) + V(Y) \pm 2Cov(X,Y)$** (1)

Also show what happens when X and Y are independent of each other?

When X and Y are independent, then they are not related and so Cov(X,Y) = 0, i.e. $V(X \pm Y) = V(X) + V(Y)$ (1)

9. ANOVA method for a linear regression gives following:

[4]

$$\text{TSS} = \sum_{i=1}^{n}(y_i - \bar{y})^2,$$

where TSS is the total variation in the data. Show that

$$\text{TSS} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

Solution:

1

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}\left(y_i^{\,2} + \bar{y}^2 - 2\bar{y}y_i\right) = \sum_{i=1}^{n} y_i^2 + \sum_{i=1}^{n}\bar{y}^2 - 2\,\bar{y}\sum_{i=1}^{n} y_i =$$

$$\sum_{i=1}^{n} y_i^2 + n\bar{y}^2 - 2\,\bar{y}\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} y_i^2 + n\frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n^2} - 2\,\frac{\sum_{i=1}^{n} y_i}{n}\sum_{i=1}^{n} y_i =$$

2

$$\sum_{i=1}^{n} y_i^2 + \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} - 2\frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n} = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$

1

i.e. $$\text{TSS} = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}$$ **Q.E.D.**