/48

STAT 2509 C
Assignment#1
SOLUTION

1. [2] Choose a right answer:

a) The set of all objects or measurements of interest to the sample collector is
(i) a statistic (ii) a population (iii) a sample (iv) a parameter

(ii) a population 1

b) A numerical measure for a sample is referred to as
(i) a population (ii) a parameter (iii) a statistic (iv) a sample

(iii) a statistic 1

2. [1] Which one of the following summary measures is affected most by outliers (or extreme values)?
(i) the first quartile (ii) the median (iii) the mean (iv) the interquartile range IQR.

(iii) the mean 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 & discrete** 1/2 1/2

b) marital status of people: **purely categorical (or qualitative)** 1

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

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

e) average daily temperature during the month of January: **quantitative & continuous** 1/2 1/2

f) letter grade obtained on a statistics test: **purely categorical (or also can be quantitative & ranked)** 1

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

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

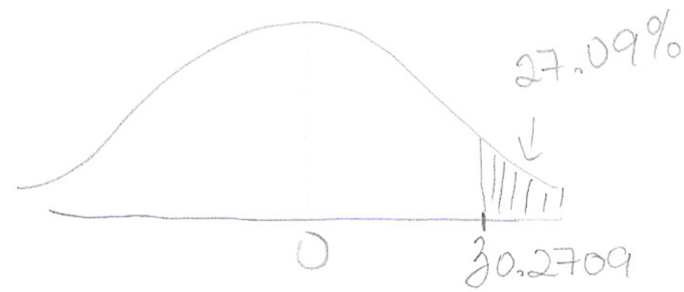
(i) $\bar{x}$ - **statistics** 1 (ii) σ^2 - **parameter** 1 (iii) μ - **parameter** 1

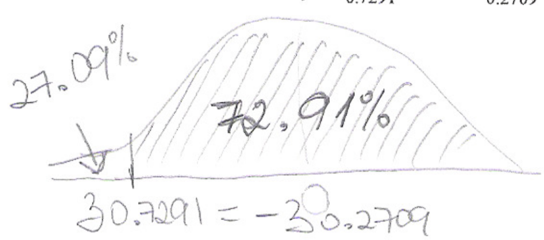
(iv) s^2 - **statistics** 1 (v) β_1 - **parameter** 1 (vi) $\hat{\beta}_0$ - **statistics** 1

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

a) $z_{0.2709}$ = **0.61** 1

b) $z_{0.7291}$ = - $z_{0.2709}$ = **- 0.61** 1



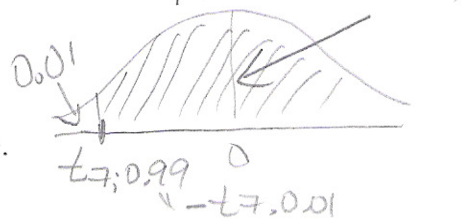


c) $z_{0.002}$ = **2.88** 1

d) $t_{7;0.01}$ = **2.998** 1

e) $-t_{7;0.01}$ = **- 2.998** 1

f) $t_{7;0.99} = -t_{7;0.01}$ = **- 2.998** 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) $\bar{x} \pm 1.15\sigma/\sqrt{n} \Rightarrow z_{\alpha/2}$ **= 1.15** $\Rightarrow \alpha/2 =$ **0.125** $\Rightarrow \alpha =$ **0.25** $\Rightarrow 1-\alpha =$ **0.75**

$\therefore$ **75% 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. For any hypothesis test:

[9]

a) Explain what the null and alternative hypotheses are.

H_a **(alternative hypothesis) - is the one we want to show; i.e. we are trying to find sufficient evidence for** 1

H_0 **(null hypothesis) – is negating the statement in** H_a**. It is a "fall-back" hypothesis. It tests against** H_a**.** 1

b) Write down the appropriate alternative hypotheses and give the formula for the each test statistic, if any, for the following null hypothesis testing situations.

(i) $H_0: \mu = \mu_0$ n = 20, s = 45 population normally distributed

1/2

$H_a: \mu \neq \mu_0$ **test-statistics:** $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ 1/2

(ii) $H_0: \mu \leq \mu_0$ n = 80, s = 29 population not normal

1/2 1/2

$H_a: \mu > \mu_0$ **test-statistics:** $t \cong z = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ **since Central Limit Theorem applies here (i.e. n = 80 is large** $\Rightarrow$ $\bar{X}$ **is approx. distributed with** $N\left(\mu, \dfrac{\sigma^2}{n}\right)$ **and we replace** σ^2 **with** s^2**)**

(iii) $H_0: \mu = \mu_0$ n = 15, σ = 25 population not normal

1/2 1/2

$H_a: \mu \neq \mu_0$ **no test-statistics can be used here, since population is not normal and n is small (i.e. n < 30), so C.L.T. does not apply, either**

(iv) $H_0: \mu = \mu_0$ n = 15, s = 36 population not normal

1/2

$H_a: \mu \neq \mu_0$ **no test-statistics can be used here, since population is not normal and n is small (i.e. n < 30), so C.L.T. does not apply, either**

1/2

(v) $H_0: \mu \geq \mu_0$ n = 10, σ = 16 population normal

1/2

$H_a: \mu < \mu_0$ **test-statistics:** $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ 1/2

(vi) $H_0: \mu \leq \mu_0$ n = 60, σ = 81 population normal

1/2

$H_a: \mu > \mu_0$ **test-statistics:** $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$ 1/2

(vii) $H_0: \mu = \mu_0$ n = 200, s = 25 population not normal

1/2

$H_a: \mu \neq \mu_0$ **test-statistics:** $t \cong z = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$ **since Central Limit Theorem**

1/2

applies here (i.e. n = 200 is large $\Rightarrow$ $\bar{X}$ **is approx. distributed with** $N\left(\mu, \dfrac{\sigma^2}{n}\right)$ **and we replace** σ^2 **with** s^2**)**

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

[7]

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

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

9. ANOVA method for a linear regression gives following:

[3]

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

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

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

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

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

1

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