//45

STAT 2509 A

Assignment #1

(Review of STAT 2507)

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

[1] **2.** 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

[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 leaves on a Maple tree **quantitative and discrete** 1/2 1/2
b) Province or Territory in which a person lives **purely categorical** 1
c) daily temperatures for a months of September 1/2 **quantitative and continuous** 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 **quantitative and continuous**
f) letter grade obtained on a statistics test **purely categorical (or categorical and ranked)** 1
g) number of milk bags bought by a family weekly **quantitative and discrete** 1/2 1/2

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

(i) $\bar{x}$ - **statistic** 1
(ii) σ^2 - **parameter** 1
(iii) μ - **parameter** 1
(iv) s^2 - **statistic** 1
(v) β_0 - **parameter** 1
(vi) $\hat{\beta}_1$ - **statistic** 1

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

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) $t_{11;0.025}$ = **2.201** (1)

e) $-t_{11;0.025}$ = − **2.201** (1)

f) $t_{11;0.975}$ = $-t_{11;0.025}$ = − **2.201** (1)

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

a) What is the confidence level for the interval

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

$\therefore$ **97.5% C.I. for** μ (1)

(ii) $\bar{x} \pm 2.58\,\sigma/\sqrt{n}$ $\Rightarrow z_{\alpha/2} = \mathbf{2.58} \Rightarrow \alpha/2 = \mathbf{0.005} \Rightarrow \alpha = \mathbf{0.010} \Rightarrow 1-\alpha = \mathbf{0.99}$

$\therefore$ **99% C.I. for** μ (1)

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

$\therefore$ **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% $\Rightarrow 1-\alpha = \mathbf{0.8968} \Rightarrow \alpha = \mathbf{0.1032} \Rightarrow \alpha/2 = \mathbf{0.0516} \Rightarrow z_{\alpha/2} = \mathbf{1.63}$ (1)

(ii) 97.96% $\Rightarrow 1-\alpha = \mathbf{0.9796} \Rightarrow \alpha = \mathbf{0.0204} \Rightarrow \alpha/2 = \mathbf{0.0102} \Rightarrow z_{\alpha/2} = \mathbf{2.32}$ (1)

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

c) Would a 90% C.I. be narrower or wider than the 89.68% C.I. in b)?

(1)

90% C.I. would be wider than 89.68% C.I. since 90% would have wider span (i.e. it covers larger interval of values) as it has larger error $z_{\alpha/2}\,\sigma/\sqrt{n}$.

[8] **7.** 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.** (1/2) (1/2)

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

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

[7] **8.** If k is a constant and X and Y are random variables, then

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)

[2] **9.** 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:

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

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

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

2

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.**