STAT 2509 C

Assignment #1

(Review of STAT 2507)

DUE: January 20th, 2015 (to be handed in during the class)

1. 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

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

2. 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.

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%
b) marital status of people
c) weight of a newborn in kg
d) rating of a professor as: excellent, good, fair, poor
e) average daily temperature during the month of January
f) letter grade obtained on a statistics test
g) number of chairs with green upholstery in a conference room

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

(i) $\bar{x}$ (ii) σ^2 (iii) μ (iv) s^2 (v) β_1 (vi) $\hat{\beta}_0$

5. Find the following values from the tables:

a) $z_{0.2709}$ b) $z_{0.7291}$ c) $z_{0.002}$ d) $t_{7;0.01}$ e) $-t_{7;0.01}$ f) $t_{7;0.99}$

(Note that the first subscript of t is the degree of freedom of t distribution)

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

a) What is the confidence level for the interval
(i) $\bar{x} \pm 1.96\,\sigma/\sqrt{n}$ (ii) $\bar{x} \pm 2.24\,\sigma/\sqrt{n}$ (iii) $\bar{x} \pm 1.15\,\sigma/\sqrt{n}$

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% (ii) 99.20% (iii) 75.40%

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

7. For any hypothesis test:

a) Explain what the null and alternative hypotheses are.
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
(ii)	$H_0: \mu \leq \mu_0$	n = 80, s = 29	population not normal
(iii)	$H_0: \mu = \mu_0$	n = 15, σ = 25	population not normal
(iv)	$H_0: \mu = \mu_0$	n = 15, s = 36	population not normal
(v)	$H_0: \mu \geq \mu_0$	n = 10, σ = 16	population normal
(vi)	$H_0: \mu \leq \mu_0$	n = 60, σ = 81	population normal
(vii)	$H_0: \mu = \mu_0$	n = 200, s = 25	population not normal

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

a) (i) E(*k*)=?, (ii) E(*kX*)=?, (iii) E(*X*± *Y*)=?
b) (i) V(*k*)=?, (ii) V(*kX*)=?, (iii) V(*X*± *Y*)=? Also show what happens when *X* and *Y* are independent of each other?

9. ANOVA method for a linear regression gives following:

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

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

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