STAT 2509A Assignment 1:

Lab. Section: STAT 2509-A3
Name: Connor Raymond Stewart
Student ID: 101041125

① Choose a right Answer:

a) A Numerical Measure for a Population is refered to as:
ii) a Parameter

b) A Numerical Measure for a Sample is Refered to as:
iii) a Statistic

② Which of the Following are Measures of Central Tendency:

— Measures of Central tendency are measures along the horizontal axis which locate the centre of the distribution:

- thus, the mean, median, & mode all are ways of finding the Centre.

So, the Answer is $\boxed{\text{b) mean \& median}}$

- Note that Std. dev. measures spread, Not Centre & the range measures spread, not Centre; thus both a & c both <u>Can't</u> be the Correct Solution.

③ Identify the Following as: qualitative, Categorial & ranked, quantitative & discrete, or quantitative & Continuous:

a) Mecury Concentration in a Sample of Tuna:
Quantitative & Continuous
Ex.: 20.2mg Mercury Per. 1 ton of Tuna is a Countable & Continuous quantity

b) Fast-food Establishment Preferred by a Student (McDonald, Burger King, A&W):
The fast food establishments Cannot be placed in a rank as the Students Choice doesn't mean anything w/ reguards to the other options
$\boxed{\therefore \text{Purely Categorical}}$

c) Score (0-100) on a Placement Examination:
There are only So many Scores & Partial marks a Student Can get on an exam:
$\boxed{\therefore \text{Quantitative \& Discrete}}$

d) Taste Ranking (excellent, good, fair, poor):
Since the choices are all relative to each other, values Can be assigned to each, i.e. excellent is one & Poor is 4:
$\boxed{\therefore \text{Categorical \& Ranked}}$

e) Colour of Rose bush:
Since one Colour means nothing to another, there's no rank:
∴ Purely Categorical

f) The Number of Defective Lightbulbs in a Package of 4 Bulbs:

There's a finite number of bulbs, & the # of defective bulbs is a countable quantity:

$\therefore$ Quantitative & Discrete

g) Dress Size: 3,5,7,9,11,13,15,17:

There's a finite # of dress sizes from the above list:

$\boxed{\therefore \text{ Quantitative \& Discrete}}$

④ Classify Each of the following as either a parameter or a statistic:

i) $\bar{x}$:

$\bar{x}$ represents the sample mean $\left(\bar{x}=\frac{\Sigma x_i}{n}\right)$, that is the mean of the sample in question:

Since this is from the sample measure, it's a Statistic

ii) σ^2:

σ^2 represents the population varience $\left(\sigma^2=\frac{\Sigma(x_i-\mu)^2}{N}\right)$:

$\therefore$ It's a Parameter

iii) μ:

μ represents the Population mean:

$\therefore$ It's a Parameter

iv) s^2:

s^2 represents the sample Variance $\left(s^2=\frac{\Sigma(x_i-\bar{x})^2}{n-1}\right)$

$\therefore$ It's a Statistic

V) B_1:

B_1 represents the slope of a linear equation $(y=B_0+B_1x)$

the equation's based of a trend over a whole population, rather then it being an estimate,

$\therefore$ It's a Parameter

~~vi) $\hat{B}_0$~~

Vi) $\hat{B}_0$: $\hat{B}_0$ is an estimated sample-based least squares result $(\hat{B}_0=\bar{y}-\hat{B}_1\bar{x})$

$\bar{x}$ & $\bar{y}$ are sample means, like in i) above, thus $\hat{B}_0$ is a sample quantity

$\therefore$ It's a Statistic

⑤ Find the following values from the tables:

a) Due to mound-shaped dist., use $1-0.025=0.975$

from 0 on the z-table, look to row 20 column 7

thus, $\boxed{z_{0.025}=1.96}$

b) Since $z_{0.975}=-z_{0.025}$ we know $z_{0.975}=-(1.96)=-1.96$

We can find this in the table w/:

find $1-0.975=0.025$

From 0 on the Z-table, look up 20 rows on Column 7 Hilroy thus, $\boxed{Z_{0.975} = -1.96}$

both Answers result in the Same Sol$^{\underline{n}}$

c) Since $Z_{0.05}$, we use $1-0.05=0.95$

from 0 on the z-table, look down 17 Rows and at Columns 5 & 6

We can take the midpoint between 0.9495 & 0.9505 So,

$\boxed{Z_{0.05} = 1.645}$

d) $\underline{t_{0.10;4}:}$ (assuming one-tailed, unspecified)

df=4 & $t_{0.10}$ So Check row 4 Column $t_{.100}$:

$\boxed{t_{0.10;4} = 1.533}$ (2.132 if two-tailed)

e) $\underline{-t_{0.10;4}:}$

df=4 & $t_{0.1}$ So like above but w/ negative Sign:

$\boxed{-t_{0.1;4} = -1.533}$

f) $\underline{t_{0.90;4}:}$

df=4 & $t_{0.90}$ So check row 4 column's $t_{0.100}$ & $t_{0.050}$

Since $t_{0.90}$ is greater then $t_{0.5}$ & the t-dist. is roughly mound-Shaped:

$\boxed{t_{0.90;4} = -t_{0.10;4} = -1.533}$

6) If K is a Constant & X & y are Random Variables, then:

a)

i) $E(K) = K$

↳ the expected value of Constant K is the value K

ii) $E(KX) = KE(X)$ [Such that $E(x) = \mu_x$ So $KE(x) = K\mu_x$]

↳ the expected value of a random variable is the variables mean

iii) $E(x \pm y) = E(x) \pm E(y)$

↳ Such that $E(x) = \mu_x$ & $E_y = \mu_y$ So $E(x) \pm E(y) = \mu_x \pm \mu_y$]

b)

i) $V(K) = 0$

↳ the varience of a Constant must be zero as theirs No variation

ii) $V(KX) = K^2 V(X)$

↳ Varience Comes in Squared units

iii) $V(x \pm y) = V(x) + V(y) \pm 2Cov(X, y)$

7. Consider a Normal Population Distribution w/ the Value of σ Known:

a) What is the Confidence level for the Interval:

i) $\alpha = 1 - x$ & $Z_{\alpha/2} = 1.15$ $[\bar{x} \pm 1.15\, \sigma/\sqrt{n}]$

by mound-Shaped dist.

Z is -1.15 When $Z_{0.1251}$ So $\alpha/2 = 0.8749 \rightarrow \alpha/2 = 0.1251 \rightarrow \alpha = 0.2502$

So, $-0.2502 = 1 - x$ thus:

x is -0.7498 or Simply, x is 74.98% Confidence

ii) $\alpha = 1 - x$ & $Z_{\alpha/2} = 2.58$ $[\bar{x} \pm 2.58\, \sigma/\sqrt{n}]$

take complement

Z is -2.58 When $Z_{0.0049}$ So $\alpha/2 = 0.9951 \rightarrow \alpha = 0.0049(2) = 0.0098$

So, $0.0098 = 1 - x$ thus:

x is -0.9902 or Simply, x is 99.02% Confidence

iii) $\alpha = 1 - x$ & $Z_{\alpha/2} = 3.09$ $[\bar{x} \pm 3.09\, \sigma/\sqrt{n}]$

from mound-shaped dist

Z is -3.09 when $Z_{0.0010}$ So $\alpha/2 = 0.999 \rightarrow \alpha = 0.001(2) = 0.002$

So, $0.002 = 1 - x$ thus:

x is 0.998 or Simply, x is 99.8% Confidence

b) What value of Z in the Confidence interval Formula:

$(\bar{x} - Z_{\alpha/2}\, \sigma/\sqrt{n},\ \bar{x} + Z_{\alpha/2}\, \sigma/\sqrt{n})$ Results in a Confidence level of

i) 89.68%

$\alpha = 1 - 0.8968 = 0.1032$ So $\alpha/2$ is 0.0516

$Z_{0.0516} \Rightarrow -1.63$ Corrosponds to 0.0516 thus $Z_{0.0516} = 1.63$

$\boxed{\therefore \text{ the Value of } Z \text{ is } 1.63}$

ii) 99.20%

$\alpha = 1 - 0.992 = 0.008$ So $\alpha/2$ is 0.004

$Z_{0.004} \Rightarrow -2.65$ Corresponds to 0.004 thus $Z_{0.004} = 2.65$

$\boxed{\therefore \text{ the value of } Z \text{ is } 2.65}$

iii) 75.40%

$\alpha = 1 - 0.754 = 0.246$ So $\alpha/2$ is 0.123

$Z_{0.123} \Rightarrow -1.16$ Corrosponds to 0.123 thus $\boxed{Z_{0.123} = 1.16}$

$\boxed{\therefore \text{ the value of } Z \text{ is } 1.16}$

c) Would a 90% Confidence Interval be narrower or wider than the 99.20% Confidence Interval in b), Why?

The 90% Interval Would be narrower then the 99.2% Interval

This is because the higher the Confidence, the Larger the range of Possible outcomes is. A lower interval must have a more narrow range of values as there must be less outcomes to Choose from.

How would we Make 99.2% C.I. of the Same width as the 90% C.I.

- we Could increase the # of Samples, n, or decrease the std. deviation, σ
- Since the C.I. is $(\bar{x} + Z_{\alpha/2}\, \sigma/\sqrt{n},\ \bar{x} - Z_{\alpha/2}\, \sigma/\sqrt{n})$, the range of $\bar{x} \pm Z_{\alpha/2}\, \sigma/\sqrt{n}$ decreases proportionally as $(\sigma/\sqrt{n})$ decreases, which occurs when we do the above to n & σ.

⑧ For any Hypothesis Test:

a) Explain what the Null & Alternative Hypothesis are:

Null Hypothesis:
This is the Contradiction to the alternative hypothesis
It assumes the opposite of what the researcher is predicting
It states there's no significent difference between the populations
Essentally, if the data's relation's weak then the null hypothesis is chosen.
Since it is easier to make a false positive over a false negative, the null hypothesis is always chosen when the relations below the Confidence threashold. This means Ho gets the benifet of dought.

Alternative Hypothesis:
This is the hypothesis the researcher is attempting to support
This holds the position that somthing new/different is happening in reguards to the null hypothesis.
Since false positives are easily made, Ha must be proven w/in a specified Confidence (P) value, meaning unlike Ho, Ha doesn't have the benefit of dought.
So, we must reject Ho to accept Ha, but Ha dosn't need to be rejected to default to Ho.

b) Write down the alt. hypothesis & give the formula for each test statistic, if any, for the following null hypothesis testing situations:

i) $\boxed{H_a: \mu \neq \mu_0}$

Since $n<30$ & the Pop. is normally distributed, we may use the t-test:

$$\boxed{t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}}$$

Also, σ is unknown, so no z-test

ii) $\boxed{H_a: \mu > \mu_0}$

Since $n>30$ & the pop. is not normally dist. & σ is unknown:
we may use the t-test:

$$\boxed{t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}}$$

iii) $\boxed{H_a: \mu \neq \mu_0}$

The Population is not normally distributed & $n<30$ w/ S unknown:

$\boxed{\text{No Statistic Can be used for this}}$

(Hypergeometric, Not t/z)

iv) $\boxed{H_a: \mu \neq \mu_0}$

The pop. is not normally dist., $n<30$, & σ is unknown:

No Statistic Can be used for this

(hypergeometric, Not t/z)

v) $\boxed{H_a: \mu < \mu_0}$

The pop. is normally dist., $n<30$, & σ is Known, we may use the z-test:

$$\boxed{Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}}$$

vi) $\boxed{H_a: \mu > \mu_0}$

The pop.'s normally dist., $n>30$, σ is Known, so we may use the z-test:

$$\boxed{Z = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}}}$$

vii) $\boxed{H_a: \mu \neq \mu_0}$

The pop.'s not normally dist., $n>30$, & σ is unknown, so we may use the t-test:

$$\boxed{t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}}$$

9 ANOVA Method for linear Regression Gives the Following:

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

Where TSS is the total Variation in the data. Show that:

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

Since TSS measures the total Sum of Squares, it consists of 2 parts:

I) $\sum_{i=1}^{n}(y_i^2)$: this is the Sum for all the y values squared

II) $\frac{(\sum_{i=1}^{n} y_i)^2}{n}$: this is called the Correction for the mean

I tells us the Sum of the Squares & II Corrects for the mean this lets us find the Sum of the Squares of the difference between the dependent var. & its mean.

We Can Prove $TSS = \sum_{i=1}^{n} y_i^2 - \frac{(\sum_{i=1}^{n} y_i)^2}{n}$:

$$TSS = \sum_{i=1}^{n}(y_i-\bar{y})^2 \quad \& \quad \bar{y} = \frac{\sum y_i}{n}$$

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

for: $-2\sum_{i=1}^{n} y_i\bar{y} + \sum_{i=1}^{n}\bar{y}^2$

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

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

For $\frac{1}{n}\sum_{i=1}^{n}\left[\sum_{i=1}^{n} y_i \sum_{i=1}^{n} y_i\right]$:

let $\left[\sum_{i=1}^{n} y_i \sum_{i=1}^{n} y_i\right] = K$ then:

$$\frac{1}{n}\sum_{i=1}^{n} K = \cancel{\frac{1}{n}}(K\cancel{n}) = K$$

thus:

$$\frac{1}{n}\sum_{i=1}^{n}\left[\sum_{i=1}^{n} y_i \sum_{i=1}^{n} y_i\right] = \sum_{i=1}^{n} y_i \sum_{i=1}^{n} y_i$$

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

$$= -\frac{\sum_{i=1}^{n}(y_i)\sum_{i=1}^{n}(y_i)}{n}$$

So:

$$\sum_{i=1}^{n} y_i^2 - 2\sum_{i=1}^{n} y_i\bar{y} + \sum_{i=1}^{n}\bar{y}^2$$

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

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

therefore:

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

thus, it is proven that TSS equates to the above form