Name _____
Date _____ Period __

SELECTED TERMS, SYMBOLS AND FORMULAS FOR BIO-STATISTICS

Population - All of the individuals about which inferences are to be made.
Sample - A limited number of observations out of an entire population.
Variable - The actual property measured by individual observations.
$\sum$ - The capital Greek sigma. Means the sum of the items indicated.
df - Degrees of freedom
n - The sample size. The number of observations or measurements in a sample.
x - An individual observation or measurement in a sample.
Mean (x) - The average = sum of all the data divided by the total number of data points.
Median - The middle measurement or value in a set of observations or measurements.
Mode - The value that occurs the most often in a set of observations or measurements.
Maximum - The largest value in a set of observations or measurements.
Minimum - The smallest value in a set of observations or measurements.
Range - The difference between the maximum and minimum values in a set of observations.
Frequency (f) - The number of individual occurrences of a particular observation.
Frequency Distribution - An arrangement of data to show the frequency of each observation.
Normal Curve - (Normal Distribution) A symmetrical, bell-shaped curve in which the mean, mode and median are all at the same point.
Null Hypothesis - The hypothesis that assumes that there will be no difference between control and experimental data.
Standard Deviation (s) - A measure of the variability (spread or dispersion) that exists within a set of observations or measurements. It expresses the average deviation from the mean that exists in a sample. The larger the standard deviation the greater the variability of the data and the greater the error of the estimate. In any set of measurements from a normal population the following relationships hold:

(x) +/- one (s) contains 68.26% of the items
(x) +/- two (s) contains 95.46 % of the items
(x) +/- three (s) contains 99.73%

Standard Deviation

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

conversely,

50% of the items fall between (x) +/- 0.67 (s)
95% of the items fall between (x) +/- 1.96 (s)
99% of the items fall between (x) +/- 2.58 (s)

Statistics, statistics, I love statistics!

Measures of Central Tendency (Mean, Median, Mode)

1) Determine your resting heart rate (beats/minute) by finding your pulse, counting the number of beats for 15 seconds and multiplying by 4. Show calculations below:

2) Write down the class data for boys and girls separately in the table below.

Resting Heart Rates	
Boys	

Resting Heart Rates	
Girls	

3) Calculate the mean, median, and mode for boys and girls separately. Show work!

<u>Boys</u>

Mean: _____

Median: _____

Mode: _____

<u>Girls</u>

Mean: _____

Median: _____

Mode: _____

Measures of Variability (range, standard deviation)

4) Use the heart rate data already collected to determine the range and standard deviation for both boys and girls separately. Show work!

Boys	Girls
Range: _____	Range: _____
Standard Deviation: _____	Standard Deviation: _____

Standard Deviation Calculation Steps:

a) Calculate the sample mean
b) Construct a table to determine the sum of (deviation)^2
c) Use the standard deviation formula to solve for "s"

The Null Hypothesis and Statistical Tests of Significance

How does a researcher make a research hypothesis?

a) Develop a research question: Is there a difference in effectiveness between Method A and Method B for teaching statistics to high school students?

b) Restate it as a research hypothesis: There will be a difference between the effectiveness of Method A and the effectiveness of Method B for teaching statistics to high school students.

The statistician is unable to determine probabilities regarding two populations. He/she can only deal with one normal distribution of which he knows all the properties (mean, standard deviation). He/she must restate the hypothesis in a form that can be analyzed using only one normal distribution.

c) must be restated as the null hypothesis: There is no difference in effectiveness between Method A and Method B for teaching statistics to high school students.

To the statistician this is the same as saying that the mean score of Group A and the mean score of Group B came from the same population of mean scores, and the difference between them is due to sampling error.

You must decide between two hypotheses:

a) H_1 (research hypothesis) - there is a difference between the groups

b) H_o (null hypothesis) - there is no difference between the groups

Question and Hypothesis Writing Practice

For each case below write:

a) a research question
b) a research hypothesis (H_1)
c) the null hypothesis (H_0)

1) You are interested in finding out if plants grow better in two different colored lights. You have a variety of colored lights and 300 pea plants to work with.
   a)

   b)

   c)

2) A parent group wants to know if the **SALA** (Slingerland method) approach to learning language arts works better with first graders than the **"See It, Say It"** method.
   a)

   b)

   c)

3) A cookie company wants to know what color package makes their cookies most appealing to shoppers.
   a)

   b)

   c)

Now what? Use a Test of Significance!

Tests of significance are used to determine if the observed difference between control and experimental data is significant or should be attributed to chance or normal variation. Examine an example of the effectiveness two different methods of teaching statistics has on learning:

	Method A	Method B
# of students (n)	50	50
mean	75	78
standard deviation (s)	7	8

It appears as though Method B is superior to Method A because the sample mean for B is higher. But these are not population means and from our previous discussions we know that sampling error may be involved.

If we took two random samples from the same population the sample means would be different because of sampling error. The question now becomes "how much of a ***difference*** in means we need to assume the two means are from ***different*** populations."

Statisticians ask "can the difference between our two sample means be attributed to random error in our sampling, or do students taught by one method actually learn more than those taught by the other method, so that we are in effect dealing with two different populations?"

Reject or NOT Reject – That is the Question!

The statistician wants to make a decision whether or not to reject the null hypothesis. He/she decides not to reject the null hypothesis if there is a high probability that the difference between the two sample means could have resulted from sampling error. He/she decides to reject the null hypothesis if there is a low probability that the difference between the two sample means could have resulted from sampling error (the sample means did not come from the same distribution of sample means and thus a real difference exists between the test scores of students receiving Method A and Method B).

Remember, a decision **NOT** to reject the null hypothesis means only that the data obtained in our samples were not sufficiently different to enable us to conclude that it didn't happen by chance. (We failed to detect a significant difference between the two groups!!)

T-test - used when the total # of subjects (n) < 30

Steps:

a) Calculate the **degrees of freedom** (df):

df = (# of experimental subjects -1) + (# of control subjects -1)

b) Determine the **"table" t-value** by using the **Table t-distribution chart** at either 0.05 or 0.01 level of significance (depending on what is asked) and find the number that corresponds to your calculated degrees of freedom

c) Use the data to solve for the **"calculated" t-value** using the following formula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{{s_1}^2}{n_1} + \frac{{s_2}^2}{n_2}}}$$

d) Compare calculated "t" with table t-value by plugging the **"table" t-value** and **"calculated" t-value** into the equation below:

(-) "table" t-value < "calculated" t-value < (+) "table" t-value

- ✓ If this is TRUE than DO NOT REJECT the H_0.
- ✓ If this is **NOT** true than **REJECT** the H_0. (Meaning there is no difference between your two sets of data).

Example Problem

Research Question: Is there is a significant difference between the salaries of Wage A and Wage B employees at the "Genes R Us" Corporation?

* Use 0.05 level of significance

	Wage A	Wage B
number in sample (n)	10	17
mean	55	60
standard deviation (SD)	4.5	4.2

Write the Null Hypothesis (H_0):

Calculations (show formula):

"Table" t-distribution chart

95% confidence ↓

df	0.05	0.01
1	12.706	63.657
2	4.303	9.925
3	3.182	5.841
4	2.776	4.604
5	2.571	4.032
6	2.447	3.707
7	2.365	3.499
8	2.306	3.355
9	2.262	3.250
10	2.228	3.169
11	2.201	3.106
12	2.179	3.055
13	2.160	3.012
14	2.145	2.977
15	2.131	2.947

95% confidence ↓

df	0.05	0.01
16	2.921	2.120
17	2.898	2.110
18	2.878	2.101
19	2.861	2.093
20	2.845	2.086
21	2.831	2.080
22	2.819	2.074
23	2.807	2.069
24	2.797	2.064
25	2.787	2.060
26	2.779	2.056
27	2.771	2.052
28	2.763	2.048
29	2.756	2.045
30	2.750	2.042

Conclusion: