STAT 202 Name:

Lab #1 – Sampling Distributions and Probabilities

**SAMPLING DISTRIBUTIONS**

Use the data "internetandlife" from my website. This consists of variables "Life Expectancy" (in years) and "Internet Users" (per 100 people) for 182 countries.

**I. First study the population:**

1. Make a histogram of the `InternetUsers` and `LifeExpectancy` variable. These are the *population distributions* of the variable Describe this distribution.

2. Using StatCrunch (Stat/Summary Stats), compute the mean, variance, standard deviation, median, minimum, and maximum of the number of hours the variables. Round the results to two decimal places.

| | |
|---|---|
| Mean = | Median = |
| Variance = | Minimum = |
| Standard deviation = | Maximum = |

**II. Now study one random sample.**

3. Take ONE random sample of size 20 from the population of students. To do this, go to the `Data` menu and select `Sample columns`. Highlight all four rows. Enter 20 for the `sample size` and 1 for the `number of samples`. Enter 30 for `use single fixed seed` (this allows me to replicate your results). Be sure to click on `Sample all columns at one time,` then click on `Compute!`. A window opens telling you that a new columns have been added. What are the first 5 countries in your sample?

4. Describe the histogram of `Sample(LifeExpectancy) & Sample(InternetUsers).`

5. Find the mean, variance, standard deviation, median, minimum, and maximum of `Sample(TV). T`hese are *sample statistics*.

| | |
|---|---|
| Mean = | Median = |
| Variance = | Minimum = |
| Standard deviation = | Maximum = |

6. How do the sample statistics compare to the population values in (3) above?

**III. Explore Multiple Samples from the Population – The Sampling Distribution**

7. Now take 1000 different random samples of size 20 from the population of students. To do this, go to the Data menu and select Sample columns. Highlight the variables Enter 20 for the sample size and 1000 for the number of samples. Select Stacked with a sample id & Sample all columns at one time. Enter 30 for use single fixed seed. Click on Sample Column(s). A window opens telling you that two new columns have been added to the table (Sample(TV) and Sample).

8. Find the mean of InternetUsers & LifeExpectancy within each of these samples. To do this, go to the Statistics menu and select Summary Stats → Columns. Highlight the second listing of InternetUsers & LifeExpectancy and under Group by choose Sample. (This column identifies which of the 1000 samples the value is a member.) Click on Next and highlight only Mean. (Deselect the others.) Click on Store in data table. Click Calculate. Then click cancel when it asks you to bin. You should have two new columns in the data table, one corresponding to the Mean within each of these samples, the other corresponding to the sample number.

9. Describe the histogram of the Mean variable for each of InternetUsers & LifeExpectancy. Plot a QQ plot to see how close it is to normal.

10. Find the mean and variance of the Mean variables.