Computer Assignment 3

Note well: If not submitted, top quiz score will be omitted from the quiz average. In order to facilitate the learning of SPSS, collaboration is acceptable; it is indeed encouraged. But, you must be the author of your entire report.

1 Purpose

By using SPSS, we will create sampling distributions and, by way of the Central Limit Theorem, a set of 20 confidence intervals. This assignment will show you the random nature of confidence intervals and will demonstrate the very important concept that $100(1-\alpha)\%$ of them will, on the average, contain the population mean. Of course, in general, we cannot identify which confidence intervals contain the mean, but in this assignment we will follow a procedure by which the true population mean will be known so as to make it possible for you to count exactly how many of the confidence intervals contain the mean.

To the end of constructing such a set of confidence intervals, we will first generate 20 sets of 30 random numbers which follow a known probability distribution and, therefore, we can find, as you have done numerous times, the expected value and variance of the associated random variable. We will then pretend that each set of random numbers is actually a set of experimental data of a population with unknown mean and variance. Your computed 20 confidence intervals will be plotted on a graph, and you will count how many of them contain the theoretical mean.

2 Generate Random Numbers Using SPSS

1. Invoke SPSS
2. Under the `Data` menu select `Insert Variable`. Call it `v1`.
3. Select `Insert Case` from the `Datat` menu. A missing value symbol "." should appear in the first cell under variable `v1`. Highlight that cell, copy, and paste it in 29 highlighted cells under `v1` so that 30 empty cells will exist under `v1`.
4. Under the `Transform` menu, select `Compute`. In the resulting screen, put `v1` in the box labeled `Target Variable`. In the list of functions, scroll down and select `RV.UNIFORM`. (Press the arrow boxes in the window with your cursor). We will generate 30 random numbers from a $U(a, b)$ by means of this SPSS function. **Each student will have his own personal uniform distribution to use. Refer to your GWID number**. For `a` use the minimum digit in the last four digits of your GWID. For `b`, use the maximum digit in the last four digits of your GWID. Insert these minimum and maximum values into the question marks which appear in the function that you selected.

5. After you clicked `OK`, acknowledge the `change existing variable` warning, so that 30 numbers will appear under `v1`. You will notice that they are decimal numbers. The uniform distribution which SPSS provides is the one for a continuous random variable. Let us generate random numbers for a discrete random variable instead as if we are tossing a fair die.

6. Go back to `Compute` in the `Transform` menu. Put your cursor in the box where `RV.UNIFORM` is shown and type in the box the following command:
   `RND(RV.UNIFORM(a,b))` where `a` and `b` are the numbers from your GWID. The function `RND` will round the uniform random numbers to produce integers. However, we need to compensate for round-off error on the end points; so type again so that you will see: `RND(RV.UNIFORM(a-.4999,b .4999))` where `a` and `b` came from your GWID. Click `OK`, acknowledge that you will overwrite the values for `v1` to produce 30 random numbers from a U(a,b) discrete distribution.

7. Generate a second column of random numbers like before but call the `Target Variable` `v2` (everything in the function box can be left the same).

8. Continue the process until you have generated 20 columns of 30 random numbers from a discrete uniform distribution. Choose any variable, like `v17`. Make a picture of the set of 30 random numbers by selecting `Histogram` under the `Graph` window and by specifying a variable of your choice, like `v17`. Any one of the 30 variables may be selected.

   Answer the following theoretical questions.

Question 1 : *Why is the histogram not perfectly flat?*

Question 2 : *Let V be a random variable which is distributed as a discrete uniform distribution beginning at a and ending at b where a and b are the same numbers which you chose in step 4. Compute the expected value and variance of V. Note:* $E[V] = \frac{a+b}{2}$ *and* $var(V) = \frac{(b-a)(b-a+2)}{12}$

Question 3 : *Define* $\bar{V} = \frac{\sum_{i=1}^{20} V_i}{20}$. *Compute the theoretical values for* $E[\bar{V}]$ *and* $var(\bar{V})$.

3 Observe Central Limit Theorem

According to the Central Limit Theorem, if $V_1,\ V_2,\ \ldots\ , V_n$ are i.i.d. then $\frac{\bar{V}-E[\bar{V}]}{\sqrt{var(\bar{V})}} \to N(0,1)$ as $n \to \infty$

Use `Compute` in the `Transform` menu. `Reset` the compute screen. Use `vbar` for the target variable and `Mean` for the function. Type so that you will see:
`Mean(v1,v2,...,v20)` where you need to fill in the `...` with the rest of the variables. Likewise, compute the standard deviation, `s`, but, to save typing, replace `Mean` with `SD`. Make sure that you do not overwrite the column for `vbar`. Find the average of all 600 random numbers and call it, `muhat`. Use the `Descriptive Statistics` menu to get the average of all the `vbar`'s; this average will be `muhat`. You will use `s` and `muhat` in step 3.

Question 4 : *What is the value of* `muhat`?

1. In Question 2, we computed μ and σ^2, then we found $E[\bar{V}]$ and $var(\bar{V})$. We will now compute the values for the variable `zscore` by using `Compute`. Enter `zscore` for the target variable and, for the function, type so that you will see: `(vbar-?)/sqrt(??)`. Replace `?` with the numerical value of $E[\bar{V}]$ and `??` with the value of $var(\bar{V})$ which were found for Question 3.

2. Make a histogram of `zscore` using the `Histogram` plot utility. The plot should look like a normal distribution with mean 0 and variance 1.

3. Now assume that μ and σ^2 are, in fact, unknown. We will compute the sample z–score. Enter `samplez` for the target variable. Compute values for the variable `samplez` by using `(vbar-muhat)/(s/sqrt(20))`. Remember to enter manually the values of `muhat` and of `s` which you calculated above. Make a histogram of `samplez`.

   Notice that the histogram resembles the shape of a Normal distribution, but `samplez` is neither a Normal nor a Student's t distribution because the V_i's are not distributed as a $N(\mu, \sigma^2)$, but instead the V_i's are distributed as a certain discrete Uniform distribution based on your GWID. Nonetheless, the resemblance of the histogram with a Normal distribution is due to the Central Limit Theorem. Regardless of the distribution of V_i, as $n \to \infty$, the distribution of `samplez` will converges to N(0,1).

Question 5 : *Assume that* $\bar{V} \sim N(muhat, \sigma^2)$ *where* $\sigma^2 = var(\bar{V})$ *which was found in Question 3, find two numbers, a and b, symmetric about* `muhat` *such that* $P(a \leq \bar{V} \leq b) = 1 - \alpha$. *Use* $\alpha = .05$

4 90% Confidence Intervals

Because of a peculiarity with SPSS, we cannot simply calculate the means and variances of the twenty columns of random numbers under the variables v1, ..., v20. Instead we have to take a circuitous route. The functions of SPSS compute across columns but not down columns; therefore, it is necessary to transpose rows and columns as in:

$$\begin{bmatrix} a & d \\ b & e \\ c & f \end{bmatrix} \rightarrow \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$$

1. Save your worksheet in case you make a big mistake later.

2. Select the `Transpose` in the `Data` menu. Highlight and select the twenty variables: v1, ..., v20. The other variables which are not selected will be lost, hence the advice to save your worksheet.

3. After the transpose has taken place, your worksheet will consist of 30 columns and 20 rows. As you have done already, use the function `Mean` in the `Compute` menu to calculate the sample means of the 30 variables, var001, ... var030. Use `xbar` as the target variable.

4. Calculate the sample standard deviation by erasing the word `Mean` and typing in its place the function `SD`. Use `s` for the target variable.

5. Reset the `Compute` screen. Use the letter, `a`, for the target variable and in the function box type: `xbar-s/sqrt(30)*Idf.t(.95,29)` where the function `Idf.t` will find the quantile for Student's t distribution with 29 degrees of freedom at an $\frac{\alpha}{2} = .05$. Similarly, calculate the upper limit, `b`.

5 Graph Confidence Intervals

Question 6 : *What is the value of the population mean*?

You should be able to count by inspection the number of confidence intervals which contain the population mean. Nothing beats a picture, however. In the menu `Graph`, select the `High Low` chart, followed by `Simple high-low-close` and `Values of individual cases`. For `High` use `b` and for `Low` use `a` which you calculated in Section 4 item 5.

6 Report

Include in your report the following:

1. Answers to the questions.

2. The graph of the column of random numbers made in Section 2 item 8.

3. The graphs of the histograms using $\bar{V}$ which were made in Section 3 items 2 and 3.

4. The transposed array of random numbers with `xbar`, `s`, and the lower and upper limits of the confidence intervals.

5. The graph of the confidence intervals made in Section 5, and draw on it by hand the line which represents the true population mean and answer how many 90% confidence intervals contain the population mean?