



- You may use Excel for all kinds of calculation.
  - For example, if you are required to attain the mean of  $\{2, 3, 4\}$ , you will not get the score by just providing the final result 3. You might have acquired the result by applying a function to the cells that contains 2, 3 and 4 in Excel. The file you used to get values should also be submitted for all the problems that require calculations.
  - You should submit two files – the pdf file that contains the answers and the Excel file that you used for calculation. Please make sure that all the files are ordered and readable.
  - The full score for each of the questions is 5 points. The total score is 250.
1. Determine whether the following statements are true or false and also explain why do you think so. Simply writing true or false will not give the score. All the statements have the word “always” and hence one easy way to explain why it is false is to provide an example that the statement does not apply.
    - (a) For any given two arrays  $A$  and  $B$ ,  $\text{COVARIANCE.P}(A, B) \leq \text{COVARIANCE.S}(A, B)$  always holds.
    - (b) For any given array  $A$ ,  $\text{STDEV.P}(A) \leq \text{STDEV.S}(A)$  always holds.
    - (c) Standard deviations are always non-negative.
    - (d) For any given data, the standard deviation is always smaller than the variance.
    - (e) If  $\mathbb{E}[X] = \mathbb{E}[Y]$  and  $\text{Var}[X] = \text{Var}[Y]$  hold for random variables  $X$  and  $Y$ , the probability distribution of  $X$  and  $Y$  are always the same.
    - (f) Suppose you have attained a sample and Findd the sample mean. Then, it is always the case that the Findd sample mean is different from the unknown population mean.
    - (g) It is always the case that the calculations of the coefficient of correlation are the same whether the given data is a population or a sample.
    - (h) If the variance of random variables  $X$  and  $Y$  are the same, the standard deviations of them are always the same.
    - (i) The probability of a given event  $A$  is always less than or equal to 100%.
    - (j) The distribution of a sample mean can always be approximated by some normal distribution.
    - (k) Suppose you are calculating a 95% confidence interval for the population mean. If you already know the population standard deviation and use it in constructing the confidence interval, the width of the interval is always the same regardless of the value of the sample mean.
    - (l) Suppose you already know the population standard deviation and are curious of the population mean. You obtained a sample and constructed two 95% confidence intervals: one with the population standard deviation and the other with the sample standard deviation. Then, the second interval is always wider than the first one.

2. There are 54 million Americans whose native tongue is not English. The following table shows the detailed information on Americans whose native tongue is not English. Suppose we will randomly select one individual from the pool of 54 million Americans.

Native tongue	Millions of Americans
Spanish	32.1
Chinese	2.8
Tagalog	1.6
Vietnamese	1.3
French	1.1
Korean	0.9
Other	14.2

- (a) Find the probability that the native tongue of the selected individual is Spanish.
- (b) Find the probability that the native tongue of the selected individual is a language other than Spanish.
- (c) Find the probability that the native tongue of the selected individual is Vietnamese or Korean.
3. Consider a random variable  $X$  that has the following distribution function.

Realization ( $x$ )	Probability ( $\mathbb{P}(x)$ )
-3	0.02
-2	0.05
-1	0.11
0	0.15
1	0.42
2	0.13
3	0.07
4	0.05

- (a) Find  $\mathbb{P}\{X \leq 0\}$
- (b) Find  $\mathbb{P}\{X < -1\}$
- (c) Find the mean of  $X$ .
- (d) Find the variance of  $X$ .
- (e) Find the standard deviation of  $X$ .
4. Suppose a random variable  $X$  follows a normal distribution with  $\mu = 410$  and  $\sigma = 30$ . Answer the following questions.
- (a) Find  $\mathbb{P}\{X \leq 400\}$ .
- (b) Find  $\mathbb{P}\{X < 400\}$ .
- (c) Find  $\mathbb{P}\{X \geq 400\}$ .
- (d) Find  $\mathbb{P}\{300 \leq X \leq 400\}$
- (e) Find the value of  $a$  that satisfies  $\mathbb{P}\{X \leq a\} = 0.2$
- (f) Find the value of  $a$  that satisfies  $\mathbb{P}\{X > a\} = 0.2$
5. Suppose a random variable  $X$  follows a normal distribution with  $\mu = 300$  and  $\sigma = 20$ . This time, answer the following questions only by using `NORM.S.DIST()` and `NORM.S.INV()` functions. Do not use `NORM.DIST()` or `NORM.INV()` functions.
- (a) Find  $\mathbb{P}\{X \leq 310\}$ .
- (b) Find the value of  $a$  that satisfies  $\mathbb{P}\{X \leq a\} = 0.3$ .
6. Suppose a random variable  $X$  has the mean of 50 and the standard deviation of 5.  $\bar{X}$  is a sample mean from a sample of the size 16.
- (a) Find  $\mathbb{E}[\bar{X}]$ .
- (b) Find the standard deviation of  $\bar{X}$ .
7. Suppose a random variable  $X$  has the mean of 100 and the standard deviation of 9.  $\bar{X}$  is a sample mean from a sample of the size 36.
- (a) Find  $\mathbb{P}\{\bar{X} \leq 98\}$ .
- (b) Find  $\mathbb{P}\{97 \leq \bar{X} \leq 103\}$ .

8. Suppose a random variable  $X$  is normally distributed with the mean of 100.  $\bar{X}$  is a sample mean and  $s$  is a sample standard deviation from a sample of size 9.

- (a) Find  $\mathbb{P}\{\frac{\bar{X}-\mu}{s/\sqrt{9}} \leq 1\}$
- (b) Find the value of  $a$  that satisfies  $\mathbb{P}\{\frac{\bar{X}-\mu}{s/\sqrt{9}} \leq a\} = 0.8$

9. You are interested in the population mean of the daily demand. You had obtained a sample of the demands as follows. Suppose that the daily demand is following a normal distribution.

122	107	110	105
88	134	124	136
85	93	94	86

- (a) Find the sample mean.
- (b) Find the 99% confidence interval of the mean with the assumption that the population standard deviation is 20.
- (c) Find the 95% confidence interval of the mean with the assumption that the population standard deviation is 18.
- (d) Find the sample standard deviation.
- (e) Find the 99% confidence interval of the mean using the sample standard deviation.
- (f) Find the 95% confidence interval of the mean using the sample standard deviation.

10. Use “GSS2014.xlsx” file for this problem. You can visit

<http://gss.norc.oregonstate.edu/Get-Documentation>

and download “Entire GSS Cross-Section Codebook” to get information on the abbreviation of the data. From page 1, you will see the explanations on each of the abbreviations and the description of each of the codes is given from page 130 of the codebook.

- (a) Is General Social Survey data a sample or a population? Why do you think so?
- (b) Determine the category (interval, ordinal, or nominal) of the following data. You do not need to provide explanations.

RACE, CHILDS, CLASS, AGEKDBRN, INCOME, MAEDUC, PARTYID, TAX

- (c) Will it be valuable to find the mean of “WRKSTAT?” Why do you think so?
- (d) Find the mean of any data that interests you. For example, if you are interested in INCOME, you can get the mean of them. Since I have already mentioned it, use data other than INCOME. Note that the mean should be meaningful.
- (e) Find the variance of the data you used in the previous question.
- (f) Find the standard deviation of the data you used in the previous question.
- (g) Now choose two data which you think have a linear relationship and find the covariance of them.
- (h) Find the coefficient of correlation of two data you chose in the previous question.
- (i) For the data you found the mean, obtain 95% confidence interval of the mean. We will assume that you already know the population standard deviation. Since we do not actually know it, simply pick a value for the population standard deviation on your own and use it to make the confidence interval.
- (j) Find 95% confidence interval of the mean using the standard deviation you calculated from the data. This time assume that you do not know the population standard deviation and also assume that the population distribution is close to a normal distribution.