Investigating the Randomness of Numbers
Part III: Expectations of the Distribution

If a population from which a sample is selected is normally distributed, most of the sample values would be expected to fall in what location? (in the middle of the distribution)

If a population from which a sample is selected is positively skewed (skewed to the right), most of the sample values would be expected to fall in what location? (toward the lower values)

TRICK QUESTION: If a population from which a sample is selected is uniformly distributed, most of the sample values would be expected to fall in what location?

This is a trick question because there is no location in which most of the values are expected to fall. They are expected to be evenly distributed throughout the population.

Students should recognize that the population from which these samples were taken is uniformly distributed. Therefore, they should ***expect*** to see sample values that are generally evenly distributed throughout the population. If the population is subdivided into groups, whether the counts of sample values within those groups are as ***expected*** can be assessed using the chi-square test for goodness of fit.

NOTES TO THE TEACHER: Discussions of the number of groups and the merits of each should occur. Likely candidates are 5 and 10 groups of uniform size, because the expected counts for such groups (10 and 5, respectively) are easily calculated. Additionally, students should recognize that the number of groups should not exceed 10. If such were the case, the expected counts for all groups (if uniform in size) would be less than 5. The chi-square test is not robust when such is the case. Using 10 groups rather than 5 may lead to greater deviations from expectations, but the larger number of degrees of freedom may likely compensate for the increased value of the test statistic. In the end, students may wish to calculate the test statistic using both 5 and 10 groups to see whether the number of groups affects the decision.

To make it easier for the students to count the observed values in each of the aforementioned groups, the lists can be sorted using the following process.

1. If students would like to preserve the original list, first copy the list to be sorted into List 6 as follows.
    - Cursor to the name of List 6.
    - Select **OPTN**.
    - Select **LIST** (**F1**).
    - Select **List** (**F1**).
    - Enter the number of the list to be copied.
    - **EXE**cute.

2. To order from the **STAT** menu
    - Access **other options** not shown (**F6**).
    - Sort in ascending order: **SRT-A** (**F1**).
    - A prompt asks for input of the number of lists to be sorted.
    - If more than one list is to be sorted, a prompt asks for the base (initial) list to be sorted and then the others.

NOTES TO TEACHER: You may wish to have students work in pairs when the frequencies are counted. The students should compare results to assure that agreement exists. When 5 groups of uniform size are used, the observed counts from list 1 and the actual values of the numbers in those groups are shown below.

1. (1 – 200) (**10** numbers): 45, 93, 101, 102, 107, 116, 137, 145, 169, 172
2. (201 – 400) (**10**): 225, 231, 261, 262, 282, 296, 303, 333, 3346, 398
3. (401 – 600) (**6**): 405, 450, 471, 558, 562, 600
4. (601 – 800) (**13**): 625, 645, 665, 666, 670, 671, 678, 690, 697, 728, 741, 746, 757
5. (801 – 1000) (**11**): 808, 825, 832, 866, 870, 884, 915, 929, 954, 965, 982

With 5 groups of uniform size, the expected count in each group is 10 (50 numbers divided by 5 groups). The calculation of the chi-square test statistic with 5 groups is simple enough to be done by hand:
$(10 – 10)^2/10 + (10 – 10)^2/10 + (6 – 10)^2/10 + (13 – 10)^2/10 + (11 – 10)^2/10 =$
$0 + 0 + 1.6 + 0.9 + 0.1 = 2.6$

If 10 groups are used, students may wish to use lists to calculate the test statistic using the following process.

1. If students wish to preserve the original lists, it will be necessary to change list files ("file drawers").
    - Be sure that the counts for each group have been recorded.
    - Select **MENU**.
    - Select **LIST**.
    - Select **SET UP** (**shift-MENU**).
    - The List File currently used is shown. Select any unused List File by selecting the appropriate F-key.
    - **EXE**cute. The selected List File is shown. If already used, select another.
2. Enter the counts in an empty list.
3. Cursor to the name in another list.
4. Enter a left-parenthesis.
5. Select **OPTN**.
6. Select **LIST** (**F1**).
7. Select **List** (**F1**).
8. Enter the number of the list in which the counts are found.
9. Enter the minus sign.
10. Enter the expected count (which is constant at 5 in the case of 10 equal-sized groups).
11. Square the result ($\boldsymbol{x}^2$).
12. Enter the division sign.
13. Enter the expected count, which is again 5 in this case.

14. Sum the values in this list as follows.
    - Select **MENU**.
    - Select **RUN**.
    - Select **OPTN**.
    - Select **LIST** (**F1**).
    - Select **other options** not shown (**F6**).
    - Again select **options** not shown (**F6**).
    - Select **Sum** (**F1**).
    - Select **other options** not shown (**F6**).
    - Select **List** (**F1**).
    - Enter the number of the list whose values are to be summed.
    - **EXE**cute.

After having calculated the chi-square test statistic either by hand or using the calculator, the students will use the results of the chi-square test to form a decision regarding the null hypothesis that the distribution of values within the groups is as expected. The chi-square test is accessed as follows.

1. Select **MENU**.
2. Select **STAT**.
3. Select **DIST**ribution (**F5**).
4. Select **CHI**-square (**F3**).
5. Select the chi-square cumulative distribution: **Ccd** (**F2**).
6. The lower value is your chi-square test statistic. Store using **EXE**.
7. The upper value is a suitably large number; 100 will generally suffice. Store.
8. The degrees of freedom is the number of groups minus 1; store.
9. **EXE**cute.
10. The probability of the given test statistic or a more extreme value is give.

NOTE TO TEACHER: The p-value for the chi-square test for List 1 should be 0.62682.

Neither the z-test of the mean nor the chi-square test of the distribution of frequencies for List 1 showed cause to reject the null hypotheses that the mean was equal to the population mean or the distribution of frequencies was as expected. Therefore, there is no evidence to suggest that the sample is not random ***based on these tests***.

Students should perform the chi-square tests for the other four lists in the same manner.

NOTE TO THE TEACHER: Using the same 5 groups that were used for List 1, the chi-square test statistics and the results of the chi-square tests are given below.

1. List 2: test statistic = 45; p-value is extremely small.
2. List 3: test statistic = 160; p-value nearly zero (note the need to adjust upper).
3. List 4: test statistic is 0; p-value is 1.
4. List 5: test statistic is 0; p-value is 1.

The *z*-test did not give reason to question the randomness of either List 2 or List 3, but the chi-square test did. The *z*-test gave reason to question List 4, but the chi-square test did not (although the p-value of 1 should set off the same type of alarm that the p-values of 1 for the *z*-tests for Lists 2 and 3 set off). Neither the *z*-test nor the chi-square test gave cause to question List 5, but that peculiar p-value occurred once again. Perhaps List 5 should be scrutinized with another test: the eyeball test.

Look again at the numbers in List 5. Nearly every number ends in 5. Although no single list of random numbers is more likely to occur than any other, the likelihood of such a large proportion of last digits ending in 5 is highly unlikely to occur. The likelihood could be measured using another chi-square test measuring the observed distribution of frequencies of last digits, but that should not be necessary. Common sense should prevail, and the notion that List 5 is likely to be random should be rejected.