Chapter 2

# Applying the Concepts: Excel Instructions

## Estimate the Probability of an Event

1. Open the die\_roll\_simulation\_data.csv dataset in Excel. (We cannot set seeds for randomization in Excel, so ignore the instruction to set the seed to 777; instead, we will use this pre-seeded data.)
2. Highlight the data in Column A.
3. Click on Insert Statistic Chart→Histogram.
4. Right-click the chart’s X-axis, select Format Axis
5. Select the ‘Number of bins’ option button and type 20.
6. In cell B2, type “Number of times a six came up 29 times:”
7. In cell C2, type:

=COUNTIF(A:A,29)

1. In cell B3, type “Probability of times a six came up 29 times:”
2. In cell C3, type:

=COUNTIF(A:A,29)/100

## Make a Statistical Inference

No tools are used in this Activity.

## 

## Activity - Evaluate a Hypothesis

1. Open the school.csv dataset in Excel.
2. Verify that the first row of Column D is labeled “AvgCourseDifficulty.”
3. In Cell G2, type:

=COUNTIF(D:D,"<2")

1. In Cell G3, type:

=COUNT(D:D)

1. In Cell G4, type:

=G2/G3

1. Cell G4 gives the proportion of students.

Question 3 cannot be addressed easily in Excel.

## Perform T-tests on Observational Data

1. Open the volleyball.csv dataset in Excel.
2. Verify that the first row of Column B is labeled “hit\_perc”.
3. Calculate Sample Mean and Standard Deviation:
   1. In Cell I2, type:

=AVERAGE(B:B)

as the sample mean.

* 1. In Cell H2, type: “Sample mean:”
  2. In Cell I3, type:

=STDEV.S(B:B)

as the sample standard deviation.

* 1. In Cell H3, type “Standard deviation:”

1. Determine Sample Size:
   1. In Cell I4, type:

=COUNT(B:B)

using the number of observations in the sample.

* 1. In Cell H4, type “Sample Size:”

1. Set Hypothesized Mean:
   1. In Cell I5, type 30 represents 30% average hitting percentage.
   2. In Cell H5, type “Hypothesized Mean”
2. Calculate t-statistic:
   1. In Cell I6, type:

=(I2 - I5)/(I3/SQRT(I4))

based on the formula:



* 1. In Cell H6, type “t-value:”

1. Find p-value:
   1. In Cell I7, type:

=T.DIST(I6,I4-1,FALSE)

* 1. In Cell H7, type “p-value:”

## Draw Conclusions from a Correlation

1. Open the icecream1.csv dataset in Excel.
2. Click “Data” tab
3. Click “Data Analysis” on the far right
4. Click “Correlation”
5. Click “OK”
6. Input Range → B2:C101
7. Click “OK”
8. See the correlation.
9. Copy the value from the new sheet in Cell B4 to the ‘icecream1’ sheet
   1. In Cell G2, paste the value from the other sheet (this is the correlation)
   2. In Cell F2, type “Correlation:”

Calculating the SBI p-value for a correlation is not currently available in Excel.

1. Open the icecream2.csv dataset in Excel.
2. Click “Data” tab
3. Click “Data Analysis” on the far right
4. Click “Correlation”
5. Click “OK”
6. Input Range → B2:D101
7. Click “OK”
8. Find the correlations.
9. Copy the value from the new sheet in Cell B4 to the ‘icecream2’ sheet
   1. In Cell G2, paste the value from the other sheet (this is the correlation)
   2. In Cell F2, type “Correlation:”

Calculating the SBI p-value for a correlation is not currently available in Excel.

## Draw Conclusions from Observational Studies

1. Open the school.csv dataset in Excel.
2. Verify that the first row of Column E is labeled “DailyFruitVeg”
3. Calculate Sample Mean and Standard Deviation:
   1. In Cell I2, type:

=AVERAGE(E:E)

as the sample mean.

* 1. In Cell H2, type “Sample mean:”
  2. In Cell I3, type:

=STDEV.S(E:E)

as the sample standard deviation.

* 1. In Cell H3, type “Standard deviation:”

1. Determine Sample Size:
   1. In Cell I4, type:

=COUNT(E:E)

using the number of observations in the sample.

* 1. In Cell H4, type “Sample Size:”

1. Set Hypothesized Mean:
   1. In Cell I5, type 2.21.
   2. In Cell H5, type “Hypothesized Mean”
2. Calculate t-statistic:
   1. In Cell I6, type:

=(I2 - I5)/(I3/SQRT(I4))

based on the formula:



* 1. In Cell H6, type “t-value:”

1. Find p-value:
   1. In Cell I7, type:

=T.DIST(I6,I4-1,TRUE)

* 1. In Cell H7, type “p-value:”

## Identify Type I and Type II Errors

No tools are used in this Activity.

## 

## Identify an A/B Test’s Null and Alternative Hypotheses

No tools are used in this Activity.

## Perform T-tests on A/B Testing Data

1. Open the portfolio.csv dataset in Excel.
2. Verify that the first row of Column A is labeled “treatment”.
3. Calculate the A treatment participant counts:
   1. In Cell F2, type:

=COUNTIF(A2:A101, “A”)

* 1. In Cell E2, type “A treatment participant counts:”

1. Calculate the B treatment participant counts:
   1. In Cell F3, type:

=COUNTIF(A2:A101, “B”)

* 1. In Cell E3, type “B treatment participant counts:”

1. Calculate the A treatment’s view time mean:
   1. In Cell F4, type:

=AVERAGEIF(A2:A101,"A",B2:B101)

* 1. In Cell E4, type “A treatment’s view time mean:”

1. Calculate the B treatment’s view time mean:
   1. In Cell F5, type:

=AVERAGEIF(A2:A101,"B",B2:B101)

In Cell E5, type “B treatment’s view time mean:”

1. Calculate the p-value of the T-Test for view time
   1. In Cell F8, type:

=T.TEST(B2:B51, B52:B101, 2, 2)

* 1. In Cell E7, type “P-Value for view time:”

## Identify Type I and Type II Errors in A/B Testing Results

No tools are used in this Activity.