Leigh Tesfatsion
Last Revised: 10 April 2013

Experimental Design:
Basic Concepts and Terminology

An **EXPERIMENTAL FRAMEWORK** consists of a structural environment within which a user can systematically vary one or more key structural features (*treatment factors*).

Example: Schelling Demo (Chris Cook's demonstration software for the Schelling Segregation Model)

https://www2.econ.iastate.edu/tesfatsi/demos/schelling/schellhp.htm

Key features of the Schelling Demo that are currently **maintained** include: agents reside in an 8×8 checkerboard world.

Key structural features of the Schelling Demo that can be **systematically varied** by the user as treatment factors include:

- the number of agent classes (1, 2, or 3);
- the relative size of the different agent classes;
- the form of the "happiness rule" for each agent;
- the initial location of the agents.

Given an experimental framework, an **HYPOTHESIS** consists of
a conjecture of the following form:

> *If the treatment factor(s) specified by the user take form A, then the
> experimental outcomes the user observes will take form B.*

An **EXPERIMENTAL DESIGN** consists of a careful description of how a particular hypothesis can be experimentally tested.

This requires: (a) an explicit specification of the treatment factors to be tested; (b) the specific range of values over which these treatment factors will be tested; (c) the manner in which observations will be generated, recorded, and reported; and (d) the criteria that will be used to evaluate the degree to which the observations appear to support the hypothesis or to be inconsistent with the hypothesis.

Example 1: Outcome Prediction for a Given Structural Specification

Illustrative Hypothesis for the Schelling Demo: If the number of agent classes is 2, and if these classes each contain 30 agents, and if each agent is unhappy (tries to move) unless he resides in a neighborhood in which at least 37.5% of his neighbors are in his class, then (regardless of the initial locations of the agents) the city will evolve into a highly segregated pattern.

Example 2: Sensitivity of Schelling Demo Outcomes to Changes in Model Structure

Suppose (a) the number of agent classes is 2; (b) each class contains 30 agents; and (c) all agents are randomly distributed at the beginning of the experiment. Consider three happiness rules H that systematically vary with regard to their tolerance level, as follows: rule H1 displays a high tolerance of outsiders; rule H2 displays only moderate tolerance of outsiders; and rule H3 displays a high intolerance of outsiders.

Illustrative Hypothesis for the Schelling Demo:

If the happiness rule for each agent is systematically varied from H1, to H2, and finally to H3, the degree of segregation observed to evolve in the Schelling Demo will correspondingly increase.

Experimental Design (Hypothesis Testing):

- For each configuration of the model's treatment factors that you wish to test, conduct multiple *runs* of the model, where each run is associated with a different user-specified setting of a *random seed* for the pseudo-random number generator (PNG).

  ### Important Note:

  - Given any specific setting for the random seed of a given PNG, this PNG will generate a "random number" each time the PNG is invoked within a run.
  - The entire sequence of "random numbers" generated from a specific setting for the random seed for a given PNG is completely determined by the random seed and the PNG, and so can be exactly replicated by later users.
  - Consequently, for each experimental run, the setting for the random seed should be captured and recorded along with all user-specified treatment-factor settings and all user-specified or default settings for maintained parameter values and simulation control options.
  - Saving this information for a run permits others to replicate the run exactly at a later time for verification or other purposes.

- For each run, record the degree of segregation displayed by the resulting agent location pattern.

- Report "descriptive statistics" that summarize these experimental segregation findings.

- As explained more carefully below, these descriptive statistics would typically include (at a minimum) the sample mean value, the sample standard deviation, and possibly also a histogram, for the degree of segregation observed across runs.

- Based on these descriptive statistics, draw conclusions regarding whether or not your hypothesis appears to be supported by your observations.

SAMPLE MEAN AND STANDARD DEVIATION

Suppose you have conducted N runs of an experiment using N different seed values for your pseudo-random number generator.

Suppose you have recorded N observations X_1, X_2,..., X_N regarding some quantifiable experimental outcome of interest X (e.g., the degree of segregation).

Definition:

The **SAMPLE MEAN (OR AVERAGE)** value of N observations $X_1, X_2, \ldots, X_N$ is defined to be

$$Mean_X \;\; = \;\; \frac{\Sigma_{i=1}^{N} X_i}{N} \;\; = \;\; \frac{[X_1 + X_2 + \ldots + X_N]}{N} \;\; . \qquad (1)$$

Definition:

The "N" definition for the **SAMPLE STANDARD DEVIA-TION** of N observations $X_1, X_2, \ldots, X_N$ is as follows:

$$SSD_X(N) \;=\; \left(\frac{\Sigma_{i=1}^{N}[X_i - MeanX]^2}{N} \right)^{1/2} \;.\qquad (2)$$

Thus, $SSD_X(N)$ measures the average degree of dispersion of the N observations about their sample mean.

The smaller the sample standard deviation, the greater the confidence one has that the sample mean provides a reliable prediction for the type of experimental outcome one can expect to observe.

For later possible use, note the following "N" representation for the **SAMPLE VARIANCE**:

$$SVAR_X(N) \;=\; [SSD_X(N)]^2 \;=\; \frac{\Sigma_{i=1}^{N}[X_i]^2}{N} - [MeanX]^2 \;.\; (3)$$

Technical Note:

Some researchers prefer to use the following "N-1" definition of the sample standard deviation due to technical considerations:

$$SSD_X(N-1) \;=\; \left(\frac{\Sigma_{i=1}^{N}[X_i - MeanX]^2}{N-1}\right)^{1/2}. \qquad (4)$$

For later possible use, note the following "N-1" representation for the **SAMPLE VARIANCE**:

$$SVAR_X(N-1) = [SSD_X(N-1)]^2 = \frac{\Sigma_{i=1}^{N}[X_i]^2}{N-1} - \frac{N[MeanX]^2}{N-1}$$

Given certain additional regularity conditions, $SVAR_X(N-1)$ provides an unbiased estimator for the true variance whereas $SVAR_X(N)$ results in bias. On the other hand, $SVAR_X(N)$ is often easier to motivate and to use in calculations.

For further discussion of this issue, see:

https://www2.amstat.org/publications/jse/secure/v8n3/weldon.cfm

HISTOGRAM

The sample mean and sample standard deviation are informative when observations exhibit a "central tendency," that is, when observations cluster around their sample mean value with a relatively small sample standard deviation. However, these descriptive statistics can be misleading if observations do not exhibit a central tendency.

For this reason it is good experimental practice to construct and view the "histogram" for any sample of observations before deciding exactly what descriptive statistics to report about these observations.

Suppose you have conducted N runs of an experiment using N different seed values for your pseudo-random number generator.

Suppose, also, that you have recorded N observations $X_1, X_2,..., X_N$ for some quantifiable experimental outcome of interest X.

For example, X could denote the degree of segregation ranging from $X = 0$ (no segregation) to $X = 1$ (highest possible segregation).

Definition:

The **HISTOGRAM** H for N observations with regard to some quantifiable experimental outcome of interest, X, is a plot of the possible X values against the frequency with which these X values were actually observed among the N observations.

HISTOGRAM ... CONTINUED

More precisely, H is constructed as follows.

1. First, construct a two-dimensional graph G with the horizontal axis consisting of all possible X values and the vertical axis consisting of the range of possible frequencies (from 0 to 100%) with which these X values could be observed in the given sample of N observations.

2. Second, for each possible X value, let $\#X$ denote the number of times within the given sample of N observations that this X value was actually observed.

3. Third, divide $\#X$ by N (the total sample size) to get the corresponding frequency $F(X)$ with which each X value was actually observed in the sample.

4. Fourth, in your graph G, draw a plot of the points $(X, F(X))$ for each possible X value. This plot constitutes the desired histogram H.

EXAMPLE: Suppose the number of observations is $N = 200$. Suppose within this sample that the value $X = 0$ was observed 50 times (implying $F(0) = 25\%$), the value $X = 0.25$ was observed 100 times (implying $F(0.25) = 50\%$), and the value $X = 0.75$ was observed 50 times (implying $F(0.75) = 25\%$). Then the histogram H for this sample of observations is the plot of the points $(0, 25\%)$, $(0.25, 50\%)$, and $(0.75, 25\%)$ in a graph of X against $F(X)$.