Obtain Descriptive Statistics Using Excel

1. Open Excel

2. Load the Analysis ToolPak by clicking **File** → **Option** →**Add-ins** → **Analysis Toolpak** → **Go**

3. Enter data by hand, or open an existing data file

4. Click **Data** → **Data Analysis** →**Descriptive Statistics** where

   (a) *Mean* is $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ which measures the central location

   (b) *Sample variance* is $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2)$ which measures the variability (or spread)

   (c) *Standard deviation* is the positive square root of sample variance $s = \sqrt{s^2}$, and it also measures variability

   (d) *Standard Error* is $\frac{s}{\sqrt{n}}$ which is the standard deviation of sample mean $\bar{x}$

   (e) *Sum* is $\sum_{i=1}^{n} x_i$

   (f) *Count* is n

5. **Data Analysis** also allows you to run simple regression.

Obtain Joint Distribution (Contingency Table) Using Stata

Suppose you are analyzing the survey data for a pizza restaurant. There are two random variables, gender and feedback. Gender takes on one for male and zero for female. Feedback takes on one if the feedback is good, and takes on zero if the feedback is bad. The owner of restaurant wants to know whether gender matters for the feedback. Statistically, this requires checking whether feedback and gender are independent, i.e., whether $P(Y = y, X = x) = P(Y = y)P(X = x),\ \forall y, x$

1. Open Stata
2. Click **File** → **Open**, and choose the file 311_survey.dta
3. Generate joint and marginal distributions by typing in command window

   ```
   tabulate gender feedback
   ```

   ```
   tabulate gender feedback, cell
   ```

   Probability (cell percentage) is reported with option cell
4. Obtain conditional distribution $P(feedback|gender)$, or row percentage, by using command

   ```
   tabulate gender feedback, row
   ```
5. Feedback is independent of gender if the column for feedback=1 is proportional to the column for feedback=0. Can you prove it?
6. Command

   ```
   tabulate gender feedback, chi2
   ```

   reports the Pearson's chi-squared test. The null hypothesis is that feedback and gender are independent. The null hypothesis is rejected if the p-value is less than 0.05.
7. Alternatively you can run a regression to check whether gender matters for feedback. The command is

   ```
   reg feedback gender
   ```