

AAE374 Discussion Section #3
Representing Distributions with Numbers:
Measures of Central Tendency and Statistical Dispersion

In our first meeting, we showed that histograms are a graphical display of tabulated frequencies in which the height of each class/category represents the proportion of cases that fall into each of the different classes.

They are useful because they reveal important information regarding which classes have higher frequencies (i.e. which values have a higher probability of occurrence), how spread out the data is, possible outliers, and modes, among others features.

We can capture these features by a few important statistics called *summary statistics*. They convey as many features as possible about the data in a simple way, and thus provide summaries about the data.

Summary statistics usually are divided in three categories:

- i. Measures of central tendency (location): mode, mean, and median
- ii. Measures of statistical dispersion: range, standard deviation, and variance
- iii. Measures for the shape of the distribution: skewness, kurtosis

Today we will concentrate on the first two classes of summary statistics: measures of central tendency and statistical dispersion. Please download statistics.xls.

1) Measures of Central Tendency

- a) Mode: The most frequent value in a data set or, in histograms, the class that has the highest bar (the point where the distribution has its peak).
- b) Median: The median is the middle observation or 50th percentile value
- c) Arithmetic Mean: the sum of the observations divided by the number of observations.
 - i) The mean \bar{X} is defined as

$$(a) \bar{X} = \frac{1}{n}(X_1 + X_2 + \dots + X_n) = \frac{1}{n} \sum_i X_i.$$

2) Measures of Statistical Dispersion (Spread of a Distribution)

- a) Range: length of the smallest interval which contains all the data
- b) Variance: Measure of dispersion which averages the squared distance of every data point from the mean of the data. Conveys the scale or degree in which the data is spread out.
 - i) The variance σ^2 is then defined as

$$\sigma^2 = \frac{1}{n-1} \sum_i (X_i - \bar{X})^2$$

- c) Standard Deviation: the root-mean-square deviation of the values from their mean, or

simply the square root of the variance.

i) The standard deviation is then defined as

$$\sigma = \sqrt{\frac{1}{n-1} \sum_i (X_i - \bar{X})^2} = \sqrt{\sigma^2}$$

d) Absolute deviations ($|X_i - \bar{X}|$) can also be measured and averaged if necessary.

3) Exercise “convergence or divergence?” - GDP.xls

a) Data

- i) GDP per capita (ppp, constant 2000 international \$)
- ii) GDP per capita growth rate (%)
- iii) Food export (% of merchandise exports)

b) Instructions

- i) Divide the countries into two groups (high and low) depending on the GDP per capita (\$10K).
- ii) Compare the mean, s.d., min, max of the growth rates and share of food exports.
- iii) Make histograms

4) Problem Set 1...

a) was a thrill to grade. Overall, people did a good job but we will grade future assignments harder.

- i) Analysis is important! These are difficult questions and a few words are not enough to address the key issues. Although, concisely hitting the main points is also appreciated.
- ii) Show your work.
- iii) A well organized problem set is appreciated – consider using the word doc.

b) Major problems:

- i) 1.c: comparative advantage driven by factor endowments and factor intensity
- ii) 1.d: marginal productivity - e.g., $(25)^{0.2}(41)^{0.8} - (25)^{0.2}(40)^{0.8} = 0.73$.
- iii) 2.c.3: Diminishing returns
- iv) 2.e.2: Compare at equilibrium levels
- v) 2.f.3/4: Factor Price Equalization - returns to labor and capital.