

AAE374 Discussion Section #4
Representing Distributions with Numbers:
Measures of Central Tendency and Statistical Dispersion

We last 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 a higher frequencies (i.e. which values have a higher probability of occurrence), how spread out is the data or its range, possible outliers, and modes among others features.

We can capture these features by a few important statistics called *summary statistics*. They convey as much features from the data as possible 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.

Download the excel -file statistics.xls

1. Measures of Central Tendency:

- i. Mode: The most frequent value in a data set or in terms of the histogram the class that has the highest bar (the point where the distribution has its peak).
- ii. Median: The median is the middle observation or 50th percentile value
- iii. Arithmetic Mean: the sum of the observations divided by the number of observations

The mean \bar{X} is defined as

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

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

- i. Range: length of the smallest interval which contains all the data
- ii. 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.

The variance σ^2 is then defined as

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

- iii. Standard Deviation: the root-mean-square deviation of the values from their mean, or as the square root of the variance.

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

- iv. The absolute deviation is instead the typical deviation, lying somewhere between the smallest and largest deviation $|X_i - \bar{X}|$.

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

Data

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

Instructions

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