Introduction & Review

author: ECONOMIC FORECASTING date: Summer 2017 autosize: true

Measures of Central Tendency

Moments allow us to describe a probability distribution. Expected Value is a measure of the distribution's central tendency.

$$\mu = E(X) = \begin{cases} \sum_x X \cdot p(X) \text{ if the variable is discrete} \\ \int_x X \cdot f(x)\, dx \text{ if the variable is continuous} \end{cases}$$

Sample measure (estimator) of E(X) is:

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

which is commonly known as the sample mean. * Minimum Variance Unbiased (MVU)

Measures of Dispersion

The second moment is a measure of the dispersion of the distribution about its mean.

$$\sigma^2 = V(X) = \begin{cases} \mathbf{E}[(X-\mu)^2] \text{ if the variable is discrete} \\ \int (x-\mu)^2\ f(x)\, dx \text{ if the variable is continuous} \end{cases}$$

Sample measure of V(X):

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

NOTE: the standard error, s, transforms this measure into the appropriate units.

Correlation

Many times, we would like to measure how two random variables are related.

$$\mathbf{C}(X,Y) = \mathbf{E}[(X-\mu_X)(Y-\mu_Y)]$$

Scaling this to fall between -1 and 1 yields:

$$\rho(X,Y) = \frac{\mathbf{C}(X,Y)}{\sqrt{\sigma_X^2 \sigma_Y^2}}$$

which is the **correlation coefficient**.

Normal Distribution

N(μ, σ^2) The shape is determined by two factors: - Expected Value (μ), and - Variance (σ^2)

Properties - Centered on μ - Symmetrical about μ - Area above μ and below μ are each equal to 1/2

Central Limit Theorem

Suppose X_i is a sequence of independent and identically distributed (i.i.d) random variables $E(X_i) = \mu$ and $V(X_i) = \sigma^2 < \infty$, Then, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ This is critical since we rarely know the distribution from whuch the data is generated.

t-Distribution

Properties - Symmetrical about 0 and bell-shaped - Standard deviation is determined by the number of degrees of freedom - As degrees of freedom $\to \infty$, standard deviation $\to 1$ - As degrees of freedom $\to \infty$, the curve approaches the probability curve of the standard normal distribution

Hypothesis Testing

Allows us to know if our estimations are statistically significant We can use standard deviations from the mean to determine statistical significance.

$$\begin{aligned}
\mu \pm \sigma &= 68.2\% \text{ of the area} \\
\mu \pm 2\sigma &= 95.4\% \text{ of the area} \\
\mu \pm 3\sigma &= 99.8\% \text{ of the area}
\end{aligned}$$

Steps for Hypothesis Testing

1. Formulate Hypothesis
2. Collect data and compute test statistic
3. Determine sampling distribution of test statistic
4. Compute the probability that a value of the sample statistic at least as large as the one observed could have been drawn from the sampling distribution
5. If the probability is high, do NOT reject the null hypothesis. If the probability is low, then reject the null hypothesis.

Formulating a Hypothesis, Part I

$$\begin{aligned}
\mathbf{H_0}&: \text{null hypothesis} \\
\mathbf{H_1}&: \text{alternative hypothesis}
\end{aligned}$$

Example: Suppose someone claims that the average height of a man in the U.S. is 5'9". What are the null and alternative hyptheses?

Formulating a Hypothesis, Part II

H_0: null hypothesis
H_1: alternative hypothesis

Example: Suppose someone claims that the average height of a man in the U.S. is 5'9".

What are the null and alternative hyptheses?

H_0: The average height of men in the U.S. is 5'9"
H_1: The average height of men in the U.S. is NOT 5'9"

Mathematically:

$$\mathbf{H_0} : Height_{men} = 5'9"$$
$$\mathbf{H_1} : Height_{men} \neq 5'9"$$

Testing the Hypothesis, Part I

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

n= 225 $\bar{X}$ = 69.1333, μ_0= 69 $s = 1 \implies t = 2$ and $Pr > |t|$ = 'a little less than 10 percent' So do we reject the null hypothesis???

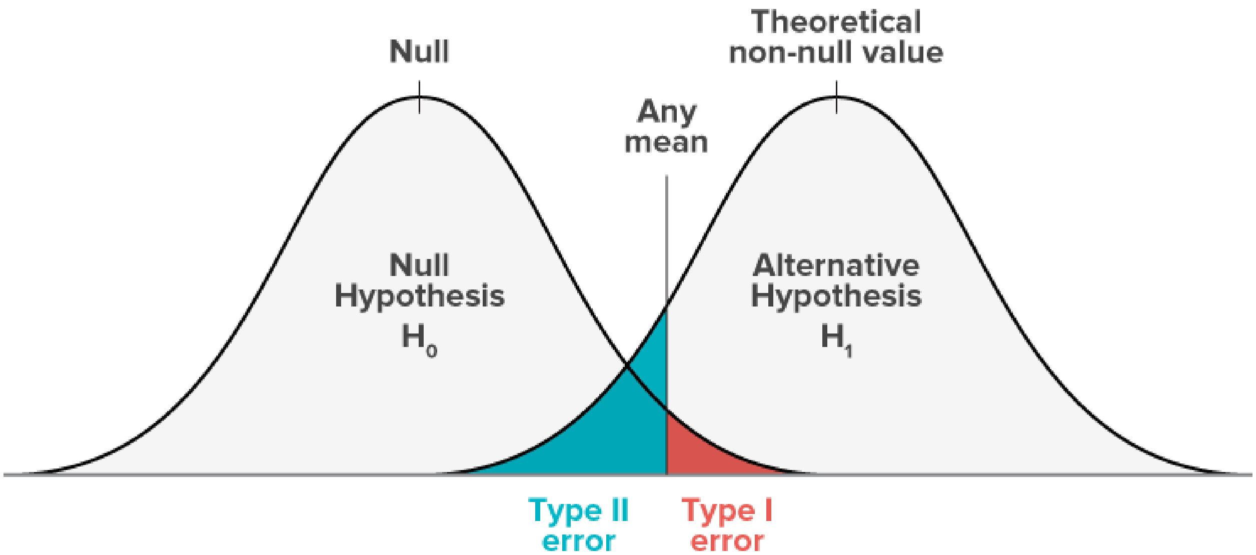
Testing the Hypothesis, Part II

$$t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$$

n= 225 $\bar{X}$ = 69.1333, μ_0= 69 $s = 1 \implies t = 2$ and $Pr > |t|$ = 'a little less than 10 percent' So do we reject the null hypothesis??? It depends...

Error Types

	Do NOT reject H_0	Reject H_0
H_0 is true	Correct Decision	Type I Error
H_0 is false	Type II Error	Correct Decision



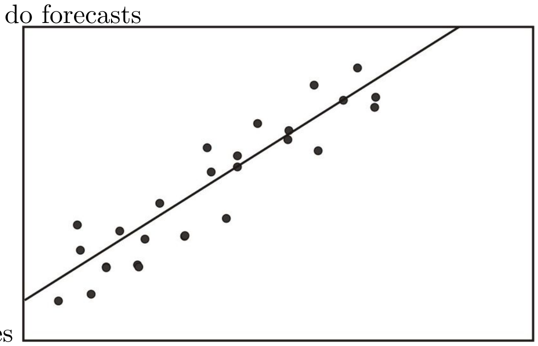
Type I Error: "Convicting the Innocent" Type II Error: "Freeing the Guilty"

Assumptions of the CLR

1. Non-stochastic Xs 2. X and Y have a linear relationship 3. E(ϵ) = 0 4. $N > k$ 5. Spherical errors

Fitting a Straight Line

- Allows us to do forecasts



- Least Squares

Least Squares

Used to calculate the equation of a straight line that minimizes the sum of squared errors.

$$SSE = \sum \left(Y - \widehat{Y}\right)^2$$

$\widehat{Y}$ can be written as $b_0 + b_1\mathbf{X}$. Thus,

$$SSE = \sum (Y - b_0 - b_1\mathbf{X})^2$$

and we can test whether $b_1 = 0$ via a t-test

What is a Time Series?

A random variable measured sequentially in time is known as a **time series**. Examples: - Monthly rainfall in Akron, Ohio from January 1997 to May 2012 - GDP of the U.S. from 1945 to 2017 - Quarterly unemployment rate in the U.S from Q1 2004 through Q3 2009

Notation:

$$\{x_1, x_2, x_3, ...\} \text{ or } \{x_t\}$$

Typically a time series is **serially correlated**.

Horizontal

Data fluctuates around a certain level, E(X), and is stationary about its mean.

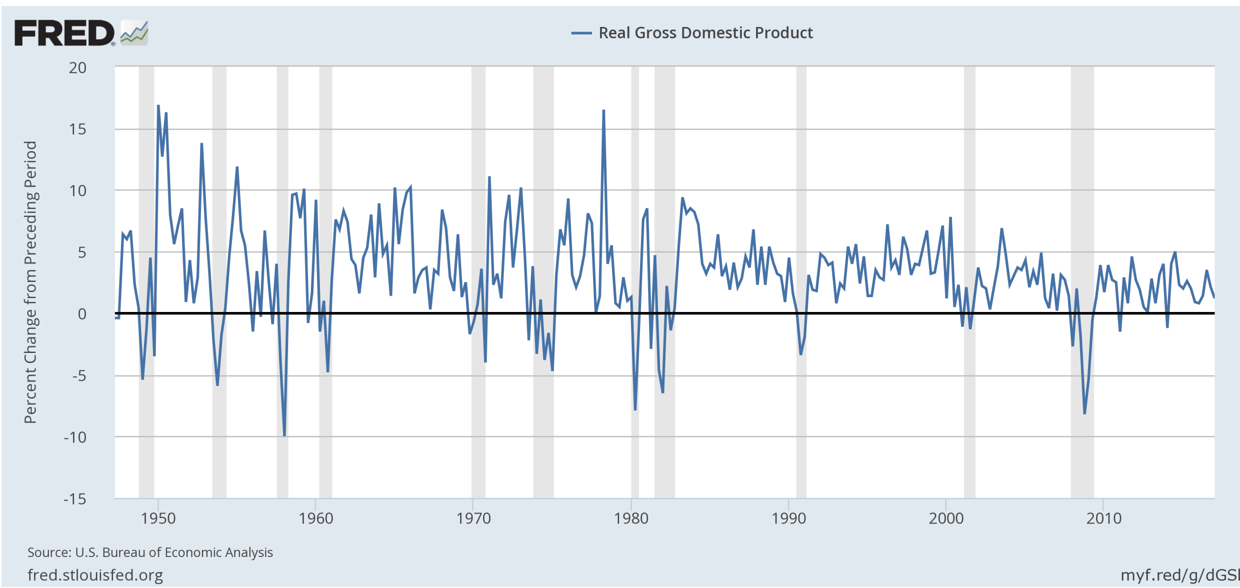


Figure 1: alt text

Trend

Growth or decline of a variable over several periods.

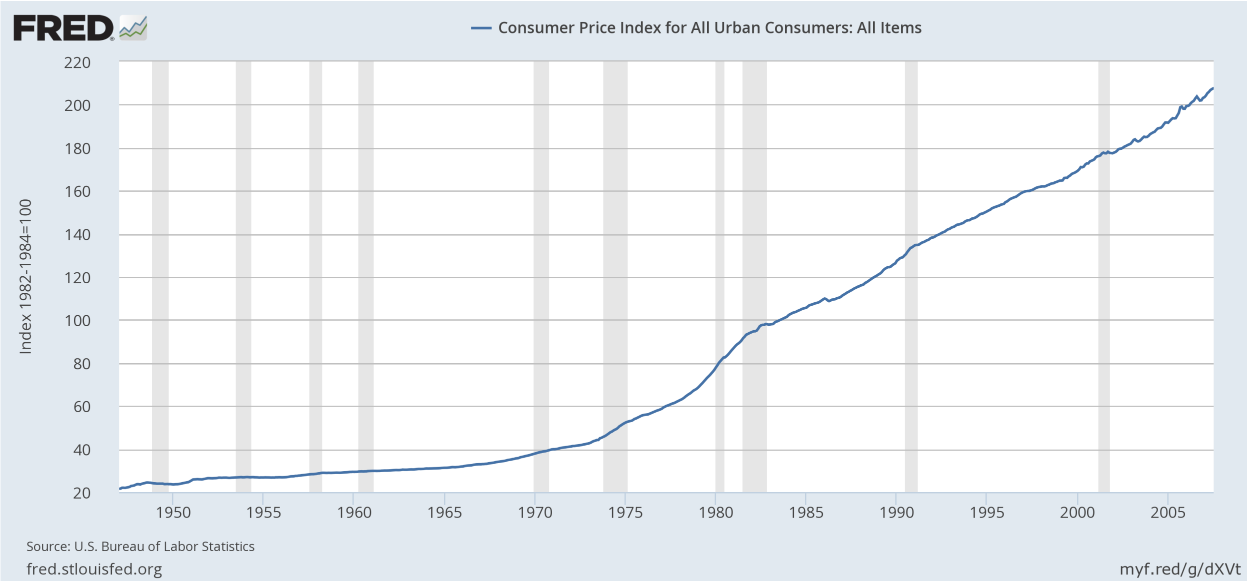


Figure 2: alt text

Cyclical

Wavelike pattern around a trend. “Caused” by general economic conditions.

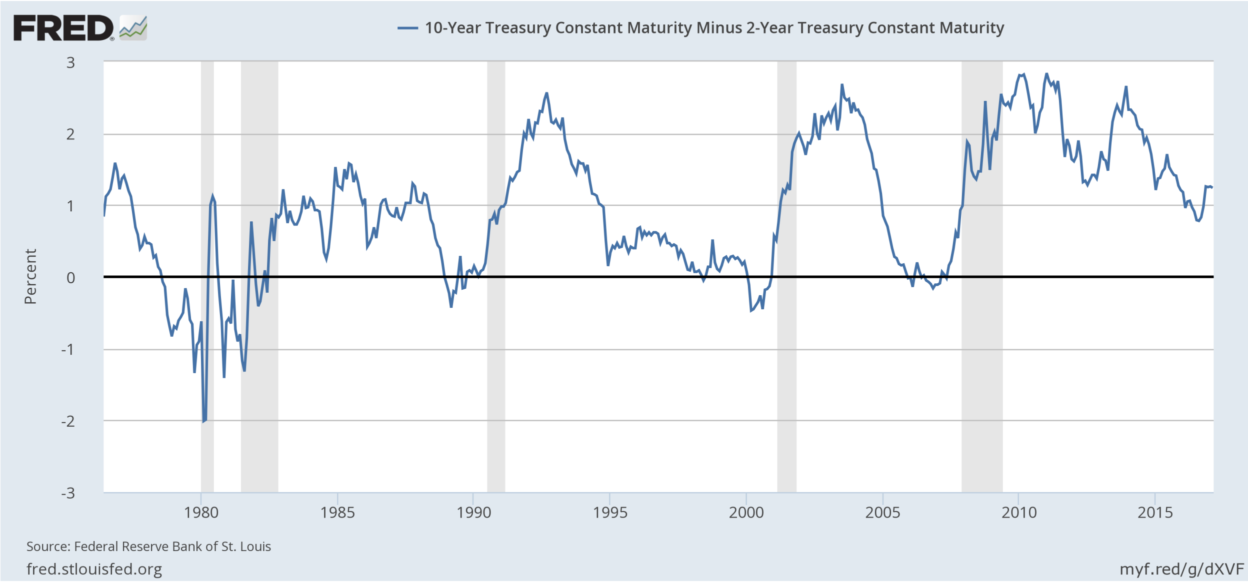


Figure 3: alt text

Seasonal

Pattern changes that appear year after year.

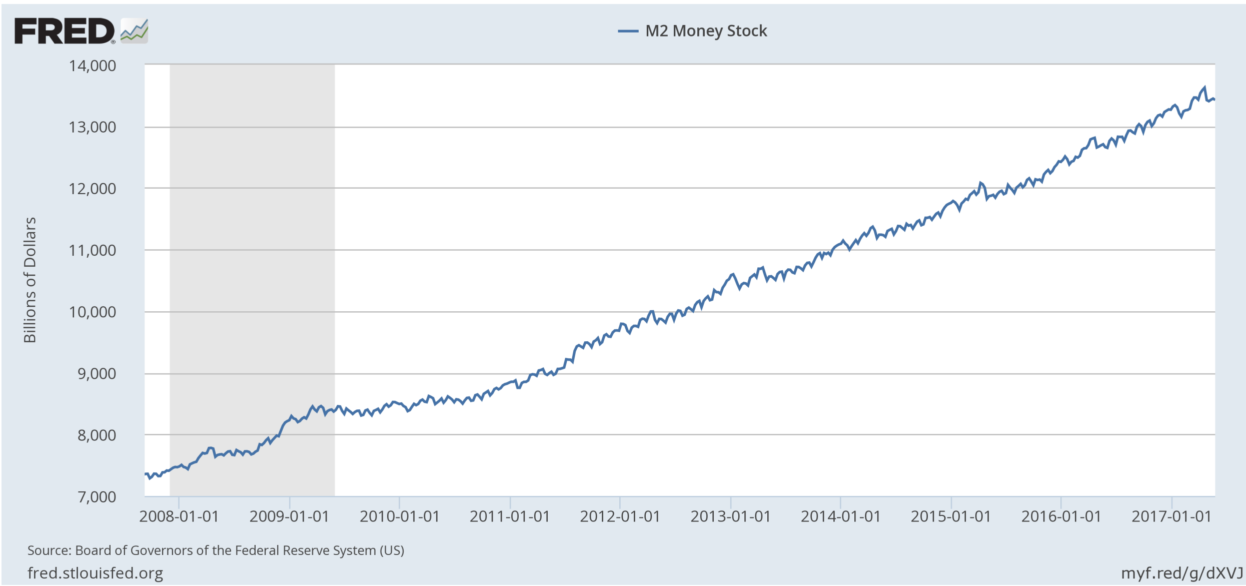


Figure 4: alt text

Intro to R

R is a language - Easily implemented to perform statistical tasks - Extendable - FREE!

We will be using RStudio! - Integrated Devlopment Environment (IDE) for R - Also FREE!

Resources for R

There are resources available on Springboard. - R for Beginners - R for SAS and SPSS Users

. . . and resources available on the internet. - r-project.org - stackoverflow.com - r-bloggers.com