ECO 6416: Simple Regression

Joshua L Eubanks

University of Central Florida

Linear Equations

A linear equation in "slope-intercept" form has the following form:

$$y = mx + b$$

where y is the dependent variable, x is the independent variable, m is the slope, b is the intercept.

A linear equation allows us to perform the following tasks:

- summarize the relationship among variables in an equation
- predict variable from data on related variables
- sensitivity of variable to a related variable

Advantages of Regression vs Descriptive Statistics

Descriptive Statistics:

- Univariate analysis describes the center, shape, and spread
- Bivariate analysis describes direction, strength, and form

Regression Analysis:

- Both simple and multivariate regression measure and explain the variation of the dependent variable (y)

Key Terms in Bivariate Analysis

- **Scatterplot:** a two dimensional plot of bivariate-cross section data
- **Time series plot:** a two dimensional plot with a quantitative time series variable on the vertical axis and time on the horizontal axis
- **Trend Line:** a straight line showing the linear fit between a time series variable and time itself
- **Trend rate:** slope of trend line
- Use samples to draw conclusions about the population

Regression Equations

$$\hat{y} = \beta_0 + \beta_1 * x_1 + \varepsilon$$

- $\hat{y}$ is the dependent variable, also referred to as *response* or *predicted* variable
- x is the independent variable, also referred to as *explanatory* or *predictor* variable
- β_0 is the intercept
- β_1 is the slope
- ε is the random disturbance, which is unknown

Analysis of Regression Equations

Prediction and forecasting (Thing 1: Plug and Chug): plug in new observations into regression equation and calculate the predicted value of the dependent variable. The margin of error for 95% CI is 2*SEE.

Marginal Analysis (Thing 2): Multiply the change in one independent variable by its slope to find the predicted change in the dependent varaible. The margin of error will be covered in Unit 2

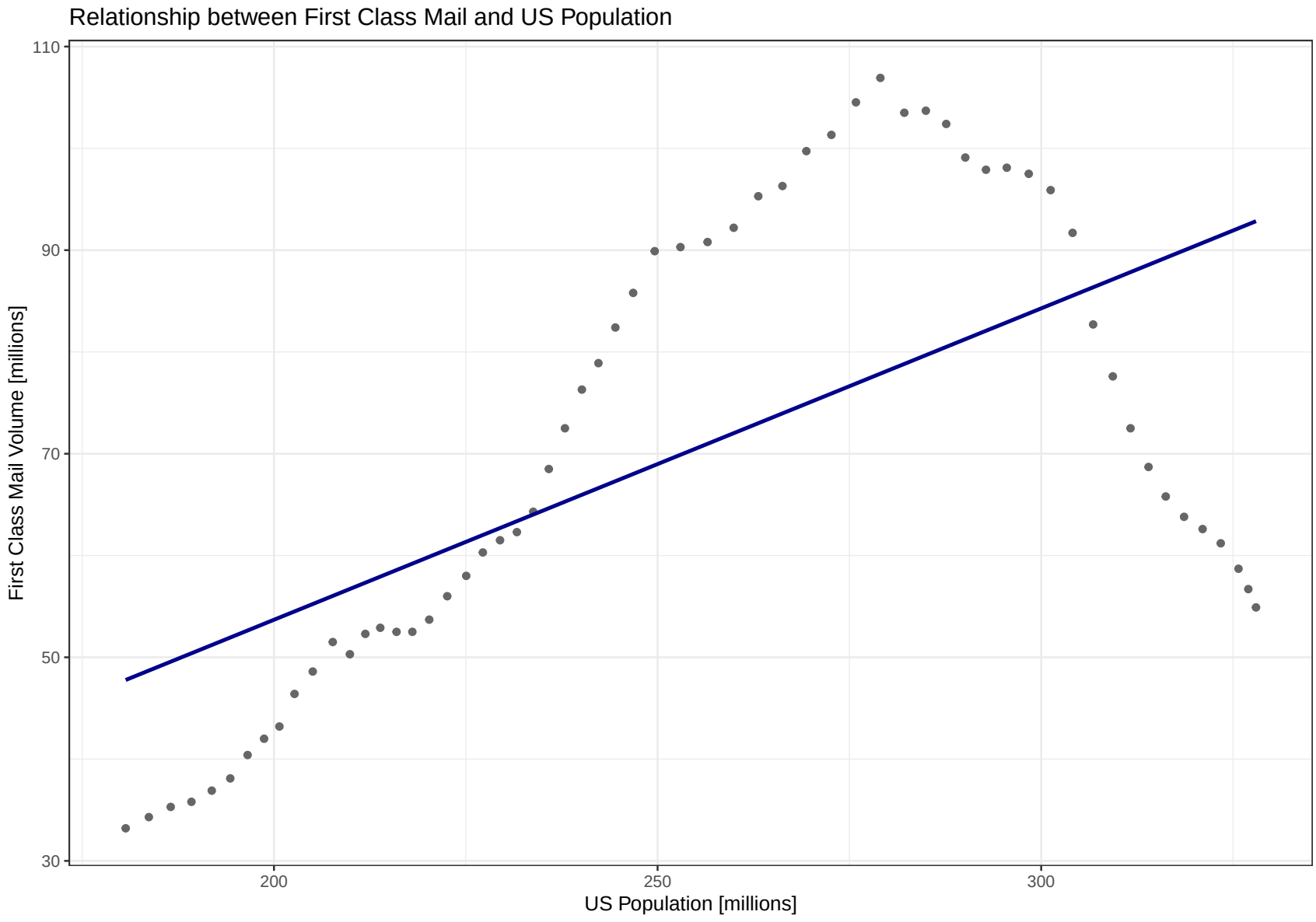
Trend Analysis

When time is the independent variable the equation becomes:

$$\hat{y} = \beta_0 + \beta_1 * \mathit{Time} + \varepsilon$$

- Time trends: upward, downward, or no trend
- Trend rates only apply to time trends
- Trend rate is the coefficient of the time variable, or the slope regression equation (β_1)

Example: First Class Mail Volume



Cross Section Regression

```
##
## Call:
## lm(formula = `1stClVol` ~ PopUSA, data = mail)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -37.946 -12.636  -3.501  16.314  29.054
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.50080   13.40499  -0.560    0.578
## PopUSA       0.30593    0.05202   5.881 2.14e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## Residual standard error: 17.97 on 58 degrees of freedom
## Multiple R-squared:  0.3736, Adjusted R-squared:  0.3628
## F-statistic: 34.59 on 1 and 58 DF,  p-value: 2.136e-07
```

Thing 1

Based off this regression, the equation would look like:

$$1stClVol = -7.5 + 0.31 * PopUSA$$

Thing one would look like: If the population is expected to be 300 million, what is the expected amount of first class mail?

$$Expected1stClVol = -7.5 + 0.31 * 300 = 85.5$$

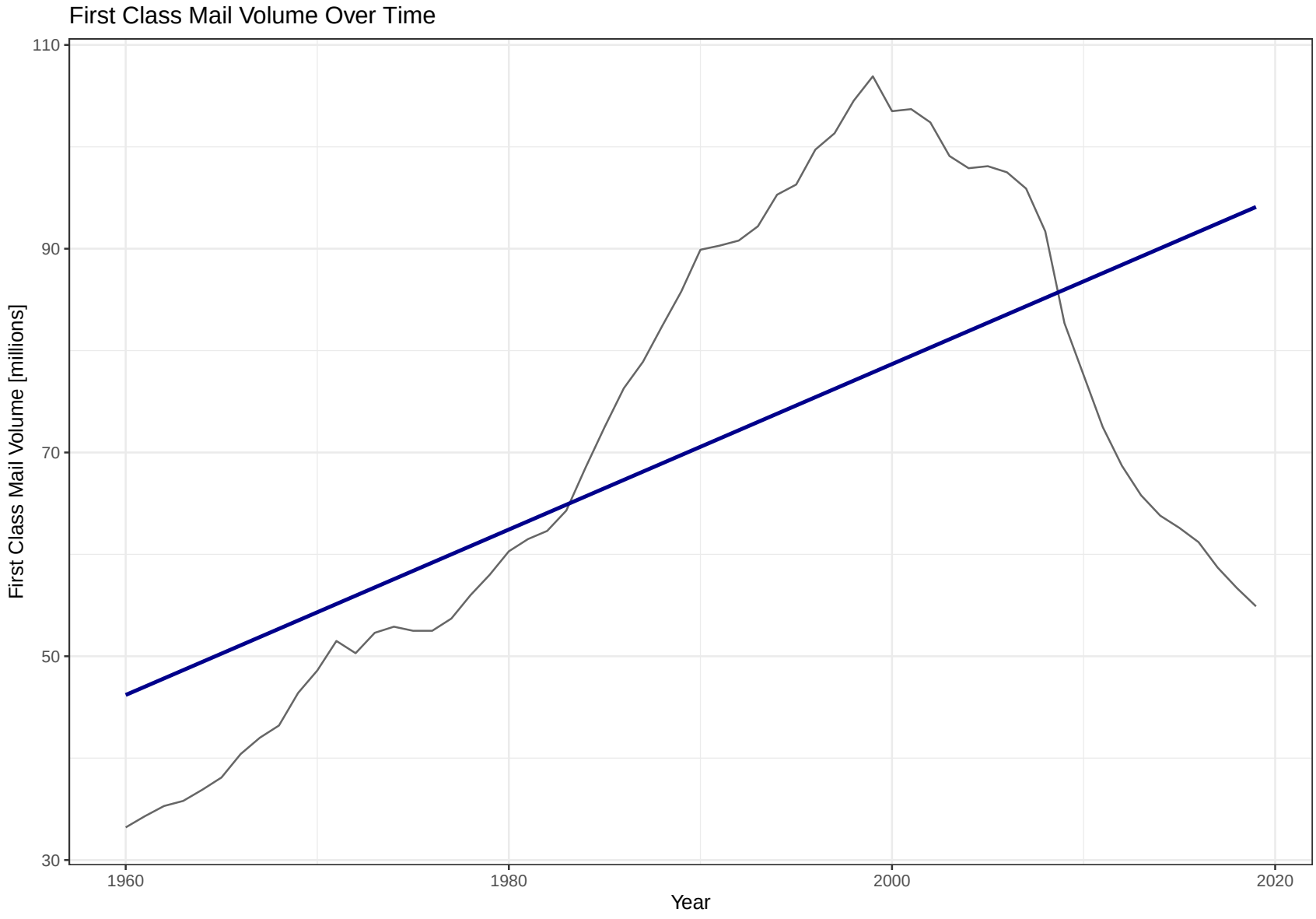
with a margin of error of $2 * 17.97 = 35.94$

Thing 2

If the expected population is going to increase by 10 million in the next year, then

$$\Delta 1stClVol = 0.31 * 10 = 3.1$$

Time Series


First Class Mail Volume Over Time
First Class Mail Volume [millions]
110
90
70
50
30
1960
1980
2000
2020
Year

Time Series Regression

```
##
## Call:
## lm(formula = `1stClVol` ~ Time, data = mail)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -39.200 -11.044  -3.453  15.998  29.058
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  45.3886     4.6117   9.842 5.58e-14 ***
## Time          0.8119     0.1315   6.174 7.02e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1
##
## Residual standard error: 17.64 on 58 degrees of freedom
## Multiple R-squared:  0.3966, Adjusted R-squared:  0.3862
## F-statistic: 38.12 on 1 and 58 DF,  p-value: 7.017e-08
```

Thing 1

Based off this regression, the equation would look like:

$$1stClVol = 45.4 + 0.81 * YearsSince1960$$

Thing one would look like: two years from now, what is the expected amount of first class mail?

$$Expected1stClVol = 45.4 + 0.81 * (60 + 2) = 95.62$$

with a margin of error of $2 * 17.64 = 35.28$

Thing 2

What is the quarterly rate of growth in first class mail volume?

$$\Delta 1stClVol = 0.81 * .25 = .20$$