- Unit 1: Introduction to Probability
- Unit 2: Probability Distributions
- Unit 3: Statistical Inference
- **Unit 4: Introduction to Linear Regression**
- Unit 5: Regression Analysis
- Unit 6: Regression Modeling

# Announcements

- Past:

  Unit 3 Team Assignment solutions available and scores posted.

- Present:

  Unit 4 Individual Assignment **due Monday, 23:59 ET** (Durham local time).

- Future:

  Unit 5 materials and team assignment available.

- **Measuring Dependence** OF PAIRS OF UNCERTAINTIES

  Covariance: quantifying linear dependence, $\mathrm{Cov}\,[X, Y] = \mathbb{E}\,[(X - \mathbb{E}\,[X])(Y - \mathbb{E}\,[Y])]$

  Correlation: measuring the strength of linear dependence, $\mathrm{Corr}\,(X, Y) = \mathrm{Cov}\,[X, Y] \,/\, \sqrt{\mathrm{Var}\,[X]\,\mathrm{Var}\,[Y]}$

  $-1 \leq \mathrm{Corr}\,(X, Y) \leq 1$

  If $\mathrm{Corr}\,(X, Y) \neq 0$, then $X$ and $Y$ are dependent

# Unit 4: Dependence and Introduction to Regression

- **Measuring Dependence**

  Covariance: quantifying linear dependence, $\mathrm{Cov}\left[X, Y\right] = \mathbb{E}\left[(X - \mathbb{E}\left[X\right])(Y - \mathbb{E}\left[Y\right])\right]$

  Correlation: measuring the strength of linear dependence, $\mathrm{Corr}\left(X, Y\right) = \mathrm{Cov}\left[X, Y\right] / \sqrt{\mathrm{Var}\left[X\right] \mathrm{Var}\left[Y\right]}$

  $-1 \leq \mathrm{Corr}\left(X, Y\right) \leq 1$

  If $\mathrm{Corr}\left(X, Y\right) \neq 0$, then $X$ and $Y$ are dependent

- **Linear regression: population model** $\mathbf{Y} = \beta_\mathbf{0} + \beta_\mathbf{1}\mathbf{X} + \epsilon$, **where** $\epsilon \sim \mathbf{N}(\mathbf{0}, \sigma_\epsilon^\mathbf{2})$
  - conceptual description of how $Y$ linearly depends on $X$
  - regression line: $Y = \beta_0 + \beta_1 X$; error term: $\epsilon \sim N(0, \sigma_\epsilon^2)$
  - $Y$ linearly depends on $X$ only if $\beta_1 \neq 0$
  - accuracy of regression is captured by $\sigma_\epsilon$ (standard deviation of the error term)

- **Measuring Dependence**

  Covariance: quantifying linear dependence, $\mathrm{Cov}\left[X, Y\right] = \mathbb{E}\left[(X - \mathbb{E}\left[X\right])(Y - \mathbb{E}\left[Y\right])\right]$

  Correlation: measuring the strength of linear dependence, $\mathrm{Corr}\left(X, Y\right) = \mathrm{Cov}\left[X, Y\right] / \sqrt{\mathrm{Var}\left[X\right]\mathrm{Var}\left[Y\right]}$

  $-1 \leq \mathrm{Corr}\left(X, Y\right) \leq 1$

  If $\mathrm{Corr}\left(X, Y\right) \neq 0$, then $X$ and $Y$ are dependent

- **Linear regression: population model** $\mathbf{Y} = \beta_\mathbf{0} + \beta_\mathbf{1}\mathbf{X} + \epsilon$, where $\epsilon \sim \mathbf{N}(\mathbf{0}, \sigma_\epsilon^2)$
  - conceptual description of how $Y$ linearly depends on $X$
  - regression line: $Y = \beta_0 + \beta_1 X$; error term: $\epsilon \sim N(0, \sigma_\epsilon^2)$
  - $Y$ linearly depends on $X$ only if $\beta_1 \neq 0$
  - accuracy of regression is captured by $\sigma_\epsilon$ (standard deviation of the error term)

- **Linear regression: sample model** $\widehat{\mathbf{Y}} = \widehat{\beta}_\mathbf{0} + \widehat{\beta}_\mathbf{1}\mathbf{X} + \epsilon$, where $\epsilon \sim \mathbf{N}(\mathbf{0}, \mathrm{SE}_{\mathrm{reg}}^2)$
  - use sample data to estimate the intercept and slope coefficients of the regression line
  - use sample data to estimate the standard deviation of the error term $\sigma_\epsilon$
  - $\mathrm{SE}_{\mathrm{reg}}$ is the estimate of the overall accuracy of the regression, i.e. $\mathrm{SE}_{\mathrm{reg}}$ estimates $\sigma_\epsilon$
  - each coefficient estimate has its own SE

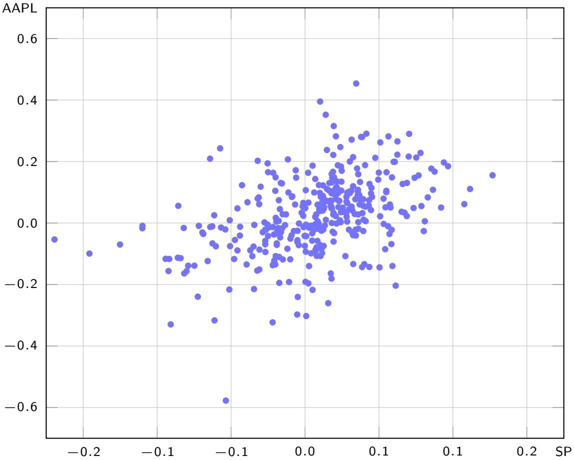
# Apple vs. S&P500 returns

Data set `StocksMonthlyReturns_Data` contains monthly returns for APPL and for the SP500 index.

- Are the Apple monthly returns related to the S&P500 monthly returns?
- What is the estimate for the regression line equation? What is the standard error of regression?
- What is the average change in APPL return if SP return increases by 1%? (marginal rate of change)
- What is the beta for Apple against the S&P500 index?
- Confidence that the beta for Apple is larger than one?
- Confidence that APPL return is positive in a month in which SP return is 10%?

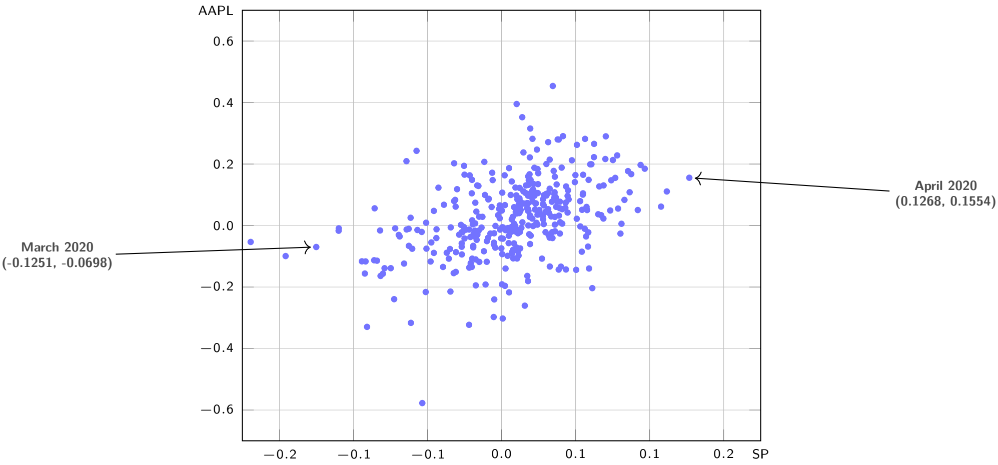
# Apple vs. S&P500 returns

Are the Apple monthly returns related to the S&P500 monthly returns?



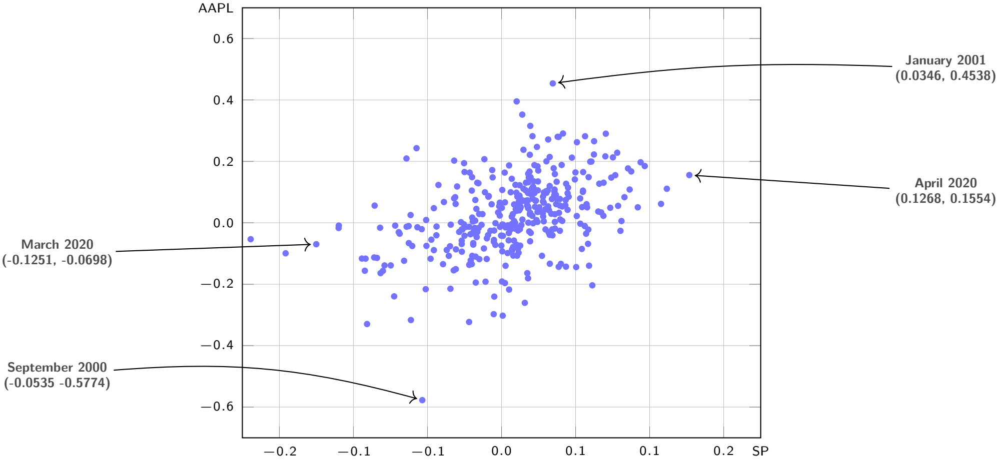
# Apple vs. S&P500 returns

Are the Apple monthly returns related to the S&P500 monthly returns?



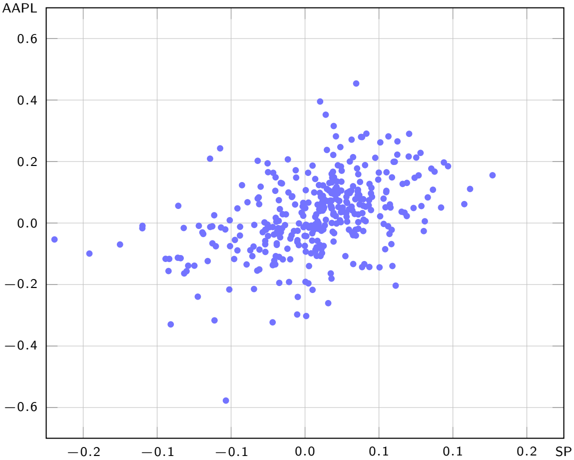
# Apple vs. S&P500 returns

Are the Apple monthly returns related to the S&P500 monthly returns?



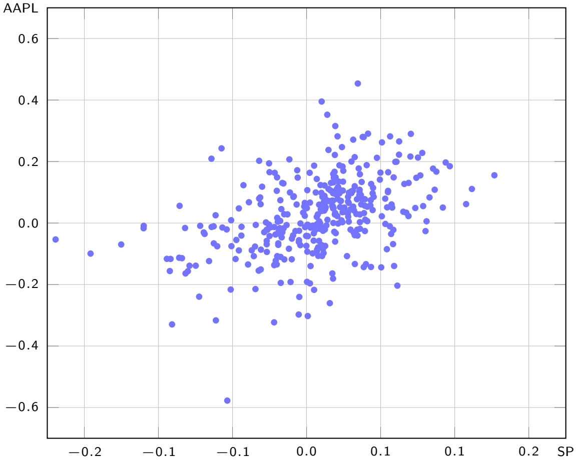
# Apple vs. S&P500 returns

Are the Apple monthly returns related to the S&P500 monthly returns?



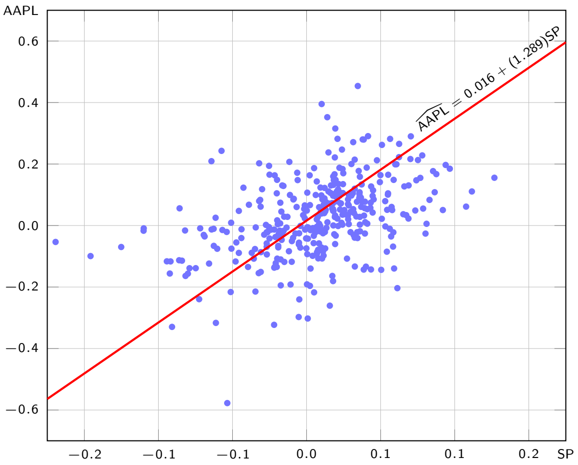
# Apple vs. S&P500 returns

Are the Apple monthly returns related to the S&P500 monthly returns? $\mathrm{Corr}\,(\text{AAPL}, \text{SP}) = 0.435$



# Apple vs. S&P500 returns

Are the Apple monthly returns related to the S&P500 monthly returns? $\text{Corr}\,(\text{AAPL}, \text{SP}) = 0.435$



# Apple vs. S&P500 returns

What is the estimate for the regression line equation? What is the standard error of regression?

- Population model: $\text{AAPL} = \beta_0 + \beta_1 \text{SP} + \epsilon$, with $\epsilon \sim N(0, \sigma_\epsilon^2)$
  - **Dependent variable:** AAPL (Monthly return on Apple stocks)
  - **Independent variable:** SP (Monthly return on the S&P500 index)

# Apple vs. S&P500 returns

What is the estimate for the regression line equation? What is the standard error of regression?

- Population model: AAPL $= \beta_0 + \beta_1$SP $+ \epsilon$, with $\epsilon \sim N(0, \sigma_\epsilon^2)$
  - **Dependent variable:** AAPL (Monthly return on Apple stocks)
  - **Independent variable:** SP (Monthly return on the S&P500 index)
- Sample model: $\widehat{\text{AAPL}} = 0.016 + (1.289)$SP, with $\epsilon \sim N(0, 0.113^2)$

| | | | | | | |
|---|---|---|---|---|---|---|
| **Dependent Variable:** | AAPL | | | | | |
| **Independent Variables:** | SP | | | | | |
| **Regression Statistics** | | | | | | |
| | **R Square** | **Adj.RSqr** | **Std.Err.Reg.** | **# Cases** | **# Missing** | **t(2.5%,366)** |
| | 0.189 | 0.187 | 0.113 | 368 | 0 | 1.966 |
| **Summary Table** | AAPL = 0.016 + 1.289 SP | | | | | |
| **Variable** | **Coeff** | **Std.Err.** | **t-Stat.** | **P-value** | **Lower95%** | **Upper95%** |
| Intercept | 0.016 | 0.005953 | 2.673 | 0.008 | 0.004204 | 0.028 |
| SP | 1.289 | 0.140 | 9.239 | 0.000 | 1.015 | 1.564 |

# Apple vs. S&P500 returns

What is the estimate for the regression line equation? What is the standard error of regression?

- **Population model:** AAPL $= \beta_0 + \beta_1 \text{SP} + \epsilon$, with $\epsilon \sim N(0, \sigma_\epsilon^2)$
  - **Dependent variable:** AAPL (Monthly return on Apple stocks)
  - **Independent variable:** SP (Monthly return on the S&P500 index)
- **Sample model:** $\widehat{\text{AAPL}} = 0.016 + (1.289)\text{SP}$, with $\epsilon \sim N(0, 0.113^2)$

| | |
|---|---|
| **Dependent Variable:** | AAPL |
| **Independent Variables:** | SP |

**Regression Statistics**

| R Square | Adj.RSqr | Std.Err.Reg. | # Cases | # Missing | t(2.5%,366) |
|---|---|---|---|---|---|
| 0.189 | 0.187 | 0.113 | 368 | 0 | 1.966 |

**Summary Table** AAPL $=$ 0.016 $+$ 1.289 SP

| Variable | Coeff | Std.Err. | t-Stat. | P-value | Lower95% | Upper95% |
|---|---|---|---|---|---|---|
| Intercept | 0.016 | 0.005953 | 2.673 | 0.008 | 0.004204 | 0.028 |
| SP | 1.289 | 0.140 | 9.239 | 0.000 | 1.015 | 1.564 |

## Apple vs. S&P500 returns

What is the estimate for the regression line equation? What is the standard error of regression?

- **Population model:** $\text{AAPL} = \beta_0 + \beta_1 \text{SP} + \epsilon$, with $\epsilon \sim N(0, \sigma_\epsilon^2)$
  - **Dependent variable:** AAPL (Monthly return on Apple stocks)
  - **Independent variable:** SP (Monthly return on the S&P500 index)
- **Sample model:** $\widehat{\text{AAPL}} = 0.016 + (1.289)\text{SP}$, with $\epsilon \sim N(0, 0.113^2)$

**Dependent Variable:** AAPL
**Independent Variables:** SP

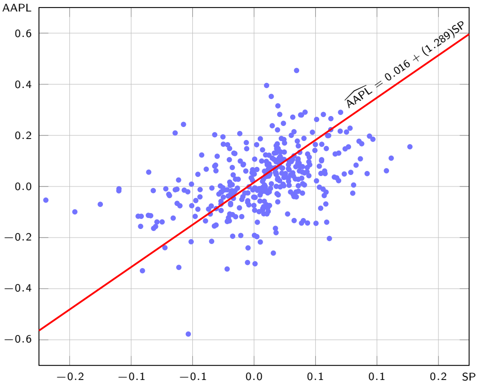
**Regression Statistics**

| R Square | Adj.RSqr | Std.Err.Reg. | # Cases | # Missing | t(2.5%,366) |
|---|---|---|---|---|---|
| 0.189 | 0.187 | 0.113 | 368 | 0 | 1.966 |

**Summary Table** AAPL = 0.016 + 1.289 SP

| Variable | Coeff | Std.Err. | t-Stat. | P-value | Lower95% | Upper95% |
|---|---|---|---|---|---|---|
| Intercept | 0.016 | 0.005953 | 2.673 | 0.008 | 0.004204 | 0.028 |
| SP | 1.289 | 0.140 | 9.239 | 0.000 | 1.015 | 1.564 |

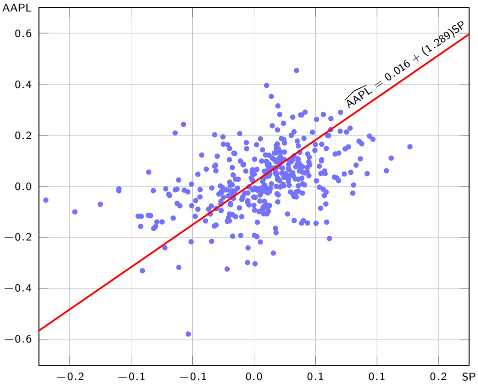
# More on the standard error of regression




## Standard error of regression:

- Best guess for the standard deviation of the error term $\sigma_\epsilon$ (recall that the error term $\epsilon \sim N(0, \sigma_\epsilon^2)$)
- Graphically, it indicates how large the deviations around the regression line are
- The smaller $\sigma_\epsilon$ the better accuracy of the model predictions

# More on the standard error of regression




## Standard error of regression:

- Best guess for the standard deviation of the error term $\sigma_\epsilon$ (recall that the error term $\epsilon \sim N(0, \sigma_\epsilon^2)$)
- Graphically, it indicates how large the deviations around the regression line are
- The smaller $\sigma_\epsilon$ the better accuracy of the model predictions

In the Apple vs. S&P500 context: **Std.Err.Reg = 0.113**

- The standard error of regression is a measure of the unsystematic risk (a.k.a. the idiosyncratic risk)
- The unsystematic risk is inherent to a specific company and is not explained by fluctuations of S&P500

## Apple vs. S&P500 returns

What is the average change in APPL return if SP return increases by 1%? (marginal rate of change)

$= 0.01$

- Population model: $\text{AAPL} = \beta_0 + \beta_1 \text{SP} + \epsilon$, where $\epsilon \sim N(0, \sigma_\epsilon^2)$

$$AAPL_{NEW} = \beta_0 + \beta_1 SP_{NEW} + \varepsilon$$

$$= \beta_0 + \beta_1 (SP_{OLD} + 0.01) + \varepsilon$$

$$= \beta_0 + \beta_1 SP_{OLD} + \beta_1 (0.01) + \varepsilon$$

$$= AAPL_{OLD} + \boxed{\beta_1 (0.01)} + \varepsilon$$

average AAPL change

Average error $= 0$

## Apple vs. S&P500 returns

What is the average change in APPL return if SP return increases by 1%? (marginal rate of change)

- Population model: $\text{AAPL} = \beta_0 + \beta_1 \text{SP} + \epsilon$, where $\epsilon \sim N(0, \sigma_\epsilon^2)$
  - If the S&P500 return equals $r$ (that is, if $\text{SP} = r$), then the Apple return is

$$\text{AAPL} = \beta_0 + \beta_1(r) + \epsilon$$

  - If the S&P500 return changes to $r + 0.01$ (that is, if $\text{SP} = r + 0.01$) and $\epsilon' \sim N(0, \sigma_\epsilon^2)$, then the new Apple return is

$$\text{AAPL}_{\mathrm{new}} = \beta_0 + \beta_1(r + 0.01) + \epsilon'$$

  - The average change in the Apple return is the expected value of the difference between the Apple returns above:

$$\mathbb{E}\left[\text{AAPL}_{\mathrm{new}} - \text{AAPL}\right] = \beta_1(0.01) + \underbrace{\mathbb{E}\left[\epsilon' - \epsilon\right]}_{=0} = \beta_1(0.01)$$

# Apple vs. S&P500 returns

What is the average change in APPL return if SP return increases by 1%? (marginal rate of change)

$c$

- **Population model:** $\text{AAPL} = \beta_0 + \beta_1 \text{SP} + \epsilon$, where $\epsilon \sim N(0, \sigma_\epsilon^2)$
  - If the S&P500 return equals $r$ (that is, if $\text{SP} = r$), then the Apple return is
    $$\text{AAPL} = \beta_0 + \beta_1(r) + \epsilon$$
  - If the S&P500 return changes to $r + 0.01$ (that is, if $\text{SP} = r + 0.01$) and $\epsilon' \sim N(0, \sigma_\epsilon^2)$, then the new Apple return is
    $$\text{AAPL}_{\text{new}} = \beta_0 + \beta_1(r + 0.01) + \epsilon'$$
  - The average change in the Apple return is the expected value of the difference between the Apple returns above:
    $$\mathbb{E}\left[\text{AAPL}_{\text{new}} - \text{AAPL}\right] = \beta_1(0.01) + \underbrace{\mathbb{E}\left[\epsilon' - \epsilon\right]}_{=0} = \beta_1(0.01)$$

- **Best estimate for the average change:** estimate $\beta_1$ with its point estimate $\widehat{\beta}_1 = 1.289$ from the regression output and multiply it with the amount of change $c$:
  $$1.289 \times c$$

# The (finance) beta of Apple against the S&P500 index

What is the beta for Apple against the S&P500 index?

- If $c$ is the change in the S&P500 return, then the best estimate for the average change in the Apple return is equal to $(1.289)c$ (Note: this is regardless of the initial value of the S&P500 return)

- Positive or negative changes in APPL tend to be 1.289 times larger than positive or negative changes in SP

**1.289 is the *beta* of Apple against the S&P500 index; it is just the slope of the regression line**

# The (finance) beta of Apple against the S&P500 index

What is the beta for Apple against the S&P500 index?

- If $c$ is the change in the S&P500 return, then the best estimate for the average change in the Apple return is equal to $(1.289)c$ (Note: this is regardless of the initial value of the S&P500 return)
- Positive or negative changes in APPL tend to be 1.289 times larger than positive or negative changes in SP

**1.289 is the *beta* of Apple against the S&P500 index; it is just the slope of the regression line**

**The (finance) beta:**

- The beta is a financial measure of risk that quantifies how the individual investment (APPL) and the market benchmark (SP) returns move together
- The beta represents the systematic risk of the investment (APPL) that is inherent to the entire market (SP).
- Beta above 1.0: investment is more risky than the market benchmark
- Beta below 1.0: investment is less risky than the market benchmark

# Apple vs. S&P500 returns

Confidence that the beta for Apple is larger than one?

| | | | | | | |
|---|---|---|---|---|---|---|
| **Dependent Variable:** | AAPL | | | | | |
| **Independent Variables:** | SP | | | | | |
| **Regression Statistics** | | | | | | |
| | **R Square** | **Adj.RSqr** | **Std.Err.Reg.** | **# Cases** | **# Missing** | **t(2.5%,366)** |
| | 0.189 | 0.187 | 0.113 | 368 | 0 | 1.966 |
| **Summary Table** | AAPL = 0.016 + 1.289 SP | | | | | |
| **Variable** | **Coeff** | **Std.Err.** | **t-Stat.** | **P-value** | **Lower95%** | **Upper95%** |
| Intercept | 0.016 | 0.005953 | 2.673 | 0.008 | 0.004204 | 0.028 |
| SP | 1.289 | 0.140 | 9.239 | 0.000 | 1.015 | 1.564 |

# Apple vs. S&P500 returns

Confidence that the beta for Apple is larger than one?

**Dependent Variable:** AAPL
**Independent Variables:** SP

**Regression Statistics**

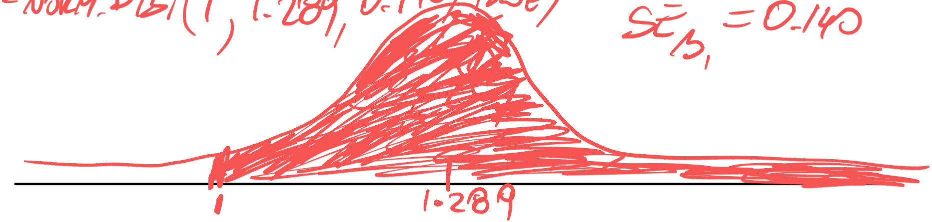
| R Square | Adj.RSqr | Std.Err.Reg. | # Cases | # Missing | t(2.5%,366) |
|---|---|---|---|---|---|
| 0.189 | 0.187 | 0.113 | 368 | 0 | 1.966 |

**Summary Table** AAPL = 0.016 + 1.289 SP

| Variable | Coeff | Std.Err. | t-Stat. | P-value | Lower95% | Upper95% |
|---|---|---|---|---|---|---|
| Intercept | 0.016 | 0.005953 | 2.673 | 0.008 | 0.004204 | 0.028 |
| SP | 1.289 | 0.140 | 9.239 | 0.000 | 1.015 | 1.564 |

1-NORM.DIST(1, 1.289, 0.140, TRUE)

$SE_{B_1} = 0.140$



## Apple vs. S&P500 returns

Confidence that APPL return is positive in a month in which SP return is 10%?

**Dependent Variable:** AAPL
**Independent Variables:** SP

**Regression Statistics**

| R Square | Adj.RSqr | Std.Err.Reg. | # Cases | # Missing | t(2.5%,366) |
|---|---|---|---|---|---|
| 0.189 | 0.187 | 0.113 | 368 | 0 | 1.966 |

**Forecasted : AAPL**

$$\text{AAPL} = 0.016 + 1.289\ \text{SP}$$

| | SP | Forecast | StErrFst | Lower95% | Upper95% |
|---|---|---|---|---|---|
| Fcst# 1 | 0.100 | 0.145 | 0.113 | -0.078 | 0.368 |

# Apple vs. S&P500 returns

Confidence that APPL return is positive in a month in which SP return is 10%?

| | | | | | | |
|---|---|---|---|---|---|---|
| **Dependent Variable:** | AAPL | | | | | |
| **Independent Variables:** | SP | | | | | |
| **Regression Statistics** | | | | | | |
| | **R Square** | **Adj.RSqr** | **Std.Err.Reg.** | **# Cases** | **# Missing** | **t(2.5%,366)** |
| | 0.189 | 0.187 | 0.113 | 368 | 0 | 1.966 |
| **Forecasted : AAPL** | | AAPL = 0.016 + 1.289 SP | | | | |
| | **SP** | **Forecast** | **StErrFst** | **Lower95%** | **Upper95%** | |
| Fcst# 1 | 0.100 | 0.145 | 0.113 | -0.078 | 0.368 | |

APPL return when SP=.1

$SE_{FCST} = 0.113$

1-NORM.DIST(0, 0.145, 0.113, TRUE)

# Unit 4 Individual Assignment

- Due Monday, 23:59 ET (Durham local time).
- Six questions: includes multiple choice and numeric response formats
- No need to show your work, i.e., no supporting documents needed.

Assignment questions:

- Question 1: correlation
- Questions 2-3: sample model estimates (regression line, standard error of regression)
- Question 4: confidence based on coefficient estimate (about finance beta)
- Questions 5-6: forecast and confidence based on a forecast