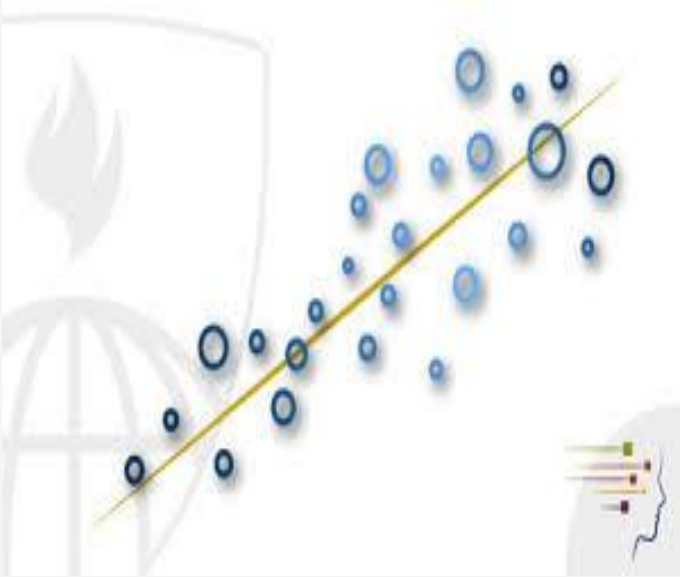
LINEAR REGRESSION AND CORRELATION ANALYSIS

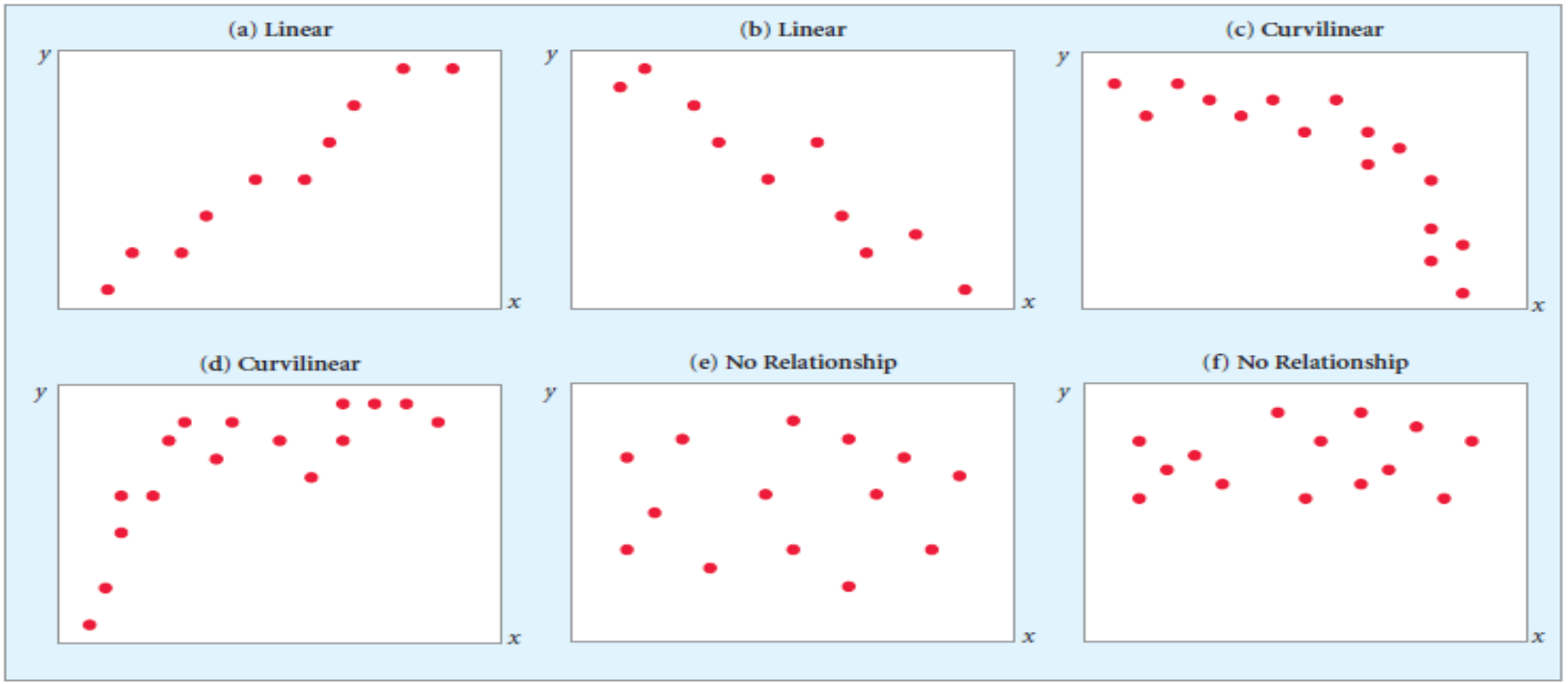
- SCATTER PLOTS AND CORRELATION
- SIMPLE LINEAR REGRESSION ANALYSIS
- USES FOR REGRESSION ANALYSIS



WHY NEED TO KNOW?

- Regression and correlation analysis are most used statistical procedure used by business decision makers for analyzing the relationship between two variables.
- Common visual tool for two-variable relationship: scatter plots (or scatter diagram)
- Scatter plot: A two-dimensional plot showing the values for the joint occurrence of two quantitative variables.
  - Dependent (response) variable, *y* – a variation wish to explain
  - Independent (explanatory) variable, x – used to explain variation in dependent variable

Two- variable relationship



(a) and (b) examples of strong *linear* (or straight line) relationships between x and y

(c) and (d) examples of relationship between the x and y variable is nonlinear.

(e) and (f) examples in which there is no identifiable relationship between the two variables. This means that as x increases, y sometimes increases and sometimes decreases but with no particular pattern.

CORRELATION ANALYSIS

- What can we measure?
  - Strength of the relationship: weak, moderate or strong
  - Direction of the relationship: positive or negative or no relationship
  - Significance test for correlation: correlation exists or not

CORRELATION COEFFICIENT

- A quantitative measure of the strength of the linear relationship between two variables.
- The correlation ranges from -1.0 to +1.0.
  - A correlation of 1.0 indicates a perfect linear relationship
  - A correlation of 0 indicates no linear relationship.
- The more correlation differs from 0.0, the stronger the linear relationship between the two variables.
- Sign (+ or -) of the correlation coefficient indicates the direction of relationship

CORRELATION COEFFICIENT

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{[n(\sum x^2) - (\sum x)^2][n(\sum y^2) - (\sum y)^2]}}$$ OR $$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{[\sum(x - \overline{x})^2][\sum(y - \overline{y})^2]}}$$

where:

r = Sample correlation coefficient
n = Sample size
x = value of the independent variable
y = value of the dependent variable

also called as Pearson Product Moment Correlation

However, because the calculations are rather tedious and long, need to have computer/statistical tool to perform the computation.

SIGNIFICANCE TEST FOR THE CORRELATION

- **What it is?:** a hypothesis testing to determine whether there exists relationship between the variables.
- Hypothesis: using the ρ (rho) notation

  H_0: ρ = 0 (no correlation)

  H_A: ρ ≠ 0 (correlation exists)
- Test statistics for correlation:

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} \qquad df = n-2$$

where:

t = Number of standard errors r is from 0

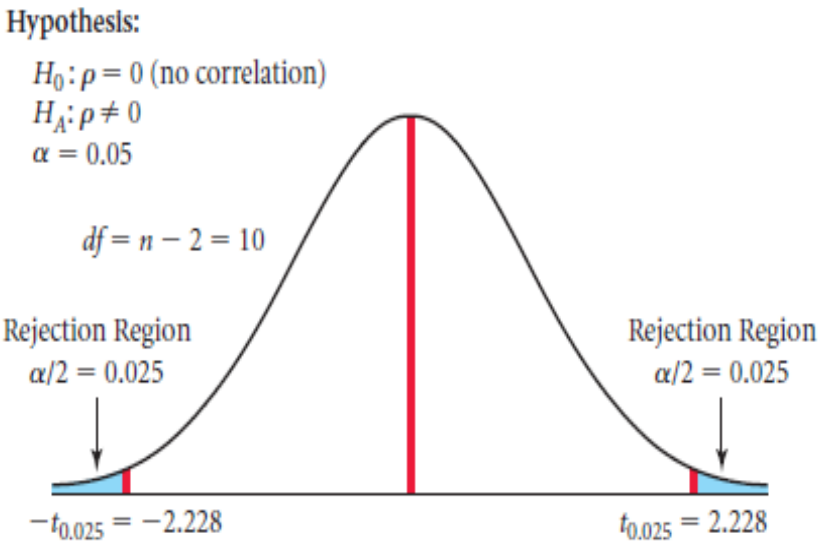
r = Sample correlation coefficient

n = Sample size

SIGNIFICANCE TEST FOR THE CORRELATION

Decision rule: calculate a t value and compare it to the critical value, t in the t-distribution table. Since it involve with 2 tailed;

- **Lower tail**
  - if the $-t \geq$ critical value, $-t_{\alpha/2}$ therefore accept H_o
  - if the $-t <$ critical value, $-t_{\alpha/2}$ therefore reject H_o
- **Upper tail**
  - if the $t \geq$ critical value, $t_{\alpha/2}$ therefore reject H_o
  - if the $t <$ critical value, $t_{\alpha/2}$ therefore accept H_o



CORRELATION ANALYSIS- EXAMPLE

A student intern at the investment firm of McMillan & Associates was given the assignment of determining whether there is a positive correlation between the number of individual stocks in a client's portfolio and the annual rate of return for the portfolio. The intern selected a simple random sample of 10 client portfolios and determined the number of individual company stocks and the annual rate of return earned by the client on his or her portfolio. To determine whether there is a statistically significant positive correlation between the two variables. The following sample data were obtained:

Number of Stocks	Rate of Return
9	0.13
16	0.16
25	0.21
16	0.18
20	0.18
16	0.19
20	0.15
20	0.17
16	0.13
9	0.11

Steps to solution:

Specify the population parameter of interest:
The intern wishes to determine whether the number of stocks is positively correlated with the rate of return earned by the client. The parameter of interest is, therefore, the population correlation, ρ

Formulate the null and the alternative hypothesis

$$H_0 : \rho \leq 0$$
$$H_A : \rho > 0 \text{ (claim)}$$

Construct the rejection region and decision rule
This is one-tailed test with upper tail (right hand) of sampling distribution. With 0.05 level of significance, the degree of freedom, n-2 = 10 – 2 = 8 degree of freedom. The t-critical value, $t_{0.05}$= 1.860
Decision rule: If $t > 1.860$, then reject H_0, otherwise accept H_0

Compute the correlation coefficient and test statistics
compute the sample correlation coefficient using software tools
r = 0.7796

Steps to solution:

Compute the test statistic

This problem will use t-test statistics.

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.7796}{\sqrt{\frac{1-0.7796^2}{10-2}}} = 3.52$$

The decision rule: If $t > 1.860$, then reject H_0, otherwise accept H_0

Since $3.52 > 1.860$, therefore H_0 is rejected

Draw a conclusion

Because the null hypothesis is rejected, the sample data do support that there is a positive linear relationship between the number of individual stocks in a client's portfolio and the portfolio's rate of return.

CORRELATION ANALYSIS- EXAMPLE (R-programming)

The investment firm Harmonic Investments wants to manage the pension fund of a major Chicago retailer. For their presentation to the retailer, the Harmonic analysts want to use correlation analysis to know the relationship between profits and number of employees for 50 Fortune 500 companies in the firm's portfolio. The data for the analysis are contained in the file **Fortune50**.

SOLUTION:

Scatter plot for independent and dependent variable

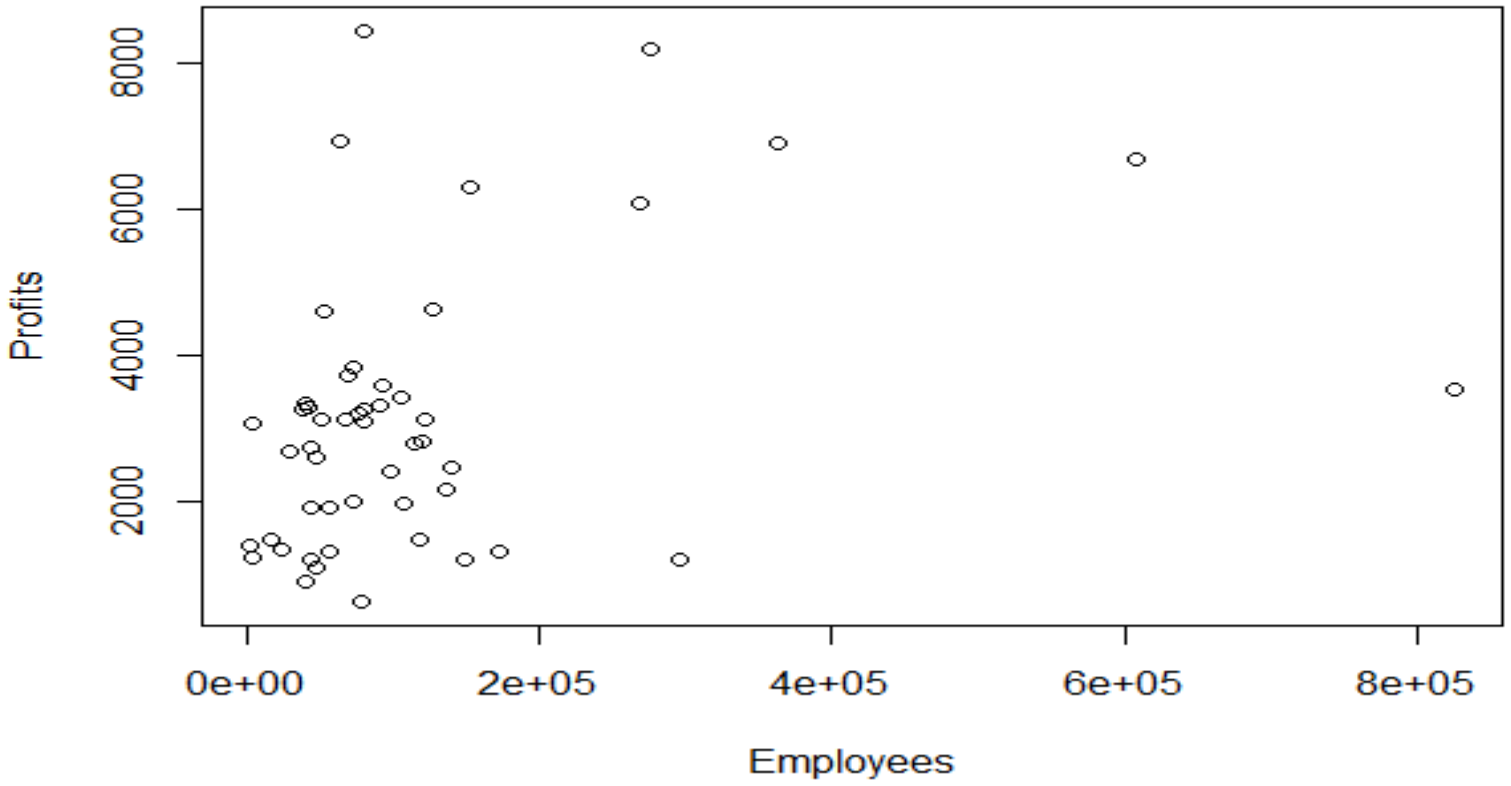
y axis (dependent) – profits; x axis (independent) – employees

Significance test for the correlation and calculate the correlation coefficient (pearson product moment correlation or r-value)

Significance test: To test either is there any relationship between variable "Profits" with variable "Employees".

r-value: to see the strength of the relationship

> plot(Profits ~ Employees, data=Fortune50)



summary: scatter plot shown fairly linear relationship

Formulate the null and the alternative hypothesis

$H_0 : \rho = 0$

$H_A : \rho \neq 0$ (claim)

```
> with(Fortune50, cor.test(Profits, Employees))

        Pearson's product-moment correlation

data:  Profits and Employees
t = 2.7057, df = 48, p-value = 0.009409
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.09504911 0.58308036
sample estimates:
      cor
0.3637726
```

correlation coefficient value:
r = 0.363 indicate weak positive linear relationship

p-value < α = 0.05, therefore, reject H_0 which indicate there is a correlation (relationship) between the variables

SIMPLE LINEAR REGRESSION ANALYSIS

THE REGRESSION MODEL & ASSUMPTIONS

- Statistical method which is used to analyze the relationship → Regression Analysis
- When we have 2 variables → referred to simple regression analysis
- When the 2 variables (independent & dependent) have linear relationship → referred as simple linear regression

THE REGRESSION MODEL & ASSUMPTIONS

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where:

y = Value of the dependent variable
x = Value of the independent variable
β_0 = Population's y intercept
β_1 = Slope of the population regression line
ε = Random error term

Linear Component

Random error component - maybe positive, zero or negative

THE REGRESSION MODEL & ASSUMPTIONS

$$y = \beta_0 + \beta_1 x + \varepsilon$$

where:

y = Value of the dependent variable
x = Value of the independent variable
β_0 = Population's y intercept
β_1 = Slope of the population regression line
ε = Random error term

Population regression coefficient

β_1(slope): can be either positive, zero or negative.
Example:

1. if (β_1=12), it indicate that for a 1 unit increase in x we can expect an average 12 unit increase in y
2. if (β_1=-12), it indicate that for a 1 unit decrease in x we can expect an average 12 unit decreases in y

THE REGRESSION MODEL & ASSUMPTIONS

$$y = \beta_0 + \beta_1 x + \varepsilon$$

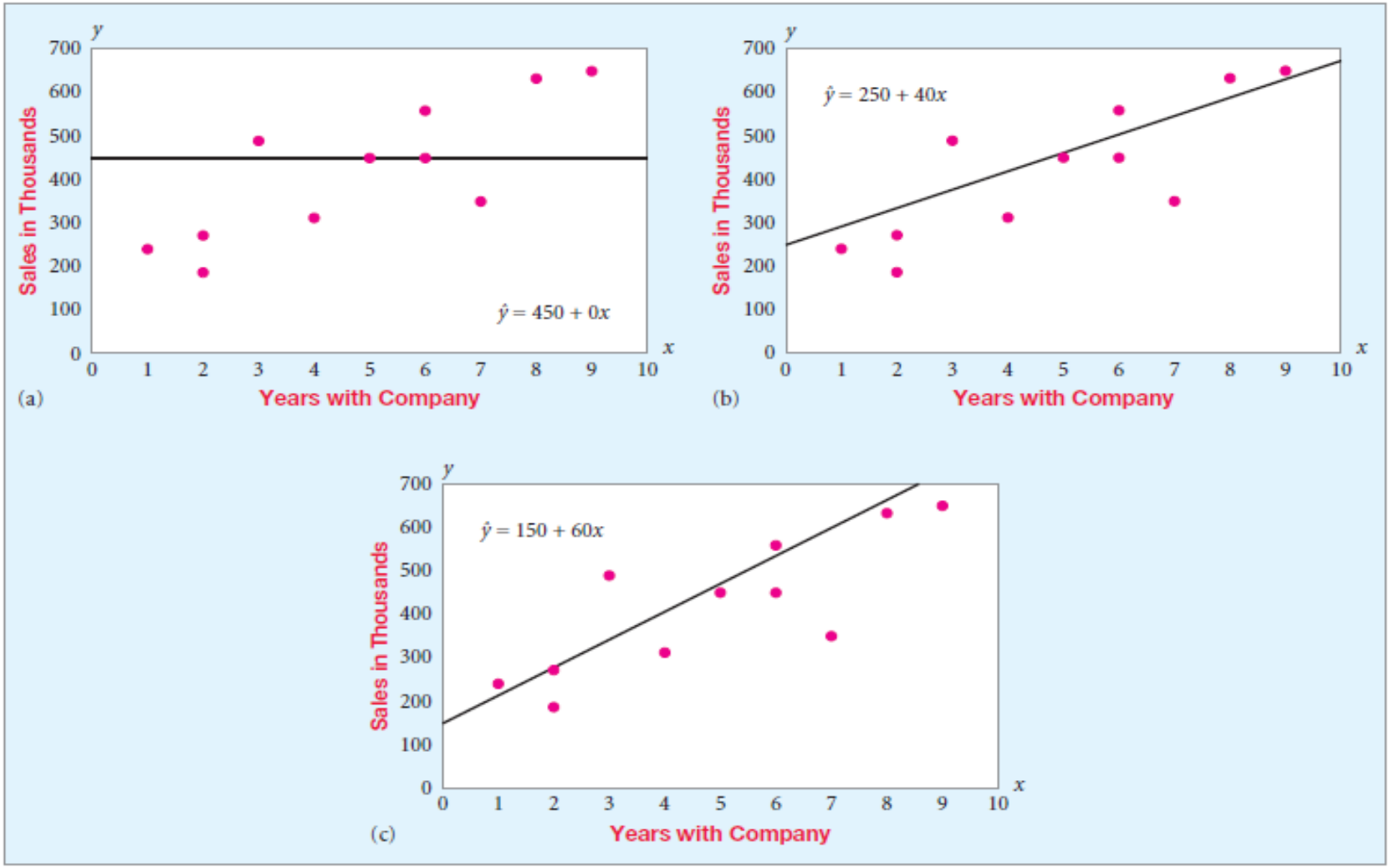
where:

y = Value of the dependent variable
x = Value of the independent variable
β_0 = Population's y intercept
β_1 = Slope of the population regression line
ε = Random error term

β_0(intercept):
- The population's *y* intercept, indicates the mean value of *y* when x is 0.

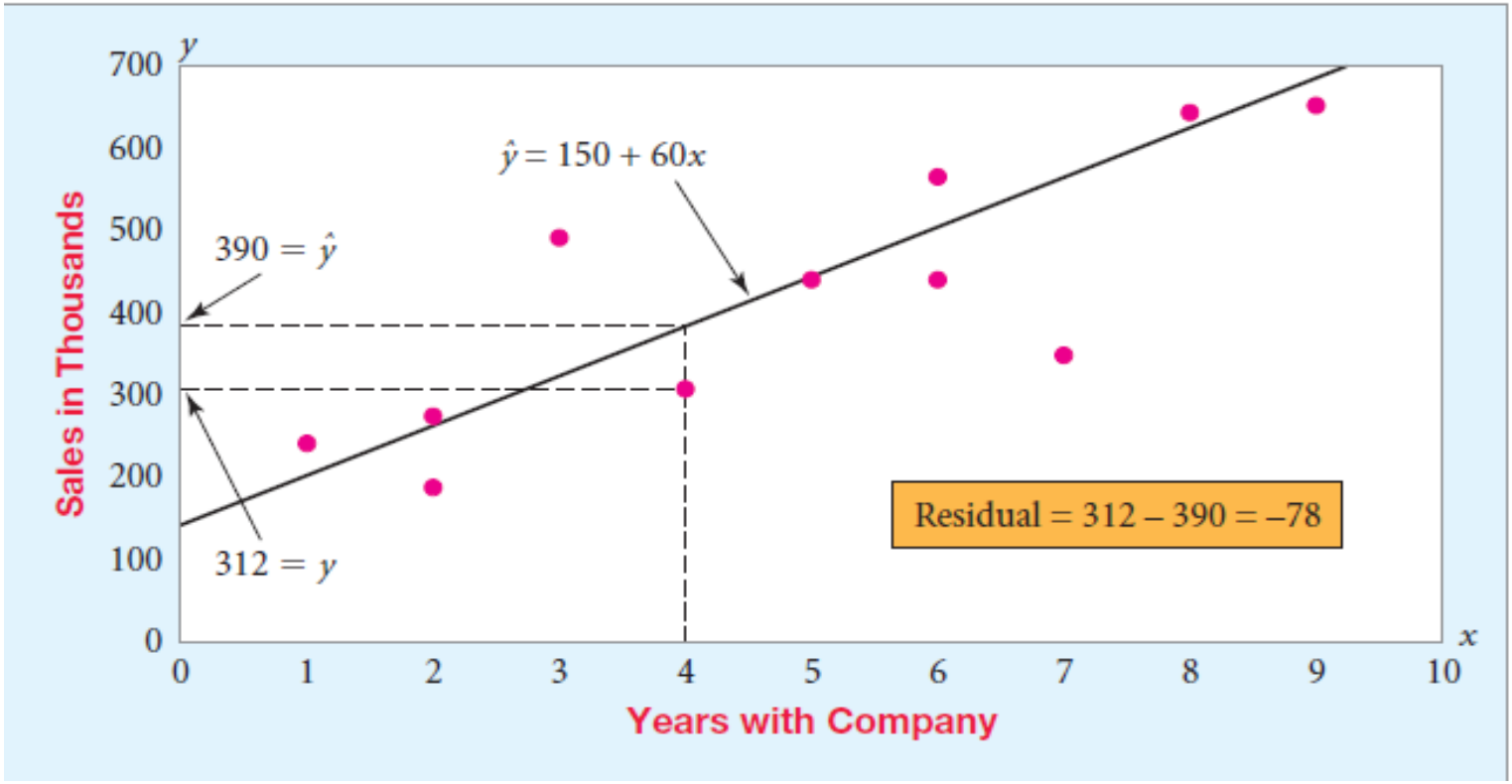
REGRESSION ANALYSIS

- To estimate regression equation (model), we need to identify the value of β_0 and β_1.
- From the regression model we can identify the regression line.
  - regression line: through the sample data is the best estimate of the population regression line
  - we must establish a criterion for selecting the best line
    - the criteria is called as LEAST SQUARES CRITERION
    - Least Square criterion: criteria to determine a regression line that minimize the sum of squares prediction error
  - Prediction error ➔ Residual (the difference between the actual value of the dependent variable (y) and the value predicted ($\hat{y}$) by the regression model).



Possible regression line - We must establish a criterion for selecting the best line (**least square criterion**)

REGRESSION ANALYSIS



data value with $x = 4$*:*

$$\hat{y} = 150 + 60(4) = 390$$

RESIDUAL ANALYSIS

$\hat{y} = 150 + 60x$

			Residual	
x	$\hat{y}$	y	$y - \hat{y}$	$(y - \hat{y})^2$
3	330	487	157	24,649
5	450	445	−5	25
2	270	272	2	4
8	630	641	11	121
2	270	187	−83	6,889
6	510	440	−70	4,900
7	570	346	−224	50,176
1	210	238	28	784
4	390	312	−78	6,084
2	270	269	−1	1
9	690	655	−35	1,225
6	510	563	53	2,809
				$\Sigma = 97,667$

an example of manual calculation for residual

Figure: an example of computer calculation for regression

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.8325
R Square	0.6931
Adjusted R Square	0.6624
Standard Error	92.1055
Observations	12

SSE = 84834.2947

ANOVA

	df	SS	MS	F	Significance F
Regression	1	191600.6720	191600.6	22.5853	0.0008
Residual	10	84834.2947	8483.429		
Total	11	276434.9167			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	175.8288	54.9899	3.1975	0.0095	53.3037	298.3539
Years with Midwest	49.9101	10.5021	4.7524	0.0008	26.5100	73.3102

Estimated regression equation is
$\hat{y} = 175.8288 = 49.9101(x)$

SIMPLE REGRESSION ANALYSIS- EXAMPLE (R-programming)

The investment firm Harmonic Investments wants to manage the pension fund of a major Chicago retailer. Correlation analysis have been done and it shows that the relationship between profits and number of employees for 50 Fortune 500 companies in the firm's portfolio are positive linear relationship. Next, the Harmonic wants to use simple linear regression analysis for their presentation to the retailer. The data for the analysis are contained in the file **Fortune50**.

SOLUTION:

Identify independent (x) and dependent (y) variable

Calculate the correlation coefficient (r) and the linear regression equation/model

Regression Analysis

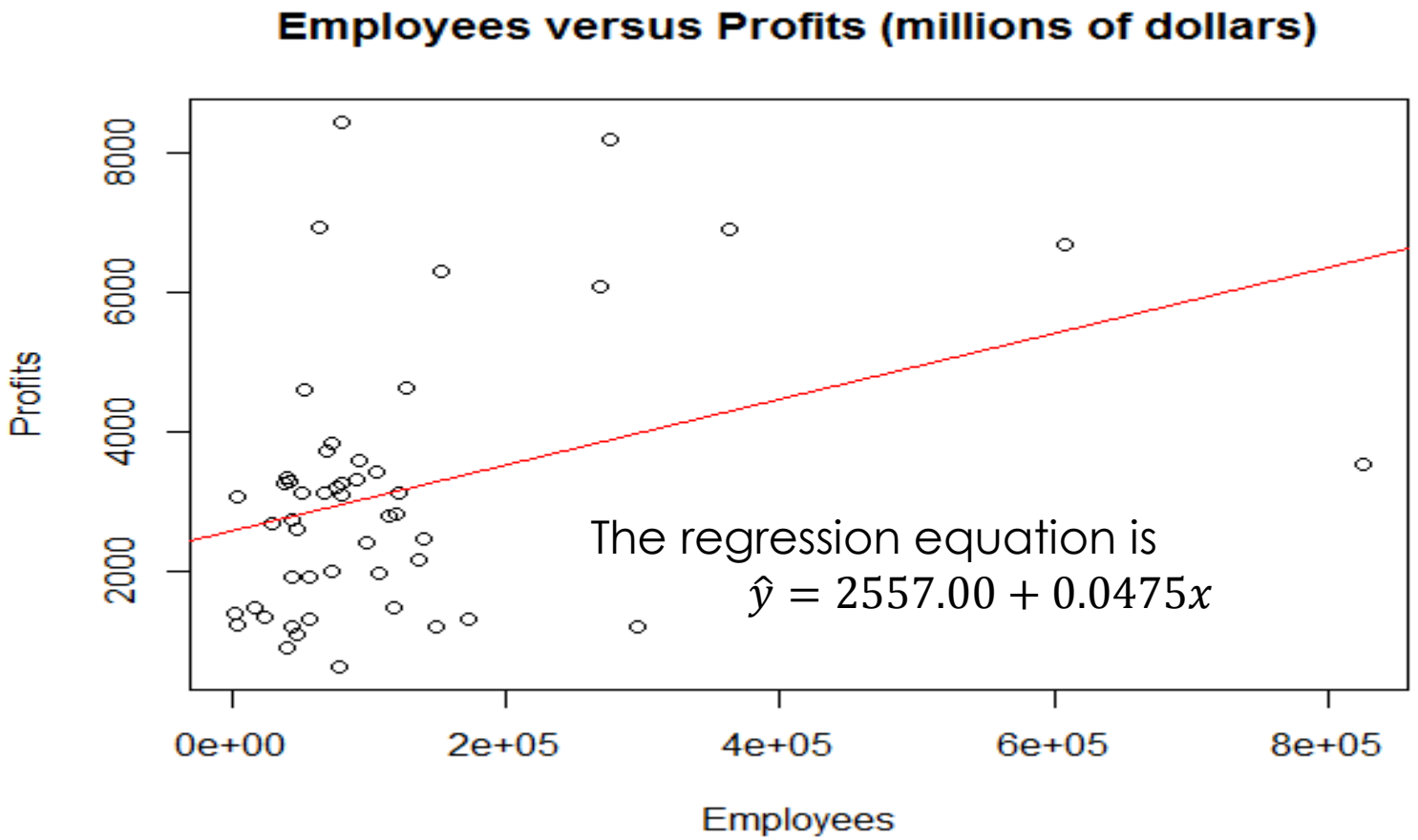
```
> lm.out = lm(Profits ~ Employees, data=Fortune50)     # name the output
> lm.out

Call:
lm(formula = Profits ~ Employees, data = Fortune50)

Coefficients:
(Intercept)    Employees
  2.557e+03    4.759e-03

> summary(lm.out)

Call:
lm(formula = Profits ~ Employees, data = Fortune50)

Residuals:
     Min       1Q   Median       3Q      Max
-2957.1 -1176.0    -38.0    566.8   5522.4

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.557e+03  3.283e+02   7.788 4.61e-10 ***
Employees   4.759e-03  1.759e-03   2.706  0.00941 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1794 on 48 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.1323,    Adjusted R-squared:  0.1143
F-statistic: 7.321 on 1 and 48 DF,  p-value: 0.009409
```

β_1 - slope

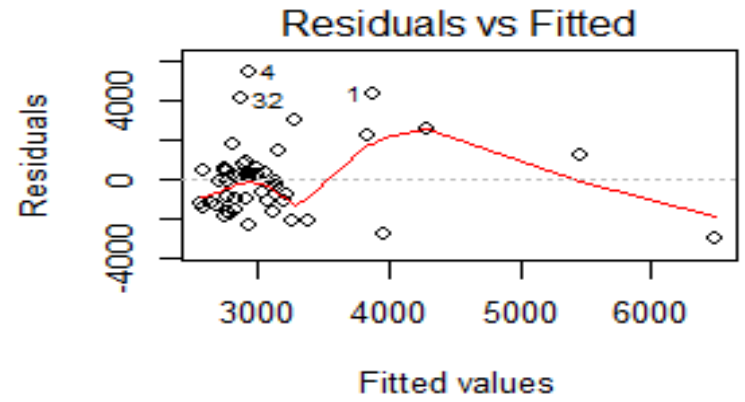
β_0 - intercept

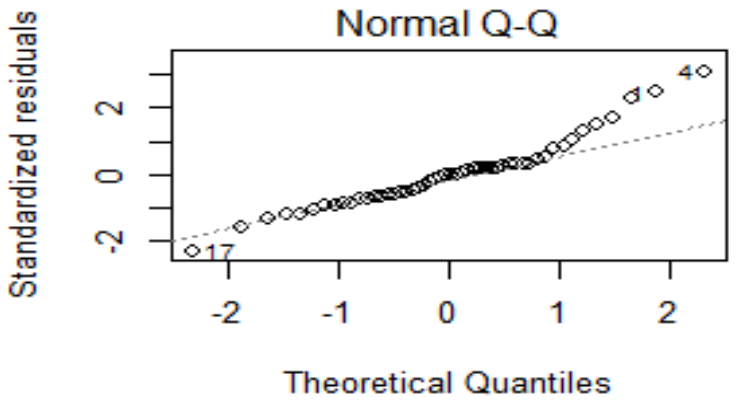
regression coefficient

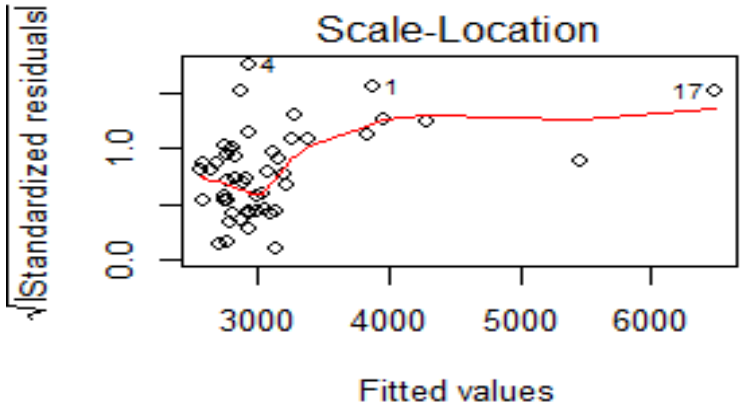
The regression equation is $\hat{y} = 2557.00 + 0.0475x$

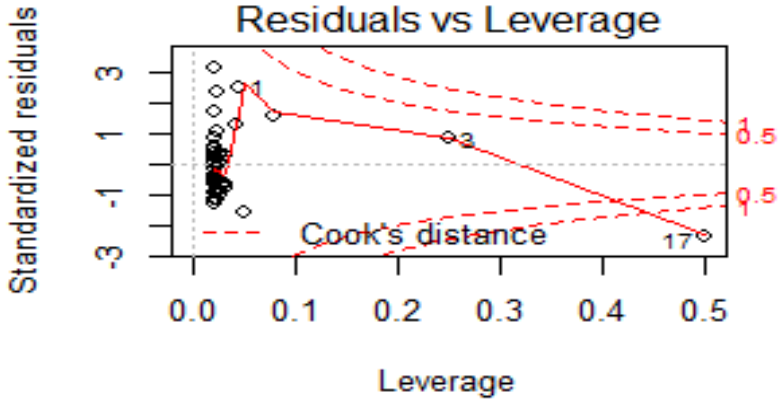



```
> plot(Profits ~ Employees, data=Fortune50, main="Employees versus
Profits (millions of dollars)")
> abline(lm.out, col="red")
```










```
> par(mfrow=c(2,2)) # partition the graphics device
> plot(lm.out)
```

REGRESSION ANALYSIS PLOT

- Other graphic display of how good the model fit is can be achieved as follows:
  - Residual vs fitted - Points that tend towards being outliers are labeled
  - Normal Q-Q - to see our residuals normally distributed
  - Residual vs Leverage - Labeled points on this plot represent cases we may want to investigate as possibly having undue influence on the regression relationship.

SIGNIFICANCE TEST IN REGRESSION ANALYSIS

- Why need significance test?:
  - Regression coefficient compute from sample data is point estimate. Thus, it subject to sampling error.
- There are 3 equivalent significant test in regression analysis:
  1. Test for significance of the correlation between x and y. ---- cover in correlation analysis
  2. Test for significance of the coefficient of determination. ---- next slide
  3. Test for significance of the regression slope coefficient. ---- next next slide ☺

SIGNIFICANCE TEST FOR THE CORRELATION OF DETERMINATION, R^2

- Coefficient of Determination, R^2:
  - The portion of the total variation in the dependent variable that is explained by its relationship with the independent variable.

Coefficient of Determination, R^2

$$R^2 = \frac{SSR}{SST}$$

SIGNIFICANCE TEST FOR THE CORRELATION OF DETERMINATION, R^2

- R^2 value:
  - the value is between 0 and 1.0
  - R^2 with value 1.0 would respond to a situation in which the regression line would pass through each of the points in the scatter plot.

- How the decision makers use R^2 ?:
  - to indicate how well the linear regression line fits the (x,y) data points.
  - the better the fit, the closer R^2 will be to 1.0
  - R^2 will be close to 0 when there is a weak linear relationship

SIGNIFICANCE TEST FOR THE CORRELATION OF DETERMINATION, R^2

- Others measurements in Regression Analysis:
  - SST (total sum of squares): used to measure variation in the dependent variable (y).
  - SSR (sum of square regression): used to measure variation attributed to the relationship between the x's and y's
  - SSE (sum square errors): used to measure the variation attributed to the error. It should be minimize

example from data set Fortune50

```
> options(show.signif.stars=F)
> anova(lm.out)                              # shows an ANOVA table
Analysis of Variance Table

Response: Profits
           Df     Sum Sq  Mean Sq F value   Pr(>F)
Employees   1   23548019 23548019  7.3206 0.009409
Residuals  48  154400512  3216677
```

SST = SSR + SSE

SIGNIFICANCE TEST FOR THE CORRELATION OF DETERMINATION, R^2

- Hypothesis testing to determine whether the correlation of determination have value of ≠ 0
- Hypothesis: using the ρ^2 (rho) notation

$$H_0: \rho^2 = 0$$
$$H_A: \rho^2 > 0$$

- Test statistics for correlation of determination: using the F-test statistics and compare the critical value from F distribution table.

$$F = \frac{\frac{SSR}{1}}{\frac{SSE}{(n-2)}} \quad df = (D_1 = 1, D_2 = n-2)$$

where:

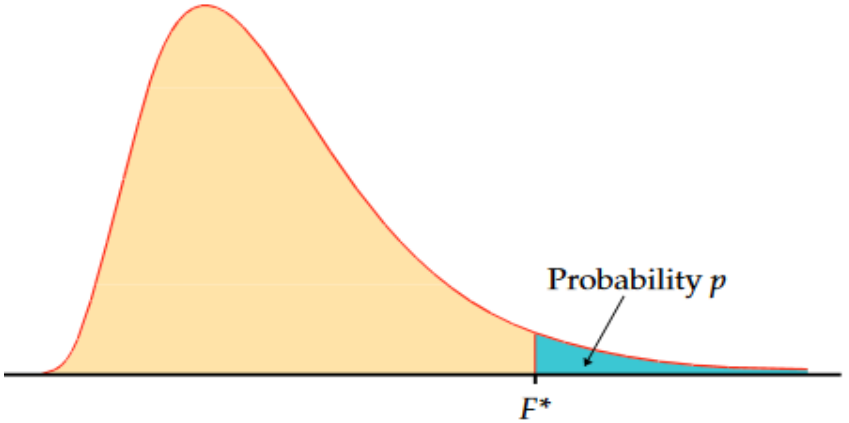
SSR = Sum of squares regression
SSE = Sum of squares error

Test Statistic for Significance of the Coefficient of Determination

SIGNIFICANCE TEST FOR THE CORRELATION OF DETERMINATION, R^2

Decision rule: calculate a F value and compare it to the critical value, F in the f-distribution table.

Since H_A: $\rho^2 > 0$: **Upper tail**



- if the $F \geq$ critical value,$F_\alpha$therefore reject H_o
- if the $F <$ critical value,$F_\alpha$therefore accept H_o

p-value
if the $p \geq$ critical value, α therefore accept H_o
if the $p <$ critical value,αtherefore reject H_o

example from data set Fortune50

```
> lm.out = lm(Profits ~ Employees, data=Fortune50)     # name the output
> lm.out

Call:
lm(formula = Profits ~ Employees, data = Fortune50)

Coefficients:
(Intercept)     Employees
  2.557e+03     4.759e-03

> summary(lm.out)

Call:
lm(formula = Profits ~ Employees, data = Fortune50)

Residuals:
    Min      1Q  Median      3Q     Max
-2957.1 -1176.0   -38.0   566.8  5522.4

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.557e+03  3.283e+02   7.788 4.61e-10 ***
Employees   4.759e-03  1.759e-03   2.706  0.00941 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1794 on 48 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.1323,    Adjusted R-squared:  0.1143
F-statistic: 7.321 on 1 and 48 DF,  p-value: 0.009409
```

p < critical value, α therefore reject H_o

R^2 value

correlation coefficient, r = $\sqrt{R^2} = \sqrt{0.1323} = 0.3637$

$R^2 = r^2$

example from data set Fortune50

Conclusion:
We reject the null hypothesis and conclude the population coefficient of determination (ρ^2) is greater than zero. This mean the independent variable (x) explains a significant proportion of the variation in the dependent variable (y)

SIGNIFICANCE TEST OF THE REGRESSION SLOPE COEFFICIENT

- Why need to test the regression slope coefficient?
  - we are interested in determining whether the population regression slope coefficient is 0.
  - A slope of 0 would imply that there is no linear relationship between x and y variables and that the x variable
- Hypothesis null and alternative to be tested are:

$$H_0: \beta_1 = 0$$
$$H_A: \beta_1 \neq 0$$

Simple Linear Regression Test Statistic for Test of the Significance of the Slope

$$t = \frac{b_1 - \beta_1}{s_{b_1}} \qquad df = n - 2$$

where:

b_1 = Sample regression slope coefficient
β_1 = Hypothesized slope (usually $\beta_1 = 0$)
s_{b_1} = Estimator of the standard error of the slope

Simple Regression Estimator for the Standard Error of the Slope

$$s_{b_1} = \frac{s_\varepsilon}{\sqrt{\sum (x - \bar{x})^2}}$$

where:

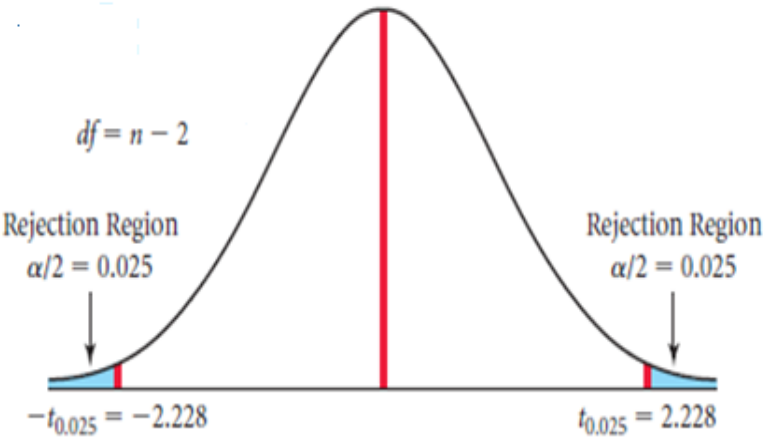
s_{b_1} = Estimate of the standard error of the least squares slope

$s_\varepsilon = \sqrt{\frac{SSE}{n-2}}$ = Sample standard error of the estimate (the measure of deviation of the actual y-values around the regression line)

SIGNIFICANCE TEST OF THE REGRESSION SLOPE COEFFICIENT

Decision rule: calculate a t value and compare it to the critical value, t in the t-distribution table. Since it involv with 2 tailed;

- **Lower tail**
  - if the -$t \geq$ critical value, -$t_{\alpha/2}$ therefore accept H_o
  - if the -$t <$ critical value , -$t_{\alpha/2}$ therefore reject H_o
- **Upper tail**
  - if the $t \geq$ critical value,$t_{\alpha/2}$therefore reject H_o
  - if the $t <$ critical value,$t_{\alpha/2}$therefore accept H_o



p-value
if the $p \geq$ critical value, α therefore accept H_o
if the $p <$ critical value,αtherefore reject H_o

example from data set Fortune50

```
> options(show.signif.stars=F)
> anova(lm.out)                                  # shows an ANOVA table
Analysis of Variance Table

Response: Profits
           Df    Sum Sq  Mean Sq F value   Pr(>F)
Employees   1  23548019 23548019  7.3206 0.009409
Residuals  48 154400512  3216677
```

The F-value and the p-value for testing weather the regression slope = 0.0

Decision: P-value < $\alpha = 0.05$, therefore H_0 is rejected

example from data set Fortune50

```
> lm.out = lm(Profits ~ Employees, data=Fortune50)     # name the output
> lm.out

Call:
lm(formula = Profits ~ Employees, data = Fortune50)

Coefficients:
(Intercept)    Employees
  2.557e+03    4.759e-03

> summary(lm.out)

Call:
lm(formula = Profits ~ Employees, data = Fortune50)

Residuals:
    Min      1Q  Median      3Q     Max
-2957.1 -1176.0   -38.0   566.8  5522.4

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 2.557e+03  3.283e+02   7.788 4.61e-10 ***
Employees   4.759e-03  1.759e-03   2.706  0.00941 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1794 on 48 degrees of freedom
  (2 observations deleted due to missingness)
Multiple R-squared:  0.1323,    Adjusted R-squared:  0.1143
F-statistic: 7.321 on 1 and 48 DF,  p-value: 0.009409
```

standard error of the regression slope = 0.00176

the calculated t-statistics and p-value for testing whether the regression slope = 0

CONFIDENCE INTERVAL ESTIMATE FOR THE REGRESSION SLOPE

- Why need confidence interval estimate for β_1?
  - to describe the relationship between the independent variable and the dependent variable.

$$b_1 \pm ts_{b_1}$$

or equivalently,

$$b_1 \pm t\frac{s_\varepsilon}{\sqrt{\Sigma(x-\bar{x})^2}} \qquad df = n-2$$

where:

s_{b_1} = Standard error of the regression slope coefficient

s_ε = Standard error of the estimate

Confidence Interval Estimate for the Regression Slope, Simple Linear Regression

PROBLEM:

In the summer of 2006, gasoline prices soared to record levels in the United States, heightening motor vehicle customers' concern for fuel economy. Analysts at a major automobile company collected data on a variety of variables for a sample of 30 different cars and small trucks. Included among those data were the Environmental Protection Agency (EPA)'s highway mileage rating and the horsepower of each vehicle. The analysts were interested in the relationship between horsepower (x) and highway mileage (*y*). The data are contained in the file **Automobiles**.

Identify the following:

1. Correlation coefficient analysis
2. Regression coefficient
3. Regression equation model
4. Regression plot
5. Test for significance of the correlation between x and *y* at 95%
6. Test for significance of the coefficient of determination at 95%
7. Test for significance of the regression slope coefficient at 95%
8. Confidence interval estimate for the regression slope at 95%

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

Source

Console ~/

```
> library(readxl)
> Automobiles <- read_excel("D:/AIS_STUFF/SUBJEK PENGAJARAN/MANB1123_Business Stat for DS/excel_data/Automobiles.xlsx")
> View(Automobiles)
> automobiles.out=lm(mileage~HorsePower,data = Automobiles)
> summary(automobiles.out)

Call:
lm(formula = mileage ~ HorsePower, data = Automobiles)

Residuals:
   Min     1Q Median     3Q    Max
-6.596 -2.846  1.132  2.633  6.689

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 31.165764   1.933211  16.121 1.06e-15 ***
HorsePower  -0.028563   0.008214  -3.477  0.00167 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.553 on 28 degrees of freedom
  (43 observations deleted due to missingness)
Multiple R-squared:  0.3016,    Adjusted R-squared:  0.2766
F-statistic: 12.09 on 1 and 28 DF,  p-value: 0.001673

> confint(automobiles.out,'HorsePower',level = 0.95,interval="confidence")
                2.5 %      97.5 %
HorsePower -0.04538933 -0.01173679
>
```

RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins

Source

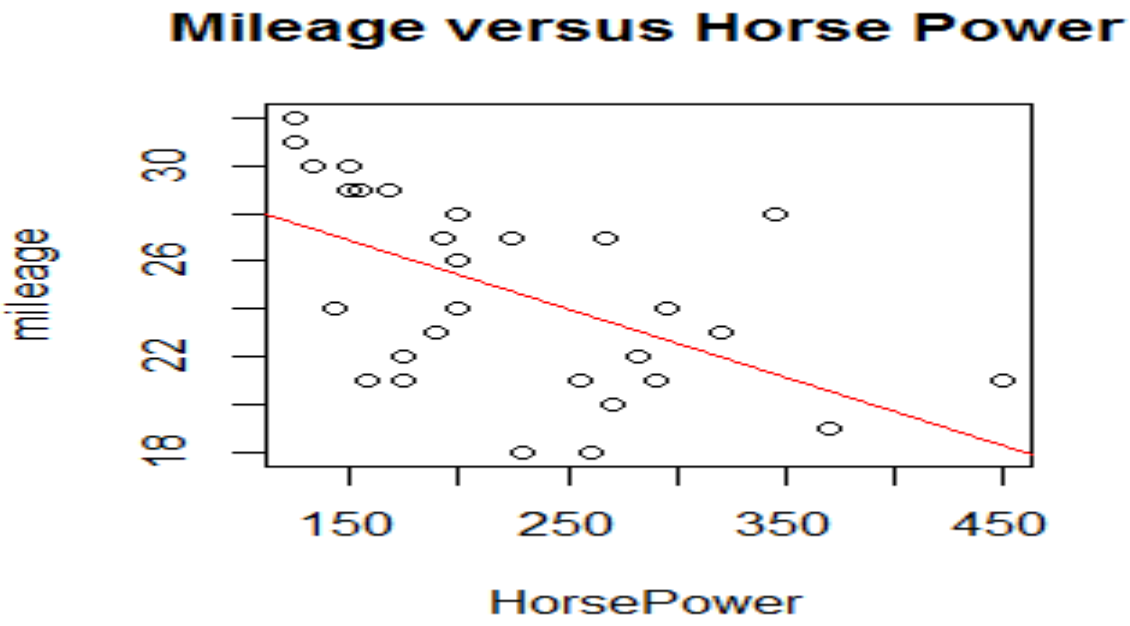
Console ~/

```
> library(readxl)
> Automobiles <- read_excel("D:/AIS_STUFF/SUBJEK PENGAJARAN/MANB1123_Business Stat for DS/excel_data/Automobiles.xlsx")
> View(Automobiles)
> automobiles.out=lm(mileage~HorsePower,data = Automobiles)
> summary(automobiles.out)

Call:
lm(formula = mileage ~ HorsePower, data = Automobiles)

Residuals:
   Min     1Q Median     3Q    Max
-6.596 -2.846  1.132  2.633  6.689

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 31.165764   1.933211  16.121 1.06e-15 ***
HorsePower  -0.028563   0.008214  -3.477  0.00167 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.553 on 28 degrees of freedom
  (43 observations deleted due to missingness)
Multiple R-squared:  0.3016,    Adjusted R-squared:  0.2766
F-statistic: 12.09 on 1 and 28 DF,  p-value: 0.001673

> confint(automobiles.out,'HorsePower',level = 0.95,interval="confidence")
                2.5 %      97.5 %
HorsePower -0.04538933 -0.01173679
>
```

Summary for CI estimate for regression slope:

Thus, at the 95% confidence level, based on the sample data, the analysts for the car company can conclude that a 1-unit increase in horsepower will result in a drop in mileage by an average amount between 0.012 and 0.045 miles per gallon.

```
> plot(mileage ~ HorsePower, data=Automobiles, main="Mileage versus Horse Power")
> abline(automobiles.out, col="red")
```



Solution via R-programming

```
> with(Automobiles, cor.test(mileage, HorsePower))

        Pearson's product-moment correlation

data:  mileage and HorsePower
t = -3.4772, df = 28, p-value = 0.001673
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.7592287 -0.2354968
sample estimates:
       cor
-0.549173

> |
```

THANK YOU

Do more exercise!!!!