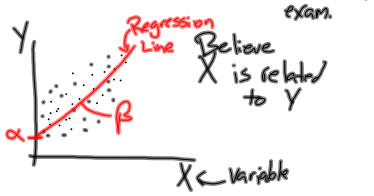
Linear Regression ← Not be on exam.

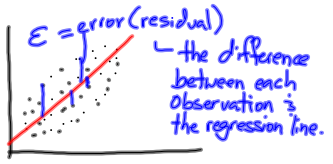


α = value of Y when X=0 (aka intercept)

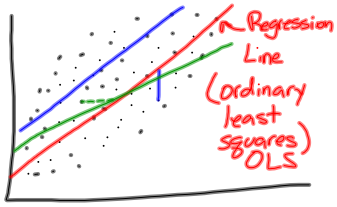
β = slope of the relationship between X & Y

$y = \alpha + \beta X + \varepsilon$ ← error term

Ex: does right-to-work laws influence the percent of workers in the private sector who are unionized?



Regression lines are the minimal residual.



Variable

STATE = state

DENS = density (%) of public workers (aka gov't) who are unionized.

* RTW = whether a state has RTW laws (=1 - yes)

* PVT = the % of private workers who are unionized.

COMP = whether public workers can unionize.

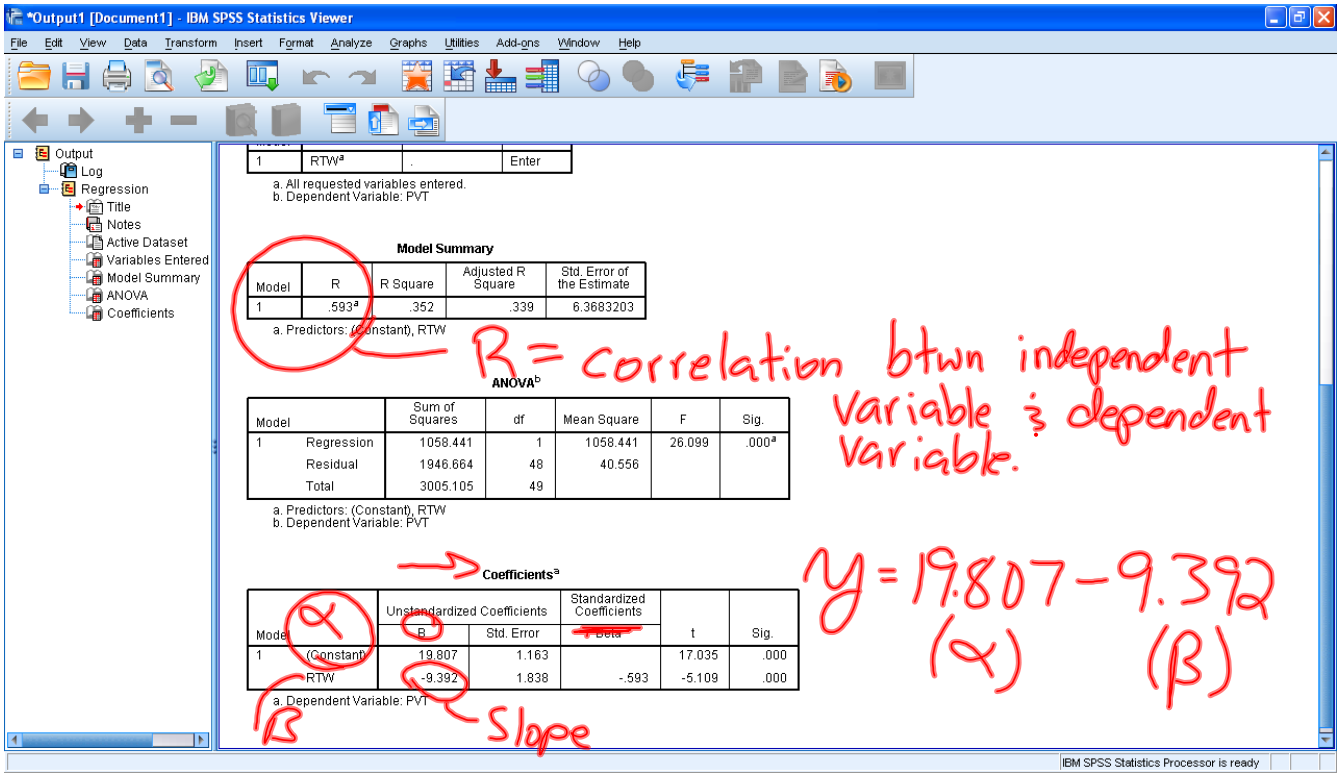
Is PVT related to RTW?

$y = \alpha + \beta X + \varepsilon$

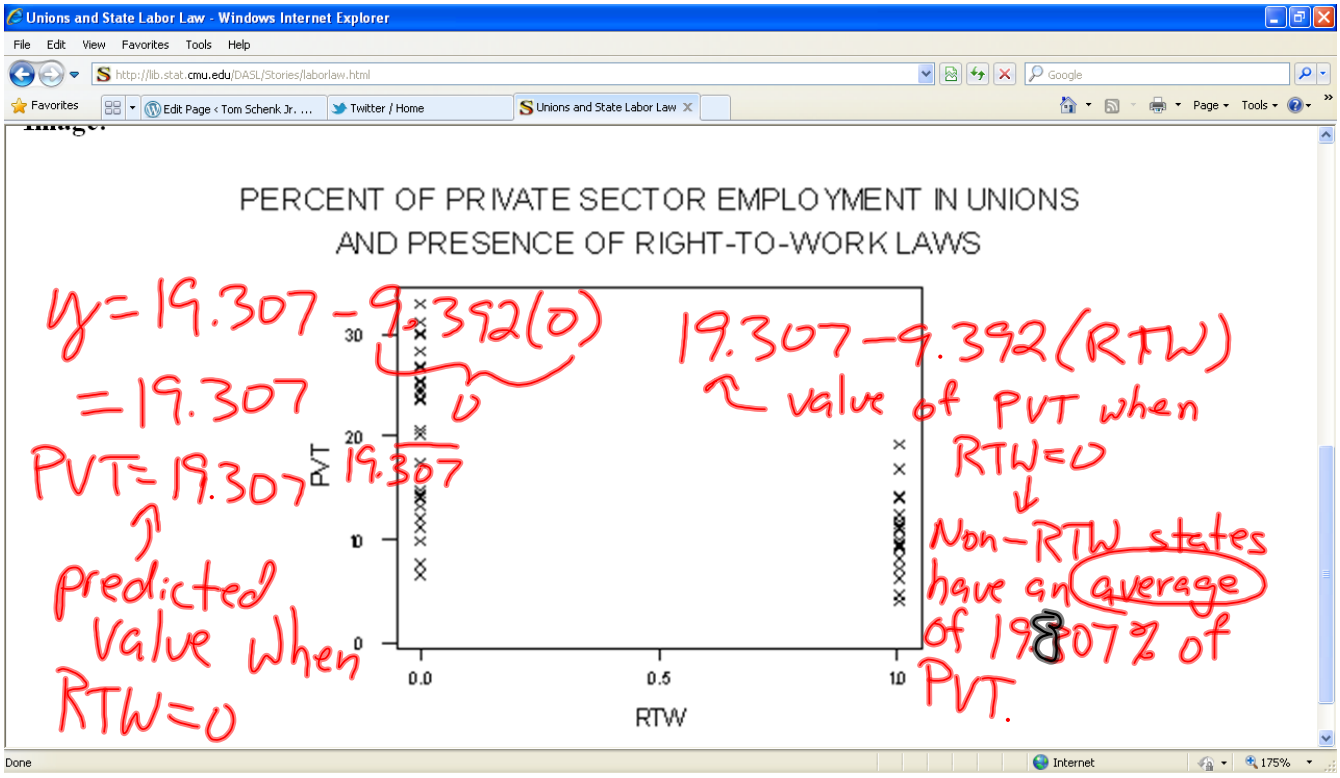
$PVT = \alpha + \beta(RTW) + \varepsilon$

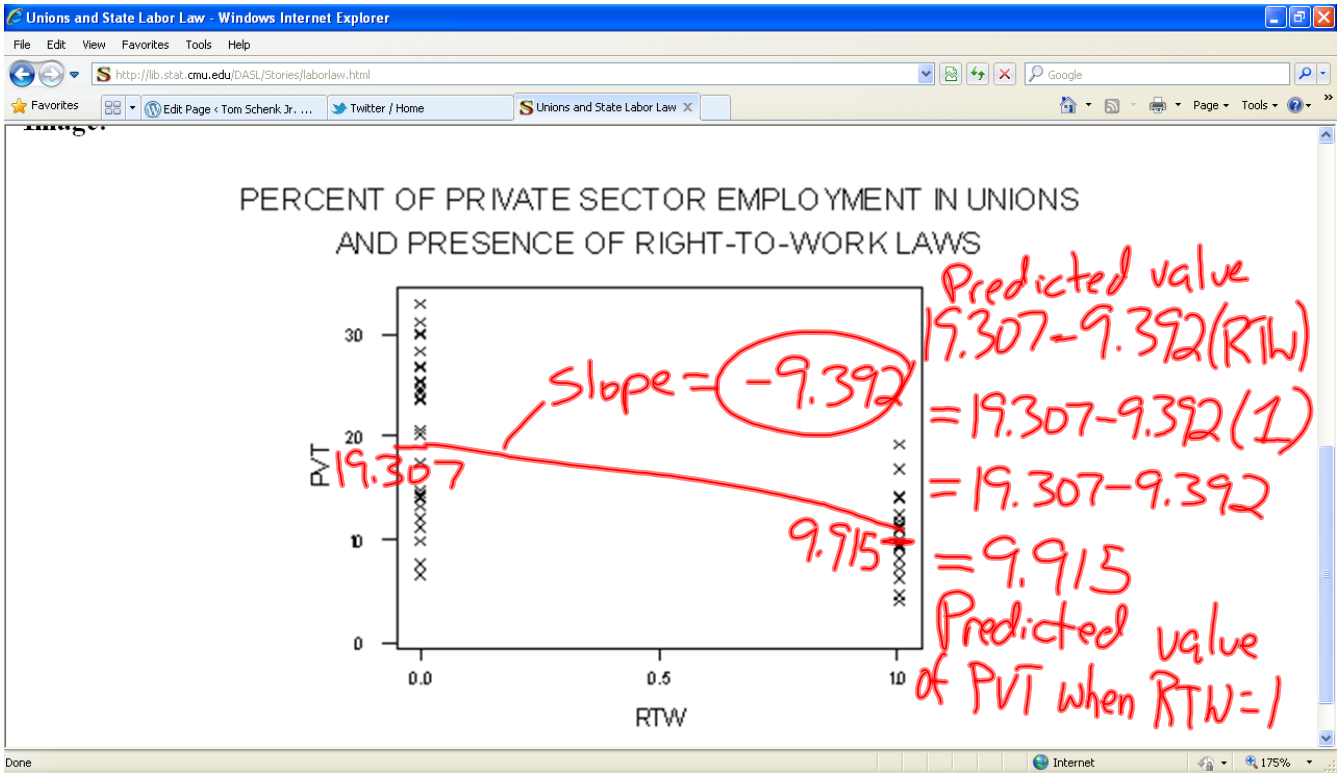
PVT: Dependent variable

RTW: Independent variable


*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help
Output
Log
Regression
Title
Notes
Active Dataset
Variables Entered
Model Summary
ANOVA
Coefficients
1 RTW[a] . Enter
a. All requested variables entered.
b. Dependent Variable: PVT
Model Summary
Model R R Square Adjusted R Square Std. Error of the Estimate
1 .593[a] .352 .339 6.3683203
a. Predictors: (Constant), RTW
R = correlation btwn independent variable & dependent variable.
ANOVA[b]
Model Sum of Squares df Mean Square F Sig.
1 Regression 1058.441 1 1058.441 26.099 .000[a]
Residual 1946.664 48 40.556
Total 3005.105 49
a. Predictors: (Constant), RTW
b. Dependent Variable: PVT
Coefficients[a]
Model Unstandardized Coefficients B Unstandardized Coefficients Std. Error Standardized Coefficients Beta t Sig.
1 (Constant) 19.807 1.163 17.035 .000
RTW -9.392 1.838 -.593 -5.109 .000
a. Dependent Variable: PVT
α
β
Slope
Y = 19.807 - 9.392
(α) (β)
IBM SPSS Statistics Processor is ready

$$y = 19.807 - 9.392(RTW)$$


Unions and State Labor Law - Windows Internet Explorer
File Edit View Favorites Tools Help
http://lib.stat.cmu.edu/DASL/Stories/laborlaw.html
Favorites
Edit Page ‹ Tom Schenk Jr. ...
Twitter / Home
Unions and State Labor Law
Page Tools
PERCENT OF PRIVATE SECTOR EMPLOYMENT IN UNIONS
AND PRESENCE OF RIGHT-TO-WORK LAWS
ŷ = 19.307 − 9.392(0)
= 19.307
PVT = 19.307
predicted value when RTW = 0
19.307 − 9.392(RTW)
value of PVT when RTW = 0
Non-RTW states have an average of 19.307% of PVT.
PVT
RTW
0.0
0.5
1.0
0
10
20
30
Done
Internet
175%


Unions and State Labor Law - Windows Internet Explorer
File Edit View Favorites Tools Help
http://lib.stat.cmu.edu/DASL/Stories/laborlaw.html
Favorites
Edit Page ‹ Tom Schenk Jr. ...
Twitter / Home
Unions and State Labor Law
Page Tools
PERCENT OF PRIVATE SECTOR EMPLOYMENT IN UNIONS
AND PRESENCE OF RIGHT-TO-WORK LAWS
PVT
RTW
0
10
20
30
0.0
0.5
1.0
19.307
9.915
Slope = -9.392
Predicted value
19.307 - 9.392(RTW)
= 19.307 - 9.392(1)
= 19.307 - 9.392
= 9.915
Predicted value of PVT when RTW = 1
Done
Internet
175%

- Regressions allow us to calculate predicted values.

Predicted values can be shown to be the average based on independent variables.

Residuals can provide us more information about the error and "fit" of the regression.

Ex: Look at IA (10)

$RTW_{IA} = 1$

$PVT_{IA} = 16.9$

Does IA values equal our prediction?

$= 19.807 - 9.392(1)$

$= 10.415$

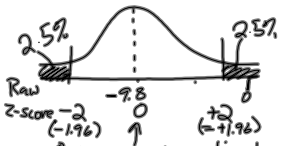
Residual = Actual − Predicted

$(16.9) - 10.415$

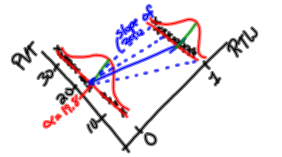
$= 6.485 = \text{residual} = \varepsilon_{IA}$

How do we capture error?

→ Z-scores



Our regression estimates are simply the "average" estimate based on data.



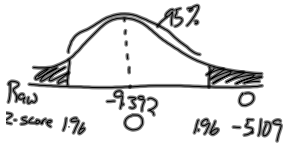
Standard error is the range of possible values for α & β.

- A positive slope means a positive relationship.
- A negative slope means a negative relationship.
- A slope of zero means no relationship.

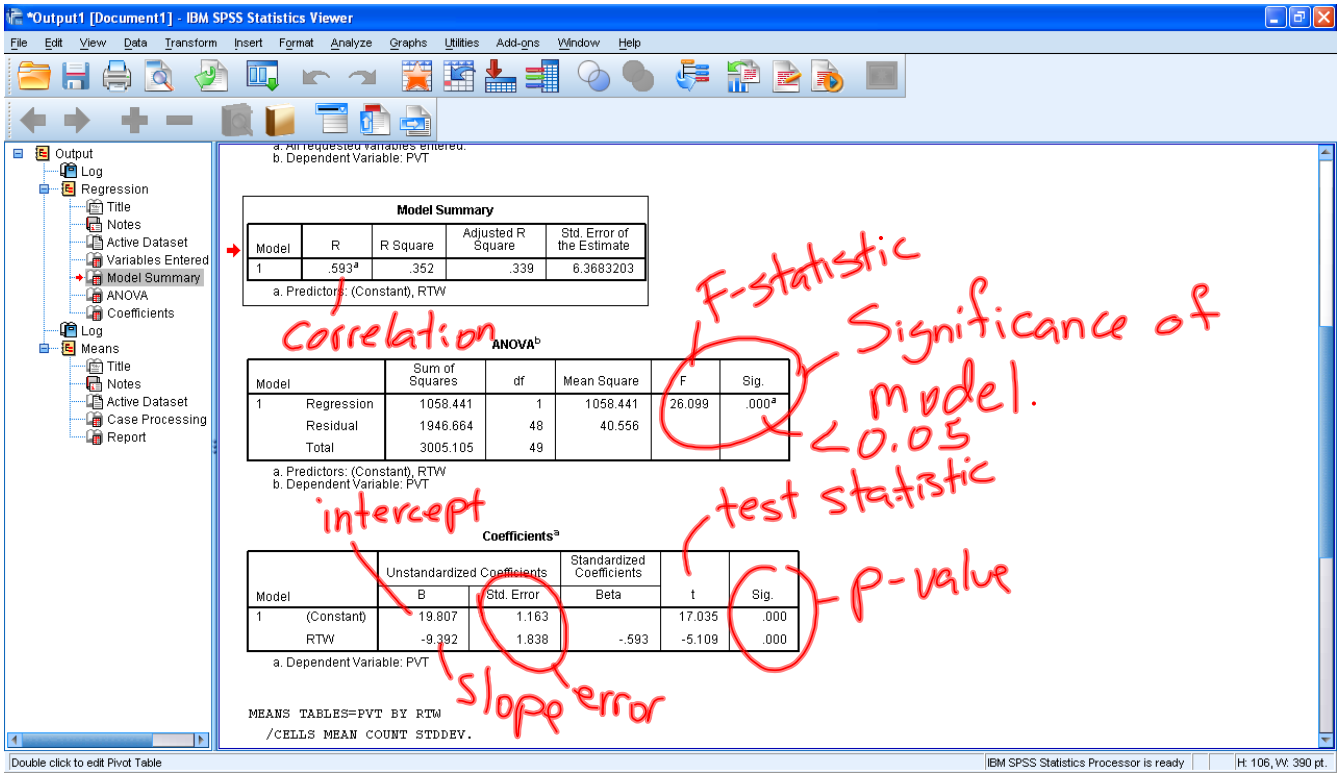
$= \text{test statistic} = \frac{\text{coefficient}}{\text{Standard error of coefficient}}$

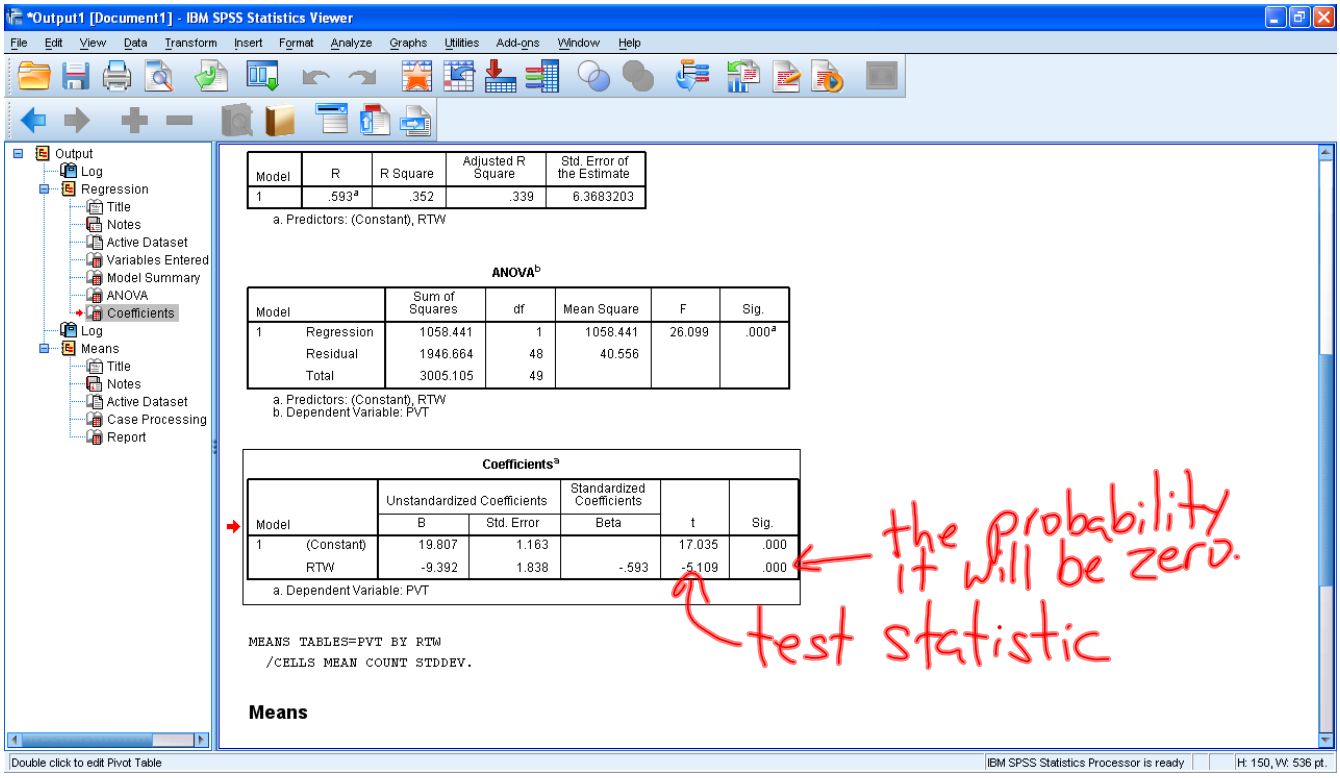
e.g. $\frac{\beta}{SE \text{ of } \beta}$

ex: $\frac{-9.392}{1.838} = -5.109$



Because the absolute value of the test statistic is greater than 1.96, then β is statistically significant. (less than a 5% chance of having a value of zero.)

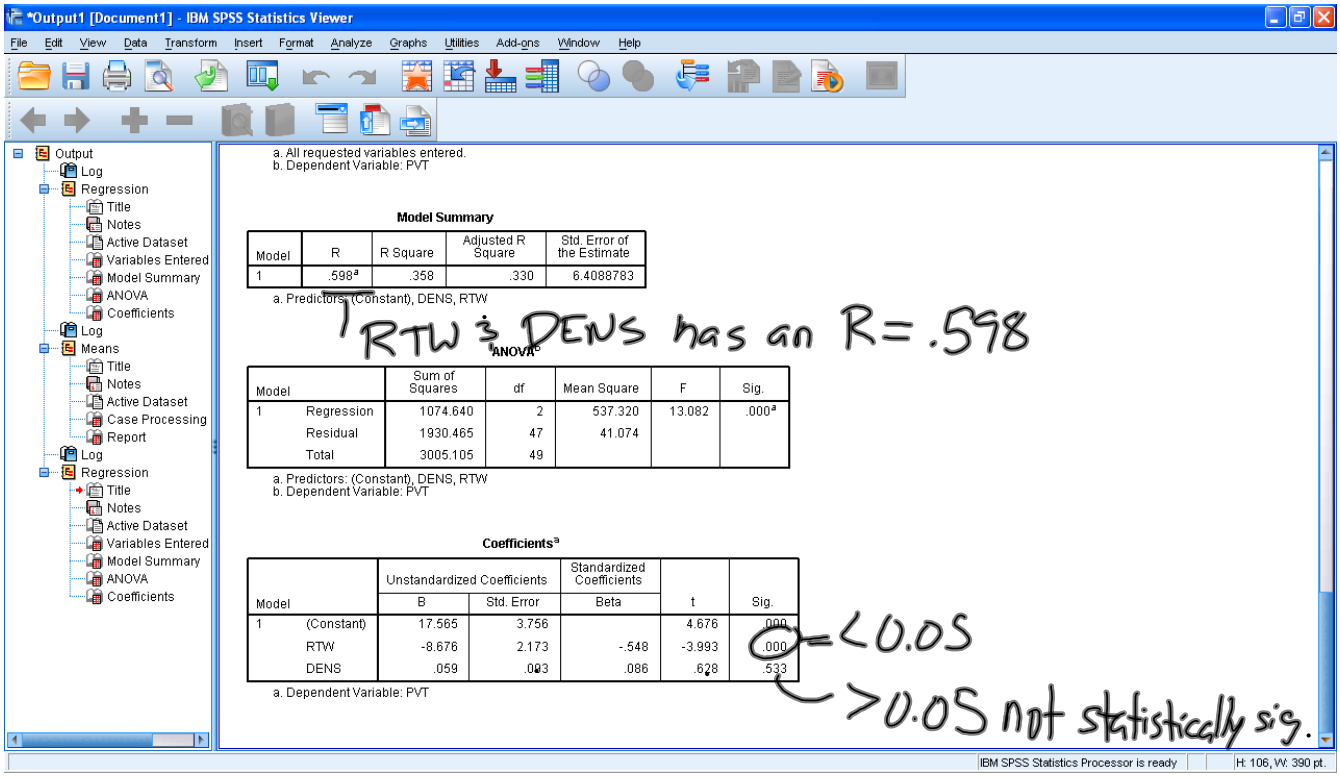

*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help
Output
Log
Regression
Title
Notes
Active Dataset
Variables Entered
Model Summary
ANOVA
Coefficients
Log
Means
Title
Notes
Active Dataset
Case Processing
Report
b. Dependent Variable: PVT
Model Summary
Model | R | R Square | Adjusted R Square | Std. Error of the Estimate
1 | .593[a] | .352 | .339 | 6.3683203
a. Predictors: (Constant), RTW
correlation
ANOVA[b]
Model | | Sum of Squares | df | Mean Square | F | Sig.
1 | Regression | 1058.441 | 1 | 1058.441 | 26.099 | .000[a]
Residual | 1946.664 | 48 | 40.556
Total | 3005.105 | 49
a. Predictors: (Constant), RTW
b. Dependent Variable: PVT
F-statistic
Significance of model.
< 0.05
intercept
test statistic
Coefficients[a]
Model | | Unstandardized Coefficients B | Unstandardized Coefficients Std. Error | Standardized Coefficients Beta | t | Sig.
1 | (Constant) | 19.807 | 1.163 | | 17.035 | .000
RTW | -9.392 | 1.838 | -.593 | -5.109 | .000
a. Dependent Variable: PVT
p-value
slope
error
MEANS TABLES=PVT BY RTW
/CELLS MEAN COUNT STDDEV.
Double click to edit Pivot Table
IBM SPSS Statistics Processor is ready
H: 106, W: 390 pt.

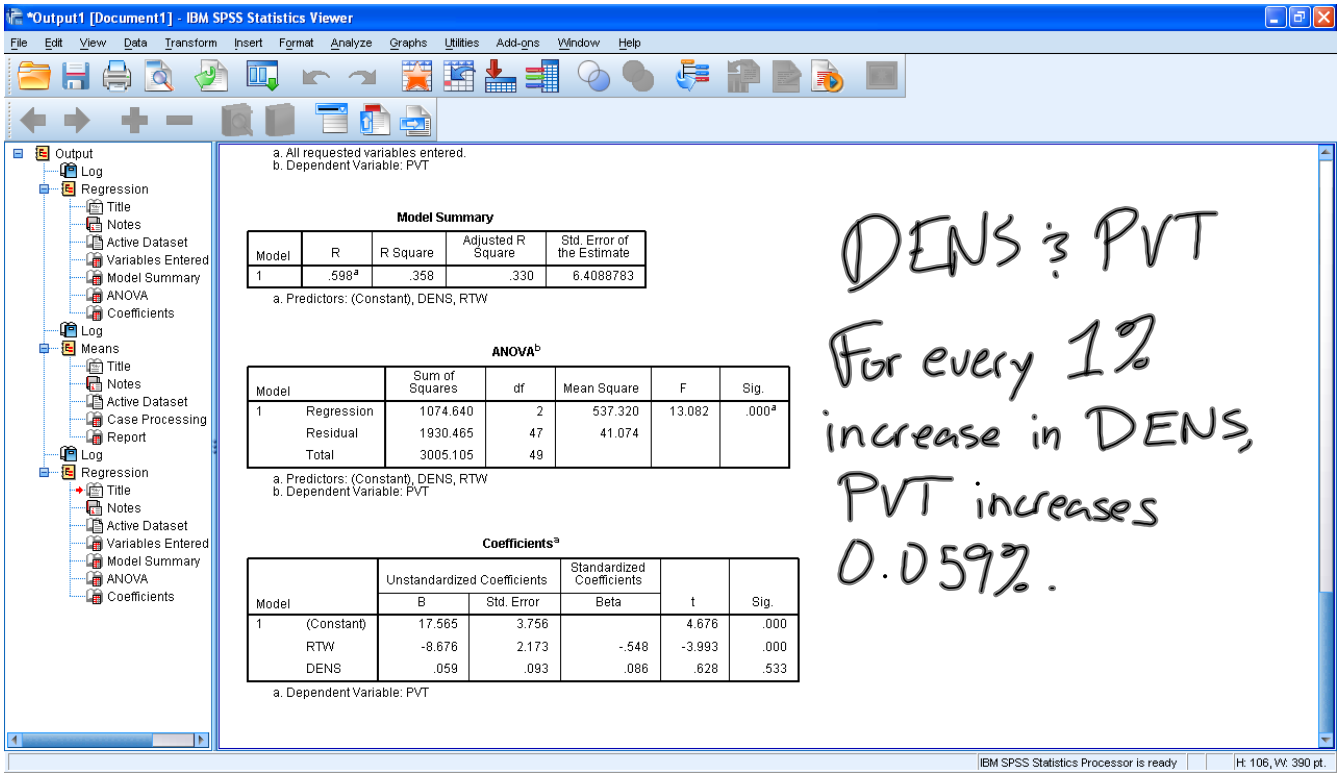

*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help
Output
Log
Regression
Title
Notes
Active Dataset
Variables Entered
Model Summary
ANOVA
Coefficients
Log
Means
Title
Notes
Active Dataset
Case Processing
Report
Model | R | R Square | Adjusted R Square | Std. Error of the Estimate
1 | .593[a] | .352 | .339 | 6.3683203
a. Predictors: (Constant), RTW
ANOVA[b]
Model | | Sum of Squares | df | Mean Square | F | Sig.
1 | Regression | 1058.441 | 1 | 1058.441 | 26.099 | .000[a]
Residual | 1946.664 | 48 | 40.556
Total | 3005.105 | 49
a. Predictors: (Constant), RTW
b. Dependent Variable: PVT
Coefficients[a]
Model | | Unstandardized Coefficients B | Unstandardized Coefficients Std. Error | Standardized Coefficients Beta | t | Sig.
1 | (Constant) | 19.807 | 1.163 | | 17.035 | .000
RTW | -9.392 | 1.838 | -.593 | -5.109 | .000
a. Dependent Variable: PVT
the probability it will be zero.
test statistic
MEANS TABLES=PVT BY RTW
/CELLS MEAN COUNT STDDEV.
Means
Double click to edit Pivot Table
IBM SPSS Statistics Processor is ready
H: 150, W: 536 pt.

Multiple Regression

- Linear regression w/ several independent variables.
  - Correlation only reveals a relationship between two variables.
    - PVT may be influenced by RTW, and by DENS (the % of public workers unionized)
    - Isolates the correlation between multiple variables and the dependent variable.

$$PVT = \alpha + \underbrace{\beta_1}_{\text{Slope for RTW \& PVT}} RTW + \underbrace{\beta_2}_{\text{Slope for DENS \& PVT.}} DENS$$


*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help
Output
Log
Regression
Title
Notes
Active Dataset
Variables Entered
Model Summary
ANOVA
Coefficients
Log
Means
Title
Notes
Active Dataset
Case Processing
Report
Log
Regression
Title
Notes
Active Dataset
Variables Entered
Model Summary
ANOVA
Coefficients
a. All requested variables entered.
b. Dependent Variable: PVT
Model Summary
Model | R | R Square | Adjusted R Square | Std. Error of the Estimate
1 | .598[a] | .358 | .330 | 6.4088783
a. Predictors: (Constant), DENS, RTW
RTW & DENS has an R=.598
ANOVA[b]
Model | Sum of Squares | df | Mean Square | F | Sig.
1 Regression | 1074.640 | 2 | 537.320 | 13.082 | .000[a]
Residual | 1930.465 | 47 | 41.074
Total | 3005.105 | 49
a. Predictors: (Constant), DENS, RTW
b. Dependent Variable: PVT
Coefficients[a]
Model | Unstandardized Coefficients B | Unstandardized Coefficients Std. Error | Standardized Coefficients Beta | t | Sig.
1 (Constant) | 17.565 | 3.756 | | 4.676 | .000
RTW | -8.676 | 2.173 | -.548 | -3.993 | .000
DENS | .059 | .093 | .086 | .628 | .533
a. Dependent Variable: PVT
<0.05
>0.05 not statistically sig.
IBM SPSS Statistics Processor is ready
H: 106, W: 390 pt.


*Output1 [Document1] - IBM SPSS Statistics Viewer
File Edit View Data Transform Insert Format Analyze Graphs Utilities Add-ons Window Help
Output
Log
Regression
Title
Notes
Active Dataset
Variables Entered
Model Summary
ANOVA
Coefficients
Log
Means
Title
Notes
Active Dataset
Case Processing
Report
Log
Regression
Title
Notes
Active Dataset
Variables Entered
Model Summary
ANOVA
Coefficients
a. All requested variables entered.
b. Dependent Variable: PVT
Model Summary
Model | R | R Square | Adjusted R Square | Std. Error of the Estimate
1 | .598[a] | .358 | .330 | 6.4088783
a. Predictors: (Constant), DENS, RTW
ANOVA[b]
Model | | Sum of Squares | df | Mean Square | F | Sig.
1 | Regression | 1074.640 | 2 | 537.320 | 13.082 | .000[a]
Residual | 1930.465 | 47 | 41.074
Total | 3005.105 | 49
a. Predictors: (Constant), DENS, RTW
b. Dependent Variable: PVT
Coefficients[a]
Model | | Unstandardized Coefficients B | Unstandardized Coefficients Std. Error | Standardized Coefficients Beta | t | Sig.
1 | (Constant) | 17.565 | 3.756 | | 4.676 | .000
RTW | -8.676 | 2.173 | -.548 | -3.993 | .000
DENS | .059 | .093 | .086 | .628 | .533
a. Dependent Variable: PVT
DENS & PVT
For every 1% increase in DENS, PVT increases 0.059%.
IBM SPSS Statistics Processor is ready
H: 106, W: 390 pt.