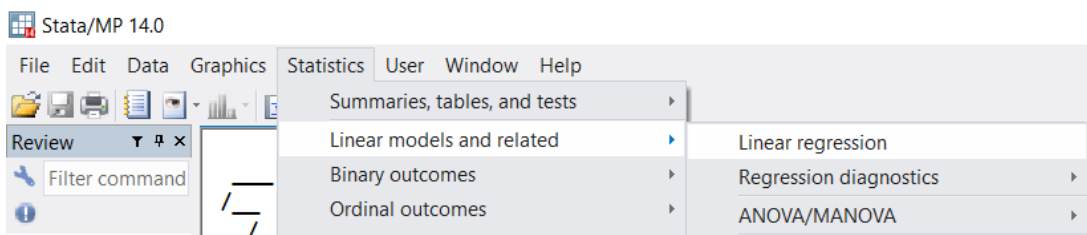


Introduction to the Regression Model

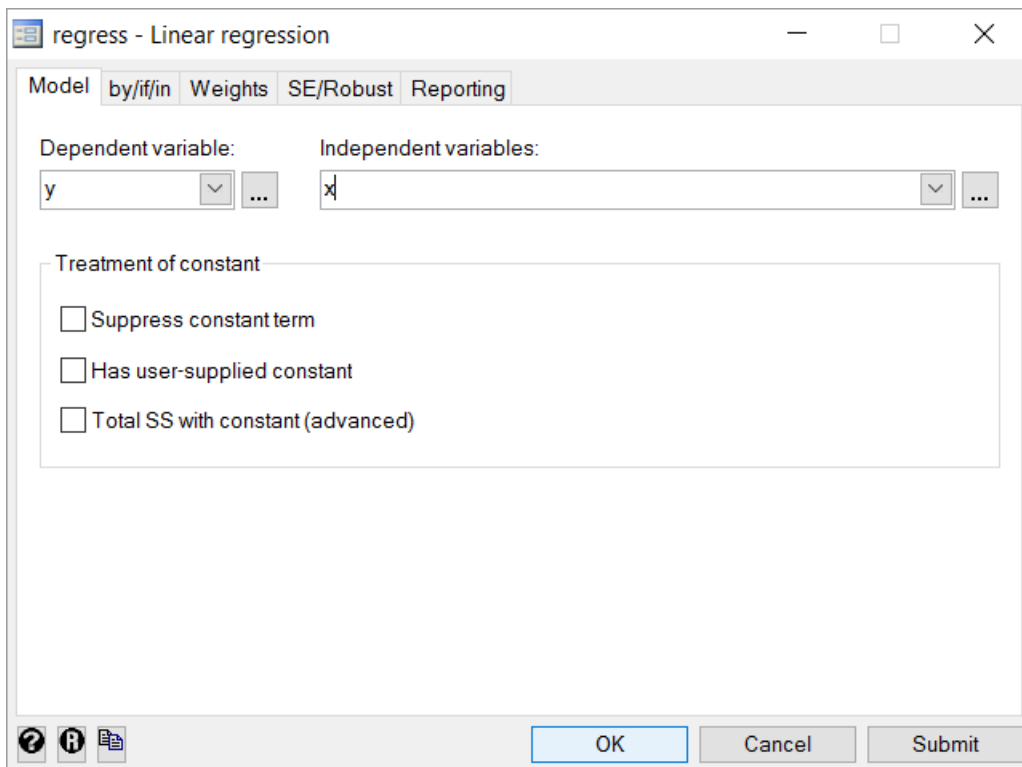
Simple Regression Model

CAPM model: $Y_i = \beta_0 + \beta_1 X_{1i} + u_i$

In this example, we estimate the least-squares regression of Y_i on X_i . To run a regression, from menu bar go to Statistics, choose Linear models and related, select Linear regression.



From regress – Linear regression window, specify dependent and independent variables in Dependent variable: and Independent variables: boxes, then, click OK.



The results will be as follows:

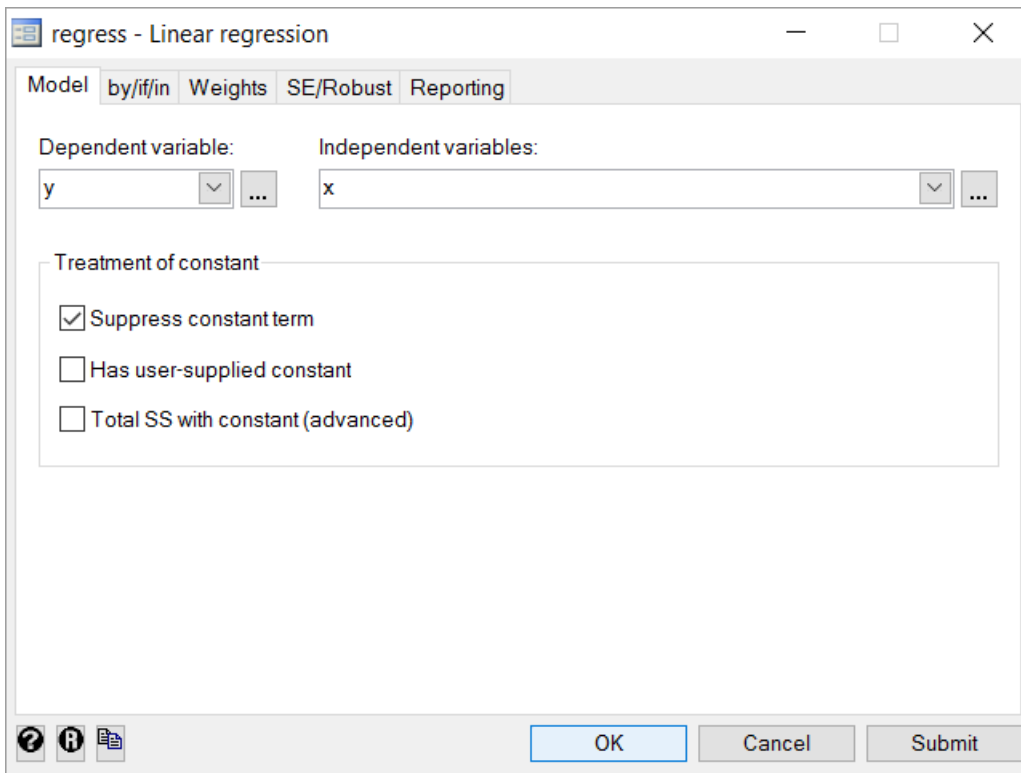
```
. reg y x1
```

Source	SS	df	MS	Number of obs	=	120
Model	247.518473	1	247.518473	F(1, 118)	=	64.86
Residual	450.31445	118	3.81622416	Prob > F	=	0.0000
				R-squared	=	0.3547
				Adj R-squared	=	0.3492
Total	697.832923	119	5.86414221	Root MSE	=	1.9535

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	.9605023	.1192646	8.05	0.000	.7243259 1.196679
_cons	.380914	.1783442	2.14	0.035	.0277441 .734084

```
. est store withcon
```

To run the model without intercept term, check Suppress constant term box.



```
. reg y x1, nocon
```

Source	SS	df	MS	Number of obs	=	120
Model	249.166211	1	249.166211	F(1, 119)	=	63.39
Residual	467.723299	119	3.93044789	Prob > F	=	0.0000
				R-squared	=	0.3476
				Adj R-squared	=	0.3421
Total	716.88951	120	5.97407925	Root MSE	=	1.9825

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	.9636218	.1210272	7.96	0.000	.7239758 1.203268

```
. est store wocon
```

```
. est table withcon wocon, star(0.1 0.05 0.01) stat(N rss F r2 r2_a)
```

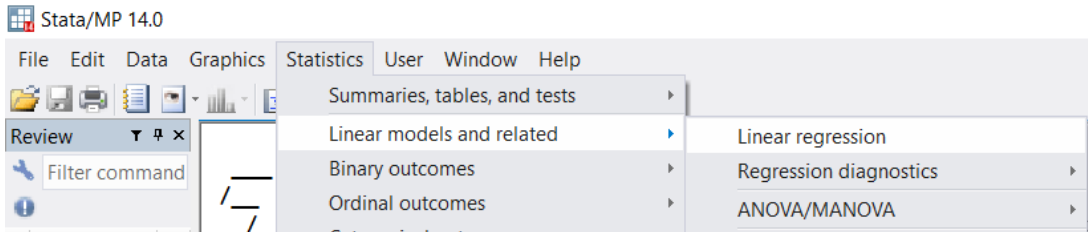
variable	withcon	wocon
x1	.96050229***	.96362177***
_cons	.38091403**	
N	120	120
rss	450.31445	467.7233
F	64.859521	63.393847
r2	.35469589	.34756571
r2_a	.34922722	.34208307

Legend: * p<.1; ** p<.05; *** p<.01

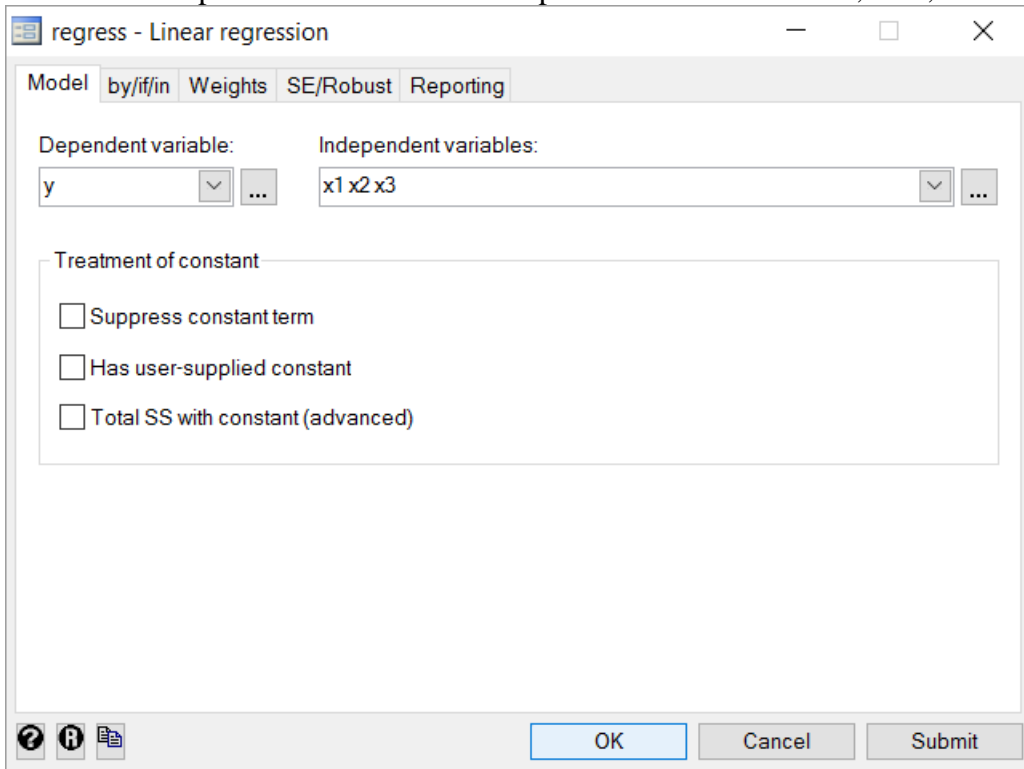
Multiple Regression Model

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

To run a regression, from menu bar go to Statistics, choose Linear models and related, select Linear regression.



From regress – Linear regression window, specify dependent and independent variables in Dependent variable: and Independent variables: boxes, then, click OK.



The results will be as follows:

```
. reg y x1 x2 x3
```

Source	SS	df	MS	Number of obs	=	120
Model	447.276407	3	149.092136	F(3, 116)	=	69.03
Residual	250.556516	116	2.15996997	Prob > F	=	0.0000
				R-squared	=	0.6410
				Adj R-squared	=	0.6317
Total	697.832923	119	5.86414221	Root MSE	=	1.4697

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
x1	.8614489	.0906645	9.50	0.000	.6818765 1.041021
x2	.4118276	.0581692	7.08	0.000	.2966163 .527039
x3	.2799795	.0419203	6.68	0.000	.196951 .363008
_cons	.1439872	.1368248	1.05	0.295	-.1270116 .4149859

Hypotheses Testing

To test whether CAPM or FF model is more appropriated or should we eliminate x_2 and x_3 out of the model, or $H_0: \beta_2 = \beta_3 = 0$.

```
. test x2 x3
```

- (1) x2 = 0
- (2) x3 = 0

```

F( 2, 116) = 46.24
Prob > F = 0.0000

```

Dummy Variable Regression Models

ANOVA Model

In the study of seasonal effect (Monday and Friday effects), ANOVA model is estimated using both models with and without intercept term:

Model with Intercept term:
$$Y_t = \beta_1 + \beta_2 D_{2t} + \beta_3 D_{3t} + \beta_4 D_{4t} + \beta_5 D_{5t} + u_t$$

Model without Intercept term:
$$Y_t = \beta_1 D_{1t} + \beta_2 D_{2t} + \beta_3 D_{3t} + \beta_4 D_{4t} + \beta_5 D_{5t} + u_t$$

where: Y_t = Return on asset at time t .

D_{1t} = 1 on Monday and = 0 otherwise.

D_{2t} = 1 on Tuesday and = 0 otherwise.

D_{3t} = 1 on Wednesday and = 0 otherwise.

D_{4t} = 1 on Thursday and = 0 otherwise.

D_{5t} = 1 on Friday and = 0 otherwise.

. reg y d2 d3 d4 d5

Source	SS	df	MS			
Model	.00681682	4	.001704205	Number of obs =	995	
Residual	.221894717	990	.000224136	F(4, 990) =	7.60	
Total	.228711537	994	.000230092	Prob > F =	0.0000	
				R-squared =	0.0298	
				Adj R-squared =	0.0259	
				Root MSE =	.01497	

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d2	.0044637	.0015009	2.97	0.003	.0015184	.0074089
d3	.0044351	.0015009	2.96	0.003	.0014898	.0073803
d4	.0040519	.0015009	2.70	0.007	.0011067	.0069972
d5	.0082592	.0015009	5.50	0.000	.005314	.0112045
_cons	-.0040211	.0010613	-3.79	0.000	-.0061037	-.0019385

. reg y d1 d2 d3 d4 d5, noconstant

Source	SS	df	MS			
Model	.006865358	5	.001373072	Number of obs =	995	
Residual	.221894717	990	.000224136	F(5, 990) =	6.13	
Total	.228760075	995	.00022991	Prob > F =	0.0000	
				R-squared =	0.0300	
				Adj R-squared =	0.0251	
				Root MSE =	.01497	

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d1	-.0040211	.0010613	-3.79	0.000	-.0061037	-.0019385
d2	.0004426	.0010613	0.42	0.677	-.0016401	.0025252
d3	.000414	.0010613	0.39	0.697	-.0016687	.0024966
d4	.0000308	.0010613	0.03	0.977	-.0020518	.0021134
d5	.0042381	.0010613	3.99	0.000	.0021555	.0063207

Interpretations of the estimated results of the two models are as follows:

Model	With intercept term	Without intercept
Monday	$\hat{Y}_t = \hat{\beta}_1 = -0.004021$	$\hat{Y}_t = \hat{\beta}_1 = -0.004021$
Tuesday	$\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 = -0.004021 + 0.004464 = 0.000443$	$\hat{Y}_t = \hat{\beta}_2 = 0.000443$
Wednesday	$\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_3 = -0.004021 + 0.004435 = 0.000414$	$\hat{Y}_t = \hat{\beta}_3 = 0.000414$
Thursday	$\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_4 = -0.004021 + 0.004052 = 0.000031$	$\hat{Y}_t = \hat{\beta}_4 = 0.000031$
Friday	$\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_5 = -0.004021 + 0.008259 = 0.004238$	$\hat{Y}_t = \hat{\beta}_5 = 0.004238$

For the purpose of testing only Monday and Friday effect, the model can be stated as:

Model to test Monday and Friday effect: $Y_t = \beta_0 + \beta_1 D_{1t} + \beta_5 D_{5t} + u_t$

```
. reg y d1 d5
```

Source	SS	df	MS			
Model	.006795783	2	.003397892	Number of obs =	995	
Residual	.221915754	992	.000223705	F(2, 992) =	15.19	
Total	.228711537	994	.000230092	Prob > F =	0.0000	
				R-squared =	0.0297	
				Adj R-squared =	0.0278	
				Root MSE =	.01496	

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d1	-.0043169	.0012243	-3.53	0.000	-.0067194	-.0019144
d5	.0039423	.0012243	3.22	0.001	.0015399	.0063448
_cons	.0002958	.0006121	0.48	0.629	-.0009055	.001497

Interpretation of the estimated result of the model is as follow:

Model	
Tuesday Wednesday Thursday	$\hat{Y}_t = \hat{\beta}_0 = 0.000296$
Monday	$\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 = 0.000296 - 0.004317 = -0.004021$
Friday	$\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_5 = 0.000296 - 0.003942 = 0.004238$

ANCOVA Model

If the model also includes quantitative independent variable (X_t), the model can be stated as:

Model with intercept dummy:
$$Y_t = \beta_0 + \beta_1 X_{1t} + \gamma_0 D_{1t} + \lambda_0 D_{5t} + u_t$$

This model can be interpreted as:

Model for Tuesday, Wednesday and Thursday:
$$Y_t = \beta_0 + \beta_1 X_{1t} + u_t$$

Model for Monday:
$$Y_t = (\beta_0 + \gamma_0) + \beta_1 X_{1t} + u_t$$

Model for Friday:
$$Y_t = (\beta_0 + \lambda_0) + \beta_1 X_{1t} + u_t$$

```
. reg y x1 d1 d5
```

Source	SS	df	MS			
Model	.007808227	3	.002602742	Number of obs =	995	
Residual	.220903311	991	.000222909	F(3, 991) =	11.68	
				Prob > F =	0.0000	
				R-squared =	0.0341	
				Adj R-squared =	0.0312	
Total	.228711537	994	.000230092	Root MSE =	.01493	

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
x1	1.90e-06	8.91e-07	2.13	0.033	1.50e-07	3.65e-06
d1	-.0043131	.0012221	-3.53	0.000	-.0067113	-.0019149
d5	.0039385	.0012221	3.22	0.001	.0015403	.0063368
_cons	-.0708331	.0333809	-2.12	0.034	-.1363383	-.0053278

Model with intercept and slope dummy variables:

$$Y_t = \beta_0 + \gamma_0 D_{1t} + \lambda_0 D_{5t} + \beta_1 X_{1t} + \gamma_1 D_{1t} X_{1t} + \lambda_1 D_{5t} X_{1t} + u_t$$

This model can be interpreted as:

Model for Tuesday, Wednesday and Thursday:
$$Y_t = \beta_0 + \beta_1 X_{1t} + u_t$$

Model for Monday:
$$Y_t = (\beta_0 + \gamma_0) + (\beta_1 + \gamma_1) X_{1t} + u_t$$

Model for Friday:
$$Y_t = (\beta_0 + \lambda_0) + (\beta_1 + \lambda_1) X_{1t} + u_t$$

```
. g x1d1 = x1*d1
```

```
. g x1d5 = x1*d5
```

```
. reg y d1 d5 x1 x1d1 x1d5
```

Source	SS	df	MS			
Model	.008022894	5	.001604579	Number of obs =	995	
Residual	.220688644	989	.000223143	F(5, 989) =	7.19	
				Prob > F =	0.0000	
				R-squared =	0.0351	
				Adj R-squared =	0.0302	
Total	.228711537	994	.000230092	Root MSE =	.01494	

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d1	-.0541205	.086225	-0.63	0.530	-.2233254	.1150844
d5	.0576593	.0862319	0.67	0.504	-.1115592	.2268777
x1	1.92e-06	1.15e-06	1.67	0.096	-3.39e-07	4.18e-06
x1d1	1.33e-06	2.30e-06	0.58	0.564	-3.19e-06	5.85e-06
x1d5	-1.43e-06	2.30e-06	-0.62	0.533	-5.95e-06	3.08e-06
_cons	-.0716146	.0431142	-1.66	0.097	-.1562204	.0129911

Chow Test vs Dummy Variable Technique

If we would like to check whether the model should be separated into two sub-periods (2000-2001 and 2002-2004), the models can be stated as:

Model 2000-2001: $Y_t = \beta_0 + \beta_1 X_{1t} + u_{1t}$

Model 2002-2004: $Y_t = \gamma_0 + \gamma_1 X_{1t} + u_{2t}$

Chow Test: $H_0: \beta_0 = \gamma_0 \text{ and } \beta_1 = \gamma_1$

Dummy Variable Regression Model

Model with intercept and slope dummy:

$$Y_t = \beta_0 + \delta_0 D_{YEARt} + \beta_1 X_{1t} + \delta_1 D_{YEARt} X_{1t} + u_t$$

where: $D_{YEARt} = 1$ for 2002-2004 and $= 0$ for 2000-2001.

The model can be interpreted as:

Model 2000-2001: $Y_t = \beta_0 + \beta_1 X_{1t} + u_{1t}$

Model 2002-2004: $Y_t = (\beta_0 + \delta_0) + (\beta_1 + \delta_1) X_{1t} + u_{2t}$

$H_{01}: \delta_0 = 0 \text{ and } \delta_1 = 0$

```
. g x1dyear = x1*dyear
```

```
. reg y dyear x1 x1dyear
```

Source	SS	df	MS	Number of obs =	995
Model	.001849335	3	.000616445	F(3, 991) =	2.69
Residual	.226862202	991	.000228923	Prob > F =	0.0450
				R-squared =	0.0081
				Adj R-squared =	0.0051
Total	.228711537	994	.000230092	Root MSE =	.01513

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
dyear	.2727136	.1521381	1.79	0.073	-.0258363 .5712635
x1	8.49e-06	3.60e-06	2.36	0.019	1.42e-06 .0000156
x1dyear	-7.38e-06	4.10e-06	-1.80	0.072	-.0000154 6.66e-07
_cons	-.3139704	.1329766	-2.36	0.018	-.5749185 -.0530223

```
. test dyear x1dyear
```

```
( 1) dyear = 0
( 2) x1dyear = 0
```

```
F( 2, 991) = 1.80
Prob > F = 0.1657
```

Distinction between the two methods is that while Chow test is to test whether all estimators are the same or not, dummy variables technique is to separately test equality of each estimator for different sub-period -- some estimators may be different while some may not.

Interaction Effect

Model with interaction effect dummy:

$$Y_t = \beta_0 + \gamma_0 D_{1t} + \lambda_0 D_{JANt} + \delta_0 D_{1t} D_{JANt} + \beta_1 X_{1t} + u_t$$

where: $D_{1t} = 1$ for Monday and $= 0$ otherwise.

$D_{JANt} = 1$ for January and $= 0$ otherwise.

The meaning of $D_{1t} D_{JANt}$

$D_{1t} D_{JANt} = 1$ for Monday in January and $= 0$ otherwise.

```
. g d1djan = d1*djan
```

```
. reg y d1 djan d1djan x1
```

Source	SS	df	MS			
Model	.006773371	4	.001693343	Number of obs =	995	
Residual	.221938166	990	.00022418	F(4, 990) =	7.55	
				Prob > F =	0.0000	
				R-squared =	0.0296	
				Adj R-squared =	0.0257	
Total	.228711537	994	.000230092	Root MSE =	.01497	

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d1	-.0059289	.0012667	-4.68	0.000	-.0084147	-.0034432
djan	.0019612	.0017065	1.15	0.251	-.0013876	.0053099
d1djan	.004773	.0036262	1.32	0.188	-.0023429	.0118889
x1	2.08e-06	8.98e-07	2.32	0.021	3.21e-07	3.85e-06
_cons	-.0769402	.033645	-2.29	0.022	-.1429639	-.0109166

Proof – Oneway ANOVA & ANOVA Model

. oneway y day, tab

day	Summary of Y		Freq.
	Mean	Std. Dev.	
1	-.00402112	.01629203	199
2	.00044255	.01373011	199
3	.00041396	.01598835	199
4	.00003082	.0147649	199
5	.00423812	.01389622	199
Total	.00022087	.01516879	995

Analysis of Variance					
Source	SS	df	MS	F	Prob > F
Between groups	.00681682	4	.001704205	7.60	0.0000
Within groups	.221894717	990	.000224136		
Total	.228711537	994	.000230092		

Bartlett's test for equal variances: $\chi^2(4) = 9.7365$ Prob> $\chi^2 = 0.045$

. reg y d2 d3 d4 d5

Source	SS	df	MS	Number of obs	=	995
Model	.00681682	4	.001704205	F(4, 990)	=	7.60
Residual	.221894717	990	.000224136	Prob > F	=	0.0000
Total	.228711537	994	.000230092	R-squared	=	0.0298
				Adj R-squared	=	0.0259
				Root MSE	=	.01497

y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
d2	.0044637	.0015009	2.97	0.003	.0015184	.0074089
d3	.0044351	.0015009	2.96	0.003	.0014898	.0073803
d4	.0040519	.0015009	2.70	0.007	.0011067	.0069972
d5	.0082592	.0015009	5.50	0.000	.005314	.0112045
_cons	-.0040211	.0010613	-3.79	0.000	-.0061037	-.0019385