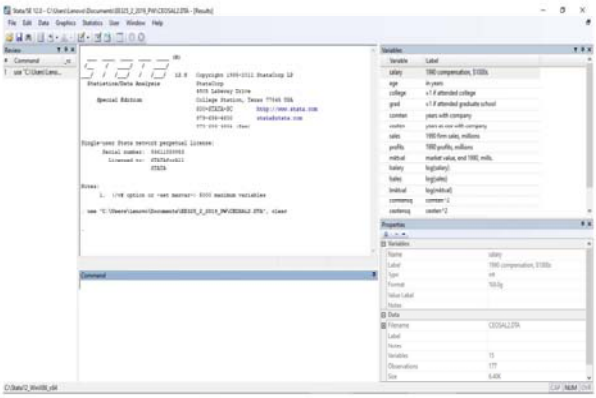
An Introduction on Basic Commands

This version: as of 31 Mar 2020

Instruction

- Log on to VDI by your username and password.
- Start the statistical software named 'STATA/MP16'
- Click file-> open-> CEOSAL2.DTA (located in your folder named 'Documents', perhaps)



CEOSAL2.DES

```
salary    age       college   grad      comten    ceoten    sales     profits
mktval    lsalary   lsales    lmktval   comtensq  ceotensq  profmarg

  Obs:   177

  1. salary                  1990 compensation, $1000s
  2. age                     in years
  3. college                 =1 if attended college
  4. grad                    =1 if attended graduate school
  5. comten                  years with company
  6. ceoten                  years as ceo with company
  7. sales                   1990 firm sales, millions
  8. profits                 1990 profits, millions
  9. mktval                  market value, end 1990, mills.
 10. lsalary                 log(salary)
 11. lsales                  log(sales)
 12. lmktval                 log(mktval)
 13. comtensq                comten^2
 14. ceotensq                ceoten^2
 15. profmarg                profits as % of sales
```

Variables

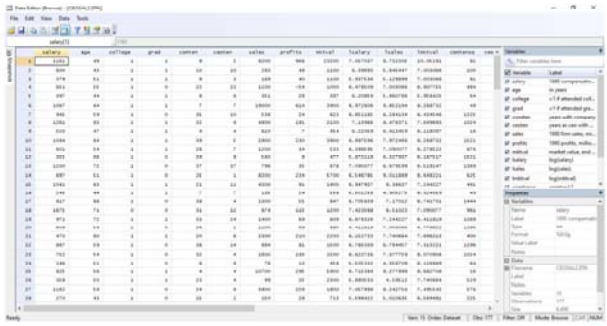
Variable	Label
salary	1990 compensation, $1000s
age	in years
college	=1 if attended college
grad	=1 if attended graduate school
comten	years with company
ceoten	years as ceo with company
sales	1990 firm sales, millions
profits	1990 profits, millions
mktval	market value, end 1990, mills.
lsalary	log(salary)
lsales	log(sales)
lmktval	log(mktval)
comtensq	comten^2
ceotensq	ceoten^2

describe

```
. des

Contains data from C:\Users\Lenovo\Documents\EE325_2_2019_PW\CEOSAL2.DTA
  obs:           177
 vars:            15                          17 Aug 1999 23:14
 size:         6,549
-------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-------------------------------------------------------------------------------
salary          int    %9.0g                  1990 compensation, $1000s
age             byte   %9.0g                  in years
college         byte   %9.0g                  =1 if attended college
grad            byte   %9.0g                  =1 if attended graduate school
comten          byte   %9.0g                  years with company
ceoten          byte   %9.0g                  years as ceo with company
sales           float  %9.0g                  1990 firm sales, millions
profits         int    %9.0g                  1990 profits, millions
mktval          float  %9.0g                  market value, end 1990, mills.
lsalary         float  %9.0g                  log(salary)
lsales          float  %9.0g                  log(sales)
lmktval         float  %9.0g                  log(mktval)
comtensq        int    %9.0g                  comten^2
ceotensq        int    %9.0g                  ceoten^2
profmarg        float  %9.0g                  profits as % of sales
-------------------------------------------------------------------------------
Sorted by:
```

browse



codebook

```
. codebook

-------------------------------------------------------------------------------
salary                                                1990 compensation, $1000s
-------------------------------------------------------------------------------

                  type:  numeric (int)

                 range:  [100,5299]                   units:  1
         unique values:  166                      missing .:  0/177

                  mean:   865.864
             std. dev:   587.589

           percentiles:        10%       25%       50%       75%       90%
                               358       471       707      1119      1550

-------------------------------------------------------------------------------
age                                                                   in years
-------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [33,86]                      units:  1
         unique values:  39                       missing .:  0/177

                  mean:   56.4294
             std. dev:   8.42189

--more--
```

describe salary

```
. describe salary

              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
salary          int    %9.0g                  1990 compensation, $1000s
```

describe salary age

```
. describe salary age

              storage  display     value
variable name   type   format      label      variable label
------------------------------------------------------------------------------
salary          int    %9.0g                  1990 compensation, $1000s
age             byte   %9.0g                  in years
```

summarize

```
. sum
```

Variable	Obs	Mean	Std. Dev.	Min	Max
salary	177	865.8644	587.5893	100	5299
age	177	56.42938	8.42189	33	86
college	177	.9717514	.1661523	0	1
grad	177	.5310734	.5004492	0	1
comten	177	22.50282	12.29473	2	58
ceoten	177	7.954802	7.150826	0	37
sales	177	3529.463	6088.654	29	51300
profits	177	207.8305	404.4543	-463	2700
mktval	177	3600.316	6442.276	387	45400
lsalary	177	6.582848	.6060894	4.60517	8.576274
lsales	177	7.231025	1.432086	3.367296	10.84545
lmktval	177	7.39941	1.133414	5.958425	10.72327
comtensq	177	656.6836	577.1227	4	3364
ceotensq	177	114.1243	212.566	0	1369
profmarg	177	6.42011	17.86074	-203.0769	47.45763

```
.
```

summarize salary

```
. summarize salary
```

Variable	Obs	Mean	Std. Dev.	Min	Max
salary	177	865.8644	587.5893	100	5299

sum salary

```
. sum salary
```

Variable	Obs	Mean	Std. Dev.	Min	Max
salary	177	865.8644	587.5893	100	5299

summarize salary, detail

```
. sum salary, detail

                    1990 compensation, $1000s
-------------------------------------------------------------
      Percentiles      Smallest
 1%           129            100
 5%           270            129
10%           358            174       Obs                 177
25%           471            185       Sum of Wgt.         177

50%           707                      Mean           865.8644
                        Largest       Std. Dev.      587.5893
75%          1119           2220
90%          1550           2265       Variance       345261.2
95%          1798           2792       Skewness       2.998603
99%          2792           5299       Kurtosis       20.40129
```

sum if salary >866

```
. sum if salary>866
```

Variable	Obs	Mean	Std. Dev.	Min	Max
salary	70	1385.514	615.4349	867	5299
age	70	57.8	8.107252	33	73
college	70	.9571429	.2039973	0	1
grad	70	.5428571	.5017567	0	1
comten	70	22.41429	11.6923	2	42
ceoten	70	8.714286	6.445182	1	37
sales	70	6373.857	8551.766	149	51300
profits	70	409.2714	567.3626	-271	2700
mktval	70	6730.786	9220.275	398	45400
lsalary	70	7.174854	.3149622	6.765039	8.575274
lsales	70	7.998154	1.354099	5.003946	10.84545
lmktval	70	8.076301	1.215686	5.986452	10.72327
comtensq	70	637.1571	512.7209	4	1764
ceotensq	70	116.8857	196.9	1	1369
profmarg	70	6.228583	9.041723	-48.13499	19.8895

sum salary if age<=40

```
. sum salary if age<=40
```

Variable	Obs	Mean	Std. Dev.	Min	Max
salary	8	1027	843.1478	310	2792

sum if grad==1

```
. sum if grad==1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
salary	94	864.2128	501.3924	100	2265
age	94	55.45745	8.1553	38	86
college	94	1	0	1	1
grad	94	1	0	1	1
comten	94	19.87234	11.48934	2	41
ceoten	94	7.265957	6.145504	0	26
sales	94	3964.915	5734.282	36	28000
profits	94	244.9043	414.3908	-438	1900
mktval	94	4342.66	6949.829	387	45400
lsalary	94	6.590308	.6125932	4.60517	7.72533
lsales	94	7.34306	1.527586	3.583519	10.23996
lmktval	94	7.558841	1.226916	5.958425	10.72327
comtensq	94	525.5106	470.1377	4	1681
ceotensq	94	90.15957	142.0795	0	676
profmarg	94	6.16246	23.31955	-203.0769	47.45763

sum if grad==0

```
. sum if grad==0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
salary	83	867.7349	675.2212	174	5299
age	83	57.53012	8.631108	33	80
college	83	.939759	.2393792	0	1
grad	83	0	0	0	0
comten	83	25.48193	12.56239	3	58
ceoten	83	8.73494	8.10914	0	37
sales	83	3036.301	6466.221	29	51300
profits	83	165.8434	391.1396	-463	2700
mktval	83	2759.59	5740.459	390	42900
lsalary	83	6.574399	.60218	5.159055	8.575274
lsales	83	7.104143	1.313251	3.367296	10.84545
lmktval	83	7.21885	.9939086	5.966147	10.66663
comtensq	83	805.241	649.7235	9	3364
ceotensq	83	141.2651	269.5968	0	1369
profmarg	83	6.711907	8.233072	-48.13499	23.94958

tabulate salary

```
. tabulate salary
```

1990 compensation, $1000s	Freq.	Percent	Cum.
100	1	0.56	0.56
129	1	0.56	1.13
174	1	0.56	1.69
199	1	0.56	2.26
218	1	0.56	2.82
245	1	0.56	3.39
246	1	0.56	3.95
264	1	0.56	4.52
270	1	0.56	5.08
300	1	0.56	5.65
308	1	0.56	6.21
310	1	0.56	6.78
323	1	0.56	7.34
345	1	0.56	7.91
348	1	0.56	8.47
350	1	0.56	9.04
355	1	0.56	9.60
358	2	1.13	10.73
363	1	0.56	11.30
369	1	0.56	11.86
375	1	0.56	12.43
377	1	0.56	12.99
379	1	0.56	13.56
--more--			

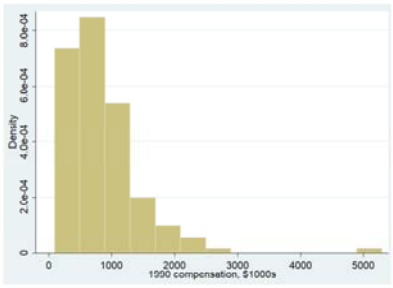
1301	1	0.56	85.31
1336	1	0.56	85.88
1340	1	0.56	86.44
1350	1	0.56	87.01
1373	1	0.56	87.57
1392	1	0.56	88.14
1444	1	0.56	88.70
1487	1	0.56	89.27
1500	1	0.56	89.83
1550	1	0.56	90.40
1600	1	0.56	90.96
1627	1	0.56	91.53
1630	1	0.56	92.09
1676	1	0.56	92.66
1700	1	0.56	93.22
1739	1	0.56	93.79
1749	1	0.56	94.35
1780	1	0.56	94.92
1795	1	0.56	95.48
1946	1	0.56	96.05
2021	1	0.56	96.61
2102	1	0.56	97.18
2199	1	0.56	97.74
2220	1	0.56	98.31
2265	1	0.56	98.87
2792	1	0.56	99.44
5299	1	0.56	100.00
Total	177	100.00	

tabulate college

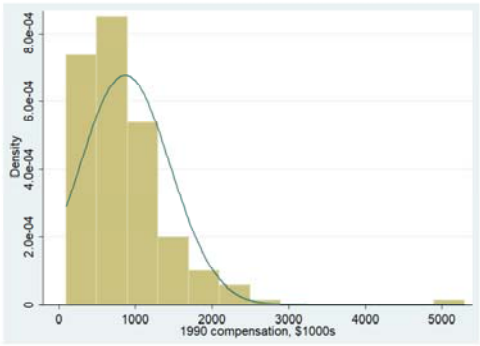
```
. tabulate college
```

=1 if attended college	Freq.	Percent	Cum.
0	5	2.82	2.82
1	172	97.18	100.00
Total	177	100.00	

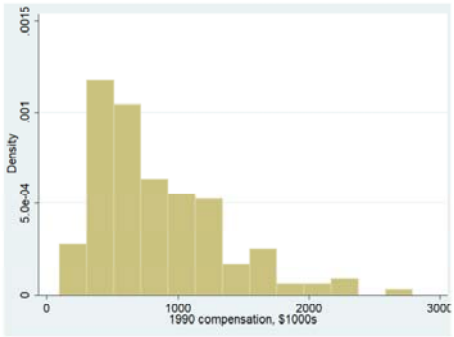
hist salary



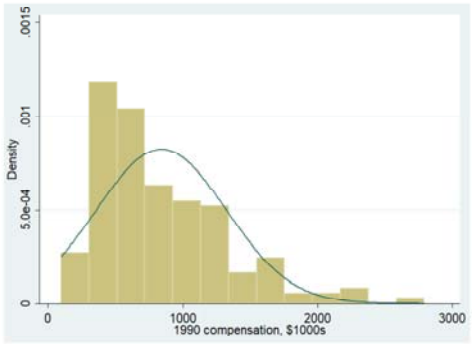
hist salary, normal



hist salary if salary<4000



hist salary if salary<4000, normal

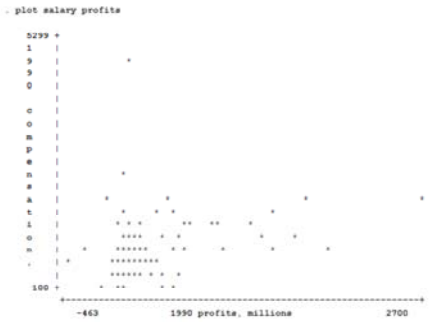


correlate salary profits sale

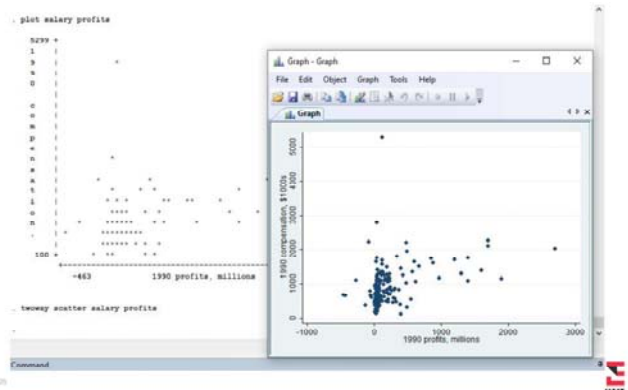
```
. correlate salary profits sales
(obs=177)

             |   salary  profits    sales
-------------+---------------------------
      salary |   1.0000
     profits |   0.3939   1.0000
       sales |   0.3802   0.7983   1.0000
```

plot salary profits



. twoway scatter salary profits



correlate salary profits sale, cov

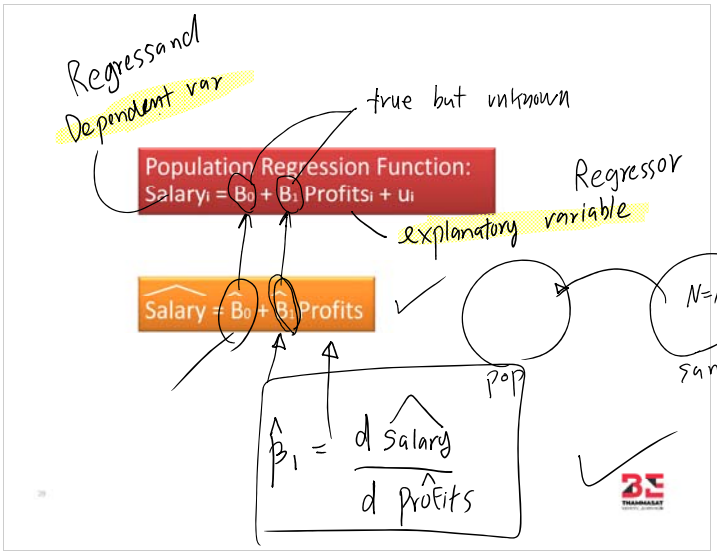
```
. correlate salary profits sales, cov
(obs=177)
```

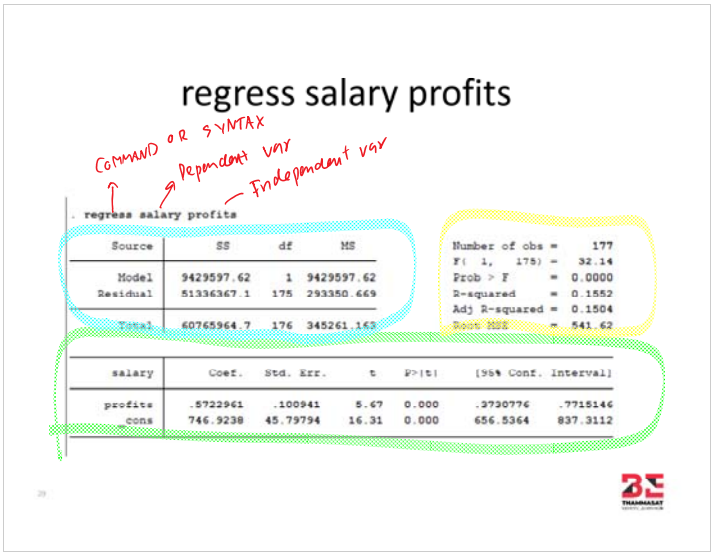
	salary	profits	sales
salary	345261		
profits	93618.1	163583	
sales	1.4e+06	2.0e+06	3.7e+07

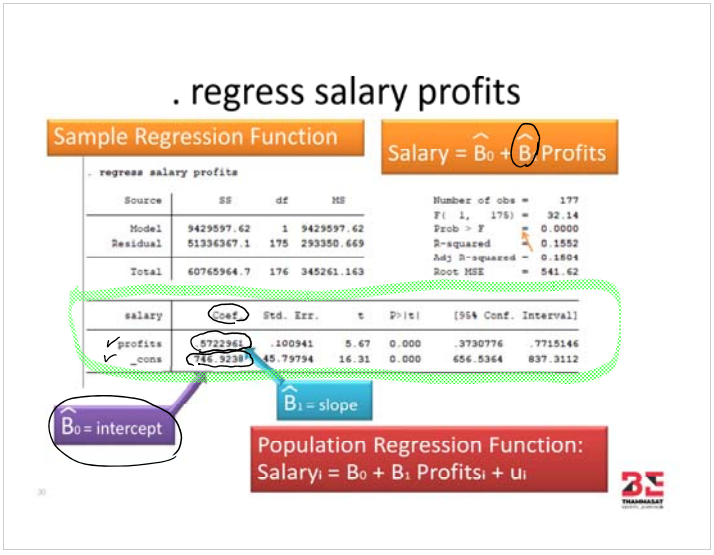
STATA

BACHELOR OF ECONOMICS
THAMMASAT UNIVERSITY

Simple Regression Analysis using STATA


Regressand
Dependent var
true but unknown
Population Regression Function:
Salaryi = B0 + B1 Profitsi + ui
Regressor
explanatory variable
Salary = B0 + B1 Profits
N=177
sample
pop
β1 = d Salary / d Profits


regress salary profits
COMMAND OR SYNTAX
Dependent var
Independent var
. regress salary profits
Source | SS df MS
Model 9429597.62 1 9429597.62
Residual 51336367.1 175 293350.669
Total 60765964.7 176 345261.163
Number of obs = 177
F(1, 175) = 32.14
Prob > F = 0.0000
R-squared = 0.1552
Adj R-squared = 0.1504
Root MSE = 541.62
salary | Coef. Std. Err. t P>|t| [95% Conf. Interval]
profits .5722961 .100941 5.67 0.000 .3730776 .7715146
_cons 746.9238 45.79794 16.31 0.000 656.5364 837.3112


. regress salary profits
Sample Regression Function
Salary = B0 + B1 Profits
. regress salary profits
Source | SS df MS
Model 9429597.62 1 9429597.62
Residual 51336367.1 175 293350.669
Total 60765964.7 176 345261.163
Number of obs = 177
F(1, 175) = 32.14
Prob > F = 0.0000
R-squared = 0.1552
Adj R-squared = 0.1504
Root MSE = 541.62
salary | Coef. Std. Err. t P>|t| [95% Conf. Interval]
profits .5722961 .100941 5.67 0.000 .3730776 .7715146
_cons 746.9238 45.79794 16.31 0.000 656.5364 837.3112
B1 = slope
B0 = intercept
Population Regression Function:
Salaryi = B0 + B1 Profitsi + ui

Salary = 746.972 + 0.572* Profits

$\hat{\beta}_0$ + $\hat{\beta}_1$ profits

$\hat{B}_0$ = intercept

$\hat{B}_1$ = slope

```
. regress salary profits

      Source |       SS       df       MS              Number of obs =     177
-------------+------------------------------           F(  1,   175) =   32.14
       Model |  9429597.62     1  9429597.62            Prob > F      =  0.0000
    Residual |  51336367.1   175  293350.669           R-squared     =  0.1552
-------------+------------------------------           Adj R-squared =  0.1504
       Total |  60765964.7   176  345261.163           Root MSE      =  541.62

------------------------------------------------------------------------------
      salary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     profits |   .5722961    .100941     5.67   0.000     .3730776    .7715146
       _cons |   746.9238   45.79794    16.31   0.000     656.5364    837.3112
------------------------------------------------------------------------------
```

Std. Err. = Standard Error Which is an estimate of the standard deviation of $\hat{B}_0$ and $\hat{B}_1$

t = t-statistics. Tests the null hypothesis that the coefficient = 0, i.e. $\hat{B}_0 = 0$, $\hat{B}_1 = 0$.

P > |t| Is the 2-tailed p-value for testing the null hypothesis that the coefficient = 0.

p-value

The chance that this interval can capture the true B_0 or B_1 is 95%

$\hat{t} = \frac{\hat{\beta}_i}{se(\hat{\beta}_i)}$

③ = ① / ②

w/ $\alpha = 0.05$, if our p-value < 0.05, then we reject H_0

H_0: $\beta_1 = 0$ (profits ↛ salary)

H_1: $\beta_1 \neq 0$ (profit → salary)

For $n > 30$, decision rule is...

If $\hat{t} > 1.96$, we can reject $H_0: \beta_i = 0$ w/ 95% confidence level.

SSE (Explained Sum of Squares)
SSR (Residual Sum of Squares)
SST(Total Sum of Squares)

TSS = ESS + RSS (dusamti) (Wooldridge)

STT = SSE + SSR

SUM OF SQUARES SS/df

$\sum \hat{u}_i^2$

ESS RSS TSS

ANOVA TABLE (Ch5)

k=1 (# of slope coefficient)

n

F = MS OF ESS / MS OF RSS

FOR JOINT TEST IN MULTIPLE REGRESSION MODEL

Ex: $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$

H_0: $\beta_1 = \beta_2 = 0$

H_1: β_1 and $\beta_2 \neq 0$ at the same time

one-tailed test

The chance that this interval can capture the true B_0 or B_1 is 95%

test profits = 0

H_0: $\beta_1 = 0$

H_1: $\beta_1 \neq 0$

```
One-sample t test

------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
profits  |     177    207.8305    30.40065    404.4543    147.8338    267.8272
------------------------------------------------------------------------------
    mean = mean(profits)                              t =   6.8364
Ho: mean = 0                      degrees of freedom =      176

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

H_0: $\beta_1 < 0$, H_1: $\beta_1 \geq 0$ — one-tailed test

H_0: $\beta_1 = 0$, H_1: $\beta_1 \neq 0$ — two-tailed test

H_0: $\beta_1 > 0$, H_1: $\beta_1 \leq 0$ — one-tailed test

R-squared, Adj. R-squared are assessments of the goodness of fit. 0 ≤ R-squared ≤ 1

Number of observations (N)

F-test, test for the null hypothesis that all the coefficients = 0

```
. regress salary profits

      Source |       SS       df       MS              Number of obs =     177
-------------+------------------------------           F(1, 175)     =    0.00
       Model |  9429597.62     1  9429597.62           Prob > F      =  0.0000
    Residual |  51336367.1   175  293350.669           R-squared     =  0.1553
-------------+------------------------------           Adj R-squared =  0.1505
       Total |  60765964.7   176  345261.163           Root MSE      =  541.62

------------------------------------------------------------------------------
      salary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     profits |   .5722961    .100941     5.67   0.000     .3730776    .7715146
       _cons |   746.9238   45.79794    16.31   0.000     656.5364    837.3112
------------------------------------------------------------------------------
```

predict y_hat, xb

$\hat{salary}$

	mktval	ltsalary	ltsales	lmktval	contemsq	costemsq	profmarg	y_hat
1	23200	7.067037	8.732305	10.05191	81	4	15.18068	1298.762
2	1100	6.39693	5.645447	7.003066	100	100	16.96113	774.394
3	1100	5.937536	5.129899	7.003066	81	9	23.66864	769.8157
4	1000	6.478509	7.003066	6.907755	484	484	-4.909091	716.0198
5	167	6.20859	5.860786	5.958425	64	36	7.977206	762.9481
6	3900	6.972606	8.852194	8.268732	49	49	3.231579	1098.314
7	623	6.851185	6.284134	6.434546	1225	100	4.477612	760.6589
8	2100	7.13966	8.476371	7.649693	1024	64	3.979167	856.2324
9	414	6.22059	6.413459	6.118097	16	16	1.147541	750.9239
10	3900	6.997596	7.972466	8.268732	1521	25	7.951035	878.5519
11	133	6.398595	7.090077	6.279522	676	49	2.833333	766.3819
12	477	6.872118	6.327937	6.167517	1521	64	1.428571	751.5022
13	858	7.090077	6.679599	6.519147	1369	1369	4.396985	766.9542
14	5700	8.646785	9.011889	8.648211	625	1	2.853658	880.8411
15	1400	6.947937	8.38687	7.244227	441	121	2.116279	789.0027
16	358	5.881259	4.905275	6.324359	49	49	17.77778	760.8588
17	847	6.706639	7.17012	6.741701	1444	16	4.230769	778.4001
18	1200	7.423099	6.51323	7.090077	961	144	17.09231	862.7979
19	609	6.875326	7.244227	6.411819	1089	576	4.928571	786.4122
20	880	6.411819	7.003066	6.779922	1296	1	6.272727	796.4122
21	5100	8.537714	7.740664	8.537714	400	36	9.130435	887.104
22	1500	6.765039	6.784457	7.313221	1296	196	9.162896	793.2798
23	3200	6.622736	7.377759	8.070906	1024	16	12.0625	817.577
24	458	6.508332	6.354709	6.126869	64	64	16.66667	754.3636
25	5900	8.715184	9.277999	8.682708	16	16	2.757009	915.7512
26	2300	5.880533	4.59512	7.740664	529	16	25.25253	761.2312
27	1800	7.057898	8.242756	7.495542	576	36	5.947368	876.2628
28	713	5.598422	5.010635	6.569481	225	4	18.66667	762.9481

$\hat{salary} = \hat{\beta}_0 + \hat{\beta}_1 profits$

$\hat{u}_i = salary_i - \hat{salary}_i$

$\sum \hat{u}_i^2 = RSS$

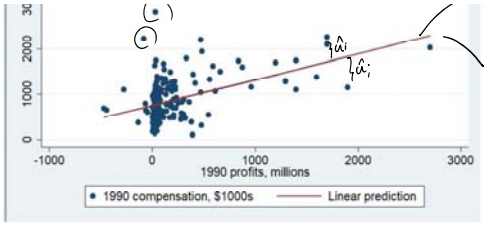
predict u_hat, residual

= salary − $\hat{salary}$

twoway scatter y_hat profits || line y_hat profits

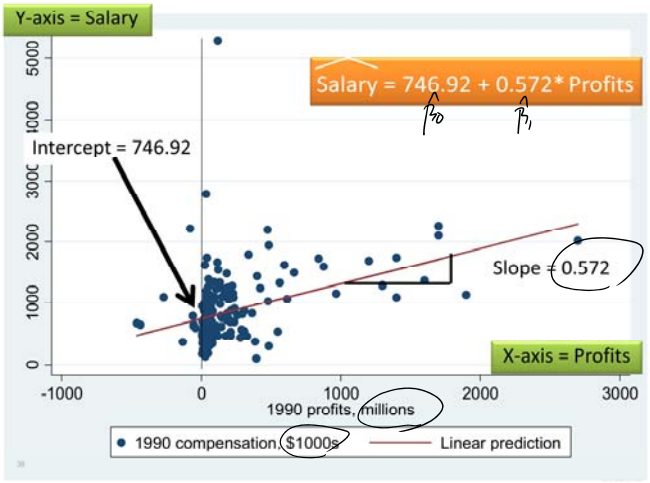
Fitted line

$\hat{salary}$ = 746.92 + 0.5722 profits



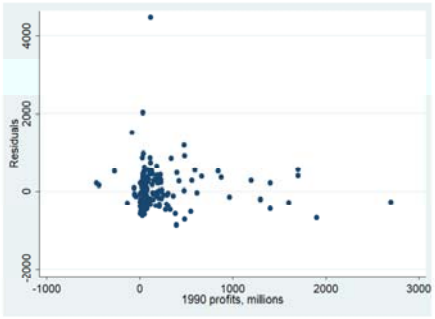
Salary = 746.92 + 0.572 Profits

$\hat{\beta}_0$ $\hat{\beta}_1$



if profits ↑ by 1 million dollar,
ceo's salary, on average, will
rise by $0.572 \times 1000 = 572$ dollars/YR,
holding all other factors constant.

. twoway scatter u_hat profits



correlate profits u_hat

X_i $\hat{u}_i$

```
. correlate profits u_hat
(obs=177)
```

	profits	u_hat
profits	1.0000	
u_hat	-0.0000	1.0000

$\hat{\beta}_1 = 0.572$

$cov(x_i, \hat{u}_i) = 0$

To check if the OLS estimation makes Xi uncorrelated with u_hat.

generate lsalary = log(salary)
generate lprofits = log(profits)
generate profitsq = profits^2

- Type: generate log_salary= log(salary)
- Type: gen log_profits=log(profits)
- Type: gen profits_sq= profits^2
- Type: regress salary profits profits_2

regress salary profits Vs. regress salary profits sales

or Two-variable regression

→ simple regression

```
. regress salary profits

      Source |       SS       df       MS              Number of obs =     177
-------------+------------------------------           F(  1,   175) =   32.14
       Model |  9429597.62     1  9429597.62            Prob > F      =  0.0000
    Residual |  51336367.1   175  293350.669           R-squared     =  0.1552
-------------+------------------------------           Adj R-squared =  0.1504
       Total |  60765964.7   176  345261.163           Root MSE      =  541.62

------------------------------------------------------------------------------
      salary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     profits |   .5722961    .100941     5.67   0.000     .3730776    .7715146
       _cons |   746.9238   45.79794    16.31   0.000     656.5364    837.3112
------------------------------------------------------------------------------
```

→ multiple regression

```
. regress salary profits salav

      Source |       SS       df       MS              Number of obs =     177
-------------+------------------------------           F(  2,   174) =   17.45
       Model |  10153943.8     2  5076971.43            Prob > F      =  0.0000
    Residual |  50611021.3   174  290873.689           R-squared     =  0.1671
-------------+------------------------------           Adj R-squared =  0.1575
       Total |  60765964.7   176  345261.163           Root MSE      =  539.33

------------------------------------------------------------------------------
      salary |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     profits |   .3620587     .16689     2.17   0.031     .0326694    .691448
      salav |   .0175915   .0119861     1.47   0.144    -.0040841    .0392671
       _cons |   728.8717   47.01705    15.50   0.000     636.0746    821.6689
------------------------------------------------------------------------------
```

P-value > 0.05

$H_0: \beta_2 = 0$

$t < 1.96$ ∴ → cannot reject H_0: $\beta_2 = 0$ hypothesized value

help