PUB – POS 316
Week 14b

Multiple regression
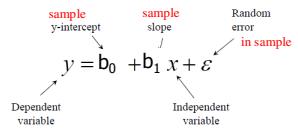
Navid Ghaffarzadegan
navidg@gmail.com
Last updated – Jan 1, 10

Introduction

- Simple regression:

$$y = b_0 + b_1\, x + \varepsilon$$

sample y-intercept; sample slope; Random error in sample

Dependent variable; Independent variable

- Regression line :
  SATM=403+22.72*HighSchoolMath

Introduction

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.382981
R Square	0.146675
Adjusted R	0.128519
Standard E	92.39064
Observatio	49

ANOVA

	df	SS	MS	F	gnificance F
Regression	1	68959.52	68959.52	8.07864	0.006606
Residual	47	401193.5	8536.031		
Total	48	470153			

	Coefficients	andard Err	t Stat	P-value	Lower 95%	Upper 95%
Intercept	403.2045	67.38424	5.983661	2.85E-07	267.6449	538.7641
X Variable	22.72341	7.994741	2.842295	0.006606	6.640067	38.80675

- SATM=403+22.72*HighSchoolMath

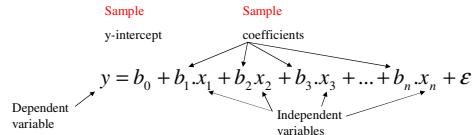- What if other variables are also important? Like gender… How can we control for that?

Multiple regression

- Multiple regression:

$$y = b_0 + b_1.x_1 + b_2 x_2 + b_3.x_3 + ... + b_n.x_n + \varepsilon$$

Sample y-intercept; Sample coefficients

Dependent variable; Independent variables

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.614968
R Square	0.378186
Adjusted R	0.35115
Standard E	79.72069
Observatio	49

ANOVA

	df	SS	MS	F	gnificance F
Regression	2	177805.1	88902.54	13.98853	1.8E-05
Residual	46	292347.9	6355.389		
Total	48	470153			

	Coefficients	andard Err	t Stat	P-value	Lower 95%	Upper 95%
Intercept	545.4428	67.54245	8.075555	2.29E-10	409.4869	681.3986
X Variable	-94.7772	22.90181	-4.13842	0.000147	-140.876	-48.6783
X Variable	22.1296	6.899878	3.207246	0.00244	8.240871	36.01834

Simple Reg: - SATM=403+22.72*HighSchoolMath

Multiple Reg: - SATM=545-94*Gender+22.12*HighSchoolMath

- We can even add more variables to the right side

Multiple regression

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.650391
R Square	0.423009
Adjusted R	0.384543
Standard E	77.64221
Observatio	49

ANOVA

	df	SS	MS	F	gnificance F
Regression	3	198878.9	66292.97	10.99694	1.54E-05
Residual	45	271274.1	6028.312		
Total	48	470153			

	Coefficients	andard Err	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	pper 95.0%
Intercept	651.2714	86.78105	7.504765	1.83E-09	476.4854	826.0574	476.4854	826.0574
X Variable	-112.376	24.20945	-4.64183	3E-05	-161.136	-63.6158	-161.136	-63.6158
X Variable	27.78912	7.370251	3.770444	0.000473	12.94467	42.63357	12.94467	42.63357
X Variable	-15.5316	8.306982	-1.86971	0.068039	-32.2628	1.199484	-32.2628	1.199484

Simple Reg: - SATM=403+22.72*HighSchoolMath

Multiple Reg: - SATM=545-94*Gender+22.12*HighSchoolMath

Multiple Reg: - SATM=651-112*Gender+27.78*HighSchoolMath-15.53*HighSchool Science

- We can still add more variables to the right side!

Multiple regression
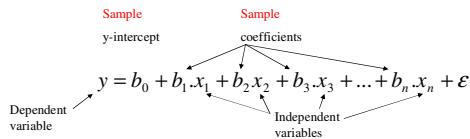
- Multiple regression:

Sample y-intercept   Sample coefficients

$$y = b_0 + b_1.x_1 + b_2.x_2 + b_3.x_3 + ... + b_n.x_n + \varepsilon$$

Dependent variable

Independent variables

Introduction

That's it! For the semester…

A few examples

Summary

- <u>An example:</u> (Q11.22 in the book) Online stock trading has increased recently, and online brokerages are accumulating huge assets. We would like to know if their *Market Share* or their *number accounts* are influencing their level of *assets*.

- What should we do?

Summary

SUMMARY OUTPUT

Regression Statistics

Multiple R	0.971357
R Square	0.943535
Adjusted R	0.927402
Standard E	20.5216
Observatio	10

ANOVA

	df	SS	MS	F	gnificance F
Regression	2	49260.5	24630.25	58.48526	4.28E-05
Residual	7	2947.952	421.136		
Total	9	52208.45			

	Coefficients	andard Err	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-21.4532	10.2432	-2.09439	0.074486	-45.6745	2.768113
X Variable	1.157541	1.34401	0.861259	0.417618	-2.02054	4.33562
X Variable	0.075594	0.011733	6.44314	0.000353	0.047851	0.103337

- Market Share = -21.45 +1.157*Mshare +0.075*Accts

Summary

SUMMARY OUTPUT

Regression Statistics

Multiple R	0.971357
R Square	0.943535
Adjusted R	0.927402
Standard E	20.5216
Observatio	10

ANOVA

	df	SS	MS	F	gnificance F
Regression	2	49260.5	24630.25	58.48526	4.28E-05
Residual	7	2947.952	421.136		
Total	9	52208.45			

	Coefficients	andard Err	t Stat	P-value	Lower 95%	Upper 95%
Intercept	-21.4532	10.2432	?	?	?	?
X Variable	1.157541	1.34401	?	?	?	?
X Variable	0.075594	0.011733	?	?	?	?

- Market Share = -21.45 +1.157*Mshare +0.075*Accts

Summary

- What we need to know:
  - 1. When to conduct a regression.
  - 2. To use excel to conduct regression.
  - 3. To interpret the results.
  - 4. To know how to get t-value and test significance of coefficients and confidence intervals (if t or p or both are not given)