





М	ultip	le R	egres	sion				
	Collected Data:							
ld	Price	Area	Taxes	Acreage	Rooms			
1	145	15	1.9	2.00	5			
2	228	38	3.0	3.60	11			
3	150	23	1.4	1.80	9			
4	130	16	1.4	0.53	7			
5	160	16	1.5	0.50	7			
6	114	13	1.8	0.31	7			
7	142	20	2.4	0.75	9			
8	265	24	4.0	2.00	7	and the second of		
				MA 214	1	5		

















Multiple regression model							
ANOVA table for multiple regression							
	Source	SS	DF	MS	F		
	Regression	SSR	р — 1	MSR	$F = \frac{MSR}{MSE}$		
	Error	SSE	<i>n</i> – p	MSE	MOL		
	SSTO	SSTO	n – 1				
$SSTO = \sum_{i=1}^{n} (y_i - \overline{y})^2$							
$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$							
$SSR = \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2 = SSTO - SSE$							
			MA 214			15	





















# Some important issues in multiple<br/>gcossionMuticollinearityWhen the covariates are highly correlated among themselves,<br/>difficulties can occur in carrying out inferences for individual<br/>covariates. In addition, correlation among covariates makes the<br/>interpretations of the coefficients of the covariates questionable.<br/>This phenomenon is called multicollinearity.

## Some important issues in multiple regression

#### Consequences of multicollinearity:

• In the presence of multicollinearity, the estimated regression coefficients  $\{\beta_{i}, i = 1, 2, ..., n\}$  tend to have large variability. This can result in extreme situations where the F-test shows a significant overall regression relation between the dependent variable and the covariates, but then the t-tests for individual  $\beta_i$ 's are not significant... an apparent contradiction!

MA 214

## <section-header><section-header><section-header><text>

## Some important issues in multiple regression

#### Selection of independent variables

Suppose in a regression model we have (p - 1) covariates. The question is how to come up with a subset of these covariates that provide essentially the same information about the dependent variable as the whole set of covariates currently in the model. The objective is to come up with a short list of covariates that adequately describe, predict and control the dependent variable with minimal multicollinearity.

MA 214

29

27







Sor reg	ne importa ression	nt issues i	in multiple	
<b>R</b> <sup>2</sup> <b>pro</b> ■ Cor (X <sub>1</sub> , sub	<b>cedure</b> isider the following e: $X_2, X_3$ ) in the model. sets of the covariates	xample: Suppose v Carrying out regre s, we get the follow	ve have three covaria ession on all possible ing <i>R</i> ² values.	tes
	# of Covariates	Covariates	R <sup>2</sup>	
	1	<i>X</i> <sub>1</sub>	0.32	
	1	X <sub>2</sub>	0.51	
	1	X <sub>3</sub>	0.42	
	2	X <sub>1</sub> , X <sub>2</sub>	0.64	
	2	X <sub>1</sub> , X <sub>3</sub>	0.72	
	2	X <sub>2</sub> , X <sub>3</sub>	0.61	
	3	$X_1, X_2, X_3$	0.74	
<ul> <li>Wh</li> </ul>	ich of these models is	s the "best"?		
		MA 214		33



## Some important issues in multiple regression

#### Stepwise Regression

- In multiple regression models, the response variable Y may depend on many covariates.
- Stepwise regression is a variable selection technique that tries to find the subset of the covariates that best predicts the response variable Y.
- One advantage of the stepwise procedure is that the variable selection is carried out automatically.

MA 214

35

Some important issues in multiple regression

#### Stepwise Regression

- However, the procedure is complex. It involves a series of hypothesis testing procedures that decide whether to include or not to include a variable in the model. There are three variations of the stepwise procedure:
  - Forward selection: begin with the smallest possible regression model, i.e., with just one covariate, and gradually work up to the multiple regression model incorporating the largest number of significantly important independent variables.

MA 214

36

## Some important issues in multiple regression

#### Stepwise Regression

- Backward elimination: begin with all covariates and gradually eliminate variables one by one until one reaches the point where all remaining covariates are significantly important.
- Stepwise: begin in the same way as forward selection, but each time a variable is added all variables in the model are examined to see if any should be eliminated at that step. This is a computationally more demanding procedure.

MA 214

37

### Some important issues in multiple regression

#### Stepwise Regression (Forward Selection)

- Begin by performing a normal multiple regression. If all variables are shown as significant (P-values < α), then STOP – the complete model is good.
- But if F-test is significant, but one or more of the pvalues for the t-tests are high (showing some of the covariates are not significant), forward stepwise regression can be used to develop the best model that contains some of the variables as follows.

MA 214

38













Otl	Other Model Selection Criterion: Adjusted R <sup>2</sup>							
Con	sider the ex	ample discu	ussed earlie	r. Suppose	n=10.			
	# of Covariates	Covariates	R <sup>2</sup>	Adj_R²				
	1	<i>X</i> <sub>1</sub>	0.32	0.24				
	1	X <sub>2</sub>	0.51	0.45				
	1	X <sub>3</sub>	0.42	0.35				
	2	$X_{1}, X_{2}$	0.64	0.54				
	2	$X_{1}, X_{3}$	0.72	0.64				
	2	X <sub>2</sub> , X <sub>3</sub>	0.61	0.50				
	3	$X_1, X_2, X_3$	0.74	0.61				
	MA 214							









Some important issues in multiple regression				
Some additional topics in regression <ul> <li>Example</li> </ul>				
The academic advisor continues	Y	<b>X</b> 1	<b>X</b> <sub>2</sub>	
her study on the salaries of recent	50	2	Male	
graduates and records several	42	10	Female	
additional observations. In total	55	1	Male	
16 individuals:	30	3	Female	
	63	5	Male	
Y Annual income of a graduate.	33	6	Female	
$X_1$ : Number of years of experience.				
$X_2$ : Gender of the graduate.		1	1	
MA 214			51	































