

11.1. Correlation And Regression

a.

- Sample covariance
- Sample correlation coefficient
- Scatter plot

b. Limitations to Correlation analysis

- Outliers
- Spurious correlation
- Nonlinear relationships

c. Hypothesis testing of correlation coefficient

d. Variables in a linear regression

Dependent (Y)

- Explained variable
- Endogeneous variable
- Predicted variable

Independent (X)

- Explanatory variable
- Exogenous variable
- Predicting variable

e1. Assumptions underlying linear regression

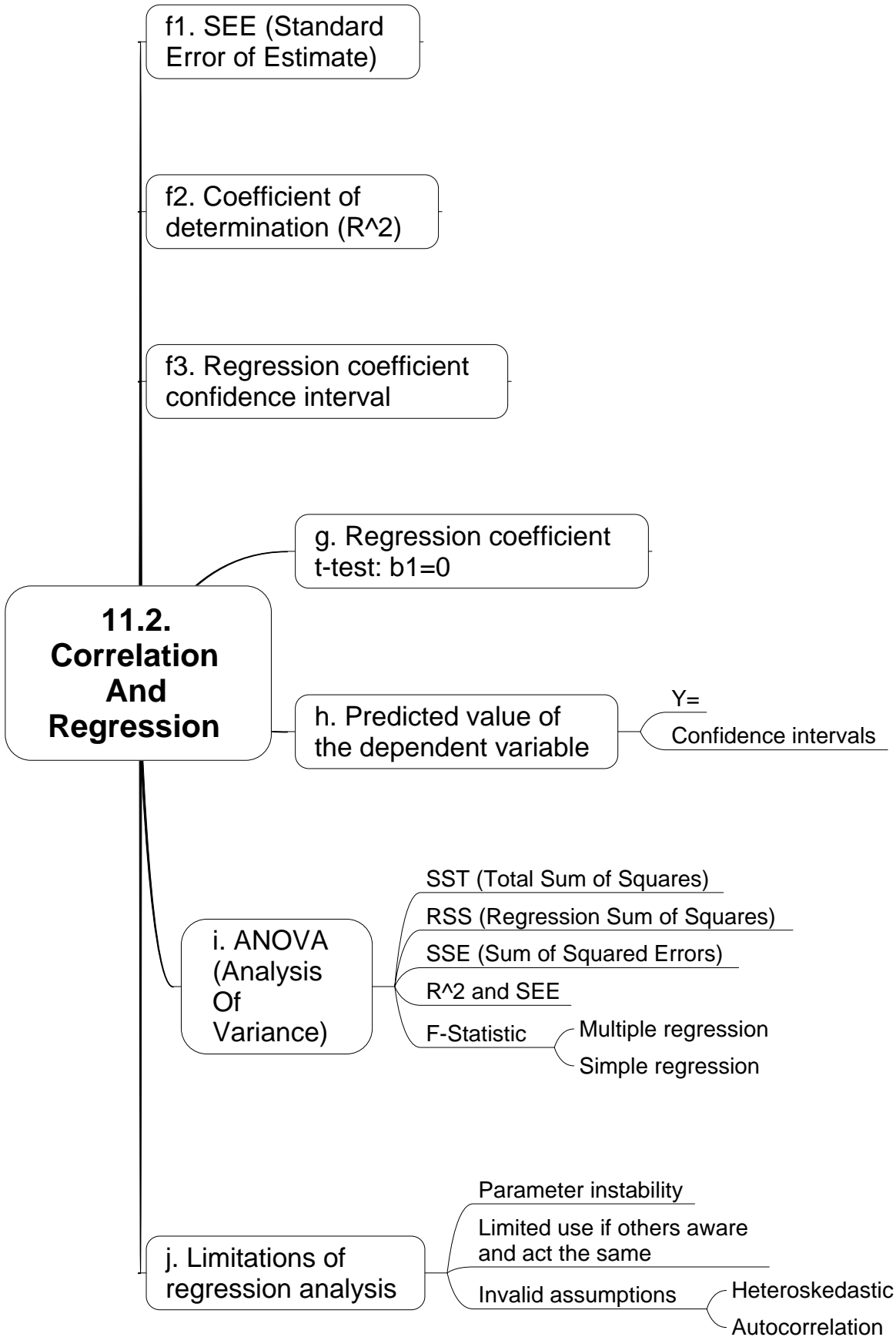
- linear relationship
- independent variable uncorrelated with residuals
- expected value of residual term = 0
- variance of residual term is constant
- residual term is independently distributed
- residual term is normally distributed

e2. Simple linear regression model

- Sum of Squared Errors (SSE)
- Ordinary Least Squares (OLS)

e3. Regression coefficients

- Slope coefficient
- Intercept



EXAMPLE READING 11: (Excel output)

Observation	X	Y
1	12	50
2	13	54
3	10	48
4	9	47
5	20	70
6	7	20
7	4	15
8	22	40
9	15	35
10	23	37

<i>Regression Statistics</i>	
Multiple R	0.47512
R Square	0.22574
Adjusted R Square	0.12896
Standard Error	15.05668
Observations	10

ANOVA					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Significance F</i>
Regression	1	528.77	528.77	2.33	0.17
Residual	8	1,813.63	226.70		
Total	9	2342.4			

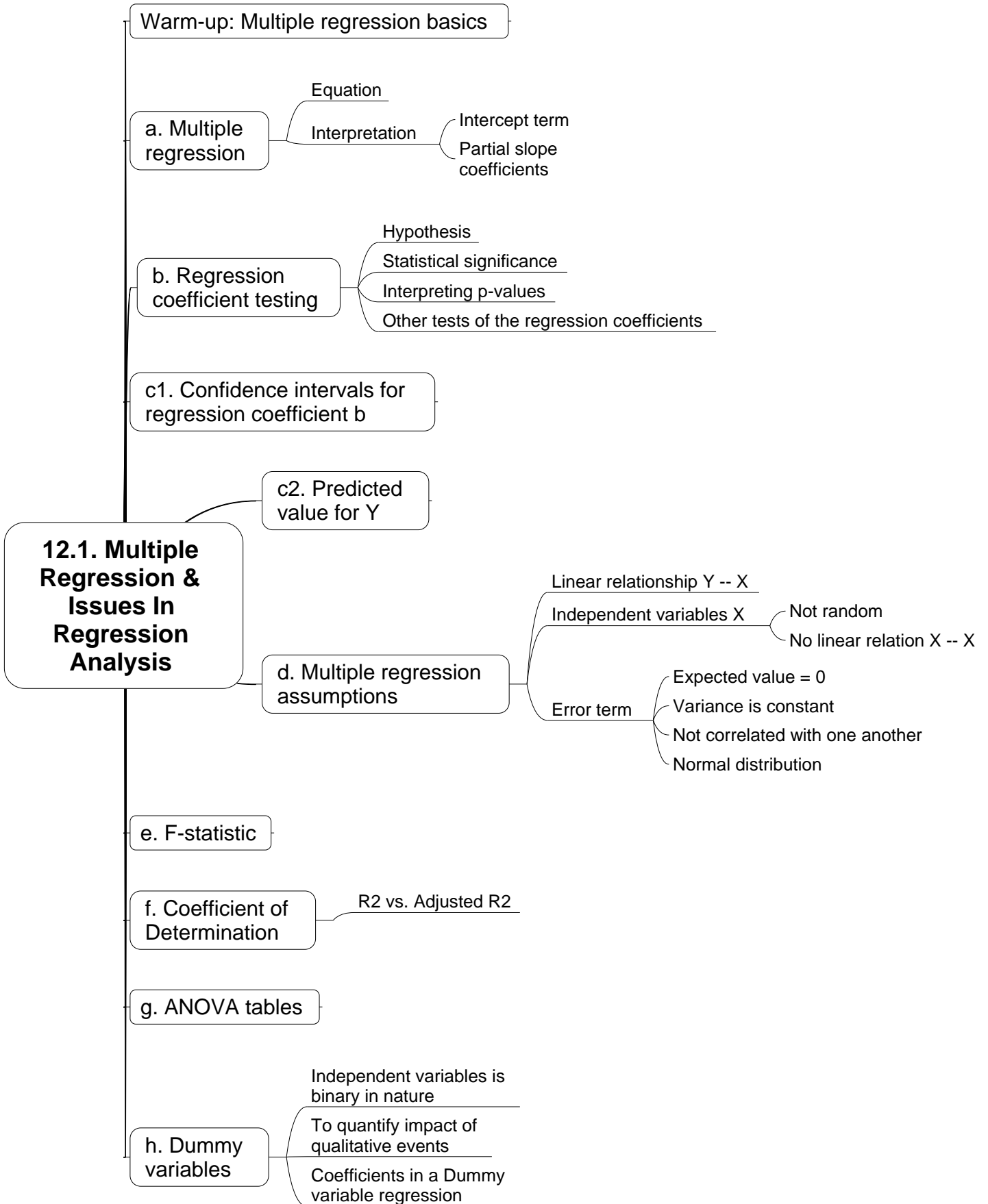
	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>	<i>Lower 95%</i>	<i>Upper 95%</i>	<i>Lower 95.0%</i>	<i>Upper 95.0%</i>
Intercept	25.5586	11.5324	2.2163	0.0575	-1.0351	52.1523	-1.0351	52.1523
X	1.1883	0.7780	1.5272	0.1652	-0.6059	2.9824	-0.6059	2.9824

RESIDUAL OUTPUT

<i>Observation</i>	<i>Predicted Y</i>	<i>Residuals</i>	<i>Standard Residuals</i>
1	39.8176	10.1824	0.7173
2	41.0059	12.9941	0.9154
3	37.4411	10.5589	0.7438
4	36.2529	10.7471	0.7571
5	49.3236	20.6764	1.4565
6	33.8764	-13.8764	-0.9775
7	30.3116	-15.3116	-1.0786
8	51.7001	-11.7001	-0.8242
9	43.3824	-8.3824	-0.5905
10	52.8884	-15.8884	-1.1192

PROBABILITY OUTPUT

<i>Percentile</i>	<i>Y</i>
5	15
15	20
25	35
35	37
45	40
55	47
65	48
75	50
85	54
95	70



Warm-up: Why multiple regression isn't easy as it looks

Assumption violations

i1. Heteroskedasticity

- What is it?
 - Unconditional
 - Conditional
- Effects on regression analysis
- Detecting heteroskedasticity
- Correcting heteroskedasticity

i2. Serial correlation (autocorrelation)

- What is it?
 - Positive
 - Negative
- Effects on regression analysis
- Detecting
- Correcting

j. Multicollinearity

- is
- Effects on regression analysis
- Detecting
- Correcting

12.2. Multiple Regression & Issues In Regression Analysis

Warm-up: Model specification

k. Model misspecification

- Subcategory 1: Misspecified functional form
 - Misspecification 1: Omitting a variable
 - Misspecification 2: Variables should be transformed
 - Misspecification 3: Incorrectly pooling data
- Subcategory 2: explanatory variables correlated with error term
 - Misspecification 4: use lagged Y as X
 - Misspecification 5: Forecasting the past
 - Misspecification 6: Measuring independent variables with error
- Subcategory 3: misspecifications resulting in nonstationarity

l. Models with qualitative dependent variables

- Probit and logit models
- Discriminant models

m. Interpreting regression results

LOS 12 i,j: Assumption violations

Assumption			
Violation	Heteroskedasticity Phương sai không đồng nhất	Serial Correlation (Autocorrelation) Tự tương quan	Multicollinearity Đa cộng tuyến
Definition	2 types: . Unconditional: . Conditional:	(esp. in time series) 2 types: . Positive: . Negative:	
Detecting	. Residual plots (Đồ thị phần dư): . Breusch-Pagan test:	. Residual plots: . DW (Durbin-Watson) test:	. High R^2 , reject F-test but not any t-tests . Rule of thumb:
Effects on regression analysis	. Standard errors: . t-test: . F-test:	. Positive: data cluster → standard errors too..... → t-stat too → Negative: data diverge → . F-test: unreliable	
Correcting	. Adjust standard errors: Robust std. errors White-corrected std. errors Heteroskedasticity-consistent std. errors →recalculate t-stats	. Adjust standard errors: Hansen-White std. errors (correct both heteroskedasticity & autocorrelation) Serial correlation consistent →recalculate t-stats . Improve specification (include seasonal terms)	. Omit 1 or more variables (not easy, must use stepwise regression)
NOTES:	. Regression analysis tests (t-tests, F-tests): H_0 : bad model (Reject H_0 →good model) . Assumption tests: H_0 : no violation (Fail to reject H_0 → good model)		

