Simple Linear Regression

Lecture Outline

- Regression: A classical view
- Regression as an ANOVA
- Regression diagnostics
- Reading

Objectives of Regression

- The primary objective of regression is to develop a linear relationship between a *response variable* and a *regressor* for the purposes of prediction

Objectives of Regression

- Regression is sometimes used to estimate the functional linear relationship between two variables

However, such an approach assumes that a functional linear relationship exists, and alternative approaches (functional regression) are superior.

Objectives of Regression

- Regression is sometimes interpreted as establishing causality

Only manipulative experiments can establish causality. Regression establishes a predictive relationship but does NOT on its own provide evidence of causality.

Example: Positive regression between Crime Rate and Number of Churches in cities.

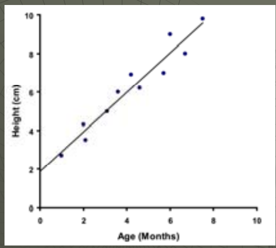
Regression and Correlation

- Regression establishes a predictive relationship between response variable Y and regressor X. The regressor is fixed, that is, its values are under the control of the experimenter.
- Correlation provides a measure of how strongly two variables X_1 and X_2 vary together (covary). Both X_1 and X_2 are free to vary naturally.

Regression Terminology

- The *response variable* is sometimes called the *dependent variable*
- The *regressor* is sometimes called the *independent variable* or the *covariate*

Regression Terminology

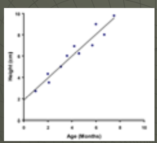


Response Variable:
Height (cm)

Regressor:
Age (months)

Regression line is uniquely defined by its slope and its Y intercept

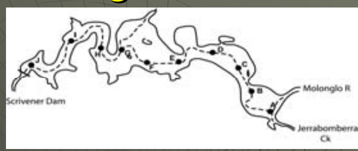
Statistical Regression



	Statistic	Parameter
Slope	B_1	β_1
Intercept	B_0	β_0
Equation	$Y = B_0 + B_1X$	$Y = \beta_0 + \beta_1$

The sample line is an estimate of a true but unknown parametric line

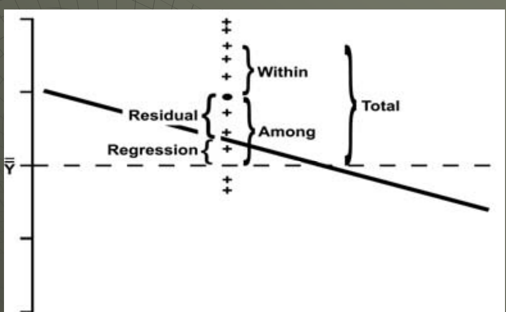
Regression as ANOVA



SITE									
A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**1**	**J**
43	25	23	32	17	23	14	13	15	13
28	28	24	32	21	21	18	26	15	15
43	28	30	32	18	17	14	18	14	14
28	25	32	33	17	18	16	15	12	13
42	25	25	32	25	19	14	15	17	16
43	25	28	29	17	24	9	14	19	19
40	26	23	26	18	14	14	17	15	16
35	25	25	38	14	17	26	15	14	15
42	23	26	27	15	17	10	11	16	11
43	25	27	29	15	18	15	14	14	15

More than one Y value for each X value

Regression as ANOVA



Regression as ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Among Sites	9	6025.04	669.4489	52.62	0.0001
Regression	1	4812.73	4812.730	31.76	0.0005
Residual	8	1212.31	151.5388	11.91	0.0001
Within	90	1145.00	12.72222		
Total	**99**	**7170.04**			

More than one Y value for each X value

Regression as ANOVA

Mean Squares and what the estimate

Source	Expected MS
Among Sites	$\sigma^2 + n\sigma_A^2$
Regression	$\sigma^2 + n\sigma_{res}^2 + n\sigma_{reg}^2$
Residual	$\sigma^2 + n\sigma_{res}^2$
Within	σ^2

F-test for variation Among Sites?

What does it mean?

F-test for Regression?

What does it mean?

F-test for Residual?

What does it mean?

Regression as ANOVA

Mean Squares and what the estimate

Source	Expected MS
Among Sites	$\sigma^2 + n\sigma_A^2$
Regression	$\sigma^2 + n\sigma_{res}^2 + n\sigma_{reg}^2$
Residual	$\sigma^2 + n\sigma_{res}^2$
Within	σ^2

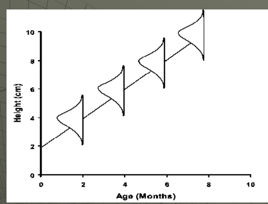
If you have only one Y for each X, what is it you can test and what can you not test?

Regression as ANOVA

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Regression	1	4812.73	4812.730	31.76	0.0005
Residual	8	1212.31	151.5388		
Total	**99**	**7170.04**			

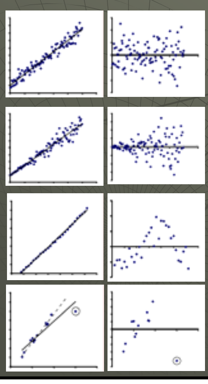
One Y value for each X value

Regression Assumptions



- Heterogeneity of Variances
- Normality of errors

Residual Analysis



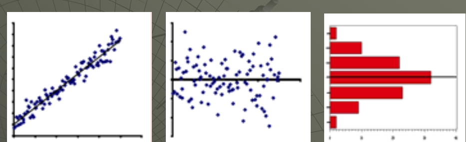
Homogeneity of Variances

Heterogeneity of Variances

Serial dependence

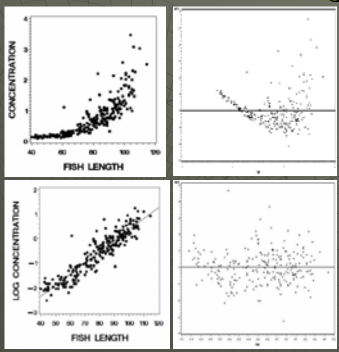
Influential aberrant outlier

Residual Analysis



Test of Normality

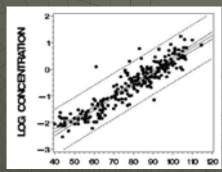
Curvilinear Regression

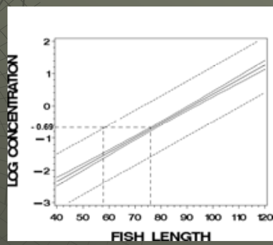


Pre-Transformation

Post-Transformation

Confidence Limits





Reading

- Module 6: Key concepts
- Regression as an ANOVA
- Worked Examples