

Methods of Data Analysis II

Derek Li

Contents

1	Review	2
1.1	Linear Regression	2
1.2	ANOVA	2

1 Review

1.1 Linear Regression

Recall the assumptions of linear regression:

1. Errors are independent (observations are independent);
2. Errors are identically distributed with $\mathbb{E}[\varepsilon_i] = 0$;
3. Homoscedasticity: $\text{Var}[\varepsilon_i] = \sigma^2$;
4. A straight-line relationship exists between ε_i and y_i .

1.2 ANOVA

Analysis of variance (ANOVA) is used to test differences between two or more means and to test general rather than specific differences among means. We test

$$H_0 : \mu_1 = \dots = \mu_n \text{ against } H_1 : \text{At least one mean is different from the others.}$$

Recall the assumptions of ANOVA:

1. Errors are independent;
2. Errors are normally distributed with $\mathbb{E}[\varepsilon_i] = 0$;
3. Homoscedasticity.

Note that the normality assumption can be relaxed if sample size is large (Central Limit Theorem). The normality assumption is most important when n is small, highly non-normal or small effect size.

In fact, ANOVA is just a specific case of the general linear model.