Stat 206: Linear Models

Lecture 5

October 9, 2019

Analysis of Variance Approach

The basic idea of ANOVA is to attributing variation in the data to different sources.

- In regression, the variation in the observations Y_i is attributed to:
  -
  -
- ANOVA is performed through:
  - Partitioning sums of squares;
  - Partitioning degrees of freedoms;

Analysis of Variance Approach

The basic idea of ANOVA is to attributing variation in the data to different sources.

- In regression, the variation in the observations Y_i is attributed to:
  - variation of the error terms – random fluctuation
  - variation of the values of the predictor variable(s) – dispersion of X
- ANOVA is performed through:
  - partitioning sums of squares
  - partitioning degrees of freedoms

Partition of Total Deviations

- **Total deviations:** Difference between Y_i and the sample mean $\overline{Y}$:

$$Y_i - \overline{Y}, \quad i = 1, \cdots, n.7$$

- Total deviations can be decomposed into the sum of two terms:

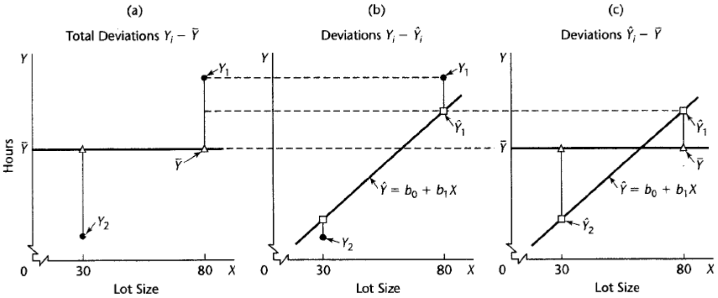
i.e., the *deviation of observed value around the fitted regression line* – and the *deviation of fitted value from the mean*.

Partition of Total Deviations

- **Total deviations:** Difference between Y_i and the sample mean $\overline{Y}$:
$$Y_i - \overline{Y}, \quad i = 1, \cdots, n.$$
- Total deviations can be decomposed into the sum of two terms:
$$Y_i - \overline{Y} = (Y_i - \widehat{Y}_i) + (\widehat{Y}_i - \overline{Y}), \qquad i = 1, \ldots, n,$$
i.e., the *deviation of observed value around the fitted regression line (residual)* and the *deviation of fitted value from the mean*.

Figure: Partition of total deviation.

FIGURE 2.7 Illustration of Partitioning of Total Deviations $Y_i - \bar{Y}$—Toluca Company Example (not drawn to scale; only observations Y_1 and Y_2 are shown).



From Applied Linear Statistical Models by Kutner, Nachtsheim, Neter and Li

Decomposition of Total Variation

Decomposition of Total Variation

- Taking sum of squares of the total deviations and noting that the sum of the cross product terms vanishes:

$$\sum_{i=1}^{n}(Y_i - \overline{Y})^2 = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2 + \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2.$$

- Decomposition of total variation:

$$SSTO = SSE + SSR$$

and

$$d.f.(SSTO) = d.f.(SSE) + d.f.(SSR).$$

Sum of Squares

- **Total sum of squares (SSTO)**:

  This is the variation of the observed Y_is around their sample mean.
- **Error sum of squares (SSE)**:

  This is the variation of the observed Y_is around the fitted regression line.

Sum of Squares

- **Total sum of squares (SSTO)**:

$$SSTO := \sum_{i=1}^{n} (Y_i - \overline{Y})^2, \quad d.f.(SSTO) = n - 1.$$

  This is the variation of the observed Y_is around their sample mean.

- **Error sum of squares (SSE)**:

$$SSE = \sum_{i=1}^{n} (Y_i - \widehat{Y}_i)^2, \quad d.f.(SSE) = n - 2.$$

  This is the variation of the observed Y_is around the fitted regression line.

- **Regression sum of squares (SSR)**:

  This is the variation of the fitted values around the sample mean. The the fitted regression slope and the dispersion in X_is, the larger is SSR.
- $SSR = SSTO - SSE$ is the effect of X in the variation in Y through linear regression.
- In other words, SSR is the in predicting Y by utilizing the predictor X through a linear regression model.

What is $\frac{1}{n}\sum_{i=1}^{n}\widehat{Y}_i$?

- **Regression sum of squares (SSR)**:

$$SSR = \sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y})^2 = \hat{\beta}_1^2 \sum_{i=1}^{n} (X_i - \overline{X})^2, \quad d.f.(SSR) = 1.$$

  This is the variation of the fitted values around the sample mean: the larger the fitted regression slope and the more dispersion in X_is, the larger is SSR.

- $SSR = SSTO - SSE$ is the effect of X in reducing the variation in Y through linear regression.
- In other words, SSR is the reduction in uncertainty in predicting Y by utilizing the predictor X through a linear regression model.

What is $\frac{1}{n}\sum_{i=1}^{n} \widehat{Y}_i$?

Expected Values of SS

- Expected values of SS:

What is $E(SSTO)$?

- **Mean squares (MS): = SS / df(SS)**

$$MSE = \frac{SSE}{\text{d.f.}(SSE)} = \frac{SSE}{n-2}, \qquad MSR = \frac{SSR}{\text{d.f.}(SSR)} = \frac{SSR}{1}.$$

- Expected values of MS:

$$E(MSE) = \sigma^2, \qquad E(MSR) = \sigma^2 + \beta_1^2 \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Expected Values of SS

- Expected values of SS:

$$E(SSE) = (n-2)\sigma^2, \quad E(SSR) = \sigma^2 + \beta_1^2 \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

What is $E(SSTO)$?

- **Mean squares (MS): = SS / df(SS)**

$$MSE = \frac{SSE}{\text{d.f.}(SSE)} = \frac{SSE}{n-2}, \qquad MSR = \frac{SSR}{\text{d.f.}(SSR)} = \frac{SSR}{1}.$$

- Expected values of MS:

$$E(MSE) = \sigma^2, \qquad E(MSR) = \sigma^2 + \beta_1^2 \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

Under Normal error model.

- $SSE \sim \sigma^2\chi^2_{(n-2)}$
- SSE and SSR are independent.
  Notes: Recall SSE and $\hat{\beta}_1$ are independent.

What is the distribution of SSR and $SSTO$ when $\beta_1 = 0$?

F Test

- $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$.
- F ratio:

$$F^* = \frac{MSR}{MSE} = \frac{SSR/1}{SSE/(n-2)}.$$

  - F^* fluctuates around $1 + \frac{\beta_1^2 \sum_{i=1}^n (X_i - \overline{X})^2}{\sigma^2}$.
  - A large value of F^* means evidence against H_0.
- Null distribution of F^*:

$$F^* \underset{H_0:\beta_1=0}{\sim} F_{1,n-2}.$$

Notes: Use the fact that if $Z_1 \sim \chi^2_{(df_1)}$, $Z_2 \sim \chi^2_{(df_2)}$ *and* Z_1, Z_2 *independent, then* $\frac{Z_1/df_1}{Z_2/df_2} \sim F_{df_1,df_2}$.

- Decision rule at level α:

$$\text{reject } H_0 \text{ if} \quad F^* > F(1-\alpha; 1, n-2),$$

where $F(1-\alpha; 1, n-2)$ is the $(1-\alpha)$-percentile of the $F_{1,n-2}$ distribution.

- **In simple linear regression, the F-test is equivalent to the t-test for testing $H_0 : \beta_1 = 0$ versus $H_a : \beta_1 \neq 0$.**
Check the following.
    - $F^* = (T^*)^2$ where $T^* = \frac{\hat{\beta}_1}{s(\hat{\beta}_1)}$ is the T-statistic.
    - $F(1-\alpha; 1, n-2) = t^2(1-\alpha/2; n-2)$.

ANOVA Table

ANOVA table for simple linear regression.

Source of Variation	SS	d.f.	MS=SS/d.f.	F^*
Regression	$SSR = \sum_{i=1}^{n}(\widehat{Y}_i - \overline{Y})^2$	d.f.$(SSR) = 1$	$MSR = SSR/1$	$F^* = MSR/MSE$
Error	$SSE = \sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2$	d.f.$(SSE) = n-2$	$MSE = SSE/(n-2)$	
Total	$SSTO = \sum_{i=1}^{n}(Y_i - \overline{Y})^2$	d.f.$(SSTO) = n-1$	$MSTO = SSTO/(n-1)$	

Heights

$n = 928$, $\overline{X} = 68.31578$, $\overline{Y} = 68.08227$, $\sum_i X_i^2 = 4334058$, $\sum_i Y_i^2 = 4307355$, $\sum_i X_i Y_i = 4318152$, $\hat{\beta}_1 = 0.637$, $\hat{\beta}_0 = 24.54$.

$$
\begin{aligned}
SSTO &= \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = \sum_{i=1}^{n} Y_i^2 - n(\overline{Y})^2 \\
&= 4307355 - 928 \times 68.08227^2 = 5893. \\
SSR &= \sum_{i=1}^{n} (\widehat{Y}_i - \overline{Y})^2 = \hat{\beta}_1^2 \sum_{i=1}^{n} (X_i - \overline{X})^2 \\
&= 0.637^2 \times \left[4334058 - 928 \times 68.31578^2\right] = 1234. \\
SSE &= SSTO - SSR = 4659.
\end{aligned}
$$

Heights (Cont'd)

Source of Variation	SS	d.f.	MS=SS/d.f.	F^*
Regression	$SSR = 1234$	d.f.$(SSR) = 1$	$MSR = 1234$	$F^* = MSR/MSE = 245$
Error	$SSE = 4659$	d.f.$(SSE) = 926$	$MSE = 5.03$	
Total	$SSTO = 5893$	d.f.$(SSTO) = 927$	$MSTO = 6.36$	

- Test whether there is a linear association between parent's height and child's height. Use significance level $\alpha = 0.01$.
- $F(0.99; 1, 926) = 6.66 < F^* = 245$, so reject $H_0 : \beta_1 = 0$ and conclude that there is a significant linear association between parent's height and child's height.
- Recall $T^* = 15.66$, $t(0.995; 926) = 2.58$ and check:

$$15.66^2 = 245, \quad 2.58^2 = 6.66.$$

Coefficient of Determination R^2

- R^2 is a descriptive measure for linear association between X and Y:
$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}.$$
- **R^2 is the of the variation in Y by explaining Y using X through a linear regression model.**
- Heights.
$$R^2 = \frac{1234}{5893} = 0.209.$$
20% of variation in child's height may be "explained" by
.

Coefficient of Determination R^2

- R^2 is a descriptive measure for linear association between X and Y:

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}.$$

- **R^2 is the proportional reduction of the variation in Y by explaining Y using X through a linear regression model.**
- Heights.

$$R^2 = \frac{1234}{5893} = 0.209.$$

20% of variation in child's height may be "explained" by the variation in parent's height.

Properties of R^2

- If all observations Y_is fall on one straight line, then
  - The predictor variable X accounts for in the observations Y_is.
- If the fitted regression line is horizontal, i.e., $\hat{\beta}_1 = 0$, then
  - The predictor variable X is in explaining the variation in the observations Y_is.
  - There is linear association between X and Y in the data.

Properties of R^2

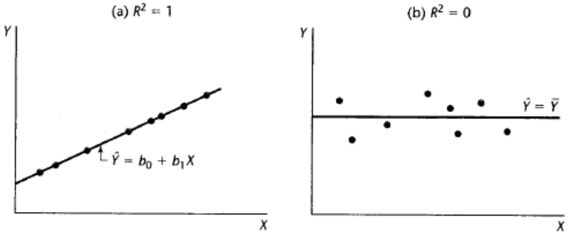
Since $0 \leq SSE, SSR \leq SSTO$, it follows:

$$0 \leq R^2 \leq 1.$$

- All observations Y_is fall on one straight line $\iff SSE = 0$ $\iff R^2 = 1$.
  - The predictor variable X accounts for all variation in the observations Y_is.
- The fitted regression line is horizontal, i.e., $\hat{\beta}_1 = 0$ $\iff SSR = 0 \iff R^2 = 0$.
  - The predictor variable X is of no use in explaining the variation in the observations Y_is.
  - There is no evidence of linear association between X and Y in the data.

Figure:

FIGURE 2.8 Scatter Plots when $R^2 = 1$ and $R^2 = 0$.



From Applied Linear Statistical Models by Kutner, Nachtsheim, Neter and Li