Lecture 5: The Multiple Regression Model

Shuo Liu

UCLA Summer School Econ 103

July 10, 2017

Outline

Introduction

- We refer to a model with more than one explanatory variables as **multiple regression model**;
- The results (e.g. estimation method, properties of estimators and etc) can be easily extended to multiple regression model.

Introduction

Example: we initially hypothesize that **sales** revenue is linearly related to **price** and **advertising expenditure**. The model is:

$$SALES = \beta_1 + \beta_2 PRICE + \beta_3 ADVERT + e \tag{1}$$

- How to interpret β_2: the change in sales (in \$1000) when the price index $PRICE$ increases by one unit (in \$1), and advertising expenditure $ADVERT$ is held constant

$$\beta_2 = \left.\frac{\triangle SALES}{\triangle PRICE}\right|_{ADVERT \text{ is held constant}} = \frac{\partial SALES}{\partial PRICE}$$

- Similar way to interpret β_3.

$$\beta_3 = \left.\frac{\triangle SALES}{\triangle ADVERT}\right|_{PRICE \text{ is held constant}} = \frac{\partial SALES}{\partial ADVERT}$$

- Similar to simple linear regression model, we add a random error term e.

Introduction

More generally, the multiple regression model with K population parameters is:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_K x_K + e \tag{2}$$

- For $\forall k = 2, \cdots, K$, β_k measures the effect of one unit change in x_k on the expected value of y, when all the other variables are held constant

$$\beta_k = \frac{\partial E(y)}{\partial x_k}, \quad \forall k \tag{3}$$

- The assumptions are identical to those that we made for **simple linear regression** model:
    1. $E(e_i) = 0$ for $\forall i$
    2. $Var(e_i) = \sigma^2$, i.e. homoskedasticity, for $\forall i$
    3. $Cov(e_i, e_j) = 0$ for $\forall i \neq j$
    4. $e_i \sim N(0, \sigma^2)$ for $\forall i$ (optional)

Introduction

- Assumptions on e are also assumptions on y (still assume x is nonrandom and y is random)
  1. $E(y_i) = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i}$ for $\forall i$
  2. $Var(y_i) = Var(e_i) = \sigma^2$ for $\forall i$
  3. $Cov(y_i, y_j) = Cov(e_i, e_j) = 0$ for $\forall i \neq j$
  4. $y|_{x_2, x_3} \sim N[(\beta_1 + \beta_2 x_2 + \beta_3 x_3), \sigma^2]$
- Another assumption on explanatory variables:

  Any one of the explanatory variables is not an exact linear function of the others. (intuition: no explanatory variable is redundant and no explanatory variable only contains information that can also be obtained from other explanatory variables)
- **Comment**: If the above extra assumption is violated–a situation/condition that we call **exact collinearity**–the OLS method fails and Stata will not run the regression.

Assumptions of the Multiple Regression Model (assume K population parameters)

- MR1. $y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_K x_{iK} + e_i, \forall i = 1, 2, \cdots, n$
- MR2. $E(y_i) = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + \cdots + \beta_K x_{iK} \Longleftrightarrow E(e_i) = 0$
- MR3. $Var(y_i) = Var(e_i) = \sigma^2$
- MR4. $Cov(y_i, y_j) = Cov(e_i, e_j) = 0$ for $\forall i \neq j$
- MR5. The values of each x_{ki} $k = 2, 3, \cdots, K$ and $i = 1, 2, \cdots, n$ are not random and are not exact linear functions of the other explanatory variables
- MR6. $y|_{x_2, x_3} \sim N[(\beta_1 + \beta_2 x_2 + \beta_3 x_3), \sigma^2] \Longleftrightarrow e \sim N(0, \sigma^2)$

Estimating the Parameters of the Multiple Regression Model

- We discuss the estimation in the context of the following model:

$$y_i = \beta_1 + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i \quad \forall i = 1, 2, \cdots, n$$

- Similar as simple linear regression model, we still search for optimal (b_1, b_2, b_3) as estimators of $(\beta_1, \beta_2, \beta_3)$ to minimize the **sum of squares of error** function $S(b_1, b_2, b_3)$:

$$S(b_1, b_2, b_3) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_1 - b_2 x_{i2} - b_3 x_{i3})^2 \qquad (4)$$

Estimating the Parameters of the Multiple Regression Model

- Now we have **three control variables**, we need **three first order conditions** through taking partial derivatives of objective function w.t. b_1, b_2 and b_3.

$$\frac{\partial S(b_1, b_2, b_3)}{\partial b_1} = -2\sum_{i=1}^{n}(y_i - b_1 - b_2 x_{i2} - b_3 x_{i3}) = 0 \tag{5}$$

$$\frac{\partial S(b_1, b_2, b_3)}{\partial b_2} = -2\sum_{i=1}^{n}x_{i2}(y_i - b_1 - b_2 x_{i2} - b_3 x_{i3}) = 0 \tag{6}$$

$$\frac{\partial S(b_1, b_2, b_3)}{\partial b_3} = -2\sum_{i=1}^{n}x_{i3}(y_i - b_1 - b_2 x_{i2} - b_3 x_{i3}) = 0 \tag{7}$$

Estimating the Parameters of the Multiple Regression Model

- **Why for $\forall k = 2, 3, \cdots, K$, x_k (the observations on x_k is $\{x_{ki}\}_{i=1}^{n}$), cannot be the exact linear functions(combinations) of the other explanatory variables:**
  If for example, $x_{3i} = c + ax_{2i}$ for $\forall i = 1, 2, \cdots, n$ where c and a are constants, then the first order conditions (5)-(7) will reduct to two effective conditions.

$$\begin{aligned}
\frac{\partial S(b_1, b_2, b_3)}{\partial b_3} &= -2\sum_{i=1}^{n} x_{i3}(y_i - b_1 - b_2 x_{i2} - b_3 x_{i3}) \\
&= -2\sum_{i=1}^{n}(c + ax_{2i})(y_i - b_1 - b_2 x_{i2} - b_3 x_{i3}) \\
&= -2\left\{c\sum_{i=1}^{n}(y_i - b_1 - b_2 x_{i2} - b_3 x_{i3})\right. \\
&\left. + a\sum_{i=1}^{n} x_{i2}(y_i - b_1 - b_2 x_{i2} - b_3 x_{i3})\right\} = 0
\end{aligned}$$

Estimating the Parameters of the Multiple Regression Model

- Again (b_1, b_2, b_3) are OLS estimators of $(\beta_1, \beta_2, \beta_3)$, which can be regarded as functions of sample data;
- Report of estimated multiple regression model is usually like:

$$\underset{(Se)}{\widehat{SALES}} = \underset{(6.35)}{118.91} - \underset{(1.096)}{7.908} PRICE + \underset{(0.683)}{1.863} ADVERT, \quad R^2 = 0.448 \quad (8)$$

- How to interpret:

  1. Negative coefficient on PRICE suggests that demand is price elastic. **Holding advertising constant**, an increase in price of \$1 will lead to a decrease in monthly revenue of \$7908;
  2. The coefficient on advertising is positive. **Holding price constant**, an increase in advertising expenditure of \$1000 will lead to an increase in sales revenue of \$1863;

  Including an intercept term improves the fitting of data even when it is not directly interpretable

Properties of OLS Estimators

- Firstly, as in simple linear regression, we need to estimate the constant variance of random error term σ^2, recall that $Var(e_i) = E(e_i^2) = \sigma^2$;
- After getting the fitted model, we calculate residuals:

$$\hat{e_i} = y_i - \hat{y_i} = y_i - (b_1 + b_2 x_{i2} + b_3 x_{i3}) \tag{9}$$

- Estimator of σ^2:

$$\hat{\sigma}^2 = \frac{1}{n-K} \sum_{i=1}^{n} \hat{e_i}^2 = \frac{1}{n-K} SSE \tag{10}$$

where K is the number of population parameters to estimate.

Properties of OLS Estimators

Gauss Markov Theorem: For the multiple regression model, if assumptions MR1-MR5 hold, then the least squares estimators are the best linear unbiased estimators (BLUE) of the parameters.

Variance of Estimators

- We can show that:

$$Var(b_2) = \frac{\sigma^2}{(1 - r_{23}^2) \sum_{i=1}^{n} (x_{2i} - \bar{x}_2)^2} \tag{11}$$

where

$$r_{23} = \frac{\sum_{i=1}^{n} (x_{2i} - \bar{x}_2)(x_{3i} - \bar{x}_3)}{\sqrt{\sum_{i=1}^{n} (x_{2i} - \bar{x}_2)^2} \sqrt{\sum_{i=1}^{n} (x_{3i} - \bar{x}_3)^2}} \tag{12}$$

Properties of OLS Estimators

Comments:

- Larger variance σ^2 (variance of random error term/uncertainty of the whole model) lead to larger variances of least square estimator b_2;
- Larger sample size n imply smaller variances of the least square estimator b_2;
- Larger variation in an explanatory variable around its mean will lead to smaller variance of least square estimator b_2;
- Larger correlation between x_2 and x_3 leads to larger variance of b_2. If $r_{23} = 1$, then $Var(b_2) = \infty$. This simply means that β_2 cannot be estimated, which is the problem of **collinearity**.

Properties of OLS Estimators

Variance-covariance matrix:

$$Cov(b_1, b_2, b_3) = \begin{pmatrix} Var(b_1) & Cov(b_1, b_2) & Cov(b_1, b_3) \\ Cov(b_1, b_2) & Var(b_2) & Cov(b_2, b_3) \\ Cov(b_1, b_3) & Cov(b_2, b_3) & Var(b_3) \end{pmatrix} \tag{13}$$

Properties of OLS Estimators

The Distribution of the Least-Square Estimator:

Consider the general model:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_K x_{Ki} + e_i \quad \text{for } \forall i \tag{14}$$

- If we have assumption MR6, then the random errors follow normal distribution $N(0, \sigma^2)$, then

$$y_i|_{x_{2i}, x_{3i}, \cdots, x_{Ki}} \sim N([\beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_K x_{Ki}], \sigma^2) \tag{15}$$

- If the errors are not normally distributed, then the least squares estimators are approximately normally distributed in **large samples**.

Properties of OLS Estimators

- Since the least square estimators $b_k, k = 1, 2, \cdots, K$, are linear functions of the dependent variable (random) and independent variable (nonrandom), it follows that the least square estimators are also normally distributed.

$$b_k \sim N(\beta_k, Var(b_k)) \quad k = 1, 2, \cdots, K \tag{16}$$

- We can further transform b_k to standard normal variable Z:

$$\frac{b_k - \beta_k}{\sqrt{Var(b_k)}} \sim N(0, 1) \quad k = 1, 2, \cdots, K \tag{17}$$

- Replacing the theoretical variance term σ^2 in $Var(b_k)$, we will get the t-distribution statistic:

$$t = \frac{b_k - \beta_k}{\sqrt{\hat{Var}(b_k)}} = \frac{b_k - \beta_k}{Se(b_k)} \sim t_{(n-K)} \tag{18}$$

the degree of freedom of t-distribution now is $n - K$.

Properties of OLS Estimators

- Besides each parameter separately, we can also form linear combination of parameters:

$$\lambda = c_1\beta_1 + c_2\beta_2 + \cdots + c_K\beta_K = \sum_{k=1}^{K} c_k\beta_k \tag{19}$$

- Then we have

$$t = \frac{\hat{\lambda} - \lambda}{Se(\hat{\lambda})} = \frac{\sum_{k=1}^{K} c_k b_k - \sum_{k=1}^{K} c_k\beta_k}{Se(\sum_{k=1}^{K} c_k b_k)} \sim t_{(n-K)} \tag{20}$$

- If $K = 3$, we have:

$$Se(c_1b_1 + c_2b_2 + c_3b_3) = \sqrt{V\hat{a}r(c_1b_1 + c_2b_2 + c_3b_3)} \tag{21}$$

where

$$V\hat{a}r(c_1b_1 + c_2b_2 + c_3b_3) = c_1^2 V\hat{a}r(b_1) + c_2^2 V\hat{a}r(b_2) + c_3^2 V\hat{a}r(b_3) \tag{22}$$

$$+2c_1c_2C\hat{o}v(b_1, b_2) + 2c_1c_3C\hat{o}v(b_1, b_3) + 2c_3c_2C\hat{o}v(b_3, b_2)$$

Confidence Interval

- To construct X level of confidence interval, we determine the $\alpha = 1 - X$, then find **two** symmetric critical values, $t_c = t_{(1-\frac{\alpha}{2}, n-K)}$ (and $-t_c$) s.t.

$$P(-t_c < t < t_c) = 1 - \alpha = X \tag{23}$$

- Replace t-statistic with its expression and do transformation:

$$P(-t_c < \frac{b_k - \beta_k}{Se(b_k)} < t_c) = 1 - \alpha = X \tag{24}$$

$\Longleftrightarrow$

$$P(b_k - t_c * Se(b_k) < \beta_k < b_k + t_c * Se(b_k)) = 1 - \alpha = X \tag{25}$$

- X-level confidence interval for β_k is:

$$[b_k - t_c * Se(b_k), b_k + t_c * Se(b_k)] \tag{26}$$

Confidence Interval

- Another form of formula of $100(1-\alpha)\%$ confidence interval:

$$[b_k - t_{(1-\frac{\alpha}{2},n-K)} * Se(b_k), b_k + t_{(1-\frac{\alpha}{2},n-K)} * Se(b_k)] \tag{27}$$

- Go back to the SALES example (explained by PRICE and ADVERT):

$$SALES = \beta_1 + \beta_2 PRICE + \beta_3 ADVERT + e \tag{28}$$

Using the sample data, we have $b_2 = -7.908$ and $Se(b_2) = 1.096$, so that the 95% confidence interval is:

$$[-7.908 - 1.993 * 1.096, -7.908 + 1.993 * 1.096] = [-10.093, -5.723]$$

How to interpret? decreasing price by \$1 will lead to an increase in revenue somewhere between \$5,723 and \$10,093.

Confidence Interval

- **Example for Linear Combination of Parameters:** Suppose we want to increase advertising expenditure (ADVERT) by \$800 and drop the price by 40 cents. Unit of ADVERT is \$1000 and unit of PRICE is \$1 (100 cents). Then construct 90% confidence interval for the change in expected sales:

$$\begin{aligned}\lambda &= E(SALES_1) - E(SALES_0) \\ &= [\beta_1 + \beta_2(PRICE_0 - 0.4) + \beta_3(ADVERT_0 + 0.8)] \\ &- [\beta_1 + \beta_2 PRICE_0 + \beta_3 ADVERT_0] \\ &= -0.4\beta_2 + 0.8\beta_3\end{aligned}$$

- A point estimate would be:

$$\hat{\lambda} = -0.4b_2 + 0.8b_3 = 4.6532 \tag{29}$$

- 90% confidence interval ($\alpha = 0.1$) is

$$[\hat{\lambda} - t_{(1-0.05,n-3)}Se(\hat{\lambda}), \hat{\lambda} + t_{(1-0.05,n-3)}Se(\hat{\lambda})] \tag{30}$$

Hypothesis Testing

Components:

- A null hypothesis H_0;
- An alternative hypothesis H_1;
- **A test statistic**(can follow t distribution, χ^2 distribution, F distribution and etc) and obtain the value of test statistic **conditional on H_0 is true**
- The significance level α is known. Then the first way to test: one or two critical values and construct rejection region; Second way: use value of test statistic to calculate p-value and compare p-value with significance level α.
- Conclusion

Hypothesis Testing

Test each parameter separately (test significance for each explanatory variable $x_k, k = 2, 3, \cdots, K$)

- Null Hypothesis:

$$H_0 : \beta_k = 0$$

- Alternative Hypothesis:

$$H_1 : \beta_k \neq 0$$

- Value of Test Statistic Conditional on H_0 is True:

$$t = \frac{b_k}{Se(b_k)} \sim t_{(n-K)}$$

- **First Way** is to Use Critical Values with Significance Level α:

$$t_c = t_{(1-\frac{\alpha}{2}, n-K)} \quad \text{and} \quad -t_c = -t_{(1-\frac{\alpha}{2}, n-K)}$$

- **Second Way** is to Use p-Value:

$$p = P(t_{(n-K)} > t) + P(t_{(n-K)} < -t)$$

Other Forms of Multiple Regression Model

1. Polynomial Equations

- Sometimes we are interested in polynomial equations such as the quadratic one:

$$y = \beta_1 + \beta_2 x + \beta_3 x^2 + e \tag{31}$$

or the cubic one:

$$y = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 + e \tag{32}$$

Other Forms of Multiple Regression Model

- The general function is:

$$E(y|x) = \beta_1 + \beta_2 x + \beta_3 x^2 + \beta_4 x^3 + \cdots + \beta_p x^p \tag{33}$$

the slope is:

$$\frac{dE(y|x)}{dx} = \beta_2 + 2\beta_3 x + 3\beta_4 x^2 + \cdots + p\beta_p x^{p-1} \tag{34}$$

Other Forms of Multiple Regression Model

Why do we want polynomial equations? Go back to SALES example, the **linear sales model with constant slope as β_3 cannot capture the diminishing returns in adversiting expenditure**, if instead we use:

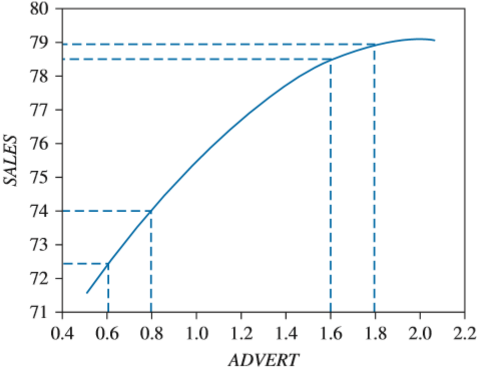
$$SALES = \beta_1 + \beta_2 PRICE + \beta_3 ADVERT + \beta_4 ADVERT^2 + e \tag{35}$$

the change in expected sales due to a change in advertising is:

$$\frac{\triangle E(SALES)}{\triangle ADVERT}|_{\text{Holding PRICE constant}} = \frac{\partial E(SALES)}{\partial ADVERT} = \beta_3 + 2\beta_4 ADVERT \tag{36}$$

If β_4 is negative, what does it imply for marginal effect of ADVERT on SALES?

Other Forms of Multiple Regression Model



Other Forms of Multiple Regression Model

Then we can use the "marginal analysis" (MR=MC) to estimate the optimal monthly advertising expenditure $ADVERT_0$:

- Marginal revenue from more advertising is: $MR = \beta_3 + 2\beta_4 ADVERT$;
- Marginal cost is assumed to be \$1;
- **Based on "MR=MC"**:

$$\beta_3 + 2\beta_4 ADVERT_0 = 1 \tag{37}$$

  where $ADVERT_0$ is the optimal level of advertising
- Use OLS estimates, a point estimate of $ADVERT_0$ is:

$$\widehat{ADVERT}_0 = \frac{1 - b_3}{2b_4} = 2.014 \tag{38}$$

  implying that the optimal monthly advertising expenditure is \$2014.

Other Forms of Multiple Regression Model

What if we want to give confidence interval and/or do hypothesis testing about $ADVERT_0$? We need to consider "nonlinear function" of population parameters

- Suppose $\lambda = \frac{1-\beta_3}{2\beta_4}$ and

$$\hat{\lambda} = \frac{1-b_3}{2b_4} \rightarrow \hat{\lambda} \approx \frac{\partial \hat{\lambda}}{\partial b_3} \times b_3 + \frac{\partial \hat{\lambda}}{\partial b_4} \times b_4 \tag{39}$$

- How to calculate the variance of $\hat{\lambda}$ approximately? (Delta Method)

$$Var(\hat{\lambda}) = \left(\frac{\partial \hat{\lambda}}{\partial b_3}\right)^2 Var(b_3) + \left(\frac{\partial \hat{\lambda}}{\partial b_4}\right)^2 Var(b_4) \tag{40}$$

$$+2\left(\frac{\partial \hat{\lambda}}{\partial b_3}\right)\left(\frac{\partial \hat{\lambda}}{\partial b_4}\right) Cov(b_3, b_4)$$

- The partial derivatives are:

$$\frac{\partial \hat{\lambda}}{\partial b_3} = -\frac{1}{2b_4}, \quad \frac{\partial \hat{\lambda}}{\partial b_4} = -\frac{1-b_3}{2b_4^2} \tag{41}$$

Other Forms of Multiple Regression Model

- Assume through OLS estimation, we get:

$$\underset{(Se)}{\widehat{SALES}} = \underset{(6.8)}{109.72} - \underset{(1.046)}{7.64} PRICE + \underset{(5.556)}{12.15} ADVERT - \underset{(0.941)}{2.77} ADVERT^2 \quad (42)$$

  Then we get:

$$b_3 = 12.15, \quad b_4 = -2.77, \quad \hat{Var}(b_3) = 5.556^2, \quad \hat{Var}(b_4) = 0.941^2 \quad (43)$$

  Suppose we also get $\hat{Cov}(b_3, b_4) = 3.289$, then we can calculate $\hat{Var}(\hat{\lambda})$ and $Se(\hat{\lambda})$.

- We can further report the 95% confidence interval for $ADVERT_0$:

$$[\hat{\lambda} - t_{(0.975,n-4)} Se(\hat{\lambda}), \hat{\lambda} + t_{(0.975,n-4)} Se(\hat{\lambda})] \quad (44)$$

2. Interaction Variables

- Suppose that we wish to study the effect of income and age on an individual's expenditure on pizza:

$$PIZZA = \beta_1 + \beta_2 AGE + \beta_3 INCOME + e \tag{45}$$

- The fitted model is:

$$\widehat{PIZZA} = 342.88 - 7.576AGE + 1.832INCOME \tag{46}$$

Other Forms of Multiple Regression Model

- It would seem more reasonable to assume that as a person grows older, his or her marginal propensity to spend on pizza declines (as age increases by 1 year, there is less of extra dollar to be expected to be spent on pizza);
- **The marginal effect of income depends on the age of the individual**;
- One way to account for such interactions is to include an **interaction variable** that is the product of the two variables involved.

Other Forms of Multiple Regression Model

- We will add the interaction variable $(AGE \times INCOME)$ to the regression model
- New model is:

$$PIZZA = \beta_1 + \beta_2 AGE + \beta_3 INCOME + \beta_4 (AGE \times INCOME) + e \quad (47)$$

- The implications of this revised model is:

$$\frac{\partial E(PIZZA)}{\partial AGE} = \beta_2 + \beta_4 INCOME \quad (48)$$

$$\frac{\partial E(PIZZA)}{\partial INCOME} = \beta_3 + \beta_4 AGE \quad (49)$$

- For example, we estimate $b_4 = -0.1232$

Other Forms of Multiple Regression Model

- The estimated marginal effect of age on pizza expenditure for two individuals (one with \$25000 income and one with \$90000 income) is:

$$\frac{\partial \hat{E}(PIZZA)}{\partial AGE} = -2.977 - 0.1232 INCOME$$
$$= \begin{cases} -6.06 & \text{for } INCOME = 25 \\ -14.07 & \text{for } INCOME = 90 \end{cases}$$

- We expect that an individual with \$25000 income will reduce pizza expenditure by \$6.06 per year, and the individual with \$90000 income will reduce pizza expenditure by \$14.07 per year.

Measuring Goodness-of-Fit

- The coefficient of determination is:

$$\begin{aligned} R^2 &= \frac{SSR}{SST} \\ &= \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \\ &= 1 - \frac{SSE}{SST} \\ &= 1 - \frac{\sum_{i=1}^{n}\hat{e_i}^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \end{aligned}$$

- The predicted values of y is:

$$\hat{y}_i = b_1 + b_2 x_{2i} + \cdots + b_K x_{Ki}, \quad \forall i = 1, 2, \cdots, n \tag{50}$$

Measuring Goodness-of-Fit

Assume we get $R^2 = 0.448$ for the SALES example, how to interpret that?

$$\underset{(Se)}{\widehat{SALES}} = \underset{(6.35)}{118.91} - \underset{(1.096)}{7.908}PRICE + \underset{(0.683)}{1.863}ADVERT, \quad R^2 = 0.448 \tag{51}$$

- 44.8% of the variation in sales revenue is explained by the variation in the price and the level of advertising expenditure;
- 55.2% of the variation in sales revenue is left unexplained and is due to variation in the error term;
- If adding the square of advertising to the model increases R^2 to 0.508, that is, an additional 6% of variation in sales is explained by including this variable.

Measuring Goodness-of-Fit

Comment: If the model does not include an intercept term (β_1), the measure R^2 is no longer appropriate

- If without an intercept term,

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 \neq \sum_{i=1}^{n}(\hat{y_i} - \bar{y})^2 + \sum_{i=1}^{n}\hat{e_i}^2 \tag{52}$$

  so $SST \neq SSR + SSE$

- So it is good to include intercept term in our assumed model form