Restrictions and Hypothesis Tests

Professor V.M. Auken, PhD

INTERVAL ESTIMATION: SOME BASIC IDEAS

Now in statistics the reliability of a point estimator is measured by its standard error. Therefore, instead of relying on the point estimate alone, we may construct an interval around the point estimator, say within two or three standard errors on either side of the point estimator, such that this interval has, say, 95% probability of including the true parameter value. This is roughly the idea behind interval estimation.

To be more specific, assume that we want to find out how "close" is, say, $\hat{\beta}_2$ to β_2. For this purpose we try to find out two positive numbers δ and α, the latter lying between 0 and 1, such that the probability that the **random interval** $(\hat{\beta}_2 - \delta, \hat{\beta}_2 + \delta)$ contains the true β_2 is $1 - \alpha$. Symbolically,

$$\Pr(\hat{\beta}_2 - \delta \leq \beta_2 \leq \hat{\beta}_2 + \delta) = 1 - \alpha$$

INTERVAL ESTIMATION: SOME BASIC IDEAS

Such an interval, if it exists, is known as a **confidence interval;** $1 - \alpha$ is known as the **confidence coefficient;** and α $(0 < \alpha < 1)$ is known as the **level of significance.**[2] The endpoints of the confidence interval are known as the **confidence limits** (also known as *critical* values), $\hat{\beta}_2 - \delta$ being the **lower confidence** *limit* and $\hat{\beta}_2 + \delta$ the **upper confidence** *limit*. In passing, note that in practice α and $1 - \alpha$ are often expressed in percentage forms as 100α and $100(1 - \alpha)$ percent.

[2]Also known as the **probability of committing a Type I error.** A Type I error consists in rejecting a true hypothesis, whereas a Type II error consists in accepting a false hypothesis. (This topic is discussed more fully in **App. A.**) The symbol α is also known as the **size of the (statistical) test.**

HYPOTHESIS TESTING

Two-Sided or Two-Tail Test

To illustrate the confidence-interval approach, once again we revert to the consumption–income example. As we know, the estimated marginal propensity to consume (MPC), $\hat{\beta}_2$, is 0.5091. Suppose we postulate that

$$H_0: \beta_2 = 0.3$$

$$H_1: \beta_2 \neq 0.3$$

that is, the true MPC is 0.3 under the null hypothesis but it is less than or greater than 0.3 under the alternative hypothesis. The null hypothesis is a simple hypothesis, whereas the alternative hypothesis is composite; actually it is what is known as a **two-sided hypothesis.** Very often such a two-sided alternative hypothesis reflects the fact that we do not have a strong a priori or theoretical expectation about the direction in which the alternative hypothesis should move from the null hypothesis.

HYPOTHESIS TESTING

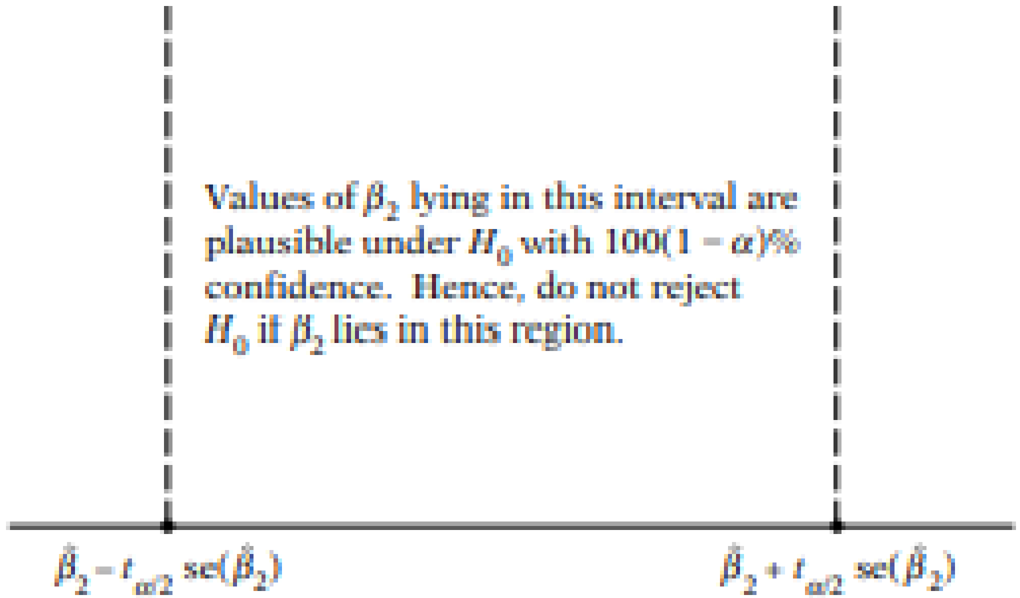
Is the observed $\hat{\beta}_2$ compatible with H_0?

We know that in the long run intervals like (0.4268, 0.5914) will contain the true

β_2 with 95 percent probability.

Consequently, in the long run (i.e., repeated sampling) such intervals provide a range or limits within which the true β2 may lie with a confidence coefficient of, say, 95%. Thus, the confidence interval provides a set of plausible null hypotheses. Therefore, if β2 under H0 falls within the 100(1 − α)% confidence interval, we do not reject the null hypothesis; if it lies outside the interval, we may reject it.7 This range is illustrated schematically in Figure.

HYPOTHESIS TESTING



A 100(1 − α)% confidence interval for β_2.

HYPOTHESIS TESTING

Decision Rule: Construct a $100(1-\alpha)\%$ confidence interval for β_2. If the β_2 under H_0 falls within this confidence interval, do not reject H_0, but if it falls outside this interval, reject H_0.

Following this rule, for our hypothetical example, H0: ß2 = 0.3 clearly lies outside the 95% confidence interval given in $0.4268 \le \beta_2 \le 0.5914$

Therefore, we can reject the hypothesis that the true MPC is 0.3, with 95% confidence. If the null hypothesis were true, the probability of our obtaining a value of MPC of as much as 0.5091 by sheer chance or fluke is at the most about 5%, a small probability.

In statistics, when we reject the null hypothesis, we say that our finding is statistically significant. On the other hand, when we do not reject the null hypothesis, we say that our finding is **not statistically significant**.

Consistency

This is the property of *strong consistency:* the estimator converges almost surely to the true value. If we has used a weak LLN (defined in terms of convergence in probability), we would have *(simple, weak) consistency.*

- The consistency proof does not use the normality assumption.

Joint confidence region for β_0 and β_1

A joint confidence region for β_0 and β_1 can also be found. Such region will provide a $100(1-\alpha)\%$ confidence that both the estimates of β_0 and β_1 are correct. Consider the centered version of the linear regression model

$$y_i = \beta_0^* + \beta_1(x_i - \bar{x}) + \varepsilon_i$$

where $\beta_0^* = \beta_0 + \beta_1\bar{x}$.

The least squares estimators of β_0^* and β_1 are $b_0^* = \bar{y}$ and $b_1 = \frac{s_{xy}}{s_{xx}}$, respectively.

Using the results that

$$E(b_0^*) = \beta_0^*,$$

$$E(b_1) = \beta_1,$$

$$Var(b_0^*) = \frac{\sigma^2}{n},$$

$$Var(b_1) = \frac{\sigma^2}{s_{xx}}.$$

When σ^2 is known, then the statistic

$$\frac{b_0^* - \beta_0^*}{\sqrt{\frac{\sigma^2}{n}}} \sim N(0,1)$$

and

$$\frac{b_1 - \beta_1}{\sqrt{\frac{\sigma^2}{s_{xx}}}} \sim N(0,1).$$

Joint confidence region for β_0 and β_1

Moreover, both the statistics are independently distributed. Thus

$$\left(\frac{b_0^* - \beta_0^*}{\sqrt{\frac{\sigma^2}{n}}}\right)^2 \sim \chi_1^2$$

and

$$\left(\frac{b_1 - \beta_1}{\sqrt{\frac{\sigma^2}{s_{xx}}}}\right)^2 \sim \chi_1^2$$

are also independently distributed because b_0^* and b_1 are independently distributed. Consequently, the sum of these two

$$\frac{n(b_0^* - \beta_o^*)^2}{\sigma^2} + \frac{s_{xx}(b_1 - \beta_1)^2}{\sigma^2} \sim \chi_2^2.$$

Joint confidence region for β_0 and β_1

Since

$$\frac{SS_{res}}{\sigma^2} \sim \chi^2_{n-2}$$

and SS_{res} is independently distributed of b_0^* and b_1, so the ratio

$$\frac{\left(\frac{n(b_0^* - \beta_0^*)^2}{\sigma^2} + \frac{s_{xx}(b_1 - \beta_1)^2}{\sigma^2}\right)\Big/ 2}{\left(\frac{SS_{res}}{\sigma^2}\right)\Big/ (n-2)} \sim F_{2,n-2}.$$

Substituting $b_0^* = b_0 + b_1\bar{x}$ and $\beta_0^* = \beta_0 + \beta_1\bar{x}$, we get

$$\left(\frac{n-2}{2}\right)\left[\frac{Q_f}{SS_{res}}\right]$$

where

$$Q_f = n(b_0 - \beta_0)^2 + 2\sum_{i=1}^{n} x_i(b_0 - \beta_1)(b_1 - \beta_1) + \sum_{i=1}^{n} x_i^2(b_1 - \beta_1)^2.$$

Joint confidence region for β_0 and β_1

Since

$$P\left[\left(\frac{n-2}{2}\right)\frac{Q_f}{SS_{res}} \le F_{2,n-2}\right] = 1-\alpha$$

holds true for all values of β_0 and β_1, so the $100(1-\alpha)\%$ confidence region for β_0 and β_1 is

$$\left(\frac{n-2}{2}\right)\frac{Q_f}{SS_{res}} \le F_{2,n-2;\alpha}.$$

This confidence region is an ellipse which gives the $100(1-\alpha)\%$ probability that β_0 and β_1 are contained simultaneously in this ellipse.

Analysis of variance

The technique of analysis of variance is usually used for testing the hypothesis related to equality of more than one parameters, like population means or slope parameters.

It is more meaningful in case of multiple regression model when there are more than one slope parameters.

This technique is discussed and illustrated here to understand the related basic concepts and fundamentals which will be used in developing the analysis of variance in the next module in multiple linear regression model where the explanatory variables are more than two.

Analysis of variance

A test statistic for testing $H_0 : \beta_1 = 0$ can also be formulated using the analysis of variance technique as follows.

On the basis of the identity $y_i - \hat{y}_i = (y_i - \overline{y}) - (\hat{y}_i - \overline{y})$,

the sum of squared residuals is

$$S(b) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$= \sum_{i=1}^{n} (y_i - \overline{y})^2 + \sum_{i=1}^{n} (\hat{y}_i - \overline{y}_i)^2 - 2\sum_{i=1}^{n} (y_i - \overline{y})(\hat{y}_i - \overline{y}).$$

Further consider $\sum_{i=1}^{n} (y_i - \overline{y})(\hat{y}_i - \overline{y}) = \sum_{i=1}^{n} (y_i - \overline{y}) b_1 (x_i - \overline{x})$

$$= b_1^2 \sum_{i=1}^{n} (x_i - \overline{x})^2$$

$$= \sum_{i=1}^{n} (\hat{y}_i - \overline{y})^2.$$

Analysis of variance

Thus we have $\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2.$

The term $\sum_{i=1}^{n}(y_i-\bar{y})^2$ i

If all observations y_i are located on a straight line, then in this case

of squares of *y* (i.e., $SS_{corrected}$) or total sum of squares denoted as s_{yy}.

The term $\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$ describes the deviation: observation minus predicted value, viz., the residual sum of squares, i.e.:

$$SS_{res}=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$$

whereas the term $\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$ describes the proportion of variability explained by regression,

$$SS_{reg}=\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2.$$

Analysis of variance

If all observations y_i are located on a straight line, then in this case

$$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 0 \text{ and thus } SS_{corrected} = SS_{reg}.$$

Note that SS_{reg} is completely determined by b_1 and so has only one degrees of freedom. The total sum of squares $s_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$ has $(n - 1)$ degrees of freedom due to constraint $\sum_{i=1}^{n}(y_i - \bar{y}) = 0$ and SS_{res} has $(n - 2)$ degrees of freedom as it depends on b_0 and b_1.

All sums of squares are mutually independent and distributed as χ^2_{df} with *df* degrees of freedom if the errors are normally distributed.

Analysis of variance

The mean square due to regression is

$$MS_{reg} = \frac{SS_{reg}}{1}$$

and mean square due to residuals is

$$MSE = \frac{SS_{res}}{n-2}.$$

The test statistic for testing $H_0 : \beta_1 = 0$ is

$$F_0 = \frac{MS_{reg}}{MSE}.$$

If $H_0 : \beta_1 = 0$ is true, then MS_{reg} and MSE are independently distributed and thus $F_0 \sim F_{1,n-2}$.

The decision rule for $H_1 : \beta_1 \neq 0$ is to reject H_0 if $F_0 > F_{1,n-2;1-\alpha}$

at α level of significance. The test procedure can be described in an Analysis of Variance table.

Analysis of variance

Analysis of variance for testing $H_0 : \beta_1 = 0$

Source of variation	Sum of squares	Degrees of freedom	Mean square	F
Regression	SS_{reg}	1	MS_{reg}	MS_{reg} / MSE
Residual	SS_{res}	$n-2$	MSE	
Total	s_{yy}	$n-1$		

Some other forms of SS_{reg}, SS_{res} and s_{yy} can be derived as follows:

The sample correlation coefficient then may be written as

$$r_{xy} = \frac{s_{xy}}{\sqrt{s_{xx}}\sqrt{s_{yy}}}.$$

Moreover, we have $b_1 = \frac{s_{xy}}{s_{xx}} = r_{xy}\sqrt{\frac{s_{yy}}{s_{xx}}}.$

Analysis of variance

The estimator of σ^2 in this case may be expressed as

$$s^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2$$

$$= \frac{1}{n-2} SS_{res}.$$

Various alternative formulations for SS_{res} are in use as well:

$$SS_{res} = \sum_{i=1}^{n}[y_i - (b_0 + b_1 x_i)]^2$$

$$= \sum_{i=1}^{n}[(y_i - \bar{y}) - b_1(x_i - \bar{x})]^2$$

$$= s_{yy} + b_1^2 s_{xx} - 2b_1 s_{xy}$$

$$= s_{yy} - b_1^2 s_{xx}$$

$$= s_{yy} - \frac{(s_{xy})^2}{s_{xx}}.$$

Analysis of variance

Using this result, we find that $SS_{corrected} = s_{yy}$

and

$$\begin{aligned} SS_{reg} &= s_{yy} - SS_{res} \\ &= \frac{(s_{xy})^2}{s_{xx}} \\ &= b_1^2 s_{xx} \\ &= b_1 s_{xy}. \end{aligned}$$

Goodness of fit of regression

It can be noted that a fitted model can be said to be good when residuals are small. Since SS_{res} is based on residuals, so a measure of quality of fitted model can be based on SS_{res}. When intercept term is present in the model, a measure of goodness of fit of the model is given by

$$R^2 = 1 - \frac{SS_{res}}{s_{yy}} = \frac{SS_{reg}}{s_{yy}}.$$

This is known as the **coefficient of determination**. This measure is based on the concept that how much variation in *y*'s stated by s_{yy} is explainable by SS_{reg}. and how much unexplainable part is contained in SS_{res}. The ratio SS_{reg} / s_{yy} describes the proportion of variability that is explained by regression in relation to the total variability of *y*. The ratio SS_{res} / s_{yy} describes the proportion of variability that is not covered by the regression.

Goodness of fit of regression

It can be seen that

$$R^2 = r_{xy}^2$$

where r_{xy} is the simple correlation coefficient between x and y. Clearly $0 \le R^2 \le 1$, so a value of R^2 closer to one indicates the better fit and value of R^2 closer to zero indicates the poor fit.

QUESTIONS!