Linear Statistical Modeling Methods with SAS

Simultaneous Inferences for SLR Models

Xuemao Zhang
East Stroudsburg University

February 21, 2024

Outline

- Joint Estimation of β_0 and β_1
- SLR with Fixed Intercept

Joint Estimation of β_0 and β_1

- Joint Estimation of β_0 and β_1 is often very important in applications. The joint estimation allows the analyst to gain some insight in the precision of the point estimators $\widehat{\beta}_0$ and $\widehat{\beta}_1$.
- When the $100(1-\alpha)\%$ confidence region is found, the implication is that with *repeated experiments*, and thus repeated computations of the confidence region, $100(1-\alpha)\%$ of such computed regions will contain both β_0 and β_1. One way to construct a confidence region is to construct two separate confidence intervals for β_0 and β_1 respectively such that the joint coverage is at least $100(1-\alpha)\%$.
- It is of interest to have confidence intervals of the type

$$\widehat{\beta}_0 \pm \text{constant} \times (\text{standard error of } \widehat{\beta}_0)$$

$$\widehat{\beta}_1 \pm \text{constant} \times (\text{standard error of } \widehat{\beta}_1)$$

  and have the constant determined so that we have confidence at least $100(1-\alpha)\%$ that β_0 and β_1 are simultaneously in the interval.

Bonferroni Joint Confidence Intervals

The Bonferroni procedure for developing joint confidence intervals for β_0 and β_1, with a specified **family confidence coefficient** is very simple: each **statement confidence coefficient** is adjusted to be higher than $1-\alpha$ so that the family confidence coefficient is at least $1-\alpha$.

Definition. We say that a confidence interval **covers**, if it contains the true parameter.

Let $P(B_0)$ be the probability that the confidence interval of β_0 covers and let $P(B_1)$ be the probability that the confidence interval of β_1 covers. (Note $\overline{B_0}$ is the event that the CI for β_0 does not cover and $\overline{B_1}$ is the event that the CI for β_1 does not cover).

$$\begin{aligned}
P(B_0 \cap B_1) =& P(B_0) + P(B_1) - P(B_0 \cup B_1)\\
=& 1 - P(\overline{B_0}) + 1 - P(\overline{B_1}) - P(B_0 \cup B_1)\\
=& 1 - P(\overline{B_0}) - P(\overline{B_1}) - P(\overline{B_0 \cup B_1})\\
\geq& 1 - \left[P(\overline{B_0}) + P(\overline{B_1})\right]
\end{aligned}$$

Bonferroni Joint Confidence Intervals

Conclusion:

$$\boxed{P(B_0 \cap B_1) \geq 1 - P(\overline{B_0}) - P(\overline{B_1})}$$

Now, $P(\overline{B_0}) = P(\overline{B_1}) = \alpha$, and thus we have

$$\boxed{P(B_0 \cap B_1) \geq 1 - 2\alpha.}$$

Now if the confidence coefficient for the confidence intervals of β_0 and β_1 are both 95% (1-0.05), then the confidence coefficient for joint estimation of β_0 and β_1 will be at least $1 - 2 \times 0.05 = 90\%$. Thus, at least 90% of such pairs of intervals will simultaneously cover the two parameters.

Conclusion: A $100(1-\alpha)\%$ joint confidence interval of β_0 and β_1 is given by

$$\boxed{\begin{array}{lll} \widehat{\beta}_0 \pm & t_{\alpha/4, n-2} \times (\text{standard error of} & \widehat{\beta}_0) \\ \widehat{\beta}_1 \pm & t_{\alpha/4, n-2} \times (\text{standard error of} & \widehat{\beta}_1) \end{array}}$$

Bonferroni Joint Confidence Intervals

The Bonferroni procedure is a general one that can be applied in many cases.

General Bonferroni Inequality: For k such events A_i, $i = 1, \ldots, k$,

$$P(A_1 \cap A_2 \cap \cdots A_k) \geq 1 - \sum_{i=1}^{k} P(\overline{A_i}).$$

Thus if $P(\overline{A_i}) = \alpha/k$, we get

$$P(A_1 \cap A_2 \cap \cdots A_k) \geq 1 - \sum_{i=1}^{k} \frac{\alpha}{k} = 1 - \alpha.$$

Example: Bonferroni Joint Confidence Intervals

Example (Exercise 1.22). Sixteen batches of the plastic were made, and from each batch one test item was molded and the hardness measured at some specific point in time. The results are shown in PlasticHardness.txt; X is elapsed time in hours, and Y is hardness in Brinell units. Obtain 90% Bonferroni joint confidence intervals for β_0 and β_1.

```
data plastic;
input y x @@;
datalines;
199.0  16.0  205.0  16.0  196.0  16.0
200.0  16.0  218.0  24.0  220.0  24.0
215.0  24.0  223.0  24.0  237.0  32.0
234.0  32.0  235.0  32.0  230.0  32.0
250.0  40.0  248.0  40.0  253.0  40.0
246.0  40.0
;
run;
```

Example: Bonferroni Joint Confidence Intervals

```
proc reg data=plastic;
model y = x;
ods output ParameterEstimates=PE;
run;

proc print data=PE;
run;
```

Example: Bonferroni Joint Confidence Intervals

```
data _null_; /*Starts a data step that does not create a new dataset
   set PE;
   if _n_ = 1 then do;
      call symput('Int', put(estimate, BEST6.));
      /*Assigns the value of the estimate variable to the macro vari
      call symput('Int_stderr', put(stderr, BEST6.));
   end;
   else do;
      call symput('Slope', put(estimate, BEST6.));
      call symput('Slope_stderr', put(stderr, BEST6.));
   end;
run
```

Example: Bonferroni Joint Confidence Intervals

- macro variables need to be resolved using double ampersands && within a data step.

```
data intervals;
tvalue = tinv(1-0.025,14);
b0lo=&&Int-(tvalue*&&Int_stderr); /*lower bound for b0*/
b0up=&&Int+(tvalue*&&Int_stderr); /*upper bound for b0*/
b1lo=&&Slope-(tvalue*&&Slope_stderr); /*lower bound for b1*/
b1up=&&Slope+(tvalue*&&Slope_stderr); /*upper bound for b1*/
run;

proc print data=intervals;
var b0lo b0up b1lo b1up;
run;
```

Regression with Fixed Intercept

The SLR model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, i = 1, \ldots, n$$

presumes that the data analyst requires the estimation of both slope and intercept from the data. A lot of times, the intercept β_0 is a fixed known constant. Then only β_1 needs to be estimated.

By the method of $\{$**least squares**$\}$**, we need to find $\hat{\beta}_1$ to minimize**

$$\textbf{SSE}(\hat{\beta}_1) = \sum_{i=1}^{n} (Y_i - (\beta_0 + \hat{\beta}_1 x_i))^2.$$

Using calculus, $\hat{\beta}_1$ must satisfy

$$\frac{\partial}{\partial \hat{\beta}_1} \left[\sum_{i=1}^{n} (y_i - \beta_0 - \hat{\beta}_1 x_i)^2 \right] = 0$$

which reduces to

$$\sum_{i=1}^{n} (y_i - \beta_0 - \hat{\beta}_1 x_i) x_i = 0.$$

Therefore, the estimator $\hat{\beta}_1$ is

$$\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i - \beta_0 \sum_{i=1}^{n} x_i}{\sum x_i^2}}$$

Regression with Fixed Intercept

For the special case of simple linear regression through the origin, the estimator $\hat{\beta}_1$ reduces to

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} y_i x_i}{\sum x_i^2} = \sum_{i=1}^{n} \frac{x_i}{\sum x_i^2} y_i.$$

Under the assumptions of uncorrelated random errors (independence) and homogeneous variance,

$$E(\hat{\beta}_1) = \frac{\sum_{i=1}^{n} (\beta_0 + \beta_1 x_i) x_i - \beta_0 \sum_{i=1}^{n} x_i}{\sum x_i^2} = \frac{\sum_{i=1}^{n} \beta_1 x_i^2}{\sum x_i^2} = \beta_1;$$

$$Var(\hat{\beta}_1) = \sigma^2 \sum_{i=1}^{n} \frac{x_i^2}{\left(\sum x_i^2\right)^2} = \frac{\sigma^2}{\sum x_i^2}.$$

Therefore, if we assume that Y_i are independent $N(0, \sigma^2)$ normal random variables, $i = 1, \ldots, n$, then

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum x_i^2}\right).$$

Regression with Fixed Intercept

Therefore,

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma/\sqrt{\sum x_i^2}} \sim N(0,1).$$

And thus

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{MSE}/\sqrt{\sum x_i^2}} \sim t_{\mathbf{n-1}},$$

where the expression of MSE is

$$\boxed{MSE = \frac{\sum_{i=1}^n (y_i - \widehat{y}_i)^2}{n-1} = \frac{\sum_{i=1}^n (y_i - \beta_0 - \hat{\beta}_1 x_i)^2}{n-1}}$$

since the fitted values are

$$\widehat{y}_i = \beta_0 + \hat{\beta}_1 x_i, i = 1, \ldots, n.$$

ANOVA for Fixed Intercept Case

Consider the model

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, i = 1, \dots, n,$$

where β_0 is a known constant.

Then

$$\boxed{\sum_{i=1}^{n}(y_i - \beta_0)^2 = \sum_{i=1}^{n}(\widehat{y}_i - \beta_0)^2 + \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2.}$$

That is, we are dealing with a partition of variation around β_0 rather than around $\overline{y}$.

The **fitted values** are

$$\widehat{y}_i = \beta_0 + \hat{\beta}_1 x_i, \quad i = 1, \dots, n.$$

Thus the component

$$\boxed{\sum_{i=1}^{n}(\widehat{y}_i - \beta_0)^2 = \hat{\beta}_1^2 \sum_{i=1}^{n} x_i^2}$$

which corresponds to SSR.

ANOVA for Fixed Intercept Case

- Recall that

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{MSE}/\sqrt{\sum x_i^2}} \sim t_{\mathbf{n-1}},$$

  where $MSE = \dfrac{\sum_{i=1}^{n}(y_i - \beta_0 - \hat{\beta}_1 x_i)^2}{n-1}$.

- Under the null hypothesis $H_0 : \beta_1 = 0$,

$$\boxed{\frac{\hat{\beta}_1^2 \sum x_i^2}{MSE} \sim F_{\mathbf{1,n-1}}.}$$

Again, we observe that the square root of the F-statistic is the t-statistic.

Example: Regression with Fixed Intercept

The following SAS code fit a regression model without intercept.

```
proc reg data=plastic;
model y = x/noint clb alpha=0.01;
run;
```

License