Stat 206: Linear Models

Lecture 4

October 7, 2019

ReCap: Sampling Distributions of LS Estimators

Under the Normal error model:

- $\hat{\beta}_0, \hat{\beta}_1$ are normally distributed:

$$\hat{\beta}_0 \sim N(\beta_0, \sigma^2\{\hat{\beta}_0\}), \quad \hat{\beta}_1 \sim N(\beta_1, \sigma^2\{\hat{\beta}_1\}).$$

- SSE/σ^2 follows a χ^2 distribution with $n-2$ degrees of freedom, denoted by $\chi^2_{(n-2)}$.
- Moreover, SSE is independent with both $\hat{\beta}_0$ and $\hat{\beta}_1$ (because residuals e_i's are independent with $\hat{\beta}_0$ and $\hat{\beta}_1$).

Recap: Confidence Interval

$(1-\alpha)$-Confidence interval of β_1:

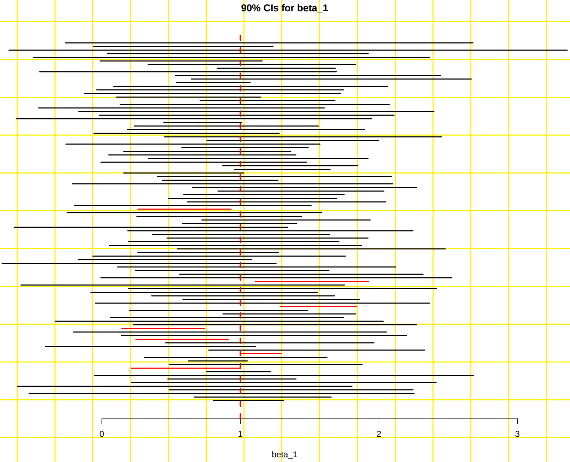
$$\hat{\beta}_1 \pm t(1-\alpha/2; n-2)s\{\hat{\beta}_1\},$$

where $t(1-\alpha/2; n-2)$ is the $(1-\alpha/2)$th percentile of $t_{(n-2)}$.

How to construct confidence intervals for β_0?

Interpretation of Confidence Intervals

Figure: A Simulation Study



Heights

- Recall $n = 928$, $\overline{X} = 68.316$, $\sum_i X_i^2 = 4334058$, $\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n} X_i^2 - n(\overline{X})^2 = 3038.761$. Also

$$\hat{\beta}_0 = 24.54, \ \ \hat{\beta}_1 = 0.637, \ \ MSE = 5.031.$$

So

$$s\{\hat{\beta}_1\} = \sqrt{\frac{5.031}{3038.761}} = 0.0407.$$

- 95%-confidence interval of β_1:

$$\begin{aligned} 0.637 \pm t(0.975; 926) \times 0.0407 &= 0.637 \pm 1.963 \times 0.0407 \\ &= [0.557, 0.717]. \end{aligned}$$

We are 95% confident that the regression slope is in between 0.557 and 0.717.

T-tests

- Null hypothesis: $H_0 : \beta_1 = \beta_1^{(0)}$, where $\beta_1^{(0)}$ is a given constant.
- T-statistic:

- **Null distribution** of the T-statistic:

$$\text{Under } H_0 : \beta_1 = \beta_1^{(0)}, \quad T^* \sim t_{(n-2)}.$$

Decision rule at significance level α.

- **Two-sided alternative** $H_a : \beta_1 \neq \beta_1^{(0)}$: Reject H_0 if and only if $|T^*| > t(1 - \alpha/2; n - 2)$, or equivalently, reject H_0 if and only if pvalue$:= P(|t_{(n-2)}| > |T^*|) < \alpha$.
- **Left-sided alternative** $H_a : \beta_1 < \beta_1^{(0)}$: Reject H_0 if and only if $T^* < t(\alpha; n - 2)$, or equivalently, reject H_0 if and only if pvalue$:= P(t_{(n-2)} < T^*) < \alpha$.
- **Right-sided alternative** $H_a : \beta_1 > \beta_1^{(0)}$: Reject H_0 if and only if $T^* > t(1 - \alpha; n - 2)$, or equivalently, reject H_0 if and only if pvalue$:= P(t_{(n-2)} > T^*) < \alpha$.

The decision rule depends on the form of

Heights

Test whether there is a linear association between parent's height and child's height. Use significance level $\alpha = 0.01$.

- The hypotheses: $H_0 : \beta_1 = 0$ *vs.* $H_a : \beta_1 \neq 0$.
- T statistic: $T^* = \frac{\hat{\beta}_1 - 0}{s\{\hat{\beta}_1\}} = \frac{0.637}{0.0407} = 15.7$.
- Critical value: $t(1 - 0.01/2; 928 - 2) = 2.58$. Since the observed $T^* = |15.7| > 2.58$, reject the null hypothesis at level 0.01.
- Or the pvalue $= P(|t_{(926)}| > |15.7|) \approx 0$. Since *pvalue* $< \alpha = 0.01$, reject the null hypothesis at level 0.01.
- Conclude that there is a significant association between parent's height and child's height.

Estimation of Mean Response

Given $X = X_h$, the mean response is $E(Y_h) =$

- An unbiased point estimator for $E(Y_h)$ is :

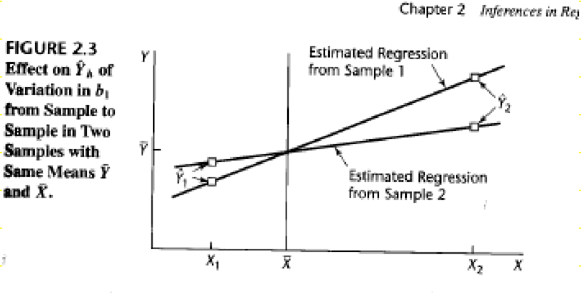
- Variance of $\widehat{Y}_h$ is:

  Notes: Use the fact that $\overline{Y}$ and $\hat{\beta}_1$ are uncorrelated.

- The larger the sample size and/or the larger the dispersion of X_is, the

  the variance of $\widehat{Y}_h$.

Figure: Effects of the distance of X_h from $\overline{X}$ on variability of $\widehat{Y}_h$.



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

The further is X_h from $\overline{X}$, the

is the variance of $\widehat{Y}_h$: The variability in the estimated slope $\hat{\beta}_1$ has a effect on $\widehat{Y}_h$ when X_h is further away from the sample mean $\overline{X}$.

- Standard error of $\widehat{Y}_h$:

$$s\{\widehat{Y}_h\} = \sqrt{MSE\left[\frac{1}{n} + \frac{(X_h - \overline{X})^2}{\sum_{i=1}^n (X_i - \overline{X})^2}\right]}.$$

- Under the Normal error model, $\widehat{Y}_h$ is normally distributed.
  - Studentized quantity:

$$\frac{\widehat{Y}_h - E(Y_h)}{s(\widehat{Y}_h)} \sim t_{(n-2)}.$$

  - $(1-\alpha)$- C.I.

$$\widehat{Y}_h \pm t(1-\alpha/2; n-2)s(\widehat{Y}_h).$$

  - The half-width of $(1-\alpha)$- C.I., $t(1-\alpha/2; n-2)s(\widehat{Y}_h)$, with the confidence coefficient $(1-\alpha)$ and the standard error $s(\widehat{Y}_h)$.

Heights

Estimate the average height of children of 70*in* tall parents.

- Recall: $n = 928$, $\overline{X} = 68.316$, $\sum_{i=1}^{n}(X_i - \overline{X})^2 = 3038.761$, $\widehat{E(Y)} = 24.54 + 0.637X$ and $MSE = 5.031$.
- $\widehat{Y}_h = 24.54 + 0.637 \times 70 = 69.2$.
- Standard error:

$$s\{\widehat{Y}_h\} = \sqrt{5.031 \times \left\{\frac{1}{928} + \frac{(70 - 68.316)^2}{3038.761}\right\}} = 0.1.$$

- 95%-confidence interval of $E(Y_h)$:

$$69.2 \pm 1.8831 \times 0.1 = [68.96, 69.35], \quad t(0.975; 926) = 1.8831.$$

- We are 95% confident that the average height of children of 70*in* parents is between $[68.96in, 69.35in]$.

Prediction of New Observation

Predict a $Y_{h(new)}$ of the response variable corresponding to a given level of the predictor variable $X = X_h$.

- $Y_{h(new)} = \beta_0 + \beta_1 X_h + \epsilon_h$.
  - This is a new observation, so ϵ_h is assumed to be with ϵ_is.
  - Consequently, $Y_{h(new)}$ is with the observed Y_is.
- The **predicted value** for $Y_{h(new)}$

Distinction between prediction and mean estimation.

- $Y_{h(new)}$ is a “moving target” as it is a random variable. On the contrary, $E(Y_h)$ is a fixed non-random quantity.
- There are sources in of variations in the prediction process:

Note the difference between $s^2\{\widehat{Y_h}\}$ and $s^2\{pred_h\}$.

Prediction Intervals

- Studentized quantity:

$$\frac{\widehat{Y}_h - Y_{h(new)}}{s(pred_h)}.$$

- Under the Normal error model, it follows a $t_{(n-2)}$ distribution.
- $(1-\alpha)-$ prediction interval of $Y_{h(new)}$:

$$\widehat{Y}_h \pm t(1-\alpha/2; n-2)s(pred_h).$$

- Prediction interval is

  than the corresponding confidence interval of the mean response.
- With sample size becomes very large, the width of the confidence interval tend to

  but this would not happen for the prediction interval.

Heights

What would be the predicted height of the child of a 70*in* tall couple?

- Predicted height: $\widehat{Y}_h = 24.54 + 0.637 \times 70 = 69.2$. Standard error:

$$s\{pred_h\} = \sqrt{5.031 \times \left\{\mathbf{1} + \frac{1}{928} + \frac{(70 - 68.316)^2}{3038.761}\right\}} = 2.25.$$

- 95% prediction interval:

$$69.2 \pm 1.8831 \times 2.25 = [64.75, 73.56], \quad t(0.975; 926) = 1.8831.$$

- We are 95% confident that the child's height will be in between $[64.75in, 73.56in]$.

Extrapolation

Extrapolation occurs when predicting the response variable for values of the predictor variable lying

the range of the observed data.

- Every model has a range of validity. Particularly, a model may be inappropriate when it is extended outside of the range of the observations upon which it was built.
- Extrapolations are often much

  reliable than interpolation and need to be handled with caution, even though they can be of more interests to us (e.g. fortune telling).
- In the Heights example: Extrapolation would happen if we use the fitted regression line to predict heights of children of

  parents.

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;

Partition of Total Deviations

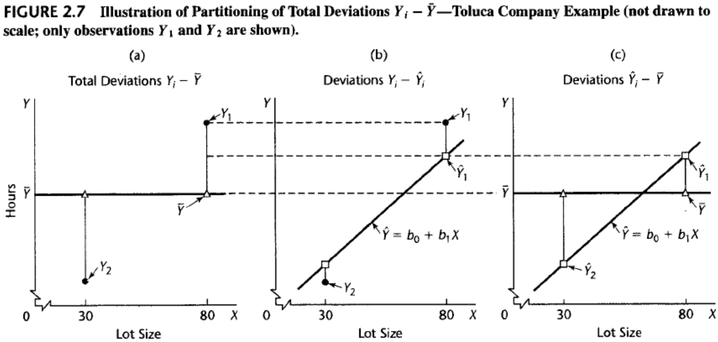
- **Total deviations:** Difference between Y_i and the sample mean $\overline{Y}$:

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

- 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*.

Figure: Partition of total deviation.



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

Decomposition of Total Variation

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.

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

  This is the variation of the fitted values around the sample mean. The the 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 $\frac{1}{n}\sum_{i=1}^{n}\widehat{Y}_i$?

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:

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

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.*

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	$SSR = 1234$	$\text{d.f.}(SSR) = 1$	$MSR = 1234$	$F^* = MSR/MSE = 245$
Error	$SSE = 4659$	$\text{d.f.}(SSE) = 926$	$MSE = 5.03$	
Total	$SSTO = 5893$	$\text{d.f.}(SSTO) = 927$	$MSTO = 6.36$	

- 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:

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