STAC67: Regression Analysis

Lecture 4

Sohee Kang

Jan. 21, 2021

Properties of Least Squares Estimates

- $$\hat{\beta}_1 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} = \frac{\sum(X_i - \bar{X})Y_i}{\sum(X_i - \bar{X})^2}$$

  $$= \sum k_i Y_i \quad , \quad k_i = \frac{X_i - \bar{X}}{\sum(X_i - \bar{X})^2} = \frac{X_i - \bar{X}}{S_{XX}}$$

- The constants k_i have several interesting properties:

  1. $\sum_{i=1}^n k_i = 0$
  2. $\sum_{i=1}^n k_i X_i = \frac{\sum(X_i - \bar{X})X_i}{S_{xx}} = \frac{\sum(X_i - \bar{X})X_i}{S_{XX}} = \frac{S_{XX}}{S_{XX}} = 1$
  3. $\sum_{i=1}^n k_i^2 = \frac{\sum(X_i - \bar{X})^2}{S_{XX}^2} = \frac{1}{S_{XX}}$

Chapter 2: Inferences in Regression

Sampling Distribution of Least Square Estimators

Everything you obtain from sample is a RV.

- Expected value of LSE

1. $E(\hat{\beta}_1) = E\left(\sum k_i Y_i\right) = \sum k_i E(Y_i) = \sum k_i (\beta_0 + \beta_1 X_i)$
   $= \beta_0 \sum k_i + \beta_1 \sum k_i X_i = 0 + \beta_1 = \beta_1$

2. $E(\hat{\beta}_0) = E(\bar{Y} - \beta_1 \bar{X}) = E(\bar{Y}) - \beta_1 E(\bar{X}) =$
   $= \beta_0 + \beta_1 \bar{X} - \beta_1 \bar{X} = \beta_0$

3. $E(\hat{\sigma}^2) = E\left(\frac{\sum e_i^2}{n-2}\right) = \sigma^2$

Sampling Distribution of Least Square Estimators

- Variance of least squre estimators

1. $Var(\hat{\beta}_1) = Var(\sum k_i Y_i) = \sum k_i^2 Var(Y_i) = \frac{\sigma^2}{S_{xx}}$

2. $Var(\hat{\beta}_0) = Var(\bar{Y} - \hat{\beta}_1 \bar{X}) = Var(\bar{Y}) + \bar{X}^2 Var(\hat{\beta}_1)$

$$= \frac{\sigma^2}{n} + \bar{X}^2 \frac{\sigma^2}{S_{xx}}$$

$$= \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}} \right)$$

$\sigma^2 \approx \frac{SSE}{n-2}$ when not known.

Sampling Distribution of Least Square Estimators

- Theorem 2

  1. $b_0 = E(\hat{\beta}_0) = \beta_0, \quad b_1 = E(\hat{\beta}_1) = \beta_1$

  2. $$Var(\hat{\beta}_0) = \left(\frac{1}{n} + \frac{\bar{X}^2}{S_{xx}}\right)\sigma^2, \quad Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$$

- The variances in theorem 2, depends on the unknown value of σ^2. If we replace σ^2 with unbiased estimator, $\hat{\sigma}^2$ and take the squre root, we get the standard error of the estimators:

$$\hat{\sigma}^2 = \frac{SSE}{n-2} \qquad SE(\beta_1) = \sqrt{\frac{\sigma^2}{S_{xx}}}$$

Standard error

$$SE(\hat{\beta}_0) = S_{b_0} = \sqrt{\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}\,\hat{\sigma}$$

$$SE(\hat{\beta}_1) = S_{b_1} = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} = \frac{\hat{\sigma}}{\sqrt{S_{xx}}}$$

Inference about Regression Parameters

- Simple Linear Model with **Normal Errors**

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

where

- β_0 and β_1 are parameters
- ϵ_i are i.i.d with normal distribution with mean 0 and variance, σ^2.

$$\epsilon_i \sim N(0, \sigma^2)$$

- Under this model,

$$Y_i | X_i \sim N(\beta_0 + \beta_1 X_i, \sigma^2)$$

Inference about Regression Parameters

- Inference about β_1

$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right) \sim \frac{\hat{\beta}_1 - \beta_1}{\frac{\sigma}{S_{xx}}} \sim N(0,1)$

- σ is unknown, and must be estimated.

$$T = \frac{\hat{\beta}_1 - \beta_1}{\left(\frac{\hat{\sigma}}{\sqrt{S_{xx}}}\right)} \sim t(n-2)$$

- Review: the Student's t-distribution with k degrees of freedom can be defined as the random variable T

$$T = \frac{Z}{\sqrt{V/k}} \sim t_{(k)},$$

where,

① $Z \sim N(0,1)$

② $V \sim \chi^2_{(k)}$

③ Z, V are independent

Sampling ditribution of β_1

- How to prove

$$\frac{\hat{\beta}_1 - \beta_1}{S_{\hat{\beta}_1}} \sim t(n-2)?$$

1. $$\frac{\hat{B}_1 - B_1}{SE(\hat{B}_1)} = \frac{\hat{B}_1 - B_1}{\sqrt{Var(\hat{B}_1)}} \Bigg/ \sqrt{\frac{SE(\hat{B}_1)^2}{Var(\hat{B}_1)}}$$

Thm: $\frac{SSE}{\sigma^2} \sim \chi^2(n-2)$ and is independent of $\hat{B}_0, \hat{B}_1$

2. $$\frac{\hat{B}_1 - B_1}{\sqrt{Var(\hat{B}_1)}} \sim N(0,1) \quad ① \checkmark$$

3. $$\frac{(n-2)SE(\hat{B}_1)^2}{Var(\hat{B}_1)} \sim \chi^2(n-2) \quad ② \checkmark$$

4. Conclude ①, ② are independent. $\checkmark$

Inference about Regression Parameters

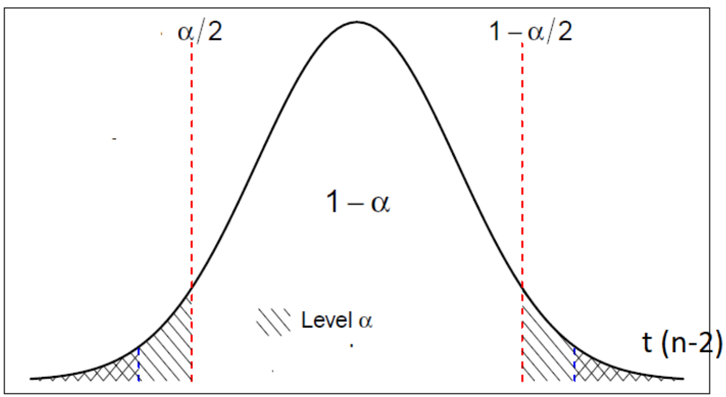
- **Theorem 3: Sampling distributions**

$$\frac{\hat{\beta}_0 - \beta_0}{SE(\hat{\beta}_0)} \sim t(n-2), \quad \frac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \sim t(n-2)$$

- **Confidence Interval for β_0 and β_1**

Since $\frac{\hat{\beta}_j - \beta_j}{SE(\hat{\beta}_j)} \sim t(n-2),\ j = 0, 1$

$(1-\alpha) = P\left(t_{\alpha/2}^{(n-2)} < \frac{\hat{\beta} - \beta}{SE(\hat{\beta})} < t_{1-\frac{\alpha}{2}}^{(n-2)}\right)$



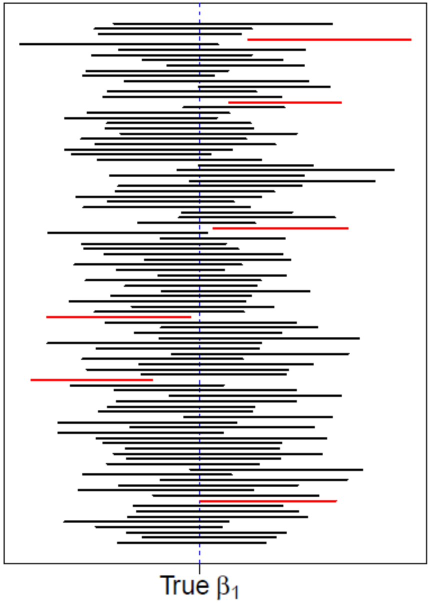
Confidence Intervals

For $100(1-\alpha)\%$ CI

$$\hat{\beta}_j \pm t(1-\alpha/2; n-2)SE(\hat{\beta}_j), j = 0, 1$$

Why should we care about confidence intervals?

- The confidence interval completely captures the information in the data about the parameter.



① Center is your estimate

② Length is how sure you are about your estimate

③ Any value outside would be rejected by a test.

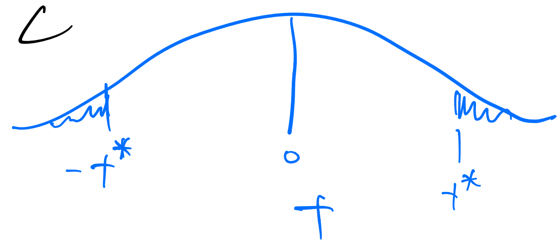
Hypothesis test concerning β_1

1. Two-sided test:

$$H_0 : \beta_1 = \beta_1^0 \quad vs \quad H_a : \beta_1 \neq \beta_1^0$$

2. Test statistics:

$$t^* = \frac{\hat{\beta}_1 - \beta_1^0}{SE(\hat{\beta}_1)}$$



3. **Decision Rule**:
   - P-value approach: we can compute the

$$p - value = 2 \times P(t_{(n-2)} \geq |t^*|)$$

The probability of observing value more extreme than t^*.

If $P - value < \alpha$, reject H_0
where α is the level of significance, or probability of type I error, $\alpha =$ Pr(Reject H_0 | H_0 is true).
   - Critical value approach:
If $|t^*| > t(1 - \alpha/2; n-2)$, Reject H_0
If $|t^*| \leq t(1 - \alpha/2; n-2)$, Accept H_0

Exercise (Crime Rate)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	20517.5999	3277.64269	6.259865	0.00e+00
X	-170.5752	41.57433	-4.102897	9.57e-05

- Test whether or not there is a linear association between crime rate and percentage of high school graduates, using a t test with $\alpha = 0.01$. State the hypotheses, decision rule, and conclusion.

$H_0 : \beta_1 = 0 \qquad H_1 : \beta_1 \neq 0$

$$t^* = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{-170.57520 - 0}{41.47433} = -4.103$$

$$\text{p-value} = 2 \cdot \underbrace{P(t_{(82)} \geq 4.103)}_{\text{rcode: pt(-4.103, 82)}} = 9.57e^{-05}$$

Since p-val $\ll \alpha$, so we reject H_0.

Exercise (Crime Rate)

② Critical Val Approach.

$|t^*| = 4.103 > \underbrace{t^{(82)}_{(1-\frac{0.01}{2})}}_{\text{rcode: } qt(0.995, 82)} = 2.6371$

$\therefore$ we reject H_0.

- Estimate β_1 with a 99 % confidence interval. Interpret the interval estimate.

$\hat{\beta}_1 \pm t^{(82)}_{0.995} SE(\hat{\beta}_1)$

$\in (-280.33, -60.82)$

Thus: We are 99% confidence we estimate the mean $\hat{\beta}_1$ is between $(-280.33, -60.82)$.

or

we are 99% confident there IS a liner relationship since $0 \notin CI$.

2.4 Interval Estimation of $E(Y|X = x_0)$

Suppose that x_0 is a new value of x for which we want to do prediction.

1. Estimation of $\mu_0 = E[Y|X = x_0]$
2. Prediction of Y value for an individual with $X = x_0$

- We use fitted regression model to do both of these
- Estimation of μ_0

$$\mu_0 = \beta_0 + \beta_1 X_0$$

Let's derive the variance formula:

Derivation of $Var(\hat{Y}_0)$

Confidence Interval for $\mu_0 = E(Y|X = x_0)$

- A $(1-\alpha)\%$ confidence interval for μ_0:

$$\hat{Y}_0 \pm t(1-\alpha/2; n-2)SE(\hat{Y}_0)$$

- Exercise: Obtain 95% confidence interval for the mean crime rate for states of high school graduate rate of 80%.

```
new.data = data.frame(X=80)
predict(fit, new.data, interval="confidence")
```

```
##        fit      lwr      upr
## 1 6871.585 6347.116 7396.054
```

2.5 Prediction of new observation

The value of Y for an individual with $X = x_0$ is:

$$Y_0 = \beta_0 + \beta_1 X + \epsilon_0$$

- Estimate of Y_0: $\hat{Y}_0 = \hat{\beta}_0 + \hat{\beta}_1 X_0 = \hat{\mu}_0$
- Prediction error: $e_0 = Y_0 - \hat{Y}_0$
- $Var(e_0) = Var(Y_0 - \hat{Y}_0)$

- $100 \times (1 - \alpha)\%$ **Prediction Interval** for a new observation $Y_{0(new)}$ with $X = x_0$ is:

$$\hat{Y}_0 \pm t(1 - \alpha/2; n - 2) s_{\{pred\}}$$

Exercise

- Exercise: Obtain 95% prediction interval for the crime rate for states of high school graduate rate of 80%.

```r
new.data = data.frame(X=80)
predict(fit, new.data, interval="prediction")
```

```
##        fit     lwr      upr
## 1 6871.585 2154.92 11588.25
```

Graph Prediction Intervals (using ggplot)

```
Crime.Pred <- predict(fit, interval="prediction")
```

```
## Warning in predict.lm(fit, interval = "prediction"): predictions on current data refer to _future_ responses
```

```
new.df = cbind(Crime,Crime.Pred)
library(ggplot2)
ggplot(data=new.df, aes(X, Y)) +
geom_point()+
geom_line(aes(y=lwr), color="red", linetype="dashed")+
geom_line(aes(y=upr), color="red", linetype="dashed")+
geom_smooth(method=lm, se=TRUE) + theme(plot.margin=unit(c(-1, 8, 7, 4), "cm"))
```