STAT 410 - Linear Regression

Lecture 3

Meng Li

Department of Statistics

Sep. 5, 2017

Properties of an estimator - a review (STAT 310) I

- Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a random sample taken from the distribution $f_\theta(\cdot)$ where θ is the unknown parameter.
- An estimator $\delta(\boldsymbol{X})$ is a function of the sample $\boldsymbol{X}$
- Two desirable properties of estimators under repeated experiments:
  - "Overall accuracy": $\mathrm{E}(\delta(\boldsymbol{X}))|\theta)$ is close to θ.
  - "Precision": $\mathrm{Var}(\delta(\boldsymbol{X})|\theta)$ is small.
- An estimator $\delta(\boldsymbol{X})$ for θ is said to be unbiased if its overall accurate for all possible values of θ. That is

$$\mathrm{E}(\delta(\boldsymbol{X})|\theta) = \theta, \qquad \text{for all } \theta. \tag{1}$$

  - The bias of $\delta(\boldsymbol{X})$ given θ is defined to be

$$B_\delta(\theta) = \mathrm{E}(\delta(\boldsymbol{X})|\theta) - \theta. \tag{2}$$

  - $B_\delta(\theta) > 0$: $\delta(X)$ tends to overestimate θ.
  - $B_\delta(\theta) < 0$: $\delta(X)$ tends to underestimate θ.
- Neither unbiasedness nor precision alone is enough. More generally, the goal is to have estimators *likely* to take values close to the *unknown fixed* parameter.
- We use a loss function $L(\theta, a)$ as a notion of distance between the estimate and the parameter, and choose an estimator that *minimizes the expected loss of $\delta(X)$*.
- Under the squared error loss, i.e., $L(\theta, a) = (\theta - a)^2$, this leads to the mean squared error (MSE):

$$MSE_\delta(\theta) = \mathrm{E}\{(\delta(X) - \theta)^2 | \theta\} = \mathrm{Var}(\delta|\theta) + B_\delta(\theta)^2. \qquad (3)$$

- This is sometimes referred to as the *Bias-Variance trade-off*. We want estimators that strike a balance between small bias and small variability.

Properties of an estimator - a review (STAT 310) III

- Note that bias, variance and MSE are computed using only information available *before the experiment*, not the observed value of the data.
- These are the *average* distances between the estimator $\delta(\boldsymbol{X})$ and the parameter θ *if the experiment is repeated many times* under fixed parameter value θ. Recall that this is the sampling/frequentist perspective in contrast with the Bayesian viewpoint.
- The estimator $\delta(\boldsymbol{X})$ is chosen using information available before the experiment. The estimate given data $\boldsymbol{X} = \boldsymbol{x}$ is simply the plug-in value of that estimator, i.e., $\delta(\boldsymbol{x})$.
- Under the sampling perspective, do we know anything about how far our realized estimate $\delta(\boldsymbol{x})$ is from the underlying θ? The answer is No.

Notation

- Note on notation: lower case (x, y) vs. upper case (X, Y).
  - In many contexts (such as STAT310), we use the upper case for a random variable and lower case for its realization.
  - In STAT410, we use the lower case y's for **both** random variables and the corresponding realizations unless stated otherwise.
  - The regressor x can be viewed as fixed constants thoroughly for our purposes:
    - Even with a random design for x's, our model is conditional on x's thus x's can be viewed as given.

Properties of estimators in SLR

- LS estimators $(\hat{\beta}_0, \hat{\beta}_1)$
  - $(\hat{\beta}_0, \hat{\beta}_1)$ are unbiased estimators of (β_0, β_1):
$$\mathrm{E}(\hat{\beta}_0) = \beta_0, \quad \mathrm{E}(\hat{\beta}_1) = \beta_1. \tag{4}$$
  - (Marginal) variances of $(\hat{\beta}_0, \hat{\beta}_1)$:
$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}, \quad \mathrm{Var}(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right). \tag{5}$$
  - Both $\hat{\beta}_0$ and $\hat{\beta}_1$ are linear combinations of y_i.
  - LS estimators are the Best Linear Unbiased Estimators (BLUE), known as the **Gauss-Markov theorem**.
- The estimator of variance $\widehat{\sigma^2} = \sum_{i=1}^n e_i^2/(n-2)$ is unbiased.
  - $\sum_{i=1}^n e_i^2$: residual (error) sum of squares, denoted as SS_{Res}.
  - $\sum_{i=1}^n e_i^2/(n-2)$: residual mean square, denoted as MS_{Res}.

Inferences on model parameters

- Gaussian assumption on the error term:

$$\varepsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2). \tag{6}$$

- All the previous moment properties do not depend on this assumption.
- We need this normal assumption in order to make **inferences** on parameters such as:
  - Hypothesis testing
  - Interval estimation
- Under the assumption of (6), we obtain that

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right), \quad \hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{S_{xx}}\right), \tag{7}$$

and

$$\frac{(n-2)\widehat{\sigma^2}}{\sigma^2} = \frac{SS_{Res}}{\sigma^2} \sim \chi^2_{n-2}. \tag{8}$$

t-test for the slope

- Suppose we wish to test the hypothesis that the slope equals a constant, say β_{10}.
- The hypotheses are

$$H_0 : \beta_1 = \beta_{10}, \quad H_1 : \beta_1 \neq \beta_{10} \tag{9}$$

- Test statistic:

$$T = \frac{\hat{\beta}_1 - \beta_{10}}{\text{se}(\hat{\beta}_1)}, \tag{10}$$

  where $\text{se}(\hat{\beta}_1) = \sqrt{\widehat{\sigma^2}/S_{xx}} = \sqrt{MS_{Res}/S_{xx}}$.
  - Recall: standard error (se) of an estimator is its estimated standard deviation.
- $T \sim t_{n-2}$ under H_0 - Null distribution of T.
- Reject H_0 if $|T| > t_{\alpha/2,n-2}$, where $t_{\alpha/2,n-2}$ is the upper $\alpha/2$ percentage of t_{n-2}.
- P-value = $2(1 - F_{t_{n-2}}(|T|))$, where $F_{t_{n-2}}$ is the CDF of t_{n-2}.

t-test for the intercept

- Suppose we wish to test the hypothesis that the intercept equals a constant, say β_{00}.
- The hypotheses are

$$H_0 : \beta_0 = \beta_{00}, \quad H_1 : \beta_0 \neq \beta_{00} \tag{11}$$

- Test statistic:

$$T = \frac{\hat{\beta}_0 - \beta_{00}}{\text{se}(\hat{\beta}_0)}, \tag{12}$$

where $\text{se}(\hat{\beta}_0) = \sqrt{MS_{Res}\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}$.

- $T \sim t_{n-2}$ under H_0 - Null distribution of T.
- Reject H_0 if $|T| > t_{\alpha/2,n-2}$, where $t_{\alpha/2,n-2}$ is the upper $\alpha/2$ percentage of t_{n-2}.
- P-value = $2(1 - F_{t_{n-2}}(|T|))$, where $F_{t_{n-2}}$ is the CDF of t_{n-2}.

Testing Significance of Regression

$$H_0 : \beta_1 = 0, \quad H_1 : \beta_1 \neq 0$$

- This tests the **significance of regression**; that is, is there a linear relationship between the response and the regressor.
- Failing to reject $\beta_1 = 0$, implies that there is no linear relationship between y and x.

Interval estimation of parameters

- $100(1-\alpha)\%$ Confidence interval for the Slope:

$$\hat{\beta}_1 - t_{\alpha/2,n-2} se(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2,n-2} se(\hat{\beta}_1).$$

- $100(1-\alpha)\%$ Confidence interval for the Intercept:

$$\hat{\beta}_0 - t_{\alpha/2,n-2} se(\hat{\beta}_0) \leq \beta_0 \leq \hat{\beta}_0 + t_{\alpha/2,n-2} se(\hat{\beta}_0).$$

- $100(1-\alpha)\%$ Confidence interval for σ^2:

$$\frac{(n-2)MS_{Res}}{\chi^2_{\alpha/2,n-2}} \leq \sigma^2 \leq \frac{(n-2)MS_{Res}}{\chi^2_{1-\alpha/2,n-2}}.$$

Interval estimation of the mean response

- Let x_0 be the level of the regressor variable at which we want to estimate the mean response, i.e.

$$\mathrm{E}(y|x_0) = \mu_{y|x_0} = \beta_0 + \beta_1 x_0.$$

- Point estimate: $\widehat{\mathrm{E}(y|x_0)} = \hat{\mu}_{y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$
- Variance of $\hat{\mu}_{y|x_0}$:

$$\begin{aligned}
\mathrm{Var}(\hat{\mu}_{y|x_0}) &= \mathrm{Var}(\hat{\beta}_0 + \hat{\beta}_1 x_0) = \mathrm{Var}(\bar{y} + \hat{\beta}_1 (x_0 - \bar{x})) \\
&= \mathrm{Var}(\bar{y}) + \mathrm{Var}(\hat{\beta}_1 (x_0 - \bar{x})) \qquad (13) \\
&= \frac{\sigma^2}{n} + \frac{\sigma^2 (x_0 - \bar{x})^2}{S_{xx}} = \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right),
\end{aligned}$$

where (13) uses the fact that Cov$(\bar{y}, \hat{\beta}_1 (x_0 - \bar{x})) = 0$.

- $100(1-\alpha)\%$ confidence interval for $\mathrm{E}(y|x_0)$:

$$\hat{\mu}_{y|x_0} - t_{\alpha/2,n-2}\sqrt{MS_{Res}\left(\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right)} \leq \mathrm{E}(y|x_0)$$
$$\leq \hat{\mu}_{y|x_0} + t_{\alpha/2,n-2}\sqrt{MS_{Res}\left(\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right)}.$$

Prediction interval of new observations

- Suppose we wish to construct a prediction interval on a future observation, y_0 at a particular level of x, say x_0.
- Point estimate: $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$.
- The confidence interval on the mean response at this point is not appropriate for this situation. (Why?)
- $y_0 = E(y|x = x_0) + e_0$ thus is more uncertain than $E(y|x = x_0)$.
- It can be shown:

$$\mathrm{E}(y_0 - \hat{y}_0) = 0, \qquad \mathrm{Var}(y_0 - \hat{y}_0) = \sigma^2 \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right),$$

which uses the fact that y_0 is independent of $\hat{y}_0$.

- $100(1-\alpha)\%$ prediction interval on y_0:

$$\hat{y}_0 - t_{\alpha/2, n-2} \sqrt{MS_{Res} \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)} \le y_0$$

$$\le \hat{y}_0 + t_{\alpha/2, n-2} \sqrt{MS_{Res} \left(1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right)}.$$