Lecture 3 Stat 331.

Last class : Simple Linear Regression

$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

① Least-squares estimation:

$$\underset{\beta_0,\beta}{\arg\min}\ S(\beta_0,\beta_1) = \underset{\beta_0,\beta_1}{\arg\min} \sum_{i=1}^{n} [y_i - \underbrace{\beta_0 - \beta_1 x_i}_{\beta_0 + \beta x_i \Rightarrow \mu_i}]^2$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{S_{xy}}{S_{xx}}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$\Rightarrow \hat{\beta}_0, \hat{\beta}_1$ are estimates based on our data.

② $\hat{\beta}_0$: estimate of average response when $x=0$.

$\hat{\beta}_1$: estimate of the average change in response for every unit increase in x.

③ How do we estimate σ^2? Using residuals : $e_i = y_i - \hat{\mu}_i$

$= y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$

$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2 \Rightarrow \hat{\sigma}^2 = \frac{1}{n-2} SS(Res)$$ (SS(Res) : sums of sq. of residuals).

Maximum Likelihood Estimation

Suppose that a R.V. Y has a probability density function $f(y;\theta)$ (or $P(Y=y;\theta)$ if Y is discrete) and we want to estimate the parameter θ that characterizes this dist^n.

Likelihood Function : $L(\theta|y) = f(y;\theta)$

- Represents "likelihood" of observing y when parameter is θ ($P(Y=y;\theta)$ for pmf)
- find the value θ that makes the observed data most likely, hence "maximum likelihood"

Maximum Likelihood Estimator (MLE) : $\hat{\theta} = \underset{\theta}{\arg\max}\ L(\theta|y)$

It is often easier to work with the log-likelihood function:

$$\ell(\theta|y) = \log L(y|\theta) \quad (\text{natural } \log)$$

Derivative of the log-likelihood function is score function:

$$S(\theta|y) = \partial \ell(\theta|y)/\partial\theta = \ell'(\theta|y)$$

Solve for θ in $S(\theta|y)=0$ to get MLE $\hat{\theta}$.

In SLR, $(Y_1, \dots, Y_n)$ are independent. Furthermore $Y_i \sim N(\beta_0+\beta_1 x_i, \sigma^2)$, $i=1,\dots,n$.

In this context, the likelihood function for $\theta=(\beta_0, \beta_1, \sigma^2)$:

$$L(\theta|\vec{y}) = \prod_{i=1}^{n} f(y_i;\theta)$$

$$= \prod_{i=1}^{n} (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{\frac{-(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^2}\right\}$$

In addition, $\ell(\theta|\vec{y}) = \sum_{i=1}^{n}\left[-\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y_i-\beta_0-\beta_1 x_i)^2\right]$

$S(\theta|\vec{y})=0$:

$$\frac{\partial \ell}{\partial \beta_0} = 0 \Rightarrow \sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)=0 \quad \text{--- (1)}$$

$$\frac{\partial \ell}{\partial \beta_1} = 0 \Rightarrow \sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)x_i=0 \quad \text{--- (2)}$$

$$\frac{\partial \ell}{\partial \sigma^2} = 0 \Rightarrow \frac{-n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2=0 \quad \text{--- (3)}$$

Note: ① Equations (1) and (2) are identical to equations that were used to solve for the least-squares estimators (LSE)

$\therefore$ MLE and LSE are identical.

② For σ^2, solving (3) gives us $\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_1 x_i)^2 = \frac{1}{n}\sum_{i=1}^{n} e_i^2$ this is different from $\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2$ (from LS). In fact, $\hat{\sigma}^2_{MLE}$ is a biased estimator for σ^2, but $\hat{\sigma}^2$ (from LS) is unbiased. Therefore, the $\hat{\sigma}^2_{MLE}$ is not used in practice.

Aside: "Estimator" vs. "Estimate"

"Estimator": Rule of calculating an estimate of the quantity of interest.

- It is a function of R.V. $\therefore$ it is a R.V.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

"Estimate": realization of an estimator based on the observed data. It is a number/constant. In salary data, e.g., $\hat{\beta}_1 = 9450$.

Study the distribution of $\hat{\beta}_1, \hat{\beta}_1$ (LS estimators)

How do we calculate $E(\hat{\beta}_0)$, $Var(\hat{\beta}_0)$, $E(\hat{\beta}_1)$, $Var(\hat{\beta}_1)$, $Cov(\hat{\beta}_1, \hat{\beta}_0)$

Aside: Suppose $Y_i \sim N(\mu_i, \sigma_i^2)$ (all independent)

Then $\sum_{i=1}^{n} a_i Y_i \sim N\left(\sum_{i=1}^{n} a_i\mu_i, \sum_{i=1}^{n} a_i^2\sigma_i^2\right)$ (can be shown via MGF).

① Study the sampling dist^n of $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})Y_i - \cancel{\bar{Y}\sum_{i=1}^{n}(x_i-\bar{x})}^{0}}{\sum_{i=1}^{n}(x_i-\bar{x})x_i - \cancel{\bar{x}\sum_{i=1}^{n}(x_i-\bar{x})}^{0}}$$

$$= \underbrace{\sum_{i=1}^{n}(x_i-\bar{x})}_{\tilde{a}_i} Y_i \Big/ \underbrace{\sum_{i=1}^{n}(x_i-\bar{x})x_i}_{c} \qquad \text{let } a_i = \tilde{a}_i / c$$

$$= \sum_{i=1}^{n} a_i Y_i \qquad \therefore \hat{\beta}_1 \sim N(?, ?)$$

$$E(\hat{\beta}_1) = \sum_{i=1}^{n} a_i E(Y_i) = \sum_{i=1}^{n} a_i \mu_i$$

$$= \sum_{i=1}^{n} a_i(\beta_0 + \beta_1 x_i)$$

$$= \sum_{i=1}^{n}(x_i-\bar{x})(\beta_0+\beta_1 x_i) \Big/ \sum_{i=1}^{n}(x_i-\bar{x})x_i$$

$$= \Big[\beta_0 \underbrace{\sum_{i=1}^{n}\cancel{(x_i-\bar{x})}}_{=0} + \beta_1 \underbrace{\sum_{i=1}^{n}\cancel{(x_i-\bar{x})x_i}}\Big] \Big/ \underbrace{\sum_{i=1}^{n}\cancel{(x_i-\bar{x})x_i}}$$

$$= \beta_1$$

$\therefore$ We showed that $\hat{\beta}_1$ is an unbiased estimator of β_1

$$Var(\hat{\beta}_1) = \sum_{i=1}^{n} a_i^2 \, Var(Y_i)$$
$$= \sigma^2 \sum_{i=1}^{n} a_i^2$$
$$= \sigma^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 \Big/ \left[\sum_{i=1}^{n} (x_i - \bar{x}) x_i\right]^2$$
$$= \sigma^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 \Big/ \left[\sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})\right]^2$$
$$= \sigma^2 \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sigma^2 / S_{xx}$$

$\therefore \hat{\beta}_1 \sim N(\beta_1, \sigma^2/S_{xx})$

② Similarly, $\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right)$. (assignment problem)

Hint: Use fact that $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$.

and show $Cov(\bar{Y}, \hat{\beta}_1) = 0$

$\therefore \hat{\beta}_0$ is an unbiased estimator of β_0.

③ Covariance between $\hat{\beta}_0$ and $\hat{\beta}_1$: $Cov(\hat{\beta}_0, \hat{\beta}_1) = -\sigma^2 \bar{x} / S_{xx}$.

(assignment problem).

Summary : Under $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$, we have

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2/S_{xx}) \implies \frac{(\hat{\beta}_1 - \beta_1)}{\underbrace{\sqrt{\sigma^2/S_{xx}}}_{\text{S.D. of } \hat{\beta}_1}} \sim N(0,1) \text{ (standard Normal)}.$$

$$\hat{\beta}_0 \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right). \implies \frac{(\hat{\beta}_0 - \beta_0)}{\underbrace{\sqrt{\sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}}_{\text{SD of } \hat{\beta}_0}} \sim N(0,1)$$

Standard Error :

$$SE(\hat{\beta}_1) = \sqrt{\frac{\hat{\sigma}^2}{S_{xx}}} \qquad SE(\hat{\beta}_0) = \sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}$$

(replaced σ^2 in SD above with their estimates).

Aside. Define $T = Z/\sqrt{U/k}$ where $Z \sim N(0,1)$ $U \sim \chi^2_k$

then when Z and U are independent $\rightarrow$ $T \sim t_{k}$. $\rightarrow$ df

(Def^n: A R.V. $W = \sum_{i=1}^{k} Z_i^2 \Rightarrow W \sim \chi^2_k$). $\rightarrow$ df $\quad Z_i \sim N(0,1)$ iid.

$\underline{\text{Fact}}$. $\hat{\sigma}^2(n-2)/\sigma^2 = SS(Res)/\sigma^2 \sim \chi^2_{n-2}$

$\underline{\text{Goal}}$: $\dfrac{\hat{\beta}_1 - \beta_1}{SE(\hat{\beta}_1)} \sim ?$ $\qquad SE(\hat{\beta}_1) = \sqrt{\dfrac{\hat{\sigma}^2}{S_{xx}}} = \dfrac{\hat{\sigma}}{\sqrt{S_{xx}}}$

$$\frac{(\hat{\beta}_1-\beta_1)/(\sigma/\sqrt{S_{xx}})}{\underbrace{(\hat{\sigma}/\sqrt{S_{xx}})/(\sigma/\sqrt{S_{xx}})}_{\sqrt{\hat{\sigma}^2/\sigma^2}}} = \frac{\overbrace{(\hat{\beta}_1-\beta_1)/(\sigma/\sqrt{S_{xx}})}^{N(0,1)\to(a)}}{\sqrt{\underbrace{\frac{\hat{\sigma}^2(n-2)}{\sigma^2}}_{\chi^2_{n-2}\to(b)}\frac{1}{(n-2)}}} \sim t_{n-2}$$

(Can show that (a) and (b) are independent)

$$\therefore \frac{\hat{\beta}_1-\beta_1}{SE(\hat{\beta}_1)} \sim t_{n-2} \qquad \left(\frac{\hat{\beta}_0-\beta_0}{SE(\hat{\beta}_0)} \sim t_{n-2}\right)$$

$\underline{\text{Next class}}$: Use the dist^n of $\hat{\beta}_1$ (t-distribution results) to show

(1) Construct confidence intervals for β_1

(2) Conduct Hypothesis test for β_1 (e.g. if $\beta_1 = 0$)

We'd like to show the following:

$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2$ $\quad E(\hat{\sigma}^2) = \sigma^2$ (for unbiasedness)

Sketch Proof:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 \rightarrow \text{Sub. in } \hat{\beta}_0$$

$$= \sum_{i=1}^{n}[y_i - \bar{y} + \hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i]^2$$

$$= \sum_{i=1}^{n}[(y_i - \bar{y}) - (x_i - \bar{x})\hat{\beta}_1]^2 \rightarrow \hat{\beta}_1 = S_{xy}/S_{xx}$$

$$= S_{yy} - 2\frac{S_{xy}}{S_{xx}}S_{xy} + \frac{S_{xy}^2}{S_{xx}} \quad \left(S_{yy} = \sum_{i=1}^{n}(Y_i - \bar{Y})^2\right)$$

$$\therefore \sum_{i=1}^{n} e_i^2 = S_{yy} - S_{xy}^2/S_{xx}$$

Moreover,

$$E(S_{yy}) = (n-1)\sigma^2 + \beta_1^2 S_{xx} \quad \text{(can be shown)}$$

$$E(S_{xy}^2) = \sigma^2 S_{xx} + S_{xx}^2\beta_1^2 \quad \text{(can be shown)}$$

Then,

$$E(S_{yy} - S_{xy}^2/S_{xx}) = \sigma^2(n-2)$$

$$\therefore E(\hat{\sigma}^2) = \sigma^2(n-2)/(n-2) = \sigma^2$$