Last Time:

$$\sum_{i=1}^{n}\left[Var(y_i-\bar{y})+[E(y_i-\bar{y})]^2\right]-S_{xx}\left[\frac{\sigma^2}{S_{xx}}+\beta_1^2\right]$$

$$=\sum_{i=1}^{n}\left[\frac{n-1}{n}\sigma^2+\frac{(n-1)^2}{n}\sigma^2\right]$$

$$=\sum_{i=1}^{n}\left[\frac{n-1}{n}\sigma^2(1+n-1)\right]$$

$$=\sum_{i=1}^{n}(n-1)\sigma^2-\sigma^2-S_{xx}\beta_1^2$$

$$=\sum_{i=1}^{n}(n-2)\sigma^2-S_{xx}\beta_1^2 \xrightarrow{\text{Somehow}} (n-2)\sigma^2\ldots$$

Define $MRSS=\frac{RSS}{n-2}$, the $E(MRSS)=\sigma^2$. So $\hat{\sigma}^2=MRSS$,

Degrees of freedom: number of cases minus the number of parameters in the mean function ($n-2$ for SLR)

$s^2=\frac{\sum(x_i-\bar{x})^2}{n-1}$, $\frac{(n-1)s^2}{\sigma^2}\sim\chi^2_{n-1}$ ($n-1$ degrees of freedom)

$y=\beta_0+\beta_1x_1+\ldots+\beta_px_p$, $dof=n-(p+1)$

5 Estimated Variance

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

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

$$\Rightarrow \hat{Var}(\hat{\beta}_1)=\frac{\hat{\sigma}^2}{S_{xx}}$$

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

3. Confidence Interval and Tests

Assumptions for SLR.

① Linearity

② $E(\varepsilon_i)=0$, $Var(\varepsilon_i)=\sigma^2$ (Unknown, Constant) (homoscedasticity....)

③ X_i fixed, Y_i is random variable. $E(Y_i)=\beta_0+\beta_1 x_i$ $Var(Y_i)=\sigma^2$

④ ε_i are uncorrelated, therefore Y_i are uncorrelated (Independence)

⑤ ε_i are normally distributed. $\varepsilon_i \overset{iid}{\sim} N(0,\sigma^2)$ (Normality)

Under these conditions, $Y_i \sim N(\beta_0+\beta_1 x_i, \sigma^2)$

Since $\hat{\beta}_0, \hat{\beta}_1$ are both linear estimators, they both follow Normal Distribution

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

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

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

$$se(\hat{\beta}_0)=\sqrt{\hat{\sigma}^2(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}})}$$

$$\frac{\hat{\beta}_1-\beta_1}{\sqrt{\frac{\sigma^2}{S_{xx}}}} \sim N(0,1) \Rightarrow \frac{\hat{\beta}_1-\beta_1}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} \sim t(n-2)$$

$$\frac{\hat{\beta}_0-\beta_0}{\sqrt{\sigma^2(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}})}} \sim N(0,1) \Rightarrow \frac{\hat{\beta}_0-\beta_0}{\sqrt{\hat{\sigma}^2(\frac{1}{n}+\frac{\bar{x}^2}{S_{xx}})}} \sim t(n-2)$$

$$\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \qquad \frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}} \sim t(n-1)$$

Tests

① For β_1 $H_0: \beta_1 = \beta_1^*$
$H_a: \beta_1 \neq \beta_1^*$

test statistic: $t = \dfrac{\hat{\beta}_1 - \beta_1^*}{\sqrt{\frac{\hat{\sigma}^2}{S_{xx}}}} \sim t(n-2)$

rejection region: reject H_0 if $t > t_{\frac{\alpha}{2}, n-2}$
or $t < -t_{\frac{\alpha}{2}, n-2}$

② For β_0: $H_0: \beta_0 = \beta_0^*$
$H_a: \beta_0 \neq \beta_0^*$

test statistic: $t = \dfrac{\hat{\beta}_0 - \beta_0^*}{\sqrt{\tilde{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}} \sim t(n-2)$

rejection region: reject H_0 if $t > t_{\frac{\alpha}{2}, n-2}$
or $t < -t_{\frac{\alpha}{2}, n-2}$

Specifically, test $\beta_1 = 0$. If null hypothesis is not rejected, there is no significant linear relationship.

$100(1-\alpha)\%$ CI for β_1:

$$\hat{\beta}_1 \pm t_{\alpha/2, n-2} \cdot \sqrt{\frac{\tilde{\sigma}^2}{S_{xx}}}$$

$100(1-\alpha)\%$ CI for β_0:

$$\hat{\beta}_0 \pm t_{\frac{\alpha}{2}, n-2} \cdot \sqrt{\tilde{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)}$$

* In fact, under the conditions:

$$\frac{(n-2)MRSS}{\sigma^2} \sim \chi^2(n-2)$$

$100(1-\alpha)\%$ CI for σ^2

$$\left(\frac{(n-2)\,MRSS}{\chi^2_{\frac{\alpha}{2}, n-2}}, \frac{(n-2)\,MRSS}{\chi^2_{\frac{\alpha}{2}, n-2}}\right)$$

$P(-Z_{\alpha/2} < \_\_ < Z_{\alpha/2}) = 1-\alpha$

$P(\_\_ < N < \_\_)\, 1-\alpha$

For the mean response $E(Y|X=x_0)$

$$E(Y|X=x_0)=\beta_0+\beta_1 x_0$$

$$\hat{E}(Y|X=x_0)=\hat{\beta}_0+\hat{\beta}_1 x_0$$

$$Var(Y|X=x_0)=Var(\hat{\beta}_0+\hat{\beta}_1 x_0)=\sigma^2\left(\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right)$$

$$\hat{Var}(Y|X=x_0)=\hat{\sigma}^2\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]$$

$$\Rightarrow Se(Y|X=x_0)=\sqrt{\hat{\sigma}^2\left[\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}\right]}$$

$100(1-\alpha)\%$ CI of $E(Y|X=x_0)$ is:

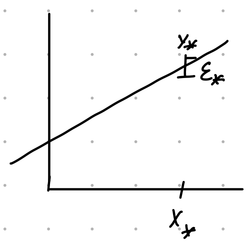
$$(\hat{\beta}_0+\hat{\beta}_1 x_0)\pm t_{\frac{\alpha}{2},n-2}\cdot Se(Y|X=x_0)$$

For prediction with a new value x_* the true value of the prediction is

$$Y_*=\beta_0+\beta_1 X_*+\varepsilon_*$$

its estimated value is

$$\hat{Y}_*=\hat{\beta}_0+\hat{\beta}_1 x_*+\varepsilon_*$$



$$Var(\hat{Y}_*)=Var(\hat{\beta}_0+\hat{\beta}_1 x_*)+Var(\varepsilon_*)$$

$$=\sigma^2\left(\frac{1}{n}+\frac{(x_*-\bar{x})^2}{S_{xx}}\right)+\sigma^2$$

$$=\sigma^2\left(1+\frac{1}{n}+\frac{(x_*-\bar{x})^2}{S_{xx}}\right)$$

$$se(\hat{Y}_*)=\sqrt{\sigma^2\left(1+\frac{1}{n}+\frac{(x_*-\bar{x})^2}{S_{xx}}\right)}$$

$100(1-\alpha)\%$ for Y_* is

$$\hat{\beta}_0+\hat{\beta}_1 x_*\pm t_{\frac{\alpha}{2},n-2}\cdot Se(\hat{Y}_*)$$