$$\vec{Y} = X\vec{\beta} + \vec{\varepsilon}$$

homoskedastic means same variance

$$\underbrace{n\left\{\begin{bmatrix} Y \end{bmatrix}\right.}_{1} = \underbrace{n\left\{\begin{bmatrix} \quad X \quad \end{bmatrix}\right.}_{p} \; p\left\{\underbrace{\begin{bmatrix} \beta \end{bmatrix}}_{1}\right. + \underbrace{n\left\{\begin{bmatrix} \varepsilon \end{bmatrix}\right.}_{1}$$

Standard Assumption about error

$$\vec{\varepsilon} \sim N_n(\vec{0}, \sigma^2 I)$$

Idempotent, Normal, Homoscedastic

"same variance"

$\vec{\beta}, \sigma^2$ are unknown parameters

$$\Rightarrow \vec{Z} = \frac{1}{\sigma}\vec{\varepsilon} \sim N_n(\vec{0}, I)$$

$$\frac{1}{\sigma^2}\vec{\varepsilon}^T\vec{\varepsilon} \sim \chi^2_n$$

with 30 minutes of matrix algebra we can find that the multivariate least squares estimate for $\vec{\beta}$ is

$\wedge$ for estimate

$$\hat{\vec{\beta}} = (X^T X^{-1})^{-1} \vec{X}^T \vec{Y} = (X^T X)^{-1} X^T (X\vec{\beta} + \varepsilon)$$

$$= (X^T X^{-1})^T X^T X \vec{\beta} + (X^T X)^{-1} X^T \vec{\varepsilon}$$

$$= \vec{\beta} + (X^T X)^{-1} X^T \vec{\varepsilon}$$

$$\vec{B} = N_p\left(\vec{\beta}, \left((X^T X^{-1})^T (\sigma^2 I)\left((X^T X)^{-1} X^T\right)^T\right)\right.$$

$$= N_p\left(\vec{\beta}, \sigma^2 (X^T X)^{-1}\right)$$

$$\Rightarrow E\left(\hat{\vec{B}}\right) = \vec{\beta}$$ so $\hat{\vec{B}}$ is unbiased estimator

$$\xRightarrow{\text{margining}} \hat{\beta}_K \sim N\left(\beta_K, \sigma^2 (X^T X)^{-1}_{KK}\right)$$

$$\Longrightarrow \frac{\hat{\beta}_K - \beta}{\sigma\sqrt{(X^T X)^{-1}_{KK}}} \sim N(0,1)$$

Student said $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ is approx. same as $\frac{\bar{X}-\mu}{s/\sqrt{n}}$

Students problem: σ is unknown, use estimate instead

$$\frac{1}{\sigma^2}\vec{\varepsilon}^T\vec{\varepsilon} = \frac{1}{\sigma^2}\vec{\varepsilon}^T P \vec{\varepsilon} + \frac{1}{\sigma^2}\vec{\varepsilon}^T (I-P)\vec{\varepsilon}$$

where

$P := X(X^T X)^{-1} X^T$ ← orthogonal projection matrix onto p dim.

$\mathrm{rank}[P] = p$

(take n-dimensional input makes it into p-dimensional)

$I-P$ is an orthogonal projection matrix onto the remaining $n-p$ dimensions

$\mathrm{rank}[I-P] = n-p$

need $PP = P$: $\quad PP = X(X^T X)^{-1} X^T X (X^T X)^{-1} X^T$

$PP = P$ ✓ P is idempotent

need $(I-P)(I-P) = (I-P)$: $\quad (I-P)(I-P) = II - PI - IP + PP$

$= I - P - P + P$

$(I-P)(I-P) = I - P$ ✓ $I-P$ is idempotent

so Cochrain's Theorem applies...

$$\frac{1}{\sigma^2}\vec{\varepsilon}^{\,T}\vec{\varepsilon} = \frac{1}{\sigma^2}\vec{\varepsilon}^{\,T} \underset{\chi^2_p}{P}\vec{\varepsilon} + \frac{1}{\sigma^2}\vec{\varepsilon}\underset{\chi^2_{n-p}}{(I-P)}\vec{\varepsilon}$$

and are indep. by
Cochrain's Theorem
so can do:

$$\chi^2_p + \chi^n_{n-p} \sim \chi^n_n$$

$$\vec{\varepsilon}^{\,T} P\vec{\varepsilon} = \vec{\varepsilon}^{\,T} PP\vec{\varepsilon} = (P\vec{\varepsilon}^{\,T})\, P\vec{\varepsilon}$$

$$\vec{\varepsilon}(I-P)\vec{\varepsilon} = \vec{\varepsilon}^{\,T}(I-P)(I-P) = ((I-P)\vec{\varepsilon})^T((I-P)\vec{\varepsilon})$$

$$P\vec{\varepsilon} = P(\vec{Y} - X\vec{\beta}) = P\vec{Y} - P\vec{X}\beta$$

$$= \underbrace{X(X^TX)^{-1}X^T\vec{Y}}_{X\hat{\vec{\beta}}} - X\underbrace{(X^TX^{-1})X^T\beta}_{X}$$

$$= X\hat{\vec{\beta}} - X\vec{\beta}$$

$$= X(\hat{\vec{\beta}} - \vec{\beta})$$

so $\vec{\varepsilon}^{\,T} P\vec{\varepsilon} = \vec{\varepsilon}^{\,T} PP\varepsilon = (P\vec{\varepsilon}^{\,T})\, P\vec{\varepsilon}$

$$= (\hat{\vec{\beta}} - \vec{\beta})^T X^T X(\hat{\vec{\beta}} - \vec{\beta})$$

$$\vec{Y} - X\hat{\vec{\beta}} \approx \vec{Y} - X\vec{\beta}$$

$$\vec{P\varepsilon} = P(\vec{Y} - X\vec{\beta}) = P\vec{Y} - PX\vec{\beta}$$
$$= \underbrace{X(X^TX)^{-1}X^T\vec{Y}}_{\vec{\hat{\beta}}} - X\underbrace{(X^TX)^{-1}X^TX}_{I}\beta$$

$$P\vec{\varepsilon} = X(\vec{\hat{\beta}} - \vec{\beta})$$

P is idempotent
P = PP

$$\vec{\varepsilon}^T P\vec{\varepsilon} = \vec{\varepsilon}^T PP\vec{\varepsilon} = (P\vec{\varepsilon})^T P\vec{\varepsilon}$$
$$= (\vec{\hat{\beta}} - \vec{\beta})^T X^TX(\vec{\hat{\beta}} - \vec{\beta})$$

Y's predicted by model

$\vec{\hat{Y}} = X\hat{\vec{\beta}}$

$E = \vec{Y} - \vec{\hat{Y}}$

residuals

$$(I-P)\vec{\varepsilon} = (I-P)(Y - X\vec{\beta}) =$$
$$= I\vec{Y} - P\vec{Y} - IX\vec{\beta} + PX\vec{\beta}$$
$$= \vec{Y} - P\vec{Y} - X\vec{\beta} + X\vec{\beta}$$
$$= \vec{Y} - X\vec{\hat{\beta}}$$
$$= \vec{E}$$

$P\vec{Y} = \hat{\vec{\beta}}$

residuals (measure difference between predicted and actual $\vec{Y}$)

$$\vec{Y} - X\vec{\beta} \approx \vec{Y} - X$$

I-P is idempotent

$$\vec{\varepsilon}^T(I-P)\vec{\varepsilon} = \vec{\varepsilon}^T(I-P)(I-P)\vec{\varepsilon}$$
$$= ((I-P)\vec{\varepsilon})^T(I-P)\vec{\varepsilon}$$
$$= \vec{E}^T\vec{E} = \sum_{i=1}^{n} e$$

if $\vec{E} = (e_1, e_2, \ldots, e_n)$

$\vec{E} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$

$$= SSE \text{ (sum of squared error)}$$

$e_i = y_i - \hat{y}_i$

SSE indep. of $(\vec{\hat{\beta}} - \vec{\beta})^T X^TX(\vec{\hat{\beta}} - \vec{\beta})$

$\Rightarrow$ RMSE and $\vec{\hat{\beta}}$ are indep.

MSE is unbiased estimator for σ^2 (RMSE for σ)

$$E\left[\frac{1}{\sigma^2} SSE\right] = n-p \qquad E\left[\frac{SSE}{n+p}\right] = \sigma^2$$

$$E\left[\frac{1}{\sigma^2}\frac{SSE}{n+p}\right] = 1 \qquad MSE = \frac{SSE}{n+p} \qquad MSE \approx \sigma^2$$

mean squared error $\sqrt{MSE} = RMSE$

$\xRightarrow{\text{margining}} \hat{\beta}_K \sim N(\beta_K, \sigma^2 (X^T X^{-1})_{KK})$

MSE $\approx \sigma^2$

RMSE $\approx \sqrt{\text{MSE}} \approx \sigma$ "root mean squared error"

SSE indep of $(\hat{\vec{\beta}}-\vec{\beta})^T X^T X (\hat{\vec{\beta}}-\vec{\beta})$

$\Rightarrow$ RMSE, $\hat{\vec{\beta}}$ are indep.

$$\frac{\hat{\beta}_K - \beta_K}{\sigma \sqrt{(X^T X)^{-1}_{KK}}} \sim N(0,1)$$

$$\frac{\hat{\beta}_K - \beta_K}{\text{RMSE}\sqrt{(X^T X)^{-1}_{KK}}} = \frac{1}{\sqrt{\text{MSE}\,(X^T X)^{-1}_{KK} \frac{\sigma^2}{\sigma^2}}}$$

$$= \frac{\dfrac{\hat{\beta}_K - \beta_K}{\sigma\sqrt{(X^T X)^{-1}_{KK}}}}{\sqrt{\dfrac{\text{MSE}}{\sigma^2}\cdot\dfrac{n-p}{n-p}}} = \frac{\overbrace{\dfrac{\hat{\beta}_K - \beta_K}{E\sqrt{(X^T X)^{-1}_{KK}}}}^{Z}}{\underbrace{\sqrt{\dfrac{SSE/\sigma^2}{n-p}}}_{\sqrt{\frac{\chi^2_{n-p}}{n-p}}}}$$

$$\sim T_{n-p}$$

$\leftarrow$ justification for using T-test for linear regression

null hyp.

$H_0 : \beta_K = \text{value}$ (if T statistic is too big, reject this)

$$\frac{(\hat{\vec{\beta}}-\vec{\beta})^T X^T X(\hat{\vec{\beta}}-\vec{\beta}) / p}{SSE / n-p}$$

$$= \frac{\frac{1}{\sigma^2}\vec{\varepsilon}^{\,T} P \vec{\varepsilon} / p}{\frac{1}{\sigma^2}\vec{\varepsilon}^{\,T}(I-P)\vec{\varepsilon} / n-p} \sim F_{p, n-p}$$

Tells you how "good" the ~~linear~~ regression is

F test does all $\vec{\beta}$s at ~~once~~, not just one value, unlike T-test

~~(combined T is F^2)~~

$H_0 : \vec{\beta} = \text{value}$

This is justification for the omnibus F-test for linear regression

(F te