Chapter 0: Review

Chapter 2: Simple Linear Regression

$\mathbf{E[y|x]}= \mu_{y|x}$ =E$[\beta_0+\beta_1 x+\epsilon]= \beta_0+\beta_1 x$ $\mathbf{V[y|x]}= \sigma^2_{y|x}$ =$\mathbf{V}[\beta_0+\beta_1 x+\epsilon]= \sigma^2$ $\hat\beta_0= \bar y-\hat\beta_1\bar x$ $\hat\beta_1= \frac{S_{xy}}{S_{xx}}=\frac{\sum_{i=1}^n(x_i-\bar x)y_i}{\sum_{i=1}^n(x_i-\bar x)^2}$

$\mathbf{E}[\hat\beta_1]= \sum_{i=1}^n c_i \mathrm{E}[y_i]=\beta_0\sum_{i=1}^n c_i+\beta_0\sum_{i=1}^n c_i x_i=\beta_1$ $\mathbf{V}[\hat\beta_1]= \sum_{i=1}^n c_i^2(\sigma^2)=\sigma^2\frac{\sum_{i=1}^n(x_i-\bar x)^2}{S_{xx}^2}=\frac{\sigma^2}{S_{xx}}$

$\mathbf{E}[\hat\beta_0]= \beta_0$ $\mathbf{V}[\hat\beta_0]= \sigma^2(\frac{1}{n}+\frac{\bar x^2}{S_{xx}})=V[\bar y-\beta_1\bar x)]=V[\bar y]+x^2V[\hat\beta_1]-cov(\bar y,\hat\beta_1)$ $\mathbf{c}_i= \frac{x-\bar x}{S_{xx}}$

$\mathbf{SS}_{res}= \sum_{i=1}^n(y_i-\hat y_i)^2=\sum_{i=1}^n\epsilon_i^2$ $\mathbf{SS}_T= \sum_{i=1}^n y_i^2-n\bar y^2, n-1$ df $\mathbf{SS}_{Reg}= \hat\beta_1 S_{xy}$, if df = 1, then $= MS_{Res}$

$\mathbf{MS}_{res}= \sigma^2=\frac{SS_{res}}{n-2}$

Hypothesis Testing (Regression)

Reject H$_0$ if $|t_0| \geq t_{\frac{\alpha}{2},n-2}$ where $t_0=\frac{\hat\beta_1-\beta_{10}}{\sqrt{\frac{MS_{res}}{S_{xx}}}}$ Failing to reject H$_0$: $\beta_i=0$ implies no rlshp between x and y. E$[y_i]$ $=\beta_1$x $+\beta_0$

$F_0=\frac{MS_{Reg}}{MS_{res}}=t_0^2$ Reject if $F_0>F_\alpha,1,n-1$ CI: $\hat\beta_1-t_{\frac{\alpha}{2},n-2}se(\hat\beta_{10})<\hat\beta_{10}<\hat\beta_1+t_{\frac{\alpha}{2},n-2}se(\hat\beta_{10})$ se$(\hat\beta_1)= \sqrt{\frac{MS_{res}}{S_{xx}}}$, se$(\hat\beta_0)= \sqrt{V[\hat\beta_0]}$

$R^2= 1-\frac{SS_{res}}{SS_T}=\frac{SS_{Reg}}{SS_T}$ $R^2_{adj}=1-\frac{SS_{res}(n-1)}{SS_T(n-k-1}$

Chapter 3: Multiple Linear Regression

$\underset{n\times1}{y}=\underset{n\times p}{x}\times\underset{p\times1}{\beta}+\underset{n\times1}{\epsilon}$ where p = k + 1, p is the total number of betas (or parameters), k is the number of regressor variables.

$\epsilon\sim$N(0, σ^2I) where I is the identity matrix whatever size E[y] = xβ V[y] = V[ϵ] = σ^2I y $\sim$N(xβ , σ^2I)

Least Square Estimate for β and σ^2

$S(\beta)=\sum_{i=1}^n\epsilon^2=\epsilon'\epsilon=(y-x\beta)'(y-x\beta)=y'y-2\beta'x'y+\beta'x'x\beta$

$\hat\beta=(x'x)^{-1}x'y$ E$[\hat\beta]$ = E$[(x'x)^{-1}x'y]$ = E$[(x'x)^{-1}x'(x\beta+\epsilon)]$ = β V$[\hat\beta]$ = $(x'x)^{-1}\sigma^2$ = cσ^2 V$[\hat\beta_j]$ = $c_{jj}\sigma^2$ E$[\beta_j]$ = β_j

$\hat\beta_j\sim$N(β_j, $c_{jj}\sigma^2$) $\hat y=x\hat\beta=x(x'x)^{-1}x'y=Hy$ E$[\hat y]$ = E$[x\hat\beta]$ = xβ V$[\hat y]$ = V$[x\hat\beta]$ = xV$[\hat\beta]$x$'$ = x(x$'$x)$^{-1}$x$'\sigma^2$ = Hσ^2

$\hat y\sim$N(xβ , Hσ^2) $\hat y_j\sim$N(x$_j\beta$, h$_{jj}\sigma^2$), where $h_{jj}=x_j'(x'x)^{-1}x_j$ $x_j=[x_{j0},x_{j1},...x_{jk}]$ and $\hat\epsilon=y-\hat y=y-Hy=(I-H)y$

$\hat{\sigma^2}(estimator)=\frac{SS_{res}}{n-p}=MS_{res}$ where p = k + 1 = the number of parameters (i.e. β's: β_0, β_1, ... β_k)

$SS_{res}(\mathbf{n - p})=(y-x\hat\beta)'(y-x\hat\beta)=y'y-2\hat\beta'X'y+\hat\beta'x'x\hat\beta=y'y-\hat\beta'x'y$ $SS_{Reg}(\mathbf{k})=\hat\beta'x'y-\frac{(\sum_{i=1}^n y_i)^2}{n}$ $SS_T(\mathbf{n - 1})=y'y-\frac{(\sum_{i=1}^n y_i)^2}{n}$

$MS_{res}=\frac{SS_{res}}{n-k-1}$ $MS_{Reg}=\frac{SS_{Reg}}{k}$ $MS_T=\frac{SS_T}{n-1}$

If $\frac{SS_{res}}{\sigma^2}\sim\chi^2_{n-k-1}$ and SS$_{res}$, SS$_{Reg}$ are indep, then $F_0=\frac{\frac{SS_{Reg}}{k}}{\frac{SS_{Res}}{n-k-1}}=\frac{MS_{Reg}}{MS_{res}}$

error = (I - H)y = (I - H)ϵ E[MS$_{Res}$] = σ^2 E[MS$_{Reg}$] = $\sigma^2+\frac{\beta^{*'}x_c'x_c\beta^*}{k\sigma^2}$ where $\beta^*=(\beta_1,\beta_2,...\beta_k)$ and x_c is the center

We reject H$_0$ if F$_0$ > F$_{\alpha,k,n-k-1}$

Testing Individual Coefficients: If H$_0$: β_j = 0 is not rejected then delete it: $t_0=\frac{\hat\beta_j}{\sqrt{\sigma^2c_{jj}}}=\frac{\hat\beta_j}{se(\hat\beta_j)}$ reject if $|t_0|$ > $t_{\frac{\alpha}{2},n-k-1}$

Confidence Intervals

σ^2**known**: $\hat\beta_j\sim N(\beta_j,c_{jj}\sigma^2)\longrightarrow\frac{\hat\beta_j-\beta_j}{\sqrt{c_{jj}\sigma^2}}\sim N(0,1)$ or, if variance is unknown, $\hat\beta_j\sim N(\beta_j,c_{jj}MS_{res})\longrightarrow\frac{\hat\beta_j-\beta_j}{\sqrt{c_{jj}MS_{res}}}$or $\frac{\hat\beta_j-\beta_j}{se(\hat\beta_j)}\sim t_{n-p}$

Then the variance estimator is $\hat{\sigma^2}=MS_{res}=\frac{SS_{res}}{n-p}\sim\chi^2_{n-p}$ So, the (1 - α) **confidence interval** for β_j is $\hat\beta_j\pm t_{\frac{\alpha}{2},n-p}se(\hat\beta_j)$

$\hat y_j\sim$N(x$_j\beta$, h$_{jj}\sigma^2$), so $\frac{\hat y_j-x_j\beta}{\sqrt{h_{jj}\sigma^2}}\sim N(0,1)$ $\frac{\hat y_j-x_j\beta}{\sqrt{h_{jj}MS_{res}}}\sim t_{n-p}$ where σ^2estimates MS$_{res}$

A 1 - α confidence interval for E[y$_0$|x$_0$] is $\hat y_0\pm t_{\frac{\alpha}{2},n-p}\sqrt{x_0'(x'x)^{-1}x_0\hat{\sigma^2}}$ or $\hat y_0\pm t_{\frac{\alpha}{2},n-p}\sqrt{x_0'(x'x)^{-1}x_0MS_{res}}$

Chapter 4: Model Testing

Properties of residuals: mean 0, $MS_{res}=\sum_{i=1}^n\frac{(\epsilon_i-\bar\epsilon)^2}{n-p}=\sum_{i=1}^n\frac{\epsilon_i^2}{n-p}=\frac{SS_{res}}{n-p}$

Scaling Residuals: Standardized Residuals: $d_i=\frac{\epsilon_i}{\sqrt{MS_{res}}}$ Studentized: $r_i=\frac{\epsilon_i}{\sqrt{MS_{res}(1-h_{ii})}}$, $\mathbf{V}[\epsilon_i]$ = σ^2(1 - h$_{ii}$), $\mathbf{cov}(\epsilon_i, \epsilon_j)$ = $-\sigma^2h_{ij}$

Other model testing: plot x$_i$ and x$_j$: linear rln means high corr.

Formal test for lack of fit: Assuming everything is tested and ideal, to test for linearity, we use: $SS_{res}=SS_{PE}+SS_{LOF}$

F$_0$ = $\frac{SS_{LOF}/(m-2)}{SS_{PE}/(n-m)}$ = $\frac{MS_{LOF}}{MS_{PE}}$ E$[MS_{LOF}]$ = E$[MS_{PE}]$ = σ^2, where m is num regressors, n is num samples V$[\bar y]$ = $\frac{p\sigma^2}{n}$ (indpure e)