write up $AX \sim N(A\mu, A\Sigma A')$

Inference (tests & confidence intervals) based on the normal distribution

---

$$Y = X\beta + \varepsilon \qquad \varepsilon \sim N_n(0, \sigma^2 I_n)$$

$$Y \sim N_n(X\beta, \sigma^2 I_n)$$

$$\hat\beta = \underbrace{(X'X)^{-1}X'}_{A}Y \qquad E(\hat\beta) = E((X'X)^{-1}X'Y)$$

$$= (X'X)^{-1}X'EY$$

$$= (X'X)^{-1}X'X\beta = \beta \quad \text{unbiased}$$

$$\operatorname{cov}(\hat\beta) = (X'X)^{-1}X'\sigma^2 I_n\left[(X'X)^{-1}X'\right]'$$

$$= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}$$

So $\hat\beta \sim N_{k+1}(\beta, \sigma^2(X'X)^{-1})$

$$\hat Y = X\hat\beta \sim N_n(X\beta, \sigma^2 X(X'X)^{-1}X')$$

Residuals

$$\hat\varepsilon = Y - \hat Y \qquad \text{Vertical distances from pts to plane}$$

Errors in "prediction"

$$\hat{\varepsilon} = Y - \hat{Y} = Y - X\hat{\beta} = Y - X(X'X)$$

$$= IY - X(X'X)^{-1}X'Y$$

$$= (I - X(X'X)^{-1}X')Y$$

A

$$E(\hat{\varepsilon}) = E(Y) - E(X\hat{\beta}) = X\beta - X\beta = 0$$

$$\mathrm{cov}(\hat{\varepsilon}) = (I - X(X'X)^{-1}X')\sigma^2 I_n (I - X(X'X)^{-1}X')$$

$$= \sigma^2(I - X(X'X)^{-1}X' - X(X'X)^{-1}X'$$

$$+ X(X'X)^{-1}\underbrace{X'X(X'X)^{-1}}_{I}X')$$

$$= \sigma^2(I - X(X'X)^{-1}X')$$

~~Consider test statistics for~~

- ~~Single regression coefficents $\hat{\beta}_j$~~
- ~~More general SETS of linear combos~~

Now show that $\hat{\beta}$ & $\hat{\varepsilon}$ are independent

Like $\bar{x}$ & s^2 ind – a little surprising

Independence of $\hat{\beta}$ & $\hat{\varepsilon}$

.3

$$V(Y) = \sigma^2 I_n = E(Y-\mu_y)(Y-\mu_y)'$$
$$= E(YY') - \mu_y\mu_y' - \mu_y\mu_y' + \mu_y\mu_y'$$
$$= E(YY') - \mu_y\mu_y' = E(YY') - X\beta\beta'X', \text{ so}$$
$$E(YY') = \sigma^2 I_n + X\beta\beta'X'$$

$$\hat{\beta} = (X'X)^{-1}X'Y \qquad \hat{Y} = X\hat{\beta} \qquad e = Y - \hat{Y}$$

$$c(e,\hat{\beta}) = E\{(e-0)(\hat{\beta}-\beta)'\} = E\{e\hat{\beta}' - e\beta'\}$$
$$= E(e\hat{\beta}') = E\{(Y-\hat{Y})\hat{\beta}'\}$$
$$= E\{Y\hat{\beta}' - \hat{Y}\hat{\beta}'\} \text{ where } \hat{\beta}' = ((X'X)^{-1}X'Y)' = Y'X(X'X)^{-1}$$
$$= E\{YY'X(X'X)^{-1} - X\hat{\beta}\hat{\beta}'\}$$
$$= \underbrace{E\{YY'\}X(X'X)^{-1}} - X(X'X)^{-1}X'\underbrace{E\{YY'\}X(X'X)^{-1}}$$
$$= \left(I - X(X'X)^{-1}X'\right)E\{YY'\}X(X'X)^{-1}$$
$$= \left(I - X(X'X)^{-1}X'\right)\left(\sigma^2 I_n + X\beta\beta'X'\right)X(X'X)^{-1}$$
$$= \left(I - X(X'X)^{-1}X'\right)\left(\sigma^2 X(X'X)^{-1} + X\beta\beta'\underbrace{X'X(X'X)^{-1}}_{I}\right)$$
$$= \sigma^2 X(X'X)^{-1} + X\beta\beta' - X\underbrace{(X'X)^{-1}X'\sigma^2 X}_{I}(X'X)^{-1} + X\underbrace{(X'X)^{-1}X'X}_{I}\beta\beta'$$
$$= 0$$

Distribution of $SSE = \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta})$

Actually, distribution of $\frac{1}{\sigma^2} SSE$

Given

(1) If $U = W_1 + W_2$ and

- W_1 and W_2 are independent
- $W_1 \sim \chi^2(\nu_1)$
- $U \sim \chi^2(\nu_1 + \nu_2)$

$\Rightarrow W_2 \sim \chi^2(\nu_2)$

(2) From least squares calculation,

$$(Y - X\beta)'(Y - X\beta) = \underbrace{(Y - X\hat{\beta})'(Y - X\hat{\beta})}_{= SSE} + (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta)$$

(3) $Y \sim N_n(X\beta, \sigma^2 I_n)$

$\hat{\beta} \sim N_{k+1}(\beta, \sigma^2 (X'X)^{-1})$

(4) If $Y \sim N_p(\mu, \Sigma)$ then $(Y - \mu)'\Sigma^{-1}(Y - \mu) \sim \chi^2(p)$

(5) $\hat{\beta}$ and SSE are independent

From 2,

$$\frac{1}{\sigma^2}(Y-X\beta)'(Y-X\beta) = (Y-X\beta)'\frac{1}{\sigma^2}I_n(Y-X\beta)$$

$$= (Y-X\beta)'\left[\sigma^2 I_n\right]^{-1}(Y-X\beta) \overset{4}{\sim} \chi^2(n)$$

$$\shortparallel$$

$$\frac{1}{\sigma^2}SSE + \frac{1}{\sigma^2}(\hat\beta-\beta)'X'X(\hat\beta-\beta)$$

$$= \frac{1}{\sigma^2}SSE + \underbrace{(\hat\beta-\beta)'\left[\sigma^2(X'X)^{-1}\right]^{-1}(\hat\beta-\beta)}_{\sim \chi^2(k+1) \text{ by } 4}$$

And by 5 these 2 terms (functions of SSE & $\hat\beta$) are independent.

So by 1, $\frac{1}{\sigma^2}SSE \sim \chi^2(n-k-1)$

t-distribution for linear combinations of regression coefficients: $L = a'\hat{\beta}$

- Basis of tests and confidence/prediction intervals
- Includes single regression coefficients

Recall If $Z \sim N(0,1)$ & $W \sim \chi^2(\nu)$ are independent

$$T = \frac{Z}{\sqrt{W/\nu}} \sim t(\nu)$$ This is a definition

---

$$a'\hat{\beta} \sim N(a'\beta, \sigma^2 a'(X'X)^{-1}a), \text{ so}$$

$$Z = \frac{a'\hat{\beta} - a'\beta}{\sqrt{\sigma^2 a'(X'X)^{-1}a}} \sim N(0,1)$$

Independent of $W = \frac{1}{\sigma^2} SSE$, so

$$T = \frac{Z}{\sqrt{W/(n-k-1)}} \sim t(n-k-1)$$

$$\|$$

$$\frac{\dfrac{a'\hat{\beta} - a'\beta}{\sqrt{\sigma^2 a'(X'X)^{-1} a}}}{\sqrt{\dfrac{SSE}{\sigma^2 (n-k-1)}}} = \frac{a'\hat{\beta} - a'\beta}{s\sqrt{a'(X'X)^{-1} a}}$$

save

see p. 205 for this notation

where $s^2 = \dfrac{SSE}{n-k-1}$ ALSO called MSE

In general MEAN SQUARE $=$ $\dfrac{\text{Sum of Squares}}{\text{Degrees of Freedom}}$

Application: Testing a single regression coefficient using $H_0: \beta_j = 0$

Recall the interpretation

$$E(Y_i \mid \underset{\sim}{x}_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_j x_{ij} + \cdots + \beta_k x_{ik}$$

$$\frac{\partial}{\partial x_{ij}} E(Y_i \mid \underset{\sim}{x}_i) = \beta_j \quad , \text{ or re-write}$$

$$E(Y_i \mid \underset{\sim}{x}_i) = \Big(\beta_0 + \sum_{\ell \neq j} \beta_\ell x_{i\ell}\Big) + \beta_j x_{ij}$$

For constant values of the other IVs, β_j is the slope of the line relating x_{ij} to $E(Y_i)$

So we say Controlling for the other variables, is X_j related to Y? (Meaning $E(Y)$)

$$T = \frac{a'\hat{\beta} - a'\beta}{s\sqrt{a'(X'X)^{-1}a}} \qquad \text{Here } a' = (0, 0, \ldots, \underset{\text{Position } j}{1}, 0 \ldots 0)$$

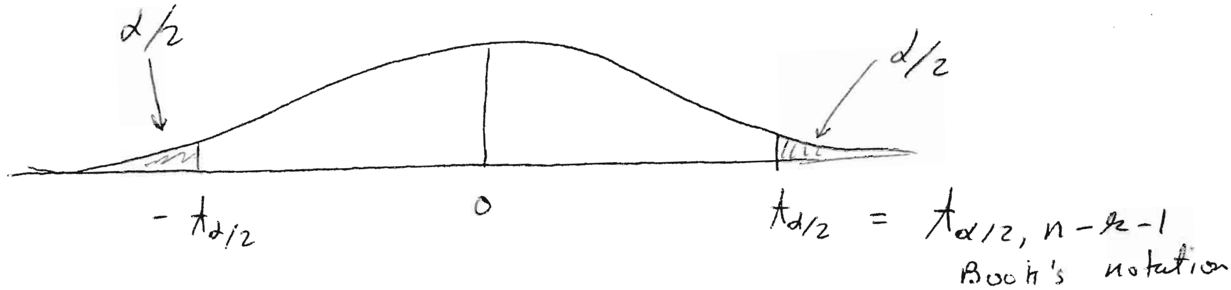
$$= \frac{\hat{\beta}_j - \beta_j}{s\sqrt{g_{jj}}} \overset{H_0: \beta_j = 0}{=} \frac{\hat{\beta}_j}{s\sqrt{g_{jj}}}$$

where $G = [g_{ij}] = (X'X)^{-1}$

Notation on p. 205

Controlling for parents' income, is education still related to income?

Confidence Intervals



$$T = \frac{\hat{\beta}_j - \beta_j}{s\sqrt{g_{jj}}} \sim t(n-k-1), \text{ so}$$

$$1-\alpha = P\left[-t_{\alpha/2} \le \frac{\hat{\beta}_j - \beta_j}{s\sqrt{g_{jj}}} \le t_{\alpha/2}\right]$$

$$\overset{\text{work}}{=} P\left[\hat{\beta}_j - t_{\alpha/2}\, s\sqrt{g_{jj}} \le \beta_j \le \hat{\beta}_j + t_{\alpha/2}\, s\sqrt{g_{jj}}\right]$$

Book says (p. 210) that this probability statement applies BEFORE the sample is selected. AFTER the sample is selected, the endpoints are no longer random.

Pretty good.

Analysis of Variance

Decomposition of SS

Without any IVs, best guess (prediction) of each observation is the sample mean $\bar{Y}$.

Total variation to explain with IVs is

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$= \sum_{i=1}^{n} (\underbrace{y_i - \hat{y}_i}_{a_i} + \underbrace{\hat{y}_i - \bar{y}}_{b_i})^2$$

$$= \underbrace{\sum (y_i - \hat{y}_i)^2}_{SSE} + 2\sum (y_i - \hat{y}_i)(\hat{y}_i - \bar{y}) + \underbrace{\sum (\hat{y}_i - \bar{y})^2}_{SSR}$$

will show this $= 0$, using $\sum (y_i - \hat{y}_i) = 0$ from first normal equation

9.11

$$\sum_{i=1}^{n}(y_i-\hat{y}_i)(\hat{y}_i-\bar{y})$$

$$=\sum_{i=1}^{n}(y_i-\hat{y}_i)\hat{y}_i-\sum_{i=1}^{n}(y_i-\hat{y}_i)\bar{y}$$

$$=\sum_{i=1}^{n}(y_i-\hat{y}_i)\hat{y}_i-\bar{y}\underbrace{\sum_{i=1}^{n}(y_i-\hat{y}_i)}_{=0}$$

NOW MATRIX NOTATION

$$=(Y-\hat{Y})'\hat{Y}=Y'\hat{Y}-\hat{Y}'\hat{Y}=Y'X\hat{\beta}-(X\hat{\beta})'X\hat{\beta}$$

$$=Y'X\hat{\beta}-\hat{\beta}'\underbrace{X'X(X'X)^{-1}}_{I}X'Y$$

$$=Y'X\hat{\beta}-[(X'X)^{-1}X'Y]'X'Y$$

$$=Y'X(X'X)^{-1}X'Y-Y'X(X'X)^{-1}X'Y=0$$

So

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2+\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$$

$$SST = SSE + SSR$$

To explain — unexplained — must be Explained

SSE & SSR are Independent

9.12

$SSE = (Y - X\hat{\beta})'(Y - X\hat{\beta})$ &

$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2 = (X\hat{\beta} - \underset{\sim}{1}\bar{Y})'(X\hat{\beta} - 1\bar{Y})$

Because $\sum_{i=1}^{n} \hat{Y}_i = \sum_{i=1}^{n} Y_i$, (HW)

$\bar{Y} = \frac{1}{n} 1'\hat{Y} = \frac{1}{n} 1' X\hat{\beta}$, and

$X\hat{\beta} - \underset{\sim}{1}\bar{Y} = X\hat{\beta} - \underbrace{\underset{\sim}{1}}_{n \times 1} \frac{1}{n} \underbrace{\underset{\sim}{1}'}_{1 \times n} \underbrace{X}_{n \times (k+1)} \underbrace{\hat{\beta}}_{(k+1) \times 1}$, and

SSR is a function of $\hat{\beta}$

Since $\hat{\beta}$ & SSE are independent, functions of them are independent, and

SSE & SSR are independent

$SST = SSR + SSE$

Proportion of variation explained by the regression is $R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST}$

Between zero & one

- $R^2 = 0$ means $SSR = 0 = \sum(\hat{y}_i - \bar{y})^2$

  Least squares plane is just $y = \bar{y}$

- $R^2 = 1$ means $SSE = 0$, all points exactly on least squares plane

In practice, these extremes don't happen

Is R^2 bigger than you would expect by chance if regression were useless? $H_0: \beta_1 = \cdots = \beta_k = 0$

9.14

ANOVA SUMMARY TABLE

(Compare p. 188)

Source	df	SS	MS	F
Regression	k	SSR	SSR/k	$\frac{MSR}{MSE}$
Error	$n-k-1$	SSE	SSE/$n-k-1$	
Total	$n-1$	SST		

Recall $R^2 = \frac{SSR}{SST} = \frac{SSR}{SSR+SSE}$, so big F goes with big R^2

Recall if $U_1 \sim \chi^2(r_1)$, $U_2 \sim \chi^2(r_2)$ ind.

$$F = \frac{U_1/r_1}{U_2/r_2} \sim F(r_1, r_2) \quad \text{Def}$$

If $H_0: \beta_1 = \cdots = \beta_k = 0$ is true

① $Y_1, \ldots Y_n \overset{iid}{\sim} N(\beta_0, \sigma^2)$ & $\frac{\Sigma(Y_i - \bar{Y})^2}{\sigma^2} = \frac{SST}{\sigma^2} \sim \chi^2(n-1)$

② $\frac{SSE}{\sigma^2} \sim \chi^2(n-k-1)$, always

③ SSR & SSE independent: Proof omitted

④ So $\frac{SSR}{\sigma^2} \sim \chi^2(n-1-(n-k-1)) = \chi^2(k)$

⑤ And $U_1 = \frac{SSR}{\sigma^2}$, $r_1 = k$ plus $U_2 = \frac{SSE}{\sigma^2}$, $r_2 = n-k-1$

⑥ Mean Squares are ind χ^2, divided by df