Christensen p346

21] $Y \sim F(d_1, d_2) \sim F(d_1, d_2, 0)$.

Let $X_1 \sim \chi^2(d_1, \gamma)$ $\perp\!\!\!\perp$ $X_2 \sim \chi^2(d_2, 0)$.

Then $W = \dfrac{X_1/d_1}{X_2/d_2} \sim F(d_1, d_2, \gamma)$.

22] If $H_0: \underset{\sim}{\beta_2} = \underset{\sim}{0}$ is true then $\gamma = 0$.

proof) Under H_0, $E(\underset{\sim}{Y}) = X\underset{\sim}{\beta} = X_1 \underset{\sim}{\beta_1}$ and

$$2\sigma^2\gamma = (X\underset{\sim}{\beta})^T (P - P_1) X\underset{\sim}{\beta} = \underset{\sim}{\beta_1}^T X_1^T (P - P_1) X_1 \underset{\sim}{\beta_1}$$

$= 0$ since $P X_1 = X_1 = P_1 X_1$

(or $X_1^T P = X_1^T P_1 = X_1^T$).

23] $$\hat{\gamma} = \frac{\hat{\underset{\sim}{Y}}^T (P - P_1) \hat{\underset{\sim}{Y}}}{2\,MSE}.$$

More results

24] Suppose $\underset{\sim}{Y} \perp\!\!\!\perp \underset{\sim}{u} \mid \underset{\sim}{u}^T \underset{\sim}{\beta_u}$ eg

$Y_i = \beta_0 + \underset{\sim}{u_i}^T \underset{\sim}{\beta_u}$ or $Y_i = m(\underset{\sim}{x_i}^T \underset{\sim}{\beta}) + \varepsilon_i$ or GLMs,

If the $\underset{\sim}{u_i}$ are iid from an elliptically contoured distribution such as the $N_{p-1}(\underline{0}, \Sigma)$ distribution, then often the OLS estimator $\hat{\underline{\beta}}_S \xrightarrow{P} c\,\underline{\beta}_U$ for some $c \neq 0$.

25) points 80) – 83) in Exam 2 review suggest that the partial F test is a large sample test for many models of the form $Y \perp\!\!\!\perp \underline{u} \mid \underline{u}^T \underline{\beta}_U$ if $H_0: L\,\underline{\beta}_U = \underset{\text{matrix}}{0}$ is true where $\underline{\beta} = \begin{pmatrix} \beta_0 \\ \underline{\beta}_U \end{pmatrix}$ and the $\underline{u}_i$ are iid from an elliptically contoured distribution.

$A = [\,\underline{0} \quad L\,]$

26) Under regularity conditions, the partial F test is a large sample test for the AR(p) model. See 84) on Exam 2 review.

↓ exam3 materal

§4.4

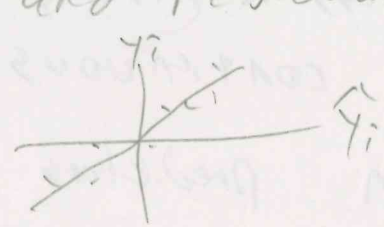
27} p111 If the linear model has a constant β_0 then the coefficient of multiple determination $R^2 = [\text{corr}(Y_i, \hat{Y}_i)]^2$

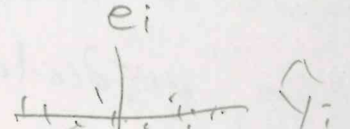
where $\text{corr}(Y_i, \hat{Y}_i)$ = sample correlation of Y_i and $\hat{Y}_i$.

28} If there is no intercept in the model, 27} is just one of several defintions for R^2. So R^2 will depend on the computer package.

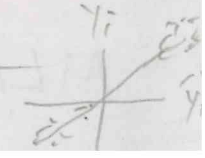
29} i) $0 \le R^2 \le 1$ but small R^2 does not imply that the linear model is bad.

ii) R^2 does not have much meaning if the response and residual plots do not look good





iii) R^2 is too high if n is small, p is close to n, or there are 2 or more separated clusters in the response plot



30) p111 total sum of squares (corrected for mean)

$$SSTO = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

regression sum of squares

$$SSR = \sum_{i=1}^{n} (\hat{Y}_i - \bar{Y})^2$$

residual or error sum of squares

$$SSE = RSS = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} e_i^2$$

If β_0 is in the model, $SSTO = SSE + SSR$.

31) If β_0 is in the model and $SSTO \neq 0$

p111

$$R^2 = \frac{SSR}{SSTO} = 1 - \frac{SSE}{SSTO}.$$

32) $\underline{Y} = X\underline{\beta} + \underline{\epsilon}$ is a <u>linear model</u>

i) multiple linear regression (MLR) at least one predictor is continuous

ii) experimental design predictors are coded often 0's and 1's or -1's and 1's

MLR output has the ANOVA table for

the ANOVA F test : Ho $\beta_1 = \cdots = \beta_{p-1} = 0$

Ho the nontrivial predictors are not needed in the MLR model so

$Y_i = \beta_0 + \varepsilon_i$ (location model with LS estimator $\hat{\beta}_0 = \overline{Y}$)

ANOVA table

source		df	SS	MS	F	pvalue
model	regression	p-1	SSR	MSR	$F_0 = \frac{MSR}{MSE}$	
error	residual	n-p	SSE	MSE		

MS = SS/df

33) If Ho above is true and $Y_i \sim N(\beta_0, \sigma^2)$,

then $E(R^2) = \frac{p-1}{n-1}$ and $V(R^2) = \frac{2(p-1)(n-p)}{(n-1)^2(n+1)}$

p. 113

ex) n = 100, p = 51 $\rightarrow$ $ER^2 = .5$ when none of the predictors are good.

34) $R^2 = 1$ if $n = p$ interpolation so $Y_i = \hat{Y}_i$

not a good model

want $n \geq 10p$ so $ER^2 \approx \frac{p-1}{10p-1} \approx .1$ when

Ho of 32) is true.

35) p112 For the Anova F test with a constant in the model, the test statistic

$$F_A = \frac{R^2}{1-R^2} \frac{n-p}{p-1}$$

$F_A \sim F_{p-1, n-p}$ if H_0 is true and $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

$F_A \overset{\sim}{\sim} F_{p-1, n-p}$ if H_0 is true and the zero mean iid ϵ_i are from a large class of error distributions in that the Anova F test is a large sample test.

proof) The reduced model is $\underset{\sim}{Y} = \underset{\sim}{1} \beta_0 + \underset{\sim}{\epsilon}$

with $\hat{\beta}_0 = \overline{Y}$ so $RSS(R) = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 = SSTO$

$$\text{so } F_A = F_R = \frac{(RSS(R) - RSS(F))/(p-1)}{RSS(F)/(n-p)}$$

$$= \frac{SSTO - SSE}{SSE} \frac{n-p}{p-1} = \frac{SSR}{SSE} \frac{n-p}{p-1}$$

and $\frac{R^2}{1-R^2} = \frac{SSR/SSTO}{SSE/SSTO} = \frac{SSR}{SSE}$. $\square$

36) R^2 does not decrease and usually increases as predictors are added to the linear model.

Skip §4.5, 4.6, 4.7

Ch 5 §5.1

1) If $\hat{\underset{\sim}{\theta}} \sim AN_p(\underset{\sim}{\theta}, \Sigma_n)$ then

$\underset{\sim}{a}'\hat{\underset{\sim}{\theta}} \sim AN_1(\underset{\sim}{a}'\underset{\sim}{\theta}, \underset{\sim}{a}'\Sigma_n\underset{\sim}{a})$ and a large sample $100(1-\delta)\%$ confidence interval (CI) for $\underset{\sim}{a}'\underset{\sim}{\theta}$ is $\underset{\sim}{a}'\hat{\underset{\sim}{\theta}} \pm t_{d_n, 1-\delta} \, SE(\underset{\sim}{a}'\underset{\sim}{\theta})$

$= \underset{\sim}{a}'\hat{\underset{\sim}{\theta}} \pm t_{d_n, 1-\delta}\sqrt{\underset{\sim}{a}'\hat{\Sigma}_n\underset{\sim}{a}}$ where $d_n \to \infty$ as $n \to \infty$.

2) Full rank OLS model $\hat{\underset{\sim}{\beta}} \sim AN_p(\underset{\sim}{\beta}, MSE(X^TX)^{-1})$.

proof sketch

OLS CLT $\sqrt{n}(\hat{\underset{\sim}{\beta}} - \underset{\sim}{\beta}) \xrightarrow{D} N_p(\underset{\sim}{0}, \sigma^2 W)$ where

$$\frac{\underline{X}^T\underline{X}}{n} \to \underline{W}^{-1} \qquad \text{so } \hat{\underline{W}}^{-1} = \frac{\underline{X}^T\underline{X}}{n}$$

and $\hat{\underline{W}} = n(\underline{X}^T\underline{X})^{-1}$. so

$$\sqrt{n}(\hat{\underline{\beta}} - \underline{\beta}) \sim AN_p[\underline{0}, MSE\, n(\underline{X}^T\underline{X})^{-1}]$$

$$\hat{\underline{\beta}} - \underline{\beta} \sim AN_p[\underline{0}, MSE(\underline{X}^T\underline{X})^{-1}]$$

and $\hat{\underline{\beta}} \sim AN_p[\underline{\beta}, MSE(\underline{X}^T\underline{X})^{-1}]$ *

3) so $\underline{a}'\hat{\underline{\beta}} \sim AN_1[\underline{a}'\underline{\beta}, MSE\, \underline{a}'(\underline{X}^T\underline{X})^{-1}\underline{a}]$

and a large sample $100(1-\delta)\%$ CI for $\underline{a}'\underline{\beta}$ is

$$\underline{a}'\hat{\underline{\beta}} \pm t_{n-p,1-\delta}\sqrt{MSE\, \underline{a}'(\underline{X}'\underline{X})^{-1}\underline{a}}$$

where $P(T \le t_{n-p,1-\delta}) = 1-\delta$ if $T \sim t_{n-p}$.

(Text uses $t_{n-p}^{\frac{\alpha}{2}}$ where $P(T > t_{n-p}^{\alpha/2}) = \frac{\alpha}{2}$ and $\alpha = \delta$.)

4) p119 A large sample $100(1-\delta)\%$ CI for β_i uses $\underset{\sim}{a}^T = (0, \ldots, 0, 1, 0, \ldots, 0)$
$i+1$ position

for $i = 0, \ldots, p-1$.

Then $\underline{a}^T (X^T X)^{-1} \underline{a}$ is the $(i+1)$th diagonal element $\underbrace{d_{ii}}$ of $(X^T X)^{-1}$. So the CI
awkward notation
$(i+1)$th diagonal element, $i = 0, \ldots, p-1$.

is $\hat{\beta}_i \pm t_{n-p, 1-\delta} \sqrt{MSE} \sqrt{d_{ii}}$.

5) p119-120 Suppose there are k $100(1-\delta_S)\%$ CIs, (single CI) and want the overall familywise confidence (probability before gathering data) that all k CIs contain their θ_i. Want $100(1-\delta_T)\%$ confidence that all k CIs contain their θ_i.

Here $\delta_T = P$ [at least one of the k CIs does not contain its θ_i, before gathering data] is called the family wise error rate,

Can get procedures where

$1-\delta_T \geq 1-\delta$

Note: assume $\epsilon \sim N_n(0, \sigma^2 I)$ for 6} - 9}.

6} p121 Bonferroni t intervals

take $\delta_S = \frac{\delta}{k}$. Then

$1-\delta_T \geq 1 - k\frac{\delta}{k} = 1-\delta.$

So if $k=5$ and $\delta = 0.1$ use $\delta_S = \frac{0.1}{5} = 0.02$

So make 5 98% CIs to ensure

that they all contain their θ_i with

at least 90% confidence.

Problem: CIs are too wide if k is large.

7} could have $k=p$ and want CIs

for $\beta_0, \beta_1, \ldots, \beta_{p-1}$
$\theta_1 \quad \theta_2 \quad \theta_p$

8) Scheffé's method

a) Let $\underset{\sim}{A}$ be a $d \times p$ matrix of rank d, let $\underset{\sim}{\underset{d\times1}{\Phi}} = A\underset{\sim}{\beta}$. Then you can make as many CIs for $\underset{\sim}{h}'\underset{\sim}{\Phi}$ as you want (even after looking at the data), with familywise error rate $\geq 1-\delta$, if the CIs are

$$\underset{\sim}{h}'\hat{\underset{\sim}{\Phi}} \pm \sqrt{d\, F_{d,n-p,1-\delta}}\ \sqrt{MSE}\ \sqrt{\underset{\sim}{h}' A (X'X)^{-1} A' \underset{\sim}{h}}$$

where $P\left(F_{d,n-p} \leq F_{d,n-p,1-\delta}\right) = 1-\delta$.

If $A' = [\underset{\sim}{a_1}\ \underset{\sim}{a_2} \cdots \underset{\sim}{a_d}]$, each $\underset{\sim}{a_i}'\underset{\sim}{\beta}$ is a $\underset{\sim}{h}'\underset{\sim}{\Phi}$.

proof take $\underset{\sim}{h} = (0,\ldots,0,1,0,\ldots,0)^T$. Then $\underset{\sim}{\Phi} = A\underset{\sim}{\beta} = \begin{pmatrix} a_1'\beta \\ a_2'\beta \\ \vdots \\ a_d'\beta \end{pmatrix} = \begin{pmatrix} \phi_1 \\ \vdots \\ \phi_d \end{pmatrix}$

9) If $d=k=p$ and $A=I_p$ then $\underline{\Phi}=\underline{\beta}$ and Scheffe's CIs for $\underline{h}'\underline{\beta}$ have the form

$$\widehat{\underline{h}'\underline{\beta}} \pm \sqrt{p\, F_{p,n-p,1-\delta}}\ \sqrt{MSE}\ \sqrt{\underline{h}'(X'X)^{-1}\underline{h}}$$

have familywise confidence $1-\delta_T \geq 1-\delta$.

10) 6)–9) assume normality, but the large sample familywise confidence $1-\delta_{T,n} \to 1-\delta_T \geq 1-\delta$ as $n \to \infty$ for a large class of 0 mean iid error distributions.

11) p124 If $d=k=p$, Scheffe's CIs are longer than Bonferroni CIs. If k is a lot bigger than d, then Scheffe's CIs are shorter. Scheffe's CIs allow data snooping: you can decide on the $\underline{h}'\underline{\beta}$ to use after getting the data. usually need to decide on the $\underline{h}'\underline{\beta}$ before

getting the data.

12} Let $\underset{d\times 1}{\underset{\sim}{\Phi}} = \underset{d\times p}{A}\,\underset{p\times 1}{\underset{\sim}{\beta}}$. A large sample $100(1-\delta)\%$ <u>confidence region</u> C_n for $\underset{\sim}{\Phi}$ is a set such that

$$P(\underset{\sim}{\Phi} \in C_n) \to 1-\delta \text{ as } n \to \infty.$$

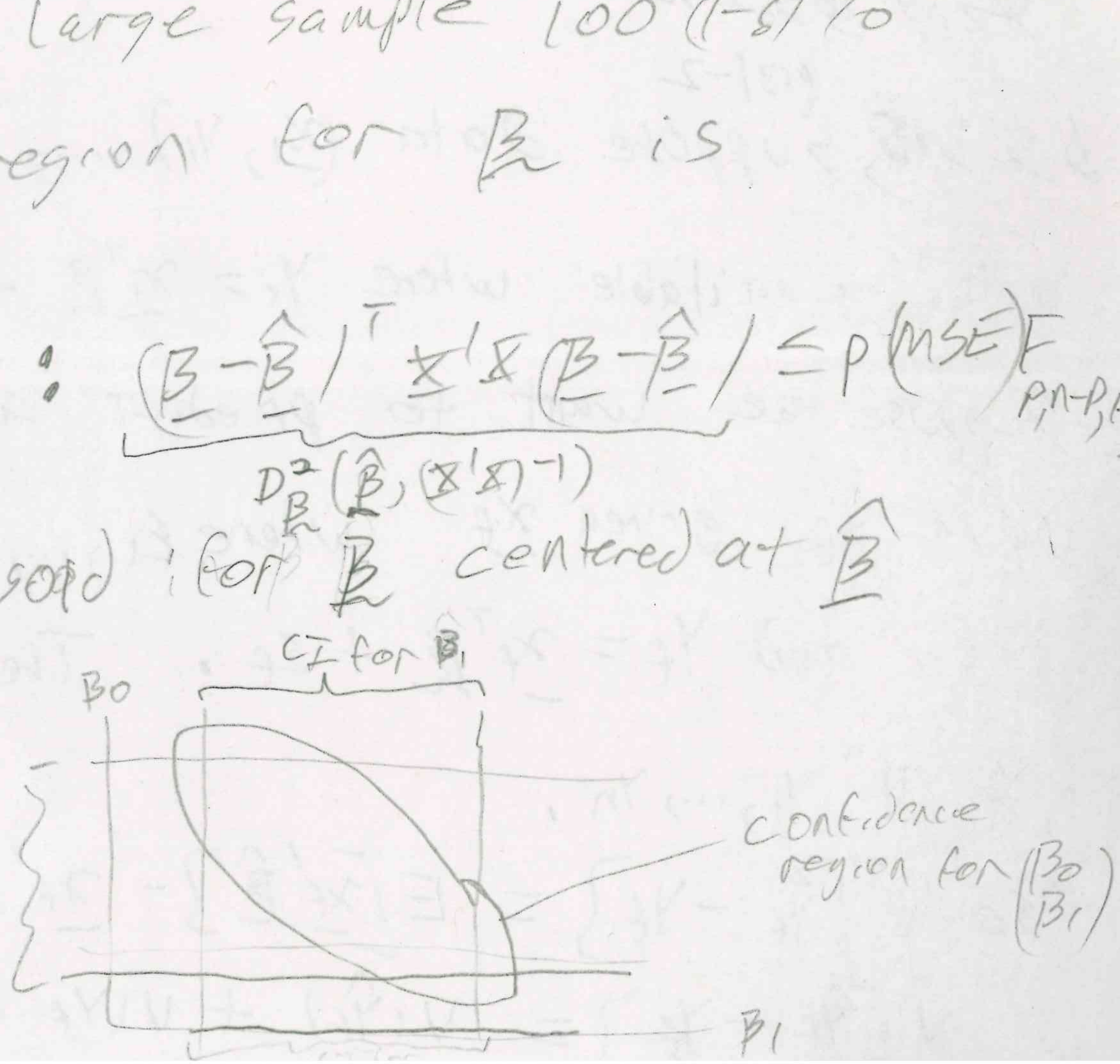
13} Often $d = r = p$ so $\underset{\sim}{\Phi} = \underset{\sim}{\beta}$ and $A = I$.

Then a large sample $100(1-\delta)\%$ confidence region for $\underset{\sim}{\beta}$ is

$$C_n = \left\{ \underset{\sim}{\beta} : \underbrace{(\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}})^T X'X (\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}})}_{D^2_{\underset{\sim}{\beta}}(\hat{\underset{\sim}{\beta}},\, (X'X)^{-1})} \le p\,(MSE)\, F_{p,n-p,1-\delta} \right\}$$

a hyperellipsoid for $\underset{\sim}{\beta}$ centered at $\hat{\underset{\sim}{\beta}}$

ex p128



$Y = \underset{\sim}{x}'\underset{\sim}{\beta} + \varepsilon$

14) St §5.2 / p129 A large sample $100(1-\delta)\%$ CI

for $E[Y]_x = E[Y|\underset{\sim}{x}] = \underset{\sim}{x}'\underset{\sim}{\beta}$ is

$$\underbrace{\underset{\sim}{x}'\hat{\underset{\sim}{\beta}}}_{\hat{Y}} \pm t_{n-p,1-\delta}\sqrt{MSE}\sqrt{\underbrace{\underset{\sim}{x}'(X'X)^{-1}\underset{\sim}{x}}_{h_{\underset{\sim}{x}} = h_i \text{ if } \underset{\sim}{x} = \underset{\sim}{x}_i}}.$$

want $h_{\underset{\sim}{x}} \le \max h_i$ so there is no extrapolation.

Skip §5.2.2

p131-2

§5.3 15) Suppose data $(\underset{\sim}{x}_1, Y_1), \dots, (\underset{\sim}{x}_n, Y_n)$

are available where $Y_i = \underset{\sim}{x}_i^T\underset{\sim}{\beta} + \varepsilon_i$.

Suppose we want to predict a future value Y_f given $\underset{\sim}{x}_f$ where $\varepsilon_1, \dots, \varepsilon_n, \varepsilon_f$ are iid, and $Y_f = \underset{\sim}{x}_f^T\underset{\sim}{\beta} + \varepsilon_f$. Then

$Y_f \perp\!\!\!\perp Y_1, \dots, Y_n$.

so $E[\hat{Y}_f - Y_f] = E[\underset{\sim}{x}_f'\hat{\underset{\sim}{\beta}}] - \underset{\sim}{x}_f'\underset{\sim}{\beta} = 0$

$V(\hat{Y}_f - Y_f) = V(\hat{Y}_f) + V(Y_f)$ since