Jack Podori
STAT 6008
Final Exam
5/6/22

* Sorry, I didn't realize we were supposed to print out the final. I also don't have a printer at home.

Part I: Multiple Choice

(1) a
(2) c
(3) c
(4) c
(5) d
(6) d
(7) a
(8) d
(9) d
(10) b

Part II: Multiple select

(11) (b, e)
(12) c

(13) (b, c, d, e)

Part III : Long Answer

14.) (a) Full $y_i = \beta_0 + \beta_1 \text{Youth} + \beta_2 \text{EDU} + \beta_3 \text{Expenditure Year 0} + \beta_4 \text{Labour Force} + \beta_5 SS + \beta_6 W + \epsilon_i$

Reduced: $y_i = \beta_0 + \beta_1 \text{Youth} + \beta_2 \text{ExpYear 0} + \epsilon_i$

- $H_0: \beta_2 = \beta_4 = \beta_5 = \beta_6 = 0$
- H_a: Not all $\beta_i = 0$ $(i \in \{2,4,5,6\})$

$$F = \frac{(RSS_{reduced} - RSS_{Full}) / (df_{reduced} - df_{Full})}{(RSS_{Full} / df_{Full})}$$

$RSS = (\text{Residual SE})^2 * df_{E}$

- $RSS_{Red} = (20.53)^2 \cdot 44 = 18545.1596$
- $RSS_{Full} = (21.15)^2 \cdot 40 = 17892.9$
- $RSS_F / df_{Full} = 17892.9/40 = 447.3225$
- $df_{Red} - df_{Full} = 4$

$$F^* = \frac{(18545.1596 - 17892.9)/4}{447.3225} \approx 0.3645$$

- $P(F_{0.05,4,40} > 0.3645) = 0.8322904$
- We fail to reject the null, which means we don't have evidence that the full model is better at explaining the variability in our response variable than the reduced model.

(b) Yes, looking at the pairs plot and the sample correlation matrix, there is some evidence that there would be problems w/ multicolinearity if we fit the full model. We can see, looking at both the pairs plot & sample correlation matrix that some of our EVs specifically the pairs (Youth, Wage), (Expenditure Year 0, Wage) have strong linear relations to eachother and pairs (Youth, Edu), (Youth, Expenditure Year 0) (education, Labour Force), (education, wage) all have moderately strong linear relations w/ one another.

Part III (continued)

14.) (continued)

(c) According to the BIC criterion, we should choose the model that includes the $3 = p+1$ coefficients w/ the two EVs being Youth & ExpenditureYear0. This is the model that minimizes BIC.

(d) Yes, for the model w/ two EV's Lasso and BIC select the same model. The reason for this is that for a fixed model size p, all our variable selection methods will choose the same variables to include in the model. The differences between our variable selection methods are where the algorithms "stop" (how many vars the choose to include).

15.)

(a) $H_0: B_2 = B_4 = B_5 = B_6 = 0$

H_a: Not all $B_i = 0$ $(i \in \{2, 4, 5, 6\})$

Test stat: Residual Deviance$_{Red}$ − Residual Deviance$_{Full}$

$= 47.579 - 45.259 = 2.32$

$$\boxed{P(\chi^2_{0.05,4} > 2.32) = 0.677302}$$

- We fail to reject the null. We don't have evidence at the $\alpha = 0.05$ level that the full model is better at explaining the variability in our response variable (in this case the log odds of the crime rate being $> \frac{1 \text{ crime}}{10{,}000 \text{ people}}$) than the reduced model.

15.) (continued)

(b) Using the reduced model, compute the estimated probability that Y > 100 when youth = 130 & ExpYearO = 60.

$$\log\left(\frac{\theta}{1-\theta}\right) = -15.01848 + 0.06812(\text{Youth}) + 0.06809(\text{ExpYearO})$$

$$\log\left(\frac{\theta(130,60)}{1-\theta(129,60)}\right) = -15.01848 + 0.06812(130) + 0.06809(60)$$

$$\log\left(\frac{\theta(x)}{1-\theta(x)}\right) = -2.07748$$

$$\frac{\theta(x)}{1-\theta(x)} = e^{-2.07748} \iff \theta(x) = e^{-2.07748} - \theta(x)e^{-2.07748}$$

$$\iff \theta(x) + \theta(x)e^{-2.07748} = e^{-2.07748}$$

$$\iff \theta(x) = e^{-2.07748} / (1+e^{-2.07748})$$

$$\Rightarrow \boxed{\theta(x_1 = 130, x_2 = 60) = 0.111305}$$

(c) Using the reduced model: Compute a 95% CI for the odds when comparing two states w/ Youth = 130, one w/ ExpYearO = 62 & the other w/ ExpYearO = 60.

- $\log\left(\frac{\theta(130,62)}{1-\theta(130,62)}\right) = -1.9413$
- $\log\left(\frac{\theta(130,60)}{1-\theta(130,60)}\right) = -2.07748$

[• Note: • this is the same as a 95% CI for $2\hat{\beta}_2$:
• $var(2\hat{\beta}_2) = 4var(\hat{\beta}_2) \Rightarrow SE(2\hat{\beta}_2) = 2SE(\hat{\beta}_2)$]

- $\log(OR) = -1.9413 - (-2.07748) = 0.13618$
- 95% CI: $\log(OR) = 0.13618 \pm 1.96(2 * 0.02152)$
$= (0.0517824, 0.2205776)$

• exponentiating endpts to get odds:

$$\boxed{95\%\ CI\ (OR):\ (1.053,\ 1.247)}$$

1(a) $Y_t = \beta_0 + \beta_1 X_t + e_t$ w/ $e_t = \rho e_{t-1} + v_t$; $v_i \overset{iid}{\sim} N(0, \sigma_v^2)$

(a) Show that $var(e_t) = \sigma_e^2 = \frac{\sigma_v^2}{1-\rho^2}$

- $\sigma_e^2 = var(e_t) = E[e_t^2] - \cancel{E[e_t]}^2 = E[e_t^2] = \sigma_e^2$

$= E[(\rho e_{t-1} + v_t)^2] = E[\rho^2 e_{t-1}^2] + E[v_t^2] + 2\rho \cancel{E[e_{t-1}]}\cancel{E[v_t]}$

$= \rho^2 E[e_{t-1}^2] + E[v_t^2]$ (Note: $var(v_t) = E[v_t^2] - \cancel{E[v_t]}^2 \Rightarrow E[v_t^2] = \sigma_v^2$)

$= \rho^2 \sigma_e^2 + \sigma_v^2$

$\sigma_e^2 = \rho^2 \sigma_e^2 + \sigma_v^2 \iff \sigma_e^2 - \rho^2 \sigma_e^2 = \sigma_v^2$

$\iff \boxed{\sigma_e^2 = \sigma_v^2/(1-\rho^2)}$

(b) Show that $corr(e_t, e_{t-1}) = \rho$.

- $cor(e_t, e_{t-1}) = \frac{cov(e_t, e_{t-1})}{\sqrt{var(e_t)\, var(e_{t-1})}} = \frac{cov(e_t, e_{t-1})}{\sigma_e^2}$

- $cov(e_t, e_{t-1}) = E[e_t \cdot e_{t-1}] - \cancel{E[e_t]}E[e_{t-1}]$

$= E[(\rho e_{t-1} + v_t) e_{t-1}]$

$= E[\rho e_{t-1}^2 + v_t e_{t-1}]$

$= \rho E[e_{t-1}^2] + \cancel{E[v_t]}E[e_{t-1}]$

$= \rho \sigma_e^2$

$$\boxed{\bullet\ corr(e_t, e_{t-1}) = \frac{cov(e_t, e_{t-1})}{\sigma_e^2} = \frac{\rho \sigma_e^2}{\sigma_e^2} = \rho}$$

(c) The confidence intervals for β_1 would not be valid. If we were to ignore correlation the standard error of $\hat{\beta}_1$ will be wrong. If we have positive correlation, $\rho > 0$, then our estimate of the $SE(\hat{\beta}_{1LS})$ will be smaller that it should be which will mean the CI's, $\hat{\beta}_{1LS} \pm t^*_{n-p-1} SE(\hat{\beta}_{1LS})$, will be too narrow (e.g. the actual coverage probability of our $100(1-\alpha)\%$ CI will be less than its stated value.