**Stat 5870, section 2** HW 12 answers

1 point each question. 20 points if Q0 is with 0 points.

0. What software – Not graded (change from what was on the assignment. Done to keep points to 20).

1. Palm Beach Co votes.

a) JMP and R order (AICc): Browne2000 Buchanan96p Bush2000 Clinton96 Nader2000 Perot96 TotalReg

SAS order (AIC): Bush2000 Nader2000 Browne2000 Clinton96 Perot96 Buchanan96p TotalReg

Note: JMP users might get different best AIC model with only 6 variables: Browne2000, Buchanan96p, Bush2000, Nader2000, Perot96, TotalReg

b) Bush2000

Note: If you got a model with more terms (e.g. Bush2000, Perot96, Buchanan96p), you were probably using SAS and not asking it to save enough models.

c) Yes, the model with Nader2000 Browne2000 total2000 Clinton96 Perot96 totalreg is a reasonable alternative to the model with Bush2000. The AIC for the suggested model is within 2 of the best model (in 1a). Models with AIC (or BIC) within 2 units of the best are usually considered reasonable alternatives.

d) No, the model in 1c is not reasonable when using BIC. It is more than 4 units from the best.

Note (not needed in answer): It is not extremely unlikely, though (i.e. not more than 10 from the best).

e) 389 votes (or 388 votes).

Notes: computed by fitting the model from 1a with seven predictor variables: predicting log votes in Palm Beach County (prediction = 5.963), then exponentiating to get a predicted number of votes. Exp(5.963) is 388.7, rounded to 389.

388 also accepted for full credit (388 obtained by rounding a bit more before exponentiating).

\*\*If the best AIC model in 1a has only 6 variables(no Clinton96), you will get exp(6.1292) = 459.07, rounded to 459.

f) f3: Compute the prediction interval. You are comparing a single observation, so want the uncertainty in the prediction of a single observation.

Notes: You can not just compare the observed value to the prediction. The predicted value has uncertainty that can not be ignored. If that uncertainty is large, quite a large deviation may be quite ordinary. Since the prediction is evaluating a single observation, you need the prediction interval, not the confidence interval.

I did not ask you to compute the prediction interval. Using the 7 variable best AIC model, the 95% prediction interval is (exp 5.006, exp 6.921) = (149, 1013). This is far from including the recorded Buchanan vote (3407 votes).

g) Using the best BIC model from 1b, the predicted # Buchanan votes is 592. Also accepted: 589, 590 etc.

I would say not similar because 592 and 389 are more than 100 votes different. Similar also accepted if supported by something reasonable, e.g., 100 votes different is small compared to the observed number (3407)

Note: If you used Bush2000 Perot96 Buchanan96p (see Q 2b answer), you got 713 votes. Even more different than the AIC answer.

h) R and JMP users ( AICc value): 1a: 8.769, SAS users (AIC value): 8.497, 1b (BIC/SBC): 11.279, full: 8.325

full model (all variables) has the smallest SSE

i) R and JMP users ( AICc value): 1a: 9.842, SAS users (AIC value): 9.629, 1b (BIC/SBC): 11.672, full: 10.195

AICc (or AIC) model has the smallest PRESS statistic

j) Yes, you expect PRESS to be larger than SSE. Two possible explanations:

PRESS quantifies precision predicting new observations, which is harder than predicting observations in the data set.

PRESS uses observations twice. Using an observation to fit a model gives predictions that are closer to the observed values.

k) Model with the smallest PRESS statistic because predicting Palm Beach Co is predicting a new observation.

2) wage differences

a) – 0.222, se = 0.0486

b) –0.227, se = 0.0399

c) (R, JMP or SAS): Age, Age2, EdCode, JobClass, MaritalStatus, MetropolitanStatus, Region

d) Age, Age2, EdCode2

e) Sex: –0.225, se = 0.040 (or 0.0399)

f) Yes, all three estimates are very similar

g) No, the se for the sex only model is quite a bit larger than that for either multiple regression model.

h) (-0.304, -0.147) OR (-0.30, -0.15)

i) 20% smaller, 95% ci: (26%, 14%)

Computed as exp(-0.225) = 0.798, exp(-0.30) = 0.738, exp(-0.147) = 0.863, converted to a percent decrease by

100 ( 1- 0.798), 100 ( 1 – 0.738) and 100 ( 1-0.863)