Stat 112
Fall 24 – Mahtash Esfandiari and Jiale Han
Assessment

Question	Possible points	Actual points
One	40	_____
Two	40	_____
Three	20	_____
Total	100	_____

This is a closed book exam and you can have access to a cheat sheet written on front and back. You are not supposed to use the Internet or your phone. Any misconduct will be reported to the Dean's office and it will be dealt with based on university regulations.

Question one. Given table one, table two, and plot one, answer questions a to g. (35 points)

Definition of variables (data is about UCLA undergrads)
academicenvp = confidence in academic confidence
classcomfort= uncomfortable, somewhat comfortable, comfortable, very comfortable
welcomingenvp = feeling welcome

a) **Interpret the coefficient for very comfortable (bolded) within context. 5 points**

b) **Interpret the coefficient of welcoming (bolded) within context. 5 Points**

c) **State the question underlying plot one within context. 5 points**

d) **Interpret plot one within context– 5 points**

e. **Interpret R-squared within context. 5 points**

f. In multiple linear regression models they generally report MSE. What is MSE a measure of? 5 points

g. If you wanted to find MSE for m1, how would you do it? 5 points

Outputs for question one

Relevant Model

> model<-lm(academicenvp~classcomfort*welcomingenvp)

> summary(model)

Table one : classcomfort

uncomfortable	somewhat	comfortable	very comfortable
380	1172	2913	910

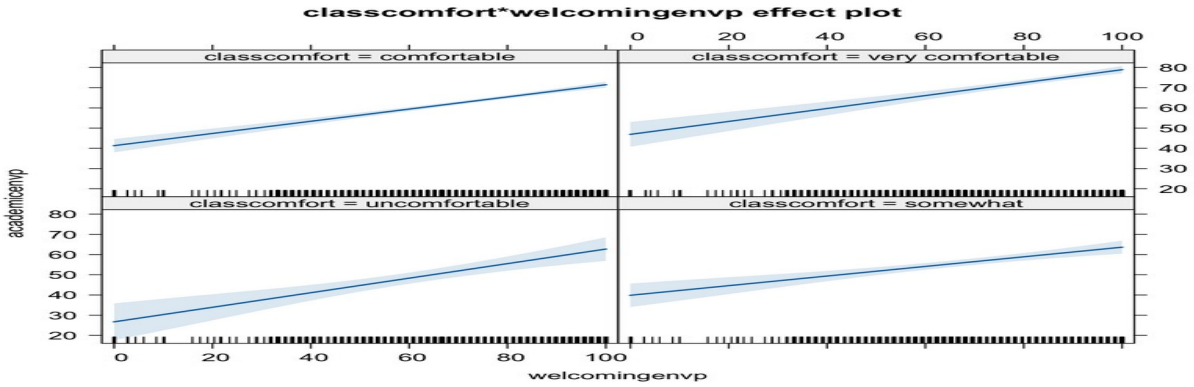
Table two: Summary table resulting from the model

Coefficient	coefficient	Standard error	t -value	Pr(>|t|
Intercept	38.70	1.63	23.74	0.000
Very comfortable	**13.88**	**3.81**	**3.64**	**0.000**
comfortable	-3.78	3.26	-1.16	0.174
Somewhat comfortable	3.52	2.56	1.36	0.174
welcomingenvp	**0.30**	**0.02**	**13.02**	**0.000**

very comfortable*welcoming	-0.01	0.05	-0.23	0.816
Comfortable*welcoming	0.07	0.07	1.50	0.133
Somewhat*welcomingp	-0.05	0.05	-1.38	0.167

Residual standard error: 13.87 on 1974 degrees of freedom
Multiple R-squared: 0.2783, Adjusted R-squared: 0.2758
F-statistic: 108.8 on 7 and 1974 DF, p-value: < 2.2e-16

Plot one: Interaction effect between class comfort and feeling welcome



Question two. Given m0, m, table one, table two, table three, plot one, plot two, answer the following questions.

Variables of the study
Divrespectp = feeling of respect for diversity (numerical)
Excluengligh = feeling of exclusion on the basis of spoken English (No, Yes)
Exclurace = feeling of exclusion on the basis of spoken race (No, Yes)
Exclucountry = feeling of exclusion based on the country of origin (No, Yes)

> m0<-glm(exclurace~1,family="binomial")
> summary(m0)
Null deviance: 1267.4 on 938 degrees of freedom
Residual deviance: 1267.4 on 938 degrees of freedom
AIC: 1269.4

m1<-
glm(exclurace~divrespectp+excluenglish*exclucountry,family="binomial")
Table one – output resulting from m1
> m1<-
glm(exclurace~divrespectp+excluenglish*exclucountry,family="binomial")

> **summary(m1)**

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)
(Intercept)	0.48185	0.30308	1.590	0.111866
divrespectp	**-0.01802**	0.00399	-4.517	6.27e-06 ***
excluenglish1	1.22049	0.21272	5.738	9.60e-09 ***
exclucountry1	0.93199	0.25404	3.669	0.000244 ***
excluenglish1:exclucountry1	-0.06255	0.41573	-0.150	0.880396

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null deviance: 1267.4 on 938 degrees of freedom
Residual deviance: 1130.3 on 934 degrees of freedom

Question 2.1

Using m0 and m1, if you were the TA for this class, how would you explain the difference between null residual and model residual. 6 points.

_____

_____

_____

Question 2.2 Given m1:

a. **What is the research question underlying m1? Please be within context. 5 points**

_____

_____

b. **Interpret the 95% confidence interval of the odds for exclusion on the basis of spoken English WITHIN CONTEXT. (see table one – it is bolded) – 5 points**

Question 2.3

a. **If you were to draw the plot of odds, which interval would you expect not to include one and why? 5 points**

---

b. **Interpret 0.84 within context. 5 points**
> exp(-0.01802*10) – notice that -0.01802 is from table one (it is bolded)
[1] 0.84

Table two. Table of odds, 95% confidence interval for odds, and P-value for the coefficient of m1

Predictor	odd	95% CI for odd	p-value
Respect for diversity	0.98	0.97-0.99	0.000
Exclusion on the basis of spoken English (Yes)	**3.39**	**2.24-5.17**	**0.000**
Exclusion on the basis of country of origin (Yes)	2.54	1.54 -4.19	0.000
Interaction between exclusion based on spoken language and country of origin	0.94	0.41-2.14	0.880

Table three: The actual and the predicted (based on m1) number of UCLA students who felt exclusion on the basis of race.

Predicted value	Actual frequency of feeling excluded on the basis of race	
	No	Yes
No	477	213
Yes	82	167

Question 2.3 Given table three

a. **Compute accuracy of the model. Show your calculations. Find the final answer. 7 points**

b. **Did the model do better with sensitivity or specificity? Choose one. 7 points**

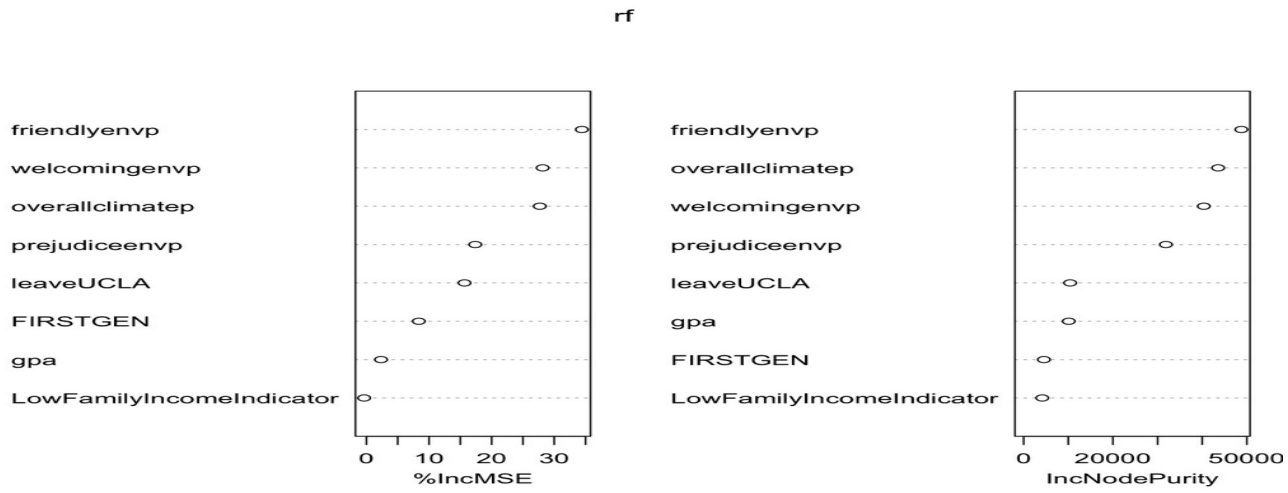
Sensitivity _____

Specificity _____
Explain why?

Question three: Given the following plot one, output one, and output two, answer the following questions. **15 points**

Variables in the study:
Academicenvp = academic confidence (numerical) – **This is the outcome**
Friendlyenvp = friendliness of our environment (numerical)
Overallclimatep = overall UCLA climate (numerical)
Prejudicenvp = feeling of prejudice (numerical)
Firtgen = (yes,no)
gpa(categorical – 4 levels)
lowfamilyincomeindicator (Yes,no)

Plot one



Model one

> model one<-
lm(academicenvp~friendlyenvp+welcomingenvp+overallclimatep+prejudiceenvp
+leaveUCLA+FIRSTGEN+gpa+LowFamilyIncomeIndicator)
> summary(m1)
Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	26.26467	4.50794	5.826	6.97e-09	***
friendlyenvp	0.18798	0.03772	4.983	7.01e-07	***
welcomingenvp	0.14510	0.02203	6.586	6.31e-11	***
overallclimatep	0.24984	0.03106	8.044	1.80e-15	***
prejudiceenvp	-0.13607	0.03403	-3.999	6.68e-05	***
leaveUCLAyes	-5.99603	1.01437	-5.911	4.23e-09	***
FIRSTGENyes	1.49963	0.80724	1.858	0.0634	.
gpa3 - 3.49	0.81395	1.01130	0.805	0.4210	
gpa3.5 and above	1.26818	1.00601	1.261	0.2077	
gpaBelow 2.49	-1.81451	1.65066	-1.099	0.2718	

LowFamilyIncomeIndicatorNot Low Income -1.76750 0.77539 -2.279 0.0228 *
Residual standard error: 13 on 1449 degrees of freedom
Multiple R-squared: 0.3808, Adjusted R-squared: 0.3765,

Model two

```
> modeltwo<-
lm(academicenvp~friendlyenvp+welcomingenvp+overallclimatep+prejudiceenvp
+leaveUCLA)
> summary(m6)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	26.31595	4.44605	5.919	4.04e-09 ***
friendlyenvp	0.18508	0.03783	4.893	1.11e-06 ***
welcomingenvp	0.14032	0.02199	6.381	2.35e-10 ***
overallclimatep	0.25485	0.03111	8.191	5.62e-16 ***
prejudiceenvp	-0.12879	0.03405	-3.783	0.000162 ***
leaveUCLAyes	-6.28320	1.00773	-6.235	5.91e-10 ***

Residual standard error: 13.05 on 1454 degrees of freedom
Multiple R-squared: 0.3739, Adjusted R-squared: 0.3717

a) **What does plot one show? 5 points**

b) **What do you conclude from it? 5 points**

c) **If you were the statistician on this project, would you recommend model one or model two? Choose one. 10 points**

Model one _____ Model two _____

Give two reasons why you choose one model one over model two or model two over model one.

_____

_____

_____