Chris Kalra

Econ 610

Homework 11

1a) Because the baseline probability for any team to win a game (β0) if neither team is a favorite to win (that is, *spread*=0) is 0.5. Therefore, any team labeled as the “favorite” when *spread*=0 has a probability of 0.5 of winning

1b) 

Hypothesis test: H0: β0 = 0.5, and HA: β0 ≠ 0.5



This t value results in a p-value less than .01, so we conclude that β0 is statistically significantly different from 0.5 at the 1% level of significance

1c) With a p-value of less than .01, *spread* is statistically significant at the 1% level. If *spread* = 10, then  (*favwin=1 | spread = 10)* = 0.57965 + 10\*(0.01937) = 0.7706

1d) 

Notice that if *spread* = 0, then the regression equation simplifies down to . However, we can see in from the table that the z value associated with  is -0.102, along with a p-value of 0.918, so we fail to reject the null hypothesis (H0: β0 = 0), and thus conclude it is reasonable for β0 to be = 0

1e) We find that the probability under the probit model is 0.8196, which is noticeably greater than our estimated probability under the LPM

1f) We have 3 restrictions, which is why our degrees of freedom is 3. We find that the p-value from the  is 0.6061, with a test statistic of 1.841. Therefore, we fail to reject the null hypothesis, and thus conclude that *favehome, fav25,* and *und25* are all jointly statistically insignificant, meaning that once *spread* has been taken into account, these three variables do not help us to better predict the probability of *favwin* occurring

2a) 

For nonwhites, the estimated probability of loan approval is 0.7078

For whites, the estimated probability of loan approval is 0.9084

The LPM is  , which gives the same results as the probit model. This is because, in the data, the proportion of nonwhites approved for loans is .7078, and the proportion of whites approved for loans is .9084

2b) Even with all of these variables taken into account, *white* is still extremely statistically significant, with a p-value of less than .001, showing strong evidence of discrimination against nonwhites

2c) With the probit model, the estimate of the coefficient on *white* is 0.520254. With the probit model, the estimate of the coefficient on white is 0.93776, which is about 1.8x the size. Likewise, the standard error is about 1.785x times the size as well

2d) For the probit model, the marginal effect of *white* is 0.0864, which means ceteris paribus, someone who is white has a 0.0864 greater probability of being approved for a loan than someone who is nonwhite

For the logit model, the marginal effect of *white* is 0.0828, which means ceteris paribus, someone who is white has a 0.0828 greater probability of being approved for a loan than someone who is nonwhite

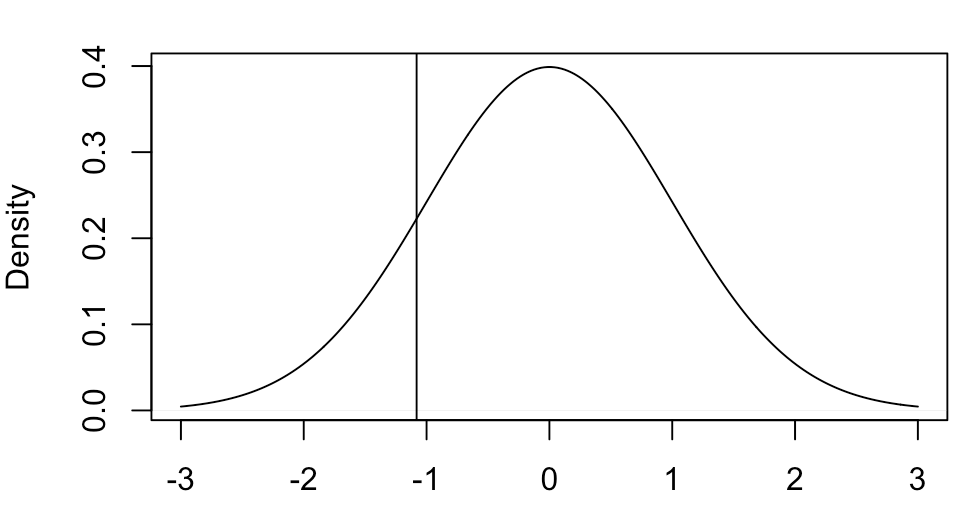
3a) For each additional mailing per year to individual i, the probability that individual i will respond increases by 0.062

3b) 

3c) Logit probabilities are approximately 1.6 times larger than Probit probabilities, due to the logistical distribution having heavier tails than the normal distribution (the probit model uses the normal distribution)

3d) *resplast* and *propresp* are extremely jointly statistically significant, with a Chi-square test statistic of 880.69 on 2 degrees of freedom, resulting in a p-value less than .0001

3e) The probability is approximately .1398, with its cutoff value at –1.081



Where the shaded area represents the probability

3f) Because I am not a sadist, I would use the “margins” package. If my GLM in R was called “*regr3f*”, the command would be *>margins(regr3f)*. It would need to be coded as a GLM with the proper link (either *probit* or *logit*, whichever I would be interested in)

If I wanted to do it by hand, I would have to plug each value into the APE equation:



Where g() is the pdf of either the logistic or probit model

3g) For either case, it would be: 

Where g() is the pdf of either the logistic or probit model. Furthermore, we would need to plug in the  values, which change between the logit and probit model, in order to find the values

3h) i) They are both bounded between 0 and 1, unlike linear probability models, which can give probabilities below 0 or greater than 1

ii) We have to worry about heteroscedasticity with linear probability models, but it is not of concern with logit or probit models

*Code (italics) and* Output (Monaco)

1b) *vegas=data*

*regr1b=lm(favwin~spread, data=vegas)*

*summary(regr1b)*

Call:

lm(formula = favwin ~ spread, data = vegas)

Residuals:

Min 1Q Median 3Q Max

-0.9836 -0.1192 0.1519 0.3069 0.4037

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.576949 0.028235 20.434 < 2e-16 \*\*\*

spread 0.019366 0.002339 8.281 9.32e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4017 on 551 degrees of freedom

Multiple R-squared: 0.1107, Adjusted R-squared: 0.1091

F-statistic: 68.57 on 1 and 551 DF, p-value: 9.324e-16

*2\*(1-pnorm(2.725340889))*

[1] 0.006423513

1d) *regr1d = glm(favwin ~ spread, family=binomial(link="probit"), data=vegas)*

*summary(regr1d)*

Call:

glm(formula = favwin ~ spread, family = binomial(link = "probit"),

data = vegas)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.69141 0.09976 0.45805 0.83300 1.12244

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) -0.01059 0.10349 -0.102 0.918

spread 0.09246 0.01212 7.631 2.32e-14 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 605.50 on 552 degrees of freedom

Residual deviance: 527.12 on 551 degrees of freedom

AIC: 531.12

Number of Fisher Scoring iterations: 5

1e) *pnorm(-0.01059 + 0.09246\*10)*

[1] 0.8196442

1f) *regr1f=glm(favwin~spread + favhome + fav25 + und25, family=binomial(link="probit"), data=vegas)*

*onefscore = 2\*(logLik(regr1f) - logLik(regr1d)) ; onefscore*

'log Lik.' 1.840843 (df=3)

*1-pchisq(onefscore, df=3)*

'log Lik.' 1.840843 (df=3)

2a) *loans=data*

*regr2a=glm(approve ~ white, family=binomial(link="probit"), data=loans)*

*summary(regr2a)*

Call:

glm(formula = approve ~ white, family = binomial(link = "probit"),

data = loans)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.1864 0.4384 0.4384 0.4384 0.8314

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 0.54695 0.07544 7.251 4.15e-13 \*\*\*

white 0.78395 0.08671 9.041 < 2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1480.7 on 1988 degrees of freedom

Residual deviance: 1401.8 on 1987 degrees of freedom

AIC: 1405.8

Number of Fisher Scoring iterations: 4

*pnorm(0.54695)*

[1] 0.7077935

*pnorm(0.54695 + 0.78395 )*

[1] 0.908389

*regr2aLinear=lm(approve ~ white, data=loans)*

*summary(regr2aLinear)*

Call:

lm(formula = approve ~ white, data = loans)

Residuals:

Min 1Q Median 3Q Max

-0.90839 0.09161 0.09161 0.09161 0.29221

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.70779 0.01824 38.81 <2e-16 \*\*\*

white 0.20060 0.01984 10.11 <2e-16 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3201 on 1987 degrees of freedom

Multiple R-squared: 0.04893, Adjusted R-squared: 0.04845

F-statistic: 102.2 on 1 and 1987 DF, p-value: < 2.2e-16

2b) *regr2b=glm(approve ~ white + hrat + obrat + loanprc + unem + male + married + dep + sch + cosign + chist + pubrec + mortlat1 + mortlat2 + vr, family=binomial(link="probit"), data=loans)*

*summary(regr2b)*

Call:

glm(formula = approve ~ white + hrat + obrat + loanprc + unem +

male + married + dep + sch + cosign + chist + pubrec + mortlat1 +

mortlat2 + vr, family = binomial(link = "probit"), data = loans)

Deviance Residuals:

Min 1Q Median 3Q Max

-3.1018 0.2369 0.3430 0.4884 1.9897

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 2.062330 0.316308 6.520 7.03e-11 \*\*\*

white 0.520254 0.096866 5.371 7.84e-08 \*\*\*

hrat 0.007876 0.007026 1.121 0.26227

obrat -0.027693 0.006147 -4.505 6.64e-06 \*\*\*

loanprc -1.011956 0.240265 -4.212 2.53e-05 \*\*\*

unem -0.036685 0.017680 -2.075 0.03799 \*

male -0.037000 0.109883 -0.337 0.73632

married 0.265745 0.094724 2.805 0.00502 \*\*

dep -0.049575 0.039065 -1.269 0.20443

sch 0.014648 0.095415 0.154 0.87799

cosign 0.086064 0.240886 0.357 0.72088

chist 0.585278 0.095602 6.122 9.24e-10 \*\*\*

pubrec -0.778742 0.126984 -6.133 8.65e-10 \*\*\*

mortlat1 -0.187628 0.257164 -0.730 0.46563

mortlat2 -0.494354 0.325883 -1.517 0.12927

vr -0.201062 0.081478 -2.468 0.01360 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1476.0 on 1970 degrees of freedom

Residual deviance: 1200.5 on 1955 degrees of freedom

(18 observations deleted due to missingness)

AIC: 1232.5

Number of Fisher Scoring iterations: 5

2c) *regr2c=glm(approve ~ white + hrat + obrat + loanprc + unem + male + married + dep + sch + cosign + chist + pubrec + mortlat1 + mortlat2 + vr, family=binomial(link="logit"), data=loans)*

*summary(regr2c)*

Call:

glm(formula = approve ~ white + hrat + obrat + loanprc + unem +

male + married + dep + sch + cosign + chist + pubrec + mortlat1 +

mortlat2 + vr, family = binomial(link = "logit"), data = loans)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.9549 0.2545 0.3458 0.4768 2.0827

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 3.80171 0.59467 6.393 1.63e-10 \*\*\*

white 0.93776 0.17290 5.424 5.84e-08 \*\*\*

hrat 0.01326 0.01288 1.030 0.30313

obrat -0.05303 0.01128 -4.702 2.58e-06 \*\*\*

loanprc -1.90495 0.46041 -4.138 3.51e-05 \*\*\*

unem -0.06658 0.03281 -2.029 0.04242 \*

male -0.06639 0.20642 -0.322 0.74776

married 0.50328 0.17799 2.828 0.00469 \*\*

dep -0.09073 0.07333 -1.237 0.21598

sch 0.04123 0.17840 0.231 0.81723

cosign 0.13206 0.44608 0.296 0.76720

chist 1.06658 0.17121 6.230 4.67e-10 \*\*\*

pubrec -1.34067 0.21736 -6.168 6.92e-10 \*\*\*

mortlat1 -0.30988 0.46351 -0.669 0.50378

mortlat2 -0.89468 0.56857 -1.574 0.11559

vr -0.34983 0.15372 -2.276 0.02286 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1476 on 1970 degrees of freedom

Residual deviance: 1201 on 1955 degrees of freedom

(18 observations deleted due to missingness)

AIC: 1233

Number of Fisher Scoring iterations: 5

2d) install.packages("margins")

library("margins")

summary(margins(regr2b, variables="white"))

factor AME SE z p lower upper

white 0.0864 0.0159 5.4209 0.0000 0.0552 0.1176

summary(margins(regr2c, variables="white"))

factor AME SE z p lower upper

white 0.0828 0.0151 5.4894 0.0000 0.0532 0.1124

3d) *ThreeD = 2\*(-2401.8537 + 2842.1976) ; ThreeD*

[1] 880.6878

*1-pchisq(ThreeD, df= 2)*

[1] 0

3e) *pnorm(-1.807 + .195\*1 + .267\*0 + 2.124\*.25)*

[1] 0.1398486

*qnorm(pnorm(-1.807 + .195\*1 + .267\*0 + 2.124\*.25))*

[1] -1.081

*plot(density(0, 1, kernel=c("gaussian")), xlab="", main="")*

*abline(v=-1.081)*