Anthony Cunningham

STAT 3200

Due May 3

**Homework 10**

#1. A. > unique(distress)

[1] 0 1 \*“distress” is a binary variable

B. P(o-ring distress) = 1/(1 – exp[-(B0 + Btempxtemp)])

Assumptions: -independent observations

- temperature is linearly related to the log of the odds of o-ring distress

C. > glm.out = glm(distress ~ temp, family=binomial(logit))

> summary(glm.out)

Coefficients:

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

(Intercept) 15.0429 7.3786 2.039 0.0415 \*

temp -0.2322 0.1082 -2.145 0.0320 \*

\* P(O-ring distress) = 1/(1 – e-(15.0429 – 0.2322x))

\* Maximum Likelihood Estimation is used to fit the model.

D. B0-hat = the log odds of O-ring distress when the temperature is 0 degrees

1/(1 + exp[-B0-hat]) = 0.999 = the probability of O-ring distress at temp = 0 degrees

Btemp-hat = the log of the odds ratio for O-ring distress when temp increases by 1 degree

eBtemp-hat = odds ratio between O-ring distress after a 1-degree increase and before a 1-degree increase

E. H0: Btemp=0 Ha: Btemp≠0 Z = -2.145 p-value = 0.032

Conclusion: We reject H0. Therefore, there is evidence that temperature is a significant predictor for the probability of O-ring distress.

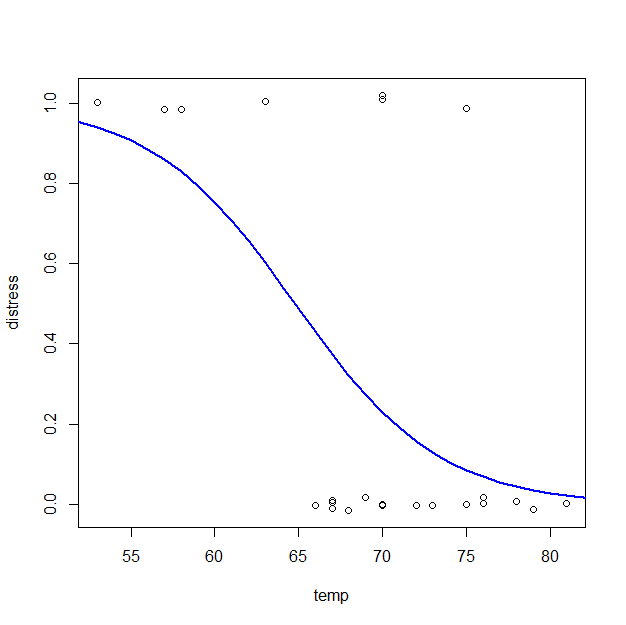
F. *P*b(*distressi* = 1 *| tempi*) = 1/(1 + exp(-(B0 + B1xi))) = 1/(1 + exp(-15.0429 + 0.2322xi))

> plot(temp, jitter(distress, factor=0.1), ylab="distress")

> xvalues=seq(50, 85)

> yvalues=predict(glm.out, list(temp=xvalues), type="response")

> lines(xvalues, yvalues, lwd=2, col="blue")



The probability of O-ring distress continually decreases on the interval of temp = [50,85].

The curve looks like the latter part of a hill, generally modeling a decrease in O-ring distress at higher temperatures.

G. *P*b(*distressi* = 1 *| tempi = 31*) = 1/(1 + exp(-15.0429 + 0.2322(31)) = 0.9996088

\* If we want to minimize the probability of O-ring distress, then I would suggest that we do not launch Challenger on a day where the predicted temperature is 31 degrees, according to the estimated model fit.

H. 1/(1 + exp(-15.0429 + 0.2322xi)) = 0.5 🡪 xi = (ln(2) + B0)/-B1 =

(ln(2) + 15.0429)/0.2322 = 67.68 degrees

#2. A. (Baseline) P-hat(Y=1|Sex=0) = 1/(1 – exp(-(B0 + B1Fati)))

P-hat(Y=1|Sex=1) = 1/(1 – exp(-((B0 + B2) + B1Fati)))

Assumptions: -independent observations

- fat consumption is linearly related to the log of the odds of high cholesterol after accounting for sex effect

B. > glm.both = glm(cholCat ~ fat + sex, family=binomial(logit))

> summary(glm.both)

Coefficients:

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

(Intercept) -4.759162 0.563834 -8.441 <2e-16 \*\*\*

fat 0.065729 0.007826 8.399 <2e-16 \*\*\*

sex 1.356750 0.552130 2.457 0.014 \*

\*P-hat(Y=1|Sex=0) = 1/(1 – exp(-4.759162 + 0.065729Fati))

\*P-hat(Y=1|Sex=1) = 1/(1 – exp(-3.402412 + 0.065729Fati))

C. exp(Bsex-hat) = The estimated ratio between the odds of a male having high cholesterol and a female’s odds of having high cholesterol at a given fat consumption level.

exp(Bfat-hat) = The estimated ratio between the odds of having high cholesterol after a 1-gram increase in fat consumption per day and the odds of having high cholesterol at a constant fat consumption per day (ie without a 1-gram increase) at a given sex.

\* Yes, the interpretations remain the same even with interaction. The only differences will be in the estimated values of the coefficients and the model itself, which will carry the extra interaction term.

D. \* chi-square test stat w/ df = df(null deviance) – df(residual deviance) = 435.54 – 279.58 = 155.96 (df = 314 – 312 = 2)

> chi.sq = glm.both$null.deviance - glm.both$deviance

> pchisq(chi.sq, 2, lower.tail=FALSE)

[1] 1.36227e-34 \* p-value

> glm.null = glm(cholCat~1, family=binomial(logit))

> anova(glm.null, glm.both, test="Chisq")

Analysis of Deviance Table

Model 1: cholCat ~ 1

Model 2: cholCat ~ fat + sex

Resid. Df Resid. Dev Df Deviance Pr(>Chi)

1 314 435.54

2 312 279.58 2 155.96 < 2.2e-16

H0: Bfat=Bsex=0 Ha: at least one Bi not equal to 0

X2\* = 155.96 p-value = 1.36 \* 10^-34

Conclusion: We reject H0. We have evidence that at least one of the predictors in the model significantly predicts the probability of whether someone has high cholesterol.

E. H0: Bfat=0 Ha: Bfat≠0 Z\* = 8.399 p-value <2 \* 10^-16

Conclusion: We reject H0. We have evidence that fat intake has a significant effect on the probability of whether one has high cholesterol, after accounting for their sex.

F. H0: Bsex=0 Ha: Bsex≠0 Z\* = 2.457 p-value = 0.014

Conclusion: We reject H0 at alpha=0.05 significance level. We have evidence that one’s sex has a significant effect on the probability of high cholesterol, after holding for the person’s fat intake.

G. \*P-hat(Y=1|Sex=0) = 1/(1 – exp(-4.759162 + 0.065729Fati))

\*P-hat(Y=1|Sex=1) = 1/(1 – exp(-3.402412 + 0.065729Fati))

> plot(fat, jitter(cholCat, factor=0.25), main="Logistic Regression ANCOVA",

+ xlab="fat", ylab="cholCat (yes=1,no=0)", type="n")

> points(fat[sex==1], jitter(cholCat[sex==1], factor=0.25), pch=16, col="red")

> points(fat[sex==0], jitter(cholCat[sex==0], factor=0.25), col="black")

> fitted.function.male = function(x)

+ {1/(1+exp(-(glm.both$coefficients[1] +

+ glm.both$coefficients[2]\*x + glm.both$coefficients[3])))}

> fitted.function.female = function(x)

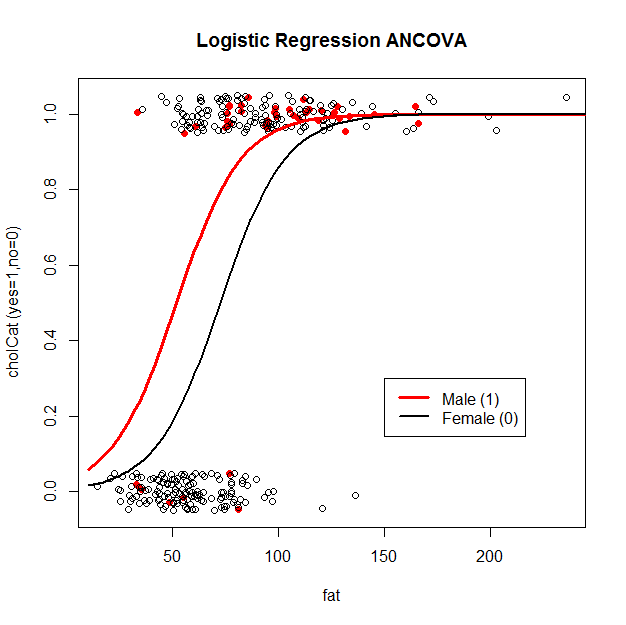
+ {1/(1+exp(-(glm.both$coefficients[1] +

+ glm.both$coefficients[2]\*x)))}

> curve(fitted.function.male, 10, 250, add=T, lwd=3, col="red")

> curve(fitted.function.female, 10, 250, add=T, lwd=2, col="black")

> legend(150, 0.3, c("Male (1)", "Female (0)"), lwd=c(3,2), col=c("red","black"))



The male group is shifted to the left from the female group.

H. P-hat(Y=1|Fat=0,Sex=0) = 1/(1 – exp(-4.759162)) = 0.0085

I. > glm.full = glm(cholCat ~ fat + sex + fat:sex, family=binomial(logit))

> summary(glm.full)

Coefficients:

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

(Intercept) -4.674853 0.587978 -7.951 1.85e-15 \*\*\*

fat 0.064513 0.008187 7.880 3.28e-15 \*\*\*

sex 0.541829 1.924729 0.282 0.778

fat:sex 0.012351 0.028011 0.441 0.659

> anova(glm.both, glm.full, test="Chisq")

Analysis of Deviance Table

Model 1: cholCat ~ fat + sex + fat:sex

Model 2: cholCat ~ fat + sex

Resid. Df Resid. Dev Df Deviance Pr(>Chi)

1 311 279.37

2 312 279.58 1 0.2117 0.6454

No, interaction is not significant in this model (p-value = 0.6454). The main effects are sufficient.