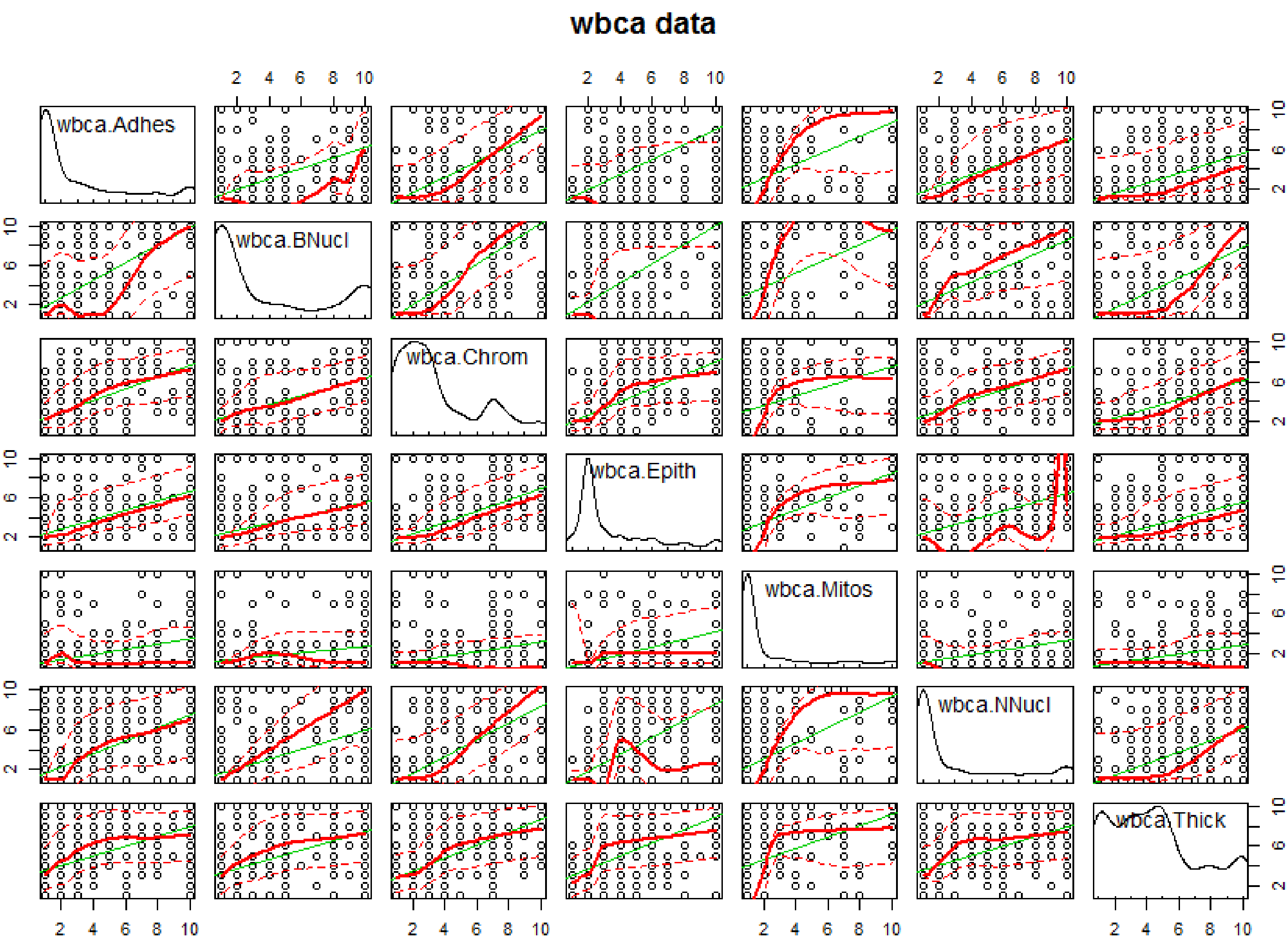
Logistic Regression Homework, STAT5120, Allen Baumgarten

Load the faraway package and then type "wbca" at the prompt…but without the quotes. Read about this dataset here: https://cran.r-project.org/web/packages/faraway/faraway.pdf. There are 681 cases of potentially cancerous tumors of which 238 are actually malignant. Determining whether a tumor is really malignant is traditionally determined by an invasive surgical procedure. The purpose of this study was to determine whether a new procedure called fine needle aspiration which draws only a small sample of tissue could be effective in determining tumor status.

(a) Fit a binomial regression with Class as the response and the other nine variables as predictors. Just do

```
model1 <- glm(Class ~ ., data = wbca, family=binomial)
summary(model1)
```

A scatterplot matrix was constructed to examine these variables. Output for an all-inclusive model shows that five of the nine variables are statistically significant.



Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.48282	-0.01179	0.04739	0.09678	3.06425

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	11.16678	1.41491	7.892	2.97e-15	***
Adhes	-0.39681	0.13384	-2.965	0.00303	**
BNucl	-0.41478	0.10230	-4.055	5.02e-05	***
Chrom	-0.56456	0.18728	-3.014	0.00257	**
Epith	-0.06440	0.16595	-0.388	0.69795	
Mitos	-0.65713	0.36764	-1.787	0.07387	.
NNucl	-0.28659	0.12620	-2.271	0.02315	*
Thick	-0.62675	0.15890	-3.944	8.01e-05	***
UShap	-0.28011	0.25235	-1.110	0.26699	
USize	0.05718	0.23271	0.246	0.80589	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 881.388 on 680 degrees of freedom
Residual deviance: 89.464 on 671 degrees of freedom
AIC: 109.46

Number of Fisher Scoring iterations: 8

Report the residual deviance and associated degrees of freedom. Can this information be used to determine if this model fits the data well (think in terms of sample sizes)? If so, what does the chi-squared test for the residual deviance indicate?

Joseph Hilbe comments on the Null and Residual deviances by saying that, "…an intercept model only, that is, a model with no predictors, is the called the *null deviance*."[1]

The residual deviance and df for this model iteration are 89.464 on 671 df, respectively. Df for the null deviance are the number of observations minus 1 (n-1). The null deviance and residual deviance can be used to calculate the Likelihood Ratio Test (G^2) which is basically a difference in the model deviances, i.e., for our model of interest vs. a null model with an intercept only. In this case, we get 881.388 – 89.464 = 791.9 which compared to a χ^2 provides strong evidence for us to reject the null hypothesis that our betas = 0 and provide no information on our predicted variable.

(b) Use AIC as the criterion in best-subsets method to determine the best subset of variables. This is easy- just do

```
reduced <- step(model1)
summary(reduced)
```

Alan Agresti remarks that, "The AIC judges a model by how close its fitted values tend to be from the true mean values, in terms of expected value. Even though a simple model farther from the true relationship than is a more

[1] Hilbe, Joseph M., *Practical Guide to Logistic Regression*, (CRC Press: 2016), 54.

complex model, it may be preferred because it tends to provide better estimates of certain characteristics, such as cell probabilities. Thus, the optimal model is the one that tends to have fit closest to the true values. Akaike defined closeness in terms of a Kullback-Leibler measure of distance…With a sample, this criterion selects the model minimizes

AIC = -2(maximized log likelihood – number of parameters in model)[2]

Dr. Hilbe's comments on the AIC, however, suggest care in using it: "The AIC is perhaps the most well-known and well used information statistic in current research. What may seem surprising to many readers is that there are a plethora of journal articles detailing studies proving how poor the AIC test is in assessing which of two models is the better fitted. Even Akaike himself later developed another criterion which he preferred to the original. However, it is his original 1973 version that is used by most researchers and that is found in most journals to assess comparative model fit. The traditional AIC statistic is found in two versions:

AIC = -2L + 2k or -2(L – k)

or

AIC = (-2L + 2k)/n or 2(L – k)/n

where L is the log-likelihood model, k is the number of parameter estimates in the model, and n is the number of observations in the model."[3]

Models were split into subsets as follows. The lowest AIC statistic was shown to be around 105 in the select step-wise model below under part (c).

Start: AIC=109.46
Class ~ Adhes + BNucl + Chrom + Epith + Mitos + NNucl + Thick + UShap + USize

	Df	Deviance	AIC
- USize	1	89.523	107.52
- Epith	1	89.613	107.61
- UShap	1	90.627	108.63
<none>		89.464	109.46
- Mitos	1	93.551	111.55
- NNucl	1	95.204	113.20
- Adhes	1	98.844	116.84
- Chrom	1	99.841	117.84
- BNucl	1	109.000	127.00
- Thick	1	110.239	128.24

Step: AIC=107.52
Class ~ Adhes + BNucl + Chrom + Epith + Mitos + NNucl + Thick + UShap

	Df	Deviance	AIC
- Epith	1	89.662	105.66
- UShap	1	91.355	107.36

[2] Agresti, Alan, Categorical Data Analysis, 3rd ed., (New Jersey: John Wiley & Sons, 2013), 212.

[3] Hilbe, 58.

<none>		89.523	107.52
- Mitos	1	93.552	109.55
- NNucl	1	95.231	111.23
- Adhes	1	99.042	115.04
- Chrom	1	100.153	116.15
- BNucl	1	109.064	125.06
- Thick	1	110.465	126.47

Step: AIC=105.66
Class ~ Adhes + BNucl + Chrom + Mitos + NNucl + Thick + UShap

	Df	Deviance	AIC
<none>		89.662	105.66
- UShap	1	91.884	105.88
- Mitos	1	93.714	107.71
- NNucl	1	95.853	109.85
- Adhes	1	100.126	114.13
- Chrom	1	100.844	114.84
- BNucl	1	109.762	123.76
- Thick	1	110.632	124.63

(c) Use the reduced model (the one from part (b)) to predict the outcome (malignant or not) for a patient with Adhes = 1, BNucl = 1, Chrom = 3, Epith = 2, Mitos = 1, NNucl = 1, Thick = 4, UShap = 1, USize = 1. Also, give a 95% confidence interval for your prediction. Our reduced model is seen to be:

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.44161	-0.01119	0.04962	0.09741	3.08205

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	
(Intercept)	11.0333	1.3632	8.094	5.79e-16	***
Adhes	-0.3984	0.1294	-3.080	0.00207	**
BNucl	-0.4192	0.1020	-4.111	3.93e-05	***
Chrom	-0.5679	0.1840	-3.085	0.00203	**
Mitos	-0.6456	0.3634	-1.777	0.07561	.
NNucl	-0.2915	0.1236	-2.358	0.01837	*
Thick	-0.6216	0.1579	-3.937	8.27e-05	***
UShap	-0.2541	0.1785	-1.423	0.15461	

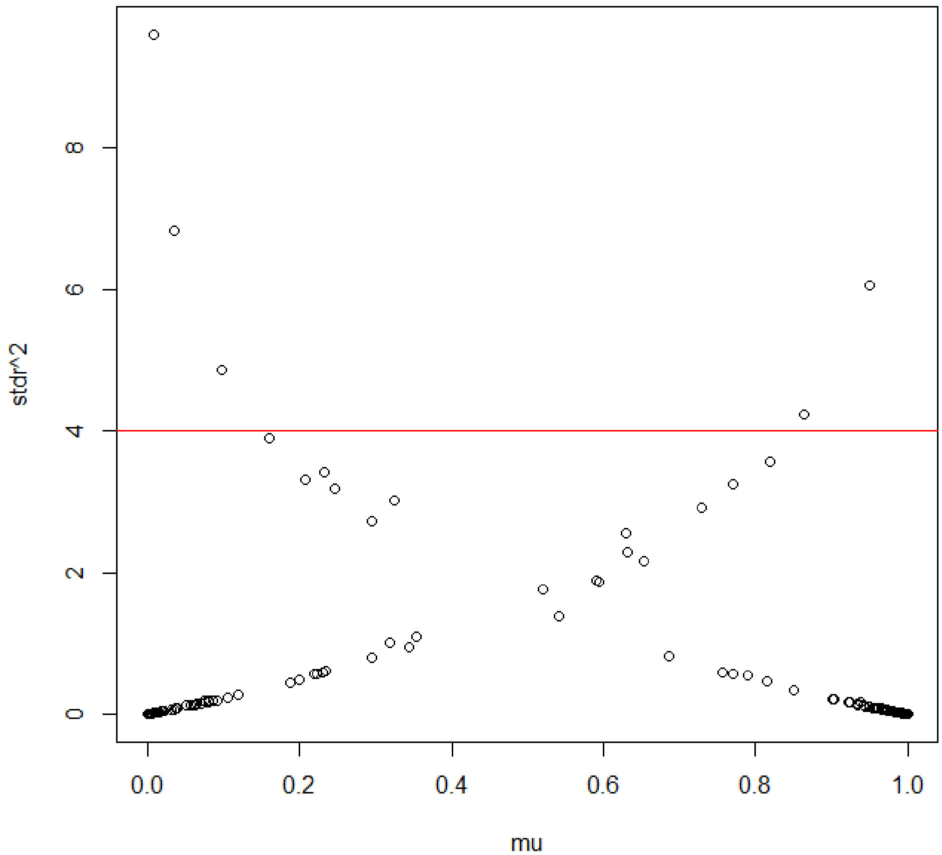
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 881.388 on 680 degrees of freedom
Residual deviance: 89.662 on 673 degrees of freedom
AIC: 105.66
Number of Fisher Scoring iterations: 8
BIC: 141.8503

Hilbe advises that, "The Schwartz Bayesian information criterion (BIC) is the most used BIC test found in the literature…My recommendation is to test models with both [the AIC and BIC]. If the values substantially differ, it is likely the model is mis-specified."[4] Here we see that the AIC and BIC so seem to differ somewhat. Residuals analysis would be in order but we will forgo that for the time being.

There are some residuals that seem to greatly skewed as seen in the standardized residuals plot:



We will use this reduced model to predict the outcome of a patient with the values above. What we will have is the log-odds of having (or not having) a cancerous tumor, given these values. Our model is:

Y = 11.033 + -0.398(Adhes) + -0.419(BNucl) + -0.568(Chrom) + -0.646(Mitos) + -0.292(NNucl) + -0.622(Thick) + -0.254(Ushap)

Our proposed values above suggest the likelihood of cancerous tumor expression to be:

Y = 11.033 + -0.398(1) + -0.419(1) + -0.568(3) + -0.646(1) + -0.292(1) + -0.622(4) + -0.254(1)
Y = 11.033 – 0.398 – 0.419 – 1.704 – 0.646 – 0.292 – 2.488 – 0.254
Y = rt

[4] Hilbe, 60.

(d) Suppose a tumor is classified as benign if p > 0.5 and classified as malignant if p < 0.5 (remember "1" means benign and "0" indicates malignant for the Class variable). Compute the number of errors of both types that will be made if this method is applied to the current data with the reduced model. Also, compute the percentage of classifications that result in each kind of error (that is, compute the error rates- false positive and false negative).

(e) Suppose we move the cut-off point to 0.9 so that p < 0.9 indicates malignant and p > 0.9 indicates benign. What are the new error counts and rates?

(f) It can be misleading to use the same data to fit a model and test its predictive ability. So split the original dataset into two parts: assign every third record to the test set and all the others to the training set. Use the training set to determine a good model (repeat parts (a) and (b)). Then use the test set to assess predictive performance (repeat parts (d) and (e)).

(g) Discuss how you could search for the "best" cut-off point to use for classifying tumors. Then write an R program to carry it out. What is the "best" cut-off value? What do you mean by "best"?

APPENDIX: R SCRIPTS

Part (a):

```
>scatterplotMatrix(~wbca$Adhes+wbca$BNucl+wbca$Chrom+wbca$Epith+wbca$Mitos+wbca$NNucl+wbca$Thic
k, data=wbca, main="wbca data")
> model1 <- glm(Class ~ ., data = wbca, family=binomial)
> summary(model1)

Call:
glm(formula = Class ~ ., family = binomial, data = wbca)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.48282  -0.01179   0.04739   0.09678   3.06425

Coefficients:
            Estimate  Std. Error  z value  Pr(>|z|)
(Intercept) 11.16678   1.41491     7.892   2.97e-15 ***
Adhes       -0.39681   0.13384    -2.965   0.00303 **
BNucl       -0.41478   0.10230    -4.055   5.02e-05 ***
Chrom       -0.56456   0.18728    -3.014   0.00257 **
Epith       -0.06440   0.16595    -0.388   0.69795
Mitos       -0.65713   0.36764    -1.787   0.07387 .
NNucl       -0.28659   0.12620    -2.271   0.02315 *
Thick       -0.62675   0.15890    -3.944   8.01e-05 ***
UShap       -0.28011   0.25235    -1.110   0.26699
USize        0.05718   0.23271     0.246   0.80589
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance:     881.388  on 680  degrees of freedom
```

Residual deviance: 89.464 on 671 degrees of freedom
AIC: 109.46

Number of Fisher Scoring iterations: 8

Part (b):

```
> reduced <- step(model1)
Start:  AIC=109.46
Class ~ Adhes + BNucl + Chrom + Epith + Mitos + NNucl + Thick + UShap + USize

         Df  Deviance   AIC
- USize   1    89.523  107.52
- Epith   1    89.613  107.61
- UShap   1    90.627  108.63
<none>         89.464  109.46
- Mitos   1    93.551  111.55
- NNucl   1    95.204  113.20
- Adhes   1    98.844  116.84
- Chrom   1    99.841  117.84
- BNucl   1   109.000  127.00
- Thick   1   110.239  128.24

Step:  AIC=107.52
Class ~ Adhes + BNucl + Chrom + Epith + Mitos + NNucl + Thick + UShap

         Df  Deviance   AIC
- Epith   1    89.662  105.66
- UShap   1    91.355  107.36
<none>         89.523  107.52
- Mitos   1    93.552  109.55
- NNucl   1    95.231  111.23
- Adhes   1    99.042  115.04
- Chrom   1   100.153  116.15
- BNucl   1   109.064  125.06
- Thick   1   110.465  126.47

Step:  AIC=105.66
Class ~ Adhes + BNucl + Chrom + Mitos + NNucl + Thick + UShap

         Df  Deviance   AIC
<none>         89.662  105.66
- UShap   1    91.884  105.88
- Mitos   1    93.714  107.71
- NNucl   1    95.853  109.85
- Adhes   1   100.126  114.13
- Chrom   1   100.844  114.84
- BNucl   1   109.762  123.76
- Thick   1   110.632  124.63
```

Part (c):

```
> summary(reduced)
Call:
glm(formula = Class ~ Adhes + BNucl + Chrom + Mitos + NNucl + Thick + UShap, family = binomial, data = wbca)

Deviance Residuals:
    Min        1Q     Median      3Q       Max
-2.44161  -0.01119   0.04962   0.09741   3.08205

Coefficients:
            Estimate   Std. Error  z value   Pr(>|z|)
(Intercept) 11.0333    1.3632      8.094     5.79e-16 ***
Adhes       -0.3984    0.1294      -3.080    0.00207 **
BNucl       -0.4192    0.1020      -4.111    3.93e-05 ***
Chrom       -0.5679    0.1840      -3.085    0.00203 **
Mitos       -0.6456    0.3634      -1.777    0.07561 .
NNucl       -0.2915    0.1236      -2.358    0.01837 *
Thick       -0.6216    0.1579      -3.937    8.27e-05 ***
UShap       -0.2541    0.1785      -1.423    0.15461
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance:      881.388  on 680  degrees of freedom
Residual deviance:   89.662  on 673  degrees of freedom
AIC:  105.66

Number of Fisher Scoring iterations:  8


> wbca_regmod <- glm(wbca$Class ~ wbca$Adhes + wbca$BNucl + wbca$Chrom + wbca$Mitos + wbca$NNucl +
wbca$Thick + wbca$UShap, data=wbca, family = binomial)
> summary(wbca_regmod)

current_model <- wbca_regmod  #Run glm(), assign to this object; also update lines 96, 103 and 114
current_data <- wbca  #Be sure this points to the current data source

G2 <- current_model$null.deviance - current_model$deviance
EL50 <- -(-12.3508)/0.4972  # Input the EL50 = -a/B (negated intercept / Beta)

# Print summary of results and coefficients
print(summary(current_model))
writeLines("Deviance Statistic:")
print(G2)
writeLines("")
writeLines("Coefficients Exponentiated:")
print(exp(coef(current_model)))
writeLines("")
```

```
writeLines("")
writeLines("95% CI's (Wald):")
print(confint.default(current_model))
writeLines("")
writeLines("")
writeLines("95% CI's (Wald) Exponentiated:")
print(exp(confint.default(current_model)))
writeLines("")
writeLines("")
writeLines("95% CI's (Profile/Likelihood Ratio):")
print(confint(current_model))
writeLines("")
writeLines("")
writeLines("95% CI's (Profile/Likelihood Ratio) Exponentiated:")
print(exp(confint(current_model)))
writeLines("")
writeLines("")

# Runs a Pearson GOF test.  If p>.05 then model is probably well-fitted though check for overdispersion, AIC, etc.
pr <- sum(residuals(current_model, type = "pearson")^2)
df <- current_model$df.residual
p_value <- pchisq(pr, current_model$df.residual, lower=F)
print(matrix(c("Pearson Chi GOF", "Chi2", "df", "p-value", "Parameters", round(pr,4),df,round(p_value,4)),
ncol=2))
writeLines("")
writeLines("")

# Print standardized residuals (full model)
mu <- current_model$fitted.value
dr <- resid(current_model, type = "deviance")
hat <- hatvalues(current_model)
stdr <- dr/sqrt(1-hat)
windows()
plot(mu, stdr^2)
abline(h = 4, col = "red")

# Predicted values (odds) of y given a continuous predictor
predict <- predict(current_model)
fit <- current_model$fitted.values

# Calculate standard errors of the linear predictor
lpred <- predict(current_model, newdata = current_data, type = "link", se.fit = TRUE)
up <- lpred$fit + (qnorm(.975) * lpred$se.fit)
low <- lpred$fit - (qnorm(.975) * lpred$se.fit)
eta <- lpred$fit
upci <- current_model$family$linkinv(up)
mu <- current_model$family$linkinv(eta)
loci <- current_model$family$linkinv(low)
writeLines("Lower CI:")
print(summary(loci))
```

```
writeLines("")
writeLines("Mean CI:")
print(summary(mu))
writeLines("")
writeLines("Upper CI:")
print(summary(upci))
writeLines("")

# Bayseian Information Criterion
library(COUNT)
writeLines("AIC and BIC Statistics:")
print(modelfit(current_model))

# Graph of probabilities of continuous predictor.  Remove #s and > and write in continuous variable name to the right of $s:
# > layout(1)
# > plot(current_data$BMI, mu, col = 1)
# > lines(current_data$BMI, loci, col = 2, type = 'p')
# > lines(current_data$BMI, upci, col = 3, type = 'p')

# To check for overdispersion in the model:  scaled and robust/sandwhich se's:
coef <- current_model$coefficients
se <- sqrt(diag(vcov(current_model)))
coefse <- data.frame(coef, se)
writeLines("")
pr <- resid(current_model, type = "pearson")
pchi2 <- sum(residuals(current_model, type = "pearson")^2)
disp <- pchi2/current_model$df.residual
scse <- se*sqrt(disp)
library(sandwich)
rmodel <- glm(current_data$Class ~ current_data$Adhes + current_data$BNucl + current_data$Chrom + current_data$Mitos + current_data$NNucl + current_data$Thick + current_data$UShap, family = binomial, data = current_data) #Update variables!!
# WITH FACTOR:  rmodel <- glm(current_data$y ~ + current_data$weight+ current_data$los + factor(current_data$type), family = binomial, data = current_data) #Update variables!!
rse <- sqrt(diag(vcovHC(rmodel, type = "HC0")))
newcoefse <- data.frame( coef, se, scse, rse)
print(newcoefse)

# Quasibinomial model:  check that coefficients are identical or nearly so.  If so, then good fit:
quasibinomialmod <- glm(current_data$Class ~ current_data$Adhes + current_data$BNucl + current_data$Chrom + current_data$Mitos + current_data$NNucl + current_data$Thick + current_data$UShap, family = quasibinomial, data = current_data) #Update variables!!
# WITH FACTOR:  quasibinomialmod <- glm(current_data$died ~ + current_data$white + current_data$hmo + current_data$los +  factor(current_data$type), family = quasibinomial, data = current_data) #Update variables!!
writeLines("")
writeLines("")
writeLines("Quasibinomial model:")
print(summary(quasibinomialmod))
```

```
> source("Log_REG_working.r")

Call:
glm(formula = wbca$Class ~ wbca$Adhes + wbca$BNucl + wbca$Chrom +
    wbca$Mitos + wbca$NNucl + wbca$Thick + wbca$UShap, family = binomial,
    data = wbca)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.44161  -0.01119   0.04962   0.09741   3.08205

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  11.0333     1.3632   8.094 5.79e-16 ***
wbca$Adhes   -0.3984     0.1294  -3.080  0.00207 **
wbca$BNucl   -0.4192     0.1020  -4.111 3.93e-05 ***
wbca$Chrom   -0.5679     0.1840  -3.085  0.00203 **
wbca$Mitos   -0.6456     0.3634  -1.777  0.07561 .
wbca$NNucl   -0.2915     0.1236  -2.358  0.01837 *
wbca$Thick   -0.6216     0.1579  -3.937 8.27e-05 ***
wbca$UShap   -0.2541     0.1785  -1.423  0.15461
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 881.388  on 680  degrees of freedom
Residual deviance:  89.662  on 673  degrees of freedom
AIC: 105.66

Number of Fisher Scoring iterations: 8

Deviance Statistic:
[1] 791.7264

Coefficients Exponentiated:
 (Intercept)    wbca$Adhes   wbca$BNucl   wbca$Chrom   wbca$Mitos    wbca$NNucl   wbca$Thick    wbca$UShap
6.190351e+04 6.713727e-01 6.575746e-01 5.667417e-01 5.243413e-01 7.471176e-01 5.370696e-01 7.756406e-
01


95% CI's (Wald):
                   2.5 %         97.5 %
(Intercept)      8.3615066   13.70515770
wbca$Adhes   -0.6519565   -0.14490519
wbca$BNucl   -0.6190357   -0.21935844
wbca$Chrom   -0.9285735   -0.20712955
wbca$Mitos   -1.3578036    0.06657864
wbca$NNucl  -0.5338499   -0.04921541
wbca$Thick   -0.9311288   -0.31212640
```

```
wbca$UShap  -0.6038995    0.09576752


95% CI's (Wald) Exponentiated:
                2.5 %       97.5 %
(Intercept) 4279.1369847 8.955181e+05
wbca$Adhes     0.5210254 8.651043e-01
wbca$BNucl     0.5384634 8.030338e-01
wbca$Chrom     0.3951169 8.129143e-01
wbca$Mitos     0.2572251 1.068845e+00
wbca$NNucl     0.5863432 9.519760e-01
wbca$Thick     0.3941086 7.318890e-01
wbca$UShap     0.5466757 1.100503e+00


95% CI's (Profile/Likelihood Ratio):
Waiting for profiling to be done...
              2.5 %      97.5 %
(Intercept)  8.7277704 14.15772490
wbca$Adhes  -0.6722270 -0.15465473
wbca$BNucl  -0.6339696 -0.22932801
wbca$Chrom  -0.9534347 -0.22625284
wbca$Mitos  -1.2507890 -0.01303244
wbca$NNucl  -0.5466091 -0.05972893
wbca$Thick  -0.9635139 -0.33585675
wbca$UShap  -0.6291944  0.07595514


95% CI's (Profile/Likelihood Ratio) Exponentiated:
Waiting for profiling to be done...
                2.5 %       97.5 %
(Intercept) 6171.9518030 1.408062e+06
wbca$Adhes     0.5105703 8.567109e-01
wbca$BNucl     0.5304818 7.950677e-01
wbca$Chrom     0.3854150 7.975164e-01
wbca$Mitos     0.2862788 9.870521e-01
wbca$NNucl     0.5789095 9.420199e-01
wbca$Thick     0.3815498 7.147255e-01
wbca$UShap     0.5330210 1.078914e+00


     [,1]             [,2]
[1,] "Pearson Chi GOF" "Parameters"
[2,] "Chi2"            "227.9781"
[3,] "df"              "673"
[4,] "p-value"         "1"


Lower CI:
   Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
```

0.000000 0.000915 0.966333 0.621669 0.990807 0.997395

Mean CI:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000000	0.007241	0.989560	0.650514	0.997888	0.999605

Upper CI:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000022	0.0732503	0.9970059	0.6850585	0.9995259	0.9999401

AIC and BIC Statistics:
$AIC
[1] 105.6618

$AICn
[1] 0.1551568

$BIC
[1] 141.8503

$BICqh
[1] 0.1805181

	coef	se	scse	rse
(Intercept)	11.0333322	1.3632013	0.79341251	1.1988868
wbca$Adhes	-0.3984309	0.1293522	0.07528577	0.1510720
wbca$BNucl	-0.4191971	0.1019604	0.05934312	0.1216056
wbca$Chrom	-0.5678516	0.1840452	0.10711828	0.1603045
wbca$Mitos	-0.6456125	0.3633695	0.21148886	0.3238224
wbca$NNucl	-0.2915327	0.1236335	0.07195737	0.1128367
wbca$Thick	-0.6216276	0.1579117	0.09190800	0.1314029
wbca$UShap	-0.2540660	0.1784898	0.10388488	0.2047329

Quasibinomial model:

Call:
glm(formula = current_data$Class ~ current_data$Adhes + current_data$BNucl +
current_data$Chrom + current_data$Mitos + current_data$NNucl +
current_data$Thick + current_data$UShap, family = quasibinomial,
data = current_data)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.44161	-0.01119	0.04962	0.09741	3.08205

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	11.03333	0.79351	13.904	< 2e-16	***

```
current_data$Adhes -0.39843    0.07530  -5.292 1.64e-07 ***
current_data$BNucl -0.41920    0.05935  -7.063 4.07e-12 ***
current_data$Chrom -0.56785    0.10713  -5.300 1.57e-07 ***
current_data$Mitos -0.64561    0.21152  -3.052  0.00236 **
current_data$NNucl -0.29153    0.07197  -4.051 5.70e-05 ***
current_data$Thick -0.62163    0.09192  -6.763 2.94e-11 ***
current_data$UShap -0.25407    0.10390  -2.445  0.01473 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasibinomial family taken to be 0.338836)

    Null deviance: 881.388  on 680  degrees of freedom
Residual deviance:  89.662  on 673  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 8
```