-Assumptions-

Ethan Shen and Steven Herrera

11/27/2018

Linear Regression Assumptions

```
final.base.model <- model.selected.interactions
kable(tidy(final.base.model), format = "markdown", digits = 3)
```

term	estimate	std.error	statistic	p.value
(Intercept)	10.392	3.518	2.954	0.003
minutes	-0.008	0.003	-2.844	0.004
ht	-0.048	0.019	-2.547	0.011
rankpoints	0.000	0.000	5.441	0.000
ace	0.110	0.040	2.764	0.006
df	-0.243	0.070	-3.474	0.001
bpSaved	-0.075	0.029	-2.548	0.011
surfaceGrass	5.285	8.299	0.637	0.524
surfaceHard	-8.048	4.294	-1.874	0.061
ht:surfaceGrass	-0.043	0.045	-0.940	0.347
ht:surfaceHard	0.040	0.023	1.710	0.087
ace:surfaceGrass	0.164	0.080	2.048	0.041
ace:surfaceHard	-0.021	0.045	-0.469	0.639
df:surfaceGrass	0.425	0.132	3.205	0.001
df:surfaceHard	0.107	0.082	1.297	0.195

## Binned Plots with Residuals vs Predicted

```
ten <- ten %>% mutate(Residuals = residuals.glm(final.base.model,type="response"),
                      Predicted = predict.glm(final.base.model,type="response"))

binnedplot(ten$Predicted, ten$Residuals,xlab="Predicted Probabilities",
           ylab="Residuals",main="Binned Residuals vs. Predicted Probabilities")
```
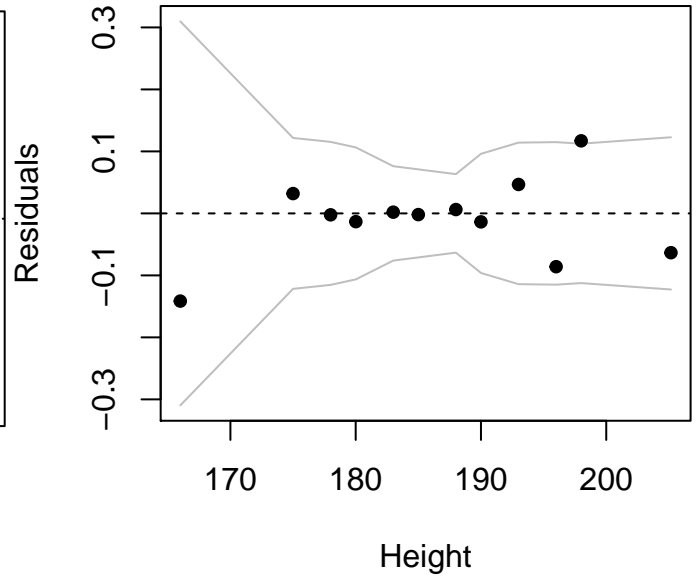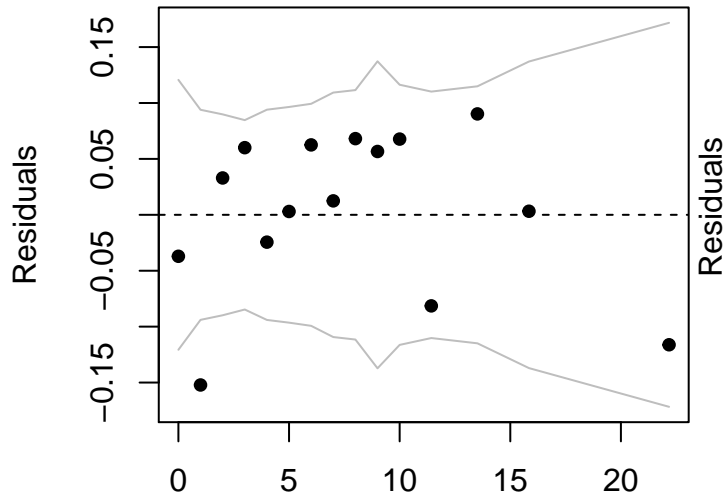
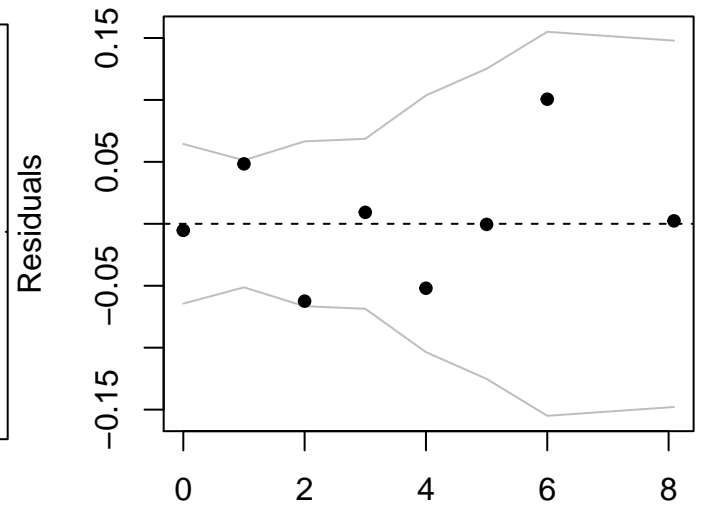# Binned Residuals vs. Predicted Probabilities
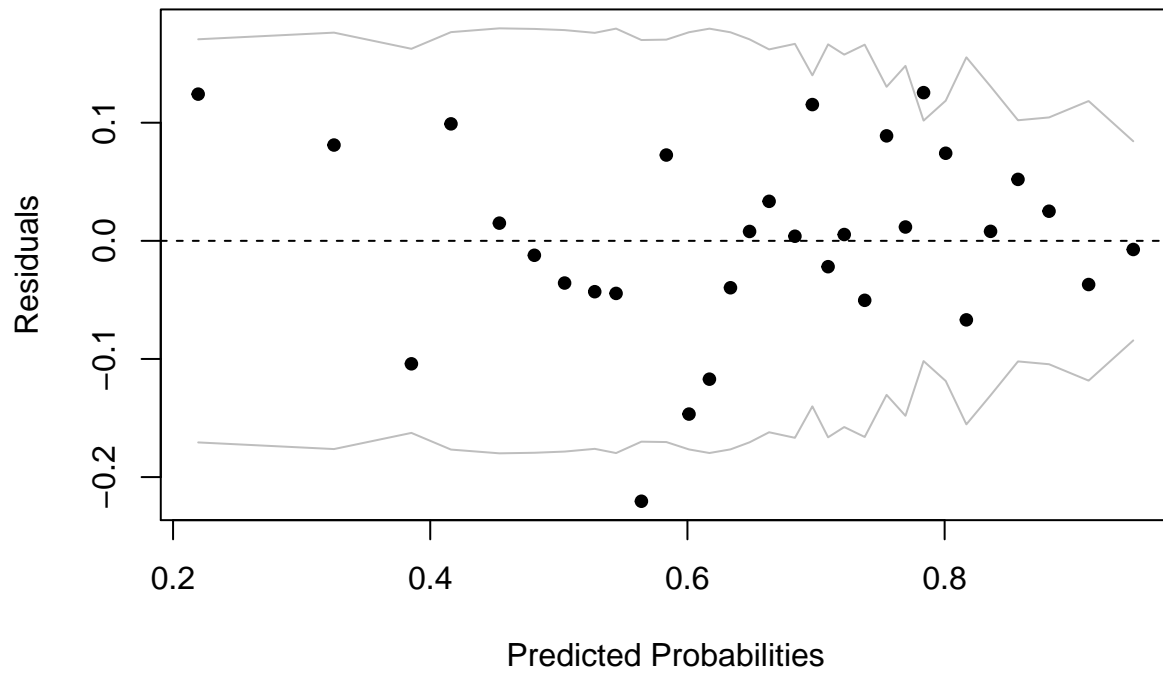
Binned Residuals vs. Minutes
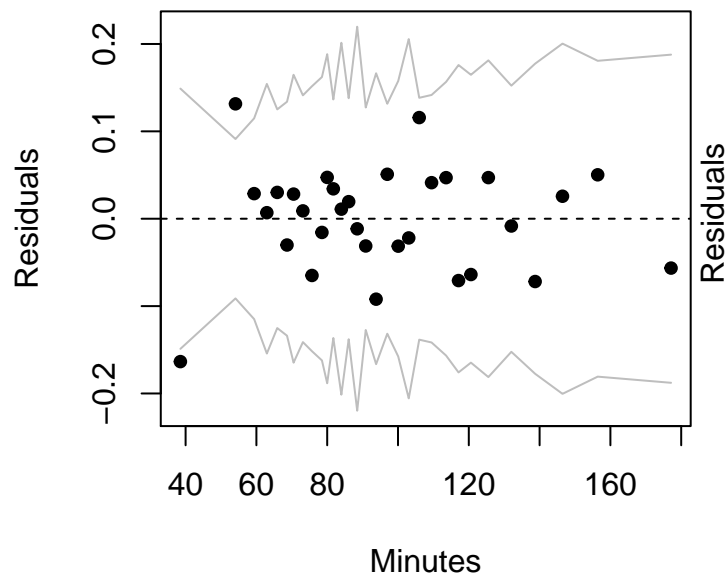
Binned Residuals vs. Height

Binned Residuals vs. Aces
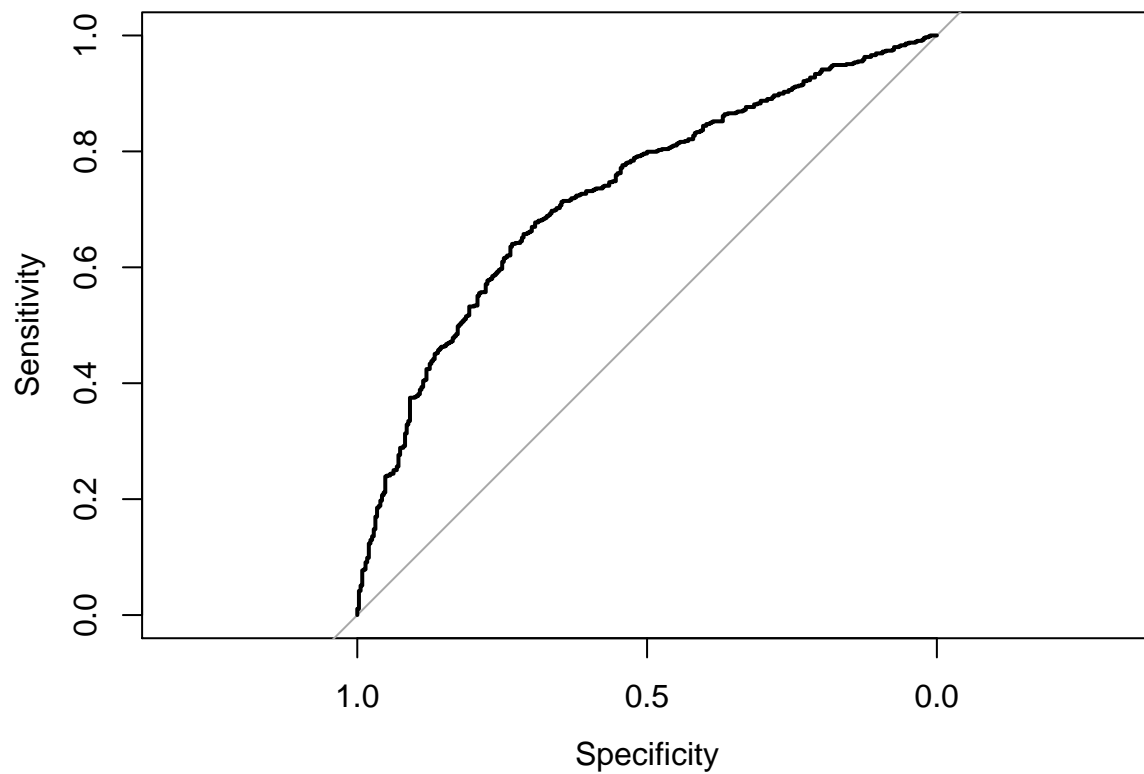

Binned Residuals vs. Double Faults

Binned Residuals vs. Saved break points

Looking at the binned residual plots, we see that all of the plots except for the binned residuals vs. saved break point have random scatter. The binned residuals vs. saved break shows a pattern. This is a violation of the assumptions.

```
ROC.ten <- roc(ten$status,ten$Predicted,plot=T)
```

```
ROC.ten$auc
```

```
## Area under the curve: 0.7268
```

```
threshold = 0.30
table(ten$status, ten$Predicted > threshold)
```

```
##
##     FALSE TRUE
##   0   26  326
##   1   13  635
```

```
(326 + 13)/(14+13+326+635)
```

```
## [1] 0.3431174
```