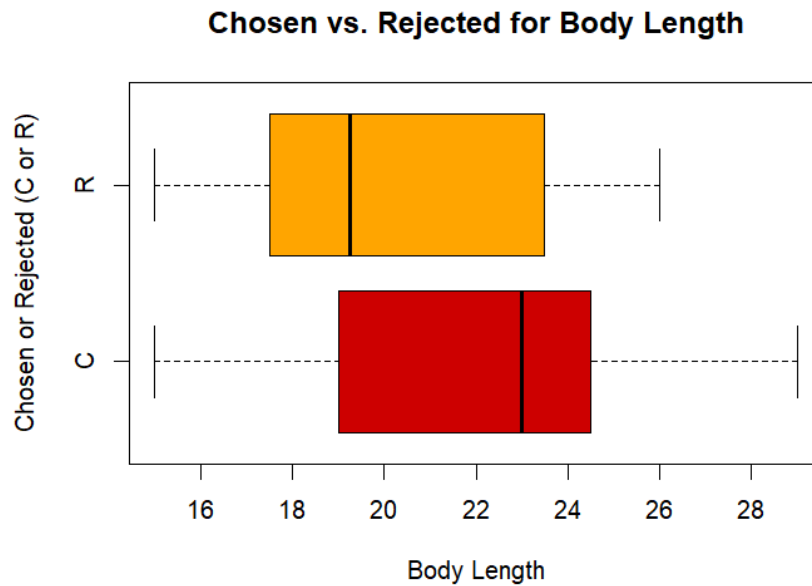


## #19 Final Part

*Gabriella Weis*

### 1. Comparing distributions with boxplots:



Chosen: "C"

Min: 15

Q1: 19

Median: 23

Mean: 22.41

Q3: 24.25

Max: 29

Rejected: "R"

Min: 15

Q1: 17.75

Median: 19.25

Mean: 20.07

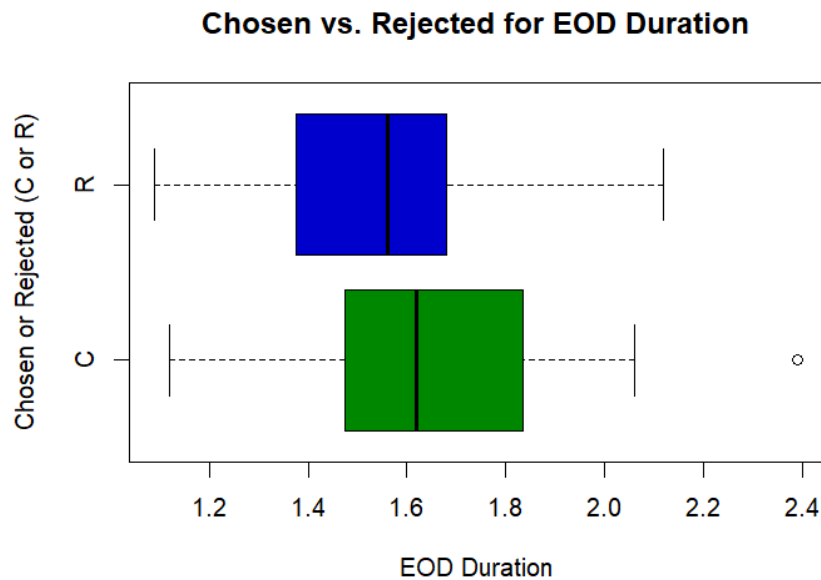
Q3: 23.25

Max: 26

Center: The centers of each boxplot are far from each other, considering the variation in the dataset; C has a median value of 23, and R has a median value of 19.25.

Shape: While the R boxplot appears to be skewed right, the C boxplot appears to be skewed left.

Spread: While the C boxplot has a range from 15-29, or 14, the R boxplot has a range from 15-26, or 11. Therefore, C has more variation/spread.



Chosen: "C"

Min: 1.120

Q1: 1.482

Median: 1.620

Mean: 1.660

Q3: 1.827

Max: 2.390

Rejected: "R"

Min: 1.090

Q1: 1.377

Median: 1.560

Mean: 1.546

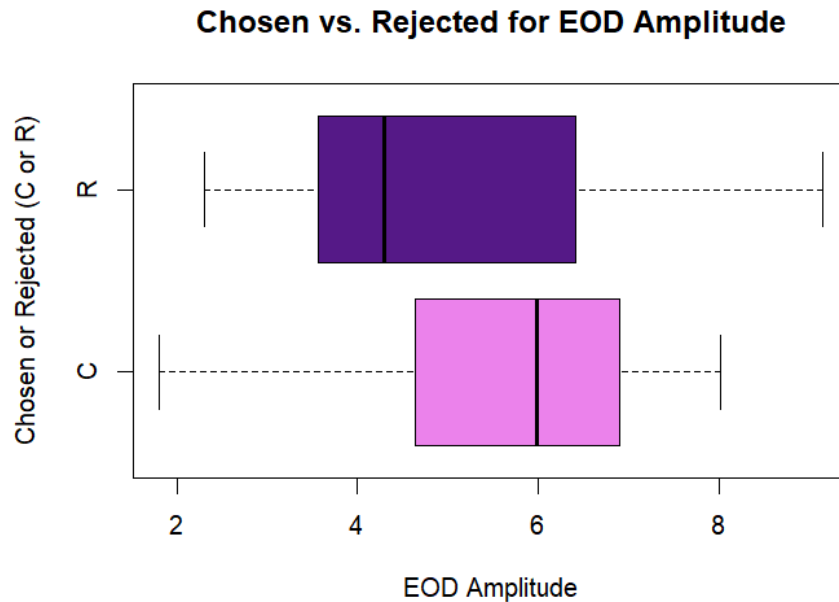
Q3: 1.680

Max: 2.120

Center: The centers of the data are fairly similar, considering the variation in the dataset; C has a median of 1.620, and R has a median of 1.560.

Shape: While the R boxplot appears to be approximately normally distributed, the C boxplot appears to be slightly skewed to the right with an outlier at 2.390.

Spread: While the C boxplot has a range from 1.120-2.390, or 1.27 the R boxplot has a range from 1.090-2.120, or 1.03. Therefore, C has more variation/spread.



Chosen: "C"  
 Min: 1.810  
 Q1: 4.657  
 Median: 5.985  
 Mean: 5.633  
 Q3: 6.900  
 Max: 8.020

Rejected: "R"  
 Min: 2.310  
 Q1: 3.585  
 Median: 4.300  
 Mean: 5.024  
 Q3: 6.242  
 Max: 9.140

Center: The centers of the data are fairly far from each other, considering the variation in the dataset; C has a median of 5.985, and R has a median of 4.300.

Shape: While the R boxplot appears to be skewed right, the C boxplot appears to be skewed left.

Spread: While the C boxplot has a range from 1.810-8.020, or 6.21 the R boxplot has a range from 2.310-9.140, or 6.83. Therefore, R has more variation/spread.

## 2. Finding statistical significance with t-tests:

### Body Length:

Null hypothesis: true difference in means between group C and group R for body length is equal to 0.

Alternative hypothesis: true difference in means between group C and group R for body length is not equal to 0.

$\alpha = 0.05$

$t = 2.3939$

$df = 53.751$

$p\text{-value} = 0.02019$

Summary: Since the  $p\text{-value} (0.02019) < \alpha (0.05)$ , we reject the null hypothesis. Therefore, this suggests convincing evidence that the true difference in means between group C and group R for body length are not equal to zero.

### Duration:

Null hypothesis: true difference in means between group C and group R for duration is equal to 0.

Alternative hypothesis: true difference in means between group C and group R for duration is not equal to 0.

$\alpha = 0.05$

$t = 1.6576$

$df = 52.905$

$p\text{-value} = 0.1033$

Summary: Since the  $p\text{-value} (0.1033) > \alpha (0.05)$ , we fail to reject the null hypothesis. Therefore, this does not suggest convincing evidence that the true difference in means between group C and group R for duration are unequal to zero.

Amplitude:

Null hypothesis: true difference in means between group C and group R for amplitude is equal to 0.

Alternative hypothesis: true difference in means between group C and group R for amplitude is not equal to 0.

$\alpha = 0.05$

$t = 1.2834$

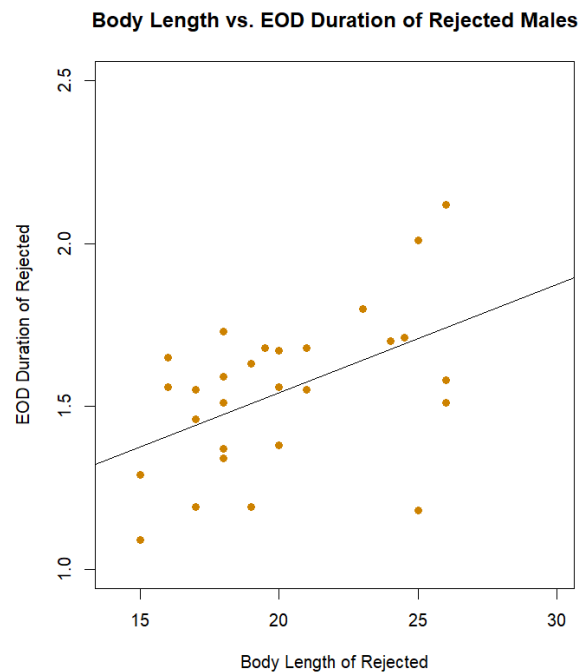
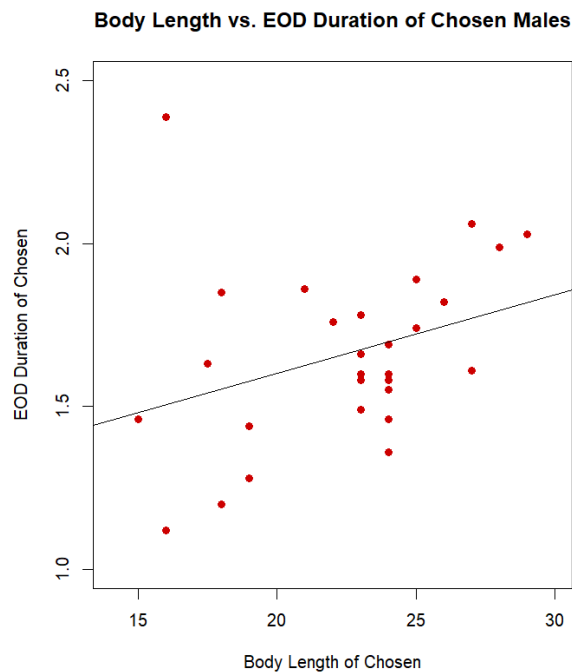
$df = 53.023$

$p\text{-value} = 0.2049$

Summary: Since the  $p\text{-value} (0.2049) > \alpha (0.05)$ , we fail to reject the null hypothesis.

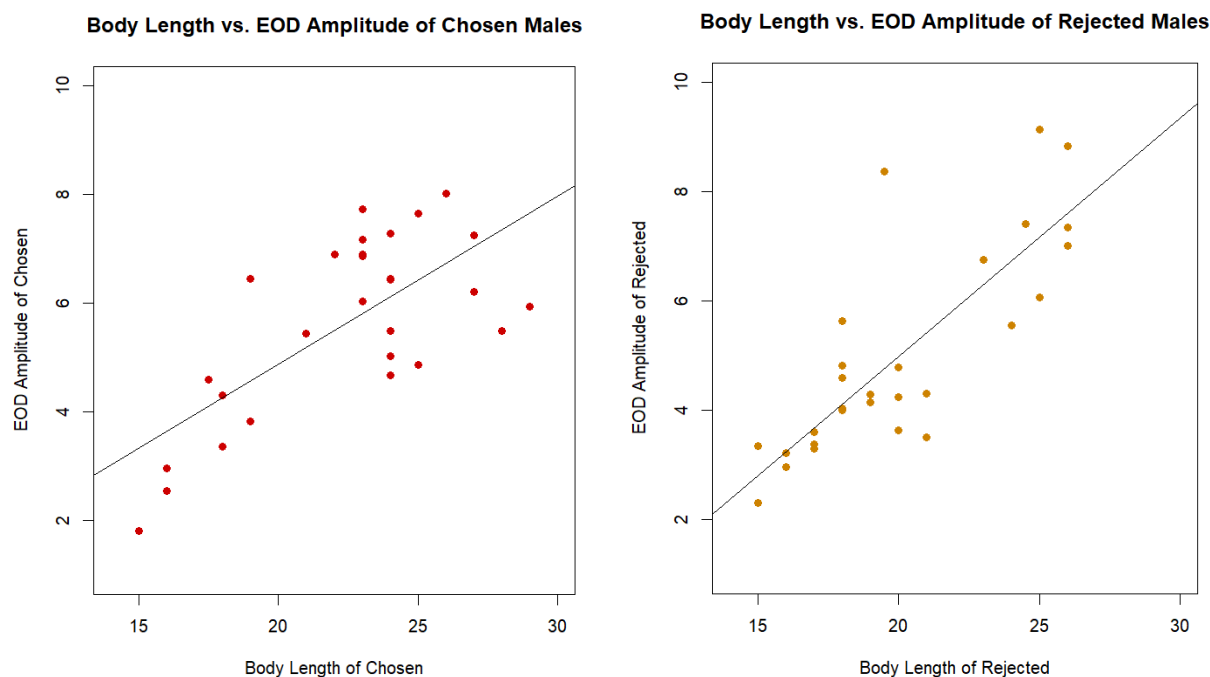
Therefore, this does not suggest convincing evidence that the true difference in means between group C and group R for amplitude are unequal to zero.

### 3. Characteristic combinations and analysis:



The slopes for the two graphs appear very similar on the scatter plots, and are 0.02421 for chosen and 0.03319 (steeper) for rejected. The graph of chosen males has a higher y-intercept (1.11736) than that of the rejected (0.87955). As for their correlations, the graph of chosen has an  $R^2$  of 0.11, and the graph of rejected has an  $R^2$  of 0.2409. See computer output below, including df, f-statistic, p-value, and five number summary.

Body Length vs. EOD Duration (Chosen)	Body Length vs. EOD Duration (Rejected)																																																																																																																		
<p>Call: lm(formula = DurationC ~ BodyLenC)</p> <p>Residuals:</p> <table><tr><td>Min</td><td>1Q</td><td>Median</td><td>3Q</td></tr><tr><td>-0.38477</td><td>-0.15164</td><td>-0.01741</td><td>0.12428</td></tr><tr><td>Max</td><td></td><td></td><td></td></tr><tr><td>0.88523</td><td></td><td></td><td></td></tr></table> <p>Coefficients:</p> <table><tr><td></td><td>Estimate</td><td>Std. Error</td><td>t value</td></tr><tr><td>Pr(&gt; t )</td><td></td><td></td><td></td></tr><tr><td>(Intercept)</td><td>1.11736</td><td>0.30688</td><td>3.641</td></tr><tr><td>0.00118 **</td><td></td><td></td><td></td></tr><tr><td>BodyLenC</td><td>0.02421</td><td>0.01351</td><td>1.792</td></tr><tr><td>0.08472 .</td><td></td><td></td><td></td></tr><tr><td>---</td><td></td><td></td><td></td></tr><tr><td>Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</td><td></td><td></td><td></td></tr><tr><td>Residual standard error: 0.2653 on 26 degrees of freedom</td><td></td><td></td><td></td></tr><tr><td>Multiple R-squared: 0.11, Adjusted R-squared: 0.07574</td><td></td><td></td><td></td></tr><tr><td>F-statistic: 3.213 on 1 and 26 DF, p-value: 0.08472</td><td></td><td></td><td></td></tr></table>	Min	1Q	Median	3Q	-0.38477	-0.15164	-0.01741	0.12428	Max				0.88523					Estimate	Std. Error	t value	Pr(> t )				(Intercept)	1.11736	0.30688	3.641	0.00118 **				BodyLenC	0.02421	0.01351	1.792	0.08472 .				---				Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				Residual standard error: 0.2653 on 26 degrees of freedom				Multiple R-squared: 0.11, Adjusted R-squared: 0.07574				F-statistic: 3.213 on 1 and 26 DF, p-value: 0.08472				<p>Call: lm(formula = DurationR ~ BodyLenR)</p> <p>Residuals:</p> <table><tr><td>Min</td><td>1Q</td><td>Median</td><td>3Q</td><td>Max</td></tr><tr><td>-0.5293</td><td>-0.1433</td><td>0.0206</td><td>0.1323</td><td>0.3775</td></tr></table> <p>Coefficients:</p> <table><tr><td></td><td>Estimate</td><td>Std. Error</td><td>t value</td></tr><tr><td>Pr(&gt; t )</td><td></td><td></td><td></td></tr><tr><td>(Intercept)</td><td>0.87955</td><td>0.23537</td><td>3.737</td></tr><tr><td>0.000925 ***</td><td></td><td></td><td></td></tr><tr><td>BodyLenR</td><td>0.03319</td><td>0.01156</td><td>2.872</td></tr><tr><td>0.008008 **</td><td></td><td></td><td></td></tr><tr><td>---</td><td></td><td></td><td></td></tr><tr><td>Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</td><td></td><td></td><td></td></tr><tr><td>Residual standard error: 0.2119 on 26 degrees of freedom</td><td></td><td></td><td></td></tr><tr><td>Multiple R-squared: 0.2409, Adjusted R-squared: 0.2117</td><td></td><td></td><td></td></tr><tr><td>F-statistic: 8.249 on 1 and 26 DF, p-value: 0.008008</td><td></td><td></td><td></td></tr></table>	Min	1Q	Median	3Q	Max	-0.5293	-0.1433	0.0206	0.1323	0.3775		Estimate	Std. Error	t value	Pr(> t )				(Intercept)	0.87955	0.23537	3.737	0.000925 ***				BodyLenR	0.03319	0.01156	2.872	0.008008 **				---				Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				Residual standard error: 0.2119 on 26 degrees of freedom				Multiple R-squared: 0.2409, Adjusted R-squared: 0.2117				F-statistic: 8.249 on 1 and 26 DF, p-value: 0.008008			
Min	1Q	Median	3Q																																																																																																																
-0.38477	-0.15164	-0.01741	0.12428																																																																																																																
Max																																																																																																																			
0.88523																																																																																																																			
	Estimate	Std. Error	t value																																																																																																																
Pr(> t )																																																																																																																			
(Intercept)	1.11736	0.30688	3.641																																																																																																																
0.00118 **																																																																																																																			
BodyLenC	0.02421	0.01351	1.792																																																																																																																
0.08472 .																																																																																																																			
---																																																																																																																			
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1																																																																																																																			
Residual standard error: 0.2653 on 26 degrees of freedom																																																																																																																			
Multiple R-squared: 0.11, Adjusted R-squared: 0.07574																																																																																																																			
F-statistic: 3.213 on 1 and 26 DF, p-value: 0.08472																																																																																																																			
Min	1Q	Median	3Q	Max																																																																																																															
-0.5293	-0.1433	0.0206	0.1323	0.3775																																																																																																															
	Estimate	Std. Error	t value																																																																																																																
Pr(> t )																																																																																																																			
(Intercept)	0.87955	0.23537	3.737																																																																																																																
0.000925 ***																																																																																																																			
BodyLenR	0.03319	0.01156	2.872																																																																																																																
0.008008 **																																																																																																																			
---																																																																																																																			
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1																																																																																																																			
Residual standard error: 0.2119 on 26 degrees of freedom																																																																																																																			
Multiple R-squared: 0.2409, Adjusted R-squared: 0.2117																																																																																																																			
F-statistic: 8.249 on 1 and 26 DF, p-value: 0.008008																																																																																																																			



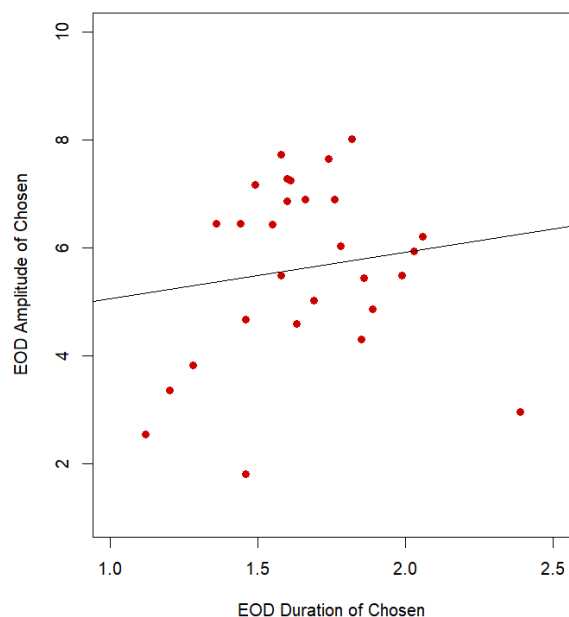
The **slopes** for the two graphs appear very similar on the scatter plots, again, and are 0.30844 for chosen and 0.43495 (steeper) for rejected. The graph of chosen males has a higher **y-intercept** (-1.27921) than that of the rejected (-3.70646). As for their correlations, the graph of chosen has an  $R^2$  of 0.4976, and the graph of rejected has an  $R^2$  of 0.6569. See computer output below, including **df**, **f-statistic**, **p-value**, and **five number summary**.

Body Length vs. EOD Amplitude (Chosen)	Body Length vs. EOD Amplitude (Rejected)																																																																				
<p>Call: lm(formula = AmplitudeC ~ BodyLenC)</p> <p>Residuals:</p> <table><tr><th>Min</th><th>1Q</th><th>Median</th><th>3Q</th><th>Max</th></tr><tr><td>-1.8572</td><td>-0.9604</td><td>0.2131</td><td>1.1029</td><td>1.9150</td></tr></table> <p>Coefficients:</p> <table><tr><th></th><th>Estimate</th><th>Std. Error</th><th>t value</th></tr><tr><td>Pr(&gt; t )</td><td></td><td></td><td></td></tr><tr><td>(Intercept)</td><td>-1.27921</td><td>1.38080</td><td>-0.926</td></tr><tr><td>0.363</td><td></td><td></td><td></td></tr><tr><td>BodyLenC</td><td>0.30844</td><td>0.06079</td><td>5.074</td></tr><tr><td>2.76e-05 ***</td><td></td><td></td><td></td></tr></table> <p>--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</p> <p>Residual standard error: 1.194 on 26 degrees of freedom Multiple R-squared: 0.4976, Adjusted</p>	Min	1Q	Median	3Q	Max	-1.8572	-0.9604	0.2131	1.1029	1.9150		Estimate	Std. Error	t value	Pr(> t )				(Intercept)	-1.27921	1.38080	-0.926	0.363				BodyLenC	0.30844	0.06079	5.074	2.76e-05 ***				<p>Call: lm(formula = AmplitudeR ~ BodyLenR)</p> <p>Residuals:</p> <table><tr><th>Min</th><th>1Q</th><th>Median</th><th>3Q</th><th>Max</th></tr><tr><td>-1.9175</td><td>-0.5289</td><td>-0.2273</td><td>0.4681</td><td>3.5950</td></tr></table> <p>Coefficients:</p> <table><tr><th></th><th>Estimate</th><th>Std. Error</th><th>t value</th></tr><tr><td>Pr(&gt; t )</td><td></td><td></td><td></td></tr><tr><td>(Intercept)</td><td>-3.70646</td><td>1.25562</td><td>-2.952</td></tr><tr><td>0.00661 **</td><td></td><td></td><td></td></tr><tr><td>BodyLenR</td><td>0.43495</td><td>0.06165</td><td>7.056</td></tr><tr><td>1.71e-07 ***</td><td></td><td></td><td></td></tr></table> <p>--- Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1</p> <p>Residual standard error: 1.131 on 26 degrees of freedom Multiple R-squared: 0.6569, Adjusted</p>	Min	1Q	Median	3Q	Max	-1.9175	-0.5289	-0.2273	0.4681	3.5950		Estimate	Std. Error	t value	Pr(> t )				(Intercept)	-3.70646	1.25562	-2.952	0.00661 **				BodyLenR	0.43495	0.06165	7.056	1.71e-07 ***			
Min	1Q	Median	3Q	Max																																																																	
-1.8572	-0.9604	0.2131	1.1029	1.9150																																																																	
	Estimate	Std. Error	t value																																																																		
Pr(> t )																																																																					
(Intercept)	-1.27921	1.38080	-0.926																																																																		
0.363																																																																					
BodyLenC	0.30844	0.06079	5.074																																																																		
2.76e-05 ***																																																																					
Min	1Q	Median	3Q	Max																																																																	
-1.9175	-0.5289	-0.2273	0.4681	3.5950																																																																	
	Estimate	Std. Error	t value																																																																		
Pr(> t )																																																																					
(Intercept)	-3.70646	1.25562	-2.952																																																																		
0.00661 **																																																																					
BodyLenR	0.43495	0.06165	7.056																																																																		
1.71e-07 ***																																																																					

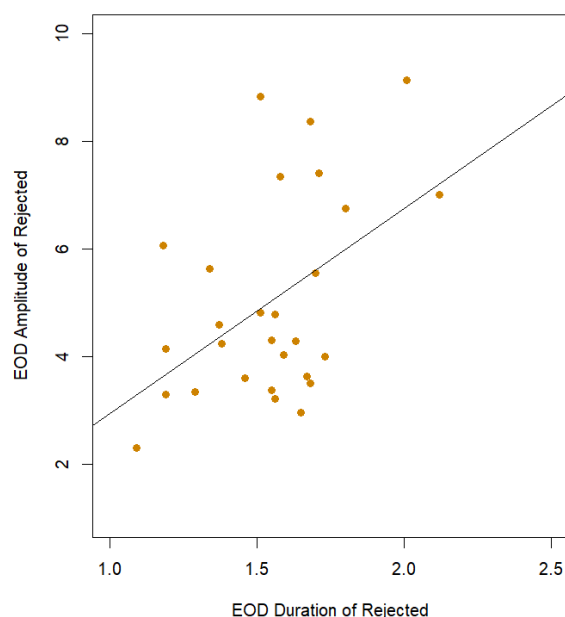
R-squared: 0.4782  
 F-statistic: 25.75 on 1 and 26 DF,  
 p-value: 2.763e-05

R-squared: 0.6437  
 F-statistic: 49.78 on 1 and 26 DF,  
 p-value: 1.714e-07

EOD Duration vs. EOD Amplitude of Chosen Males



EOD Duration vs. EOD Amplitude of Rejected Males



The slopes for the two graphs appear different on the scatter plots, and are 0.8653 for chosen and 3.8100 (much steeper) for rejected. The graph of chosen males has a much higher y-intercept (4.1968) than that of the rejected (-0.8657). As for their correlations, the graph of chosen has an  $R^2$  of 0.02088, and the graph of rejected has an  $R^2$  of 0.2305. See computer output below, including df, f-statistic, p-value, and five number summary.

EOD Duration vs. EOD Amplitude (Chosen)	EOD Duration vs. EOD Amplitude (Rejected)																																																								
Call: lm(formula = AmplitudeC ~ DurationC)	Call: lm(formula = AmplitudeR ~ DurationR)																																																								
Residuals:	Residuals:																																																								
<table><tr><td>Min</td><td>1Q</td><td>Median</td><td>3Q</td><td>Max</td></tr><tr><td>-3.6501</td><td>-0.9835</td><td>0.1136</td><td>1.2723</td><td>2.2483</td></tr></table>	Min	1Q	Median	3Q	Max	-3.6501	-0.9835	0.1136	1.2723	2.2483	<table><tr><td>Min</td><td>1Q</td><td>Median</td><td>3Q</td><td>Max</td></tr><tr><td>-2.4609</td><td>-1.1033</td><td>-0.2448</td><td>0.9232</td><td>3.9525</td></tr></table>	Min	1Q	Median	3Q	Max	-2.4609	-1.1033	-0.2448	0.9232	3.9525																																				
Min	1Q	Median	3Q	Max																																																					
-3.6501	-0.9835	0.1136	1.2723	2.2483																																																					
Min	1Q	Median	3Q	Max																																																					
-2.4609	-1.1033	-0.2448	0.9232	3.9525																																																					
Coefficients:	Coefficients:																																																								
<table><tr><td></td><td>Estimate</td><td>Std. Error</td><td>t value</td></tr><tr><td>Pr(&gt; t )</td><td></td><td></td><td></td></tr><tr><td>(Intercept)</td><td>4.1968</td><td>1.9547</td><td>2.147</td></tr><tr><td>0.0413 *</td><td></td><td></td><td></td></tr><tr><td>DurationC</td><td>0.8653</td><td>1.1621</td><td>0.745</td></tr><tr><td>0.4632</td><td></td><td></td><td></td></tr><tr><td>---</td><td></td><td></td><td></td></tr></table>		Estimate	Std. Error	t value	Pr(> t )				(Intercept)	4.1968	1.9547	2.147	0.0413 *				DurationC	0.8653	1.1621	0.745	0.4632				---				<table><tr><td></td><td>Estimate</td><td>Std. Error</td><td>t value</td></tr><tr><td>Pr(&gt; t )</td><td></td><td></td><td></td></tr><tr><td>(Intercept)</td><td>-0.8657</td><td>2.1342</td><td>-0.406</td></tr><tr><td>0.68835</td><td></td><td></td><td></td></tr><tr><td>DurationR</td><td>3.8100</td><td>1.3651</td><td>2.791</td></tr><tr><td>0.00971 **</td><td></td><td></td><td></td></tr><tr><td>---</td><td></td><td></td><td></td></tr></table>		Estimate	Std. Error	t value	Pr(> t )				(Intercept)	-0.8657	2.1342	-0.406	0.68835				DurationR	3.8100	1.3651	2.791	0.00971 **				---			
	Estimate	Std. Error	t value																																																						
Pr(> t )																																																									
(Intercept)	4.1968	1.9547	2.147																																																						
0.0413 *																																																									
DurationC	0.8653	1.1621	0.745																																																						
0.4632																																																									
---																																																									
	Estimate	Std. Error	t value																																																						
Pr(> t )																																																									
(Intercept)	-0.8657	2.1342	-0.406																																																						
0.68835																																																									
DurationR	3.8100	1.3651	2.791																																																						
0.00971 **																																																									
---																																																									



Signif. codes: 0 '***' 0.001 '**' 0.01 *' 0.05 '.' 0.1 ' ' 1  Residual standard error: 1.666 on 26 degrees of freedom Multiple R-squared: 0.02088, Adjusted R-squared: -0.01678 F-statistic: 0.5544 on 1 and 26 DF, p-value: 0.4632	Signif. codes: 0 '***' 0.001 '**' 0.01 *' 0.05 '.' 0.1 ' ' 1  Residual standard error: 1.693 on 26 degrees of freedom Multiple R-squared: 0.2305, Adjusted R-squared: 0.2009 F-statistic: 7.79 on 1 and 26 DF, p-value: 0.009714
---	--

#### 4 & 5. Conclusion questions:

4. The single characteristic difference between chosen and rejected males that feels most significant is body length of the fish. The only t-test where we were able to reject the null hypothesis and conclude significance was the one comparing the means of body length between chosen and rejected males, adopting the alternative hypothesis that the true difference in these means is not equal to zero. This is visible via the boxplots we created and their corresponding five number summaries, which we used to compare shape and spread of characteristics; the boxplot comparing body length between chosen and rejected males possesses the greatest difference between the means and medians for all boxplots comparing variables between chosen and rejected males. Body length also tends to be on the greater side for chosen than rejected males, as depicted in the above scatter plots.

5. The explanatory variable in this experiment is fish body length, as it is the only variable out of the three that the conductors of the experiment manipulated. As detailed in the experiment, a different female was placed in a tank between two males that differed in lengths by at least four centimeters. Also, we must assume random assignment happened for the experimental design to be sound. With this condition met, the study is an experimental one with controlled treatments and random selection, and we can safely conclude that male body length causes female mate selection. A clear example of this is the boxplot for body length, which demonstrates that body length for chosen males tends to be much higher than those for rejected males. Therefore, body length seems to be a good predictor for whether male fish will be chosen or rejected.