

Investigating the Effect of Gender on Freestyle Swimming Performance: Implications for Transgender Athletes in Female Sports

TMA4267 Linear Statistical Models Spring 2023

Diana Dorobantu

April 25, 2023

Abstract

The issue of transgender athletes competing in female sports has sparked intense debate in recent years, with concerns raised about the potential advantages that transgender female athletes may have over their cisgender counterparts. Using Design of Experiments (DOE) methodology, the aim of this project is to explore the impact of gender on freestyle swimming performance, which is a critical area of concern in this debate. The study will involve analyzing the time laps of male and female swimmers to gain insights into the differences in technique and speed between the two genders. The experiment will utilize a k-factor design with two levels per factor to systematically evaluate the effects of each factor and their interactions on the time taken to complete the task. The experimental design will be generated using R/R-Studio, and a randomized experimental plan will be followed. The time taken to complete the task will be recorded and analyzed using statistical techniques such as Analysis of Variance (ANOVA) and regression analysis. The results of this study will provide valuable information on the extent to which gender affects performance in freestyle swimming and its potential implications for transgender athletes in female sports.

Issue addressed

In recent years, the issue of transgender athletes in female sports has been a topic of intense debate. One particular area of concern is freestyle swimming, where gender may play a significant role in performance. As a swimmer myself, I have witnessed firsthand the differences in technique and speed between male and female swimmers. In this project, I aim to explore the effect of gender on freestyle swimming and its implications for transgender athletes in female sports. By examining the time laps of male and female swimmers, I hope to gain a deeper understanding of the impact gender has on performance in this sport.

Selection of Factors and Levels

Based on the problem described, the relevant factors could be: Gender (Male [1] vs. Female[0]), Warm-up time (short [0] vs. long[1]), Breathing technique (bilateral[2] vs. unilateral[1]) and Distance (50m vs. 100m).

It is possible that there may be interactions between some of the factors, such as the effect of breathing technique being more pronounced for longer distances or for one gender versus the other.

The levels for each factor could be determined based on common practice or existing research. For example, warm-up times could be set to 5 minutes and 10 minutes, as those are commonly used warm-up durations. Bilateral breathing and unilateral breathing could be the two levels for the breathing technique factor, as

those are the two most common techniques used by swimmers. For distance, 50m and 100m could be used as those are common race distances.

To control that the factors are at the desired level, the experiment could be conducted in a controlled environment such as a swimming pool with standardized conditions. The swimmers could be given specific instructions on the warm-up time and breathing technique to use. The distance swam could be measured precisely using timing equipment and markers on the pool deck.

Selection of Response Variable

The response variable of interest is the Time it takes for the swimmer to complete the designated distance. There may be other response variables of interest such as: Heart Rate, Oxygen Consumption, Stroke Count or Stroke Rate, but for the purposes of this study I consider that Time is the most relevant response variable.

The response variable can be measured using accurate timing equipment such as a stopwatch or touch pads. The experiment could also be recorded using video equipment to allow for further analysis and measurement. For this trial a stopwatch was used.

The accuracy of the measurements can be ensured by using standardized timing devices, as well as conducting multiple trials to obtain an average time for each factor combination. The accuracy of the measurements can also be improved by having multiple trained individuals measure the time to reduce measurement errors.

Implementation of the experiment

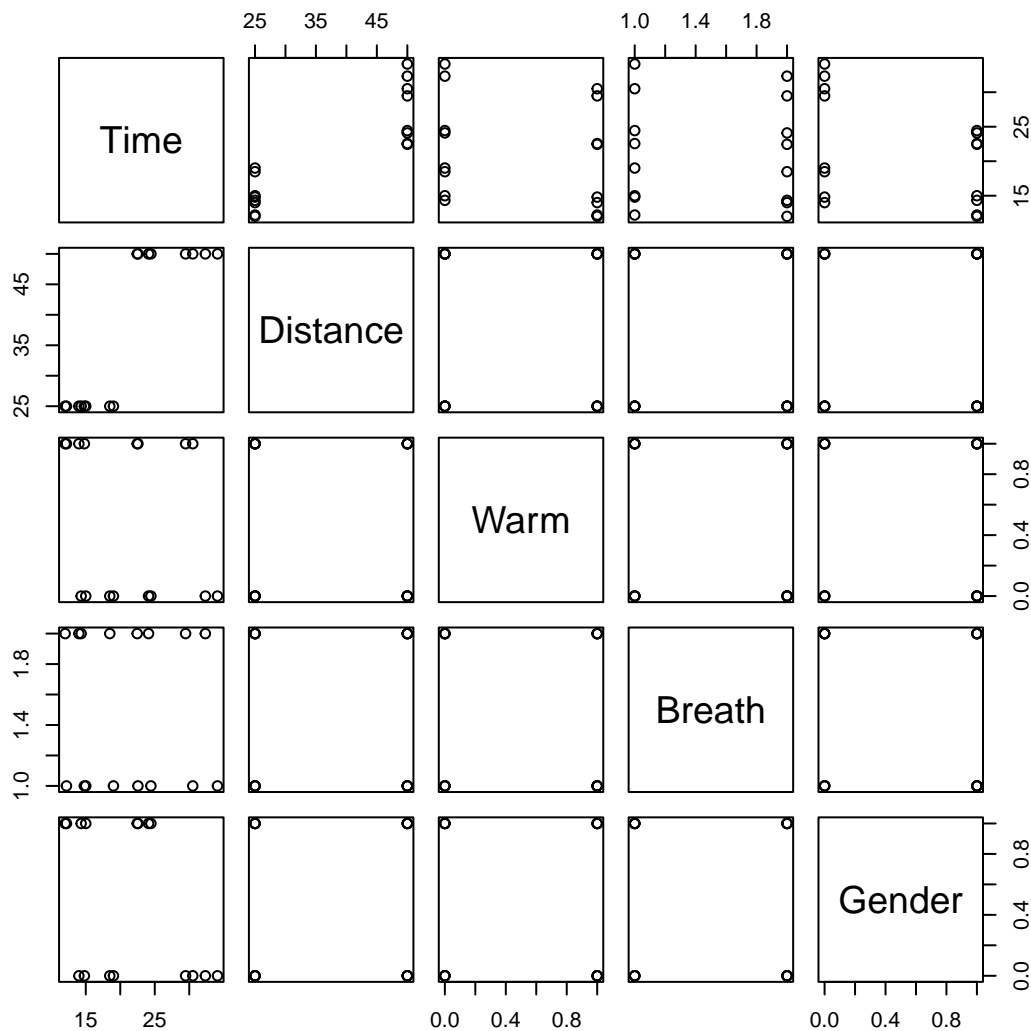
To ensure the validity of the experiment, it is crucial to adhere to a randomization process when assigning treatments to experimental units. Any deviation from proper randomization can introduce bias and decrease the accuracy of the results. It is also important to ensure that each trial is a genuine replicate, performed independently and constituting a full trial, to account for the total variability of the experiment and increase the reliability of the results. Without genuine replicates, the variability in the results may not be fully accounted for, leading to less accurate conclusions. In my particular experiment, I conducted the trials over multiple mornings and then calculated the average of the runs to increase the accuracy of the results.

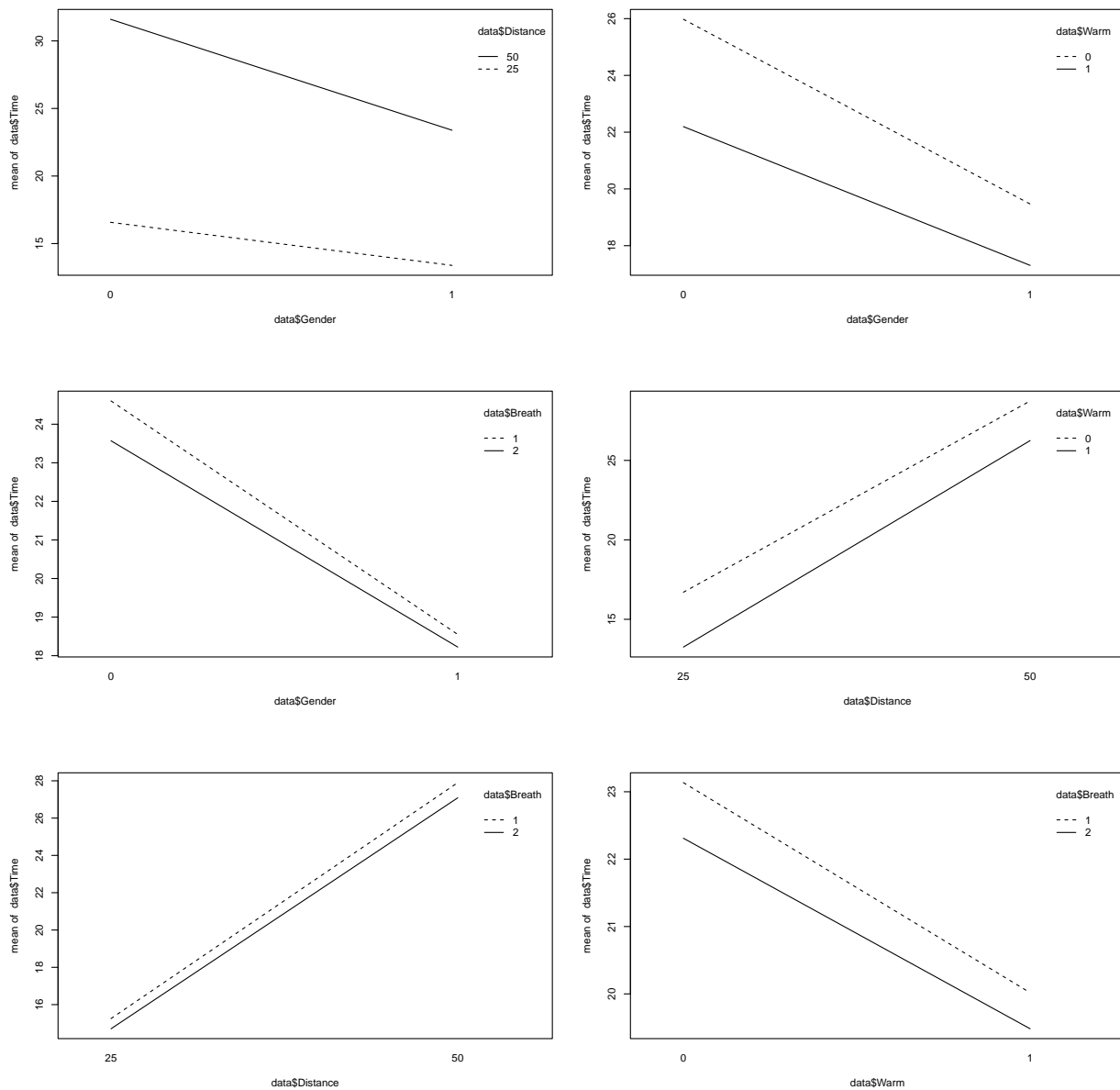
##	Run	Time	Distance	Warm	Breath	Gender
## 1	1	19.01	25	0	1	0
## 2	2	34.09	50	0	1	0
## 3	3	14.80	25	1	1	0
## 4	4	30.51	50	1	1	0
## 5	5	18.48	25	0	2	0
## 6	6	32.33	50	0	2	0
## 7	7	14.00	25	1	2	0
## 8	8	29.47	50	1	2	0
## 9	9	15.00	25	0	1	1
## 10	10	24.44	50	0	1	1
## 11	11	12.20	25	1	1	1
## 12	12	22.56	50	1	1	1
## 13	13	14.33	25	0	2	1
## 14	14	24.10	50	0	2	1
## 15	15	12.01	25	1	2	1
## 16	16	22.45	50	1	2	1

Choice of design:

A 2k factorial design may be appropriate based on the number of factors and levels involved. The required resolution of the design will depend on the expected level of interaction between the factors. If there is a possibility of a high degree of interaction, a higher resolution design may be needed. A blocked design may be desirable to control for extraneous variables that could influence the results. In my experiment, I conducted all trials in the morning following a week of preparation, including exercise, sleep, and diet, to ensure a fair trial. Replicates might also be necessary to ensure the reliability of the results and to estimate experimental error. In my case, I conducted multiple runs and calculated the average for the final result. Depending on the specific research question and available resources, a 2k-p fractional factorial design or other designs may also be considered.

Analysis of data





Modelling data

Model 1

→ linear model that includes all four predictor variables (Distance, Warm, Breath, and Gender)

```
##
## Call:
## lm(formula = Time ~ ., data = data)
##
## Residuals:
```

```
##      Min      1Q  Median      3Q      Max
## -2.0300 -1.1256  0.2362  1.0956  1.9200
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   7.82000    1.88662   4.145  0.00163 **
## Distance      0.50060    0.03294  15.199  9.92e-09 ***
## Warm         -2.97250    0.82339  -3.610  0.00410 **
## Breath       -0.68000    0.82339  -0.826  0.42644
## Gender       -5.70000    0.82339  -6.923  2.51e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.647 on 11 degrees of freedom
## Multiple R-squared:  0.9638, Adjusted R-squared:  0.9506
## F-statistic: 73.16 on 4 and 11 DF,  p-value: 7.481e-08
```

- Multiple R-squared value (0.9638) means that approximately 96.38% of the variation in “Time” can be explained by the predictor variables.

- F-statistic represents the overall significance of the model, and its associated p-value is 7.481e-08 which very small, indicating that the model as a whole is a good fit for the data.

- Intercept coefficient (7.82000) represents the estimated average value of “Time” when all other predictor variables are zero.

- coefficients for “Distance”(0.50060), “Warm” (-2.97250), “Breath”(-0.68000) and “Gender”(-5.70000) indicate the estimated change in “Time” associated with a one-unit increase in each predictor variable, holding all other predictor variables constant; p-values associated with each coefficient suggest that “Distance”, “Gender” and “Warm” are significant predictors of “Time,” whereas “Breath” is not.

Model 2

→ quadratic model that includes the four predictor variables and all their two-way interactions (Distance:Warm, Distance:Breath, Distance:Gender, Warm:Breath, Warm:Gender, and Breath:Gender)

```
##
## Call:
## lm(formula = Time ~ .^2, data = data)
##
## Residuals:
##      1      2      3      4      5      6      7      8
## -0.21625  0.18375 -0.01875  0.05125  0.29875 -0.26625 -0.06375  0.03125
##      9     10     11     12     13     14     15     16
##  0.12375 -0.09125  0.11125 -0.14375 -0.20625  0.17375 -0.02875  0.06125
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   5.32625    0.81541   6.532  0.001258 **
## Distance      0.59780    0.01917  31.189  6.36e-07 ***
## Warm         -5.65750    0.61861  -9.146  0.000262 ***
## Breath       -0.78000    0.47917  -1.628  0.164494
## Gender       -0.03000    0.61861  -0.048  0.963198
## Distance:Warm  0.03840    0.01107   3.470  0.017848 *
```

```
## Distance:Breath -0.01060    0.01107  -0.958 0.382110
## Distance:Gender -0.20100    0.01107 -18.164 9.30e-06 ***
## Warm:Breath      0.29000    0.27665   1.048 0.342524
## Warm:Gender      1.62000    0.27665   5.856 0.002058 **
## Breath:Gender     0.70500    0.27665   2.548 0.051369 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2766 on 5 degrees of freedom
## Multiple R-squared:  0.9995, Adjusted R-squared:  0.9986
## F-statistic: 1075 on 10 and 5 DF,  p-value: 1.091e-07
```

· Multiple R-squared value (0.9995) means that approximately 99.95% of the variation in “Time” can be explained by the predictor variables.

· F-statistic represents the overall significance of the model, and its associated p-value is 1075 which very large, indicating that there is a large difference between the means of the groups being compared, therefore the model is not significant.

Model 3

→ cubic model that includes the four predictor variables, all their two-way interactions, and all their three-way interactions (Distance:Warm:Breath, Distance:Warm:Gender, Distance:Breath:Gender, and Warm:Breath:Gender)

```
##
## Call:
## lm(formula = Time ~ .^3, data = data)
##
## Residuals:
##      1      2      3      4      5      6      7      8      9     10
## -0.0775  0.0775  0.0775 -0.0775  0.0775 -0.0775 -0.0775  0.0775  0.0775 -0.0775
##     11     12     13     14     15     16
## -0.0775  0.0775 -0.0775  0.0775  0.0775 -0.0775
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.92750    1.38419   2.837  0.2157
## Distance          0.63380    0.03452  18.360  0.0346 *
## Warm             -4.97500    1.69085  -2.942  0.2086
## Breath            0.23500    0.86300   0.272  0.8307
## Gender            1.93500    1.69085   1.144  0.4572
## Distance:Warm      0.02280    0.04113   0.554  0.6777
## Distance:Breath    -0.03680    0.02148  -1.713  0.3363
## Distance:Gender    -0.25080    0.04113  -6.098  0.1035
## Warm:Breath        -0.33000    1.02815  -0.321  0.8023
## Warm:Gender         1.92000    1.35126   1.421  0.3904
## Breath:Gender       -0.77000    1.02815  -0.749  0.5908
## Distance:Warm:Breath  0.01480    0.02480   0.597  0.6575
## Distance:Warm:Gender -0.01320    0.02480  -0.532  0.6886
## Distance:Breath:Gender 0.03760    0.02480   1.516  0.3712
## Warm:Breath:Gender   0.13000    0.62000   0.210  0.8684
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.31 on 1 degrees of freedom
## Multiple R-squared:  0.9999, Adjusted R-squared:  0.9982
## F-statistic: 612 on 14 and 1 DF, p-value: 0.03167
```

- Multiple R-squared value (0.9999) means that approximately 99.99% of the variation in “Time” can be explained by the predictor variables.
- F-statistic represents the overall significance of the model, and its associated p-value is 612 which very large, indicating this model is not significant as well.

Model 4

→ full model that includes the four predictor variables and all their interactions up to four-way interaction (Distance:Gender:Warm:Breath)

```
##
## Call:
## lm(formula = Time ~ .^4, data = data)
##
## Residuals:
## ALL 16 residuals are 0: no residual degrees of freedom!
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      3.2300         NaN      NaN      NaN
## Distance          0.6524         NaN      NaN      NaN
## Warm            -3.5800         NaN      NaN      NaN
## Breath           0.7000         NaN      NaN      NaN
## Gender           3.3300         NaN      NaN      NaN
## Distance:Warm    -0.0144         NaN      NaN      NaN
## Distance:Breath  -0.0492         NaN      NaN      NaN
## Distance:Gender  -0.2880         NaN      NaN      NaN
## Warm:Breath     -1.2600         NaN      NaN      NaN
## Warm:Gender     -0.8700         NaN      NaN      NaN
## Breath:Gender   -1.7000         NaN      NaN      NaN
## Distance:Warm:Breath  0.0396         NaN      NaN      NaN
## Distance:Warm:Gender  0.0612         NaN      NaN      NaN
## Distance:Breath:Gender 0.0624         NaN      NaN      NaN
## Warm:Breath:Gender  1.9900         NaN      NaN      NaN
## Distance:Warm:Breath:Gender -0.0496         NaN      NaN      NaN
##
## Residual standard error: NaN on 0 degrees of freedom
## Multiple R-squared:  1, Adjusted R-squared:  NaN
## F-statistic:  NaN on 15 and 0 DF, p-value: NA
```

- Since we have no replicates, we have no way of assessing the significance of any of the coefficients of this full model with all interactions.
- All of the estimated coefficients and standard errors are missing and replaced with “NaN”. This is due to a perfect fit of the model to the data, resulting in all residuals being zero. Therefore, the model explains all of the variation in the data and the estimated coefficients are not meaningful. The adjusted R-squared value is

also not reported because it requires non-zero residual degrees of freedom. The F-statistic is not meaningful because the denominator degrees of freedom are zero. Overall, this output suggests that the model is not useful for making inferences or predictions on new data.

Conclusion and Recommendations

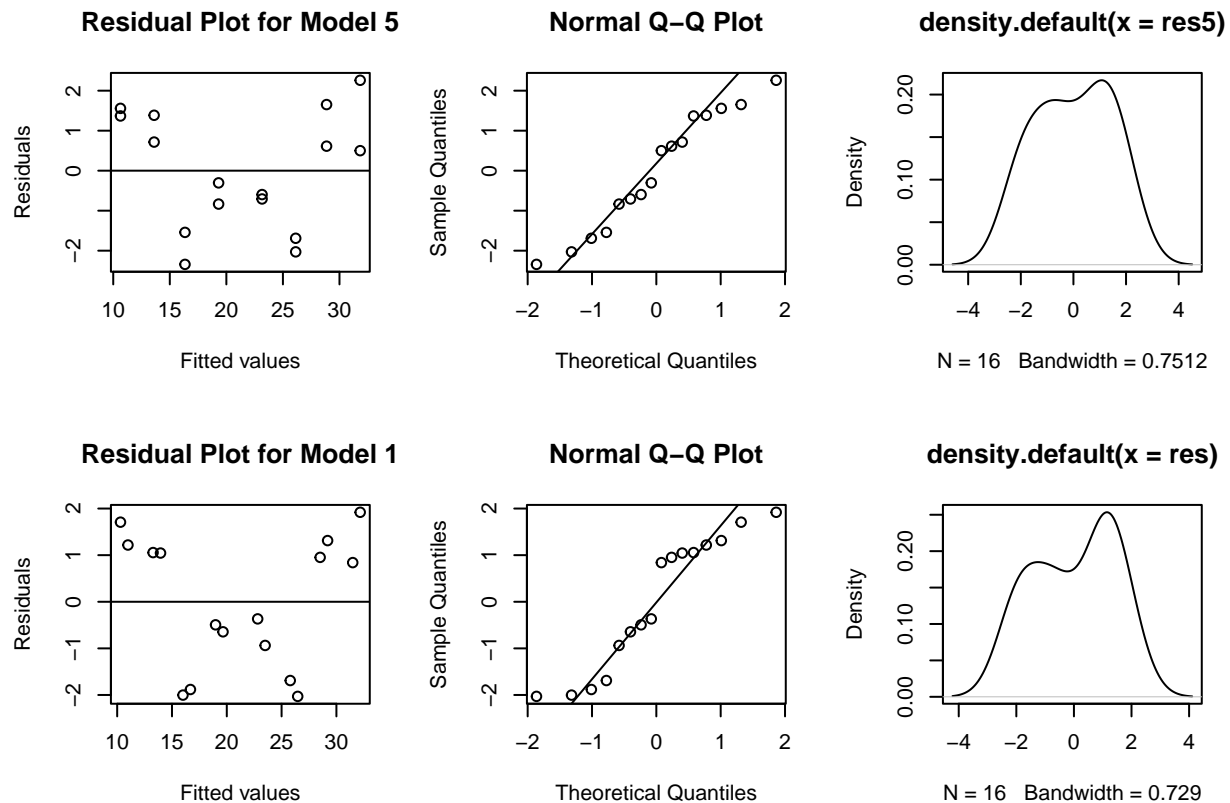
From the statistical analysis of Model 1 we saw that “Breath” it is not a significant predictor, therefore we create another model excluding it.

Model 5

→ linear model that includes three predictor variables (Distance, Warm, Gender)

```
##
## Call:
## lm(formula = Time ~ Distance + Warm + Gender, data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3425 -1.0119  0.0975  1.3719  2.2600
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.8000     1.4071   4.833  0.00041 ***
## Distance      0.5006     0.0325  15.405 2.87e-09 ***
## Warm         -2.9725     0.8124  -3.659  0.00327 **
## Gender        -5.7000     0.8124  -7.016 1.40e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.625 on 12 degrees of freedom
## Multiple R-squared:  0.9615, Adjusted R-squared:  0.9519
## F-statistic: 99.97 on 3 and 12 DF,  p-value: 9.349e-09
```

- (1) In the Residual Plots we can see that the spread of the residuals tends to be higher for lowest and highest fitted values in both models, but it doesn't look serious enough that we would need to make any changes to the models.
- (2) In the Q-Q Plots we can see that the residuals tend to stray from the line quite a bit near the tails in both models, which could indicate that they're not normally distributed.
- (3) In the Density Plot for Model 1 we can see that the density roughly follows a bell shape which enforce that the residuals are not normally distributed. In comparison, Model 5 is more suitable as follows more a bell shape, but moving forward we can decide to perform a transformation on the data to ensure that the residuals are more normally distributed.



ANOVA

```
## Analysis of Variance Table
##
## Model 1: Time ~ Distance + Warm + Breath + Gender
## Model 2: Time ~ Distance + Warm + Gender
##   Res.Df    RSS Df Sum of Sq    F Pr(>F)
## 1      11 29.831
## 2      12 31.680 -1   -1.8496 0.682 0.4264
```

· In this case, Model 1 has 11 Residual Degrees of Freedom and a Residual Sum of Squares of 29.831, while Model 5 has 12 Residual Degrees of Freedom and a Residual Sum of Squares of 31.680. The change in degrees of freedom between the two models is -1, indicating that Model 1 has one more parameter than Model 2. The change in Residual Sum of Squares between the two models is -1.8496, indicating that Model 1 fits the data better than Model 5. However, the F-statistic of 0.682 is not significant at the 0.05 level (p-value of 0.4264 > 0.05). Therefore, the ANOVA table shows that the difference between the two models is not statistically significant, as the F-statistic is not large enough to reject the null hypothesis that the two models are equivalent. This suggests that the “Breath” variable does not provide significant additional explanatory power for the response variable Time beyond that provided by the other three predictor variables included in Model 5.

To summarize our experiment, Model 1 emerged as the optimal choice, and we can assert that gender plays a highly influential role in performance within this particular sport. Furthermore, the other explanatory models also support this notion, emphasizing the crucial implications of gender on performance in freestyle swimming.