

# Exam 2

Denis Ostroushko

2022-12-08

## Problem 1

We begin this problem by summarizing the data available to us, and continue this summary into the **1 - A** Section. The data set contains 112 observations for 112 participants in the study.

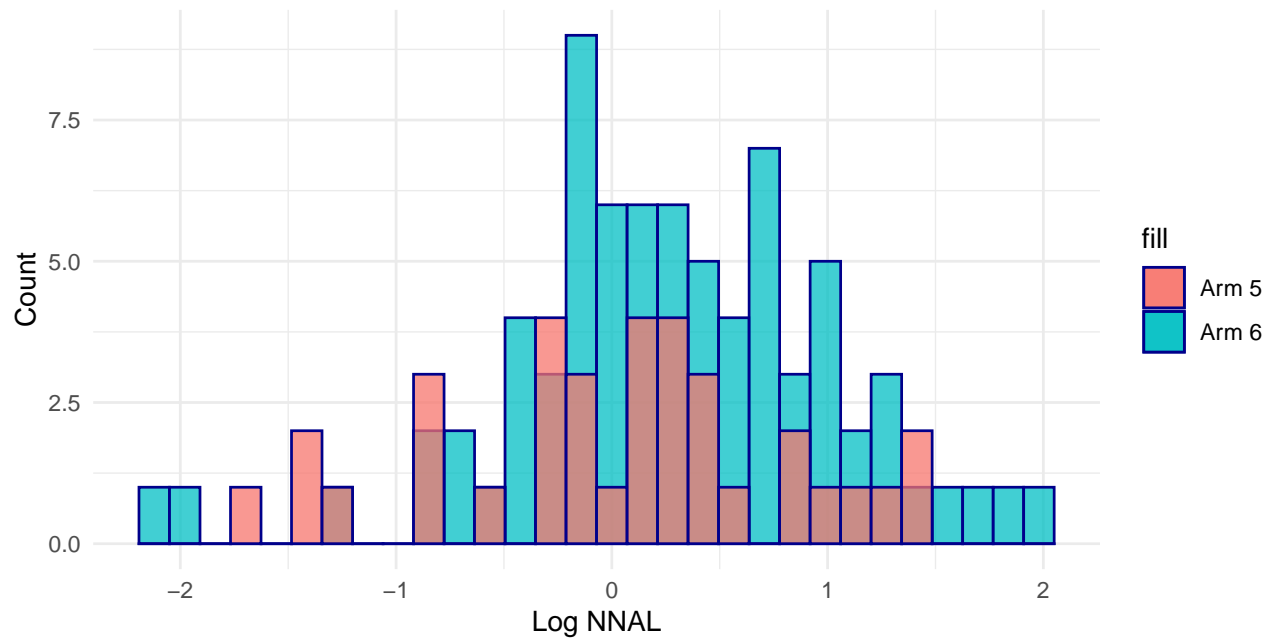
One of the participants has a missing value of baseline NNAL measurement. This variable is important for our analysis, therefore we will omit this observation. The final data set includes 111 observations.

### 1 - A

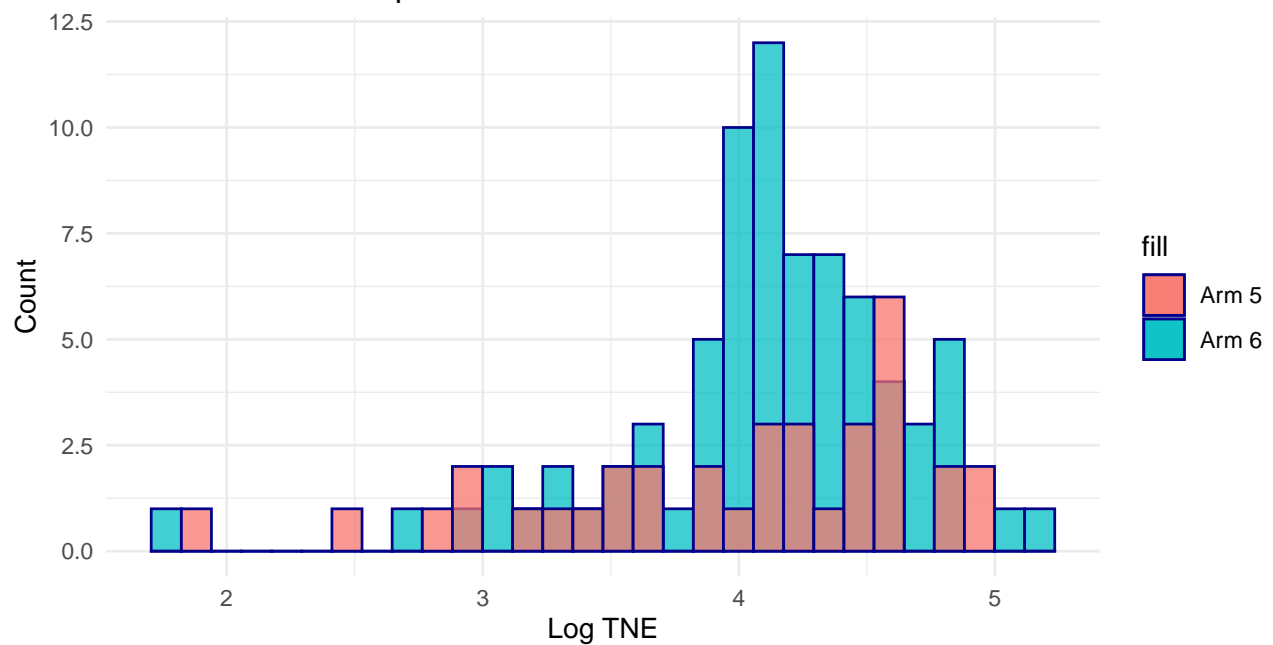
Problem 1-A asks us to fit a logistic regression model using two log-transformed baseline measurements. We are interested in evaluating how the two variables are balanced between the two experiment arms, arm 5 and arm 6.

Arm	N	Baseline Log NNAL		Baseline Log TNE	
		Mean	SD	Mean	SD
5	35	0.0265297	0.8009900	3.963363	0.7431000
6	76	0.2704350	0.7677647	4.097586	0.5577623

Baseline Log NNAL Measurements  
Between Two Groups



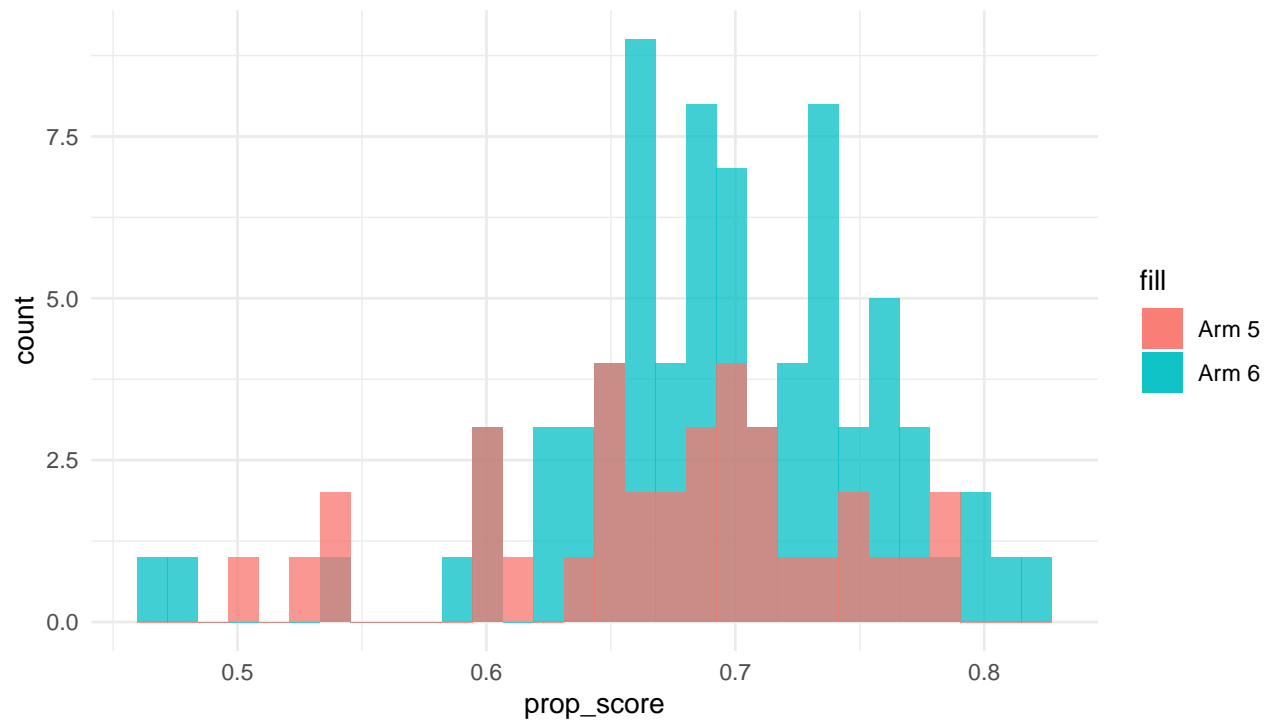
Baseline Log TNE Measurements  
Between Two Groups



**Fit the Model**

Model statement

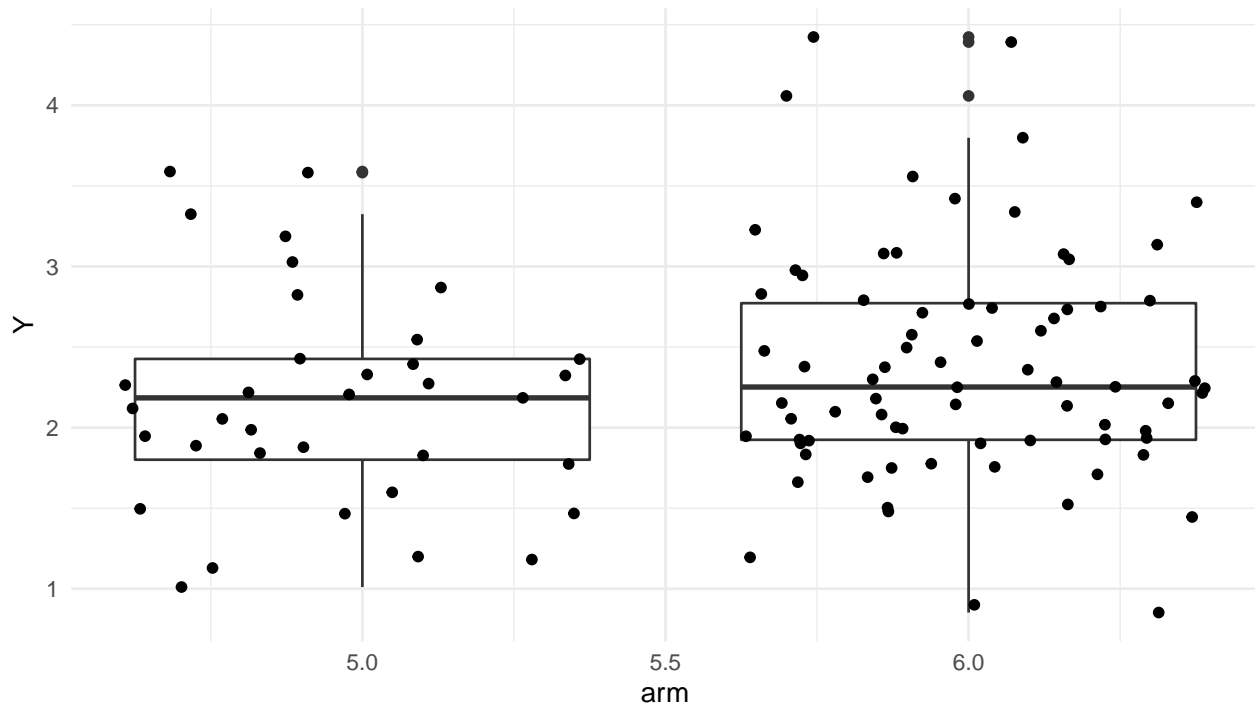
**Summary and plot by group**



### 1 - B

```
## # A tibble: 2 x 4
##   arm     N Mean  SD
##   <dbl> <int> <dbl> <dbl>
## 1     5    35  2.17 0.668
## 2     6    76  2.38 0.712
```

Evaluate variances and outliers



F Test for variances - two tailed test at  $\alpha = 0.05$ , 95% confidence level

[https://www.itl.nist.gov/div898/handbook/eda/section3/eda359.htm#:~:text=An%20F%2Dtest%20\(Snedecor%20and,the%20](https://www.itl.nist.gov/div898/handbook/eda/section3/eda359.htm#:~:text=An%20F%2Dtest%20(Snedecor%20and,the%20)

No outliers, no stat. difference between groups in F tests  $\rightarrow$  conduct a T test

```
## [1] -0.06284174
```

```
## [1] 0.4929838
```

```
##      t
```

```
## 1.543411
```

```
## [1] 0.1272329
```

**1 - C**

effect size

Based on some literature review, many psychology and social science data analysis methods refer to this method as Cohen's d. All sources I reviewed state that in practice effect sizes between 0.2 and 0.5 are considered as Medium size. I will rely on a test instead because we have more accessible and straightforward way to get a confidence interval, without using a Delta Method and other method from statistical theory.

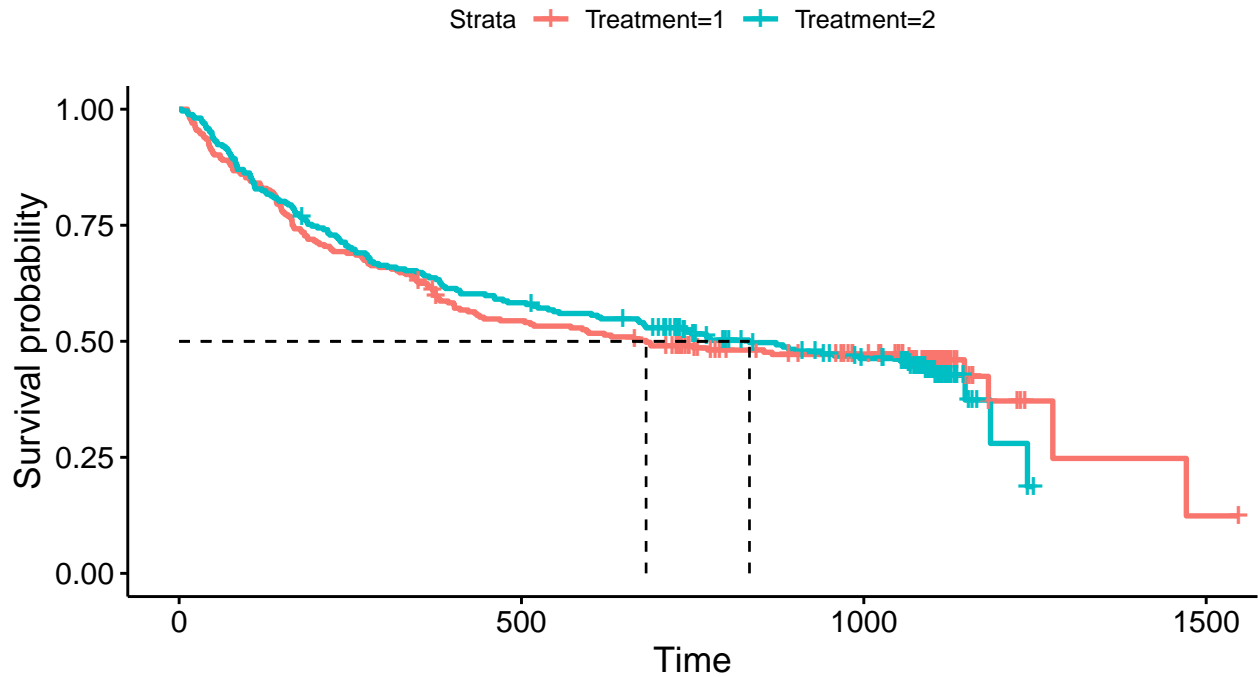
## Problem 2

```
## [1] 551
```

```
## [1] 25
```

Comment on outliers and describe the data set

## 2 - A



Add model statement here

Predictor	Estiamte	Exponentiated Estiamte	Standard Error	Z Value	P value
Treatment	0.002623	1.002626	0.118131	0.022204	0.982285

Interpret the absolute lack of difference between the two groups

## 2 - B

Predictor	Estiamte	Exponentiated Estiamte	Standard Error	Z Value	P value
GENDER2	-0.087586	0.916140	0.120380	-0.727583	0.466869
age	0.017906	1.018067	0.003979	4.499976	0.000007
Treatment	0.032936	1.033485	0.118417	0.278139	0.780906
Race	-0.202888	0.816370	0.196934	-1.030232	0.302901
'HLA-Match'	-0.508781	0.601228	0.185197	-2.747244	0.006010

## 2 - C

Make a table with model estimates from

Model	Predictor	Estiamte	Exponentiated Estiamte	Standard Error	Z Value	P value
Full	Treatment	0.032936	1.033485	0.118417	0.278139	0.780906
Treatment Only	Treatment	0.002623	1.002626	0.118131	0.022204	0.982285