Portfolio 1, Study Group 10

Jesper Fischer, Lasse Hansen & Sarah Hvid Andersen

```r
#Loading packages
pacman::p_load(sjPlot, tidyverse, broom)
```

1a

```r
#importing the data.
sleep <-  read.csv("sleepstudy.csv")
#selecting the data from subject 308
p308 <- subset(sleep, Subject == 308)
#making a model with the data from subject 308
m1 <-  lm(Reaction ~ Days, data = p308)
summary(m1)
```

```
##
## Call:
## lm(formula = Reaction ~ Days, data = p308)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -106.397   -4.098    9.688   22.269   61.674
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   244.19      28.08   8.695 2.39e-05 ***
## Days           21.77       5.26   4.137  0.00326 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 47.78 on 8 degrees of freedom
## Multiple R-squared:  0.6815, Adjusted R-squared:  0.6417
## F-statistic: 17.12 on 1 and 8 DF,  p-value: 0.003265
```

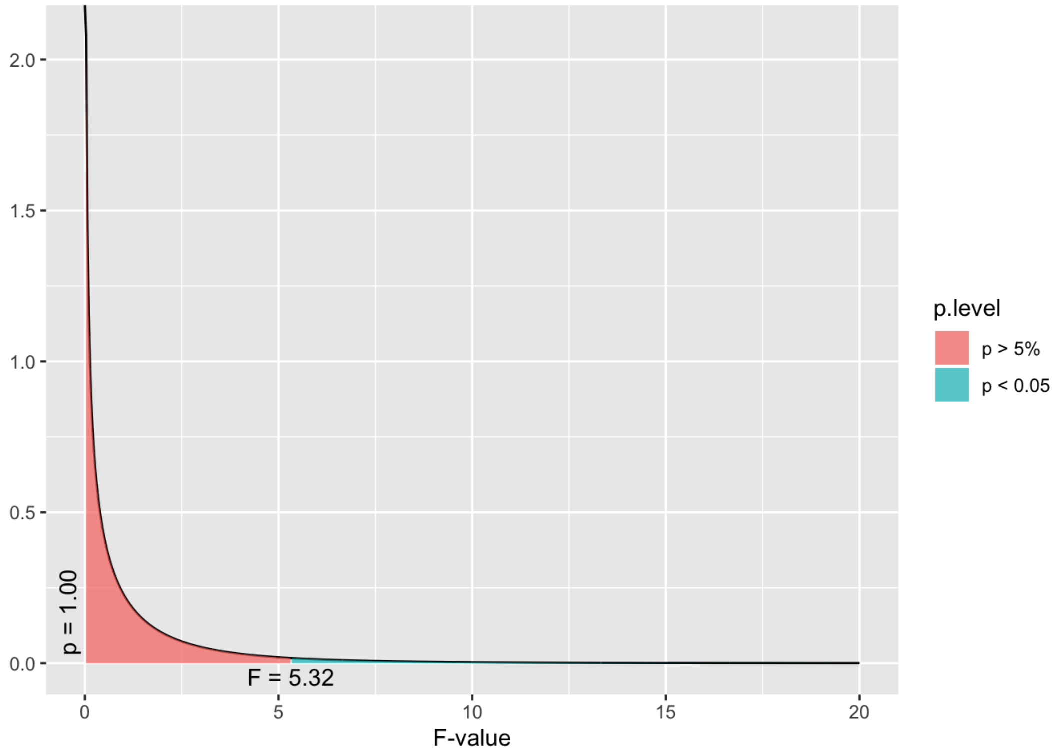
Days participant 308 was sleep deprived explained a significant proportion of variance in reaction time, R^2 = .64, F(1, 8) = 17.12, p = .003.

1b

As reported above the model degrees of freedom = 1 and the residual degrees of freedom = 8.

1c + 1d

```r
dist_f(f=0, deg.f1 = 1, deg.f2 = 8,xmax=20)
```



A regression with these degrees of freedom becomes statistically significant when the f-value gets above 5.32 at $p < .05$.

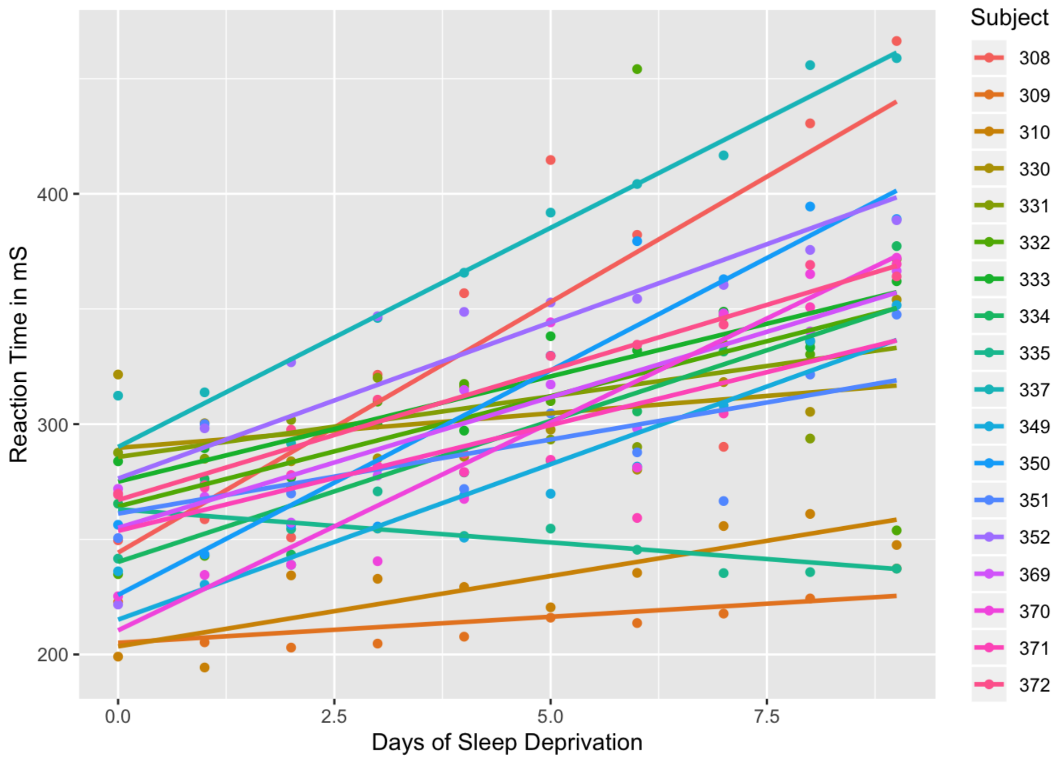
2a

```
#getting the coefficients (slope and intercept) for each participant:
fitted_models <-  sleep %>% group_by(Subject) %>% do(model = lm(Reaction ~ Days, data = .))
#putting the data, into a tibble
m2 <-  fitted_models %>% tidy(model)
m2
```

```
## # A tibble: 36 x 6
## # Groups:   Subject [18]
##    Subject term        estimate std.error statistic  p.value
##      <int> <chr>          <dbl>     <dbl>     <dbl>    <dbl>
##  1     308 (Intercept)   244.      28.1        8.70 2.39e- 5
##  2     308 Days           21.8      5.26       4.14 3.26e- 3
##  3     309 (Intercept)   205.       5.22      39.3  1.93e-10
##  4     309 Days            2.26     0.977      2.31 4.93e- 2
##  5     310 (Intercept)   203.       7.24      28.1  2.78e- 9
##  6     310 Days            6.11     1.36       4.51 1.98e- 3
##  7     330 (Intercept)   290.      13.1       22.1  1.85e- 8
##  8     330 Days            3.01     2.45       1.23 2.55e- 1
##  9     331 (Intercept)   286.      13.8       20.7  3.05e- 8
## 10     331 Days            5.27     2.58       2.04 7.55e- 2
## # … with 26 more rows
```

2b

```
#making subject a factor:
sleep$Subject <-  as.factor(sleep$Subject)
#plotting Reaction time as a function of days sleep deprieved with each subject having their own regression
line:
ggplot(sleep, aes(Days, Reaction, colour = Subject))+geom_point()+geom_smooth(method = "lm", se=FALSE) + xla
b("Days of Sleep Deprivation") + ylab("Reaction Time in mS")
```



2c

```
#puting the degrees of freedom into the dataframe with the regression coefficients
m2$t.value <- m2$statistic
m2$statistic <- NULL
m2$df_model <-  1
m2$df_residual <-  8
m2
```

```
## # A tibble: 36 x 8
## # Groups:   Subject [18]
##    Subject term        estimate std.error  p.value t.value df_model df_residual
##      <int> <chr>          <dbl>     <dbl>    <dbl>   <dbl>    <dbl>       <dbl>
##  1     308 (Intercept)   244.      28.1   2.39e- 5    8.70        1           8
##  2     308 Days           21.8      5.26  3.26e- 3    4.14        1           8
##  3     309 (Intercept)   205.       5.22  1.93e-10   39.3         1           8
##  4     309 Days            2.26     0.977 4.93e- 2    2.31        1           8
##  5     310 (Intercept)   203.       7.24  2.78e- 9   28.1         1           8
##  6     310 Days            6.11     1.36  1.98e- 3    4.51        1           8
##  7     330 (Intercept)   290.      13.1   1.85e- 8   22.1         1           8
##  8     330 Days            3.01     2.45  2.55e- 1    1.23        1           8
##  9     331 (Intercept)   286.      13.8   3.05e- 8   20.7         1           8
## 10     331 Days            5.27     2.58  7.55e- 2    2.04        1           8
## # … with 26 more rows
```

2d

```
#We want to look at the effects sleep deprivation has on reaction time therefore we only look at the slopes.
The intercepts only show the relationship between the mean reaction time and the reaction time when the part
icipants were 0 days sleep deprived.

m3 <-  filter(m2, term == "Days")

#We only want those that displayed a statistically significant effect from sleep deprivation:
m4 <-  filter(m3, p.value < 0.05)
m4
```

```
## # A tibble: 14 x 8
## # Groups:   Subject [14]
##    Subject term  estimate std.error    p.value t.value df_model df_residual
##      <int> <chr>    <dbl>     <dbl>      <dbl>   <dbl>    <dbl>       <dbl>
##  1     308 Days     21.8      5.26  0.00326       4.14        1           8
##  2     309 Days      2.26     0.977 0.0493        2.31        1           8
##  3     310 Days      6.11     1.36  0.00198       4.51        1           8
##  4     333 Days      9.14     1.37  0.000158      6.66        1           8
##  5     334 Days     12.3      2.26  0.000635      5.41        1           8
##  6     337 Days     19.0      1.80  0.00000553   10.6         1           8
##  7     349 Days     13.5      1.54  0.0000229     8.75        1           8
##  8     350 Days     19.5      2.68  0.0000862     7.27        1           8
##  9     351 Days      6.43     2.51  0.0332        2.57        1           8
## 10     352 Days     13.6      2.81  0.00131       4.83        1           8
## 11     369 Days     11.3      1.74  0.000186      6.51        1           8
## 12     370 Days     18.1      2.66  0.000138      6.80        1           8
## 13     371 Days      9.19     2.76  0.0104        3.33        1           8
## 14     372 Days     11.3      1.24  0.0000172     9.09        1           8
```

14 individuals displayed a statistically significant effect of sleep deprivation.

3a + 3b

We used an one tailed t-test because we were asked wether the slopes are larger than zero. there is strong theoretical background that the slopes are positive (reationtime increases as days sleep deprived increases). Therefore it is justifyable to use a one-tailed t-test even though the risk of comitting a type 1 error is increased. A t-test was used because we want to compare two means (0 and the mean of the reaction time). This test assumes normally distributed data, which is tested below with visual inspection and by using the shapiro wilks test.

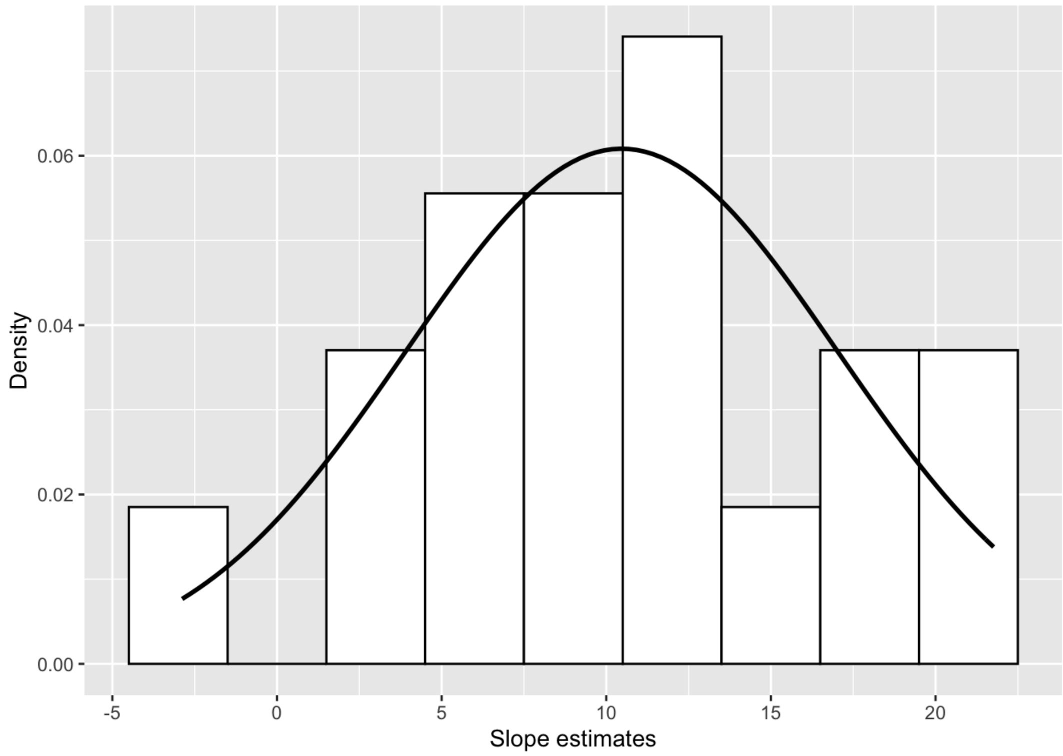
```r
#first accessing the slopes from the previous exercise
est <-  m3$estimate

#Putting them into a dataframe
estimates <-  as.data.frame(est)

#testing whether the slopes are normally distributed:
shapiro.test(est)
```

```
## 
##  Shapiro-Wilk normality test
## 
## data:  est
## W = 0.97881, p-value = 0.9369
```

```r
ggplot(estimates, aes(est)) + geom_histogram(aes(y=..density..), colour = "black", fill = "white", binwidth = 3) + labs(x ="Slope estimates", y = "Density") + stat_function(fun = dnorm, args = list(mean = mean(estimates$est), sd = sd(estimates$est)), colour= "black", size = 1)
```



As can be seen the data looks kind of normally distributed, and the shapiro wilks test is non-significant indicating that the reaction times are not significantly different from a perfect normal distribution.

One can therefore perform the t-test:

```
#Performing the t-test and telling it to perform a one-sided kind
test <-  t.test(est, mu = 0, alternative = "two.sided")
test
```

```
##
##  One Sample t-test
##
## data:  est
## t = 6.7715, df = 17, p-value = 3.264e-06
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   7.205956 13.728615
## sample estimates:
## mean of x
##  10.46729
```

```
#Calculating the effect size r
r <-  sqrt(test$statistic^2/((test$statistic^2)+test$parameter))
r
```

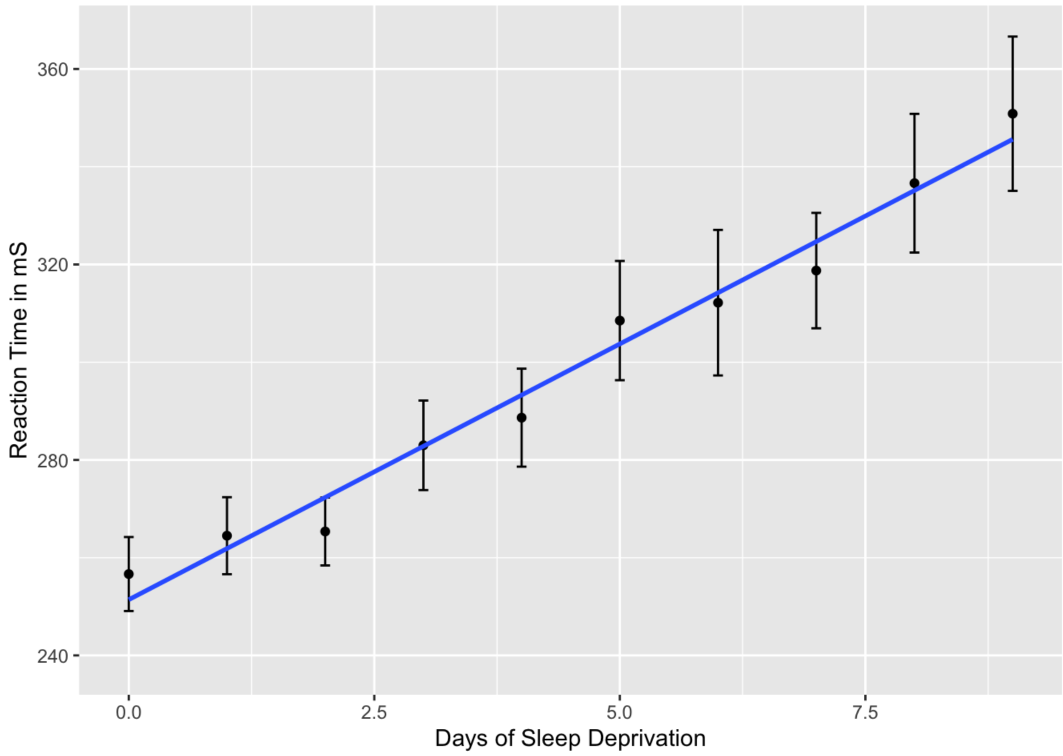
```
##         t
## 0.8541239
```

3c

Our slopes are shown to be statistically significantly different from 0, $t(17)=6.77$ $p < .001$, $r = 0.85$

3d

```
#Making a plot of Reaction time as a function of Days. And adding the mean reaction time with standard error
bars for each day
ggplot(sleep, aes(x = Days, y = Reaction))+geom_point(stat = "summary", fun.y = mean)+ stat_summary(fun.data
= mean_se, geom = "errorbar", color = 'black', width = 0.1)+geom_smooth(method = "lm", alpha=0) + xlab("Day
s of Sleep Deprivation") + ylab("Reaction Time in mS")
```



Voluntary bonus task:

```
#Adding 10% noise.
sleep$Reaction.Noise <-  sleep$Reaction+sleep$Reaction*runif(10, min = -0.05, max = 0.05)
```

```
#Getting statistics from the newly derived data
fitted_models1 <-  sleep %>% group_by(Subject) %>% do(model1 = lm(Reaction.Noise ~ Days, data = .))

#putting the data, into a tibble(dataframe)
m5 = fitted_models1 %>% tidy(model1)

#Filtering out the slopes
m6 = filter(m5, term == "Days")

#We only want those that displayed a statistically significant effect from sleep deprivation:
m7 = filter(m6, p.value < 0.05)

m7
```

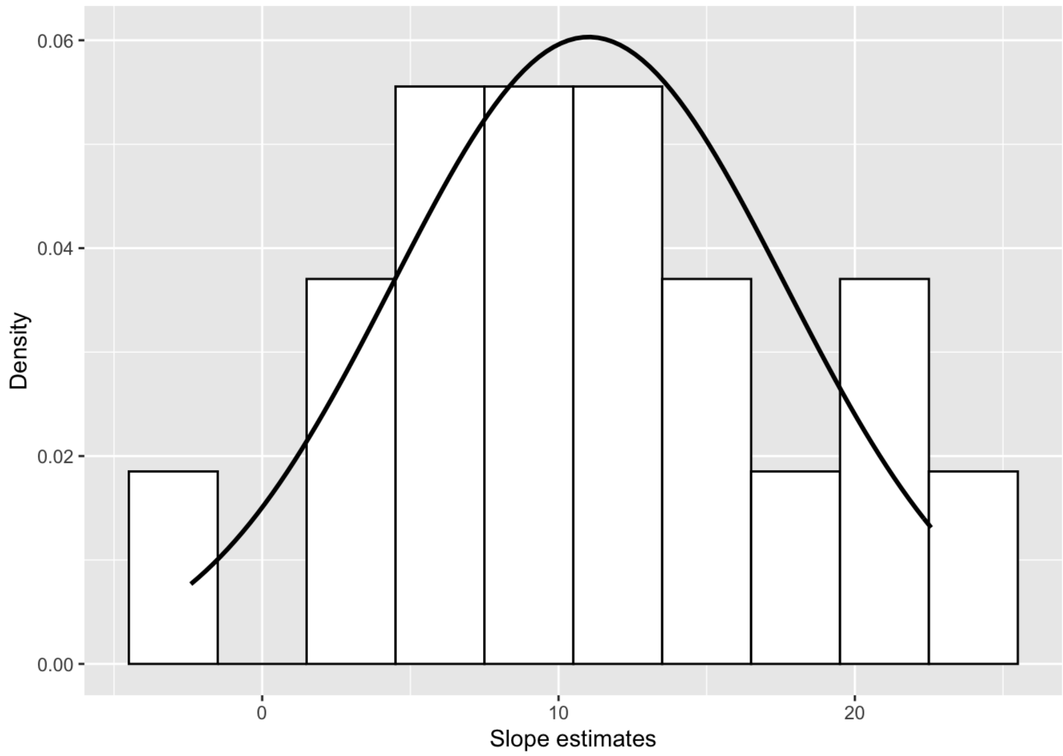
```
## # A tibble: 13 x 6
## # Groups:   Subject [18]
##    Subject term  estimate std.error statistic   p.value
##    <fct>   <chr>    <dbl>     <dbl>     <dbl>     <dbl>
##  1 308     Days     22.6       5.75      3.93 0.00438
##  2 310     Days      6.44      1.43      4.51 0.00197
##  3 333     Days      9.73      2.01      4.83 0.00130
##  4 334     Days     12.9       2.69      4.78 0.00139
##  5 337     Days     19.7       2.13      9.26 0.0000150
##  6 349     Days     14.0       2.08      6.72 0.000149
##  7 350     Days     20.0       3.15      6.35 0.000220
##  8 351     Days      7.11      2.89      2.46 0.0391
##  9 352     Days     14.3       2.77      5.15 0.000870
## 10 369     Days     11.9       1.98      6.02 0.000316
## 11 370     Days     18.6       2.88      6.44 0.000201
## 12 371     Days      9.76      3.15      3.10 0.0146
## 13 372     Days     11.8       1.86      6.36 0.000219
```

```
#Saving the estimates and putting them into a dataframe
est1 = m6$estimate
estimates1 = as.data.frame(est1)

#Testing whether the slopes are normally distributed:
shapiro.test(est1)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  est1
## W = 0.9803, p-value = 0.953
```

```
#Plotting a histogram of the new data
ggplot(estimates1, aes(est1)) + geom_histogram(aes(y=..density..), colour = "black", fill = "white", binwidt
h = 3) + labs(x ="Slope estimates", y = "Density") + stat_function(fun = dnorm, args = list(mean = mean(est
imates1$est1), sd = sd(estimates1$est1)), colour= "black", size = 1)
```



As can be seen the data looks kind of normally distributed, and the shapiro wilks test is non-significant indicating that the reaction times are not significantly different from a perfect normal distribution.

One can therefore perform the t-test with the new data:

```
#Dping a t-test and telling it to perform a one-sided type
test1 = t.test(est1, mu = 0, alternative = "two.sided")
test1
```

```
##
##  One Sample t-test
##
## data:  est1
## t = 7.0679, df = 17, p-value = 1.888e-06
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   7.729651 14.308084
## sample estimates:
## mean of x
##  11.01887
```

```r
#Calculating the effect size
r1 = sqrt(test1$statistic^2/((test1$statistic^2)+test1$parameter))
r1
```

```
##         t
## 0.8637695
```

As can be seen there is only 13 statistically significant slopes after adding 10% white noise.

The results show that adding 10% white noise can have an influence on the individual level, 13 slopes being statistically significant meaning that one person after adding the 10% noise did not seem to have their reaction time influenced by sleep deprivation, eventhough they seemed to be influenced by sleep deprivation without the 10% noise.

On a group level adding the 10% noise had no meaningful effect.