

# Project 0

Jacob Kebe

## Introduction

This analysis concerns a dataset of cortisol and DHEA levels measured over the course of the day using a saliva capturing method. Participants took samples over the course of 3 days, and were instructed to collect samples at waking, before lunch, and 30 and 600 minutes after waking. We have data on Subject IDs, collection date and sample number, cortisol and DHEA (nmol/L), sample time and wake time recorded by participant in booklet, and sample time recorded by an electronic cap in the saliva collection interface. We are interested in testing 3 hypotheses. First we are interested in the agreement between times recorded in the booklet and times recorded by the electronic cap, and whether there is a bias in the booklet times. Statistically, we are interested in whether the slope of the relationship between the times is significantly different than 1, and whether the mean difference is significantly different than 0. We are also interested in what proportion of samples fall within a 15 minute and 30 minute window around the 30 minutes and 600 minutes since waking marks. Finally, we want to study the change in cortisol and DHEA over time. The clinical hypothesis is that cortisol levels spike and peak 30 minutes after waking and decrease for the rest of the

day. We are interested in whether the slope of the relationship between cortisol and time is significantly different before and after the 30 minute mark. # Methods Data was processed in the following ways. Variables for minutes since waking recorded by booklet and cap were calculated using the provided times ('minWakeBook' and 'minWakeMEM', respectively). Data input as an empty string, and cortisol and DHEA outliers (anything greater than or equal to 80 for cortisol or 5.205 for DHEA in nmol/L) were all replaced by NAs in order to properly call functions necessary for the analysis. A second dataset only including observations for collection samples 2 and 4 (30 min and 600 min since waking) was created to answer the second research question.

To analyze the agreement between booklet times and cap times, we conducted linear mixed effect regression with cap times as a function of booklet times, accounting for a random intercept for each subject. Using a mixed effect model accounted for the within subject variation due to each subject having measurements over several days. We tested both the intercept and slope of the model as evidence of bias and agreement (or lack thereof) in the model.

We calculated the proportion of samples whose recorded cap time was within the 15 minute window and 30 minute window of the ideal minutes since waking. We also calculated this proportion for the booklet times.

For the third research question, a linear model of cortisol as a function of minutes since waking was proposed. Violation of linear regression assumptions was assessed since cortisol is a nonnegative right-skewed variable. We found that both residual plots and quantile plots were more indicative of adherence to assumptions when cortisol was log-transformed.

Similarly, we log transformed DHEA. For cortisol, we used a piecewise linear mixed effect regression model with a knot at 30 minutes, allowing for the slope of the relationship between cortisol and time to differ before and after the 30 minute mark, and adjusting for the repeated measures for each subject with a random intercept. We were interested in the coefficients for the time variable before and after 30 minutes. Because the scatterplot of log DHEA as a function of time was indicative of a negative linear relationship, we fit a linear mixed effect model of log DHEA as a function of time with a random intercept for each subject. We considered the coefficient for the time variable to characterize the relationship between DHEA and time. For the linear mixed effect models we tested individual coefficients using the t-test from the lmerTest package using Satterwaite's approximation for degrees of freedom. -appropriateness of approach in answering research question.

```
#load data
cortdat <- read.csv("~/Downloads/Project0_Clean_v2.csv")

#load packages
library(naniar)
library(tidyr)
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
```

```

##      filter, lag

## The following objects are masked from 'package:base':
##      intersect, setdiff, setequal, union

library(olsrr)

## 

## Attaching package: 'olsrr'

## The following object is masked from 'package:datasets':
##      rivers

#fill in reported wake time for all rows.

cortdat <- cortdat %>%
  mutate(Sleep.Diary.reported.wake.time = na_if(Sleep.Diary.reported.wake.time, ""))
  fill(Sleep.Diary.reported.wake.time, .direction="down")

##Calcute time since waking

#1. convert time variables to POSIXct so we can take differences

cortdat$timeBook <- as.POSIXct(cortdat$Booket..Clock.Time, format="%H:%M")
cortdat$timeMEM <- as.POSIXct(cortdat$MEMs..Clock.Time, format="%H:%M")
cortdat$timeWake <- as.POSIXct(cortdat$Sleep.Diary.reported.wake.time, format="%H:%M")

```

```

#2. minutes since waking up for MEM and Booklet

cortdat$minWakeMEM <- as.numeric(difftime(cortdat$timeMEM,cortdat$timeWake, units="mins"))

cortdat$minWakeBook <- as.numeric(difftime(cortdat$timeBook,cortdat$timeWake, units="mins"))

#Booklet time minus MEM time

cortdat$disc <- as.numeric(difftime(cortdat$timeBook,cortdat$timeMEM, units="mins"))

#replace "" and 9999 with NA

cortdat <- cortdat %>%
  mutate(Booket..Clock.Time = na_if(Booket..Clock.Time, ""))
  %>%
  mutate(MEMs..Clock.Time = na_if(MEMs..Clock.Time, ""))

#replace impossible cortisol and DHEA levels with missing to remove outliers.

is.na(cortdat$Cortisol..nmol.L.) <- cortdat$Cortisol..nmol.L. >= 80

is.na(cortdat$DHEA..nmol.L.) <- cortdat$DHEA..nmol.L. >= 5.205

#Q1: Use linear mixed effect regression to regress Booklet time since waking on MEM time

mod1 <- lmerTest::lmer(minWakeMEM ~ minWakeBook + (1|SubjectID), data=cortdat)

summary(mod1)

## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]

## Formula: minWakeMEM ~ minWakeBook + (1 | SubjectID)

```

```

##      Data: cortdat

## 

## REML criterion at convergence: 2786.5

## 

## Scaled residuals:

##      Min       1Q   Median       3Q      Max
## -4.3591 -0.2958 -0.1806 -0.0200  5.9957

## 

## Random effects:

## Groups      Name        Variance Std.Dev.
## SubjectID (Intercept) 44.84     6.696
## Residual           985.03    31.385
## Number of obs: 285, groups: SubjectID, 31

## 

## Fixed effects:

##             Estimate Std. Error      df t value Pr(>|t|)    
## (Intercept) 1.017e+01 2.878e+00 6.362e+01  3.534 0.000769 ***
## minWakeBook 9.910e-01 7.029e-03 2.627e+02 140.978 < 2e-16 ***
## ---      
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## 

## Correlation of Fixed Effects:

##          (Intr)

```

```
## minWakeBook -0.628
```

#Q2: Proportions of adherence

```
cortdat$idealSampleTime <- ifelse(cortdat$Collection.Sample == 1 | cortdat$Collection.Sa
```

#1 if difference between booklet recorded time and ideal time is less than 7.5 minutes

```
cortdat$adherence7.5 <- ifelse(abs(cortdat$minWakeBook-cortdat$idealSampleTime)<=7.5, 1,
```

```
cortdat$adherence15 <- ifelse(abs(cortdat$minWakeBook-cortdat$idealSampleTime)<=15, 1, 0)
```

#Filter only Collection Samples 2,4 (30m since wake, 600m since wake)

```
cortdatq2 <- cortdat[cortdat$Collection.Sample == 2 | cortdat$Collection.Sample == 4,]
```

```
x7.5 <- sum(!is.na(cortdatq2$adherence7.5))
```

```
x15 <- sum(!is.na(cortdatq2$adherence15))
```

```
y <- length(cortdatq2$SubjectID)
```

```
prop7.5 <- prop.test(x7.5,y)
```

```
prop15 <- prop.test(x15,y)
```

#Q3 piecewise linear regression of Cortisol

```
cortdat0<-cortdat[cortdat$minWakeBook<=30,]
```

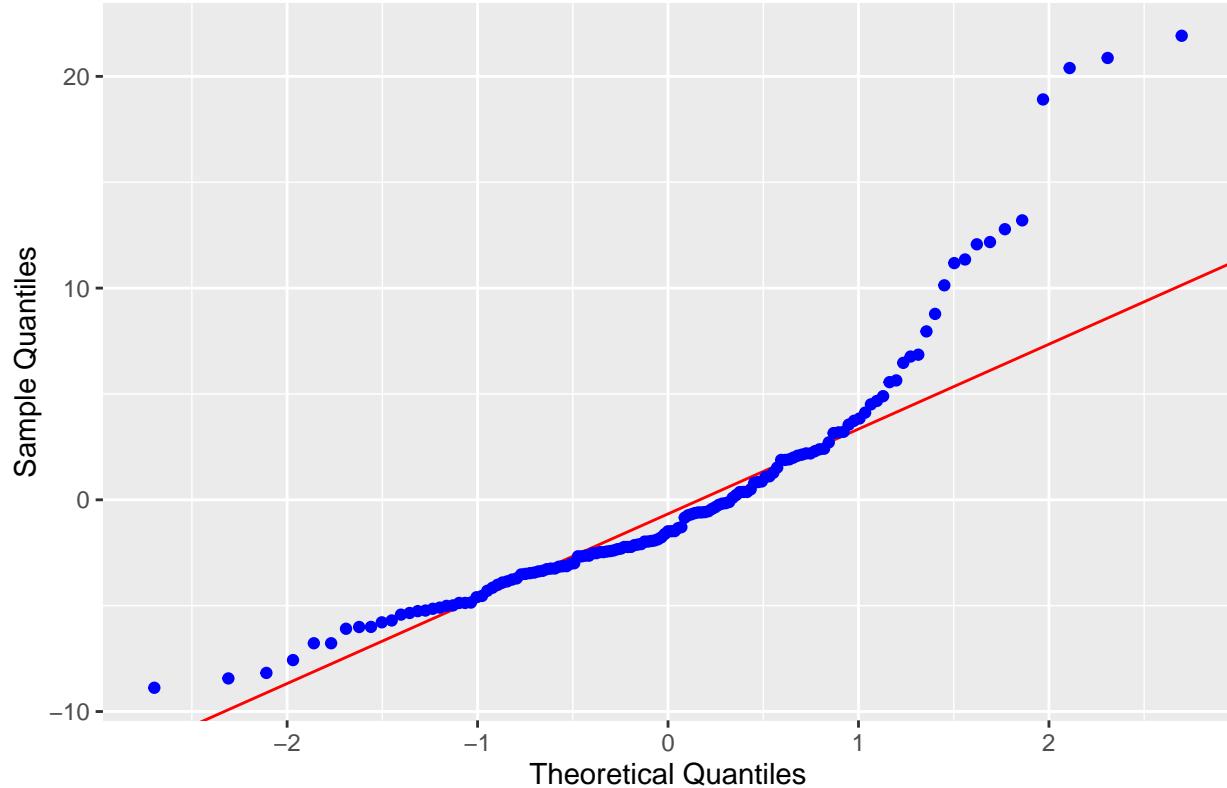
```
cortdat1<-cortdat[cortdat$minWakeBook>30,]
```

```
modknot1 <- lm((Cortisol..nmol.L.) ~ minWakeBook, data=cortdat0)
```

```
modknot2 <- lm((Cortisol..nmol.L.) ~ minWakeBook, data=cortdat1)
```

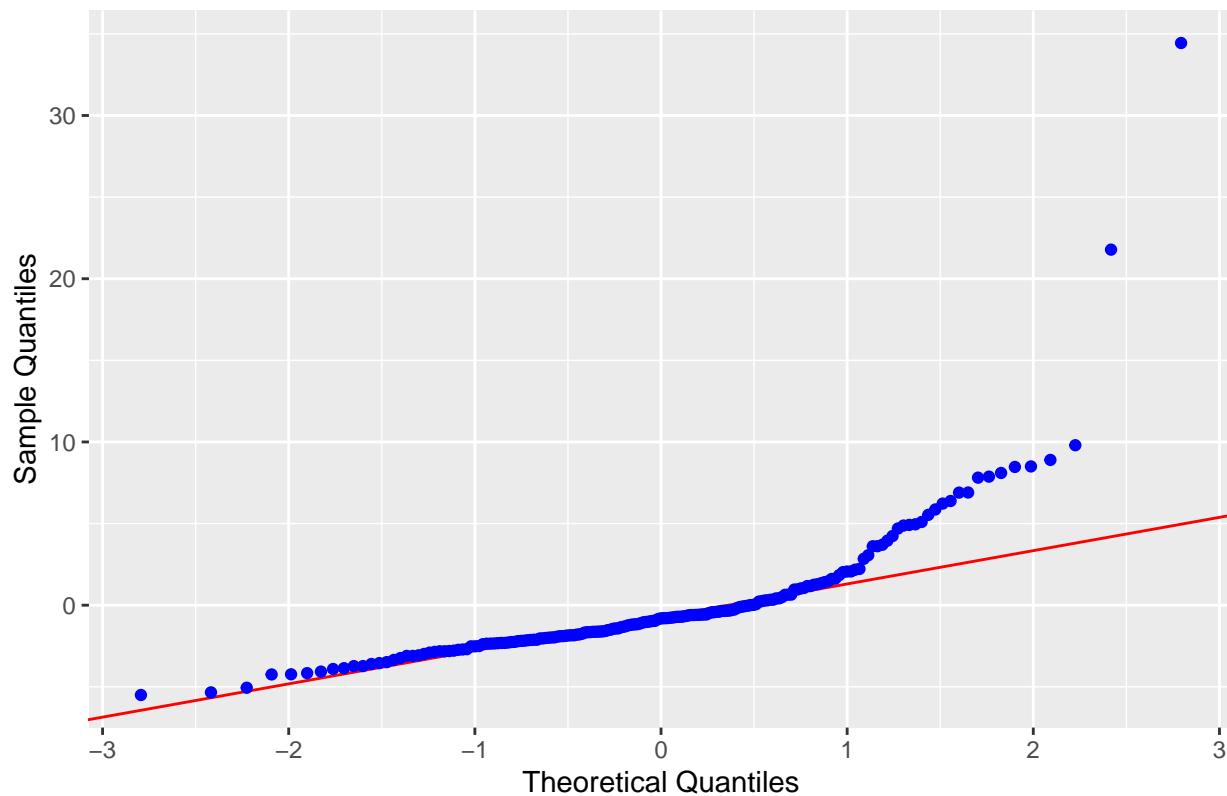
```
ols_plot_resid_qq(modknot1)
```

Normal Q–Q Plot



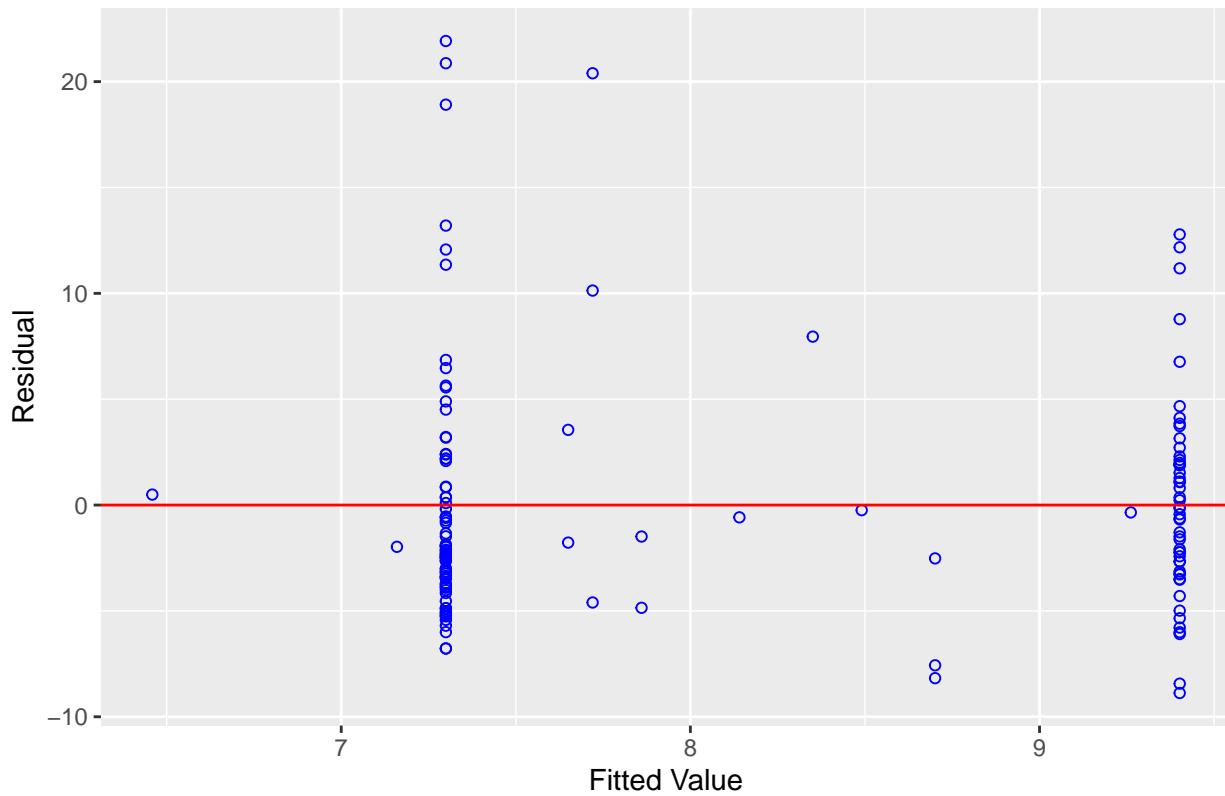
```
ols_plot_resid_qq(modknot2)
```

## Normal Q–Q Plot



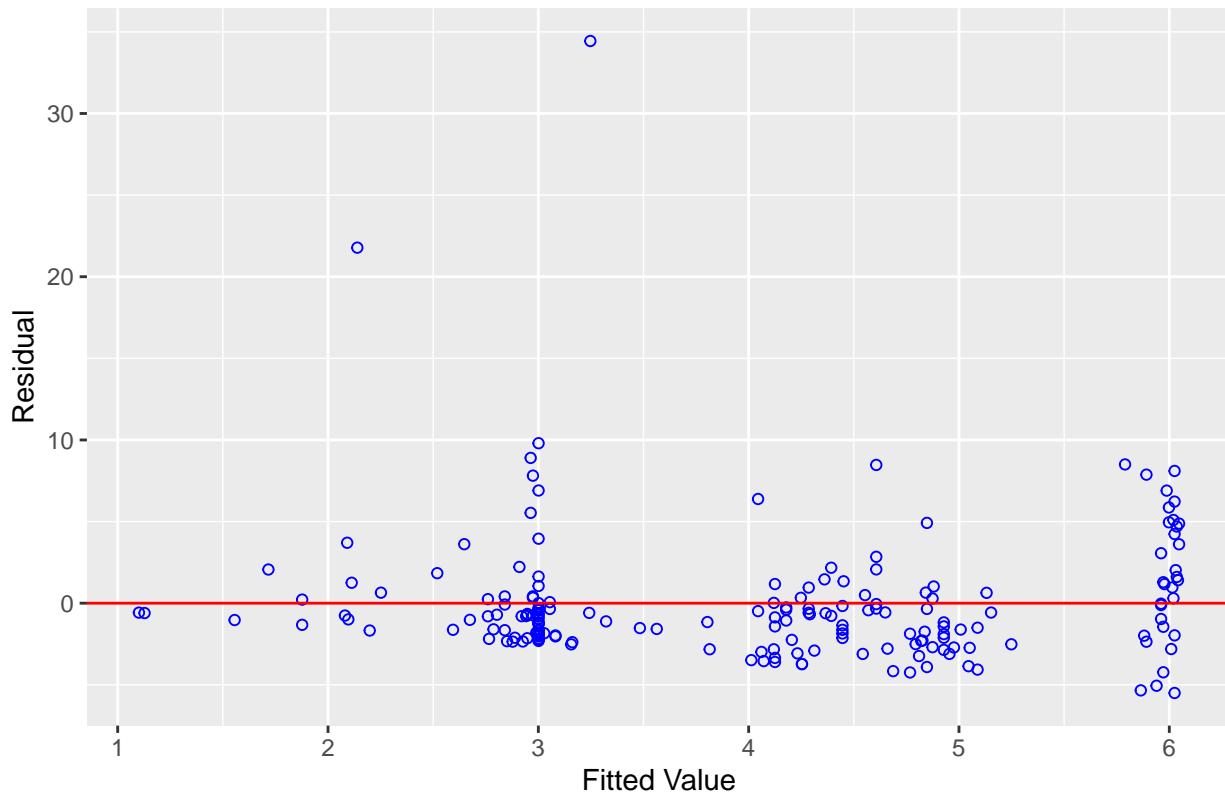
```
ols_plot_resid_fit(modknot1)
```

### Residual vs Fitted Values



```
ols_plot_resid_fit(modknot2)
```

## Residual vs Fitted Values



#residual plots show that using the log of cortisol is better!

```
library(lme4)
```

```
## Loading required package: Matrix

##
## Attaching package: 'Matrix'

## The following objects are masked from 'package:tidyR':
##      expand, pack, unpack
```

```

library(splines)

q3_mod <- lmer(log(Cortisol..nmol.L.)~bs(minWakeBook, knots=c(30), degree=1)+(1|SubjectID),
                 data=cortdat)

summary(q3_mod)

## Linear mixed model fit by REML ['lmerMod']

## Formula: log(Cortisol..nmol.L.) ~ bs(minWakeBook, knots = c(30), degree = 1) +
##           (1 | SubjectID)

## Data: cortdat

## 

## REML criterion at convergence: 805.1

## 

## Scaled residuals:

##      Min       1Q   Median       3Q      Max 
## -3.3140 -0.5084  0.0531  0.5756  3.7601

## 

## Random effects:

## Groups      Name        Variance Std.Dev.
## SubjectID (Intercept) 0.1045    0.3233
## Residual            0.5790    0.7609

## Number of obs: 335, groups: SubjectID, 31

##

```

```

## Fixed effects:

##                                     Estimate Std. Error t value
## (Intercept)                   1.6555    0.1339 12.365
## bs(minWakeBook, knots = c(30), degree = 1)1  0.1683    0.1570  1.072
## bs(minWakeBook, knots = c(30), degree = 1)2 -1.7648    0.1658 -10.645
## 

## Correlation of Fixed Effects:

##          (Intr) b(WB,k=c(30),d=1)1
## b(WB,k=c(30),d=1)1 -0.800
## b(WB,k=c(30),d=1)2 -0.574  0.413

```