



## Project 0

Joseph Froelicher

September 15, 2021

### Contents

<b>INTRODUCTION</b>	<b>2</b>
<b>METHODS</b>	<b>2</b>
Data . . . . .	2
Statistical Analysis . . . . .	3
<b>RESULTS</b>	<b>3</b>
Measurement Agreement . . . . .	3
Adherence . . . . .	4
Cortisol and DHEA vs. Time . . . . .	6
Cortisol . . . . .	6
DHEA . . . . .	7
<b>CONCLUSION</b>	<b>8</b>
Limitations . . . . .	9
<b>APPENDIX</b>	<b>10</b>

# INTRODUCTION

Cortisol and DHEA are two hormones that regulate our stress system. The researchers are interested in the trends of Cortisol and DHEA over time, as well as the effectiveness of a novel collection device for collecting saliva on strips of filter paper (called the Saliva Procurement and Integrated Testing Booklet). Thirty-one health control subjects collected saliva samples four times a day for three days using the novel device. The device also validates collection times with an electronic monitoring cap. The researchers are concerned with three questions, is there agreement between subject's recordings of sampling times and the electronic monitoring caps recorded times, are subjects adhering accurately to the 30 and 600 minute recording times, and lastly what is the change in Cortisol and DHEA over time. More specifically, are the times recorded by the subjects in the booklet different than those recorded by the electronic monitoring device, are what proportion of patients are adhering to the 30 and 600 minute time points with regard to seven and a half minute and 15 minute cutoffs, and lastly is there a difference in cortisol levels and DHEA levels over time and more specifically before the 30 minute time point to after the 30 minute time point.

# METHODS

## Data

Cortisol and DHEA levels were collected at 4 time points for 31 subjects. The time points were recorded using both methods as previously mentioned. The data contained several outliers and missing values due to study limitations. Time points containing missing values were not used in this analysis. Additionally, several subjects were left out of the analysis due to unrealistic values for Cortisol or DHEA as identified by the researchers. See the appendix for the identification number of the subjects removed. Six subjects cortisol or DHEA levels were removed, as well as there were 87 time points with missing data for either the booklet recording time or the electronic monitoring device recording. There were  $n = 279$  observations total used for the regression analysis to address the question of how hormone levels change over time.

	Time Point	1	2	3	4
n		72	66	74	73
Cortisol (nmol/L)		$7.57 \pm 6.48$	$8.95 \pm 4.49$	$3.31 \pm 2.19$	$4.68 \pm 11.45$
DHEA (nmol/L)		$1.73 \pm 1.19$	$1.18 \pm 1.02$	$0.55 \pm 0.49$	$0.51 \pm 0.68$

**Table 1.** Summary statistics for each of the hormone measurements, Cortisol and DHEA.

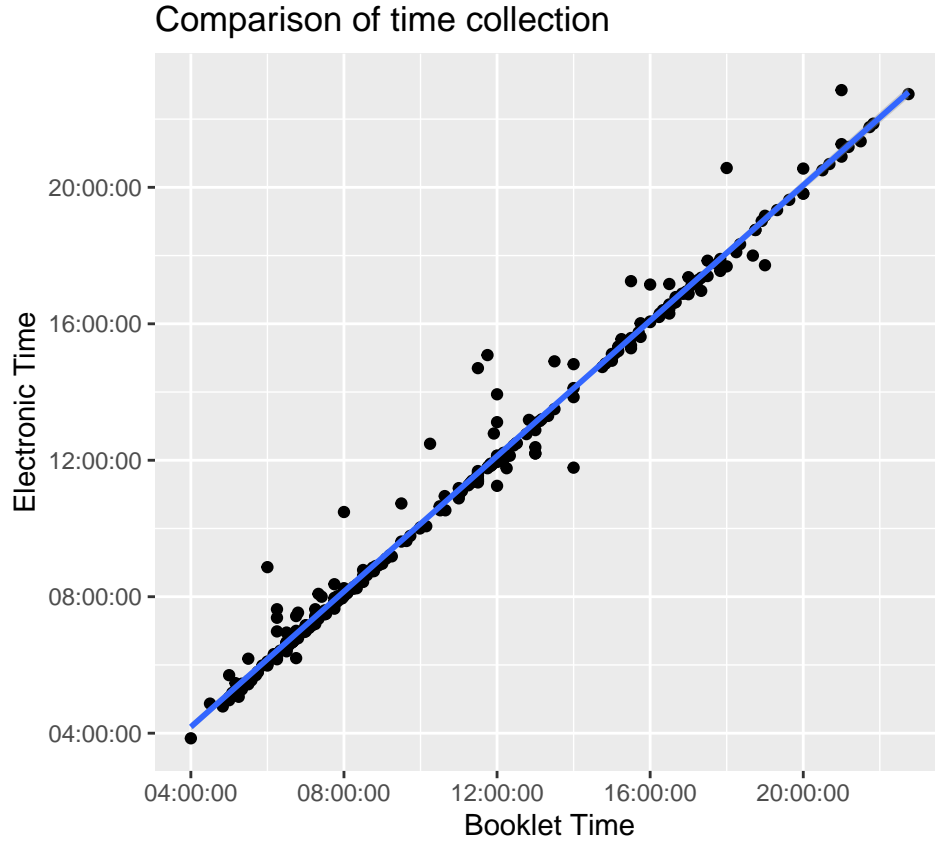
## Statistical Analysis

With regard to the first hypothesis, of whether there is a difference of recorded times between the two methods of recording time of collection, a simple Pearson’s correlation was determined to be appropriate to address the hypothesis of concern. For the second hypothesis, of whether subjects are adhering accurately to the 30 and 600 minute collection times, the data were inspected visually. And lastly, for the third hypothesis, of whether there is a difference between subject Cortisol and DHEA levels before and after the 30 minute time point, a piecewise linear regression with random intercepts to account for correlation was performed to test the difference in slopes from before to after 30 minutes.

## RESULTS

### Measurement Agreement

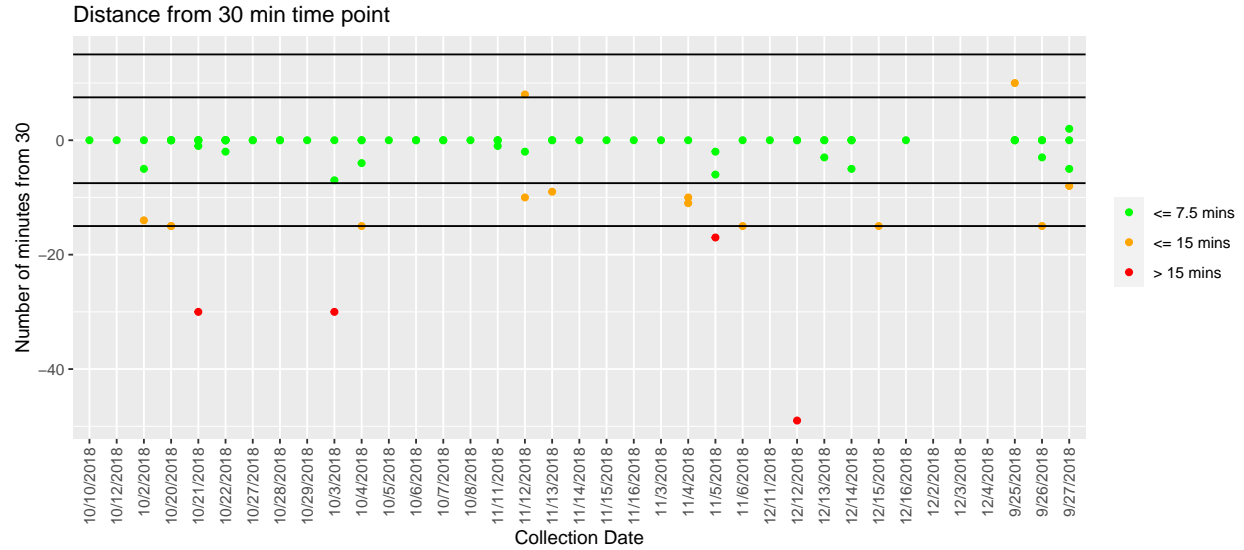
There were two methods of collection for measuring Cortisol and DHEA. In general there was good agreement between the two methods (Pearson’s Correlation = 0.9930776). As seen below in figure 1, there may be concern for the measurement accuracy of the Electronic caps due to the number of points above the  $x = y$  line.



**Figure 1.** Booklet recorded time vs. electronic monitoring cap recorded time for each time point and each subject.

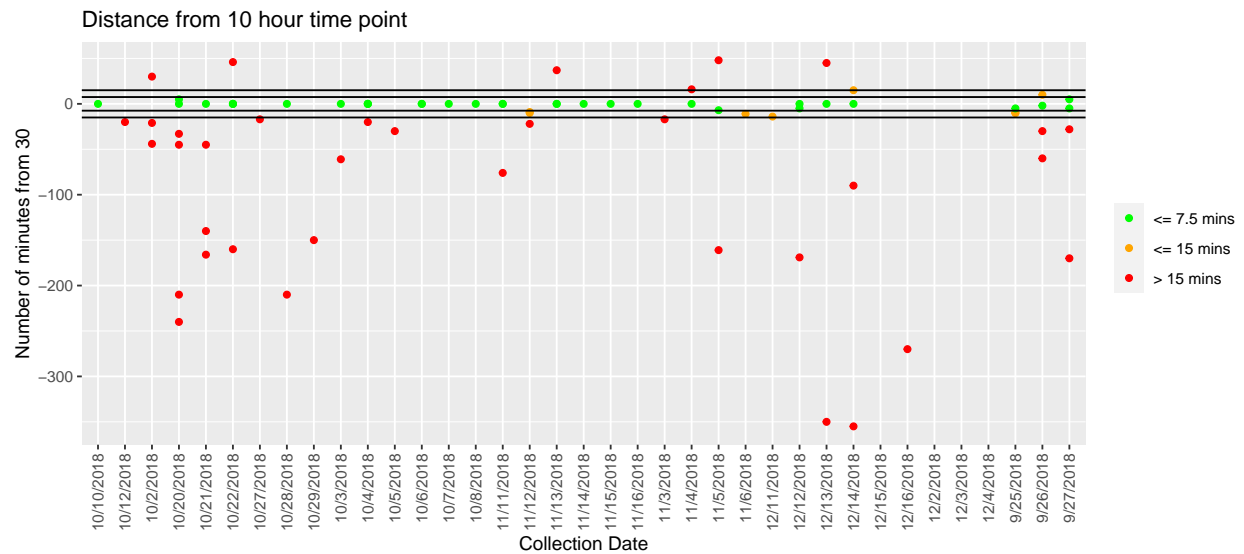
## Adherence

To examine the adherence of subjects to the two time points of interest, 30 and 600 minutes, the investigators decided to visually inspect as well as not the percentage of subjects exceeding  $\pm 7.5$  minutes and  $\pm 15$  minutes from each of the time points (see figures 2 and 3). The booklet time was determined to be more accurate by the researchers due to a malfunction in the electronic monitoring device, thus the booklet recording times were used to determine adherence.



**Figure 2.** Adherence to the 30 minute time point

Adherence to the 30 minute time point was relatively good. After removing missing values, 80.65% of subjects were within 7 minutes and 87.1% of subjects were within 15 minutes according to their booklet recording time.



**Figure 3.** Adherence to the 600 minute time point.

Adherence to the 600 minute time point was relatively good. After removing missing values, 52.69% of subjects were within 7 minutes and 69.89% of subjects were within 15 minutes according to their booklet recording time.

## Cortisol and DHEA vs. Time

The two hormones of interest, Cortisol and DHEA were examined for both linear and piecewise linear trends over time. Each of these models handles correlation between time points by allowing for random intercepts. DHEA and Cortisol were expected to have similar trends, however, our results indicated that while there is no significant difference between the slope of Cortisol before and after the thirty point time point, there was a significant difference between before and after 30 mins for the hormone DHEA (see Tables 2., 3. & 4.). Due to lack of normality in the data, the data were log transformed. Results were back transformed and are interpreted as such.

### Cortisol

	Estimate	SE	Z	P-value	Lower	Upper
Intercept	5.627	0.111	15.583	0.000	4.528	6.993
time	1.004	0.004	0.921	0.357	0.996	1.012
knot	0.994	0.004	-1.385	0.166	0.986	1.002

**Table 2.** Results of piecewise regression for Cortisol level (nmol/L) with a knot tied at 30 minutes.

The piecewise regression did not result in a significant difference between before and after the 30 minute time point for cortisol level (nmol/L).

	Estimate	SE	Z	P-value	Lower	Upper
Intercept	6.1992	0.0867	21.0438	0	5.2305	7.3474
time	0.9981	0.0002	-10.9796	0	0.9978	0.9985

**Table 3.** Results of linear regression for Cortisol level (nmol/L).

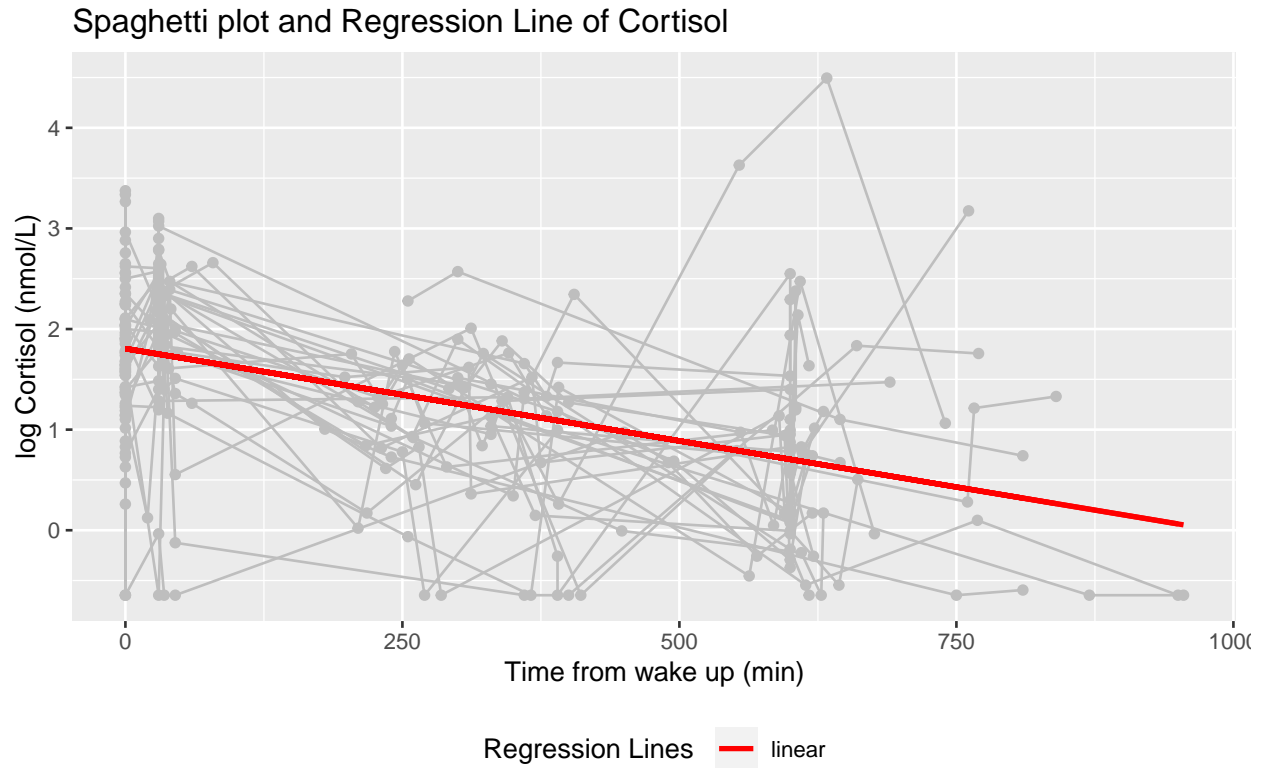
On average, Cortisol levels are decreasing at a rate of 0.998 (0.9979, 0.9981) times per minute before thirty minutes, and a rate of 1.018 (1.011, 1.025) times higher after 30 mins than they are before.

## DHEA

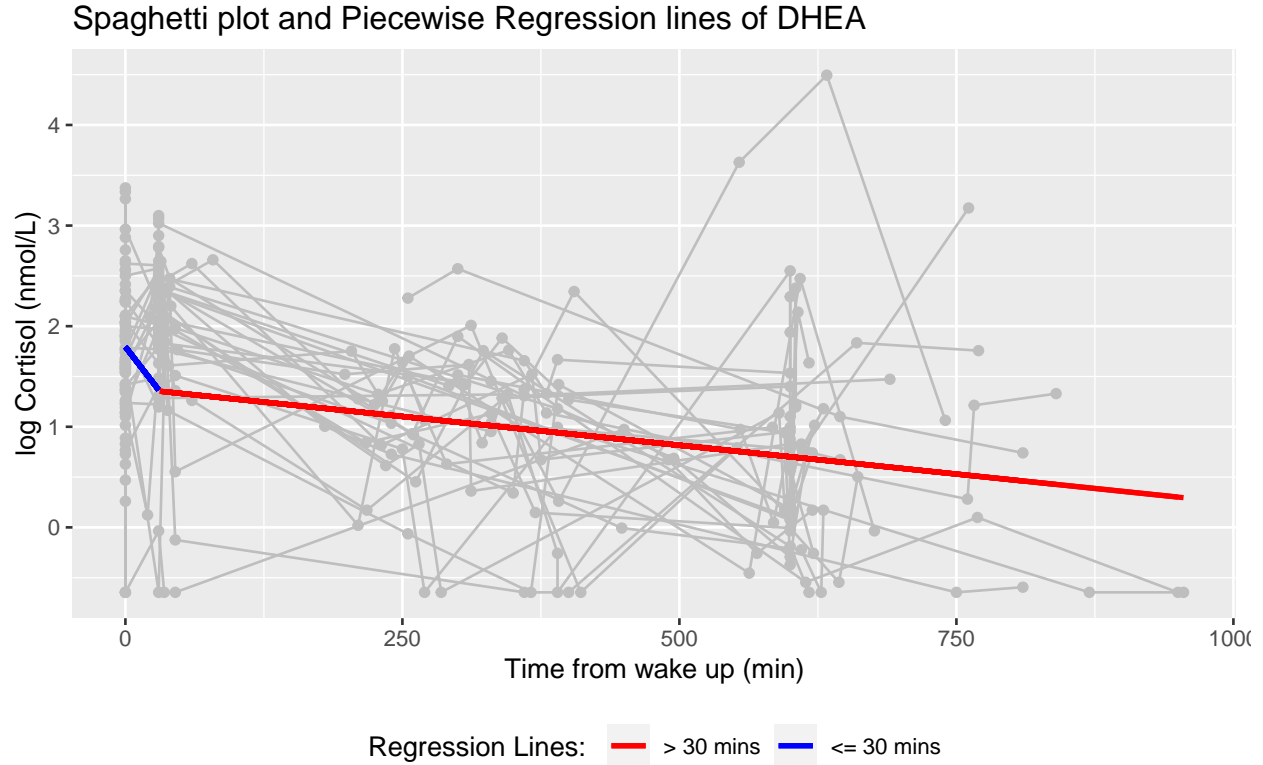
	Estimate	SE	Z	P-value	Lower	Upper
Intercept	1.355	0.122	2.495	0.013	1.067	1.720
time	0.981	0.003	-5.890	0.000	0.975	0.987
knot	1.018	0.003	5.326	0.000	1.011	1.025

**Table 4.** Results of piecewise regression for DHEA level (nmol/L) with a knot tied at 30 minutes.

On average, DHEA levels are decreasing at a rate of 0.981 (0.975, 0.987) times per minute before thirty minutes, and a rate of 1.018 (1.011, 1.025) times higher after 30 mins than they are before.



**Figure 4.** Plot of linear regression for Cortisol level (nmol/L).



**Figure 4.** Plot of piecewise regression for DHEA level (nmol/L) with a knot tied at 30 minutes.

## CONCLUSION

In the scope of our three scientific questions, is there agreement between subject's recordings of sampling times and the electronic monitoring caps recorded times, are subjects adhering accurately to the 30 and 600 minute recording times, and lastly what is the change in Cortisol and DHEA over time, an analysis was performed and yielded the aforementioned results. There seemed to be generally good agreement between the electronic monitoring device and the booklet recordings. However, the researchers should be skeptical of the subjects adherence to the 30 and 600 minute time point protocols. And lastly there was a significant change in the rate of decline of DHEA level in subjects from before 30 minutes to after. The results of the regression indicate an interesting finding, and warrant further analyses of the relationship between the two hormones, cortisol and DHEA, and time.



## Limitations

The study involves several limitations. The first of which is a traditional problem when collecting repeated measures, there is missing data for certain time points, and thus the analysis was performed on unbalanced data. Secondly, as was addressed in the second research question, the researchers should be skeptical of adherence to the recording time points. Lastly, while the researchers were interested in the changes in cortisol and DHEA over time, the question addressed was the difference between before the 30 min time point to after the 30 minute time point. In a future analysis, considering that hormone levels are continuous over time, it may be of interest to model the data using a continuous function, rather than piece wise linear functions.

## APPENDIX

<https://github.com/BIOS6624-UCD/bios6624-JoeFroelicher>

```
# (C) Joseph H. Froelicher "https://scholar.google.com/citations?user=2uQPY3AAAAAJ&hl=en&oi=ao"
library(dplyr)
library(finalfit)
library(ggplot2)
library(janitor)
library(jstable)
library(kableExtra)
library(kimisc)
library(lme4)
library(lubridate)
library(sjPlot)
library(survival)
library(tableone)
library(tidyr)
library(tidyverse)

data = read_csv("../data_raw/Project0_Clean_v2.csv") %>% clean_names()

data_table0 = data[!is.na(data$booket_clock_time) & !is.na(data$me_ms_clock_time) & data$subject_id != 0,]
colnames(data_table0)[3] = "Collection Time Point"
colnames(data_table0)[13] = "Cortisol (nmol/L)"
colnames(data_table0)[14] = "DHEA (nmol/L)"

table0 = CreateTableOne2(vars=c("Cortisol (nmol/L)", "DHEA (nmol/L)"), strata = "Collection Time Point", data = data_table0)
colnames(table0)[1] = "Time Point"
kable(table0[,1:5], booktabs = TRUE)

### Question 1 ###
cor = cor(hms.to.seconds(data$booket_clock_time), hms.to.seconds(data$me_ms_clock_time), use = "pairwise.complete.obs")

figure1 = ggplot(data, aes(x = booket_clock_time, y = me_ms_clock_time)) +
  geom_point() +
  geom_smooth(method=lm) +
  labs(x = "Booklet Time", y = "Electronic Time", title = "Comparison of time collection")

figure1

### Question 2 ###
# get the two data sets that contain second and fourth time points for looking at distance from 30 mins
data_time2 = data[data$collection_sample == 2,]
data_time2$distance = (30 - data_time2$booklet_sample_interval_decimal_time_mins)
data_time2$distance_cat = vector(mode = "integer", length = length(data_time2$subject_id))

for (i in 1:dim(data_time2)[1]) {
  if (!is.na(data_time2$distance[i]) & abs(data_time2$distance[i]) <= 7.5) {
    data_time2$distance_cat[i] = 0
  } else if (!is.na(data_time2$distance[i]) & abs(data_time2$distance[i]) <= 15) {
    data_time2$distance_cat[i] = 1
  }
}
```

```

    } else if (!is.na(data_time2$distance[i])) {
      data_time2$distance_cat[i] = 2
    }
  }

figure2 = ggplot(data_time2, aes(x = collection_date, y = distance, color = factor(distance_cat))) +
  geom_point() +
  ggtitle("Distance from 30 min time point") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1), legend.title = element_blank()) +
  geom_hline(yintercept = 7.5) +
  geom_hline(yintercept = 15) +
  geom_hline(yintercept = -15) +
  geom_hline(yintercept = -7.5) +
  ylab("Number of minutes from 30") +
  xlab("Collection Date") +
  scale_color_manual(labels = c("<= 7.5 mins", "<= 15 mins", "> 15 mins"), values = c("green", "orange"))

figure2

percent7_30 = dim(data_time2[data_time2$distance_cat == 0,])[1] / dim(data_time2)[1]
percent15_30 = (dim(data_time2[data_time2$distance == 1,])[1] + dim(data_time2[data_time2$distance_cat == 1,])[1]) / dim(data_time2)[1]

data_time4 = data[data$collection_sample == 4,]
data_time4$distance = (600 - data_time4$booklet_sample_interval_decimal_time_mins)
data_time4$distance_cat = vector(mode = "integer", length = length(data_time4$subject_id))

for (i in 1:dim(data_time2)[1]) {
  if (!is.na(data_time4$distance[i]) & abs(data_time4$distance[i]) <= 7.5) {
    data_time2$distance_cat[i] = 0

  } else if (!is.na(data_time4$distance[i]) & abs(data_time4$distance[i]) <= 15) {
    data_time4$distance_cat[i] = 1

  } else if (!is.na(data_time4$distance[i])) {
    data_time4$distance_cat[i] = 2
  }
}

figure3 = ggplot(data_time4, aes(x = collection_date, y = distance, color = factor(distance_cat))) +
  geom_point() +
  ggtitle("Distance from 10 hour time point") +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust=1), legend.title = element_blank()) +
  geom_hline(yintercept = 7.5) +
  geom_hline(yintercept = 15) +
  geom_hline(yintercept = -15) +
  geom_hline(yintercept = -7.5) +
  ylab("Number of minutes from 30") +
  xlab("Collection Date") +
  scale_color_manual(labels = c("<= 7.5 mins", "<= 15 mins", "> 15 mins"), values = c("green", "orange"))

figure3

percent7_600 = dim(data_time4[data_time4$distance_cat == 0,])[1] / dim(data_time4)[1]

```

```

percent15_600 = (dim(data_time4[data_time4$distance == 1,])[1] + dim(data_time4[data_time4$distance_cat
### Question 3 ###
# remove outliers and handle missing values labeled as 9999
data[data$cortisol_ug_dl == 9999,]$cortisol_ug_dl = NA
data[data$subject_id == c(77, 101, 241, 242, 246, 249),]$dhea_nmol_l = NA

# question 3 models
cort_fit1 = lmer(log(cortisol_nmol_l) ~ (1 | subject_id) + (booklet_sample_interval_decimal_time_mins),
summary(cort_fit1)
AIC(cort_fit1)

cort_fit2 = lmer(log(cortisol_nmol_l) ~ (1 | subject_id) + (booklet_sample_interval_decimal_time_mins) +
summary(cort_fit2)
AIC(cort_fit2)

dhea_fit1 = lmer(log(dhea_nmol_l) ~ (1 | subject_id) + (booklet_sample_interval_decimal_time_mins), data
summary(dhea_fit1)
AIC(dhea_fit1)

dhea_fit2 = lmer(log(dhea_nmol_l) ~ (1 | subject_id) + (booklet_sample_interval_decimal_time_mins) + I(
summary(dhea_fit2)
AIC(dhea_fit2)

Vcov = vcov(cort_fit2, useScale = FALSE)
betas = fixef(cort_fit2)
se = sqrt(diag(Vcov))
zval = betas / se
pval = 2 * pnorm(abs(zval), lower.tail = FALSE)
lower = exp(betas + qnorm(1-0.975)*se)
upper = exp(betas - qnorm(1-0.975)*se)

table1 = round(cbind(exp(betas), se, zval, pval, lower, upper), 3)
colnames(table1) = c("Estimate", "SE", "Z", "P-value", "Lower", "Upper")
rownames(table1) = c("Intercept", "time", "knot")
kable(table1, booktabs = T)

Vcov = vcov(cort_fit1, useScale = FALSE)
betas = fixef(cort_fit1)
se = sqrt(diag(Vcov))
zval = betas / se
pval = 2 * pnorm(abs(zval), lower.tail = FALSE)
lower = exp(betas + qnorm(1-0.975)*se)
upper = exp(betas - qnorm(1-0.975)*se)

table2 = round(cbind(exp(betas), se, zval, pval, lower, upper), 4)
colnames(table2) = c("Estimate", "SE", "Z", "P-value", "Lower", "Upper")
rownames(table2) = c("Intercept", "time")
kable(table2, booktabs = T)

Vcov = vcov(dhea_fit2, useScale = FALSE)
betas = fixef(dhea_fit2)
se = sqrt(diag(Vcov))

```

```

zval = betas / se
pval = 2 * pnorm(abs(zval), lower.tail = FALSE)
lower = exp(betas + qnorm(1-0.975)*se)
upper = exp(betas - qnorm(1-0.975)*se)

table3 = round(cbind(exp(betas), se, zval, pval, lower, upper), 3)
colnames(table3) = c("Estimate", "SE", "Z", "P-value", "Lower", "Upper")
rownames(table3) = c("Intercept", "time", "knot")
kable(table3, booktabs = T)

# question 3 plots
cort_pred = predict(lm(log(cortisol_nmol_l) ~ (booklet_sample_interval_decimal_time_mins), data = data[

figure4 = ggplot(data, aes(x = booklet_sample_interval_decimal_time_mins, y = log(cortisol_nmol_l), gro
  geom_line(col = "gray") +
  geom_point(col = "gray") +
  ylab("log Cortisol (nmol/L)") +
  xlab("Time from wake up (min)") +
  ggtitle("Spaghetti plot and Regression Line of Cortisol") +
  geom_line(
    data = data[!is.na(data$booklet_sample_interval_decimal_time_mins) & !is.na(data$cortisol_nmol_l)],
    aes(y = cort_pred, x = booklet_sample_interval_decimal_time_mins, color = "red"),
    size = 1
  ) +
  scale_color_manual(name = "Regression Lines ", values = c("red"), labels = c("linear")) +
  theme(legend.position = "bottom")

figure4

# dheia plot with regression lines
dhea_pred1 = predict(lm(log(dhea_nmol_l) ~ (booklet_sample_interval_decimal_time_mins), data = data[data
dhea_pred2 = predict(lm(log(dhea_nmol_l) ~ (booklet_sample_interval_decimal_time_mins), data = data[data
dhea_pred2 = dhea_pred2 + 1/3
dhea_pred2 = dhea_pred2 + 1.45
dhea_pred1 = dhea_pred1 + 1.5

figure5 = ggplot(data, aes(x = booklet_sample_interval_decimal_time_mins, y = log(cortisol_nmol_l), gro
  geom_line(col = "gray") +
  geom_point(col = "gray") +
  ylab("log Cortisol (nmol/L)") +
  xlab("Time from wake up (min)") +
  ggtitle("Spaghetti plot and Piecewise Regression lines of DHEA") +
  geom_line(
    data = data[data$booklet_sample_interval_decimal_time_mins <= 30 & !is.na(data$booklet_sample_inter
    aes(y = dhea_pred1, x = booklet_sample_interval_decimal_time_mins, color = "red"),
    size = 1
  ) +
  geom_line(
    data = data[data$booklet_sample_interval_decimal_time_mins > 30 & !is.na(data$booklet_sample_inter
    aes(y = dhea_pred2, x = booklet_sample_interval_decimal_time_mins, color = "blue"),
    size = 1
  ) +
  scale_color_manual(name = "Regression Lines: ", values = c("red", "blue"), labels = c("> 30 mins", "<

```

```
theme(legend.position = "bottom")  
figure5
```