# Body temperature investigation (Group 0.0)

B007007

Block 3 (semester 2, week 2)

## 1 Introduction

The data available at https://uoepsy.github.io/data/BodyTemperatures.csv comprise measurements of the body temperature (`BodyTemp`, in Celsius) and pulse rate (`Pulse`) for a random sample of 50 healthy individuals. We are interested in estimating the mean body temperature for all healthy humans and testing whether this is different from the commonly thought value of 37 °C. As such, for the purpose of this investigation, we will only focus on the variable `BodyTemp`. While the variable `Pulse` includes 5 missing values, the corresponding rows do have values for the `BodyTemp` variable and as such are not excluded from the analysis.

## 2 Analysis

Figure 1(a) shows that the body temperatures in the sample of 50 healthy individuals roughly follow a bell-shaped distribution, with most values between 36.5 and 37.5 °C and fewer in the tails of the distribution. No values are lower than 35.75 or larger than 38.22 as shown by Table 1. The boxplot in Figure 1(b) highlights an outlying point. However, upon closer inspection, that value is a plausible body temperature and as such is not excluded from the analysis.

The average body temperature in the sample is $M = 36.81$ °C ($SE = 0.06$), see Table 1. Furthermore, we are 95% confident that the average body temperature for a healthy individual is between 36.69 °C and 36.93 °C.

Throughout the report, a significance level of $\alpha = .05$ is used. To investigate whether the average body temperature of all healthy individuals differs from the commonly thought value of 37 °C, we performed a two-sided one sample t-test. The sample data provide very strong evidence against the null hypothesis and in favour of the alternative one that the mean body temperature significantly differs from the hypothesised value of 37 °C, $t(49) = -3.14, p = .003$, two-sided. As such, we reject the null hypothesis and decide that the average body temperature has changed. In particular, from the confidence interval, we can say that this is between 0.07 and 0.31 lower than the hypothesised value.

## 3 Discussion

To be added in the next weeks.

Table 1: Descriptive statistics of body temperatures (°C)

| n | M | SD | Min | Max |
|---|---|---|---|---|
| 50 | 36.81 | 0.43 | 35.78 | 38.22 |

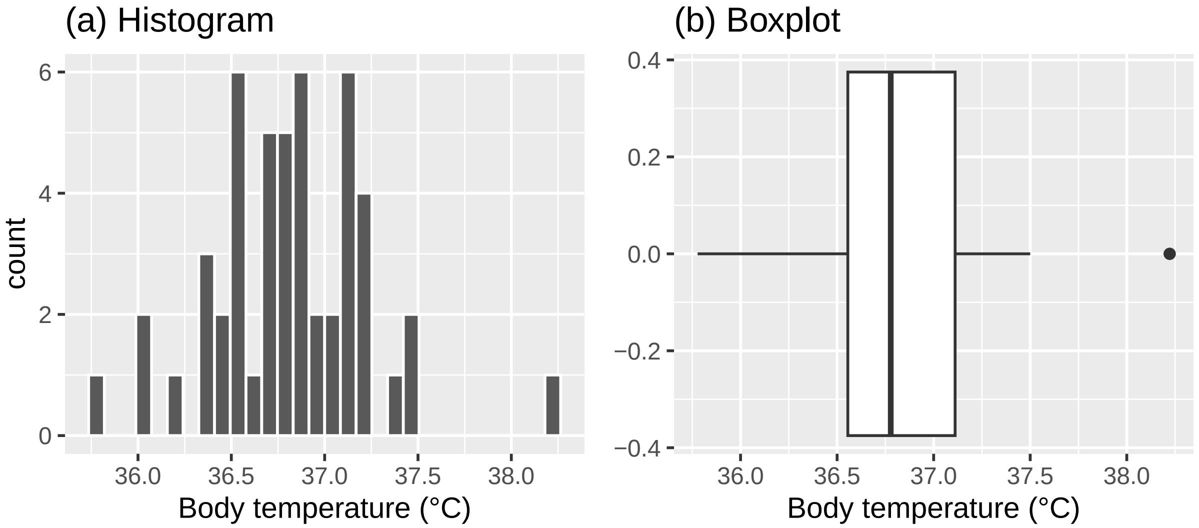

Figure 1: Distribution of body temperatures (°C)

# 4 Appendix B: R code

```
knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE)
# Packages
library(tidyverse)
library(patchwork)
library(kableExtra)


# Week 1 code
temp_data <- read_csv("https://uoepsy.github.io/data/BodyTemperatures.csv")

head(temp_data)      # top six rows of the data: head(DATA, n = 6) by default
glimpse(temp_data)   # or str()
summary(temp_data)   # quick summary of the variables in the data
dim(temp_data)       # number of rows and columns in the data

# the two computations below only do the computation and print the result

temp_data |>
    drop_na()
    # drops all rows with NAs in any of the columns, even columns which
    # we don't use. Would throw away information in this case.

temp_data |>
    drop_na(BodyTemp)
    # drops rows where there are NAs only the column of interest (BodyTemp)

# Because there are no NAs in the BodyTemp variable, the data won't change in
# this specific example - it still has 50 rows.
# If there were any NAs in the variable of interest (BodyTemp here), this would
```

```
# be the correct way to remove the rows with NAs in the variable of interest.

# to store the result of a computation, use the assignment operator <-

temp_data <- temp_data |>
  drop_na(BodyTemp)

# there's two variables in the data, but we are only analysing BodyTemp, so
# let's subset the data and only select that column

temp_data <- temp_data |>
  select(BodyTemp)

dim(temp_data)  # 50 rows and 1 column only now


# Distribution of BodyTemp
plt.h <- ggplot(temp_data, aes(x = BodyTemp)) +
  geom_histogram(colour = 'white') +
  labs(x = "Body temperature (°C)",
       title = "(a) Histogram")
plt.h

plt.b <- ggplot(temp_data, aes(x = BodyTemp)) +
  geom_boxplot() +
  labs(x = "Body temperature (°C)",
       title = "(b) Boxplot")
plt.b

plt.h | plt.b


# Option 1: with a descriptives table
stats <- temp_data |>
  summarise(
    n = n(),
    M = mean(BodyTemp),
    SD = sd(BodyTemp),
    Min = min(BodyTemp),
    Max = max(BodyTemp)
  )
stats

tstar <- qt(c(0.025, 0.975), df = stats$n - 1)
tstar

stats$M + tstar * (stats$SD / sqrt(stats$n))


# Option 2: creating each variable (not using a descriptives table)
xbar <- mean(temp_data$BodyTemp)
n <- nrow(temp_data)
s <- sd(temp_data$BodyTemp)
```

```
se <- s / sqrt(n)
tstar <- qt(c(0.025, 0.975), df = n - 1)
xbar + tstar * se


# Week 2 code

# Step 1. Specify the null and alternative hypothesis
# H0 : mu = 37
# H1 : mu not = 37
mu0 <- 37

# Step 2. Compute the observed value of the t-statistic
# t_obs = (xbar - mu0) / SE, where SE = s / sqrt(n)

t_obs <- (stats$M - mu0) / (stats$SD / sqrt(stats$n))
t_obs

# Step 3. Compute the p-value
pvalue <- 2 * pt(abs(t_obs), df = stats$n - 1, lower.tail = FALSE)
pvalue

# Step 4. Make a decision by comparing the p-value to alpha (significance level)
# We use alpha = .05
# - Reject H0 if pvalue <= alpha
# - Do not reject H0 if pvalue > alpha

pvalue <= 0.05

# Step 5. Writing up the result in context
plt.h | plt.b
kbl(stats, booktabs = TRUE, digits = 2,
    caption = "Descriptive statistics of body temperatures (°C)")
```