Body temperature investigation

DAPR1 staff

1 Introduction

The data available at https://uoepsy.github.io/data/BodyTemp.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`.

2 Analysis

Throughout the entire report, a significance level of 5% is used.

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 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.

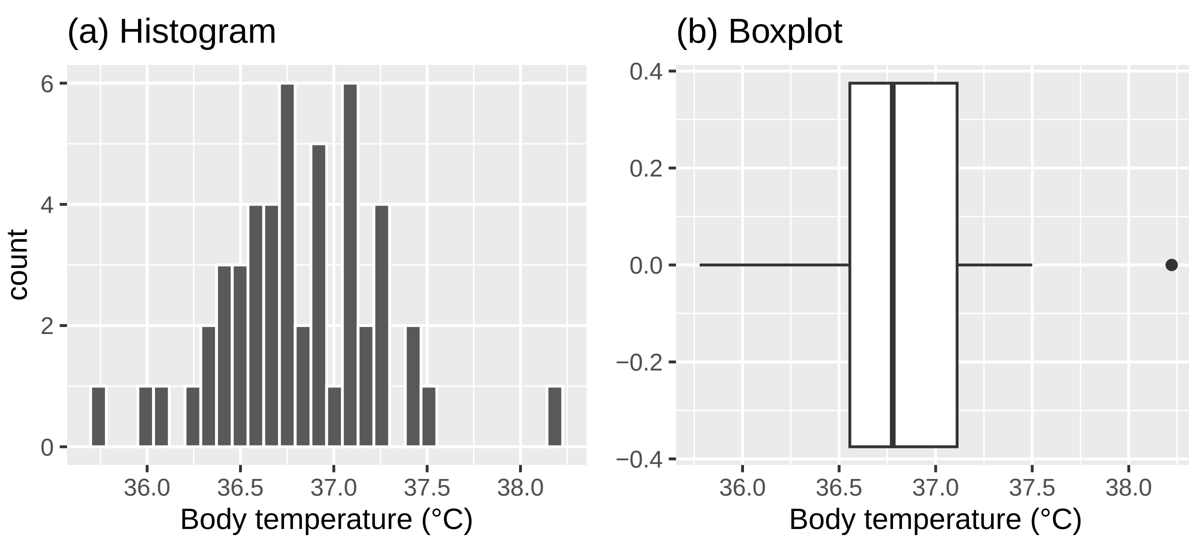

Figure 1: Distribution of body temperatures (°C)

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

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

n	M	SD	Min	Max
50	36.81	0.43	35.78	38.22

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 commonly thought value of 37 °C. However, a Cohen's $D = -0.44$ indicates a small to medium effect size and that the result may not be of practical importance.

The sample data did not raise any concerns of violations of the t-test assumptions of independence and normality of the sampling distribution of the mean. First, measurements were taken on a random sample from the population of all healthy individuals, hence independence was met. Second, the histogram in Figure 1(a) displays a roughly bell shaped-distribution and the QQ-plot in Figure 2 shows agreement between the sample and theoretical normal quantiles, as evidenced by the points closely following the perfect agreement line. We also performed a Shapiro-Wilk test against the null hypothesis that, in the population, the data follow a normal distribution. The sample data did not provide sufficient evidence to reject the null hypothesis ($W = 0.97$, $p = .31$) and, as such, we conclude that the population data are normal. From this, we can also assume that the sampling distribution of the mean follows a normal distribution.

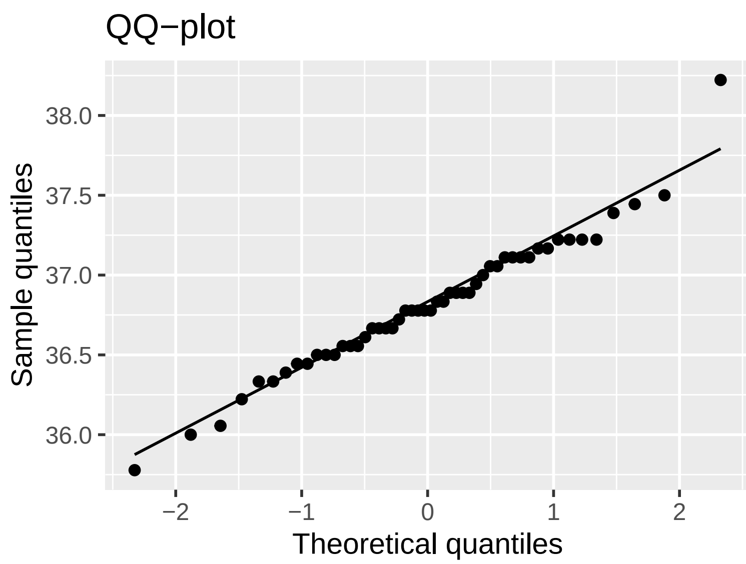

Figure 2: QQ-plot of body temperatures sample (°C)

3 Discussion

This report estimated the mean body temperature of all healthy humans and tested whether, at the 5% significance level, that value is different from the long thought 37 °C.

The sample data indicated that the population mean body temperature significantly differs from 37 °C and, specifically, that it may be lower than that value. The population mean value is somewhere between 36.69 °C and 36.93 °C, with 95% confidence. However, the departure from the hypothesised value was not found to be large enough to be of any practical importance for decision makers.

4 Appendix: R code

```
knitr::opts_chunk$set(echo = FALSE, message = FALSE, warning = FALSE)
##### week 1 code
library(tidyverse)
library(patchwork)
library(kableExtra)

temp_data <- read_csv("https://uoepsy.github.io/data/BodyTemp.csv")

temp_data <- temp_data %>%
  select(BodyTemp)

dim(temp_data)
head(temp_data)
glimpse(temp_data)   # str()
summary(temp_data)

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

plt.d <- ggplot(temp_data, aes(x = BodyTemp)) +
  geom_density() +
  labs(x="Body temperature (°C)",
       title = "Density")
plt.d

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$n - 1)
stats$M + tstar * (stats$SD / sqrt(stats$n))
```

```r
# Option 2: creating each variable
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 null and alternative hypotheses
# H0 : mu = 37
# H1 : mu not equal to 37

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

t_obs <- (stats$M - 37) / (stats$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)
# - Reject H0 if pvalue <= alpha
# - Do not reject H0 if pvalue > alpha

pvalue <= 0.05

# Step 5. Writing up


##### week 3 code

# Step 1. Specify null and alternative hypothesis

# H0 : mu = 37
# H1 : mu not equal to 37

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

t_obs <- (stats$M - 37) / (stats$SD / sqrt(stats$n))
t_obs

# Step 3. Computing the critical values for the appropriate null distribution
# df = n - 1 = 49

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

```
# Step 4. Make a decision by comparing the observed t-statistic to the
# critical values -t^* and +t^*

t_obs
tstar

# Step 5. Writing up


###### week 5 code

# Effect size: Cohen's D
D <- (stats$M - 37) / stats$SD
D

# Assumptions

# 1. Random sample / independence

# 2a. Normality: let's check first for normality of the population data
plt.qq <- ggplot(temp_data, aes(sample = BodyTemp)) +
    geom_qq() +
    geom_qq_line() +
    labs(x = "Theoretical quantiles",
         y = "Sample quantiles",
         title = "QQ-plot")

(plt.h | plt.d) / (plt.b | plt.qq)

shapiro.test(temp_data$BodyTemp)

# 2b. Normality: if previous one didn't hold, is sample size large enough?
dim(temp_data)

plt.h | plt.b
kbl(stats, booktabs = TRUE, digits = 2,
    caption = "Descriptive statistics of body temperatures (°C)")
plt.qq
```