

## 1 The shape of body part measurements

The human body comes in various shapes and sizes. However, as daVinci knew, there are certain proportions that are consistent throughout. For this project two data sets are used which contain various measurements of human bodies.

To download the data sets issue these commands:

- > source("http://www.math.csi.cuny.edu/st/R/normtemp.R")
- > source("http://www.math.csi.cuny.edu/st/R/fat.R")

The normtemp data set<sup>1</sup> contains measurements of normal body temperature for 300 healthy adults in the variable temperature. The variable gender records the gender of the subject, and hr the heart rate in beats per minute.

The fat data set<sup>2</sup> contains many measurements of human bodies that can be done with a tape measure (circumference measurements), for instance the variable wrist contains measurements of wrist size in centimeters. Additionally, the variable body.fat contains body fat measurements.

After downloading the data sets, they may be attached so that the variable names are visible from the command line.

```
> attach(normtemp)
```

```
> attach(fat)
```

## 2 *t*-based statistical Inferences

What is normal body temperature? Common wisdom is that it is 98.6 °F. The normtemp data set may be used to test this assumption using a test of significance.

## 2.1 Basics of significance tests

We will test the *null hypothesis* that the distribution of normal body temperatures is normal with mean  $\mu = 98.6$  and unspecified variance, against the *alternative hypothesis* that the mean is less than this. That is

$$H_0: \mu = 98.6, \qquad H_A: \mu < 98.6.$$

<sup>&</sup>lt;sup>1</sup>This data set was contributed to the *Journal of Statistical Education* by Allen L. Shoemaker, http: //www.amstat.org/publications/jse/v4n2/datasets.shoemaker.html

<sup>&</sup>lt;sup>2</sup>This data set was contributed to the *Journal of Statistical Education* by Roger W. Johnson, http: //www.amstat.org/publications/jse/v4n1/datasets.johnson.html.

As we assume the data is normally distributed, we will use the T-statistic as a test statistic:

$$T = \frac{\bar{x} - \mu}{s / \sqrt{n}} = \frac{\text{observed} - \text{expected}}{\mathbf{SE}}.$$

Finally, we'll use  $\alpha=0.05$  for the level of significance.

Question 1: An assumption for using the *t*-statistic is that the data be normally distributed, or n be large enough. In this case, n is quite large, but nonetheless, verify graphically that the data appears to come from a normal distribution. Which graph did you use, and how did it show normality.

We can find the p-value with the following commands:

```
> xbar = mean(temperature)
> s = sd(temperature)
> n = length(temperature)
> T.obs = (xbar - 98.6)/(s/sqrt(n))
> pt(T.obs, df = n - 1)
```

```
[1] 1.205316e-07
```

We see that the *p*-value is tiny, about  $10^{-7}$ , and is much less than the significance level  $\alpha$ . The *p*-value is computed using the alternative hypothesis to specify what values of the test statistic are more extreme than the observed value, T.obs. In this case, as smaller values of  $\bar{x}$  support  $H_A$ , we see that smaller values are more extreme.

Alternatively, one can find the critical value(s) for this problem using  $H_A$  and  $\alpha$ . For this problem, the lone critical value corresponds to having  $\alpha \cdot 100\%$  of the area to the left of the critical value. That is, the critical value is the 0.05 quantile of the *t*-distribution and is found with

> qt(0.05, df = n - 1)
[1] -1.656752
> T.obs

[1] -5.454823

One can see that **T.obs** is much less than the critical value.

For this problem, the computer can do all of the work with the t.test() function. To use this function, one must specify the data (temperature), the null hypothesis (mu=98.6), and the alternative hypothesis (alt="less"). The function returns the observed value of T, the degrees of freedom, the *p*-value and more:

```
> t.test(temperature, mu = 98.6, alt = "less")
One Sample t-test
data: temperature
t = -5.4548, df = 129, p-value = 1.205e-07
alternative hypothesis: true mean is less than 98.6
95 percent confidence interval:
        -Inf 98.35577
sample estimates:
mean of x
    98.24923
```

You are left to compare the *p*-value to  $\alpha$  if so desired.

The alternative hypothesis may be specified three ways to the t.test function:  $H_A$ :  $\mu < \mu_0$  is specified with alt="less",  $H_A : \mu > \mu_0$  is specified with alt="greater", and  $H_A : \mu \neq \mu_0$  is specified with alt="two.sided".

Question 2: Some have suggested that  $98.2^{\circ}F$  is a better value for the mean body temperature. Do a test of significance of

$$H_0: \mu = 98.2, \qquad H_A: \mu \neq 98.2.$$

Is the difference statistically significant at the  $\alpha = 0.05$  level?

Question 3: Do females have a higher body temperature? We will ask this by testing if the mean body temperature for females is 98.2°F or is it more. That is

$$H_0: \mu = 98.2, \qquad H_A: \mu > 98.2.$$

You can extract the temperatures for just the females in the study with this command:

```
> females = temperature[gender == "female"]
```

1. First check that using the *t*-statistic is appropriate for this data. That is, either the data appears to come from a normal population or the data set (n) is large enough.

Is the statistic appropriate and why?

2. Now apply t.test() to find the *p*-value. Are the differences statistically significant at a level of  $\alpha = 0.05$ .

Question 4: A watch band maker assumes that the average wrist size is seven inches. We can use the wrist variable to check this assumption using a significance test of

 $H_0: \mu = 7$  inches,  $H_A: \mu \neq 7$  inches.

- 1. Check that a t-test is appropriate for this data, by verifying normality of the data, or a large n. Why is the t-test appropriate?
- 2. As the measurements in wrist are in centimeters, convert the null and alternative hypotheses into centimeters (1 inch = 2.54 cms).
- 3. Test to see if the difference between the mean wrist size of the data is statistically significant at the  $\alpha = 0.05$  level.

Question 5: The body mass index, BMI, is a measurement of weight divided by height squared in metric units. "Normal" values are in the range [18,5,25]. Do a significance test at the  $\alpha = 0.05$  level to see if the mean BMI is 25 versus an alternative that it is more.

- 1. First plot the data in BMI and argue if the *t*-test is appropriate for this data and why.
- 2. Write down the null and alternative hypotheses

$$H_0: H_A:$$

- 3. What is the *p*-value?
- 4. Is the difference statistically significant?

Question 6: A manufacturer of one-size-fits-all sweatpants wishes to know whether the mean waist size (the data is stored in the variable **abdomen**) has changed from the last time it designed a pair of sweatpants. Before, the manufacturer assumed the mean waist size was 34 inches. Do a significant test at the  $\alpha = 0.05$  level, testing

$$H_0: \mu = 34, \qquad H_A: \mu > 34.$$

- 1. Why is a *t*-test approporiate for the data in abdomen?
- 2. Rewrite the null and alternative hypotheses using cms (1 inch = 2.54 centimeters).
- 3. What is the *p*-value?
- 4. Is the difference statistically significant?

Question 7: The mean height of 17-year old Finnish boys is estimated to be 180cms. Suppose the data collected in the **height** variable is a random sample from adult males in the United States. Is the mean height statistically different from that of full grown 17-year olds from Finland at the  $\alpha = 0.05$  level of significance?

- 1. Specify the null and alternative hypotheses using inches.
- 2. Comment on the appropriateness of the t-test?
- 3. What is the *p*-value?