Dr Isabella Gollini

Probability and Statistics

Solution Example 2 – Lab – The testing of hypotheses and t-tests

A pharmaceutical company wishes to determine whether its new allergy product (A) is any better at reducing the level of a certain histamine in the blood stream than its current product (B). Two independent random samples of individuals were drawn from groups of people using product A and product B, respectively, and their histamine levels (in mg per cubic litre) were recorded. The data are given below.

Product A:	16.61	15.38	15.70	17.58	16.66	17.13
Product B:	18.66	19.52	16.98	18.19	17.20	

1. State carefully the statistical model that underlies an appropriate analysis, specifying in particular the unknown parameters.

2. Stating in terms of the model parameters the hypotheses that you are testing, write down a test statistic to investigate whether the mean level of the histamine for Product A is less than for Product B. Find the corresponding p-value and draw conclusions.

Solution
Load the data:

```
prA <- c(16.61, 15.38, 15.70, 17.58, 16.66, 17.13)
prB <- c(18.66, 19.52, 16.98, 18.19, 17.20)
```

1. It is assumed that the data for product A are a random sample from a normal $N(\mu_A, \sigma^2)$ distribution and the data for product B are an independent random sample from a normal $N(\mu_B, \sigma^2)$ distribution, where the unknown parameters are the mean μ_A, the mean μ_B and the common variance σ^2.

2. We test the null hypothesis $H_0 : \mu_A = \mu_B$ against the alternative $H_1 : \mu_A < \mu_B$, so that we have a one-tail test.

   Before proceeding to the test we calculate the sample mean and sample variance for the two samples:

	A	B
mean	16.51	18.11
variance	0.6990	1.1005

   The `R` code is:

```r
mean(prA); var(prA)

## [1] 16.51
## [1] 0.69896

mean(prB); var(prB)

## [1] 18.11
## [1] 1.1005
```

An unbiased estimate of σ^2 is given by the pooled estimate:

$$\frac{5(0.6990) + 4(1.1005)}{9} = 0.8774.$$

The test statistic is given by

$$\begin{aligned} t &= \frac{\bar{x} - \bar{y}}{\sqrt{s^2\left(\frac{1}{n} + \frac{1}{m}\right)}} \quad \sim t_{n+m-2} \\ &= \frac{16.51 - 18.11}{\sqrt{0.8774\left(\frac{1}{6} + \frac{1}{5}\right)}} \quad \sim t_{6+5-2} \\ &= -2.821. \quad \sim t_9 \end{aligned}$$

By looking at the table of the distribution function (or by looking at the table of the percentage points) of the t-distribution with 9 degrees of freedom we find that $p = F(-2.821) = 1 - F(2.821) = 1 - 0.990 = 0.010$, so that the value of the t-statistic is just about significant at the 1% level. So there is very strong evidence that the mean level of the histamine is lower for Product A than for Product B.

The `R` code is:

```r
t.test(prA, prB,
       alternative = "less",
       var.equal = TRUE)

##
##  Two Sample t-test
##
## data:  prA and prB
## t = -2.8208, df = 9, p-value = 0.01001
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##        -Inf -0.5602492
## sample estimates:
## mean of x mean of y
##     16.51     18.11
```