Problem 1. One Sample t-test

(Dalgaard, 2003 page 83) Here is an example concerning daily energy intake in KJ for 11 women (Altman, 1991, p.183).
5620, 5470, 5640, 6810, 6390, 6515, 6805, 7515, 7515, 8230, 8770

(a) Calculate the mean, and the standard deviation of these data.

(b) Before using R, test the following two set of hypotheses:
   \[ H_0: \mu = 7725 \]
   \[ H_a: \mu \neq 7725 \]
   Secondly,
   \[ H_0: \mu = 7725 \]
   \[ H_a: \mu \geq 7725 \]

(c) Create a 90%, a 95%, and a 99% confidence interval for the mean population of the daily energy intake of women.

(d) Read through the R-codes(8) handout.

(d) Now, answer the previous questions using R.

Note that all you need to do here is to employ the command

\[ \text{t.test(intake)} \]

or for a one-sided test:

\[ \text{t.test(intake,alternative="greater")} \]

or for a 90% confidence interval:

\[ \text{t.test(intake,conf.level = 0.90)} \]
Problem 2. Paired t-test

The price of fish varies by species and time. The average price received by fishermen and vessel owners for several species of fish increased from 41 cents per pound in 1970 to $1.10 per pound in 1980.

<table>
<thead>
<tr>
<th>Type of Fish</th>
<th>Price in 1970</th>
<th>Price in 1980</th>
</tr>
</thead>
<tbody>
<tr>
<td>COD</td>
<td>13.1</td>
<td>27.3</td>
</tr>
<tr>
<td>FLOUNDER</td>
<td>15.3</td>
<td>42.4</td>
</tr>
<tr>
<td>HADDOCK</td>
<td>25.8</td>
<td>38.7</td>
</tr>
<tr>
<td>MENHADEN</td>
<td>1.8</td>
<td>4.5</td>
</tr>
<tr>
<td>OCEAN PERCH</td>
<td>4.9</td>
<td>23</td>
</tr>
<tr>
<td>SALMON, CHINOOK</td>
<td>55.4</td>
<td>166.3</td>
</tr>
<tr>
<td>SALMON, COHO</td>
<td>39.3</td>
<td>109.7</td>
</tr>
<tr>
<td>TUNA, ALBACORE</td>
<td>26.7</td>
<td>80.1</td>
</tr>
<tr>
<td>CLAMS, SOFT-SHELLED</td>
<td>47.5</td>
<td>150.7</td>
</tr>
<tr>
<td>CLAMS, BLUE HARD-SHELLED</td>
<td>6.6</td>
<td>20.3</td>
</tr>
<tr>
<td>LOBSTERS, AMERICAN</td>
<td>94.7</td>
<td>189.7</td>
</tr>
<tr>
<td>OYSTERS, EASTERN</td>
<td>61.1</td>
<td>131.3</td>
</tr>
<tr>
<td>SEA SCALLOPS</td>
<td>135.6</td>
<td>404.2</td>
</tr>
<tr>
<td>SHRIMP</td>
<td>47.6</td>
<td>149</td>
</tr>
</tbody>
</table>

(a) Read the R-codes handout (8).

(b) Test the hypothesis that suggests the mean difference between the price of the fish in 1970 and 1980 is zero. Conduct a two-sided test and also a one sided test with the alternative hypothesis suggesting that "the population price of the fish has increased from 1970 to 1980".

(c) Create a 95% confidence interval around the mean difference of the price.

Note that all you need to do here is to employ the command

\[ \text{t.test}(1970, 1980, \text{paired}=T) \]
Problem 3.1 Two-sample t-test

In your R prompt type:

```r
library(ISwR)
```

This way, you will append the datasets in Dalgaard to your software. Now, type:

```r
data(energy)
attach(energy)
energy
```

This should give you the dataset called `energy` that has 22 rows and 2 columns. It contains data on the energy expenditure in groups of lean and obese women. First, note that due to the command `attach` the dataset `energy` will be the working dataset in R so you do not need to retype the name of the dataframe each time you want to manipulate it.

Let’s do a side-by-side plot of energy expenditure in terms of stature:

```r
plot(expend~stature)
```

Now:

(a) Not using R, perform a t-test to compare the population means of the energy expenditure between the two groups.

(b) Now, use R to address the above question.

   **Note that all you need to do here is to employ the command:**

   ```r
t.test(expend~stature)
   ```

(c) Using R, create a 90%, a 95% and 99% confidence interval for the mean differences.
Problem 3.2 Comparison of Variances

Remember the $F$–test? We can use that test to compare the variances of the two groups. In other words, we want to test:

$H_0: \sigma_1 = \sigma_2$

$H_a: \sigma_1 \neq \sigma_2$

This can be done quite simply using the following command:

```r
var.test(expend~stature)
```

Try this! and report the results.
Problem 4 z-table, t-table in R

Previously, we showed that the areas under the $z$–table can be obtained using $R$. For example, to get the area below 1.96 in a $z$–curve it is sufficient to type:

```r
> pnorm(1.96)
[1] 0.9750021
```
or, equivalently, the cut-off point below which 97.5% of the standard normal curve is covered is 1.96:

```r
> qnorm(0.975)
[1] 1.959964
```

We can do the same with the other two distributions we studied namely, $t$ and $F$ distributions. To practice this, let’s revisit the first problem in the current homework:

1. (Moore and McCabe, 1998) The one-sample $t$ statistic from a sample of $n = 30$ observations for the two-sided test of :

   \[ H_0 : \mu = 64 \]
   \[ H_a : \mu \neq 64 \]

   has the value $t_{obs} = 1.12$.

   (a) What are the degrees of freedom for $t$? (b) Locate two critical values $t^*$ from the t-table that bracket $t_{obs}$. What are the right-tail probabilities $p$ for these two values? (c) Relying on your answer to the previous section, how would you report the $P$–value for this test? (d) Is the value $t = 1.12$ statistically significant at the 10% level? At the 5% level? (e) Find the exact $P$–value using R.

to calculate the $p$-value for this problem we only need to type:

```r
> 2*(1-pt(1.12,29))
[1] 0.2719041
```

In other words, the $p$-value for this problem is 0.27. what happened? Analyse the R-command!