Performing significance tests with R.

Using R with pmg makes it fairly easy to compute the p-value of a significance test. For the significance tests discussed in class we have the following steps

1. Determine H_o and H_A, the null and alternative hypotheses.
2. Decide on a test statistic
3. Compute the observed value of the test statistic from the sample data.
4. Find the p-value **or** compare the observed value to the critical value.

The computer allows us to do just that, and not have to do the compuations that can be messy.

To illustrate first we need to start things up:

1. Start R by double clicking on its icon from the desktop
2. Start pmg by issuing the command

   ```
   > library(pmg)
   ```

3. Minimize R's main window.
4. Load some data sets. First load the MASS package by using the dialog under the File > Load package ... dialog.
5. Then open the "Data > Load data set..." dialog and load the following data sets:

   mtcars, Cars93, cats, DDT, and michelson

   [You could do this from the command area with:

   ```
   > library(MASS)
   > data(c("mtcars", "Cars93", "cats", "DDT", "michelson"))
   ```

   ]

Now your data sets should appear in the pmg window, similar Figure 1.

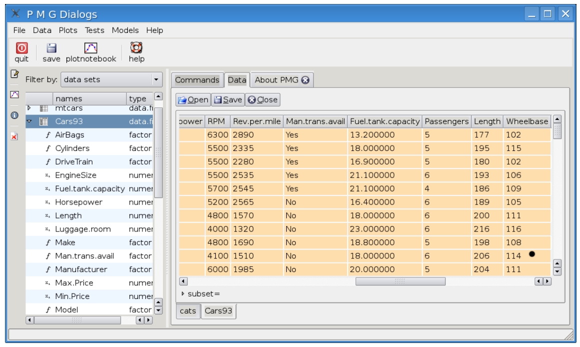

Figure 1: PMG dialog window after some data sets are loaded

1 Performing a one-sample *t*-test

The *t*-test is the name used when the test statistic is

$$T = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

The assumptions are that the data is normal.

Suppose we wish to test whether the mean weight of cars from 1973 is more than 3150 pounds. Since the scale is in 1000s of pounds we with to perform a significance test of

$$H_o : \mu = 3.15 \qquad H_A : \mu > 3.15$$

To do this easily use the "Dynamic tests" dialog under the Tests menu. Once opened select a test. In this case we want the "1-sample t.test." In Figure 2 the two-sample test is shown.

By selecting this test, you have done step 2, choosing a test statistic. Next we add in the sample information so that the computer can compute step 3. Do this by finding the `wt` variable in the `mtcars` data set and dragging into the space labeled "Drop variable(s) here." (Alternately you can click that space, type `mtcars$wt` and then hit enter.)

Finally we have to specify the null and the alternative. This is done by changing the pop-up value from "not equal to" to "greater than" and clicking on the 0 and typing in 3.15 then enter.

If you do this properly, the *p*-value should be computed for you. If you get 0.3500412 you are in business.

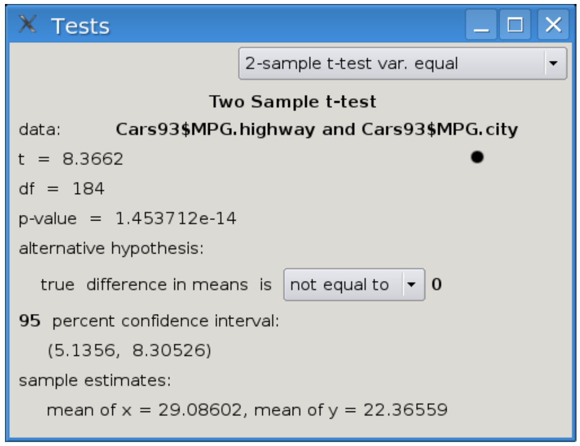

Figure 2: Dynamic tests dialog for two-sample t.test

1.1 Problems

1. Follow the steps above. The output has other values. Explain what these mean

   (a) $t = 0.3887992$

   (b) $\mathrm{df} = 31$

   (c) mean of x = 3.21725

2. The variable `mpg` records the miles per gallon. Do a significance test of

   $$H_0 : \mu = 20, \quad \mu \neq 20$$

   What is the p-value?

3. The data set `cats` has the variable `Bwt` recording the body weight of some cats used in an experiment. The weights are recorded in kilograms. Test the following

   $$H_0 : \mu = 2.5 \qquad H_A : \mu > 2.5$$

   What is the p-value.

4. The data set `DDT` records DDT amounts on random samples of Kale. Suppose mean of 3.15 is considered normal, but one larger than 3.15 is not. Do a significance test to see if the difference is statistically significant. (First write down the null and alternative, then find the p-value.)

5. The DDT data set is small, 15 observations. In this case, you need to assume the population in question is normally distributed. This means that the sample should have a density plot that is roughly bell-shaped. Open the "Lattice Explorer" dialog under the "Plots" menu and make a density plot. Is the data roughly normal?

2 two-sample tests

We learned two two sample tests. First if the two samples are large, then we can perform test with test statistic

$$T = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

Next, if the samples are small but we assume the variances are equal and normal populations then we use

$$T = \frac{\bar{x_1} - \bar{x_2}}{s_p\sqrt{1/n_1 + 1/n_2}}$$

The first one is computed using "2-sample t-test" the latter using "2-sample t-test var equal".

Has mileage improved between 1983 and 1993? We will use the data from `mpg` in the `mtcars` data set and the data `MPG.city` from `Cars93`. To be precise we are testing

$$H_0 : \mu_1 = \mu_2 \qquad \mu_1 > \mu_2,$$

where μ_1 is the population mean for the 1993 mileage.

To compute the p-value, select the 2-sample t-test test as in Figure 2. Then drag first `MPG.city` and then `mpg` to the area "Drop variable(s) here". Since the null hypothesis is always the same (the difference $\mu_1 - \mu_2 = 0$) we don't fill that in, but we do select "greater than" to specify the alternative (being careful to note the order the variables were dropped.)

2.1 Problems

1. What is the p value for the above test? Do you accept or reject?

2. Do a test of `MPG.city` to `MPG.highway` with

$$H_0 : \mu_1 = \mu_2 \qquad H_o : \mu_1 < \mu_2$$

What is the p-value? Do you accept or reject?

3. Did the weight of cars increase between 1973 and 1993? Use the variables `wt` from `mtcars` and `Weight` from `Cars93` to test. Specify

$$H_0 : \qquad H_A :$$

   and find the p-value.

4. Sometimes the data is not suitable for the dialog, and other methods must be used. Take for instance the `cats` data set. The variable `Sex` records the gender of the cats. Is there a difference in weights between the male and female cats. That is test

$$H_0 : \mu_f = \mu_m \qquad H_A : \mu_f \neq \mu_m$$

   To do so, you can type the following command in the "Commands" area

```
> t.test(Bwt ~ Sex, cats)
```

   Find the p-value and report it as well compare it to