620-270  Applied Statistics

Computer Lab 3

In this lab session you will use MINITAB to

- perform inference (confidence intervals and hypothesis tests) on paired samples.
- perform inference on two independent samples, assuming equal or unequal variances.
- construct plots to examine relationships between variables.

1. Comparing two samples: paired

An assessment was made of 8-year-old radiata pine trees in a plantation in north-eastern Victoria, to examine whether infection caused by the disease “pine needle blight” was greater in trees bordering large gaps than in the remainder of the plantation. Ten plots were assessed, each containing an area with a gap, and a fully stocked area without a gap. The disease scores for the twenty assessed areas were as follows:

<table>
<thead>
<tr>
<th>Plot</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trees bordering gap</td>
<td>4.08</td>
<td>3.25</td>
<td>3.83</td>
<td>3.80</td>
<td>4.27</td>
<td>2.67</td>
<td>4.00</td>
<td>3.44</td>
<td>2.13</td>
<td>1.44</td>
</tr>
<tr>
<td>Fully stocked area</td>
<td>3.25</td>
<td>3.33</td>
<td>3.17</td>
<td>3.00</td>
<td>3.08</td>
<td>2.42</td>
<td>3.42</td>
<td>3.08</td>
<td>2.25</td>
<td>1.20</td>
</tr>
</tbody>
</table>

Enter the data into columns C1 and C2 in MINITAB (no need to enter plot numbers, as the rows in MINITAB are already numbered).

Because the data are from paired samples (the pairs corresponding to plots), we are interested in the differences. Calculate the differences and store them in C3, using either **Calc → Calculator** or by typing `LET C3 = C1 - C2` in the session window.

Test the hypothesis (at the 0.05 significance level) that the gaps have no effect on disease levels, by applying

**Stat → Basic Statistics → 1-Sample t**

to column C3. This will also give you a confidence interval for the mean difference. Make sure that the confidence level is set at 95% by clicking on the **Options** button.

An even easier way to perform a paired-sample t-test (without having to calculate differences) is by

**Stat → Basic Statistics → Paired t**

selecting columns C1 and C2. Note the command generated by this menu operation, **Paired C1 C2**. Typing this would be as easy as using the menu.

Use the menu again, and this time tick Dotplot of differences under the **Graphs** button. Also change the confidence level to 99% under **Options**. On the dotplot, note how $H_0$ is outside even a 99% CI.

2. Comparing two samples: unpaired, equal variances

The following data are the weights (in kg) of AFL players wearing the numbers 1 to 15 for Carlton and Collingwood (some numbers are not represented with Collingwood):

<table>
<thead>
<tr>
<th>Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carlton</td>
<td>99</td>
<td>78</td>
<td>83</td>
<td>89</td>
<td>88</td>
<td>85</td>
<td>92</td>
<td>92</td>
<td>86</td>
<td>85</td>
<td>101</td>
<td>72</td>
<td>81</td>
<td>89</td>
<td>92</td>
</tr>
<tr>
<td>Collingwood</td>
<td>72</td>
<td>92</td>
<td>91</td>
<td>72</td>
<td>92</td>
<td>80</td>
<td>77</td>
<td>81</td>
<td>80</td>
<td>97</td>
<td>76</td>
<td>85</td>
<td>102</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Assume that these weights are normally distributed, and that the sample from each club is essentially random, i.e. numbers are allocated independently of weight.

Although the numbers create pairs, these are arbitrary and of no statistical consequence. It is therefore an unpaired two-sample problem.

Enter the weights into columns C4 and C5. Test whether the mean weights of Carlton and Collingwood players are the same (using $\alpha = 0.05$), with

**Stat → Basic Statistics → 2-Sample t**, selecting columns C4 and C5 under **Samples in different columns**. Tick the box **Assume equal variances**.

Examine the output. Not only is there information regarding the hypothesis test (test statistic, $P$-value, etc.), but there are summary statistics for both samples and a 95% CI for the mean difference.

Now we will perform the analysis with all the weights in one column, and the “subscripts” indicating the teams in another. Firstly, give columns C4 and C5 the headings **Carl** and **Coll** respectively. Then stack these columns into C6, with team subscripts in C7, using **Manip → Stack → Stack Columns**.

Now proceed again with

**Stat → Basic Statistics → 2-Sample t**, this time clicking on **Samples in one column**.

Examine the output: it is the same as before, except for the generated **MINITAB** command, which is different because the weights are all in one column.

3. **Comparing two samples: unpaired, unequal variances**

Now we will not assume equal variances. This involves calculating the degrees of freedom for an approximate $t$ distribution using the rather complicated formula given in lectures. Of course, **MINITAB** does this for you — all you need to do is ensure that the **Assume equal variances** box isn’t ticked.

Perform the analysis assuming unequal variances, and examine the output to see how it differs from the analysis involving equal variances.

Run it again with boxplots of the two teams (tick on this under the **Graphs** button).

4. **Plotting relationships between variables, and plotting distributions**

The following measurements represent salinity (%) and nitrate concentration ($\mu$M/L) of eight water samples.

<table>
<thead>
<tr>
<th>Salinity</th>
<th>35.43</th>
<th>36.10</th>
<th>35.74</th>
<th>35.3</th>
<th>35.4</th>
<th>35.91</th>
<th>35.48</th>
<th>36.28</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nitrate</td>
<td>30.0</td>
<td>24.2</td>
<td>25.4</td>
<td>29.8</td>
<td>30.7</td>
<td>24.0</td>
<td>28.5</td>
<td>22.7</td>
</tr>
</tbody>
</table>

Enter the data into columns C8 and C9, and name them.

Create a scatterplot of the data using the command **plot c9*c8**. Note that the column specified first is plotted on the vertical axis.

Now create the plot using the menu **Graph → Plot**, filling in the Y and X variables. There are many options available for plotting, some of which we will use later.
The plot command can also be used to construct a probability density function. Try the following commands to plot the pdf of a $\chi^2_3$ distribution:

```
MTB > set c10
DATA> 0:20/0.1
DATA> end
MTB > pdf c10 c11;
SUBC> chis 5.
MTB > plot c11*c10;
SUBC> connect.
```

The statement 0:20/0.1 creates a variable ranging from 0 to 20, in increments of 0.1. This is stored in C10. The pdf is then generated for each value in C10 and stored in C11. The subcommand connect joins the points by straight lines.

Create a plot of a $t_{10}$ distribution. Remember that a $t$ distribution is symmetrical about 0, so let the values range from (say) -5 to 5.

5. Correlation coefficient

For the nitrate vs salinity relationship, have a look at the scatterplot and guess the value of the correlation coefficient $r$.

Find $r$ using the command correlate c8 c9. Note that you only need the first four letters of a MINITAB command. Now find $r$ using the menu by

Stat → Basic Statistics → Correlation.

Try the following and see if it changes the value of $r$:

(a) Interchange the variables, i.e. type correlate c9 c8.

(b) Increase all the salinity values by 5, i.e.

```
MTB > let c12 = c8 + 5
MTB > corr c12 c9
```

(c) Multiply all the nitrate values by 3.

When you have finished, don't forget to close MINITAB, and shut down Windows.

On the reverse side of this page, there are some instructions on saving work in the labs, to retrieve later at home.