Sample size calculations

In this practical we will work further on sample size calculations and start cross-over clinical trials. In the first section, we will study the power of a statistical test as a function of the sample size, and then in the following section, we will learn about 2x2 cross-over clinical trials.

The R commands you will need for the power plot and other calculations are given below. Remember to keep using the help() command.

1. A clinical trial was conducted to compare the cariostatic effect, the ability to prevent tooth decay, of two different toothpastes in children of a given age. The response variable is the change in dmfs after two years: dmf stands for decayed, missing, filled and is in relation to the teeth. A mean dmf increment of 0.5 is considered by the investigators as a worthwhile treatment difference. Previous research has established that the dmfs increment in two years for this age group is 1.25 dmfs.

(a) Plot the power $1 - \beta$ as a function of the sample size $n$ for (two-sided) $2\alpha$ significance levels of 0.01, 0.05, 0.1, 0.2.

Describe the relationship between the power to detect a true increase in height and the sample size for increasing levels of statistical significance.

(b) From your plot, how many children would be required for an 85% chance of detecting the specified increase of 0.5cm as significant at the 1% level?

(c) Again using your plot, if there were at most 500 children available for the study, what is the maximum achievable power?

(d) Use R to obtain exact answers to (b) and (c) above.

```r
delta <- 0.5
sigma <- 1.25
alpha <- 0.01
n <- seq(1, 500, length=100)

# Sample size formula for comparing two means
power0 <- pnorm(delta/sigma*sqrt(n/2) - qnorm(1 - alpha/2))

# Plotting the power curve
plot(n, power0, type="l", col="cyan", ylim=c(0,1))
```

Assign values for $\delta_1$ and $\sigma$, and set $\alpha = 0.01$ initially.

Generate a sequence of length 100 of possible values of $n$.

$\Phi(z)$ where $Z \sim \text{N}(0,1)$.

Gives $\Phi^{-1}(z)$.

Enter the formula obtained by rearranging the familiar sample size formula for comparing two means $n = 2 \left( \frac{z_\alpha + z_\beta}{\delta_0} \right)^2$.

Create a plot of power against $n$, using lines.
You can obtain a smoother plot using smoothing splines.

\[
power0.ss <- \text{smooth.spline}(n, power0)
\]
\[
\text{plot(power0.ss, type="l",}
\]
\[
lty=2, \text{ col="blue"},
\]
\[
\text{ylim=c(0,1))}
\]

Update \(\alpha\) and create another set of values for power.

\[
\alpha <- \ldots
\]
\[
power1 <- \ldots
\]
\[
\text{lines}(n, \text{ power1, lty=2,}
\]
\[
col="\text{green}"
\]
\[
\vdots
\]
\[
\text{abline(h=0.85)}
\]
\[
\text{abline(v=250)}
\]

Add a title to the plot. It is recommended that you put your name on the top of any plots you print in order to avoid confusion.

\[
\text{title("This is my title")}
\]

Use the mouse pointer to read \(n\) points off a plot.

\[
\text{locator(}n\text{)}
\]

Add a legend to your plot. This legend is for the first 2 curves, power0 and power1.

\[
\text{legend(350, 0.2,}
\]
\[
c("\text{power0", "power1"),}
\]
\[
\text{col=c("cyan", "green"),}
\]
\[
lty=1:2)}
\]

Cross-over trials

In this practical, we will analyse data obtained from a \(2 \times 2\) cross-over trial comparing two treatments for mild to acute bronchial asthma. The treatments were single doses of two active drugs which we will label \(A\) and \(B\). The response of interest was the forced expired volume (in litres) in one second \((FEV_1)\). Half of the available patients were to receive the treatments in the order \(AB\), and the other half were to receive them in the order \(BA\). However, not all participants completed the trial and so the final set of results contains measurements on 8 patients who received the treatments in the order \(AB\) and 9 patients who received them in order \(BA\).

The baseline \(FEV_1\) measurement \(z_{ij1}\) was taken during the run-in period immediately prior to administering the first treatment. Then, two and three hours later, \(FEV_1\) measurements were taken again and the average of these is \(y_{ij1}\). After a suitable wash-out
period during which a second baseline measurement $z_{ij2}$ was recorded, the second treatment was administered and measurements again taken at two and three hours to give the average value in the second period $y_{ij2}$.

The dataset for this session is in the file `fev.dat` which you can download from MyUni in the usual way.

Forced expiratory volume ($FEV_1$) in litres from a $2 \times 2$ crossover trial comparing two treatments A and B for mild to acute bronchial asthma.

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$AB$</td>
<td>$BA$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Patient</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 1</th>
<th>Period 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>z_{1j1}</td>
<td>y_{1j1}</td>
<td>z_{1j2}</td>
<td>y_{1j2}</td>
<td>z_{2j1}</td>
<td>y_{2j1}</td>
<td>z_{2j2}</td>
</tr>
<tr>
<td>1</td>
<td>1.09</td>
<td>1.28</td>
<td>1.24</td>
<td>1.33</td>
<td>1.74</td>
<td>3.06</td>
</tr>
<tr>
<td>2</td>
<td>1.38</td>
<td>1.60</td>
<td>1.90</td>
<td>2.21</td>
<td>2.41</td>
<td>2.68</td>
</tr>
<tr>
<td>3</td>
<td>2.27</td>
<td>2.46</td>
<td>2.19</td>
<td>2.43</td>
<td>3.05</td>
<td>2.60</td>
</tr>
<tr>
<td>4</td>
<td>1.34</td>
<td>1.41</td>
<td>1.47</td>
<td>1.81</td>
<td>1.20</td>
<td>1.48</td>
</tr>
<tr>
<td>5</td>
<td>1.31</td>
<td>1.40</td>
<td>0.85</td>
<td>0.85</td>
<td>1.70</td>
<td>2.08</td>
</tr>
<tr>
<td>6</td>
<td>0.96</td>
<td>1.12</td>
<td>1.12</td>
<td>1.20</td>
<td>1.89</td>
<td>2.72</td>
</tr>
<tr>
<td>7</td>
<td>0.66</td>
<td>0.90</td>
<td>0.78</td>
<td>0.90</td>
<td>0.89</td>
<td>1.94</td>
</tr>
<tr>
<td>8</td>
<td>1.69</td>
<td>2.41</td>
<td>1.90</td>
<td>2.79</td>
<td>2.41</td>
<td>3.35</td>
</tr>
<tr>
<td>9</td>
<td>0.96</td>
<td>1.16</td>
<td>1.01</td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Enter R and read in the data as follows:

```r
> fev <- read.table("fev.dat", header=T)  # Read the FEV1 dataset into a data frame, and look at the data.
> attach(fev)
> objects(fev)
> objects(2)
```

Your task is to analyse these data by answering the following questions, and thereby reach a conclusion regarding the difference between treatments A and B.

The R commands you need for each step are set out on the attached sheet. Remember to keep using the `help()` command.

1. Produce the **subject profiles** plot for each group. Produce the **groups-by-periods** plot.

   For each plot, describe what you see. Comment on any interesting features such as trends or patterns, or any unusual observations.

2. Obtain the **subject differences versus totals plot**. Comment on and interpret any observed features, and give a tentative conclusion.
3. Analyse the data using an appropriate sequence of \( t \)-tests, interpreting the results of each test. What is your conclusion regarding a possible difference between the treatments \( A \) and \( B \)?

**Assignment 2**: hand in your plots, a summary of your analysis of the data using \( t \)-tests, and your conclusions.

**Due**: by noon on Friday 17th September (Week 8).
Subject profiles plots

These can be generated using the following sequence of commands.

```r
> plot(1, 1, xlab="Periods", ylab="Response",
     main="Group 1 Subject Profiles", xlim=c(0.5, 2.5),
     ylim=range(c(y1, y2)), type="n")
```

This produces the base plot to which we will add the points and lines. Consequently we don’t need to put any actual data on the plot at this stage, so I have just used the coordinates (1,1). The `xlab=` and `ylab=` arguments specify the text labels to be placed on the x- and y-axes respectively. The `main=` argument specifies the main title of the plot. The arguments `xlim=` and `ylim=` specify the bounds for the x- and y-axes, and the `range()` function returns a vector of size two, consisting of the minimum and the maximum of the data specified. Finally, the argument `type="n"` tells `plot()` that we don’t want to plot the data—just the axes and labels.

```r
> for (i in seq(along=Group)[Group == "AB"])
     lines(c(1, 2), c(y1[i], y2[i]), type="b")
```

The `for` command works just like in other computing languages: in this case `i` takes on each value of the expression to the right of `in` in turn. The expression `seq(along=Group)` generates a sequence from 1 to the total length of `Group` (this is equivalent to writing `1:length(Group)`), and the bit in square brackets asks for only the observations from group 1 (AB).

The `lines()` function plots points in the same way as `plot()`, except that it adds lines to an existing plot rather than creating a new one (you saw how to do this in the last computing practical). The `type="b"` argument plots both lines and points.

Now repeat the commands with appropriate modifications to generate the subject profiles plot for Group 2 (BA).

Tables of means, medians, sums

You will need the group means in each period for the groups-by-periods plot. The command

```r
> mean1 <- tapply(y1, list(Group), mean)
```

gives a table of `y1` by `Group` (or whatever is contained in the `list()` function call) and applies the function `mean`. Any other function can be used, such as `median()` or `sum()`.

Have a look at `mean1` and check the means are correct. Modify the above command to obtain `mean2`, the group means in Period 2.
Groups-by-Periods Plot

Here, we will create the plot first, then add the appropriate labels.

```r
> plot(c(1,2), c(mean1[1], mean2[2]), ylim=c(1.0,3.0), xlab="Period", ylab="Mean", main="Groups-by-periods plot", type="b")
> lines(c(1,2), c(mean1[2], mean2[1]), type="b")
```

In order to add text to a plot, you will need to use the following function.

```r
> text(c(1,1,2,2), c(mean1, mean2), labels=c("1A", "2B", "1B", "2A"))
```

The `text()` function adds text to a plot. If the `labels=` argument is omitted, a sequence of integers starting at one will be used. Note that you could use `locator(4)`.

Subject differences versus totals plot

Firstly, you need to create the totals \( t \) and differences \( d \) for each group. For example, for Group 1, these are

```r
> t1 <- (y1 + y2)[Group=="AB"]
> d1 <- (y2 - y1)[Group=="AB"]
```

Also find the subject totals and differences \( (P2 - P1) \) for Group 2 and call them \( t2 \) and \( d2 \).

You can plot two sets of points on one set of axes using a different plotting character for each set. First, create the plot axes using `plot()`, using the argument `pch=0`.

```r
> plot(t1, d1, pch=0, xlab="Totals", ylab="Differences", main="Differences-vs-totals Plot", xlim=range(c(t1,t2)), ylim=range(c(d1,d2)))
```

You may also need to specify `xlim=` and `ylim=` explicitly, so that points added later will fit on the plot. (Check this by looking at the ranges of the totals and differences.)

Now add points, using a different plotting character:

```r
> points(t2, d2, pch=3)
```

Remember that you saw how to produce convex hulls and add them to plots in the first introductory R exercise.

```r
> hull.1 <- chull(t1, d1)
> hull.2 <- chull(t2, d2)
> polygon(t1[hull.1], d1[hull.1], density=0)
> polygon(t2[hull.2], d2[hull.2], density=10)
```
**t-tests**

$t$-tests are simple in R. For example,

```r
> t.test(t1, t2, conf.level=0.95)
```

This command will perform a two-sample $t$-test, giving a $p$-value, estimates of the means, and a confidence interval corresponding to the value given in the `conf.level=` argument. To find a $p$-value given a test statistic $T_s$ on $df$ degrees of freedom, use

```r
> pt(Ts, df)
```

Other $t$-distribution functions are `rf()`, `qf()`, and `df()`. You can look these up using `help()`.

**Miscellanea**

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>pf(Fs, df1, df2)</code></td>
<td>Find the p-value given an $F$-statistic $Fs$ on $df1$ and $df2$ degrees of freedom.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>qqnorm()</code></td>
<td>This function produces normal probability plots.</td>
</tr>
</tbody>
</table>

Patty Solomon
August 2003