Testing for Differences: One sample, Independent Samples and Paired Samples t-tests

Boniface Karia

February 17, 2016

Objectives:

- Recognise when to use a one sample t-test, independent samples t-test, paired samples t-test;
- Explain and check the assumptions and conditions for each test;
- Run and interpret a ‘One sample t-test’ to show a sample mean is different from some hypothesised value;
- Run and interpret an ‘independent samples t-test’ to show the difference between two groups on one attribute;
- Run and interpret a ‘paired samples t-test’ to show the difference between two attributes as assessed by one sample

One Sample t-test

Hypothesis Testing for a Single Population Mean;

- The One-Sample T Test compares the mean score of a sample to a known value. Usually, the known value is a population mean.

One-sample t-test for the mean

We test the hypothesis H_0: $\mu = \mu_0$ using the statistic

$$t_{n-1} = \frac{\bar{y} - \mu_0}{SE(\bar{y})},$$

where the standard error of $\bar{y}$ is: $SE(\bar{y}) = \frac{s}{\sqrt{n}}$.

When the conditions are met and the null hypothesis is true, this statistic follows a Student's t-model with $n - 1$ degrees of freedom. We use that model to obtain a P-value.

Assumptions and Conditions

- Independence Assumption The data values should be independent.
- Randomisation Condition The data arise from a random sample.
- 10% Condition The sample size should be no more than 10% of the population.
- Normal Population Assumption The data should be from a population that follows a normal model.
  Why do we need to ensure the data is not extremely skewed?
  How can we check these assumptions?

Example: One Sample t-test

The mean yield for a certain crop is 8 kg/m2. This year, a new pesticide was applied and the yield from nine randomly selected plots are:
9.3 8.2 7.9 8.8 9.4 8.6 8.9 9.5 9.0
Any evidence that yield has increased? - Test the Ho: $\mu = 8.0$ against Ha: $\mu > 8.0$
We create a simple vector of yield using the function c()

```
yield<-c(9.3,8.2,7.9,8.8,9.4,8.6,8.9,9.5,9.0) # create an obj
yield    # to show contents of object 'yield'
```

```
## [1] 9.3 8.2 7.9 8.8 9.4 8.6 8.9 9.5 9.0
```

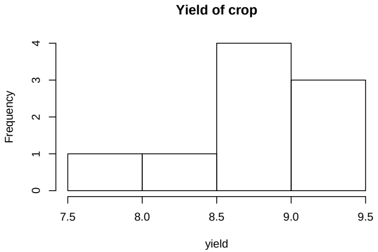
Exploring the data

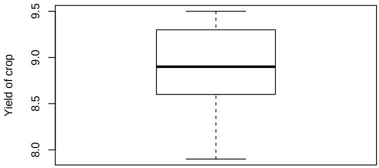
Before undertaking any statistical analysis it is good to explore the data. We start with numerical summaries.

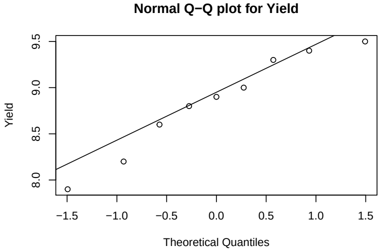
```
summary(yield)  # to summarise the variable 'yield'
```

```
##     Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    7.900   8.600   8.900   8.844   9.300   9.500
```

What is the estimated mean yield from the sample? ....
Before we do the t-test we check the distribution of the data using histogram, boxplot, normal qqplot or stem and leaf.







What is your comment regarding the normality assumption on the data?

Cont

Finally, we might want a more formal test of agreement with normality (or not). R provides the Shapiro-Wilk test:

```
shapiro.test(yield)          # do a Shapiro-Wilk test on 'yiel
```

Alternatively, we use the Kolmogorov-Smirnov test using ks.test() function

```
ks.test(yield, "pnorm", mean = mean(yield),
        sd = sqrt(var(yield)))
```

Here we are testing the null hypothesis: the data are sampled from the normal distribution. What is your comment on the test results?
Note: Use normality tests with caution:
Let us perform the one sample t-test on the data 'yield' specifying the alternative hypothesis is "greater" as follows:

Cont

```r
t.test(yield,mu=8,alt="greater")
```

```
##
##  One Sample t-test
##
## data:  yield
## t = 4.6819, df = 8, p-value = 0.000789
## alternative hypothesis: true mean is greater than 8
## 95 percent confidence interval:
##  8.509051      Inf
## sample estimates:
## mean of x
##  8.844444
```

```r
# specify hypothesized mean mu=8
 #by default alternative is 2-sided
```