Tutorial Pack 3

(To be completed during LW3 tutorial)

(90 minutes)

LEARNING OBJECTIVES

- **To provide practical experience of using R to carry out analysis of A/B, Multivariate and Margin of Error Tests**
- **To develop students understanding of key search engine algorithms**

LEARNING OUTCOMES

- **By the end of this tutorial students will have;**
  - **Performed the ANOVA in R**
  - **Calculated the margin of error of an online poll**
  - **Revised key information retrieval and search engine concepts.**
  - **Developed an understanding of the advantages and disadvantages of competing search engine algorithms and information retrieval models.**

RESOUCRES AND TOOLS REQUIRED

- **R or RStudio**
- **LW2/3 lecture slides**

Our previous experiment collected data on users' behaviour indirectly. In the LW3 lecture we also mentioned ways in which we can measure user needs and preferences more directly, i.e., using **polls** and **surveys.** The intention of carrying out a poll is to sample opinion from a subset of users and use that information to draw conclusions about the wider population of users thinks.

Although in principle it is *possible* to ask every user, in practice it is not feasible. This could be due to various factors, including users not wishing to answer or not everyone accessing the site whilst the poll is running.

Let's consider an ecommerce website uses an algorithm to determine which products it should recommend to users based on their previous purchase history. Suppose the ecommerce website updates its algorithm and wants to find out if users find its recommendations more useful. One way could be to carry out a simple "Yes it is useful" or "No, it is not useful" poll.

Suppose the company carried out the experiment fairly and asked random users to vote "Yes" or "No". The total number of people who took part in this poll is 357.

N= 357	**Yes, useful**	**No, not useful**
Answered	192	165

Based on this set of results it appears the new algorithm (53.78%) is slightly favoured over the previous one (46.21%). However, since our poll represents the thoughts from a small sample of all users how can we determine how indicative this result is of what the entire userbase thinks.

The **margin of error** of a poll tells us by how much we can expect the estimated proportion in our sample who voted "Yes" to differ (above and below) from the true population value (assuming all users were asked). It can be expressed through eq1 as shown below.

$$n = p(1-p)\left(\frac{Z}{E}\right)^2 \qquad (\mathrm{E}q1)$$

In this case,

n ~ sample size,

p ~ proportion we expect to respond yes

Z ~ value from the standard normal distribution (usually 1.96)

E ~ desired margin of error

Questions

Using your knowledge of polls, answer the following questions.

Qnum	Question	ANSWER
1	What happens to the sample size required if we are very confident that most people will find the recommendations useful?	
2	Assuming that we are 60% confident in our newly developed product recommendation algorithm, how many users would we need to poll to have a margin of error of 1%.	
3	Does this number change if we are only 40% confident but still want to achieve a margin of error of 1%?	

Sometimes we may be interested in analysing the results of a poll retrospectively. In this case we cannot control how many people took part but would still like to know the margin of error associated with the result of the poll. In this case we can rearrange eq1 to eq4 to derive the margin of error (unknown) in terms of the other (known) variables.

$$n = p(1-p)\left(\frac{Z}{E}\right)^2 \qquad (Eq1)$$

By rearranging (Eq1)

$$1 = \frac{p(1-p)\left(\frac{Z}{E}\right)^2}{n} \qquad (Eq2)$$

$$1 = \sqrt{\frac{p(1-p)}{n}}\frac{Z}{E} \quad (Eq3) \quad \Rightarrow \quad E = \sqrt{\frac{p(1-p)}{n}}Z \Rightarrow \quad (Eq4)$$

Questions

Using your knowledge polls, answer the following question.

Qnum	Question	ANSWER
4	Determine the margin of error of the original poll and calculate the upper and lower bounds for the proportion of people that found the new algorithm useful.	

In the learning week 2 lecture, we performed **one way analysis of variance** (ANOVA) to determine whether the average time spent on a page was equal across three different versions of a page, versions A, B and C. The data from that experiment is repeated below.

Page Version	Time Spent on page (seconds)	Average time
A	56	
A	12	
A	33	
A	2	
A	64	33.4
B	14	
B	65	
B	24	
B	32	
B	7	28.4
C	45	
C	5	
C	21	
C	3	
C	43	23.4

For ANOVA, the null hypothesis is that there is no statistically significant difference in the means of any two independent groups. Where more than two groups are used, ANOVA analysis cannot tell us which groups differ the most. For that we need to also combine the ANOVA test with the **TukeysHSD** (Honest Significant Difference) test.

Like all statistical tests, the ANOVA test has several assumptions that need to be satisfied before it can be used. These include:

- The dependant variable being numerical
- That the independent variables represent two or more categorical groups
- Independence of observations
- The dependant variable for each categorical group following a normal distribution
- Homogenous variance of the dependant variable for each categorical group

In R the analysis of variance is carried out using the **aov**() function.

Questions

Suppose that you work for an online charity. The charity collects donations through a special page on its website. Prior to asking users to enter their credit card information the charity shows them a special informational webpage that highlights how their donation will be used.

Over time the charity has experimented with different designs for this page to increase the average amount donated.

The table below shows the amount (in GBP) donated across four different versions of the special informational webpage obtained by a fair experiment over the last 24 hours.

Version W	Version X	Version Y	Version Z
7	15	5	5
6	5	2	10
4	9	5	7
8	5	25	2
1	12	18	5
2	16	35	7.50
17	12	10	9
15	16	40	12
12	22	25	15
1	12	15	5
7	8	5	25
9	7.5	10	14
4	10	5	5

Using the lecture slides from LW2 as a guide, attempt to answer the following questions.

Qnum	Question	ANSWER
5	Input this data into R and determine the average amount donated by users having seen each of the four different versions of the informational page	
6	Plot the above data using a box and whisker diagram.	
7	Carry out an analysis of variance using the that from the above experiment to determine whether one of the pages is more effective by the average amount donated.	
8	Discuss the results of your test and if necessary, carry out further statistical tests to identify which specific page is more successful in attracting higher average donations.	

Questions

Qnum	Question	ANSWER
9	Suppose a search engine provides 15 results for a given query but only 3 are commonly clicked by search users. What is the search engines precision in this case?	
10	When designing a search engine algorithm, how might the practitioner test the search engines recall performance?	
11	The query parser is an important component of modern search engines. Suppose the phrase “cheap holidae ticket” was entered into a search engine. How might the query parser interpret this phrase?	
12	Give **THREE** specific examples of techniques search engines might use to determine the usefulness of a site to a given search query.	
13	Explain why the use of probabilistic informational retrieval models might bias the search results. What are the potential countermeasures?	