MSc Web and Social Media Analytics

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 how KPIs (Key Performance Indicators) are created and applied in practice**

LEARNING OUTCOMES

- **By the end of this tutorial students will have;**
  - **Made use of different R functions to manipulate and analyse data**
  - **Proposed a set of KPI given for a given situation.**
  - **Formulated A/B Tests**
  - **Performed the Fishers Exact Test in R**
  - **Performed the ANOVA in R**
  - **Calculated the margin of error of an online poll**

RESOUCRES AND TOOLS REQUIRED

- **R or RStudio**
- **LW2 lecture slides**
- **LW1 and LW2 tutorial packs**

Today we are going to explore the R programming language and develop skills in analysing user activity data. R is a very powerful open-source statistical package that can be used to analyse data from a variety of sources and in different formats. You can install R for free and use it with no restrictions. Both R and RStudio are available through appsanywhere.westminster.ac.uk

IMPORTANT:

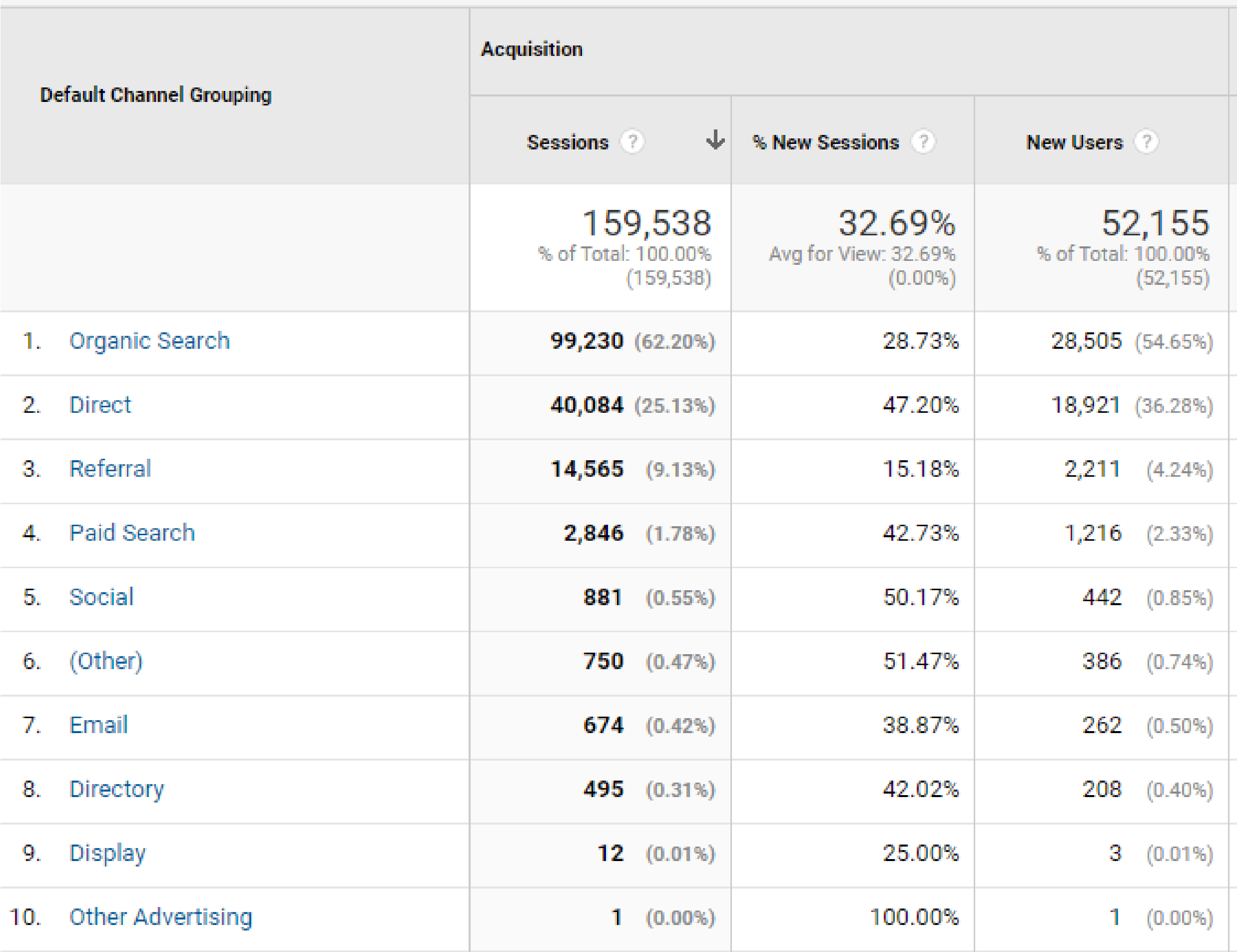
The pack is designed for you to go at your own speed and gets progressively more difficult. At the end of each section there is a series of practice questions – you should attempt to answer **all questions**.

Any questions you do not complete today should be completed before your next tutorial.

Study Notes

In LW2, we discussed the importance of KPIs (Key Performance Indicators) as part of the process of conducting web analytics.

The following is a set of three reports generated by Google Analytics for the University's Website.



Default Channel Grouping	Acquisition		
	Sessions ↓	% New Sessions	New Users
	159,538 % of Total: 100.00% (159,538)	32.69% Avg for View: 32.69% (0.00%)	52,155 % of Total: 100.00% (52,155)
1. Organic Search	**99,230** (62.20%)	28.73%	28,505 (54.65%)
2. Direct	**40,084** (25.13%)	47.20%	18,921 (36.28%)
3. Referral	**14,565** (9.13%)	15.18%	2,211 (4.24%)
4. Paid Search	**2,846** (1.78%)	42.73%	1,216 (2.33%)
5. Social	**881** (0.55%)	50.17%	442 (0.85%)
6. (Other)	**750** (0.47%)	51.47%	386 (0.74%)
7. Email	**674** (0.42%)	38.87%	262 (0.50%)
8. Directory	**495** (0.31%)	42.02%	208 (0.40%)
9. Display	**12** (0.01%)	25.00%	3 (0.01%)
10. Other Advertising	**1** (0.00%)	100.00%	1 (0.00%)

Figure 1 Acquisition Report

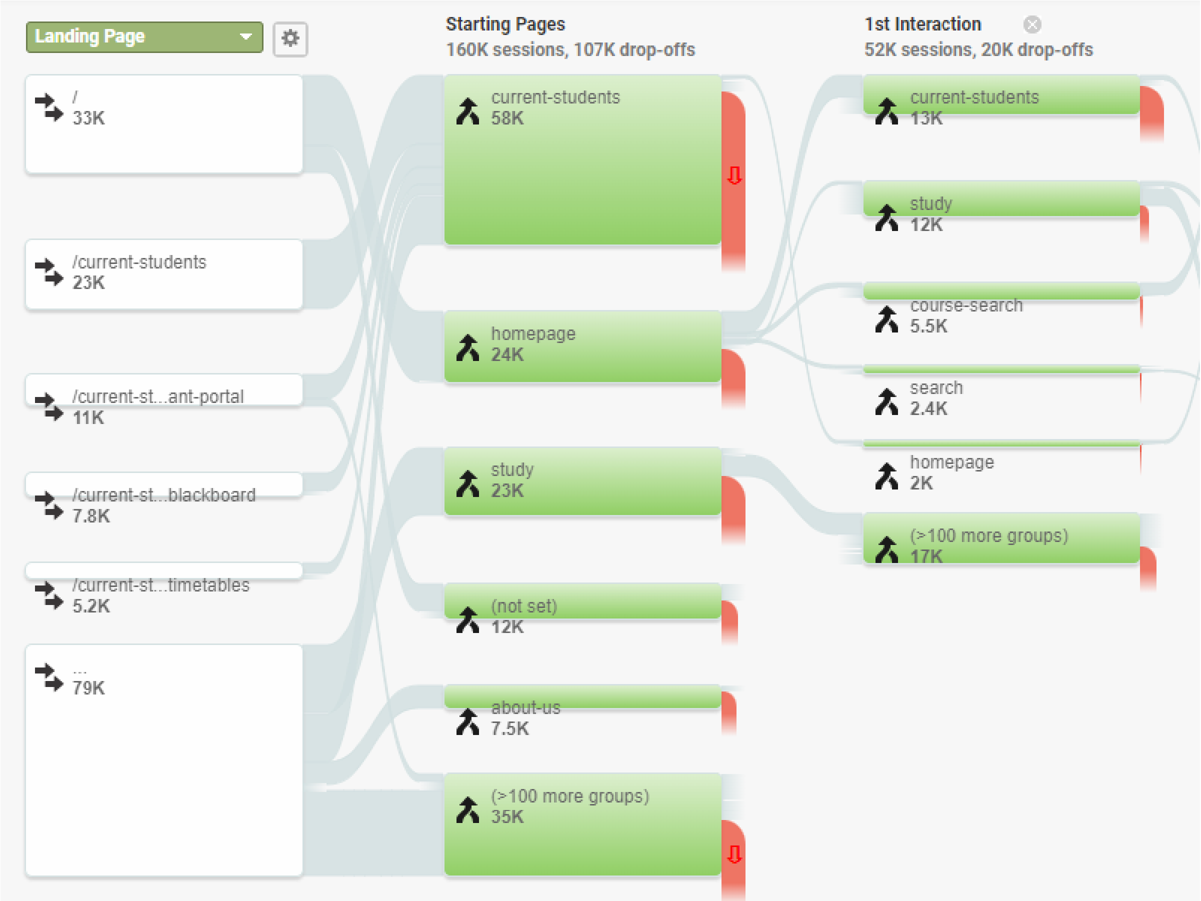
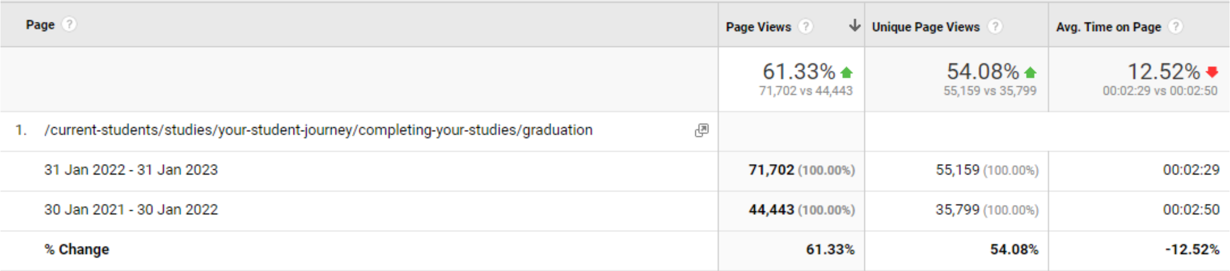


Figure 2 Behaviour Funnel Report



Page	Page Views	Unique Page Views	Avg. Time on Page
	61.33% ▲ 71,702 vs 44,443	54.08% ▲ 55,159 vs 35,799	12.52% ▼ 00:02:29 vs 00:02:50
1. /current-students/studies/your-student-journey/completing-your-studies/graduation			
31 Jan 2022 - 31 Jan 2023	**71,702** (100.00%)	55,159 (100.00%)	00:02:29
30 Jan 2021 - 30 Jan 2022	**44,443** (100.00%)	35,799 (100.00%)	00:02:50
% Change	**61.33%**	**54.08%**	**-12.52%**

Figure 3 Engagement Report

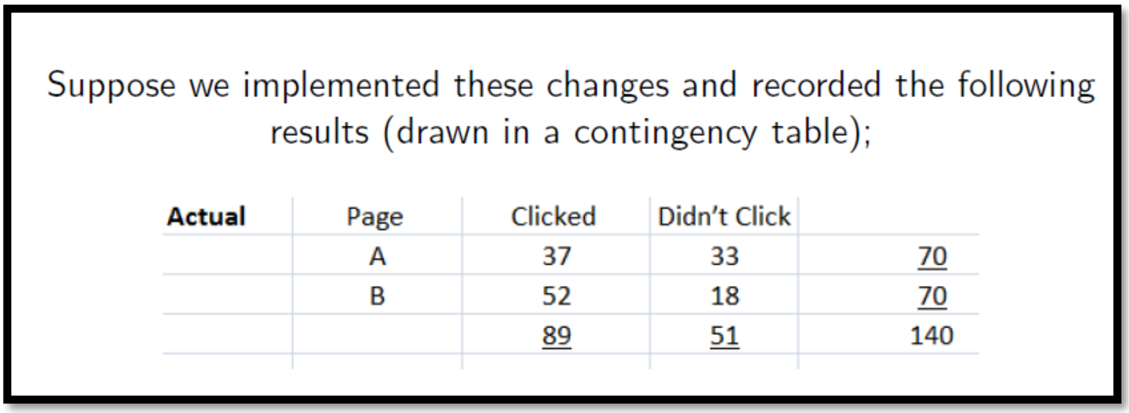
Questions

Use your knowledge of the University's long-term objectives and goals to answer the following questions.

Qnum	Question	ANSWER
1	Study the reports carefully. Create three well-defined KPIs that you believe are relevant to the University considering the above reports.	
2	For one of your KPIs, design an experiment that could be used to measure the impact of any potential changes. In your answer you **SHOULD** state your reasoning, how you will ensure the experiment is fair, what metric you are trying to improve and what statistical test you could use.	

In the LW2 lecture we explored used the Fishers Exact Test (FET) to analyse the results of an A/B test. In this case, we wanted to know which page was more optimal in driving a greater number of clicks to the weekend closures page.

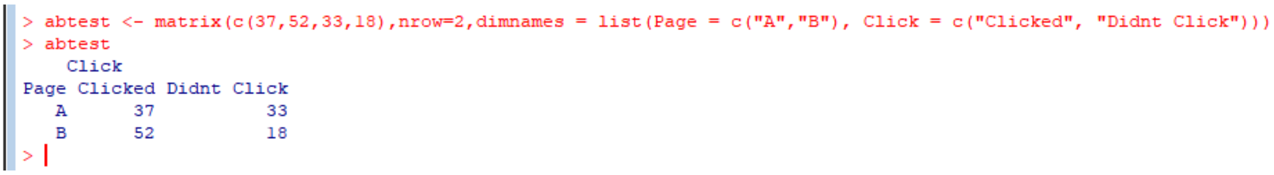
The results of that experiment are shown in the contingency table below.



Suppose we implemented these changes and recorded the following results (drawn in a contingency table);

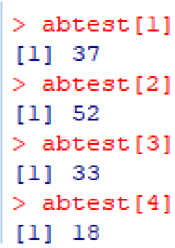
Actual	Page	Clicked	Didn't Click	
	A	37	33	70
	B	52	18	70
		89	51	140

To conduct the test in R we can use the **fisher.test**() function. The FET function in R expects to receive the data to test within a matrix, thus before we can conduct the test the results have to be entered in the correct form as shown below.



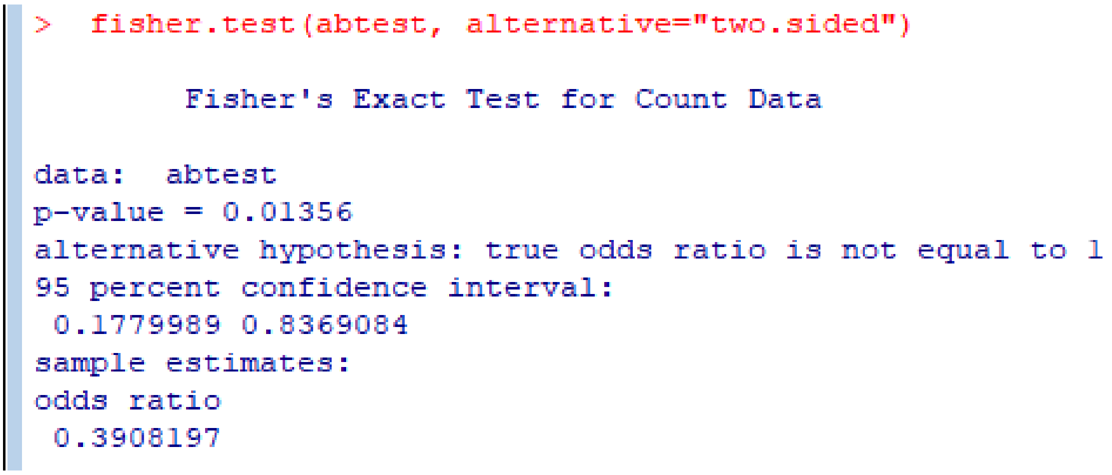
```
> abtest <- matrix(c(37,52,33,18),nrow=2,dimnames = list(Page = c("A","B"), Click = c("Clicked", "Didnt Click")))
> abtest
    Click
Page Clicked Didnt Click
   A      37          33
   B      52          18
> 
```

As with vectors, individual elements within a matrix can be accessed using index notation.



```
> abtest[1]
[1] 37
> abtest[2]
[1] 52
> abtest[3]
[1] 33
> abtest[4]
[1] 18
```

The fisher.test() function can now be applied using the variable “abtest” as the first parameter. Furthermore, we specify that a two-sided test should be performed, since we want to test both the case that page B is better or indeed worse than page A in terms of clicked generated.



```
>  fisher.test(abtest, alternative="two.sided")

        Fisher's Exact Test for Count Data

data:  abtest
p-value = 0.01356
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.1779989 0.8369084
sample estimates:
odds ratio
 0.3908197
```

The p-value indicates the probability that we could have observed this set of results under the assumption that the page viewed has no impact on the tendency of users to click on the link to the planned closure page. This is the same as saying that the odds ratio is equal to 1. If the odds ratio for two events occurring is 1 then by definition, they must have the same probability of occurring.

Since the estimated odds ratio is lower than 1, the probability of clicking on the weekend closure page having viewed page A is 39% lower than if the person had previously viewed version B. Thus, based on the p-value (<=0.05) and the odds ratio (<1) we would favour version B over version A.

Odds, Odds ratio and probability

Suppose that there is a 25% chance of rain tomorrow. This is the same as saying that the odds of it raining are 3:1 since (hence a probability of 1/(3+1) => 25%) or 1 chance in 4.

Odds ratios on the otherhand are used to compare the relative change in the odds (also by definition the probability) of an event under different circumstances. If odds of event occuring in scenario A is 5:1 and the odds of the same event occuring under scenario B is 10:1 then the odds ratio between these scenarios is (1/6)/(1/11) => 1.83. The probability of the event occuring under scenario A releative to B is 1.83 times greater.

The **chi-squared test**, like the fisher's exact test, is another example of a non-parametric test that can be used to test the relationships between categorical variables. Although it can be used with small sample sizes it tends to provide more reliable results than the fishers exact test at intermediate and large sample sizes. In R the chi-squared test is run by invoking the chisq.test() function.

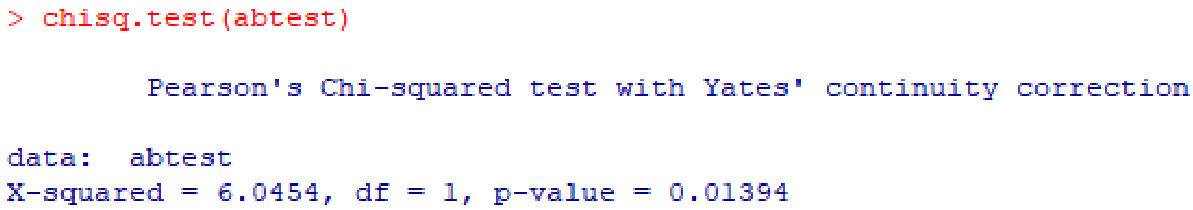
Like the FET, the chi-squared test should only be used if the following assumptions are satisfied.

- The input and output variables are categorical (e.g. Clicked/Didn't click)
- The observations are independent of one another (e.g. The people in the experiment don't influence each other's choices)
- Rows in the contingency table are mutually exclusive (e.g. A person saw either page A or B but not both)
- There is a minimum of 5 observations in each cell of the contingency table
- The null hypothesis of the test is that the variables are not correlated with one another

Non-parametric statistical tests

A statistical test used to determine whether there is enough evidence to reject an underlying assumption surrounding a process or event. Parametric statistical tests, in contrast to non-parametric statistical tests, make much stronger assumptions about the distribution from which data gathered originates. If their assumptions can be satisfied, parametric tests tend to enable more reliable conclusions being drawn. In many cases we may not know of indeed have enough data to asssume the distribution of a set of data and hence must resort to using non-parametric tests.

The results of the chi-squared test applied to our original problem are shown below. As you can see the p-value is lower that our 5% significance level, hence we conclude that there is sufficient evidence to reject the null hypothesis that the version of the page viewed is independent of whether the user clicks on the planned engineering works link.



```
> chisq.test(abtest)

        Pearson's Chi-squared test with Yates' continuity correction

data:  abtest
X-squared = 6.0454, df = 1, p-value = 0.01394
```

Questions

Using your knowledge of R and statistical tests answer the following questions.

<table>
<tr><th>Qnum</th><th>Question</th><th>ANSWER</th></tr>
<tr><td>3</td><td>Suppose an online store is experimenting with offering a $5 discount to new customers on orders over $50. The aim would be to test whether there is a significant increase in the number of orders.

Formulate and justify the use of an A/B test and state how you will ensure that the experiment is carried out fairly? What, if any, assumptions would need to be satisfied?</td><td></td></tr>
<tr><td>4</td><td>Suppose you carried out the experiment and obtained the following results:

<table><tr><td></td><td colspan="2">Placed Order?</td></tr><tr><td></td><td>Yes</td><td>No</td></tr><tr><td>Not Given discount</td><td>38</td><td>142</td></tr><tr><td>Given Discount</td><td>46</td><td>134</td></tr></table>
Input this problem in R and carry out an appropriate statistical test to determine whether a relationship exists.</td><td></td></tr>
<tr><td>5</td><td>Does your conclusion change if a different statistical test is used?</td><td></td></tr>
<tr><td>6</td><td>What are the odds of placing an order under the two different scenarios.</td><td></td></tr>
<tr><td>7</td><td>What is the odds ratio of placing an order when a discount is given versus when no discount is given?</td><td></td></tr>
</table>

Our previous experiment collected data on users' behaviour indirectly. In the LW2 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.

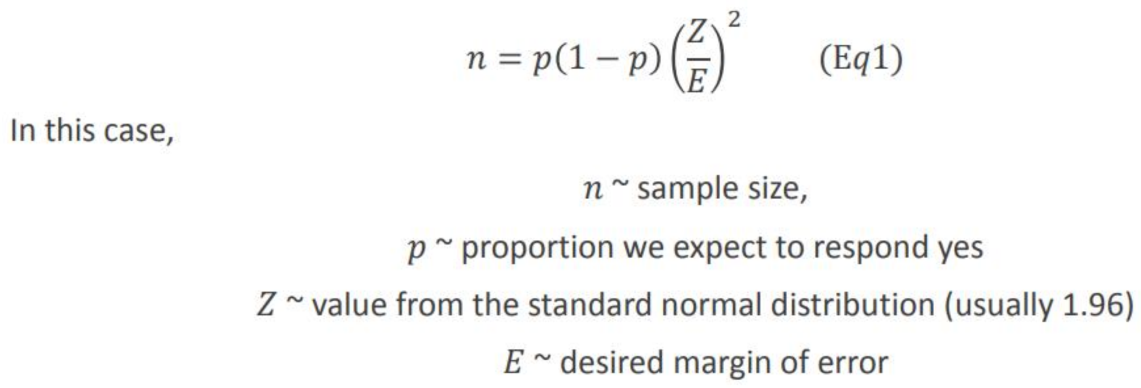
Let's consider an ecommerce website uses an algorithm to determine which products it should recommend to users based on their previous purchase history. Suppose the ecommerce website updates its algorithm and wants to find out if users find its recommendations more useful. One way could be to carry out 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.



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

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
8	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%.	
9	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.

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

By rearranging (Eq1)

$$1 = \frac{p(1-p)\left(\frac{Z}{E}\right)^2}{n} \qquad (\mathrm{E}q2)$$

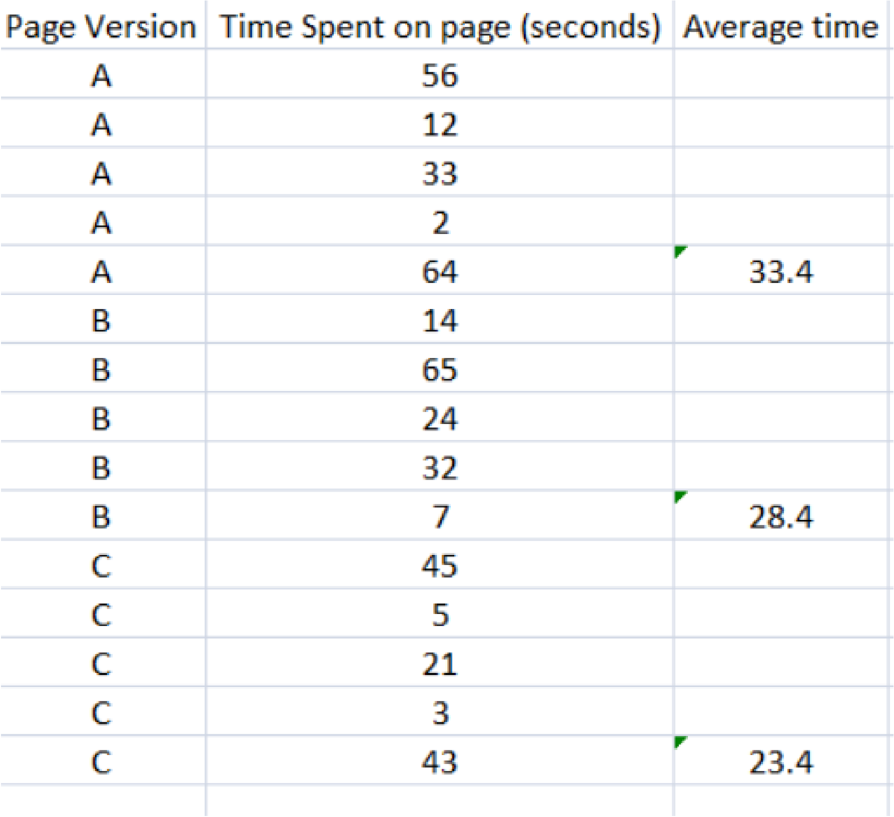
$$1 = \sqrt{\frac{p(1-p)}{n}}\frac{Z}{E} \qquad (\mathrm{E}q3) \quad \Rightarrow \quad E = \sqrt{\frac{p(1-p)}{n}}Z \ \Rightarrow \quad (\mathrm{Eq}4)$$

Questions

Using your knowledge polls, answer the following question.

Qnum	Question	ANSWER
10	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

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

Qnum	Question	ANSWER
11	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	
12	Plot the above data using a box and whisker diagram.	
13	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.	
14	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.	