**The Scenario**  
  
The US federal election is five months away. You have just been hired as a member of the Democratic Party, as a Data Analyst intern, in charge of optimizing the methods in which the campaign is run.   
  
Traditionally, whenever a web-user visits the website, they are asked for demographic data (party affiliation, gender etc.). An email is then sent to the user, with the following slogan: “It’s Time for a Change” as the heading in the body of the message. The email campaign has received mixed reviews, and it’s effectiveness has been brought into question. Recently one of the campaign managers has suggested two other slogans in the body of the email:   
  
Variant 2: “When the Time Comes, Vote Democratic”  
Variant 3: “Democrats for a Change”  
  
With the campaign running low on funds, you have been asked to combine your knowledge of web-architecture with A/B testing to evaluate the overall email campaign, and decide which slogan is best.   
  
The ‘Session 4 AB Testing Lab (Election)’ file contains data on 15774 users who have visited the website, as well as the user data which users have entered. The following data is available:  
  
Gender: The gender the user reports when they visit the site.  
Age: The age group the user reports when they visit the site.

Variant: The slogan the user is shown (Variant 1, 2, 3).

Party: The party which the user reports they support when they visit the site.  
Contribute: Whether the user contributes any funds to the campaign.

ContributionAmt: If the user contributes to the campaign, the dollar amount they contribute.  
EmailSubscribe: If the user signs up to receive email updates on the campaign.

ShareLink: If the user shares the website link on social media.

ShareMedium: If the user shares the website link, which social media outlet do they use.

派對：用戶在訪問網站時報告他們支持的派對。

Contribute：用戶是否為活動貢獻了任何資金。

ContributionAmt：如果用戶為活動做出貢獻，則他們貢獻的美元金額。

EmailSubscribe：如果用戶註冊接收有關活動的電子郵件更新。

ShareLink：如果用戶在社交媒體上分享網站鏈接。

ShareMedium：如果用戶分享網站鏈接，他們使用哪個社交媒體渠道。

**You are tasked with answering the following questions:**

1. Using ShareLink, as the measure of ‘best’ (someone sharing is better than someone not sharing), figure out which variant is the best variant to use. Should the campaign use a different variant as opposed to staying with Variant 1?
2. Would your answer differ if the benchmark for our p-value was 1% instead of the default 5%?
3. With your answer for Question 1, is party affiliation a confounding/lurking variable? Justify your answer, and support your answer with data.
4. Would your answer to Q3 differ if the benchmark for our p-value was 1% instead of the default 5%?
5. Using EmailSubscribe, as the measure of ‘best’, figure out which variant is the best variant to use.
6. Would your answer differ if the benchmark for our p-value was 1% instead of the default 5%?
7. Calculate the average amount a user donates for each variant. If our goal is the receive the most amount of money from users, which variant would we want to use? Why?
8. Is there a specific age group within the Variant you selected in Question 7, which should be targeted if we are wanting to receive the most amount of money? Note, we are not interested in the number of people within the segment who donate, but rather how much on average is donated. \*Hint, you might need to use Excel modeling here, as opposed to just using AB Testing.
9. One of the campaign coordinators suggests that it doesn’t matter how much someone donates, as long as they donate. The thinking is that down the line, someone who donates once will likely donate again. Based on this logic, if our goal is to get the most number of donations (not necessarily the highest amount), which variant should we choose?
10. Based on your answer in question 9, is party affiliation a confounding variable?
11. Based on your answers above, overall which variant do you believe is the best? Support your answer quantitatively (\*Hint – try to figure out which variant meets most of the criteria for the questions posed).
12. Your calculations show that if the wrong link is shared to a potential Democratic voter, there is a strong chance it could result in them changing their vote to Independent or Republican. If you are currently using a p-Value threshold of 5%, should you consider changing the threshold? If so, would you increase or decrease your p-Value threshold? Why?
13. Let’s assume you wanted to re-run your calculations to find which age group & gender combination would result in the highest donations for each variant. Does this calculation make sense, and would this be possible with A/B testing? If so, how? (You don’t need to actually calculate this, just comment on how you would do it if, and if it’s possible to do).
14. For you’re A/B tests in Question 1 and 5, if you found that there was an issue with the data collection process, and therefore a 3% margin of error in the data (i.e. each record has a 3% chance of being wrong), would this impact the acceptable p-value threshold you would use?
15. For Question 1 and 5, assume you were told of an additional uncounted 5000 completed surveys, which showed a similar average and standard deviation as the data you currently had. Would this impact your z-score? And therefore your p-value? In which way (would the z-score be higher and p-value lower? Vice-versa?)