Final Exam 2021 – Quantitative Methods for Business and Social Science (BPOLO2010E)

Home assignment (UC)

BSc in International Business and Politics

Copenhagen Business School

01 June 2021

Student Number:

Examiner: Zoltan Fazekas

Pages: 10 (33 with front page, references, and appendix)

Characters: 25.625

Reference System: APA 7th edition

Question 1: Unconscious biases

1.1 An implicit antipublic sector bias can have practical implications and lead to resource constraints for the public sector. Thus, experiments like these are needed to figure out how to correct this bias and change implicit attitudes. This paper shows that information only affects implicit attitudes in the short run, thus, the public sector should inform more often or try alternatives. This knowledge benefits political decision making, as they are often under budget constraints or time limits and need to know who to target. The need depends on the possibility of drawing causal conclusions. To do this, we need to control for confounders: Done through a Randomized Controlled Experiment (RCT), as it can isolate the effects of a treatment variable on the basis of randomization and having a counterfactual ("No Information"-group) (Imai, 2017a). The randomization of assigning subjects to treatment and control groups ensures that the groups are identical on all parameters except for the treatment. Causal conclusions can thus be made by comparing the control and treatment groups' performance rating of the USPS. Experiment I tests whether positive information decreases the impact of implicit attitudes on explicit attitudes. It also tests if information has short- or long-lasting effects and can thus illuminate the longitudinal effects. Experiment II examines whether USPS advertising overrides implicit attitudes and can thus help draw causal conclusions regarding information's effect on implicit attitudes on public sector evaluations (Marvel, 2016). They differ in amount of treatment groups; meaning that Experiment II also shows the effect of the type of information received (quantitative or qualitative). The IATs in the two experiments also differ: I uses stereotypes (fast/slow) and II uses preferences (good/bad), which adds value to both experiments, shown by the fact that the IAT results differ between these two experiments.

1.2 a) IATs measure associations by making individuals attribute 2 concept categories with 2 attribute categories as fast as possible. We can infer implicit attitudes by comparing times in the IAT. Implicit attitudes are deeply ingrained mental associations of a concept with one's feelings. They affect explicit attitudes and are not accessible to introspection, thus, to measure someone's implicit attitude, you need IATs. IATs can overcome social desirability bias, since they measure implicit and not explicit attitudes, which may be of importance in areas like the public sector where anti-public messages are common.

b) Internal validity is high if causal assumptions are satisfied, and this study has a great basis for causal conclusions as it uses randomization when assigning treatment, both information- and time-wise. Thus, looking at the research design, this study resembles great internal validity. External validity measures the extent to which the conclusions can be generalized beyond the study. This study has relatively small sample sizes and was subject to item non-response, where 2-day-lag-respondents in Experiment I did not rate USPS' performance, which limits the generalizability. Likewise, the use of MTurk creates sample selection bias, as their subjects have volunteered themselves (majority is male, white, educated, democratic, and liberal) and are paid to answer, which makes the sample non-random nor representative. The study however defends the use of MTurk based on the

samples being more diverse than similar samples. Likewise, since the subjects supposedly are more favourable towards the public sector, the results are conservative and thus more generalizable. Since the study looks for causal effects and not to describe population parameters, a more representative sample should not be crucial to the generalizability. Thus, this study's external validity is acceptable, as it is unlikely to generate false positives.
Reliability is the consistency of a measure, and the large standard deviations in this study indicate uncertainty and lack of consistency, as well as experiment I and III's mean IAT scores are relatively low, which suggest weak, implicit attitudes - This should though be seen in the light of the conservative results. The study itself is quite valid and could be performed again and again, suggesting acceptable reliability. All in all, the study has relatively high validity and acceptable reliability making the findings of the study both generalizable and causal.

1.3 In Experiment II, we see that for every 1-unit increase in the IAT score, we will on average see a decrease in the performance ratings of 0.2. This tells us that the more negative the implicit attitude is, the lower the USPS performance rating. This supports H1 that claims that individuals' implicit antipublic attitudes about the USPS will factor into their evaluations of the USPS' performance. This coefficient is however not statistically significant, as p-value $= 0.551 > \alpha = 0.05$, and the standard error, 0.34, is greater than the coefficient in absolute terms. The Information and Advertising coefficients, 1.20 and 0.51, shows the average treatment effects of the two treatment groups compared to the control group. They are both statistically significant and have relatively low standard errors. Thus, these coefficients tell us that if one is in the Information treatment group (a binary variable, thus 1-unit increase), then one's USPS performance rating will increase by 1.20, and if one is in the Advertising treatment group, we associate a 0.51-unit increase in the USPS performance rating on a scale of 0-7. Both interaction terms are negative, $Information * IAT$ being -0.13 and $Advertising * IAT$ -0.70, which suggest that being in either of the treatment groups and increasing the IAT score by 1 unit is associated by a decrease in the performance rating of USPS. This counter H2 that claims that favourable performance information will decrease the influence of Implicit Attitudes when evaluating USPS' performance. The counter-intuitive findings in Table 3 prompted the researchers to reconsider the variables used. Importantly, looking at the standard errors, 0.5 and 0.44, which are large considering the coefficients, and the p-values, p=0.79 and p=0.12, we see that these findings are not statistically significant. The $R^2 = 0.17$ is however of acceptable size in this model, which means that the predictors explain 17% of the variance in the outcome variable.

Question 2: Self-reported COVID-19 hygiene-relevant routine behaviours

2.1 The nationally representative data is used for a randomized, controlled survey experiment (here a subset sample size: $n = 442$). I use a two-tailed, one-sample hypothesis test using H_0 = The proportion of women in the high anchor group is 0.48, $H_0 = 0.48$ and H_a = The proportion of women in the high anchor group is different from 0.48, $H_a \neq 0.48$. Our test statistic is the sample proportion of women in the high anchor group: $\bar{X}_n = \frac{Sample\ prop.}{Sample\ size} = \frac{232}{442} = 0.525$ and I use a 95%-confidence level. Using the central limit theorem, the reference distribution properties under the H_0 is $\bar{X}_n \sim N(0.48, \frac{0.48(1-0.48)}{442})$. First, I calculate the standard error (SE) under H_0:

$$SE = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.48(1-0.48)}{442}} = 0.024.$$

Next, I find the z-score for our sample estimate ($\bar{X}_n = 0.525$):

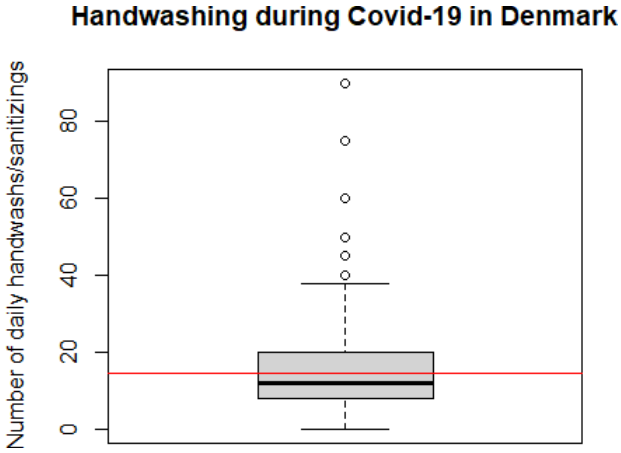
$$\text{z} - \text{score} = \frac{\bar{X}_n - \mu_0}{SE} = \frac{0.525 - 0.48}{0.024} = 1.889$$

As the z-score 1.889 is smaller than the critical value 1.96, it is within the normal distribution and we retain H_0. Likewise, looking at the p-value, which is the probability of observing a 52.5% proportion of women in the high anchor group: $p - value = 2\phi(-|1.889|) = 0.059$, we retain our H_0, as p-value$=0.059 > \alpha = 0.05$. This can also be seen by looking at the 95% confidence interval [0.478, 0.571], as $H_0 = 0.48$ is within this interval. This means that during repeated data generating processes, 95% of the time this confidence interval would bracket the true value parameter, the average proportion of women. With an α level of 0.1, however, we fail to retain the null hypothesis H_0, as $p = 0.059 < \alpha = 0.1$. In this case, the 90%-confidence interval becomes [0.486, 0.564] not including $\mu_0 = 0.48$. This conclusion would also be made if comparing the z-score 1.889 to the critical value, which is 1.64 with a 90%-confidence level, as it indicates that our observed value is outside the 90%-confidence level normal distribution. These step-by-step results are identical to the output from a two-sided R proportion test if we do not apply Yates' continuity correction.

<u>2.2</u> The age variable is measured on a continuous interval, so I use a t-distribution test. Using the difference-in-means estimator as test statistic gives the following hypotheses: H_0 = The average age in the low anchor group does not differ between government and opposition supporters, $\mu_0 = \mu_1$, and H_a = The average age of government and opposition supporters differ, $\mu_0 \neq \mu_1$. There is a difference between the age means, 54.2 and 55.5, but it is not statistically significant since the p-value$= 0.49 > \alpha = 0.05$. This is also seen looking at the 95% confidence interval [-4.763, 2.284] as it contains zero, meaning we retain the H_0 of zero average treatment effect. Our conclusion will not change using an $\alpha = 0.01$, since our p-value is still significantly greater than the level of test $p = 0.49 > \alpha = 0.01$. Thus, the mean age of government supporters is not statistically, significantly different from the mean age of opposition supporters in the low anchor group.

<u>2.3</u> Outcome_handwash is a continuous interval variable and shows the number of handwashes the day prior. The range is 90 based on 884 observations. The median is 12 (black line), the mean is

14.64 (red line), and the central tendency and distribution is visualized in the boxplot. The boxplot shows an interquartile range of 12, which shows a strong central tendency around the mean, and that there are several outliers in the higher end of the range. This interquartile range is small considering there are two different treatments and the range of 90. This boxplot underlines the great spread of the observations in the data. The SE of the mean shows the degree to which we expect the mean to deviate from its expected value:



$$SE\ of\ sample\ mean = \frac{Standard\ deviation}{\sqrt{n}} = \frac{9.907}{29.732} = 0.333$$

The SE of the mean is 0.333, which is quite small considering the mean=14.64 and the range of 90. Standard deviation measures the spread of the distribution; thus, on average, the number of handwashes is 9.9 away from the mean. The 95% confidence interval is:

$$CI_{95\%} = [\bar{X}_n - z_{a/2} * SE, \quad \bar{X}_n + z_{a/2} * SE\ ]$$

$$CI_{95\%} = [14.64 - 1.96 * 0.333, \quad 14.64 + 1.96 * 0.333\ ] = [13.987, 15.292]$$

which indicates that in repeated data generating processes, 95% of the time this confidence interval $[13.987, 15.292]$ would bracket the true mean of number of handwashes. A univariate summary plot of the two anchor treatment groups can be seen in Appendix A, which sets the stage for 2.4.

2.4 A two-sample Student's T-test based on H_0 = The difference between treatment group means is 0, $\mu_0 = \mu_1$, shows that the means of the two anchor groups are 18.43 (high anchor) and 10.86 (low anchor). The test shows that the true difference-in-means is not equal to 0, as the p-value < 0.01 < $\alpha = 0.01$. Similarly, the 95%-confidence interval [6.36, 8.78] does not bracket 0, underlining the rejection of H_0. A linear regression model shows the same results (see table). We expect one in the low anchor treatment group to report 10.68 handwashes and increasing the treatment variable by one unit, meaning switching to the high anchor group (High anchor = 1, low anchor = 0), increases the expected outcome with 7.57 to 18.43 $(7.57 + 10.86)$. This treatment coefficient matches the confidence interval found in the t-test: [6.36, 8.78]. The coefficient and intercept have relatively low SEs and are both statistically significant, which matches our findings in the t-test. In conclusion, the anchor made a difference for the self-reported handwashing count, which is

	Dependent variable:
	outcome_handwash Handwash
Treatment(High anchor)	7.57*** (0.62)
Constant	10.86*** (0.44)
Observations	884
R^2	0.15
Adjusted R^2	0.15
Note:	*p<0.1; **p<0.05; ***p<0.01

also the conclusion made in the research paper (Hansen et al., 2021). This anchor bias is an important finding and should be considered when doing surveys, since political decisions often are based on survey experiment data.

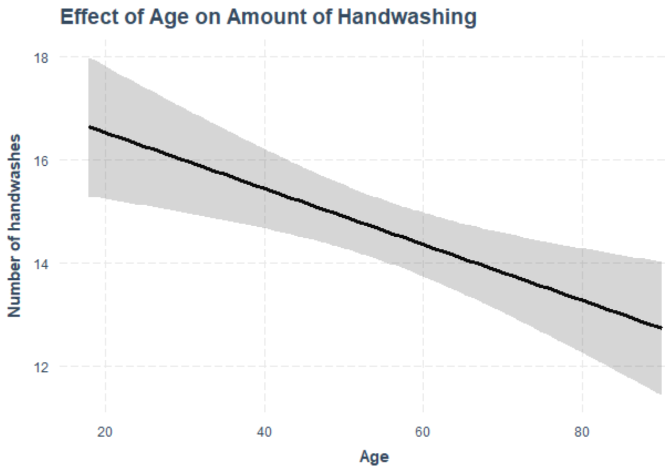
2.5 I expect the effect to be small since the anchor treatment only indirectly points in a direction of an appropriate close contact number. The outcome may also be affected by respondents being more aware of their close contacts than hand hygiene because of the social distancing restrictions during the time of the survey. Conducting the same two types of test as in 2.4 with a 95%-confidence level, I can reject H_0= The difference in means between the low and high anchor treatment groups regarding close contacts is 0. This is because the p-value$= 0.02 < \alpha = 0.05$, the confidence interval [0.25, 3.18] not containing 0, and the statistical significance of the intercept (p<0.01) and the Treatment Coefficient (p<0.05). It is noticeable in the linear regression model that the SE of the Treatment coefficient is significant compared to the coefficient itself. Likewise, the model has a very small R^2 which means that the treatment variable only explains 1% of the variance in the outcome variable, number of close contacts. In conclusion, the mean close contact score is generally lower for the low anchor group, 6.71, compared to the high anchor group, $8.42 (= 6.71 + 1.71)$, partly because of the anchor treatment. This supports my expectation of the scope of the anchor treatment's effect on one's number of close contacts.

	Dependent variable:
	outcome_closecontact Close Contact
Treatment(High anchor)	1.71^{**} (0.74)
Constant	6.71^{***} (0.53)
Observations	884
R^2	0.01
Note:	*p<0.1; **p<0.05; ***p<0.01

2.6 Extending the model from 2.4 to a multiple linear regression model results in the coefficients on the right. This causes the anchor treatment variable to increase slightly and the intercept to increase significantly (10.86 to 14.91). Both are statistically significant with p-values below 0.01 and relatively low SEs (around 8%). The Sex predictor is binary, and the coefficient represents the change in number of handwashes if the subject is male; on average males would self-report -2.64 less handwashes than females. The government support coefficient tells us that if the respondent supports the government (government supporter = 1), then the expected outcome increases by 0.26. This is however not statistically significant since the SE is close to three times the size of the coefficient (0.62) and the p-value> 0.1. The Age coefficient is negative, which means that increasing one's age by 1 year results in a decrease in handwashes by -0.05. Since the age variable is continuous and the range is large (18-90) this means that the oldest respondents,

	Dependent variable:
	outcome_handwash Handwash
Treatment(High anchor)	7.62^{***} (0.61)
Age	-0.05^{***} (0.02)
Sex	-2.64^{***} (0.61)
Gov. support	0.26 (0.62)
Constant	14.91^{***} (1.13)
Observations	884
R^2	0.18
Adjusted R^2	0.17
Note:	*p<0.1; **p<0.05; ***p<0.01

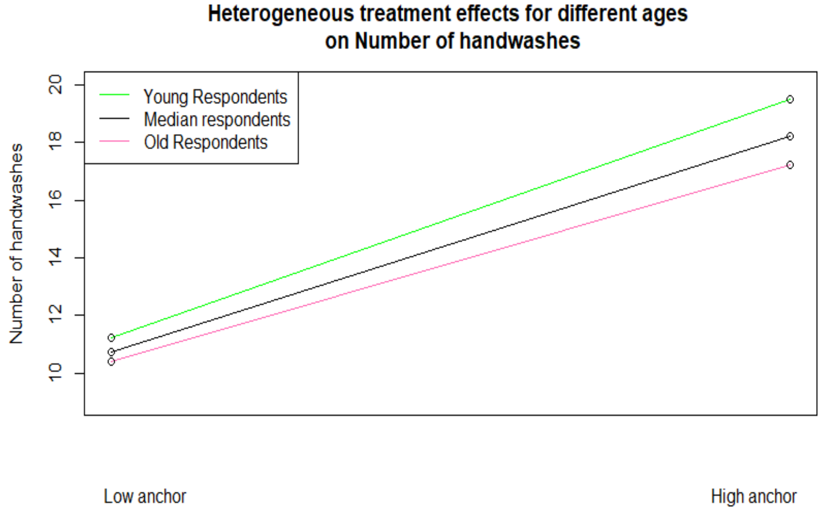
ceteris paribus, wash their hands -3.6 $(= 90 - 18 * (-0.05))$ times less than respondents at the age of 18 (See graph). The 95%-confidence drawn in grey shows that there is a statistically significant relationship (as one cannot make the solid line horizontal within the 95%-confidence interval space), which is backed up by the Age coefficient being statistically significant ($p < 0.01$). The R^2 is 18%, which means that the model can now explain 18% of the variance in the outcome variable compared to the 15% in the simpler model. Looking at the adjusted $R^2 = 1 - \frac{(1-R^2)(n-1)}{n-p-1}$ (which accounts for the number of predictors through the degrees of freedom correction), we see an increase as well (15% to 17%), which means that the model's new predictors help explain more variation in the outcome than without them.



2.7 To examine heterogeneous treatment effects for different ages, I extend the model from 2.6 to a multiple linear regression model with an interaction term: $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \in$ to $Outcome\ Handwash = \alpha + \beta_1 * Treatment + \beta_2 * Age + \beta_3 * (Treatment * \ Age) + \beta_4 * Sex + \beta_5 * Gov.support + \in$. I can now examine the direction and magnitude of the treatment effect: β_3 represents the additional average treatment effect dependent on age and represents how the effect of the anchor treatment is conditional on age (Imai, 2017e).

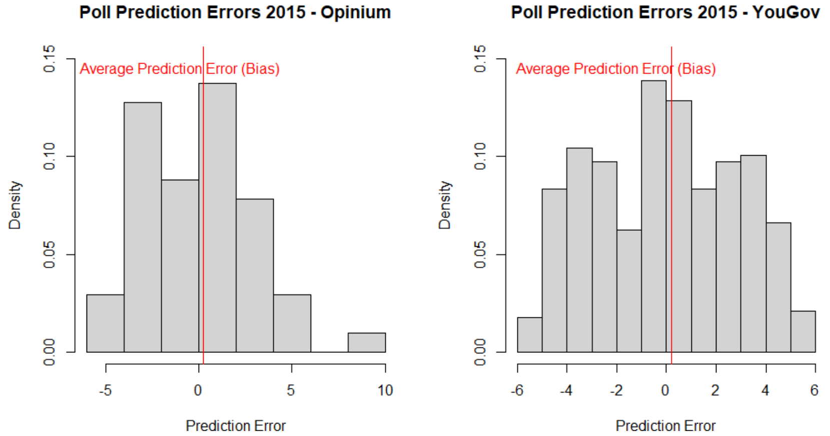
	Dependent variable:
	outcome_handwash Handwash
Treatment(High anchor)	10.39*** (1.92)
Age	-0.03 (0.02)
Sex	-2.66*** (0.61)
Gov. support	0.22 (0.62)
Treatment*Age	-0.05 (0.03)
Constant	13.61*** (1.42)
Observations	884
R^2	0.18
Adjusted R^2	0.17
Note:	*p<0.1; **p<0.05; ***p<0.01

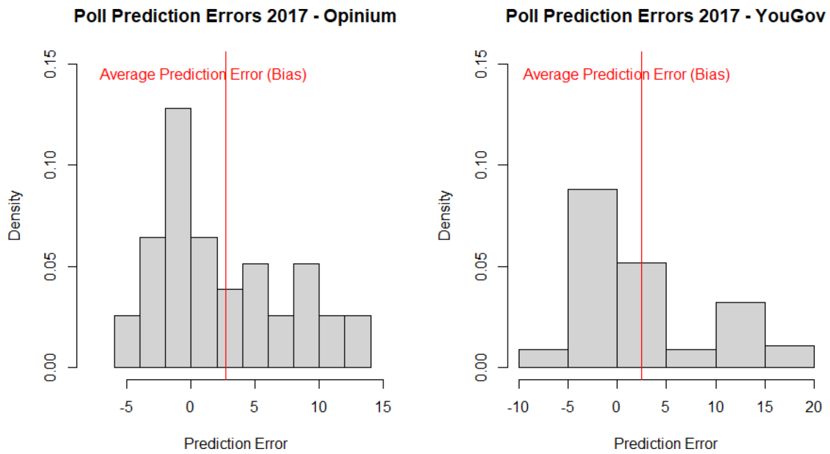
Compared to the model in 2.6, the intercept decreases to 13.61 and the Treatment coefficient increase to 10.39. This indicates that when the respondent is in the high anchor group, the expected number of handwashes increases with 10.39, ceteris paribus. It is noticeable that the SEs have increased (1.42 and 1.92) in this model, but that they are still statistically significant. The Age, Sex, and Gov.support coefficients are similar to before, but the Gov.support and now Age coefficient are not statistically significant. The interaction term Treatment*Age is negative, which tells us that if the respondent is in the high anchor group and we increase age by one 1 year, the expected number of handwashes decrease by -0.05. This should be seen in combination with the Age coefficient, since the average effect equals $\beta_2 + \beta_3 = -0.03 + (-0.05) = -0.08$, whereas the effect for the low anchor group equals $\beta_2 = -0.03$. The interaction term coefficient is however not statistically significant, which requires us to retain the null hypothesis saying that the coefficient is equal to zero, meaning that a change in these predictors cannot with certainty be associated with changes in the outcome variable. This matches the findings in the research paper, where they conclude that the marginal treatment effect is similar for all age groups (Hansen et al., 2021). The heterogeneous treatment effects for different ages on the number of self-reported handwashes is visualized in the plot, which supports the weak findings, since the treatment effect and slope for the different age groups is somewhat similar and without much variation.



Question 3: Polls and election results

3.1 To examine the accuracy of the two polling companies' predictions, I subset the data into elections and polling companies and find the average prediction error ($PE = Actual\ Outcome - Predicted\ Outcome$) aka the bias (Imai, 2017d). The 2015 histograms show that the poll PEs varies widely from one poll to another. Opinium's PEs are relatively small and larger PEs are less likely to occur. YouGov's distribution is more even, which suggest that larger PEs occur more often. The bias for Opinium's polls is 0.224 and YouGov's is 0.188 (red, vertical lines). This suggest that YouGov is slightly better at predicting the 2015 UK election than Opinium. In 2017, we see that the distribution of PEs has widened for both polling companies and that the election in general was harder to predict. In 2017, Opinium's bias equals 2.703 and YouGov's 2.503, which again suggests that the companies' biases are similar in scope and direction, but that YouGov seem slightly better.





YouGov has close to 6 and 3 times more polls than Opinium in 2015 and 2017, which may cause smaller PEs, since they can cancel each other out. Thus, I look at the Root-Mean-Squared Error (RMSE) that represents the average magnitude of the PE (Imai, 2017c): $RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} PE_i^2}$.

As expected, RMSEs for 2015 and 2017 show that Opinium, 2015: 2.972 and 2017: 5.930, in fact did better than YouGov, 2015: 3.078 and 2017: 7.441. This also matches the distribution of PEs in the graphs of Poll Prediction Errors for 2015 and 2017.

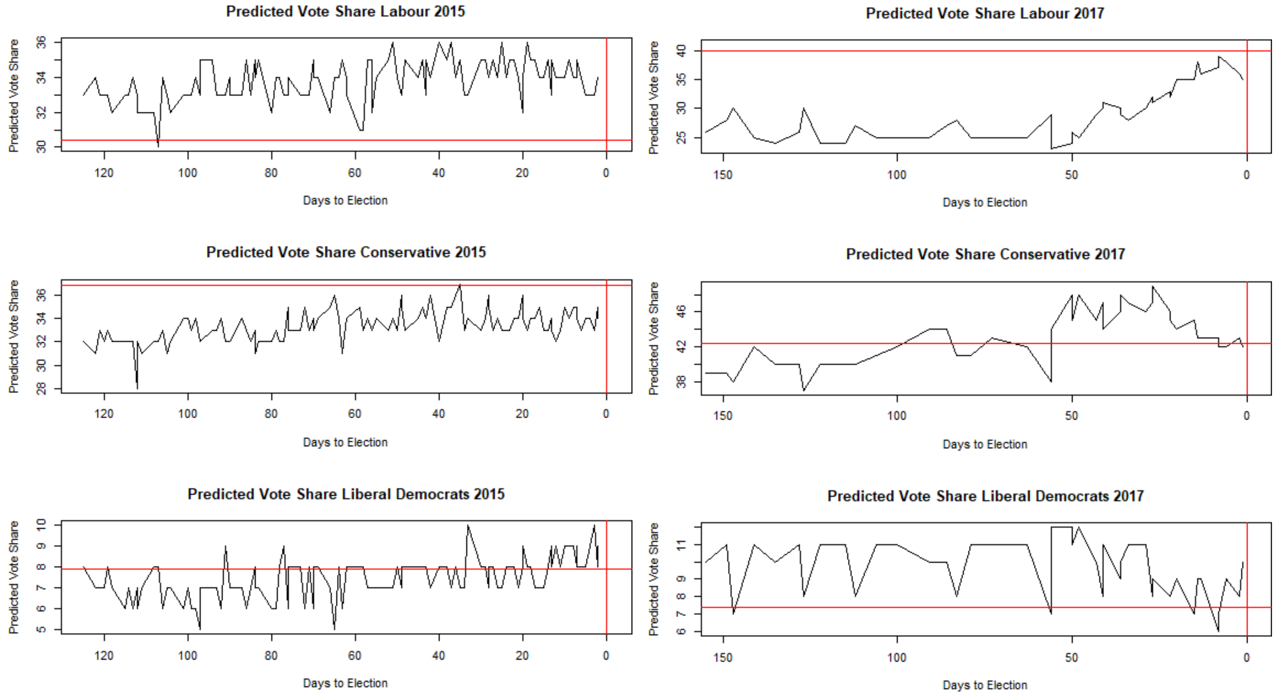
The most difficult parties to predict are found by looking at both polling houses at the same time. Parties with high biases and RMSEs indicate a high degree of uncertainty. As seen in the table, Conservatives, 3.785, was the most difficult party to predict in 2015, as the average magnitude of each poll PE on Conservatives is 3.785%-points. They are closely followed by Labour with 3.564%-points. Liberal Democrats have a smaller RMSE = 1.052, which may be caused by their significantly lower vote share possibly affected by the bipolar competitive landscape between Labour and Conservatives. In 2017, the most difficult party to predict was Labour (RMSE = 11.48), as Conservatives and Liberal Democrats had relatively low RMSEs (3.047 and 2.668). Compared to 2015, these RMSE are quite high. This greater uncertainty may be caused by the election being very close (40% and 42.4%) and the turmoil in the British political landscape caused by the Brexit-election in 2016. These findings also match the graphs found in 3.2.

Table 3.1		**Labour**	**Conservatives**	**Liberal Democrats**
2015	Bias	-3.379	3.528	0.431
	RMSE	3.564	3.785	1.052
2017	Bias	10.477	-0.668	-2.123
	RMSE	11.480	3.047	2.668

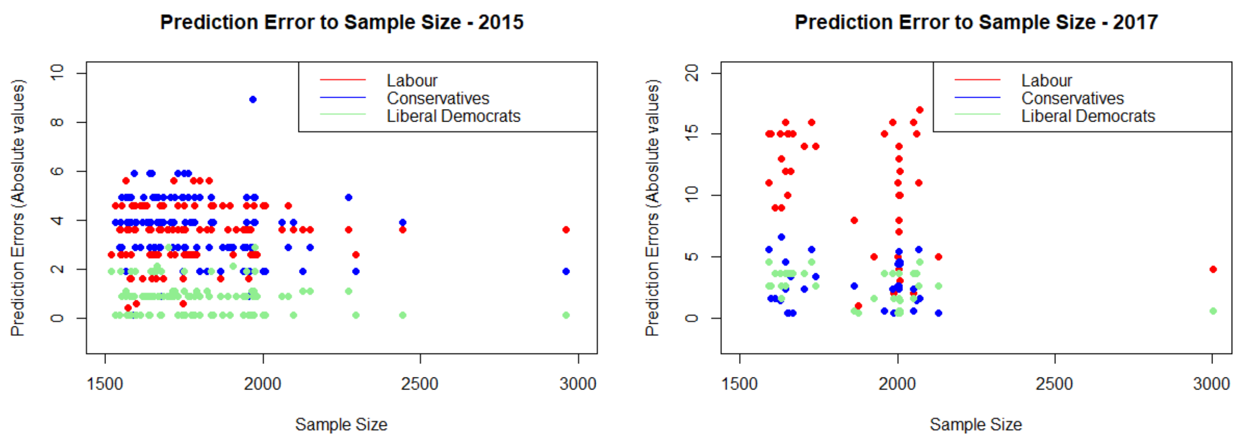
<u>3.2</u> To examine the exact quality of predictions, I look into the bias and RMSE of three periods leading up to election day. The timeline can be seen in the plot (next page), where the intercept of the red lines indicates the actual election result. The bias for the period leading up to the election shifts over time: the first $1/3$ in 2015 (125 days in total) have a bias of 0.946, the middle 0.133, and the last period leading up to the election a bias of -0.402 (See Table 3.2). This shows that the quality of the prediction changes closer to the election date, that the polls both over- and underestimate, and that there is more uncertainty closest to election in 2015 than in the middle period. This is backed up by the RMSE and suits the 2015 Predicted Vote Share graphs of the 3 election parties. Looking at 2017 (150 days in total), we see a different pattern based on the RMSE (8.381 < 8.535 > 5.573), where the middle period was the hardest to predict, and the period closest to election contains the predictions of highest quality. This matches one's expectations of the graphs on Predicted Vote Share 2017, where the latest polls seem close to the actual election results. Noticeably, the RMSE in 2017 does not follow the same pattern as the bias, which, as discussed in 3.1, indicates that the polls in 2017 predicted both too high and too low resulting in a low bias and high RMSE (e.g. "Middle

Table 3.2		**First period**	**Middle period**	**Period leading up to election**
2015	Bias	0.946	0.133	-0.402
	RMSE	3.345	2.853	2.977
2017	Bias	4.661	3.933	1.086
	RMSE	8.381	8.535	5.573

Period” for 2017 or the Predicted Vote Share Conservatives 2017 graph). Thus, the quality of the prediction, as seen on the much higher RMSEs and biases in 2017, and the periods with the highest quality predictions changed within and between the two years.



3.3 The Law of Large Numbers states that the sample average, here the prediction for the three parties, will converge towards the population average as the sample size increases (Imai, 2017b). In other words, we expect the PE to get closer and closer to 0, as the sample size increases: $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to E(X)$. The Sample Size variable ranges from 1520 to 3002, with a median of 1763 and mean of 1819. This suggests outliers (See graphs below). I have taken the absolute values of the PEs for the 3 parties each year and plotted them compared to their sample sizes. This makes it possible for me to look at the scope of PEs as well as look at the direction of the correlations. The sample size of the polls is more evenly distributed in 2015 than 2017, perhaps because of there being



only 44 polls in 2017 to 113 in 2015. The distribution of the three parties and their PEs also matches the findings in Table 3.1, where Liberal Democrats stand out in 2015 as easier to predict and where the 2017 Labour predictions were very uncertain. In 2015, the poll for Labour with the lowest PE only has a sample size of 1570, for Conservatives that number is 1583. For Liberal Democrats, there are 44 polls with the same low magnitude of prediction error, where the sample size varies from 1532 to 2960. These numbers thus indicate that there is no obvious relationship between low uncertainty and sample size. In 2017, for Labour, the lowest PE poll sample size is 1875. Conservatives' lowest PE poll sample sizes varies from 1651 to 2130. Liberal Democrats follow the same pattern as Conservatives with sample sizes ranging from 1875 to 2007. Remarkable here is that only once out of the six scenarios has a poll with the smallest PE been the one with the largest sample size, which is the last poll before election in 2015 for Liberal Democrats. When looking at the correlations between the absolute PEs and sample size, the negative correlations indicate that there is indeed a correlation between better performing polls and sample size, which means that a larger sample size is associated with a smaller PE. The magnitude of these correlations is however quite low (around -0.2), which indicates a weak relationship, which matches our findings when we looked at sample sizes above. The positive, but very weak correlation for Labour in 2015 actually indicates the opposite relationship as expected. The correlation for Labour and Liberal Democrats in 2017 are relatively strong correlations, which matches the pattern in the Prediction Error to Sample Size 2017 plot, as the 3002-sample sized poll have a low PE compared to the rest of the poll PEs regarding the same party with lower sample sizes. Correlation is not causality in an observational study, thus from these correlations I can simply observe that there seems to be a stronger association of better performance of polls to larger sample sizes in 2017 than in 2015, but also that the correlations point in many directions. The lack of consistency across elections may be caused by the days-to-election element, meaning that there should in theory be less uncertainty closer to the election date. Thus, if the polls with the smallest sample sizes are made in the beginning of one election period but not the other, then the correlation will differ between the two years biased by the timing of the polls. Another reason for lack of consistency could be the number of polls per period. The Law of Large Numbers indicate that a larger sample size, here the number of polls, will make the sample average converge towards the expectation. Thus, looking at our two years, 2015 would according to the Law of Large Numbers make a better indication about the actual correlation than 2017, assuming no timing-bias for either year, thus suggesting that in this prediction experiment, the correlation between polls with larger sample sizes and performance is generally relatively weak.

Table 3.3		**Labour**	**Conservatives**	**Liberal Democrats**
2015	Correlation	0.087	-0.204	-0.161
2017	Correlation	-0.396	-0.169	-0.452

Bibliography

Hansen, P. G., Larsen, E. G., & Gundersen, C. D. (2021). Reporting on one's behavior: a survey experiment on the nonvalidity of self-reported COVID-19 hygiene-relevant routine behaviors. *Behavioural Public Policy*, 1–18. https://doi.org/DOI: 10.1017/bpp.2021.13

Imai, K. (2017a). Causality. In *Quantitative Social Science: An Introduction* (pp. 32–66). Princeton University Press.

Imai, K. (2017b). Large sample theorems. In *Quantitative Social Science: An Introduction* (pp. 300–306). Princeton University Press.

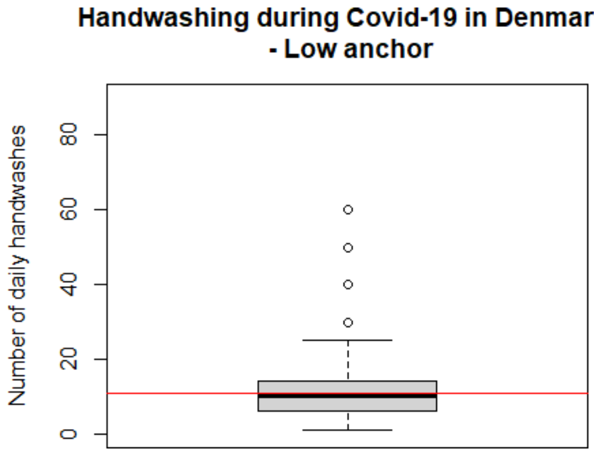
Imai, K. (2017c). Linear Regression. In *Quantitative Social Science: An Introduction* (pp. 139–161). Princeton University Press.

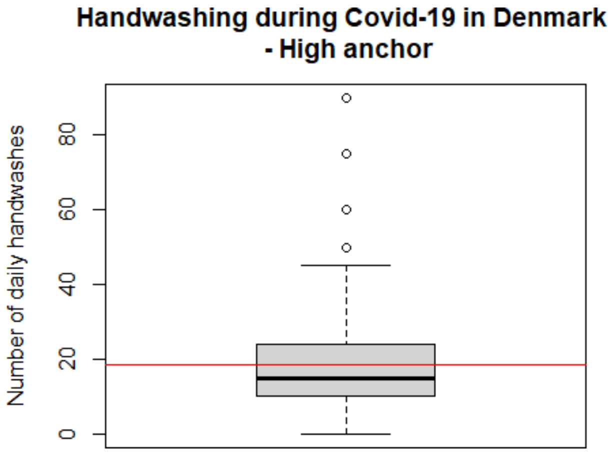
Imai, K. (2017d). Prediction. In *Quantitative Social Science: An Introduction* (pp. 123–138). Princeton University Press.

Imai, K. (2017e). Regression and Causation. In *Quantitative Social Science: An Introduction* (pp. 161–180). Princeton University Press.

Marvel, J. D. (2016). Unconscious bias in citizens' evaluations of public sector performance. *Ournal of Public Administration Research and Theory*, *26*(1), 143–158. https://doi.org/10.1093/jopart/muu053

Appendix A: Handwashing during Covid-19 in Denmark - Anchor Treatment Groups





Appendix B: Coding

```
#Loading the data:
setwd("~/R")
covid <- read.csv("covid.csv")
dim(covid)
names(covid)

load("poll-uk.Rdata")
results
polls
dim(polls)

#------------------------------------------Question 2.1 ----------------------------------------------
table(covid$treat_handwash)

#One-sample, two-sided test for a proportion:
covid.high.anchor <- subset(covid, covid$treat_handwash == "high")
table(is.na(covid.high.anchor$male))
table(covid.high.anchor$male)
cv <- qnorm(0.975, 0,1)
cv

#Sample proportion, X.bar:
n <- 442
x.bar.prop.women <- 1-mean(covid.high.anchor$male)
x.bar.prop.women
#Standard error for the significance test aka the standard deviation of the sampling distribution:
se.prop.women <- sqrt((0.48*(1-0.48))/n)
```

```
se.prop.women

#Z-score for our sample estimate (0.528):

z.score.prop.women <- ((x.bar.prop.women-0.48)/se.prop.women)

z.score.prop.women

#P-value:

pvalue.prop.women <- 2*pnorm(x.bar.prop.women, mean = 0.48, sd = se.prop.women, lower.tail =
FALSE)

#Or:

pvalue.prop.women <- 2*pnorm(z.score.prop.women, lower.tail = FALSE)

pvalue.prop.women


#R function version (prop.test) of calculations above:

prop.test(232, n = 442, p=0.48, conf.level = 0.95, correct = FALSE)

prop.test(232, n = 442, p=0.48, conf.level = 0.90, correct = FALSE)


#.................................Question 2.2 ............................................

#Two-sample, two-sided test for a proportion:

covid.low.anchor <- subset(covid, covid$treat_handwash == "low")

table(is.na(covid.low.anchor$age))


t.test(covid.low.anchor$age[covid.low.anchor$gov == 1],
       covid.low.anchor$age[covid.low.anchor$gov == 0])


#.................................Question 2.3 ............................................

#Summary of the outcome_handwash variable:

summary(covid$outcome_handwash)

range(covid$outcome_handwash)

nrow(covid)
```

```
table(is.na(covid$outcome_handwash))


#Visualization of the outcome_handwash variable:
boxplot(covid$outcome_handwash,
        main = "Handwashing during Covid-19 in Denmark",
        ylab = "Number of daily handwashs/sanitizings")
abline(h = mean(covid$outcome_handwash), col = "red")


#Standard error of the mean:
sem.handwash <- sd(covid$outcome_handwash)/sqrt(884)
sem.handwash
sd(covid$outcome_handwash)


#Looking at the two anchor treatment groups:
par(mfrow=c(1,2))
boxplot(covid.low.anchor$outcome_handwash,
        main = "Handwashing during Covid-19 in Denmark \n - Low anchor",
        ylim = c(0, 90),
        ylab = "Number of daily handwashes")
abline(h = mean(covid.low.anchor$outcome_handwash), col = "red")
boxplot(covid.high.anchor$outcome_handwash,
        main = "Handwashing during Covid-19 in Denmark \n - High anchor",
        ylab = "Number of daily handwashes")
abline(h = mean(covid.high.anchor$outcome_handwash), col = "red")


#..........................................Question 2.4 ...................................................
#Student's T-test:
t.test(covid.high.anchor$outcome_handwash,
```

```
    covid.low.anchor$outcome_handwash)

#Linear regression:
covid$binary.treatment <- ifelse(covid$treat_handwash == "high", 1, 0)
m.anchor.treatment <- lm(outcome_handwash ~ binary.treatment, data = covid)
m.anchor.treatment

library(stargazer)
writeLines(capture.output(
stargazer(m.anchor.treatment,
          digits = 2,
          font.size = "scriptsize",
          single.row = TRUE,
          column.labels=c("Handwash"),
          keep.stat = c("n", "rsq", "adj.rsq"),
          covariate.labels =c("Treatment(High anchor)"),
          model.names=FALSE, type="html"
  )), "table-11.htm")

#............................................Question 2.5 .....................................................
#Summary of the variable:
summary(covid$outcome_closecontact)
range(covid$outcome_closecontact)
table(is.na(covid$outcome_closecontact))
sem.contact <- sd(covid$outcome_closecontact)/sqrt(884)
sem.contact

#Student's T-test:
```

```
t.test(covid$outcome_closecontact[covid$treat_handwash == "high"],
       covid$outcome_closecontact[covid$treat_handwash == "low"])

#Linear regression:
m.anchor.treatment.contact <- lm(outcome_closecontact ~ binary.treatment, data = covid)
m.anchor.treatment.contact

#Printing the linear regression table:
writeLines(capture.output(
  stargazer(m.anchor.treatment.contact,
            digits = 2,
            font.size = "scriptsize",
            single.row = TRUE,
            column.labels=c("Close Contact"),
            keep.stat = c("n", "rsq"),
            covariate.labels =c("Treatment(High anchor)"),
            model.names=FALSE, type="html"
  )), "table-7.htm")

#.................................Question 2.6 ........................................
#Extending the linear regression from 2.4:
m.anchor.treatment.extended <- lm(outcome_handwash ~ binary.treatment + age + male + gov,
data = covid)
m.anchor.treatment.extended

#Table of the model:
library("texreg")
screenreg(m.anchor.treatment.extended)
```

```
#Printing the extended linear regression table:
writeLines(capture.output(
stargazer(m.anchor.treatment.extended,
          digits = 2,
          font.size = "scriptsize",
          single.row = TRUE,
          column.labels=c("Handwash"),
          keep.stat = c("n", "rsq", "adj.rsq"),
          covariate.labels =c("Treatment(High anchor)", "Age", "Sex", "Gov. support"),
          model.names=FALSE, type="html"
  )), "table-12.htm")

range(covid$age)

#Visualization:
install.packages("jtools")
library("jtools")

effect_plot(
  m.anchor.treatment.extended,
  pred = age,
  centered = "all",
  plot.points = FALSE,
  interval = TRUE,
  data = NULL,
  at = NULL,
  int.type = c("confidence", "prediction"),
```

```
  int.width = 0.95,
  y.label = "Number of handwashes",
  x.label = "Age",
  main.title = "Effect of Age on Amount of Handwashing",
  colors = "black")


#.......................................Question 2.7 ..................................................
#Extending the multiple linear regression model from 2.6 to a linear regression model with an inter-
action term:
m.heterogeneous <- lm(outcome_handwash ~ binary.treatment * age + male + gov, data = covid)
m.heterogeneous


writeLines(capture.output(
  stargazer(m.heterogeneous,
          digits = 2,
          font.size = "scriptsize",
          single.row = TRUE,
          column.labels=c("Handwash"),
          keep.stat = c("n", "rsq", "adj.rsq"),
          covariate.labels =c("Treatment(High anchor)", "Age", "Sex", "Gov. support", "Treat-
ment*Age"),
          model.names=FALSE, type="html"
  )), "table-10.htm")


#Visualization:
quantile(covid$age)
#Young age = 25% quantile:
data_age_25q <- data.frame(binary.treatment = seq(from = 0, to = 1, by = 1),
                           age = 41.75,
```

```
                    gov = mean(covid$gov),
                    male = mean(covid$male))
pred.age.25q <- predict(m.heterogeneous, newdata = data_age_25q, interval = "confidence")
cbind.age.25q <- cbind(data_age_25q, pred.age.25q)

#Median age = median:
data_age_median <- data.frame(binary.treatment = seq(from = 0, to = 1, by = 1),
                    age = median(covid$age),
                    gov = mean(covid$gov),
                    male = mean(covid$male))
pred.age.median <- predict(m.heterogeneous, newdata = data_age_median, interval = "confidence")
cbind.age.median <- cbind(data_age_median, pred.age.median)

#Old age = 75% quantile:
data_age_75q <- data.frame(binary.treatment = seq(from = 0, to = 1, by = 1),
                    age = 70,
                    gov = mean(covid$gov),
                    male = mean(covid$male))
pred.age.75q <- predict(m.heterogeneous, newdata = data_age_75q, interval = "confidence")
cbind.age.75q <- cbind(data_age_75q, pred.age.75q)

#Plot:
par(mfrow=c(1,1))
plot(covid$outcome_handwash ~ covid$binary.treatment,
    xlab = "Low anchor
High anchor",
    ylab = "Number of handwashes",
    ylim = c(9,20),
```

```
    xaxt = "n",
    type = "n",
    main = "Heterogeneous treatment effects for different ages \n on Number of handwashes")
lines(cbind.age.25q$binary.treatment, cbind.age.25q[, "fit"], col = "green")
lines(cbind.age.median$binary.treatment, cbind.age.median[, "fit"], col = "black")
lines(cbind.age.75q$binary.treatment, cbind.age.75q[, "fit"], col = "hot pink")

legend(x = "topleft", legend = c("Young Respondents", "Median respondents", "Old Respondents"),
     lty = 1,
     col = c("green", "black", "hot pink"))
points(0, 11.23172, type = "p", col = "black")
points(1, 19.51299, type = "p", col = "black")
points(0, 10.38663, type = "p", col = "black")
points(1, 17.23899, type = "p", col = "black")
points(0, 10.74561, type = "p", col = "black")
points(1, 18.20494, type = "p", col = "black")

#..................................Question 3.1 ........................................
table(polls$house)
summary(polls)
polls$date <- as.Date(polls$poll_date)

#Subsetting the data to only 2015 and 2017:
p15 <- subset(polls, polls$poll_date < "2017-01-04")
p17 <- subset(polls, polls$poll_date > "2015-05-05")

#Subsetting the data to look at each firm in 2015:
#2015:
```

```
opinium.15 <- subset(p15, house == "Opinium")
yougov.15 <- subset(p15, house == "YouGov")

#Errors Opinium 2015:
error.op.lab <- 30.4-opinium.15$vote_lab
error.op.con <- 36.9-opinium.15$vote_con
error.op.libdem <- 7.9-opinium.15$vote_libdem
pred.err.op.15 <- c(error.op.lab, error.op.con, error.op.libdem)
mean(pred.err.op.15)

par(mfrow=c(1,2))
hist(pred.err.op.15, freq = FALSE,
    ylim = c(0.0, 0.15),
    xlim = c(-6,10),
    xlab = "Prediction Error",
    main = "Poll Prediction Errors 2015 - Opinium")
abline(v = mean(pred.err.op.15), col = "red")
text(-1, 0.145, "Average Prediction Error (Bias)", col = "red")

#Errors YouGov 2015:
error.you.lab <- 30.4-yougov.15$vote_lab
error.you.con <- 36.9-yougov.15$vote_con
error.you.libdem <- 7.9-yougov.15$vote_libdem
pred.err.you.15 <- c(error.you.lab, error.you.con, error.you.libdem)
mean(pred.err.you.15)

hist(pred.err.you.15, freq = FALSE,
    ylim = c(0.0, 0.15),
```

```
     xlab = "Prediction Error",
     main = "Poll Prediction Errors 2015 - YouGov")
abline(v = mean(pred.err.you.15), col = "red")
text(-2, 0.145, "Average Prediction Error (Bias)", col = "red")



#Subsetting the data to look at each firm in 2017:
#2017:
opinium.17 <- subset(p17, house == "Opinium")
yougov.17 <- subset(p17, house == "YouGov")

#Errors Opinium 2017:
error.op.lab.17 <- 40.0-opinium.17$vote_lab
error.op.con.17 <- 42.4-opinium.17$vote_con
error.op.libdem.17 <- 7.4-opinium.17$vote_libdem
pred.err.op.17 <- c(error.op.lab.17, error.op.con.17, error.op.libdem.17)
mean(pred.err.op.17)

par(mfrow=c(1,2))
hist(pred.err.op.17, freq = FALSE,
     ylim = c(0.0, 0.15),
     xlim = c(-8, 15),
     xlab = "Prediction Error",
     main = "Poll Prediction Errors 2017 - Opinium")
abline(v = mean(pred.err.op.17), col = "red")
text(1, 0.145, "Average Prediction Error (Bias)", col = "red")

#Errors YouGov 2017:
```

```
error.you.lab.17 <- 40.0-yougov.17$vote_lab
error.you.con.17 <- 42.4-yougov.17$vote_con
error.you.libdem.17 <- 7.4-yougov.17$vote_libdem
pred.err.you.17 <- c(error.you.lab.17, error.you.con.17, error.you.libdem.17)
mean(pred.err.you.17)

hist(pred.err.you.17, freq = FALSE,
    ylim = c(0.0, 0.15),
    xlab = "Prediction Error",
    main = "Poll Prediction Errors 2017 - YouGov")
abline(v = mean(pred.err.you.17), col = "red")
text(1, 0.145, "Average Prediction Error (Bias)", col = "red")


#Root-Mean-Squared Errors 2015:
#Opinium:
sqrt(mean(pred.err.op.15^2))
#YouGov:
sqrt(mean(pred.err.you.15^2))

#Root-Mean-Squared Errors 2017:
#Opinium:
sqrt(mean(pred.err.op.17^2))
#YouGov:
sqrt(mean(pred.err.you.17^2))
```

```
#Bias for the three parties 2015:
predictions.2015 <- c(
  lab = mean(p15$vote_lab),
  con = mean(p15$vote_con),
  libdem = mean(p15$vote_libdem))
results.2015 <- c(30.4, 36.9, 7.9)
bias.2015 <- results.2015-predictions.2015
bias.2015

#RMSE 2015:
#Now with own function:
RMSE = function(result, pollprediction){
  sqrt(mean((result - pollprediction)^2))
}

RMSE.2015 <- c(
  lab = RMSE(results$res_lab[results$election == "2015"], p15$vote_lab),
  con = RMSE(results$res_con[results$election == "2015"], p15$vote_con),
  libdem = RMSE(results$res_libdem[results$election == "2015"], p15$vote_libdem)
)
RMSE.2015

#Bias for the three parties 2017:
predictions.2017 <- c(
  lab = mean(p17$vote_lab),
  con = mean(p17$vote_con),
  libdem = mean(p17$vote_libdem))
results.2017 <- c(40, 42.4, 7.4)
```

```
bias.2017 <- results.2017-predictions.2017
bias.2017

#RMSE 2017:
RMSE.2017 <- c(
  lab = RMSE(results$res_lab[results$election == "2017"], p17$vote_lab),
  con = RMSE(results$res_con[results$election == "2017"], p17$vote_con),
  libdem = RMSE(results$res_libdem[results$election == "2017"], p17$vote_libdem)
)
RMSE.2017

#-----------------------------------Question 3.2 ---------------------------------------------
#Creating days-to-election variables:
#Actual election dates are 2015-05-07 and 2017-06-08:
p15$days_to_election <- as.Date("2015-05-07") - as.Date(p15$date)
p17$days_to_election <- as.Date("2017-06-08") - as.Date(p17$date)

#2015:
par(mfrow=c(3,1))
#Labour:
plot(p15$days_to_election, p15$vote_lab,
     xlab = "Days to Election",
     ylab = "Predicted Vote Share",
     xlim =c(125, -1), pch = 16,
     type = "l",
     main = "Predicted Vote Share Labour 2015")
abline(h = 30.4, col = "red")
abline(v = 0, col = "red")
```

```
#Conservative:
plot(p15$days_to_election, p15$vote_con,
    xlab = "Days to Election",
    ylab = "Predicted Vote Share",
    xlim =c(125, -1), pch = 16,
    type = "l",
    main = "Predicted Vote Share Conservative 2015")
abline(h = 36.9, col = "red")
abline(v = 0, col = "red")

#Liberal Democrats:
plot(p15$days_to_election, p15$vote_libdem,
    xlab = "Days to Election",
    ylab = "Predicted Vote Share",
    xlim =c(125, -1), pch = 16,
    type = "l",
    main = "Predicted Vote Share Liberal Democrats 2015")
abline(h = 7.9, col = "red")
abline(v = 0, col = "red")


#2017:
par(mfrow=c(3,1))
#Labour:
plot(p17$days_to_election, p17$vote_lab,
    xlab = "Days to Election",
    ylab = "Predicted Vote Share",
```

```
    ylim = c(23, 41),
    xlim =c(150, -1), pch = 16,
    type = "l",
    main = "Predicted Vote Share Labour 2017")
abline(h = 40.0, col = "red")
abline(v = 0, col = "red")

#Conservative:
plot(p17$days_to_election, p17$vote_con,
    xlab = "Days to Election",
    ylab = "Predicted Vote Share",
    xlim =c(150, -1), pch = 16,
    type = "l",
    main = "Predicted Vote Share Conservative 2017")
abline(h = 42.4, col = "red")
abline(v = 0, col = "red")

#Liberal Democrats:
plot(p17$days_to_election, p17$vote_libdem,
    xlab = "Days to Election",
    ylab = "Predicted Vote Share",
    xlim =c(150, -1), pch = 16,
    type = "l",
    main = "Predicted Vote Share Liberal Democrats 2017")
abline(h = 7.4, col = "red")
abline(v = 0, col = "red")
```

```
#Biases for different time periods - 2015:
first.polls.15 <- subset(p15, subset = (p15$days_to_election > 80))
middle.polls.15 <- subset(p15, subset = (p15$days_to_election > 40 & p15$days_to_election <=
80))
last.polls.15 <- subset(p15, subset = (p15$days_to_election <= 40))

#Bias and RMSE first period 2015:
error.lab.2015.f <- 30.4-first.polls.15$vote_lab
error.con.2015.f <- 36.9-first.polls.15$vote_con
error.libdem.2015.f <- 7.9-first.polls.15$vote_libdem
pred.err.all.2015.f <- c(error.lab.2015.f, error.con.2015.f, error.libdem.2015.f)
mean(pred.err.all.2015.f)
sqrt(mean(pred.err.all.2015.f^2))

#Bias and RMSE middle period 2015:
error.lab.2015.m <- 30.4-middle.polls.15$vote_lab
error.con.2015.m <- 36.9-middle.polls.15$vote_con
error.libdem.2015.m <- 7.9-middle.polls.15$vote_libdem
pred.err.all.2015.m <- c(error.lab.2015.m, error.con.2015.m, error.libdem.2015.m)
mean(pred.err.all.2015.m)
sqrt(mean(pred.err.all.2015.m^2))

#Bias and RMSE last period 2015:
error.lab.2015.l <- 30.4-last.polls.15$vote_lab
error.con.2015.l <- 36.9-last.polls.15$vote_con
error.libdem.2015.l <- 7.9-last.polls.15$vote_libdem
pred.err.all.2015.l <- c(error.lab.2015.l, error.con.2015.l, error.libdem.2015.l)
mean(pred.err.all.2015.l)
```

```
sqrt(mean(pred.err.all.2015.l^2))

#Biases for different time periods - 2017:
first.polls.17 <- subset(p17, subset = (p17$days_to_election > 100))
middle.polls.17 <- subset(p17, subset = (p17$days_to_election > 50 & p17$days_to_election <=
100))
last.polls.17 <- subset(p17, subset = (p17$days_to_election <= 50))

#Bias first period 2017:
error.lab.2017.f <- 40-first.polls.17$vote_lab
error.con.2017.f <- 42.4-first.polls.17$vote_con
error.libdem.2017.f <- 7.4-first.polls.17$vote_libdem
pred.err.all.2017.f <- c(error.lab.2017.f, error.con.2017.f, error.libdem.2017.f)
mean(pred.err.all.2017.f)
sqrt(mean(pred.err.all.2017.f^2))

#Bias middle period 2017:
error.lab.2017.m <- 40-middle.polls.17$vote_lab
error.con.2017.m <- 42.4-middle.polls.17$vote_con
error.libdem.2017.m <- 7.4-middle.polls.17$vote_libdem
pred.err.all.2017.m <- c(error.lab.2017.m, error.con.2017.m, error.libdem.2017.m)
mean(pred.err.all.2017.m)
sqrt(mean(pred.err.all.2017.m^2))

#Bias last period 2017:
error.lab.2017.l <- 40-last.polls.17$vote_lab
error.con.2017.l <- 42.4-last.polls.17$vote_con
error.libdem.2017.l <- 7.4-last.polls.17$vote_libdem
```

```
pred.err.all.2017.l <- c(error.lab.2017.l, error.con.2017.l, error.libdem.2017.l)
mean(pred.err.all.2017.l)
sqrt(mean(pred.err.all.2017.l^2))

#..........................................Question 3.3 ..........................................................
#Sample size:
summary(polls$sample)
summary(p15$sample)
sd(p15$sample)
summary(p17$sample)
sd(p17$sample)

#Sample size to prediction error - 2015:
p15$error.lab.2015 <- 30.4-p15$vote_lab
p15$error.con.2015 <- 36.9-p15$vote_con
p15$error.libdem.2015 <- 7.9-p15$vote_libdem

table(p15$error.lab.2015)
table(p15$sample, p15$error.lab.2015)
table(p15$error.con.2015)
table(p15$sample, p15$error.con.2015)
table(p15$error.libdem.2015)
table(p15$sample, p15$error.libdem.2015)

#Sample size to prediction error - 2017:
p17$error.lab.2017 <- 40-p17$vote_lab
p17$error.con.2017 <- 42.4-p17$vote_con
p17$error.libdem.2017 <- 7.4-p17$vote_libdem
```

```
table(p17$error.lab.2017)
table(p17$sample, p17$error.lab.2017)
table(p17$error.con.2017)
table(p17$sample, p17$error.con.2017)
table(p17$error.libdem.2017)
table(p17$sample, p17$error.libdem.2017)

#Correlation between sample size and prediction errors - 2015:
par(mfrow=c(1,2))
plot(p15$sample, abs(p15$error.lab.2015),
     xlim = c(1500, 3000),
     xlab = "Sample Size",
     ylim = c(-1, 10),
     ylab = "Prediction Errors (Aboslute values)",
     main = "Prediction Error to Sample Size - 2015",
     pch = 16,
     col = "red")
points(p15$sample, abs(p15$error.con.2015), pch = 16, col = "blue")
points(p15$sample, abs(p15$error.libdem.2015), pch = 16, col = "light green")
legend(x = "topright", legend = c("Labour", "Conservatives", "Liberal Democrats"),
       col = c("red", "blue", "light green"),
       lty = 1)

cor(p15$sample, abs(p15$error.lab.2015))
cor(p15$sample, abs(p15$error.con.2015))
cor(p15$sample, abs(p15$error.libdem.2015))
```

```
#Correlation between sample size and prediction errors - 2017:
plot(p17$sample, abs(p17$error.lab.2017),
     xlim = c(1500, 3002),
     xlab = "Sample Size",
     ylim = c(-2, 20),
     ylab = "Prediction Errors (Aboslute values)",
     main = "Prediction Error to Sample Size - 2017",
     pch = 16,
     col = "red")
points(p17$sample, abs(p17$error.con.2017), pch = 16, col = "blue")
points(p17$sample, abs(p17$error.libdem.2017), pch = 16, col = "light green")
legend(x = "topright", legend = c("Labour", "Conservatives", "Liberal Democrats"),
       col = c("red", "blue", "light green"),
       lty = 1)

cor(p17$sample, abs(p17$error.lab.2017))
cor(p17$sample, abs(p17$error.con.2017))
cor(p17$sample, abs(p17$error.libdem.2017))
```