Quantitative Methods in Political Science
Fall 2021

Midterm Exam

Online Group A

(50 points total)

Name: _____

Please do not turn this page until you are told to do so.

Please read carefully: You have 90 minutes to complete the exam. There are three blank pages at the end of your exam paper that you can use for notes during the exam (notes will not be evaluated). No other notes are permitted. Please write clearly and provide answers only in the spaces provided.

Good luck!

Grading Table (for instructor use only)

	Part 1.1	**Part 1.2**	**Part 2**	**Part 3**	**Total**
Points					

1 Interpretation of regression models (23 points)

1. In a recent publication in the *American Journal of Political Science*, Samara Klar and Alexandra McCoy study the effect of partisan biases when evaluating sexual misconduct allegations in politics in the context of the *#MeToo* movement in the US using an experimental design. The authors present each survey participant a vignette in which the accuser claims that the harasser offered the accuser a high-paying job in the election campaign if the accuser agreed to have an affair with the harasser. The harasser, however, denies all allegations, calling them "unequivocally false."

   The study authors manipulate three different components of the vignette: the partisanship of the accuser (*Democratic/Republican*), the partisanship of the harasser (*Democratic/Republican*), and whether a woman accuses a man or vice versa (no same-sex scenario!). After reading the fictitious scenario in the vignette, respondents were asked to report whether they consider the harasser to be guilty or not guilty of this allegation ("presumed guilt") using a 7-point response scale, ranging from "Definitely not guilty" (= 1) to "Definitely guilty" (= 7). The authors code three dummy variables, the partisanship of the accuser and the partisanship of the harasser according to whether it matches that of the respondent (= 1), and whether a woman accuses a man (= 1).

TABLE 2 Presumption of Guilt Regressed on Each Treatment

	Full Sample β (s.e.)	Democrats B (s.e.)	Republicans B (s.e.)	Men B (s.e.)	Women β (s.e.)
Same partisanship as harasser	−0.48**	−0.46**	−0.49**	−0.59**	−0.36**
	(0.10)	(0.13)	(0.15)	(0.15)	(0.12)
Same partisanship as accuser	0.15	0.14	0.18	0.16	0.15
	(0.10)	(0.13)	(0.15)	(0.15)	(0.12)
Woman accuses man	0.26**	0.48**	0.01	0.33*	0.20†
	(0.10)	(0.13)	(0.15)	(0.15)	(0.12)
N	748	403	345	387	361
Adj. R^2	0.04	0.06	0.03	0.04	0.03

**p < 0.01; *p < 0.05; †p < 0.1.

The following questions deal with the model of the full sample (first model):

(a) **(2 points)** Write down the systematic component of the (first) model. And what is missing here compared to a typical systematic component?

(b) **(3 points)** Interpret the coefficient of the *Woman accuses man* dummy.

(c) **(2 points)** The authors claim: "Partisans view alleged harassers from their own party as less guilty than alleged harassers from the opposing party." Do you agree given the presented evidence? Explain your answer in one or two sentences.

(d) **(2 points)** The authors furthermore claim that "partisanship of the accuser does not drive biased evaluations of the misconduct allegation. ($H_0 : \beta_{accuser} = 0$)". Using the coefficient estimate and the standard error for the *Same partisanship as accuser* dummy, construct a (rough) 95 percent confidence interval for its effect on presumed guilt (use the rule of thumb for the calculation of the confidence interval). Based on the confidence interval, do you agree with them? Explain your answer in one sentence.

(e) **(3 points)** How strong should we expect this partisan bias to be? Let's define the partisan bias as a first-difference of the expected evaluation of presumed guilt when comparing respondents that rather support the same party as accuser *and* harasser to those that support the opposite party. Calculate the first-difference (you only have to calculate the quantity of interest, **not** its uncertainty) based on the estimates of the first model.

(f) **(1 points)** Let's go back to (a): What is the consequence of not having a typical systematic component?

2. In the same study, the authors extend their analysis to test whether more polarized individuals – i.e. respondents that have very different feelings towards both parties, the Democrats and the Republicans – engage in more partisan-motivated evaluations. The authors therefore construct a *feeling thermometer differential* as the difference between the thermometer rating for the party they prefer more minus the thermometer rating for the other party. This thermometer differential ranges from 0 ("parties are evaluated equally") to 6 ("largest possible difference between the parties"). Before the experiment the survey already measured demographic traits and the extent to which respondents support the *#MeToo* movement. The authors estimate the following model:

TABLE 3 Presumption of Guilt Regressed on Partisanship of Harasser × Thermometer Differential

	Full Sample B (s.e.)
Same partisanship as harasser	−0.08
	(0.18)
Same partisanship as accuser	0.15
	(0.10)
Woman accuses man	0.25*
	(0.10)
Thermometer Differential	0.02
	(0.02)
Same partisanship as harasser ×	−0.08**
Thermometer Differential	(0.03)
Support for #MeToo Movement	0.20**
	(0.05)
Age	−0.01**
	(0.00)
Education	−0.14**
	(0.05)
Gender	−0.02
	(0.10)
Party identification	−0.01
	(0.03)
Ideology	0.05
	(0.03)
Constant	3.97
	(0.34)
N	723
Adj. R^2	0.09

**p < 0.01; *p < 0.05; †p < 0.1.

(a) **(2 points)** What is the conditional effect of *Thermometer Differential* on the presumption of guilt, i.e., the effect of *Thermometer Differential* conditional on the *Same partisanship as harasser'* dummy (<u>not</u> its uncertainty)?

(b) **(2 points)** What is the interpretation of the p-value of the *Woman accuses man* coefficient? (No calculation required!)

(c) **(2 points)** What is the problem if we would measure *Party identification* and *Ideology* after the treatment and include them as covariates in the above model rather than the pre-treatment measures the authors use?

(d) **(1 points)** Suppose we drop a covariate from the above model, say *Age*, re-estimate the new model, but still obtain the same R^2. Would the adjusted R^2 of this new model increase, decrease or does it not matter? Explain (you do not have to calculate anything)!

(e) **(3 points)** Given the above model, what is the expected difference on the "presumed guilt" scale for respondents who have the same partisanship as the harasser in their vignette if they are not polarized (i.e. they evaluate both parties equally) as opposed to have a value of 2 on the *Thermometer Differential*, all else equal? Calculate this first-difference (you only have to calculate the quantity of interest, **not** its uncertainty)

2 Basic Statistics (7 Points)

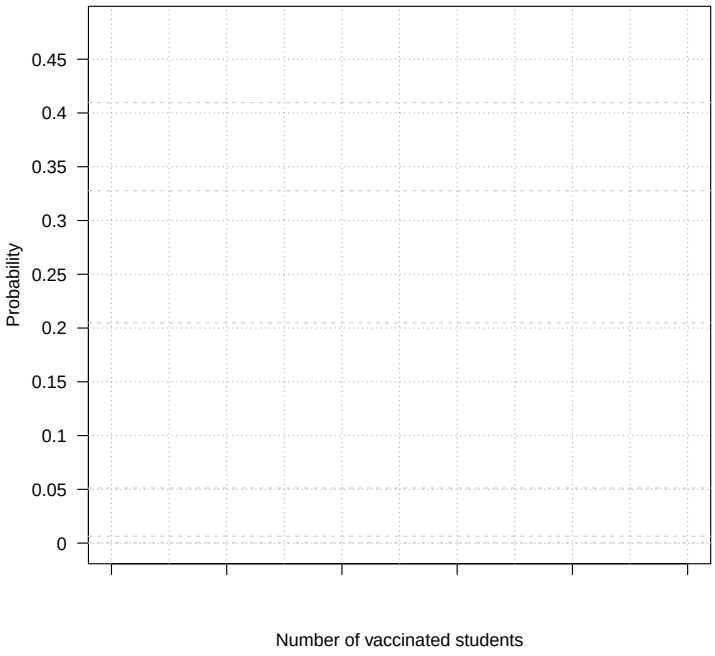
$0.8 = \frac{2^3}{10}$ of the students at a university are vaccinated. You randomly select five students.

1. **(4 points)** Draw the probability function of the number of students that are vaccinated in the group of five students, given the vaccination rate.

   Hint 1: $\binom{5}{0} = 1$, $\binom{5}{1} = 5$, $\binom{5}{2} = 10$, $\binom{5}{3} = 10$, $\binom{5}{4} = 5$, $\binom{5}{5} = 1$,

   Hint 2: $2^3 = 8$; $2^5 = 32$; $2^7 = 128$; $2^9 = 512$; $2^{11} = 2048$; $2^{13} = 8192$; $2^{15} = 32768$

   Hint 3: $x^2 \times x^3 = x^5$; $(x^2)^3 = x^{2\times 3} = x^6$



2. **(1 points)** Calculate the expectation and the variance of the distribution.

3. **(2 points)** What is the probability that there are at least two but less than four vaccinated students in the group?

3 Key terms and concepts (20 points)

Choose the correct answer(s).

Each correct answer is worth 2 points. Check only one box per question. Wrong answers, or answers with multiple checks will receive zero points. Should you need to make a correction, circle the checked box with your final answer and leave the other answer checked.

1. When we use the term *99% confidence interval* for a sample mean, ...

   - ☐ we mean that there is a 99% probability that this confidence interval covers the true population mean.
   - ☐ we expect that any given 99% confidence interval from a random sample will contain the true population parameter.
   - ☐ we can expect that this 99% confidence interval is one of the intervals that could contain the population parameter and we could verify if it does by repeating the experiment a large number of times.
   - ☐ if that 99% confidence interval does not include value x, we can say that there is a statistically significant difference between that value x and the sample mean at the 99% confidence level.
   - ☐ we know that the true population parameter and 99% of the sample means will fall inside this confidence interval, should we repeat the experiment a large number of times.

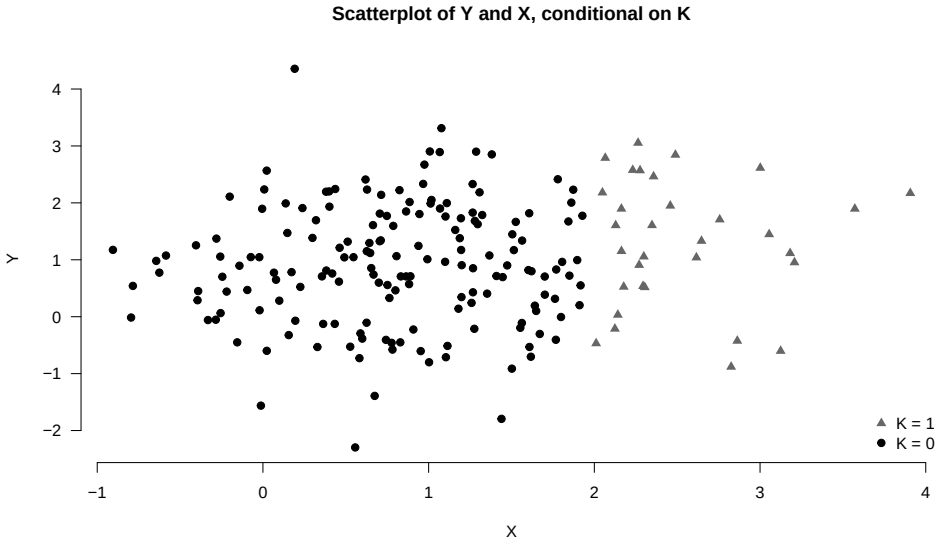
2. When calculating the expected values via simulation, ...

   - ☐ we only incorporate the fundamental uncertainty when calculating the quantity of interest and leave out the estimation uncertainty.
   - ☐ we only incorporate the estimation uncertainty when calculating the quantity of interest and leave out the fundamental uncertainty.
   - ☐ we take random draws from a sampling distribution of each coefficient separately, thus incorporating the uncertainty associated with the estimation of our model.
   - ☐ we expect the fundamental uncertainty to cancel out in the long run since $Var(\epsilon|X) = \sigma^2$, and for this reason, we may omit it from the calculation of expected values.
   - ☐ we assume that each independent variable in our model follow a joint multivariate normal distribution.

3. The residual variance...

   - ☐ is the quantity that depicts the variance that cannot be explained by the stochastic component of the model.
   - ☐ is inversely related to the variance of the OLS estimator (i.e., the larger the residual variance, the smaller the variance of regression coefficients).
   - ☐ is the quantity that depicts the variance that cannot be explained by the systematic component of the model.
   - ☐ is the quantity required to calculate the first differences via simulation from expected values.

4. Consider the following scatterplot of outcome variable Y and treatment X. The observations are colored according to a binary post-treatment variable K, which is not correlated with Y. Which of the following statements is correct?



   - ☐ K seems to be a confounder and should be controlled for if we want to estimate the total effect of X on Y to avoid omitted variable bias.
   - ☐ K seems to be a collider, and including it in the model may increase the statistical and/or substantive significance of the coefficient for X.
   - ☐ K seems to be a collider, and including it in the model may decrease the statistical significance of the coefficient for X.
   - ☐ We would not be able to distinguish between a collider and a confounder in a scatterplot.

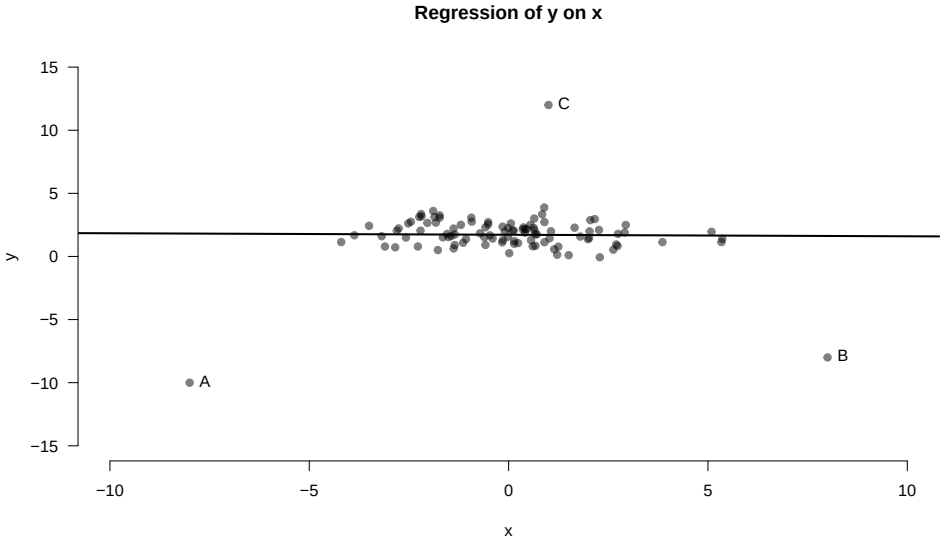
5. What does *holding Z constant* part in interpreting the regression coefficient of X in $Y = \hat{\beta}_0 + \hat{\beta}_1 X + \hat{\beta}_2 Z$ imply?

   - ☐ We are comparing the effect of X on Y within the same values of Z.
   - ☐ We obtain the coefficient for X independent of Z, i.e. coefficient $\hat{\beta}_1$ only corresponds to the part of the variance in Y that is jointly explained by X and Z.
   - ☐ Z is assumed to have constant variance in this model.
   - ☐ All of the statements above are correct.
   - ☐ None of these statements are correct.

6. Bootstrapping approach to calculating confidence intervals...

   - ☐ requires a large sample for us to be able to assume that the distribution can be approximated via the normal distribution.
   - ☐ requires us to treat the sample as if it is the population.
   - ☐ means you are performing sampling without replacement multiple times to generate the sampling distribution and calculate the empirical quantiles of this distribution.
   - ☐ means all of the above.
   - ☐ means none of the above.

7. Consider the following scatterplot of a bivariate linear regression model of y on x. Observations A, B and C are labelled in the plot. Which of the following statements is correct?



   - ☐ Observations A, B, and C are influential outliers.
   - ☐ Observation B is an influential outlier, but not A and C.
   - ☐ Observations A and B are influential outlier, but not C.
   - ☐ Neither of the observations A, B and C are influential outliers.

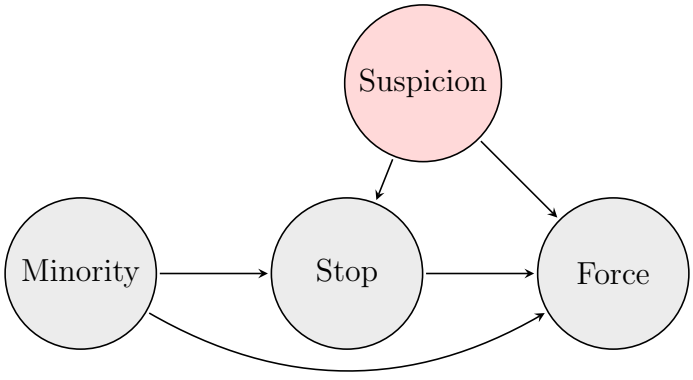
8. The marginal effect of D on Y (i.e. $\frac{\delta Y}{\delta D}$) is given by $\beta_1 + \beta_4 Z$. Which of the following equations might be the corresponding regression equation?

   - ☐ $Y = \beta_0 + \beta_1 D + \beta_2 Z + \beta_3 X + \beta_4 XZ + \epsilon$
   - ☐ $Y = \beta_0 + \beta_1 X + \beta_2 D + \beta_3 Z + \beta_4 XZ + \epsilon$
   - ☐ $Y = \beta_0 + \beta_1 D + \beta_2 Z + \beta_3 X + \beta_4 DZ + \epsilon$
   - ☐ $Y = \beta_0 + \beta_1 D + \beta_2 X + \beta_3 DX + \epsilon$

9. You estimate the model $y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$, expecting that this specification allows you to capture the unbiased effect of x_1 on y. All variables have different scales, and there are 150 observations in the dataset. Which statement is correct?

   - ☐ If coefficient $\hat{\beta}_3$ has a p-value > 0.05, we can drop x_3 from the model because it is not important.
   - ☐ The model has 147 degrees of freedom.
   - ☐ If you center all independent variables, only $\hat{\beta}_0$ will change its absolute value and/or sign.
   - ☐ The regression predictions (fitted values) for any observation in our data will differ between the models estimated with centered and uncentered variables.
   - ☐ In this model, if $\hat{\beta}_1 > \hat{\beta}_2$, the effect of x_1 on y is larger than the effect of x_2 on y.

10. A team of researchers is interested in the effects of racially biased policing, especially how the race of the civilian (minority status) involved in an encounter impacts police behavior (use of force). To conduct an empirical analysis, the team relies on administrative records. They are concerned about the impact of unobserved influences, U, such as the suspicion or perceived threat of a police officer. A record of police force can only be made when an individual is stopped by the police.



Which of the following statements is true?

- ☐ *Stop* is a mediating variable of *Minority* status and *Force*, and thus a bad control.
- ☐ *Minority* status is a confounder of *Stop* and *Force*, and thus a good control variable.
- ☐ *Minority* status is a confounder of *Stop* and *Force*, and thus a bad control variable.
- ☐ *Stop* is a mediating variable of *Minority* status and *Force*, and thus a good control.
- ☐ All of the above.
- ☐ None of the above.

Notes (will not be evaluated):

Notes (will not be evaluated):

Notes (will not be evaluated):