QM 2021: Midterm Mock Exam

(50 points total)

2021

Instructions

Please read carefully: You have **90 minutes** to complete the exam. You may use a calculator to solve mathematical problems in this mock exam. In the real exam, easy numbers will be presented to you, so that no calculator is necessary nor permitted in the exam. Please make sure to concisely **number your answers** so that they can be matched with the corresponding questions.

1 Interpretation of regression models (22/50 points)

1.1 Education and Income (10 points)

A researcher examines the relationship between education and yearly income. Her publication contains the following two tables:

Variable	Mean	Std. Dev.
Age	51	15
Female	0.7	(dummy)
Education	10	4

	Coef.	Std. Err.
Intercept	-9.94	2.10
Age	1.03	0.03
Female	-4.09	1.03
Education	1.04	0.12

1. **(1 points)** Write down the systematic component of the model.
2. **(2 points)** Interpret the coefficients for *Age* and *Female*.
3. **(2 points)** The author claims that she "cannot reject the null hypothesis that *Education* has no effect on *Income* ($H_0 : \beta_{\text{Education}}$=0)' '. Using the coefficient estimate and the standard error for *Education*, construct a (rough) 95 percent confidence interval for the effect of *Education* on *Income* (you may use the rule of thumb for the calculation of the confidence interval). Based on the confidence interval, do you agree with the author? Explain your answer in one sentence.
4. **(2 points)** Calculate the expected value of *Income* for an observation with the following covariate values: $Age = \overline{Age}$, $Female = 0$, $Eduation = \overline{Education}$ (you only have to calculate the quantity of interest, **not** its uncertainty).
5. **(3 points)** Calculate the first difference in $Income$ between low and high values of *Education*, holding all other variables constant at their means (again, you only have to calculate the quantity of interest, **not** its uncertainty). Use $\overline{Education}\pm$ one standard deviation for low and high values of *Education*, respectively.

1.2 Public Opinion and Public Policy (12 points)

How well does public policy represent public opinion in U.S. states? In a recent publication in the *American Journal of Political Science*, Simonovits et al. investigate this question by examining the relationship between the preferred minimum wage of states' citizen and the actual minimum wages in states. If states are responsive to their citizens' preferences, preferences for higher minimum wages should be associated with higher minimum wages within the respective state. Moreover, Simonovits et al. test whether this relationship varies between states with and without access to direct democracy. They include the following variables in their model.

- The dependent variable *State Minimum Wage* is measured as a state's actual minimum wage **in US$ per hour**.
- *Average Preference* measures the average preferred minimum wage **in US$ per hour** in each US state.
- *Citizen Initiatives* is a **dummy variable** that measures a State's access to direct democracy. It is coded 1 if a state allows citizen initiatives and 0 otherwise.

Below, you see some estimates that mimick Simonovits et al.'s findings:

	Coef.	Std. Err.
Intercept	0.605	0.885
Average Preference	0.905	0.085
Citizen Initiative	0.057	1.530
Average Preference × Citizen Initiative	0.057	0.033

1. **(2 points)** What is the effect of *Average Preference* in states with access to direct democracy (not its uncertainty)? Interpret it for an audience that is not versed in statistics.
2. **(3 points)** Citizens in Ohio and Iowa both on average prefer a minimum wage of US$ 10. The state of Ohio allows citizen initiatives while Iowa does not. What minimum wages can we expect in both states (again, you do not have to calculate the uncertainty)?
3. **(1 points)** The p-value of the interaction term is $p = 0.088$. What can we conclude given a significance level of $\alpha = .05$?
4. **(2 points)** Consider the coefficient of *Average Preference*. Construct a (rough) 95% confidence interval. Can we reject the null hypothesis that in states *without* access to direct democracy, citizen average preference has no effect on the state's minimum wage?
5. **(1 point)** The authors claim that their results support the hypothesis that states are *not perfectly responsive* to citizen preferences. A reviewer remarks that in order to test this, showing that the coefficient of *Average Preference* is significantly different

from zero is incorrect. What would be the correct null hypothesis? Test this null hypothesis for states without access to direct democracy using the confidence interval you constructed in the previous exercise.

6. **(3 points)** Suppose you have requested the replication materials from the authors. You rerun the model and extract the variance-covariance matrix of the coefficients, from which you learn that $\text{Cov}(\hat{\beta}_{\text{Average Preference}}, \hat{\beta}_{\text{Average Preference}\times\text{Citizen Initiative}}) = 0.001$. Calculate the 95% confidence interval for the marginal effect of *Average Preference* in states *with* access to direct democracy.

2 Probability theory (8/50 points)

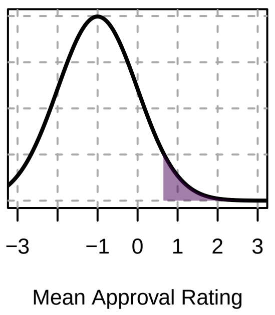
2.1 Popularity of Donald Trump (5 points)

A researcher wants to find out how popular Donald Trump is among US voters. To do this, she runs a survey among 1,000 randomly sampled voters in the US. She finds that Donald Trump's mean popularity is -1.0 and the standard error of the estimate is 0.5.

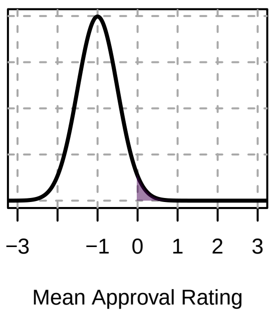
1. **(3 points)** Below you see six distributions with shaded areas. Which distribution satisfies the following two properties:

- It shows the sampling distribution of the estimate.
- The top 5% of the distribution are shaded.

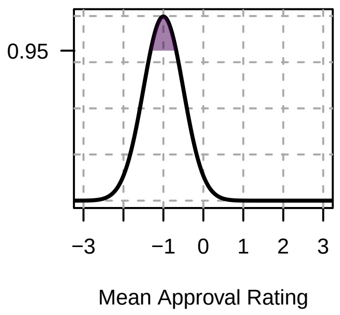
a.



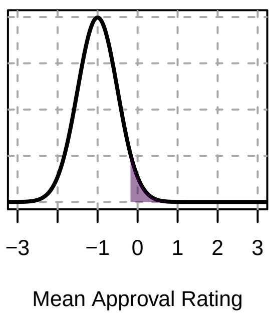
b.



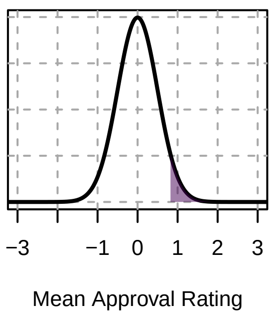
c.



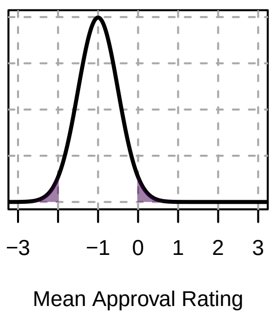
d.



e.



f.



2. **(2 points)** Donald Trump claims to have a positive average popularity rating and refers to results from another survey of 1,000 randomly sampled US voters. Imagine you could draw 1,000 random samples from the population of US voters, each containing 1,000 voters. Suppose that the researcher estimated the true sampling distribution, in how many samples would you expect Donald Trump to obtain a positive average popularity rating?

2.2 Biden voters (3 points)

Suppose you have sampled four voters from Wyoming. Each voter has a 3/10 probability of voting for Biden.

1. **(2 points)** What is the probability that exactly 3 out of 4 voters vote for Biden?
2. **(1 point)**What is the expected value for the number of Biden voters in the sample?

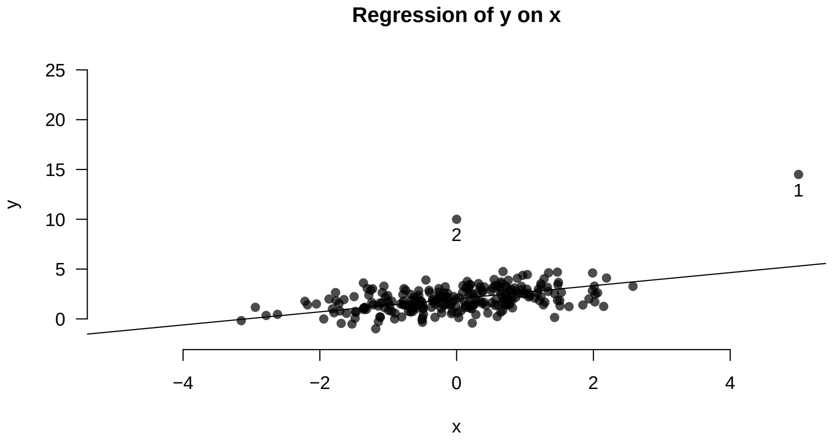
Hint: $\binom{4}{0} = 1$, $\binom{4}{1} = 4$, $\binom{4}{2} = 6$, $\binom{4}{3} = 4$, $\binom{4}{4} = 1$

3 Key terms and concepts (20/50 points)

3.1 Single-choice questions (16 points)

Each correct answer is worth **2 points**. Select only one answer per question. Wrong answers or answers with multiple selected answers will receive zero points.

1. Suppose you run two linear regression models: (1) $Y = \beta_0^* + \beta_1^* X_1 + \epsilon^*$, and (2) $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$, where X_2 is a potential confounder. Omitted variable bias occurs if:
   a. X_2 affects Y and is correlated with X_1, and $\beta_1 - \beta_1^* = 0$.
   b. X_2 affects Y and is uncorrelated with X_1, and $\beta_1 - \beta_1^* = 0$.
   c. X_2 affects Y and is uncorrelated with X_1, and $\beta_1 - \beta_1^* \neq 0$.
   d. X_2 affects Y and is correlated with X_1, and $\beta_1 - \beta_1^* \neq 0$.
2. What is the marginal effect of X on Y (i.e. $\frac{\delta Y}{\delta X}$) given the following equation: $Y = \beta_0 + \beta_1 X + \beta_2 W + \beta_3 Z + \beta_4 XW + \beta_5 XZ + \beta_6 WZ + \beta_7 XWZ + \epsilon$?
   a. None of the other answers.
   b. $\beta_1 + \beta_4 W$.
   c. $\beta_1 + \beta_6 + \beta_7 X$.
   d. $\beta_1 + \beta_5 Z$.
   e. $\beta_1 + \beta_4 W + \beta_5 Z + \beta_7 WZ$.
3. Consider the following scatterplot of a bivariate linear regression model of y on x. Observations 1 and 2 are labeled in the plot. Which of the following statements is correct?



   a. Observation 1 is an influential outlier.
   b. Neither observation 1 nor 2 are influential outliers.
   c. Observations 2 is an influential outlier.
   d. Observations 1 and 2 are influential outliers.

4. Simulating expected values from a regression model takes several steps. Which one is *not* part of it?
   a. Obtain coefficients and variance-covariance matrix from the regression.
   b. Choose interesting covariate values a.k.a set a scenario.
   c. Calculate expected values using the sampling distribution of the coefficients and the scenario.
   d. Randomly choose a multivariate normal distribution in order to approximate the sampling distribution.
   e. All of the other answers are required.
5. What causes attenuation bias?
   a. Measurement error in X (independent variable).
   b. High correlation between independent variables.
   c. Measurement error in Y (dependent variable).
   d. None of the other answers.
   e. Omitting an important independent variable.
6. The following regression table shows estimates of the effects of students' grades and self-esteem on happiness. Which of the following statements is true?

	Dependent variable:		
	Happiness		
	(1)	(2)	(3)
Grades	0.396^{***}		0.040
	(0.111)		(0.110)
Self-Esteem		0.654^{***}	0.635^{***}
		(0.086)	(0.101)
Constant	2.857^{***}	2.051^{***}	1.904^{***}
	(0.693)	(0.448)	(0.605)
Observations	100	100	100
R^2	0.115	0.372	0.373
Adjusted R^2	0.106	0.366	0.360
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

a. *Grades* and *Happiness* are independent (i.e., uncorrelated) variables.

b. *Grades* explain some variation in *Happiness*, but not beyond their correlation with *Self-Esteem*.

c. Omitting the variable *Grades* severely biases the effect of *Self-Esteem* on *Happiness* downwards.

d. *Grades* improve the explanatory power of our model over and beyond *Self-Esteem*.

7. Let X_1 and X_2 be two different values of an independent variable. A first-difference is. . .

a. a quantity of interest from a regression model defined as $Y|X_1 - Y|X_2$.

b. a quantity of interest from a regression model defined as $E(Y|1) - E(Y|0)$.

c. a quantity of interest from a regression model defined as $E(Y|X_1)/E(Y|X_2)$.

d. None of the other answers.

e. a quantity of interest from a regression model defined as $E(Y|X_1) + E(Y|X_2)$.

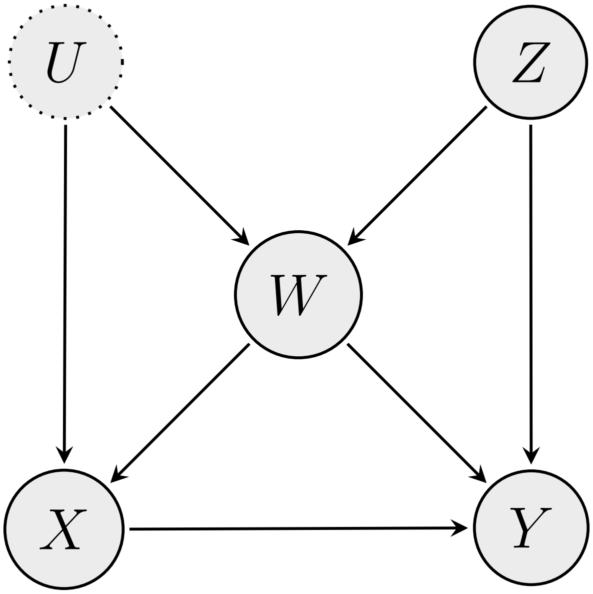
8. When we use the term '95% confidence interval' for a sample mean, we mean that. . .

a. the true population parameter is contained in the interval about 95 times if we were to repeat the experiment one hundred times.

b. there is a 95 percent probability that the population mean is within the interval.
c. any given 95 percent confidence interval from a random sample will contain the true population mean.
d. we can reject the null hypothesis with 95 percent confidence.
e. None of the other answers.

3.2 Causal graphs (4 points + 2 bonus points)

You are interested in estimating the causal effect of X on Y. The causal relationships between these two variables, additional variables W and Z, and an unobservable variable U are illustrated in the causal graph below.



Note: In the following questions, one or multiple choices may be correct.

1. **(2 points)** Below, you see all paths that connect X and Y. Please indicate *all* paths that are *non-causal*.

   a. `X <- W -> Y`
   b. `X <- U -> W -> Y`
   c. `X <- U -> W <- Z -> Y`
   d. `X <- W <- Z -> Y`
   e. `X -> Y`

2. **(2 points)** To retrieve a causal effect of X on Y, all *causal paths* running from X to Y must be open while all *non-causal paths* must be blocked. Can this be achieved? If so, how? The causal effect of X on Y…

   a. …can be retrieved by controlling for Z.
   b. …can be retrieved by controlling for both W and Z.
   c. …cannot be retrieved at all.
   d. …can be retrieved by controlling for neither W nor Z.
   e. …can be retrieved by controlling for W.

3. **(Bonus question, 2 points)** Suppose Z was also an unobservable variable and could thus not be possibly controlled for. Would this change your answer to the previous question? If so, why? Explain in 1-2 sentences.