Data 102, Fall 2023
Midterm 2

- You have **110 minutes** to complete this exam. There are **5 questions**, totaling **40 points**.
- You may use **two** 8.5×11 sheet of handwritten notes (front and back), and the provided reference sheet. No other notes or resources are allowed.
- You should write your solutions inside this exam sheet.
- You should write your Student ID on every sheet (in the provided blanks).
- Make sure to write clearly. We can't give you credit if we can't read your solutions.
- Even if you are unsure about your answer, it is better to write down something so we can give you partial credit.
- We have provided a blank pages of scratch paper at the **end** of the exam. No work on this page will be graded.
- You may, without proof, use theorems and facts given in the discussions or lectures, **but please cite them**.
- We don't answer questions individually. If you believe something is unclear, bring your question to us and if we find your question valid we will make a note to the whole class.
- Unless otherwise stated, no work or explanations will be graded for multiple-choice questions.
- Unless otherwise stated, you must show your work for free-response questions in order to receive credit.

Last name	
First name	
Student ID (SID) number	
Berkeley email	
Name of person to your left	
Name of person to your right	

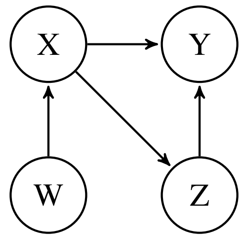
Honor Code [1 pt]:
As a member of the UC Berkeley community, I act with honesty, integrity, and respect for others. I am the person whose name is on the exam, and I completed this exam in accordance with the Honor Code.

Signature: ____

1 True or False [7 Pts]

For each of the following, determine whether the statement is true or false. For this question, no work will be graded and no partial credit will be assigned.

For parts (a) and (b), consider the following causal DAG:



(a) [1 Pt] When estimating the causal effect of Z on Y, the set of nodes $\{W, X\}$ satisfies the backdoor criterion and should therefore be treated as confounders.

◯ True ◯ False

(b) [1 Pt] When estimating the causal effect of X on Z, W satisfies the exclusion criterion and could therefore be used as an instrumental variable.

◯ True ◯ False

(c) [1 Pt] Value iteration is a dynamic programming algorithm for estimating the reward function given the best sequence of state/action pairs.

◯ True ◯ False

(d) [1 Pt] In a neural network, using backpropagation and stochastic gradient descent will always find the best set of parameters to minimize the loss on the training set.

◯ True ◯ False

(e) [1 Pt] In datasets with outliers, random forests will usually achieve worse test set accuracy than decision trees because bootstrap is sensitive to the presence of outliers.

◯ True ◯ False

(f) [1 Pt] In a Markov Decision Process (MDP), the reward for any action is conditionally independent of previous rewards obtained, given the current state.

◯ True ◯ False

(g) [1 Pt] Posterior normal approximations can be used for GLM coefficients regardless of whether the prior for those coefficients is normal or not.

◯ True ◯ False

2 Clickbait for Fun and Profit [6 Pts]

Jason has been experimenting with three different headline styles for a digital media site to maximize user engagement, measured by click-through rates. He is using multi-armed bandits to pick one of three different headline styles to show to website visitors, and has obtained the following data on how many clicks each style received so far:

Style	Times shown	Clicks received
A	30	6
B	10	2
C	20	5

The company knows that Jason has been using one of the algorithms Explore-then-Commit (ETC), Uniform Confidence Bound (UCB) or Thompson Sampling (TS). For each of the following, determine whether the statement is true or false. **For parts (a) - (d), no work will be graded.**

(a) [1 Pt] Based on the number of times each style has been shown, Jason is definitely not using ETC (as defined in class).

◯ True ◯ False

(b) [1 Pt] If using UCB with any confidence level δ, then he won't present A to the next visitor.

◯ True ◯ False

(c) [1 Pt] If Jason has been using UCB with a very small nonzero confidence level δ, then he will present Style C to the next visitor.

◯ True ◯ False

(d) [1 Pt] If Jason has been using UCB with a confidence level δ very close to 1, then he will present Style B to the next visitor.

◯ True ◯ False

(e) [2 Pts] Suppose Jason is using Thompson sampling with uniform priors. Fill in the blanks in the sequence of steps below to describe how he should determine what style to show to the next visitor. **No work outside the blanks will be graded.**

1. Draw one sample each from the following Beta distributions:

   (1) Beta(____, 25) (2) Beta(3, ____) (3) Beta(____, ____)

2. If the first sample is the largest, show Style ____.

   If the second sample is the largest, show Style ____.

   If the third sample is the largest, show Style ____.

3 Regressions on a Crime Dataset [9 Pts]

Consider the following dataset on arrests from the Introductory Econometrics book by Wooldridge, with information for $n = 2,725$ adult men on the following variables:

- `narr86` (y): Number of arrests in the year 1986 (this variable equals zero for 1,970 of the 2,725 men in the dataset)
- `pcnv` (x_1): Proportion of previous arrests that led to a conviction
- `tottime` (x_2): Total time (in months) in prison since turning 18
- `inc86` (x_3): Legal income in 1986 (in hundreds of dollars)
- `qemp86` (x_4): Number of quarters employed in 1986
- `black` (x_5): Binary variable which equals 1 if the individual is Black and 0 otherwise

(a) [2 Pts] We fit Poisson Regression (MODEL ONE) and Negative Binomial Regression (MODEL TWO) to this data in order to explore the relationship between y and x_1, x_2, x_3, x_4, x_5. The summaries of these model fits are given below.

```
                              MODEL ONE summary
==============================================================================
Dep. Variable:                 narr86   No. Observations:                 2725
Model:                            GLM   Df Residuals:                     2719
Model Family:                 Poisson   Df Model:                            5
Link Function:                    Log   Scale:                          1.0000
Method:                          IRLS   Log-Likelihood:                -2284.2
Date:                Sun, 05 Nov 2023   Deviance:                       2893.2
Time:                        23:37:58   Pearson chi2:                 4.25e+03
No. Iterations:                     6   Pseudo R-squ. (CS):             0.1093
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.5271      0.058     -9.053      0.000      -0.641      -0.413
pcnv          -0.4209      0.084     -4.987      0.000      -0.586      -0.255
tottime        0.0004      0.006      0.074      0.941      -0.010       0.011
inc86         -0.0086      0.001     -8.288      0.000      -0.011      -0.007
qemp86        -0.0030      0.029     -0.105      0.916      -0.059       0.053
black          0.4945      0.069      7.189      0.000       0.360       0.629
==============================================================================
```

```
                              MODEL TWO summary
==============================================================================
Dep. Variable:                 narr86   No. Observations:                 2725
Model:               NegativeBinomial   Df Residuals:                     2719
Method:                           MLE   Df Model:                            5
Date:                Sun, 05 Nov 2023   Pseudo R-squ.:                 0.04711
Time:                        23:38:02   Log-Likelihood:                -2182.8
converged:                       True   LL-Null:                       -2290.7
Covariance Type:            nonrobust   LLR p-value:                 1.174e-44
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.5038      0.074     -6.817      0.000      -0.649      -0.359
pcnv          -0.4806      0.103     -4.651      0.000      -0.683      -0.278
tottime        0.0027      0.007      0.383      0.702      -0.011       0.017
inc86         -0.0083      0.001     -7.174      0.000      -0.011      -0.006
qemp86        -0.0135      0.035     -0.387      0.698      -0.082       0.055
black          0.4926      0.088      5.570      0.000       0.319       0.666
alpha          1.0191      0.115      8.826      0.000       0.793       1.245
==============================================================================
```

Based only on the information given in the above summaries, which of these two models should be preferred for this dataset? You must justify your answer to receive credit.

◯ Poisson (MODEL ONE)

◯ Negative Binomial (MODEL TWO)

Justification:

(b) [2 Pts] Which of the following are correct interpretations of the coefficients of MODEL TWO? Select all answers that apply.

☐ A. For each additional month in prison since turning 18, the model predicts a 0.0027 increase in the log-odds of being arrested (if all other variables are held fixed).

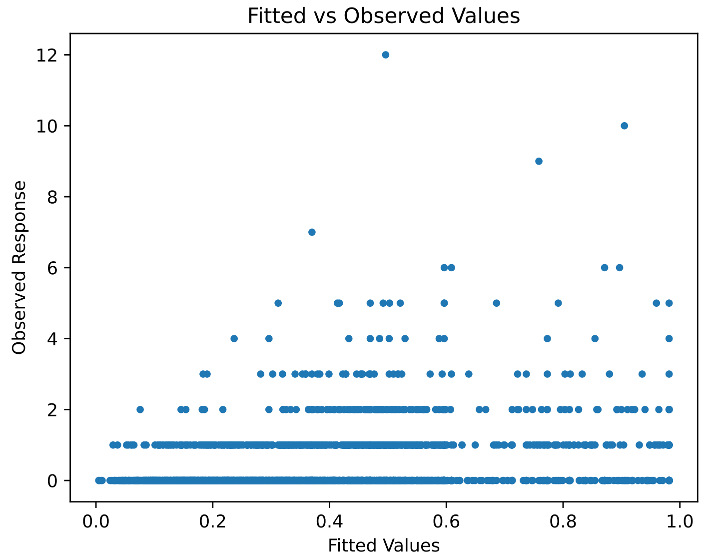
☐ B. It is possible that a 90% confidence interval for the coefficient of `inc86` could include 0.

☐ C. It is possible that a 99% confidence interval for the coefficient of `inc86` could include 0.

(c) [2 Pts] Consider now a third model (MODEL THREE) obtained by Negative Binomial regression for y on x_1, x_3, x_5 (in other words, the two variables `tottime` and `qemp86` are now dropped). The summary for this model is given below. Is MODEL THREE preferable to both MODEL ONE and MODEL TWO? You must provide a **numerical** justification for your answer to receive credit.

```
                          MODEL THREE Summary
==============================================================================
Dep. Variable:                 narr86   No. Observations:                 2725
Model:               NegativeBinomial   Df Residuals:                     2721
Method:                           MLE   Df Model:                            3
Date:                Mon, 06 Nov 2023   Pseudo R-squ.:                 0.04704
Time:                        00:40:27   Log-Likelihood:                -2182.9
converged:                       True   LL-Null:                       -2290.7
Covariance Type:            nonrobust   LLR p-value:                 1.884e-46
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.5167      0.062     -8.334      0.000      -0.638      -0.395
pcnv          -0.4776      0.103     -4.635      0.000      -0.680      -0.276
inc86         -0.0086      0.001    -10.860      0.000      -0.010      -0.007
black          0.4982      0.088      5.672      0.000       0.326       0.670
alpha          1.0174      0.115      8.821      0.000       0.791       1.243
==============================================================================
```

(d) [3 Pts] Based on the information above and the following plot of the observed response values (y) and the fitted values $\hat{y}$ given by MODEL TWO, which of the following statements are true? Select all answers that apply.



- ☐ A. MODEL TWO is unlikely to be useful for predicting the number of arrests when the number of arrests is large (e.g., 4 or more).
- ☐ B. The fitted values for MODEL ONE will include many values larger than 1.
- ☐ C. Consider the men for whom $y = 0$ (i.e., they have not been arrested in 1986), but their fitted value from MODEL TWO is more than 0.6. Almost all of them are Black, or have spent many years in prison since turning 18, or both.

4 Gambler vs Casino [9 Pts]

Bugsy goes to the casino and plays n independent games of chance. Each game, he has probability p of winning, and probability $1 - p$ of losing, where $0 < p \leq 0.5$. Let X be the number of times he wins. The casino uses concentration inequalities to guarantee that Bugsy won't win too often: in other words, they want to bound the probability that X/n is not much larger than p, with probability at least 0.95.

(a) [2 Pts] Using Chebyshev's inequality, show that

$$\mathbb{P}\left(\frac{X}{n} < p + \sqrt{\frac{20p(1-p)}{n}}\right) \geq 0.95$$

Hint: you should use the fact that if $\mathbb{P}(|a - b| < c) \leq q$*, then* $\mathbb{P}(a - b < c) \leq q$*.*

(b) [3 Pts] Using Hoeffding's inequality, show that

$$\mathbb{P}\left(\frac{X}{n} < p + \sqrt{\frac{\ln 20}{2n}}\right) \geq 0.95.$$

(c) [2 Pts] The casino manager argues that Hoeffding's inequality will always produce a "better" guarantee: in other words, that for the same probability 0.95, that $p+\sqrt{\frac{\ln 20}{2n}} < p+\sqrt{\frac{20p(1-p)}{n}}$. Find a value of p that proves the casino manager wrong. You do not need to find all such values of p: providing one is enough.

Hint: $\ln(20) \approx 3$

(d) [2 Pts] Which of the following statements are true? Select all answers that apply.

- ☐ A. The bound from Chebyshev's inequality is "tight": in other words, the true probability $\mathbb{P}\left(\frac{X}{n} < p + \sqrt{\frac{20p(1-p)}{n}}\right)$ is always at most 0.96.
- ☐ B. The bound from Hoeffding's inequality is "tight": in other words, the true probability $\mathbb{P}\left(\frac{X}{n} < p + \sqrt{\frac{\ln 20}{2n}}\right)$ is always at most 0.96.
- ☐ C. If we use Chernoff's bound with the Binomial MGF and the optimal value of λ, then we will obtain a better bound than the result from parts (a) or (b).

5 Can Taking Data 102 Make You Happier? [8 Pts]

The Data 102 staff want to determine whether taking Data 102 causes an increase in student happiness. They collect the following data from a random sample of Data Science majors:

- `happiness`: a number from 0 to 10 indicating how happy the student is
- `ds102`: whether or not the student has taken Data 102 (binary)
- `GPA`: the student's GPA
- `ds140`: whether or not the student has taken Data 140 (binary)

(a) [2 Pts] They assume that GPA is the only confounding variable for the effect of taking Data 102 on happiness. Which of the following statements, if true, would make this assumption incorrect? Select all answers that apply.

☐ A. Happiness causes an increase in GPA

☐ B. Taking Data 140 increases the chances of a student taking Data 102, and also increases their happiness

☐ C. Students who work together on a Data 102 project are likely to increase each others' happiness

For the remainder of this question, assume that they are correct, and that GPA is the only confounding variable for the effect of taking Data 102 on happiness.

(b) [3 Pts] Suppose their sample only contains four students, and their data is as follows. Compute the IPW estimate for the average treatment effect. You do not need to simplify arithmetic expressions to receive full credit.

`happiness`	`ds102`	`GPA`	`ds140`	propensity score
4	0	3.2	1	0.8
9	1	2.7	0	0.9
5	1	3.1	0	0.2
6	0	3.4	1	0.4

(c) [3 Pts] Alan learns that several semesters ago, Data 102 enrollment was cut in half just before the semester started. Half the students were removed at random, and were enrolled in an alternative course. He then discovers that some of the students who were removed got re-enrolled, because they didn't have prerequisites for the alternative course.

Can the staff use the random enrollment decision as an instrumental variable? For each of the assumptions necessary for instrumental variables, specify whether the random enrollment decision satisfies the criterion, and explain why or why not.

This page has been intentionally left blank.

6 Congratulations [0 Pts]

Congratulations! You have completed Midterm 2.

- **Make sure that you have written your student ID number on *every other page* of the exam.** You may lose points on pages where you have not done so.
- Also ensure that you have **signed the Honor Code** on the cover page of the exam for 1 point.
- If more than 10 minutes remain in the exam period, you may hand in your paper and leave. If $\leq$ 10 minutes remain, please **sit quietly** until the exam concludes.

[Optional, 0 pts] What's on your mind?

Midterm 2 Reference Sheet

Useful Distributions:

Distribution	Support	PDF/PMF	Mean	Variance	Mode
$X \sim \text{Poisson}(\lambda)$	$x = 0, 1, 2, \ldots$	$\frac{\lambda^x e^{-\lambda}}{x!}$	λ	λ	$\lfloor \lambda \rfloor$
$X \sim \text{Binomial}(n, p)$	$x \in \{0, 1, \ldots, n\}$	$\binom{n}{x} p^x (1-p)^{1-x}$	np	$np(1-p)$	$\lfloor (n+1)p \rfloor$
$X \sim \text{Beta}(\alpha, \beta)$	$0 \leq x \leq 1$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha}{\alpha+\beta} \frac{\beta}{\alpha+\beta} \frac{1}{\alpha+\beta+1}$	$\frac{\alpha-1}{\alpha+\beta-2}$
$X \sim \text{Gamma}(\alpha, \beta)$	$x \geq 0$	$\frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$	$\frac{\alpha}{\beta}$	$\frac{\alpha}{\beta^2}$	$\frac{\alpha-1}{\beta}$
$X \sim \mathcal{N}(\mu, \sigma^2)$	$x \in \mathbb{R}$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$	μ	σ^2	μ
$X \sim \text{Exponential}(\lambda)$	$x \geq 0$	$\lambda \exp(-\lambda x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	0

Conjugate Priors: For observations x_i, $i = 1, \ldots, n$:

Likelihood	Prior	Posterior
$x_i \mid \theta \sim \text{Bernoulli}(\theta)$	$\theta \sim \text{Beta}(\alpha, \beta)$	$\theta \mid x_{1:n} \sim \text{Beta}\left(\alpha + \sum_i x_i, \beta + \sum_i (1 - x_i)\right)$
$x_i \mid \mu \sim \mathcal{N}(\mu, \sigma^2)$	$\mu \sim \mathcal{N}(\mu_0, 1)$	$\mu \mid x_{1:n} \sim \mathcal{N}\left(\frac{\sigma^2}{\sigma^2+n}\left(\mu_0 + \frac{1}{\sigma^2}\sum_i x_i\right), \frac{\sigma^2}{\sigma^2+n}\right)$
$x_i \mid \lambda \sim \text{Exponential}(\lambda)$	$\lambda \sim \text{Gamma}(\alpha, \beta)$	$\lambda \mid x_{1:n} \sim \text{Gamma}\left(\alpha + n, \beta + \sum_i x_i\right)$

Generalized Linear Models

Regression	Inverse link function	Likelihood
Linear	identity	Gaussian
Logistic	sigmoid	Bernoulli
Poisson	exponential	Poisson
Negative binomial	exponential	Negative binomial

Some powers of e:

x	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
$y = e^x$	1.05	1.11	1.22	1.35	1.49	1.65	1.82	2.01	2.23	2.46	2.72