Econ 300 Section 42841 Exam I

University of Illinois at Chicago

2023-09-27

Name: _____

Instructions

Please read the following before starting the exam:

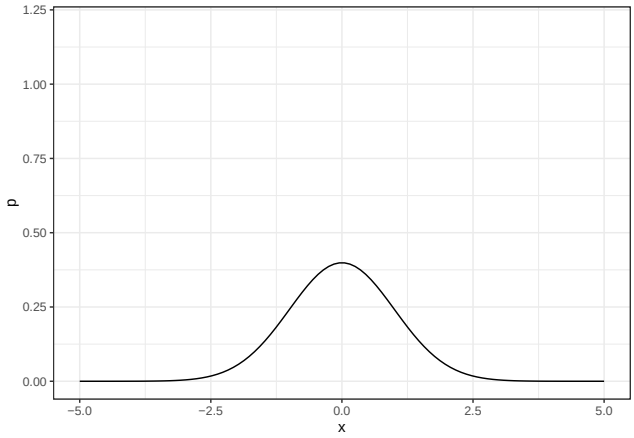
1. The exam is worth 100 points and consists of 6 multiple choice questions and 2 free response questions. Point values are included next to questions - please allocate your time appropriately
2. Please show all your work for full credit.
3. You may not communicate with classmates or other people about the exam.
4. Laptops, phones, and other electronics devices are not allowed during the exam
5. You may consult a single sheet of notes for the exam
6. A list of formulas is provided, not all of them will be relevant
7. The exam is due at the end of class. No exams will be accepted after this time.

Formulas / Notation

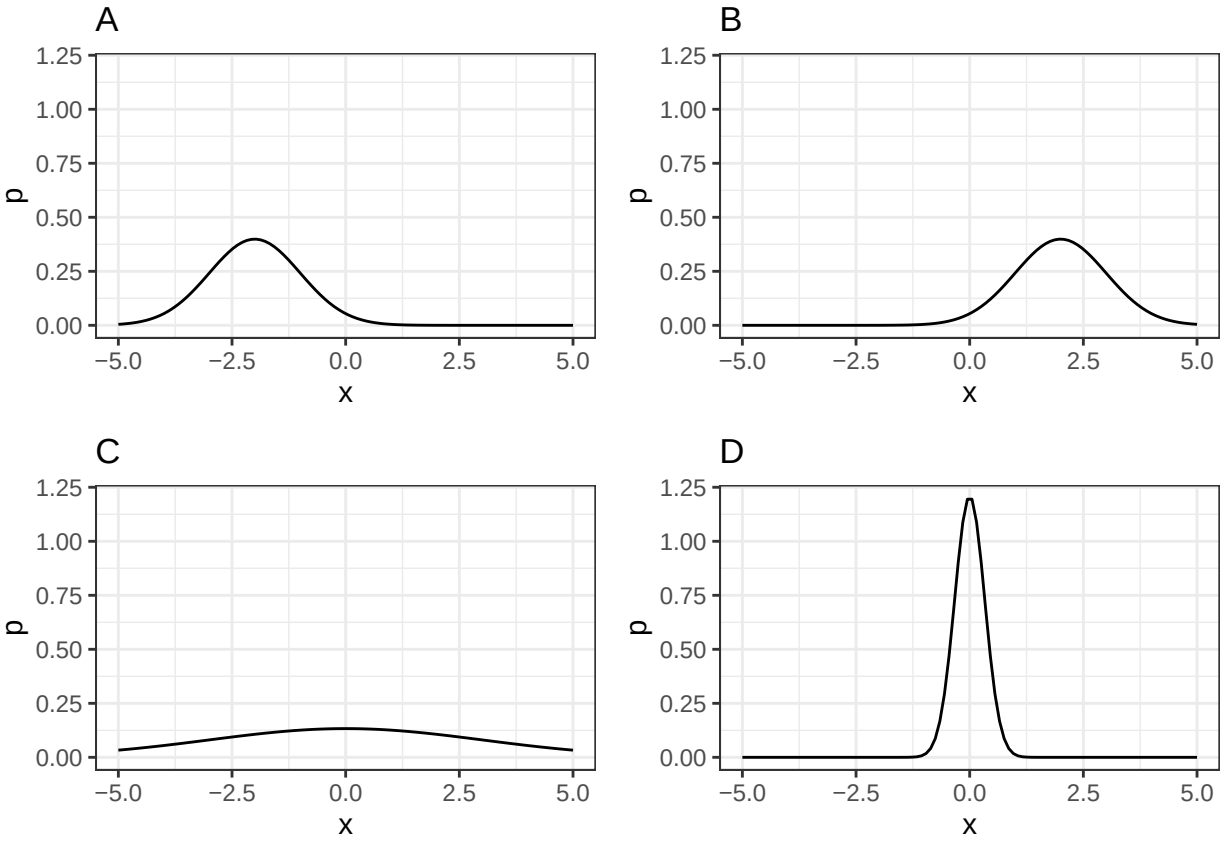
- $\beta_0, \beta_1, \varepsilon$ will refer to the intercept, slope, and error of a regression, respectively
- $E[x] = \frac{1}{n}\sum_i x_i$
- $var(x) = E[(x-\mu)^2]$
- $cov(x,y) = E[(x - E[X])(y - E[y])$
- $\rho_{xy} = \frac{cov(x,y)}{\sigma_x \sigma_y}$
- $\hat{\sigma}^2 = \sum_i (y_i - \hat{y}_i)^2$
- $\hat{\beta}_1 = cor(x,y)\frac{\sigma_y}{\sigma_x}$
- $E[\hat{\beta}_1] = \beta_1 + cor(x,\varepsilon)\frac{\sigma_\epsilon}{\sigma_x}$
- $var(\hat{\beta}) = \frac{\hat{\sigma}^2}{n\sigma_x^2}$

Question 1 (5 points)

You start with the following distribution

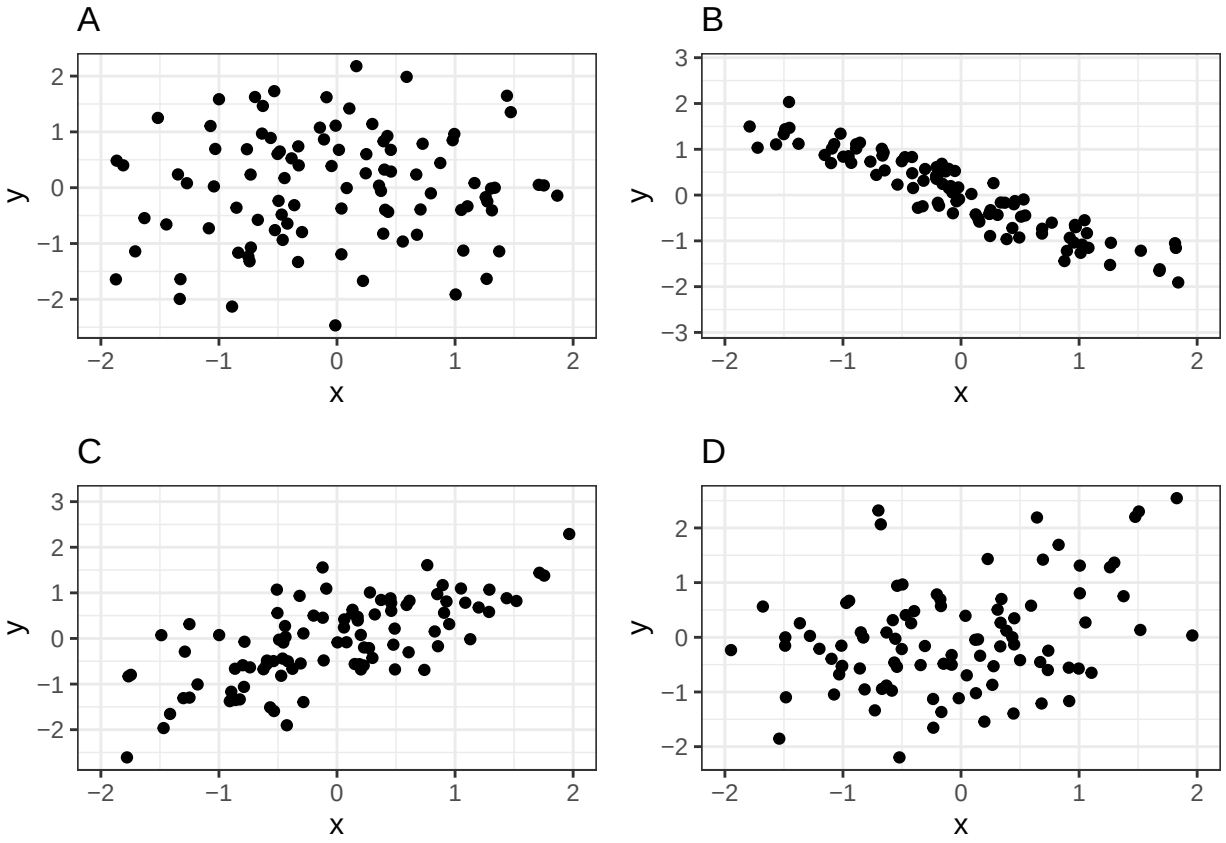


Which of the following distributions has a standard deviation 3 times as large as the first (note that the axes are all the same)?



Question 2 (5 points)

For which of the following plots is cor(x,y) the most positive



Question 3 (5 points)

The central limit theorem states that

- A Large populations tend to be normally distributed
- B As your sample size grows large, your estimate of the sample mean becomes unbiased
- C As your sample size grows large, the variance of your sample mean approaches zero
- D As your sample size grows large and you have independent and identically distributed samples from a finite variance distribution, your sample mean approaches a normal distribution

Question 4 (5 points)

You run a regression of y on x and find that $R^2 = 0.64$. Which of the following is true:

- A Since R^2 is high we know our estimate of $\hat{\beta}_1$ is probably unbiased
- B Since R^2 is low we know our estimate of $\hat{\beta}_1$ is probably biased
- C 64% of the variation in y is explained by x
- D Since R^2 is positive we know that as x increases y also tends to increase

Question 5 (5 points)

The true relationship between x and y is given by $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. We estimate $\hat{\beta}_1$. What does it mean for $\hat{\beta}_1$ to be biased?

- A $\hat{\beta}_1$ is biased whenever there are large error terms in the model
- B $\hat{\beta}_1$ is biased whenever $\hat{\beta}_1 \neq \beta_1$
- C $\hat{\beta}_1$ is biased whenever $E[\hat{\beta}_1] \neq \beta_1$
- D $\hat{\beta}_1$ is not biased under OLS regression

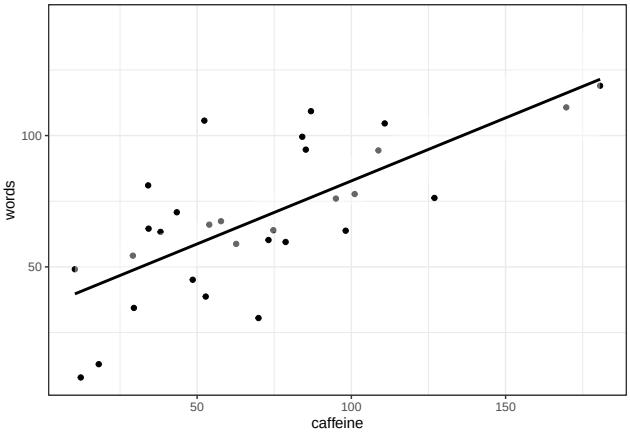
Question 6 (5 points)

When will the OLS regression given by $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ be biased?

- A When $cor(x, y) \neq 0$
- B When $cor(x, \varepsilon) \neq 0$
- C When $cor(y, \varepsilon) \neq 0$
- D OLS regression is not biased

Question 7 (35 points, 5 points per part)

You are estimating the effect of caffeine consumption (measured in mg) on the length of students' answers to a free response exam question (measured in words). You obtain information about students caffeine consumption immediately prior to the 8:00 AM exam via a survey question at the end of the exam. You use the regression equation $words_i = \beta_0 + \beta_1 caffeine_i + \varepsilon_i$, which you plot below along with the line of best fit, which has $\hat{\beta}_0 = 30, \hat{\beta}_1 = 0.5$



- A What is the dependent variable in this regression?

- B What is the predicted essay length for a student who took a Vivarin pill containing 200mg of caffeine prior to class?

- C Timmy consumes 200mg of caffeine and proceeds to hand in his exam after 15 minutes, having written 10 words in total. Calculate $\hat{\varepsilon}$ for this observation

- D You decide to swap your mountain dew (55mg of caffeine) for a red bull (80mg of caffeine). How many additional words are you expected to write on your essay?

- E What can be said about the correlation between caffeine and essay length (is it positive, negative, or zero)?

- F Do you expect your estimate of $\hat{\beta}_1$ in this regression to be biased? Why or why not?

- G You wish to test whether your error term is correlated with caffeine consumption, so you calculate $cor(caffeine, \hat{\varepsilon})$ using the residuals. Is this a good idea? Why or why not?

Question 8 (35 points)

You are estimating the effect of receiving paid time off (PTO, measured in days of work) on number of hours worked per year. You use data from several large companies in Chicago, and measure each employee's total amount of PTO and number of hours worked over a 5 year period

- A Write the regression equation / estimating equation / core model (5 points)

- B Suppose you obtain an estimate of $\hat{\beta}_0 = 1800, \hat{\beta}_1 = 10$. Interpret $\hat{\beta}_0$ and $\hat{\beta}_1$ within the context of this research question (10 points)

- C List two factors that are included in the error term and explain why they're in the error term (10 points)

- D Suppose that full-time salaried employees are more likely to be given PTO as part of their compensation package. Will this bias the estimate of $\hat{\beta}_1$? Why or why not? (10 points)

Extra Credit (11 points)

- For question 8, it is reasonable that being awarded paid time off will have a heterogeneous response based on employee type, e.g. highly compensated employees or employees subject to overtime pay may not reduce their labor supply in response to PTO while other workers may reduce their supply. Explain how β_1 accounts for this information and whether $\hat{\beta}_1$ is biased as a result of this (5 points)

- The Gauss-Markov theorem states that OLS regression gives the best linear unbiased estimate of β_1. Explain what defines the best estimate of β_1, and list at least 1 assumption required for Gauss-Markov to hold (5 points)

- How much caffeine, in mg, did you consume prior to this exam (1 point)