ECON306 – Final exam

2013 · 6 · 24

Name: _____ PSU ID: _____

There are 50 questions worth 2 points each. You have 1h 50min to answer all of them. Don't forget to write your name and PSU ID (e.g. bxs5142).

Part I. For problems 1–6, consider the random variables x, y with the following joint distribution

	$y=1$	$y=2$	$y=3$	$y=4$	$y=5$
$x=-1$	0.00	0.15	0.01	0.07	0.02
$x=0$	0.10	0.07	0.03	0.00	0.30
$x=1$	0.01	0.03	0.01	0.15	0.05

1. Find the marginal distribution of x.

2. Compute the expected value of x.

3. Compute the variance of x. [Hint: what is the expected value of x^2?]

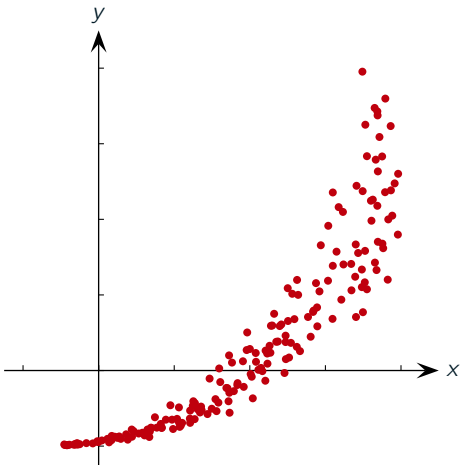
4. Find the distribution of x conditional on $y = 3$.

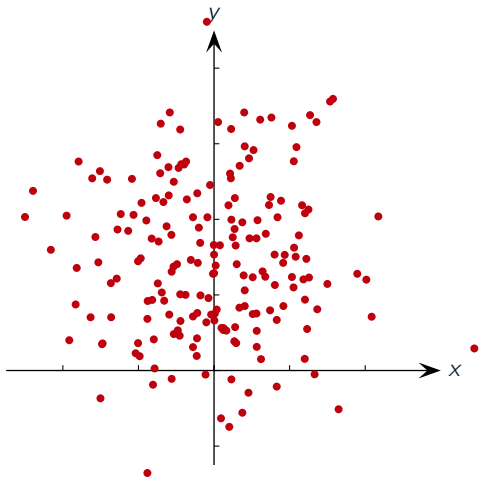
5. Compute the expected value of x conditional on $y = 3$.

6. Are x and y independent? Briefly justify your answer (1–2 sentences).

Part II. (Questions 7-30) For each of the following scatterplots involving random samples for variables x and y, answer the following questions:

- Do x and y appear to be independent?
- Would the OLS intercept estimate $\hat{\beta}_0$ be positive, negative, or close to 0?
- The relationship between x and y appears to be positive and significant, negative and significant, or insignificant?
- Would the R^2 coefficient of the OLS estimated model be big (> 0.8), small (< 0.2), or intermediate?
- Does the data generating process appear to be linear? If not, which transformation could you use to obtain a better fit?
- Is there any classical assumption other than correct specification which appears to be violated? (Write at most one.)





7. Independence: _____

8. β_0: _____

9. β_1: _____

10. R^2: _____

11. Transformation: _____

12. Assumptions: _____

13. Independence: _____

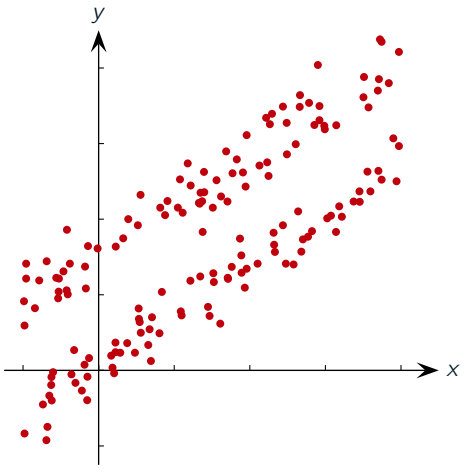
14. β_0: _____

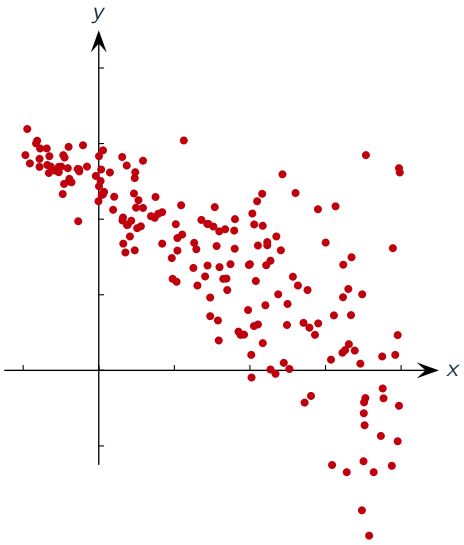
15. β_1: _____

16. R^2: _____

17. Transformation: _____

18. Assumptions: _____





19. Independence: _____

20. β_0: _____

21. β_1: _____

22. R^2: _____

23. Transformation: _____

24. Assumptions: _____

25. Independence: _____

26. β_0: _____

27. β_1: _____

28. R^2: _____

29. Transformation: _____

30. Assumptions: _____

Part III. For problems 31–38, consider the following model for the weight of a person:

$$\text{WEIGHT}_i = \beta_0 + \beta_1\,\text{HEIGHT}_i \cdot \text{WAIST}_i + \beta_2\,\text{BACK}_i \cdot \text{HEIGHT}_i + \beta_3\,\text{NECK}_i \cdot \text{HEIGHT}_i + \varepsilon_i$$

where:

$\text{WEIGHT}_i =$ weight measured in kilograms
$\text{HEIGHT}_i =$ height measured in centimeters
$\text{NECK}_i =$ neck girth measured in centimeters (indicates build)
$\text{BACK}_i =$ width, measured from shoulder to shoulder in centimeters (indicates bone structure)
$\text{WAIST}_i =$ waist diameter measured in centimeters (indicates build)

Suppose that you run a linear regression and obtain the following results (some values are missing)

weigh t	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
height*waist	-0.1632209			0.971		
height*neck	69.47699	2.290621				
height*back	33.42678				27.87425	38.97932
_cons	1.905555	7.177806	0.27	0.791	-12.17987	15.99098

31. Write down the estimated model (equation).

32. What is the average difference in weight between people with the same values for NECK, WAIST and BACK, whose HEIGHT differs by one inch?

33. Which variables are significant with 95% confidence?

34. How can you explain that the coefficient of WAIST · HEIGHT is negative?

35. What do you think would happen to R^2 if we excluded WAIST · HEIGHT from the regression?

36. What do you think would happen to *the adjusted* R^2? Briefly justify your answer (1–2 sentences).

37. What would happen to $\text{SE}(\hat{\beta}_1)$ if we increased the sample size?

Part IV. For problems 38–44, consider the following estimated model

$$\widehat{\text{WAGE}}_i = \hat{\beta}_0 + \hat{\beta}_1 \, \text{EXP}_i + \hat{\beta}_2 \, \text{DEG}_i + \hat{\beta}_3 \, \text{EXP}_i \cdot \text{DEG}_i$$

where:

$$\begin{aligned} \text{WAGE}_i &= \text{annual income in thousands of dollars} \\ \text{EXP}_i &= \text{years of professional experience} \\ \text{DEG}_i &= 1 \text{ if the individual has a college degree and 0 otherwise} \\ \text{SKILL}_i &= \text{(unobserved) innate ability} \end{aligned}$$

38. What sign (positive or negative) would you expect each *slope* coefficient to have? Briefly justify your answer (1–2 sentences per coefficient).

39. For people with 10 years of experience, what is the difference in average income between those who went to college and those who didn't?

40. How would the income of people with no college education and 10 years of experience change on average if they had gone to college? (They would have 4 less years of experience.)

41. There may be endogeneity because we omitted the variable SKILL. Which variables would you expect to be endogenous? Which estimates (from $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$) would you expect to be biased?

42. For each of the endogenous variables, would you expect a positive or a negative bias?

43. Provide an example of an instrumental variable that could be used to correct this bias.

44. What are the three requirements that a valid instrument should satisfy?

Part V. For problems 45–50, consider the following estimated model for Geico (auto insurance company):

$$\widehat{\text{SWITCH}}_i = 0.055 + 0.001\ \text{AD}_i + 0.152\ \text{ACC}_i - 0.011\ \text{INC}_i + \varepsilon_i$$

where:

$\text{SWITCH}_i =$ 1 if a person switched from his current insurance provider to Geico and 0 otherwise

$\text{AD}_i =$ 1 if the person received personalized advertisement from Geico and 0 otherwise

$\text{ACC}_i =$ 1 if the person made an insurance claim on the previous year

$\text{INC}_i =$ annual household income in thousands of dollars

45. What is the effect of personalized advertisement on the probability of switching to Geico?

46. What is the probability of switching for a person who received personalized advertisement, made no insurance claim last year, and whose household annual income is 60,000 USD?

47. How can you interpret this number? Are there any difficulties involved?

48. How could you transform the model to avoid these difficulties?

49. Do you see any potential source of endogeneity?

50. How could you solve it?