Review Questions

Econ 103

Spring 2018

About This Document

These questions are the "bread and butter" of Econ 103: they cover the basic knowledge that you will need to acquire this semester to pass the course. There are between 10 and 15 questions for each lecture. After a given lecture, and before the next one, you should solve all of the associated review questions. To give you an incentive to keep up with the course material, all quiz questions for the course will be randomly selected from this list. For example Quiz #1, which covers lectures 1–2, will consist of one question drawn at random from questions 1–10 and another drawn at random from questions 12–24 below. We will not circulate solutions to review questions. Compiling your own solutions is an important part of studying for the course. We will be happy to discuss any of the review questions with you in office hours or on Piazza, and you are most welcome to discuss them with your fellow classmates. Be warned, however, that merely memorizing answers written by a classmate is a risky strategy. It may get you through the quiz, but will leave you woefully unprepared for the exams. There is no curve in this course: to pass the exams you will have to learn the material covered in these questions. Rote memorization will not suffice.

Lecture #1 – Introduction

1. Define the following terms and give a simple example: *population*, *sample*, *sample size*.
2. Explain the distinction between a *parameter* and a *statistic*.
3. Briefly compare and contrast *sampling* and *non-sampling* error.
4. Define a *simple random sample*. Does it help us to address sampling error, non-sampling error, both, or neither?
5. A drive-time radio show frequently holds call-in polls during the evening rush hour. Do you expect that results based on such a poll will be biased? Why?

6. Dylan polled a random sample of 100 college students. In total 20 of them said that they approved of President Trump. Calculate the margin of error for this poll.
7. Define the term *confounder* and give an example.
8. What is a randomized, double-blind experiment? In what sense is it a "gold standard?"
9. Indicate whether each of the following involves experimental or observational data.
   (a) A biologist examines fish in a river to determine the proportion that show signs of disease due to pollutants poured into the river upstream.
   (b) In a pilot phase of a fund-raising campaign, a university randomly contacts half of a group of alumni by phone and the other half by a personal letter to determine which method results in higher contributions.
   (c) To analyze possible problems from the by-products of gas combustion, people with with respiratory problems are matched by age and sex to people without respiratory problems and then asked whether or not they cook on a gas stove.
   (d) An industrial pump manufacturer monitors warranty claims and surveys customers to assess the failure rate of its pumps.
10. Based on information from an observational dataset, Amy finds that students who attend an SAT prep class score, on average, 100 points better on the exam than students who do not. In this example, what would be required for a variable to *confound* the relationship between SAT prep classes and exam performance? What are some possible confounders?

Lecture #2 – Summary Statistics I

11. For each variable indicate whether it is nominal, ordinal, or numeric.
    (a) Grade of meat: prime, choice, good.
    (b) Type of house: split-level, ranch, colonial, other.
    (c) Income
12. Explain the difference between a histogram and a barchart.
13. Define *oversmoothing* and *undersmoothing*.
14. What is an *outlier*?
15. Write down the formula for the sample mean. What does it measure? Compare and contrast it with the sample median.

16. Two hundred students took Dr. Evil's final exam. The third quartile of exam scores was 85. Approximately how many students scored *no higher* than 85 on the exam?

17. Define *range* and *interquartile range*. What do they measure and how do they differ?

18. What is a boxplot? What information does it depict?

19. Write down the formula for variance and standard deviation. What do these measure? How do they differ?

20. Suppose that x_i is measured in inches. What are the units of the following quantities?
    (a) Sample mean of x
    (b) Range of x
    (c) Interquartile Range of x
    (d) Variance of x
    (e) Standard deviation of x

21. Evaluate the following sums:
    (a) $\sum_{n=1}^{3} n^2$
    (b) $\sum_{n=1}^{3} 2^n$
    (c) $\sum_{n=1}^{3} x^n$

22. Evaluate the following sums:
    (a) $\sum_{k=0}^{2} (2k+1)$
    (b) $\sum_{k=0}^{3} (2k+1)$
    (c) $\sum_{k=0}^{4} (2k+1)$

23. Evaluate the following sums:
    (a) $\sum_{i=1}^{3} (i^2+i)$

(b) $\displaystyle\sum_{n=-2}^{2}(n^2-4)$

(c) $\displaystyle\sum_{n=100}^{102} n$

(d) $\displaystyle\sum_{n=0}^{2}(n+100)$

24. Express each of the following using Σ notation:

(a) $z_1 + z_2 + \cdots + z_{23}$

(b) $x_1y_1 + x_2y_2 + \cdots + x_8y_8$

(c) $(x_1 - y_1) + (x_2 - y_2) + \cdots + (x_m - y_m)$

(d) $x_1^3 f_1 + x_2^3 f_2 + \cdots + x_9^3 f_9$

Lecture #3 – Summary Statistics II

25. Show that $\displaystyle\sum_{i=m}^{n}(a_i + b_i) = \sum_{i=m}^{n} a_i + \sum_{i=m}^{n} b_i$. Explain your reasoning.

26. Show that if c is a constant then $\displaystyle\sum_{i=m}^{n} cx_i = c\sum_{i=m}^{n} x_i$. Explain your reasoning.

27. Show that if c is a constant then $\displaystyle\sum_{i=1}^{n} c = cn$. Explain your reasoning.

28. Mark each of the following statements as True or False. You do not need to show your work if this question appears on a quiz, although you should make sure you understand the reasoning behind each of your answers.

(a) $\displaystyle\sum_{i=1}^{n}(x_i/n) = \left(\sum_{i=1}^{n} x_i\right)/n$

(b) $\displaystyle\sum_{k=1}^{n} x_k z_k = z_k \sum_{k=1}^{n} x_k$

(c) $\displaystyle\sum_{k=1}^{m} x_k y_k = \left(\sum_{k=1}^{m} x_k\right)\left(\sum_{k=1}^{m} y_k\right)$

(d) $\displaystyle\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{m} y_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} x_i y_j$

(e) $\left(\sum_{i=1}^{n} x_i\right) / \left(\sum_{i=1}^{n} z_i\right) = \sum_{i=1}^{n} (x_i/z_i)$

29. Show that $\sum_{i=1}^{n}(x_i - \bar{x}) = 0$. Justify all of the steps you use.

30. Re-write the formula for skewness in terms of the z-scores $z_i = (x_i - \bar{x})/s$. Use this to explain the original formula: why does it involve a cubic and why does it divide by s^3?

31. How do we interpret the sign of skewness, and what is the "rule of thumb" that relates skewness, the mean, and median?

32. What is the distinction between μ, σ^2, σ and $\bar{x}, s^2, s$? Which corresponds to which?

33. What is the empirical rule?

34. Define *centering*, *standardizing*, and *z-score*.

35. What is the sample mean $\bar{z}$ of the z-scores $z_1, \ldots, z_n$? Prove your answer.

36. What is the sample variance s_z^2 of the z-scores $z_1, \ldots, z_n$? Prove your answer.

37. Suppose that $-c < (a - x)/b < c$ where $b > 0$. Find a lower bound L and an upper bound U such that $L < x < U$.

38. Compare and contrast *covariance* and *correlation*. Provide the formula for each, explain the units, the interpretation, etc.

39. Suppose that x_i is measured in centimeters and y_i is measured in feet. What are the units of the following quantities?

    (a) Covariance between x and y

    (b) Correlation between x and y

    (c) Skewness of x

    (d) $(x_i - \bar{x})/s_x$

Lecture #4 – Regression I

40. In a regression using height (measured in inches) to predict handspan (measured in centimeters) we obtained $a = 5$ and $b = 0.2$.

    (a) What are the units of a?

    (b) What are the units of b?

(c) What handspan would we predict for someone who is 6 feet tall?

41. Plot the following dataset and calculate the corresponding regression slope and intercept *without* using the regression formulas.

x	y
0	2
1	1
1	2

42. Write down the optimization problem that linear regression solves.

43. Prove that the regression line goes through the means of the data.

44. By substituting $a = \bar{y} - b\bar{x}$ into the linear regression objective function, derive the formula for b.

45. Consider the regression $\widehat{y} = a + bx$.
    (a) Express b in terms of the sample covariance between x and y.
    (b) Express the sample correlation between x and y in terms of b.

46. What value of a minimizes $\displaystyle\sum_{i=1}^{n}(y_i - a)^2$? Prove your answer.

47. Suppose that $s_{xy} = 30$, $s_x = 10$, $s_y = 6$, $\bar{y} = 12$, and $\bar{x} = 4$. Calculate a and b in the regression $\widehat{y} = a + bx$.

48. Suppose that $s_{xy} = 30$, $s_x = 10$, $s_y = 6$, $\bar{y} = 12$, and $\bar{x} = 4$. Calculate c and d in the regression $\widehat{x} = c + dy$. Note: we are using y to predict x in this regression!

49. A large number of students took two midterm exams. The standard deviation of scores on midterm #1 was 16 points, while the standard deviation of scores midterm #2 was 17 points. The covariance of the scores on the two exams was 124 points squared. Linus scored 60 points on midterm #1 while Lucy scored 80 points. How much higher would we predict that Lucy's score on the midterm #2 will be?

50. Suppose that the correlation between scores on midterm #1 and midterm #2 in Econ 103 is approximately 0.5. If the regression slope when using scores on midterm #1 to predict those on midterm #2 is approximately 1.5, which exam had the larger *spread* in scores? How much larger?

Lecture #5 – Basic Probability I

51. What is the definition of probability that we will adopt in Econ 103?

52. Define the following terms:
    (a) *random experiment*
    (b) *basic outcomes*
    (c) *sample space*
    (d) *event*

53. Define the following terms and give an example of each:
    (a) *mutually exclusive events*
    (b) *collectively exhaustive events*

54. Suppose that $S = \{1, 2, 3, 4, 5, 6\}$, $A = \{2, 3\}$, $B = \{3, 4, 6\}$, and $C = \{1, 5\}$.
    (a) What is A^c?
    (b) What is $A \cup B$?
    (c) What is $A \cap B$?
    (d) What is $A \cap C$?
    (e) Are A, B, C mutually exclusive? Are they collectively exhaustive?

55. A family has three children. Let A be the event that they have less than two girls and B be the event that they have exactly two girls.
    (a) List all of the basic outcomes in A.
    (b) List all of the basic outcomes in B.
    (c) List all of the basic outcomes in $A \cap B$
    (d) List all of the basic outcomes in $A \cup B$.
    (e) If male and female births are equally likely, what is the probability of A?

56. Let $B = A^c$. Are A and B mutually exclusive? Are they collectively exhaustive? Why?

57. State each of the three axioms of probability, aka the *Kolmogorov Axioms*.

58. Suppose we carry out a random experiment that consists of flipping a fair coin twice.
    (a) List all the basic outcomes in the sample space.
    (b) Let A be the event that you get at least one head. List all the basic outcomes in A.
    (c) List all the basic outcomes in A^c.
    (d) What is the probability of A? What is the probability of A^c?

59. Calculate the following:

(a) $5!$

(b) $\dfrac{100!}{98!}$

(c) $\dbinom{5}{3}$

60. (a) How many different ways can we choose a President and Secretary from a group of 4 people if the two offices must be held by different people?

    (b) How many different committees with two members can we form a group of 4 people, assuming that the order in which we choose people for the committee doesn't matter.

61. Suppose that I flip a fair coin 5 times.

    (a) How many basic outcomes contain exactly two heads?

    (b) How many basic outcomes contain exactly three tails?

    (c) How many basic outcomes contain exactly one heads?

    (d) How many basic outcomes contain exactly four tails?

62. Explain why $\dbinom{n}{r} = \dbinom{n}{n-r}$.

63. Suppose that I choose two distinct numbers at random from the set $\{1, 2, 3, 4, 5, 6, 7, 8, 9\}$. What is the probability that both are odd?

Lecture #6 – Basic Probability II

64. State and prove the *complement rule*.

65. State the *multiplication rule*, and compare it to the definition of conditional probability.

66. Mark each statement as TRUE or FALSE. If FALSE, give a one sentence explanation.

    (a) If $A \subseteq B$ then $P(A) \geq P(B)$.

    (b) For any events A and B, $P(A \cap B) = P(A)P(B)$.

    (c) For any events A and B, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

67. Suppose that $P(B) = 0.4$, $P(A|B) = 0.1$ and $P(A|B^c) = 0.9$.

    (a) Calculate $P(A)$.

    (b) Calculate $P(B|A)$.

68. Define statistical independence. How is it related to conditional probability, and what does it mean intuitively?

69. State and prove the law of total probability for $k = 2$.

70. Find the probability of getting *at least* one six if you roll a fair, six-sided die three times.

71. Suppose a couple decides to have three children. Assume that the sex of each child is independent, and the probability of a girl is 0.48, the approximate figure in the US.
    (a) How many basic outcomes are there for this experiment? Are they equally likely?
    (b) What is the probability that the couple has *at least one* girl?

72. Let A and B be two arbitrary events. Use the addition rule and axioms of probability to establish the following results.
    (a) Show that $P(A \cup B) \leq P(A) + P(B)$. (This is called *Boole's Inequality.*)
    (b) Show that $P(A \cap B) \geq P(A) + P(B) - 1$. (This is called *Bonferroni's Inequality*)

73. Let A and B be two mutually exclusive events such that $P(A) > 0$ and $P(B) > 0$. Are A and B independent? Explain why or why not.

74. Molly the meteorologist determines that the probability of rain on Saturday is 50%, and the probability of rain on Sunday is also 50%. Adam the anchorman sees Molly's forecast and summarizes it as follows: "According to Molly we're in for a wet weekend. There's a 100% chance of rain this weekend: 50% on Saturday and 50% on Sunday." Is Adam correct? Why or why not?

75. Suppose I throw two fair, six-sided dice once. Define the following events:

$$\begin{aligned} E &= \text{The first die shows 5} \\ F &= \text{The sum of the two dice equals 7} \\ G &= \text{The sum of the two dice equals 10} \end{aligned}$$

    (a) Calculate $P(F)$.
    (b) Calculate $P(G)$.
    (c) Calculate $P(F|E)$.
    (d) Calculate $P(G|E)$.

Lecture #7 – Basic Probability III / Discrete RVs I

76. What is the base rate fallacy? Give an example.

77. Derive Bayes' Rule from the definition of conditional probability.

78. What are two names for the *unconditional* probability in the numerator of Bayes' rule?

79. When is it true that $P(A|B) = P(B|A)$? Explain.

80. Of women who undergo regular mammograms, two percent have breast cancer. If a woman has breast cancer, there is a 90% chance that her mammogram will come back positive. If she does *not* have breast cancer there is a 10% chance that her mammogram will come back positive. Given that a woman's mammogram has come back positive, what is the probability that she has breast cancer?

81. The Triangle is a neighborhood that once housed a chemical plant but has become a residential area. Two percent of the children in the city live in the Triangle, and fourteen percent of these children test positive for excessive presence of toxic metals in the tissue. For children in the city who do not live in the Triangle, the rate of positive tests is only one percent. If we randomly select a child who lives in the city and she tests positive, what is the probability that she lives in the Triangle?

82. Three percent of *Tropicana* brand oranges are already rotten when they arrive at the supermarket. In contrast, six percent of *Sunkist* brand oranges arrive rotten. A local supermarket buys forty percent of its oranges from *Tropicana* and the rest from *Sunkist*. Suppose we randomly choose an orange from the supermarket and see that it is rotten. What is the probability that it is a *Tropicana*?

83. Define the terms *random variable*, *realization*, and *support set*.

84. What is the probability that a RV takes on a value outside of its support set?

85. What is the difference between a *discrete* and *continuous* RV?

86. What is a *probability mass function*? What two key properties does it satisfy?

Lecture #8 – Discrete RVs II

87. Define the term *cumulative distribution function* (CDF). How is the CDF of a discrete RV X related to its pmf?

88. Let X be a RV with support set $\{-1, 1\}$ and $p(-1) = 1/3$. Write down the CDF of X.

89. Write out the support set, pmf, and CDF of a Bernoulli(p) RV.

90. Define the term *parameter* as it relates to a random variable. Are parameters constant or random?

91. Let X be a RV with support set $\{0, 1, 2\}$, $p(1) = 0.3$, and $p(2) = 0.5$. Calculate $E[X]$.

    Let X be a discrete RV. Define the expected value $E[X]$ of X. Is $E[X]$ constant or random? Why?

92. Suppose X is a RV with support $\{-1, 0, 1\}$ where $p(-1) = q$ and $p(1) = p$. What relationship must hold between p and q to ensure that $E[X] = 0$?

93. Let X be a discrete RV and a, b be constants. Prove that $E[a + bX] = a + bE[X]$.

94. Suppose that $E[X] = 8$ and $Y = 3 + X/2$. Calculate $E[Y]$.

95. Suppose that X is a discrete RV and g is a function. Explain how to calculate $E[g(X)]$. Is this the same thing as $g\left(E[X]\right)$?

96. Let X be a RV with support set $\{-1, 1\}$ and $p(-1) = 1/3$. Calculate $E[X^2]$.

97. Let X be a RV with support set $\{2, 4\}$, $p(2) = 1/2$ and $p(4) = 1/2$. Mark each of the following claims as TRUE or FALSE, either by appealing to a result from class, or by directly calculating both sides of the equality.
    (a) $E[X + 10] = E[X] + 10$
    (b) $E[X/10] = E[X]/10$
    (c) $E[10/X] = 10/E[X]$
    (d) $E[X^2] = \left(E[X]\right)^2$
    (e) $E[5X + 2]/10 = \left(5E[X] + 2\right)/10$

Lecture #9 – Discrete RVs III

98. Define the *variance* and *standard deviation* of a RV X. Are these constant or random?

99. Explain how to use our formula for $E[g(X)]$ to calculate the variance of a discrete RV.

100. Write down the shortcut formula for variance, and use it to calculate $Var(X)$ where $X \sim \text{Bernoulli}(p)$.

101. Let X be a random variable and a, b be constants. Prove that $Var(a+bX) = b^2 Var(X)$.

102. Define the Binomial(n, p) RV in terms of independent Bernoulli trials, and write down its support set and probability mass function.

103. Substitute $n = 1$ into the pmf of a Binomial(n, p) RV and show that you obtain the pmf of a Bernoulli(p) RV.

104. A multiple choice quiz has 12 questions, each of which has 5 choices. To pass you need to get at least 8 of them correct. Nina forgot to study, so she simply guesses at random.

   (a) Let the random variable X denote the number of questions that Nina gets correct on the quiz. What kind of random variable is X? Specify all parameter values.

   (b) Calculate the probability that Nina passes the quiz.

Lecture #10 – Discrete RVs IV

105. Define the *joint probability mass function* p_{XY} of two discrete RVs X and Y and list its two key properties.

106. What is the difference between a joint pmf and a marginal pmf? Can you calculate a marginal pmf from a joint? How? Can you calculate a joint pmf from a marginal pmf?

107. Suppose that X is a random variable with support $\{1, 2\}$ and Y is a random variable with support $\{0, 1\}$ where X and Y have the following joint pmf:

$$\begin{aligned} p_{XY}(1,0) &= 0.20, \quad & p_{XY}(1,1) &= 0.30 \\ p_{XY}(2,0) &= 0.25, \quad & p_{XY}(2,1) &= 0.25 \end{aligned}$$

   (a) Express the joint pmf in a table with X in the *rows*, as we did in class.

   (b) Using the table, calculate the marginal pmfs of X and Y.

108. The question relies upon the following joint pmf:

		X	
		0	1
Y	1	0.1	0.2
	2	0.3	0.4

   (a) Calculate the conditional pmf of Y given that $X = 0$.

   (b) Calculate the conditional pmf of X given that $Y = 2$.

109. The question relies upon the following joint pmf:

		X	
		0	1
Y	1	0.1	0.2
	2	0.3	0.4

(a) Calculate $E[XY]$.

(b) Calculate $Cov(X, Y)$.

110. This question relies on the following joint pmf:

		Y		
		-1	0	1
X	0	1/9	1/9	0
	1	2/9	1/9	1/9
	2	0	1/9	2/9

(a) Calculate $p_Y(0)$.

(b) Calculate $p_{X|Y}(2|0)$.

(c) Calculate $E[XY]$.

(d) Are X and Y independent? Why or why not?

111. Prove the shortcut formula for variance: $Var(X) = E[X^2] - (E[X])^2$.

112. Prove that $Cov(X, Y) = E[XY] - E[X]E[Y]$. Hint: the steps are similar to our derivation of the shortcut formula for variance from class.

113. Let X and Y be two random variables with $E[X] = 2$ and $E[Y] = 1$. Calculate $E[X - Y]$.

114. Suppose $E[X] = 2$ and $Var(X) = 5$. Calculate $E[X^2]$.

115. Let X and Y be two random variables with $Var(X) = 2$, $Var(Y) = 1$, and $Cov(X, Y) = 0$. Calculate $Var(X - Y)$.

116. Let X and Y be two random variables with $Var(X) = \sigma_X^2$, $Var(Y) = \sigma_Y^2$, and $Cov(X, Y) = \sigma_{XY}$. If a, b, c are constants, what is $Var(cX + bY + a)$?

117. Suppose that X and Y are two random variables with correlation $\rho = 0.3$, and standard deviations $\sigma_X = 4$ and $\sigma_Y = 5$.

(a) Calculate $Cov(X, Y)$.

(b) Let $Z = (X + Y)/2$. Calculate $Var(Z)$.

118. What does it mean for a sequance of random variables $X_1, X_2, \ldots, X_n$ to be "independent and identically distributed (iid)?"

119. Mark each statment as TRUE or FALSE. If FALSE, explain.

(a) The expected value of a sum $E[X_1 + X_2 + \cdots + X_n]$ is *not* in general equal to the sum of the expected values $E[X_1] + E[X_2] + \cdots + E[X_n]$. But when $X_1, X_2, \ldots, X_n$ are independent then the two are equal.

(b) The variance of a sum $Var(X_1 + X_2 + \cdots + X_n)$ is always equal to the sum of the variances $Var(X_1) + Var(X_2) + \cdots + Var(X_n)$.

120. Suppose that $X \sim \text{Binomial}(n, p)$.

    (a) Explain how X can be defined in terms of Bernoulli(p) RVs.

    (b) Using the preceding part, derive $E[X]$.

    (c) Using the preceding part, derive $Var(X)$.

121. Suppose that $X \sim \text{Binomial}(9, 1/3)$ and $Y \sim \text{Binomial}(4, 1/2)$. Calculate $E[(Y - X)/2]$.

Lecture #11 – Continuous RVs I

122. If X is a continuous RV and a, b are constants, how do we calculate $P(a \leq X \leq b)$?

123. What are the two properties of a probability density function?

124. True or False: since $f(x)$ is a probability, $0 \leq f(x) \leq 1$.

125. How is the PDF of a continuous RV related to its CDF?

126. Let X be a continuous RV with CDF F. Express $P(-2 \leq X \leq 4)$ in terms of F.

127. Let X be a continuous RV with CDF F. Express $P(X \geq x_0)$ in terms of F.

128. Suppose that X is a continuous RV with support set $[-1, 1]$.

    (a) Is 2 a possible realization of this RV?

    (b) What is $P(X = 0.5)$?

    (c) True or False: $P(X \leq 0.3) = P(X < 0.3)$. Explain.

129. Let X be a Uniform$(0, 1)$ RV. Calculate the CDF of X.

130. Let X be a Uniform$(0, 1)$ RV. Calculate $Var(X)$.

131. Let X be a Uniform(a, b) RV. Calculate $E[X]$

132. Let X be a continuous RV with support $[0, 1]$ and $f(x) = Cx^2(1 - x)$. Find C.

133. Let X be a continuous RV with support $[0, 1]$ and $f(x) = 3x^2$. Find the CDF of X.

134. Let X be a continuous RV with support $[0, 1]$ and $f(x) = 3x^2$. Calculate $Var(X)$.

135. Let X be a continuous RV with support $[0, 1]$ and $f(x) = 3x^2$. Calculate the probability that X takes a value in the interval $[0.2, 0.8]$.

136. Let X be a RV with support set $[-2, 2]$ and the following CDF:

$$F(x_0) = \begin{cases} 0, & x_0 < -2 \\ x_0/4, & -2 \leq x_0 \leq 2 \\ 1, & x_0 > 2 \end{cases}$$

(a) Calculate the PDF of X.

(b) Is X an example of the RVs from the lecture slides? If so, which one? Be sure to specify the values of any and all parameters of the distribution.

Lecture #12 – Continuous RVs II