

# 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 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)  $\sum_{i=1}^3 (i^2 + i)$

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

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

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

24. Express each of the following using  $\Sigma$  notation:

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

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

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

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

## Lecture #3 – Summary Statistics II

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

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

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

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

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

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

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

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

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  $\mu, \sigma^2, \sigma$  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  $z_1, \dots, z_n$ ? Prove your answer.
36. What is the sample variance  $s_z^2$  of the z-scores  $z_1, \dots, z_n$ ? 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  $\hat{y} = a + bx$ .
- Express  $b$  in terms of the sample covariance between  $x$  and  $y$ .
  - Express the sample correlation between  $x$  and  $y$  in terms of  $b$ .
46. What value of  $a$  minimizes  $\sum_{i=1}^n (y_i - a)^2$ ? Prove your answer.
47. Suppose that  $s_{xy} = 30$ ,  $s_x = 5$ ,  $s_y = 3$ ,  $\bar{y} = 12$ , and  $\bar{x} = 4$ . Calculate  $a$  and  $b$  in the regression  $\hat{y} = a + bx$ .
48. Suppose that  $s_{xy} = 30$ ,  $s_x = 5$ ,  $s_y = 3$ ,  $\bar{y} = 12$ , and  $\bar{x} = 4$ . Calculate  $c$  and  $d$  in the regression  $\hat{x} = c + dy$ . 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)  $\frac{10!}{98!}$
  - (c)  $\binom{5}{3}$
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  $\binom{n}{r} = \binom{n}{n-r}$ .
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

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:
- $E$  = The first die shows 5
  - $F$  = The sum of the two dice equals 7
  - $G$  = The sum of the two dice equals 10
- (a) Calculate  $P(F)$ .
  - (b) Calculate  $P(G)$ .
  - (c) Calculate  $P(F|E)$ .
  - (d) Calculate  $P(G|E)$ .

## Lecture #7

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

87. Define the term *cumulative distribution function* (CDF). How is the CDF of a discrete random variable  $X$  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 random variable. 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 random variable 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  $g(E[X])$ ?
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 random variable 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)  $E[X^2] = (E[X])^2$
  - (e)  $E[5X + 2]/10 = (5E[X] + 2)/10$

## Lecture #9

98. Define the *variance* and *standard deviation* of a random variable  $X$ .
99. Fill in the missing details from class to calculate the variance of a Bernoulli Random Variable *directly*, that is *without* using the shortcut formula.
100. Substitute  $n = 1$  into the pmf of a Binomial( $n, p$ ) RV and show that you obtain the pmf of a Bernoulli( $p$ ) RV.
101. Let  $X$  be a RV that takes on the values 1, 2, and 3 with equal probability and never takes on any other values.

- (a) What is the CDF of  $X$ ? Write it out as a piecewise function.
  - (b) Calculate the expected value of  $X$ .
102. Suppose that  $X_1, X_2, \dots, X_n \sim \text{iid Bernoulli}(p)$ . Define  $S_n = \sum_{i=1}^n X_i$ .
- (a) Write down the pmf and support of  $X_1$ .
  - (b) Calculate  $E[X_1]$ .
  - (c) Calculate  $E[X_1^2]$ .
  - (d) Calculate  $\text{Var}(X_1)$  using the shortcut formula.
  - (e) What kind of RV is  $S_n$ ? Write down its pmf and support.
103. Suppose we carry out a sequence of independent Bernoulli trials, each with probability of success  $p$ , and stop as soon as we get the first success.
- (a) What is the probability that we get a success on our first trial?
  - (b) What is the probability that we get our first success on the *second* trial? (That is, what is the probability of a Failure followed by a Success?)
  - (c) What is the probability that we get our first success on the  $n$ th trial?
  - (d) Suppose that we define a random variable  $X$  that equals the trial number of the first success in a sequence of independent Bernoulli trials, each with probability  $p$  of success. This is the definition of a Geometric( $p$ ) random variable. What is the probability mass function  $p(x) = P(X = x)$  of  $X$ ? What is the support of this random variable?

## These will be used as extension problems

104. This question refers to the prediction market example from lecture. Imagine it is October 2012. Let  $O$  be a contract paying \$10 if Obama wins the election, zero otherwise, and  $R$  be a contract paying \$10 if Romney wins the election, zero otherwise. Let  $\text{Price}(O)$  and  $\text{Price}(R)$  be the respective prices of these contracts.
- (a) Suppose you *buy* one of each contract. What is your profit?
  - (b) Suppose you *sell* one of each contract. What is your profit?
  - (c) What must be true about  $\text{Price}(O)$  and  $\text{Price}(R)$ , to prevent an opportunity for statistical arbitrage?
  - (d) How is your answer to part (c) related to the Complement Rule?
  - (e) What is the implicit assumption needed for your answers to parts (a)–(c) to be correct? How would your answers change if we were to relax this assumption?

105. You have been entered into a very strange tennis tournament. To get the \$10,000 Grand Prize you must win at least two sets *in a row* in a three-set series to be played against your Econ 103 professor and Venus Williams alternately: professor-Venus-professor or Venus-professor-Venus according to your choice. Let  $p$  be the probability that you win a set against your professor and  $v$  be the probability that you win a set against Venus. Naturally  $p > v$  since Venus is much better than your professor! Assume that each set is independent.
- Let W indicate win and L indicate lose, so that the sequence WWW means you win all three sets, WLW means you win the first and third set but lose the middle one, and so on. Which sequences of wins and losses land you the Grand Prize?
  - If you elect to play the middle set against Venus, what is the probability that you win the Grand Prize?
  - If you elect to play the middle set against your professor, what is the probability that you win the Grand prize?
  - (5 points) To maximize your chance of winning the prize, should you choose to play the middle set against Venus or your professor?
106. Rossa and Rodrigo are playing their favorite game: matching pennies. The game proceeds as follows. In each round, both players flip a penny. If the flips match (TT or HH) Rossa gets one point; if the flips do not match (TH or HT) Rodrigo gets one point. The game is best of three rounds: as soon as one of the players reaches two points, the game ends and that player is declared the winner. Since there's a lot of money on the line and graduate students aren't paid particularly well, Rossa secretly alters each of the pennies so that the probability of heads is  $2/3$  rather than  $1/2$ . In spite of Rossa's cheating, the individual coin flips remain independent.
- (6 points) Calculate the probability that Rossa will win the first round of this game.
  - Calculate the probability that the game will last for a full three rounds.
  - Calculate the probability that Rodrigo will win the game.
  - Yiwen is walking down the hallway and sees Rodrigo doing his victory dance: clearly Rossa has lost in spite of rigging the game. Given that Rodrigo won, calculate the probability that the game lasted for three rounds.
107. A plane has crashed in one of three possible locations: the mountains ( $M$ ), the desert ( $D$ ), or the sea ( $S$ ). Based on its flight path, experts have calculated the following prior probabilities that the plane is in each location:  $P(M) = 0.5$ ,  $P(D) = 0.3$  and  $P(S) = 0.2$ . If we search the mountains then, given that the plane is actually there, we have a 30% chance of *failing* to find it. If we search the desert then, given that the plane is actually there, we have a 20% chance of *failing* to find it. Finally, if we search the sea

then, given that the plane is actually there, we have a 90% chance of *failing* to find it. Naturally if the plane is *not* in a particular location but we search for it there, we will not find it. You may assume that searches in each location are independent. Let  $F_M$  be the event that we *fail* to find the plane in the mountains. Define  $F_D$  and  $F_S$  analogously.

- (a) (10 points) We started by searching the mountains. We did not find the plane. What is the conditional probability that the plane is nevertheless in the mountains? Explain.
- (b) (20 points) After failing to find the plane in the mountains, we searched the desert, and the sea. We did not find the plane in either location. After this more exhaustive search what is the conditional probability that the plane is in the mountains? Explain.

108. I have an urn that contains two red balls and three blue balls. I offer you the chance to play the following game. You draw one ball at a time from the urn. Draws are made at random and *without replacement*. You win \$1 for each red ball that you draw, but lose \$1 for each blue ball that you draw. You are allowed to stop the game at any point. Find a strategy that ensures your expected value from playing this game is *positive*.
109. An ancient artifact worth \$100,000 fell out of Indiana Jones's airplane and landed in the Florida Everglades. Unless he finds it within a day, it will sink to the bottom and be lost forever. Dr. Jones can hire one or more helicopters to search the Everglades. Each helicopter charges \$1000 per day and has a probability of 0.9 of finding the artifact. If Dr. Jones wants to maximize his *expected value*, how many helicopters should he hire?