Study guide for first midterm exam

Joshua Loftus

Suggested studying practices:

- Read the lecture notes posted on the course page. You can mostly skip over `R` code, since **there will be no coding on the exam**. You might want to look at the output from the code, like for some of the plots, since that may help understand the given example. These notes should be sufficient to answer any exam problems, but if you want additional reading to help with your understanding of the material in the notes see the next two references.
- Any topics covered in lectures may appear on the exam. This includes, but is not limited to: RCTs vs observational studies, summary statistics, probability definitions, addition and multiplication rules, conditional probability, independence, Bayes' rule, Bernoulli and Binomial random variables, expectation and variance, and linearity.
- If you are referencing the textbook Statistics by Freedman, Pisani, and Purves, the relevant chapters for material we have covered so far are 1-4, 13-15, and sections 1-2 of chapter 18 (but note that they refer to probability distribution functions as "probability histograms", a term we have not used).
- Another possible reference is the OpenIntro Statistics book available free here: https://drive.google.com/file/d/0B-DHaDEbiOGkc1RycUtIcUtIelE/view in particular, Chapter 2 on probability and 3.4.1 on the Binomial distribution. This book includes exercises and solutions.

Practice questions

1. A certain company sells a supplement for cold relief. To prove the supplement is effective, they commission a lab to do a study. Participants suffering from common cold symptoms are recruited into the study and given the supplement for free. They are asked to record if their symptoms improved after taking the supplement. At the conclusion of the study, about 70% of the participants said their symptoms improved after taking the supplement, another 25% didn't respond to the follow up, and 5% said their symptoms got worse. Do you think the study was effective at proving the supplement relieves symptoms? Why or why not?

2. Use the definition $\mathbf{Var}(X) = E[(X - E[X])^2]$ to show $\mathbf{Var}(X) = E[X^2] - E[X]^2$.

3. Fact: If X and Y are independent then $E[XY] = E[X]E[Y]$. Use this fact to show that if X and Y are independent then $\mathbf{Var}(X + Y) = \mathbf{Var}(X) + \mathbf{Var}(Y)$.

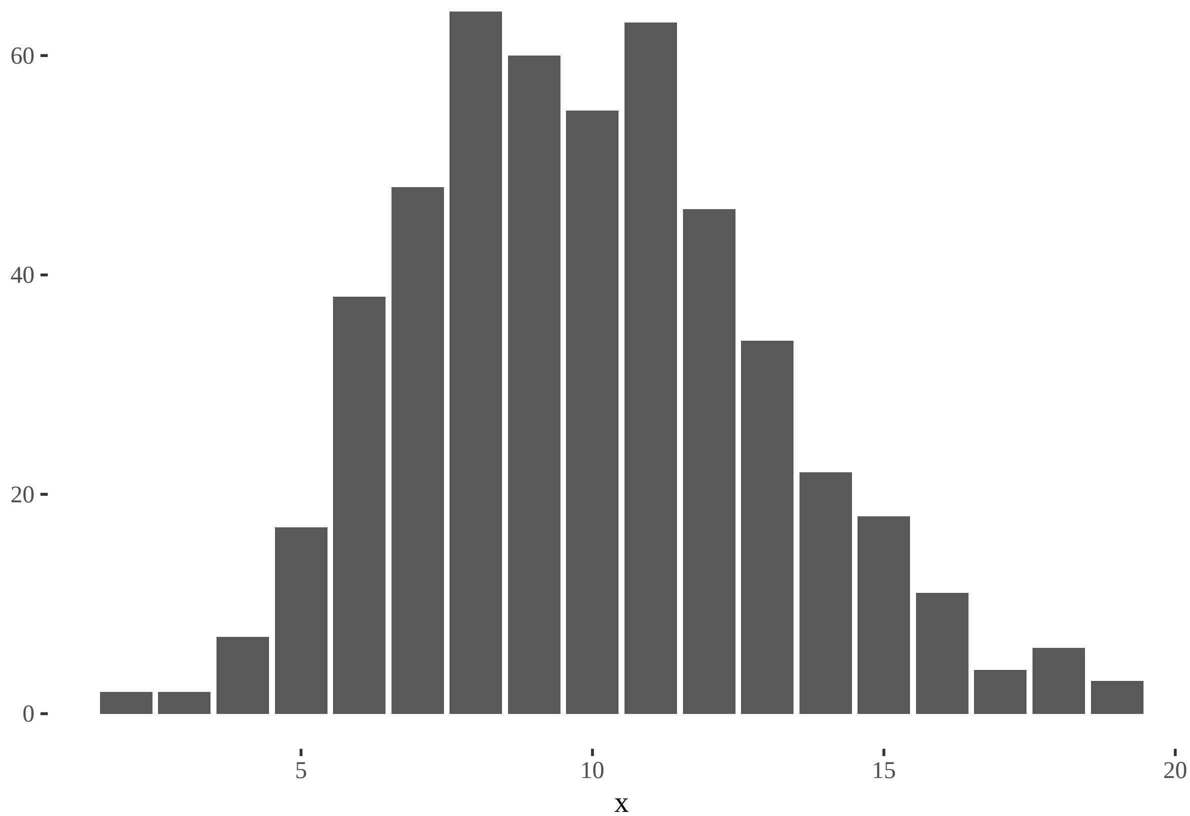
4. Suppose $U \sim \mathbf{Bin}(n, p)$ and $V \sim \mathbf{Bin}(m, p)$ are independent Binomials. What are $E[U + V]$ and $\mathbf{Var}(U + V)$?

5. According to the Department of Health and Human Services, African American children are 4 times more likely to be admitted to the hospital for asthma compared to non-Hispanic white children. A certain politician concludes from this that asthma must have a genetic explanation related to race, therefore we cannot expect any public health policy to reduce the difference in health outcomes. Leaving aside *ethical* questions, can you find any *statistical* errors in his argument?

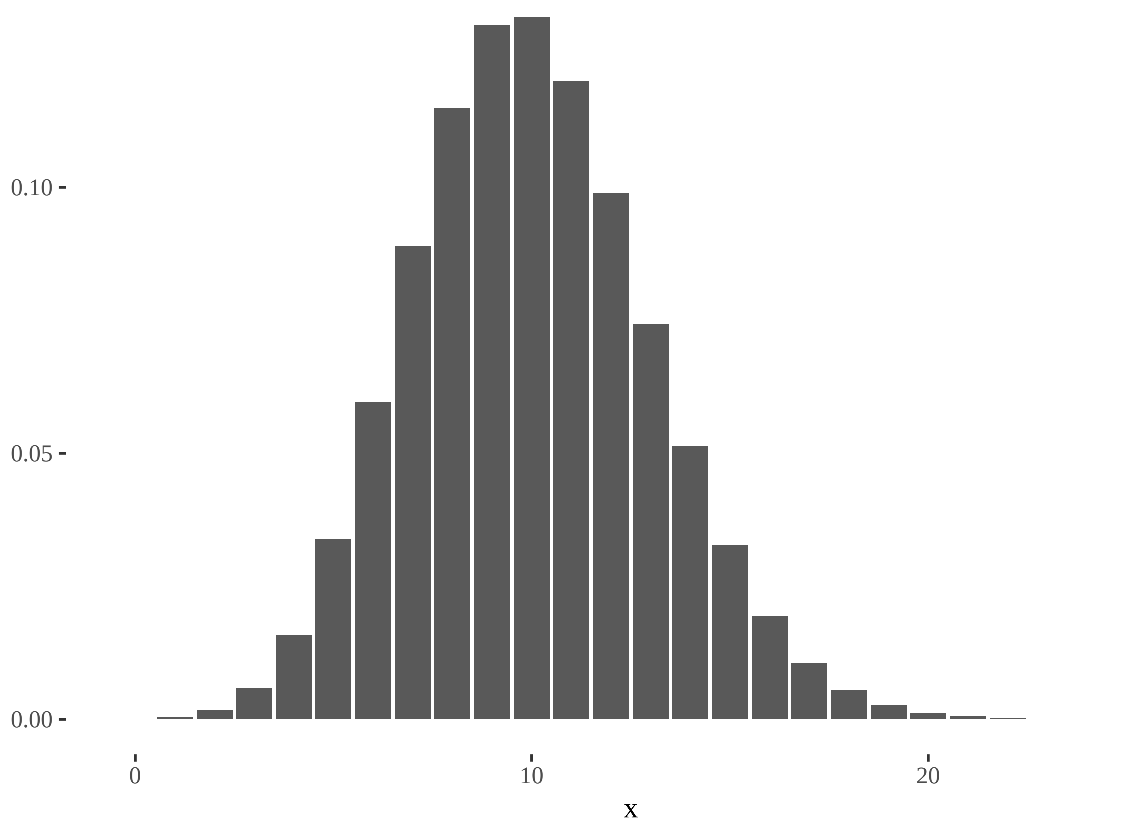
6. A credit card company's algorithm uses *big data* and *machine learning* to detect fraudulent transactions and automatically warn customers that their card may have been stolen. The algorithm is very accurate: $P(\text{detection}|\text{fraud}) = 0.999$, but it sometimes makes errors $P(\text{detection}|\text{normal activity}) = 0.01$. You receive a message that the bank has detected fraudulent activity on your account. You wonder: what is the probability that your account has experienced fraudulent activity given that the bank's algorithm says it was detected. Could you calculate this probability, or do you need more information?

7. You and your friend are sharing a pizza. There are 8 slices, half are cheese and half are pepperoni. As you are eating, each time you take a new slice you pick it randomly. You and your friend each finish half of the pizza. Is Bin$(4, 1/2)$ a good model for the number of pepperoni slices that you eat? Explain.

8. One of the following plots is the probability distribution function (pdf) of a Binomial random variable, and one is a histogram of random samples from the same Binomial distribution. Which one is the pdf, first or second? Can you guess the expected value? Do you think $\sigma^2 = 4$ is a good guess for the variance, or too high/low?



60
40
20
0
5
10
15
20
x



0.10
0.05
0.00
0
10
20
x

9. A dice game: roll a regular 6-sided die. If it lands on 6, you win. If it lands on 1, you lose. If it lands anywhere in between, you roll the dice a second time, and if the sum of the two rolls is greater than 6 you win. What is the probability that you win *on the second roll*? What is the probability that you win on any roll?

10. Suppose $W_1, W_2, \ldots, W_{100}$ are independent and identically distributed, with $E[W_1] = 1 + \mu/2$ and Var$(W_1) = 4$. What is the standard deviation of $\bar{W}$? Is $\bar{W}$ an unbiased estimator of μ?

11. Continuing problem 11 above, suppose that instead of a sample of size 100 we now continue gathering new observations of W_i, for $i = 101, 102, \ldots$. How large does the sample have to be for the standard deviation of $\bar{W}$ to be as low as 1/100?