Lab Normal

Abhi Thanvi, Paul Holaway

July 7th, 2022

Contents

Lab Normal

Welcome

Just like learning a new spoken language, you will not learn the language without practice. Labs are an important part of this course. Collaboration on labs is **extremely encouraged**. If you find yourself stuck for more than a few minutes, ask a neighbor or course staff for help. When you are giving help to your neighbor, explain the **idea and approach** to the problem without sharing the answer itself so they can figure it out on their own. This will be better for them and for you. For them because it will stick more and they will have a better understanding of the concept. For you because if you can explain it to other students, that means you understand it better too.

The Idea of this Lab

The idea behind this lab is to understand different distributions and how to calculate respective probabilities for each distribution. These distributions are a huge part of the statistical world, especially considering how many of the probability questions boil down to identifying distribution. In this class, we focus on small, yet important slice of the distributions. This lab consists of Binomial, Bernoulli, and Normal distribution and it is key to understand the importance of identifying the differences among these distributions and how to calculate their respective probabilities. Distributions might seem intimidating at first, but with some practice and an open mind, it is very doable. We personally think you all are more than capable at succeeding with the distributions we present! :)

"The key to understanding statistics is to understand the distribution. . . of the topics in the textbook" - Silly Abhi

Problem 1: TSA Pre-Check!

It is the year 2030, and the airport officers have become cautious of some Data Scientist who are a menace to the society because of how knowledgeable they are. The officers said it cannot be Abhi, because "he ain't a menace". Abhi feels offended by this and wants prove he is good with distributions! He feels rusty, but he has his apple watch where he can see some answers. Help him out in identifying some distributions. Help Abhi prove that he can also be a "menace to the society" ;)

Abhi: Hey guys and girls! Thanks for the help, I am running out of time so let's dive in!

Question 1: Bob has 39 coins in his pocket. He makes a bet with his sister, Jill, of how he can randomly pull a coin out of his pocket, toss it in the air, and get heads. Jill is naive and thinks this is impossible! She thinks the chances of Bob pulling it off is very low. Consider the success to be Bob pulling it off successfully. What kind of a distribution is this and why?

Answer: This is a Bernoulli distribution, because there is only one trial (flipping one coin) and there are only two outcomes (heads or tails) with one being "success" (heads) and one being "fail" (tails).

Question 2: What is the probability of Bob pulling it off successfully? Jill is young so she needs to see some numbers plugged into equation. So, show your work!

Answer: The probability of Bob succeeding is .50, because the probability of getting heads on a fair coin is 50% always.

$P(x) = p^x(1-p)^{1-x}$ $x = 1$ because getting heads is considered "success", p is .50 because that's the probability of getting a "heads" on a fair coin. $P(heads) = 0.50^1(1-0.50)^{1-1} = 0.50^1 * 0.50^0 = 0.50 * 1 = 0.50$

Question 3: Is Jill correct about the chances of Bob successfully pulling it off is very low?

Answer 3: No, Jill is incorrect about the chances, because there's a significant chance (50 percent!) that Bob will succeed. However, there is also an equally significant chance that Bob can fail, too, so Bob shouldn't rashly be making a bet like this!

Question 4: Bob is smart, he wants to increase his chances of winning the bet. He changes his bet and states that he will now be pulling a coin and tossing them 10 times. He says that he will get 3 heads and 7 tails, in any order of the 10 tosses. How has Bob changed the distribution?

Answer 4: Bob has made this a binomial distribution, since even though there is still a success (heads) and failure (tails) with the same probabilities (0.5 each), he's doing more than 1 trial (10) and predicting a specific number of heads and tails from those tosses.

Question 5: What is the probability of Bob winning the bet. In other words, what is the probability of getting 3 Heads in some order within the 10 tosses. Consider heads to be success and tails to be failure.

*Hint: To do 5! (factorial) in R, you can code `factorial(5)`. **Answer 5:** Let the discrete variable X represent the number of heads that Bob tosses. $P(X = 3) = \binom{10}{3} * 0.5^3(1 - 0.5)^{10-3}$

```
#Using R is optional, but it is helpful
choose(10, 3) * 0.5^3 * (1-0.5)^(10-3)
```

```
## [1] 0.1171875
```

Question 6: Oh boy! Did Bob just put an axe on his feet? In other words, did Bob really increase his chances of winning the bet when he changed the conditions of his bet?

Answer 6: Once Bob changed the conditions of the bet, his chances for winning plummeted. The probability of getting EXACTLY three heads out of ten tosses (≈ 0.1171875) is much smaller than the probability of getting a head on one flip (0.5).

Problem 2: The Normal Life

Question 1: Rex and Stacy have taken the SAT and wanted to compare their scores. The issue is that Rex took his SAT when the SAT was out of 2400. Stacy's SAT was the newer one and out of 1600. Think about how we can compare their scores. Discuss this with your group and report how you can compare their scores.

Answer: You could compare their scores by comparing the z-scores of their scores. In other words, you could subtract the mean score in their year from their score (to center their score distribution at zero) and then divide their scores by the standard deviation of the scores from their year (to make the spread of their scores equal, so that both scores can be analyzed on the same scale - not on separate 1600 and 2400 scales as was shown before). We should do this instead of simply dividing Rex's score by 1.5 to bring him to the 1600 scale (as would seem to be the easier method), because there's a chance that his SAT was harder or easier than Stacy's SAT - if so, his score would be influenced by this strategy. We need to employ the use of both mean and standard deviation in the z-score method so that we can see the spread of both SAT scores without any bias from test difficulty changes.

Question 2: Rex received a 2100 on the test. College Board stated that the mean of scores during Rex's year was 1950 and standard deviation of 120. Calculate the z-score for Rex.

Answer: $Z = \frac{x-\mu}{\sigma} = \frac{2100-1950}{120} = \frac{150}{120} = 1.25$

Question 3: Stacy received a 1500 on the test. College Board stated that the mean of scores during Stacy's year was 1350 and standard deviation of 100. Calculate the z-score for Stacy.

Answer: $Z = \frac{x-\mu}{\sigma} = \frac{1500-1350}{100} = \frac{150}{100} = 1.5$

Question 4: What is the probability that Rex and Stacy would receive a score that is at least what they got? We already have the Z-scores so this part should be quick.

Answer: (Student Response Here)
$P(Rex > 2100) = 1 - P(Rex < 2100) = 1 - P(Z < 1.25) = 1 - 0.8944 = 0.1056$
$P(Stacy > 1500) = 1 - P(Stacy < 1500) = 1 - P(Z < 1.5) = 1 - 0.9332 = 0.0668$

Question 5: Who did better on the SAT? Why?

Answer: Stacy did better on the SAT, because there's a much lower probability that she could've gotten a better score. In addition, she had a higher z-score than Rex, which means that on an equal scale and with adjustment for variation in test difficulty between the tests (which z-score testing accomplishes through the use of adjusting scores based on μ and σ), Stacy did better on the test.

Feedback

Hey this is Abhi! As the halfway mark is completed, I would like to know (again) whether you are liking the course or you hate your summer because of us (hopefully not!).

Please give some feedback of what you like about the course and what you would like to change about this course! Do you like something in particular or do not like something? We are open to all ideas and want to have the best time together! Have a great weekend and Paul and I will see you on the other side :)

Feedback: The course is great! It's really fun and it's nice seeing some hands-on explanations! Maybe it would be nice to see more of these visuals (like maybe, for explaining CLT, could you use an online simulation like an online quincunx to show how binomial distribution approaches the Gaussian?) and some more math units (like more about distributions). Awesome!

https://www.mathsisfun.com/data/quincunx.html (Set size to 15 and speed to 200)

Submission

Once you have finished your lab...

1. Go to the top left and click `File` and `Save`.
2. Click on the `Knit` button to convert this file to a PDF.
3. Submit **BOTH** the `.Rmd` file and `.pdf` file to Blackboard by 11:59 PM tonight.