Lady Tasting Coffee: A Case Study in Experimental Design

by

Jacinth Maynard, Department of Mathematics, Lock Haven University of Pennsylvania
Mary Puterbaugh Mulcahy, Division of Biological and Health Sciences, University of Pittsburgh at Bradford
Daniel Kermick, Division of Biological and Health Sciences, University at Pittsburgh at Bradford

Part I – Coffee Shop Wager

Characters

Model—an attractive young woman with impeccable taste, working as a successful model for an advertising firm.

Escort—a tall, dark and handsome young man, working as a marketing and survey researcher for the same advertising firm.

Older Gentleman—an adjunct professor of biostatistics at a local college and a coffee house regular; dressed in slightly-aged tweed jacket.

Setting

The red sports car makes a quick stop in front of the Philadelphia coffeehouse after an evening on the town. A sleek model and her tall, dark, and handsome escort gracefully exit the car and approach the counter where the escort purchases two cups of house coffee. At the condiments table, the escort proceeds to pour milk into the two coffees, and the following dialogue results.

Dialogue

Model:	Oh, no. I'll need a fresh cup.
Escort:	You like milk in your coffee, right?
Model:	Yes, I like milk in my coffee, but only, and I say only, if it is added to the cup first.
Escort (laughing):	Oh, come on, that's ridiculous; you can't possibly tell the difference. Coffee is coffee whether you add the milk first or second.
Model:	Of course I can tell the difference; you've no right to laugh at me.
Older Gentleman:	Pardon me for overhearing your conversation. Actually, the lady may be able to tell the difference.
Escort:	You're kidding me, right?
Older Gentleman:	Such claims have been made before. A woman made that claim at an afternoon tea party in Cambridge in the 1920s. She stated that

she could always tell whether milk was added before or after the tea was poured in the cup. The famous statistician Sir Ronald Fisher was at high tea* that afternoon and immediately designed an experiment to test the woman's palate. Rumor has it that she shocked the guests because she correctly told Sir Fisher whether milk had been added first or second after tasting multiple cups of tea. The event became famous because Sir Fisher used it to explain the basics of experimental design in one of the first textbooks ever published on the topic.

Escort: Well that's a good story, but it also sounds suspiciously like a story with no basis in reality, invented by an imaginative professor writing a textbook. You certainly haven't convinced me to buy the lady another cup of coffee.

Older Gentleman: From what I've read, it was a real event, and furthermore, humans have very sensitive taste buds. I think she deserves a new cup.

Model: Now that the coffee is cold, I certainly deserve a fresh cup.

Older Gentleman: Ma'am I fully agree. In fact, I think you deserve several cups of fresh coffee! Let's set up our own experiment to determine whether this lady can or can not discern whether milk was added to the cup before or after the coffee.

Model: How fun! Yes, let's see who is right with our own tea test, I mean coffee test.

Older Gentleman: Clever pun, ma'am. You know the real student's t-test was developed in the early 1900s by a taste tester of sorts at the Guinness Brewery in Dublin, Ireland. Only he wasn't tasting tea either. Sir, are you a gambling man? How about a wager? If our experiment shows that she can tell the difference, you will pay for the coffee. If the experiment does not demonstrate her discerning palate, I will pay for the fresh cups.

Escort: You're on! But I want the rules hashed out before she starts sipping. I mean, how many cups are we talking about? And what if she's wrong for just a few cups? I don't want to pay for the coffee just because she is a good guesser.

Older Gentleman: Fisher would agree completely. Even the smallest experiment requires forethought and planning. You must tell me, sir, just how sure do you want to be that she isn't guessing?

*Going all the way back to the 1600s, many English people have enjoyed an afternoon tea. The phrase "high tea" refers to that traditional practice of people gathering in the late afternoon to drink tea, usually served with some sort of meat, bread, crackers, or other food.

Questions

Imagine that you are going to design and perform the experiment described in the dialogue.

1. What is the hypothesis that will be tested in this experiment?
2. Why is it important to offer the model more than just two cups (one with the milk added first and one with the milk added second)? Explain your answer.
3. How many cups of coffee should the model taste? Explain your answer.
4. Describe exactly how the cups should be prepared. Does every cup need to be exactly the same in every way except the order of the addition of milk and coffee? Can you actually make every cup identical? Explain your answer fully.
5. In what order should the cups be presented? What method or decision rules might you use to decide which cup is to be offered first, second, etc.?
6. How do you recommend that the characters decide if the model is able to tell whether the milk was added to the cup before or after the coffee? (In other words, how many cups does she have to correctly evaluate for you to conclude that she really can tell the difference?) Explain your choice.
7. Without looking it up in a textbook or online, provide your own definition of "experimental design."

Part II – Tasting Tea

Read Fisher's essay entitled "Mathematics of a Lady Tasting Tea," available online at http://legacy.library.ucsf.edu/tid/fqi22e00/pdf; then answer the questions below.

Questions

The questions below ask you to compare your answers from Part I to the explanations found in Fisher's essay.

1. Referring to your answers to the questions in Part I, was the hypothesis you chose different than the null hypothesis given by Fisher in his essay? Explain your answer.
2. What reason did you give for why it is important to offer the model more than just two cups (one with the milk added first and one with the milk added second)? Was your answer the same as Fisher's answer? Based on the essay, please describe Fisher's answer to this question.
3. How many cups did you say the model would taste? How many cups did Fisher say that the "tea lady" in the story should taste? Please describe fully Fisher's answer to this question, including any mathematical considerations. Was your answer the same as Fisher's answer? If not, how is it different?
4. Before reading Fisher's essay, did you think that every cup needed to be exactly the same in every way except the order of the addition of milk and coffee? Does Fisher believe that every cup should be prepared identically? Describe Fisher's explanation for how to deal with uncontrollable variation among cups.
5. Before reading Fisher's essay, what method did you recommend for choosing the order that the cups should be presented? Was your answer different than what Fisher recommends in his essay? Please describe Fisher's explanation for how to choose the cup order.
6. In Part I, you answered the question of how one should decide if the model is able to tell whether the milk was added to the cup before or after the coffee. (In other words, how many cups does she have to correctly evaluate for you to conclude that she really can tell the difference?) Compare your answer to Fisher's answer and then describe Fisher's answer fully.
7. Your instructor will provide you with a textbook definition of experimental design. Was your definition complete? Your answers to all of the questions in Part I probably differed in small or large ways from Fisher's proposed design of a tea tasting experiment. Are there any differences that changed your perception of statistics and experimental design? If yes, describe how Fisher's essay enlightened you. Even if you did not find that Fisher's essay changed your views, make a short summary list of the important design concepts that you think are emphasized in Fisher's essay.
8. Following the guidance of your instructor, use Part III of the story or, alternatively, if your instructor has provided materials, set up a mock event similar to the tea tasting experiment. Instead of tea, you might consider seeing if your classmates can tell the difference between 1% and 2% milk, between two brands of bottled water, between two brands of flavored diet or regular soda, or some other simple taste comparison. Adhere to the principles of Fisher's paper, and draw conclusions based on your results, using Fisher's rules.
9. Rumor has it that the real lady who had tea with Fisher on an afternoon in the 1920s was able to accurately tell whether the milk was added first or second every single time she was offered a new cup. Can you conclude from her success that most people can tell the difference between milk-first and tea-first cups? Briefly describe the design of an experiment that would test this broader question.

Part III – Tasting Coffee

Setting

The escort and the older gentleman prepared four cups of coffee to which they added the milk first and four cups of coffee in which they added the milk second. The two men attempted to make sure that the cups were identically prepared as much as possible in terms of the amount of coffee, milk, etc. By writing the numbers 1–8 on slips of paper and pulling the numbers out of the hat, the escort and older gentleman randomly decided that cups #2, 3, 6, and 7 would be prepared with the milk added first and cups #1, 4, 5, and 8 would be prepared with the coffee added first. The model has just finished sipping all cups (presented in order 1–8), and she has written down her decision.

Dialogue

Older Gentleman:	Are we all in agreement that we will use Fisher's rules for deciding the conclusion for the experiment?
Escort:	Yes, we are. Please tell us which cups you think had the milk added first.
Model (smiling broadly):	It is so obvious that cups #2, 3, 4, and 7 had the milk first.
Older Gentleman:	Are you sure? Is this your final answer?
Model:	Yes, those are the cups that had the milk added first.
Escort:	Ha! Based on Fisher's rules, you can't really tell the difference! The milk was added first to cups #2, 3, 6, and 7. Hot dog, mister! YOU will have to pay for the cups!
Model:	But I got 6 of the 8 correct! That's pretty good! Who is this Fisher guy to say I can't do it?
Older Gentleman:	I have to agree ma'am that you may be able to tell the difference, but we did decide beforehand that we would use Fisher's rules, and his rules are quite strict. We can only conclude that you can tell the difference if there was less than a 5% chance that your success could be explained by good guessing.
Model:	But I didn't guess! I really can tell; it seems unfair that you accuse me of guessing just because I made one mistake!
Escort:	We aren't really accusing you of guessing; the 5% rule is the common cutoff that is used in many statistical analyses.
Model:	Fisher's rules are too strict. Give me another two cups and, if I get it right this time, my success rate will be better than the 5% guess rate. I'll show you that I really am able to tell the difference!
Escort:	Hold on! It would be cheating to change the plan now. No more cups. Please, let's quit while I'm ahead!

Questions

1. Did the characters correctly follow Fisher's rules when they concluded that the results do not allow us to conclude that the woman can tell the difference between the two types of cups?
2. Do you think the lady really can tell the difference between the milk-first versus the milk-second cups?
3. Which of the following sentences do you think most accurately and clearly states the conclusion of the experiment? Propose your own statement if you find flaws in all of the statements below.
   a. The model cannot tell the difference between the milk-first and the milk-second cups.
   b. There is a 5% or greater chance that the woman guessed her answers.
   c. At the 5% level, the model cannot significantly tell the difference between the milk-first and the milk-second cups.
4. Would it be acceptable to add two more cups now? Why or why not? What is the value of deciding the experimental design before you begin an experiment and not changing it in the middle of the experiment?
5. Do you think Fisher's rules are too strict? Why or why not?
   a. Would you feel the same way about his rules if we were testing whether a monitor at a nuclear power plant can really recognize elevated levels of radiation?
   b. Would you feel the same way if we were testing whether a new children's vitamin caused increased risk of kidney dysfunction?
   c. In order to answer these questions better, read the more detailed descriptions of these two hypothetical experiments given below. For each design, state the null hypothesis. There are two types of mistakes that statisticians can make at the conclusion of a significance test: they can incorrectly reject a true null hypothesis (Type I error) or they can incorrectly fail to reject a false null hypothesis (Type II error). The 5% rule ensures that a Type I error is never made at a greater rate than 5%, and the likelihood of making a Type I error is often inversely related to the likelihood of making a Type II error. Type II error rates are usually not controlled in scientific experiments, and can be considerably higher than 5%. For the following two experiments, what would be the human ethical consequence of making Type I and Type II mistakes in your conclusion at the end of the experiment? Would you recommend the 5% rule for these experiments? Why or why not?
      i. Nuclear Monitor Experimental Design: The monitor is exposed to eight environments (four high and four low radiation levels) and the experimenter records whether the monitor warning light comes on or not.
      ii. Vitamin Experimental Design: Eight children are given a placebo and eight children are given the new vitamin. Urine from the children is collected and the pH in the urine is monitored and compared between the two groups. Both high and low pH are indications of kidney dysfunction.
6. Return again to the coffee-tasting experiment. Does the outcome described prove that the model cannot tell the difference between the two types of cups of coffee?