**GER1000 2020 Sem 1**

**Quiz 7 and solutions**

1) A coin manufacturer claims that he has produced a biased coin with P(H) = 0.4 and P(T) = 0.6 where P(H) denotes the probability of the coin landing on heads and P(T) denotes the probability of the coin landing on tails. Brad takes this coin and tosses it 10 times. He obtains heads on the first 8 tosses and tails on the last two tosses. He decides to do a hypothesis test to see if there is enough evidence to reject the manufacturer’s claim. Which one of the following statements should he adopt as his null hypothesis?

1. P(H) = 0.4
2. P(H) = 0.5
3. P(H) = 0.8
4. P(H) = 0.2
5. P(H) = 0.6

*Explanation: The null hypothesis should be the claim of the manufacturer, that P(H) = 0.4. It should not be based on the data, so (C) is wrong. The other options are irrelevant. Refer to Chapter 5 Unit: “P-values” slides 2-6.*

2) Once upon a time, in a land far away, there lived a statistician by the name of Ronald Fisher. The town in which he lived, had a certain lady by the name of Muriel, who claimed that she had a special ability. She claimed that by merely tasting a cup of tea, she could identify whether milk had been added before the tea or tea had been added before the milk in the brewing process. Now Fisher was highly suspicious about whether this lady was in fact making a genuine claim. Fisher decided to prepare 8 cups of tea. For 4 of the cups, he added milk before adding the tea and for the remaining 4 cups he added tea before adding the milk. In all other aspects, the preparation of the 8 cups were identical. He then presented the 8 cups to Muriel for drinking and she identified every cup correctly.

From here, we assume the following

* If Muriel had no special ability, she guesses randomly and her chance of guessing each cup correctly is 0.5.
* Muriel did not see how any of the 8 cups of tea was prepared.

What can Fisher conclude at 5% level of significance?

1. There is enough evidence to reject the null hypothesis and conclude that Muriel has **no** special ability.
2. There is enough evidence to reject the alternative hypothesis and conclude that Muriel has a special ability.
3. There is enough evidence to reject the null hypothesis and conclude that Muriel has a special ability.
4. There is enough evidence to reject the alternative hypothesis and conclude that Muriel has **no** special ability.

*Explanation: Refer to Chapter 5 Unit: “P-values” slides 3 and 4. So the chance of her guessing all 8 cups correctly, assuming the null is true, is (1/2)8 = 0.00391, which is less than the specified level of significance of 5%. This gives us enough evidence to reject the null hypothesis and conclude that Muriel does in fact have a special ability. Look up Ronald Fisher’s lady tasting tea experiment on Wikipedia if you’re curious to find out more.*

3) The swab test is a rather uncomfortable process that one has to go through to determine whether a person has been infected with Covid-19. A researcher decides to develop a new test to detect Covid-19 in humans. This test has a sensitivity of 0.99. He administers the test in a town of 100 000 people of whom 1% are known to have Covid-19. The partial results are summarized in the contingency table below.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Positive | Negative | Row total |
| Covid-19 | 990 | 10 | 1 000 |
| No Covid 19 |  |  | 99 000 |
| Column total |  |  | 100 000 |

What can be said about the specificity, assuming the researcher obtained that r(Covid-19 | Positive) = 99% for his test?

1. The specificity is equal to 99%.
2. The specificity is more than 99%.
3. The specificity is less than 99%.

*Explanation: In order for r(Covid-19 | positive) = 99%, the number of false positives should be 10. (the total number of positives is 1000).* *Then the number of true negatives is 98 990. Hence the specificity has to be higher than 99% (in fact very close to 100%!) Use the following contingency table to check the value of r(negative | no covid 19).*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Positive | Negative | Row total |
| Covid-19 | 990 | 10 | 1000 |
| No Covid-19 | 10 | 98 990 | 99 000 |
| Column total | 1000 | 99 900 | 100 000 |

*Refer to Uncertainty Chapter on “Testing rare events” Slide* *9 and 10 for information on false positives and false negatives, as well as how to generate the contingency table.*

4) It is known that numerical variables *X* and *Y* are positively correlated with *r* = 0.12. Given that *Z* = 3*X*, which of the following statements is true?

(I) The correlation between *Z* and *Y* is 0.36.

(II) The correlation between *Z* and *Y* is 0.04.

1. (I) only
2. (II) only
3. Neither (I) nor (II)

*Explanation: The r-value is not affected by a change of scale in general. In particular, it is not affected when any of the variables in the regression equation is multiplied by a positive constant, therefore neither statement is true. Refer to Chapter 2 slide 52.*

*--- Refer to the following scenario for the next two questions: ---*

Benny is a messy student who keeps all his colored socks in a box. The box contains a total of 4 blue and 2 yellow socks. On the first day, while running late for class, he randomly selects (without replacement) two socks out of the box to wear. Assume the socks are indistinguishable from one another apart from color.

5) What is the probability that he will end up wearing a matching pair of socks on the first day to class?

1. 1/3
2. 2/5
3. 7/18
4. 7/15

*Explanation: To get a matching pair of socks, he needs to get either 2 blue socks, or 2 yellow socks. Using the addition rule, P(match) = P(2 blue socks) + P(2 yellow socks) = (4/6)(3/5) + (2/6)(1/5) = 7/15. Refer to Chapter 5, unit 5, slide 4*. *Alternatively, using the complement rule, P(no match) = (4/6)(2/5) + (2/6)(4/5) = 8/15. P(match) = 1 – P(no match) = 7/15. Refer to Chapter 5, Unit 5, slide 4, and Chapter 5, Unit 2, slides 5-7.*

6) The next day, Benny selects his socks again, but this time you are told that the first sock he has picked from the box is blue. What can you say about the probability he will end up wearing a matching pair of socks to class, in comparison to the first day?

1. The probability of him wearing matching socks is unchanged
2. The probability of him wearing matching socks is higher than that of the first day
3. The probability of him wearing matching socks is lower than that of the first day

*Explanation: P(2blue|first blue) = (4/6)(3/5)/(4/6) = (3/5). Notice that this answer is simply the probability of him choosing one blue sock, out of the remaining 5 socks. This is because you have already fixed the event of one sock being blue, so the conditional probability reduces down to 3/5 when picking the other sock.*

*Interestingly, this answer is greater than any of the 4 options in the other question on Benny’s matching socks. Refer to Chapter 5, Unit 5, slides 1-10.*

7) A player can choose to participate in the following game or not. It costs $100 per entry to play the game each time. After paying $100, a player can choose between options A, B or C, with a 25%, 50%, and 30% chance of winning respectively. The reward for each option is shown in the table below:

|  |  |  |  |
| --- | --- | --- | --- |
| Option | A | B | C |
| Payoff ($) | 125 | 100 | 250 |

Which is the best option if a player wishes to end up with the most money based on expected value? Assume that this player sticks to the same option across the entire series of games.

1. Option A
2. Option B
3. Option C
4. The player chooses not to participate in the game

*Explanation: Calculate the expected profit for each option after entry: Option A = 0.25\*$125 = $31.25. Option B = 0.5\*$100 = $50. Option C = 0.3\*$250 = $75. All options give a lower expected return than the entry fee of $100, so it’s better to just keep the $100. It does not matter how many games you play; in fact, the more you play, the more you end up losing on average regardless of which option picked. Refer to Chapter 5, Unit 3, slides 1-9.*

8) Cheryl suspects that there is a higher percentage of girls in the Faculty of Arts as compared to other faculties. To prove her point, she collects data on the total number of males and females from the Faculty of Arts, as shown below. Based on the findings, what can Cheryl conclude?

|  |  |  |
| --- | --- | --- |
|  | Males | Females |
| Number | 400 | 1625 |

1. The rate(males|Arts) is less than 22%.
2. The rate(females|Arts) is greater than 75%.
3. Females are positively associated with being in the Faculty of Arts.
4. I, II, III
5. I, II only
6. I only
7. II only

*Explanation: I is true, because Rate(males|Arts) = 400/(1625+400) = 19.75%. II is true because Rate(females|Arts) = 1625/(1625+400) = 80.25%. III is inconclusive based on the numbers alone, because checking for association requires a comparison with the numbers from other faculties. You would need to compare the rate of females in Arts to that of other faculties to reach such a conclusion. For example, there could well be 90% females in other faculties, which will then render this statement false. Refer to Chapter 1, Unit 6, slide 4.*

9) A study is interested in the association between heart disease and cancer. A random sample of 500 people with no heart disease and 500 people with heart disease was taken and their cancer status was recorded down. Which of the following statements must be correct?

1. This study gives a good estimate of the population risk of heart disease.
2. The sample rate of cancer can be calculated from this study.
3. I only
4. II only
5. Both I and II
6. Neither I nor II

*Explanation: I is false because this is a case control study with respect to heart disease, and the sample rate of heart disease has already been fixed at 50% by design. There are no grounds to estimate the population risk here. II is true regardless of how the subjects were* *sampled, because in this case the statement is interested in only the sample. Refer to Chapter 4, Unit 2, slide 7.*