You have completed

Practice Quiz 4 (Chapter 4)

You scored **11/26**

Have not received the confirmation Email? Click here to re-send.

Display chosen by module coordinator

SCROLL DOWN TO VIEW DETAILED RESULTS

Chapter 4 Questions

1. A football match is played between Singapore and Italy, at the Singapore National Stadium. 90% of the spectators in the stadium are wearing red. Among the spectators in the stadium wearing red, 80% are rooting for Singapore. Among the spectators in the stadium not wearing red, 80% are rooting for Singapore.

Among all the spectators in the stadium, a spectator is randomly chosen with every spectator having the same chance of being selected.

Let A be the event that the chosen spectator is rooting for Singapore, and B be the event that the chosen spectator is wearing red. Which of the following statements, regarding the 2 events A and B is true?

(1 mark)

You scored 1 / 1 mark

- The 2 events are mutually exclusive and independent.
- The 2 events are neither independent nor mutually exclusive.
- The 2 events are independent but not mutually exclusive.

The 2 events are mutually exclusive but not independent.

General Comments

By the basic rule on rates, observe that

P(Root for Singapore | Wear red) = P(Root for Singapore | Does not wear red) = P(Root for Singapore) = 0.8.

Since

P(Root for Singapore) = P(Root for Singapore | Wear red) = 0.8, rooting for Singapore and wearing red are independent events.

Here

P(Root for Singapore) + P(Wear red) = 1.7,

which is not equal to P(Root for Singapore or Wear red), since this probability is less than or equal to 1. Hence rooting for Singapore and wearing red are not mutually exclusive events.

2. A nutritionist wants to estimate the average mass of a particular tortilla chip brand, Naritos. He decides to collect a sample of 100 packets of Naritos chips, using the method described as follows: He went down to the nearest factory that produces Naritos chips and measured the average mass of the first 100 packets of Naritos chips produced on that day. You can assume the mass of the packaging is negligible.

Using the formula

$$\overline{x} \pm t^* \times \frac{s}{\sqrt{n}},$$

he obtained a 95% confidence interval for the population mean as [676.3, 723.7]. Which of the following statements must be true?

(I) If we collect 100 samples of 100 packets of Naritos chips using the same sampling method, about 95 of the samples will have confidence intervals containing the population average mass of Naritos chips.

(II) We can conclude that there is a 95% chance that the population average mass of Naritos chips lies within [676.3, 723.7].

(1 mark)

You scored 0 / 1 mark

Only (II).

Only (I).

Neither (I) nor (II).

Both (I) and (II).

General Comments

Recall that

sample statistic = population parameter + bias + random error.

For a confidence interval to be valid, it is with the assumption that bias is approximately zero, thus

sample statistic = population parameter + random error.

However, in this scenario, we do not use a probability sampling method when acquiring our sample, so selection bias will interfere with the sample's estimate of the population mean mass of Naritos chips.

It is possible that many samples out of the 100 samples collected have a sample mean that understates or overstates the true population mean by a large amount.

For instance, it is possible that the first 100 packets of chips produced each day are a lot heavier than the average packet of chips.

Thus, we do not expect to see 95 out of 100 of the confidence intervals generated containing the true population mean. Statement (I) is false.

Furthermore, we do not use the word 'chance' that the population mean lies within a given range, thus statement (II) is also false.

3. A study was conducted on 1000 students in a secondary school on 11th January 2021. On that day, of these 1000 students, 80% of the students took the MRT (mass rapid transit) to school. Among those students that took the MRT to school, 10% were late for school. Among those students who did not take the MRT to school, 20% were late for school.

Among those students who participated in the study and were late for school, a student was randomly chosen, with every student having the same chance of being selected. What is the probability (correct to 1 decimal place) that the chosen student did not take the MRT?

(1 mark)

You scored 1 / 1 mark

- 10.0%
- 33.3%
- 30.0%
- 20.0%

General Comments

The total number of students is 1000 and P(take MRT) = 0.8 and P(did not take MRT) = 0.2, hence 800 students took MRT and 200 students did not.

Among 800 students who took MRT to school, 10% were late. That is, P(late|took MRT) =0.1, so there were 80 students who took the MRT and were late.

Among 200 students who did not take MRT to school, 20% were late. So P(late|did not take MRT)=0.2, so there were 40 students who did not take the MRT and were late.

Put the calculated figures in a 2x2 table as shown below.

	Late	Not late	Row total
Took MRT	80	720	800
Did not take MRT	40	160	200
Column total	120	880	1000

From the table, we can calculate

P(did not take MRT|late) = 40/120=33.3% (1 decimal place).

4. A poll conducted before a local election gives a 95% confidence interval for the percentage of voters who support candidate X as (54%, 60%). Based on the same poll result, which of the following can potentially be a 99% confidence interval for the percentage of voters who support candidate X?

(1 mark)

You scored 1 / 1 mark

(56%, 58%)

(52%, 62%)

(54%, 60%)

General Comments

A 99% confidence interval should be wider than the 95% confidence interval for the same sample.

5. Suppose we want to test if a coin is biased towards heads. We decide to toss the coin 10 times and record the number of heads. We shall assume the independence of coin tosses, so that the 10 tosses constitute a probability experiment.

Let X denote the number of heads occurring in 10 independent tosses of the coin. We will carry out a hypothesis test with X as the test statistic.

Let H be the event that the coin lands on head, in a single toss. We set our hypothesis to be

- $H_0 : P(H) = 0.5,$
- $H_1 : P(H) > 0.5.$

Suppose in our execution of the 10 tosses, we observe 4 heads. This means $X = 4$ is the test result we observe.

Recall the definition of p-value to be 'the probability of obtaining a test result at least as extreme as the one observed, assuming the null hypothesis is true.'

What is the range of test results ''at least as extreme as the one observed", in this scenario?

(1 mark)

You scored 0 / 1 mark

$5 \leq X \leq 10.$

$0 < X < 5.$

$4 \leq X \leq 10.$

$0 \leq X \leq 4.$

General Comments

Note that in the context of p-value computation, "at least as extreme" is interpreted as "at least as favourable to the alternative hypothesis". In this scenario, the greater the value X assumes, the more favourable the case is to the alternative hypothesis. Hence, the answer should be all values of X greater than or equal to the observed value.

6. A player rolls a fair six-sided die twice. You can assume the rolls are independent. We define the following events:

A: The first roll shows numbers 1 or 2.

B: The second roll shows numbers 5 or 6.

C: The sum of the two rolls is less than or equal to 7.

Consider the following statements:

(I) P(B | C) > P(B).

(II) P(A and C) = P(A) x P(C).

Which of the statements above must be true?

(1 mark)

You scored 0 / 1 mark

Neither (I) nor (II).

Only (II).

Only (I).

Both (I) and (II).

General Comments

Statement (I) is false since

$P(B|C) = \frac{P(B \text{ and } C)}{P(C)} = \frac{\frac{3}{36}}{\frac{7}{12}} = \frac{1}{7}$,

while P(B)=$\frac{1}{3}$. Statement (II) is false because $P(A \text{ and } C) = \frac{11}{36}$ and P(A) = $\frac{1}{3}$ and P(C) = $\frac{7}{12}$.

7. There are 5 red balls and 5 blue balls in a bag. Mary picked two balls from the bag, one after another such that the first ball picked was not replaced in the bag before the second ball was picked. She will win a prize if the two balls she picked are of the same colour. The first ball she picked was red. What is the probability that Mary wins a prize?

(1 mark)

You scored 1 / 1 mark

1/9

1/2

4/9

2/9

General Comments

The answer is simply the condititional probability

P(2nd red | 1st red) $= \frac{(5/10)*(4/9)}{(5/10)} = \frac{4}{9}$.

Notice that this answer is simply the probability of choosing 1 red ball, out of the remaining 9 balls in the bag. This is because we have already fixed the event of the first ball being red, so the conditional probability reduces to 4/9 when picking the second ball conditional probability reduces to 4/9 when picking the second ball.

8. A standard deck of 52 playing cards comprises 4 suits (Clubs, Diamonds, Hearts, Spades), with each suit comprising of 13 cards of distinct ranks (A, 2, 3, 4, 5, 6, 7, 8, 9, 10, J, Q, K).

What is the probability that a card randomly selected from this deck is a '3', given that it is a 'Hearts'?

(1 mark)

You scored 1 / 1 mark

None of the other options.

$\frac{1}{52}$

$\frac{1}{4}$

$\frac{1}{13}$

General Comments

Let event A be 'drawing a 3' and event B be 'drawing a 'Hearts''. Note that each card has equal chance of being selected. As a result,

$P(A \text{ and } B) = \frac{1}{52}$ and

$P(B) = \frac{1}{4}$, so

$P(A|B) = P(A \text{ and } B)/P(B) = \frac{1}{52}/\frac{1}{4} = \frac{1}{13}$.

9. A machine produces tempered glass for mobile phones. The tempered glass produced are either grade A (good), grade B (minor defects) or grade C (unacceptable). The glass produced are put through an inspection machine that is able to detect any grade C glass and discard it. We may assume that the machine makes no mistakes in the detection process, meaning that given a piece of grade C glass, the probability of it being discarded is 1. At the same time, given a piece of grade A or B glass, the probability of it being discarded is 0.

Suppose the machine produces grade A glass 90% of the time, grade B glass 2% of the time and grade C glass 8% of the time. Which of the following is the probability (computed to 3 significant figures), denoted by p, that a piece of glass that is not discarded by the inspection machine is grade A?

(1 mark)

You scored 1 / 1 mark

0.900

0.978

0.920

There is not enough information for p to be computed.

General Comments

We want to compute the probability P(grade A | not grade C). By conditional probability,

P(grade A | not grade C) = P(grade A and not grade C) / P(not grade C) = P(grade A) / P(grade A or B) = 0.9/0.92 = 0.978.

Note that a piece of glass that is (grade A and not grade C) is certainly (grade A). Similarly a piece of glass that is (not grade C) is certainly (grade A or grade B).

10. Suppose we want to test if a coin is biased towards tails. We decide to toss the coin 10 times and record the number of heads. We shall assume the independence of coin tosses, so that the 10 tosses constitute a probability experiment.

Let X denote the number of heads occurring in 10 independent tosses of the coin. We will carry out a hypothesis test with X as the test statistic.

Let H be the event that the coin lands on head, in a single toss. We set our hypothesis to be

- $H_0 : P(H) = 0.5,$
- $H_1 : P(H) < 0.5.$

Suppose in our execution of the 10 tosses, we observe 7 heads. This means $X = 7$ is the test result we observe.

Recall the definition of p-value to be 'the probability of obtaining a test result at least as extreme as the one observed, assuming the null hypothesis is true.' Which of the following is/are possible test result(s) "at least as extreme as the one observed", in this scenario? Select all that apply.

(1 mark)

You scored 1 / 1 mark

- [ ] $X = 2.$
- [ ] $X = 6.$
- [ ] $X = 4.$
- [ ] $X = 10.$
- [ ] $X = 8.$

General Comments

Note that in the context of p-value computation, “at least as extreme” is interpreted as “at least as favourable to the alternative hypothesis”. In this scenario, the smaller the value X assumes, the more favourable the case is to the alternative hypothesis. Hence, the answer should be all possible values of X lesser than or equal to the observed value.

11. Which of the following statements is/are true of normal distributions? Select all that apply.

(1 mark)

You scored 0 / 1 mark

- [ ] A normal distribution can be right-skewed.

A normal distribution can be left-skewed.

$N(1,4)$ has a standard deviation of 2.

The density curve of a normal distribution is symmetrical about its mode.

General Comments

Since the density curve of any normal distribution is symmetrical about its mean, and its mean is equal to its mode, it is symmetrical about its mode. A symmetrical distribution cannot be left-skewed or right-skewed. Lastly, since N(1,4) has variance 4, it has standard deviation equal to $\sqrt{4} = 2$.

12. A game is played using a fair six-sided die, a pawn and a simple board as shown below. (A pawn is a chess piece.)

S	1	2	3	4	5	E

Initially, the pawn is placed on square S. The game is played by throwing the die and moving the pawn back and forth in the following manner:

S 1 2 3 4 5 E 5 4 3 2 1 2 3 4 5 E 5 4 ....

Thus, for example if the first and second throws of the die give a "5" and "4" respectively, the final position of the pawn will be on square "3", because the first throw would send the pawn to square "5", and the second throw would then send the pawn from square "5" to square "3".

The game will stop only when the pawn stops at square "E'" after a die roll, passing by "E" **does not** end the game.

Let X denote the number of throws of the die required to move the pawn such that it stops at square "E". Which of the following statements is/are true?

(I) P(X=2) = $\frac{5}{36}$.

(II) The events X=1 and X=2 are mutually exclusive.

(1 mark)

You scored 1 / 1 mark

Both (I) and (II).

Only (I).

Neither (I) nor (II).

Only (II).

General Comments

The event X=2 can only happen when we do not roll a 6 on the first roll, but our first two rolls add up to 6. Note that it is impossible to hit "E" a second time in two die rolls because it is 16 moves away from the starting point. To win in 2 rolls, we need one of the following pairs to happen:

(1,5), (2,4), (3,3), (4,2), (5,1).

Here the x in (x, y) is the number obtained on the first roll and y is the number obtained on the second roll. This corresponds to 5 events out of 36, so the probability is $\frac{5}{36}$ and thus statement (I) is correct. Statement (II) is also correct because it is not possible to end the game when the die is rolled once AND the die is rolled twice.

13. A fair coin is tossed three times. We say that A is the event of getting at least 2 heads. Likewise, B denotes the event of getting no heads and C is the event of getting heads on the second toss.

Consider the following cases:

(I) Events A and B;

(II) Events A and C;

(III) Events B and C.

Which of the cases above contains events that are mutually exclusive?

(1 mark)

You scored 0 / 1 mark

Only (I).

Only (III).

Only (II) and (III).

Only (II).

Only (I) and (III).

General Comments

Event A can be denoted by {THH,HHT, HTH, HHH}. Similarly, event B is denoted by {TTT} and event C by {THT, HHH, HHT, THH}. Events B and C are mutually exclusive because there is no intersection, i.e., both B and C cannot happen at the same time. The same can be said for events A and B.

14. 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.

Out of 10 independent tosses, Brad observes 8 heads and 2 tails. Based on these data, he decides to do a hypothesis test to see if there is enough evidence to reject the manufacturer's claim. Which of the following statements should he adopt as his null hypothesis?

(1 mark)

You scored 0 / 1 mark

P(H)=0.6.

P(H)=0.5.

P(H)=0.8.

P(H)=0.4.

General Comments

Whenever we want to gather evidence to reject a claim, the null hypothesis should maintain that the claim does not differ from the truth. In this case, it should state P(H)=0.4. It should not be based on the data, so P(H)=0.8 is wrong. The other options are irrelevant.

15. The sample space S of a probability experiment comprises the 30 days of September 2021. Let A denote the first 10 days of September 2021. Let B denote 1st September 2021.

Which of the following statements is/are correct? Select all that apply.

(1 mark)

You scored 0 / 1 mark

- [ ] Both A and B are outcomes of S.
- [ ] A is an outcome of S and B is an event of S.
- [ ] Both A and B are events of S.
- [ ] A is an event of S and B is an outcome of S.

General Comments

A is a sub-collection of S so it is an event of S, and B is an element of S so it is an outcome of S. Since we also regard outcomes as events, both A and B are events of S.

16. Let the random variable X denote the number of heads obtained in 3 independent tosses of a fair coin. Which of the following tables correctly displays the probabilities over the possible values of X?

(1 mark)

You scored 0 / 1 mark

X	0	1	2	3
Probability	0.125	0.375	0.375	0.125

X	0	1	2	3
Probability	0.25	0.25	0.25	0.25

X	0	1	2	3
Probability	0.333	0.167	0.167	0.333

X	0	1	2	3
Probability	0.125	0.5	0.25	0.125

General Comments

Listing all possible outcomes, the sample space is {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}. Since the coin is fair, and the tosses are independent, we have uniform probability within the sample space. This implies

$P(X = n) =$ number of outcomes containing n heads $\times \frac{1}{8}$,

which we can use to fill up the table.

17. A group of students want to find out if there is any association between staying in a hall and being late for class in NUS in a particular month. If students are late for at least 5 classes, they are considered "late for class" for that month. After collecting a simple random sample of 1000 students which has no bias, they found that 200 out of 350 students who stay in hall are late for class, while 390 out of 650 students who do not stay in a hall are late for class.

A Chi-squared test was done to test for association between staying in a hall and being late for class at 5% level of significance. The p-value derived from the Chi-squared test is 0.3809.

Which of the following statements is/are true? Select all that apply.

(1 mark)

You scored 0 / 1 mark

- [ ] There is negative association between staying in a hall and being late at the sample level.
- [ ] Since the p-value is more than 0.05, we cannot conclude that there is an association between staying in a hall and being late at the population level.
- [ ] There is positive association between staying in a hall and being late at the sample level.
- [ ] Since the p-value is more than 0.05, we can conclude that there is an association between staying in a hall and being late at the population level.

General Comments

To check for association, between staying in hall and being late at the sample level, we compare

rate(Late for class | Staying in hall) = $\frac{200}{350} < \frac{390}{650}$ = rate(Late for class | Not staying in hall).

Hence, we see that staying in hall is negatively associated with being late at the sample level. As we are doing a Chi-squared test to see if there is an association at the population level, recall that the null hypothesis states that there is no association at the population level, while the alternative hypothesis states that there is an association at the population level. Since the p-value is greater than the level of significance, we do not reject the null hypothesis. Hence, we cannot conclude that there is an association between staying in hall and being late at the population level.

18. Based on a random sample of 200 staff members in NUS, a 95% confidence interval for the proportion of all NUS staff who went on a vacation in 2018 for at least 5 days is given as (0.33, 0.59). Which of the following statements must be true?

(I) If another sample of size 500 is drawn using the same sampling method, for the same confidence level, the confidence interval will be wider than (0.33, 0.59).

(II) A maximum of 59% of all NUS staff went on vacation in 2018 for at least 5 days.

(1 mark)

You scored 1 / 1 mark

Only (II)

Only (II).

Both (I) and (II).

Neither (I) nor (II).

Only (I).

General Comments

Statement (I) is false since a bigger sample size will likely result in a narrower confidence interval for the same confidence level. Statement (II) is also false as a confidence interval does not give us any insight on the maximum possible value of the population proportion.

19. Tom selects a child at random from a population of children. Let

- A be the event a child of age < 3 is selected;
- B be the event a child of age < 5 is selected.

It is known that P(A)>0. Which of the following must be true?

(I) P(A or B) < P(A) + P(B).

(II) P(A) $\leq$ P(B),

(1 mark)

You scored 0 / 1 mark

Only (I).

Only (II).

Both (I) and (II).

Neither (I) nor (II).

General Comments

Every child of age < 3 is also of age < 5. Thus, event B must occur if event A occurs. This means that P(A) $\leq$ P(B), that is statement (II) is true.

Furthermore, since whenever event A occurs, event B must occur, we have P(A or B) = P(B). So comparing P(A or B)=P(B) and P(A) + P(B), we conclude that P(A or B) < P(A) + P(B) since P(A) is strictly bigger than 0. Thus, statement (I) is true.

20. A 95% confidence interval, constructed using a simple random sample, for the population mean number of children per household in Country Z is (1.21, 4.67). Which of the following statements is/are true? Select all that apply.

(1 mark)

You scored 1 / 1 mark

- [ ] The probability that the population mean number of children per household in Country Z is between 1.21 and 4.67 is 0.95.
- [ ] We are 95% confident that the population mean number of children per household in Country Z is between 1.21 and 4.67.
- [ ] 95% of all samples of the same size and sampling procedure should have sample mean number of children per household between 1.21 and 4.67.
- [ ] Taking 100 different samples of the same size and using the same sampling procedure and computing confidence intervals for each sample in the same way, approximately 95 of the intervals will contain the true population mean..
- [ ] 95% of all households in Country Z have between 1.21 and 4.67 children.

General Comments

It is wrong to say that the probability of the population mean is between 1.21 and 4.67 is 0.95. The population mean is either between 1.21 and 4.67 or it is not. There is no probability/chance involved. Confidence intervals are constructed based on sample means and when 100 intervals are constructed from the 100 samples, all the 100 intervals should capture their own sample means. Lastly, confidence intervals give us inference about the population mean and not on each individual household.

21. A researcher is interested to know if smoking and heart disease are associated with each other in the population of Singapore. The researcher takes a census of the population in Singapore with a 100% response rate. The researcher conducts a chi-squared test on the census data at 5% significance level and obtains a p-value of 0.001.

Which of the following is a valid conclusion?

(1 mark)

You scored 0 / 1 mark

- Since p-value is less than 0.05, the null hypothesis is **rejected** at the 5% significance level, and the researcher concludes that smoking and heart disease **are not associated** with each other in the population.
- Since p-value is less than 0.05, the null hypothesis is **rejected** at the 5% significance level, and the researcher concludes that smoking and heart disease **are associated** with each other in the population.
- None of the other options is a valid conclusion.
- Since p-value is less than 0.05, the null hypothesis is **not rejected** at the 5% significance level, and the researcher concludes that smoking and heart disease **are not associated** with each other in the population.
- Since p-value is less than 0.05, the null hypothesis is **not rejected** at the 5% significance level, and the researcher concludes that smoking and heart disease **are associated** with each other in the population.

General Comments

The researcher took a census of the population, not a sample of the population. There is no estimation involved in this scenario; only a measurement of the population, so hypothesis testing is irrelevant. Hypothesis tests on testing a claim about the population should be conducted on probability-based samples, not a census.

22. There are 5 bags that are the same, except that 2 are coloured red and 3 are coloured blue. Each of the bags contains 4 balls that are identical, except that 3 are coloured yellow and 1 is coloured green. Let A be the event that a randomly selected bag is red, and B be the event that a ball randomly selected from the chosen bag is yellow. You are given that P(A and B) = 0.3.

What can we say about the events A and B?

(I) The two events are mutually exclusive.

(II) The two events are independent.

(1 mark)

You scored 1 / 1 mark

- Only (II).
- Neither (I) nor (II).
- Both (I) and (II).
- Only (I).

General Comments

Note that P(A) = 2/5 and

P(A | B) = P(A and B) / P(B) = 0.3 / 0.75 = 2/5 = P(A).

Thus, the two events are independent. The two events are not mutually exclusive since they can occur at the same time (picking a yellow ball from a red bag).

23. The swab test detects whether a person has been infected with COVID-19. A researcher developed a new test to detect COVID-19 in humans and the test has a specificity of 0.90. He administers the test in a town of 100,000 people, in which 1% are known to have COVID-19, as indicated in the contingency table below.

	Positive	Negative	Row Total
COVID-19			1,000

No COVID-19			99,000
Column Total			100,000

What can be said about the sensitivity of the test, assuming that the researcher obtained

rate(COVID-19 | Negative)=1/298

for his test?

(1 mark)

You scored 0 / 1 mark

- The sensitivity is more than 80%.
- The sensitivity is less than 80%.
- The sensitivity is equal to 80%.

General Comments

Note that

Specificity = rate(Negative | No COVID-19) = 0.90,

so the number of true negatives is 89,100. Assuming that rate(COVID-19 | Negative)=1/298, the total number of negatives is

$$89100 \div \frac{(298-1)}{298} = 89400,$$

and the number of false negatives is 300. Then the number of true positives is 700. Hence,

sensitivity = rate(Positive | COVID-19) = 0.7.

We can use the following contingency table to check the values.

	Positive	Negative	Row Total
COVID-19	700	300	1000
No COVID-19	9900	89100	99000
Column Total	10600	89400	100 000

24. Two simple random samples, sample A and sample B, of NUS students were taken. Sample A consisted of 100 NUS students, while sample B consisted of 250 NUS students. For both samples A and B, the sample proportion of NUS students that visit campus at least once a week from January 2021 to March 2021 is 0.8.

For sample A, a 95% confidence interval for the proportion of NUS students that visit campus at least once a week from January 2021 to March 2021 was constructed to be [0.75, 0.85].

Which of the following statements must be true? Select all that apply.

(1 mark)

You scored 0 / 1 mark

- [ ] A maximum of 85% of all NUS students visit campus at least once a week from January 2021 to March 2021.
- [ ] If a 90% confidence interval for the proportion of NUS students that visit campus at least once a week from January 2021 to March 2021 is constructed using sample A, the width of the confidence interval will be narrower than [0.75, 0.85].
- [ ] If a 95% confidence interval for the proportion of NUS students that visit campus at least once a week from January 2021 to March 2021 is constructed using sample B, the width of the confidence interval will be narrower than [0.75, 0.85].
- [ ] A minimum of 75% of all NUS students visit campus at least once a week from January 2021 to March 2021.
- [ ] The population proportion of NUS students that visit campus at least once a week from January 2021 to March 2021 must be either in the 95% confidence interval of sample A or the 95% confidence interval of sample B.

General Comments

The 90% confidence interval computed from sample A will be narrower than [0.75, 0.85], because both the sample proportion $p^* = 0.8$ and the sample size $n = 100$ are the same as the ones used to arrive at [0.75, 0.85], whereas the z^* score of 1.645 is smaller than the z^* score for 95% confidence intervals.

If sample A and sample B have the same sample proportion $p^* = 0.8$, and both samples' confidence intervals are constructed at 95% confidence level, using the same z^* score of 1.96, then the difference in confidence intervals depends only on the sample size n. Since sample B is larger than sample A, sample B's confidence interval will be narrower than sample A's confidence interval.

On the other hand, the confidence interval does not inform us of the range of the proportion of NUS students that visit campus at least once a week, nor can we be sure that the population proportion lies within any particular confidence interval generated.

25. Which of the following statements about the p-value is/are true? Select all that apply.

(1 mark)

You scored 0 / 1 mark

- [ ] The p-value gives the probability that the null hypothesis is true.
- [ ] At a 5% significance level, a p-value smaller than 0.05 provides sufficient evidence that the null hypothesis should be rejected.
- [ ] The p-value is the probability of obtaining results at least as favorable to the alternative hypothesis as the collected data, computed based on the assumption that the null hypothesis is true.
- [ ] At a 5% significance level, a p-value larger than 0.05 provides sufficient evidence that the null hypothesis should be rejected.

General Comments

In a hypothesis test, the null hypothesis is assumed to be true, we neither prove nor provide evidence that the null hypothesis is true. A significance level is a threshold we set, so that any p-value below it indicates there is sufficient evidence to reject the null hypothesis.

26. A researcher takes a simple random sample from Country X's population to estimate its unemployment rate. The results showed that a 95% confidence interval for the population unemployment rate was between 0.18 and 0.22.

Which of the following statements correctly interprets the results?

(1 mark)

You scored 0 / 1 mark

- If many samples of the same size were collected by the same procedure, and their respective confidence intervals calculated in the same way, about 95% of these samples will have the sample unemployment rate lie within the samples' respective confidence intervals.
- If many samples of the same size were collected by the same procedure, and their respective confidence intervals calculated in the same way, about 95% of these samples will have the sample unemployment rate lie between 0.18 and 0.22.
- If many samples of the same size were collected by the same procedure, and their respective confidence intervals calculated in the same way, about 95% of these samples will have the population unemployment rate lie between 0.18 and 0.22.
- If many samples of the same size were collected by the same procedure, and their respective confidence intervals calculated in the same way, about 95% of these samples will have the population unemployment rate lie within the samples' respective confidence intervals.

General Comments

By the interpretation of confidence intervals via repeated sampling, we conclude that about 95% of the samples will contain the population parameter (in this case population unemployment rate) within their respective confidence intervals.