UNIVERSITI BRUNEI DARUSALAM

CODE AND MODULE'S NAME:

SM-4290 PROJECT RESEARCH

RESEARCH TITLE:

Investigation on UBD Students' Confidence in

Hypothesis Test Concepts

STUDENT/ RESEARCHER'S NAME:
Nur Aida Afiqah binti Zaini (17B3029)
Mathematics
Faculty of Science

SUPERVISOR:

Dr Haziq Jamil
Assistant Professor in Statistics
Mathematical Sciences, Faculty of Science
Universiti Brunei Darussalam

SUBMISSION DATE:

25th April 2021

Table of Contents

1. Abstract

Background: Hypothesis testing is an essential tool in statistical inference, mainly to produce well-written research. However, past studies have shown that students hold misconceptions about these statistical inferences. **Methods:** This study investigated Universiti Brunei Darussalam (UBD)'s undergraduate students' confidence of the hypothesis testing concept. Besides, it also addressed the most common error of statistical inference concerning hypothesis test, p-values and the level of significance. This study was done using a two-phase, mixed method approach with an online questionnaire and a follow-up interview. Chi square test used to analyze the multiple-choice questions of the survey. Follow-up interview was used to provide insight of students' understanding of this concept. **Results:** Main finding shows that there is a weak performance of the survey. Follow-up interviews show that students do not have a deep understanding of hypothesis testing and held some common misconceptions. **Conclusion:** The recommendations are that future study should focus solely on statistical students regarding hypothesis testing with more questions. With the outcome, it is hoped that this research could help diagnose any common error or misconceptions of statistical knowledge of UBD undergraduate students.

Keywords: Hypothesis Testing-Misconceptions

2. Introduction

The use of statistics is essential to effectively conduct research or to make predictions in a variety of fields such as education and business. There are two parts of statistics which are descriptive statistics such as mean and median; and statistical inference in relation with hypothesis testing.

Students are expected to encounter introductory statistics as early as high school. This can be seen in Cambridge General Certificate of Education Ordinary Level (GCE O) examinations in high school and further in GCE Advanced Level (GCE A) examination prior to university level. For examples in GCE O Statistic, one of the syllabi aims is for learners to be able to "interpret both primary statistical information and the conclusions of statistical analysis" whereas in the GCE A Mathematics, one of the aims in the Statistics section is to be able to "understand the nature of a hypothesis test and the terms null hypothesis, alternative hypothesis, significance level, rejection region (or critical region), acceptance region and test statistic". Hence, introductory statistics are also taught in order to prepare individuals with using hypothesis testing when conducting research.

Numerous studies have shown that students have misconceptions on statistical inference for the past three decades. In today's world, statistics are made easier by means of advanced technology and machine learning. Certain software such as R and Statistical Package for the Social Sciences (SPSS) has been developed to assist students in understanding statistical processes. This emphasises the need for the ideas of statistical reasoning and literacy and conceptual understanding in statistics that is not performable by machines. However, hypothesis testing has been expressed as one of the most misunderstood topics (Vallecillos, 2000). Due to the common misconceptions, students are weaker in using hypothesis testing to interpret experimental data compared to numerical calculations (Nickerson, 2000). It is also reported that there are three major aspects of the statistical hypothesis that are found to be difficult for students, namely, to understand the concept of a hypothesis test, to interpret p-values and to interpret the significance level.

With regards to the previous statement, there have been extensive studies documenting students' most common misconceptions of hypothesis testing. This results in a drastic impact within the research community (Sotos et al, 2009).

In light of this matter, the primary goal of this study was to investigate Universiti Brunei Darussalam (UBD) undergraduate students' confidence of hypothesis testing. The study also aimed to answer the following sub-research questions:

1) Is there any correlation between confidence in student's knowledge of hypothesis testing and class performance (such as CGPA)?
2) What are the most common errors in the hypothesis test concept made by UBD undergraduate students?
3) Do the final year UBD undergraduate students understand hypothesis testing concepts more confidently compared to non-final year students?

3. Literature Review

A review of the literature suggested that students have trouble with understanding hypothesis testing concepts (Vallecillos, 2000; Sotos et al, 2009; Smith, 2008; Burns-Childers et al, 2017, R.Mevarech, 1983). It is also noted that some researchers were testing the knowledge of introductory statistic undergraduate students with various education background as presented by Smith (2008), Sotos et al (2009) and Burns-Childers (2017). Researches agree that even when they are able to perform all the steps of the hypothesis test, misconceptions still occur amongst students as they were unable to explain the reasoning behind these steps. This can be seen from Burns-Childers et al (2017) study that tested student's understanding in test statistics which stated that "many students used a procedural approach to hypothesis testing which includes students plugging in correct values in a formula" but failed to fully understand the real concept.

According to Batanero (2000), the common misinterpretations circulate around the significance level, interpreting p-values and the role of the null and alternative hypotheses. This leads to several researchers to investigate further into the three aspects of hypothesis testing. Sotos et al (2009) identified that first year university undergraduates who took introductory statistic are more confident in the hypothesis testing concept but showed significantly lower confidence in interpreting the significance level.

Sotos et al (2009) also said that "gathering as much as possible about students' misconceptions can be very helpful for the development of new assessment techniques that acknowledge those common errors". This is due to the fact that different statistics educators have different way of teaching styles and finding the ideal technique might help tackle misconceptions.

4. Research Methodology

4.1 Overall methods

This study was conducted using a two-phase, mixed method approach of data collection and analysis. The first method was a quantitative phase by disseminating a self-administered survey and a qualitative phase followed this by conducting follow-up interviews.

The study was conducted at Universiti Brunei Darussalam, situated at Jalan Tungku in Brunei Darussalam. There are a total of twelve academic faculties and institutes with only six undergraduate faculties in UBD. It is not known what fraction of the UBD undergraduate students has basic knowledge in introductory statistics. Therefore, the non-probability sampling method was utilized in this study. This method indicates that the odds of being chosen for a sample is unknown or cannot be calculated. Thus, study subjects were recruited from different backgrounds of statistics experience and were from different undergraduate faculties.

Prior to data collection, items were determined from previous studies of researchers in the discipline. The research items for the quantitative phase were sculptured into multiple-choice questions (MCQ) along with rating scales. The same survey would be used in the qualitative phase. In order to refine the survey items, a pilot survey was then carried out by two students.

The primary focus of this pilot study was to get an approximate completion time on top of ensuring the appropriate intensity of the survey. Initially, the survey consisted of 10 questions from a previous study by Sotos et al (2008) and Smith (2007). The pilot survey was completed in an average of 15 minutes and the follow-up interviews took over an hour. This pilot survey showed that the two students were only able to focus on the first seven questions and were then throwing answers without reasoning on the last three questions due to exhaustion. This also occurred in the follow-up interview as it seemed that students felt tired when explaining their reasoning after the 45 minutes mark. In addition, it was pointed out from the pilot interview that some items using Sotos and Smith study were given similar repetitive explanations. As a result, the finalized survey contained only five questions from Smith (2007) study that could also be utilized in the qualitative phase. To obtain a satisfactory response rate, it was ensured that the survey could be done under 10 minutes whereas the interview would only take a maximum of 45 minutes.

At the final section of the survey, respondents were inquired if they were willing to participate in a follow-up interview. Those who agreed to be interviewed were asked to submit an email address so they could be contacted solely for the interview.

4.2 Quantitative Phase

The survey was disseminated in February 2021 for a month, through an online link using Google Form. UBD emails and a social media platform, WhatsApp were utilized to share the survey link with UBD undergraduate students.

The questionnaire comprises a set of five multiple-choice questions to measure students' understanding of hypothesis testing. Each wrong answer describes the misconception that might be held by students. With the motive to investigate the distinct relationship of hypothesis testing apprehension and confidence, a ten-point Likert scale also accompanies each question. The Likert scale denotes how confident students were about each answer (0-0% confident correct answer, 10-100% confident correct answer). Each question was classified into four categories of hypothesis testing shown in Table 1.

Question	Category
1	Recognize Applicability
2	Recognize Applicability
3	Generate Hypothesis
4	Conclusion
5	Decision Rule

Table 1 shows the category of multiple-choice questions of survey.

Analysis from the survey was done using R software and Microsoft Excel. Descriptive statistics such as mean of total scores and confidence level were calculated. On top of the main purpose, the results of the survey were also used to address sub-question 1 and 3. Sub-question 1, the correlation between confidence in students' knowledge of hypothesis testing and their CGPA, was calculated using chi square test whereas sub-question 3, final year students versus non final year, was calculated from the descriptive summary statistics of students' confidence of correct answers. In addition, chi square test was also used to find any other correlations between two covariates. Prior to that, 'confidence level' were divided into High (8-10), Medium(4-7) and Low scores (0-3),

'cGPA' were divided into High (3.4-4), Medium (1.8-3.3) and Low (0-1.7) cgpa and statistical experience were divided into Before (studied statistic prior to university), After (only studied statistics in university) and None (zero statistics experience).

4.3 Qualitative Phase

In order to obtain deeper insight into students' understanding of hypothesis concepts, a qualitative approach has been made by means of interviews. Interviewees were recruited from the final section of the survey.

All the interviews were conducted in March 2021 in a span of 10 days. Students were asked to explain their thought process of answering the multiple-choice questions. At the start of the interview, they were provided with a pen, a sheet of paper and their individual surveys. Interviewees were given two minutes to recall their memory before discussing each item. They were then asked the reason of choosing their answers and/or omitting other options. In addition, they were also allowed to change their answers throughout the interview. All interview sessions were recorded and properly transcribed to ensure the accuracy of this phase.

Furthermore, the follow-up interview helped to answer sub-question 2. The qualitative phase would provide information as to 'what' and 'why' the common errors are made i.e the common misconceptions of hypothesis testing.

5. Results

5.1 Quantitative Phase

5.1.1 Suryey Respondents

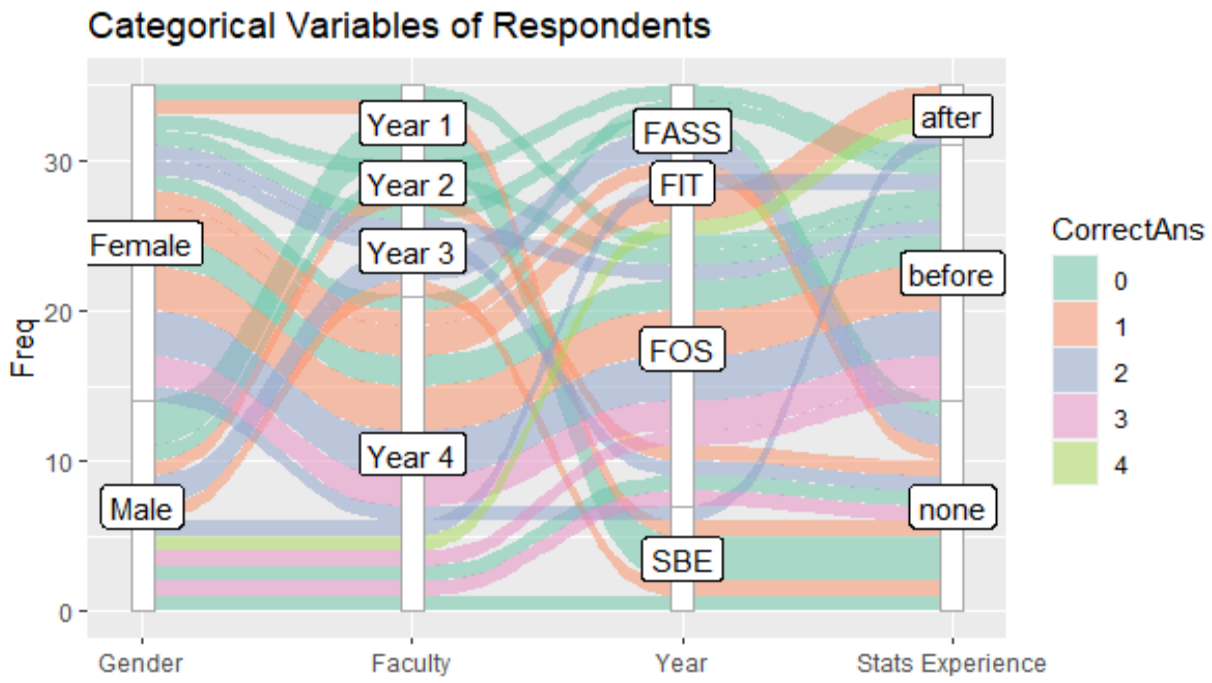

Figure 1: Overall categorical variables of respondents with correct answers

5.1.2 Multiple Choice Questions

This section will talk about the findings from the multiple choice questions. There was a total of 30 responses of the survey. The overall analysis of MCQ scores is presented in Table 2.

Summary	Score	Percent scores (out of 5)
Minimum score	0	0
Maximum score	4	80 %
Median	1.0	20 %
Mean	1.2	24 %

Table 2: The descriptive statistics of survey scores.

The mean total of correct answers is 1.2 points. The result ranged from 0 to 4 points, with only 1 student getting 4 points. A total of 83.3% of the respondents achieved a score of 3 and below. Forty percent of the participants scored 0 on this survey.

To identify the weakest item, analysis was done according to Smith's Framework Category of the five chosen item. Analysis of the items are reported in Table 3 along with the correct (**bolded**) and most common wrong answer of each item.

Question	Category	Option	Frequency	Percent (*n=30)*	Correct & Common Wrong
1	Recognize Applicability	**a**	**9**	**30.0**	**a**
		b	5	16.7	
		c	5	16.7	
		d	11	36.7	d
2	Recognize Applicability	**a**	**6**	**20.0**	**a**
		b	9	30.0	
		c	11	36.7	c
		d	4	13.3	
3	Generate Hypothesis	**a**	**9**	**30.0**	**a**
		b	15	50.0	b
		c	6	20.0	
4	Conclusion	a	6	20.0	
		b	10	33.3	b
		c	**8**	**26.7**	**c**
		d	6	20.0	
5	Decision Rule	a	9	30.0	a
		b	5	16.7	
		c	**7**	**23.3**	**c**
		d	9	30.0	d

Table 3: Frequency of multiple-choice scores.

As shown in Table 3, it is evident that there is a weak performance on the MCQ with correct answers ranging from 20.0% to 30.0%. Question 1 and 3 have the highest percentage of correct answers with 30.0% each. Surprisingly, only question 3 has three chosen answers with option b as the highest percentage of the wrong answer while respondents chose from all possible answer choices for other questions. Most importantly, it is also noted that the percentage of most common wrong answers of all five questions are greater than the correct choice.

5.1.3 Correlation between two variables

One way to investigate the main goal and sub-question 1 was to assess the correlation between two covariates of the study. Pairwise variables were tabulated into a two-way contingency table to perform a chi-square test. Table 4 presents p-values of the chi square test between two variables with the hypotheses as follows:

H_0: Data is evenly distributed (Independence of row/column)

H_1: Data is not evenly distributed (Non-independence)

Address	Correlation between -		Chi square p-value
Main goal	Confidence Level	Correct-Wrong Answers	0.5537
Sub-question 1		cGPA	0.01149 *
		Statistics Experience	0.4948
Sub-research 3		4^{th} yr vs Non-4^{th} year	0.005497 *
	Correct-Wrong	cGPA	0.2704
		Statistics Experience	0.0994
	CGPA	Statistics Experience	0.2799

Table 4: Chi-square test of paired variables.

Data from table 4 shows that only the correlation between confidence level and cGPA has a p-value of less than 0.05 (0.01149). The null hypothesis was rejected and thus, there is evidence at a 5% level that confidence level and cGPA are dependent to each other. The five other pairs were found to be independent of each other except confidence level and final year vs non final year.

5.1.4 Confidence Level

Figure 2 shows boxplots of confidence level of right vs wrong answers of each item. It can be seen that for questions 3, the median of confidence level of respondents who got correct answers are equal to the confidence level of those with wrong answers. This applies to question 5 too. The medians confidence level of respondents who got question 1, 2 and 4 correct are higher than those with wrong answer. However, it can also be seen the median of all these levels are between 5 and 7.5 which surprisingly do not differ too much. Statistically, this is because there are no correlations between confidence level and correct or wrong answer as shown in table 4.

Figure 2: Confidence level of Right vs Wrong Answers

Figure 3 shows the confidence level of final year vs non final year of each questions. Generally, it can be seen that 4th year students have higher confidence in answering the survey questions. There is a correlation between year of study and level of confidence as shown in table 4.

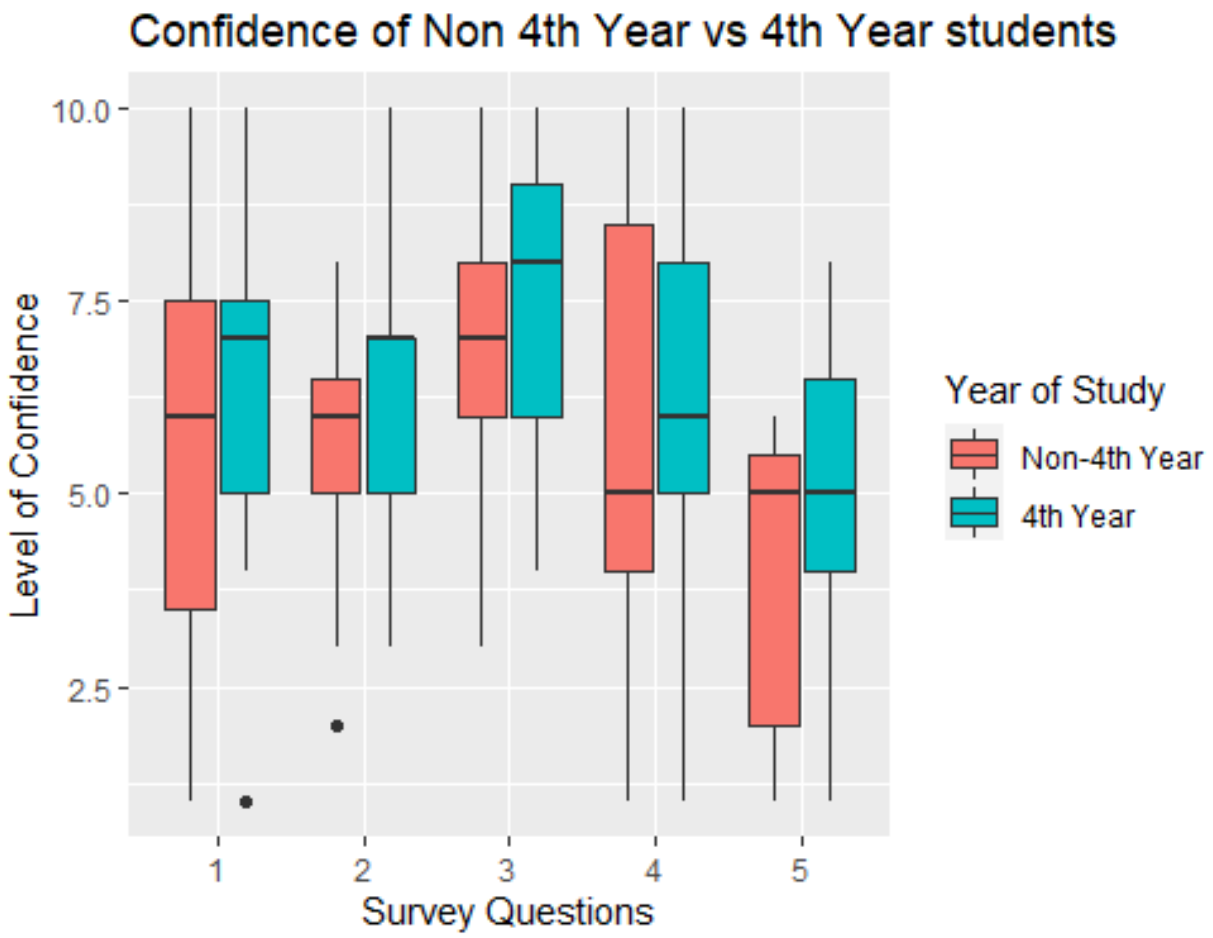

Figure 3: Confidence Level of Final Year vs Non Final Year students

5.2 Qualitative Phase

5.2.1 Characteristic of Interviewees

Number	Gender	Major	Year	Statistical Experience
7	Female	Mathematics	4	Prior & during uni
10	Male	Entrepreneurship	3	Prior & during uni
14	Male	Mathematics	4	None
19	Female	Mathematics	4	Only during uni
20	Female	Biology	4	Prior & during uni
28	Female	Mathematics	3	Prior & during uni

5.2.2 Interviews

This section will present the data obtained from the follow-up interviews to qualitatively examine the study's main goal and research sub-question 2 (common errors made by UBD undergraduate students). The analyses will provide insights of students' understanding of hypothesis question and if any, presence of misconceptions. For each study item, excerpts from transcribed interviews are included to analyse the responses. Statements made by the students will be preceded by an "S" and an "I" denotes statements made by interviewer.

Item 1: Recognize Applicability

1. Which of the following is most likely to be answered by a study that requires statistical hypothesis testing?
   a. Do athletes have a lower GPA than other students?
   b. What equation predicts a student's GPA from his/her IQ?
   c. What are the typical costs for full-time university students in Brunei?
   d. Do the 11:50am Korean language students perform better than the 2:00pm Korean language students this semester?

Option a is the correct answer. This item shows a poor performance of the respondents with only 30% of them getting the correct answer while option choice d was seen to be the most famous choice (36.7%). Option b (16.7%) and c (16.7%) do not need hypothesis testing or simply do not meet the criteria. However, this question tests students' capability to recognize applicability of hypothesis testing. This means that more than half of the respondents understand that hypothesis testing is useful to compare between two groups. Option d implies that students might not understand the power of the test of inferencing samples to a population. This is because option d itself can be answered without hypothesis testing as the population is known to be small. Small population does not need sampling and hence, it can be directly measured from the mean of both groups.

Of the six interviewees, 2 chose a, 3 chose d and 1 chose b. Throughout the interviews, 1 person changed from b to a and 1 person changed from a to d. Overall, by the end of the interview, 4 interviewees chose d and 2 picked a.

In the interviews, students gave similar explanations for option b and c stating that these two options were not comparing between two groups. Some of their explanations are shown below:

Excerpt: Interviewee 7

S – I omit b because it's an equation so it doesn't really ask about hypothesis testing and c is too broad..

Excerpt: Interviewee 14

S - For study that needs statistical hypothesis testing is usually for comparison and sorts..um and uh..I can actually use hypothesis testing for question b but I only use hypothesis testing to see if the equation is actually good or should I say good um.. just the goodness of fit ..and c is uh the (reads option c) that's..that doesn't require any hypothesis testing as you can certainly use um.. the mean or average and stuff.

Excerpt: Interviewee 28

S – Hypothesis is to know whether they're the one that perform better or not (I - so comparing) yeah comparison. I didn't chose b because it only compares equation.

I – How about c?

S - For c ..I think it's because there's no sample, only cost

Most of the interviewees admit that they were confused between option a or d. Interestingly, interviewees who answered d mentioned about the two stated times or that the times made option d more specific. Explanations are given below:

Excerpt: Interviewee 11

I – Why did you chose d?

S – Um..because d has the most statistical data, more specific like 1150 and 2pm

Excerpt: Interviewee 20

S – From what I know hypothesis testing.. that means I need number comparing number between 2 things. Out of all choices I would get 2 numbers so d is the answer.

Excerpt: Interviewee 7

I – Can I know why did you change from a to d?

S – Umm.. it is stats hypothesis testing but the test ..um thats in general. I think the time for option d could not be significant enough or ….(inaudible) to give difference in performance so by the end of the day they're still studying at home not.. just during the class. Meanwhile for athletes, that like that's them- its different than…yeah I don't think I can explain it better.

Interviewees 11 and 20 chose d because the option was more specific. Interviewee 20 implied that the two times could be used for hypothesis testing to compare the groups. Interviewee 7 thought

that the stated time does not give any significance but still chose d because a was too broad. However, Interviewee 14 gave a different explanation as to picking option a and said that the two groups in d were from the same sample. Interviewee 28 also thought that similarly.

Excerpt: Interviewee 7

S - and for part d (reads d inaudibly) that is uh.. it is also comparison group but uhh it's considered as from the same sample

I – From the same sample, why?

S - That's like splitting sample into 2 groups.. it's just 1150am class and compared with 2pm class.. it is virtually the same

Excerpt: Interviewee 28

S – I'll choose a because that's like it is comparing students..non athletes and athletes but I'm still stuck in between a and d…but I think I choose a because it's like the athletes are also students so it's like they're doing different things than other students..

Summary

All the interviewees recognize that statistical hypothesis testing is useful to compare between two groups or two samples. However, they did not understand that hypothesis testing is used to infer population. Those who answered option d did not understand that a small known population, in this case, does not need statistical hypothesis testing. This result was similar to Smith (2008) study. Smith (2008) stated that *"They did not mention the role of inference in statistical hypothesis testing. They did not mention that statistical hypothesis tests provide a means of dealing with the variability that exists between and among samples."*

Item 2: Recognize Applicability

2. Which of the following statements is the *best justification* for using a statistical hypothesis test to answer the question: Are female students as successful as male students in college mathematics?
   a. It allows you to talk about uncertainty associated with your decision.
   b. It allows you to use only a sample of students to prove something for all students.
   c. It allows you to calculate means and use statistical tables to answer the questions.
   d. It allows you to find and prove the answer using mathematical calculation.

Option a is the correct answer. Only 20% of the respondents answered this correctly while option c was the most popular choice with 36.7% of the respondents. 30% of them chose b and 13.3% answered d.

This item also tests whether students perceive statistical hypothesis testing to be a proof if they choose b or d. Students who believed that statistical hypothesis testing provide proofs held a misconception that sample proves a statement for a poulation.

Option c is true but is not the best justification among the four choices. Students who chose option c might understand that hypothesis testing is not a proof. This question also assessed whether students realize that statistical hypothesis testing provide a way to talk about uncertainty associated with a decision or if students think that the test is only to compare means and looking up values in statistical tables.

Of the six interviewees, options a, c and d were chosen by two people each. Throughout the interviews, one person changed from c to d, one person changed from d to a and one person changed from d to b. Overall, at the end of the session, 3 interviewees chose a and options b, c and d were chosen by one person each.

Some interviewees thought that statistical hypothesis testing is only about procedures or more to mathematical. Some of their explanations are shown below:

Excerpt: Interviewee 7

S- I probably change it to d (I- Why?) because c is just comparing between the means but for d, it goes more in depth..it goes more deeper like finding p-values.

Excerpt: Interviewee 10

S – Because it says to calculate means and use statistical table to answer the question so it's..it is more of a analyzing data than analyzing uh…qualitative data, it's more to quantitative data

I - So why did you omit option a?

S - a is more to qualitative, b is also more to qualitative and d is a bit confusing

I - Why is it confusing?

S- I, for one, don't know how you can find and prove the answer using mathematical calculation yeah I don't how it's related..it doesn't ..I mean statistical doesn't mean it's necessary related to calculation right? It's more about statistics.

Excerpt: Interviewee 20

S - d is because.. it's just.. that's the answer…hmm ..since its asking about hypothesis uh that's something to do with maths right and d mention something about maths . hmm maybe the answer is c.

During the session, interviewee 20 switched answer from d to b. Both options are related to the misconception of a sample proving a statement for population. However, interviewee 20 only changed answer after requesting interviewer for the definition of statistical hypothesis testing. This indicates that interviewee might not be familiar with statistical hypothesis testing at all even if he/she has statistical experience prior to and during university. Meanwhile, during the interview, interviewee 10 mentioned that he/she had a short statistical hypothesis class prior to university level although he/she stated no prior statistic experience in the survey.

However, three interviewees chose answer a with their explanations below:

Excerpt: Interviewee 14

S – and the reason why I chose a because it allows you to talk about uncertainty associated with your decision so when uh.. I take sample of this population and I ..(inaudible) hypothesis testing and I should have a level of uncertainty so if I expand the sample into population, I'm confident that I have 5% chance or 1% that I could be wrong.

Excerpt: Interviewee 28

S - I think I change to a because um.. hypothesis testing is basically talking about the.. the concepts of whether for this case female students successful than male students so there's uncertainty there so um I think if I eliminate b ..it's like this is correct but to prove, it's a vague sentence so maybe it's to prove the uncertainty…and for c (reads) I think this is also the steps under it, so the final answer is uncertainty for me. All of these (other options) are the steps to get to hypothesis test.

Excerpt: Interviewee 19

S – I chose a because as I said statistical hypothesis testing is to know if our plan is true or not so I eliminate b, c and d. Option c & d because more to calculation

I – What do you mean by calculation?

S – It's true that we use calculator for hypothesis testing but um…how to say this.. calculation um…its calculation, we haven't done proving them …I'm not confident between a and b but I still stick to a because it is my decision to choose null or claim

Interviewees 14 and 28 realize that statistical hypothesis testing provide a way to talk about uncertainty associated with a decision. Interviewee 19 chose option a because of the "associated

with your decision" to choose own's hypotheses. However, none of them understand that hypothesis testing does not prove any statement for the population.

From Item 2, some interviewees were able to understand that hypothesis testing is more than just a quantitative tool. However, even if they chose the right answer, students in general do not understand that hypothesis testing is not a proof but instead, an inferential method.

Item 3: Generate Hypothesis

3. A researcher would like to establish that humanities students have higher CGPA than science students. Which of the following null hypotheses should be tested? The mean CGPA of humanities students is:
   a. greater than that of science students.
   b. greater than or equal to that of science students.
   c. less than or equal to that of science students.
   d. less than that of science students.

The correct answer is c and answer a is the alternative hypothesis. Only 30% of the respondents answered this item correctly while half of the respondents chose the answer b. This item tests whether students are able to generate hypotheses in this scenario. In addition, this question also addressed if students realize that null hypothesis should contain *equality*.

Within the interviewees group, 3 people chose b, 2 people chose c and 1 person chose d. During the interviews, 1 person switched from b to a and 2 interviewees change from c to b. By the end of the interview session, 5 people chose b, 1 person picked a and no one answered c and d.

Only interviewee 7 chose the answer a with the explanation below;

> **Excerpt: Interviewee 7**
>
> S - I think I'm changing it to a. For the question, the humanities have higher cgpa than science students so that means….the means of cgpa should be more than the science students which eliminates c and d and then difference equals to ..(inaudible) that means the cgpa is equal does not mean humanities has higher cgpa so a.

Surprisingly, the rest of the interviewees chose the same answer. Some of their explanations are shown below:

Excerpt: Interviewee 28

S – b because the null hypothesis is like.. it's a statement so here the statement is like 'humanities have higher cgpa than science students' so from that we can assume that the null hypothesis is that humanities have higher cgpa than science.

Excerpt: Interviewee 19

S - Change to b because from the question.. humanities students have higher cgpa than science students so the null hypothesis should be greater than or equal to..same like the question, because of the statement. It's not (option) a because it's possible science students have same cgpa with humanities cgpa so null is b.

Only two of the interviewees mentioned that null should contain equal or nullity. The excerpts are presented below:

Excerpt: Interviewee 14

I - You answered b, why?

S - I actually omit a and d because for null hypothesis, its better to have the word equal so we can get rid of it so I'm stuck between answer b or c. so ..hmm but one thing im sure is there need to be an equal word inside the answer for the null…

Excerpt: Interviewee 20

S - I got confused with that. I just realized null is 0 so it should .. since the hypothesis we are testing is to establish that humanities students have higher cgpa than science students so the null hypothesis should be that science students has higher cpga than humanities so im not sure which one should be the null hypothesis.

The high percentage of respondents picking option b implies that students are well-familiar that null hypotheses should contain equality. This is also evident from the interviews. Some interviewees acknowledge that the null and alternative hypotheses should contradict each other. However, question 3 shows that interviewees are not able to generate hypotheses clearly as they did not realize the statement in the question could be the alternative instead.

Item 4: Decision Rule

4. A study tested the claim that: 'Transfer students are less successful at UBD than students admitted as first-time freshmen'. Information from samples is shown below. What is the most reasonable inference about the population of all first semester students that can be drawn from this information?

	Transfer Admits	Freshman Admits
n	50	50
mean GPA	2.5	2.8

a. There are equal numbers of transfer and first-time freshman students.

b. The mean first semester GPA of all freshman admits is 0.3 greater than that of all transfer admits.

c. It is unlikely that the first semester GPA of all transfer admits equals that of all freshman admits.

d. The mean first semester GPA of all university students is (2.5+2.8)/2 or about 2.65.

Option c is the correct answer. Only 26.7% of the respondents picked the correct answer whereas the top choice is option b (33.3%). This item tests students' understanding of inferencing a sample to the population.

Within the interviewees, two people answered a, two people answered b, one person chose c and one person chose d. Throughout the interviews, one interviewee switched from a to c and another one switched from a to d. Overall, two answered b, two answered c and two answered d. No one chose option a.

Interviewees who got the correct answer correctly understood hypothesis testing as the means of inferential tool. Some explanations are shown below:

Excerpt: Interviewee 7

S - For a I thought maybe I was rushing that time but since this is a sample so we don't know the actual number so a is not…For b it's generalizing for all freshman all transfer

I – so you're saying that's wrong why?

S – like for this sample it could be like this but for another sample it could be different. For d, it's just wrong and for c, I think because its very unlikely for freshman students to have the same gpa for all transfers.

Excerpt: Interviewee 14

S – option a (reads) it is true but it has nothing to do with the inference of the population.hmm..d.. I mean yeah d is true but it doesn't say anything about the inference and b is wrong because it doesn't include the 5% sig level because wait I wont say its wrong I'll say its incomplete because it doesn't say the (reads) at a sig level of 5% so it means that there's a 5% chance that the statement is not true. For c (reads) that is true bcs even if it is unlikely it can still happen and if it happens it belongs to the 5%.

Some interviewees find option c to be vague or could not find the relation of option c and the question. Some examples of an explanations are:

Excerpt: Interviewee 20

S - Im most confident with d because it's a fact. It's not c hmm im not confident with that. I don't get uh like the info from the table is not equals to c. I don't know how c got it from the table.

Excerpt: Interviewee 19

S – It's either b or c but I still stick to b because of the information given like there's 0.3 but c..it's not related to the information

I - Why is it not related?

S - Like uh it's a vague sentence.

This item shows that four of six interviewees do not have a deep understanding of hypothesis testing as an inferential tool. A few of interviewees thought that the statement from samples could prove a population which is a common misconception.

Item 5: Conclusion

5. To test the hypothesis that private schools are more effective than public schools, researchers plan to compare the mean starting salaries of private and public school graduates. But they cannot agree on whether to test the results at a significance level of 0.05 ($\alpha = 0.05$) or at a significance level of 0.10 ($\alpha = 0.10$). What effect will using 0.10 rather than 0.05 have on the study?

   a. Using 0.10 will result in a greater chance that they will incorrectly retain the hypothesis.

b. Using 0.10 will result in a greater chance that the null hypothesis is actually true.

c. Using 0.10 will result in a greater chance that they will incorrectly reject the null hypothesis.

d. Using 0.10 will result in a greater chance that the alternative hypothesis is actually true.

The correct answer, c, was answered by 23.3% of the respondents. Both options a and d were equally the top answers, with 30% each. This item test the understanding of significance level.

Within the interviewees, three people chose c and the other 3 options were chosen by 1 person each. Throughout the interview, one person changed from a to c, one person changed from c to b and one person switched from d to c. Overall, at the end of session, four people picked c and 2 people picked b.

Some interviewees managed to explain significance level correctly by drawing a normal distribution with rejection regions on both side of tails. However, those who do not have a deep understanding of this either thought that null hypothesis can be true or that they do not know understand what alpha values are for. Interestingly, one interviewee mixed confidence level and significance level.

6 Discussion

As stated in Introduction, the primary goal of this study was to investigate Universiti Brunei Darussalam (UBD) undergraduate students' confidence of hypothesis testing. The major finding of the study was the presence of poor understanding of hypothesis testing as this was apparent from the low performance of survey. However, this leads to the discovery of potential hypothesis testing misconceptions that was further assessed in the qualitative study.

In regard to sub-question research 1, our result shows that there is a correlation between confidence in student's knowledge of hypothesis testing and cGPA. This was calculated using chi square test. However, the test also shows that there is no correlation between confidence level and correct or wrong answers. In general, the result shows that students with higher cGPA are generally more confident with their answers of the study whether the answers are true or false.

As for sub-question research 2, the most common errors made by UBD undergraduate students is to generate hypotheses. Due to the fact that majority of the students do not held strong understanding, it was clear that a high proportion of students thought that hypothesis testing is used to provide proof of population from samples. In addition, data from the survey and interview implies that students also thought that hypothesis testing are more to procedural way to prove null hypothesis. However, these misconceptions were also addressed in Smith (2008) and Sotos et al (2009). For instance, Sotos et al (2009) states that "..the two major misconceptions for the concept of hypothesis test were just as popular among our students: 20% of them believed that a hypothesis test is a mathematical (logical) proof of the null hypothesis"

Regarding sub-research 3, it was found that there is a correlation between confidence level of hypothesis testing of final year UBD undergraduate students vs non final year. Quantitative result shows that final year student has a higher mean of confidence level than non-final year.

Limitation and Recommendations

First of all, access of exam results of statistic classes was one of the limitations of this study. In order to get a similar data, cGPA of students were asked in the survey. This data was needed to answer one of the sub-question research. Students' cGPA could only be obtained from the students themselves. This results in a lower response rate of the survey as students prefer not to share their exact cGPA. A better recommendation is to use intervals of cGPA levels instead.

Secondly, the reduced number of questions to ease the survey questions results in a harder analysis. Pilot study shows that student might be tired after answering 7 statistical questions but smaller number of questions was harder to be analysed in fear of biasness. A recommendation for this problem is to have a target population of statistical students only. This way, both the response rate and questions can be increased to give a better effective research.

7 Conclusion

With the outcome, it is hoped that this research could help lecturers to diagnose any common error or misconceptions of statistical knowledge among UBD undergraduate students.

8 Acknowledgements

I would like to thank my supervisor, Dr Haziq Jamil, who always give fruitful knowledge during my research study. I thank all my respondents and interviewees for participating in my research study.

References

Batanero, C. (2000). Controversies around the role of statistical tests in experimental research. Mathematical thinking and learning, 2(1-2), 75-97.

Burns-Childers, A., Rock, L., Chamberlain Jr, D., Kemp, A., Meadows, L., Stalvey, H., & Vidakovic, D. (2017). Students' Understanding of Test Statistics in Hypothesis Testing. In Proceedings of the 20th annual conference on Research in Undergraduate Mathematics Education (pp. 82-92). SIGMAA on RUME San Diego, CA.

Dolor, J., & Noll, J. (2015). USING GUIDED REINVENTION TO DEVELOP TEACHERS' UNDERSTANDING OF HYPOTHESIS TESTING CONCEPTS. Statistics Education Research Journal, 14(1).

Mevarech, Z. R. (1983). A deep structure model of students' statistical misconceptions. Educational studies in mathematics, 14(4), 415-429.

Nickerson, R. S. (2000). Null hypothesis significance testing: a review of an old and continuing controversy. Psychological methods, 5(2), 241

Smith, T. M. (2008). An investigation into student understanding of statistical hypothesis testing (Doctoral dissertation).

Sotos, A. E. C., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2009). How confident are students in their misconceptions about hypothesis tests?. Journal of Statistics Education, 17(2).

UBD (n.d). Quick Facts. Universiti Brunei Darussalam. Retrieved 2020, November 12 from https://ubd.edu.bn/about/quick-facts/.

UBD (n.d). University Research Ethics Committee (n.d). Universiti Brunei Darussalam. Retrieved 2020, November 12 from https://ubd.edu.bn/research/support-for researchers/university-research-ethics-committee.html.

Vallecillos, A. (2000). Understanding of the logic of hypothesis testing amongst university students. Journal für Mathematik-Didaktik, 21(2), 101-123

Supporting Materials

P-value = 0.5537	High	Medium	Low
Wrong	29	67	18
Correct	10	23	3

Table: Contingency table of Confidence level and Correct/Wrong Answers

P-value: 0.2704	High	Medium	Low
Wrong	34	72	8
Correct	6	28	2

Table: Contingency table of CGPA and Correct/Wrong Answers

P-value: 0.09445	Only in University	Prior to University	None
Wrong	8	62	44
Correct	7	18	11

Table: Contingency table of Statistical Experience and Correct/Wrong Answers

P-value: 0.01149	High cGPA	Med cGPA	Low cGPA
High	7	32	0
Medium	31	51	8
Low	2	17	2

Table: Contingency table of Confidence Level and cGPA

P-value: 0.4948	Only in University	Prior to University	None
High	5	24	10
Medium	8	47	35
Low	2	9	10

Table: Contingency table of Confidence Level and Statistical Experience

P-value: 0.2799	Only in University	Prior to University	None

High	0	5	3
Medium	2	11	7
Low	1	0	1

Table: Contingency table of cGPA and Statistical Experience

Appendix A: Survey

Appendix B: Survey data