Journal of Pedagogical Sociology and Psychology
Volume 4, Issue 1, 2022
https://doi.org/10.33902/JPSP.202212259

Research Article

Undergraduate students' proficiencies in solving bivariate normal distribution problems

Bosire Nyambane Onyancha[1] and Ugorji I. Ogbonnaya[2]

[1]*University of South Africa, Unisa, South Africa;* [2]*University of Pretoria, South Africa.*

Correspondence should be addressed to Bosire Nyambane Onyancha nyambaneo@gmail.com
Received 2 June 2021; Revised 24 November 2021; Accepted 23 January 2022

This study explored undergraduate students' proficiencies in solving bivariate normal distribution (BND) problems in a Kenyan university. The study followed a case study design and qualitative research approach. One hundred and seventy-five undergraduate statistics students in a Kenyan university participated in the study. Data was collected using an achievement test. Content analysis of the students' solutions to test questions revealed that majority of the students were not proficient in solving BND problems with respect to calculating; (i) the probability of a normal distribution given the mean and variance of a variable, (ii) the mean of a normal distribution given the variance and the probability of a variable, (iii) the mean and variance of the joint distribution, and hence the probability of the variable given the conditional distribution of a variable, and (iv) the mean and standard deviation of two random variables given a bivariate random density function. It is recommended that the basic statistical concepts relevant to learning the BND be thoroughly revised before formally teaching BND.

Keywords: Bivariate normal distribution, joint distribution, probability, random variables, statistics

1. Introduction

A normal distribution (also referred to as Gaussian distribution) is a probability function in which the values of a variable are symmetric around the central mean (Bhandari, 2021). For a normally distributed variable, the mean and median are equal, and 68% of the data falls within 1 standard deviation, 95% of the scores lie within 2 standard deviations of the mean and 100% (precisely 99.74%) of the scores lie within 3 standard deviations of the mean (Gordon, 2006). A standard normal distribution has a mean equal to 0 and a variance of 1 and it is denoted as Z~N(0,1), (Rouaud, 2017). The normal distribution is an important probability distribution in statistics because many psychological and educational variables are normally distributed (McLeod, 2019). The normal distribution and the bivariate normal distribution (BND) form part of the content of most statistical courses in the Kenyan university as per the university's course outlines.

A bivariate normal distribution is a joint normal distribution involving two random and normally distributed variables (Weisstein, 2002). The BND function is obtained from the product of two marginal distribution functions (probability density function) for two random and normally distributed variables, say X and Y. A random variable X is said to be normally distributed if its probability density function is $f(x) = \frac{1}{\sigma_x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2}$ and a random variable Y has a normal distribution if its probability density function is $f(y) = \frac{1}{\sigma_y\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2}$.

Then the joint probability density function $f(x,y)$ for two dependent random and normally distributed variables X and Y with functions $f(x)$ and $f(y)$ respectively, is given by;-

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) - \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$$

where the correlation coefficient $\rho = Cor(x, y) = \frac{V_{xy}}{\sigma_x \sigma_y}$, $0 \leq x \leq \infty, 0 \leq y \leq \infty, \sigma_x > 0, \sigma_y > 0$ and $-1 \leq \rho \leq 1$, $\mu_x = E(x)$, $\mu_y = E(y)$, $\sigma_x = standard\ deviatin\ of\ X$, $\sigma_y = standard\ deviatin\ of\ Y$.

A bivariate normal distribution finds application in many fields of life including economics, engineering, actuarial science, and medicine (Grover, Sabharwal & Mittal, 2014; Lin, Dou, Kuriki, & Huang, 2014). For instance, BND could be used to find the relationship between the marks obtained by students in a class in two subjects, and the sales revenue and advertising expenditure of various branches of a company in a particular period.

In the Kenyan university, where this research was conducted, the BND is taught as a major part of the content in Statistics courses. The students encounter difficulties in solving normal distribution and BND problems as evidenced by their dismal performances in statistics courses in which the content is examined. Shojaie, Aminghafari and Mahammadpour (2012) acknowledged that students experience difficulties in introductory courses in probability and statistics which include finding the joint distribution of a function and calculation of bivariate expectation. Similarly, some other studies (Harradine, Batanero & Rossman, 2011; Lugo-Armenta & Pino-Fan, 2021; Sotos, Vanhoof, Van den Noortgate & Onghena, 2007) acknowledge that statistics in general is challenging for students at various levels of education.

Memnun, Ozbilen and Dinc (2019) observed that many students experienced difficulties in solving probability problems and failed to apply different probability concepts in problem-solving. Similarly, Arum, Kusmayadi, and Pramudya (2018) showed that students experience difficulties in solving probabilistic problems. This can probably contribute to difficulties students experience in further learning of advanced statistical concepts. According to Garfield and Ben-Zvi (2008), many students view statistics as a difficult topic that is unpleasant to learn. These difficulties in basic probability and statistics concepts could be the reason for the students' challenges in solving advanced statistical problems, such as BND problems. Adequate foundational knowledge of a normal random variable forms the basis for effective learning of BND (Flury, 2013; Tacq, 2010) and thus possible improved proficiency in solving BND problems among undergraduate students in the Kenyan university. Koparan (2015) indicated that students encounter difficulties in understanding and evaluating variance concepts, which may essentially hinder the learning of other related advanced statistical concepts, for example, the BND. Kachapova and Kachapov (2012) acknowledged that students develop misconceptions about a random variable as much as it is a fundamental concept in probability theory that is not intuitive.

Ideally, students need to possess excellent knowledge and computational skills of the probability and mean of a normal distribution to correctly solve BND problems. Literature seems very sparse on undergraduate students' learning of bivariate normal distribution and consequently, their proficiency in solving BND problems. No study was found that focused on students' difficulties or proficiencies in solving BND problems. Hence, this study explored undergraduate students' proficiency in:

a) calculating the probability of a normal distribution given the mean and variance of a variable.

b) calculating the mean of a normal distribution given variance and the probability of a variable.

c) calculating the mean and variance of the joint distribution and hence the probability of the variable given the conditional distribution of a variable.

d) calculating the mean and standard deviation of two random variables given a bivariate random density function.

2. Mathematical Proficiency

The study draws from Kilpatrick, Swafford and Findell's (2001) mathematical proficiency framework. The framework consists of five components or strands, namely; conceptual understanding (which deals with the comprehension of mathematical concepts, operations, and relations), procedural fluency (which entails skills in carrying out procedures flexibly, efficiently and appropriately), strategic competence (which deals with the ability to formulate, represent and solve mathematical problems), adaptive reasoning (which entails the capacity for logical thought, reflection, explanation, and justification), and productive disposition (which involves habitual inclination to see mathematics as sensible, useful and worthwhile, coupled with a belief in diligence and one's efficacy), (Gloves, 2012; Kilpatrick, Swafford & Findell, 2001; Schoenfeld & Kilpatrick, 2008). Kilpatrick, Swafford, and Findell (2001), further assert that the five components of mathematical proficiency are interwoven and interdependent in the development of proficiency in mathematics. Hence, the five strands of mathematical proficiency are needed for one to be proficient in mathematical problem-solving. In this study, however, we believe that the students' proficiencies in solving BND problems lie heavily on their conceptual understanding, procedural fluency, strategic competence, and adaptive reasoning.

3. Methodology

3.1. Research Design

The study adopted a case study design and qualitative research approach. Creswell (2014) defines a case study design as a research inquiry that explores an in-depth analysis of an event, an activity, or a process of one or more individuals to collect detailed information about a problem of study.

3.2. Participants

The participants were 175 (113 males and 62 females) comprising 143 second-year and 32 third-year undergraduate statistics students in a Kenyan university. The students had completed the probability and statistics III unit where the BND content was covered. The students were of mixed abilities. They were conveniently chosen to participate in the study based on their consent to be part of the study. The study was made open for all students to participate in.

The normal distribution and the BND are taught as a major part of the content in Statistics in the Kenyan university. The students experienced difficulties in solving BND problems evidenced by the below-average performance in probability and statistics III which covers BND content.

3.3. Data Collection

Data was collected using a BND test. The test items were developed by the first author with the guidance of a statistics lecturer at the Kenyan university. The BND test comprised of four questions; two questions were from the normal distribution and two other questions from the bivariate normal distribution.

Question 1 of the BND test, asked the students to calculate the probability of a normal distribution given the mean and variance of a variable. The knowledge of calculating the probability of a normal distribution is applied in solving the probability of a joint distribution of a BND using standard normal distribution table. The students were required to standardize a normal variable and use the standard normal distribution table in finding the probability of a normal distribution.

Question 2 required the students to calculate the mean of a normal distribution variable given the variance and probability of a variable. The probability of a joint distribution of BND uses the basic foundational knowledge of the probability of a normal distribution variable of standardizing the normal scores, in solving questions. The question required the students to obtain a standardized normal value from the normal distribution table given a probability of 1%, and then substitute the Z-value, the standard deviation, and the value of the variable in the expression, $Z = \left(\frac{X-\mu}{\delta}\right)$ to obtain the mean of the variable.

Question 3 asked the students to calculate the mean and variance of two random variables of a BND and use the values to determine the probability of a joint distribution. The question evaluated the conditional mean and variance aspects of a BND and the calculation of the probability of a BND. The students were required to use the conditional mean and variance equations of a joint distribution of a BND to calculate the mean and variance, thereby obtaining the Z-score value, which is subsequently used to find the probability of the probability of the joint distribution.

Question 4 required the students to solve for the mean and standard deviation of two random variables of a joint distribution given a bivariate random density function. The question examined the use of the BND equation in solving for the parameters of a given random density function. The question required the students to compare and equate the general BND function to the bivariate random density function, their parameters, and variables, to obtain the mean and standard deviation of the two random variables of a joint distribution.

Validity of the BND test was ensured by a statistics lecturer in the university who evaluated the questions to ascertain that they were well constructed and covered all aspects required for testing undergraduate students' proficiency in solving BND problems. The lecturer established that the test items were within the university's statistics course outline and measured the intended content. The reliability of the BND test was computed using Cronbach's alpha (α) and was found to be 0.826. Thus, the reliability of the BND test was within the required range and therefore good for use in the research study.

3.4. Data Analysis

Data analysis involved deductive content analysis (Krippendorff, 2018) of the students' solutions to the test questions. This helped to identify their proficiency in solving the questions. Descriptive statistical calculations of frequencies of the students' solutions of BND questions classified as completely correct, partially correct, incorrect, and not attempted were done. Students who solved each of the questions correctly were regarded to be proficient, while those who partially solved the questions correctly were said to be averagely proficient. And those who either solved the BND questions wrongly or failed to attempt the questions were regarded to be non-proficient. The descriptive analysis of frequencies of solutions of the students thus gave insight into their level of proficiency in solving BND problems in the Kenyan university.

4. Findings

The results of the data analysis are presented in order of the test questions. This is followed by a summary the students' proficiencies in solving the BND test items.

4.1. Calculation of the Probability of a Normal Distribution given the Mean and Variance of a Variable

The students were asked to solve the following question to explore their proficiency in calculating the probability of a normal distribution given the mean and variance of a variable.

Question 1

Given that, $X \sim N(1,4)$, find $P(1 \leq X \leq 3)$

From the question, the variable X is normally distributed with mean and variance equal to 1 and 4 respectively. The students were expected to calculate the standard deviation (square root of variance), then use it, the mean, and the value of the variable (1 and 3) to find a standardized normal value, Z. This value is used to read the probability of a normally distributed variable from the standard normal distribution table.

For a normal distribution variable $X \sim N(\mu, \sigma^2)$, μ is the mean of the variable and σ^2 is the variance of the variable. From the given question, the mean is 1 and the variance is equal to 4. The square root of the variance, σ^2 equals to the standard deviation, σ. From the question, the following is given: $\mu = 1 \; and \; \sigma = \sqrt{4} = 2, \; since \; \sigma^2 = 4, X = 1 \; and \; X = 3$.

The probability of a variable between $X = 1 \; and \; X = 3$ is expected to be solved as follows:

Hence, $P(1 \leq X \leq 3) = P(\frac{1-1}{2} \leq \frac{X-1}{2} \leq \frac{3-1}{2})$

$= P\left(\frac{0}{2} \leq \frac{X-1}{2} \leq \frac{2}{2}\right)$

$= P(0 \leq Z \leq 1)$

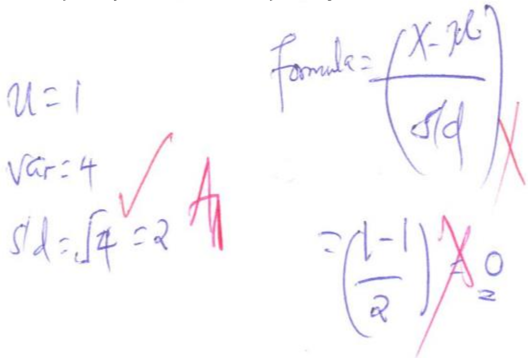
$= 0.3413$ (Read from the standard normal distribution table)

Therefore, $P(1 \leq X \leq 3) = 0.3413$.

Out of 175 undergraduate students, only 51 students answered the question correctly, 47 students solved the partially correct, 28 students solved the question wrongly and 49 students did not attempt the question. Figure 1 shows the work of one student who solved the question partially correctly. The student calculated the standard deviation correctly but did not solve the standard normal value, Z, which is used to find the probability of variable X.

Figure 1

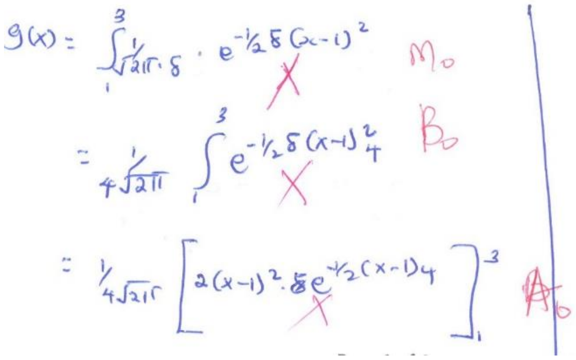
Example of a students' partly correct solution to question 1



Another student (Figure 2), solved the question by wrongly attempting to integrate the probability density distribution function of a normal variable by substituting $X = 1\ and\ X = 3$ as limits of integration.

Figure 2

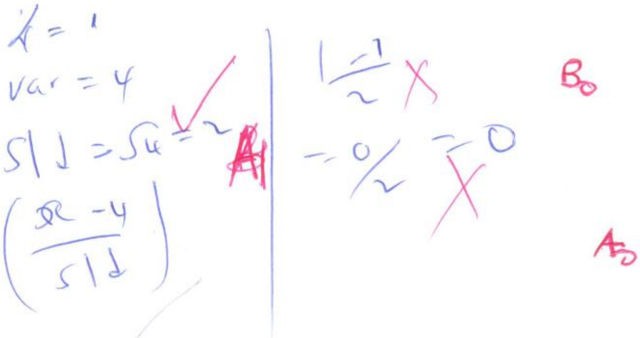
Example of a students' wrong solution to question 1



Yet another student (Figure 3) was able to determine the mean, variance and the standard deviation but could not find the probability of the variable X. This implies that the student had inadequate understanding of calculating the standardized normal value of a variable to help find the probability of a normal distribution. Similar errors were observed in the solutions of the question by other students.

Figure 3

Example of a students' wrong solution to question 1



The errors committed by students in solving this question can be attributed to their insufficient knowledge, and their failure to recall normal distribution concepts. Generally, the students performed poorly in solving the probability of a normal distribution which demonstrated their lack of conceptual understanding and procedural fluency proficiency in solving normal distribution questions. It further demonstrated their lack of adequate knowledge of a normal distribution which possibly hinders their learning and understanding of the BND.

4.2. Calculation of the Mean of a Normal Distribution given Variance and the Probability of a Variable

The students were tasked to solve the following question to explore their ability in calculating the mean of a normal distribution given variance and the probability of a variable.

Question 2

> In regulating the blue dye for mixing paint, a machine can be set to discharge an average of μcm/can of paint. The amount discharged is $\mathrm{N}(\mu, 0.16)$. If more than 6cm is discharged into the paint can, the shade of blue is unacceptable. Determine the setting μ so that only 1% of the cans of paint will be unacceptable.

The question required that the undergraduate students calculate the mean of a random variable given variance and the probability of a variable. The students were expected to use the standard normal distribution table to find a value that gives a probability of 1% = 0.01 with a standard deviation of 0.4, $(\sqrt{0.16})$ and whose variable, $X = 6$. The standard normal value, $Z = 2.33$, the standard deviation, $\delta = 0.4$, and the variable $X = 6$ are then substituted in the expression, $Z = \left(\frac{X-\mu}{\delta}\right)$. The value, μ of the variable is obtained. The students' expected solution is as shown below;-

Let X be the amount of dye discharged to the can, then $X \sim N(\mu, 0.16)$, that is, the amount of die discharged is normally distributed with mean, μ, variance of 0.16 and a probability of 0.01(1%).

Therefore, the mean μ is determined such that; $P(X > 6) = 0.01$

$$0.01 = P(x > 6) = P\left(\frac{x-\mu}{0.4} > \frac{6-\mu}{0.4}\right) = P(Z > \frac{6-\mu}{0.4})$$

From the standard normal distribution table, a probability of 0.01 gives a standardized normal value of 2.33

Hence, the expression $\frac{6-\mu}{0.4}$ = 2.33

On solving;

$6 - \mu$ = 0.4(2.33)

And μ = 6 – 0.4(2.33)

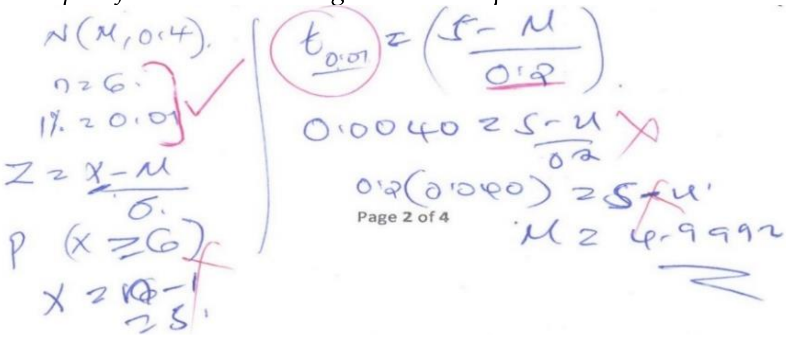
= 6 – 0.932

= 5.068

Hence the mean should be set at μ = 5.068, so that only 1% of the cans of paint can be unacceptable.

The result showed that no student solved the question correctly, 4 students solved the question wrongly and 171 students did not attempt the question. Figure 4 below shows a sampled students' wrong solution of the question. The student was unable to obtain the standard normal value, Z value from standard normal distribution table. The student also used a wrong standard deviation value of 0.2 instead of 0.4. Further, it is evident that the student had a wrong interpretation of the normal distribution problem.

Figure 4
Example of a students' wrong solution to question 2



The inability of students to calculate the mean of a normal distribution given variance and probability of a variable, which is a basic concept of BND, demonstrated their difficulties in solving normal distribution questions. This further showed a lack of strategic competence, conceptual understanding, and procedural fluency on the part of students in solving normal distribution problems.

Ideally, students with inadequate knowledge of mean and probability of a normal distribution experience difficulties in learning BND, an advanced concept of the normal distribution.

4.3. Calculation of the Mean and Variance of a joint Distribution and the Probability of the Variable given a joint Distribution of a Variable

The students were asked the question below to explore their know-how in calculating the mean and variance of two random variables and hence the probability of a joint distribution.

Question 3

In a certain population of married couples, the height x ft of husband and the height y of the wife have a bivariate normal distribution with parameters; the mean height of husband, μ_1=5.8 ft, mean height of wife, μ_2=5.3 ft, standard deviations, $\delta_1 = \delta_2$=0.2ft and correlation coefficient, $\rho = 0.6$.

i) Find the conditional p.d.f of y given that x is 6.3 ft.

ii) What is the probability that the height of the wife lies between 5.28 ft and 5.92 ft given that the height of the husband is 6.3 ft?

This question required students to calculate the mean and standard deviation of a wife given that of a husband, using the conditional mean and variance equations of a joint distribution. The students could further use these values to calculate the probability of a joint distribution of a variable from standard normal distribution table. The expected solution to the question is as follows:

The conditional mean of a variable given another variable for a joint distribution is given by;

Mean, $E(x/Y=y) = \mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)$ and;

The conditional variance of a variable given another variable for a joint distribution is given by;

Variance, $Var(x/Y=y) = \sigma_x^2(1-\rho^2)$

The standard deviation is given by the square root of variance.

$$\text{Standard deviation} = \sqrt{{\delta_2}^2(1-\rho^2)}$$

$$Therefore, the\ mean, E(x/Y=y) = \mu_2 + \rho\frac{\delta_2}{\delta_1}(x-\mu_1)$$

$$substituting\ the\ values, \mu_2 = 5.3, \rho = 0.6, \delta_1 = \delta_2 = 0.2, \mu_1 = 5.8\ and\ x = 6.3:$$

$$E(x/Y=y) = 5.3 + 0.6\left[\frac{0.2}{0.2}\right](6.3-5.8)$$

$$= 5.3 + 0.6(0.5)$$

$$= 5.3 + 0.3$$

$$= 5.6ft$$

$$And\ substituting\ \rho = 0.6, \delta_2 = 0.2, in\ the\ \ equation, \delta = \sqrt{{\delta_2}^2(1-\rho^2)} leads\ to;$$

$$\delta = \sqrt{0.2^2 \times (1 - 0.6^2)}$$

$$\delta = 0.2 \times \sqrt{(1 - 0.6^2)} = 0.2 \times \sqrt{1 - 0.36}$$

$$\delta = 0.2 \times \sqrt{0.64} = 0.2 \times 0.8$$

$$\delta = 0.16$$

Finding the conditional probability that the height of the wife lies between the limits given when the height of the husband is 6.3 ft results to;

$$P(5.28 \leq Y \leq 5.92\ /x = 6.3) = \emptyset\left(\frac{5.92-5.6}{0.16}\right) - \emptyset\left(\frac{5.28-5.6}{0.16}\right)$$

$$= \emptyset(2) - \emptyset(-2)$$

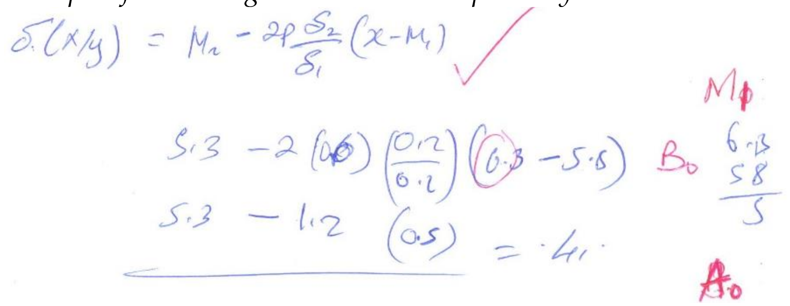
And hence, $P(5.28 \leq Y \leq 5.92\ /x = 6.3)$ = 1.954 – 1

= 0.954

This question was solved correctly by one student, 3 students answered it partially correct, 19 students solved it wrongly and 152 students did not attempt the question. A sampled students' work in figure 5 shows correct substitution of values in a wrong formula $E(x/Y) = \mu_x - 2\rho\frac{\sigma_x}{\sigma_y}(y - \mu_y)$ instead of using $E(x/Y) = \mu_x + \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y)$. Thus the student failed to calculate the correct mean, variance, standard deviation, and probability of the joint distribution.

Figure 5

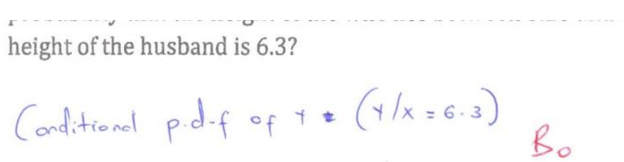
Example of an undergraduate students' partially correct solution to question 3



Another student (Figure 6) was not able to state the conditional mean and variance of a variable given another variable. This means that the student could not proceed with the solution of the question thus the student was not proficient in solving the BND question.

Figure 6

Example of a students' wrong solution to question 3



The students (Figures 5 and 6) had a poor grasp of the conditional mean and variance equations of the joint distribution of a bivariate normal distribution. The student (Figure 5) could not state the correct equations to use. The student used a wrong sign (minus) instead of a plus in the formula for calculating the conditional mean of one variable given another. The student (Figure 6) had difficulties in using the conditional mean and variance equations in solving this question. Similar errors were observed from majority of the other students who were unable to state and use the conditional mean and variance equations in solving the question. This implies that the students lacked conceptual understanding, procedural fluency and adaptive reasoning proficiency in solving BND problems.

4.4. Calculation of the Mean and Standard Deviation of two random Variables given a Bivariate Random Density Function

The undergraduate students were asked the following question to explore their proficiency in solving the mean and standard deviation of two random variables of a joint distribution given a bivariate random density function.

Question 4

Given that the bivariate random density function of y is:

$$f(x,y) = \frac{1}{2\pi\delta_1\delta_2\sqrt{1-\rho^2}} e^{-\frac{1}{1.02}\{(x+2)^2-2.8(x+2)(y-1)+4(y-1)^2\}}$$

Find the values of:

a) μ_1 and μ_2, the expected values of X and Y respectively,
b) δ_1 and δ_2, the standard deviations of random variables X and Y respectively.

The students were expected to equate the general bivariate normal distribution function, to the given bivariate random density function. By equating the corresponding, coefficients, the expected values, μ_1 and μ_2, and the standard deviations, δ_1 and δ_2 the two random variables can be obtained. The students' expected solution is as illustrated below;

$$\frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)-\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]}$$
$$= \frac{1}{2\pi\delta_1\delta_2\sqrt{1-\rho^2}} e^{-\frac{1}{1.02}\{(x+2)^2-2.8(x+2)(y-1)+4(y-1)^2\}}$$

a) The exponential function of the general bivariate normal distribution equation is equated to the corresponding exponential function of the given bivariate random density function to solve for the mean and variance of the two random variables. This results to:

$$e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)-\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]} = e^{-\frac{1}{1.02}\{(x+2)^2-2.8(x+2)(y-1)+4(y-1)^2\}}$$

$$(x+2)^2 = \frac{(x-\mu_1)^2}{\sigma_1^2}$$

Finding the square root on both sides leads to;

$x + 2 = x - \mu_1$,

Hence the mean, $\mu_1 = -2$

Similarly, $(y-1)^2 = \frac{(y-\mu_2)^2}{\sigma_2^2}$,

Meaning $y - 1 = y - \mu_2$

And the mean, $\mu_2 = 1$

b) $e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2-2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)-\left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]} = e^{-\frac{1}{1.02}\{(x+2)^2-2.8(x+2)(y-1)+4(y-1)^2\}}$

c) $\frac{(x-\mu_1)^2}{\sigma_1^2} = (x+2)^2$

When the coefficients of a variable X are equated, it results to;

$$\frac{1}{\sigma_1^2} = 1, \text{Hence}, \sigma_1^2 = 1;$$
$$\text{and } \sigma_1 = 1$$

$$Also,\ \frac{(y-\mu_2)^2}{\sigma_2^2} = 4(y-1)^2,$$

The coefficients of the variable Y are equated and lead to;

this implies that $\frac{1}{\sigma_2^2} = 4$ And therefore, $\sigma_2^2 = \frac{1}{4}$;

$$\text{Hence, } \sigma_2 = \frac{1}{2} = 0.5$$

In this question, 7 students answered question (a) correctly, 12 students solved it wrongly and 156 students did not attempt the question. Figure 7 shows a sampled student's wrong solution of the question. The student expanded the sub equations having the X and Y variables, of the BND exponent function.

Figure 7

Example of a students' wrong solution to question 4a

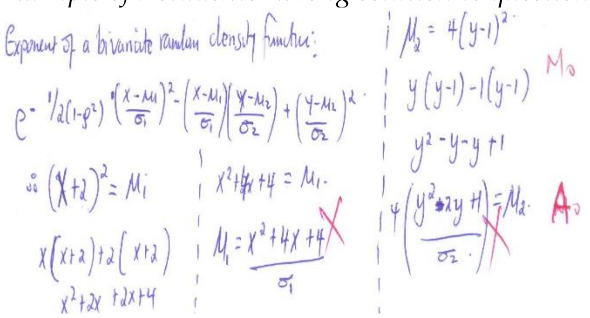
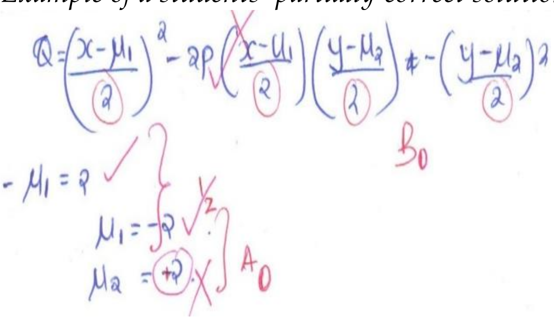


Figure 8 shows another student who correctly equated the exponential functions, but solved only the mean, μ_1, correctly. The error committed by the student was using a wrong exponential function with a denominator of 2 instead of δ_1 and δ_2, for the two random variables.

Figure 8

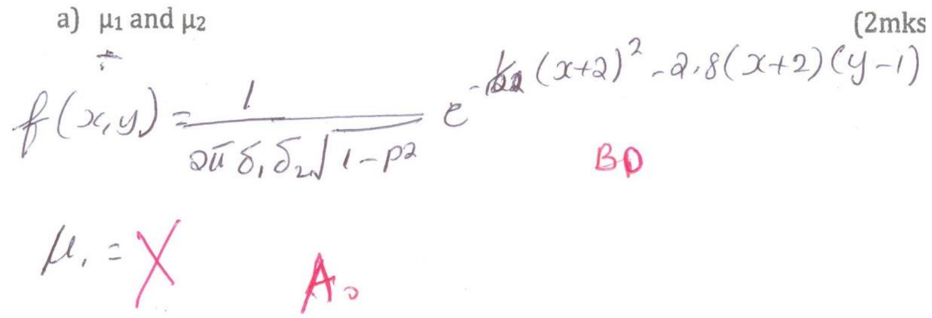
Example of a students' partially correct solution to question 4a



Yet another student (Figure 9) could only state the bivariate normal density function as given in the question but could not proceed to compare the coefficients and variables to the general bivariate normal distribution equation in order to obtain the mean and variance of the two random variables. Many other students committed similar errors.

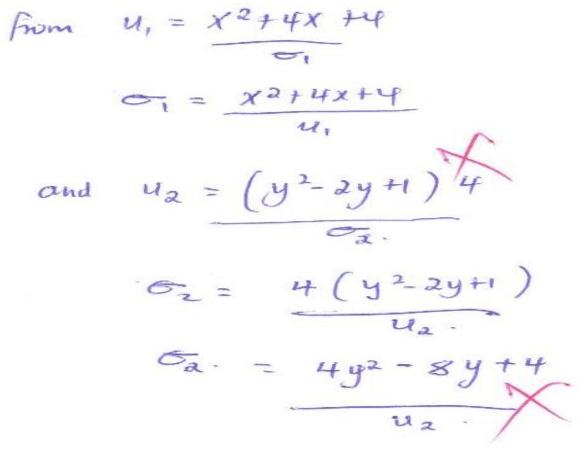
Figure 9

Example of a students' wrong solution to question 4a



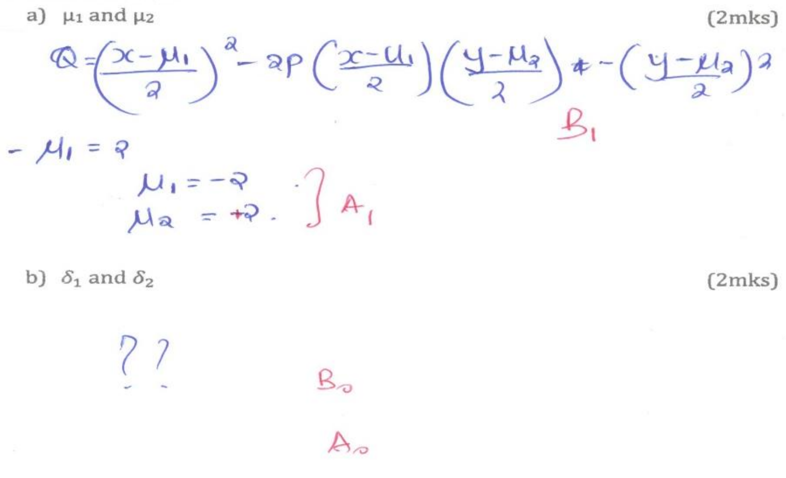
As shown in figure 10, one student wrongly attempted to solve the question by expanding exponential sub equations for the given bivariate random density function instead of equating the corresponding coefficients and variables.

Figure 10
Example of a students' solution to question 4b



Another student (Figure 11) solved question 4b correctly but could not solve question 4b. The student only wrote question marks on the working space illustrating the difficulties experienced in solving the question. The student was not proficient in comparing the coefficients and variables of the general BND and the given density function to obtain the standard deviation of one variable given another variable to solve the BND question. Similar errors were observed among other undergraduate students who did the test.

Figure 11
Example of a students' solution to question 4b



The undergraduate students' underwhelming performance on the BND problems demonstrates their poor grasp of BND concepts. Furthermore, the dismal performance shows the difficulties these students encountered in interpreting BND function and solving BND problems. Thus, a lack of strategic competence, conceptual understanding, procedural fluency, and adaptive reasoning in solving BND problems was manifested among undergraduate students.

Table 1
A summary of the students' proficiencies in solving the BND questions (n=175)

Question No.	*No. of Completely correct solutions*	*No. of partially correct solution*	*No. of Completely incorrect solutions*	*No. not attempted*
Q-1	51	47	28	49
Q-2	0	1	3	171
Q-3a	1	3	19	152
Q-3b	1	0	7	167
Q-4a	7	0	12	156
Q-4b	0	0	6	169

5. Discussion

The results of the study showed that the undergraduate students were not proficient in calculating the probability of a normal distribution variable, the mean for a normally distributed variable, the conditional mean, variance, standard deviation, and probability of a joint distribution, and the mean and standard deviation of two random variables of a BND given a bivariate random density function.

The study result also showed that most of the students were not able to calculate the probability of a normal distribution variable when given the mean and variance of the variable, except for the 51 students (29.14%) who solved the question correctly. The students were also not proficient in computing the mean of a normal distribution function when given the standard deviation and the probability of a variable. These demonstrated the students' lack of foundational knowledge of the normal distribution. Batanero, Tauber, and Sanchez (2004) observed students had difficulties in interpreting probabilities under the normal curve and solving different problems involving the normal distribution. The normal distribution and the marginal density of a normal distribution form part of the basic knowledge for effective learning of the bivariate normal distribution. The formula for the normal probability density function looks fairly complicated, and to use it, one needs to know the population mean and standard deviation (Bhandari, 2021).

The findings further showed that the majority of the students were not proficient in solving BND problems as demonstrated by their inabilities to calculate the conditional mean, variance, standard deviation, and probability of a joint distribution. The students had difficulties in extracting the values of the parameters given, to compute the mean and variance of a variable given another variable of a joint distribution. This study agrees with Koparan (2015) who indicated that students encounter difficulties in understanding and evaluating variance concepts.

Further, majority of the students were not proficient in solving BND problems as revealed by their failure to equate and compare the general BND function to the given bivariate random density function, to obtain the expected value of X (μ_1), the expected value of Y (μ_2), the standard deviation of a random variable X (δ_1) and the standard deviation of a random variable Y (δ_2) for the two dependent random variables of a joint distribution. This is a clear attestation of students' difficulties in understanding BND concepts, leading to incompetence in solving BND problems in the Kenyan university. The difficulties and incompetence manifested in solving BND problems was a common problem among the students. The findings of this study are parallel with the findings of Shojaie, Aminghafari & Mahammadpour (2012) that students experience difficulties in finding the joint distribution of a function and calculating the bivariate expectation. The BND function may be long and complicated for some students to comprehend and consequently, the students could not relate the general BND equation to the given bivariate random density function for them to compute the mean and the standard deviation of the two random variables of a joint distribution.

As evidenced by the results of this study, majority of the students were not proficient in solving the normal distribution questions. Consequently, the undergraduate students were also not proficient in solving bivariate normal distribution problems. The students' incompetence in solving BND questions can be attributed to their lack of adequate foundational knowledge of a normal distribution. It implies that students had a disconnection between the bivariate normal distribution and the normal distribution concepts which consequently led to students' lack of proficiency in solving BND problems.

6. Conclusion

The study investigated undergraduate students' proficiencies in solving BND problems in a Kenyan university. It was found that the students were not proficient in solving BND problems with regards to calculating: (i) the probability of a normal distribution given the mean and variance of a variable, (ii) the mean of a normal distribution given variance and probability of a variable, (iii) the mean and variance of a joint distribution and hence the probability of the variable given the conditional distribution of a variable, and (iv) the mean and standard deviation of two random variables given a bivariate random density function.

7. Recommendations

To improve undergraduate students' learning of BND and their proficiency in solving BND problems, the authors suggest that lecturers should embrace elaborate revision which anchors on the mean and probability of a normal distribution, and the normal distribution density function before introducing or teaching BND in probability and statistics. This entails learning from known basic statistical concepts to unknown advanced concepts in BND. This enhances students' adequate grasp of fundamental statistical concepts that would help them to meaningfully learn BND. Students' lack of understanding of basic statistical concepts, results to difficulties in learning advanced concepts, which consequently lead to incompetence in solving questions on the advanced concepts. To remedy students' difficulties in solving statistics problems, we recommend that lecturers guide the students in critical analyses of statistical equations, for the students to conceptually understand the parameters of the statistical equations. Furthermore, students should be exposed to more worked examples on the normal distribution and BND which could essentially improve their proficiency in solving BND problems.

In addition, it is recommended that the Kenyan university should consider teaching BND as an introductory course under the multivariate distributions unit, in the third year of the undergraduate study, thereby enabling the students to meaningfully contextualize the knowledge.

Future studies should look to address the missing link between foundational concepts and the BND to foster undergraduate students' proficiency and excellence on BND.

Acknowledgements. This paper is the extended version of a contribution to the second International Conference of Pedagogical Research (ICOPR) which was held on May 15-16, 2021 with the hostage of Universiti Utara Malaysia.

Funding: No funding source is reported for this study.

Declaration of interest: No conflict of interest is declared by authors.

References

Arum D. P., Kusmayadi T. A., & Pramudya I. (2018). Students' difficulties in probabilistic problem-solving. *Journal of Physics. Conference series.* IOP publishing. https://doi.org/10.1088/1742-6596/983/1/012098

Batanero C., Tauber L. M., & Sanchez V. (2004). Students' reasoning about the normal distribution. In Ben-Zvi D., Garfield J. (Eds). *The challenge of developing statistical literacy, reasoning and thinking* (pp. 257-276). Springer. https://doi.org/10.1007/1-4020-2278-6_11

Bengtsson, M. (2016). How to plan and perform a qualitative study using content analysis. *NursingPlus Open*, *2*, 8-14. https://doi.org/10.1016/j.npls.2016.01.001

Bhandari, P. (2021). Understanding normal distributions. Retrived May 11, 2021, from https://www.scribbr.com/statistics/normal-distribution/

Cain, M. K., Zhang, Z., & Yuan, K. H. (2017). Univariate and multivariate skewness and kurtosis for measuring nonnormality: Prevalence, influence and estimation. *Behavior Research Methods, 49*, 1716–1735. https://doi.org/10.3758/s13428-016-0814-1

Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches*. (4th ed.). Sage.

Flury, B. (2013). A first Course in Multivariate Statistics. Springer texts in Statistics. Springer Science+Business Media, LLC

Garfield, J. B, & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York, NY: Springer.

Gauvrit, N., & Morsanyi, K. (2014). The equiprobability bias from a mathematical and psychological perspective. *Advances in cognitive psychology, 10*(4), 119–130. https://doi.org/10.5709/acp-0163-9

Gloves S., (2012). Developing mathematical proficiency. *Journal of Science and Mathematics Education. 35*(2), 119-145. Retrieved on 18th November, 2021, from, http://www.recsam.edu.my/sub_JSMESEA/images/journals/YEAR2012/dec2012vol-2/119-145.pdf

Gordon, S. (2006). *The normal distribution*. Mathematics learning centre, University of Sydney. Retrieved May 16, 2021, from https://www.sydney.edu.au/content/dam/students /documents/mathematics-learning-centre/normal-distribution.pdf

Grover,G., Sabharwal, A. & Mittal, J. (2014). Application of Multivariate and Bivariate Normal Distributions to Estimate Duration of Diabetes, *International Journal of Statistics and Applications, 4* (1), 46-57. doi: 10.5923/j.statistics.20140401.05

Harradine, A., Batanero, C., & Rossman, A. (2011). Students and teachers' knowledge of sampling and inference. In *Teaching Statistics in school mathematics-challenges for teaching and teacher education.* Springer. Dordrecht. 235-246

Kachapova, F., & Kachapov, I. (2012). Students' misconceptions about random variables. *International Journal of Mathematical Education, 43 (7),* 963-971. https://doi.org/10.1080/0020739X.2011.644332

Khazanov, L. & Prado, L. (2010). Correcting students' misconceptions about probability in an introductory college statistics course: *ALM International Journal, 5*(1), 23-35.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). Adding it up: Helping children learn Mathematics. *National Research Council. ISBN: 0-309-50524-0*, 480. Retrieved on 16th November, 2021, from, http://www.nap.edu/catalog/9822.html

Koparan, T. (2015). Difficulties in learning and teaching statistics: Teacher views. *International Journal of Mathematical Education in Science and Technology, 46*(1), 94-104. https://doi.org/10.1080/0020739X.2014.941425

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology* (4th ed.). California: Sage.

Lin, G.D., Dou, X., Kuriki, S. & Huang, J. (2014). Recent developments on the construction of bivariate distributions with fixed marginals. *Journal of Statistical Distributions and Applications, 1*(14), 1-23. https://doi.org/10.1186/2195-5832-1-14

Lugo-Armenta, J. G., & Pino-Fan, L. R. (2021). Inferential Statistical Reasoning of Math Teachers: Experiences in Virtual Contexts Generated by the COVID-19 Pandemic. *Education Sciences, 11*(7), 363. https://doi.org/10.3390/educsci11070363

McLeod, S. A. (2019). What is a normal distribution in statistics? Retrieved September 16, 2019, from https://www.simplypsychology.org/normal-distribution.html.

Memnun D. S., Ozbilen O., & Dinc E. (2019). A qualitative research on the difficulties and failures about probability concepts of high school students. *Journal of Educational Issues, 5*(1), 1-19. https://doi.org/10.5296/jei.v5i1/14146

Rouaud M. (2017). Probability, Statistics and Estimation: Propagation of uncertainties in experimental measurement. Retrieved April 28, 2021, from http://www.incertitudes.fr/book.pdf

Schoenfeld, A. H., & Kilpatrick, J. (2008). Towards a theory of proficiency in teaching mathematics. In D. Tirosh & T. Wood (Eds.), *The international handbook of mathematics teacher education: Tools and processes in mathematics teacher education. 2*, 321–354. Rotterdam, The Netherlands: Sense Publishers.

Shojaie, S. R. H., Aminghafari, M., & Mahammadpour, A. (2012). On the expected absolute value of a bivariate normal distribution. *Journal of Statistics Theory and Applications, 11*(4), 371-377.

Sotos, A.E.; Vanhoof, S., Van den Noortgate, W. & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review, 2*(2), 98-113. https://doi.org/10.1016/j.edurev.2007.04.001

Tacq, J. (2010). *Multivariate normal distributions.* Elseiver.

Weisstein, E. W. (2002). Bivariate normal distribution. From *MathWorld-- A wolfram web resource.* Retrieved October 25, 2019, from, http://mathworld.wolfram.com/BivariateNormalDistribution.html