

PROJECT 4

MATH 5900 – ADVANCED DATA ANALYSIS

April 29, 2024

TAIWO JEGEDE

E00755963

POWER AND SAMPLE SIZE

PURPOSE OF THE ANALYSIS

The purpose of this analysis is to perform a power analysis and determine the appropriate sample size for a study comparing the effectiveness of a flipped classroom approach to teaching evidence-based nursing (EBN) compared to a traditional classroom approach. The key research question is whether the flipped classroom approach leads to improved knowledge and self-efficacy in EBN practices compared to the traditional approach.

DESCRIPTION OF THE DATA

The data used in this analysis comes from a quasi-experimental study published in PLOS ONE. The study involved an experimental group that received the flipped classroom intervention and a control group that received the traditional classroom instruction. Both groups were assessed on their knowledge and self-efficacy in EBN practices at three time points: pre-test, post-test, and one-month follow-up.

The key variables in the data are:

- Group (experimental vs. control)
- Time (pre-test, post-test, follow-up)
- Knowledge score
- Self-efficacy score

The sample sizes for the two groups were 75 and 76 participants, respectively. The means and standard deviations of the knowledge scores at each time point were provided, which allows for the calculation of effect sizes.

PROPOSED ANALYSIS

For this analysis, we will focus on the knowledge scores and use a 2-sample t-test to compare the experimental and control groups. The primary statistical method used in this analysis is the two-sample t-test. This test is used to determine if two population means are equal. The specific steps in the analysis include:

- Calculation of the pooled standard deviation, which is a weighted average of the standard deviations of the two groups.
- Calculation of the effect size (d), which is the difference between the two means divided by the pooled standard deviation.
- Use of the `pwr.t.test` function in R to calculate the required sample size for a range of power values (0.70, 0.75, 0.80, 0.85, 0.90, 0.95) and significance levels (0.05 and 0.01).

$$\text{pooled_standard_deviation} = \sqrt{((75 - 1) * 18.55^2 + (76 - 1) * 18.85^2) / (75 + 76 - 2)} = 27.72791$$

Calculating the effect size (Cohen's d) using the pre-test and post-test means

$$d = (59.21 - 65.33) / 18.70 = -0.2207162$$

After running the code to calculate the power of a paired t-test, we got a power=0.4708306: This is the calculated power of the test, which is 47.08%. In other words, if we conduct this paired t-test with these parameters, there is a 47.08% chance of detecting a true difference between the groups, assuming it actually exists. This is a very low power when we consider all things, so we will need to increase the sample size, but then how do we know the perfect sample size to use and at what power. So, we use a for loop, where the alpha will be 0.05 and 0.01 and for different power, we then get the sample sizes for these.

ANALYSIS RESULT

We conducted a power analysis to determine the required sample size to detect this effect size with 80% power and a significance level of 0.05 using a 2-sample t-test. This can be done using the `pwr.t.test()` function in R

```
pwr.t.test(n=sample_s1, d=d, sig.level=0.05, type="paired", alternative="two.sided")
```

We then repeated the power analysis for different power levels (0.70, 0.75, 0.80, 0.85, 0.90, 0.95) and significance levels (0.05, 0.01) to explore the tradeoffs between sample size, power, and significance

The power analysis results are as follows:

For a significance level of 0.05:

- For 70% power, the required sample size is 129 per group.
- For 75% power, the required sample size is 144 per group.
- For 80% power, the required sample size is 163 per group.
- For 85% power, the required sample size is 186 per group.
- For 90% power, the required sample size is 218 per group.
- For 95% power, the required sample size is 269 per group.

For a significance level of 0.01:

- For 70% power, the required sample size is 201 per group.
- For 75% power, the required sample size is 220 per group.
- For 80% power, the required sample size is 243 per group.
- For 85% power, the required sample size is 271 per group.

- For 90% power, the required sample size is 309 per group.
- For 95% power, the required sample size is 369 per group.

These results indicate that to achieve 80% power to detect the observed effect size of 0.22 with a significance level of 0.05, a sample size of 163 per group would be required. Increasing the power or decreasing the significance level would require larger sample sizes.

We also plotted a power vs sample size plot to help us even understand the result better. From the plot, I also observed that, for both significance levels, as the sample size increases, the power of the test also increases and this is because larger samples provide more information, making it easier to detect a true effect if one exists. And for a given power, a lower significance level (0.01) requires a larger sample size compared to a higher significance level (0.05). This is because a lower significance level means a stricter criterion for rejecting the null hypothesis, thus requiring more evidence (i.e., a larger sample size).

The plot provides specific power values for different combinations of sample sizes and significance levels. For example, for a power of 0.8 and a significance level of 0.05, the required sample size is approximately 163.

It is very important that we note that in the context of hypothesis testing, power is the ability of a test to detect an effect, given that the effect actually exists. So, a test with high power is more reliable as it reduces the likelihood of a Type II error (failing to reject the null hypothesis when it is false).

CONCLUSION

The power analysis conducted in this study provides important insights for designing a future study comparing the effectiveness of a flipped classroom approach to teaching evidence-based nursing (EBN) versus a traditional classroom approach.

The observed effect size of Cohen's $d = -0.2207162$ suggests a small to medium difference in knowledge scores between the flipped classroom and traditional classroom groups. This relatively small effect size indicates that a sufficiently large sample size will be required to detect a statistically significant difference. So, to achieve 80% power to detect the observed effect size of -0.2207162 with a significance level of 0.05, a sample size of 163 participants per group would be required. Increasing the desired power or decreasing the significance level would necessitate even larger sample sizes.

Also, our analysis explored the sample size requirements across a range of power levels (70% to 95%) and significance levels (0.05 and 0.01). This highlights the important tradeoffs that researchers must consider when designing the study. Higher power and more stringent significance levels require larger sample sizes, which may not always be feasible.

Based on these findings, i can recommend that we should aim for a sample size of at least 163 participants per group (experimental and control) to achieve 80% power to detect the observed effect size at a significance level of 0.05, and also we should consider increasing the sample size further to 186 per group to achieve 85% power, which would provide more robust statistical evidence.

When we do this, the research team can make informed decisions about the appropriate study design and sample size requirements to effectively evaluate the impact of the flipped classroom approach on EBN knowledge and self-efficacy.

REFERENCE

Chu, T.-L., Wang, J., Monrouxe, L., Sung, Y.-C., Kuo, C., Ho, L.-H., & Lin, Y.-E. (2019). The effects of the flipped classroom in teaching evidence based nursing: A quasi-experimental study. PLOS ONE, 14(1), e0210606. <https://doi.org/10.1371/journal.pone.0210606>

APPENDIX

```
library('pwr')
```

```
library('ggplot2')
```

```
# Load required packages
```

```
install.packages("pwr")
```

```
library(pwr)
```

```
m1 = 65.33
```

```
m2 = 59.21
```

```
sd1 = 18.55
```

```
sd2 = 18.85
```

```
sample_s1 = 75
```

```
sample_s2 = 76
```

```
pooled_standard_deviation = sqrt(((sample_s1 - 1) * sd1^2 + (sample_s2 - 1) * sd2^2)/  
sample_s1 + sample_s2 - 2)
```

```
pooled_standard_deviation
```

```
d = (m2 - m1) / pooled_standard_deviation
```

```
d
```

```
library(pwr)
```

```

pwr.t.test(n=sample_s1 ,d=d,sig.level=0.05,type="paired",alternative="two.sided")

power_values = c(0.70,0.75,0.80,0.85,0.90,0.95)

# Calculate the required sample sizes for alpha = 0.05

for (power in power_values) {

  n <- pwr.t.test(d = d, power = power, sig.level = 0.05, type = "paired")$n

  print(paste("For power =", power, "and alpha = 0.05, the required sample size is", round(n)))

}

# Calculate the required sample sizes for alpha = 0.01

for (power in power_levels) {

  n <- pwr.t.test(d = d, power = power, sig.level = 0.01, type = "paired")$n

  print(paste("For power =", power, "and alpha = 0.01, the required sample size is", round(n)))

}

sample_sizes_005 = c(129, 144, 163, 186, 218, 269)

sample_sizes_001 = c(201, 220, 243, 271, 309, 369)

# Create a data frame

df <- data.frame(

  Power = rep(power_levels, 2),

  SampleSize = c(sample_sizes_005, sample_sizes_001),

```

```

Alpha = rep(c("0.05", "0.01"), each = length(power_levels))

)

# Create the plot

ggplot(df, aes(x = SampleSize, y = Power, color = Alpha)) +

  geom_line() +

  labs(title= "Power vs Sample Size", x = "Sample Size", y = "Power", color = "Significance
Level") +

  theme_minimal()

```

