Introduction

The dataset reports metrics of 240 Cornerstone classes of the year 2021-2022 in Minerva University. My research question is whether classes of two courses Formal Analysis (FA) and Multimodal Communication (MC) have significantly different averages of hand raises. Prior analysis showed that FA classes have approximately twice as many hand raises as MC classes. I will further examine how this considerable difference is reflected on the population of all Cornerstone classes of all years 2015 - 2022 of Minerva University. To enhance students' experience, understanding of the hand raising metric is important to make evaluation in curriculum or professors' performance (Böheim et al). Therefore, I will conduct a difference of means test, then calculate both statistical and practical significance, and investigate the relationship between the null hypothesis and the confidence interval of that difference.

Dataset

The dataset contains metrics of 240 Cornerstone classes for the 2021-2022 academic year. Those were randomly sampled by Forum (Appendix A).

Dependent variable (y)	The average number of hand raises. This variable is counted by the total number of hand raises from all students in one class, its unit is "hand raise". Then, the average is calculated by taking the mean of the total number of hand raises across each course's 60 classes. It is a quantitative discrete variable because given a finite interval, the number of hand raises is also finite; more importantly, they are natural numbers. Without time constraint, one can count with a clear preceding and succeeding order.
Independent	The type of course.

variable	There are two subgroups (x1= "Formal Analyses" and x2 = "Multimodal Communications"). Each course is labeled and distinguished from the others by its name.This is a qualitative nominal variable because the classes were put into the “course” categories that cannot be ranked in order.

Analysis

Hypotheses

Defining hypotheses:

Hypotheses	Natural language	Statistical notation
Null hypothesis	The difference in the average of hand raises between FA classes and MC classes is equal to 0	(μ no. hand raises | FA classes) - (μ no. hand raises | MC classes) = 0
Alternative hypothesis	The difference in the average of hand raises between FA classes and MC classes is not equal to 0	(μ no. hand raises | FA classes) - (μ no. hand raises | MC classes) ≠ 0

I have no assured prediction or interest in the direction of the difference between two courses; therefore, I use a 2-tailed test.

To determine the significance level, I define type I and type II error:

- Type I error: I conclude that there is a difference when in fact there is no difference in the average of hand raises between FA and MC classes.
- Type II error: I conclude that there is no difference when in fact there is a difference in the average hand raises between FA and MC classes.

There is no substantial difference in the danger of committing these two types of error; therefore, I will set the significance level to the default $\alpha = 0{,}05$ and confidence interval of 95%, which means I only reject the null hypothesis if it has a p-value less than 0,05. I do not use Bonferroni correction because I am doing just one test on the dataset.

Summary statistics

The dataset was read by Python by the pandas package. See Appendix A and B for general and detailed summary statistics.

Table 1: Summary statistics for the number of hand raises (hand raises per class) for two sample groups: FA classes and MC classes		
	FA classes (y | x1)	MC classes (y | x2)
Count	$n1 = 60$	$n2 = 60$
Mean	$\bar{x}1 = 110.05$	$\bar{x}2 = 64$
Median	103	55
Mode	83, 141	37, 55
Standard Deviation	$s1 = 48.52$	$s2 = 35.77$
Range	199	170

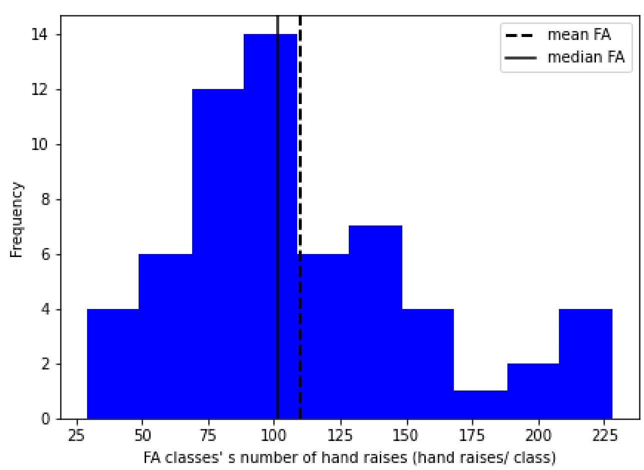


Figure 1. Sample distribution of FA classes

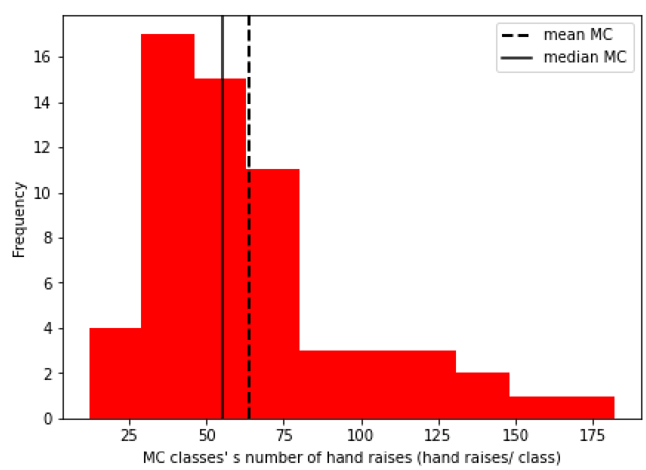


Figure 2. Sample distribution of MC classes

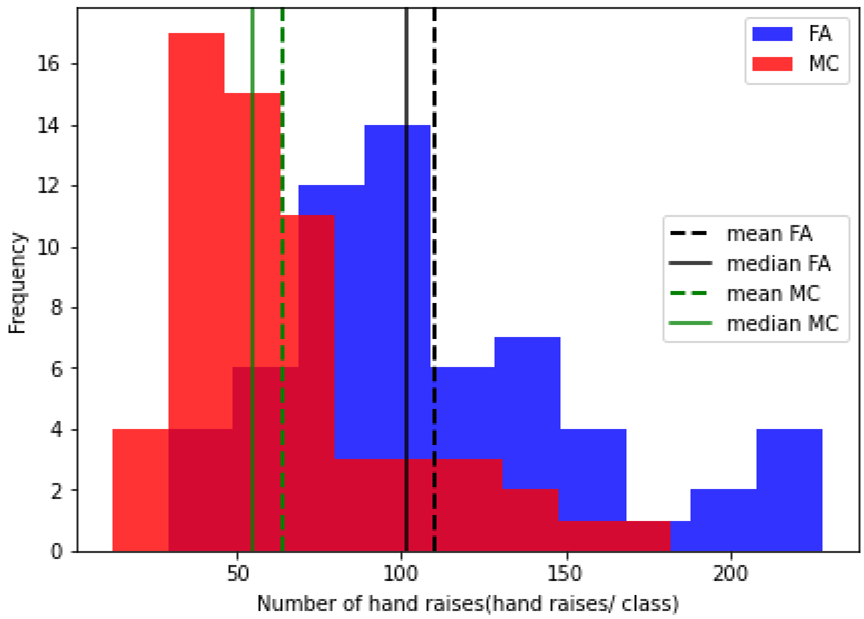


Figure 3. Sample distributions of both FA and MC classes

It is shown that (y|x1) is larger than (y|x2) because the mean, median and mode of x1 is higher than that of x2. Secondly, both sample distributions are skewed to the right because their medians are both smaller than their means; hence, there are outliers on the right. However, x2's histogram with a lower sample variance and higher standard deviation, is more skewed than x1.

Conditions for Inference

Since the population standard deviation is unknown, I use the t-distribution as the sampling distribution and calculate the standard error for inference of corresponding unknown population parameter. Certain conditions of the Central Limit Theorem must be met:

(1) *A random sample:* Since the metrics are randomly pulled from Forum, this condition is met.

(2) *The sampling distribution is normal:* The sample sizes of both subgroups are 60, which is larger than the condition of 30. Even though the sample distributions are

skewed, I assumed that the population distribution will not be too highly skewed to diminish the effect of the large sample size.

(3) *Two groups need to be independent of each other:* For sampling without replacement, I estimate that the population size is 6400 in Appendix C. Since the sample size of both subgroups is 60, which is smaller than 5% of the population (320), this condition is met since removing each observation doesn't change the probability of the next draw in a finite population. Therefore, I do not use the Finite Population Correction Factor (FPC) to avoid overestimating the standard error.

(4) *The number of hand raises between classes of each subgroup is independent of each other.* The number of hand raises in one class does not affect the number of hand raises in other classes when picked at random.

Difference of Means Test (Appendix D)

To assess statistical significance I compute the T-score = 5.92, standard error = 7.78 and degrees of freedom df = 59. The T-score of 5.92 means that the averages of hand raises of FA classes and MC classes are 5.92 standard deviation away from each other. It results in a two-tailed p-value of approximately 0 < 0.05. This p-value represents the probability of getting a sample this extreme or more given the null is true. That means the probability of the observed difference as an area under the sampling distribution lies in the heavily extreme small tails. Therefore, the difference of means rejects the null hypothesis, and the observation in the sample distribution is less likely to happen by chance or sampling variance. The result is statistically significant.

Assessing practical significance, I measure effect size by Cohen's d formula because it allows comparisons between studies with different variables. The result d = 1.08 indicates that the difference is very large and meaningful in real life. This is aligned with the extremely small p-value. Overall, the difference in average of hand raises between FA classes and MC classes is statistically and practically large.

Confidence Intervals of difference of means (Appendix E)

To further examine what range of values the difference could take as a plausible estimation of population parameters, I constructed a confidence interval for the difference of means: (y|x1) - (y|x2) = [30.67, 61.49]. Notably, the null hypothesis is outside of the confidence interval, which might indicate a statistical significance. This provides more evidence for the prior hypothesis testing with p-value; also, the direction of the difference towards the null and the size of difference. In the long run, if more samples and more difference confidence intervals are conducted, 95% of those will capture the population mean.

Statistical power

What is the probability of rejecting a null hypothesis given that the alternative difference of means M = 46 (point estimate) is true? (Appendix F)

Following the power function, I calculate the observed sample mean (X1 = -15.12 and X2 = 15.12) and the Z statistic (Z1 = -7.93 and Z2 = -4.01) when rejecting the null hypothesis. The sum of two areas under two tails that are outside of the Type II error zone is the power. The result of approximately 99% shows a very high probability that the test correctly rejects the null hypothesis when a specific alternative hypothesis is true. However, a 99% power means my test is highly sensitive to true effects, including very small ones. This may lead to finding statistically significant results with very little usefulness in the real world, and an increase in Type I error.

What is the minimum number of participants needed to detect an effect size of levels 1.08 at a 5% significance level and a 90% power? (Appendix F)

Based on the target distance between the center of the null and alternative distributions in terms of the standard error, the desired sample size is subtracted from the equation. In conclusion, a sample size of 36 would yield a 90% power with expected meaningful significance in result.

Results and Conclusions

Result (p = 0, d = 1.08, [30.67, 61.49]) suggests that FA classes are 95% likely to have a significantly larger number of hand raises than MC classes in Minerva across all academic years. The Minerva academic team should further investigate the cause of this gap.

These conclusions are a generalization induction because I made inference about the population as statistically analogous from one sample dataset given. Yet they are strong because the test has a large statistical power. The larger the power, the stronger the induction. In this test, a medium-large sample size (n = 60) will make the standard error smaller, a small p-value (p = 0) would increase the probability of rejecting the null, and a big effect size (d = 1.08) would create a large distance in two distributions of difference of means (Null is true and Alternative is true). All of which decreases the probability of having a Type II error. Given much evidence and satisfied conditions for the Central Limit Theorem, the conclusion about the population is likely to follow. However, the reliability could not be ensured due to assumptions about the skewness of the population. Moreover, one could research with a sample size of 36 to re-evaluate the accuracy of the result.

Reflection

1. For validity, given that all my premises (calculation of T-score, p-value, etc) are correct, the conclusion must be true. A sound argument also requires true premises. I would compare the consistency between results and conduct multiple calculations of a metric with different formulas or in different ways (python, online calculator…).
2. The feedback I received from my professor or TA is to effectively provide more accurate definitions or meanings of the concept. I have incorporated my own understanding and interpretations of T-store, p-value, confidence interval, power, etc.
3. Acknowledgement:

   I apply knowledge from the OpenStats textbook and multiple in-class activities (algorithms, codes…). Moreover, special thanks to Larrysa and Godson for correcting my coding mistakes in constructing the confidence interval (I used T-score instead of the t critical value).

APPENDIX

APPENDIX A: IMPORT, GENERAL DATASET AND VARIABLE INFORMATION

(1) Dataset import and display

```
#APPENDIX A: GENERAL DESCRIPTIVE STATISTIC FOR 2 VARIABLES
# import relevant packages and libraries
import pandas as pd
import matplotlib.pyplot as plt
from scipy import stats
import numpy as np

# import the data using pandas and read into a dataframe
df = pd.read_csv ("https://docs.google.com/spreadsheets/d/e/2PACX-1vRM3xx4yt1Mdy9Z5lvfqCdzyjrTCRfcgXFjm7BMQIRE9DBx1RhmROcojAMDV7N
df = df[["semester", "course", "section", "class", "num_hand_raises"]]
df.head(10)
```

	semester	course	section	class	num_hand_raises
0	Fall 2021	Formal Analyses	MW@09:00AM San Francisco	CS50 Session 7 - (4.1) Fallacy Detection	222
1	Fall 2021	Formal Analyses	TTh@05:00PM San Francisco	CS50 Session 15 - (8.2) Distributions of Discr...	139
2	Fall 2021	Formal Analyses	TTh@11:00AM San Francisco	CS50 Session 24 - (13.2) Difference of Means T...	103
3	Fall 2021	Formal Analyses	MW@07:00AM San Francisco	CS50 Session 1 - (1.1) Critical Thinking	55
4	Fall 2021	Formal Analyses	MW@07:00AM San Francisco	CS50 Session 8 - (4.2) Logic Synthesis	92
5	Fall 2021	Formal Analyses	MW@07:00AM San Francisco	CS50 Session 21 - (11.2) Confidence Intervals ...	95
6	Fall 2021	Formal Analyses	TTh@11:00AM San Francisco	CS50 Session 18 - (10.1) Sampling Distribution...	76
7	Fall 2021	Formal Analyses	MW@11:00AM San Francisco	CS50 Session 20 - (11.1) Confidence Intervals ...	146
8	Fall 2021	Formal Analyses	MW@11:00AM San Francisco	CS50 Session 12 - (7.1) Correlation	194
9	Fall 2021	Formal Analyses	TTh@09:00AM San Francisco	CS50 Session 19 - (10.2) Sampling Distribution...	79

(2) General descriptive statistics and data visualization for 2 variables

(a) Dependent variable

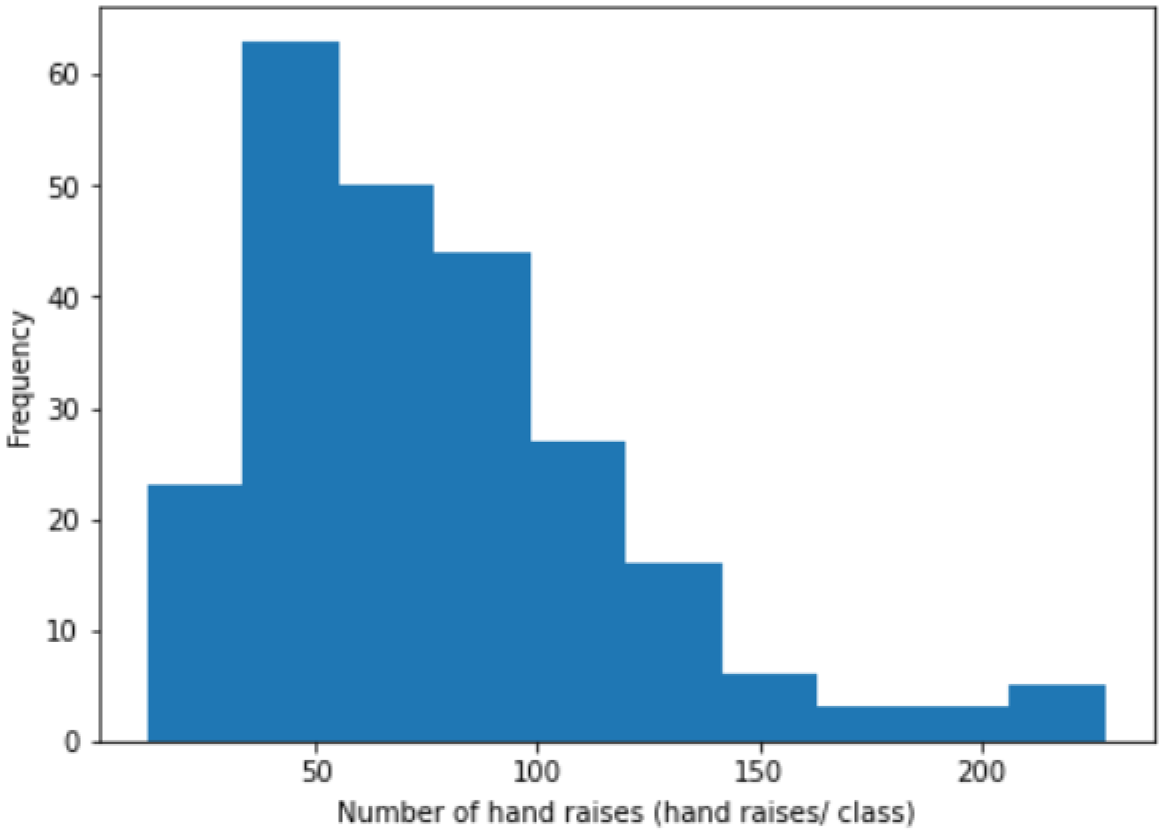
```
#APPENDIX A - DEPENDENT VARIABLE
# print the summary statistic of the dependent variable
print("The decriptive statistics for the variable Number of hand raises are: \n", df["num_hand_raises"].describe())
```

```
The decriptive statistics for the variable Number of hand raises are:
 count    240.000000
mean      77.570833
std       42.268359
min       12.000000
25%       49.000000
50%       69.000000
75%       98.250000
max      228.000000
Name: num_hand_raises, dtype: float64
```

```
#APPENDIX A - DEPENDENT VARIABLE
# print the median and mode of the dependent variable
print("Dependent variable's median = ", df["num_hand_raises"].median())
print("Dependent variable's mode = ", df["num_hand_raises"].mode())
```

```
Dependent variable's median =  69.0
Dependent variable's mode =  0    37
1    49
Name: num_hand_raises, dtype: int64
```

```
# create histogram for the dependent variable
plt.figure(figsize=(7,5))
plt.hist(df["num_hand_raises"])
plt.xlabel("Number of hand raises (hand raises/ class)")
plt.ylabel ("Frequency")
plt.show()
```

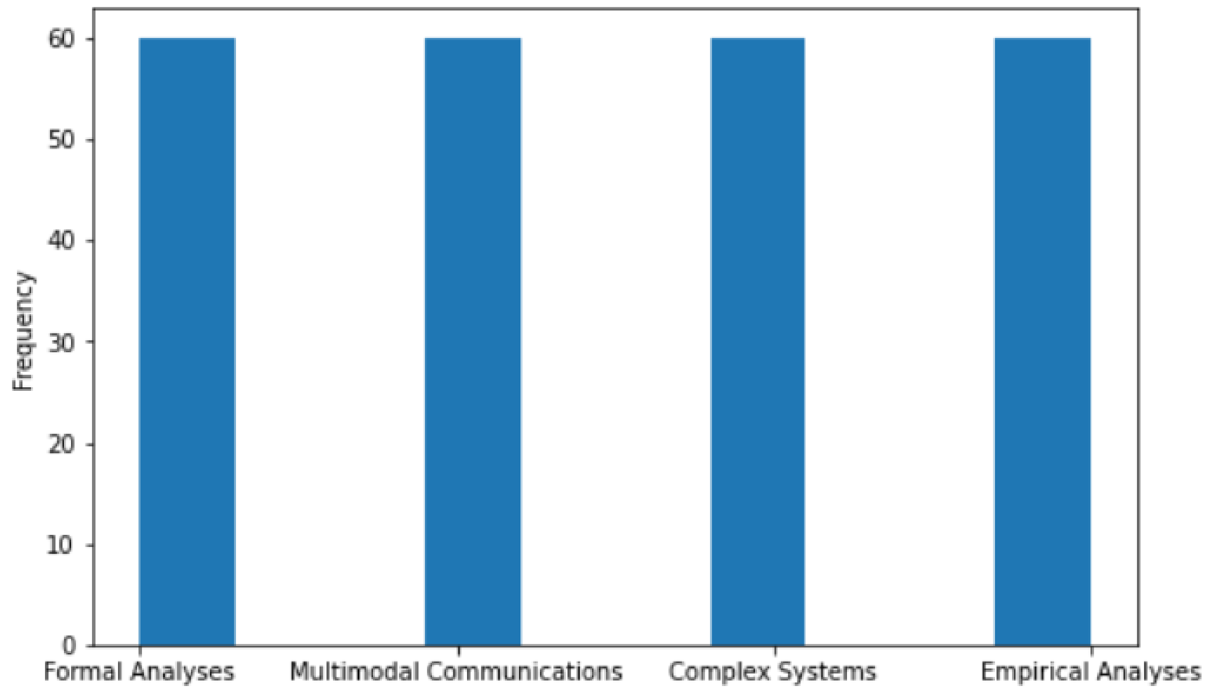


(b) Independent variable

```
#APPENDIX A - INDEPENDENT VARIABLE
# print the summary statistic of the independent variable
print("The decriptive statistics for the variable Type of course are: \n", df["course"].describe())
```

```
The decriptive statistics for the variable Type of course are:
 count                 240
unique                  4
top       Formal Analyses
freq                   60
Name: course, dtype: object
```

```
# create histogram for the independent variable
plt.figure(figsize=(8,5))
plt.hist(df["course"])
plt.ylabel ("Frequency")
plt.show()
```



APPENDIX B: DESCRIPTIVE STATISTICS FOR TWO SUBGROUPS

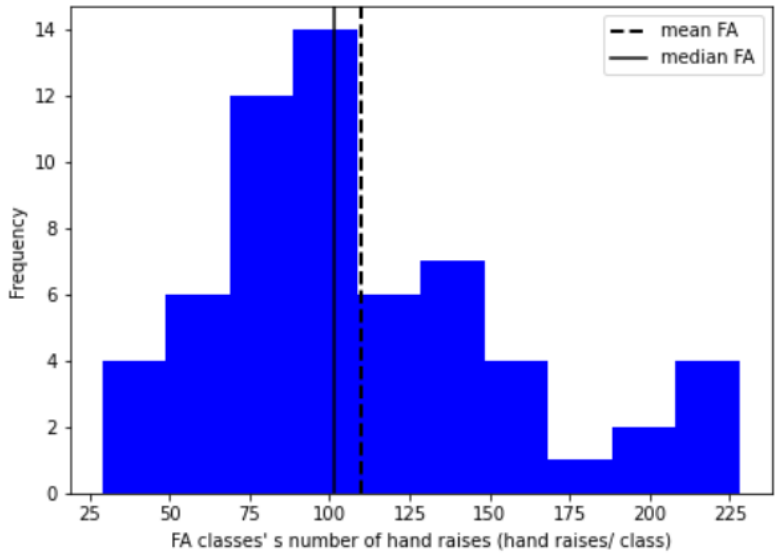
(1) Subgroup 1: FA classes

```python
# APPENDIX B: SPECIFIC DESCRIPTIVE STATISTIC FOR 2 SUBGROUPS
# Assign data to variables
FA = df.loc[df["course"] == "Formal Analyses", "num_hand_raises"]
MC = df.loc[df["course"] == "Multimodal Communications", "num_hand_raises"]

# Subgroup 1
print("The decriptive statistics for the subgroup FA are: \n", FA.describe())
```

```
The decriptive statistics for the subgroup FA are:
 count     60.00000
mean     110.05000
std       48.51644
min       29.00000
25%       77.00000
50%      101.50000
75%      139.50000
max      228.00000
Name: num_hand_raises, dtype: float64
```

```python
def histogram (data, data_name, hist_color, bin):
    plt.figure(figsize=(7,5))
    #assign parameters by histogram factors that need to be customized to various dataset
    plt.hist(data, bin, color = hist_color)
    plt.xlabel("{} classes\' s number of hand raises (hand raises/ class)". format(data_name))
    plt.ylabel('Frequency')
    # Add mean and median lines to the graph.
    plt.axvline(np.mean(data), color='k', linestyle='dashed', linewidth=2, label = "mean {}".format(data_name))
    plt.axvline(np.median(data), color='k', linestyle='solid', linewidth=1.5, label = "median {}".format(data_name))
    # format method is used to pass a string argument into another string
    plt.legend()
    print("The skewness of {} is {}". format(data_name, data.skew()))
    if data.skew() <= -1 or data.skew() >= 1:
        print("This is a strong skew")
    elif -1 < data.skew() <= -0.5 or 0.5 <= data.skew() <1:
        print("This is a moderate skew")
    else:
        print ("The data is fairly symmetrical")
    plt.show()

histogram (FA, "FA", "blue", 10)
```

```
The skewness of FA is 0.843962326834049
This is a moderate skew
```



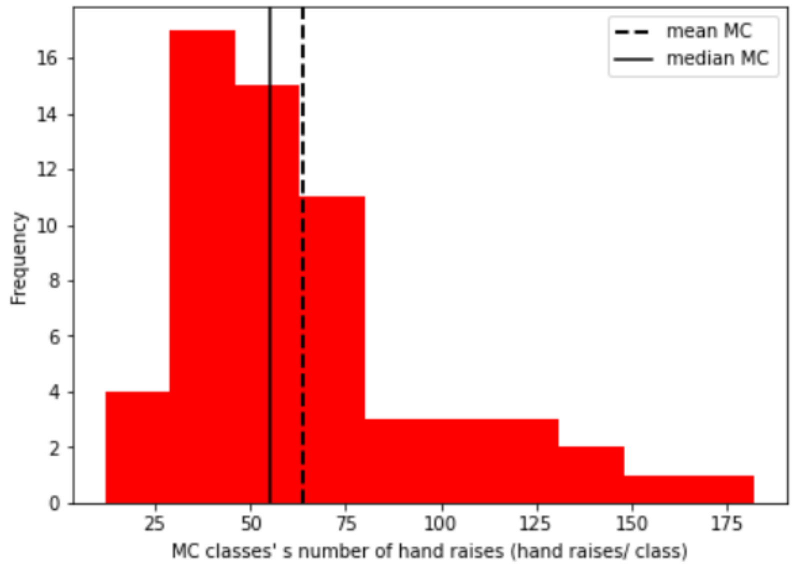
(2) Subgroup 2: MC classes

```
# Subgroup 2
print("The decriptive statistics for the subgroup MC are: \n", MC.describe())
```

```
The decriptive statistics for the subgroup MC are: 
 count     60.000000
mean      63.966667
std       35.766174
min       12.000000
25%       37.750000
50%       55.000000
75%       77.500000
max      182.000000
Name: num_hand_raises, dtype: float64
```

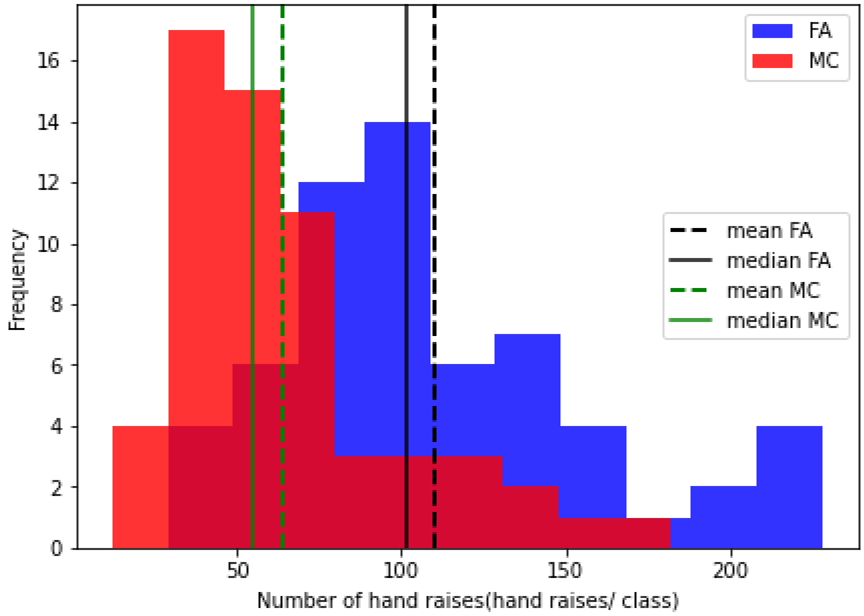
```
histogram (MC, "MC", "red", 10)
```

```
The skewness of MC is 1.20997892596191
This is a strong skew
```



(3) Sample distributions of both subgroups on the same graph

```
plt.figure(figsize=(7,5))
#plot 2 sample distribution on the same graph
plt.hist(FA, bin, color = "blue",  align = "mid", alpha = 0.8)
plt.hist(MC, bin,  color = "red", align = "mid", alpha = 0.8)

plt.xlabel('Number of hand raises(hand raises/ class)')
plt.ylabel('Frequency')
legend1= plt.legend(labels = ["FA", "MC"], loc = "upper right")
#plt.axvline is used to draw line
plt.gca().add_artist(legend1) #I added this line to create 2 lengends in the same graph
plt.axvline(np.mean(FA), color='k', linestyle='dashed', linewidth=2, label = "mean FA")
plt.axvline(np.median(FA), color='k', linestyle='solid', linewidth=1.5, label = "median FA")
plt.axvline(np.mean(MC), color='g', linestyle='dashed', linewidth=2, label = "mean MC")
plt.axvline(np.median(MC), color='g', linestyle='solid', linewidth=1.5, label = "median MC")

plt.legend (loc = "center right")
plt.show()
```



APPENDIX C: POPULATION ESTIMATE

Since the population's unit is the number of Cornerstone classes. I used the data of my first semester to generalize to a larger population.

The assumption I made is that all other years will have roughly the same number of sessions per semester. I assigned it to number 25 because that was my last class of this semester. The generalization takes into account underestimation and overestimation.

```
# Estimate population size
def population_size (a, b, c, d, e):
    population = 1
    # I used function to make the estimation more robust with different version of estimation
    total_number_of_session_per_course_per_semester = a
    number_of_courses = b
    number_of_classes_having_the_same_session_in_1_week = c
    number_of_semester = d
    number_of_all_academic_years_2015_2022 = e
    population *= a * b * c * d * e
    print ("Population size estimated = ", population)
population_size (25, 4, 4, 2, 8)
```

```
Population size estimated =  6400
```

APPENDIX D: DIFFERENCE OF MEANS TEST

```python
# APPENDIX D: DIFFERENCE OF MEANS TEST
def difference_of_means_test(data1,data2,tails):
    # the len function returns the sample size of two subgroups
    n1 = len(data1)
    n2 = len(data2)

    #the mean function returns the average of hand raises of two subgroups
    x1 = np.mean(data1)
    x2 = np.mean(data2)
    point_estimate = x2 - x1
    #point estimate is the mean of the sampling distribution, which is the same as the population mean

    #numpy calculates the sample standard deviation of two subgroups
    s1 = np.std(data1,ddof=1)
    # Bessel’s correction uses n-1 in denominator to take into account of small sample size
    s2 = np.std(data2,ddof=1)

    # calculate the standard error of the mean difference
    SE = np.sqrt(s1**2/n1 + s2**2/n2)
    # Tscore is a conversion of the difference between 2 means in a standardardize unit of standard error
    Tscore = np.abs(point_estimate)/SE #
    # the degree of freedom = chosen sample size -1
    # choose the smaller sample size between 2 subgroup
    df = min(n1,n2) - 1
    # convert Tscore into the probability as an area under the curve.
    # 2 tailed-test requires doubling the result because the probability is equally distributed between 2 tails
    pvalue = tails * (1 - stats.t.cdf(Tscore,df))
    # Calculate the effect size based on Cohen's d formula
    SDpooled = np.sqrt((s1**2*(n1-1) + s2**2*(n2-1))/ (n1 + n2 -2))
    Cohensd = point_estimate/SDpooled

    print("point estimate =", round(point_estimate, 2))
    print ("Standard error SE = ", round(SE, 2))
    print('T-score =',round(Tscore,2))
    print('P-value p =', round(pvalue,2))
    print('Effect size d =', round(Cohensd,2))
    print( s1, s2)
difference_of_means_test(MC, FA, 2) #call the function
```

```
point estimate = 46.08
Standard error SE =  7.78
T-score = 5.92
P-value p = 0.0
Effect size d = 1.08
35.76617409004698 48.516439639097975
```

APPENDIX E: CONFIDENCE INTERVAL OF DIFFERENCE OF MEANS

```
# APPENDIX E: CONFIDENCE INTERVAL OF DIFFERENCE OF MEANS
def confidence_interval(data1,data2,tails):
    # the len function returns the sample size of two subgroups
    n1 = len(data1)
    n2 = len(data2)

    #the mean function returns the average of hand raises of two subgroups
    x1 = np.mean(data1)
    x2 = np.mean(data2)
    point_estimate = x2 - x1

    #numpy calculates the sample standard deviation of two subgroups
    s1 = np.std(data1,ddof=1) # Bessel’s correction uses n-1 in denominator to take into account of small sample size
    s2 = np.std(data2,ddof=1)

    # calculate the standard error of the mean difference
    SE = np.sqrt(s1**2/n1 + s2**2/n2)

    # The difference of means test requires the degree of freedom formula of both subgroups
    df_pooled = n1 + n2 - 2
    # the t critical value is used in the formular of confidence level of difference of means test
    t_critical = stats.t.ppf(0.05/tails, df_pooled)

    # confidence interval is expressed as equal intervals from t critical to the point estimate in standardardized standard error
    Confidence_interval = [round(point_estimate + t_critical*SE,2), round(point_estimate- t_critical*SE,2)]
    print("point estimate =", round(point_estimate, 2))
    print ("t critical value t*=", round(t_critical,2))
    print ("Standard error SE = ", round(SE, 2))
    print("Confidence interval = ", Confidence_interval)
confidence_interval (MC, FA, 2)
```

```
point estimate = 46.08
t critical value t*= -1.98
Standard error SE =  7.78
Confidence interval =  [30.67, 61.49]
```

APPENDIX F: STATISTICAL POWER CALCULATION

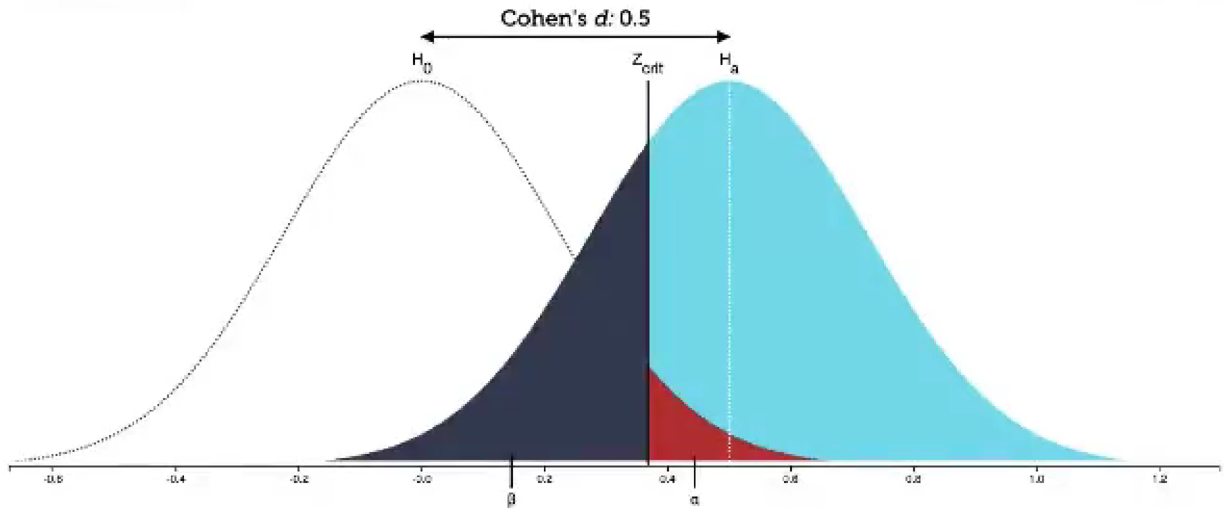
(1) Calculate the power

```
# APPENDIX F: STATISTICAL POWER CALCULATION
def power_calculation(data1,data2, null_mean):
    n1 = len(data1)
    n2 = len(data2)
    SE = np.sqrt(s1**2/n1 + s2**2/n2)
    x1 = np.mean(data1)
    x2 = np.mean(data2)
    point_estimate = x2 - x1
    df = min(n1, n2) - 1
    # I calculate the left-tailied and right-tailed area seperatly (significance level = 0.025 for each tail).
    # For significance level = 0.05, we would reject H0 if the difference is in the lower 2.5% or upper 2.5% tail:
    left_tailed_zscore = stats.norm.ppf(.025)   # = -1.96
    right_tailed_zscore = stats.norm.ppf(.975)  # = 1.96

    # I reject the null hypothesis if 1.96 <= Z <= -1.96 <=> 1.96 <= (the true population mean - null mean) /SE <= -1.96
    #The true population mean = point estimate = (+/-)1.96 * SE + null mean
    X1 = left_tailed_zscore * SE + null_mean
    X2 = right_tailed_zscore * SE + null_mean

    # Call the values that reject the null hypothesis given the alternative is true: reject value
    # Left-tail: P (Reject vallue <= X1 | alternative is true) =>  P (Z1 <= (X1 - point estimate)/SE)) => Left-tailed p-value
    Z1 = (X1 - point_estimate)/SE
    Left_tailed_percentile = stats.t.cdf(Z1, df)
    #Right-tail: P ( X >= X2 | M = point estimate) =  P (Z2 <= (X2 - point estimate)/SE)) => Right-tailed p-value
    Z2 = (X2 - point_estimate)/SE
    Right_tailed_pvalue = 1 - stats.t.cdf(Z2,df)

    # Power = 1 - β (the probability commiting type II error).
    # Type II error accepting the null hypothesis even if the alternative is true
    power = Left_tailed_percentile + Right_tailed_pvalue
    print ("X1 = {} and X2 = {}". format (round(X1,2), round(X2,2)))
    print ("Z1 = {} and Z2 = {}". format (round(Z1,2), round(Z2,2)))
    print( "Power = ", round (power,5))
power_calculation (MC, FA, 0)
```

```
X1 = -15.12 and X2 = 15.12
Z1 = -7.93 and Z2 = -4.01
Power =  0.99991
```

(2) Calculate the sample size

I start by identifying the Z-score that would give me a power of 90%, which would be the blue area in the picture below:



The result is 1.28. Additionally, the rejection region always extends 1.96 x SE from the center of the null distribution for alpha = 0.05. This allows me to calculate the target distance between the center of the null and alternative distributions in terms of the standard error:

1.28 x SE + 1.96 x SE = 3.24 x SE

I also want the distance between the null and alternative distributions' centers to equal the minimum effect size of interest,46 (point estimate), which allows me to set up an equation between this di↵erence and the standard error:

46 = 3.24 x SE

As calculated before, the formula of SE is $SE(\overline{X}_1 - \overline{X}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ As I am interesting in calculating the total number of sample size, n1 and n2 are simplified as n. Note that s1 and s2 has been calculated in APPENDIX D as 35.77 and 48.52

46 = 3.24 x sqrt ((s1 ^2 + s2 ^2)/ 0.5n)

0.5n = (3.24^2)/ 46^2 x (35.77**2 + 48.52**2)

n = (3.24^2)/ 46^2 x (35.77**2 + 48.52**2) *2

n = 36 (approximately)

```python
print("The Z-score resulting in power of 90% = {}". format(stats.norm.ppf(0.9)))
print("The sample size = {}". format((3.24**2 / 46**2) * (35.77**2 + 48.52**2) * 2))
```

```
The Z-score resulting in power of 90% = 1.2815515655446004
The sample size = 36.05383157852553
```

FOOTNOTE

#professionalism: I structure the report and the appendix well. I followed a 2-round checking of grammar mistakes, cohesion and logic.

#distributions: I justified the use of t-distribution, and analyzed every condition for the Central Limit Theorem with assumptions. I made a clear distinction between sample distribution when visualizing the histogram of two subgroups, sampling distribution in the difference of means test, confidence interval… and the population distribution when I made assumptions about the skewness and made inferences about population parameters.

#probability: I explained p-value, significance test and confidence interval in terms of probability. I stated the conditional probability when I made distinction between the hypothesis test and power calculation: the hypothesis test will be given that the null hypothesis is true while the power calculation is the opposite

#significance: I discussed practical and statistical significance. I made a connection between the consistency in the small p-value and the large effect size to emphasize the strength of the argument. I use the confidence interval to back up the difference between 2 subgroups, and emphasize on the significance.

#confidence interval: I use the confidence interval to back up for the difference between 2 subgroups when doing the testing with p-value, and emphasize the significance of the difference. I emphasize how the confidence intervals provide more information than the testing with p-value.

#dataviz: I created multiple histograms to explain the variables and the subgroups.

#descriptivestats: I use stats to explain the conclusion of the graph with interpretation of the median-mean relationship, the outlier. Use PYthon to calculate the skewness.

#induction: I discussed multiple reasons to a strong and reliable induction in the relationship with power: the large power, the effect size… and suggest a relevant sample size.

#organization: I stated the research question, the objective and the outline at the introduction. The conclusion tied back to the introduction.

#variable: I explain why variable fit their types and how it is relevant to the quesition

Reference

Böheim, Ricardo, et al. “Student Hand-Raising as an Indicator of Behavioral Engagement and Its Role in Classroom Learning.” *Contemporary Educational Psychology*, June 2020, p. 101894, 10.1016/j.cedpsych.2020.101894.

Prof Terrana and summer 2022 FA team. “Cornerstone Metrics Data By Class 2021-2022, Sample”.