This course covers:

1. Variables
2. Research Questions
3. Understanding your data and checking assumptions
4. Selecting and interpreting inferential statistics (basic)

A research process begins with an issue or problem of interest to the researcher.

Key elements in a research problem are the variables.

A variable is defined as a characteristics or attribute of the participants or situation for a given study that 1) can be measured or observed by the researcher, and 2) has different values or varies among participants or situation.

For example: gender, age, teaching style, achievement in math, interpersonal communication skills, type of intervention or treatment are variables.
If a concept has only one value in a particular study – it is not a variable.

For example: gender is not a variable if all participants in a study are female. Ethnic group is not a variable if all participants are Malays.

In quantitative research, variables are commonly divided into independent (I.V.), dependent (D.V.), and extraneous variables.

There are two types of I.V.: 1) active, and 2) attribute. It is important to distinguish between these types when we discuss the results of a study.

An active or manipulated I.V. is a variable that manipulated by the researcher or someone at least one level which is given to a group of participants, within a specified period of time during the study.

For instance, to study the effect of a new type of teaching style, on student performance.

To be considered an active I.V., the treatment should be given after the study is planned.

Randomized experimental and quasi-experimental studies have an active I.V.
Attribute or measured I.V. is an I.V. that cannot be manipulated, yet is a major focus of the study.

The values of the I.V. are preexisting attributes of the persons or their going environment that are not systematically changed during the study.

For example, level of parental education, SES, gender, age, IQ, and ethnic group.

Studies with only attribute I.V. are called non-experimental studies.

SPSS uses a variety of terms for the I.V. such as factor and grouping variable.

When we analyze data from a research study, the statistical analysis does not differentiate whether the I.V. is an active or an attribute variable.

However, there is a crucial difference in interpretation – the need to demonstrate that a given intervention or treatment causes a change in behavior or performance.

Only the approaches that have an active I.V. can provide data that allow one to infer that the I.V. caused the change or difference in the D.V.
Although non-experimental studies (those with attribute I.V.) are limited in what can be said about causation, they can lead to solid conclusions about the differences between groups and about associations between variables.

SPSS uses the term values (levels, groups, samples, or categories) to describe the several options or categories of a variable.

These values are not necessarily ordered.

For example, gender (attribute I.V.) could be diagrammed as follows:
The D.V. is an attribute or characteristic that is dependent on or influenced by the I.V. - assumed to measure or assess the effect of the I.V.

They may be called the outcome, effect, criterion or consequence variables.

D.V. must have at least two values.

For instance, test scores, leadership skills of principles, and ratings on questionnaire.

SPSS uses a number of terms for the D.V. such as Dependent list and test variable.

Research hypotheses are predictive statements about relationship between variables.

Research questions are similar to hypotheses, except that they do not entail specific predictions and are phrased in question format.

For example:
Students who take only one test per day will score better on standardized tests than will students who take two test in one day. - Hypothesis

Is there a difference in students’ scores on a standardized test if they took two tests in one day versus taking only one test on each of two days? - RQ
Morgan, Griego, and Gloeckner (2001) divide research question into three broad types: descriptive, difference, and associational.

**Type of question / hypothesis**
- Descriptive
- Difference
- Associational

**Specific purpose**
- Summarize data
- Compare groups
- Find strength of association, relate variables

**General purpose**
- Description
- Explore relationship between variables

Descriptive RQ's are to describe or summarize data, without trying to generalized to a larger population of individuals.

Difference RQ's are to compare scores (on the D.V.) of two or more different groups, each of which is composed of individuals with one of the values or levels on the I.V.

This type of question attempts to demonstrate that groups are not the same on the D.V.

Associational RQ's are to associate or relate two or more variables.
This type of question attempt to see how two or more variables covary or how one or more variable enables one to predict another variable.

Complex RQ’s (multivariate) involve more than two variables at a time.

Exercise. IV or DV?

1. Awareness of quality control affects quality outcomes.
2. Children who are blocked from reaching their goals exhibit more aggressive behavior than children not blocked.
3. Graduate students who have completed a how-to-study course will make significantly higher GPA’s than graduate students who have never taken such a course.
4. Under intangible reinforcement conditions, middle-class children will learn significantly better than lower-class children.

After the data entered into SPSS, the first step to complete is exploratory data analysis (EDA).

EDA is used to examine and get to know your data:

1. to see if there are problems in the data such as outliers, non-normal distributions, problems with coding, missing values, and/or errors inputting the data.
2. to examine the extent to which the assumptions of the statistics that you plan to use are met.
3. to get basic information regarding demographics of participants.
4. to examine the relationships between variables to determine how to conduct the hypothesis testing analyses.
There are many ways to check for errors. For example:

1. Compare the minimum and maximum values for each variable in Descriptives output.

2. Examine the means and standard deviations to see if they look reasonable, given what you know about the variables.

3. Examine the N column to see if any variables have a lot of missing data.

4. Look for outliers in the data.

Every inferential statistical test has assumptions.

Statistical assumptions are much like the directions for appropriate use of a product found in an owner’s manual.

Assumptions explain when it is and isn’t reasonable to perform specific test.

Check homogeneity of variances (standard deviation squared) of the groups to be compared are substantially different for the t test and ANOVA.

SPSS provides the Levene test to check this assumption and it provides ways to adjust the results if the variances are significantly different.
Normality – the frequency distribution would look like a symmetrical bell-shaped or normal curve.

It is important to check skewness.

Using SPSS output, SPSS recommends that you divide the skewness by its standard error.

If the result is less than 2.5 (which is approximately the $p = .01$) then skewness is not significantly different from normal.

A simple guideline is that if the skewness is less than plus or minus one, the variable is at least approximately normal.

Descriptive Statistics for Nominal (Dichotomous) or Qualitative Variables

Frequency distribution are tabular or graphical presentations of data that show each category for a variable and the frequency of the category occurrence in the data set.

Percentage for each category are often reported instead of, or in addition to, the frequencies.

The mode is the frequency of central tendency associated with a qualitative variable – distributions of variables may be unimodal, bimodal, or multimodal.
It is important to see if the means make sense, to examine the range of the data, and to check the shape of the distribution (i.e., skewness value).

Usually, Boxplots are used to examine variables – method of graphically representing ordinal and scale data.

A boxplot (box-and-whisker plot) is a graphic with information on one variable, much like a frequency polygon.

A boxplot gives you the mean, median, range, interquartile range, and skew of the distribution with just one picture.

A descriptive statistics report gives you fairly complete story of the data coz you

1. understand data better,

2. communicate with others who use statistics, and

3. persuade others.

The most interesting descriptive statistics reports are those that compare two or more distributions scores.
As for telling story, these points is suggested:

1. Form of distributions
2. Central tendency
3. Overlap of the two distributions
4. Interpretation of the effect size index
5. Of course, you should arrange the points so that your story is told well.

Example of writing a descriptive report.

The difference between height of women and men is about 5 inches (The mean and median height of women is 64.6 inches; the mean and median height of men is 69.8 inches). Although there is some overlap of the two distributions, the middle 50 percent of the men are all taller than the middle 50 percent of the women. In each distribution the heights are distributed symmetrically.
Other example.

<table>
<thead>
<tr>
<th></th>
<th>Self</th>
<th>Usage</th>
<th>User</th>
<th>Barrier</th>
<th>Inst</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2.99</td>
<td>.735</td>
<td>2.64</td>
<td>.779</td>
<td>3.31</td>
</tr>
<tr>
<td>Female</td>
<td>3.37</td>
<td>.593</td>
<td>2.97</td>
<td>.762</td>
<td>3.69</td>
</tr>
</tbody>
</table>

Female participant’s mean of the variable *self-reported knowledge* was 3.37 compared to 2.99 for males. Thus females said their *self-reported knowledge* was a little above “moderate” while males said “moderate”.

---

**Selecting Inferential Statistics**

When a group comparison or difference question is asked, the I.V. and design can be classified as between groups or within subjects.

Between-groups design are design where each participant in the study is in one and only one condition or group.

For example, in a study investigating the influenced of fathers’ education on student performance, there may be three groups (or levels or values) of the I.V., *father’s education*. 
If the researcher wished to have 20 participants in each group, then 60 participants would be needed to carry out the research.

In within-subjects designs (repeated measures designs), each of the conditions or levels of the I.V. are somehow connected to each of the other conditions or levels of the I.V.

These designs also include examples where the participants are matched by the researcher or in some natural way (e.g., twins, husband and wife, or mother and child).

Comparing performance on the same D.V. assessed before and after intervention is a common example of a repeated measure design.

**Selection of inferential statistics**

How do you decide which of the many possible inferential statistics to use?

Table 1 shows an appropriate selection of inferential statistics for basic, two variables, difference questions or hypotheses.

Table 2 shows an appropriate statistics for basic, two variables, associational questions or hypotheses.
Table 1. Selection of an Appropriate Statistics for Basic, Two Variables, Difference Questions or Hypotheses

<table>
<thead>
<tr>
<th>Single Measurement of IV</th>
<th>2 Levels</th>
<th>3 Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between (Ind) Within (Repeated)</td>
<td>Between</td>
<td>Within</td>
</tr>
<tr>
<td>Approx. normal, equal variances or mutually exclusive, independent groups</td>
<td>t-test or one-way ANOVA</td>
<td>One-Way ANOVA</td>
</tr>
<tr>
<td>ordinal or continuous, non-normal, independent, or mutually exclusive groups</td>
<td>Mann-Whitney</td>
<td>Wilcoxon</td>
</tr>
<tr>
<td>Nominal, independent, or mutually exclusive groups</td>
<td>χ²-test</td>
<td>McNemar</td>
</tr>
</tbody>
</table>

Table 2. Selection of an Appropriate Statistics for Basic, Two Variables, Associational Questions or Hypotheses

<table>
<thead>
<tr>
<th>Scale of Measurement of IV &amp; DV</th>
<th>Relate Two Variables or Scores for the same or Related Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Both normal, scale, no assumptions not mutually violated</td>
<td>Pearson (r) or Binomial Regress</td>
</tr>
<tr>
<td>Both at least ordinal or assumptions mutually violated</td>
<td>Kendall Tau or Spearman</td>
</tr>
<tr>
<td>Both are nominal</td>
<td>Cramér's Phi or Cramer's V</td>
</tr>
</tbody>
</table>
Interpreting the Results of Statistical Test

The SPSS computations produce a number or calculated value based on the specific data in your study.

The calculated value is compared to a critical value (found in a statistical table) that takes into account the degrees of freedom.

Then SPSS provides a probability value called significant.

If the probability is **less than** the preset alpha (usually .05), we can conclude that the results are statistically significant.

If \( p < \alpha \), reject \( H_0 \). Thus, if \( \alpha = .05 \) and \( p \leq .05 \), reject \( H_0 \).

If \( \alpha = .05 \) and \( p > .05 \), retain \( H_0 \). Of course, if \( p = .03 \), or .01, or .001, reject \( H_0 \).

If the probability is .051 or greater, retain \( H_0 \).

When \( H_0 \) is rejected, the difference is described as **statistically significant**.

When \( H_0 \) is retained - the difference is not significant.
If the p value is sufficiently small, one rejects $H_0$ and accepts $H_a$.

Most studies require very small p values, such as $p < .05$, before concluding that the data sufficiently contradict $H_0$ to reject it.

In such cases, results are said to be significant at the .05 level.

This means that if the null hypothesis were true, the chance of getting such extreme results as in the sample data would be no greater than 5%.

A statistically significant outcome does not give information about the strength or size of the outcome.

Therefore, it is important to know, in addition to information on statistical significance, the size of the effect.

Effect size is the strength of the relationship between the I.V. and the D.V., and/or the magnitude of the difference between levels of the I.V. with respect to the D.V.

If one compares two groups, the $d$ family effect size can be computed by subtracting the mean of the second group from the mean of the first group and dividing by the pooled standard deviation of both groups.
For $d$, Jacob Cohen (1969) proposed the conventions that follow—seem to be universally accepted:

- small effect, $d = .20$
- medium effect, $d = .50$
- large effect, $d = .80$

but $d$ can be more than 1.

Example of interpreting Independent $t$ Test.

<table>
<thead>
<tr>
<th></th>
<th>$M$</th>
<th>SD</th>
<th>$t$</th>
<th>df</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Years of teaching</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2.64</td>
<td>1.65</td>
<td>2.468</td>
<td>120.72</td>
<td>.015</td>
</tr>
<tr>
<td>Female</td>
<td>2.04</td>
<td>1.07</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-reported knowledge with ICT</td>
<td></td>
<td></td>
<td>-2.983</td>
<td>121</td>
<td>.003</td>
</tr>
<tr>
<td>Male</td>
<td>2.99</td>
<td>.74</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>3.37</td>
<td>.59</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For the variable years of teaching experience, the appropriate $t (120.72) = 2.468$, $p < .015$. The average years of teaching for male faculty members ($M = 2.64$) is significantly higher than the score for females ($M = 2.04$). The effect size $d$ is approximately .417, which is small to medium. For the variable self-reported knowledge, the appropriate $t (121) = -2.983$, $p < .003$. Based on means, female faculty members have higher self-reported knowledge with information and learning technology ($M = 3.37$) than males ($M = 2.99$). The variable self-reported knowledge with information and learning technology has medium to large effect size, $d$ is approximately .558.
Example of interpreting Pearson Correlation.

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Years of teaching</td>
<td></td>
<td>-.191*</td>
<td>-.132</td>
</tr>
<tr>
<td>2. Self-reported knowledge with ICT</td>
<td></td>
<td></td>
<td>.719***</td>
</tr>
<tr>
<td>3. Usage of ICT</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The strongest positive correlation, which would be considered a much larger than typical effect size, was between the variable *self-reported knowledge* of information and learning technology and the variable *usage* of information and learning technology, $r(121) = .719$, $p < .001$. This means that faculty members who had relatively high self-reported knowledge were likely to use information and learning technology in the teaching processes.

<table>
<thead>
<tr>
<th>PC Ownership</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Males</td>
</tr>
<tr>
<td>Females</td>
</tr>
<tr>
<td>Totals</td>
</tr>
</tbody>
</table>

There are 44 males who own a personal computer out of 73 male participants. Of the 50 female participants, 40 of them own a personal computer. The table shows the chi-square results and indicates that males and females are significantly different on whether they own a personal computer or not ($\chi^2 = 5.33$, $df = 1$, $N = 123$, $p < .021$). Female faculty members were more likely to own a personal computer than males. Phi is -.208 indicates the effect size is considered small to medium according to Cohen (1988).