LECTURE 7: IMPORTANT CONCEPTS IN HYPOTHESIS TESTING

While it is possible to conduct the different inferential statistical tests manually, it is easier and more efficient to use a software package for statistical analysis. Currently, IBM SPSS is among the most widely used software packages for statistical analysis. N.B. The electronic copy of **Lecture 6** contains direct links to YouTube tutorials demonstrating the necessary steps for conducting the different inferential statistical tests on SPSS.

1. Null and alternative hypotheses

Before conducting inferential statistical tests on SPSS, the researcher should consider two hypotheses. They are called the null hypothesis (H_0) and the alternative hypothesis (H_a). These hypotheses contain opposing viewpoints. ***The null hypothesis (H_0)*** is a hypothesis that says there is no statistical significance between the two variables. ***The alternative hypothesis (H_a)***, on the other hand, indicates that there is a statistical significance between the two variables being measured.

Example

In a study investigating the effects of English audiobooks on EFL learners' listening comprehension, the null hypothesis would be: **H_0** = English audiobooks have no significant effect on EFL learners' listening comprehension. Meanwhile, the alternative hypothesis would be: **H_a** = English audiobooks have a significant effect on EFL learners' listening comprehension.

2. Interpreting the significance level (alpha value)

After conducting the inferential test in SPSS, the software generates the significance level. This is also known as the *Alpha level (α)*. The p-value is the probability that the null hypothesis is true (correct). In other words, the p-value is the probability of making the wrong decision or that the results are due to random chance.

The p-value generated by the statistical tests in SPSS (i.e., $p< 0.05$) is a fractional number. To better understand this number, we should multiply it by 100 to get 5%. This number means that there is a 5% (or lower) that the null hypothesis is false. Usually:

- If $P \leq 0.05$, the results are statistically significant. It indicates that there is a 5% chance or less that the null hypothesis is correct.
- If $P > 0.05$, the results are not statistically significant. This indicates that there is more than a 5% chance that the null hypothesis is correct. Therefore, it is more likely that the results are due to chance.

The table below demonstrates the necessary decisions in each scenario.

Result	**Decision**
$P \leq 0.05$	Reject the null hypothesis (i.e., the results are statistically significant)
$P > 0.05$	Accept the null hypothesis (i.e., the results are not statistically significant)

3. Bad practices in advanced statistical testing

There are number of bad practices that the researchers should avoid when conducting inferential statistical tests. In this lecture, we focus on type 1 and type 2 errors, p-hacking, and publication bias.

I. Type I and Type II errors

"In statistical hypothesis testing, a type I error is the mistaken rejection of an actually true null hypothesis (also known as a "false positive" finding or conclusion; example: "*an innocent person is convicted*"), while a type II

error is the mistaken acceptance of an actually false null hypothesis (also known as a "false negative" finding or conclusion; example: "*a guilty person is not convicted*")"[1].

	TRUE NULL HYPOTHESIS	FALSE NULL HYPOTHESIS
Reject H_0	**Type 1 error (False positive)** [*finding significant results when there are no significant results*]	Correct decision
Accept H_a	Correct decision	**Type 2 error (False negative)** [*finding no significant results when there are significant results*]

To avoid these errors, the researcher should choose the appropriate inferential tests (i.e., parametric or non-parametric test) based on the type of data, sample size, and distribution of results.

II. P-hacking

"P-hacking (also known as *significance chasing* or *selective inferencing*) is the misuse of data analysis to find patterns in data that can be presented as statistically significant"[2]. This can also be done by only presenting the part of data that can generate significant results. In addition to being an unethical practice, p-hacking increases the chance of committing a type 1 (i.e., finding significant results when there are no significant results).

III. Publication bias

Publication bias is a type of bias that occurs when the outcome of a study influences the researcher's decision on publishing or withholding the findings.

PRACTICE

Indicate which advanced statistical test should be used in the following scenarios.

SCENARIO	TEST
1. To investigate the effectiveness of teaching English vocabulary with songs, you compared between an experimental group listening to songs with subtitles, a group listening to songs without subtitles, and group receiving traditional teaching. The study was conducted with a large sample and the data were normally distributed.	One-way ANOVA
2. You want to compare the grammatical accuracy of 60 adult Algerian EFL learners and 60 adult Tunisian EFL learners. The data is not normally distributed.	Mann-Whitney U test
3. You want to understand the relationship between the number of years living abroad and the pronunciation intelligibility of 8 ESL learners. The data are normally distributed.	Spearman correlation
4. You want to compare the pronunciation accuracy of an Experimental group involving 48 ESL learners who have been practicing with computer-assisted language learning technologies and that of a control group involving 40 EFL learners who have been practicing in a traditional classroom setting. The data are normally distributed.	Independent t-test
5. To investigate the effects of English podcasts on the listening comprehension of 7 EFL learners, you conduct a pre-test that is followed by 6 weeks of Podcasts listening sessions and end the training with a post-test.	Wilcoxon Rank sum test
6. You tested the lexical complexity in the speech of 12 EFL learners during 4 occasions. The data was normally distributed.	Friedman test

[1] Shuttleworth, M., & Wilson, L. (2021). Type I Error and Type II Error - Experimental Errors in Research. Retrieved 27 November 2021, from https://explorable.com/type-i-error

[2] Smith, G. D., & Ebrahim, S. (2002). Data dredging, bias, or confounding. BMJ (Clinical research ed.), 325(7378), 1437–1438. https://doi.org/10.1136/bmj.325.7378.1437