Worksheet A

Determining Type I and Type II Errors

In a statistical test of a hypothesis, researchers can come to two types of incorrect conclusions. The hypothesis can be inappropriately rejected (this is called type I error), or the hypothesis might be inappropriately accepted (a type II error). The Greek letter α is used to denote the probability of type I error, and the letter β is used to denote the probability of type II error. Consider the information in the table below:

		Decision (based on observed sample)	
		Accept H_o	*Reject H_o*
Outcome	*Null Hypo. (H_o) is True*	Correct Decision	Type I (α/Alpha) Error
	Alt. Hypo. (H_1) is True	Type II (β/Beta) Error	Correct Decision

Now, for each of the following scenarios identify (using ordinary language) what would constitute the null hypothesis, the alternative hypothesis, Type I error, and Type II errors.

Example A. A jury must decide whether a person accused of setting a deadly fire is guilty of arson.

H_0:____

H_1:____

	Accept H_o	*Reject H_o*
Null Hypo. (H_o) is True	Correct Decision	
Alt. Hypo. (H_1) is True		Correct Decision

Example B. A woman takes a home pregnancy test.

H_0:____

H_1:____

	Accept H_o	*Reject H_o*
Null Hypo. (H_o) is True	Correct Decision	
Alt. Hypo. (H_1) is True		Correct Decision

Example C. A college student pulls an all-nighter to study for a midterm exam.

H_0:____

H_1:____

	Accept H_o	*Reject H_o*
Null Hypo. (H_o) is True	Correct Decision	
Alt. Hypo. (H_1) is True		Correct Decision

Worksheet A

In which of the above situations might a false positive (or a Type I error) be the most egregious error? A false negative (or a Type II error)?

9.3. Interpreting *p*-values

P-value and its interpretation for experimental research is very similar to correlational research (see Exercise 7.2.). In experimental research, p represents a probability that H_0 is correct (that is, there is no difference in the means). Stated a different way, it is the probability of obtaining a difference as large (or larger) as observed by random sampling from the identical populations. Just like with correlation coefficient, *p*-value of 0.03 does not tell that there is 97% chance to replicate the results or that there is 97% chance that the difference you observed is real.

There is another important consideration. Imagine that you are testing a new therapy for treating depression, and when you compared experimental group to control group, you obtained a *p*-value of 0.10. Since this value is higher than 0.05, you cannot reject H_0; but, does this mean than H_0 is true? That is, does this mean that the new therapy has no effect?

Statistically, if you want to demonstrate that a treatment has an effect, you begin by assuming that the treatment has no effect (H_0). You then use this assumption to calculate a *p*-value (the probability of obtaining a treatment effect at least as strong as what was observed by a random chance). A small *p*-value contradicts this assumption and lets you reject the null hypothesis. Trying to prove the null using a *p*-value is, therefore, trying to prove that H_0 is true based on the assumption that it is true. This is why statisticians often say that you can never prove the null, you can only reject it. Put in another way, absence of evidence is not an evidence of absence – failure to reject the null does not mean that it must be true.

Now, using an ordinary language provide an interpretation for each of the scenarios below. Assume that the threshold for *p*-value is equal to 0.05; that is, all $p < 0.05$ is statistically significant and $p \geq 0.05$ are not statistically significant.

Example A. Participants rated a person who purchases environmentally friendly products as more cooperative than a person who purchases conventional products (M = 4.75, SD = 1.37, vs. M = 3.62, SD = 1.76), $t(57) = 2.76$, $p = .008$). (From Mazar & Zhong, 2010).

Example B. No significant difference has been found between young adults from lesbian and heterosexual single-mother households in the proportion who had experienced sexual attraction to someone of the same gender (9 of 25 vs. 4 of 20 respectively; $t(42) = 1.22$, $p = 0.11$). (From Tasker & Golombok, 1995).

Worksheet A

Example C. When told that they will be completing a problem-solving exercise for a study of general aspects of cognitive processes, women (M = .58) and men (M = .53) were equally accurate, $F < 1$, $p >$ 0.05. When given the same test but told that they will be completing a standardized test for a study of gender differences in mathematics performance, women (M = .36) were less accurate than men (M = .64), $F(1, 103) = 13.21$, $p < .01$. (From Johns et al., 2005).