LING82100: midterm

(Due 4/1)

1 Experimental design

Please read the following description of a study and indicate:

- whether it is an *experiment*, a *quasi-experiment*, or an *observational study*
- which variables are *dependent variables* (or *outcomes*) and which are *independent variables* (or *predictors*)
- whether it is *within-subjects* or *between-subjects*

  A team of developmental disorders researchers want to know whether autism spectrum disorder (ASD) or specific language impairment (SLI) affects short-term verbal memory. They recruit a sample of 110 children, ages 4–8. At the start of the study the children are diagnosed by an expert panel for autism spectrum disorder and specific language impairment. There are 51 children with ASD, 21 children with SLI, and 44 children with typical development (TD). Children then complete a non-word repetition task in which they listen to a recording of a made-up word of 1-4 syllables, and then attempt to repeat it. These repetitions are recorded, and an annotator counts the number of phoneme errors for each example. The team uses statistical analysis to test whether the number of errors are the same in all three diagnostic groups (ASD, SLI, and TD), while controlling for the length of the nonsense word (in syllables).

2 Standard error and confidence intervals

A continuously-valued sample of size $n = 64$ has $\bar{X} = 1.78$ and $s = .62$.

- What is the standard error of the sample mean?
- What is the 95% confidence interval for the sample mean?

(Hint: you may want to use R for this step.)

3 Null hypothesis significance testing

Let us suppose that we obtain a p-value of .02. Is this significant or non-significant at $\alpha = .05$? Now imagine that there is no population-level difference. Is this an example of Type I error, Type II error, or neither?

4 Power analysis

According to the pwr library, a two-sided two-sample Welch's t-test with $n_1 = 30$, $n_2 = 45$, $d = .8$ and $\alpha = .01$ has a power of .771:

```
> pwr.t2n.test(n1 = 30, n2 = 45, d = .8, sig.level = .01)

     t test power calculation

             n1 = 30
             n2 = 45
              d = 0.8
      sig.level = 0.01
          power = 0.770613
    alternative = two.sided
```

Which of the following would increase power of this test? (Indicate all that apply.)

- Increasing the effect size
- Increasing the sample sizes
- Decreasing the α-level

5 Reporting test results

Below, you will see the R commands and outputs for three statistical tests. For each, write a brief paragraph (one or two sentences) which reports the test. Your report should include sample means or medians (where appropriate), the test statistic, degrees of freedom (where appropriate), the p-value rounded to three significant digits, and whether or not one can reject the null hypothesis at $\alpha = .01$.

5.1 Fisher exact test

The following example, taken from a textbook by R. Fisher, tests the null hypothesis that monozygotic ("identical") twins are no more or less likely to both be convicted of a crime than dizygotic ("fraternal") twins.

```
> convictions <- matrix(c(2, 10, 15, 3),
+                       nrow = 2,
+                       dimnames = list(c("Dizygotic", "Monozygotic"),
+                                       c("Convicted", "Not convicted")))
> fisher.test(convictions)

        Fisher's Exact Test for Count Data
```

```
data:  convictions
p-value = 0.0005367
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.003325764 0.363182271
sample estimates:
odds ratio
0.04693661
```

5.2 Welch t-test

The following tests the null hypothesis that *I. versicolor* and *I. virginica* have the same sepal widths (cf. Homework 04).

```
> by.species <- split(iris, iris$Species)
> versicolor <- by.species$versicolor
> virginica <- by.species$virginica
> t.test(versicolor$Sepal.Width, virginica$Sepal.Width)

        Welch Two Sample t-test

data:  versicolor$Sepal.Width and virginica$Sepal.Width
t = -3.2058, df = 97.927, p-value = 0.001819
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.33028364 -0.07771636
sample estimates:
mean of x mean of y
    2.770     2.974
```

5.3 Kendall τ_b test

The following tests the null hypothesis that there is no correlation between a 7-point wellformedness rating and degree of clausal embedding (1–3).

```
> embedding <- read.csv("embedding.csv")
> with(embedding, cor.test(rating, depth, method = "kendall"))

        Kendall's rank correlation tau

data:  rating and depth
z = -1.2687, p-value = 0.2046
alternative hypothesis: true tau is not equal to 0
sample estimates:
       tau
-0.1927072
```