

# LING82100: final exam

(Due 5/22)

Please submit your solution as a PDF file using GitHub Classroom. Your solution are due at 11:59pm on Friday, 5/22. Please email the instructor with any questions you have about the exam.

## 1 Scales of measurement

State whether the following types of variables are *categorical*, *ordinal*, *interval*, or *ratio*:

- GRE score
- word frequency
- teaspoons of baking powder
- species of *Iris*
- city gas mileage (in miles per gallon)
- grammatical acceptability ratings, collected using a 7-point Likert scale (1 labeled “completely ungrammatical”, 7 labeled “fully grammatical”)

## 2 Study design

Please read the following description of a study and indicate

- whether it is an *experiment*, a *quasi-experiment*, or an *observational study*,
- whether the primary manipulation is *within-subjects* or *between-subjects*, and
- what sorts of *random intercepts* and *random slopes* would be appropriate, if any.

A psycholinguist is interested whether morphologically complex words—namely those with inflectional affixes—take longer to “name”, i.e., to read out loud, than uninflected words. 45 students, part of a PSYCH 101 subject pool, participate in the study. During the roughly 30-minute session, each subject is presented with a randomly ordered sequence containing both inflected and uninflected words, shown on an LCD screen, and asked to say the word out loud as fast as possible. A special voice-activated microphone is used to measure reaction time. The experimenter tracks subject ID, age, and gender, trial number, and the length, neighborhood density, word frequency, and base frequency of the words presented for naming.

### 3 Standard error and confidence intervals

A continuously valued sample of size  $n = 50$  has a mean of  $\bar{X} = 22.1$  and  $s = 1.2$ .

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

### 4 Power analysis

At some given  $\alpha$ -level, study  $A$  has power .5 and study  $B$  has power .6.

- Does  $A$  have a higher or lower probability of type I error (i.e., the rejection of a true null hypothesis) than  $B$ ?
- Does  $A$  have a higher or lower probability of type II error (i.e., the failure to reject a false null hypothesis) than  $B$ ?

### 5 Test interpretation

Below, you will see a brief description of a null hypothesis test and associated R output. You are to convert the R output into something that can be reported for publication. **Do not forget to**

- round appropriately,
- report all relevant test statistics, degrees of freedom (where appropriate), and associated  $p$ -values,
- and state whether the result or results are significant at  $\alpha = .05$ .

Additional instructions will be provided for each test/scenario.

#### 5.1 The binomial test

Bottles of Jeppson's Malört proudly announce that

...only 1 out of 49 men will drink Jeppson's Malört.

Let this be our null hypothesis. Then, suppose that, in a random sample of 8 men, 2 men will drink Jeppson's Malört. This gives rise to the following binomial test:

```
> binom.test(2, 8, 1 / 50)
```

```
Exact binomial test
```

```
data: 2 and 8
```

```
number of successes = 2, number of trials = 8, p-value = 0.01034
alternative hypothesis: true probability of success is not equal to 0.02
95 percent confidence interval:
 0.03185403 0.65085579
sample estimates:
probability of success
                0.25
```

In prose, report the results, including

- the null hypothesis parameter
- the test statistic  $x$ ,
- the sample size  $n$ ,
- the  $p$ -value,
- the 95% confidence interval,

and whether or not we should reject the null hypothesis.

**Hint** Pay attention to the distinction between the population parameter (here, the probability of a man drinking Jeppson's Malört, according to the bottle), and the hypothesis test  $p$ -value.

## 5.2 The $t$ -test

Suppose we wish to test whether Dutch regular verbs (e.g., *printen/printte* 'to print') are longer or shorter than irregular verbs (e.g., *bidden/bad* 'to pray'). To test the null hypothesis that there is no difference in lengths between the two groups, suppose we perform the following  $t$ -test.

```
> t.test(length ~ regularity, data = d)

Welch Two Sample t-test

data:  length by
t = 1.8896, df = 270.56, p-value = 0.05988
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.006751323  0.329168463
sample estimates:
mean in group irregular    mean in group regular
                4.534591                4.373383
```

In prose, report the results, including

- what type of  $t$ -test was used,

- the test statistic,
- the degrees of freedom,
- the  $p$ -value,
- the 95% confidence interval,

and whether or not we should reject the null hypothesis.

### 5.3 Logistic regression

Suppose we elicit two tokens of the word *street* from pedestrians in Columbus, OH, coding whether the initial sibilant is [s] or [ʃ] (str). Then, suppose we fit a logistic regression model with main effects of speakers' hypothesized social class (WC: working class; LMC: lower middle class; UMC: upper middle class) and whether each rendition is spoken with less or more emphasis (emphatic).

```
> contrasts(d$class) <- contr.sum
> contrasts(d$class)
      [,1] [,2]
LMC     1    0
UMC     0    1
WC     -1   -1
> contrasts(d$emphatic) <- contr.sum
> contrasts(d$emphatic)
      [,1]
less     1
more    -1
> r <- glm(str ~ class + emphatic, data = d, family = binomial)
> summary(r)
...
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.4981     0.2010   7.453 9.09e-14 ***
class1        -0.3711     0.2382  -1.558  0.119
class2         1.2624     0.3145   4.014 5.98e-05 ***
emphatic1     -0.7957     0.1797  -4.428 9.51e-06 ***
...
> lrtest(update(r, . ~ . - class), r)
Likelihood ratio test

Model 1: str ~ emphatic
Model 2: str ~ emphatic + class
  #Df LogLik Df  Chisq Pr(>Chisq)
1    2 -121.52
```

	Coef.	SE	$\chi^2$	$p(\chi^2)$
(Intercept)	...	...		
Class:			...	...
WC	...			
LMC	...	...		
UMC	...	...		
Emphasis:			...	...
less	...	...		
more	...			

Table 1: Sample results table for subsection 5.3.

```

2   4 -109.73  2 23.582   7.57e-06 ***
...
> lrtest(update(r, . ~ . - emphatic), r)
Likelihood ratio test

Model 1: str ~ class
Model 2: str ~ emphatic + class
  #Df  LogLik Df  Chisq Pr(>Chisq)
1    3 -120.89
2    4 -109.73  1 22.328  2.299e-06 ***
...
> summary(glht(r, linfct = mcp(class = "Tukey"))
          Simultaneous Tests for General Linear Hypotheses
...
Linear Hypotheses:
              Estimate Std. Error z value Pr(>|z|)
UMC - LMC == 0   1.6335     0.5023   3.252  0.0034 **
WC - LMC == 0   -0.5203     0.3642  -1.429  0.3217
WC - UMC == 0   -2.1538     0.5091  -4.230 <0.001 ***
...

```

Please populate Table 1, remembering to round appropriately. Then, in prose, report the Tukey test results for class using the “inequality” notation we used for Homework 7.

## 5.4 Mixed effects linear regression

Suppose we record productions of three /t/ and three of /d/ from a group of English and Spanish speakers, measuring voice onset time (VOT). Then, suppose we fit a mixed effects linear model with main effects of language and phoneme, and random intercepts for participant and item (i.e., word).

```

> contrasts(d$language) <- contr.sum
> contrasts(d$language)
      [,1]
english    1
spanish   -1
> contrasts(d$phoneme) <- contr.sum
> contrasts(d$phoneme)
      [,1]
d       1
t      -1
> r <- lmer(vot ~ language + phoneme + (1 | participant) + (1 | item), d)
> summary(r)
...
Fixed effects:
              Estimate Std. Error t value
(Intercept)   4.706      3.134   1.502
language1    27.778      3.115   8.917
phoneme1     -26.927      3.109  -8.661
...
> lrtest(update(r, . ~ . - language), r)
Likelihood ratio test

Model 1: vot ~ phoneme + (1 | participant) + (1 | item)
Model 2: vot ~ language + phoneme + (1 | participant) + (1 | item)
  #Df LogLik Df  Chisq Pr(>Chisq)
1   5 -2700.1
2   6 -2686.6  1 26.939  2.099e-07 ***
...
> lrtest(update(r, . ~ . - phoneme), r)
Likelihood ratio test

Model 1: vot ~ language + (1 | participant) + (1 | item)
Model 2: vot ~ language + phoneme + (1 | participant) + (1 | item)
  #Df LogLik Df  Chisq Pr(>Chisq)
1   5 -2699.8
2   6 -2686.6  1 26.387  2.794e-07 ***
...

```

Please populate Table 2, remembering to round appropriately.

	Coef.	SE	$\chi^2$	$p(\chi^2)$
(Intercept)	...	...		
Language:			...	...
English	...	...		
Spanish	...			
Phoneme:			...	...
/t/	...			
/d/	...	...		

Table 2: Sample results table for subsection 5.4.