3 Data collection & samples: Assignment

Instructions

Submit your answers to the **n** first exercises through Aula Global. The remainder of your assignments concern self-studies.

Describing samples

You want to study whether speed of second language acquisition changes depending on the first language of learners. To this end, you recruit a sample of 50 monolingual first year students enrolled at UPF who just started their first class of Catalan.

1. Is this a representative sample?

You want to study whether the speed of reaching B1 in Catalan changes depending on the first language of learners. To this end, you recruit a sample of 50 monolingual first year students enrolled at UPF who just started their first class of Catalan.

2. Is this a representative sample?

You want to study whether the speed of reaching B1 in Catalan by first year students at UPF changes depending on the first language of learners. To this end, you recruit a sample of 50 monolingual first year students enrolled at UPF who just started their first class of Catalan.

3. Is this a representative sample?

You want to study whether the speed of reaching B1 in Catalan by first year students at UPF changes depending on the first language of learners. To this end, you recruit a sample of all monolingual first year students enrolled at UPF who just started their first class of Catalan.

4. Is this a representative sample?

You want to study whether the speed of reaching B1 in Catalan by monolingual first year students at UPF changes depending on the first language of learners. To this end, you recruit a sample made up of all monolingual first year students enrolled at UPF who just started their first class of Catalan.

5. Is this a representative sample?

You want to study whether the speed of reaching B1 in Catalan by monolingual first year students at UPF changes depending on the first language of learners. To this end, you recruit a sample made up of all monolingual first year students enrolled at UPF who just started their first class of Catalan and 50 monolingual first year students enrolled at UPF who just started their first class of German.

6. Is this a representative sample?

Simulating samples

Here are some simulated samples of the number of gestures that speakers used in a formal vs. informal context an experimental trial.[1] Let's assume that speakers use an average of 55 gestures when speaking with a friend (informal context) and 45 when speaking with a professor (formal), and to be Poisson distributed.

You are tasked to determine what sample sizes, if any, from the six simulated below, give a good approximation of the true mean difference between formal and informal contexts. More precisely, determine which of the sample sizes, if any, gives an estimate at least within 4 average gesture of the true mean difference of the contexts.

Sample 1 (1 subject per condition)

```
set.seed(101) #this makes the random draws be the same each time

sample1_informal <- rpois(n = 1, lambda = 55)   #1 draw from Poisson distribution with lambda = 55
sample1_formal <-  rpois(n = 1, lambda = 45) #1 draw from Poisson distribution with lambda = 45
print(mean(sample1_informal) - mean(sample1_formal))
```

```
## [1] 5
```

Sample 2 (5 subjects per condition)

```
set.seed(11) #this makes the random draws be the same each time

sample2_informal <- rpois(n = 5, lambda = 55)   #5 draw from Poisson distribution with lambda = 55
sample2_formal <-  rpois(n = 5, lambda = 45) #5 draw from Poisson distribution with lambda = 45
print(mean(sample2_informal) - mean(sample2_formal))
```

```
## [1] 4.2
```

Sample 3 (15 subjects per condition)

```
set.seed(191) #this makes the random draws be the same each time

sample3_informal <- rpois(n = 15, lambda = 55)   #15 draw from Poisson distribution with lambda = 55
sample3_formal <-  rpois(n = 15, lambda = 45) #15 draw from Poisson distribution with lambda = 45

print(mean(sample3_informal) - mean(sample3_formal))
```

```
## [1] 15.26667
```

Sample 4 (40 subjects per condition)

```
set.seed(101) #this makes the random draws be the same each time

sample4_informal <- rpois(n = 40, lambda = 55)   #40 draw from Poisson distribution with lambda = 55
sample4_formal <-  rpois(n = 40, lambda = 45) #40 draw from Poisson distribution with lambda = 45
```

[1] We will look at the actual data from this experiment in future sessions (**???**).

```r
print(mean(sample4_informal) - mean(sample4_formal))
```

```
## [1] 8.7
```

Sample 5 (100 subjects per condition)

```r
set.seed(101) #this makes the random draws be the same each time

sample5_informal <- rpois(n = 100, lambda = 55)   #100 draws from Poisson distribution with lambda = 55
sample5_formal <-  rpois(n = 100, lambda = 45) #100 draws from Poisson distribution with lambda = 45
print(mean(sample5_informal) - mean(sample5_formal))
```

```
## [1] 9.67
```

Sample 6 (1000 subjects per condition)

```r
set.seed(101) #this makes the random draws be the same each time

sample6_informal <- rpois(n = 1000, lambda = 55)   #1000 draws from Poisson distribution with lambda =
sample6_formal <-  rpois(n = 1000, lambda = 45) #1000 draws from Poisson distribution with lambda = 45
print(mean(sample6_informal) - mean(sample6_formal))
```

```
## [1] 10.462
```

7. What is the smallest sample size that gives an estimate at least within 4 average gestures of the true mean difference?
8. What is the smallest sample size that suggests that there is a difference between gestures in formal and informal contexts?

Self-study

Answer these four questions concerning your analysis report:

1.1 Your general research question (e.g., how does a learner's first language affect learning a second language?) 1.2 Your specific research question (e.g., Is there a difference between the speed in which Spanish and German monolinguals learn Catalan?) 1.3 What kind of data would you use to address (1.2) if you had unlimited resources? 1.4 What kind of data are you planning to use to address (1.2) within the scope of this class?

Elaborate with as much detail as you (currently) have but it is perfectly acceptable if you answer in bullet points and single sentences. You will be asked to submit these answers for next week's assignment. A randomly selected peer will give you feedback on them

References