11.2: Simulation Overview

The basic idea here is actually pretty simple. You make some assumptions about how many subjects will be in your design (set N), you make some assumptions about the distributions that you will be sampling your scores from, then you use R to fabricate fake data according to the parameters you set. Once you build some simulated data, you can conduct a statistical analysis that you would be planning to run on the real data. Then you can see what happens. More importantly, you can repeat the above process many times. This is similar to conducting a replication of your experiment to see if you find the same thing, only you make the computer replicate your simulation 1000s of times. This way you can see how your simulated experiment would turn out over the long run. For example, you might find that the experiment you are planning to run will only produce a "signficant" result 25% of the time, that's not very good. Your simulation might also tell you that if you increase your N by say 25, that could really help, and your new experiment with N=25 might succeed 90% of the time. That's information worth knowing.

Before we go into more simulation details, let's just run a quick one. We'll do an independent samples \(t\)-test. Imagine we have a study with N=10 in each group. There are two groups. We are measuring heart rate. Let's say we know that heart rate is on average 100 beats per minute with a standard deviation of 7. We are going to measure heart rate in condition A where nothing happens, and we are going to measure heart rate in condition B while they watch a scary movie. We think the scary movie might increase heart rate by 5 beats per minute. Let's run a simulation of this:

```
group_A <- rnorm(10,100,7)
group_B <- rnorm(10,105, 7)
t.test(group_A,group_B,var.equal = TRUE)
```

run | restart | restart & run all

```
        Two Sample t-test

data:  group_A and group_B
t = -1.7061, df = 18, p-value = 0.1052
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.434802   1.185828
sample estimates:
mean of x mean of y
 98.20342 103.32791
```

We sampled 10 scores from a normal distribution for each group. We changed the mean for group_b to 105, because we were thinking their heart rate would be 5 more than group A. We ran one \(t\)-test, and we got a result. This result tells us what happens for this one simulation.

We could learn more by repeating the simulation 1000 times, saving the \(p\)-values from each replication, and then finding out how many of our 1000 simulated experiments give us a significant result:

```
save_ps<-length(1000)
for(i in 1:1000){
  group_A <- rnorm(10,100,7)
  group_B <- rnorm(10,105, 7)
  t_results <- t.test(group_A,group_B,var.equal = TRUE)
  save_ps[i] <- t_results$p.value
}
prop_p<-length(save_ps[save_ps<0.05])/1000
print(prop_p)
```

run | restart | restart & run all

```
[1] 0.344
```

Now this is more interesting. We found that 34.4% of simulated experiments had a \(p\)-value less than 0.05. That's not very good. If you were going to collect data in this kind of experiment, and you made the correct assumptions about the mean and standard deviation of the distribution, and you made the correct assumption about the size of difference between the groups, you would be planning to run an experiment that would not work-out most of the time.
What happens if we increase the number of subject to 50 in each group?

```
save_ps<-length(1000)
for(i in 1:1000){
  group_A <- rnorm(50,100,7)
  group_B <- rnorm(50,105, 7)
  t_results <- t.test(group_A,group_B,var.equal = TRUE)
  save_ps[i] <- t_results$p.value
}
prop_p<-length(save_ps[save_ps<0.05])/1000
print(prop_p)
```

run | restart | restart & run all

```
[1] 0.957
```

Ooh, look, almost all of the experiments are significant now. So, it would be better to use 50 subjects per group than 10 per group according to this simulation.
Of course, you might already be wondering so many different kinds of things. How can we plausibly know the parameters for the distribution we are sampling from? Isn't this all just guess work? We'll discuss some of these issues as we move forward in this chapter.
