Week 8 Self-directed Tutorial Solutions

Andrew Perfors

This document contains the exercises (with solutions) that you can do in place of an in-person tutorial this week. Please ask any questions about them on the discussion board or as part of the Q&A.

It seems evident that most people want to both rescue LFB and Foxy *and* attack the others, so Gladly and Doggie decide to get a group together to go to Otherland. The plan is to travel together at the border, and then split up. While one group creates a diversion by attacking the main government buildings in Otherland, the other group will search for LFB and Foxy.

In order to decide who will go in each group, they decide to measure each person’s level of *spy ability* and *aggressiveness*. The idea is to put the best spies in the group that will search for LFB and Foxy and the most aggressive people in the group that attacks.

The *spy* measure has been normed to have a mean of 14.4 and a standard deviation of 2.3, while the *aggressiveness* measure normally has a mean of 54.3 and a standard deviation of 6.3.

1. Wiggly Tail has a *spy* score of 14.6 and an *aggressiveness* score of 55.0. Assuming each person should go in the group which corresponds to their highest z-score, which group should Wiggly Tail go with? Why do we use z-scores for this instead of just the raw scores? (You can put your calculations in the R code chunk below).

spy <- (14.6-14.4)/2.3  
aggressiveness <- (55.0-54.3)/6.3

*ANSWER: Their spy z-score s 0.087, and aggressiveness is 0.11. They should go in the group that attacks because they are relatively more aggressive than they are a good spy. Using z-scores is sensible because it lets us compare things across different scales.*

1. The data frames dattack contains the *aggressiveness* scores for all the people who ended up deciding to join the attacking group, while dsearch contains the *spy* scores for all of the people who joined the search group. (Both of these are in the column called *score*). Manually calculate the t test statistic for each of them, assuming the true population mean under the null hypothesis is 14.4 for *spy* and 54.3 for *aggressiveness*.

attackTstat <- (mean(dattack$score)-54.3)/(sd(dattack$score)/sqrt(nrow(dattack)))  
attackTstat

## [1] 3.20087

searchTstat <- (mean(dsearch$score)-14.4)/(sd(dsearch$score)/sqrt(nrow(dsearch)))  
searchTstat

## [1] 2.905894

1. Use a one-sample t-test to determine whether each of the groups scores higher than the normed mean for their respective measures (i.e., whether the attack group has a higher mean *aggressiveness* score than 54.3 and the search group has a higher mean *spy* score than 14.4). Was your test statistic calculation in #2 correct? What do the results indicate about the null hypothesis? How might you interpret these results?

t.test(dsearch$score,mu=14.4)

##   
## One Sample t-test  
##   
## data: dsearch$score  
## t = 2.9059, df = 55, p-value = 0.005267  
## alternative hypothesis: true mean is not equal to 14.4  
## 95 percent confidence interval:  
## 14.80897 16.62656  
## sample estimates:  
## mean of x   
## 15.71776

t.test(dattack$score,mu=54.3)

##   
## One Sample t-test  
##   
## data: dattack$score  
## t = 3.2009, df = 60, p-value = 0.002192  
## alternative hypothesis: true mean is not equal to 54.3  
## 95 percent confidence interval:  
## 55.40147 59.07183  
## sample estimates:  
## mean of x   
## 57.23665

*ANSWER: Both of these are significant. This probably indicates that they did a pretty good job at getting above-average people along their respective measures into the appropriate groups.*

Having split into groups, and preparing for action, everyone sets off. Doggie, Gladly, and the others are nervous but excited. They hope to be heroes, rescuing LFB and Foxy and solving everyone’s food problems.

After a few hours of traveling, it is late afternoon and everyone is getting tired. Suddenly, Gladly hears a noise in the forest ahead of them, and he shouts: “Who goes there?”

For a long tense moment there is no sound except for the soft rustle of the leaves in the trees. Then a gruff voice shouts back: “Show yourselves. This is Otherland.”

Doggie tenses and holds up his staff, preparing to use it if necessary. “We’re looking for one of our friends, who you kidnapped,” he shouts in return. “Where is she?”

Another long moment. Everyone slowly crowds behind Doggie and Gladly, holding their weapons at the ready.

The leaves part, revealing a large group of Others, led by the biggest bear anybody of them has ever seen. All of the Others have their own weapons, which they are holding ready as well. The bear growls. “If you want to fight, let’s fight,” he says.

Doggie is trembling with fear, but he doesn’t run. He lowers his staff and he can hear Gladly shifting beside him, preparing for a battle.

“It doesn’t have to be this way,” says Gladly. “Just tell us where our friends are and what you’ve been doing with our food.”

“What *we’ve* been doing with *your* food?” the huge bear asks incredulously. “What we did to *your* friends? You’re the one who sent spies to us initially. We’ve done nothing to your food.”

A snake slithers out from behind the bear and hisses, “We should be asking you what you’ve been doing to our food. We’re starving over here.”

“Impossible!” Doggie blurts out. “We’re running out too! Where could it be going, if you’re not stealing it?”

A guitar (Doggie blinks incredulously at this) sticks its… head? strut? out from the other side of the bear and says, “You seriously want us to believe that?”

“Er…” Quackers interrupts. “Are you.. a guitar?”

The guitar sighs. “Yes, and my name is Kevin, and WOW what a surprise I’m a sentient guitar, blah blah blah,” it says irritatedly. “Can we get back to the point?”

A unicorn standing next to the bear says, “What *is* the point, anyway? I don’t think they have our food, Super Size. LFB and Foxy were probably telling the truth.”

“LFB and Foxy?!?” exclaims Gladly. “You spoke to them? Where are they? Are they safe?”

“Er.. yeah,” says a voice from behind the crowd of Others. It’s a very *familiar* voice, a voice that makes the hearts of all of the people from Bunnyland leap a bit. There’s a collective holding of breaths, and then the welcome faces of LFB and Foxy – both a bit tired and bedraggled looking, but safe and sound – peek out next to the unicorn. “I’m okay,” LFB says, and hanging back behind her a bit, Foxy nods as well.

“You’re safe!” Doggie charges forward to give LFB a hug, and then pauses when everyone from Otherland bristles and raises their weapons. “Are.. *are* you okay?” he asks more uncertainly. “Are you prisoners?”

Foxy sighs. “We’re not prisoners,” she says. “And we’ve been starting to work together to try to sort this out.”

LFB adds somewhat pointedly, “But it’s been hard to convince them to trust us entirely, and I have to say you haven’t helped matters by charging in with all these weapons and things.”

“We were trying to rescue *you*!” says Gladly.

“And I appreciate it, I really do. But don’t you see what this looks like? Do you blame them for being a little wary?”

“What, are you on their side now?” asks Doggie.

“There is no *side*,” says LFB. “And, honestly? Take a look. Most of the Others are way bigger than us, and have more and better weapons. Do you really want to be fighting them?”

Doggie squints skeptically over. “Sure, maybe a little,” he says. But they don’t have a *huge* advantage.”

The unicorn laughs. “Seriously? You’re delusional. We massively outweigh you.”

Before this can go further, Gladly bursts out, “Why don’t we measure ourselves and see?”

It is quickly agreed. Everybody on each side passes around a paper on which all each person writes down their weight and how many weapons they are carrying.

The results are found in the tibble dw, which has been loaded for you. It contains four columns:

*person* indicates the id of the person

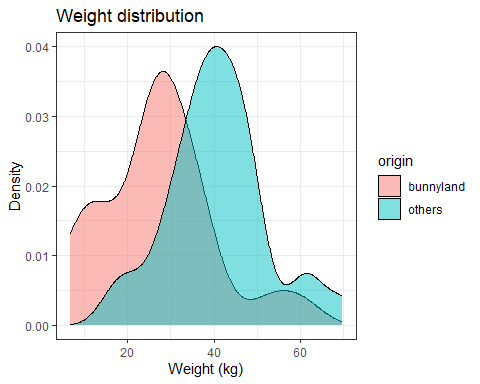
*origin* indicates whether they are from the Others or Bunnyland

*weight* indicates their weight in kilos

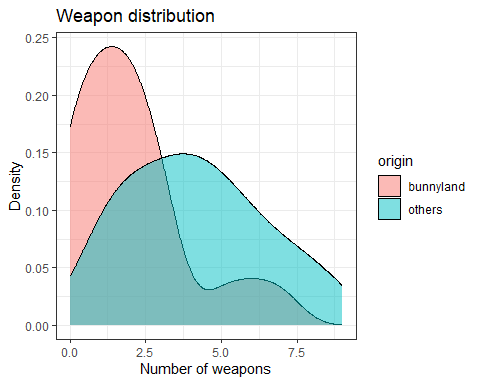
*weapons* indicates how many weapons they are carrying

As always, first we look at our data. This time, just for variety, we’ll use density plots. What do the distribution of weapons and weights look like among the two groups? Press play to see.

dw %>%   
 ggplot(mapping=aes(x=weight,fill=origin)) +   
 geom\_density(alpha=0.5) +  
 theme\_bw() +  
 labs(title="Weight distribution",x="Weight (kg)",y="Density")



dw %>%   
 ggplot(mapping=aes(x=weapons,fill=origin)) +   
 geom\_density(alpha=0.5) +  
 theme\_bw() +  
 labs(title="Weapon distribution",x="Number of weapons",y="Density")



The first question everyone has is whether the mean weights of the two groups (Bunnyland and Others) are significantly different from each other. Since you’re comparing the mean of two groups, you want to do an independent-samples t-test.

1. The code chunk below calculates the mean and n of each of the two groups. Use these numbers to calculate the test statistic you would find if you did a Student t-test, assuming an estimated population standard deviation of 12.155.

dw\_sum <- dw %>%  
 group\_by(origin) %>%  
 summarise(mean = mean(weight),  
 n = n()) %>%  
 ungroup()  
  
dw\_sum

## # A tibble: 2 × 3  
## origin mean n  
## <fct> <dbl> <int>  
## 1 bunnyland 27.1 24  
## 2 others 40.7 29

# Put your calculation here:  
(27.1-40.7)/(12.155\*(sqrt((1/24)+(1/29))))

## [1] -4.054624

Is this test statistic larger or smaller if the data are more unlikely?

*ANSWER: It is larger the more unlikely the data are.*

1. What is the difference between a Student t-test and a Welch t-test? Why is it suggested that you generally use a Welch?

*ANSWER: A Student t-test assumes homogeneity of variance: that both groups have the same variance. It’s generally best to default to Welch because it doesn’t make this assumption, and this assumption is often not true.*

1. Run a Welch independent-samples t-test to see whether the mean of the weights between the two groups is different. What do you find? What does this mean for your research hypothesis?

# Complete code:  
t.test(weight ~ origin, data=dw)

##   
## Welch Two Sample t-test  
##   
## data: weight by origin  
## t = -4.0198, df = 46.54, p-value = 0.0002116  
## alternative hypothesis: true difference in means between group bunnyland and group others is not equal to 0  
## 95 percent confidence interval:  
## -20.494403 -6.820654  
## sample estimates:  
## mean in group bunnyland mean in group others   
## 27.07833 40.73586

*ANSWER: There is a significant difference between the average weight in the two groups, t(46.54)=-4.02, p=.0002. This suggests that the Others weigh significantly more than the people from Bunnyland, on average - perhaps Doggie and Gladly might be wiser not to fight them!*

1. Calculate and interpret the effect size for the difference in weights between groups. (Note: use the cohensD() function but include the argument method="unequal" because this corresponds to the Welch t-test which presumes unequal variances and we want the cohensD to match the t-test).

cohensD(formula = weight ~ origin, data=dw, method="unequal")

## [1] 1.115417

*ANSWER: Cohen’s D is 1.12, which is pretty large (especially for the kind of studies that are often found in psychology).*

Technically speaking, before running any test you should have checked the assumptions, but we wanted you to practice running a t-test before doing so. That said, let’s do the analysis properly when it comes to exploring whether the two groups have significantly different mean numbers of weapons (the *weapons* variable). Today we’ll focus on the assumption of normality. (As the Welch test doesn’t rely on the assumption that variances of the two groups are homogeneous, we won’t directly test the homogeneity of variance assumption today, but you will learn about this later in the context of ANOVA).

1. What test would you use to determine whether the distributions are normal? Apply this test to the weapons data. What do your results indicate?

# we should test the normality of each group taken individually, which would be done this way:  
  
shapiro.test(dw$weapons[dw$origin=='bunnyland'])

##   
## Shapiro-Wilk normality test  
##   
## data: dw$weapons[dw$origin == "bunnyland"]  
## W = 0.85779, p-value = 0.003061

shapiro.test(dw$weapons[dw$origin=='others'])

##   
## Shapiro-Wilk normality test  
##   
## data: dw$weapons[dw$origin == "others"]  
## W = 0.94878, p-value = 0.1702

*ANSWER: The normality of scores in each group were tested using the Shapiro-Wilk test. It was not significant in the Otherland group, W=0.949, p=.170, indicating that the distribution of scores was normal in that group. However, the Shapiro-Wilk test was significant in the Bunnyland group, W=0.858, p=.003, indicating that the distribution of scores was not normal in that group.*

1. Do the appropriate test to compare the mean number of weapons in each of the two groups, given your results from #8. Explain what the results of that test are and why you ran it.

wilcox.test(weapons ~ origin, data=dw)

## Warning in wilcox.test.default(x = DATA[[1L]], y = DATA[[2L]], ...): cannot  
## compute exact p-value with ties

##   
## Wilcoxon rank sum test with continuity correction  
##   
## data: weapons by origin  
## W = 154.5, p-value = 0.0004956  
## alternative hypothesis: true location shift is not equal to 0

*ANSWER: Because the distribution of scores in the Bunnyland group are not normal, we need to run the Wilcoxon. The results indicate that the two groups are significantly different from each other, W=154.5, p=.0005. In other words, the Others appear to have significantly more weapons per person than the folks from Bunnyland. (Note, by the way, that the Wilcoxon alone only tells you that there is a significant difference between the Others and the people from Bunnyland; it doesn’t tell you WHO has more weapons. To figure that out, we created a figure – you could also create a summary tibble where you group by origin. You always do want to figure out the direction (i.e., which one is greater), since the interpretation is massively different depending on that, but that is a separate step.*

After looking at this data, Gladly and Doggie and the rest of the folks from Bunnyland seems somewhat chastened. “We don’t have much chance against the others,” Gladly murmurs sadly.

“Maybe that means *we* should attack *you*!” shouts the snake. He turns to Super Size, the enormous bear. “We could totally win,” he says.

“Just because we can doesn’t mean we should,” says Super Size. “But I don’t know what else we can do. Even if it’s not their fault, we’re starving, and they have food we could eat. I can’t help but care more about my people than strangers.”

“Oh will you all please STOP?” LFB finally shouts. “You’re both doing the same horrible thing. The only reason you want to attack the other group – *both* of you – is that you see each other as strangers, as different. And that makes it seem okay. But look” – and at this LFB turns toward the Bunnyland contingent – “you guys see a scary big bear, but I see my new friend Super Size, who is the most cuddly and kind person you’ll ever met. You see a weird guitar and I see Kevin, who is surly sometimes but has a huge heart hiding a lot of hurt. You see a weird unicorn and I see Rainbow, who was nice to me from the very beginning even when for all she knew I was an evil enemy come to destroy her.”

At this point, Foxy turns to the Others. “Same for you. You see an aggressive dog, and I see my friend Doggie, who yes is impulsive but is also the most courageous person I’ve ever met. You see a mean bear and I see Gladly, who was willing to risk his life just for the merest chance of saving us.”

LFB implores: “We need to stop seeing each other as strangers and start seeing each other as people. That’s the only solution: we need to work *together* to solve this, which means we have to start trusting.”

Everyone shuffles around, but it’s obvious that they’re thinking. Finally Rainbow, the unicorn, asks “That’s a good idea but *how* do we do it? We can’t just manufacture trust out of nowhere.”

A little blue penguin speaks up. “Is there some kind of trust exercise or something we could do?”

Gladly says, “Yeah! I know! Do you know the game ‘two truths and a lie’? We could play that.”

“I love that game,” says Quackers. “You talk to people one on one, and tell them two truths about yourself and one lie and they have to guess which is the lie. It’s a good way to get to know people fast.”

“That sounds great,” says LFB. “Let’s play it - we all have to talk to somebody, or maybe multiple people, from the other group. We can measure our trust in each other before and after the game to see if it worked.”

After a bit of conferring, everyone agrees. The dataset dt contains the result. It has three columns, as follows:

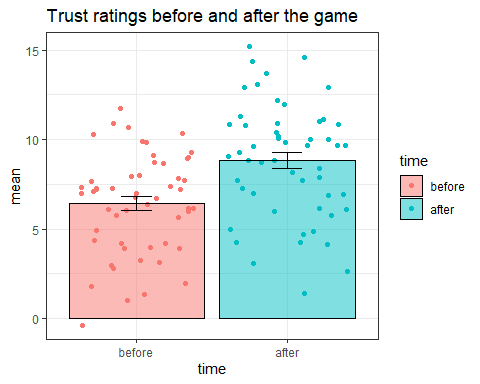
*person*: each of the people

*before*: that person’s answer to the following question before doing any trust game: “On a scale of 0 to 20, where 0 is ‘not at all’ and 20 is ‘completely’, how much do you generally trust the people from the other group?”

*after*: that person’s answer after doing the trust game with one or more people from the other group.

The figure below shows the data.

dt\_long <- dt %>%  
 pivot\_longer(cols=c(before,after),names\_to = "time", values\_to="answer")  
  
# reorder the factor levels so it makes sense in the plot  
dt\_long$time <- factor(dt\_long$time, levels = c("before","after"))  
  
dt\_sum <- dt\_long %>%  
 group\_by(time) %>%  
 summarise(mean = mean(answer),  
 sd = sd(answer),  
 n = n(),  
 sderr = sd/sqrt(n)) %>%  
 ungroup()  
  
dt\_sum %>%   
 ggplot(mapping=aes(x=time,y=mean,fill=time)) +  
 geom\_jitter(data=dt\_long,mapping=aes(x=time,y=answer,color=time)) +  
 geom\_col(alpha=0.5,colour="black") +  
 geom\_errorbar(mapping=aes(ymin=mean-sderr,ymax=mean+sderr),width=0.2) +  
 theme\_bw() +  
 labs(title = "Trust ratings before and after the game",ylab = "Trust rating (higher is better)")



It *looks* like there’s been a change, but is this different from one might have expected if there wasn’t really one? To test this we will want to use a paired-sample t-test.

1. If we assume the relevant assumptions are met, why is a paired-sample t-test appropriate in this case? What is the statistical null hypothesis and the statistical alternative hypothesis? (Assume a two-tailed test).

*ANSWER: We’re getting the same measure from each person at different times so the samples aren’t really independent of each other - instead they are paired, or repeated. We want to see if the difference between before and after is significant. Our statistical null hypothesis says that this difference is not different from zero; the alternative says that it is.*

1. Run the paired-sample t-test, and interpret the result.

# Two ways of running the t-test:  
t.test(answer ~ time, data=dt\_long, paired=TRUE)

##   
## Paired t-test  
##   
## data: answer by time  
## t = -9.6038, df = 52, p-value = 4.12e-13  
## alternative hypothesis: true mean difference is not equal to 0  
## 95 percent confidence interval:  
## -2.942521 -1.925404  
## sample estimates:  
## mean difference   
## -2.433962

t.test(dt$before, dt$after, paired=TRUE)

##   
## Paired t-test  
##   
## data: dt$before and dt$after  
## t = -9.6038, df = 52, p-value = 4.12e-13  
## alternative hypothesis: true mean difference is not equal to 0  
## 95 percent confidence interval:  
## -2.942521 -1.925404  
## sample estimates:  
## mean difference   
## -2.433962

*ANSWER: The test is significant, which suggests that the trust games were associated with a significant increase in trust.*