Math 152 - Statistical Theory - Homework 11

Ethan Ashby

Due: Tuesday, November 10, 2020, midnight PDT

1: PodQ Describe one thing you learned from someone in your pod this week (it could be: content, logistical help, background material, R information, etc.) 1-3 sentences.

Elena told me that some of the problems would go a lot more smoothly if I checked out some passages in the textbook. Reading those chapters made things SO much easier!

7: R - inflated errors An unethical experimenter desires to test the following hypotheses:

$H_0 : \theta = \theta_0$
$H_1 : \theta \neq \theta_0$

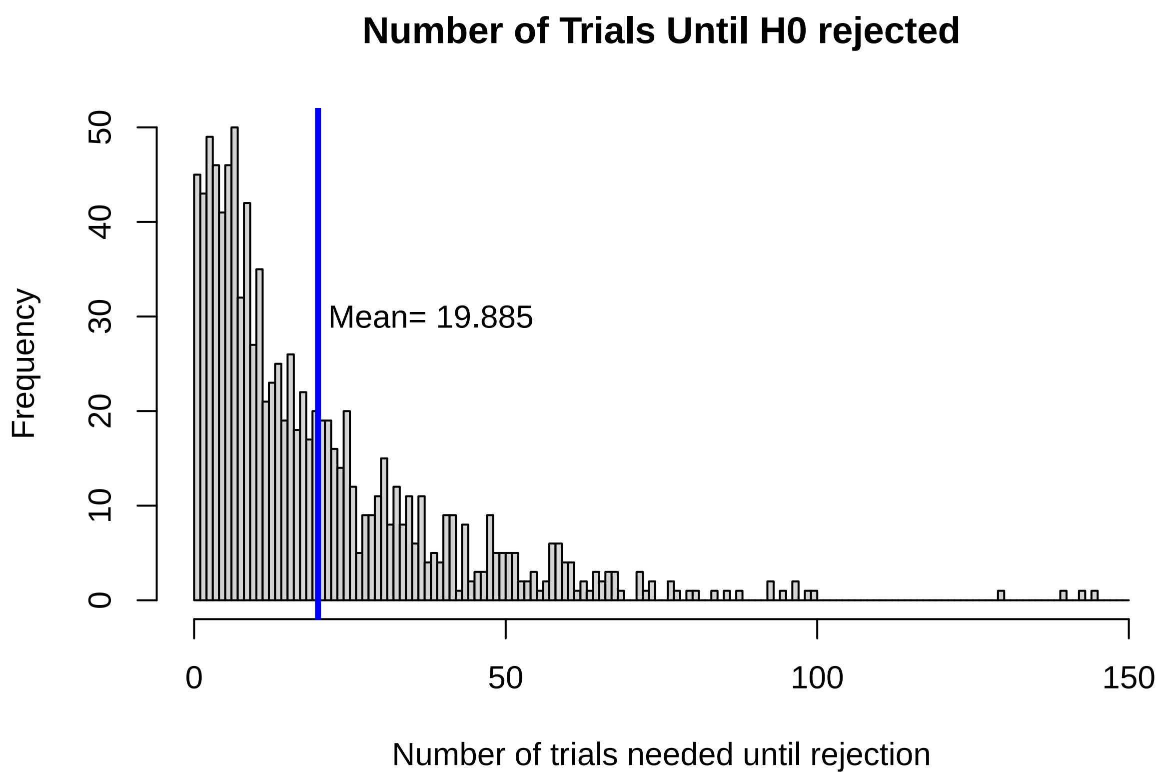
She draws a random sample $X_1, X_2, \ldots, X_n$ from a distribution with the pdf $f(x|\theta)$, and carries out a test of size α_0. If this test does not reject H_0, she discards the sample, draws a new independent random sample of n observations, and repeats the test based on the new sample. She continues drawing new independent samples in this way until she obtains a sample for which H_0 is rejected.

c. Do a simulation in R to corroborate your answer to b. Assume the X_i are normally distributed with mean μ unknown and variance 1. Use $\alpha_0 = 0.05$. Figure out a way to display your results so that they form part of a convincing argument. (If you tell me what you want to display, I'm happy to tell you the R code to display it.)

```r
library(tidyverse)
'%!in%' <- function(x,y)!('%in%'(x,y))
set.seed(11)

num_samps=100
alpha=0.05
mu_hidden=10 #the null mean (hidden from our CI construction)
vec_of_stats<-c()
for (i in 1:1000){
  vec<-c()
  while(FALSE %!in% vec){
    samp<-rnorm(num_samps, mu_hidden, sd=1)
    ub_ci=mean(samp)+qnorm(1-(alpha)/2)*(1/sqrt(num_samps))
    lb_ci=mean(samp)+qnorm(alpha/2)*(1/sqrt(num_samps))
    vec<-c(vec, between(mu_hidden, lb_ci, ub_ci))
  }
  vec_of_stats<-c(vec_of_stats, length(vec))
}

hist(vec_of_stats, breaks=seq(0,150,1), main="Number of Trials Until H0 rejected", xlab="Number of trial
abline(v=mean(vec_of_stats), col="blue", lwd=3)
text(38, 30, paste("Mean=", mean(vec_of_stats), sep=" "))
```

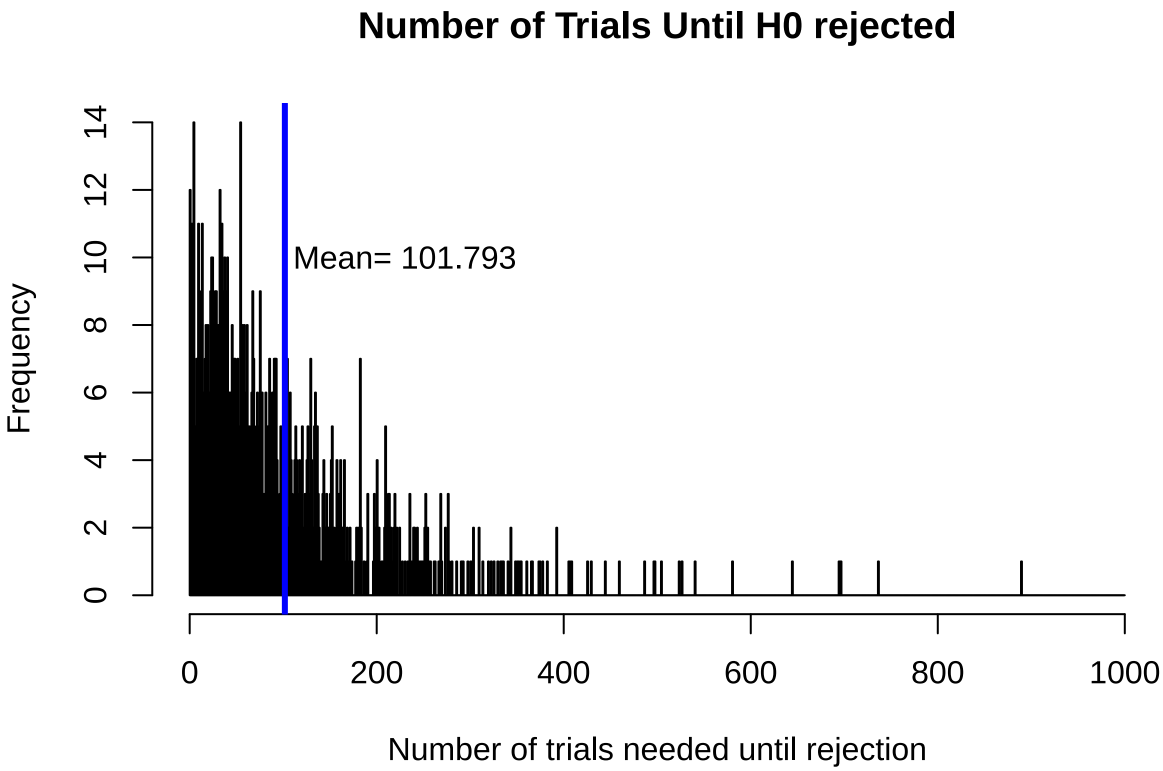


My simulation works by generating 100 random datapoints from a normal distribution with μ=10 and σ^2=1 (the data distribution under the null). Then I construct a normal confidence interval around the sample mean, and see if my CI captures μ. I repeat this procedure until my confidence interval fails to capture μ, i.e. meaning I reject the null. I stop and record how many trials it took to reject H_0. I repeat the previous 1000 times, generating 1000 statistics representing the number of trials needed to reject H_0. I initialized my procedure with $\alpha = 0.05$, and lo and behold, as I predicted, the mean number of trials needed to reject H_0 is 19.885, or really close to $\frac{1}{\alpha}$.

```r
set.seed(11)

num_samps=100
alpha=0.01
mu_hidden=10 #the null mean (hidden from our CI construction)
vec_of_stats<-c()
for (i in 1:1000){
  vec<-c()
  while(FALSE %!in% vec){
    samp<-rnorm(num_samps, mu_hidden, sd=1)
    ub_ci=mean(samp)+qnorm(1-(alpha)/2)*(1/sqrt(num_samps))
    lb_ci=mean(samp)+qnorm(alpha/2)*(1/sqrt(num_samps))
    vec<-c(vec, between(mu_hidden, lb_ci, ub_ci))
  }
  vec_of_stats<-c(vec_of_stats, length(vec))
}

hist(vec_of_stats, breaks=seq(0,1000,1), main="Number of Trials Until H0 rejected", xlab="Number of tria
abline(v=mean(vec_of_stats), col="blue", lwd=3)
text(230, 10, paste("Mean=", mean(vec_of_stats), sep=" "))
```



When I repeat the simulation with $\alpha = 0.01$, I get the same behavior, where the mean of my sampling distribution (i.e. the expected number of trials I need to run to get a rejection) is 101.8, which is very close to $\frac{1}{\alpha}$.