

## Does the treatment reduce pain for arthritis patients?

#1. Import the data into R from the file `arthritis.csv` (comma-separated file on Learn, import with function `read.csv()`).

```
# import data
art <- read.csv("arthritis.csv")
```

#2. Define a set of hypotheses to answer the research question.

Null Hypothesis: There is no difference between the population mean pain level of control and treatment.

Alternative Hypothesis: There is difference between the population mean pain level of control and treatment.

#3. Perform a t-test. Report the test statistic, degree of freedom, and the p-

```
# t-test ( $H_A: \mu_C - \mu_T > 0$ )
(ttest <- t.test(art$Control, art$Treatment, alternative="two.sided"))

##
## Welch Two Sample t-test
##
## data: art$Control and art$Treatment
## t = 1.1172, df = 21.975, p-value = 0.276
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -6.065995 20.232662
## sample estimates:
## mean of x mean of y
## 78.83333 71.75000
```

The test statistic is 1.1172. The degree of freedom is 21.975. The p-value is 0.276.

#4. Perform a permutation test (permutations) for the previously stated hypothesis.

```
CT <- c(art$Control, art$Treatment)
#x = treatment
#y = control
nx = length(art$Treatment)
ny = length(art$Control)

teststat = (mean(art$Control) - mean(art$Treatment)) / sqrt(var(art$Treatment)/nx
+ var(art$Control)/ny)
teststat

## [1] 1.117235
```

```

#---
set.seed(31878039)
N=10000
set.seed(31878039)
tstat = numeric(N)
for (i in 1:N) {
  sCT <- sample(CT, replace=F)
  ctrl <- sCT[1:12]
  trt <- sCT[-c(1:12)]
  tstat[i] = (mean(ctrl)-mean(trt))/sqrt(var(trt)/nx + var(ctrl)/ny)
}
pvalue = mean(tstat >= teststat)
pvalue

## [1] 0.1347

```

#5. How many unique permutations are possible?

```

# number of possible permutations
choose(24, 12)

## [1] 2704156

```

There are 2704156 possible unique permutations.

#6. Perform an exact permutation test using all possible permutations.

```

# exact permutation test
set.seed(31878039)
combi <- combn(24, 12)
tsstat <- numeric(length=ncol(combi))
for (i in 1:ncol(combi)){
  ctrl <- CT[combi[,i]]
  trt <- CT[-combi[,i]]
  tsstat[i] <- (mean(ctrl) - mean(trt)) / sqrt(var(trt)/nx + var(ctrl)/ny)
}
pvalexact <- mean(tsstat >= teststat)

alpha=0.05
tsstat = sort(tsstat)
tcrit = tsstat[ceiling((1-alpha)*N)]

print(c(mean(art$Control)-mean(art$Treatment), teststat, tcrit, pvalexact))

## [1] 7.0833333 1.1172347 -3.0722768 0.1389151

```

## Sample size estimation [5 marks]

#7. Write down the set of hypotheses for the new study.

Null Hypothesis: There is no difference between the population mean pain level of control and treatment.

Alternative Hypothesis: There is difference between the population mean pain level of control and treatment.

#8. Estimate the means and standard deviations given the data in the file `arthritis.csv` of the previous question.

```
# sample estimates
mC <- mean(art$Control)
mT <- mean(art$Treatment)
sdC <- sd(art$Control)
sdT <- sd(art$Treatment)
```

#9. Simulate the power for the t-test and permutation test ( permutations) for a sample of 24 () patients to reject the null hypothesis. Explain the results.

```
# define permutation test function
set.seed(31878039)
permtest <- function(x, y, B=1000){
  CT <- c(x, y)
  stat <- replicate(B, {
    sCT <- sample(CT, replace=FALSE)
    ctrl <- sCT[1:length(x)]
    trt <- sCT[-c(1:length(x))]
    (mean(ctrl) - mean(trt)) / sqrt(var(ctrl)/length(x) + var(trt)/length(y))
  })
  mean(stat >= t.test(x, y, alternative="two.sided")$statistic)
}

# simulate p-values
pvals <- replicate(1000, {
  x <- rnorm(12, mC, sdC)
  y <- rnorm(12, mT, sdT)
  tpval <- t.test(x, y, alternative="two.sided")$p.value
  ppval <- permtest(x, y, B=1000)
  c(tpval, ppval)
})
# estimate power
mean(pvals[1,] < 0.01)

## [1] 0.045

mean(pvals[2,] < 0.01)

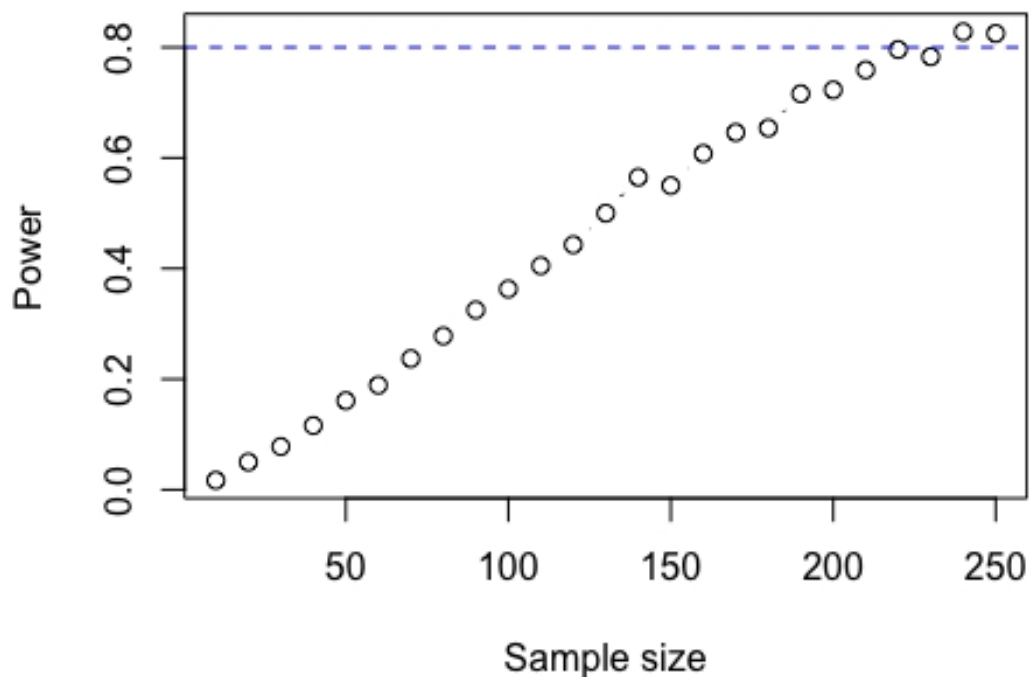
## [1] 0.088
```

Estimated power is about 45% of the time that the null hypothesis would be correctly rejected when using t-test.

Estimated power is about 88% of the time that the null hypothesis would be correctly rejected when using permutation test.

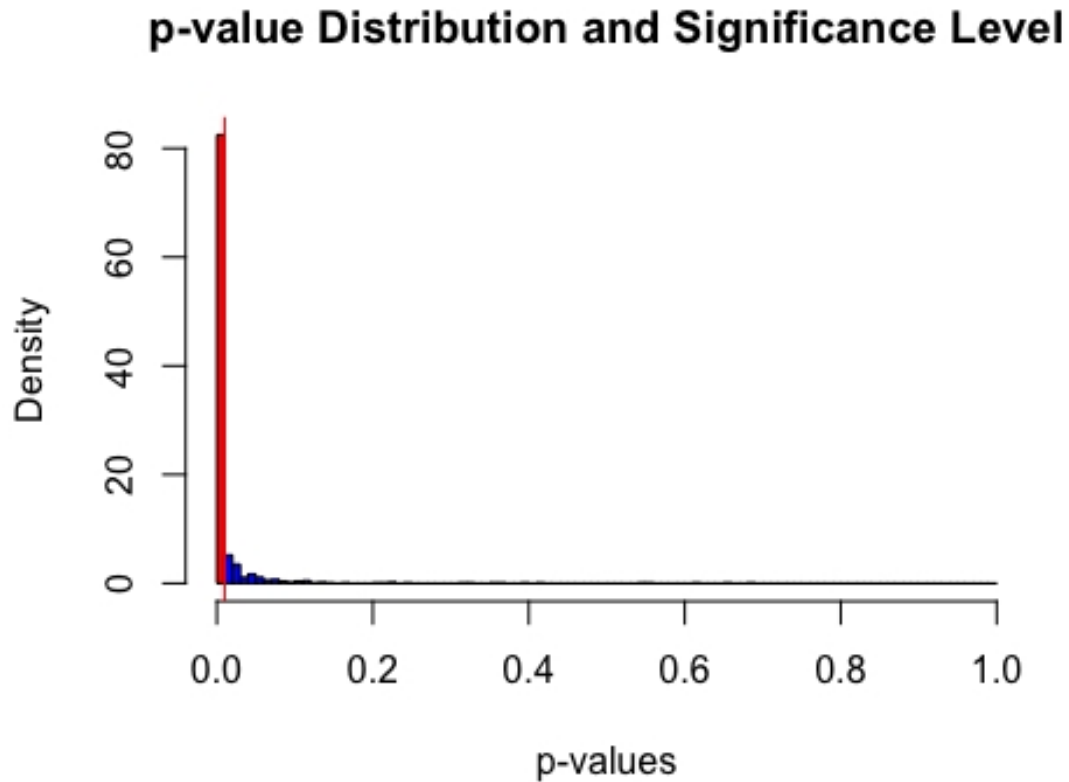
#10. Simulate the power only for the t-test for a range of sample sizes: , assuming equal sample sizes in the treatment and control group ().

```
# simulate t-test power for a range of sample sizes
set.seed(31878039)
nseq <- seq(10, 250, by=10)
power <- numeric(length=length(nseq))
for (i in 1:length(nseq)){
  n <- nseq[i]/2
  pvals <- replicate(1000, {
    x <- rnorm(n, mC, sdC)
    y <- rnorm(n, mT, sdT)
    t.test(x, y, alternative="two.sided")$p.value
  })
  # estimate power
  power[i] <- mean(pvals < 0.01)
}
plot(nseq, power, type="b", xlab="Sample size", ylab="Power")
abline(h=0.8, lty=2, col="blue2")
```



```
hbreaks = seq(0, 1.0, 0.01)
upperbreaks= hbreaks[-1]
hist(pvals, breaks=hbreaks, prob=T, col=ifelse(upperbreaks<=0.01,
```

```
"red","blue"),  
  xlab="p-values",main="p-value Distribution and Significance Level")  
abline(v=0.01,col="red")
```



#11. How many patients do we at least need approximately in our sample to reject the null hypothesis with a probability of 0.8?

```
# estimated sample size  
min(nseq[power > 0.8])  
## [1] 240
```

We would need at least 240 patients to reject the null hypothesis.