

```
library(ggplot2)
library(magrittr)
library(bootstrap)
```

```
# 1. bootstrap approach to sampling dists, conf int, hyp test
```

```
# a. data
```

```
protein <- c(12.06, 11.16, 11.35, 11.89, 12.49, 12.19, 11.89, 12.47, 12.42,
            11.57, 12.2, 11.04, 12.17, 12.82, 11.81, 11.86, 11.75, 11.82,
            12.17, 11.63, 11.54, 12.76, 12.2, 12.13, 12.08, 12.56, 12.77,
            13.12, 12.15, 12.07, 11.48, 11.61, 12.28, 12.38, 11.67, 11.67,
            11.55, 12.16, 12.92, 11.85, 12.53, 12.29, 12.06, 12.06, 12.01,
            12.81, 11.78, 11.66, 11.4, 12.33, 12.21, 11.93, 12.71, 11.65,
            12.32, 12.52, 11.84, 12.56, 13.72, 11.29)
```

```
# b. What is the point estimate for the standard deviation of protein content for the population of
# bags of this brand of flour?
```

```
# can't use below? don't know sampling dist of sample stv?
```

```
# "sample stv is point estimate for pop stv"
```

```
sd(protein)
#x <- sample(protein,replace = TRUE)
#sd(x)
```

```
# c. boots()
```

```
boots <- function(x, n){
  i <- sample(1:length(x), size = n, replace = TRUE)
  sd(x[i])
}
```

```
boots(protein,60)
```

```
# d. histogram
```

```
#replicate(10000,boots(protein,60)) %>%
```

```
# qplot( binwidth = 0.01)
```

```
h <- replicate(10000,boots(protein,60))
```

```
qplot(h, binwidth = 0.01)
```

```
#?replicate
```

```
# e. 95% bootstrap conf int using perentile method
```

```
#?quantile
```

```
#quantile(h)
```

```
#str(quantile(h,c(0.025, 0.975)))
```

```
#str(quantile(h,c(0.025, 0.975))[1])
```

```
#quantile(h,c(0.025, 0.975))[[1]]
```

```
quantile(h,c(0.025, 0.975))
```

```
# f. hyp test, sigma = 0.5
```

```
# There is insufficient evidence to support rejecting
```

```
# the null hypothesis that the population standard deviation
```

```
# is equal to 0.5, as a 95% confidence interval constructed
```

```
# via the bootstrap quantile method includes the value 0.5.
```

```
# 2. bootstrap package
```

```
?bootstrap
```

```
# a. repeat 1d
```

```
strap <- bootstrap(protein, nboot = 10000, theta = sd)
```

```

# b. str(strap)

str(strap)

qplot(strap$thetastar,binwidth = 0.01)


# c. quantile method
quantile(strap$thetastar,c(0.025, 0.975))

# There is insufficient evidence to support rejecting
# the null hypothesis that the population standard deviation
# is equal to 0.5, as a 95% confidence interval constructed
# via the bootstrap quantile method includes the value 0.5.


# 5. state park ranger
times <- c(3.6, 3.3, 1.2, 3.6, 2.3, 5.6, 3.4, 3.5, 3.1, 2.6)
my_data <- data.frame(
  orig_times = times
)
my_data$adj_times <- my_data$orig_times - 3.0
my_data$dist <- abs(my_data$adj_times)
#head(my_data)
#?rank
my_data$rank = rank(my_data$dist, ties.method = "a")
my_data$group = rep("P",length(my_data$orig_times))
my_data$group[my_data$adj_times < 0] = "N"
tn <- sum(my_data$rank[my_data$group == "N" ])
tp <- sum(my_data$rank[my_data$group == "P" ])
my_data
n = length(times)
Z <- (min(tn, tp) - (n*(n+1))/4) / sqrt( (n*(n+1)*(2*n +1))/24 )

```

```

#denom <- sqrt( (n*(n+1)*(2*n +1))/24 )
#Z <- Z/ denom
2*(1-pnorm(abs(Z)))

#X <- tp - (n*(n+1))/4
#X/denom
#2*(1-pnorm(abs(X/denom)))

#?wilcox.test
#wilcox.test(times, mu = 3.0, exact = FALSE, correct = FALSE)

# There is insufficient evidence to reject the null hypothesis
# that the average time spent in the state park is 3 hours
# (Wilcoxon signed-rank test, p-val: 0.48 ).

```