

Non-parametric Tests

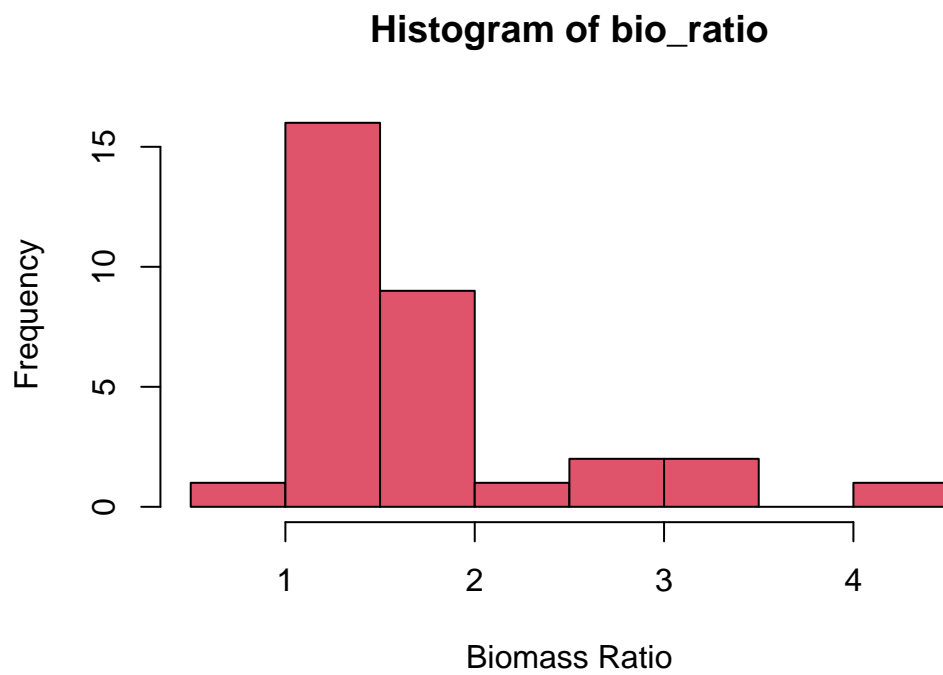
```
library(tidyverse)
library(BSDA) #for the sign test function
```

Checking Normality

Example of marine biomass ratio taken from 32 marine organisms. Biomass ratio is used to test the protection effect of biological reserves used to conserve marine species. It is the total mass of all marine plants and animals per unit area of reserve divided by the same quantity in the unprotected control. If the biomass ratio equals one, then the reserve has no protection had no effect.

```
# define data
bio_ratio = c(1.34, 1.96, 2.49, 1.27, 1.19, 1.15, 1.29, 1.05, 1.10, 1.21, 1.31,
              1.26, 1.38, 1.49, 1.84, 1.84, 3.06, 2.65, 4.25, 3.35, 2.55, 1.72,
              1.52, 1.49, 1.67, 1.78, 1.71, 1.88, 0.83, 1.16, 1.31, 1.40)

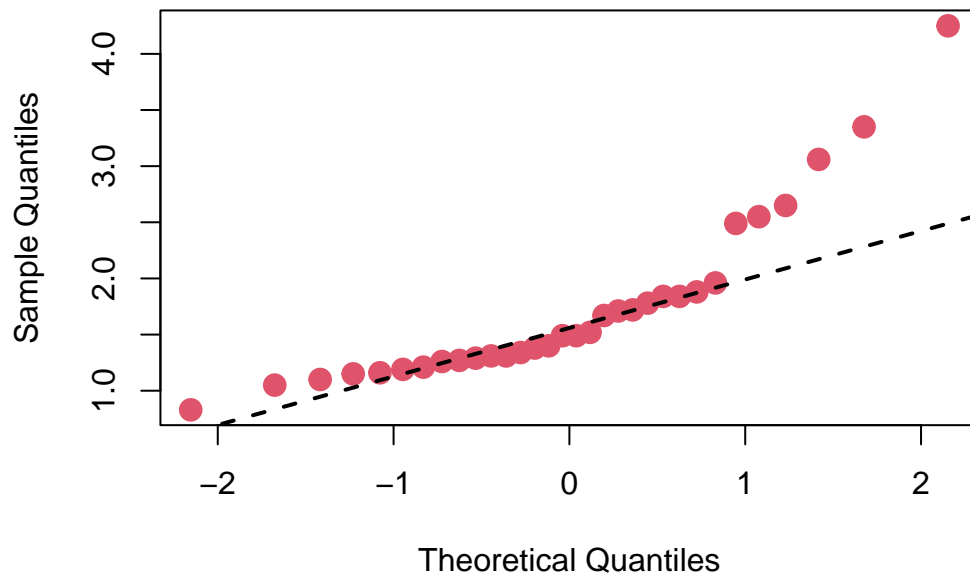
# Histogram - notice the severe right skew
hist(bio_ratio, xlab="Biomass Ratio", freq=T, col=2)
```



```
# Create a quantile-quantile plot (QQplot)
qqnorm(bio_ratio, col=2, pch=19, cex=1.5)

# Add a straight line which passes through the first and third quartiles.
qqline(bio_ratio, col = 1,lwd=2,lty=2)
```

Normal Q-Q Plot



```
# Perform Shapiro-Wilk test  
shapiro.test(bio_ratio)
```

Shapiro-Wilk normality test

```
data:  bio_ratio  
W = 0.81751, p-value = 8.851e-05
```

```
# reject the null, again evidence towards non-normality
```

Non-parametric Sign Test

This test uses the signs of the differences, but not the magnitudes.

H_0 : the median of the diff distribution is zero

Test the effect of linoleic acid supplementation on systolic blood pressure. A group of 17 adults with a diet high in linoleic acid were assessed at baseline and 4 weeks later.

Note: You will need to install the BSDA package in order to use the function SIGN.test.
`install.packages("BSDA")`

```
# define data
base_SBP = c(119.67, 100, 123.56, 109.89, 96.22, 133.33, 115.78, 126.39,
             122.78, 117.44, 111.33, 117.33, 120.67, 131.67, 92.39, 134.44,
             108.67)
post_SBP = c(117.33, 98.78, 123.83, 107.67, 95.67, 128.89, 113.22, 121.56,
             126.33, 110.39, 107, 108.44, 117, 126.89, 93.06, 126.67, 108.67)

# perform test
SIGN.test(base_SBP, post_SBP, md=0)
```

Dependent-samples Sign-Test

```
data:  base_SBP and post_SBP
S = 13, p-value = 0.02127
alternative hypothesis: true median difference is not equal to 0
95 percent confidence interval:
 0.5567996 4.7765495
sample estimates:
median of x-y
      2.56
```

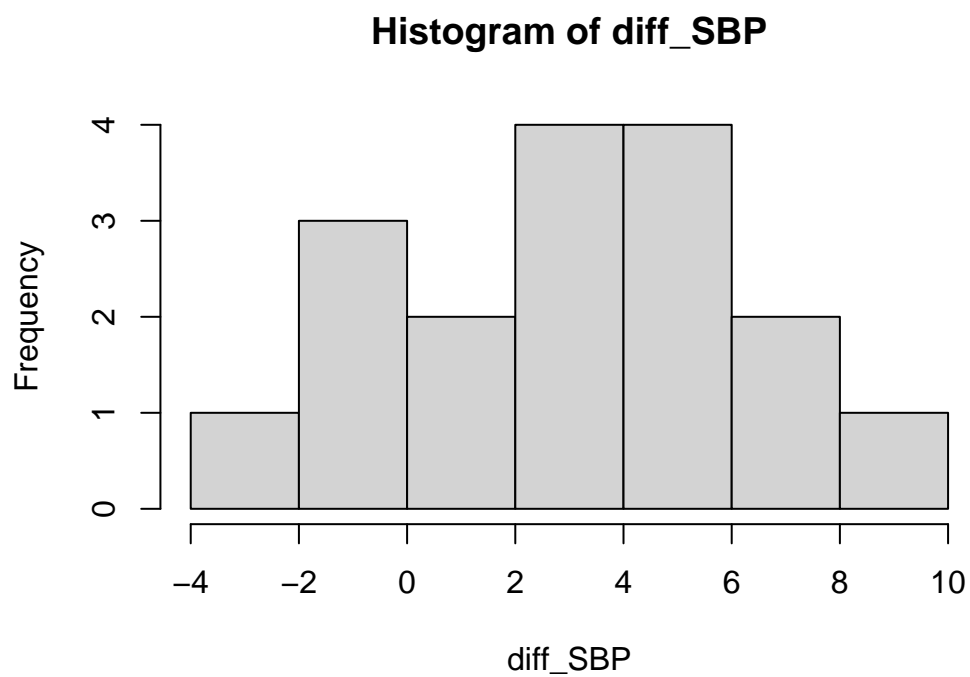
Achieved and Interpolated Confidence Intervals:

	Conf.Level	L.E.pt	U.E.pt
Lower Achieved CI	0.8565	1.2200	4.4400
Interpolated CI	0.9500	0.5568	4.7765
Upper Achieved CI	0.9510	0.5500	4.7800

Non-parametric Wilcoxon-Signed Rank Test

```
# Calculate the differences
diff_SBP = base_SBP-post_SBP

hist(diff_SBP)
```



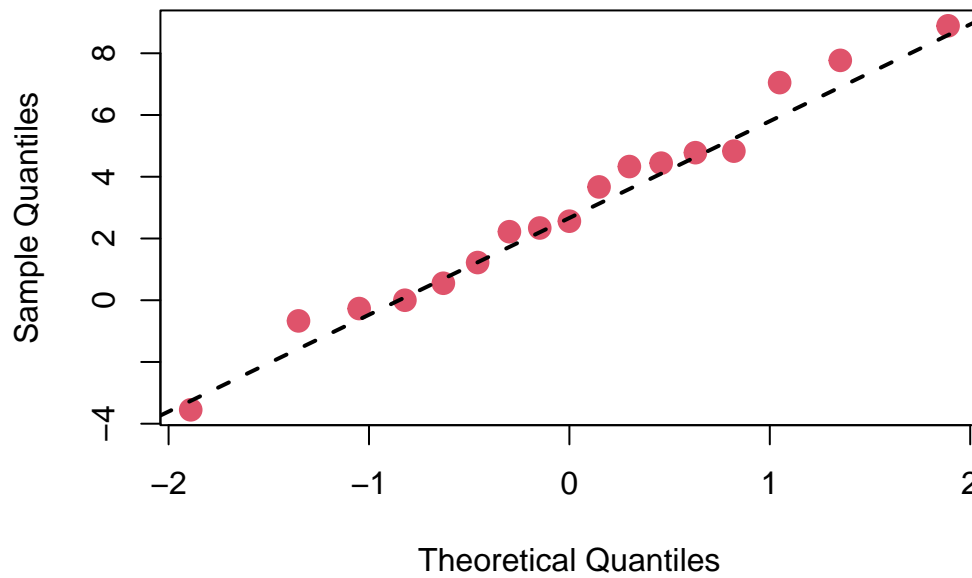
```
# perform test  
shapiro.test(diff_SBP)
```

Shapiro-Wilk normality test

```
data: diff_SBP  
W = 0.98228, p-value = 0.975
```

```
qqnorm(diff_SBP, col=2, pch=19, cex=1.5)  
qqline(diff_SBP, col = 1,lwd=2,lty=2)
```

Normal Q-Q Plot



```
# perform test  
wilcox.test(diff_SBP)
```

Warning in wilcox.test.default(diff_SBP): cannot compute exact p-value with zeroes

Wilcoxon signed rank test with continuity correction

```
data: diff_SBP  
V = 124, p-value = 0.004107  
alternative hypothesis: true location is not equal to 0
```

```
# equivalently...  
wilcox.test(base_SBP, post_SBP, paired=T)
```

Warning in wilcox.test.default(base_SBP, post_SBP, paired = T): cannot compute exact p-value with zeroes

Wilcoxon signed rank test with continuity correction

data: base_SBP and post_SBP

V = 124, p-value = 0.004107

alternative hypothesis: true location shift is not equal to 0

Non-parametric Wilcoxon-Rank Sum Test: Two Independent Groups

Compare the length of hospital stay for patients with the same diagnosis admitted at two different hospitals.

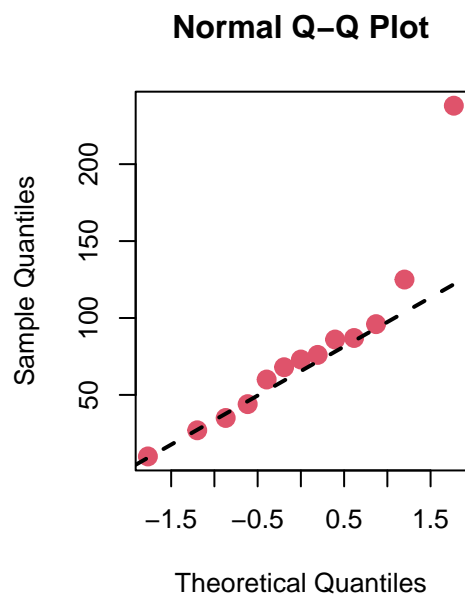
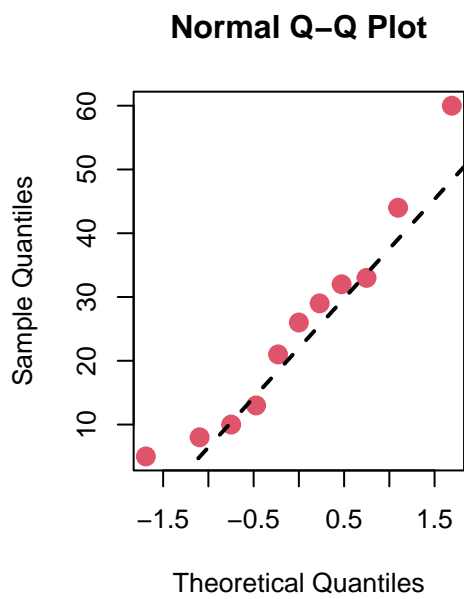
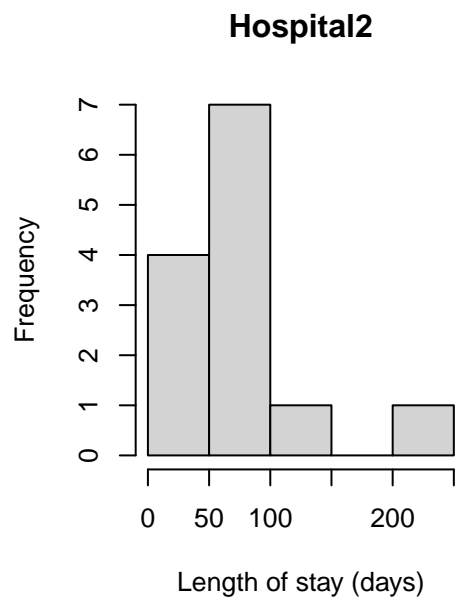
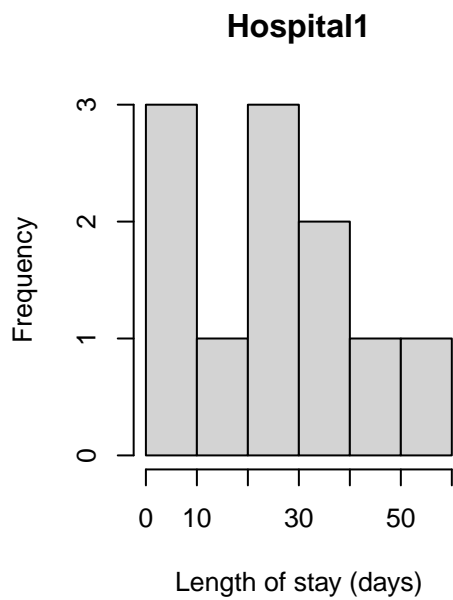
```
# define data
hosp1 = c(21,10,32,60,8,44,29,5,13,26,33)
hosp2 = c(86,27,10,68,87,76,125,60,35,73,96,44,238)

# check normality
par(mfrow=c(2,2))
hist(hosp1, xlab="Length of stay (days)", freq=T, main="Hospital1")

hist(hosp2, xlab="Length of stay (days)", freq=T, main="Hospital2")

qqnorm(hosp1, col=2, pch=19, cex=1.5)
qqline(hosp1, col = 1,lwd=2,lty=2)

qqnorm(hosp2, col=2, pch=19, cex=1.5)
qqline(hosp2, col = 1,lwd=2,lty=2)
```



```
# perform test
wilcox.test(hosp1, hosp2, mu=0)
```



```
Warning in wilcox.test.default(hosp1, hosp2, mu = 0): cannot compute exact p-  
value with ties
```

Wilcoxon rank sum test with continuity correction

```
data: hosp1 and hosp2  
W = 17.5, p-value = 0.001925  
alternative hypothesis: true location shift is not equal to 0
```

Notice that the test statistic T1 is different from what we got in the slides (83.5) Test statistic:
W = 17.5, p-value = 0.001925

Why is that? For Wilcoxon Rank Sum test only, R is calculating the statistic without $n1(n1+1)/2$ term.

There is a note in the function documentation: “The literature is not unanimous about the definitions of the Wilcoxon rank sum and Mann-Whitney tests. The two most common definitions correspond to the sum of the ranks of the first sample with the minimum value subtracted or not: R subtracts and S-PLUS does not, giving a value which is larger by $m(m+1)/2$ for a first sample of size m.”

If we want to get the same value of T1 per our formula in the slides, we need to add the $n1(n1+1)/2$ term.

```
res = wilcox.test(hosp1, hosp2, mu=0)
```

```
Warning in wilcox.test.default(hosp1, hosp2, mu = 0): cannot compute exact p-  
value with ties
```

```
res$statistic = res$statistic + 11*(11+1)/2  
  
res
```

Wilcoxon rank sum test with continuity correction

```
data: hosp1 and hosp2  
W = 83.5, p-value = 0.001925  
alternative hypothesis: true location shift is not equal to 0
```