

t-test and Wilcoxon (outliers)

Generation

Let us generate a sample with an outlier.

```
set.seed(1)
x <- rnorm(100, 0, 1)
xnew <- c(x, 100)
print(c("x", "mean=", mean(x), mean(xnew), "sd=", sd(x), sd(xnew)), quote=FALSE)
```

```
## [1] x          mean=          0.108887366914655 1.09790828407392
## [5] sd=          0.898199359660041 9.97963397631546
```

```
y <- rnorm(100, 1, 1)
ynew <- c(y, 100)
print(c("y", "mean=", mean(y), mean(ynew), "sd=", sd(y), sd(ynew)), quote=FALSE)
```

```
## [1] y          mean=          0.962191923434069 1.94276428062779
## [5] sd=          0.957879069588428 9.90061078448061
```

t-test

```
t.test(x, y)
```

```
##
##  Welch Two Sample t-test
##
## data:  x and y
## t = -6.4983, df = 197.19, p-value = 6.492e-10
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.1122615 -0.5943477
## sample estimates:
## mean of x mean of y
## 0.1088874 0.9621919
```

```
t.test(xnew, ynew)
```

```
##
##  Welch Two Sample t-test
##
## data:  xnew and ynew
## t = -0.60399, df = 199.99, p-value = 0.5465
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.603106 1.913394
## sample estimates:
## mean of x mean of y
## 1.097908 1.942764
```

```
wilcox.test(x, y)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  x and y
## W = 2635, p-value = 7.587e-09
## alternative hypothesis: true location shift is not equal to 0
```

```
wilcox.test(xnew, ynew)
```

```
##
##  Wilcoxon rank sum test with continuity correction
##
## data:  xnew and ynew
## W = 2735.5, p-value = 1.256e-08
## alternative hypothesis: true location shift is not equal to 0
```

Rank analogue of t-test