DATA2002

Wilcoxon signed-rank test

Garth Tarr

Wilcoxon signed-rank test

Normal approximation to the Wilcoxon signed-rank test statistic

Wilcoxon signed-rank test

What was wrong with the sign test?

- The sign test ignores a lot of information (inefficient use of data; low power).
- How can we use more information than just the sign for data with a symmetric, but possibly non-normal, distribution?

Put another way

- Suppose the sample $X_1, X_2, \ldots, X_n$ are drawn from a population symmetric with respect to mean μ (or median).
- We test the hypotheses: $H_0\colon\ \mu = \mu_0$ vs $H_1\colon\ \mu > \mu_0,\ \mu < \mu_0,\ \mu \neq \mu_0$.
- The t-test and Z-test assume a normal distribution without a long tail (outliers).
- They use all magnitude information from the normal curve.
- On the other hand, the sign test discards all data information on magnitude and hence it has low power.

Ranks to the rescue

Many non-parametric tests are based not on the data, but on their ranks.

To find the ranks for a set of data:

- Arrange the data in ascending order
- Assign a rank of 1 to the smallest observation, 2 to the second smallest, etc.
- For tied observations (in blue or red in the table below), assign each the average of the corresponding ranks

Sample	**8**	**5**	**10**	**2**	**5**	**8**	**8**	**6**
Ordered sample	2	5	5	6	8	8	8	10
Successive ranks	1	2	3	4	5	6	7	8
Assigned ranks	1	2.5	2.5	4	6	6	6	8

Thinking about magnitude

- Under the symmetric distribution assumption with mean μ_0 from H_0, half of the $d_i = x_i - \mu_0$ should be negative and half positive and the expected counts are both $n/2$.
- Under the null hypothesis, the positive and negative d_i should be of similar magnitude and occur with equal probability (on average).
- If we rank the absolute values of d_i in ascending order, the expected rank sums for the negative and positive d_i should be nearly equal.

Wilcoxon signed-rank test

We need to define the following quantities:

- $D_i = X_i - \mu_0$ for $i = 1, 2, \ldots, n$
- $R_1, \ldots, R_n$ be the ranks of $|D_1|, |D_2|, \ldots, |D_n|$
- W^+ be the sum of the ranks R_i corresponding to positive D_i
- W^- be the sum of the ranks R_i corresponding to negative D_i
- Let $W = \min(W^+, W^-)$

Calculations of w^+ and w

When we observe the data we have $d_i = x_i - \mu_0$ with ranks (of the absolute values), $r_1, \ldots, r_n$ for $|d_1|, \ldots, |d_n|$.

$$w^+ = \sum_{i\colon d_i > 0} r_i \qquad \text{and} \qquad w^- = \sum_{i\colon d_i < 0} r_i.$$

We should

- reject $H_0\colon\ \mu = \mu_0$ in favour of $H_1\colon\ \mu > \mu_0$ if w^+ is large enough
- reject $H_0\colon\ \mu = \mu_0$ in favour of $H_1\colon\ \mu < \mu_0$ if w^+ is small enough
- reject $H_0\colon\ \mu = \mu_0$ in favour of $H_1\colon\ \mu \neq \mu_0$ if $w = \min(w^+, w^-)$ is small enough

Workflow

Suppose $X_i, \ldots, X_n$ are drawn from some population that follows a symmetric distribution. Given a significance level α, we want to test on the population mean, μ.

- **Hypothesis:** $H_0\colon\ \mu = \mu_0$ vs $H_1\colon\ \mu > \mu_0,\ \mu < \mu_0,\ \mu \neq \mu_0$
- **Assumptions:** X_i are independently sampled from a symmetric distribution.
- **Test statistic:** $W^+ = \sum_{i\colon D_i > 0} R_i$ for one-sided or $W = \min(W^+, W^-)$ for two-sided
- **Observed test statistic:** w^+ for one-sided or $w = \min(w^+, w^-)$ for two-sided
- **p-value:**
  - $P(W^+ \geq w^+)$ for $H_1\colon\ \mu > \mu_0$
  - $P(W^+ \leq w^+)$ for $H_1\colon\ \mu < \mu_0$
  - $2P(W^+ \leq w)$ for $H_1\colon\ \mu \neq \mu_0$
- **Decision:** If the p-value is less than α, there is evidence against H_0. If p-value is greater than α, the data are consistent with H_0.

Calculation of p-value: no ties

- Calculate the difference sample $|d_1|, \ldots, |d_n|$ where $d_i = x_i - \mu_0$.
- Calculate the observed sum of the positive ranks:

$$w^+ = \sum_{i:\, d_i>0} r_i$$

- The *exact* p-value $P(W^+ \geq w^+)$ for w^+ is

$$P(W^+ \geq w^+) = P(W^+ \leq n(n+1)/2 - w^+)$$

Notes:

- $W^+ + W^- = 1 + 2 + \ldots + n = n(n+1)\frac{1}{2} \Rightarrow W^- = n(1+n)\frac{1}{2} - W^+$
- Hence, under the null hypothesis, $\mathrm{E}(W^+) = n(1+n)\frac{1}{4}$
- Can also show[2] (if there are no ties) that $\mathrm{Var}(W^+) = n(n+1)(2n+1)/24$

[1] W Summing consecutive integers

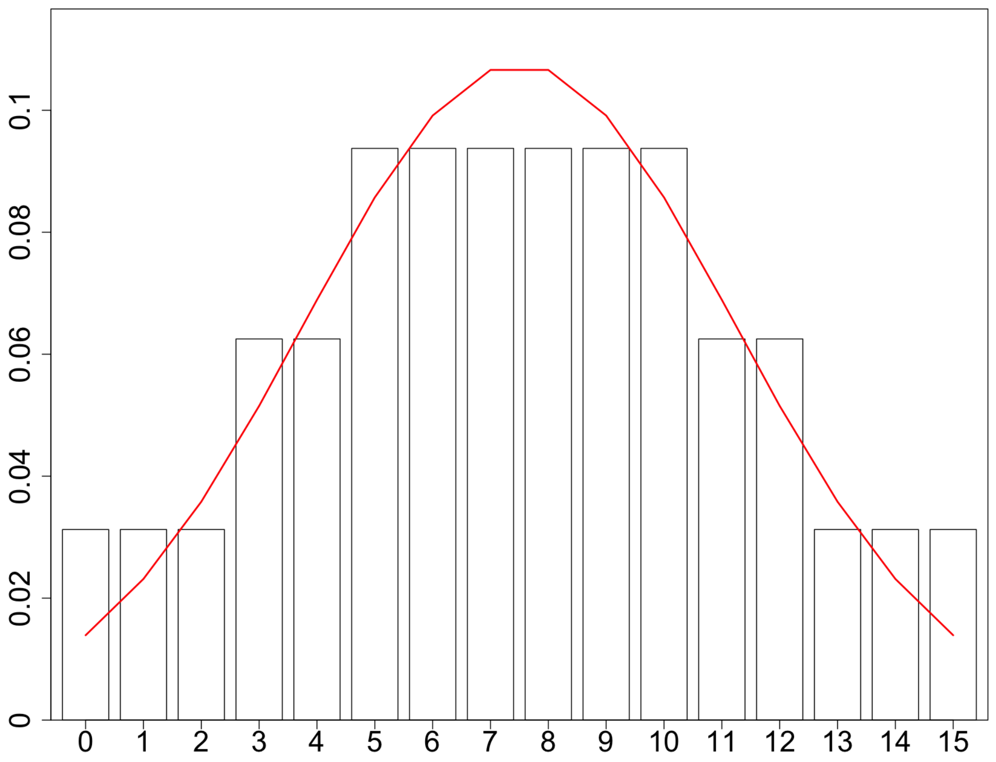
[2] See Larsen and Marx (2012; Theorem 14.3.2)

Calculating the p-value

We can use the `dsignrank()` function to inspect the distribution of the test statistic the Wilcoxon signed-rank test.

```
n = 5 # sample size
# possible values for the sum of
# the positive ranks
q = 0:(n * (n + 1)/2)
probs = dsignrank(q, n)
names(probs) = q
mu = n * (n + 1)/4
s2 = n * (n + 1) * (2 * n + 1)/24
```

```
library(plotrix)
plotrix::barp(dsignrank(q,n), names.arg = q,
              ylim=c(0,0.11))
lines(dnorm(q, mu, sqrt(s2)),
      col = "red", lwd = 2)
```



The **red line** is a normal distribution curve using the mean and variance from the previous slide.

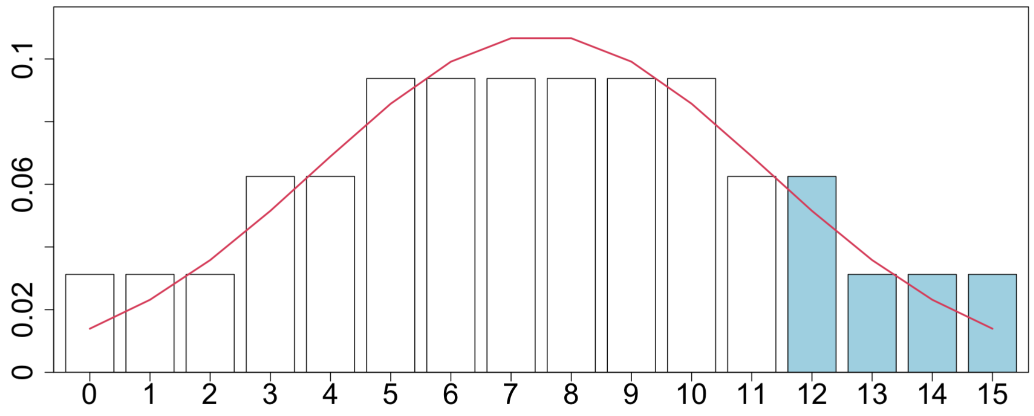
Calculating the p-value

In a sample of size $n = 5$ we observe $w^+ = 12$.

$$P(W^+ \geq 12)$$

```r
c(psignrank(12 - 1, n, lower.tail = FALSE),
  1 - psignrank(12 - 1, n),
  sum(dsignrank(12:15, n)))
```

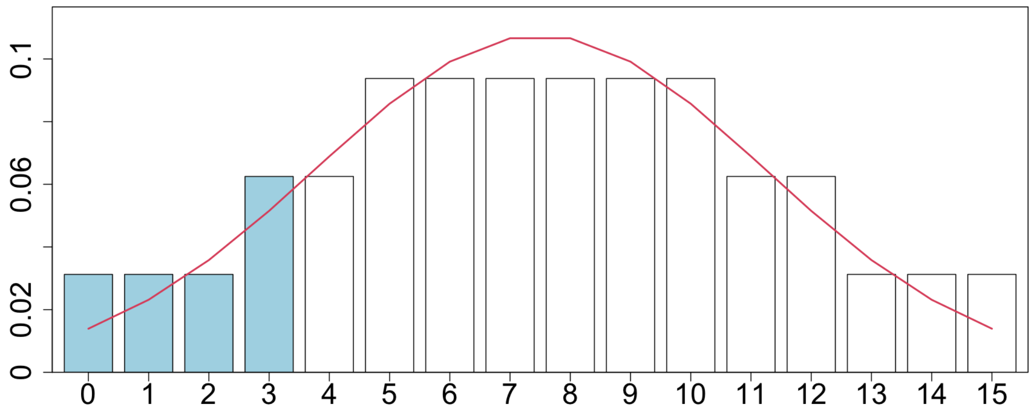
```
## [1] 0.15625 0.15625 0.15625
```



$$P(W^+ \geq 12) = P(W^+ \leq 15 - 12) = P(W^+ \leq 3)$$

```r
dsignrank(0:3, n)
c(psignrank(3, n), sum(dsignrank(0:3, n)))
```

```
## [1] 0.03125 0.03125 0.03125 0.06250
## [1] 0.15625 0.15625
```



Weight gain

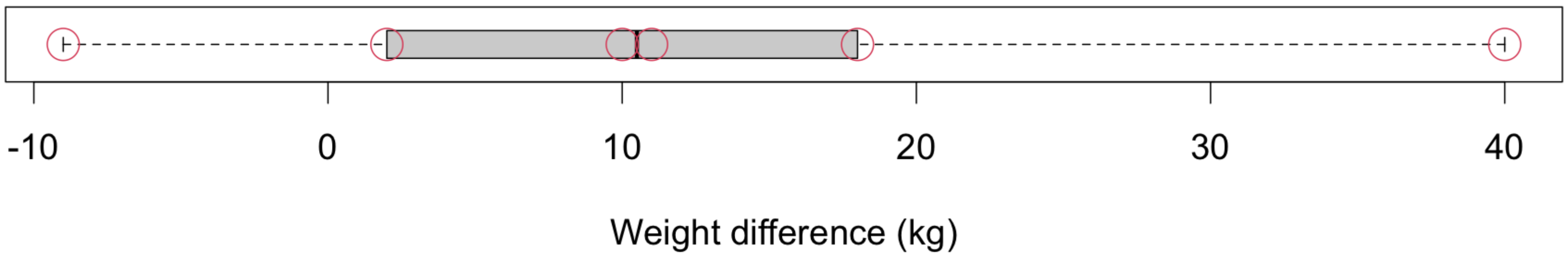
Weights of six pairs of twins on diets Y and X are

```
y = c(85, 69, 81, 112, 77, 86)
x = c(83, 78, 70, 72, 67, 68)
d = y-x
d
```

```
## [1]  2 -9 11 40 10 18
```

Is there a weight gain in taking diet Y as compared with diet X?

```
boxplot(d, xlab = "Weight difference (kg)", horizontal = TRUE)
points(d, rep(1, length(d)), col = 2, cex = 2)
```



The **tidy** way

```r
w_calc = data.frame(
  dif = d,
  absDif = abs(d),
  rankAbsDif = rank(abs(d)),
  signrank = sign(d)*rank(abs(d))
  )
w_calc
```

```
##   dif absDif rankAbsDif signrank
## 1   2      2          1        1
## 2  -9      9          2       -2
## 3  11     11          4        4
## 4  40     40          6        6
## 5  10     10          3        3
## 6  18     18          5        5
```

```r
w_calc %>%
  filter(signrank>0) %>%
  summarise(sum(signrank)) %>%
  pull()
```

```
## [1] 19
```

The **base** way

```r
rbind(
  dif = d,
  absDif = abs(d),
  rankAbsDif = rank(abs(d)),
  signrank = sign(d)*rank(abs(d))
  )
```

```
##            [,1] [,2] [,3] [,4] [,5] [,6]
## dif           2   -9   11   40   10   18
## absDif        2    9   11   40   10   18
## rankAbsDif    1    2    4    6    3    5
## signrank      1   -2    4    6    3    5
```

```r
signrank = sign(d)*rank(abs(d))
sum(signrank[signrank>0])
```

```
## [1] 19
```

- **Hypothesis:** $H_0\colon\ \mu_d = 0$ vs $H_1\colon\ \mu_d > 0$
- **Assumptions:** D_i are independently sampled from a symmetric distribution.
- **Test statistic:** $W^+ = \sum_{i:D_i>0} R_i$ where R_i are the ranks of $|D_1|, |D_2|, \dots, |D_n|$. Under H_0, $W \sim \mathrm{WSR}(n)$.
- **Observed test statistic:** $w^+ = 1 + 4 + 6 + 3 + 5 = 19$
- **p-value:**

$$\begin{aligned} P(W^+ \geq w^+) &= P(W^+ \geq 19) \\ &= P(W^+ \leq 6(6+1)/2 - 19) \\ &= P(W^+ \leq 2) = \texttt{psignrank(2, 6)} \\ &= 0.047. \end{aligned}$$

- **Decision:** The p-value is (just) less than 0.05, therefore there is some evidence against the null hypothesis that the diets are are equally effective and we conclude that diet Y does appear to be associated with higher weight gain than diet X.

```r
psignrank(2,6)
```

```
## [1] 0.046875
```

```r
wilcox.test(d, alternative = "greater")
```

```
##
##     Wilcoxon signed rank exact test
##
## data:  d
## V = 19, p-value = 0.04688
## alternative hypothesis: true location is greater than 0
```

```r
wilcox.test(y, x, alternative = "greater", paired = TRUE)
```

```
##
##     Wilcoxon signed rank exact test
##
## data:  y and x
## V = 19, p-value = 0.04688
## alternative hypothesis: true location shift is greater than 0
```

Compare with the alternative approaches

Paired t-test

```r
t.test(d, alternative = "greater")
```

```
##
##     One Sample t-test
##
## data:  d
## t = 1.7783, df = 5, p-value = 0.06774
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  -1.597222       Inf
## sample estimates:
## mean of x
##        12
```

Sign test

```r
c(sum(d > 0), sum(d != 0))
```

```
## [1] 5 6
```

```r
binom.test(c(5,1), p = 0.5,
           alternative = "greater")
```

```
##
##     Exact binomial test
##
## data:  c(5, 1)
## number of successes = 5, number of trials = 6,
## p-value = 0.1094
## alternative hypothesis: true probability of succe
## 95 percent confidence interval:
##  0.4181966 1.0000000
## sample estimates:
## probability of success
##              0.8333333
```

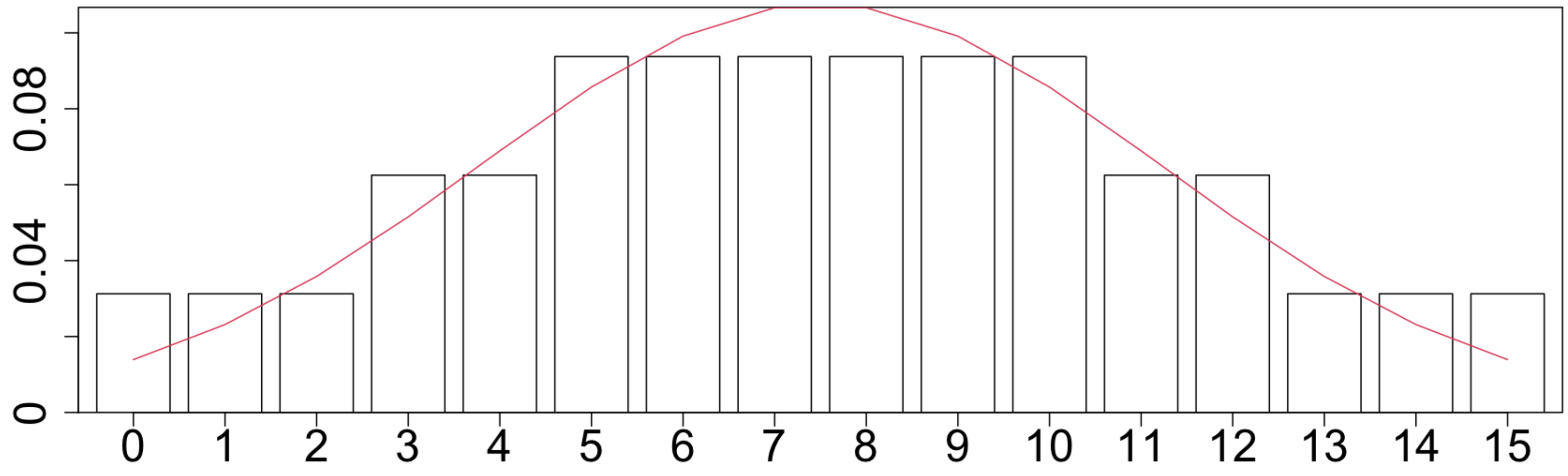
Normal approximation to the Wilcoxon signed-rank test statistic

Normal approximation $n = 5$

```r
n = 5
mu = n*(n+1)/4; s2 = n*(n+1)*(2*n+1)/24; q = 0:(n*(n+1)/2)
c(mu, s2, max(q))
```

```
## [1]  7.50 13.75 15.00
```

```r
plotrix::barp(dsignrank(q,n),names.arg = q,ylim=c(0,max(dnorm(q,mu,sqrt(s2)))+.0001), cex = 2)
points(dnorm(q,mu,sqrt(s2)),col = 2,type = 'l')
```

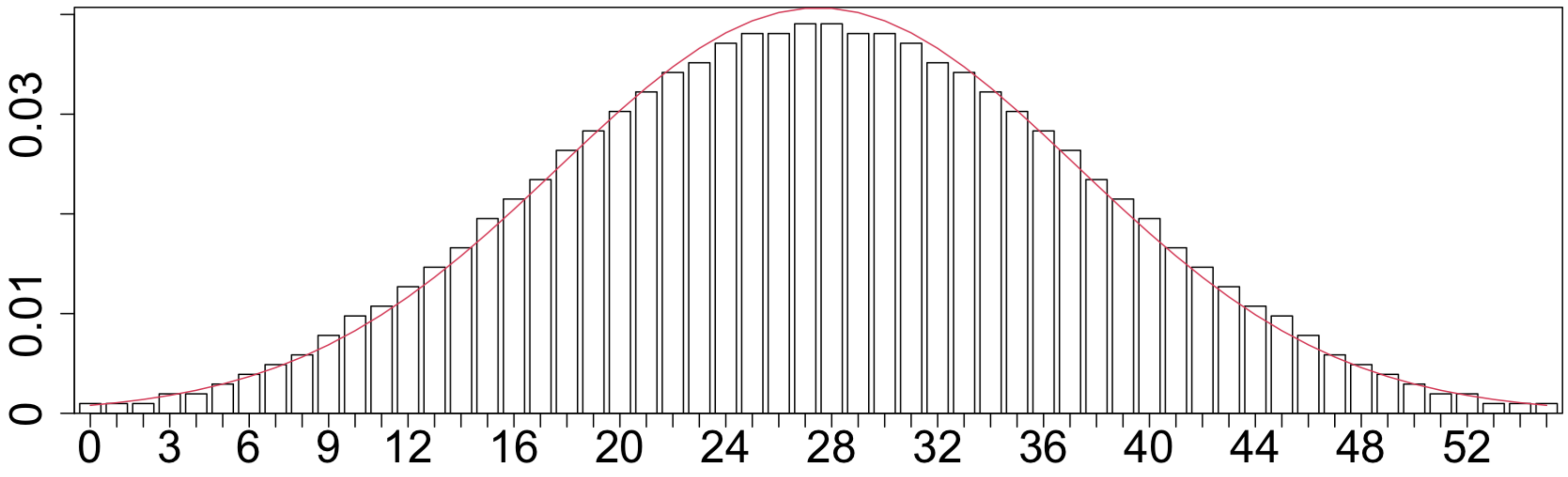


Normal approximation $n = 10$

```
n = 10
mu = n*(n+1)/4; s2 = n*(n+1)*(2*n+1)/24; q = 0:(n*(n+1)/2)
c(mu, s2, max(q))
```

```
## [1] 27.50 96.25 55.00
```

```
plotrix::barp(dsignrank(q,n),names.arg = q,ylim=c(0,max(dnorm(q,mu,sqrt(s2)))+.0001), cex = 2)
points(dnorm(q,mu,sqrt(s2)),col = 2,type = 'l')
```

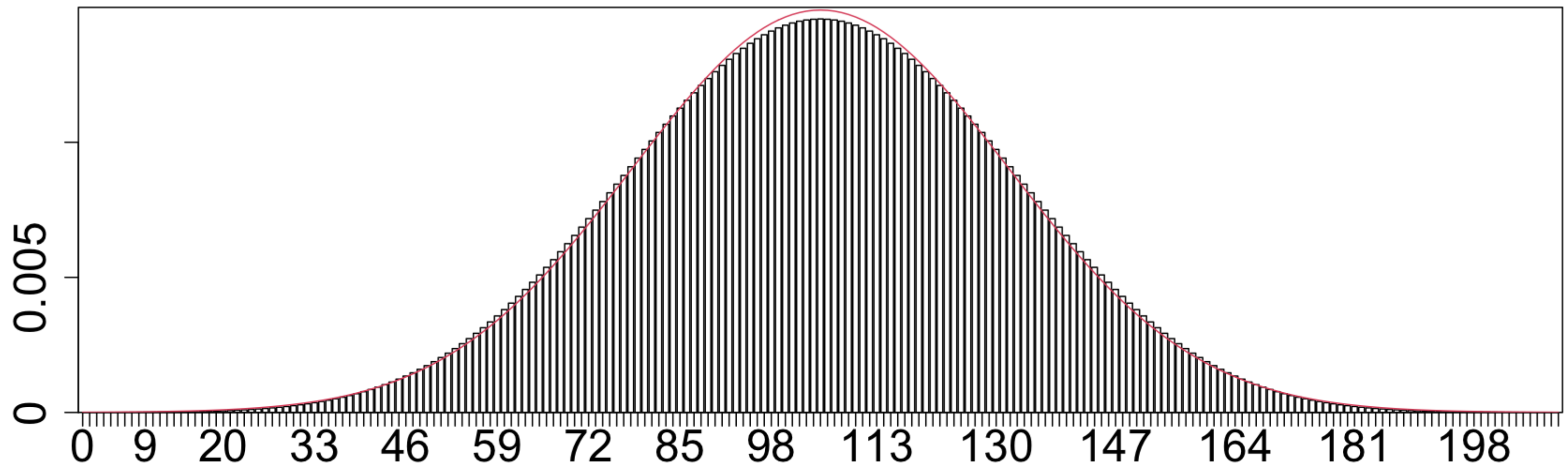


Normal approximation $n = 20$

```
n = 20
mu = n*(n+1)/4; s2 = n*(n+1)*(2*n+1)/24; q = 0:(n*(n+1)/2)
c(mu, s2, max(q))
```

```
## [1] 105.0 717.5 210.0
```

```
plotrix::barp(dsignrank(q,n),names.arg = q,ylim=c(0,max(dnorm(q,mu,sqrt(s2)))+.0001), cex = 2)
points(dnorm(q,mu,sqrt(s2)),col = 2,type = 'l')
```



Normal approximation

For **large enough** n, we can use a normal distribution to approximate the distribution of the Wilcoxon sign rank test statistic.

I.e. in large samples (without ties),

$$W^+ \sim N\left(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24}\right), \quad \text{approximately.}$$

Hence the large sample test statistic is,

$$T = \frac{W^+ - \mathrm{E}(W^+)}{\sqrt{\mathrm{Var}(W^+)}} \sim N(0,1),$$

where $\mathrm{E}(W^+) = \dfrac{n(n+1)}{4}$ and $\mathrm{Var}(W^+) = \dfrac{n(n+1)(2n+1)}{24}$.

Bus waiting times

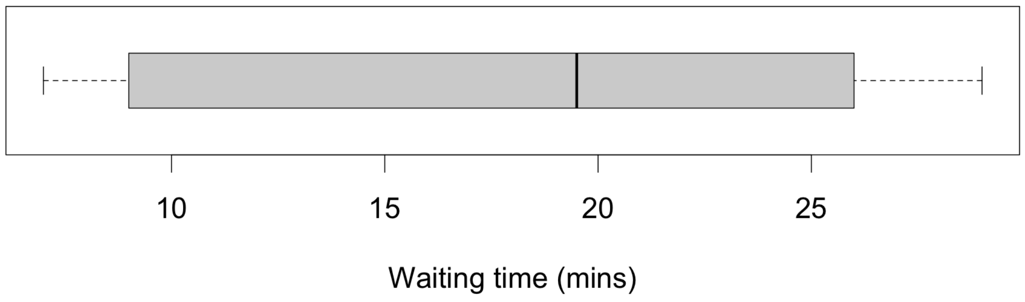
The following data are waiting times for the 370 bus in minutes for 10 randomly selected passengers:

```r
bus = c(25, 19, 9, 27, 8, 7, 26, 12, 29, 20)
```

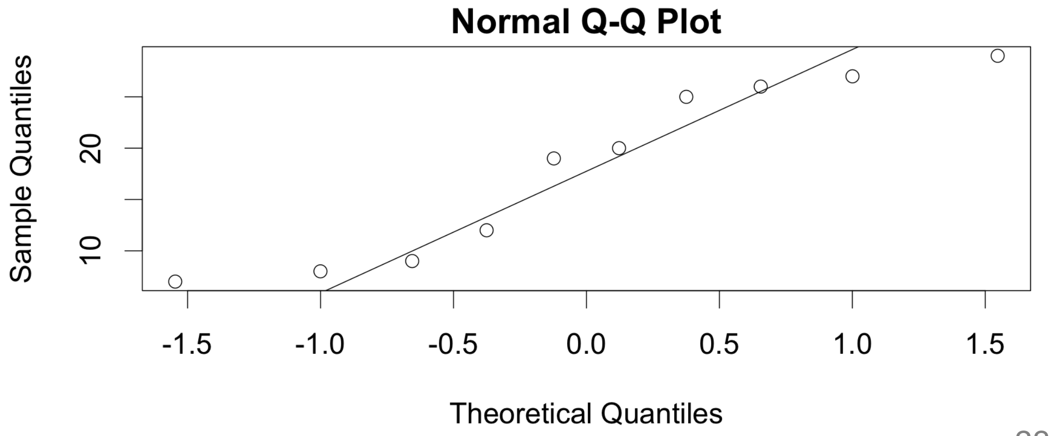
The bus authority claims a typical wait time of 15 minutes. Do these data suggest a different typical wait time?

The standard approach is a one-sample t-test to test $H_0\colon\ \mu = 15$.

```r
boxplot(bus, horizontal = TRUE,
        xlab = "Waiting time (mins)")
```




```r
qqnorm(bus); qqline(bus)
```



- **Hypothesis:** $H_0\colon\ \mu = 15$ vs $H_1\colon\ \mu \neq 15$
- **Assumptions:** X_i are independently sampled from a symmetric distribution.
- **Test statistic:** $W = \min(W^+, W^-)$ where $W^+ = \sum_{i:D_i>0} R_i$, $W^- = \sum_{i:D_i<0} R_i$, $D_i = X_i - 15$ and R_i are the ranks of $|D_1|, |D_2|, \ldots, |D_n|$. Under H_0, $W^+ \sim \text{WSR}(10)$, a symmetric distribution with mean $\text{E}(W^+) = \dfrac{n(n+1)}{4} = 27.5$ and $\text{Var}(W^+) = \dfrac{n(n+1)(2n+1)}{24} = 96.25$.
- **Observed test statistic:** found by
  - Determine difference sample $D_i = X_i - \mu_0$
  - Assign the signed ranks of D_i
  - Calculate w^+, the sum of the positive ranks and w^-, the sum of the negative ranks.
  - We have a two sided alternative, so the observed test statistic is $w = \min(w^+, w^-)$

Test statistic

X_i	$D_i = X_i - 15$	**Sign**	$\|D\|$	**Rank**	**Signed rank**
25	10	+	10	7	7
19	4	+	4	2	2
9	-6	-	6	4	-4
27	12	+	12	9	9
8	-7	-	7	5	-5
7	-8	-	8	6	-6
26	11	+	11	8	8
12	-3	-	3	1	-1
29	14	+	14	10	10
20	5	+	5	3	3

$w_+ = 7 + 2 + 8 + 9 + 10 + 3 = 39$
$w_- = |-4 + -5 + -6 + -1| = 16$

Test statistic: $w = \min(w^+, w^-) = 16$

If H_0 is true, W^+ comes from a symmetric distribution with mean,

$$\mathrm{E}(W^+) = \frac{n(n+1)}{4} = \frac{10 \times 11}{4} = 27.5$$

and variance,

$$\mathrm{Var}(W^+) = \frac{n(n+1)(2n+1)}{24} = 96.25$$

Observed test statistic:

$$t_0 = \frac{w - \mathrm{E}(W^+)}{\sqrt{\mathrm{Var}(W^+)}} = \frac{16 - 27.5}{\sqrt{96.25}} = -1.172$$

We have a test statistic $t_0 = -1.172$ so the approximate p-value is

$$\begin{aligned} 2P(W^+ \leq 16) &\approx 2P\left(Z \leq \frac{16 - \mathrm{E}(W^+)}{\sqrt{\mathrm{Var}(W^+)}}\right) \\ &= 2P\left(Z \leq \frac{16 - 27.5}{\sqrt{96.25}}\right) \\ &= 2P(Z \leq -1.172) \\ &= \texttt{2*pnorm(-1.172)} \\ &= 0.241 \end{aligned}$$

Because the p-value is large, there is no evidence to reject H_0. Therefore there is no evidence to dispute the bus authority's claim of a typical wait time of 15 minutes.

Compare to the exact p-value:

```
2*psignrank(16, 10)
```

```
## [1] 0.2753906
```

```
wilcox.test(bus - 15)
```

```
##
##      Wilcoxon signed rank exact test
##
## data:  bus - 15
## V = 39, p-value = 0.2754
## alternative hypothesis: true location is not equa
```

We have a similarly large p-value using the exact distribution. I.e. the approximation is working well even when $n = 10$.

Normal approximation with ties

As we've seen, we can approximate W^+ by a normal distribution, *NOT the data* X_i.

The p-value is approximately given by

$$\text{p-value} \approx P\left(Z \geq \frac{w^+ - \mathrm{E}(W^+)}{\sqrt{\mathrm{Var}(W^+)}}\right) \quad \text{for } H_1\colon\ \mu > \mu_0$$

$$\text{p-value} \approx P\left(Z \leq \frac{w^+ - \mathrm{E}(W^+)}{\sqrt{\mathrm{Var}(W^+)}}\right) \quad \text{for } H_1\colon\ \mu < \mu_0$$

$$\text{p-value} \approx 2P\left(Z \geq \left|\frac{w^+ - \mathrm{E}(W^+)}{\sqrt{\mathrm{Var}(W^+)}}\right|\right) \quad \text{for } H_1\colon\ \mu \neq \mu_0.$$

where **in general**,

$$\mathrm{E}(W^+) = \frac{1}{2}\sum_{i\colon d_i \neq 0} r_i \quad \text{and} \quad \mathrm{Var}(W^+) = \frac{1}{4}\sum_{i\colon d_i \neq 0} r_i^2$$

When there are no ties and zeros this simplifies to the expression we saw earlier: $\mathrm{E}(W^+) = n(n+1)/4$ and $\mathrm{Var}(W^+) = n(n+1)(2n+1)/24$

Smoking

Blood samples from 11 individuals before and after they smoked a cigarette are used to measure aggregation of blood platelets.

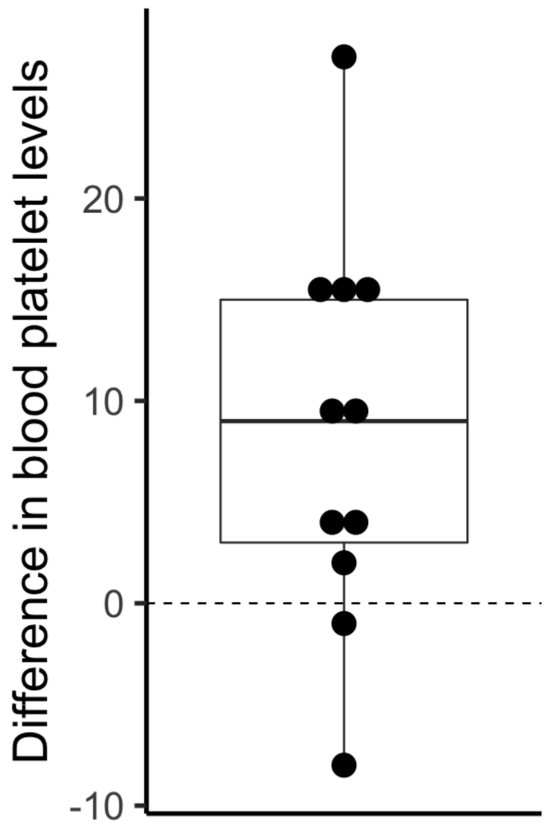
```
before = c(25, 25, 27, 44, 30, 67, 53, 53, 52,
after =  c(27, 29, 37, 36, 46, 82, 57, 80, 61,
df = data.frame(before, after,
  difference = after-before)
df
```

```
##    before after difference
## 1      25    27          2
## 2      25    29          4
## 3      27    37         10
## 4      44    36         -8
## 5      30    46         16
## 6      67    82         15
## 7      53    57          4
## 8      53    80         27
## 9      52    61          9
## 10     60    59         -1
## 11     28    43         15
```

Is the aggregation affected by smoking?

```r
library(ggplot2)
p = ggplot(df, aes(x="", y=difference)) +
  geom_boxplot() +
  geom_dotplot(binaxis = "y", stackdir = "center") +
  theme_classic(base_size = 24) +
  geom_hline(yintercept = 0, linetype='dashed') +
  labs(y = 'Difference in blood platelet levels')+
  theme(axis.title.x=element_blank(),
        axis.text.x=element_blank(),
        axis.ticks.x=element_blank())
```

```r
p
```



```r
library(dplyr)
names(df)
```

```
## [1] "before"     "after"      "difference"
```

```r
df = df %>% dplyr::mutate(
  absDif = abs(difference),
  rankAbsDif = rank(absDif),
  srank = sign(difference)*rank(abs(difference))
)
df
```

```
##    before after difference absDif rankAbsDif srank
## 1      25    27          2      2        2.0   2.0
## 2      25    29          4      4        3.5   3.5
## 3      27    37         10     10        7.0   7.0
## 4      44    36         -8      8        5.0  -5.0
## 5      30    46         16     16       10.0  10.0
## 6      67    82         15     15        8.5   8.5
## 7      53    57          4      4        3.5   3.5
## 8      53    80         27     27       11.0  11.0
## 9      52    61          9      9        6.0   6.0
## 10     60    59         -1      1        1.0  -1.0
## 11     28    43         15     15        8.5   8.5
```

```r
w_p = sum(df$srank[df$srank > 0])
w_p
```

```
## [1] 60
```

```r
w_m = sum(-df$srank[df$srank < 0])
w_m
```

```
## [1] 6
```

```r
w = min(w_p, w_m)
w
```

```
## [1] 6
```

- **Hypothesis:** $H_0\colon\ \mu_d = 0$ vs $H_1\colon\ \mu_d \neq 0$
- **Assumptions:** D_i are independently sampled from a symmetric distribution.
- **Test statistic:** $W^+ = \sum_{i:D_i>0} R_i$ where R_i are the ranks of $|D_1|, |D_2|, \ldots, |D_n|$. Under H_0, $W^+ \sim \mathrm{WSR}'(11)$, the WSR dist. with $n = 11$ and the set of ties as given.
- **Observed test statistic:** $w = \min(w^+, w^-) = 6$ because $w^+ = 60$, $w^- = 6$
- **p-value:** Since the $\mathrm{WSR}'(11)$ distribution is unknown, it is approximated by normal with
$$\mathrm{E}(W^+) = \frac{n(n+1)}{4} = \frac{11(11+1)}{4} = 33 \text{ and}$$
$$\mathrm{Var}(W^+) = \frac{1}{4}\sum_{i=1}^{11} r_i^2 = \frac{1}{4}[(-2)^2 + \cdots + (-8.5)^2] = \frac{506}{4} = 126.25.$$
$$\text{p-value} = 2\,P(W^+ \le 6) \simeq 2P\left(Z \le \tfrac{6-33}{\sqrt{126.25}}\right) = 2P(Z \le -2.403) = 2 \times 0.008 = 0.016$$
- **Decision:** the p-value is less than 0.05, hence there is evidence against H_0.

```r
ew = sum(df$rankAbsDif)/2
varw = sum((df$rankAbsDif)^2)/4
c(w, ew, varw)
```

```
## [1]   6.00  33.00 126.25
```

```r
t0 = (w - ew)/sqrt(varw)
p_value = 2 * pnorm(t0)
c(t0, p_value)
```

```
## [1] -2.40296846  0.01626259
```

```r
wilcox.test(df$difference)
```

```
## Warning in wilcox.test.default(df$difference): cannot
## compute exact p-value with ties

##
##     Wilcoxon signed rank test with continuity correction
##
## data:  df$difference
## V = 60, p-value = 0.01835
## alternative hypothesis: true location is not equal to 0
```

```r
wilcox.test(df$difference, correct = FALSE)
```

```
## Warning in wilcox.test.default(df$difference, correct =
## FALSE): cannot compute exact p-value with ties

##
##     Wilcoxon signed rank test
##
## data:  df$difference
## V = 60, p-value = 0.01626
## alternative hypothesis: true location is not equal to 0
```

```r
t.test(df$difference, alternative = "two.sided")
```

```
##
##      One Sample t-test
##
## data:  df$difference
## t = 2.9065, df = 10, p-value = 0.01566
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   1.97332 14.93577
## sample estimates:
## mean of x
##  8.454545
```

```r
binom.test(c(2,9), 0.5)
```

```
##
##      Exact binomial test
##
## data:  c(2, 9)
## number of successes = 2, number of trials = 11,
## p-value = 0.06543
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.0228312 0.5177559
```

We could also manually calculate the p-value for the sign test:

```r
2 * pbinom(2, 11, 0.5)
```

```
## [1] 0.06542969
```

See end of Lecture 13 for full details on the t-test and sign test for the smoking data.

Final notes

A few extra notes about the Wilcoxon signed-rank test:

- Since we assume that the distribution is symmetric, the hypotheses can also be stated in terms of the **median** (rather than the mean).
- The p-value from a Wilcoxon signed-rank test will typically be smaller than the p-value of a sign test on the same data. Using the information in the ranks, the test becomes much more **powerful** in detecting differences from μ_0, and almost as powerful as the one sample t-test.

Further reading

Larsen and Marx (2012; section 14.3).

Larsen, R. J. and M. L. Marx (2012). *An Introduction to Mathematical Statistics and its Applications*. 5th ed. Boston, MA: Prentice Hall. ISBN: 978-0-321-69394-5.

Wickham, H., M. Averick, J. Bryan, W. Chang, L. D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, et al. (2019). "Welcome to the tidyverse". In: *Journal of Open Source Software* 4.43, p. 1686. DOI: 10.21105/joss.01686.