

# DATA2002

## Sign test

Garth Tarr



Paired  $t$ -tests (revision)

Checking for normality

Sign test

# Paired $t$ -tests (revision)

# Rats

❓ Does a biochemical substance have an inhibitive effect on muscular growth?

For each of 10 rats:

- One hind leg muscle was regularly injected with the **biochemical substance**
- The corresponding muscle on the other hind leg was regularly injected with a **harmless placebo**
- At the end of 6 months the weights of the muscles were measured (in grams) and recorded as follows:



Rat ID	1	2	3	4	5	6	7	8	9	10
Biochemical	1.7	2.0	1.7	1.5	1.6	2.4	2.3	2.4	2.4	2.6
Placebo	2.1	1.8	2.2	2.2	1.5	2.9	2.9	2.4	2.6	2.5

# Paired $t$ -test

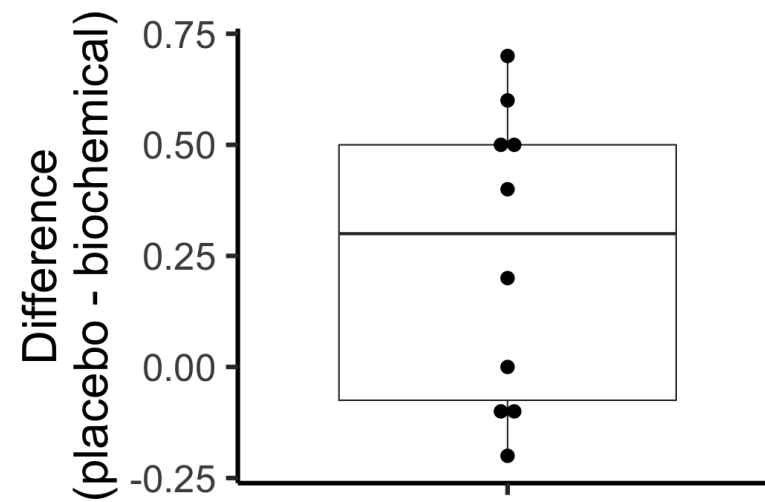
```
rat = data.frame(
  bio = c(1.7, 2.0, 1.7, 1.5, 1.6, 2.4, 2.3, 2.4, 2.4, 2.6),
  pla = c(2.1, 1.8, 2.2, 2.2, 1.5, 2.9, 2.9, 2.4, 2.6, 2.5)
) %>% # d = placebo - biochemical
  mutate(d = pla - bio)
rat %>% summarise(mean = mean(d), sd = sd(d))
```

```
##      mean      sd
## 1 0.25 0.3308239
```

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

```
##
##      One Sample t-test
##
## data:  rat$d
## t = 2.3897, df = 9, p-value = 0.02029
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  0.05822761      Inf
## sample estimates:
## mean of x
```

```
rat %>% ggplot() +
  aes(x = "", y = d) +
  geom_boxplot() +
  geom_dotplot(
    binaxis = "y",
    stackdir = "center") +
  labs(x = "",
       y = "Difference\n(placebo - biochemical)")
```



# Paired $t$ -test

Let  $\mu_d$  be the population mean difference in muscle weight between the treated leg and the placebo leg (difference = placebo - treatment).

- **Hypothesis:**  $H_0: \mu_d = 0$  vs  $H_1: \mu_d > 0$
- **Assumptions:**  $D_i$  are independent and identically distributed (iid)  $N(\mu, \sigma^2)$ .
- **Test statistic:**  $T = \frac{\bar{D} - \mu_d}{S_d / \sqrt{n}}$ . Under  $H_0$ ,  $T \sim t_{n-1}$
- **Observed test statistic:**  $t_0 = \frac{0.25}{0.33 / \sqrt{10}} = 2.39$
- **p-value:**  $P(t_9 \geq 2.39) = 0.02$
- **Conclusion:** The p-value is less than 0.05, therefore we reject the null hypothesis at the 5% level of significance and conclude that the biochemical substance does inhibit muscle growth.



# Critical values and rejection regions

Our test statistic followed a  $t$  distribution with  $n - 1 = 9$  degrees of freedom. Hence we could also have made our decision by looking at the **critical value** (equivalently, the **rejection region**).

The shaded area to the right is 0.05.

The critical value,  $c$ , is:

```
qt(0.95, df = 9)
```

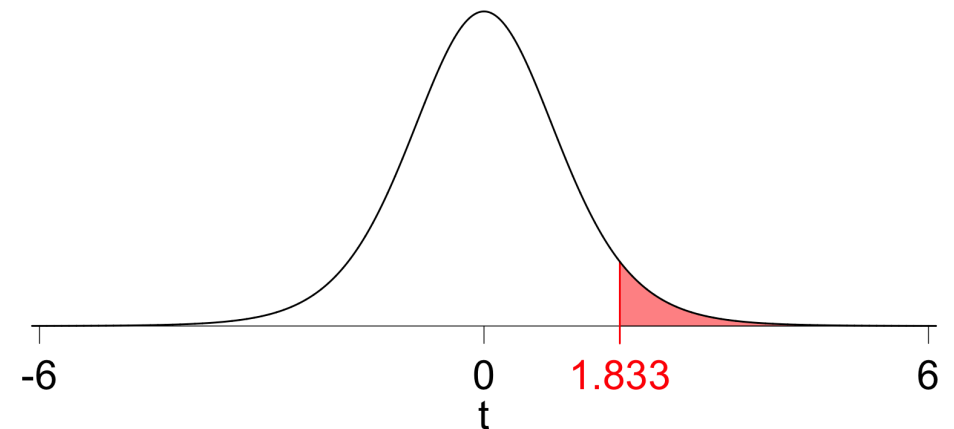
```
## [1] 1.833113
```

The **rejection region** is the part of the real line greater than  $c = 1.833$ .

If we observe a test statistic at least as big as 1.833, we reject the null hypothesis.

Our **observed test statistic** was  $t_0 = 2.39$ , which is larger than 1.833, hence we reject the null hypothesis at the 5% level of significance.

Probability density function for  $T \sim t(9)$



❓ Why is the rejection region only in the upper tail of the  $t$  distribution?



# Rejection region for the sample mean

Instead of operating on the scale of the test statistic, we can map back to the original scale and write the rejection region in terms of the sample mean.

The rejection region using test statistic:

$$t_0 = \frac{\bar{d} - \mu_0}{s_d / \sqrt{n}} \geq c$$

Noting that,

$$\begin{aligned} \alpha &= P\left(\frac{\bar{D} - \mu_0}{S_d / \sqrt{n}} \geq c\right) \\ &= P\left(\bar{D} \geq \mu_0 + cs_d / \sqrt{n}\right), \end{aligned}$$

we define a rejection region on the measurement scale as:

$$\{\bar{D} : \bar{D} \geq \mu_0 + cs_d / \sqrt{n}\} \quad \text{for} \quad H_1: \mu_d > \mu_0.$$

```
qt(0.95, df = 9)*sd(rat$d)/sqrt(nrow(rat))
```

```
## [1] 0.1917724
```

$$\mu_0 + cs_d / \sqrt{n} = 1.833 \times 0.33 / \sqrt{10} = 0.19$$

**Rejection region on the measurement scale:**

$$\{\bar{d} : \bar{d} \geq 0.19\}$$

Hence, we would reject the null hypothesis when we observe a sample mean difference greater than 0.19 grams.

In the present case, the sample mean difference was 0.25 grams, so we reject the null hypothesis.



# Confidence interval for the sample mean

For a **two-sided alternative hypothesis**, the confidence interval is:

$$\bar{d} \pm c \frac{s_d}{\sqrt{n}}$$

Here the alternative hypothesis is  $H_1: \mu_d > 0$  and so we have a **one sided confidence interval**. The lower bound for the one sided confidence interval is:

$$\bar{d} - c \frac{s_d}{\sqrt{n}} = 0.25 - 1.83 \times \frac{0.33}{\sqrt{10}} \approx 0.06$$

Hence the 95% confidence interval is  $(0.06, \infty)$ .

② Using the confidence interval, how do we know whether or not to reject the null hypothesis,  
 $H_0: \mu_d = 0$ ?

# Assumptions

I NEED HELP  
MAKING UNREALIS-  
TIC ASSUMPTIONS TO  
SUPPORT A BUSINESS  
CASE FOR A BAD IDEA.

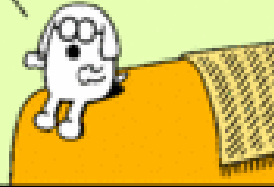
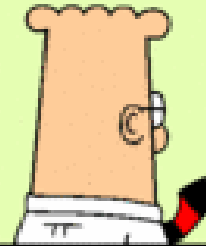


EASY.



www.dilbert.com scottadams@aol.com

THERE'S A HOLE IN THE  
BACK OF OUR WARDROBE  
CLOSET THAT LEADS TO  
A MAGICAL WORLD OF  
PREPOSTEROUS BUSINESS  
ASSUMPTIONS.



2-14-06 © 2006 Scott Adams, Inc./Dist. by UFS, Inc.

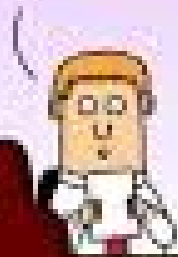
WE DON'T  
HAVE A  
WARDROBE  
CLOSET.



ASSUME  
WE DO.



AND MY  
REVENUE  
FORECAST  
SAYS...



DID YOU  
MAKE ANY  
ASSUMP-  
TIONS?

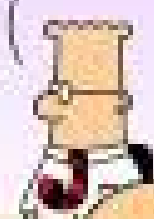


Dilbert.com DilbertCartoonist@gmail.com

I MADE  
A LOT OF  
THEM.



THEN  
WE DON'T  
BELIEVE  
YOUR  
FORECAST.



5-9-12 © 2012 Scott Adams, Inc./Dist. by Universal Uclick

CAN I  
TELL YOU  
ABOUT IT  
ANYWAY?



DO WHAT-  
EVER MAKES  
YOU FEEL  
LESS  
ABSURD.



# Normality

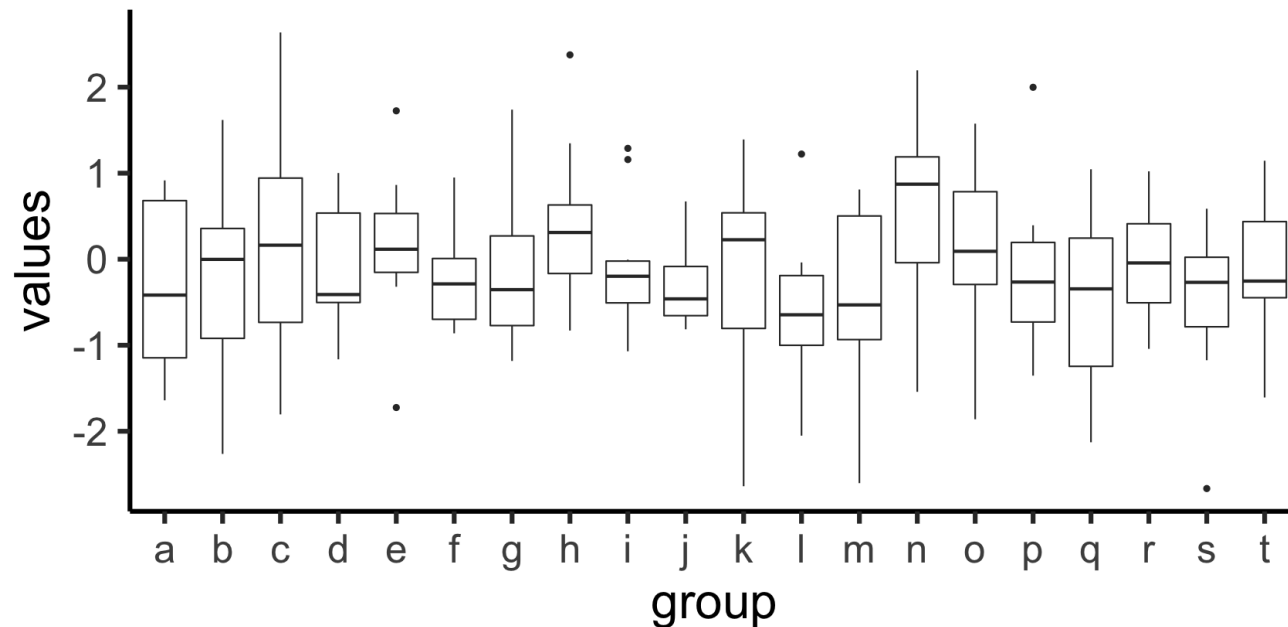
- The assumption that your data are **sampled from a normal population** arises quite often.
- If you have a **large enough** sample size, then the normality assumption is not as important as you can usually rely on the central limit theorem to ensure your test statistic at least approximately follows a  $t$ -distribution.
- In **small samples** it can be difficult to tell whether or not your sample comes from a normal population!

You can check the normality assumption using:

- **Boxplots** (looking for symmetry)
- **QQ plots** (looking for a straight line)
- Formal hypothesis tests?

# Checking for normality: boxplots

```
set.seed(2019)
values = rnorm(200)
group = rep(letters[1:20], each = 10)
dat = data.frame(values, group)
library(ggplot2)
ggplot(dat, aes(x = group, y = values)) + geom_boxplot()
```



These are all boxplots of 10 observations drawn from a normal population.

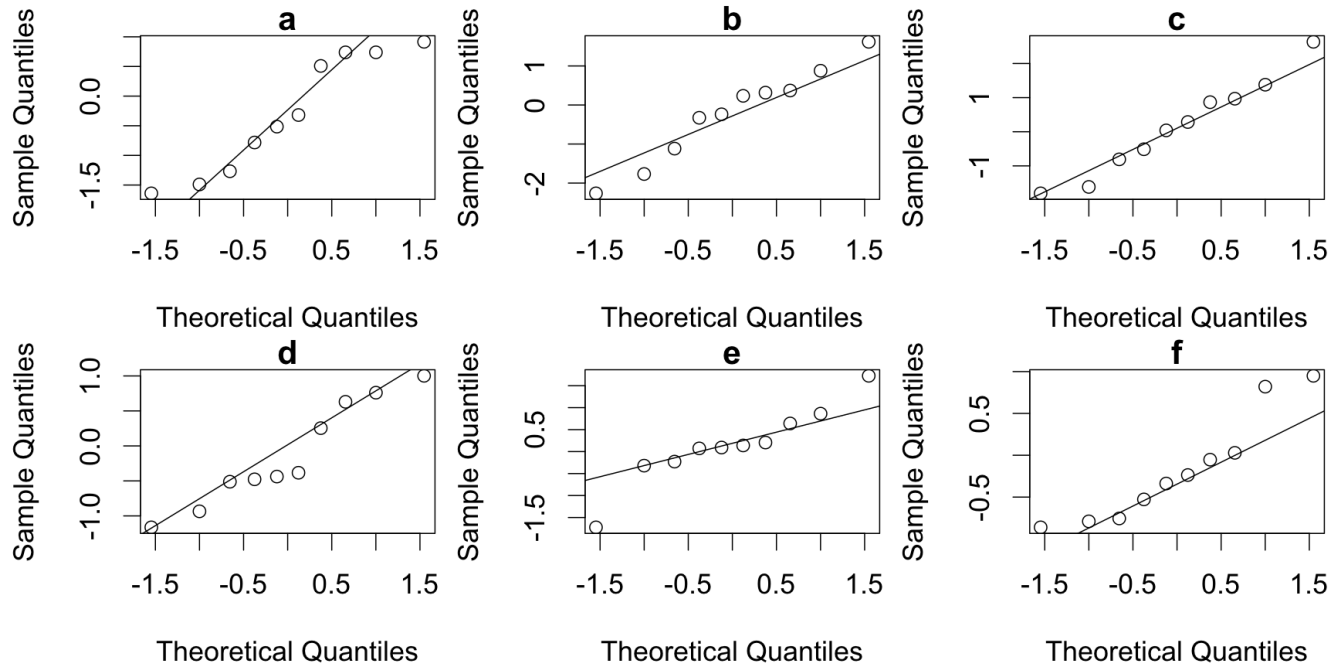
If we're checking for normality in a boxplot, then we're mostly looking for **symmetry**.

# Checking for normality: QQ plots (base)

```
par(mfrow = c(2,3), cex = 1.5, mar = c(4,4,1.1,0.5))  
for(i in letters[1:6]){  
  qqnorm(dat$values[dat$group == i], main = i)  
  qqline(dat$values[dat$group == i])  
}
```

These are QQ-plots, each with 10 observations drawn from a normal population.

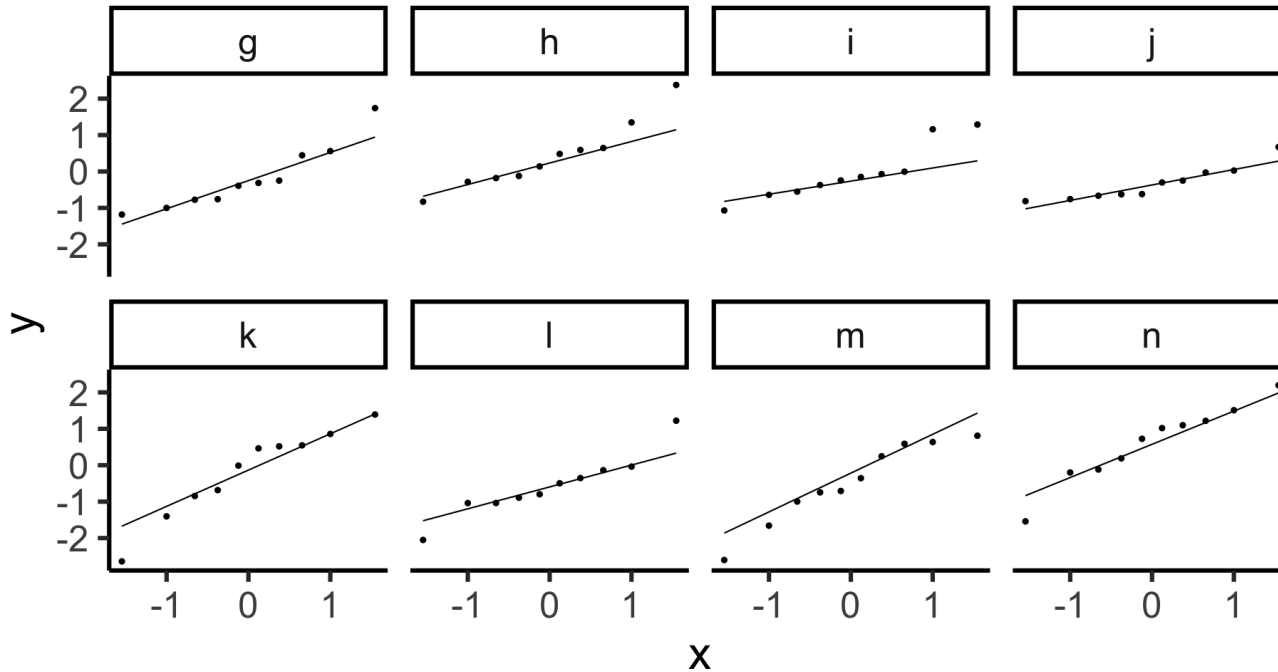
If we're checking for normality in a QQ plot, then we're mostly looking for **points that lie reasonably close to the line.**



# Checking for normality: QQ plots (ggplot2)

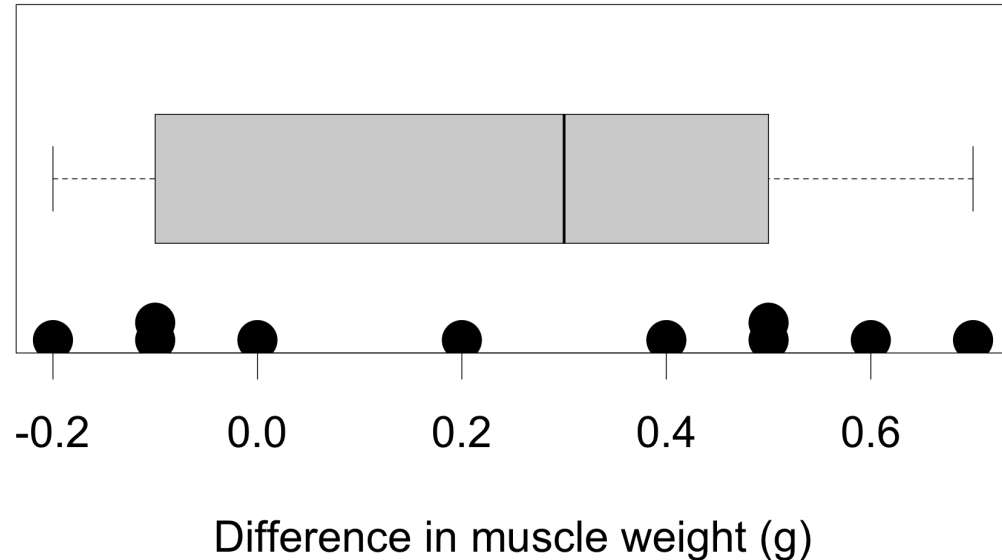
```
library(dplyr)
dat %>% dplyr::filter(group %in% letters[7:14]) %>%
  ggplot(aes(sample = values, group = group)) + geom_qq() +
  facet_wrap(~group, nrow = 2)
```

These are also QQ-plots, each with 10 observations drawn from a normal population.



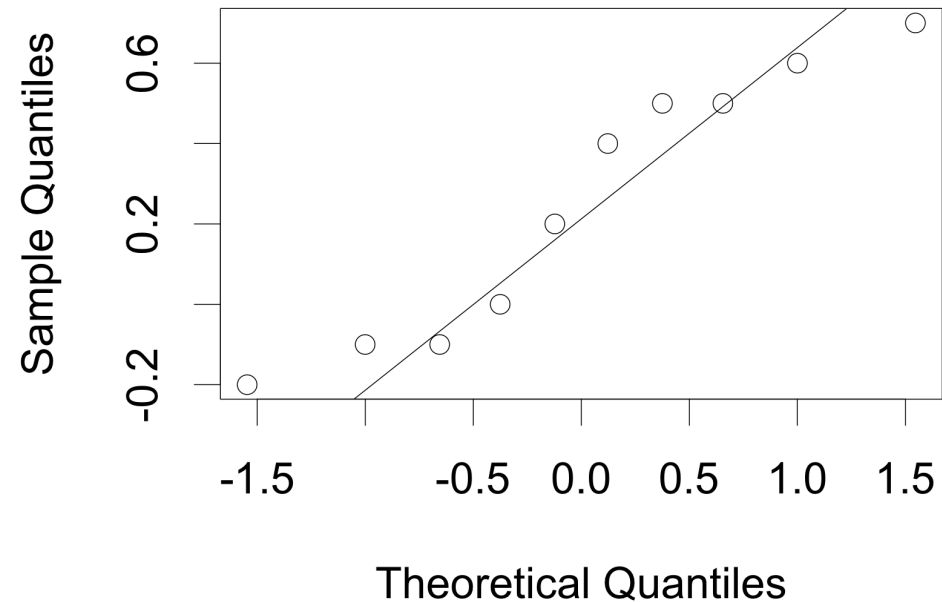
# Rat data: normality (base graphics)

```
par(cex = 3, mar = c(4, 0.5, 0.5, 0.5))
boxplot(rat$d, horizontal = TRUE,
        xlab="Difference in muscle weight (g)")
stripchart(rat$d, pch = 20, at = 0.5,
           add = TRUE, method="stack")
```



```
par(cex=3)
qqnorm(rat$d)
qqline(rat$d)
```

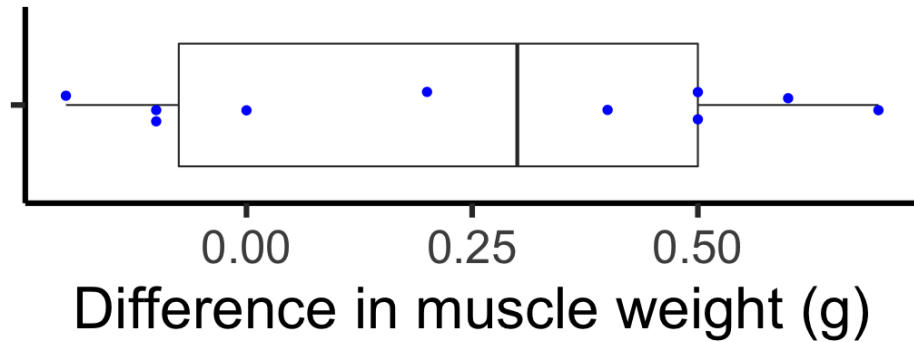
**Normal Q-Q Plot**



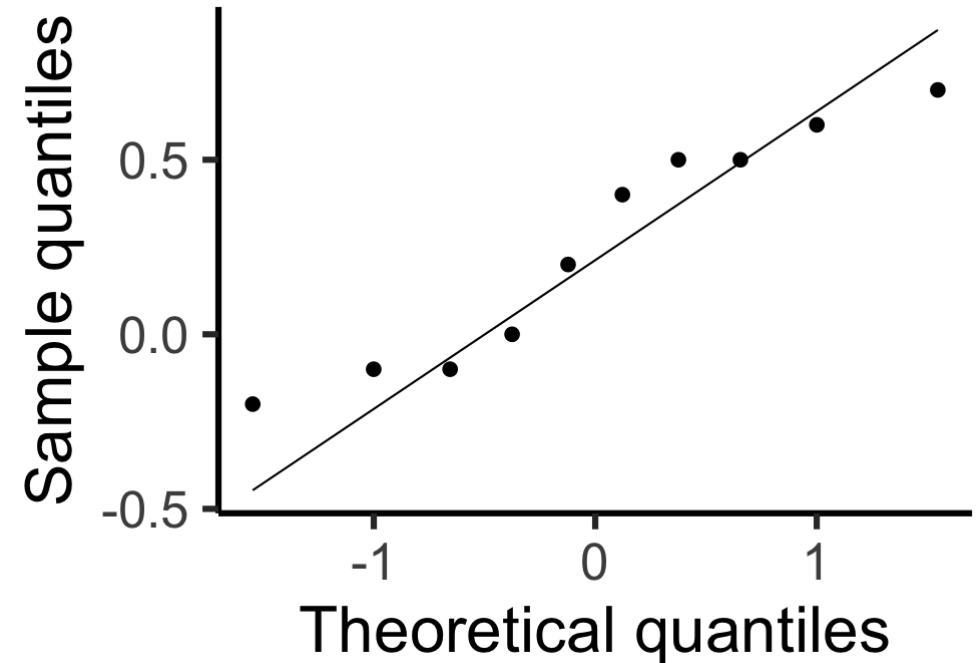


# Rat data: normality (ggplot2)

```
ggplot(rat, aes(x = "", y = d)) +  
  geom_boxplot() +  
  geom_jitter(width = 0.1,  
              size = 2,  
              colour = "blue") +  
  labs(x = "",  
       y = "Difference in muscle weight (g)") +  
  coord_flip()
```



```
ggplot(rat, aes(sample = d)) +  
  geom_qq(size = 3) +  
  geom_qq_line() +  
  labs(x = "Theoretical quantiles",  
       y = "Sample quantiles")
```



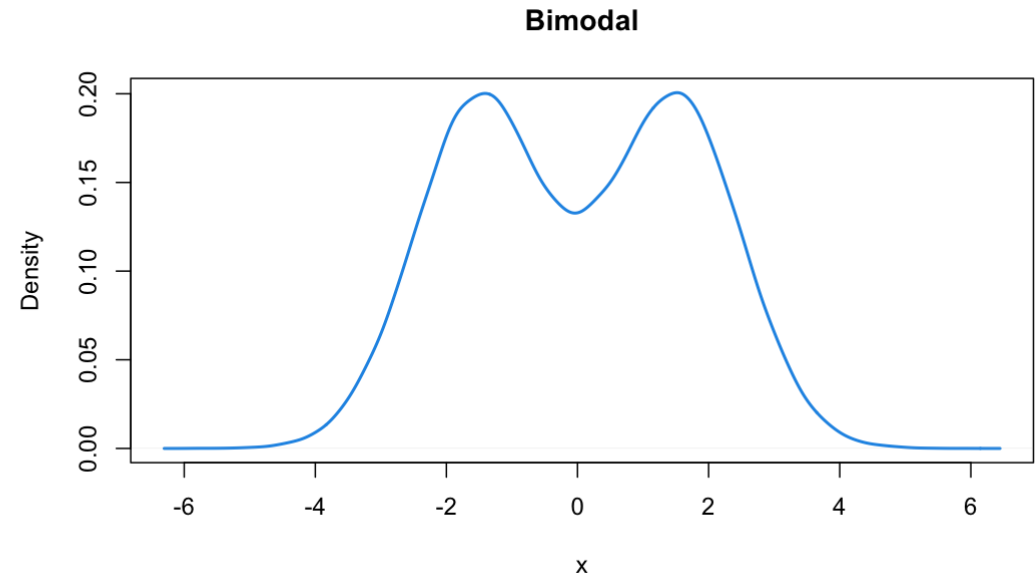
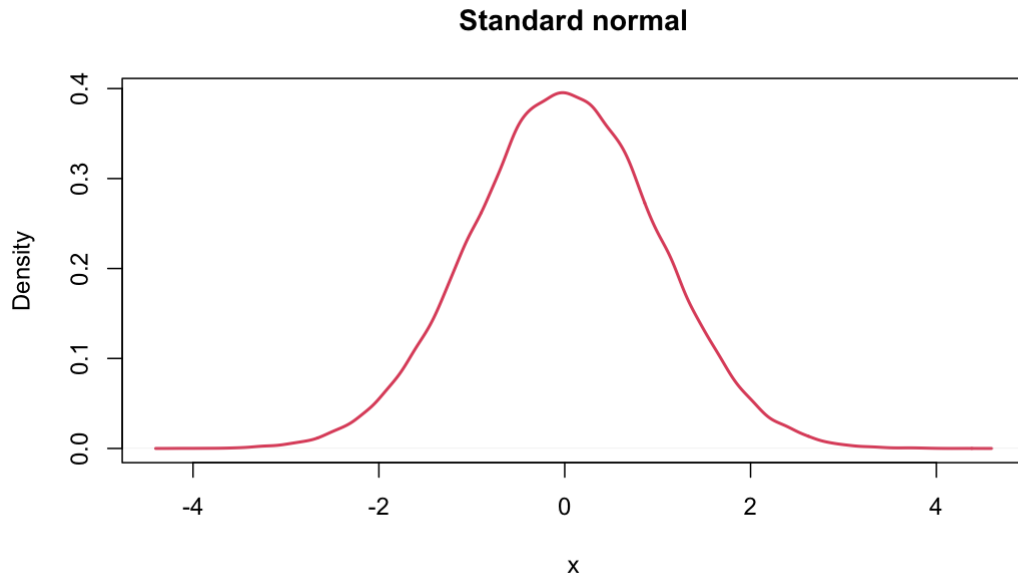
Sign test

# Sign test

- Suppose a sample  $X_1, \dots, X_n$  are independently sampled from a continuous distribution with mean  $\mu$ .
- We want to test  $H_0: \mu = \mu_0$ .
- If the distribution is **symmetric** about  $\mu_0$  under  $H_0$ , then  $D_i = X_i - \mu_0$  should scatter around 0
  - i.e.  $D_i$  is equally likely to be positive or negative.
- If  $H_0$  is true, the probability  $p_+$ , of getting a positive  $D_i$  is 0.5.
- The **sign test** reduces to a binomial test of proportions.
- The sign test is a **nonparametric** test as no assumption on the data distribution is made except symmetry (though we do still require the independence assumption).

# Symmetric distributions

```
set.seed(123)
n = 100000
par(cex = 2, mfrow = c(1,2))
plot(density(rnorm(n)),
     main = 'Standard normal', ylab = 'Density', xlab = 'x', col = 2, lwd = 2)
plot(density(c(rnorm(n,-1.5),rnorm(n,1.5))),
     main = 'Bimodal', ylab = 'Density', xlab = 'x', col = 4, lwd = 2)
```

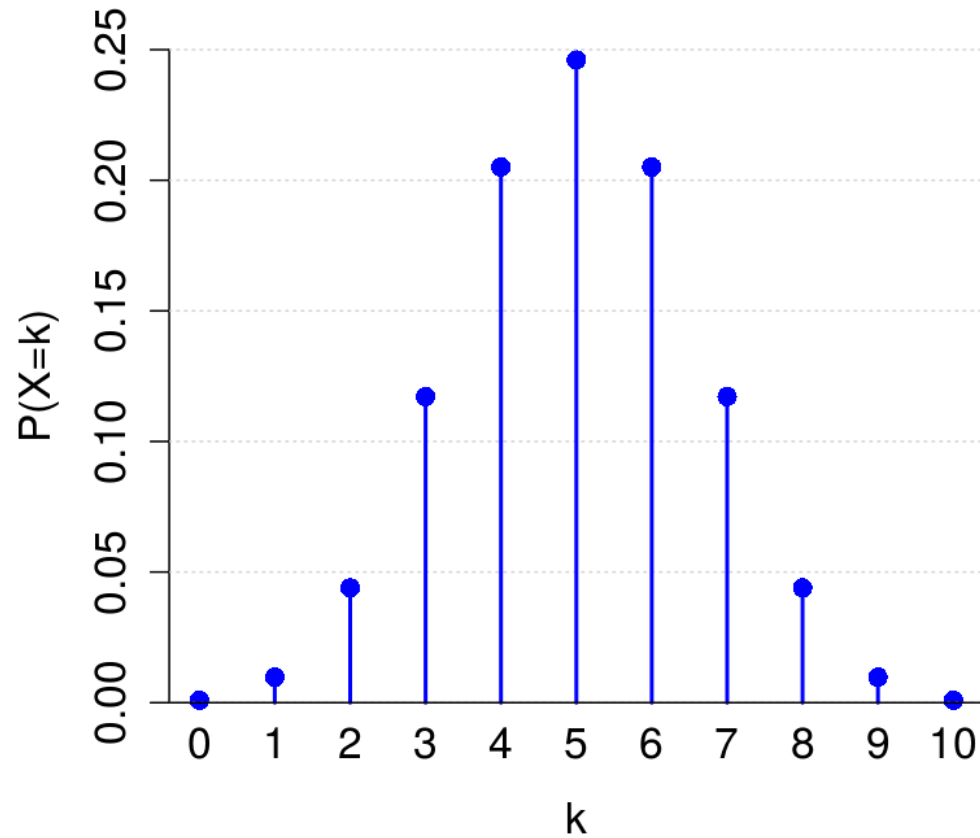


Suppose  $X_1, \dots, X_n$  are drawn from some symmetric population with mean  $\mu$ .

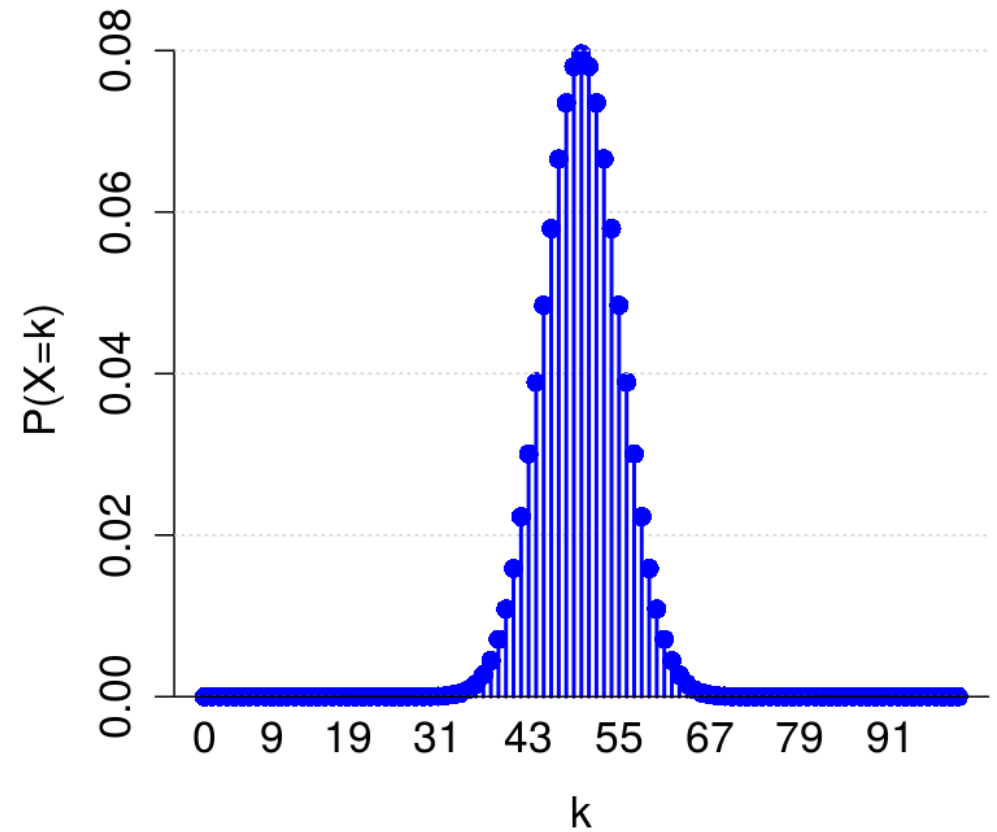
- **Hypothesis:**  $H_0: \mu = \mu_0$  vs  $H_1: \mu > \mu_0, \mu < \mu_0$  or  $\mu \neq \mu_0$ . We can also write it in terms of the probability of seeing a positive difference,  $H_0: p_+ = \frac{1}{2}$  vs  $H_1: p_+ > \frac{1}{2}, p_+ < \frac{1}{2}$  or  $p_+ \neq \frac{1}{2}$ .
- **Assumptions:**  $X_i$  are independently sampled from a symmetric distribution.
- **Test statistic:**  $T = \#(D_i > 0)$  where  $D_i = X_i - \mu$ . Under  $H_0$ ,  $T \sim B(n, \frac{1}{2})$  where  $n$  is the number of non-zero differences.
- **Observed test statistic:**  $t_0 = \#(d_i > 0)$
- **p-value:**
  - $H_1: \mu < \mu_0$  we use a lower tail p-value  $P(T \leq t_0)$
  - $H_1: \mu > \mu_0$  we use an upper tail p-value  $P(T \geq t_0)$
  - $H_1: \mu \neq \mu_0$  and  $t_0 < \frac{n}{2}$  we use two times the lower tail probability  $2P(T \leq t_0)$
  - $H_1: \mu \neq \mu_0$  and  $t_0 > \frac{n}{2}$  we use two times the upper tail probability  $2P(T \geq t_0)$
- **Decision:** If p-value  $< \alpha$  there is evidence against  $H_0$ .

# Binomial distribution

Probability mass function for  $X \sim B(10, 0.5)$



Probability mass function for  $X \sim B(100, 0.5)$



# Sign test for paired data

- As we saw last week, **paired data** are very common.
- For example before/after trials, studies on twins, left/right arm freckles count.
- We have data of the form

First variable: $X$	$X_1$	$X_2$	$\dots$	$X_n$
Second variable: $Y$	$Y_1$	$Y_2$	$\dots$	$Y_n$
<hr/>				
Difference: $D$	$D_1$	$D_2$	$\dots$	$D_n$

where  $D_i = X_i - Y_i$  for  $i = 1, \dots, n$ .

- We often want to see if there is a difference between the distributions of the  $X$  and  $Y$  populations.
- If we can safely assume that the differences are **normally distributed**, we can use a paired  $t$ -test, to test if the population mean difference equals zero,  $H_0: \mu_d = 0$ .
- If we do not feel comfortable making the normality assumption, we can still analyse the differences using the **sign test**.

# Sign test for paired data

- We use the sign of the differences and ignore their magnitude

⇒ test reduces to a **test of proportions**

- Based on the proportion of positive differences,  $p_+$ .
- The null hypothesis corresponding to **no difference** between the two populations is  $H_0: p_+ = \frac{1}{2}$ , i.e. there is an equal proportion of positive and negative differences.
- Under the null hypothesis, the symmetry assumption of the sign test is guaranteed, so when considering a sign test on the differences we only need to assume that the differences are **independent**.

❓ What happens when the before and after measurements are identical?

The test of proportions is for binomial data with two possible outcomes only (yes/no, success/fail, etc). Thus, we will discard differences which are exactly zero.



# Rat muscle weights

```
rat %>% mutate(  
  pos_d = d > 0  
)
```

```
##      bio pla      d pos_d  
## 1  1.7 2.1  0.4  TRUE  
## 2  2.0 1.8 -0.2 FALSE  
## 3  1.7 2.2  0.5  TRUE  
## 4  1.5 2.2  0.7  TRUE  
## 5  1.6 1.5 -0.1 FALSE  
## 6  2.4 2.9  0.5  TRUE  
## 7  2.3 2.9  0.6  TRUE  
## 8  2.4 2.4  0.0 FALSE  
## 9  2.4 2.6  0.2  TRUE  
## 10 2.6 2.5 -0.1 FALSE
```

Count the number of positive differences (TRUE = 1, FALSE = 0)

```
sum(rat$d > 0)
```

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

Count the number of non-zero differences

```
sum(rat$d != 0)
```

```
## [1] 9
```



Let  $p_+$  be the probability of a positive difference between the treated leg and the placebo leg (difference = placebo - treatment).

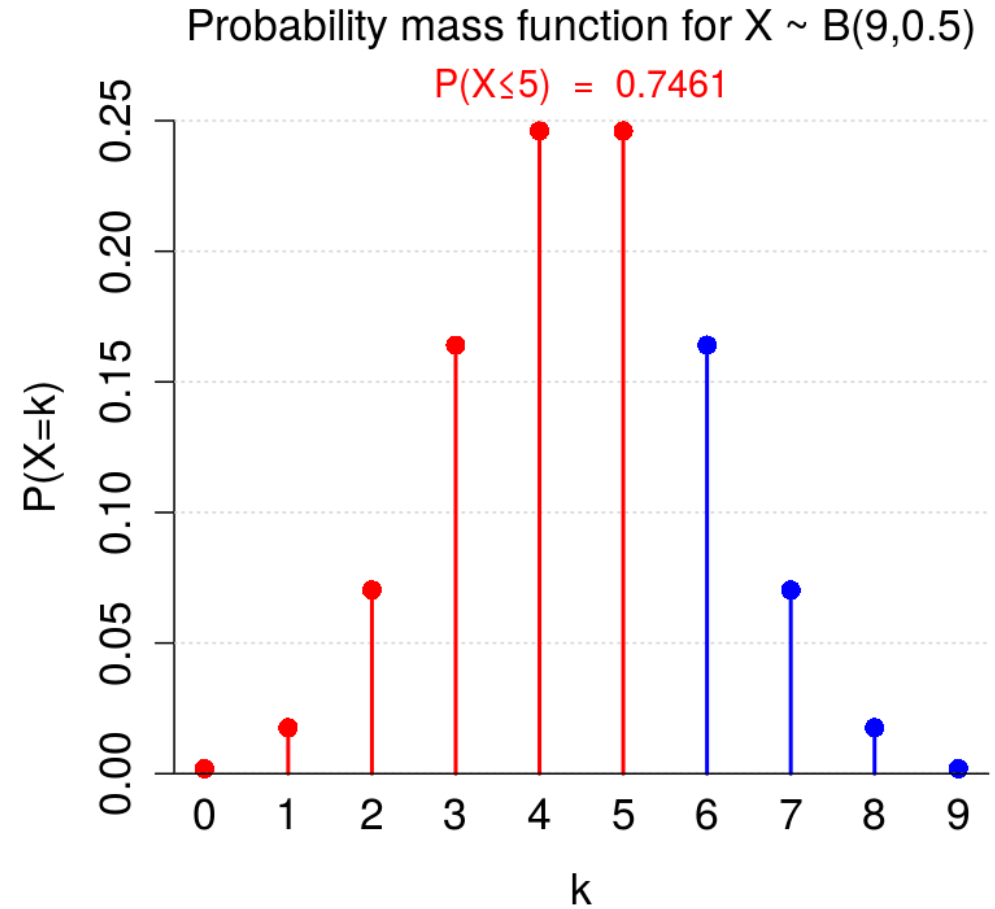
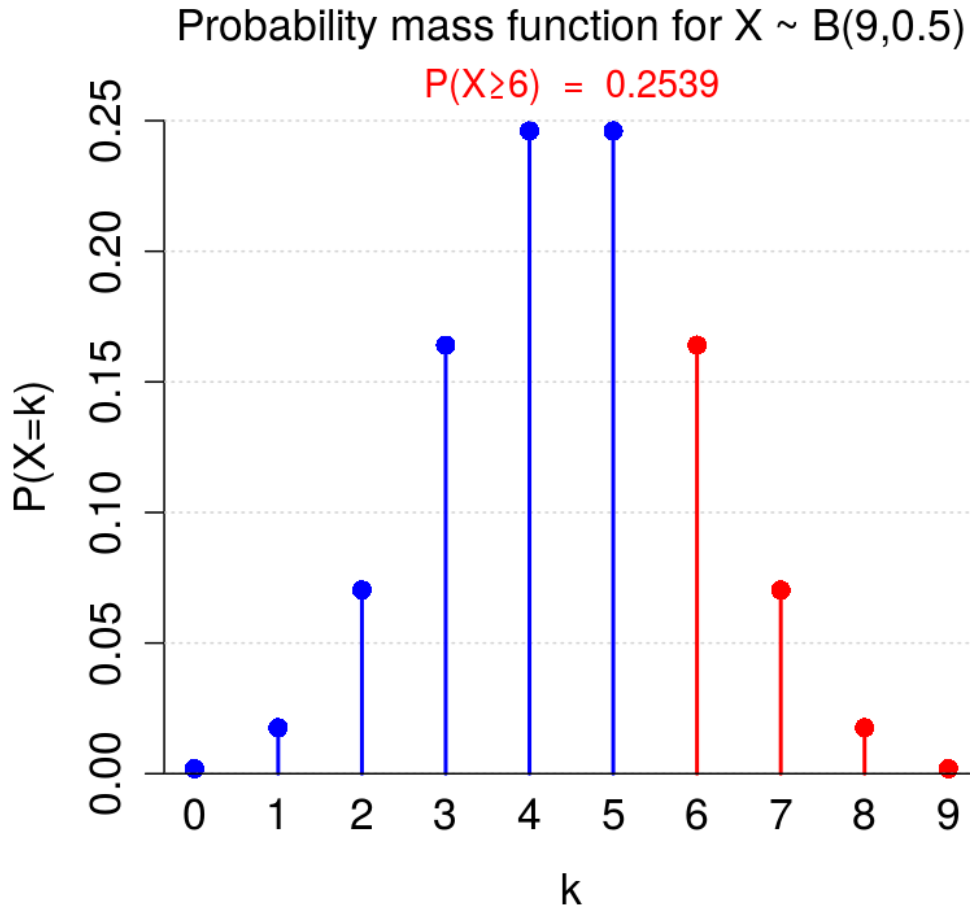
- **Hypothesis:**  $H_0: p_+ = \frac{1}{2}$  vs  $H_1: p_+ > \frac{1}{2}$
- **Assumptions:** Differences,  $D_i$ , are independent.
- **Test statistic:** Let  $T$  be the number of positive differences out of the 9 non-zero differences. Under  $H_0$ ,  $T \sim B(9, \frac{1}{2})$ . I.e. under  $H_0$ ,  $T$  follows a binomial distribution with  $n = 9$  and  $p = 0.5$ .
- **Observed test statistic:** We observed  $t_0 = 6$  positive differences in the sample.
- **p-value:** probability of getting a test statistic as or more extreme than what we observed,

$$\begin{aligned} P(T \geq 6) &= 1 - P(T \leq 5) \\ &= 1 - \text{pbinom}(5, \text{size} = 9, \text{prob} = 1/2) \approx 0.2539 \end{aligned}$$

- **Conclusion:** As the p-value is greater than 0.05, the data are consistent with the null hypothesis at the 5% level of significance. There is no significant difference between the biochemical and the placebo.

# Binomial distribution

Let  $T \sim B(9, 0.5)$  then,  $P(T \geq 6) = 1 - P(T \leq 5)$ .





```
s = sign(rat$d)[sign(rat$d) != 0]
table(s)
```

```
## s
## -1  1
##  3  6
```

```
binom.test(c(6, 3), p = 0.5, alternative = "greater")
```

```
##
##      Exact binomial test
##
## data:  c(6, 3)
## number of successes = 6, number of trials = 9, p-value = 0.2539
## alternative hypothesis: true probability of success is greater than 0.5
## 95 percent confidence interval:
##  0.3449414 1.0000000
## sample estimates:
## probability of success
##           0.6666667
```

```
# equivalently could specify number of successes and the total number of trials
# binom.test(6, n = length(s), p = 0.5, alternative = "greater")
```



# Conflicting results

- The p-value using the sign test is 0.254.
- The p-value using the  $t$ -test is 0.020.

```
t.test(rat$d, mu = 0, alternative = "greater")
```

```
##
##      One Sample t-test
##
## data:  rat$d
## t = 2.3897, df = 9, p-value = 0.02029
## alternative hypothesis: true mean is greater than 0
## 95 percent confidence interval:
##  0.05822761      Inf
## sample estimates:
## mean of x
##      0.25
```

- In this case the paired  $t$ -test and the sign test give conflicting results.
- This is not uncommon when the sample size is small.

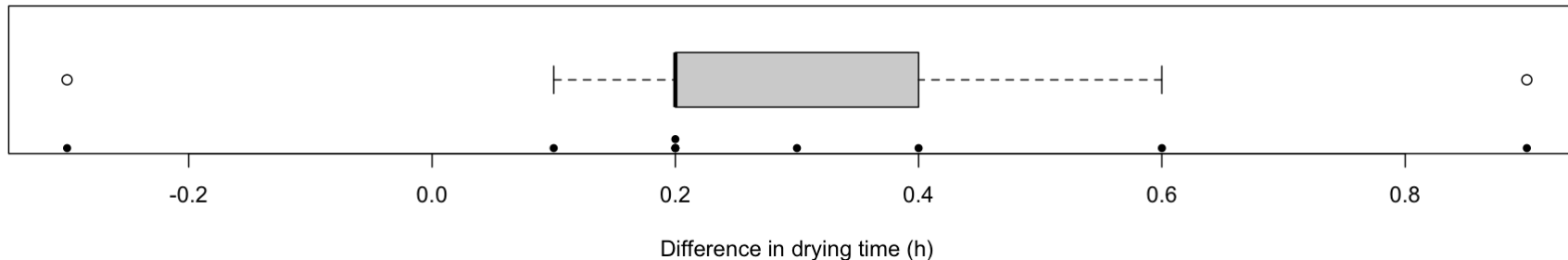
# Paint drying times

A paint supplier claims that a new additive will reduce the drying time of acrylic paint. To test this claim 10 panels of wood are painted: one half with the original paint formula and one half with the paint having the new additive. The drying times in hours are:

```
new_paint = c(6.4,5.8,7.4,5.5,6.3,7.8,8.6,8.2,7.0,4.9)
old_paint = c(6.6,5.9,7.8,5.7,6.0,8.4,8.8,8.4,7.3,5.8)
d = old_paint - new_paint
sort(d)
```

```
## [1] -0.3  0.1  0.2  0.2  0.2  0.2  0.3  0.4  0.6  0.9
```

```
par(mar = c(4,0.5,0.5,0.5))
boxplot(d, horizontal = TRUE, xlab = "Difference in drying time (h)")
stripchart(d, pch = 20, at = 0.5, add = TRUE, method = "stack")
```



# Paint drying times

Let  $p_+$  denote the probability of a positive difference.

- $H_0: p_+ = \frac{1}{2}$  against  $H_1: p_+ > \frac{1}{2}$ .
- Assume differences are independent. Let  $\alpha = 0.05$ .
- Let  $T$  denote the number of positive differences. There are 10 non-zero differences. Thus under  $H_0$ ,  $T \sim B(10, \frac{1}{2})$ .
- We observe  $t_0 = 9$  positive differences out of the 10 non-zero ones.
- Thus, the p-value is:

$$\begin{aligned} P(T \geq t_0) &= P(T \geq 9) \\ &= 1 - P(T \leq 8) \\ &= 1 - \text{pbinom}(8, 10, 1/2) \approx 0.0107. \end{aligned}$$

- As the p-value is less than  $\alpha = 0.05$ , we reject  $H_0$  and conclude that the new additive is effective in reducing the drying time of the paint.

# Remarks

## Statistical Thinking

What are some of the benefits and drawbacks of using the sign test?



# Remarks

- The sign test **ignores a lot of the information** in the sample but it can be applied in **quite general** situations. We only use the *sign* of the  $d_i$  and ignore their *magnitude* in the sign test.
- The sign test does not depend on the distribution of the data! For this reason sometimes these types of tests are called **nonparametric**.
- If the normality assumption is satisfied, the  $t$ -test is more **powerful**, in the sense that it will reject the null hypothesis when the alternative hypothesis is true more often than the sign test.
- The sign test can be used to test if a **single sample** is taken from a **continuous distribution** that is **symmetric about** its population mean  $\mu$ .
- The sign test is more *robust*, i.e. less affected by outlying large or small observations, than a  $t$ -test.
- If the sample is reasonably believed to come from a normal population, you should use the more powerful  $t$ -test instead of a sign test.

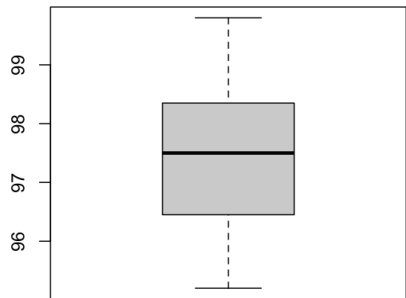
Additional practice questions

# Moisture retention

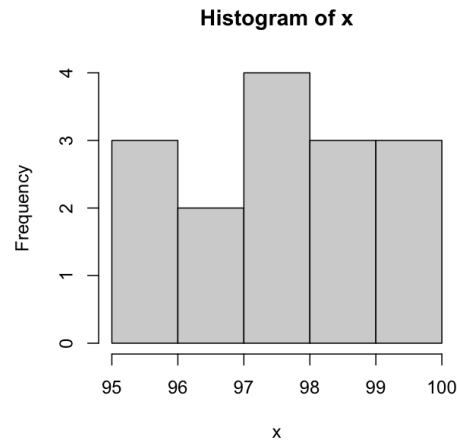
The following data are 15 measurements of moisture retention (%) using a new sealing system. The system is expected to be better (greater retention) than the previous system, for which the mean retention was 96%.

```
x = c(97.5, 95.2, 97.3, 96.0, 96.8, 99.8, 97.4, 95.3, 98.2, 99.1, 96.1, 97.6, 98.2, 98.5, 99.4)
```

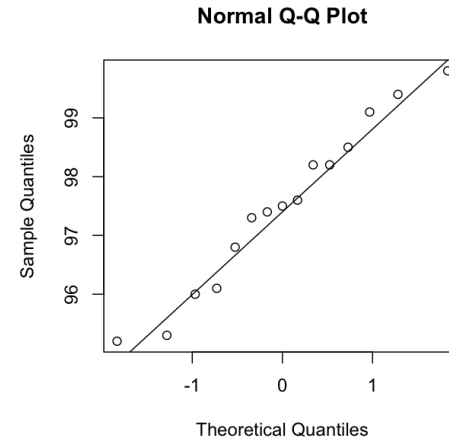
```
boxplot(x)
```



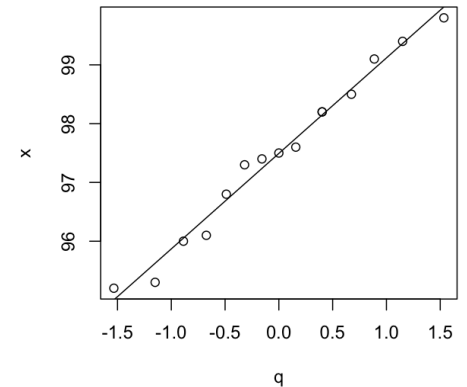
```
hist(x)
```



```
qqnorm(x)  
qqline(x)
```



```
p = rank(x)/(length(x)+1)  
q = qnorm(p)  
plot(q,x)  
abline(lm(x~q))
```



# Moisture retention

The following data are 15 measurements of moisture retention (%) using a new sealing system. The system is expected to be better (greater retention) than the previous system, for which the mean retention was 96%.

```
x = c(97.5, 95.2, 97.3, 96.0, 96.8, 99.8, 97.4, 95.3, 98.2, 99.1, 96.1, 97.6, 98.2, 98.5, 99.4)
```

The sign of the differences  $d_i = x_i - \mu_0 = x_i - 96$  are:

```
sign(x - 96)
```

```
## [1]  1 -1  1  0  1  1  1 -1  1  1  1  1  1  1  1
```

```
sum(sign(x - 96) == 1)
```

```
## [1] 12
```

```
sum(sign(x - 96) != 0)
```

```
## [1] 14
```

Let  $p_+$  be the probability of a positive difference. The *sign test* for the mean % of moisture retention using a new sealing system is:

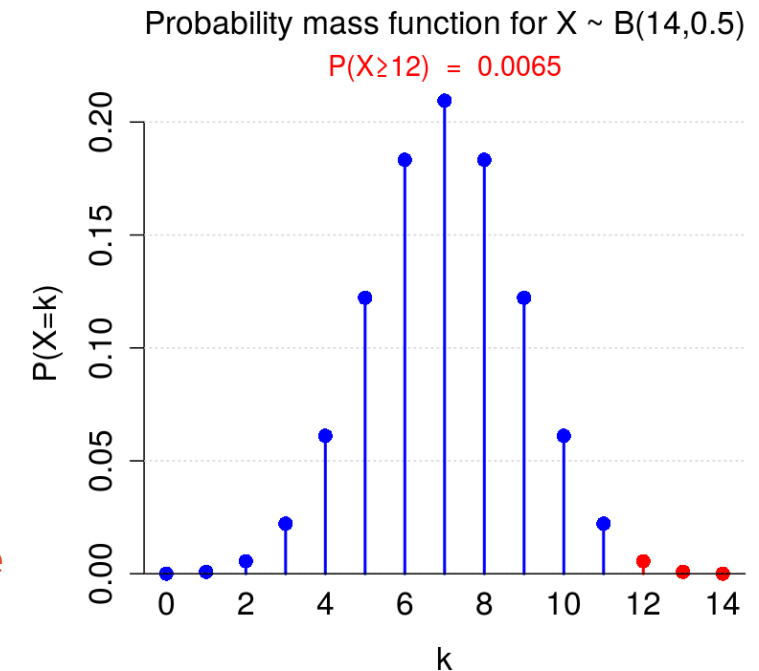
- **Hypothesis:**  $H_0: p_+ = \frac{1}{2}$  vs  $H_1: p_+ > \frac{1}{2}$
- **Assumptions:** Differences,  $D_i$ , are independent.
- **Test statistic:**  $T = \#(D_i > 0)$  where  $D_i = X_i - \mu$ . Under  $H_0$ ,  $T \sim B(14, 0.5)$ .

- **Observed test statistic:**  $t_0 = \#(d_i > 0) = 12$ .

- **p-value:**

$$\begin{aligned} P(T \geq 12) &= P(T = 12) + P(T = 13) + P(T = 14) \\ &= \sum_{i=12}^{14} \binom{14}{i} 0.5^i 0.5^{14-i} \\ &= 0.5^{14} \left[ \binom{14}{12} + \binom{14}{13} + \binom{14}{14} \right] = 0.0065. \end{aligned}$$

- **Decision:** Since the p-value is less than 0.05 there is strong evidence against the null hypothesis. Hence, we conclude that the retention rate is greater than 96%.



## "Manual" calculation

```
x = c(97.5, 95.2, 97.3, 96.0, 96.8, 99.8, 97.4, 95.3, 98.2, 99.1,  
mu0 = 96 # hypothesised parameter value  
d = x - mu0 # vector of differences  
n = sum(d != 0) # number of non-zero differences  
t0 = sum(d > 0) # count of positive differences  
ps = t0/n # proportion of positive differences  
p0 = 0.5 # proportion of positive differences under the null hypot  
p_value = 1 - pbinom(q = t0 - 1, size = n, prob = p0)  
p_value
```

```
## [1] 0.006469727
```

## Using `binom.test()`

```
binom.test(t0, n, p = 0.5, alternative = "greater", conf.level = 0.95)
```

```
##  
##      Exact binomial test  
##  
## data:  t0 and n  
## number of successes = 12, number of trials = 14, p-value = 0.00647  
## alternative hypothesis: true probability of success is greater than 0.5  
## 95 percent confidence interval:  
##  0.6146103 1.0000000  
## sample estimates:  
## probability of success  
##           0.8571429
```



## Compare with `t.test()`

```
t.test(x, mu = mu0, alternative = "greater")
```

```
##  
##      One Sample t-test  
##  
## data:  x  
## t = 4.0658, df = 14, p-value = 0.0005784  
## alternative hypothesis: true mean is greater than 96  
## 95 percent confidence interval:  
##  96.84642      Inf  
## sample estimates:  
## mean of x  
##  97.49333
```





# 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 %>% arrange(difference)
```

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

Is platelet aggregation affected by smoking?

```
apply(df, 2, mean) %>% round(2)
```

before	after	difference
42.18	50.64	8.45

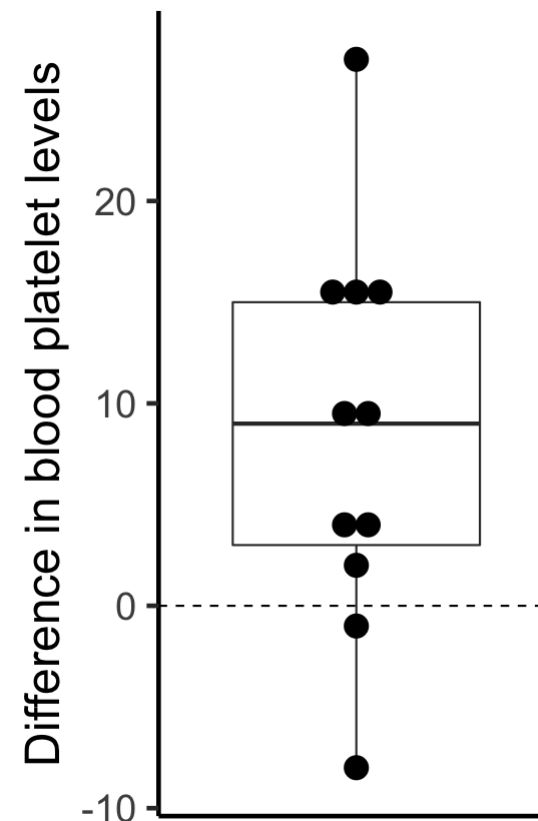
```
apply(df, 2, sd) %>% round(2)
```

before	after	difference
15.61	18.90	9.65



```
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())
```

p





Let  $X_i$  and  $Y_i$  be the blood platelet aggregation levels for the  $i^{th}$  person before and after smoking, respectively. Define the change in person  $i$ 's platelet aggregation levels as  $D_i = Y_i - X_i$  and the population mean change in platelet aggregation levels as  $\mu_d$ . Also let  $p_+$  be the proportion of positive differences. Testing if  $p_+ = 0.5$  is the same as testing if  $\mu_d = 0$ .

Sign test for zero mean difference of the aggregation of blood platelets before and after smoking:

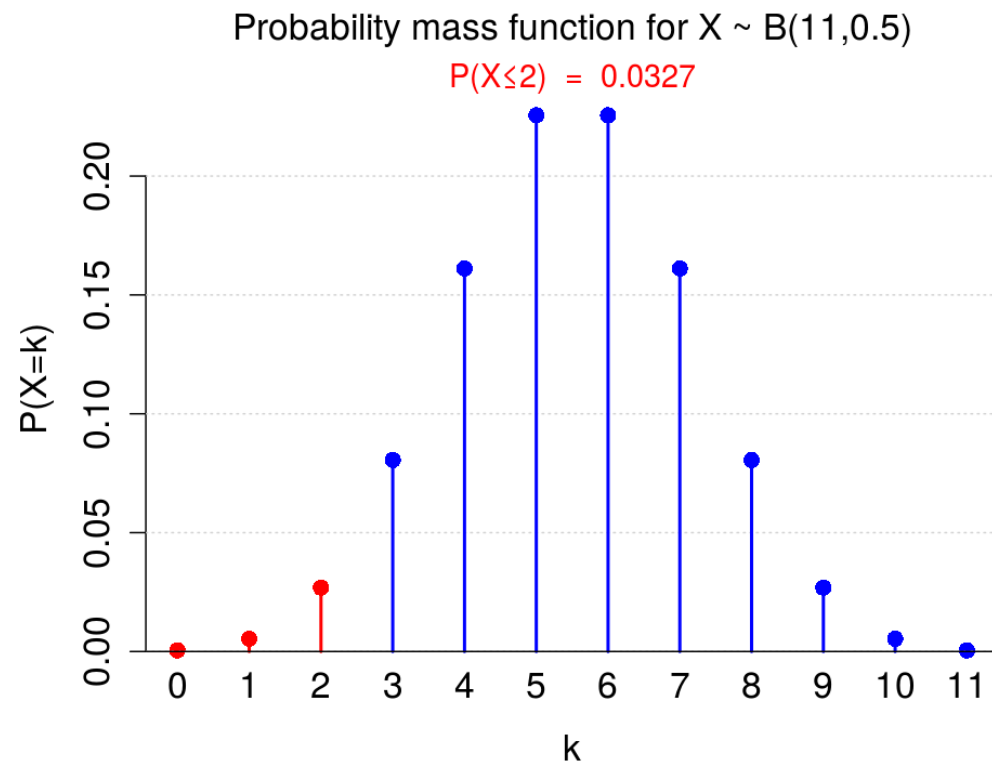
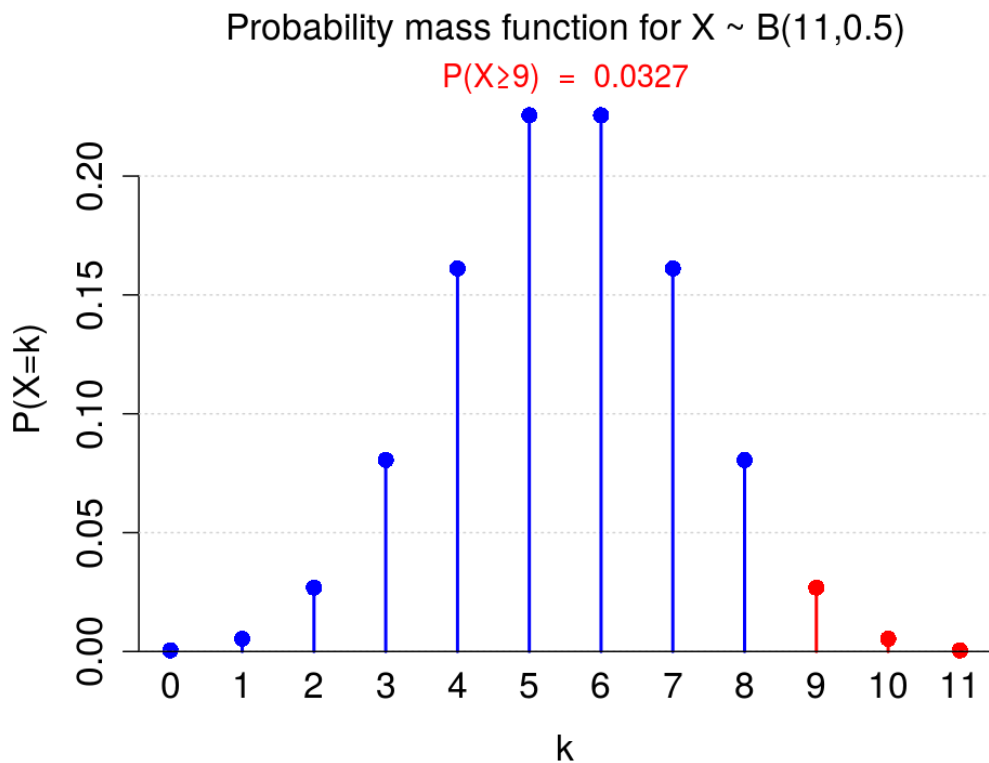
- **Hypothesis:**  $H_0: p_+ = \frac{1}{2}$  vs  $H_1: p_+ \neq \frac{1}{2}$
- **Assumptions:** Differences,  $D_i$ , are independent.
- **Test statistic:**  $T = \#(D_i > 0)$ . Under  $H_0$ ,  $T \sim B(11, \frac{1}{2})$ .
- **Observed test statistic:**  $t_0 = \#(d_i > 0) = 9$ . Large and small values of  $t_0$  will argue against  $H_0$ .
- **p-value:**  $2P(T \geq 9) = 2 \times \sum_{i=9}^{11} \binom{11}{i} 0.5^i 0.5^{11-i} = 2 \times 0.5^{11} \left[ \binom{11}{9} + \binom{11}{10} + \binom{11}{11} \right] = 0.065$
- **Decision:** The p-value is (slightly) larger than 0.05, so we do not reject  $H_0$ . Hence the data are consistent with the null hypothesis that platelet aggregation is not affected by smoking.

This result is in contrast to the t-test which had p-value of 0.016.



The p-value calculation is based on getting a result **as or more extreme** than the observed test statistic. We observed 9 positive differences, but the alternative hypothesis is "not equal to" so we also need to consider **extreme** at the other end of the distribution.

Hence, the p-value is  $P(T \geq 9) + P(T \leq 2) = 2P(T \geq 9)$  because the binomial distribution is symmetric when  $p = 0.5$ .





```
d = after - before
n = length(d)
t0 = sum(d > 0)
c(n, t0)
```

```
## [1] 11 9
```

```
2 * (1 - pbinom(t0 - 1, n, 0.5))
```

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

```
binom.test(t0, n, 0.5)
```

```
##
##      Exact binomial test
##
## data:  t0 and n
## number of successes = 9, number of trials = 11, p-value = 0.06543
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.4822441 0.9771688
## sample estimates:
## probability of success
##      0.8181818
```



Compare with the paired  $t$ -test from last week:

- **Hypothesis:**  $H_0: \mu_d = 0$  vs  $H_1: \mu_d \neq 0$

- **Assumptions:**  $D_i$  are iid\*  $N(\mu, \sigma^2)$ .

- **Test statistic:**  $T = \frac{\bar{D} - \mu_0}{S_d / \sqrt{n}}$ . Under  $H_0$ ,

$$T \sim t_{n-1}$$

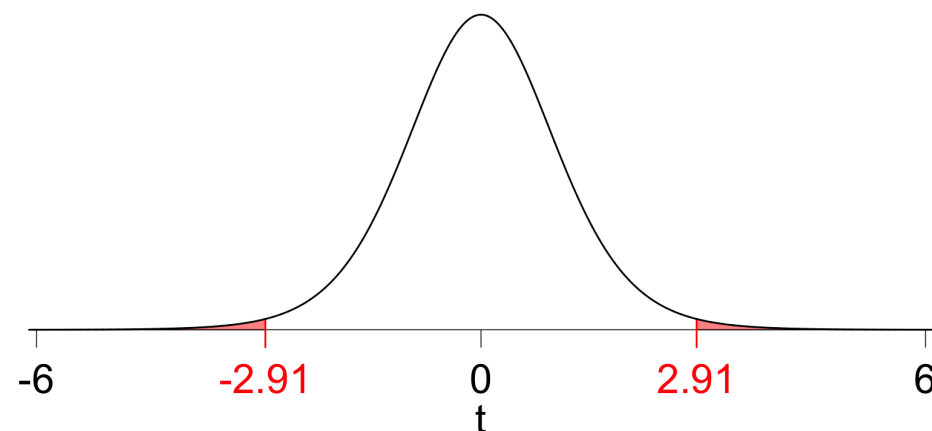
- **Observed test statistic:**

$$t_0 = \frac{8.45}{9.65 / \sqrt{11}} = 2.91$$

- **p-value:**  $2P(t_{10} \geq |2.91|) = 0.016$ .

- **Decision:** As the p-value is small, there is evidence against the null hypothesis. There is evidence that blood platelet aggregation levels increase after smoking.

Probability density function for  $T \sim t(10)$



\* iid is short for "independent and identically distributed"

# Further reading

Larsen and Marx (2012; section 14.2).

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.

Phipps, M. and M. Quine (2001). *A Primer Statistics*. 4th edition. Pearson Education Australia. ISBN: 1740096266.

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](https://doi.org/10.21105/joss.01686).