Non–parametric tests

MF9130E – Introductory Course in Statistics
08.05.2023

Chi Zhang
chi.zhang@medisin.uio.no
Oslo Center for Biostatistics and Epidemiology
Department of Biostatistics, UiO

Outline

8:30-9:15	Exploratory data analysis II
9:30-10:15	Transformations, non-parametric tests
Demontration & Practice	**Demonstration in R**

Lab notes for today:
(under *R Lab and Code* tab)

<u>EDA II</u>

<u>Non-parametric tests</u>

Tranformation

Instead of analysing the original data, x1, x2, …, analyse **transformed** data instead

Useful when transformed data are closer to a **normal** distribution

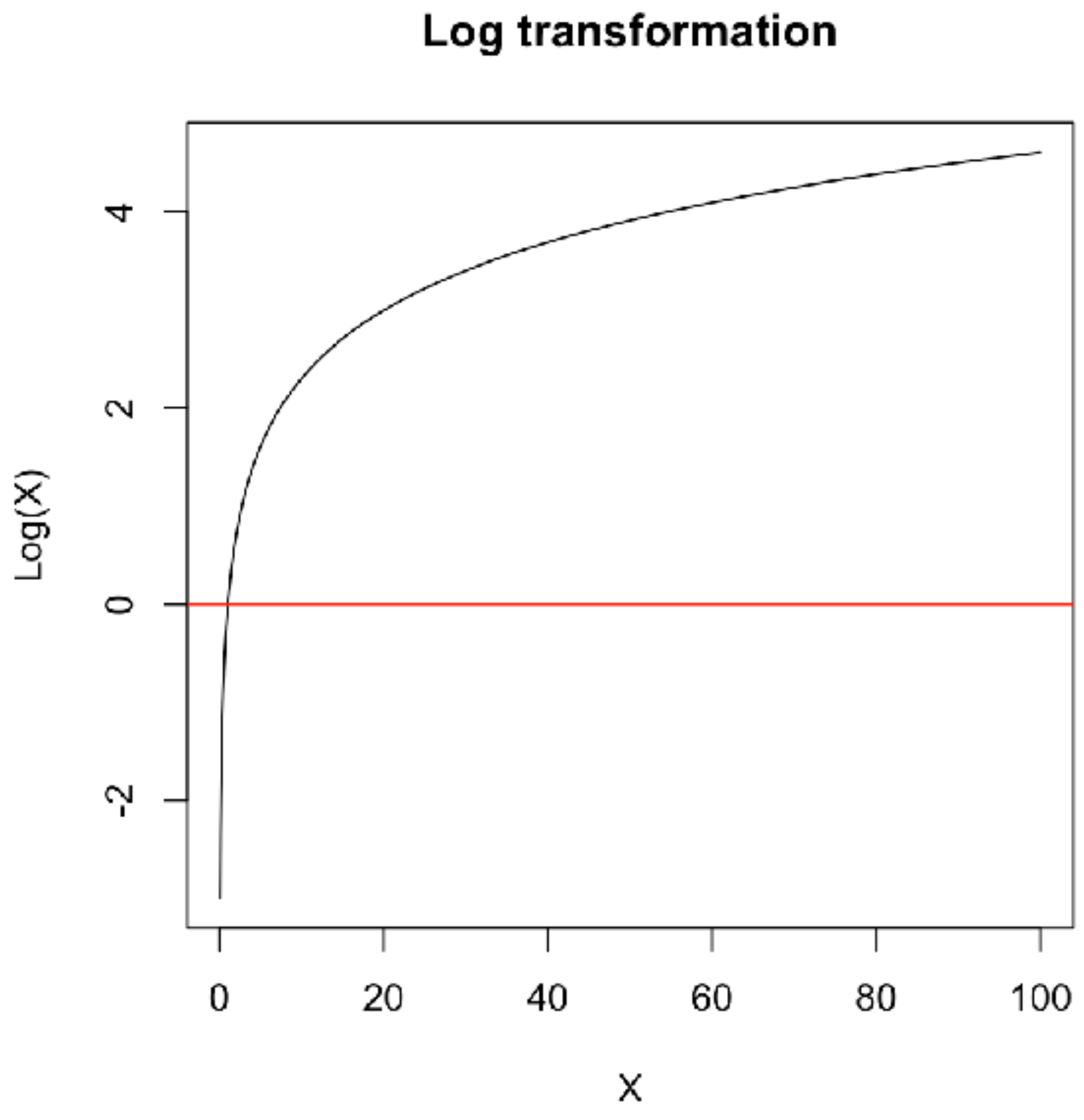
Transformation *could* give
- normally distributed data;
- more linear relationship
- more similar variances in different groups

Natural logarithm is commonly used.
Denoted as **ln**, or simply **log**

x_new = log(x)

log(e) = 1, where e = 2.718282
log(exp(x)) = x
log(1) = 0

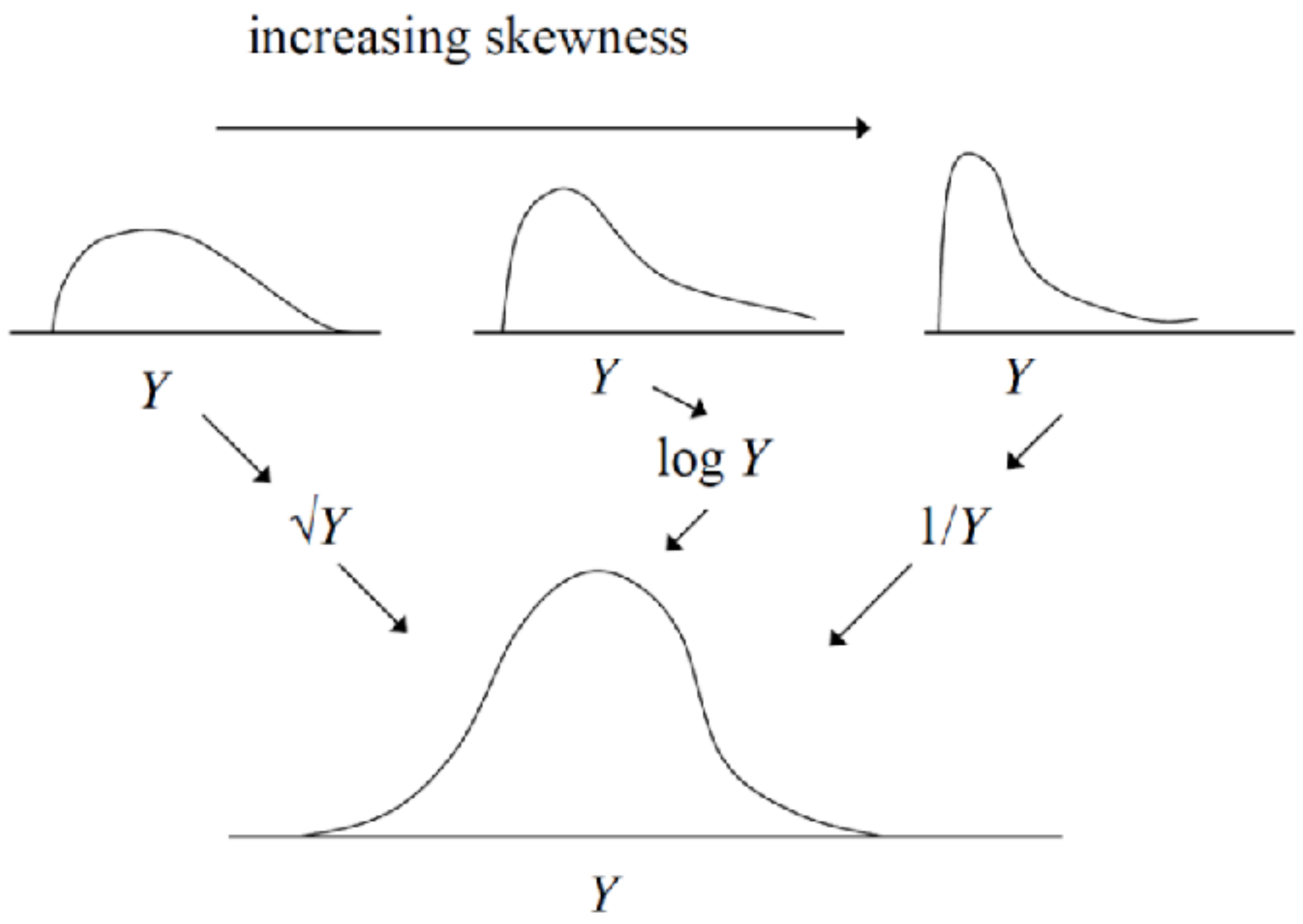
Can make **right-skewed** data (with large values) closer to normal



Tranformation

Tukey's Ladder of Powers

Power	Transformation	Name
3	y^3	Cubic
2	y^2	Square
1	**y (no transformation)**	**(Original data)**
1/2	y^(1/2) = sqrt(y)	Square root
	log(y)	Logarithm (Log)
-1	1/y	Reciprocal
-2	1/(y^2)	Reciprocal square
…		



There are other data transformation: e.g. Box-Cox transformation

Sometimes transformation are not enough. Choose alternative methods!

Non–parametric methods

Parametric methods/tests are based on **probability distributions**, have **parameters**

- Normal distribution: mean and variance;
- binomial distribution: n and p
- student-t and chi-square distribution: degree of freedom

Non-parametric methods/tests do not assume a specific parametric distribution

Signs, ranks are used; most useful for **small** datasets.

Aalen 8.8, Kirkwood and Sterne 30.2

Non-parametric methods
Sign test, Wilcoxon signed rank test for paired samples
Wilcoxon rank sum test for two indpendent samples

We have discussed the confidence interval for mean, and t-test for sample means (paired data, independent data)

Confidence interval for **median**

Test for paired data:
- Sign test
- **Wilcoxon signed-rank test**

Test for two independent samples:
- Mann-Whiteney test / **Wilcoxon rank-sum test**

You do not need to compute by hand. Use statistical program!

Dataset: length of hospital stay (liggetid)

Data collected at Ullevål hospital; 1139 observations, 21 variables
Main outcome of interest: **length of hosptal stay** (liggetid)
Here we compare length of stay based on: **stroke**

(Lab notes: non-parametric teste exercise 2)

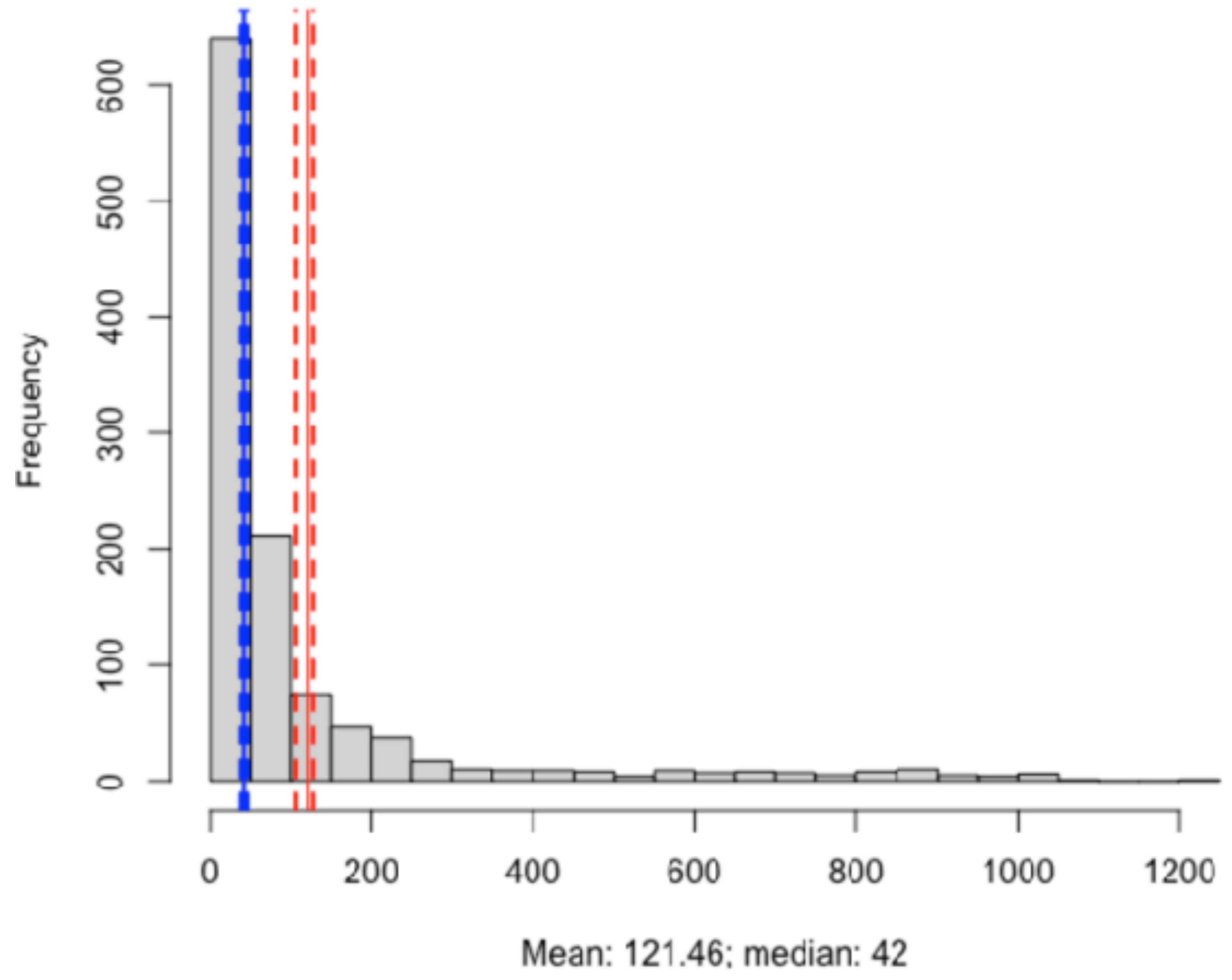
	faar	fmaan	fdag	innaar	innmaan	inndag	utaar	utmaan	utdag	kjoenn	kom_fra	slag	alder	liggetid	lnliggti	kom_fra2
327	1909	1	21	1985	11	28	85	12	20	mann	2	0	7[illegible]	22	3.0910425	1
328	1908	1	16	1985	12	3	85	12	17	kvinne	3	0	7[illegible]	14	2.6390573	0
329	1902	9	8	1985	12	4	85	12	17	kvinne	1	0	8[illegible]	13	2.5649494	0
330	1900	1	2	1985	12	5	85	12	16	mann	1	0	8[illegible]	11	2.3978953	0
331	1911	12	20	1985	12	12	87	3	9	kvinne	2	0	7[illegible]	452	6.1136822	1
332	1901	9	25	1985	12	13	87	1	22	kvinne	2	0	8[illegible]	405	6.0038871	1
333	1908	7	29	1985	12	13	85	12	19	kvinne	2	0	7[illegible]	6	1.7917595	1
334	1906	4	8	1985	12	20	87	1	1	kvinne	2	0	7[illegible]	377	5.9322452	1
335	1893	2	10	1985	12	23	87	1	21	kvinne	1	0	9[illegible]	394	5.9763509	0
336	1899	10	4	1986	1	3	86	1	29	mann	4	NA	8[illegible]	26	3.2580965	0
337	1893	12	23	1986	1	3	86	9	3	kvinne	2	NA	9[illegible]	243	5.4930614	1
338	1911	4	6	1986	1	3	86	2	3	kvinne	1	NA	7[illegible]	31	3.4339872	0
339	1908	8	21	1986	1	3	86	3	19	kvinne	1	NA	7[illegible]	75	4.3174881	0
340	1906	12	14	1986	1	7	86	4	2	kvinne	6	NA	8[illegible]	85	4.4426513	0
341	1913	10	26	1986	1	9	86	4	7	kvinne	2	NA	7[illegible]	88	4.4773368	1
342	1908	3	28	1986	1	9	86	3	14	kvinne	2	NA	7[illegible]	64	4.1588831	1
343	1905	9	17	1986	1	10	86	2	12	kvinne	5	NA	8[illegible]	33	3.4965076	0
344	1903	6	23	1986	1	10	86	1	17	mann	4	NA	8[illegible]	7	1.9459101	0
345	1895	2	12	1986	1	14	86	1	17	kvinne	1	NA	9[illegible]	3	1.0986123	0
346	1907	2	4	1986	1	14	86	4	1	kvinne	2	NA	7[illegible]	77	4.3438054	1
347	1909	9	4	1986	1	15	86	3	5	kvinne	2	NA	7[illegible]	49	3.8918203	1

Confidence interval for median

What is the **median** and 95% confidence interval for length of stay (liggetid)?
How does it compare to **mean** and 95% CI?

The lower and upper 95% confidence limits are the

$$\frac{n}{2} - 1.96\frac{\sqrt{n}}{2} \quad \text{and} \quad 1 + \frac{n}{2} + 1.96\frac{\sqrt{n}}{2}$$ -th ranked values




```
> n
[1] 1139
> lwr_rank <- n/2 - 1.96*(sqrt(n)/2)
> upr_rank <- 1 + n/2 + 1.96*(sqrt(n)/2)
> c(lwr_rank, upr_rank)
[1] 536.4259 603.5741
> c(quantile(los, lwr_rank/n), quantile(los, upr_rank/n))
47.09622% 52.99158%
       38        46
> # from package:
> # install.packages('DescTools')
> DescTools::MedianCI(los, conf.level = 0.95)
median lwr.ci upr.ci
    42     38     46
attr(,"conf.level")
[1] 0.956129
```

n=1139
536-th ranked value (47-th percentile): 38
603-th ranked value (52-th percentile): 42
Recall that median is 50th percentile: half data greater than median, half smaller!

Tests to compare 2 groups (continuous)

	Normal distributed	**Not Normal distributed**
One sample / Paired samples	One-sample t-test Paired-sample t-test	Sign test; **Wilcoxon signed-rank test**
Two independent samples	Two-sample t-test	**Mann-Whitney U / Wilcoxon rank-sum test**
Three + samples	ANOVA (not in this course)	Kruskal-Wallis (not in this course)

```
t.test(x1, x2, paired = T)

t.test(x, y, paired = F)

wilcox.test(x1, x2, paired = T)

wilcox.test(x, y, paired = F)
```

Rank based methods (intuition)

(Without going into too much technical details: we focus more on selecting the right method rather than computation)

Subj	Pre	Post	Difference	Sign	Rank
1	5260	3910	1350	+	6
2	5470	4220	1250	+	4
3	5640	3885	1755	+	10
4	6180	5160	1020	+	3
5	6390	5645	745	+	2
6	6515	4680	1835	+	11
7	6805	5265	1540	+	8.5
8	7515	5975	1540	+	8.5
9	7515	6790	725	+	1
10	8230	6900	1330	+	5
11	8770	7335	1435	+	7

Compare pre vs post values.
Paired samples (pre and post on the same subject)

Compute the difference: pre - post

Null hypothesis: no difference pre vs post.

Sign test: how likely are all 11 differences turn out positive, when we assume no difference?

Wilcoxon signed rank test:
compare positive ranks and negative ranks
non-parametric counterpart of **paired t-test**

Tedious to compute by hand: use R!

Rank based methods (intuition)

Group 1: 1.3, 1.5, 2.1, 3.2, 4.3
Group 2: 3.4, 4.9, 6.3, 7.1
Test if the two groups are different

Group 1	Group 2	Rank
1.3		1
1.5		2
2.1		3
3.2		4
	3.4	5
4.3		6
	4.9	7
	6.3	8
	7.1	9
Rank sum: 16	Rank sum: 29	

Mann-Whitney / Wilcoxon rank sum test

Two independent samples: compare n1 measurements from one group, with n2 measurements from another group

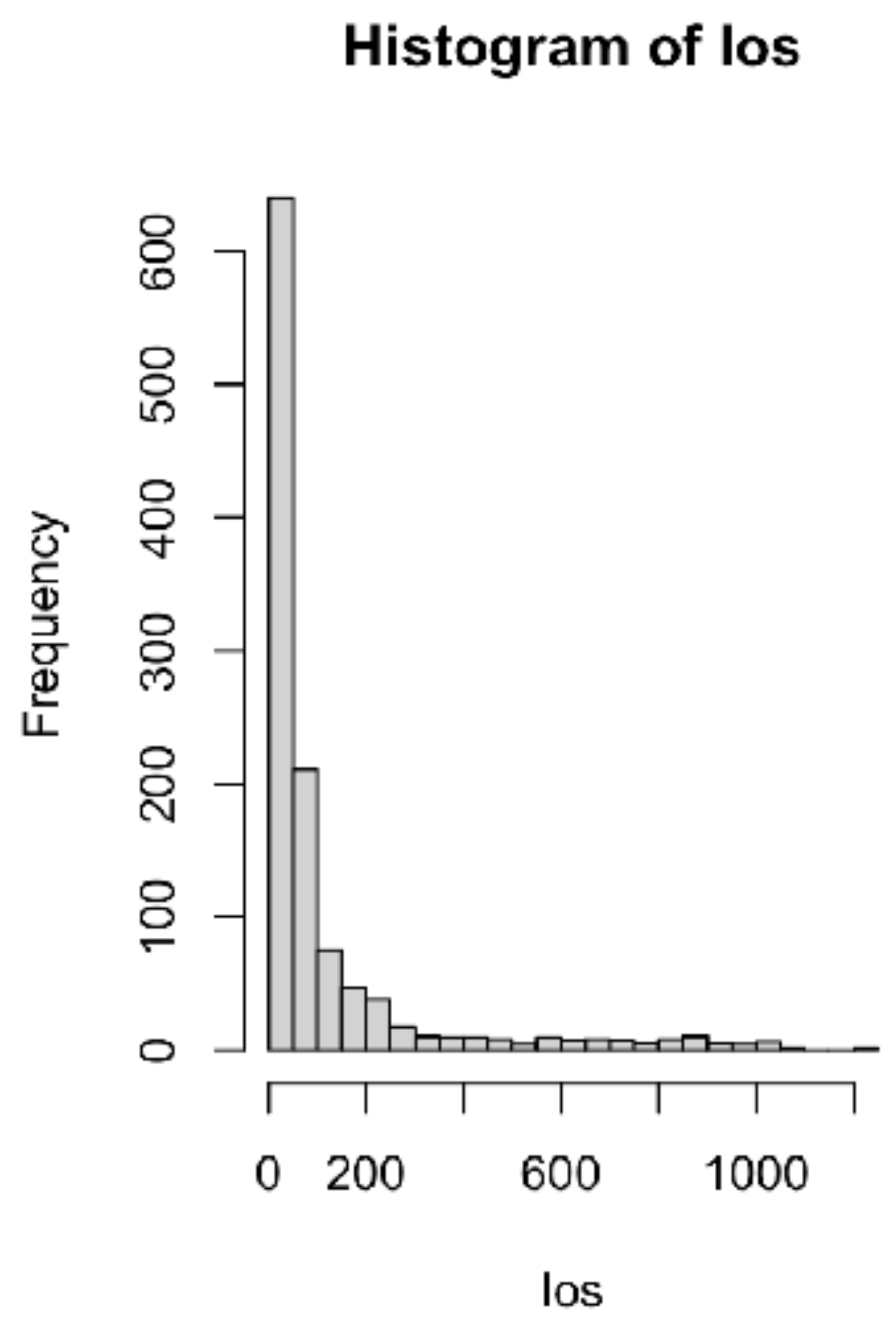
It is the non-parametric counterpart of **two-sample t-test**

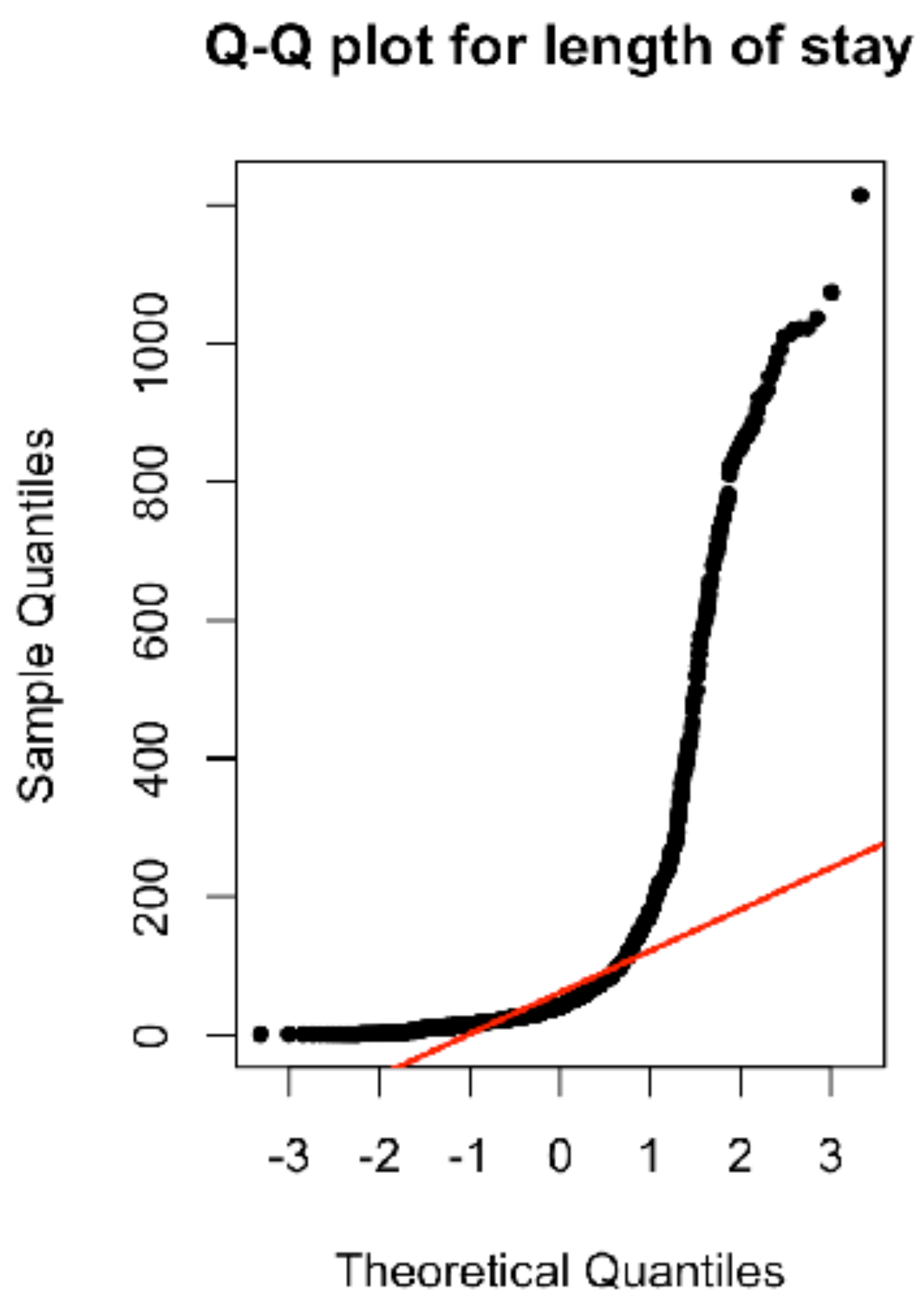
H0: the distributions of both groups are equal

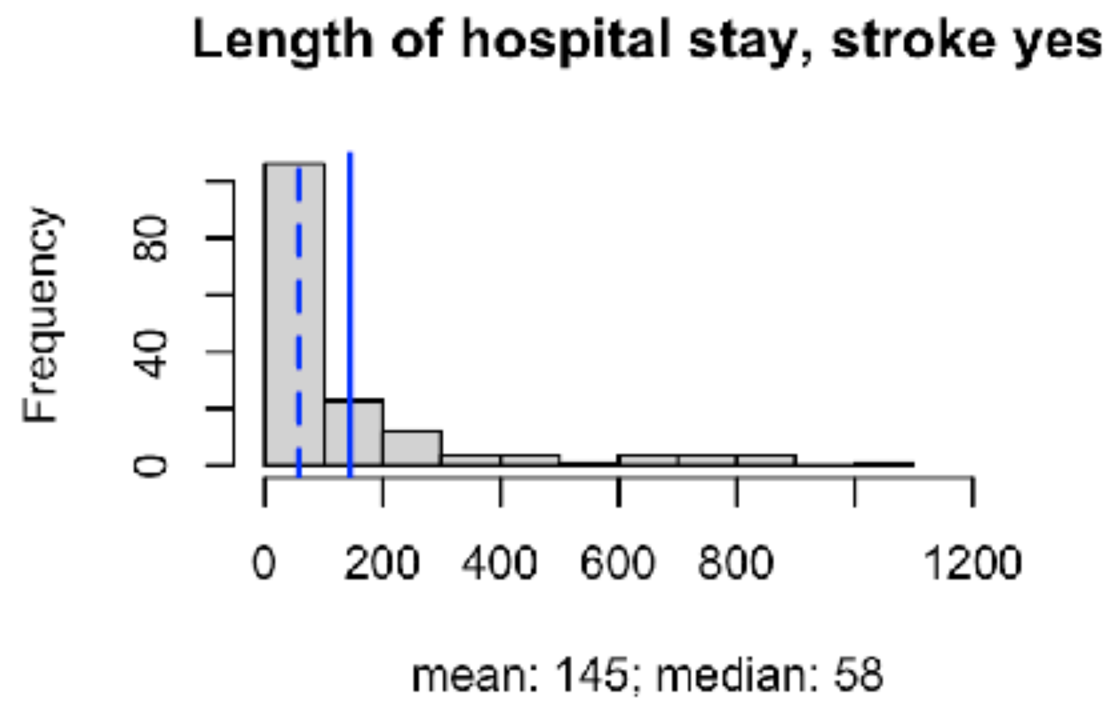
Test for equal medians, when two distributions can be assumed to be identical form

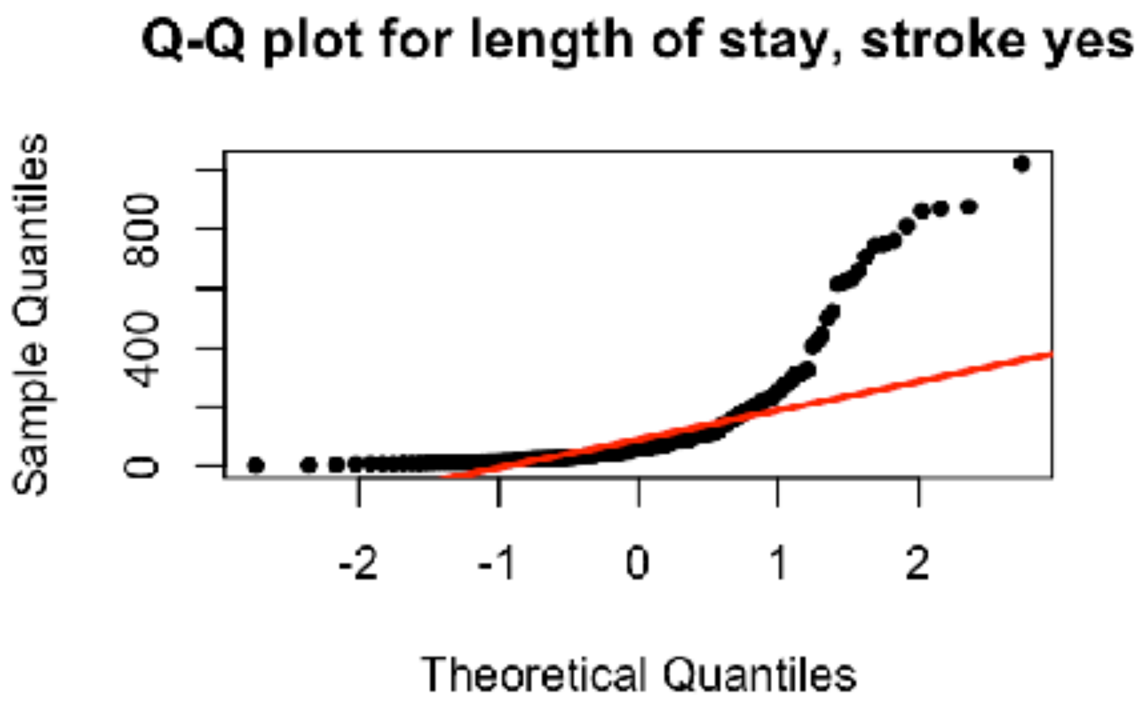
Length of stay data

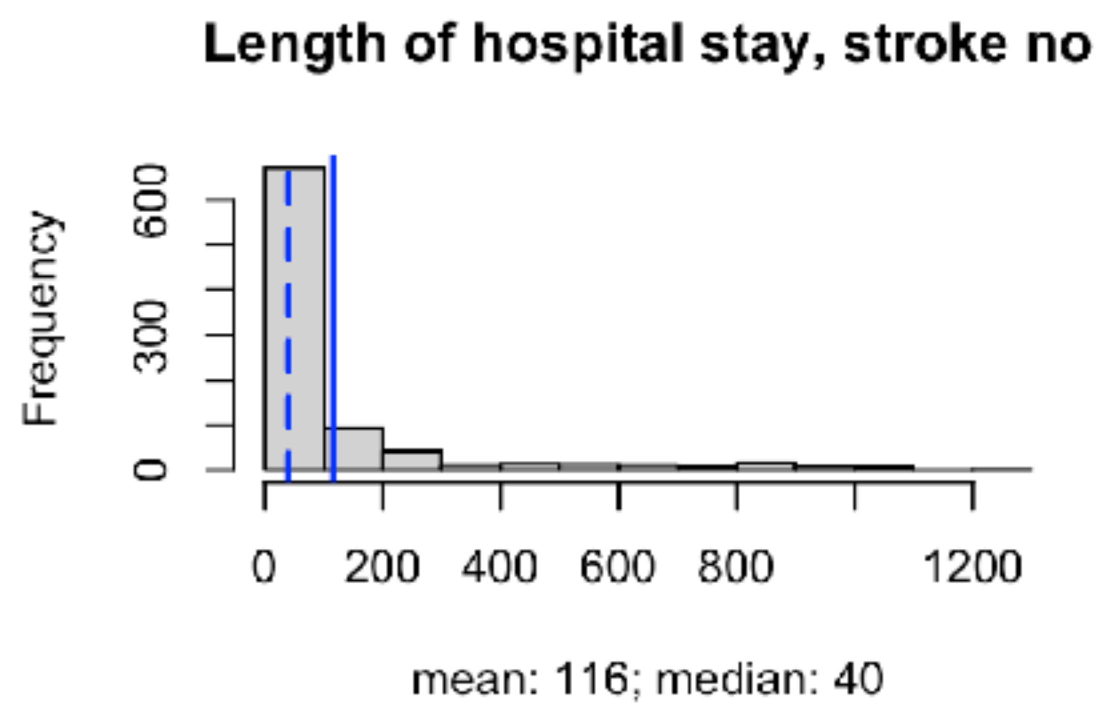
Check the distribution of los (length of stay)
Right skewed; far from normal

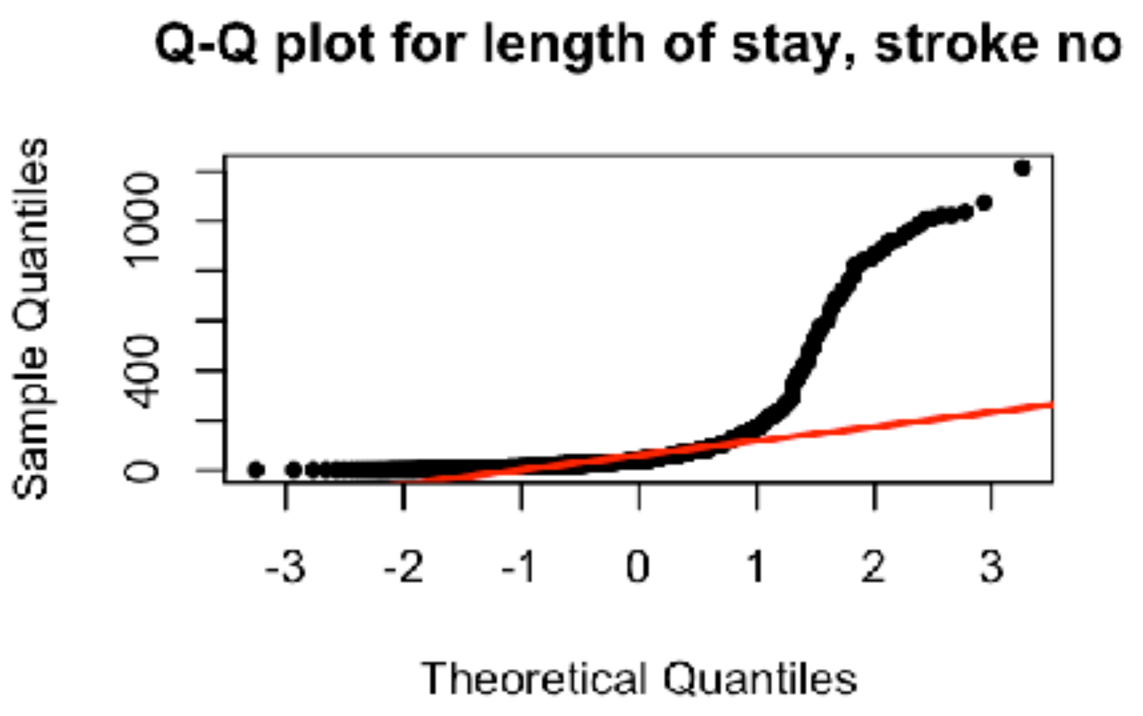












Length of stay data t–test vs Wilcoxon rank sum test

Similar to t-test, we need to know whether the samples are **paired or not**.

How do you know this?

(Here we have independent samples: different patients!)

Do a **Wilcoxon rank sum test** and **t-test** on the skewed data: notice the different conclusions

```
# non parametric test
# not matched; independent samples
# wilcoxon rank sum
wilcox.test(ligt_s1, ligt_s0)
```

```
	Wilcoxon rank sum test with continuity correction

data:  ligt_s1 and ligt_s0
W = 84060, p-value = 0.001361
alternative hypothesis: true location shift is not equal to 0
```

```
# t-test on skewed data
t.test(ligt_s1, ligt_s0)
```

```
	Welch Two Sample t-test

data:  ligt_s1 and ligt_s0
t = 1.5847, df = 220.8, p-value = 0.1145
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.879646 63.359853
sample estimates:
mean of x mean of y 
 145.0123  116.7722 
```

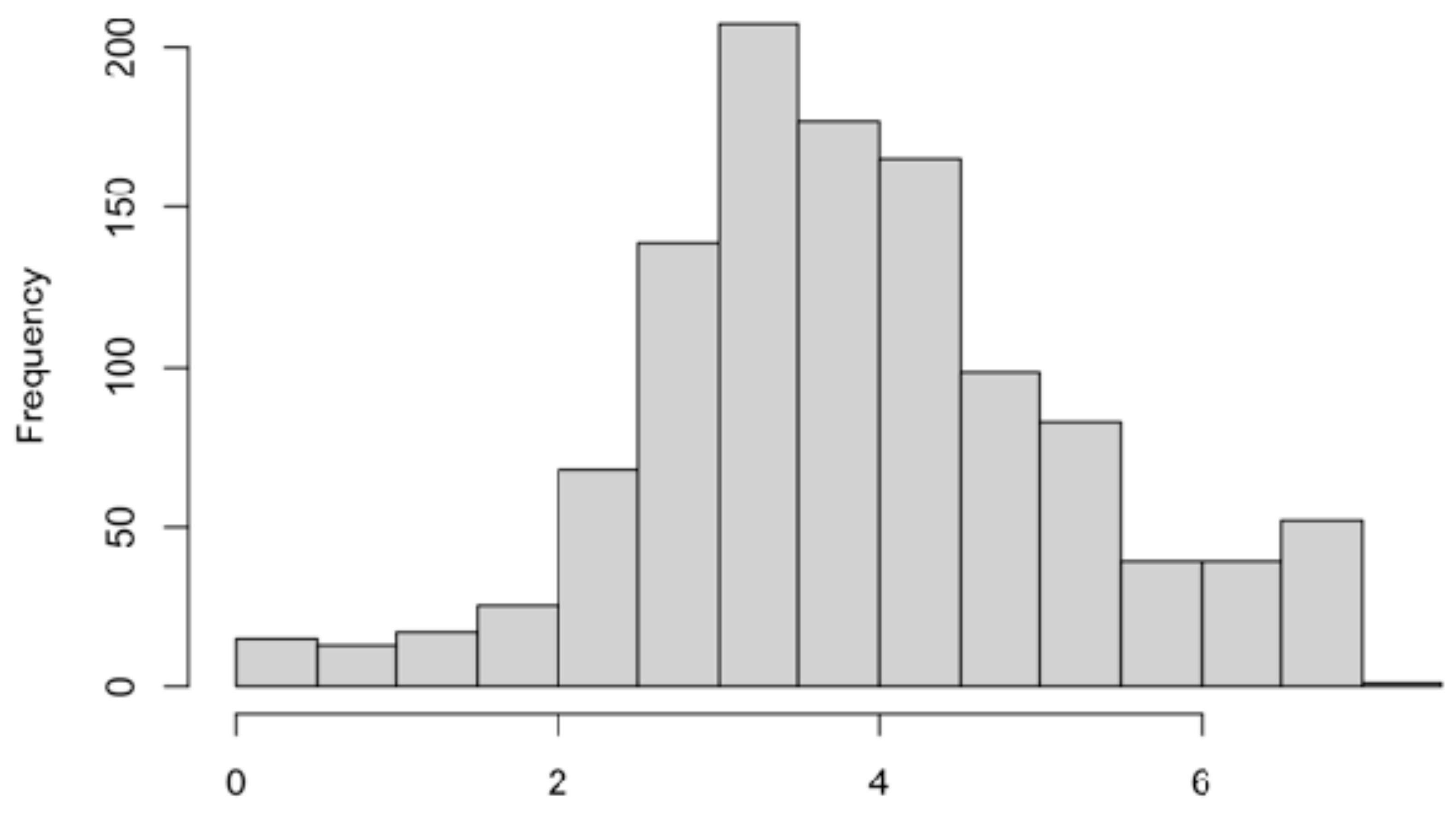
Length of stay data: log transform

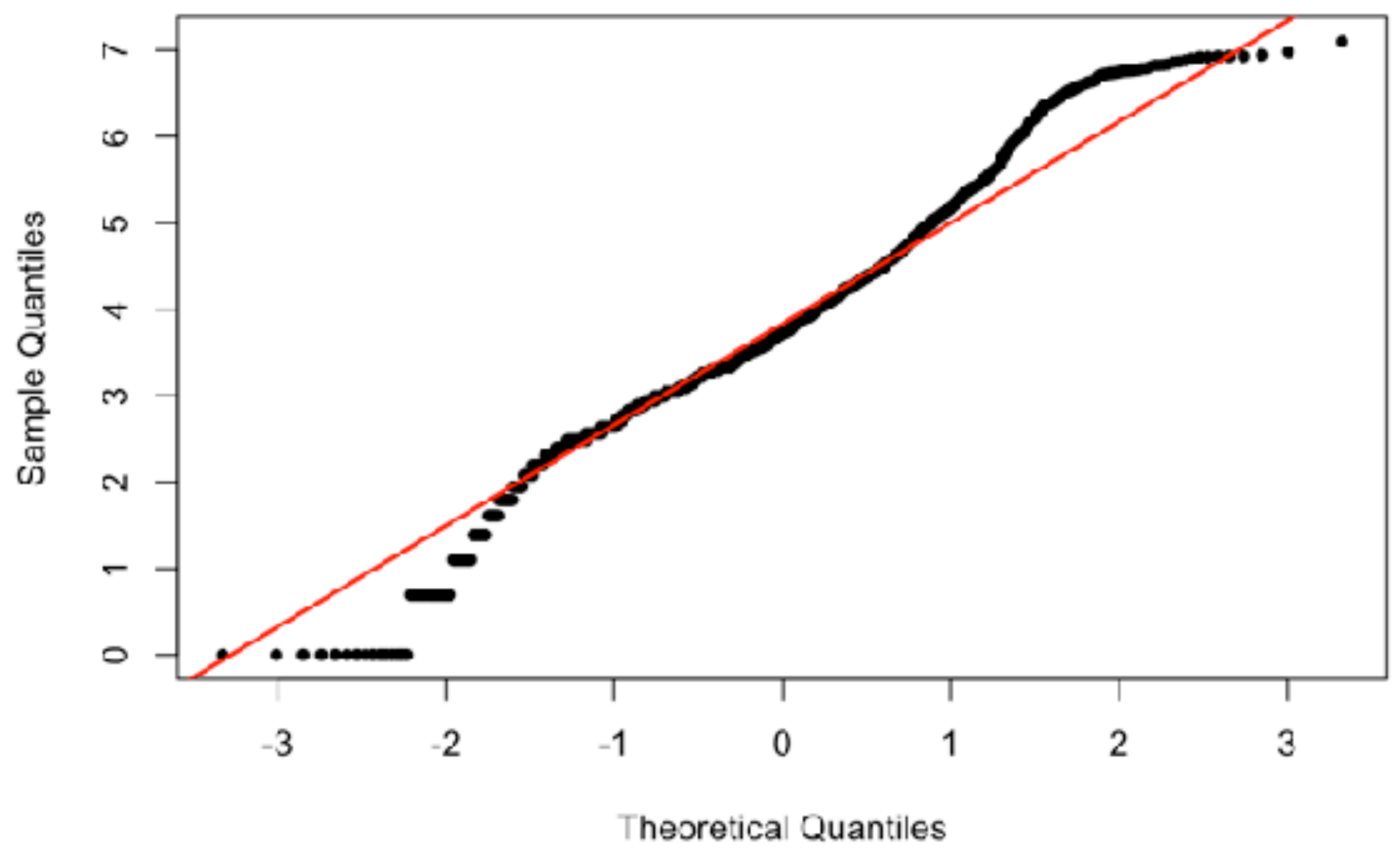
Now we try to repeat the same comparison, but on **log-transformed** length of stay data (lnliggti)
Here log-transformed data is already provided; if not, you can compute it by exp(liggetid)

(The interpretation will be different)

	faar	fmaan	fdag	innaar	innmaan	inndag	utaar	utmaan	utdag	kjoenn	kom_fra	slag	alder	liggetid	lnliggti	kom_fra2
327	1909	1	21	1985	11	28	85	12	20	mann	2	0	76	22	3.0910425	1
328	1908	1	16	1985	12	3	85	12	17	kvinne	3	0	77	14	2.6390573	0
329	1902	9	8	1985	12	4	85	12	17	kvinne	1	0	83	13	2.5649494	0
330	1900	1	2	1985	12	5	85	12	16	mann	1	0	85	11	2.3978953	0
331	1911	12	20	1985	12	12	87	3	9	kvinne	2	0	74	452	6.1136822	1
332	1901	9	25	1985	12	13	87	1	22	kvinne	2	0	84	405	6.0038871	1
333	1908	7	29	1985	12	13	85	12	19	kvinne	2	0	77	6	1.7917595	1
334	1906	4	8	1985	12	20	87	1	1	kvinne	2	0	79	377	5.9322452	1
335	1893	2	10	1985	12	23	87	1	21	kvinne	1	0	92	394	5.9763509	0
336	1899	10	4	1986	1	3	86	1	29	mann	4	NA	87	26	3.2580965	0
337	1893	12	23	1986	1	3	86	9	3	kvinne	2	NA	93	243	5.4930614	1
338	1911	4	6	1986	1	3	86	2	3	kvinne	1	NA	75	31	3.4339872	0
339	1908	8	21	1986	1	3	86	3	19	kvinne	1	NA	78	75	4.3174881	0
340	1906	12	14	1986	1	7	86	4	2	kvinne	6	NA	80	85	4.4426513	0
341	1913	10	26	1986	1	9	86	4	7	kvinne	2	NA	73	88	4.4773368	1
342	1908	3	28	1986	1	9	86	3	14	kvinne	2	NA	78	64	4.1588831	1
343	1905	9	17	1986	1	10	86	2	12	kvinne	5	NA	81	33	3.4965076	0
344	1903	6	23	1986	1	10	86	1	17	mann	4	NA	83	7	1.9459101	0
345	1895	2	12	1986	1	14	86	1	17	kvinne	1	NA	91	3	1.0986123	0
346	1907	2	4	1986	1	14	86	4	1	kvinne	2	NA	79	77	4.3438054	1
347	1909	9	4	1986	1	15	86	3	5	kvinne	2	NA	77	49	3.8918203	1

Length of stay data: log transform






```r
# separate data
logligt_s1 <- loglos[which(!is.na(stroke) & stroke == 1)]
logligt_s0 <- loglos[which(!is.na(stroke) & stroke == 0)]

t.test(logligt_s1, logligt_s0, paired = F)
```

```
    Welch Two Sample t-test

data:  logligt_s1 and logligt_s0
t = 3.4218, df = 237.88, p-value = 0.0007321
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1587474 0.5895572
sample estimates:
mean of x mean of y
 4.180085  3.805933
```

Antibody data (Exercise 3)

20 samples: before and after measurements -> paired data

A report summarized the results as **t = 1.8, p>0.05**

Is it the correct conclusion?

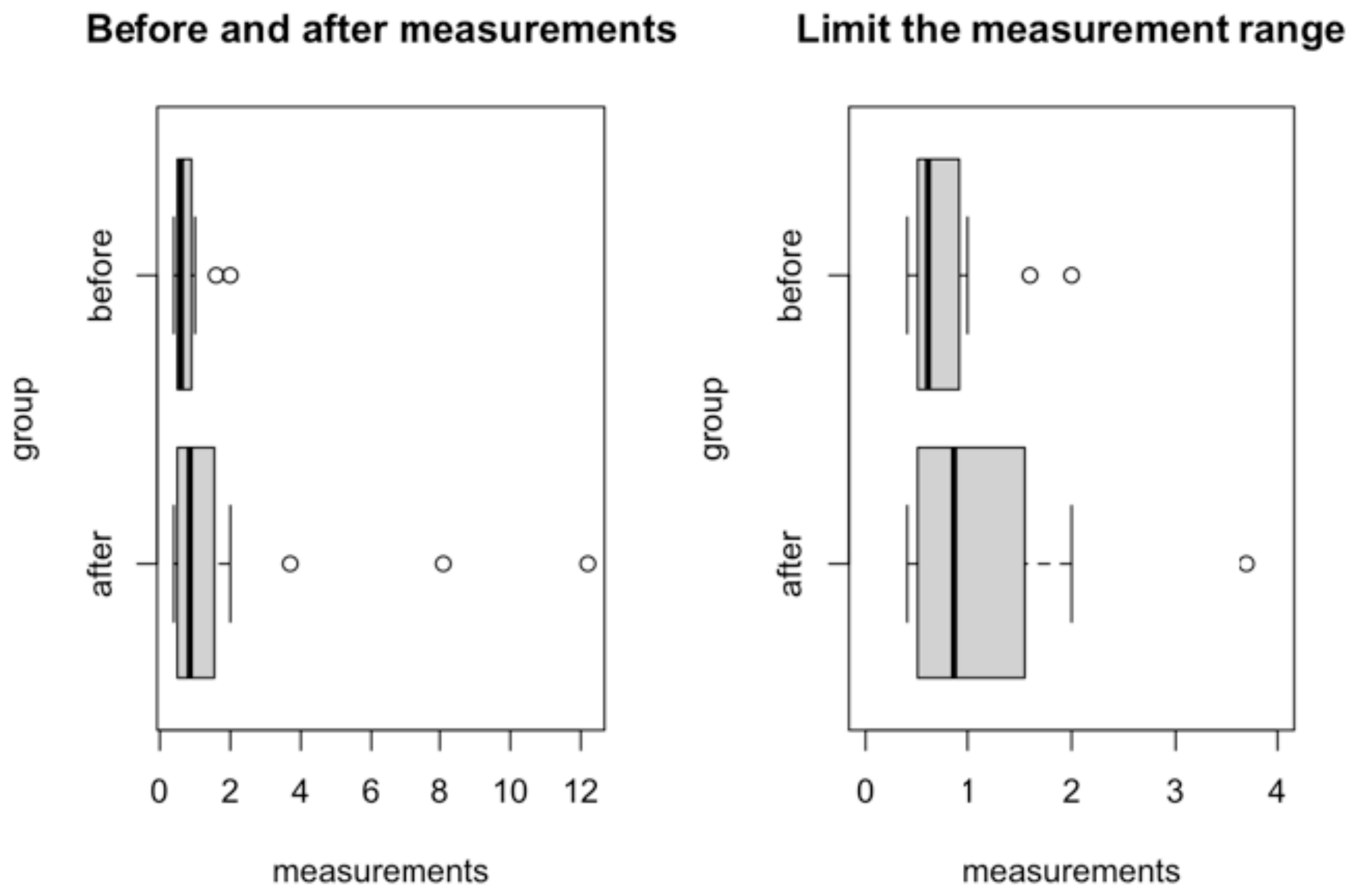
Do some exploratory analysis, and reproduce the result from the t-test.

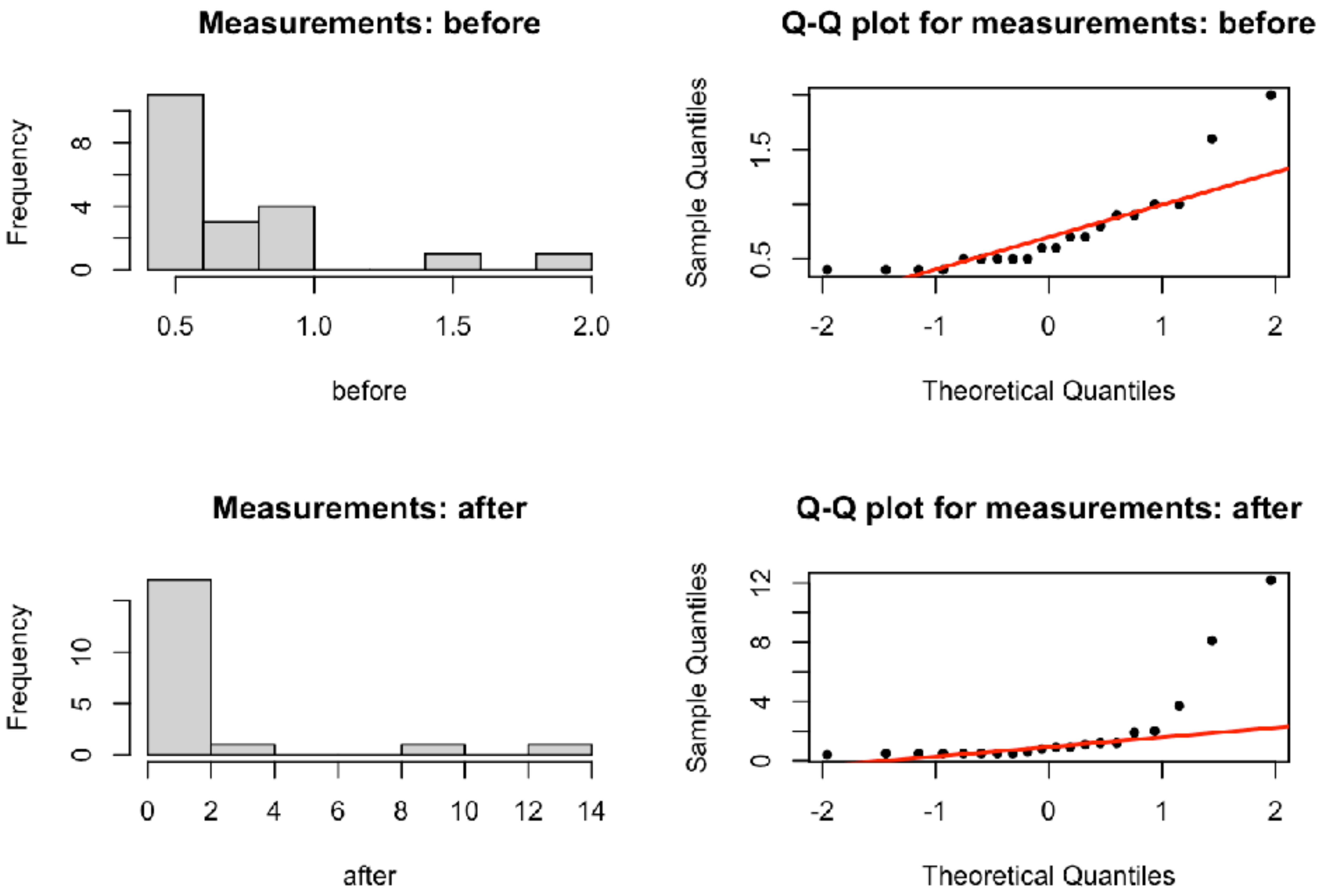
	before	after
1	0.4	0.4
2	0.4	0.5
3	0.4	0.5
4	0.4	0.9
5	0.5	0.5
6	0.5	0.5
7	0.5	0.5
8	0.5	0.5
9	0.5	0.5
10	0.6	0.6
11	0.6	12.2
12	0.7	1.1
13	0.7	1.2
14	0.8	0.8
15	0.9	1.2
16	0.9	1.9
17	1.0	0.9
18	1.0	2.0
19	1.6	8.1
20	2.0	3.7

Antibody data (Exercise 3)

Box plot: before and after suggests difference in distributions

Histogram and QQplot indicates non-normality





Antibody data (Exercise 3)

```r
# reproduce the result from t-test
t.test(antibody$before, antibody$after, paired = T)
```

```
	Paired t-test

data:  antibody$before and antibody$after
t = -1.8498, df = 19, p-value = 0.07996
alternative hypothesis: true mean difference is not equal to 0
95 percent confidence interval:
 -2.5151563  0.1551563
sample estimates:
mean difference 
          -1.18 
```

```r
# what happens if you use independent two-samples test?
wilcox.test(before, after, paired = F) # p = 0.14
```

```
Warning in wilcox.test.default(before, after, paired = F): cannot compute exact
p-value with ties

	Wilcoxon rank sum test with continuity correction

data:  before and after
W = 146.5, p-value = 0.1454
alternative hypothesis: true location shift is not equal to 0
```

```r
# matched data (paired)
# one-sample test (signed rank)

wilcox.test(before, after, paired = T)
```

```
Warning in wilcox.test.default(before, after, paired = T): cannot compute exact
p-value with ties

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

	Wilcoxon signed rank test with continuity correction

data:  before and after
V = 2, p-value = 0.00412
alternative hypothesis: true location shift is not equal to 0
```

Be careful with what test you use!

Do EDA and plot your data!