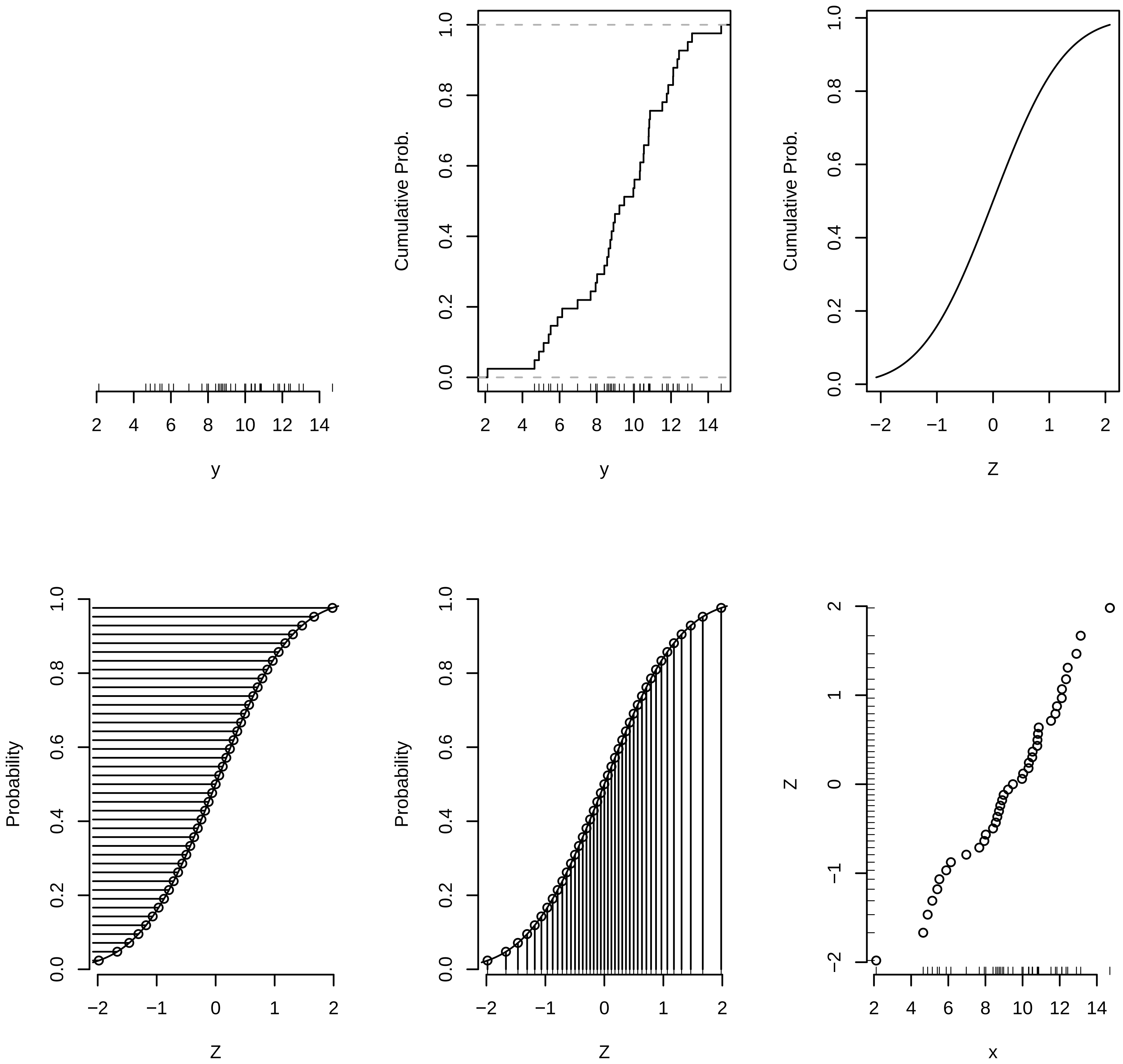
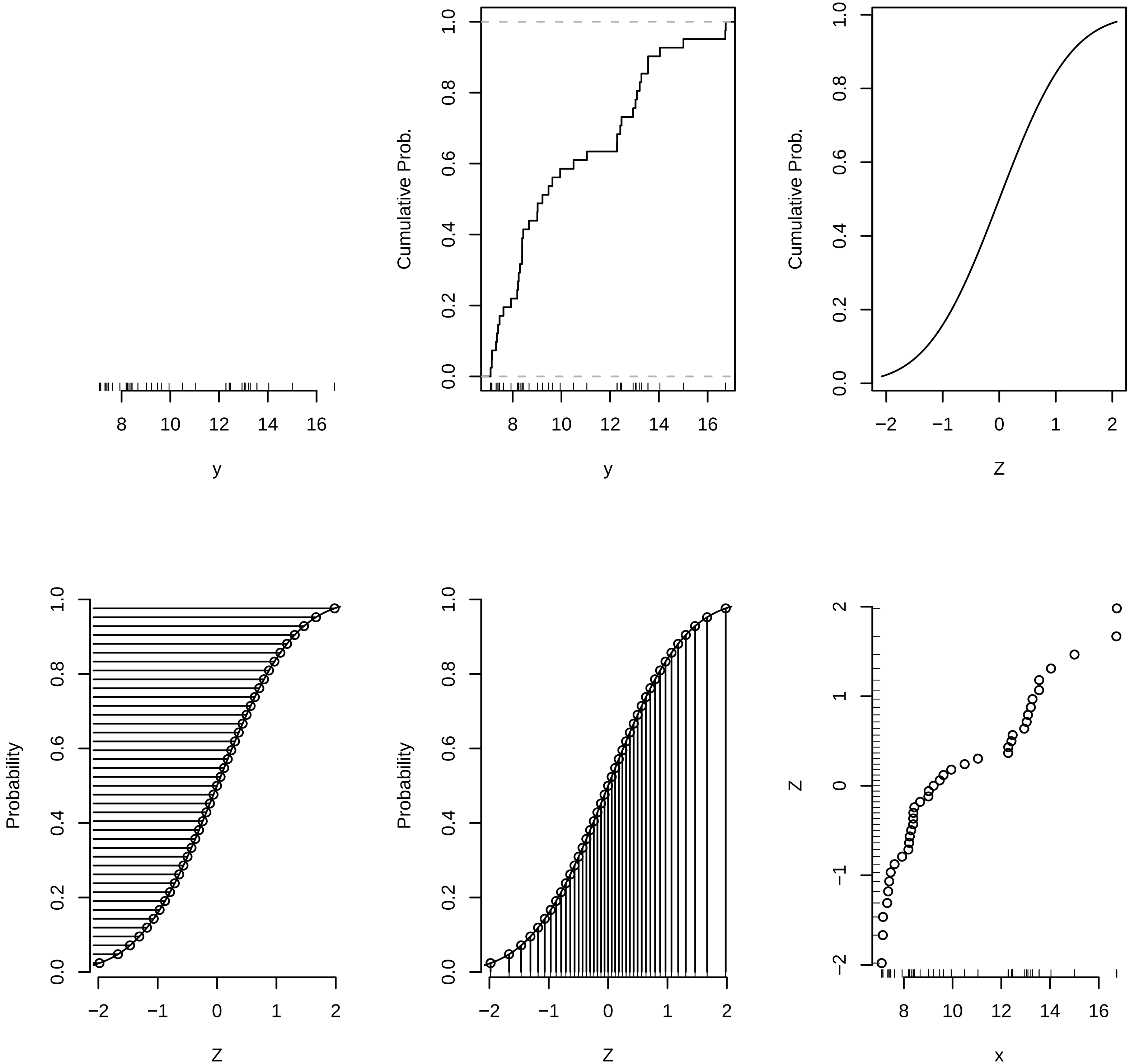
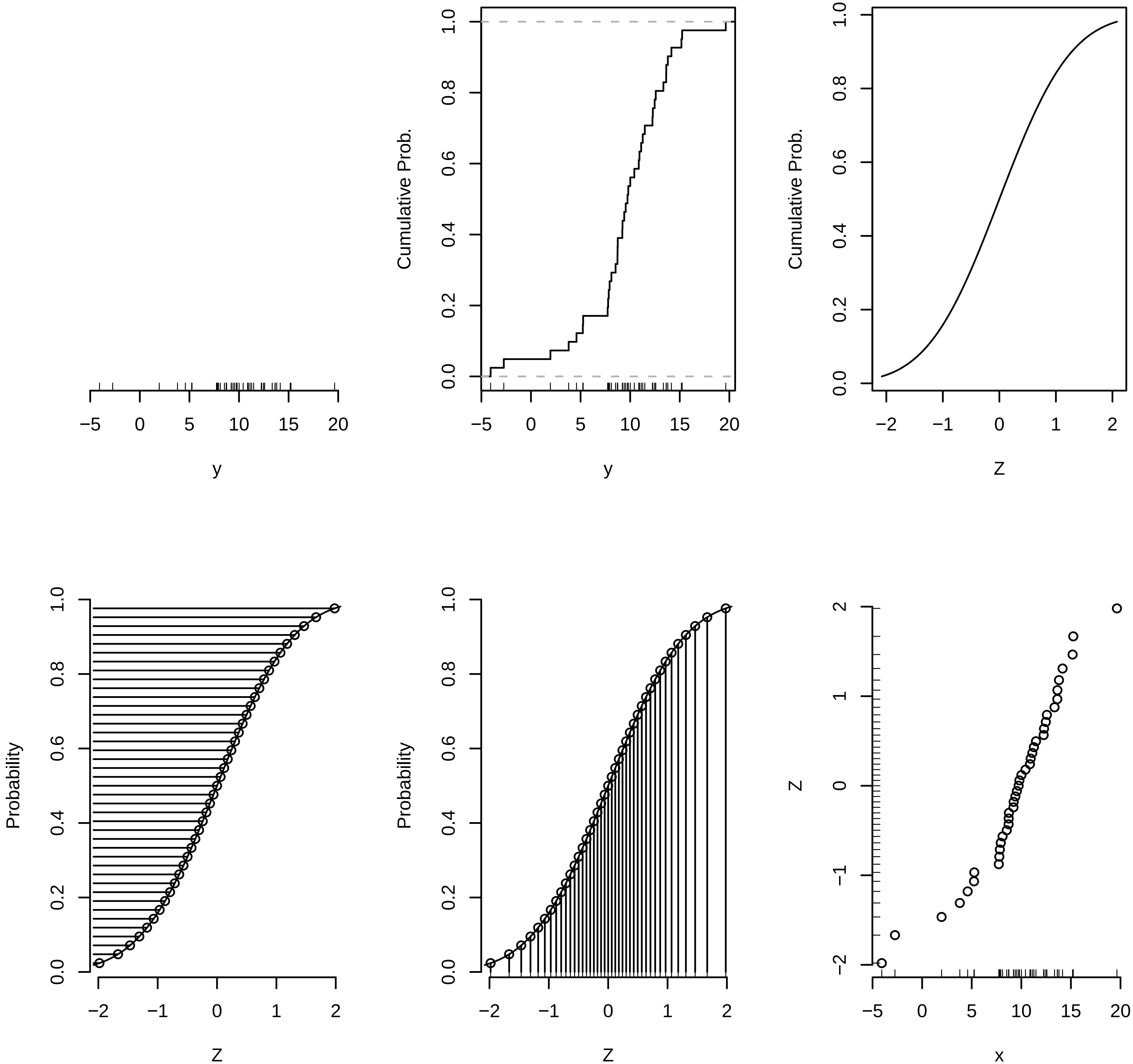


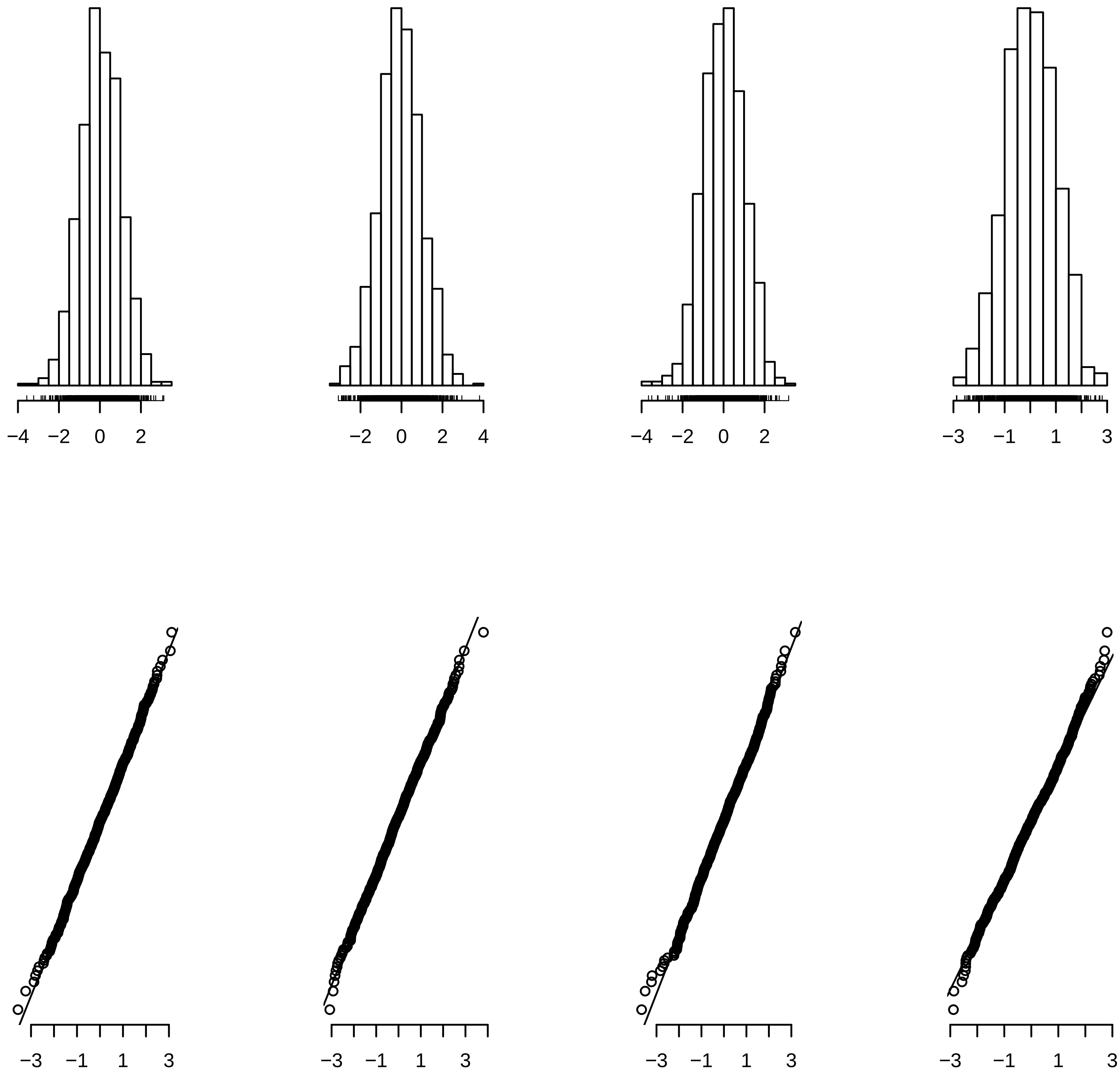
y
Cumulative Prob.
y
Cumulative Prob.
Z
Probability
Z
Probability
Z
Z
x



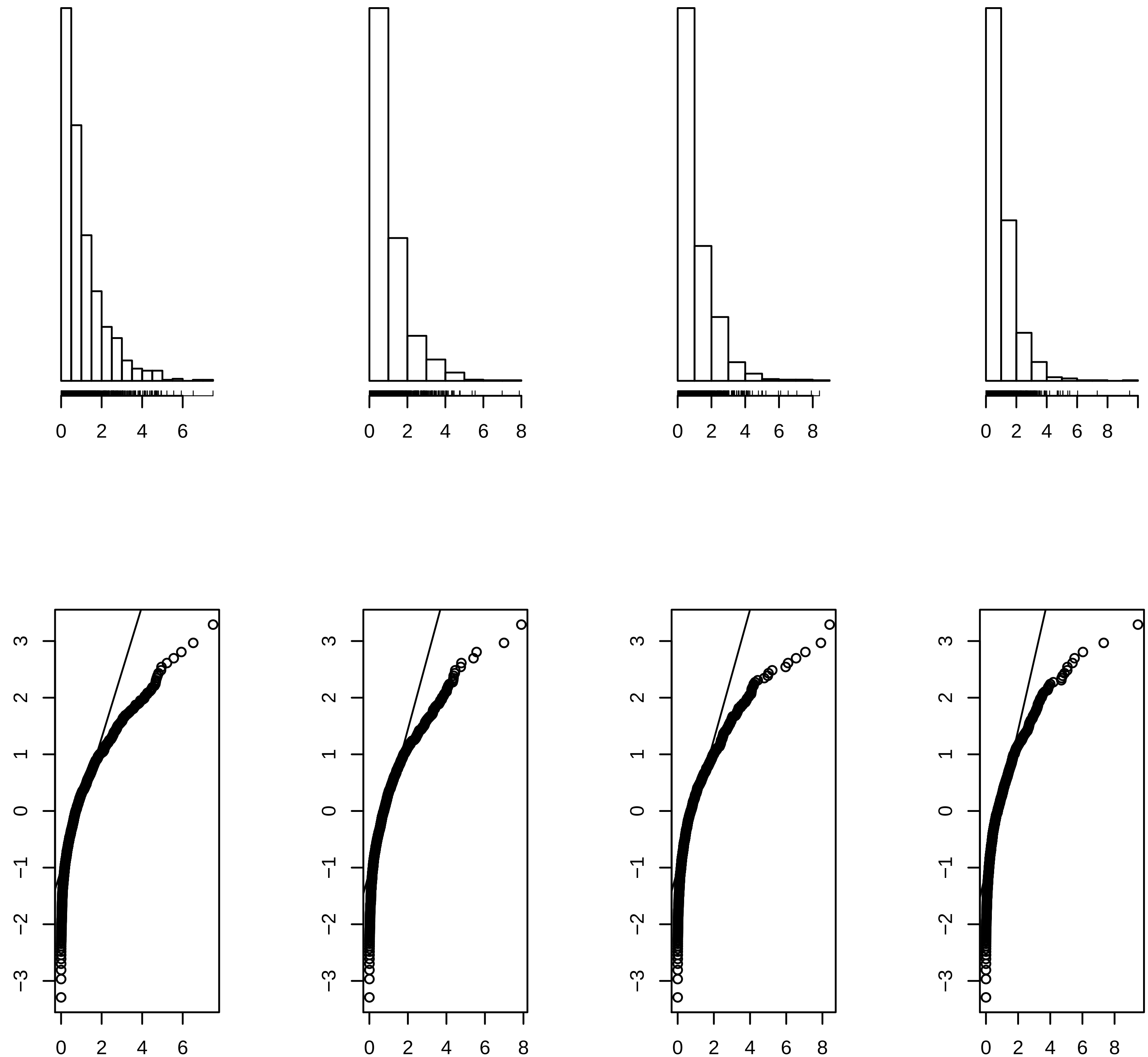
8 10 12 14 16
y
Cumulative Prob.
0.0 0.2 0.4 0.6 0.8 1.0
8 10 12 14 16
y
Cumulative Prob.
0.0 0.2 0.4 0.6 0.8 1.0
−2 −1 0 1 2
Z
Probability
0.0 0.2 0.4 0.6 0.8 1.0
−2 −1 0 1 2
Z
Probability
0.0 0.2 0.4 0.6 0.8 1.0
−2 −1 0 1 2
Z
Z
−2 −1 0 1 2
8 10 12 14 16
x



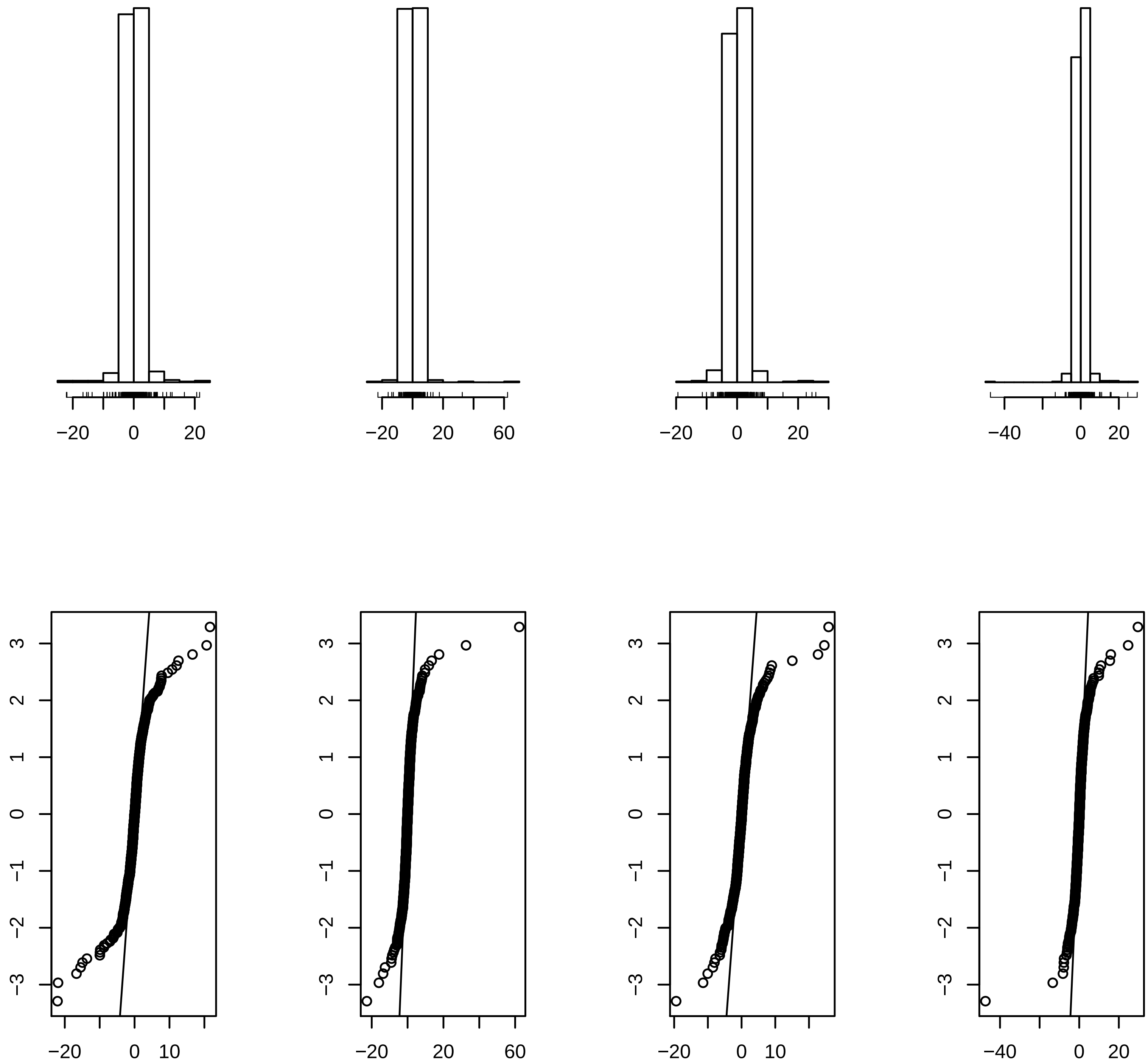
−5 0 5 10 15 20
y
Cumulative Prob.
0.0 0.2 0.4 0.6 0.8 1.0
−5 0 5 10 15 20
y
Cumulative Prob.
0.0 0.2 0.4 0.6 0.8 1.0
−2 −1 0 1 2
Z
Probability
0.0 0.2 0.4 0.6 0.8 1.0
−2 −1 0 1 2
Z
Probability
0.0 0.2 0.4 0.6 0.8 1.0
−2 −1 0 1 2
Z
Z
−2 −1 0 1 2
−5 0 5 10 15 20
x



−4 −2 0 2
−2 0 2 4
−4 −2 0 2
−3 −1 1 3
−3 −1 1 3
−3 −1 1 3
−3 −1 1 3
−3 −1 1 3



0 2 4 6
0 2 4 6 8
0 2 4 6 8
0 2 4 6 8
−3 −2 −1 0 1 2 3
0 2 4 6
−3 −2 −1 0 1 2 3
0 2 4 6 8
−3 −2 −1 0 1 2 3
0 2 4 6 8
−3 −2 −1 0 1 2 3
0 2 4 6 8



−20
0
20
−20
20
60
−20
0
20
−40
0
20
3
2
1
0
−1
−2
−3
−20
0
10
−20
20
60
−20
0
10
−40
0
20

Do antibiotic-resitant bacteria grow more slowly in the absence of antibiotic?

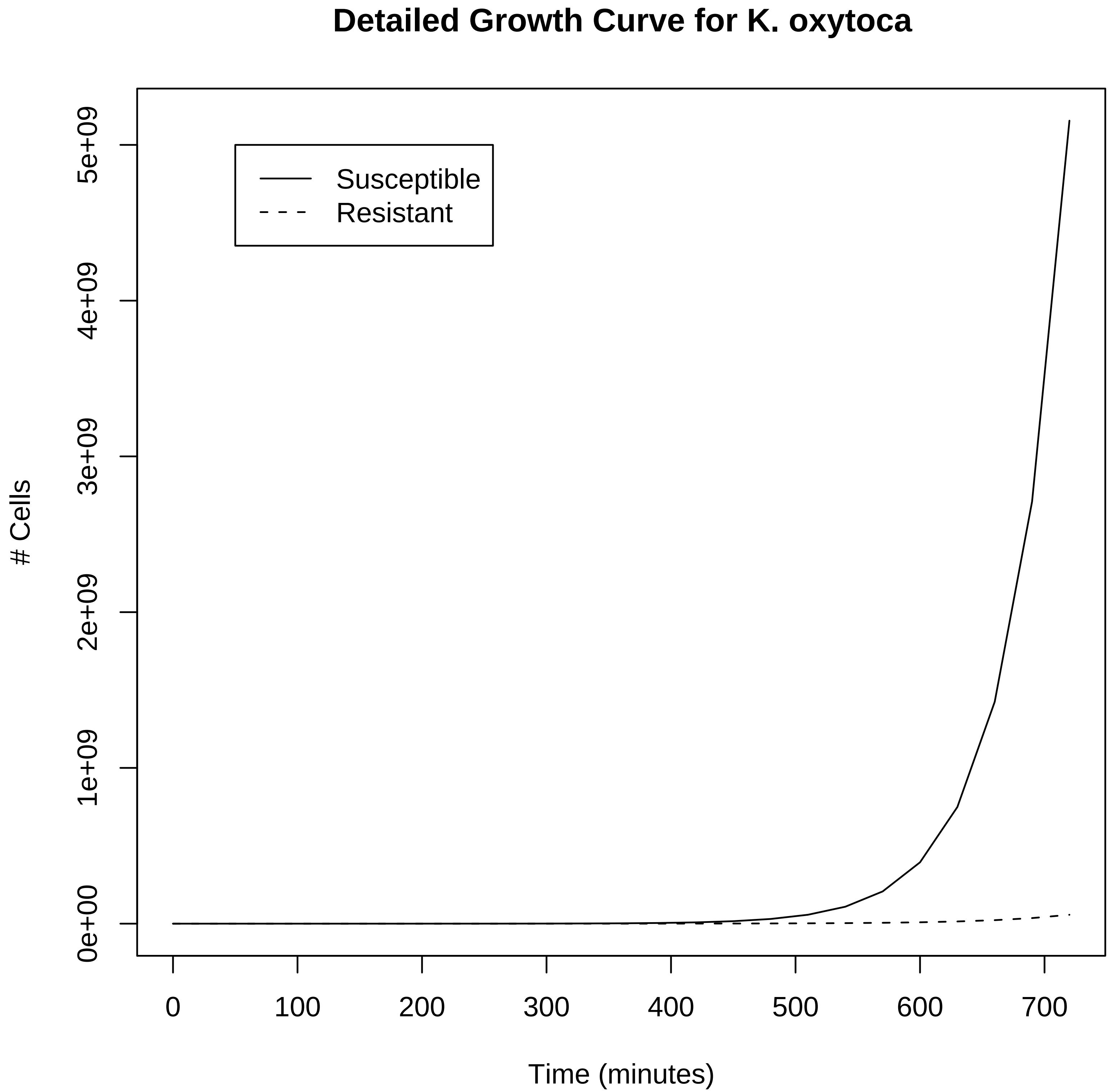
Microbiologists measured the ratio of bacteria before and after 12 hours of growth. They compared antibiotic resistant forms of 8 different bacteria (the next slides do not show real data)

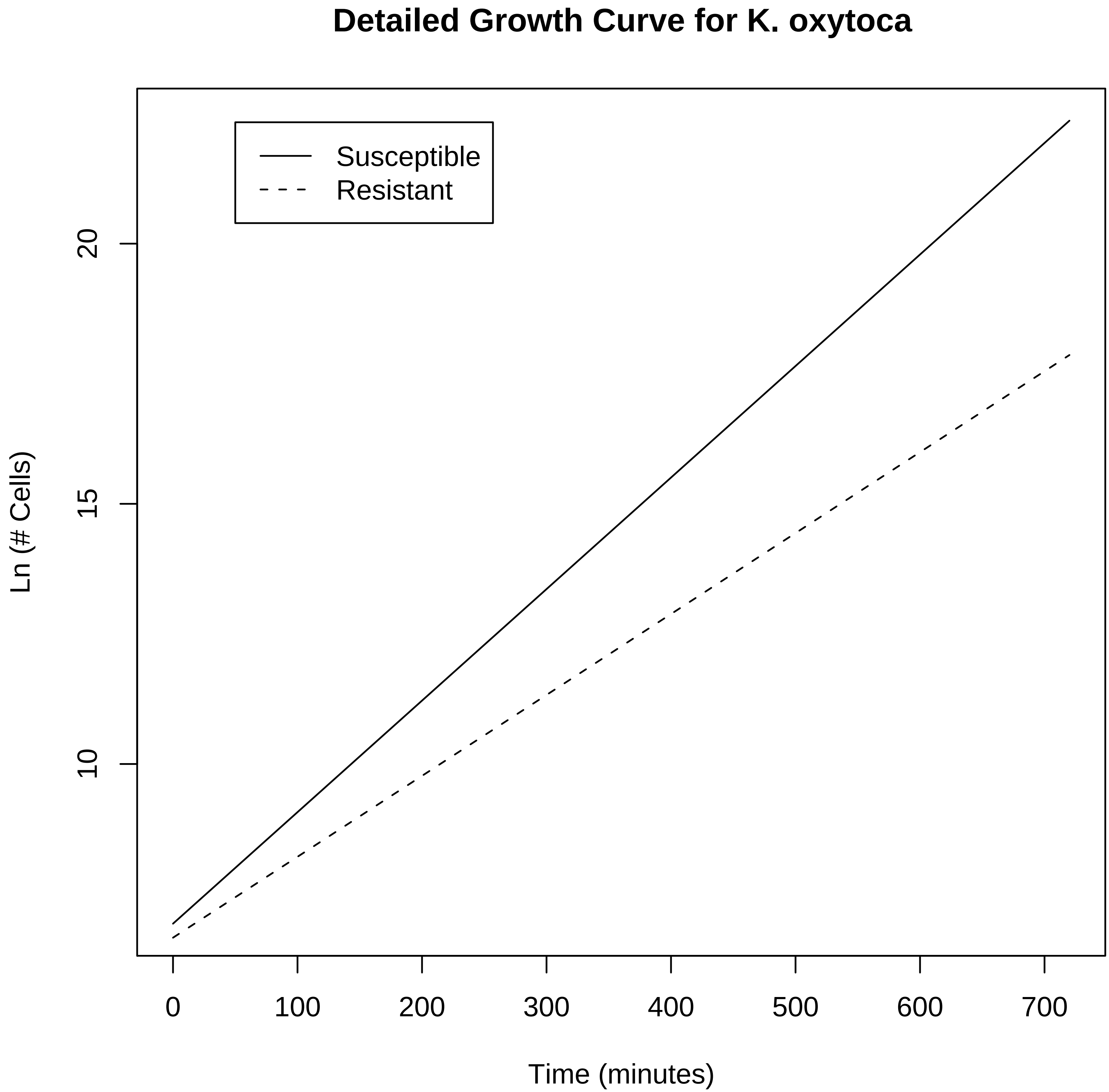
Data are the # cells after 12 hours growth/# cells before 12 hours growth (and then divided by 1 million).

	# cells before	# cells after	"growth" ratio / 1,000,000
Susceptible *K. oxytoca*	1,024	5.15×10^9	5.03
Resistant *K. oxytoca*	782	5.47×10^6	0.07
Susceptible *S. dysenteriae*	1,543	1.72×10^9	11.18
Resistant *S. dysenteriae*	2,237	9.62×10^7	4.96
...			

Data as ratios of the number of cells after growth (divided by 1 million):

	# cells ratio non-resistant	# cells ratio resistant	Differences
E. coli	5.93	7.26	-1.33
S. enterica	2.81	0.03	2.78
K. oxytoca	5.03	0.07	4.96
S. dysenteriae	11.18	0.43	10.75
P. vulgaris	2.70	0.35	2.35
S. marcescens	2.90	1.47	1.43
C. freundii	1.64	1.58	0.07
C. koseri	0.38	0.30	0.09


Detailed Growth Curve for K. oxytoca
Susceptible
Resistant
Cells
0e+00
1e+09
2e+09
3e+09
4e+09
5e+09
0
100
200
300
400
500
600
700
Time (minutes)


Detailed Growth Curve for K. oxytoca
Susceptible
Resistant
Ln (# Cells)
10
15
20
0
100
200
300
400
500
600
700
Time (minutes)

Same data, but measured in doubling time:

	doubling time non-resistant	doubling time resistant	Differences
E. coli	32.00	31.59	0.41
S. enterica	33.61	47.73	-14.12
K. oxytoca	32.34	44.57	-12.23
S. dysenteriae	30.75	38.46	-7.71
P. vulgaris	33.70	39.09	-5.39
S. marcescens	33.54	35.15	-1.61
C. freundii	34.87	34.97	-0.10
C. koseri	38.83	39.61	-0.78
	$\bar{y}_1 = 33.70$ $s_1^2 = 5.8802$ $s_1 = 2.4249$	$\bar{y}_2 = 38.90$ $s_2^2 = 27.6850$ $s_2 = 5.2617$	$\bar{d} = -5.19$ $s_d^2 = 32.1756$ $s_d = 5.6724$ $s_p^2 = 16.7826$

$$r = \frac{\#\text{cells after 12 hours}}{\#\text{ cells before}} = \frac{S}{S_0}$$

If t is the number of minutes takes to double, then:

$$S = S_0 2^{(720/t)}$$

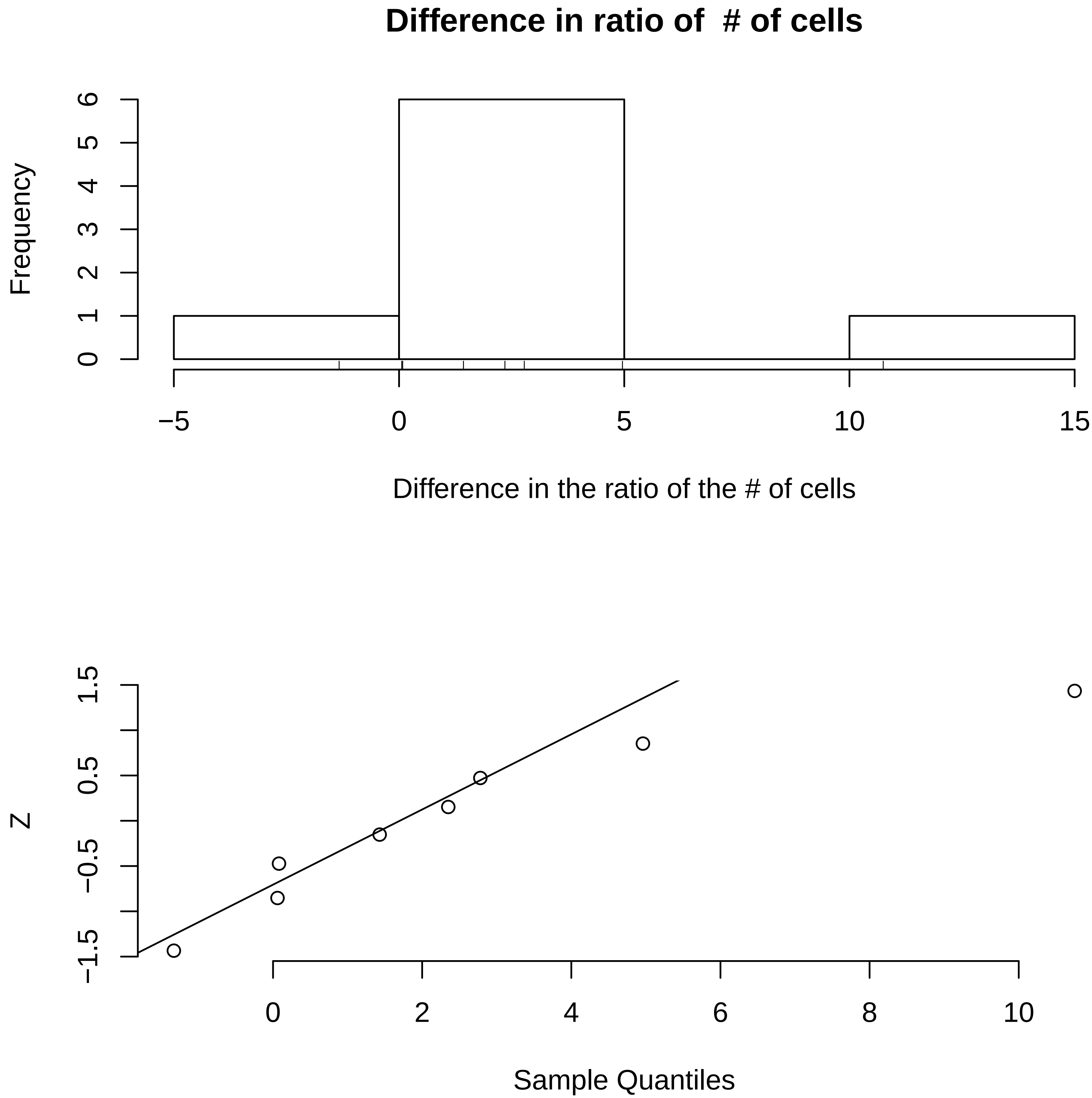
$$\begin{aligned} r &= \frac{S}{S_0} \\ &= 2^{(720/t)} \end{aligned}$$

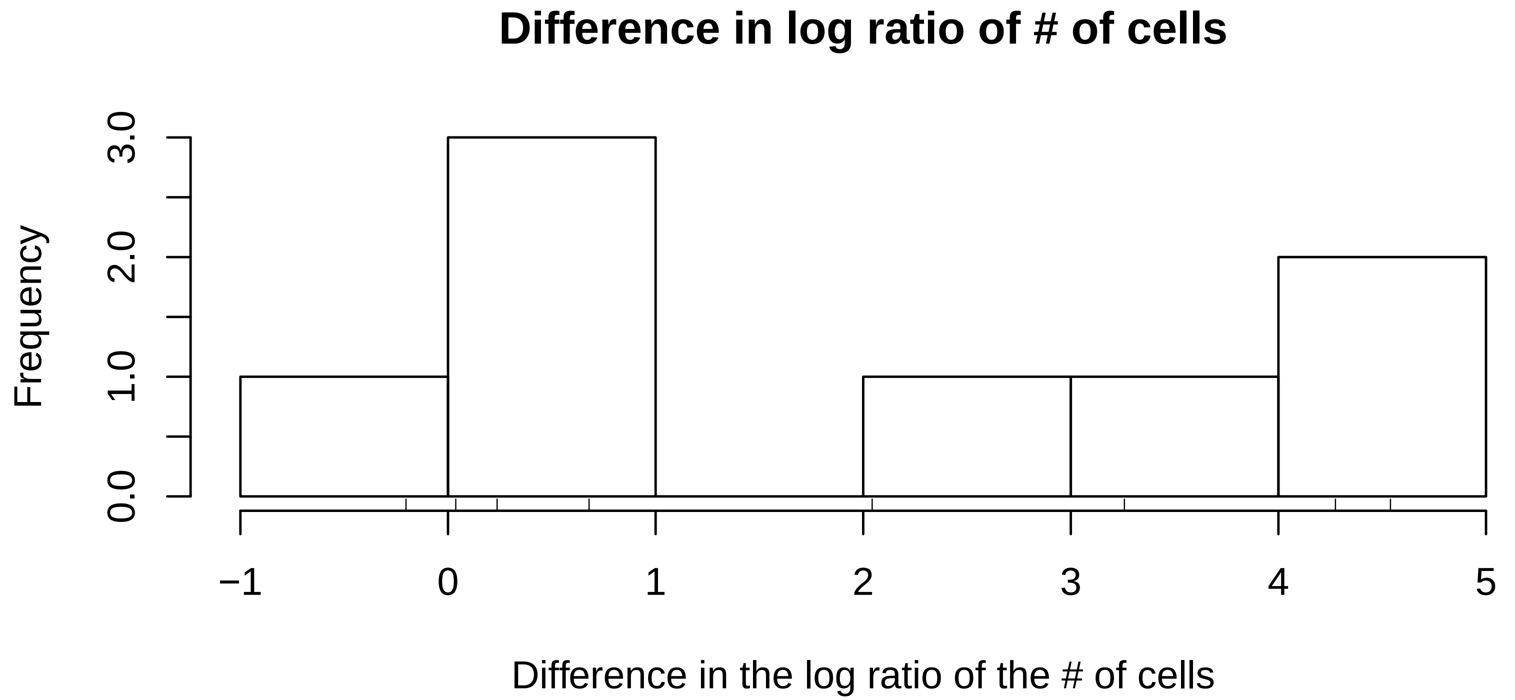
Slight changes in t lead to huge (compounding) differences in r. The Central Limit Theorem does not hold for multiplicative effects!

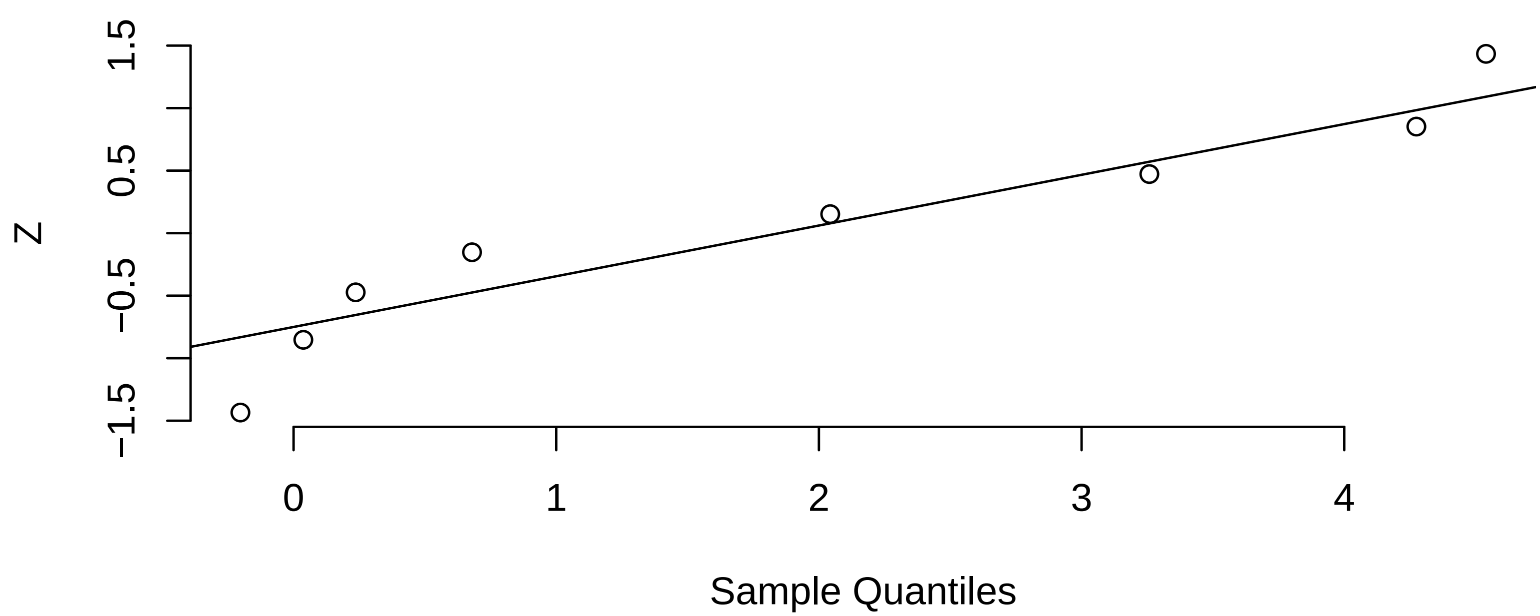
The distribution of r across bacteria is not even close to normal!

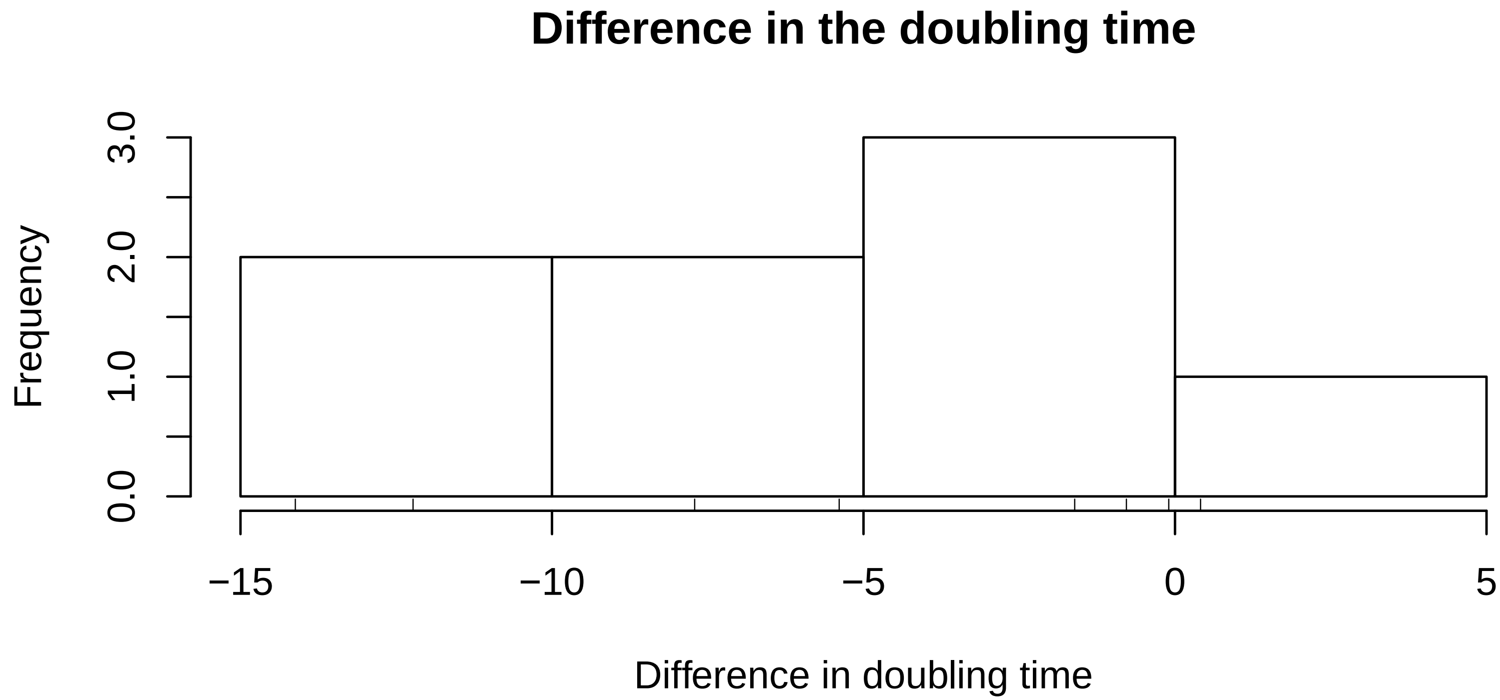
$$\begin{aligned} \log r &= \log 2^{(720/t)} \\ &= (720/t) \log(2) \end{aligned}$$

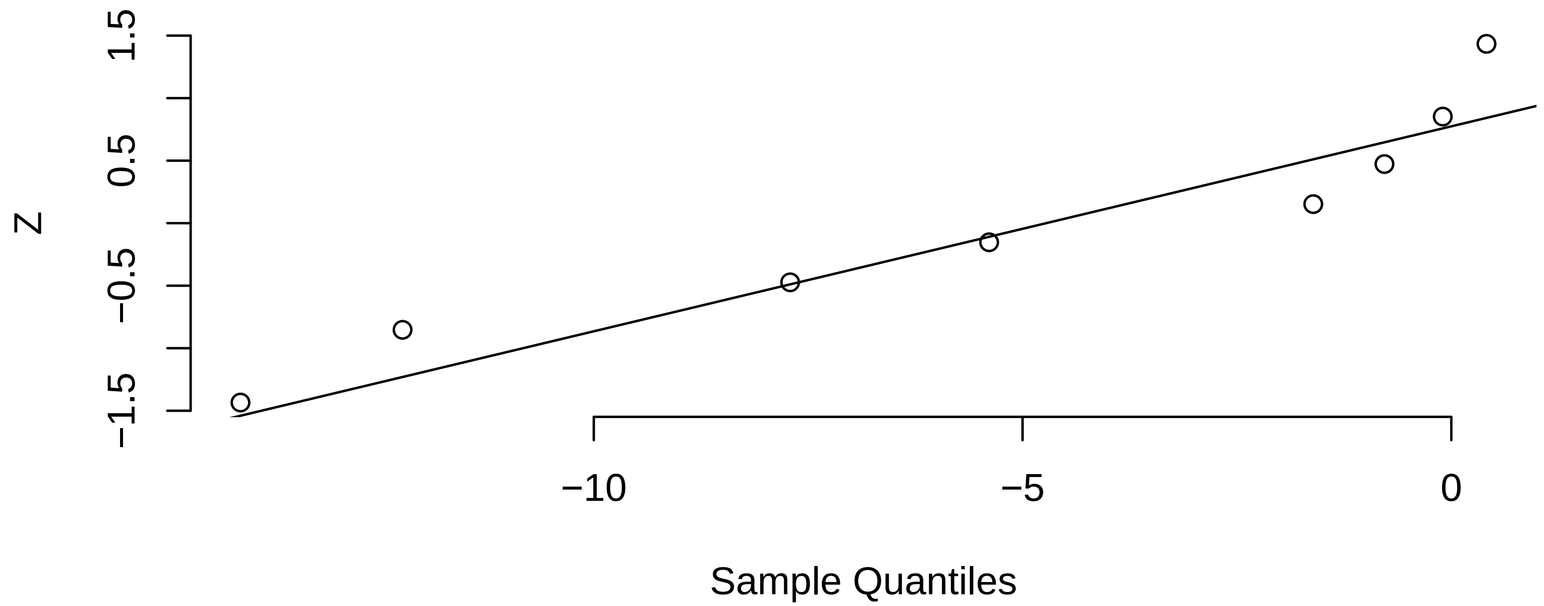
The distribution of $\log r$ is close to normal.


Difference in ratio of # of cells
Frequency
0
1
2
3
4
5
6
−5
0
5
10
15
Difference in the ratio of the # of cells
Z
−1.5
−0.5
0.5
1.5
0
2
4
6
8
10
Sample Quantiles


Difference in log ratio of # of cells
Frequency
0.0
1.0
2.0
3.0
−1
0
1
2
3
4
5
Difference in the log ratio of the # of cells


Z
−1.5
−0.5
0.5
1.5
0
1
2
3
4
Sample Quantiles


Difference in the doubling time
Frequency
0.0
1.0
2.0
3.0
−15
−10
−5
0
5
Difference in doubling time


Z
−1.5
−0.5
0.5
1.5
−10
−5
0
Sample Quantiles

	ln[# cells ratio] non-resistant	**ln**[# cells ratio] resistant	Differences
E. coli	1.78	1.98	-0.20
S. enterica	1.03	-3.51	4.54
K. oxytoca	1.62	-2.66	4.27
S. dysenteriae	2.41	-0.84	3.26
P. vulgaris	0.99	-1.05	2.04
S. marcescens	1.06	0.39	0.68
C. freundii	0.49	0.46	0.04
C. koseri	-0.97	-1.20	0.24
	$\bar{y}_1 = 1.05$ $s_1^2 = 1.0137$ $s_1 = 1.0068$	$\bar{y}_2 = -0.80$ $s_2^2 = 3.1053$ $s_2 = 1.7622$	$\bar{d} = 1.86$ $s_d^2 = 3.8008$ $s_d = 1.9496$

$$t = \frac{\bar{d} - 0}{\left(\frac{s_d}{\sqrt{n}}\right)} = \frac{1.86}{0.686} = 2.711$$

$$t_0.02(2), 7 = 3.00$$

$$t_0.05(2), 7 = 2.36$$

$$0.02 < P < 0.05$$

95% Confidence interval:

$$1.86 - (2.36)0.686 < \ln \mu_s - \ln \mu_r < 1.86 - (2.36)0.686$$

$$0.24 < \ln \mu_s - \ln \mu_r < 3.48$$

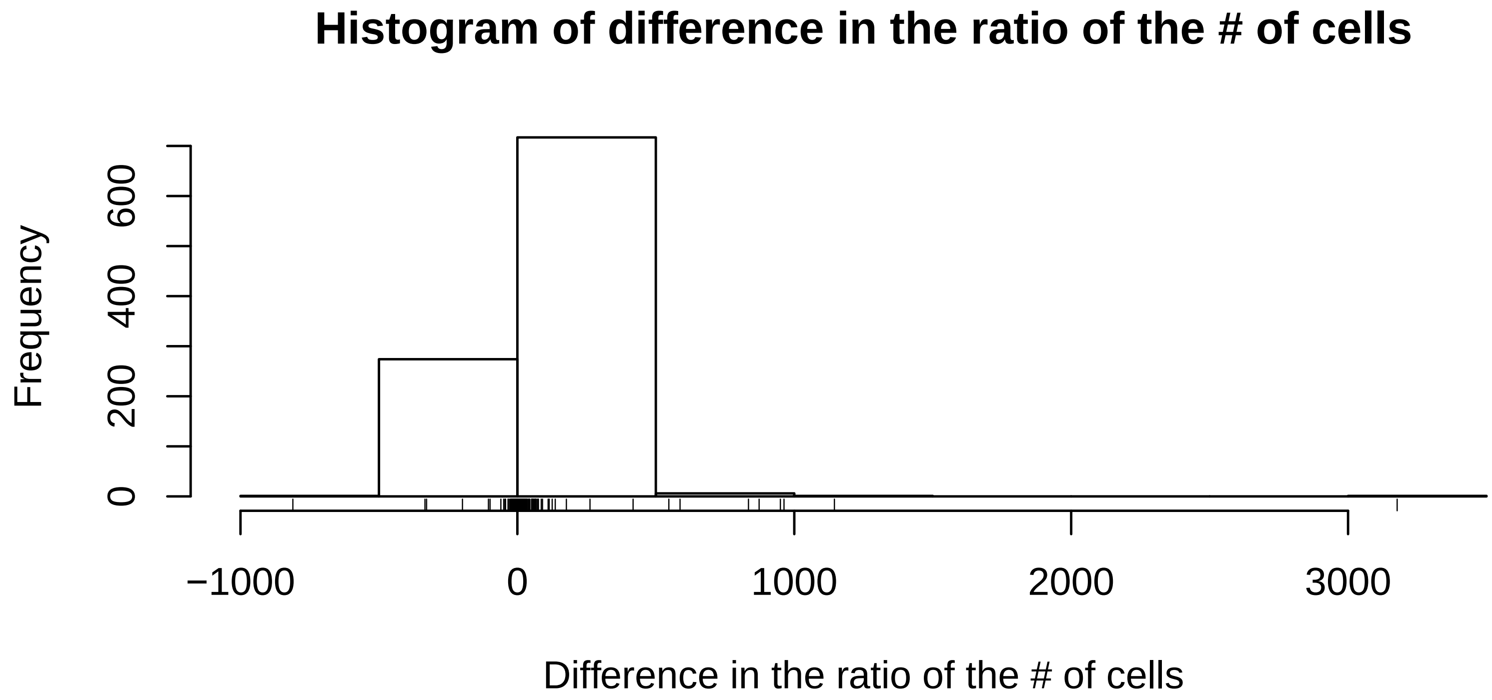
Back-transform to get confidence interval in terms of un-transformed data.

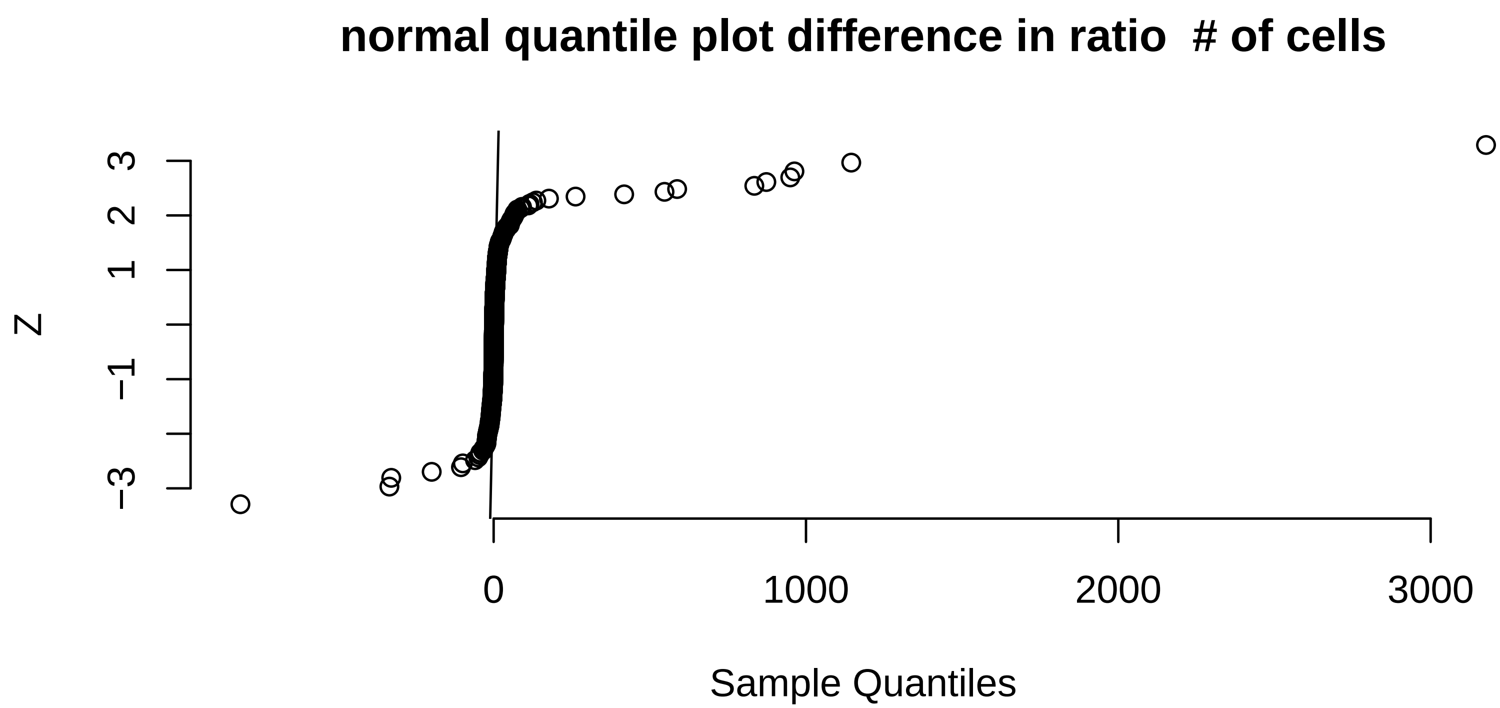
We are 95% confident that:

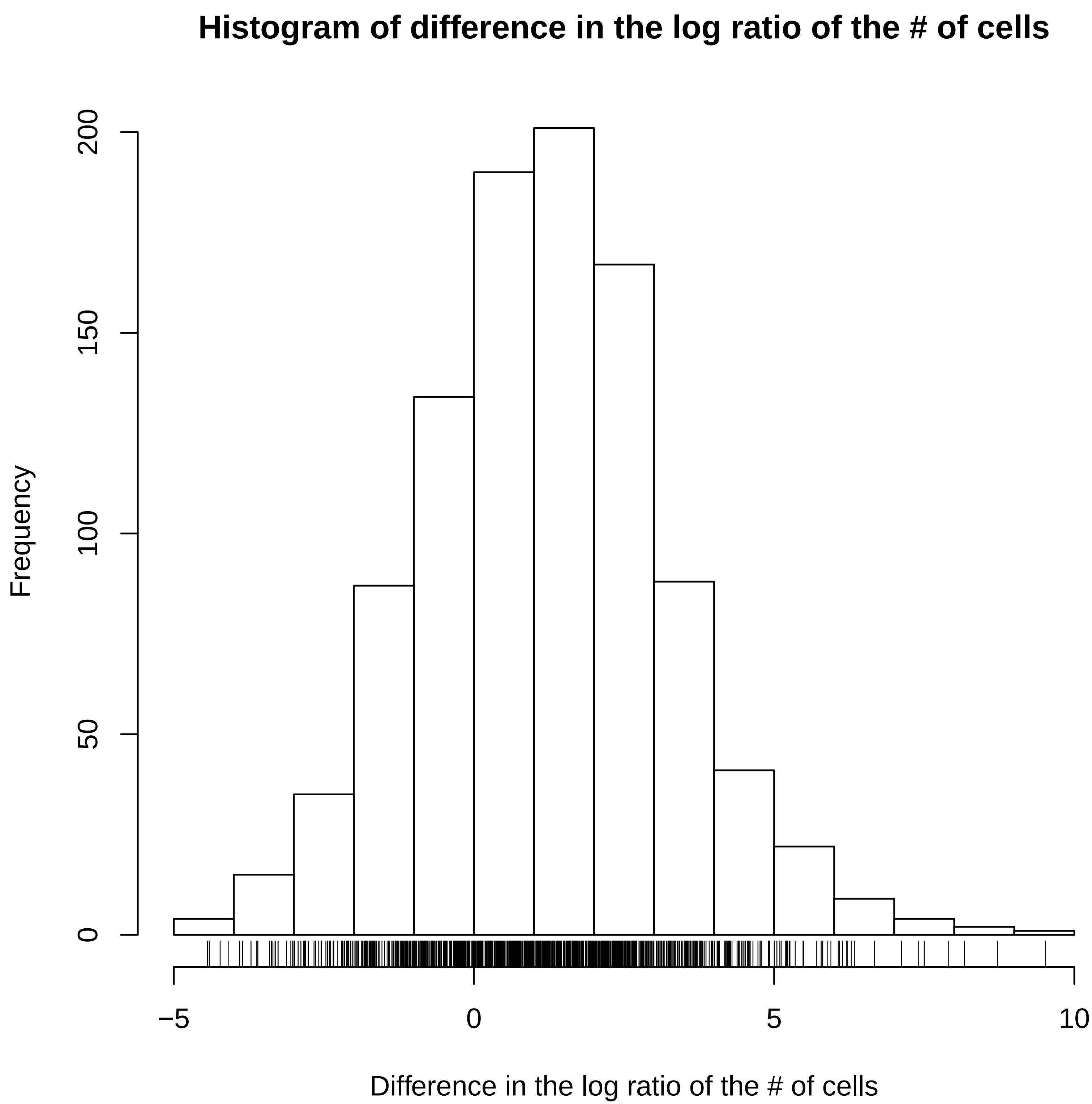
$$0.24 < \ln \mu_s - \ln \mu_r < 3.48$$

$$e^{0.24} < e^{\ln \mu_s - \ln \mu_r} < e^{3.48}$$

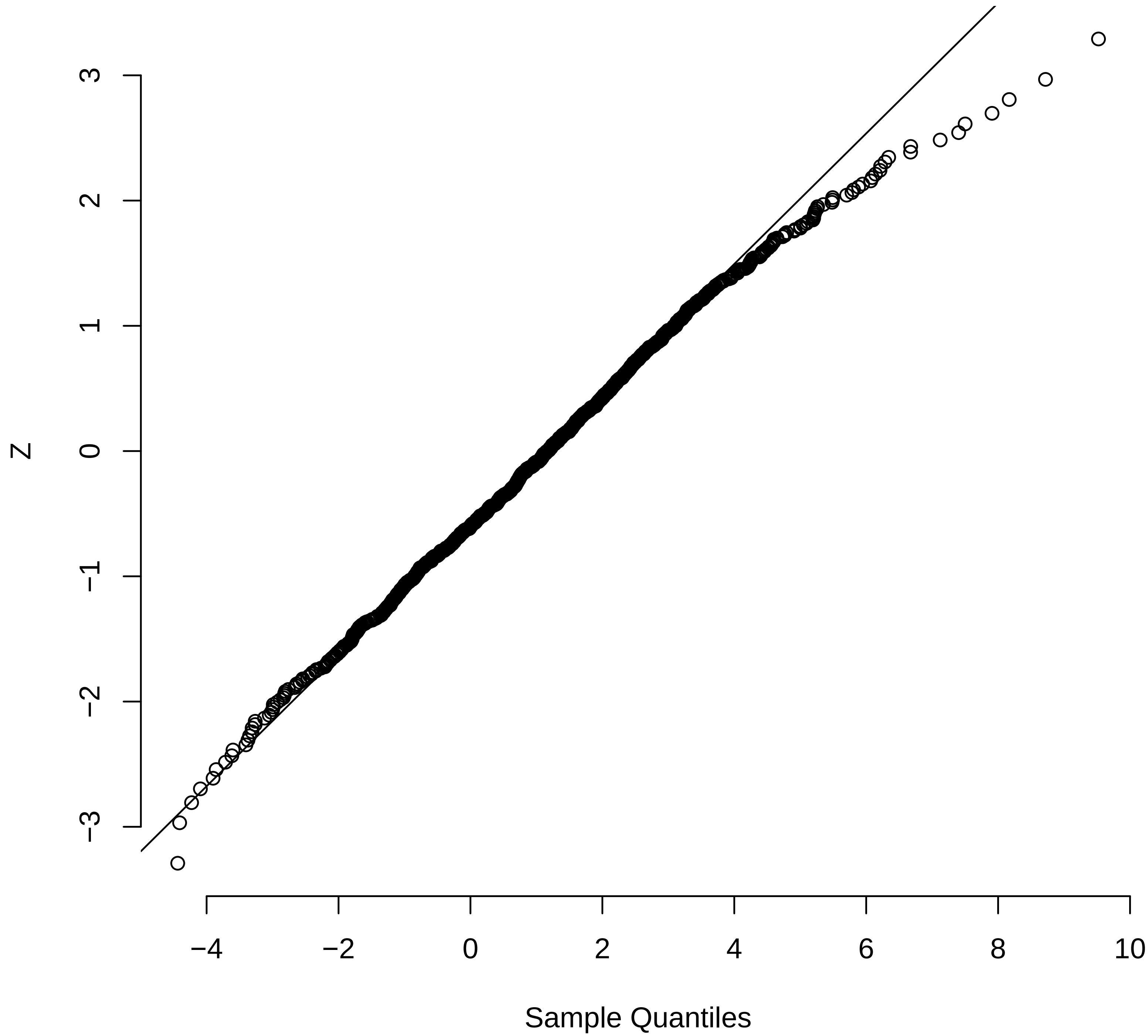
$$1.27 < \frac{\mu_s}{\mu_r} < 32.42$$


Histogram of difference in the ratio of the # of cells
Frequency
0
200
400
600
−1000
0
1000
2000
3000
Difference in the ratio of the # of cells


normal quantile plot difference in ratio # of cells
Z
−3
−1
1
2
3
0
1000
2000
3000
Sample Quantiles


Histogram of difference in the log ratio of the # of cells
Frequency
0
50
100
150
200
−5
0
5
10
Difference in the log ratio of the # of cells

normal quantile plot difference in log ratio # of cells



Transformation of variables can result in a set of data that satisfy the assumptions of our hypothesis testing machinery.

Test for normality $\rightarrow$ transform $\rightarrow$ Hypothesis test $\rightarrow$ back transform conclusions or confidence interval.

Transformations are common – some variables are typically measured in a log scale (dB for sound volume, pH, Richter scale)

H_0	Normal test stat.	Normal table/dist.	Non-parametric test stat.	Non-parametric table/dist.
$\mu_0 = 5.2$, $\sigma_0 = 3.4$	$Z = \frac{\bar{Y} - \mu_0}{\left(\frac{\sigma_0}{\sqrt{n}}\right)}$	Z		
$\mu_0 = 8.3$	$t = \frac{\bar{Y} - \mu_0}{\left(\frac{s}{\sqrt{n}}\right)}$	$t_{\alpha(2),n-1}$		
unpaired data $\mu_1 = \mu_2$	$t = \frac{\bar{Y_1} - \bar{Y_2}}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$	$t_{\alpha(2),n_1+n_2-2}$	U	$U_{\alpha(2),n_1,n_2}$
paired data $\mu_1 = \mu_2$ $\mu_1 - \mu_2 = 0$	$t = \frac{\bar{d}}{\left(\frac{s_d}{\sqrt{n_1}}\right)}$	$t_{\alpha(2),n_1-1}$	Sign test	Binomial distribution

- Probability for ranges of values under a normal.
- Connections between binomial and Normal.
- Confidence intervals for the mean or the difference between means.
- differences between standard normal and the t-distributions
- log-transformations when data are right-skewed.

$$x = e^y \text{ then } y = \ln[x]$$

(ln[x] is often denoted log[x]).

Are there more disciplinary problems with elementary school children on days that are within 1 day of a full moon?

Study design: Ask teachers to record the number of times a student has to be reprimanded. For each student studied, the teacher reports the average number of problems per day in the "near full moon" time frame and the "other times" time frame.

I just made up the data that follow

Average # of disciplinary problems per day

Student	within one day of full moon	other times
1	3.33	0.27
2	3.67	0.59
3	2.27	0.32
4	3.33	0.19
5	3.33	1.26
6	3.67	0.11
7	4.67	0.30
8	2.67	0.40
9	6.00	1.59
10	4.33	0.60
11	3.33	0.65
12	0.67	0.69
13	1.33	1.26
14	0.33	0.23
15	2.00	0.38

Student	within one day of full moon	other times	difference
1	3.33	0.27	3.06
2	3.67	0.59	3.08
3	2.27	0.32	1.95
4	3.33	0.19	3.14
5	3.33	1.26	2.07
6	3.67	0.11	3.56
7	4.67	0.30	4.37
8	2.67	0.40	2.27
9	6.00	1.59	4.41
10	4.33	0.60	3.73
11	3.33	0.65	2.68
12	0.67	0.69	−0.02
13	1.33	1.26	0.07
14	0.33	0.23	0.1
15	2.00	0.38	1.62

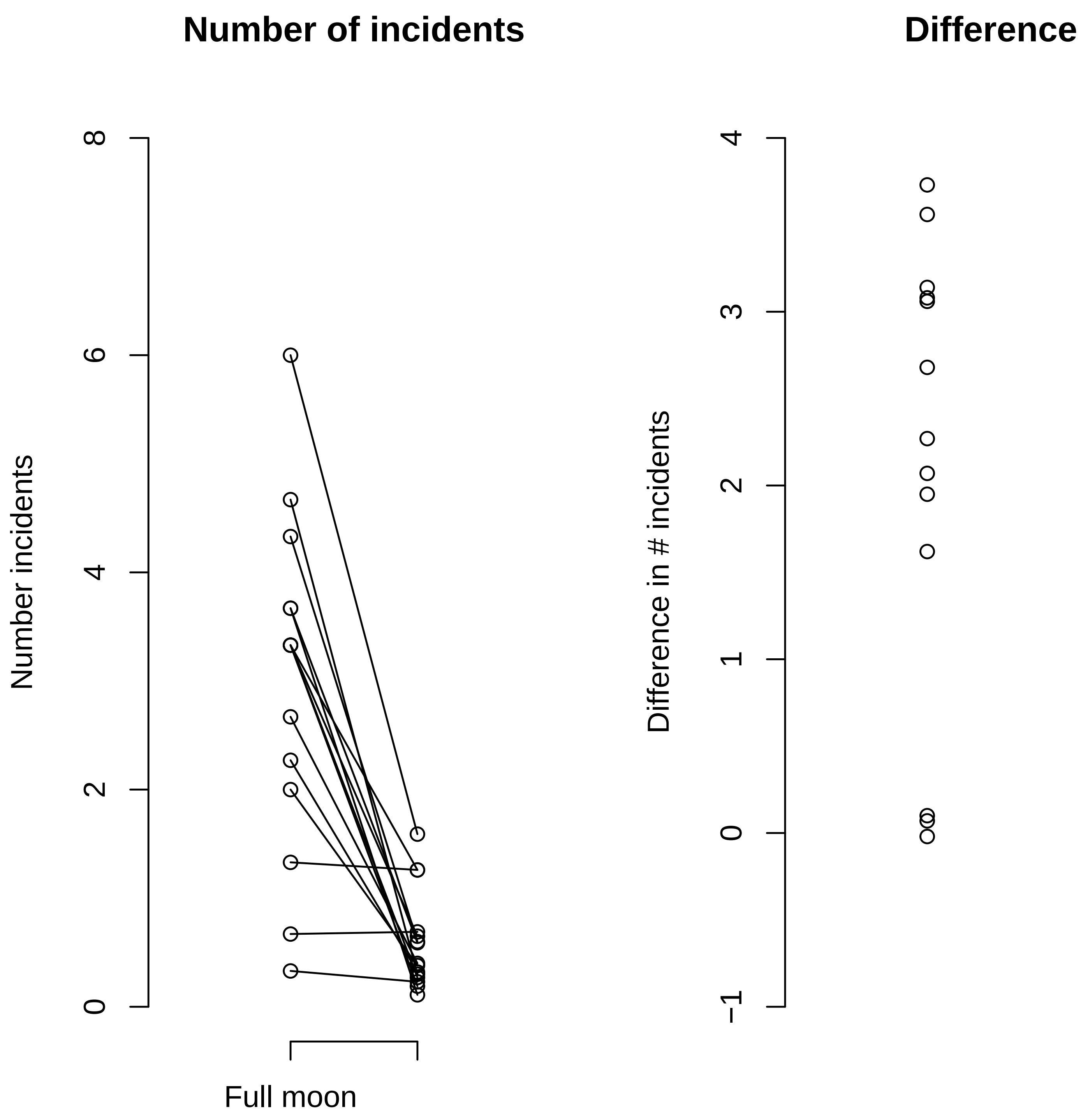
H_0: the mean number of incidents is the same near the full moon and other times:

$$\mu_f = \mu_o$$

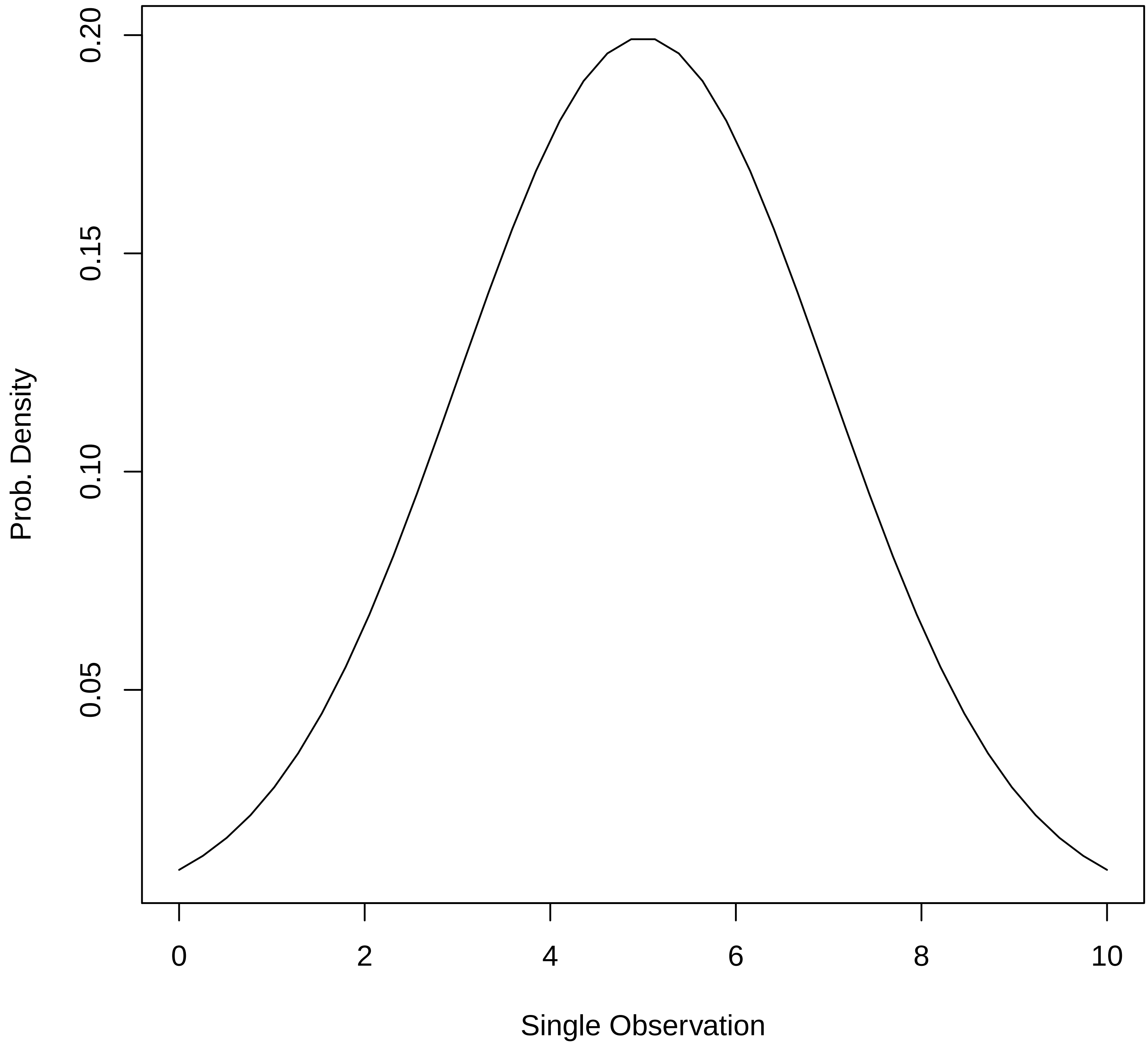
H_A:

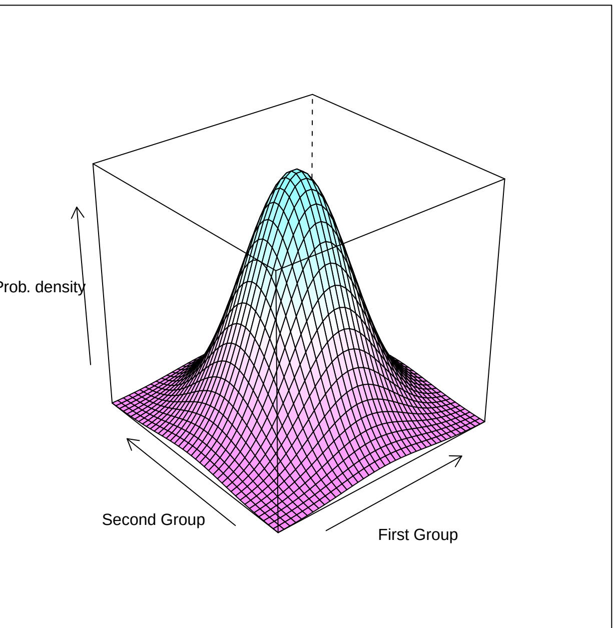
$$\mu_f \neq \mu_o$$

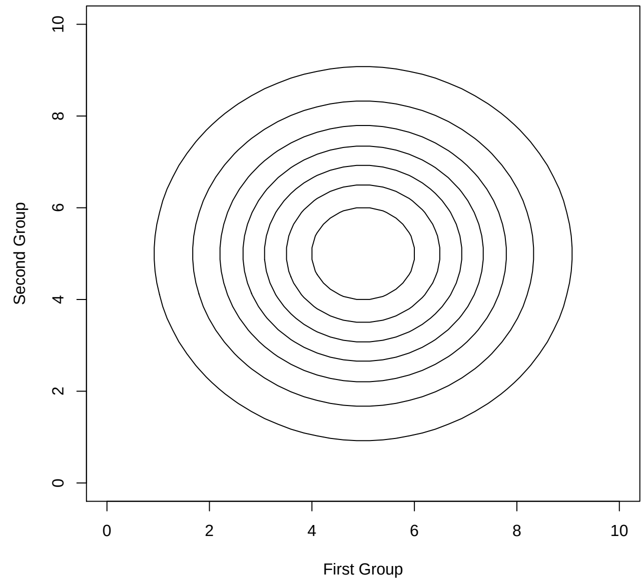
a paired t-test seems appropriate, but . . .

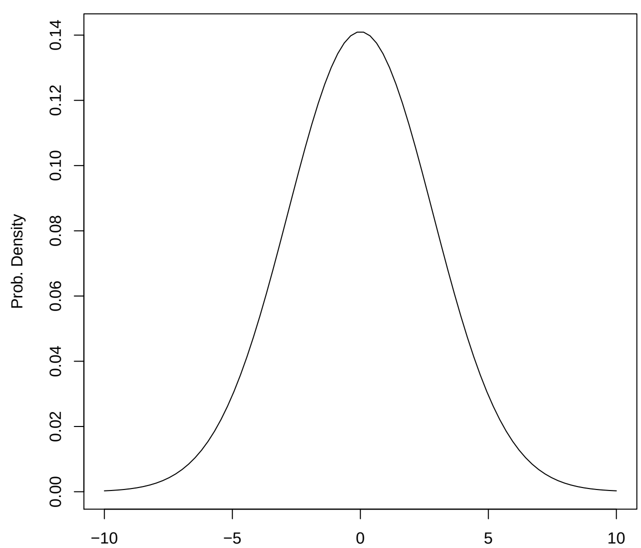


Number of incidents
Number incidents
0
2
4
6
8
Full moon
Difference
Difference in # incidents
−1
0
1
2
3
4

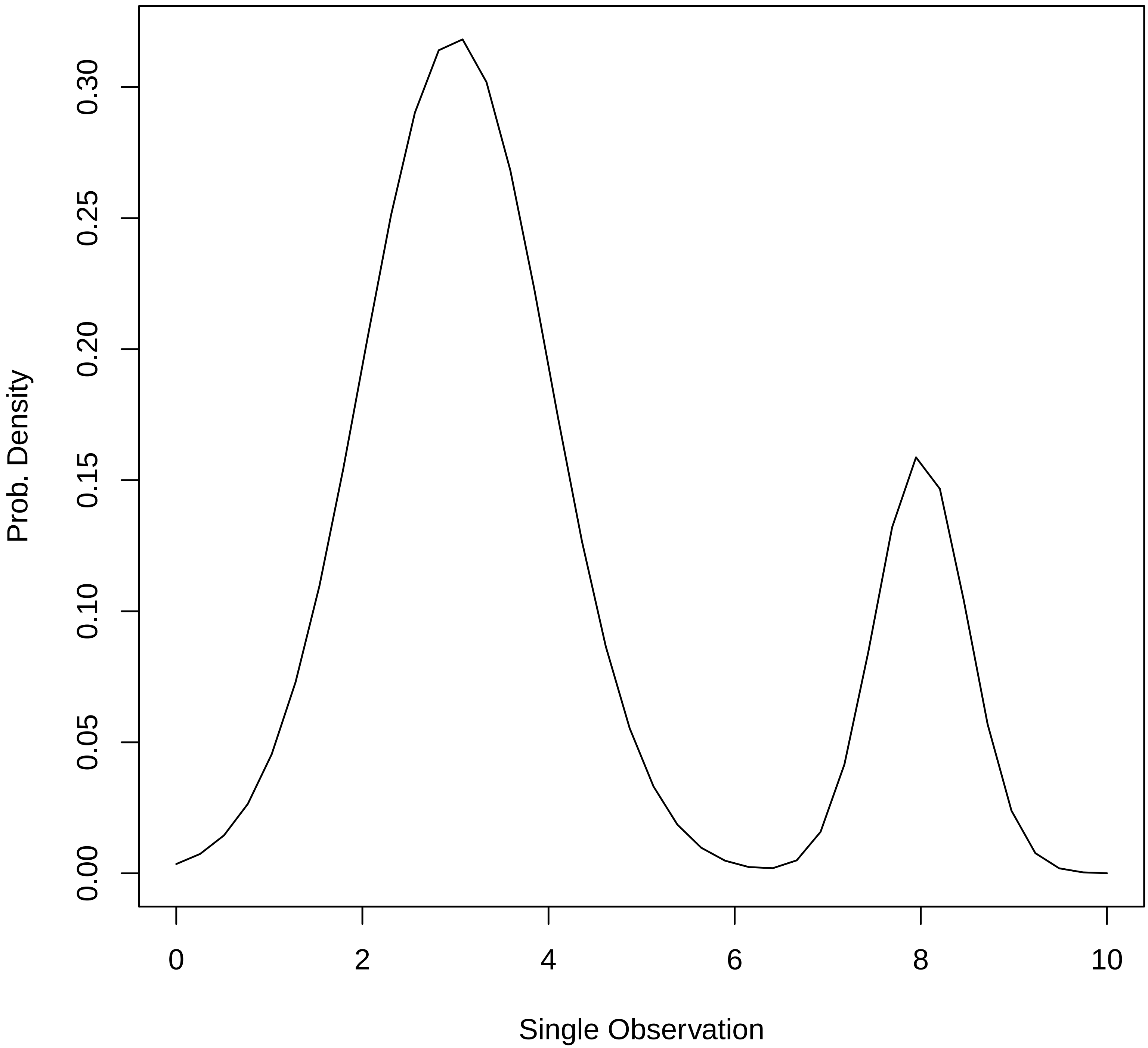

0.05
0.10
0.15
0.20
Prob. Density
0
2
4
6
8
10
Single Observation



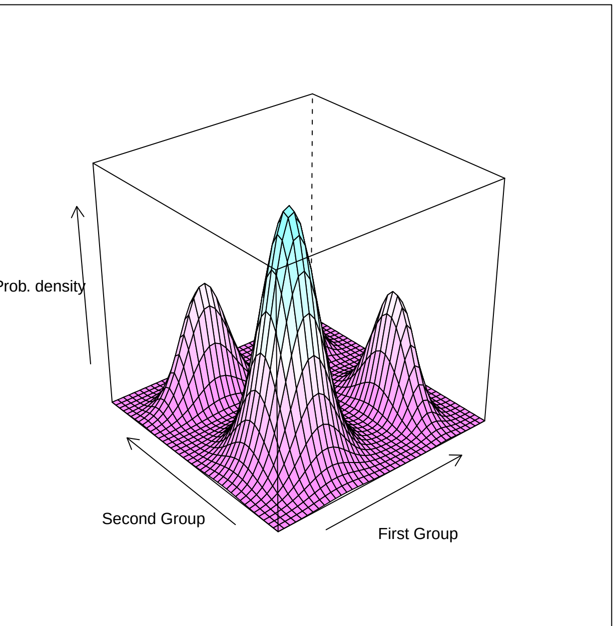


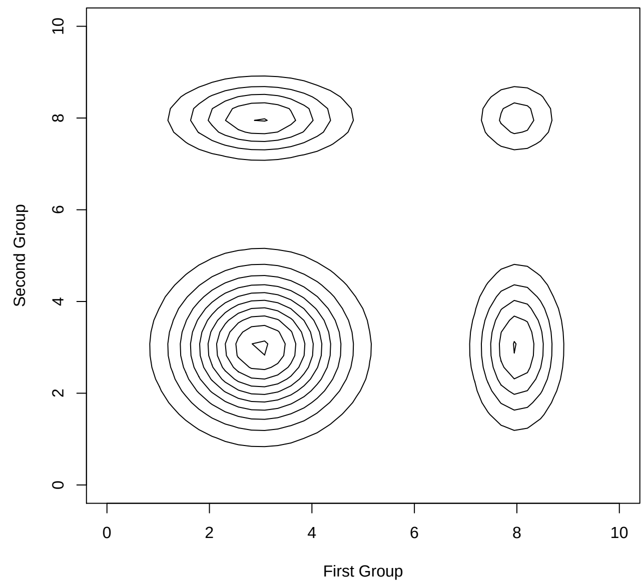


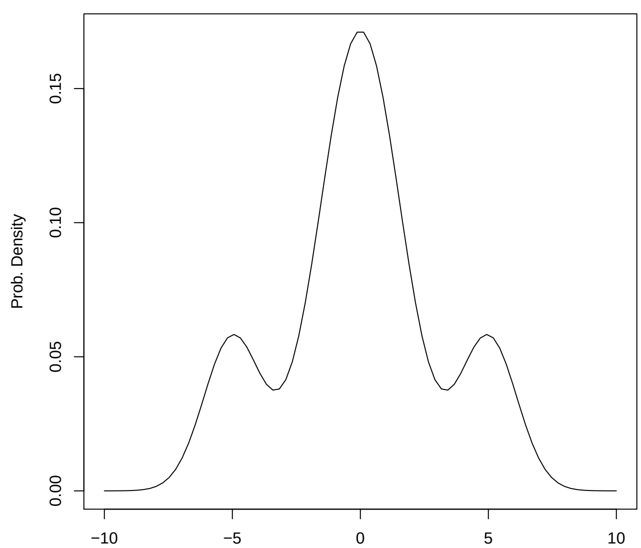
Null distribution of difference



Prob. Density
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0
2
4
6
8
10
Single Observation







Null distribution of difference

We **cannot** assume normality if:

- examination of the data shows strong deviations from normality (multiple modes, skew. . .), OR
- the null hypothesis is incompatible with normality.

If we cannot assume normality, we can:

- transform the data to a form that *is* approximately normal,
- use (less-powerful) non-parametric techniques that do not assume normality.

Student	within one day of full moon	other times	difference	sign of diff
1	3.33	0.27	3.06	+
2	3.67	0.59	3.08	+
3	2.27	0.32	1.95	+
4	3.33	0.19	3.14	+
5	3.33	1.26	2.07	+
6	3.67	0.11	3.56	+
7	4.67	0.30	4.37	+
8	2.67	0.40	2.27	+
9	6.00	1.59	4.41	+
10	4.33	0.60	3.73	+
11	3.33	0.65	2.68	+
12	0.67	0.69	-0.02	−
13	1.33	1.26	0.07	+
14	0.33	0.23	0.1	+
15	2.00	0.38	1.62	+

$n = 15$

count of positive differences $= x = 14$

What is the probability that we would get 14 out of 15 positive differences if there was no difference between days near the full moon and other times?

H_0: the median difference between the two groups is 0.

If the median difference is 0, then

$$\Pr(d > 0) = 0.5$$

What is the probability of a result as extreme as 14 out of 15 differences being in the same direction?

$x = 14$ $\qquad$ $n = 15$ $\qquad$ $p = 0.5$

$$\Pr(x \geq 14 \text{ or } x \leq 1) = \Pr(x = 14) + \Pr(x = 15) + \Pr(x = 1) + \Pr(x = 0)$$

$$2\left[\binom{15}{0}\left(\frac{1}{2}\right)^{15}\left(\frac{1}{2}\right)^{0} + \binom{15}{1}\left(\frac{1}{2}\right)^{14}\left(\frac{1}{2}\right)^{1}\right] \approx 0.001$$

We reject the null hypothesis that the full moon has no effect on student behavior. Paired comparisons of 14 out of 15 students showed more incidents on days near a full moon. Under the null hypothesis, the median difference in the number of disciplinary incidents in the days near a full moon and the number of disciplinary incidents at other times should be 0. The proportion of comparisons with increased rate of disciplinary incidents($\hat{p} = 14/15$) is too great to be explained by sampling error alone ($P \approx 0.001$).

Sign test:

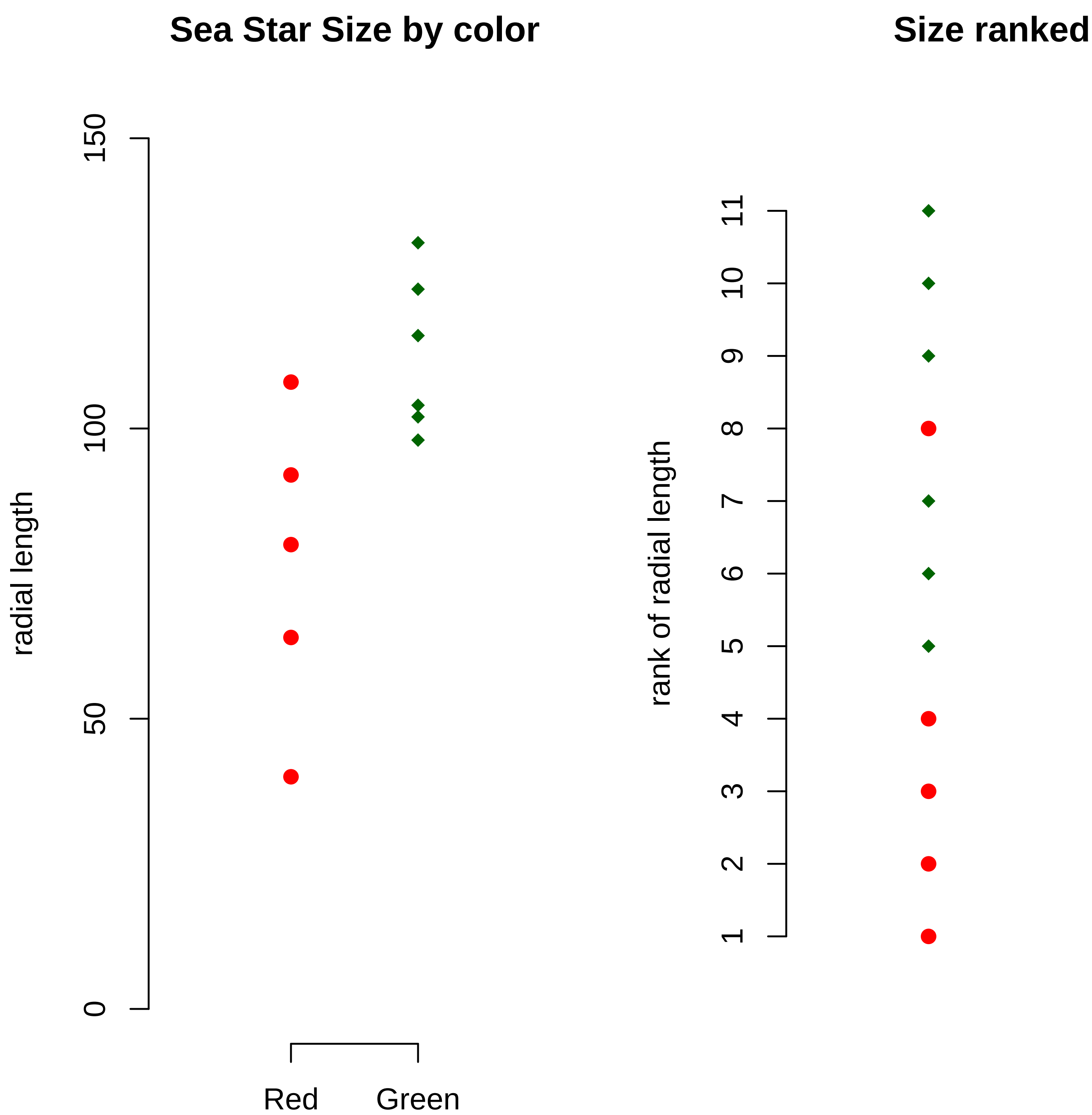
- convert the differences to + or − and treat the + as successes in a binomial test with $p_0 = 0.5$
- test statistic = the # of + differences
- assumes a random sample
- does **not** assume normality of the *differences*
- H_0: median=0

From Glover and Mitchell

Size (mm) of sea stars of different colors:

Red		Green
108		102
64		116
80		98
92		132
40		104
		124

How can we assess if this difference is significant without assuming some specifics about a distribution?



Sea Star Size by color
Size ranked
radial length
0
50
100
150
Red
Green
rank of radial length
1
2
3
4
5
6
7
8
9
10
11

Table 1: Student's t-distribution

df $\boldsymbol{\alpha}$**(2)**	**0.2**	**0.1**	**0.05**	**0.02**	**0.01**	**0.001**	**0.0001**
1	3.08	6.31	12.71	31.82	63.66	636.62	6366.20
2	1.89	2.92	4.30	6.96	9.92	31.60	99.99
3	1.64	2.35	3.18	4.54	5.84	12.92	28.00
4	1.53	2.13	2.78	3.75	4.60	8.61	15.54
5	1.48	2.02	2.57	3.36	4.03	6.87	11.18
6	1.44	1.94	2.45	3.14	3.71	5.96	9.08
7	1.41	1.89	2.36	3.00	3.50	5.41	7.88
8	1.40	1.86	2.31	2.90	3.36	5.04	7.12
9	1.38	1.83	2.26	2.82	3.25	4.78	6.59
10	1.37	1.81	2.23	2.76	3.17	4.59	6.21
11	1.36	1.80	2.20	2.72	3.11	4.44	5.92
12	1.36	1.78	2.18	2.68	3.05	4.32	5.69
13	1.35	1.77	2.16	2.65	3.01	4.22	5.51
14	1.35	1.76	2.14	2.62	2.98	4.14	5.36
15	1.34	1.75	2.13	2.60	2.95	4.07	5.24
16	1.34	1.75	2.12	2.58	2.92	4.01	5.13
17	1.33	1.74	2.11	2.57	2.90	3.97	5.04
18	1.33	1.73	2.10	2.55	2.88	3.92	4.97
19	1.33	1.73	2.09	2.54	2.86	3.88	4.90
20	1.33	1.72	2.09	2.53	2.85	3.85	4.84

Table 1: Mann-Whitney U-distribution critical values for $\alpha(2) = 0.05$

	n_1												
n_2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
3	–	–	15	17	20	22	25	27	30	32	35	37	40
4	–	16	19	22	25	28	32	35	38	41	44	47	50
5	15	19	23	27	30	34	38	42	46	49	53	57	61
6	17	22	27	31	36	40	44	49	53	58	62	67	71
7	20	25	30	36	41	46	51	56	61	66	71	76	81
8	22	28	34	40	46	51	57	63	69	74	80	86	91
9	25	32	38	44	51	57	64	70	76	82	89	95	101
10	27	35	42	49	56	63	70	77	84	91	97	104	111
11	30	38	46	53	61	69	76	84	91	99	106	114	121
12	32	41	49	58	66	74	82	91	99	107	115	123	131
13	35	44	53	62	71	80	89	97	106	115	124	132	141
14	37	47	57	67	76	86	95	104	114	123	132	141	151
15	40	50	61	71	81	91	101	111	121	131	141	151	161

Table 2: Mann-Whitney U-distribution critical values for $\alpha(2) = 0.01$

	n_1												
n_2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
3	–	–	–	–	–	–	27	30	33	35	38	41	43
4	–	–	–	24	28	31	35	38	42	45	49	52	55
5	–	–	25	29	34	38	42	46	50	54	58	63	67
6	–	24	29	34	39	44	49	54	59	63	68	73	78
7	–	28	34	39	45	50	56	61	67	72	78	83	89
8	–	31	38	44	50	57	63	69	75	81	87	94	100
9	27	35	42	49	56	63	70	77	83	90	97	104	111
10	30	38	46	54	61	69	77	84	92	99	106	114	121
11	33	42	50	59	67	75	83	92	100	108	116	124	132
12	35	45	54	63	72	81	90	99	108	117	125	134	143
13	38	49	58	68	78	87	97	106	116	125	135	144	153
14	41	52	63	73	83	94	104	114	124	134	144	154	164
15	43	55	67	78	89	100	111	121	132	143	153	164	174

Step 1: Sort the data

Step 2: Calculate the rank-sum for a category:

$$R_1 = 1 + 2 + 3 + 4 + 8 = 18$$

Category	Value	Rank
Red	40	1
Red	64	2
Red	80	3
Red	92	4
Green	98	5
Green	102	6
Green	104	7
Red	108	8
Green	116	9
Green	124	10
Green	132	11

Step 3:

$$\begin{aligned} U_1 &= n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \\ &= 5(6) + \frac{5(6)}{2} - 18 \\ &= 27 \end{aligned}$$

Step 4:

$$\begin{aligned} U_2 &= n_1 n_2 - U_1 \\ &= 5(6) - 27 \\ &= 3 \end{aligned}$$

Step 5:

$$\begin{aligned} U &= \max(U_1, U_2) \\ &= \max(27, 3) = 27 \end{aligned}$$

Step 6: Look up U_{n_1,n_2} is the appropriate table.

$$U_{0.05(2),5,6=27}$$
$$U_{0.01(2),5,6=29}$$

$$P < 0.05$$

We reject the null hypothesis that the green and red sea stars are drawn from identical distributions. The radial length for green sea stars were longer (median = 110mm, $n = 6$) than red sea stars (median = 80mm, $n = 5$). Based on a Mann-Whitney U test ($U = 27$), the difference in ranks was too large to be explained by chance ($P = 0.05$).

The U_1 statistic measures the number of pair-wise comparisons between samples from population 1 and population 2 in which the sample from population 1 had a smaller value. U_2 is the number of these comparisons for which the population 2 sample was smaller:

	green					
	98	102	104	116	124	132
40	●	●	●	●	●	●
64	●	●	●	●	●	●
80	●	●	●	●	●	●
92	●	●	●	●	●	●
108	●	●	●	●	●	●

$U_1 = 27$ $U_2 = 3$

Mann-Whitney's U is the larger of U_1 and U_2. If the samples from both populations have similar values, then the ranking of all data points will mix together samples from each population. The result will be both U_1 and U_2 will be similar in value, and have a value that is not high (and not close to the critical value).

If there is a strong tendency for samples from population 1 to be smaller than those from population 2, then that population with smaller values will have a large U_1 statistic. If there is a strong tendency for samples from population 2 to be smaller, then U_2 will be large. Whenever one population dominates the low ranks, and the other dominates the high ranks then U will be large. This is signal that the samples are drawn from populations with different distributions.

From Samuels and Witmer

Resting Human β-endorphin levels in two groups (measured in pg/mL):

Jogger HBE	New exerciser HBE
39	70
40	47
32	54
60	27
19	31
52	42
41	37.1
32	41
13	9
37	18
28	33
	23
	49
	41
	59

$$R_1 = 2+4+7+9.5+9.5+12+14+15+17+22+25 = 137$$

$$n_1 = 11 \qquad n_2 = 15$$

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

$$U_1 = 11(15) + \frac{11(12)}{2} - 137 = 94$$

$$U_2 = n_1 n_2 - U_1$$

$$U_2 = 11(15) - 94 = 71$$

$$U = \max(U_1, U_2) = \max(94, 71) = 94$$

$$U_{0.05(2),11,15} = 121$$

HBE	raw rank	ranks with ties
9	1	1
13	2	2
18	3	3
19	4	4
23	5	5
27	6	6
28	7	7
31	8	8
32	9	$(9+10)/2$
32	10	$(9+10)/2$
33	11	11
37	12	12
37.1	13	13
39	14	14
40	15	15
41	16	$(16+17+18)/3$
41	17	$(16+17+18)/3$
41	18	$(16+17+18)/3$
42	19	19
47	20	20
49	21	21
52	22	22
54	23	23
59	24	24
60	25	25
70	26	26

HBE	raw rank	ranks with ties
9	1	1
13	2	2
18	3	3
19	4	4
23	5	5
27	6	6
28	7	7
31	8	8
32	9	9.5
32	10	9.5
33	11	11
37	12	12
37.1	13	13
39	14	14
40	15	15
41	16	17
41	17	17
41	18	17
42	19	19
47	20	20
49	21	21
52	22	22
54	23	23
59	24	24
60	25	25
70	26	26

We cannot reject the hypothesis that radial length of β-endorphin levels for joggers (n=11, median = 12pg/mL) and people who have just begun exercising (n = 15, median = 17 pg/mL) are identical. Using a Mann-Whitney U test (U = 94), and the tendency for joggers to have lower resting endorphin levels can be explained by sampling error even if there is no difference in endorphin levels between the population of joggers and those who just started exercising ($P > 0.05$)