ECE4010J RC5

Overview

- 1. Fisher Test
- 2. Neyman-Pearson Decision Theory
- 3. Single Sample Tests for the Mean and Variance
- 4. Non-parametric Single Sample Test for Median

Fisher Test

- Hypotheses and Testing:

definition: reject or fail to reject statements (hypotheses) based on statistical data.
We will make a hypothesis about an unknown parameter θ: $\theta = \theta_0$. The θ_0 here is a given value called null value. The hypothesis is called null hypothesis, denoted as H_0
example: given a set of data $X_1, X_2, \ldots, X_n$ from a normal distribution X,

- (with $\sigma^2 = 10$ already known) "$\mu = 20$"
- (with σ^2 unknown) "$\sigma^2 \geq 6$"
- (with σ^2 unknown) "$\mu \leq 22$"

The basic idea of hypothesis testing is, you first suppose H_0 is true, and you find the probability to get the data based on H_0. If it's very small(the data is very wierd if H_0 holds), then you reject H_0.

Fisher Test

- P-value:

The intuitive interpretation of P-value is the probability that H_0 gives out such a "wield" sample.
For example, $X \sim N(\mu, 5)$ The null hypothesis is $H_0 : \mu \leq 26$. We get a sample and the mean is $\overline{x}$. If $\overline{x} \leq 26$, then obviously H_0 can't be rejected. If $\overline{x} > 26$, we start to have evidence against H_0, and the probability we're interested in is

$$P[\overline{X} > \overline{x} | H_0 \text{ is true}] = P[\overline{X} > \overline{x} | \mu \leq 26] \leq P[\overline{X} > \overline{x} | \mu = 26]$$

The righthand side is the P-value of this hypothesis. It's $1 - \Phi(\frac{\overline{x}-\mu}{\sigma/\sqrt{n}})$

Fisher Test

- P-value for z-test:

  A hypothesis test on a normal sample, with overall variance known, and the null hypothesis takes one of the three forms:

  1. $H_0 : \mu \leq \mu_0$
  2. $H_0 : \mu \geq \mu_0$
  3. $H_0 : \mu = \mu_0$

  is called a z-test. 1) and 2) are called one-tailed test, 3 is called two-tailed test.

  The P-value for 1) is $1 - \Phi(\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}})$.

  The P-value for 2) is $1 - \Phi(\frac{\mu_0-\overline{x}}{\sigma/\sqrt{n}})$

  The P-value for 3) is $2(1 - \Phi(\frac{|\overline{x}-\mu_0|}{\sigma/\sqrt{n}}))$

Fisher Test

- Example for z-test:

The diameters of bolts are known to have a standard deviation of 0.0001 inch. A random sample of 10 bolts yields an average diameter of 0.2546 inch.

i) Test the hypothesis that the true mean diameter of bolts equals 0.255 inch, using $\alpha = 0.05$.

H0 : $\mu = 0.255$ at $\alpha = 0.05$. $Z = \frac{\bar{X} - 0.255}{\sigma / \sqrt{n}}$

In[1]:= $Z = \frac{0.2546 - 0.255}{0.0001 / \sqrt{10}}$

Out[1]= -12.6491

In[4]:= InverseCDF[NormalDistribution[], $1 - \frac{0.05}{2}$]

InverseCDF[NormalDistribution[], $\frac{0.05}{2}$]

Out[4]= 1.95996

Out[5]= -1.95996

we reject H0

Fisher Test

- Conclusion and Discussion:

If the P-value(often denoted as α) is small, then we have evidence that H_0 is false. In this situation, we say H_0 is rejected at significance level of α. Otherwise, we fails to reject H_0.
The threshold is 5%, however, further study need to be conducted. A small P-value doesn't necessarily means that H_0 is indeed false.

$a \to b$, $P[b]$ is low, doesn't mean $P[a]$ is low.

Also, it seems that when n is big enough, the P-value of two-tailed test will always be small.

Neyman-Pearson Decision Theory

We now need two hypotheses. One is the null hypothesis H_0, the other is H_1, called alternative hypothesis. In NPDT, we aim at finding the correct hypothesis, rather than only against H_0.
Example: $H_0 : \mu = 40$; $H_1 : |\mu - 40| \geq 1$
Two indicator:

- α, called the significance level or error of the I kind, means $P[\text{reject } H_0|H_0 \text{ is true}]$.
- β, called error of the II kind, means $P[\text{accept } H_0|H_1 \text{ is true}]$. $1-\beta$ is called power.

Neyman-Pearson Decision Theory

We now need two hypotheses. One is the null hypothesis H_0, the other is H_1, called alternative hypothesis. In NPDT, we aim at finding the correct hypothesis, rather than only against H_0.
Example: $H_0 : \mu = 40$; $H_1 : |\mu - 40| \geq 1$
Two indicator:

- α, called the significance level or error of the I kind, means $P[\text{reject } H_0 | H_0 \text{ is true}]$.
- β, called error of the II kind, means $P[\text{accept } H_0 | H_1 \text{ is true}]$. $1 - \beta$ is called power.

	Do not reject H_0	Reject H_0
H_0 is true	Correct Decision	Incorrect Decision: Type I error α
H_0 is false	Incorrect Decision: Type II error β	Correct Decision

Neyman-Pearson Decision Theory

- Critical Region

The critical region means, if the statistic falls into this region, the H_0 is rejected at the significance level of α. The critical region must be fixed before data are obtained.

1. $H_0 : \mu = \mu_0$, critical region is $\overline{x} \neq \mu_0 \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$
2. $H_0 : \mu \leq \mu_0$, critical region is $\overline{x} > \mu_0 + z_{\alpha}\frac{\sigma}{\sqrt{n}}$
3. $H_0 : \mu \geq \mu_0$, critical region is $\overline{x} < \mu_0 - z_{\alpha}\frac{\sigma}{\sqrt{n}}$

Neyman-Pearson Decision Theory

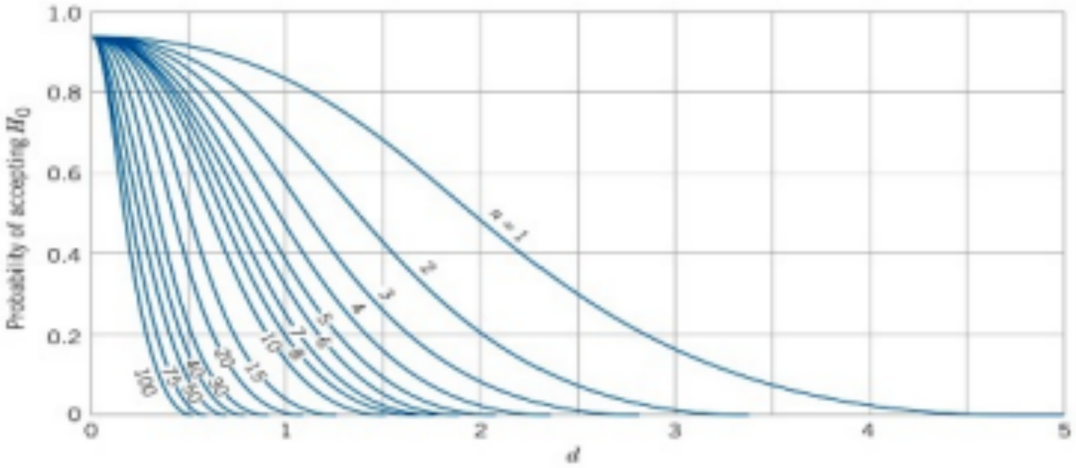
- β and sample size

β can be controlled by increasing sample size n. For a two-tailed null hypothesis, if we need to get a desired β, we need $n \geq \frac{(z_{\alpha/2}+z_{\beta})^2\sigma^2}{\delta^2}$. Here δ is the interval between H_0 and H_1.

Neyman-Pearson Decision Theory

- OC curve

OC Curve is a plot of actual μ vs. $P[H_0$ is accepted$]$.
For a two-tailed z-test, when $\mu = \mu_0$, the value is $1 - \alpha$ since there's α probability that we reject H_0 when H_0 is correct.
When μ is the bounded value of H_1, the value is β since now H_1 is true.
Notice that for different n, the curve is different.
In practice our OC curve is a plot of d vs $P[H_0$ is accepted$]$. $d := \frac{\mu-\mu_0}{\sigma}$.
This is the OC curve of two-tailed z-test.



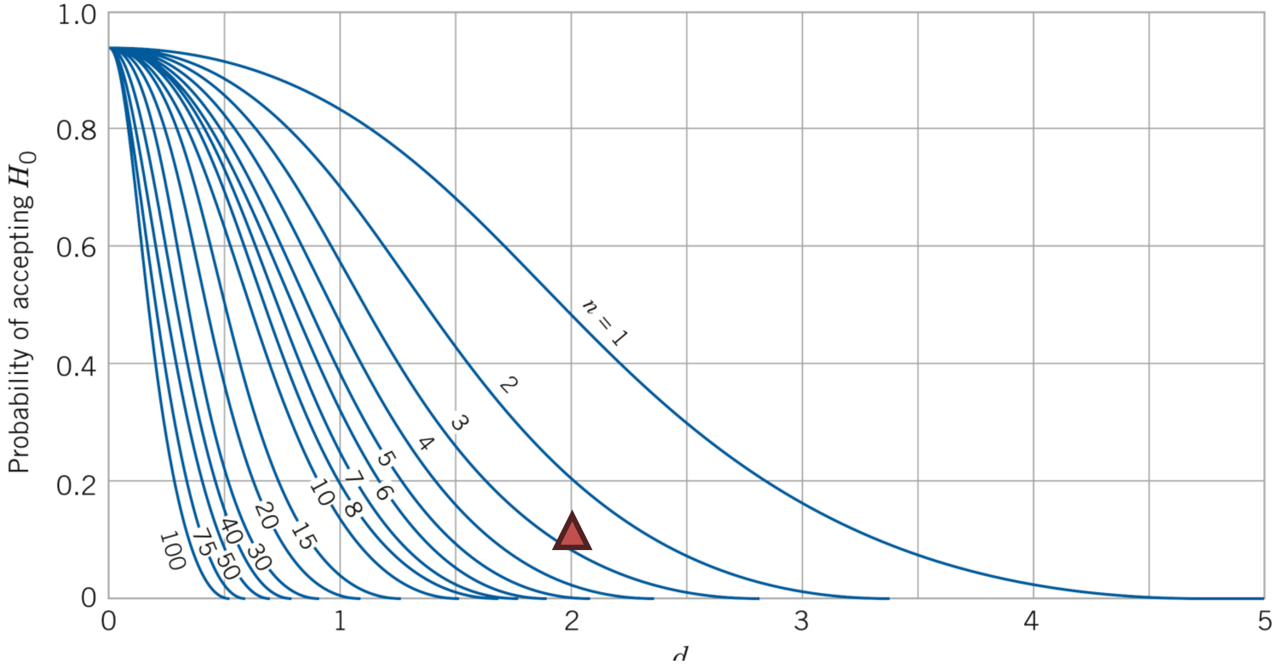
Neyman-Pearson Decision Theory

- Example for OC curve

The diameters of bolts are known to have a standard deviation of 0.0001 inch. A random sample of 10 bolts yields an average diameter of 0.2546 inch.

i) Test the hypothesis that the true mean diameter of bolts equals 0.255 inch, using $\alpha = 0.05$.

ii) What size sample would be necessary to detect a true mean bolt diameter of 0.2552 inch or more with a probability of at least 0.90, assuming $\alpha = 0.05$?



OC Curves for two-sided tests based the normal distribution with $\alpha = 0.05$.

$$d = \frac{\mu - \mu_0}{\sigma}$$

In[16]:= $d = \frac{0.2552 - 0.255}{0.0001}$

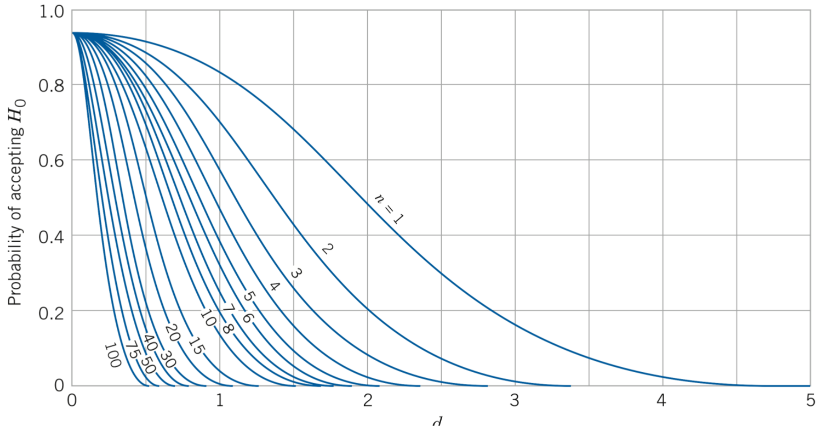
Out[16]= 2.

n = 3 is sufficient.

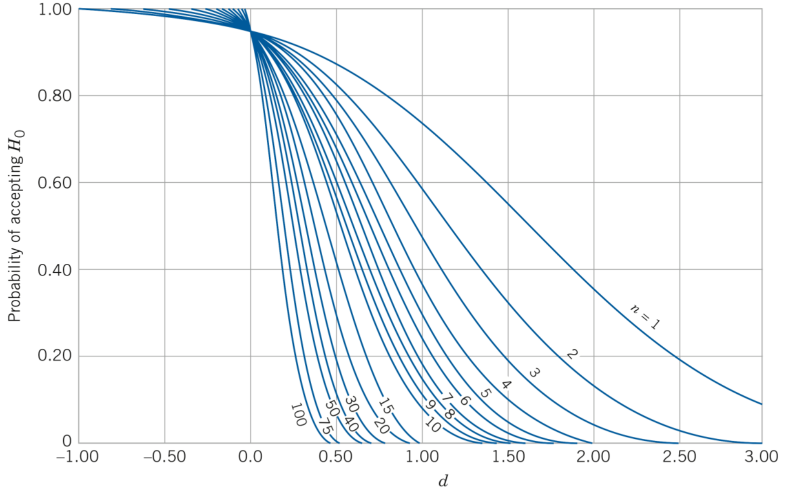
Neyman-Pearson Decision Theory

- More on OC curve
- Z-test

$$d = \frac{\mu - \mu_0}{\sigma}$$



OC Curves for two-sided tests based the normal distribution with $\alpha = 0.05$.

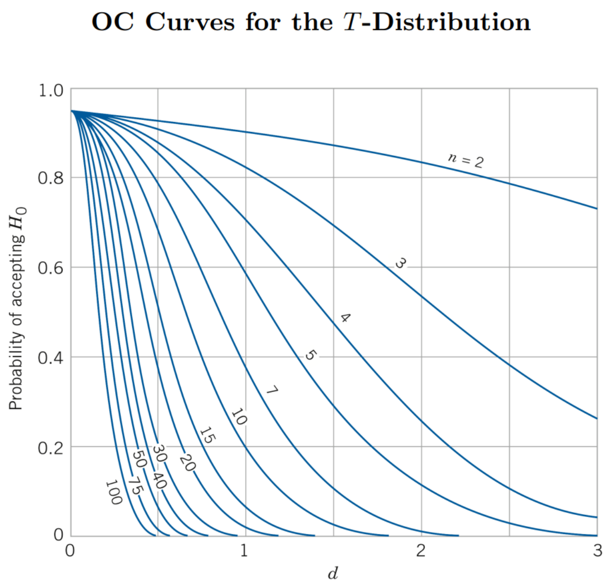


OC Curves for one-sided tests based the normal distribution with $\alpha = 0.05$.

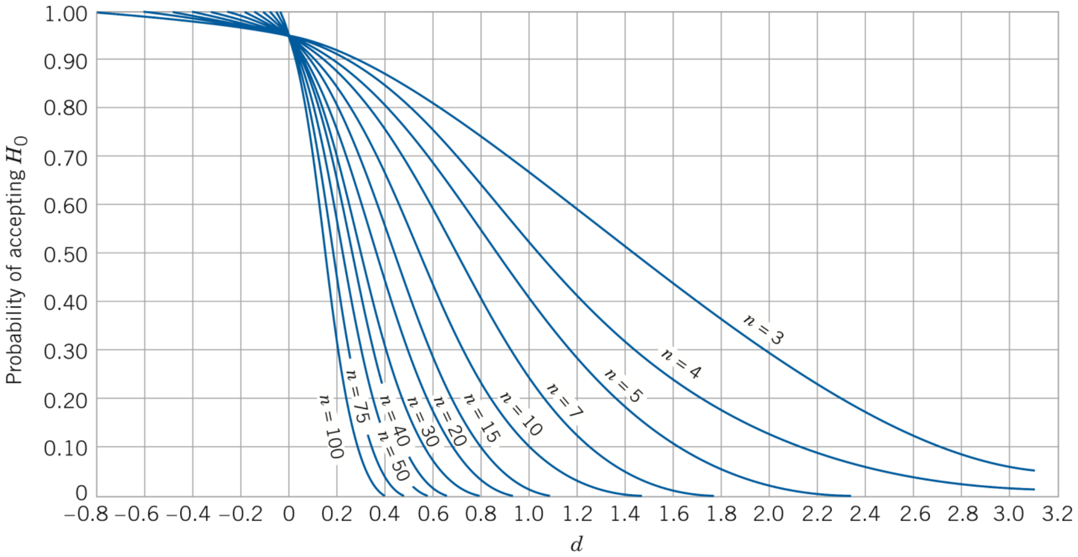
Neyman-Pearson Decision Theory

- More on OC curve
- T-test

$$d = \frac{|\mu - \mu_0|}{s}$$



OC Curves for two-sided tests based the T-distribution with $\alpha = 0.05$.



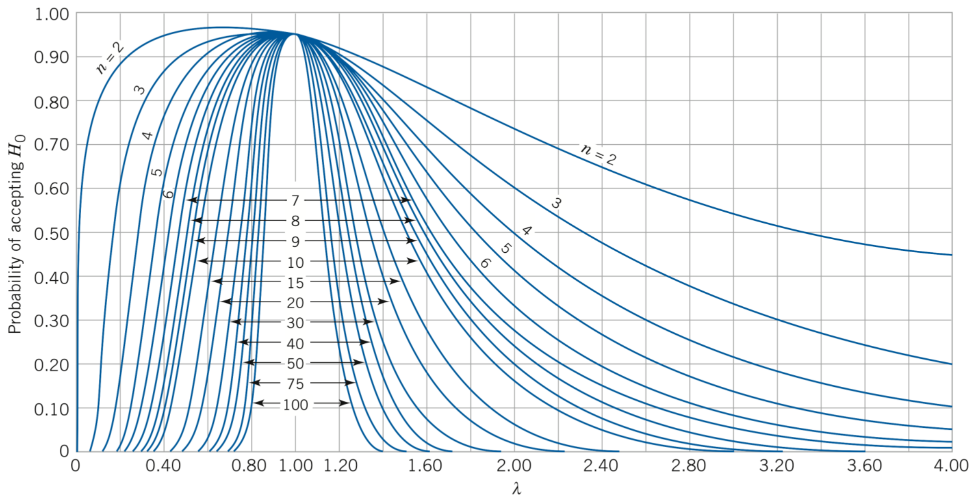
OC Curves for one-sided tests based the T-distribution with $\alpha = 0.05$.

- More on OC curve

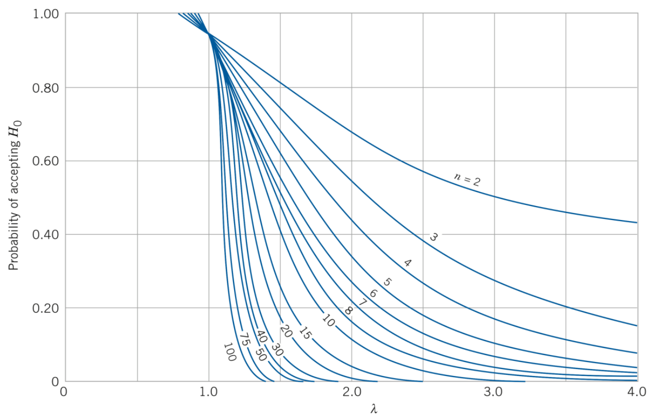
Note that the OC curves for the left- and right-tailed chi-squared distributions are distinct!

$$\lambda = \frac{\sigma}{\sigma_0}.$$

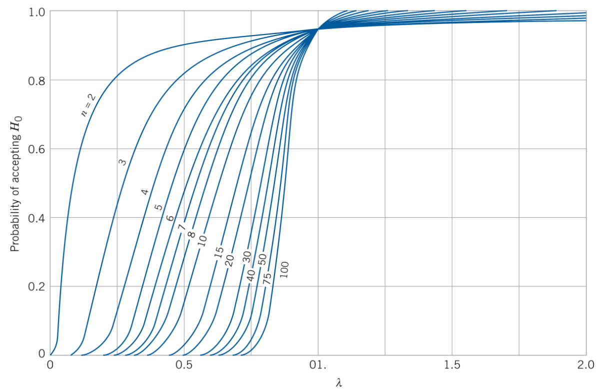
OC Curves for the Chi-Squared-Distribution



OC Curves for two-sided tests based the chi-squared-distribution with $\alpha = 0.05$.



OC Curves for one-sided (upper-tail) tests based on the chi-squared-distribution with $\alpha = 0.05$.



OC Curves for one-sided (lower-tail) tests based on the chi-squared-distribution with $\alpha = 0.05$.

Neyman-Pearson Decision Theory

- Summary

(i) Select appropriate hypotheses H_1 and H_0 and a test statistic;

(ii) Fix α and β for the test;

(iii) Use α and β to determine the appropriate the sample size;

(iv) Use α and the sample size to determine the critical region;

(v) Obtain the sample statistic; if the test statistic falls into the critical region, reject H_0 at significance level α and accept H_1. Otherwise, accept H_0.

Single Sample Tests for the Mean and Variance

- Test Statistic

When we suppose H_0 is true, some statistic will follow a certain distribution. We call such statistic a **test statistic**. For example, in z-test, $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ is the test statistic, which should follow a standard normal distribution. If the sample value is far from 0, then H_0 is rejected.
In the following hypothesis testing, we always use this kind of method.

Single Sample Tests for the Mean and Variance

- T-test
- Sample from a normal distribution (if large sample, no need to be normal)
- Probably don't know the overall variance
- Want to do a test on mean

For the null hypothesis $\mu = \mu_0$
Now suppose we don't know the overall variance σ^2. We need to use S^2 instead. And the test statistic is $T_{n-1} = \frac{\overline{X}-\mu_0}{S/\sqrt{n}}$, which follows a student-T distribution.
Then, the critical region at significance level α is given by
$|T_{n-1}| > t_{\alpha/2,n-1}$, or $\overline{x} \neq \mu_0 \pm t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}$

For $H_0 : \mu \leq \mu_0$, the critical region at significance level α is given by
$T_{n-1} > t_{\alpha,n-1}$, or $\overline{x} > \mu_0 + t_{\alpha,n-1}\frac{S}{\sqrt{n}}$

For $H_0 : \mu \geq \mu_0$, the critical region at significance level α is given by
$T_{n-1} < -t_{\alpha,n-1}$, or $\overline{x} < \mu_0 - t_{\alpha,n-1}\frac{S}{\sqrt{n}}$

Single Sample Tests for the Mean and Variance

- Chi-squared Test
- Sample from a normal distribution (This is a must)
- Probably don't know the overall variance
- Want to do a test on variance

Chi-squared test is used to test the hypotheses that are related to variance or standard deviation.
The test statistic is $\chi^2_{n-1} = \frac{(n-1)S^2}{\sigma_0^2}$. At significance level α, the critical region is:

- $H_0 : \sigma = \sigma_0$: $\chi^2_{n-1} > \chi^2_{\alpha/2,n-1}$ or $\chi^2_{n-1} < \chi^2_{1-\alpha/2,n-1}$
- $H_0 : \sigma \leq \sigma_0$: $\chi^2_{n-1} > \chi^2_{\alpha,n-1}$
- $H_0 : \sigma \geq \sigma_0$: $\chi^2_{n-1} < \chi^2_{1-\alpha,n-1}$

For β, you can still read it from the OC curve. Notice that $\lambda = \frac{\sigma}{\sigma_0}$

Non-parametric Single Sample Test for Median

- Sign Test for the Median
- Flexible

20.2. Sign Test. Let $X_1, \ldots, X_n$ be a random sample of size n from an arbitrary continuous distribution and let

$$Q_+ = \#\{X_k : X_k - M_0 > 0\}, \qquad Q_- = \#\{X_k : X_k - M_0 < 0\}.$$

We reject at significance level α

- $H_0 : M \leq M_0$ if $P[Q_- \leq k \mid M = M_0] < \alpha$,
- $H_0 : M \geq M_0$ if $P[Q_+ \leq k \mid M = M_0] < \alpha$,
- $H_0 : M = M_0$ if $P[\min(Q_-, Q_+) \leq k \mid M = M_0] < \alpha/2$.

Non-parametric Single Sample Test for Median

- Wilcoxon Signed Rank Test

20.4. Wilcoxon Signed Rank Test. Let $X_1, \dots, X_n$ be a random sample of size n from a symmetric distribution. Order the n absolute differences $|X_i - M|$ according to magnitude, so that $X_{R_i} - M_0$ is the R_ith smallest difference by modulus. If ties in the rank occur, the mean of the ranks is assigned to all equal values.

Let

$$W_+ = \sum_{R_i>0} R_i, \qquad |W_-| = \sum_{R_i<0} |R_i|.$$

We reject at significance level α

- $H_0\colon M \le M_0$ if $|W_-|$ is smaller than the critical value for α,
- $H_0\colon M \ge M_0$ if W_+ is smaller than the critical value for α,
- $H_0\colon M = M_0$ if $W = \min(W_+, |W_-|)$ is smaller than the critical value for $\alpha/2$.

20.5. Example. Returning to the previous example, we want to test $H_0\colon M \le 3.5$ and have the following observations, ordered from smallest to largest:

X_i	$X_i - M_0$	R_i	X_i	$X_i - M_0$	R_i
3	-0.5	-5.5	2	-1.5	-13
3	-0.5	-5.5	5	1.5	$+13$
3	-0.5	-5.5	5	1.5	$+13$
3	-0.5	-5.5	5	1.5	$+13$
4	0.5	$+5.5$	5	1.5	$+13$
4	0.5	$+5.5$	1	-2.5	-18
4	0.5	$+5.5$	6	2.5	$+18$
4	0.5	$+5.5$	6	2.5	$+18$
4	0.5	$+5.5$	6	2.5	$+18$
4	0.5	$+5.5$	6	2.5	$+18$

We calculate the sum of the negative ranks,

$$w_- = -5.5 - 5.5 - 5.5 - 5.5 - 13 - 18 = -53.$$

Consulting a table, the critical value for $n = 20$ and $\alpha = 0.05$ is 60. For $\alpha = 0.01$ it is 43. Since $|w_-|$ lies between these values, the P-value of the test is between 1% and than 5%, most likely around 2%-3%.

n	p 0.05	0.025	0.01	0.005
18	47	40	33	28
19	54	46	38	32
20	60	52	43	37

Non-parametric Single Sample Test for Median

- Wilcoxon Signed Rank Test Example

Exercise 1.

A company wants to test whether a new assembly line procedure increases the physical stress on its workers. It selects eleven workers to work for one day using each of the assembly line procedures. At the end of each day, their pulse frequency is measured:

Procedure 1	X	63	65	71	75	72	75	68	74	62	73	72
Procedure 2	Y	80	78	96	87	88	96	82	83	77	79	71

It is thought that the median pulse frequency is higher in Procedure 2 than in Procedure 1.

We perform a paired test and consider M_{Y-X}. Then we test

$$H_0\colon M_{Y-X} \leq 0.$$

M0

We will reject H_0 if $|W_-|$ is small. **(1/2 Mark)** We calculate $Y - X$:

$Y - X$	17	13	25	12	16	21	14	9	15	6	-1

(1/2 Mark) We can see from the table of $Y - X$ that there is only a single negative value of $Y - X$, which has rank 1. Therefore,

$$|W_-| = 1.$$

What is W+?

W+ = 2+3+4+…+10

(1/2 Mark) and $W = |W_-| = 1$. **(1/2 Mark)** According to the table for the Wilcoxon signed-rank test, we reject H_0 at the 5% level of significance if $W < 14$, so we can here reject H_0. **(1/2 Mark)**

n	p = 0.05	0.025	0.01	0.005
5	1			
6	2	1		
7	4	2	0	
8	6	4	2	0
9	8	6	3	2
10	11	8	5	3
11	14	11	7	5
12	17	14	10	7
13	21	17	13	10
14	26	21	16	13
15	30	25	20	16
16	36	30	24	19
17	41	35	28	23