6.0 A STATISTICAL TESTS

WE HAVE A RANDOM VARIABLE ξ OF WHICH WE ONLY KNOW n NUMERICAL VALUES X_i. THE SET OF X_i IS CALLED SAMPLE.

WE MAKE HYPOTHESIS ON THE DISTRIBUTION OF ξ, FOR EXAMPLE WE MAY SAY THAT ITS MEDIAN IS η_0. WE CALL THIS A NULL HYPOTHESIS H_0. THE ALTERNATIVE HYPOTHESIS H_1 IS, IN THIS CASE, EITHER $\eta \neq \eta_0$ OR $\eta > \eta_0$. IN THE FORMER CASE WE WILL CHECK THE HYPOTHESIS USING A TWO-TAILED TEST, IN THE LATTER A ONE-TAILED TEST.

THE P-VALUE OF A TEST IS THE PROBABILITY THAT, ASSUMING H_0 TO BE TRUE, THE SAMPLE IS NOT COMPLIANT WITH H_0 DUE TO BAD LUCK.

$$P\text{-VALUE} = P\left(\text{GETTING THE SAME OR A WORSE RESULT} \,\middle|\, H_0 \text{ IS TRUE}\right)$$

FIXING A SIGNIFICANCE LEVEL, USUALLY 5%, IF P-VALUE IS SMALLER WE REJECT HYPOTHESIS H_0, OTHERWISE WE DO NOT REJECT H_0

NOT REJECTING H_0 DOES NOT MEAN THAT IT IS TRUE, SIMPLY THAT WE DO NOT HAVE ENOUGH EVIDENCE TO SAY THAT IT IS FALSE. SINCE WE DO NOT WANT TO COMMIT THE ERROR OF REJECTING H_0 WHEN IT MAY BE TRUE, WE AVOID REJECTING WHENEVER WE DO NOT HAVE ENOUGH EVIDENCE.

6.0 B) SIGN TEST

TWO-TAILED

$$\begin{cases} H_0: & \eta = \eta_0 \\ H_1: & \eta \neq \eta_0 \end{cases}$$

FOR EXAMPLE THE SAMPLE IS
$\{1; 5; 8; 9; 10; 15; 25; 28\}$
$n = 8$

WE TAKE $\eta_0 = 6$

VALUES ON THE LEFT OF η_0 $S^- = 2$ ON THE RIGHT $S^+ = 6$

WE HAVE AN EQUAL OR WORSE RESULT WHENEVER WE HAVE $S^- = 2$ AND $S^+ = 6$ OR $S^- = 1$ AND $S^+ = 7$ OR $S^- = 0$ AND $S^+ = 8$ BUT ALSO WHEN $S^- = 6$ AND $S^+ = 7$, $S^- = 7$ AND $S^+ = 8$, $S^- = 8$ AND $S^+ = 0$.
THEREFORE WE HAVE TWO "TAILS", ONE WITH $S^- \leq 2$ AND THE OTHER WITH $S^- \geq 6$

$$\text{P-VALUE} = P(S^- \leq 2) + P(S^- \geq 6)$$

IF H_0 IS TRUE THE PROBABILITY THAT A SAMPLE'S ELEMENT FALLS ON THE LEFT, THUS INCREASING S^-, IS 50%. OTHERWISE IT DOES NOT INCREASE S^- WITH PROB. 50%.
THE SUM OF ELEMENTS ON THE LEFT, S^-, IS THUS A SUM OF BERNOULLI. THEREFORE $S^- = B(n, 50\%)$

$$\text{P-VALUE} = P\left(B\left(8, \tfrac{1}{2}\right) \leq 2\right) + P\left(B\left(8, \tfrac{1}{2}\right) \geq 6\right)$$

SINCE A BINOMIAL WITH $p = \frac{1}{2}$ IS SYMMETRIC

$$\text{P-VALUE} = 2P\left(B\left(8, \tfrac{1}{2}\right) \leq 2\right)$$

THIS VALUE CAN BE CALCULATED MANUALLY WITH:

6.0 C

$$P(B(n,p)=x)=\binom{n}{x}p^x\cdot(1-p)^{n-x}=\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

OR, USING BINOMIAL DISTRIBUTION TABLES WITH $p=\frac{1}{2}$, PAYING ATTENTION AT THE DIFFERENCE BETWEEN CUMULATIVE AND NON CUMULATIVE TABLES,

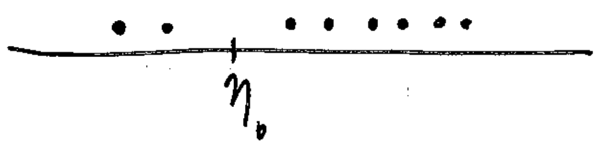
IN OUR CASE $\quad$ $p\text{-VALUE}=2\cdot P\left(B\left(8,\frac{1}{2}\right)\le 2\right)=2\cdot 0{,}1445=0{,}289$

IF $\alpha=30\%$, H_0 IS REJECTED.

IF $\alpha=20\%$, H_0 IS NOT REJECTED, WHICH MEANS THAT WE DO NOT WANT TO COMMIT THE ERROR OF REJECTING IT IN CASE IT MIGHT BE TRUE.

AFTER PERFORMING THE TWO-TAILED TEST, WE MAY PERFORM ONE-TAILED TEST. WE HAVE TWO ONE-TAILED TESTS:

$$\begin{cases} H_0: \eta \le \eta_0 \\ H_1: \eta > \eta_0 \end{cases} \qquad \begin{cases} H_0: \eta \ge \eta_0 \\ H_1: \eta < \eta_0 \end{cases}$$



USING AGAIN S^-, WE HAVE:

$p\text{-VALUE}=P(S^-\le 2)=0{,}1445$

THIS CASE IS VERY SIMILAR TO THE TWO-TAILED TEST. IN FACT, IN ONE OF THE ONE-TAILED TESTS, p-VALUE IS ALWAYS ONE HALF.

WITH $\alpha=20\%$ H_0 IS REJECTED.

WITH $\alpha=5\%$ H_0 IS NOT REJECTED.

$p\text{-VALUE}=P(S^-\ge 2)=0{,}9649$

THIS CASE IS VERY DIFFERENT FROM THE TWO-TAILED TEST. LOOK AT THE DATA: THEY ARE VERY "GOOD", SO A WORSE RESULT IS VERY EASY TO FIND, AND THEREFORE H_0 IS FOR SURE NOT REJECTED.

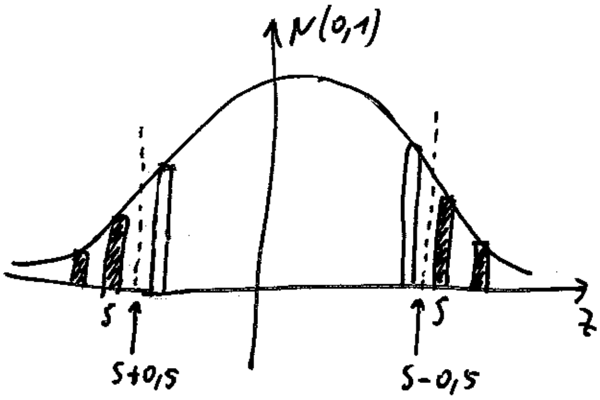
6.0 D

SIGN TEST FOR LARGE SAMPLE

SINCE S^- AND S^+ DISTRIBUTE LIKE $B(n, \frac{1}{2})$ WE CAN APPROXIMATE FOR $n \geq 10$ $B(n, \frac{1}{2})$ WITH A NORMAL.

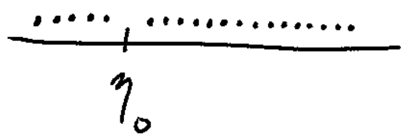
$$z = \frac{S - \frac{n}{2}}{\sqrt{n/4}} \sim N(0,1).$$

PAYING ATTENTION THAT S SHOULD BE CORRECTED USING $S+0,5$ OR $S-0,5$ DEPENDING ON WHETHER WE ARE CONSIDERING THE LEFT OR RIGHT TAIL.



FOR EXAMPLE, $n=20$ $S^-=5$ $S^+=15$, $P(S^- \leq 5) \approx$

$$\approx P\left(z < \frac{5,5-10}{\sqrt{5}}\right) = P(z < -2,013) = 0,0222$$



TWO-TAILED TEST HAS p-VALUE = 0,0444

ONE-TAILED TESTS HAVE p-VALUE $= P(S^- \leq 5) \approx 0,0222$

FOR $\begin{cases} H_0: \eta \leq \eta_0 \\ H_1: \eta > \eta_0 \end{cases}$ AND p-VALUE $= P(S^- \geq 5) \approx P\left(z > \frac{4,5-10}{\sqrt{5}}\right) =$

$= P(z > -2,907) = 0,9982$ FOR $\begin{cases} H_0: \eta \geq \eta_0 \\ H_1: \eta < \eta_0 \end{cases}$ (WHICH HAS CLEARLY A HIGHLY PROBABLE H_0)

6.0 E

WILCOXON SUM RANK TEST

WE USE W. SUM RANK TEST WHEN THE SAMPLE IS DIVIDED IN TWO GROUPS.

WE ORDER THE SAMPLE ACCORDING TO THE VALUE OF THE CONSIDERED VARIABLE AND ASSIGN SCORES FROM 1 TO n.

VARIABLE	3	5	6	6,5	7	9	9	13	19	23
GROUP	A	A	B	B	A	A	B	B	A	A
SCORE	1	2	3	4	5	6,5	6,5	8	9	10

WHEN THERE ARE TIES, WE ASSIGN AN AVERAGE SCORE.

$SUM_A = 1 + 2 + 5 + 6{,}5 + 9 + 10 = 33{,}5$ $\qquad$ $SUM_B = 3 + 4 + 6{,}5 + 8 = 21{,}5$

WE NOW TAKE THE GROUP WITH FEWEST MEMBERS (NOT THE GROUP WITH SMALLEST SCORE). IN THIS CASE B.

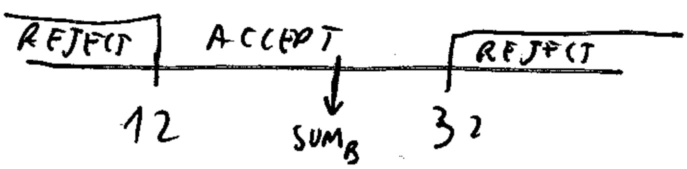
TWO-TAILED TEST

$$\begin{cases} H_0: & A \sim B \\ H_1: & A \nsim B \end{cases}$$

WE SIMPLY LOOK IN TABLES FOR $n_{SMALLER} = 4$ AND $n_{LARGER} = 6$

FOR $\alpha = 5\%$, IF WE HAVE A "ONE-TAILED TEST" TABLE, WE USE $2{,}5\%$.

IN THIS CASE



SUM_B IS IN THE MIDDLE $\Rightarrow$ H_0 IS NOT REJECTED.

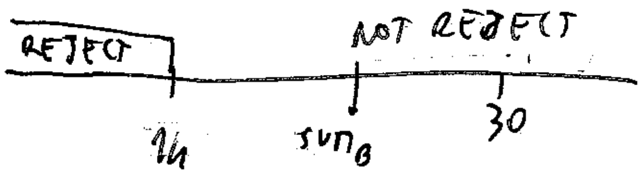
6.0.F

ONE-TAILED TEST

$$\begin{cases} H_0: \text{A SMALLER OR EQUAL B} \\ H_1: \text{A LARGER B (A SHIFTED TO THE RIGHT)} \end{cases}$$

WE ALWAYS TAKE THE GROUP WITH FEWER ELEMENTS, LOOK FOR $\alpha = 5\%$ IN THE TABLE

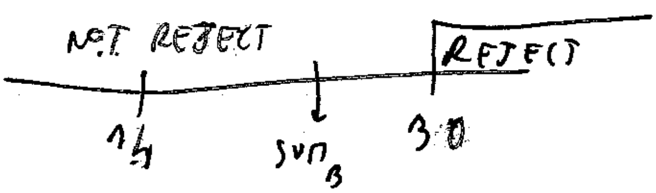
IN ORDER TO CHOOSE WHICH TAIL, WE NOTE THAT TO SATISFY H_0 WE WANT SUM_B TO BE LARGE, THEREFORE



H_0 IS NOT REJECTED

$$\begin{cases} H_0: \text{A LARGER OR EQUAL B} \\ H_1: \text{A SMALLER B (A SHIFTED TO THE LEFT)} \end{cases}$$

IN ORDER TO SATISFY H_0 WE WOULD LIKE SUM_B TO BE SMALL (WHICH MEANS B SHIFTED TO THE LEFT, THEREFORE A SHIFTED TO THE RIGHT)



H_0 IS NOT REJECTED

6.0 6 | WILCOXON SUM-RANK TEST FOR LARGE SAMPLE

WHEN BOTH GROUPS ARE AT LEAST 13, WE CAN USE A LARGE SAMPLE APPROXIMATION.

$$z = \frac{sum_i - n_i \cdot \frac{(n_A+n_B+1)}{2}}{\sqrt{\frac{n_A n_B (n_A+n_B+1)}{12}}} \sim N(0,1)$$

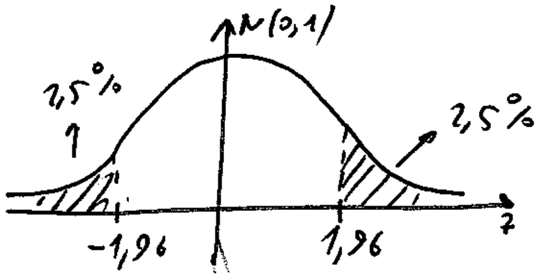
WHERE $i=A$ OR $i=B$ DEPENDING ON THE CONSIDERED ITEM

FOR EXAMPLE $n_A = 47$ $n_B = 36$

$sum_B = 1543$

$$z_B = \frac{1543 - 36 \cdot \frac{36+47+1}{2}}{\sqrt{\frac{36 \cdot 47 \cdot (36+47+1)}{12}}} \approx 0,2839$$

TWO-TAILED TEST WITH $\alpha = 5\%$

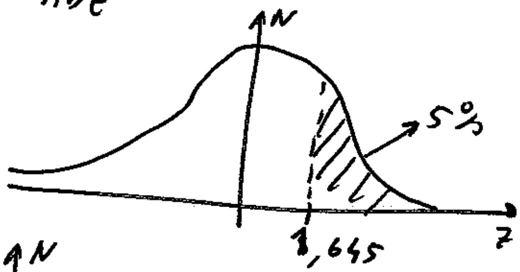


z_B FALLS IN THE MIDDLE $\Rightarrow H_0$ NOT REJECTED

ONE-TAILED TESTS WITH $\alpha = 5\%$:

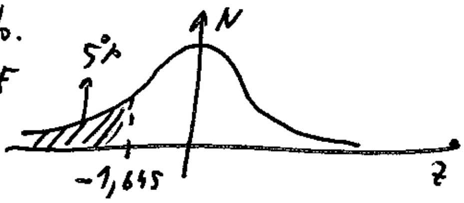
$\begin{cases} H_0: \text{A equal or right-shifted to B} \\ H_1: \text{A left-shifted to B} \end{cases}$

WE WANT SUM_B TO BE SMALL TO SATISFY H_0. THEREFORE REJECTION REGION IS IN THE "LARGE" SIDE



$\begin{cases} H_0: \text{A equal or left shifted} \\ H_1: \text{A right shifted} \end{cases}$

WE WANT B TO BE ON THE RIGHT, THEREFORE WITH LARGE RANKS, TO SATISFY H_0. THEREFORE REJECTION IS ON THE "SMALL" SIDE



6.0 H WILCOXON SIGNED-RANK TEST

WE USE W. SIGNED-RANK TEST WHEN THE SAME SAMPLE HAS TWO VALUES FOR EACH ELEMENT.

TWO-TAILED TEST

$$\begin{cases} H_0 : A = B \\ H_1 : A \neq B \end{cases}$$

SUBJECT	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
A	36,4	35,9	37,2	37	36,5	37,2	37,1	37
B	36,5	36,8	38	37,3	36,3	37,1	37,2	37,1

WE CALCULATE THE DIFFERENCE, REMEMBERING HOW WE DID IT.

$B - A =$ +0,1 +0,9 +0,8 +0,3 −0,2 −0,1 +0,1 +0,1

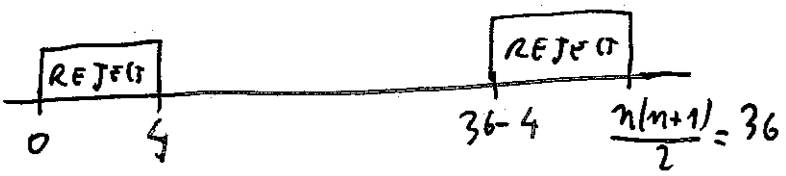
NOW WE ASSIGN RANKS TO ABSOLUTE VALUES, USING THE SIGN TO DIVIDE INTO GROUPS

RANK	2,5	8	7	6	5	2,5	2,5	2,5
SIGN	+	+	+	+	−	−	+	+

$SUM_+ = 2,5 + 8 + 7 + 6 + 2,5 + 2,5 = 28,5$

$SUM_- = 5 + 2,5 = 7,5$

WE LOOK IN THE TABLE FOR $\alpha = 5\%$ AND $n = 8$ FINDING 4



WHEN BOTH VALUES FALL IN THE CENTER, H_0 IS ACCEPTED.
HOWEVER, DUE TO THE SYMMETRICITY OF THIS TEST, IT IS ENOUGH TO CHECK THAT THE SMALLER IS NOT IN THE LEFT REJECTION WHICH AUTOMATICALLY IMPLIES THAT THE LARGER IS NOT IN THE RIGHT REJECTION.

6.0 I)

THEREFORE IT IS ENOUGH TO CHECK THAT THE SMALLEST SCORE (AND NOT THE ONE WITH FEWEST CASES AS IN W. SUM-RANK) IS NOT IN THE LEFT REJECTION

H_0 IS ACCEPTED

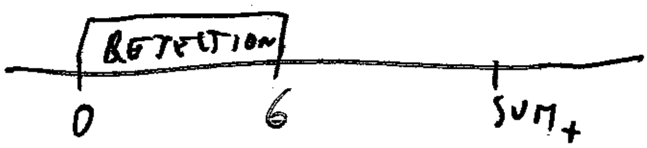
ONE-TAILED TESTS WITH $\alpha = 5\%$:

$$\begin{cases} H_0: A \leq B \\ H_1: A > B \end{cases}$$

CRITICAL VALUE IS 5

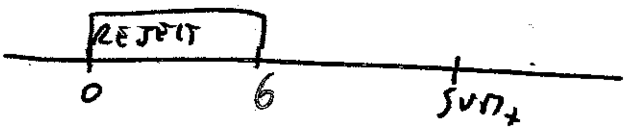
IF WE WANT TO USE ONLY THE LEFT REJECTION REGION, INSTEAD OF SELECTING THE APPROPRIATE REJECTION REGION, THIS TIME WE HAVE TO SELECT THE APPROPRIATE STATISTICS BETWEEN SUM_+ AND SUM_-

CHOOSING SUM_-, IN ORDER TO SATISFY H_0 $B-A \geq 0$, IT MUST BE SMALL, THEREFORE IT HAS REJECTION REGION ON THE RIGHT.
WE CHOOSE THEREFORE SUM_+, WHOSE REJECTION REGION IS ON THE LEFT

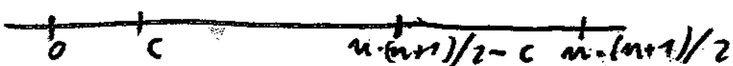


$$\begin{cases} H_0: A \geq B \\ H_1: A < B \end{cases}$$

CHOOSING SUM_-, TO SATISFY H_0 $B-A \leq 0$, IT MUST BE LARGE, THEREFORE REJECTION REGION IS ON THE RIGHT. THUS, SUM_+ IS THE CORRECT STATIS



NOTE: WE COULD DO THESE ONE-TAILED TESTS AS USUAL CHOOSING THE LEFT OR RIGHT TAIL INSTEAD OF CHOOSING SUM_+ OR SUM_-. IN THIS CASE WE SHOULD HOWEVER USE BOTH AREAS



6.0 J | LARGE SAMPLE

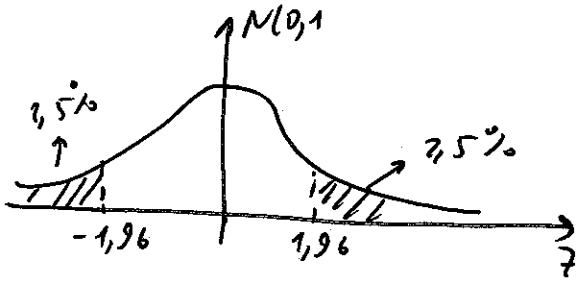
WHEN n IS LARGER THAN 30,

$$z = \frac{SUM - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \sim N(0,1)$$

FOR EXAMPLE $SUM_- = 83,5$ $SUM_+ = 381,5$ $n = 30$ DIFF = A−B

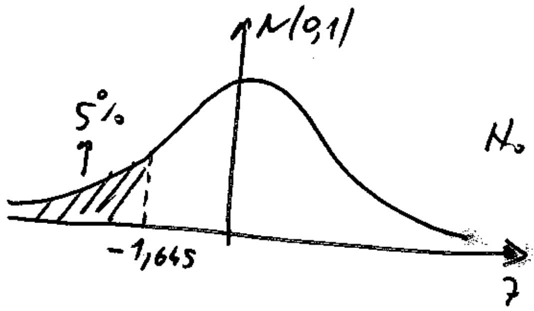
$z \approx -3,06$

$$\begin{cases} H_0: A = B \\ H_1: A \neq B \end{cases}$$



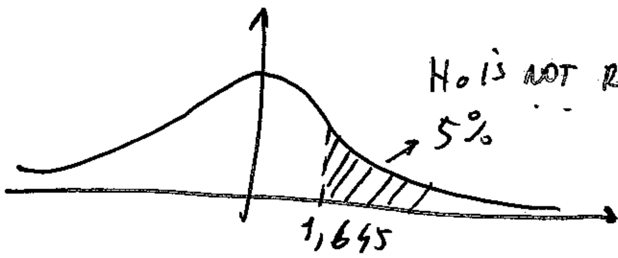
H_0 IS REJECTED

$$\begin{cases} H_0: A \leq B \\ H_1: A > B \end{cases}$$ → WE WANT SUM_- LARGE



H_0 IS REJECTED

$$\begin{cases} H_0: A \geq B \\ H_1: A < B \end{cases}$$ → WE WANT SUM_- SMALL



H_0 IS NOT REJECTED

NOTE THAT LARGE SAMPLE ONE-TAILED TEST IS DONE AS USUAL

IMPORTANT: WHENEVER WE DO SIGNED-TEST WE CAN ALSO DO SUM TEST CONSIDERING THE TWO VARIABLE VALUES AS TWO GROUPS OF SAMPLE'S ELEMENTS.

6.0 K KRUSKAL WALLIS

TEST WHETHER A VARIABLE DISTRIBUTION IS THE SAME ON <u>MORE</u> THAN TWO SAMPLE'S TYPES

<u>EXAMPLE</u>: 3 GROUPS, A, B, C, WITH 10 ELEMENTS

C; A; C; A; A; C; B; A; B AND C; A; A; A; C; C

C; A; B; B; C; A; B; B; C; B; C; A; B; B; B

WE PUT THE DATA IN THE SAME SAMPLE, AND RANK THEM FROM 1 TO 30 $n = 30$ $K = 3$

SCORE $A = 120$ SCORE $B = 270,5$ SCORE $C = 134.5$

H_0: DISTRIBUTIONS ARE THE SAME
H_1: DISTRIBUTIONS ARE NOT THE SAME

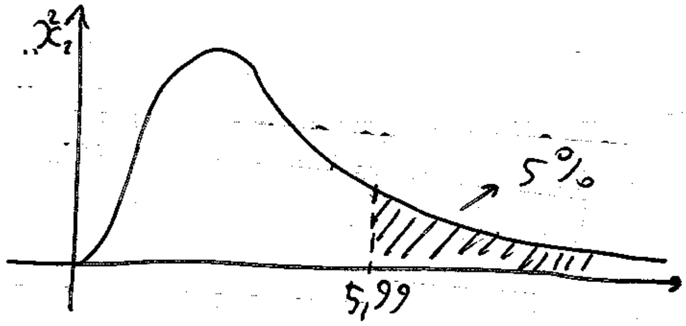
$n_A = 10$ $n_B = 10$ $n_C = 10$
$\alpha = 5\%$

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{K} \frac{R_j^2}{n_j} - 3(n+1) = \frac{12}{n(n+1)} \sum_{j=1}^{K} n_j \left(R_j - \bar{R} \right)^2$$

$H = 6.097$ THE PERFECT RESULT WOULD BE $H = 0$

H IS DISTRIBUTED AS χ^2 WITH K-1 DEGREES OF FREEDOM

$\chi^2_{0.05,2} = 5.99142$



H_0 IS REJECTED.

6.1

0,78 0,51 0,23 3,79 0,77 0,98 0,96 0,89

1) is $\eta = 1$? HW4) is $\eta \geq 1$?

2) is $\eta \leq 1$? $\alpha = 5\%$

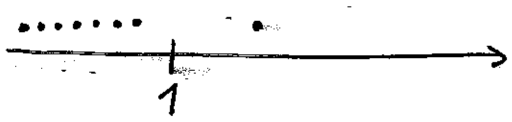
HW 3) is $\eta \leq 0,8$?

$$\begin{cases} H_0: \eta = 1 \\ H_1: \eta \neq 1 \end{cases} \qquad S^- = 7 \qquad S^+ = 1$$

$p\text{-VALUE} = 2 \cdot P(S^+ \leq 1) \approx 2 \cdot 0,03516 = 0,07032 = 7,032\%$

WE DO NOT REJECT H_0

$$\begin{cases} H_0: \eta \leq 1 \\ H_1: \eta > 1 \end{cases}$$

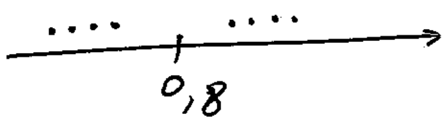


THE RESULT IS ALREADY "GOOD". A WORSE ONE WOULD BE $S^+ \geq 1$

$p\text{-VALUE} = P(S^+ \geq 1) \approx 0,9961 = 99,61\%$ WE ACCEPT H_0

IN FACT, DATA WERE SO EVIDENT THAT H_0 IS CORRECT, THAT WE GOT A VERY LARGE p-VALUE.

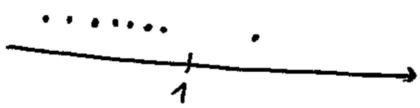
$$\begin{cases} H_0: \eta \leq 0,8 \\ H_1: \eta > 0,8 \end{cases} \qquad S^- = 4 \qquad S^+ = 4$$



A WORSE RESULT IS WHEN $S^- \leq 4$

$p\text{-VALUE} = P(S^- \leq 4) = 63,7\%$ H_0 IS NOT REJECTED

$$\begin{cases} H_0: \eta \geq 1 \\ H_1: \eta < 1 \end{cases}$$



A WORSE RESULT IS SIMPLY WHEN $S^+ \leq 1$

$p\text{-VALUE} = P(S^+ \leq 1) \approx 0,03516 = 3,516\%$ H_0 IS REJECTED

□

6.2A

SAYS

FROM A CONTINOUS DISTRIBUTION WE EXTRACT THE FOLLOWING SAMPLE

4,31 3,99 3,43 3,72 2,78 2,65 3,13 2,53 2,92 3,76

IS THE MEDIAN $= 3{,}80$? $\alpha = 10\%$

HW: ALSO WITH LARGE SAMPLE APPROXIMATION

$$\begin{cases} H_0 : \eta_x = 3{,}80 \\ H_1 : \eta_x \neq 3{,}80 \end{cases}$$

IF H_0 IS TRUE $\eta_x = 3{,}80 \Rightarrow$ 50% OF CASES ON THE LEFT OF 3,80 AND 50% ON THE RIGHT

HERE WE HAVE 8 ON THE LEFT AND 2 ON THE RIGHT

$$p\text{-VALUE} = 2\left(P(S^- = 8) + P(S^- = 9) + P(S^- = 10)\right) =$$

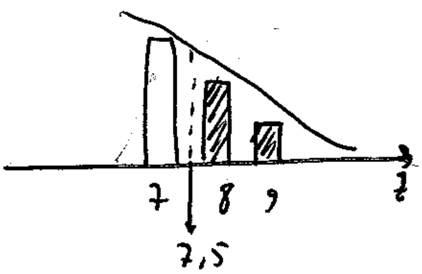
$$= 2\binom{10}{8}\frac{1}{2^{10}} + 2\binom{10}{9}\frac{1}{2^{10}} + 2\binom{10}{10}\frac{1}{2^{10}} = \frac{1}{2^9}\left(\frac{10!}{8!\,2!} + \frac{10!}{9!\,1!} + 1\right) =$$

$$= \frac{1}{2^9}(45 + 11) = \frac{56}{2^9} = \frac{28}{2^8} = \frac{14}{2^7} = \frac{7}{2^6} \approx 0{,}109$$

WE DO NOT REJECT H_0 ,

USING INSTEAD LARGE SAMPLE APPROXIMATION:

$$p\text{-VALUE} = 2 \cdot P(S^- \geq 8)$$



6.2 B

$$P\text{-VALUE} \approx 2 \cdot P\left(Z > \frac{5 - 0{,}5 - 0{,}5 \cdot 10}{\sqrt{10/4}}\right) = 2 \cdot P(Z > 1{,}58) \approx$$

$$\approx 2 \cdot 0{,}0571 = 0{,}1142$$

H_0 IS ACCEPTED.

VALUE IS VERY CLOSE BUT NOT EXACT, SINCE $n = 10$ IS THE MINIMUM VALUE FOR WHICH

$$0 < p \pm 3 \sqrt{\frac{p(1-p)}{n}} < 1$$ IS SATISFIED.

IN THIS CASE
$$p \pm 3 \sqrt{\frac{p(1-p)}{n}} = 0{,}5 \pm 3 \sqrt{\frac{0{,}25}{10}} = 0{,}5 \pm 0{,}47 =$$

$$= \begin{cases} 0{,}97 \\ 0{,}03 \end{cases}$$

$\square$

6.3

TWO SET OF ECONOMISTS MAKE FORECASTING ON INFLATION RATE

A) 3,1 4,8 2,3 5,6 0 2,9

B) 4,4 5,8 3,9 8,7 6,3 10,5 10,8

DO A AND B HAVE THE SAME DISTRIBUTION ?
$\alpha = 5\%$

WE USE WILCOXON RANK SUM TEST TWO-TAILED FOR SMALL SAMPLES

RANK	VAL	SET
1	0	A
2	2,3	A
3	2,9	A
4	3,1	A
5	3,9	B
6	4,4	B
7	4,8	A
8	5,6	A
9	5,8	B
10	6,3	B
11	8,7	B
12	10,5	B
13	10,8	B

SET A SCORE = 1 + 2 + 3 + 4 + 7 + 8 = 25

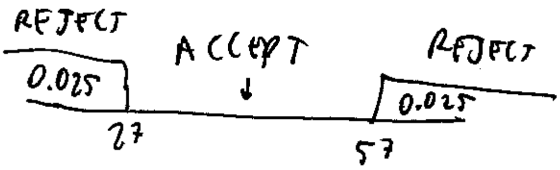
SET B SCORE = 5 + 6 + 9 + 10 + 11 + 12 + 13 = 66

WE TAKE THE SET WITH FEWER SUBJECTS

$H_0: A \sim B$

$H_1: A \nsim B$

TO ACCEPT IT, WHEN WE PUT $\alpha = 0{,}05$, A SCORE MUST BE BETWEEN 27 AND 57 (VALUES FOR $\alpha = 0.025$)



WE REJECT H_0 HYPOTHESIS

□

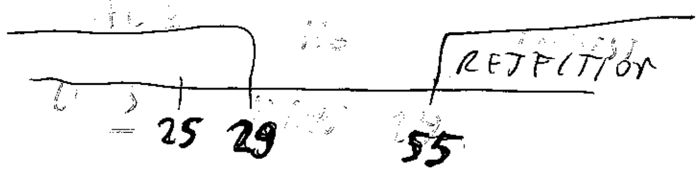
6.4 THE SAME SET OF ECONOMISTS, PROVE THAT DISTRIBUTION "A" IS NOT SHIFTED TO THE RIGHT WITH RESPECT TO DISTRIBUTION "B" $\alpha = 5\%$

WE USE WILCOXON RANK SUM TEST
ONE-TAILED FOR SMALL SAMPLES

SET A SCORE = 25   SET B SCORE = 66

$$\begin{cases} H_0: A \leq B \\ H_1: A > B \end{cases}$$

WITH $\alpha = 0{,}05$



⇓

WE DO NOT REJECT H_0

D

(6.5) WE WANT TO TEST TWO FERTILIZERS A AND B

A PRODUCES CROPS PER HECTARE

HW 37, 40, 33, 29, 42, 33, 35, 28

WHILE B 65, 35, 47, 52

WITH $\alpha = 0{,}1$ TEST WHETHER THEY ARE EQUALLY EFFICIENT

WE USE WILCOXON SUM RANK TEST

28 A	1
29 A	2
33 A	3,5
33 A	3,5
35 A	5,5
35 B	5,5
37 A	7
40 A	8
42 A	9
47 B	10
52 B	11
65 B	12

$$\text{SUM } A = 1+2+3{,}5+3{,}5+5{,}5+7+8+9 = 39{,}5$$

$$\text{SUM } B = 5{,}5+10+11+12 = 38{,}5$$

$n_A = 8 \quad n_B = 4$

$$\begin{cases} H_0: A \sim B \\ H_1: A \not\sim B \end{cases}$$

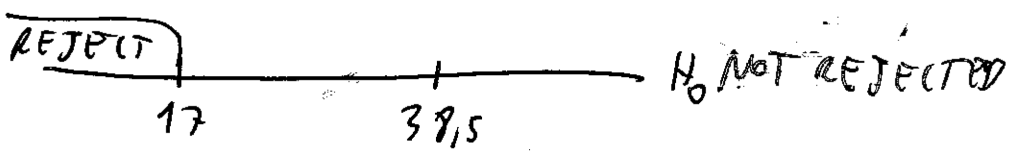
FOR 4/8 AND $\alpha = 0{,}1$ WE HAVE $T_L = 15$ $T_U = 37$

$\Downarrow$

WE REJECT H_0 SINCE $38{,}5 > 37$

<u>NOTE</u> : IF WE PUT $\begin{cases} H_0: A \leq B \\ H_1: A > B \end{cases}$

WE WANT B TO HAVE A LARGE SCORE TO SATISFY H_0.



$\square$

6-6 PROVE THAT PRICES FOR HOUSES HAS NOT INCREASED THIS YEAR WITH $\alpha = 5\%$.

	This year		Last year	
1	$1.033.613	T	$324.625	L
2	$1.036.494	T	$439.080	L
3	$358.462	T	$135.607	L
4	$266.081	T	$444.821	L
5	$1.640.898	T	$440.286	L
6	$1.026.114	T	$223.515	L
7	$1.714.962	T	$356.739	L
8	$1.568.402	T	$360.697	L
9	$1.322.752	T	$753.232	L
10	$1.643.641	T	$148.408	L
11	$942.094	T	$476.334	L
12	$837.461	T	$178.165	L
13	$346.841	T	$413.302	L
14	$874.395	T	$408.840	L
15	$1.634.329	T	$278.033	L
16	$1.609.172	T	$125.527	L
17	$217.584	T	$297.383	L
18	$303.575	T	$515.574	L
19	$1.723.826	T	$215.702	L
20	$1.142.444	T	$546.076	L
21	$786.239	T	$1.172.569	L
22	$567.131	T	$277.654	L
23	$967.001	T	$331.212	L
24	$765.582	T	$538.585	L
25	$362.392	T	$873.764	L
26	$583.035	T	$296.540	L
27	$1.508.420	T	$779.144	L
28	$445.971	T	$693.151	L
29	$1.537.249	T	$341.496	L
30	$225.847	T	$673.082	L
31	$560.983	T	$502.065	L
32	$400.807	T	$185.066	L
33	$1.520.335	T	$378.280	L
34	$2.344.521	T	$872.115	L
35	$1.470.307	T	$253.032	L
36	$158.764	T	$643.238	L
37			$497.542	L
38			$906.152	L
39			$453.571	L
40			$546.165	L
41			$995.621	L
42			$532.845	L
43			$121.773	L
44			$1.209.890	L
45			$237.892	L
46			$1.064.689	L
47			$788.157	L

$$\begin{cases} H_0: T \le L \\ H_1: T > L \end{cases}$$

$n_T = 36 \quad n_L = 47$

WE USE WILCOXON SUM RANK TEST ONE-TAILED FOR LARGE SAMPLES

IF WE PUT DATA IN ORDER AND ASSIGN SCORES, WE OBTAIN

SCORE L: 1943

SCORE T: 1543

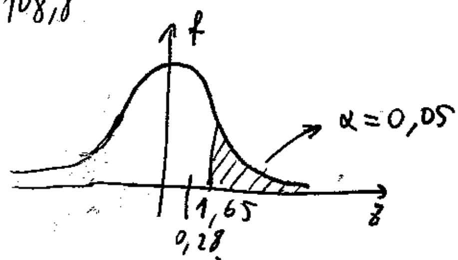
$$z = \frac{t_T - n_T \frac{(n_L + n_T + 1)}{2}}{\sqrt{\frac{n_L n_T (n_L + n_T + 1)}{12}}}$$ IS STANDARD NORMAL

$t_T = 1543 \quad n_L = 47 \quad n_T = 36$

$\Downarrow$

$$z_T = \frac{1543 - 36 \cdot \frac{84}{2}}{\sqrt{47 \cdot 36 \cdot \frac{84}{12}}} = \frac{1543 - 1512}{\sqrt{11844}} = $$

$$\approx \frac{31}{108{,}8} \approx 0{,}2849$$



H_0 NOT REJECTED

6.7 10 STUDENTS ARE ASKED TO RANK
2 PROFESSORS

PROF 1	6	8	4	9	4	7	6	5	6	8
PROF 2	4	5	5	8	1	9	2	3	7	2

IS THE EVALUATION EQUIVALENT ?

FOR THIS PROBLEM WE CAN USE WILCOXON SUM RANK TEST, OR, SINCE THE DATA ARE COUPLED, WE CAN USE WILCOXON SIGNED RANK TEST ON THE DIFFERENCE OF EACH COUPLE OF EVALUATION.

DIFF 2 3 −1 1 3 −2 4 2 −1 6

WE USE WILCOXON SIGN RANK TEST TWO-TAILED
FOR SMALL SAMPLES

SO WE CALCULATE THE ABSOLUTE VALUE AND RANK IT

ABS DIFF 2 3 1 1 3 2 4 2 1 6

RANK	2	2	2	5	5	5	7,5	7,5	9	10
POSIT/NEGAT	−	+	−	+	−	+	+	+	+	+
ABS. VALUE	1	1	1	2	2	2	3	3	4	6

$$\begin{cases} H_0 : P1 \sim P2 \\ H_1 : P1 \not\sim P2 \end{cases}$$

$SUM_- = 2+2+5 = 9$ $SUM_+ = 2+5+5+7,5+7,5+9+10 =$
$= 46$

T_0 WITH $n=10$ AND $\alpha = 0,05$ is 8

REJECT | SUM− SUM+
8

WE DO NOT REJECT H_0

(6-8) A 30 PATIENTS RECEIVE A MEDICINE TO INCREASE BLOOD PRESSURE. PRESSURE AFTER MINUS PRESSURE BEFORE ARE:

20 30 −10 10 30 −20 40 20 −10 60 80 40 0 30 50 −10 −30
0 20 10 70 −10 −20 30 50 60 30 10 0 10

IS PRESSURE AFTER NOT LARGER THAN PRESSURE BEFORE?

WE USE WILCOXON SIGNED RANK TEST ONE-TAILED FOR LARGE SAMPLES

RANK	2	7,5	14	19,5	23,5	25,5	27,5	29	30
POS/NEG	− − −	− + − − + − + +	+ − + + −	+ + + − + +	+ +	+ +	+ +	+	+
ABS VAL	0 0 0	10 10 10 10 10 10 10 10	20 20 20 20 20	30 30 30 30 30 30	40 40	50 50	60 60	70	80

HERE $\begin{cases} H_0: P.A. \leq P.B. & \text{(ZERO OR NEGATIVE VALUES)} \\ H_1: P.A. > P.B. & \text{(POSITIVE VALUES)} \end{cases}$

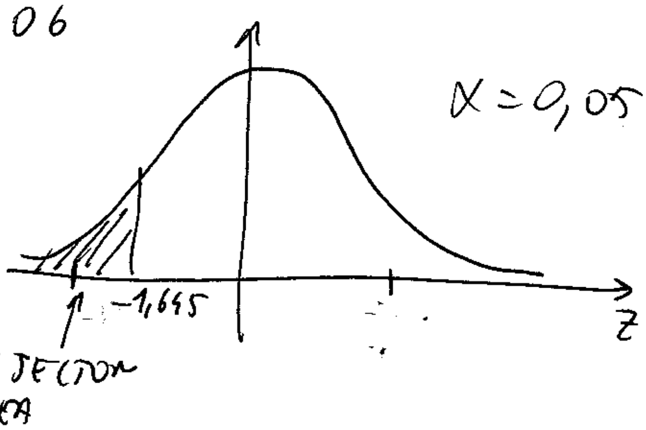
$t_- = 2+2+2+7{,}5\cdot 4+14\cdot 2+19{,}5 = 83{,}5$

$t_+ = 7{,}5\cdot 4+14\cdot 3+19{,}5\cdot 5+23{,}5\cdot 2+25{,}5\cdot 2+27{,}5\cdot 2+29+30 =$
$= 381{,}5$

$n = 30$ WE USE $z_- = \dfrac{t_- - n(n+1)/4}{\sqrt{n(n+1)\dfrac{(2n+1)}{24}}} = \dfrac{83{,}5 - 232{,}5}{\sqrt{2363{,}75}} =$

$= \dfrac{-149}{48{,}61} = -3{,}06$

$\alpha = 0{,}05$



WE REJECT H_0 ⇒ THE MEDICINE DOES NOT WORK WITH CONFIDENCE OF 95%.

□

6.8 B TWO-TAILED TEST

$$\begin{cases} H_0: \ P.A. = P.B. \\ H_1: \ P.A. \neq P.B. \end{cases}$$

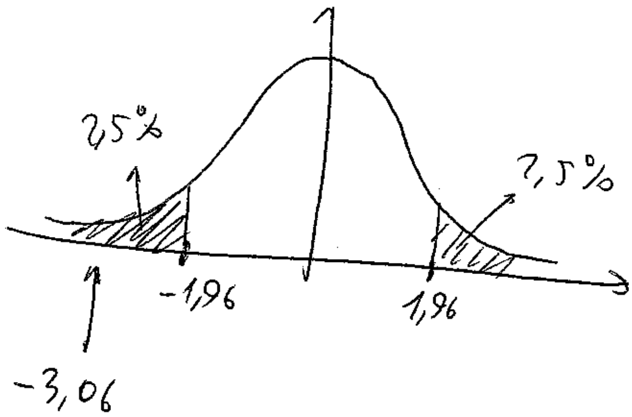
WE HAVE THE PROBLEM OF ZEROS !

THEY CLEARLY CONTRIBUTE TOWARDS H_0, SO WE MAY COUNT THEM IN THE SCORE WHICH IS LOWEST. OR WE MAY CONSIDER THEM HALF POSITIVE AND HALF NEGATIVE

$SUM_- = 83,5$ $SUM_- = 80,3$

$SUM_+ = 381,5$ $SUM_+ = 384,5$

$n = 30$ $Z_- = -3,06$ OR $Z_- = -3,12$



WE REJECT H_0

(6.9) WE HAVE 3 POPULATIONS A, B, C.
WITH $n_A = n_B = n_C = 15$ AND WITH DATA
WHICH CAN BE RANKED AS
$R_A = 230$ $R_B = 440$ $R_C = 365$

$\alpha = 0,05$ DO THESE DISTRIBUTIONS DIFFER?

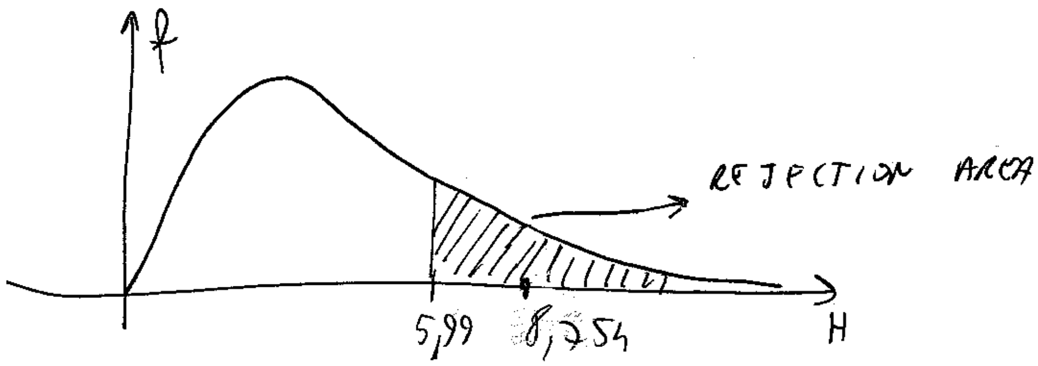
$H_0: A \sim B \sim C$
$H_1: A \nsim B$ OR $A \nsim C$ OR $B \nsim C$

WE USE KRUSKAL-WALLIS TEST

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{3} \frac{R_j^2}{n_j} - 3(n+1) = \frac{12}{45 \cdot 46} \left(\frac{230^2}{15} + \frac{440^2}{15} + \frac{365^2}{15} \right) - 3 \cdot 46 =$$

$$= \frac{12}{45 \cdot 46 \cdot 15} (379725) - 138 = 8,754$$

H IS DISTRIBUITED LIKE A χ_2. WITH $\alpha = 0,05$ $\chi = 5,99$



WE REJECT H_0

D

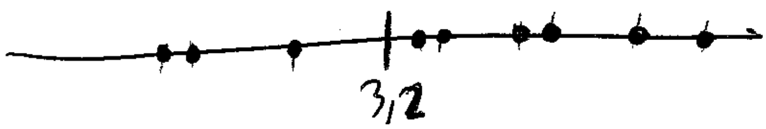
6.10

4,31 3,72 2,78 3,25 2,65 3,23 4,53 2,92 3,76

HW ARE EXTRACTED FROM CONTINUOUS DISTRIBUTIO

IS MEDIAN $\geq 3,2$?

$$\begin{cases} H_0: \eta \geq 3,2 \quad (\eta \geq 3,2) \\ H_1: \eta < 3,2 \end{cases}$$



$S^- = 3$ $\qquad$ $S^+ = 6$

THE IDEAL SITUATION WOULD BE S^- EQUAL TO 4,5 OR SMALLER (SINCE A SMALL S^- IS AN EVIDENCE THAT OUR η_0 IS TOO MUCH ON THE LEFT AND THE REAL ONE SHOULD BE ON THE RIGHT)

p-VALUE = P(THIS CASE OR A WORSE ONE) =

$$= P(S^- = 3) + P(S^- > 3) = P(S^- \geq 3) =$$

$$= 1 - P(S^- \leq 3) = 1 - 0,2539 = 0,7461$$

WITH α UP TO 0,7461 $\quad H_0$ IS NOT REJECTED

$\Downarrow$

MEDIAN.. CAN BE $\geq 3,2$

$\triangleright$

6.11

HW: 7 STUDENTS HAVE PASSED MATH AND LAW. WITH THESE GRADES:

LAW	30	29	25	26	28	27	30
MATH	25	28	26	27	29	30	24

IS EACH STUDENT EQUALLY GOOD IN MATH AND LAW? $\alpha = 0.05$

DIFF	5	1	-1	-1	-1	-3	6

SIGN	+	+	-	+	-	-	-
ABS. VALUE	6	5	3	1	1	1	1
RANK	7	6	5	2,5	2,5	2,5	2,5

$$\begin{cases} H_0: \text{LAW} \sim \text{MATH} \\ H_1: \text{LAW} \nsim \text{MATH} \end{cases}$$

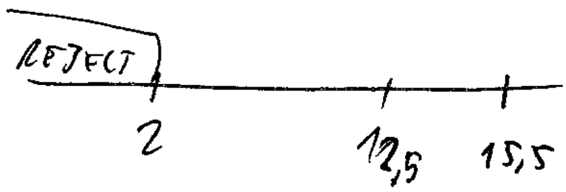
POS SUM = 15.5 NEG SUM = 12.5

$n = 7$ CRITICAL VALUE = 2

NEG SUM > CRITICAL VALUE

$\Downarrow$

WE DO NOT REJECT H_0



6.12

THE MANUFACTURER OF CD PLAYERS HAS ESTABLISHED THAT THE MEDIAN TIME TO FAILURE IS 5250 HOURS. A SAMPLE OF 20 CD-PLAYERS IS TESTED AND 14 EXCEEDED 5250 HOURS WHILE 6 FAILED BEFORE.

IS THERE EVIDENCE THAT THE MEDIAN IS NOT 5250 HOURS?

$\alpha = 10\%$

SIGN TEST
$$\begin{cases} H_0 & \eta = 5250 \\ H_1 & \eta \neq 5250 \end{cases}$$

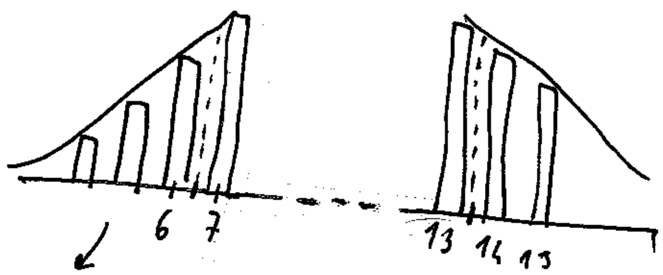
$$S^+ = 14 \qquad S^- = 6$$

IF WE USE SMALL SAMPLE:

$$p\text{-VALUE} = P(S^+ \geq 14) + P(S^+ \leq 6) = 2 \cdot P(S^+ \leq 6) = \text{FROM BINOMIAL TABLES}$$
$$= 2 \cdot 0{,}058 = 0{,}116$$

IF WE USE LARGE SAMPLE:

$$Z = \frac{S^+ \pm 0{,}5 - 0{,}5 \cdot 20}{0{,}5\sqrt{20}}$$

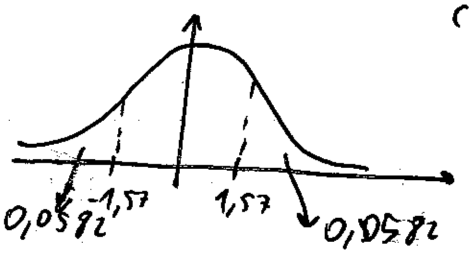


FOR 6 WE CHOOSE +0,5

FOR 14 WE CHOOSE -0,5

$$Z^+ = \frac{14 - 0{,}5 - 10}{0{,}5\sqrt{20}} \approx 1{,}57$$

$$Z^- = \frac{6 + 0{,}5 - 10}{0{,}5\sqrt{20}} \approx -1{,}57$$



p-VALUE OF TWO-TAILED TEST IS 0,116

WITH $\alpha = 10\%$ WE CANNOT REJECT $\eta = 5250$