

## Index of 7-L: One and two discrete samples, Non-parametric methods

Page	Title
1	Practical information
2	Inference for proportions – Overview
3	Choice of method for proportions
4	Inference for 1 proportion – Details
5	1 proportion: Aphid drops
6	Inference for 2 independent proportions – Details
7	2 independent proportions: Echinacea for common cold
8	Nonparametric (distribution-free) methods
9	1 sample: Sign test
10	McNemar's test
11	Ranks
12	2 samples: Wilcoxon–Mann–Whitney test
13	Example for 2-sample W-M-W test
14	1 sample: Wilcoxon's signed rank test
15	Example for 1-sample Wilcoxon test
16	Notes on home assignment I
17	Mid-term practical info
18	Summary notes

## PRACTICAL INFORMATION

### Today's lecture:

- inference for **one** and **two proportions**:<sup>1</sup>
  - \* similar breakdown of designs as for continuous data,
  - \* mix of familiar  $z$ -type inference and new approaches,
  - \* inference for 2 proportions continues in Session 8 (Two-way tables),
- non-normal data (6L–9) and **nonparametric methods** for one and two (continuous) samples — **sign test** and **rank-based tests**:<sup>2</sup>
  - \* the sign test can be understood as a binomial test,<sup>3</sup>
  - \* for rank-based tests, all calculations are based on statistical software.

### Scheduling notes:

- first home assignment done, second home assignment posted (due March 6),
- information about **mid-term exam**:
  - \* Thursday March 5, 9:00-10:00pm, AVC Lecture Theatre B,
  - \* new course syllabus page posted (for mid-term),
  - \* see (mid-term) exams from previous years for examples.
- **optional**: if you want to use own data for home assignment 4, you should start preparing the data and project outline (due 3/19), see homepage project guidelines.

---

<sup>1</sup> PSLS 4e: Chapters 19-20; S: Chapter 8-9 (parts); IPS 7e: Chapter 8.

<sup>2</sup> PSLS Supplementary Chapter 27 on rank-based tests; **note**: all references to hand calculations can be skipped.

<sup>3</sup> Despite the sign test not being covered in PSLS and S texts, it is still included in the course!

## INFERENCE FOR PROPORTIONS – OVERVIEW

**Basic assumption:** binomial setting(s)  $\Rightarrow$  binomial distribution  $B(n, p)$  for number of “successes”, where  $n$  = number of “trials” and  $p$  = probability of success.

Same **3 fundamental designs** as for quantitative data:

- **one sample** — one binomial distribution, parameter of interest is  $p$ ,
- **two independent samples** — two binomial distributions  $\Rightarrow$  focus on  $p_1 - p_2$ ,<sup>4</sup>
- **two paired samples** — not in textbooks, but discussed as a sign test.<sup>5</sup>

**Statistical inference:**

- **estimation:** always use sample proportions,
- several approaches for **confidence intervals:**
  - \* classical<sup>6</sup> approximation of  $B(n, p)$  by  $N(np, \sqrt{np(1-p)})$ ,
  - \* “plus four” (Wilson) adjustment of classical approach,
  - \* “exact” based on binomial distribution (1-sample setting only),
- several approaches for **tests:**
  - \* classical<sup>6</sup>  $z$ -test approximation,
  - \* exact based on binomial or other distributions.

---

<sup>4</sup>  $p_1 - p_2$  is the **risk difference**, rather than risk ratio or odds ratio (Epidemiology).

<sup>5</sup> Two paired samples often lead to McNemar’s test or to  $\kappa$  (kappa)-calculations.

<sup>6</sup> “Classical” (also “normal approximation”) refers to methods based on the standard normal ( $z$ ) distribution.

## CHOICE OF METHOD FOR PROPORTIONS

**Issue:** several methods exist for both CI and test across all designs  $\Rightarrow$  we need guidelines to choose a good method.

**Principle:** choice of method should be based on data dimensions, with separate guidelines for different inferential procedures (CI and test) and for different designs.

PSLS **guidelines:** <sup>7</sup>

Design	Method	Conditions (all must be met)	(Notes)
1-sample $(n, \hat{p})$	classical CI	$n\hat{p} \geq 15; n(1-\hat{p}) \geq 15$	$n\hat{p} \sim \#$ positives $n(1-\hat{p}) \sim \#$ negatives
	“plus four” CI	$n \geq 10$	
	“exact” CI	no conditions	
$H_0 : p = p_0$	z-test	$np_0 \geq 10; n(1-p_0) \geq 10$	1-sample exact methods based on binomial distrib.
	exact test	no conditions	
2-sample indep. $(n_1, \hat{p}_1, n_2, \hat{p}_2)$	classical CI	$n_1\hat{p}_1 \geq 10; n_1(1-\hat{p}_1) \geq 10;$ $n_2\hat{p}_2 \geq 10; n_2(1-\hat{p}_2) \geq 10$	2-sample exact test (Fisher’s exact test) not in textbooks $\rightarrow$ Session 8
	“plus four” CI	$n_1 \geq 5; n_2 \geq 5$	
$H_0 : p_1 = p_2$ (combined $\hat{p}$ )	z-test	$n_1\hat{p} \geq 5; n_1(1-\hat{p}) \geq 5;$ $n_2\hat{p} \geq 5; n_2(1-\hat{p}) \geq 5$	not in textbooks $\rightarrow$ Session 8
	exact test	no conditions	

<sup>7</sup> Same guidelines as in IPS 7e; the coverage in S is too limited to be of practical use.

## INFERENCE FOR 1 PROPORTION – DETAILS

- **Data:**  $X$  = number of “successes” in a binomial setting.
- **Model:**  $X \sim B(n, p)$ .
- **Estimation:**  $\hat{p} = X/n$ ,  $SE_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})/n}$ .
- **Confidence intervals** for  $p$  with confidence level  $1-\alpha$ :
  - \* **classical approx.:**  $\hat{p} \pm z^* SE_{\hat{p}}$ ,  $z^* = z_{1-\alpha/2}$ ,
  - \* **plus four approx.:**  $\tilde{p} \pm z^* SE_{\tilde{p}}$  (same formula, but add 2 successes and 2 failures<sup>8</sup>),
  - \* **“exact”**<sup>9</sup>: based on binomial distribution (software only, usually the default),

Recommendations — **“exact”**: always conservative (too wide); **classical**: may be very poor (too narrow) in small samples; **plus four**: generally good approximation,
- **Test** of  $H_0: p = p_0$  (where  $p_0$  is a known value), against one-/two-sided alternative  $H_a$ ,
  - \* **classical**, approximate  $z$ -test:  $z = (\hat{p} - p_0) / \sqrt{p_0(1-p_0)/n} \approx N(0,1)$  under  $H_0$ ,
  - \* **exact**: based on binomial distribution, e.g.,
    - $H_a: p > p_0$ :  $P = P(X \geq X_{\text{obs}})$ ,
    - $H_a: p \neq p_0$ :  $P = 2 \min\{P(X \geq X_{\text{obs}}), P(X \leq X_{\text{obs}})\}$ ,<sup>10</sup>

Recommendations: **exact** generally preferable, but **classical** almost same in large samples.

<sup>8</sup> This surprising method is usually referred to Agresti & Coull (1998); different versions exist (e.g., “Wilson”).

<sup>9</sup> Different methods exist, the classical Clopper-Pearson CI/method is known to be quite conservative in small samples and when  $p$  is near 0 or 1, but improvements exist; Minitab version 22 uses a “Blaker adjustment”.

<sup>10</sup> Other formulae exist, but this is the simplest one; see Supplementary Exercise 8.85 (solution).

## 1 PROPORTION: APHID DROPS

Aphid landings on their feet or back:<sup>11</sup>

- **Data:** 19 out of 20 aphids dropped at height 20 cm landed on their feet,
- **Model:** binomial setting  $\sim B(20, p)$ , where  $p$  = probability of feet landing,
- **Estimation:**  $\hat{p} = 19/20 = 0.95$ ,  $SE_{\hat{p}} = \sqrt{\frac{0.95 \cdot 0.05}{20}} = 0.049$ ,
- **95% CI for  $p$ :**
  - \* **classical:**  $0.95 \pm 1.96 \times 0.049 = (0.854, 1.046)$ ,
  - \* **plus four:**  $0.875 \pm 1.96 \times 0.068 = (0.743, 1.007)$ ,<sup>12</sup>
  - \* **“exact”** (Minitab 21/Stata): (0.751, 0.999); Minitab 22: (0.761, 0.997),
- **Test** of  $H_0: p = 0.5$  against  $H_a: p > 0.5$ :
  - \* **classical:**  $z = (0.95 - 0.5) / \sqrt{\frac{0.5(1-0.5)}{20}} = 4.025$ , and  $P = P(Z > 4.025) = 0.000028$ ,
  - \* **exact** using  $B(20, 0.5)$  and software/formula:
$$P = P(X \geq 19) = P(X = 19) + P(X = 20) = 0.000020,$$
- **Conclusion:** clear evidence of non-random landings,
  - \* “plus four” and exact CIs similar and preferable (“classical” CI awful),
  - \* exact test preferable (but  $z$ -test not too far off).

<sup>11</sup> Pea aphids drops were videotaped after release; Ribak et al. (2013), *Current Biology* 23, R102-103; PSLs 4e Ex. 19.6.

<sup>12</sup>  $\tilde{p} = (19+2)/(20+4) = 0.875$ , and  $SE(\tilde{p}) = \sqrt{0.875(1-0.875)/24} = 0.068$ .

## INFERENCE FOR 2 INDEPENDENT PROPORTIONS – DETAILS

- **Data:**  $X$  and  $Y$  = number of “successes” in two independent binomial settings.
- **Model:**  $X \sim B(n_1, p_1)$  and  $Y \sim B(n_2, p_2)$ ,  $X$  and  $Y$  independent,
- **Estimation:**

$$\hat{p}_1 = X/n_1, \hat{p}_2 = Y/n_2, D = \hat{p}_1 - \hat{p}_2,$$

$$SE_D = \sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2}.$$
- **Confidence interval** for  $p_1 - p_2$  with confidence level  $1 - \alpha$ :
  - \* **classical** approx.:  $\hat{p}_1 - \hat{p}_2 \pm z^* SE_D$ ,  $z^* = z_{1-\alpha/2}$ ,
  - \* **plus four** approx.:  $\tilde{p}_1 - \tilde{p}_2 \pm z^* SE_{\tilde{D}}$  (same formula, but add 1 success and 1 failure in each sample!),

Recommendations — **classical**: may be very poor in small samples; **plus four**: generally good approximation,
- **Test** (approximate) of  $H_0: p_1 = p_2 (= p)$ , against one-/two-sided alternative  $H_a$ ,
  - \* **estimate common value  $p$** :  $\hat{p} = (X + Y)/(n_1 + n_2)$  – total number of successes divided by the total number of trials,
  - \* “pooled” standard error under  $H_0$ :  $SE_{D_p} = \sqrt{\hat{p}(1-\hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ ,
  - \* **test statistic**:  $z = (D - 0)/SE_{D_p} = (\hat{p}_1 - \hat{p}_2)/SE_{D_p}$ ,
  - \*  **$P$ -values** from  $N(0,1)$  the usual way.

## 2 INDEPENDENT PROPORTIONS: ECHINACEA FOR COMMON COLD

Development of cold in Echinacea and control groups:<sup>13</sup>

- **Data:** after exposure, 88 out of 103 persons in control group, and 44 out of 48 in a treatment group developed a cold,
- **Estimation:**

tx :	$\hat{p}_1 = 44/48 = 0.917,$	$SE(\hat{p}_1) = 0.040,$
control :	$\hat{p}_2 = 88/103 = 0.854,$	$SE(\hat{p}_2) = 0.035,$
diff :	$\hat{p}_1 - \hat{p}_2 = 0.062,$	$SE(\hat{p}_1 - \hat{p}_2) = 0.053,$
	$\tilde{p}_1 - \tilde{p}_2 = 0.052,$	$SE(\tilde{p}_1 - \tilde{p}_2) = 0.055,$ <sup>14</sup>
- **95% CI for  $p_1 - p_2$ :**
  - \* **classical:**  $0.062 \pm 1.96 \times 0.053 = (-0.041, 0.166),$
  - \* **plus four:**  $0.052 \pm 1.96 \times 0.055 = (-0.056, 0.160),$
- **Test of  $H_0: p_1 = p_2$  against  $H_a: p_1 \neq p_2$ :**
  - \* **classical:**  $z = (\hat{p}_1 - \hat{p}_2)/SE_{D_p} = 0.062/0.058 = 1.075,$ <sup>15</sup> and  $P = 2 \times P(Z > 1.075) = 0.282,$
  - \* **alternative methods** → Session 8.

**Conclusions:**

- only little difference between confidence intervals despite violation of guideline for classical method,
- $P$ -value so large that we can be confident there is no evidence against  $H_0$  (despite only narrowly meeting the guideline); observed effect is in the **opposite direction** of what one might have hoped.

<sup>13</sup> Experimental study on efficacy of Echinacea product against common cold; Turner et al. (2005), *New England Journal of Medicine* 353, 341-348.; PSLs 4e Exercise 20.3.

<sup>14</sup> Calculations:  $SE(\tilde{p}_1 - \tilde{p}_2) = \sqrt{(0.90 \cdot (1-0.90)/50 + 0.8476 \cdot (1-0.8476)/105)} = 0.055$ , where we used the values  $\tilde{p}_1 = (44+1)/(48+2) = 0.90$ , and  $\tilde{p}_2 = (88+1)/(103+2) = 0.8476$ .

<sup>15</sup> Calculations:  $SE_{D_p} = \sqrt{0.874(1-0.874)(1/48 + 1/103)} = 0.058$ , where pooled  $\hat{p} = (44+88)/(48+103) = 0.874$ .

## NONPARAMETRIC (DISTRIBUTION-FREE) METHODS

- no parametric statistical model (involving particular distribution type),
- still assumptions of **i.i.d. samples** and possibly of particular features of distributions,
- **classical methods** — the only ones in the course syllabus!
  - \* mostly based on **ranks**, that is, the **relative magnitude of observations**, where it does not matter how much  $X_2 > X_1$ , only that  $X_2 > X_1$ ,
  - \* analyses computable by hand (tedious for large data), but reference distributions require special tables or large-sample approximations,
  - \* all methods in the course **available in Minitab/Stata/R**,
  - \* **advantages**: no distribution assumptions, robust, “simple to use”...
  - \* **disadvantages**:  
some loss of information compared to good parametric model, problems with getting good estimates and confidence intervals (**what to estimate?**), not available beyond the very simplest designs,
- alternative: **modern, computer-intensive methods**:
  - \* **resampling/permutation/bootstrap** methods,<sup>16</sup>
  - \* very flexible and powerful, but **not** so simple to use.<sup>17</sup>

---

<sup>16</sup> Recommended by PSL/IPS; described in IPS Supplementary Chapter 16.

<sup>17</sup> Minitab 19-21 includes features for 1 and 2 samples under Calc-Resampling, rather than in the Stat menu.

## 1 SAMPLE: SIGN TEST

**Example** (Visual receptive fields, 6L–11):

- neural activity at 9 recordings of both Spontaneous activity (SA) and Response (R),
- analyze differences (R–SA) (ordered): -7.5 -2.5 12.5 13.3 14.2 16.7 26.7 34.2 44.2.

**Sign test** for **null hypothesis**  $H_0$ : median = known value ( $m_0$ ):

- **Model**:  $X_1, \dots, X_n$  i.i.d. from continuous distribution,
- **Test procedure**:
  - \* test statistic:  $Y$  = number of  $X$ 's  $> m_0$ ,
  - \* disregard  $X$ 's =  $m_0$ , let  $n_1$  = number of  $X$ 's  $\neq m_0$ ,
  - \* **under  $H_0$** :  $Y \sim B(n_1, 0.5) \Rightarrow$  corresponds to **testing  $H_0: p=0.5$**  in the binomial distribution  $B(n_1, p)$  for  $Y$ ,
  - \*  **$P$ -values** from binomial distribution, e.g.  $P = P(Y \geq Y_{\text{obs}})$  for  $H_a$ : median  $> m_0$ ,<sup>18</sup>
- **Confidence interval** for the median: in Minitab/Stata.

**Example** — interest is in testing  $H_0$ : median = 0 vs.  $H_a$ : median  $> 0$ :

- no differences = 0; out of 9 differences, 7 are  $> 0$ ,
- $P = P(Y \geq 7) = 0.090$ , where  $Y \sim B(9, 0.5)$ ,
- cannot reject  $H_0$ : no evidence of higher activity for R by the test.

<sup>18</sup> Also the alternative hypotheses have equivalent binomial formulations, e.g. here  $H_a$ :  $p > 0.5$ .

## McNEMAR'S TEST

= sign test for paired binary data<sup>19</sup> (or paired proportions); **note**: not in course curriculum.

**Example:** Varicose veins and overweight:

- 122 pairs of brothers, one overweight and one normal weight, with records of presence or absence of varicose veins,

Group	var. veins	
	+ (1)	- (0)
normal wt	23	99
overwt	30	92

Normal weight	Overweight	
	+ var. veins (1)	- var. veins (0)
+ var. veins (1)	19	4
- var. veins (0)	11	88

- hypothesis of interest: same proportion of varicose veins among normal weight and overweight persons? — observed proportions:

$$\text{normal weight: } \hat{p} = 23/122 = 0.19,$$

$$\text{overweight: } \hat{p} = 30/122 = 0.25.$$

**Test procedure:**

- code each “success” as 1, and each “failure” as 0,
- compute differences  $D_i$  (e.g. normal weight – overweight) within each pair  $i$ :
  - \*  $D_i = 0$ : same outcome (either 1 or 0) in both pair members,
  - \*  $D_i = 1$ : success in first pair member, failure in second,
  - \*  $D_i = -1$ : failure in first pair member, success in second,
- disregard all  $D_i = 0$ ; let  $n_1 = \# (D_i = 1 \text{ or } D_i = -1)$ ; assume  $Y = \# (D_i = 1) \sim B(n_1, p)$ ; and test  $H_0 : p = 0.5$  against  $H_a : p \neq 0.5$ ,
- example:**  $Y_{\text{obs}} = 4$ ;  $Y \sim B(15, p)$ ; and  $P = 2 \times P(Y \leq 4) \approx 2 \cdot (0 + 0 + 0.003 + 0.014 + 0.042) = 0.12$  (binomial table); **conclusion:** no statistical evidence against  $H_0$ .

<sup>19</sup> Different versions of McNemar's test exist; the one described here gives an exact  $P$ -value based on the binomial distribution, and is generally recommended.

## RANKS

Values/numbers  $x_1, \dots, x_n$ .

- **order values** by increasing magnitude:

$$x_{(1)} \leq x_{(2)} \leq \dots x_{(n)}, \quad \text{where } \text{rank}(x_{(i)}) = i$$

i.e., rank =  $i$  when value is the  $i$ th smallest among all values,

- **ties** (several values equal): use average rank among all tied values,
- it is sometimes possible to assign ranks, even if data only partially observed (**left-/right-censored**: smaller/greater than or equal to a cut-off).

**Example** (constructed data):

data	2.2	3.1	1.9	2.2	2.0	5.0
ordered data	1.9	2.0	2.2	2.2	3.1	5.0
ranks	1	2	3.5 <sup>a</sup>	3.5 <sup>a</sup>	5	6

<sup>a</sup> average rank computed as:  $3.5 = (3 + 4)/2$

- the value 5.0 is much larger than the others but that is not reflected (strongly) in the ranks,
- if an additional observation was partially observed and only known to be  $> 5$  (i.e., right-censored at 5), then its rank would be 7,
- sum of ranks = 21 (generally, among  $n$  values the sum of ranks equals  $n(n+1)/2$ ).

## 2 SAMPLES: WILCOXON–MANN–WHITNEY TEST

**Wilcoxon rank sum test** (PSLS/IPS terminology, also often Mann–Whitney test):

- **Model:**  $X_1, \dots, X_{n_1}$  and  $Y_1, \dots, Y_{n_2}$  independent and i.i.d. samples from distributions  $\text{Dist}_X$  and  $\text{Dist}_Y$ , respectively,
- **Hypotheses** — two possible setups:
  - (1)  $H_0: \text{Dist}_X = \text{Dist}_Y$  (same distribution),  $H_a: \text{Dist}_X \neq \text{Dist}_Y$ ,<sup>20</sup>
  - (2) assuming “ $\text{Dist}_X = \text{Dist}_Y + \Delta$ ” (distributions differ only in position):  $H_0: \Delta = 0$  (corresponding to  $\text{median}_X = \text{median}_Y$ ) vs. one- or two-sided alternatives  $H_a$ ,
- **Test procedure:**
  - \* **rank all observations** as if a single sample,
  - \* **test statistic:**  $W$  = sum of ranks for  $X$ -sample,
  - \* **under  $H_0$ :** distribution of  $W$  has **no easy form**
    - **tabulated** in special tables for small  $n_1, n_2$ , when there are **no ties**,
    - **software** may give exact values, or use different types of **approximations** (in Minitab/Stata/R), with improving accuracy for increasing sample size,
- **Confidence interval** for  $\text{median}_X - \text{median}_Y$  (**valid under  $\Delta$ -assumption**) → software,
- recommended to check for similar spread and skewness in the two **distributions of ranks**.<sup>21</sup>

<sup>20</sup> More specific wording of  $H_a$ :  $\text{Dist}_X$  is systematically larger than  $\text{Dist}_Y$ , or vice versa (for a two-sided  $H_a$ ); see Chapter 27 of PSLS.

<sup>21</sup> Fagerland & Sandvik (2009), *Statistics in Medicine* 28, 1487-1497.

EXAMPLE FOR 2-SAMPLE W-M-W TEST

Parasite burdens of calves in Lithuania:

pasture	Data values									
safe	0	8	8	10	26	34	38	44	46	
infected	20	30	30	36	50	52	54	70	70	100
both	0	8	8	10	20	26	30	30	34	36
samples	38	44	46	50	52	54	70	70	100	
	Ranks									
blue ~	1	2.5	2.5	4	5	6	7.5	7.5	9	10
safe	11	12	13	14	15	16	17.5	17.5	19	

Nonparametric analysis:

- **Model:** two independent samples, assume also that distributions differ only in position ( $\Delta$ -assumption),
- **test statistic:**  $W$  = sum of ranks in safe sample = **61** (or 129 for infected sample),
- approximate **P-value** = 0.020 (Minitab/R) or 0.018 (Stata),
- **95% CI** for median difference (infected–safe): (6.0, 46.0).

Normal distribution analysis (from Lecture 6):

- **Estimation:**  $\hat{\mu}_1 = 51.2$ ,  $\hat{\mu}_2 = 23.8$ ,  $SE(\hat{\mu}_1 - \hat{\mu}_2) = \sqrt{s_1^2/10 + s_2^2/9} = 9.59$ ,
- **test statistic:**  $t = (\hat{\mu}_1 - \hat{\mu}_2)/SE(\hat{\mu}_1 - \hat{\mu}_2) = 2.86$ ; **P-value** = 0.011, from  $t(16)$ ,
- **95% CI** for mean difference (infected–safe): (7.1, 47.8).

## 1 SAMPLE: WILCOXON'S SIGNED RANK TEST

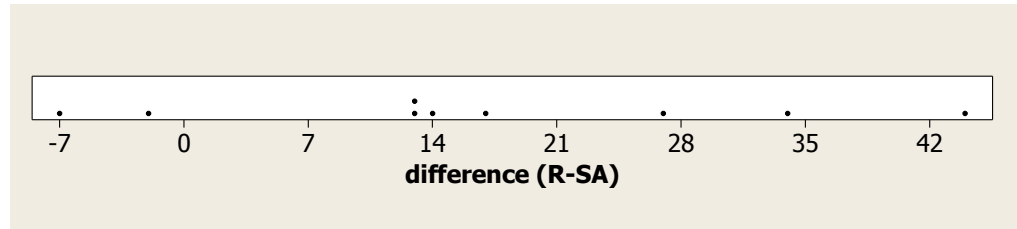
Wilcoxon's test for null hypothesis  $H_0$ : median = known value ( $m_0$ ):

- **Model:**  $X_1, \dots, X_n$  i.i.d. sample from a continuous, **symmetric** distribution,  
— **note:** the assumed symmetry is an **extra assumption** compared to the sign test,
- **Alternative hypotheses**  $H_a$ : either one- or two-sided,
- **Test procedure:**
  - \* let  $R_i = X_i - m_0$ , and disregard observations with  $R_i = 0$
  - \* rank the  $|R_i|$ 's, and let  $S_i = \text{rank of } |R_i|$ ,
  - \* **idea:** if, for example, true median  $> m_0$ , then there will be both **more and larger ranks** for observations  $> m_0$ ,
  - \* **test statistic:**  $W^+ = \text{sum of } S_i\text{'s for positive observations (i.e., } R_i > 0, \text{ corresponding to } X_i > m_0)$ ,
  - \* **under  $H_0$ :** distribution of  $W^+$  has **no easy form**
    - **tabulated** in special tables for small  $n$ , when there are **no ties**,
    - **software** may give exact values (e.g., R), or use different types of **approximations** (Minitab/Stata), with improving accuracy for increasing sample size,
- **Confidence interval** for the median: in Minitab/R.

## EXAMPLE FOR 1-SAMPLE WILCOXON TEST

Visual receptive fields (6L–11):

- dotplot for differences (R–SA):



- data values ( $X_i$ ) and ranks ( $S_i$ ; blue  $\sim X_i > 0$ ):

$X_i$	-7.5	-2.5	12.5	13.3	14.2	16.7	26.7	34.2	44.2
$ R_i $	7.5	2.5	12.5	13.3	14.2	16.7	26.7	34.2	44.2
$S_i$	2	1	3	4	5	6	7	8	9

- assume the distribution (of differences) to be symmetric about its median,
- $H_0$ : median = 0 vs.  $H_a$ : median > 0,
- $W^+ = 3 + 4 + 5 + 6 + 7 + 8 + 9 = 42$ ,  $W^- = 3$ ,
- **P-value**: 0.012 (Minitab) / 0.021 (Stata) / 0.010 (R),
- **95% CI for median** (Minitab/R): (3.3, 30.5),
- Wilcoxon test is significant and preferable here (assumed symmetry seems ok).

**Summary** — comparison Wilcoxon vs. sign test:

Wilcoxon test is stronger (in fact, the sign test is quite weak), but additionally assumes the distribution to be symmetric.

## NOTES ON HOME ASSIGNMENT I

- **Overall impression** — good, with some variation in detail and effort,
- **Formatting** — generally nice; it is okay to include the text of the assignment but it is obviously your own answers that count (so beware that you don't make them too brief),
- **Coverage** — make sure you answer all the specific questions in the text (which are not displayed as a list),
- **Methods** — generally good, but beware that any methods or recommendations not covered in the course material need to be explained and referenced,
- **Terminology:**
  - \* we never “accept  $H_0$ ” — statistical tests give evidence against  $H_0$ , not in favour,
- Some **problems/mistakes seen repeatedly:**
  - \* *suspected* outliers (obtained by the  $1.5 \times \text{IQR}$  rule and visible in the boxplot) are not automatically *true* outliers: you have to use your judgement about whether these data points belong with the others (1L–6),
  - \* a normality test can aid in the assessment of whether a sample looks normal, but it should not be interpreted rigidly at a significance level of 5% (4L–8) (as with any other statistical test, but maybe more importantly here).

**Bottom line** — a good start (and lots of good marks).

## MID-TERM EXAM PRACTICAL INFO

- **mark:** **optional** for 15% of course mark; that is, you decide **after you have received your mark** if you want to use it or not (**if you use it**, your final exam will be shorter),
- **all aids (books and notes) are allowed**, including calculators, but only limited use of computers and computing/communication devices,<sup>22</sup>
- **duration:** 1 hour, without extensions!

**One question/assignment** with possible types of (sub)questions (possibly involving multiple choice):

- choice of statistical model and analysis —
  - \* carry out analysis when calculations manageable (see below),
  - \* or base analysis on Minitab print + extra calculations,
  - \* or outline analysis if calculations not manageable,
- probability calculations manageable (see below).

**Calculations manageable by hand (calculator):**

- simple probabilities (e.g.,  $1-p$ ,  $(1-p)^n$ , simple binomial),
- probabilities in normal distribution (incl. standardization to  $N(0, 1)$ ),
- $z/t$ -tests and CIs (given estimates or calculated statistics),
- one-sample proportion (**note:** two-sample proportions not included),
- no data entry into calculator or large summations.

---

<sup>22</sup> The details are written out in the Instructions for home assignments and exams, at the homepage.

## SUMMARY NOTES

### Key words and concepts:

- proportion data — modelled by binomial distributions,
  - \* always **estimate** by sample proportions,
  - \* **CI methods for proportions**:
    - classical (based on  $z$ -distribution), “plus four”, “exact” (1-sample only),
    - choice between methods, based on  $n$  and  $\hat{p}$ ,
  - \* **test methods for proportions**:
    - classical (based on  $z$ -distribution), exact (based on binomial or other distributions),
    - choice between methods, based on  $n$  and  $\hat{p}$ ,
- **nonparametric tests**:
  - \* characterized by no distribution (normality) assumptions,
  - \* often focusing on **median**(s) instead of mean(s),
  - \* many methods exist — the VHM 801 course covers:
    - **sign test** for 1-sample → test of  $H_0 : p = 0.5$  in  $B(n, p)$ ,
    - **rank-based tests** for 1 sample (Wilcoxon signed rank) and 2 independent samples (Mann-Whitney): all calculations by software, beware of assumptions for the test/model about distribution shape.