

## Bonus Assignment 3 – Sample Comparison

Qixun Qu

901001-5551

qixun@student.chalmers.se

### Problem A

The data in this problem is in pairs, the difference between each pair (data of bottle-fed babies subtracts data of breast-fed babies) is computed first as shown in Table 1. Plot differences versus both durations of breast-fed babies and durations of bottle-fed babies respectively in Figure 1 and Figure 2. From Figure 1, it is not obvious that differences are related to the durations of breast-fed infants. In contrast, in Figure 2, there is a tendency that difference becomes larger as the increasing of durations of bottle-fed babies. Therefore, the hypothesis that breast-fed babies tend to have less prolonged effusions than bottle-fed babies is probably true.

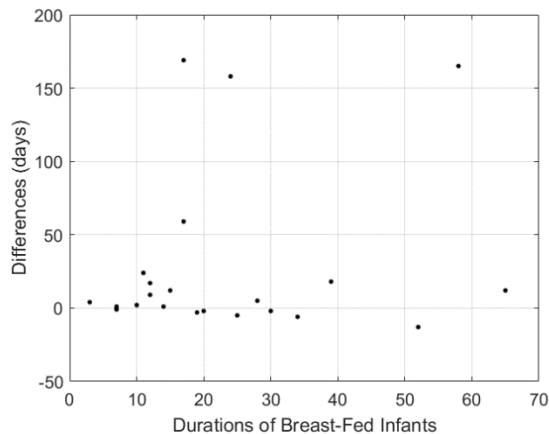


Figure 1. Difference vs. Durations  
of Breast-Fed Infants

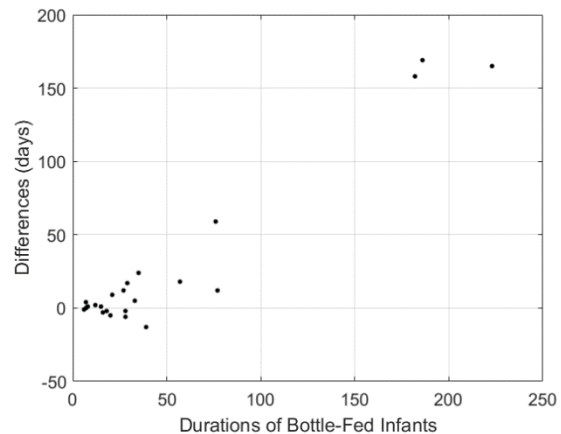


Figure 2. Difference vs. Durations  
of Bottle-Fed Infants

### Problem B

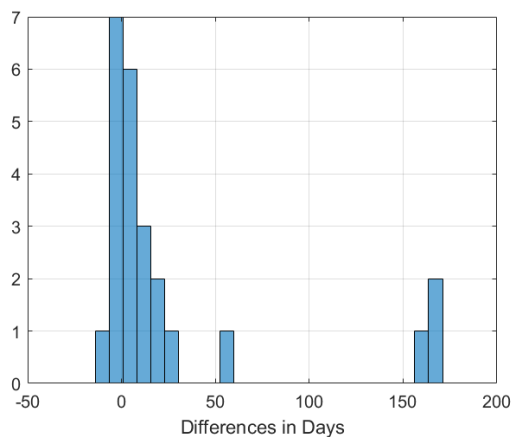


Figure 3. Distribution of Differences

The distribution of all differences are shown in Figure 3. It can be seen that it is not normal distribution due to some high values. Thus, it is more appropriate to use a nonparametric test. Since the data is in pairs, signed rank test will be carried out in next step.

The null hypothesis  $H_0$  is that the median of differences is 0, in other words, there is no difference between durations of breast-fed babies and durations of bottle-fed babies. First, calculate the signed rank for each pair, which is shown in Table 1. The test statistics  $W_+$  is 215. Under  $H_0$ , its mean and variance are computed as Equation 2 and 3. In this case,  $n$  is equal to 23, since there are 23 non-zero differences.  $n$  is larger than 20, the normalized test statistics should be used as Equation 4. After which, the two-sided P-value can be obtained as **0.0192**, indicating that the null hypothesis is ought to be rejected.

$$W_+ = \sum rank(D_i) = 215, \forall D_i > 0 \quad \text{Eq. 1} \quad E(W_+) = \frac{n(n+1)}{4} = 138 \quad \text{Eq. 2}$$

$$Var(W_+) = \frac{n(n+1)(2n+1)}{24} = 1081 \quad \text{Eq. 3} \quad Z = \frac{W_+ - E(W_+)}{\sqrt{Var(W_+)}} = 2.3420 \quad \text{Eq. 4}$$

Breast-Fed	Bottle-Fed	Difference	Absolute Value	Sign	Rank
20	18	-2	2	-1	6
11	35	24	24	1	19
3	7	4	4	1	9
24	182	158	158	1	21
7	6	-1	1	-1	2.5
28	33	5	5	1	10.5
58	223	165	165	1	22
7	7	0	0	0	0
39	57	18	18	1	18
17	76	59	59	1	20
17	186	169	169	1	23
12	29	17	17	1	17
52	39	-13	13	-1	16
14	15	1	1	1	2.5
12	21	9	9	1	13
30	28	-2	2	-1	6
7	8	1	1	1	2.5
15	27	12	12	1	14.5
65	77	12	12	1	14.5
10	12	2	2	1	6
7	8	1	1	1	2.5
19	16	-3	3	-1	8
34	28	-6	6	-1	12
25	20	-5	5	-1	10.5

Table 1. The Sign and Rank of Differences between Two Samples

2017/03/03