

Assignment 3 MVE155 Statistical inference

Yankun Xu (940630-0237)
yankun@student.chalmers.se

1. Question a

We have 24 pairs of infants, there two observations, breast-fed and bottle-fed, in each pair. What we need to do is to compare these two groups.

First of all, we could use boxplot to examine data. As figure1 shown, the right data (bottle-fed) is more skewed, and its median is slightly larger than left. In addition, both data have outliers, but right has three much larger outliers. Therefore, we could believe that durations of bottle-fed are usually longer than breast-fed.

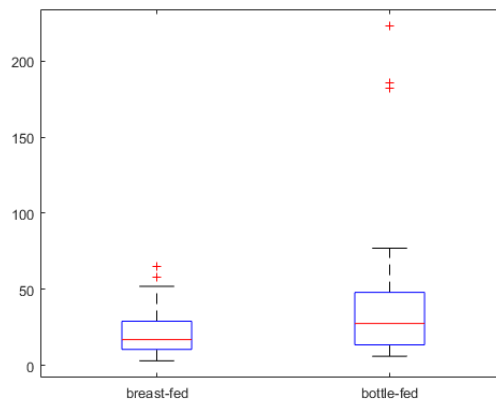


Figure 1. Boxplot of two groups

Then, we could examine normal quantiles plot (thicker plot is bottle-fed data, thinner plot is breast-fed data). Figure2 also shows the skewness of the distribution, t-test which assumes an approximate normal distribution test is not suitable for this case.

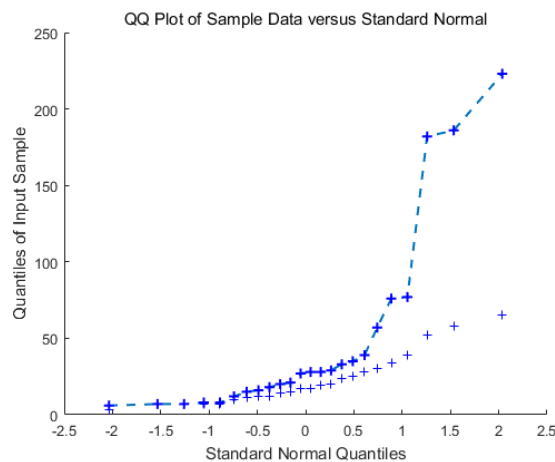


Figure 2. Normal probability plots

Then, we can use a plot of breast-fed group vs bottle-fed group. Figure3 indicates that bottle-fed group becomes larger along with breast-fed group being larger, these two groups are correlated, so a paired test is reasonable.

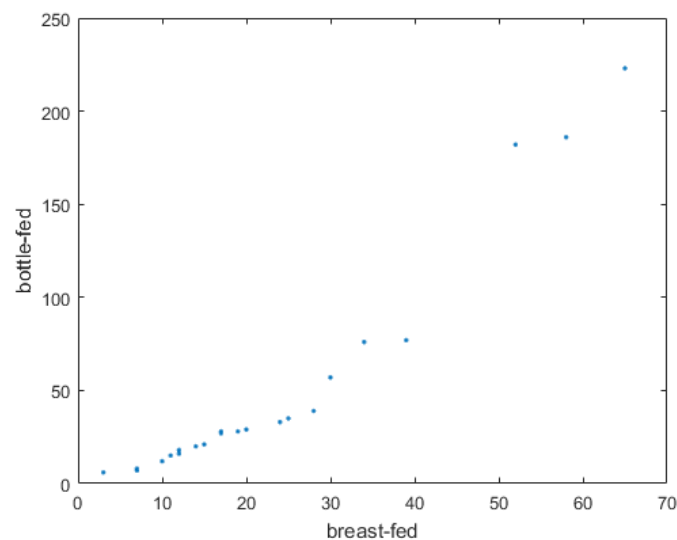


Figure 3. Plot of breast-fed group vs bottle-fed group

2. Question b

As analysis indicates in question a, a non-parametric test method is better under the condition of data doesn't fit normal distribution well. In non-parametric test method, rank sum test which is used for independent case and signed rank test which is used for paired case are two popular methods to test data. Based on figure3 and its analysis, paired non-parametric test – signed rank test should be used in this case.

The null hypothesis H_0 : distribution of D (differences of two groups: bottle-fed minus breast-fed) is symmetric about 0.

Test statistics: $W_+ = \sum \text{rank}(|D_i|) \cdot I(D_i > 0)$ or $W_- = \sum \text{rank}(|D_i|) \cdot I(D_i < 0)$.

For $n \geq 20$, use the normal approximation of null distribution with $\mu_W = \frac{n(n+1)}{4}$

and $\sigma_W^2 = \frac{n(n+1)(2n+1)}{24}$.

In Matlab, we could use command `signrank` directly to calculate the 2-sided P-value and test conclusion with 5% significant level. The result is that 2-sided P-value is 1.91% and reject the null hypothesis in favor of $\alpha = 5\%$. Actually, I also calculate the P-value by hand because I don't know how to rank the difference following the rules of signed rank test, I get $W_+ = 215$ and combine $n=23$ to calculate $Z = 2.34$. two-sided P-value is 1.92%

Therefore, two groups are different.