201702896 이하연

Rx

n=42 new treatment therapy 21 0
standrad treatment therapy 21 1

1. The following data set called "anderson.dat" consists of remission survival times on 42 leukemia patients, half of whom get a certain new treatment therapy and the other half of whom get a standard treatment therapy. The exposure variable of interest is treatment status ($Rx = 0$ if new treatment, $Rx = 1$ if standard treatment). Two other variables for control as potential confounders are log white blood cell count (i.e., logwbc) and sex. Failure status is defined by the relapse variable (0 if censored, 1 if failure). The data set is listed as follows:

Censored 0/failure 1

| Subj | Survt | Relapse | Sex | log WBC | $Rx$ | |
|---|---|---|---|---|---|---|
| 1 | 35 | 0 | 1 | 1.45 | 0 | low |
| 2 | 34 | 0 | 1 | 1.47 | 0 | low |
| 3 | 32 | 0 | 1 | 2.20 | 0 | low |
| 4 | 32 | 0 | 1 | 2.53 | 0 | medium |
| 5 | 25 | 0 | 1 | 1.78 | 0 | low |
| 6 | 23 | 1 | 1 | 2.57 | 0 | medium |
| 7 | 22 | 1 | 1 | 2.32 | 0 | medium |
| 8 | 20 | 0 | 1 | 2.01 | 0 | low |
| 9 | 19 | 0 | 0 | 2.05 | 0 | low |
| 10 | 17 | 0 | 0 | 2.16 | 0 | low |
| 11 | 16 | 1 | 1 | 3.60 | 0 | high |
| 12 | 13 | 1 | 0 | 2.88 | 0 | medium |
| 13 | 11 | 0 | 0 | 2.60 | 0 | medium |
| 14 | 10 | 0 | 0 | 2.70 | 0 | medium |
| 15 | 10 | 1 | 0 | 2.96 | 0 | medium |
| 16 | 9 | 0 | 0 | 2.80 | 0 | medium |
| 17 | 7 | 1 | 0 | 4.43 | 0 | high |
| 18 | 6 | 0 | 0 | 3.20 | 0 | high |
| 19 | 6 | 1 | 0 | 2.31 | 0 | medium |
| 20 | 6 | 1 | 1 | 4.06 | 0 | high |
| 21 | 6 | 1 | 0 | 3.28 | 0 | high |
| 22 | 23 | 1 | 1 | 1.97 | 1 | low |
| 23 | 22 | 1 | 0 | 2.73 | 1 | medium |
| 24 | 17 | 1 | 0 | 2.95 | 1 | medium |
| 25 | 15 | 1 | 0 | 2.30 | 1 | low |
| 26 | 12 | 1 | 0 | 1.50 | 1 | low |
| 27 | 12 | 1 | 0 | 3.06 | 1 | high |
| 28 | 11 | 1 | 0 | 3.49 | 1 | high |
| 29 | 11 | 1 | 0 | 2.12 | 1 | low |
| 30 | 8 | 1 | 0 | 3.52 | 1 | high |
| 31 | 8 | 1 | 0 | 3.05 | 1 | high |
| 32 | 8 | 1 | 0 | 2.32 | 1 | medium |
| 33 | 8 | 1 | 1 | 3.26 | 1 | high |
| 34 | 5 | 1 | 1 | 3.49 | 1 | high |
| 35 | 5 | 1 | 0 | 3.97 | 1 | high |
| 36 | 4 | 1 | 1 | 4.36 | 1 | high |
| 37 | 4 | 1 | 1 | 2.42 | 1 | medium |
| 38 | 3 | 1 | 1 | 4.01 | 1 | high |
| 39 | 2 | 1 | 1 | 4.91 | 1 | high |
| 40 | 2 | 1 | 1 | 4.48 | 1 | high |
| 41 | 1 | 1 | 1 | 2.80 | 1 | medium |
| 42 | 1 | 1 | 1 | 5.00 | 1 | high |

a. Suppose we wish to describe KM curves for the variable logwbc. Because logwbc is continuous, we need to categorize this variable before we compute KM curves. Suppose we categorize logwbc into three categories—low, medium, and high—as follows:

low (0–2.30), $n = 11$;
medium (2.31–3.00), $n = 14$;
high (>3.00), $n = 17$.

Based on this categorization, compute and graph KM curves for each of the three categories of logwbc. ~~(You may use a computer program to assist you or you can form three tables of ordered failure times and compute KM probabilities directly.)~~

p.3

b. Compare the three KM plots you obtained in part a. How are they different?

p.4

c. Below is an edited printout of the log–rank test comparing the three groups.

| Group | Events observed | Events expected |
|---|---|---|
| 1 | 4 | 13.06 |
| 2 | 10 | 10.72 |
| 3 | 16 | 6.21 |
| Total | 30 | 30.00 |

Log–rank = chi2(2) = 26.39
P-value = Pr > chi2 = 0.0000

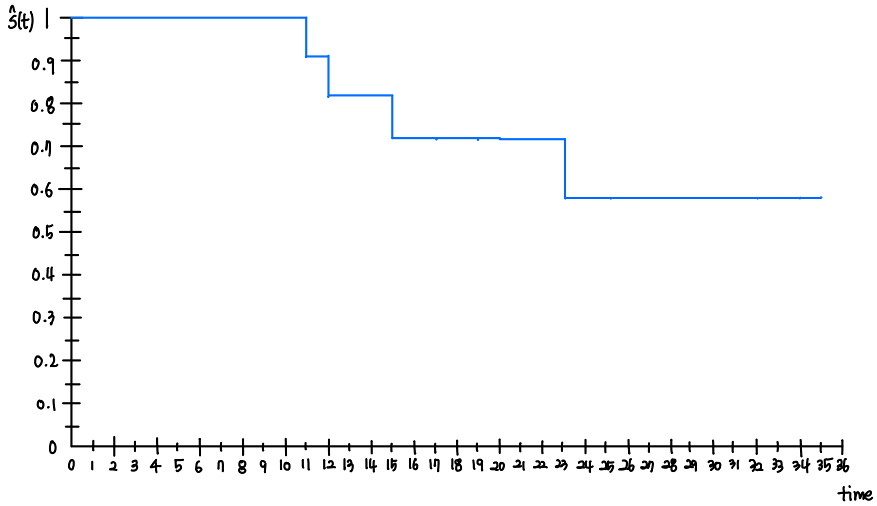
What do you conclude about whether or not the three survival curves are the same?

Ho reject, three survival curves의 difference가 존재

1 - a.

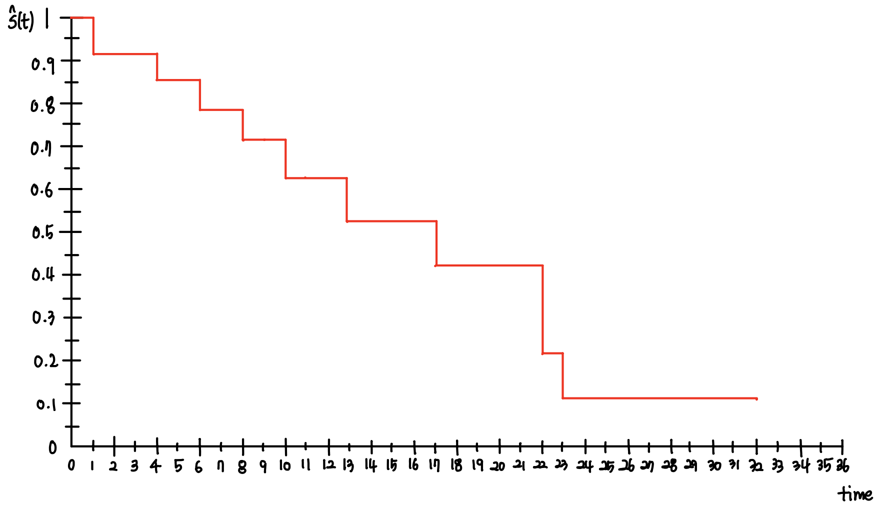
Group 1 (log WBC - low)

| $t_{(j)}$ | $n_j$ | $m_i$ | $q_i$ | $\hat{S}(t_{(j)})$ |
|---|---|---|---|---|
| 0 | 11 | 0 | 0 | 1 |
| 11 | 11 | 1 | 0 | $1 \times \frac{10}{11} = 0.9091$ |
| 12 | 10 | 1 | 0 | $.9091 \times \frac{9}{10} = 0.8182$ |
| 15 | 9 | 1 | 0 | $.8182 \times \frac{8}{9} = 0.7273$ |
| 17 | 8 | 0 | 1 | $.7273 \times \frac{8}{8} = 0.7273$ |
| 19 | 7 | 0 | 1 | $.7273 \times \frac{7}{7} = 0.7273$ |
| 20 | 6 | 0 | 1 | $.7273 \times \frac{6}{6} = 0.7273$ |
| 23 | 5 | 1 | 0 | $.7273 \times \frac{4}{5} = 0.5818$ |
| 25 | 4 | 0 | 1 | $.5818 \times \frac{4}{4} = 0.5818$ |
| 32 | 3 | 0 | 1 | $.5818 \times \frac{3}{3} = 0.5818$ |
| 34 | 2 | 0 | 1 | $.5818 \times \frac{2}{2} = 0.5818$ |
| 35 | 1 | 0 | 1 | $.5818 \times \frac{1}{1} = 0.5818$ |



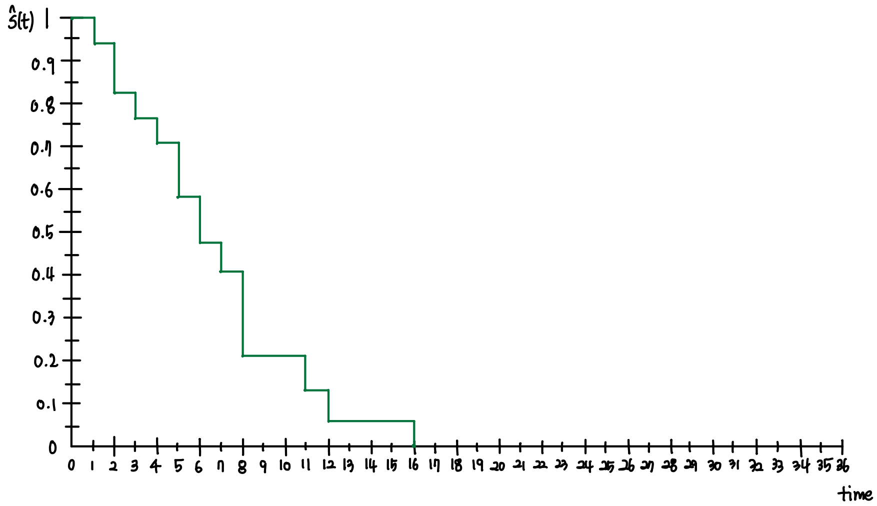

Group 2 (log WBC - medium)

| $t_{(j)}$ | $n_j$ | $m_i$ | $q_i$ | $\hat{S}(t_{(j)})$ |
|---|---|---|---|---|
| 0 | 14 | 0 | 0 | 1 |
| 1 | 14 | 1 | 0 | $1 \times \frac{13}{14} = 0.9286$ |
| 4 | 13 | 1 | 0 | $.9286 \times \frac{12}{13} = 0.8571$ |
| 6 | 12 | 1 | 0 | $.8571 \times \frac{11}{12} = 0.7857$ |
| 8 | 11 | 1 | 0 | $.7857 \times \frac{10}{11} = 0.7143$ |
| 9 | 10 | 0 | 1 | $.7143 \times \frac{10}{10} = 0.7143$ |
| 10 | 9 | 1 | 1 | $.7143 \times \frac{8}{9} = 0.6349$ |
| 11 | 7 | 0 | 1 | $.6349 \times \frac{7}{7} = 0.6349$ |
| 13 | 6 | 1 | 0 | $.6349 \times \frac{5}{6} = 0.5291$ |
| 17 | 5 | 1 | 0 | $.5291 \times \frac{4}{5} = 0.4233$ |
| 22 | 4 | 2 | 0 | $.4233 \times \frac{2}{4} = 0.2116$ |
| 23 | 2 | 1 | 0 | $.2116 \times \frac{1}{2} = 0.1058$ |
| 32 | 1 | 0 | 1 | $.1058 \times \frac{1}{1} = 0.1058$ |

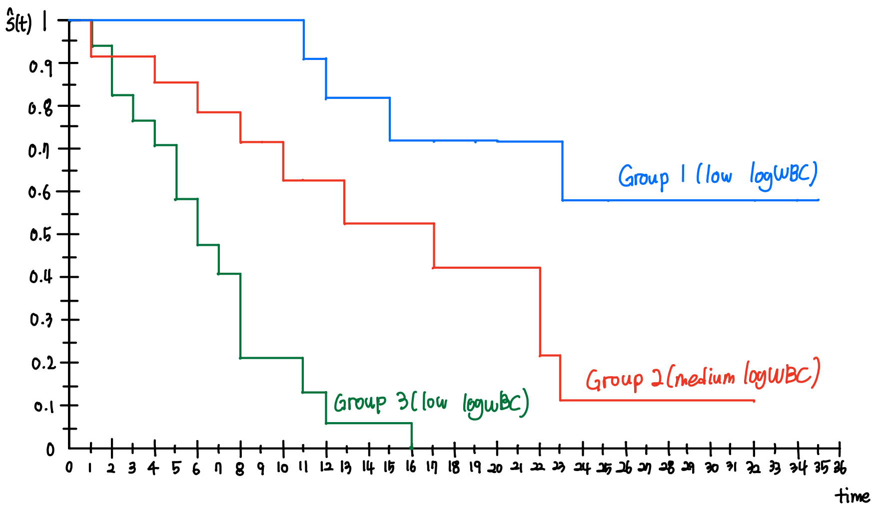



Group 2 (log WBC - high)

| $t_{(j)}$ | $n_j$ | $m_i$ | $q_i$ | $\hat{S}(t_{(j)})$ |
|---|---|---|---|---|
| 0 | 17 | 0 | 0 | 1 |
| 1 | 17 | 1 | 0 | $1 \times \frac{16}{17} = 0.9412$ |
| 2 | 16 | 2 | 0 | $.9412 \times \frac{14}{16} = 0.8235$ |
| 3 | 14 | 1 | 0 | $.8235 \times \frac{13}{14} = 0.7647$ |
| 4 | 13 | 1 | 0 | $.7647 \times \frac{12}{13} = 0.7059$ |
| 5 | 12 | 2 | 0 | $.7059 \times \frac{10}{12} = 0.5882$ |
| 6 | 10 | 2 | 1 | $.5882 \times \frac{8}{10} = 0.4706$ |
| 7 | 7 | 1 | 0 | $.4706 \times \frac{6}{7} = 0.4034$ |
| 8 | 6 | 3 | 0 | $.4034 \times \frac{3}{6} = 0.2017$ |
| 11 | 3 | 1 | 0 | $.2017 \times \frac{2}{3} = 0.1345$ |
| 12 | 2 | 1 | 0 | $.1345 \times \frac{1}{2} = 0.0672$ |
| 16 | 1 | 1 | 0 | $.0672 \times \frac{0}{1} = 0$ |



# 1 - b.




전반적인 흐름으로 보았을 때, (time 0-2 제외) Group 3, Group 2, Group 1 순으로 $\hat{S}(t)$가 작다.
따라서 log WBC가 낮을수록 생존확률이 낮다고 추정할 수 있다.
또한 시간이 갈수록 각 Group 간의 effect 차이가 큰 것도 볼 수 있다.