Week4 Report

Xiaohui Li, Yuxin Ma

In this week, there are several tasks that we have done. Firstly, we update one of the conclusion we made last week about the fitness to the $\log(K)$ and $\sqrt{K}$. Secondly, we do the isotonic regression to the power curve to make it much smoother and keep non-decreasing, specifically we increase repetition times to make our results more accurate and we try to find the accuracy efficiency trade-off optimal way to do this. Thirdly, we do the analysis about scaling when we control the ratio $r_1 = \frac{Y_1}{\sigma_1}$ and $r_k = \frac{\sigma_k}{\sigma_1}$ to be stable. Finally, we change the expression of P_combined value by taking the minimum, maximum, median and mean value of P-values.

1 Critical Ratio Analysis

We have found a new finding different from last week about the fitness to the critical ratio by $\log(K)$ and $\sqrt{K}$. The 80% power related critical ratio $r_1 = \frac{Y_1}{\sigma_1}$ actually follows a **squared root K** form, and the squared root model highly explains the increase pattern, with R-squared amounting to **0.991**. Previously we controlled for the subgroup unit number size (N/K) and increase K to get a logarithm increasing form. This time we consider controlling for the total experiment unit number N rather than N/K, which seems to make more sense. In this case after changing size K the N/K will also vary, leading to the changing Yk obs and sk's distribution. The plots and results are as follow.
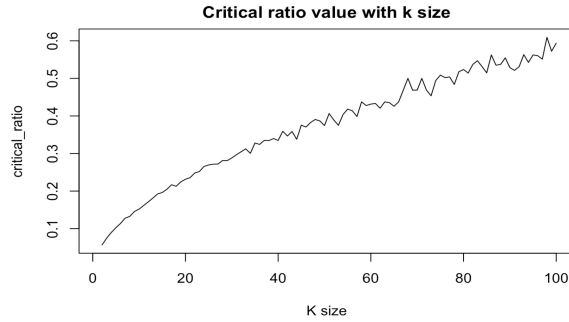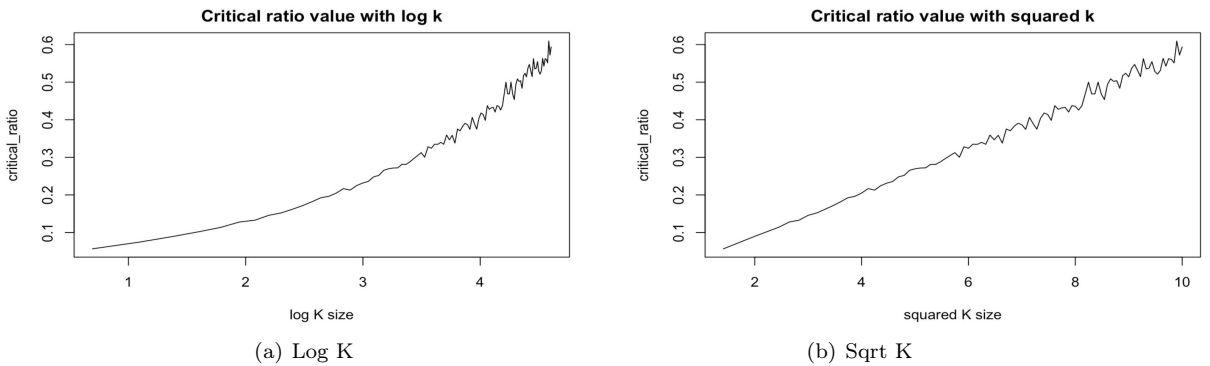

Figure 1.1: Critical Ratio for K

(a) Log K

(b) Sqrt K

Figure 1.2: Logarithm K and squared root K

We further apply the linear regression to compare $\sqrt{K}$ and $\log K$.

```
Call:
lm(formula = critical_ratio ~ logk)

Residuals:
     Min       1Q    Median       3Q      Max
-0.05399 -0.03174 -0.01112  0.02862  0.15321

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.206102   0.017507  -11.77   <2e-16 ***
logk         0.158025   0.004642   34.05   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.03936 on 97 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.9228,    Adjusted R-squared:  0.922
F-statistic:  1159 on 1 and 97 DF,  p-value: < 2.2e-16


Call:
lm(formula = critical_ratio ~ sqrtk)

Residuals:
      Min        1Q    Median        3Q       Max
-0.030777 -0.005996  0.000198  0.006513  0.042270

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.0426060  0.0042328  -10.07   <2e-16 ***
sqrtk        0.0615901  0.0005927  103.91   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01336 on 97 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.9911,    Adjusted R-squared:  0.991
F-statistic: 1.08e+04 on 1 and 97 DF,  p-value: < 2.2e-16
```

Figure 1.3: Regression Summary

Note the $\sqrt{K}$'s β coefficient is more statistically significant and larger, while the r^2 of log is also larger than $\log k$, this supports the reasoning that critical ratio is logarithm increase. In particular, the r squared for $\sqrt{K}$ model is high as 0.9911, indicating a potential theoretical relationship.

2   Isotonic Regression

To make the power non-decreasing, which will be a more reasonable result, we use isotonic regression to the power and get a new plot where the isotonic curve is represented by the red lines. For this plot, we hold r_k and σ to be constant while varying r_1 and Y_1.
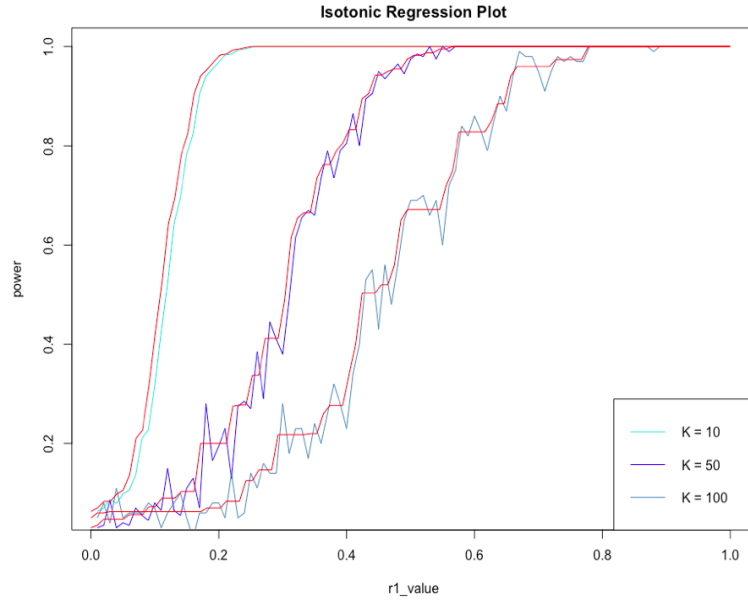
Figure 2.1: Isotonic Regression Plot

3 Repetition Times

To make our analysis much more reasonable, we try to increases the repetition times of our experiment. But the ways to increase the repetition times is very important. Simply just increasing N is actually not the best choice since the σ will also increase at the same time, leading to the less reasonable comparison between the experiments. So we come up with other ways to increase the repetition times, in the sense of bootstrap sampling and independent repetition experiments, deriving more sample p-values and taking the mean of them.

From a power derivation with one p-value matrix, we could at least get $\frac{N}{K}$ number for getting power. However, given the assumption of independence, we could repeatedly randomly sample to increase the repetition number. (rep_10, rep_50, rep_n, boot_50 increase from 10 times to 50*n times) In addition, we could repeat the whole power derivation process and get more independent results. (cv_10, cv_50) The results is shown in the following.
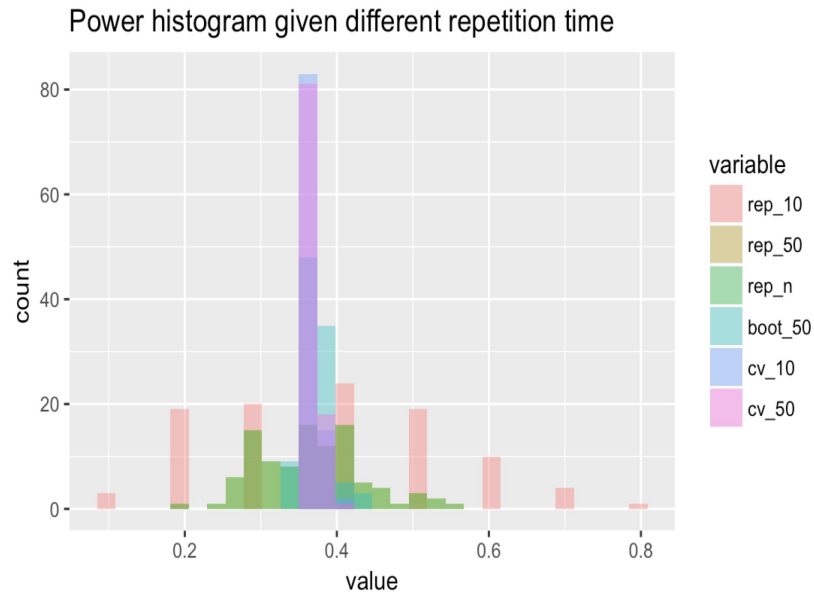

Figure 3.1: Power histogram given different ways to increase the repetition times

Note as the repetition time increases, the power samples are more closely distributed. And the repeated independent experiment is better off than bootstrap sampling.

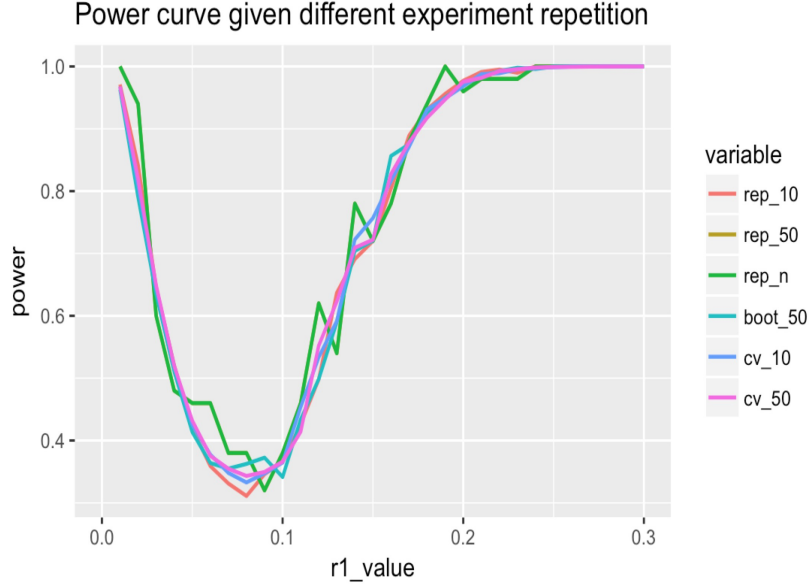We next test the power curve given different repetition times.

Figure 3.2: Power curve given different ways to increase the repetition times

The independent method (cv_10 and cv_50) has more smooth curve than others. Also the variance of bootstrap method is 0.0003649825, while independent repetition is 9.204492e-05 and 4.950972e-05 respectively for 10 times and 50 times. Since the 50 times are much time costing than 10 times while the disparity of the two is not that much, we choose the repetition time of 10 as our applied power derivation choice.

4   Scaling

For this part, we hold $r_1 = \frac{Y_1}{\sigma_1}$ and $r_k = \frac{\sigma_k}{\sigma_1}$ to be constant, which means that $\frac{Y_1}{\sqrt{\sigma_1^2 + \sigma_k^2}}$ is constant. Under this condition, we try to plot Y ranging from 0 to 1 by 0.01 step, as well as 1.2Y, 1.5Y and 2Y. But at the same time, we find the the ratio r_1 also has significant impact on the power, leading to the different power pattern of the plot. So we also make different plots by r_1 ranges from 0.01 to 0.3.

The good news is that when we hold r_1 and r_k constant, no matter how Y_1 changes, the power pattern seems to be the same. But from the plot we can see that when $r_1 = 0.01$, it turns out to be a "S-shape" curve; in $r_1 = 0.03$ and $r_1 = 0.05$, the horizontal part of the curve tends to oscillates more and the power, from the y-axis we can see, becomes smaller. When $r_1 = 0.08$, the plot seems all to be the horizontal oscillating part, and the power is around the smallest value, where maximum equals to around 0.3. When r_1 ranges from 0.1 to 0.2, we can see that the power all oscillates and increases closed to 1. And from Figure 4.2, we can see that all power equals to 1.

To the further discussion of r_1, we hold Y_1 constant while varying r_1 to see its power pattern as follow.

We can see that it corresponds to the plot before as the power first decreases and then increases. And if Y_1 increases, the power will become more stable.

5   P-combined Analysis

Before, we take the minimum of p-values to generate p_combined value, and this time, we take different ways to generate p_combined by taking the maximum, median and mean values. Also, we take treatment size K into consideration.
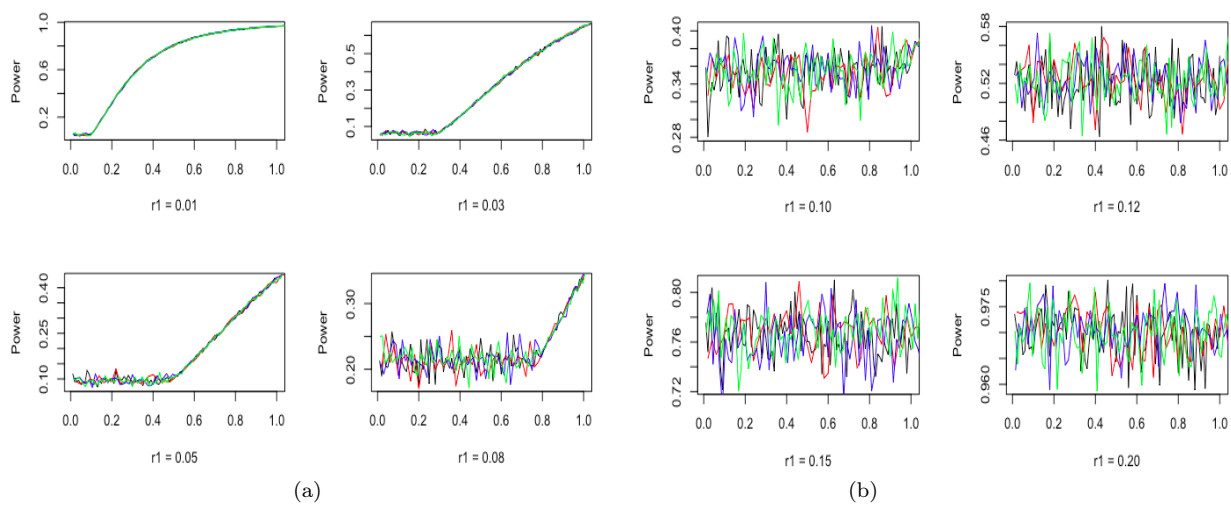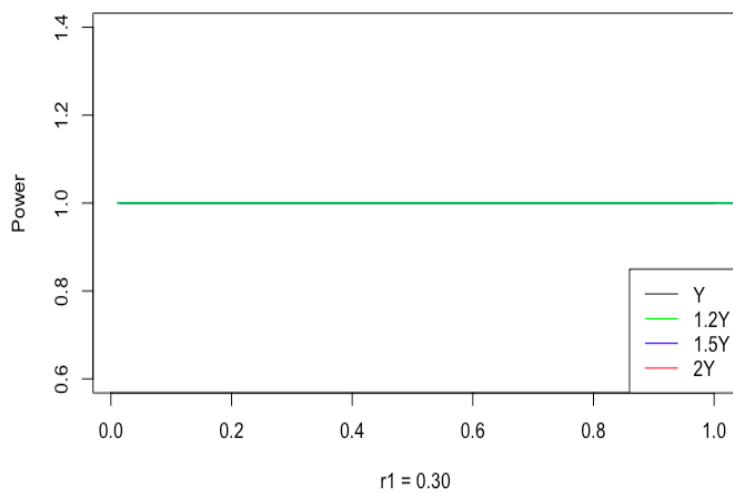
(a)   (b)

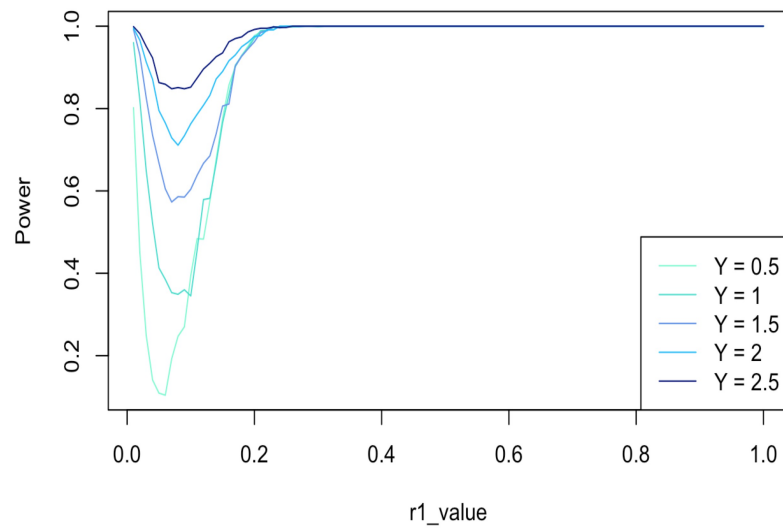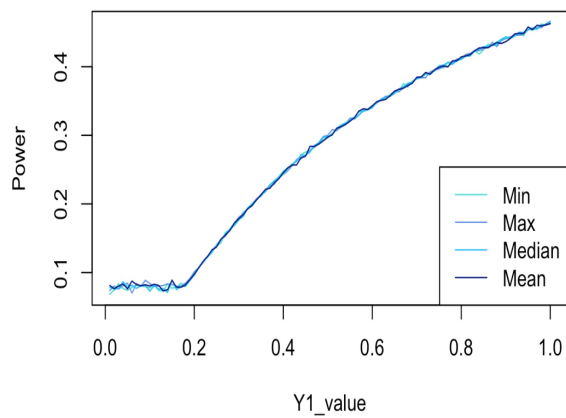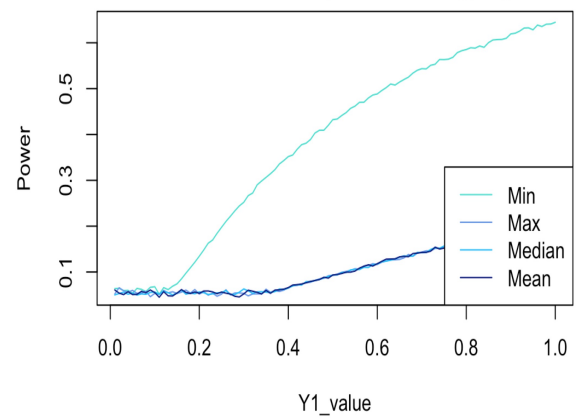Figure 4.1: Scaling Plot

Figure 4.2: Scaling Plot

Figure 4.3: r_1 Plot

(a) K=2

(b) K=3

Figure 5.1: Power plot of different type of p_combined
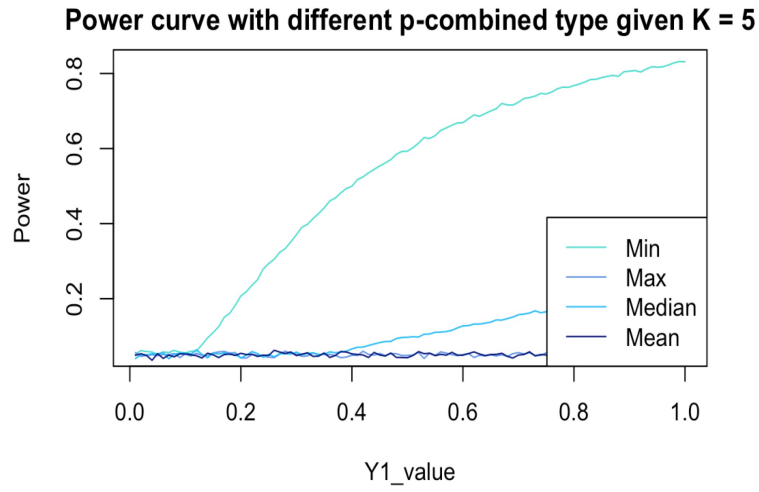
Power curve with different p-combined type given K = 5

Figure 5.2: Power plot when K=5 by different type of p_combined

From the plot we can see that as $K = 2$, all four types of p_combined seems to have the same pattern. But when K increases to 3, the minimum still keeps the "S-shape" curve but other three curves become lower. When $K = 5$, actually including all values bigger than 5 according to our experiments, the minimum curve still keeps its pattern; the maximum and mean curve just not increase any more and stay horizontally; the median curve has a small increasing trend, a little higher than the mean and maximum curve but much lower than the minimum curve. So we can see the the minimum curve is the most stable one.