Protocol and Power Analysis

Fall 2023 Stat 158: Experimental Design Project
Group2: Lisa Li, Yunfei Peng, Dingda Yang

I. Description

We plan to conduct a **Completely Randomized** experiment in which we evaluate the effect of question layout, order, and time pressure on college students' calculation ability. At each run, each subject will be asked to complete a calculation test with 15 questions within 4 minutes, under a certain treatment combination of the factor levels. And the accuracy of each subject will be recorded.

II. Power Calculation

2.1 Definition of sample size

Since we will run a CR[3] design experiment and will assign treatments to students in Lab section, our experimental units will be subjects (i.e., students). In this case, we define our sample size as the number of subjects in the experiment.

2.2 Estimation of σ_{err}^2

Based on our dry run which follows the procedure described in the initial proposal, we made some changes in the procedure of our experiment. Then we conducted the trial run based on the new protocol.

Firstly, we created 3 quizzes of similar difficulty and the same 4-minute time limitation with the help of online test tools. And then, we chose 3 statistics-major students (2 from our group), who have similar mathematical educational background, to conduct the trial run, taking 3 measurements from each of them under the same treatment combination (i.e., layout: *one at a time*; order: *easy to hard*; timer display: *with countdown timer*). Finally, we obtained the results from the online test tool, and calculated the accuracy respectively. Therefore, we got 9 observations in the trial run. The pilot data is as follows (to simplify the results, we keep the results in the form of fraction instead of percentages):

subject1: 14/15, 1, 1

subject2: 14/15, 12/15, 1

subject3: 14.5/15, 1, 14/15

Since we will run a CR[3] design experiment, we assume that there is little variation between subjects so that we took the variance from the pilot data to calculate the estimated variance for error. And we get the estimated $\hat{\sigma}_{err}^2$ = 0.0042 with $\hat{\sigma}_{err}$ = 0.065, which is quite reasonable.

2.3 Choosing alternative hypothesis

From the pilot data, we see that the results range from 12/15 to 1. Since our main goal is to figure out whether the form of tests will influence students' capacity of calculation, the effect of the three factors won't be too big. That is to say, when we only change the levels in one specific factor and keep other factors fixed, answering one more question correctly can be considered as the effect caused by that factor. In this case, for the factors which have two levels, we construct the alternative hypotheses under zero sum constraints:

Effects of factor *layout of questions*: $H_A: \alpha_1 = \frac{1}{15}, \alpha_2 = -\frac{1}{15}$

Effects of factor *display of timer*: $H_A: \gamma_1 = \frac{1}{15}, \gamma_2 = -\frac{1}{15}$

For the factor *order of questions*, we think that the psychological pressure is likely to be greater in *hard to easy* level so that we put a larger weight on that level, keeping the other 2 levels with the same weights. In this case, for the factor *order of questions*, we construct the alternative hypothesis under zero sum constraints:

Effects of factor *order of questions*: $H_A: \beta_1 = \frac{1}{15}, \beta_2 = -\frac{2}{15}, \beta_3 = \frac{1}{15}$

2.4 Power curve for the experiment

The factor *layout of questions* (vertical dashed noting n = 3, 4 from left to right):

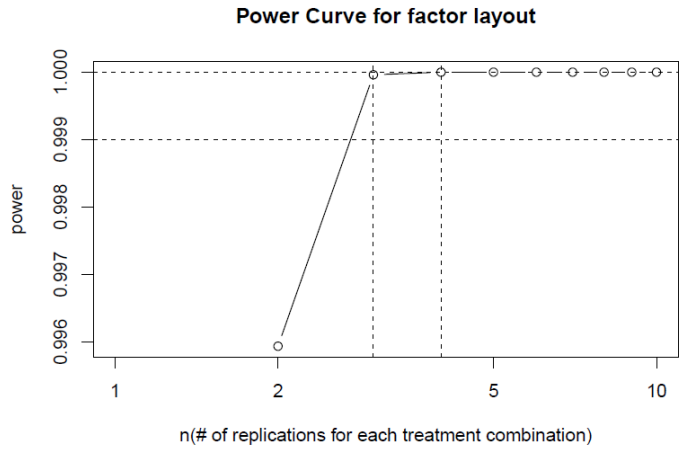


Figure 1 Power curve for factor layout (R code in detail is in appendix)

The factor *order of questions* (vertical dashed noting n = 3, 4 from left to right):

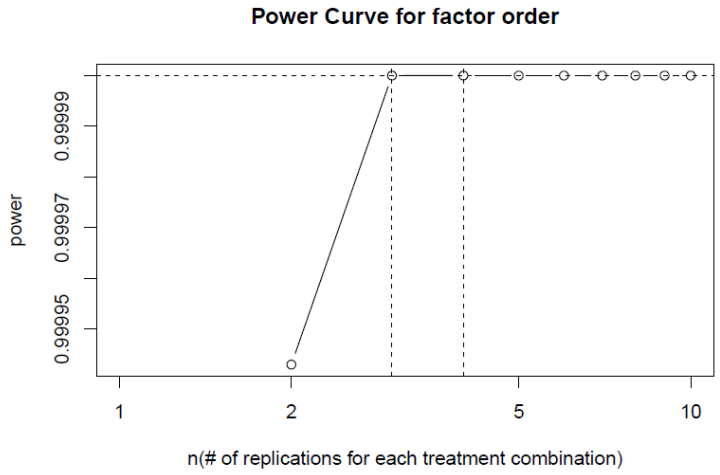


Figure 2 Power curve for factor layout (R code in detail is in appendix)

The factor *display of timer* (vertical dashed noting n = 3, 4 from left to right):

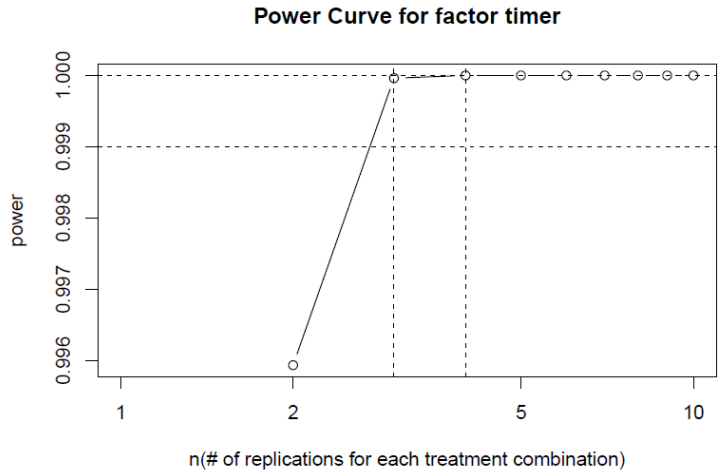


Figure 3 Power curve for factor timer (R code in detail is in appendix)

2.5 Final sample size and Power

Based on the power curve above, it seems that we can have a very high power (greater than 0.99) for all 3 factors using only a small number of replications in each treatment combination. In this case, considering the feasibility of the experiment, we would choose to have 3 replications for each treatment combination. In other words, we will have $3 \times 12 = 36$ subjects in total for our experiment. For the factor *layout of questions*, this sample size will give us a power of 0.99996; for the factor *order of questions*, it will give us a power of 1; for the factor *display of timer*, it will give us a power of 0.99996.

III. Protocol of Experiment

3.1 Preparations before experiment

Before conducting experiment, we need to make up questions for the calculation test and determine the testing time carefully. In order to guarantee the difficulty and intensity, there are 15 questions in our test, including addition, subtraction, multiplication, division, integral operations, and derivatives and subjects will be asked to complete the test within 4 minutes. One certain test will be shown in section 5.1 in the appendix part and all treatment combinations will be enforced based on questions in this certain test.

We are going to use www.wenjuan.com to make up the questions and collect the response. This website is intended for collecting data or creating online test, which is professional and convenient because it can simulate online test environment and we can create tests according to our treatment combinations. For example, one of our treatment combinations is one at a time/easy to hard/with countdown timer, so its corresponding test interface can be like that (Figure 4):

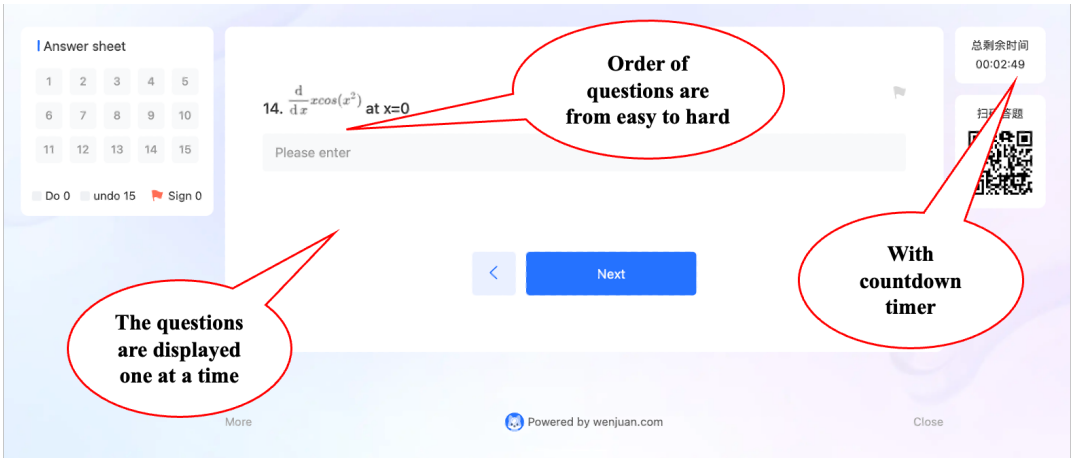


Figure 4 The interface of the test under one treatment combination

In order to randomly assign the units to our treatment combinations, we need to number all the treatment combinations from 1 to 12 (seen in appendix) in advance. Since there will be 3 replications in our experiment, we can conduct the R command `sample(rep(1:12,3))` to randomly assign 36 units to the treatment combinations. The specific procedure of assignment will be elaborated in section 3.2. So the final data will be recorded like that (Table 2):

Table 1 Table for recording response (only the first six rows are displayed and the full table will be shown in the appendix part)

Run	***#Treatment***	***Layout of Questions***	***Order of Questions***	***Countdown Timer***	***Accuracy (%)***	***Notes***
1	4	one at a time	hard to easy	no		

2	4	one at a time	hard to easy	no		
3	10	on single page	hard to easy	no		
4	9	one at a time	easy to hard	yes		
5	5	one at a time	random	yes		
6	6	one at a time	random	no		

Prior to the experiment, we also need to prepare a laptop for subjects to take tests, several pens and draft sheets for them to do calculation, and chocolates for incentive and rewards. In addition, the laptop and the test link should be checked in advance to ensure good working.

3.2 Implementation of experiment

There should be 3 replications in our experiment so 36 units, that is, 36 students in STAT 158 are expected to participate in our test. So we need to use both two labs on Nov.20th, 2023. We have a total of 36 runs in our experiment. At a single run, the procedure will be: (We only discussed about the situation where students will come to our station one by one on that day.)

***Step(i)* Assigning Treatment Combination**

When the subject comes to our station, we will randomly assign him/her a number, which stands for one treatment combination, according to the outcome of the R command `sample(rep(1:12,3))` and the order in which the subject comes to our station.

***Step(ii)* Providing Test Equipment and Instructions**

The subject will be equipped with a laptop, a pen, and a draft sheet at our station. One of our group members will assign the corresponding test link to the subject and read the instructions for the subject, which will also be displayed at the original interface of the test. At that time, the test interface will be like that:

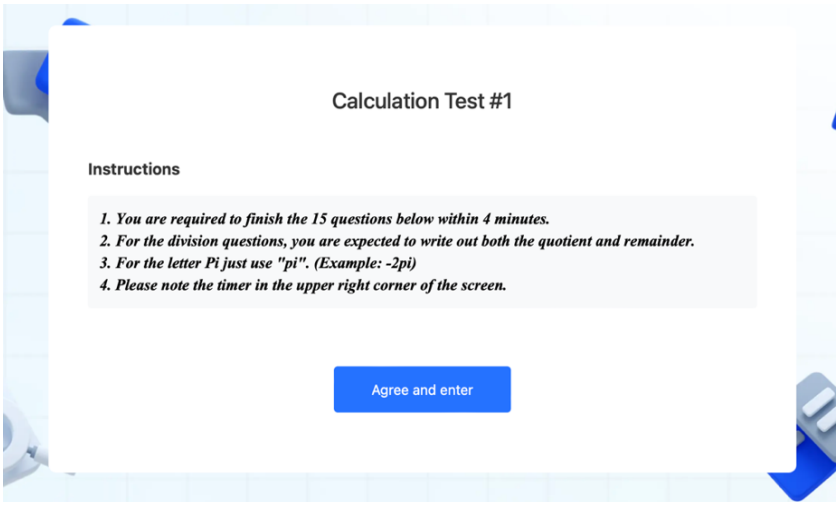


Figure 5 The original interface of the test under treatment combination #1

***Step(iii)* Initiating the Test**

After the subject is ready, he/she can click “Agree and enter” and start the test. At the same time, one of our group members will start the stopwatch. Throughout the test, any attempts at cheating, such as using calculators, will be prevented. After four minutes, we will stop the stopwatch and ask the subject to submit his/her test.

***Step(iv)* Recording the Data**

The accuracy of the test will be automatically recorded in this website. Any special circumstances during the experiment can be recorded in "Notes" column in the table. After the run, the participant will be rewarded with chocolates and they will be asked to share their feedback and impressions of this experiment.

3.3 Attention points

There are several steps that should be adhered to during the preparation and implementation of the experiment:

- Ensure that the difficulty, intensity and breadth of test aligns with our research objectives and effectively measures students' abilities. That’s why we need to do dry runs and make changes after dry runs; (The specific changes after dry runs will be elaborated in next chapter.)
- Check the testing links on www.wenjuan.com and the laptop in advance to ensure the stability of the online testing environment. This includes ensuring that participants can smoothly access the testing interface and preventing any potential technical issues;
- Since the countdown timer on the top right of the screen is very small, it is important to instruct participants to take note of its presence if the participants are assigned to the treatment combination including countdown timer.

IV. Changes After Dry Run

Before finalizing the protocol, we conducted a dry run on each of three group members. In order to test each level of each factor, we introduced three rounds in our dry run and we chose three treatment combinations which involve each level of each factor. At each round, the subject was asked to complete 20 calculating questions within 5 minutes using Yunfei's ipad, under certain treatment combination. The questions for dry run are shown in appendix. The result of dry run is listed in the following table (Table 3).

Table 2 Result of dry run

Group Member	***Layout***	***Order***	***Timer***	***Accuracy(%)***
Yunfei Peng	one at a time	easy to hard	yes	70
Dingda Yang	one at a time	hard to easy	no	90
Lisa Li	at single page	random	yes	90

Based on the dry run, we have made two main changes. Firstly, we cut down the number of questions and testing time from 20 questions in 5 minutes to 15 questions in 4 minutes. The reason for this kind of change is that we may not have enough time to get all 36 people to finish the experiment during the lab sections if we set the testing time to 5 minutes. Accordingly, we have reduced the number of questions from 20 to 15. This change is also aimed at letting participants maintain a high level of concentration and prevent them from slacking off, lack of deliberation or even giving up the test especially when facing some complicated integral and derivatives calculation.

Secondly, we decided to use an online test platform (www.wenjuan.com) to make up questions, carry out tests and collect the results, instead of using ipad or paper as we thought before. This change is firstly aimed at cutting down the logistical cost for our experiment because this platform has the function of adding timer, reshuffling the questions and displaying questions one at a time so that we don't need to enforce treatment manually. The second reason is that this platform can simulate a true online test environment and also provide participants with a user-friendly experience, which help us prevent unwanted intervenes and better evaluate the effect of factors of interest on students' test performance.

V. Appendix

5.1 One certain test

This test is under the treatment combination of on a single page/easy to hard/with countdown timer.

Calculation Test #1

Instructions:
1. You are required to finish the 15 questions below within 4 minutes.
2. For the division questions, you are expected to write out both the quotient and remainder.
3. For the letter Pi just use "pi". (Example: -2pi)
4. Please note the timer in the upper right corner of the screen.

1. 15+11
2. 21+12
3. 34-22
4. 63-21
5. 42-25
6. 36+47
7. 123-88
8. 162+189
9. 13×14
10. 58×39
11. 462÷8=___...___
12. 693÷17=___...___
13. $\int_0^{\sqrt{\pi}} xcos(x^2)dx$
14. $\frac{d}{dx}xsin(2x)$ evaluate at x=0
15. $3\int_0^1 x^2e^{x^2}\,dx$

5.2 Numbering of treatment combination

#Treatment Combination	*Layout of Questions*	*Order of Questions*	*Countdown Timer*
1	one at a time	easy to hard	yes
2	one at a time	easy to hard	no
3	one at a time	hard to easy	yes
4	one at a time	hard to easy	no
5	one at a time	random	yes
6	one at a time	random	no
7	on a singe page	easy to hard	yes
8	on a singe page	easy to hard	no
9	on a single page	hard to easy	yes
10	on a single page	hard to easy	no
11	on a single page	random	yes
12	on a single page	random	no

5.3 R code for random assignment

```
> set.seed(1234)
```

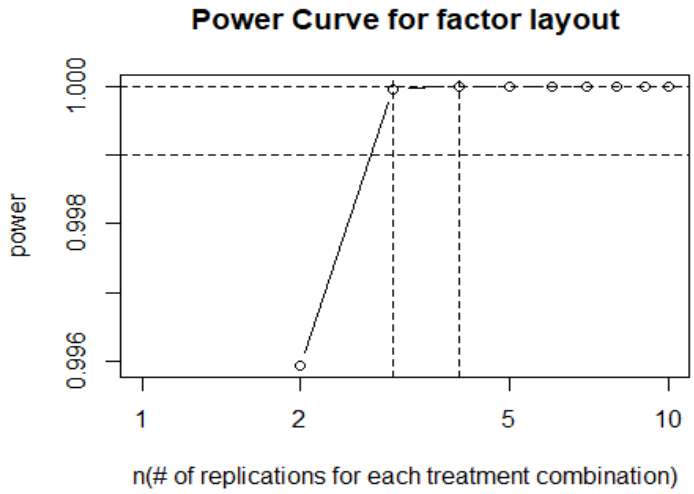
```
> sample(rep(1:12,3))
```

```
 [1]  4  4 10  9  5  6 11  4  2  7 10  7  3  2  8 11  5  2  8  3  7 12
1 10  8 12  6  6  9  9  5  1  1 11  3 12
```

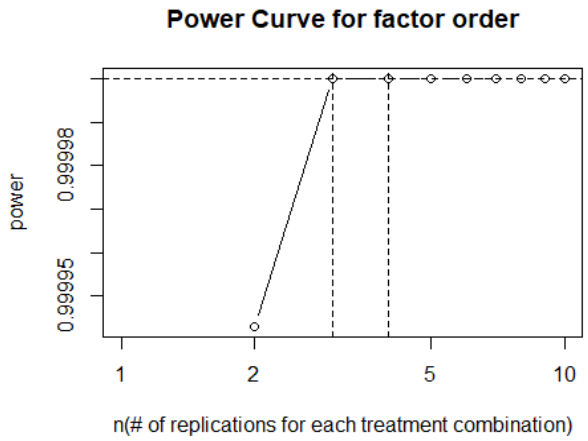
5.4 R code for power calculation

```
# Estimate sigma-square

pilot <- c(14/15, 1, 1, 14/15, 12/15, 1, 14.5/15, 1, 14/15)
var(pilot)

## [1] 0.004197531

# power curve for factor layout
nseq <- 1:10
groupmeans <- c(1/15, -1/15)
sigmasquare <- var(pilot)
power <- sapply(nseq,function(n){
    nperlevel <- n * 3 * 2
```

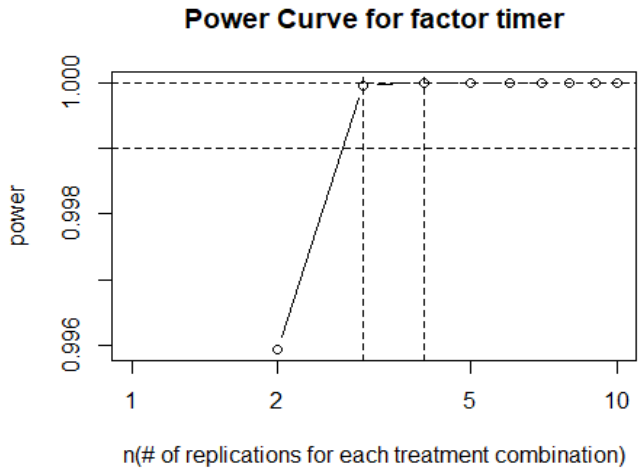
```
    ncpseq <- nperlevel * sum(groupmeans^2)/sigmasquare
    dfDenom <- 3 * 2 * 2 * (n-1)
    fcutoff <- qf(1-0.05,df=1,df2=dfDenom)
    1-pf(fcutoff,df1=1,df2=dfDenom,ncp=ncpseq)
})
plot(nseq,power,log="x",type="b",
     xlab="n(# of replications for each treatment combination)",
     main="Power Curve for factor layout")
abline(h=c(0.999, 1), v = c(3, 4), lty=2)
```




```
# power curve for factor order
groupmeans <- c(1/15, -2/15, 1/15)
power <- sapply(nseq,function(n){
    nperlevel <- n * 2 * 2
    ncpseq <- nperlevel * sum(groupmeans^2)/sigmasquare
    dfDenom <- 3 * 2 * 2 * (n-1)
    fcutoff <- qf(1-0.05,df=2,df2=dfDenom)
    1-pf(fcutoff,df1=2,df2=dfDenom,ncp=ncpseq)
})
plot(nseq,power,log="x",type="b",
     xlab="n(# of replications for each treatment combination)",
     main="Power Curve for factor order")
abline(h=c(0.999, 1),v = c(3, 4), lty=2)
```




```r
# power curve for factor timer
groupmeans <- c(1/15, -1/15)
power <- sapply(nseq,function(n){
    nperlevel <- n * 3 * 2
    ncpseq <- nperlevel * sum(groupmeans^2)/sigmasquare
    dfDenom <- 3 * 2 * 2 * (n-1)
    fcutoff <- qf(1-0.05,df=1,df2=dfDenom)
    1-pf(fcutoff,df1=1,df2=dfDenom,ncp=ncpseq)
})
plot(nseq,power,log="x",type="b",
     xlab="n(# of replications for each treatment combination)",
     main="Power Curve for factor timer")
abline(h=c(0.999, 1),v = c(3, 4), lty=2)
```



As we can see in the three plots, when we have 3 or 4 replications for each treatment combination, we've already had a pretty high power for the experiment. In this case, when considering feasibility, we will choose **3** as the number of replications in each treatment combination.

```r
# power for factor layout
groupmeans <- c(1/15, -1/15)
power_layout <- sapply(3, function(n){
  nperlevel <- n * 3 * 2
    ncp <- nperlevel * sum(groupmeans^2)/sigmasquare
    dfDenom <- 3 * 2 * 2 * (n-1)
    fcutoff <- qf(1-0.05,df=1,df2=dfDenom)
    1-pf(fcutoff,df1=1,df2=dfDenom,ncp=ncp)
})
power_layout
```

```
## [1] 0.9999614
```

```r
# power for factor order
groupmeans <- c(1/15, -2/15, 1/15)
power_order <- sapply(3, function(n){
  nperlevel <- n * 2 * 2
    ncp <- nperlevel * sum(groupmeans^2)/sigmasquare
    dfDenom <- 3 * 2 * 2 * (n-1)
    fcutoff <- qf(1-0.05,df=2,df2=dfDenom)
    1-pf(fcutoff,df1=2,df2=dfDenom,ncp=ncp)
})
power_order
```

```
## [1] 1
```

```r
# power for factor timer
groupmeans <- c(1/15, -1/15)
power_timer <- sapply(3, function(n){
  nperlevel <- n * 3 * 2
    ncp <- nperlevel * sum(groupmeans^2)/sigmasquare
    dfDenom <- 3 * 2 * 2 * (n-1)
    fcutoff <- qf(1-0.05,df=1,df2=dfDenom)
    1-pf(fcutoff,df1=1,df2=dfDenom,ncp=ncp)
})
power_timer
```

```
## [1] 0.9999614
```

5.5 Table for recording response

Run	#treatment combination	display	order	timer	accuracy(%)	Notes
1	4	one	hard to easy	no		
2	4	one	hard to easy	no		
3	10	full	hard to easy	no		
4	9	full	easy to hard	yes		
5	5	one	random	yes		
6	6	one	random	no		
7	11	full	random	yes		
8	4	one	hard to easy	no		
9	2	one	easy to hard	no		
10	7	full	easy to hard	yes		
11	10	full	hard to easy	no		
12	7	full	easy to hard	yes		
13	3	one	hard to easy	yes		
14	2	one	easy to hard	no		
15	8	full	easy to hard	no		
16	11	full	random	yes		
17	5	one	random	yes		
18	2	one	easy to hard	no		
19	8	full	easy to hard	no		
20	3	one	hard to easy	yes		
21	7	full	easy to hard	yes		
22	12	full	random	no		
23	1	one	easy to hard	yes		
24	10	full	hard to easy	no		
25	8	full	easy to hard	no		
26	12	full	random	no		
27	6	one	random	no		
28	6	one	random	no		
29	9	full	easy to hard	yes		
30	9	full	easy to hard	yes		
31	5	one	random	yes		
32	1	one	easy to hard	yes		
33	1	one	easy to hard	yes		
34	11	full	random	yes		
35	3	one	hard to easy	yes		
36	12	full	random	no		

5.6 Questions for dry run

You are requested to complete the test below within five minutes.

Please write the quotient and remainder of a division operation.

1. 15 + 14
2. 21 + 15
3. 73 - 24
4. 53 + 36
5. 85 - 31
6. 29 + 38
7. 67 + 26
8. 71 - 83
9. 36 + 45
10. 88 + 127
11. 26 x12
12. 38 x 22
13. 476 ÷ 24
14. 55 × 55
15. 2023 × 71
16. $\int x\sin\left(x^2\right)dx$
17. $\frac{d}{dx}\left(x\sin(x)\right)$
18. $\frac{d}{dx}\left(\frac{x^2}{sin(x)}\right)$
19. $\int xe^x dx$
20. $\int x\cos(x)dx$