Simulation in Statistics

Investigation of one-way ANOVA assumptions by simulation

Gan Luan

What is simulation?

Simulation is a way to model random events, such that the simulated outcomes closely match real-word outcomes. By observing simulated outcomes, researchers gain insight on the real word.

Example

Suppose LeBron James can make 80% of his free throws.

Q: If he goes to free throw line 5 times, how likely he made all 5 free throws?

Simulation:

- State the question:

  What is the probability that LeBron could make all 5 free throws?

- Plan:

  - ✓ We use integers randomly generated from 0 to 9; 0-7 means he makes it; 8-9 means he misses it
  - ✓ 1 trail is 5 shots: generate 5 integers

- Run 10 trails:

  [0, 9, 8, 1, 9] [1, 0, 7, 4, 4] [7, 8, 9, 9, 4] [9, 8, 6, 8, 9]
  [9, 2, 3, 5, 6] [0, 2, 6, 7, 4] [2, 3, 2, 3, 6] [4, 6, 2, 5, 1]
  [4, 0, 3, 2, 7] [0, 9, 5, 1, 4]

- Conclusion:

  The probability that LeBron could make all 5 free throws is 0.5.

Simulation and one-way ANOVA assumptions

one-way ANOVA assumptions

- Randomization
- Equal variance
- Normal assumption

Investigate equal variance assumption by stimulation

- State the question:

  How the one-way ANOVA conclusion will be affected if samples are from distributions that have different variances?

Plan:

- Suppose three treatments generating results following different distributions

  Treatment i: result i ~ $N(\mu_i, \sigma_i^2)$ (i=1, 2, 3)

- Sample i of size 30 is from $N(\mu_i, \sigma_i^2)$ as the response of treatment i (i=1, 2, 3)

- Perform one-way ANOVA for sample 1, sample 2, and sample 3 and make decision of whether rejecting the null hypothesis or not.

- Compare the decision from ANOVA and the real relationship of μ_1, μ_2, and μ_3 and get the conclusion of whether this ANOVA result is correct or not.

- This is considered as one trail.

➢ Plan:

Suppose we have

treatment 1: result 1 ~ N(0, 1)

treatment 2: result 2 ~ N(0, 1)

treatment 3: result 3 ~ N(0, **2**)

Generate three samples from N (0,1), N (0, 1), and N (0, **2**) respectively, label as sample 1, 2, 3

Perform one-way ANVOA with sample 1, 2, 3.

- If the decision from ANOVA is that means of 3 samples are same, then ANOVA result is correct
- Otherwise the ANOVA result is wrong

- **Run 1000 trails (when true means are equal):**

Suppose we have

treatment 1: result 1 ~ N(0, 1)

treatment 2: result 2 ~ N(0, 1)

treatment 3: result 3 ~ N(0, **2**)

In 1000 trails, 945 trails made right decision, 45 trails made wrong decision.

- **Conclusion:**

In this case, we conclude that 94.5% of chance we can make right decision.

More general:

Suppose we have

treatment 1: result 1 ~ N(0, 1)

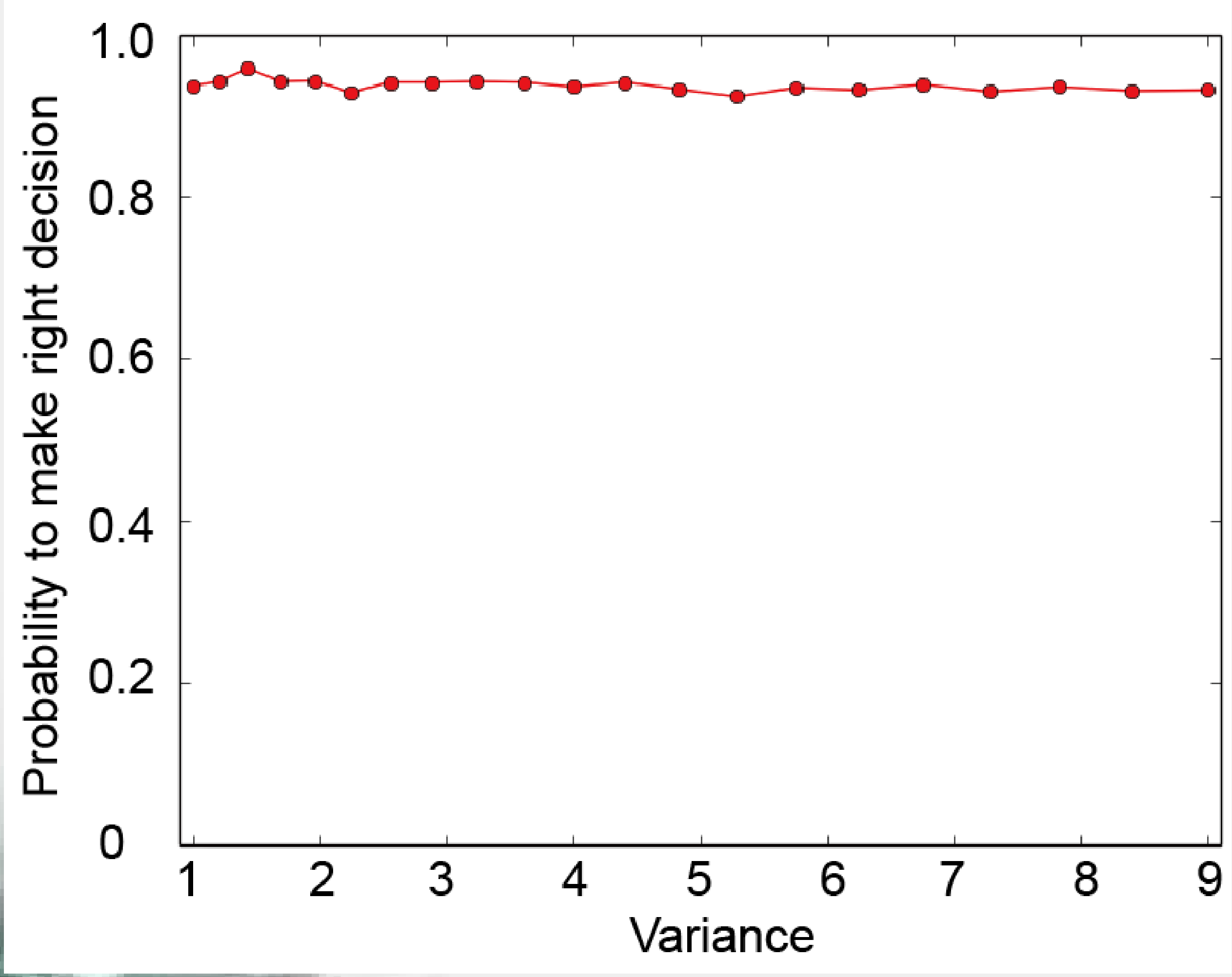
treatment 2: result 2 ~ N(0, 1)

treatment 3: result 3 ~ N(0, **variance**)

Variance be 1, 1.2, 1.4, …, 8.8, 9

Run 1000 trails for each variance, and calculate the rate we get the right conclusion.

When true means are equal



ANOVA is robust to unequal variance if the true means are equal.

- **Run 1000 trails (when true means are unequal):**

  Suppose we have

  treatment 1: result 1 ~ N(0, 1)

  treatment 2: result 2 ~ N(0, 1)

  treatment 3: result 3 ~ N(**1**, **2**)

  In 1000 trails, 922 trails reject the null hypothesis, 78 trails accept the null hypothesis.

- **Conclusion:**

  In this case, we conclude that 92.2% of chance we can make right decision.

More general:

Suppose we have

treatment 1: result 1 ~ N(0, 1)

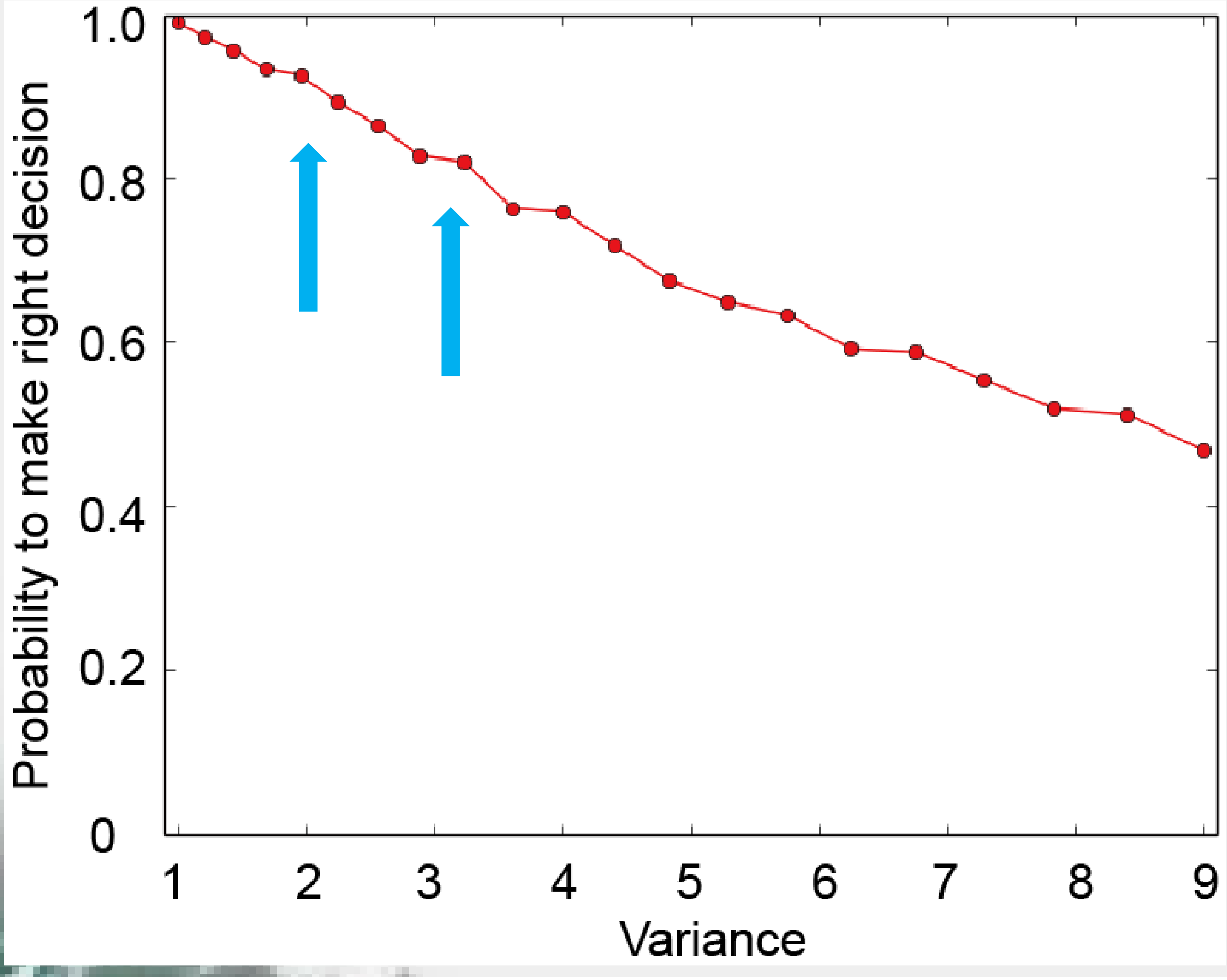
treatment 2: result 2 ~ N(0, 1)

treatment 3: result 3 ~ N(**1**, **variance**)

Variance be 1, 1.2, 1.4, …, 8.8, 9

Run 1000 trails for each variance, and calculate the rate we get the right conclusion.

When true means are unequal (the case we care about)



ANOVA is sensitive to unequal variance if the true means are unequal.

Investigate normal assumption by simulation

Suppose we have

treatment 1: result 1 ~ N(0, 1)

treatment 2: result 2 ~ N(0, 1)

treatment 3: result 3 ~ **Distribution 3 (mean, r)**

Distribution 3 is a mixture distribution of normal distribution and uniform distribution, with the ratio of uniform distribution be r.

Pdf of distribution 3 = (1-r)*normal_pdf + r* uniform_pdf $(0.5<r<1)$

The value of r represents of the extent of violation of normal assumption.

Investigate normal assumption by stimulation

when true means are equal

Suppose we have

treatment 1: result 1 ~ N(0, 1)

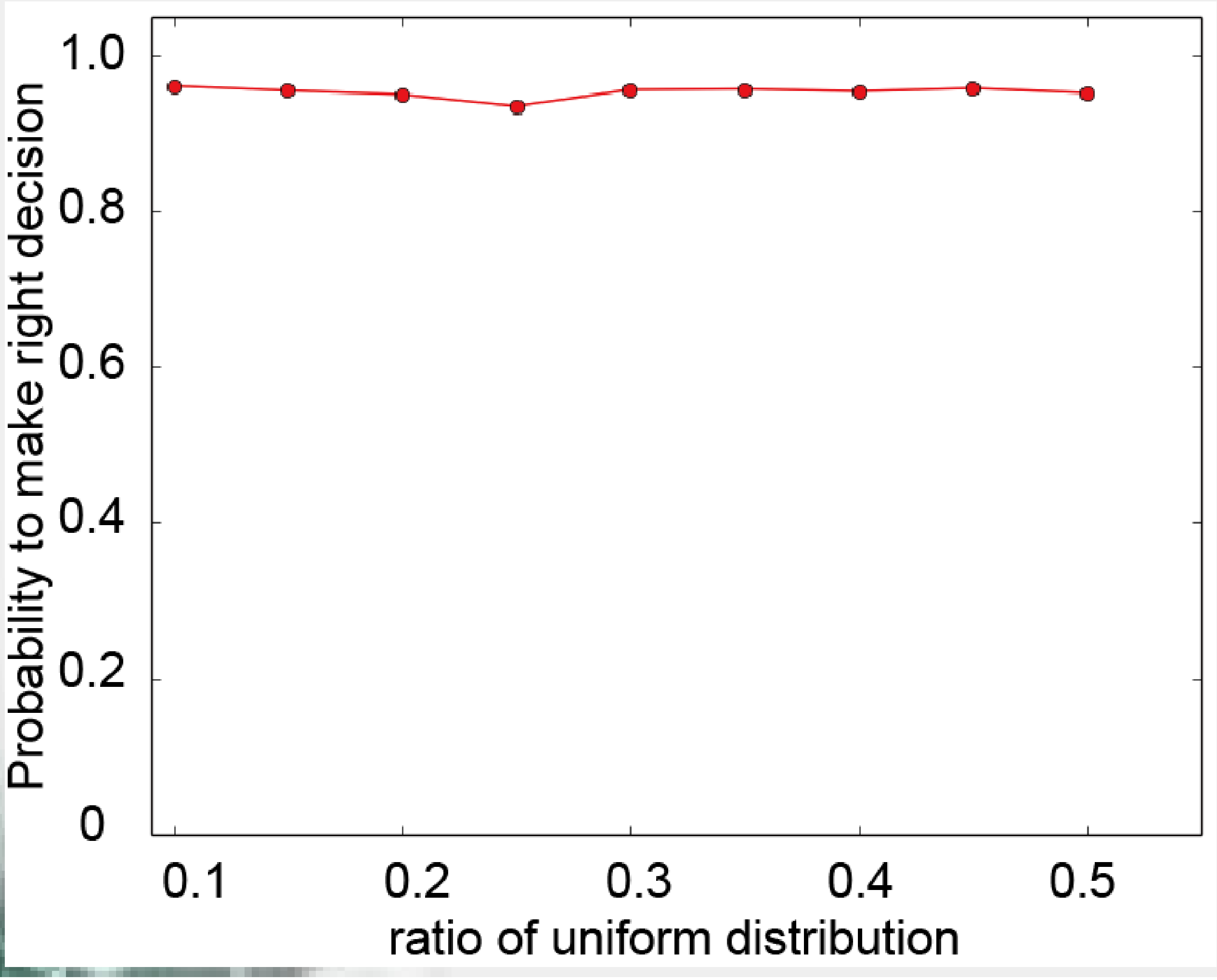
treatment 2: result 2 ~ N(0, 1)

treatment 3: result 3 ~ **Distribution 3 (0, r)**

r be 0.1, 0.15, … , 0.5

Run 1000 trails for each r, and calculate the rate we get the right conclusion.

When true means are equal



ANOVA is robust to non-normality if the true means are equal.

Investigate normal assumption by stimulation when true means are unequal

Suppose we have

treatment 1: result 1 ~ N(0, 1)

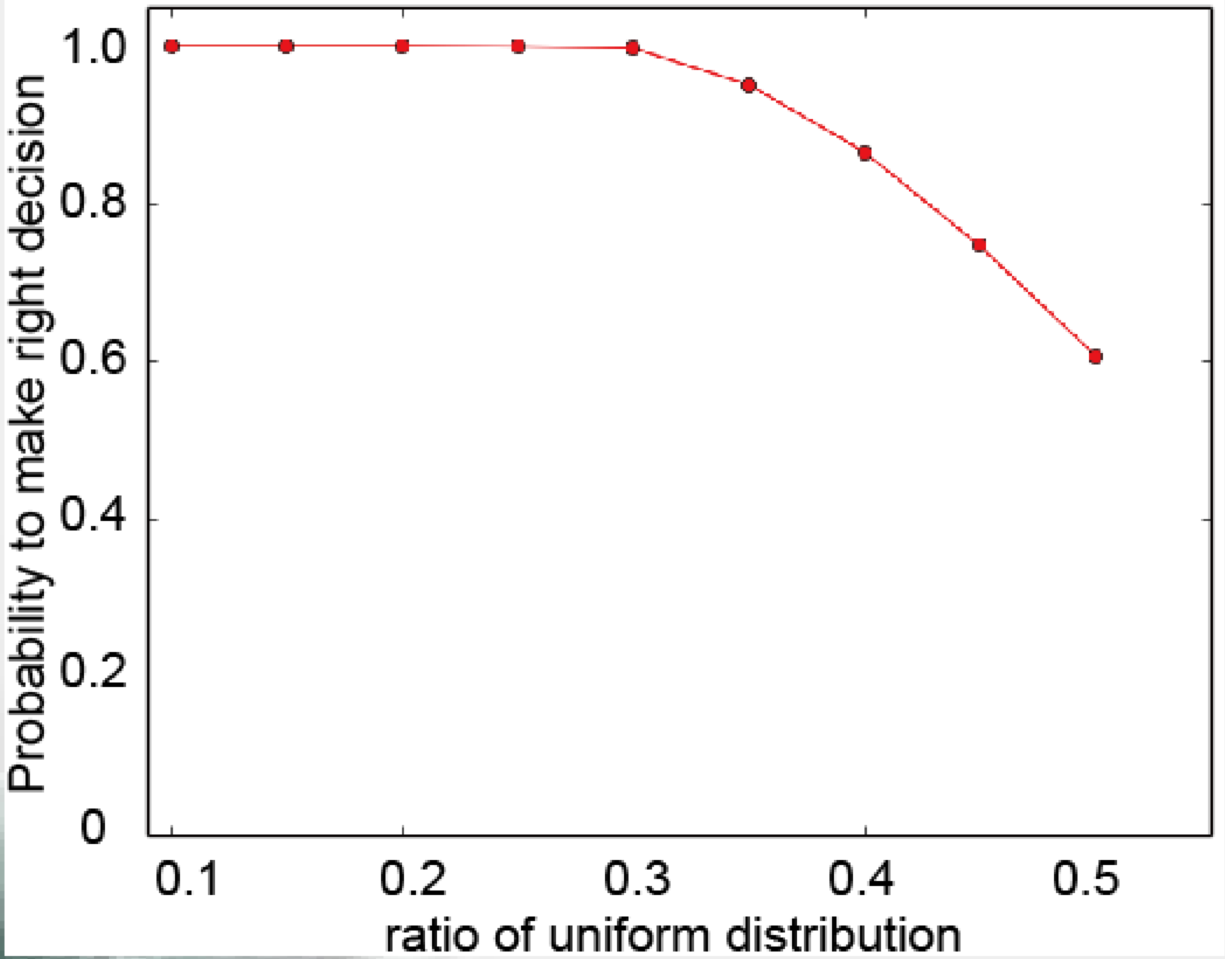
treatment 2: result 2 ~ N(0, 1)

treatment 3: result 3 ~ **Distribution 3 (1, r)**

r be 0.1, 0.15, … , 0.5

Run 1000 trails for each r, and calculate the rate we get the right conclusion.

When true means are unequal (the case we care about)



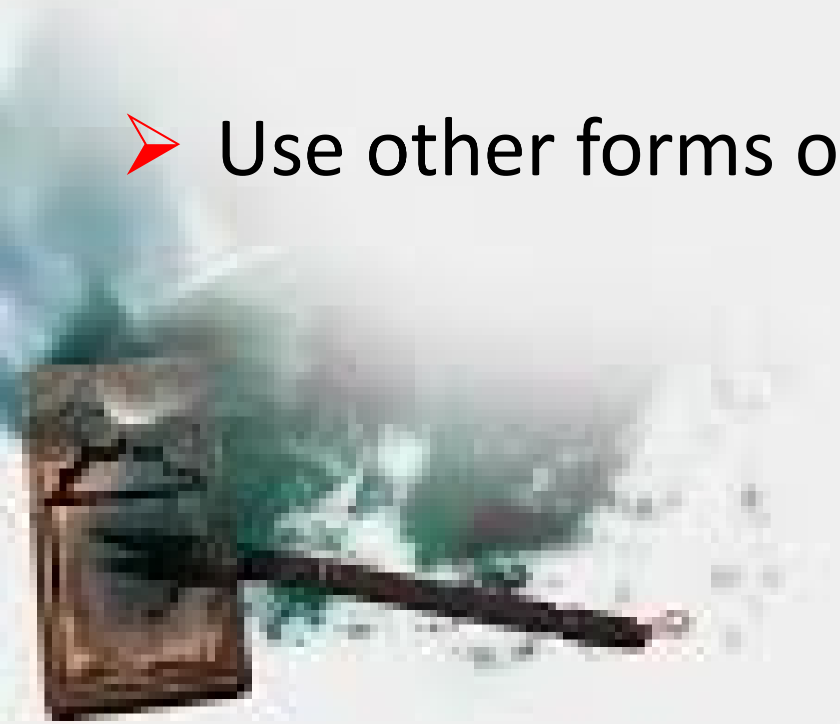
ANOVA is robust to non-normality if the true means are unequal.

Summary

One-way ANOVA is robust to normal assumption but sensitive to equal variance assumption.

Future plan

- Use more groups instead of three groups
- Use other values for difference between means
- Use other forms of non-normal distributions

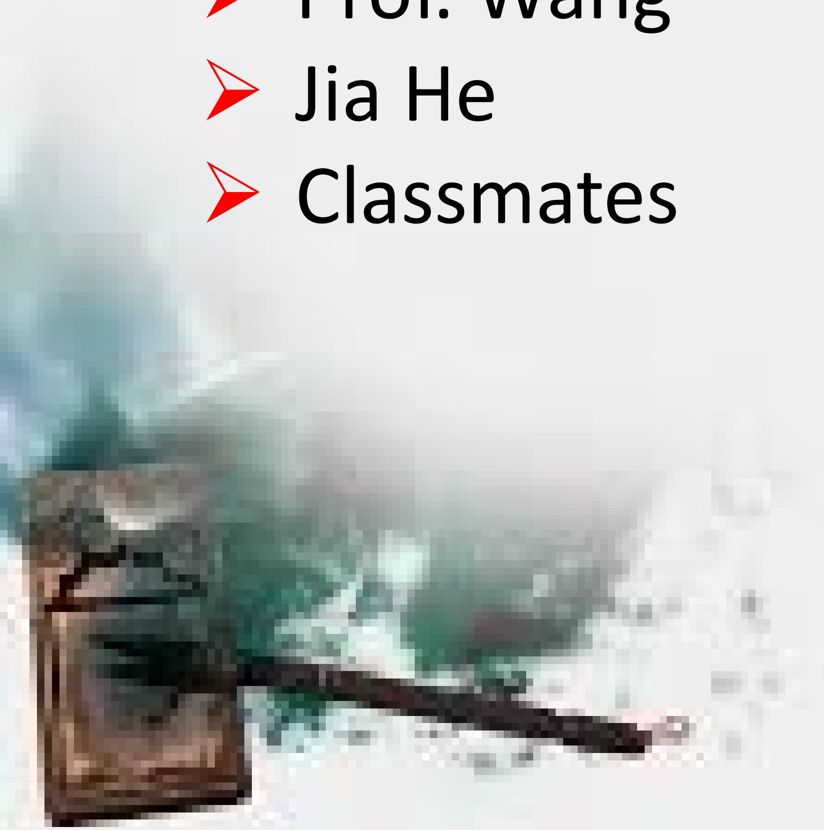


Take home message

- The idea of simulation in statistics
- Intuitive idea of ANOVA assumptions
- The program can be modified to study other statistical questions

Acknowledgement

- Prof. Dhar
- Prof. Loh
- Prof. Wang
- Jia He
- Classmates



Thank you!