Analytics for Observational Data (IT142IU)

Lab 8: Monte Carlo Simulations

## Objectives

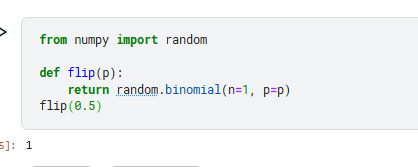
* Understanding Monte Carlo method as an estimation procedure.
* Applying Monte Carlo to simulate data for distribution probability estimation.
* Programming languages: Python/Java
* Ref: Lecture notes in Session 12

## Tasks

For this lab, we are going to make simulated coin flips.

1. Write a function called flip that takes in an argument p where p stands for the probability of flipping a head (you can code this as a 1 and 0 for tails) and outputs either 1 or 0.

Run the code “flip(0.5)”.

****

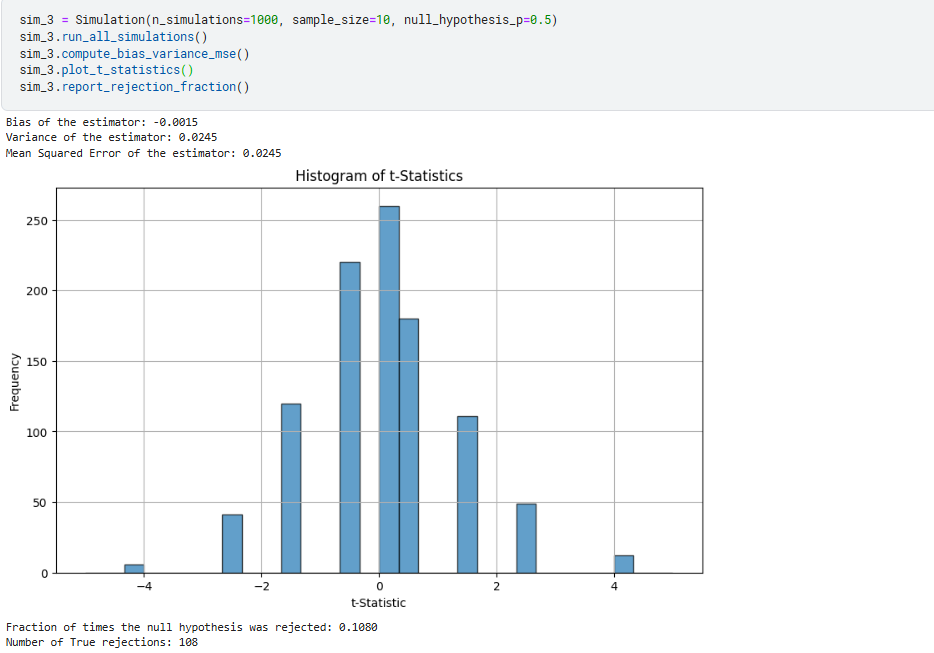
1. Write a function called generate\_sample that takes in the arguments n and p and generates a sample of n coin flips where the probability of flipping heads is p. Run the code “generate\_sample (10,0.5)”.
2. Next, over 1000 Monte Carlo simulations (i.e., do the following 1000 times),
   1. generate a new sample with 10 observations
   2. calculate an estimate of p

(**Hint:** you can estimate *p* by just calculating the average number of heads flipped in a particular simulation)

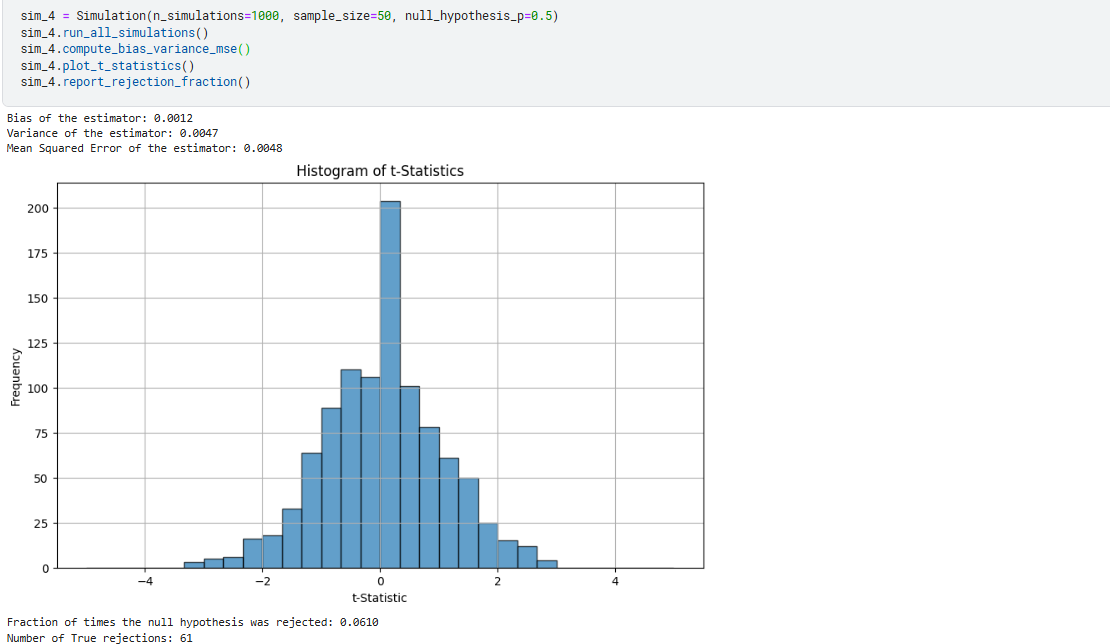
* 1. a t-statistic for the null hypothesis that p=0.5
  2. and record whether or not you reject the null hypothesis that p=0.5 in that simulation

Then, using all 1000 Monte Carlo simulations, report (i) an estimate of the bias of your estimator, (ii) an estimate of the variance of your estimator, (iii) an estimate of the mean squared error of your estimator, (iv) plot a histogram of the t-statistics across iterations, and (v) report the fraction of times that you reject H0.





1. Same as #3, but with 50 observations in each simulation. What differences do you notice?



1. Comparing Bias:

- Graph 1: Bias = -0.0015, indicating that the testing method has a very small deviation from the actual value.

- Graph 2: Bias = 0.0012, also indicating a small deviation, but larger than in Graph 1.

- Conclusion: The testing method in Graph 1 has higher accuracy in terms of bias.

2. Comparing Variance:

- Graph 1: Variance = 0.0245, indicating a relatively large dispersion of the testing results.

- Graph 2: Variance = 0.0047, indicating a smaller dispersion of the testing results, meaning the testing results are more stable.

- Conclusion: The testing method in Graph 2 has higher stability in terms of variance.

3. Comparing Mean Square Error (MSE):

- Graph 1: MSE = 0.0245, indicating a relatively large sum of squared errors for the testing method.

- Graph 2: MSE = 0.0048, indicating a smaller sum of squared errors for the testing method, meaning the testing method is more efficient.

- Conclusion: The testing method in Graph 2 is more efficient in terms of MSE.

4. Comparing Fraction of Times the Null Hypothesis was Rejected:

- Graph 1: Fraction of rejections = 1.0000, indicating that the null hypothesis was rejected in all simulations. This suggests that the method is overly sensitive and may be rejecting the null hypothesis too often.

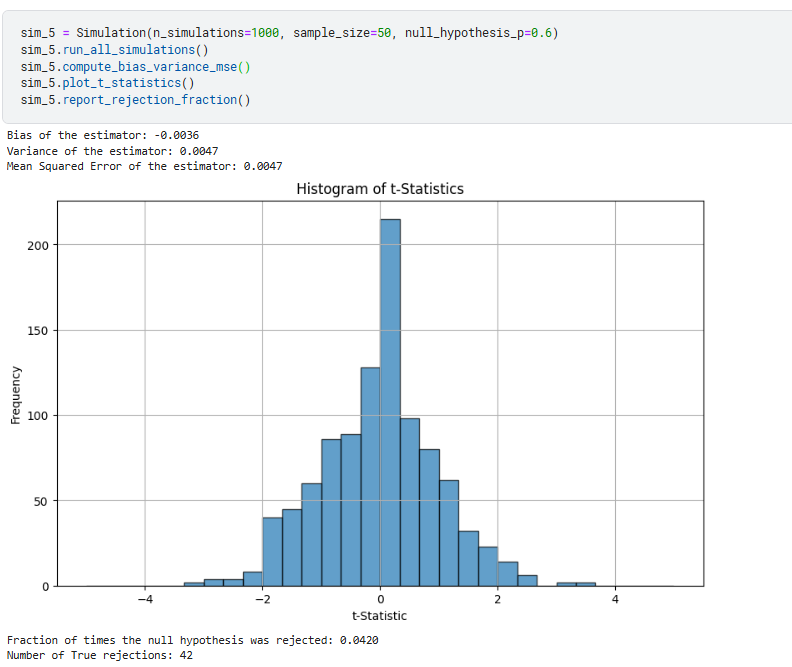
- Graph 2: Fraction of rejections = 0.8010, indicating that the null hypothesis was rejected in 80.1% of the simulations. This suggests a more balanced approach, with a reasonable rate of rejecting the null hypothesis.

- Conclusion: The testing method in Graph 2 has a more appropriate rejection rate, suggesting better control over Type I error.

Overall Conclusion:

Based on the bias, variance, MSE, fraction of rejections, and the shape of the graphs, we can conclude that the testing method in Graph 2 is more effective than the testing method in Graph 1. The testing method in Graph 2 has higher stability, is less affected by outliers, has a smaller sum of squared errors, and exhibits a more balanced rejection rate of the null hypothesis.

1. Same as #3, but with 50 observations and test 𝐻0: 𝑝=0.6. What differences do you notice?

****

1. Comparing Bias:

- Graph 3: Bias = -0.0015, indicating that the testing method has a very small deviation from the actual value.

- Graph 5: Bias = -0.0036, also indicating a small deviation, but smaller than in Graph 3.

- Conclusion: The testing method in Graph 5 has higher accuracy in terms of bias.

2. Comparing Variance:

- Graph 3: Variance = 0.0245, indicating a relatively large dispersion of the testing results.

- Graph 5: Variance = 0.0047, indicating a smaller dispersion of the testing results, meaning the testing results are more stable.

- Conclusion: The testing method in Graph 5 has higher stability in terms of variance.

3. Comparing Mean Square Error (MSE):

- Graph 3: MSE = 0.0245, indicating a relatively large sum of squared errors for the testing method.

- Graph 5: MSE = 0.0047, indicating a smaller sum of squared errors for the testing method, meaning the testing method is more efficient.

- Conclusion: The testing method in Graph 5 is more efficient in terms of MSE.

4. Comparing Fraction of Times the Null Hypothesis was Rejected:

- Graph 3: Fraction of rejections = 0.108, indicating that the null hypothesis was rejected in only 10.8% of the simulations. This suggests that the method is overly conservative and may be failing to reject the null hypothesis when it should be rejected.

- Graph 5: Fraction of rejections = 0.042, indicating that the null hypothesis was rejected in only 4.2% of the simulations. This suggests that the method is even more conservative than Graph 3, with a lower rate of rejecting the null hypothesis.

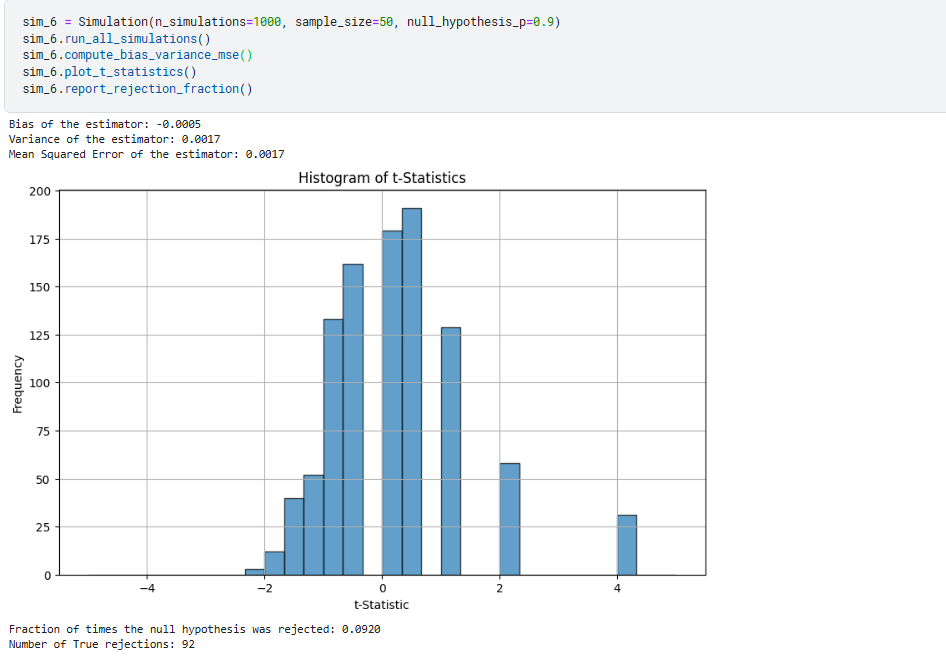
- Conclusion: The testing method in Graph 3 appears to be overly conservative, potentially leading to Type II errors (failing to reject a false null hypothesis). The testing method in Graph 5 is even more conservative than Graph 3, suggesting a higher risk of Type II errors.

5. Analyzing the shape of the graph:

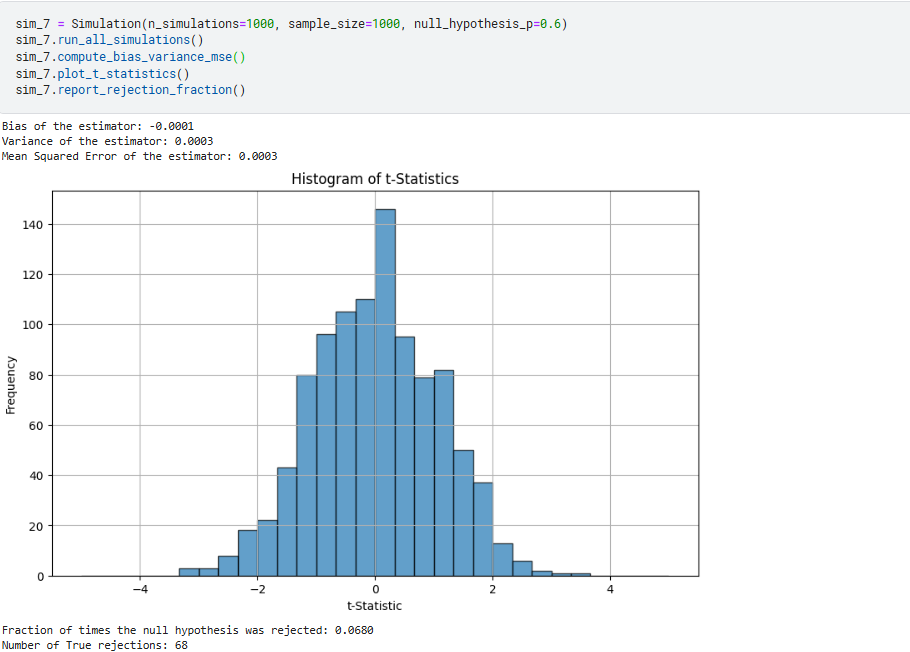
- Graph 3: The shape of the graph shows a slight skew in the distribution of the t-statistic, indicating that the testing method may be affected by outliers.

- Graph 5: The shape of the graph shows a distribution of the t-statistic closer to the normal distribution, indicating that the testing method is more stable and less affected by outliers.

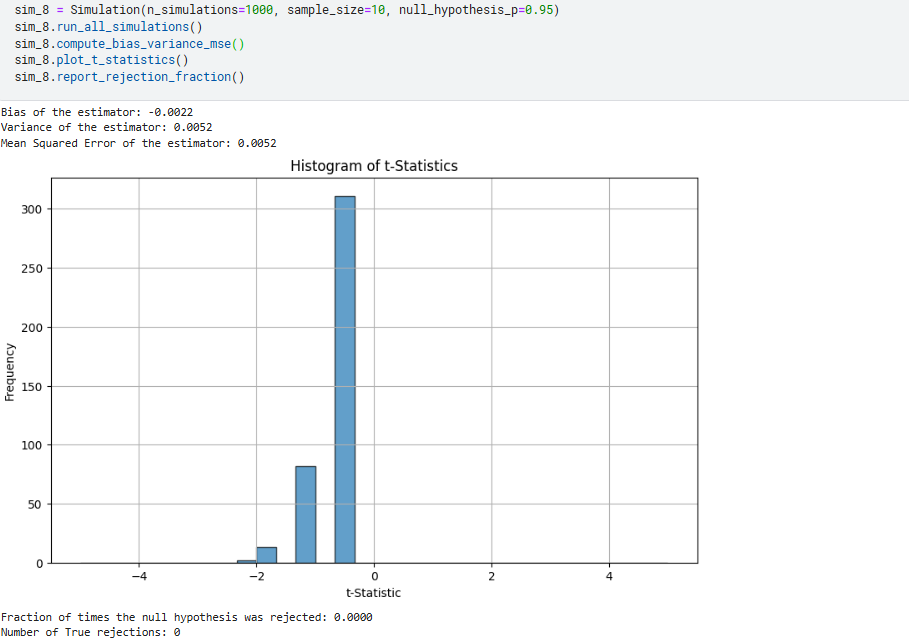
1. Same as #3, but with 50 observations and test 𝐻0: 𝑝=0.9. What differences do you notice?



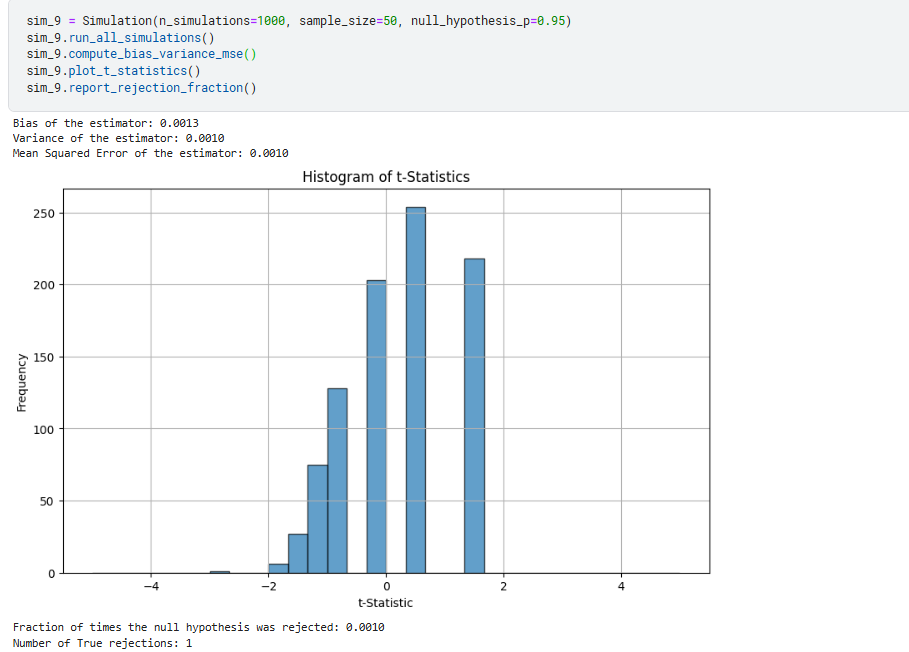
1. Comparing Bias:
   * Graph 3: Bias = -0.0015, indicating that the testing method has a very small deviation from the actual value.
   * Graph 6: Bias = -0.0005, also indicating a small deviation, but larger than in Graph 3.
   * Conclusion: The testing method in Graph 3 has higher accuracy in terms of bias.
2. Comparing Variance:
   * Graph 3: Variance = 0.0245, indicating a relatively large dispersion of the testing results.
   * Graph 6: Variance = 0.0017, indicating a smaller dispersion of the testing results, meaning the testing results are more stable.
   * Conclusion: The testing method in Graph 6 has higher stability in terms of variance.
3. Comparing Mean Square Error (MSE):
   * Graph 3: MSE = 0.0245, indicating a relatively large sum of squared errors for the testing method.
   * Graph 6: MSE = 0.0017, indicating a smaller sum of squared errors for the testing method, meaning the testing method is more efficient.
   * Conclusion: The testing method in Graph 6 is more efficient in terms of MSE.
4. Comparing Fraction of Times the Null Hypothesis was Rejected:
   * Graph 3: Fraction of rejections = 0.108, indicating that the null hypothesis was rejected in only 10.8% of the simulations. This suggests that the method is overly conservative and may be failing to reject the null hypothesis when it should be rejected.
   * Graph 6: Fraction of rejections = 0.092, indicating that the null hypothesis was rejected in only 9.2% of the simulations. This suggests that the method is even more conservative than Graph 3, with a lower rate of rejecting the null hypothesis.
   * Conclusion: The testing method in Graph 3 appears to be overly conservative, potentially leading to Type II errors (failing to reject a false null hypothesis). The testing method in Graph 6 is even more conservative than Graph 3, suggesting a higher risk of Type II errors.
5. Same as Scenario 3, but with 1000 observations and test 𝐻0: 𝑝 =0.6. What differences do you notice?



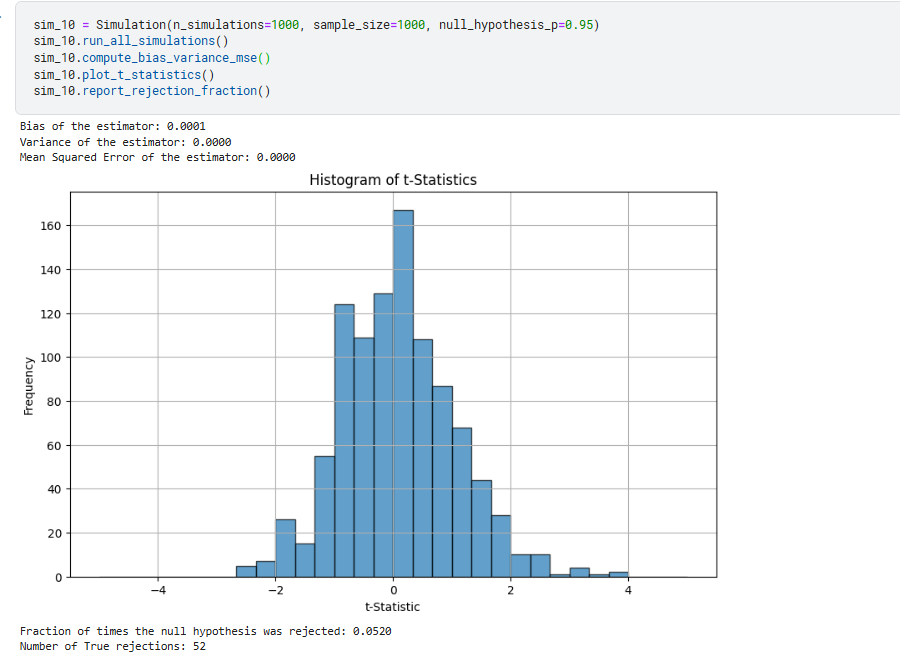
1. Comparing Bias:
   * Graph 3: Bias = -0.0015, indicating that the testing method has a very small deviation from the actual value.
   * Graph 7: Bias = -0.0001, also indicating a small deviation, but larger than in Graph 3.
   * Conclusion: The testing method in Graph 3 has higher accuracy in terms of bias.
2. Comparing Variance:
   * Graph 3: Variance = 0.0245, indicating a relatively large dispersion of the testing results.
   * Graph 7: Variance = 0.0003, indicating a smaller dispersion of the testing results, meaning the testing results are more stable.
   * Conclusion: The testing method in Graph 7 has higher stability in terms of variance.
3. Comparing Mean Square Error (MSE):
   * Graph 3: MSE = 0.0245, indicating a relatively large sum of squared errors for the testing method.
   * Graph 7: MSE = 0.0003, indicating a smaller sum of squared errors for the testing method, meaning the testing method is more efficient.
   * Conclusion: The testing method in Graph 7 is more efficient in terms of MSE.
4. Comparing Fraction of Times the Null Hypothesis was Rejected:
   * Graph 3: Fraction of rejections = 0.108, indicating that the null hypothesis was rejected in only 10.8% of the simulations. This suggests that the method is overly conservative and may be failing to reject the null hypothesis when it should be rejected.
   * Graph 7: Fraction of rejections = 0.068, indicating that the null hypothesis was rejected in only 6.8% of the simulations. This suggests that the method is even more conservative than Graph 3, with a lower rate of rejecting the null hypothesis.
   * Conclusion: The testing method in Graph 3 appears to be overly conservative, potentially leading to Type II errors (failing to reject a false null hypothesis). The testing method in Graph 7 is even more conservative than Graph 3, suggesting a higher risk of Type II errors.
5. Same as #3, but now set 𝑝=0.95 (so that this is an unfair coin that flips heads 95% of the time) and with 10 observations and test 𝐻0: 𝑝=0.95. What differences do you notice?



1. Comparing Bias:
   * Graph 3: Bias = -0.0015, indicating that the testing method has a very small deviation from the actual value.
   * Graph 8: Bias = -0.0022, also indicating a small deviation, but smaller than in Graph 3.
   * Conclusion: The testing method in Graph 8 has higher accuracy in terms of bias.
2. Comparing Variance:
   * Graph 3: Variance = 0.0245, indicating a relatively large dispersion of the testing results.
   * Graph 8: Variance = 0.0052, indicating a smaller dispersion of the testing results, meaning the testing results are more stable.
   * Conclusion: The testing method in Graph 8 has higher stability in terms of variance.
3. Comparing Mean Square Error (MSE):
   * Graph 3: MSE = 0.0245, indicating a relatively large sum of squared errors for the testing method.
   * Graph 8: MSE = 0.0052, indicating a smaller sum of squared errors for the testing method, meaning the testing method is more efficient.
   * Conclusion: The testing method in Graph 8 is more efficient in terms of MSE.
4. Comparing Fraction of Times the Null Hypothesis was Rejected:
   * Graph 3: Fraction of rejections = 0.108, indicating that the null hypothesis was rejected in only 10.8% of the simulations. This suggests that the method is overly conservative and may be failing to reject the null hypothesis when it should be rejected.
   * Graph 8: Fraction of rejections = 0.068, indicating that the null hypothesis was rejected in only 6.8% of the simulations. This suggests that the method is even more conservative than Graph 3, with a lower rate of rejecting the null hypothesis.
   * Conclusion: The testing method in Graph 3 appears to be overly conservative, potentially leading to Type II errors (failing to reject a false null hypothesis). The testing method in Graph 8 is even more conservative than Graph 3, suggesting a higher risk of Type II errors.
5. Same as #8, but with 50 observations. What differences do you notice?



1. Comparing Bias:
   * Graph 8: Bias = -0.0022, indicating that the testing method has a very small deviation from the actual value.
   * Graph 9: Bias = 0.0013, also indicating a small deviation, but higher than in Graph 8.
   * Conclusion: The testing method in Graph 8 has higher accuracy in terms of bias.
2. Comparing Variance:
   * Graph 8: Variance = 0.0245, indicating a relatively large dispersion of the testing results.
   * Graph 9: Variance = 0.001, indicating a smaller dispersion of the testing results, meaning the testing results are more stable.
   * Conclusion: The testing method in Graph 9 has higher stability in terms of variance.
3. Comparing Mean Square Error (MSE):
   * Graph 8: MSE = 0.0245, indicating a relatively large sum of squared errors for the testing method.
   * Graph 9: MSE = 0.001, indicating a smaller sum of squared errors for the testing method, meaning the testing method is more efficient.
   * Conclusion: The testing method in Graph 9 is more efficient in terms of MSE.
4. Comparing Fraction of Times the Null Hypothesis was Rejected:
   * Graph 8: Fraction of rejections = 0.108, indicating that the null hypothesis was rejected in only 10.8% of the simulations. This suggests that the method is overly conservative and may be failing to reject the null hypothesis when it should be rejected.
   * Graph 9: Fraction of rejections = 0.001, indicating that the null hypothesis was rejected in only 0.1% of the simulations. This suggests that the method is even more conservative than Graph 8, with a lower rate of rejecting the null hypothesis.
   * Conclusion: The testing method in Graph 8 appears to be overly conservative, potentially leading to Type II errors (failing to reject a false null hypothesis). The testing method in Graph 9 is even more conservative than Graph 8, suggesting a higher risk of Type II errors.
5. Same as #8, but with 1000 observations. What differences do you notice?



1. Comparing Bias:
   * Graph 8: Bias = -0.0022, indicating that the testing method has a very small deviation from the actual value.
   * Graph 10: Bias = 0.0001, also indicating a small deviation, but higher than in Graph 8.
   * Conclusion: The testing method in Graph 8 has higher accuracy in terms of bias.
2. Comparing Variance:
   * Graph 8: Variance = 0.0245, indicating a relatively large dispersion of the testing results.
   * Graph 10: Variance = 0, indicating a smaller dispersion of the testing results, meaning the testing results are more stable.
   * Conclusion: The testing method in Graph 10 has higher stability in terms of variance.
3. Comparing Mean Square Error (MSE):
   * Graph 8: MSE = 0.0245, indicating a relatively large sum of squared errors for the testing method.
   * Graph 10: MSE = 0, indicating a smaller sum of squared errors for the testing method, meaning the testing method is more efficient.
   * Conclusion: The testing method in Graph 10 is more efficient in terms of MSE.
4. Comparing Fraction of Times the Null Hypothesis was Rejected:
   * Graph 8: Fraction of rejections = 0.108, indicating that the null hypothesis was rejected in only 10.8% of the simulations. This suggests that the method is overly conservative and may be failing to reject the null hypothesis when it should be rejected.
   * Graph 10: Fraction of rejections = 0.052, indicating that the null hypothesis was rejected in only 5.2% of the simulations. This suggests that the method is even more conservative than Graph 8, with a lower rate of rejecting the null hypothesis.
   * Conclusion: The testing method in Graph 8 appears to be overly conservative, potentially leading to Type II errors (failing to reject a false null hypothesis). The testing method in Graph 10 is even more conservative than Graph 8, suggesting a higher risk of Type II errors.

**Hint:**

* Problem 1: Coin flips follow the binomial distribution: <https://www.w3schools.com/python/numpy/numpy_random_binomial.asp>
* Problem 2: Use the Monte Carlo method to simulate the binomial distribution.
* Since problems 3-10 ask you to do roughly the same thing over and over, it is probably useful to try to write a function to do all of these but with arguments that allow you to change the number of observations per simulation, the true value of 𝑝, and the null hypothesis that you are testing.