# Lecture 30 The Behaviour of The p-value

BIO210 Biostatistics

Xi Chen

Fall, 2025

School of Life Sciences
Southern University of Science and Technology

南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

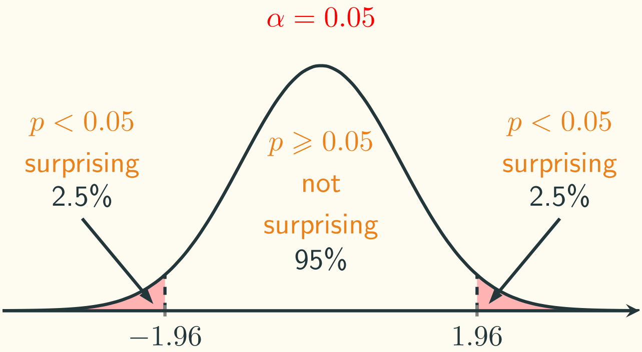
# Why p-values Are Successful In Science

In some sense it offers a first line of defense against being fooled by randomness, separating signal from noise.

– Benjamini, 2016

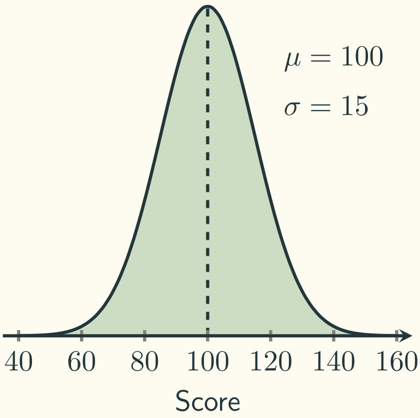
# Why p-values Are Successful In Science

- $p$-value $= \mathbb{P}($ observed data or more extreme $\mid H_0$ is true$)$: How surprising the data is, assuming there is no effect?
- $p$-value calculation: using the distribution of the test statistics, which is based on the sampling distribution.

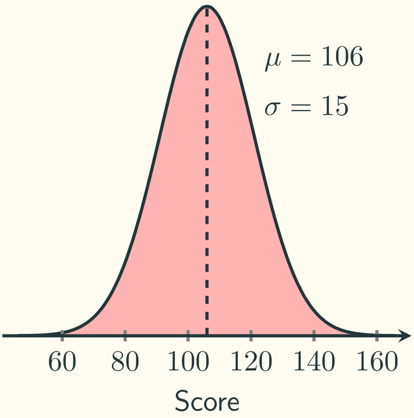


# Test Scores In Math Exams - Two Populations

Students ($\leqslant 5$ hours per week)




Students ($> 5$ hours per week)



# Test Scores In Math Exams - Hypothesis Testing

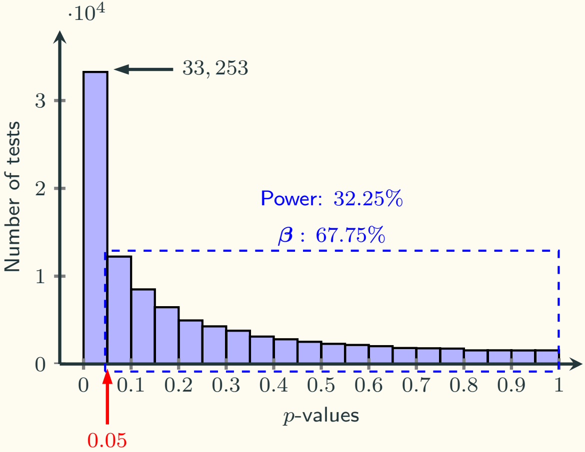
**Question:** are the mean scores the same between students who study $\leqslant 5$ hours a week and those who study $> 5$ hours a week?

# Simulation Setup

```
import numpy as np
from scipy.stats import norm
from scipy.stats import ttest_ind as tt

np.random.seed(42)                        # set seed for reproducible results
pop1 = norm(loc=100, scale=15)            # set up population 1 (study hours < 5)
pop2 = norm(loc=106, scale=15)            # set up population 2 (study hours > 5)
```

# Two-Sample $t$-tests

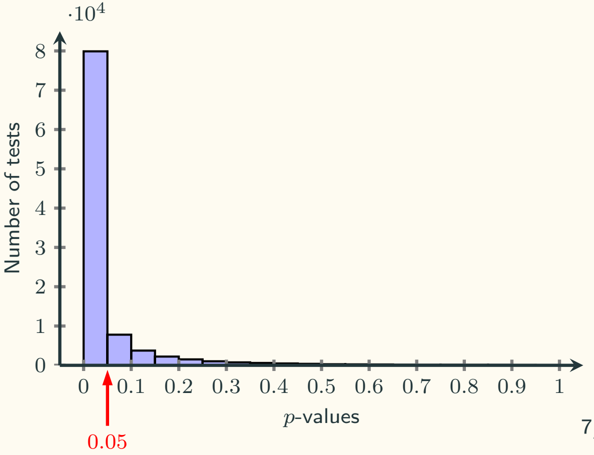
```
pvals = np.zeros((100000,))

for i in range(100000):
    s1 = pop1.rvs(size=30)
    s2 = pop2.rvs(size=30)
    ts, p = tt(s1, s2, equal_var = True)
    pvals[i] = p
```



# Distribution of $p$-values

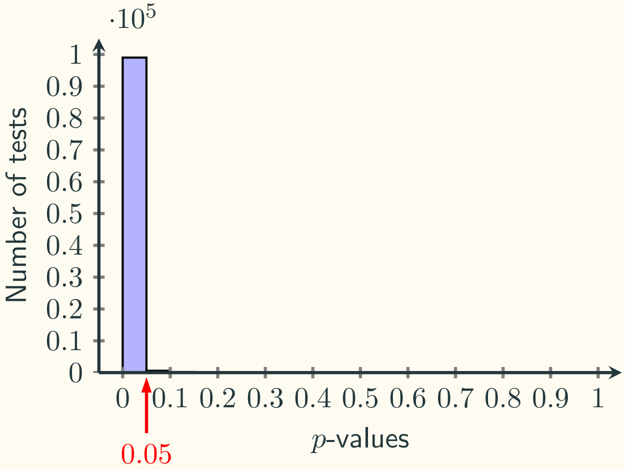
We want to increase our power to 80%: $n = \left[\dfrac{(1.96 + 0.842) \times \sqrt{450}}{106 - 100}\right]^2 = 99$

```
pvals = np.zeros((100000,))

for i in range(100000):
    s1 = pop1.rvs(size=99)
    s2 = pop2.rvs(size=99)
    ts, p = tt(s1, s2, equal_var = True)
    pvals[i] = p
```

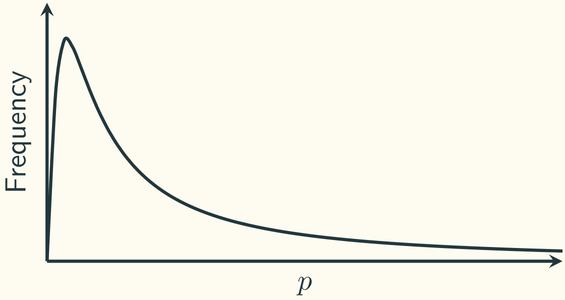


# Distribution of $p$-values

Sample size: $n = 231$



# Distribution of $p$-values When $H_1$ Is True

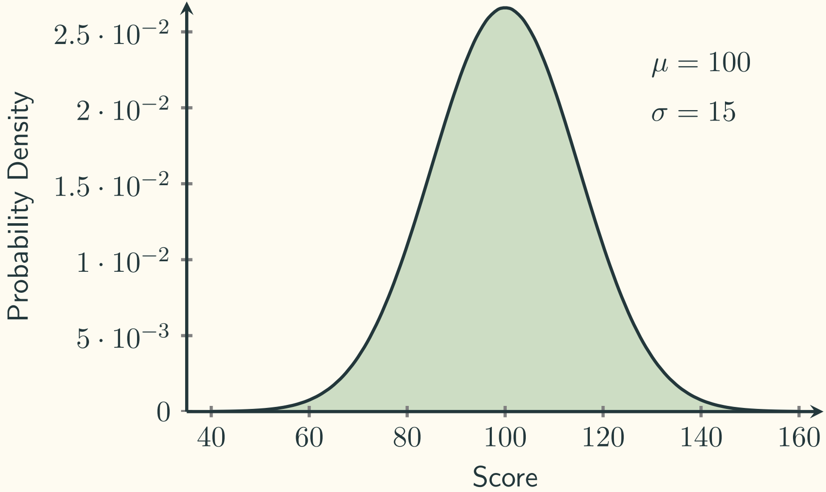



When $H_1$ is true, the distribution of $p$-values are skewed to the right, and the shape depends on the power.

What is the distribution of $p$-values when $H_0$ is true ?

# Test Scores In Math Exams - Drinking Coke

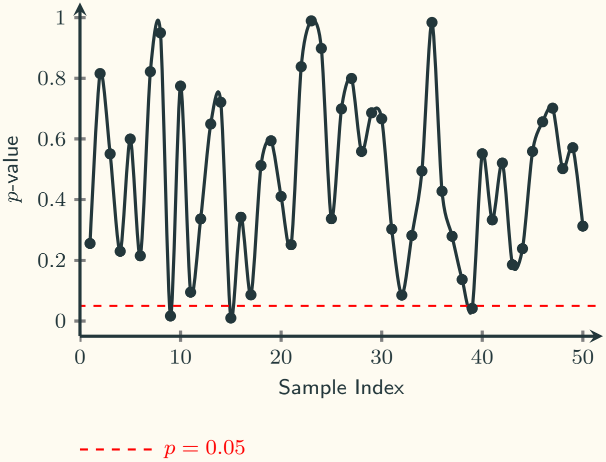
Students ($\geqslant 355$ mL Coke per week)



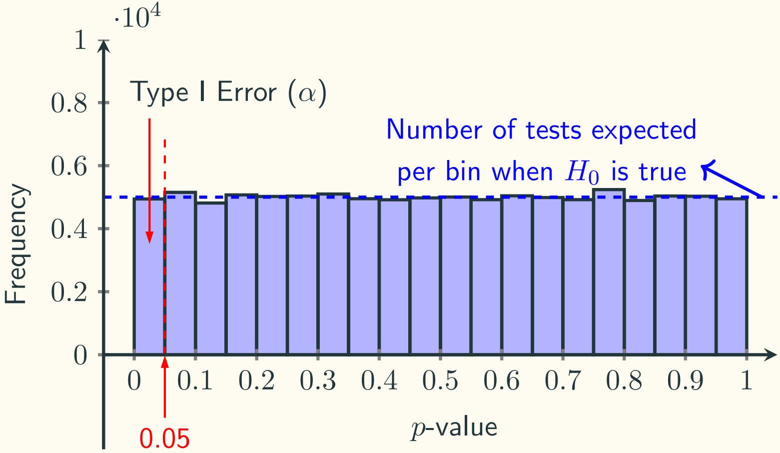

Take samples of sizes $n = 100$

# $p$-value Fluctuation

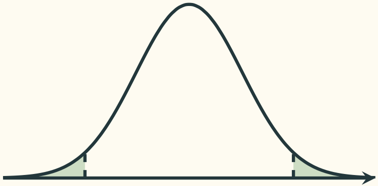
```python
pvals = np.zeros((100000,))

for i in range(100000):
    s1 = pop1.rvs(size=100)
    s2 = pop1.rvs(size=100)
    ts, p = tt(s1, s2, equal_var = True)
    pvals[i] = p
```



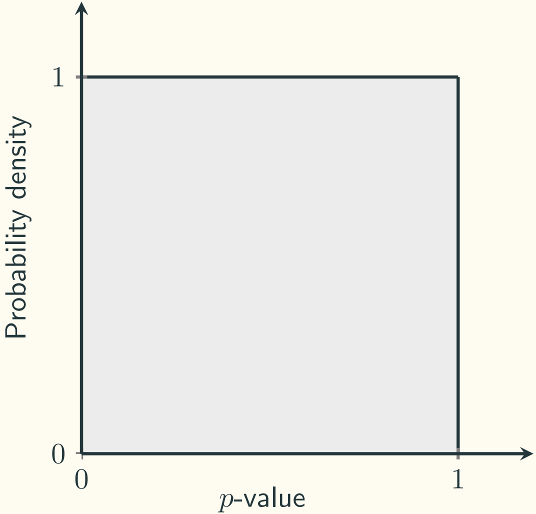
# Distribution of $p$-values When $H_0$ is true

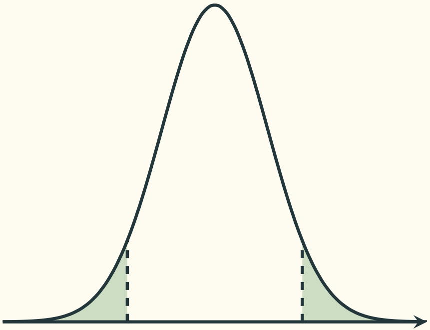



$p$-value: $\mathbb{P}\,(\text{data} \mid H_0 \text{ is true})$

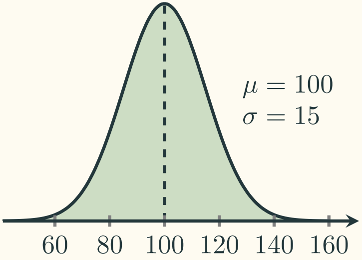


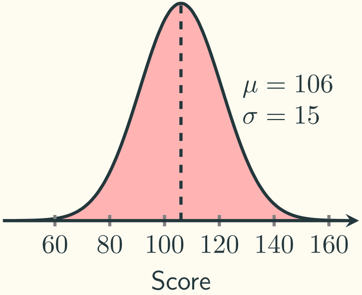
# $p$-values Are Uniformly Distributed When $H_0$ Is True



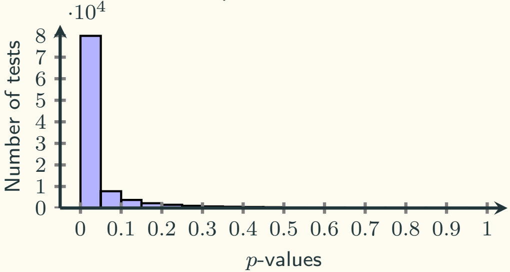



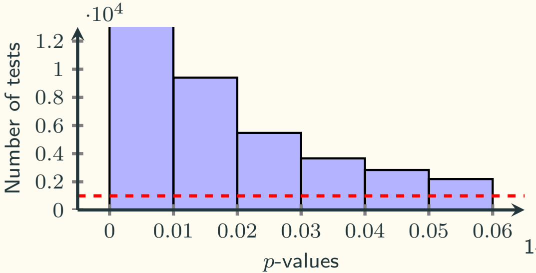
# More $p$-value Distribution When $H_1$ Is True



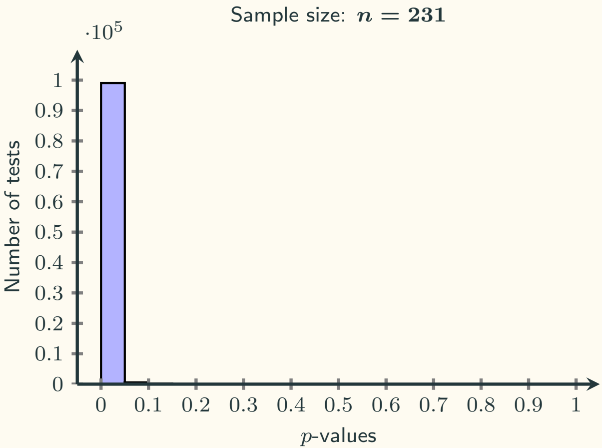

Sample Size: $\boldsymbol{n = 99}$



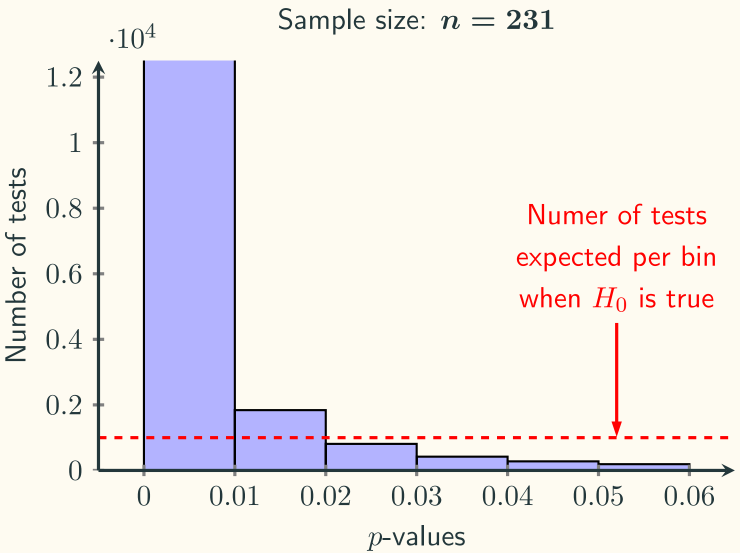



# Interpreting $p$-value When The Power Is High

```
pvals = np.zeros((100000,))

for i in range(100000):
    s1 = pop1.rvs(size=231)
    s2 = pop2.rvs(size=231)
    ts, p = tt(s1, s2, equal_var = True)
    pvals[i] = p
```



# Interpreting $p$-value When The Power Is High




- **Question:** In this case, if you get a $p = 0.045$ or $p = 0.035$, which one is more likely to be true? $H_0$ or $H_1$?

# Lindley's Paradox (1957)

- In the simulations, we know $H_0$ is true or not, but in the real world, we don't know. When we have very high power, use an $\alpha$ level of $0.05$, and find a $p$-value of $p = 0.045$, the data is surprising, assuming the null hypothesis $H_0$ is true, but it is even more surprising, assuming the alternative hypothesis $H_1$ is true. This shows how a significant $p$-value is not always evidence for the alternative hypothesis.

- A result can be unlikely when the null hypothesis is true, but it can be even more unlikely assuming the alternative hypothesis is true, and power is very high. For this reason, some researchers have suggested using lower $\alpha$ levels in very large sample sizes, and this is probably sensible advice. Other researchers have suggested using Bayesian statistics, which is also sensible advice.