Lecture 30 The Behaviour of The p-value

BIO210 Biostatistics

Xi Chen

Spring, 2022

School of Life Sciences
Southern University of Science and Technology

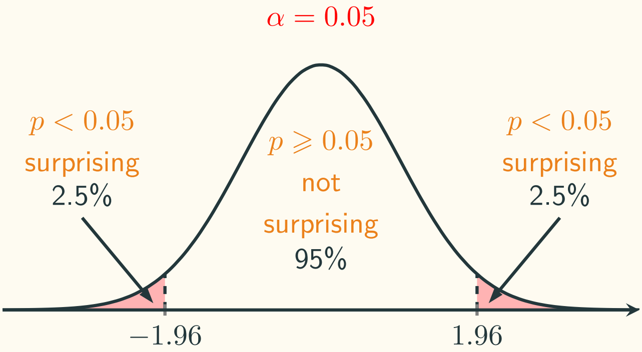
南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

Why p-values Are Successful In Science

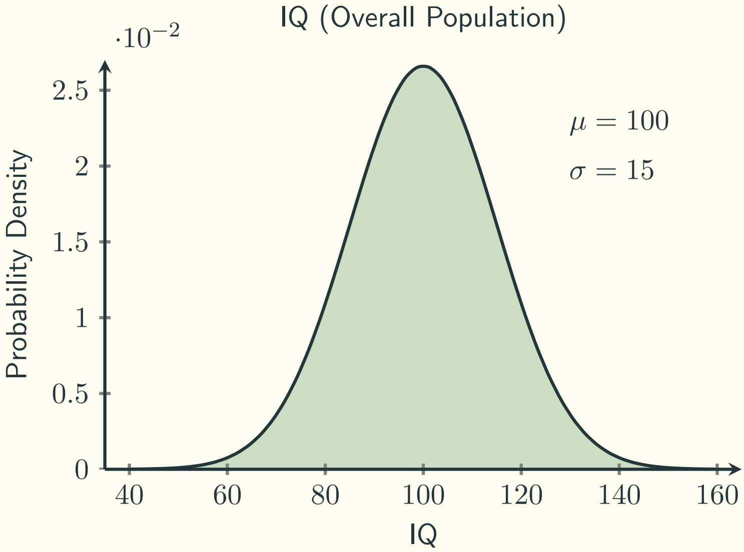
> In some sense it offers a first line of defense against being fooled by randomness, separating signal from noise.
>
> – Benjamini, 2016

Why p-values Are Successful In Science

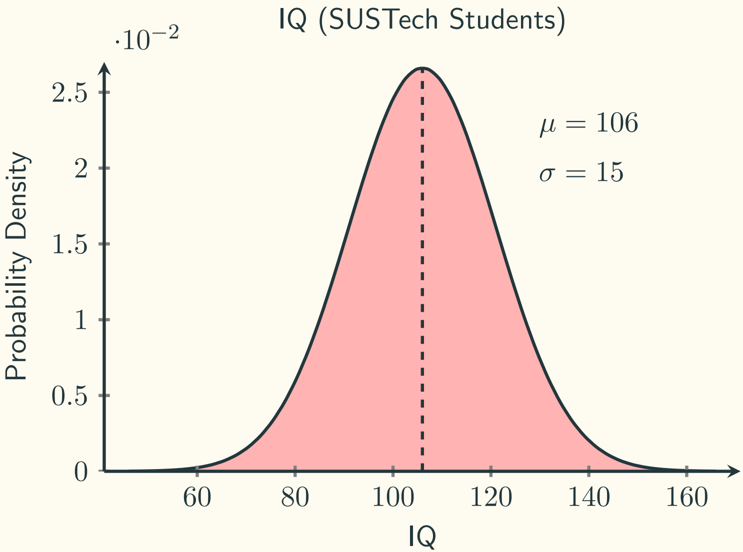
- p-value $= P($ observed data or more extreme $\mid H_0$ is true$)$: How surprising the data is, assuming there is no effect?
- p-value calculation: using the distribution of the test statistics, which is based on the sampling distribution.



IQ Distribution



IQ Distribution



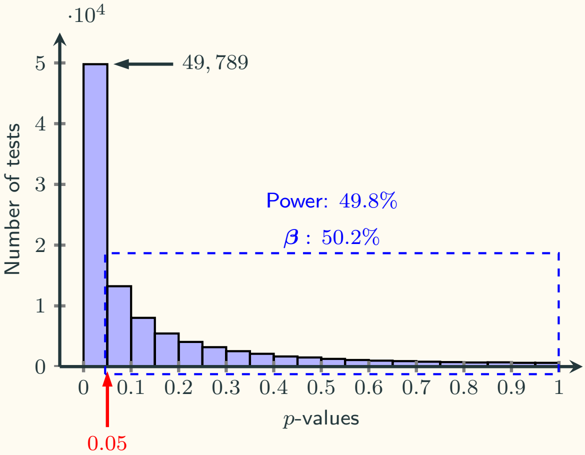
Take a sample of size $n = 26$

We ask:

Is $\mu = 100$?

One Sample t-test

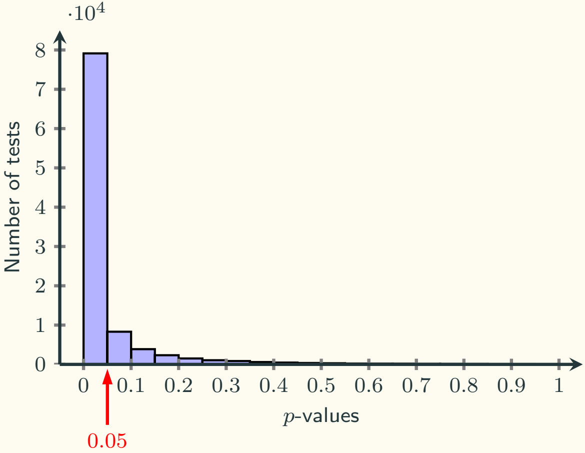
```python
import numpy as np
from scipy.stats import ttest_1samp as tt

np.random.seed(42)

pvals = np.zeros((100000,))

for i in range(100000):
    sample = np.random.normal(loc=106,
                              scale=15,
                              size=26)
    ts, p = tt(sample, popmean=100)
    pvals[i] = p
```



Distribution of p-values

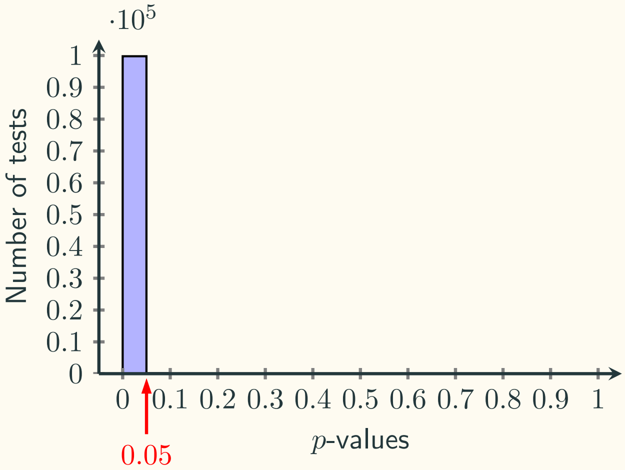
We want to increase our power to 80%: $n = \left[\frac{(1.96 + 0.842) \times 15}{106 - 100}\right]^2 = 50$

```python
import numpy as np
from scipy.stats import ttest_1samp as tt

np.random.seed(42)

pvals = np.zeros((100000,))

for i in range(100000):
    sample = np.random.normal(loc=106,
                              scale=15,
                              size=50)
    ts, p = tt(sample, popmean=100)
    pvals[i] = p
```

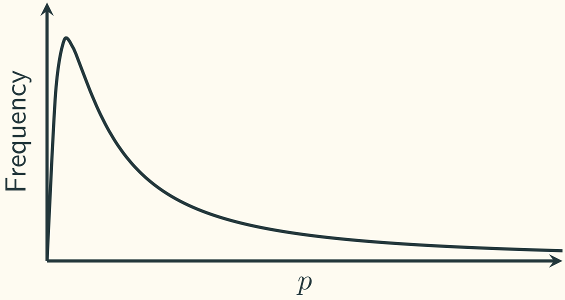


Distribution of p-values

Sample size: $n = 144$



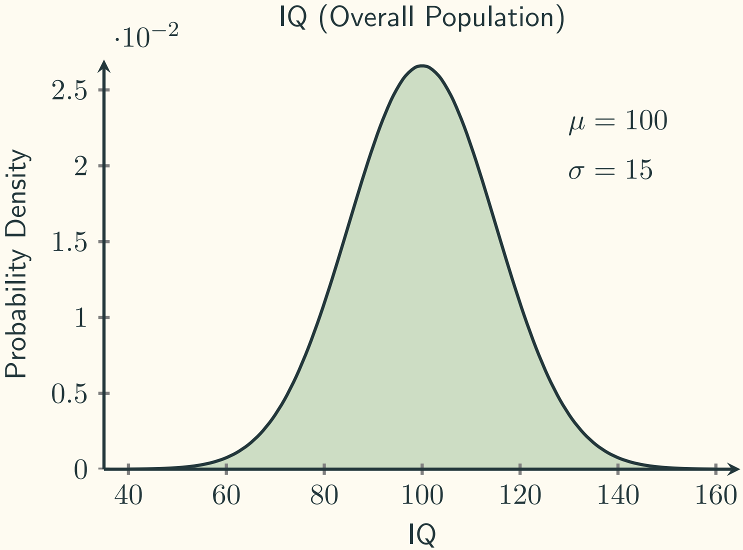
Distribution of p-values When H_1 Is True



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 ?

IQ Distribution



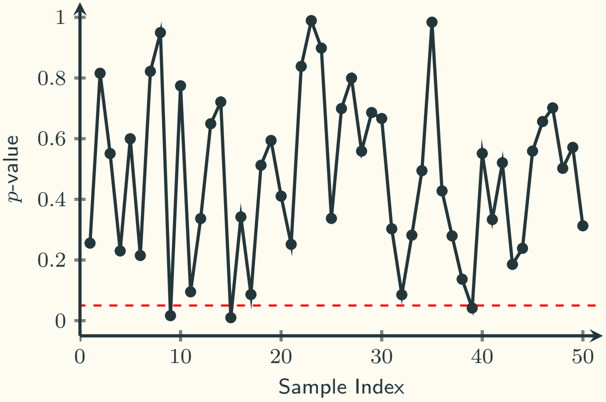
Take samples of size $n = 100$

We ask:

Is $\mu = 100$?

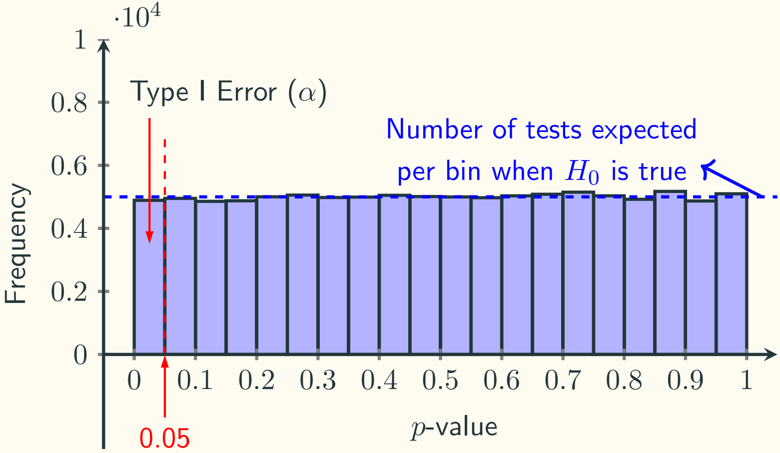
p-value Fluctuation

```python
import numpy as np
from scipy.stats import ttest_1samp as tt

np.random.seed(42)

pvals = np.zeros((100000,))

for i in range(100000):
    sample = np.random.normal(loc=100,
                              scale=15,
                              size=100)
    ts, p = tt(sample, popmean=100)
    pvals[i] = p
```

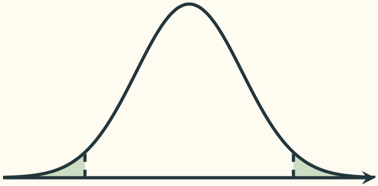


$p = 0.05$

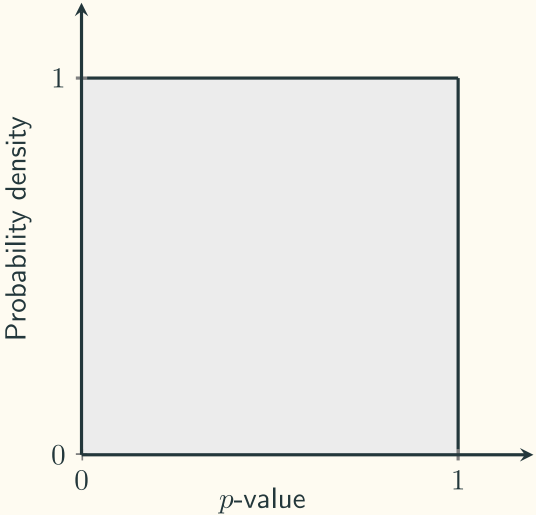
Distribution of p-values When H_0 is true

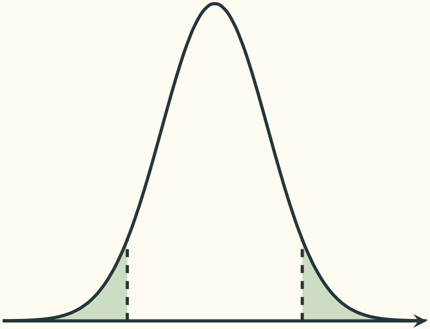


p-value: $P(\text{data} \mid H_0 \text{ is true})$

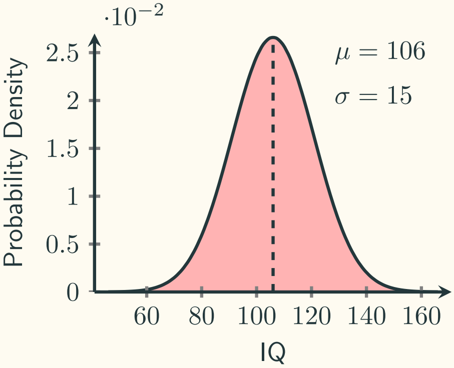


p-values Are Uniformly Distributed When H_0 Is True



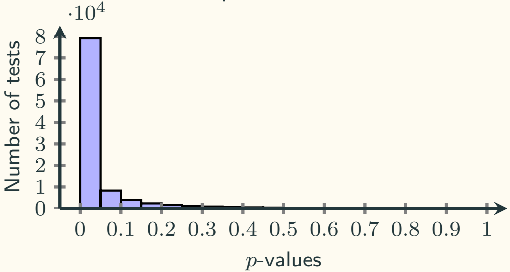


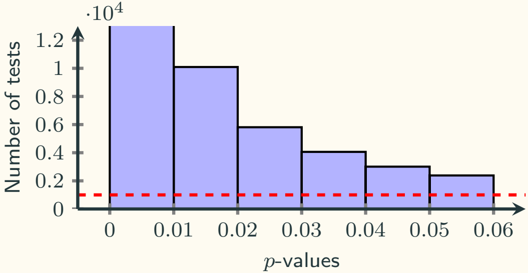
More p-value Distribution When H_1 Is True



Take samples and ask: Is $\mu = 100$?

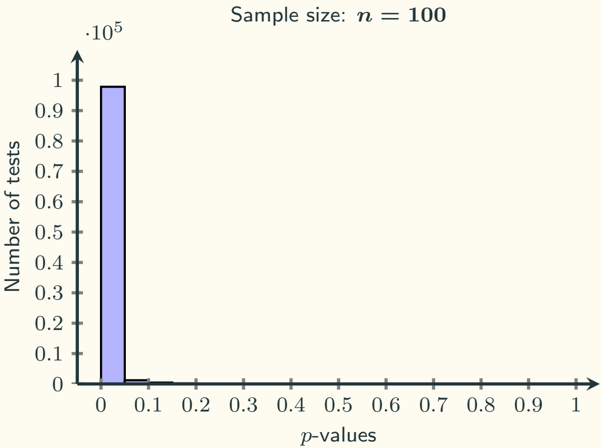
Sample Size: $\boldsymbol{n = 50}$



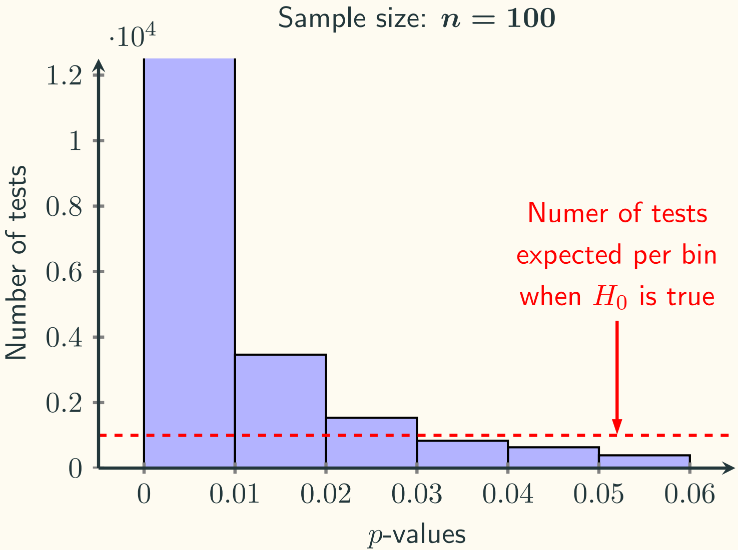


Interpreting p-value When The Power Is High

```python
import numpy as np
from scipy.stats import ttest_1samp as tt

np.random.seed(42)

pvals = np.zeros((100000,))

for i in range(100000):
    sample = np.random.normal(loc=106,
                              scale=15,
                              size=100)
    ts, p = tt(sample, popmean=100)
    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 α 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 α levels in very large sample sizes, and this is probably sensible advice. Other researchers have suggested using Bayesian statistics, which is also sensible advice.