Hypothesis Testing

Hypothesis Testing

- **Hypothesis Testing** is the application of statistical methods to real-world questions.
- We start with an assumption, called the **null hypothesis**
- We run an experiment to test this null hypothesis

Hypothesis Testing

- Based on the results of the experiment, we either **reject** or **fail to reject** the null hypothesis
- If the null hypothesis is rejected, then we say the data supports another, mutually exclusive **alternate hypothesis**
- We never “PROVE” a hypothesis!

Framing the Hypothesis

- How do we frame the question that forms our null hypothesis?
- At the start of the experiment, the null hypothesis is assumed to be true.
- If the data fails to support the null hypothesis, only then can we look to an alternative hypothesis

Framing the Hypothesis

If testing something assumed to be true,
the null hypothesis can reflect the assumption:

Claim: *"Our product has an average shipping weight of 3.5kg."*

Null hypothesis: average weight = 3.5kg

Alternate hypothesis: average weight ≠ 3.5kg

Framing the Hypothesis

The null hypothesis should contain an equality (=, ≤ ,≥):

average shipping weight = 3.5kg $\boxed{H_0\colon\ \mu = 3.5}$

The alternate hypothesis should not have an equality (≠,<,>):

average shipping weight ≠ 3.5kg $\boxed{H_1\colon\ \mu \neq 3.5}$

Hypothesis Testing

- So what lets us reject or fail to reject the null hypothesis?

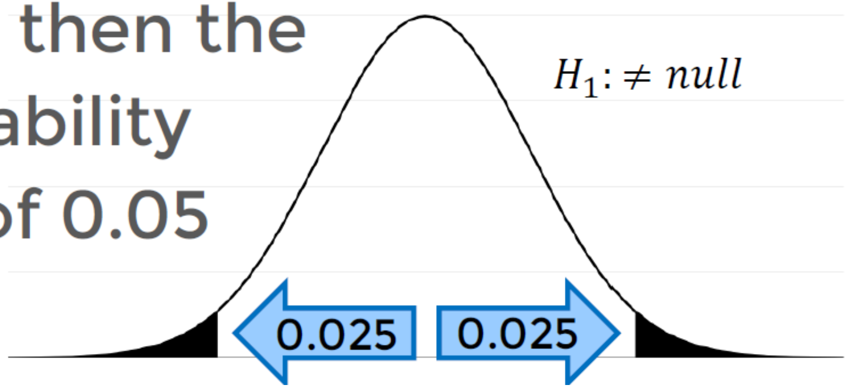
Hypothesis Testing

- We run an experiment and record the result.
- **Assuming our null hypothesis is valid,** if the probability of observing these results is very small (inside of 0.05) then we reject the null hypothesis.
- Here 0.05 is our **level of significance**

$$\alpha = 0.05$$

Hypothesis Testing - Tails

- The level of significance α is the area inside the *tail(s)* of our null hypothesis.
- If $\alpha = 0.05$ and the alternative hypothesis is *not equal to* the null, then the two tails of our probability curve *share* an area of 0.05



Hypothesis Testing – P-value Test

In a **traditional test**:

- take the level of significance
- use it to determine the critical value
- compare the test statistic to the critical value

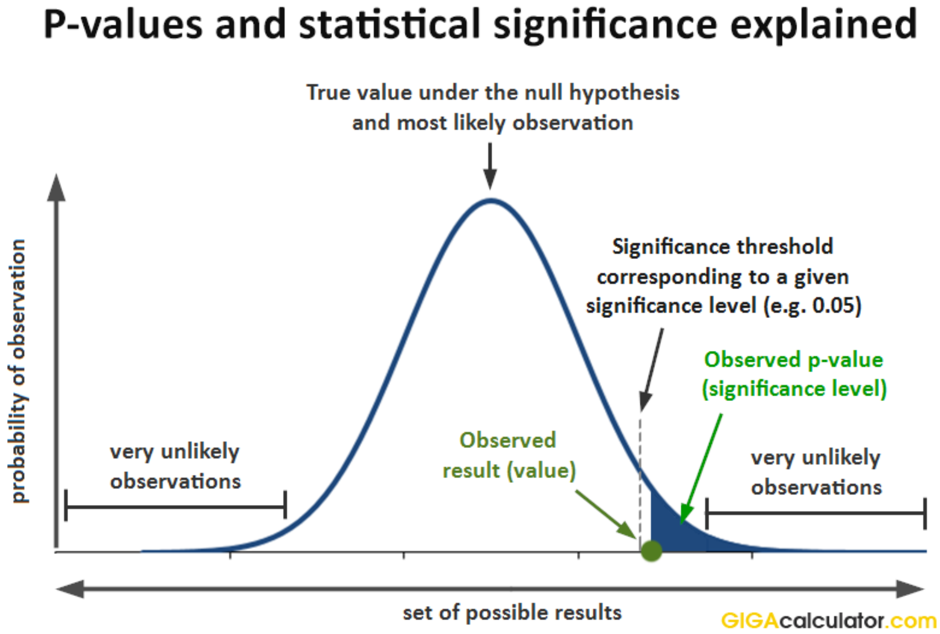
In a **P-value test**:

- take the test statistic
- use it to determine the P-value
- compare the P-value to the level of significance α

Hypothesis Testing – P-value Test

“If the P-value is low, the null must go!”
reject $H0$

“If the P-value is high, the null must fly!”
fail to reject $H0$



Extra: https://userpage.fu-berlin.de/soga/200/2070_hypothesis_tests/20713_The_Critical_Value_and_the_p-Value_Approach_to_Hypothesis_Testing.html

Testing Exercise:

$$\mu = 3.125$$
$$\sigma = 0.700$$

- A company is looking to improve their website performance.
- Currently pages have a mean load time of 3.125 seconds, with a standard deviation of 0.700 seconds.
- They hire a consulting firm to improve load times.

- Management wants a 99% confidence level
- A sample run of 40 of the new pages has a mean load time of 2.875 seconds.
- Are these results statistically faster than before?

$$\boxed{\begin{gathered}\mu = 3.125\\ \sigma = 0.700\\ \alpha = 0.01\\ n = 40\\ \bar{x} = 2.875\end{gathered}}$$

1. State the null hypothesis:

$$H_0\colon \mu \geq 3.125$$

2. State the alternative hypothesis:

$$H_1\colon \mu < 3.125$$

3. Set a level of significance:

$$\alpha = 0.01$$

4. Determine the test type:

left tail | right tail | two tail

TRADITIONAL METHOD:

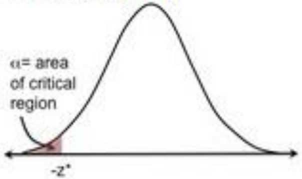
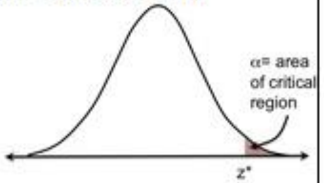
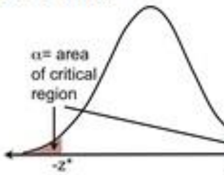
5. Test Statistic:

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{2.875 - 3.125}{0.7/\sqrt{40}} = -2.259$$

$$\mu = 3.125$$
$$\sigma = 0.700$$
$$\alpha = 0.01$$
$$n = 40$$
$$\bar{x} = 2.875$$

STANDARD NORMAL DISTRIBUTION: Table Values Represent AREA to the LEFT of the Z score.

Z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00			
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00			
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00			
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00			
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00			
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00			
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00			
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00			
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00			
-3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00			
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00			
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00			
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00			
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00			
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00			
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00			
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00			
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101

α level	One-tailed (left)	One-tailed (right)	Two-tailed
			
α = 0.05	z = -1.64	z = 1.64	z = ± 1.96
α = 0.01	z = -2.33	z = 2.33	z = ± 2.57
α = 0.001	z = -3.08	z = 3.08	z = ± 3.32

6. Critical Value:

z-table lookup on 0.01 $z = -2.325$

$$Z = -2.259$$
$$z = -2.325$$

TRADITIONAL METHOD:

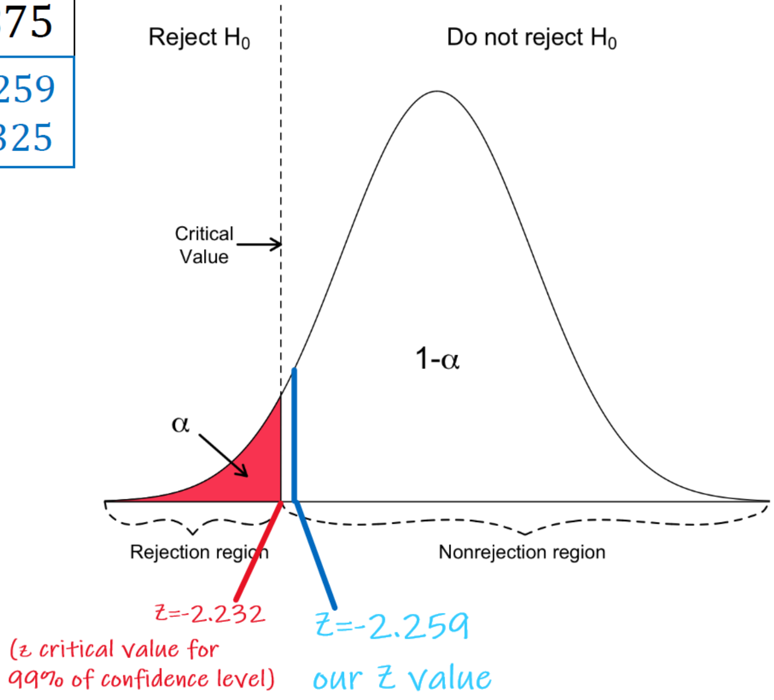
7. Fail to Reject the Null Hypothesis

Since $-2.259 > -2.325$, the test statistic falls outside the rejection region

We can't say that the new web pages are statistically faster.

$\mu = 3.125$
$\sigma = 0.700$
$\alpha = 0.01$
$n = 40$
$\bar{x} = 2.875$

$Z = -2.259$
$z = -2.325$



P-VALUE METHOD:

5. Test Statistic:

$$Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{2.875 - 3.125}{0.7/\sqrt{40}} = -2.259$$

6. P-Value:

z-table lookup on -2.26 $P = 0.0119$

$\mu = 3.125$
$\sigma = 0.700$
$\alpha = 0.01$
$n = 40$
$\bar{x} = 2.875$

$Z = -2.259$
$P = 0.0119$

P-VALUE METHOD:

7. Fail to Reject the Null Hypothesis

Since $0.0119 > 0.01$, the P-value is greater than the level of significance α

We can't say that the new web pages are statistically faster.

$\mu = 3.125$
$\sigma = 0.700$
$\alpha = 0.01$
$n = 40$
$\bar{x} = 2.875$

$\boxed{Z = -2.259}$
$\boxed{P = 0.0119}$

Type I and Type II Errors

- Often in medical fields (and other scientific fields) hypothesis testing is used to test against results where the "truth" is already known.

- For example, testing a new diagnostic test for cancer for patients you have already succesfully diagnosed by other means.

Type I and Type II Errors

- In this situation, you already know if the Null Hypothesis is True or False.
- In these situations where you already know the "truth", then you would know its possible to commit an error with your results .

Type I and Type II Errors

- This type of analysis is common enough that these errors already have specific names:

- Type I Error
- Type II Error

Type I and Type II Errors

- If we **reject** a null hypothesis that should have been supported (Actually **True**),
- we've committed a

  Type I Error

  $H_0: There\ is\ no\ fire$

Pull the fire alarm, only to find out there really was no fire.

Type I and Type II Errors

- If we **fail to reject** a null hypothesis that should have been rejected (actually **False**)
- we've committed a

Type II Error

$\boldsymbol{H_0}: There\ is\ no\ fire$

Don't pull the fire alarm, only to find there really is a fire.

$$H_0: Not\ pregnant$$

$$H_1: Are\ pregnant$$