Topic 5: Hypothesis Testing

Optional Reading: Chapter 9

Xiner Zhou

Department of Statistics

University of California, Davis

- Test about a population mean
- Test about a population proportion
- Test about the difference between two population means
- Test about the difference between two population proportions

Approval?

Suppose you’re a scientist working at one of the big pharmaceutical companies.
Your team have developed a drug
To get FDA approval, you need to show that the drug meet the minimum governmental potency standards.

?

$\mu \leq$ $\mu >$

mini government potency

Statistical hypothesis testing problem

The Reasoning behind Hypothesis Testing = Court Trial

In trying a person for a crime, court must decide between two options
- Innocent?
- Guilty ?

Accused

Assume: innocent

Prosecutor: collect evidence = data

Judge: Is there enough evidence against innocence?

- Yes: court reject the innocence hypothesis → Guilty!
- No: there was not enough evidence to conclude guilty → Not guilty!

Let's first introduce some abstract notions, then see where they fit into the "court trial" process.

The two competing hypotheses:

- Alternative hypothesis H_a
  - The hypothesis that the researcher wishes to support

- Null hypothesis H_0
  - The opposite or contradiction of the alternative hypothesis

There are two possible conclusions

- Reject the null hypothesis H_0 and conclude that the alternative hypothesis H_a is true

- We do not reject the null hypothesis H_0, or equivalently, we accept the null hypothesis H_0

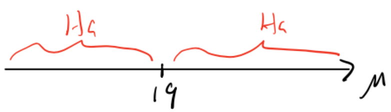
You wish to show that the average hourly wage of carpenters in the state of California is different from \$19, which is the national average. This is the alternative hypothesis, written as

$$H_a : \mu \neq 19$$

The null hypothesis is

$$H_0 : \mu = 19$$

You would like to reject the null hypothesis, thus concluding that the California mean is not equal to \$19.



Two-tailed or two-sided hypothesis:
no directional difference matters to us as indicated in the alternative hypothesis, only care about "not equal"

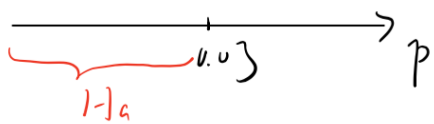
A milling process currently produces an average of 3% defectives. You are interested in showing that a simple adjustment on a machine will decrease p, the proportion of defectives produced in the milling process. Thus, the alternative hypothesis is

$$H_a : p < .03$$

and the null hypothesis is

$$H_0 : p = .03$$

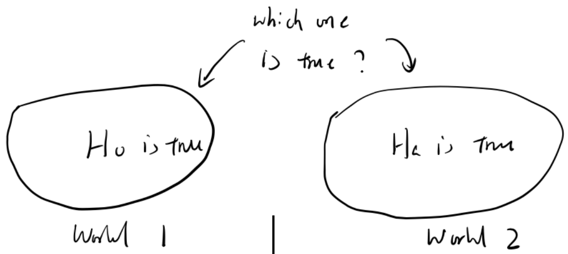
If you can reject H_0, you can conclude that the adjusted process produces fewer than 3% defectives.



One-tailed or one-sided hypothesis:

Directional difference matters to us, we are interested in detecting “falling into one side of a particular value”.

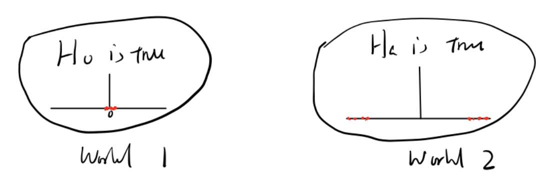
How to make a decision scientifically and objectively ?

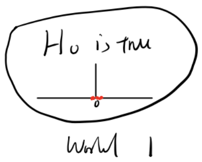


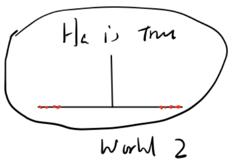
The decision to Reject H_0 / Accept H_0 is based on sample drawn from the population of interest

For each type of problem, we propose a Test Statistic

- is a single number calculated from the sample
- The test statistic is designed such that:
  - If the null hypothesis H_0 is true, it should be close to 0
  - If the alternative hypothesis H_a is true, it should deviate away from 0, i.e. large in absolute value







So,

- A large value of test statistic means: You have strong evidence to support alternative hypothesis H_a
- A small value of test statistic means: You don't have strong evidence to support alternative hypothesis H_a

Now, presented with data, you're pressed to make the final decision.

But, how do you decide a test statsitic just calculated is

Large enough so we should conclude alternative hypothesis is true ?

Small enough so we should conclude null hypothesis is true ?

There are two equivalent approaches:

1. P-value approach
2. Critical value approach

The Reasoning behind Hypothesis Testing = Court Trial

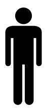
When you have a research question that needs to be verified = you are doing a court trial

In trying **a person** for a crime, **court** must decide between two options

The hypothesis you try to prove

Statistical machinery

- **Innocent?** Null hypothesis H_0 is true
- **Guilty ?** Alternative hypothesis H_a is true



Accused

Assume: innocent

We always start by assuming null hypothesis = the innocent hypothesis

Prosecutor: collect evidence = data

You collect the data

Test statistic & its associated concepts

Judge: Is there enough evidence against innocence?

- Yes: court reject the innocence hypothesis → **Guilty!**

  Conclusion: your alternative hypothesis is accepted !
  What you want to prove is accepted !

- No: there was not enough evidence to conclude guilty → **Not guilty!**

  Conclusion: your null hypothesis is accepted !
  What you want to prove is not accepted !

Let's illustrate our first hypothesis testing through a concrete example.

Average weekly earning for female social workers is $670.

Question: Do men in the same profession have higher salary than women?

Data: a random sample of n=40 male social workers

$\bar{X} = \$725,\ S = \$102$

How to conceptualize this problem?

- What claim or hypothesis or statement we want to prove?
- What is the alternative hypothesis?
- What is the null hypothesis?

Hypothesis Testing about a population mean:

A population has mean μ. We want to test a hypothesis of the form:

$H_0 : \mu = \mu_0$ versus $H_a : \mu > \mu_0$

Some hypothesized value

One-sided hypothesis

Negation of alternative ⬅ Always put what you want to prove or the question in the alternative!

Why $H_0 : \mu = \mu_0$?

First formulate H_a which is what we want to prove, then the opposite to H_a seems to be $H_0 : \mu \leq \mu_0$

In fact: $H_0 : \mu = \mu_0$ is equivalent to $H_0 : \mu \leq \mu_0$

- If H_a is accepted, then it means evidence support $\mu > \mu_0$, then it will certainly reject $\mu < \mu_0$, since it is even more implausible than $\mu = \mu_0$
- If H_a is not accepted, since all we care about is H_a holds or not, we don't care about whether $\mu = \mu_0$ or $\mu < \mu_0$, so there is no difference for us to write $H_0 : \mu = \mu_0$ or $H_0 : \mu \leq \mu_0$

But writing $H_0 : \mu = \mu_0$ will simplify the thought process.

What test statistic to use?

We are going to use the point estimate and its sampling distribution for a population mean

Point estimate for μ is: sample mean $\bar{X}$

↓

In the court trial, we always start by assuming that $H_0 : \mu = \mu_0$ is true

↓

The CLT gives us the sampling distribution of sample mean: $\bar{X} \sim N(\mu_0, (\frac{\sigma}{\sqrt{n}})^2)$

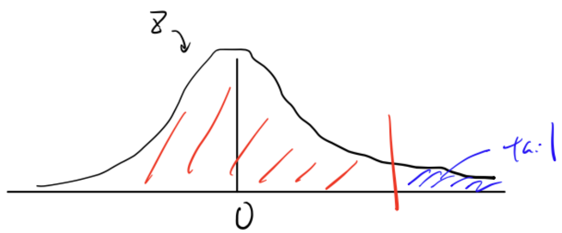
↓

By standardization, we get a Z test statistic:

$$Z_{obs} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$$

Distance between sample mean and population mean
under null hypothesis, in the unit of standard deviation

The behavior of Test Statistic under null and alternative hypothesis:

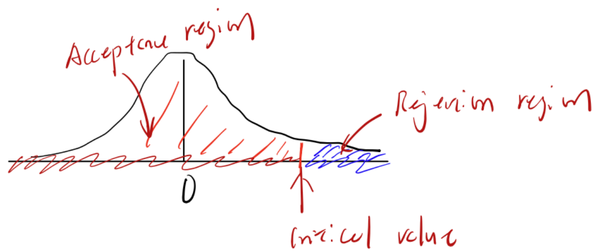


We expect most data fall into the red region, if the null hypothesis is true

Tail blue region indicates extremely unlikely events if null hypothesis is true,

When will it happen?

- When alternative hypothesis is true: when the population mean μ is indeed larger than what we assumed

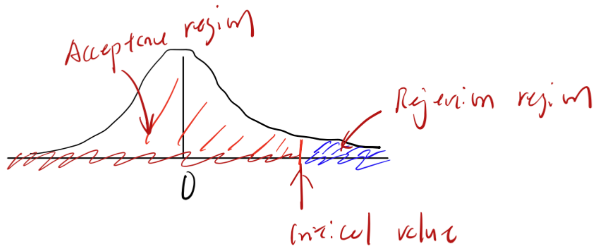


Acceptance region:
Values that support null hypothesis

Rejection region:
Values that support alternative hypothesis

Critical value:
The value that separates the acceptance region and the rejection region

Question: So how do we decide where to cut off the acceptance and rejection region?
Where should we put the critical value?



Question: So how do we decide where to cut off the acceptance and rejection region?
Where should we put the critical value?

If critical value too small:
Might falsely reject null hypothesis when the null hypothesis is actually true

Type I error

We don't want to make this type of mistake, so we control this type of risk/probability to be small:

$$\alpha = P(\text{Type I Error}) = P(\text{falsely reject } H_0 \text{ when it is true})$$

This measure of risk is called significance level or level of significance

- Depend on how much risk you are willing to take of making an incorrect decision
- Usually a small number $\alpha = 0.05, 0.01$

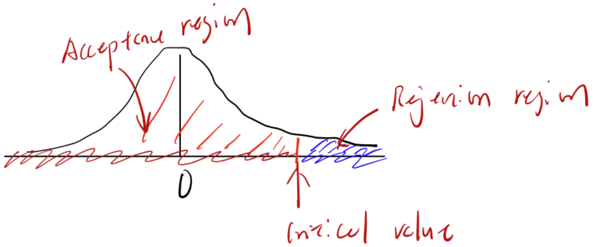
Approach 1: critical value approach

If we know our risk tolerance, i.e. the significance level α, then we can find the critical value.

Once we find the critical value, we know the acceptance and rejection region.

Then:

- if the test statistic fall into acceptance region, we do not reject null hypothesis
- if the test statistic fall into rejection region, we reject null hypothesis



The CLT gives us the sampling distribution of sample mean: $\bar{X} \sim N(\mu_0, (\frac{\sigma}{\sqrt{n}})^2)$

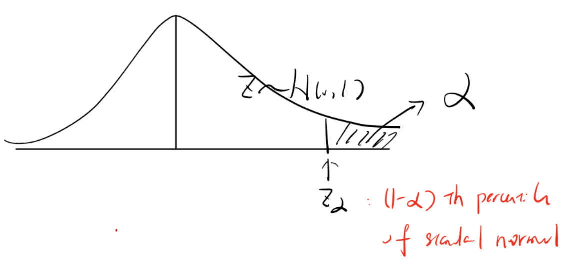
By standardization, we get a Z test statistic:

$$Z_{obs} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1) \text{ if null hypothesis is true}$$

The critical value can be found :

$P(Z \geq \text{critical value}) = \alpha$

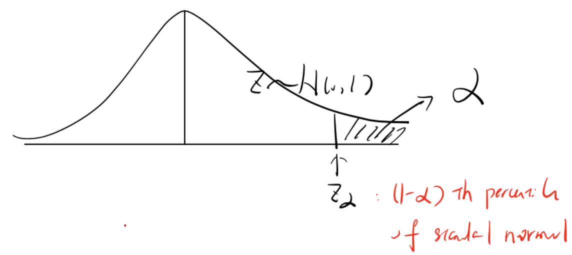
$\Rightarrow$ critical value $= Z_\alpha$

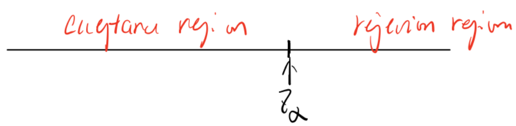


The critical value can be found :

$P(Z \geq \text{critical value}) = \alpha$

$\Rightarrow$ critical value $= Z_\alpha$





What we need to do:

Calculate test statistic $Z_{obs} = \dfrac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

If $Z_{obs} > Z_\alpha$, reject H_0 and conclude that H_a is true
If $Z_{obs} < Z_\alpha$, accept H_0 and conclude that H_0 is true

Summarize : critical value approach

Compare test statistic with the critical value

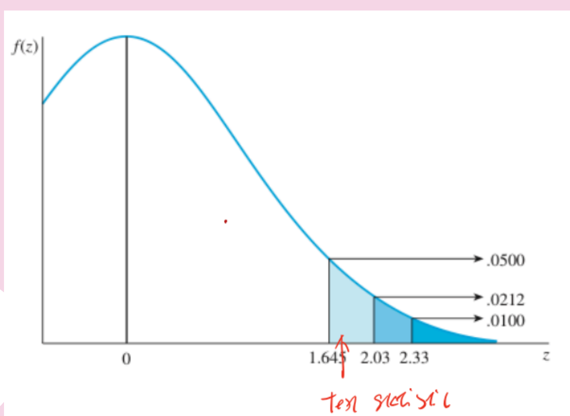
Based on significance level α, which controls the risk of making wrong decision

High risk tolerant

Low risk tolerant

We have to recalculate the critical values for people with different risk tolerance, and make different decisions. Too much repetitive work!

Risk tolerance: 0.01 -> 0.02 -> 0.05



Decision: null is true -> null is true -> alternative is true

Approach 2: P-value approach

We hope that: we don't have to go back and recalculate everything for different risk tolerances.

What is P-value?

- **Formal definition: the smallest significance level for which the null hypothesis can be rejected.**

- **More useful definition: If the null hypothesis is true, what is the probability of observing the point estimate as large as or even more extreme than the one we observe from current sample**

  - Point estimate: corresponding to the parameter of interest
    - sample mean $\bar{X}$ for population mean μ, sample proportion for population proportion p

For one-sided hypothesis problem:

$$\underbrace{p - value = P(Z \geq Z_{obs})}_{\text{Area under the curve of standard normal to the right tail of the calculated test statistic}}$$

P-value: measures the strength of evidence against null hypothesis

- Small p-value:
  - strong evidence that null hypothesis should be rejected
  - Alternative is true

- Large p-value:
  - Little or no evidence that null hypothesis should be rejected
  - Null is true

What we need to do:

Calculate p-value
If p-value $\leq \alpha$, reject H_0 and conclude that H_a is true
Ifp-value $> \alpha$, accept H_0 and conclude that H_0 is true

Summarize : p-value approach

Compare p-value with specified significance level

Summary: Hypothesis Testing about a Population Mean μ

Step 1: null hypothesis
$H_0 : \mu = \mu_0$

Step 2: Alternative hypothesis

One-sided
$H_a : \mu > \mu_0$ or $H_a : \mu < \mu_0$

Two-sided
$H_a : \mu \neq \mu_0$

Choice of one-sided or two-sided hypothesis:

Depend on what do you want to detect in specific problem.

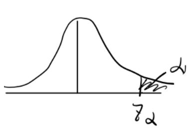
If pollution level higher than certain level cause great health risk, then the research question will naturally ask you to conclude if polution level is greater than some level. In this case, it will be one-sided.

Step 3: calculate test statistic
$$Z_{obs} = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

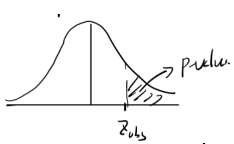
Step 4: critical value or p-value approach

One-sided:

Critical value

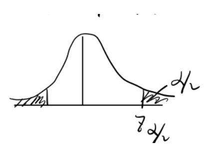


P-value

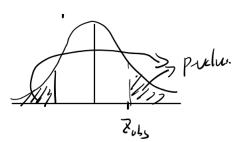


Two-sided:

Critical value



P-value



Step 5: Decision and Conclusion

Reject H_0 if:

$Z_{obs} \geq Z_\alpha$

p-value $\leq \alpha$

$|Z_{obs}| \geq Z_\alpha$

p-value $\leq \alpha$

Let's illustrate our first hypothesis testing through a concrete example.

Average weekly earning for female social workers is \$670.

Question: Do men in the same profession have higher salary than women?

Data: a random sample of n=40 male social workers

$\bar{X} = \$725,\ S = \$102$

The researchers want to set significance level $\alpha = 0.01$

Step 1: null, alternative hypotheses

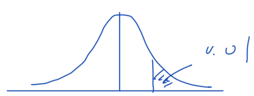
$H_0: \mu = 670$ vs $H_a: \mu > 670$

Step 2: test statistic

$$Z_{obs} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{725 - 670}{102/\sqrt{40}} = 3.41$$

Step 3:

critical value approach:



$Z_{\alpha=0.01} = 2.33$

Acceptance region

rejection region

p-value approach: p-value $= P(Z \geq 3.41)$

$= 0.0003$

Step 4: conclusion and decision:

<u>Critical value approach:</u>

Since the observed test statistic falls in the rejection region,
We reject H_0 and conclude that the average weekly earnings for male social workers are significantly higher than the female social workers.

<u>P-value approach:</u>

Since p-value =0.0003< 0.01,
We reject H_0 and conclude that the average weekly earnings for male social workers are significantly higher than the female social workers.

Note: we present fact as evidence, but fairness or explanation or reasons why the disparity exists, is not a statistical question.

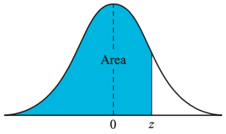


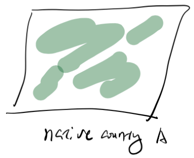
TABLE 3 **Areas under the Normal Curve**

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

TABLE 3 *(continued)*

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5 871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Giant hogweed



Average density in native country was found to be 5 plants/square meters.

Scientists: Does the invaded country has average density that is different from the native country, at significance level 0.05?

Sample: $n = 50, \bar{X} = 11.17, S = 3.9$

Step 1 : null , alternative hypotheses

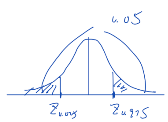
$H_0 : \mu = 5$ vs $H_a : \mu \neq 5$

Step 2 : test statistic

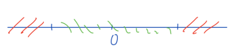
$$Z_{obs} = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{11.17 - 5}{3.9/\sqrt{50}} = 11.19$$

Step 3 :

critical value approach :



$$\underbrace{Z_{0.975}}_{1.96} = \underbrace{-Z_{0.025}}_{= -1.96}$$



rejection

acception

p-value approach : p-value $= P(|Z| \geq 11.19)$

$$= 2 \times \underbrace{P(Z \leq -11.19)}_{= 4.6 \times 10^{-29} \text{ denote as } 0^+}$$

$$= 0^+$$

Step 4: conclusion and decision:

Critical value approach:

Since the observed test statistic falls in the rejection region,
We reject H_0 and conclude that the average density in invaded country is significantly different from native country.

P-value approach:

Since p-value =0+ (meaning a really small number close to 0) < 0.05,
We reject H_0 and conclude that the average density in invaded country is significantly different from native country.

Does College Pay Off? An article in Time describing various aspects of American life indicated that higher educational achievement paid off! College grads work 7.4 hours per day, fewer than those with less than a college education. [2] Suppose that the average work day for a random sample of $n = 100$ individuals who had less than a 4-year college education was calculated to be $\bar{x} = 7.9$ hours with a standard deviation of $s = 1.9$ hours.

Use the p-value approach to test the hypothesis that the average number of hours worked by individuals having less than a college degree is greater than individuals having a college degree. At what level can you reject H_0 ?

Step 1 : null , alternative hypotheses

$H_0: \mu = 7.4$ vs $H_a: \mu > 7.4$

Step 2 : test statistic

$$Z_{obs} = \frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{7.9 - 7.4}{\frac{1.9}{\sqrt{100}}} = 2.63$$

Step 3 :

p-value approach : p-value $= P(Z \geq 2.63)$

$= 0.004$

< 0.05

Step 4: conclusion and decision:

<u>P-value approach:</u>

Since p-value $=0.004<0.05$,
We reject H_0 and conclude that the average number of hours worked by individuals having less than a college degree is greater than individuals with a college degree.