Outline

- General concept
- Type of errors and power
- z-test for the mean, variance known
- t-test for the mean, variance unknown
- test for sample proportion
- Confidence interval and hypothesis test
- P-value

Hypotheses Testing

The **hypothesis test** has **two** possible but contradictory truths, written in terms of Hypotheses:

H_0 **Null Hypothesis**

- The *prior belief* with the data.

H_1 **Alternative Hypothesis**

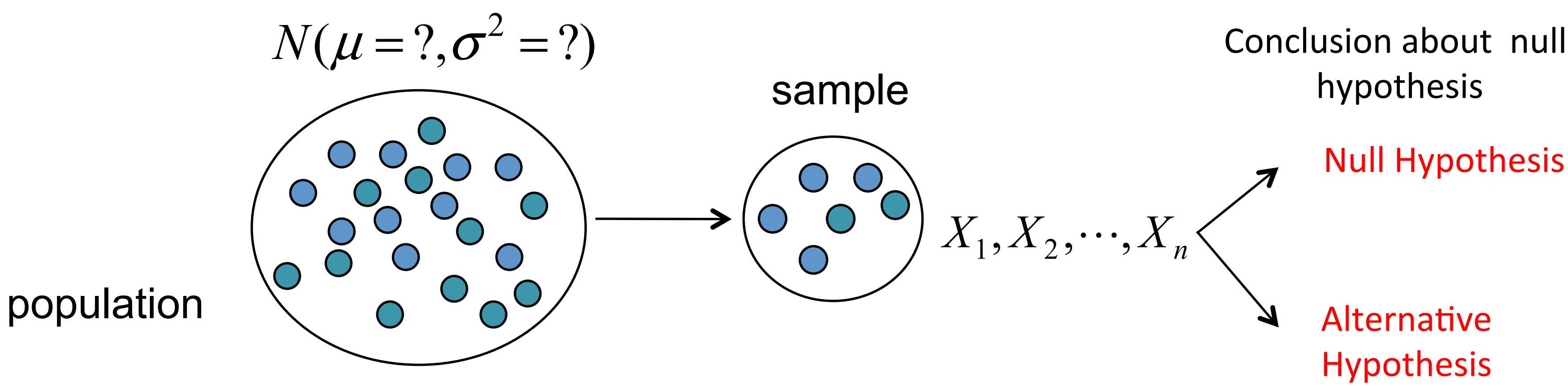
- The alternative belief we have with the data
- Varies from problem to problem

Goal of hypothesis testing is to use data to tell which hypothesis is true.

Mathematical formulation

- For example, comparing the mean and variance of samples from normal distribution

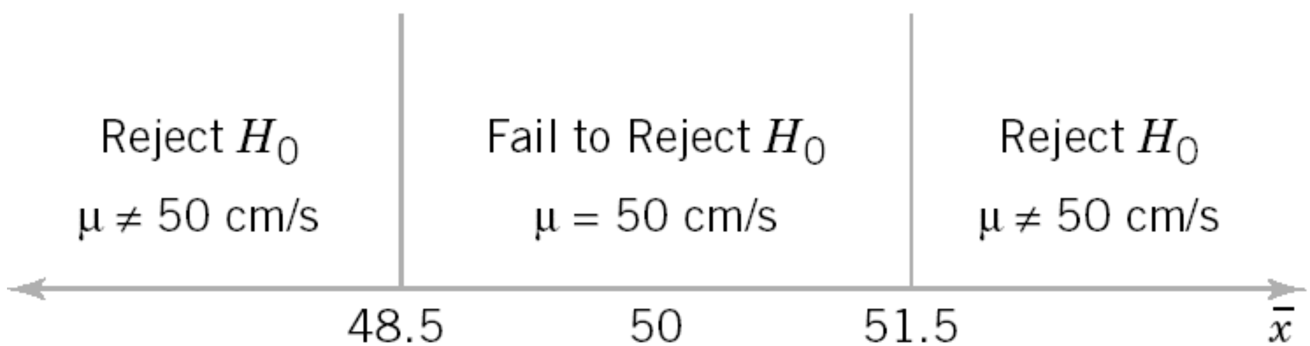
Null Hypothesis	$H_0 : \mu = \mu_0$	$H_0 : \sigma^2 = \sigma_0^2$
Alternative Hypothesis	$H_1 : \mu \neq \mu_0$	$H_1 : \sigma^2 \neq \sigma_0^2$



Example

- Example: measure the diameter of a batch of screws

$$H_0\text{: } \mu = 50 \text{ centimeters}$$
$$H_1\text{: } \mu \neq 50 \text{ centimeters}$$



Examples of Hypotheses Test

(1) Is the coin fair? Prior assumption: coin is fair. Let p = P(heads),

$$\begin{cases} H_0: \quad p = \dfrac{1}{2} & \text{(our prior belief)} \\ H_1: \quad p \neq \dfrac{1}{2} & \text{(the opposite of our prior belief)} \end{cases}$$

(2) Apple produces a batch of iPhones, with battery life (X) with mean μ, variance = σ^2

Is the variability of battery under control?

Here the variability is under control when it is smaller than a known value σ_0^2

$$\begin{cases} H_0: \quad \sigma^2 \leq \sigma_0^2 \\ H_1: \quad \sigma^2 > \sigma_0^2 \end{cases}$$

(3) Does this batch of iPhones have sufficient battery life?

$$\begin{cases} H_0: \quad \mu > \mu_0 \\ H_1: \quad \mu \leq \mu_0 \end{cases}$$

Class Activity

(1) Is the coin fair? $P=P(\text{heads})$

A. H_0: $P = ½$ **B. H_A: $P = ½$** **C. H_0: $P \neq ½$**

(2) A machine produces product (X) with mean μ , variance σ^2

Is the variability under control?

A. H_A: $\sigma^2 \leq \sigma_0^2$ **B. H_0: $\sigma^2 > \sigma_0^2$** **C. H_0: $\sigma^2 \leq \sigma_0^2$**

Do we support the hypothesis that the machine produce an item of a size larger than a known $\mathbf{\mu_0}$?

A. $H_A : \mu \leq \mu_0$ **B. $H_0 : \mu > \mu_0$** **C. $H_0 : \mu \leq \mu_0$**

Two-sided versus one-sided

Two-sided Alternative Hypothesis

X: inner-diameter of screws

$$H_0: \mu = 50$$ null hypothesis

$$H_1: \mu \neq 50$$ alternative hypothesis

One-sided Alternative Hypotheses

$$H_0: \mu = 10$$
$$H_1: \mu < 10$$

X: customers' waiting time in a bank

$$H_0: \mu = 100$$
$$H_1: \mu > 100$$

X: time to failure of machines

Simple versus composite

Simple Hypothesis

Testing two possible values of the parameter

$$H_0 : \mu = 12$$ null hypothesis

$$H_1 : \mu = 24$$ alternative hypothesis

Remaining quantity in a vending machine, 1 dozen or 2 dozens

Composite Hypotheses

Testing a range of values

$$H_0 : \mu = 10$$

$$H_1 : \mu < 10$$

X: customers' waiting time in a bank

$$H_0 : \mu = 50$$

$$H_1 : \mu \neq 50$$

Average diameter of screw

Outline

- General concept
- Type of errors and power
- z-test for the mean, variance known
- t-test for the mean, variance unknown
- test for sample proportion
- Confidence interval and hypothesis test
- P-value

Errors in Hypothesis Testing

α = P(**type I error**) = P(**Reject H_0, when H_0 is true**)

This error rate can be set low enough to ensure the test is "safe".

β = P(**type II error**) = P(**Accept H_0, when H_1 is true**)

truth

decision

	H_0 is True	H_0 is False
Accept H_0	Correct Decision	Type II Error
Reject H_0	Type I Error	Correct Decision

Court room

Suppose you are the prosecutor in a courtroom trial. The defendant is either guilty or not. The jury will either convict or not.

		truth	
		Not Guilty	Guilty
decision	Free-of-guilt	Right decision	Wrong decision
	Convict	Wrong decision	Right decision

Significant level

The classic hypothesis test

- fixes α (type I error) to be some small (tolerable) value
- accept the corresponding β (type II error)

  that results from this.

The level fixed for α is called **significance level.**

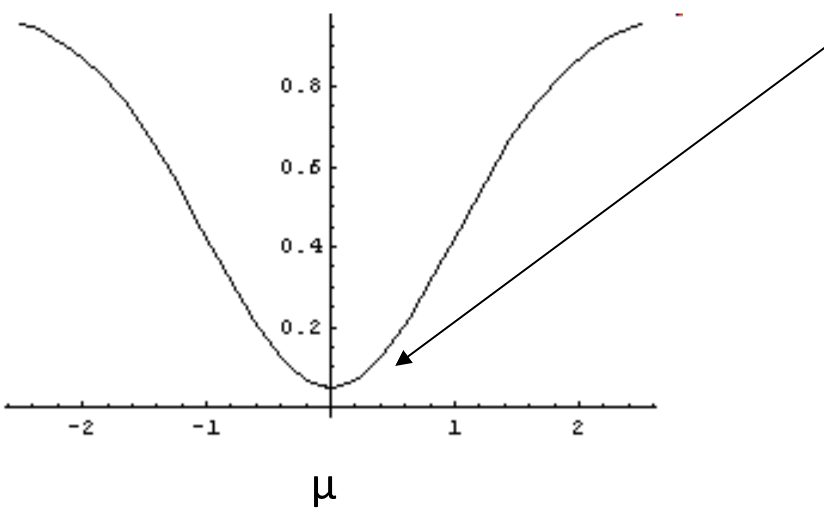
Typical value for α: 0.1, 0.05, 0.001

Statistical power

Power = $1 - \beta$ = P(reject H_0 when H_1 is true)

Example: power function for the test $H_0: \mu = 0$

$H_1: \mu \neq 0$



At $\mu = 0$, H_0 is true, so this point is equal to α, the type I error

Y-axis shows power, X-axis is the value of μ.

Class Activity 2

In 1990, a study on the weight of students at GT provided an average weight of μ = 160 lbs. We would like to test our belief that the GT student weight average did not decrease in the past 15 years.

1. What is the alternative hypothesis?

A. H_1: $\mu = 160$ B. H_1: $\mu > 160$ C. H_1: $\mu < 160$

2. Test H_0: $\mu = 160$ vs. H_1: $\mu < 160$. What is $P(\text{Reject } H_0 \mid \mu = 160)$?

A. Type I error B. Type II error C. Power

3. Test H_0: $\mu = 160$ vs. H_1: $\mu < 160$. What is $P(\text{Reject } H_0 \mid \mu < 160)$?

A. Type I error B. Type II error C. Power

Outline

- General concept
- Type of errors and power
- z-test for the mean, variance known
- t-test for the mean, variance unknown
- test for sample proportion
- P-value

One-sided z-test

For the GT student average example, we want to test

$$H_0 : \mu = 160$$
$$H_1 : \mu < 160$$

The conventional way to test this hypothesis is to find the test for which the type-I error is fixed at a particular value (e.g., $\alpha = 0.01$, 0.05, 0.10).

Sample mean $\overline{x}$ is a good estimator for μ, so here we use

$\overline{x}$ as the test statistic.

We should reject the null hypothesis when $\overline{x}$ is small

Determine threshold from significance level

The test is to reject H_0 if $\overline{x}$ < b.

To determine the threshold b,

$$\alpha = P(\bar{X} < b | \mu = 160) = P(Z < \frac{b\text{-}160}{\sigma/\sqrt{n}})$$

This determines b because $\dfrac{b\text{-}160}{\sigma/\sqrt{n}} \equiv \Phi^{-1}(\alpha)$

$$b = \left(\sigma/\sqrt{n}\right)\Phi^{-1}(\alpha) + 160$$

Numerical values

$$b = \left(\sigma / \sqrt{\mathrm{n}}\right)\Phi^{-1}(\alpha) + 160$$

If α = 0.05, σ = 10, n = 100,

$$\Phi^{-1}(0.05) = -1.645$$

We reject H_0 when

$$\bar{\mathrm{x}} < \left(\sigma / \sqrt{\mathrm{n}}\right) z_{\alpha} + 160 = -10 / \sqrt{100} \times 1.645 + 160 = 158.355$$

Note that even when $158.355 \leq \bar{\mathrm{x}} < 160$

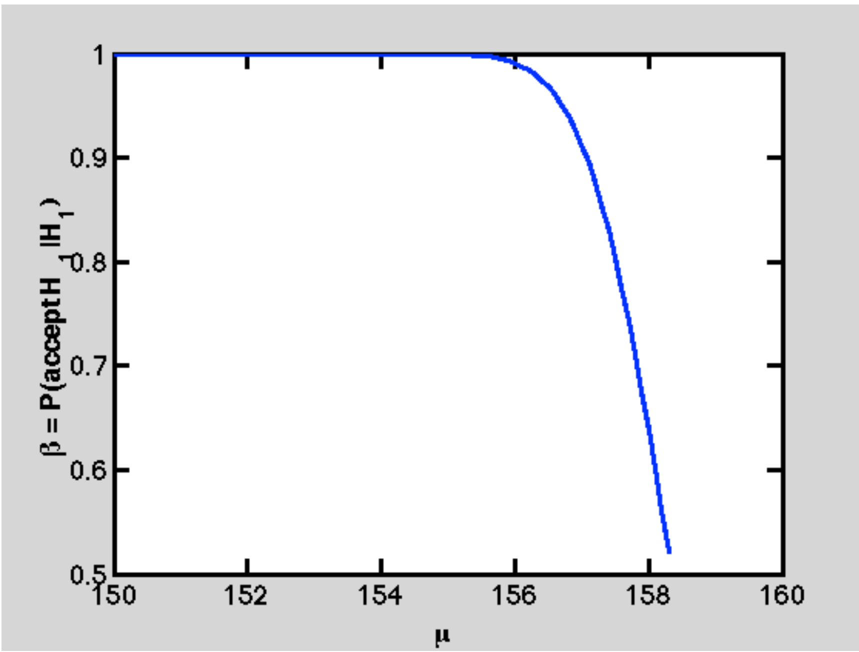
we still accept H_0 => we take into account data variability.

Power

- Suppose $\mu = 150$ (so H_1 is true)
- Power when $\mu = 150$

$$P\{\overline{x} < 158.355\} = P\left\{\frac{\overline{x}\text{-}150}{10/\sqrt{100}} < \frac{158.355 - 150}{10/\sqrt{100}}\right\} = \Phi(8.355) \approx 1$$

Entire Power Function = Φ(158.355-μ)



Two-sided z-test

For the GT student average example, suppose now we want to test that the average weight has not changed in the past 15 years

$$H_0 : \mu = 160$$

$$H_1 : \mu \neq 160$$

$$\bar{x} \notin [160 - k, 160 + k]$$

Reject H_0 when

$$\alpha = \mathrm{P}(|\bar{x}\text{-}160| > k | \mu = 160) = \mathrm{P}(|\mathrm{Z}| > \frac{k}{\sigma / \sqrt{\mathrm{n}}})$$

$$\frac{k}{\sigma / \sqrt{\mathrm{n}}} = z_{\alpha/2}$$

α = 0.05, $z_{0.025} = 1.96$

σ = 10, n = 100, $k = z_{\alpha/2} \sigma / \sqrt{\mathrm{n}} = 1.96$

Reject H_0 when $\bar{x} \notin [158.04, 161.96]$

Outline

- General concept
- Type of errors and power
- z-test for the mean, variance known
- t-test for the mean, variance unknown
- test for sample proportion
- Confidence interval and hypothesis test
- P-value

t-test, two-sided

For the GT student average example, we want to test

$$H_0 : \mu = 160$$
$$H_1 : \mu \neq 160$$

We do not know the true variance, but just the sample variance s.

Note that before we test using **standardized test statistics**

Known variance

z-test

$$\left| \frac{\bar{x} - 160}{\sigma / \sqrt{n}} \right| > k$$

Unknown variance

t-test

$$\left| \frac{\bar{x} - 160}{s / \sqrt{n}} \right| > k$$

Determine threshold for t-test

The test is to reject H_0 if $\left|\frac{\bar{x}-160}{s/\sqrt{n}}\right| > k$

To determine the threshold b,

$$\alpha = \mathrm{P}\left(\left|\frac{\bar{x}-160}{s/\sqrt{n}}\right| > k \mid \mu = 160\right) = \mathrm{P}(|\mathrm{T}| > \mathrm{k}) \Rightarrow k = t_{\alpha/2,n-1}$$

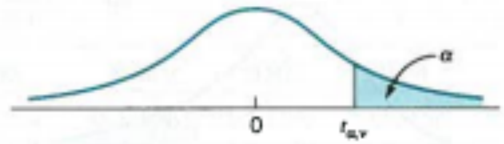
$\tilde{\Phi}(x)$ CDF of t-random variable

We reject H_0 when $\left|\frac{\bar{x}-160}{\boxed{s}/\sqrt{n}}\right| > \boxed{t_{\alpha/2,n-1}}$

In comparison, when variance is known, we reject H_0 when

$$\left|\frac{\bar{x}-160}{\boxed{\sigma}/\sqrt{n}}\right| > \boxed{z_{\alpha/2}}$$

Percentage Points of the t Distribution[a]



ν	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0025	0.001	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	0.265	0.727	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.49	4.019	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.587
11	0.260	0.697	1.363	1.796	2.20	2.718	3.106	3.497	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.992
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.135	3.527	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646

Example: coke machine two-side t-test

A soft-drink machine at a steak house is regulated so that the amount of drink dispensed is approximately normally distributed with a mean of 200 ml.

The machine is checked periodically by taking a sample of 9 drinks and computing the average content. If for one batch of samples variance s = 30 ml, and mean is $\overline{x}$ = 215 ml, do we believe the machine is operating running OK, for a significant level $\alpha = 0.05$?

Formulation: coke machine

$$H_0 : \mu = 200$$

$$H_1 : \mu \neq 200$$

Test statistics $\left| \dfrac{\bar{x} - 200}{s / \sqrt{n}} \right|$

Threshold $k = t_{0.025,8} = 2.306$

Reject H_0 when $\left| \dfrac{\bar{x} - 200}{s / 3} \right| > 2.306$

With our data, $\left| \dfrac{\bar{x} - 200}{s / \sqrt{n}} \right| = \left| \dfrac{210 - 200}{30 / 3} \right| = 1 < 2.306$

So machine is running OK.

Outline

- General concept
- Type of errors and power
- z-test for the mean, variance known
- t-test for the mean, variance unknown
- test for sample proportion
- Confidence interval and hypothesis test
- P-value

Test on population proportion

Sample proportion, X ~ BIN(n, p)

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p \neq p_0 \end{cases}$$

Based on CLT approximation, n > 30

Under H_0, X approximately normal random variable

$$X \sim \mathrm{N}\left(\ np_0, \quad np_0(1-p_0) \ \right)$$

standardized test statistics $\dfrac{X - np_0}{\sqrt{np_0(1-p_0)}} \sim \mathrm{N}\left(\ 0, \quad 1 \ \right)$

Test with significance level: α

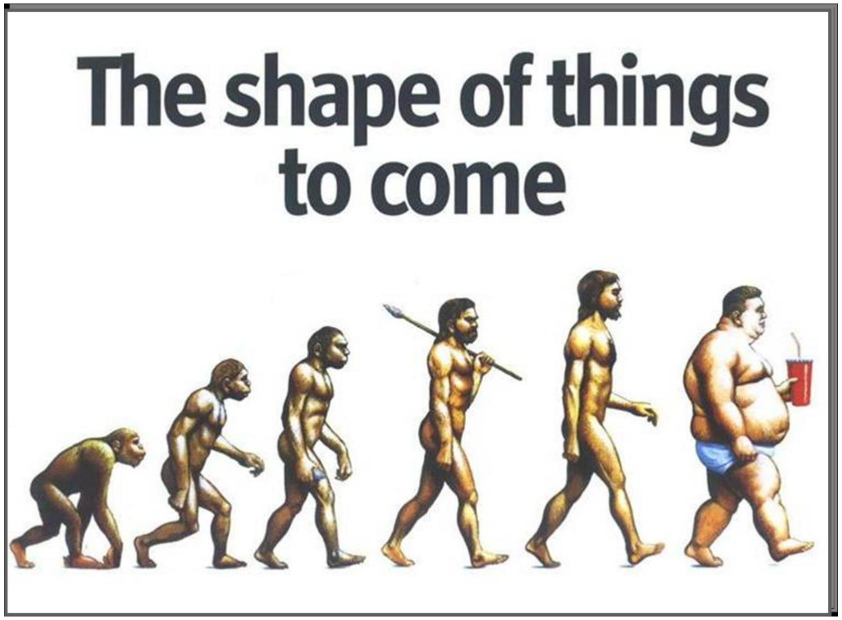
Reject H_0 when $\left| \dfrac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \right| > z_{\alpha/2}, \hat{p} = X/n$

Example: obesity

The Associated Press (October 9, 2002) reported that 1276 individuals in a sample of 4115 adults were found to be obese.

A 1998 survey based on people's own assessment revealed that 20% of adult Americans considered themselves obese.

Does the recent data suggest that the true proportion of adults who are obese is consistent with the self-assessment survey, with significance level 0.05?



Example continued

$$\begin{cases} H_0 : p = 0.2 \\ H_1 : p \neq 0.2 \end{cases}$$

Data from 2002 study: 1276 of 4115 were reported with obesity

$$\hat{p} = 1276 / 4115 = 0.31$$

$$\alpha = 0.05$$

$$z_{0.025} = 1.96$$

$$\left| \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0) / n}} \right| = \left| \frac{0.31 - 0.2}{\sqrt{0.2(1 - 0.2) / 4115}} \right| = 17.74 > 1.96$$

Reject the null hypothesis: the proportion is not 0.2 (very likely to be greater than 0.2)

Summary of Tests

Summary: test for mean

Null Hypothesis

$$H_0 : \mu = \mu_0$$

Test Statistic

$$\bar{x}$$

Significance level: α

Alternative Hypothesis	Known Variance **H0 is rejected if**	Unknown Variance **H0 is rejected if**
$H_1 : \mu \neq \mu_0$	$\lvert\bar{x} - \mu_0\rvert > z_{\alpha/2}\sigma / \sqrt{n}$	$\lvert\bar{x} - \mu_0\rvert > t_{\alpha/2,n-1} s / \sqrt{n}$
$H_1 : \mu > \mu_0$	$\bar{x} > \mu_0 + z_\alpha \sigma / \sqrt{n}$	$\bar{x} > \mu_0 + t_{\alpha,n-1} s / \sqrt{n}$
$H_1 : \mu < \mu_0$	$\bar{x} < \mu_0 - z_\alpha \sigma / \sqrt{n}$	$\bar{x} < \mu_0 - t_{\alpha,n-1} s / \sqrt{n}$

Summary: test for sample proportion

Null Hypothesis

$$H_0 : p = p_0$$

Test Statistic

$$\frac{\hat{p} - p_0}{\sqrt{p_0\left(1 - p_0\right)/n}}$$

Significance level: α

Alternative Hypothesis	**H0 is rejected if**
$H_1 : p \neq p_0$	$\left\lvert \frac{\hat{p} - p_0}{\sqrt{p_0\left(1 - p_0\right)/n}} \right\rvert > z_{\alpha/2}$

Outline

- General concept
- Type of errors and power
- z-test for the mean, variance known
- t-test for the mean, variance unknown
- test for sample proportion
- p-value
- Confidence interval and hypothesis test

Motivation for p-value

Think about the GT student average weight example.

Assume average weight is 160lb. Based on the test, we reject H_0 if sample average is less than 158.35, and this gives significance level 0.05 (i.e. chance of making type I error = 0.05)

Consider

case1: sample average = 150lb

case2: sample average = 155lb

In both cases we reject H_0

However, if sample sample average = 150lb, we should **reject H_0 with more confidence** than if sample average = 155lb.

In other words, when sample average = 150 lb, the chance for H_0 to happen is smaller.

p-value

- p-value = probability of observing some data even more "extreme" than the given data

- It is a measure of the null hypothesis plausibility based on the samples

- The smaller the p-value is, the less likely H_0 is true

  ⇨ Reject H_0 when p-value is small

⇨ Test can also be performed as rejecting H_0 when p-value is less than prescribed significant level

Convention: decisions based on P-value

- If **P-value < 0.01**
  - H_0 <u>is not plausible</u>/ H_1 is supported

- If **P-value > 0.1**
  - H_0 <u>is</u> plausible

- If **0.01 < P-value < 0.1**
  - we have some evidence that H_0 <u>is not plausible</u> but we need further investigation

1. p-value for one-sided Test

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu > \mu_0$$

$$\text{Test statistic: } Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

observed value of test statistic: $\bar{x}$

$$\text{p-value: } P(\bar{X} > \bar{x}) = P\left(Z > \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}\right)$$

$$= 1 - \Phi\left(\frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}\right)$$

Example: p-value for one-sided Test

Claim: mean battery life of a cellphone is 9 hours.
Observed mean battery life: 8.5 hours, from 44 observations,
σ = 1 hour
Test: the mean battery life of a cellphone exceeds 10 hours

$$H_0: \mu = 9$$

$$H_1: \mu < 9$$

Test statistic: $Z = \dfrac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$

observed value of test statistic: $\bar{x} = 8.5$

p-value: $P(\bar{X} < \bar{x}) = P\left(Z < \dfrac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}\right)$

$$= \Phi\left(\frac{8.5 - 9}{1/\sqrt{44}}\right) = 0.0095$$

Example: p-value nicotine content

A cigarette manufacturer claims that the average nicotine content of a brand of cigarettes is at most 1.5.

1. What is the alternative hypothesis?

A. H_A : m ≠ 1.5 **B. H_A : m < 1.5** **C. H_A : m > 1.5**

2. What is the p-value? Denote

A. P(Z ≥ z) **B. P(Z ≤ z)** **C. P(|Z| ≥|z|)**

$$Z = \sqrt{n}(\bar{X} - \mu_0)/S \text{ and } z = \sqrt{100}(1.7 - 1.5)/1.3$$

Calculation for nicotine example

- We observe the nicotine content for 100 cigarettes with a sample mean 1.7 and sample standard error 1.3.
- Observing data more extreme:

$$P(\bar{X} > \sqrt{100}(1.7-1.5)/1.3)$$
$$= P(\sqrt{n}(\bar{X}-\mu_0)/S > \sqrt{100}(1.7-1.5)/1.3)$$
$$= P(t_{n-1} > 1.5385) = 0.0620$$

- Not a very small p-value

2. p-value for two-sided Test

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

$$\text{Test statistic: } Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

observed value of test statistic: $\bar{x}$

$$\text{p-value: } P\left(\left| \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}} \right| > \left| \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \right| \right)$$

$$= P\left(|Z| > \left| \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \right| \right)$$

$$= 1 - 2\Phi\left(\frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \right)$$

Example: compute p-value

$$n = 13, \bar{x} = 2.879, \sigma = 0.325$$

$$H_0 : \mu = 3$$

$$H_1 : \mu \neq 3$$

p-value

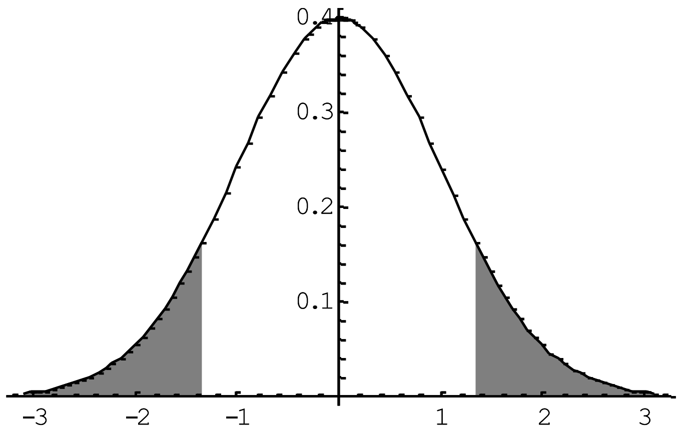
$$= P(\text{getting } \bar{x} \text{ even further away from 3 than } \bar{x} = 2.879)$$

$$= \mathrm{P}(|\,\overline{\mathrm{X}} - 3\,| > |\,3 - 2.879\,|\ )$$

$$= \mathrm{P}(\ |\mathrm{Z}|\ > \frac{.121}{.325/\sqrt{13}})$$

$$= \mathrm{P}(\ |\mathrm{Z}|\ > 1.34)$$

$$= 2\Phi(\text{-}1.34) = 0.18$$



Example: p-value for two-sided test

Test to see if the mean is significantly far from 90. The sample mean is 87.9 with known standard deviation of 5.9 for a sample of size 44.

1. What is the alternative hypothesis?

A. H_A : m ≠ 90 B. H_A : m < 90 C. H_A : m > 90

2. At a significance level $\alpha = 0.1$, what is your decision?

A. strongly suggest H_A B. H_0 is plausible C. strongly support H_0

3. What is the p-value? Denote $Z = \sqrt{n}(\bar{X} - \mu_0)/S$ and $z = \sqrt{44}(87.9 - 90)/5.9$

A. $P(Z \geq z)$ B. $P(Z \leq z)$ C. $P(|Z| \geq |z|)$

Comparison of significance level and p-values

- Two ways of reporting results of a hypothesis test

- 1: report the results of a hypothesis test is to state that the null hypothesis was or was not rejected at a specified level of significance. This is called **fixed significance level** testing.

- 2: The *p*-value is the probability that the test statistic will take on a value that is **at least as extreme** as the observed value of the statistic when the null hypothesis H_0 is true.

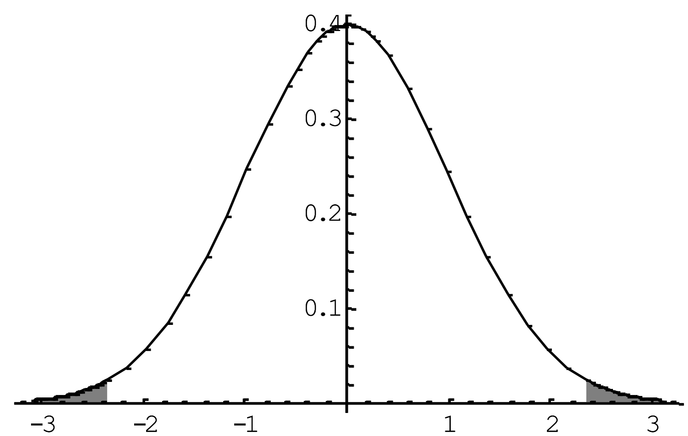
Relate p-value and significance level: example

$$H_0: \mu = 90$$
$$H_1: \mu \neq 90$$

$$\text{reject if } |z| > z_{\frac{\alpha}{2}} \text{ where } z = \frac{\bar{x}-90}{90/\sqrt{n}}$$

Note: if we have an α = 0.10 level test, then $Z_{.05} = 1.645$ and we would reject H_0 when $|Z| > 1.645$.

That means the actual p-value < 0.10.

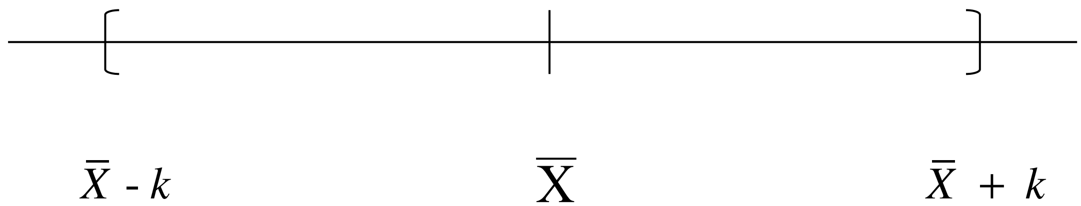


Outline

- General concept
- Type of errors and power
- z-test for the mean, variance known
- t-test for the mean, variance unknown
- test for sample proportion
- P-value
- Confidence interval and hypothesis test

Connection between confidence Intervals and Hypothesis Test

CI :



We believe the true parameter $\mu_0 \in [\bar{X} - k, \bar{X} + k]$

with probability $1 - \alpha$, $\bar{X} \in [\mu_0 - k, \mu_0 + k]$

Test :

If we wish to test
$$H_0 : \mu = \mu_o$$
$$H_1 : \mu \neq \mu_o$$

If $\overline{X}$ not in $[\mu_0 - k, \mu_0 + k]$, reject H_o

Under H_0, making error with probability $1 - \alpha$

More Examples

Two types of approaches to hypothesis testing

Fixed significance level	p-value
Decide decision statistics Reject H0, when Decision statistic >< threshold	Calculate p-value: what is the probability to observe a statistic more "extreme" than data The smaller p-value is, the more likely H1 is true → reject H0

Procedure of hypothesis test (sec. 9.1.6)

1. Set the significance level (.01, .05, .1)
2. Set null and alternative hypothesis
3. Determine other parameters
4. Decide type of the test

   - test for mean with known variance (z-test)
   - test for mean with unknown variance (t-test)
   - test for sample proportion parameter

6. Use data available:

   - perform test to reach a decision
   - and report p-value

Example 1: change of average height? z-test

The average height of females in the freshman class at GT has been 162.5 cm with a a standard deviation of 6.9 cm.

Is there reason to believe that there has been a **change** in the average height if a random sample of 50 females in the present freshman class has an average height of 165.2cm? Assume the standard deviation remains the same, and significance level 0.05.

Solution to example 1

1. Significance level: $\alpha = 0.05$
2. Set null and the alternative

   $H_0 : \mu = 162.5$

   $H_1 : \mu \neq 162.5$
3. Determine other parameters:
4. Decide type of test: $n = 50, \sigma = 6.9$

   known variance, test for mean, use z-test

   Reject H_0 if $\left|\overline{x} - \mu_0\right| > z_{\alpha/2}\sigma / \sqrt{n}$

   Plug in values: $z_{0.025} = 1.96$

   Reject H_0 if $\left|\overline{x} - 162.5\right| > 1.96 \times 6.9 / \sqrt{50} = 1.9126$
5. Use data to perform test:

$\overline{x} = 165.2, \left|\overline{x} - 162.5\right| = 2.7 > 1.9126$ ➡ Reject H_0

6. Report p-value:

$$\begin{aligned}
&\text{p-value} \\
&= P(|Z| > |165.2 - 162.5| / (6.9 / \sqrt{50})) \\
&= \ P(\ |Z| > 2.7669) \\
&= 2(1 - \Phi(2.7669)) = 0.0056
\end{aligned}$$

Example 2: effective hours of drug t-test

In a study for the effective hours of certain drug, the sample average time for n = 9 individual was 6.32 hours and the sample standard deviation was 1.65 hours.

It has previously been assumed that the average adaptation time was at least 7 hours.

Assuming effective hours to be normally distributed, does the data contradict the prior belief?

Solution to example 2

1. Significance level: $\alpha = 0.05$
2. Set null and the alternative

   $H_0 : \mu = 7$

   $H_1 : \mu \neq 7$
3. Determine other parameters: $n = 9, s = 1.65$
4. Decide type of test:

   unknown variance, test for mean, use t-test

   Reject H_0 if $|\bar{x} - \mu_0| > t_{\alpha/2,n-1} s / \sqrt{n}$

   Plug in values: $t_{0.025,8} = 2.306$

   Reject H_0 if $|\bar{x} - 7| > 2.306 \times 1.65 / \sqrt{9} = 1.2683$
5. Use data to perform test:

   $\bar{x} = 6.32, |\bar{x} - 7| = 0.68 < 1.2683$ ➡ Accept H_0
6. Report p-value:

6. Report p-value:

$$\text{p-value}$$

$$= P(|T_8| > |6.2 - 7| / (1.65 / \sqrt{9}))$$

$$= P(|T_8| > 1.4545)$$

$$= 2 \times 0.0562 = 0.1124$$

Table V Percentage Points $t_{\alpha,\nu}$ of the t Distribution

ν \ α	.40	.25	.10	.05	.025	.01	.005	.0025	.001	.0005
1	.325	1.000	3.078	6.314	12.706	31.821	63.657	127.32	318.31	636.62
2	.289	.816	1.886	2.920	4.303	6.965	9.925	14.089	23.326	31.598
3	.277	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.271	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.267	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.265	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.029	4.785	5.408
8	.262	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041

Example 3: sample proportion

In a survey from 2000, it has been found that 33% of the adults favored paying traffic ticket rather than attending traffic school.

Now we did a online survey and found out from a sample of 85 adults, 26 favor paying traffic ticket.

Do we have reason to believe that the proportion of adults favoring paying traffic tickets has decreased today with a significant level 0.05?

Summary: test for sample proportion

Null Hypothesis

$$H_0 : p = p_0$$

Significance level: α

Test Statistic

$$\frac{\hat{p} - p_0}{\sqrt{p_0\left(1 - p_0\right)/n}}$$

Alternative Hypothesis	H0 is rejected if
$H_1 : p \neq p_0$	$\left\lvert \frac{\hat{p} - p_0}{\sqrt{p_0\left(1 - p_0\right)/n}} \right\rvert > z_{\alpha/2}$

Solution to example 3

1. Significance level: $\alpha = 0.05$
2. Set null and the alternative

   $$H_0 : p = 0.33$$

   $$H_1 : p \neq 0.33$$
3. Determine other parameters: $n = 85$
4. Decide type of test:

   sample proportion test

   Reject H_0 if $$\left| \frac{\hat{p} - p_0}{\sqrt{p_0 \left(1 - p_0\right) / n}} \right| > z_{\alpha/2}$$

Plug in values: $z_{0.025} = 1.96$

Reject H_0 if $$\left| \frac{\hat{p} - 0.33}{\sqrt{0.33 \left(1 - 0.33\right) / 85}} \right| > 1.96$$

Solution to example 3

5. Use data to perform test:

$$\hat{p} = 26/85 = 0.3059$$

$$\left| \frac{0.3059 - 0.33}{\sqrt{0.33(1-0.33)/85}} \right| = 0.4725 < 1.96 \Rightarrow \text{Accept } H_0$$

6. Report p-value:

$$\text{p-value} = P\left(|Z| > \left| \frac{0.3059 - 0.33}{\sqrt{0.33(1-0.33)/85}} \right|\right)$$

$$= P(|Z| > 0.4725)$$

$$= 2(1 - \Phi(0.4725)) = 0.6366$$

Summary for performing hypothesis test

- Follow the procedure

- Examples:
  - Test for normal mean, variance known, use z-test
  - Test for normal mean, variance unknown, use t-test
  - Test for sample proportion, using normal approximation and z-test