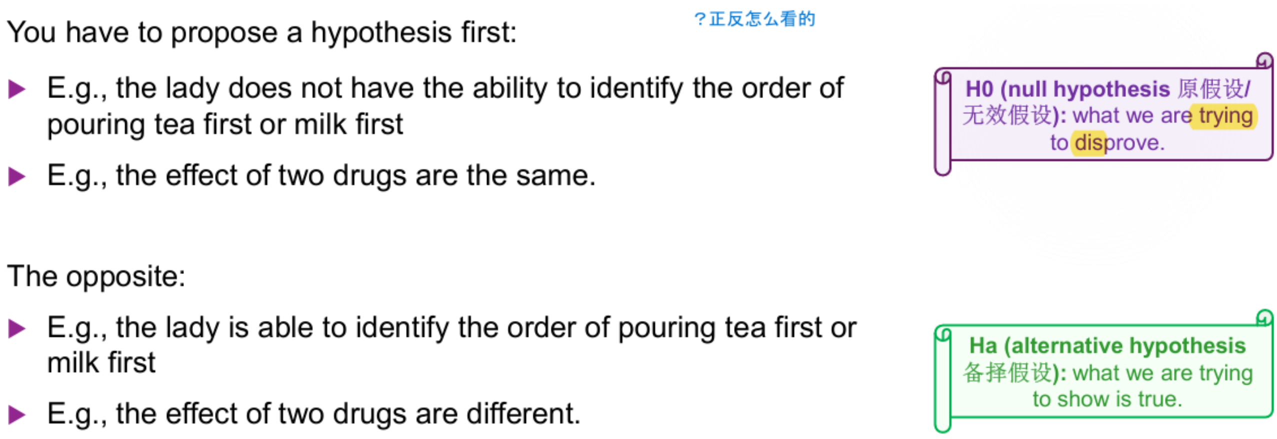
术语

H0& Ha

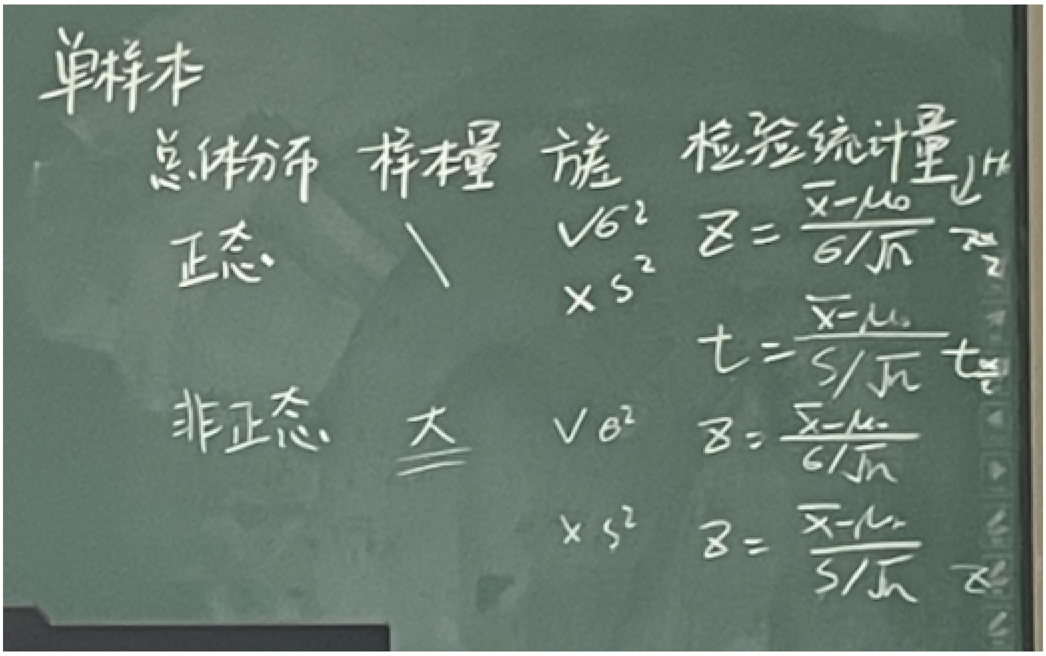


等号放在H0那边（？）

Rejection region=critical region

```
![[Pasted image 20240420091216.png]]
![[Pasted image 20240420091222.png]]
```

test statistic检验统计量&critical value (临界值/阈值)

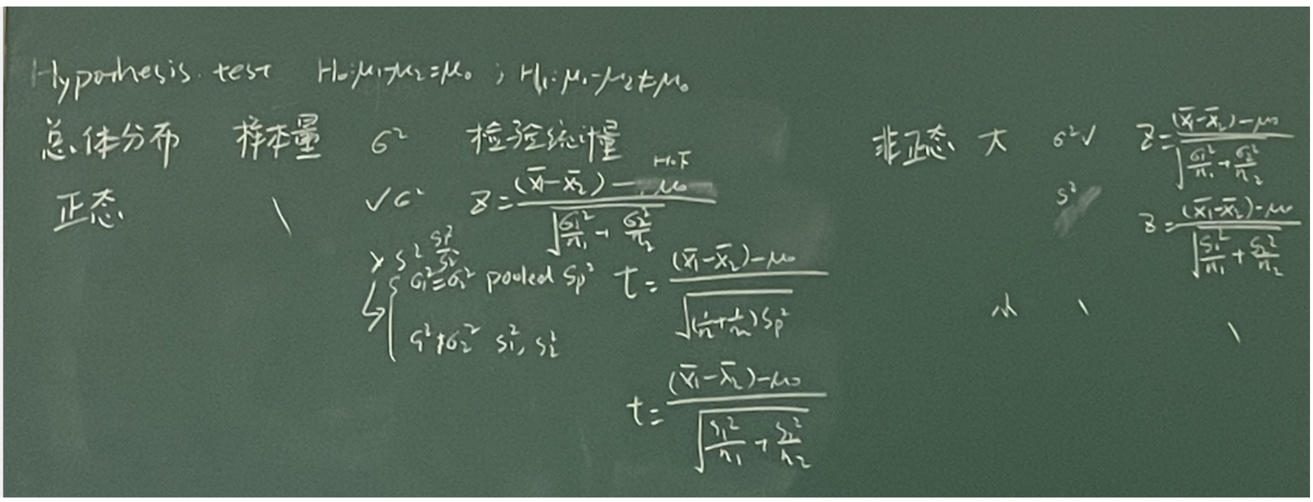


4.2 Summary: Hypothesis tests for one mean

H0 : μ = μ0, Ha : μ ≠ μ0

Population Distribution	Sample Size	Population Variance	Test Statistic
Normal	Any	σ^2 known	$z_{obs} = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$
	Any	σ^2 unknown uses s^2, df=n-1	$t_{obs} = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$
Not Normal/ Unknown	Large	σ^2 known	$z_{obs} = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$
	Large	s^2 unknown uses s^2	$z_{obs} = \frac{\bar{X}-\mu_0}{s/\sqrt{n}}$
	Small	Any	Non-parametric methods

Check first



5.2 Summary: Hypothesis tests for a difference of two means H0 : μ1 − μ2 = μ0, Ha : μ1 − μ2 ≠ μ0

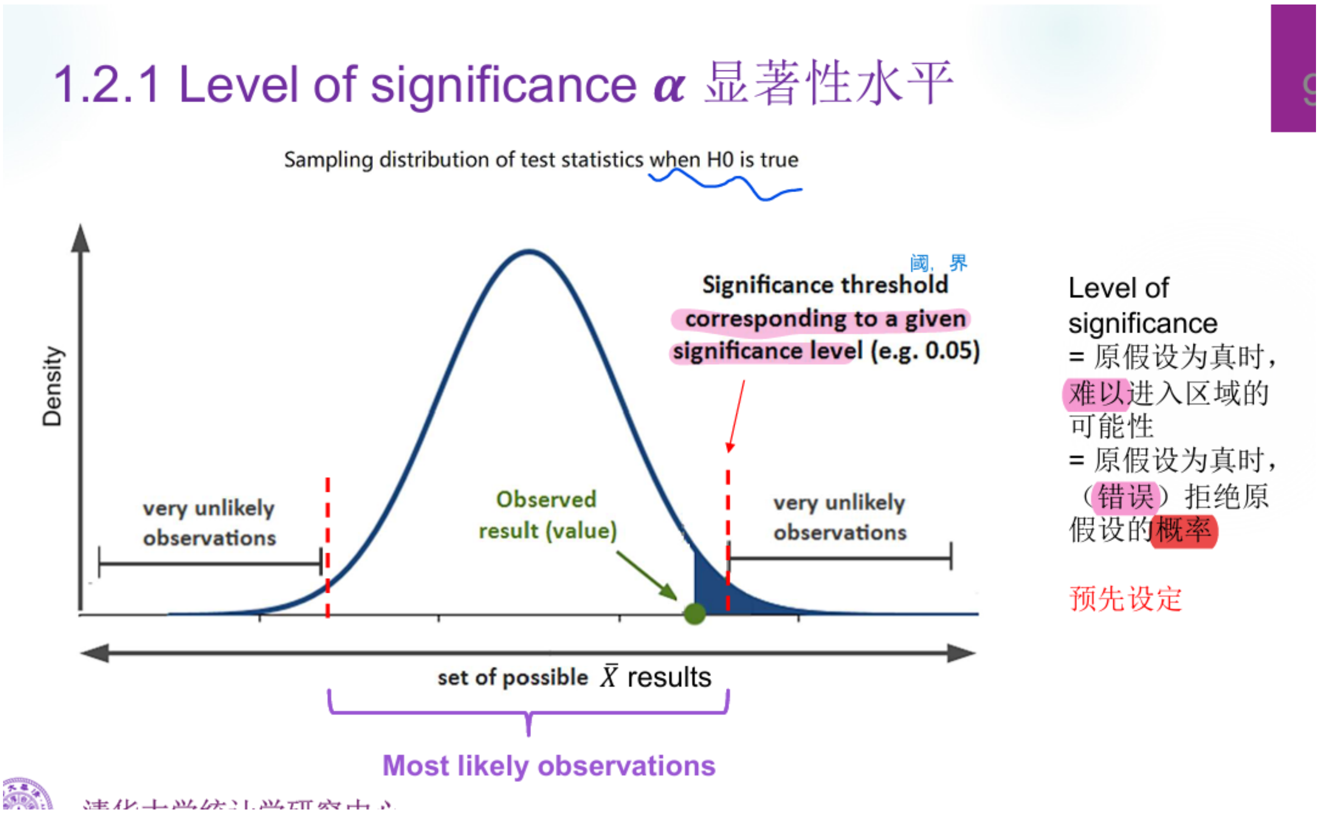
Population Distribution	Sample Size	Population Variances	Test Statistic
Normal	Any	Known	$z_{obs} = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
	Any	unknown assume $\sigma_1^2 = \sigma_2^2$, $df = n_1 + n_2 - 2$	$t_{obs} = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}$
	Any	unknown assume $\sigma_1^2 \neq \sigma_2^2$, $df = \nu$	$t_{obs} = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

5.2 Summary: Hypothesis tests for a difference of two means H0 : μ1 − μ2 = μ0, Ha : μ1 − μ2 ≠ μ0

Population Distribution	Sample Size	Population Variances	Test Statistic
Not Normal/ Unknown	Large	Known	$z_{obs} = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
	Large	unknown	$z_{obs} = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$
	small	Any	Nonparametric Methods

level of significance显著性水平

α是概率，在图像中表示为面积

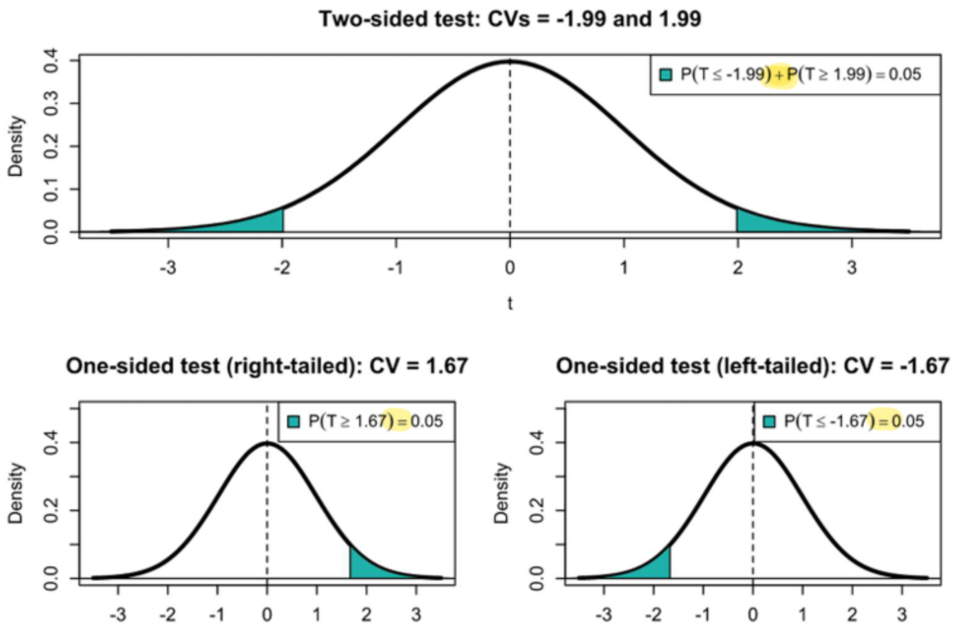


Two-sided test vs. one-sided test

- α的分配不同

1. **Two-sided test**: *Is the average cholesterol level of patients from this population* ***different*** *from 5.0 mmol/L?*
   - $H_0: \mu = 5 \text{ versus } H_1: \mu \neq 5$
2. **One-sided test**: *Is the average cholesterol level of patients from this population* ***greater than*** *5.0 mmol/L?*
   - $H_0: \mu = 5 \text{ versus } H_1: \mu > 5$
3. **One-sided test**: *Is the average cholesterol level of patients from this population* ***less than*** *5.0 mmol/L?*
   - $H_0: \mu = 5 \text{ versus } H_1: \mu < 5$

1.3 Two-sided test vs. one-sided test (2)



p-value

- 有多离谱
- p-value depends on the null hypothesis

1.4 P-values

- Two results in hypothesis testing: reject or ~~accept~~ do not reject H_0, but we don't know how strong the evidence of the rejection or non-rejection.
- Example
  - Let $X_1, \cdots, X_{10}$ be a random sample from $N(\mu, 1)$. Test the hypothesis
$$H_0: \mu = 0 \text{ vs. } H_a: \mu \neq 0$$
  - Take the level of significance $\alpha = 0.05$ and the rejection region
$$D = \{(X_1, \cdots, X_{10}): |\sqrt{10}\bar{X}| > 1.96\} = \{(X_1, \cdots, X_{10}): |\bar{X}| > 0.62\}$$
  - For one sample, $\bar{X} = 0.48$, do not reject H_0
  - For another sample, $\bar{X} = 0.12$, do not reject H_0

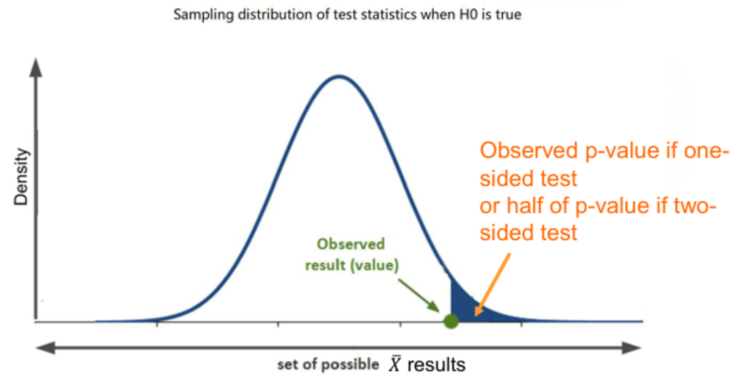
?

You need P-value

1.4 P-values (2)

$\sqrt{10}$ 是什

- Assuming H_0 *is true* (当H_0为真时): $U = \sqrt{10}\bar{X} \sim N(0,1)$, $\bar{X} \sim N(0, 1/10)$
- Let $\bar{x}_0$ be the observed value of $\bar{X}$



$$\boldsymbol{p = P(|\bar{X}| > |\bar{x}_0| \mid H_0)} = P\left(\sqrt{10}|\bar{X}| > \sqrt{10}|\bar{x}_0| \mid H_0\right) = P\left(|U| > \sqrt{10}|\bar{x}_0| \mid H_0\right)$$

- 在图像边缘

1.4 P-values (3)

假设检验的p值是假设H0为真，检验统计量至少与观察到的检验统计量一样极端的概率

- ***P*-value** for a hypothesis test is the probability of the test statistic being at least as extreme as the observed test statistics, *assuming* H_0 *is true.*

- P-value is another way of looking at a test in an effort to assess how strong (or weak) the rejection of a hypothesis (证据的充分程度)

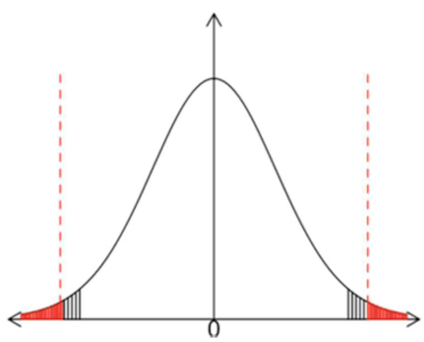
- P-value depends on the **null hypothesis**

- 双边检验的pvalue要算两份

1.4 P-values (4)

- Two-side test : H_0: $\theta = \theta_0 \leftrightarrow H_1: \theta \neq \theta_0$, rejection region $R = \{|T(\boldsymbol{X})| > c\}$, defind P-value $p = P(|T(\boldsymbol{X})| > |t(\boldsymbol{x})| \mid H_0)$

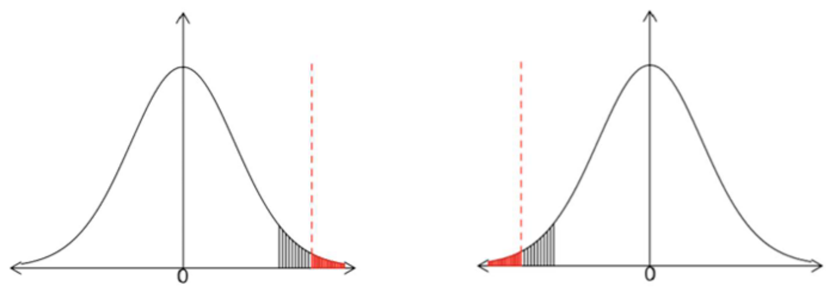
P-value of two side U-test



1.4 P-values (5)

- One side test: H_0: $\theta \le \theta_0 \leftrightarrow H_1$: $\theta > \theta_0$, rejection region $R = \{T(\boldsymbol{X}) > c\}$, define the P-value $p = P(T(\boldsymbol{X}) > t(\boldsymbol{x}) \mid H_0)$
- One side test: H_0: $\theta \ge \theta_0 \leftrightarrow H_1$: $\theta < \theta_0$, rejection region $R = \{T(\boldsymbol{X}) < c\}$, define the P-value $p = P(T(\boldsymbol{X}) < t(\boldsymbol{x}) \mid H_0)$

P-value of one side U-test



应用

1.4 P-values (6)

- **Small p-value indicates**:
  - ➢ Seeing a very rare event (due to chance)
  - ➢ or H_0 is not true
  - ➢ 原则：小概率事件在一次试验中是几乎不可能发生。 因此, **H0** 非真更有可能。
- **Decision rule:** Reject H_0 if P-value $\le \alpha$ level of significance, the difference is statistically significant
- The smaller the P-value, the stronger rejection of the null hypothesis and vice versa, e.g., P-value = 0.01 vs. P-value = 1e-8
- Strong rejection of the null hypothesis is also referred to as the result being *highly statistically significant*
- The P-value is NOT the probability that the null hypothesis is true.

P-value ≠ P(null hypothesis|data)

注意点

```
![[Pasted image 20240420092001.png]]
```

1.4 P-values (8)

使用和解释*p*值的原则。

1. ***p*值可以表示数据与一个特定的统计模型是否相容；**
   例如在原假设通常用来假设一个效应不存在，如两组之间没有差异，两个因素没有相关性。此时*p*值越小，数据与原假设的不相容性(incompatibility)越大，可以解释为这些数据怀疑或否定了原假设。
2. ***p*值不能代表假说为真的概率，也不代表数据完全是由随机因素造成的概率；**
   *p*值是所得数据与解释之间关系的说明，而不是对解释本身的说明。
3. **科研结论、商业决定和政策制定不能完全凭*p*是否小于一个特定的值来决定；**
   重大决策与结论中，需要考虑诸多因素，如实验设计、数据质量、外部证据、假设的合理性等等，不能只由*p*值决定Yes or No的问题。
4. **正确的推理需要全面的报告和透明度；**
   正确的科学推理，需要研究者公布研究中包含的所有假设，所有数据收集的决定，所有进行的统计分析和所有*p*值。
5. **一个*p*值，或者显著性，不能表示一个效应的大小，或者一个结果的重要性；**
   *p*值大小不代表效应大小。再微小的效应，达到一定的样本量和测量精度，都能得到小的*p*值；再大的效应，在样本量和测量精度不那么高的时候，也可能只能得到普普通通的*p*值。
6. ***p*值本身不能作为判断一个模型或假说的良好量度。**
   单独的*p*值只能提供有限信息。用一个略小于0.05的*p*值来拒绝原假设就难以有说服力；相反，一个相对较大的*p*值也不能说就赞成原假设。当有其它方法可选时，数据分析不应该以一个简单的*p*值计算作为结束。

Type I and II errors

There are two mistakes we can make in a hypothesis test.

- Type I error: H_0 is rejected but in reality H_0 is true False positive 假阳性
- Type II error: H_0 is not rejected but in reality H_0 is false False negative 假阴性

Type I error
(false positive)

Type II error
(false negative)

	Do not reject H_0	Reject H_0
H_0 is true	No Error	Type I error
H_0 is false	Type II error	No Error

1.5.1 Control risks (1)

- The probability of type I error is denoted by α (same as the level of significance), i.e., 一类错误发生率

$$\alpha = \text{Prob}\,(\text{reject } H_0 | H_0 \text{ is true}).$$

- The type II error is denoted by β, i.e., 二类错误发生率

$$\beta = \text{Prob}\,(\text{fail to reject } H_0 | H_a \text{ is true at some value}).$$

- The ideal situation is both type I and type II error are 0, which means we can always make correct decision. However, the reality is that every decision we make will have associated error probability.

1.5.1 Control risks (2)

- In practice, if we try to decrease the type I error, the type II error will increase, and vice versa.
- Researchers should consider the **consequences** of type I error and type II errors to help determine significance level.

【讨论】 以下两种诊断的两类错误分别会带来什么后果，哪一种错误比较严重？

EX1. 癌症诊断

EX2. 新冠诊断

α-pre assigned

1.5.1 Control risks (3)

25

In general, choose a test such that

First Step: the probability of type I error is controlled at a pre-assigned level of significance α

Second Step: the probability of type II error is made as small as possible (it has maximum power among all tests with level of significance $\leq \alpha$)

Type I error is controlled at level α, e.g., $\alpha = 0.05$, i.e., the test with significance level α will have type I error $\leq \alpha$

Steps to a hypothesis testing-单样本

1. State the null and alternative hypotheses
2. Collect data and calculate test statistic assuming H0 is true
3. Assuming the H0 is true, calculate the p-value (commonly used) or rejection region
4. Draw conclusion. We either reject H0 or do not reject H0.

Step 1: hypothesis

Step 2: Calculate testing statistic

- check

▶ Check $np_0 = 300 * 0.2 > 10$ and $n(1 - p_0) > 10$. If so, we can use large sample approximation.

-
- 计算z0，注意这里用的是p0（因为前提是H0=T，所以这个已经给定的（假设的）p0是已知的了）

Assuming H0 is true, the test statistic is

$$z_0 = \frac{\hat{p} - p_o}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{50/300 - 0.20}{\sqrt{\frac{50/300(1-50/300)}{300}}} = -1.55$$

-

Step 3: Calculate p-value

P-value = P(Z< -1.55) = 0.061

Alternative hypothesis	Hypothesis type	p-value formula
$H_a : p < p_0$	Left-tail hypothesis	$P(Z < z_0)$
$H_a : p > p_0$	Right-tail hypothesis	$P(Z > z_0)$
$H_a : p \neq p_0$	Two-tail hypothesis	$2P(Z < -\lvert z_0 \rvert)$

```
> prop.test(x=50, n=300,p=0.20, alternative = "less",correct=FALSE)

	1-sample proportions test without continuity correction

data:  50 out of 300, null probability 0.2
X-squared = 2.0833, df = 1, p-value = 0.07446
alternative hypothesis: true p is less than 0.2
95 percent confidence interval:
 0.0000000 0.2050048
sample estimates:
        p
0.1666667
```

NOTE: This is not test based on normal approximation, but gives the same direction (Chi-square test).

Step 4: Conclusion

比较pvalue和α

- We use α to denote the level of significance.
- Level of significance is determined before you see the data.
- In practice, we usually set α = 0.05. Other common choices are α = 0.01, or α = 0.1.
- We simply compare the p-value with the α level, we reject H0 if the p-value is less than α; and do not reject H0 is the p-value is greater than or equal to α.

另一种方法-Confidence Interval Approach

Recall that a $100(1-\alpha)\%$ C.I. is

$$\left(\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad \hat{p} + z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right).$$

$\hat{p} = 0.167$, SE = $\sqrt{0.167 * 0.833/300} = 0.0215$

95% CI is $(0.167 \pm 1.96 * 0.0215) = (0.125, 0.209)$

Conclusion?

Hypothesis tests for a difference of two proportions (two independent samples)-大样本

Step 1: hypothesis

1. Write hypotheses.

H0: $p_a - p_c = 0$

Ha: $p_a - p_c \neq 0$

Step 2: Calculate testing statistic

- check

2. Check any necessary assumptions

$$n_a\widehat{p}_a = 124$$

$$n_a(1-\widehat{p}_a) = 240 - 124 = 116$$

$$n_c\widehat{p}_c = 90$$

$$n_c(1-\widehat{p}_c) = 260 - 90 = 170$$

These counts are all at least 10 so we can use the normal approximation method.

-

3. Calculate an appropriate test statistic.

Pooled estimate $\widehat{p} = \dfrac{\widehat{p}_1 n_1 + \widehat{p}_2 n_2}{n_1 + n_2}$ $\qquad \widehat{p} = \dfrac{124+90}{240+260} = \dfrac{214}{500} = 0.428$

Standard Error of $\hat{p}_1 - \hat{p}_2$: $SE_0 = \sqrt{\frac{\widehat{p}(1-\widehat{p})}{n_1} + \frac{\widehat{p}(1-\widehat{p})}{n_2}} = \sqrt{\widehat{p}(1-\widehat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$

$$SE_0 = \sqrt{0.428(1-0.428)\left(\frac{1}{240}+\frac{1}{260}\right)} = 0.04429$$

Test Statistic for Two Independent Proportions

$$z = \frac{\widehat{p}_1 - \widehat{p}_2}{SE_0} \qquad z = \frac{\frac{124}{240} - \frac{90}{260}}{0.04429} = 3.850$$

Under H0, $p_1 = p_2 = p$
(i. e., unknown proportion)
Then $\hat{p}_1 \sim AN\left(p, \frac{p(1-p)}{n_1}\right)$
$\hat{p}_2 \sim AN\left(p, \frac{p(1-p)}{n_2}\right)$.
Since sample from adult and sample from children are independent,
Var($\hat{p}_1$- $\hat{p}_2$)= $\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}$. But p is unknown, we estimated by pooled estimate $\hat{p}$, which gives us the SE($\hat{p}_1$- $\hat{p}_2$).

- 清华大学统计学研究中心
-

Step 3: Calculate p-value

4. Calculate p-value associated with the test statistic and make conclusions.

Our p-value will be the area of the z distribution more extreme than 3.850.

$p = 0.0000591 \times 2 = 0.0001182$ < 0.05

Reject the null hypothesis.

In the population of ice cream customers, the proportion of adults who prefer cones is different from the proportion of children who prefer cones in the population.

Step 4: Conclusion

Small sample proportions-Exact binomial test

step1 hypothesis

step2 calculate testing statistic

- check

  Since 29*0.10 < 5, we cannot use normal approximation in this case.

  Instead, the exact binomial test is more appropriate.

  NOTE: Exact test can be used when sample size is large!

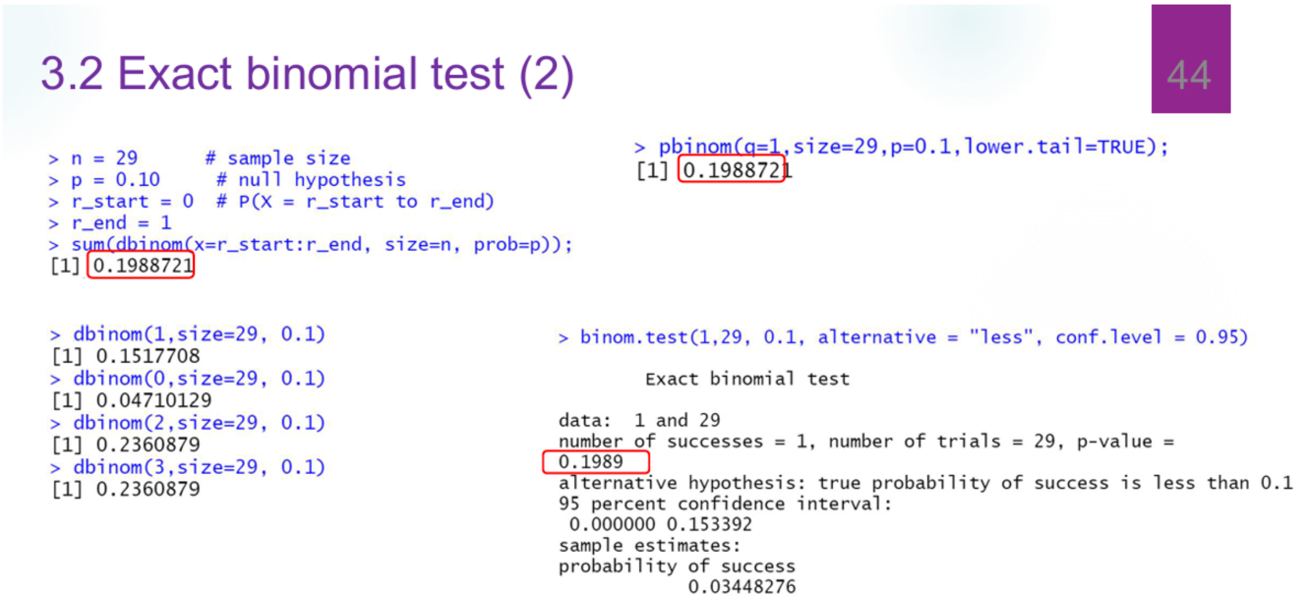
  The sample's proportion $\hat{p} = \frac{x_0 \text{ (from our sample)}}{n}$ =

  The sampling distribution of $\hat{p}$ is complicated. The exact binomial test uses the "method of *p*-values".

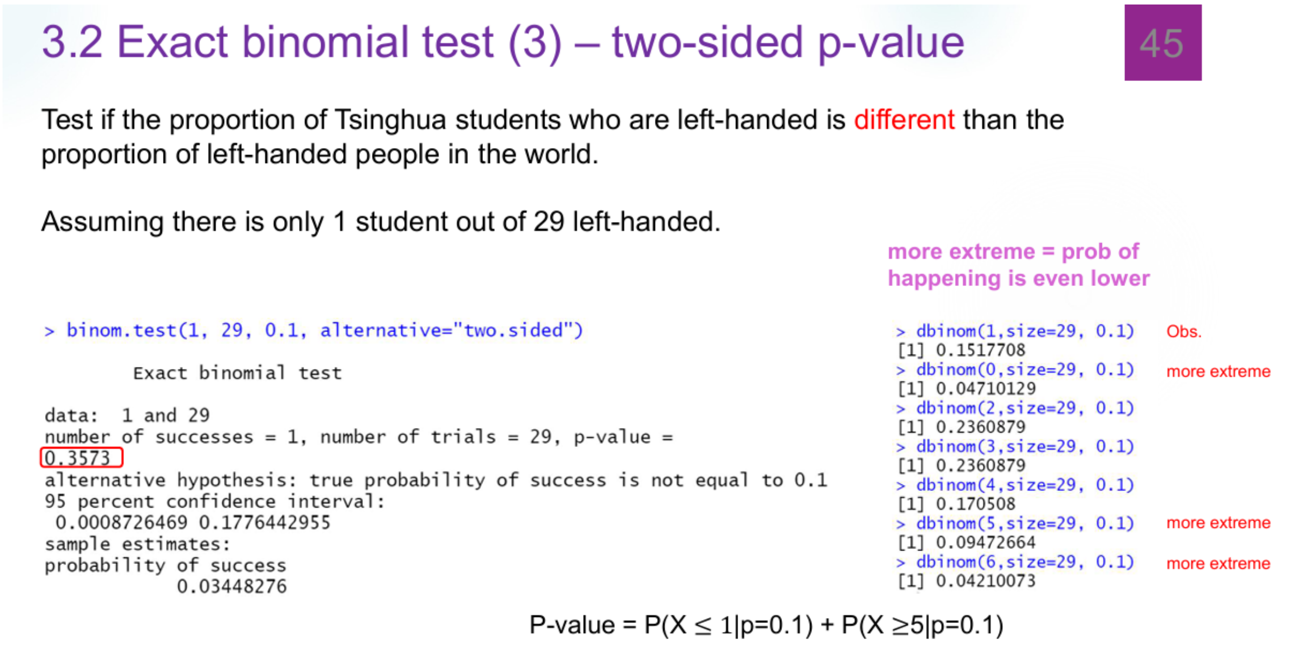
  $$p\text{-values} = P(X \le x_0 | p=0.1) = \sum_{k=0}^{x_0} C_n^k p^k (1-p)^{n-k}$$

-

exact binomial test



step3 calculate p-value



step4 conclusion

Two-proportion difference – small sample scenario

We will discuss in the non-parametric testing section.

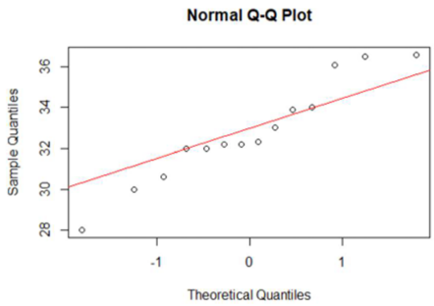
Hypothesis test for a single mean-大样本

- check

32.2	32.2	32.3	30.0

1. Check normality:

```
> wings<-c(33.9, 33.0, 30.6, 36.6, 36.5, 34.0, 36.1, 32.0, 28.0,
+          32.0, 32.2, 32.2, 32.3, 30.0)
> qqnorm(wings); qqline(wings, col = 2)
```



清华大学统计学研究中心

- Set up a two-sided test of

$$
\begin{aligned}
H_0 &: \mu = 3000 \\
\text{vs. } H_a &: \mu \neq 3000
\end{aligned}
$$

- Let $\alpha = 0.05$ denote a 5% significance level

- Calculate the test statistic:

$$z_{obs} = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{2500 - 3000}{1000/\sqrt{10}} = -1.58$$

- What does this mean? Our observed mean is 1.58 standard errors below the hypothesized mean
- The test statistic is the standardized value of our data assuming the null hypothesis is true!
- Question: If the true mean is 3000 grams, is our observed sample mean of 2500 "common" or is this value unlikely to occur?

4.1 Hypothesis test for a single mean - Example

- Based on our significance level ($\alpha = 0.05$) and assuming H_0 is true, how "far" does our sample mean have to be from $H_0 : \mu = 3000$ in order to reject?
- Critical value $= z_c$ where $2 \times P(Z > |z_c|) = 0.05$
- In our example, $z_c = 1.96$
- The rejection region is any value of our test statistic that is less than -1.96 or greater than 1.96
- Decision should be the same whether using the p-value or critical / rejection region

- An alternative approach for the two sided hypothesis test is to calculate a 100(1-α)% confidence interval for the mean
- We are 95% confident that the interval (1880, 3120) contains the true population mean μ

$$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{10}} \rightarrow 2500 \pm 1.96\frac{1000}{\sqrt{10}}$$

- The hypothetical true mean 3000 is a plausible value of the true mean given out data
- We cannot say that the true mean is different from 3000

Hypothesis tests for a difference of two means (two independent samples)-大样本

- check

5.1 Example: Hypothesis tests for a difference of two means – check assumptions and conditions

① **Independence within groups:**

- Both the smoker and the nonsmoker mothers are randomly sampled.
- 50 < 10% of all smoker mothers and 100 < 10% of all non-smoker mothers. ?

We can assume that the birth weight of a baby born to a smoking mother is independent of another one born to a smoking mother, and the birth weight of a baby born to a nonsmoking mother is independent of another one born to such mother as well.

② **Independence between groups**

Mothers who smoked during pregnancy are independent of mothers who did not smoke.

③ **Normality:**

The distributions are somewhat skewed but the sample sizes are at least 50. The sampling distribution of the average difference will be nearly normal as well.

-

H0 : $\mu_s = \mu_{ns}$.

Ha : $\mu_s \neq \mu_{ns}$.

5.1 Example: Hypothesis tests for a difference of two means – sampling distribution

Standard error of the difference between two sample means (large sample and non-normality scenario)

$$SE_{(\bar{x}_1-\bar{x}_2)} = \sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$$

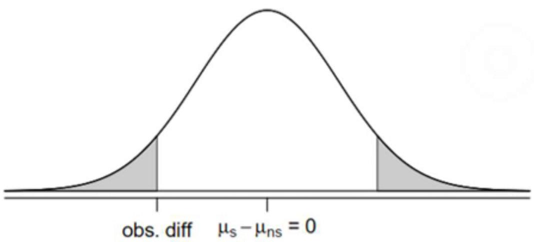
$$\begin{aligned} SE_{(\bar{x}_s-\bar{x}_{ns})} &= \sqrt{\frac{s_s^2}{n_s}+\frac{s_{ns}^2}{n_{ns}}} \\ &= \sqrt{\frac{1.43^2}{50}+\frac{1.60^2}{100}} \\ &= 0.26 \end{aligned}$$

Recall: The variance of the difference between two independent random variables is the sum of their variances.

The sampling distribution:

$$(\bar{x}_s - \bar{x}_{ns}) \sim N\left((\mu_s - \mu_{ns}) = 0, SE_{\bar{x}_s-\bar{x}_{ns}} = 0.26\right)$$

$$\begin{aligned} p - value &= P\left(|\bar{x}_s - \bar{x}_{ns}| > 0.04 \mid (\mu_s - \mu_{ns}) = 0\right) \\ &= P\left(|Z| > \frac{0.04-0}{0.26}\right) \\ &= P(|Z| > 1.54) \\ &= 2 \times (1 - 0.9382) \\ &= 0.1236 \end{aligned}$$



Paired Samples

- check

- Conduct a one-sided test at the α= 0.05 level of significance.
- The null hypothesis is $\mu_1 = \mu_2$
- and the alternative is $\mu_1 \neq \mu_2$
- Use these data to create a new set of observations that represent the differences within each pair:

$$\begin{aligned} d_1 &= x_{11} - x_{12} \\ d_2 &= x_{21} - x_{22} \\ d_3 &= x_{31} - x_{32} \\ &\vdots \\ d_n &= x_{n1} - x_{n2} \end{aligned}$$

n pairs

- Remember to check the normality of the difference!

5.4 Paired Samples – nothing new

- The mean of the set of differences is $\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$;
- This sample mean provides a point estimate for the true difference in population means $\mu_1 - \mu_2$.
- The standard deviation of the differences is $s_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}}$.
- Denote $\delta = \mu_1 - \mu_2$
- The null hypothesis is $H_0: \delta = 0$

  and the alternative as $H_A: \delta \neq 0.$
- $t = \frac{\bar{d} - \delta}{s_d / \sqrt{n}}$; If the null hypothesis is true, this quantity has a t-distribution with n - 1 degrees of freedom.

- $\bar{x}_{diff}$: sample mean of the differences = **-0.95**

  s: sample standard deviation of the differences = **1.317**

  n: sample size (i.e. number of pairs) = **20**
- **t =** $\bar{x}_{\text{diff}} / (\text{sdiff}/\sqrt{n}) = -0.95 / (1.317/\sqrt{20})$ = -3.226 df = 19
- P-value = 0.00445
- Since this p-value is less than our significance level α = 0.05, we reject the null hypothesis. We have sufficient evidence to say that the mean max vertical jump of players is different before and after participating in the training program.