生物统计（1）假设检验，方差分析

俞章盛

教授　生命学院生物信息与生物统计系
客座教授　耶鲁大学生物统计系

Outline

- 生物统计的作用
- 假设检验
- ANOVA

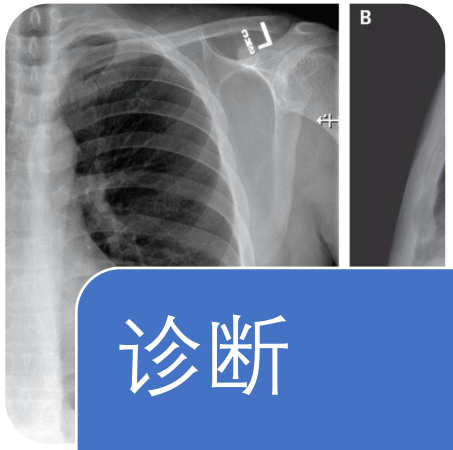
Outline

- 生物统计的作用
- 假设检验
- ANOVA

新冠肺炎

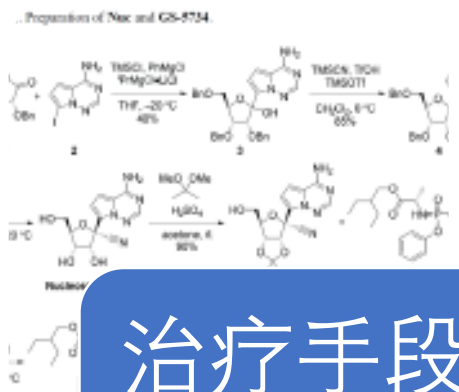
诊断

- 核酸检测
- 评价诊断



诊断

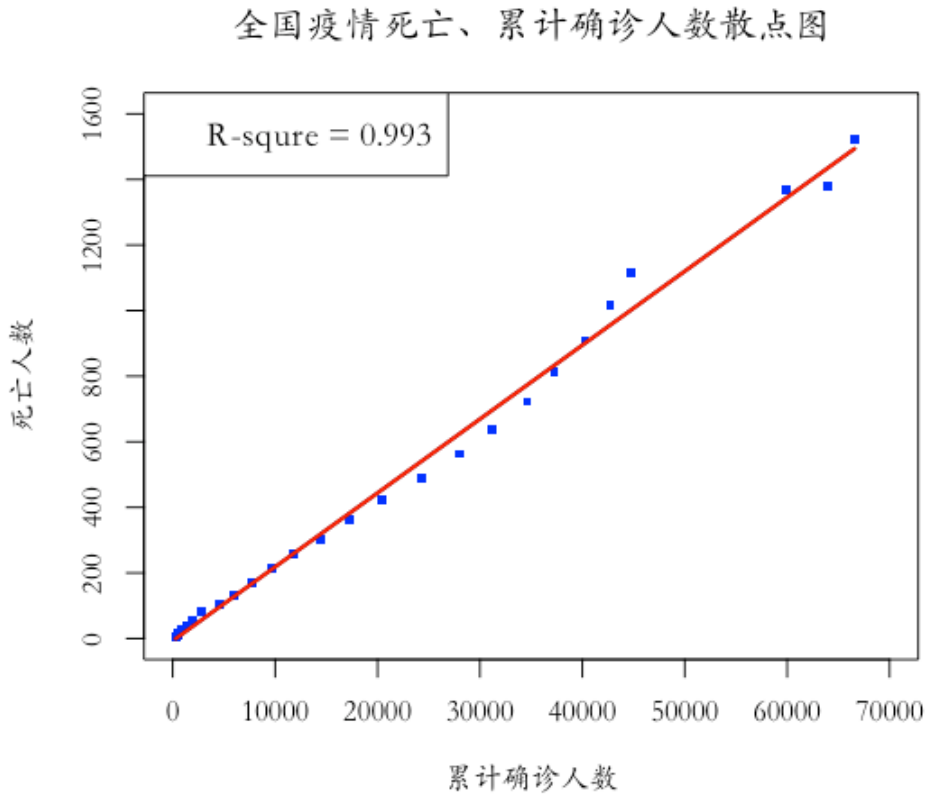
- 影像
- 评价诊断



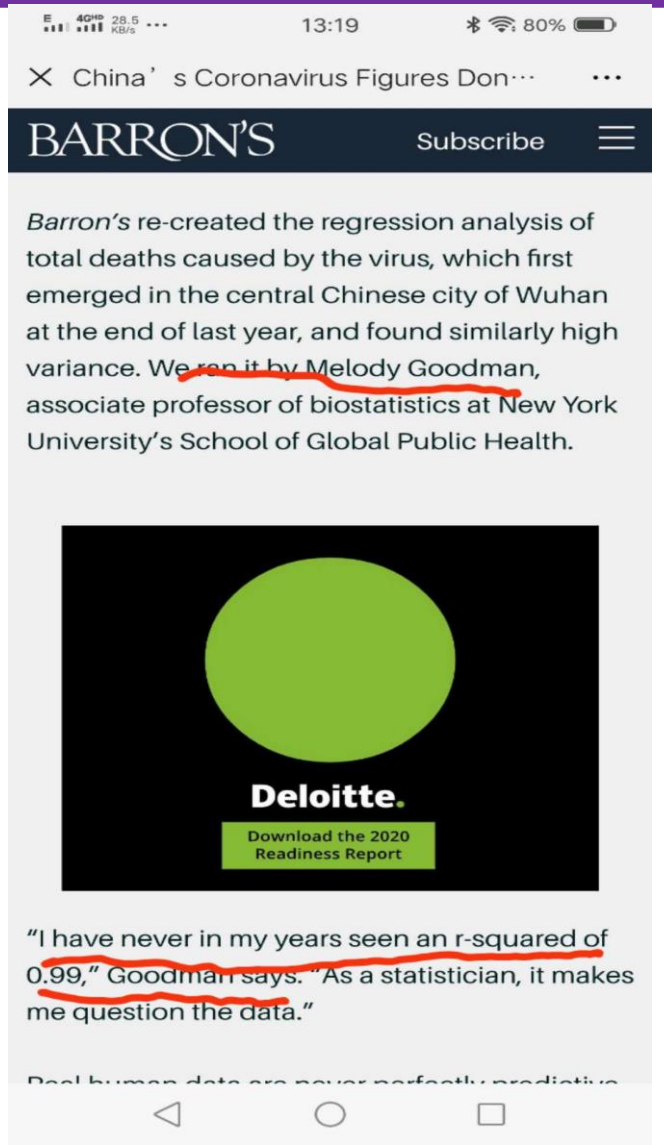
治疗手段

- 瑞德希韦
- 疗效评价

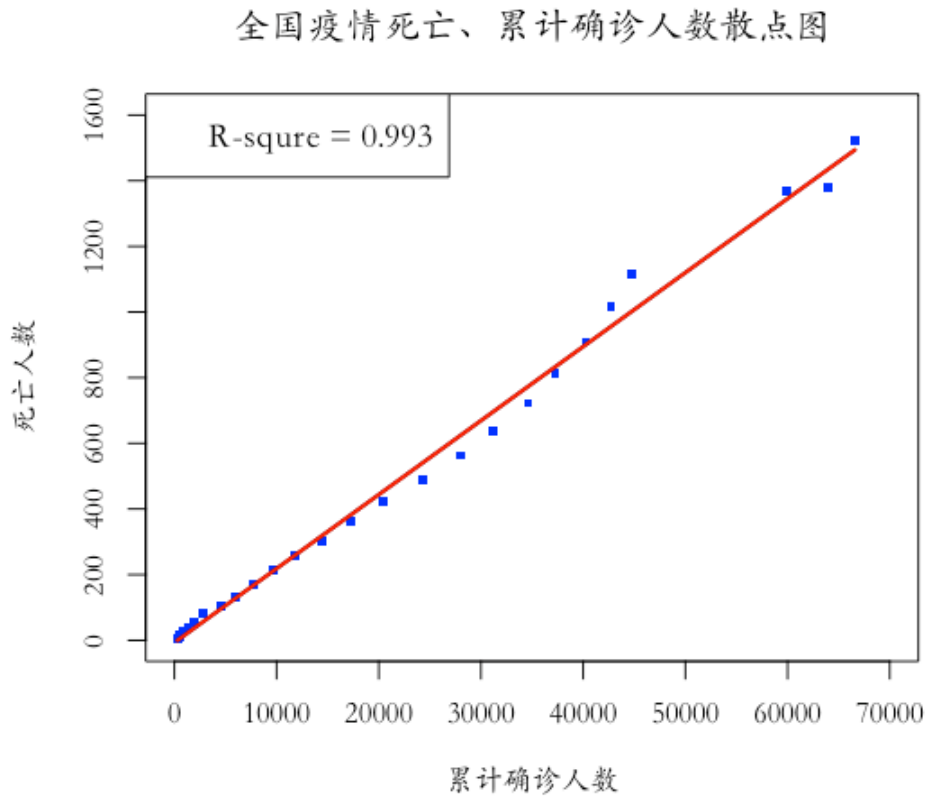
如何理解累计数据

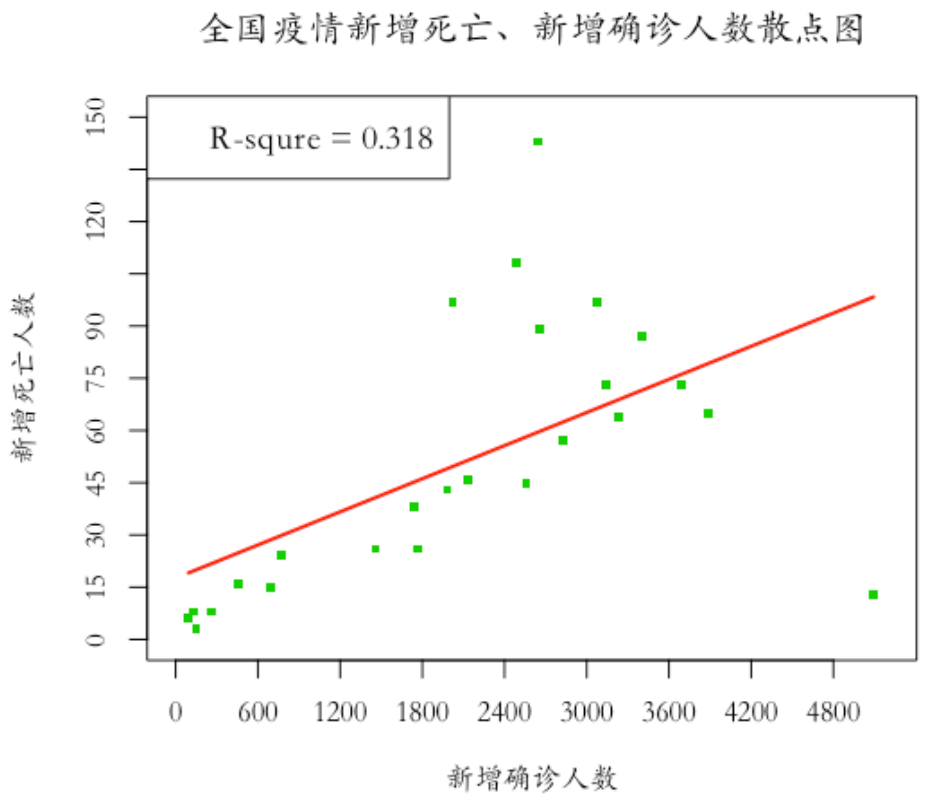


R square=1表示近似于完美，
too good to be true?



如何理解累计数据





R square=1表示近似于完美，
too good to be true?

如何理解统计数据

"Essentially, all models are wrong, but some are useful."
--- Box, George E. P.; Norman R. Draper

1. 产生一千个数据集
2. 每个数据集有100个新增死亡和新增确诊的线性回归关系。他们的R-square=0.504;
3. 在每个数据集中获得累积死亡和累计确诊;
4.计算累计死亡，累计确诊的R-Square.

新增死亡与新增确诊模型
R-square=0.504

累计死亡与累计确诊模型
R-square=0.999

累计变量在每个观测点上是不独立的，这与一遍回归分析模型的假设相违背

如何评价药效

	返美	工作	工作	在家	急诊	住院										
	1月15日	1月16日	1月17日	1月18日	1月19日	1月20日	1月21日	1月22日	1月23日	1月24日	1月25日	1月26日	1月27日	1月28日	1月29日	1月30日
发病日	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
体温（摄氏度）			自觉发烧	自觉发烧	37.2	37.9	39	39.4	39.1	39.4	38.8	39.4	37.3	36.8	36.8	36.3
咳嗽																
流鼻涕																
疲劳																
恶心																
呕吐																
腹泻																
腹部不适																
启用remdesivir																
住院天数						1	2	3	4	5	6	7	8	9	10	11

人民的希望真的有效吗？？

On hospital day 8 (illness day 12), the patient's clinical condition improved. Supplemental oxygen was discontinued, and his oxygen saturation values improved to 94 to 96% while he was breathing ambient air. The previous bilateral lower-lobe lesions

关于药效的信号是否是真实存在的？

- Holshue et al. (2020 NEJM)

如何评价药效-随机化临床试验

- Severe 2019-nCoV Remdesivir Randomized Clinical Trial(ClinicalTrials.gov Identifier: NCT04257656)

如何评价药效-随机化临床试验

- A Trial of Lopinavir–Ritonavir in Adults Hospitalized with Severe Covid-19 (NEJM 2020)
- A randomized, controlled, open-label trial involving hospitalized adult patients with confirmed SARS-CoV-2 infection
- Experimental drug: Standard care+ lopinavir+ritonavir (400 mg and 100 mg, respectively) twice a day for 14 days
- Control: Standard care.
- 199病人随机分配到试验组（99）和对照组（100）
- 比较Time to Clinical Improvement (TTCI) [Censored at Day 28]
- Improvement: 2级， 6death, 5 ECMO, 4 ICU, 3 Hospitalization (Oxy), 2 Hosp(no Oxy), 1 Discharge

Patients' Characteristics

Table 1. Demographic and Clinical Characteristics of the Patients at Baseline.*

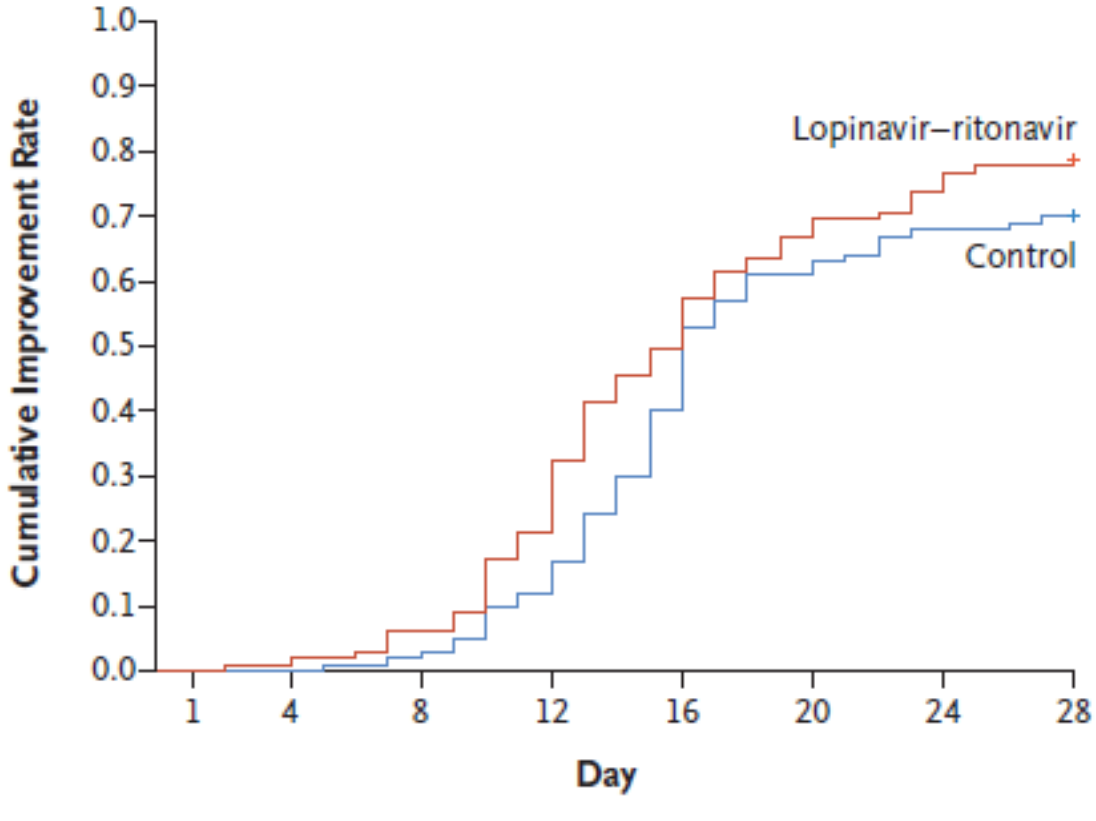
Characteristic	Total (N=199)	Lopinavir–Ritonavir (N=99)	Standard Care (N=100)
Age, median (IQR) — yr	58.0 (49.0–68.0)	58.0 (50.0–68.0)	58.0 (48.0–68.0)
Male sex — no. (%)	120 (60.3)	61 (61.6)	59 (59.0)
Coexisting conditions — no. (%)			
Diabetes	23 (11.6)	10 (10.1)	13 (13.0)
Cerebrovascular disease	13 (6.5)	5 (5.1)	8 (8.0)
Cancer	6 (3.0)	5 (5.1)	1 (1.0)
Body temperature, median (IQR) — °C	36.5 (36.4–36.8)	36.5 (36.4–37.0)	36.5 (36.5–36.8)
Fever — no. (%)	182 (91.5)	89 (89.9)	93 (93.0)
Respiratory rate >24/min — no. (%)	37 (18.8)	21 (21.6)	16 (16.0)
Systolic blood pressure <90 mm Hg — no. (%)	2 (1.0)	2 (2.0)	0
White-cell count ($\times 10^{-9}$/liter) — median (IQR)	7.0 (5.1–9.4)	7.3 (5.3–9.6)	6.9 (4.9–9.1)
4–10 $\times 10^{-9}$/liter — no. (%)	137 (70.3)	64 (67.4)	73 (73.0)
<4 $\times 10^{-9}$/liter — no. (%)	20 (10.3)	12 (12.6)	8 (8.0)
>10 $\times 10^{-9}$/liter — no. (%)	38 (19.5)	19 (20.0)	19 (19.0)

Outcomes

Table 3. Outcomes in the Intention-to-Treat Population.*

Characteristic	Total (N=199)	Lopinavir–Ritonavir (N=99)	Standard Care (N=100)	Difference†
Time to clinical improvement — median no. of days (IQR)	16.0 (15.0 to 17.0)	16.0 (13.0 to 17.0)	16.0 (15.0 to 18.0)	1.31 (0.95 to 1.80)‡
Day 28 mortality — no. (%)	44 (22.1)	19 (19.2)§	25 (25.0)	−5.8 (−17.3 to 5.7)
Earlier (≤12 days after onset of symptoms)	21 (23.3)	8 (19.0)	13 (27.1)	−8.0 (−25.3 to 9.3)
Later (>12 days after onset of symptoms)	23 (21.1)	11 (19.3)	12 (23.1)	−3.8 (−19.1 to 11.6)
Clinical improvement — no. (%)				
Day 7	8 (4.0)	6 (6.1)	2 (2.0)	4.1 (−1.4 to 9.5)
Day 14	75 (37.7)	45 (45.5)	30 (30.0)	15.5 (2.2 to 28.8)
Day 28	148 (74.4)	78 (78.8)	70 (70.0)	8.8 (−3.3 to 20.9)
ICU length of stay — median no. of days (IQR)	10 (5 to 14)	6 (2 to 11)	11 (7 to 17)	−5 (−9 to 0)
Of survivors	10 (8 to 17)	9 (5 to 44)	11 (9 to 14)	−1 (−16 to 38)
Of nonsurvivors	10 (4 to 14)	6 (2 to 11)	12 (7 to 17)	−6 (−11 to 0)
Duration of invasive mechanical ventilation — median no. of days (IQR)	5 (3 to 9)	4 (3 to 7)	5 (3 to 9)	−1 (−4 to 2)
Oxygen support — days (IQR)	13 (8 to 16)	12 (9 to 16)	13 (6 to 16)	0 (−2 to 2)
Hospital stay — median no. of days (IQR)	15 (12 to 17)	14 (12 to 17)	16 (13 to 18)	1 (0 to 2)
Time from randomization to discharge — median no. of days (IQR)	13 (10 to 16)	12 (10 to 16)	14 (11 to 16)	1 (0 to 3)
Time from randomization to death — median no. of days (IQR)	10 (6 to 15)	9 (6 to 13)	12 (6 to 15)	−3 (−6 to 2)

Outcomes



No. at Risk

Lopinavir–ritonavir	99	98	93	78	50	33	26	22
Control	100	100	98	88	60	39	32	30

Figure 2. Time to Clinical Improvement in the Intention-to-Treat Population.

生物统计学的作用

- 新药研发中的药效评价
- 诊断试剂，医疗器械准确性评价(新冠病毒的核酸检测)
- 临床研究设计，分析
- 健康大数据挖掘

Outline

- 生物统计的作用
- 假设检验
- ANOVA

Introduction

- Hypothesis Testing
  - Decide about the value of a parameter based on some preconceived idea and collected data.

- Example: Suppose you want to test if Drug A is better or not in lowering deaths in cancer patients compared to the Drug B. You need to decide on two possibilities

  - There is no difference between the drugs. Or nothing important is happening. This is called the **null hypothesis**.

  - There is a difference between the drugs. Or there is something important happening here. This is called the **alternative hypothesis**.

How does Hypothesis Testing Work?

- Similar to a courtroom trial. In trying a person for a crime, the jury needs to decide between one of two possibilities:
  - The person is innocent.
  - The person is guilty.
- To begin with, the person is assumed innocent.
  - Null hypothesis: the person is innocent
- The prosecutor presents evidence, trying to convince the jury to reject the original assumption of innocence, and conclude that the person is guilty.

Casey Anthony Trial

- Casey Anthony is a Florida woman who was acquitted of killing her 2-year old daughter.
- Two hypothesis for the trial
  - Casey Anthony is innocent.
  - Casey Anthony is guilty.
- The prosecutor uses evidence in form of DNA, witnesses, and so forth.
- In the trial, the prosecutors couldn't disprove the null hypothesis that she is innocent with the evidence they had. So the jurors had no choice but to accept the null hypothesis that she is not guilty.
  - Juror speaks on ABC News: http://www.youtube.com/watch?v=Npei98z8EE4&feature=relmfu

Statistical Tests of Hypothesis

1. **The null hypothesis, H_0:**
   - Assumed to be true until we can prove otherwise.

2. **The alternative hypothesis, H_a:**
   - Will be supported if we can disprove H_0

Court trial:	Pharmaceuticals:
H_0: innocent	H_0: No difference between Drug A and Drug B
H_a: guilty	H_a: Difference between Drug A and Drug B

Statistical Tests of Hypothesis

3. The test statistic and its *p*-value

- **Test statistic**: A single statistic calculated from the sample which will allow us to reject or not reject H_0.
- **P-value**: A probability, calculated from the test statistic that measures whether the test statistic is **likely** or **unlikely**, assuming H_0 is true.

4. The rejection region:

- A rule that tells us for which values of the test statistic the null hypothesis should be rejected.

5. Conclusion:

- Either "Reject H_0" or "Do not reject H_0", along with a statement about the reliability of your conclusion.

Statistical Tests of Hypothesis

- How do you decide when to reject H_0?
  - Depends on the **significance level, α**, the maximum tolerable risk you want to have of making a mistake, if you decide to reject H_0.
  - Usually, the significance level is **α = .01** or **α = .05**.

- If p-value < **α, we will reject** H_0, and claim Ha is true.
- Note that "Not Reject H_0" does not mean H_0 is true
  - Not guilty ≠ innocent

Example

The mayor of a small city claims that the average income in his city is $35,000 with a standard deviation of $5000. We take a sample of 64 families, and find that their average income is

1-2. We want to test the hypothesis:

H_0: $\mu = 35{,}000$ (mayor is correct) versus

H_a: $\mu \neq 35{,}000$ (mayor is wrong)

Start by assuming that H_0 is true and $\mu = 35{,}000$.

Example

3. The best estimate of the population mean μ is the sample mean, \$30,000:
   - From the Central Limit Theorem the sample mean has an approximate normal distribution with mean μ = 35,000 and standard error SE = 5000/8 = 625.
   - The sample mean, \$30,000 lies z = (30,000 – 35,000)/625 = -8 standard deviations below the mean.
   - The probability of observing a sample mean this far from μ = 35,000 (assuming H_0 is true) is *nearly zero*.

Example

4. From the Empirical Rule, values more than three standard deviations away from the mean are considered extremely unlikely. Such a value would be extremely unlikely to occur if indeed H_0 is true, and would give reason to reject H_0.
5. Since the observed sample mean, \$30,000 is so unlikely, we choose to reject H_0: $\mu = 35{,}000$ and conclude that the mayor's claim is incorrect.

Results of Hypothesis Test

- Possible results when conduct a hypothesis test

	Population (Truth)	
Result of Test	$\mu = \mu_0$	$\mu \neq \mu_0$
Do Not Reject	Correct	Incorrect
Reject	Incorrect	Correct

- Significance Level (α): A predetermined probability

  $\alpha = P(\text{reject } H_0 \mid H_0 \text{ is true}) = \text{Type I error rate}$

  In most situations $\alpha = 0.05$ or $\alpha = 0.01$
- Interpretation of $\alpha = 0.05$: If the null hypothesis is true and a study is repeated a large number of times, the true null hypothesis will be rejected incorrectly 5% of the time.
- P(don't reject H_0 | Ha is true)= Type II error rate

单样本假设检验(阅读)

Two-Sided Hypothesis t-test for a Mean

- Example: The population of male industrial workers in London who have never experienced a major coronary event has mean systolic blood pressure (SBP) 136 mm Hg and mean diastolic blood pressure (DBP) 84 mm Hg. You might be interested in determining whether these values are the same as those for the population of industrial workers who have suffered a coronary event.

(a) A sample of 86 workers who suffered a coronary event

mean SBP 143 mm Hg, SD 24.4 mm Hg.

Test H_0: $\mu = 136$ mm Hg vs. H1: $\mu \neq 136$ mm (two-sided)

$\alpha = 0.05$ level

μ: mean SBP of the population of industrial workers who have suffered a coronary event.

One sample hypotheses: we only perform sampling from those with an event.

Two-Sided Hypothesis Test for μ

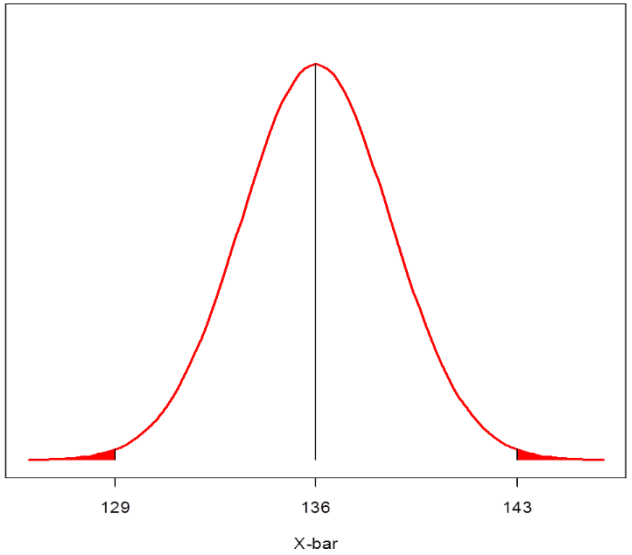
- Answer:

  H_0: $\mu = \mu_0 = 136$ mm Hg vs. H_A: $\mu \neq \mu_0$

  Sample mean: 143 mm Hg,

  Sample SD: 24.4 mm Hg,

  n=86



p-value=P(observing a difference between sample mean and μ_0 as extreme as or more extreme than the observed sample mean given the null hypothesis is true)

p-value=P(observing a difference between sample mean and μ_0 as extreme as or more extreme than 143 (against H_0) given H_0 is true)

Extreme Values Against H0 (Two-Sided)

Under the null hypothesis H_0: $\mu = \mu_0$, $t = \frac{\bar{X} - \mu_0}{s/\sqrt{n}}$ has t distribution with n-1 df, and it is the test statistic. The df=86-1=85.

H_0: $\mu = \mu_0 = 136$ mm Hg vs. H_A: $\mu \neq \mu_0$

Because $$\frac{\bar{X} - \mu_0}{s/\sqrt{n}} = \frac{143-136}{24.4/\sqrt{86}} = 2.66$$

$$p-value = P\left(\frac{|\bar{X} - \mu_0|}{s/\sqrt{n}} \geq \frac{|143 - \mu_0|}{s/\sqrt{n}} \;\middle|\; H_0 \text{ is true} \right)$$

$$= P\left(\frac{\bar{X} - \mu_0}{s/\sqrt{n}} \geq \frac{143-\mu_0}{s/\sqrt{n}} \;\middle|\; H_0 \text{ is true} \right) + \left(\frac{\bar{X} - \mu_0}{s/\sqrt{n}} \leq -\frac{143-\mu_0}{s/\sqrt{n}} \;\middle|\; H_0 \text{ is true} \right)$$

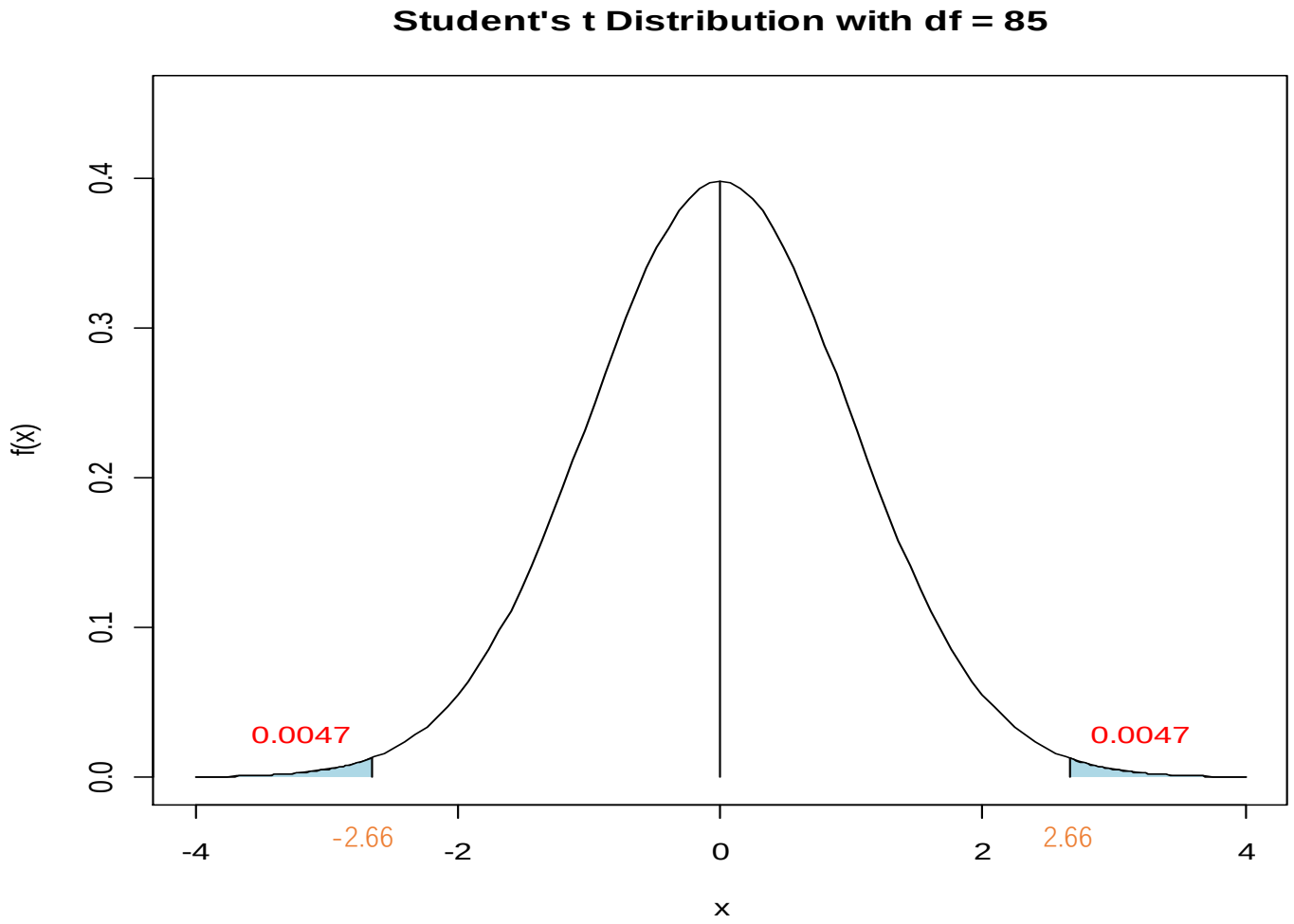
$$= P(t \geq 2.66 \mid H_0 \text{ is true}) + P(t \leq -2.66 \mid H_0 \text{ is true})$$

$$= 0.0047 + 0.0047$$

$$= 0.0094 \leq \alpha$$

which is less than the significance level α = 0.05. Therefore, the test is significant and H_0 is *rejected*

Calculate p-value for t Distribution



Two-Sided Hypothesis Test for μ

H_0: $\mu = \mu_0 = 136$ mm Hg vs. H_A: $\mu \neq \mu_0$

This sample of 86 workers provides strong evidence against the null hypothesis H_0: $\mu = 136$ mm Hg. In other words, the data are not compatible with H_0

Conclusion:

Since the p-value = 0.0094 < 0.05, we *reject* H_0 at the 0.05 significance level and conclude that the mean SBP for workers who have experienced a major coronary event is significantly different (actually higher) from the mean for workers who have not.

Summary of One-sample Two-Sided t-test

- General set-up: μ and σ are unknown
  - Null hypothesis H_0: $\mu = \mu_0$
  - Alternative hypothesis H_A: $\mu \neq \mu_0$ (two-sided)
  - Test statistic:

$$t = \frac{\overline{X} - \mu_0}{s/\sqrt{n}}$$

    which has a t distribution with n-1 degrees of freedom. So this test is known as (one sample) *t-test*
  - *p-value* = $P(t_{n-1} \geq |t|) + P(t_{n-1} \leq -|t|)$, where $|t|$ is the absolute value of the observed t calculated from the sample
  - Compare p-value with significance level α and make conclusion

Two-Sided Z-test: Example

- Example: The general population of US males aged between 20 and 74 has mean serum cholesterol level of 211 mg/100 ml, with standard deviation 46 mg/100 ml. Use a random sample of 12 hypertensive smokers with mean serum cholesterol level 217 mg/100 ml to test whether the subpopulation of men who are hypertensive smokers has the same mean serum cholesterol level at 0.05 significance level

H_0: $\mu = \mu_0 = 211$ mg/100 ml vs. H_A: $\mu \neq \mu_0$

Because the population s.d. is known as $\sigma = 46$ mg/100 ml, a z-test is used

Test statistic

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{217 - 211}{46/\sqrt{12}} = 0.45$$

$$p = P(Z \geq 0.45) + P(Z \leq -0.45)$$
$$= 0.326 \times 2$$
$$= 0.652 > \alpha$$

Do NOT reject H_0

Chapter 11两样本t检验

Chapter 11. Comparison of Two Means

11.1 Paired Samples

11.2 Independent two-sample

(a. Equal Variance)

Paired samples: an example

- 50 hypertensive patients were recruited to test the effect of an hypertension drug.
- At enrollment, the systolic blood pressure of 50 patients were measured.
- They were given the medicine for 2 weeks and come back for post-treatment visit. Their blood pressure were measured again.

- Question: Is there any evidence that the drug reduce the patients' blood pressure ?

Data structure

	Pre-treatment SBP	Post-Treatment SBP	Difference
Sample	x_{11}	x_{12}	$d_1 = x_{11} - x_{12}$
	x_{21}	x_{22}	$d_2 = x_{21} - x_{22}$
	x_{31}	x_{32}	$d_3 = x_{31} - x_{32}$
	$\vdots$	$\vdots$	$\vdots$
	x_{n1}	x_{n2}	$d_n = x_{n1} - x_{n2}$
	$\bar{x}_1$	$\bar{x}_2$	$\bar{d}_n = \bar{x}_1 - \bar{x}_2$
Population	μ_1	μ_2	$\delta = \mu_1 - \mu_2$

Pros and cons of this design

- In this study, the two measurements (pre- and post treatment SBP) are from the same patient. The advantage of this type of design lies in that each patient serves as his own "control" and the biological variation is filtered out. Patients who have higher blood pressure tend to have higher blood pressure after treatment too. Hence, biological factors (eg. age) resulting higher SBP will not confound the effect of treatment.
- It is not appropriate to consider the two measurements from the same patient to be independent since one measurement provide substantial information on the other one.

Paired samples: the framework

Suppose (x_{i1}, x_{i2}) is a paired measures on the same experiment unit i, $i = 1,2,...n$. We want to conduct a hypothesis testing or construct a confidence interval for $\delta = \mu_1 - \mu_2$.

Other examples: Angina example in textbook; desire to learn statistics before and after one takes P551 :)

Assumption: d_i is independently normaly distributed

Strategy: treat $d_1, d_2, ...d_n$ as a random sample and the problem becomes a one-sample problem

Paired samples (two-sided): what to do

Hypothesis testing (type I error = α):

Two-sided test:

$H_0: \ \mu_1 = \mu_2$ vs. $H_a: \ \mu_1 \neq \mu_2$

In other words,

$H_0: \ \delta = \delta_0$ vs. $H_a: \ \delta \neq \delta_0$

$(1-\alpha)\times 100\%$ confidence interval for $\mu_1 - \mu_2$, or δ.

Paired samples: how to do

Hypothesis testing :

$$T = \frac{\bar{d}_n - \delta_0}{s_n / \sqrt{n}}, \qquad \bar{d}_n = \frac{\sum_{i=1}^{n} d_i}{n}, s_n^2 = \frac{\sum_{i=1}^{n} (d_i - \bar{d}_n)^2}{n-1}$$

Reject H_0 if $|T| > t_{n-1,\alpha/2}$

Two - sided $(1-\alpha)\times 100\%$ confidence interval for δ :

$$(\bar{d}_n - t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}, \bar{d}_n + t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}})$$

Paired samples (One-sided): what to do

Hypothesis testing (type I error = α):

(1)

$H_0: \mu_1 \leq \mu_2$ vs. $H_a: \mu_1 > \mu_2$

In other words,

$H_0: \delta \leq \delta_0$ vs. $H_a: \delta > \delta_0$

(2)

$H_0: \mu_1 \geq \mu_2$ vs. $H_a: \mu_1 < \mu_2$

In other words,

$H_0: \delta \geq \delta_0$ vs. $H_a: \delta < \delta_0$

.

Paired samples: how to do

Hypothesis testing :

$$T = \frac{\bar{d}_n - \delta_0}{s_n / \sqrt{n}}, \qquad \bar{d}_n = \frac{\sum_{i=1}^{n} d_i}{n}, s_n^2 = \frac{\sum_{i=1}^{n} (d_i - \bar{d}_n)^2}{n-1}$$

(1) Reject H_0 if $T > t_{n-1,\alpha}$

(2)Reject H_0 if $T < -t_{n-1,\alpha}$

***One* - sided (1 - α) × 100% confidence interval for δ :**

(1) $(\bar{d}_n - t_{n-1,\alpha} \frac{s_n}{\sqrt{n}}, \infty)$

(2)$(-\infty, \bar{d}_n + t_{n-1,\alpha} \frac{s_n}{\sqrt{n}})$

The SBP example: hypothesis testing

1. Hypothesis

H_0: $\delta \le 0$ $(\mu_1 \le \mu_2)$, the drug doesn't reduce patients' SBP.

H_a : $\delta > 0$ $(\mu_1 > \mu_2)$, the drug does reduce SBP.

2. Type I error $\alpha = 0.05$

3. Calculate $\bar{d}_n = \bar{x}_1 - \bar{x}_2 = 145 - 120 = 25$ and $s_n = \sqrt{\dfrac{\sum_{i=1}^{n}(d_i - \bar{d}_i)^2}{n-1}} = 55$

4. Compute test statistic $T = \dfrac{\bar{d}_n - \delta_0}{s_n / \sqrt{n}} = \dfrac{25}{55/\sqrt{50}} = 3.21$

5. Find $t_{n\text{-}1,\alpha} = t_{50,0.05} = 1.676$

6. Since 3.21>1.676, the null hypothesis is rejected

with 5% Type I error. Hence, with 95% confidence, the drug reduce the patients' SBP.

The SBP example: confidence interval

One-sided 95% confidence interval with a lower bound for $\delta = \mu_1 - \mu_2$

$$(25 - t_{n-1,\alpha}\frac{55}{\sqrt{50}}, \infty)$$

$$= (25 - 1.676 \times \frac{55}{\sqrt{50}}, \infty)$$

$$= (11.96, \infty)$$

Two-sided 95% confidence interval for $\delta = \mu_1 - \mu_2$

$$(25 - t_{n-1,\alpha/2}\frac{55}{\sqrt{50}}, 25 + t_{n-1,\alpha/2}\frac{55}{\sqrt{50}})$$

$$= (25 - 2.00 \times \frac{55}{\sqrt{50}}, 25 + 2.00 \times \frac{55}{\sqrt{50}})$$

$$= (9.44, 40.56)$$

Hence, we are 95% confident that the interval (9.44,40.56) covers the reduction of SBP before and after treatment.

Two-sample t Test (unpaired, or independent samples)

- Interested in comparing the means of two populations (e.g. normal vs. diseased)
- Draw a random sample from each of the two populations (sample size can be different)
- Make inference based on the two samples

Two-sample *t* Test (unpaired): assumptions

- Two random samples, one from each population
- The two samples are independent to each other
- Normally distributed
- Variances of the two populations are approximately equal

		Group 1	Group 2
Population	Mean	μ_1	μ_2
	Standard Deviation	$\sigma_1 = \sigma$	$\sigma_2 = \sigma$
Sample	Mean	$\bar{x}_1$	$\bar{x}_2$
	Standard Deviation	s_1	s_2
	Sample Size	n_1	n_2

Two sample *t* Test (unpaired): what to do

Let $\delta = \mu_1 - \mu_2$

Hypothesis testing (type I error = α) :

(a) One - sided test : $H_0: \ \delta \le \delta_0$ vs. $H_a: \delta > \delta_0$

(b) One - sided test : $H_0: \ \delta \ge \delta_0$ vs. $H_a: \delta < \delta_0$

(c) Two - sided test : $H_0: \ \delta = \delta_0$ vs. $H_a: \delta \ne \delta_0$

(1 - α) × 100% confidence interval for δ

Pooled estimate of the variance

We need to estimate the common standard deviation σ based on the two samples. It can be estimated by a weighted sum of the sample standard deviation s_1 and s_2 :

$$s_p = \sqrt{\frac{n_1 - 1}{n_1 + n_2 - 2} s_1^2 + \frac{n_2 - 1}{n_1 + n_2 - 2} s_2^2}$$

s_p is called the **pooled** estimate of the standard deviation. The weight $\frac{n_1 - 1}{n_1 + n_2 - 2}$ and $\frac{n_2 - 1}{n_1 + n_2 - 2}$ reflect the proportion of information contribution from each sample. Larger sample size provides more information regarding σ and is weighted more.

Test statistic

Hypothesis testing of (a), (b) and (c) is based on the following statistic:

$$T = \frac{(\bar{x}_1 - \bar{x}_2) - \delta_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

T has a t distribution with $n_1 + n_2 - 2$ degrees of freedom when $\delta = \mu_1 - \mu_2 = \delta_0$. One can calculate T and compare it with relevant quantile of the t distribution with $n_1 + n_2 - 2$ degrees of freedom for hypothesis testing. Confidence interval for δ can be constructed based on $\bar{x}_1 - \bar{x}_2$, $s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$ and relevant quantile of the t distribution with $n_1 + n_2 - 2$ degrees of freedom.

Two sample *t* Test (unpaired): how to do

Hypothesis testing :

(a) Reject H_0 if $T > t_{n_1+n_2-2,\alpha}$

(b) Reject H_0 if $T < -t_{n_1+n_2-2,\alpha}$

(c) Reject H_0 if $T > t_{n_1+n_2-2,\alpha/2}$ or $T < -t_{n_1+n_2-2,\alpha/2}$

Two - sided (1 - α) × 100% confidence interval for δ :

$$\left(\bar{x}_1 - \bar{x}_2 - t_{n_1+n_2-2,\alpha/2} s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}}, \bar{x}_1 - \bar{x}_2 + t_{n_1+n_2-2,\alpha/2} s_p \sqrt{\frac{1}{n_1}+\frac{1}{n_2}} \right)$$

Two sample *t* Test (unpaired): example –p.265

- Measurements of serum iron for two samples of children: healthy and suffering from cystic fibrosis
- Interested in whether or not the mean serum levels of the two populations are identical
- A random sample of size 9 and a random sample of size 13 are drawn from the healthy population and the diseased population, respectively.

Serum Iron level example: data

		Group 1 (normal)	Group 2 (diseased)
Population	Mean	μ_1	μ_2
	Standard Deviation	σ	σ
Sample	Mean	$\bar{x}_1 = 18.9$	$\bar{x}_2 = 11.9$
	Standard Deviation	$s_1 = 5.9$	$s_2 = 6.3$
	Sample Size	$n_1 = 9$	$n_2 = 13$

Serum level example: hypothesis testing

1. Hypothesis

   $H_0 : \delta = 0 \ (\mu_1 = \mu_2)$, or no difference in the mean serum iron levels for the two groups

   $H_a : \delta \neq 0 \ (\mu_1 \neq \mu_2)$, or difference in the mean serum iron levels for the two groups

2. Type I error $\alpha = 0.05$

3. Calculate $s_p = \sqrt{\dfrac{(9-1)(5.9)^2 + (13-1)(6.3)^2}{9+13-2}} = 6.14$

4. Compute test statistic $T = \dfrac{18.9 - 11.9 - 0}{6.14\sqrt{1/9 + 1/13}} = 2.63$

5. Find $t_{9+13-2,0.025} = t_{20,0.025} = 2.086$

6. $2.63 > 2.086$, the null hypothesis is rejected with Type I error = 0.05. The difference in the mean serum iron levels between the two groups is statistically significant at level 0.05. It appears that children with cystic fibrosis suffer from an iron dificiency.

Serum level example: confidence interval

Two-sided 95% confidence interval for $\delta = \mu_1 - \mu_2$

$$(18.9 - 11.9 - 2.08 \times 6.14 \times \sqrt{1/9 + 1/13}, 18.9 - 11.9 + 2.08 \times 6.14 \times \sqrt{1/9 + 1/13})$$

$$= (1.4, 12.6)$$

Hence, we are 95% confident that the interval (1.4,12.6) covers the true difference in mean serum iron levels for the two populations.

For hypotheses: $H_0 : \mu_1 = \mu_2$, we find the interval does NOT cover the null hypotheses δ=0. Therefore we reject the null hypothese with 95% confidence.

Two-sample t Test (unpaired): unequal variances

When $\sigma_1 \neq \sigma_2$, s_1 and s_2 can be used to estimate σ_1 and σ_2, the test statistic is then

$$T = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

However, the distribution of T when $\delta = \delta_0$ is rather complex. We approximate it by a t distribution. To calculate the degrees of freedom, we first compute

$$v = \frac{[s_1^2/n_1 + s_2^2/n_2]^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}.$$

v is then rounded to the nears interger, which is the degrees of freedom for the t distribution. Hypothesis testing and confidence interval can be carried out in a similar way as in the equal variance scenario. When v is large, it is roughly a standard normal distribution.

Chapter 12.1 Analysis of Variance (ANOVA)

A Roadmap

❖Testing Population Means

- -Single group

  one sample z-test / t-test;

  compare the mean to a constant

- -Two groups

  two sample t-test / paired t-test

  compare the means directly or compare the mean difference to a constant

- -More than two groups

  to test if all means are equal;

  method: **AN**alysis **O**f **VA**riance (ANOVA)

Example: Study of Lung Function

- Subjects from three different medical centers:
  - The John Hopkins University School of Medicine
  - The Rancho Los Amigos Medical Center
  - St. Louis University School of Medicine

- We need to ensure that patients from different sites are comparable in terms of some baseline characteristics before merging them into one group.

- One of the characteristics of interest is the pulmonary function at the entry of the study.

Study of Lung Function Cont.

- Forced Expiratory Volume at 1 second (FEV_1) as a measurement of pulmonary function
- We want to test whether or not the mean FEV_1 values from the three sites are the same
- Question: how to construct such a test?

Hypotheses

$$X_{1i} \sim N(\mu_1, \sigma^2),\ X_{2i} \sim N(\mu_2, \sigma^2), X_{3i} \sim N(\mu_3, \sigma^2)$$

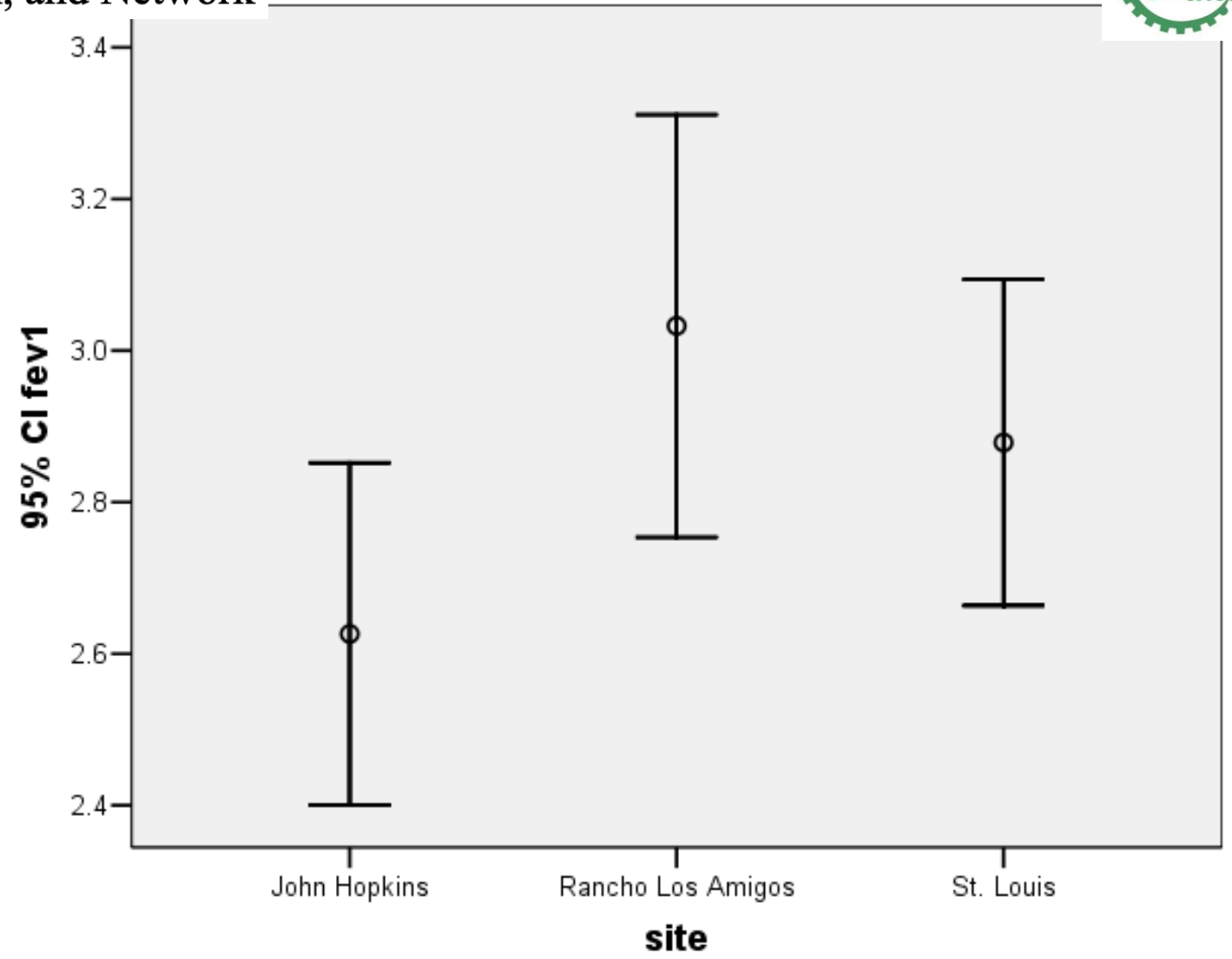
Hypotheses :

Let μ_1, μ_2 and μ_3 be the mean FEV_1 values for patients from the three sites.

$H_0 : \mu_1 = \mu_2 = \mu_3$

H_a : at least one pair of the means are not identical

A natural thought : can we use two - sample t test to solve this problem? For instance, evaluate all possible pairs of means by two sample t test. If so, what about α, how do we control it?


95% CI fev1
3.4
3.2
3.0
2.8
2.6
2.4
John Hopkins
Rancho Los Amigos
St. Louis
site

Two-sample *t* test?

John Hopkins	**Rancho Los Amigos**	**St. Louis**
(group 1)	(group 2)	(group 3)
$n_1 = 21$	$n_2 = 16$	$n_3 = 23$
$\bar{x}_1 = 2.63$ liters	$\bar{x}_2 = 3.03$ liters	$\bar{x}_3 = 2.88$ liters
$s_1 = 0.496$ liters	$s_2 = 0.523$ liters	$s_3 = 0.498$ liters

There are three possible comparisons : group 1 vs. group 2, group1 vs. group 3, and group 2 vs. group3.

Hypothetical procedure :

1. Conduct two - sample *t* test for each of the three possible pairs
2. Reject $H_0 : \mu_1 = \mu_2 = \mu_3$ if one of the three two - sample *t* tests is significant at $\alpha = 0.05$

Result :

- group 1 vs. group 2 : $p = 0.02$
- group 1 vs. group 3 : $p = 0.10$
- group 2 vs. group 3 : $p = 0.36$

Hence, reject the null hypothesis

Difficulties

$H_{01}: \mu_1 = \mu_2; H_{02}: \mu_1 = \mu_3; H_{03}: \mu_2 = \mu_3$

*Type I error rate α=Pr[reject $\mu_1 = \mu_2 = \mu_3 \mid \mu_1 = \mu_2 = \mu_3$ is true]

=Pr[reject any of $H_{01}, H_{02}, H_{03} \mid H_{01}, H_{02}, H_{03}$ is true]

Pr[fail to reject all of $H_{01}, H_{02}, H_{03} \mid H_{01}, H_{02}, H_{03}$ is true]

= Pr[fail to reject in all three tests$|H_{01}, H_{02}, H_{03}$ is true]

=Pr[fail to reject $H_{01}|H_{01}$true] Pr[fail to reject $H_{02}|H_{02}$true] Pr[fail to reject $H_{03}|H_{03}$true]

$= (1\text{-}0.05)^3 = 0.857.$

Hence, α = Pr[reject any one of $H_{01}, H_{02}, H_{03} \mid H_{01}, H_{02}, H_{03}$ is true]

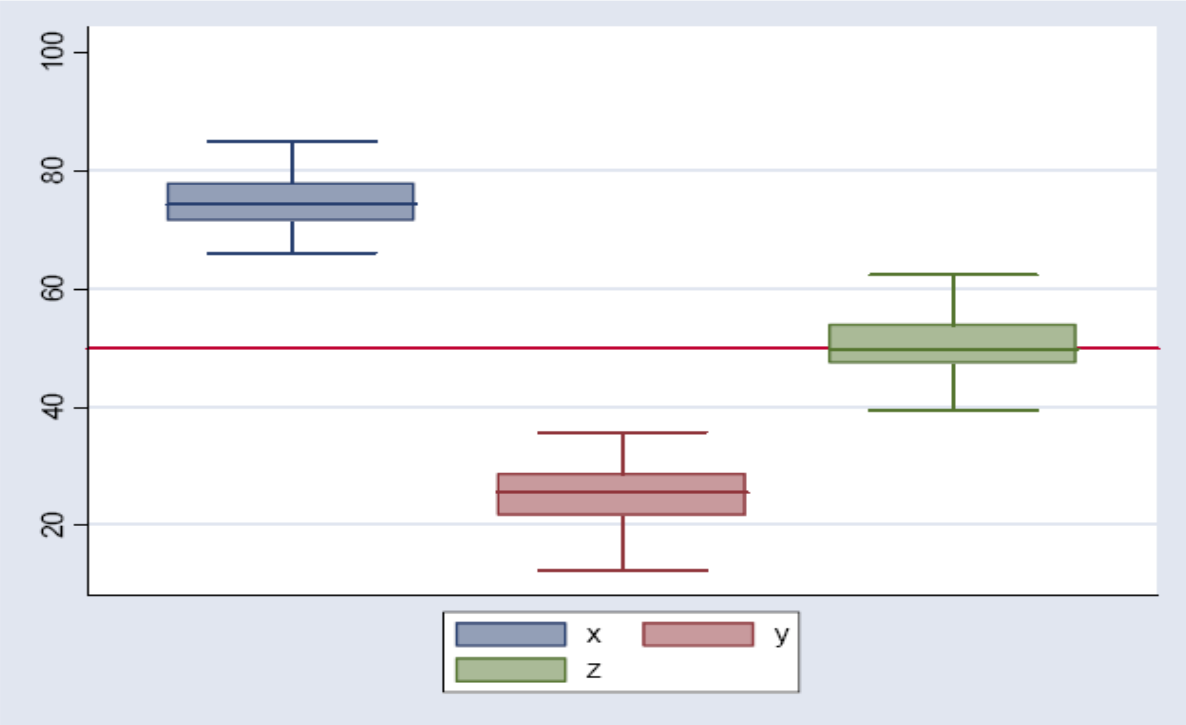
$= 1\text{-}$ Pr[fail to in all three tests $\mid H_0$ is true] =

$= 1\text{-}\ 0.857 = 0.143 > 0.05$

* Actually the three t tests are not independent $\rightarrow$ more complex
* **Need a procedure in which the type I error can be controlled**

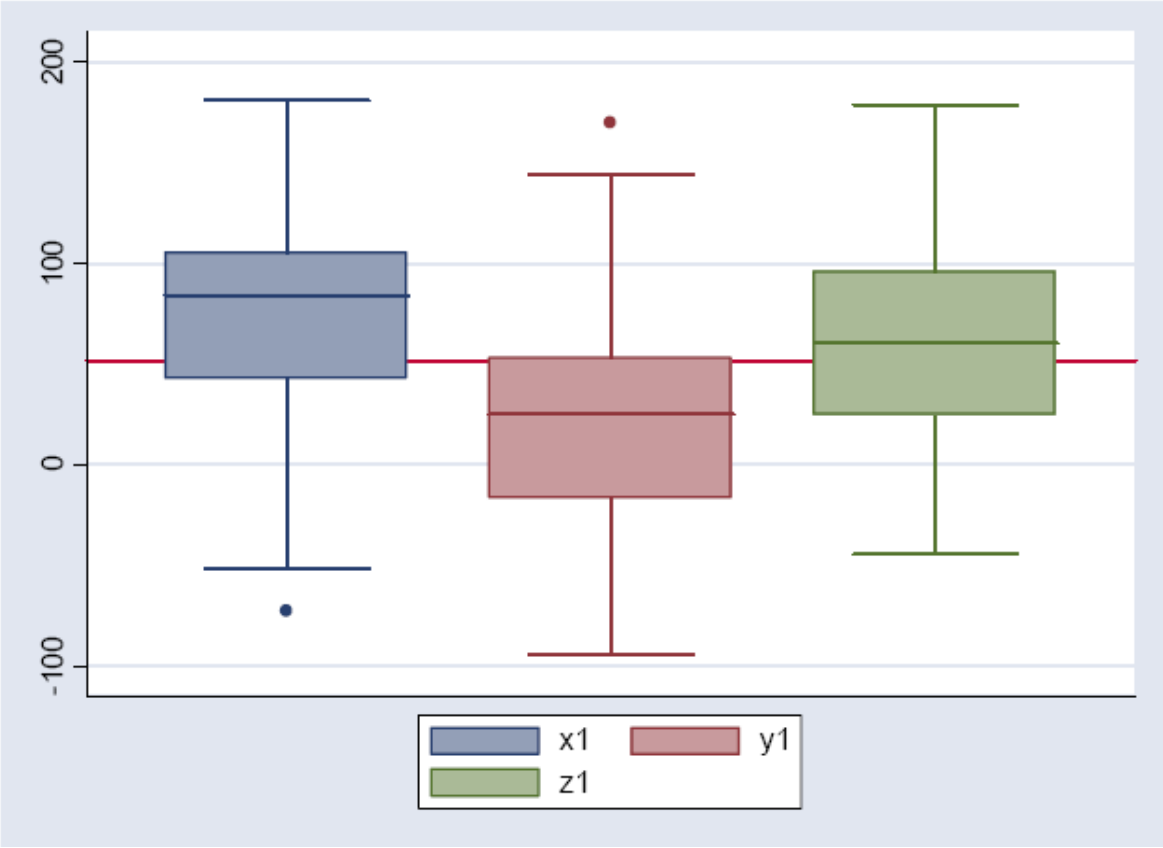
Intuition on ANOVA

- Case 1: Small within group variation, large between/within ratio
- Means of (x1, y1, z1) are (75 25 50)



Intuition on ANOVA

- Case 2: Large within group variation, small between/within ratio
- Means of (x21, y2, z2) are (75 25 50)



One-way Analysis of Variance

- Based on two sources of variation
  - Within group variation: the variation of the individual values around their population means
  - Between group variation: the variation of the population means around the overall mean
- Compare the two sources of variations: if the between group variation is larger relative to the within group variation, it indicates that the population means are different
- How large is large?

One-way ANOVA Framework

		Group 1	Group 2	...	group k
Population	Mean	μ_1	μ_2	...	μ_k
	Standard Deviation	σ_1	σ_2	...	σ_k
Sample	Mean	$\bar{x}_1$	$\bar{x}_2$	...	$\bar{x}_k$
	Standard Deviation	s_1	s_2	...	s_k
	Sample Size	n_1	n_2	...	n_k

$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$

H_a : at least one pairs of means are not identical

Type I error $= \alpha$

Assumptions

- A random sample is drawn from each population independently
- The variances of the populations are the same:

$$\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$$

- Each population is approximately normal

Measure of Variability

- Within group variation: pooled estimate of the common variance σ^2

  Within group Sum of Square $(SS_w) = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2$

  Within group Mean Square $(MS_w) = \dfrac{SS_w}{n-k} = \dfrac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_k - 1)s_k^2}{n-k}$,

  and $n = n_1 + n_2 + \cdots + n_k$

- Between group variation:

  Between group Sum of Square $(SS_B) = n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_k(\bar{x}_k - \bar{x})^2$

  Between group Mean Square $(MS_B) = \dfrac{SS_B}{k-1} = \dfrac{n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \cdots + n_k(\bar{x}_k - \bar{x})^2}{k-1}$,

  $\bar{x} = \dfrac{n_1\bar{x}_1 + n_2\bar{x}_2 + \cdots + n_k\bar{x}_k}{n}$

- Total variation: total Sum of Square $(\mathrm{SS}_T) = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2 = SS_w + SS_B$

Variability Ratio vs. Hypotheses

- Compare the within group variation and between group variation:

$$F = \frac{MS_B}{MS_W}$$

Under H_0, F is close to 1.If there is a diffence among the means, then the between group variation is larger than the within group variation and F is larger than 1.

Hence, reject the null hypothesis when F is large.

The F distribution

- Under the null hypothesis F has a F distribution with $\mathbf{k-1}$ and $\mathbf{n-k}$ degrees of freedom, which correspond to the nominator and denominator, respectively. We denote this distribution by $F_{k-1,n-k}$.
- Reject the null hypothesis if $F > F_{k\text{-}1,n\text{-}k,\alpha}$. Hence, the type I error is α. Here, $F_{k\text{-}1,n\text{-}k,\alpha}$ denotes the value that cuts off an area of α in the upper tail of the F distribution with two parameters as k-1 and n-k
- Properties of F distribution
  - Only takes positive values
  - Not symmetric
  - The shape of the curve is determined by the two degree of freedom parameters

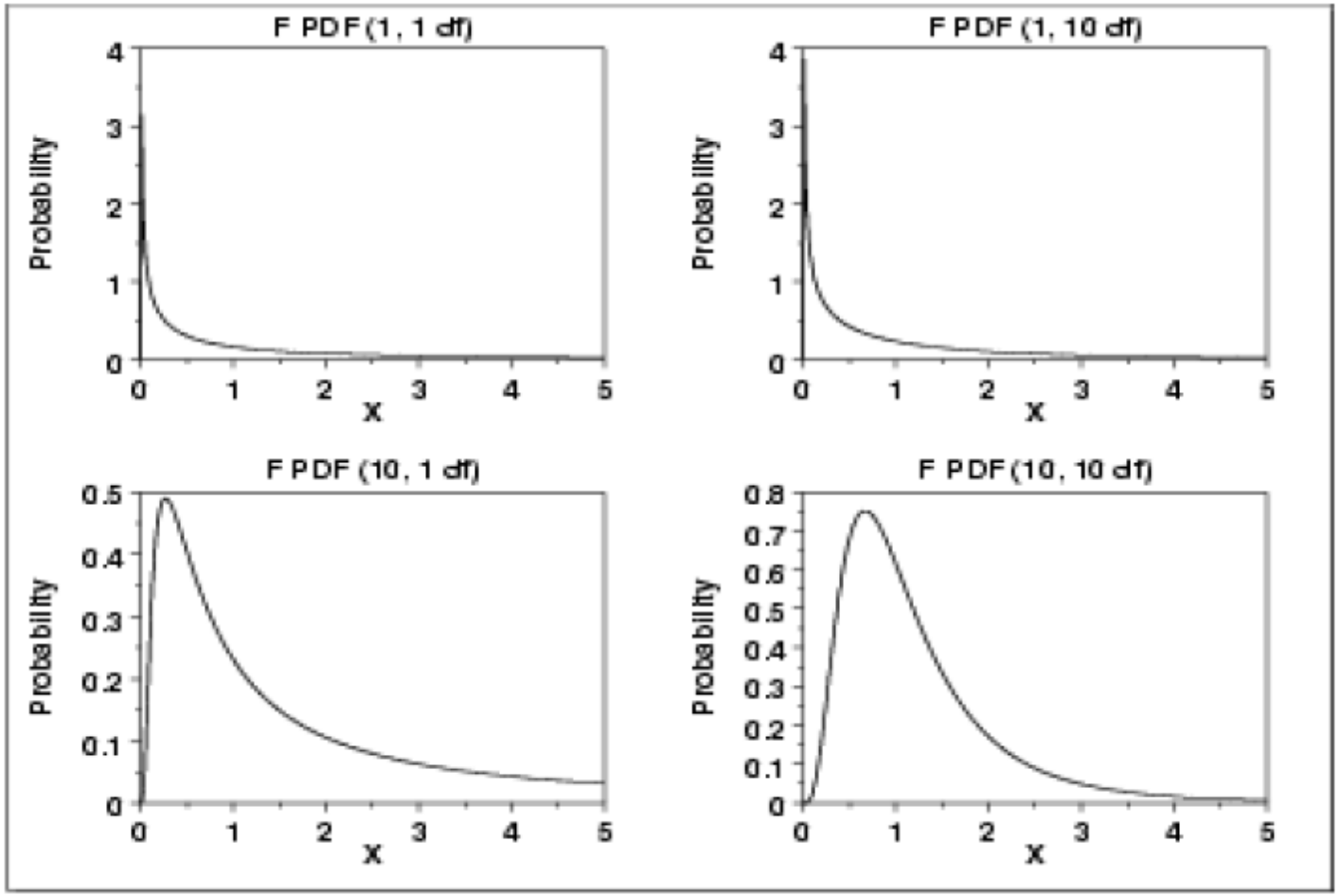

F PDF (1, 1 df)
F PDF (1, 10 df)
F PDF (10, 1 df)
F PDF (10, 10 df)
Probability
X

TABLE A.5
Percentiles of the *F* distribution

Denominator df	Area in Upper Tail	Numerator Degrees of Freedom (df)										
		1	2	3	4	5	6	7	8	12	24	∞
2	0.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.41	9.45	9.49
	0.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.41	19.45	19.50
	0.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.41	39.46	39.50
	0.010	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.42	99.46	99.50
	0.005	198.5	199.0	199.2	199.3	199.3	199.3	199.4	199.4	199.4	199.5	199.5
	0.001	998.5	999.0	999.2	999.3	999.3	999.3	999.4	999.4	999.4	999.5	999.5
3	0.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.22	5.18	5.13
	0.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.74	8.64	8.53
	0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.34	14.12	13.90
	0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.05	26.60	26.13
	0.005	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.39	42.62	41.83
	0.001	167.0	148.5	141.1	137.1	134.6	132.9	131.6	130.6	128.3	125.9	123.5
4	0.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.90	3.83	3.76
	0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.91	5.77	5.63
	0.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.75	8.51	8.26
	0.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.37	13.93	13.46
	0.005	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	20.70	20.03	19.32
	0.001	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	47.41	45.77	44.05
5	0.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.27	3.19	3.10
	0.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.68	4.53	4.36
	0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.52	6.28	6.02
	0.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	9.89	9.47	9.02
	0.005	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.38	12.78	12.14
	0.001	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	26.42	25.13	23.79
6	0.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.90	2.82	2.72
	0.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.00	3.84	3.67
	0.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.37	5.12	4.85
	0.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.72	7.31	6.88
	0.005	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.03	9.47	8.88
	0.001	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	17.99	16.90	15.75

(continued)

TABLE A.5
(continued)

Denominator df	Area in Upper Tail	Numerator Degrees of Freedom (df)										
		1	2	3	4	5	6	7	8	12	24	∞
7	0.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.67	2.58	2.47
	0.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.57	3.41	3.23
	0.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.67	4.41	4.14
	0.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.47	6.07	5.65
	0.005	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.18	7.64	7.08
	0.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	13.71	12.73	11.70
8	0.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.50	2.40	2.29
	0.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.28	3.12	2.93
	0.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.20	3.95	3.67
	0.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.67	5.28	4.86
	0.005	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.01	6.50	5.95
	0.001	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.19	10.3	9.33
9	0.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.38	2.28	2.16
	0.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.07	2.90	2.71
	0.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	3.87	3.61	3.33
	0.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.11	4.73	4.31
	0.005	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.23	5.73	5.19
	0.001	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	9.57	8.72	7.81
10	0.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.28	2.18	2.06
	0.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.91	2.74	2.54
	0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.62	3.37	3.08
	0.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.71	4.33	3.91
	0.005	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.66	5.17	4.64
	0.001	21.04	14.91	12.55	11.28	10.48	9.93	9.52	9.20	8.45	7.64	6.76
12	0.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.15	2.04	1.90
	0.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.69	2.51	2.30
	0.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.28	3.02	2.72
	0.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.16	3.78	3.36
	0.005	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	4.91	4.43	3.90
	0.001	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.00	6.25	5.42

One-way ANOVA Table

Source of variation	SS	d.f.	MS	F Statistic
Between	SS_B	K-1	MS_B	MS_B/MS_W
Within	SS_W	n-K	MS_W	
Total	SS_T	n-1		

Decision rule:

Application to FEV_1

- $SS_W = (n_1-1)s_1^2 + (n_2-1)s_2^2 + (n_3-1)s_3^2 = (21-1)(0.496)^2 + (16-1)(0.523)^2 + (23-1)(0.498)^2$

  $=14.478 \text{ liters}^2$

$$MS_W = \frac{SS_W}{n_1+n_2+n_3-3} = \frac{14.478}{21+16+23-3}$$

$$= 0.254 \text{ liters}^2$$

- $\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2 + n_3\bar{x}_3}{n_1+n_2+n_3} = \frac{21(2.63)+16(3.03)+23(2.88)}{21+16+23} = 2.83 \text{ liters}$

$$SS_B = n_1(\bar{x}_1-\bar{x})^2 + n_2(\bar{x}_2-\bar{x})^2 + n_3(\bar{x}_3-\bar{x})^2 = 21(2.63-2.83)^2 + 16(3.03-2.83)^2 + 23(2.88-2.83)^2$$

$$=1.538 \text{ liters}^2$$

$$MS_B = \frac{SS_B}{3-1} = \frac{1.538}{3-1}$$

$$= 0.769 \text{ liters}^2$$

$$SST = SS_B + SS_W = 1.538 + 14.478 = 16.016$$

Application to FEV_1

Source of variation	SS	d.f.	MS	F Statistic
Between	1.538	2	0.769	3.028
Within	14.478	57	0.254	
Total	16.016			

- $F = \frac{MS_B}{MS_B} = \frac{0.769}{0.254} = 3.028 < 3.159 = F_{3-1,60-3,0.05} = F_{2,57,0.05}$
- We do not have enough evience to reject the null hypothesis that the mean FEV_1 values of the three sites are identical.

```
> pulmonary_function<-read.csv("C:/YU/teaching/致远交叉创新/data/pulmonary
function.csv",
+                                          header=TRUE)
> fev1<- as.numeric(as.character(pulmonary_function$fev1))
Warning message:
强制改变过程中产生了NA
> center<-as.factor(pulmonary_function$center)
> fev.aov<-aov(fev1~center)
> summary(fev.aov)
             Df Sum Sq Mean Sq F value Pr(>F)
center        2  1.583  0.7914   3.115  0.052 .
Residuals    57 14.480  0.2540
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
3 observations deleted due to missingness
```

Another ANOVA Example

- Children Health and Development Study (CHDS) was conducted in CA. This data used in this question is a subset of CHDS containing information about 680 newborn babies.
- The outcome of interest is the body length of the newborn babies and we are interested in various factors from the parents which would affect the outcome. A particular important risk factor is the mother's smoking behavior.
- The information has been recorded as a discrete variable with three levels—nonsmoker (msomke_cat=0), light smoker (msmoke_cat=1), heavy smoker (msmoke_cat=2), the summary data is shown as below

Data

	Summary of length		
msmoke_cat	Mean	Std. Dev.	Freq.
0	20.448819	.94625574	381
1	20.12426	.97097929	169
2	19.984615	1.0037497	130
Total	20.279412	.9821018	680

Data

Assuming that the length (measured in inches) follow normal distributions in the three groups:

(a) If the scientific question is to figure out whether the mean birth lengths are the same for the three groups, what would be an appropriate test to perform? Specify the null and alternative hypotheses and the test statistics.

Source	SS	df	MS	F
Between groups	26.30	?	13.15	?
Within groups	628.61	677	?	
Total	?	679	-------	

- There are four cells missing in ANOVA table (marked with “?”), please fill them in:

  d.f. for between group=_____

  SS for total=_____

  MS for within group=_____

  F-ratio=_____

- After you complete the above table, perform the test for the hypothesis as specified in part 1) at the level of =0.05. State your conclusion.
  - If the test is performed at the level of =0.01, what would be your conclusion?

A Special Case

When $k = 2$, F test is equivalent to two-sided two-sample t test.

$$MS_W = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2} = s_p^2$$

$$MS_B = \frac{n_1(\bar{x}_1-\bar{x})^2+n_2(\bar{x}_2-\bar{x})^2}{2-1} = n_1\left(\bar{x}_1-\frac{n_1\bar{x}_1+n_2\bar{x}_2}{n_1+n_2}\right)^2+n_2\left(\bar{x}_2-\frac{n_1\bar{x}_1+n_2\bar{x}_2}{n_1+n_2}\right)^2 = n_1\left(\frac{n_2(\bar{x}_1-\bar{x}_2)}{n_1+n_2}\right)^2+n_2\left(\frac{n_1(\bar{x}_2-\bar{x}_1)}{n_1+n_2}\right)^2$$

$$= \frac{n_1n_2^2(\bar{x}_1-\bar{x}_2)^2+n_2n_1^2(\bar{x}_1-\bar{x}_2)^2}{(n_1+n_2)^2} = \frac{n_1n_2(\bar{x}_1-\bar{x}_2)^2(n_1+n_2)}{(n_1+n_2)^2} = \frac{n_1n_2(\bar{x}_1-\bar{x}_2)^2}{n_1+n_2} = \left(\frac{1}{n_1}+\frac{1}{n_2}\right)(\bar{x}_1-\bar{x}_2)^2$$

$$F = \frac{MS_B}{MS_W} = \frac{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)(\bar{x}_1-\bar{x}_2)^2}{s_p^2} = \frac{(\bar{x}_1-\bar{x}_2)^2}{s_p^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = T^2$$

$F > F_{1,n-2,\alpha} \Leftrightarrow T > \sqrt{F_{1,n-2,\alpha}}$ or $T < -\sqrt{F_{1,n-2,\alpha}} \Rightarrow$ Two-sided two-sample t test

Hence, $F_{1,n-2,\alpha} = t^2_{n-2,\alpha/2}$

Two-way ANOVA

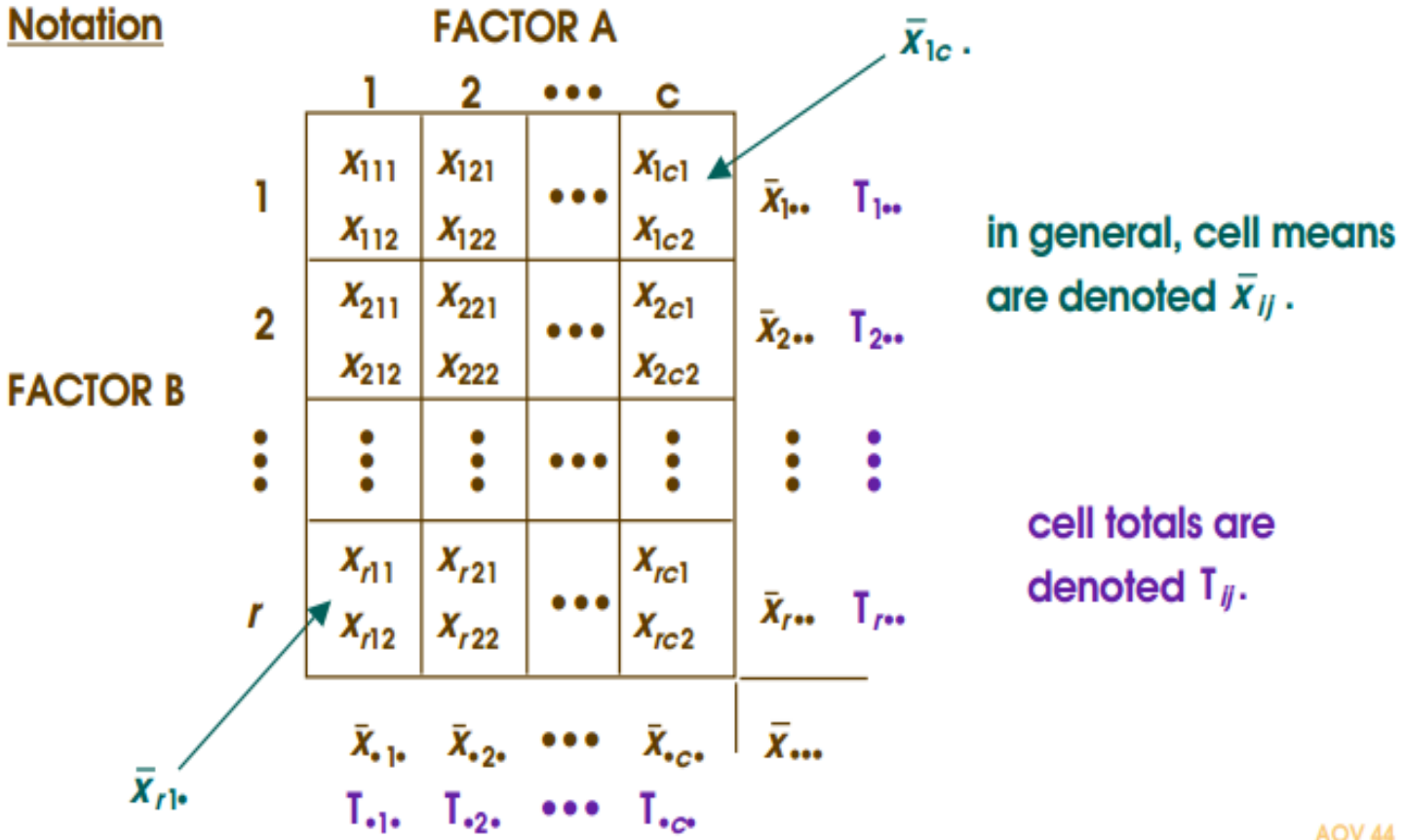
- Example
- In our study on diets, suppose two factors were of interest: diet and exercise level. A total of 24 hamsters (of the same initial weight) were available, and two were assigned at random to each combination of Diet and exercise level.

Exercise \ Diet	Diet 1	Diet 2	Diet 3
None	17, 17	15, 21	17, 21
Mild	17, 19	20, 20	22, 22
Moderate	12, 12	19, 25	18, 22
Heavy	20, 22	24, 32	21, 25

$T_{...} = 480$

$\bar{x}_{...} = 20$

Two-way ANOVA



Two-way ANOVA

- Testing for difference among the two categories– using one way ANOVA

None/D1	None/D2	None/D3	Mild/D1 •••	Heavy/D1	Heavy/D2	Heavy/D3
17	15	17	17	20	24	21
17	21	21	19	22	32	25
$T_{j\cdot}$ 34	36	38	36	42	56	46
$\bar{x}_{j\cdot}$ 17	18	19	18	21	28	23

H_0: 12 means are equal

$F = 29.82/8 = 3.73$ and $F_{.95}(11,12) \approx 2.72$

$\therefore$ reject H_0

$$EMS = rn\sigma_c^2 + cn\sigma_r^2 + n\sigma_{rc}^2 + \sigma_e^2$$

ANOVA

Source	SS	df	MS	F
Between 12 means	328	11	29.82	3.73
Within	96	12	8	
Total	424	23		

$EMS = \sigma_e^2$

Two-way ANOVA

- Testing for interaction and the main factor– using two way ANOVA

$$\sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n}(x_{ijk}-\bar{x}_{...})^2 = cn\sum_{i=1}^{r}(\bar{x}_{i..}-\bar{x}_{...})^2 + rn\sum_{j=1}^{c}(\bar{x}_{.j.}-\bar{x}_{...})^2$$

SST — SSBR — SSBC

$$+ n\sum_{i=1}^{r}\sum_{j=1}^{c}(\bar{x}_{ij.}-\bar{x}_{i..}-\bar{x}_{.j.}+\bar{x}_{...})^2 + \sum_{i=1}^{r}\sum_{j=1}^{c}\sum_{k=1}^{n}(x_{ijk}-\bar{x}_{ij.})^2$$

INTERACTION SSI — SSW

Source	SS	df	MS	EMS	F
Between	328	11			
Exercise	144	3	48	$\sigma_\varepsilon^2 + cn\sigma_r^2$	6
Diets	112	2	56	$\sigma_\varepsilon^2 + rn\sigma_c^2$	7
ExD	72	6	12	$\sigma_\varepsilon^2 + n\sigma_{rxc}^2$	1.5
Within	96	12	8	σ_ε^2	
Total	424	23			

Two-way ANOVA

- Testing for interaction and the main factor– using two way ANOVA

Tests

H_{01}: exercise levels are the same i.e., $H_{01}: \mu_{1\cdot} = \mu_{2\cdot} = \mu_{3\cdot} = \mu_{4\cdot}$

$F_{.95}(3,12) = 3.49$ and $F = 6 > 3.49$ ∴ **reject H_0**

H_{02}: diets are the same i.e., $H_{02}: \mu_{\cdot 1\cdot} = \mu_{\cdot 2\cdot} = \mu_{\cdot 3\cdot}$

$F_{.95}(2,12) = 3.88$ and $F = 7 > 3.88$ ∴ **reject H_0**

H_{03}: there is no interaction between exercise level and diet

$F_{.95}(6,12) = 3.00$ and $F = 1.5 < 3.00$ ∴ **cannot reject H_0**

$$\frac{n(n+1)}{2} = \frac{15(16)}{2} = 120 = R_1 + R_2 + R_3$$

Nonparameric ANOVA

- The Kruskal-Wallis One-Way ANOVA
- Example: A certain chemical analysis of the urine of 15 infants yielded values
- as shown below:

I Term Infants	II Preterm Infants	III Preterm Infants With Acidosis (at 1-3 weeks of age)
4.5 (5)	3.2 (1)	7.3 (13)
3.9 (2)	4.6 (6)	8.4 (15)
5.0 (9)	5.1 (10)	6.9 (12)
4.8 (7)	4.9 (8)	8.2 (14)
4.1 (3)	4.3 (4)	6.2 (11)
$R_1 = 26$	$R_2 = 29$	$R_3 = 65$

$$\frac{n(n+1)}{2} = \frac{15(16)}{2} = 120 = R_1 + R_2 + R_3$$

Krusal-Wallis One-way ANOVA

- H0 :the three groups are identical w. r. t. this particular urine determination
- Ha :at least one of the three groups differ w.r. t. location in this variable

1. Combine all scores into a single series and rank them
2. Compute

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} R_j^2 / n_j - 3(n+1) \text{ where } n = n_1 + n_2 + n_3 = \frac{12}{15(16)} \left[26^2/5 + 29^2/5 + 65^2/5 \right] - 3(15+1) = 9.42$$

3. Look up value in Table O of Daniel if $n1$, $n2$, $n3$ are all $\leqslant 5$

Here Pr ($H \geqslant 8$) = .009 ⇒ Pr ($H \geqslant 9.42$) < .009

∴ reject H 0 at, say, the α = .05 level

Krusal-Wallis One-way ANOVA

- The K-W test has an asymptotic relative efficiency of .955 compared to the F test if the assumptions underlying the F test hold.

Krusal-Wallis One-way ANOVA

- The K-W test has an asymptotic relative efficiency of .955 compared to the F test if the assumptions underlying the F test hold.
- Example: The following table shows the net book value of equipment capital per bed for a sample of hospitals from each of 4 types of hospitals. Does the average net book value of equipment capital differ among the four types of hospital.

Krusal-Wallis One-way ANOVA

Type of Hospital			
A	B	C	D
1735 (5)	5260 (22)	2790 (11)	3475 (17)
1520 (2)	4455 (19)	2667 (8.5)	3115 (13)
1476 (1)	4480 (20)	2655 (7)	3050 (12)
1688 (3)	4325 (18)	2400 (6)	3125 (14)
1702 (4)	5075 (21)	2755 (10)	3275 (15)
2667 (8.5)			3300 (16)
R_j 23.5	100	42.5	87

H 0 : all hospitals are the same with respect to location

H a: there are differences w.r.t. location among the hospitals

$$H = \frac{12}{n(n+1)}\sum_{j=1}^{k}\frac{R_j^2}{n_j} - 3(n+1) = \frac{12}{22(23)}\left[\frac{23.5^2}{6} + \frac{100^2}{5} + \frac{42.5^2}{5} + \frac{87^2}{6}\right] - 3(23) = 19.1$$

here $H \sim x^2(3)$ $\Pr\left(x^2(3) > 12.838\right) = .005$

$$\text{p} - \text{value} = \Pr\left[\chi^2(3) > 19.1\right] < .005$$

∴ reject H0

at this point, we would do multiple comparisons. [out of the scope of this course].

Friedman Two-way ANOVA

- For Two-way ANOVA model, when normal distribution is not applicable, we can use Friedman Two-way ANOVA as a nonparametric test.
- Example :Nine subjects with bronchial asthma participated in an experiment to evaluate the relative effectiveness of three drugs. The following table shows the change in FEV1 (forced expired volume in 1 second) values (expressed as liters) two hours after drug administration.

Friedman Two-way ANOVA

H_0:There is no difference in the drugs within a block (patient)

H_1:Not all drugs are equal (at least one drug yields larger values than at least one of the others)

	DRUG		
Subject	A	B	C
1	.00 (1)	.13 (2)	.26 (3)
2	.04 (1)	.17 (2)	.23 (3)
3	.02 (1)	.20 (2)	.21 (3)
4	.02 (1)	.27 (3)	.19 (2)
5	.04 (1)	.11 (2)	.36 (3)
6	.03 (1)	.18 (2)	.25 (3)
7	.05 (1)	.21 (2)	.32 (3)
8	.02 (1)	.23 (2)	.38 (3)
9	.00 (1)	.24 (2)	.30 (3)
	$R_1 = 9$	$R_2 = 19$	$R_3 = 26$

$$\sum R_j = n\left(\frac{k(k+1)}{2}\right)$$

$$= 9\left(\frac{3(4)}{2}\right) = 54$$

Friedman Two-way ANOVA

H_0:There is no difference in the drugs within a block (patient)

H_1:Not all drugs are equal (at least one drug yields larger values than at least one of the others)

Steps

(1) Rank treatments within each block. If ties, give average rank within block.

(2) Sum ranks for each treatment over all subjects.

(3) Compute

$$x_r^2 = \frac{12}{nk(k+1)}\sum_{j=1}^{k}\left(R_j\right)^2 - 3n(k+1) = \frac{12}{9(3)(4)}\left[9^2+19^2+26^2\right] - 3(9)(4) = 16.22$$

(4) Look up p-value in Table Pa or Pb from Daniel.

There we see that the p-value =.000011 ∴ certainly reject H0

(5) If *n* and *k* are too large for the table, then

$$x_r^2 \sim x^2(k-1)$$

Friedman Two-way ANOVA

The asymptotic efficiency of the Friedman test relative to the F-test is .955k/k+1 under normality. It may be much better when distributions are non-normal.