脑功能统计学法

Wei Liao (廖伟), Ph.D.

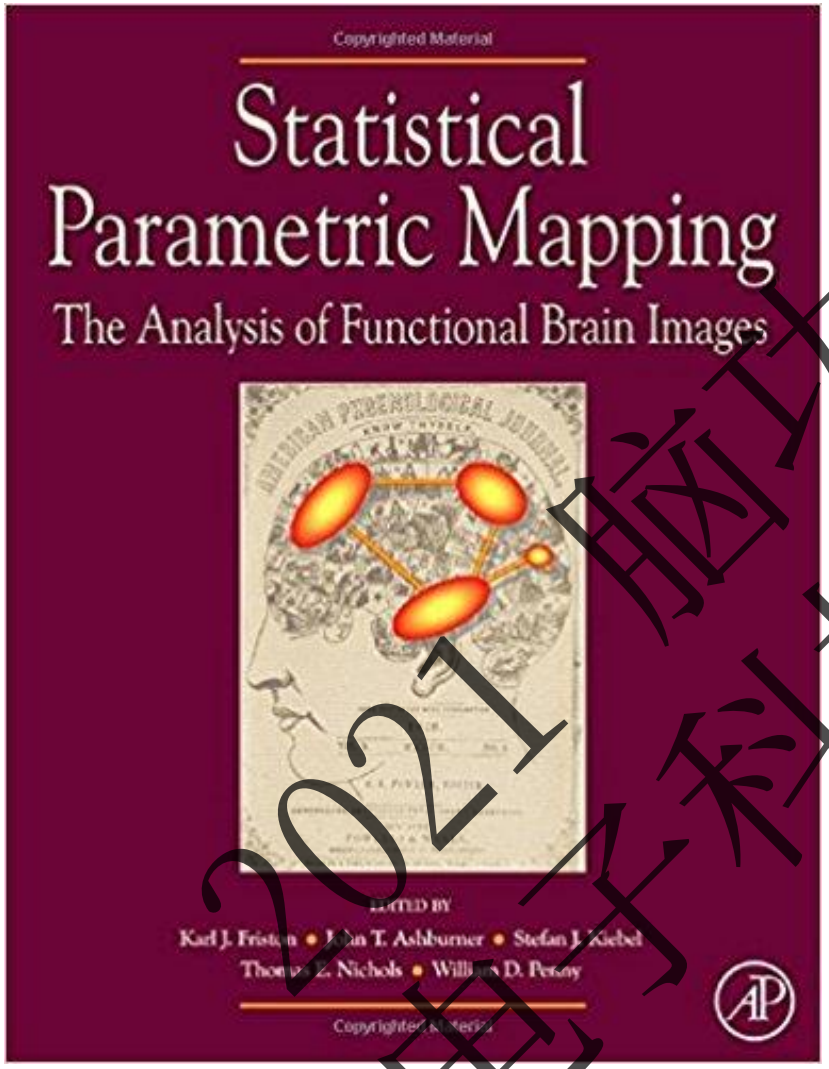
weiliao@uestc.edu.cn

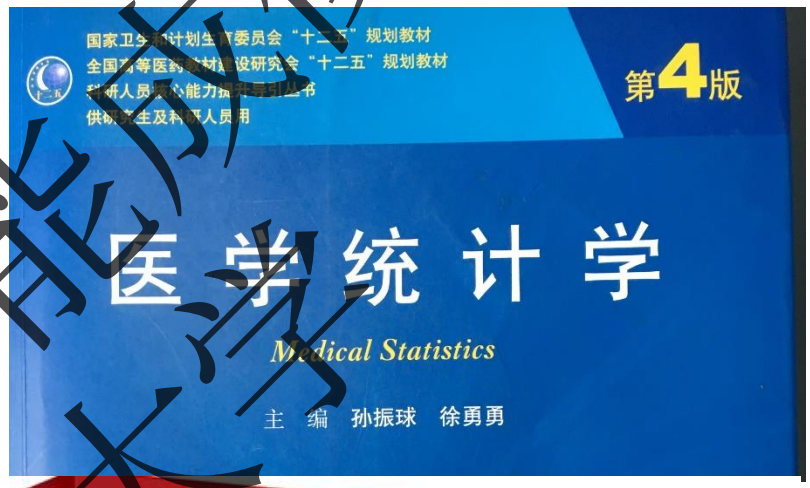
School of Life Science and Technology
University of Electronic Science and Technology of China

大纲

- 广义线性模型(GLM)
- 基本统计方法
- 高级统计方法

推荐阅读





Warm-up

- 目的:

  统计描述(descriptive statistics)

  统计推断(inferential statistics)

- 资料类型:

(1)数值变量(numerical variable) [定量或准确测量的变量，表现为数值大小的不同]

(2)分类变量(categorical variable) [只有不相容的类别或属性]

① 无序分类变量(计数资料): [无大小之分的属性或类别]

性别：两类无序变量
血型：四类无序变量

② 有序分类变量(等级资料): [各属性或类别有程度之分]

疾病程度：轻、中、重
治疗效果：无效、显效、好转、治愈

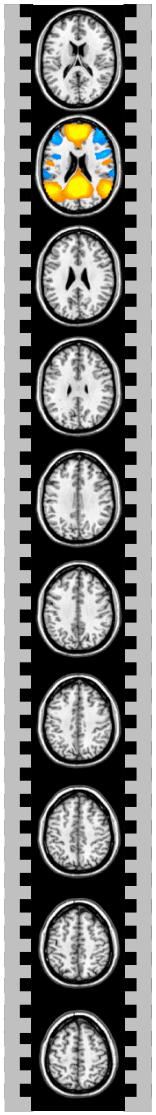
Warm-up

- 医学统计方法:

①差异性比较
- 两组—t-test
- 多组—ANOVA

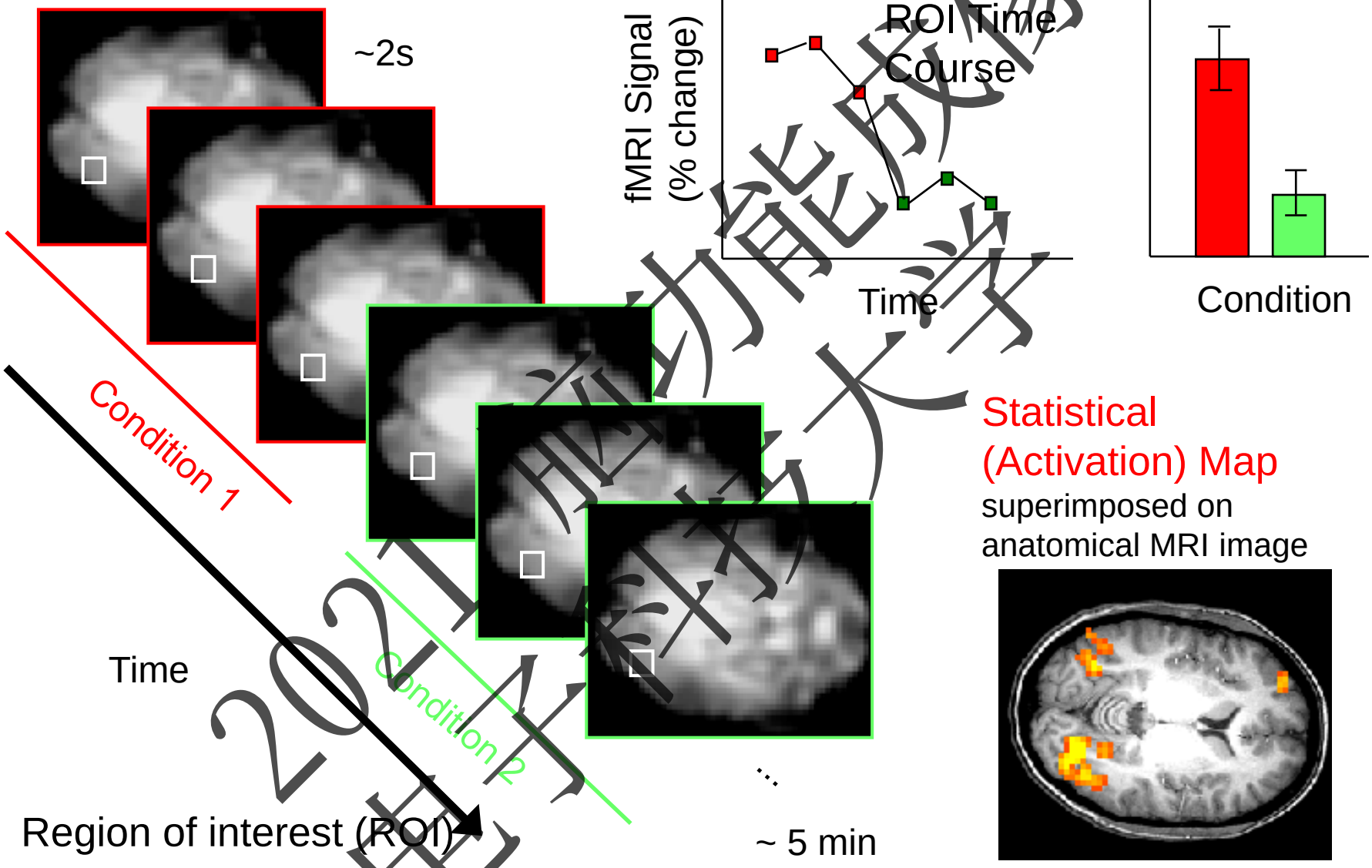
②相关性分析
- 定量资料
- 定性资料

③影响性分析
- 结局为定量资料—线性回归
- 结局为定性资料—Logistic回归
- 结局为生存资料—Cox回归

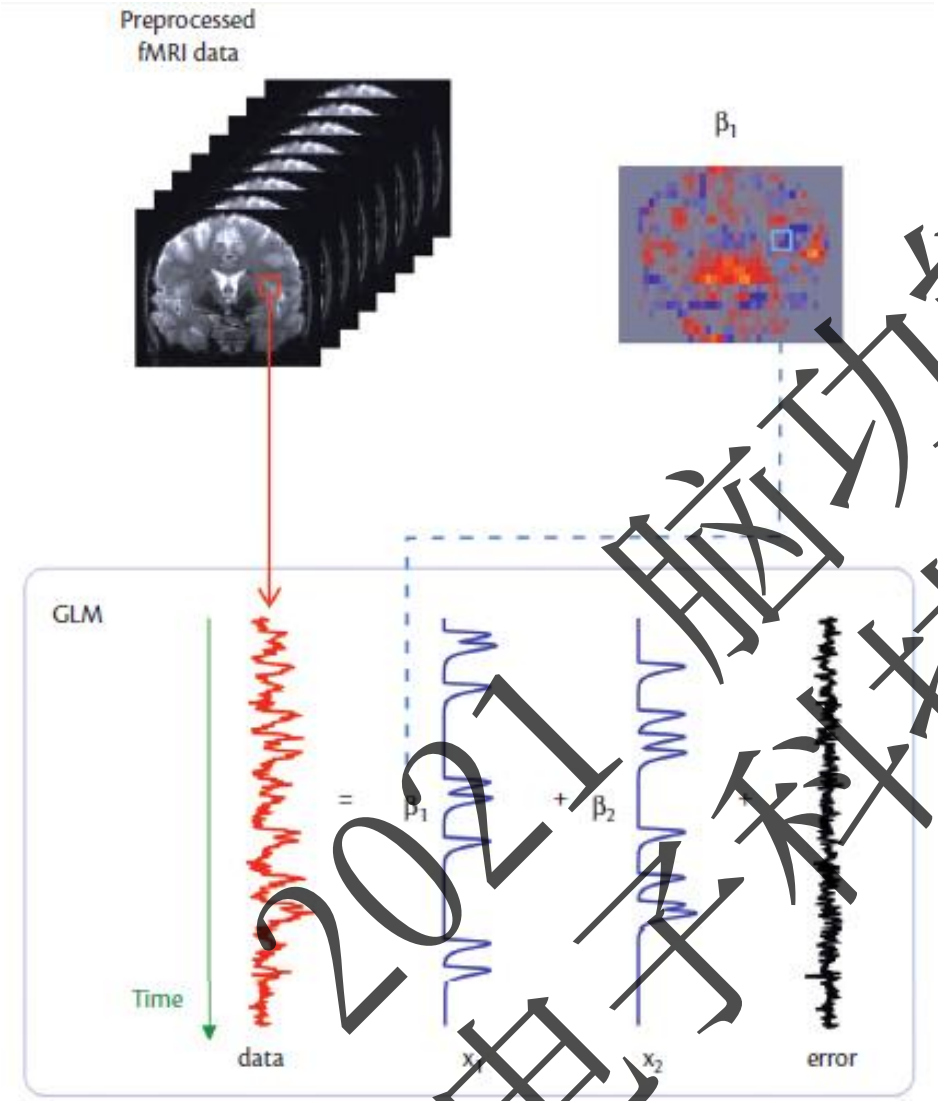


- Background (GLM)
- t-test
- One-way ANOVA
- Two-way ANOVA
- Issues

Statistical Parametric Mapping

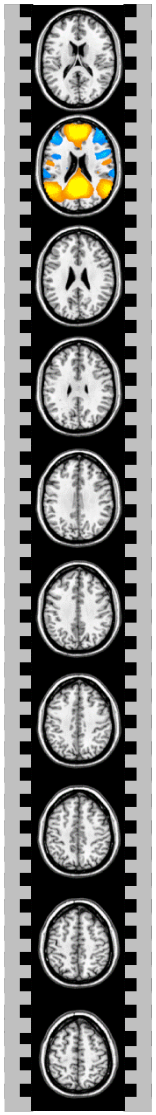


General Linear Model (GLM)



model (X) containing a set of regressors (X1 and X2). One beta value (representing an amplitude) is calculated for each of the regressors included in the model (i.e., Β 1 for X1 and Β2 for X2), and what is left over is the error (i.e., the residuals).

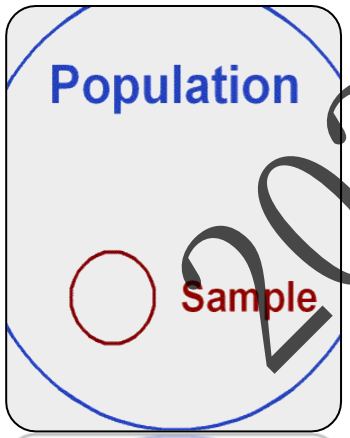
This is typically done for all voxels separately, known as a voxelwise analysis, and the results (beta values, or probabilities based on them) are stored, displayed, and possibly analyzed further in the form of voxelwise maps (images, as shown for Β1 in the top right.).

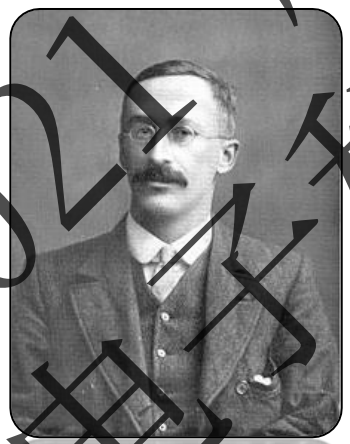


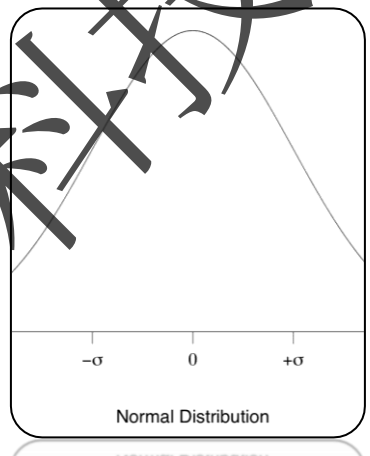
- Background (GLM)
- t-test
- One-way ANOVA
- Two-way ANOVA
- Issues

t-test—Background

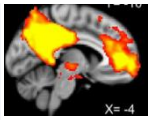
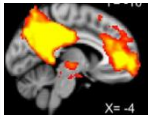
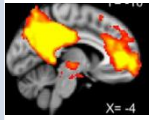
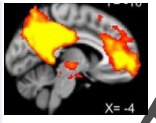
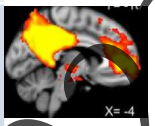
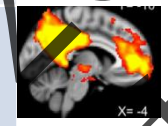
- Samples vs Populations
- Descriptive vs Inferential
- William Sealy Gosset ('Student')
- Distributions, probabilities and P-values
- Assumptions of t-tests







t-test—one sample/group *t*-test

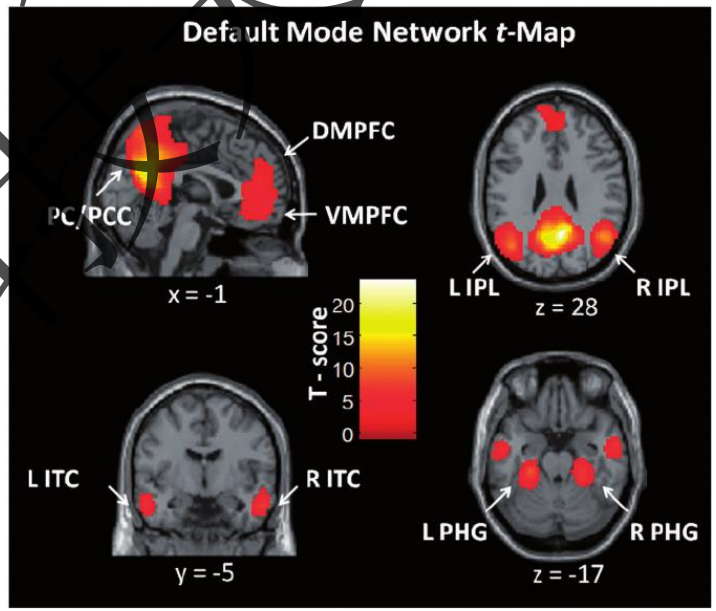
Group 1	
S1	
S2	
S3	
S4	
S5	
…	
Sn	

- **Is there a significant correlation of PCC for one group**

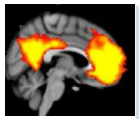
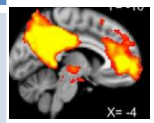
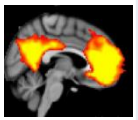
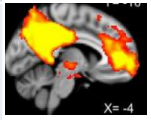
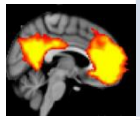
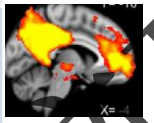
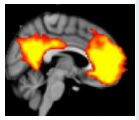
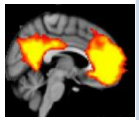
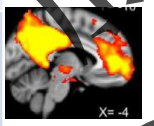
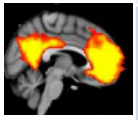
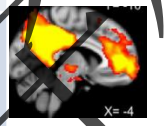
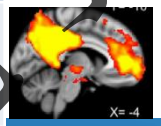
H0: μ=μ0=0, no corr.

H1: μ≠μ0=0, a significant corr.

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



t-test—two-sample/group *t*-test

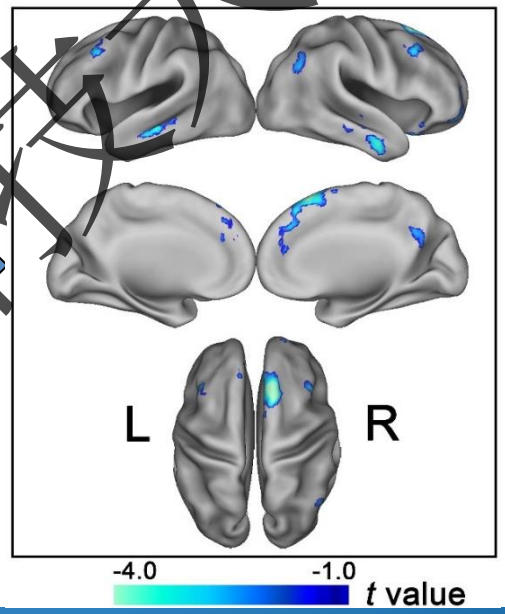
	Group 1	Group 2	
	P1	S1	
	P2	S2	
	P3	S3	
	P4	S4	
	P5	S5	
	…	…	
	Pn	Sn	
		Sm	

- Is there different connectivity of PATIENTs and CONTROLs.

Between-subjects design

H0: μ1=μ2, no difference

H1: μ1≠μ2, significant diff.



$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

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

t-test—Paired/matched *t*-test(I)

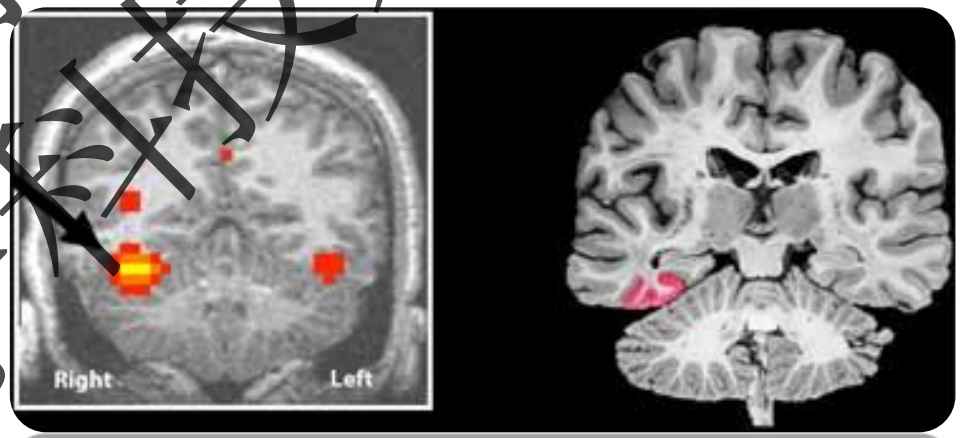

Cond 1	Cond 2
S1	S1
S2	S2
S3	S3
S4	S4
S5	S5
…	…
Sn	Sn

- **Is there different activation of the FFG for faces vs objects**

Within-subjects design

H0: μ1=μ2, There is no difference in activation of the FFG during face vs object stimuli

H1: μ1≠μ2, There is a significant difference in activation of the FFG during face vs object stimuli



t-test—Paired/matched *t*-test(II)

Pre-treatment (Baseline)	Post-treatment (T1)
Group 1	Group 2
P1_t0	P1_t1
P2_t0	P2_t1
P3_t0	P3_t1
P4_t0	P4_t1
P5_t0	P5_t1
…	…
Pn_t0	Pn_t1

- Is there different connectivity of Time 0 and Time 1.

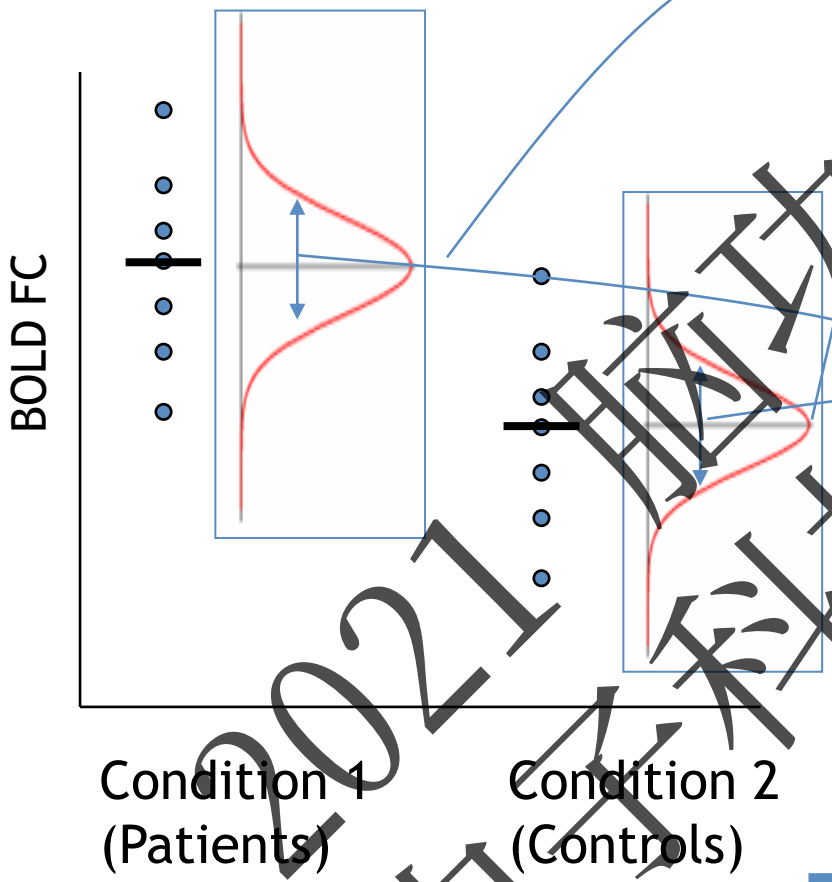
Within-subjects design

H0: μ1=μ2, no difference
H1: μ1≠μ2, significant diff.

t-test—Calculating t

* Independent Samples t-test

- t = differences between sample means / standard error of sample means



$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}}$$

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

- The exact equation varies depending on which type of t-test used

t-test—Alternatives

- 1 Sample t-test (sample vs. hypothesized mean) $t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}},$
- 2 Sample t-test (group1/condition 1 vs group2/condition 2)

	Independent Samples	Related Samples also called dependent means test
Interval measures/ parametric	Independent samples t-test $t = \frac{\overline{x}_1 - \overline{x}_2}{s_{\overline{x}_1 - \overline{x}_2}}$	Paired samples t-test $t = \frac{\overline{X}_D - \mu_0}{s_D/\sqrt{n}}.$
Ordinal/ non-parametric	Mann-Whitney U-Test	Wilcoxon test

t-test—Degrees of Freedom (*df*)

- The number of 'entities' that are free to vary when estimating t
- *n* – 1 (for paired sample t; one-sample)
- *n1*+*n2*-2 (for two-sample)
- Larger sample or no. of observations = more df

Putting it all together…

- t (*df*) = t: t-value, p : p-value

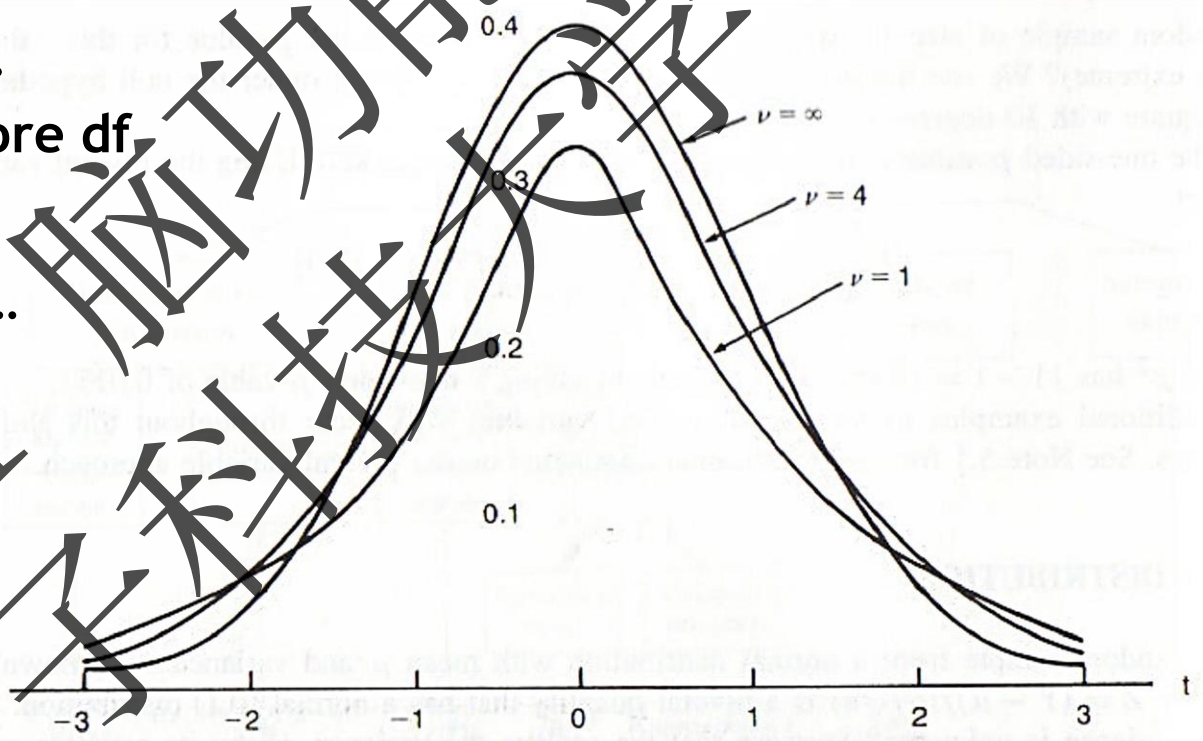
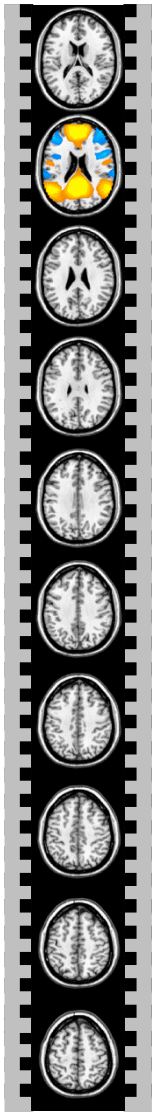


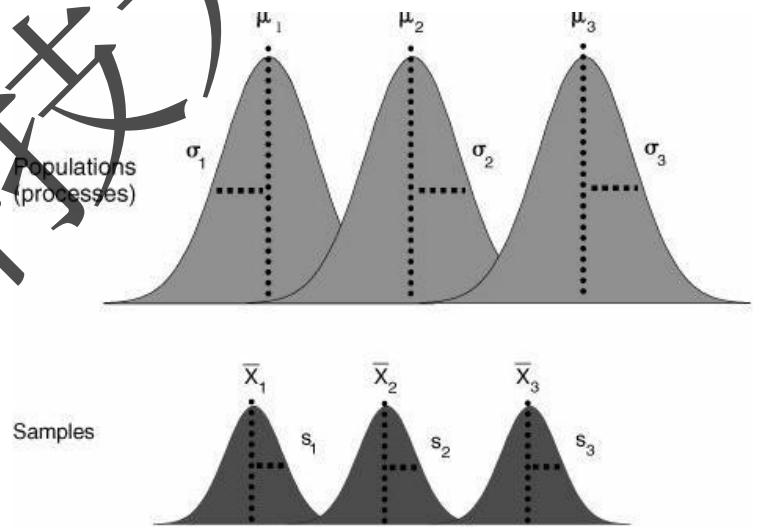
Figure 5.2 Student *t*-distribution with one, four, and ∞ degrees of freedom.



- Background (GLM)
- t-test
- One-way ANOVA
- Two-way ANOVA
- Issues

ANOVA- Analysis of Variance

- More than 2 groups and/or conditions
  - e.g. objects, faces and bodies
- Do this without inflating the Type I error rate
- Still compares the differences in means between groups/conditions but it uses the variance of data to calculate if means are significantly different (H_A)
- Tests the null hypothesis that the means are the same via the F- test
- Extra assumptions



How? The F-statistic

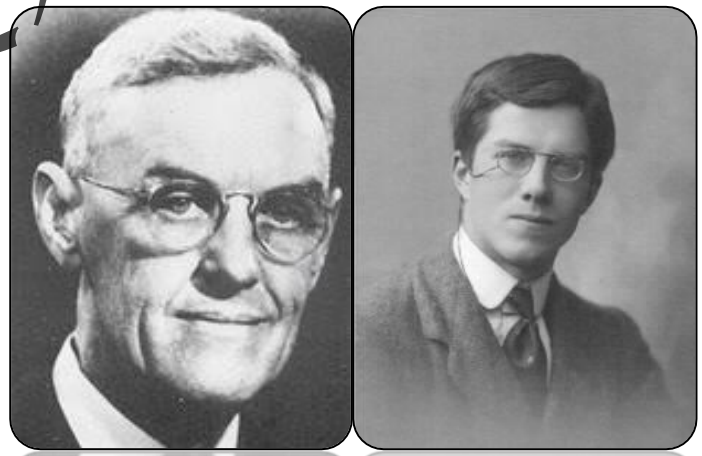
F-ratio = MS_M / MS_R

- By comparing the variance ($SS_T = SS_M + SS_R$)
  SS_T (variability between scores)
  SS_M (variability explained by model)
  SS_R (variability due to individual difference)

÷ df_M ÷ df_R

- F- ratio
  - Magnitude of the difference between the different conditions
  - p-value associated with F is probability that differences between groups could occur by chance if null-hypothesis is correct
  - need for post-hoc testing / planned contrasts (ANOVA can tell you if there is an effect but not where)

$$F = \frac{\text{explained variance}}{\text{unexplained variance}},$$



① Sir Ronald Aylmer Fisher

One-way ANOVA—Between Groups

One Factor, 3+ levels

Level 1	Level 2	Level 3
M1	S1	HC1
M2	S2	HC2
M3	S3	HC3
M4	S4	HC4
M5	S5	HC5
…	…	…
Mn	Sn	HCn
	Sm	HCm
		…
		HCk

- Is there different connectivity among MDD, SZ and HC.

Between-groups design

H0: μ1=μ2=μ3, no difference
H1: Means are not all equal.

df_m=? df_m=g-1

df_R =? df_R =N-g

One-way ANOVA—Repeated

One Factor, 3+ levels

- Is there different connectivity among Time 0, 1 and 2.

within-groups design

H0: μ1=μ2=μ3, no difference
H1: Means are not all equal.

Pre-treatment (Baseline)	Post-treatment (2 weeks)	Post-treatment (4 weeks)
Level 1	Level 2	Level 3
M1_T0	M1_T1	M1_T2
M2_T0	M2_T1	M2_T2
M3_T0	M3_T1	M3_T2
M4_T0	M4_T1	M4_T2
M5_T0	M5_T1	M5_T2
…	…	…
M7_T0	M7_T1	M7_T2

$$F = \frac{MS_{between}}{MS_{error}}$$

F(2,12)=

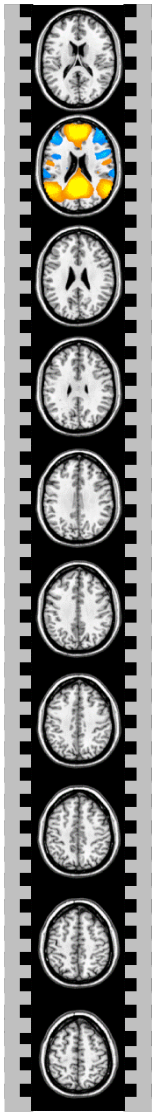
$$df_{Between} = a - 1 = 3 - 1 = \mathbf{2}$$

$$df_{Within} = N - a = 21 - 3 = \mathbf{18}$$

$$df_{Subjects} = s - 1 = 7 - 1 = \mathbf{6}$$

$$df_{Error} = df_{Within} - df_{Subjects} = 18 - 6 = \mathbf{12}$$

$$df_{Total} = N - 1 = 21 - 1 = \mathbf{20}$$



- Background (GLM)
- t-test
- One-way ANOVA
- Two-way ANOVA
- Issues

Two-way ANOVA

2 way (_ x _) ANOVA and even 3 way ANOVA

- Two or more factors and many levels:

Type	2-way ANOVA for independent groups	repeated measures ANOVA	mixed ANOVA
Participants			

	Condition I	Condition II
Task I	Participant group A	Participant group B
Task II	Participant group C	Participant group D

	Condition I	Condition II
Task I	Participant group A	Participant group A
Task II	Participant group A	Participant group A

	Condition I	Condition II
Task I	Participant group A	Participant group B
Task II	Participant group A	Participant group B

Two-way ANOVA—Between Groups

A / B	Level 1	Level 2
Level 1	s_1 s_2 s_3 s_4 s_5	s_{16} s_{17} s_{18} s_{19} s_{20}
Level 2	s_6 s_7 s_8 s_9 s_{10}	s_{21} s_{22} s_{23} s_{24} s_{25}

健康不吸烟	健康吸烟
精分不吸烟	精分吸烟

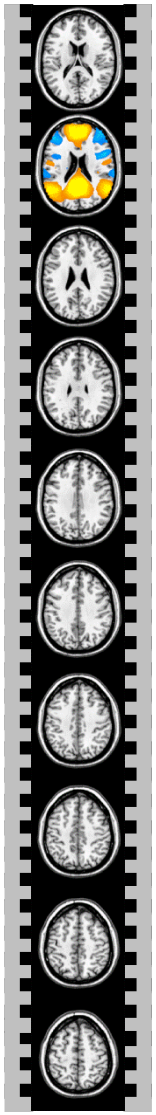
- Simple effect (4)
- Main effect (2)
- Interaction (1)

Two-way ANOVA—Repeated

A / B	Level 1	Level 2
Level 1	s_1 s_2 s_3 s_4 S_5	s_1 s_2 s_3 s_4 S_5
Level 2	s_1 s_2 s_3 s_4 S_5	s_1 s_2 s_3 s_4 S_5

Two-way ANOVA—Mixed

A / B	Level 1	Level 2
Level 1	s_1 s_2 s_3 s_4 s_5	s_6 s_7 s_8 s_9 s_{10}
Level 2	s_1 s_2 s_3 s_4 s_5	s_6 s_7 s_8 s_9 s_{10}



- Background (GLM)
- t-test
- One-way ANOVA
- Two-way ANOVA
- Issues

Issues

- Model selection (be careful!)
- Post-hoc analysis
- Multiple comparisons

推荐阅读

thebmj Research Education News & Views Campaigns

Endgames » Statistical Question

Units of sampling, observation, and analysis

BMJ 2015 ; 351 doi: https://doi.org/10.1136/bmj.h5396 (Published 09 October 2015)
Cite this as: *BMJ* 2015;351:h5396

Clinical Review & Education

JAMA Guide to Statistics and Methods

Logistic Regression Diagnostics
Understanding How Well a Model Predicts Outcomes

NATURE METHODS | THIS MONTH

Points of significance: Importance of being uncertain

研究生论文写作指导（1411117001）

2022春