心理统计
第十七讲：初级心理统计复习

严超赣

Chao-Gan Yan, Ph.D.

yancg@psych.ac.cn
http://rfmri.org/yan

Institute of Psychology, Chinese Academy of Sciences

1

描述统计和推论统计

- 描述统计（Descriptive statistics） 总结，组织，和使数据简单化的统计程序。
- 推论统计（Inferential statistics） 使我们能够通过对样本的研究将其结果推广于总体。

2

自变量和因变量 (Independent and Dependent Variables)

- 一个好的研究设计是来考察是否某个特定的变量(IV) 影响另一个变量 (DV).
  - 自变量(*independent variable)* 是原因. 在真实验中,**自变量是研究者所操纵的.**
  - 因变量(*dependent variable)* 是测量自变量效果的变量,需要是可观察的和可测量的.

3

参数和统计量

- 参数（parameter）-- 描述总体的数值。参数可以从总体的一系列测量中推论得到。
- 统计量 （statistic）描述样本的数值。统计量可以从一次测量中获得。
- 往往用不同的数学符号来表示参数和统计量 。如总体平均数用μ来表示，而样本平均数用$\bar{x}$来表示。
- 取样误差（Sampling error）样本统计量与相应的总体参数之间的差距。

4

数据的测度

- 等距测度vs比例测度
  - **两个邻近范畴之间的距离都是相等的**
  - 二者都可以使用参数检验
  - 只有对于比例测度,"0"才表示不具有所测量的属性
  - 只有对于比例测度,才可进行**乘除运算**

5

在各种测度的数据描述相关关系

- 命名等级:
- 顺序等级:
- 等距/比例等级:

6

次数分布 – 最简单的描述统计

- 描述统计的目的：简化和整理数据的表达。
- 次数分布（Frequency Distribution）：是指一批数据在某一量度的每一个类目所出现的次数情况
- 组织此类数据的第一种方法是：建立次数分布表
- 次数分布表和次数分布图就是表达一组数据是如何在某一度量上分布的

7

对称分布（*symmetrical distribution*）

- 可以画一条垂直线穿过分布的中央，使得分布的一边恰是另一边的镜象。

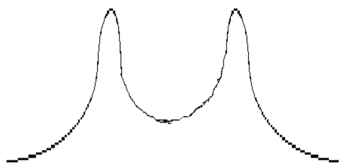
8

偏态分布（*skewed distribution*）

- 分数堆积在分布的一端，而另一端成为比较尖细的尾端（tail）
  - 偏态分布尾端向右的称为正偏态（*positively skewed*）(因为其尾端指向正数)
  - 偏态分布尾端向左的称为负偏态（*negatively skewed*）.

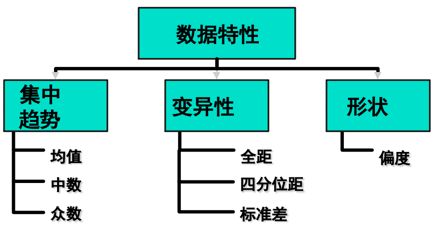
9

双峰分布（bi-modal distribution ）



10

数据的特性



11

11

Z 分数的意义

- 标准分数（*standard score*） 是一种转换分数提供其分布位置的信息. Z 分数是标准分数的一种。
- *z-score* 指出了每个X 值在分布中的精确位置。z-score 的符号(+ 或 -) 表明其比均值大或小. z-score 的数值部分用X 与 μ间标准差个数的形式指出了其与均值的距离。

$$Z= (X- \mu) /\sigma \rightarrow X = (Z)(\sigma) + \mu$$

12

取样分布和样本均值的分布

- 样本均值分布 (distribution of sample mean)：
  - 总体中可抽取的所有可能的特定容量(n)的随机样本的均值的集合。
- 抽样分布 （sampling distribution ）：
  - 总体中可抽取的所有可能的特定容量分布的统计量所形成的统计分布。
  - 样本均值分布是抽样分布的特例
- 为了对总体均值作出最佳估计,我们所要做的就是考察所有可能的样本（n一定，这点很重要）然后根据其特性作出预测。

13

样本均值分布的特性1：形状

- 样本均值的分布形状接近正态分布.
- 当 *n* 较大时(30 以上), 样本均值的分布几乎是完全的正态分布. 而不在乎原始分布的形状
- 如果在同一总体中选择一组样本, 大部分均值应当堆积在总体均值 μ 附近
- 如果不是这样，取样一定有偏差

14

样本均值分布的特性 2：均值

- 这些样本均值的平均应该等于总体均值. 样本均值的平均 叫做 X 的期望值. 期望值的意思因为这个值会在总体均值 μ 的附近.
- 如果 *n* 足够大, 那么分布是正态, 也一定是对称和单峰, 则mean = median = mode

15

样本均值分布的特性 3：变异性

- 样本均值分布的变异性用标准误来描述
- 定义：样本均值分布的标准差 叫做 X 的标准误 （standard error of *X*；SE）

X 的标准误= σ_X = X 与 μ 的标准距离.
  - 这个统计量描述了与均值的标准(或称典型，平均)距离. 在这里，它也是样本均值 X 和 总体均值 μ 的差值.
  - 这个统计量的主要目的和用途是告诉我们样本均值对总体均值的估计是否准确. 换言之，取样误差是多大.
- 标准误的数值取决于两个特征: 总体方差和样本容量

16

中心极限定律 (Central Limit Theorem)

- 综合了样本均值分布的特性：形状，均值，方差
- 定律：对于任何均值为μ ，标准差为σ的总体，样本容量为n的样本均值的分布，随着n趋近无穷大时，会趋近均值为μ，标准差为 $\sigma/\sqrt{n}$ 的正态分布
- 因此，当 n 足够大时(30或以上)：

$$\bar{x} \sim N(\mu\ ,\ \frac{\sigma}{\sqrt{n}})$$

17

假设检验

- 假设检验 （Hypothesis testing）是一种用样本数据来评价有关总体的某一假设的可置信性的推论程序。
- 以样本为依据，对总体作出判断.

18

步骤总结

1. 陈述虚无假设和研究假设.
2. 设定决策标准找出临界.
   如a（alpha）=.05
3. 收集数据，计算样本统计量.
4. 作出决策.
   如果得到的z分数落在拒绝区域中，就拒绝H_0

19

I类错误和II类错误

		真相	
		没有效应 H_0 正确	真实效应 H_0 错误
统计决策	拒绝 H_0	I类错误	正确
	接受 H_0	正确	II类错误

20

单尾检验（one-tailed test）与双尾检验（two-tailed test）

- 单尾检验：
  - 假设处理会在某一特定方向上造成差异 (即，处理会使均值增加).
- 双尾检验：
  - 假设检验最提出的方式是作一个更一般的假设：处理应当改变均值，或增加或减少.

21

21

Z检验的步骤

1. 陈述 H_0 和 H_1； 确定显著性标准 α = ?
2. 确定检验是单尾还是双尾
3. 确定临界 z 分数
4. 计算样本的实际 z 分数
5. 比较样本的实际 z 分数与临界 z 分数
6. 对 H_0 作出结论

22

22

z检验的前提（Assumptions）

1. 随机样本 - 样本必须对总体有代表性。随机取样有助于确保取样的代表性
2. 独立观察 – 也与样本代表性有关，每个观察应该与所有其它观察是独立的。一个特定的观察的概率应当保持恒定
3. σ保持恒定 - 原总体的标准差必须保持恒定. 为什么？一般的说，处理就是假定对总体中的每一个个体都加上(或减去) 一个常数。所以总体的均值可能因处理而导致变化。但是，记住对每一个个体都加上(或减去)一个常数 并不改变其标准差
4. 取样样本是相对正态的 – 或者因为原始观察的样本是相对正态的，或者因为中心极限定理(或二者都有)
   违反以上任何一个前提会严重地危及依据样本对总体作出推论的有效性(应付种种违反前提的情况，其它类型的推论统计需要用到)。

23

23

I 类错误, II类错误, & 效力的关系

		ACTUAL SITUATION	
		NO EFFECT, H_0 TRUE	EFFECT EXISTS, H_0 FALSE
EXPERIMENTER'S DECISION	Reject H_0	Type I error α	Dc $1-\beta$ ct
	Retain H_0	Decision correct	Type II β

24

24

提高效力的途径

- 增大样本量.
- 增加处理效应.
- 减少误差.
- 降低 alpha 水平.
- 采用单尾检验.

代价

值得推荐，因为不需花代价

-- 增加I类错误

-- 必须选择正确的尾端

25

25

t 与 z 的不同适用条件

σ已知	σ未知
$\sigma_M = \frac{\sigma}{\sqrt{n}}$	$s_M = \frac{s}{\sqrt{n}}$
$z = \frac{M-\mu}{\sigma_M}$	$t = \frac{M-\mu}{s_M}$

26

26

自由度(degree of freedom)

- 自由度描述了样本中可以自由变化的分数的数目。因为样本均值对于样本中的分数值构成了限制，所以样本有 *n* - 1 个自由度 。
- n 的数目越大, 样本对总体的代表性越好, 也就意味着s 是σ 的更好估计值。
- 其对考验统计量的意义是： t-分布的形状是样本容量 n的函数。 更确切地说， t-分布的形状是自由度df 的函数。 n 的数目越大(或df越大), t-分布就越接近正态分布.

27

27

t检验的步骤

1. 陈述 H_0 和 H_1； 确定显著性标准 α = ?
2. 确定检验是单尾还是双尾
3. 确定检验的自由度df
4. 查表求临界 t 分数
5. 计算样本的实际 t 分数
6. 比较样本的实际 t 分数与临界 t 分数
7. 对 H_0 作出结论

➢ (tobs = 计算出的 t 分数；tcrit =表中的临界 t 分数)

28

28

独立样本t检验的假设

- 虚无假设
  - 独立样本所来自的两个总体的均值之间没有显著的差异，即所抽取的两个样本来自同一个总体
  - H_0： μ_1-μ_2=0
- 备择假设
  - 独立样本所来自的两个总体的均值之间有显著差异
  - H_1： μ_1- $\mu_2 \neq 0$

29

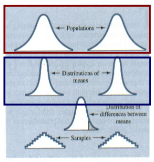
29

总体方差估计

- 单样本的T检验

$$s^2 = \frac{SS}{n-1} = \frac{SS}{df}$$

$$s_{\bar{x}}^2 = \frac{s^2}{n}$$



- 独立样本的T检验
  - 总体方差的合并估计值S_{pooled}^2
    - 两个样本可能大小不等，所以是加权平均
    - 权重不是样本数目，是自由度（样本数目-1）

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2}$$

注意只有两样本方差大体相等（即满足方差同质性）时才可以用

    - 均值分布的方差的计算

$$s_1^2 = \frac{s_p^2}{n_1} \quad s_2^2 = \frac{s_p^2}{n_2}$$

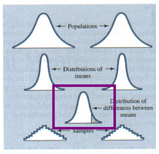
30

30

标准误和t的计算

- 单样本t检验
  - 标准误

$$s_{\bar{X}} = \frac{s}{\sqrt{n}} = \sqrt{\frac{s^2}{n}}$$



- 独立样本t检验
  - 均值差异样本的方差和标准差
    - 均值差异样本的方差是总体1的均值分布的方差加上总体2的均值分布的方差

$$\boxed{s_{\bar{X}_1-\bar{X}_2}{}^2 = s_1{}^2 + s_2{}^2}$$

$$s_1{}^2 = \frac{s_p{}^2}{n_1} \qquad s_2{}^2 = \frac{s_p{}^2}{n_2}$$

$$s_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{S_p^2}{n_1} + \frac{S_p^2}{n_2}}$$

31

31

t统计量的计算

- 单样本的t检验

$$t = \frac{\bar{X} - \mu}{s_{\bar{X}}}$$

- 独立样本的t检验

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1-\bar{X}_2}} = \frac{(\bar{X}_1 - \bar{X}_2)}{s_{\bar{X}_1-\bar{X}_2}}$$

32

32

独立样本t检验的统计前提

1. 观察间彼此独立（同前）
2. 两个总体均为正态分布(同前)
3. 两个总体具有相等的方差. 这称为方差同质性（homogeneity of variance）
   - 回顾合并样本方差的公式，只有方差总体方差大体相等时才可以用；如果差异太大，则不能用
   - 方差同质性的Hartley’s F-max 检验

$$F_{\max} = \frac{s_1{}^2}{s_2{}^2}$$

   - 拇指原则: 对于小样本 (n < 10)，如果一个样本方差 (s^2) 比另一个大4倍以上, 大概不会满足方差同质性假设。对于大一些的样本, 如果一个样本方差 (s^2) 比另一个大2倍以上, 多半会违反方差同质性前提.

33

33

独立样本t检验的效应大小

- 效应大小 (effect size, ES)
  - 两个总体分布的重叠程度
    - ES越大，重叠程度越小，效应越明显
    - ES越小，重叠程度越大，效应越不明显

$$EffectSize = \frac{X_1 - X_2}{S_p}$$

ES=(37-31)/4.472=1.34

34

34

相关样本的意义

- 被试内设计，两组数据不存在组间差异
- 匹配组t-test 是有一个或若干个特征使得被试内两两建立联系，这种联系是实验前建立的，分析数据时已匹配成对
- 匹配能大大减少个体间的误差，提高统计效力

控制了数据中个体差异引起的误差

35

35

相关样本t检验的逻辑

- 相关样本t统计量的计算是基于样本分数的差异，而不是原始分数
- 每一对对应数据的差异D构成了一个差异样本
- 这个差异样本的
  - 平均数 $D = X_1 - X_2$
  - 方差

$$s^2 = \frac{SS_D}{df} = \frac{\sum D^2 - \frac{(\sum D)^2}{n}}{df}$$

36

36

相关样本t的计算

- 单样本t检验

$$t=\frac{\bar{X}-\mu}{s_{\bar{X}}}$$

$$\Uparrow$$

$$s_{\bar{X}}=\sqrt{\frac{s^2}{n}}$$

$$\Uparrow$$

$$s^2=\frac{SS}{n-1}=\frac{SS}{df}$$

$$\Uparrow$$

$$SS=\sum X^2-\frac{(\sum X)^2}{N}$$

- 相关样本t检验

$$t=\frac{\bar{D}-\mu_D}{s_{\bar{D}}}$$

$$\Uparrow$$

$$\boxed{s_{\bar{D}}=\sqrt{\frac{s^2}{n}}}$$

$$\Uparrow$$

$$s_D{}^2=\frac{SS_D}{n-1}=\frac{SS_D}{df}$$

$$\Uparrow$$

$$SS_D=\sum D^2-\frac{(\sum D)^2}{N}$$

37

37

	样本数据	假设总体参数	样本标准差	标准误	t-统计量
单样本t-检验	$\bar{X}$	μ	$s^2=\frac{SS}{df}$	$\sqrt{\frac{s^2}{n}}$	$t=\frac{\bar{X}-\mu}{s_{\bar{x}}}$
相关样本t-检验	$\bar{D}$	μ_D	$s^2=\frac{SS_D}{df}$	$\sqrt{\frac{s^2}{n}}$	$t=\frac{\bar{D}-\mu_D}{s_{\bar{D}}}$
独立样本t-检验	$\bar{X}_1-\bar{X}_2$	$\mu_1-\mu_2$	$s_p^2=\frac{SS_1+SS_2}{df_1+df_2}$	$\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}$	

$$t=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{s_{\bar{x}_1-\bar{x}_2}}$$

38

估计

- 样本 → 总体
  - 假设检验
  - 估计
- 影响估计的因素
  - 样本容量：2个或100个
  - 估计的类型：点估计或区间估计
    - 6两
    - 0.5两-4斤
    - 在精确度与置信区间，我们有一个trade-off

39

估计的逻辑

- 估计的逻辑与假设检验不同
  - 假设检验：试图否定虚无假设。
  - 估计： 对于总体参数的值作有根据的猜测。
- 何时需要作估计
  - 想了解总体的基本信息，但不能测量到所有个体，所以抽取一个样本
  - 如果已经知道处理有效应，进而想知道效应有多大
  - 经过假设检验后拒绝了H_0
    - "我们拒绝处理没有造成差别，但我们希望知道到底有多少差别。"

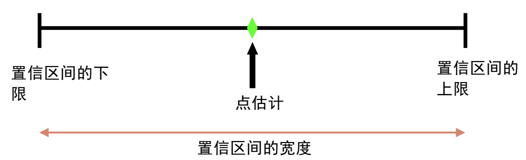
40

两类总体均值估计

- 均值的点估计（*point estimates*）: 用单一数值作为未知数量的估计
- 均值的区间估计（*interval estimates*）： 用某一数值的范围作为未知数量的估计.
  - 置信区间（confidence intervals；C.I.）当一个区间与一个特定的置信度 (或概率)一起出现时，称为置信区间.
- 两类估计都由同样的方程所决定，其差别是对于点估计，只计算一个单一的数值，但对于区间估计,应计算两点之间的区间。

41

点估计和区间估计



42

一般公式

- 置信区间的一般公式:

点估计 ± (临界值)(标准误)

43

μ的置信区间(σ已知)

- 前提
  - 总体标准差σ已知
  - 总体正态分布
  - 如果总体不是正态,用一个大样本
- 置信区间的估计值:

$$\boxed{\bar{X} \pm Z\frac{\sigma}{\sqrt{n}}} \qquad Z = \frac{X-\mu}{\sigma_X}$$

- $\bar{X}$ 是点估计值

  Z 尾端α/2是是正态分布临界值

  $\sigma/\sqrt{n}$ 是标准误

44

单因素设计和因素设计

- 在方差分析中,因素就是自变量.
- 包含一个组间自变量的研究称为单因素设计（single-factor design）.
- 具有多于一个自变量研究称为因素设计（factorial design）.
- 构成因素的个别处理条件称为因素的水平.
- 上述研究称为因素设计, 两个组间因素,每一个因素有 3 个水平 (称为 3 X 3 组间设计).

45

1. ANOVA的逻辑

- 与假设检验的逻辑是同样的, 只是具体内容有变化
- step 1: 陈述 H_0 (和H_1 ??)，确定标准: a = ?
- step 2: ANOVA 检验总是单尾
- step 3: 指出检验的df (有两个df)
- step 4: 查表找出临界 F统计量
- step 5: 对于样本，计算F统计量
- step 6: 比较 F统计量和临界 F统计量
- step 7: 对于H_0 作出结论

46

计算样本的F统计量观测值

- 两个独立样本的t检验

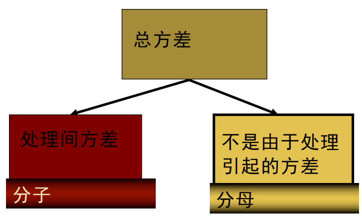
$$t_{obs} = \frac{\text{得到的样本均值间差异}}{\text{期望的机会差异}}$$

- 对于 ANOVA检验统计量（称为 F比率）类似

$$F = \frac{\text{样本均值间方差（差异）}}{\text{期望的误差方差}}$$

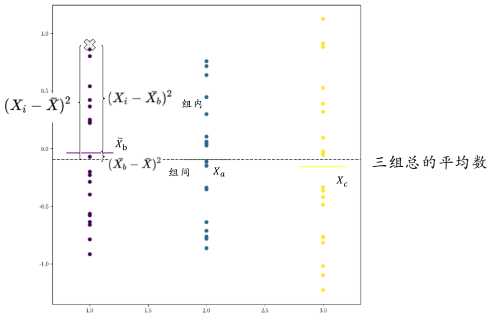
47

方差的分解



48

方差的分解



49

方差的分解

假定H0为真，则各组之间不存在差异 $\bar{X}_b \longrightarrow \bar{X}$

$$\sum(X_i-\bar{X})^2=\sum(X_i-\bar{X}_b)^2+\sum(\bar{X}_b-\bar{X})^2$$

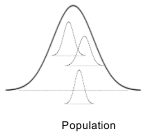
$$\downarrow \qquad \downarrow$$

$$\sum(X_i-\bar{X})^2 \qquad 0$$

$$SS_T=SS_W+SS_B$$

$$\downarrow \qquad \downarrow$$

$$SS_T \qquad 0$$



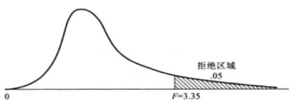
50

F比率

*F*值

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

当H0为真时，F值服从一个广为人知的概率分布。为了纪念R.A.Fisher，这个分布称为F分布。

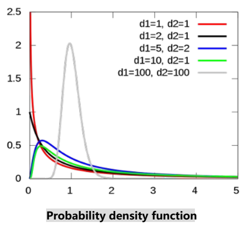


（F分布，自由度为(2,27)）

51

F比率

*F*分布



前提条件：

1.总体正态

观测值来自正态分布的总体。

2.变异的同质性

$\sigma_1^2=\sigma_2^2=\ldots=\sigma_k^2$

3.独立性

52

方差分析表

- 将结果总结到方差分析表中

表 12.6 例 12.2 的方差分析表

来源	SS	df	MS	F	η_p^2
组间	30	2	15		
组内	16	12	1.333	11.25	0.65
总和	46	14			

53

4. 事后检验（Post hoc tests）

- ANOVA 的结果是检验H_0: $\mu_1=\mu_2=\mu_3$,这是一个两点 (拒绝/不拒绝) 决策. 并未提供哪个备择假设得到支持. 也就是说, 只知道一些组与其它组不同, 但并知道差别在哪些组之间.
- 所以从ANOVA得到显著差异的结果 (拒绝H_0)后,一定要做作一些事后检验. 事后检验使我们能够比较各组, 发现差异产生在什么地方.
- 事后检验就是比较每一个处理组与另一个处理组, 一次比较两个. 这称为成对比较.

54

Tukey's HSD 检验

- 此检验要求各组有相等的样本容量.
- 可以计算出单一的值确定处理均值间的最小差异，考查此差异在统计上是否显著.
- HSD = q * sqrt（$MS_{组内}$/n）
- q 值可以从表中查出(附表5). 需要用到K和 $df_{组内}$, 以及α_{EW}

55

Scheffe检验

- 特别适用于n 不等的情况
- 这是保守的检验(降低 I类错误的风险, 但增加II类错误的风险).
- 用F比率检验差异. 重新计算两项的 $MS_{组间}$, 每次只检验一个比较.注意:用整体的 $df_{组间}$ 和整体的$MS_{组内}$.

56

方差分析与t-检验的关系

- 差异间独立样本 t-检验与两个水平的单因素组间 ANOVA有何区别?
- 没有. F比率 = t^2
- 差异间t-检验和 ANOVA, t-检验是考察两个均值间的差异，ANOVA 是考察方差. 如果只有两个组, t 统计量的平方就是F统计量

57

重复测量的ANOVA能够处理数据的类型

- 在上例中有一个自变量 (称为组内因素): 时间.
- ANOVA 亦可用于分析同时包含组间和组内因素的混合设计

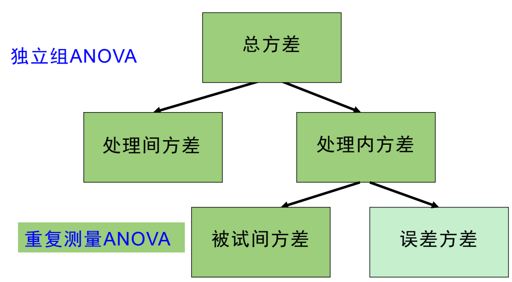
58

重复测量ANOVA的逻辑

- 和方的分解分两个阶段
  - 首先考虑方差的来源.
    - a) 处理间变异 ：处理效应 (处理造成的差异)
      - i. 处理效应
      - ii. 实验误差
  - 重复测量设计在每一种处理条件用同样的个体，所以处理间变异不包含个体差异。
    - b) 处理内变异
      - i. 被试间方差
      - ii. 实验误差

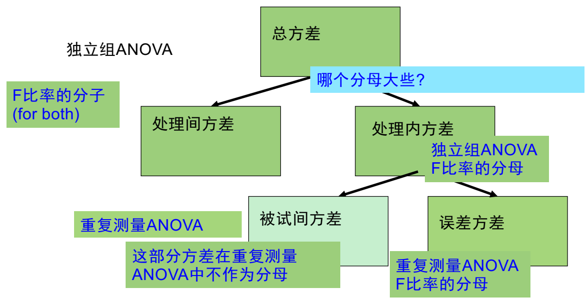
59

ANOVA间关系的比较



60

ANOVA间关系的比较



61

和方的分割

$$SS_{总和} = SS_{组间} + SS_{组内}$$

$$= SS_{组间} + SS_{被试间} + SS_{误差}$$

$$\sum (X_i - X)^2 = \sum (X_T - X)^2 + \sum (X_i - X_T)^2$$

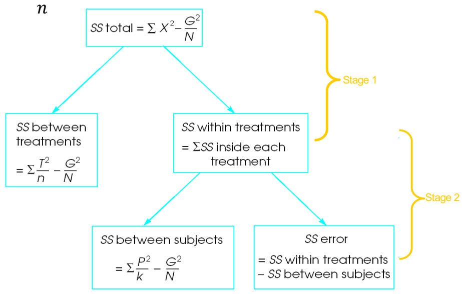
$$= \sum (X_T - X)^2 + \sum (X_P - X)^2 + Error$$

62

和方的分割

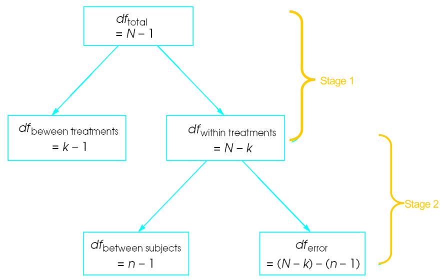
$$SS = \sum (X - X)^2$$

$$SS = \sum X^2 - \frac{(\sum X)^2}{n}$$



63

自由度的分割



64

重复测量ANOVA的逻辑

$$F = \frac{处理间变异}{实验误差} = \frac{MS_{组间}}{MS_{误差}}$$

65

方差分析表

表 12.9　例 12.3 的方差分析表

来源	SS	df	MS	F	η_p^2
组间	14	2	7	7	0.70
组内	18	9			
被试间	12	3			
误差	6	6	1		
总和	32	11			

66

Tukey's HSD 检验

- 可以计算出单一的值确定处理均值间的最小差异，考查此差异在统计上是否显著.
- HSD = q * sqrt（$MS_{组内}$/n）
- q 值 可以从表中查出(附表5). 需要用到K和$df_{组内}$, 以及α_{EW}
- 用误差的和方，自由度， 均方代替原来分母项的组内和方，自由度， 均方
- HSD = q * sqrt（$MS_{误差}$/n）
- q 值可以从表中查出(附表5). 需要用到K和$df_{误差}$, 以及α_{EW}

67

方差分析的效应大小

独立样本方差分析

$$\eta^2 = \frac{SS_{组间}}{SS_{组间} + SS_{组内}}$$

$\eta^2 = r^2$

r^2 的大小	效应的评估
$0.01 < r^2 < 0.09$	小的效应
$0.09 < r^2 < 0.25$	中等效应
$r^2 > 0.25$	大的效应

68

方差分析的效应大小

重复测量方差分析

$$F = \frac{处理间变异}{实验误差} = \frac{MS_{组间}}{MS_{误差}}$$

$$\eta^2 = \frac{SS_{组间}}{SS_{组间} + SS_{误差}}$$

$$\text{Cohen's } f \quad f = \sqrt{\frac{F_{effect} \times df_{effect}}{df_{error}}}$$

$$f = \sqrt{\frac{\eta^2}{1 - \eta^2}}$$

69

方差分析的统计效力

		ACTUAL SITUATION	
		NO EFFECT, H_0 TRUE	EFFECT EXISTS, H_0 FALSE
EXPERIMENTER'S DECISION	Reject H_0	Type I error α	De $1-\beta$ ct
	Retain H_0	Decision correct	Type II β

70

方差分析的统计效力

表 12.13 ANOVA 的效应和效力换算表

表 12.13.1 三组被试

效应大小 / 效力 / 每组人数	0.10	0.25	0.40
10	0.07	0.20	0.45
20	0.09	0.38	0.78
30	0.12	0.55	0.93
40	0.15	0.68	0.98
50	0.18	0.79	0.99
100	0.32	0.98	1.0

表 12.14 用 ANOVA 做 $\alpha=0.05$ 的假设检验达到 80%的统计效力所需的被试

效应大小 / 被试数 / 组数	0.10	0.25	0.40
3 组	322	52	21
4 组	274	45	18
5 组	240	39	16

G*Power: https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower

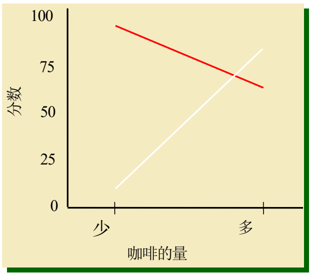
71

复杂设计的统计分析

- 2 x 2 的因素设计
  - 因素1的主效应
  - 因素2的主效应
  - 因素1和因素2的交互作用
- 上述三个检验是彼此独立的

72

交互作用和两个因素的主效应



73

和方的两阶段分解

- 和方的第一阶段分解:
  - 总和方= 处理间和方 + 处理内和方
- 和方的第二阶段分解:
  - 处理间和方
    = A的主效应+ B的主效应+AXB交互作用

74

和方的两阶段分解

- 和方的第一阶段分解:
  - 总和方= 处理间和方 + 处理内和方
- 和方的第二阶段分解:
  - 处理间和方
    = A的主效应+ B的主效应+AXB交互作用

$$\sum (X_{ij} - X)^2 = \sum (X_{ij} - X)^2 + \sum (X_{ij} - X_{ij})^2$$

$$= \sum (X_i - X)^2 + \sum (X_{.j} - X)^2 + Interaction + \sum (X_{ij} - X_{ij})^2$$

75

和方的第二阶段分解

- 和方的第一阶段分解:
  - 总和方= 处理间和方 + 处理内和方
- 和方的第二阶段分解:
  - 处理间和方
    = A的主效应+ B的主效应+AXB交互作用
- 三个F比率可以表达为:
  - F(A的主效应)= $\frac{\text{A的主效应方差}}{\text{误差方差}}$
  - F(B的主效应)= $\frac{\text{B的主效应方差}}{\text{误差方差}}$
  - F(AXB)= $\frac{\text{AXB交互作用}}{\text{误差方差}}$

76

方差分析表

- 与单因素 ANOVA 方差分析表类似

来源	df	SS	MS	F
A	a-1	SS_A	SS_A/df_A	MS_A/MS_E
B	b-1	SS_B	SS_B/df_B	MS_B/MS_E
AB	(a-1)(b-1)	SS_{AB}	SS_{AB}/df_{AB}	MS_{AB}/MS_E
误差	N- ab	SS_E	SS_E/df_E	
总和	N-1	SS_T	SS_T/df_T	

77

如何作假设检验

- 先考察交互作用
  - **如果交互作用显著, 主效应就难于解释**
- 然后考察主效应
- 一定要作图以直观表示结果

78

交互作用

- 在因素设计的方差分析中，如果一个因素对因变量的影响因另一个因素的不同水平而不同，我们就说这两个因素有交互作用。
- 一旦交互作用显著，应先解释交互作用，然后再解释主效应。

79

二因素ANOVA

- 二因素的方差分析，交互作用显著，就不需要作事后检验，而需要作简单主效应分析
- 交互作用不显著，如果主效应显著，因素有3个或3个以上水平，则需要作事后检验
- 无论交互作用是否显著，都需画图

80

相关的概念

- 相关是度量和描述两个变量之间关系的一种统计技术.
- 数据要求：一定要有至少两个变量，两组分数。

81

应用相关的研究情境

- 预测 – 如果两个变量间有强相关，我们就可以根据一个变量的值，预测另一个变量的值。
  - 如，如果知道某些人格特征，可以预测员工绩效
- 相容效度 – 如果发明新的心理测验（测验A），想知道它是否测量了X，就需要知道测验A 是否与X相关.
- 效标关联效度 – 如果发明新的量表，管理潜能量表来预测晋升所需时间，这个量表分数应当与晋升所需时间相关。
- 重测信度 – 如果对同一组被试两次用相同的测验，将两组分数做相关. 如果测验是可信的，两次测验应当得到相似的结果，产生高相关
- 理论验证 – 比如验证社交技能与焦虑的相关

82

相关表明变量 X 和 Y之间关系的3个特征

- 1) 关系的方向
  - 正相关（正数）意味着两个变量向相同的方向变化. 亦即，一个变量增加，另一个变量也增加.
  - 负相关（负数）意味着两个变量向相反的方向变化. 亦即，一个变量增加，另一个变量反而减少.
- 2) 关系的形式
  - 本课集中讨论线性(直线)相关，但两变量的关系也有其他形式
- 3) 关系的程度
  - 相关也度量了X 和 Y间关系的强度. 相关系数的值 在-1 和 +1之间. 0相关意味着没有关系. +1 意味着"完全的正相关" 之间 两个，-1 意味着完全的负相关.

83

相关的解释

- 它表明了：一个变量的方差中，由X和Y间的相关解释的方差的比例
- 当r=0.7时，Y变异的一部分能由X推出，r^2=0.49，即Y 49%的变异能够由X推出

84

积差相关的效应大小

- Cohen's Convention：
  - r=.10 小的效应
  - r=.30 中等效应
  - r=.50 大的效应

85

将相关的概念数量化

- 我们主要讨论两种相关，Pearson 积差相关,Spearman 等级相关.
  - $r = \dfrac{\text{X 和 Y 共同变化的程度}}{\text{X 和 Y各自变化的程度}} = \dfrac{\text{X和Y的协方差}}{\text{X和Y各自的方差}}$
  - 共变意味着随着X 变化，Y 也变化.
  - r = 1.0（或 -1.0）即"完全的相关".意味着分子分数等于分母分数。

86

离差的乘积和(SP)

- 定义公式：SP = $\sum(X-\overline{X})(Y-\overline{Y})$
- 对于每一点与X 和 Y的平均值的差，即离差，求两个离差的乘积，再求和
- SP的计算公式：$\sum XY - \dfrac{\sum X \sum Y}{n}$

87

乘积和(SP)与和方(SS)公式

- 非常相似，其区别是 SS只有一个变量（X），SP 有两个变量（X 和 Y）.

	和方（SS）	乘积和（SP）
定义公式	$\sum(X-\overline{X})^2$	$\sum(X-\overline{X})(Y-\overline{Y})$
计算公式	$\sum X^2 - \dfrac{(\sum X)^2}{n}$	$\sum XY - \dfrac{\sum X \sum Y}{n}$

88

Pearson 相关的计算

- 也称积差相关（product-moment correlation）
- r = $\dfrac{SP}{\sqrt{SSxSSy}}$
- 分子SP是X和Y协方差的指标.分母是X和Y各自的变异

89

相关系数的显著性检验

- 总体参数 ρ，样本统计量 r.
- 虚无假设和备择假设
  - 双侧:
    H_0: $\rho = 0$, X和Y 之间无相关
    H_1: $\rho \neq 0$
  - 单侧:
    没有正相关. H_0: $\rho \leq 0$ H_1: $\rho > 0$
    df = n - 2
- 查表或 $t = \dfrac{r}{\sqrt{\dfrac{1-r^2}{N-2}}}$

90

等级相关 (Spearman相关, 斯皮尔曼相关)

- 一种非参数检验
- 用于顺序型数据，非线性数据
- r_s =1 - 6ΣD²/n(n²-1)
  - D – 各自排序后的等级差

91

相关系数---肯德尔和谐系数

- 等级相关的一种
- 适用资料：
  - 适用于k个评价者，评价多个事物的等级变量资料
  - 多用于评分者信度分析

$$W=\frac{\sum R_i^2-\frac{(\sum R_i)^2}{N}}{\frac{1}{12}K^2(N^3-N)}$$

92

相关系数---点二列相关

- 适用资料：
  - 一列为正态等距变量
  - 另一列为二分命名变量
  - 常用于试卷的信度分析
- 公式：

$$r_{pb}=\frac{\bar{x}_p-\bar{x}_q}{S_t}\cdot\sqrt{pq}$$

93

rpb与t之间的关系

- $r^2pb=t^2/(t^2+df)$，df=N1+N2-2

94

回归方程

$$\hat{Y}=.22(X)+.68$$

95

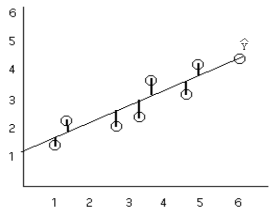
最佳拟合线

- 目标是使误差最小。即, 这条线与所有的数据点最近，是最佳拟合线.
- 回归线是给定X，a 和 b，用公式 (线性方程)来预测Y 的值。 我们的目标是找出一条线，以对Y作最佳估计. 即, 这条线使得所有Y 值的估计误差最小.

96

最小平方法（least-squares solution）

- 考察每个点, 比较Y的观测值与Y的预测值 $\hat{Y}$ 称为 "Y-hat")



距离 = $Y - \hat{Y}$

$SS_{误差}$ = 误差平方和 = 公式

$$\sum (Y - \hat{Y})^2$$

97

回归线

- 最佳拟合线斜率的公式是:
  $b = SP/SS_X = r\, S_Y/S_X$
- 最佳拟合线截距的公式是:
  $a = Y - bX$

98

回归方程估计的标准误

- 首先计算误差的平方和
  $$SS_{误差} = \sum (Y - \hat{Y})^2$$
- 将误差的平方和除以自由度 (即误差的方差, 或误差的均方)
  - $SS_{误差} / df$
  - $df = n - 2$
- 为求得估计的标准误，将误差的方差取平方根 (类似标准差)
- 最后得到公式: $S_{误差} = \sqrt{\frac{SS_{error}}{df}} = \sqrt{\frac{\sum (Y - \hat{Y})^2}{n-2}} = \sqrt{\frac{SS_{error}}{df}} = \sqrt{MS_{error}}$

99

标准误与相关系数之间的关系

- $SS_{误差} = (1 - r^2)SS_Y$

100

一元线性回归的假设检验

Source	df	SS	MS	F
Model	1	$\sum(\hat{y}_i - \bar{y})^2$	SS_{Model}/df_{Model}	MS_{Model}/MS_{Error}
Error	N - 2	$\sum(y_i - \hat{y}_i)^2$	SS_{Error}/df_{Error}	
Total	N - 1	$\sum(y_i - \bar{y})^2$		

101

一元线性回归的统计前提

1. 模型设定假定(线性预设)
2. 正交预设
3. 残差方差齐性预设
4. 正态分布预设

102

参数与非参数检验

- 参数检验
  - 用于等比/等距型数据
  - 对参数的前提：正态分布和方差同质
- 非参数检验
  - 不用对参数进行假设
  - 对分布较少有要求，也叫distribution-free tests
  - 用于类目/顺序型数据
  - 没有参数检验敏感，效力低
  - 因此在二者都可用时，总是用参数检验

103

卡方检验的基本原理

- 样本中得到的不同类别的频数称为观察频数$f_{observed}$，按期望分布计算得到的频数称为期望频数或理论频数$f_{expected}$
- $\chi^2 = \sum \frac{(f_{observed} - f_{expected})^2}{f_{expected}}$
  - χ^2反映了实际的观察频数与期望频数的偏离程度

104

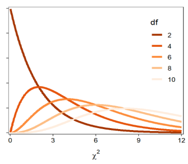
卡方检验的基本原理

- $\chi^2 = \sum \frac{(f_{observed} - f_{expected})^2}{f_{expected}}$
  - χ^2反映了实际的观察频数与期望频数的偏离程度。
    - $f_{observed} = f_{expected}$，$\chi^2 = 0$
    - 相差越小，χ^2越小
    - 相差越大，χ^2越大

105

卡方检验的基本原理

- 卡方分布
  - 由自由度(df)决定的正偏态分布。

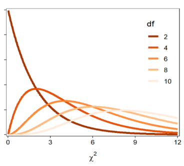


https://derekogle.com/Book107/ChiSquare.html

106

卡方检验的基本原理

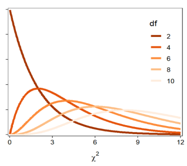
- 卡方分布
  - χ^2都是正值。
  - χ^2分布呈现正偏态，右端无限延伸，不与基线相交。



107

卡方检验的基本原理

- 卡方分布
  - 自由度越大，分布越趋向对称，极限分布为正态分布。
  - 几个χ^2 变量的和仍然服从χ^2 分布（可加性）。



108

卡方检验的基本原理

- 卡方检验：衡量模型与实际观察到的数据的比较情况。
- χ^2 取决于实际值与观测值之间的差异大小、自由度以及样本大小。
- χ^2 是在给定样本大小和分类变量的频数的情况下，比较预期结果和实际结果之间差异的大小。常用于分析分类变量的差异，尤其是那些名义变量。

109

卡方检验的应用

- Goodness of fit to a distribution:
  - χ^2 可以用来检验频率的观测分布和理论分布之间的拟合优度。
    - 实际的观察频数与期望频数间差异是否显著，从而确定观测分布和理论分布是否相符，又称为单因素χ^2检验。

110

卡方检验的应用

- Test for independence:
  - χ^2 也可以用来检验两个变量是相关的还是独立的。
    - 研究分类变量或等级变量是否相关，用于两个多项分类变量的计数资料的分析。

111

卡方匹配度检验的公式

- f_e=pn
- df =C-1
- $\chi^2 = \sum[(f_o-f_e)^2/f_e]$
  - F_o：观察次数
  - f_e：期望次数
  - C：类目的个数
  - X^2：统计量

112

卡方独立性检验

- 检验行和列的两个变量彼此有无关联
- 是命名型变量，顺序型变量相关的计算方法

113

卡方独立性检验的公式

- $\chi^2 = \sum[(f_o-f_e)^2/f_e]$
- f_e= (row total) (column total)/n,
- df = (R-1) (C-1)
  - F_o：观察次数
  - f_e：期望次数
  - R：行类目的个数　　C：列类目的个数
  - χ^2：统计量

114

χ^2与效应大小（effect size）

- Phi系数，范围0至1，是一种多元相关系数
- 在2×2列联表时，

$$\Phi = \sqrt{\frac{\chi^2}{N}}$$

- 在多于2×2列联表时，

$$Cramer's\ \Phi = \sqrt{\frac{\chi^2}{(N)(df_{Smaller})}}$$

115

Phi系数:Cohen's convension

- 当dfsmall=1时，
  - Φ=0.10表示小的效应，Φ=0.30表示中等的效应，Φ=0.50表示高的效应.
- 当dfsmall=2时，
  - Φ=0.07表示小的效应，Φ=0.21表示中等的效应，Φ=0.35表示高的效应.
- 当dfsmall=3时，
  - Φ=0.06表示小的效应，Φ=0.17表示中等的效应，Φ=0.29表示高的效应.

116

数据类型及适用的统计检验

数据类型	统计检验			
等比/等距型	独立样本 t 检验	相关样本 t 检验	方差分析	积差相关
顺序型	曼-惠特尼 U 检验 (The Mann-Whitney U Test)	维尔克松 T 检验 (The Wilcoxon T Test)	克-瓦式单向方差分析	等级相关
类目型	卡方独立性检验	符号检验法	卡方匹配度性检验	卡方独立性检验

117

曼-惠特尼U检验 (The Mann-Whitney U Test)

- 用于两个独立样本的检验
- 顺序型数据
  - 如果两个样本的差异是反映真实的总体差异，将两个样本合并后，所有的分数排序为一条线，然后来自一个样本的分数应当集中于线的一端，而来自另一个样本的分数应当集中于线的另一端
  - 如果两个样本没有差异，那么两个样本合并后大分数和小分数应当均匀地混合在一起，因为没有理由假定一组分数会大于另一组

118

Mann-Whitney U 检验

- 独立样本的非参数检验
- $U_A = n_A n_B + [n_A(n_A+1)/2] - \sum R_A$
- $U_B = n_A n_B + [n_B(n_B+1)/2] - \sum R_B$
- 选择较小的U
- $U_A + U_B = n_A * n_B$
- 如果 $U_{obs} <= U_{crit}$, 才能拒绝H_0

119

用曼-惠特尼U作假设检验

- 曼-惠特尼检验的虚无假设：两处理之间无系统差异。
- 如果Mann-Whitney U=0，其中一个样本不得分，两个样本无重叠，有最大的差异
- 当两个样本越接近时， Mann-Whitney U越大
- 所以，如果 $U_{obs} <= U_{crit}$, 才能拒绝H_0 （与参数检验正好相反）

120

符号检验

- 符号检验：
  - 比较两个相关（配对）样本的差异，数据来自顺序量表，它将中位数作为集中趋势的量度，两样本每对数据之差用正，负号表示，做单样本的二项分布检验判断正负号数是否存在显著性差异。利用正，负号的数目对某种假设做出判定的非参数检验方法。

121

维尔克松T检验 (The Wilcoxon T Test)

- 检验重复测量设计的两种处理条件之间的差别
- 对差异分数样本的绝对值进行排序
  - 虚无假设认为两处理间没有显著差异，如果虚无假设正确，样本数据之间的差异都是由机会造成。所以正的和负的差异应当均匀地混合在一起
  - 相反，如果两处理间有系统差异，会造成持续出现正的差异或持续出现负的差异

122

Wilcoxon T Test

- 相关样本的非参数检验
- 将差异分数排序，忽略正负号 (+或 -)，然后分别计算正的差异分数的秩次和 以及 负的差异分数的秩次和。
- Wilcoxon *T* 就是较小的那个和
- 如果 $T_{obs} <= T_{crit}$, 才能拒绝H_0

123

克-瓦式单向方差分析

- H=12/（N（N+1））∑R^2/n- 3（N+1）
  R：每一组数据的等级和
  自由度:k-1

124

费里德曼双向方差分析

$$\chi_r^2 = \frac{12}{nk(k+1)}\sum R^2 - 3n(k+1)$$

125

置换检验

- 置换检验(Permutation test)是Fisher于20世纪30年代提出的一种基于大量计算，利用样本数据的随机排列，进行统计推断的方法。
- 因其对总体分布自由，应用较为广泛，特别适用于总体分布未知的小样本资料，以及某些难以用常规方法分析资料的假设检验问题。

126