心理统计

第八讲：独立样本T检验与相关样本T检验

严超赣

Chao-Gan Yan, Ph.D.

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

Institute of Psychology, Chinese Academy of Sciences

1

1

独立样本T检验与相关样本T检验

1 使用背景概述

2 独立样本的T检验

3 相关样本与独立样本的区分

4 相关样本的T检验

2

2

心理学家想解决什么问题

- 心理咨询是否有疗效？
- 看暴力的影片能够影响儿童的攻击行为？
- 经过训练的小白鼠能够更快的走出迷宫？
- 高收入者和低收入者对经济危机的看法有没有不同？
- ……

3

3

一个简单实验设计

总体　　定义总体

样本　　从总体中取随机样本1 (实验组)　　从总体中取随机样本2 (控制组)

自变量　　对于实验组被试给予A处理　　对于控制组被试不给予A处理

因变量　　测量被试在任务Q 上的行为指标　　测量被试在任务Q 上的行为指标

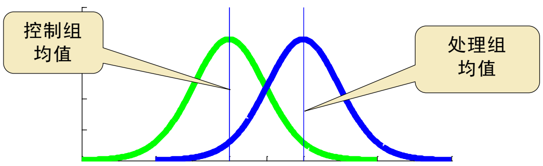
描述统计　　计算任务 Q的平均分数 $\bar{x}_E$　　计算任务 Q的平均分数 $\bar{x}_C$

推论统计　　比较 $\bar{x}_E$, $\bar{x}_C$ 用推论统计确定处理A是否影响了任务Q的分数

4

4

t-test 评价两组均值是否在统计上有差异



5

5

解决这些问题的实验设计

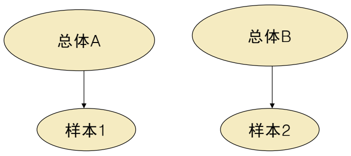
- 独立样本的研究设计（被试间的设计）---独立样本t检验
  - 教师甲在1班使用教学方法A ，在2班使用教学方法B，两种教学方法哪种更有效？
- 相关样本的研究设计---相关样本t检验
  - 重复测量(被试内设计)
    - 教师乙在1班使用教学方法A后得到一组考试成绩，使用了新的教学方法B后得到另一组考试成绩，两种教学方法哪种更有效？
  - 匹配组设计
    - 教师丙在1班使用教学方法A后得到一组考试成绩，在C班使用了新的教学方法B后得到另一组考试成绩，两班学生在性别比例、年龄、智商水平、初始成绩上都相互匹配，两种教学方法哪种更有效？

6

6

独立样本的t检验的逻辑（1）

- 已知两个样本，总体的均值和方差未知

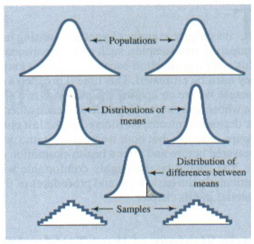


- 目标：估计总体均值的差异

7

独立样本的t检验的逻辑（2）

- 关键：弄清样本均值差异分布的来源



从两个待比较的总体中抽取样本

两个样本均值的分布

从两个样本均值分布中各取一个均值，并求两者差异，重复n次

样本均值差异的分布

8

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- $\mu_2 \neq 0$

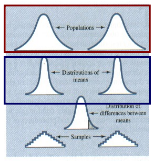
9

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

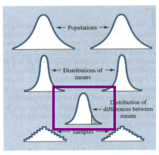
$$s_1^2 = \frac{s_p^2}{n_1} \qquad s_2^2 = \frac{s_p^2}{n_2}$$

10

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

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

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

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

11

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

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

12

差异总结

- 与单样本T检验有三点不同之处
  - 比较的分布是均值差异的分布
  - 确定t的临界值是基于两个样本的自由度
  - 比较分布的样本分数是基于两个分数之差

13

13

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.

14

14

独立样本t检验的例题

- 一位发展心理学家要考察8岁男孩和8岁女孩言语技能的差异。抽取了10个男孩和10个女孩，给予他们标准化的言语能力测验。数据如下：

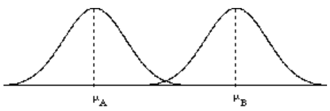
女孩	男孩
$n_1 = 10$	$n_2 = 10$
$\bar{X}_1 = 37$	$\bar{X}_2 = 31$
$SS_1 = 150$	$SS_2 = 210$

15

15

独立样本t检验的步骤（1）

- step 1：建立假设
  - 陈述 H0 和 H1；确定显著性标准: α= ?
  - 对两个不同的总体作出假设



  - 如假设有两个总体 1 (男孩) 和总体 2 (女孩)。我们想知道言语能力有无性别差异。

    H0：假设男孩和女孩的言语能力没有差异。

    H0：$\mu1 = \mu2$ 或 $\mu1 – \mu2 = 0$

    H1：假设男孩和女孩的言语能力有差异.

    H1：$\mu1 \neq \mu2$ 或 $\mu1 – \mu2 \neq 0$

16

16

独立样本t检验的步骤（2）

- step 2 确定检验的方向
  - 假设没有方向性：双尾检验
    - H_0：$\mu_1 = \mu_2$
    - H_1：$\mu_1 \neq \mu_2$
  - 假设有方向性：单尾检验
    - 如我们想知道男人的身高是否比女人高
    - H_0：$\mu_1 \leq \mu_2$
    - H_1：$\mu_1 > \mu_2$

17

17

独立样本t检验的步骤（3）

- Step 3 确定自由度
  - 从自由度的概念开始
    - 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 。
    - 我们在用两个样本,每一个样本各代表一个总体
    - 所以我们需要用总体参数的估计，因此必须考虑自由度
  - 如何计算df？
    - 样本 1: $n_1 - 1$
    - 样本 2: $n_2 - 1$
    - 所以对于两个样本 $df = n_1 + n_2 - 2$

18

18

独立样本t检验的步骤（4）

- Step 4：查表求临界值
  - 例：男孩和女孩的言语能力是否存在差异？
  - 根据三个因素确定
    - 显著性水平α
    - 自由度df
    - 检验的方向
  - 教材附表2
  - 上例中：
    - 双尾, α= 0.05, df = 18
    - tcrit =?　2.101

独立样本的t检验步骤（5）

- Step 5 进行方差同质性检验
  - 方差同质性的Hartley's F-max 检验

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

  - 上例中方差同质性检验
    - SS_1＝150.0；
    - SS_2＝210.0；
    - F-max=s_1^2/s_2^2=SS1/SS2=7/5<2
    - 根据拇指原则可认为两样本方差同质

女孩	男孩
$n_1 = 10$	$n_2 = 10$
$\bar{X}_1 = 37$	$\bar{X}_2 = 31$
$SS_1 = 150$	$SS_2 = 210$

STEP 6.1:　求均值差异

$$\bar{X}_1 - \bar{X}_2 = 6$$

女孩	男孩
$n_1 = 10$	$n_2 = 10$
$\bar{X}_1 = 37$	$\bar{X}_2 = 31$
$SS_1 = 150$	$SS_2 = 210$

STEP 6.2:　计算合并方差

$$s_p^2 = \frac{SS_1 + SS_2}{df_1 + df_2} = \frac{150 + 210}{(10-1)+(10-1)} = \frac{360}{18} = 20$$

女孩	男孩
$n_1 = 10$	$n_2 = 10$
$\bar{X}_1 = 37$	$\bar{X}_2 = 31$
$SS_1 = 150$	$SS_2 = 210$

STEP 6.3: 计算标准误

$$SE = \sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}} = \sqrt{\frac{20}{10} + \frac{20}{10}} = \sqrt{4} = 2$$

女孩	男孩
$n_1 = 10$	$n_2 = 10$
$\bar{X}_1 = 37$	$\bar{X}_2 = 31$
$SS_1 = 150$	$SS_2 = 210$

STEP 6.4: 计算t统计量

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}} = \frac{(37-31)-0}{2} = 3$$

独立样本的t检验步骤（7）

- Step7 比较样本的实际t-分数与临界t-分数
  - $t_{obs} = 3$ $t_{crit} = 2.101$
- step8 对 H0 作出结论
  - 因为观察到的 t 统计量大于临界 t统计量, 所以我们有理由拒绝H0，即男孩和女孩的言语能力有显著差异.

25

25

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

26

26

独立样本t检验的效力：在.05水平作假设检验的power

	样本容量	效应大小		
		0.20	0.50	0.80
●	（单尾）10人	.11	.29	.53
	20人	.15	.46	.80
	30人	.19	.61	.92
	40人	.22	.72	.97
	50人	.26	.80	.99
	100人	.41	.97	1
●	（双尾）10人	.07	.18	.39
	20人	.09	.33	.69
	30人	.12	.47	.86
	40人	.14	.60	.94
	50人	.17	.70	.94
	100人	.29	.94	1

27

27

研究设计中的样本容量

例：在.05水平作假设检验power为80%所需的样本容量

	效应大小		
	0.20	0.50	0.80
● 单尾	310	50	20
● 双尾	393	64	26

28

28

练习1

- 研究者想了解疲劳对于心理警觉性的影响.
  第1组24小时剥夺睡眠,
  第2组正常睡眠.
- 然后测试其看到屏幕上亮点的成绩。
  两组数据是:
  $n_1 = 5$, $n_2 = 10$, $X_1 = 35$, $X_2 = 24$,
  $SS_1 = 120$, $SS_2 = 270$

29

29

独立样本t检验结果的报告

- 剥夺睡眠组看到亮点的反应时（M = 35.0, SD = 5.48）高于正常睡眠组的反应时（M = 24.0, SD = 5.48）。两组的差异显著，t（13）=-3.67， p < .05 （双尾检验）。

30

30

相关样本与独立样本的区别

- 例题：为比较两家超市的物价，小张到超市A随机买了100件物品，又到超市B随机买了100件物品。小李开列了一张30件物品的清单，到超市A， B分别买这30件物品。

（1）小张，小李的数据分别使用了什么样的设计？
需要采用什么统计分析？
（2）小张，小李的比较谁能得到更可靠的结果？

31

31

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

32

32

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

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

33

33

相关样本t的计算

- 单样本t检验

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

$$\Uparrow$$

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

$$\Uparrow$$

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

$$\Uparrow$$

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

- 相关样本t检验

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

$$\Uparrow$$

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

$$\Uparrow$$

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

$$\Uparrow$$

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

34

34

例1: 9名被试治疗后自我接受分数有无显著增加?

	自我接受分数	
被试	现在 *X1*	治疗前 *X2*
1	69	67
2	73	63
3	74	67
4	70	64
5	70	61
6	75	66
7	73	60
8	68	63
9	69	63

35

35

相关样本t检验的步骤（1）

- Step 1 ：陈述假设；确定显著性水平

H_0 ：治疗后自我接受分数无显著增加，μ_D<=0

H_1 ：治疗后自我接受分数有显著增加，μ_D>0

α=0.05

36

36

相关样本t检验的步骤（2）

- Step2: 确定检验是双尾还是单尾
  单尾检验
- Step3：确定检验的自由度
  $\boxed{\begin{aligned} df &= n - 1 \\ &= 9 - 1 \\ &= 8 \end{aligned}}$
- Step4：查表求临界t分数
  $t_{crit} = 1.86$

37

37

相关样本的t检验步骤（3）

- Step 5：计算样本的实际t分数
  - 要领：求D → $(\Sigma D)^2$
  - 求D^2 → ΣD^2

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

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

38

38

	自我接受分数			
被试	现在X_1	治疗前X_2	$D = X_1 - X_2$	D^2
1	69	67	2	4
2	73	63	10	100
3	74	67	7	49
4	70	64	6	36
5	70	61	9	81
6	75	66	9	81
7	73	60	13	169
8	68	63	5	25
9	69	63	6	36
Σ			Σ D=67	Σ D^2= 581
平均值	71.22	63.78	D =7.44	

39

39

相关样本的T检验步骤（4）

- 样本方差

$$s^2 = \frac{SS_D}{df} = \frac{\sum D^2 - \frac{(\sum D)^2}{n}}{df} = \frac{581 - \frac{67^2}{9}}{8} = 10.28$$

- 标准误

$$s_{\bar{D}} = \sqrt{\frac{s^2}{n}} = \sqrt{\frac{10.28}{9}} = 1.07$$

- 实际t分数

$$t = \frac{\bar{D} - \mu_D}{s_{\bar{D}}} = \frac{7.44 - 0}{1.07} = 6.95$$

40

40

相关样本的T检验步骤（5）

- Step6: 比较样本的实际t分数与临界t分数
  - $t_{obs} = 6.95 > t_{crit} = 1.86$
- Step7：对 H0 作出结论
  - 因为观察到的 t 统计量大于临界t统计量，所以我们有理由拒绝H_0。
  - 得出结论：治疗后自我接受分数有显著增加

41

41

相关样本t检验的统计前提

- 在每一种处理条件内，观察都彼此独立
- 差异分数的总体分布是正态的
- 不需要考虑方差同质性

42

42

相关样本t检验的效应大小

- 效应大小＝差异样本均值/差异样本方差
- ES＝ $\dfrac{\overline{D}}{S_D}$

43

43

相关样本t检验的效力：在.05水平作假设检验的power

	样本容量	效应大小		
		0.20	0.50	0.80
●	（双尾）10人	.09	.32	.66
	20人	.14	.59	.93
	30人	.19	.77	.99
	40人	.24	.88	1
	50人	.29	.94	1
	100人	.55	1	1
●	（单尾）10人	.15	.46	.78
	20人	.22	.71	.96
	30人	.29	.86	1
	40人	.35	.93	1
	50人	.40	.97	1
	100人	.63	1	1

44

44

研究设计中的样本容量

- 在.05水平作相关样本t 检验power为80%所需的样本容量

	效应大小		
	0.20	0.50	0.80
双尾	196	33	14
单尾	156	26	12

45

45

相关样本t检验结果的报告

- 广告播出前被试对产品的评定 (M = 11.71，SD =4.86) 与广告播出后被试对产品的评定 (M = 14.86，SD = 2.83) 没有显著差异，t（6）=-2.26， p > .05 （双尾检验）。

46

46

相关样本t检验的练习

- 在 Everitt (1994)的研究中，17个患神经性厌食症的女孩治疗前后分别称了体重。计算了每个被试分数的差异。

体重的差异

n = 17
$\overline{D}$ = 7.26
s_D = 7.16

治疗后体重有无显著增加？.计算效应大小

47

47

体重的差异

n = 17
$\overline{D}$ = 7.26
s_D = 7.16

$$SE = \frac{7.16}{\sqrt{17}} = 1.74$$

$$t(16) = \frac{7.26 - 0}{1.74} = 4.17$$

$$t_{crit} = 2.12$$

$$d = \frac{7.26}{7.16}$$

$$d = 1.01$$

48

48

独立样本t检验与相关样本t检验的比较

- 例题：启动效应是认知心理学关于无意识的重要内容，研究者发现，不仅文字具有启动效应，图像也可产生语义启动。为了对语义启动量进行比较，以为研究者用了等量且具有相同意义的图片和词语作为启动材料，要求被试来完成语义归类的任务。结果如下：
- 结果 1 2 3 4 5 6 7 8 9 10 11 12
- 图片 50 42 36 55 58 40 49 53 58 38 25 48
- 词语 54 48 35 58 62 45 48 51 56 43 32 55

- 测量对象：对12名被试重复测量
- 测量对象：2组独立的被试，每组12人

49

49

独立样本t检验和相关样本t检验的比较

- 相关样本t检验的结果：
  - tobs=2.91>tcrit,拒绝虚无假设，认为图片与词语启动量有显著差异
- 独立样本t检验的结果：
  - tobs=0.75<tcrit,接受虚无假设，认为图片与词语启动量无显著差异

50

50

是什么导致结果的不同

$$\sigma^2_{(x-y)} = \sigma^2_{(x)} - 2r\sigma_{(x)}\sigma_{(y)} + \sigma^2_{(y)}$$

- 对于独立样本而言，r=0，中间项等于0；
- 对于相关样本而言，|r|>0,中间项大于0；

- 注意：并非所有的研究都是相关设计好

51

51

	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	$\bar{X}_1-\bar{X}_2$	$\mu_1-\mu_2$	$s_p^2=\frac{SS_1+SS_2}{df_1+df_2}$	$\sqrt{\frac{s_p^2}{n_1}+\frac{s_p^2}{n_2}}$	

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

52

作业

1. 一位研究者对长子与次子的心理特征感兴趣。他在一年级大学生中随机抽取了10个长子和20个非长子对其施测自尊量表。10个长子在量表上的平均分是X = 48， SS=670。 20个非长子的平均分是 X = 41，SS=1010。这些数据表明两组间是否有显著差异？用α= .01 的显著性水平作假设检验。

53

53

作业

2. 在认知失调理论的经典实验中，Festinger （1959）和他的同事让40名大学生被试参加一个非常枯燥乏味的实验。完成实验后指示这些被试对其他人说这是一个有趣的实验，劝其参加。将这些被试随机分成两组。其中一组的20人每人给1美元的报酬（低报酬组），另一组每人给20美元的报酬（高报酬组）。之后，让每个学生评定实验的有趣程度（高分表示认为实验比较有趣）。下面是一组虚构的数据：

低报酬组				高报酬组			
3	3	4	6	1	2	5	2
5	5	5	7	3	5	4	5
8	5	4	8	2	3	4	4
2	6	4	4	1	2	3	3
6	7	5	5	5	1	1	3

认知失调理论预测低报酬组比高报酬组更容易以为实验真的有趣。因为这样比较容易让他们认知协调。那些得到足够报酬的被试则不需要改变态度，因此其观点更容易反映真实的情况。以上数据有没有支持这个预测？（用α= .01 的显著性水平）

54

54

作业

3. 以下数据给出了两个职业组的样本在Catttel （1973）16因素人格量表中的轻松-紧张维度上的得分（分数越低，表明越轻松）。

作家	7	7	6	9	8	8	7	9	5	3	6	8	7	9
飞行员	4	2	2	3	1	5	4	3	2	2	6	2	5	3

以上数据显示这两个职业组在轻松-紧张维度上的得分有显著差异吗？（用α= .05 的显著性水平）

55

55

作业

4. 为了确定日常的体育锻炼多大的运动量适宜，一位研究者用了7组被试，将其年龄，性别，体重，健康状况等有关变量加以匹配。其中一组被试每星期锻炼2小时，另一组被试每星期锻炼5小时，一段时间后让医生评定其健康状况，得到以下数据。这些数据是否说明运动量对健康有影响？

被试对	锻炼2小时组	锻炼5小时组
A	15	18
B	12	14
C	16	12
D	9	11
E	13	14
F	16	16
G	17	16

56

56

作业

5. 下表是感觉剥夺1小时前后测得7名被试的听阈。感觉剥夺实验是否对被试的听阈有显著影响？

被试	前测	后测
A	31	30
B	34	31
C	29	29
D	33	29
E	35	32
F	32	34
G	35	28

57

57

作业

6. 以下是比较两种处理条件下作业成绩的数据：

处理1	处理2	差距
10	11	1
2	5	3
1	2	1
15	18	3
7	9	2
X = 7	**X** = 9	**X** = 2
SS=134	SS=150	SS=4

a) 假设以上数据来自独立指标设计的实验，用 t 检验来确定两种处理条件下作业成绩有无显著差异（用Alpha = .05）

b) 假设以上数据来自重复指标设计的实验，用 t 检验来确定两种处理条件下作业成绩有无显著差异（用Alpha = .05）

c) 如何解释a）b）二者的不同结论

58

58