第12章 连续变量的统计推断（一）
—— t 检验

学习目标

- 了解 t 检验中涉及的几个基本原理。
- 掌握单样本 t 检验、两独立样本 t 检验、配对样本 t 检验的方法，并能读懂分析结果。

主要内容

- **12.1 *t* 检验概述**
- **12.2 样本均数与总体均数的比较**
  - 单样本t检验——检验单个样本来自的总体的均数
- **12.3 成组设计两样本均数的比较**
  - 独立样本t检验——成组设计的两个样本，对两个总体均数进行比较
- **12.4 正态性、方差齐性的考察与应对策略**
- **12.5 配对设计样本均数的比较**
  - 配对样本t检验——配对设计的两个样本，对两个总体均数进行比较

12.1 t 检验概述

t 分布的发明者

- W. S. Gosset，戈赛特，英国数学家，1876-1937，在1908年以笔名“Student”发表了一篇关于 t 分布的论文，开创了利用小样本计量资料进行统计推断的先河。
- t 分布的概率密度曲线看上去有些像标准正态分布，但是高峰矮一些，而且尾巴长一些。

几个结论

- 1. 当总体服从正态分布N(μ, σ^2)时，从该总体中抽取样本容量为n的样本，通过概率论中的中心极限定理可知样本均值服从正态分布：

$$\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$$

进一步将其标准化，可得：

$$\frac{\overline{X}-\mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

注：一般来讲，由于总体的σ^2未知，该结论在SPSS中用不到。

几个结论

- **2.** 总体不服从正态分布**(**不能是强烈的偏态分布，总体均值**μ,** 总体方差**σ²)**，但当样本容量比较大时（大于**50**），通过概率论中的中心极限定理可知样本均值近似服从正态分布：

$$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$$

进一步将其标准化，可得：

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

注：一般来讲，由于总体的**σ²**未知，该结论在**SPSS**中用不到。

几个结论

- **3. 当总体服从正态分布时（总体均值为μ），从该总体中抽取样本容量为n的样本，样本标准差为s，则样本均值服从t分布：**

$$\frac{\overline{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$$

注：这是单样本t检验的理论基础。

几个结论

- **4.** 当总体不服从正态分布时（不能是强烈的偏态分布，总体均值为μ），从该总体中抽取样本容量为n的样本(n>50)，样本标准差为s，则样本均值服从t分布：

$$\frac{\overline{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$$

注：这是单样本t检验的理论基础。

几个结论

5. 设 $X_1, X_2, \cdots, X_n$ 与 $Y_1, Y_2, \cdots, Y_m$ 分别是来自正态总体 $X \sim N(\mu_1, \boxed{\sigma^2})$ 与 $Y \sim N(\mu_2, \boxed{\sigma^2})$ 的相互独立的简单随机样本.

则 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim N(\mu_1, \frac{\sigma^2}{n})$ $\bar{Y} = \frac{1}{m}\sum_{j=1}^{m} Y_j \sim N(\mu_2, \frac{\sigma^2}{m})$

$$\Longrightarrow \bar{X} - \bar{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{m})$$

$$\Longrightarrow \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} \sim N(0,1)$$

几个结论

$$\frac{(n-1)S_1^2}{\sigma^2} \sim \chi^2(n-1), \quad \frac{(m-1)S_2^2}{\sigma^2} \sim \chi^2(m-1)$$

$$\Rightarrow \frac{(n-1)S_1^2}{\sigma^2} + \frac{(m-1)S_2^2}{\sigma^2} \sim \chi^2(n+m-2)$$

又 $\overline{X}-\overline{Y}$ 与 $\frac{(n-1)S_1^2}{\sigma^2} + \frac{(m-1)S_2^2}{\sigma^2}$ 相互独立

几个结论

$$\frac{\dfrac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{\sigma^2}{n}+\dfrac{\sigma^2}{m}}}}{\sqrt{\dfrac{\dfrac{(n-1)S_1^2}{\sigma^2}+\dfrac{(m-1)S_2^2}{\sigma^2}}{n+m-2}}}$$

$$=\boxed{\frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\dfrac{1}{n}+\dfrac{1}{m}}\sqrt{\dfrac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}}}\sim t(n+m-2)}$$

注：这是两独立样本t检验的理论基础。

几个结论

6. 设 $X_1, X_2, \cdots, X_n$ 与 $Y_1, Y_2, \cdots, Y_m$ 分别是来自正态总体 $X \sim N(\mu_1, \boxed{\sigma_1^2})$ 与 $Y \sim N(\mu_2, \boxed{\sigma_2^2})$ 的相互独立的简单随机样本.

令 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ $\qquad$ $\overline{Y} = \frac{1}{m}\sum_{j=1}^{m} Y_j$

$$S_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad S_2^2 = \frac{1}{m-1}\sum_{j=1}^{m}(Y_j - \overline{Y})^2$$

几个结论

则 $\dfrac{(n-1)S_1^2}{\sigma_1^2} \sim \chi^2(n-1)$, $\dfrac{(m-1)S_2^2}{\sigma_2^2} \sim \chi^2(m-1)$

$$\boxed{\dfrac{S_1^2 / \sigma_1^2}{S_2^2 / \sigma_2^2} \sim F(n-1, m-1)}$$

若 $\sigma_1 = \sigma_2$ 则 $\dfrac{S_1^2}{S_2^2} \sim F(n-1, m-1)$

注：这是两独立样本t检验中方差齐性检验的理论基础。

12.2 样本均数与总体均数的比较

- 单个样本均数检验问题是一种关于总体均数的假设检验问题，这种问题中只有一个样本，研究目的是推断这个样本对应的总体的均数是否等于（或大于、或小于）某个已知的数值。

单样本均值 t 检验步骤

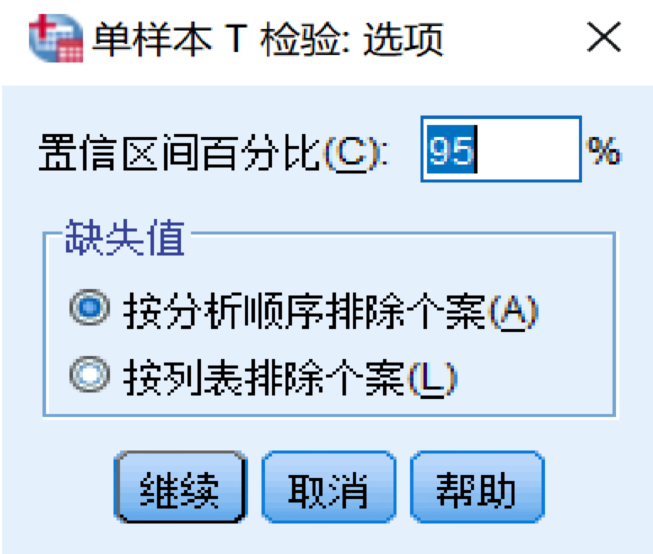
- 写出假设。H_0：$\mu = \mu_0$；H_1：$\mu \neq \mu_0$
- 选定统计量。$$\frac{\overline{X}-\mu}{s/\sqrt{n}} \sim t(n-1)$$
- 确定显著性水平（检验水准）α。
- 计算样本事件的概率P。
- 根据P与的α的大小比较判断是拒绝原假设还是不拒绝原假设。

案例

- 数据文件：CCSS_Sample.sav
- 要求：检验2007年4月一线城市的总信心指数均值是否为100
- 具体过程：
  - 选择个案
  - “分析”→“比较平均值”→“单样本 *t* 检验”
  - 将“index1”选入“检验变量”框
  - “检验值”设置为100

案例


单样本 T 检验
Qa3
Qa4
Qa8
Qa9
Qa10
Qa16
现状指数 [index1a]
预期指数 [index1b]
time = 200704 (Fl...
检验变量(T):
总指数 [index1]
选项(O)...
Bootstrap...
检验值(V): 100
确定
粘贴(P)
重置(R)
取消
帮助


单样本 T 检验: 选项
置信区间百分比(C): 95 %
缺失值
按分析顺序排除个案(A)
按列表排除个案(L)
继续
取消
帮助

样本标准差$S=\sqrt{\sum_{i=1}^{n}\frac{(X_i-\overline{X})^2}{n-1}}$

样本均数标准误差等于$S/\sqrt{n}$

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
总指数	300	98.3363	18.92074	1.09239

One-Sample Test						
	Test Value = 100					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
总指数	-1.523	299	.129	-1.66367	-3.8134	.4861

统计量$t=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$

自由度

双尾P值

$\overline{X}-\mu_0$

$\mu-\mu_0$的95%的置信区间

从分析结果可知，P值大于α（0.05），所以不拒绝H_0，即认为2007年4月一线城市的总信心指数仍为100！

单样本 t 检验中的其他问题

- 一、总体均数置信区间与 t 检验的一致性
  - 从置信区间角度分析出假设检验的结论。通过样本数据可以得到总体均数的95%的置信区间（96.1864, 100.4861），而假设的总体均数为100，落在上面的区间内，从而认为样本来自的总体的均数与假设的总体均数在显著性水平 $\alpha = 0.05$ 条件下无显著差异。
  - 从 t 检验角度看结论。P = 0.129 > 0.05，故不拒绝H_0（$\mu = \mu_0$），认为2007年4月一线城市总信心指数在显著性水平 $\alpha = 0.05$ 条件下与100无显著差异。

单样本 t 检验中的其他问题

- 二、方法的适用条件
  - 1、若样本量足够大（超过50），任何分布的总体均适用，但要求分布不是强烈的偏态。
  - 2、若样本量较小，正态分布的总体才适用。

因为只有在这样的条件下才会出现 t 分布！

12.3 成组设计两样本均数的比较

推断两个样本是否来自相同的总体，更具体地说，是要检验两样本所来自的总体的均数是否有显著性差异。

说明

- 两独立样本 t 检验在推导过程中除了要求总体服从正态分布外，还要求两样本各自所在总体的方差无显著性差异，即方差齐性（两个正态总体方差的检验）。

两独立样本均值 *t* 检验步骤

- 写出假设。**H_0: $\mu_1 = \mu_2$; H_1: $\mu_1 \neq \mu_2$**
- 选定统计量。

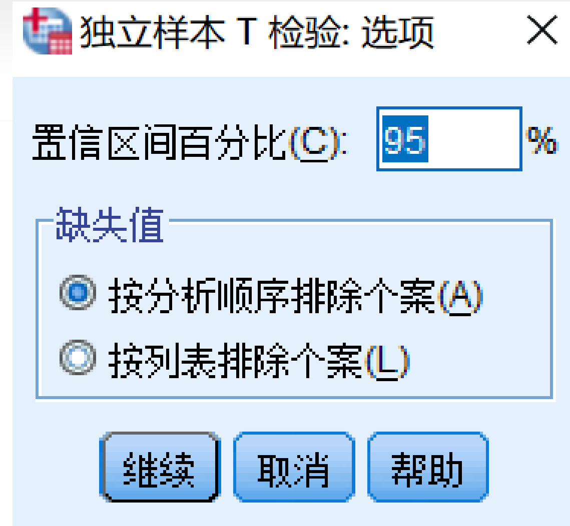
$$\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{1}{n}+\frac{1}{m}}\sqrt{\frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}}} \sim t(n+m-2)$$

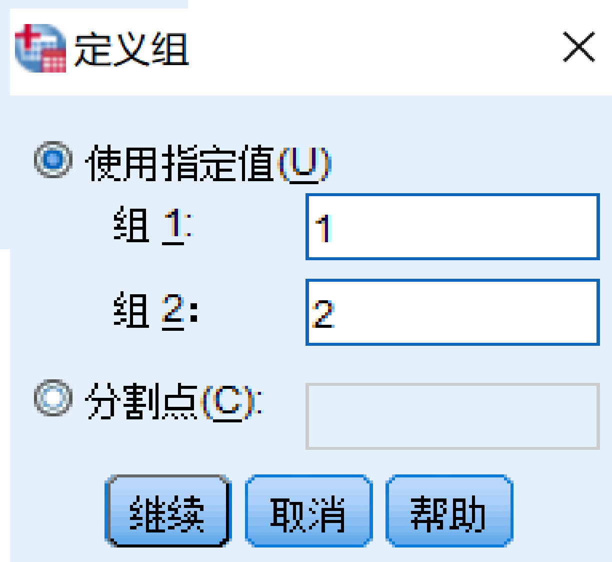
- 确定显著性水平**α**。
- 方差齐性检验
- 计算样本事件的概率**P**。
- 根据**P**与的**α**的大小比较判断是拒绝原假设还是不拒绝原假设。

案例

- 数据文件：CCSS_Sample.sav
- 要求：检验2007年4月家庭年收入较高的家庭（年收入大于4.8万元）和家庭年收入较低的家庭总信心指数是否存在显著差异
- 具体过程：
  - 选择个案
  - “分析”→“比较平均值”→“独立样本t检验”
  - 将“index1”选入“检验变量”框
  - 将“Ts9”选入“分组变量”，点击“定义组”按钮，“组1”框中填入1，“组2”框中填入2


独立样本 T 检验
A10. 那么与现在...
A16. 对于大宗耐...
Qs9
Qa3
Qa4
Qa8
Qa9
Qa10
Qa16
现状指数 [index1a]
检验变量(T):
总指数 [index1]
分组变量(G):
Ts9(? ?)
定义组(D)...
选项(O)...
Bootstrap...
确定
粘贴(P)
重置(R)
取消
帮助


独立样本 T 检验: 选项
置信区间百分比(C): 95 %
缺失值
按分析顺序排除个案(A)
按列表排除个案(L)
继续
取消
帮助


定义组
使用指定值(U)
组 1: 1
组 2: 2
分割点(C):
继续
取消
帮助

案例

Group Statistics					
	家庭收入2级	N	Mean	Std. Deviation	Std. Error Mean
总指数	Below 48,000	110	90.7458	21.23893	2.02505
	Over 48,000	145	104.4475	14.92637	1.23957

P<0.05, 因此方差不齐性

Independent Samples Test					
		Levene's Test for Equality of Variances		t-test for Equality of Means	
		F	Sig.	t	df
总指数	Equal variances assumed	11.930	.001	-6.047	253
	Equal variances not assumed			-5.771	186.197

方差齐性检验

方差齐性时的t检验

方差非齐性时的t检验

P<0.05, 因此拒绝原假设，即认为有显著差异

Independent Samples Test					
		t-test for Equality of Means			
		Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
					Lower
总指数	Equal variances assumed	.000	-13.70173	2.26593	-18.16421
	Equal variances not assumed	.000	-13.70173	2.37431	-18.38574

方差齐性时的t检验

方差非齐性时的t检验

案例

Independent Samples Test		
		t-test for Equality of Means
		95% Confidence Interval of the Difference
		Upper
总指数	Equal variances assumed	-9.23924
	Equal variances not assumed	-9.01771

适用条件

- 独立性：各观察值之间是相互独立的。对结果的影响较大，但一般没问题。
- 正态性：各个样本均来自正态总体。有一定的耐受能力，可以通过直方图等进行观察，偏的不厉害就行。
- 方差齐性：各个样本所在总体的方差相等。相对而言对结论的影响较大，需要进行方差齐性检验。方差齐性和非齐性时有不同的t检验。

12.4 正态性、方差齐性的考察与应对策略

正态性的考察方法

- 图形化考察方法
  - 直方图、箱图、茎叶图、P-P图、Q-Q图
- 偏态系数和峰态系数
  - “探索”过程中可计算这两个指标
- 专用的正态性假设检验方法
  - “探索”过程中，点击“绘图”按钮，勾选“带检验的正态图”，则在分析时，会绘制Q-Q图，并提供数据正态性的两种假设检验结果：
    - K-S检验：适用于样本量较大的情形，SPSS建议样本量大于5000时使用
    - S-W检验：适用于小样本或非整数加权样本

方差齐性的考察方法

- 两独立样本 *t* 检验过程中会自动进行方差齐性检验
- **SPSS默认采用Levene's方法**
  - 该方法对正态性假设是稳健的

数据不符合适用条件时的应对策略

- 样本量均衡性
- 变量变换：对数转换、平方根转换、平方根反正弦转换、平方变换、倒数变换、**Box-Cox**变换
- 校正检验

12.5 配对设计样本均数的比较

将受试对象按情况相近者配对（或者自身进行配对），分别给予两种处理，以观察两种处理效果有无显著性差别。

说明

- 在配对设计得到的样本数据中，每对数据之间都有一定的相关，如果采用成组的 t 检验就无法利用这种关系，浪费了大量统计信息。
- 对于这种情况，统计学上的解决办法是求出每对的差值，通过检验该差值总体均数是否为 0，就可以得知两种处理有无差异。可以看出，本质还是单样本 t 检验。
- 虽然功能实际上和单样本 t 检验重复，但数据输入格式不同。

配对样本均值 t 检验步骤

- 写出假设。H_0：$\mu_d = 0$；H_1：$\mu_d \neq 0$
- 选定统计量。

$$\frac{\bar{d}-\mu_d}{s/\sqrt{n}} \sim t(n-1)$$

- 确定显著性水平α。
- 计算样本事件的概率P。
- 根据P与的α的大小比较判断是拒绝原假设还是接受原假设。

案例

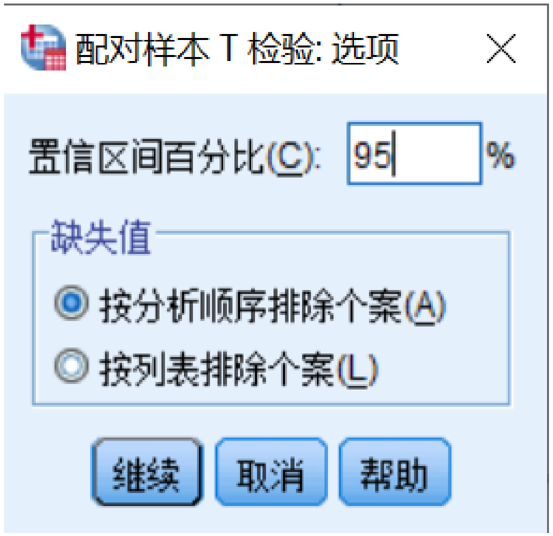
- 数据文件：CCSS_pair.sav。2007年4月的信心指数为index1、index1a和index1b，2007年12月对应的信心指数为index1n、index1an和index1bn。
- 要求：检验2007年12月的信心指数相对于2007年4月是否发生了显著的变化。
  – 这是一个配对设计，应当采用配对设计差值的 *t* 检验来进行分析。
  – 按照配对 *t* 检验对数据格式的要求，这里的数据应当每个变量（一列）代表一个组，而每条记录（一行）代表一对数据。

案例

- 具体过程：
  - “分析”→“比较平均值”→“配对样本t检验”
  - 将“index1”和“index1n”成对选入“成对变量”框
  - 将“index1a”和“index1an”成对选入“成对变量”框
  - 将“index1b”和“index1bn”成对选入“成对变量”框

案例

配对样本 T 检验
成对变量(V):
月份 [time]
id
s2.性别 [s2]
s3.年龄 [s3]
Qa3
Qa4
Qa8
Qa10
Qa16
index1
index1a
对
Variable1
Variable2
1 [index1] [index1n]
2 [index1a] [index1...
3 [index1b] [index1...
4
选项(O)...
Bootstrap...
确定
粘贴(P)
重置(R)
取消
帮助



配对样本 T 检验: 选项
置信区间百分比(C): 95 %
缺失值
按分析顺序排除个案(A)
按列表排除个案(L)
继续
取消
帮助

案例

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	index1	98.7872	88	16.15275	1.72189
	index1n	96.0357	88	20.10295	2.14298
Pair 2	index1a	94.5386	88	22.42476	2.39049
	index1an	98.5402	88	26.69120	2.84529
Pair 3	index1b	101.1220	88	19.60309	2.08970
	index1bn	94.6557	88	22.14602	2.36077

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	index1 & index1n	88	.264	.013
Pair 2	index1a & index1an	88	.182	.089
Pair 3	index1b & index1bn	88	.305	.004

相关性分析

H_0：一次项系数为0

案例

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	index1 - index1n	2.75149	22.21977	2.36864	-1.95643	7.45940	1.162	87	.249
Pair 2	index1a - index1an	-4.00163	31.57745	3.36617	-10.69225	2.68899	-1.189	87	.238
Pair 3	index1b - index1bn	6.46630	24.69302	2.63229	1.23435	11.69826	2.457	87	.016

- 结论
  - 对于预期指数index1b， P = 0.016 < 0.05, 故可以认为预期指数发生了显著性变化，又因为差值均数为正，因此可以认为2007年12月预期指数下降了。
  - 对于总指数index1，P = 0.249 > 0.05，故可以认为总指数并没有发生显著性变化。
  - 对于现状指数index1a，P = 0.238 > 0.05，故可以认为现状指数并没有发生显著性变化。

THE END