社会统计学及SPSS软件应用
STATISTICS WITH SPSS

Instructor:王荣欣

Email: rxwang@qq.com

周二3-4节、单周周四3-4节，3A106-2

2020年10月5日

CONTENTS

THE BASIC OF HYPOTHESIS TESTING

- Null Hypothesis Significance Testing （显著度检验）
  - 零假设（null hypothesis），null隐含了nullify（放弃），提出零假设的目的就是为了放弃它。
  - 零假设的内容通常是：在总体中，两个变量之间的关系是0或与0没有值得关注的（显著的）差异。
- Neyman-Pearson Lemma

Null Hypothesis Significance Testing(1)

- The goal is to find a decision rule about whether the null hypothesis should be **ruled out** based on the available data.
- When the null hypothesis is rejected, this means you decide that the corresponding alternative hypothesis is accepted.

NULL HYPOTHESIS SIGNIFICANCE TESTING(2)

- 一组平均数与所有可能平均数的比较，抽到这一组在所有可能性中的概率。
- 计算test statistic（例如t值），看test statistic所对应的概率。这个概率是以零假设为前提推出来的。
- Statistical significance: **if P-value is small enough**, then we say the results are statistically significant.

NULL HYPOTHESIS SIGNIFICANCE TESTING(3)

- 假设检验是在假设“待检验的假设为真”的前提下，计算观测结果发生的概率。
- 当观测结果发生的概率很低时，可以得出原假设不成立的结论。
- 对某个待检验假设，统计分析用significant（显著的）这个词来表示结果发生的概率很低。

1. 在假定H_0成立的条件下，计算某个统计量的值，并确定它的概率分布。
2. 计算由样本得到的统计量的值所发生的概率，又称之为显著性水平（significance level），一般用α表示。
3. 若统计量的值所发生的概率低于我们事先设定的概率标准（如0.10、0.05和0.01），就说明统计显著，于是倾向于拒绝或否定原假设。（通过统计的显著性检验，推翻原假设）

DECISION RULE

- A natural strategy is to try to determine how small the sample mean must be in order to reject the hypothesis that μ is greater than or equal to 24.
- But rather than work with $\bar{y}$, it is more convenient to work with

$$Z = \frac{\bar{y} - 24}{\frac{\sigma_Y}{\sqrt{n}}}$$

- The issue of whether the sample mean is sufficiently small to reject the null hypothesis is the same as whether Z is sufficiently small.（距离$|Z|$越大，越不可能）

1 z test：Z检验法，又称正态分布检验
   - 为什么是Z分布？
     如果样本均值组成的sampling distribution（抽样分布）是正态分布，那么将该正态分布**标准化**后可得到Z分布。
2 t test：若不知道总体标准差，用样本标准差s代替总体方差，此时样本均值的sampling distribution是t分布，即t检验。
   - 当自由度超过20时，t值（T-value）的分布近似正态分布，平均值是0，标准差是1。自由度越大，t值的分布就越接近正态分布。

z值的计算假定总体平均值与标准差是已知的，t值假定样本的标准差已知。

否定域的含义是什么？

否定域与显著性水平（1）

- 否定域在整个抽样分布中所占的比例，叫做显著性水平α（或显著度α），代表样本的统计值落在否定域的可能性。
  - 显著度（level of significance）是指我们承担了多大的犯一类错误的风险。
- The p-value is just the smallest α value for which the null hypothesis is rejected.
- Alternatively, the p-value is the probability of a Type I error if the observed value of Z is used as a critical value.

否定域与显著性水平（1）

- 否定域在整个抽样分布中所占的比例，叫做显著性水平α（或显著度α），代表样本的统计值落在否定域的可能性。
  - 显著度（level of significance）是指我们承担了多大的犯一类错误的风险。
- The p-value is just the smallest α value for which the null hypothesis is rejected.
- Alternatively, the p-value is the probability of a Type I error if the observed value of Z is used as a critical value.

假设检验的两类错误

- Type I error or α error （弃真）：零假设H_o实际上是正确的，却被否定了。
- Type II error or β error（纳伪）：零假设H_o实际上是错误的，却没有被否定。

假设检验的两类错误

- Type I error or α error （弃真）：零假设H_o实际上是正确的，却被否定了。
- Type II error or β error（纳伪）：零假设H_o实际上是错误的，却没有被否定。

假设检验的两类错误（2）

1. A Type I error is made when we reject a null hypothesis when it is true, and Type II error is made when we do not reject a null hypothesis when in fact the alternative hypothesis holds.
2. A Type I legal error is to falsely convict an innocent person（冤枉好人）, and a Type II error is find someone ”not guilty” when in fact they did commit the crime（放过坏人）.

这两种说法是否一致？

1. “在95%水平上显著”
2. “冒了5%犯一类错误的风险”

否定域与显著性水平（2）

- A critical value is the value used to determine whether the null hypothesis should be rejected.
- If you want the probability of a Type I error to be α=0.025, the critical value is -1.96.
- If α=0.005, then the critical value is -2.58. [1]

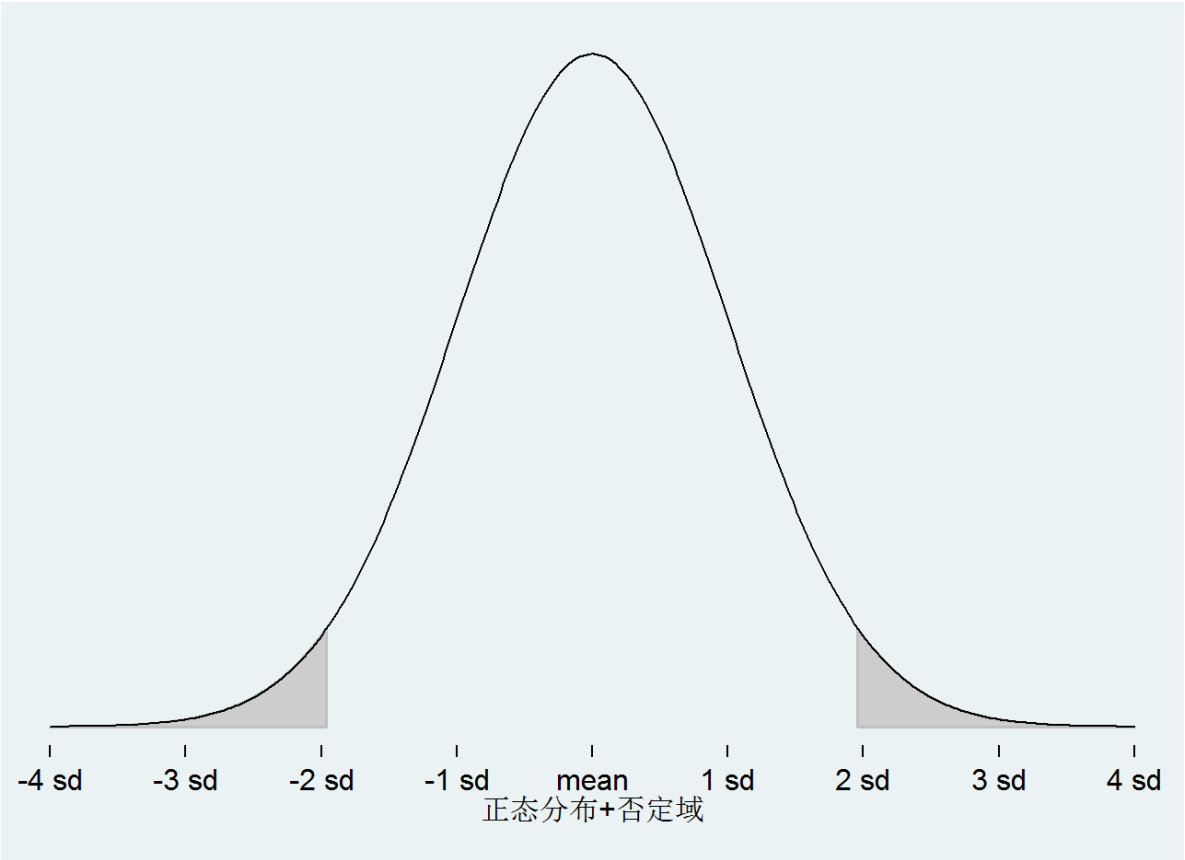

Figure 2.1: 否定域

否定域与显著性水平（3）

1. Decision rule 1: Reject the null hypothesis H_o: $\mu \geq \mu_0$ if Z $\leq c$, where c is the α quantile of a standard normal distribution.
2. Decision rule 2: Reject the null hypothesis H_o: $\mu \leq \mu_0$ if Z $\geq c$
3. Decision rule 3: Reject the null hypothesis H_o: $\mu = \mu_0$ if Z $\leq -c$ or $Z \geq c$
   - eg: Compare $z = \frac{\bar{y}-\mu}{\frac{\sigma}{\sqrt{n}}}$, if Z is outside the range of -1.96 and 1.96, the hypothesis is rejected.

假设检验

参数检验

1 单样本的假设检验
- 单样本Z检验
- 单样本T检验
- 单样本χ^2检验（适用总体方差和标准差）

2 双样本的假设检验
- Paired-samples test
- Independent sample test

3 多样本的假设检验
- One-way ANOVA
- Two-way or Multi-way ANOVA

非参数检验（eg. 卡方检验）

单样本的假设检验

研究问题：总体的平均年龄μ是否为24岁？

1. H_0: 总体参数是24。
2. H_1: 总体参数不是24（可能大于24，也有可能小于24）。

T TEST

Dependent variable is continuous, and independent variable is dichotomous(eg. 男性/女性、城市/农村).

1. One-sample T test
2. Paired-samples T test
3. Independent sample T test

Test the significance by $*p \leq 0.1, **p \leq 0.05, ***p \leq 0.01$, and if it is significant, it means findings from a sample can be found in the population.

ONE-SAMPLE T TEST

The purpose of the t test for a single **sample mean** is to determine whether the mean for a random sample of participants differs significantly from a known value (**population mean**) or a hypothetical value （假定总体均值知道）.

PAIRED-SAMPLES T TEST

A researcher drew a random sample from a population and administered a depression scale to the sample. This administration of the scale yielded **pre-test scores**, for which the researcher computed a mean. Then, the researcher administered a new antidepressant drug to the sample. Next, the researcher administered the depression scale again, which yielded **posttest scores**. As a result, for each pretest score earned by an individual, there is an associated posttest score for the same individual. These sets of scores are paired scores.

通过对比自变量两个平均数的差异，判断这种差异在总体中是否存在，即是否具有显著性，从而判断自变量和因变量之间是否有关系。

ONE-SAMPLE T TEST

Example 2.1

假设有人提出*1988*年全国城市居民平均收入为*3800*元，而在*1988*年的*CHIP*数据中，我们发现居民的年收入均值为*3687*元，标准差为*3489*元。
那么在*0.05*的显著性水平下，这一样本结果与*3800*元的提法一致吗？

1. 首先，建立$H_0 : \mu = 3800$ 和 $H_1 : \mu \neq 3800$
2. 根据样本数据，计算T检验统计量
3. stata 命令：ttest earn =3800
4. 结论：否定H_0，即认为全国城市居民平均收入不是3800元。

```
. ttest earn=3800

One-sample t test
```

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf. Interval]	
earn	20421	3687.005	24.4163	3489.139	3639.147	3734.862

```
    mean = mean(earn)                                     t =  -4.6279
Ho: mean = 3800                         degrees of freedom =    20420

   Ha: mean < 3800               Ha: mean != 3800                 Ha: mean > 3800
 Pr(T < t) = 0.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 1.0000
```

Figure 2.2: One-sample T Test

Example 2.2

在1988年的CHIP数据中，有女性9700人，其年收入均值为3329元，标准差为3197元；有男性10721人，其年收入均值为4011元，标准差为3704元。

那么在0.05的显著性水平下，是否存在收入的性别差异？

1. 首先，建立$H_0 : \mu_1 = \mu_2$ 和 $H_1 : \mu_1 \neq \mu_2$
2. 检验方差是否相等（F检验）
3. 根据样本数据，计算T检验统计量
4. stata 命令：ttest earn, by(sex) unequal
5. 结论：否定H_0，即认为男性和女性的年收入均值是不相等

N_1=9700，$\bar{y_1}$=3329元，s_1=3197；N_2=10721，$\bar{y_2}$=4011元，s_2=3704元
那么，在0.05 的显著性水平下，是否存在收入的性别差异？

1. 首先，建立$H_0 : \mu_1 = \mu_2$ 和 $H_1 : \mu_1 \neq \mu_2$
2. 若满足H_0，男性和女性所有可能样本平均收入的分布为同一分布，这一分布的均值应该相等。
3. 尽管从同一分布得到男性和女性样本平均收入可能不同，但二者的距离不应该很远。若二者的距离太远，我们就认为在H_0下是不可能的，即推翻H_0。那么，多远叫“远”？
4. 若t值超过了设定的临界点（即$p < 0.05$），我们就认为二者的距离已经足够远了，从而推翻H_0。

```
. ttest earn, by(sex) unequal

Two-sample t test with unequal variances
```

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
male	10721	4010.848	35.77724	3704.456	3940.718	4080.978
female	9700	3329.074	32.45828	3196.77	3265.449	3392.699
combined	20421	3687.005	24.4163	3489.139	3639.147	3734.862
diff		681.7748	48.30684		587.0895	776.4601

```
    diff = mean(male) - mean(female)                              t =  14.1134
Ho: diff = 0                     Satterthwaite's degrees of freedom =  20373.8

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

Figure 2.3: Two-sample T Test