

t 检验总结与应用

BWB

目录

1	单样本 t 检验	3
1.1	1.1 目的	3
1.2	1.2 应用场景	3
1.3	1.3 数据要求	3
1.4	1.4 假设	3
1.5	1.5 HATPC 步骤	3
1.5.1	1.5.1 假设 (Hypotheses)	3
1.5.2	1.5.2 假设条件 (Assumptions)	3
1.5.3	1.5.3 检验统计量 (Test Statistic)	3
1.5.4	1.5.4 P 值计算 (P-value)	4
1.5.5	1.5.5 结论 (Conclusion)	4
1.6	1.6 说明	4
1.7	1.7 示例	4
1.8	1.8 R 代码示例	4
2	双样本 t 检验	6
2.1	1.1 目的	6
2.2	1.2 应用场景	6
2.3	1.3 数据要求	6
2.4	1.4 假设	6
2.5	1.5 HATPC 步骤	6
2.5.1	1.5.1 假设 (Hypotheses)	6
2.5.2	1.5.2 假设条件 (Assumptions)	6
2.5.3	1.5.3 检验统计量 (Test Statistic)	7
2.5.4	1.5.4 P 值计算 (P-value)	7
2.5.5	1.5.5 结论 (Conclusion)	7
2.6	1.6 说明	7
2.7	1.7 示例	7
2.8	1.8 R 代码示例	8

3	配对 t 检验	10
3.1	1.1 目的	10
3.2	1.2 应用场景	10
3.3	1.3 数据要求	10
3.4	1.4 假设	10
3.5	1.5 HATPC 步骤	10
3.5.1	1.5.1 假设 (Hypotheses)	10
3.5.2	1.5.2 假设条件 (Assumptions)	10
3.5.3	1.5.3 检验统计量 (Test Statistic)	10
3.5.4	1.5.4 P 值计算 (P-value)	11
3.5.5	1.5.5 结论 (Conclusion)	11
3.6	1.6 说明	11
3.7	1.7 示例	11
3.8	1.8 R 代码示例	12
4	随机变量及其性质	13
4.1	期望和方差	13
4.2	随机变量的运算	13
4.3	R 实现	13
4.4	示例：测试啤酒瓶容量	14
4.4.1	步骤 1：陈述假设	14
4.4.2	步骤 2：收集数据并计算置信区间	14
4.4.3	步骤 3：解释结果	14
4.4.4	步骤 4：得出结论	15
4.5	结果可视化	15
5	功效分析	17
5.1	功效的定义	17
5.2	Cohen's d 效应量	17
5.3	R 中的功效计算	17
6	统计概念可视化	18
7	统计概念可视化	18
7.1	正态分布和置信区间	18

1 单样本 t 检验

1.1 1.1 目的

单样本 t 检验用于检验一个样本的均值是否与已知的总体均值有显著差异。

1.2 1.2 应用场景

- 验证某产品的平均性能是否达到标准。
- 判断实验结果是否与预期值一致。

1.3 1.3 数据要求

- 样本是随机抽取的。
- 数据来自正态分布或样本量较大 ($n \geq 30$)。
- 样本观测值相互独立。

1.4 1.4 假设

- 原假设 $H_0: \mu = \mu_0$
- 备择假设 $H_1: \mu \neq \mu_0$ (双尾检验) 或 $\mu < \mu_0$ (左尾检验)、 $\mu > \mu_0$ (右尾检验)

1.5 1.5 HATPC 步骤

1.5.1 1.5.1 假设 (Hypotheses)

- 原假设 $H_0: \mu = \mu_0$
- 备择假设 H_1 : 根据研究目的确定

1.5.2 1.5.2 假设条件 (Assumptions)

- 样本是随机独立抽样的。
- 样本来自正态分布或样本量较大。

1.5.3 1.5.3 检验统计量 (Test Statistic)

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$$

其中:

- \bar{X} : 样本均值
- μ_0 : 总体均值 (原假设中的值)
- S : 样本标准差
- n : 样本容量

1.5.4 1.5.4 P 值计算 (P-value)

- 自由度 $df = n - 1$
- 根据 t 分布计算 P 值, 或使用统计软件直接获得。

1.5.5 1.5.5 结论 (Conclusion)

- 若 P 值小于显著性水平 α , 则拒绝原假设。
- 做出关于总体均值的统计推断。

1.6 1.6 说明

单样本 t 检验通过比较样本均值与已知的总体均值, 判断两者之间是否存在显著差异。在实践中, 通常用于验证某一特定参数是否达到预期标准。例如, 检验某品牌电池的平均寿命是否达到厂商声称的数值。

1.7 1.7 示例

问题描述:

假设我们有一款电池, 厂商声称其平均寿命为 1000 小时。我们随机抽取了 15 个电池进行测试, 得到的寿命数据如下 (单位: 小时):

950, 970, 930, 960, 940, 980, 920, 955, 965, 945, 935, 975, 985, 925, 950

我们想检验这些电池的平均寿命是否低于厂商声称的值。

1.8 1.8 R 代码示例

下面的代码使用 R 语言的 `t.test` 函数对电池寿命数据进行单样本 t 检验。

Listing 1: 单样本 t 检验的 R 实现

```
# 样本数据
battery_life <- c(950, 970, 930, 960, 940, 980, 920, 955, 965, 945, 935, 975, 985,
  925, 950)
```

```
# 执行单样本 t 检验
t_test_result <- t.test(battery_life, mu = 1000, alternative = "less")
print(t_test_result)

# 显著性水平
alpha <- 0.05

# 判断是否拒绝原假设
if (t_test_result$p.value < alpha) {
  cat("拒绝原假设，有足够证据表明平均寿命低于 1000 小时。\\n")
} else {
  cat("无法拒绝原假设，缺乏证据表明平均寿命低于 1000 小时。\\n")
}
```

解答：

运行上述代码，我们得到：

- t 值约为 -8.78，自由度为 14。
- P 值非常小（约为 1.04×10^{-7} ），小于显著性水平 0.05。

因此，我们拒绝原假设，认为电池的平均寿命显著低于 1000 小时。

2 双样本 t 检验

2.1 1.1 目的

双样本 t 检验用于比较两个独立样本的均值是否有显著差异。

2.2 1.2 应用场景

- 比较两种教学方法对学生成绩的影响。
- 评估两种药物的疗效差异。

2.3 1.3 数据要求

- 两组样本彼此独立。
- 两组数据来自正态分布或样本量较大。
- 方差齐性（如果假设方差不等，需使用 Welch's t 检验）。

2.4 1.4 假设

- 原假设 $H_0: \mu_1 = \mu_2$
- 备择假设 $H_1: \mu_1 \neq \mu_2$ （双尾检验）

2.5 1.5 HATPC 步骤

2.5.1 1.5.1 假设 (Hypotheses)

- 原假设 $H_0: \mu_1 = \mu_2$
- 备择假设 $H_1: \mu_1 \neq \mu_2$

2.5.2 1.5.2 假设条件 (Assumptions)

- 两组样本是独立的随机样本。
- 两组数据来自正态分布。
- 方差齐性（可通过方差齐性检验验证）。

2.5.3 1.5.3 检验统计量 (Test Statistic)

当方差相等时：

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

其中，合并标准差 S_p 为：

$$S_p = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}$$

当方差不等时，使用 Welch's t 检验：

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

2.5.4 1.5.4 P 值计算 (P-value)

- 自由度：方差相等时， $df = n_1 + n_2 - 2$ ；方差不等时，使用 Welch-Satterthwaite 公式计算自由度。
- 根据 t 分布计算 P 值。

2.5.5 1.5.5 结论 (Conclusion)

- 若 P 值小于显著性水平 α ，则拒绝原假设。
- 做出关于两组均值差异的统计推断。

2.6 1.6 说明

双样本 t 检验用于比较两组独立样本的均值，判断两者之间是否存在显著差异。在教育研究中，可以用于比较不同教学方法的效果；在医学研究中，可以用于比较不同治疗方法的疗效。

2.7 1.7 示例

问题描述：

某研究想比较 A、B 两种教学方法对学生考试成绩的影响。随机选取了两组学生，A 组和 B 组的成绩数据如下（单位：分）：

A 组 ($n_1 = 15$):

85, 88, 90, 87, 86, 89, 91, 84, 88, 90, 86, 87, 89, 85, 88

B 组 ($n_2 = 15$):

83, 80, 85, 82, 81, 84, 83, 79, 82, 80, 81, 83, 82, 80, 84

我们想检验两组学生的平均成绩是否有显著差异。

2.8 1.8 R 代码示例

下面的代码使用 R 语言的 `t.test` 函数对两组成绩数据进行双样本 `t` 检验。

Listing 2: 双样本 `t` 检验的 R 实现

```
# A 组成绩
group_A <- c(85, 88, 90, 87, 86, 89, 91, 84, 88, 90, 86, 87, 89, 85, 88)

# B 组成绩
group_B <- c(83, 80, 85, 82, 81, 84, 83, 79, 82, 80, 81, 83, 82, 80, 84)

# 检验方差齐性
var_test <- var.test(group_A, group_B)
print(var_test)

# 根据方差齐性结果选择 var.equal 参数
if (var_test$p.value > 0.05) {
  var_equal <- TRUE
} else {
  var_equal <- FALSE
}

# 执行双样本 t 检验
t_test_result <- t.test(group_A, group_B, var.equal = var_equal)
print(t_test_result)

# 显著性水平
alpha <- 0.05

# 判断是否拒绝原假设
if (t_test_result$p.value < alpha) {
  cat("拒绝原假设, 有足够证据表明两组平均成绩有显著差异.\n")
} else {
  cat("无法拒绝原假设, 缺乏证据表明两组平均成绩有显著差异.\n")
}
```

解答:

运行上述代码, 我们得到:

- 方差齐性检验 P 值大于 0.05，认为方差相等。
- t 值约为 8.25，自由度为 28。
- P 值非常小（约为 2.14×10^{-9} ），小于显著性水平 0.05。

因此，我们拒绝原假设，认为两组平均成绩存在显著差异。

3 配对 t 检验

3.1 1.1 目的

配对 t 检验用于比较两个相关样本（同一对象的两个测量值）的均值差异。

3.2 1.2 应用场景

- 测试治疗前后患者的健康指标变化。
- 评估培训前后员工技能的提升。

3.3 1.3 数据要求

- 数据成对存在，且每对之间是相关的。
- 差值近似正态分布或样本量较大。
- 各对差值彼此独立。

3.4 1.4 假设

- 原假设 $H_0: \mu_D = 0$
- 备择假设 $H_1: \mu_D \neq 0$ （双尾检验）或根据研究目的确定方向

3.5 1.5 HATPC 步骤

3.5.1 1.5.1 假设 (Hypotheses)

- 原假设 $H_0: \mu_D = 0$
- 备择假设 $H_1: \mu_D > 0$ （药物有效降低血压）

3.5.2 1.5.2 假设条件 (Assumptions)

- 成对数据的差值是独立的随机样本。
- 差值近似正态分布。

3.5.3 1.5.3 检验统计量 (Test Statistic)

$$t = \frac{\bar{D}}{S_D / \sqrt{n}}$$

其中：

- \bar{D} : 差值的样本均值
- S_D : 差值的样本标准差
- n : 配对样本的数量

3.5.4 1.5.4 P 值计算 (P-value)

- 自由度 $df = n - 1$
- 根据 t 分布计算 P 值。

3.5.5 1.5.5 结论 (Conclusion)

- 若 P 值小于显著性水平 α , 则拒绝原假设。
- 做出关于差值均值的统计推断。

3.6 1.6 说明

配对 t 检验用于比较同一对象在不同条件下的测量结果, 例如治疗前后的健康指标。通过分析成对差值, 判断是否存在显著变化。

3.7 1.7 示例

问题描述:

一组 10 名患者在服用某药物前后血压的测量值如下 (单位: mmHg):

患者编号	服药前血压	服药后血压	差值 (前 - 后)
1	150	140	10
2	160	145	15
3	155	150	5
4	148	138	10
5	152	140	12
6	158	147	11
7	149	135	14
8	153	142	11
9	151	139	12
10	154	143	11

表 1: 患者服药前后血压及差值

我们想检验药物是否有效降低了患者的血压。

3.8 1.8 R 代码示例

下面的代码使用 R 语言的 `t.test` 函数对患者服药前后的血压进行配对 t 检验。

Listing 3: 配对 t 检验的 R 实现

```
# 服药前后血压数据
before <- c(150, 160, 155, 148, 152, 158, 149, 153, 151, 154)
after <- c(140, 145, 150, 138, 140, 147, 135, 142, 139, 143)

# 计算差值
diff <- before - after

# 检查差值的正态性 (可选)
shapiro.test(diff)

# 执行配对 t 检验
t_test_result <- t.test(before, after, paired = TRUE, alternative = "greater")
print(t_test_result)

# 显著性水平
alpha <- 0.05

# 判断是否拒绝原假设
if (t_test_result$p.value < alpha) {
  cat("拒绝原假设, 有足够证据表明药物有效降低血压。\\n")
} else {
  cat("无法拒绝原假设, 缺乏证据表明药物有效。\\n")
}
```

解答:

运行上述代码, 我们得到:

- t 值约为 11.87, 自由度为 9。
- P 值非常小 (约为 1.16×10^{-7}), 小于显著性水平 0.05。

因此, 我们拒绝原假设, 认为药物有效降低了患者的血压。

4 随机变量及其性质

4.1 期望和方差

对于随机变量 X :

$$E(X) = \mu \text{ (均值)}$$

$$Var(X) = \sigma^2 \text{ (方差)}$$

$$SD(X) = \sigma \text{ (标准差)}$$

4.2 随机变量的运算

对于独立的随机变量 X_1, X_2, \dots, X_n :

$$E(X_1 + X_2 + \dots + X_n) = E(X_1) + E(X_2) + \dots + E(X_n)$$

$$Var(X_1 + X_2 + \dots + X_n) = Var(X_1) + Var(X_2) + \dots + Var(X_n)$$

对于常数 c 和随机变量 X :

$$E(cX) = cE(X)$$

$$Var(cX) = c^2 Var(X)$$

4.3 R 实现

Listing 4: R 中的随机变量运算

```
# 生成随机变量
X1 <- rnorm(1000, mean = 10, sd = 2)
X2 <- rnorm(1000, mean = 5, sd = 1)

# 计算均值和方差
mean_X1 <- mean(X1)
var_X1 <- var(X1)
mean_X2 <- mean(X2)
var_X2 <- var(X2)

# 随机变量的和
X_sum <- X1 + X2
mean_sum <- mean(X_sum)
var_sum <- var(X_sum)

# 打印结果
cat("E(X1) =", mean_X1, "\n")
```

```
cat("Var(X1) =", var_X1, "\n")
cat("E(X2) =", mean_X2, "\n")
cat("Var(X2) =", var_X2, "\n")
cat("E(X1 + X2) =", mean_sum, "\n")
cat("Var(X1 + X2) =", var_sum, "\n")
```

4.4 示例：测试啤酒瓶容量

让我们考虑一个实际例子，我们想测试啤酒瓶的平均容量是否与声称的 375 毫升有显著差异。

4.4.1 步骤 1：陈述假设

- 原假设 (H_0): 真实平均容量为 375 毫升 ($\mu = 375$)
- 备择假设 (H_1): 真实平均容量不是 375 毫升 ($\mu \neq 375$)

4.4.2 步骤 2：收集数据并计算置信区间

我们将使用 R 来分析给定的数据并计算 95% 置信区间。

Listing 5: R 中的置信区间和假设检验

```
# 给定数据
beer_data <- c(374.8, 375.0, 375.3, 374.8, 374.4, 374.9)

# 计算样本均值
sample_mean <- mean(beer_data)
print(paste("样本均值:", round(sample_mean, 4)))

# 执行t检验并计算置信区间
test_result <- t.test(beer_data, mu = 375, conf.level = 0.95)

# 打印结果
print(test_result)
```

4.4.3 步骤 3：解释结果

让我们查看输出：

```
[1] "样本均值: 374.8667"
```

单样本 t 检验

数据: beer_data

$t = -0.70711$, $df = 5$, $p\text{-value} = 0.5108$

备择假设: 真实均值不等于 375

95 percent 置信区间:

374.5183 375.2151

样本估计:

x 的均值

374.8667

从这个输出, 我们可以解释:

- 样本均值是 374.87 毫升。
- 95% 置信区间是 [374.52 毫升, 375.22 毫升]。
- p 值为 0.5108, 大于我们通常使用的 0.05 显著性水平。

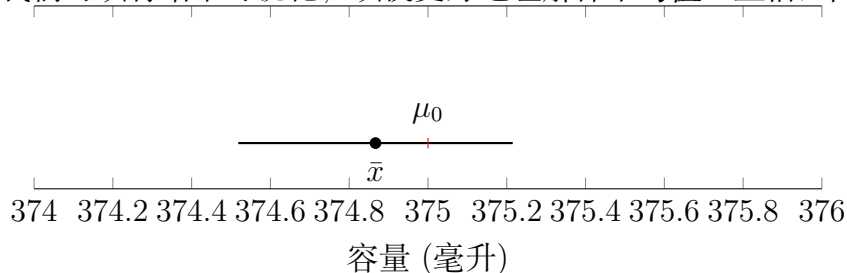
4.4.4 步骤 4: 得出结论

基于这些结果, 我们可以得出结论:

- 由于 p 值 (0.5108) 大于 0.05, 我们未能拒绝原假设。
- 我们没有足够的证据表明真实平均容量与 375 毫升有显著差异。
- 95% 置信区间 (374.52 毫升到 375.22 毫升) 包含了 375 毫升, 进一步支持了我们的结论。
- 尽管样本均值 (374.87 毫升) 略小于 375 毫升, 但考虑到我们的样本大小和变异性, 这个差异在统计学上并不显著。

4.5 结果可视化

我们可以将结果可视化, 以便更好地理解样本均值、置信区间和假设值之间的关系。



在这个可视化中:

- 蓝线表示 95% 置信区间。

- 蓝点表示样本均值。
- 红色竖线表示假设的总体均值（375 毫升）。

如我们所见，假设值 375 毫升落在我们的置信区间内，这与我们未能拒绝原假设的决定一致。样本均值接近 375 毫升，且差异在统计学上并不显著。

5 功效分析

5.1 功效的定义

统计功效是正确拒绝错误原假设的概率。它受以下因素影响：

- 样本大小
- 效应量
- 显著性水平 (α)

5.2 Cohen's d 效应量

Cohen's d 是一种标准化的效应量度量：

$$d = \frac{|\mu_1 - \mu_2|}{\sigma}$$

其中 μ_1 和 μ_2 是两组的均值， σ 是合并标准差。

5.3 R 中的功效计算

Listing 6: R 中 t 检验的功效计算

```
library(pwr)

# 计算功效
power_result <- pwr.t.test(n = 30, d = 0.5, sig.level = 0.05, type = "two.sample",
  alternative = "two.sided")

# 打印结果
print(power_result)

# 计算80%功效所需的样本量
sample_size_result <- pwr.t.test(power = 0.8, d = 0.5, sig.level = 0.05, type = "two
  .sample", alternative = "two.sided")

print(sample_size_result)
```

6 统计概念可视化

7 统计概念可视化

7.1 正态分布和置信区间

