One-sample Inference

Applied Statistics

Fall 2025

目录

大纲 Outline

1. 总体均值的推断 Inference for the Mean of a Population
2. 总体方差的推断 Inference for the population variance
3. 推断方法的稳健性 Robustness of an inference procedure

1 总体均值的推断

- 在使用数据估计总体均值时，总体方差 σ 通常也未知。
  When using data to estimate the population mean, it is common that population variance is also unknown.

- 我们使用样本标准差 s 来估计总体方差 σ。
  We use sample standard deviation s to estimate population variance σ.

- **标准误 (Standard Error, SE)**: 是从数据中估计出的样本统计量的标准差。
  Standard Error (SE): the standard deviation of a sample statistic estimated from the data.

- 对于样本均值的标准误：
  For sample mean:
  $$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$$

1.0.1 t 统计量与 t 分布 The t/ Statistic and t Distribution

- 由于估计 σ 引入了额外的变异性，以下新统计量不再服从正态分布。
  Estimating introduces extra variability, so the new statistic below does not follow a normal distribution anymore.
  $$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

- 该统计量服从自由度为 $n-1$ 的 t 分布。
  The statistic follows the t distribution with $n-1$ degrees of freedom.

1.1 学生 t 分布 Student t Distribution

1.1.1 t 分布的定义 Definition

- 随机变量 t 具有自由度为 df 的 t 分布，定义为：
  Random variable t has a t distribution with df degrees of freedom:
  $$t = \frac{z}{\sqrt{\chi^2/df}}$$
  其中 z 服从标准正态分布，χ^2 服从自由度为 df 的卡方分布，且 z 和 χ^2 独立。
  where z has a standard normal distribution and χ^2 has a chi-squared distribution with df degrees of freedom. z and χ^2 are independent.

1.1.2 t 分布的推导 Derivation of the t Distribution

从正态样本推导 t 统计量 Derivation of t statistic from Normal Sample

假设从 $N(\mu, \sigma)$ 总体中抽取一个简单随机样本，样本量为 n。

Suppose that a simple random sample of size n is drawn from an $N(\mu, \sigma)$ population.

样本均值的抽样分布可以表示为：

Then the sampling distribution for $\bar{x}$ becomes:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = \frac{\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)s^2}{\sigma^2}/(n-1)}}$$

已知：

- $\frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
- $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$
- $\bar{x}$ 和 s^2 独立（对于正态总体）
  $\bar{x}$ and s^2 are independent (given normal population).

根据 t 分布的定义，上述比值服从自由度为 $n-1$ 的 t 分布。

By definition of t distribution, the above ratio follows a t distribution with $n-1$ degrees of freedom.

1.1.3 t 分布的性质 Properties of the t Distribution

- 学生 t 分布是“钟形”的，关于均值 0 对称。
  The Student t distribution is ”mound” shaped and symmetrical about its mean of 0.

- 与正态分布相比，t 分布有更厚的尾部。
  Student t distribution has fatter tails compared to the normal distribution.

- 随着自由度的增加，t 分布趋近于标准正态分布。
  As the number of degrees of freedom increases, the t distribution approaches the standard normal distribution.

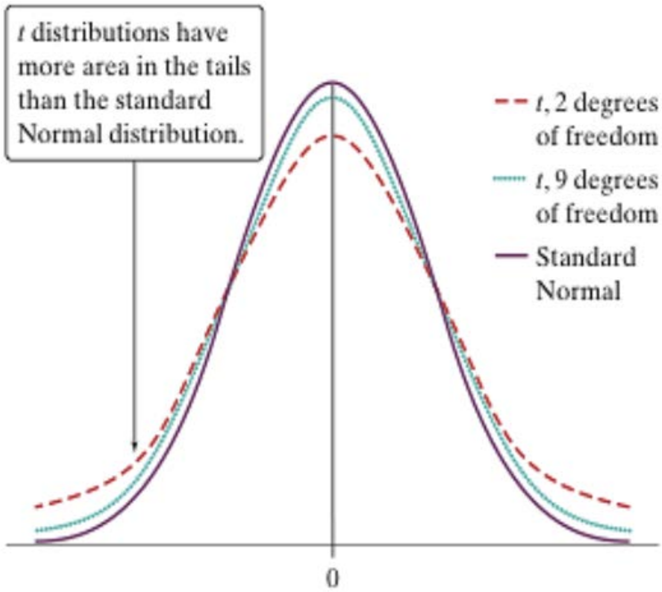


图 1: t 分布与正态分布的比较 Comparison of t Distribution and Normal Distribution

1.1.4 t 分布的历史 History of the t Distribution

- t 分布由威廉 • 戈塞特 (William S. Gosset) 于 1908 年发现。
  The t distributions were discovered in 1908 by William S. Gosset.

- 戈塞特是吉尼斯啤酒公司的统计学家，公司禁止员工发表与酿酒相关的发现。
  Gosset was a statistician employed by the Guinness brewing company, which prohibited its employees from publishing their discoveries that were brewing-related.

- 在这种情况下，公司允许他以笔名"Student" 发表一篇不涉及酿酒的例子。
  The company let him publish under the pen name "Student" using an example that did not involve brewing.

- 为了纪念他，t 分布常被称为"Student's t"。
  The t distribution is often called "Student's t" in his honor.

1.2 单样本 t 置信区间 One-sample t Confidence Interval

1.2.1 公式 Formula

- 当 σ 未知时，误差边际 $z * \sigma/\sqrt{n}$ 变为 $t * s/\sqrt{n}$。
  When σ is unknown, the margin of error $z * \sigma/\sqrt{n}$ becomes $t * s/\sqrt{n}$.

- 我们使用自由度为 $df = n - 1$ 的 t 分布来查找临界值，而不是标准正态分布。
  We use a t distribution with $df = n - 1$ instead of a standard normal to find the

critical value.

- $1-\alpha$ 置信区间为：
  The $1-\alpha$ confidence interval is now:
  $$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$$

1.2.2 性质与解释 Properties and Interpretation

- 当总体分布为正态时，该置信区间的精确置信水平为 $1-\alpha$。
  This confidence interval is exactly $1-\alpha$ when the population distribution is Normal.
- 对于大样本 (n 较大)，即使总体分布非正态，该区间也近似正确。
  It is approximately correct for large n in other cases.
- 误差边际的估计考虑到了来自估计方差的增加的”变异性”，因此比基于 z 统计量的区间更宽。
  The estimation of the margin of error accounts for the increased ”variability” from the estimated variance, making it larger than the one based on z-statistic.

1.2.3 实践示例：夏普比率 Practice: the Sharpe ratio

夏普比率置信区间计算 Sharpe Ratio Confidence Interval Calculation

夏普比率衡量超额风险的平均回报 $\frac{r-r_F}{s}$。
The Sharpe ratio measures the average return for excess risk $\frac{r-r_F}{s}$.
已知：
Given:

- 样本量 $n = 100$
  Sample size is 100.
- 平均夏普比率 $\bar{x} = 0.45$
  The average Sharpe ratio is 0.45.
- 样本标准差 $s = 0.3$
  The sample has a standard deviation of 0.3.

使用 t 分布计算并解释总体均值的 90% 置信区间。
Calculate and interpret the 90% confidence interval for the population mean using t-distribution.
计算：

1. 置信区间为：

The confidence interval will be:

$$\bar{x} \pm t_{0.05}(99)\frac{s}{\sqrt{n}} = 0.45 \pm 1.66\frac{0.3}{\sqrt{100}} = 0.45 \pm 0.0498$$

2. 注意到 $t_{0.05}(99) = 1.66$ 非常接近 $z_{0.05} = 1.65$
   Notice that $t_{0.05}(99) = 1.66$ is very close to $z_{0.05} = 1.65$.

解释：我们有 90% 的把握认为总体平均夏普比率在 0.4002 到 0.4998 之间。
Interpretation: We are 90% confident that the population mean Sharpe ratio is between 0.4002 and 0.4998.

1.3 单样本 t 检验 The One-sample t Test

1.3.1 检验统计量 Test Statistic

- 当总体标准差未知时，我们估计标准误为 $\frac{s}{\sqrt{n}}$。
  When the population standard deviation is unknown we estimate the standard error as $\frac{s}{\sqrt{n}}$.
- 我们使用的检验统计量为：
  We use the test statistic:
  $$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$
  其中 μ_0 是原假设中假设的总体均值。
  where μ_0 is the hypothesized population mean in the null hypothesis.
- 该统计量服从自由度为 $n-1$ 的 t 分布。
  This statistic has a $t(n-1)$ distribution.

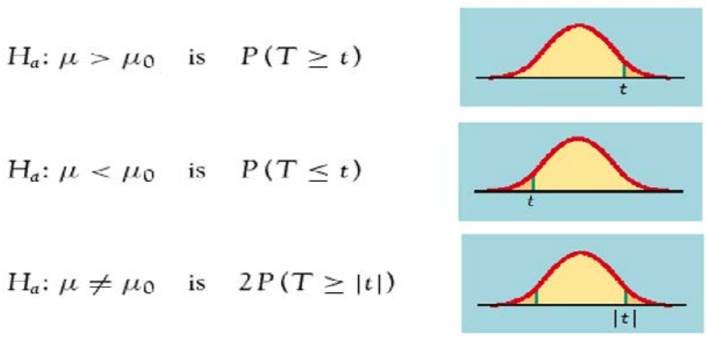
1.3.2 P 值计算 P-value Calculation

- P 值基于与备择假设一致的 $t(n-1)$ 分布计算。
  We find the P-value based on the $t(n-1)$ that is consistent with the alternative hypothesis.
- 如果总体分布为正态，这些 P 值是精确的；对于大样本，即使总体非正态也近似正确。
  These P-values are exact if the population distribution is Normal and are approximately correct for large n in other cases.

1.3.3 不同备择假设下的 P 值公式 P-value Formulas for Different Alternatives

备择假设 Alternative Hypothesis	符号表示 Notation	P 值公式 P-value Formula
右尾检验 Right-tailed test	$H_a : \mu > \mu_0$	$P(T \geq t)$
左尾检验 Left-tailed test	$H_a : \mu < \mu_0$	$P(T \leq t)$
双尾检验 Two-tailed test	$H_a : \mu \neq \mu_0$	$2P(T \geq \|t\|)$

表 1: 不同假设下的 P 值计算公式 P-value Calculation Formulas for Different Hypotheses



1.3.4 示例：互联网手机服务收费 Example: Internet Cell Service Rates

手机服务收费比较 Cell Service Rate Comparison

背景：互联网手机服务提供商声称他们的费率低于传统手机服务公司（如 AT&T）。
Background: Internet cell service providers argue that their rates are lower than conventional cell service companies such as AT&T.

已知：

- AT&T 客户的月平均话费为 \$17.09（标准差已知）。
  The mean and standard deviation of monthly cell bills for AT&T customers are \$17.09.
- 从互联网手机服务中随机抽取 100 名客户，平均话费为 \$16.42，标准差为 \$3.87。
  Take a random sample of 100 customers of the internet cell services and the mean bill is \$16.42 with a standard deviation \$3.87.

问题：

1. 我们能否得出互联网手机服务费率低于 AT&T 的结论？
   Can we conclude that the rates are lower than AT&T?

2. 我们能否得出 AT&T 的话费与互联网竞争对手存在差异的结论？
   Can we conclude that there is a difference between AT&T's bills and the internet competitor?

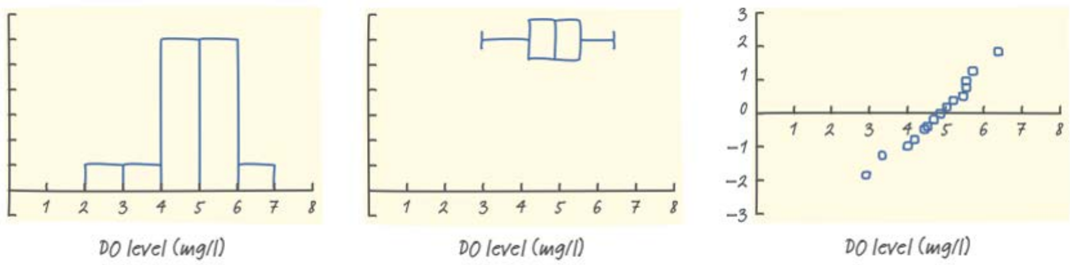
解答思路：

1. 对于第一个问题，使用左尾检验：$H_0 : \mu = 17.09$, $H_a : \mu < 17.09$。
2. 对于第二个问题，使用双尾检验：$H_0 : \mu = 17.09$, $H_a : \mu \neq 17.09$。
3. 计算 t 统计量：$t = \frac{16.42-17.09}{3.87/\sqrt{100}} = \frac{-0.67}{0.387} \approx -1.73$。
4. 查找 P 值或与临界值比较得出结论。

1.4 检验条件的检查 Checking the Required Conditions

1.4.1 必要条件 Required Conditions

- **随机抽样 Random sampling:** 样本必须是来自总体的简单随机样本 (SRS)。
  The sample must be a SRS from the population.
- **正态性/大样本 Normality or Large Sample:** 可以通过绘制直方图、箱线图、正态分位数图或 Q-Q 图来检查" 正态" 形状，特别是对于小样本。
  We can draw a histogram, a boxplot, a Normal quantile plot or a Q-Q plot to check how "normal" the shape is especially for small samples.



1.4.2 影响推断的关键因素 Factors that Strongly Matter

- **随机抽样 Random sampling:** 样本必须是来自目标总体的 SRS，这样所有的概率规则才适用。
  The sample must be a SRS from the population. Then all the probability rules will apply.
- **异常值和偏度 Outliers and skewness:** 它们强烈影响均值，从而影响 t 检验。然而，随着样本量的增加，由于中心极限定理（$\bar{x}$ 接近正态）和大数定律（s 将是 σ

的准确估计），它们的影响会减弱。

They strongly influence the mean and therefore the t procedures. However, their impact diminishes as the sample size gets larger because of the central limit theorem ($\bar{x}$ is close to Normal) and the law of large numbers (s will be an accurate estimate of σ).

- 除小样本情况外，数据来自目标总体的 SRS 这一条件比总体分布为正态的条件更重要。

  Except in the case of small samples, the condition that the data are an SRS from the population of interest is more important than the condition that the population distribution is Normal.

1.5 t 检验的稳健性 Robustness of t Procedures

1.5.1 实用指南 Practical Guidelines

样本量 Sample Size	建议 Recommendation
样本量 < 15	当数据看起来接近正态时使用 t 检验。
样本量至少 15 且小于 40	可以使用 t 检验，除非存在异常值或强偏度。
大样本 ≥ 40	即使对于明显偏斜的分布，也可以使用 t 检验。

表 2: t 检验的稳健性指南 Robustness Guidelines for t Procedures

1.6 推断分布的选择 Choice of Inference Distributions

抽样来源	小样本统计量	大样本统计量
正态分布，方差已知	z	z
正态分布，方差未知	t	t^*
非正态分布，方差已知 Nonnormal distribution with known variance	不可用 not available not available	z z
非正态分布，方差未知 Nonnormal distribution with unknown variance	不可用 not available not available	t^* t^*

表 3: 不同情况下的推断分布选择 Choice of Inference Distributions for Different Situations

注： 表中 t^* 表示对于大样本，即使总体非正态，t 检验也近似有效。

Note: t^* in the table indicates that for large samples, the t-test is approximately valid even if the population is not normal.

2 总体方差的推断 Inference About Population Variance

2.1 研究动机 Motivation

- **方差作为风险度量 Variance as a measure of risk:** 比较两个股票投资组合的方差，以选择风险较小的一个。
  comparing the variance of two stock portfolios to select one with a smaller risk.
- **质量控制 Quality control:** 减小产品重量、尺寸或体积的方差。
  to reduce the variance of products including weight, size, or volume.
- 研究的参数是总体方差 σ^2。
  The parameter to investigate is the population variance σ^2.
- 方差推断仅适用于正态总体。
  The inference applies only to normal populations.

2.2 卡方分布 Chi-Squared Distribution

2.2.1 定义与性质 Definition and Properties

- v 个独立标准正态随机变量平方和构成一个自由度为 v 的卡方分布：
  The sum of v independent standard normal random variables' squares forms a χ^2 distribution with v degrees of freedom:

$$\chi^2 = \sum_{i=1}^{v} Z_i^2 \sim \chi^2(v)$$

- 期望和方差：
  Expectation and variance:

$$E(\chi^2) = v$$

$$V(\chi^2) = 2v$$

- 卡方分布不是对称的。
  The chi-squared distribution is not symmetrical.
- 平方运算强制非负值：寻找 $P(\chi^2 < 0)$ 是无意义的。
  The square forces non-negative values: finding $P(\chi^2 < 0)$ is illogical.

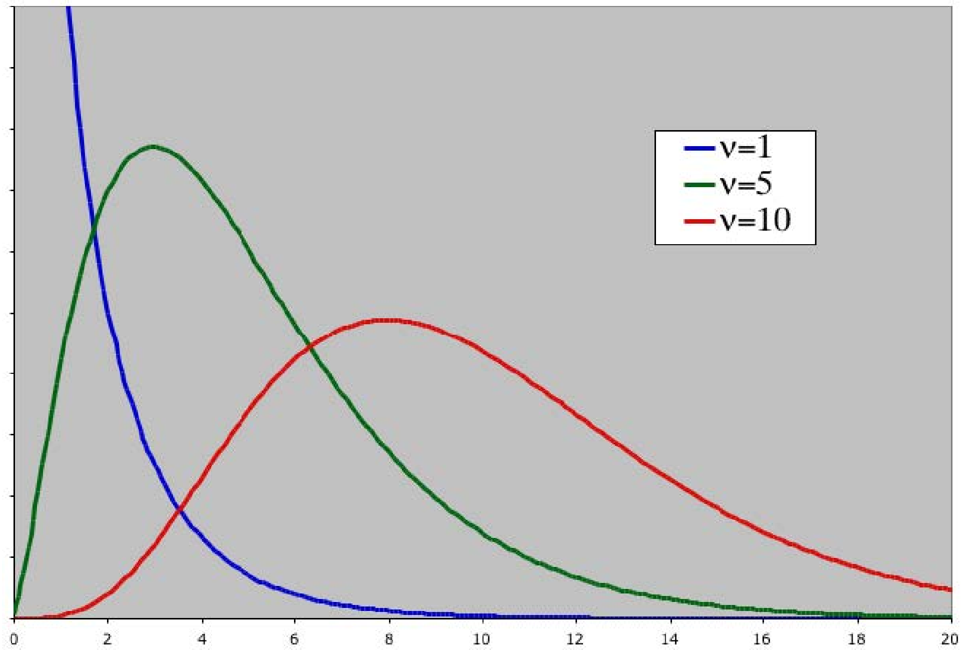


图 2: 不同自由度的卡方分布 Chi-Squared Distribution with Different Degrees of Freedom

2.3 样本方差的性质 Properties of Sample Variance

- 样本方差 s^2 是总体方差 σ^2 的无偏、一致且有效的点估计量。
  The sample variance s^2 is an unbiased, consistent, and efficient point estimator for σ^2.

- 如果 $X_i \sim N(\mu, \sigma^2)$，则：
  If $X_i \sim N(\mu, \sigma^2)$, then:
$$\sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$$

- 对于正态总体，有：
  For normal populations, we have:
$$\frac{\sum (X_i - \bar{x})^2}{\sigma^2} = \frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$$

2.4 总体方差的假设检验 Testing for a Population Variance

2.4.1 检验统计量 Test Statistic

- 常用的检验是左尾检验 $H_a : \sigma^2 < \sigma_0^2$（例如质量控制中希望方差减小）。
  Commonly a left tail test $H_a : \sigma^2 < \sigma_0^2$. Why?

- 检验统计量为：
  The test statistic is:
  $$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$
  服从自由度为 $n-1$ 的卡方分布。
  which has a chi-squared distribution with $n-1$ degrees of freedom.

- 对于左尾检验，P 值是 χ^2 统计量小于或等于计算值的概率。
  For a left-tailed test, the p-value is $P(\chi^2 \leq \text{calculated value})$.

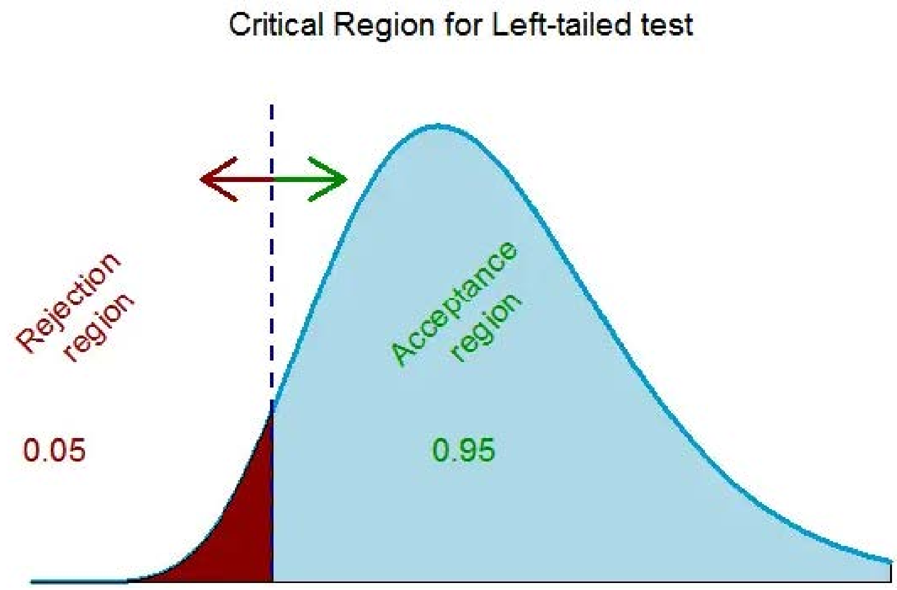


图 3: 左尾检验

2.4.2 实践示例：容器填充机器 Practice: Container-filling Machines

容器填充方差检验 Container-filling Variance Test

背景：容器填充机用于包装各种液体，包括牛奶、软饮料和油漆。理想情况下，液体量应只有轻微变化。

Container-filling machines are used to package a variety of liquids; including milk, soft drinks, and paint. Ideally, the amount of liquid should vary only slightly.

声称：一种新型机器声称可以如此一致地填充 1 升容器，以至于填充量的方差将小于 1 毫升。

A new type of machine boasts that this machine can fill 1 liter (1,000 ml) containers so consistently that the variance of the fills will be less than 1ml.

数据：

- 样本方差 $s^2 = 0.633$ ml^2
  The sample has a variance of 0.633.

- 样本量 $n = 25$
  The sample size is 25.

问题：在 5% 的显著性水平下，这些数据是否允许总裁做出这一声称？
Do these data allow the president to make this claim at the 5% significance level?
解答思路：

- 假设：$H_0 : \sigma^2 = 1$, $H_a : \sigma^2 < 1$（左尾检验）
- 检验统计量：$\chi^2 = \frac{(25-1)\times 0.633}{1} = 24 \times 0.633 = 15.192$
- 临界值：$\chi^2_{0.95}(24) \approx 13.848$
- 决策：由于 $15.192 > 13.848$，不在拒绝域，不能拒绝 H_0。
- 结论：在 5% 水平下，没有足够证据支持方差小于 1ml 的声称。

2.5 总体方差的估计 Estimation for a Population Variance

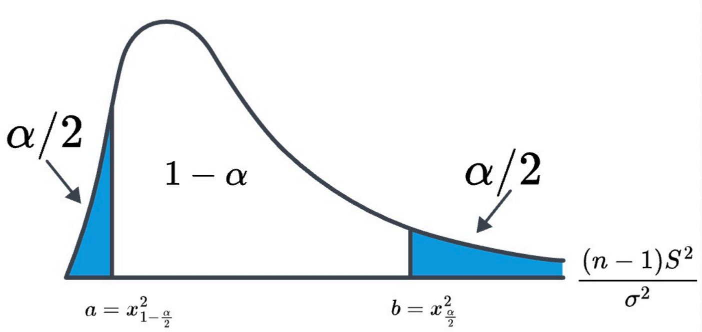
2.5.1 方差的置信区间 Confidence Interval for Variance

- 基于样本量为 n 的样本，总体方差 σ^2 的 $1-\alpha$ 置信区间为：
  Confidence interval of a population variance σ^2 based on a sample of size n is given by:

$$\text{下限 Lower limit} = \frac{(n-1)s^2}{\chi^2_{\alpha/2}}$$

$$\text{上限 Upper limit} = \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

  其中 $\chi^2_{\alpha/2}$ 和 $\chi^2_{1-\alpha/2}$ 是自由度为 $n-1$ 的卡方分布的临界值。
  where $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ are critical values from the chi-squared distribution with $n-1$ degrees of freedom.



2.5.2 标准差的置信区间 Confidence Interval for Standard Deviation

- 总体标准差 σ 的置信区间是方差区间的平方根：
  Confidence interval of a population standard deviation σ is the square root of the variance interval:
  $$\text{下限 Lower limit} = \sqrt{\frac{(n-1)s^2}{\chi^2_{\alpha/2}}}$$
  $$\text{上限 Upper limit} = \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}}$$

2.5.3 实践示例：投资组合回报 Practice: Portfolio Returns

投资组合回报标准差置信区间 Portfolio Returns Standard Deviation CI

背景：在 10 年期间，一个投资组合年回报的标准差为 15%。年回报服从正态分布。

During a 10-year period, the standard deviation of annual returns on a portfolio was 15 percent. The annual returns are normally distributed.

问题：求年回报总体标准差的 95% 置信区间。

Find a 95% confidence interval for the population standard deviation of the annual return.

解答：

- 已知：$n = 10$, $s = 0.15$, $\alpha = 0.05$
- 自由度：$df = 10 - 1 = 9$
- 临界值：$\chi^2_{0.025}(9) \approx 19.023$, $\chi^2_{0.975}(9) \approx 2.700$
- 置信区间：
  $$\sqrt{\frac{(10-1)\times 0.15^2}{19.023}} \le \sigma \le \sqrt{\frac{(10-1)\times 0.15^2}{2.700}}$$
  $$\sqrt{\frac{9\times 0.0225}{19.023}} \le \sigma \le \sqrt{\frac{9\times 0.0225}{2.700}}$$
  $$0.103 \le \sigma \le 0.274$$
- 解释：我们有 95% 的把握认为年回报的总体标准差在 10.3% 到 27.4% 之间。

3 稳健性 Robustness of Inference Procedures

3.1 稳健性的定义 Definition of Robustness

- 如果一个统计推断程序所需的概率计算对假设的违反不敏感，则该程序被称为稳健的。
  A statistical inference procedure is called robust if the required probability calculations are insensitive to violations of the assumptions made.

- 当使用程序的条件被违反时，置信水平或 P 值不会发生太大变化。
  The confidence level or P-value does not change very much when the conditions for the use of the procedure are violated.

- 例如，均值的 t 检验和方差的 χ^2 检验都只在总体精确服从正态分布时才完全正确。
  For example, both t procedures for means and χ^2 procedures for variances are exactly correct only when the population is distributed exactly normally.

3.2 模拟研究：t 检验的稳健性 Simulation Study: Robustness of t-test

t 检验的稳健性模拟 Robustness Simulation of t-test

方法：

- 从均匀分布 $U(0,1)$ 中生成 1000 个样本，每个样本量 $n = 10$。
  Generate 1000 samples of size 10 from a uniform population $U(0,1)$.

- 进行双尾 t 检验 $H_0 : \mu = 0.5$，显著性水平 $\alpha = 0.05$。
  Run a two-tail t-test $H_0 : \mu = 0.5$ at 5% significance level.

- 记录 1000 个 P 值中小于 0.05 的比例。
  Record the proportion of p-values less than 0.05 among these 1000 tests.

结果：

变量 Variable	观测数 Obs	均值 Mean	标准差 Std. Dev.	最小值 Min	最大值 Max
type1	1,000	.046	.2095899	0	1

解释：即使对于非正态总体（均匀分布）和小样本（$n = 10$），t 检验的第一类错误率（0.046）非常接近名义水平 0.05，表明 t 检验对正态性假设的偏离是稳健的。
Interpretation: Even for a non-normal population (uniform) and small sample ($n =$

10), the Type I error rate of the t-test (0.046) is very close to the nominal level 0.05, indicating that the t-test is robust to deviations from the normality assumption.

3.3 模拟研究：χ^2 检验的稳健性 Simulation Study: Robustness of χ^2 test

χ^2 检验的稳健性模拟 Robustness Simulation of χ^2 test

方法：

- 从均匀分布 $U(0,1)$ 中生成 1000 个样本，样本量分别为 $n = 10$ 和 $n = 30$。
  Generate 1000 samples from a uniform population $U(0,1)$ with sample sizes $n = 10$ and $n = 30$.
- 进行双尾 χ^2 检验 $H_0 : \sigma = 0.288675135$（均匀分布 $U(0,1)$ 的真实标准差），显著性水平 $\alpha = 0.05$。
  Run a two-tail χ^2 test $H_0 : \sigma = 0.288675135$ (the true standard deviation of $U(0,1)$) at 5% significance level.

结果：

样本量 Sample Size	变量 Variable	观测数 Obs	均值 Mean	标准差 Std. Dev.	最小值 Min
$n = 10$	type1	1,000	.007	.0834144	0
$n = 30$	type1	1,000	.002	.044699	0

解释：两种情况下拒绝原假设的比例（0.007 和 0.002）都远低于 5%，因此显著性水平相当不准确。卡方方差检验对正态性假设的偏离不稳健。
Interpretation: In both cases the proportions of the tests rejecting the null are much lower than 5%, hence the significance level is quite inaccurate. χ^2 procedures for variances are not robust when the population is not normal.

3.4 总结与比较 Summary and Comparison

- 对于正态总体，t 检验和方差检验在真实原假设下具有"准确"的第一类错误率。
  For normal populations, the t-test and variance test given a true null hypothesis has the "accurate" Type I error rate at the significance level.
- 没有真实数据是完全正态的。
  No real data is exactly normal.
- 均值的 t 检验对总体正态性的偏离是稳健的，特别是当样本量较大时。

t procedures for means are robust to deviations from the normality of the population, especially when sample sizes are large.

- 方差的 χ^2 检验则不是稳健的。
  χ^2 procedures for variances are not robust.

4 拓展内容：单样本推断的高级主题 Extended Content: Advanced Topics in One-sample Inference

4.1 Bayesian 方法的单样本推断 Bayesian One-sample Inference

贝叶斯视角的单样本推断 Bayesian Perspective on One-sample Inference

与频率派方法的对比 Comparison with Frequentist Approach:

- **频率派 Frequentist:** 参数是固定的未知常数，基于抽样分布进行推断。
  Parameters are fixed unknown constants, inference based on sampling distribution.
- **贝叶斯 Bayesian:** 参数是随机变量，具有先验分布，基于后验分布进行推断。
  Parameters are random variables with prior distributions, inference based on posterior distribution.

贝叶斯单样本均值的推断 Bayesian Inference for One-sample Mean:

- 假设：$X_i \sim N(\mu, \sigma^2)$，其中 σ^2 已知。
  Assume: $X_i \sim N(\mu, \sigma^2)$ with σ^2 known.
- 先验分布：$\mu \sim N(\mu_0, \tau_0^2)$
  Prior distribution: $\mu \sim N(\mu_0, \tau_0^2)$.
- 后验分布：$\mu|\text{data} \sim N(\mu_n, \tau_n^2)$，其中：
  Posterior distribution: $\mu|\text{data} \sim N(\mu_n, \tau_n^2)$, where:

$$\mu_n = \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{x}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}$$

$$\tau_n^2 = \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)^{-1}$$

- 后验均值是先验均值和样本均值的加权平均。
  The posterior mean is a weighted average of the prior mean and the sample mean.

优势与劣势 Advantages and Disadvantages:

- **优势：** 自然地结合先验信息；直接提供参数的概率陈述。
  Advantages: Naturally incorporates prior information; provides direct probability statements about parameters.
- **劣势：** 先验选择的主观性；计算可能复杂。
  Disadvantages: Subjectivity in prior choice; computation can be complex.

4.2 非参数方法：符号检验与 Wilcoxon 符号秩检验 Nonparametric Methods: Sign Test and Wilcoxon Signed-Rank Test

非参数单样本检验 Nonparametric One-sample Tests

适用情况 When to use:

- 数据明显偏离正态分布时
  When data clearly depart from normality
- 样本量很小且无法评估正态性时
  When sample size is very small and normality cannot be assessed
- 数据是序数尺度（如排名）时
  When data are on ordinal scale (e.g., ranks)

符号检验 Sign Test:

- 检验中位数是否等于某个特定值。
  Tests whether the median equals a specified value.
- 只使用观测值与假设中位数之差的符号（正或负）。
  Uses only the sign (positive or negative) of the differences between observations and hypothesized median.
- 检验统计量：正号的数量（或负号的数量）。
  Test statistic: Number of positive signs (or negative signs).
- 在原假设下，正号数量服从二项分布 $Bin(n, 0.5)$。
  Under null hypothesis, number of positive signs follows binomial distribution $Bin(n, 0.5)$.
- **优点：** 简单，对异常值稳健。
  Advantages: Simple, robust to outliers.

- **缺点：**只使用符号信息，效率较低。
  Disadvantages: Uses only sign information, less efficient.

Wilcoxon 符号秩检验 Wilcoxon Signed-Rank Test:

- 检验中位数是否等于某个特定值。
  Tests whether the median equals a specified value.
- 使用观测值与假设中位数之差的符号和大小（通过排名）。
  Uses both the sign and magnitude (via ranks) of the differences.
- 步骤：
  Procedure:
  1. 计算每个观测值与假设中位数的差。
     Compute differences between each observation and hypothesized median.
  2. 取绝对值并排序。
     Take absolute values and rank them.
  3. 将正差值的秩和作为检验统计量。
     Sum ranks of positive differences as test statistic.
- **优点：**比符号检验更有效，因为使用了更多信息。
  Advantages: More efficient than sign test as it uses more information.
- **缺点：**假设分布对称。
  Disadvantages: Assumes symmetric distribution.

4.3 效应量：Cohen's d 与 Hedges' g Effect Size: Cohen's d and Hedges' g

- 统计显著性与实际重要性不同。统计检验告诉我们效应是否存在，而效应量告诉我们效应有多大。
  Statistical significance is different from practical importance. Statistical tests tell us if an effect exists, while effect size tells us how large the effect is.
- **Cohen's d:** 用于单样本 t 检验的效应量度量。
  Cohen's d: Effect size measure for one-sample t-test.

$$d = \frac{\bar{x} - \mu_0}{s}$$

其中 μ_0 是原假设值，s 是样本标准差。
where μ_0 is the null hypothesis value, s is the sample standard deviation.

- **解释 Interpretation:**
  - $|d| \approx 0.2$: 小效应 Small effect
  - $|d| \approx 0.5$: 中等效应 Medium effect
  - $|d| \approx 0.8$: 大效应 Large effect
- **Hedges' g:** Cohen's d 的小样本校正。
  Hedges' g: Small-sample correction for Cohen's d.

$$g = d \times \left(1 - \frac{3}{4n - 9}\right)$$

  其中 n 是样本量。
  where n is the sample size.

- **重要性 Importance:**
  - 避免仅依赖 p 值做出决策。
    Avoid relying solely on p-values for decision making.
  - 在样本量很大时，即使效应很小也可能显著。
    With large sample sizes, even small effects can be significant.
  - 报告置信区间和效应量提供更完整的信息。
    Reporting confidence intervals and effect sizes provides more complete information.

总结 Summary

- **单样本推断 One-sample Inference:**
  - 涉及使用单个样本对总体参数（如均值、方差）进行推断。
    Involves using a single sample to make inferences about population parameters (e.g., mean, variance).
  - 当总体方差未知时，使用 t 分布进行均值的推断。
    When population variance is unknown, use t-distribution for inference about the mean.
  - 当总体为正态分布时，使用卡方分布进行方差的推断。
    When population is normal, use chi-squared distribution for inference about variance.

- **关键概念 Key Concepts:**
  - **标准误 Standard Error:** 样本统计量的标准差，$SE_{\bar{x}} = s/\sqrt{n}$。
    Standard deviation of a sample statistic, $SE_{\bar{x}} = s/\sqrt{n}$.
  - t **统计量** ***t*****-statistic:** $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$，服从 $t(n-1)$ 分布。
    $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$, follows $t(n-1)$ distribution.
  - t **分布** ***t*****-distribution:** 比正态分布尾部更厚，随着自由度增加趋近于正态分布。
    Has fatter tails than normal distribution, approaches normal as degrees of freedom increase.
  - **卡方统计量 Chi-squared statistic:** $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$，用于方差检验。
    $\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$, used for variance tests.

- **置信区间 Confidence Intervals:**
  - 均值：$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$（$\sigma$ 未知）
    Mean: $\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$ (σ unknown)
  - 方差：$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right)$
    Variance: $\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right)$
  - 标准差：取方差区间的平方根。
    Standard deviation: Square root of variance interval.

- **假设检验 Hypothesis Testing:**

  - 均值检验：使用 t 统计量，根据备择假设方向计算 P 值。

    Test for mean: Use t-statistic, calculate p-value based on direction of alternative.

  - 方差检验：使用卡方统计量，通常关注方差是否减小。

    Test for variance: Use chi-squared statistic, often concerned with whether variance has decreased.

  - 决策：比较 P 值与显著性水平 α。

    Decision: Compare p-value with significance level α.

- **条件检查与稳健性 Condition Checking and Robustness:**

  - **随机抽样 Random sampling:** 最重要条件，确保样本代表总体。

    Most important condition, ensures sample represents population.

  - **正态性 Normality:** 可通过图形方法检查；对于 t 检验，随着样本量增加对正态性偏离更稳健。

    Can be checked with graphical methods; for t-test, more robust to deviations from normality as sample size increases.

  - **t 检验的稳健性 Robustness of t-test:** 对小样本非正态性相当稳健，特别是当分布对称时。

    Quite robust to non-normality for small samples, especially when distribution is symmetric.

  - **卡方检验的稳健性 Robustness of χ^2-test:** 对正态性偏离不稳健。

    Not robust to deviations from normality.

- **实用指南 Practical Guidelines:**

  - 样本量 <15：仅在数据接近正态时使用 t 检验。

    Sample size < 15: Use t-test only when data appear normal.

  - 样本量 15-40：可使用 t 检验，除非有异常值或强偏度。

    Sample size 15-40: t-test can be used except with outliers or strong skewness.

  - 样本量 ≥ 40：即使分布明显偏斜，也可使用 t 检验。

    Sample size ≥ 40: t-test can be used even with clearly skewed distributions.

- **分布选择 Distribution Choice:**

情况 Scenario	小样本 Small Sample	大样本 Large Sample
正态，σ 已知	z	z
正态，σ 未知	t	t
非正态，σ 已知	不可用	z
非正态，σ 未知	不可用	t

表 4: 推断分布选择总结 Summary of Inference Distribution Choices

- **高级主题 Advanced Topics:**
  - **贝叶斯推断 Bayesian inference:** 将参数视为随机变量，结合先验信息。
    Treats parameters as random variables, incorporates prior information.
  - **非参数检验 Nonparametric tests:** 当正态性假设严重违反时使用，如符号检验、Wilcoxon 符号秩检验。
    Used when normality assumption is severely violated, e.g., sign test, Wilcoxon signed-rank test.
  - **效应量 Effect size:** 量化效应大小，如 Cohen's d，避免仅依赖统计显著性。
    Quantifies magnitude of effect, e.g., Cohen's d, avoids relying solely on statistical significance.
- **关键公式 Key Formulas:**
  - 标准误：$SE_{\bar{x}} = \frac{s}{\sqrt{n}}$
  - t 统计量：$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$
  - 均值置信区间：$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$
  - 卡方统计量：$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$
  - 方差置信区间：$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right)$

本章核心要点 Core Takeaways

- **理解 t 分布的重要性 Understand the importance of t-distribution:** 当总体方差未知时，t 分布提供了比正态分布更准确的推断。
  Provides more accurate inference than normal distribution when population variance is unknown.
- **正确选择检验分布 Choose the correct test distribution appropriately:** 根据样本量、正态性假设和是否知道方差来选择 z 检验或 t 检验。
  Choose between z-test and t-test based on sample size, normality assumption,

and whether variance is known.

- **检查假设条件 Check assumptions:** 特别是随机抽样和正态性（对于小样本）。

  Especially random sampling and normality (for small samples).

- **理解稳健性概念 Understand robustness:** t 检验对正态性偏离相对稳健，而卡方方差检验则不稳健。

  t-test is relatively robust to deviations from normality, while chi-squared variance test is not.

- **解释置信区间和 P 值 Interpret confidence intervals and p-values correctly:** 置信区间提供估计范围，P 值提供反对原假设的证据强度。

  Confidence intervals provide range of estimates, p-values provide strength of evidence against null hypothesis.

- **考虑效应量 Consider effect size:** 统计显著不等于实际重要，报告效应量提供更完整的信息。

  Statistical significance does not equal practical importance, reporting effect size provides more complete information.

- **了解替代方法 Know alternative methods:** 当正态性假设严重违反时，考虑非参数方法。

  Consider nonparametric methods when normality assumption is severely violated.

- **应用于实际问题 Apply to practical problems:** 单样本推断在质量控制、金融风险评估、科学研究等领域有广泛应用。

  One-sample inference has wide applications in quality control, financial risk assessment, scientific research, etc.