Chaper 6 - Confidence Intervals

Daxiang Na（那达翔）

2022-10-06

Contents

Q1: For sampling distribution, it can have sampling size n and sampling time m, what determines if it follows CLT? n or m? If we sample for 1 time and 100 times, each time with same size n, does that makes a difference? Another way to ask this question: if we sample for 20 times ($m = 3$), but each time with sample size $n = 1000$, does that follow CLT?

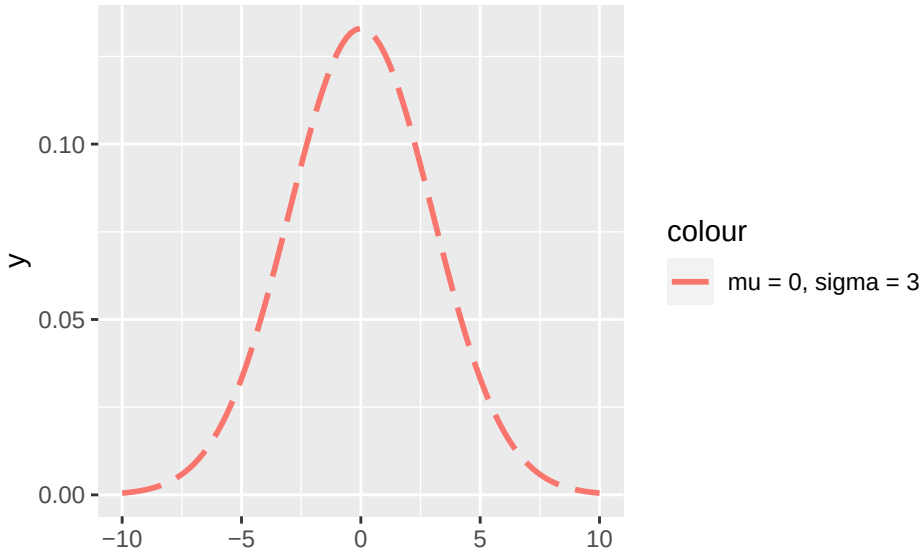
1 CLT and Sampling Distribution

2 Margin of Error and Sample Size

Estimated sample size needed when confidence interval (CI) is given:

$$n = \left\lceil \frac{z_{\alpha/2}^2 \cdot \sigma^2}{m^2} \right\rceil$$

3 Normal Distribution Curve with ggplot 正态分布曲线

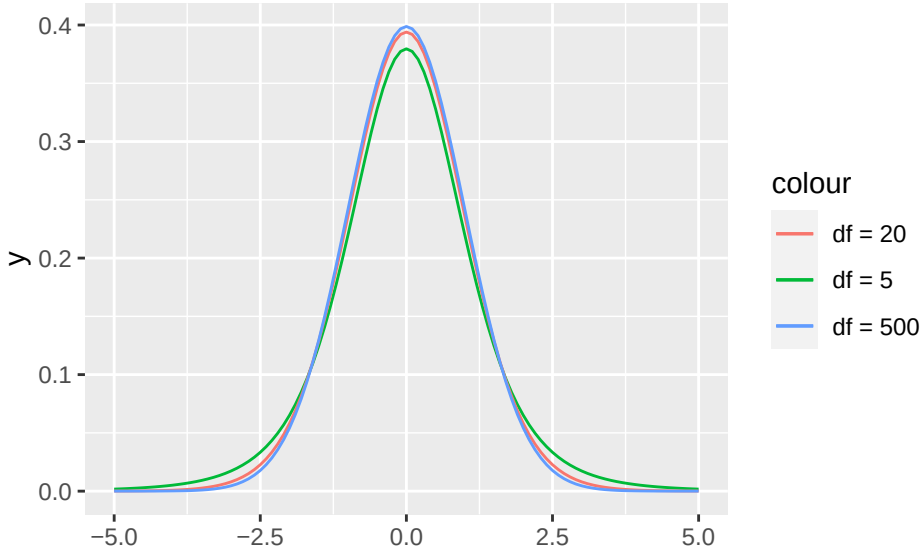


4 T distribution T 分布

When σ^2 is also unknown, we substitute the sample variance s^2 and use the t distribution instead of the normal distribution.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

t distribution curve with ggplot, T 分布曲线



This t-statistic has a t distribution with n − 1 degrees of freedom

p-values for t-tests

- We calculate our p-value as follows, for each of the three types of tests (*t-tests*):
- One-sided, lower-tailed hypothesis ($H_1 : \mu < \mu_0$):
  - `pt(t,df)`
- One-sided, upper-tailed hypothesis ($H_1 : \mu > \mu_0$):
  - `1-pt(t,df)`
- Two-sided hypothesis ($H_1 : \mu \neq \mu_0$):
  - If $z \leq 0$: `2*pt(t,df)`
  - If $z > 0$: `2*(1-pt(t,df))`