Statistical Methodology for Software Engineers

tutorial 3

Hadas Lapid, PhD

Contents

1. Z distribution – confidence interval

- Create a random normal sample of 500 under the assumption of μ=20 and σ=5.
- Show the sample distribution with 25 bins
- Build the confidence interval at 93% confidence for the ***population*** mean and given population standard deviation, σ.

- Reminder: $C.I. = \bar{x} \pm Z_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$
- Hint: what is the size of α?

2. Z distribution realization

- Under normal distribution assumption
- Researchers of early childhood development investigated the number of words in the vocabularies of young children. They found that:

Age	**Mean**	**Standard deviation**
18 months	μ=85 words	σ=80 words
24 months	μ=275 words	σ=120 words

Two children showed the following vocabularies:
X1(18m) = 93 words
X2(24m) = 287 words
Which of them has larger vocabulary, relative to their age?
Show it by calculating their respective z statistics
Show it by calculating their respective p-values (pnorm)

Hint: $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

3. t distribution

Calculate the **critical t value** for constructing the expectation value around a mean at 98% confidence for a sample with n=15 observations.

Hint: use qt()
Hint2: you are not asked to calculate the C.I, only t

4. t distribution

- A sample of 18 observations has s.d. of 1.1
- It is hypothesized that the sample comes from a population with mean 4.5 and unknown population standard deviation.
- Find the critical X needed for deciding whether the sample mean rises from the hypothesized population in 98% confidence.

Recall: when σ is unknown replace t with z and s with σ

$$\bar{X}_c = \mu_0 + t_{1-\alpha}(df) \cdot \frac{s}{\sqrt{n}}$$

5. Central limit theorem

Create a random uniform sample of 3000 observations. (use runif() function between 0 and 100 and data.frame())
Using the central limit theorem, find the approximated population mean.

<u>To do this:</u>

- Iteratively pick 30 observations over 100 iterations (sample(nrow(data)))
- Calculate the mean of the sampled observations and update a vector of means with this sample mean
- When done Show vector of means histogram
- Calculate the mean of the means vector