t-Distribution

CSU, Chico Math 314

2018-10-12

outline

outline

Recap: Confidence Intervals

Confidence intervals and hypothesis tests (as of yet) require knowledge of σ

$$\bar{X} \pm z^* \cdot \sigma_{\bar{X}}, \quad \textit{and} \qquad z = \frac{\bar{X} - \mu_0}{\sigma_{\bar{X}}}.$$

outline

Confidence Intervals, from now on

Confidence intervals and hypothesis tests (from now on) do not require knowledge of σ

$$\bar{X} \pm t^*_{df} \cdot s_{\bar{X}}, \quad \textit{and} \qquad t_{df} = \frac{\bar{X} - \mu_0}{s_{\bar{X}}}.$$

outline

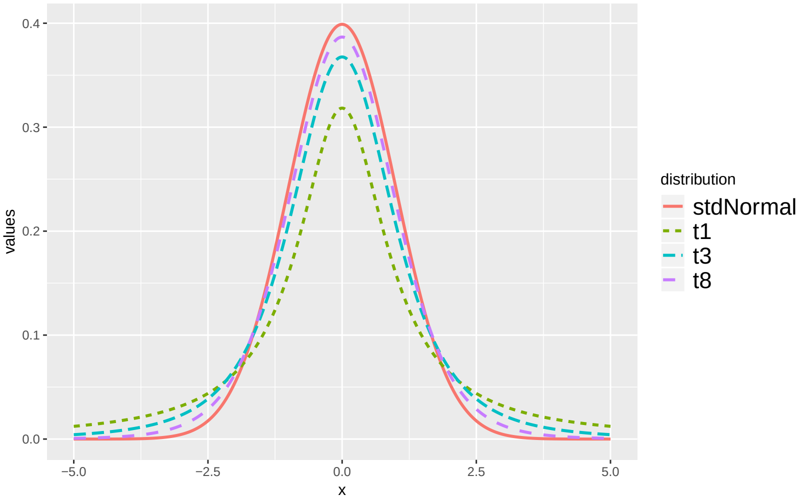
t-distribution, definition

The t-**distribution** is the name of a new distribution function, which accounts for the added uncertainty of having to estimate σ in the process of estimating μ.

t-distribution

The t-distribution is a (standard) normal-looking probability density function that is always centered at zero, has fatter tails than the normal distribution, and has a single parameter, **degrees of freedom**.

t, plot



0.4
0.3
0.2
0.1
0.0
values
−5.0
−2.5
0.0
2.5
5.0
x
distribution
stdNormal
t1
t3
t8

t, degrees of freedom

degrees of freedom

The degrees of freedom describe the heaviness of the tails of the t-distribution. The larger the degrees of freedom, the more closely the distribution looks like a standard normal, $N(0, 1)$.

t, df for sample mean

If the sample has n observations and we are examining a single mean, then we use the t-distribution with $df = n - 1$ degrees of freedom.

t, notes

Notes on the t-distribution

- watch for skew
- heavy tails of t account for estimation of σ
- must standardize random variable of interest

t, in R

The function in R to deal with the t-distribution follow the same pattern

```
?pt
```

outline

t, confidence intervals

Confidence intervals under the t-distribution are nearly identical,

$$\bar{X} \pm t^*_{df} \cdot s_{\bar{X}}.$$

t, confidence interval example

To investigate the average mercury content in dolphin muscle, we collect a sample of 19 dolphins from the Taiji area in Japan. We calculate a sample mean of $4.4\mu g$ of mercury per gram of muscle with a sample standard deviation of 2.3. Make a 95% confidence interval of the mercury content found in Taiji dolphins.

t, CI example

Put the pieces of the puzzle together and interpret.

```r
xbar <- 4.4  # mean(x)
n <- 19      # ifelse(anyNA(x), sum(!is.na(x)), length(x))
s <- 2.3     # sd(x)
t <- qt(c(0.025, 0.975), n-1)
xbar + t*s/sqrt(n)

## [1] 3.291435 5.508565
```

t, CI example

We are 95% confident the poulation average mercury content of muscles in Taiji area dolphins is between 3.29 and $5.51\mu g/g$.

t, hypothesis tests

Hypothesis tests under the t-distribution are nearly identical,

$$t_{df} = \frac{\bar{X} - \mu_0}{s_{\bar{X}}}.$$

t, HT example

The population mean running time for The Cherry Blossom Race in 2006 was 93.29 minutes. We want to test whether or not participants in the 2012 Cherry Blossom Race are getting faster or slower, versus the other possibility that there has been no change. The sample mean and sample standard deviation of the sample of 100 runners from the 2012 Cherry Blossom Race are 95.61 and 15.78 minutes, respectively. Set up and evaluate the appropriate hypothesis test using the t-distribution, and compare your answer to the normal distribution.

t, HT example

First set up the null and alternative hypotheses and choose a level of significance.
We test

$$H_0: \quad \mu_{2012} = 93.29$$
$$H_A: \quad \mu_{2012} \neq 93.29$$

with $\alpha = 0.05$.

t, HT example

Hypothesis testing follows the same framework. We calculate the test statistic, now calling it t because we will use the t-distribution.

```
n <- 100
t <- (95.61 - 93.29)/(15.78/sqrt(n))
```

t, HT example

Calculate p-value from test statistics under both t and normal distribution.

```
2*(1-pt(abs(t), n-1))

## [1] 0.1446745

2*(1-pnorm(abs(t))) # only for comparison

## [1] 0.1415034
```

t-distribution, take away

- When σ unknown, replace Z with t-distribution.
  - watch for skew.
- single mean degrees of freedom, $df = n - 1$
- large degrees of freedom makes t-distribution look like standard normal
- must standardize random variables with t
- safe bet: use t-distribution instead of Z everywhere!

outline

references I

David M Diez, Christopher D Barr, and Mine Cetinkaya-Rundel. *OpenIntro Statistics*. CreateSpace independent publishing platform, third edition, 2015.

Hadley Wickham. *ggplot2: elegant graphics for data analysis*. Springer Science and Business Media, 2009.