Statistics

Week 4 Recitation

ESD, SUTD

Term 5, 2017

Exercise 1: Confidence Interval

You are given 130 body temperatures (in °F) in the *Excel* spreadsheet.

1. Are the body temperatures normally distributed? (Q-Q plot.)
2. Find the 99% two-sided confidence interval for the true mean body temperature. (t-distribution or z-distribution?)
3. For this population, is the benchmark temperature of 98.6°F appropriate as the mean?

Reminder

$$\left[\bar{X} - Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \tag{1}$$

$$\left[\bar{X} - t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}, \bar{X} + t_{n-1,1-\alpha/2}\frac{s}{\sqrt{n}}\right] \tag{2}$$

Do the exercise in excel first. If you have spare time, do it in R.

Exercise 1: Confidence interval (Discussion)

Use t-distribution. But z-distribution is a good approximation.

- Unknown population variance σ^2
- Normal distribution
- Large sample size

Hypothesis testing

- With confidence interval
- With p-value

Normality Test:

H_0: the data are normally distributed
H_1: the data are not normally distributed

Exercise 1: Confidence interval (Discussion)

t-distribution and standard normal distribution

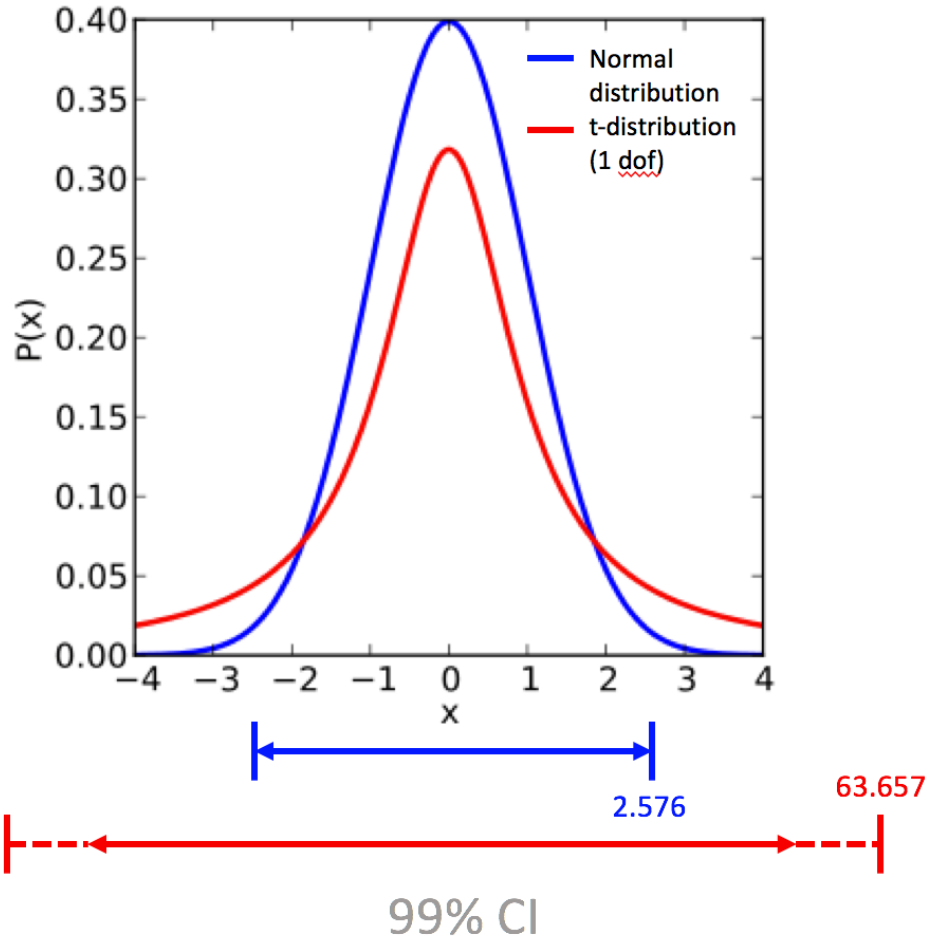
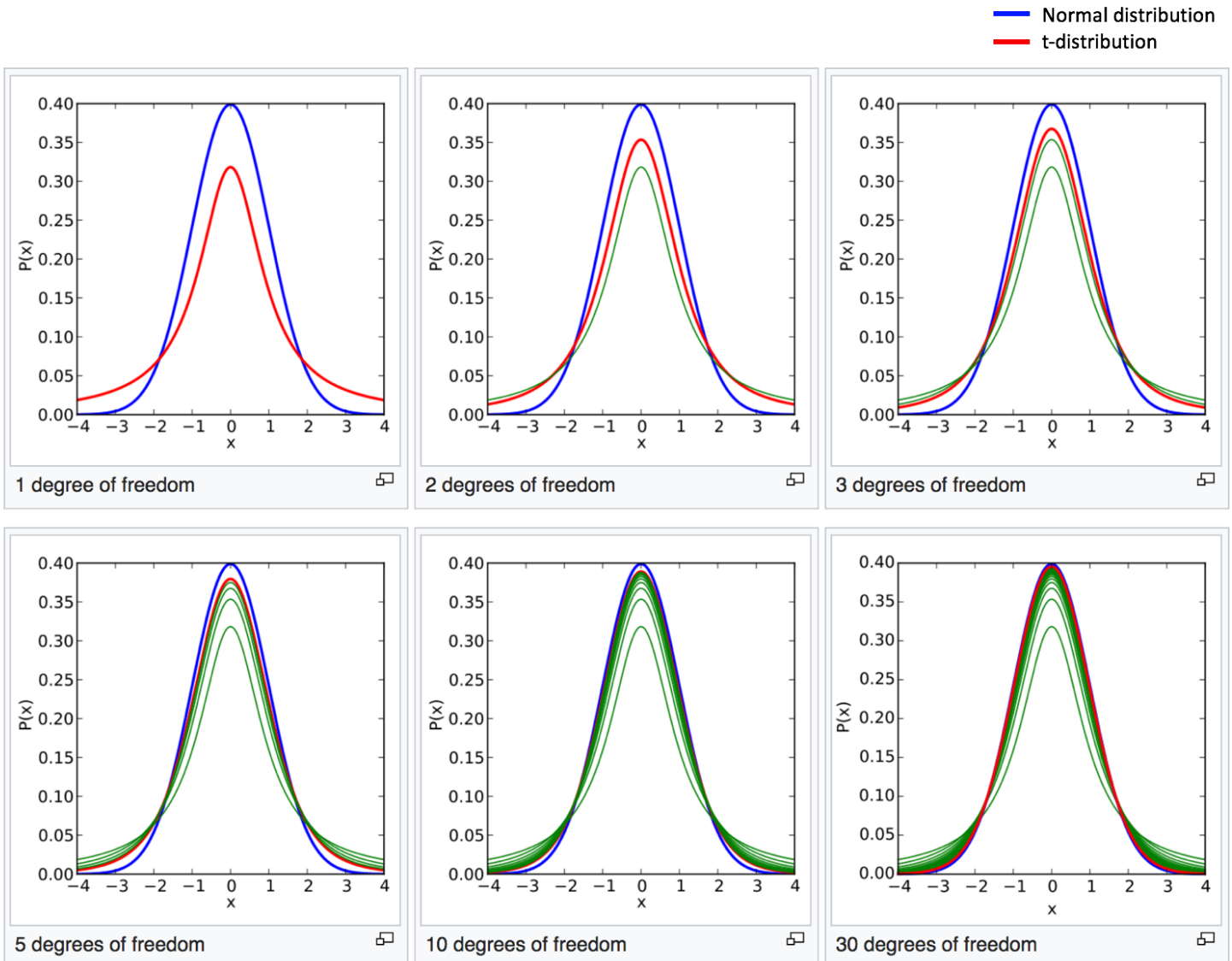


Image ref: Wikipedia Student's t-distribution

Exercise 1: Confidence interval (Discussion)

t-distribution approaches to normal distribution when dof. -> ∞



Exercise 2: One-Sided CI and Hypothesis Testing

A new training technique is believed to improve running times. After a month of training with this technique, six runners from a team recorded times of 50.1, 50.3, 50.3, 51.2, 51.5, 51.6 (s).

1. Assume that the running times are normally distributed. Find the 99% upper bound for the average running time after the training.

2. Is this significantly lower than the team average before the technique was introduced, $52s$? Perform a hypothesis test at the 1% significance level.

3. What if the old average was $51.7s$?

Some ideas for the project

Analyze existing data:

- Descriptive statistics
- Data visualization: side-by-side box plots, histogram, and etc.
- Q-Q plots to check if data fits distribution (normal, exponential, . . .)
- Forecasting and smoothening using moving average
- Confidence intervals, hypothesis testing
- Later in the term: curve fitting (regression), etc.

Some ideas for the project

Collect new data - design a survey

A survey should:

- Be voluntary and anonymous
- Contain a clear statement of its purpose
- State how the data will be used, and whether it will be destroyed or retained afterwards

If you plan to conduct a formal survey for the project, check with us first.