# F19 STA 100 A01 Discussion 07

*Yishan Huang*

*2019/11/12*

Discussion Time: Tuesday 8:00 – 8:50 am, Haring Hall 1204.

Notes: https://github.com/Hahahuo-13316/sta100-a01-fall19

Office hour: Tuesday 12:00 – 1:00 pm, Mathematical Sciences Building 1117.

Email: yishuang@ucdavis.edu

Quiz: This Friday.

## Confidence intervals

### Confidence interval for population mean $\mu$

- Assume a population that is normally distributed with mean $\mu$ and standard deviation, $\sigma$, which are both unknown. Take a population with sample size $n$ as $Y_1, Y_2, \ldots, Y_n$, then we know that

  $$\frac{\bar{Y}-\mu}{s/\sqrt{n}} \sim t_{n-1},$$

  where $s$ is the estimate of $\sigma$

  $$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$

  and the denominator $s/\sqrt{n}$ is the standard error of the sample mean $\mu$. And, $t_{n-1}$ is the Student's $t$-distribution with $(n-1)$ degrees of freedom.

- A confidence interval of $\mu$ with **confidence level** $(1-\alpha)$ is an interval calculated by the sample that, when the sample is randomly taken from the population, the probability that $\mu$ lies in the interval is $(1-\alpha)$.

- How to find a confidence interval? We know that

  $$\mathbb{P}\left(-t_{n-1}^{(\alpha/2)} \leq \left|\frac{\bar{Y}-\mu}{s/\sqrt{n}}\right| \leq t_{n-1}^{(\alpha/2)}\right) = 1-\alpha.$$

  That is,

  $$\mathbb{P}\left(\mu \text{ within } \bar{Y} \pm t_{n-1}^{(\alpha/2)} \cdot \frac{s}{\sqrt{n}}\right) = 1-\alpha.$$

  Hence, the $(1-\alpha)$ confidence interval of $\mu$ is

  $$\left[\bar{Y} - t_{n-1}^{(\alpha/2)} \cdot \frac{s}{\sqrt{n}},\ \bar{Y} + t_{n-1}^{(\alpha/2)} \cdot \frac{s}{\sqrt{n}}\right].$$

- For large sample ($n \geq 30$), we can substitute the student $t$-distribution with $(n-1)$ degrees of freedom, by the standard normal distribution. That is, the interval is like

  $$\left[\bar{Y} - z^{(\alpha/2)} \cdot \frac{s}{\sqrt{n}},\ \bar{Y} + z^{(\alpha/2)} \cdot \frac{s}{\sqrt{n}}\right].$$

- How to interpret the confidence interval? We are $(1-\alpha)\cdot 100\%$ percent confident that the mean of the population $\mu$ is between $[a,b]$ (if $[a,b]$ is the confidence interval that you have calculated). We should notice that $(1-\alpha)\cdot 100\%$ **is not the probability that** $[a,b]$ **contains** $\mu$, because when the sample fixed, the interval is also fixed and hence has no more randomness.
- In this course, unless further announced, the population follows normal distribution so we could always use the formula above.
- If we need some specified precision, we can use the form above to solve $n$. For example, if we need to find the smallest $n$ such that the length of $(1-\alpha)$ confidence interval of $\mu$ is less than 0.2, then we can do
$$2\cdot z^{(\alpha/2)}\cdot\frac{s}{\sqrt{n}}\leq 0.2,\quad \text{i.e., } n\geq(10s\cdot z^{(\alpha/2)})^2.$$

**A special case: confidence interval of a proportion**

- Let $p$ be that proportion, then it is a special case where all the $Y_i$ can only take value within $\{0,1\}$. If the sample size is large, then although the population is not normally distributed, we could still assume that and hence the $(1-\alpha)$ confidence interval is
$$\hat{p}\pm z^{(\alpha/2)}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$
Where $\hat{p}$ is the proportion in that sample.

**Confidence interval for difference**

- The standard error of $\bar{Y_1}-\bar{Y_2}$ is
$$\mathrm{SE}_{\bar{Y_1}-\bar{Y_2}}=\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}},$$
and the $(1-\alpha)$ confidence interval of $\mu_1-\mu_2$ is
$$(\bar{Y_1}-\bar{Y_2})\pm z^{(\alpha/2)}\cdot\mathrm{SE}_{\bar{Y_1}-\bar{Y_2}}.$$