Stat 213: Intro to Statistics 10
Small Sample Inference

H. Kim

Fall 2007

recall

- Significant tests and confidence intervals for μ are based on

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

  - when the population is normal
  - when the population is non-normal and $n \geq 30$
  - when σ unknown, and $n \geq 30$

- What if the population is normal, but n is not large with unknown σ,

  What is the sampling distribution of

$$\frac{\overline{X} - \mu}{\hat{\sigma}/\sqrt{n}} \quad ?$$

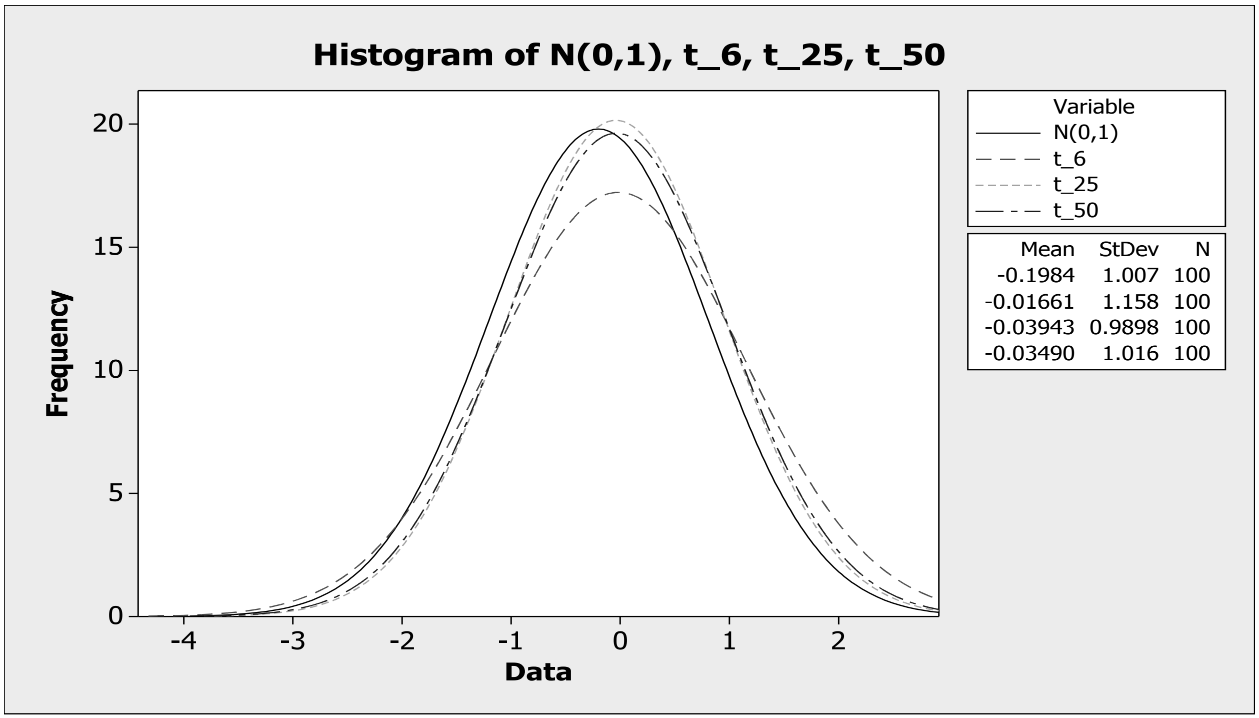
Gosset's t (Student's t)

- The model is always bell-shaped, but the details change with different sample sizes due to the estimated standard error, $se(\bar{X}) = \frac{\hat{\sigma}}{\sqrt{n}}$, the shape of the sampling model changed
- The Student's t-models form a whole family of related distributions that depend on a parameter known as `degree of freedom` (d.f.)
- When conditions are met, the `standardized sample mean`

$$t_{n-1} = \frac{\overline{X} - \mu}{\hat{\sigma}/\sqrt{n}}$$

follows a Student's t model with $(n-1)$ d.f.

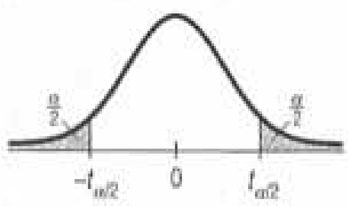
Normal and t- distributions



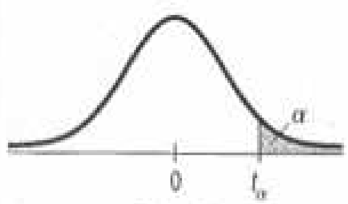
t table

- $t_{(5,0.05)} = 2.571$, $t_{(19,0.05)} = 2.093$

Two tail probability		0.20	0.10	0.05	0.02	0.01	
One tail probability		0.10	0.05	0.025	0.01	0.005	
Table T	df						df
Values of t_α	1	3.078	6.314	12.706	31.821	63.657	1
	2	1.886	2.920	4.303	6.965	9.925	2
	3	1.638	2.353	3.182	4.541	5.841	3
	4	1.533	2.132	2.776	3.747	4.604	4
	5	1.476	2.015	2.571	3.365	4.032	5
	6	1.440	1.943	2.447	3.143	3.707	6
	7	1.415	1.895	2.365	2.998	3.499	7
	8	1.397	1.860	2.306	2.896	3.355	8
	9	1.383	1.833	2.262	2.821	3.250	9
	10	1.372	1.812	2.228	2.764	3.169	10
	11	1.363	1.796	2.201	2.718	3.106	11
	12	1.356	1.782	2.179	2.681	3.055	12
	13	1.350	1.771	2.160	2.650	3.012	13
	14	1.345	1.761	2.145	2.624	2.977	14
	15	1.341	1.753	2.131	2.602	2.947	15
	16	1.337	1.746	2.120	2.583	2.921	16
	17	1.333	1.740	2.110	2.567	2.898	17
	18	1.330	1.734	2.101	2.552	2.878	18
	19	1.328	1.729	2.093	2.539	2.861	19



Two tails



One tail

summary

- If X: Normal with known μ and σ^2,

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}$$

- If X: non-Normal with either known or unknown μ and σ^2 and $n > 30$, by CLT

$$\frac{\overline{X} - \mu}{\hat{\sigma}/\sqrt{n}} \sim \mathcal{N}$$

- If X: Normal with unknown σ^2 and $n < 30$,

$$\frac{\overline{X} - \mu}{\hat{\sigma}/\sqrt{n}} \sim t_{n-1}$$

- If X: non-Normal with either known or unknown σ^2 and $n < 30$,

???

small sample inference: CI for μ

- $(1-\alpha)100\%$ confidence interval for μ

$$\left(\overline{X} - t_{n-1,\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}, \quad \overline{X} + t_{n-1,\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}}\right)$$

- example:

  A random sample is 38, 38, 40, 45, 28, 30, 27, 41, 29, 26

  - 95% confidence interval for μ

  $\overline{X} = 34.2$, $\hat{\sigma}^2 = 47.524$, $n = 10$, $\frac{\hat{\sigma}}{\sqrt{10}} = 2.18$

small sample inference: testing $H_0 : \mu = \mu_0$

- the sample is randomly selected from a normal distributed population
- $H_0 : \mu = \mu_0$ vs $H_a : \mu > \mu_0, \quad \mu < \mu_0, \quad \mu \neq \mu_0$
- test statistic:

$$t_{n-1} = \frac{\overline{X} - \mu_0}{\hat{\sigma}/\sqrt{n}}$$

- rejection regions
  - rejection regions for $H_0 : \mu > \mu_0 \quad \{t_{n-1} > t_{n-1,\alpha}\}$
  - rejection regions for $H_0 : \mu < \mu_0 \quad \{t_{n-1} < -t_{n-1,\alpha}\}$
  - rejection regions for $H_0 : \mu \neq \mu_0$

$$\{t_{n-1} < -t_{n-1,\alpha/2} \ \text{ or } \ t_{n-1} > t_{n-1,\alpha/2}\}$$

- conclusion

example

- $H_0 : \mu = 37$ vs $H_a : \mu \neq 37$

  $\overline{X} = 34.2$, $\hat{\sigma}^2 = 47.524$, $n = 10$, $\frac{\hat{\sigma}}{\sqrt{10}} = 2.18$

example

- A manufactory process produced ball bearings. It is known that the diameter of ball bearing is normally distributed. Based on a random sample of 9 observations with the mean of 10.02 3 mm and standard deviation of 0.12 mm. Find a 99% confidence interval for the population mean diameter of ball bearing produced by this process.

example

- Body temperatures of 25 intertidal crabs placed in air at 24.5^oC yielded a mean of 24.71^oC with standard deviation of 1.34^oC. Test at $\alpha = 0.05$ if the mean body temperature for this species of crab differs from the ambient air temperature. Assume that the body temperature for this species of crab is normally distributed.

review: sample size

- we should have some idea of how large a margin of error we need to be able to draw conclusions or detect a different we want to see
  1. determine the `standard error` (se) of the point estimator
  2. choose a `margin of error` (K) and a `confidence coefficient` $(1-\alpha)$
  3. solve $z_{1-\alpha/2} se(\overline{X}) = K$ for n
     - for the mean, μ: $z_{1-\alpha/2}\frac{\hat{\sigma}}{\sqrt{n}} = K$
     - for the proportion, p: $z_{1-\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = K$

review: example

- The standard deviation of the population is known as 5.4711, and the margin of error is equal to 1.8. What is the minimum sample size at the 90 % level of confidence ?
  - $z_{0.95} = 1.645, \sigma = 5.4711, K = 1.8$
  - $n = 24.9998 \approx 25$

$$\begin{aligned} t_{24,0.1/2} \times \frac{5.4711}{\sqrt{n}} &= 1.8 \\ 1.711 \times \frac{5.4711}{\sqrt{n}} &= 1.8 \end{aligned}$$

$n = 27.046 \approx 28$