MATH 124 Spring 2005

Lecture: 22
Date: May 12, 2005

Skills you should acquire from this lecture:

- Concept of a Standard Error (SE)
- Statistical inference for population mean when population standard deviation is not known.
- How to use the t-distribution table for confidence intervals.
- How to put bounds on a P-value using the t-distribution table.

Related readings in the textbook:

- Section 7.1

We have talked about confidence intervals and hypothesis testing for μ when σ was known. Today we address the situation when σ is unknown.

Standard error

When we estimate the standard deviation of a statistic using data the result is called the <u>standard error</u> of the statistic.

eg	Standard deviation	Standard error
$\bar{X}$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$SE_{\bar{X}} = \frac{s}{\sqrt{n}}$ ← sample standard deviation
$\hat{p}$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

eg Suppose we have a sample of data

5.1, 3.2, 4.5, 4.9, 5.7

$$\bar{X} = \frac{5.1 + 3.2 + 4.5 + 4.9 + 5.7}{5} = 4.68$$

$s = ?$ $\quad \Sigma X_i^2 = 113$

$$s = \sqrt{\frac{113 - 5(4.68)^2}{5-1}} = 0.9338 \text{ (4dp)}$$

so what is $SE_{\bar{X}}$?

$$SE_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{.9338}{\sqrt{5}} = .4176 \text{ (4dp)}$$

Inference in this situation (for μ from $N(\mu, \sigma)$ with σ unknown)

Recall that $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ which has $N(0,1)$ distribution is the basis of

inference when $\sigma/\sqrt{n}$ is known.

what happens if we replace $\sigma/\sqrt{n}$ with the SE?

Call this

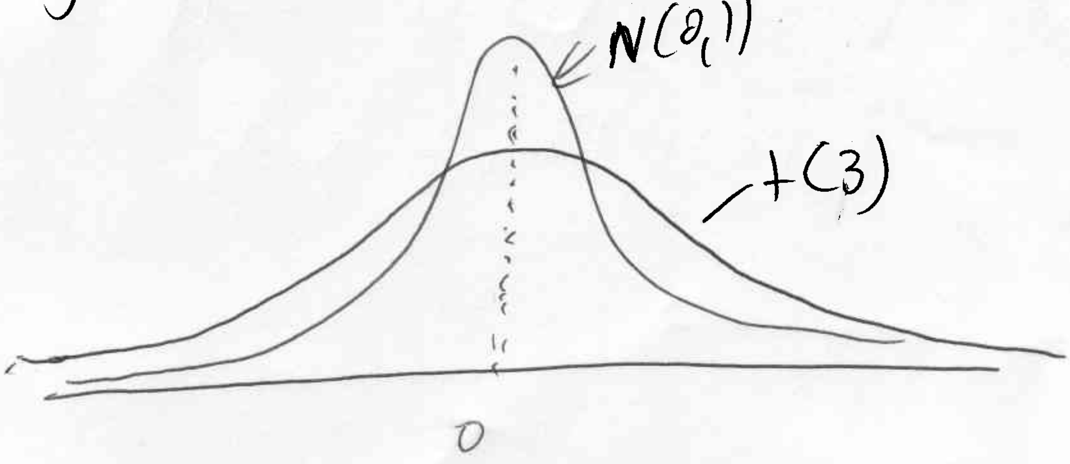
$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

this statistic does not have ~~the~~ standard normal distribution. Instead it ~~has~~ is distributed according to something called the "t distribution" with $n-1$ "degrees of freedom".

what is ~~this~~ t-distribution?

- Symmetric bell shaped curve centered at 0
- has only one parameter k which we call the degrees of freedom. ($k = 1, 2, 3, \ldots$)
- is more spread out then the standard normal curve
- gets less spread out as k increases
- when k approaches infinity the distribution approaches the $N(0,1)$ curve

eg



There is a table in the back of the book that allows us to look up areas under the t-distribution.

Confidence interval for μ (when σ unknown)

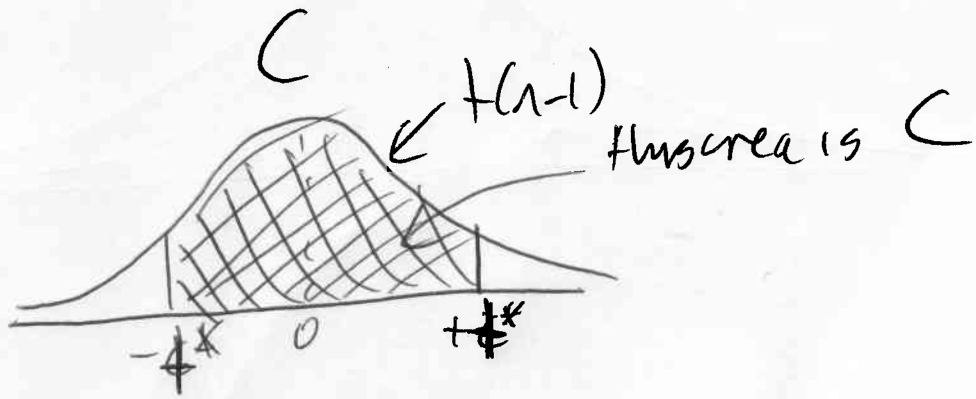
A level C confidence interval for μ is given by

$$\bar{x} \pm t^* \frac{s}{\sqrt{n}}$$

where t^* is the value from the $t(n-1)$ density curve with area C between $-t^*$ and t^*

ie

$$P(-t^* < T < t^*) = C$$



Using our data from before what is a 95% CI for μ? (Assume it is from a normal distribution)

Recall $\bar{x} = 4.68$ $SE_{\bar{x}} = .4176$ so

the 95% CI is given by

$$4.68 \pm 2.776\,(.4176)$$

↑ this number is from t distribution

$$4.68 \pm 1.1592$$

So 95% CI for μ is

(3.52, 5.84).

Hypothesis tests for μ (when σ is unknown)

We proceed here in much the same way as we did when we knew σ the only difference is in how we compute the test statistic and which distribution we use

$H_0: \mu = \mu_0$

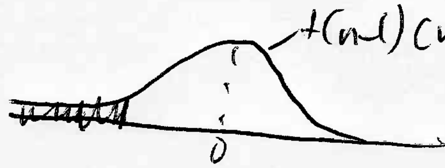
$H_A: \mu \neq \mu_0$

test statistic.

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

has t distribution with $n-1$ degrees of freedom

Pvalues come from the t-distribution table.

Alternative	P-value	
$H_A: \mu \neq \mu_0$	$2P(T > \lvert t \rvert)$	
$H_A: \mu > \mu_0$	$P(T > t)$	
$H_A: \mu < \mu_0$	$P(T < t)$	

Example

Suppose we want to test whether or not the population from which we sampled our data has mean of 5 then

$$H_0: \mu = 5$$

$$H_A: \mu \neq 5$$

The test statistic would be

$$t = \frac{4.68 - 5}{.9338/\sqrt{5}}$$ which has t-distribution with $5-1=4$ df

$$= -.7663 \text{ (4dp)}$$

$$\text{Pvalue} = 2P(T > |-.7663|)$$
$$= 2P(T > .7663)$$

Since we cannot look this up exactly instead we put bounds on it. From t table df=4 line we find

$$.741 < .7663 < .941$$

$$.25 > P(T > .7663) > .20$$ ← converting into upper tail probabilities

$$.5 > 2P(T > .7663) > .4$$ ← The p-value we want

Since the Pvalue is between .4 and .5 we have no evidence to say that the mean is not 5.