

# Lecture 26

(1)

We have talked about confidence intervals and hypothesis testing for  $\mu$  when  $\sigma$  was known. Today we address the situation when  $\sigma$  is unknown.

## Standard error

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

eg

Standard deviation

Standard error

$\bar{X}$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

$$SE_{\bar{X}} = \frac{S}{\sqrt{n}} \leftarrow \begin{array}{l} \text{sample} \\ \text{standard} \\ \text{deviation} \end{array}$$

$\hat{p}$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

$$SE_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

eg Suppose we have a sample of data

5.1, 3.2, 4.5, 4.9, 5.7

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

$$s = \sqrt{\sum x_i^2} = 113$$

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

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

$$SE_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{0.9338}{\sqrt{5}} = 0.4176 \text{ (4dp)}$$

Inference in this situation (for  $\mu$  from  $N(\mu, \sigma)$  with  $\sigma$  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? (3)

Call this

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

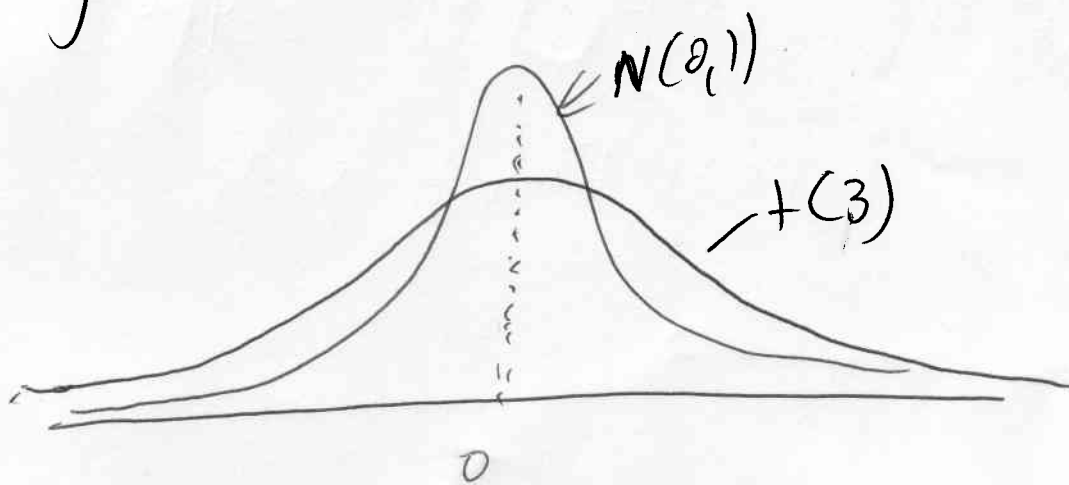
this statistic does not have the standard normal distribution. Instead it ~~is~~ is distributed according to something called the "t distribution" with  $n-1$  "degrees of freedom".

what is this t-distribution?

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

eg

(4)



There is a table in the back of the book that allows us to look up areas under the  $t$ -distribution. We will discuss how to use it at the problem solving session on Friday.

Confidence interval for  $\mu$  (when  $\sigma$  unknown)

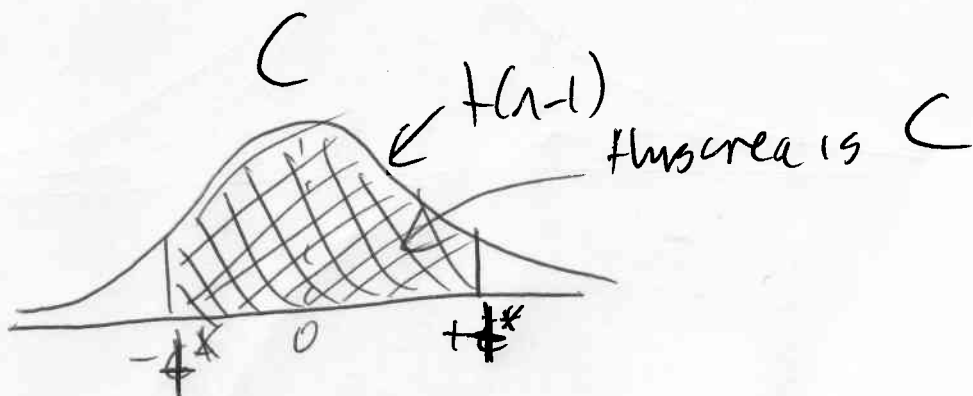
A level  $C$  confidence interval for  $\mu$  is given by

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

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

12

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



using our data from before what is a 95% CI for  $\mu$ ? (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  $\mu$  is

$$(3.52, 5.84).$$

# Hypothesis tests for $\mu$ (when $\sigma$ is unknown)

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

$$H_0: \mu = \mu_0$$

$$H_A: \mu \neq \mu_0$$

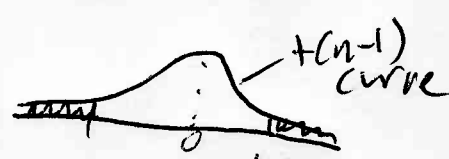
test statistic.

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

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

P-values come from the t-distribution table.

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



$H_A: \mu > \mu_0$	$P(T > t)$
--------------------	------------



$H_A: \mu < \mu_0$	$P(T < t)$
--------------------	------------

