MATH1019 Linear Algebra and Statistics for Engineers

Lecture 4: Estimation & Hypothesis Testing

Overview: In Lecture 3 we learned how to estimate population parameters by using sample information. Now we will consider the second formal manner of making inferences from a sample to population using *tests of hypotheses*.

Motivation: Hypothesis testing arises as a natural consequence of the scientific method. The scientist observes nature or a process, formulates a theory, and then tests theory against observation. The experimenter theorises that a population parameter takes on a certain value. A sample is then selected and the observation is compared with theory (hypothesis). A decision is then made whether to reject or accept the theory depending on how compatible the observations are with the theory.

Learning outcomes

In today's lecture we will learn how to:

- Determine sample size for a given accuracy
- Determine confidence interval when σ is unknown
- Set up and evaluate tests of hypothesis

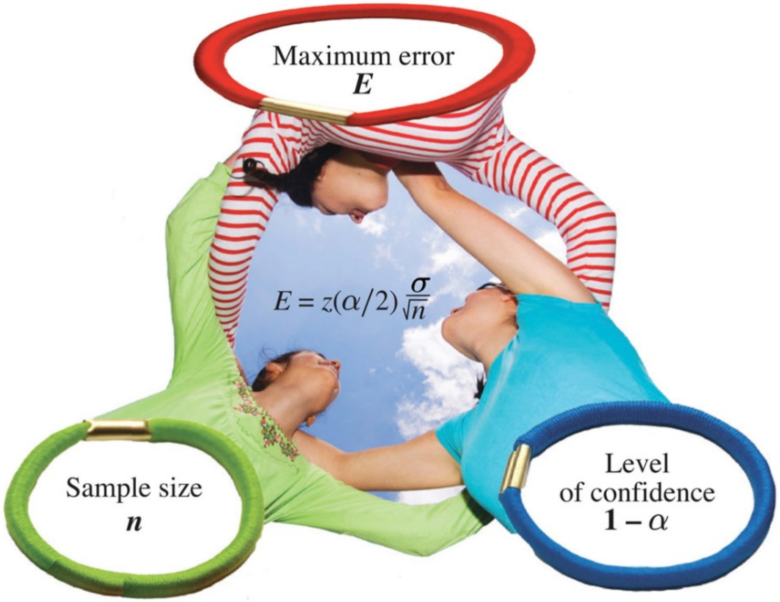
Selecting Sample Size

- You can select the sample size that will guarantee a desired confidence level for a fixed margin of error.
- Let *E* be the margin of error

$$E = \frac{z_{\alpha/2}\sigma}{\sqrt{n}} \quad \Rightarrow \quad \sqrt{n} = \frac{z_{\alpha/2}\sigma}{E}$$

$$n = \left(\frac{z_{\alpha/2}\sigma}{E}\right)^2$$

- **Always round UP**



Example

To assess the accuracy of a laboratory scale a standard weight is weighed repeatedly. The scale readings are normally distributed with unknown mean. The standard deviation is known to be 0.0002gm. How many measurements must be made to get a margin of error of ±0.0001 with 98% confidence?

$$n = \left(\frac{z_{\alpha/2}\sigma}{E} \right)^2 = \left(\frac{2.326(0.0002)}{0.0001} \right)^2 = 21.64$$

So we choose n = 22

Tests of hypothesis

- The null hypothesis states a claim that we assume to be true unless we can find sufficient evidence to indicate otherwise.
- The alternative hypothesis is a statement that we hope or suspect is true but we need sufficient evidence in its support before we are willing to accept it.

The two types of error

	Accept H_0	Reject H_0
H_0 is true	Correct	Type I error (*false alarm*)
H_A is true	Type II error (*lost opportunity*)	correct

- Of these, type I error is more serious and we try to keep this at or below a pre-specified margin of error.

One and two sided alternatives

- The alternative hypothesis can be either one-sided, like $\mu < \mu_0$ or $\mu > \mu_0$, or two-sided, like $\mu \neq \mu_0$.
- The alternative hypotheses are formed depending on what we are trying to prove, or hope, fear or suspect to be true.
- We must form the hypotheses before we collect the data.

Examples - Setting up Hypotheses

- The mean area of new apartments in East Perth is advertised as $520m^2$. You think that they are smaller than advertised.
- Your car averages 10km per litre. A new fuel is supposed to increase the mileage.
- The diameter of a bolt is supposed to be 6mm. If it is either too big or too small the nut will not fit properly

- Our tests are usually based on

$$\boxed{\frac{\text{statistic - parameter}}{\text{standard error}}}$$

- For example, say we suspect that 500g packets of cereal are being under-filled.

- We cannot accuse the company unless we feel quite confident that we are correct in our assertion.

- The alternative hypothesis would be

$$H_A: \mu < 500.$$

- All calculations are carried out assuming that the null hypothesis is true so it is always in the form "=”. In this case it is $H_0: \mu = 500$.

- Note: This is called a one-sided test as we are not concerned if we get more than the nominal amount of cereal.

- To test the hypothesis, we take a simple random sample and calculate the sample mean.
- We now want to know whether the sample provides enough evidence to suggest

  $\mu < 500$.

Logic:

- If $\mu = 500$ we expect some random fluctuations about 500, but how far from 500 do we need it to be before we can say it is unlikely that the sample came from a population with mean 500?

Test Statistic

- We calculate the test statistic
- That is, we standardize the $\bar{x}$ value and obtain a z-value that tells us how many standard errors our observed value is from the hypothesised mean.

$$z = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}}$$

Value under Null hypothesis

Known value of variance of random variable

P-value

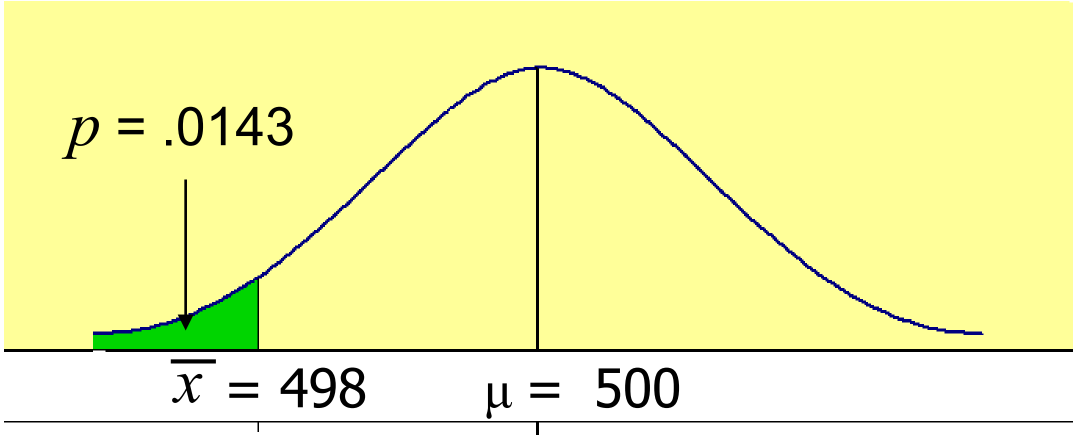
- From the tables (or calculator) find the probability of obtaining a value as extreme or more extreme as this. This probability is called the *p*-value.

- If the p-value is small then we say that we have enough evidence to reject the null hypothesis (i.e. there is only a small chance of the null hypothesis being correct so we are willing to back the alternative hypothesis).

Example

- A random sample of size 30 gives a sample mean of 498g. Given that the population standard deviation is 5, is there enough evidence to suggest that the packages are underweight?
- H_0: $\mu = 500$
- H_A: $\mu < 500$
- We calculate $P\left(\bar{x} < 498\right)$ assuming that H_0 is true.

- That is, $$\bar{x} \sim N\left(500, \tfrac{5}{\sqrt{30}}\right)$$
- So if $\mu = 500$ the probability of obtaining a sample mean less than 498 is 0.0143.
- Is this a small enough probability to say that we think the null hypothesis is wrong? (It is a value that would occur 1.43% of the time by chance alone when in fact H_0 is true).



Level of Significance

- We often decide in advance how small a p-value we need to reject H_0. This is called the level of significance (α).

- Common values for α are 0.05 and 0.01. i.e. the data gives evidence against H_0 that is so strong that it would happen no more than 5% (1%) of the time.

Conclusion at 5% level of significance

- The p-value is 0.0143 which is less than 0.05 so at the 5% level of significance we reject H_0 and state that the cereal boxes on average contain less than 500gm.

- *Note that in making the statement there is a chance that we are wrong but that chance is very small.*

Conclusion at 1% level of significance

- The p-value is 0.0143 which is greater than 0.01 so at the 1% level of significance we conclude that we have insufficient evidence to support the claim that the average weight of the boxes is less than 500gm.

- *Note we are NOT saying that the boxes do contain 500gm just that we have insufficient evidence to support the alternative hypothesis.*

P-values for two-sided tests

- Remember that the p-value is the probability of observing as much or more extreme a value as the observation. So if the above test had been two-sided, i.e. $\mu \neq 500$, the p-value would have been $2(0.0143) = 0.0286$.

Classical approach to hypothesis testing

1. Decide on the level of significance.
2. Draw a bell curve
3. Mark the critical value and region on the curve
4. Calculate the value of the test statistic
5. Put the calculated value on the bell curve
6. State your conclusion

Example

The TIV Telephone Company provides long-distance telephone service in an area. According to the company's records, the average length of all long-distance calls placed through this company in 1999 was 12.44 minutes with a standard deviation of 2.65 minutes.

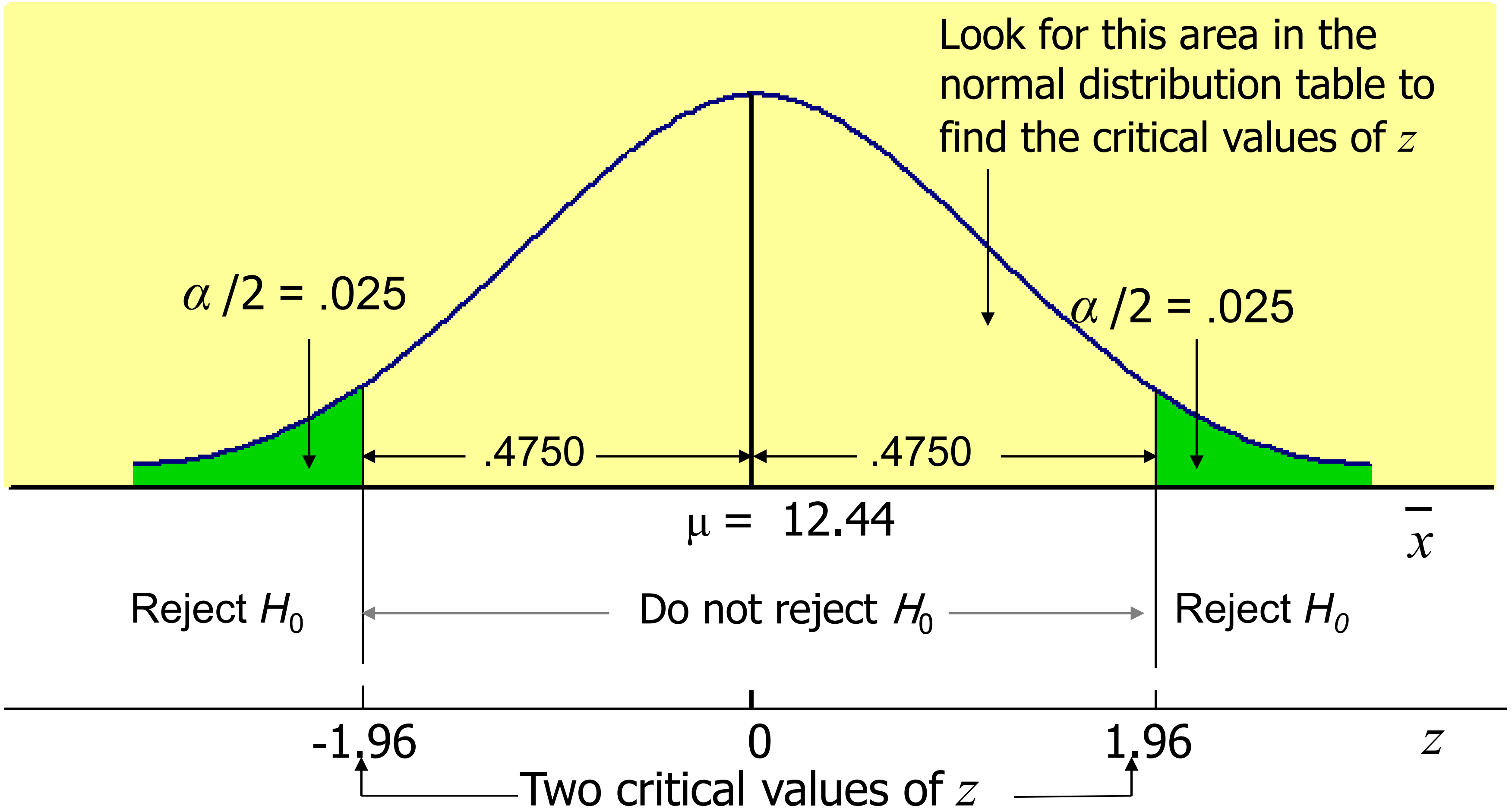
The company's management wanted to check if the mean length of the current long-distance calls is different from 12.44 minutes.

A sample of 150 such calls placed through this company produced a mean length of 13.71 minutes. Using the 5% significance level, can you conclude that the mean length of all current long-distance calls is different from 12.44 minutes?

Solution

- $H_0 : \mu = 12.44$
  - The mean length of all current long-distance calls is 12.44 minutes
- $H_A : \mu \neq 12.44$
  - The mean length of all current long-distance calls is different from 12.44 minutes

- $\alpha = .05$
- The $\neq$ sign in the alternative hypothesis indicates that the test is two-tailed
- Area in each tail = $\alpha / 2 = .05 / 2 = .025$
- The z values for the two critical points are -1.96 and 1.96


Look for this area in the normal distribution table to find the critical values of z
α/2 = .025
α/2 = .025
.4750
.4750
μ = 12.44
x̄
Reject H0
Do not reject H0
Reject H0
-1.96
0
1.96
z
Two critical values of z

Solution cont.

Test statistic:

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = \frac{13.71 - 12.44}{.21637159} = 5.87$$

Conclusion:

- The value of z = 5.87
  - It is greater than the critical value
  - It falls in the rejection region
- Hence, we reject H_0

t-test

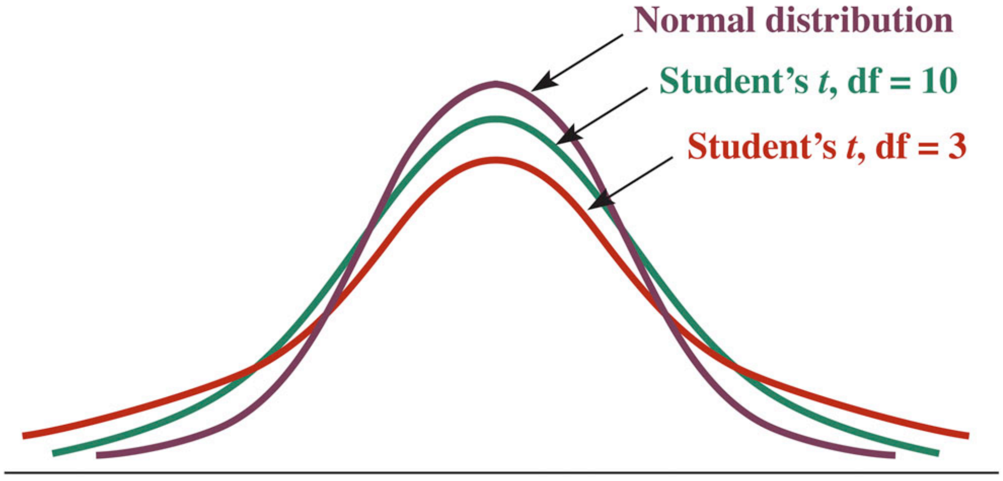
In reality the population standard deviation (σ) will not be known. We therefore use *s* as an estimate of σ, but now

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

will not be normally distributed. It will follow the *t*-distribution which is more variable than *z* but $\rightarrow$*z* as $n \rightarrow \infty$.

t distributions

- Suppose that an SRS of size n is drawn from a N(μ,σ) population.
- Then the one sample *t* statistic given earlier has the *t* distribution with *n*-1 degrees of freedom.
- Each *t* distribution is different and depends on the sample size. It is specified by its degrees of freedom.



When σ is unknown

- There are two major differences between inferential methods for the two cases of known and unknown σ.

- Firstly, we replace σ by its estimator *s*, the sample standard deviation.

- Secondly, we can no longer use normal critical value $z_{\alpha/2}$.

When σ is unknown

- Instead we use the values obtained from the t-distribution tables.

- Remember that these values depend not only on α, but also on the degrees of freedom, which is n - 1.

- For small n, the population from which you are sampling must be normally distributed

Example

A manufacturer of small appliances employs a market research firm to estimate retail sales of its products by gathering information from a sample of retail stores. This month an SRS of 50 stores in the sales region finds that these stores sold an average of 24 of the manufacturer’s hand mixers, with standard deviation 11.

(a) Give a 95% confidence interval for the mean number of mixers sold by all stores in the region.

(b) The distribution of sales is strongly right skewed, because there are many smaller stores and a few very large stores. The use of *t* in (a) is reasonably safe despite this violation of the normality assumption. Why?

Solution

(a) d.f. = *n* - 1 = 49;

$$\overline{x} \pm t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

$$= 24 \pm (\ldots\ldots) \frac{11}{\sqrt{50}}$$

$$= 24 \pm$$

$$= (\underline{\qquad\qquad}, \underline{\qquad\qquad})$$

CLT

(b) As sample size is large, the sample mean will be approximately normal even though the parent population is not. Thus it is safe to use *t*-distribution.

Example

The level of phosphate in the blood varies normally over time. A dialysis patient had his levels monitored on six visits to the clinic. Is there evidence that his level is above 4.8 which is the top of the range for healthy people?

5.6 5.1 4.6 4.8 5.7 6.4

H_0: $\mu = 4.8$ and H_A: μ ___ 4.8

- The sample mean and the sample standard deviation are 5.367 and 0.665 respectively.

- Test statistic: $t = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

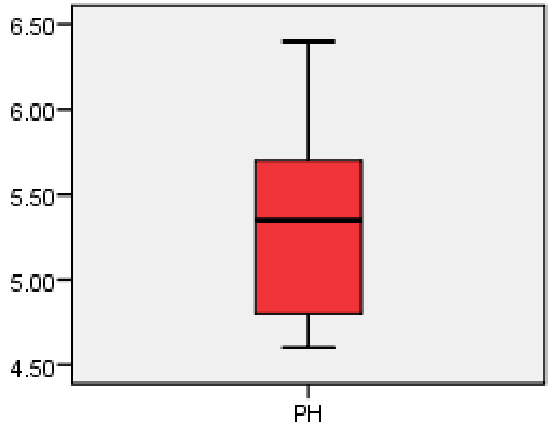
$$= \frac{5.367 - 4.8}{.665/\sqrt{6}} = 2.089$$

- P-value: P(t > 2.089) (___ *df*)

  ______ <p-value< _______

- Conclusion: At 5% level of significance, ______ H_0 and conclude that the patient's phosphate level is__________________

Checking Underlying Assumptions



Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
PH	.156	6	.200*	.955	6	.783

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

- No extreme skewness, no outliers
- No departure from Normality

$\Rightarrow$Test is valid

EXAMPLE

Here are data on a group of people who contracted botulism, a form of food poisoning that can be fatal. The variables recorded are the person's age in years, the incubation period (the time in hours between eating the infected food and the first sign of illness), and whether the person survived (S) or died (D).

Person	1	2	3	4	5	6	7	8	9
Age	29	39	44	37	42	17	38	43	51
Incubation	13	46	43	34	20	20	18	72	19
Outcome	D	S	S	D	D	S	D	S	D
Person	10	11	12	13	14	15	16	17	18
Age	30	32	59	33	31	32	32	36	50
Incubation	36	48	44	21	32	86	48	28	16
Outcome	D	D	S	D	D	S	D	S	D

Example cont.

Test if the mean incubation period for the population is more than 30 days →One sample T-Test

Null Hypothesis: mean incubation period is 30 days (or $H_0: \mu = 30$)

Alternative Hypothesis: mean incubation period is more than 30 days (or $H_A: \mu > 30$)

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Incubation	18	35.7778	19.7917	4.6650

One-Sample Test

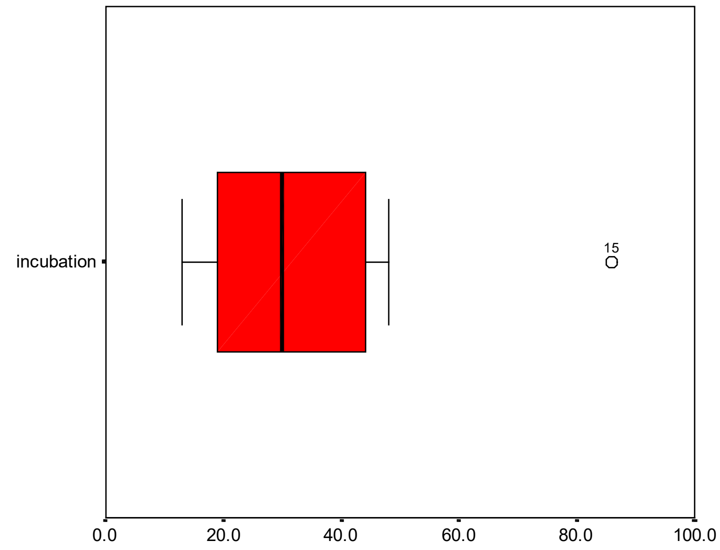
	Test Value = 30					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
incubation	.640	17	.531	2.7222	-6.2481	11.6926

P-value for test is 0.531/2 = **0.2655**

Why?

Example cont.

Checking Underlying Assumptions



Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
incubation	.187	18	.098	.847	18	.010**

**. This is an upper bound of the true significance.

a. Lilliefors Significance Correction

- No extreme skewness or many outliers
- No gross departure from Normality
- Test is valid

Example cont.

CONCLUSION

Note that the proposed test is one tailed, hence **p value** for the test is = (0.531)/2 = 0.2655 > 0.05; hence we have **enough evidence** to conclude that the mean incubation period is 30 days (**do not reject H_0**). In other words mean incubation period as high as 35.778 would occur about 26.55% of time if the population mean incubation period was 30 days. In report you would write; the data do not support the claim that the mean incubation period is more than 30 days (t=0.640, df=17, p=0.266, one-tailed).

Lecture Summary

- Sample size $n = \left(\dfrac{z_{\alpha/2}\sigma}{E} \right)^2$
- Confidence interval for an unknown σ:

$$\overline{x} \pm t_{n-1,\frac{\alpha}{2}} \frac{s}{\sqrt{n}}$$

- Test statistic: $t = \dfrac{\overline{x} - \mu}{s/\sqrt{n}}$
- P-value: the probability of observing as much or more extreme a value as the observation.