Estimation

Estimation

- Estimation is the process of using sample data to draw inferences about the population

Sample information	Inferences	Population parameters
$\bar{x}, s^2$	$\longrightarrow$	μ, σ^2

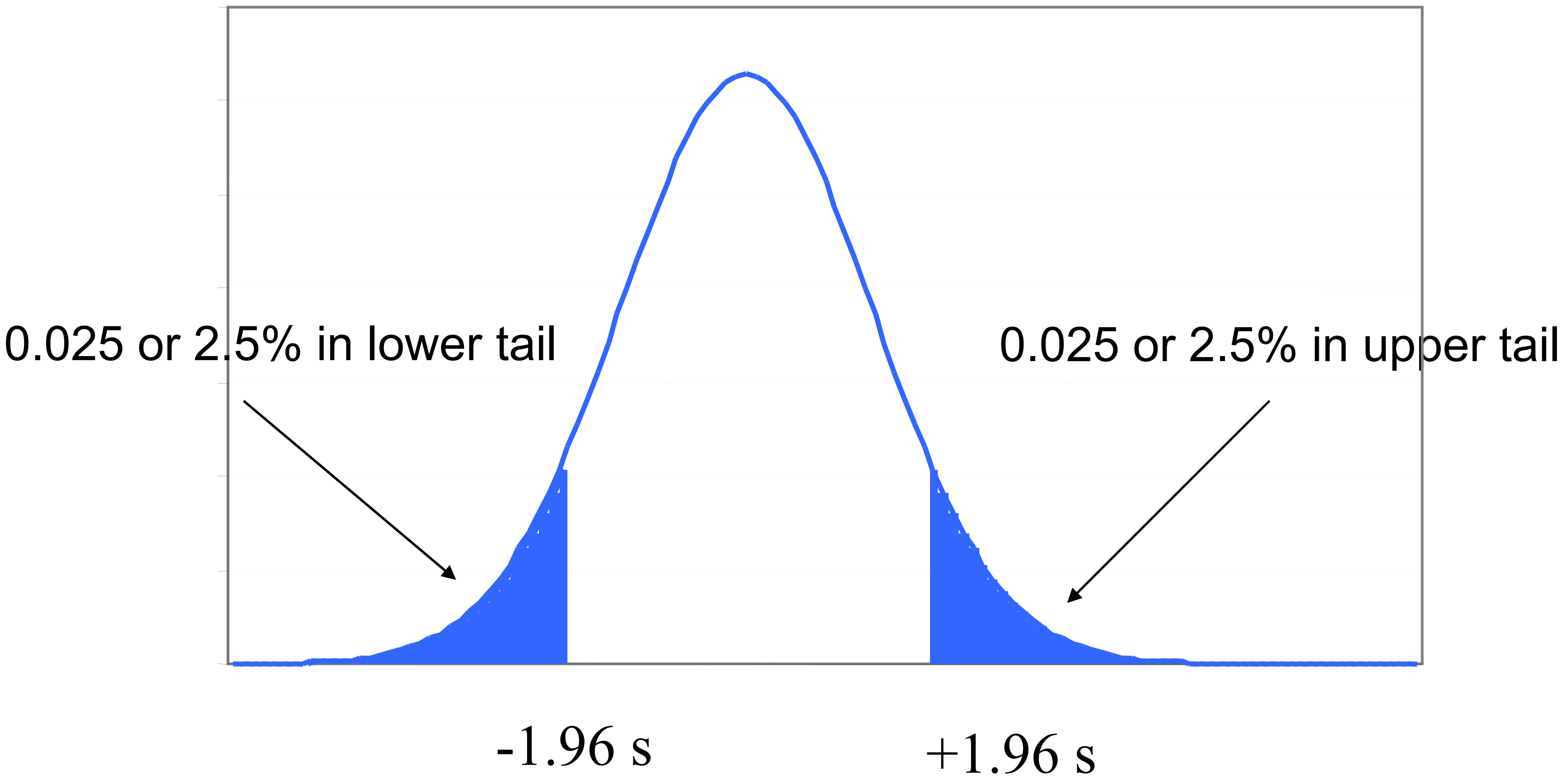
Point and Interval Estimates

- Point estimate – a single value
  - E.g. the temperature tomorrow will be 23°
- Interval estimate – a range of values, expressing the degree of uncertainty
  - E.g. the temperature tomorrow will be between 21° and 25°

Estimating a Mean (Large Samples)

- Point estimate – use the sample mean (unbiased)
- Interval estimate – sample mean ± ‘something’
- What is the something?
- Go back to the distribution of $\overline{x}$

Normal Distribution



Standardised Normal Table

z=1.96 defines 2.5% of observations in each tail so 95% in-between ±1.96 standard errors of the mean

z	**0.00**	..	**0.06**	**0.07**
0.0	0.5000		0.4761	
0.1	0.4602		0.4364	
1.5	0.0668		0.0606	0.0630
1.9	⋮	⋮	0.0250	⋮

The 95% Confidence Interval

- Recall the distribution of the sample mean

$$\bar{x} \sim N(\mu, \sigma^2 / n)$$

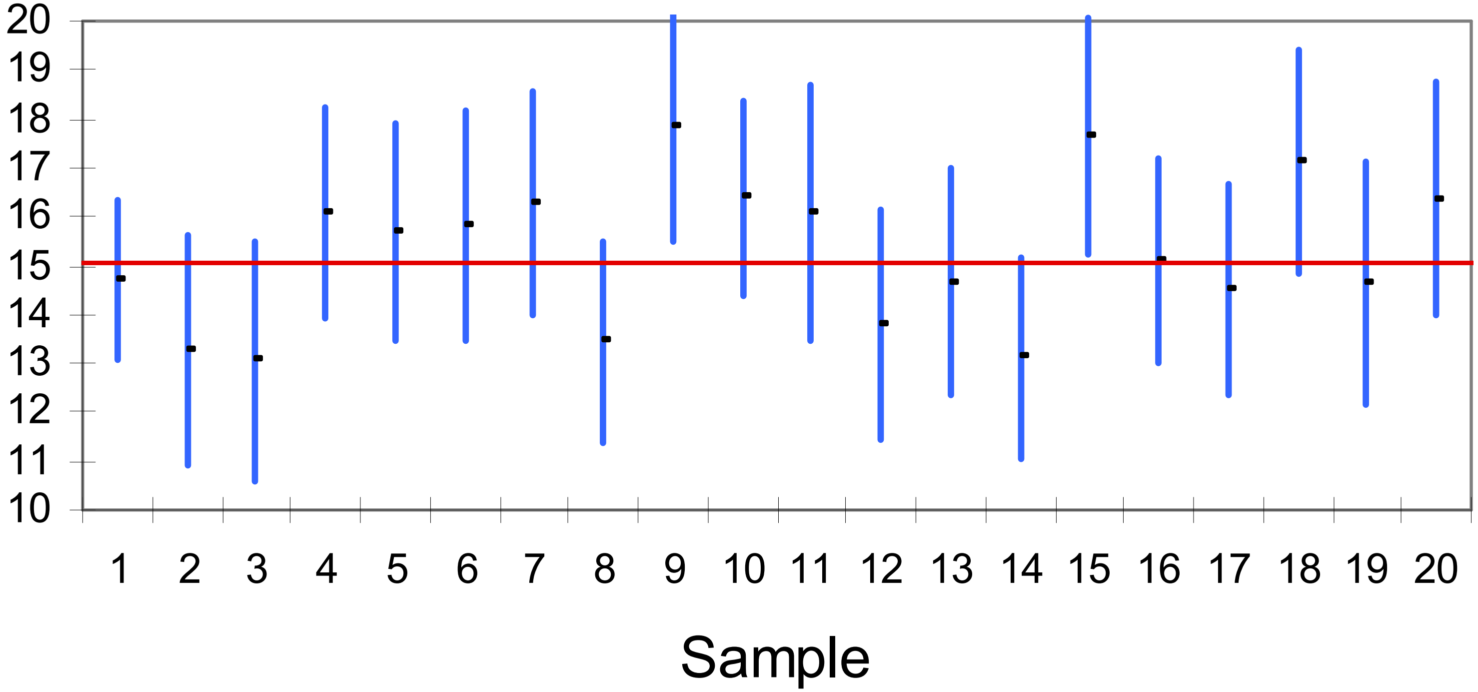
- Hence the 95% probability interval is

$$\Pr(\mu - 1.96\sqrt{\sigma^2/n} \leq \bar{x} \leq \mu + 1.96\sqrt{\sigma^2/n} = 0.95$$

- Rearranging this gives the 95% confidence interval for our estimate of the true population mean

$$[\bar{x} - 1.96\sqrt{\sigma^2/n} \leq \mu \leq \bar{x} + 1.96\sqrt{\sigma^2/n}]$$

What is a Confidence Interval?



One sample out of 20 (5%) does not contain the true mean, 15.

Example: Estimating Average Wealth

- Sample data:
  - $\overline{x}$ = 130 (in £000)
  - $s^2 = 50{,}000$
  - $n = 100$
- Estimate μ, the population mean

Example: Estimating Average Wealth (cont.)

- Point estimate: 130 (uses the sample mean)
- Interval estimate

$$\bar{x} \pm 1.96 \times \sqrt{s^2/n}$$
$$= 130 \pm 1.96 \times \sqrt{50{,}000/100}$$
$$= 130 \pm 43.8 = [86.2, 173.8]$$

- so we are 95% confident that the true mean lies somewhere between £86,200 and £173,800

Using Different Confidence Levels

- The 95% confidence level is a convention
- The 99% confidence interval is calculated by adding and subtracting 2.57 standard errors (instead of 1.96) to the point estimate.
- A higher level of confidence implies a wider interval

Estimating a Proportion

- Similar principles
  - The sample proportion provides an unbiased point estimate
  - The 95% CI is obtained by adding and subtracting 1.96 standard errors
- We need to know the sampling distribution of the sample proportion, where p is the sample proportion, π the population proportion

$$p \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$

Example: Unemployment

- Of a sample of 200 men, 15 are unemployed. What can we say about the true proportion of unemployed men?
- Sample data
  - p = 15/200 = 0.075
  - n = 200

Example: Unemployment (cont.)

- Point estimate: 0.075 (7.5%)
- Interval estimate:

$$p \pm 1.96 \times \sqrt{\frac{p(1-p)}{n}}$$

Standard error of proportion

$$= 0.075 \pm 1.96 \times \sqrt{\frac{0.075 \times 0.925}{200}}$$

$$= 0.075 \pm 0.037 = [0.038, 0.112]$$

Estimation With Small Samples: Using the *t* Distribution

- Recall that if
  - The sample size is small (<25 or so), and
  - The true variance σ^2 is unknown
- Then the *t* distribution should be used instead of the standard Normal.

Example: Beer Expenditure

- A sample of 20 students finds an average expenditure on beer per week of £12 with a standard deviation of £8. Find the 95% CI estimate of the true level of expenditure of students.
- Sample data:

$$\bar{x} = 12, s = 8, n = 20$$

Example: Beer Expenditure (cont.)

- The 95% CI is given by

Just use a value from t-table instead of normal

$$\bar{x} \pm t_{\alpha/2, n-1} \sqrt{s^2/n},$$

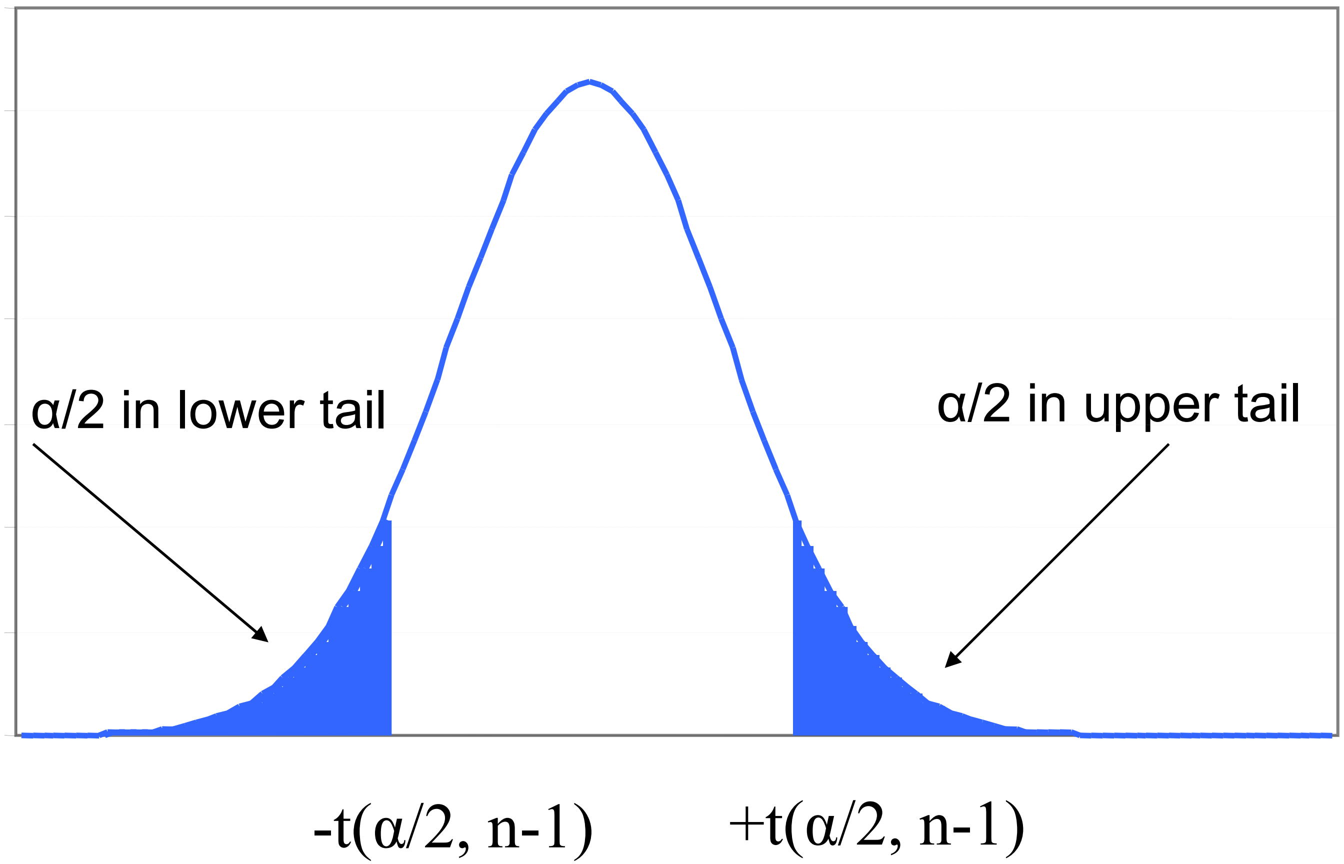
$$t_{0.025,19} = 2.093$$

$$= 12 \pm 2.093 \sqrt{8^2/20}$$

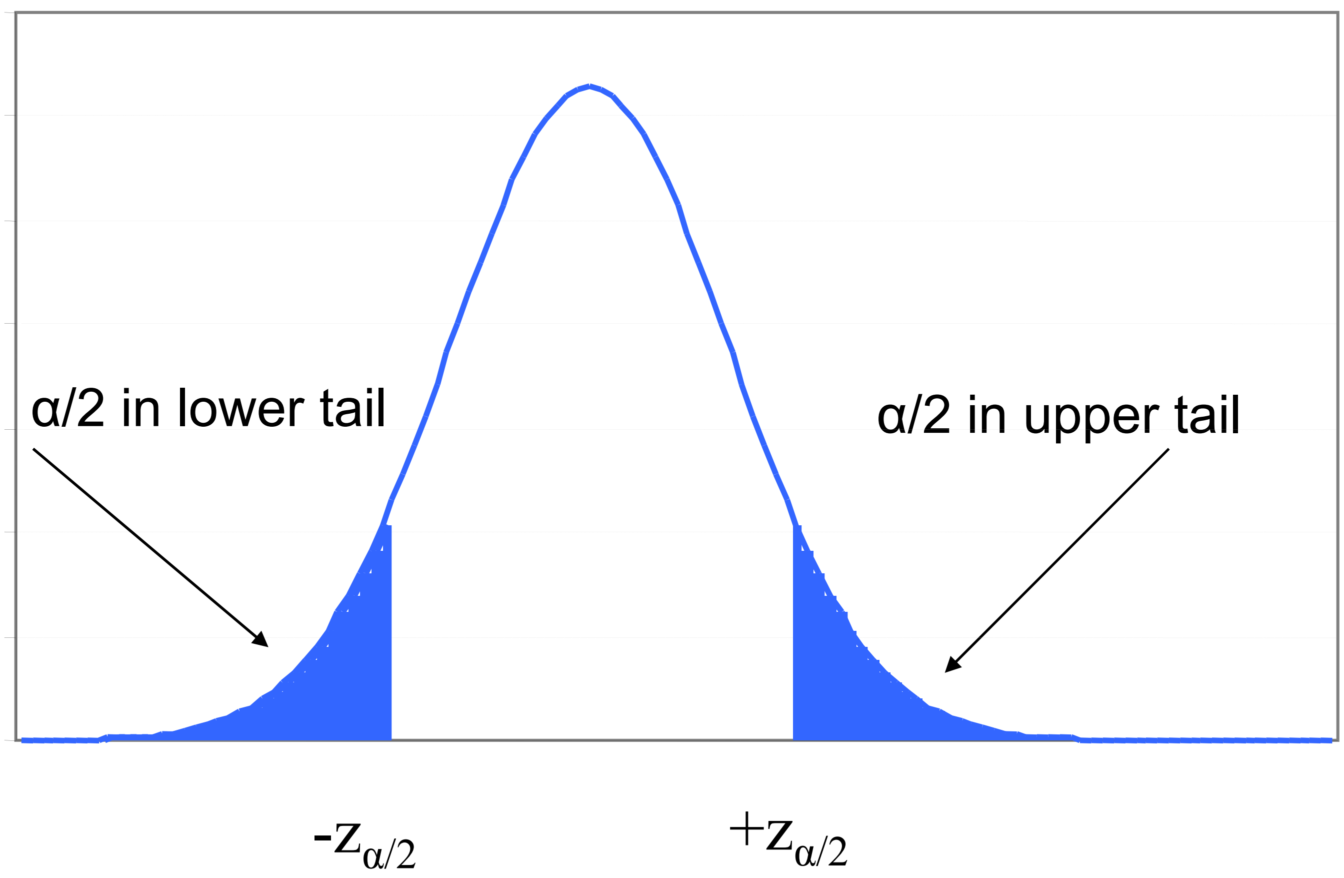
$$= 12 \pm 3.7 = [8.3, 15.7]$$

- The *t* value of t_{19} = 2.093 is used instead of *z* =1.96

100(1-α)% Confidence Intervals: Small Samples



100(1-α)% Confidence Intervals: Large Samples



Estimating the Difference of Two Means

- Example: A survey of holidaymakers found that on average women spent 3 hours per day sunbathing, while men spent 2 hours

- $n_w = n_m = 36$, $s_w = 1.1$, $s_m = 1.2$

- Estimate the true difference between men and women in sunbathing habits

Same Principles as Before…

- Obtain a point estimate from the samples

$$\bar{x}_w - \bar{x}_m = 3 - 2 = 1$$

- Add and subtract 1.96 standard errors to obtain the 95% CI
  - We just need the appropriate formulae $\boxed{\sqrt{\frac{s_w^2}{n_w} + \frac{s_m^2}{n_m}}}$

$$1 \pm 1.96\sqrt{\frac{1.1^2}{36} + \frac{1.2^2}{36}}$$

$$= 1 \pm 0.7 = [0.3, 1.7]$$

- The difference between women and men's sunbathing is between 20 mins. and 1 hour and 42 mins.

Confidence Intervals: General Steps

- Decide on the probability $100(1-\alpha)\%$ you want to associate with the confidence interval
- Determine the sampling distribution of the random variable in question (mean, proportion, differences between means, proportions) and NB that this sometime varies by sample size
- Look up the value of $z_{\alpha/2}$ or $t_{\alpha/2}$ in the relevant tables
- the $100(1-\alpha)\%$ CI is obtained by adding and subtracting z or t standard errors

Summary

- The sample mean and proportion provide unbiased estimates of the true values
- The 95% confidence interval expresses our degree of uncertainty about the estimate
- The point estimate ± 1.96 standard errors provides the 95% interval in large samples
- for small samples, need to use t with n-1 degrees of freedom.