INTRODUCTION TO STATISTICAL MODELLING

STAT2507D

Chapter 8 - 2

Large Sample Estimation

FALL 2014: STAT2507D
Carleton University

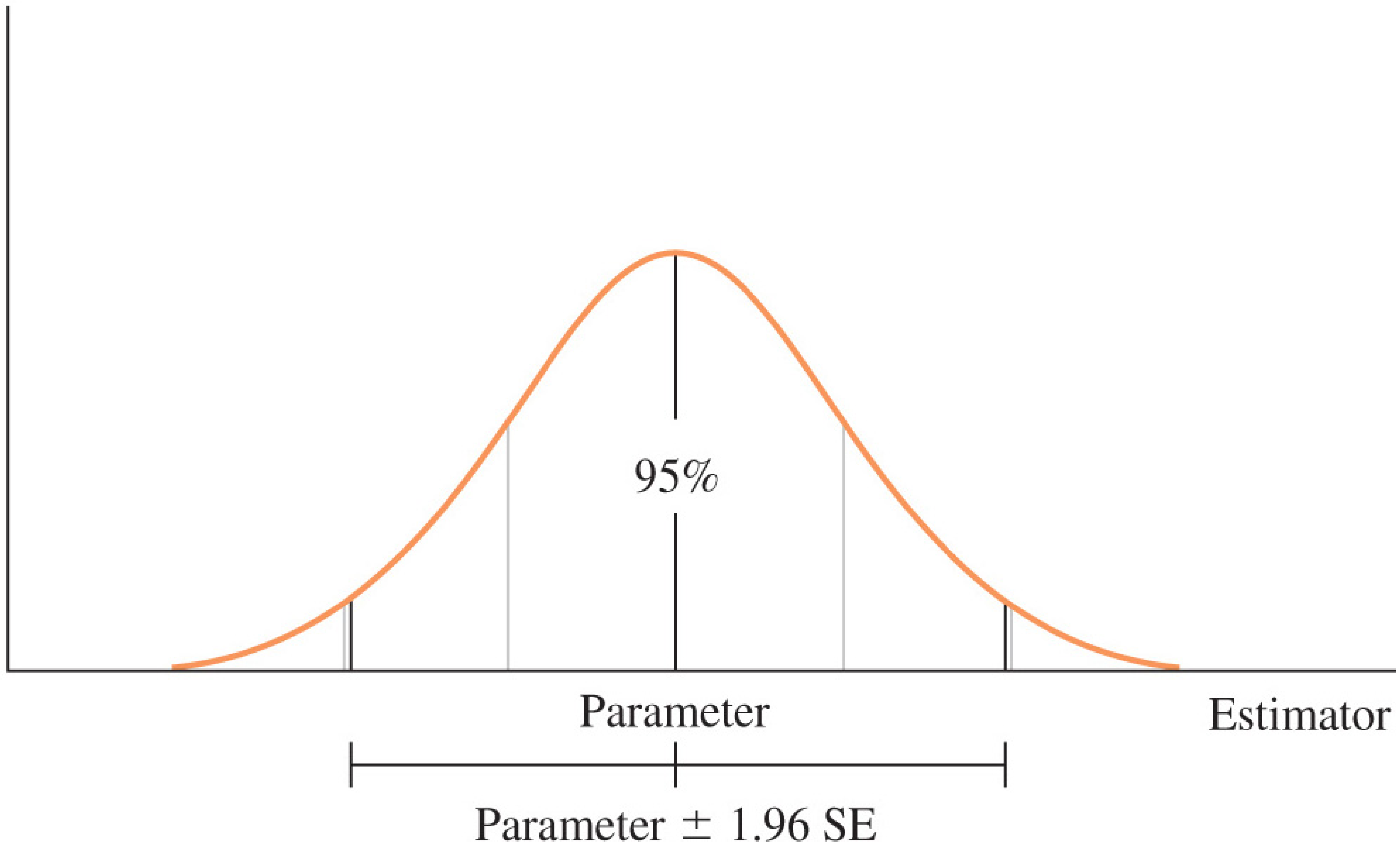
INTERVAL ESTIMATION

- Interval estimator is a rule for calculating two numbers – say, a and b – to create an interval that you are fairly certain contains the parameter of interest.
- "Fairly sure" means "with high probability", measured using the confidence coefficient, $1-\alpha$
- Usually, $1-\alpha = 0.90, 0.95, 0.98, 0.99$

INTERVAL ESTIMATION

- Sampling distribution of a point estimator is approximately normal
- Margin of Error = 1.96*SE (95% confidence coefficient)
- If we know the value of the parameter, then we do not have variable centre
- $Parameter \pm 1.96 * SE$

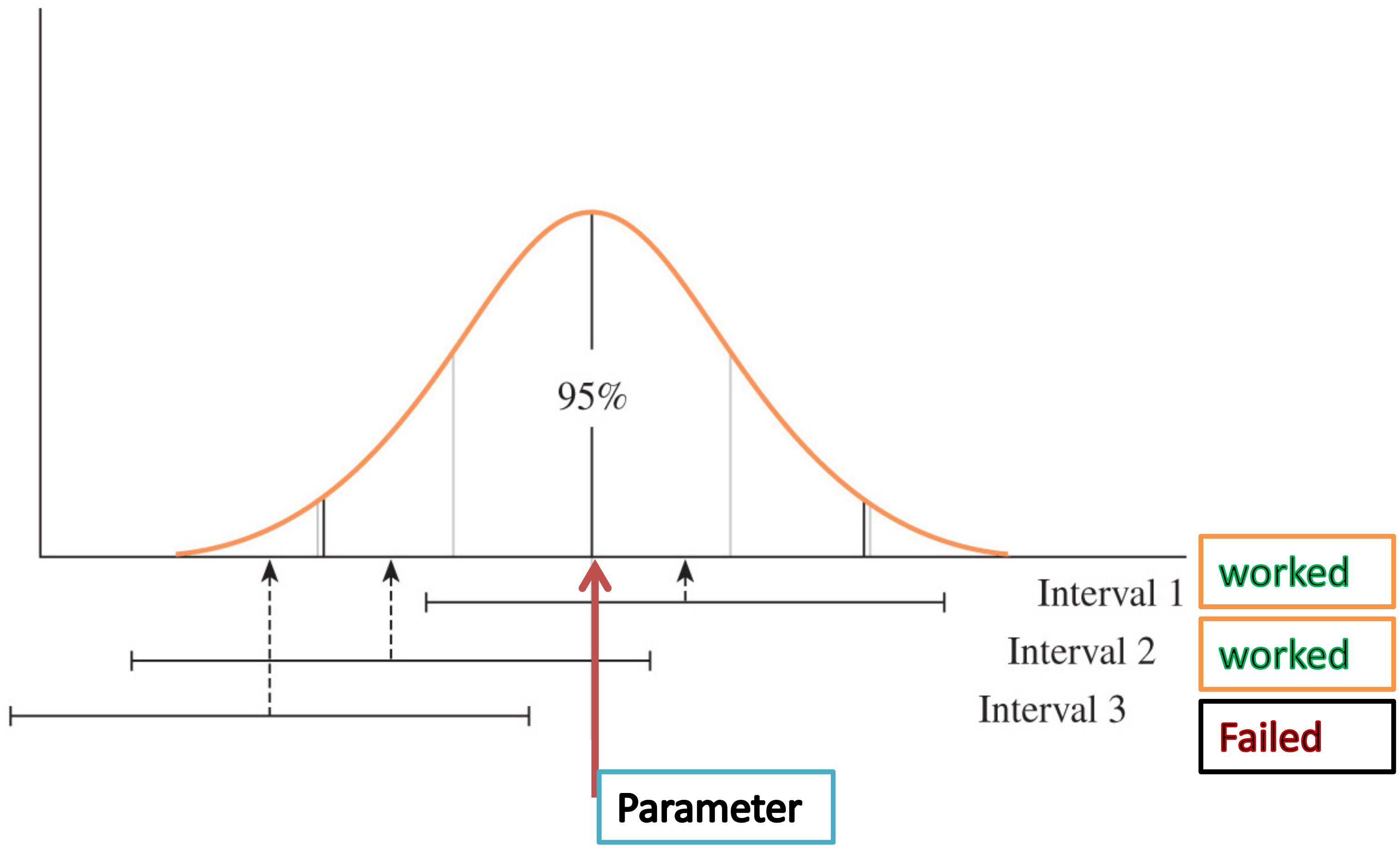
INTERVAL ESTIMATION



INTERVAL ESTIMATION

- Usually, we don’t know the value of the parameter. We use an estimator for the parameter
  - The value taken by the estimator changes depending the selected sample
  - Estimator ± 1.96SE
  - Hence, this has a variable centre

INTERVAL ESTIMATION



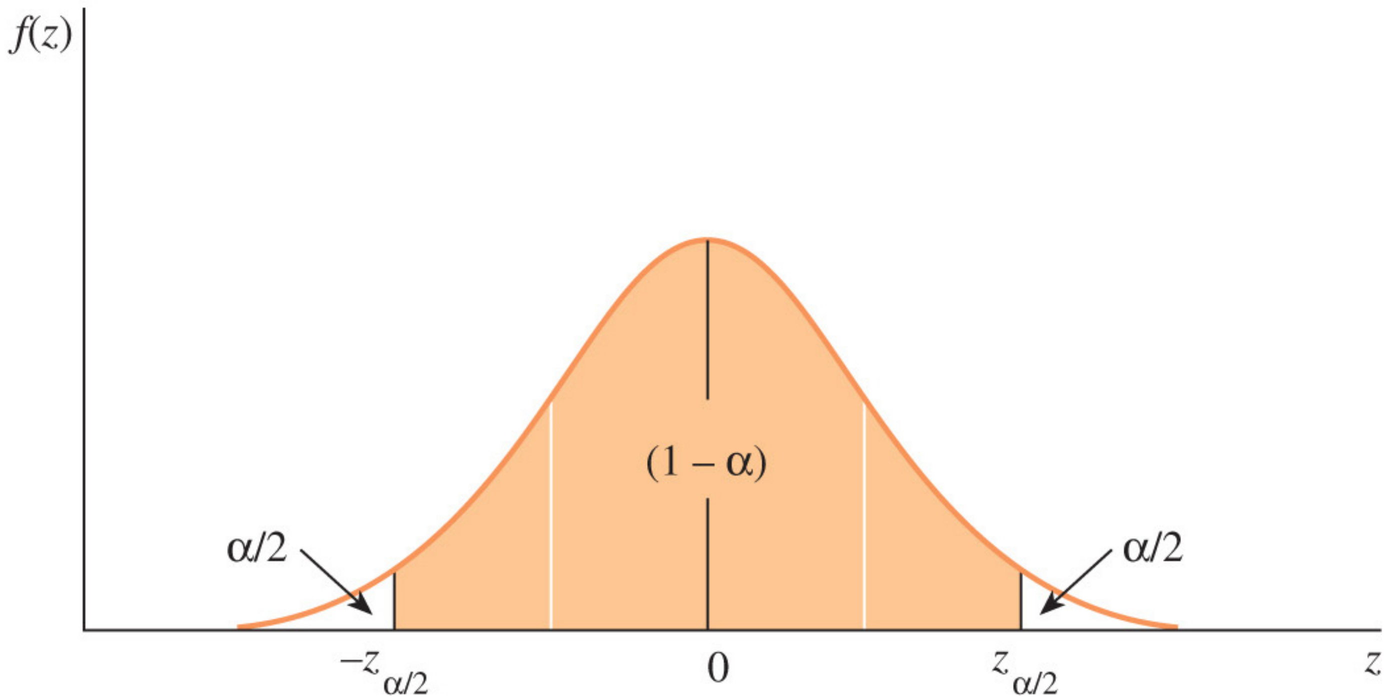
CHANGING THE CONFIDENCE LEVEL

- To change the confidence coefficient of 0.95 to (1-α), pick a value of Z that puts area (1-α) in the center of the Z distribution:
- 100(1-α)% Confidence interval:

$$Estimator \pm Z_{\frac{\alpha}{2}} * \mathrm{SE}$$

- Where $Z_{\frac{\alpha}{2}}$ is the z-value with an area α/2 in the right tail of a standard normal distribution.

CHANGING THE CONFIDENCE LEVEL



Z-VALUES FOR VARIOUS CONFIDENCE LEVELS

- Estimator$\pm$ $Z_{\alpha/2}$ SE

$(1-\alpha)$	α	$\alpha/2$	$(1-\alpha/2)$ (+ z table)	$Z_{\alpha/2}$
0.90	0.10	0.05	0.95	1.645
0.95	0.05	0.025	0.975	1.96
0.98	0.02	0.01	0.99	2.33
0.99	0.01	0.005	0.995	2.58

A (1-α)100% Large Sample Confidence Interval

- (Point Estimator) ± $Z_{\alpha/2}$(standard error of the estimator)
- Where $Z_{\alpha/2}$ is the z-value with an area of $\alpha/2$ in the right tail of a standard normal distribution.
- This formula gives two values: the lower confidence limit (LCL) and the upper confidence limit (UCL)
- LCL = (Point Estimator) – $Z_{\alpha/2}$(SE)
- UCL = (Point Estimator) + $Z_{\alpha/2}$(SE)

Large Sample Confidence Interval for Mean

- Large sample confidence interval for population mean μ
  - Practical problems often lead to the estimation of μ, the mean of the population of quantitative measurements. E.g. Average demand for a certain product; average achievement of students at certain university

LARGE SAMPLE CONFIDENCE INTERVAL FOR MEAN

- Large sample confidence interval for population mean μ
  - When n is large, the sample mean $\bar{x}$ is the best estimator for the population mean μ.
  - Sampling distribution of $\bar{x}$ is approximately normal (by CLT), confidence interval can be estimated using approach seen earlier

Large Sample Confidence Interval for Mean

- A (1-α)100% large-sample confidence interval for a population mean
  - when σ known

$$\bar{x} \pm Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

- Where $Z_{\alpha/2}$ is the z-value corresponding to an area α/2 in the upper tail of a standard normal z distribution; n – sample size; σ – Standard deviation of the sampled population
- Above is confidence interval for quantitative population

Large Sample Confidence Interval for Mean

- Another way to find the large-sample confidence interval for a population mean μ using

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \xRightarrow{standardizing} Z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

$$P\left(-Z_{\frac{\alpha}{2}} \le Z \le Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

$$P\left(-Z_{\frac{\alpha}{2}} \le \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \le Z_{\frac{\alpha}{2}}\right) = 1 - \alpha$$

Large Sample Confidence Interval for Mean

- Previous equation can be re-written as follows:

$$P\left(\bar{x} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

- In repeated sampling, this random interval will contain the population mean μ with probability (1-α)

Large Sample Confidence Interval for Mean

- A (1-α)100% large-sample confidence interval for a population mean μ
  - when σ unknown

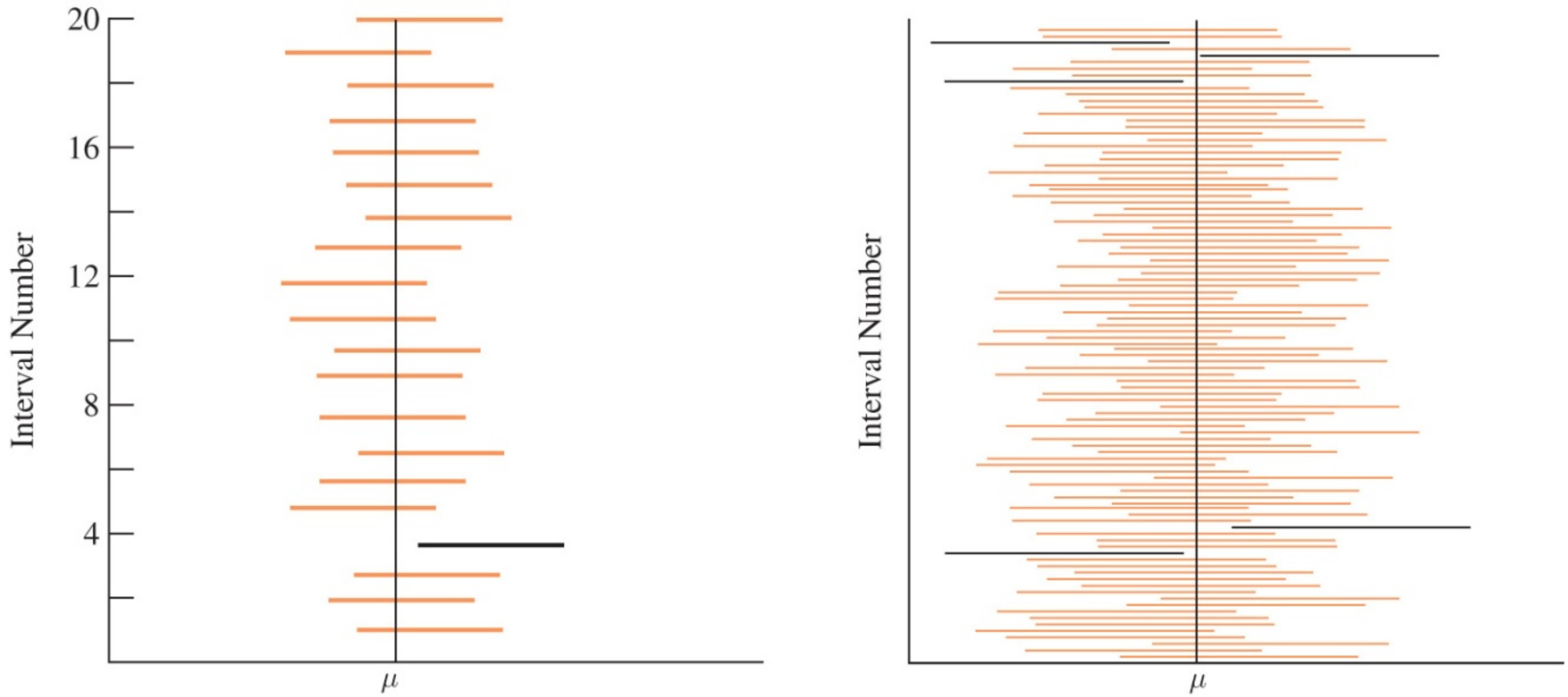
$$\bar{x} \pm Z_{\alpha/2}\frac{s}{\sqrt{n}}$$

$$P\left(\bar{x} - Z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + Z_{\frac{\alpha}{2}}\frac{s}{\sqrt{n}}\right) = 1 - \alpha$$

INTERPRETING CONFIDENCE INTERVAL

- What does it mean to say that you are "95% confident" that the true value of the population mean μ is within this given interval?
- When 20 such intervals are constructed, each using different sample information, 95% or 19 out of 20 will contain μ within their upper and lower bounds.

INTERPRETING CONFIDENCE INTERVAL



(a) 20 intervals

(b) 100 intervals

INTERPRETING THE CONFIDENCE INTERVAL

- Two desirable characteristics of a good confidence interval
  - It is as narrow as possible. The narrower the interval, the more exactly you have located the estimated parameter (SE is as small as possible)
  - It has a large confidence coefficient, near 1. The larger the confidence coefficient, the more likely it is that the interval will contain the estimated parameter.

EXERCISE

8.5) A random sample of n measurements is selected from a population with unknown mean μ and known standard deviation $\sigma = 10$. Calculate the width of a 95% confidence interval for μ for these values on n: 100, 200, 400.

EXERCISE

8.6) Referring to previous exercise, what effect does each of these actions have on the width of a confidence interval?

a. Double the sample size

b. Quadruple the sample size

EXERCISES

8.7) Refer to the previous exercises.

a) Calculate the width of a 90% confidence interval for μ when $n = 100$

b) Calculate the width of a 99% confidence interval for μ when $n = 100$

c) Compare the widths of 90% and 99% confidence intervals for μ. What effect does increasing the confidence coefficient have on the width of the confidence interval?

EXERCISES

8.8) Pure rain falling through clean air registers a pH value of 5.7 (pH is a measure of acidity: 0 is acid; 14 is alkaline). Suppose water samples from 40 rainfalls are analysed for pH, $\bar{x}$ and s are equal to 3.7 and 0.5, respectively. Find a 99% confidence interval for the mean pH in rainfall and interpret the interval. What assumptions must be made for the confidence interval to be valid?

LARGE SAMPLE CONFIDENCE INTERVAL FOR PROPORTION

- In some experiments or sample surveys, estimation of the proportion of people or objects in a large group that posses a certain characteristic is needed.
  - Proportion of seeds that germinate
  - The proportion of ‘likely’ voters who plan to vote for a particular political candidate

Large Sample Confidence Interval for Proportion

- Each is a practical example of binomial experiment and the parameter to be estimated is the binomial proportion p. When the sample size, n is large, the sample proportion

$$\hat{p} = \frac{x}{n} = \frac{Total\ number\ of\ successes}{Total\ number\ of\ trials}$$

LARGE SAMPLE CONFIDENCE INTERVAL FOR PROPORTION

- A (1-α)% large-sample confidence interval for a population proportion p

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{pq}{n}}$$

- Where $Z_{\alpha/2}$ is the z-value corresponding to an area α/2 in the right tail of a standard normal z distribution. When p and q are unknown, they are estimated using best point estimators $\hat{p}$ and $\hat{q}$

LARGE SAMPLE CONFIDENCE INTERVAL FOR PROPORTION

- Since p and q are unknown, they are estimated using the best point estimators: $\hat{p} = \frac{x}{n}$ and $\hat{q} = 1 - \hat{p}$
- When sample size is large enough so that $n\hat{p} > 5, and\ n\hat{q} > 5$, normal approximation to binomial distribution is valid.

EXERCISES

8.9) A random sample of $n = 300$ observations from a binomial population produced $x = 263$ successes. Find a 90% confidence interval for p and interpret the interval

EXERCISE

8.10) Based on the online survey of a representative national sample of 1006 Canadian adults, conducted, 55% think multiculturalism has been "very good" or "good" for Canada. Further, the survey found 54% want Canada to be a melting pot where immigrants assimilate and blend into Canadian society. The poll reported a margin of error – which measures sampling variability – is +/-3.1%, 19 times out of 20.

EXERCISE

a. Construct a 90% confidence interval for the proportion of Canadian adults who think the multiculturalism has been “very good” or “good” for Canada
b. Construct a 90% confidence interval for the proportion of Canadian adults who want Canada to be a melting pot where immigrants assimilate and blend into Canadian society
c. How did the researchers calculate the margin of error for this survey? Confirm that their margin of error is correct.