



Confidence Interval

Professor Dr Abubakir M. Saleh

Biostatistics NUR304

5th week



Outline

- The sampling variability of a percentage (Proportion).
- The sampling variability of a mean (Quantitative data).



Objectives

At the end of this lecture, students should be able to :

- Calculate confidence interval of a percentage.
- Calculate confidence interval of a mean



Introduction

- **Confidence intervals**, are used to express the statistical uncertainty of an estimate obtained from the data.
- Today's lectures introduce this concept for analysis of categorical data and quantitative measurements.
- Confidence intervals are linked by the concept of sampling variability.
- In the lecture we will illustrate this concept with practical examples.



The sampling variability of a percentage (proportion)

“Qualitative Data”



95% Confidence interval (CI) for a percentage

- When we observe a percentage in a random sample (e.g. 37%), it would be useful to be able to give an interval of possible values within which the true population percentage might lie.
- 95% CI is usually used to show the interval where the true population percentage lies with 95% degree of certainty (or confidence)

95% Confidence interval for a percentage

- A 95% confidence interval for a percentage=
 $p \pm 1.96 \times \text{standard error of } p$

Usually written

$$p \pm 1.96 \times \text{SE}(p).$$

$$\text{SE}(P) = \sqrt{\frac{p \times (100 - p)}{n}}$$

$$\text{So the 95\% CI } (p) = p - 1.96 \times \sqrt{\frac{p \times (100 - p)}{n}} \quad \text{to} \quad p + 1.96 \times \sqrt{\frac{p \times (100 - p)}{n}}$$

p: percentage

n: number of observations (sample)

- If the 335 men were a random sample of all men in Erbil the true population percentage of smokers in Erbil has 95% confidence interval *(37.9% were smokers)*

$$37.9 \pm 1.96 \times \sqrt{\frac{37.9 \times (100-37.9)}{335}}$$

$$37.9 - 1.96 \times \sqrt{\frac{37.9 \times 62.1}{335}} = 32.7\%$$

$$37.9 + 1.96 \times \sqrt{\frac{37.9 \times 62.1}{335}} = 43.1\%$$

i.e. from 32.7% to 43.1%.

- These two values are the lower and upper confidence limits, respectively.
- The population percentage of smokers is very likely to lie between 32.7% to 43.1%.



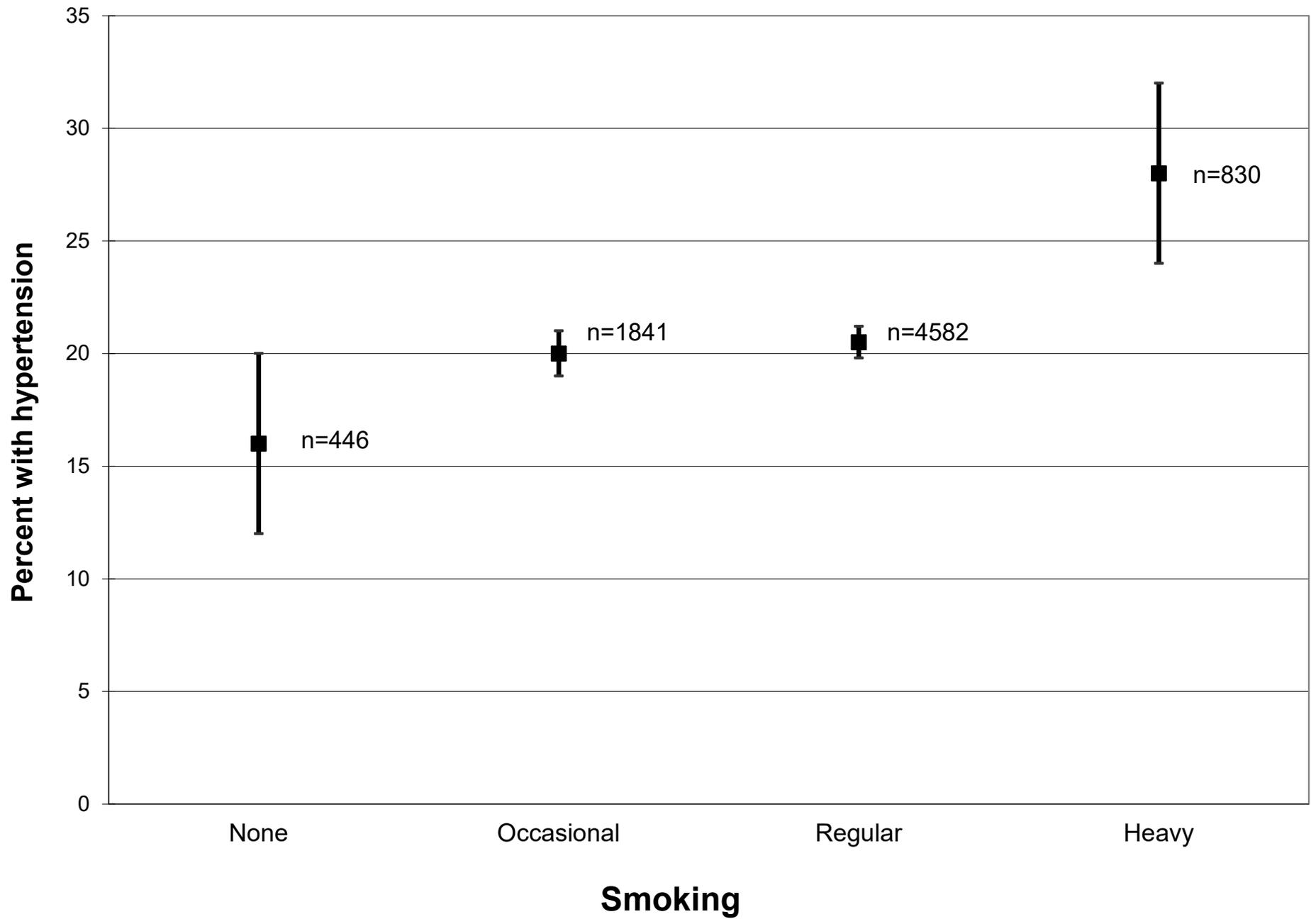
Interpretation of 95% CI

- 95% confidence interval is the most common statistical technique for displaying the degree of uncertainty that should be attached to any percentage.
- We are 95% confident that the true population percentage lies within this interval (i.e. 32.7% to 43.1%).
- Remember there exists a 5% risk that the true population percentage lies **outside** the interval.



Display of Confidence Intervals

- If we have percentage for two or more groups, we can display the percentages and their 95% confidence interval in a graph
- In a survey of over 7,000 men their smoking habits were classified into four categories: none, occasional, regular and heavy.
- The percentages of men in each group with systolic hypertension, together with the 95% confidence limits, are shown in the following figure





- The squares show the observed percentages, while the vertical lines show the 95% confidence intervals.
- Note how the confidence intervals become very narrow for a large sample (e.g. the 4582 regular smokers).
- Also, note that if two confidence intervals do not overlap, (e.g. heavy versus regular smokers) this is evidence of a real difference between two groups.



The sampling variability of a mean

“Quantitative data”

Notation

- A random sample of size (n) is taken from the population of interest.
- The mean (\bar{x}) and standard deviation (SD) of the quantitative variable in the sample are calculated
- E.g. of quantitative variables; height, weight, Blood pressure, serum cholesterol, haemoglobin level.

Question

- On the basis of \bar{X} and SD, what can we say about true population mean ?

- Example

The mean haemoglobin level of 25 persons sampled randomly from a population living in a town, with the following results:

Mean = 13.6 gm/dl

SD = 4.3

- What can be said about the true mean of haemoglobin level in this population?

Confidence Interval for a Mean

- To make inferences about the true mean of population, we construct a **confidence interval** using the same approach as that used for proportions.
- **Standard error of the mean**, or $SE(\bar{X}) = \frac{SD}{\sqrt{n}}$
- 95% CI of mean = $\bar{x} \pm 1.96SE(\bar{x})$
- $\bar{x} \pm 1.96 \times \frac{SD}{\sqrt{n}}$

- **Example**

- In the haemoglobin level example,

$$n = 25, \quad \bar{x} = 13.6, \quad SD = 4.3$$

- The 95% confidence interval:

$$13.6 \pm (1.96 \times 4.3/\sqrt{25})$$

which is 11.9 to 15.3 gm/dl

Example

- The weight of nine children in school is:

32, 32, 31, 30, 28, 28, 27, 25, 18 kg

Mean = 27.9 kg

Standard deviation = 4.4

- *What is the confidence interval of the mean?*

The number of children is 9. We can use the formula to calculate the standard error:

$$4.4/\sqrt{9}$$

- This becomes:

- $SE = 4.4/3 = 1.467$

- The 95% confidence interval for the mean is:

sample mean - 1.96 standard errors' to 'sample mean + 1.96 standard errors'

or:

$$27.9 - 1.96 \times 1.467 \text{ to } 27.9 + 1.96 \times 1.467$$

- So, the 95% confidence interval of the mean is: *from 25.0 to 30.8 kg*

- We are 95% confident that the true mean weight of school children is between 25 and 30.8Kg

Summary

- Sample proportion or mean is only an estimate of population proportion or mean
- Confidence interval is one way to see how accurate our sample result (proportion or mean) is.

95% Confidence interval:

- Percentage $p - 1.96 \times \sqrt{\frac{p \times (100 - p)}{n}}$ to $p + 1.96 \times \sqrt{\frac{p \times (100 - p)}{n}}$
- Mean $\bar{x} \pm 1.96 \times \frac{SD}{\sqrt{n}}$

Q1. In a survey of contraceptive use, a sample of 1200 women in a town found 25% were current users of contraception.

What is the 95% confidence interval ?

The standard error of this estimate is 1.25%.

The 95% confidence interval is from 22.55% to 27.45%

Interpret this 95%CI

We are 95% confident that 22.55 to 27.45% of women in this town use contraceptive

Q. The mean birth weight of a representative sample of 153 newborns is 3.250 Kg and the SD is 0.428 Kg. A 95% confidence interval for the population mean birth weight is:

from 3.182 to 3.318 Kg

Which one is true

- a) about 95% of the individual newborn birth weights are between 3.182 and 3.318 Kg
- b) the mean birth weight for these 153 newborns is probably between 3.182 and 3.318 Kg
- c) the mean of the population from which the 153 newborns came is between 3.182 and 3.318 Kg
- d) there is a 95% probability that the mean birthweight of the population from which the 153 newborns came is from 3.182 and 3.318 Kg

- Answer **d**



References

- [Essential Medical Statistics](#), by Betty Kirkwood & Jonathan Sterne
(Published by Blackwell)
[Statistics Without Tears](#), a Primer for Non-mathematicians, by Derek Rowntree
(Published by Penguin)

Questions?