



The course material is available under the following copyright agreements and protected by Creative Commons



# Strengthening the Occupational Health Expertise and Scientific Performance of Public Health Institution of Turkey



The project leading to this presentation has received funding from the EU Horizon 2020 Research and Innovation Programme under agreement No 692188. This presentation reflects only the author's views. The Research Executive Agency under the power of the European Commission is not responsible for any use that may be made of the information it contains.



## 3.2 Introduction to Statistics

# Learning Objectives

1. Understand why we need statistics
2. Explain basic concepts in statistics like: case, population, rate, sample, random
3. Know the difference between incidence and prevalence, between mortality and morbidity
4. Understand why and when statistical inference is needed
5. Be able to interpret existing descriptive epidemiologic literature
6. Be able to look critically at statistics



# What is a case?

**“a health-related state or event”**

- Occupational Disease
- Work-related Disease
- Work-related Health Complaints/ ill Health

A case definition must be:

- Clearly stated
- Easy to use and measure in a standard manner by different people



# What is the population at risk?

Total number of people under study

- All workers
- All workers in a certain region
- All workers in a certain industrial sector
- All workers with a certain occupation, age, gender...

Often a (random) sample is used because national data are not available.



# Why use rates?

**DAILY NEWS** LEADING NEWS SOURCE FOR TURKEY AND THE REGION

HOME | TURKEY | ECONOMY | WORLD | SPORTS | LIFE | ARTS/CULTURE | OPINION | VI

Sub Categories: » [HOMEPAGE](#) / [TURKEY](#) / [LOCAL](#)

## At least 200 workers died of work related reasons in June: Report

ISTANBUL

[Like](#) 5 [Tweet](#) [G+1](#) 0 [Pin it](#) [Share on Facebook](#) [Email](#) [Print](#)



# What are rates?

Number of cases in a given period (nominator)

---

Number of people in the population at risk in the same  
period (denominator)



# Measuring Disease Frequency

## Incidence and prevalence

Prevalence = counting all existing cases at a specified time

Incidence = counting new events in a specified time period

What are pro's and con's of both measures?

When to use these measures?



# Case fatality

- is a measure of disease severity
- is defined as the proportion of cases with a specified disease or condition who die within a specified time.

Number of deaths from diagnosed  
cases in a given period

---

Number of diagnosed cases in the same period



# Mortality and morbidity

- Mortality rate or death rate is useful for investigating diseases with a high case-fatality
- Morbidity or illness rate can be measured through
  - outpatient and primary health care consultations
  - specialist services (such as injury treatment)
  - registers of disease events.
    - hospital admissions and discharges



# Why use statistical inference?



“Data don’t make any sense,  
we will have to resort to statistics.”

# Why use statistical inference?

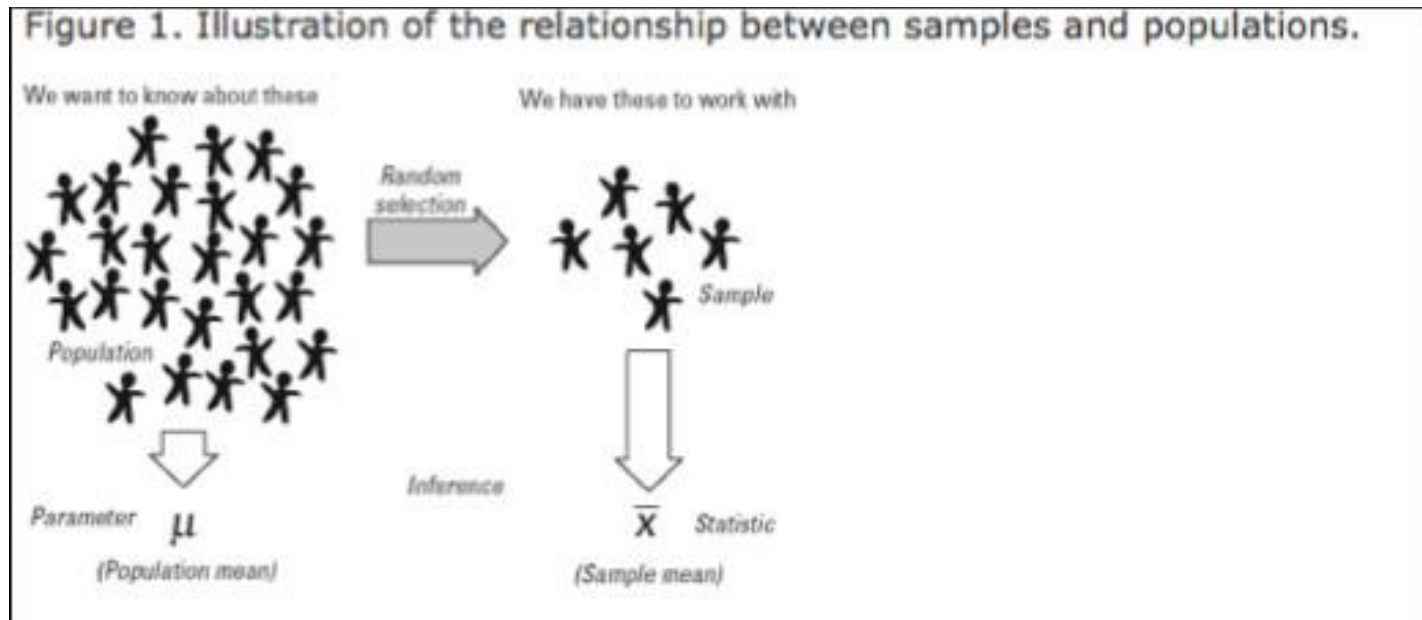
The use of samples to make inferences about populations. It is possible that the sample differs from the population and the outcome does not reflect the 'true value'.

The smaller the sample the more difficult it will be to find a statistically significant outcome

With large samples it will be easier to find a statistically significant outcome, that might not be relevant in real life

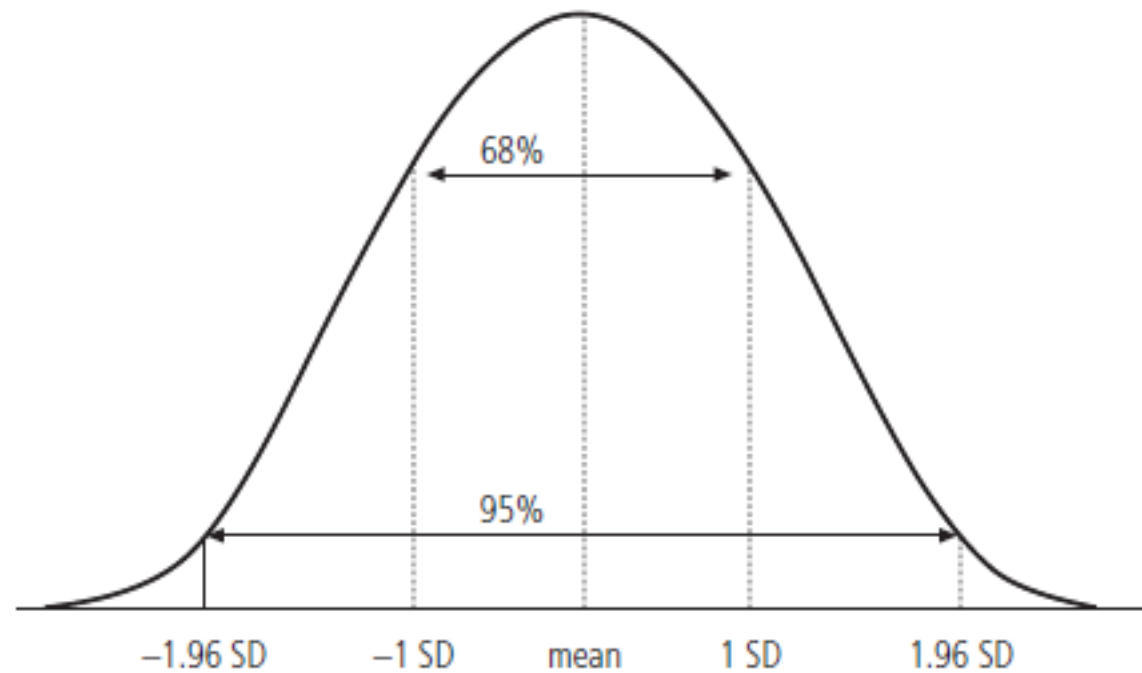


# Why use statistical inference?



# What is statistical inference?

Figure 4.3. The normal distribution curve



# What is statistical inference?

## Statistical significance

p-values are calculated and presented to indicate the degree to which an outcome has been determined by coincidence

$p < 0.05$  indicated that there is less than 5% chance that the outcome is the result of coincidental factors

$P = 0.05$  or larger: The outcome is not statistically significant



PUBLICATION BIAS

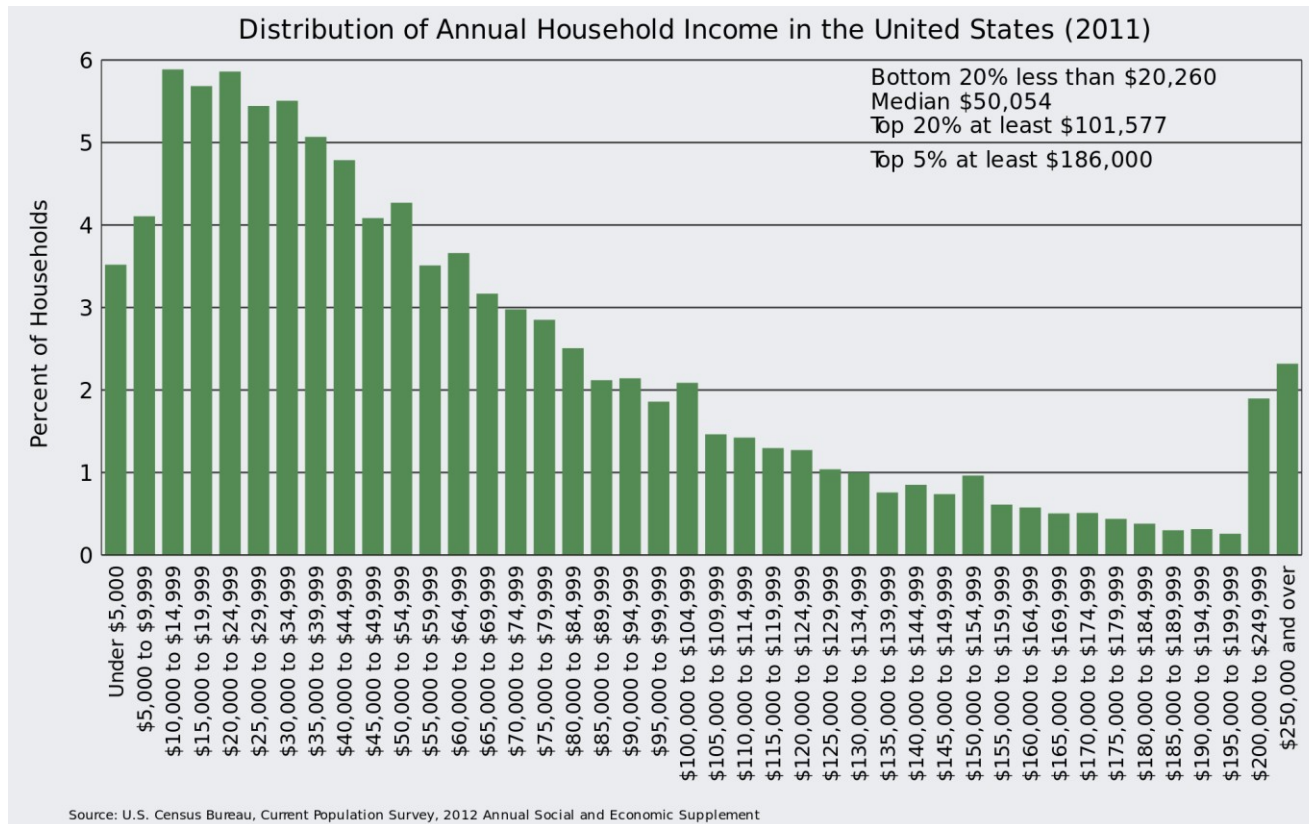
# What is statistical inference?

## 95% Confidence intervals (CI)

- Give the same information as the p-value
- the outcome = 1.3 and 95% CI = 0.9–1.8.
- the ‘true’ value of the observed outcome will lie inside this confidence interval



# What is statistical inference?



# How to present data?

## Tables, graphs and charts

Graphs and charts are:

- simple and clear
- easy to remember
- can show complex relationships

Tables are:

- Used for complex data
- Very specific or precise data



# How to present data?

Figure 3.3. Deaths during heat wave in Paris, 2003<sup>5</sup>

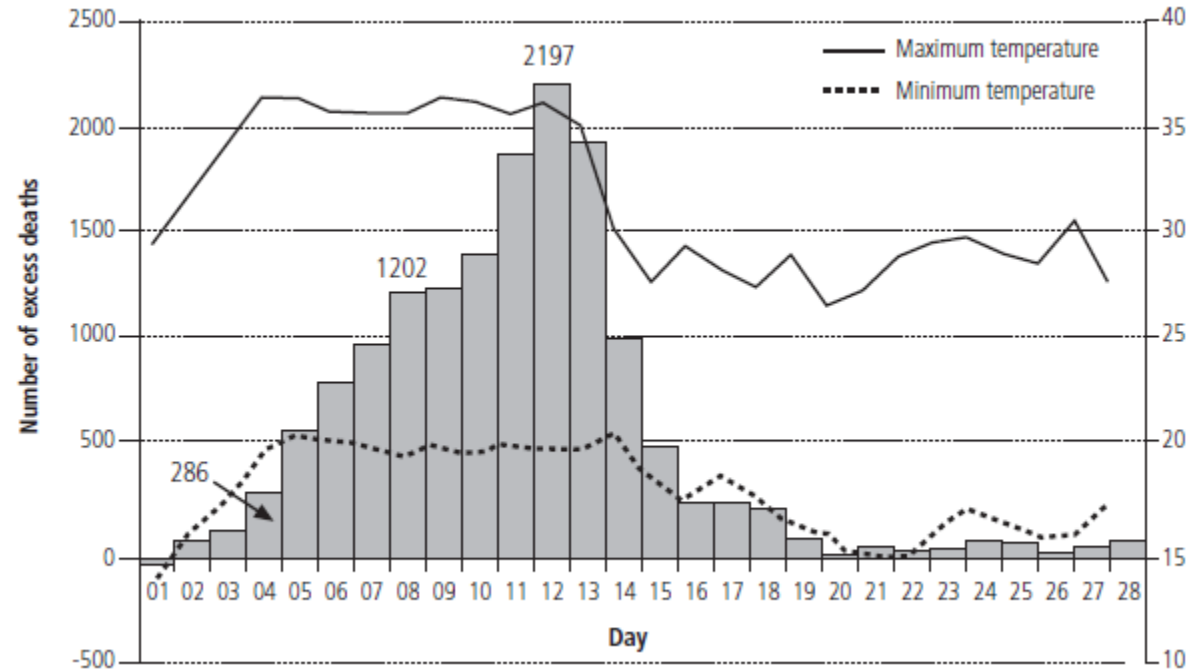


Table 2.8. Directly standardized male death rates from respiratory infections, and the ranking of five countries using three different standard populations<sup>30</sup>

Country	Age-standardized rate (per 100 000)			Ranking of countries by age-standardized rate		
	Segi	European	WHO world	Segi	European	WHO world
Australia	6.3	10.1	7.9	5	5	5
Cuba	27.2	44.2	34.6	4	4	4
Mauritius	45.2	72.6	56.6	3	3	3
Singapore	71.9	120.8	93.3	2	1	1
Turkmenistan	114.2	87.9	91.2	1	2	2



# Summary

- To interpret the number of cases you need the number of people in the population at risk and calculate the rate
- To measure case frequency you can use incidence and prevalence
- Case fatality rate is used to give information on the severity
- In low case fatality rate diseases you can use morbidity rates
- Statistical inference is used to make use of data; the results of samples are extrapolated to the population
- P-values and 95% Confidence intervals are used to deal with coincidence



# How to lie with statistics?

## Assignment for two

Please go to

<https://faculty.washington.edu/chudler/stat3.html>

And read what is written there.

Prepare in Powerpoint at least two examples of statistics that are difficult to interpret or even misleading:

- Find them in the ESPrIT repository
- Or make up your own examples

You have 30 minutes.

