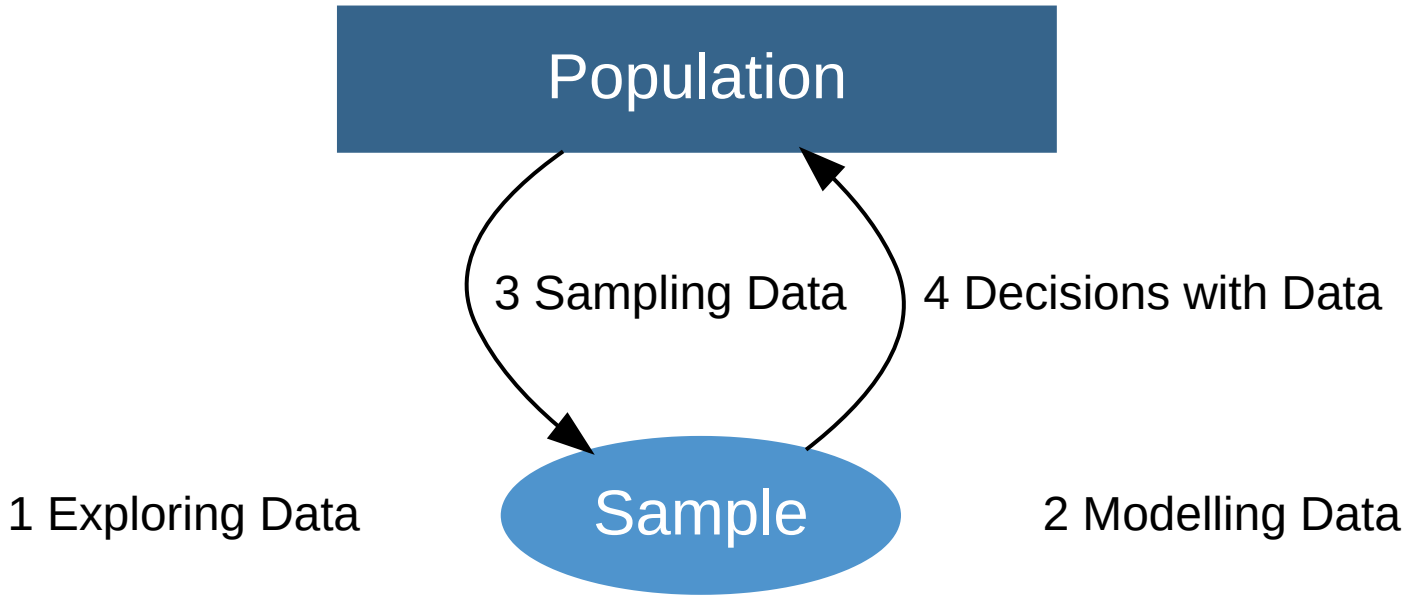
Observational Studies

Exploring Data | Design of Experiments

Unit Overview



Module1 Exploring Data

Design of Experiments

Where did the data come from & can we make reliable conclusions?

Data & Graphical Summaries

What type of data do we have & how can we visualise it?

Numerical Summaries

What are the main features of the data?

Observational Studies

Data Story | Does smoking cause cancer?

Observational Studies

4 Precautions

Summary

Data Story

Does smoking cause cancer?

Quit Victoria Prevalence Data in Motion Graphics

MARCH 13 2012

Health issue irrelevant, tobacco firms tell court

Lenore Taylor

Big tobacco companies have told the High Court they ''deny the content'' of documents lodged by the federal government making the case that smoking causes lung cancer.

MOST POPULAR

1 Crocodile attacks man in NT's Daly River region

Does smoking cause cancer?

According to the Australian Cancer Council
“Tobacco smoking is the largest preventable cause of cancer, responsible for more cancer deaths in Australia than any other single factor. It is also directly responsible for many heart and lung diseases”.

However, in a hearing to the Australian High Court in 2012 disputing the introduction of cigarette plain packaging with health warnings:

- British American Tobacco was prepared to accept that there are serious health consequences caused by smoking;
- Imperial Tobacco responded “some people say that …”

🔗 SMH

Observational Studies

The Need for Observational Studies

- In observational studies, the assignment of subjects into treatment and control groups, is outside the control of the investigator.
- Many research questions require an **observational study**, rather than a controlled experiment.
  - To study the effects of smoking, investigators cannot choose which subjects will be in the Treatment Group!
  - Most educational research uses observational studies.
- The **conclusions** of observational studies require great care.

📖 An **observational study** is one in which the investigator cannot use randomisation for allocation to groups.

4 Precautions

1. Observational studies can’t establish causation.

Study: Smoking Dramatically Increases Liver Cancer Risk

Article date: November 2, 2011

By Stacy Simon

A new study reinforces the link between liver cancer and the risk factors of smoking, obesity, and heavy drinking.

Researchers from the US and Europe studied 125 liver cancer patients to determine what risk factors were contributing to their disease. They compared them to 229 people without cancer who were matched by age, gender and other factors. The participants were all part of a European study group that was formed so researchers could investigate the role of biological, dietary, lifestyle and environmental factors in the development of cancer and other chronic diseases.

They found that almost half the cases of liver cancer in the study were associated with smoking, 16% were associated with obesity and 10% were associated with heavy alcohol consumption. Almost 21% of cases were associated with hepatitis C and 13% with hepatitis B.

The most common type of liver cancer, hepatocellular carcinoma, is a leading cause of cancer deaths worldwide. In many sub-Saharan African and Southeast Asian countries, it is the most common type of cancer. It’s less common in the United States. Worldwide, the major risk factors for liver cancer are long-term infections with hepatitis B virus and hepatitis C virus. People with these infections are more likely to develop cirrhosis, a disease in which liver cells become damaged and are replaced by scar tissue. People with cirrhosis have an increased risk of liver cancer. In the US, most liver cancer is associated with alcohol-related cirrhosis and possibly non-alcoholic fatty liver disease.

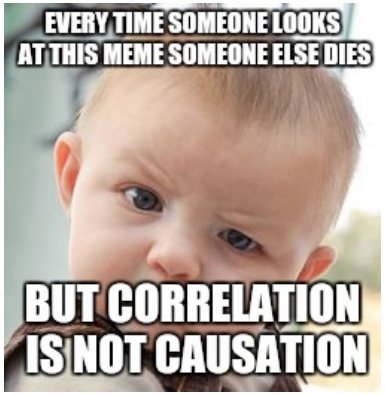
In an accompanying editorial, Morris Sherman, MD and Josep M. Llovet, MD clarify that smoking by itself does not cause liver cancer, but that it dramatically increases the risk, especially for people who have other risk factors, such as hepatitis B or C virus.

They conclude, “We should be counseling our patients who have other risk factors for hepatocellular carcinoma to quit smoking.”

🔗 Cancer Council

- Observational studies can establish association (that one thing is linked to another).
  - Association may **suggest** causation.
  - But association does not **prove** causation.

Example: Comparing smokers and non-smokers, there is a higher rate of liver cancer among the smokers. Hence, if you smoke, you are more likely to get liver cancer, but this does not imply that smoking causes liver cancer!



Statistical Thinking

What could explain the fact that smokers have a higher rate of liver cancer?

- Smokers tend to drink more alcohol than non-smokers, and excessive alcohol consumption causes liver cancer.
- So the effect of smoking is confounded (mixed-up) with the effect of alcohol consumption.

2. Observational Studies can have misleading hidden confounders.

- **Confounding** occurs when the Treatment and Control Groups differ by some third variable (other than the treatment) which influences the response that is studied.
- Confounders can be hard to find, and can mislead about a cause and effect relationship.
- Confounding (or lurking) variables can be introduced into a randomised study:
  - **selection bias**: introduced by investigators in the selection of subjects for treatment. eg select healthier subjects for surgery.
  - **survivor bias**: caused by dropout of some subjects, eg an “improvement” is due to the dropout of worst subjects who do not respond to the treatment.
  - **adherers** and **non-adherers**: subjects who adhere to the treatment/program tend to be more compliant and healthier.

🔗 Examples

Definitions

Confounding occurs when the seeming effect of the treatment is also caused by other factors.

Selection bias occurs when some subjects are more likely to be chosen to be in the study than others.

Survivor bias is when certain types of subjects finish the study.

Adherer bias occurs when certain types of subjects keep taking the treatment (or placebo), as opposed to the **non-adherers**.

Example of Adherer bias

- The Coronary Drug Project was a randomised controlled double-blind experiment conduced between between 1966-1975.
- It studied the effectiveness of Clofibrate, a compound claimed to lower abnormally high concentrations of fats and cholesterol in the blood.
- Patients taking less than 80% of prescribed Clofibrate were classified as “Non-adherers”.

Table 1. The clofibrate trial. Numbers of subjects, and percentages who died during 5 years of followup. Adherers take 80% or more of prescription.

	Clofibrate		*Placebo*	
	Number	*Deaths*	*Number*	*Deaths*
Adherers	708	15%	1,813	15%
Non-adherers	357	25%	882	28%
Total group	1,103	20%	2,789	21%

Note: Data on adherence missing for 38 subjects in the clofibrate group and 94 in the placebo group. Deaths from all causes.

Source: The Coronary Drug Project Research Group, "Influence of adherence to treatment and response of cholesterol on mortality in the Coronary Drug Project," *New England Journal of Medicine* vol. 303 (1980) pp. 1038–41.

Comparison 1 : Looks like Clofibate works

Comparison 2 : Looks like Clofibate doesn't work

Comparison 3 : Adherers are different to Non-adherers.

Source: Statistics, Freedman, Pisani, and Purves (2007), p14.

Notice:

- For the Treatment Group (Clofibrate): Comparing the lower mortality rate for Adherers (15%) to Non-adherers (25%) may suggest the drug works.
- However, comparing overall mortality rate for Treatment (20%) to Control (21%) suggests it doesn’t work.

Answer:

- Comparing the mortality rate for Adherers (15%) to Non-adherers (25%) for both Treatment and Control Groups, shows that more Non-adherers die even when they are not on the drug. Hence, this suggests that Adherers are healthier than Non-adherers and this is the reason for lower mortality rates (not the drug).
- As patients choose to be Adherers or Non-adherers (ie it is not a random allocation), this is an observational study.

Statistical Thinking

Suppose a study finds that having yellow fingertips is associated with lung cancer. Does having yellow fingertips cause lung cancer?

- Smoking is a confounding factor.
- People with yellow fingertips tend to smoke, and smoking appears to be linked to lung cancer.

A study finds that smokers tend to have higher rates of lung cancer. Does smoking cause lung cancer?

- Possible confounders are diet or alcohol or a gene for smoking.

Strategy for dealing with confounders

- Sometimes we can make the groups more comparable by dividing them into subgroups with respect to the confounder.

- If alcohol consumption is a potential confounding factor for smoking's affect on liver cancer, we can divide our subjects into 3 groups:
  - heavy drinkers
  - medium drinkers
  - light drinkers.
- This is called **controlling** for alcohol consumption.

Then we have 3 separate comparisons:

- heavy drinking: smokers vs non-smokers
- medium drinking: smokers vs non-smokers
- light drinking: smokers vs non-smokers

Statistical Thinking

What are the limitations of this strategy?

- This strategy is limited by our ability to identify all confounders and then divide the study by the confounders.
- This explains the long time to establish that smoking causes lung cancer. Researchers needed to control for factors such as health, fitness, diet, lifestyle and environment.

3. Observational studies with a confounding variable can lead to Simpson's Paradox

- **Simpson's Paradox** (or the reversing paradox) was first mentioned by British statistician Udny Yule in 1903. It was named after Edward H. Simpson.
- Sometimes there is a clear trend in **individual** groups of data that disappears when the groups are **pooled** together.
  - It occurs when relationships between percentages in subgroups are reversed when the subgroups are combined, because of a confounding or lurking variable.
  - The association between a pair of variables (X,Y) reverses sign upon conditioning of a third variable Z, regardless of the value taken by Z.

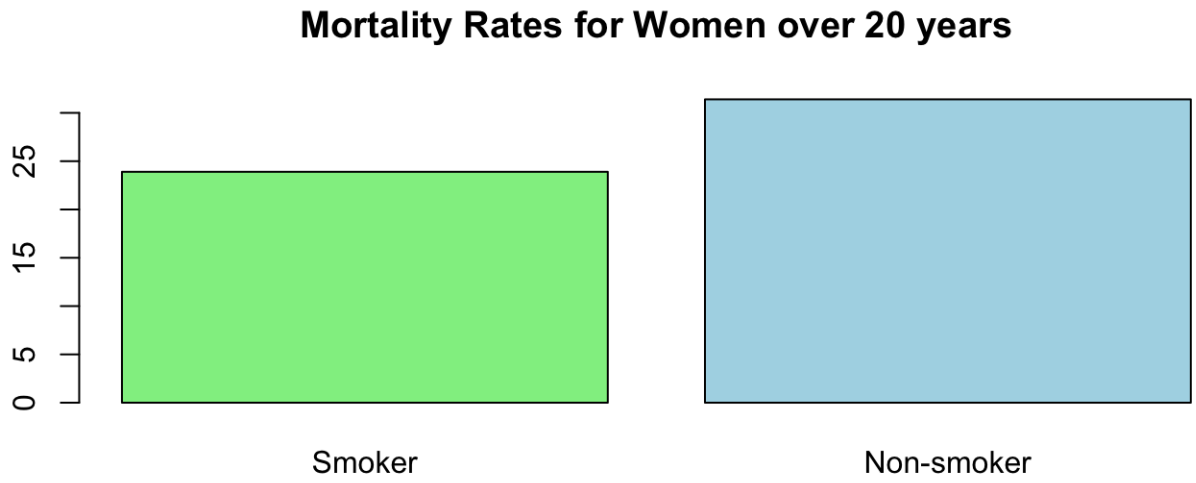
Maths: Simpson's Paradox

Simpson's Paradox and Smoking

- A famous study shows Simpson's Paradox in analysing the effect of smoking on mortality rates in women.
- The data came from 2 studies:
  - inital data from a 1 in 6 survey from an electoral roll in a mixed urban and rural area near Newcastle upon Tyne in the UK.
    - follow-up data 20 years later.
- The study concentrated on the 1314 women who were either smokers or non-smokers (in the full data, only 162 had stopped smoking and only 18 did not record their status).

Initial Results

Status	Died	Survived	Total	Mortality Rate
Smoker	139	443	582	23.9%
Non-smoker	230	502	732	31.4%
Total	369	945	1314	28.1%



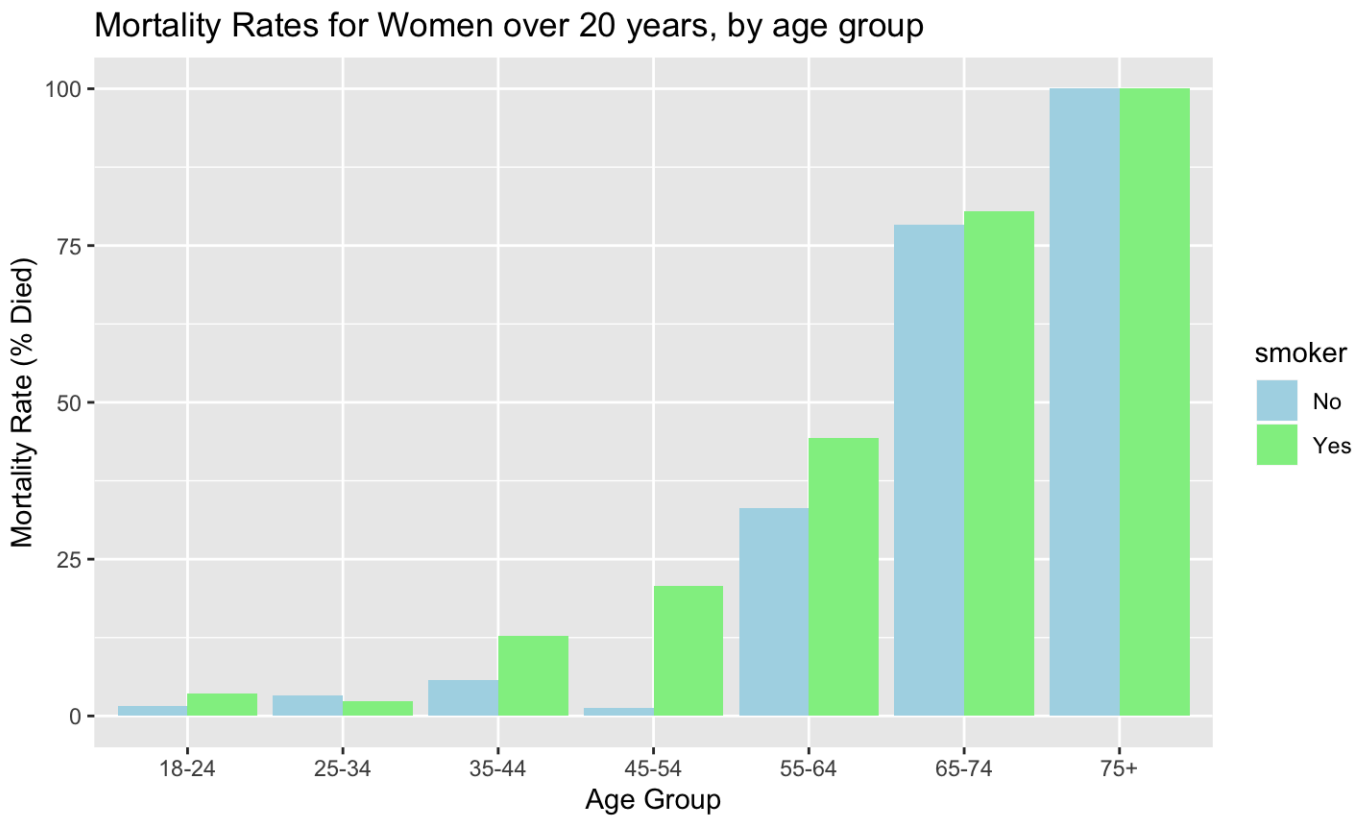
Statistical Thinking

What does this data seem to say?

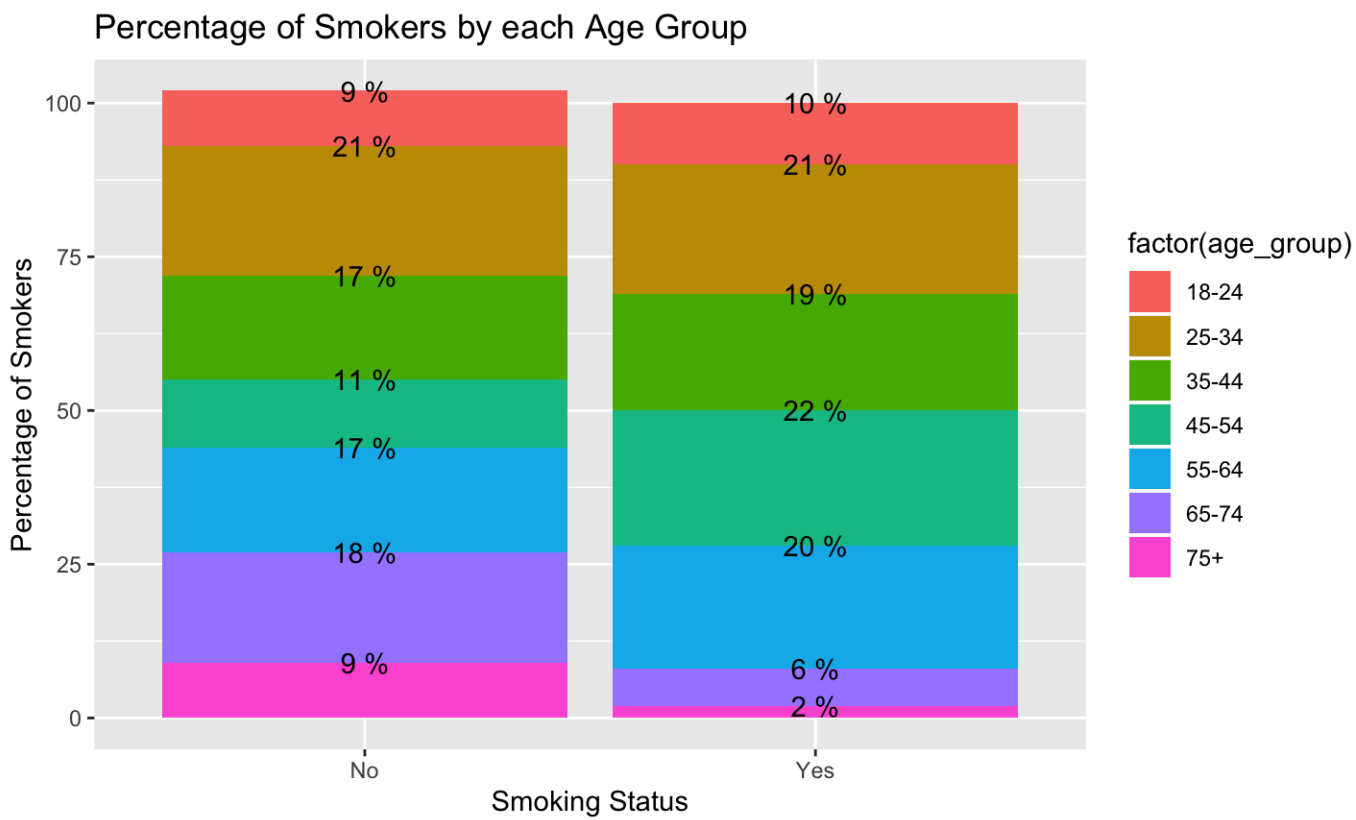
- It seems to imply that smoking has a ‘protective effect’.
- Do smokers live longer?

Now examine by age group.

Age Group	Smokers Died	Smokers Survived	Non-Smokers Died	Non-Smokers Survived
18-24	2	53	1	61
25-34	3	121	5	152
35-44	14	95	7	114
45-54	27	103	12	66
55-64	51	64	40	81
65-74	29	7	101	28
75+	13	0	64	0



Now we see that smokers have higher mortality rates for virtually every age group!


Percentage of Smokers by each Age Group
Percentage of Smokers
100
75
50
25
0
9 %
21 %
17 %
11 %
17 %
18 %
9 %
10 %
21 %
19 %
22 %
20 %
6 %
2 %
No
Yes
Smoking Status
factor(age_group)
18-24
25-34
35-44
45-54
55-64
65-74
75+

Statistical Thinking

Now what does this summary of data reveal?

- As there are many more young women who smoked than older women, and as younger women are expected to live longer than older women, adding all the groups together makes smoking appear to be beneficial.
- This is a classic example of the Simpson's Paradox phenomenon; it shows that a trend present within multiple groups can reverse when the groups are combined.

🔗 Definition 🔗 Animation 🔗 Video 🔗 More examples

4. Observational studies result from using an historical control

- Some studies present as a controlled experiment, but on further examination, there is a **historical control**, where **time** is a confounding variable.
- Investigators might compare the effect of a new medication on **current** patients, with an old medication on **past** patients. The Treatment Group (new drug) and the historical Control Group (old drug) may differ in aspects beside the treatment.
- Controlled experiments need to be performed in the same time period (contemporaneously).

Take care

There are 3 different uses of the word ‘control’:

1. A control = a subject who did not get the treatment.
2. A controlled experiment = a study where the investigators allocate subjects to the 2 groups.
3. Controlling for confounders = trying to reduce the influence of confounding variables.

Summary

Many statistical studies involve observational data and so we need to be very careful with interpretation errors such as confusing assocation for causation, misleading confounders, Simpson's Paradox and historical controls.

Key terms

observational study, association, causation, confounders, lurking variables, selection bias,survivor bias, adherer, non-adherer, controlling, Simpson's Paradox, historical control, contemporaneous control

Further Thinking

UTS Science in Focus: The Paradox of Probability

UTS