ECON10005 Quantitative Methods 1

Tutorial in Week 3

Zheng Fan

The University of Melbourne

Introduction

Zheng Fan

- Ph.D student in Economics at Unimelb

Don't be shy if you need help

- visit Ed Discussion Board
- Go for lecturer's consultation sessions: see Canvas
- In case of special considerations, consult Stop 1
- For admin issues contact Chin via qm1-economics@unimelb.edu.au
- Email me with subject code titled: fan.z@unimelb.edu.au

Important notes

Quiz and attendance mark (10%):

- To get 1 mark for each week, you have to pass the test and attend the tutorial (no half). Out of 11 weeks, you can get maximal 10 points.

Data Analysis Report (10%): Draft Task (3%)

- On Canvas: Assignment Group Registration (due by 24 March 5pm)
- Data set will be sent via Canvas announcement from me.
- Due in week 6, on 6 April 2pm Thursday

Mid-Semester Test 1 (10%):

- In week 5, 30 March Thursday at Wilson Hall (and Kwong Lee Dow)
- Keep a look out for an Canvas announcement from Chin
- Register a session from 3 time slots that suits your time

Summary

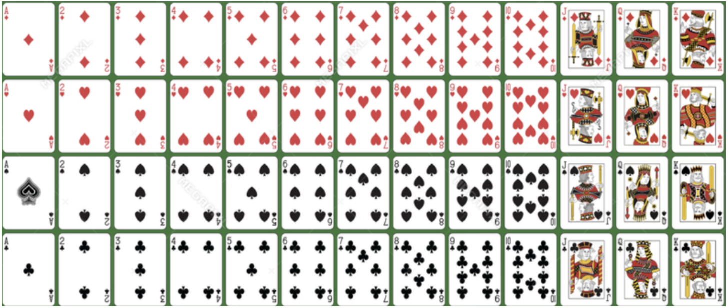
- Correlation: scatter plot, covariance, correlation $[-1, 1]$, causality
- Random experiment: sample space (set of all possible outcomes), event (collection of outcomes)
- Probability: complement, intersection, union, addition rule, joint probability, marginal probability, independence events, mutually exclusive events, conditional probability, Bayes' theorem, law of total probability

A. Mean, Standard Deviation and Correlation

Discuss the descriptive statistics from Part A of the Pre-quiz questions.

Let's review this question in Excel.

B. Probability: Poker



Events:

- A: card is Jack of Hearts
- B: card is a picture card
- C: card is red
- $D = B \cup C$ (new in addition to pre-quiz) $P(D)$, $P(A \cap D)$, $P(A \mid D)$?

B. Probability: Poker

- Events:

  A: Jack of Hearts, B: a picture card, C: red, $D = B \cup C$

- Known:

$$\begin{array}{lll} P(A) = 1/52 & P(B) = 12/52 & P(C) = 26/52 = 1/2 \\ P(A \cap B) = 1/52 & P(A \cap C) = 1/52 & P(B \cap C) = 6/52 \end{array}$$

B. Probability: Poker

- Events:

  A: Jack of Hearts, B: a picture card, C: red, $D = B \cup C$

- Known:

$$\begin{array}{lll} P(A) = 1/52 & P(B) = 12/52 & P(C) = 26/52 = 1/2 \\ P(A \cap B) = 1/52 & P(A \cap C) = 1/52 & P(B \cap C) = 6/52 \\ P(D) = 32/52 & P(A \cap D) = 1/52 & \\ P(A \mid B) = 1/12 & P(A \mid C) = 1/26 & P(A \mid D) = 1/32 \end{array}$$

B. Probability: Poker

- Events:

  A: Jack of Hearts, B: a picture card, C: red, $D = B \cup C$

- Known:

$$
\begin{array}{lll}
P(A) = 1/52 & P(B) = 12/52 & P(C) = 26/52 = 1/2 \\
P(A \cap B) = 1/52 & P(A \cap C) = 1/52 & P(B \cap C) = 6/52 \\
P(D) = 32/52 & P(A \cap D) = 1/52 & \\
P(A \mid B) = 1/12 & P(A \mid C) = 1/26 & P(A \mid D) = 1/32
\end{array}
$$

- Are the following pairs of events independent?

  A and B

B. Probability: Poker

- Events:

  A: Jack of Hearts, B: a picture card, C: red, $D = B \cup C$

- Known:

$$\begin{array}{lll} P(A) = 1/52 & P(B) = 12/52 & P(C) = 26/52 = 1/2 \\ P(A \cap B) = 1/52 & P(A \cap C) = 1/52 & P(B \cap C) = 6/52 \\ P(D) = 32/52 & P(A \cap D) = 1/52 & \\ P(A \mid B) = 1/12 & P(A \mid C) = 1/26 & P(A \mid D) = 1/32 \end{array}$$

- Are the following pairs of events independent?

  A and B (no) $P(A \cap B) \neq P(A) \cdot P(B)$

  A and C

B. Probability: Poker

- Events:
  A: Jack of Hearts, B: a picture card, C: red, $D = B \cup C$
- Known:

$$\begin{array}{lll} P(A) = 1/52 & P(B) = 12/52 & P(C) = 26/52 = 1/2 \\ P(A \cap B) = 1/52 & P(A \cap C) = 1/52 & P(B \cap C) = 6/52 \\ P(D) = 32/52 & P(A \cap D) = 1/52 & \\ P(A \mid B) = 1/12 & P(A \mid C) = 1/26 & P(A \mid D) = 1/32 \end{array}$$

- Are the following pairs of events independent?
  A and B (no) $P(A \cap B) \neq P(A) \cdot P(B)$
  A and C (no) $P(A \cap C) \neq P(A) \cdot P(C)$
  B and C

B. Probability: Poker

- Events:
  A: Jack of Hearts, B: a picture card, C: red, $D = B \cup C$
- Known:

$$\begin{array}{lll} P(A) = 1/52 & P(B) = 12/52 & P(C) = 26/52 = 1/2 \\ P(A \cap B) = 1/52 & P(A \cap C) = 1/52 & P(B \cap C) = 6/52 \\ P(D) = 32/52 & P(A \cap D) = 1/52 & \\ P(A \mid B) = 1/12 & P(A \mid C) = 1/26 & P(A \mid D) = 1/32 \end{array}$$

- Are the following pairs of events independent?

A and B (no) $P(A \cap B) \neq P(A) \cdot P(B)$

A and C (no) $P(A \cap C) \neq P(A) \cdot P(C)$

B and C (yes) $P(B \cap C) = P(B) \cdot P(C)$

A and D

B. Probability: Poker

- Events:
  A: Jack of Hearts, B: a picture card, C: red, $D = B \cup C$

- Known:

$$\begin{array}{lll} P(A) = 1/52 & P(B) = 12/52 & P(C) = 26/52 = 1/2 \\ P(A \cap B) = 1/52 & P(A \cap C) = 1/52 & P(B \cap C) = 6/52 \\ P(D) = 32/52 & P(A \cap D) = 1/52 & \\ P(A \mid B) = 1/12 & P(A \mid C) = 1/26 & P(A \mid D) = 1/32 \end{array}$$

- Are the following pairs of events independent?

A and B (no) $P(A \cap B) \neq P(A) \cdot P(B)$

A and C (no) $P(A \cap C) \neq P(A) \cdot P(C)$

B and C (yes) $P(B \cap C) = P(B) \cdot P(C)$

A and D (no) $P(A \cap D) \neq P(A) \cdot P(D)$

B. Probability: Poker

- Events:

  A: Jack of Hearts, B: a picture card, C: red, $D = B \cup C$

- Known:

$$
\begin{array}{lll}
P(A) = 1/52 & P(B) = 12/52 & P(C) = 26/52 = 1/2 \\
P(A \cap B) = 1/52 & P(A \cap C) = 1/52 & P(B \cap C) = 6/52 \\
P(D) = 32/52 & P(A \cap D) = 1/52 & \\
P(A \mid B) = 1/12 & P(A \mid C) = 1/26 & P(A \mid D) = 1/32
\end{array}
$$

- Discuss which of following would be "most informative" about A

  B has occurred; C has occurred; D has occurred

B. Probability: Poker

- Events:
  A: Jack of Hearts, B: a picture card, C: red, $D = B \cup C$

- Known:

$$
\begin{array}{lll}
P(A) = 1/52 & P(B) = 12/52 & P(C) = 26/52 = 1/2 \\
P(A \cap B) = 1/52 & P(A \cap C) = 1/52 & P(B \cap C) = 6/52 \\
P(D) = 32/52 & P(A \cap D) = 1/52 & \\
P(A \mid B) = 1/12 \ > & P(A \mid C) = 1/26 \ > & P(A \mid D) = 1/32
\end{array}
$$

- Discuss which of following would be "most informative" about A

  B has occurred; C has occurred; D has occurred

C. Probability: COVID-19 Test

1. Suppose that there are 15,000 active cases of COVID-19 in Melbourne, out of a population of 5 million people. Now imagine we select a Melbourne resident at random and give them a test for COVID-19, and it comes back positive. What is the probability that this person actually has COVID-19? For the purposes of your calculation that the test has false positive rate of 1 out of 1000 and a false negative rate of 0.2.

- Events:

C. Probability: COVID-19 Test

1. Suppose that there are 15,000 active cases of COVID-19 in Melbourne, out of a population of 5 million people. Now imagine we select a Melbourne resident at random and give them a test for COVID-19, and it comes back positive. What is the probability that this person actually has COVID-19? For the purposes of your calculation that the test has false positive rate of 1 out of 1000 and a false negative rate of 0.2.

- Events:

$$\begin{array}{ll} C = \text{has COVID} & \bar{C} = \text{no COVID} \\ T = \text{tests positive} & \bar{T} = \text{negative} \end{array}$$

C. Probability: COVID-19 Test

1. Suppose that there are 15,000 active cases of COVID-19 in Melbourne, out of a population of 5 million people. Now imagine we select a Melbourne resident at random and give them a test for COVID-19, and it comes back positive. What is the probability that this person actually has COVID-19? For the purposes of your calculation that the test has false positive rate of 1 out of 1000 and a false negative rate of 0.2.

- Base rate:

C. Probability: COVID-19 Test

1. Suppose that there are 15,000 active cases of COVID-19 in Melbourne, out of a population of 5 million people. Now imagine we select a Melbourne resident at random and give them a test for COVID-19, and it comes back positive. What is the probability that this person actually has COVID-19? For the purposes of your calculation that the test has false positive rate of 1 out of 1000 and a false negative rate of 0.2.

- Base rate:

Active cases	Population	P(C)	$P(\bar{C})$
15,000	5,000,000	0.003	0.997

C. Probability: COVID-19 Test

1. Suppose that there are 15,000 active cases of COVID-19 in Melbourne, out of a population of 5 million people. Now imagine we select a Melbourne resident at random and give them a test for COVID-19, and it comes back positive. What is the probability that this person actually has COVID-19? For the purposes of your calculation that the test has false positive rate of 1 out of 1000 and a false negative rate of 0.2.

- False positive rate:

C. Probability: COVID-19 Test

1. Suppose that there are 15,000 active cases of COVID-19 in Melbourne, out of a population of 5 million people. Now imagine we select a Melbourne resident at random and give them a test for COVID-19, and it comes back positive. What is the probability that this person actually has COVID-19? For the purposes of your calculation that the test has false positive rate of 1 out of 1000 and a false negative rate of 0.2.

- False positive rate:

$$P(T|\bar{C}) = 0.001 \qquad P(\bar{T}|\bar{C}) = 0.999$$

C. Probability: COVID-19 Test

1. Suppose that there are 15,000 active cases of COVID-19 in Melbourne, out of a population of 5 million people. Now imagine we select a Melbourne resident at random and give them a test for COVID-19, and it comes back positive. What is the probability that this person actually has COVID-19? For the purposes of your calculation that the test has false positive rate of 1 out of 1000 and a false negative rate of 0.2.

- False negative rate:

C. Probability: COVID-19 Test

1. Suppose that there are 15,000 active cases of COVID-19 in Melbourne, out of a population of 5 million people. Now imagine we select a Melbourne resident at random and give them a test for COVID-19, and it comes back positive. What is the probability that this person actually has COVID-19? For the purposes of your calculation that the test has false positive rate of 1 out of 1000 and a false negative rate of 0.2.

- False negative rate:

$$P(\bar{T}|C) = 0.2 \qquad P(T|C) = 0.8$$

C. Probability: COVID-19 Test

1. Collection of all the key information required

- Events:

$$C = \text{has COVID} \qquad \bar{C} = \text{no COVID}$$
$$T = \text{tests positive} \qquad \bar{T} = \text{negative}$$

- Base rate:

Active cases	Population	P(C)	$P(\bar{C})$
15,000	5,000,000	0.003	0.997

- False positive rate:

$$P(T|\bar{C}) = 0.001 \qquad P(\bar{T}|\bar{C}) = 0.999$$

- False negative rate:

$$P(\bar{T}|C) = 0.2 \qquad P(T|C) = 0.8$$

- Now we have everything to compute $P(C|T) = ?$

C. Probability: COVID-19 Test

1. (continued) Now we have everything to compute $P(C|T) =?$

$$
\begin{aligned}
P(C|T) &= \frac{P(T|C)P(C)}{P(T)} && \text{(Bayes' rule)} \\
&= \frac{P(T|C)P(C)}{P(T|C)P(C) + P(T|\bar{C})P(\bar{C})} && \text{(LoTP)} \\
&= \frac{0.8 \cdot 0.003}{0.8 \cdot 0.003 + 0.001 \cdot 0.997} \\
&= 0.707
\end{aligned}
$$

C. Probability: COVID-19 Test

2. How does your answer to question 1 change if there are actually 30,000 active COVID-19 cases?

- Events:

$$C = \text{has COVID} \qquad \bar{C} = \text{no COVID}$$
$$T = \text{tests positive} \qquad \bar{T} = \text{negative}$$

- Base rate:

Active cases	Population	P(C)	$P(\bar{C})$
30,000 ~~(15,000)~~	5,000,000	0.006	0.994

- False positive rate:

$$P(T|\bar{C}) = 0.001 \qquad P(\bar{T}|\bar{C}) = 0.999$$

- False negative rate:

$$P(\bar{T}|C) = 0.2 \qquad P(T|C) = 0.8$$

C. Probability: COVID-19 Test

2. (continued) Now we have everything to compute $P(C|T)$

$$
\begin{aligned}
P(C|T) &= \frac{P(T|C)P(C)}{P(T)} && \text{(Bayes' rule)} \\
&= \frac{P(T|C)P(C)}{P(T|C)P(C) + P(T|\bar{C})P(\bar{C})} && \text{(LoTP)} \\
&= \frac{0.8 \cdot 0.006}{0.8 \cdot 0.006 + 0.001 \cdot 0.994} \\
&= 0.828
\end{aligned}
$$

C. Probability: COVID-19 Test

3. How does your answer to question 1 change if the false positive rate is only 1 out of 5000?

- Events:

$$C = \text{has COVID} \qquad \bar{C} = \text{no COVID}$$
$$T = \text{tests positive} \qquad \bar{T} = \text{negative}$$

- Base rate:

Active cases	Population	P(C)	$P(\bar{C})$
15000	5000000	0.003	0.997

- False positive rate:

$$P(T|\bar{C}) = 0.0002 \ \cancel{(0.001)} \qquad P(\bar{T}|\bar{C}) = 0.9998$$

- False negative rate:

$$P(\bar{T}|C) = 0.2 \qquad P(T|C) = 0.8$$

C. Probability: COVID-19 Test

3. (continued) Now we have everything to compute $P(C|T)$

$$
\begin{aligned}
P(C|T) &= \frac{P(T|C)P(C)}{P(T)} && \text{(Bayes' rule)} \\
&= \frac{P(T|C)P(C)}{P(T|C)P(C) + P(T|\bar{C})P(\bar{C})} && \text{(LoTP)} \\
&= \frac{0.8 \cdot 0.003}{0.8 \cdot 0.003 + 0.0002 \cdot 0.997} \\
&= 0.923
\end{aligned}
$$

C. Probability: COVID-19 Test

4. Instead of applying the test as a "screening test" by randomly selecting from the entire Melbourne population, suppose the test is targeted only at those who have symptoms of respiratory tract infections (eg. cold, flu, allergy etc as well as COVID-19). Suppose that there are 500,000 people in Melbourne with such symptoms. How does this change your answer to question 1?

- Events:

$$C = \text{has COVID} \qquad \bar{C} = \text{no COVID}$$
$$T = \text{tests positive} \qquad \bar{T} = \text{negative}$$

- Base rate:

Active cases	Population	P(C)	$P(\bar{C})$
15,000	500,000 ~~(5,000,000)~~	0.03	0.97

C. Probability: COVID-19 Test

4. (continued)

- False positive rate:

$$P(T|\bar{C}) = 0.001 \quad P(\bar{T}|\bar{C}) = 0.999$$

- False negative rate:

$$P(\bar{T}|C) = 0.2 \quad P(T|C) = 0.8$$

- Now we have everything to compute $P(C|T)$

$$\begin{aligned} P(C|T) &= \frac{P(T|C)P(C)}{P(T)} && \text{(Bayes' rule)} \\ &= \frac{P(T|C)P(C)}{P(T|C)P(C) + P(T|\bar{C})P(\bar{C})} && \text{(LoTP)} \\ &= \frac{0.8 \cdot 0.03}{0.8 \cdot 0.03 + 0.001 \cdot 0.97} = 0.961 \end{aligned}$$

The end

Thanks for your attention! 😄

Feel free to leave and see you next week!