

17. 統計的應用及誤用 (Uses and Abuses of Statistics)

- 參照會考嘅出題模式，依一課根本上可以話係“唔會考”嘅。
 - 原因好簡單：因為依一課只係齋吹水、根本冇數計。

17.1. 認識抽取調查樣本的不同技巧及製作問卷的基本原則

(Recognising Different Techniques in Survey Sampling and the Basic Principles of Questionnaire Design)

- 相信大家成日都會聽到好多由政府、商業機構發表嘅統計或調查結果。
 - 例如香港家庭嘅收入中位數字、平均工作時數、各電視台嘅收視率等。
 - 依 D 統計通常都俾到我哋一個印象就係調查對象嘅全體（即總體）係“全香港”。
 - 但係人都知道依 D 調查只係訪問咗一部份人之後算。
 - ◆ 依個從總體中選取一部份人作樣本接受調查就叫“抽取調查樣本”。

17.1.1. 抽取調查樣本的不同技巧

- “抽取調查樣本”嘅技巧可以分成兩大類：
 - 概率抽樣
 - ◆ 喺概率抽樣入面，總體中每個個體被納入樣本嘅機會都係均等嘅。
 - 非概率抽樣
 - ◆ 非概率抽樣係“根據主觀判斷及經驗選取樣本，而非以隨機的方式選取樣本”。
 - 留意“總體中每個個體被納入樣本嘅機會都係均等”依句說話其實有時都要睇你嘅“調查嘅標題”同“點抽樣”等決定到。
 - ◆ 假設我想做一個“香港人是否贊成香港政府提高酒稅”嘅調查。
 - ◆ 咁如果我只係喺夜晚企喺蘭柱芳度“隨機咁”訪問經過嘅香港人，你話我個抽樣方法係唔係“概率抽樣”？好明顯唔係，因為夜晚會去蘭柱芳嘅人多數都會飲酒，而提高酒稅會對佢哋有較大影響，所以樣本除咗唔夠代表性之外仲係有偏見嘅。
 - ◆ 但假如我改為隨機咁打電話訪問香港人，咁就可以話係“概率抽樣”。

非概率抽樣

- 常見嘅“非概率抽樣”方法有：
 - 街頭方問
 - ◆ 喺街頭方問當中，訪問員通常會選取一些看來較「友善」的途人進行訪問。
 - 自發性調查（包括電話、網上）
 - ◆ 所有自發性嘅調查通常只會係嗰 D 對調查題目有較強烈意見嘅先會主動參加。
- “非概率抽樣”一般用途不大（較大型及有公信力嘅調查都唔會用佢）。
 - 咁係因為抽樣方法唔科學化、而調查嘅結果通常會存在偏差、可能不準確、並帶有誤導成分。

概率抽樣

- 常見嘅“非概率抽樣”方法有：
 - 簡單隨機抽樣法。簡單隨機抽樣法嘅做法如下：
 - ◆ 總體入面每個個體應先有一個獨一無二嘅編號。
 - 例如如果總體係香港成年人，咁我哋就可以用身份證號碼。
 - ◆ 之後隨機咁抽出千個編號。
 - ◆ 把對應隨機抽出編號嘅個體納入樣本之中。
 - 系統隨機抽樣法（或稱為等距抽樣法）。
 - ◆ 喺等距抽樣法入面，我哋首先選取一個隨機起點，並按照固定區間抽取個體，從而組成一個樣本。
 - ◆ 例如公廠要監控生產產品嘅品質，為咗減低成本工廠唔會件件貨都 check。工廠可能會由每日第 20 個製成品、每隔 300 個就抽一個出嚟做樣本檢測。
 - 層狀隨機抽樣法。
 - ◆ 喺層狀隨機抽樣法入面，總體會根據分層標準被分成不同的層。然後利用簡單隨機抽樣，從各層按比例抽取樣本。
 - ◆ 例如為確保統計包括同等數目嘅男女數目，我哋可以先將總體先分成男、女兩組，然後再用簡單隨機抽樣法由兩組中抽出樣本。

17.1.2. 製作問卷的基本原則

- 一份問卷做得好唔好會直接咁影響到結果嘅可信度及有效性。
- 而製作問卷嘅基本原則包括：
 - 問題要切合統計調查嘅目的
 - 問卷不應太長
 - 用語方面，措詞要適當和用字要精確
 - ◆ 例如用“你強烈反對政府提高煙草稅嗎？”同“你贊成政府提高煙草稅？”得出嚟嘅調查結果可能好唔同。因為有 D 人可能會覺得“雖然自己反對，但未去到強烈反對”，所以就算自己反對政府加煙草稅都會對第一條問題答“否”。
 - 問題嘅形式要適合
 - ◆ 例如想知道受訪者“有幾強烈反對或贊成政府加煙草稅”，我哋應該用“評程度評分，如 1 代表強烈反對、9 代表強烈贊成”而唔係用“是非題”。
 - 問題的次序編排要合理
 - 避免問題帶有引導被訪者取向或偏重某一回應的成分
 - 避免使用複合或帶有雙重反義的發問方式
 - ◆ 例如“你不贊成政府不提高煙酒稅嗎？”應寫成“你贊成政府提高煙酒稅？”
 - 問卷定稿前，應以一部份可能被訪人士為對象進行測試
 - ◆ 從中可能會發現到受訪者不明白某些詞彙或問題