

Inlämningsuppgift 3: Statistisk undersökning

22 maj 2022

I den här inlämningsuppgiften ska du genomföra en statistisk undersökning. Det börjar med att fundera över, vad det är du vill undersöka, planera experimentet och analysen, samla in data och sedan dra slutsatser med hjälp av metoderna du lärde dig i kursen.

Planering

1. Välj vad du vill undersöka och hur du vill analysera datan samt hur du ska samla in datan/ta ett stickprov ur din population:
 - Vilken population har du i åtanke och är stickprovet representativt för denna?
 - Vilken parameter/vilka parametrar av fördelningen i populationen vill du skatta eller dra slutsatser om?
 - Är datan tillräcklig omfattande/enkel för att genomföra din planerad analys?
 - Vilken typ av data kommer du behöva (kategorisk, numerisk, både och)?
2. Stickprovet borde bestå av oberoende observationer. Hur kan du samla in datan, så att det är ett rimligt antagande?
3. För att kunna applicera våra metoder är det nödvändigt att antingen datan själv är normalfördelad eller stickprovet tillräckligt stort så att centrala gränsvärdessatsen kan användas. Motivera dina antaganden! Fundera lite extra om din idé duger ur denna synvinkel, då analysen kommer vara ointressant/för svår om datamängden inte lämpar sig. Du får gärna stämma av med mig innan du sätter igång om ditt upplägg är rimligt.
4. Bestäm hur du vill lägga upp och genomföra analysen:
 - Fundera över vilka slutsatser du vill dra.
 - Har du några förutfattade meningar om vilka värden för parametern eller vilka fördelningsantaganden som är rimliga (\rightarrow särskilt viktigt för hypotesprövning) innan du genomför experimentet och samlar in data?
 - Välj en rimlig konfidensgrad (för ditt konfidensintervall) eller signifikansnivå (för din hypotesprövning).
 - Om din analys bygger på antagandet att populationen själv är normalfördelad är det nödvändigt att motivera/ bekräfta detta (t.ex. med ett stapeldiagram av den insamlade datan).

Genomförandet

1. Samla in datan. Ta hänsyn till orimliga värden/svar och värden som saknas (ofullständiga svar t.ex.).
2. Börja analysen:
 - Hur mycket tycker du att inkludering/exkludering av saknade eller uppenbarligen felaktiga svar/värden kan påverka resultatet?
 - Borde man ifrågesätta antaganden du gjorde i planeringen (så som oberoende eller normalfördelade mätvärden)? Hur mycket tror du att det kan påverka analysen/resultaten i så fall?
3. Dra slutsatser: Räkna fram konfidensintervall, teststatistika samt P -värden och formulera dina slutsatser och resultat med två olika läsarkretsar i åtanke:
 - En rapport för någon med statistisk kunskap som kan relatera till motsvarande begrepp och analysmetoder och
 - en sammanfattning för någon som inte har någon som helst förkunskap om statistiska metoder.

Ledtrådar

Det viktigaste med uppgiften är att göra en korrekt analys, inte att gå in på en extra klurig fråga. Därför är det viktigare att spendera en del tid på planering och förberedelser istället för att genomföra ett orimligt stort antal experiment/intervjuer.

Saker du kan tänka på:

- (a) Du har kanske redan från början en viss population i åtanke. Det är viktigt att den är varken för omfattande eller för trång (vill du dra slutsatser om en vanlig Chalmers student är det varken lämpligt att fråga GU-studenter eller bara kurskamrater).
- (b) Stickprovet ska vara representativt för populationen (t.ex. till lika stora delar kvinnor och män - om inte kvinnliga studenter är din population).
- (c) Se till att du har ett tillräckligt stort stickprov för att kunna genomföra din analys (kanske några mätvärden måste förkastas) och motivera alla antaganden (att värden kommer från samma fördelning, är oberoende etc.)

- (d) Om du intervjuar: Säkerställ att dina frågor är tydliga, otvetydiga och neutrala så att de inte leder personen som svarar i en viss riktning. Dubbelkolla med någon kamrat om ditt frågeformulär duger innan du släpper det. Kom ihåg att du kan ladda upp frågeformuläret på discord eller kursens Canvas-sida (under Diskussioner) för att samla in svar.
- (e) Det är viktigt att genomtänka analysen och metoderna du vill använda innan du samlar in data (t.ex. lämpar sig ett konfidensintervall för att skatta parametervärden, ett χ^2 -test för att hypotespröva om datan verkligen är likformigt fördelad; vill du jämföra olika populationer behöver du ett stickprov ur vardera - inte alla metoder fungerar med alla datamängder helt enkelt).
- (f) Börja tidigt med att skriva på rapporten och försök använda det som verktyg och vägvisare för att planera, strukturera och genomföra både ditt experiment och din analys.
- (g) Din rapport borde innehålla:
- frågeställningen som du försöker besvara
 - metoder och verktyg du använder dig av (inte ju fler desto bättre, men minst två olika ska det vara, t.ex. kan du välja ifrån konfidensintervall för medelvärde, varians, proportioner, hypotesprövning av olika slag, jämförelse av olika populationer osv.)
 - en beskrivning av populationen och motsvarande antaganden
 - en beskrivning av datainsamlingen (t.ex. intervjuprocessen eller frågeformuläret)
 - en väldokumenterad beräkning av motsvarande testvariabler och konfidensintervall
 - en interpretation av resultaten
 - ett ställningstagande till problem med datainsamlingen/undersökningen som skulle kunna påverka resultaten och därmed slutsatserna
 - en kort sammanfattning för en okunnige.