

Tentamen i Undersökningsmetodik (3.5 hp)

Kurs: Statistik och dataanalys II

2023-08-15

Skrivtid:	kl. 9.00 - 13.00 (4 timmar)
Godkända hjälpmedel:	Miniräknare utan lagrade formler och lagrad text
Vidhäftade hjälpmedel:	inga

- Tentamen består av två uppgifter uppdelade i deluppgifter. Maximalt antal poäng anges per deluppgift.
 - För var och en av deluppgifterna som följer i Uppgift 1 och 2 behöver du inte skriva långt och mycket, bara det viktigaste. Som vägledning kan du sikta på ca $\frac{1}{2}$ - 1 A4 per deluppgift, max! Inga beräkningar behövs, men du kan illustrera med bilder och skisser om det underlättar. Även formler, dvs. matematisk framställning, går bra om du tycker att det är relevant.
 - Maxpoäng är 100 och för godkänt resultat krävs minst 50 poäng. Betygsgränser:
 - A: 90 – 100
 - B: 80 – 89
 - C: 70 – 79
 - D: 60 – 69
 - E: 50 – 59
 - Fx: 40 – 49
 - F: 0 – 40
- OBS! Fx och F är underkända betyg som kräver omexamination. Studenter som får betyget Fx kan alltså inte komplettera för högre betyg.
- Svar med fullständiga redovisningar och motiveringar ska lämnas.
 - Använd endast skrivpapper som tillhandahålls i skrivsalen.
 - Max en uppgift per blad, t.ex. uppgift 1a-d på ett eller flera blad men börja på nytt blad för nästa uppgift 2a-d osv.
 - För full poäng på en uppgift krävs tydliga, utförliga och väl motiverade förklaringar.
 - Kontrollera dina svar! Slarvfel kan ge poängavdrag!
 - Lösningförslag läggs ut på Athena kort efter tentamen.

LYCKA TILL!

Uppgift 1. (50p)

- a) Förklara vilka villkor som måste vara uppfyllda för att en urvalsdesign ska vara ett sannolikhetsurval. Ge ett exempel på en urvalsdesign som inte är ett sannolikhetsurval (vi gick igenom tre sådana i samband med undervisningen) och förklara varför det inte är ett sannolikhetsurval. Det går bra att hitta på något eget exempel också om du vill. (10p)
- b) Förklara vad CASM-modellen handlar om och beskriv kortfattat de fyra stegen som definierar modellen. Ge sedan exempel på en s.k. *stabil attitydfråga* och en *temporär attitydfråga* och hur modellen måste justeras när man har den senare av de två. (10p)
- c) Förklara begreppet *mätfel* och varför man vill undvika det. Förklara kortfattat begreppen *reliabilitet* och *validitet*. Ge slutligen två exempel på hur mätfel kan uppstå. (10p)
- d) SNI, NACE, SSYK, SUN och SEI är exempel på *klassifikationsstandarder*. Välj ut två av dessa, ange vad akronymen (dvs. förkortningen) står för, beskriv sedan kortfattat vad de beskriver. Ange två fördelar med att använda standarder för klassificering. Välj ut två av dessa, ange vad akronymen står för och beskriv kortfattat vad de avser (du behöver inte detaljförklara hur de är uppbyggda). (10p)
- e) Det finns ett antal s.k. *kvalitetskriterier* som enligt lag ska beaktas och deklarerats i Sveriges officiella statistik. Välj två av dessa kriterier och beskriv dem kortfattat. Finns det någon målkonflikt mellan de två kriterierna? Motivera ditt svar. (10p)

Uppgift 2. (50p)

Ett konsultföretag där du arbetar genomförde under år 2022 en undersökning kring sexuella trakasserier på arbetsplatsen bland jurister (inklusive advokater, biträdande jurister mm.). Man utgick ifrån en lista sammanställd av Advokatsamfundet år 2017 bestående av $N = 12\,000$ jurister där 30% var kvinnor.

Från listan drogs ett sannolikhetsurval bestående av $n = 1000$ jurister som alla fick en enda enkel fråga att besvara: "Förekommer det sexuella trakasserier på din arbetsplats?". Datainsamlingen skedde telefonintervjuer. Av de tillfrågade var det $n_s = 200$ som svarade på frågan, man fick alltså 80% bortfall.

Svaren redovisades sedan efter kön (dvs. den juridiska definitionen av kön; kvinna eller man) och svaret på frågan (Ja/Nej). Följande tabell redovisar resultaten som kom in:

	Ja	Nej	Totalt
Kvinnor	60	60	120
Män	20	60	80
Totalt	80	120	200

- Vad skulle du säga är målpopulation för den här undersökningen? Vad är rampopulationen? Förklara kortfattat vad som menas med täckningsfel och förklara om det kan finnas täckningsfel i den här undersökningen? Motivera! (10p)
- I statistiska undersökningar brukar man skilja på primärdata och sekundärdata. Förklara dessa begrepp kortfattat! Hur skulle du karaktärisera de två redovisade variablerna ovan, vilken eller vilka är primär- respektive sekundärdata? Ange slutligen en fördel och en nackdel med primär- respektive sekundärdata. (10p)
- Reflektera över hur frågan är formulerad och vilka möjliga svarsalternativ som fanns att välja bland. Kan frågeformuleringen och möjliga svarsalternativ resultera i ett missvisande resultat? Hur skulle du konstruera frågan och svarsalternativen? Motivera! (10p)
- Du noterar att 60% (120 av 200) av de svarande är kvinnor men att endast 30% i listan var kvinnor. Vad kan skillnaden bero på? Ange minst två möjliga orsaker och förklara om de kan resultera i snedvridna resultat och slutsatser. (10p)
- En kollega till dig vill gå ut med ett pressmeddelande som säger att det förekommer sexuella trakasserier på 40% (80 av 200) av advokatbyråerna i Sverige. Förklara för din kollega som saknar utbildning i statistik att det finns minst tre fel med den slutsatsen. Här räcker det om du kan ange två problem. (10p)