

Tentamen i Undersökningsmetodik (3.5 hp)

Kurs: Statistik och dataanalys II

2024-01-10

Skrivtid: kl. 8.00 - 12.00 (4 timmar)

Godkända hjälpmedel: inga

Vidhäftade hjälpmedel: inga

- Tentamen består av tre uppgifter uppdelade i deluppgifter. Maximalt antal poäng anges per deluppgift.
 - OBS! För var och en av deluppgifterna 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. (45p)

- a) Ge tre skäl till varför urvalsundersökningar ofta är ett bättre alternativ än en *totalundersökning*. Förklara sedan vilka villkor som måste vara uppfyllda för att en urvalsdesign ska vara ett *sannolikhetsurval*. Förklara slutligen kortfattat varför man ofta föredrar sannolikhetsurval framför icke-sannolikhetsurval. (15p)
- b) Förklara begreppen *målpopulation* och *rampopulation*. Vilka problem kan uppstå om de inte är identiska? Förklara genom exempel varför de ofta inte är identiska. (10p)
- c) Man brukar i samband med statistiska undersökningar tala om *data*, *metadata* och *paradata*. Förklara dessa begrepp och hur de ska användas, gärna genom exempel. (10p)
- d) Förklara för en som inte kan så mycket statistik vad som menas med *primär-* respektive *sekundärdata* och ange en för- och en nackdel med att kombinera datakällor. Förklara även fördelarna med *klassifikationsstandarder* i detta sammanhang. (10p)

Uppgift 2. (35p)

- a) Definiera vad som menas med *partiellt bortfall* respektive *objektsbortfall*. Förklara kortfattat varför bortfall är ett problem. Ge sedan två exempel på åtgärder som man kan ta till för att förebygga eller hantera bortfallet och förklara kortfattat hur dessa åtgärder kan minska bortfallet eller mildra dess effekter. (15p)
- b) Förklara hur begreppet *TSE* definieras. Skulle du påstå att det är lätt eller svårt att mäta och bedöma TSE i en undersökning? Förklara! OBS! Du behöver inte i detalj förklara varje enskild del av TSE men lista dem gärna. (10p)
- c) Förklara vad som menas med *statistisk röjandekontroll*. Ge ett exempel på en metod man kan använda för att minska risken för röjande. Vad kan konsekvenserna bli om statistikproducenter inte beaktar behovet av statistisk röjandekontroll i sin verksamhet? (10p)

Uppgift 3. (20p)

Inför införandet av euron som gemensam valuta inom EU genomförde ett undersökningsinstitut i Sverige en metodstudie, gissningsvis någon gång före år 2001, för att studera hur olika frågeformuleringar kan påverka resultaten. Två olika versioner av "samma" fråga ställdes, gissningsvis till två olika grupper av individer:

Version A:

Inom EU arbetar man för att bilda en valutaunion EMU, med gemensam valuta för de länder som är med i EU. Är du för eller emot att Sverige går med i EMU?

Svarsfördelningen blev: "För" 38 %, "Emot" 48 %, "Vet ej" 13 %.

Version B:

Sannolikt kommer Belgien, Holland, Luxemburg, Italien, Portugal, Spanien, Irland, Frankrike, Österrike, Tyskland och Finland gå med i EMU från start. Om det blir så, tycker du då att Sverige också bör gå med i EMU eller tycker du inte det?

Svarsfördelningen blev: "Bör gå med" 50 %, "Bör inte gå med" 42 %, "Vet ej" 8 %.

- a) Jämför frågeformuleringarna med varandra, vad skiljer dem åt? Varför blev det så olika resultat? Kan formuleringen ha spelat roll för skillnaden i andelen "Vet ej" svar? Skulle du rekommendera någon av de två versionerna som mer tillförlitlig? Förklara! (10p)
- b) Ange två olika datainsamlingsmetoder som skulle kunna användas för den här studien om den gjordes idag med dagens teknik och möjligheter. Ange en fördel och en nackdel med respektive metod och jämför. Vilken metod skulle du välja? Motivera! (10p)