

Skriftlig tentamen i Undersökningsmetodik inom fortsättningskursen i statistik, måndagen den 20 december 2004

Skrivtid: 4 timmar

Hjälpmedel: miniräknare, bifogat formelblad.

Ange ditt fullständiga personnummer (dock ej ditt namn) på varje inlämnat pappersark.

Kom ihåg att en motiverad lösning/lösning med angivna formella uttryck kan ge poäng även om beräkningarna inte är helt korrekta.

Lös vänligen inte mer än en uppgift per inlämnat pappersark.

Skrivningsgenomgång: måndagen den 10 januari 2005, kl. 10 i sal B705.

Uppgift 1

I. Det förekommer benämningar “kalla” resp. “varma” datainsamlingsmetoder. Vad syftar benämningarna på; vilka metoder tillhör “kalla” resp. “varma” typen? (2p)

II. Vilka olika motiv för deltagande i en stickprovsundersökning kan det finnas hos en respondent? (3p)

III. Vad menas med *primacy effect* och *recency effect*? Hur uppstår dessa och hur kan man minska deras inflytande? (5p)

Uppgift 2

I. Vad är ett sannolikhetsurval? (2p)

II. Vad menas med en inklusionssannolikhet? (2p)

III. Vad menas med en estimator? (2p)

IV. På formelbladet ges det ett uttryck för variansskattning av en andelskattning vid OSU/uå. Visa, genom härledning, att detta uttryck är ett specialfall av uttrycket för variansskattning av en medelvärdesskattning vid OSU/uå, nämligen det fall som gäller då Y antar endast värden 0 eller 1. Starta från uttrycket för variansskattning av medelvärdet (också på formelbladet). (4p)

Uppgift 3

I. π -estimatorn

a) Vad är, enligt den s k designbaserade ansatsen, det enda slumpmässiga förknippat med ett urval (ideala situationen förutsätts: inga täcknings-, mät-, osv fel förekommer)?

b) Säg att vi använder en urvalsdesign som till element i tilldelar urvalssannolikheten π_i . Om vi med I_i betecknar den stokastiska variabeln i a ovan för elementet i , vad har I_i för fördelning?

c) Vad är $E[I_i]$?

- d) Skriv upp π -estimatoren i dess allmänna form.
 e) Visa, genom att starta från uttrycket i d, att π -estimatoren är väntevärdesriktig. (5p)

II.

a) Vi avser att göra en stickprovsundersökning, med OSU/uå som urvalsmetod, samt skatta med s.k. π -estimatoren (enligt kursboken, den vanliga skattningsmetoden vid OUS/uå). Förutsatt mycket stor population ($N \geq 10^6$), så att:

- ändlighetskorrektionen kan anses vara 1,
- kvoten $N/(N-1)$ kan anses vara 1,
- övriga uttryck där N står i nämnaren kan anses vara 0,
- σ (den sanna standardavvikelsen av y) och s (den skattade standardavvikelsen av y)

kan anses vara lika,

vilken minsta stickprovsstorlek behöver vi för att få en 95%-ig felmarginal ('bound of error', $B \approx 2 \times \sqrt{\hat{V}(\hat{\mu})}$) inte bli större än 0.03. Vi tror att vi vet att standardavvikelsen av y är $\sigma = 1$. (2p)

b) Säg att vi, i en stickprovsundersökning, har bestämt oss för att ha en stickprovsstorlek på $n = 4500$ individer. Standardavvikelsen för kvinnor och för män förefaller dock skilja sig för vår undersökningsvariabel, vilket kan leda till försämrad precision om vi inte gör ett stratifierat urval. Hur ska vi fördela stickprovsstoren på strata, om vi tror att vi vet att standardavvikelsen av undersökningsvariabeln är för kvinnor $\sigma_K = 1$ och för män $\sigma_M = 2$. Det är lika många kvinnor som män i populationen, som i övrigt inte har någon tredje sorts individer. (3p)

Uppgift 4

a) Vi har gjort en stickprovsundersökning i situationen liknande den i uppgift 3/II/a (men, vi tar inte över några av där nämnda förutsättningar). Populationsstorleken var faktiskt 10^6 individer, och stickprovsstorleken 4500. Inget bortfall förekom, inget mätfel heller. Vi erhöll följande data ($\bar{y}$ och s^2 betecknar medelvärde resp. varians av undersökningsvariabeln i stickprovet): $\bar{y} = 3$, $s^2 = 3.76$. Skatta medelvärde för undersökningsvariabeln i populationen, samt ge dess 95%-igt konfidensintervall. (5p)

b) Situationen som i a ovan, dock efter att ha tittat på respondenternas kön inser vi att det råkade vara så att mycket fler män än kvinnor blev valda att delta i undersökningen. I populationen var det lika många kvinnor som män, men i stickprovet fick vi 1500 kvinnor och 3000 män. (Mycket osannolikt i verkligheten om alla svarade; däremot kan respondentvägran ha lett till detta utfall, och undersökaren blir kanske dragen till att tillämpa samma resonemang där; men vi betraktar nu det förstnämnda fallet även om det är osannolikt.) Vi erhöll följande data (där t ex $\bar{y}_K$ och s_K^2 betecknar medelvärde resp. varians av undersökningsvariabeln bland kvinnor i stickprovet): $\bar{y}_K = 3.5$, $\bar{y}_M = 2.75$, $s_K^2 = 1.21$, $s_M^2 = 4.84$. Skatta medelvärde för undersökningsvariabeln i populationen, med hänsyn tagen till antalet kvinnor och män i populationen, samt ge dess 95%-igt konfidensintervall. Vi får nu bortse från ändlighetskorrektionerna. (5p)

Uppgift 5

I ett företag är 2112 personer anställda. Dessa jobbar i flexibla arbetslag vars storlek anpassas efter jobben. Varje anställd tillhör ett och endast ett arbetslag; det finns just för tillfället 88 arbetslag. På grund av flexibiliteten vet man inte på förhand hur stora arbetslagen är.

För att skatta totalt antal arbetade timmar i veckan i företaget väljer man då ut med hjälp av systematiskt urval 8 arbetslag ur en slumpmässigt ordnad populationsram innehållande de 88 lagen. För varje lag antecknar man hur många anställda arbetar i lagen samt hur många timmar sammanlagt dessa anställda har arbetat under referensveckan. Därpå kan man konstatera följande:

Arbetslag	Antal anställda	Summa arbetade timmar
1	31	1395
2	14	420
3	20	700
4	16	880
5	42	1680
6	29	870
7	27	1080
8	13	455

Skatta totalen för arbetade timmar i företaget under referensveckan och ge tillhörande 95%-igt konfidensintervall. (10p)