

## Tentamen i Surveymetodik 732G26

Måns Magnusson

22 mars 2013, kl. 8.00-12.00

Surveymetodik med uppsats, 15 hp  
Kandidatprogrammet i Statistik och dataanalys  
VT2013

---

### Instruktioner

- **Hjälpmedel:**

- Lohr, S: *Sampling- Design and analysis* (anteckningar får finnas).
- Miniräknare.

- **Jourhavande lärare:**

Måns Magnusson

- **Poänggränser:**

Skrivningen ger maximalt 20 poäng. För betyget godkänt krävs normalt 12 poäng och för betyget väl godkänt krävs 16 p.

- **Övrig information:**

Samtliga siffror i examen är fiktiva.

Lycka till!

---



1. Arbetsmiljöverket vill undersöka risken för arbetsskador på arbetsplatser med hög risk för arbetsskador. De har genomfört en undersökning och dragit ett slumpmässigt urval om 600 personer från en population på 91122 individer som yrkesarbetar (för enkelhetsskull har samtliga personer svarat).

I undersökningen ställs frågan om hur många gånger respondenten har råkat ut för en arbetsskada under det senaste året som gjort att de måste stanna hemma från arbetet minst en dag. Bland de svarande fås medelvärdet 0.125 och standardavvikelsen 0.454.

Arbetsmiljöverket är även intresserade av hur stor andel av respondenterna som varit med om en arbetsskada (överhuvudtaget) det senaste året. Undersökningen visar att denna andel är 0.093.

- (a) Beräkna en totalskattning,  $\hat{t}$ , med tillhörande konfidensintervall (95 %) för det totala antalet arbetsskador i hela populationen. **2p.**
  - (b) Beräkna en skattning av andelen som varit med om en arbetsskada,  $\hat{p}$ , med tillhörande konfidensintervall (95 %). **1p.**
  - (c) Baserat på denna undersökning vill Arbetsmiljöverket genomföra en ny studie och är intresserade av att få ett konfidensintervall (95 %) för  $\bar{y}$  som är högst  $\bar{y} \pm 0.02$ . Beräkna (baserat på de data du fått i denna undersökning) vilken urvalsstorlek ( $n$ ) som behövs (minst) för att få detta konfidensintervall i den nya studien. **2p.**
2. Istället för att genomföra den nya studien beslutar sig arbetsmiljöverket för att genomföra en stratifierad studie eftersom det finns stora skillnader i arbetsskador mellan olika yrkesgrupper (det är dock ingen skillnad i kostnad mellan de olika strata). De väljer därför att stratifiera populationen av personer som arbetar på arbetsplatser med hög risk (ex. undersköterskor, maskinoperatörer, byggnadsarbetare och lärare) respektive lägre risk för arbetsskador.
- De personer som arbetar i yrken med högre risk för arbetsskador utgör 36473 individer och gruppen med lägre risk utgör resten av populationen. Från den tidigare undersökningen har medelvärdet uppskattats till 0.29 med standardavvikelsen 0.669 för de med högre risk för arbetsskador och 0.031 med standardavvikelsen 0.215 för gruppen med lägre risk.

- (a) Eftersom båda grupperna är ungefär lika stora föreslås att undersökningen ska genomföras genom med 300 respondenter i varje strata, d.v.s.  $n = 600$ . Baserat på denna urvalsstorlek och resultaten från den första undersökningen, beräkna punktskattning av totalen  $\hat{t}$  med tillhörande konfidensintervall (95 %). Vad får du för designeffekt av att använda detta stratifierade urval? **2p.**
- (b) Innan de beslutar sig för att genomföra denna undersökning frågar de dig hur du tycker att de ska fördela urvalet mellan de två strata för att få en så bra skattning som möjligt. Vilken metod väljer du och varför? Med den metod du valt, beräkna hur stort urval du tar i respektive strata. Beräkna punktskattning och konfidensintervall (95 %) för  $\hat{t}$ . Beräkna även designeffekten för denna urvalsdesign. **3p.**

3. I en annan studie om arbetsmiljö som genomförs i Linköping har de som genomfört undersökningen valt att dra ett urval av arbetsplatser. Urvalet har dragits från en ram av samtliga arbetsplatser med fler än 20 anställda. Från respektive arbetsplats har sedan maximalt 30 anställda valts ut slumpmässigt (med OSU) för att delta i studien. Dessa har sedan intervjuats avseende sin arbetsmiljö.

Vi känner sedan tidigare till att det i Linköping finns totalt 327 arbetsplatser av denna storlek och att totalt 27012 anställda arbetar på dessa arbetsplatser. I tabellen nedan framgår resultatet från dessa intervjuer när det gäller antalet arbetsskador.

	Anställda	Intervjuade	Medelvärde	Standardavvikelse
1	215	30	0.23	0.50
2	25	25	0.08	0.28
3	33	30	0.20	0.41
4	27	27	0.30	0.87
5	21	21	0.14	0.36
6	113	30	0.20	0.61

Table 1: Antal arbetsskador på de undersökta arbetsplatserna

- (a) Utgå (initialt) från att de personer som intervjuats gällande arbetsskador är samtliga personer som arbetar vid respektive företag (d.v.s. att antalet intervjuade = antalet anställda). Beräkna en skattning av det genomsnittliga antalet arbetsskador i populationen med tillhörande konfidensintervall (95 %). **2 p.**
- (b) Ta nu hänsyn till att endast ett mindre urval (i vissa fall) av de anställda vid varje arbetsplats har intervjuats gällande arbetsskador och ta hänsyn till den extra osäkerhet detta medför och beräkna  $\hat{y}$ , med tillhörande konfidensintervall (95 %). **3p.**
4. I Norröpings kommun har en opinionsundersökning genomförts bland den röstberättigande befolkningen. Undersökningen har genomförts med ett vanligt OSU. En av de frågor som undersöktes i denna undersökning var antalet personer som skulle rösta på socialdemokraterna om det vore val idag. Tyvärr var din ram bristfällig och den enda information du har i ramen är könet på respondenten. I tabellen nedan framgår resultatet från undersökningen.

	Bortfall	Svarande	Population	Antal (s)-väljare
Män	160	232	50851	87
Kvinnor	158	255	51772	93

Table 2: Resultat: Opinionsundersökning i Norrköping

- (a) Ignorera för tillfället bortfallet och skatta andelen i Norrköping som skulle rösta på socialdemokraterna om det vore val idag, tillsammans med tillhörande konfidensintervall (95 %). **1p.**

- (b) Du inser dock att bortfallet är allt för stort för att inte analyseras vidare. Med hjälp av R genomför du en mindre bortfallsanalys med ett  $\chi^2$ -test som ger följande resultat:

```
Pearson's Chi-squared test with Yates' continuity correction

data:  Resultat[, 1:2]
X-squared = 0.45, df = 1, p-value = 0.5026
```

Vad är din slutsats? Vilket antagande skulle du göra om bortfallet i detta fall och varför? **1p.**

- (c) Oavsett vilket antagande du gjort om bortfallet vill du kalibrera dina resultat för bortfallet som uppkommit. Beräkna desgnvikten  $w_i$  och kalibreringsvikten  $g_i$  och använd dessa för att beräkna en totalskattning  $\hat{t}_{yr}$  av antalet personer som skulle rösta på socialdemokraterna om det vore val idag. Jämför med resultatet i uppgift a) (du måste räkna om andelen till total). Vad är din slutsats? **3p.**

## Appendix

### NORMAL CUMULATIVE DISTRIBUTION FUNCTION

$x$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986