

## Tentamen i Surveymetodik 732G26

Måns Magnusson

21 mars 2014, kl. 8.00-12.00

Surveymetodik med uppsats, 15 hp  
Kandidatprogrammet i Statistik och dataanalys  
VT2014

### Instruktioner

- **Hjälpmedel:**

- Lohr, S: *Sampling- Design and analysis* (anteckningar får **inte** finnas, men sidflärpar är tillåtet).
- Miniräknare.

- **Jourhavande lärare:**

Måns Magnusson

- **Poänggränser:**

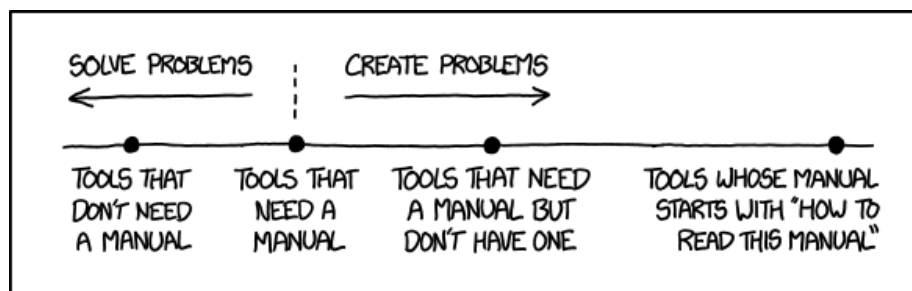
Skrivningen ger maximalt 20 poäng. För betyget godkänt krävs normalt 12 poäng och för betyget väl godkänt krävs 16 p.

- **Övrig information:**

Samtliga siffror i examen är fiktiva.

Är det så att någon siffra skulle saknas för att kunna lösa uppgiften, skriv då tydligt ut att du saknar denna information, anta ett gofityckligt värde för denna storhet och lös uppgiften med detta antagande.

Lycka till!



**Uppgift 1**

Pensionsmyndigheten vill undersöka hur många personer i åldersgruppen 50-61 år som skulle kunna tänka sig att arbeta vidare efter pensionen. I Sverige bor 1288778 personer i denna åldersgrupp. Pensionsmyndigheten vill undersöka hur många personer som kan (och vars kroppar klarar) att arbeta vidare efter 61 år ålder på 100% (då det som tidigast är möjligt att gå i pension). De är dessutom intresserade av hur många år efter 61 års ålder personer skulle kunna tänka sig att fortsätta arbeta heltid.

Myndigheten valde en urvalsstorlek på 1250 och av dessa deltog 772 personer i undersökningen som genomfördes med obundet slumpmässigt urval. Resultaten de fick var att 0.79 skulle kunna tänka sig arbeta efter 61 års ålder och i genomsnitt skulle de då kunna arbeta i ytterligare 3.14 år (med en standardavvikelse på 2.6).

I denna undersökning antar vi Missing completely at random (MCAR).

- a) Beräkna en totalskattning av hur många i populationen som kan tänka sig arbeta efter 61 års ålder med tillhörande konfidensintervall 90 %. Ignorera bortfallet. **2p**.
- b) Om vi är intresserade av att producera ett konfidensintervall för  $\bar{y}$  (genomsnittligt antal år respondenten kan tänka sig arbeta efter 61 års ålder), är urvalsstorleken tillräckligt stor för att konfidensintervallet ska vara approximativt normalfördelade? Varför eller varför inte? **2p**.
- För denna uppgift kan följande storhet vara av intresse:

$$\sum_{i \in \mathcal{S}} \frac{(y_i - \bar{y})^3}{n} = 9.78$$

där  $y$  är antalet år respondenten kan tänka sig arbeta heltid efter 61 års ålder.

- c) Beräkna inklusionssannolikheten  $\pi_i$  för respondenterna i denna undersökning. **1p**.

**Uppgift 2**

- a) Visa att

$$\hat{V}(\hat{p}) = \left(1 - \frac{n}{N}\right) \frac{\hat{p}(1 - \hat{p})}{n - 1}$$

är ett specialfall av

$$\hat{V}(\bar{y}) = \left(1 - \frac{n}{N}\right) \frac{s^2}{n}$$

där  $y$  bara kan ta värdena  $\{0, 1\}$  och  $\hat{p} = \bar{y} = \sum_i \frac{y_i}{n}$ . **2p**

- b) Inför valet 2014 genomförs 100 opinionsundersökningar med runt cirka 2 000 deltagare i varje undersökning. Antag att det bor 6 000 000 röstberättigade i Sverige.
- i) Vad är sannolikheten att du kommer delta i en undersökning? **1p**
- ii) Vad är sannolikheten att du **inte** kommer delta i någon av undersökningarna? (Antag att de är oberoende av varandra) **1p**

### Uppgift 3

Svenska Brukshundsföreningen vill genomföra en medlemsundersökning bland sina 60 000 medlemmar. Tanken är att dra ett urval på 500 hushåll från föreningens register och sedan genomföra en telefonintervju med de som inkluderats i urvalet.

- a) Baserat på förslaget till undersökning ovan. Förklara följande begrepp genom att exemplifiera hur begreppen skulle kunna användas i studien ovan. Varje begrepp ger **0.5 p.**
- i) Övertäckning
  - ii) Obundet slumpmässigt urval med återläggning
  - iii) Missing completely at random (MCAR)
  - iv) Statistikens tillgänglighet
  - v) Stratifierat urval
  - vi) pps/ $\pi$ ps-urval
  - vii) Imputering
- b) Nämn tre sätt det skulle vara möjligt att förbygga bortfall i denna undersökning. **1.5p**

## Uppgift 4

Bil Sweden är intresserade av att undersöka hur mycket kostnader olika typer av motorcykelägare har för sin motorcykel under det senaste året. De är intresserade av de sammanlagda kostnaderna ( $y$ ) och skickar därför ut en postenkät till ett stratifierat slumpmässigt urval av storlek 1000 till de ägare av de 283631 motorcyklar som var i trafik under 2013. Stratifiering gjordes baserat på motorcykelns årsmodell. Kostnaderna i kr per respondent ( $c$ ) är lika för alla (postenkät). Totalt deltog  $n_r = 605$  motorcykelägare i undersökningen.

Det fanns också intresserade av hur många som haft mer än kr i 4000 utgifter för sin motorcykel ( $p$ ).

Undersökningen gav följande resultat:

	$N_h$	$n_h$	$n_{rh}$	$\bar{y}_h$	$s_h$	$p_h$	$c_h$
2010 -	41361	146	93	3982.60	166.86	0.43	12
2000 - 2010	107289	378	235	3617.71	649.59	0.28	12
< 2000	134981	476	277	4299.55	1414.89	0.56	12
Samtliga	283631	1000	605	3985.98	1086.45	0.43	

- a) Baserat på resultatet ovan beräkna en punktskattning  $\bar{y}_U$  med tillhörande konfidensintervall (99%). **2.5p**
- b) Beräkna designvikterna för respektive strata. **1.5p**

De vill nu genomföra en ny undersökning och konsulterar dig. I budgeten har de lagt in att de kan skicka ut totalt 1000 enkäter. De står och väger mellan lika, proportionell, optimal och Neymanallokering.

- c) De är intresserade av att jämföra skillnader mellan olika strata (sett som redovisningsgrupper). Vilken allokering föreslår du? Beräkna urvalsstorleken i respektive strata med denna allokering. **1p.**
- d) Om de istället skulle vilja att urvalet ska ha samma fördelning mellan strata som fördelningen ser ut i populationen. Vilken allokering skulle du då rekommendera? **1p.**

## 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