



Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2010-10-09
Sal (1) Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och <u>ringa in</u> vilken sal som avses	TER1
Tid	8-12
Kurskod	732G19
Provkod	TENT
Kursnamn/benämning Provnamn/benämning	Utredningskunskap I Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	5
Jour/Kursansvarig Ange vem som besöker salen	Kalle Wahlin
Telefon under skrivtiden	0709-719096
Besöker salen ca kl.	9.15
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Carita Lilja, 1463, carita.lilja@liu.se
Tillåtna hjälpmedel	Valfri räknedosa, till tentamen vidhäftad formelsamling
Övrigt	
Vilken typ av papper ska användas, rutigt eller linjerat	Rutigt
Antal exemplar i påsen	10

Tentamen

Linköpings Universitet, Institutionen för datavetenskap, Statistik

Kurskod och namn: 732G19 Utredningskunskap I, samt 732G02, 732G90 och 732G04

Datum och tid: 2010-10-09, 8-12

Jourhavande lärare: Kalle Wahlin

Tillåtna hjälpmedel: Valfri räknedosa, till tentamen vidhäftad formelsamling.

Betygsgränser: Tentamen omfattar totalt 20p. Godkänt från 12p, väl godkänt från 16p.
Siffrorna i uppgifterna är påhittade.

Redovisa och motivera tydligt alla dina lösningar!

Uppgift 1 (7p)

Svenska Dagbladet har genomfört en läsarundersökning i Stor-Stockholm. Urvalet bestod av 1000 personer dragna som ett OSU bland de totalt 1 200 000 invånarna äldre än 16 år i området.

- a) Svaren på den första frågan, "Läser du regelbundet Svenska Dagbladet?", fördelade sig enligt

Svar	Antal
Ja	540
Nej	460
Totalt	1000

Bestäm ett 95% konfidensintervall för andelen läsare i Stor-Stockholm bland samtliga personer äldre än 16 år. (2p)

- b) Bestäm ett 95% konfidensintervall för antalet läsare i Stor-Stockholm bland samtliga personer äldre än 16 år. (1p)
- c) Samma fråga ställdes i en liknande undersökning, även här utförd som ett OSU om 1000 personer, bland de 6 000 000 personer äldre än 16 år som bor i Sverige utanför Stor-Stockholm. Resultatet var

Svar	Antal
Ja	150
Nej	850
Totalt	1000

Betrakta Stor-Stockholm och övriga Sverige som två stratum och bestäm ett 95% konfidensintervall för andelen personer i Sverige som läser Svenska Dagbladet bland de som är äldre än 16 år. (3p)

- d) Hur stort stickprov skulle ha dragits ur respektive stratum om istället proportionell allokering hade använts? (1p)

Uppgift 2 (2p)

En fackförening önskar undersöka hur många av medlemmarna som är nöjda med sina arbetstider. I en förundersökning tillfrågades ett OSU av medlemmarna kring om de var nöjda, och 70% svarade ja. I den kommande undersökningen önskar man bilda ett 95% konfidensintervall för andelen nöjda, och man vill ha en felmarginal om högst 10 procentenheter. Hur stort urval bör man dra?

Uppgift 3 (7p)

En person har köpt 2500 tulpanlökar förpackade i 100 plastpåsar. Han misstänkte att partiet innehåller ett stort antal ruttna lökar, och för att skatta andelen ruttna lökar drog han ett OSU om 5 påsar. Följande resultat erhöles.

Påse	Antal lökar	Antal ruttna lökar
1	25	15
2	20	10
3	30	24
4	25	20
5	35	21

- Bestäm en kvotskattning av andelen ruttna lökar i partiet om 2500 lökar. (2p)
- Bilda baserat på kvotskattning ett 95% konfidensintervall för andelen ruttna lökar i partiet. (3p)
- Bestäm en väntevärdesriktig punktskattning av andelen ruttna lökar i partiet. (2p)

Uppgift 4 (2p)

Föreliggande är en population om följande element.

2 3 5 7 11 13

Man önskar dra ett systematiskt urval om 2 element ur populationen. Ange samtliga möjliga systematiska urval ur populationen, om urvalsramen ej görs cyklisk.

Uppgift 5 (2p)

- Vilken typ av information bör ingå i ett introduktionsbrev vid en enkätundersökning? (1p)
- Redogör kortfattat för Hansen-Hurvitz bortfallsplan. (1p)

Formelblad i Surveymetodik ht 2007 (732G04, 732G02-C, 732G90-C)

Normalfördelningskvantiler:

$1 - \alpha$	$z_{\alpha/2}$	z_{α}
0.90	1.645	1.28
0.95	1.96	1.645
0.99	2.576	2.326

1 Obundet slumpmässigt urval

1.1 Urvalsvarians(er)

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n} \right) = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n \cdot \bar{x}^2 \right)$$

$$s^2 = \frac{n}{n-1} \cdot p \cdot (1-p).$$

$$\sum x^2 = (n-1) \cdot s^2 + n \cdot (\bar{x})^2$$

1.2 Konfidensintervall

$$\mu : \quad \bar{x} \pm z_{\alpha/2} \cdot \sqrt{\frac{s^2}{n} \cdot \left(1 - \frac{n}{N}\right)}$$

$$\tau : \quad N \cdot \bar{x} \pm z_{\alpha/2} \cdot \sqrt{N^2 \cdot \frac{s^2}{n} \cdot \left(1 - \frac{n}{N}\right)}$$

$$P : \quad p \pm z_{\alpha/2} \cdot \sqrt{\frac{p(1-p)}{n-1} \cdot \left(1 - \frac{n}{N}\right)}$$

Ändlighetskorrektionen $(1 - \frac{n}{N})$ kan utelämnas om $\frac{n}{N} \leq 0.05$.

1.3 Urvalsdimensionering

$$n_0 \geq \frac{4 \cdot (z_{\alpha/2})^2 \cdot \sigma^2}{B^2}$$

$$n_0 \geq \frac{4 \cdot (z_{\alpha/2})^2 \cdot P(1-P)}{B^2}$$

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

Samband mellan konfidensintervallbredder: $B_{\tau} = N \cdot B_{\mu}$

1.4 Redovisningsgrupper

Punktskattningar

$$\bar{x}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} x_{1i}$$

$$s_1^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n_1} (x_{1i} - \bar{x}_1)^2 = \frac{1}{n_1 - 1} \left(\sum_{i=1}^{n_1} x_{1i}^2 - n_1 \cdot (\bar{x}_1)^2 \right)$$

Konfidensintervall

$$\mu_1 : \bar{x}_1 \pm z_{\alpha/2} \sqrt{\frac{s_1^2}{n_1} \left(1 - \frac{n_1}{N_1}\right)}$$

$$P_1 : p_1 \pm z_{\alpha/2} \sqrt{\frac{p_1(1-p_1)}{n_1 - 1} \left(1 - \frac{n_1}{N_1}\right)}$$

$$\tau_1 : N \cdot \bar{x}' \pm z_{\alpha/2} \cdot \sqrt{N^2 \cdot \frac{s'^2}{n} \left(1 - \frac{n}{N}\right)}$$

$$A_1 : N \cdot p' \pm z_{\alpha/2} \cdot \sqrt{N^2 \cdot \frac{p'(1-p')}{n-1} \left(1 - \frac{n}{N}\right)}$$

där

$$x'_i = \begin{cases} x_i & \text{om elementet tillhör redovisningsgruppen} \\ 0 & \text{annars} \end{cases}$$

och

$$\bar{x}' = \frac{1}{n} \sum_{i=1}^n x'_i$$

$$s'^2 = \frac{1}{n-1} \sum_{i=1}^n (x'_i - \bar{x}')^2 = \frac{1}{n-1} \left(\sum_{i=1}^n (x'_i)^2 - n \cdot (\bar{x}')^2 \right)$$

$$p' = \frac{\text{Antal element från redovisningsgruppen i urvalet med } x = 1}{n}$$

1.5 Ramproblem

Konfidensintervall vid övertäckning

$$\mu_0 : \bar{x}_0 \pm z_{\alpha/2} \cdot \sqrt{\frac{s_0^2}{n_0} \cdot \left(1 - \frac{n}{N}\right)}$$

$$P_0 : p_0 \pm z_{\alpha/2} \cdot \sqrt{\frac{p_0(1-p_0)}{n_0 - 1} \cdot \left(1 - \frac{n}{N}\right)}$$

För totaler:

$$\widehat{N}_0 \cdot \bar{x}_0 \pm z_{\alpha/2} \cdot \sqrt{N^2 \cdot \frac{1}{n-1} \cdot \left(\sum_{i=1}^{n_0} x_i^2 - \frac{1}{n} \left(\sum_{i=1}^{n_0} x_i \right)^2 \right) \cdot \left(1 - \frac{n}{N}\right)}$$

$$\widehat{N}_0 \cdot p_0 \pm z_{\alpha/2} \cdot \sqrt{N^2 \cdot \frac{n_0 \cdot p_0}{n-1} \cdot \left(1 - \frac{n_0 \cdot p_0}{n}\right) \cdot \left(1 - \frac{n}{N}\right)}$$

Replikatproblem

$$\hat{\tau} = \frac{1}{n} \sum_{i=1}^n \frac{x_i}{M_i/N}$$

$$\hat{\mu} = \frac{\sum_{i=1}^n \frac{x_i}{M_i}}{\sum_{i=1}^n \frac{1}{M_i}}$$

2 Stratifierat urval

2.1 Punktskattningar:

$$\hat{\mu} = \bar{x}_{st} = W_1 \cdot \bar{x}_1 + W_2 \cdot \bar{x}_2 + \dots + W_L \cdot \bar{x}_L = \sum_{i=1}^L W_i \cdot \bar{x}_i$$

$$\hat{\tau} = N \cdot \bar{x}_{st}$$

$$\hat{P} = p_{st} = W_1 \cdot p_1 + W_2 \cdot p_2 + \dots + W_L \cdot p_L = \sum_{i=1}^L W_i \cdot p_i$$

2.2 Konfidensintervall

$$\mu : \bar{x}_{st} \pm z_{\alpha/2} \cdot \sqrt{\sum_{i=1}^L W_i^2 \cdot \frac{s_i^2}{n_i} \cdot \left(1 - \frac{n_i}{N_i}\right)}$$

$$\tau : N \cdot \bar{x}_{st} \pm z_{\alpha/2} \cdot \sqrt{N^2 \cdot \sum_{i=1}^L W_i^2 \cdot \frac{s_i^2}{n_i} \cdot \left(1 - \frac{n_i}{N_i}\right)}$$

$$P : p_{st} \pm z_{\alpha/2} \cdot \sqrt{\sum_{i=1}^L W_i^2 \cdot \frac{p_i \cdot (1-p_i)}{n_i - 1} \cdot \left(1 - \frac{n_i}{N_i}\right)}$$

$$A : N \cdot p_{st} \pm z_{\alpha/2} \cdot \sqrt{N^2 \cdot \sum_{i=1}^L W_i^2 \cdot \frac{p_i \cdot (1-p_i)}{n_i - 1} \cdot \left(1 - \frac{n_i}{N_i}\right)}$$

2.3 Proportionell allokering

$$n_i = n \cdot W_i = n \cdot \frac{N_i}{N}$$

$$n_0 \geq \frac{4 \cdot (z_{\alpha/2})^2 \cdot \sum_{i=1}^L W_i \cdot \sigma_i^2}{B^2}$$

$$n_0 \geq \frac{4 \cdot (z_{\alpha/2})^2 \cdot \sum_{i=1}^L W_i \cdot P_i \cdot (1-P_i)}{B^2}$$

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}}$$

2.4 Fullständig optimal allokering

$$n_i = n \cdot \frac{W_i \cdot \sigma_i / \sqrt{c_i}}{\sum_{j=1}^L W_j \cdot \sigma_j / \sqrt{c_j}} = n \cdot \frac{N_i \cdot \sigma_i / \sqrt{c_i}}{\sum_{j=1}^L N_j \cdot \sigma_j / \sqrt{c_j}}$$

Fix kostnad:

$$n \geq \frac{(C_{max} - C_0) \cdot \sum_{i=1}^L (N_i \cdot \sigma_i / \sqrt{c_i})}{\sum_{i=1}^L (N_i \cdot \sigma_i \cdot \sqrt{c_i})}$$

Fix konfidensintervallbredd:

$$n \geq \frac{\left(\sum_{i=1}^L (W_i \cdot \sigma_i \cdot \sqrt{c_i}) \right) \cdot \sum_{i=1}^L (W_i \cdot \sigma_i / \sqrt{c_i})}{B^2 / (4 \cdot (z_{\alpha/2})^2) + (1/N) \cdot \sum_{i=1}^L (W_i \cdot \sigma_i^2)}$$

2.5 Neyman-allokering

$$n_i = n \cdot \frac{W_i \cdot \sigma_i}{\sum_{j=1}^L W_j \cdot \sigma_j} = n \cdot \frac{N_i \cdot \sigma_i}{\sum_{j=1}^L N_j \cdot \sigma_j}$$
$$n \geq \frac{\left(\sum_{i=1}^L (W_i \cdot \sigma_i) \right)^2}{B^2 / (4 \cdot (z_{\alpha/2})^2) + (1/N) \cdot \sum_{i=1}^L (W_i \cdot \sigma_i^2)}$$

Vid 0/1-data gäller $\sigma_i = \sqrt{P_i \cdot (1 - P_i)}$.

3 Enstegs klusterurval

3.1 OSU av kluster

Väntevärdesriktig skattning

$$\hat{\mu}_u = \frac{N}{M_0} \cdot \bar{T}$$

$$\text{K.I.: } \hat{\mu}_u \pm z_{\alpha/2} \cdot \sqrt{\left(\frac{N}{M_0}\right)^2 \cdot \frac{s_T^2}{n} \cdot \left(1 - \frac{n}{N}\right)}$$

$$\hat{\tau} = M_0 \cdot \hat{\mu}_u$$

$$\text{K.I.: } M_0 \cdot \hat{\mu}_u \pm M_0 \cdot z_{\alpha/2} \cdot \sqrt{\left(\frac{N}{M_0}\right)^2 \cdot \frac{s_T^2}{n} \cdot \left(1 - \frac{n}{N}\right)}$$

Kvotskattning, $\hat{\mu}_R = \bar{x}_{cl}$

$$\bar{x}_{cl} = \frac{\sum_{i=1}^n T_i}{\sum_{i=1}^n M_i}$$

$$\widehat{\text{Var}}(\bar{x}_{cl}) = \left(1 - \frac{n}{N}\right) \cdot \frac{1}{n \cdot (n-1)} \cdot \frac{1}{(\bar{M})^2} \cdot \left(\sum_{i=1}^n T_i^2 + (\bar{x}_{cl})^2 \cdot \sum_{i=1}^n M_i^2 - 2 \cdot \bar{x}_{cl} \cdot \sum_{i=1}^n T_i \cdot M_i\right)$$

$$\text{K.I. : } \bar{x}_{cl} \pm z_{\alpha/2} \cdot \sqrt{\left(1 - \frac{n}{N}\right) \cdot \frac{1}{n \cdot (n-1)} \cdot \frac{1}{(\bar{M})^2} \cdot \left(\sum_{i=1}^n T_i^2 + (\bar{x}_{cl})^2 \cdot \sum_{i=1}^n M_i^2 - 2 \cdot \bar{x}_{cl} \cdot \sum_{i=1}^n T_i \cdot M_i\right)}$$

3.2 PPS-urval av kluster

$$\hat{\mu}_{pps} = \frac{1}{n} \sum_{i=1}^n \bar{x}_i$$

$$\text{K.I. } \bar{x}_{pps} \pm z_{\alpha/2} \cdot \sqrt{\frac{1}{n \cdot (n-1)} \cdot \left(\sum_{i=1}^n (\bar{x}_i)^2 - n \cdot (\bar{x}_{pps})^2\right)}$$

$$\hat{P}_{pps} = p_{pps} \frac{1}{n} \sum_{i=1}^n p_i$$

$$\text{K.I. } p_{pps} \pm z_{\alpha/2} \cdot \sqrt{\frac{1}{n \cdot (n-1)} \cdot \left(\sum_{i=1}^n (p_i)^2 - n \cdot (p_{pps})^2\right)}$$

3.3 Urvalsdimensionering

pps-urval

$$n \geq \frac{4 \cdot (z_{\alpha/2})^2 \cdot \sigma_{x_i}^2}{B^2}$$

4 Tvåstegs klusterurval

4.1 OSU av kluster

Väntevärdesriktiga skattningar

$$\hat{\mu}_{u2} = \left(\frac{N}{M_0}\right) \cdot \widehat{T} = \left(\frac{N}{M_0}\right) \cdot \frac{1}{n} \cdot \sum_{i=1}^n M_i \cdot \bar{x}_i$$

$$\hat{P}_{u2} = \left(\frac{N}{M_0}\right) \cdot \frac{1}{n} \cdot \sum_{i=1}^n M_i \cdot p_i$$

Kvotskattningar

$$\hat{\mu}_{R2} = \frac{\sum_{i=1}^n \hat{T}_i}{\sum_{i=1}^n M_i} = \frac{\sum_{i=1}^n M_i \cdot \bar{x}_i}{\sum_{i=1}^n M_i}$$

$$\hat{P}_{R2} = \frac{\sum_{i=1}^n M_i \cdot p_i}{\sum_{i=1}^n M_i}$$

4.2 PPS-urval

$$\hat{\mu}_{pps2} = \frac{1}{n} \cdot \sum_{i=1}^n \bar{x}_i$$

$$Var(\widehat{\mu}_{pps2}) = \frac{s_{\bar{x}_i}^2}{n} = \frac{1}{n \cdot (n-1)} \cdot \left(\sum_{i=1}^n (\bar{x}_i)^2 - n \cdot (\bar{x}_{pps})^2 \right)$$

5 Bortfallsstratumansatsen

Punktskattningar och konfidensintervall

$$\hat{\mu}_{st} = \frac{n'_S}{n} \cdot \bar{x}_S + \frac{n'_B}{n} \cdot \bar{x}_B$$

$$\hat{P}_{st} = \frac{n'_S}{n} \cdot p_S + \frac{n'_B}{n} \cdot p_B$$

$$\hat{\mu}_{st} \pm z_{\alpha/2} \cdot \sqrt{\frac{n'_S}{n} \cdot \frac{s_S^2}{n} + \left(\frac{n'_B}{n}\right)^2 \cdot \frac{s_B^2}{n_B} + \frac{1}{n} \cdot \left(\frac{n'_S}{n} \cdot (\bar{x}_S - \hat{\mu}_{st})^2 + \frac{n'_B}{n} \cdot (\bar{x}_B - \hat{\mu}_{st})^2\right)}$$

$$\hat{P}_{st} \pm z_{\alpha/2} \cdot \sqrt{\left(\frac{n'_S}{n}\right)^2 \cdot \frac{p_S \cdot (1-p_S)}{n'_S - 1} + \left(\frac{n'_B}{n}\right)^2 \cdot \frac{p_B \cdot (1-p_B)}{n'_B - 1} + \frac{1}{n} \cdot \left(\frac{n'_S}{n} \cdot (p_S - \hat{P}_{st})^2 + \frac{n'_B}{n} \cdot (p_B - \hat{P}_{st})^2\right)}$$

Slutligt bortfall

$$b = (\text{Bortfallsandel i fas 1}) \times (\text{Bortfallandel i fas 2})$$

$$= \frac{n'_B}{n} \cdot \frac{n''_B}{n_B}$$

där n''_B = antal bortfall i fas 2-urvalet.