Arkiv

Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2011-06-08
Sal (4) Om tentan går i flera salar ska du bifoga ett försättsblad till varje sal och <u>ringa in</u> vilken sal som avses	U3 U4 U6 U7
Tid	14-18
Kurskod	732G19
Provkod	TENT
Kursnamn/benämning **Provnamn/benämning**	Utredningskunskap I Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	6
Jour/Kursansvarig Ange vem som besöker salen	Karl Wahlin
Telefon under skrivtiden	0709-719096
Besöker salen ca kl.	ca kl 15
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	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	

60

Tentamen

Linköpings Universitet, Institutionen för datavetenskap, Statistik

---

Kurskod och namn:	732G19 Utredningskunskap I, samt 732G04 Surveymetodik
Datum och tid:	2011-06-08, 14-18
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 (4p)

I en långsträckt kommun ville man uppskatta hur många vuxna innevånare som var positivt inställda till att bygga en simhall. Simhallen skulle ligga i ena änden av kommunen, men en skattehöjning för att finansiera bygget skulle drabba alla kommuninnevånare. Populationen indelades i fyra grupper, efter hur långt ifrån den tänkta byggplatsen man bor. Sammanlagt 400 personer utvaldes genom proportionell allokering och därpå följande OSU och följande resultat erhölls.

Grupp	Antal vuxna personer i gruppen	Antal positiva i stickprovet
1	4000	50
2	1000	40
3	2000	50
4	3000	50

Beräkna ett 90% konfidensintervall för antalet vuxna i kommunen som är positivt inställda till byggandet av en simhall.

Uppgift 2 (2p)

Antag att man vid en undersökning väljer ett stickprov om 300 personer och att man får ett bortfall om 20%. Av de svarande är det 90 som svarat ja på en viss fråga. Ange teoretiskt minimi- och maximivärde för den procentuella andelen ja-svarare i hela stickprovet om 300 personer.

Uppgift 3 (4p)

Distributionsföretaget som sköter utdelning av morgontidningar i en stad vill undersöka det genomsnittliga antalet morgontidningar per hushåll. De totalt 4000 hushållen i staden delas in i 400 kluster om 10 hushåll vardera, och ett OSU om 4 kluster dras. Följande resultat erhålles.

Kluster	Antal morgontidningar										Totalt
1	1	2	1	3	3	2	1	4	1	1	19
2	1	3	2	2	3	1	4	1	1	2	20
3	2	1	1	1	1	3	2	1	3	1	16
4	1	1	3	2	1	5	1	2	3	1	20

Beräkna det genomsnittliga antalet morgontidningar per hushåll och bilda ett 95% konfidensintervall.

Uppgift 4 (4p)

En fabrik tillverkar spolarvätska, som distribueras i 1-liters plastflaskor. Vikten för en fylld flaska är en slumpvariabel då det finns en slumpmässig variation i mängden vätska som fylls i en flaska. En återförsäljare har beställt ett parti om 250 flaskor. För att kontrollera sitt parti drar återförsäljaren slumpmässigt 30 flaskor ur partiet och kontrollväger dessa. Genom att subtrahera plastflaskans vikt, som kan anses konstant, kommer hon fram till att genomsnittsinnehållet i stickprovet är 1.01 liter med en standardavvikelse om 0.02 liter.

a) Beräkna ett 95% konfidensintervall för genomsnittsinnehållet i flaskorna i hela partiet om 250 flaskor. (2p)

b) Beräkna ett 95% konfidensintervall för hur mycket extra eller för lite spolarvätska samtliga flaskor i hela partiet innehåller. (2p)

Uppgift 5 (2p)

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

Uppgift 6 (4p)

Man vill undersöka hur stora utgifter hushåll med barn i en viss region har för livsmedel per vecka. Man har tillgång till en lista över de 250 familjerna i regionen, men man vet inte vilka av hushållen som har hemmavarande barn. Man drar ett OSU om 50 familjer och finner att 42 av familjerna har åtminstone ett hemmavarande barn. För dessa beräknar man följande, där X är utgifter (i dollar) för livsmedel per vecka:

$$\sum_{i=1}^{42} x_{1i} = 1720 \qquad \sum_{i=1}^{42} x_{1i}^2 = 72200$$

Beräkna den genomsnittliga utgiften för livsmedel per vecka bland hushåll med barn i regionen och bilda ett 95% konfidensintervall.

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:

$$\widehat{N_0}\cdot\bar{x}_0 \pm z_{\alpha/2}\cdot\sqrt{N^2\cdot\frac{\frac{1}{n-1}\cdot\left(\sum_{i=1}^{n_0} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n_0} x_i\right)^2\right)}{n}\cdot\left(1-\frac{n}{N}\right)}.$$

$$\widehat{N_0}\cdot p_0 \pm z_{\alpha/2}\cdot\sqrt{N^2\cdot\frac{\frac{n_0\cdot p_0}{n}\cdot\left(1-\frac{n_0\cdot p_0}{n}\right)}{n-1}\cdot\left(1-\frac{n}{N}\right)}.$$

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:

$$\begin{aligned}
\hat{\mu} &= \bar{x}_{st} = W_1 \cdot \bar{x}_1 + W_2 \cdot \bar{x}_2 + \ldots + W_L \cdot \bar{x}_L = \sum_{i=1}^{L} W_i \cdot \bar{x}_i \\
\hat{\tau} &= N \cdot \bar{x}_{st} \\
\hat{P} &= p_{st} = W_1 \cdot p_1 + W_2 \cdot p_2 + \ldots + W_L \cdot p_L = \sum_{i=1}^{L} W_i \cdot p_i
\end{aligned}$$

2.2 Konfidensintervall

$$\begin{aligned}
\mu &: \quad \bar{x}_{st} \pm z_{\alpha/2} \cdot \sqrt{\sum_{i=1}^{L} W_i^2 \cdot \frac{s_i^2}{n_i} \cdot \left(1 - \frac{n_i}{N_i}\right)} \\
\tau &: \quad N \cdot \bar{x}_{st} \pm z_{\alpha/2} \cdot \sqrt{N^2 \cdot \sum_{i=1}^{L} W_i^2 \cdot \frac{s_i^2}{n_i} \cdot \left(1 - \frac{n_i}{N_i}\right)} \\
P &: \quad p_{st} \pm z_{\alpha/2} \cdot \sqrt{\sum_{i=1}^{L} W_i^2 \cdot \frac{p_i \cdot (1 - p_i)}{n_i - 1} \cdot \left(1 - \frac{n_i}{N_i}\right)} \\
A &: \quad N \cdot p_{st} \pm z_{\alpha/2} \cdot \sqrt{N^2 \cdot \sum_{i=1}^{L} W_i^2 \cdot \frac{p_i \cdot (1 - p_i)}{n_i - 1} \cdot \left(1 - \frac{n_i}{N_i}\right)}
\end{aligned}$$

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

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

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

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

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

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

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

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

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

3.2 *PPS*–urval av kluster

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

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

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

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

3.3 Urvalsdimensionering

***pps*–urval**

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

4 Tvåstegs klusterurval

4.1 OSU av kluster

Väntevärdesriktiga skattningar

$$\hat{\mu}_{t2} = \left(\frac{N}{M_0}\right) \cdot \widehat{\overline{T}} = \left(\frac{N}{M_0}\right) \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} M_i \cdot \bar{x}_i$$

$$\hat{P}_{t2} = \left(\frac{N}{M_0}\right) \cdot \frac{1}{n} \cdot \sum_{i=1}^{n} M_i \cdot p_i$$

Kvotskattningar

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

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

4.2 *PPS*–urval

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

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

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.