Exp

Försättsblad till skriftlig tentamen vid Linköpings Universitet

Datum för tentamen	2010-01-14
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	KÅRA
Tid	14-18
Kurskod	732G19
Provkod	TENT
Kursnamn/benämning **Provnamn/benämning**	Utredningskunskap I Tentamen
Institution	IDA
Antal uppgifter som ingår i tentamen	4
Jour/Kursansvarig Ange vem som besöker salen	Lotta Hallberg
Telefon under skrivtiden	
Besöker salen ca kl.	16.00
Kursadministratör/kontaktperson (namn + tfnr + mailaddress)	Carita Lilja, 1463, carli@ida.liu.se
Tillåtna hjälpmedel	Räknedosa, med tentan vidhäftad formelsamling
Övrigt	G=12, VG=16
Vilken typ av papper ska användas, rutigt eller linjerat	
Antal exemplar i påsen	6

Tentamen i Utredningskunskap, 2010-01-14

Skrivtid: kl: 14-18
Hjälpmedel: Räknedosa, med tentan vidhäftad formelsamling
Jourhavande lärare: Lotta Hallberg
Redovisa och motivera kort alla dina lösningar.

All data i denna tenta är påhittad.

Vid en skidanläggning så finns många turister som hyr stugor. För att förbättra för turisterna så görs vissa undersökningar. Se uppgifterna nedan.

1

Vid skidanläggningen så finns en skidskola i utförsåkning. Skolan har fyra nivåer som kallas grön, blå, röd och svart, där grön är för nybörjare och svart för mycket erfarna åkare. Då 1500 turister genomgått utbildningen drog man slumpmässigt ett urval om 150 deltagare för att ta reda på deras åsikt om utbildningen. Man ställde därför frågan till de utvalda om de var nöjda med utbildningen. Innan man drog urvalet så stratifierade man på typ av nivå. Resultat:

Stratum	*Antal nöjda i urvalet*	N_i	n_i
Grön	51	590	60
Blå	39	490	50
Röd	18	310	30
Svart	5	110	10
		1500	150

a) Förklara hur ett stratifierat urval går till och varför det kan tänkas vara lämpligt att stratifiera på nivå här. 1p

b) Skatta andelen som är nöjda med utbildningen bland de 1500 som deltagit. Bilda ett 95% konfidensintervall. 3p

c) Om man vill göra om undersökningen nästa säsong hur stora urval bör man då dra ur varje stratum om man använder Neyman-allokering. Låt totala urvalsstorleken vara 150 även denna gång. Anta vidare att totala antalet deltagare i varje nivå är densamma som detta år och att andelen nöjda är ungefär densamma. 2p

d) Skatta med ett 90% konfidensintervall totala antalet nöjda deltagare i gruppen på svart nivå. (Normalapproximera även om det inte är tillåtet här) 2p

2

a) Ett fackförbund för anställda inom skidanläggningar i landet vill göra en undersökning om hur de anställda ställer sig till ett nytt arbetstidsavtal. Om man vill dra ett OSU bland de 7500 anställda och därefter bilda ett konfidensintervall för andelen positiva till avtalet med en konfidensgrad på 95% och en felmarginal på högst 5%, hur stort urval ska man då minst dra? 2p

b) Anta att undersökningen ovan genomförs och man skattar andelen positiva till 65%. Bilda ett 95% konfidensintervall för andelen positiva bland alla medlemmar. 2p

3
Man vill undersöka intresset hos turisterna för deltagande i en viss tävling. Totalt finns 10 stugbyar i området och man drar 4 stycken med sannolikheter proportionellt mot antalet stugor i byn. Sk pps-urval. Sedan dras slumpmässigt 10% av stugorna ur de 4 byarna. Man frågar sedan i varje stuga i urvalet om intresse finns för att delta i tävlingen. Resultat:

By nr	*Totala antalet stugor i byn*	*Antal positiva bland de utvalda*
1	51	2
2	32	2
3	68	5
4	80	5

a) Skatta med ett 95% konfidensintervall andelen stugor i de 10 byarna som är positiva till att delta i tävlingen. (Att en stuga är positiv menar jag att minst en person i stugan är positiv.) 3p
b) Om dragningen av byar skett med ett OSU istället, hur stor skattas då andelen positiva stugor till? (ej KI) 2p

4
En pulka-tävling ska arrangeras. För att kunna göra en skattning på hur lång tid tävlingen kommer ta att genomföra så provåker 15 slumpmässigt valda personer backen och de får en total tid på 7 minuter och 50 sekunder och standardavvikelsen blir 25 sekunder. Totalt har 50 personer anmält sig till tävlingen. Man räknar med att det tar först 10 minuter innan man kan starta och sedan är går det 1 minut mellan starterna. Till sist tar prisutdelningen 10 minuter.
a) Skatta hur lång tid det tar att genomföra tävlingen. 1p
b) Bilda ett 95% konfidensintervall för tävlingens totala tid. (uteslut ändlighetskorrektionen) 2p

Formelblad i Surveymetodik (732G04, 732G02-C, 732G90-C) 732G19

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

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

Konfidensintervall

$$
\begin{aligned}
\mu_1 &: \bar{x}_1 \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}\left(1-\frac{n_1}{N_1}\right)} \\
P_1 &: p_1 \pm z_{\alpha/2}\sqrt{\frac{p_1(1-p_1)}{n_1-1}\left(1-\frac{n_1}{N_1}\right)} \\
\tau_1 &: N\cdot\bar{x}' \pm z_{\alpha/2}\cdot\sqrt{N^2\cdot\frac{s'^2}{n}\left(1-\frac{n}{N}\right)} \\
A_1 &: N\cdot p' \pm z_{\alpha/2}\cdot\sqrt{N^2\cdot\frac{p'(1-p')}{n-1}\left(1-\frac{n}{N}\right)}
\end{aligned}
$$

där

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

och

$$
\begin{aligned}
\bar{x}' &= \frac{1}{n}\sum_{i=1}^{n} x_i' \\
s'^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i'-\bar{x}')^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n}(x_i')^2 - n\cdot(\bar{x}')^2\right) \\
p' &= \frac{\text{Antal element från redovisningsgruppen i urvalet med } x=1}{n}
\end{aligned}
$$

1.5 Ramproblem

Konfidensintervall vid övertäckning

$$
\begin{aligned}
\mu_0 &: \quad \bar{x}_0 \pm z_{\alpha/2}\cdot\sqrt{\frac{s_0^2}{n_0}\cdot\left(1-\frac{n}{N}\right)} \\
P_0 &: \quad p_0 \pm z_{\alpha/2}\cdot\sqrt{\frac{p_0(1-p_0)}{n_0-1}\cdot\left(1-\frac{n}{N}\right)}
\end{aligned}
$$

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

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

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

Kvotskattningar

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

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

4.2 *PPS*–urval

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

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

5 Bortfallsstratumansatsen

Punktskattningar och konfidensintervall

$$\begin{aligned} \hat{\mu}_{st} &= \frac{n'_S}{n} \cdot \bar{x}_S + \frac{n'_B}{n} \cdot \bar{x}_B \\ \hat{P}_{st} &= \frac{n'_S}{n} \cdot p_S + \frac{n'_B}{n} \cdot p_B \end{aligned}$$

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

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

Slutligt bortfall

$$\begin{aligned} b &= (\text{Bortfallsandel i fas 1}) \times (\text{Bortfallandel i fas 2}) \\ &= \frac{n'_B}{n} \cdot \frac{n''_B}{n_B} \end{aligned}$$

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