KARLSTADS UNIVERSITET

12307683

# TENTAMEN / EXAMINATION

Fylls i av **student** / *To be completed by the* ***student***

| | |
|---|---|
| Skriv anonymiseringskoden på samtliga svarsblad / *Write your anonymity code on each sheet* | Anonymiseringskod / *Anonymity code*: NEGC47-0006-RDO |
| Provbenämning / *Exam name*: Tentamen Undersökningsmetodik | Oanmäld: |
| Kurskod / *Course code*: NEGC47 | Modul / *Module*: 2010 |
| Tentamensdatum / *Examination date*: 2021-01-14 | |
| Jag har tagit del av regler som gäller vid tentamen / *I have read the current rules for examinations* ☐ Ja / *Yes* | Antal inlämnade blad med anonymiseringskod / *Number of sheets with anonymity code*: |

Fylls i av **skrivvakt** / *To be completed by the* ***invigilator***

| | |
|---|---|
| Kontroll av legitimation / *Identification checked* ☐ Ja / Yes | Härmed intygas att kontroller utförts / *This is to certify that the checks have been carried out* |
| Kontroll av inlämnade blad / *Answer sheets checked* ☐ Ja / Yes | Tydlig sign. / *Signature* |
| Inlämningstid / *Time of submission*: __ : __ | ........................................ |

Fylls i av **lärare** / *To be completed by the* ***examiner***

*Bedömning av uppgifter / Questions attempted*

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | ~ |
|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 3,5 | 8,5 | 7 | 9 | 4 | | | | | |
| 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | ~ |
| | | | | | | | | | | |
| 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | ~ |
| | | | | | | | | | | |

| | |
|---|---|
| Totalt antal poäng / *Total points*: 42 | Examin. lärare / Kursansvarig signatur / *Signature of the examiner* |
| Betyg / *Grade*: G | Namnförtydligande / *Clarification of the signature* |

12307683

Försättsbladet ska alltid lämnas in även om ingen uppgift behandlats / *Examination should always be submitted even if no questions are answered*

Utskriven 2021-01-15 kl. 15:12:55

NEGC47-0006-RDO

Fråga 1

1 a) Man gör en kvalitetsdeklaration för att kunna bedöma värdet av informationen i undersökningen. Man gör användarna uppmärksamma på fel som kan göra informationen mindre användbar i ger bättre beslutsunderlag och gör det lättare för användaren att kunna precisera kvalitetskrav inför nya undersökningar.

Man gör alltså användaren uppmärksam på i vilka situationer materialet i undersökningen är tillräckligt som beslutsunderlag och när det inte är tillräckligt. Kvalitetsdeklarationen gör att man minskar risken att man som användare "litar blint" på informationen för att den ser bra ut.

I PSU-rapporten ingår komponenterna:

- Innehåll
- Tillförlitlighet
- Aktualitet
- Tillgänglighet
- Jämförbarhet

4p

1b) De som hade varit röstberättigade om det hade varit riksdagsval andra söndagen i september aktuellt undersökningsår och som är bokförda i Sverige.

1p

1 c) Registret över totalbefolkning (RTB)

Registret innehåller aktuell information eftersom det uppdateras dagligen

→

NEGC47 - 0006 - RDO

Fråga 1

1 c) Ramen innehåller värden på bakgrunds- och hjälpvariabler (demografiska variabler som kön, ålder, boendeort etc.) 2p

1 d) Bortfall och mätning anses ha störst betydelse för resultatet.

(svar helt eller delvis saknas för vissa personer)

(frågor missförstås eller är svårbesvarade pga. minnesproblematik) 1p

1 e) Antingen genom datorstödda telefonintervjuer eller genom webbenkät. 1p

1 f) Eftersom undersökningen genomförts med huvudsakligen samma metod och samma huvudsakliga definitioner sedan 1972 (utom nov 1981 t.o.m. nov. 1983) är statistiken jämförbar över tid. 1p

NEGC47-0006-RDO

Fråga 2

2 a) Obundet slumpmässigt urval (OSU) är teoretiskt enkelt men kan ge dålig precision, stratifierat urval kan ge bättre precision än vad OSU gör och är bra vid sneda urval. Om man använder systematiskt urval kan man få en hög precision om man använder en lämpligt sorterad ram. Gruppurval är billigt och kräver ingen ram men ger ofta sämre precision än vad OSU gör.

När det kommer till hur stickprovet fördelas mellan strata (allokering) får man större precisionsförbättring ju mer stratummedelvärdena skiljer sig åt och avviker från det totala medelvärdet vid proportionellt stratifierat urval i jämförelse med OSU med återläggning

Neymanallokering kallas också optimal allokering och ger högst precision men den är svår att tillämpa praktiskt eftersom man då måste ha kännedom om standardavvikelserna innan allokeringen görs. 3,5 p

2 b) Genom att öka stickprovsstorleken ökar precisionen. 0 p

NEGC47-0006-RDO

Fråga 3

| Hushåll | A | B | C | |
|---|---|---|---|---|
| 1 | 52 | 51 | 28 | 131 |
| 2 | 52 | 40 | 48 | 140 |
| 3 | 63 | 45 | 57 | 165 |
| 4 | 52 | 47 | 44 | 143 |
| $y_{i.}$ | 219 | 183 | 177 | 579 |

3a) Randomiserat blockförsök

Modell: $y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}$ $\begin{cases} i = A, B, C \ (1, 2, 3) & \text{Diskmedel} \\ j = 1, 2, 3, 4 & \text{Hushåll} \end{cases}$

Antaganden: $\varepsilon_{ij}$ oberoende, NF, homoskedastisk

$H_0: \tau_A = \tau_B = \tau_C = 0$

$H_1$: Något $\tau_i \neq 0$

$\alpha = 0{,}05 \qquad F_0 = \dfrac{MS_{Tr}}{MS_E}$

$$SS_T = \sum\sum y_{ij}^2 - \frac{1}{N} y_{..}^2 = 28789 - \frac{1}{12} \cdot 579^2 = 852{,}25$$

$52^2 + 52^2 + 63^2 + 52^2 + 51^2 + \ldots + 44^2$

$$SS_{Tr} = \frac{\sum y_{i.}^2}{b} - \frac{1}{N} y_{..}^2 = \frac{219^2 + 183^2 + 177^2}{4} - \frac{1}{12} \cdot 579^2 = 258$$

$$SS_{Bl} = \frac{\sum y_{.j}^2}{a} - \frac{1}{N} y_{..}^2 = \frac{131^2 + 140^2 + 165^2 + 143^2}{3} - \frac{1}{12} \cdot 579^2 = 208{,}25$$

$$SS_E = SS_T - SS_{Tr} - SS_{Bl} = 852{,}25 - 258 - 208{,}25 = 386$$

| Variationskälla | SS | df | MS | fo |
|---|---|---|---|---|
| Tr | 258 | 2 | 258/2 = 129 | 129/64,3 ≈ 2,01 |
| Bl | 208,25 | 3 | 208,25/3 ≈ 69,42 | 69,42/64,3 ≈ 1,08 |
| E | 386 | 6 | 386/6 ≈ 64,3 | |
| T | 852,25 | 11 | | |

$H_0$ förkastas om $f_0 > f_{0{,}05,\,2,\,6} = 5{,}14$

$2{,}01 < 5{,}14$ så $H_0$ kan ej förkastas.

På 5% signifikansnivå kan vi påstå att det inte finns någon skillnad i medelvärde mellan diskmedlen.

7p

NEGC47-0006-RDO

Fråga 3

3b) I diagram 9 tycker jag att residualerna följer linjen väl, möjligtvis ligger de något högt i början och något lågt i slutet, men jag skulle ändå påstå att det ser ut som att de är normalfördelade. I tabell 2 testas normalfördelningen, signifikansnivån är högre än 0,05, vilket innebär att nollhypotesen inte kan förkastas → residualerna är på 5% sig nivå normalfördelade.

Antagandet om normalfördelning är alltså uppfyllt.

/R

I diagram 5 kan vi se att Diskmedel 1 verkar ha lägst varians (spridning), likaså i boxplotten (3), det tyder på att variansen inte är lika. I tabell 10 testas detta och sig. 0,364 > 0,05 vilket innebär att H0 inte kan förkastas, variansen beror inte på värdena på de oberoende variablerna → variansen är olika

/R

/V H0: Lika varians kan ej förkastas.

Oberoende?

1,5p

NEG(47-0006-RDO

Fråga 4

4 a) $H_0: \mu_i = \mu_j \quad \forall\ i,j,\ i \neq j$

$H_1: \mu_i \neq \mu_j$

$\alpha = 0{,}05$

Teststatistika $|\bar{y}_{i\cdot} - \bar{y}_{\cdot j}|$

$H_0$ förkastas om $|\bar{y}_{i\cdot} - \bar{x}_{j\cdot}| > LSD$

$$LSD = \underset{\alpha/2}{t_{0{,}025,6}} \sqrt{MS_E\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} = 2{,}45 \cdot \sqrt{64{,}3\left(\frac{1}{4} + \frac{1}{4}\right)} = 13{,}89$$

| i | j | $\lvert\bar{y}_{i\cdot} - \bar{y}_{\cdot j}\rvert$ | Slutsats | |
|---|---|---|---|---|
| A | B | 9 | $H_0$ förkastas ej | $(9 < 13{,}89)$ |
| A | C | 10,5 | $H_0$ förkastas ej | $(10{,}5 < 13{,}89)$ |
| B | C | 1,5 | $H_0$ förkastas ej | $(1{,}5 < 13{,}89)$ |

$\bar{y}_A = 219/4 = 54{,}75$

$\bar{y}_B = 183/4 = 45{,}75$

$\bar{y}_C = 177/4 = 44{,}25$

På 5% signifikansnivå finns ingen skillnad i medelvärde mellan någon av diskmedlen.

/R

7p

4 b)

0p

NEGC47 - 0006 - RDO

Fråga 5

| | A | C |
|---|---|---|
| | 52 | 28 |
| | 57 | 52 |
| | 59 | 54 |
| | 69 | 67 |
| | 59 | 52 |
| | 70 | 74 |
| | 53 | 48 |
| | 66 | 57 |
| $\Sigma$ | 485 | 432 |

5a) $\bar{x} = \frac{1}{n} \sum y_i$

$\bar{x} = \frac{1}{8} \cdot 485 = 60{,}625$

$\boxed{n = 8 \quad N = 120}$

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

$s^2 = \frac{1}{8-1}\left(29741 - \frac{1}{8} \cdot 485^2\right) = \frac{1}{7}\left(29741 - 29403{,}12\right.$

$s^2 = \frac{1}{7} \cdot (29741 - 29403{,}125) \approx 48{,}268$

$\hat{V}(\bar{y}) = \left(1 - \frac{8}{120}\right) \cdot \frac{48{,}268}{8} = 5{,}63$

95% KI: $60{,}625 \pm \underline{1{,}96} \cdot \sqrt{5{,}63}$ t-värde

$55{,}97 \leq \mu \leq 65{,}28$

Med 95% ~~sannolikhet~~ säkerhet är medelvärdet för populationen mellan 55,97 och 65,28.

Antagande: Normalfördelning 4,5p

5b) $X_{tot} = 6283$

$N_i = 120$

$\bar{X}_c = 6283/120 \approx 52{,}36$

$\bar{x}_c = 432/8 = 54$

$p_{kvot} = \frac{485}{432} \approx 1{,}12$

$\bar{Y}_{kvot} = \frac{48{,}268}{54} \cdot 52{,}36 = 46{,}802$

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

$\bar{Y}_{kv} = \frac{\bar{y}}{\bar{x}} \cdot \mu_x$

$= \frac{60{,}625}{54} \cdot 52{,}36$

$s^2 = \frac{\sum (A_i - p_{kv} \cdot C_i)^2}{n-1}$

$$s^2 = \frac{(52 - 1{,}12 \cdot 28)^2 + (57 - 1{,}12 \cdot 52)^2 + \ldots + (66 - 1{,}12 \cdot 57)^2}{8-1} = 91{,}13$$

NEBC47-0006-RDD

Fråga 5

5 b forts) $s^2_{kv} = 91{,}13$

$$D(\bar{y}_{kv}) = \left(1 - \frac{n}{N}\right)\frac{s^2}{n} = \left(1 - \frac{8}{120}\right)\frac{91{,}13}{8} = 10{,}63$$

/ok

4,5 p

NEGC47-0006-RDO

Fråga 6

| | Ålder | $N_i$ | $\bar{X}_i$ | $s_i$ |
|---|---|---|---|---|
| 1 | 55-64 | 400 | 4 | 3 |
| 2 | 40-54 | 1000 | 10 | 4 |
| 3 | 25-39 | 600 | 9 | 2 |

$n_1 = 40$

$n_2 = 100$

$n_3 = 60$

6 a) $\bar{x}_{st} = \sum_{i=1}^{L} \frac{N_i}{N} \cdot \bar{x}_i = \frac{400}{2000} \cdot 4 + \frac{1000}{2000} \cdot 10 + \frac{600}{2000} \cdot 9 = 0{,}8 + 5 + 2{,}7 = 8{,}5$

$$\hat{V}(\bar{x}) = \sum \left(\frac{N_i}{N}\right)^2 \left(1 - \frac{n_i}{N_i}\right) \frac{s_i^2}{n_i} = \left(\frac{400}{2000}\right)^2 \left(1 - \frac{40}{2000}\right) \frac{3^2}{40} + \left(\frac{1000}{2000}\right)^2 \left(1 - \frac{100}{2000}\right) \frac{4^2}{100}$$

$$+ \left(\frac{600}{2000}\right)^2 \left(1 - \frac{60}{2000}\right) \frac{2^2}{60} = 0{,}00882 + 0{,}038 + 0{,}00582$$

$$= 0{,}05264$$

95% KI: $8{,}5 \pm 1{,}96 \cdot \sqrt{0{,}05264} \rightarrow 8{,}05 \leq \mu \leq 8{,}95$

4p

Med 95% ~~sannolikhet~~ säkerhet kan man påstå att den genomsnittliga tiden som de anställda spenderar på gymmet är mellan 8,05 och 8,95 h.

6 b) $n \geq \dfrac{\sum N_i^2 \sigma_i^2 / w_i}{N^2 \frac{B^2}{c^2} + \sum N_i \sigma_i^2}$, $w_i = \dfrac{n_i}{n}$

Du har räknat för PSU

$$n \geq \frac{\frac{400^2 \cdot 3^2}{40/200} + \frac{1000^2 \cdot 4^2}{100/200} + \frac{600^2 \cdot 2^2}{60/200}}{2000^2 \cdot \frac{0{,}45^2}{1{,}96^2} + 400 \cdot 3^2 + 1000 \cdot 4^2 + 600 \cdot 2^2}$$

$B = \dfrac{8{,}95 - 8{,}05}{2} = 0{,}45$

$$= \frac{44\,000\,000}{232\,849{,}646} = 188{,}96 \approx 189$$

För att få samma precision som 6a med OSU behövs ett stickprov på minst 189 anställda.

0p

NEGC47-0006-RDO

Fråga 6

6c)

$$n_i = \frac{N_i \sigma_i}{\sum N_j \sigma_j} \cdot n$$

$$n_1 = \frac{400 \cdot 3}{400 \cdot 3 + 1000 \cdot 4 + 600 \cdot 2} \cdot 200 = 37{,}5 \approx 38$$

$$n_2 = \frac{1000 \cdot 4}{400 \cdot 3 + 1000 \cdot 4 + 600 \cdot 2} \cdot 200 = 125$$

$$n_3 = \frac{600 \cdot 2}{400 \cdot 3 + 1000 \cdot 4 + 600 \cdot 2} \cdot 200 = 37{,}5 \approx 38$$

För att få samma precision som i a-uppgiften med Neymann-allokering hade man tagit 38 anställda mellan 55 och 64 år, 125 anställda mellan 40 och 54 år och 38 anställda mellan 25 och 39 år.

Du har bara räknat på allokering utifrån n=200

0p