Lösningsförslag till Tentamen 100608 i 732G26 Surveymetodik med uppsats

1. a) $\hat{t} = N \cdot \bar{y} = 55560 \cdot (5540/200) = 55560 \cdot 27.70 = 1539012$ timmar

   99% konfidensintervall (z = 2.576):

$$N \cdot \bar{y} \pm 2.576 \cdot N \cdot \sqrt{\left(1-\frac{n}{N}\right) \cdot \frac{s^2}{n}} \Rightarrow 55560 \cdot \left(27.70 \pm 2.576 \cdot \sqrt{\left(1-\frac{200}{55560}\right) \cdot \frac{15^2}{200}}\right)$$

$$\Rightarrow 1539012 \pm 151531 = (1387481, 1690543)$$

   b) Bredden för intervall för $t \leq 280000$ ➔ Bredden för intervall för $\mu \leq 5.04$.

$$n_0 \geq \frac{4 \cdot 2.576^2 \cdot 15^2}{5.04^2} \approx 236$$

$$\Rightarrow n = \frac{236}{1+\frac{236}{55560}} \approx 235$$

   c) Sätt de 10 individernas värden till 0 och räkna om medelvärde och standardavvikelse.
   Beteckna övertäckningens antal, medelvärde och standardavvikelse med $n_ö$, $\bar{x}_ö$ resp. $s_ö$.

$$\bar{u} = \frac{5540 - 10 \cdot 28.5}{200} = 26.275$$

$$s_u = \sqrt{\frac{(n-1) \cdot s^2 + n \cdot \bar{x}^2 - \left((n_ö - 1) \cdot s_ö^2 + n_ö \cdot \bar{x}_ö^2\right) - n \cdot \bar{u}^2}{n-1}} =$$

$$= \sqrt{\frac{199 \cdot 15^2 + 200 \cdot 27.70^2 - \left(9 \cdot 10.9^2 + 10 \cdot 28.5^2\right) - 200 \cdot 26.275^2}{199}} \approx 16$$

   Ny punktskattning: $55560 \cdot 26.275 \approx 1459839$

   99% K.I. $55560 \cdot \left(26.275 \pm 2.576 \cdot \sqrt{\left(1-\frac{200}{55560}\right) \cdot \frac{16^2}{200}}\right) \approx 1459839 \pm 161633 =$

$$= (1298206, 1621472)$$

2. a) Stratifierat urval. Punktskattning av medeltal:

$$\bar{x}_{str} = \frac{N_1}{N} \cdot \bar{x}_1 + \frac{N_2}{N} \cdot \bar{x}_2 = 0.969 \cdot \frac{3350}{150} + 0.071 \cdot \frac{1800}{50} \approx 24.197$$

   Punktskattning av total

   99% konfidensintervall

$$\bar{x}_{str} \pm 2.576 \cdot \sqrt{\left(\frac{N_1}{N}\right)^2 \cdot \left(1-\frac{n_1}{N_1}\right) \cdot \frac{\hat{p}_1 \cdot (1-\hat{p}_1)}{n_1 - 1} + \left(\frac{N_2}{N}\right)^2 \cdot \left(1-\frac{n_2}{N_2}\right) \cdot \frac{\hat{p}_2 \cdot (1-\hat{p}_2)}{n_2 - 1}}$$

   Eftersom populationen antas vara mycket stor, så blir även N_1 och N_2 mycket stora och Ändlighetskorrektionerna kan därför utelämnas (dvs. sättas till 1).

   ➔

$$0.639 \pm 1.96 \cdot \sqrt{(0.58)^2 \cdot \frac{(116/150)\cdot(34/150)}{149} + (0.42)^2 \cdot \frac{(68/150)\cdot(82/150)}{149}}$$

$$\Rightarrow 0.639 \pm 0.051 = (0.588, 0.690)$$

b) Optimal allokering blir här Neyman-allokering eftersom man kan anta lika stora kostnader.

$$n_h = n \cdot \frac{N_h \cdot S_h}{N_1 \cdot S_1 + N_2 \cdot S_2} = n \cdot \frac{\frac{N_h}{N} \cdot S_h}{\frac{N_1}{N} \cdot S_1 + \frac{N_2}{N} \cdot S_2}$$

S_h approximeras med $\sqrt{\frac{150}{149} \cdot \hat{p}_h \cdot (1 - \hat{p}_h)}$ från den första undersökningen. (Det är OK att använda $\sqrt{\hat{p}_h \cdot (1 - \hat{p}_h)}$ också)

n = 300 ➔

$$n_1 = 300 \cdot \frac{0.58 \cdot \sqrt{\frac{150}{149} \cdot \frac{116}{150} \cdot \frac{34}{150}}}{0.58 \cdot \sqrt{\frac{150}{149} \cdot \frac{116}{150} \cdot \frac{34}{150}} + 0.42 \cdot \sqrt{\frac{150}{149} \cdot \frac{68}{150} \cdot \frac{82}{150}}} \approx 161$$

$$\Rightarrow n_2 = 300 - 161 = 139$$

3. Tvåstegs klusterurval

a) Eftersom klusterstorlekarna varieras så pass mycket är en kvotskattning lämpligast

$$\hat{\bar{y}}_r = \frac{\sum_{i \in S} M_i \cdot \bar{y}_i}{\sum_{i \in S} M_i} = \frac{50 \cdot 47 + 135 \cdot 59 + 40 \cdot 40 + 73 \cdot 46 + 109 \cdot 49}{50 + 135 + 40 + 73 + 109} = \frac{20614}{407} \approx$$

$$\approx 50.65$$

$$\hat{V}(\hat{\bar{y}}_r) = \frac{1}{\bar{M}^2} \cdot \left[\left(1 - \frac{n}{N}\right) \cdot \frac{1}{n} \cdot \frac{1}{n-1} \cdot \sum_{i \in S} \left(M_i \cdot \bar{y}_i - M_i \cdot \hat{\bar{y}}_r\right)^2 + \right.$$

$$\left. + \frac{1}{n \cdot N} \cdot \sum_{i \in S} M_i^2 \cdot \left(1 - \frac{m_i}{M_i}\right) \cdot \frac{s_i^2}{m_i} \right]$$

$$\sum_{i\in S}\left(M_i\cdot\bar{y}_i-M_i\cdot\hat{\bar{y}}_r\right)^2=\left(50\cdot 47-50\cdot\frac{20614}{407}\right)^2+\left(135\cdot 59-135\cdot\frac{20614}{407}\right)^2+$$

$$+\left(40\cdot 40-40\cdot\frac{20614}{407}\right)^2+\left(73\cdot 46-73\cdot\frac{20614}{407}\right)^2+\left(109\cdot 49-109\cdot\frac{20614}{407}\right)^2\approx$$

$$\approx 1633267.8$$

$$\sum_{i\in S}M_i^2\cdot\left(1-\frac{m_i}{M_i}\right)\cdot\frac{s_i^2}{m_i}=50^2\cdot\left(1-\frac{20}{50}\right)\cdot\frac{18^2}{20}+135^2\cdot\left(1-\frac{20}{135}\right)\cdot\frac{19^2}{20}+$$

$$+40^2\cdot\left(1-\frac{20}{40}\right)\cdot\frac{17^2}{20}+73^2\cdot\left(1-\frac{20}{73}\right)\cdot\frac{22^2}{20}+109^2\cdot\left(1-\frac{20}{109}\right)\cdot\frac{20^2}{20}\approx$$

$$\approx 603736.05$$

$$\Rightarrow$$

$$\hat{V}\left(\hat{\bar{y}}_r\right)=\frac{1}{\left(\frac{50+135+40+73+109}{5}\right)^2}\cdot\left[\left(1-\frac{5}{230}\right)\cdot\frac{1}{5}\cdot\frac{1}{4}\cdot 1633267.8+\right.$$

$$\left.+\frac{1}{5\cdot 230}\cdot 603736.05\right]\approx 12.14$$

➔ 95% konfidensintervall

$$50.65\pm 1.96\cdot\sqrt{12.14}\approx 50.65\pm 6.83\approx(43.82\,,57.48)$$

Med väntevärdesriktig skattning fås

$$\hat{\bar{y}}_{unb}=\frac{\hat{t}_{unb}}{M_0}=\frac{1}{M_0}\cdot\frac{N}{n}\cdot\sum_{i\in S}M_i\cdot\bar{y}_i=\frac{1}{15512}\cdot\frac{230}{5}\cdot 20614\approx 61.13$$

$$\hat{V}\left(\hat{\bar{y}}_{unb}\right)=\left(\frac{1}{M_0}\right)^2\cdot\left[N^2\cdot\left(1-\frac{n}{N}\right)\cdot\frac{1}{n}\cdot\frac{1}{n-1}\cdot\sum_{i\in S}\left(\hat{t}_i-\frac{\hat{t}_{unb}}{N}\right)^2+\right.$$

$$\left.+\frac{N}{n}\cdot\sum_{i\in S}M_i^2\cdot\left(1-\frac{m_i}{M_i}\right)\cdot\frac{s_i^2}{m_i}\right]=$$

$$=\left(\frac{1}{M_0}\right)^2\cdot\left[N^2\cdot\left(1-\frac{n}{N}\right)\cdot\frac{1}{n}\cdot\frac{1}{n-1}\cdot\sum_{i\in S}\left(M_i\cdot\bar{y}_i-\frac{M_0}{N}\cdot\hat{\bar{y}}_{unb}\right)^2+\right.$$

$$\left.+\frac{N}{n}\cdot\sum_{i\in S}M_i^2\cdot\left(1-\frac{m_i}{M_i}\right)\cdot\frac{s_i^2}{m_i}\right]$$

$$\sum_{i\in S}\left(M_i\cdot\bar{y}_i-\frac{M_0}{N}\cdot\hat{\bar{y}}_{unb}\right)^2=\left(50\cdot 47-\frac{15512}{230}\cdot 60.13\right)^2+\left(135\cdot 59-\frac{15512}{230}\cdot 60.13\right)^2+$$

$$+\left(40\cdot 40-\frac{15512}{230}\cdot 60.13\right)^2+\left(73\cdot 46-\frac{15512}{230}\cdot 60.13\right)^2+\left(106\cdot 49-\frac{15512}{230}\cdot 60.13\right)^2\approx$$

$$\approx 26005136.2$$

$$\Rightarrow$$

$$\hat{V}\left(\hat{\bar{y}}_{unb}\right)=\left(\frac{1}{15512}\right)^2\cdot\left[230^2\cdot\left(1-\frac{5}{230}\right)\cdot\frac{1}{5}\cdot\frac{1}{4}\cdot 26005136.2+\frac{230}{5}\cdot 603736.05\right]\approx 279.76$$

95% konfidensintervall

$$61.13\pm 1.96\cdot\sqrt{279.76}\approx 61.13\pm 32.78=(28.35,93.91)$$

b) För att urvalet skall vara självvägt måste urvalsvikterna vara lika stora för varje individ.
Urvalsvikterna är

$$w_{ij}=\frac{N}{n}\cdot\frac{M_i}{m_j}=\frac{230}{5}\cdot\frac{M_i}{20}$$ som varierar från kluster till kluster

➔ Nej, urvalet är inte självvägt!

c) PPS-urval ger

$$\hat{\bar{y}}_{pps}=\bar{\bar{y}}=\frac{1}{n}\cdot\sum_{i\in S}\bar{y}_i=\frac{1}{5}\cdot(47+59+40+46+49)=48.20$$

$$\hat{V}\left(\hat{\bar{y}}_{pps}\right)=\frac{s_{\bar{y}}^2}{n}=\frac{1}{n}\cdot\frac{1}{n-1}\cdot\sum_{i\in S}\left(\bar{y}_i-\bar{\bar{y}}\right)^2=\frac{1}{n}\cdot\frac{1}{n-1}\cdot\left(\sum_{i\in S}\bar{y}_i^2-n\cdot\bar{\bar{y}}^2\right)=$$

$$=\frac{1}{5}\cdot\frac{1}{4}\cdot\left(47^2+59^2+40^2+46^2+49^2-5\cdot 48.20^2\right)=47.7$$

➔ 95% konfidensintervall:

$$48.20\pm 1.96\cdot\sqrt{47.7}\approx 48.20\pm 13.54=(34.66,61.74)$$

4. a) Bortfallsuppföljning enligt Hansen-Hurvitz:

$$\hat{\bar{y}} = \frac{n_R}{n} \cdot \bar{y}_R + \frac{n_M}{n} \cdot \bar{y}_M = \frac{n_R}{n} \cdot \hat{p}_R + \frac{n_M}{n} \cdot \hat{p}_M = \frac{147}{400} \cdot \frac{109}{147} + \frac{253}{400} \cdot \frac{30}{30} = 0.905$$

$$\hat{V}\left(\hat{\bar{y}}\right) = \frac{n_R - 1}{n-1} \cdot \frac{s_R^2}{n} + \frac{n_M - 1}{n-1} \cdot \frac{s_M^2}{\nu \cdot n} + \frac{1}{n-1} \cdot \left[\frac{n_R}{n} \cdot \left(\bar{y}_R - \hat{\bar{y}}\right)^2 + \frac{n_M}{n} \cdot \left(\bar{y}_M - \hat{\bar{y}}\right)^2 \right]$$

$$\nu = \frac{30}{253} \,;\, s_R^2 = \frac{n_R}{n_R - 1} \cdot \hat{p}_R \cdot \left(1 - \hat{p}_R\right);\, s_R^2 = \frac{\nu \cdot n_R}{\nu \cdot n_R - 1} \cdot \hat{p}_M \cdot \left(1 - \hat{p}_M\right)$$

$\Rightarrow$

$$\hat{V}\left(\hat{\bar{y}}\right) = \frac{146}{399} \cdot \frac{\frac{147}{146} \cdot \frac{109}{147} \cdot \frac{38}{147}}{400} + \frac{252}{399} \cdot \frac{\frac{30}{29} \cdot 1 \cdot 0}{30} +$$

$$+ \frac{1}{299} \cdot \left[\frac{147}{400} \cdot \left(\frac{109}{147} - 0.905 \right)^2 + \frac{253}{400} \cdot \left(1 - 0.905\right)^2 \right] \approx 0.0002284$$

➔ 95% konfidensintervall:

$0.905 \pm 1.96 \cdot \sqrt{0.0002284} \approx 0.905 \pm 0.030 = \left(0.875, 0.935\right)$

b) Poststratifierad skattning:

$N = 1352 + 1831 + 3370 + 5479 = 12032$

$$\bar{y}_{post} = \sum_{h=1}^{4} \frac{N_h}{N} \cdot \bar{y}_{hR} = \frac{1352}{12032} \cdot \frac{14}{17} + \frac{1831}{12032} \cdot \frac{12}{19} + \frac{3370}{12032} \cdot \frac{36}{55} + \frac{5479}{12032} \cdot \frac{47}{56} \approx 0.754$$

Den poststratifierade skattningen avviker tydligt från skattningen i a)-uppgiften. Detta tyder på att urvalet inte är MAR utan snarare NMAR. De som inte svarat på undersökningen kan möjligen vara sådana som fått anställning och inte känner något behov av att upplysa sin tidigare arbetsgivare om detta.