Lösningsförslag till Tentamen 100813 i 732G26 Surveymetodik med uppsats

1. a) $\hat{t} = N \cdot \bar{y} = 20000 \cdot 7300 = 146$ miljoner kronor
   95% konfidensintervall (z = 1.96):
   $$N \cdot \bar{y} \pm 1.96 \cdot N \cdot \sqrt{\left(1-\frac{n}{N}\right) \cdot \frac{s^2}{n}} \Rightarrow 20000 \cdot \left(7300 \pm 1.96 \cdot \sqrt{\left(1-\frac{100}{20000}\right) \cdot \frac{2600^2}{100}}\right)$$
   $$\Rightarrow 146.00 \pm 10.17 \text{ milj. kronor} = (135.83, 156.17) \text{ milj. kronor}$$

   b) Bredden för intervall för $t \leq 20$ milj ➔ Bredden för intervall för $\mu \leq 1000$.
   $$n_0 \geq \frac{4 \cdot 1.96^2 \cdot 2600^2}{1000^2} \approx 104$$
   $$\Rightarrow n = \frac{104}{1+\frac{104}{20000}} \approx 104$$

   c) (i) Efternamnens begynnelsebokstav har knappast något med spenderingsviljan att göra ➔ Systematiskt urval ger ungefär samma noggrannhet som OSU
   (ii) Årsinkomsten torde vara positivt korrelerad med spenderingsviljan ➔ Ett systematiskt urval är sannolikt mer representativt än ett OSU ➔ Bättre noggrannhet än vid OSU
   (iii) Stor risk för periodicitet i registret ➔ Ett systematiskt urval kan slå helt fel och ge värden med mycket liten varians men stor bias ➔ Sämre noggrannhet än vid OSU

2. a) Stratifierat urval. Punktskattning av medeltal:
   $$\hat{p}_{str} = \frac{N_1}{N} \cdot \hat{p}_1 + \frac{N_2}{N} \cdot \hat{p}_2 = \frac{6}{20} \cdot \frac{39}{100} + \frac{14}{20} \cdot \frac{58}{200} = 0.32$$
   Punktskattning av totalantalet
   $$\hat{t}_{str} = N \cdot \hat{p}_{str} = 20000 \cdot 0.32 = 6400$$
   99% konfidensintervall för totalantalet
   $$N \cdot \bar{y}_{str} \pm N \cdot 2.576 \cdot \sqrt{\left(\frac{N_1}{N}\right)^2 \cdot \left(1-\frac{n_1}{N_1}\right) \cdot \frac{\hat{p}_1 \cdot (1-\hat{p}_1)}{n_1 - 1} + \left(\frac{N_2}{N}\right)^2 \cdot \left(1-\frac{n_2}{N_2}\right) \cdot \frac{\hat{p}_2 \cdot (1-\hat{p}_2)}{n_2 - 1}}$$
   ➔
   $$6400 \pm$$
   $$20000 \cdot 2.576 \cdot \sqrt{\left(\frac{6}{20}\right)^2 \cdot \left(1-\frac{100}{6000}\right) \cdot \frac{0.39 \cdot 0.61}{99} + \left(\frac{14}{20}\right)^2 \cdot \left(1-\frac{200}{14000}\right) \cdot \frac{0.29 \cdot 0.71}{199}}$$
   $$\Rightarrow 6400 \pm 1375 = (5025, 7775)$$

b) Fullständig optimal allokering:

$$n_h = n \cdot \frac{N_h \cdot S_h \Big/ \sqrt{c_h}}{N_1 \cdot S_1 \Big/ \sqrt{c_1} + N_2 \cdot S_2 \Big/ \sqrt{c_2}} =$$

S_h approximeras med $s_h = \frac{n_h}{n_h - 1} \cdot \hat{p}_h \cdot (1 - \hat{p}_h)$ från den första undersökningen.

$c_1 = 2 \cdot c_2$

$n = 300$ ➔

$$n_1 = 300 \cdot \frac{6000 \cdot \frac{100}{99} \cdot 0.39 \cdot 0.61 \Big/ \sqrt{2 \cdot c_2}}{6000 \cdot \frac{100}{99} \cdot 0.39 \cdot 0.61 \Big/ \sqrt{2 \cdot c_2} + 14000 \cdot \frac{200}{199} \cdot 0.29 \cdot 0.71 \Big/ \sqrt{c_2}} =$$

$$= 300 \cdot \frac{6000 \cdot \frac{100}{99} \cdot 0.39 \cdot 0.61 \Big/ \sqrt{2}}{6000 \cdot \frac{100}{99} \cdot 0.39 \cdot 0.61 \Big/ \sqrt{2} + 14000 \cdot \frac{200}{199} \cdot 0.29 \cdot 0.71} = 78$$

$$\Rightarrow n_2 = 300 - 78 = 222$$

3. Tvåstegs klusterurval

a) Kvotskattning:

$$\hat{\bar{y}}_r = \frac{\sum_{i \in S} M_i \cdot \bar{y}_i}{\sum_{i \in S} M_i} = \frac{16 \cdot 2.7 + 20 \cdot 3.2 + 12 \cdot 3.8 + 26 \cdot 3.5 + 10 \cdot 4.2}{16 + 20 + 12 + 26 + 10} = \frac{285.8}{84} \approx 3.40 \text{ tkr} = 3400 \text{ kr}$$

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

$m_i / M_i = 0.5.$

$\bar{M} = 84/5 = 16.8$

$$\Rightarrow \frac{1}{16.8^2} \cdot \left[\left(1 - \frac{5}{900}\right) \cdot \frac{1}{4} \cdot \frac{1}{3} \cdot \left((16 \cdot 2.7 - 16 \cdot 3.4)^2 + (20 \cdot 3.2 - 20 \cdot 3.4)^2 + (12 \cdot 3.8 - 12 \cdot 3.4)^2 + \right.\right.$$

$$\left. + (26 \cdot 3.5 - 26 \cdot 3.4)^2 + (10 \cdot 4.2 - 10 \cdot 3.4)^2 \right) +$$

$$\left. + \frac{1}{5 \cdot 900 \cdot 16.8^2} \cdot 0.5 \cdot \left(16^2 \cdot \frac{0.3^2}{8} + 20^2 \cdot \frac{0.5^2}{10} + 12^2 \cdot \frac{0.7^2}{6} + 26^2 \cdot \frac{0.4^2}{13} + 10^2 \cdot \frac{0.9^2}{5}\right)\right] \approx$$

≈ 0.0691

➔ 95% konfidensintervall

$$3.40 \pm 1.96 \cdot \sqrt{0.0691} \approx 3.40 \pm 0.52 \approx (2.88, 3.92) \text{ tkr} = (2880, 3920) \text{ kr}$$

b) PPS-urval ger

$$\hat{\hat{y}}_{pps} = \bar{\bar{y}} = \frac{1}{n} \cdot \sum_{i \in S} \bar{y}_i = \frac{1}{5} \cdot (2.7 + 3.2 + 3.8 + 3.5 + 4.2) = 3.48$$

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

$$= \frac{1}{5} \cdot \frac{1}{4} \cdot \left(2.7^2 + 3.2^2 + 3.8^2 + 3.5^2 + 4.2^2 - 5 \cdot 3.48^2\right) = 0.0654$$

➔ 95% konfidensintervall:

$$3.48 \pm 1.96 \cdot \sqrt{0.0654} \approx 3.48 \pm 0.50 = (2.98, 3.98) \text{ tkr} = (2980, 3980) \text{ kr}$$

4. a) Regressionsskattning:

$$\hat{\bar{y}}_{reg} = \bar{y} + \hat{B}_1 \cdot (\bar{x}_U - \bar{x})$$

$$\hat{B}_1 = \frac{r \cdot s_y}{s_x} = \frac{0.7 \cdot 2600}{6400} \approx 0.2844$$

$$\Rightarrow \hat{\bar{y}}_{reg} = 7300 + 0.2844 \cdot \left(\frac{559000000}{20000} - 27300 \right) \approx 7485 \text{ kronor}$$

$$\Rightarrow \hat{t}_{reg} = 20000 \cdot 7485 = 149.7 \text{ miljoner kronor}$$

95% konfidensintervall:

$$SE\left(\hat{t}_{reg}\right) = N \cdot \sqrt{\left(1 - \frac{n}{N}\right) \cdot \frac{1}{n} \cdot s_y^2 \cdot \left(1 - r^2\right)}$$

$$\Rightarrow$$

$$149.7 \cdot 10^6 \pm 1.96 \cdot 20000 \cdot \sqrt{\left(1 - \frac{100}{20000}\right) \cdot \frac{1}{100} \cdot 2600^2 \cdot \left(1 - 0.7^2\right)} \approx$$

$$\approx 149.7 \cdot 10^6 \pm 7.3 \cdot 10^6 = (142.4, 157.0) \text{ miljoner kronor}$$

b) Det kan inte förutsättas att en person med månadsinkomst 0 inte skulle kunna spendera pengar på semestern. Vederbörande kan ju ha sparade pengar. Därför bör inte kvotskattning användas som ju förutsätter att sambandet går genom origo.

c) Även här kan regressionstekniken användas. Låt hela urvalet utgöra ”tillfällig” population och den del av urvalet som besvarat frågan om spenderade pengar ”tillfälligt” urval.
Korrelationskoefficienten 0.7 kan bara ha beräknats på basis av de observationer där bägge variablerna är mätta och kan därför fortfarande användas.

⇨ $\bar{x}_U = 27300$, $\bar{y} = 7300$, $s_y = 2600$, $\bar{x} = 26900$, $s_x = 6100$, $r = 0.7$

$$\Rightarrow \quad \hat{B}_1 = \frac{r \cdot s_y}{s_x} = \frac{0.7 \cdot 2600}{6100} \approx 0.2984 \Rightarrow \hat{\bar{y}}_{U,\,just} = 7300 + 0.2984 \cdot (27300 - 26900) = 7419$$