

Opgave IV.5

Interessevariabel

$y_i = 1$ hvis patienten indenfor året har klaget over ondt i maven, $= 0$ ellers.

$\bar{y}_{kl-R} = \text{antal med ondt i maven} / \text{antal patienter} = 0.047$

et-trins klynge : 93 læger vælges ud af 3500 praktiserende læger. For de 93 udvalgte læger observeres for alle patienter om de har henvendt sig med ondt i maven eller ej.

Mangler oplysninger om $N_k = \text{antal patienter pr. læge}$. Derfor antages at alle N_k er ens og dermed lig med $5100000/3500 = 1500$ (ca.), idet N sættes til 5100000. Desuden kendes variationen i andelen med ondt i maven ikke mellem de forskellige læger (nogle lægers patientgrupper kunne have større frekvens af mavesmerter end andre). Derfor antages også at disse andele er ens, $P = 0.047$ som beregnet ovenfor.

Dermed bliver variansen på antal patienter med ondt i maven pr. praksis

$$\text{var}(Y_k) = N_k P(1 - P) = 67.2$$

Da klyngestørrelserne er ens er $kl-1 = kl-R$. På grund af usikkerheden om de opgivne tal beregnes dog

$$\bar{y}_{kl-1} = \frac{3500}{5100000} \times \frac{1}{93} \times 5830 = 0.043 .$$

Altså en lidt anden værdi end 0.047.

$$\text{Var}(\bar{y}_{kl-1}) = (3500/5100000)^2 (3500 - 93)/(3500 \times 93) 67.2 = 0.006^2 .$$

Simpel tilfældig udvælgelse af $n = 123610$ giver

$$\frac{5100000 - 123610}{(5100000 - 1) 123610} 0.047(1 - 0.047) = 0.006^2 .$$

Der er altså ingen væsentlig forskel på præcisionen ved simpel tilfældig udvælgelse og den anvendte klyngeudvælgelse, da forholdene antages helt ens i alle klynger.

Opgave V.2

$N = 2300000$
 $n = 1000$

1. Approksimation betegner bidraget

$$\frac{N_h - 1}{N - 1} S_h^2$$

2. Approksimation betegner bidraget

$$\frac{N_h - 1}{N - 1} S_h^2 \left(1 + \frac{N - n}{nN} \frac{N}{N - 1} \frac{N - N_h}{N_h} \right)$$

Disse bidrag er angivet i tabellen.

S_h	1. Approks	2. Approks
0.47	0.0751	0.0753
0.56	0.1035	0.1037
0.65	0.0634	0.0637
0.64	0.0532	0.0536
0.62	0.0192	0.0196
Sum	0.3144	0.3159

Dermed er variansen ved den groveste approksimation

$$\frac{N - n}{Nn} \sum_{h=1}^6 \frac{N_h - 1}{N - 1} S_h^2 = \frac{2300000 - 1000}{230000 \times 1000} 0.3144 = 0.0177^2$$

Variansen ved den forbedrede approksimation er

$$\frac{N-n}{Nn} \sum_{h=1}^6 \frac{N_h-1}{N-1} S_h^2 \left(1 + \frac{N-n}{nN} \frac{N}{N-1} \frac{N-N_h}{N_h} \right) = \frac{2300000-1000}{230000 \times 1000} 0.3159 = 0.0178^2$$

I følge formelen for variansanalyseopspaltningen er $S^2 = 0.3144 + 0.0710 = 0.3854$, idet variansen mellem strata let kan beregnes.

Dermed er variansen ved simpel tilfældig udvælgelse uden vægtning givet ved

$$\frac{N-n}{nN} S^2 = 0.0196$$