

Stikprøveusikkerhed

Usikkerhedens størrelse afhænger af selve procenttallet og stikprøvens størrelse. Hvis universet er relativt lille, som det typisk forekommer i business-to-business undersøgelser, afhænger usikkerheden også af universets størrelse.

Sikkerhedsgrænser for store universer

Usikkerheden kan angives som 95% sikkerhedsgrænse. Det procenttal, som man havde fået, hvis man havde spurgt hele universet, kaldes for *det sande procenttal*. Afstanden mellem det sande procenttal og det observerede procenttal er med 95% sikkerhed mindre end sikkerhedsgrænsen. Med andre ord: Det sande procenttal ligger med 95% sikkerhed i intervallet mellem det observerede procenttal minus sikkerhedsgrænsen og det observerede procenttal plus sikkerhedsgrænsen. For store universer (f. eks. hele Danmarks befolkning på 15 år eller derover) er sikkerhedsgrænserne for forskellige procenter og forskellige stikprøvestørrelser vist i nedenstående tabel. Tabellen er beregnet ud fra følgende formel:

$$1,96 \sqrt{\frac{\text{procent}(100 - \text{procent})}{\text{stikprøvestørrelse}}}$$

Stikprøve- størrelse	5% eller 95%	10% eller 90%	15% eller 85%	20% eller 80%	25% eller 75%	30% eller 70%	35% eller 65%	40% eller 60%	45% eller 55%	50%
50	6,0	8,3	9,9	11,1	12,0	12,7	13,2	13,6	13,8	13,9
75	4,9	6,8	8,1	9,1	9,8	10,4	10,8	11,1	11,3	11,3
100	4,3	5,9	7,0	7,8	8,5	8,9	9,3	9,6	9,8	9,8
150	3,5	4,8	5,7	6,4	6,9	7,3	7,6	7,8	8,0	8,0
200	3,0	4,2	4,9	5,5	6,0	6,4	6,6	6,8	6,9	6,9
250	2,7	3,7	4,4	5,0	5,4	5,7	5,9	6,1	6,2	6,2
300	2,5	3,4	4,0	4,5	4,9	5,2	5,4	5,5	5,6	5,7
400	2,1	2,9	3,5	3,9	4,2	4,5	4,7	4,8	4,9	4,9
500	1,9	2,6	3,1	3,5	3,8	4,0	4,2	4,3	4,4	4,4
600	1,7	2,4	2,9	3,2	3,5	3,7	3,8	3,9	4,0	4,0
700	1,6	2,2	2,7	3,0	3,2	3,4	3,5	3,6	3,7	3,7
800	1,5	2,1	2,5	2,8	3,0	3,2	3,3	3,4	3,5	3,5
900	1,4	2,0	2,3	2,6	2,8	3,0	3,1	3,2	3,3	3,3
1.000	1,4	1,9	2,2	2,5	2,8	3,0	3,0	3,0	3,1	3,1
1.200	1,2	1,7	2,0	2,3	2,5	2,6	2,7	2,8	2,8	2,8

Eksempel:

Fra et meget stort antal personer er 200 svarpersoner udvalgt simpelt tilfældigt. Disse 200 personer interviewes, f.eks. med hensyn til, om de er forbrugere af en bestemt vare. 80 af dem svarer bekræftende, dvs. 40%. Det fremgår af tabellen, at sikkerhedsgrænsen ligger ved 6,8%. Det betyder, at det er næsten sikkert (95% sandsynlighed), at området mellem de to grænser 33,2% (40%-6,8%) og 46,8% (40%+6,8%) vil indeholde procenten af bekræftende, såfremt man havde spurgt hele Danmarks befolkning på 15 år og derover.

Sikkerhedsgrænser for mindre universer

Hvis universet er mindre og stikprøven udgør en betydelig del af universet, skal der tages hensyn til universets størrelse. Man beregner derfor sikkerhedsgrænser for procenttallene med samme teknik som ovenfor, men bruger følgende formel i stedet for:

$$1,96 \sqrt{\frac{\text{procent}(100 - \text{procent})}{\text{stikprøvestørrelse} - 1} \cdot \frac{\text{universets størrelse} - \text{stikprøvestørrelse}}{\text{universets størrelse} - 1}}$$

Eksempel:

Fra 1000 personer er 200 svarpersoner udvalgt simpelt tilfældigt. For eksempel er de 1000 personer ansat i et firma, og de 200 af dem bliver interviewet med hensyn til medarbejdertilfredshed. På et af spørgsmålene svarer 80 af dem bekræftende, dvs. 40%. Sikkerhedsgrænsen er derfor

$$1,96 \sqrt{\frac{40(100 - 40)}{200 - 1} \cdot \frac{1000 - 200}{1000 - 1}} = 6,1$$

Det betyder, at det er næsten sikkert (95% sandsynlighed), at området mellem de to grænser 33,9% (40%-6,1%) og 46,1% (40%+6,1%) vil indeholde procenten af bekræftende, såfremt man havde spurgt alle 1000 medarbejdere.

Sammenligning mellem to stikprøver i store målgrupper

I andre sammenhænge har man behov for at vurdere, om to procenttal fra forskellige stikprøver med sikkerhed kan siges at være forskellige, eller om en observeret forskel blot beror på tilfældigheder. Lad p_1 og p_2 betegne de to procenttal, og lad n_1 og n_2 betegne de to stikprøvestørrelser. Hvis de to målgrupper, som stikprøverne er taget fra, er store, er de to procenttal med 95% sikkerhed forskellige, hvis forskellen mellem dem overskrider tallet:

$$1,96 \sqrt{\frac{p_1(100 - p_1)}{n_1} + \frac{p_2(100 - p_2)}{n_2}}$$

Eksempel:

I en busundersøgelse har man fundet, at der i hovedstadsområdet er 20% enlige, mens andelen kun er 16% i det øvrige land. Er forskellen mellem de to tal signifikant? Der er blevet interviewet 250 personer i hovedstadsområdet og 750 i det øvrige land. Teststørrelsen er

$$1,96\sqrt{\frac{20(100-20)}{250} + \frac{16(100-16)}{750}} = 5,6$$

Siden forskellen mellem de to procenttal (20%-16%=4%) er mindre end dette tal, kan de to procenttal ikke med sikkerhed siges at være forskellige.

Sammenligning mellem to stikprøver i mindre målgrupper

Hvis målgrupperne er mindre, og stikprøverne udgør en betydelig del af målgrupperne, skal der tages hensyn til målgruppestørrelsen. Lad p_1 og p_2 betegne de to procenttal, lad n_1 og n_2 betegne de to stikprøvestørrelser og lad N_1 og N_2 betegne de to målgruppestørrelser. De to procenttal er med 95% sikkerhed forskellige, hvis forskellen mellem dem overskrider tallet:

$$1,96\sqrt{\frac{p_1(100-p_1)}{n_1-1} \cdot \frac{N_1-n_1}{N_1-1} + \frac{p_2(100-p_2)}{n_2-1} \cdot \frac{N_2-n_2}{N_2-1}}$$

Eksempel:

Af de 1000 ansatte i et firma er 600 ansat i produktionen og 400 i salg og administration. Der udvælges tilfældigt 100 svarpersoner i hver gruppe. I produktionen svarer 40 af dem bekræftende på et spørgsmål, dvs. 40%, og i salg og administration 50, dvs. 50%. Teststørrelsen er:

$$1,96\sqrt{\frac{40(100-40)}{100-1} \cdot \frac{600-100}{600-1} + \frac{50(100-50)}{100-1} \cdot \frac{400-100}{400-1}} = 12,3$$

Siden forskellen mellem de to procenttal (50%-40%=10%) er mindre end dette tal, kan de to procenttal ikke med sikkerhed siges at være forskellige.