



Av **Ellen M. Kulset**, førsteamanuensis, Norges Handelshøyskole (NHH) og Høgskolen i Sørøst-Norge

STIKKPRØVER I REVISJON: TEST AV KONTROLLER¹

Stikkprøver i revisjon defineres gjerne som «Anvendelsen av revisjonshandlinger på mindre enn 100 prosent av enhetene i en populasjon som er relevant for revisjonen, under forutsetning av at alle enheter har en mulighet til å bli valgt ut, for å skaffe revisor et betryggende grunnlag for å kunne trekke konklusjoner om hele populasjonen» (ISA 530, pkt. 5a).

Et eksempel på en slik stikkprøve vil være når vi sender ut saldoforespørslers til et utvalg av de kundene som skylder penger til vår revisjonskunde per 31.12 og bruker bevisene vi henter inn gjennom denne revisjonshandlingen til å konkludere ikke bare på om de kundefordringene vi har kontrollert eksisterer, men også på om de vi ikke har kontrollert gjort det. Stikkprøver kan anvendes ved bevisinnhenting knyttet til ulike revisjonspåstander og benyttes både i forbindelse med kontroll- og detaljtester.

Stikkprøver regnes som et kraftfullt revisjonsverktøy ettersom man kan skaffe bevis for revisjonspåstander knyttet til store populasjoner ved å kontrollere et relativt sett lite antall av enhetene i populasjonen. Jeg vil i denne artikkelen ha fokus på bruk av statistikk i forbindelse med revisors kontrolltesting og vil ved hjelp av et enkelt eksempel vise hvordan man kan gjennomføre slik testing i praksis. Jeg vil også bruke statistiske tabeller utarbeidet til bruk i revisjonssammenheng for å evaluere enkelte tommelfingerregler som ofte benyttes av revisor når han gjennomfører tester av et selskaps internkontroll. Artikkelen format tilsier at fremstillingen vil måtte bli relativt overfladisk, og jeg viser derfor til boken «Auditing & Assurance Services», kapittel 8, for en grundigere fremstilling av temaet.

Valg av revisjonshandling: stikkprøve eller ikke?

Moderne informasjonsteknologi har ført til at vi ikke lenger har det samme behovet for stikkprøver som tidligere: Regnskapssystemer er i økende grad automatiserte, og vi kan da istedenfor å bruke stikkprøver, teste de generelle kontrollene i systemet (inkludert de programendringene som har funnet sted) og deretter supplere med å teste et mindre antall transaksjoner enn vi ville ha gjort i en tradisjonell stikkprøve. Det er for revisor også i stadig større grad mulig å laste ned store, komplette kundedata-sett istedenfor å bygge på utvalg. Dette betyr at revisor i større grad enn tidligere har mulighet til å teste alle enhetene i en populasjon, eventuelt teste alle de enhetene som har en eller flere egenskaper som antas å innebære større risiko for feil.

Disse endringene betyr imidlertid ikke at det ikke lenger er behov for stikkprøver: mange av de kontrollene et selskap utfører er fremdeles manuelle (f. eks. manuelle avstemminger), det er en rekke revisjonstester som krever at revisor fysisk undersøker eiendeler (f. eks. revisjon av eksistens av fysiske eiendeler), revisor må ofte vurdere revisjonsbevis som han har innhentet fra ulike tredjeparter (f. eks. bruk av saldoforespørslers ol.), og det kan f. eks. tenkes at revisor ønsker å teste enkelte av de generelle kontrollene hos en kunde ved hjelp av stikkprøver.

¹ Denne artikkelen er basert på foredraget «Stikkprøver i revisjon - fokus på kontrolltesting» som forfatteren avholdt på Midt-Norge kommunerevisorforening sitt nyttårsmøte 26/01/17. En del endringer er innarbeidet slik at fremstillingen passer bedre til artikkelformatet. Artikkelen bygger ellers særlig på kapittel 8 i boken «Auditing & Assurance Services» av Eilifsen et al. 2014. Lesere som ønsker ytterligere innsikt i problemstillingene som behandles i artikkelen vises til Eilifsen et al., ev. til spesiallitteraturen, f. eks. Lillestøls bok «Statistiske metoder i revisjon».

«Vi skiller gjerne mellom stikkprøver som har en statistisk angrepsvinkel og stikkprøver med en skjønsmessig angrepsvinkel.»

Vi skiller gjerne mellom stikkprøver som har en statistisk angrepsvinkel og stikkprøver med en skjønsmessig angrepsvinkel. Hovedforskjellen mellom de to tilnæringsmåtene er at en revisor som benytter en statistisk angrepsvinkel bruker sannsynlighetsteori når han bestemmer stikkprøvestørrelse, velger hvilke enheter som skal med i stikkprøven og vurderer funnene fra stikkprøven. Dette betyr at revisor oppnår en hensiktsmessig stikkprøvestørrelse med tilhørende mulighet til å begrunne hvorfor akkurat denne størrelsen er valgt og ikke minst at utvalgsrisikoen² blir kvantifiserbar, dvs. revisor kan benytte resultatet fra testen til å konstruere et konfidensintervall³ rundt estimatet på avvik/feil i populasjonen der det faktiske avviket/feilen med en gitt sannsynlighet befinner seg innenfor dette intervallets grenser. ISA 530 «Stikkprøver i revisjon» tillater at revisor benytter både statistisk og ikke-statistisk baserte stikkprøver. Anekdotiske bevis tilsier at vi i Norge i den senere tid har sett en utvikling både mot noe mer bruk av statistiske metoder samt utstrakt bruk av enkle tommelfingerregler i forbindelse med kontrolltesting der stikkprøvestørrelsen er basert på underliggende statistisk teori.

Attributtstikkprøve – test av kontroller

Når revisor skal gjennomføre en test av en av et selskaps interne kontroller for å finne ut om denne fungerer slik at han kan vurdere å bygge på den i sitt revisjonsarbeid, tester han et utvalg ganger kontrollen ble utført og bruker resultatet fra denne testen til å lage et estimat på hvor mange ganger

kontrollen ikke ble utført som planlagt/foreskrevet i perioden han tester. Revisor ønsker altså kunnskap om andelen revisjonseenheter i en populasjon som har en bestemt egenskap (kontrollen ble utført som foreskrevet eller ikke) og trekker en stikkprøve bestående av n enheter for å estimere hvor mange ganger kontrollen ikke fungerte i hele populasjonen. Dersom vi trekker denne stikkprøven ved hjelp av tilfeldig trekking, vet vi fra sannsynlighetsteorien at antall avvik i stikkprøven (antall ganger kontrollen ikke fungerte) er hypergeometrisk fordelt. Vi kan bruke denne kunnskapen til å avgjøre hvor stor stikkprøve som er ønskelig og for å lage et estimat på en øvre konfidensgrense, dvs. et estimat på hvor stort populasjonsavviket kan være med et ønsket konfidensnivå c , f. eks. dersom vi har observert null avvik i stikkprøven.

Jeg skal nå ved hjelp av et eksempel vise hvordan vi kan gjennomføre en statistisk basert stikkprøve i forbindelse med test av en kontroll. La oss ta utgangspunkt i følgende eksempel:

Du er ansatt som revisor i revisjonsselskapet Brødrene Hansen. I forbindelse med revisjonen av NBB AS er du bedt om å teste om du kan stole på en av selskapets interne kontroller. Testen går ut på å fastslå om det for selskapets salgsfakturaer eksisterer autoriserte ordre og pakksedler. Selskapet har totalt sendt ut 10.000 fakturaer. Et avvik foreligger når det ikke kan fremvises en autorisert ordre og/eller en pakkseddel for fakturaen som kontrolleres. For å finne ut om det er grunn til å tro at mer enn 5 prosent av de 10.000 fakturaene mangler dokumentasjon, blir du bedt om å teste om det for et *tilfeldig utvalg fakturaer* eksisterer tilfredsstillende dokumentasjon. Det er ønskelig at risikoen for feilaktig å konkludere med at kontrollen fungerer når den ikke gjør det er relativt lav (10 prosent) ettersom det er planlagt å stole endel på denne kontrollen (det skal ikke hentes inn så mange andre bevis knyttet til påstanden). Basert på erfaring fra tidligere, har du ikke grunn til å forvente at du finner avvik i utvalget.⁴

⁴ Merk at dette er en type kontroll som kan være automatisert, og i så tilfelle vil det kunne være tilstrekkelig med en mye mindre omfattende kontroll enn den som blir foreslått i eksempelet.

² Utvalgsrisiko defineres i ISA 530, pkt. 5c som «Risikoen for at den konklusjonen revisor trekker på grunnlag av et utvalg, kan være forskjellig fra den konklusjonen revisor ville ha trukket dersom hele populasjonen var gjenstand for den samme revisjonshandlingen».

³ Begrepet konfidensnivå kan vi definere som komplementet til utvalgsrisikoen, dvs. sannsynligheten for at den konklusjonen revisor trekker på grunnlag av et utvalg, er den samme som den konklusjonen revisor ville ha trukket, dersom hele populasjonen var gjenstand for den samme revisjonshandlingen. Fra statistikkfaget har vi at «Et konfidensintervall for en ukjent parameter θ er et intervall med grenser som med en gitt sannsynlighet c kalt konfidensnivået, omslutter θ » (Lillestøl 1991, p. 204). Disse grensene kaller vi for øvre og nedre konfidensgrense.

Forventet avvikshyppighet i populasjonen (i prosent)	Tolererbar avvikshyppighet (i prosent)									
	3	4	5	6	7	8	9	10	15	20
0,00	76	57	45	38	32	28	25	22	15	11
0,25	129	96	77	64	55	48	42	38	25	18
0,50	129	96	77	64	55	48	42	38	25	18
0,75	129	96	77	64	55	48	42	38	25	18
1,00	176	96	77	64	55	48	42	38	25	18
1,25	221	132	77	64	55	48	42	38	25	18

Tabell 1 – Stikkprøvestørrelse som en funksjon av tolererbar og forventet avvikshyppighet ved et ønsket konfidensnivå på 90 prosent (hentet fra Eilifsen et al. (2014) tabell 8-6)

Hvor stort utvalg vil du kontrollere? Hvordan vil du velge enhetene i utvalget? Hvordan kan du ta hensyn til utvalgsrisiko?

For å bestemme passende stikkprøvestørrelse, er det utviklet tabeller der vi finner stikkprøvestørrelse som en funksjon av hvor mange avvik vi forventer i en populasjon, hvor mange avvik vi kan tolerere i denne populasjonen uten at dette avviket vil få oss til å konkludere med at internkontrollen ikke fungerer (og dermed ikke kan bygges på) og hvilket konfidensnivå⁵ vi ønsker. Dersom vi, som i eksempelet på forrige side, forventer ingen avvik, tolererer en avvikshyppighet på 5 prosent og ønsker et konfidensnivå på 90 prosent, finner vi av tabell 1 over (hentet fra tabell 8-6 i boken «Auditing & Assurance Services») en stikkprøvestørrelse på 45. I denne tabellen finner vi passende stikkprøvestørrelse uavhengig av populasjonsstørrelse, og tabellen egner seg så sant populasjonen består av flere enn ca. 500 enheter.

Dersom populasjonen består av mindre enn 500 enheter, kan vi benytte en korreksjonsformel⁶ for å nedjustere størrelsen på stikkprøven noe, men hva gjør vi dersom vi ønsker å gjennomføre en stikkprøve og populasjonen er svært liten, f. eks. hvis kontrollen vi tester gjennomføres på ukentlig basis? AICPA har utviklet en tabell, se tabell 2 over (AICPA 2014, side 43), som kan benyttes når kontrollfrekvensen er lav. I denne tabellen foreslås det at man ved ukentlige kontroller benytter en stikkprøvestørrelse på fra 5-9.

Hvis vi tar utgangspunkt i tabell 1 over, ser vi at en så lav stikkprøvestørrelse som 9 ikke finnes i

Kontrollfrekvens og populasjonsstørrelse	Utvalgsstørrelse
Kvartalsvis (4)	2
Månedlig (12)	2-4
Bimånedlig (24)	3-8
Ukentlig (52)	5-9

Tabell 2 - Stikkprøvestørrelser ved små populasjoner

tabellen. En stikkprøvestørrelse på 9 innebærer dermed implisitt f. eks. at vi ved et ønsket konfidensnivå på 90 prosent, har lagt til grunn at det ikke er forventet noen avvik i populasjonen, samtidig som vi vil akseptere kontrollen som velfungerende så lenge den ikke utføres feil i flere enn 10 av årets uker (tolererbart avvik 20 prosent). Dersom vi ikke tolererer flere enn f. eks. 4 avvik i populasjonen (tilnærmet 8 prosent avvik) før vi anser dette som en ikke fungerende kontroll, må vi velge en stikkprøvestørrelse på 19⁷ (beregnet ved bruk av korreksjonsfaktor), dvs. en dobbelt så omfattende test som den AICPA foreslår i den øvre delen av kontrollintervallet i tabell 2. Alternativt kan vi for å komme frem til en stikkprøvestørrelse på 9 ved ukentlige kontroller f. eks. implisitt legge til grunn et betydelig lavere konfidensnivå⁸ (f. eks. 60 prosent) kombinert med et noe lavere tolererbart avvik (f. eks. 11 prosent).

Vi ser videre av tabell 2 over at det foreslås at man kontrollerer 2 av kontrollene ved kvartalsvise kontroller. Ifølge Jacoby og Hitzig (2011) krever imidlertid slike kontroller en stikkprøvestørrelse på 4 (dvs. kontroll av hele populasjonen) selv hvis man bare ønsker et lavt konfidensnivå. Dette synspunktet kan underbygges på følgende måte: anta at kontrollen ikke fungerte i et av kvartalene og at revisor har bestemt seg for å sjekke to av de fire kvartals-

5 Eilifsen et al. (2014) sier følgende om valg av forutsetninger på sidene 268-271: Valg av konfidensnivå bør foretas basert på faktorer som en kontos viktighet, hvor viktig påstanden kontrolltesten skal gi bevis for er, hvor vanskelig det er å utføre kontrollen samt hvor mye man planlegger å stole på kontrollen. Tolererbart avvik settes gjerne basert på hvor viktig kontrollen man tester er slik at hvis kontrollen anses viktig settes avviket til 3-5 prosent mens man ellers setter det i intervallet 6-10 prosent. Når det gjelder forventet avvikshyppighet tar man gjerne utgangspunkt i hva man har avdekket av avvik i forbindelse med tidligere revisjoner.

6 Vi kan benytte følgende korreksjonsformel: Stikkprøvestørrelse etter korreksjon = $n * \sqrt{1 - \frac{28}{52}}$ der n er stikkprøvestørrelsen før korreksjon og N er populasjonsstørrelse.

7 Stikkprøvestørrelse etter korreksjon = $28 * \sqrt{1 - \frac{28}{52}} = 19$

8 For å finne passende stikkprøvestørrelse dersom man ønsker et lavere konfidensnivå enn 90 prosent, kan man f. eks. benytte ulike former for revisjonsprogramvare. Stikkprøvestørrelsen på 9 ved et konfidensnivå på 60 prosent og et tolererbart avvik på 11 prosent er beregnet ved hjelp av revisjonsprogramvaren ACL.

«Dersom vi ønsker å gjennomføre en kontrolltest i samsvar med statistisk teori, er vi imidlertid nødt til å trekke et *tilfeldig utvalg*.»

vise kontrollene. Det vil nå være seks ulike stikkprøver som kan trekkes og kontrollen som ikke ble utført korrekt vil kun komme med i tre av disse. Avviket vil altså bare avdekkes med en sannsynlighet på 50 prosent. Jacoby og Hitzig viser videre i sin artikkel at når revisor bruker stikkprøvestørrelsene i tabell 2 foran, vil sannsynligheten for å avdekke et eksisterende avvik være lavere ved de av kontrollene som utføres sjelden enn ved de som utføres f. eks. ukentlig, se ellers Jacoby and Hitzig (2011) for en grundigere gjennomgang av problemstillingene knyttet til små utvalg.

I sum har vi at stikkprøvestørrelsene som fremkommer i tabell 2 gir relativt lite bevis isolert sett. Bruken av disse tommelfingerreglene forsvares gjerne med at ved sjeldent forekommende kontroller, vil effekten av andre beviskilder være større enn for kontroller som gjennomføres oftere. Det er imidlertid viktig at revisor er klar over denne problemstillingen. Det bør også påpekes at revisor ved slike sjeldent forekommende kontroller gjerne velger å ha fokus på de kontrollene som er utført per 31.12. Gitt at kontrollene er kumulative, vil man kunne få relativt høy bevisverdi på den måten, og det er ikke nødvendigvis behov for å teste kontrollene som er utført i løpet av året.

Det fremgår av ISA 530, Vedlegg 4, at det er flere ulike metoder som kan brukes for å velge ut hvilke enheter som skal kontrolleres (utvalgseenhetene). Dersom vi ønsker å gjennomføre en kontrolltest i samsvar med statistisk teori, er vi imidlertid nødt til å trekke et *tilfeldig utvalg*. Per definisjon betyr dette at alle utvalg av den ønskede størrelse er like sannsynlige, slik at «alle utvalgseenheter i populasjonen skal ha den samme sannsynligheten for å komme med i stikkprøven, men også at enhver kombinasjon med et gitt antall utvalgseenheter har samme sjanse for å bli trukket ut» (Lillestøl, 1996, side 20). I praksis kan vi for å trekke ut enhetene generere tilfeldige tall ved hjelp av Excel eller en app. Disse tallene vil dersom de genereres ved hjelp av en datamaskinbasert algoritme ikke være tilfeldige i teoretisk forstand, men for vårt formål er dette «tilfeldig nok». Et alternativ er å benytte tilfeldige tall generert fra nettsiden random.org som skal være generert ved hjelp av en fysisk tilfeldig prosess, dvs. de er tilfeldige også i teoretisk forstand.

Utvalgsstørrelse	Faktisk antall avdekkede avvik			
	0	1	2	3
50	4,6	7,6	10,3	12,9
55	4,1	6,9	9,4	11,8
60	3,8	6,4	8,7	10,8
70	3,3	5,5	7,5	9,3
80	2,9	4,8	6,6	8,2
90	2,6	4,3	5,9	7,3
100	2,3	3,9	5,3	6,6
120	2,0	3,3	4,4	5,5
160	1,5	2,5	3,3	4,2
200	1,2	2,0	2,7	3,4

Tabell 3 – Øvre konfidensgrense ved et ønsket konfidensnivå på 90 prosent (hentet fra Eilifsen et al. (2014) tabell 8-9)

La oss nå av pedagogiske hensyn anta at vi valgte å kontrollere et utvalg på 50 enheter og at det for alle de kontrollerte salgsfakturaene eksisterer autoriserte ordrer og pakksedler. Det beste estimatet vi nå har på avviksprosent i populasjonen er null, og vi kan lese av fra en annen tabell, tabell 3 over, at øvre konfidensgrense er på 4,6 prosent. Rent konkret betyr dette at basert på hva vi har funnet i utvalget vårt, er sjansen 90 prosent for at den faktiske avviksandelen i populasjonen er lavere enn 4,6 prosent.

Vi sammenligner nå denne øvre konfidensgrensen med tolererbart avvik og finner at vi kan konkludere med at vi *ikke* har grunn til å tro at flere enn 5 prosent av salgene mangler dokumentasjon: sagt på en annen måte, denne internkontrollen kan vi stole på. Hvis vi hadde avdekket ett avvik i utvalget, er vårt beste estimat på avvik i populasjonen 2 prosent (1/50) og vi kan lese av i tabell 3 foran at den øvre konfidensgrensen er på 7,6 prosent. Ettersom denne grensen er høyere enn det tolererbare avviket betyr dette at vi nå ikke har tilstrekkelig sikkerhet for at det ikke er færre enn 5 prosent avvik i populasjonen, og vi må i denne situasjonen konkludere med at internkontrollen ikke fungerer så godt at vi kan stole på den i revisjonsarbeidet vårt. I praksis har vi i så fall to ulike muligheter: dersom vi tror at vi har vært «uheldige» med stikkprøven og bestemt tror at det ikke er

«Vi er gjerne spesielt bekymret for førstnevnte type feil ettersom vi da risikerer at vi ender opp med gal revisjonsberetning mens vi i sistnevnte tilfelle kan ende opp med å revidere for mye med for høye kostnader som konsekvens.»

flere avvik i populasjonen enn at vi kan stole på internkontrollen, kan vi velge å utvide stikkprøven. Alternativet er å gjennomføre «mer» substansrevisjon enn det vi initielt hadde planlagt ved å endre «type, tidspunkt og/eller omfang» av substansrevisjonshandlingene knyttet til den/de påstandene internkontrolltesten var ment å skulle hente inn bevis for.

Ettersom vi har gjennomført en utvalgsbasert test står vi overfor to ulike typer risiko for feilkonklusjon. I det førstnevnte eksempelet risikerer vi at vi feilaktig har akseptert at internkontrollen fungerer når den ikke gjør det (type II-feil), og i det sistnevnte eksempelet risikerer vi at vi feilaktig har forkastet at internkontrollen fungerer (type I-feil) når den faktisk egentlig fungerer. Vi er gjerne spesielt bekymret for førstnevnte type feil ettersom vi da risikerer at vi ender opp med gal revisjonsberetning mens vi i sistnevnte tilfelle kan ende opp med å revidere for mye med for høye kostnader som konsekvens.

Når det gjelder bruk av stikkprøver i forbindelse med offentlig revisjon, følger det av ISSAI 1530, pkt. 3, at revisor her kan ha ytterligere plikter når det gjelder både hvilket arbeid som skal utføres og hvilke plikter man har med tanke på rapportering av eventuelle funn. I praksis kan det for eksempel tenkes at man kan bli pålagt å gjennomføre en grundigere revisjon enn «vanlig», dvs. at man må bruke et høyere konfidensnivå i forbindelse med stikkprøvene man gjennomfører, eller at man må rapportere både en nedre og en øvre konfidensgrense, f. eks. som en del av pålagte forvaltningsrevisjonsoppgaver, jf. kommuneloven § 78, pkt. 2. Et eksempel på det siste kan være at man blir bedt om å kartlegge responstid knyttet til en kommunal tjeneste, og at man i den forbindelse rapporterer et beste estimat på responstiden samt hvilket konfidensintervall som med en sannsynlighet på f. eks. 95 prosent (avhengig av valgt konfidensnivå) inneholder den faktiske responstiden slik at dette intervallet kan brukes for å vurdere om responstiden er i samsvar med den kommunen er pålagt å sørge for.

Avslutning

Jeg har i denne artikkelen med et enkelt eksempel vist hvordan man kan gjennomføre en stikkprøve

med en statistisk angrepsvinkel samt kort diskutert noen av «tommelfingerreglene» som per i dag benyttes av revisor når han tester en kundes internkontroll. Vi har sett at revisor bør se på bruken av disse tommelfingerreglene med et kritisk blikk ettersom de gjerne gir relativt lite bevis isolert sett. Avslutningsvis vil jeg benytte sjansen til å understreke at kjennskap til statistiske metoder gir revisor et verktøy som han kan benytte i sitt daglige revisjonsarbeid («utvider verktøykassen»), og at dette bidrar til at revisor bedre kan reflektere rundt egen praksis. ☺

Referanser

American Institute of Certified Public Accountants. 2014. *Audit Sampling*. New York: AICPA

ISA 530 Stikkprøver i revisjon

ISSAI 1530 Audit Sampling

Eilifsen, Aa, W. F. Messier, S. M. Glover and D. F. Prawitt. 2014. *Auditing and assurance services*. Berkshire: Mc Graw Hill Education.

Jacoby, J. E. and N. B. Hitzig. 2011. Auditing Internal Controls in Small Populations. *The CPA Journal*. December: 34-36.

Lillestøl, J. 1991. *Sannsynlighetsregning og statistikk med anvendelser*. 4. utgave. Oslo: Bedriftsøkonomens Forlag.

Lillestøl, J. 1995. *Statistiske metoder i revisjon*. Oslo: Cappelen.

Ellen M. Kulset er siviløkonom NHH, statsautorisert revisor og har en PhD i Business Economics fra NHH. Hun har tidligere arbeidet som manager i KPMG og arbeider nå som førsteamanuensis ved NHH og Høgskolen i Sørøst-Norge.