

Test av internkontroll ved hjelp av en stikkprøve:

Kan revisor stole på kundens interne kontroller?



Førsteamanuensis
Ellen M. Kulset
Universitetet i Sørøst Norge og
Norges Handelshøyskole



Professor Emeritus
Jostein Lillestøl
Norges Handelshøyskole

Artikkelen fokuserer på et sentralt beslutningsproblem: kan revisor stole på kundens interne kontroller eller ikke? Vi presenterer noe av logikken og teorien bak stikkprøvekontroller og prøver samtidig å besvare enkelte spørsmål som ofte stilles i forbindelse med dette.

ISA 530 Stikkprøver i revisjon

Bruk av stikkprøver reguleres av ISA 530 *Stikkprøver i revisjon*, og det åpnes for bruk av både statistiske- og skjønnsbaserte stikkprøver. Det legges imidlertid til grunn at for eksempel stikkprøvestørrelser ikke skal avhenge av om revisor rent formelt benytter en statistisk angrepsvinkel eller ikke, se ISA 530 pkt. A 9 og 11.

Tematilsyn i 2016

Finanstilsynet gjennomførte i 2016 et tematilsyn hos 65 norske revisjonsselskaper (inkludert de fem store) der de undersøkte selskapers bruk av stikk-

prøver (Finanstilsynet 2016). Kulset (2019) analyserer Finanstilsynets funn og finner blant annet at noen av stikkprøvestørrelsene som er i bruk i forbindelse med test av kontroller, ikke synes å bygge på statistisk teori. Basert på dette funnet er denne artikkelens første siktemål å presentere noe av den teoretiske støtten norske revisorer tidvis kan ha bruk for når de anvender stikkprøver for å teste kundens internkontroll.

Pensumlitteraturen som benyttes i revisjonsfaget i dag knyttet til stikkprøver, har en praktisk tilnærming til emnet: studentene skal lære ulike metoder som de skal kunne bruke i praksis, men det blir i liten grad lagt til rette for at studentene skal kunne forstå hvordan disse metodene bygger på det de tidligere har lært i statistikkfaget.¹ Denne artikkelens andre siktemål er å svare på noen av de spørsmålene revisjonsstudenter ofte stiller ved å bruke den kunnskapen studentene tidligere skal ha tilegnet seg, først og fremst i metodefaget statistikk. Artikkelen henvender seg altså både til revisjonsstudenter og praktiserende revisorer. Artikkelen vil kun fokusere på enkelte problemstillinger knyttet til attributtsampling (behandlet f.eks. i Eilifsen et al., 2014 i kapittel 8) og

fremstillingen er illustrert ved bruk av Excel.

Et sentralt beslutningsproblem – estimere avvikshyppighet

Denne artikkelen fokuserer på et sentralt beslutningsproblem som revisor ofte står overfor: skal han/hun stole på en kundes interne kontroller eller ikke? I denne situasjonen ønsker revisor gjerne å bruke en stikkprøve for å lage et estimat med tilhørende konfidensgrenser på andelen ganger som en internkontroll ikke har fungert i en gitt periode (avvikshyppighet).

Revisjonshandlingen

Fra et teoretisk ståsted kan revisjonshandlingen beskrives som følger: Vi har en populasjon bestående av N elementer (her hver gang en kontroll skulle vært utført i en gitt periode). I denne populasjonen er M elementer spesielle (her at kontrollen ikke er utført som den skal, dvs. vi har et kontrollavvik). Revisor ønsker å akseptere kundens kontroll dersom andelen kontrollavvik $a = M/N$ i populasjonen er liten (dvs. under/på et akseptabelt nivå). Her er a ukjent, men a kan estimeres ut fra en stikkprøve med $\hat{a} = Y/n$ der Y er antall kontrollavvik i stikkprøven og n er stikkprøvestørrelsen. Det trekkes et tilfeldig utvalg bestående av n elementer fra populasjonen, og vi har at Y er hypergeometrisk fordelt med parametre (N, M, n) , se f. eks. Ubøe (2016) side 132.

¹ En grundig innføring i det teoretiske grunnlaget for statistiske metoder i revisjon er tilgjengelig i boken «Statistiske metoder i revisjon» av Lillestøl (2000), som tidligere var pensum på revisorstudiet ved NHH. Forfatterne retter en stor takk til professor Aasmund Eilifsen og professor Tonny Stenheim for nyttige innspill i forbindelse med tidligere versjoner av denne artikkelen. Ellen M. Kulset takker Sparebankstiftelsen Nøtterøy-Tønsberg for forskningsmidler tildelt USN til bruk i prosjekter innenfor fagfeltene regnskap og revisjon.

Stikkprøvestørrelse – estimatets presisjon

Stikkprøvens størrelse er bestemmende for hvor presist estimatet å er ved at større stikkprøvestørrelse gir et «smalere» konfidensintervall. Estimatets presisjon uttrykkes ved at det beregnes en øvre grense for andelen kontrollavvik (ensidig konfidensintervall) med en gitt sannsynlighetsgaranti (konfidensnivå).

Dersom den øvre grensen for andelen ligger under en fastlagt vesentlighetsgrense (tolererbar avvikshyppighet), aksepteres den testede interne kontrollen, i motsatt fall aksepteres den ikke.² Når revisor skal bestemme nødvendig stikkprøvestørrelse, bør det tas hensyn til både risikoen for å overvurdere hvor godt internkontrollen fungerer (beta-risiko) og risikoen for å undervurdere hvor godt internkontrollen fungerer (alfa-risiko).

Beta- og alfa-risiko

Med beta-risiko i forbindelse med kontrolltesting forstår vi risikoen for at vi, basert på resultatet i stikkprøven, konkluderer med at internkontrollen fungerer tilfredsstillende når den faktisk ikke gjør det. Med alfa-risiko forstår vi risikoen for at vi, basert på funnene i stikkprøven, konkluderer med at interkontrollen ikke fungerer slik den skal (dvs. vi kan ikke bygge på denne) selv om den faktisk fungerer tilfredsstillende.³

Beta-risiko spesielt viktig i praksis

I praksis vil revisor gjerne kombinere test av kontroll for en påstand med en analytisk substanshandling eller en detaljtest. Dette betyr at beta-risikoen blir spesielt viktig: hvis revisor basert på funnet i en kontrolltest finner at en kontroll kan stoles på selv om denne kontrollen egentlig ikke fungerer, vil han/hun ikke foreta de endringene det er behov for i revisjonsprogrammet med tanke på endring av type, tidspunkt eller omfang av substanshandlingene. I ytterste konsekvens kan revisor i denne situasjonen ende opp med å konkludere feil i revisjonsberetningen fordi han/hun ikke har hentet inn tilstrekkelig med revisjonsbevis.

Alfa-risikoen er derimot mindre viktig ettersom revisor, dersom han/hun konkluderer med at internkontrollen ikke fungerer, vil måtte utføre mer substanstesting enn planlagt. Alfa-risiko blir dermed først og fremst et spørsmål om revisjonseffektivitet.

Stikkprøvestørrelse og øvre grense

Pragmatiske forenklinger trengs

Når revisor skal foreta en test av kontroll ved hjelp av en stikkprøve, er første utfordring å bestemme stikkprøvestørrelsen.

Denne vil kunne avhenge av følgende fem faktorer:

1. populasjonsstørrelsen
2. tolererbar avvikshyppighet
3. forventet avvikshyppighet
4. tolererbar risiko for overvurdering (beta-risiko)
5. tolererbar risiko for undervurdering (alfa-risiko)

For generell teori og alternative metoder se Lillestøl (2000), kapittel 2 og 9. I praksis gjøres ofte pragmatiske forenklinger, og det er dette som er i fokus i det følgende.

Stikkprøvestørrelse ved hjelp av tabeller

Revisorstudenter lærer å bestemme stikkprøvestørrelsen ved å slå opp i tabeller eller ved å bruke en passende revisjonsprogramvare. Vi finner slike tabeller f.eks. som tabell 8–5 og 8–6 i Eilifsen lærebok eller i AICPAs guide som tabell A-1 og A-2, her gjengitt i utdrag som tabell 1 og 2 (bakerst i artikkelen). Studentene lærer her at stikkprøvestørrelsen skal avhenge av (1) ønsket konfidensnivå (2) tolererbar avvikshyppighet og (3) forventet avvikshyppighet.⁴

Vi ser for eksempel av tabell 2 at dersom revisor ønsker et konfidensnivå på 90 %, forventer en avvikshyppighet lik 0 % og tolererer en avvikshyppighet på 10 %, er det passende med en stikkprøvestørrelse på 22.

Finanstilsynets forslag til stikkprøvestørrelser

Finanstilsynet gjennomførte i 2016 et tematisyn med fokus på stikkprøver i revisjon og fant da at de fem store revisjonsselskapene for kontroller som utføres flere ganger pr. dag benytter stikkprøvestørrelser i intervallet 14 til 50 ved test av kontroll (Finanstilsynet 2016).

Basert på funnene i undersøkelsen legger Finanstilsynet til grunn at norske revisorer bør benytte stikkprøvestørrelser i dette intervallet for kontroller som utføres én gang eller oftere pr. dag dersom det er ønskelig med moderat sikkerhet.

Vi ser av tabell 2 at disse stikkprøvestørrelsene er egnet for eksempel dersom revisor ønsker et konfidensnivå på 90 %, forventer en avvikshyppighet lik null og tolererer avvik i intervallet fra 4 % til 15 %. Finanstilsynet spesifiserer ikke nærmere hvordan man skal forstå begrepet «moderat sikkerhet», men ved å legge tabell 2 til grunn er en mulig fortolk-

2 I statistisk terminologi er konfidensnivået til et konfidensintervall for en ukjent parameter definert som sannsynligheten for at intervallet vil inneholde parameterens sanne verdi. I revisjonssammenheng skal en slik konfidensgaranti fungere som sikkerhet for at den faktiske avvikshyppigheten i populasjonen ikke overskrider den tolererbare avvikshyppigheten, jf. ISA 530, vedlegg 2.

3 I boken «Auditing and assurance services», Eilifsen et al. (2014) brukes betegnelsene type I feil og type II feil for henholdsvis undervurdering (alfa-risiko) og overvurdering (beta-risiko). Dette samsvarer med begrepene i teorien for statistisk hypotesetesting med forventet akseptabel andel kontrollavvik som nullhypotese.

4 Tolererbar avvikshyppighet settes lik den avvikshyppigheten revisor er villig til å akseptere at finnes i populasjonen uten at han/hun vil si at internkontrollen ikke kan bygges på i det videre revisjonsarbeidet. Forventet avvikshyppighet kan f.eks. settes basert på tidligere erfaring hos kunden. I praksis setter revisor gjerne denne til null med de konsekvensene dette har for alfa-risikoen revisor opererer med. Dette vil vi komme tilbake til.

ning at dette for gitte forutsetninger kan tilsi et konfidensnivå på 90 %.⁵

Vanlige spørsmål

Revisorstudenter stiller gjerne varianter av følgende spørsmål:

1. Hvorfor skal stikkprøvestørrelsen avhenge av akkurat de tre faktorene listet opp over?
2. Hvordan beregnes egentlig stikkprøvestørrelsene som fremkommer i tabellene i læreboken?
3. Hvordan beregnes øvre konfidensgrense slik denne fremkommer i lærebokens tabeller?

Vi skal i denne artikkelen forsøke å besvare disse spørsmålene ved å presentere ulike talleksempler kombinert med å bruke formelverket i Excel. Vi vil først benytte tilnærmede binomiske sannsynligheter istedenfor de hypergeometriske ettersom AICPA's tabellverk er basert på binomiske sannsynligheter. Deretter ser vi kortfattet på bruk av den hypergeometriske sannsynlighetsfordelingen. For en generell gjennomgang av teorien som ligger bak tabellverket, viser vi til «Technical Notes on the AICPA Audit Guide Audit Sampling» (Stewart, 2012).

Forenklinger som gjøres

Forenklingen til at stikkprøvestørrelsen avhenger av bare tre størrelser istedenfor fem kan forklares som følger:

For det første kan man trygt se bort fra populasjonsstørrelsen, så lenge populasjonen er tilstrekkelig stor og stikkprøven er liten i forhold til denne. Dette er fordi det i dette tilfellet spiller liten rolle både teoretisk og i praksis om den tilfeldige trekningen foregår med tilbakelegging, som svarer til binomisk situasjon med parametre (n, a) , istedenfor uten tilbakelegging, som svarer til hypergeometrisk situasjon med parametre (n, M, N) .

Videre innebærer beregningsmåten at det tabellene omtaler som konfidensnivå ved bruk av statistisk terminologi, er det samme som én minus beta-risikoen, slik at garantier om risikoen for feilaktig aksept blir tatt hensyn til via størrelsen konfidensnivå.

Hva med alfa-risikoen? Det sies eksplisitt i boken til Eilifsen (side 261) at det ikke tas hensyn til alfa-risiko. Dette har den fordel at tabelleringen blir enklere, når revisor bare har tre størrelser å ta stilling til ved oppslag i tabellen istedenfor fire, eventuelt fem med populasjonsstørrelsen. Det er dette som kommer til uttrykk i tabell 8–5 og 8–6 i Eilifsens lærebok, der en må spesifisere vesentlig og forventet prosentandel kontrollavvik og ønsket konfidensnivå. Som sagt ovenfor er dette det samme som én minus beta-risikoen. Vi skal se senere at tabellene implisitt gir en viss garanti for alfa-risikoen likevel.

Betydningen av de tre størrelsene

Hvilken rolle spiller så de tre størrelsene i tabellene ved bestemmelsen av stikkprøvestørrelsen? Vårt primære ønske er at konfidensintervallet fanger opp eventuelle vesentlige kontrollavvik, dvs. at det er liten sannsynlighet for at den øvre konfidensgrensen for andelen kontrollavvik havner under vesentlighetsgrensen (i praksis det vi definerer som tolererbar avvikshyppighet), når populasjonen i realiteten har kontrollavvik på eller over denne grensen.

Bredden av konfidensintervallet, her avstanden fra estimert andel kontrollavvik i populasjonen til den tilhørende øvre konfidensgrense, kalles presisjonen. Jo større ønsket presisjon (dvs. mindre avstand), desto større stikkprøvestørrelse trengs.

For å være ressurseffektiv kan stikkprøvestørrelsen planlegges ut fra situasjonen «på grensen», der ønsket presisjon settes lik tolererbar avvikshyppighet minus forventet avvikshyppighet. Konstruert slik, innebærer det at man ser bort fra alfa-risikoen. For å ta kontroll også på denne må vi sørge for at det er stor sannsynlighet for at den øvre konfidensgrensen havner nedenfor vesentlighetsgrensen (tolererbar avvikshyppighet), i de tilfellene der kundens kontrollsystem er akseptabelt og som forventet. Det betyr at vi må øke presisjonen, noe som krever en litt større stikkprøve.

Konfidensintervaller i tabellene

Vi skal merke oss at konfidensintervaller i tabellene blir konstruert slik at det samsvarer med å akseptere kundens kontrollsystem dersom antall kontrollavvik Y er mindre enn eller lik et kritisk antall c , som kan være null. Dette er grunnlaget for beregningene i eksemplene som følger i denne artikkelen, som avviker fra konfidensintervaller basert på standardavvik og normaltilnærming.

Eksempel

La oss anta at vi har en populasjon bestående av 1000 enheter, at det er en faktisk avvikshyppighet i populasjonen lik 6 % ($a=0,06$), at vi trekker et utvalg på 60 enheter fra populasjonen ($n=60$) og at vi noterer antall avvik i stikkprøven (Y). La oss videre anta at vi ikke finner noen avvik ($Y=0$) i utvalget.

Ettersom denne populasjonen er relativt stor samtidig som utvalget er relativt lite, kan vi nå bruke funksjonen BINOM.FORDELING.N i Excel for å finne ut hvor sannsynlig det er at vi skal få dette utfallet i en slik stikkprøve. Vi taster inn⁶ parameterverdiene $Y = 0$ (antall_s = 0), $n = 60$ (forsøk = 60), $a = 0,06$ (sannsynlighet_s = 0,06) og ber om å få den binomiske punktsannsynligheten (kumulativ = USANN).

Vi får da:

1. $P(Y=0) = \text{Bin}(0; 60; 0,06) = 0,0244$ der Bin står for binomisk punktsannsynlighet.

⁵ Se Kulset (2019) for en nærmere analyse av funnene i Finanstilsynets rapport.

⁶ Vi oppgir her i parentes de parameterverdiene som Excel ber om.

Sannsynligheten for å ikke finne avvik i stikkprøven med de forutsetningene vi har lagt til grunn, er altså 0,0244 eller 2,44 %, og vi kan si at det observerte utfallet er relativt usannsynlig. Hvis vi gjentar slike undersøkelser over tid under de samme betingelsene (gjennomfører et stort antall stikkprøver fra denne populasjonen), vil vi forvente at vi i 97,56 % av tilfellene finner ett eller flere avvik i stikkprøven.

Ukjent faktisk avvikshyppighet

La oss nå ta utgangspunkt i den samme populasjonen som over, men la oss anta at vi ikke kjenner den faktiske avvikshyppigheten og at vi ønsker å gjennomføre en stikkprøve med en stikkprøvestørrelse lik 60 for å lage et estimat på avvikshyppigheten med en tilhørende øvre konfidensgrense. Ettersom vi ikke fant noen avvik i stikkprøven, vil vårt beste estimat på avvikshyppigheten i populasjonen være lik null. For å finne en øvre konfidensgrense bruker vi formel (1) (over) med ulike forutsetninger for faktisk avvikshyppighet (a) og ser av tabellen under at sannsynligheten for ikke å finne noen avvik i stikkprøven på 60, $P(Y=0)$, reduseres, jo høyere faktisk avvikshyppighet vi har i populasjonen.

Avviksandel i populasjonen (a)	0,01	0,02	0,03	0,04	0,05	0,06
$P(Y=0)$	0,547	0,298	0,161	0,086	0,046	0,024

Anta at vi ønsker et konfidensnivå på 95 %, som svarer til beta-risikoen 5 %. Dette betyr at dersom vi gjentar denne stikkprøvetesten mange ganger, forventer vi at den faktiske avvikshyppigheten overskrider den anslåtte øvre grensen (feilaktig aksept) i kun 5 % av tilfellene. Vi ser av tabellen over at dersom det er en avvikshyppighet lik 5 % i populasjonen, og vi ikke finner noen avvik i stikkprøven, vil sannsynligheten for ingen avvik være tilnærmet 5 %. Dette betyr at vi enten har observert et utfall som er relativt usannsynlig der faktisk avvikshyppighet er minst 5 %, eller at faktisk avvikshyppighet er høyst 5 %. Med den siste tolkningen kan 5 % betraktes som en øvre grense for avvikshyppigheten i populasjonen.

Beregne stikkprøvestørrelsen

Hittil har vi bare forutsatt at vi ønsker en stikkprøvestørrelse lik 60. Dersom vi nå ønsker å bestemme hva stikkprøvestørrelsen bør være for ulike forutsetninger, kan vi også benytte funksjonen BINOM.FORDELING.N i Excel.

Faktisk avvikshyppighet i populasjonen er ukjent (det er jo derfor vi vil gjennomføre en stikkprøve), antall avvik i stikkprøven er ukjent inntil stikkprøven er gjennomført og populasjonsstørrelsen er kjent, og forholdsvis stor i forhold til aktuell stikkprøvestørrelse. For å kunne løse ut for n må vi altså bestemme hvilken vesentlighetsgrense (tolererbart avvik) vi ønsker oss, vi må anslå hvor mange avvik vi forventer å finne i stikkprøven og bestemme oss for hvilket konfidensnivå vi finner passende.

Eksempel med tolererbar avvikshyppighet på 6 %

Vi vil fortsatt anta at det kun er stikkprøve uten kontrollavvik som gir grunnlag for å akseptere kundenes kontrollsystem. Dersom vi nå ønsker at sannsynligheten for ikke å observere avvik i stikkprøven, $P(Y=0)$, ved en avvikshyppighet lik 6 % skal være på 5 % (det vil si et konfidensnivå på 95 % med tilhørende beta-risiko på 5 %), kan vi løse ut for n i (2) ved hjelp av en prøv- og feil-metodikk⁷, og vi finner en passende stikkprøvestørrelse på 49.⁸

$$2. P(Y=0) = \text{Bin}(0; n; 0,06) = 0,05 \text{ for } n = 49$$

Vi gjenfinner stikkprøvestørrelsen på 49 i tabell 1 ved å slå opp for forventet avvikshyppighet lik 0 %, tolererbar avvikshyppighet lik 6 % og konfidensnivå lik 95 %.

Mer generelt har vi dermed at dersom vi tolererer avvik på 6 %⁹, ønsker et konfidensnivå lik 95 % og kun ønsker å bygge på internkontrollen i revisjonsarbeidet vårt dersom vi ikke finner avvik i stikkprøven, kan vi sette passende stikkprøvestørrelse (n) til den minste n slik at

$$3. \text{Bin}(0; n; 0,06) \leq 1-x$$

der x er ønsket konfidensnivå og i vårt eksempel

$$4. \text{Bin}(0; n; 0,06) \leq 0,05$$

Det kritiske antall

La oss nå anta at vi vil se på internkontrollen som velfungerende dersom vi finner høyst et lite antall c i stikkprøven, her kalt det kritiske antall. Hittil har vi hatt $c=0$. Vi skal nå se på tilfellene $c=1$ og $c=2$. Generelt foregår nå bestemmelsen av stikkprøvestørrelsen n for gitt valgt kritisk antall c ved å velge den minste n slik at

$$5. \text{CBin}(Y=c, n, a=\text{tol}) \leq 1-x \text{ der tol} = \text{tolererbar avvikshyppighet og CBin er kumulativ binomisk sannsynlighet.}$$

Dette kan foregå ved samme prøv- og feilmetodikk som ovenfor. I eksemplet med tolererbart avvik 6 % og 95 % konfidensnivå har vi da følgende: Dersom internkontrollen vil bli betraktet som velfungerende dersom det er null eller ett avvik i stikkprøven, vil en passende stikkprøvestørrelse være 78. Dersom vi ønsker å kunne se på internkontrollen

⁷ Med prøv- og feilmetodikk mener vi her at vi først gjør et anslag på hva en passende stikkprøvestørrelse kan være og deretter tester ut flere forskjellige stikkprøvestørrelser til vi har kommet så nær opptil 0,05 som mulig. Denne prøv- og feilmetodikken kan være nyttig i bruk for å forstå hvordan stikkprøvestørrelsen fremkommer. Når revisor i praksis skal komme frem til passende stikkprøvestørrelse, vil det naturligvis være mer fornuftig å benytte en passende programvarepakke for å finne løsningen (f.eks. ACL, IDEA el.), eventuelt kan revisor selv programmere algoritmen han/hun har brukt for å slippe å «prøve og feile» hver gang, f.eks. i Excel. En mulig VBA-kode finner vi i «Audit Sampling: Technical notes».

⁸ Vi bruker formelen BINOM.FORDELING.N i Excel og finner at sannsynligheten for parameterverdiene (0;48;0,06) er lik 0,0513 og for (0;49;0,06) er lik 0,04822, dvs. minste n er lik 49.

⁹ Med tolererbar avvikshyppighet lik 6 % mener vi at hvis beregnet avvikshyppighet i populasjonen inklusiv avsetning til utvalgsrisiko (øvre grense) er større enn 6 %, vil vi konkludere med at internkontrollen ikke fungerer (ikke kan bygges på).

som velfungerende dersom vi finner null, en eller to avvik i stikkprøven, vil tilhørende stikkprøvestørrelse være 103.

For å finne disse stikkprøvestørrelsene i Excel bruker vi fremdeles funksjonen BINOM.FORDELING.N. For $Y = 1$ taster vi for eksempel inn parameterverdiene $Y = 1$ (antall_s = 1), $a = 0,06$ (sannsynlighet_s = 0,06) og prøver oss frem med ulike verdier for n . Løsningen finner vi for den n der den kumulative binomiske sannsynligheten (kumulativ = SANN) er tilnærmet lik 0,05, dvs.:

6. $P(Y \leq 1) = \text{CBin}(1; n; 0,06) = 0,0479 \approx 0,05$ for $n = 78$ der CBin står for kumulativ binomisk sannsynlighet.

Valg av stikkprøvestørrelse

Spørsmålet blir naturligvis om vi skal velge en stikkprøvestørrelse på 49, 78 eller 103. Ved første øyekast synes det fristende å velge minst mulig stikkprøve, men vil dette alltid være et fornuftig valg?¹⁰

La oss anta at det er en avvikshyppighet på 1 % i populasjonen, at vi fortsatt tolererer 6 % avvik, at vi har valgt en stikkprøvestørrelse på 49 og at vi finner ett avvik i stikkprøven.

Sannsynligheten for å finne ett eller flere avvik i en stikkprøve på 49 enheter med disse forutsetningene vil være $P(Y \geq 1) = 0,389$,¹¹ dvs. disse utfallene er relativt sannsynlige, men vi vil likevel konkludere med at internkontrollen ikke fungerer ettersom vi kan lese av en øvre grense på ca. 9,2 % i tabell 3 for n lik 50.

Med en faktisk avvikshyppighet lik 1 % (merk at denne i praksis naturligvis er ukjent) fungerer internkontrollen i dette eksemplet mer enn tilfredsstillende, men vi vil likevel ende opp med ikke å bygge på internkontrollen ettersom øvre grense er høyere enn tolererbar avvikshyppighet.

Dersom vi finner ett avvik i et utvalg på 100, kan vi imidlertid konkludere med at internkontrollen fungerer (øvre grense på 4,7 %). Vi ser altså at en lav stikkprøvestørrelse vil føre til en høy alfa-risiko, dvs. en høy risiko for å konkludere med at internkontrollen ikke fungerer når den faktisk gjør det.

Valget vi tar mellom en stikkprøve på 49 (med kritisk antall $c = 0$), 78 (med kritisk antall $c = 1$) eller 103 enheter (med kritisk antall $c = 2$), som i eksemplet over, skal ifølge tabellverket avhenge av hvilken avvikshyppighet vi forventer i populasjonen. I vårt eksempel betyr det at dersom vi ikke forventer avvik, bør vi benytte en stikkprøvestørrelse på 49. Dersom vi forventer en avvikshyppighet på 1,25 % eller lavere (men større enn null), velger vi en stikkprøve på 79

og dersom vi forventer en avvikshyppighet i populasjonen på 1,5 % eller 1,75 %, velger vi en stikkprøvestørrelse på 103. Dette betyr implisitt at vi ikke bare tar hensyn til beta-risiko, men også alfa-risiko når vi velger stikkprøvestørrelse via vårt anslag på forventet avvikshyppighet.

Vi ser altså at stikkprøvestørrelsen avhenger av tolererbar avvikshyppighet, forventet avvikshyppighet og ønsket konfidensnivå. Selv om AICPA-tabellen ikke tar eksplisitt omsyn til alfa-risikoen som input, gir de angitte stikkprøvestørrelser en viss trygghet innenfor det tabellområdet som er mest aktuelt i praksis. Om ønskelig kan vi eksplisitt beregne alfa-risikoen som

7. Alfa-risiko = $1 - \text{CBin}(Y=c, n, a=\text{forventet})$

I vårt eksempel med vesentlig andel kontrollavvik lik 6 %, og med antatt forventet andel lik 1 %, ville vi valgt stikkprøvestørrelse $n=78$ med tilhørende kritisk antall $c=1$. Da blir alfa-risikoen $1 - \text{CBin}(1; 78; 0,01) = 0,184$.¹² De tre presenterte valg (n, c) av stikkprøvestørrelse (n) og kritisk antall (c), gir følgende alfa-risikoer: For (48, 0) lik 38,9 %, for (79, 1) lik 18,7 % og for (103, 2) lik 8,5 %. Her vil $n=79$ fremstå som det mest aktuelle alternativet, gitt betraktning om kostnadseffektivitet, i samsvar med tabelloppslaget ovenfor.¹³

Vurdere resultatet av stikkprøven – øvre grense

Etter at stikkprøvestørrelsen (n) er fastlagt og resultatet fra stikkprøven (Y) er kjent, kan beslutning om aksept eller ikke trekkes dersom tabellen også har angitt det kritiske antall (c). Uten denne bruker vi tabeller som angir øvre konfidensgrense med ønsket konfidensnivå for den faktiske andelen kontrollavvik basert på antall avdekkede avvik (Y) for den gitte stikkprøvestørrelsen (n).

Tabell 3 (bakerst i artikkelen) er et utdrag av en slik tabell for konfidensnivå 95 %, som kan gjenfinnes i AICPA's guide og i de fleste lærebøkene i revisjon. Den øvre grensen fremkommer ved en ligning svarende til (5), der det kritiske antallet c er erstattet med det observerte antallet Y og tolererbar avvikshyppighet «tol» er erstattet med «øvre grense». Her løser vi ut for denne, som den største andel a som oppfyller:

8. $\text{CBin}(Y, n, a=\text{øvre grense}) \leq 1-x$

Dersom vi har gjennomført en stikkprøve med n lik 60, har funnet akkurat ett avvik i stikkprøven og ønsker et konfi-

12 Alfa-risikoen ved oppslag i AICPA-tabellen vil variere fra tilfelle til tilfelle, og kan i prinsippet bli mye større, spesielt i situasjoner der forventet andel kontrollavvik er nær opp til vesentlighetsgrensen. Våre beregninger for tabellen med 95 % konfidensnivå tyder imidlertid på at den sjelden er over 30 %, og oftest langt mindre og under 10 % i det mest aktuelle tabellområdet. Dette henger sammen med at kritisk verdi c er heltall, og ingen kombinasjon (n, c) treffer beta-risikoen eksakt, slik at overoppfylling av beta-risiko kommer oss til gode i alfa-risikoen.

13 Tabeller for stikkprøvestørrelser som tar eksplisitt omsyn til alfa-risikoen som input blir lett upraktiske. Beregninger lar seg gjennomføre ved prøving og feiling med utgangspunkt i (5) og (6). Oppgaven er da å finne den kombinasjonen (n, c) som oppfyller begge risikobeskrankingene.

10 I praksis velger gjerne revisor en stikkprøvestørrelse i intervallet 20 til 60 ved testing av frekvente interne kontroller, se for eksempel revisors stikkprøvepraksis slik den rapporteres i Christensen et al. (2015) og Finanstilsynet (2016) med de konsekvensene dette har for testens forkastningsrisiko.

11 Dette er alfa-risikoen, gitt at forventet andel avvik var 1 %.

densnivå lik 95 %, får vi en øvre grense lik 7,7 %, jf. avlesning av øvre grense slik den fremkommer i tabell 3. Ved to avvik i stikkprøven får vi en øvre grense lik 10,2 %.¹⁴

Andre spørsmål

Av og til dukker det opp spørsmål om bruken av hypergeometrisk fordeling istedenfor binomisk fordeling og om bruk av tosidige konfidensintervaller.

Med enkel tilgang til programvare, er det i dag egentlig ingen grunn til å benytte binomisk tilnærming i en situasjon der populasjonsstørrelsen er gitt.¹⁵ Fremgangsmåten for å finne riktig stikkprøvestørrelse og øvre grense vil være som skissert over med riktig formel i Excel, dvs. HYPGEOM.FORDELING.N istedenfor BINOM.FORDELING.N. Formelen ser slik ut:

9. $CHyp(Y, n, M, N) \leq 1 - x$ der CHyp er kumulativ hypergeometrisk sannsynlighet.

Forskjellen fra en binomisk formel

Forskjellen fra den binomiske formelen er at nå er nøkkelparameteren antall kontrollavvik i populasjonen M istedenfor andelen kontrollavvik a, og at vi må taste inn denne sammen med populasjonsstørrelsen N. Ved planlegging av nødvendig stikkprøvestørrelse (n) bruker vi denne formelen med $Y=c$ (valgt kritisk antall) og $M=\text{tolererbart antall}$, og løser ut for den minste n som gir uttrykket høyst lik beta-risikoen $1-x$. Dette tilsvarer beregningen i formel (5) i den binomiske situasjonen.

Hvis vi nå ønsker å finne passende stikkprøvestørrelse for en populasjon på $N=1000$ med de samme forutsetningene som tidligere (kritisk antall c lik 0, tolererbar andel $a=M/N$ lik 0,06 og ønsket konfidensnivå x lik 0,95), betyr dette at tolererbar M er lik 60. Vi finner ved prøving og feiling med formelen CHyp (0, n, 60, 1000) at en passende n er lik 48, dvs. tilnærmet den samme som den (49) vi fant ved å benytte den binomiske formelen. I praksis er det altså lite å hente ved å bruke den hypergeometriske formelen når populasjonen er så vidt stor.

For gitt observert Y finner vi øvre grense for antall kontrollavvik i populasjonen ved å løse ut for det minste hele tallet M slik at ulikheten (8) er oppfylt. Med en stikkprøvestørrelse på 60, ett avvik i stikkprøven og ønsket konfidensnivå 95 %, er spesifikasjonen CHyp (1, 60, M, 1000), og vi finner den øvre grensen for M lik 75, dvs. andel 7,5 %, bare litt lavere enn 7,7 % som vi fikk med den binomiske formelen.

Dersom vi ikke har passende programvare tilgjengelig og ikke ønsker å benytte en slik prøv- og feilmethode, kan vi

f.eks. benytte VBA-programmet under, se Appendiks, for å finne passende n ved ulike parameterverdier.¹⁶

Tosidig konfidensintervall

Dersom vi ønsker et tosidig konfidensintervall,¹⁷ vil øvre konfidensgrense ved et 90 %-tosidig intervall tilsvare den øvre konfidensgrensen ved et 95 %-ensidig intervall. Den nedre grensen finner vi på tilsvarende måte som skissert tidligere ved å løse ut for største M som oppfyller ligning (10) under:¹⁸

10. $CHyp(Y-1, n, M, N) \geq x$

der CHyp er kumulativ hypergeometrisk sannsynlighet.¹⁹

Ved et ønsket tosidig nivå på 90 %, avvik i stikkprøve lik 2, n lik 60 og N lik 1000 får vi da:

11. $CHyp(1, 60, M, N) \geq 0,95$

og ved å løse ut for M finner vi nedre grense lik 0,7 % og vi rapporterer resultatet fra stikkprøven slik:

Estimat på avvikshyppighet i populasjonen: 3,3 % (= 2/60)

Nedre grense: 0,7 % Øvre grense: 9,9 %

Avsluttende kommentarer

Vi har i denne artikkelen brukt talleksempler og Excel for å svare på spørsmål revisjonsstudenter ofte stiller når de lærer om test av kontroller i revisjonssammenheng. Videre har vi presentert noe av det teoretiske grunnlaget som ligger til grunn når revisor gjennomfører en stikkprøve for å teste en internkontroll hos en kunde. For svar på andre spørsmål vises det til den mangfoldige spesiallitteraturen som finnes på dette området, både på norsk og på engelsk, og i litteraturlisten er det derfor inkludert sentrale bøker som er publisert innenfor fagfeltet som kan være interessante for artikkelens lesere.

¹⁴ Legg merke til at (5) og (7) i utgangspunktet er samme formel, som likevel uttrykker to ulike saksforhold med hver sin oppgave å løse.

¹⁵ Store populasjonsstørrelser medfører imidlertid numerisk krevende beregninger, og programvare har ofte vist seg å være upålitelig.

¹⁶ Merk at ettersom VBA ikke skiller mellom store og små bokstaver til bruk i formler, har parameteren N fått benevnelsen P i dette programmet.

¹⁷ Bruk av et tosidig konfidensintervall vil i revisjonssammenheng særlig være aktuelt i forbindelse med offentlig revisjon. Et eksempel kan være dersom revisor skal sjekke om faktisk responstid for en tjeneste er i samsvar med en forhåndsbestemt benchmark. Dersom revisors utgangspunkt er at det forventes en avvikshyppighet lik null, er det ikke aktuelt med bruk av et tosidig konfidensintervall.

¹⁸ Det finnes «hypergeometriske intervall-kalkulatorer lett tilgjengelig på nett som kan benyttes for å spesifisere øvre og nedre konfidensgrenser for en gitt stikkprøvestørrelse med tilhørende avdekkede avvik, se f.eks. www.cluster-text.com/confidence_interval.php.

¹⁹ Y-1 fordi vi her søker en sannsynlighet for $\geq Y$ som for heltall er 1- sannsynligheten for $\leq Y-1$.

Litteratur

- AICPA. 2014. *Audit Sampling*. New York: AICPA.
- Arkin, H. 1982. *Sampling methods for the auditor: an advanced treatment*. McGraw-Hill Book Company.
- Bailey, A. D. 1981. *Statistical Auditing: Review, Concepts and Problems*. New York: Harcourt Brace Jovanovich Inc.
- Christensen, B. E., Elder, R. J. og S. M. Glover. 2015. Behind the numbers: Insights into Large Audit Firms Sampling Policies. *Accounting Horizons*. Vol. 29, No. 1: 61–81.
- Eilifsen, Aa, Messier, W. F., Glover, S. M. og D. F. Prawitt. 2014. *Auditing & Assurance Services. Third International Edition*. Berkshire: McGraw Hill Education.
- Finanstilsynet. 2016. Stikkprøver i revisjon. Tematilsyn 2016.
- Kulset, E. 2019. Revisors bruk av profesjonelt skjønn i forbindelse med stikkprøvebasert testing av frekvente interne kontroller. *Trender og utfordringer i regnskap og revisjon*. Fagbokforlaget.
- Lillestøl, J. 2000. *Statistiske metoder i revisjon*. Oslo: Cappelen Akademisk Forlag.
- Roberts, D. M. 1978. *Statistical Auditing*. New York: AICPA.
- Stewart, T. R. 2012. *Technical Notes on the AICPA Audit Guide Audit Sampling*. New York: AICPA
- Wilburn. A. 1985. Wilburn. *Practical Statistical Sampling for Auditors*. Marcel Dekker, inc.
- Ubøe, J. 2016. *Statistikk for økonomifag*. Oslo: Gyldendal Akademisk.

APPENDIKS**VBA-kode for å finne passende stikkprøvestørrelse – hypergeometrisk sannsynlighetsfordeling**

Function HypgeoSample(risk As Double, Y As Double, M As Double, P As Double)

Dim n As Double

If risk <= 0 Or risk > 1 Or Y < 0 Or M <= 0 Or P < 1 Then

HypgeoSample = CVErr(x1ErrNum)

Else

n = 1

While Application.WorksheetFunction.HypGeom_Dist(Y, n, M, P, True) > risk And n <= 10 000

n = n + 1

Wend

HypgeoSample = IIf(Application.WorksheetFunction.HypGeom_Dist(Y, n, M, P, True) <= risk, n, CVErr(x1ErrNA))

End If

End Function

Tabell 1 – Stikkprøvestørrelse som en funksjon av tolererbar og forventet avvikshyppighet ved et ønsket konfidensnivå på 95 % (hentet fra Eilifsen et al. 2014 tabell 8-5)

Forventet avvikshyppighet i populasjonen	Tolererbar avvikshyppighet									
	3 %	4 %	5 %	6 %	7 %	8 %	9 %	10 %	15 %	20 %
0,00 %	99	74	59	49	42	36	32	29	19	14
0,25 %	157	117	93	78	66	58	51	46	30	22
0,50 %	157	117	93	78	66	58	51	46	30	22
0,75 %	208	117	93	78	66	58	51	46	30	22
1,00 %	*	156	93	78	66	58	51	46	30	22
1,25 %	*	156	124	78	66	58	51	46	30	22
1,50 %	*	192	124	103	66	58	51	46	30	22
1,75 %	*	227	153	103	88	77	51	46	30	22
2,00 %	*	*	181	127	88	77	68	46	30	22

Tabell 2 – Stikkprøvestørrelse som en funksjon av tolererbar og forventet avvikshyppighet ved et ønsket konfidensnivå på 90 % (hentet fra Eilifsen et al. (2014) tabell 8-6)

Forventet avvikshyppighet i populasjonen	Tolererbar avvikshyppighet									
	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 3 – Øvre konfidensgrense ved et ønsket konfidensnivå på 95 % (hentet fra Eilifsen et al. 2014 tabell 8-8)

Utvalgsstørrelse	Faktisk antall avdekkede avvik			
	0	1	2	3
50	5,9	9,2	12,1	14,8
55	5,4	8,4	11,1	13,5
60	4,9	7,7	10,2	12,5
70	4,2	6,6	8,8	10,8
80	3,7	5,8	7,7	9,5
90	3,3	5,2	6,9	8,4
100	3,0	4,7	6,2	7,6
125	2,4	3,8	5,0	6,1
150	2,0	3,2	4,2	5,1
200	1,5	2,4	3,2	3,9

Tabell 4 – Øvre konfidensgrense ved et ønsket konfidensnivå på 90 % (hentet fra Eilifsen et al. (2014) tabell 8-9)

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

BLI MEDLEM!

Fordeler for medlemmer:

- Gode rabatter på alle kurs, tidsskrifter og bøker
- Tilgang til vår faglige spørretjeneste
- Full tilgang til våre nettsider med relevant og nyttig faginformatjon
- God bankavtale – svært gode lånevilkår
- Rabatter på BMW og Mini

Hvorfor vente?
Meld deg inn i dag!

revisorforeningen