En middels god Agricola-spillers bekjennelser: Analyse av mine *live*-spill, 2014–2023

Rune Halvorsen (2024 01 02)

Innledning

Jeg ble hektet på Agricola i 2014. Fra sommeren 2014 og fram til utgangen av 2023 har jeg spilt 848 *live*-spill; 765 «fysiske» og 83 på play-agricola.com. Jeg er over middels glad i statistikk og Excel-ark, og liker å analysere data. Fra de aller første spillene jeg spilte, begynte jeg å lagre spillresultater og gjøre notater. Mange har spurt hva jeg noterer på de små, gule lappene og hva jeg bruker notatene til. Jeg har svart at planen var å analysere dem – en eller annen gang i framtida. Nå har jeg brukt et par dager i romjula på å gjøre noen enkle analyser, og i denne artikkelen oppsummerer jeg resultatene. Jeg har fått noen overraskelser og lært en god del, og har lyst til å dele resultater og kommentarer med alle som er interessert. Ikke fordi jeg tror min måte å spille på nødvendigvis er så mye å etterape, men fordi jeg er helt sikker på at resultatene kan gi grunnlag for mange interessante diskusjoner – som kan gi l'ring b[de for meg og andre.

Materialet

Innimellom spiller vi «tøysespill», «nybegynnerspill» med spesialtilpasninger, etc. Bare spill med standardoppsett er tatt med i analysen. Med det mener jeg:

– Draft 9 (altså er Fischer- og Bridge-spill utelatt)
– Ingen andre decks ann EIKWmFr, G, G4G5G6G7 (eller globuskort) og 'NorwayDeck' (AY)

Jeg har dessuten utelatt spill der *jeg* har spilt i par, mens spill der jeg har spilt aleine mot par er med.
Det er noen hull i notatene mine. For å bli med i analysen, måtte jeg ha fullstendige notater om:

– Min startposisjon
– Poeng og plassering (merk at plassering kan være 1, delt 1., 3-delt 1., 2., delt 2., 3-delt 2., 3., delt 3. og 4).
– Relativ skår (som i turneringene; forskjellen mellom min skår og gjennomsnittsskåren på bordet)
– Antall *Actions* (uten å ta hensyn til *Occs* og *Minors* som *de facto* gir ekstra *Actions*, f.eks. **AdoptiveParents**, **SlapdashRenovation**, etc.)
– Hvilket materiale huset var bygd av og hvor mange rom jeg bygde
– Hvilken runde FG-*Action* ble tilgjengelig
– Hvilken runde jeg fikk min 1. FG [i spill der jeg ikke tok FG (pga. **Cowboy&Mother**) har jeg registrert min 1. FG som FG i «runde 15»)
– Hvor mange *Occs* jeg spilte, og hvilke
– Hvor mange *Minors* jeg spilte (*pass-on minors* har jeg ikke registrert systematisk, så de er ikke talt med), og hvilke
– Hvor mange *Majors* jeg spilte, og hvilke
– Hvor mange poeng jeg fikk i hver av de 14 skåringskategoriene (inkludert *BeggingCards*)

Antallet aksepterte spill var n = 742.
I begynnelsen noterte jeg ikke når FG-*Action* ble tilgjengelig. Analyser av betydningen av dette er derfor utført på et deldatasett som består av de n = 709 spillene som inneholder disse opplysningene.

Bruksanvisning til leseren

Som forsker sitter det i ryggmargen at all dokumentasjon av materiale, metoder og resultater skal være fullstendig og nøyaktig. Datamateruealet består av tall som egner seg godt for statistisk analyse, som forutsetter kunnskap om statistiske metoder. Jeg har forsøkt å unngå statistisk stammespråk, men en del begreper er umulig å unngå. Jeg har derfor skrevet et kapittel (med grå skrift) som forklarer hva jeg har gjort, for de som måtte være interessert. For lesere uten grunnkunnskaper i statistikk vil nok alle *p*-verdiene og tallene være vanskelig å forstå fullt ut. Jeg har derfor oppsummert konklusjonene for hvert tema i rød skrift, uten bruk av tall eller vanskelige begreper. Disse skal kunne leses og forstås av alle.

Jeg har valgt å bruke de originale, engelske begrepene om grunnleggende elementer i Agricola-spillet, kort etc. Disse er konsekvent skrevet med stor forbokstav og isatt i *kursiv*. Navn på kort for *Minor Improvements* (*Minors*) og *Occupations* er skrevet i **fet skrift** for å være lette å se. Av gammel vane skriver jeg kortnavnene i ett, uten mellomrom, f.eks. **LadyInWaiting**.

Tre variabler som karakteriserer et godt spill

I nesten all Agricola-litteratur brukes seier som mål på et godt spill, og *Win%*, det vil si prosentandelen vunnete spill, brukes som mål på hvilke valg som resulterer i gode spill. Jeg har derfor også brukt *Win%* som mål på godt spill. Men som mål på godt spill, har *Win%* også klare ulemper:

– Det er et grovt mål; i hvert spill er det bare to utfall; seier eller «ikke-seier».
– Det tar ikke hensyn til at det, i de fleste sammenhenger (for eksempel turneringer), er mye bedre å bli nummer 2 enn å bli nummer 4.

I hvert enkelt spill er det poengene som avgjør. Jeg har derfor også brukt poeng, og relativ skår som er beregnet på grunnlag av poengene, til å karakterisere resultatet. Dette kommer jeg tilbake til nedenfor.

En middels god Agricola-spiller

Før jeg starter gjennomgangen av resultatene, vil jeg begrunne hvorfor jeg anser meg sjøl som en middels god Agricola-spiller. Tallene forteller nemlig sitt tydelige språk!

Av de 742 spillene, har jeg vunnet 164. Det gir en seiersprosent *Win%* = 22,10 %, eller, uttrykt på den måten som vil bli brukt i resten av denne artikkelen, en seiersfrekvens WinF = 0,2210. Av disse seirene er det 9 delte og to 3-delte. 3-delte seire er kanskje nærmest å betrakte som andreplasser? Jeg har derfor brukt tallet 163 til å karakterisere min seiersfrekvens i analysene, *WinF* = 0,2197.

Dette er jo et langt lavere tall enn man skulle forvente av en helt middels spiller, som burde hatt en seiersfrekvens mellom 0,26 og 0,27. Når jeg likevel anser meg som en middels spiller, er det fordi jeg også har langt færre fjerdeplasser enn jeg skulle hatt dersom plasseringene var helt tilfeldig fordelt, og fordi jeg har litt flere andreplasser enn tredjeplasser, enten vi tar hensyn til delte plasseringer eller ikke (Tabell 1).

Tabell 1. Fordeling av plasseringer og relativ skår i mine spill (n = 742).

Plassering	1			2			3		4
	udelt	2-delt	3-delt	udelt	2-delt	3-delt	udelt	2-delt	–
Poeng (gj.sn.)	50,01	48,44	46,00	45,34	45,32	38,33	40,03	41,31	36,36
Relativ Skår, gj.sn.	8,053	4,543	1,835	2,334	0,289	–2,457	–2,868	–3,962	–7,576
Antall spill (n)	153	9	2	202	19	3	187	16	151

Tabellen forteller også noe viktig om bruken av poeng (*Pts*) og relativ skår (*RelSc*) til å karakterisere resultatet.

Gjennomsnittet av antall poeng jeg har fått i de 742 spillene, er *Points* = 43,06. Poengforskjellen mellom udelt 1. og udelt 2.-plass er 4,7, mellom udelt 2. og 3.-plass 5,3 og melom udelt 3. og 4.-plass 3,7 poeng. Videre ser vi at det er 5,7 relativ skår (*RelSc*)-enheter mellom udelte 1.- og 2.-plasser, 5,2 enheter mellom udelte 2.- og 3.-plasser og 4,7 enheter mellom udelt 3.-plass og 4.-plass. Grovt sett kan vi derfor si at 5 poeng og 5 *RelSc*-enheter svarer til en forskjell på én plass på resultatlista, gitt at det ikke er *ties*.

Hvorfor er dette viktig? Jo, for det første er poeng en dårlig indikator på om et spill har gått bra. Jeg har vunnet spill på 34 poeng og på 64 poeng, og jeg har også blitt nummer to med 64 poeng (jeg har fått 60 poeng eller mer fem ganger)! Poengtallet gjenspeiler nivået på kortene (og spillerne!), om FG kom i runde 5, hvor mange fellesbonuskort som ble spilt, og mange andre forhold. Relativ skår jevner ut alle forskjeller mellom spill (men ikke mellom spillere), og er derfor for et mye mer robust mål på hvor godt det gikk i et spill. Men, og her kommer hovedpoenget: Både når vi ser på poeng og når vi ser på relativ skår, svarer ca. 5 enheter til en forskjell på én plass på den endelige resultatlista. Dette gir oss nøkkelen til å tolke forskjeller i *RelSc,* som jeg vil bruke som «hovedmål» på resultat i spillene: 1 *RelSc*-enhet betyr omtrent ett poeng mer sammenliknet med de andre spillerne, og at en spiller klatrer en femdel av vegen fra én plassering til neste, høyere plassering.

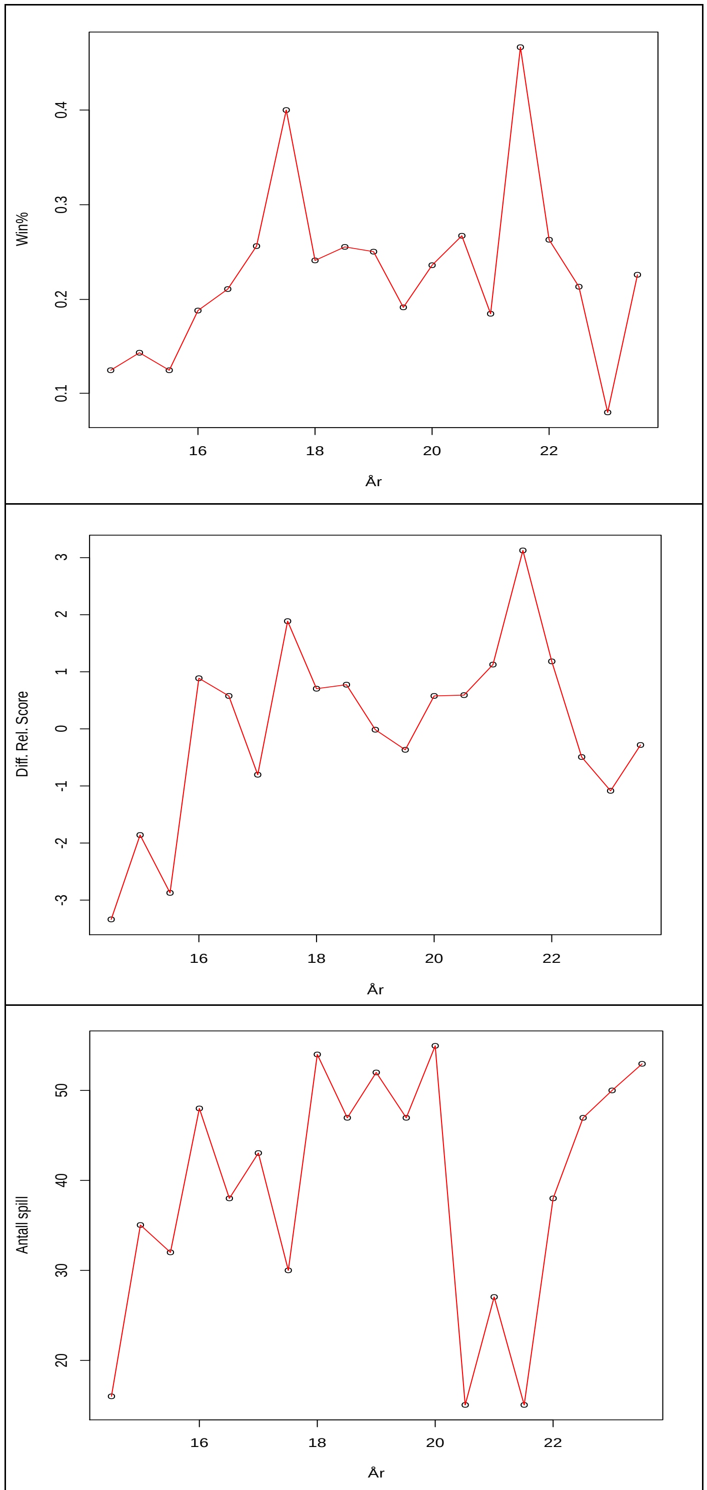

Fig. 1. (a) Gjennomsnittlig seiersfrekvens (*WinF*), (b) gjennomsnittlig relativ skår (RelSc) og (c) antall spilte spill, per halvår fra om med høsten 2014 til og med høsten 2023. Markeringene på x-aksen (16, 18 etc.) markerer vårhalvåret disse årene.

Denne middels gode Agricola-spilleren har vært middels god gjennom det meste av «karriæren» (Fig. 1). Riktignok tok det to år å få den gjennomsnittlige seiersfrekvensen opp over 0,2, men med unntak for våren 2023 da jeg bare vant 8 % av de 50 spillene jeg spilte, har seiersprosenten holdt seg nær eller over 20 % (Fig. 1b viser at det ble mange 2.-plasser). Grunnen til det dårlige resultatet våren 2023 tror jeg først og fremst skyldes at de fleste spillene var med «globus-kortene», som jeg ikke har hatt tid til å lære meg på annen måte enn gjennom å spille dem. Heldigvis tyder resultatene fra høsten 2023 på at det går rette vegen!

Toppene, høsten 2017 (40 % spill vunnet) og høsten 2021 (46,7 % vunnet), kom i halvår da jeg, av ulike grunner, spilte færre spill enn ellers. Med unntak for oppstartsperioden og muligens en effekt av at vi tok i bruk «globus-kortene», har jeg godt grunnlag for å karakterisere meg som en middels spiller gjennom mesteparten av det siste tiåret! Likevel tror jeg at jeg er en bedre spiller ved inngangen til 2024 enn jeg var, for eksempel, i 2016. Men samtidig har også alle andre blitt bedre. Jeg tror at *red queen hypothesis*, som er en sentral teori i evolusjonsbiologien (foreslått av van Valen i 1973), også gjelder for Agricola: arter må hele tiden tilpasse seg, utvikle seg og spre seg for å overleve. Eller, som den røde dronningen sa til Alice i Eventurland i Lewis Carroll's bok fra 1871 da hun skulle forklare Alice hvordan det fungerte i *Looking-Glass Land*: «Now, here, you see, it takes all the running you can do, to keep in the same place.»

Statistiske analysemetoder

Spørsmålene vi ofte stiller oss og som diskuteres om og om igjen, er av typen: «Hvilken startposisjon gir størst vinnersjanse?» «Lønner det seg å 'presse' *StartingPlayer* når jeg har bygd rom først, og FG ikke kommer i runde 5?» «Bør jeg bygge det 4. rommet?» Disse, og andre spørsmål ber om statistisk dataanalyse.

I denne artikkelen har jeg brukt «klassisk statistiske metoder», det vil si metoder som bygger på antakelser som for eksempel:

– at resultatet i ett spill ikke påvirker resultatet i andre spill (det vil si at jeg greier å «nullstille» etter en serie tap – eller seiere)
– at fordelingen av resultatene følger tilfeldighetens lover, f.eks. at verdiene for relativ skår fordeler seg likt på verdier som er større enn gjennomsnittsverdien og verdier som er mindre enn gjennomsnittsverdien

I klassisk statistikk tester vi nøytrale nullhypoteser. Det innebærer at vi starter testing med å sette opp en hypotese som beskriver resultatet, gitt at det bare er tilfeldigheter som virker inn. De nøytrale nullhypotesene som svarer til spørsmålet «Hvilken startposisjon gir størst vinnersjanse?» er:

(1) Vinnnerfrekvensen er like stor, uansett startposisjon.
(2) Relativ skår er like stor, uansett startposisjon.

Jeg har tidligere begrunnet hvorfor jeg har brukt *WinF* = 0,2197 som referanseverdi for vinnerfrekvens, altså min forventete vinnerfrevensen i et tilfeldig valgt spill jeg har spilt. Den statistiske analysen går ut på å teste om *forskjellen* mellom observert vinnerfrekvens og referanseverdien varierer mer mellom startposisjonene enn vi skulle forvente dersom det bare er tilfeldig variasjon som skaper variasjon i vinnerfrekvensen.

For relativ skår, er verdien *RelSc* = 0,29 benyttet som referanseverdi. Det er den gjennomsnittlige relative skåren for alle mine spill (n = 742). I analysene har jeg regnet om alle *RelSc*-verdier til «relativ relativ skår» (*rRelSc*), det vil si den observerte relative skåren minus 0,29. Det gjør alle analyser mye enklere, fordi den forventete forskjellen i relativ relativ skår, gitt at det bare er tilfeldig variasjon, er 0.

Fra et statistisk synspunkt, er *WinF* og *rRelSc* veldig forskjellige variabler, som må analyseres med forskjellige metoder. Egentlig er *rRelSc* enklere å analysere enn *WinF*. Jeg skal forsøke å forklare dette nærmere (de som har full kontroll på den klassiske statistikken og de som ikke er interessert i å bry seg med statistikk kan hoppe over dette).

rRelSc er en såkalt kontinuerlig variabel. Den kan i prinsippet anta alle positive og negative verdier, og i et tilfeldig valgt spill der jeg er med, er det forventet av min relative relative skår (*rRelSc*) skal være 0. For ethvert utvalg av spill kan vi teste om den gjennomsnittlige *rRelSc* for dette utvalget avviker mer fra forventningsverdien 0 enn vi skulle forvente i et tilfeldig utvalg av spill. Et eksempel på en slik testsituasjon er følgende nullhypotese: «(Gjennomsnittsverdien for) *rRelSc* er den samme når jeg ender i *Wooden Rooms* som i et tilfeldig valgt spill.» Denne hypotesen tester vi med en **såkalt ettutvalgs t-test**. I denne testen regnes det ut et mål, *t-verdien*, som uttrykker hvor stort avvik det er mellom gjennomsnittsverdien i utvalget vårt og forventningsverdien. t-verdien er en såkalt **teststatistikk**. Dersom forutsetningene for testen er oppfylt og det ikke har noen betydning for sluttresultatet hvilket materiale huset mitt er bygd i, vil t-verdien være nær 0. Den observerte t-verdien kan «oversettes» til en sannsynlighet for at utvalget skal ha en gjennomsnittsverdi for *rRelSc* som er så forskjellig fra 0 som vi har observert, gitt at hypotesen faktisk er sann (altså at det egentlig ikke er noen årsakssammenheng mellom hvilket materiale huset er bygd i og resultatet). Denne sannsynligheten kaller vi testens **p-verdi**. Alle sannsynligheter angis på en skala fra 0 (helt usannsynlig; noe som ikke *kan* skje) til 1 (helt sikkert). p-verdien bruker vi til å vurdere hypotesen. En p-verdi på 0,05 betyr at gjennomsnittsverdien i utvalget vårt er så forskjellig fra 0 at det bare er 5 % sjanse for at vi skal finne et så avvikende gjennomsnitt dersom resultatet i det store og hele ikke avhenger av hvilket materiale huset er bygd i. Jo lavere p-verdi, desto sikrere kan vi være på at det spiller noen rolle om vi har renovert vår *WoodenHut*. I mine spill endte jeg i *Wood* i 37 av 742 spill. Gjennomsnittlig rRelSc = –1,358 for disse spillene, det vil si at jeg skårer mellom 1 og 2 poeng mindre enn gjennomsnittet for alle spill. Men p-verdien på 0.2195 (21,95%) er ikke lav nok til at vi tør konkludere at det er sannsynlig at det dårligere å ende i *WoodenHut* enn i et tilfeldig valgt spill. Det finnes ingen absolutt grense for hvor lav p-verdi vi skal kreve for å tørre å konkludere at det nok må være en sammenheng, det vil si at vi forkaster hypotesen om at det ikke er noen sammenheng. Vi oppgir derfor gjerne hvilken p-verdi vi finner i testene våre, enten med tall eller i noen tilfeller som et intervall:

– $0,05 < p < 0,10$ (angis med «+»)
– $0,01 < p < 0,05$ (angis med *)
– $0,001 < p < 0,01$ (angis med **)
– $p < 0,001$ (angis med ***)

Alle andre testsituasjoner analyseres i prinsippet på samme måte. Ofte ønsker vi å sammenlikne relativ relativ skår mellom to utvalg. Nullhypotesen «(Gjennomsnittsverdiene for) *rRelSc* er ikke forskjellig for spill der jeg ender i *StoneHouse* med 3 rom og 4 rom.» Antallet spill der jeg endte i 3 og 4 steinrom, er henholdsvis 117 og 181. rRelSc-verdiene er henholdsvis rRelSc = –0,702 og 1,518, altså en forskjell på 2,220 enheter, som svarer til over 2 poeng. Med et så stort materiale (antallet observasjoner har stor betydning for hvor sikre konklusjoner vi kan trekke) og så stor forskjell, er det ingen overraskelse at vi finner p = 0,0026 for **toutvalgs t-testen** av denne hypotesen. Det er bare 0,26 % sjanse for at en så stor forskjell i relativ relativ skår bare skyldes tilfeldigheter og vi konkluderer derfor med at det jevnt over gir klart bedre resultat med 4 enn med 3 *StoneHouse Rooms*.

Det er også mulig å teste hypoteser om forskjell mellom flere enn to utvalg, som for eksempel: ««(Gjennomsnittsverdiene for) *rRelSc* er ikke forskjellig mellom spill der jeg ender i *Wooden Hut*, *Clay Hut* og *StoneHouse*.» Da bruker vi **variansanalyse** (ANOVA), teststatistikken kalles *F*, som svarer til en p-verdi. I mine spill finner vi gjennomsnittlige rRelSc-verdier på henholdsvis –1.358, –1,108 og 0,878 for hus i de tre materialene. Det svarer til p = 0,00006, som viser at det er en helt klar forskjell mellom resultatene for hus bygd i ulike materialer.

For *kontinuerlige variabler*, som for eksempel rRelSc, og *tellevariabler* som for eksempel antall *Actions*, kan vi være interessert i å teste hypotesen «Det er ingen samvariasjon mellom de to variablene». Det kan vi gjøre ved å beregne **korrelasjon**skoeffisienten mellom dem. Korrelasjonskoeffisienten *r* er et tall mellom –1 og +1. Verdien *r* = 0 betyr at det ikke er noen samvariasjon (som nullhypotesen foreskriver), verdien *r* = +1 betyr at det er en perfekt lineær sammenheng mellom variablene. Da kan verdiene for den ene variabelen y uttrykkes som en funksjon av den andre variabelen, x,

$$y = ax + b,$$

der a og b er konstanter og a er positiv. Dersom a er negativ, er $r = -1$. Til hver korrelasjonskoeffisient hører en p-verdi for test av den nøytrale nullhypotesen om at det ikke er noen sammenheng mellom de to variablene. Som for t-testen, avhenger p-verdien av antallet observasjoner, n. r-verdien sier imidlertid bare noe om hvor sterkt variablene samvarierer, og påvirkes ikke av n. For *rRelSc* og *Actions* i hele materialet (n = 742) finner vi $r = 0{,}331$. Det svarer til en p-verdi < 0,000000001. Det er som forventet; det er klart resultatet blir bedre om man får mange *Actions!*

t- og r-verdiene kan «oversettes» til et uttrykk for hvor stor andel av variasjonen i en variabel, for eksempel *rRelSc*, som er «forklart» av en annen variabel, f.eks. antallet steinhusrom. Kvardratet av korrelasjonskoeffisienten (r^2) er et uttrykk for andelen forklart variasjon i en korrelasjonsanalyse. I vårt eksempel finner vi:

$$r^2 = 0{,}331^2 = 0{,}331*0{,}331 = 0{,}110 = 11{,}0\ \%$$

Vi sier at antallet *Actions* forklarer 11 % av variasjonen i relativ relativ skår.

WinF er en frekvensvariabel, som krever andre testmetoder. I statistikksammenheng ser vi på seiersfrekvensen som et resultat av n uavhengige «forsøk» (der hvert spill er ett «forsøk»). Vi antar at sannsynligheten for seier er den samme i hvert spill. Hypotesene vi kan teste er paralleller til hypotesene vi testet for *rRelSc*: «Seiersfrekvensen *WinF* er den samme når jeg ender i *Wooden Rooms* som i et tilfeldig valgt spill» og «Seiersfrekvensen *WinF* er ikke forskjellig for spill der jeg ender i *StoneHouse* med 3 rom og 4 rom.» Fordi de statistiske egenskapene til en frekvens, som varierer mellom 0 og 1, skiller seg fra de statistiske egenskapene til en kontinuerlig variabel, må vi bruke andre tester. Parallellen til ettutvalgs t-test er **den eksakte binomialtesten**. Hypoteser som for kontinuerlige variabler analyseres med toutvalgs t-test og variansanalyse (som er to sider av samme sak), analyseres ved hjelp av **generaliserte lineære modeller (GLM) med antakelse om binomialfordelt feil**, også kalt logistisk regresjon. Også for *WinF* gir analysene informasjon om hvor stor andel av variasjonen som er forklart, uttrykt som D-enheter (D = *deviance*).

Den eksakte binomialtesten benyttes også til meta-analyse av analyseresultater; det vil si for å teste om antallet signifikante enkelttester er flere enn forventet dersom en serie nøytrale nullhypoteser var sanne. For eksempel brukes denne testen til å teste om det er flere kort (*Minors* eller *Occupations*) som har signifikant bedre rRelSc enn forventet. Gitt at vi setter ei grense på p = 0.1 i tosidige tester, vil vi forvente at 1 av 20 tester av enkeltkort har signifikant positiv og 1 av 20 tester har signifikant negativ effekt. Da kan det observerte antallet signifikante tester sammenliknes med det forventete antallet, gitt antall tester.

Kort oppsummert: Analyser med *rRelSc* (relativ relativ skår, heretter: rr-skår) som uttrykk for hvor godt spillresultatet var, er gjort med ett- og to-utvalgs t-tester, variansanalyse og korrelasjonsanalyse, analyser med *WinF* (seiersfrekvens) som mål på hvor godt resultatet ble, er utført med den eksakte binomialtesten (ett utvalg) og GLM med binomal feil (logistisk regresjon).

Alle analyser er utført med programmet R, versjon 4.3.1.

Betydningen av startposisjon

Tabell 2 viser at seiersfrekvensen er lavest i startposisjon 1 og høyest i startposisjon 3, mens relativ relativ skår er høyest i startposisjon 2. Merk at de aller fleste av spillene er spilt med 1 *Food* i kompensasjon til 4.-spiller i runde 5. Resultatene gir i hvert fall ingen indikasjoner på at *Food*-kompensasjonen er urimelig, verken at den er unødvendig eller at den gir en uforholdsmessig stor fordel.

Tabell 2. Fordeling av plasseringer og rr-skår i mine spill (n = 742).

Startposisjon	1	2	3	4
Seiersfrekvens	0,186	0,226	0,242	0,221
rr-Skår, gj.sn.	–0,033	0,087	–0,007	–0,020
Antall spill (n)	167	168	190	217

Et interessant resultat (i parentes bemerket) er at det ser ut til at jeg har hatt uflaks med startposisjonstrekningen; antallet spill som 4. og 3.-spiller er mye høyere enn forventet. *Food*-kompensasjonen ser ut til å ha vært god å ha for meg. Om det er best å være 2.- eller 3.-spiller er avhengig av hvilket mål vi legger vekt på; resultatene for rr-Skår og seiersfrekvens er motstridende. Å starte er uansett minst gunstig, sannsynligvis på grunn av at 1 *Food* mindre ikke oppveies av fordelen ved å kunne velge fritt blant alle *Actions*. Kanskje er det også en ulempe at man lett kommer bak i køen fordi 2.- eller 3.-spilleren tar *StartingPlayer* i runde 1.

Variansanalyse for test av hypotesten «(Gjennomsnittlig) *rRelSc* er ikke forskjellig mellom startposisjonene» har p = 0,9001, som viser at det ikke er grunnlag for å forkaste hypotesen. Den tilsvarende GLM-testen for *WinF* har p = 0.6228. Forskjellene mellom startposisjonene er altså ikke store nok til at det er grunnlag for å vurdere endringer i kompensasjonen.

Konklusjon: Startposisjonen har liten betydning for plasseringen.

Betydningen av når FG-*Action*kortet åpnes og min plassering i FG-køen

Tabell 3 viser, som forventet, at jeg er hissig på FamilyGrowth (heretter: FG). Jeg får første FG nesten tre ganger så ofte som jeg får FG tre runder etter at FG-*Action*kortet ble åpnet (men merk at det er mange kort som gjør at flere får FG i samme runde og at dette ikke er tatt hensyn til her). Også som forventet er det klare sammenhenger mellom når FG-*Action*kortet ble åpnet, min plassering i FG-køen og rr-Skår. Når jeg får første FG er rr-Skåren signifikant positiv uansett når FG kommer, men den er klart mest positiv når FG kommer i runde 5. Forskjellen på 3,41 enheter utgjør en forskjell på nær en hel plassering!

Tabell 3. Sammenhenger mellom når FG-*Action*kortet ble åpnet, min plassering i FG-køen og rr-Skår (n = 709). Antall spill i parentes. Fargete celler angir signifikante ettutvalgs t-tester; oransje celler angir positiv sammenheng og blå celler negativ sammenheng. Sterk farge angir tester med p < 0,05; svak farge angir 0,05 < p < 0,10.

	Min plassering i FG-køen (antall runder etter at FG-kortet kom at jeg fikk 1. FG)				
Runde FG kom	0	1	2	3	4+
5	3,41 (89)	–0,29 (46)	–1,98 (44)	–2,12 (28)	–3,52 (30)
6	2,13 (77)	–0,01 (50)	–0,63 (50)	–0,72 (23)	–0,97 (19)
7	1,26 (74)	0,71 (72)	–0,58 (56)	–3,61 (37)	0,66 (14)
n	240	168	150	88	63

Variansanalyse for test av hypotesten «(Gjennomsnittlig) *rRelSc* er ikke avhengig av når FG-*Action*kortet åpnes» har p = 0,6970. Det viser at det ikke spiller noen rolle for sluttresultatet når FG kommer, dersom vi ikke tar hensyn til når jeg fikk FG. Tabellen viser at årsaken til det er at den positive effekten av å få 1. FG når FG kommer i runde 5 oppveies av den negative effekten av å få FG i runde 7 eller 8 når FG kommer i runde 5! En toveis ANCOVA der vi inkluderer når FG-*Action*kortet ble åpnet som kategorisk variabel og min plassering i FG-køen som kontinuerlig variabel, med interaksjoner (det vil si åpning for at sammenhengen mellom når FG kom og min plassering varierer på forskjellige måter avhengig av når FG kom), viser at effekten av min plassering i FG-køen er svært sterkt signifikant (p << 0,001; forklarer 7,0 % av variasjonen i rr-Skår), og at det også er signifikant forskjellige sammenhenger avhengig av når FG-kortet åpnes:

FG-*Action* åpnet i runde 5: rrSkår = –1,79*FGplass + 2,71
FG-*Action* åpnet i runde 6: rrSkår = –0,87*FGplass + 1,61
FG-*Action* åpnet i runde 7: rrSkår = –0,98*FGplass + 1,31

Det er særlig verd å merke seg at det er klart mindre gunstig å få 2. eller 3. FG når FG kommer i runde 5 enn når FG kommer i runde 6. Når FG kommer i runde 7, spiller det liten rolle om jeg får 1. eller 2. FG. Særlig ugunstig er det å få FG i runde 10 når FG kommer i runde 7; da utgjør forskjellen i rr-Skår nesten én hel plassering! Tallene for FG 4 eller flere runder etter at FG-*Action*kortet ble åpnet, er koblet til spesielle strategier som f.eks. **Cowboy&Mother**, og spill der jeg av en eller annen grunn ikke klarer å holde på plassen i FG-køen. Interessant nok viser resultatene særlig dårlig resultat for disse strategiene når FG kommer i runde 5.

Tabell 4 viser tilsvarede relasjoner for seiersfrekvens (*WinF*). Kommer FG i runde 5 og jeg får første FG nær dobles sjansen for seier, til over 40 %! Merk at det er en del uregelmessigheter i sammenhengen mellom rr-Skår og seriersfrekvens. Det skyldes først og fremst at antallet spill (og antallet seiere) i hver kombinasjon er lav, slik at seiersfrekvensen i betydelig grad er styrt av tilfeldigheter.

Tabell 4. Sammenhenger mellom når FG-*Action*kortet ble åpnet, min plassering i FG-køen og *WinF* (n = 709). Antall seiere og antall spill er gitt i parentes. Fargete celler angir signifikante ettutvalgs t-tester; oransje celler angir positiv sammenheng og blå celler negativ sammenheng. Sterk farge angir tester med p < 0,05; svak farge angir 0,05 < p < 0,10.

	Min plassering i FG-køen (antall runder etter at FG-kortet kom at jeg fikk 1. FG)				
Runde FG kom	0	1	2	3	4+
5	0,404 (36/89)	0,261 (12/46)	0,182 (8/44)	0,143 (4/28)	0,100 (3/30)
6	0,351 (27/77)	0,200 (10/50)	0,140 (7/50)	0,174 (4/23)	0,105 (2/19)
7	0,243 (18/74)	0,167 (12/72)	0,143 (8/56)	0,108 (4/37)	0,286 (4/14)
n	240	168	150	88	63

Konklusjoner:
– Sannsynligheten for seier avhenger *ikke* av når FG-*Action*kortet åpnes
– Når FG kommer i runde 5 og jeg får første FG, er sjansen for at jeg vinner spillet nær dobbelt så stor som i et gjennomsnittsspill
– 1. FG gir økt sannsynlighet for godt resultat også når FG kommer i runde 6, og litt bedre enn gjennomsnittlig resultat når FG kommer i runde 7.
– Særlig ugunstige kombinasjoner av når FG kommer og min 1. FG synes å være når FG kommer i runde 5 og jeg får 2. eller 3. FG. At FG kommer i runde 7 og jeg får 1. FG i runde 10 er særlig ugunstig. Da er sannsynligheten for seier halvert sammenliknet med et gjennomsnittsspill.
– Strategier uten FG (**Cowboy&Mother** etc.), uten å bygge rom etc., gir økt vinnersannsynlighet dersom, og bare dersom, FG kommer i runde 7

Sammenhenger mellom strategivalg og resultat

Antall poeng i de ulike skåringskategoriene kan ses på som et uttrykk for strategivalg. Når poengene telles opp til slutt, er alle poeng like mye verdt. Men det betyr ikke at alle poeng er like lette å få, og det betyr heller ikke at alle poeng gir samme sjanse for å vinne spillet. Datamaterialet gir god mulighet for å teste ulike strategier. Tabell 5 viser sammenhenger mellom antall poeng skåret i ulike skåringskategorier og sluttresultatet, målt som rr-Skår.

Tabell 5. Sammenhenger mellom rr-skår og gjennomsnittlig antall poeng i hver skåringskategori (Gj.sn. poeng), analysert med lineær regresjonsanalyse. Koeffisient og Intercept er parameterestimater for regresjonslikningen rr-Skår = Koeffisient * Poeng + Intercept og r^2 angir andel av den totale variasjonen i rr-Skår som forklares av variabelen. Variablene er sortert etter avtakende verdi for koeffisienten (slik at variablene med størst betydning for resultatet står øverst i tabellen). Alle sammenhenger er sterkt signifikante (p << 0,001) bortsett fra *BeggingCards* som har *p* = 0.0352 (jeg fikk *BeggingCards* bare i 6 spill).

Variabel	Gj.sn. poeng	Koeffisient	Intercept	r^2
BeggingCards	–0,02	1,79	0,05	0,0046
Pasture	2,96	1,31	–3,87	0,0498
Fields	3,00	1,17	–3,52	0,0502
UnusedSpaces	–0,86	1,14	0,99	0,0645
WildBoar	1,18	1,00	–1,17	0,0463
Sheep	0,93	0,92	–0,85	0,0316
Cattle	1,37	0,82	–1,11	0,0453
Vegetable	2,34	0,82	–1,91	0,0391
Stable	0,54	0,79	–0,43	0,0201
Grain	2,00	0,67	–1,34	0,0281
Peeps	13,49	0,65	–8,78	0,0484
VictoryPoints	5,63	0,63	–3,57	0,0711
Rooms	5,79	0,47	–2,70	0,0403
BonusPoints	4,71	0,45	–2,13	0,0817

Resultatene indikerer at nest etter å unngå *BeggingCards*, er det som gir best uttelling å få mange *Pastures* og *Fields*, og å unngå *UnusedSpaces*. Alle disse skåringskategoriene har en koeffisient på over 1, det vil si at ett poeng skåret i en av disse kategoriene gir et utslag på sluttresultatet som er større enn ett poeng. Det gir mening fordi *Pastures* og *Fields* er forutsetningen for å skåre poeng i andre kategorier, og *UnusedSpaces* først og fremst gjenspeiler summen av *Pastures* og *Fields*. Ett poeng på *goods* (uavhengig av hvilken *good*) har større effekt på rr-Skår enn ett poeng på *Peeps*, *Rooms*, *VictoryPoints* og *BonusPoints*. Det skyldes sannsynligvis en kombinasjon av to ting: *Goods* kan konverteres til *Food* og de kan oppformeres til mer *Food* eller flere poeng, mens det koster å skaffe seg *VictoryPoints*, *BonusPoints* og *Rooms*. *Peeps* har en kostnad både i form av *Rooms* som må bygges og *feeding*. Merk også at dyr (og *Veggies*) gir større uttelling per poeng enn *Grain*, sannsynligvis fordi dyr, når de først er anskaffet, formerer seg og «avler» nye poeng, og fordi *Veggies* gir ett poeng per enhet. Resultatet indikerer også at *Stables* kanskje er mer verdifulle enn deres status, som jeg oppfatter som ganske lav.

Konklusjon: Poeng for *Pastures* og *Fields* er særlig verdifulle fordi de gir mulighet for å skåre poeng i andre skåringskategorier.

Strategivalg innebærer mye mer enn hvordan mine 43,06 poeng i gjennomsnitt fordeler seg på ulike skåringskategorier. Er det viktig å få mange *Actions*, spille mange kort og/eller kjøpe *majors*? Datamaterialet gir mulighet for å analysere dette med samme metodikk som for skåringskategoriene. Resultatene er oppsummert i Tabell 6.

Tabell 6. Sammenhenger mellom rr-skår og antall *Actions* og antall spilte kort i ulike kategorier, analysert med lineær regresjonsanalyse. Koeffisient og Intercept er parameterestimater for regresjonslikningen rr-Skår = Koeffisient * Poeng + Intercept og r^2 angir andel av den totale variasjonen i rr-Skår som forklares av variabelen. Variablene er sortert etter avtakende verdi for koeffisienten (slik at variablene med størst betydning for resultatet står øverst i tabellen). Alle sammenhenger er sterkt signifikante (p << 0,001). Merk at antallet *minors* ikke inkluderer *pass-on minors*.

Variabel	Gj.sn. antall	Koeffisient	Intercept	r^2
Actions	37,60	0,58	–21,64	0,1083
Majors	2,87	1,22	–2,29	0,0308
Minors	4,25	0,99	–4,21	0,0398
Occupations	4,61	0,64	–2,95	0,0173

Analysen viser, ikke overraskende, at jeg gjør det best når jeg får mange *Actions*. Det er jo en logisk følge av at det er desidert mest gunstig å få FG i runde 5, og lite gunstig å få sein FG. Resultatene viser imidlertid også at det i det store og hele er gunstig å bruke *Actions* på å spille kort og kjøpe *Majors.* At rr-Skår øker mest per *Major* som blir kjøpt, er heller ingen overraskelse fordi mange *Majors* gir mer enn ett poeng og er dyre å kjøpe. Det som kanskje er mer overraskende, i hvert fall ved første blikk, er at uttellingen per *Minor* er større enn uttellingen per *Occupation.* Det kan skyldes at kostnaden ved å spille *Occupations* (en hel *Action* og inntil 2 *Food*) er større enn kostnaden ved å spille *Minors*, som oftest spilles som del av en *dobbeltaction*. Samtidig åpner det for at *Minors* kan være viktigere for sluttresultatet enn man har lett for å tro. I et gjennomsnittsspill spilles 4,25 *Minors* som hver gir en uttelling på omtrent 1 rr-Skårenhet. Sammenliknet med et spill med verdiløse *Minors* eller uten spilte *Minors*, svarer 4,25 * 1 rr-Skårenhet omtrent til en hel plassering!

Antall *Actions* er naturligvis avhengig av når FG-*Action*kortet åpnes, men i mindre grad enn man kanskje skulle trodd; 39,29 når FG kommer i runde 5 (n = 237), 37,62 når FG kommer i runde 6 (n = 219) og 36,16 (n = 253) når FG kommer i runde 7. At forskjellene ikke er større, kan skyldes to ting (eller en kombinasjon av dem): at 1. FG betyr lengre bak i køen for 2. (og 3.) FG fordi de som kommer lengre bak i køen bygger 2 (eller flere rom), og at det faktisk ikke er mulig å få flere enn 5 *peeps*. Globuskortene åpner enda flere muligheter for at flere får FG i samme runde.

Det er stor forskjell i kvalitet mellom kortene vi drafter. Derfor kan det være interessant å se på om det er sammenhenger mellom resultatet og hvor mange kort av en gitt type som er spilt/kjøpt. Tabellene 7–9 oppsummerer resultatene.

Tabell 7. Sammenhenger mellom rr-Skår, WinF og antall *Majors* som er kjøpt. «Test» viser resultater av parvise to-utvalgs t-tester for hypotesen «rr-Skåren er ikke signifikant forskjellig mellom de to kategoriene (av antall *Majors*) som blir sammenliknet» og tilsvarende GLM-analyse for WinF. Kategorier som har minst én bokstav felles er *ikke* signifikant forskjellig på nivå α = 0,05; det vil si at testen har en p-verdi < 0,05. Kategorier med for lavt antall til at testing er mulig, er grået ut.

Antall *Majors*	0	1	2	3	4	5	6
n	18	262	297	132	28	4	1
rr-Skår	–2,596	–1,032	0,118	1,464	2,755	3,273	1,210
Test av rr-Skår	a	ab	c	d	cd		
WinF	0,056	0,168	0,215	0,326	0,357	0,250	0
Test av WinF	a	ab	bc	d	cd		

Tabell 7 viser en helt jevn økning i rr-Skår og økning i *WinF* per *Major* som blir kjøpt opp til 4. Det underbygger at det er viktig å kjøpe *Majors* for å få et godt resultat.

Tabell 8. Sammenhenger mellom rr-Skår, WinF og antall *Minors* spilt. «Test» viser resultater av parvise to-utvalgs t-tester for hypotesen «rr-Skåren er ikke signifikant forskjellig mellom de to kategoriene (av antall *Minors*) som blir sammenliknet» og tilsvarende GLM-analyse for WinF. Kategorier som har minst én bokstav felles er ikke signifikant forskjellig på nivå α = 0,05; det vil si at testen har en p-verdi < 0,05. Kategorier med for lavt antall til at testing er mulig, er grået ut.

Antall *Minors*	1	2	3	4	5	6	7	8
n	5	63	135	221	203	87	25	3
rr-Skår	–0,740	–2,080	–1,681	–0,074	0,749	1,653	3,250	1,127
Test av rr-Skår		a	a	b	bc	c	c	
WinF	0,400	0,111	0,119	0,217	0,241	0,333	0,440	0,333
Test av WinF		ab	b	ac	cd	de	e	

Tabell 8 viser et stort sprang i rr-Skår fra 3 til 4 spilte *Minors*, og at det ikke er noen tendens til at den 6. og 7. spilte *Minor*en har mindre betydning for resultatet enn de første spilte minorene. Forskjellen i rr-Skår mellom 2 og 7 spilte *Minors* er over 5 enheter- som svarer til en hel plassering, og seiersfrekvensen dobler seg fra 2 til 4 og én gang til fra 4 til 7 spilte *Minors*!

Tabell 9. Sammenhenger mellom rr-Skår, WinF og antall *Occupations* spilt. «Test» viser resultater av parvise to-utvalgs t-tester for hypotesen «rr-Skåren er ikke signifikant forskjellig mellom de to kategoriene (av antall *Occupations*) som blir sammenliknet» og tilsvarende GLM-analyse for WinF. Kategorier som har minst én bokstav felles er ikke signifikant forskjellig på nivå α = 0,05; det vil si at testen har en p-verdi < 0,05. Kategorier med for lavt antall til at testing er mulig, er grået ut.

Antall *Occupations*	1	2	3	4	5	6	7
n	3	30	120	209	199	104	77
rr-Skår	–6,707	–1,348	–0,867	–0,405	0,068	1,306	1,330
Test av rr-Skår		ab	a	a	abc	c	bc
WinF	0	0,100	0,167	0,206	0,196	0,288	0,364
Test av WinF		a	a	ab	ab	bc	c

Også for *Occupations* er det en relativt jevn økning i rr-Skår med antall spilte kort, med unntak for et stort sprang fra 5 til 6 spilte kort (p = 0.093). Dette spranget faller sammen med stort sprang i *WinF*; forskjellen mellom 3, 4 og 5 spilte *Occupations* er liten, men den 6. og 7. *Occupations* blir tydeligvis ikke spilt med mindre det er spesielle grunner for å gjøre det.

Konklusjon:
– Sjansen for et godt resultat øker relativt jevnt med økende antall *Actions* og spilte kort av hver type.
– Resultatene antyder at gode *Minors* er særlig verdifulle fordi de spilles uten *Action*-kostnad, og at det av samme grunn er viktig å være kritisk med hensyn til hvor mange og hvilke *Occupations* som blir spilt.

Stats for *Majors*

Hvilke majors bør jeg kjøpe? Tester av rr-Skår og forskjell i seiersfrekvens (WinF), det vil si differansen mellom WinF med og uten den aktuelle majoren, kan gi svar på det. Resultater av testene for *Majors* er oppsummert i Tabell 10.

Tabell 10. Forskjell i rr-Skår og seiersfrekvens mellom spill med og uten de enkelte *Majors* på hånda. Spill der en *Major* er tilbakelevert til fordel for en oppgradering er talt med blant spill med *Majoren*. Positive forskjeller betyr at rr-Skår og/eller Win% er høyere med enn uten *Majoren*. Fargete celler angir signifikante ettutvalgs t-tester; oransje celler angir positiv sammenheng og blå celler negativ sammenheng. Sterk farge angir tester med p < 0,05; svak farge angir 0,05 < p < 0,10.

Major	n	dRelSc, forskjell	p	Win%, forskjell	p
Fireplace	130	–0.158	0.7933	0.004	0.9180
CookingH	392	0.020	0.9652	–0.022	0.4650
Well	146	1.930	0.0011	0.085	0.0277
ClayO	140	0.538	0.3882	0.055	0.1580
StoneO	144	0.681	0.2543	0.038	0,3280
Joinery	108	2.181	0.0010	0.133	0.0023
Pottery	127	0.563	0.3522	0.039	0.3350
BMW	167	1.082	0.0636	0.048	0.1810

Tabellen viser at de to *Majors* som har klart størt positiv effelt på resultatet, er **Joinery** og **Well.** Resultatet for **Joinery**, som er klart signifikant, er overraskende. Et spill der jeg kjøper **Joinery** har en seierssannsynlighet på 33,3 %, mot 20,0 % uten **Joinery**! Tilsvarende tall for **Well** er 28,8 % og 20,3%. Også BMW, Pottery og ovnene har en positiv effekt på seiersfrekvensen, men den er betydelig svakere. **Fireplace** og **CookingHearth** har ingen effekt på seierssannsynligheten. Dette indikerer at det er viktig å kjøpe Well, gjerne også andre *Majors*, og at strategier som ender med én *Major* som er **Fireplace eller CookingHearth** ikke bidrar positivt til et godt sluttresultat. Hvorfor **Joinery** gir et så positivt resultat, er ikke opplagt og bør undersøkes nærmere.

Konklusjon:
– Kjøp av **Well** innebærer ofte et godt sluttresultat
– Også kjøp av **Joinery**, **BMW** og **Pottery** kjennetegner spill med et godt resultat
– Strategier som ender med **Fireplace eller CookingHearth** som eneste *Major* gir ofte et dårligere sluttresultat

Hvor mange rom bør bygges, og i hvilket materiale?

En viktig strategisk avgjørelse er om man skal gå til *Clay* eller *Stone*. Datamaterialet gir klare svar på hva som i det store og hele gir best resultat, og hvor mange rom det gir uttelling å bygge.

Tabell 11. rr-Skår og seiersfrekvens i spill som ender med hus i ulike materialer. «Test» viser resultater av parvise to-utvalgs t-tester for hypotesen «rr-Skåren/*WinF* er ikke signifikant forskjellig mellom de to kategoriene (av antall *Occupations*) som blir sammenliknet». Kategorier som har minst én bokstav felles er ikke signifikant forskjellig på nivå α = 0,05; det vil si at testen har en p-verdi < 0,05. kategorier med for lavt antall til at testing er mulig, er grået ut. Resultatene av tester for rrSkår og *WinF* er sammenfallende.

Materiale	*Wood*	*Clay*	*Stone*
rr-Skår	–1,358	–1,108	0,878
WinF	0.162	0,158	0,267
Tester	a	a	b
n	37	285	420

Tabell 11 viser at resultatet er minimalt forskjellig enten man ender i *Wood* eller *Clay*, mens det er en stor og signifikant forskjell mellom å ende i *Clay* og å ende i *Stone*. Denne forskjellen kommer sterkest til uttrykk i *WinF*, som er ca. 10 prosentpoeng høyere for *Stone* enn for de andre materialene. Grunnen til at *Wood* kommer godt ut sammenliknet med *Clay*, er nok at spill som har endt i *WoodRooms* ofte har kompensert poengene med **WoodenHutBuilder** eller at det har vært 2-romsstrategier (f.eks. med **Cowboy&Mother**, der renovering ikke har vært prioritert på grunn av at uttellingen ikke har gjort det regningssvarede, med få *Actions*.

Tabellene 12 og 13 viser resultater for hus i ulike materialer etter splitting på antall rom (jeg har ikke data for antall rom i *Wooden House*).

Tabell 12. Sammenhenger mellom materialet huset er bygd i, antall rom og rr-Skår (n = 742). Antall spill er gitt i parentes. Fargete celler angir signifikante ettutvalgs t-tester; oransje celler angir positiv sammenheng og blå celler negativ sammenheng. Sterk farge angir tester med p < 0,05; svak farge angir 0,05 < p < 0,10.

	Antall rom				
Materiale	2	3	4	5	6
Wood			–1,358 (37)		
Clay	–0,490 (10)	–2,086 (98)	–0,897 (135)	0,347 (42)	–
Stone	–0,559 (26)	–0,702 (117)	1,518 (181)	1,861 (86)	2,060 (10)
n	36	215	316	128	10

Tabell 13. Sammenhenger mellom materialet huset er bygd i, antall rom og seiersfrekvens (n = 742). Antall seiere og spill er gitt i parentes. Fargete celler angir signifikante eksakte tester; oransje celler angir positiv sammenheng og blå celler negativ sammenheng. Sterk farge angir tester med p < 0,05; svak farge angir 0,05 < p < 0,10.

	Antall rom				
Materiale	2	3	4	5	6
Wood			0,162 (6/37)		
Clay	0.200 (2/10)	0.122 (12/98)	0.156 (21/135)	0.238 (10/42)	–
Stone	0.192 (5/26)	0.231 (27/117)	0.287 (52/181)	0.302 (26/86)	0.200 (2/10)
n	36	215	316	128	10

Resultatene levner ingen tvil om at det er ugunstig å ende i *Clay* og at det er særlig ugunstig å ende i 3 *Clay Rooms*. Også 4 *Clay Rooms* er signifikant mindre gunstig enn gjennomsnittet. Sjøl om ikke 3 *Stone Rooms* gir et bedre resultat enn gjennomsnittet, er forskjellen mellom 3 *Clay Rooms* og 3 *Stone Rooms* (en økning i seiersfrekvens på over 10 prosentpoeng) så stor at det ikke er noen tvil om at det generelt er lønnsomt å renovere til stein, også når huset bare har tre rom. For hus med 4 rom er forskjellen mellom hus i *Clay* og i *Stone*, naturlig nok, enda større.

Konklusjon:
– Det gir gjennomgående langt bedre resultater med *Stone House* enn *Clay Hut*, og *Stone House* bør være målet i alle spill der det ikke er spesielle grunner til å forbli i *Clay.*
– Det gir god uttelling å bygge det 3., 4. og 5. rommet dersom huset renoveres til *Stone*.

Stats for *Occupations* og *Minors*

Materialet er akkurat stort nok til å gi en indikasjon på hvilke kort som øker og hvilke kort som reduserer sannsynligheten for at jeg gjør et godt spill, dersom kortet spilles. Sånn som dataene mine er organisert, er det et tidkrevende arbeid å sortere ut hvilke kort som er spilt i hvert enkelt spill. Jeg har derfor ikke systematisk testet alle kort, men valgt ut kort jeg er spesielt interessert i, og som jeg tror jeg har spilt nok ganger til at materialet kan si noe om mine resultater i spill der de er spilt. Når gjennomsnittlig antall kort spilt er ca. 9 per spill og totalt antall kort er ca. 550 for EIKWmFr og ca. 1 000 med globuskortene, vil et gjennomsnittskort bli draftet og spilt av meg i hvert 9/550 – 9/1000. spill, dvs. hvert 60. til 110. spill. Med

742 spilte spill, vil dermed et gjennomsnittskort bli spilt av meg mellom 7 og 12 ganger; EIK-kort naturligvis i flere spill fordi disse kortene har vært med i alle spillene, også spill med færre *Decks* (materialet omfatter også en god del EIK-spill). Omtrent 10 spill er altfor få til å si noe om kortet bidrar positivt (eller negativt) til sluttresultatet. For å gi klare resultater, bør kortet være spilt i ca. 20 eller flere spill. Tabellene 14 og 15 viser resultatene for henholdsvis *Minors* og *Occupations.*

Tabell 14. Forskjell i rr-Skår og seiersfrekvens mellom spill med og uten de enkelte *Minors* på hånda. Spill der en *Minor* er tilbakelevert, f.eks. pga. at **MadonnaStatue** ble spilt, er talt med blant spill med *Minoren*. Positive forskjeller betyr at rr-Skår og/eller *WinF* er høyere med enn uten *Majoren*. Fargete celler angir signifikante ettutvalgs t-tester; oransje celler angir positiv sammenheng og blå celler negativ sammenheng. Sterk farge angir tester med p < 0,05; svak farge angir 0,05 < p < 0,10. PWR = *Power statistic*, hentet fra ReactApp. n = 742.

Kort	PWR	n	rr-Skår, forskjell	p	*WinF*, forskjell	Antall seiere	p
AnimalPen	3,73	27	0,935	0,3302	0,041	7	0,6419
Axe	3,69	28	1,467	0,2731	0,032	7	0,6520
BarberShop	2,52	23	−0,842	0,4685	−0,047	4	0,8019
Bookshelf	3,87	26	−0,095	0,9472	0,012	6	0,8160
ClayRoof	3,98	27	1,550	0,2518	0,194	11	0,0324
ClaySupports	2,98	28	−0,541	0,6766	−0,006	6	1,0000
Downsizing	3,32	32	−1,678	0,1977	0,032	8	0,6705
DuckPond	4,52	24	0,881	0,3767	0,075	7	0,4573
FishTrap	3,50	37	0,921	0,3711	0,138	13	0,0711
ForestPasture	4,04	19	1,103	0,4771	−0,009	4	1,0000
FruitTree	4,07	19	2,021	0,1100	0,099	6	0,2810
GoosePond	4,09	27	2,696	0,0313	0,156	10	0,0651
GrainCart	2,58	31	0,343	0,7323	−0,061	5	0,5212
GrainElevator	1,97	28	0,437	0,7169	0,069	8	0,3684
Granary	3,13	15	3,885	0,0254	0,116	5	0,3450
GroundPickaxePlow	3,42	38	−1,030	0,3163	−0,038	7	0,6985
Harrow	1,18	32	−2,339	0,0570	−0,131	3	0,0906
HouseGoat	4,05	21	0,665	0,6279	0,068	6	0,4365
Ladder	1,90	32	−0,241	0,8513	−0,067	5	0,5223
LandingNet	3,01	38	1,785	0,1720	0,102	12	0,1690
Lasso	1,31	11	−2,029	0,4596	−0,038	2	0,4748
MadonnaStatue	2,90	15	2,388	0,1294	0,048	4	0,7542
MilkingStool	4,22	19	2,386	0,1287	−0,063	3	0,7814
Pelts	4,14	42	2,812	0,0152	0,120	14	0,0915
PlanterBox	4,16	47	−0,007	0,9946	0,015	11	0,8600
Quarry	4,56	22	−0,010	0,9942	0,055	6	0,6053
RidingPlow	3,48	39	0,029	0,9756	−0,015	8	1,0000
Sawhorse	3,03	51	0,059	0,9489	−0,025	10	0,8656
SleepingCorner	3,37	23	−1,627	0,0977	−0,182	1	0,0422
Spinney	2,89	39	0,850	0,3474	0,012	9	0,8473
SteelPlow	3,43	38	2,492	0,0209	−0,009	8	1,0000
SwanLake	3,94	17	1,520	0,3221	0,076	5	0,5559
SwingPlow	3,49	32	0,216	0,8455	−0,001	7	1,0000
TurnipField	3,88	20	4,439	0,0027	0,185	8	0,0603
WoodCart	2,66	62	−0,813	0,3864	−0,013	13	1,0000
WritingDesk	3,07	37	1,177	0,2793	0,053	10	0,4313

Den *Minoren* jeg har spilt desidert flest ganger, 62 (hvert ellevte spill), er **WoodCart.** Dessverre ser det ikke ut til at det har vært til særlig nytte; spill med denne minoren har litt lavere rr-Skår enn gjennomsnittet og ingen effekt på seiersfrekvensen. Min klart beste *Minor* er **TurnipField**. Spill med denne *Minoren* har nesten én hel plassering bedre resultat enn spill uten, og seiersfrekvensen er nesten dobbelt så høy! Resultatene viser dessuten at *Minors* med poeng, uten kostnad, gjennomgående har godt resultat og bidrar til å øke seierssannsynligheten: **GoosePond***,* **FruitTree** og **SwanLake** er eksempler i materialet, mens **DuckPond** kommer litt svakere ut (men også den med et positivt resultat). Resultatet for **Quarry** er svakere, kanskje fordi det er vanskelig å få den ut tidlig nok dersom man også skal få nytte av effekten (+ 3 *Stone* på *Daylabourer Action space*)?

Blant plogene står **SteelPlow** i en særstilling, mens det er minimal forskjell på **RidingPlow** og **SwingPlow**.

Kort som litt overraskende skiller seg positivt ut, er **Granary** og **LandingNet**. Resultatene minner om at **ClayRoof**, **Pelts** og **FishTrap** er gode kort. Kort som kommer overraskende dårlig ut, er **Bookshelf**, **PlanterBox**, **SleepingCorner** og **Downsizing**, mens resultatene bekrefter at **Harrow** er et svakt kort.

Av de 36 kortene i Tabell 14, har 7 en signifikant positiv effekt i ensidig test (p < 0.1 i tosidig test). Det er signifikant flere enn forventet (p = 0,0018 i eksakt binomialtest).

Korrelasjonen mellom rr-Skår og *Power* er r = 0,4919 (p = 0,0023), som viser at *Power* i det store og hele gir et godt inntrykk av hvor gode *Minors* er for meg.

Konklusjon:
– De beste *Minorene* er de som er lette å spille (ingen kostnad) og som gir 1 poeng og enkel effekt, gjerne *Food*

Tabell 15. Forskjell i rr-Skår og seiersfrekvens mellom spill med og uten de enkelte *Occupations* på hånda. Positive forskjeller betyr at rr-Skår og/eller *WinF* er høyere med enn uten *Majoren*. Fargete celler angir signifikante ettutvalgs t-tester; oransje celler angir positiv sammenheng og blå celler negativ sammenheng. Sterk farge angir tester med p < 0,05; svak farge angir 0,05 < p < 0,10. PWR = *Power statistic*, hentet fra ReactApp. n = 742.

Kort	PWR	n	rr-Skår, forskjell	p	*WinF*, forskjell	Antall seiere	p
AnimalBreeder	3,98	40	0,060	0,9576	0,032	10	0,7020
AnimalDealer	2,40	27	0,541	0,6549	–0,113	3	0,2442
ArtDirector	2,70	20	–1,201	0,3668	–0,072	3	0,5947
Baker	2,12	20	1,869	0,1032	0,020	4	1,0000
BargainingBaker	3,12	18	0,881	0,6209	0,116	6	0,2555
BerryPicker	1,29	24	0,730	0,5728	–0,012	5	1,0000
Boatswain	3,80	40	0,033	0,9761	0,053	11	0,4436
CharcoalBurner	4,63	28	0,855	0,4877	0,105	9	0,2502
Chief	3,81	14	1,084	0,5241	0,213	6	0,0969
Chief'sDaughter	3,74	17	1,747	0,2748	0,136	6	0,2356
ChurchWarden	3,52	25	4,891	0,0010	0,187	10	0,0486
CloisterDweller	3,93	16	1,966	0,1972	0,094	5	0,3682
Conservator	2,54	30	–0,130	0,9111	0,049	8	0,5116
Constable	3,03	24	–1,164	0,4430	0,075	7	0,4573
Cowboy&Mother	2,55	21	0,800	0,4247	0,019	5	0,7945
Dancer	1,48	18	–1,125	0,4421	0,002	4	1,0000
DebtCollector	3,60	11	–1,083	0,5945	0,146	4	0,2726
Diplomat	2,44	36	0,526	0,5686	0,091	11	0,2268
DoveHunter	2,94	34	–1,880	0,0634	–0,046	6	0,6803
Educator	4,20	22	–0,314	0,8330	0,008	5	1,0000
Engineer	3,63	42	–0,935	0,4245	–0,031	8	0,8520

FarmSteward	4,03	16	0,131	0,9354	−0,032	3	1,0000
FenceDealer	3,30	32	−0,870	0,4795	0,032	8	0,6705
FenceDeliveryman	2,01	20	−0,802	0,5887	−0,123	2	0,2813
FieldWarden	3,75	22	1,747	0,1041	−0,039	4	0,8013
FieldWatchman	4,35	21	−1,269	0,3564	−0,031	4	1,0000
GrainSpeculator	3,29	9	0,097	0,9724	0,002	2	1,0000
Greengrocer	2,12	32	−0,565	0,5461	−0,001	7	1,0000
Grocer	4,72	27	1,194	0,3257	0,118	9	0,1636
Guildmaster	3,20	11	0,094	0,9585	−0,038	2	1,0000
HeadOfTheFamily	3,11	22	0,207	0,8973	0,008	5	1,0000
HedgeKeeper	2,91	36	1,845	0,0622	0,002	8	1,0000
HillFarmer	2,63	21	−1,723	0,0817	−0,079	3	0,5978
Hoarder	2,71	42	0,682	0,5144	0,096	13	0,0686
HutBuilder	3,62	25	3,391	0,0171	0,228	11	0,0136
LadyInWaiting	3,27	22	3,457	0,0218	0,149	8	0,1202
Magician	2,62	11	0,371	0,8308	−0,038	2	0,4748
MarketWoman	3,82	34	1,642	0,1567	0,047	9	0,5348
Mason	4,18	23	−0,651	0,5113	−0,137	2	0,2036
MilkingHand	3,66	14	2,140	0,2585	−0,006	3	1,0000
Patron	3,92	33	0,431	0,6917	−0,071	5	0,4075
PerpStud	3,94	26	−0,344	0,8187	−0,029	5	0,3150
ReedCollector	3,10	14	−0,030	0,8695	0,068	4	0,5202
Reeve	3,55	37	0,907	0,3949	0,081	11	0,2391
Remodeler	3,01	29	−0,234	0,8346	−0,121	3	0,1771
SeasonalWorker	2,44	22	−0,643	0,7006	0,008	5	1,0000
SundayWorker	2,45	38	−0,850	0,2992	−0,038	7	0,6985
Thatcher	2,19	38	−0,629	0,5718	−0,092	5	0,2406
Tutor	3,23	29	2,001	0,2002	0,130	10	0,0686
WealthiestEuropean	2,55	12	2,434	0,0565	0,031	3	0,7334

For *Occupations* var den største overraskelsen de dårlige resultatene for tre av kortene som regnes som blant de aller beste, som draftes *by default* når de dukker opp i draften, og som «alltid» blir spilt: **Educator**, **FieldWatchman**, og **PerpetualStudent**. En rask titt på datagrunnlaget for *Power* i *ReactApp* viser at et fellestrekk for disse tre kortene er at de har høy *Power* fordi kortene alltid blir draftet og spilt, mens seierssannsynligheten er moderat, omkring 30. Det er kanskje verd å minne om at *Power* uttrykker den relative sannsynligheten for at et kort som blir delt ut, først blir draftet, dernest at blir spilt, og til sist ender opp på vinnerens bord. Det betyr at et kort kan ha høy *Power enten* fordi det har høy sannsynlighet for å bli draftet og spilt eller fordi det er stor sannsynlighet for at den som vinner, faktisk velger å spille det! De dårlige resultatene mine for disse kortene sier meg at dette er vanskelige kort å få til. Hvor ofte har jeg ikke stønnet over at **Educator** hjalp til å få ut dårlige *Occupations*, at jeg ikke fikk støtte til **FieldWatchman** og at **PerpetualStudent** aldri ga meg de *Occupations*-ene jeg trengte når jeg trengte å få dem ut? Jeg har også dårlige resultater for **Mason**, et kort som har høy *power* på grunn av høy seiersprosent. Årsaken er sannsynligvis at jeg strevde for mye med å få fire steinrom. Det er likevel verd å legge merke til at **Chief** og **Chief'sDaughter**, som favoriserer *Stone House*, har positiv effekt på resultatet. For **Constable** er resultatene motstridende; dårlig resultat for rr-Skår til tross for akseptabel seierssannsynlighet (7/24 = 29 %). Trolig skyldes det en del mislykkete forsøk på å spille kortet tidlig, som resulterte i at andre spillere fikk fem bonuspoeng uten at jeg greide det sjøl. **Constable** er altså et runde 14-kort! To andre, «vanskelige» kort jeg har nøytrale resultater for er **AnimalBreeder** og **Boatswain**.

Andre kort med høy *Power* får jeg bedre til. Mine beste *Occupations* er **ChurchWarden**, **HutBuilder** og **LadyInWaiting**. Rimelig brukbare resultater har jeg også for **Patron**, **Reeve** og **Tutor**, som hjelper til å få ut *Occupations*. Jeg merker meg også at en av de to *Occupations* jeg har spilt flest ganger (42), **Hoarder**, har positiv effekt på resultatet, mens det samme ikke kan sies om det andre favorittkortet mitt, **Engineer**, uten at jeg helt forstår dette: Det burde være bra å få *Food* for å spille *Improvements* når det å spille *Improvements* er bra! Det minner meg om at datagrunnlaget ikke er veldig stort, og at det faktisk er godt mulig at mye av det som ser ut som mønstre for enkeltkort, faktisk er resultat av tilfeldighetenes spill. Likevel merker jeg meg at mange gode kort, men ikke toppkort, med «sikker» og/eller enkel effekt, ser ut til å ha positiv effekt på resultatet mitt: **FieldWarden** og **MarketWoman** er eksempler på dette. Interessant er det også at **HedgeKeeper**, som støtter opp om tidlig *fencing*, påvirker resultatet positivt, mens **FenceDeliveryman**, som støtter sein *fencing* har negativ effekt. Det har også **FenceDealer**, som støtter *fencing*, men enten gir få *pastures* eller *unused spaces*.

Blant kort med middels *power* som jeg har godt resultat med, må nevnes **Baker** og **WealthiestEuropean**, mens en lang rekke middels kort som gir enkel *Food*, men som ikke bygger farm eller gir poeng, kommer ut som svake for meg: **Dancer**, **DoveHunter, HillFarmer**, **SeasonalWorker** og **SundayWorker**. Heller ikke *Reed*-rabatt ser ut til å være gunstig: **Thatcher** er et av mine dårligere kort, kanskje fordi det gjør det mindre attraktivt å ta RSF og dermed kjøpe *Improvements* som koster *Stone*.

Av de 50 kortene i Tabell 15, har 7 en signifikant positiv effekt i ensidig test ($p < 0.1$ i tosidig test). Det er signifikant flere enn forventet ($p = 0{,}0118$ i eksakt binomialtest).

Korrelasjonen mellom rr-Skår og *Power* er $r = 0{,}1880$, $p = 0{,}1911$, som viser at *Power* i det store og hele gir få holdepunkter for hvor gode *Occupations* er for meg.

Konklusjon:
– Mine resultater for *Occupations* samsvarer ikke godt med generelle *Stats* som for eksempel *Power*.
– Jeg har gode resultater for enkle, gode kort med «sikker» effekt (**ChurchWarden**, **HutBuilder**, **LadyInWaiting**, **MarketWoman** og **FieldWarden**), men dårlige resultater for vanskelige kort som **Educator**, **FieldWatchman** og **PerpetualStudent**.
– Kort som fremmer *Stone House* har generelt god effekt, mens kort som gir tidlig *Food* uten å bygge farm eller gi poeng, har mindre god effekt.

Sluttord

Jeg har lenge tenkt at det var på tide å analysere spillresultatene. Når jeg nå endelig har gjort det, er jeg glad for at jeg ventet så lenge. 742 spill høres mye ut, men for mange av spørsmålene jeg har ønsket å finne svar på, er konklusjonene fortsatt usikre på grunn av at materialet ikke er stort nok. Særlig gjelder det resultatene for enkeltkort, som må tas med en stor klype salt.

Analysene har gitt meg noen aha-opplevelser om egne spill. Likevel er det slett ikke sikkert at mine resultater, forklaringer og konklusjoner gjelder for andre! Noen resultater, som for eksempel sat **Joinery** er den beste *Majoren* i min hånd, minner meg igjen om at resultatene må tolkes med varsomhet. **Joinery** er ofte den siste *Majoren* som blir kjøpt, og kanskje ender den opp i min hånd i spill der jeg fikk 1. FG i runde 5 eller 6, mange *Actions* og ikke har noe bedre å gjøre i spillets sluttfase enn å kjøpe nettopp den? Det forteller ikke datamaterialet noe om.

Til tross for motforestillingene tror jeg likevel resultatene påviser en del sammenhenger med allmenn gyldighet, for eksempel: at 1 *Food* til 4. spiller er en godt balansert kompensasjon for ugunstig startposisjon; at det gjennomgående er gunstig å bygge rom først og få 1. FG; at det er svært ugunstig å sitte bakerst i FG-køen, særlig dersom FG kommer i runde 7; at *Minors* med poeng og enkelt, sikker effekt alltid er gode; og at det er gode grunner til å satse på å ende i steinhus!

Så lenge vi har et aktivt Agricola-miljø vil den røde dronningens dype innsikt fortsette å gjelde; den som blir stående stille og ikke utvikle sine ferdigheter vil sakke akterut på resultatlistene. Alle bidrag til å forklare resultatene mottas derfor med takk!