

Del 1: Flervalgsoppgaver

Sett kryss for det du mener er riktig svar rett på arket og lever denne delen sammen med besvarelsen av Del II. Husk å fylle inn kandidatnummer.

1. Hvilket av de følgende kan brukes som et synonym for begrepet «observerbar implikasjon»?

- Data
- Analyse
- Teori
- Hypotese
- Tolkning

2. Hva menes med at en kausalmodell er umettet?

- Minst én logisk tenkelig årsakspil er utelatt
- Modellen er lite kompleks
- Modellen er uferdig
- Årsakspilene i modellen går i bare en retning
- Modellen spesifiserer ikke sammenhengenes fortegn

3. Hvilket av følgende mål er IKKE et spredningsmål?

- Variasjonsbredde
- Standardavvik
- Kjikkvadrat
- Modalprosent
- Varians

4. Hvilket av følgende uttrykk angir en formel for å kalkulere t-verdien til en regresjonskoeffisient b ?

- $b/SE(b)$
- $SE(b)/b$
- $b*SE(b)$
- $b/SE(b)*N$
- $b*N/SE(b)$

5. Hva innebærer en intensiv forskningsstrategi?

- Datainnsamlingen er basert på feltarbeid
- Datainnsamlingen er basert på kvalitativt intervju materiale.
- Undersøkelsen inneholder mange enheter og få variabler
- Undersøkelsen inneholder få enheter og mange variabler
- Datainnsamlingen har foregått over et lengre tidsrom

6. Hvordan kan casestudier ifølge Andersen benyttes til tankeeksperiment?

- Case kan velges ut fra variasjon i variabler som forskeren er særlig interessert i.
- Analysen kan gjennomføres som teoretisk øvelse fra skrivebordet
- Bare en variabel tillates å variere
- Analysen inneholder alltid en kontrollgruppe
- Valg av case er definert av tidligere undersøkelser

7. Hva er fordelene med å sentrere variablene som inngår i et multiplikativt samspillsledd

- Sentrering gjør det mindre sannsynlig at estimatene i regresjonsmodellen er påvirket av utelatt variabelskjevhet
- Vi får mindre problemer med heteroskedastisitet
- Sentrering gjør det lettere å regne ut punktestimater for effekter på alle verdiene for uavhengige variable i en samspillsmodell
- Vi får typisk høyere multikolaritet etter sentrering
- Sentrering gjør det mulig å signifikant teste effekten av en variabel som inngår i samspillsleddet for gjennomsnittsverdien på den andre uavhengige variabelen

8. Vi kjører en regresjonsmodell med politiske holdninger som avhengig variabel og inntekt målt i tusen kroner og antall år med utdanning som uavhengige variable. Hvordan tolker vi regresjonskoeffisienten for inntekt?

- Effekten på politiske holdninger av å øke inntekt med tusen kroner for individer uten utdanning
- Effekten på politiske holdninger av å øke inntekt med tusen kroner
- Effekten på politiske holdninger av å øke inntekt med tusen kroner når vi holder utdanning konstant
- Effekten på politiske holdninger av å øke inntekt med én krone når vi holder utdanning konstant
- Effekten på politiske holdninger av å øke inntekt med én krone

9. Hvilket av følgende begreper refererer til kvaliteten på en operasjonell definisjon?

- Reliabilitet
- Definisjonsmessig sammenheng
- Ekstern validitet
- Definisjonsmessig validitet
- Intern validitet

10. Hva er et annet navn for økologisk feilslutning?

- Atomistisk feilslutning
- Aggregativ feilslutning
- Tidsfeilslutning
- Feilslutning som gjelder miljøpolitikk
- Feilslutning som gjelder miljøforskning

11. Hva slags data må forskeren ha for å kunne trekke slutninger om endring over tid?

- Data på intervallnivå
- Data på ordinalnivå
- Data på nominalnivå
- Synkrone data
- Diakrone data

12. Hva er en bakenforliggende variabel?

- En variabel som i en kausalmodell ligger forut for et bestemt variabelpar
- En variabel som er skjult av kausalmodellen
- En variabel som i modellen påvirker minst en annen variabel
- En variabel som er særlig sentral i forskerens analyse
- En variabel som er lite sentral i forskerens analyse

13. Konstantleddet i en bivariat regresjonsmodell er 5. Hva betyr dette?

- Konstantleddet er positivt og statistisk signifikant
- Konstantleddet er positivt og statistisk signifikant på 5 prosent nivå
- Det gjennomsnittlige restleddet i modellen er 5
- Den predikerte verdien på avhengig variabel er 5 når uavhengig variabel er 0
- Ingen av de andre svaralternativene er korrekte

14. Hva er forholdet mellom Kjikvadratet og F_i ?

- Kjikvadratet og F_i har ikke noe med hverandre å gjøre.
- Kjikvadratet er et normert mål som bygger på F_i
- F_i og Kjikvadratet uttrykker det samme
- F_i er et normert mål som bygger på Kjikvadratet
- Både F_i og Kjikvadratet er normerte mål som bygger på Cramers V

15. I en regresjonsmodell er restleddene til to typer observasjoner (A og B) negativt autokorrelerte. Hva forventer vi dersom restleddet til A viser seg å være negativt?

- Restleddet til B er negativt
- Restleddet til B er positivt
- Restleddet til B er mindre i absolutt størrelsen enn et gjennomsnittlig restledd i regresjonen
- Restleddet til B er større i absolutt størrelsen enn et gjennomsnittlig restledd i regresjonen
- Restleddet til B vil være signifikant positivt

16. Hvilken av formlene uttrykker en multivariat regresjonsmodell uten samspill?

- $Y_i = b_0 + b_1X_{1i} * b_2X_{2i} + e_i$
- $Y_i = b_0 + b_1 * X_{1i} + e_i$
- $Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + b_3(X_{1i} * X_{2i}) + e_i$
- $Y_i = b_0 + b_1X_{1i} + b_2X_{2i} + e_i$
- $Y_i = b_0 + b_1X_{1i} + e_i$

17. Hva gjør vi når vi bedriver diagnostikk på en regresjonsmodell?

- Vi signifikanstester alle de uavhengige variablene, én etter én, for å se om de bør inkluderes eller ekskluderes fra modellen
- Vi undersøker hvor sensitive eller robuste resultatene er for endringer i modellspefikasjonen
- Vi undersøker om variablene i regresjonsmodellen er målt på en reliabel og valid måte
- Vi undersøker om forutsetningene for modellen holder eller ikke
- Vi undersøker hvor mye variasjon i avhengig variabel modellen kan forklare

18. Standardfeilen til en regresjonskoeffisient er fire ganger så stor som størrelsen på regresjonskoeffisienten, og regresjonskoeffisienten har negativt fortegn. Hva er t-verdien til regresjonskoeffisienten?

- 0.25
- 4
- 0.25
- 4
- Det trengs flere opplysninger for å kunne kalkulere t-verdien

19. Hvilke tre elementer består en datamatrise av?

- Enheter, utvalg og verdier
- Signifikans, variabler og verdier
- Variabler, utvalg og verdier
- Enheter, variabler og verdier
- Signifikans, variabler og enheter

20. Hva er en strategisk test?

- En test der to plausible teorier gir helt identiske hypoteser om hvordan en empirisk sammenheng forventes å være
- En test av en hypotese som stammer fra en svært presis teori som en annen forsker har utviklet
- En test av en hypotese som er basert på grundige strategiske overveielser
- En test der to plausible teorier genererer vidt forskjellige hypoteser om hvordan en empirisk sammenheng forventes å være
- En test av en hypotese som stammer fra en svært presis teori, uavhengig av hvem som har utviklet den

21. Hva innebærer det at det er samspill mellom X og Z, når Y er avhengig variabel?

- At effekten av Z på X avhenger av verdien på Y
- At effekten av X på Z avhenger av verdien på Y
- At effekten av X på Y avhenger av verdien på Z
- At verdien på Y avhenger av effekten av Z på X
- At verdien på Y avhenger av effekten av X på Z

22. Vi har en variabel for religion med fire verdier: kristen, jøde, muslim, alle andre religioner. Vi estimerer effekten av religion på sannsynligheten for å stemme SV. Hva er den substansielle tolkningen av konstantleddet i den følgende regresjonslikningen: $Y_i = b_0 + b_1 * KRISTEN_i + b_2 * JØDE_i + b_3 * MUSLIM_i + e_i$

- Forventet sannsynlighet for å stemme SV for kristne
- Forventet sannsynlighet for å stemme SV for jøder
- Forventet sannsynlighet for å stemme SV for muslimer
- Forventet sannsynlighet for å stemme SV for alle andre religioner
- Konstantleddet har ingen substansielle tolkning her, siden uavhengig variabel er på nominalnivå.

23. Hva menes med ja-siing?

- At respondenter med sterke meninger er overrepresentert i undersøkelsen
- At spørsmålsformuleringene er ledende
- Respondenter sier seg enige i påstander de ikke har noen mening om
- At spørsmålene er ledende mot ja-svar
- Respondenter sier ja til å delta i undersøkelser

24. Vi har en regresjonsmodell med samspillsledd der $Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + b_3 (X_{1i} * X_{2i}) + e_i$. Vi finner at b_2 er statistisk signifikant på 5% nivå. Hva betyr dette?

- Sammenhengen mellom X_2 og Y er signifikant forskjellig fra null
- Effekten av X_2 på Y er signifikant forskjellig fra null
- Sammenhengen mellom X_2 og Y er signifikant forskjellig fra null når X_1 er lik 0
- Effekten av X_1 på Y er signifikant forskjellig fra null når X_2 er lik 0
- Effekten av X_1 på Y er signifikant forskjellig fra null når X_2 er på sin gjennomsnittsverdi

25. Hva er formålet med et fortolkende case?

- Formålet er en detaljert beskrivelse av en spesifikk hendelse
- Formålet er beskrivende med en gitt populasjon
- Formålet er beskrivende med et teoretisk univers
- Formålet er analytisk med et teoretisk univers
- Formålet er analytisk med en gitt populasjon

26. Hva menes med livsfaseeffekt?

- At en uavhengig variabels effekt inntreffer i faser (trappetrinnseffekt)
- At en forskjell mellom aldersgrupper skyldes forskjeller mellom gruppenes oppvekstvilkår
- At en forskjell mellom aldersgrupper skyldes aspekter ved det å bli eldre
- At en uavhengig variabels har effekt bare i visse faser i enhetenes liv
- At en forskjell mellom aldersgrupper er mangfoldig

27. Hva menes med heteroskedastisitet?

- At restleddene korrelerer systematisk
- At restleddene er ukorrelerte
- At restleddene har samme varians
- At restleddene har ulik varians
- At restleddene er normalfordelte

28. Hva menes med bruttoendring?

- All endring i samme retning i fordelingen på en bestemt variabel mellom to tidspunkter
- Samlet endring i modellen
- All endring i enhetenes verdi på en bestemt variabel mellom to tidspunkter
- Summen av de uavhengige variablenes påvirkning på den avhengige i en regresjonsmodell
- Den totale endringen som blir resultatet når en forsker har revidert en statistisk modell

29. Hva menes med en positiv sammenheng?

- Lav verdi på en variabel opptrer typisk sammen med høy verdi på en annen variabel
- Høy verdi på en variabel opptrer typisk sammen med høy verdi på en annen variabel
- Sammenhengen blir av de fleste betraktet som godt nytt
- Sammenhengen blir av forskeren betraktet som godt nytt
- Sammenhengen gir grunnlag for å anbefale politiske reformer

30. Hva er en «slutning» («inference» på engelsk)?

- Prosessen der man bruker fakta vi kjenner til å dra lærdommer om fakta vi ikke kjenner
- Prosessen der man bruker fakta vi kjenner til å dra lærdommer om en årsakssammenheng
- Prosessen der man bruker fakta vi kjenner til å gi en grundig beskrivelse av et fenomen
- Prosessen der vi bruker årsakssammenhenger vi kjenner til å dra lærdommer om årsakssammenhenger vi ikke kjenner
- Prosessen der vi bruker fakta vi kjenner til å fremsette gyldige normative påstander

Del II: Kortsvarsoppgaver

Landet Utopia skal ha presidentvalg, og det er tre kandidater – Bryant, Curry og James – som velgerne kan stemme på. En spørreundersøkelse har nettopp blitt gjennomført, og resultatene i Tabell 1 viser frekvensfordelingen for hvordan respondentene besvarte spørsmålet: «Hvilken av de følgende kandidatene vil du stemme på ved det kommende presidentvalget?»

Tabell 1: Frekvensfordeling for spørreundersøkelse 1

Svaralternativ	Antall respondenter
1) Bryant	587
2) Curry	753
3) James	421
4) «Vet ikke»	231
5) Ikke besvart	57

- a) Omgjør den absolutte frekvensfordelingen i tabell 1 til en relativ frekvensfordeling. (3p)
- b) Hva er modus for fordelingen i Tabell 1 (2p)
- c) Er det mulig å bestemme medianen for fordelingen i Tabell 1? Begrunn svaret. Om mulig, bestem medianen. (3p)
- d) Regn ut den prosentvise fordelingen blant de tre kandidatene i Tabell 1, når vi ser bort ifra respondenter som har svart «vet ikke» eller som ikke har svart (2p)

En ny spørreundersøkelse blir gjennomført, der spørsmålet lyder: «Vil du stemme på Curry som president?» Den prosentvise svarfordelingen gis i Tabell 2

Tabell 2: Prosentvis fordeling for spørreundersøkelse 2

Svaralternativ	Prosentvis fordeling (N=1802)
1) Ja	56.7%
2) Nei	30.3%
3) Vet ikke	12.1%
4) Ikke besvart	1.9%

e) **Diskuter kort om det er elementer ved spørsmålsformuleringene som kan forklare forskjeller i resultater mellom de to spørreundersøkelsene (3p)**

f) **Diskuter to andre mulige årsaker til at kan bli forskjeller i svarfordelinger mellom to spørreundersøkelser som spør respondenter om det samme fenomenet (4p).**

Som en komplementerende undersøkelse til surveyene vurderes det å gjøre kvalitative dybdeintervjuer med velgere for å belyse hvilken motivasjon de har for å stemme på Curry. To alternative strategier blir vurdert: I) Dybdeintervju av 7 Curry-velgere og II) dybdeintervju av 3 Curry-velgere, 1 Bryant-velger og 1 James-velger.

g) **Diskuter fordeler og ulemper ved å velge strategi I) fremfor II) (6p).**

h) **La oss anta at strategi II) velges. Diskuter fordeler og ulemper ved å gjennomføre strukturerte intervjuer i stedet for uformelle/ustrukturerte intervjuer i denne konteksten (6p)**

En politisk journalist fremsetter følgende hypotese: Høyere utdanning øker sannsynligheten for å stemme på Curry, fordi høyere utdanning fører til mer liberale verdier (som Curry står for).

i) **Sett opp en kausalmodell som reflekterer denne hypotesen (3p)**

Journalisten fremsetter også en annen hypotese: Høyere inntekt øker sannsynligheten for å stemme på Curry, og denne effekten er sterkere for kvinner enn for menn.

j) **Sett opp en kausalmodell som reflekterer denne hypotesen (3p)**

I etterkant av presidentvalget (som Curry vant) blir det gjort en undersøkelse på hvilke faktorer som forklarer variasjonen i valgdeltakelse mellom landets 347 valgdistrikter. Valgdeltakelse (VDT) regnes ut som prosentandel av stemmeberettigede som avga stemme ved valget. Distriktet med lavest valgdeltakelse oppnådde 37.0%, mediandistriktet oppnådde 56.5% og distriktet med høyest valgdeltakelse oppnådde 71.5%.

k) **Hva er variasjonsbredden til valgdeltakelsesvariabelen for dette datamaterialet? (2p)**

De uavhengige variablene i studien er antall stemmeberettigede i distriktet (ANT); gjennomsnittlig alder på de stemmeberettigede i distriktet (ALD); gjennomsnittlig utdanningsnivå på de stemmeberettigede i distriktet, målt i antall år med utdanning (UTD); samt gjennomsnittlig inntektsnivå i distriktet, målt i antall 1000 Euro (INT).

En anerkjent professor i statsvitenskap kritiserer måten disse uavhengige variablene er målt på, og uttaler: «I stedet for at man måler gjennomsnittsverdier for alle stemmeberettigede for å forklare forskjeller i valgdeltakelse burde man heller fokusere på gjennomsnittsverdier for de som faktisk møtte opp og stemte».

l) Diskuter uttalelsen ovenfor. Er du enig med professoren? (6p)

m) Hvilket korrelasjonsmål ville du ha benyttet for å se på den bivarierte sammenhengen mellom inntekt og valgdeltakelse, og hvorfor? (3p)

Tabell 3 viser to OLS-regresjoner. Modell 2 inkluderer alle de nevnte uavhengige variablene, operasjonalisert som angitt i teksten ovenfor, mens modell 1 utelater inntektsvariabelen, siden denne variabelen mangler data for 89 distrikter.

Tabell 3: OLS regresjoner med valgdeltakelse. T-verdier i parentes.

	Modell 1	Modell 2
ANT	-0.004 (-9.38)	-0.003 (-8.02)
ALD	0.22 (1.89)	0.31 (2.21)
UTD	1.00 (2.45)	0.09 (0.24)
INT		0.31 (2.72)
Konstant	53.99 (29.97)	51.45 (31.61)
N	347	258
R²	0.53	0.69

n) I lys av informasjonen i Tabell 3, diskuter fordeler og ulemper ved å benytte modell 1 relativt til modell 2 dersom du hovedsakelig er opptatt av å estimere effekten av utdanning på valgdeltakelse. (6p)

- o) **Skriv opp regresjonsligningen for modell 1 (3p).**
- p) **Hva er standardfeilen til UTD i modell 1? Vis utregningen (2p)**
- q) **Forklar på en intuitiv måte hvorfor du tror koeffisienten til ANT i modell 2 er såpass liten i størrelse, samtidig som t-verdien til denne koeffisienten har så høy tallverdi. (4p)**
- r) **Hva er den predikerte verdien på VDT fra Modell 1 for distrikter der det bor 5 000 stemmeberettigede, snittalderen er 40 år, og snittutdanningen er 11 år? Vis utregningen. (3p)**

Ved nærmere ettersyn viser det seg at distriktene som ligger nord i landet har gjennomgående høyere valgdeltakelse enn distrikter i sør, selv når man tar høyde for forskjeller i alle de uavhengige variablene i Modell 2.

- s) **Forklar kort hvorfor man kan forvente autokorrelasjon i Modell 2, til tross for at det ikke er tidsseriedata som benyttes (3p)**
- t) **Beskriv kort en mulig strategi for å redusere autokorrelasjonsproblemet i Modell 2 (4p).**