

Eksamen i MET200, høst 2022

Eksamensform og vurdering:

Individuell, skriftlig hjemmeeksamen (4 timer).

Eksamen består av 3 deler. Hver del har flere deloppgaver som vektes i henhold til poeng oppgitt for hver deloppgave. Eksamen vurderes med gradert karakter (A-F).

Hjelpemidler: Alle tilgjengelige hjelpemidler er tillatt. Deler av eksamen krever utregninger i Excel. Det kreves ikke kildehenvisninger.

Sensorveiledning: Det gis maksimalt 60 poeng, 15 for del 1, 15 for del 2 og 30 for del 3. Karaktersettingen tar utgangspunkt i tabellen under, men karakteren kan justeres opp eller ned etter en helhetsvurdering.

	%	Poeng
A	≥89	≥53,4
B	≥77	≥46,2
C	≥65	≥39,0
D	≥53	≥31,8
E	≥41	≥24,6
F	<41	<24,6

Kvalitativ metode (maksimalt 15 poeng)

Oppgave 1. (5 poeng)

Hva er semi-strukturerte intervjuer, og hvorfor er dette en vanlig metode å benytte i kvalitativ forskning?

Sensorveiledning:

Temaene du vil spørre om er fastlagt på forhånd, men rekkefølge bestemmes underveis. Du tilpasser deg det intervjudeltageren sier, og kan stille oppfølgingsspørsmål som kan gi dypere kunnskap. Dette forsikrer at du med intervjuene kan besvare problemstilling, men også kan følge intervjudeltagerens fortelling. Fleksibiliteten kan fremme en tillitsvekkende situasjon og gjøre at deltageren opplever det som trygt å dele erfaringer og meninger.

Denne typen intervjuer er mye brukt da det gir en retning på samtalen som knytter seg til et overordnet forskningsspørsmål, samtidig som det åpner opp for at det kan komme temaer på banen som er nye for forskeren, og som ikke ville blitt synlige om det ble brukt standardiserte spørreskjemaer.

Oppgave 2. (5 poeng)

Forklar prinsippet om konfidensialitet i kvalitativ forskning, og vis hvordan dette prinsippet kan ivaretas gjennom forskningens ulike faser.

Sensorveiledning:

De som gjøres til gjenstand for forskning, har krav på at all informasjon de gir, blir behandlet konfidensielt. Forskeren må hindre bruk og formidling av informasjon som kan skade enkeltpersoner det forskes på. Forkningsmaterialet må vanligvis anonymiseres, og det må stilles strenge krav til hvordan lister med navn eller andre opplysninger som gjør det mulig å identifisere enkeltpersoner oppbevares og tilintetgjøres (NESH, 2006). Det innebærer at forskeren må anonymisere deltakere i prosjektet når resultatene av undersøkelsen presenteres. Prinsippet omhandler også spørsmål om gjenbruk og lagring av informasjon.

Oppgave 3. (5 poeng).

Redegjør for begrepet **posisjonering**/forforståelse. Kom med et (tenkt) eksempel på en situasjon hvor forskeren bør reflektere rundt sin egen forforståelse av fenomenet som studeres

Sensorveiledning:

Reflektere over, samt skrive frem egen posisjon. Være gjennomsiiktig på egen erfaring med tema, gjennom praksis, egen historie osv. Nærhet og distanse er sentrale stikkord. Forskeren kan ha inngående kjennskap til feltet som studeres, og kanskje taus kunnskap om fenomenet. Da bør refleksjonen knyttes til hvordan forskeren kan skape nødvendig distanse til deltakernes ytringer for å unngå at meningsinnholdet tas for gitt. På den annen side, dersom forskeren nærmer seg et felt hen har lite kjennskap til vil det være hensiktsmessig å opparbeide nok kunnskap om feltet til å kunne stille spørsmål som er med på å frembringe relevant kunnskap. Slike prosesser bør redegjøres for.

Kvantitativ metode (maksimalt 15 poeng)

Oppgave 1. (5 poeng)

Du får i oppdrag å gjennomføre en studie for å undersøke om det er en sammenheng mellom motorisk utvikling i 4-6 års alder og fysisk aktivitetsnivå som ung voksen. Ta utgangspunkt i at du har tilgang på et representativt utvalg barnehagebarn og at det finnes valide målemetoder til formålet og at det er etisk forsvarlig å gjennomføre studien. Beskriv hvordan du ville designet en slik studie og hvorfor designet du har valgt bringer oss nærmest mulig et svar på om det er en sammenheng mellom motorisk utvikling og aktivitetsnivå som voksen. Sett opp et tosidig hypotesepar for studien. Beskrivelsen bør ikke overskride ½ side.

Sensorveiledning: Vi er her ute etter alle relevante momenter ved en kohort-studie. Full score krever refleksjoner rundt konfunderende faktorer og endring av adferdi

Oppgave 2. (6 poeng)

Under finner du beskrivelser av tre studier. Avgjør hvilket studiedesign de ulike studiene benytter. Begrunn kort (1-2 setninger).

- a) En gruppe forskere benytter data fra en stor nasjonal kartlegging av fysisk aktivitetsnivå blant barn og unge i Norge. I studien ble det inkludert barn i 1., 4. og 10. klasse fra skoler i hele landet. Alle barna har gått med aktivitetsmåler i 7 dager og ved bruk av spørreskjema har de svart (JA/NEI) på om de har diagnostisert astma. Forskerne ønsker å undersøke om det er en assosiasjon mellom fysisk aktivitet og astma.

Sensorveiledning: Tversnittsstudie. Eksponering og utfall er studert på samme tidspunkt og gir et øyeblikksbilde av sammenhengen

- a) En gruppe forskere kobler, ved hjelp personnummer, aktivitetsnivå registrert med akselerometer blant norske menn og kvinner i alderen 20-85 år mot data fra dødsårsaksregisteret. Aktivitetsnivået er registrert for noen år siden. Dette gjør de for å undersøke om det er en sammenheng mellom fysisk aktivitetsnivå og total dødelighet

Sensorveiledning: Kohortstudie (longitudinell, follow-up og prospektiv studie kan også godkjennes). Deltakerne kategoriseres i henhold til eksponeringsgrad ved baseline (aktivitetsnivå) og "følges" prospektivt

- b) En gruppe forskere ønsker å undersøke om 12 uker med henholdsvis intervalltrening og kontinuerlig utholdenhetstrening påvirker oksygenopptaket hos en gruppe utrente menn og kvinner. Forskerne rekrutterer deltakere inn i studien fortløpende og etter at de har samtykket blir tilfeldig fordelt mellom følgende grupper; 1) intervalltrening to ganger i uka, 2) kontinuerlig utholdenhetstrening 2 ganger i uka, eller 3) fortsette som normalt i 12 uker. Målsettingen i studien er å rekruttere 15 deltagere til hver gruppe. Deltagere måler oksygenopptaket før og etter 12 ukers perioden og endringen sammenlignes mellom de tre gruppene.

Sensorveiledning: Randomisert kontrollert studie (deltagere blir tilfeldig loddtrukket til gruppe)

Oppgave 3. (2 poeng)

Hva innebærer det at en målemetode er valid? Gi et tenkt eksempel på hvordan du kan gå frem for å undersøke om en målemetode er valid.

Sensorveiledning:

Valid: Hvorvidt man evner å måle det (teoretiske begrepet) man ønsker å måle/undersøke nøyaktig. Validitet eller kriterievaliditet kan sjekkes ved å sammenligne en metode opp mot gullstandard. Eks kan man med veieing kontrollere at vekten er kalibrert

Oppgave 4. (2 poeng)

Hva mener vi når vi sier at forskningsresultat er generaliserbart? Hvilke forutsetninger ligger til grunn for at vi kan generalisere et forskningsresultat

Sensorveiledning:

Generaliserbarhet betyr at vi mener at forskningsresultatene kan generaliseres til å gjelde for alle enhetene (populasjonen) som utvalget er trukket fra. Forutsetninger for generaliserbarhet er;

1. at utvalget ditt er representativt for populasjonen du ønsker å si noe om
2. høy nok deltagelse
3. indre og ytre validitet

Statistikk (maksimalt 30 poeng)

Statistikkdelen av eksamen er delt inn i tre deler. Hver del har et tilhørende datasett. De tre datasettene laster dere ned fra Canvas (Moduler → Eksamen - datasett).

Oppgave 1 (Bruk datasettet «datasett_basis2021_eksamen.xlsx» i Canvas)

1. Fyll ut tabellen med passende deskriptiv statistikk for de ulike variablene, separat for menn og kvinner. Rund av til to desimaler der nødvendig. Forklar hvordan du har tenkt. (4 poeng)

Variabel	Menn (n=107)	Kvinner (n=93)
LDL-kolesterol (mmol/L)*	3,03 (0,89)	2,98 (1,01)
MVPA pr dag (min)**	14,8 (5,7-37,6)	7,5 (2,3-19,0)
Røykestatus***		
Aldri	33 (31)	56 (60)
Tidligere	45 (42)	27 (29)
Nåværende	29 (27)	10 (11)

* Gjennomsnitt (SD); ** Median (Q1-Q3); *** n (%)

LDL-kolesterol er en numerisk variabel uten tegn til sterk skjevfordeling og er derfor beskrevet med gjennomsnitt og standardavvik. MVPA pr dag er en sterkt skjevfordelt, numerisk variabel og er derfor beskrevet med median og kvartiler. Røykestatus er en kategorisk variabel og er beskrevet med n (%).

2. Hva er forskjellen på gjennomsnitt og median? Vurder hvordan disse sentralmålene påvirkes av ekstreme verdier (ikke behov for Excel). (2 poeng)
- Gjennomsnittet er summen av alle observasjonene delt på antall observasjoner, og vil dermed inkludere alle verdier – også eventuelle ekstreme verdier. Avhengig av antall observasjoner og hvor ekstreme verdiene er vil gjennomsnittet bli forskjøvet i retning av disse verdiene. Medianen er den midterste verdien når variabelen er sortert etter størrelse, og vil derfor ikke påvirkes av andre verdier – heller ikke ekstreme verdier.
3. Når man konkluderer etter en statistisk test er det to typiske feil man kan gjøre når det kommer til hypotesene. Forklar hva feilene kalles og hva hver av disse innebærer (ikke behov for Excel). (2 poeng)
- Type I-feil: Falsk positiv – feilaktig forkaste nullhypotesen
 - Type II-feil: Falsk negativ – feilaktig beholde nullhypotesen
4. Hva er forutsetningene for å kunne gjøre en lineær regresjon (ikke behov for Excel)? (2 poeng)
- Den viktigste forutsetningen er at man har to numeriske variabler som har en tilnærmet lineær sammenheng, deretter stiller man følgende krav til residualene;
 - Residualenes varians er konstant
 - Residualene er uavhengige av hverandre
 - Residualene er normalfordelte

Oppgave 2 (Bruk datasettet «Vekt.xlsx» i Canvas)

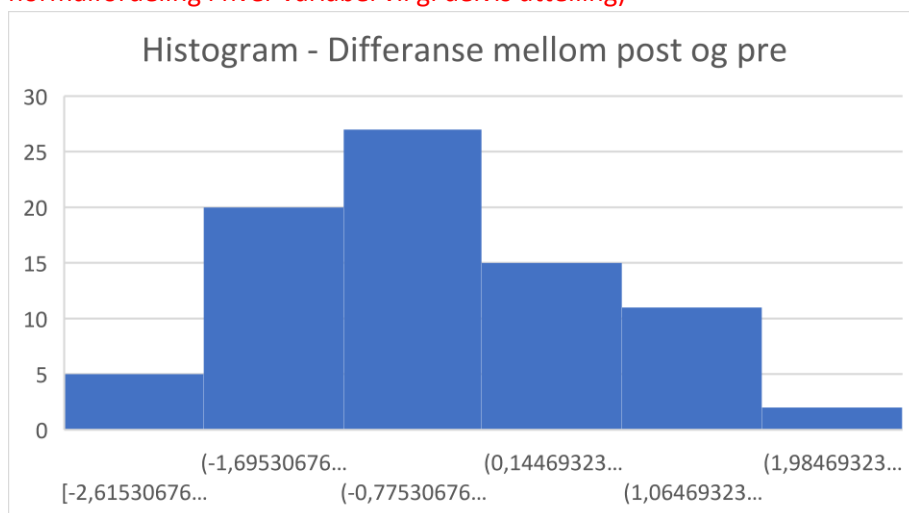
Vi har gjennomført en 12 ukers treningsintervensjon blant overvektige voksne, og vi ønsker å se på hvordan kroppsvekten til deltakerne har endret seg fra før intervensjon (Vekt_pre) til etter intervensjon (Vekt_post). Bruk tosidig p-verdi, med signifikansnivå satt til 0,05.

1. **Hva slags variabeltyper er de sentrale variablene her? (1 poeng)**
 - a. «Vekt_pre» og «Vekt_post» er begge er numeriske (kontinuerlige) variabler.
2. **Hva er passende statistisk test for å svare på problemstillingen? (1 poeng)**
 - a. Paret t-test.
3. **Gjør analysen og konkluder på problemstillingen. (4 poeng)**
 - a. Den gjennomsnittlige kroppsvekten i utvalget var omtrent 90,3 kg før intervensjon og omtrent 90,1 kg etter, noe som innebærer at utvalget i gjennomsnitt veier 0,2 kg mindre etter intervensjonsperioden. Forskjellen er ikke statistisk signifikant ved signifikansnivå 0,05, med en tosidig p-verdi på 0,185. Vi må derfor her beholde H_0 .

t-Test: Paired Two Sample for Means

	Vekt_pre	Vekt_post
Mean	90.30755	90.13852
Variance	86.03277	81.29722
Observations	80	80
Pearson Correlation	0.992767	
Hypothesized Mean Difference	0	
df	79	
t Stat	1.33794	
P(T<=t) one-tail	0.092377	
t Critical one-tail	1.664371	
P(T<=t) two-tail	0.184755	
t Critical two-tail	1.99045	

4. **Er forutsetningene for å gjøre den statistiske testen møtt? Utdyp. (2 poeng)**
 - a. Forutsetningene for å kunne gjøre en paret t-test er at differansen mellom de to variablene er omtrent normalfordelt. I dette tilfellet er differansen tilnærmet normalfordelt, og vi kan derfor si at forutsetningene er oppfylt. (Vurdering av normalfordeling i hver variabel vil gi delvis uttelling)



Oppgave 3 (Bruk datasettet «Forsikringsutbetalinger.xlsx» i Canvas)

Du har blitt leid inn som konsulent i et amerikansk helseforsikringsselskap, der de ønsker å finne ut av om de har større årlige dollarutbetalinger («Kostnad») til de som røyker enn de som ikke røyker («Røyker»). Gjør om variabelen «Kostnad» til en kategorisk variabel («Kostnad_kategori»), der du får oppgitt at kostnader under 10 000 dollar/år klassifiseres som «Lav» utbetaling, kostnader fra 10 000 til 29 999 dollar/år klassifiseres som «Middels» utbetaling, og kostnader på 30 000 dollar/år eller mer klassifiseres som «Høy» utbetaling. Du velger å sette signifikansnivået til 0,05.

1. Sett opp H_0 og H_1 for denne problemstillingen. (1 poeng)
 - a. Det er ingen forskjell i forsikringsutbetaling mellom røykere og ikke-røykere.
 - b. Det er en forskjell i forsikringsutbetaling mellom røykere og ikke-røykere.
2. Hva slags variabeltype er «Røyker»? (1 poeng)
 - a. Kategorisk (nominal).
3. Lag to krystabeller med variablene «Røyker» og «Kostnad_kategori» - en med de observerte verdiene og en med de forventede verdiene. (4 poeng)

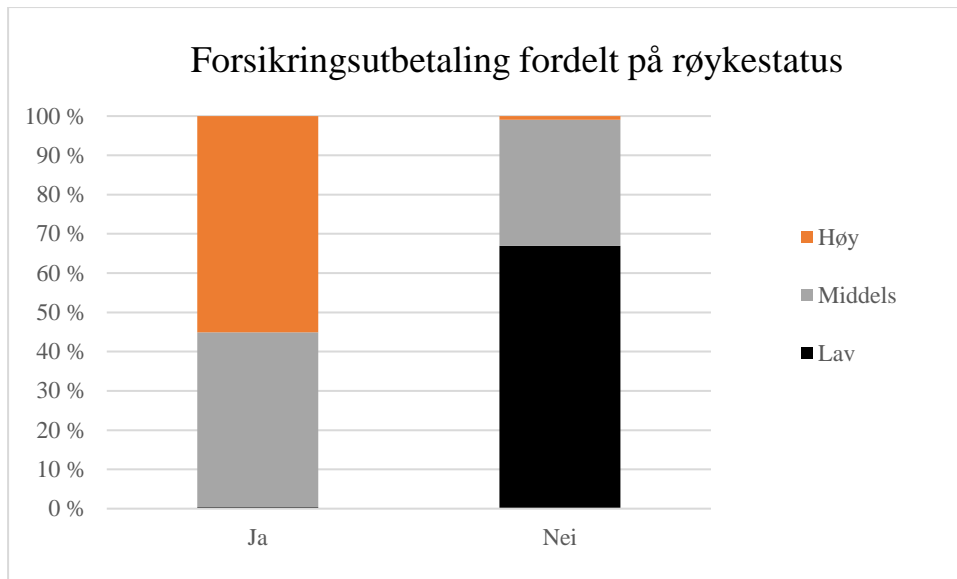
Observerte verdier

		<i>Kostnadskategori</i>			
<i>Røyker</i>	<i>Lav</i>	<i>Middels</i>	<i>Høy</i>	<i>Total</i>	
<i>Ja</i>	1	122	151	274	
<i>Nei</i>	712	342	10	1064	
<i>Total</i>	713	464	161	1338	

Forventede verdier

		<i>Kostnadskategori</i>			
<i>Røyker</i>	<i>Lav</i>	<i>Middels</i>	<i>Høy</i>	<i>Total</i>	
<i>Ja</i>	146	95	33	274	
<i>Nei</i>	567	369	128	1064	
<i>Total</i>	713	464	161	1338	

4. Bruk en egnet figur til å visualisere forskjellene mellom disse to variablene. Kalkuler også en p-verdi ved å bruke kji-kvadrat, og konkluder på problemstillingen. (4 poeng)
 - a. Den største andelen av røykere ligger i gruppen med høye utbetalinger fra helseforsikringene, mens den største andelen av ikke-røykere befinner seg i gruppen med lave utbetalinger. En kji-kvadrat test gir en p-verdi på $<0,001$, og vi kan med det si at disse forskjellene er statistisk signifikante og forkaste nullhypotesen.



5. Er forutsetningene for å kunne gjøre kji-kvadrat møtt? Forklar. (2 poeng)

- a. Forutsetningene for å kunne gjøre kji-kvadrat er at man har to kategoriske variabler, og at antall forventede verdier i hver celle er minst 5. I dette tilfellet er forutsetningene tilfredsstillt.