

Noen momenter ved vurdering av eksamen

PSY1010 – PSYC1100 våren 2021

Denne veiledningen peker på noen enkeltelementer som kan diskuteres i oppgavene. Den er ikke dekkende eller ment som en fullstendig liste, og skal ikke betraktes som absolutte krav. Oppgavene kan løses på mange måter og i tilfeller der kandidaten har benyttet en uortodoks løsning blir sensor nødt til å vurdere hvorvidt løsningen kan regnes som akseptabel.

Som en regel skal kandidaten belønnes for å vise forståelse av faget og de prinsipielle sidene av metodestoffet. I tilfeller der kandidaten trekker inn kunnskap fra andre områder av faget eller tilgrensende fagområder enn det som direkte spørres etter i det enkelte spørsmål må man vurdere relevansen. Er stoffet gjort relevant for spørsmålet skal dette honoreres. I tilfeller der stoffet ikke er særlig relevant og mer bærer preg av "jeg kan også dette" skal man se bort fra dette.

Legg vekt på forståelse og oversikt av stoffet når dere leser oppgavene. Kandidater som viser god oversikt og forståelse bør honoreres. Det er også flott hvis kandidaten klarer å illustrere poengene sine med gode eksempler.

Ved regneoppgaver bør man i tillegg til svaret også vurdere måten utregningen er gjort på. I tilfeller der galt svar åpenbart bare skyldes enkle regnefeil bør man ta hensyn til dette i vurderingen av svaret. Merk at for de som ikke har med kalkulator, er det viktig at kandidaten viser utregningene og forståelse for fremgangsmåten.

Alle fire oppgaver teller like mye og kandidaten får ikke trekk eller bonus for å velge noen framfor noen andre. Det skal heller ikke krediteres om kandidaten har svart på alle fire, i stedet for tre.

Vedrørende deloppgavene, er det med hensikt at noen oppgaver er vanskeligere enn andre. Dette gjør det lettere å sette karakter, og det er ikke meningen at sensor skal være like streng på alle.

Tre oppgaver skal være besvart. Det vil si at kandidater som kun har besvart 2 oppgaver automatisk stryker. Det gjelder også om en oppgave er så tynt besvart at den ikke viser noe kunnskap/er jevngodt med ubesvart. Dersom en oppgave ligger litt under grensen for å stå, men viser noe kunnskap, kan kandidaten allikevel stå dersom de andre oppgavene er såpass gode at de i snitt trekker eksamen opp til ståkarakter.

Studentene har våren 2021 eksamen som åpen bok-eksamen. Det vil si at alle hjelpemidler er tillatt. Det er derfor ekstra viktig å legge vekt på forståelse framfor å gjengi «riktig resultat». Studentene har fått beskjed om at de kan bruke valgfrie dataprogrammer så lenge hele besvarelsen leveres som en PDF. Eksempelvis kan man tegne på ark og skanne det, tegne med penn på nettbrett, plotte tall i dataprogram og så videre. Valg av teknisk løsning skal ikke trekke opp eller ned på karakteren.

OPPGAVE 1

Tenk deg at du skal gjennomføre en studie av effekten av en ny type kortvarig og intensiv behandling av tvangslidelse (obsessiv-kompulsiv lidelse; OCD). Deltakerne kommer fra én klinikk, hvor halvparten av pasientene som inngår i studien mottar den nye behandlingsformen, mens den andre halvparten mottar vanlig behandling. I forbindelse med dette ønsker du å bruke et spørreskjema som måler symptomer på tvangslidelse.

a) Hva er uavhengig og avhengig variabel i denne undersøkelsen?

Behandlingstype er uavhengig variabel, symptomer på OCE er avhengig variabel. Ufyllende svar kan gjerne honoreres.

b) Hvorfor er det viktig å undersøke reliabiliteten til tester, og hvordan kan man gjøre dette?

*Kandidaten bør kunne si noe om viktigheten av stabilitet i målingen for at resultatet skal kunne betraktes som et godt mål på et fenomen. Det er et pluss om kandidaten tar med forholdet mellom reliabilitet og validitet. (Kan også diskuteres under c) Har kandidaten forstått at en test kan være stabil uten at den måler det man mener at den skal måle? Pluss hvis kandidaten har forstått at man ikke opererer med absolutter. **Ulike måter å estimere reliabilitet på kan også diskuteres og når dette gjøres relevant for spørsmålet bør det honoreres.** På forelesning er det nevnt test-retest reliabilitet, reliabilitet til parallelle former, splitt - half reliabilitet, intern konsistens og inter-rater reliabilitet.*

c) Hvorfor er validitet viktig og hvilke former for validitet bør man vurdere i dette spørreskjemaet?

Vi er her ute etter validitet knyttet til psykologiske tester, som spørreskjemaet er et eksempel på. Kandidaten bør diskutere viktigheten av man faktisk måler det man skal måle. Følgende former for validitet er nevnt i pensum. Face validity, Kriteriebasert validitet, samtidig validitet, prediktiv validitet, innholdsvaliditet og begrepsvaliditet.

d) Hva kan du si om indre og ytre validitet i denne studien?

Bør i det minste si at indre validitet handler om å etablere årsak virkning og ytre validitet om generalisering. En utfordring for indre validitet her kan være at studien ikke er blindet, deltakerne forstår hvilken gruppe de er i, reaktivitet osv. Ytre validitet kan være truet av at kun en klinikk er med i studien. Her er det rom for refleksjon og oppgaven kan tas i ulike retninger.

OPPGAVE 2

Du gjennomfører en studie av effekten av en ny type kortvarig og intensiv behandling av tvangslidelse (OCD). Det er totalt 50 deltakere i studien. Halvparten ($n_1=25$) får gammel behandling, mens den andre halvparten ($n_2=25$) får den nye behandlingen. Seks måneder senere svarer de på et spørreskjema som måler symptomer på tvangslidelse. Hver deltaker får en skåre fra 1 (få symptomer) til 10 (mange symptomer).

I gruppe 1, som fikk gammel behandling, var gjennomsnittet ($\bar{x}_1$) 5,0 og variansen (s_1^2) 2,75.

I gruppe 2, som fikk ny behandling, var gjennomsnittet ($\bar{x}_2$) 3,5 og variansen (s_2^2) 2,75.

Under er noen formler du kan få bruk for i denne oppgaven:

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

$$SD = \sqrt{s^2}$$

$$SE = \frac{SD}{\sqrt{n}}$$

$$SE(D) = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} = SD \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

$$t = \frac{D}{SE(D)} = \frac{\bar{x}_1 - \bar{x}_2}{SE(D)}$$

a) Forklar kort hva varians (s^2), standardavvik (SD) og standardfeil (SE) betyr og regn ut standardavviket og standardfeilen for de to gruppene.

Varians: Kandidaten bør vise forståelse for at man at dette er mål på spredningen av skårene. Oppgaven etterspør en kort forklaring, så mye mer er derfor ikke nødvendig, men ulike ting det går an å ta med (uten at man trekker om dette ikke er med): Varians kan aldri være negativ; varians består av ulike komponenter – sann variasjon og varians som skyldes målefeil, som igjen kan bestå av systematisk og tilfeldig målefeil; man kan studere ulike kilder til varians, som i variansanalyse, forklart varians osv.

Standardavvik: Det bør framgå at kandidaten forstår at standardavviket er et mål på spredning som kan overføres direkte til den skalaen målingen er gjort på, og bra om det nevnes at standardavviket er roten av variansen. Mange vil nok forklare dette som (absolutt) gjennomsnittlig avvik fra gjennomsnittet – en slik tolkning er ok å ha med, men bonus til de som forstår at SD ikke faktisk helt sammenfallende med den tolkningen.

Standardfeil: Kandidaten bør vise at han/hun har forstått at standardfeilen er knyttet tilfeldig variasjon/hvor langt unna det 'sanne' gjennomsnittet vi kan forvente å komme/at det er et mål på feilmargin. Videre skal det honoreres om kandidaten redegjør for som at standardfeilen er beregnet på bakgrunn av standardavviket i utvalget og utvalgsstørrelsen og kan forståes som standardavviket i utvalgsfordelingen.

Bør besvares med minst 1 siffer, men flere er greit.

$$SD = \sqrt{s^2} = \sqrt{2,75} = 1,658$$

$$SE = SD / \sqrt{n} = 1,658 / 5 = 0,33$$

b) Regn ut standardfeilen for differansen mellom gruppene, $SE(D)$, og regn ut t . Forklar kort hva disse begrepene viser til. Obs: Siden variansen er lik i gruppene, er $s_p = SD$, som du regnet ut i oppgave 2a.

$$SE(D) = \sqrt{2,75} * \sqrt{1/25 + 1/25} = 1,658 * 0,283 = 0,469$$

$$t = D / SE(D) = (\text{mean}(x_1) - \text{mean}(x_2)) / SE(D) = (5,0 - 3,5) / 0,469 = 3,198$$

Standardfeil for differansen: Hvor stor forskjell vi kan forvente mellom gruppene selv om det ikke finnes noen forskjell i virkeligheten (under nullhypotesen).

t: Plassering langs t-fordelingen; avvik fra forventet verdi (0) målt i standardfeil. Hvor ekstrem skåren er. Jo høyere skåre desto mindre sannsynlig å oppstå når det ikke finnes noen sammenheng i virkeligheten.

c) Du slår opp i en tabell og finner at t-verdien i dette tilfellet tilsvarer en p-verdi på 0,002. Hva betyr dette? Hvordan ville du konkludert om studien på bakgrunn av dette?

Dette vil si at dersom nullhypotesen er sann, vil man observere resultater like eller mer ekstreme i 0,002 (altså 0,2%) av tilfellene. Videre må studentene reflektere rundt hypotesetesting. De kan godt fortelle om utvalgsfordelinger og nullfordistribusjoner. Alfa er ikke oppgitt i oppgaven, så det er et tema studentene kan drøfte. Antakelig vil man sette alfa til 0,05 eller 0,01; i begge tilfeller ville man sagt at dette er statistisk signifikant. Det vil så fall si at vi forkaster nullhypotesen, og at vi kan konkludere med at det er forskjell i symptomnivå etter behandlingsform. Hvis studentene trekker inn styrke/power og at man kan øke utvalgsstørrelsen vil det også være en relevant betraktning.

d) Tenk deg at du skulle sammenlignet flere enn to grupper. Hvordan ville du da gått fram?

Her kan man snevert si at man bør bruke variansanalyse for å redusere samlet feilrate sammenlignet med en rekke t-tester. Kandidatene kan dessuten reflektere rundt andre utfordringer knyttet til å sammenligne flere grupper, som at det krever flere ressurser, hva som kan kjennetegne de ulike gruppene osv.

OPPGAVE 3

Du ønsker å finne ut om det er en kausal sammenheng mellom hukommelsestap og Covid-19 sykdom blant et stort utvalg av personer.

a) Beskriv kort hvilket forskningsdesign du ville velge?

Oppgaven utfordrer kandidaten til å tenke selvstendig og vise forståelse. Her kan kandidaten foreslå ulike varianter av eksperimentelle design som avdekker årsaksforhold (dvs har høy indre validitet). Kandidaten bør resonnerer seg frem til å velge et solid forskningsdesign som har høy grad av eksperimentell kontroll. Betingelser til et ekte eksperiment er at man bruker kontrollgruppe, randomisering og har et pretest-posttest kontrollgruppedesign. Randomisert tilordning til grupper er et viktig kjennetegn på ekte eksperiment (tilfeldig hvem som havner i eksperimentgruppen, og hvem som havner i kontrollgruppen). Randomisert tilordning skal sikre lik fordeling av feilvarians mellom gruppene. Hvis man ikke følger en randomisert tilordningsprosedyre, utgjør det en trussel mot indre validitet.

b) Hvilke fordeler og ulemper har designet du har valgt?

Også denne oppgaven utfordrer kandidaten til å tenke selvstendig og vise forståelse. Det må her redegjøres for trusler mot indre og ytre validitet i designet de har valgt og må derfor sees i sammenheng med besvarelsen over (3a).

c) Du finner i tillegg at det er en positiv korrelasjon på 0.35 mellom hukommelsessvikt og Covid-19 sykdom. Hva betyr det?

Kandidaten bør forklare prinsippet om samvariasjon. Bra om kandidaten viser til at korrelasjon er basert på naturlig variasjon til forskjell fra eksperimentet som benytter påført variasjon under kontrollerte betingelser. I tillegg til å svare at det er positiv korrelasjon mellom variablene er det et pluss hvis kandidaten beskriver «hvordan» (at begge går i samme retning: at begge øker eller minker) og diskuterer styrkeforholdet til korrelasjonen. En r på 0,35 er på forelesning vurdert som en lav korrelasjon, men dette er konvensjoner. Fint om kandidaten poengterer at sammenhengen kan gå begge veier mellom hukommelsessvikt og Covid-19.

- d) Hvorfor skal man være forsiktig med å trekke slutninger om årsak-virkning fra korrelasjonsdata?

Her står kandidaten fritt til å forklare med egne ord. Vi har snakket mye om å skille mellom korrelasjon og kausalitet, samt korrelasjonsmetodens begrensninger: at årsaksretning kan gå begge veier og problemet med tredjevariabler og spuriøse sammenhenger. Ved antagelser om årsak må rivaliserende hypoteser utelukkes da disse kan forklare korrelasjonen, og dette er vanskelig å gjennomføre i praksis om man studerer sammenhenger mellom hukommelsestap og Covid-19 som opptrer i en naturlig setting. I et eksperiment derimot ønsker man kontroll på alt dette (ytre og indre validitet).

OPPGAVE 4

Du ønsker å gå i dybden på hvorvidt hukommelsestap etter Covid-19 sykdom kan bedres etter inntak av et nytt godkjent medikament. Du har denne gangen tilgang på kun noen få forsøkspersoner.

- a) Du lurer nå på om du skulle ha benyttet et kvalitativt kasus studie eller et $N=1$ design. Diskuter fordeler og ulemper ved de to metodetilnærmingene. Bruk den beskrevne situasjonen som eksempel.

Disse designene må ikke blandes. Kasus studiet er en kvalitativ deskriptiv teknikk hvor virkeligheten gjenspeiles og beskrives, man manipulerer ikke og dette er ikke en eksperimentell metode. Det er egnet om man ønsker å gå i dybden og undersøker sjeldne tilstander. I dette eksemplet vil det da være en person som har fått et sjeldent og alvorlig tilfelle av hukommelsestap etter Covid-19 sykdom. Det bør fremgå at kasus studiet ikke er egnet til å trekke slutninger om årsaks-sammenhenger og gir ikke svar på om hukommelsestap etter Covid-19 sykdom kan bli bedre etter inntak av nytt medikament. Kasus studie har problemer både i forhold til indre- og ytre validitet. Singel-Subject design ($N=1$) er et eksperiment med få personer og vil derfor i større grad avdekke om medikamentet fungerer på hukommelsestap. Her kan man for eksempel velge et ABAB-design, og jo flere replikasjoner (ABABAB) jo bedre, pga bedre intern validitet. Flott om carry-over effekten av medikamentene blir diskutert som utfordring her.

- b) Hvilke varianter av Single-Subject Design ($N=1$) kjenner du til?

Vi har gjennomgått tre typer: 1. Baseline design 2. Diskret forsøksdesign og 3. Dynamiske design.

- c) Beskriv noen kjente eksempler på Singel-Subject Design innen psykologisk forskning.

Her er det fort gjort å hente eksempler fra kvalitative kasus studier og ikke fra eksperimenter. De kan for eksempel nevne eksperimentene til Thorndike, Ebbinghaus, Skinner og Pavlov. Av nyere eksempler kan Moser & Moser sin hjerneforskning på rotter nevnes.

- d) Forklar kort hvorfor $N=1$ design ligner et within-subject design.

$N=1$ design ligner within-subject design ved at alle subjekter i eksperimentet gjennomgår alle nivåer av en variabel. Men i $N=1$ design sammenlignes ikke gj.snitt med andre subjekter. For single-subjects designs benyttes gjennomsnittet av mange målinger på samme person/dyr hvor subjektet er sin egen kontrollgruppe. Slik reduseres feilvarians. Stabilitetskriteriet hos forsøkspersonen er viktig, og god kontroll på ytre omgivelser for å trekke riktige konklusjoner om årsak-virkning (god indre validitet).

Lykke til med sensuren av eksamensoppgavene!

Beste hilsen,
Maria Teresa Grønning Dale og Fartein Ask Torvik