

## Noen momenter ved vurdering av eksamen PSY1010 – PSYC1100 våren 2022

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. **NB! I år ble det ikke oppgave 4d markert tydelig i Inspira som egen deloppgave. Noen studenter har da ikke svart på «4d», men på de andre deloppgavene i oppgave 4. De skal da ikke trekkes for dette i eksamensvurderingen.**

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 2022 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, plote tall i dataprogram og så videre. Valg av teknisk løsning skal ikke trekke opp eller ned på karakteren.

## OPPGAVE 1

Du har en gruppe på 60 personer som var på Utøya i 2011 og som sliter med posttraumatisk stress (PTSD). Du ønsker å undersøke om kognitiv atferdsterapi eller antidepressiva er symptomlindrende mot PTSD-symptomer eller ikke, eller om det er en kombinasjon av terapi og medikamenter som er mest effektivt.

a. Hva slags design vil du benytte her, og hva er det som kjennetegner denne type design?

*Her er det i praksis mulig å foreslå ulike design, og utfordrer kandidaten til å tenke selvstendig og vise forståelse. Når man har to uavhengige variabler som man undersøker effekten av, både hver for seg og i forhold til hverandre, er faktorielle design egnet fordi det undersøker effekter av flere enkeltfaktorer alene, og hvordan flere faktorer virker sammen: Vi ser på multiple korrelasjoner. Når vi i tillegg har så mange som 60 forsøkspersoner, er det nærliggende å foreslå et faktorielt mellomgruppe design. Hvis man for eksempel vurderer kost-nytte prinsippet sparer man forsøkspersoner ved å ha forskjellige personer i hver gruppe, enn om alle 60 gjennomgår alle intervensjoner. Den statistiske analysemetoden heter da tofaktor mellomgruppe ANOVA (variansanalyse), der vi har forskjellige personer i hver betingelse/gruppe. Hvis man for eksempel ønsker å se på grupper som enten får terapi eller medikamenter opp mot de som ikke får terapi og medikamenter blir det et 2x2 design med 4 grupper. Et slikt 2x2 faktorielt design har fire betingelser, med to uavhengige variabler (faktorer) som hver for seg har to nivåer. Det er i denne besvarelsen opp til den enkelte hvor mange nivåer de vil ha på hver faktor.*

b. Hva er avhengig (e) og uavhengig (e) variabler i dette eksempelet?

*Uavhengig variabel (årsak): medikamenter og terapi*

*Avhengig variabel (virkning): PTSD-symptomer*

c. Hva menes med hovedeffekt og interaksjonseffekt her?

*Hovedeffekt er den separate effekten av en uavhengig variabel. Ved interaksjonseffekt betyr det at effekten av én uavhengig variabel er avhengig av hvilket nivå det er på den andre uavhengige variabelen. Ved interaksjon kan vi derfor spørre: hvordan samvarierer bruk av medikamenter og terapi på grad av PTSD-symptomer?*

d. Du finner en interaksjonseffekt mellom de to behandlingsformene. Forklar hva det innebærer og lag gjerne figur og/eller tabell for å illustrere dette.

*Ved et 2x2 faktorielt design kan dette belyses ved en 4-felts tabell og /eller en figur (graf). Da vil for eksempel hovedeffekt 1 bregnes ved å regne ut forskjellene mellom kolonnegjennomsnittene og dele på 2. Hovedeffekt 2 beregnes ved å se på forskjellene på rekke(rad-)gjennomsnittene og dele på 2. Interaksjonseffekten finner vi ved å beregne forskjellene på gjennomsnittene i cellene i tabellen. Interaksjonseffektene kan ta ulike kombinasjoner (gjennomgått på forelesning), for eksempel kan interaksjonseffekt være tilstede uten a hovedeffekter der kun medikamenter og terapi sammen reduserer PTSD-symptomer, mens de to behandlingsformene hver for seg ikke har noen effekt. Merk at en to-veis interaksjon kan maskere hovedeffekter.*

## OPPGAVE 2

Man sier ofte at utvalget (sample) representerer populasjonen.

- a. Hva menes med utvalg og populasjon?
- b. Hvilke fremgangsmåter kan man benytte for å trekke et utvalg?
- c. Hvilke mål bør man rapportere for å beskrive et utvalg.
- d. Måten man trekker et utvalg på kan ha betydning for generaliserbarheten av resultatene. Forklar hvorfor.

*a) Utvalg er gruppen man utfører forskningen på. Sample er alle teoretisk mulige forsøkspersoner. Utvalget skal være en miniatyrkopi av populasjonen. Ekstra pluss hvis kandidaten forstår at sampling ikke nødvendigvis handler om personer, men om mulige observasjoner.*

*b) Forskjellen mellom sannsynlighetsutvalg og ikke-sannsynlighetsutvalg må være diskutert. Typer av samplingteknikker som er nevnt: Vanlig randomisert sampling. Vanlig stratifisert sampling og stratifisering ved hjelp av forholdstall (proporsjoner). Systematisk sampling. Cluster (klynge) sampling.*

*c) Kandidaten bør nevne og beskrive mål på spredning (range, Interkvartil range, varians og standardavvik) og sentraltendens (Gjennomsnitt, median og modalverdi). Pluss hvis kandidater har diskutert forholdet mellom de ulike målene for både spredning og sentraltendens. Hvis kandidaten viser forståelse for at standardavviket er kvadratroten av variansen og at SD kan overføres direkte til den skalaen målingen er gjort på til forskjell fra variansen som er kvadrert, skal dette honoreres selv om det også kan inngå i oppgave 3.*

*d) Problemer knyttet til generalisering og randomisert kontra ikke-randomisert sampling bør være diskutert og forstått. Kandidaten bør kunne diskutere problemene knyttet til skjev seleksjon opp mot begrepet ytre (ekstern) validitet.*

### **OPPGAVE 3**

Å forstå logikken bak slutningsstatistikk er viktig når man ønsker å trekke gode konklusjoner.

- a. Gjør rede for den generelle logikken som ligger til grunn for slutningsstatistikk. Bruk gjerne begrepene populasjon, utvalg, tilfeldighet, standardfeil, nullhypotese, alternativ hypotese og statistisk signifikans.

*Oppgaven er omfattende, og det er viktig at kandidaten viser forståelse for logikken som ligger bak slutningsstatistikk og ikke kun definerer de ulike begrepene som er oppgitt i oppgaveteksten. Sentrale poeng er at når man trekker et tilfeldig utvalg fra en populasjon, så forventer man at verdiene man beregner i utvalget (slik som gjennomsnittsverdi) vil avvike fra populasjonens «sanne» verdi (slik som populasjonsgjennomsnittet). Trekker vi mange utvalg*

av samme størrelse fra samme populasjon, får vi en utvalgsfordeling (for f.eks. gjennomsnittsverdi), som viser fordeling av tilfeldige avvik fra den «sanne» populasjonsverdien. Denne fordelingen forventes å bli normalfordelt rundt populasjonsverdien om man trekker mange slike utvalg (jfr sentralgrenseteoremet). Standardfeilen er standardavviket til denne utvalgsfordelingen, og er en funksjon av spredning (standardavvik) i populasjonen og størrelsen på utvalget. Plusspoeng om kandidaten viser til normalfordelingskurven og den prosentvise fordelingen som ligger innenfor bestemte avstander til forventningen, for eksempel at 95 % av fordelingen vil ligge +/- 1.96 standardfeil fra gjennomsnittet i en slik fordeling.

Dette kan videreføres til testing av nullhypotesen - det er viktig å få fram at det er nullhypotesen vi tester i tradisjonell hypotesetesting. Kandidaten bør få fram at slutningsstatistikk i bunn og grunn dreier seg om sannsynlighet for at verdien (parameteret) vi finner i utvalget har oppstått tilfeldig, gitt at nullhypotesen er sann. Dersom sannsynligheten er tilstrekkelig lav, forkaster vi nullhypotesen og beholder den alternative hypotesen. Kriteriet for å avvise  $H_0$  kalles for alfa nivået. Som en konvensjon brukes et sannsynlighetsnivå på 5 % eller 1%. Er sannsynligheten for resultatet mindre enn dette, sier vi at funnet er statistisk signifikant. Plusspoeng hvis kandidaten nevner at 5% grensen er en konvensjon og at et signifikant resultat ikke nødvendigvis er en stor forskjell eller en sterk effekt.

Bayesiansk tilnærming til hypotesetesting er ikke gjennomgått på kurset, men er så vidt nevnt i læreboka. Det forventes ikke at kandidatene har med dette i oppgaven. Videre er kun one-sample Z-test gjennomgått på forelesningene denne gang, slik at det ikke forventes at kandidatene skal trekke inn testobservatorer som t og F.

- b. Hva menes med Type I- og Type II-feil? Hvordan er disse relatert til hverandre?

Type I-feil er feil man kan gjøre i en studie dersom man feilaktig forkaster en sann nullhypotese. P-verdien som beregnes i tradisjonell nullhypotesting viser sannsynligheten for å gjøre en Type I-feil dersom  $H_0$  er sann. Signifikansnivået vi setter på forhånd viser hvor mye stor sannsynlighet for Type I-feil vi er villige til å tolerere. Med økende p-verdi øker sjansen for type I feil, og jo lavere p-verdien settes jo mindre sjanse for type I feil. Men da øker sjansen for type II feil (kalles også beta,  $\beta$ ).

Type II feil er at man feilaktig beholder en falsk nullhypotese. Sannsynligheten for type II feil kommer an på flere ting:

- Valgt signifikansnivå for å forkaste  $H_0$ 
  - Jo strengere signifikansnivå, jo større sannsynlighet for type II-feil
- Hvor stort utvalg ( $n$ ) vi har
  - Jo lavere  $n$ , jo vanskeligere å forkaste  $H_0$
  - Jo større  $n$ , jo lettere er det å forkaste  $H_0$
- Hvor sterk effekten er/hvor stor forskjellen er
  - Jo sterkere effekt, jo lettere å forkaste nullhypotesen

Plusspoeng om kandidaten nevner at aldri være sikre om  $H_0$  eller  $H_a$  er sann eller feil, derfor er det viktig å balansere Type I og Type II feil opp mot hverandre. Ofte vil man ha et design der man akseptere en 5% sjanse for Type I feil, men vil samtidig kunne forkaste en falsk nullhypotese 80% av gangene. Sistnevnte kalles statistisk styrke (power, og kan også symboliseres som  $1-\beta$ ). For å beregne statistisk styrke før en studie man ha et anslag på hvor sterk en effekt/ hvor stor forskjell som forventes å være og se dette i lys av hvor stor utvalgsstørrelse man trenger for å kunne forkaste nullhypotesene 80% av gangene.

c. Hva er forskjellen mellom standardavvik og standardfeil?

*Standardavviket ( $s$ ) er et mål som brukes innen beskrivende statistikk, som gir et mål på spredning av skårer innen et datasett (kan også kalles innen et utvalg). Standardavviket er et mål på gjennomsnittlig avvik fra gjennomsnittet. Eller litt enklere sagt: hvor store avvik er det fra gjennomsnittsverdien i et datasett? Standardavviket beregnes gjennom å ta kvadratroten av variansen ( $s^2$ ) i skårene innen et datasett/ett utvalg.*

$$\text{Varians: } s^2 = \frac{\sum(X-M)^2}{n-1}$$

$$\text{Standardavvik: } s = \sqrt{s^2}$$

*Kandidatene kan vise til sin besvarelse av oppgave a) siden standardfeilen også er sentralt der. Standardfeilen er et mål som brukes innen slutningsstatistikken og brukes primært i prosessen med å teste nullhypoteser og beregne konfidensintervall. Hvis der er standardfeilen til en gjennomsnittsverdi det er snakk om, kan standardfeilen sees på som et mål på avvik fra den «sanne» populasjonsgjennomsnittet når man trekker mange utvalg av samme størrelse fra en populasjon. Se også oppgave a).*

*Plusspoeng hvis det nevnes at standardfeilen ikke noe annet enn standardavviket i utvalgsfordelingen. Videre er det positivt om kandidaten redegjør for som at standardfeilen er beregnet på bakgrunn av standardavviket i utvalget og utvalgsstørrelsen:*

$$s_M = \frac{s}{\sqrt{n}}$$

#### OPPGAVE 4

En gruppe forskere finner en interessant sammenheng der 25 % av den norske befolkningen opplever vårslapphet og økt tretthet når det går mot varmere og lysere dager.

- a. Beskriv forholdet mellom **A**: symptomer på slapphet og tretthet og **B**: lyse og varme dager om korrelasjonen mellom disse to variablene var positiv, negativ og kurvlineær. Ved positiv korrelasjon vil variablene gå i samme retning, at både A og B øker eller minker. Negativ korrelasjon oppstår når variablene går i motsatt retning, dvs at variabel A øker og variabel B minker. I dette eksempelet: Høyere slapphet og tretthet er relatert til lavere grad av lyse og varme dager. Og da vil det også være slik at når A går ned går B opp. Fint om kandidaten nevner noe om at styrken på en korrelasjon uttrykkes ved hjelp av en korrelasjonskoeffisient. Ved kurvlinær korrelasjon vil A og B ha både positiv og negativ korrelasjon i ulike faser av forholdet. Her er det bonus hvis kandidaten ser at korrelasjonen vil bli overordnet svak/ikkeeksiterende hvis lineære korrelasjonsmodeller benyttes, hvor tanken er at en perfekt korrelasjon gir en rett linje. Nullkorrelasjon viser ingen relasjon mellom A og B.

- b. En observert sammenheng mellom variabel A og variabel B kan skyldes innvirkning fra en tredje variabel: **Z**. Forklar med eksempler på mulige tredjevariabler som kan være årsaken bak sammenhengen mellom økt tretthet/slapphet og varmere/lysere dager. Bruk gjerne ulike kausalmodeller når du forklarer.

*Her står kandidaten fritt til å forklare med egne ord. Viktig å trekke frem korrelasjonsmetodens begrensninger. Ved antagelser om årsak må rivaliserende hypoteser utelukkes da disse kan forklare korrelasjonen, og det er vanskelig å gjennomføre når man studerer sammenhenger som opptrer i en naturlig setting. Mulige tredjevariabler her er for eksempel pollenallergi, senere leggetid pga lysere kvelder og derfor blir man mer trøtt, flere eksamener/høyere arbeidspress vårmesteret osv. Kandidaten kan bruke de ulike kausalmodellene gjennomgått på forelesning til å illustrere for eksempel at z er bakenforliggende, mellomliggende eller har moderatoreffekt.*

- c. Det viser seg at sammenhengen skyldes ren tilfeldighet og at disse variablene ikke har et kausalt forhold til hverandre. Hva kaller vi en slik sammenheng som enten skyldes en tredjevariabel eller oppstår ved ren tilfeldighet? Kom med tre eksempler fra din hverdag på slike sammenhenger.

*Dersom korrelasjon oppstår tilfeldig og uten sammenheng, sier man sammenhengen er spuriøs. Spuriøse sammenhenger skyldes ofte tredjevariabel (z), men kan også oppstå helt tilfeldig uten at en bakenforliggende variabel skaper sammenhengen. Et eksempel på en spuriøs sammenheng som skyldes tredjevariabel kan være mengden løv som faller fra trær og mengden klær man tar på seg (z=temperatur ute), eller sammenheng mellom forekomst av hodepine dagen etter man har vært fest (z=høyt alkoholinntak eller z= lite søvn). Eller helt vilkårlig sammenheng: hvite biler og antall ulykker. Dersom man har tilgang til nok data er det mulig å finne mengder av fenomener som samvarierer selv om det ikke finnes noen forbindelse mellom dem. Denne samvariasjonen kan da forsvinne ved for eksempel å gjøre flere datamålinger.*

- d. Regresjonsanalyse kan betraktes som en videreføring av korrelasjon. Forklar hvorfor, og bruk de to variablene over som eksempel.

*Ligningen for bivariat lineær regresjon kan skrives som:  $\hat{Y} = a + bX$ . Poenget med regresjonsanalyse er at det mulig å fastsette verdien av y når man kjenner verdiene på x-variabelen. Her er prediktor den uavhengige variabelen x og kriteriet den avhengige variabelen y. De fleste vil kanskje mest naturlig tenke seg at varmere og lysere dager er X og grad av vårslapphet/tretthet er Y, hvor man spør: Hva skjer med y når x øker med én enhet? Merk at vi ikke snakker om kausalitet her.*

Lykke til med sensuren av eksamensoppgavene!

Beste hilsen,

Maria Teresa Grønning Dale, Pål Ulleberg og Knut Inge Fostervold