

Forside

Det er kun ett riktig svar per oppgave. Riktig svar gir 1-8 poeng (se hver oppgave), mens feil svar gir minus 1 poeng. Hvis oppgaven ikke besvares gis det 0 poeng.

Tillatt hjelpemiddel: Utlevert kalkulator (ikke egen)

Vedlagt: Formelark og tabeller

STA200_1_Datanivå

(1 poeng per deloppgave)

Du har gjennomført en randomisert kontrollert studie hvor du ønsker å undersøke hvorvidt det er noen effekt av et spesielt treningsopplegg. For å gjøre dette rekrutterer du 200 deltakere. Før du randomiserer deltakerne til intervensjonsgruppe (de som skal følge treningsopplegget) og kontrollgruppe (de som skal trene som vanlig), kartlegger du en rekke faktorer hos deltakerne. Følgende faktorer (variabler) ble registrert:

- Alder (år)
- Høyde (centimeter)
- Vekt (kilogram)
- Sivil status (ugift, samboer, gift, skilt, enke/enkemann)
- Antall barn
- Aktivitetsnivå (meget lavt, lavt, middels, høyt og meget høyt)
- Fysisk form (maksimalt oksygenopptak VO₂max, uttrykt som milliliter oksygen opptatt per minutt per kg kroppsvekt)

Hvilket nivå av data ligger variabelen *alder* på?

- A Nominal
- B Kontinuerlig
- C Ordinal
- D Diskret

Hvilket nivå av data ligger variabelen *sivil status* på?

- A Nominal
- B Kontinuerlig
- C Ordinal
- D Diskret

Hvilket nivå av data ligger variabelen fysisk form på?

A Nominal

B Kontinuerlig

C Diskret

D Ordinal

Hvilket nivå av data ligger variabelen Antall barn på?

A Kontinuerlig

B Nominal

C Diskret

D Ordinal

STA200_2_Deskriptiv statistikk

(2 poeng per deloppgave)

Før du starter med intervensjonen (omtalt i oppgave 1), ønsker du å undersøke hvorvidt det er noen forskjeller i aktivitetsnivå og fysisk form mellom deltakerne i intervensjonsgruppen og kontrollgruppen. Det første du ønsker å gjøre er å beskrive innsamlede data.

Hvordan vil du beskrive variabelen Aktivitetsnivå (se beskrivelse av variabelen i oppgave 1)?

- A Gjennomsnitt og standardavvik
- B Median og kvartilavvik
- C Modus
- D Histogram

Hvordan vil du beskrive variabelen Fysisk form (se beskrivelse av variabelen i oppgave 1 – du vurderer variabelen til å være normalfordelt etter å ha sjekket histogrammet)?

- A Median og kvartilavvik
- B Modus
- C Gjennomsnitt og standardavvik
- D Kun gjennomsnitt

Hvilken statistisk test vil du utføre for å undersøke hvorvidt det er statistisk signifikant forskjell i variabelen Aktivitetsnivå mellom intervensjonsgruppen og kontrollgruppen?

- A Avhengig/parret t-test
- B Mann-Whitney
- C Wilcoxon
- D Uavhengig/uparet t-test

Hvilken statistisk test vil du utføre for å undersøke hvorvidt det er statistisk signifikant forskjell i variabelen Fysisk form mellom intervensjonsgruppen og kontrollgruppen?

- A Avhengig/parret t-test
- B Mann-Whitney
- C Wilcoxon
- D Uavhengig/uparet t-test

Seksjon 4

STA200_3_Feiltype

(2 poeng)

Etter å ha gjennomført undersøkelsen beskrevet i oppgave 1, finner du at det ikke var noen effekt av treningsintervensjonen. Dette kan være sannheten, men det kan også være at din undersøkelse av ulike årsaker ikke har klart å påvise en effekt som i virkeligheten finnes. Hvilken type feil har i så fall rammet din undersøkelse?

- A Type I-feil
- B Type II-feil
- C En systematisk målefeil
- D Du har valgt feil statistisk test

Seksjon 5

STA200_4_Hypoteser

(2 poeng)

Hva er et eksempel på en nullhypotese (H_0)?

- A Treningsregime hadde effekt for kvinner, men ikke menn
- B Det er ingen forskjell i fysisk form mellom intervensjons- og kontrollgruppe etter endt intervensjon
- C Intervensjonsgruppa hadde en statistisk signifikant framgang i fysisk form, sammenliknet med intervensjonsgruppa
- D Menn tjener mer penger enn kvinner

Seksjon 6

STA200_5_Statistisk signifikans

(3 poeng)

Hva betyr det at man setter grensen for statistisk signifikans til 5%?

- A Vi godtar at vi i 20% av tilfellene feilaktig vil beholde nullhypotesen
- B Vi godtar 95% sannsynlighet for å gjøre en forkastningsfeil
- C Vi godtar maksimalt 5% sannsynlighet for Type I-feil
- D I 5% av alle tilfeller vil vi konkludere med at det var en statistisk signifikant forskjell mellom to grupper

STA200_6_Hypotesetesting

(3 poeng)

Du leser følgende i en vitenskapelig artikkel: Det var en statistisk signifikant forskjell mellom gruppene ($p=0.00023$). Hvilket av utsagnene nedenfor er **IKKE** korrekt?

- A Det var en statistisk signifikant forskjell mellom gruppene og forfatterne var mer enn 99% sikker på at de ikke hadde begått en forkastningsfeil
- B Forskjellen mellom gruppene var stor og dermed betydningsfull
- C Det var en statistisk signifikant forskjell mellom gruppene og forfatterne var mer enn 95% sikker på at de ikke hadde begått en forkastningsfeil
- D Forskjellen er statistisk signifikant, men man kan ikke uttale seg om det er en betydningsfull forskjell eller ikke

STA200_7_Kji-kvadrat

(6 poeng)

Du jobber på en skole hvor du merker at noen undervisningsrom har dårligere inneklime enn andre rom, og ønsker å undersøke om det er noen sammenheng mellom dårlig inneklime og forekomst av luftveisplager. Du gjennomfører en undersøkelse hvor du spør 118 studenter hvorvidt de opplever noen form av luftveisplager (ja/nei), samt registrerer hvorvidt de hovedsakelig sitter i et rom med godt eller dårlig inneklime. Du systematiserer dataene dine i en 2*2 krysstabell som ser slik ut:

		Har opplevd luftveisplager	
		Nei	Ja
Eksponert for dårlig inneklime	Ja	50	30
	Nei	9	29

Følgende hypoteser ble satt opp:

- H_0 : Det er ingen sammenheng mellom eksponering for dårlig inneklime og forekomst av luftveisplager
- H_1 : Det er en sammenheng mellom eksponering for dårlig inneklime og forekomst av luftveisplager

Gjennomfør en kji-kvadrat test. Hva blir verdien til kji-kvadrat og hvordan vil du konkludere?

- A** | Kji-kvadrat ligger mellom 4 og 5, og H_0 forkastes
- B** | Kji-kvadrat ligger mellom 4 og 5, og H_0 beholdes
- C** | Kji-kvadrat ligger mellom 15 og 16 og H_0 forkastes
- D** | Kji-kvadrat ligger mellom 15 og 16 og H_0 beholdes
- E** | Kji-kvadrat ligger under 3.84 og H_0 beholdes

Seksjon 9

STA200_8_Standardavvik

(2 poeng)

Hva er **ikke riktig** om et standardavvik?

- A Standardavviket sier noe om spredningen i et datasett, og skal oppgis sammen med gjennomsnittet
- B Standardavviket angir forskjellen mellom to grupper
- C Standardavviket angir verdienes gjennomsnittlige avvik fra gjennomsnittet
- D Standardavviket er kvadratroten av variansen

Seksjon 10

STA200_9_Standardavvik

(3 poeng)

I tabellen nedenfor ser du en oversikt over 6 studenters målinger av tid på 40 meter sprint. Tiden er målt i antall sekunder.

Forsøksperson	Sekunder
1	6.4
2	6.6
3	7.2
4	5.4
5	5.2
6	7.9

Hva er standardavviket for dette utvalget?

- A 0.8
- B 1.0
- C 5.4
- D 1.5

STA200_10_T-test

(5 poeng)

Som student er du glad i pizza, men ikke veldig glad i å lage mat. Det betyr at du relativt ofte bestiller pizza levert på døren. I det området hvor du bor er det to ulike leverandører av pizza: Pizza Expressen og Pizza Palace. Du synes det virker som om Pizza Expressen leverer pizzaen noe raskere enn Pizza Palace og bestemmer deg for å gå vitenskapelig til verks for å teste ut din hypotese om at de to leverandørene skiller seg med tanke på leveringstid (antall minutter fra du bestiller til det ringer på døra). Etter 3 måneder med datainnsamling (pizzaleveranser fra de to leverandørene) har du fått **11 levering fra Pizza Expressen og 9 levering fra Pizza Palace**.

Gjennomsnittlig leveringstid i minutter var 18 minutter for Pizza Expressen og 20 minutter for Pizza Palace. Etter å benyttet uavhengig t-test kommer du fram til en **observert t-verdi på 2.8**. Finn antall frihetsgrader og trekk en konklusjon. Hvordan vil du konkludere (sett grense for statistisk signifikans til 5%)?

- A | Observert p-verdi var lavere enn kritisk verdi så jeg beholder H_0
- B | Observert p-verdi var lavere enn kritisk verdi så jeg forkaster H_0
- C | Observert p-verdi var høyere enn kritisk verdi så jeg beholder H_0
- D | Observert p-verdi var høyere enn kritisk verdi så jeg forkaster H_0

STA200_11_T-test

(4 poeng for oppgave a, 2 poeng for oppgave b)

En forsker har undersøkt hvordan en gruppe elever skårer på en konsentrasjonstest før og etter et undervisningsopplegg som forskeren har en hypotese om at kan øke elevenes evne til å konsentrere seg. Testen gir elevene en skåre (fra 0 til 100) som sier i hvor stor grad de er i stand til å løse ulike oppgaver som krever konsentrasjon. Test 1 ble gjennomført før undervisningsopplegget startet og test 2 ble gjennomført etter 3 måneder med dette undervisningsopplegget.

Da forskeren ikke kan stort om statistikk, kommer han til deg og ber om hjelp. Etter at du har kikket litt på dataene vurderer du datasettet til å være normalfordelt og datanivå til å være kontinuerlig. Resultatene ser du i tabellen under.

Elev	Test 1	Test 2
1	50	58
2	61	61
3	46	52
4	42	45
5	42	51
6	59	69

a) Gjennomfør en t-test (pass på at du velger riktig type t-test). Hvor stor er t?

A $t = ca - 3.8$

B $t = ca - 3.2$

C $t = ca - 2.2$

D $t = ca - 1.6$

b) Hvordan vil du konkludere?

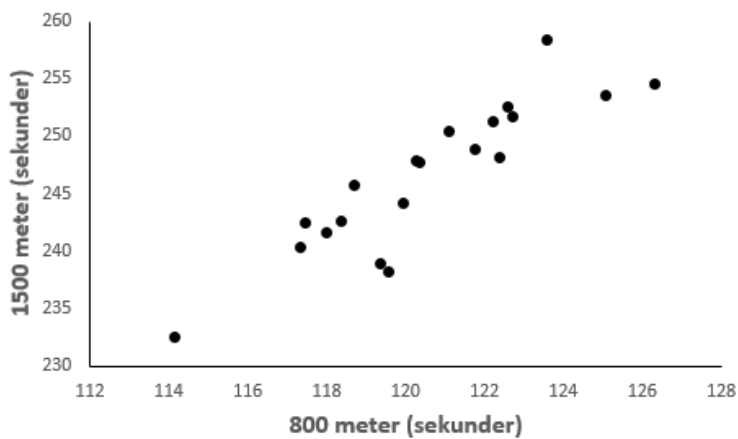
A H_0 forkastes

B H_0 beholdes

STA200_12_Korrelasjon

(2 poeng for oppgave a, 3 poeng for oppgave b)

Punktdiagrammet (scatterplott) under viser samvariasjonen mellom tid brukt på 800 meter og 1500 meter løp for et utvalg på 20 studenter. Du gjennomfører en Pearson korrelasjonstest og finner en korrelasjonskoeffisient $r=0.89$.



Figur 1. Samvariasjon mellom tid (sekunder) brukt på 800 meter (x-aksen) og 1500 meter (y-aksen)

a) Hvilket utsagn er riktig?

- A Det er en negativ korrelasjon mellom tid brukt på 800 meter og 1500 meter løp
- B Det er en positiv korrelasjon mellom tid brukt på 800 meter og 1500 meter løp
- C Man kan ikke vurdere en eventuell korrelasjon kun ved å se på plottet
- D Siden alle punktene ikke ligger på en rett linje er det ingen korrelasjon

b) Tenk deg at du løper 800 meter på 124 sekunder. Da kan du bruke en regresjonslikning basert på dataene i plottet til å predikere hvor lang tid du vil bruke på 1500 meter. Du kommer fram til en konstant (a) på 5.074 og en betakoeffisient (b; stigningstallet til regresjonslinjen) på 2.004. Ved å bruke denne informasjonen, hvor lang tid vil du bruke på å løpet 1500 meter (estimert)?

- A 253,6 sekunder
- B 222,5 sekunder
- C 248,5 sekunder
- D Det er ikke mulig å bruke slike data til å estimere noe

STA200_13_MW

(4 poeng for oppgave a, 2 poeng for oppgave b og c)

To grupper med studenter (en gruppe med økonomistudenter og en gruppe med idrettsstudenter) løper 200 meter. Resultatene ble notert i antall sekund og man ser av et histogram at dataene var noe skjevfordelte. Du ønsker å teste om økonomistudentene løper fortere enn idrettstudentene og bestemmer deg derfor for å gjennomføre en Mann-Whitney U-rangsum test for uavhengige grupper.

Gruppen med økonomistudenter (gruppe 1) bestod av 13 studenter og gruppen av idrettsstudenter (gruppe 2) bestod av 19 studenter. Du har rangert dataene, funnet korrigeret rang og regnet deg fram til at rangsum for gruppe 1 er 54 og rangsum for gruppe 2 er 81.

a) Bruk oppgitte rangsummer til å beregne Mann-Whitneys U-rangsum for begge gruppene. Hva er laveste U-verdi?

A	284
B	546
C	28
D	69

b) Slå opp i kritisk verdi for Mann Whitneys U. Hva er kritisk verdi for U?

A	65
B	45
C	72
D	52

c) Hvordan vil du konkludere?

A	H0 beholdes
B	H0 forkastes