

STUDIEÅRET 2016/2017

Individuell skriftlig eksamen

i

STA 200- Statistikk

Torsdag 27. april 2017 kl. 10.00-12.00

Hjelpemidler: Kalkulator og formelsamling som blir
delt ut på eksamen

Eksamensoppgaven består av 18 sider inkludert forsiden

Ordinær eksamen i STA 200, vår 2017

Det er kun ett riktig svar per deloppgave.

Riktig svar gir 1-10 poeng (spesifisert for hver deloppgave), **mens feil svar gir minus 1 poeng**. Hvis oppgaven ikke besvares gis det 0 poeng.

Sett en ring rundt riktig svar, og kommenter på eget ark ved vedhov. Pass på å merke eventuelle kommentarer med oppgavenummer og bokstad.

Kun utlevert kalkulator kan benyttes.

Vedlagt: Formelark og tabeller

Lykke til!

Oppgave 1

(Hver oppgave gir 2 poeng – 6 poeng totalt)

Folk oppgir forskjellige grunner til hvorfor de er fysisk aktive eller ikke. Tabellen nedenfor viser de fem hyppigst rapporterte årsakene til å være aktiv, oppgitt av et utvalg voksne og eldre.

Årsaker til å være fysisk aktiv	Mann (n)	Kvinne (n)
Forebygge helseplager	120	105
Komme i bedre form	89	80
Fysisk og psykisk velvære	75	74
Holde vekten nede	69	79

a) Hvilket nivå ligger data på?

1. Diskret
2. Ordinal
3. Kontinuerlig
4. Nominal

b) Hvordan kan man beskrive samlingen (sentralmålet) i et slikt datasett?

1. Ved å oppgi gjennomsnitt
2. Ved å oppgi typetall (modus)
3. Ved å oppgi median

c) Hvilken statistisk test skal man benytte for å undersøke hvorvidt det er forskjeller mellom kjønn og hva man rapporterer som årsaker til at man er fysisk aktiv?

1. Parret/avhengig t-test
2. Uparret/uavhengig t-test
3. Kji-kvadrat
4. Wilcoxon

Oppgave 2

(Hver oppgave gir 2 poeng – totalt 6 poeng)

Vi har målt oksygenopptaket (i ml/kg/min) til 10 studenter ved idrettslinjen på en videregående skole på Østlandet. Resultatene ser du i tabellen under.

Deltaker nr.	1	2	3	4	5	6	7	8	9	10
Oksygenopptak (ml/kg/min)	54	35	39	41	56	49	29	35	51	43

a) Hvilket nivå ligger data på?

1. Nominal
2. Kontinuerlig
3. Ordinal
4. Diskret

b) Hva er gjennomsnittlig (\bar{X}) oksygenopptak for dette utvalget?

1. $\bar{X} = 35.9$
2. $\bar{X} = 43.2$
3. $\bar{X} = 44.9$
4. $\bar{X} = 38.0$

c) Hva er standardavviket (SD) for oksygenopptak for dette utvalget?

1. SD = 7.0
2. SD = 8.0
3. SD = 9.0
4. SD = 10.0

Oppgave 3

(Hver deloppgave gir 2 poeng – totalt 10 poeng)

a) Nedenfor følger en rekke utsagn. Hvilket utsagn er riktig om KONTINUERLIGE data?

1. Før man behandler kontinuerlige data må man sjekke fordelingen til datasettet for å se om data er normalfordelte
2. Kontinuerlige data kan kun beskrives ved hjelp av median og kvartilavvik
3. For å analysere et datasett bestående av kontinuerlige data så benytter man seg av ikke-parametriske statistiske tester

b) Hva betyr det at signifikansnivået er 5% ($\alpha = 0.05$)?

1. Hvis nullhypotesen (H_0) er riktig vil vi i 5% av tilfellene godta denne
2. Vi har testet 5% av den totale populasjonen
3. Det avhenger helt av datasettet vårt og vil variere fra gang til gang

c) Nedenfor følger en rekke utsagn. Hvilket utsagn er et eksempel på en nullhypotese (H0)?

1. Det er en kjønnsforskjell i fysisk aktivitetsnivå mellom kvinner og menn
2. Det er ingen forskjell i grad i aktivitetsnivå mellom kvinner og menn
3. Finnes det en kjønnsforskjell i fysisk aktivitetsnivå?
4. Med 95% sikkerhet kan man si at det ikke er kjønnsforskjeller i fysisk aktivitetsnivå

d) Hvilket av alternativene nedenfor beskriver en Type I feil?

1. Behandlingene er ikke forskjellige, og vi finner ikke en forskjell
2. Behandlingene er ikke forskjellige, men vi finner en forskjell
3. Behandlingene er forskjellige, og vi finner en forskjell
4. Behandlingene er forskjellige, men vi finner ikke en forskjell

e) Hvilket av alternativene nedenfor beskriver en Type II feil?

1. Behandlingene er ikke forskjellige, og vi finner ikke en forskjell
2. Behandlingene er ikke forskjellige, men vi finner en forskjell
3. Behandlingene er forskjellige, og vi finner en forskjell
4. Behandlingene er forskjellige, men vi finner ikke en forskjell

Oppgave 4

(Oppgaven gir 8 poeng)

I kurset STA200 kan studentene løse e-læringsoppgaver til hver leksjon de har vært igjennom. I tabellen nedenfor finner du data om den mulige sammenhengen mellom omfang av eksamensforberedelser og prestasjon på eksamen i statistikk. Omfang av eksamensforberedelser er oppgitt som **antall løste sett med e-læringsoppgaver i forkant av eksamen** og eksamensprestasjon er oppgitt som **antall poeng på statistikkeksamen**. Data er oppgitt for fem tilfeldig valgte studenter.

Antall løste sett med e-læringsoppgaver	Antall poeng på statistikkeksamen
2	35
5	42
6	50
7	55
4	32

Regn ut Pearsons korrelasjonskoeffisient. Hvor stor er r ? (bruk maks 2 desimaler ved utregning)

1. $r = \text{ca } 0.88$
2. $r = \text{ca } 0.78$
3. $r = \text{ca } 0.95$
4. $r = \text{ca } -0.75$

Oppgave 5

(Deloppgave a gir 2 poeng, deloppgave b gir 4 poeng – totalt 6 poeng)

I forrige oppgave regnet dere ut korrelasjonen mellom antall løste oppgaver og eksamensresultat for 5 tilfeldige individer. I etterkant av eksamen ønsker du å bruke lineær regresjon til å modellere sammenhengen mellom antall løste oppgaver og eksamensresultat, og får tilgang til en database som inneholder samme informasjon som tabellen forrige oppgave, men for **alle** studentene som tok STA200 i 2017 (tenk deg at du får utlevert denne databasen, du trenger den ikke fysisk for å svare på spørsmålene under).

a) I regresjonslikningen du kommer fram til, hvilken variabel er den avhengige variabel (Y) og hvilken er den uavhengige variabel (X)?

1. Y: Eksamensresultat, X: Antall løste eksamensoppgaver
2. Y: Antall løste eksamensoppgaver, X: Eksamensresultat
3. Begge er uavhengige
4. Begge er avhengige

- b) Du regner ut at konstantleddet (a) blir 21 og stigningstallet (b) blir 4.4. Bruk denne informasjonen til å predikere eksamensresultatet til en student som har løst 3 sett med e-læringsoppgaver ved hjelp av regresjonslikningen.

$$Y = a + bX$$

1. 3 løste sett med e-læringsoppgaver gir et predikert eksamensresultat på **34.2 poeng**
2. 3 løste sett med e-læringsoppgaver gir et predikert eksamensresultat på **25.4 poeng**
3. 3 løste sett med e-læringsoppgaver gir et predikert eksamensresultat på **35.0 poeng**
4. Man kan ikke bruke lineær regresjon til å svare på dette spørsmålet

Oppgave 6

(Deloppgave a gir 6 poeng, deloppgave b og c gir 2 poeng hver – totalt 10 poeng)

En gruppe middelaldrende menn med høyt blodtrykk gjennomfører et treningsopplegg på 3 måneder. Opplegget består både av organiserte treninger og kostholdsveiledning. Deltakerne får blodtrykket sitt målt før og etter treningsintervensjonen. Verdiene (se tabellen) er oppgitt som systolisk blodtrykk (mmHg) for test 1 (før intervensjon) og test 2 (etter intervensjon). Data viser seg å være normalfordelte.

Følgende hypoteser formuleres:

H0: Gjennomsnittlig blodtrykk er likt ved test 1 og test 2

H1: Gjennomsnittlig blodtrykk er forskjellig ved test 1 og test 2

Person	Test 1 Blodtrykk (mmHg)	Test 2 Blodtrykk (mmHg)
1	171	171
2	178	176
3	174	171
4	173	172
5	175	173

a) Hvor stor er T?

1. $t = \text{ca } 1.1$
2. $t = \text{ca } 2.6$
3. $t = \text{ca } 5.4$
4. $t = \text{ca } 3.2$

b) Hva er kritisk verdi for t med 5% signifikansnivå?

1. 3.747
2. 2.776
3. 1.143
4. 3.812

c) Hva blir konklusjonen?

1. Vi forkaster H_0 (treningsintervensjonen hadde en signifikant effekt på systolisk blodtrykk)
2. Vi beholder H_0 (treningsintervensjonen hadde ikke effekt på systolisk blodtrykk)
3. Vi kan ikke konkludere kun basert på dette
4. Vi beholder H_0 , men ser en tendens til at intervensjonen har hatt en effekt

Oppgave 7

(Deloppgave a gir 6 poeng, deloppgave 2 gir 2 poeng – totalt 8 poeng)

Et høyt inntak av salt kan føre til høyt blodtrykk. En gruppe studenter ønsket å teste om dette var tilfelle i et utvalg middelaldrende menn.

De kartla kostholdsvaner og blodtrykk hos 441 menn i alderen 45-55 år. Studentene kategoriserte deltakerne etter hvorvidt de hadde et høyt eller normalt saltinntak, samt høyt eller normalt blodtrykk.

Studentene ønsket å bruke disse opplysningene til å teste ut om det er en sammenheng mellom saltinntak og blodtrykk. De setter opp de 4 kombinasjonsmulighetene i en 2*2 krysstabell:

	Høyt blodtrykk	Normalt blodtrykk	Totalt
Høyt saltinntak	118	95	213
Normalt saltinntak	95	133	228
Totalt	213	228	441

Følgende hypoteser ble formulert:

H0: Gruppene er like, det er ingen sammenheng mellom saltinntak og blodtrykk

H1: Gruppene er ulike, det er sammenheng mellom saltinntak og blodtrykk

a) Gjennomfør en Kji-kvadrat test. Hvor stor blir kji-kvadrat?

1. Kji-kvadrat = ca 8.3
2. Kji-kvadrat = ca 2.1
3. Kji-kvadrat = ca 15.3
4. Kji-kvadrat = ca 6.3

b) Hva blir riktig konklusjon?

1. Vi forkaster H_0 , det er en sammenheng mellom saltinntak og blodtrykk
2. Vi beholder H_0 , det er ingen sammenheng mellom saltinntak og blodtrykk
3. Det er ikke mulig å konkludere på grunn av at utvalget er for lite

Oppgave 8

(Oppgaven gir 2 poeng)

En bedrift på Østlandet gjennomførte en undersøkelse for å kartlegge arbeidsmiljøet blant de ansatte. Samtlige ansatte svarte på spørsmålet: «*Hvor godt trives du på arbeidsplassen?*»

Svaralternativene som de ansatte kunne krysse av var på en skala på 1-5 der 1 = svært dårlig og 5 = svært godt.

Da ledelsen fikk resultatene fra undersøkelsen så de at mange ikke trivdes på arbeidsplassen og de ønsket derfor å gjøre noe med det. De satte i gang en rekke tiltak og etter 6 måneder med ulike tiltak gjennomførte de undersøkelsen på nytt for å se om tiltakene hadde hatt en effekt.

Hvilken test ville du gjennomført for å undersøke om det var en statistisk signifikant forskjell i trivsel på arbeidsplassen før og etter tiltakene ble gjennomført?

1. T-test for uavhengige grupper
2. T-test for parrede observasjoner
3. Wilcoxon
4. Mann Whitney U-rangsum

Oppgave 9

(Deloppgave a gir 6 poeng, deloppgave b og c gir 2 poeng hver – totalt 10 poeng)

En gruppe fotballspillere (gruppe 1) og langrennsutøvere (gruppe 2) måler sitt maksimale oksygenopptak (i ml/kg/min). Dataene viser seg å være skjevfordelte. Ved hjelp av Mann-Whitney U-rangsum ønsker vi å undersøke om det er signifikant forskjell i maksimalt oksygenopptak mellom, fotballspillerne og langrenns utøvere. Vi setter opp følgende hypoteser:

H_0 : Gruppene er like

H_1 : Gruppene er forskjellige

Du har funnet korrigerert rang, og regnet deg fram til at rangsum for gruppe 1 er 44 og rangsum for gruppe 2 er 76.

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

1. $U = 8$

2. $U = 13.5$

3. $U = 15.0$

4. $U = 48$

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

1. 17

2. 19

3. 10

4. 15

c) Hva kan vi konkludere med?

1. Vi konkluderer med at H_0 forkastes

2. Vi kan ikke konkludere på bakgrunn av dette

3. Vi konkluderer med at H_0 beholdes

Oppgave 10

(Oppgaven gir totalt 4 poeng)

En kroppsøvingslærer på ungdomsskoletrinnet ønsker å bedre trivsel i kroppsøvingstimene. Læreren gjennomfører en undersøkelse hvor elevene selv rapporterer, på en skala fra 0 til 12, i hvor stor grad de trives i kroppsøvingstimene. Jo høyere verdi de rapporterer, jo bedre trives de. Etter at læreren har gjennomført undersøkelsen, iverksetter hun en rekke tiltak med hensikt å øke grad av trivsel. Tiltakene varer i 3 måneder. Etter tiltaksperioden gjennomfører hun en ny undersøkelse, og resultatene ser ut som følger:

Elev	Trivsel før tiltak (test 1)	Trivsel etter tiltak (test 2)
1	3	5
2	7	5
3	5	5
4	8	6
5	2	5
6	8	10
7	4	7
8	6	8

Følgende hypoteser settes opp:

- H_0 : Grad av trivsel ved test 1 er lik grad av trivsel ved test 2
- H_1 : Grad av trivsel ved test 1 er ikke lik grad av trivsel ved test 2

Etter at du har regnet ut Wilcoxons test for pardifferanser finner du at laveste rangsum er 6. Hvordan vil du konkludere?

1. Vi forkaster H_0 – tiltakene har hatt en effekt
2. Vi kan ikke konkludere
3. Vi beholder H_0 – tiltakene har ikke hatt en effekt