

STUDIEÅRET 2018/2019

Utsatt individuell skriftlig eksamen

i

STA 200- Statistikk

Tirsdag 27. august 2019 kl. 10.00-12.00

Hjelpemidler: kalkulator og formelsamling blir delt ut  
på eksamen

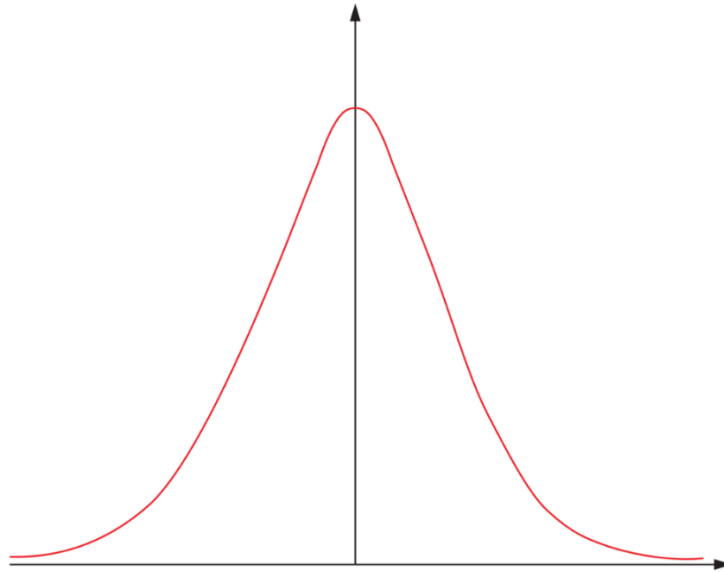
Eksamensoppgaven består av 8 sider inkludert forsiden

Sensurfrist: 17. september 2019

**DEL 1. Hva er statistikk?**

1. Hva er forskjellen på en kontinuerlig og en diskret variabel? **(1 poeng)**
2. Hva slags type variabel er kjønn (gutt/jente)? **(1 poeng)**
3. Hva er de to viktigste forutsetningene for å få et utvalg som representerer populasjonen av interesse? **(1 poeng)**
4. Se på de følgende to utsagnene. Er de like? Begrunn kort. **(1 poeng)**
  - a. Frank, Arne og Sigrid foretrekker å reise med bil. Mina foretrekker kollektivtransport.
  - b. 3 av 4 nordmenn foretrekker å reise med bil.

## DEL 2. Normalfordeling og hypotesetesting



1. Hvor stor andel av observasjonene kan forventes å falle innenfor  $\pm 2SD$  i en normalfordeling (svar i hele prosent)? **(1 poeng)**
2. Hvilken hypotese karakteriseres ved å påstå at *det finnes* en sammenheng eller forskjell? **(1 poeng)**
3. Hva er et standardavvik? **(1 poeng)**
4. Hva menes i statistikk med at signifikansnivået settes til 0,05? **(1 poeng)**
5. Hva innebærer det å gjøre en type-II feil? **(1 poeng)**

### DEL 3. Samlingsmål og spredningsmål

1. Hvilke egenskaper gjør medianen velegnet som samlingsmål til en kontinuerlig variabel med mange 'outliers'/ekstreme verdier eller med en meget skjev (skewed) fordeling? **(1 poeng)**
2. Hvilke egenskaper gjør standardavvik (SD) velegnet som spredningsmål for en normalfordelt, kontinuerlig variabel? **(1 poeng)**
3. Ta utgangspunkt i tallene under, som representerer antall daglige timer brukt til å se på TV for 7 individer. Rund av til en desimal der det er aktuelt.

3
0.5
1
2
5
4
4

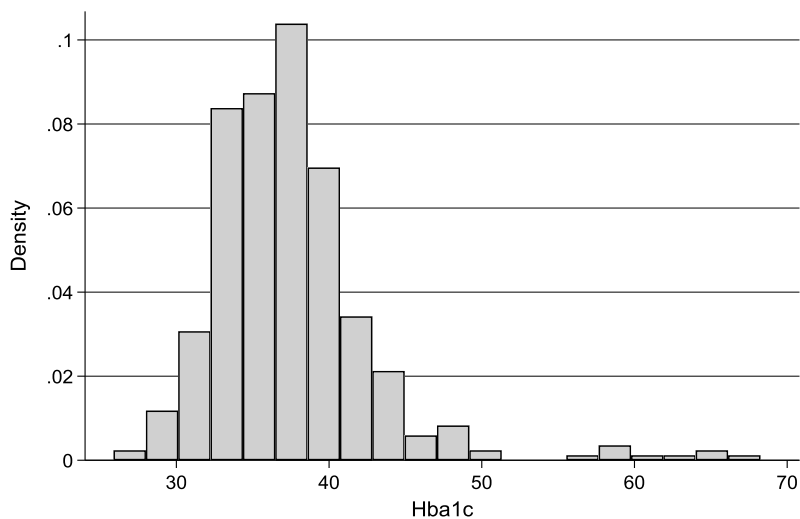
- a. Hva er gjennomsnittet? **(1 poeng)**
- b. Hva er standardavviket? **(1 poeng)**
- c. Hva er median? **(1 poeng)**

#### DEL 4. Forskjell mellom to grupper

1) Når man gjør en statistisk test av forskjeller i en kontinuerlig variabel mellom to grupper, kan observasjonene (data) være paret (paired samples) eller uparet (independent samples). Hva er forskjellen på paret og uparet data? **(1 poeng)**

2) Hba1c (et uttrykk for blodsukkernivå) er blitt målt hos 10 personer som ikke røyker og 10 personer som røyker daglig. Formålet er å undersøke om røyking påvirker blodsukker-reguleringen. Hva blir  $H_0$  i dette tilfellet? **(1 poeng)**

3) Histogrammet under viser fordelingen av Hba1c målt hos de 10 røykerne og de 10 ikke-røykerne. Argumenter for en statistisk test som vil være passende til å undersøke forskjell i Hba1c mellom røykere og ikke-røykere. Ta utgangspunkt i figuren under. **(2 poeng)**



4) Median Hba1c blant ikke-røykere var 35,0 mmol/l. Medianen for røykere var 39,0 mmol/l. Den statistiske testen resulterte i en tosidig p-verdi på 0.4. Hva må vi gjøre med  $H_0$  på bakgrunn av dette? Argumenter for svaret ditt. **(1 poeng)**

### DEL 5. Kji-kvadrat

I en studie har du samlet inn informasjon om vektstatus og diabetes. Du ønsker å undersøke om forekomsten av diabetes er forskjellig på tvers av de ulike BMI-kategoriene. For å besvare problemstillingen bestemmer du deg for å gjennomføre en kji-kvadrattest med signifikansnivå 0,05.

1. Hva slags variabeltype er det snakk om når man kan gjøre en kji-kvadrattest? **(1 poeng)**
2. Forklar kort hva kji-kvadrattesten undersøker. **(1 poeng)**
3. Under ser du en tabell med de observerte verdiene. Hvor mange prosent har diabetes i de ulike vektkategoriene (rund av til hele prosent)? **(1 poeng)**

	Diabetes		
Vektstatus	Nei	Ja	Totalsum
Normalvekt	50	6	56
Overvekt	54	12	66
Fedme	64	14	78
<b>Totalsum</b>	<b>168</b>	<b>32</b>	<b>200</b>

4. I tabellen for forventede verdier under er feltene tomme. Regn ut de forventede verdiene med riktig formel fra formelarket, og rund av til hele tall. **(1 poeng)**

	Diabetes		
Vektstatus	Nei	Ja	Totalsum
Normalvekt			56
Overvekt			66
Fedme			78
<b>Totalsum</b>	<b>168</b>	<b>32</b>	<b>200</b>

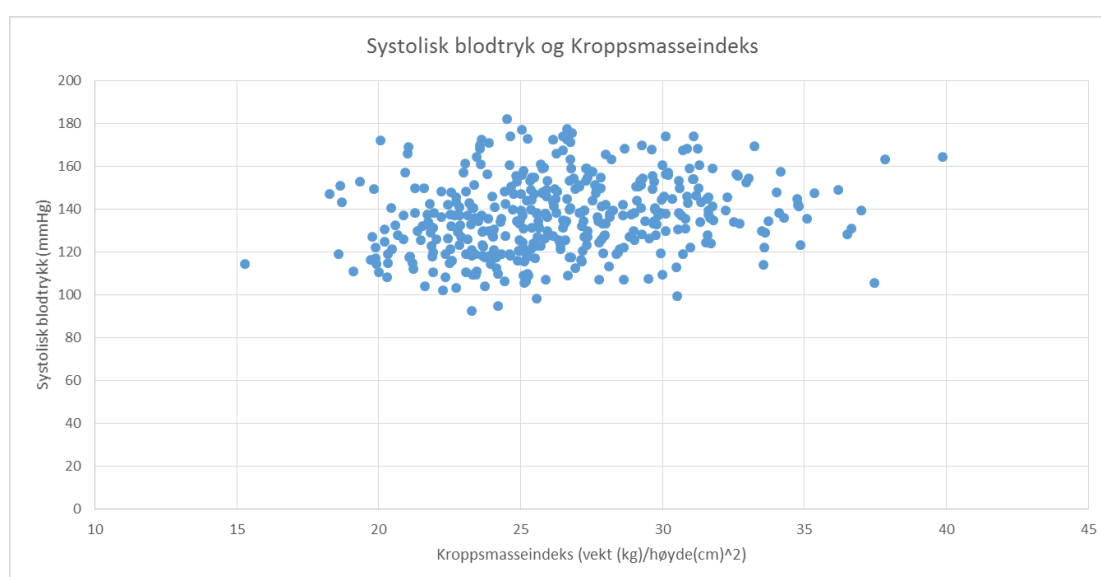
5. Du finner at p-verdien for kji-kvadrat er 0,445. Tilfredsstill denne testen forutsetningene for å gjøre kji-kvadrat? Hvordan vil du konkludere i forhold til problemstillingen? Utdyp kort. **(1 poeng)**

## DEL 6. Sammenheng mellom variabler

I en helseundersøkelse har 400 voksne og eldre mellom 40-70 år fått målt systolisk blodtrykk (SBP) og kroppsmasseindeks (BMI). Din oppgave som helseforsker er å undersøke sammenhengen mellom systolisk blodtrykk og BMI med korrelasjons- og regresjonsanalyse.

1) Hvilket datanivå (variabeltyper) er særlig egnet til korrelasjons- og regresjonsanalyse? **(1 poeng)**

2) Du lager et 'scatter-plot', eller punktdiagram, for å visualisere sammenhengen mellom systolisk blodtrykk og BMI. Beskriv sammenhengen, med utgangspunkt i scatter-plottet. **(1 poeng)**



3) En korrelasjonsanalyse mellom systolisk blodtrykk og BMI resulterer i en korrelasjonskoeffisient ( $r$ ) på 0,22. Hva betyr dette, og hva er en korrelasjonskoeffisient? **(1 poeng)**

4) Under ser du resultater fra en regresjonsmodell med systolisk blodtrykk som responsvariabel (Y) og BMI som forklaringsvariabel (X). Basert på regresjonskoeffisienten ( $b$ ); hva kan vi si om sammenhengen mellom systolisk blodtrykk og BMI? **(1 poeng)**

---

<i>Regresjonsstatistikk</i>	
Multipel R	0,219
R-kvadrat	0,048
Justert R-kvadrat	0,046
Standardfeil	17,131
Observasjoner	400

---

Variansanalyse

	<i>fg</i>	<i>SK</i>	<i>GK</i>	<i>F</i>	<i>Signifikans-F</i>
Regresjon	1	5887,8	5887,8	20,1	< 0,001
Residualer	398	116806,0	293,5		
Totalt	399	122693,7			

	<i>Koeffisienter</i>	<i>Standardfeil</i>	<i>t-Stat</i>	<i>P-verdi</i>	<i>Nederste 95%</i>	<i>Øverste 95%</i>
Skjæringspunkt ( <i>a</i> )	110,983	5,821	19,064	< 0,001	99,538	122,428
BMI ( <i>b</i> )	0,976	0,218	4,479	< 0,001	0,548	1,405

5) Hvordan vil du konkludere på spørsmålet om det er sammenheng mellom systolisk blodtrykk og BMI i dette tilfellet? Argumenter kort for svaret. **(1 poeng)**

6) Forklar hvorfor man, på bakgrunn av denne type data, bør være varsom med å konkludere med kausale sammenhenger. **(1 poeng)**