UNIVERSITETET I OSLO
DET MEDISINSKE FAKULTET

Sensorveiledning. Eksamen
ERN4000 STATISTIKK
Fredag 5. september 2026 kl. 09:00

Hjelpemidler

- Stata (på den stasjonære maskinen)
- Egen kalkulator med exp og ln, og uten trådløs kommunikasjon
- Stata Cheat Sheet
- Lærerbok: Aalen OO (red), Frigessi A, Moger TA, Scheel I, Skovlund E, Veierød MB. Statistiske metoder i medisin og helsefag, 2. utgave. Gyldendal Akademisk 2018.
  eller
  Aalen OO (red), Frigessi A, Moger TA, Scheel I, Skovlund E, Veierød MB. Statistiske metoder i medisin og helsefag. Gyldendal Akademisk 2006.

Kort forklaring til oppgavene

I oppgavene skal du foreta statistisk analyse av noen problemstillinger. I presentasjonen av resultatene er det tre statistiske mål som er viktige å få frem: estimat (f.eks. gjennomsnitt, andel, odds ratio, regresjonskoeffisient), konfidensintervall og p-verdi.

Oppgave 1

I datafilen data2026.dta finner du resultater fra en randomisert studie med 304 personer der man sammenlignet to intervensjoner, kostomlegging og et program for fysisk aktivitet, og undersøkte vektendring i løpet av 18 måneder.

Variabelnavn	**Beskrivelse**
intervensjon	Intervensjonsgruppe: 0 = Kostomlegging, 1 = Fysisk aktivitet
alder	Alder (år) ved inklusjon
kjønn	Kjønn: 0 = Kvinne, 1 = Mann
status	Samlivsform: 0 = Enslig, 1 = Samboende
høyde	Høyde (cm)
vekt0	Vekt (kg) ved inklusjon
vekt1	Vekt (kg) etter 18 måneder
vektendring	vekt0 minus vekt1
vektendring2	kategorisk variabel: 1 = vektendring ≥5 kg, 2 = vektendring <5 kg

a) *Undersøk om kjønn, samlivsform, alder, høyde og vekt er signifikant forskjellig i de to intervensjonsgruppene ved inklusjon. Oppgi resultatene i en tabell med samme oppsett som nedenfor og angi i fotnote til p-verdiene hvilke tester som er brukt for hver variabel (8). Skriv en kort tekst som begrunner valget av tester (4). Oppgi hvilke Stata kommandoer som ble brukt for å fylle inn i tabellen og for å velge tester (2).*

Variabel	Kost (n=150) n (%)	Fysisk aktivitet (n=154) n (%)	p-verdi
Kjønn			
Kvinne	85 (56.7)	99 (64.3)	
Mann	65 (43.3)	55 (35.7)	0.17[1]
Samlivsform			
Enslig	36 (24.0)	37 (24.0)	
Samboende	114 (76.0)	117 (76.0)	0.996[1]
	Gjennomsnitt (SD)	Gjennomsnitt (SD)	
Alder	65.0 (5.1)	64.7 (5.3)	0.62[2]
Høyde	170.9 (7.5)	169.7 (6.8)	0.15[2]
Vekt	95.4 (10.5)	94.7 (9.4)	0.53[2]

SD = standard avvik
[1]Kji-kvadrattest
[2]To-utvalgs t-test

Sammenligningen av de to gruppene mhp kjønn: Her har vi en 2x2 tabell og anvender kji-kvadrattest. Forutsetningen om at alle cellene i tabellen skal ha forventet antall større enn 5 er oppfylt.
Tilsvarende for samlivsform.

Sammenligning av de to gruppene mhp alder: Vi skal sammenligne to uavhengige grupper og normalitetsforutsetningen er oppfylt i begge grupper (histogram og qq/normalfordelingsplott i hver gruppe). Vi kan bruke to-utvalgs t-test.
Tilsvarende for høyde og vekt.

Stata:

```
tabulate kjønn intervensjon, column
tabulate kjønn intervensjon, chi2 expected

tabulate status intervensjon, column
tabulate status intervensjon, chi2 expected

ttest alder, by(intervensjon)
hist alder, by(intervensjon)
qnorm alder if (intervensjon==0)
qnorm alder if (intervensjon==1)

ttest høyde, by(intervensjon)
hist høyde, by(intervensjon)
qnorm høyde if (intervensjon==0)
qnorm høyde if (intervensjon==1)
```

```
ttest vekt0, by(intervensjon)
hist vekt0, by(intervensjon)
qnorm vekt0 if (intervensjon==0)
qnorm vekt0 if (intervensjon==1)
```

(Vi ba om tester og p-verdier her, men det er verdt å nevne at tabell 1 i en randomisert studie vanligvis er rent deskriptiv og ikke inkluderer p-verdier, i hht CONSORTstatement. Se item 25 avsnittet s. 41-43 i Hopewell et al. CONSORT 2025 explanation and elaboration: updated guideline for randomised trials. *BMJ* 2025;389:e081124.)

b) *Estimer forskjellen i vektendring mellom de to gruppene med 95% konfidensintervall (4). Er det signifikant forskjellig vektendring i de to gruppene? (2) Begrunn valget av test (2)*

Gruppen med kostomlegging har hatt noe større vektendring enn gruppen med fysisk aktivitet, men forskjellen er ikke signifikant. Gjennomsnittlig forskjell i vektendring er 1.05, 95% KI (-0.05, 2.15), p=0.06.

```
. ttest vektendring, by(intervensjon)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
    Kost |     150    4.673333    .4019568    4.922945    3.879061    5.467605
Fysisk a |     154     3.62013    .3893011      4.8311     2.85103    4.389229
---------+--------------------------------------------------------------------
Combined |     304    4.139803    .2808658    4.897062    3.587108    4.692497
---------+--------------------------------------------------------------------
    diff |            1.053203    .5594363               -.0476835     2.15409
------------------------------------------------------------------------------
    diff = mean(Kost) - mean(Fysisk a)                            t =   1.8826
H0: diff = 0                                     Degrees of freedom =      302

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.9696         Pr(|T| > |t|) = 0.0607          Pr(T > t) = 0.0304
```

Vi sammenligner to uavhengige grupper og normalitetsforutsetningen er oppfylt i begge grupper (histogram og qq plott i hver gruppe). Vi kan bruke to-utvalgs t-test.

Stata

```
hist vektendring, by(intervensjon)
qnorm vektendring if (intervensjon==0)
qnorm vektendring if (intervensjon==1)
```

Oppgave 2

I denne oppgaven skal du bruke datafilen kostgruppen.dta. Denne filen inkluderer bare de 150 personene som ble randomisert til gruppen som fikk kostomlegging. Variablene er de samme som i oppgave 1.

a) *Undersøk ved hjelp av lineær regresjon om det er signifikant sammenheng mellom vektendring og kjønn. (2) Angi også den estimerte regresjonskoeffisienten med 95% konfidensintervall. (2) Hvordan fortolkes regresjonskoeffisienten for kjønn? (2)*

Det er signifikant sammenheng mellom vektendring og kjønn, p=0.046. Regresjonskoeffisienten for kjønn er 1.616, 95% KI (0.030, 3.203). Gjennomsnittlig vektendring er 1.62 kg mer for menn enn kvinner.

Stata:
```
regress vektendring kjønn
```

b) *Undersøk ved hjelp av lineær regresjon om det er signifikant sammenheng mellom vektendring og alder. (2) Angi også den estimerte regresjonskoeffisienten med 95% konfidensintervall. (2)*

Det er ikke signifikant sammenheng mellom vektendring og alder, p=0.406. Regresjonskoeffisienten for alder er 0.066, 95% KI (-0.090, 0.222).

Stata:
```
regress vektendring alder
```

c) *Forskerne utførte en multippel regresjonsanalyse med vektendring som avhengig variabel, og alder og kjønn som uavhengige variabler. Utfør en slik multippel regresjonsanalyse og oppgi R^2. (2) Hva forteller R^2 deg? (2)*

R^2=0.039.
R^2 er forklart varians. Regresjonsmodellen med alder og kjønn forklarer 3.9% av variasjonen i vektendring.

Stata:
```
regress vektendring alder kjønn
```

d) *Lag en tabell som angitt nedenfor som viser estimerte regresjonskoeffisienter, konfidensintervaller og p-verdier for alder og kjønn. Både de ujusterte (2) og de justerte (2) resultatene.*

	Ujustert			Justert		
Variabel	Regresjons-koeffisient	95% KI	p-verdi	Regresjons-koeffisient	95% KI	p-verdi
Kjønn*	1.62	(0.03, 3.20)	0.05	1.89	(0.26, 3.52)	0.02
Alder	0.07	(-0.09, 0.22)	0.41	0.11	(-0.05, 0.27)	0.17

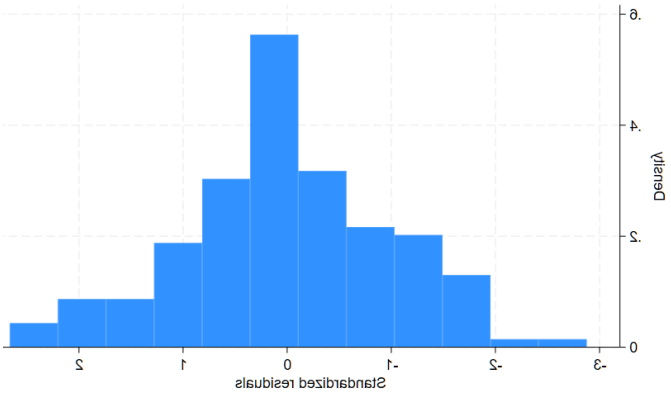
*Menn i forhold til kvinner

e) *Var forutsetningene tilstede for å gjøre analysen i d? Begrunn svaret. (3)*

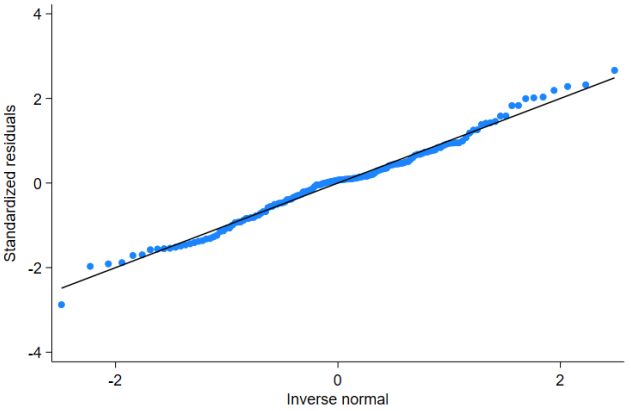
Forutsetningene er oppfylt.

```
regress vektendring alder kjønn
predict stres, rsta
```

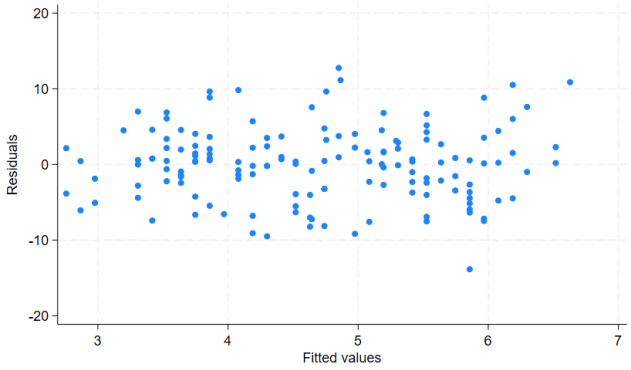
`hist stres`



`qnorm stres`



`rvfplot`



Oppgave 3

Vi skal nå bruke datafilen data2026.dta som vi brukte i oppgave 1 (alle 304 personene).

a) *Lag en 2x2 tabell med variabelen intervensjon i kolonner og variabelen vektendring2 i rader for kvinner og en tilsvarende tabell for menn. Prosentuer tabellene på en hensiktsmessig måte utfra studiens design. (4) Hva ser du fra disse tabellene? (4)*

I Stata kan du bruke
`bysort var1: tabulate var2 var3, option`
der du setter inn oppgavens variabelnavn for `var1, var2` *og* `var3`*, og bruker* `row` *eller* `column` *som* `option`.

```
. bysort kjønn: tabulate vektendring2 intervensjon, column

-> kjønn = K

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

vektendrin |     intervensjon
        g2 |      Kost  Fysisk ak |     Total
-----------+----------------------+----------
       >=5 |         34         42 |        76
           |      40.00      42.42 |     41.30
-----------+----------------------+----------
        <5 |         51         57 |       108
           |      60.00      57.58 |     58.70
-----------+----------------------+----------
     Total |         85         99 |       184
           |     100.00     100.00 |    100.00

-> kjønn = M

+-------------------+
| Key               |
|-------------------|
|     frequency     |
| column percentage |
+-------------------+

vektendrin |     intervensjon
        g2 |      Kost  Fysisk ak |     Total
-----------+----------------------+----------
       >=5 |         36         17 |        53
           |      55.38      30.91 |     44.17
-----------+----------------------+----------
        <5 |         29         38 |        67
           |      44.62      69.09 |     55.83
-----------+----------------------+----------
     Total |         65         55 |       120
           |     100.00     100.00 |    100.00
```

Hos kvinnene var det 40.0% som hadde en vektendring på ≥5 kg i kostomlegging gruppen og 42.4% i fysisk aktivitet gruppen. Hos mennene var det større forskjell på gruppene, 55.4% hadde en vektendring på ≥5 kg i kostomlegging gruppen og 30.9% i fysisk aktivitet gruppen.

b) *For kvinner skal du beregne relativ risiko (RR) for en vektendring på ≥5 kg for kostbehandling sammenlignet med fysisk aktivitet. (4) Beregn også 95% konfidensintervall. (4) For menn er RR=1.79 og 95% KI er (1.14, 2.81). Kommenter resultatene for begge kjønn. (2)*

Enten beregne i Stata:

```
. csi 34 42 51 57

                 |   Exposed   Unexposed  |      Total
-----------------+------------------------+-----------
           Cases |        34          42  |         76
        Noncases |        51          57  |        108
-----------------+------------------------+-----------
           Total |        85          99  |        184
                 |                        |
            Risk |        .4    .4242424  |   .4130435
                 |                        |
                 |      Point estimate    |    [95% conf. interval]
                 |------------------------+------------------------
 Risk difference |        -.0242424       |   -.1668062    .1183213
      Risk ratio |         .9428571       |     .666376    1.334051
 Prev. frac. ex. |         .0571429       |   -.3340511     .333624
 Prev. frac. pop |         .0263975       |
                 +-----------------------------------------------
                              chi2(1) =     0.11  Pr>chi2 = 0.7392
```

Eller regn ut selv:

	Kost (”eksponert”)	Fysisk aktivitet (”ikke-eksponert”)
≥5kg (syk)	34 (a)	42 (b)
<5kg (ikke-syk)	51 (c)	57 (d)
Total	85 (a+c)	99 (b+d)

$$RR = \frac{\hat{p}_1}{\hat{p}_0} = \frac{a/(a+c)}{b/(b+d)} = \frac{34/85}{42/99} = 0.942857143$$

95% konfidensintervall for RR

$(e^{nedre}, e^{øvre})$

$$nedre = \ln RR - 1.96 \cdot SE(\ln RR)$$
$$øvre = \ln RR + 1.96 \cdot SE(\ln RR)$$

$$\begin{aligned} SE(\ln RR) &= \sqrt{\tfrac{1}{a} - \tfrac{1}{a+c} + \tfrac{1}{b} - \tfrac{1}{b+d}} \\ &= \sqrt{\tfrac{1}{34} - \tfrac{1}{85} + \tfrac{1}{42} - \tfrac{1}{99}} \\ &= \sqrt{0.031355573} \\ &= 0.177075049 \end{aligned}$$

$$\begin{aligned} nedre &= \ln RR - 1.96 \cdot SE(\ln RR) \\ &= \ln 0.942857143 - 1.96 \cdot 0.177075049 \\ &= -0.375254307 \end{aligned}$$

Tilsvarende finnes *øvre*=0.288226596

Deretter (e^{nedre}, $e^{øvre}$) som gir (0.666, 1.334).

RR=0.94, 95% KI (0.67, 1.33)

Hos kvinner finner vi at estimert relativ risiko for en vektendring på ≥5 kg er 0.94 når vi sammenligner kostomlegging versus fysisk aktivitet. Det er ikke signifikant forskjell på de to intervensjonsgruppene mht vektendring hos kvinner (konfidensintervallet inkluderer 1). Hos mennene derimot, var estimert relativ risiko 1.79 og signifikant (konfidensintervallet inkluderer ikke 1), m.a.o. kostomlegging endret vekten mer enn fysisk aktivitet.

Oppgave 4

Vi fortsetter med datafilen data2026.dta (alle 304 personene) som vi brukte i oppgave 1 og 3.

I en randomiserte studie med en kontinuerlig variabel målt ved start og ved studieslutt, er det vanlig å bruke en regresjonsmodell der verdien ved studieslutt er avhengig variabel og verdien ved studiestart er uavhengig variabel sammen med gruppe. Vekt1 er da vår avhengige variabel med vekt 0 og intervensjon som uavhengige variabler. I tillegg skal vi her ta med kjønn som uavhengig variabel og interaksjon mellom intervensjon og kjønn. Utfør en slik multippel lineær regresjonsanalyse. Merk at siden vi har med et interaksjonsledd skal det være prefiks `i.` for både kjønn (`i.kjønn`) og intervensjon (`i.intervensjon`) i Stata. Er det signifikant interaksjon mellom intervensjon og kjønn? (4) Forklar hva det betyr? (4) Dere skal ikke tolke regresjonskoeffisientene.

```
regress vekt1 vekt0 i.intervensjon i.kjønn i.kjønn#i.intervensjon
```

```
      Source |       SS           df       MS      Number of obs   =       304
-------------+----------------------------------   F(4, 299)       =    307.96
       Model |  29071.7734         4  7267.94334   Prob > F        =    0.0000
    Residual |   7056.5366       299  23.6004568   R-squared       =    0.8047
-------------+----------------------------------   Adj R-squared   =    0.8021
       Total |    36128.31       303  119.235346   Root MSE        =     4.858

------------------------------------------------------------------------------
                vekt1 | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
----------------------+-------------------------------------------------------
                vekt0 |   .9902138   .0317367    31.20   0.000     .9277581    1.052669
                      |
         intervensjon |
     Fysisk aktivitet |   .0424749   .7183851     0.06   0.953    -1.371256    1.456206
                      |
                kjønn |
                Mann  |   -1.52082   .8582499    -1.77   0.077    -3.209796     .168155
                      |
    kjønn#intervensjon |
Mann#Fysisk aktivitet |   2.485817   1.143874     2.17   0.031     .2347537    4.736881
                      |
                _cons |  -3.080288   2.942441    -1.05   0.296    -8.870806    2.710229
------------------------------------------------------------------------------
```

Det er signifikant interaksjon mellom intervensjon og kjønn (p=0.03). M.a.o. ser det ut til at intervensjonen har ulik effekt for menn og kvinner.

Sensorveiledning:

Oppgave	Maksimal poengsum
1	22
2	21
3	18
4	8
Sum	69

Krav til bestått:

- Grensen for bestått er 45 av 69 mulige poeng.
- Besvarelsen skal vise at kandidaten har bred faglig kunnskap innen hele emnet og god kunnskap innen de mest sentrale områdene. Ingen av disse oppgavene er utenfor de mest sentrale områdene. Dette betyr at manglende/totalt utilfredsstillende besvarelse av enkelte oppgaver ikke kan kompenseres ved svært god besvarelse av andre.
- Bruken av fagterminologi er tilfredsstillende og besvarelsen er klar og presis.
- Kandidaten viser god faglig vurderingsevne og selvstendighet.