



Psykologisk institutt

Eksamensoppgave i PSY2017/PSYPRO4317

Statistikk og kvantitative forskningsmetoder

Faglig kontakt under eksamen: Christian Klöckner

Tlf.: 73 59 19 60

Eksamensdato: 11.12.014

Eksamenstid (fra-til): 09:00 – 13:00

Hjelpemiddelkode/Tillatte hjelpemidler:

Annen informasjon:

Målform/språk: Bokmål

Antall sider: 10

Antall sider vedlegg:

Kontrollert av:

Dato

Sign

ALLE OPPGAVER SKAL BESVARES!

Eksamen består av 4 oppgaver. Hvor mange poeng det er mulig å oppnå med hver oppgave og deloppgave er oppgitt bak hver oppgave.

For å få full uttelling i poeng for en oppgave må den være besvart helt riktig.

Det blir poengtrekk om svaret mangler mindre deler eller inneholder småfeil.

Svar som er feil i større omfang eller mangler store deler av svaret gir ikke poeng.

Makspoengtall er 87 poeng.

Totalpoeng tilsvarer følgende karakter

A -> 80-87 poeng

B -> 65-79 poeng

C -> 50-64 poeng

D -> 35-49 poeng

E -> 20-34 poeng

F -> mindre enn 20 poeng

Lykke til!

Oppgave 1 (totalt 16 poeng)

To uavhengige grupper av skoleelever ble spurt ved hjelp av et spørreskjema om hvordan de trives på skolen, en gruppe i første året på videregående skole, en gruppe i siste året på ungdomsskolen.

Trivsel ble målt med et spørsmål som ble besvart på et 7 poeng skala (1=svært dårlig, 7=svært bra).

I tabellen ser du verdiene til fem utvalgte elever fra hver gruppe:

Ungdomsskole	Videregående
5	7
6	6
6	2
4	3
5	4

- Beregn gjennomsnitt for de to gruppene, bestem median og modus (3 poeng: 0,5 poeng per riktig delsvar)
- Beregn standardavvik i de to gruppene og tolk dem (3 poeng: to poeng for riktig beregnede standardavvik, ett poeng for riktig tolkning)
- Beregn standardfeil til de to gjennomsnittene og 95% konfidensintervallene rundt dem. Hva forteller konfidensintervallene deg? (6 poeng: 1 poeng per riktig standardfeil og konfidensintervall, 2 poeng for riktig tolkning)
- Hvilken statistisk test vil du bruke for å teste om det er gjennomsnittsforskjell mellom de to gruppene? Gi en kort begrunnelse. (2 poeng)
- Nå antar at du hadde spurt samme gruppe elever før og etter overgangen fra ungdomsskole til videregående skole. Hvilken test ville du bruke nå? Gi en kort begrunnelse. (2 poeng)

Alle beregninger skal rundes på 3 sifrer etter komma.

Oppgave 2 (totalt 30 poeng)

I en studie med ble de følgende variabler målt i et utvalg på 435 mennesker:

PN – personlig norm (opplevd forpliktelse til å redusere bilbruk)

AC – Bevissthet av handlingskonsekvenser (hva har min bilkjøring med miljøproblemer å gjøre)

AN – Bevissthet av handlingsbehov (hvorfor er bilkjøring et problem for miljøet)

PBC – Handlingskontroll (hvor stor er evne til å kjøre kollektivt istedenfor bil)

SN – sosiale normer (forventninger av relevante andre mennesker til at man skal redusere bilbruk)

Nå er forskningsspørsmål, hvilke variabler har en signifikant påvirkning på nivået av PN og det blir kjørt en regresjonsanalyse med PN som avhengig variabel og AC, AN, PBC og SN som uavhengige variabler. Her er outputten fra SPSS:

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,432 ^a	,186	,179	,860	1,814

a. Predictors: (Constant), SN, PBC, AC, AN

b. Dependent Variable: PN

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	72,846	4	18,211	24,638	,000 ^b
	Residual	317,839	430	,739		
	Total	390,685	434			

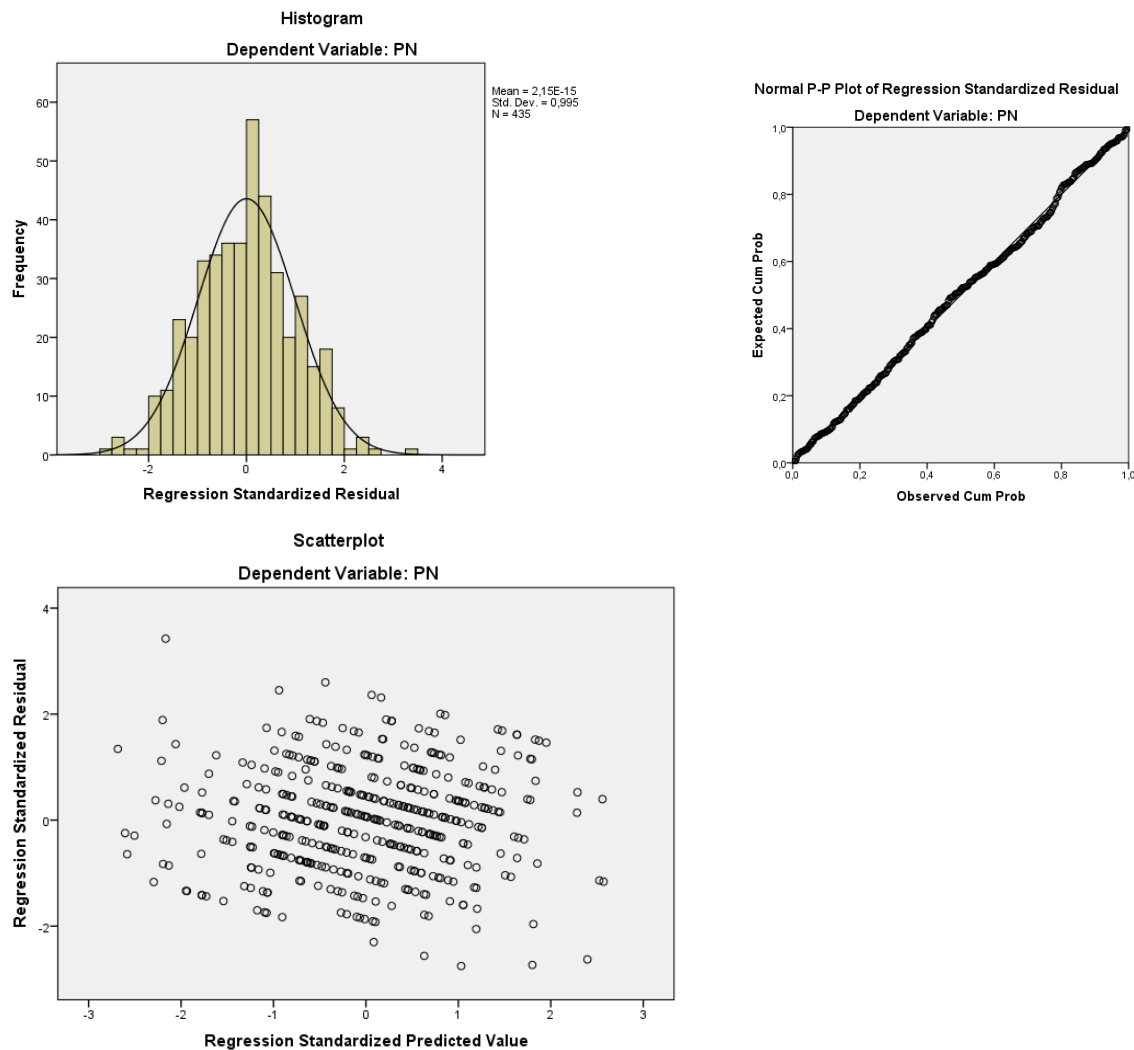
a. Dependent Variable: PN

b. Predictors: (Constant), SN, PBC, AC, AN

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	1,030	,217		4,743	,000		
	AC	,214	,059	,192	3,631	,000	,673	1,485
	AN	,221	,055	,214	3,996	,000	,662	1,511
	PBC	,005	,030	,007	,153	,879	,994	1,007
	SN	,153	,038	,180	4,041	,000	,958	1,044

a. Dependent Variable: PN



- A) Forklar kort hva en regresjonsanalyse er og hvorfor den ble brukt i denne analysen. (3 poeng)
- B) Hvilke forutsetningene om residualene må være på plass for å kjøre en regresjonsanalyse? Hvordan kan man teste om de er innfridd? Tolk informasjonen i outputen i forhold til disse forutsetningene. (8 poeng: 1 poeng per forutsetning, 0,5 poeng per korrekt nevnt test og 0,5 poeng per riktig tolkning av outputen)
- C) Hva er multikolaritet, hvorfor er den et problem for regresjonsanalysen og hvordan tolker du outputten i forhold til multikolaritetsproblematikken? (4 poeng)

- D) Hva tester ANOVA-testen i en regresjonskontekst? Hva er nullhypotesen? Hvordan tolker du outputen? (3 poeng)
- E) Skriv ned regresjonsligningen basert på regresjonstabellen og tolk koeffisientene. (4 poeng)
- F) Hva tester t-testen av de enkle koeffisientene i den tredje tabellen? Hva er nullhypotesen her? (3 poeng)
- G) Hvilken av de fire variablene har mest påvirkning på PN? Begrunn ditt valg kort. (2 poeng)
- H) Hva er R^2 og hvordan tolker du outputen? (3 poeng)

Oppgave 3 (totalt 35 poeng)

For å måle variablene i oppgave 2 ble det konstruert et spørreskjema med 13 spørsmål. Etterpå ble det kjørt en eksplorerende faktoranalyse (hovedkomponenteranalyse) med alle 430 personer som hadde svart på alle 13 spørsmål for å finne ut om spørsmålene lader på faktorer som kunne tolkes som å være de underliggende dimensjonene PN, AN, AC, PBC og SN.

- A) Bedøm om 430 deltakere er tilstrekkelig for en faktoranalyse. (2 poeng)
- B) I første trinn ble det laget en korrelasjonsmatrise, drøft om den er egnet til en faktoranalyse. (4 poeng)

Correlation Matrix

	AC1	AC2	AC3	PN1	PN2	PN3	AN1	AN2	AN3	SN1	SN2	PBC1	PBC2
Correlation AC1	1,000	,411	,379	,213	,150	,166	,221	,291	,381	,076	,084	-,015	-,004
AC2	,411	1,000	,501	,260	,210	,211	,242	,331	,488	,073	,070	,087	,068
AC3	,379	,501	1,000	,260	,189	,234	,222	,321	,558	,109	,144	,033	,026
PN1	,213	,260	,260	1,000	,379	,527	,150	,245	,265	,132	,210	-,005	,030
PN2	,150	,210	,189	,379	1,000	,352	,225	,288	,295	,119	,146	-,049	-,041
PN3	,166	,211	,234	,527	,352	1,000	,090	,168	,203	,115	,243	,070	,131
AN1	,221	,242	,222	,150	,225	,090	1,000	,320	,376	-,016	,019	,042	,037
AN2	,291	,331	,321	,245	,288	,168	,320	1,000	,514	,122	,127	,046	,004
AN3	,381	,488	,558	,265	,295	,203	,376	,514	1,000	,159	,159	,045	-,003
SN1	,076	,073	,109	,132	,119	,115	-,016	,122	,159	1,000	,472	,057	,016
SN2	,084	,070	,144	,210	,146	,243	,019	,127	,159	,472	1,000	,038	,091
PBC1	-,015	,087	,033	-,005	-,049	,070	,042	,046	,045	,057	,038	1,000	,529
PBC2	-,004	,068	,026	,030	-,041	,131	,037	,004	-,003	,016	,091	,529	1,000

- C) KMO-testen og Bartlett's-testen ga følgende resultater. Tolk dem og forklar hva testene går ut på. (6 poeng: 2 poeng for forklaring av hver test og 1 poeng for tolkningen)

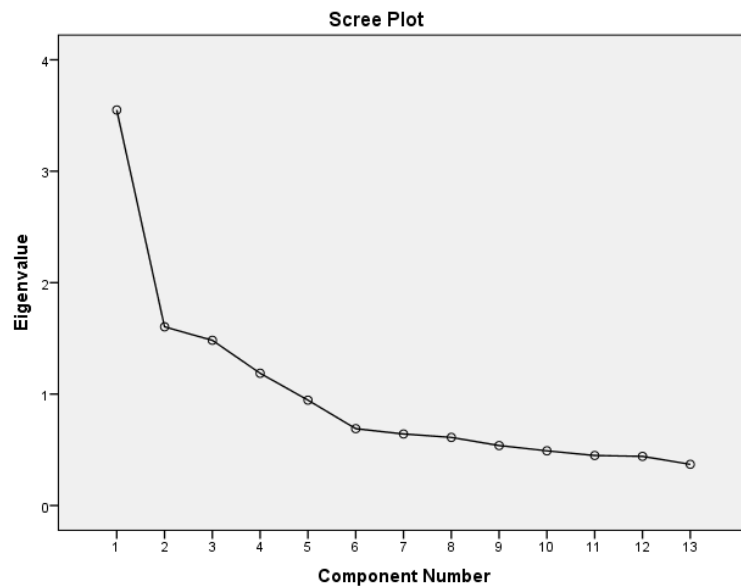
KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,774
Bartlett's Test of Sphericity	Approx. Chi-Square	1273,293
	df	78
	Sig.	,000

- D) Eigenverditabellen og Scree-plotten ser sånn ut. Argumenter for forskjellige antall faktorer som kunne trekkes ut basert på outputen og teoretiske forventninger. (8 poeng)

Total Variance Explained			
Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	3,549	27,302	27,302
2	1,604	12,336	39,638
3	1,483	11,405	51,044
4	1,187	9,130	60,174
5	,946	7,276	67,450
6	,689	5,304	72,753
7	,642	4,939	77,692
8	,611	4,701	82,393
9	,538	4,141	86,534
10	,491	3,776	90,310
11	,449	3,454	93,764
12	,441	3,394	97,158
13	,369	2,842	100,000

Extraction Method: Principal Component Analysis.

a. When components are correlated, sums of squared loadings are used to determine the number of components to retain.



- E) Fra forskerne ble det valgt å gå videre med fire faktorer. Det gir følgende tabell med kommunaliteter. Forklar kort hva kommunalitet av en variabel er i en faktoranalyse og tolk outputten. (4 poeng)

Communalities

	Initial	Extraction
AC1	1,000	,415
AC2	1,000	,543
AC3	1,000	,543
PN1	1,000	,674
PN2	1,000	,515
PN3	1,000	,699
AN1	1,000	,334
AN2	1,000	,443
AN3	1,000	,682
SN1	1,000	,749
SN2	1,000	,711
PBC1	1,000	,751
PBC2	1,000	,764

Extraction Method: Principal Component Analysis.

- F) Etter en oblimin rotasjon fikk forskerne følgende ladningsmønstrene. Forklar kort hva forskjellen mellom en ortogonal og en skjev rotasjon er og hva fordelene med den ene og den andre er. Forklar hvilken av typene en oblimin rotasjon er og hvorfor forskerne valgte den. Tolk ladningene i tabellen nedenfor (8 poeng: 2 poeng for forskjellen mellom skjev og ortogonal rotasjon, 2 poeng for begrunnelse av valget her og 4 poeng for tolkning av ladningene).

Pattern Matrix^a

	Component			
	1	2	3	4
AC1	,655	-,052	-,023	,041
AC2	,725	,093	,027	-,015
AC3	,725	,013	-,091	,006
PN1	,033	-,015	-,023	-,804
PN2	,105	-,140	,025	-,667
PN3	-,083	,134	-,041	-,839
AN1	,546	,035	,194	-,046
AN2	,631	-,024	-,059	-,064
AN3	,813	-,025	-,103	,008
SN1	,058	-,015	-,869	,067
SN2	-,010	,042	-,808	-,132
PBC1	,060	,864	-,017	,062
PBC2	-,026	,872	,000	-,053

Extraction Method: Principal Component Analysis.

Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 7 iterations.

- G) Outputen viser til slutt også følgende tabell. Hva viser den? Tolk den. (3 poeng)

Component Correlation Matrix

Component	1	2	3	4
1	1,000	,025	-,097	-,363
2	,025	1,000	-,041	-,030
3	-,097	-,041	1,000	,182
4	-,363	-,030	,182	1,000

Extraction Method: Principal Component Analysis.

Rotation Method: Oblimin with Kaiser Normalization.

Oppgave 4 (totalt 6 poeng)

Forklar hvorfor generaliserende statistikk trenger signifikanstesting ved å svare på følgende delspørsmål:

- A) Hva skjer hvis man drar et utvalg fra en populasjon? Hva er «sampling distribution of means»? (3 poeng)
- B) Hva er standardfeil og hvordan estimeres den? (3 poeng)

Sensorveiledning i PSY2017/PSY4317 – Høst 2014

Statistikk og kvantitative forskningsmetoder

ALLE OPPGAVER SKAL BESVARES!

Eksamen består av 4 oppgaver. Hvor mange poeng det er mulig å oppnå med hver oppgave og deloppgave er oppgitt i starten av hver oppgave. For å få full poeng for en oppgave må den være besvart helt riktig. Det blir poengtrekk i den graden svaret mangler mindre deler eller inneholder småfeil. Svar som er feil i større omfang eller mangler store deler av svaret gir ikke poeng. Maks poengtall er 87 poeng

Totalpoeng tilsvarer følgende karakter

A -> 80-87 poeng

B -> 65-79 poeng

C -> 50-64 poeng

D -> 35-49 poeng

E -> 20-34 poeng

F -> mindre enn 20 poeng

Oppgave 1 (totalt 16 poeng)

To uavhengige grupper av skoleelever ble spurt ved hjelp av et spørreskjema om hvordan de trives på skolen, en i første år videregående skole, en i siste år ungdomsskole. Trivsel ble målt med et spørsmål som ble besvart på et 7 poeng skala (1=svært dårlig, 7=svært bra). I tabellen ser du verdiene til fem utvalgte elever fra hver gruppe:

ungdomsskole	Videregående
5	7
6	6
6	2
4	3
5	4

- F) Beregne gjennomsnittene i de to gruppene, bestem median og modus (3 poeng – 0,5 poeng per riktig delsvar)

Gjennomsnitt ungdomsskole: $(5+6+6+4+5)/5 = 26/5 = 5,2$ 0,5 poeng

Gjennomsnitt videregående: $(7+6+2+3+4)/5 = 22/5 = 4,4$ 0,5 poeng

Median ungdomsskole: 4, 5, 5, 6, 6 -> median = 5 0,5 poeng

Median videregående: 2, 3, 4, 6, 7 -> median = 4 0,5 poeng

Modal ungdomsskole: 4, 5, 5, 6, 6 -> modal = 5 og 6 (like antall) 0,5 poeng

Modal videregående: 2, 3, 4, 6, 7 -> modal = 2, 3, 4, 6, 7 (alle har like antall) 0,5 poeng

- G) Beregne standardavvik i de to gruppene og tolk dem (3 poeng – to poeng for riktig beregnede standardavvik, ett poeng for riktig tolkning)

For å beregne standardavvik må vi først beregne varians for å holde det litt mer oversiktlig.

Varians ungdomsskole:

$$((5-5,2)^2 + (6-5,2)^2 + (6-5,2)^2 + (4-5,2)^2 + (5-5,2)^2) / (5-1) =$$

$$((-0,2)^2 + (0,8)^2 + (0,8)^2 + (-1,2)^2 + (-0,2)^2) / 4 =$$

$$(0,04 + 0,64 + 0,64 + 1,44 + 0,04) / 4 =$$

$$2,8 / 4 = 0,7$$

Standardavvik er kvadratroten av varians: kvadratroten av 0,7 = 0,837

1 poeng

Varians videregående:

$$((7-4,4)^2 + (6-4,4)^2 + (2-4,4)^2 + (3-4,4)^2 + (4-4,4)^2) / (5-1) =$$

$$((2,6)^2 + (1,6)^2 + (-2,4)^2 + (-1,4)^2 + (-0,4)^2) / 4 =$$

$$(6,76 + 2,56 + 5,76 + 1,96 + 0,16) / 4 =$$

$$17,2 / 4 = 4,3$$

Standardavvik er kvadratroten av varians: kvadratroten av 4,3 = 2,074 1 poeng

Standardavviket i gruppen fra videregående skole er betydelig større enn i gruppen fra ungdomsskole. Det betyr at elevene i videregående skole svarer mer forskjellige. 1 poeng

- H) Beregne standardfeil til de to gjennomsnittene og 95% konfidensintervallene rundt dem. Hva forteller konfidensintervallene deg? (6 poeng, 1 poeng per riktig standardfeil og konfidensintervall, 2 poeng for riktig tolkning)

Standardfeil er standardavvik delt på kvadratroten av N.

Standardfeil ungdomsskole:
 $0,837/\text{kvadratroten av } 5 = 0,837 / 2,236 = 0,374$ 1 poeng

Standardfeil videregående:
 $2,074/\text{kvadratroten av } 5 = 2,074 / 2,236 = 0,928$ 1 poeng

95% konfidensintervallene beregnes som gjennomsnitt pluss/minus 1,96 ganger standardfeil:

Ungdomsskole:
 nedre grensen: $5,2 - 1,96 * 0,374 = 4,467$
 øvre grensen: $5,2 + 1,96 * 0,374 = 5,933$ 1 poeng

Videregående:
 nedre grensen: $4,4 - 1,96 * 0,928 = 2,581$
 øvre grensen: $4,4 + 1,96 * 0,928 = 6,219$ 1 poeng

95% konfidensintervallene beskriver grensene innenfor dem 95% av alle gjennomsnitt i tilfeldige utvalg på 5 kommer til å ligge når populasjonsgjennomsnittet hadde vært den samme som den vi regnet ut i vårt utvalg. Bare 5% av gjennomsnittene i tilfeldige utvalg ville ligge utenfor disse verdiene. Konfidensintervallene beskriver derfor usikkerheten rund gjennomsnittsestimatene. 2 poeng

- I) Hvilken statistisk test vil du bruke for å teste om det er gjennomsnittsforskjell mellom de to gruppene? Gi en kort begrunnelse. (2 poeng)

Testen som egner seg til en sånn analyse er «independent samples t-test». 1 poeng
 (Er svaret ANOVA eller regresjonsanalyse med en 0/1 kodert prediktor er det også riktig)

Begrunnelsen er at de to gruppene er uavhengige av hverandre.

1 poeng

- J) Nå antar at du hadde spurt samme gruppe elever før og etter overgangen fra ungdomsskole til videregående skole. Hvilken test ville du bruke nå? Gi en kort begrunnelse. (2 poeng)

Spør man samme gruppe to ganger er testen «paired samples t-test»
(eller repeated ANOVA/RANOVA eller regresjon med robuste standardfeil)

1 poeng

Begrunnelsen er at de to gruppene er ikke uavhengig av hverandre og avhengigheten i residualene må korrigeres for.

1 poeng

Alle beregninger skal rundes på 3 sifrer etter komma.

Oppgave 2 (totalt 30 poeng)

I en studie med ble de følgende variabler målt i et utvalg på 435 mennesker:

PN – personlig norm (opplevd forpliktelse til å redusere bilbruk)

AC – Bevissthet av handlingskonsekvenser (hva har min bilkjøring med miljøproblemer å gjøre)

AN – Bevissthet av handlingsbehov (hvorfor er bilkjøring et problem for miljøet)

PBC – Handlingskontroll (hvor stor er evne til å kjøre kollektivt istedenfor bil)

SN – sosiale normer (forventninger av relevante andre mennesker til at man skal redusere bilbruk)

Nå er forskningsspørsmål, hvilke faktorer har en signifikant påvirkning på nivået av PN og det blir kjørt en regresjonsanalyse med PN som avhengig variabel og AC, AN, PBC og SN som uavhengige variabler. Her er outputten fra SPSS:

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,432 ^a	,186	,179	,860	1,814

a. Predictors: (Constant), SN, PBC, AC, AN

b. Dependent Variable: PN

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	72,846	4	18,211	24,638	,000 ^b
	Residual	317,839	430	,739		
	Total	390,685	434			

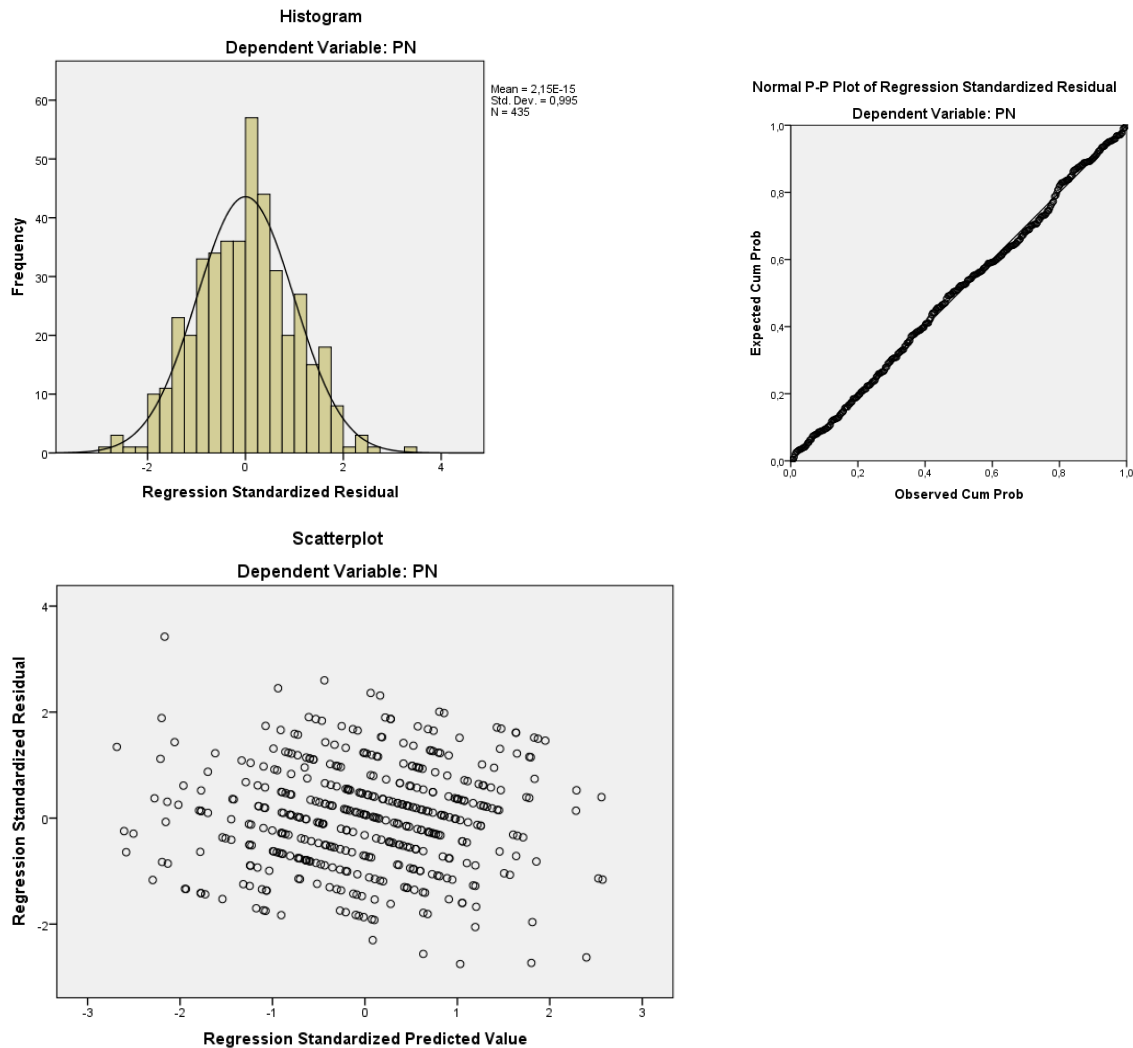
a. Dependent Variable: PN

b. Predictors: (Constant), SN, PBC, AC, AN

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	1,030	,217		4,743	,000		
	AC	,214	,059	,192	3,631	,000	,673	1,485
	AN	,221	,055	,214	3,996	,000	,662	1,511
	PBC	,005	,030	,007	,153	,879	,994	1,007
	SN	,153	,038	,180	4,041	,000	,958	1,044

a. Dependent Variable: PN



I) Forklar kort hva en regresjonsanalyse er og hvorfor den ble brukt i denne analysen. (3 poeng)

En regresjonsanalyse kvantifiserer sammenheng mellom en avhengig kontinuerlig variabel og en eller flere uavhengige variabler som kan være kontinuerlige eller dikotome. Mens man kvantifiserer sammenhengen mellom en uavhengig variabel og den avhengige variabelen kontrolleres for de andre uavhengige variablene.

I denne analysen var man ut etter å kvantifisere sammenhengen mellom den avhengige variabelen PN og de uavhengige variablene AC, AN, PBC og SN. Alle er kontinuerlige variabler. Derfor egner seg en regresjonsanalyse godt til å svare på forskningsspørsmålet.

3 poeng

- J) Hvilke forutsetningene om residualene må være på plass for å kjøre en regresjonsanalyse? Hvordan kan man teste om de er innfridd? Tolk informasjonen i outputten i forhold til disse forutsetningene. (8 poeng, 1 poeng per forutsetning, 0,5 poeng per korrekt nevnt test og 0,5 poeng per riktig tolkning av outputten)

Det finnes fire forutsetninger knyttet til residualene i en regresjonsanalyse:

- 1) residualene må ha et gjennomsnitt på 0 i populasjonen 1 poeng
- 2) residualene må være normalfordelte 1 poeng
- 3) residualene må være uavhengig av hverandre 1 poeng
- 4) residualene må ikke være forskjellig for forskjellige predikerte y -verdier (homoskedastisitet) 1 poeng

1) Kan ikke testes i populasjonen, men histogram av residualene i utvalget kan brukes for å få et inntrykk 0,5 poeng

2) kan sjekkes visuelt med P-P plotten, som må forme en linje så tett til diagonalen som mulig 0,5 poeng

3) testes med Durbin-Watson testen. Testverdien må ligge så nær 2 som mulig 0,5 poeng

4) Homoskedastisitet inspiseres visuelt med scatterplotten av standardiserte residualene over de standardiserte predikerte verdiene 0,5 poeng

Fra outputten er konklusjon følgende:

- 1) gjennomsnitt av residualene ligger nær null (histogram) 0,5 poeng
- 2) residualene er normalfordelte i følge P-P plotten (en fin linje) 0,5 poeng
- 3) Verdien fra Durbin-Watson testen er nær nok til 2 for å anta uavhengighet av residualene 0,5 poeng
- 4) Scatterplotten viser ikke et mønstre hvor variansen av residualene er systematisk forskjellig med økende eller minskende predikerte verdier, vi antar homoskedastisitet 0,5 poeng

- K) Hva er multikolinearitet, hvorfor er den et problem for regresjonsanalysen og hvordan tolker du outputten i forhold til multikolinearitetsproblematikken? (4 poeng)

Multikolinearitet er til stede når varians i en uavhengig variabel kan forklares fullstendig eller til en stor del av en eller flere av varians i de andre uavhengige variablene. Det betyr at variabelen klarer ikke å bidra unik varians i forklaring av den avhengige variabelen, fordi den deler for mye varians med andre prediktorer. Multikolinearitet er et problem i regresjonsanalysen fordi estimering av regresjonsvekt blir unøyaktig (standardeilen øker). Multikolinearitet testes med toleranseverdien eller variance inflation faktoren (VIF) som er rapportert i tabellen. Forskjellige forfattere nevner forskjellige grenseverdier, men selv om man bruker veldig strenge verdier (VIF

over 2,5 og tolerance under 0,4) så ligger verdiene i denne analysen på den sikre siden. Det foreligger derfor ingen multikolaritetsproblematikk. 4 poeng

- L) Hva tester ANOVA-testen i en regresjonskontekst? Hva er nullhypotesen? Hvordan tolker du outputten? (3 poeng)

ANOVA-testen (også kaldt OMNIBUS testen) tester nullhypotesen at regresjonsmodellen forklarer ikke mer varians enn en modell som bare bruker gjennomsnitt på PN som forventet verdi for alle personer. Det er derfor en generell test om de prediktorene som er tatt med i regresjonen til sammen forklarer en signifikant del av variansen rundt gjennomsnittet. I vårt tilfelle er testen signifikant på $p < .001$ nivå som sier at modellen med prediktorene er signifikant bedre enn gjennomsnittsmodellen. Det er derfor lov å gå videre og tolke regresjonsvektene i den neste tabellen. 3 poeng

- M) Skriv ned regresjonsligningen og tolk koeffisientene. (4 poeng)

regresjonsligningen er som følger:

$$PN = 1,030 + 0,214 * AC + 0,221 * AN + 0,005 * PBC + 0,153 * SN$$

2 poeng

Konstanten (b_0 , 1,030) beskriver skjæringspunktet, dvs. forventet PN for en person med 0 på alle prediktorene. 1 poeng

De andre koeffisientene beskriver stigningsvinkler:

Med hver enhet økning i AC øker PN med 0,214 enheter hvis alle andre prediktorer holdes konstant.

Med hver enhet økning i AN øker PN med 0,221 enheter hvis alle andre prediktorer holdes konstant.

Med hver enhet økning i PBC øker PN med 0,005 enheter hvis alle andre prediktorer holdes konstant.

Med hver enhet økning i SN øker PN med 0,153 enheter hvis alle andre prediktorer holdes konstant. 1 poeng

- N) Hva tester t-testen av de enkle koeffisientene i den tredje tabellen? Hva er nullhypotesen her? (3 poeng)

t-testene tester om en koeffisient (for eksempel b_0 eller b_1) er signifikant. Nullhypotesen er at den er 0 i populasjonen (dvs. at den har en annen verdi enn 0 bare fordi det er et tilfeldig avvik pga. at vi har et utvalg og ikke populasjonen). Blir t-testen signifikant så forkaster vi nullhypotesen og antar at verdien i populasjon ikke er 0. 3 poeng

- O) Hvilken av de fire variablene har mest påvirkning på PN? Begrunn ditt valg kort. (2 poeng)

For å sammenligne størrelsen av påvirkning av forskjellige variabler egner seg best de standardiserte koeffisientene. AN har den største og har derfor størst påvirkning, men med unntak av PBC er de veldig like uansett.

2 poeng

- P) Hva er R^2 og hvordan tolker du outputten? (3 poeng)

R^2 beskriver andel forklart varians. Så med andre ord, hvor mye av variansen rundt gjennomsnittet i PN kan forklares av mine fire prediktorer til sammen. Når man bruker mer en en prediktor bør den justerte R^2 en tolkes. Den sier at 17,9% av variansen er forklart. Det er en betydelig andel, men likevel er den allerstørste delen av variansen fortsatt ikke forklart.

3 poeng

Adjustert R^2 må nevnes for å få full poengtall.

Oppgave 3 (totalt 35 poeng)

For å måle variablene i oppgave 2 ble det konstruert et spørreskjema med 13 spørsmål. Etterpå ble det kjørt en eksplorerende faktoranalyse (hovedkomponenteranalyse) med alle 430 personer som hadde svart på alle 13 spørsmål for å finne ut om spørsmålene lader på faktorer som kunne tolkes som å være de underliggende dimensjonene PN, AN, AC, PBC og SN.

- H) Bedøm om 430 deltakere er tilstrekkelig for en faktoranalyse. (2 poeng)

En faktoranalyse krever mange deltakere for å føre til stabile resultater. Hva som betegnes som tilstrekkelig avhenger bland annet hvor mange store ladninger man finner og hvor stor kommunaliteten er. Men en tommelfingerregel er at man bør ha minst 300 deltakere. Derfor betraktes 430 som tilstrekkelig.

2 poeng

(Er svaret bare at 430 er nok uten begrunnelse gir det bare ett poeng!)

- I) I første trinn ble det laget en korrelasjonsmatrise, drøft om den er egnet til en faktoranalyse. (4 poeng)

Correlation Matrix														
	AC1	AC2	AC3	PN1	PN2	PN3	AN1	AN2	AN3	SN1	SN2	PBC1	PBC2	
Correlation	AC1	1,000	,411	,379	,213	,150	,166	,221	,291	,381	,076	,084	-,015	-,004
	AC2	,411	1,000	,501	,260	,210	,211	,242	,331	,488	,073	,070	,087	,068
	AC3	,379	,501	1,000	,260	,189	,234	,222	,321	,558	,109	,144	,033	,026
	PN1	,213	,260	,260	1,000	,379	,527	,150	,245	,265	,132	,210	-,005	,030
	PN2	,150	,210	,189	,379	1,000	,352	,225	,288	,295	,119	,146	-,049	-,041
	PN3	,166	,211	,234	,527	,352	1,000	,090	,168	,203	,115	,243	,070	,131
	AN1	,221	,242	,222	,150	,225	,090	1,000	,320	,376	-,016	,019	,042	,037
	AN2	,291	,331	,321	,245	,288	,168	,320	1,000	,514	,122	,127	,046	,004
	AN3	,381	,488	,558	,265	,295	,203	,376	,514	1,000	,159	,159	,045	-,003
	SN1	,076	,073	,109	,132	,119	,115	-,016	,122	,159	1,000	,472	,057	,016
	SN2	,084	,070	,144	,210	,146	,243	,019	,127	,159	,472	1,000	,038	,091
	PBC1	-,015	,087	,033	-,005	-,049	,070	,042	,046	,045	,057	,038	1,000	,529
	PBC2	-,004	,068	,026	,030	-,041	,131	,037	,004	-,003	,016	,091	,529	1,000

For at en korrelasjonstabell skal egne seg til en faktoranalyse må det være flere korrelasjoner som er større enn .300 (etter Field sin tommelfingerregel). Det er til stede i denne tabellen. I tillegg ønsker vi å validere et målinstrument, dvs. vi burde finne større korrelasjoner mellom variabler som hører til samme teoretiske konsept enn mellom variabler som hører til forskjellige konsepter. Stort sett finner vi også dette mønstre, men noen av korrelasjonene mellom variabler som skal måle det samme er relativt lave. 4 poeng

- J) KMO-testen og Bartlett's-testen ga følgende resultater. Tolk dem og forklar hva testene går ut på. (6 poeng: 2 poeng for forklaring av hver test og 1 poeng for tolkningen)

KMO and Bartlett's Test		
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,774
Bartlett's Test of Sphericity	Approx. Chi-Square	1273,293
	df	78
	Sig.	,000

Kaiser-Meyer-Olkin measure måler om korrelasjonene mellom enkelte par av variablene kan forklares av andre variabler i samme matrise. Med andre ord, den måler om det finnes mønstre av sammenheng i korrelasjonsmatrisen. KMO bør ligge over 0,5 og 0,8 er bra. 2 poeng

I vår analyse er KMO ,774, som er nær ,8 og derfor bra. Vi kan gå videre med faktoranalysen 1 poeng

Bartlett's testen tester nullhypotesen at korrelasjonsmatrisen er en identitetsmatrise, dvs. at alle korrelasjoner egentlig er 0 i populasjonen. En signifikant verdi betyr at vi forkaster denne nullhypotesen og kan anta at det finnes noen signifikante korrelasjoner. Testen er dessverre veldig sensitiv så at den blir egentlig alltid signifikant og forteller lite. 2 poeng

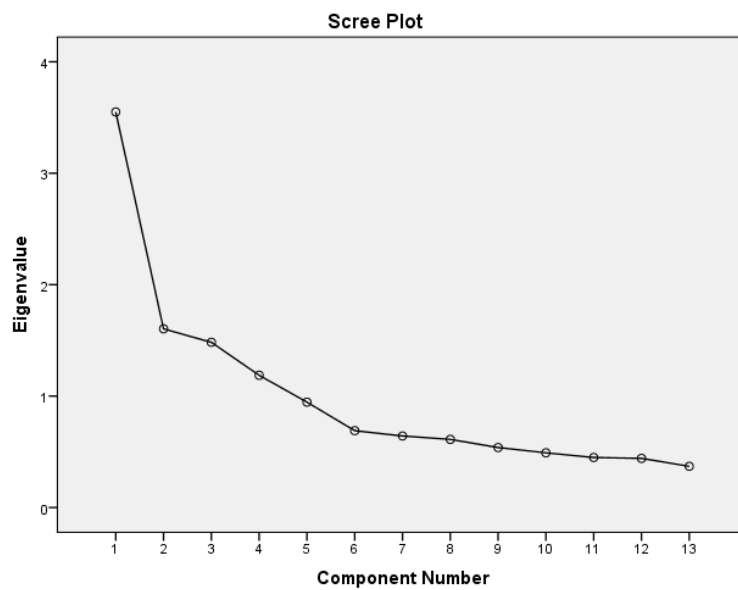
Vi har en signifikant Bartlett's test og kan derfor anta at vi har noen signifikante korrelasjoner i matrisen 1 poeng

- K) Eigenverditabellen og Scree-plotten ser sånn ut. Argumenter for forskjellige antall faktorer som kunne trekkes ut basert på outputten og teoretiske forventninger. (8 poeng)

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	3,549	27,302	27,302
2	1,604	12,336	39,638
3	1,483	11,405	51,044
4	1,187	9,130	60,174
5	,946	7,276	67,450
6	,689	5,304	72,753
7	,642	4,939	77,692
8	,611	4,701	82,393
9	,538	4,141	86,534
10	,491	3,776	90,310
11	,449	3,454	93,764
12	,441	3,394	97,158
13	,369	2,842	100,000

Extraction Method: Principal Component Analysis.

- a. When components are correlated, sums of squared load



L) Tre forskjellige kriterier kan brukes her for å bestemme antall faktorer:

1) Kaiser kriterium, som sier at man tar med alle faktorer som har en egenverdi større enn 1, dvs. fire faktorer her 2 poeng

2) Scree plot er litt tvetydig, siden kurven knekker to ganger, en gang etter 2 og en gang etter 6 faktorer. Knekken ved 6 faktorer er tydeligere, så Scree plotten forsvarer 2 faktorer bedre enn 6. 4 poeng

3) Siden faktoranalysen ble kjørt for å validere et måleinstrument og vi forventet fem faktorer, kan det også tenkes å trekke ut fem faktorer. 2 poeng

M) Fra forskerne ble det valgt å gå videre med fire faktorer. Det gir følgende tabell med kommunaliteter. Forklar kort hva kommunalitet av en variabel er i en faktoranalyse og tolk outputten. (4 poeng)

Communalities

	Initial	Extraction
AC1	1,000	,415
AC2	1,000	,543
AC3	1,000	,543
PN1	1,000	,674
PN2	1,000	,515
PN3	1,000	,699
AN1	1,000	,334
AN2	1,000	,443
AN3	1,000	,682
SN1	1,000	,749
SN2	1,000	,711
PBC1	1,000	,751
PBC2	1,000	,764

Extraction Method: Principal Component Analysis.

Kommunalitet er andelen varians i en variabel som er forklart av alle faktorer samtidig. Den beregnes ved å summere de kvadrerte ladningene av en variabel på alle faktorer. Det er ønskelig at kommunaliteter er stor så at en faktorløsning forklare mest mulig varians i hver variabel.

2 poeng

I analysen som foreligger her har de fleste variabler en stor del av varians forklart av faktorene, særlig SN1, SN2, PBC1 og PBC2. Noen variabler derimot har ganske lave kommunaliteter (særlig AN1, AC1 og AN2), dvs. det er ganske store deler av varians i disse variablene som ikke forklares av faktorene.

2 poeng

- N) Etter en oblimin rotasjon fikk forskerne følgende ladningsmønstrene. Forklar kort hva forskjellen mellom en ortogonal og en skjev rotasjon er og hva fordelene med den ene og den andre er. Forklar hvilken av typene en oblimin rotasjon er og hvorfor forskerne valgte den. Tolk koeffisientene i tabellen nedenfor (8 poeng: 2 poeng for forskjellen mellom skjev og ortogonal rotasjon, 2 poeng for begrunnelse av valget her og 4 poeng for tolkning av ladningene).

Pattern Matrix^a

	Component			
	1	2	3	4
AC1	,655	-,052	-,023	,041
AC2	,725	,093	,027	-,015
AC3	,725	,013	-,091	,006
PN1	,033	-,015	-,023	-,804
PN2	,105	-,140	,025	-,667
PN3	-,083	,134	-,041	-,839
AN1	,546	,035	,194	-,046
AN2	,631	-,024	-,059	-,064
AN3	,813	-,025	-,103	,008
SN1	,058	-,015	-,869	,067
SN2	-,010	,042	-,808	-,132
PBC1	,060	,864	-,017	,062
PBC2	-,026	,872	,000	-,053

Extraction Method: Principal Component Analysis.

Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 7 iterations.

Forskjellen mellom en ortogonal og en skjev rotasjon er at den sistnevnte tillater at faktorene er korrelerte etter rotasjonen, mens den førstnevnte beholder ukorrelerte faktorene. Fordelen med en ortogonal rotasjon er at faktorladningene er enklere å tolke, fordelen med en skjev rotasjon er at ladningsmønstrene blir tydeligere hvis faktorene i den virkelige verden burde være korrelerte.

2 poeng

Oblimin rotasjon er en skjev rotasjon og den ble trolig valgt pga. at det må forventes teoretisk at de forskjellige prediktorene korrelerer.

2 poeng

Tabellen viser at AC1-3 og AN1-3 har (relativt) store ladninger på første faktor. Det kan altså antas at disse spørsmålene måler noe felles. Det bekrefter delvis antakelser man hadde da man konstruerte måleinstrumentet, men det ser ut som AC og AN kan ikke skilles fra hverandre, spørsmålene er for like. SN1 og SN2 lader sterkt på faktor 3, PBC1 og PBC2 lader sterkt på faktor 2 og PN1-3 lader sterkt på faktor 4. Det finnes ingen krysslading av betydning (større enn .25). Så det kan konkluderes med at for PN, PBC og SN fungerer måleinstrumentet som forventet.

4 poeng

O) Outputten viser til slutt også følgende tabell. Hva viser den? Tolk den. (3 poeng)

Component Correlation Matrix

Component	1	2	3	4
1	1,000	,025	-,097	-,363
2	,025	1,000	-,041	-,030
3	-,097	-,041	1,000	,182
4	-,363	-,030	,182	1,000

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Tabellen viser korrelasjonsmønstre mellom faktorene etter en skjev rotasjon. De fleste korrelasjonene er ganske lave, men faktor 4 (PN) og faktor 1 (AN/AC) viser en mellomstor korrelasjon. Det kan peke på at en skjev rotasjon var riktig valg her. 3 poeng

Oppgave 4 (totalt 6 poeng)

Forklar hvorfor generaliserende statistikk trenger signifikanstesting ved å svare på følgende delspørsmål:

C) Hva skjer hvis man drar et utvalg fra en populasjon? Hva er «sampling distribution of means»? (3 poeng)

Siden vi som regel ikke har tilgang til hele populasjonen og det vil heller ikke være forskningsøkonomisk å spørre alle bruker vi vanligvis utvalg fra populasjoner. Når man trekker et utvalg vil utvalget ha et gjennomsnitt og et standardavvik i verdiene, men disse er ikke nødvendigvis de samme som verdiene i populasjonen, fordi pga. utvalgstrekking er noen personer med mens andre er ikke. Det fører til at både gjennomsnitt og standardfeil (og alle andre estimer basert på et utvalg) kommer til å variere hver gang man trekker et utvalg fra en populasjon. Trekker man mange utvalg så vil de fleste gjennomsnitt komme til å ligge nær populasjonsgjennomsnittet, men noen er ganske langt fra den (bare pga. tilfeldigheter). Denne fordelingen av gjennomsnitt kalles for sampling distribution av means. 3 poeng

D) Hva er standardfeil og hvordan estimeres den? (3 poeng)

Standardfeilen beskriver hvor mye gjennomsnittene i utvalg fra samme populasjon kommer til å variere rundt populasjonsgjennomsnittet når man drar mange utvalg. Standardfeilen beregnes som standardavvik av verdiene i populasjonen delt på kvadratroten av antall personer i utvalget (N). Siden standardavviket av verdiene i populasjonen er ukjent må den estimeres ut fra utvalget, noe som har vist seg i simulasjonsstudier til å være greit i utvalg større enn 30 personer. Så vi beregner standardfeilen ved å dele standardavvik av verdiene i utvalget på kvadratroten av utvalgsstørrelsen. 3 poeng