



Psykologisk institutt

Eksamensoppgave i PSY2017/PSYPRO4317

Statistikk og kvantitative forskningsmetoder

Faglig kontakt under eksamen: Christian Klöckner

Tlf.: 73 59 19 60

Eksamensdato: 29.05.2015

Eksamenstid (fra-til): 09:00 – 13:00

Hjelpemiddelkode/Tillatte hjelpemidler: Alle lærebøker og skrevne notater.

Kalkulator

Annen informasjon:

Målform/språk: Bokmål

Antall sider: 9

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 score 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.

Makspoengtall er 84 poeng.

Totalpoeng tilsvarer følgende karakter

A -> 72-84 poeng

B -> 59-71 poeng

C -> 46-58 poeng

D -> 33-45 poeng

E -> 19-32 poeng

F -> mindre enn 19 poeng

Lykke til!

Oppgave 1 (totalt 14 poeng)

Ved to bomstasjoner i Trondheim ble det gjennomført en spørreundersøkelse om hva folk synes om plassering av bomstasjonen.

Et spørsmål i undersøkelsen var om folk synes at plasseringen av bomstasjonen var rettferdig. Svaralternativene gikk fra 1=svært urettferdig til 7=svært rettferdig.

I analysen skilte man mellom mennesker som bodde innenfor bomringen og mennesker som bodde utenfor. I tillegg delte man gruppene inn etter politiske preferanser. I tabellen finner du svar fra fem deltakere i hver gruppe.

Politisk preferanse	Innenfor bomringen	Utenfor bomringen
FRP	2	1
	1	1
	3	1
	2	2
	2	1
AP	6	4
	5	3
	7	4
	4	3
	5	2

OBS: Alle beregninger skal rundes på 3 sifrer etter komma

- Beregne gjennomsnitt for alle fire delgruppe (FRP innenfor, FRP utenfor, AP innenfor, AP utenfor). Tolk gjennomsnittene. (3 poeng: 0,5 poeng per riktig gjennomsnitt; 1 poeng for riktig tolking)
- Beregne standardavvik i de fire gruppene og tolk dem (5 poeng: 1 poeng for hver riktig beregnet standardavvik, ett poeng for riktig tolkning).
- Hvilken statistisk test vil du bruke for å teste om det er gjennomsnittsforskjell mellom dem som bor innenfor bomringen og dem som bor utenfor (uansett politisk preferanse)?
Gi en kort begrunnelse. (2 poeng)
- Hvis du nå er interessert i å finne ut om forskjellen mellom innenfor og utenfor bomringen er forskjellig for FRP og AP tilhengere, hvilken test bruker du da?
Gi en kort begrunnelse. (4 poeng)

Oppgave 2 (totalt 35 poeng)

I en masteroppgave om studentenes deltakelse i resirkulering ble blant annet de følgende variablene målt i et utvalg på 700 mennesker:

BEH – Hvor mye av plastavfallet ditt resirkulerer du (1=ingenting, 7=alt)

INT – intensjon til å delta i plastresirkulering (1=svært lav, 7=svært høy)

PN – personlig norm (opplevd forpliktelse til å delta i resirkulering; 1=svært lav, 7=svært høy)

PBC – Handlingskontroll (hvor vanskelig er det å resirkulere plast; 1=svært enkelt, 7=svært vanskelig)

SN – sosiale normer (forventninger av relevante andre mennesker til at man skal resirkulere; 1=svært lav, 7=svært høy)

HABIT – Hvor sterk er resirkuleringsvanene, dvs. hvor automatisk er atferden? (1=svært svak, 7=svært sterk)

Masterstudenten kjørte en regresjonsanalyse med BEH som avhengig variabel og INT, PN, PBC, SN og HABIT 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	,706 ^a	,498	,494	1,535	1,989

a. Predictors: (Constant), HABIT, PBC, SN, INT, PN

b. Dependent Variable: BEHAVIOR

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1586,200	5	317,240	134,591	,000 ^b
	Residual	1598,097	678	2,357		
	Total	3184,297	683			

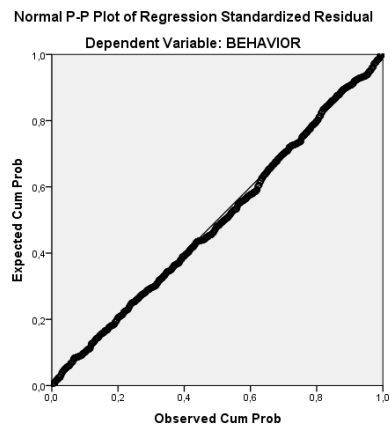
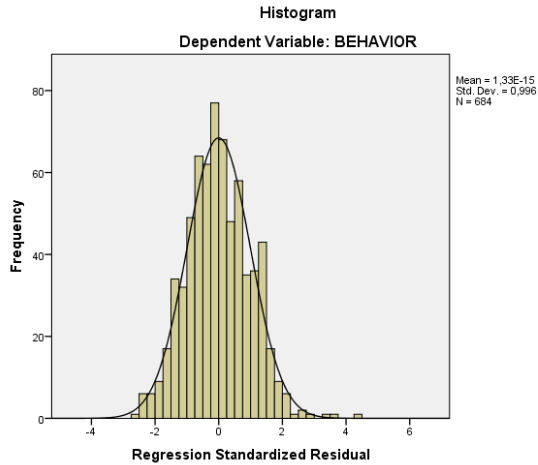
a. Dependent Variable: BEHAVIOR

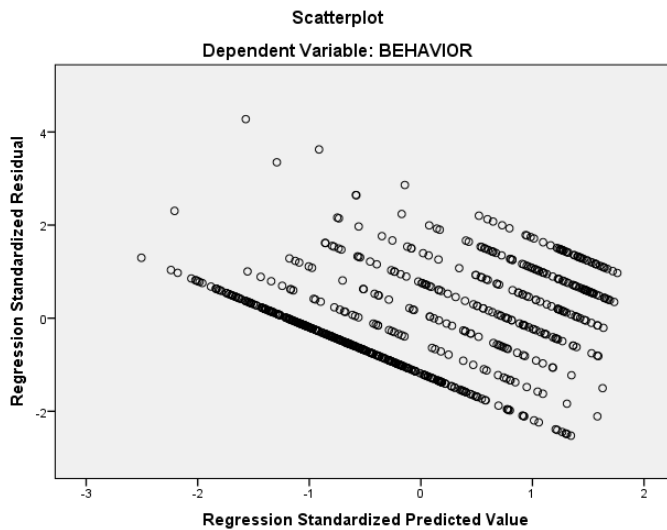
b. Predictors: (Constant), HABIT, PBC, SN, INT, PN

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
	B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1 (Constant)	2,606	,290		8,980	,000	2,036	3,176		
INT	,254	,042	,200	6,010	,000	,171	,338	,668	1,496
PBC	-,586	,030	-,563	19,537	,000	-,645	-,528	,892	1,121
SN	,047	,040	,035	1,170	,242	-,032	,126	,837	1,195
PN	,003	,058	,002	,049	,961	-,110	,116	,560	1,785
HABIT	,195	,050	,132	3,884	,000	,096	,294	,636	1,572

a. Dependent Variable: BEHAVIOR





- A) Hvorfor brukte masterstudenten en regresjonsanalyse?
Hvilket forskningsspørsmål besvarer denne analysen? (4 poeng)
- B) Skriv ned regresjonsligningen basert på regresjonstabellen og tolk koeffisientene. (4 poeng)
- C) Hvilke av de fem variablene har en signifikant påvirkning på BEH?
Hvilken har den største påvirkningen?
Begrunn ditt valg kort. (5 poeng)
- D) Hva forteller oss R^2 i tabell 1? (3 poeng)
- E) Hva betyr tallene under «95% confidence interval for B» i tabell 3? (4 poeng)
- F) Hva forteller tallene som står under «collinearity statistics»? (4 poeng)
- G) ANOVA-testen i andre tabellen er signifikant, hva betyr det? (3 poeng)
- H) Tolk informasjonen i outputen i forhold til om forutsetningene knyttet til residualene i en regresjonsanalyse er på plass. (8 poeng: 1 poeng per riktig nevnt forutsetning, 0,5 poeng per korrekt nevnt test og 0,5 poeng per riktig tolkning av outputen)

Oppgave 3 (totalt 29 poeng)

Det var studenter fra fire universiteter som deltok i studien nevnt i oppgave 2: NTNU, UIB, UIA, og UIO.

Nå var studentene interessert i å finne ut om plastresirkulering er forskjellig i de fire byene og hvis ja, hvilke byer var forskjellige.

Det ble kjørt en variansanalyse med BEH som avhengig variabel og universitet som uavhengig variabel.

Her er først de deskriptive statistikkene fra SPSS outputen:

Descriptives

BEHAVIOR

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
1 NTNU	200	3,75	2,219	,157	3,44	4,06	1	7
2 UIB	199	2,68	2,068	,147	2,39	2,97	1	7
3 UIA	100	2,86	2,165	,217	2,43	3,29	1	7
4 UIO	182	1,96	1,782	,132	1,70	2,22	1	7
Total	681	2,83	2,161	,083	2,67	2,99	1	7

- A) Hva forteller gjennomsnittene og konfidensintervallene deg om gruppeforskjell allerede før variansanalysetesten? (3 poeng)
- B) Drøft antall personer i gruppene i forhold til forutsetninger av en variansanalyse. (3 poeng)

Dette er output fra ANOVA testen:

ANOVA

BEHAVIOR

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	310,915	3	103,638	24,487	,000
Within Groups	2865,326	677	4,232		
Total	3176,241	680			

C) Tolk outputen. (4 poeng)

Levene testen leverer det følgende resultat:

Test of Homogeneity of Variances

BEHAVIOR

Levene Statistic	df1	df2	Sig.
12,205	3	677	,000

Robust Tests of Equality of Means

BEHAVIOR

	Statistic ^a	df1	df2	Sig.
Welch	25,415	3	322,330	,000
Brown-Forsythe	24,233	3	551,498	,000

a. Asymptotically F distributed.

D) Hva betyr dette? (4 poeng)

E) Når skal man bruke robuste tester (Welch, Brown-Forsythe) istedenfor den vanlige F-testen i variansanalysen?

Hvordan vurderer det i dette tilfelle? (5 poeng)

Til slutt ville studenten finne ut, hvilke universiteter var forskjellige. Derfor gjennomførte hun en post-hoc test:

Multiple Comparisons

Dependent Variable: BEHAVIOR

(I) UNIVERSITY University	(J) UNIVERSITY University	Mean Differen ce (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1 NTNU	2 UIB	1,067*	,215	,000	,51	1,62
	3 UIA	,890*	,267	,006	,20	1,58
	4 UIO	1,788*	,205	,000	1,26	2,32
2 UIB	1 NTNU	-1,067*	,215	,000	-1,62	-,51
	3 UIA	-,177	,261	,906	-,85	,50
	4 UIO	,722*	,197	,002	,21	1,23
3 UIA	1 NTNU	-,890*	,267	,006	-1,58	-,20
	2 UIB	,177	,261	,906	-,50	,85
	4 UIO	,898*	,254	,003	,24	1,56
4 UIO	1 NTNU	-1,788*	,205	,000	-2,32	-1,26
	2 UIB	-,722*	,197	,002	-1,23	-,21
	3 UIA	-,898*	,254	,003	-1,56	-,24

*. The mean difference is significant at the 0.05 level.

F) Forklar kort hva en post-hoc test er. (4 poeng)

G) Tolk resultatene fra post-hoc testen. (6 poeng)

Oppgave 4 (totalt 6 poeng)

Et viktig konsept i generaliserende statistikk er Nullhypotesetesting:

A) Hva er nullhypotesen og hvorfor trenger man den i signifikanstesting? (3 poeng)

B) Hva sier p-nivået i forhold til nullhypotesen? (3 poeng)

Eksamensoppgave i PSY2017/PSY4317 – Vår 2015

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 bak 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 små feil. Svar som er feil i større omfang eller mangler store deler av svaret gir ikke poeng. Maks poengtall er 83 poeng.

Totalpoeng tilsvarer følgende karakter

A -> 72-83 poeng

B -> 59-71 poeng

C -> 46-58 poeng

D -> 33-45 poeng

E -> 19-32 poeng

F -> mindre enn 19 poeng

Lykke til!

Oppgave 1 (totalt 13 poeng)

Ved to bomstasjoner i Trondheim ble det gjennomført en spørreundersøkelse om hva folk synes om plassering av bomstasjonen. Et spørsmål i undersøkelsen var om folk synes at plasseringen av bomstasjonen var rettferdig. Svaralternativene gikk fra 1=svært urettferdig til 7=svært rettferdig. I analysen skilte man mellom mennesker som bodde innenfor bomringen og mennesker som bodde utenfor. I tillegg delte man gruppene inn etter politiske preferanser. I tabellen finner du svar fra fem deltakere i hver gruppe.

Politisk preferanse	Innenfor bomringen	Utenfor bomringen
FRP	2	1
	1	1
	3	1
	2	2
	2	1
AP	6	4
	5	3
	7	4
	4	3
	5	2

OBS: Alle beregninger skal rundes på 3 sifrer etter komma

- A) Beregne gjennomsnitt for alle fire delgruppe (FRP innenfor, FRP utenfor, AP innenfor, AP utenfor). Tolk gjennomsnittene. (3 poeng: 0,5 poeng per riktig gjennomsnitt; 1 poeng for riktig tolking)

Gjennomsnitt FRP, innenfor:	$(2+1+3+2+2)/5 = 10/5 = 2,000$	0,5 poeng
Gjennomsnitt FRP, utenfor:	$(1+1+1+2+1)/5 = 6/5 = 1,200$	0,5 poeng
Gjennomsnitt AP, innenfor:	$(6+5+7+4+5)/5 = 27/5 = 5,400$	0,5 poeng
Gjennomsnitt AP, utenfor:	$(4+3+4+3+2)/5 = 16/5 = 3,200$	0,5 poeng

Opplevelsen at plasseringen av bomringen er rettferdig er mest utpreget blant AP-tilhengere som bor innenfor bomringen, mens den er minst utpreget blant FRP tilhengere sin bor utenfor. FRP tilhengere som bor innenfor betrakter bomringen som fortsatt urettferdig, men litt mer rettferdig enn de som bor utenfor. AP tilhengere, som bor utenfor ligger mellom FRP innenfor og AP innenfor.

1 poeng

- B) Beregne standardavvik i de fire gruppene og tolk dem (4 poeng: tre poeng for riktig beregnede standardavvik, ett poeng for riktig tolkning).

Varians FRP, innenfor:

$$\begin{aligned} & ((2-2,0)^2+(1-2,0)^2+(3-2,0)^2+(2-2,0)^2+(2-2,0)^2)/(5-1) = \\ & ((0)^2+(-1)^2+(1)^2+(0)^2+(0)^2)/4= \\ & (0+1+1+0+0)/4= \\ & 2/4=0,5 \end{aligned}$$

Standardavvik er kvadratrotten av varians: kvadratrotten av 0,5 = 0,707

1 poeng

Varians FRP, utenfor:

$$\begin{aligned} & ((1-1,2)^2+(1-1,2)^2+(1-1,2)^2+(2-1,2)^2+(1-1,2)^2)/(5-1) = \\ & ((-0,2)^2+(-0,2)^2+(-0,2)^2+(0,8)^2+(-0,2)^2)/4= \\ & (0,04+0,04+0,04+0,64+0,04)/4= \\ & 0,8/4=0,20 \end{aligned}$$

Standardavvik er kvadratrotten av varians: kvadratrotten av 0,2 = 0,447

1 poeng

Varians AP, innenfor:

$$\begin{aligned} & ((6-5,4)^2+(5-5,4)^2+(7-5,4)^2+(4-5,4)^2+(5-5,4)^2)/(5-1) = \\ & ((0,6)^2+(-0,4)^2+(1,6)^2+(1,4)^2+(-0,4)^2)/4= \\ & (0,36+0,16+2,56+1,96+0,16)/4= \\ & 5,2/4=1,3 \end{aligned}$$

Standardavvik er kvadratrotten av varians: kvadratrotten av 1,3 = 1,140

1 poeng

Varians AP, utenfor:

$$\begin{aligned} & ((4-3,2)^2+(3-3,2)^2+(4-3,2)^2+(3-3,2)^2+(2-3,2)^2)/(5-1) = \\ & ((0,8)^2+(-0,2)^2+(0,8)^2+(-0,2)^2+(-1,2)^2)/4= \\ & (0,64+0,04+0,64+0,04+1,44)/4= \\ & 2,8/4=0,7 \end{aligned}$$

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

1 poeng

Standardavviket er et mål på spredning rundt gjennomsnittet i et utvalg. I en normalfordeling vil omtrent 68% av alle verdier ligge mellom +/- et standardavvik. I vårt tilfelle er det tydelig at det er større spredning av verdiene i AP gruppene (særlig innenfor bomringen) enn i FRP gruppene, som er mer enig i bedømmelsen at bomringen er plassert urettferdig.

1 poeng

- C) Hvilken statistisk test vil du bruke for å teste om det er gjennomsnittsforskjell mellom dem som bor innenfor bomringen og dem som bor utenfor (uansett politisk preferanse)? Gi en kort begrunnelse. (2 poeng)

Siden det er gjennomsnitt i bare to uavhengige grupper som skal sammenlignes er t-testen riktig valg (enveis ANOVA eller en regresjon med dummy koding av gruppe er også mulig og gir full poenguttelling) 2 poeng

- D) Hvis du nå er interessert i å finne ut om forskjellen mellom innenfor og utenfor bomringen er forskjellig for FRP og AP tilhengere, hvilken test bruker du da? Gi en kort begrunnelse. (4 poeng)

Siden alle gruppene er uavhengige kan spørsmålet besvares med en toveis ANOVA som inkluderer de to hovedeffektene og samspill. Signifikant samspill mellom politisk preferanse og bosted vil da gi svar. (en regresjonsanalyse med dummy koding av de to grupperingsvariablene og samspill-led av de to er også korrekt og gir full uttelling). 4 poeng

Oppgave 2 (totalt 35 poeng)

I en masteroppgave om studentenes deltakelse i resirkulering ble blant annet de følgende variablene målt i et utvalg på 700 mennesker:

BEH – Hvor mye av plastavfallet ditt resirkulerer du (1=ingenting, 7=alt)

INT – intensjon til å delta i plastresirkulering (1=svært lav, 7=svært høy)

PN – personlig norm (opplevd forpliktelse til å delta i resirkulering; 1=svært lav, 7=svært høy)

PBC – Handlingskontroll (hvor vanskelig er det å resirkulere plast; 1=svært enkelt, 7=svært vanskelig)

SN – sosiale normer (forventninger av relevante andre mennesker til at man skal resirkulere; 1=svært lav, 7=svært høy)

HABIT – Hvor sterk er resirkuleringsvanene, dvs. hvor automatisk er atferden? (1=svært lav, 7=svært høy)

Masterstudenten kjørte en regresjonsanalyse med BEH som avhengig variabel og INT, PN, PBC, SN og HABIT 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	,706 ^a	,498	,494	1,535	1,989

a. Predictors: (Constant), HABIT, PBC, SN, INT, PN

b. Dependent Variable: BEHAVIOR

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1586,200	5	317,240	134,591	,000 ^b
	Residual	1598,097	678	2,357		
	Total	3184,297	683			

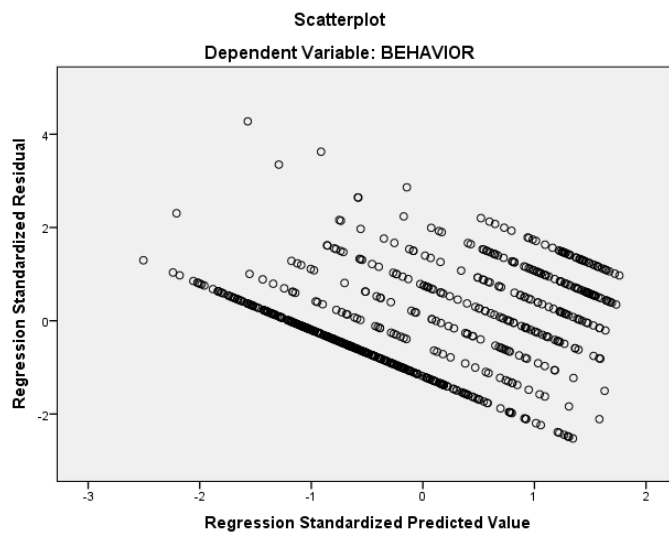
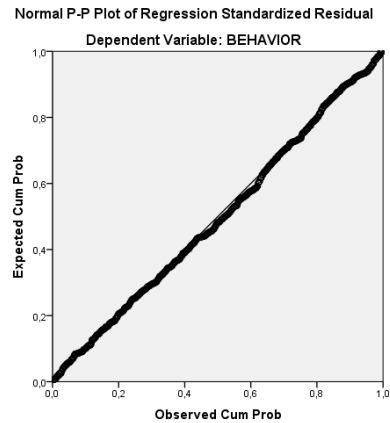
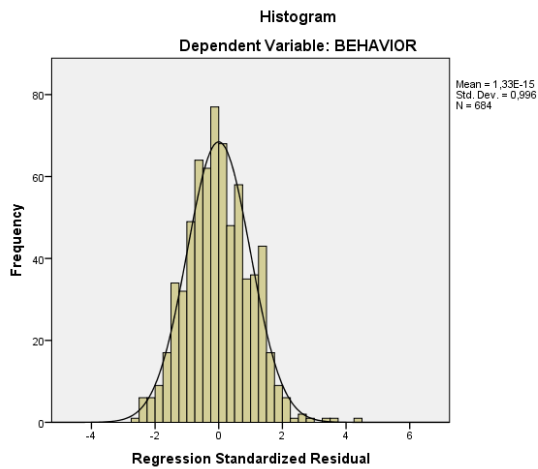
a. Dependent Variable: BEHAVIOR

b. Predictors: (Constant), HABIT, PBC, SN, INT, PN

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95,0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
		1	(Constant)	2,606			,290		8,980	,000
	INT	,254	,042	,200	6,010	,000	,171	,338	,668	1,496
	PBC	-,586	,030	-,563	-19,537	,000	-,645	-,528	,892	1,121
	SN	,047	,040	,035	1,170	,242	-,032	,126	,837	1,195
	PN	,003	,058	,002	,049	,961	-,110	,116	,560	1,785
	HABIT	,195	,050	,132	3,884	,000	,096	,294	,636	1,572

a. Dependent Variable: BEHAVIOR



- A) Hvorfor brukte masterstudenten en regresjonsanalyse? Hvilket forskningsspørsmål besvarer denne analysen? (4 poeng)

Masterstudenten hadde en kontinuerlig avhengig variabel og fem kontinuerlige uavhengige variabler. Forskningsspørsmålet var, hvor stor sammenhengene mellom de uavhengige variablene og den avhengige var, kontrollert for innflytelse av de andre uavhengige. En lineær regresjon er analysen som passer til dette spørsmålet. 4 poeng

- B) Skriv ned regresjonsligningen basert på regresjonstabellen og tolk koeffisientene. (4 poeng)

$BEH = 2,606 + 0,254 * INT - 0,586 * PBC + 0,047 * SN + 0,003 * PN + 0,195 * HAB$
(ligninger som inkluderer residual aksepteres også)

Tolkning:

En (hypotetisk) deltaker med 0 på alle uavhengige variabler ville hatt 2,606 som forventet verdi på den avhengige variabelen BEH.

Med hver enhet økning i INT øker den forventete verdien i BEH med ,254 enheter (gitt at alle andre uavhengige variabler holdes konstant). Jo større intensjonen er å resirkulere jo oftere resirkuleres plast.

Med hver enhet økning i PBC minster den forventete verdien i BEH med ,586 enheter (gitt at alle andre uavhengige variabler holdes konstant). Jo vanskeligere det er å resirkulere jo sjeldnere resirkuleres plast.

Med hver enhet økning i SN øker den forventete verdien i BEH med ,047 enheter (gitt at alle andre uavhengige variabler holdes konstant). Jo sterkere sosiale normer er å resirkulere jo oftere resirkuleres plast.

Med hver enhet økning i PN øker den forventete verdien i BEH med ,003 enheter (gitt at alle andre uavhengige variabler holdes konstant). Jo sterkere personlige normer er å resirkulere jo oftere resirkuleres plast.

Med hver enhet økning i HAB øker den forventete verdien i BEH med ,195 enheter (gitt at alle andre uavhengige variabler holdes konstant). Jo sterkere vanene er å resirkulere jo oftere resirkuleres plast.

4 poeng

- C) Hvilke av de fem variablene har en signifikant påvirkning på BEH? Hvilken har den største påvirkningen? Begrunn ditt valg kort. (5 poeng)

Tre av de uavhengige variablene har et regresjonsvekt som er signifikant forskjellig fra 0 og det er INT, PBC og HAB. Alle tre er signifikante på $p < .001$ nivået. SN og PN derimot har et p-nivå som ligger betydelig over .05. Derfor kan nullhypotesen at deres regresjonsvekt i realiteten er 0 ikke forkastes. Den med avstand største påvirkning har PBC som kan leses ut av de standardiserte regresjonsvektene. So følger intensjon og vaner.

5 poeng

- D) Hva forteller oss R^2 i tabelle 1? (3 poeng)

R^2 beskriver andel forklart varians. Så med andre ord, hvor mye av variansen rundt gjennomsnittet i BEH kan forklares av de fem prediktorene til sammen. Når man bruker mer en en prediktor bør den justerte R^2 en tolkes. Den sier at 49,4% av variansen er forklart. Det er en relativ stor andel i psykologisk atferdsforskning.

3 poeng

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

- E) Hva betyr tallene under «95% confidence interval for B» i tabell 3? (4 poeng)

95% Konfidensintervallet angir usikkerheten rundt estimatet B. Tallene betyr at estimatet fra et stort antall utvalg på samme størrelse som det vi har ville i 95% av tilfellene vært innenfor de grensene som konfidensintervallet beskriver hvis vi trekket fra en populasjon som har en sann verdi som er identisk med verdien vi regnet ut i utvalget. I vårt tilfelle: Hadde for eksempel den sanne verdien i populasjonen for regresjonsvekten av INT vært 0,254 og vi hadde trukket et stort antall utvalg på 700 fra denne populasjonen og estimert regresjonsvekten i utvalgene, så hadde

95% av disse vært mellom 0,171 og 0,338.

4 poeng

F) Hva forteller tallene som står under «collinearity statistics»? (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 (standardfeilen 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 multikolinearitetsproblematikk.

4 poeng

G) ANOVA-testen i andre tabellen er signifikant, hva betyr det? (3 poeng)

ANOVA-testen (også kaldt OMNIBUS testen) tester nullhypotesen at regresjonsmodellen forklarer ikke mer varians enn en modell som bare bruker gjennomsnitt på BEH 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

H) Tolk informasjonen i outputten i forhold til om forutsetningene knyttet til residualene i en regresjonsanalyse er på plass. (8 poeng: 1 poeng per riktig nevnt 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 veldig nær null (histogramm) 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 svært nær 2 -> uavhengighet av residualene 0,5 poeng
- 4) Scatterplotten viser ikke et mønstre hvor variansen av residualene er systematisk forskjellig med økende eller minkene predikerte verdier, vi antar homoskedastisitet 0,5 poeng