

Institutt for sosiologi og statsvitenskap
Department of sociology and political science

**Eksamensoppgave i SOS1002 Samfunnsvitenskapelig
forskningsmetode**
**Examination paper for SOS1002 Research Methods in
the Social Sciences**

**Faglig kontakt under eksamen/Academic contact during examination: Albert Simkus
og Brita Bungum**

Tlf./Phone: 73591758/73590440

Eksamensdato/Examination date: 26.05.2016

Eksamenstid/Examination time: 5 timer/5 hours

Hjelpemiddelkode/Tillatte hjelpemidler/Permitted examination support material:

Enkel kalkulator/simple calculator: Casio fx-82ES plus, Hewlett Packard HP30S, Citizen SR-270X, eller Citizen SR-270X College

Sensurdato/Grades announced on: 16.06.2016

Målform/språk/Language: Bokmål/Nynorsk/English

Antall sider/Number of pages: 4 side Bokmål/ 4 sider Nynorsk/ 4 pages English

Antall sider vedlegg Number of pages enclosed: Ingen/None

English:

ALL SECTIONS AND QUESTIONS MUST BE ANSWERED

Quantitative section

Section I.

1. Name three measures of variation and describe the differences between them.

The range is the difference between the highest and lowest scores for a single variable. The variance is the average (mean) squared deviation (or difference) of each score from the mean of the set of scores for a variable. The standard deviation is the square root of the variance. Students may also list the interquartile range – the range between the first and third quartiles of the distribution.

2. If we say that a t-test for a regression coefficient is “significant,” exactly what does this mean? Answer this in some detail and make sure you include all the important points stated precisely. It simply means that the probability that the value of the regression coefficient is 0.0 is very small, depending on the probability value used, ie. 0.05, 0.01, or 0.001. THIS TEST DOES NOT SAY ANYTHING ABOUT HOW STRONG OR IMPORTANT THE EFFECT IS, ONLY THAT THE TRUE VALUE IS UNLIKELY TO BE ZERO.

3. What is a Chi-square test used for?

Chi-square is a test-of-significance for an “association,” “relationship,” or “lack of independence” between two variables in a cross-tabulation table, usually used with variables measured on the nominal or ordinal level. It simply indicates whether it is improbable that the variables are statistically independent, not the strength of the association or relationship.

4. Explain what the “reference category” is when you use a set of dummy variables in a regression model. Explain both how it is important in interpreting the effects of the dummy variables and how it is created.

When creating dummy variables for a single theoretical variable with n categories, you must make n-1 dummy variables, one less than the number of categories with cases. When you do so, the “omitted category” becomes the “reference category,” because the regression coefficients for the dummy variables included for that theoretical variable each indicated the effect of being in those categories *relative to being in the reference category*.

5. Explain what an “interaction effect” means.

An “interaction effect” is when the effect of an independent variable, X, on a dependent variable Y, is influenced by a third variable, Z. An example would be if the variable “years of completed education” had a stronger effect on an attitude scale for women than it did for men. Interaction effects can be entered into a multiple regression model by creating a variable for the interaction by multiplying the two original independent variables together. For example, a dummy variable for “female” multiplied by a variable for “years of education.” When such an interaction variable is included in a model, the original variables included in the interaction must also be included, such as female, years of education, and female*years of education.

Section II.

The questions in this section are all based on data from a representative sample (equivalent to an approximately random sample) of adults in the country of Macedonia in 2003. The dependent variable is a scale measuring attitudes towards mores concerning sexual behavior. The scale is based on an average of responses to four questionnaire items measuring attitudes towards homosexual sex, living together without being married, and looking at pornography (these attitudes are correlated with each other in this relatively traditional society however we might think of these as involving different dimensions of values). The scale is named LIBSEXNORMS and higher values on this scale indicate that respondents had more “traditional” or “conservative” norms concerning sexual behavior.

The following independent variables were included in the survey and analyses:

MALE – A dummy variable with males coded to a value of “1”.

AGE – The respondent’s age in single years, restricted to ages 21 to 80.

AGESQ- The respondent’s years of age squared.

EDYEARS – The number of single years of formal education completed by the respondents.

ALBANIAN – A dummy variable coded “1” for self-identified ethnic Albanians, nearly all the cases coded “0” identified as ethnic Macedonians.

MARITAL STATUS – Three dummy variables for whether respondents were SINGLE, DIVORCED, or WIDOWED.

ALBxAGE – A variable computed by multiplying the variable ALBANIAN by the variable AGE.

Immediately below the questions for this section are tables you should use in answering the following questions. The first is a model summary table showing statistics for a series of three successive regression models. The second shows various regression statistics for the same set of three models.

Questions:

PART A:

1. Which model is the best model? Why? What proportion of the variance in the dependent variable is explained by this model?

Model 3 is the best model because both the improvement from Model 1 to Model 2 the improvement from Model 2 to Model 3 are statistically significant, as shown by the F-test for change in R-square in the last column of the Model Summary table. The proportion of the variance explained is R-square, shown to be .276 in the third column of the Model Summary table. Note: While Model 3 is the best model in that it includes more complicated, statistically significant effects. One can argue that they do not add very much to R-square ($0.002 + 0.016$), thus adding little of explanatory significance, while complicating the understanding of the effects. Thus, Model 1 is helpful in evaluating the simpler effects, while Model 3 does reveal more complicated – although marginally important – effects.

2. Which variables have the strongest effects and which have the weakest? Be specific in using numbers and explaining exactly what they show. (Be sure to include the significance, strength, and direction of the relationship.)

The Beta (standardized regression coefficients) may be compared to evaluate the relative importance of the independent variables (except for the dummy variables and more complicated sets of variables); the unstandardized regression coefficients should also be mentioned and interpreted, but in comparing their relative sizes, their different units of measurement must be taken into account (ie. the effect of gender is the difference between males and females, while the effect of age is the effect of one year of age).

IMPORTANT NOTE!: A serious error was made in the explanation of the direction of the meaning of the dependent variable. The beginning explanation mistakenly states that “higher values on this scale indicated that respondents had more “traditional” or “conservative” values”: this is incorrect! As consistent with the name “LIBSEXNORMS,” higher values indicate more liberal, not more conservative attitudes. For this reason, the students answers should be interpreted with leniency for the confusion caused by the fact the results are the opposite to what would be expected sociologically. The students answers SHOULD be correct, if they are based on the incorrect definition they were given, and the signs of the coefficients printed in the tables. If the students react to the fact that the directions of the effects are opposite to what would be expected sociologically, this is to their credit! If they are confused by this, they should not have their grades reduced for this reason alone.

The most important effect of all is the effect of ethnicity: the Beta for being Albanian versus Macedonian is $-.375$ in the simplest model, and is even stronger in the subsequent models. This implies that Albanians are much more liberal, in terms of the incorrect variable description; in fact, they are more conservative, to that degree. The unstandardized regression coefficient would be interpreted that Albanians were 0.723 lower on the scale ($-.723$) than were Macedonians, based on Model 1; the difference is even bigger in Models 2 and 3.

Age had the second greatest effects, with Beta = $.225$ in Model 1, and $b = .056$ (of course, the counter-intuitive directions for the effects).

Perhaps surprisingly, gender, represented by the dummy effect for MALE, has the least effect, insignificant in all models.

The effects of age and of marital status (students should recognize that “married” was the omitted reference category and interpret the effects of the other categories relative to “married” were statistically significant, though not as important as the effects of being Albanian and years of education. Again, the direction of their effects were counterintuitive, due to the reversed scale of the dependent variable, with age, being single, and being divorced seeming to be associated with less conservative views.

Extra credit for figuring out a confidence interval for one of the effects.

Here, the student only has to select one of the unstandardized coefficients for one of the independent variables in one of the models. The better students should remember from class that they can calculate the confidence interval for a regression coefficient by adding and subtracting 1.96 (or they can round this to 2.0) \times the standard error in the column next to the coefficient.

So, for example, for the effect of 0.056 for education in Model one, the confidence interval is approximately $(.005 \times 2 = .01, \text{ added and subtracted to } 0.056 = 0.046 - 0.066.)$

3. Is there a significant “interaction effect?” If so, explain what it means.

Yes, there is an interaction between being Albanian and Age. It is sufficient for students to recognize that there is an interaction between Albanian and Age. Given the error in the direction of the scale, and the inclusion of the AGESQ term, I think we should not require a correct interpretation of its meaning.

4. Are there any curvilinear effects? If so, explain what they mean.

It is sufficient for the students to recognize that adding the variable for AGE-squared allows the effect of age to be curvilinear (not as in a straight line).

5. If you were going to try to add additional independent variables, both calculated from these variables and completely new variables, what would you try?

Anything reasonable should count here, such as urban versus rural residence, religiosity, etc.

PART B: Calculate the predicted value of the attitude scale for individuals with the following sets of characteristics, from the models specified:

- Using model 1, what is the predicted value for a respondent who is a 50 year-old widowed Albanian woman with 6 years of education?

		2.496							
-0.02	0	0							
-0.005	50	-0.25							
0.056	6	0.336							
-0.723	1	-0.723							
0.203	0	0							
0.324	0	0							
-0.144	1	-0.144							
		1.715							
			2.496	0	-0.25	0.336	-0.723	-0.144	=1.715

- Using model 3, what is the predicted value for a respondent is a 50 year-old married Macedonian woman, with 6 years of education?

2.49	0	0.15	0.324	0	0	0	0	=2.964
				2.49				

-0.025	0	0
0.003	50	0.15
0.054	6	0.324
-1.426	0	0
0.218	0	0
0.331	0	0
-0.034	0	0
0	2500	0
0.016	0	0
		2.964

- Required formulas:
- $Y^* = b_0 + bX_1 + bX_2 + bX_3 + bX_4 + \dots$

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.509 ^a	.259	.257	.80213	.259	119.622	7	2395	.000
2	.511 ^b	.261	.258	.80135	.002	5.699	1	2394	.017
3	.526 ^c	.276	.274	.79304	.016	51.441	1	2393	.000

a. Predictors: (Constant), WIDOWED, DIVORCED, SINGLE, ALBANIAN, MALE Respondent's sex, EDYEARS Years of Education (with missing values assigned values from V101), AGE Age limited to 21 - 80

b. Predictors: (Constant), WIDOWED, DIVORCED, SINGLE, ALBANIAN, MALE Respondent's sex, EDYEARS Years of Education (with missing values assigned values from V101), AGE Age limited to 21 - 80, AGESQ

c. Predictors: (Constant), WIDOWED, DIVORCED, SINGLE, ALBANIAN, MALE Respondent's sex, EDYEARS Years of Education (with missing values assigned values from V101), AGE Age limited to 21 - 80, AGESQ, ALBxAGE

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.496	.112		22.244	.000
	MALE Respondent's sex	-.020	.034	-.011	-.576	.565
	AGE Age limited to 21 - 80	-.005	.001	-.085	-4.097	.000
	EDYEARS Years of Education (with missing values assigned values from V101)	.056	.005	.225	11.433	.000
	ALBANIAN	-.723	.036	-.375	-20.223	.000
	SINGLE	.203	.048	.079	4.186	.000
	DIVORCED	.324	.156	.037	2.077	.038
	WIDOWED	-.144	.069	-.040	-2.087	.037
	2	(Constant)	2.115	.195		10.858
MALE Respondent's sex		-.017	.034	-.009	-.515	.607
AGE Age limited to 21 - 80		.012	.007	.205	1.664	.096
EDYEARS Years of Education (with missing values assigned values from V101)		.055	.005	.222	11.272	.000
ALBANIAN		-.720	.036	-.373	-20.151	.000
SINGLE		.255	.053	.100	4.801	.000
DIVORCED		.323	.156	.037	2.076	.038
WIDOWED		-.116	.070	-.032	-1.659	.097
AGESQ		.000	.000	-.289	-2.387	.017
3	(Constant)	2.490	.200		12.464	.000
	MALE Respondent's sex	-.025	.034	-.013	-.736	.462
	AGE Age limited to 21 - 80	.003	.007	.051	.411	.681
	EDYEARS Years of Education (with missing values assigned values from V101)	.054	.005	.220	11.270	.000
	ALBANIAN	-1.426	.105	-.739	-13.642	.000
	SINGLE	.218	.053	.085	4.133	.000
	DIVORCED	.331	.154	.037	2.147	.032
	WIDOWED	-.034	.070	-.010	-.490	.624
	AGESQ	.000	.000	-.243	-2.021	.043
	ALBxAGE	.016	.002	.387	7.172	.000

a. Dependent Variable: LIBSEXNORMS

Eksamen sos1002 26.05 vår 2016
Sensurveiledning - kvalitativ del

Oppgave 1

En sterk besvarelse bør på en enkel måte kunne forklare fire av de seks kvalitative begrepene nedenfor, det er en styrke ved en besvarelse dersom kandidaten klarer å eksemplifisere i tillegg til forklaringen av selve begrepet

Gjør **kort** rede for **fire** av de seks begrepene. Bruk max. ½ side på hvert begrep

- a) Datagenerering – *arbeidet med “å samle inn» data. Et pluss her hvis kandidaten har fått med seg at Tjora (2012) konsekvent bruker begrepet «generering» i stedet for «innsamling» for å presisere at empiriske data ikke finnes «der ute» i utgangspunktet, men konstrueres ved forskningen.*
- b) Interaktiv observasjonsrolle – *der observatører og deltakere er åpne (synlig og tilgjengelig) for hverandre, slik at der vil være en grad av interaksjon mellom dem. Interaksjonen kan variere fra aktiv til passiv, fra eksplisitt til subtil*
- c) Intersubjektivitet- *brukes for å betegne en situasjon der flere personer skaper en felles forståelse av virkeligheten, for eksempel i løpet av et intervju*
- d) Fokuserte intervju- *brukes hos Tjora (2012) om en kortere form for semistrukturert intervju, om et tema som er avgrenset og snevret godt inn før intervjuet starter*
- e) Tekstnær koding- *brukes av Tjora(2012) om koding (første steg i SDI modellen for kvalitativ analyse) hvor en benytter ord og uttrykk direkte fra empirien som navn for kodene*
- f) Konseptuell generalisering- *generalisering ved utvikling av konsepter og presentasjon av konsepter (eksempelvis typologier) og teorier som vil ha relevans for andre tilfeller enn det (eller dem) som er studert*

Oppgave 2

Det som bør være avgjørende for bedømmingen av denne oppgaven er i hvilken grad kandidaten klarer å vise kunnskap om kvalitative metodeopplegg og hvor godt besvarelsen får frem refleksjoner, forståelse og relevante begrunnelser for opplegget

- a. Du jobber i en forskningsgruppe som har fått i oppdrag av Trondheim kommune å få mer kunnskap om hvordan ungdom forstår og erfarer lokalpolitikken i byen. Kommunen er opptatt av å få vite mer om hvilke tema ungdom opplever som viktige og hva skal til for at de vil engasjere seg mer i politikk? Forskningsgruppen blir enig om at fokusgruppeintervju kan være en hensiktsmessig metodeteknikk

Gjør rede kort rede for hva et fokusgruppeintervju er og hvorfor det kan være relevant å bruke i en slik undersøkelse. Hvilken rolle vil dere som forskerne ha i et slikt intervju?

Her bør kandidaten kunne redegjøre for at et fokusgruppeintervju er et gruppeintervju der flere deltakere inviteres til i felleskap å diskutere et eller flere temaer. Fokusgruppeintervju er en effektiv form for datagenerering fordi en utvikler intervjudata fra flere informanter samtidig. Metodeteknikken kan virke mindre truende for deltakerne enn individuelle intervjuer og i et forskningsprosjekt der informantene er unge og tema er politikk kan nettopp det å sitte sammen i en gruppe gjøre det enklere å få frem ulike oppfatninger, forståelser og erfaringer. Særlig interessant er det at diskusjoner som oppstår i fokusgruppeintervjuet kan generere spontane svar og at en kan fange opp meninger som oppstår i interaksjonen mellom informantene i gruppen. Fokusgrupper kan også ha en aktiviserende og mobiliserende effekt utover selve intervjusituasjonen, en kilde til nye tanker og ideer for de som deltar. Alle disse momentene taler til fordel for et forskningsprosjekt der tema er ungdom og engasjement for lokalpolitikk og der en ønsker økt innsikt i hva som kan mobilisere ungdom til økt politisk aktivitet. Forskerrollen i fokusgruppeintervjuet er ordstyrer (moderator). Det gjelder å lage en trygg ramme for intervjuet og styre samtalen slik at alle deltakerne får plass, kommer til orde og blir sett og hørt.

Forklar videre hvordan du ville ha gått frem med tanke på utvalg av informanter til undersøkelsen.

Her er det en stor styrke dersom kandidaten har forstått at kvalitative utvalg velges strategisk og/eller teoretisk, og at dette ikke handler om representativitet i statistisk forstand. For å sikre trygghet og god flyt i samtalen kan det noen ganger være hensiktsmessig at fokusgruppene har en viss homogenitet, eksempelvis kan det være greit at det ikke er for stor aldersforskjell på ungdommene i en gruppe. Dersom en er opptatt av å få diskusjoner mellom ulike ståsted og erfaringer kan det være bra å maksimere variasjon, men dette er en avveining som en må gjøre på bakgrunn av problemstillingen i prosjektet.

Når det gjelder utvalg så er det er også en styrke dersom kandidaten diskuterer hvor mange deltakere som bør være med i fokusgruppeintervjuene. Det anbefales gjerne et antall mellom 6 og 12 deltakere, men det viktigste er at det er mange nok til at ulike meninger blir representert og lavt nok til at deltakerne føler seg trygge

Gi i tillegg eksempler på andre kvalitative metodeteknikker som også kunne være relevante å bruke i denne undersøkelsen, og begrunn hvordan disse kan gi økt innsikt i hvordan ungdom opplever og forstår lokalpolitikken i Trondheim

Denne delen av oppgaven åpner opp for ulike løsninger. Det kan eksempelvis være relevant å få et utvalg ungdom til å skrive om hva som engasjerer dem i lokalpolitikken, hva opplever de som viktig? Det kan skje nettbasert gjennom en blogg eller gjennom andre former for skriftlige tilbakemeldinger. For å få økt innsikt i ungdoms engasjement kan det også være relevant å bruke debatter på nett og i media om dette tema. Det kan også være aktuelt å kombinere fokusgruppeintervju med dybdeintervjuer eller kanskje observasjon på debattmøter og i ungdomspolitiske organisasjoner. Gir besvarelsen andre relevante ideer/ forslag til metodeteknikker her, (som ikke er nevnt spesifikt her,) bør dette bare telle positiv

- b. I analysen av datamaterialet vil dere bruke SDI modellen (Tjora 2012) Forklar hva som er hovedhensikten med SDI modellen og hvordan dere med utgangspunkt i denne modellen kan gå frem i analysen av datamateriale?

Her er det først og fremst sentralt at det fremgår av besvarelsen at kandidaten at forstått at dette er en skjematisk modell for kvalitativ forskning hvor grunnprinsippet er en induktiv utvikling fra empiri til konsepter eller teorier med deduktive trinnvise tilbakekoblinger. Av en sterk besvarelse må det forventes at det kommer frem en forståelse av at den induktive utviklingen starter i empirien mens den stegvise deduktive handler om kvalitetssikring. Herunder følger det at målsettingen med modellen er teori(konsept)utvikling og kvalitetssikring (trinnvis deduktivt). (Tjora, s 174->) Når det gjelder hvordan SDI modellen kan brukes i analysen så er dette drøftet på forelesning der vi også har hatt en praktisk øvelse, med modellen som utgangspunkt. Det er en styrke dersom kandidaten i grove trekk kjenner til fremgangsmåten som innebærer tekstnær koding, (som kun kan skje gjennom empirien), kategorisering (samle tekstnærekoder i hovedtema) og til sist utvikling av konsepter

- c. I Tjora (2012) er *transparens* (gjennomsiktighet) og *refleksivitet* (selvbevissthet) beskrevet som to av flere kvalitetskriterier for kvalitativ forskning. Redegjør for hva som menes med disse to kriteriene og drøft kort hvordan de kan ha relevans for det undersøkelsesopplegget du foreslår ovenfor i oppgave a.

Begrepet transparens brukes i metodisk sammenheng om hvorvidt og hvordan detaljene i en studie beskrives for leserne. Høy transparens i en studie betraktes som positivt for kvalitet på forskningen. Begrepet refleksivitet er også knyttet til kvalitet på forskningen, og brukes hos Tjora om forskerens evne og vilje til å undersøke eget forskningsarbeid og hvordan personlige interesser og kunnskaper kan ha formet dette

Begge begrepene er introdusert og drøftet i kapittel 7. hos Tjora og i forelesninger. Her forventes det at kandidaten er i stand til å drøfte hvordan de ser relevans av disse kvalitetskriteriene i en undersøkelse om hvordan ungdoms forstår og erfarer lokalpolitikk. Spørsmål knyttet til gjennomføringen, valg som er gjort, styrker og svakheter ved

undersøkelsesopplegget og ved selve forskningsprosessen, kan være relevante ting å nevne i forhold til transparens. Det å problematisere eller drøfte egen forskerrolle og ståsted er særlig relevant i forbindelse med refleksivitet