Mal for sensorveiledning

Emnekode	PSY3100
Emnenavn	Forskningsmetode – kvantitativ
Emneansvarlig/oppgavegiver	Mehmet Mehmetoglu
Kvalitetssikret av	Matthias Mittner
Semester, år	Vår, 2024
Vurderingsform, lengde	Skoleeksamen, 4 timer
Tillatte hjelpemidler	Ingen

Emnets læringsutbyttebeskivelser angitt I kunnskaper, ferdigheter og generell kompetanse. (Henvisning med lenke til emnesiden på NTNUs nettsider er tilstrekkelig)	https://www.ntnu.no/studier/emner/PSY3100#tab=omEmnet
Pensum	MJ - Mehmetoglu, M. & Jakobsen, T.G. (2022). Applied statistics using Stata: A guide for the social sciences. 2ed. Sage. MV - Mehmetoglu, M. & Venturini, S. (2021). Structural equation modelling with partial least squares using Stata and R. Chapman and Hall/CRC.
Eventuelle formelle krav til besvarelsen	Ingen
Hvordan de ulike oppgavene i eksamenssettet er vektlagt	Alle de fem oppgavene skal besvares. Les oppgavene nøye og svar presist! Vekting er oppgitt ved hver oppgave.

Sensurveiledning

Alle de fem oppgavene skal besvares. Les oppgavene nøye og svar presist!

Oppgave 1 (20%)

1. Forklar kort hva regresjonsanalyse er, og hva den brukes til.

Regression analysis is a statistical technique that examines the relationship between one continuous dependent (Y) and one ($X1$) or more continuous/categorical independent variables ($X2$, $X3$, …, Xn), with the aim mainly being to test one or several hypotheses and/or make predictions.

2. Forklar kort hva som menes med *deterministisk* og *random* del av en regresjonsmodell.

A regression model consists of two parts: one deterministic and one random. The deterministic part is represented by $\beta0+\beta1Xi$ predicting $E[Yi]$, which is the same for each member of the population with the same value of X. The random part is reflected by ϵi making the population members with the same value of X vary around $E[Yi]$.

3. Forklar kort hva residual standard avvik er.

The standard deviation of the residuals shows us the 'average' distance between the observed individual Y-values (Yi) and predicted mean-Y values represented by the regression line.

4. Forklar kort hvordan vi kan teste signifikansen av en koeffisient via konfidensintervall tilnærmingen.

Since the null hypothesis assumes no effect ($\beta1=0$), we should check if the value of zero is included in the confidence interval. If the interval does not include zero, this will mean that our estimated slope is very unlikely to have come from the null hypothesis population. Thus, we would reject the null hypothesis and accept the alternative which assumes an effect.

5. Forklar kort hvilke mål vi kan bruke for å beskrive hvor godt modellen passer til dataene.

The goodness of fit of a multiple regression model can be assessed using the residual standard deviation and coefficient of determination ($R2$) measures. Briefly put, the lower the residual standard deviation or the higher $R2$ is, the better the model can be claimed to fit the data.

Oppgave 2 (20%)

1. Forklar kort hva en dummy-variable er.

A dummy variable (also called an indicator variable) takes on the values 0 and 1 to represent two different attributes respectively (e.g., male and female, unsuccessful and successful, and private and public).

2. Forklar kort med et eksempel hvordan en regresjonsmodell med én dummy og én kovariate er satt opp.

Y representing flat prices and *X*1 indicating the geographical location of the flats (dummy). In addition, *X*2 can represent a covariate, floor size, measured in square meters.

3. Forklar kort med et eksempel en interaksjon mellom to kontinuerlige variabler i en regresjonsmodell.

*X*1 represents our predictor (*health*), *X*2 represents our moderator (*age*) and E[*Yi*] represents the mean-*Y*. *Y* measures the number of hours spent working out in a month. While *age* is measured in years, *health* is measured using a six-point scale (1 = not important to 6 = important). β1 represents the coefficient on *health* when *age* is equal to zero, and β2 the coefficient on *age* when *health* is equal to zero. Finally, β3 shows the amount by which the coefficient on *health* changes as a result of a one-unit increase in *age*.

4. Forklar kort hva logistisk regresjon går ut på.

In logit regression, one estimates how much the *natural logarithm* of the *odds* for *Y* = 1 (logit) changes for each one-unit change in *X*. *P*(*Y* = 1) tells us the probability that the dependent variable (which is dichotomous, 0–1) equals 1. The probability that *Y* does not equal 1 is $P(Y \neq 1) = P(Y = 0) = 1 - P(Y = 1)$.

5. Forklar kort hva faktoranalyse brukes til.

Factor analysis is, firstly, used to reduce a large number of variables down to a meaningful and manageable number of factors that can reflect most of these variables' contents. Secondly, factor analysis is used to examine the dimensionality of a set of variables. Here, the researcher would be interested in finding out whether there exists one or more than one dimension in the variables. If the latter is the case, then the factor analysis would reveal how many and which variables belong to which dimensions. Thirdly, factor analysis is used to assess some of the psychometric properties of a multidimensional scale. Fourthly and relatedly, factor analysis is used in the early stages of a scale development.

Oppgave 3 (20%)

Nedenfor er det estimert en multippel regresjonsanalyse. Vurder hvor god modellen er. Tolk samtlige koeffisienter og vurder hvilken av prediktorene er viktigst i modellen.

```
. reg gavesum forholdtilf kvinne snill inntekt, b
```

Source	SS	df	MS
Model	2149690.01	4	537422.502
Residual	1455791.97	995	1463.1075
Total	3605481.97	999	3609.09107

Number of obs	=	1,000
F(4, 995)	=	367.32
Prob > F	=	0.0000
R-squared	=	0.5962
Adj R-squared	=	0.5946
Root MSE	=	38.251

gavesum	Coefficient	Std. err.	t	P>|t|	Beta
forholdtilf	18.86935	1.003628	18.80	0.000	.3790699
kvinne	11.11335	2.429793	4.57	0.000	.0922137
snill	4.265221	.6707547	6.36	0.000	.1282265
inntekt	.0001009	3.12e-06	32.34	0.000	.6517278
_cons	106.7275	5.090645	20.97	0.000	.

gavesum viser hvor mye en har brukt på gave til kjæresten, som har min=98 og maks=437.
forholdtilf måler hvor tilfreds en er med forholdet sitt, som har min=1 og maks=5.
kvinne representer kjønn der 0=menn og 1=kvinner.
snill måler hvor snill en vurderer kjæresten sin, som har min=1 og maks=7.
inntekt måler hvor mye en tjener i året, som har min=151.000 og maks=1.500.000.

It is expected that the candidate uses R-squared value to judge the quality of the model. Moreover, each of the coefficients including the _cons should be interpreted correctly and precisely. It is further important that the candidate uses beta (standardised) coefficients to decide the relative importance of predictors.

Oppgave 4 (20%)

Nedenfor er det estimert en interaksjonsanalyse med en kovariate. Tolk de første fire koeffisientene. For tolkningen av koeffisienten til interaksjonsleddet, benytt grafen nedenfor.

```
. reg hourly_wage c.education i.foreigner c.education#i.foreigner work_experience

      Source |       SS          df      MS        Number of obs   =    1,000
-------------+----------------------------------   F(4, 995)       =    820.95
       Model | 7692435.13          4  1923108.78   Prob > F        =    0.0000
    Residual | 2330829.83        995  2342.54254   R-squared       =    0.7675
-------------+----------------------------------   Adj R-squared   =    0.7665
       Total |   10023265        999  10033.2983   Root MSE        =      48.4

-------------------------------------------------------------------------------
         hourly_wage |  Coefficient  Std. err.      t    P>|t|   [95% conf. interval]
---------------------+---------------------------------------------------------
           education |    9.40488   .4069589    23.11   0.000    8.606284    10.20348
         1.foreigner |  -36.19582   15.75347    -2.30   0.022   -67.10966   -5.281978
                     |
foreigner#c.education|
                   1 |    -6.5818   .7717388    -8.53   0.000   -8.096222   -5.067377
                     |
     work_experience |   4.625272   .4032279    11.47   0.000    3.833998    5.416547
               _cons |   68.87201   10.33789     6.66   0.000    48.58545    89.15858
-------------------------------------------------------------------------------
```

foreigner representerer nasjonalitet der 0=ikke-utlending og 1=utlending.

education representer utdanningsnivå som har min=8 og maks=34.

work_experience representerer arbeidserfaring som har min=4 og maks=27.

hourly_wage representer timelønn som har min=90 og maks=553.

Predictive margins
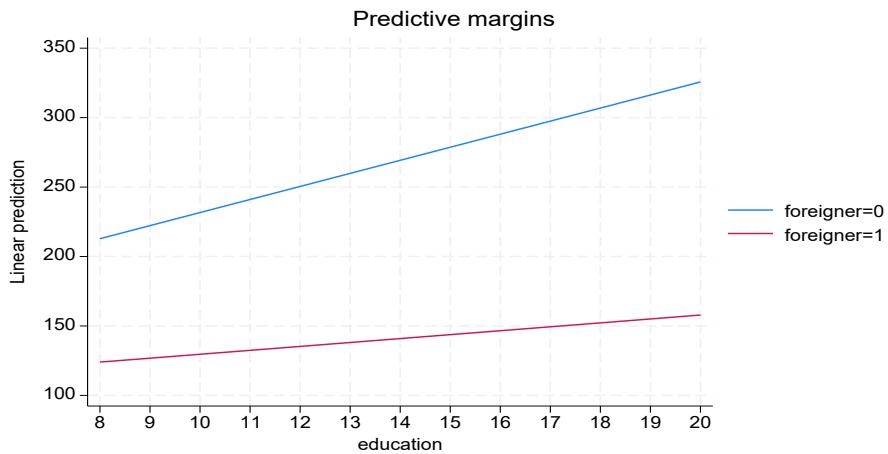
It is expected that the candidate interprets the first three coefficients correctly. However, the interpretation of the interaction term should be a given a higher weight.

Oppgave 5 (20%)

Hva prøver forskeren å finne ut med analysen nedenfor.

Vurder målemodellen og strukturmodellen basert på figuren (output) nedenfor fra plssem.

Measurement model - Standardized loadings

	Reflective: JOB_SATISF~N	Reflective: LEADERSHIP~Y	Reflective: WORK_ENVIR~T	Reflective: JOB_LOYALTY
var1	0.856			
var2	0.886			
var3	0.928			
var4		0.863		
var5		0.884		
var6		0.857		
var7			0.832	
var8			0.878	
var9			0.902	
var10				0.814
var11				0.891
var12				0.910
Cronbach	0.870	0.837	0.842	0.844
DG	0.920	0.902	0.904	0.905
rho_A	0.887	0.837	0.853	0.865

Discriminant validity - Squared interfactor correlation vs. Average variance extracted (AVE)

	JOB_SATISF~N	LEADERSHIP~Y	WORK_ENVIR~T	JOB_LOYALTY
JOB_SATISF~N	1.000	0.027	0.053	0.135
LEADERSHIP~Y	0.027	1.000	0.000	0.083
WORK_ENVIR~T	0.053	0.000	1.000	0.018
JOB_LOYALTY	0.135	0.083	0.018	1.000
AVE	0.793	0.754	0.759	0.762

Structural model - Standardized path coefficients

Variable	JOB_SATISF~N	JOB_LOYALTY
JOB_SATISF~N		0.314
		(0.000)
LEADERSHIP~Y	0.164	0.237
	(0.000)	(0.000)
WORK_ENVIR~T	0.230	0.061
	(0.000)	(0.036)
r2_a	0.078	0.189

It is expected that the candidate first explains what hypotheses the researcher is testing. For the evaluation of the measurement model, the candidate should use criteria item reliability, construct reliability, average variance extracted and discriminant validity. For the evaluation of the structural model, the candidate should use the criteria 3S (sign, significance, and size) and R-squared value.

NTNU

Fakultet for informasjonsteknologi og elektroteknikk

Karakterskala som er benyttet

Bokstavkarakter: https://innsida.ntnu.no/wiki/-/wiki/Norsk/Karakterskalaen