

Psykologisk institutt

Eksamensoppgave i PSY3100 Forskningsmetode - Kvantitativ

Faglig kontakt under eksamen: Mehmet Mehmetoglu

Tlf.: 91838665

Eksamensdato:

Eksamenstid (fra-til):

Hjelpemiddelkode/Tillatte hjelpemidler: Ingen

Annen informasjon:

Målform/språk:

Antall sider:

Antall sider vedlegg:

Kontrollert av:

Dato

Sign

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

Oppgave 1 (20%)

1. Forklar kort med et eksempel hva *regresjon med en dummy variabel* er.
2. Forklar kort med et eksempel hva *regresjon med en dummy variabel og en kovariate* er.
3. Forklar kort med et eksempel hva *regresjon med mer enn en dummy variabel* er.
4. Forklar kort med et eksempel hva *regresjon med mer enn en dummy variabel og en kovariate* er.
5. Forklar kort hvorfor ANOVA egentlig er regresjonsanalyse.

Oppgave 2 (20%)

1. Forklar kort med et eksempel hva *enkel (simple) regresjon* er.
2. Forklar kort med et eksempel hva *multiple regresjon* er.
3. Forklar kort med et eksempel hva *logistisk regresjon* er.
4. Forklar kort med et eksempel hva *faktoranalyse* er.
5. Forklar kort med et eksempel hva *pls-sem* er.

Oppgave 3 (20%)

Basert på figuren (*output*)* nedenfor, svar på følgende spørsmål.

1. Forklar hva de ulike koeffisientene i modellen forteller oss.
2. Forklar hvordan du vil sammenligne de inkluderte gruppene i modellen.
3. Forklar om (og hvorfor) du vil bruke R-squared eller Adj. R-squared for å vurdere modellens kvalitet.
4. Forklar hva *F*-testen forteller oss her.
5. Forklar om signifikanstesting basert på *p*-verdi og konfidensintervall er i samsvar.

```
. reg flat_price2 i.energy_efficiency
```

| | | | | | | |
|----------|------------|----|------------|---------------|---|---------|
| Source | SS | df | MS | Number of obs | = | 82 |
| | | | | F(2, 79) | = | 1.50 |
| Model | 5.4049e+12 | 2 | 2.7024e+12 | Prob > F | = | 0.2301 |
| Residual | 1.4261e+14 | 79 | 1.8052e+12 | R-squared | = | 0.0365 |
| | | | | Adj R-squared | = | 0.0121 |
| Total | 1.4802e+14 | 81 | 1.8274e+12 | Root MSE | = | 1.3e+06 |

| flat_price2 | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] |
|-------------------|-----------|-----------|-------|-------|----------------------|
| energy_efficiency | | | | | |
| mediocre | -448333.3 | 467046.4 | -0.96 | 0.340 | -1377966 481299.3 |
| poor | -840500 | 505708.8 | -1.66 | 0.100 | -1847088 166088.2 |
| _cons | 3660500 | 424880.4 | 8.62 | 0.000 | 2814797 4506203 |

*Prisen på leiligheter (*flat_price2*) er den avhengige og energi merking (*energy_efficiency*) er den uavhengige variabelen. Husk at *energy_efficiency* består av tre kategorier: *best*, *middels (mediocre)* og *dårlig (poor)*.

Oppgave 4 (20%)

Forklar hva de første tre koeffisientene i figuren (output) nedenfor forteller oss.

```
. reg loenn c.alder i.kvinne c.alder#i.kvinne
```

| Source | SS | df | MS | Number of obs | = | 471 |
|----------|------------|-----|------------|---------------|---|--------|
| Model | 487686.169 | 3 | 162562.056 | F(3, 467) | = | 14.42 |
| Residual | 5263212.71 | 467 | 11270.2628 | Prob > F | = | 0.0000 |
| | | | | R-squared | = | 0.0848 |
| | | | | Adj R-squared | = | 0.0789 |
| Total | 5750898.88 | 470 | 12235.9551 | Root MSE | = | 106.16 |

| loenn | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] |
|---------------------|-----------|-----------|-------|-------|----------------------|
| alder | 2.770991 | .5940404 | 4.66 | 0.000 | 1.603668 3.938314 |
| 1.kvinne | 25.72292 | 33.85417 | 0.76 | 0.448 | -40.80245 92.24829 |
| kvinne#c.alder 1 | -1.941886 | .8683014 | -2.24 | 0.026 | -3.648147 -.2356245 |
| _cons | 133.0634 | 22.63506 | 5.88 | 0.000 | 88.58418 177.5425 |

Kvinne er en dummy/dikotom variabel kodet som 0=menn og 1=kvinne. Alder er en numerisk/kontinuerlig variabel, som er målt i antall år.

Loenn (timelønn) er en numerisk/kontinuerlig variabel, som er målt i antall kroner.

Oppgave 5 (20%)

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

| Measurement model - Standardized loadings | | | | |
|---|-----------------------|---------------------------|-----------------------------|----------------------------|
| | Reflective: KULTUR | Reflective: KURIOSITET | Reflective: FERIE UTEN-S | Reflective: FERIE NORGE |
| V1A | 0.742 | | | |
| V1B | 0.769 | | | |
| V1C | 0.782 | | | |
| V2A | | 0.962 | | |
| V2B | | 0.874 | | |
| V3B | | | 0.939 | |
| V3D | | | 0.919 | |
| V3A | | | | 0.910 |
| V3E | | | | 0.706 |
| Cronbach | 0.656 | 0.829 | 0.842 | 0.517 |
| DG | 0.814 | 0.921 | 0.927 | 0.805 |

| Discriminant validity - Squared interfactor correlation vs. Average variance extracted (AVE) | | | | |
|--|--------|------------|--------------|-------------|
| | KULTUR | KURIOSITET | FERIE UTEN-S | FERIE NORGE |
| KULTUR | 1.000 | 0.006 | 0.191 | 0.138 |
| KURIOSITET | 0.006 | 1.000 | 0.048 | 0.095 |
| FERIE_UTEN-S | 0.191 | 0.048 | 1.000 | 0.466 |
| FERIE_NORGE | 0.138 | 0.095 | 0.466 | 1.000 |
| AVE | 0.584 | 0.844 | 0.863 | 0.663 |

| Structural model - Standardized path coefficients | | |
|---|------------------|------------------|
| Variable | FERIE UTEN-S | FERIE NORGE |
| KULTUR | 0.422 (0.000) | 0.349 (0.000) |
| KURIOSITET | 0.185 (0.000) | 0.281 (0.000) |
| r2_a | 0.224 | 0.215 |

p-values in parentheses

*Her bruker forskeren to typer motivasjon (KULTUR og KURIOSITET) samt interesse for å feriere i utlandet (FERIE UTENLANDS) og interesse for å feriere i Norge (FERIE I NORGE) målt ved hjelp av flere items (på en skala fra 1 til 5).

SENSORVEILEDNING

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

Oppgave 1 (20%)

5. Forklar kort med et eksempel hva *regresjon med en dummy variabel* er.

Tilsvaret er en uavhengig t-test hvor den avhengige er kontinuerlig mens den uavhengige er en variabel som inneholder to grupper/kategorier, eller en lignende forklaring. Eksempel må gis.

6. Forklar kort med et eksempel hva *regresjon med en dummy variabel og en kovariate* er.

Tilsvaret er en uavhengig t-test med en tilleggs variabel å kontrollere for, eller en lignende forklaring. Eksempel må gis.

7. Forklar kort med et eksempel hva *regresjon med mer enn en dummy variabel* er.

Tilsvaret er en uavhengig ANOVA, hvor den avhengige er kontinuerlig mens den uavhengige er en variabel som inneholder flere enn to grupper/kategorier, eller en lignende forklaring. Eksempel må gis.

8. Forklar kort med et eksempel hva *regresjon med mer enn en dummy variabel og en kovariate* er.

Tilsvaret er en uavhengig ANOVA med en tilleggs variabel å kontrollere for, eller en lignende forklaring. Eksempel må gis.

5. Forklar hvorfor ANOVA egentlig er regresjonsanalyse.

En koeffisient tilsvaret forskjellen mellom to gjennomsnitt i regresjonsanalyse, og vi sammenligner gjennomsnitt i ANOVA, eller en lignende forklaring.

Oppgave 2 (20%)

6. Forklar kort med et eksempel hva *enkel (simple) regresjon* er.

Studert sammenhengen mellom en kontinuerlig avhengig variabel og en kategorisk eller kontinuerlig uavhengig variabel, eller en lignende forklaring. Eksempel må gis.

7. Forklar kort med et eksempel hva *multiple regresjon* er.

Studert sammenhengen mellom en kontinuerlig avhengig variabel og flere kategoriske og/eller kontinuerlige uavhengige variabler, eller en lignende forklaring. Eksempel må gis.

8. Forklar kort med et eksempel hva *logistisk regresjon* er.

Studert sammenhengen mellom en kategorisk avhengig variabel og en eller flere kategoriske og/eller kontinuerlige uavhengige variabler, eller en lignende forklaring. Eksempel må gis.

9. Forklar kort med et eksempel hva *faktoranalyse* er.

Finner ut et mindre sett med underliggende faktorer som forklarer kovariansen/korrelasjonen mellom et stort sett med observerte variabler, eller en lignende forklaring. Eksempel må gis.

10. Forklar kort med et eksempel hva *pls-sem* er.

Studerer sammenhengen mellom latente variabler, eller en lignende forklaring. Eksempel må gis.

Oppgave 3 (20%)

Basert på figuren (*output*)* nedenfor, svar på følgende spørsmål.

6. Forklar hva de ulike koeffisientene i modellen forteller oss.

Sammenligning mellom mediocre og best, og poor og best må gjøres ved å bruke 3S-kriteriet.

7. Forklar hvordan du vil sammenligne de inkluderte gruppene i modellen.

Sammenligning mellom mediocre og poor må gjøres ved å bruke konfidensintervall.

8. Forklar om (og hvorfor) du vil bruke R-squared eller Adj. R-squared for å vurdere modellens kvalitet.

R-squared brukes fordi det fortsatt er snakk om kun en variabel, eller en lignende forklaring.

9. Forklar hva *F*-testen forteller oss her.

En signifikant *F*-test forteller at det er signifikant forskjell minst mellom to grupperes gjennomsnitt. I dette tilfelle, er *F*-testen ikke signifikant, eller en liknende forklaring.

10. Forklar om signifikanstesting basert på *p*-verdi og konfidensintervall er i samsvar.

Ja. Begge konfidensintervallene omfatter 0 og *p*-verdien for begge koeffisientene er over 0.05, eller en lignende forklaring.

```
. reg flat_price2 i.energy_efficiency
```

| Source | SS | df | MS | Number of obs | = | 82 |
|----------|------------|----|------------|---------------|---|---------|
| Model | 5.4049e+12 | 2 | 2.7024e+12 | F(2, 79) | = | 1.50 |
| Residual | 1.4261e+14 | 79 | 1.8052e+12 | Prob > F | = | 0.2301 |
| | | | | R-squared | = | 0.0365 |
| | | | | Adj R-squared | = | 0.0121 |
| Total | 1.4802e+14 | 81 | 1.8274e+12 | Root MSE | = | 1.3e+06 |

| flat_price2 | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] |
|-------------------------------|-----------|-----------|-------|-------|----------------------|
| energy_efficiency mediocre | -448333.3 | 467046.4 | -0.96 | 0.340 | -1377966 481299.3 |
| poor | -840500 | 505708.8 | -1.66 | 0.100 | -1847088 166088.2 |
| _cons | 3660500 | 424880.4 | 8.62 | 0.000 | 2814797 4506203 |

*Prisen på leiligheter (*flat_price2*) er den avhengige og energi merking (*energy_efficiency*) er den uavhengige variabelen. Husk at *energy_efficiency* består av tre kategorier: *best*, *middels (mediocre)* og *dårlig (poor)*.

Oppgave 4 (20%)

Forklar hva de første tre koeffisientene i figuren (output) nedenfor forteller oss.

```
. reg loenn c.alder i.kvinne c.alder#i.kvinne
```

| Source | SS | df | MS | Number of obs | = | 471 |
|----------|------------|-----|------------|---------------|---|--------|
| Model | 487686.169 | 3 | 162562.056 | F(3, 467) | = | 14.42 |
| Residual | 5263212.71 | 467 | 11270.2628 | Prob > F | = | 0.0000 |
| Total | 5750898.88 | 470 | 12235.9551 | R-squared | = | 0.0848 |
| | | | | Adj R-squared | = | 0.0789 |
| | | | | Root MSE | = | 106.16 |

| loenn | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] |
|---------------------|-----------|-----------|-------|-------|----------------------|
| alder | 2.770991 | .5940404 | 4.66 | 0.000 | 1.603668 3.938314 |
| 1.kvinne | 25.72292 | 33.85417 | 0.76 | 0.448 | -40.80245 92.24829 |
| kvinne#c.alder 1 | -1.941886 | .8683014 | -2.24 | 0.026 | -3.648147 -.2356245 |
| _cons | 133.0634 | 22.63506 | 5.88 | 0.000 | 88.58418 177.5425 |

Kvinne er en dummy/dikotom variabel kodet som 0=menn og 1=kvinne.

Alder er en numerisk/kontinuerlig variabel, som er målt i antall år.

Loenn (timelønn) er en numerisk/kontinuerlig variabel, som er målt i antall kroner.

Følgende eller en lignende forklaring:

Koeffisienten til alder (2.77) gjelder den gruppen av variabelen kvinne som er kodet 0 (dvs. men). Vi kan da si at koeffisienten (effekten) av alder på timelønn øker med 2.77 for men. Eller kan vi også si at hos men for hver enhets økning i alder medfører 2.77 kr økning i timelønn i snitt.

Koeffisienten til kvinne (25.7) gjelder dem som er 0 i alder. Siden 0 er ikke i range til variabelen alder, gir dette ikke mening å tolke.

Koeffisienten til interaksjonsleddet (-1.94) viser til forskjellen mellom koeffisienten til alder for men og kvinner. Dette betyr at koeffisienten til alder for kvinner er 1.94 lavere enn den for men.

Oppgave 5 (20%)

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

| Measurement model - Standardized loadings | | | | |
|---|-----------------------|---------------------------|-----------------------------|----------------------------|
| | Reflective: KULTUR | Reflective: KURIOSITET | Reflective: FERIE UTEN-S | Reflective: FERIE NORGE |
| V1A | 0.742 | | | |
| V1B | 0.769 | | | |
| V1C | 0.782 | | | |
| V2A | | 0.962 | | |
| V2B | | 0.874 | | |
| V3B | | | 0.939 | |
| V3D | | | 0.919 | |
| V3A | | | | 0.910 |
| V3E | | | | 0.706 |
| Cronbach | 0.656 | 0.829 | 0.842 | 0.517 |
| DG | 0.814 | 0.921 | 0.927 | 0.805 |

| Discriminant validity - Squared interfactor correlation vs. Average variance extracted (AVE) | | | | |
|--|--------|------------|--------------|-------------|
| | KULTUR | KURIOSITET | FERIE UTEN-S | FERIE NORGE |
| KULTUR | 1.000 | 0.006 | 0.191 | 0.138 |
| KURIOSITET | 0.006 | 1.000 | 0.048 | 0.095 |
| FERIE_UTEN-S | 0.191 | 0.048 | 1.000 | 0.466 |
| FERIE_NORGE | 0.138 | 0.095 | 0.466 | 1.000 |
| AVE | 0.584 | 0.844 | 0.863 | 0.663 |

| Structural model - Standardized path coefficients | | |
|---|------------------|------------------|
| Variable | FERIE UTEN-S | FERIE NORGE |
| KULTUR | 0.422 (0.000) | 0.349 (0.000) |
| KURIOSITET | 0.185 (0.000) | 0.281 (0.000) |
| r2_a | 0.224 | 0.215 |

p-values in parentheses

*Her bruker forskeren to typer motivasjon (KULTUR og KURIOSITET) samt interesse for å feriere i utlandet (FERIE UTENLANDS) og interesse for å feriere i Norge (FERIE I NORGE) målt ved hjelp av flere items (på en skala fra 1 til 5).

Følgende eller en lignende forklaring:

Målemodellen vurderes ut ifra item reliability, construct reliability, convergent validity, og discriminant validity. Som vi ser er alle faktorladningene/korrelasjonene i modellen over det anbefalte nivået 0.7. Det vil si at, minst 50 prosent av en variabels varians er fanget opp av faktoren. Videre ser vi at construct reliability er også tilstede i modellen i den forstand at D.G.Rho (tilsvarer Cronbach) er over det anbefalte nivået på 0.7. Det vil si at hver construct er homogene. Vi ser også at convergent validitet (målt i form av AVE) er også tilstede siden hver construct fanger opp i snitt 50 prosent av variansen i sine respektive items. Vi ser til slutt at alle AVE-verdiene er større enn kvadrerte korrelasjonene mellom faktorene. Dvs, diskriminant validitet er tilstede. På bakgrunn av disse funnene, kan vi fastslå at målemodellen er god. Derfor kan vi gå videre og analysere strukturmodellen.

Strukturmodellen vurderes ut ifra R-square og koeffisientene slik som det gjøres i vanlig regresjonsanalyse. Her har vi to avhengige variabler, FERIE UTENLANDS og FERIE I NORGE. Modellen bestående av KULTUR og KURIOSITET forklarer henholdsvis 23 prosent og 22 prosent av variansen i FERIE UTENLANDS og FERIE i Norge. Disse forklarte variansene kan på et generelt grunnlag sies å være akseptable. Når det gjelder tolkningen av koeffisientene, er det en positiv sammenheng mellom KULTUR og KURIOSITET og de to avhengige variablene. Det vil si, økt KULTUR/KURIOSITET medfører økt interesse for FERIE UTENLANDS og FERIE i Norge.

