

Skriftlig eksamen i Statistik

4 timers skriftlig prøve

Dette opgavesæt består af 4 delopgaver, der indgår i bedømmelsen af den samlede opgavebesvarelse med følgende omtrentlige vægte:

Opgave 1	20%
Opgave 2	35%
Opgave 3	20%
Opgave 4	25%
I alt	100%

Til prøven er alle hjælpemidler tilladte – også elektroniske.

Opgavebesvarelsen skal afleveres i henhold til skolens eksamensreglement
jf. bekendtgørelse nr. 1500 af 2. december 2016

Fredag den 9. juni 2017
Kl. 12.00 – 16.00

Baggrundsinformation

I dagspressen ses ofte artikler om, at tandlægebesøg er mere udbredt blandt ældre, kvinder og personer med høje indkomster end blandt unge, mænd og personer med lave indkomster. Det har fået en kommunes tandlæger til nærmere at undersøge indbyggernes tandlægebesøg. Tandlægerne baserede deres undersøgelse på data fra kommunen, data fra deres organisation samt en lokal og mere detaljeret undersøgelse blandt 18 tilfældigt udvalgte indbyggere over 18 år. Der blev ikke taget hensyn til typen af konsultation i tandlægernes undersøgelse.

Opgave 1

Ud fra den lokale undersøgelse ville tandlægerne finde ud af, om der var en sammenhæng mellem Antal årlige tandlægebesøg (Besøg) og henholdsvis 1: Indkomst, 2: Køn, 3: Alder og 4: Antal års uddannelse (Uddannelse).

For at undersøge en eventuel sammenhæng lavede tandlægerne en regressionsanalyse med Besøg som forklaret variabel.

I regressionen indgik som nævnt følgende forklarende variable:

Indkomst (respondentens indkomst i 1.000 kr.).

Køn (0 = mand; 1 = kvinde).

Alder (respondentens alder).

Uddannelse (antal år respondenter har brugt på uddannelse efter folkeskolen).

I bilag 1 vises følgende regressioner:

Model 1: Besøg forklaret ved Indkomst, Køn, Alder og Uddannelse.

Model 2: Besøg forklaret ved Indkomst, Køn og Alder.

Model 3: Besøg forklaret ved Indkomst og Køn.

Model 4: Besøg forklaret ved Indkomst.

I bilag 2 vises en korrelationsmatrix for de anvendte variable og i bilag 3 findes data til regressionerne.

1.1 Giv ud fra bilag 2 en vurdering af de anvendte forklarende variable.

1.2 Opstil ved hjælp af bilag 1 den bedste model til forklaring af Besøg (antal årlige tandlægebesøg).

Man besluttede at anvende model 4, hvor Besøg (antal årlige tandlægebesøg) alene forklares ved Indkomst (i 1.000 kr.).

1.3 Opstil denne model, og forklar modellens koefficienter. Vurder endvidere modellens forudsætninger ud fra plottet i Model 4.

1.4 Bestem, ud fra Model 4, et 95% forudsigelsesinterval for Besøg (antal årlige tandlægebesøg), hvis respondentens indkomst er 450 (i 1.000 kr.).

Opgave 2

Tandlægerne var i tvivl om de målte indkomster svarede til de generelle indkomster i kommunen. En undersøgelse hos kommunen viste, at indkomsterne i kommunen var normalfordelte med et gennemsnit på $\mu = 550$ svarende til 550.000 kr. og en standardafvigelse på $\sigma = 150$ svarende til 150.000 kr. (varians $\sigma^2 = 22.500$). Tabel 2.1 viser de årlige indkomster (i 1.000 kr.) fra de 18 respondenter.

Tabel 2.1 Årlige indkomster fra 18 respondenter.

410	630	310	190	250	410
205	670	320	285	185	890
610	405	555	750	260	360

2.1 Undersøg om de årlige indkomster følger en normalfordeling.

2.2 Test på 1% niveauet om variansen på de årlige indkomster kan antages at være 22.500.

2.3 Test på 5% niveauet om den gennemsnitlige årlige indkomst er under 550 (svarende til 550.000 kr.). Gør rede for signifikansniveauets betydning for din konklusion.

Tandlægerne i en nabokommune har lavet en tilsvarende undersøgelse med $n = 20$ respondenter. Den gennemsnitlige indkomst var her $\bar{x} = 550$ og variansen $s^2 = 40.000$.

2.4 Test på 5% niveauet om indkomsterne i de to kommuner i den videre analyse kan antages at have ens varianser.

2.5 Undersøg, ved hjælp af en ensidet test, om der er forskel på de gennemsnitlige indkomster i de to kommuner. Gør rede for signifikansniveauets betydning for din konklusion.

Antag i det følgende at den gennemsnitlige indkomst for indbyggerne i kommunen følger en normalfordeling med en middelværdi på $\mu = 550$ og en varians på $\sigma^2 = 22.500$.

2.6 Bestem sandsynligheden for at indkomsten hos en tilfældig valgt indbygger i kommunen er mindre end 600.

2.7 Bestem sandsynligheden for at den gennemsnitlige indkomst for 18 tilfældigt valgte indbyggere i kommunen er mindre end 600.
Sammenlign med opgave 2.6 og begrund en eventuel forskel.

Opgave 3

Tandlægerne ønskede tillige at finde ud af, om deres undersøgelse var repræsentativ med hensyn til henholdsvis alder og køn.

Tabel 3.1 nedenfor viser aldersfordelingen i tandlægernes undersøgelse (Antal) og i kommunen (angivet i %).

Tabel 3.1 Aldersfordeling i undersøgelsen og i kommunen.

Alder	Antal	Kommune
18 år – 30 år	5	30%
31 år – 50 år	7	40%
51 år –	6	30%

3.1 Undersøg ved hjælp af tabel 3.1, om tandlægernes undersøgelse er repræsentativ med hensyn til alder.

I tandlægernes undersøgelse deltog 11 mænd og 7 kvinder. I kommunen er 45% af indbyggerne mænd.

3.2 Vurder ved hjælp af en test, om andelen af mænd i tandlægernes undersøgelse svarer til andelen af mænd i kommunen.

Opgave 4

Tandlægerne mente, at der var en sammenhæng mellem alder og antal årlige tandlægebesøg. Deres organisation har tidligere lavet en større undersøgelse om blandt andet denne sammenhæng. Resultatet af organisationens undersøgelse fremgår af nedenstående tabel 4.1.

Tabel 4.1 Sammenhæng mellem alder og antal årlige tandlægebesøg (Besøg).

Alder	Besøg			
	0	1-3	4-	I alt
18 år - 30 år	406	359	35	800
31 år - 50 år	495	530	42	1.067
51 år -	332	432	36	800
I alt	1.233	1.321	113	2.667

- 4.1 Test på 5% niveauet om der kan antages at være en sammenhæng mellem Alder og Antal årlige tandlægebesøg (Besøg). Kommentér en evt. sammenhæng.
- 4.2 Test på 5% niveauet om andelen af de 800 i gruppen 18 år - 30 år, der ikke har været til tandlæge (Besøg = 0), er større end den tilsvarende andel af de 1.067 i gruppen 31 år - 50 år.
Gør rede for signifikansniveauets betydning for din konklusion.
- 4.3 Undersøg ved hjælp af et 95% konfidensinterval, om det kan antages, at over 50% ikke har været til tandlæge (Besøg = 0). Vil det ændre på din konklusion, hvis den blev baseret på et 99% konfidensinterval?

Bilag 1 Regressioner**Model 1: Besøg forklaret ved Indkomst, Køn, Alder og Uddannelse.**

RESUMEOUTPUT

<i>Regressionsstatistik</i>	
Multipel R	0,9752
R-kvadreret	0,9511
Justeret R-kvadreret	0,9360
Standardfejl	0,3829
Observationer	18

ANOVA					
	<i>fg</i>	<i>SK</i>	<i>MK</i>	<i>F</i>	<i>Signifikans F</i>
Regression	4	37,0383	9,2596	63,1494	0,0000
Residual	13	1,9062	0,1466		
I alt	17	38,9444			

	<i>Koefficienter</i>	<i>Standardfejl</i>	<i>t-stat</i>	<i>P-værdi</i>	<i>Nedre 95%</i>	<i>Øvre 95%</i>
Skæring	-1,3564	0,4800	-2,8260	0,0143	-2,3933	-0,3195
Indkomst	0,0071	0,0005	14,8822	0,0000	0,0060	0,0081
Køn	-0,4277	0,2245	-1,9049	0,0792	-0,9127	0,0574
Alder	0,0139	0,0077	1,8006	0,0950	-0,0028	0,0305
Uddannelse	-0,0054	0,0320	-0,1674	0,8696	-0,0745	0,0638

Model 2: Besøg forklaret ved Indkomst, Køn og Alder.

RESUMEOUTPUT

<i>Regressionsstatistik</i>	
Multipel R	0,9752
R-kvadreret	0,9509
Justeret R-kvadreret	0,9404
Standardfejl	0,3694
Observationer	18

ANOVA

	<i>fg</i>	<i>SK</i>	<i>MK</i>	<i>F</i>	<i>Signifikans F</i>
Regression	3	37,0342	12,3447	90,4709	0,0000
Residual	14	1,9103	0,1364		
I alt	17	38,9444			

	<i>Koefficienter</i>	<i>Standardfejl</i>	<i>t-stat</i>	<i>P-værdi</i>	<i>Nedre 95%</i>	<i>Øvre 95%</i>
Skæring	-1,4147	0,3187	-4,4388	0,0006	-2,0983	-0,7311
Indkomst	0,0070	0,0005	15,4770	0,0000	0,0061	0,0080
Køn	-0,4112	0,1947	-2,1121	0,0531	-0,8288	0,0064
Alder	0,0143	0,0069	2,0947	0,0549	-0,0003	0,0290

Model 3: Besøg forklaret ved Indkomst og Køn.

RESUMEOUTPUT

<i>Regressionsstatistik</i>	
Multipel R	0,9673
R-kvadreret	0,9356
Justeret R-kvadreret	0,9270
Standardfejl	0,4090
Observationer	18

ANOVA					
	<i>fg</i>	<i>SK</i>	<i>MK</i>	<i>F</i>	<i>Signifikans F</i>
Regression	2	36,4355	18,2177	108,9149	0,0000
Residual	15	2,5090	0,1673		
I alt	17	38,9444			

	<i>Koefficienter</i>	<i>Standardfejl</i>	<i>t-stat</i>	<i>P-værdi</i>	<i>Nedre 95%</i>	<i>Øvre 95%</i>
Skæring	-0,9007	0,2252	-4,0002	0,0012	-1,3806	-0,4208
Indkomst	0,0072	0,0005	14,4326	0,0000	0,0061	0,0083
Køn	-0,3013	0,2076	-1,4516	0,1672	-0,7438	0,1411

Model 4: Besøg forklaret ved Indkomst.

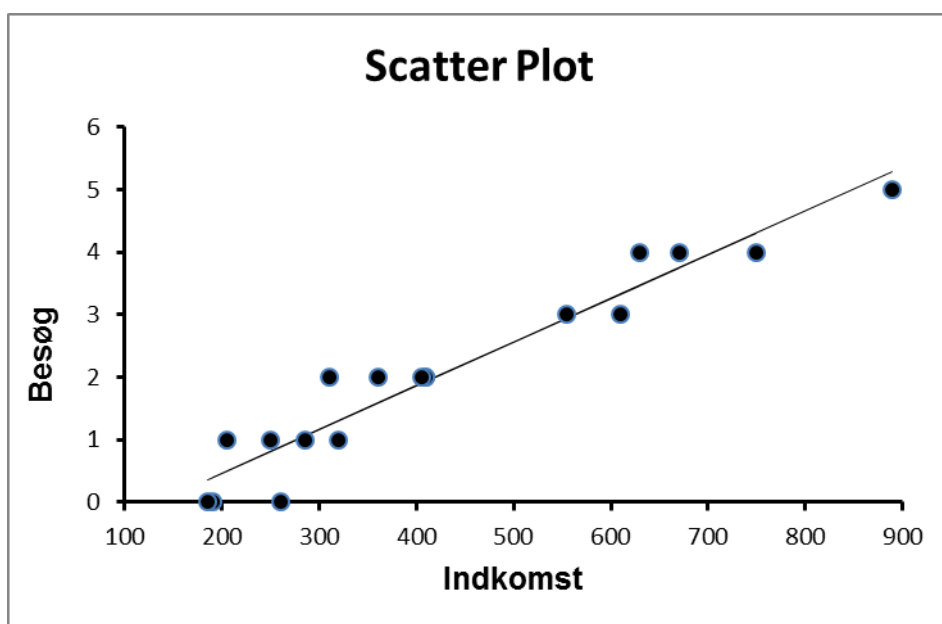
RESUMEOUTPUT

<i>Regressionsstatistik</i>	
Multipel R	0,9626
R-kvadreret	0,9265
Justeret R-kvadreret	0,9219
Standardfejl	0,4229
Observationer	18

ANOVA

	<i>fg</i>	<i>SK</i>	<i>MK</i>	<i>F</i>	<i>Signifikans F</i>
Regression	1	36,0830	36,0830	201,7605	0,0000
Residual	16	2,8615	0,1788		
I alt	17	38,9444			

	<i>Koefficienter</i>	<i>Standardfejl</i>	<i>t-stat</i>	<i>P-værdi</i>	<i>Nedre 95%</i>	<i>Øvre 95%</i>
Skæring	-0,9238	0,2322	-3,9779	0,0011	-1,4161	-0,4315
Indkomst	0,0070	0,0005	14,2042	0,0000	0,0059	0,0080



Bilag 2 Korrelationer

	<i>Besøg</i>	<i>Indkomst</i>	<i>Køn</i>	<i>Alder</i>	<i>Uddannelse</i>
Besøg	1				
Indkomst	0,9626	1			
Køn	0,2023	0,3043	1		
Alder	0,3168	0,2345	0,3208	1	
Uddannelse	-0,1304	-0,1332	-0,5110	-0,4701	1

Bilag 3 Data til regressioner

Besøg	Indkomst	Køn	Alder	Uddannelse
2	410	0	38	7
4	630	0	56	5
2	310	0	46	6
0	190	0	22	10
1	250	0	51	8
2	410	1	45	10
1	205	0	53	7
4	670	0	39	15
1	320	0	32	11
1	285	1	27	5
0	185	1	56	2
5	890	1	65	3
3	610	1	28	4
2	405	1	71	1
3	555	0	27	6
4	750	1	48	3
0	260	0	30	8
2	360	0	42	2

Besøg (antal årlige tandlægebesøg for respondenter).

Indkomst (respondentens indkomst i 1.000 kr.).

Køn (0 = mand; 1 = kvinde).

Alder (respondentens alder).

Uddannelse (antal år respondenter har brugt på uddannelse efter folkeskolen).