



Norges miljø- og biovitenskapelige
universitet
Fakultet for miljøvitenskap og teknologi
Institutt for naturforvaltning

Masteroppgave 2014
30 stp

Statistiske analyser av energiforbruk til oppvarming: Regresjonsanalyse av klimatiske påvirkningsfaktorer, og estimering av varmtvannsandel

Gregor Skaug Larsen

***“Reader, if thou to credit what is here
Art slow, 'tis no surprise, since I can scarce
Believe, who saw it all as clear as clear.”***

Dante Alighieri, La Divina Commedia, L'inferno, Canto XXIV 46-48

Forord

Med disse ord innleder jeg det foreløpige punktum for min tid som student. Den følgende masteroppgaven er kulminasjonen av seks års energirelaterte studier, hvorav de to siste årene har vært ved institutt for naturforvaltning på Norges miljø- og biovitenskapelige universitet, der jeg har studert fagspesialiseringen fornybar energi.

Jeg vil benytte anledningen til å rette en spesiell takk til Monica Havskjold, for all rettlledning, gode forslag og konstruktiv kritikk hun bidratt med som min veileder på denne oppgaven.

Jeg vil også takke min samboer (og våre to hunder) for å ha holdt ut med alle mine sene kvelder med jobbing, samt rette en takk til Martin Bruusgaard & CO AS for at de har latt meg benytte deres lokaler til å skrive oppgaven. Pappa fortjener også en stor takk for alle hans konstruktive tilbakemeldinger, samt Kjetil Eikeset for å være en inspirasjon til dannelsen.

Eventuelle feil og ufullstendige opplysninger i oppgaven er helt og holdent undertegnedes ansvar.

Helt til slutt - Kjære Mamma og Bror. Uten deres urokkelige tro på meg, ville jeg aldri vært her jeg er i dag. Deres udødelige støtte vil jeg alltid bære videre. Disse sidene dedikeres dere. Takk.

Sandvika, 8. desember 2014

.....

Gregor Skaug Larsen

Sammendrag

Denne studien tar for seg statistiske analyser av energiforbruk til oppvarming levert gjennom fjernvarme, for et utvalg bygg i Oslo. Hensikten med analysene har vært å undersøke hvorvidt meteorologiske faktorer som for eksempel vindstyrke, lufttrykk, solstråling etc. som supplement til utetemperatur, kan benyttes for å forklare energiforbruket. I tillegg er lineærregresjon analysert som metode for å beregne andelen oppvarmingsenergi som benyttes til varmt tappevann.

De meteorologiske variablene som inngår i analysen er hentet fra Meteorologisk institutts online database – ”eKlima”, og er registrert ved Blindern målestasjon i Oslo. De observerte energiforbrukene ble hentet inn av Kristensen i forbindelse med hans masteroppgave ved Norges Miljø- og Biovitenskapelige Universitet våren 2014 (Kristensen 2014).

I fremgangsmåten er det valgt å dele opp de observerte energiforbrukene i tre populasjoner – hhv. boligpopulasjon, næringspopulasjon og en totalpopulasjon. Sistnevnte utgjorde summen av boligpopulasjonen og næringspopulasjonen. Hensikten bak dette var antakelsen om at energiforbruket i boliger har annerledes karakteristikk enn energiforbruket til næringskunder, og at det ville være interessant å undersøke hvorvidt eventuelle sammenhenger funnet for totalpopulasjonen er gyldige for både bolig- og næringspopulasjonen.

Det er påvist at samtlige populasjoners forventede energiforbruk varierer med hvilken retning vinden kommer fra. Det kan forventes høyere energiforbruk ved vind fra nord og øst, enn vind fra sør og vest. Dette er sammenfallende med høyere forventede vindstyrker fra nord og øst enn fra sør og vest. I tillegg er det sammenfallende høyere forventningsverdi av energigrader ved vind fra nord og øst, enn fra sør og vest. Det er derfor konkludert med at konveksjonstapet er påvirket av hvilken retning vinden kommer fra.

Ved multippel lineærregresjon er det funnet at de klimatiske faktorene midlere vindstyrke, antall soltimer, midlere lufttrykk og energigrader er signifikante forklaringsvariabler for det observerte energiforbruket for samtlige analyserte populasjoner, og det er konkludert med at de nevnte meteorologiske omstendighetene forklarer en boligpopulasjons energiforbruk bedre enn lineærregresjon med energigrader som forklaringsvariabel alene.

Ved enkel lineærregresjon der energigrader er benyttet som forklaringsvariabel, er det funnet at denne metoden kan benyttes til estimering av energiforbruk til oppvarming av varmtvann for en boligpopulasjon. Det er likevel en svakhet ved metoden at energiforbruket til oppvarming av varmtvann anses konstant over flere år. Denne svakheten kan motvirkes i noen grad ved å analysere energiforbruket ett år om gangen.

Næringspopulasjonens energiforbruk er påvist ikke-lineært korrelert med energigrader. Kubisk estimering av energiforbruket er beregnet med høyere forklart varians enn lineærestimering. Ved forutsetning om kubisk korrelasjon er det påvist ulik temperatursensitivitet for energiforbruket til næringspopulasjonen og boligpopulasjonen. Det er i oppgaven konkludert med at ikke-lineær sammenheng og temperatursensitivitet bør utredes nærmere i fremtidige studier.

Abstract

The following master thesis presents a statistical analysis of energy delivered through district heating for a selection of buildings in Oslo, Norway. The research objective is to study whether meteorological factors (i.e. wind speed, air pressure, solar radiation) can be used as supplement to outdoor temperature to explain the energy consumption. In addition simple linear regression has evaluated as a method to estimate the percentage of heating energy spent on hot water.

The meteorological variables used in the analysis have been found in the Norwegian Meteorological Institutes online database – “eKlima”, and are registered at Blindern survey station in Oslo. The data on energy consumption has been collected by Kristensen for his master thesis at The Norwegian University of Life Sciences spring 2014 (Kristensen 2014). In the analysis procedure it has been decided to split the observed energy consumptions in three populations – residential, commercial, and a total population. The latter population consisted of the sum of energy spent in the residential and the commercial populations. The reason for this grouping of the populations was due to the assumption that commercial energy consumption has a different characteristic than residential energy consumption and it would be interesting to see whether potential meteorological explanatory variables would be valid for all populations.

The study has proven that the expected energy consumption of all populations varies according to which direction the wind blows. It can be expected higher energy consumption with wind from north and east than wind from the south and west. This correlates to higher expected wind speed and lower expected outdoor temperature with wind blowing from north and east, and lower expected wind speed and higher expected outdoor temperature with wind from the south and west. The study concludes that the convection heat loss is affected by from which direction the wind blows.

By multivariate linear regression it has been found that the climatic factors, average wind speed, number of hours with sunshine, average air pressure, and average energy degrees (difference between average outdoor temperature and 17 °C indoor temperature), are statistically significant as explanatory variables for the energy consumption in all populations.

The study concludes that these climatic variables better explain the energy consumption of the residential population than simple linear regression with energy degrees alone as explanatory variables does.

It has been found that simple linear regression can be used as a method to estimate the percentage of heat energy used for hot water in the residential population. Still it is a weakness to the method as it in this study assumes that the energy spent on heating water is constant during the observation period. This method weakness can be somewhat counter acted by analysing the energy consumption year by year.

The energy consumption of the commercial population has been proven non-linear correlated with energy degrees. Cubic estimation of the commercial energy consumption has proven to have a higher coefficient of determination than estimation by linear regression. By assumption of valid cubic correlation the energy consumption of the commercial population has been proven to have different outdoor temperature sensitivity than the residential population. The thesis concludes that a non-linear correlation and temperature sensitivity needs further study.

Innholdsfortegnelse

1 Innledning	12
1.1 Bakgrunn	12
1.2 Energiforbruk til oppvarming	13
1.3 Klimatiske omstendigheter	14
1.4 Problemstilling	15
1.5 Hva resultatene kan brukes til	16
2 Metode	17
2.1 Fjernvarme	17
2.2 Energigrader	17
2.3 Fyringssesongen	17
2.4 Forbruksdata	18
2.5 Tilfeldig utvalg og populasjon	18
2.6 Meteorologiske data	19
2.7 Forkastede observasjoner	19
2.8 Statistiske analyser	20
2.9 Statistisk fremgangsmåte	20
2.9.1 Test av normalfordeling	20
2.9.2 Test av internkorrelasjon	21
2.9.3 Test av nominal forventningsvariasjon	21
2.9.4 Multipel lineærregresjon – forklaringsvariablenes signifikans	21
2.9.5 Enkel lineærregresjon – estimering av varmtvannsandel	22
3 Resultater	24
3.1 Meteorologiske variable	24
3.1.1 Numeriske variable - Beskrivelse	24
3.1.2 Test for internkorrelasjon	25
3.1.3 Nominale variable - Beskrivelse	27
3.1.4 Vindstyrke fordelt på vindretning	27
3.1.5 Energigrader fordelt på vindretning	29
3.2 Totalpopulasjonens energiforbruk	31
3.2.1 Korrelasjon med meteorologiske variable	31
3.2.2 Energiforbruk i helg kontra hverdag	31
3.2.3 Forbruk fordelt på vindretning	32
3.2.4 Multipel lineærregresjon av numeriske, uavhengige variable	34

3.2.5 Enkel lineærregresjon mot temperatur	37
3.2.6 Estimering av varmtvannsandel, totalpopulasjon.....	39
3.3 Boligpopulasjonens energiforbruk.....	39
3.3.1 Korrelasjon med meteorologiske variable.....	39
3.3.2 Energiforbruk hverdag/helg.....	40
3.3.3 Energiforbruk fordelt på vindretning	41
3.3.4 Multippel lineærregresjon av numeriske, uavhengige variable	41
3.3.5 Enkel lineærregresjon mot temperatur	43
3.3.6 Estimering av varmtvannsandel, boligpopulasjon.....	44
3.4 Næringskundenes energiforbruk.....	45
3.4.1 Korrelasjon med meteorologiske variable.....	45
3.4.2 Energiforbruk hverdag/helg.....	46
3.4.3 Energiforbruk fordelt på vindretning	47
3.4.4 Multippel lineærregresjon av numeriske, uavhengige variable	47
3.4.5 Enkel lineærregresjon mot temperatur	49
3.4.6 Ikke-lineær regresjon mot temperatur	50
3.4.7 Estimering av varmtvannsandel, næringspopulasjon	51
3.5 Alternativ estimering av total varmtvannsandel.....	53
3.6 Estimering av årlig varmtvannsandel, næringspopulasjon.....	54
3.7 Temperatursensitivitet.....	55
4 Oppsummering.....	56
4.1 Energiforbruk og vind	56
4.2 Signifikante forklaringsvariabler.....	57
4.3 Estimering av varmtvannsandel.....	58
4.3.1 Totalpopulasjonen.....	58
4.3.2 Boligpopulasjonen.....	58
4.3.3 Næringspopulasjonen	59
4.4 Temperatursensitivitet.....	61
5 Konklusjon.....	62
Litteraturliste.....	64
Figurliste.....	66
Vedleggsliste.....	68

1 Innledning

1.1 Bakgrunn

Jordens klima er i endring. FN har uttalt at klimaendringene er den største utfordringen verden står overfor i dag (FN-sambandet 2014a). Øystater trues av havstigning (FN-sambandet 2014b), og stadig flere mennesker blir fordrevet fra sine hjem som følge av klimatiske endringer (The UN Refugee Agency 2014). Å trygge tilgangen på matressurser og rent drikkevann vil bli stadig mer utfordrende (The UN Refugee Agency 2014). Konsensus blant forskere er at jordens klima påvirkes av menneskelig aktivitet (Oreskes 2004). Det FN-ledede Intergovernmental Panel on Climate Change har anslått at antropogene drivhusgassutslipp i år 2100 var på det høyeste nivået i menneskelig historie, med 49 (+/- 5) Gt CO₂ eq/yr (gigatonn CO₂-ekvivalenter pr. år) (Edenhofer et al. 2014).

Som et ledd i bekjempingen av klimaproblematikken, har Den europeiske union (EU) innført en rekke ulike direktiver som medlemslandene plikter å følge. Eksempelvis besluttet EU-kommisjonen i desember 2008 å innføre det såkalte *fornybardirektivet* som inneholdt en rekke målsetninger for å redusere klimagassutslippene, som skal være oppfylt innen år 2020 (Rosvold 2014). Generelt kan en si at målsetningen er;

- Redusere utslipp av drivhusgasser med 20%^a
- 20% av totalt energiforbruk skal komme fra fornybare energikilder
- Energieffektiviteten skal økes med 20%

Det europeiske råd vurderte i februar 2011 at energieffektiviseringen ikke gikk helt etter planen. For å sikre at målet oppnås, ble resultatet at Europaparlamentet 11. september 2012 vedtok *energieffektiviseringsdirektivet* (EED). Direktivet ble senere vedtatt i Europarådet 4. oktober 2012 (Europaportalen 2013).

Selv om Norge ikke er medlem av EU, vil de direktiver som vedtas i Europaparlamentet få direkte konsekvenser for Norge gjennom Det europeiske økonomiske samarbeidsområde (EØS), dersom Norge fortsatt skal ha fri tilgang til det indre markedet. Uten at det går nærmere inn på det juridiske og politiske her, betyr det i praksis at Norge implementerer de EØS-relevante direktiver som vedtas i EU, alternativt med tilpasninger som ikke er til hinder for fri konkurranse. 19. desember 2011 innlemmet EØS-komiteen fornybardirektivet, der det fremgår at Norges mål blant annet er å oppnå at 67,5 % av energibruken skal være fra

^a Utslippet reduksjon er sammenliknet med utslippsmengden i år 1990

^b Totalandelen i perioden er beregnet fra verdiene i tabell 7 og tabell 9 fra Kristensens masteroppgave (se vedlegg 7)

^c Hver målepunktsid dekker ett helt bygg, av ukjent størrelse, antall beboere, og ansatte

fornybare energikilder innen år 2020 (Olje- og energidepartementet 2011). Statistisk sentralbyrå (SSB) estimerer at andelen fornybar energi vil komme opp i 73% innen år 2020 som følge av en eventuell implementering av energieffektiviseringsdirektivet (Bøeng & Rosnes 2013).

Rent praktisk kan energieffektivisering gjennomføres på flere områder, herunder transportsektoren, industrien, husholdninger m.m. Husholdninger stod for ca. 21% av det totale energiforbruket i Norge i 2011 ifølge SSB (Bøeng & Holstad 2013). NVE anslår videre at ca. 30% av husholdningenes energiforbruk dekker det el-spesifikke behovet, mens 70% av energiforbruket benyttes til oppvarming av rom (58%) og varmtvann (12%) (Magnussen et al. 2011).

1.2 Energiforbruk til oppvarming

I et bygg skal det helt generelt leveres energi til belysning, installert utstyr, oppvarming av vann og oppvarming av rom. Energien levert til belysning og utstyr antas å gi en energitjeneste i form av utstyrets funksjon, samt et varmetilskudd til bygget. Eksempelvis beregnes 60% av energien levert til belysning og utstyr i boligblokker og småhus som varmetilskudd til romoppvarming (Standard Norge 2007).

Dersom en gjør en antakelse om at oppvarming av varmtvann ikke avhenger av utetemperaturen, kan en helt basalt dele energibehovet til et bygg i to deler;

- Temperaturuavhengig
- Temperaturavhengig

Dersom en antar en lineær sammenheng mellom energiforbruk og temperatur, kan dette uttrykkes matematisk med;

Formel 1 Lineærsammenheng mellom energiforbruk og temperaturdifferanse

$$Q = \alpha + \beta \Delta T$$

Q uttrykker energiforbruket, α er konstant og uavhengig av temperatur, mens β uttrykker en kalkulert varmetapskoeffisient som er proporsjonal med temperaturdifferansen mellom inne- og utetemperatur ΔT .

Den energien som kreves til oppvarming av varmtvann, kan ansees som upåvirket av utetemperatur og er således *temperaturuavhengig*. For et utvalg kunder i Oslo er det tidligere estimert andeler av levert energi fra fjernvarme er brukt til oppvarming av varmtvann (Kristensen 2014):

Tabell 1 Estimering av energiandel fra fjernvarme til oppvarming av varmtvann (Kristensen 2014)

	Boligkunder	Næringskunder	Hele kundeutvalget ^b
År	Andel energi til varmtvann	Andel energi til varmtvann	Andel energi til varmtvann
2008	35,7 %	16,2 %	20,8 %
2009	35,5 %	14,2 %	19,1 %
2010	31,5 %	11,4 %	15,6 %
2011	35,0 %	15,7 %	20,3 %
Totalt i perioden^b	34,3 %	14,1 %	18,7 %

I estimeringene i tabell 1 er det antatt at varmtvannsforbruket er mer eller mindre konstant over ett år for en stor populasjon. Varmtvannsandelens størrelse varierer fra år til år som følge av at mengden energi til romoppvarming varierer med utetemperaturen, og at mengden energi til oppvarming av varmtvann er beregnet år for år (Kristensen 2014).

Energibehovet til romoppvarming er teoretisk sett direkte proporsjonal med temperaturdifferansen mellom inne- og utetemperatur og således *temperaturavhengig*. Fra et akademisk perspektiv kan det være høyst interessant å undersøke om *andre* klimatiske faktorer eventuelt påvirker energiforbruket ved å påvirke varmetapskoeffisienten β .

1.3 Klimatiske omstendigheter

I Osloområdet bor brorparten av beboerne i blokker og andelen eneboliger er avtakende. Tallet på blokkleiligheter økte med 30% mellom 2001 og 2011, samtidig som tallet på eneboliger økte med bare 4% i samme periode (Utne & Andersen 2011). Dette medfører at forholdet mellom totalt oppvarmet gulvflate og totalt utvendig areal er avtakende. Det faktum at andelen blokkleiligheter øker, *antas* å bidra til mindre temperaturfølsomhet i energibruken totalt sett (Pöyry Management Consulting AS 2014).

Det har vist seg vanskelig å oppdrive forskningslitteratur som konkret omhandler hvilken effekt meteorologiske omstendigheter (utelatt temperatur) direkte har på energiforbruket til

^b Totalandelen i perioden er beregnet fra verdiene i tabell 7 og tabell 9 fra Kristensens masteroppgave (se vedlegg 7)

oppvarming. Det har således vært funnet nødvendig å benytte seg av de fundamentale termofysiske prinsipper for varmeoverføring som grunnlag for hypotese.

Basert på kjente former for varmetransport som konveksjon, radiasjon, og transmisjon (Çengel 2006), er den overordnede hypotesen at energibehovet *påvirkes* av klimatiske omstendigheter. (Se konveksjonseksempel, vedlegg 3). Teoretisk sett bør konveksjonsvarmetapet øke ved økende vindhastigheter, hvilket vil gjøre det momentane energibehovet større. Motsatt bør antallet soltimer bidra til å redusere behovet for levert energi til oppvarming, som følge av økt energitilførsel gjennom solinnstråling.

Følgelig kan en således anta at naturpåvirkninger som for eksempel vind, regn, eller kraftig sol øker hhv. byggets momentane konveksjonstap, transmisjonstap og absorpsjon. Klimatiske forhold kan også potensielt endre beboeres/brukeres adferdsmønster, for eksempel ved økt varmtvannsforbruk i kalde, våte perioder. Det vil være således være interessant å undersøke hvorvidt det er en statistisk sammenheng mellom energibruk og meteorologiske data. Ettersom det benyttes døgnverdier for observasjonene hos Meteorologisk institutt, vil det ikke bli sett nærmere på variasjoner i beboere/brukeres adferdsmønster over døgnet i den videre analysen.

1.4 Problemstilling

Den følgende masteroppgaven tar for seg statistiske analyser av energi til oppvarming, sett i sammenheng med meteorologiske data for et utvalg bygg tilknyttet fjernvarme i Oslo. Hensikten er å vurdere om klimatiske omstendigheter (ut over utetemperatur) påvirker energiforbruket. I tillegg blir det sett på lineærregresjon som metode for å estimere andelen oppvarmingsenergi som benyttes til oppvarming av varmtvann. Det vitenskapelige arbeidet tar utgangspunkt i følgende problemstillinger:

”Er det grunnlag for å si at meteorologiske omstendigheter, utover temperatur, påvirker energiforbruket til romoppvarming?”

”Kan lineærregresjon benyttes som metode for å estimere andelen oppvarmingsenergi til varmtvann?”

1.5 Hva resultatene kan brukes til

Resultatene fra denne studien vil kunne tilføre økt kunnskap om potensiell sammenheng mellom vær og energiforbruk til oppvarming. Dette vil kunne brukes til å forbedre metoden for temperaturkorrigering (klimakorrigerings) av energiforbruket.

Videre studier vil kunne se på metoder for å inkorporere meteorologiske omstendigheter i beregninger av andelen energi som benyttes til oppvarming av varmtvann.

Videre vil økt kunnskap om hvorvidt meteorologiske faktorer påvirker forbruket kunne bidra til å utvikle predikasjonsmetoder for energiforbruket til oppvarming, uavhengig av energibærer. Å kunne si noe om hvordan forbruket vil utvikle seg over tid, vil være svært nyttig informasjon for å sikre leveringsstabilitet. I tillegg kan eventuelle meteorologiske sammenhenger med energiforbruk være relevante for utforming av tekniske byggeforskrifter, slik at klimatiske faktorer kan tas hensyn til ved nye konstruksjoner.

2 Metode

2.1 Fjernvarme

For den statistiske analysen er det besluttet å bruke forbruksdata fra fjernvarme. Dette begrunnes i at fjernvarme tilføres et bygg for å dekke to behov- oppvarming av rom, og oppvarming av varmt tappevann. Elektrisk energi blir, i motsetning til fjernvarme, også benyttet til å dekke forbruksenergi, som for eksempel utebelysning, hvilket er irrelevant for romoppvarming. Forbruksdata fra fjernvarme anses dermed som best egnet for å analysere energiforbruk til oppvarming, for å i størst mulig grad redusere diskrepanser i forbruksdatasettet.

2.2 Energigrader

I beregninger av bygningers energibehov etter NS3031, skal det tas hensyn til blant annet varme fra apparater og personer, såkalte *internlaster* (Standard Norge 2007). Det antas vanligvis at varmetilskudd fra internlaster utgjør ca. 3 °C økning i innnetemperaturen (Pöyry Management Consulting AS 2014). I den videre analysen tas det forbehold om konstant midlere innnetemperatur på 20 °C, hvilket medfører at temperaturdifferansen vil til enhver tid være differansen mellom observert utetemperatur og 17 °C (i hensyntatt internlaster). Denne temperaturdifferansen er ved Meteorologisk institutts online klimadatabase "eKlima" registrert som *energigrader*.

2.3 Fyringssesongen

Det er valgt å definere en tidsbegrenset fyringssesong for avgrensing av observerte variabler. Det er hovedsakelig gjort av matematiske hensyn. Observert temperaturdifferanse vil i de varme månedene av året gå mot verdien 0, og således være ugunstig å analysere som eventuell divisor. I tillegg vil en definert fyringssesong sikre at observerte energiforbruk er gjort med like forutsetninger. Fjernvarmebasert oppvarming av rom gjøres primært gjennom radiatorer/gulvvarme, og oppvarming av ventilasjonsluft. Når midlere utetemperatur kommer over 9 °C stopper oppvarming gjennom radiator/gulvvarme, og oppvarmingsbehovet dekkes gjennom ventilasjonsanlegget. Når midlere utetemperatur kryper under 11 °C gjenopptas oppvarming gjennom radiatorer/gulvvarme igjen (Pedersen et al. 2008). Fyringssesongen i Oslo er normalt fra ca. 14. september til 7. mai (Josteinm 2009).

2.4 Forbruksdata

Rådatasettet som er benyttet er samlet inn fra Hafslund av Kristensen i forbindelse med hans masteroppgave ved Norges miljø- og biovitenskapelige universitet (Kristensen 2014).

Kildematerialet dekket forbruksdata time for time fra 01.01.2008 til og med 31.12.2013, for bortimot 600 bygg i Osloområdet. Ved nærmere analyse viste deg seg at flere av målepunkt-IDene^c var duplikater, og flere var ikke kontinuerlige gjennom måleperioden.

Som forutsetning for videre analyse ble det forutsatt at målepunkt-ID skulle være unik i populasjonen, og forbruksdata skulle være kontinuerlig observert i perioden 01.01.2008 til og med 31.12.2011. Denne tidsperioden ble ansett for å være tilstrekkelig representativ for energiforbruket i videre analyse. Samtlige energiforbruk ble så inpsisert grafisk over hele overnevnte tidsperiode. Der forbruket ikke fulgte en noenlunde harmonisk svingning over året, ble målepunktsiden forkastet fra populasjonen slik at ikke vedvarende målefeil skulle være representert. Forbruksdata for hver målepunkt-ID ble så summert for hvert døgn, slik at hver målepunkt-ID hadde 365 forbruksobservasjoner pr. år (366 i 2008 grunnet skuddår). Etter sortering gjenstod et utvalg på 93 unike målepunkt-IDer. Disse ble matchet mot tariffliste fra Hafslund (Kristensen 2014), supplert av informasjon hentet fra masteroppgaven til Martin S. Kristensen, for å knytte målepunkt-IDene til kundekategoriene ”boligkunde” eller ”næringskunde”. Dette ble gjort for å muliggjøre analyse av disse forbrukskategoriene hver for seg da det er antatt at de har relativt ulike karakteristikk. Av populasjonen på 93 unike målepunkt-IDer var 18 boligkunder, og 75 næringskunder. Disse ble så tilordnet anonymiserte verdier/avn i henhold til undertegnet taushetserklæring (se vedlegg 2). Målepunkt-IDer tilknyttet boligkunder ble tilordnet verdier/avn fra ”Y1” til ”Y18”, og næringskunder ble tilordnet verdier/avn fra ”Z1” til ”Z75”

2.5 Tilfeldig utvalg og populasjon

For at næringskunder og boligkunder skulle kunne bli analysert på relativt likt grunnlag, ble det besluttet at populasjonene skulle være like store. Følgelig ble det besluttet å gjøre et tilfeldig utvalg av 18 næringskunder fra populasjonen på 75. Dette ble gjort slik at totalpopulasjonen skulle være likt vektet mellom boligkunder og næringskunder. Det tilfeldige utvalget ble gjennomført i statistikkprogrammet RStudio (se kapittel 2.8):

^c Hver målepunktsid dekker ett helt bygg, av ukjent størrelse, antall beboere, og ansatte

```
> str(id.vectors)
'data.frame':    75 obs. of  2 variables:
 $ Idn: Factor w/ 75 levels "Z1","Z10","Z11",...: 1 12 23 34 45 56 67 74 75
2 ...
 $ Idb: Factor w/ 18 levels "Y1","Y10","Y17",...: 2 12 13 14 15 16 17 18 19
3 ...

> sample(Idn, 18)
[1] Z35 Z65 Z24 Z19 Z54 Z32 Z51 Z21 Z17 Z62 Z61 Z69 Z16 Z43 Z60 Z3  Z29 Z6
```

(Linjen markert med gult angir de tilfeldig utvalgte anonymiserte målepunkt-IDene)

Øvrige næringskunder ble forkastet fra videre analyse.

Til slutt ble forbruket i observasjonsdøgnet summert for henholdsvis næringskunder, boligkunder og den totale populasjonen, og deretter sortert etter dato, fra 01.01.2008 til 31.12.2011.

2.6 Meteorologiske data

Fra Meteorologisk institutts (MI) arkivdatabase ble det hentet ut observerte døgnverdier i tidsperioden 01.01.2008 til og med 31.12.2011.

Som forklaringsvariabler for konveksjon ble det hentet ut observasjoner av vindretning (kl. 12 i observasjonsdøgnet), midlere vindhastighet, midlere lufttrykk, spesifikk luftfuktighet og relativ luftfuktighet.

Som forklaringsvariabler for radiasjon (stråling) ble det hentet ut observasjoner av midlere skydekke og antall soltimer.

Som forklaringsvariabel for transmisjon ble det hentet ut observasjoner av nedbørsmengde i observasjonsdøgnet.

Som forklaringsvariabel (og styringsvariabel) for alle typer energioverføring ble det hentet ut observasjoner av energigrader (temperaturdifferansen mellom midlere utetemperatur og 17 °C).

Samtlige observasjoner er hentet ned med tallfestet usikkerhet angitt av MI. Usikkerhet i observasjonen var skalert fra 0 til 6 der 0 er ”veldig sikker” og 6 er ”veldig usikker”.

2.7 Forkastede observasjoner

De meteorologiske observasjonene ble sortert etter dato og satt sammen med energiforbruksdataene. Det ble så valgt å forkaste en rekke observasjoner (samtlige registrerte

variabler i observasjonsdøgnet). Alle observasjoner gjort utenom fyringssesongen (se kapittel 2.3) ble forkastet. Deretter ble alle observasjoner der MI institutt hadde anslått observasjonsusikkerhet > 2 (se kapittel 2.6) forkastet fra videre analyse.

De observasjonene der det manglet observert data for en meteorologisk variabel ble ikke forkastet ettersom de statistiske metodene som ble benyttet (se kapittel 2.9) tar hensyn til manglende data.

Det endelige datasettet bestod til slutt av 901 observasjoner (døgnverdier) av energiforbrukene (se kapittel 2.4) og de meteorologiske variablene (se kapittel 2.6).

2.8 Statistiske analyser

Alle statistiske analyser er utført med statistikkprogrammet "RStudio", som benytter programmeringsspråket "R". Programmeringsspråket er amerikanskbasert, hvilket innebærer at der direkte avskrift fra RStudio er presentert i oppgaven, er punktum benyttet som desimaltegn.

Alle fremstillinger av tabeller, er gjort i Microsoft Excel, der komma er brukt som desimaltegn.

2.9 Statistisk fremgangsmåte

2.9.1 Test av normalfordeling

Samtlige numeriske variable testes for normalfordeling. Dette er for å kunne avgjøre hvilken korrelasjonstest som er best egnet. Test av normalfordeling gjøres med Shapiro-Wilks test. Shapiro-Wilk test i R, returnerer en W-koeffisient og en p-verdi. W-koeffisienten, sammen med n antall observasjoner benyttes for å finne den overnevnte p-verdien for testen.

Normalitetstestene er utført med hypotesene;

"H0 : Fordelingen er ikke ulik en normalfordeling"

"H1 : Fordelingen er ulik en normalfordeling"

H0 er forkastet dersom returnert p-verdi $< 0,05$.

2.9.2 Test av internkorrelasjon

For å unngå feilaktige sammenhenger, er variablene undersøkt for internkorrelasjon, som i praksis betyr at variablene er avhengige. Der internkorrelasjon er påvist, er det valgt å forkaste den laveste korrelasjon til responsvariabelen. Responsvariabelen er typisk den variabelen som skal forklares av andre observasjoner. Følgelig er responsvariabelen pr. definisjon en avhengig variabel. Etter responsvariabelen ble valgt, ble forklaringsvariablene definert. Ettersom disse skulle forklare responsvariabelen, må de være uavhengige av hverandre. Med bakgrunn i responsvariabelens uparametriske (ikke-normalfordelte) fordeling ble det valgt å gjennomføre Spearman korrelasjonstest der samtlige variabler ble parett. Det ble valgt å definere signifikant internkorrelasjon dersom absoluttverdien av korrelasjonskoeffisienten var større enn 0,5^d:

Formel 2 Definisjon av signifikant korrelasjon

$$\text{signifikant korrelasjon} \equiv |\text{Korrelasjon}| > 0,5$$

2.9.3 Test av nominal forventningsvariasjon

Videre ble det utført test av responsvariabelens forventningsvariasjon. Dette er gjennomført ved først å kontrollere responsvariabelens fordeling i hvert utfallsrom i den nominale^e (ikke-numeriske) variabelen. Hvorvidt fordelingene er parametriske eller ikke, sammen med antallet av utfallsrom, avgjorde hvilken statistisk test som var best egnet (se vedlegg 1). Test for ulik forventningsvariasjon er utført med følgende hypoteser:

"H0 : Forventningsverdien til responsvariabelen er ikke ulik i utfallsrommene"

"H1 : Forventningsverdien til responsvariabelen er ulik i utfallsrommene"

H0 er forkastet dersom returnert p-verdi < 0,05.

2.9.4 Multippel lineærregresjon – forklaringsvariablenes signifikans

For å påvise en sammenheng mellom energiforbruket og de uavhengige numeriske forklaringsvariablene, er det valgt å benytte multippel lineærregresjon. Lineærregresjon gjennomført i R, returnerer koeffisienter for alle ønskede forklaringsvariabler, deres

^d Korrelasjonskoeffisienten er en relativ størrelse der $-1 \leq \text{korrelasjonskoeffisient} \leq 1$.

0 indikerer ingen korrelasjon, mens -1 og 1 indikerer hhv. kraftig negativ og kraftig positiv korrelasjon

^e Nominal/Ordinal variabel er ikke-numeriske variabler med og uten rang. Eks. Ordinal – "Small", "Medium", "Large", har rangert ordning, mens nominal har urangert ordning, Eks. "gul", "grønn", "blå".

signifikansnivå, samt regresjonens statistiske signifikans som helhet. Sistnevnte returneres som en p-verdi, der statistisk signifikans er definert dersom $p < 0,05$.

De forklaringsvariablene som ikke var signifikante ble forkastet, og ny multipl lineærregresjon ble gjennomført. De signifikante forklaringsvariablene ble så pare med de kalkulerte koeffisientene, slik at estimert energiforbruk kunne uttrykkes som en lineærfunksjon av flere variable på formen:

Formel 3 Generalisert flérvareibelfunksjon

$$Y(X_{1,i}, X_{\dots,i}, X_{n,i}) = \alpha + \beta X_{1,i} + \gamma X_{\dots,i} + \delta X_{n,i}$$

Funksjonsuttrykket ble så brukt til å kalkulere et estimert energiforbruk for hver enkelt observasjonsrekke.

Til slutt ble det observerte energiforbruket (responsvariabelen) korrelasjonstestet mot det estimerte energiforbruket. I tillegg til korrelasjonstest, er forklart varians (R^2) benyttet som måleparameter på hvor god estimeringen er. Forenklet sagt er forklart varians et relativt mål på hvor godt de observerte verdiene er reproduert av estimeringen. Ved $R^2 = 1$ er de observerte verdiene perfekt forklart av estimeringen.

2.9.5 Enkel lineærregresjon – estimering av varmtvannsandel

For å estimere andelen energi som går til varmtvann ble det benyttet enkel lineærregresjon (regresjon med én forklaringsvariabel) mellom observert energiforbruk og temperaturdifferanse. Dette ble ansett som en adekvat tilnærming ettersom lineærfunksjonen med én forklaringsvariabel uttrykkes på formen;

Formel 4 Generalisert énvareibelfunksjon

$$Y(X_i) = \alpha + \beta X_i$$

hvilket er identisk med antakelsen i kapittel 1.2 om at energiforbruk til varmtvann er uavhengig av utetemperatur. I tillegg er det også antatt at varmtvannsforbruket er relativt konstant for en stor populasjon. Ved å sette energigrader = 0, ble det antatt at konstanten α uttrykte energiforbruket til varmtvann for én enkelt observasjonsrekke, som i datasettet utgjør 1 døgn.

Årsforbruket av energi til varmtvann for populasjonen ble så kalkulert ved å multiplisere α med antall døgn pr. år (366 døgn i 2008 grunnet skuddår).

Beregningene av estimert levert energi etter formel 4, og estimert energi til varmtvann er utført inkludert de forkastede observasjonsrekker (se kapittel 2.7), slik at varmtvannsforbruket

i sommermånedene får virke inn på den totale andelen over året. Totalt antall observasjoner utgjorde 1461 i perioden 01.01.2009 til 31.12.2011.

Andelen energi som er brukt til varmtvann ble kalkulert som forholdet mellom estimert årlig mengde energi til oppvarming av varmtvann, og summert årlig mengde observert levert energi fra fjernvarme.

3 Resultater

Videre følger resultatene fra studien. Undersøkelsen av de meteorologiske variable er presentert først. Deretter følger undersøkelsen av den totale populasjonen. I disse kapitlene (3.1 og 3.2) er resultatene generert i RStudio presentert slik de fremstår i dataprogrammet. I de repetitive undersøkelsene for henholdsvis boligpopulasjonen og næringspopulasjonen er dette henvist til vedlegg 6 og 7. Datasettet som er brukt som underlag for analysen bestod av totalt 901 observasjonsrekker (se kapittel 2.7). Datasettet som er benyttet for å beregne varmtvannsandeler bestod av 1461 observasjonsrekker (uten forkastede observasjonsrekker).

3.1 Meteorologiske variable

3.1.1 Numeriske variable - Beskrivelse

Variabelen *middelvind* er definert som midlere vindstyrke i observasjonsdøgnet, målt i m/s. Fordelingen av de observerte verdiene for *middelvind* kan ikke bevises normalfordelt i henhold til Shapiro-Wilk test (se vedlegg 4). I all videre korrelasjonsanalyse der midlere vindstyrke inngår, er det således benyttet Spearman korrelasjonstest (se vedlegg 1). Det er ingen manglende (ikke-registrerte) observasjoner av midlere vindstyrke i datasettet.

Variabelen *gd17* er observerte energigrader, hvilket er temperaturdifferansen mellom observert midlere utetemperatur, og 17° C.

Fordelingen av de observerte verdiene for *gd17* kan ikke bevises normalfordelt i henhold til Shapiro-Wilk test. I all videre korrelasjonsanalyse der temperaturdifferansen inngår, er det således benyttet Spearman korrelasjonstest. Det er ingen manglende observasjoner av energigrader i datasettet.

Variabelen *middelsky* er midlere skydekke i observasjonsdøgnet, angitt i 1/8 intervall.

Fordelingen av de observerte verdiene for *middelsky* kan ikke bevises normalfordelt i henhold til Shapiro-Wilk test. I all videre korrelasjonsanalyse der midlere skydekke inngår, er det således benyttet Spearman korrelasjonstest. Det er ingen manglende observasjoner av midlere skydekke i datasettet.

Variabelen *sol* er observert antall soltimer i observasjonsdøgnet.

Fordelingen av de observerte verdiene for *sol* kan ikke bevises normalfordelt i henhold til Shapiro-Wilk test. I all videre korrelasjonsanalyse der antall soltimer inngår, er det således benyttet Spearman korrelasjonstest. Det er registrert 606 observasjoner av antall soltimer i datasettet, hvilket innebærer 295 manglende observasjoner.

Variabelen *middelp* er midlere lufttrykk i observasjonsdøgnet målt i hPa.

Fordelingen av de observerte verdiene for *middelp* kan ikke bevises normalfordelt i henhold til Shapiro-Wilk test. I all videre korrelasjonsanalyse der midlere lufttrykk inngår, er det således benyttet Spearman korrelasjonstest. Det er ingen manglende observasjoner av midlere lufttrykk i datasettet.

Variabelen *nedbor* er midlere nedbørsmengde i observasjonsdøgnet, målt i mm.

Fordelingen av de observerte verdiene for *nedbor* kan ikke bevises normalfordelt i henhold til Shapiro-Wilk test. I all videre korrelasjonsanalyse der nedbørsmengde inngår, er det således benyttet Spearman korrelasjonstest. Det er registrert 501 observasjoner av nedbørsmengde, hvilket innebærer 400 manglende observasjoner.

Variabelen *spesRH* er midlere spesifikk luftfuktighet i observasjonsdøgnet, målt i g/kg

Fordelingen av de observerte verdiene for *spesRH* kan ikke bevises normalfordelt i henhold til Shapiro-Wilk test. I all videre korrelasjonsanalyse der spesifikk luftfuktighet inngår, er det således benyttet Spearman korrelasjonstest. Det er ingen manglende observasjoner av spesifikk luftfuktighet i datasettet.

Variabelen *middelRH* er midlere relativ luftfuktighet i observasjonsdøgnet, målt i prosent.

Fordelingen av de observerte verdiene for *middelRH* kan ikke bevises normalfordelt i henhold til Shapiro-Wilk test. I all videre korrelasjonsanalyse der midlere relativ luftfuktighet inngår, er det således benyttet Spearman korrelasjonstest. Det er ingen manglende observasjoner av midlere relativ luftfuktighet i datasettet.

3.1.2 Test for internkorrelasjon

De beregnede Spearmans korrelasjonskoeffisienter for de numeriske variablene er vist i tabellen under:

Tabell 2 Korrelasjonskoeffisienter for de numeriske meteorologiske variablene

VARIABEL	<i>middelvind</i>	<i>gd17</i>	<i>middelsky</i>	<i>sol</i>	<i>middelp</i>	<i>nedbor</i>	<i>spesRH</i>	<i>middelRH</i>
<i>middelvind</i>	1	-0,2197	0,1717	-0,1733	-0,3160	0,1421	0,1061	-0,2753
<i>gd17</i>	-0,2197	1	0,0320	-0,2217	0,0387	-0,1778	-0,8494	0,2441
<i>middelsky</i>	0,1717	0,0320	1	-0,8334	-0,1513	0,1907	0,2749	0,5768
<i>sol</i>	-0,1733	-0,2217	-0,8334	1	0,0307	-0,1720	-0,0870	-0,6211
<i>middelp</i>	-0,3160	0,0387	-0,1513	0,0307	1	-0,3459	-0,1405	-0,1094
<i>nedbor</i>	0,1421	-0,1778	0,1907	-0,1720	-0,3459	1	0,3339	0,2967
<i>spesRH</i>	0,1061	-0,8494	0,2749	-0,0870	-0,1405	0,3339	1	0,2368
<i>middelRH</i>	-0,2753	0,2441	0,5768	-0,6211	-0,1094	0,2967	0,2368	1

Fra tabell 2 er det observert internkorrelasjon ($|korrelasjon| > 0,5$) mellom forklaringsvariablene:

spesRH og *gd17*:

Der er observert negativ korrelasjon mellom spesifikk luftfuktighet og energigrader. Dette innebærer at observert spesifikk luftfuktighet synker når observerte energigrader øker (temperaturen synker), hvilket henger sammen med at varm luft kan holde på større mengde vann i gassform enn kald luft (Rajput 2009).

middelRH og *middelsky*:

Det er observert positiv korrelasjon mellom midlere relativ luftfuktighet og skydekke. Dette innebærer at observert midlere relativ luftfuktighet øker dersom skydekkeandelen øker. Det antas at dette skyldes at lufttemperaturen synker når deler av strålingstilførselen fra sola reflekteres i skyene. Ved synkende temperatur øker relativ luftfuktighet (Rajput 2009).

sol og *middelsky*:

Det er observert negativ korrelasjon mellom antall soltimer og skydekke. Det innebærer at økende skydekke gir færre soltimer og vice versa.

3.1.3 Nominale variable - Beskrivelse

Det er definert 3 ulike stokastiske variabler med kategorisk (ikke-numerisk) utfallsrom:

1. *vindretn* med utfallsrommene "N" (nord), "E" (øst), "S" (sør) og "W" (vest).
2. *helg* med utfallsrommene "J" (ja) og "N" (nei).

For å forenkle analysen av vindretning ble det valgt å konvertere de observerte numeriske vindretningene til nominale variabler grunnet diskontinuitet ved nordlige vindretninger (for eksempel er 359° i praksis fra nord, på samme måte som 001°).

Utfallsrommene til variabelen *vindretn* er definert slik:

- "N" er alle observasjoner av vindretning kl. 12.00 i observasjonsdøgnet der registrert vindretning er fra og med 315° til og med 44° .
- "E" er alle observasjoner av vindretning kl. 12.00 i observasjonsdøgnet der registrert vindretning er fra og med 45° til og med 134° .
- "S" er alle observasjoner av vindretning kl. 12.00 i observasjonsdøgnet der registrert vindretning er fra og med 135° til og med 224° .
- "W" er alle observasjoner av vindretning kl. 12.00 i observasjonsdøgnet der registrert vindretning er fra og med 225° til og med 314° .

vindretn er således en nominal variabel (ikke-numerisk, uten rangering).

Utfallsrommene til variabelen *helg* er definert slik:

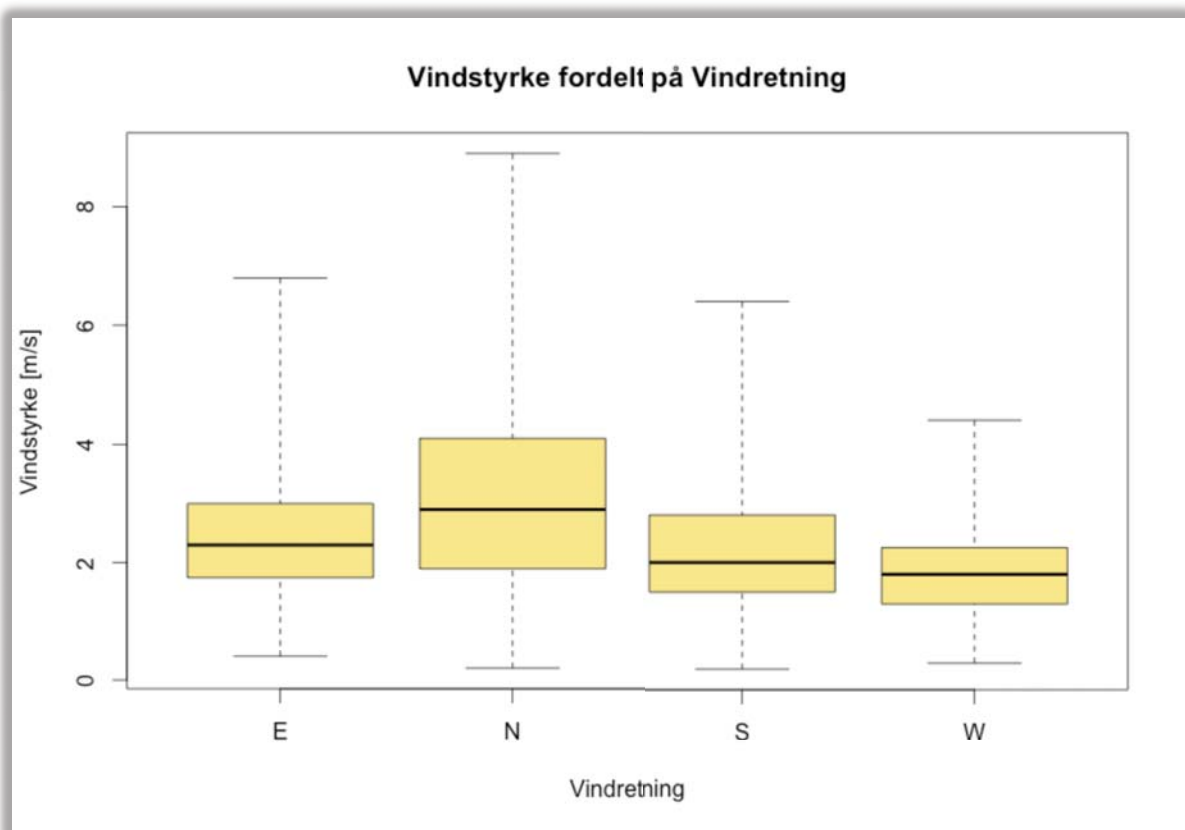
- "J" er alle observasjoner gjort på dagene lørdag eller søndag.
- "N" er alle observasjoner gjort på dagene mandag, tirsdag, onsdag torsdag eller fredag.

helg er således en nominal variabel (ikke-numerisk, uten rangering).

3.1.4 Vindstyrke fordelt på vindretning

Variabelen *middelvind* (midlere vindstyrke) er numerisk, kontinuerlig, og angitt i m/s.

Fordelingen av observert vindstyrke i utfallsrommene den nominale stokastiske variabelen *vindretn*, er vist under:



Figur 1 Boksplot av observert midlere vindstyrke fordelt på observert vindretning

Av figur 1 fremgår det visuelt at det er forskjeller i forventningsverdien (fet, horisontal, sort strek) til vindstyrken mellom himmelretningene.

Ettersom variabelen *vindretn* er nominal med fire utfallsrom, og fordelingen av *middelvind* i utfallene er ikke bevist normalfordelt i noen av utfallsrommene i hh.t. Shapiro-Wilk test (se vedlegg 4), er det benyttet Kruskal-Wallis hypotesetest, der;

H0 : Forventningsverdien til observert vindretning er ikke ulik i utfallsrommene

H1 : Forventningsverdien til observert vindretning er ulik i utfallsrommene

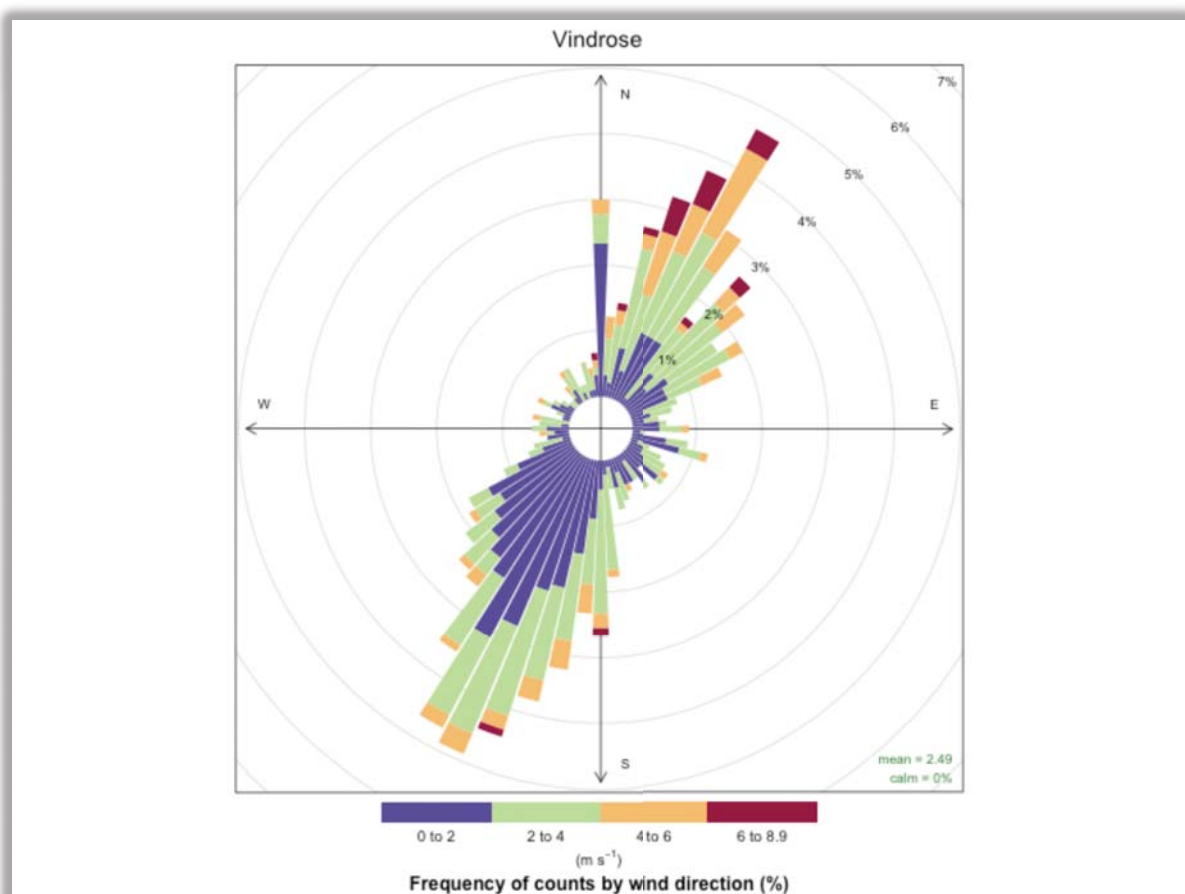
Kruskal-Wallis rank sum test

data: middelvind by vindretn

Kruskal-Wallis chi-squared = 76.6583, df = 3, p-value < 2.2e-16

H0 er forkastet da Kruskal-Wallis hypotesetest gir $p \ll 0,05$ hvilket indikerer at vindstyrken varierer med vindretningen.

Plottet sammenheng mellom vindstyrke, vindretning er vist i vindrosen under:

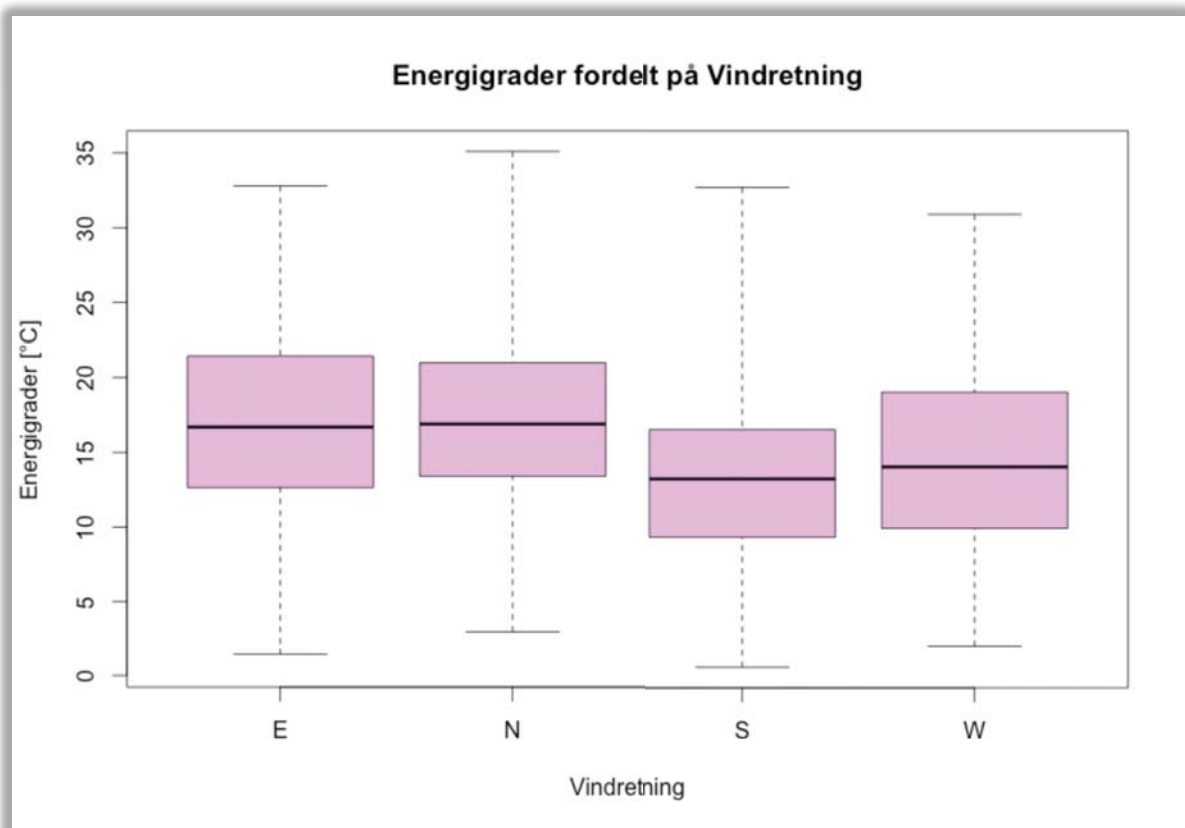


Figur 2 Vindroseplot av observert midlere vindstyrke og observert vindretning i grader

Figur 2 viser et vindroseplot av vindretning og vindstyrke. Lengden fra origo angir den relative frekvensen av observasjoner i 5°s- intervall. Fargen angir den observerte vindstyrken. Tilsynelatende er frekvensfordelingen relativt lik mellom nordlig og sørlig retning. Likefullt fremgår det tydelig at vind fra nordlig retning har høyere frekvens av de høyeste observerte vindstyrkene.

3.1.5 Energigrader fordelt på vindretning

Variabelen *gd17* er observerte energigrader, og er kun påvist normalfordelt i henhold til Shapirotest i utfallsrommet "E" til variabelen *vindretn*. Fordelingen av observerte energigrader i utfallsrommene variabelen *vindretn* er vist i figuren under:



Figur 3 Foredling av observerte energigrader på observert vindretning

Av figur 3 fremgår det visuelt at det er forskjeller i forventningsverdien til energigrader mellom de observerte vindretningene.

Ettersom variabelen *vindretn* er nominal med fire utfallsrom, og fordelingen av *gd17* i utfallene påvist normalfordelt i hh.t. Shapiro-test i ett av utfallsrommene er det benyttet Kruskal-Wallis hypotesetest, der;

H0 : Forventningsverdien til observert energigrader er ikke ulik i utfallsrommene

H1 : Forventningsverdien til observert energigrader er ulik i utfallsrommene

Kruskal-Wallis rank sum test

data: *gd17* by *vindretn*

Kruskal-Wallis chi-squared = 84.7994, df = 3, p-value < 2.2e-16

H0 er forkastet da Kruskal-Wallis hypotesetest gir $p < 0,05$ hvilket indikerer at energigrader varierer med vindretning.

3.2 Totalpopulasjonens energiforbruk

3.2.1 Korrelasjon med meteorologiske variable

Responsvariabelen *totalsum* utgjør summen av alle 36 (se kapittel 2.5) energiforbruk pr. observasjonsrekke. *totalsum* er således en numerisk, kontinuerlig variabel.

Shapiro-Wilk test av fordelingen gir en p-verdi $\ll 0,05$ og dermed ingen støtte for å si at *totalsum* er normalfordelt. Spearman-korrelasjonen for det observerte energiforbruket *totalsum* og de meteorologiske variablene er vist under:

Tabell 3 Korrelasjonskoeffisienter energiforbruk (totalpopulasjon) og numeriske meteorologiske variable

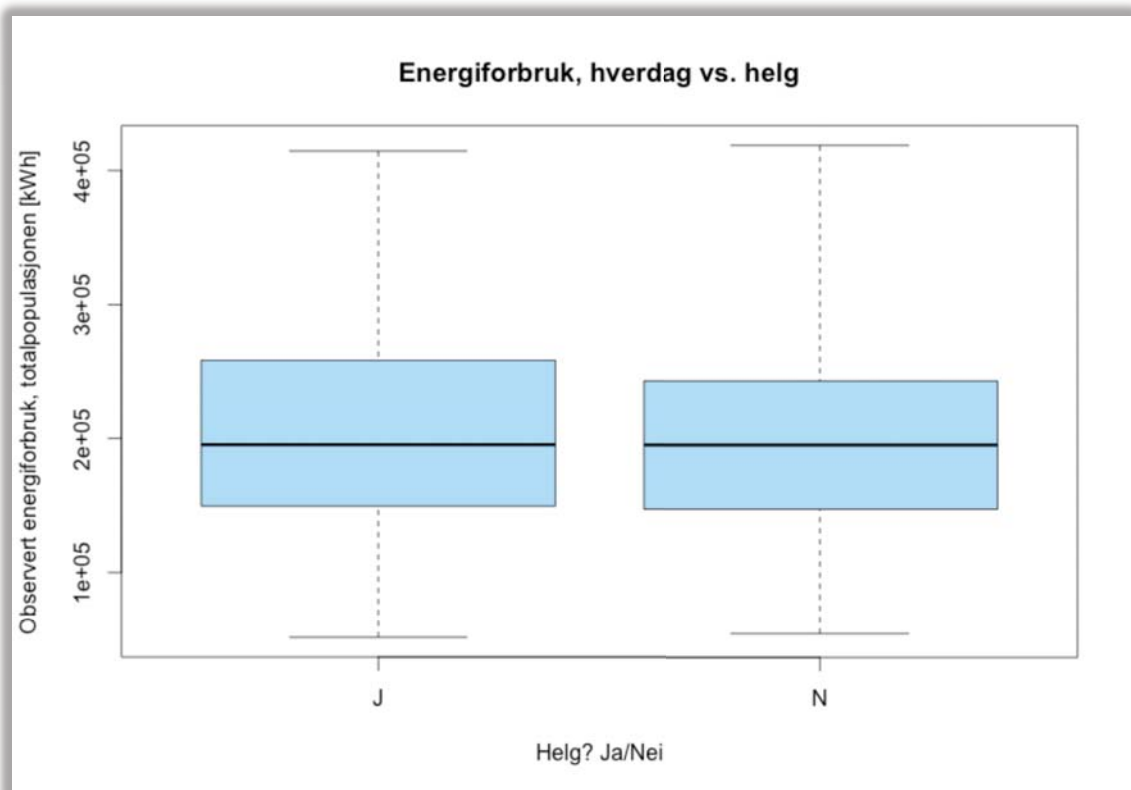
VARIABEL	<i>totalsum</i>
<i>totalsum</i>	1,0000
<i>middelvind</i>	-0,1715
<i>gd17</i>	0,9715
<i>middelsky</i>	0,0760
<i>sol</i>	-0,2879
<i>middelp</i>	0,0484
<i>nedbor</i>	-0,1636
<i>spesRH</i>	-0,8124
<i>middelRH</i>	0,2719

Kraftig korrelasjon observert mellom variablene *totalsum* og *gd17*.

Antatt internkorrelasjon (se kapittel 3.1.2) medfører at variablene *spesRH*, *middelsky* forkastes fra videre analyse (i henhold til kapittel 2.9.2).

3.2.2 Energiforbruk i helg kontra hverdag

I datasettet opptrer *helg* som en nominal variabel med utfallet "J" og "N". Fordelingen av *totalsum* på utfallene *helg* = J og *helg* = N er vist under:



Figur 4 Boksplot av totalpopulasjonens energiforbruk fordelt på hverdag/helg

Av figur 1 fremgår det visuelt liten eller ingen forskjell i energiforbruket avhengig av om det er helg eller ikke. Ettersom variabelen *helg* er nominal med 2 utfallsrom, og fordelingen av *totalsum* i hvert av utfallene ikke kan bevises normalfordelte i henhold til Shapiro-test, er det benyttet Wilcoxon hypotesetest, der;

H_0 : Forventningsverdien til observert energiforbruk er ikke ulik i utfallsrommene

H_1 : Forventningsverdien til observert energiforbruk er ulik i utfallsrommene

Wilcoxon rank sum test with continuity correction

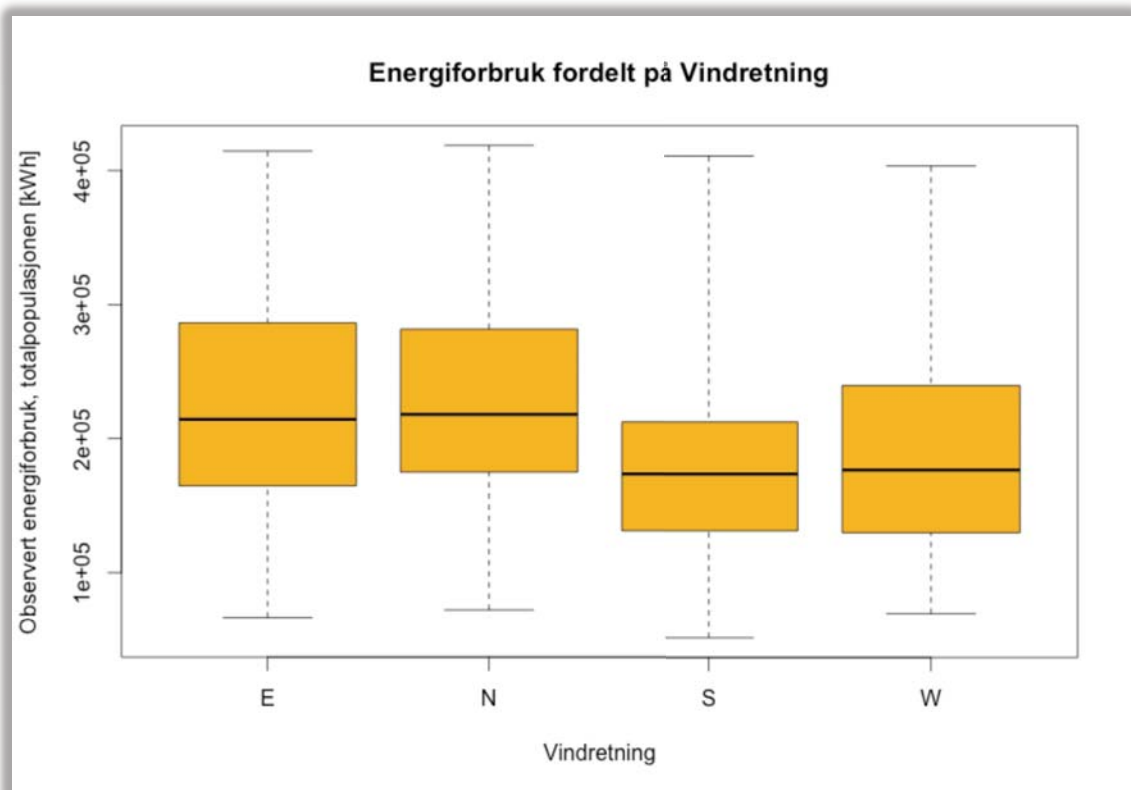
data: totalsum by helg

$W = 83499$, $p\text{-value} = 0.8328$

H_0 kan ikke motbevise etter Wilcoxon hypotesetest, da $p > 0,05$. Testen indikerer at energiforbruket ikke varierer signifikant mellom helg og hverdag.

3.2.3 Forbruk fordelt på vindretning

vindretn er en nominal variabel med utfallene "N", "E", "S", og "W". Fordelingen av *totalsum* på utfallene $vindretn = N$, $vindretn = E$, $vindretn = S$ og $vindretn = W$ er vist under:



Figur 5 Boksplot av totalpopulasjonens energiforbruk fordelt på vindretning

Av figur 5 fremgår det visuelt at det er forskjeller i totalsum mellom fordelingene. Forbruket ser ut til å være generelt høyere ved vind fra nord eller øst, og generelt lavere når vinden kommer fra sør eller vest.

Ettersom variabelen *vindretn* er nominal med 4 utfallsrom, og fordelingen av *totalsum* i hver av utfallene ikke kan bevises normalfordelte i henhold til Shapiro-Wilk test, er det benyttet Kruskal-Wallis hypotesetest, der:

H_0 : Forventningsverdien til observert energiforbruk er ikke ulik i utfallsrommene

H_1 : Forventningsverdien til observert energiforbruk er ulik i utfallsrommene

Kruskal-Wallis rank sum test

data: totalsum by vindretn

Kruskal-Wallis chi-squared = 89.1029, df = 3, p-value < 2.2e-16

H_0 er forkastet da Kruskal-Wallis test gir $p \ll 0.05$ hvilket indikerer at *totalsum* varierer med *vindretn*.

3.2.4 Multippel lineærregresjon av numeriske, uavhengige variable

Multippel lineærregresjon er benyttet for å evaluere hvorvidt forklaringsvariablene er signifikante. Der variablene har antatt internkorrelasjon (er avhengige variabler) er variabelen med den laveste korrelasjonen til *totalsum* utelatt fra regresjonen (se kapittel 3.2.1).

```
> mlr1=lm(totalsum~middelvind+gd17+sol+middelp+nedbor)
> summary(mlr1)
```

Call:

```
lm(formula = totalsum ~ middelvind + gd17 + sol + middelp + nedbor +
    middelRH)
```

Residuals:

Min	1Q	Median	3Q	Max
-62210	-11990	521	11199	60373

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-325543.61	75977.35	-4.285	2.38e-05	***
middelvind	2377.09	714.19	3.328	0.000968	***
gd17	11702.57	173.09	67.609	< 2e-16	***
sol	-1302.63	401.71	-3.243	0.001300	**
middelp	341.54	74.30	4.597	6.03e-06	***
nedbor	-116.12	175.02	-0.663	0.507460	
middelRH	53.33	123.10	0.433	0.665100	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16760 on 344 degrees of freedom
(550 observations deleted due to missingness)

Multiple R-squared: 0.939, Adjusted R-squared: 0.938

F-statistic: 882.8 on 6 and 344 DF, p-value: < 2.2e-16

Multippel lineærregresjon indikerer at variablene *nedbor*, og *middelRH* ikke er signifikante som forklaringsvariabler for næringspopulasjonens energiforbruk. 550 observasjonsrekker er slettet fra regresjonen grunnet manglende registrerte observasjoner av hhv. antall soltimer og nedbørsmengde.

Gjentatt multippel lineærregresjon der variablene *nedbor* og *middelRH* er utelatt, returnerer resultatet:

```
> mlr2=lm(totalsum~middelvind+gd17+sol+middelp)
> summary(mlr2)
```

```

Call:
lm(formula = totalsum ~ middelvind + gd17 + sol + middelp)

Residuals:
    Min       1Q   Median       3Q      Max
-69901 -10709   -173   11191   60776

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -284211.15    57232.38  -4.966 8.92e-07 ***
middelvind    2692.12     554.92    4.851 1.56e-06 ***
gd17         11584.80     127.31   90.995 < 2e-16 ***
sol          -842.15     179.80   -4.684 3.49e-06 ***
middelp       301.66      57.09    5.284 1.77e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17100 on 601 degrees of freedom
(295 observations deleted due to missingness)
Multiple R-squared:  0.9416, Adjusted R-squared:  0.9412
F-statistic: 2421 on 4 and 601 DF, p-value: < 2.2e-16

```

For hver av forklaringsvariablenes koeffisienter er $\text{Pr}(>|t|) \ll 0.05$, hvilket indikerer at alle forklaringsvariablene er statistisk signifikante. Regresjonen i sin helhet er returnert med $p \ll 0.05$, hvilket indikerer forklaringsvariablene med deres koeffisienter er statistisk signifikant. Forklart varians, R^2 er beregnet til 0,9416. Følgelig kan forbruket *totalsum* estimeres lineært med funksjonen:

Formel 5 Funksjon for totalpopulasjonens estimerte energiforbruk ved multippel lineærregresjon

$$Q_{\text{estimert}} = -284211,14 + 2692,12X_1 + 11584,8X_2 - 842,15X_3 + 301,66X_4$$

der

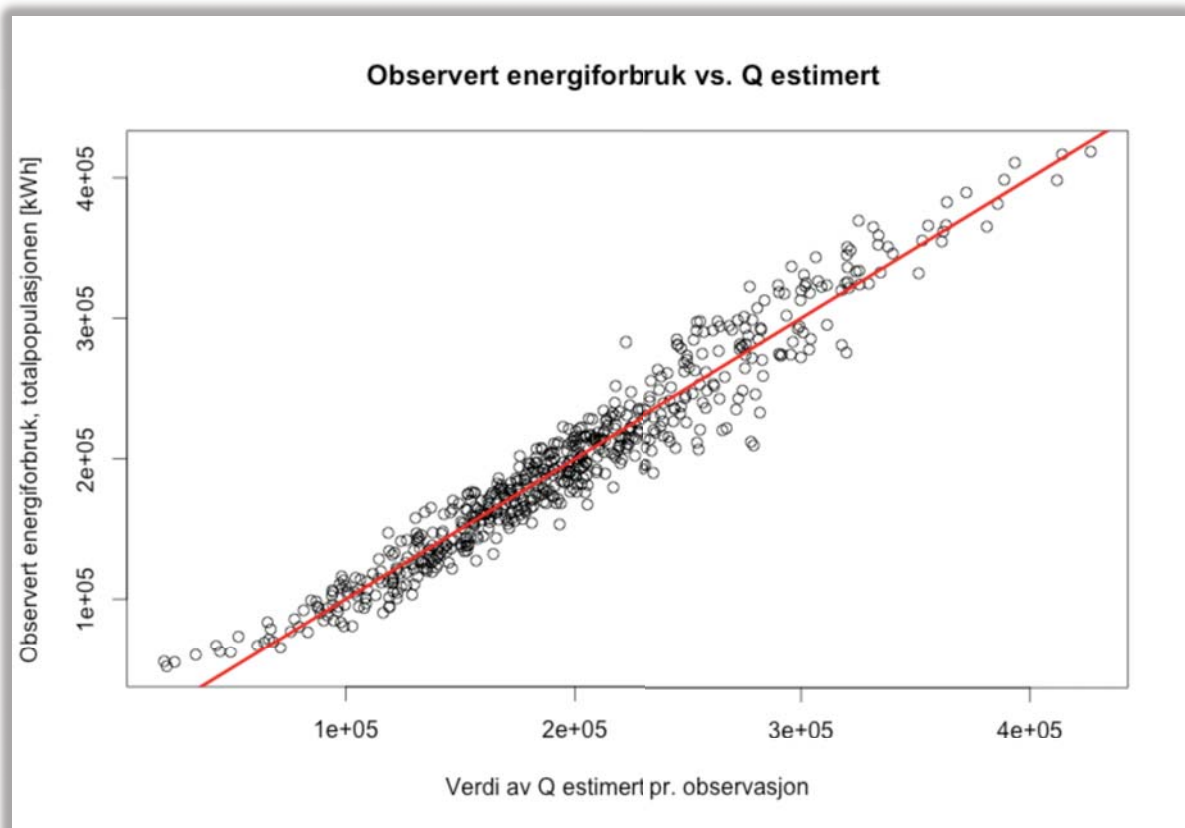
$$X_1 = \text{middelvind} = \text{midlere vindstyrke}$$

$$X_2 = \text{gd17} = \text{energigrader}$$

$$X_3 = \text{sol} = \text{antall soltimer}$$

$$X_4 = \text{middelp} = \text{midlere lufttrykk}$$

Q_{estimert} er beregnet for hver observasjonsrekke, og definert som en stokastisk variabel. Totalpopulasjonens observerte energiforbruk (y-aksen), plottet mot estimert energiforbruk ved multippel lineærregresjon (x-aksen) er vist i figuren under:



Figur 6 Totalpopulasjonens observerte energiforbruk plottet mot utregnet verdi av multippel lineærregresjon

Fra figur 6 fremgår det visuelt at totalpopulasjonens observerte energiforbruk er relativt lineært korrelert med estimert energiforbruk ved multippel lineærregresjon. For å illustrere den lineære sammenhengen er linja $Y = X$ markert med rød strek. Det observeres antydning til krumming av plottet ved lavt observert energiforbruk, som antyder forstyrrelser i lineærsammenheng ved lavt energiforbruk.

Spearman's rank correlation rho

```
data: totalsum and yfunk
S = 1144901, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.9691324
```

Spearman korrelasjonskoeffisient er kalkulert til 0,9691 med $p \ll 0,05$, hvilket indikerer kraftig korrelasjon, og statistisk signifikans.

3.2.5 Enkel lineærregresjon mot temperatur

Enkel lineærregresjon mellom observert energiforbruk og energigrader er benyttet for å estimere mengden energi som går til oppvarming av varmtvann, i henhold til antakelsen om konstant varmtvannsforbruk (se kapittel 1.2 og 2.9.5)

```
> mlr3=lm(totalsum~gd17)
> summary(mlr3)

Call:
lm(formula = totalsum ~ gd17)

Residuals:
    Min       1Q   Median       3Q      Max
-87710 -12631     510  12822  55208

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  19408.31    1621.81   11.97  <2e-16 ***
gd17         11917.54     97.79   121.87  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18290 on 899 degrees of freedom
Multiple R-squared:  0.9429, Adjusted R-squared:  0.9429
F-statistic: 1.485e+04 on 1 and 899 DF, p-value: < 2.2e-16
```

Regresjonen er signifikant beregnet med $p \ll 0,05$, og forklart varians, R^2 er beregnet til 0,9429.

Estimering av energiforbruket med energigrader som forklaringsvariabel er uttrykt ved funksjonen:

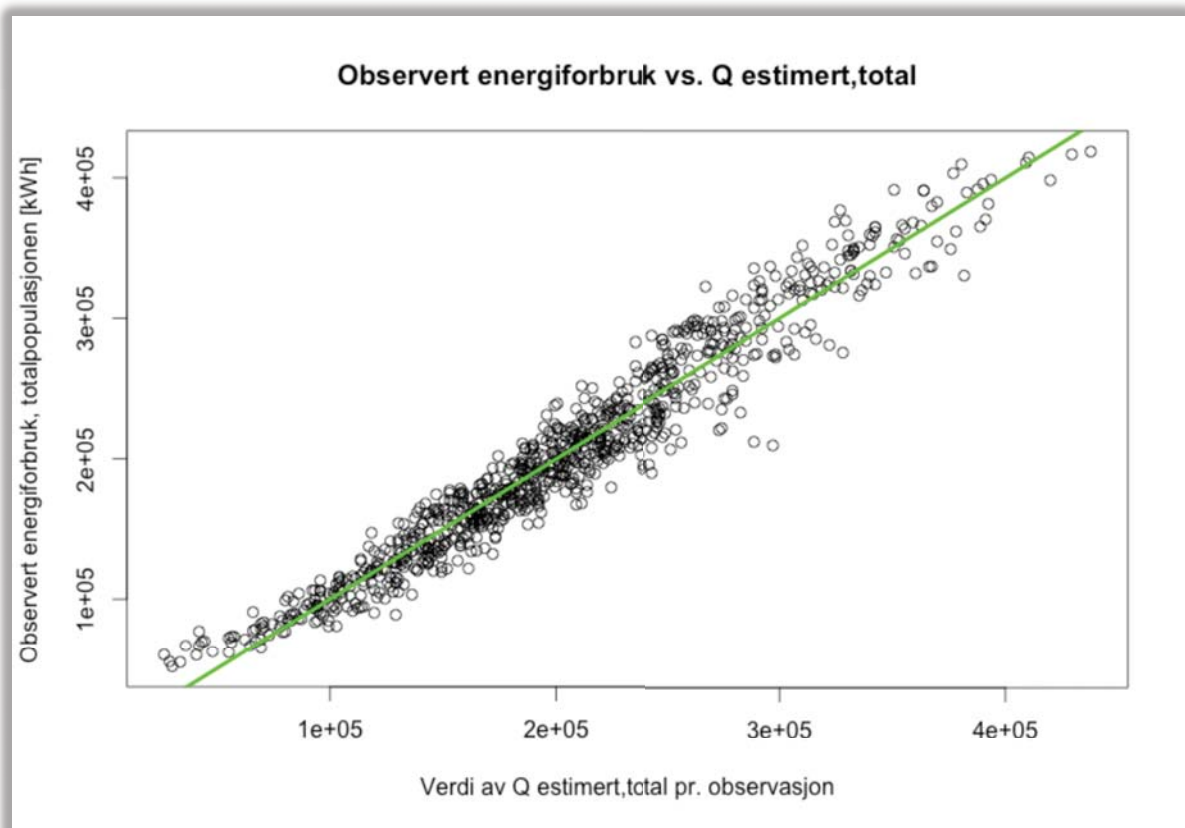
Formel 6 Funksjon for totalpopulasjonens estimerte energiforbruk ved enkel lineærregresjon

$$Q_{\text{estimert,total}} = 19408,31 + 11917,54 X_1$$

der

$$X_1 = gd17 = \text{energigrader}$$

Det observerte energiforbruket for totalpopulasjonen (y-aksen), plottet mot estimert energiforbruk ved enkel lineærregresjon (x-aksen) er vist i figuren under:



Figur 7 Totalpopulasjonens observerte energiforbruk, plottet mot utregnet verdi av enkel lineærregresjon

Av figur 7 fremgår det visuelt at observert energiforbruk er relativt lineært korrelert med estimert energiforbruk ved enkel lineærregresjon mot observerte energigrader. For å illustrere lineærsammenhengen er linja $Y = X$ er markert med grønn strek. Som i figur 6 observeres det antydning til krumming av plottet ved lavt energiforbruk.

Testet for korrelasjon:

Spearman's rank correlation rho

data: totalsum and alt.yfunk

$S = 3478164$, p-value $< 2.2e-16$

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

0.9714683

Spearman's korrelasjonskoeffisient er beregnet til 0,9714.

3.2.6 Estimering av varmtvannsandel, totalpopulasjon

Funksjonsuttrykket $Q_{\text{estimert, total}}$ (se formel 6) medfører at mengden energi til oppvarming av varmtvann pr. observasjon (pr. døgn) er 19,41 MWh i henhold til antakelsen gjort i formel 4 (se kapittel 2.9.5).

Følgelig er estimert årlig energiforbruk til oppvarming av varmtvann for hele populasjonen beregnet til:

Formel 7 Estimert årlig energimengde til oppvarming av varmtvann for totalpopulasjonen

$$Q_{\text{estimert, varmtvann, total}} = 19408,31 \text{ kWh/dag} * 365 \text{ dager/år} = 7,08 \text{ GWh/år}$$

Estimert energiforbruk til varmtvann utgjør en årlig andel av observert energimengde levert gjennom fjernvarme på:

Tabell 4 Estimert årlig energiandel til varmtvann, totalpopulasjon

	2008 (366 dager)	2009	2010	2011	Totalt
Fjernvarme totalt levert [GWh]	46,34	52,90	63,68	51,00	213,92
Estimert energi til varmtvann [GWh]	7,10	7,08	7,08	7,08	28,34
Prosentvis andel	15,3 %	13,4 %	11,2 %	13,0 %	13,3 %

3.3 Boligpopulasjonens energiforbruk

Resultatene fra analysene gjort av populasjonsgruppen ”bolig” følger i dette kapittelet. For enkelhets skyld er koding og resultater (unntatt plot) generert i RStudio presentert i vedlegg 6.

3.3.1 Korrelasjon med meteorologiske variable

Responsvariabelen y_{sum} er det summerte forbruket pr. observasjon for alle 18 boligkunder. Variabelen er numerisk, kontinuerlig og Shapiro-Wilks test for normalfordeling indikerer at det summerte forbruket ikke er normalfordelt da kalkulert p-verdi $\ll 0,05$.

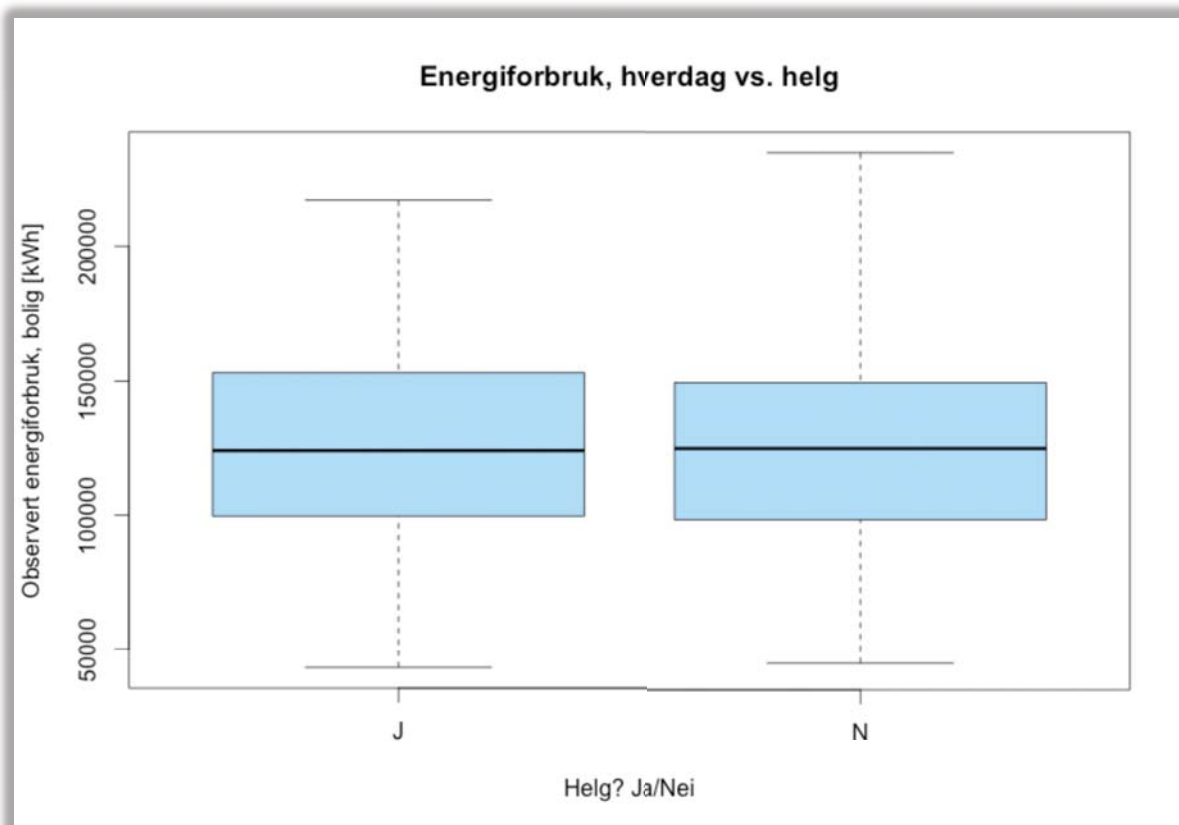
Spearmankorrelasjon til det observerte energiforbruket for boligpopulasjonen og de numeriske forklaringsvariablene er vist under:

Tabell 5 Korrelasjonskoeffisienter for energiforbruk (boligpopulasjon) og numeriske meteorologiske variable

VARIABEL	<i>y</i> sum
<i>y</i> sum	1,0000
<i>middelvind</i>	-0,1711
<i>gd17</i>	0,9833
<i>middelsky</i>	0,0666
<i>sol</i>	-0,2830
<i>middelp</i>	0,0348
<i>nedbor</i>	-0,1398
<i>spesRH</i>	-0,8284
<i>middelRH</i>	0,2606

Variablene *middelsky*, og *spesRH* er forkastet fra videre analyse (i henhold til kapittel 2.9.2).

3.3.2 Energiforbruk hverdag/helg

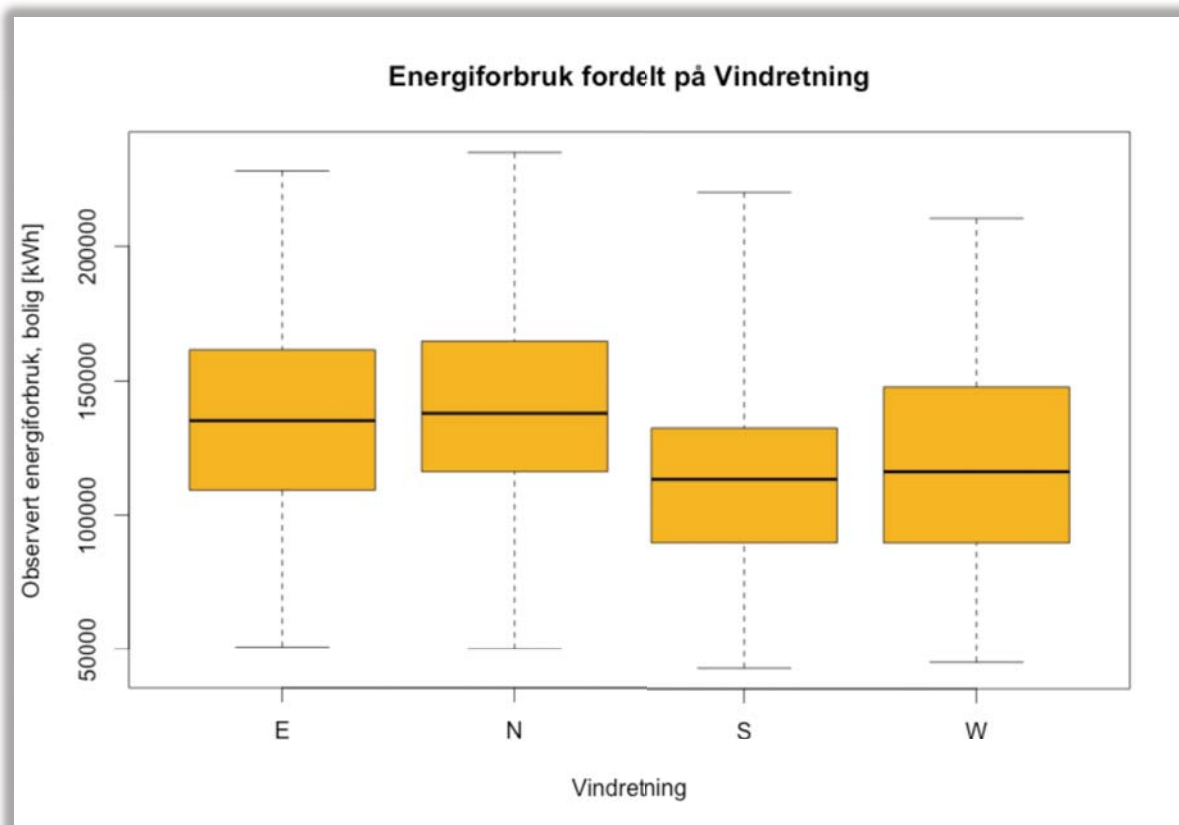


Figur 8 Bokplot av boligpopulasjonens energiforbruk fordelt på hverdag/helg

Av figur 8 fremgår det visuelt liten forskjell i forventningsverdien til boligpopulasjonens energiforbruk mellom hverdag og helg.

Wilcoxon hypotesetest indikerer ingen forskjell i forventningsverdien til boligpopulasjonens energiforbruk i fordelingene hverdag og helg, da kalkulert p-verdi = $0,7978 > 0,05$.

3.3.3 Energiforbruk fordelt på vindretning



Figur 9 Boksplot av boligpopulasjonens energiforbruk fordelt på vindretning

Av figur 9 fremgår det visuelt at det er forskjeller i forventningsverdien til boligpopulasjonens energiforbruk mellom de ulike vindretningene. Kruskal-Wallis hypotesetest indikerer at forbruket varierer med vindretning da kalkulert p-verdi $<< 0,05$.

3.3.4 Multipel lineærregresjon av numeriske, uavhengige variable

Multipel lineærregresjon av alle ikke forkastede variable mot energiforbruket *ysum* resulterer i at forklaringsvariablene *nedbor* og *middelRH* er forkastet som ikke statistisk signifikante.

Gjentatt multipel lineærregresjon resulterer i statistisk signifikante koeffisienter.

Regresjonen i sin helhet er kalkulert med $p << 0,05$ og er således statistisk signifikant.

Estimert energiforbruk som funksjon av flere variable kan uttrykkes:

Formel 8 Funksjon for boligpopulasjonens estimerte energiforbruk ved multippel lineærregresjon

$$Q_{estimert} = -38835,11 + 1068,94X_1 + 5954,65X_2 - 327,19X_3 + 71,59X_4$$

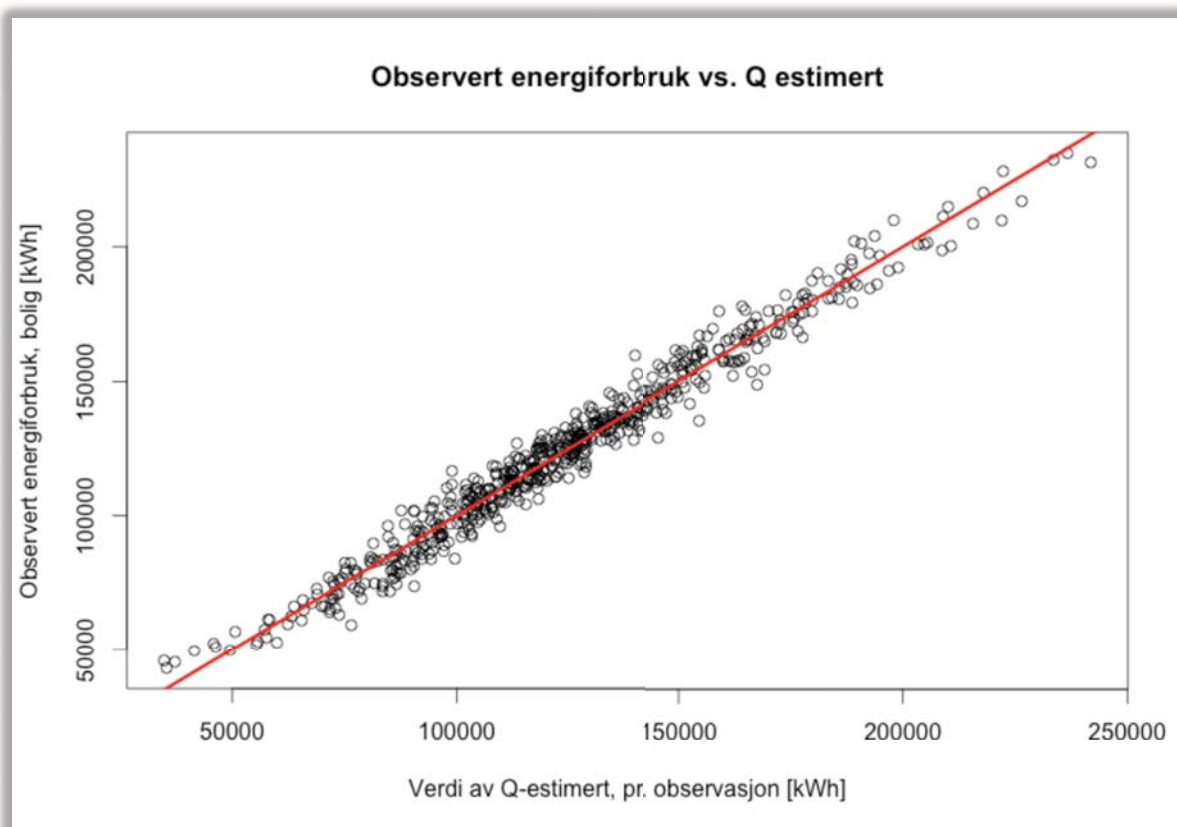
$X_1 = \text{middelvind}$

$X_2 = \text{gd17}$

$X_3 = \text{sol}$

$X_4 = \text{middelp}$

Boligpopulasjonens observerte energiforbruk (y-aksen), plottet mot estimert energiforbruk ved multippel lineærregresjon (x-aksen), er vist i figuren under:



Figur 10 Boligpopulasjonens observerte energiforbruk, plottet mot utregnet verdi av multippel lineærregresjon

Fra figur 10 fremgår det visuelt at det observerte energiforbruket for boligpopulasjonen har sterk lineær sammenheng med estimert energiforbruk ved multippel lineærregresjon. Linja $Y = X$ er markert med rød strek. Spearmans korrelasjonskoeffisient er kalkulert til $\rho = 0,9846$

med $p \ll 0,05$ hvilket indikerer kraftig korrelasjon og statistisk signifikans. Forklart varians, R^2 er beregnet til 0,9712.

3.3.5 Enkel lineærregresjon mot temperatur

Enkel lineærregresjon av boligpopulasjonens observerte energiforbruk, mot observerte energigrader $gd17$, gir statistisk signifikante koeffisienter ($\Pr(>|t| \ll 0,05)$), og er kalkulert med $p \ll 0,05$ (Se vedlegg 6). Forklart varians, R^2 for enkel lineærregresjon er beregnet til 0,9678.

Det estimerte energiforbruket for boligpopulasjonen er uttrykt lineært ved funksjonen:

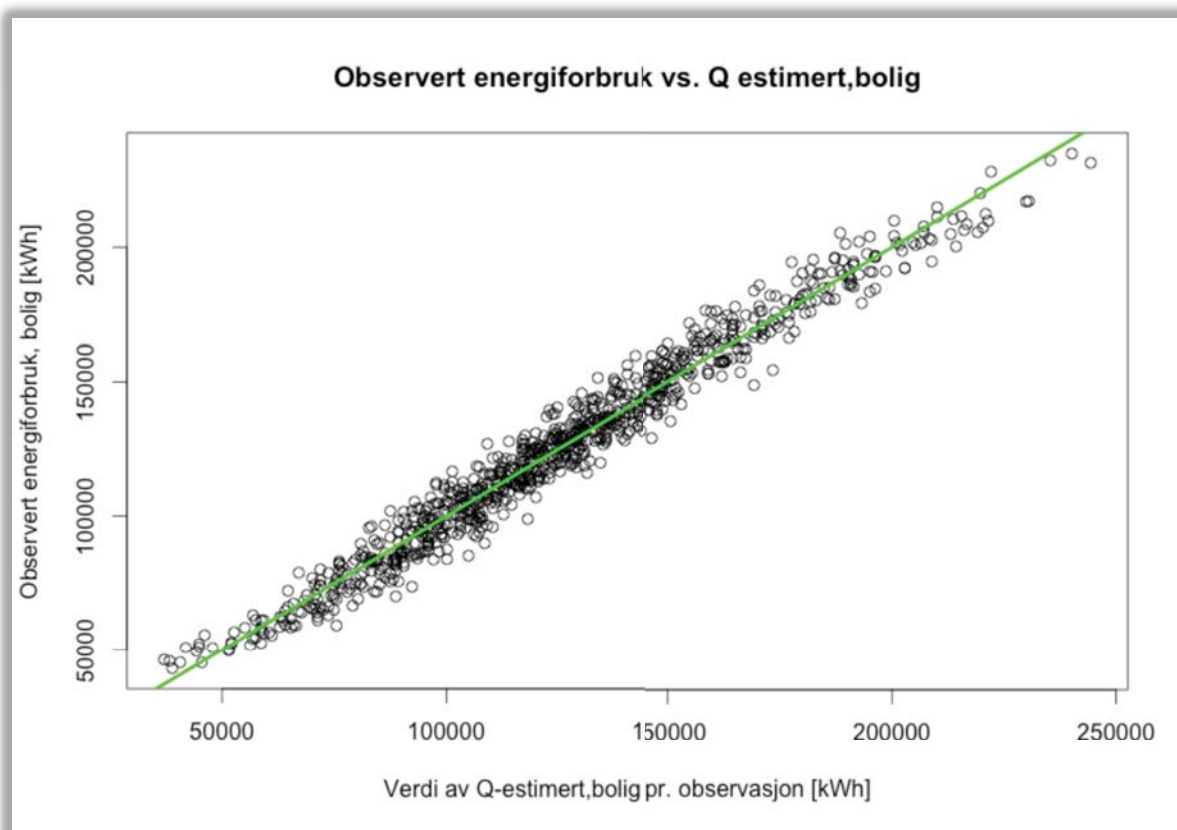
Formel 9 Funksjon for boligpopulasjonens estimerte energiforbruk ved enkel lineærregresjon

$$Q_{\text{estimert,bolig}} = 33447,16 + 6008,14 X_1$$

der

$$X_1 = gd17 = \text{energigrader}$$

Det observerte energiforbruket for boligpopulasjonen, plottet estimert energiforbruk ved enkel lineærregresjon er vist i figuren under:



Figur 11 Boligpopulasjonens observerte energiforbruk, plottet mot utregnet verdi av enkel lineærregresjon

Fra figur 11 fremgår det visuelt at det er sterk lineær sammenheng mellom boligpopulasjonens observerte energiforbruk, og det estimerte energiforbruket ved enkel lineærregresjon mot observerte energigrader. Spearmans korrelasjonskoeffisient er kalkulert til $\rho = 0,9833$, med $p \ll 0,05$, hvilket indikerer kraftig korrelasjon og statistisk signifikans.

3.3.6 Estimering av varmtvannsandel, boligpopulasjon

Estimert årlig energiforbruk til varmtvann er beregnet på bakgrunn av formel 9 (se kapittel 3.3.5) der energimengde til oppvarming av varmtvann pr. observasjon (pr. døgn) er estimert til 33,45 MWh:

Formel 10 Estimert årlig energimengde til oppvarming av varmtvann for boligpopulasjonen

$$Q_{\text{estimert, varmtvann, bolig}} = 33447,16 \text{ kWh/dag} * 365 \text{ dager/år} = 12,21 \text{ GWh/år}$$

Estimert energiforbruk til varmtvann utgjør en årlig andel av observert energiforbruk levert gjennom fjernvarme på:

Tabell 6 Estimert årlig energiandel til varmtvann, boligpopulasjon

	2008 (366 dager)	2009	2010	2011	Totalt
Fjernvarme totalt levert [GWh]	32,24	34,27	38,80	32,26	136,57
Estimert energi til varmtvann [GWh]	12,24	12,21	12,21	12,21	48,87
Prosentvis andel	37,0 %	35,6 %	31,5 %	37,8 %	35,5 %

3.4 Næringskundernes energiforbruk

Resultatene fra analysene gjort av populasjonsgruppen "næring" følger i dette kapittelet. For enkelhets skyld er koding og resultater (unntatt plot) generert i RStudio presentert i vedlegg 7.

3.4.1 Korrelasjon med meteorologiske variable

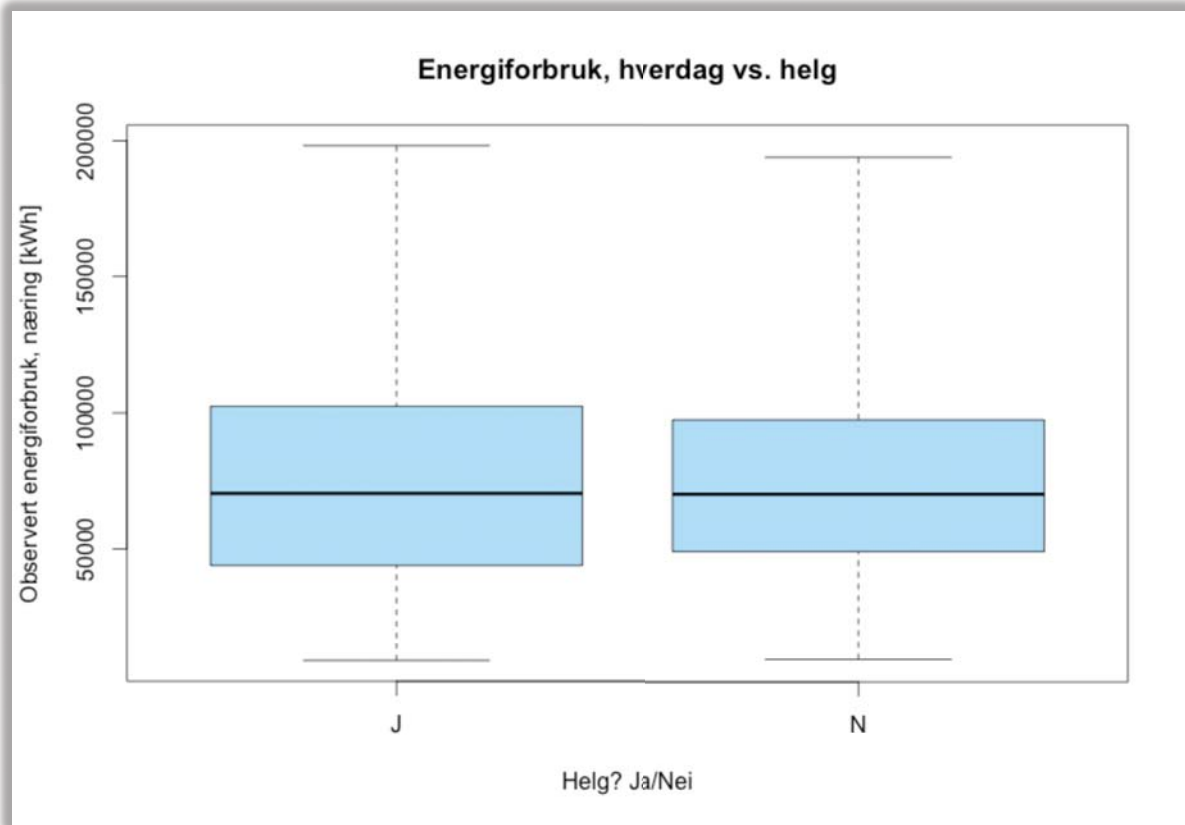
Responsvariabelen 'zsum' er det summerte forbruket pr. observasjon for alle 18 næringskunder. Variabelen er numerisk, kontinuerlig, og Shapiro-Wilks test for normalfordeling indikerer at det summerte forbruket ikke er normalfordelt da kalkulert p-verdi $< 0,05$. Spearmans korrelasjonskoeffisient mellom respons- og forklaringsvariablene er vist under:

Tabell 7 Korrelasjonskoeffisienter for energiforbruk (næringspopulasjon) og numeriske meteorologiske variable

VARIABEL	<i>zsum</i>
<i>zsum</i>	1,0000
<i>middelvind</i>	-0,1651
<i>gd17</i>	0,9262
<i>middelsky</i>	0,0839
<i>Sol</i>	-0,3132
<i>middelp</i>	0,0586
<i>nedbor</i>	-0,1699
<i>spesRH</i>	-0,7679
<i>middelRH</i>	0,2768

Variablene *middelsky*, og *spesRH* er forkastet fra videre analyse (i henhold til kapittel 2.9.2).

3.4.2 Energiforbruk hverdag/helg

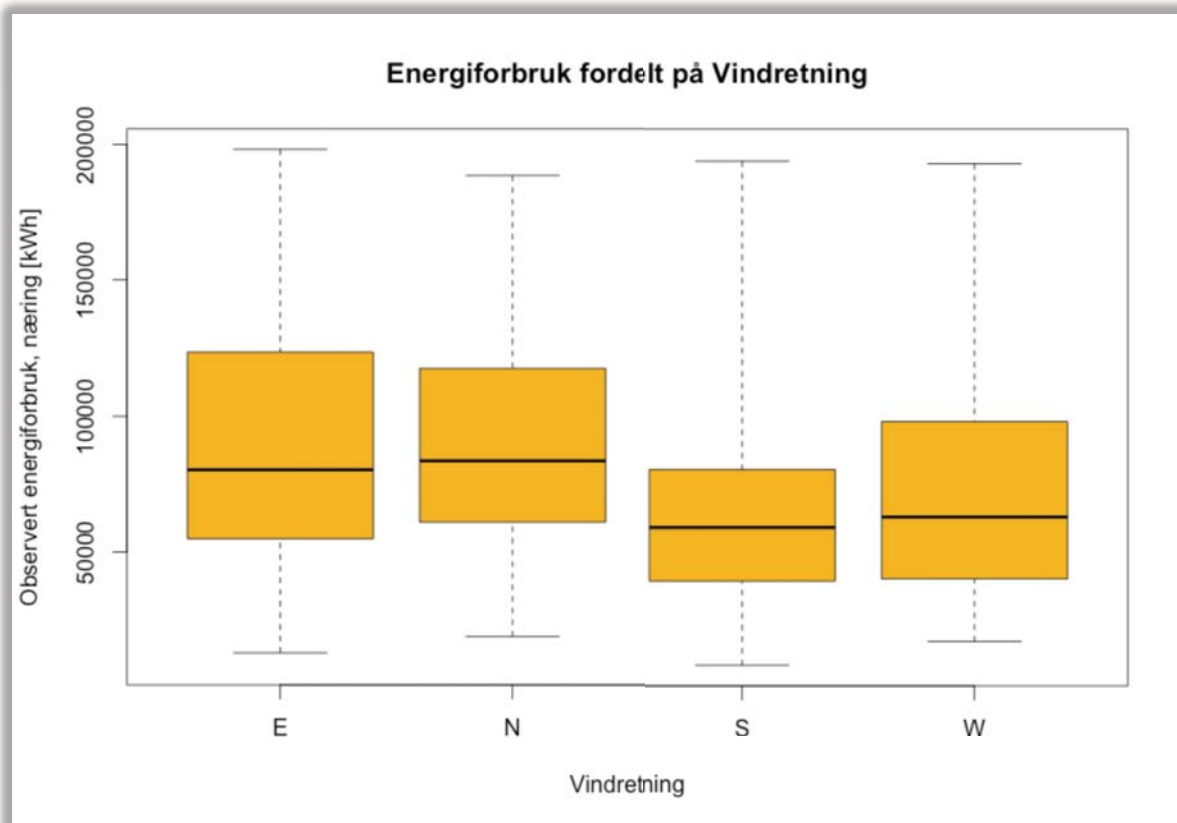


Figur 12 Boksplot av næringspopulasjonens energiforbruk fordelt på hverdag og helg

Av figur 12 fremgår det visuelt liten forskjell i forventningsverdien til boligpopulasjonens energiforbruk mellom hverdag og helg.

Wilcoxon hypotesetest indikerer ingen forskjell i forventningsverdien til boligpopulasjonens energiforbruk i fordelingene hverdag og helg, da kalkulert p-verdi = $0,7978 > 0,05$.

3.4.3 Energiforbruk fordelt på vindretning



Figur 13 Boksplot av næringspopulasjonens energiforbruk fordelt på vindretning

Av figur 9 fremgår det visuelt at det er forskjeller i forventningsverdien til boligpopulasjonens energiforbruk mellom de ulike vindretningene. Kruskal-Wallis hypotesetest indikerer at forbruket varierer med vindretning da kalkulert p-verdi $<< 0,05$.

3.4.4 Multippel lineærregresjon av numeriske, uavhengige variable

Multippel lineærregresjon av alle ikke forkastede variable mot energiforbruket y_{sum} resulterer i at forklaringsvariablene *nedbor* og *middelRH* forkastes som ”ikke statistisk signifikante”.

Gjentatt multippel lineærregresjon resulterer i statistisk signifikante koeffisienter.

Regresjonen i sin helhet er kalkulert med $p << 0,05$ og er således statistisk signifikant.

Estimert energiforbruk som funksjon av flere variable kan uttrykkes:

Formel 11 Funksjon for næringspopulasjonens estimerte energiforbruk ved multippel lineærregresjon

$$Q_{estimert} = -245376,04 + 1623,18X_1 + 5630,15X_2 - 514,96X_3 + 230,07X_4$$

der

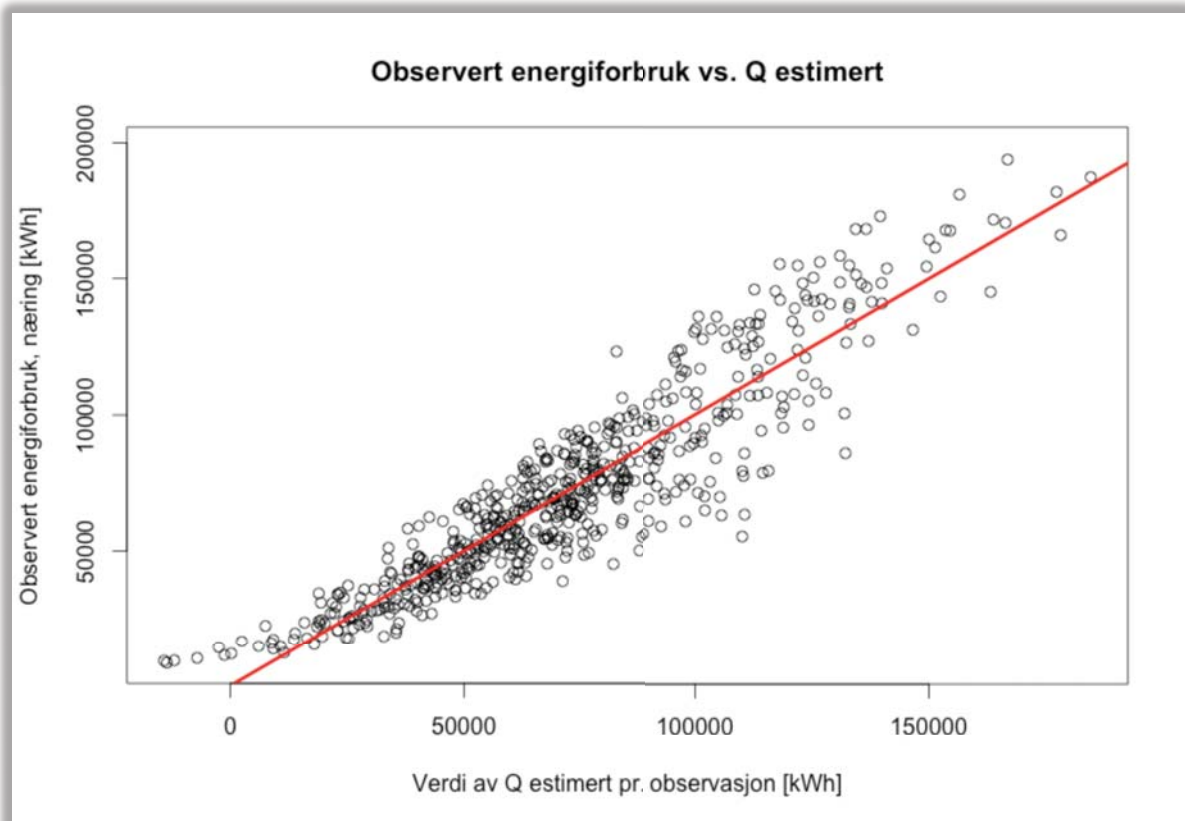
$$X_1 = \text{middelvind}$$

$X_2 = gd17$

$X_3 = sol$

$X_4 = middelp$

Næringspopulasjonens observerte energiforbruk (y-aksen), plottet mot estimert energiforbruk ved multippel lineærregresjon (x-aksen), er vist i figuren under:



Figur 14 Næringspopulasjonens observerte energiforbruk, plottet mot utregnet verdi av multippel lineærregresjon

Spearman's korrelasjonskoeffisient er kalkulert til $\rho = 0,9142$, med $p \ll 0,05$, hvilket indikerer relativt sterk korrelasjon og statistisk signifikans. Til tross for relativt sterk korrelasjon, er det observert relativt stor spredning i utfallene, i forhold til hhv. boligpopulasjonen (se figur 10) og totalpopulasjonen (se figur 6). Forklart varians, R^2 er beregnet til 0,8469.

Det observeres en noe divergerende sammenheng mellom energiforbruk og økt temperaturdifferanse, i tillegg til en relativt markant krumming av plottene ved de laveste observerte energiforbrukene, hvilket antyder ikke-lineær sammenheng med én eller flere av forklaringsvariablene.

3.4.5 Enkel lineærregresjon mot temperatur

Enkel lineærregresjon av det observerte energiforbruket mot observerte energigrader $gd17$, gir statistisk signifikante koeffisienter ($\Pr(>|t|) < 0,05$), og er kalkulert med $p < 0,05$ (Se vedlegg 7). Forklart varians, R^2 for enkel lineærregresjon er beregnet til 0,8588.

Det estimerte energiforbruket for næringspopulasjonen er uttrykt lineært ved funksjonen:

Formel 12 Funksjon for næringspopulasjonens estimerte energiforbruk ved enkel lineærregresjon

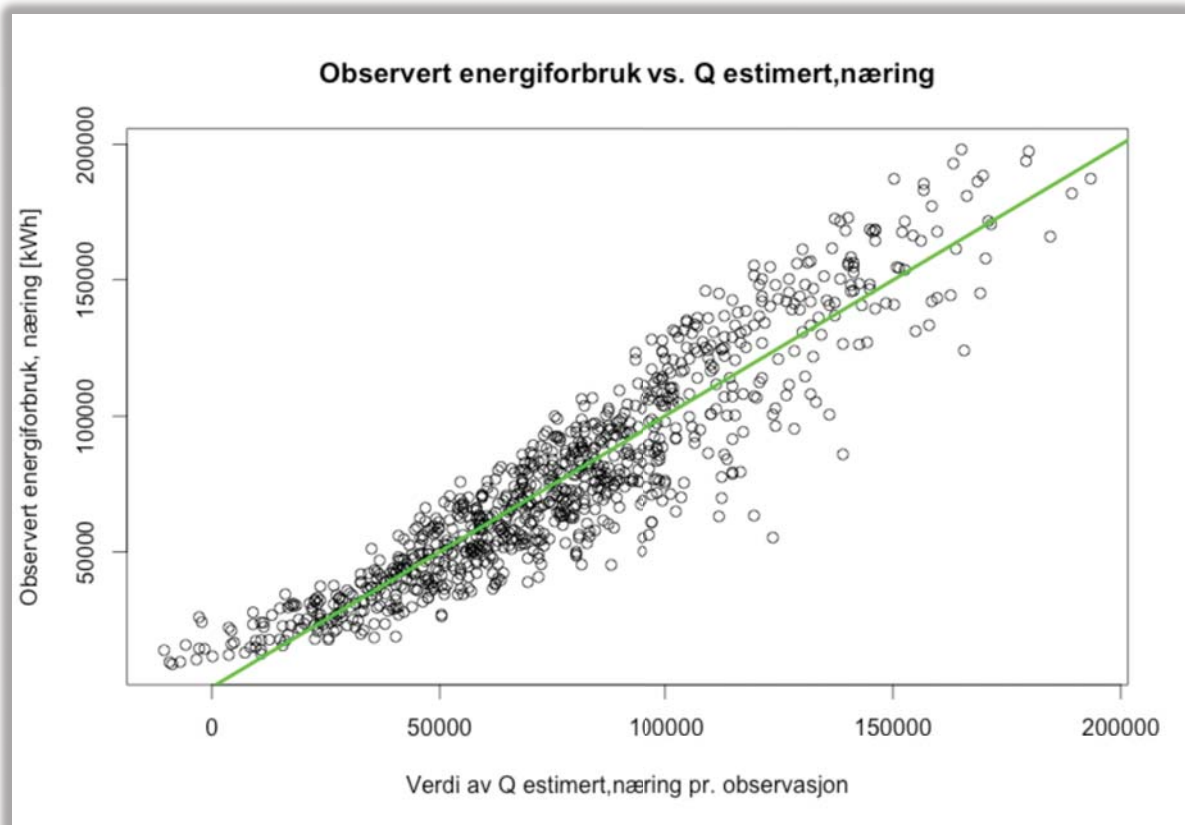
$$Q_{\text{estimert,næring}} = -14038,85 + 5909,4 X_1$$

der

$$X_1 = gd17 = \text{energigrader}$$

Det observeres negativt konstantledd for funksjonen $Q_{\text{estimert,næring}}$. Dette er ikke forenelig med antakelsen om varmtvannsandel ved lineærsammenheng (se kapittel 2.9.5).

Det observerte energiforbruket for næringspopulasjonen, plottet estimert energiforbruk ved enkel lineærregresjon er vist i figuren under:



Figur 15 Næringspopulasjonens observerte energiforbruk, plottet mot utregnet verdi av enkel lineærregresjon

Spearman's korrelasjonskoeffisient er kalkulert til $\rho = 0,9262$, med $p \ll 0,05$, hvilket indikerer relativt sterk korrelasjon og statistisk signifikans.

Til tross for relativt sterk korrelasjon, er det observert stor spredning i utfallene og det observeres en noe divergerende sammenheng mellom energiforbruk og økte energigrader. I tillegg observeres det markant krumming av plottene ved de laveste observerte energiforbrukene og antydning til krumming ved de høyeste observerte energiforbrukene. Dette bygger opp under antakelsen om ikke-lineær sammenheng mellom næringspopulasjonens observerte energiforbruk og observerte energigrader.

3.4.6 Ikke-lineær regresjon mot temperatur

I kapittel 3.4.5 ble det vist at varmtvannsandelen av energiforbruket til næringspopulasjonen ikke kan kalkuleres ved lineærregresjon etter antakelsen gjort i kapittel 2.9.7. Dersom en ser bort ifra fysikken, kan det observerte energiforbruket for næringspopulasjonen forklares matematisk (bedre kurvetilpassing) med ikke-lineær regresjon. Det viste seg at den mest statistisk signifikante beskrivelsen er en kubisk funksjon der 1.gradsleddet ikke inngår:

Call:

```
lm(formula = zsum ~ I(gd17^2) + I(gd17^3))
```

Residuals:

Min	1Q	Median	3Q	Max
-71772	-8462	871	10132	35905

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	15274.8016	1284.9226	11.89	<2e-16 ***
I(gd17^2)	332.5384	11.7391	28.33	<2e-16 ***
I(gd17^3)	-5.4460	0.3919	-13.89	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14480 on 898 degrees of freedom

Multiple R-squared: 0.8676, Adjusted R-squared: 0.8673

F-statistic: 2943 on 2 and 898 DF, p-value: < 2.2e-16

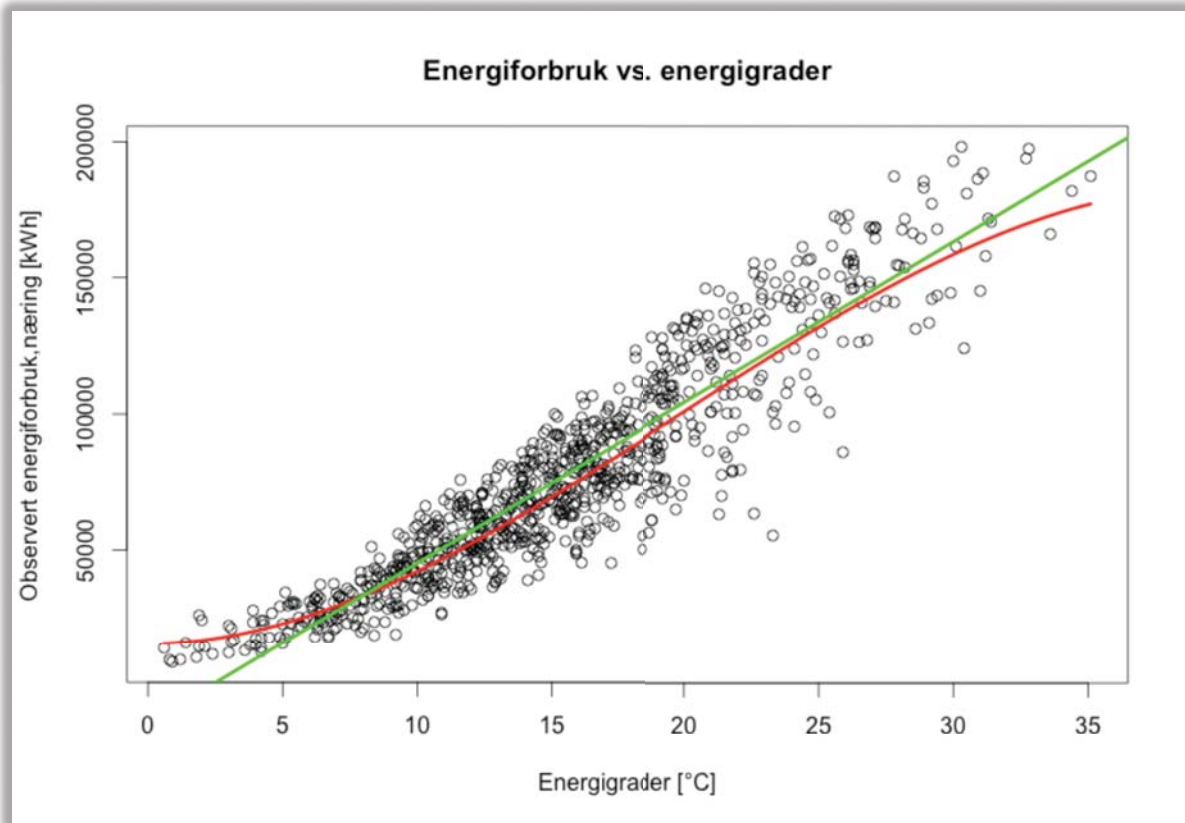
Regresjonen er statistisk signifikant da $p \ll 0,05$.

Ikke-lineær regresjon medfører at næringspopulasjonens observerte energiforbruk kan uttrykkes som funksjon av temperaturdifferanse ved funksjonen:

Formel 13 Funksjonsuttrykk for ikke-lineær sammenheng

$$Q(x)_{ikke-lineær, næring} = 15274,8 + 332,5x^2 - 5,4x^3$$

Sammenhengen mellom næringskundernes observerte energiforbruk og observerte energigrader er vist i figuren under:



Figur 16 Næringspopulasjonens observerte energiforbruk, plottet mot observerte energigrader

I figur 16 er funksjonen $Q_{\text{ikke-lineær, næring}}$ (se formel 13) markert med rød kurve. Det observeres en visuelt bedre tilpassing for $Q_{\text{ikke-lineær, næring}}$ enn ved lineærregresjon (se formel 12) som er markert i figur 16 med grønn linje.

Spearman's korrelasjonskoeffisient for ikke-lineær regresjon er kalkulert til 0,9262, med $p \ll 0,05$, hvilket er identisk med korrelasjonen kalkulert for enkel lineærregresjon (se kapittel 3.4.5). Forklart varians, $R^2 = 0,8676$, hvilket er høyere enn for lineærregresjon (se kapittel 3.4.5). Matematisk sett forklarer således den ikke-lineære tilnærmingen energiforbruket bedre enn lineærtlnærmingen.

3.4.7 Estimering av varmtvannsandel, næringspopulasjon

Dersom en setter temperaturdifferansen lik 0, vil konstantleddet kunne representere det temperaturuavhengige energiforbruket.

Ved å integrere funksjonsuttrykket for $Q_{\text{ikke-lineær,næring}}$ fra 0 energigrader til høyeste verdi av observerte energigrader, ble det beregnet et teoretisk fjernvarmeforbruk. Varmtvannsandelen for det totale energiforbruket for næringspopulasjonen ble så beregnet som forholdet mellom det teoretiske fjernvarmeforbruket, og konstantleddet i $Q_{\text{ikke-lineær,næring}}$ integrert over de samme grenseverdier:

Formel 14 Estimering av andel energi til oppvarming av varmtvann for næringspopulasjonen

$$Andel_{\text{varmtvann}} = \frac{\int_{x=\min}^{x=\max} 15247,8 \, dx}{\int_{x=\min}^{x=\max} Q_{\text{ikke-lineær,næring}} \, dx}$$

Grenseverdier for integral er:

Nedre: Energigrader = 0

Øvre: Maksimalt observerte energigrader:

> max(gd17)

[1] 35.1

$$\int_0^{35,1} 152478,8016 \, dx = 536,1 \, [MWh]$$

$$\int_0^{35,1} Q(x)_{\text{ikke-lineær,næring}} \, dx = \left[15274,8016x + \frac{332,5384}{3}x^3 - \frac{5,4460}{4}x^4 \right]_0^{35,1}$$

$$\int_0^{35,1} Q(x)_{\text{ikke-lineær,næring}} \, dx = 3263,0 \, [MWh]$$

$$Andel_{\text{varmtvann}} = \frac{536,1}{3263,0} = 16,4 \, \%$$

Det observeres at den matematiske estimeringen av varmtvannsandelen er i samme størrelsesorden som tidligere estimert varmtvannsandel for næringskunder (se tabell 1).

Andelsberegningen over er gyldig for energiforbruk observert i næringspopulasjonen fra og med 01.01.2008 til og med 31.12.2011.

Ved kurvedrøfting er det i tillegg observert følgende tilfelle;

Formel 15 Generalisert form av funksjonen "Q ikke-lineær,næring" (formel 13)

$$Q(x)_{ikke-lineær,næring} = \alpha + \beta x^2 - \gamma x^3$$

Setter 1. derivert = 0:

$$Q'(x)_{ikke-lineær,næring} = 2\beta x - 3\gamma x^2 = 0$$

$$x(2\beta - 3\gamma x) = 0 \rightarrow x = 0 \quad (1. \text{ ekstremalpunkt ved energigrader} = 0 \text{ } ^\circ\text{C})$$

$$2\beta - 3\gamma x = 0$$

$$\frac{2\beta}{3\gamma} = x$$

Setter inn verdier for β og γ (se formel 13):

$$x = \frac{2 * 332,5}{3 * 5,4} = 41,0$$

Med andre ord medfører utregningen over at maksimalt energiuttak for næringspopulasjonen skjer ved energigrader = 41 °C, hvilket tilsvarer en utetemperatur på -24, altså 4 °C lavere enn dimensjonerende utetemperatur for Oslo (-20 °C) (Setsaas 2009).

3.5 Alternativ estimering av total varmtvannsandel

Basert på det faktum at lineærtilnærmingen til varmtvannsandelen for næringskundene resulterer i negativ mengde varmtvann, er følgelig lineærtilnærmingen for den totale populasjonen ugyldig. Dette kan bevises ettersom:

Formel 16 Sammenheng mellom Q estimert ved lineærregresjon for de ulike populasjonene

$$Q_{estimert,total} = Q_{estimert,bolig} + Q_{estimert,næring} \quad (\text{formel 6, formel 9 og formel 12})$$

Dersom en videre antar at varmtvannsandelen for næringskundene i årene 2008 – 2011 utgjør 16.4% (se kapittel 3.4.7) vil det tilsvare 12,54 GWh av det observerte energiforbruket i hele perioden. Summen av det estimerte energiforbruket til varmtvann for boligpopulasjonen i samme tidsperiode utgjør 48,87 GWh.

Den estimerte andelen energi til varmtvann av den observerte totalmengden energi levert, er presentert i tabellen under:

Tabell 8 Estimert totalandel energi til oppvarming av varmtvann 2008 - 2011

	2008 - 2011
Observert totalt levert [kWh] bolig	137,57 GWh
Observert totalt levert [kWh] næring	76,35 GWh
SUM totalt levert	213,92 GWh
Estimert energi til varmtvann, bolig	48,87 GWh
Estimert energi til varmtvann, næring	12,54 GWh
SUM totalt varmtvann	61,41 GWh
Estimert andel energi til varmtvann totalpopulasjonen	28,7 %

3.6 Estimering av årlig varmtvannsandel, næringspopulasjon

Ved antatt konstant varmtvannsforbruk i perioden 2008 – 2011, er det beregnet et døgnforbruk til oppvarming av varmtvann for næringspopulasjonen på 8,6 MWh. Beregnede årlige andeler av observert levert energi gjennom fjernvarme som dekker oppvarming av varmtvann for næringspopulasjonen er presentert i tabellen under:

Tabell 9 Estimert årlig andel energi til oppvarming av varmtvann, næringspopulasjon

	2008 (366 dager)	2009	2010	2011	Totalt
Fjernvarme totalt levert [GWh]	14,10	18,63	24,88	18,74	76,35
Estimert energi til varmtvann [GWh]	3,14	3,13	3,13	3,13	12,53
Prosentvis andel	22,3 %	16,8 %	12,6 %	16,7 %	16,4 %

3.7 Temperatursensitivitet

Det er vist i kapittel 3.4.6 at næringspopulasjonens energiforbruk kan beskrives mer treffende med ikke-lineær regresjon kontra lineærregresjon. Boligpopulasjonens energiforbruk kan derimot beskrives lineært. Dette medfører at de to populasjonene har ulik temperatursensitivitet innenfor utfallsrommet av de observerte energigradene. Dette er vist ved kurvedrøfting:

1.derivert av boligpopulasjonens estimerte energiforbruk (se formel 9):

$$Q'(X_1)_{\text{estimert,bolig}} = 6008,14$$

1.derivert av næringspopulasjonens estimerte energiforbruk (se formel

$$Q'(x)_{\text{ikke-lineær,næring}} = 2 * 332,5x - 3 * 5,4x^2$$

Endringen i forbrukene er like ved betingelsen:

$$Q'(X_1)_{\text{estimert,bolig}} = Q'(x)_{\text{ikke-lineær,næring}}$$

Setter inn verdiene og setter lik 0:

$$-16,2x^2 + 665x - 6008,14 = 0$$

Røttene i 2.gradslikningen:

$$x_1 = 13,4 \text{ og } x_2 = 27,6$$

Det kan ytterligere vises at:

$$Q'(X_1)_{\text{estimert,bolig}} < Q'(x)_{\text{ikke-lineær,næring}} \text{ når } x \in < 13,4 - 27,6 >$$

Dette medfører at ved observerte energigrader mellom 13,4 °C og 27,6 °C (hvilket impliserer en utetemperatur på hhv. 3,6 °C og -10,6 °C), er næringspopulasjonens energiforbruk observert mer økende enn boligpopulasjonens energiforbruk.

For alle andre observerte energigrader er boligpopulasjonens energiforbruk mer økende enn næringspopulasjonen. Dette kan vises ved at:

$$Q'(X_1)_{\text{estimert,bolig}} > Q'(x)_{\text{ikke-lineær,næring}} \text{ når } x \text{ ikke er } \in < 13,4 - 27,6 >$$

4 Oppsummering

4.1 Energiforbruk og vind

Det er påvist en sammenheng mellom forbruk og vindretning for samtlige populasjonsutvalg. Det er signifikante forskjeller i energiforbruk ved vindretning fra nord og øst, kontra sør og vest (se definisjon av retninger i kapittel 3.1.3). Økningen i energiforbruket grunnet konveksjonsvarmetap kan teoretisk sett skyldes flere variable faktorer, herunder økt temperaturdifferanse og/eller økt vindstyrke. Fordelingen av vindstyrke på vindretningen viser at det er høyere sannsynlighet for at vind fra nord er kraftigere enn vind fra sør. Rangert etter høyeste forventningsverdi for observert vindstyrke er rekkefølgen; nord, øst, sør og vest. Det er også påvist at observerte energigrader varierer med vindretningen. Fordelingen av observerte energigrader på vindretningene indikerer at det kan forventes høyere energigrader ved vind fra nord og øst, kontra vind fra sør og vest. Vindretningens påvirkning på energiforbruket antas at hovedsakelig skyldes forskjellene i forventede energigrader ved vind fra de ulike himmelretningene, og den kraftige korrelasjonen mellom energiforbruk og energigrader.

Vindroseplottet (se kapittel 3.1.4) viser mer presist frekvensfordelingen av de registrerte vindretningene. Det er observert at fordelingen er relativt jevn rundt toppunktene ca. 30° (innenfor definert nordlig vindretning) og 205° (innenfor definert sydlig retning). Det observeres også at det er relativt lave frekvenser av registrerte vindretninger fra ca. 270° til ca. 355° , og fra ca. 75° til 170° . Således ville det gitt større variasjoner i analysen av vindstyrkefordelingen på vindretningene dersom en hadde analysert etter nord-øst (45°), sør-øst (135°), sør-vest (225°) og nord-vest (315°).

Ser en bort i fra observasjonene av retningene øst og vest (grunnet lave frekvenser av vindretningsobservasjoner), er forventningsverdien for både energiforbruket og vindstyrken vesentlig høyere ved vind fra nord enn fra sør. Det er ikke påvist en statistisk signifikant sammenheng mellom observert vindstyrke og energiforbruk, da det ved multippel lineærregresjon er vist at vindstyrken er en signifikant forklaringsvariabel for energiforbruket i samtlige populasjonsutvalg (se kapittel 3.2.4, 3.3.4 og 3.4.4).

4.2 Signifikante forklaringsvariabler

Det er påvist at energiforbruket til samtlige populasjonsutvalg statistisk sett kan forklares ved flere meteorologiske variabler. Variablene som er funnet statistisk signifikante ved multippel lineærregresjon er henholdsvis vindstyrke, antall soltimer, midlere lufttrykk, og energigrader.

Det er påvist at økende antall soltimer reduserer energiforbruket for samtlige populasjonsutvalg, da koeffisientene for soltimer i alle de tre tilfellene er negative. Koeffisientene for samtlige av de resterende forklaringsvariablene er positive. Dette indikerer at energiforbruket øker, dersom variablenes verdi øker. Således er det vist at økning i henholdsvis vindstyrke, midlere lufttrykk, eller energigrader vil enkeltvis bidra til å øke energiforbruket. Det antas at vindstyrke og midlere lufttrykk har innvirkning på konveksjonskoeffisienten (se vedlegg 3) mens energigrader > 0 er en forutsetning for varmeutveksling.

Samtlige konstantledd ved de ulike multippel lineærregresjonene var av negativ størrelse. Dette skyldes hovedsakelig at variabelen midlere lufttrykk inngår i beregningene. I motsetning til de andre meteorologiske variablene, nærmer ikke lufttrykket seg på noe tidspunkt 0. Dette, kombinert med stigningsgraden (koeffisienten) til variabelen medfører at konstantleddet i regresjonen blir negativt.

Ved å benytte de overnevnte signifikante variablene i multippel lineærregresjon, er energiforbruket til boligpopulasjonen forklart mer presist enn ved enkel lineærregresjon med energigrader som forklaringsvariabel. Dette er vist ved at forklart varians ($R^2 = 0,9712$) er større ved multippel lineærregresjon enn ved enkel lineærregresjon ($R^2 = 0,9678$). Dette gjelder ikke for totalpopulasjonen og næringspopulasjonen, hvilket antas å skyldes næringspopulasjonens antatt ikke-lineære energiforbruk (se kapittel 3.4.6) og at totalpopulasjonens observerte energiforbruk utgjør summen av hhv. boligpopulasjonens og næringspopulasjonens observerte energiforbruk.

Det slutes at de meteorologiske faktorene vindstyrke, antall soltimer, midlere lufttrykk, og energigrader mer presist forklarer energiforbruket til en boligpopulasjon, enn det forklares av energigrader alene, ettersom forklart varians for multippel lineærregresjon er større enn forklart varians ved enkel lineærregresjon. Dersom en populasjon inneholder næringskunder, er de nevnte meteorologiske faktorene mindre egnet som forklaring enn energigrader alene. Til tross for dette er de meteorologiske forklaringsvariablene funnet statistisk signifikante for samtlige populasjoner. Det er således konkludert med at vindstyrke, antall soltimer, midlere

lufttrykk, og energigrader påvirker energiforbruket til oppvarming. Det er ikke påvist i hvilken grad disse faktorene virker inn. Grunnet kompleksiteten i kurvedrøfting av multippel ikke-lineær regresjon er dette ikke utredet nærmere i studien, men kan være interessant å se nærmere på i fremtidige studier.

4.3 Estimering av varmtvannsandel

4.3.1 Totalpopulasjonen

Estimering av andelen energi til oppvarming av varmtvann, er gjort ved enkel lineærregresjon med energigrader som forklaringsvariabel, med bakgrunn i antakelsen om at energiforbruket består av én temperaturavhengig del og én temperaturuavhengig del, der sistnevnte utgjør energiforbruket til oppvarming av varmtvann (se kapittel 2.9.5). Det er påvist at denne antakelsen er relevant for boligpopulasjonen (se kapittel 3.3.5), men ikke for næringspopulasjonen (se kapittel 3.4.5), og følgelig er den ikke relevant for den totale populasjonen.

Det slutes dermed at enkel lineærregresjon som metode for å estimere energibehovet for oppvarming av varmtvann for en blandet populasjon (bolig og næring) medfører uriktige andelsberegninger.

Ved alternativ metode som tar hensyn til antakelsen om ikke-linearitet i sammenhengen mellom næringspopulasjonens observerte energiforbruk og de observerte energigrader, er andelen energi brukt til oppvarming av varmtvann estimert til 28,7% (se tabell 8). Dette estimatet er ikke direkte sammenliknbart med Kristensens estimeringer (se tabell 1) da det er tvilsomt at observasjonsfordelingen mellom næring og bolig i de to estimatene er like.

4.3.2 Boligpopulasjonen

Plottet av boligpopulasjonens observerte energiforbruk, mot det estimerte energiforbruket basert på observerte energigrader (se figur 11), viser en svært lineær sammenheng. Spearmans korrelasjonskoeffisient er signifikant beregnet til 0,9833, og forklart varians er beregnet til 0,9678. Plottet viser at det observerte energiforbruket har en liten antydning til å krumme i ekstremalpunktene ved min/maks observert temperaturdifferanse.

For boligpopulasjonen er konstantleddet ved lineærregresjon positivt, og beregnet til 33447,16. Dette medfører at energiforbruket for denne spesifikke boligpopulasjonen kan antas

å konvergere mot 33,4 MWh pr. observasjon når observerte energigrader går mot 0. Med andre ord utgjør det temperaturuavhengige energiforbruket, 33,4 MWh pr. døgn.

For boligkundene utgjør dette i årene 2008 til 2011 mellom 31,5% (2010) og 38,0% (2008) av det totale faktisk observerte energiforbruket (se tabell 6). Variasjonen i varmtvannsandelen skyldes det varierende temperaturavhengige forbruket.

Til sammenlikning estimerer Kristensen (2014) i sin masteroppgave at varmtvannsandelen for boligkundene i samme tidsperiode utgjør mellom 31,5% (2010) og 35,7% (2008) (se tabell 1).

Differansen i andelene mellom de to ulike metodene antas å skyldes at estimering i enkel lineærregresjon forutsetter at varmtvannsforbruket er relativt konstant i hele observasjonsperioden, i motsetning til Kristensens estimering der det beregnes energiforbruk for hvert år. Kristensens estimering tar dermed hensyn til endret forbrukeratferd fra år til år.

Det slutes dermed, at antakelsen om konstant energibehov til oppvarming av varmtvann for boligpopulasjonen i hele observasjonsperioden, er en potensiell svakhet ved regresjonsmetoden for estimering under forutsetning om at observasjonsperioden dekker flere år. Det antas likefullt at estimering ved lineærregresjon er adekvat dersom hver observasjonsperiode tilsvarer ett år, grunnet boligpopulasjonens sterke lineære korrelasjon med observerte energigrader. Energiandelsestimering ved observasjonsperiode tilsvarende ett år er ikke utredet i studien, og antakelsen bør studeres mer inngående, for eksempel der måling av volumforbruket av vann kan inngå.

4.3.3 Næringspopulasjonen

For næringspopulasjonen er estimering av energiforbruk til varmtvann ved enkel lineærregresjon mot temperaturdifferanse påvist som ikke relevant, da det resulterer i negativt energiforbruk til varmtvann.

Plottet mellom observert og estimert energiforbruk (se figur 15), viser en tydelig krumming i ekstremalpunktene (min/maks observerte energigrader), hvilket tyder på en ikke-lineær sammenheng mellom observert energiforbruk og energigrader. Det er også påvist en noe divergerende sammenheng mellom økte energigrader og det observerte energiforbruket. Spearmans korrelasjonskoeffisient for lineærregresjonen er signifikant beregnet til 0,9262, og forklart varians er beregnet til 0,8588. Dette er betydelig lavere enn for boligpopulasjonen (se kapittel 4.3.1) og bygger oppunder argumentasjonen om ikke-lineær sammenheng mellom

observert energiforbruk og temperaturdifferanse. Det er ikke påvist noen konkret årsak til at det er slik, men det antas at ikke-lineær sammenheng kan skyldes tekniske reguleringer, ikke-statistisk innetemperatur og/eller endringer i brukeratferd (eksempelvis kan en ansatt på sitt kontor ha installert en elektrisk varmeovn. Forutsatt at avgitt effekt fra fjernvarme er styrt for å oppnå en konstant innetemperatur, vil i praksis den elektriske ovnen overta effektbehovet dersom innehaveren av kontoret ønsker høyere innetemperatur enn fjernvarmeanlegget styres for å oppnå). Dette vil ikke fremgå spesifikt i datasettet.

Til tross for at det ikke er funnet noen termofysisk forklaring som bygger oppunder ikke-lineær sammenheng, viser det seg at energiforbruket for næringspopulasjonen kan beskrives matematisk som en kubisk funksjon av observerte energigrader (se kapittel 3.4.6).

Spearmankorrelasjonen til denne tilnærmingen er beregnet til 0,9262 (identisk med lineærregresjon). Forklart varians er beregnet til 0,8673, hvilket er noe høyere enn for lineærregresjon. Ved denne matematiske beskrivelsen av energiforbruket, er konstantleddet positivt, og beregnet til 15274,8. Dette medfører at det temperatuavhengige energiforbruket for denne spesifikke næringspopulasjonen er estimert til 15,3 MWh pr. døgn.

Ved å integrere fra energigrader = 0 °C til høyeste observasjon av energigrader = 35,1 °C (se kapittel 3.4.7), er andelen energi til oppvarming av varmtvann for næringspopulasjonen beregnet til 16,4% av det hele det observerte energiforbruket i perioden 2008 – 2011.

Ved antatt relativt konstant energibehov til oppvarming av varmtvann utgjør dette for denne spesifikke næringspopulasjonen 8,6 MWh pr. døgn, hvilket tilsvarer en årlig andel på 12,6% (2010) - 22,3% (2008). Til sammenlikning estimerer Kristensen (2014) i sin masteroppgave at varmtvannsandelen for næringskundene i samme tidsperiode utgjør mellom 11,4% (2010) og 16,2% (2008) (se tabell 1).

Differansen i andelene mellom de to ulike metodene antas å skyldes at estimering ved ikke-lineær regresjon forutsetter at varmtvannsforbruket er konstant i hele observasjonsperioden, i motsetning til Kristensens estimering der det forutsettes konstant energibehov for ett år om gangen. Kristensens estimering tar dermed hensyn til endret forbrukeratferd fra år til år.

Det er således sluttet, at antakelsen om konstant energibehov til oppvarming av varmtvann for næringspopulasjonen, er en svakhet ved den ikke lineære regresjonsmetoden for estimering under forutsetning om at observasjonsperioden dekker flere år.

4.4 Temperatursensitivitet

Ved kurvedrøfting er det vist at kubisk tilnærming til næringspopulasjonens energiforbruk resulterer i et teoretisk maksimalt energiuttak ved $-24\text{ }^{\circ}\text{C}$, hvilket er $4\text{ }^{\circ}\text{C}$ lavere enn dimensjonerende utetemperatur for Oslo-området ($-20\text{ }^{\circ}\text{C}$). Det er uvisst om reelt maksimalt energiuttak inntreffer ved $-20\text{ }^{\circ}\text{C}$, og det ville således ha vært interessant å studere næringspopulasjonens energiforbruk mer inngående, der også de reelle tekniske installasjoner kan tas i betraktning.

Ved å sammenlikne stigningskoeffisienten til funksjonen for det estimerte energiforbruket til boligpopulasjonen og stigningskurven til det estimerte energiforbruket til næringspopulasjonen, er det vist at de to populasjonene har ulik temperatursensitivitet. Energiforbruket til næringspopulasjonen er mer sensitivt (vil endres i større grad) enn boligpopulasjonen, ved endringer i utetemperaturen mellom $3,6\text{ }^{\circ}\text{C}$ og $-10,6\text{ }^{\circ}\text{C}$. Ved alle andre observerte energigrader er endringen i energiforbruket til boligpopulasjonen mer sensitivt for temperaturendringer enn næringspopulasjonen. Den observerte ulikheten i temperatursensitivitet er forutsatt at sammenhengen mellom næringspopulasjonens energiforbruk og observerte energigrader ikke er lineær.

Denne antatte ikke-lineariteten kan ikke forklares ut ifra enkle termodynamiske lover, og vil kunne være svært interessant å studere nærmere der en kan ta tekniske faktorer i betraktning.

5 Konklusjon

Studien har først tatt utgangspunkt i problemstillingen ”*Er det grunnlag for å si at meteorologiske omstendigheter, utover temperatur, påvirker energiforbruket til romoppvarming?*”.

Det er påvist at forventningsverdien til det observerte energiforbruket for samtlige populasjoner er ulik avhengig av hvilken retning vinden kommer fra. Det kan forventes høyere energiforbruk ved vind fra nord og øst, enn vind fra sør og vest. Dette er sammenfallende med høyere forventede vindstyrker fra nord og øst enn fra sør og vest. I tillegg er det sammenfallende høyere forventningsverdi av energigrader ved vind fra nord og øst, enn fra sør og vest. Det er konkludert med at konveksjonstapet er påvirket av hvilken retning vinden kommer fra.

Ved multippel lineærregresjon er det vist at flere meteorologiske variable kan benyttes som forklaringsvariabler i tillegg til energigrader. Særlig er dette gjeldende dersom populasjonsutvalget er boligkunder. For denne populasjonen er det vist at multippel lineærregresjon med variablene midlere vindstyrke, antall soltimer, midlere lufttrykk og energigrader resulterer i et estimert energiforbruk som ligger nærmere det faktisk observerte energiforbruket enn lineærregresjon der energigrader er eneste forklaringsvariabel. For en boligpopulasjon er det således grunnlag for å si at meteorologiske omstendigheter utover utetemperatur (energigrader) påvirker energiforbruket til oppvarming.

Forklart varians for hhv. næringspopulasjonen og totalpopulasjonen er lavere ved multippel lineærregresjon enn ved enkel lineærregresjon med energigrader som forklaringsvariabel. I disse to populasjonene er således det observerte energiforbruket bedre forklart av observerte energigrader alene. Det observerte energiforbruket til næringspopulasjonen er påvist som ikke-lineær korrelert med energigrader, og har følgelig innvirkning på totalpopulasjonen da sistnevnte utgjør summen av hhv. boligpopulasjonen og næringspopulasjonen. Denne ikke-lineære sammenhengen medfører at multippel lineærregresjon har lavere forklart varians enn for enkel lineærregresjon med energigrader som forklaringsvariabel.

Like fullt er de meteorologiske forklaringsvariablene funnet statistisk signifikante, hvilket medfører at disse to populasjonenes energiforbruk kan forklares påvirket av meteorologiske omstendigheter.

Videre har studien sett på problemstillingen ”*Kan lineærregresjon benyttes som metode for å estimere andelen oppvarmingsenergi til varmtvann?*”. Det fremkommer av studien at næringspopulasjonens energiforbruk følger en ikke-lineær sammenheng med energigrader. Følgelig er lineærregresjon ikke relevant som metode for å estimere andelen energi til oppvarming av varmtvann for en populasjon der næringskunder inngår. Det er ikke påvist noen årsak til den ikke-lineære sammenhengen, utover at det antas at det skyldes tekniske eller bruksmessige faktorer.

For boligpopulasjonen konkluderes det med at lineærregresjon kan benyttes til å estimere andelen energi til oppvarming av varmtvann. Det er likevel en svakhet ved estimeringsmetoden i studien at den ikke tar hensyn til endret forbrukeratferd over flere observasjonsår. Antakelsen om konstant energibehov til oppvarming av varmtvann kan motvirkes ved å analysere det observerte energiforbruket ett år om gangen.

Under forutsetning om at energiforbruket til boligpopulasjonen er lineært korrelert med energigrader, og at næringspopulasjonen er kubisk korrelert med energigrader, er det påvist ulik temperatursensitivitet i de to populasjonene. Kurvedrøfting indikerer at energiforbruket til boligpopulasjonen er mer følsomt for temperaturforandringer der utetemperaturen er fra 3,6 °C og oppover, samt ved -10,6 °C og nedover. I temperaturområdet 3,6 °C til -10,6 °C vil næringspopulasjonens energiforbruk være mer følsomt for temperaturforandringer enn boligpopulasjonens energiforbruk.

Dette følger som et implisitt resultat av studien, og det konkluderes med at dette bør utredes nærmere i fremtidige studier.

Litteraturliste

- Bøeng, A. C. & Holstad, M. (2013). Fakta om energi - Utviklingen i energibruk i Norge: Statistisk Sentralbyrå. 59 s.
- Bøeng, C. & Rosnes, O. (2013). Konsekvenser av Energieffektiviseringsdirektivet. *Rapporter*, 26/2013: Statistisk sentralbyrå. 33 s.
- Çengel, Y. A. (2006). *Heat and Mass Transfer: A Practical Approach*. 3rd utg.: McGraw-Hill. 884 s.
- Edenhofer, O., Pichs-Madruga, R., Sokona, Y., Farahani, E., Kadner, S., Seyboth, K., Adler, A., Baum, I., Brunner, S., Eickemeier, P., et al. (2014). IPCC 2014: Summary For Policymakers, In: Climate Change 2014, Mitigation of Climate Change. *Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*: Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Europaportalen. (2013). *Energieffektiviseringsdirektivet*. Tilgjengelig fra: <http://www.regjeringen.no/nb/sub/europaportalen/eos/eos-notatbasen/notatene/2011/sep/energieffektiviseringsdirektivet.html?id=744609> (lest 27.04.2014).
- FN-sambandet. (2014a). *Klima*. FN.no. Tilgjengelig fra: <http://www.fn.no/Tema/Klima> (lest 27.09.2014).
- FN-sambandet. (2014b). *Øystater trues av havstigning*. FN.no. Tilgjengelig fra: <http://www.fn.no/Tema/Klima/OEystater-trues-av-havstigning> (lest 27.09.2014).
- Josteinm. (2009). *Fyringssesong*. 16. november 2009 utg. metlex.met.no: Meteorologisk Institutt. Tilgjengelig fra: <https://metlex.met.no/wiki/Fyringssesong>.
- Kristensen, M. S. (2014). *Fjernvarmebruk i Oslo: Formålsdeling, Temperaturkorrigering og trender for et utvalg kunder*. Masteroppgave. Norges miljø- og bioteknologiske universitet, Fakultet for miljøvitenskap og teknologi, Institutt for Naturforvaltning. 80 s.
- Magnussen, I. H., Splide, D. & Killingland, M. (2011). Energibruk i fastlands-Norge. I: Energibruksseksjonen (red.). *Rapport*, 9: Norges Vassdrags- og Energidirektorat. 66 s.
- Olje- og energidepartementet. (2011). *Fornybardirektivet en del av EØS-avtalen*. Nr.: 110/11: Regjeringen Stoltenberg II. 1 s.
- Oreskes, N. (2004). The scientific consensus on climate change. *Science*, 306 (5702): 1686-1686.
- Pedersen, L., Stang, J. & Ulseth, R. (2008). Load prediction method for heat and electricity demand in buildings for the purpose of planning for mixed energy distribution systems. *Energy and Buildings*, 40 (7): 1124-1134.
- Pöyry Management Consulting AS. (2014). Evaluering av modeller for klimajustering av energibruk. *Rapport nr 7*: Norges Vassdrags- og Energidirektorat. 53 s.
- Rajput, R. K. (2009). *Engineering Thermodynamics: A Computer Approach (SI Units Version)*: Jones & Bartlett Learning.
- Rosvold, K. A. (2014). *EU 20 20 20*. Store Norske Leksikon. http://snl.no/EU_20-20-20: SNL AS (lest 26.03.2014).
- Setsaas, L. (2009). *Inneklimateknikk*. 2. utg. VVS-teknikk: Gyldendal Norsk Forlag AS. 290 s.

- Standard Norge. (2007). *NS 3031: Beregninger av bygningers energiytelse. Metode og data*: Pronorm AS. 72 s.
- The UN Refugee Agency. (2014). *The Storm Ahead*. I: UNHCR (red.): UNHCR. Tilgjengelig fra: <http://www.unhcr.org/pages/49e4a5096.html> (lest 27.09.2014).
- Utne, H. & Andersen, E. (2011). *Folke- og Boligtellingen, boliger, november*. Statistisk Sentralbyrå. Tilgjengelig fra: <http://www.ssb.no/befolkning/statistikker/fobbolig/hvert-10-aar/2013-02-26> (lest 03.Sep.2014).

Figurliste

FIGUR 1 BOKSPLOT AV OBSVERT MIDLERE VINDSTYRKE FORDELT PÅ OBSVERT VINDRETNING.....	28
FIGUR 2 VINDROSEPLOT AV OBSVERT MIDLERE VINDSTYRKE OG OBSVERT VINDRETNING I GRADER	29
FIGUR 3 FOREDLING AV OBSVERTE ENERGIGRADER PÅ OBSVERT VINDRETNING	30
FIGUR 4 BOKSPLOT AV TOTALPOPULASJONENS ENERGIFORBRUK FORDELT PÅ HVERDAG/HELG	32
FIGUR 5 BOKSPLOT AV TOTALPOPULASJONENS ENERGIFORBRUK FORDELT PÅ VINDRETNING	33
FIGUR 6 TOTALPOPULASJONENS OBSVERTE ENERGIFORBRUK PLOTTET MOT UTREGNET VERDI AV MULTIPPEL LINEÆRREGRESJON	36
FIGUR 7 TOTALPOPULASJONENS OBSVERTE ENERGIFORBRUK, PLOTTET MOT UTREGNET VERDI AV ENKEL LINEÆRREGRESJON	38
FIGUR 8 BOKSPLOT AV BOLIGPOPULASJONENS ENERGIFORBRUK FORDELT PÅ HVERDAG/HELG	40
FIGUR 9 BOKSPLOT AV BOLIGPOPULASJONENS ENERGIFORBRUK FORDELT PÅ VINDRETNING.....	41
FIGUR 10 BOLIGPOPULASJONENS OBSVERTE ENERGIFORBRUK, PLOTTET MOT UTREGNET VERDI AV MULTIPPEL LINEÆRREGRESJON	42
FIGUR 11 BOLIGPOPULASJONENS OBSVERTE ENERGIFORBRUK, PLOTTET MOT UTREGNET VERDI AV ENKEL LINEÆRREGRESJON	44
FIGUR 12 BOKSPLOT AV NÆRINGSPOPULASJONENS ENERGIFORBRUK FORDELT PÅ HVERDAG OG HELG	46
FIGUR 13 BOKSPLOT AV NÆRINGSPOPULASJONENS ENERGIFORBRUK FORDELT PÅ VINDRETNING.....	47
FIGUR 14 NÆRINGSPOPULASJONENS OBSVERTE ENERGIFORBRUK, PLOTTET MOT UTREGNET VERDI AV MULTIPPEL LINEÆRREGRESJON	48
FIGUR 15 NÆRINGSPOPULASJONENS OBSVERTE ENERGIFORBRUK, PLOTTET MOT UTREGNET VERDI AV ENKEL LINEÆRREGRESJON	49
FIGUR 16 NÆRINGSPOPULASJONENS OBSVERTE ENERGIFORBRUK, PLOTTET MOT OBSVERTE ENERGIGRADER.....	51
 FORMEL 1 LINEÆRSAMMENHENG MELLOM ENERGIFORBRUK OG TEMPERATURDIFFERANSE	13
FORMEL 2 DEFINISJON AV SIGNIFIKANT KORRELASJON	21
FORMEL 3 GENERALISERT FLÉRVARIABELFUNKSJON.....	22
FORMEL 4 GENERALISERT ÉNVARIABELFUNKSJON.....	22
FORMEL 5 FUNKSJON FOR TOTALPOPULASJONENS ESTIMERTE ENERGIFORBRUK VED MULTIPPEL LINEÆRREGRESJON	35
FORMEL 6 FUNKSJON FOR TOTALPOPULASJONENS ESTIMERTE ENERGIFORBRUK VED ENKEL LINEÆRREGRESJON.....	37
FORMEL 7 ESTIMERT ÅRLIG ENERGIMENGDE TIL OPPVARMING AV VARMTVANN FOR TOTALPOPULASJONEN	39
FORMEL 8 FUNKSJON FOR BOLIGPOPULASJONENS ESTIMERTE ENERGIFORBRUK VED MULTIPPEL LINEÆRREGRESJON	42
FORMEL 9 FUNKSJON FOR BOLIGPOPULASJONENS ESTIMERTE ENERGIFORBRUK VED ENKEL LINEÆRREGRESJON.....	43
FORMEL 10 ESTIMERT ÅRLIG ENERGIMENGDE TIL OPPVARMING AV VARMTVANN FOR BOLIGPOPULASJONEN	44
FORMEL 11 FUNKSJON FOR NÆRINGSPOPULASJONENS ESTIMERTE ENERGIFORBRUK VED MULTIPPEL LINEÆRREGRESJON ..	47
FORMEL 12 FUNKSJON FOR NÆRINGSPOPULASJONENS ESTIMERTE ENERGIFORBRUK VED ENKEL LINEÆRREGRESJON	49
FORMEL 13 FUKSJONSUTTRYKK FOR IKKE-LINEÆR SAMMENHENG.....	50
FORMEL 14 ESTIMERING AV ANDEL ENERGI TIL OPPVARMING AV VARMTVANN FOR NÆRINGSPOPULASJONEN	52

FORMEL 15 GENERALISERT FORM AV FUNKSJONEN "Q IKKE-LINEÆR,NÆRING" (FORMEL 13)	53
FORMEL 16 SAMMENHENG MELLOM Q ESTIMERT VED LINEÆRREGRESJON FOR DE ULIKE POPULASJONENE	53
FORMEL 17 NEWTONS LOV FOR KJØLING.....	71
FORMEL 18 TEMPERATURSPESIFIKK KONVEKSJON	71
TABELL 1 ESTIMERING AV ENERGIANDEL FRA FJERNVARME TIL OPPVARMING AV VARMTVANN (KRISTENSEN 2014)	14
TABELL 2 KORRELASJONSKOEFFISIENTER FOR DE NUMERISKE METEOROLOGISKE VARIABLENE	26
TABELL 3 KORRELASJONSKOEFFISIENTER ENERGIFORBRUK (TOTALPOPULASJON) OG NUMERISKE METEOROLOGISKE VARIABLE.....	31
TABELL 4 ESTIMERT ÅRLIG ENERGIANDEL TIL VARMTVANN, TOTALPOPULASJON	39
TABELL 5 KORRELASJONSKOEFFISIENTER FOR ENERGIFORBRUK (BOLIGPOPULASJON) OG NUMERISKE METEOROLOGISKE VARIABLE.....	40
TABELL 6 ESTIMERT ÅRLIG ENERGIANDEL TIL VARMTVANN, BOLIGPOPULASJON	45
TABELL 7 KORRELASJONSKOEFFISIENTER FOR ENERGIFORBRUK (NÆRINGSPOPULASJON) OG NUMERISKE METEOROLOGISKE VARIABLE.....	45
TABELL 8 ESTIMERT TOTALANDEL ENERGI TIL OPPVARMING AV VARMTVANN 2008 - 2011	54
TABELL 9 ESTIMERT ÅRLIG ANDEL ENERGI TIL OPPVARMING AV VARMTVANN, NÆRINGSPOPULASJON	54

Vedleggsliste

Vedlegg 1 Fotokopi av skjema for "Valg av statistisk test"	69
Vedlegg 2 Fotokopi av Taushetserklæring.....	70
Vedlegg 3 Redegjørelse for konveksjonstap.....	71
Vedlegg 4 Test av normalfordeling.....	72
Vedlegg 5 Utskrift fra R – totalpopulasjonen & meteorologiske variable	73
Vedlegg 6 Utskrift fra RStudio – boligpopulasjonen	87
Vedlegg 7 Uskrift fra RStudio – næringspopulajsonen	93
Vedlegg 7 Fotokopi av tabeller fra Kristensen, Martin S sin masteroppgave	100

Vedlegg 1 Fotokopi av skjema for "Valg av statistisk test"

		X = 2 levels	X ≥ 3 levels	1 X	≥ 2 X	
Y	Interval	Parametric	T tests	ANOVA	Pearson correlation	Multiple linear regression
		Non-parametric	Wilcoxon tests	Kruskal-Wallis	Spearman correlation	
	Nominal / Ordinal		Chi square tests		Generalized linear regression	
			Difference testing		Trend evaluation	
		X				

Skjemaet benyttes ved å først velge responsvariabel Y, deretter forklaringsvariabel X.

Egenskapene eller utfallsrommet til Y og X, samt type test (ulikhet eller trend) definerer hvilken analysemetode som skal benyttes.

Eks. Parametrisk fordelt Y, og X med 3 eller flere utfall, impliserer ANOVA-test (analysis of variance).

Vedlegg 2 Fotokopi av Taushetserklæring



TAUSHETSERKLÆRING

Jeg erklærer herved at all informasjon om Hafslund Varme AS sine kunder skal anonymiseres ved alle former for publisering.

Jeg har ikke lov til å bringe med meg skriftlig arbeidsmateriale eller dokumenter ut av Hafslunds lokaler uten etter å ha innhentet særlig tillatelse for dette.

Jeg har ansvaret for at eventuelle konfidensielle papirer er nedlåst og at PC/terminaler ikke viser sensitive data ved fravær fra arbeidsstedet.

Jeg er oppmerksom på at brudd på taushetsplikten så vel under engasjementet som etter oppdragets opphør vil kunne medføre erstatningsansvar.

Sandvika
Sted

21. Okt 2014
Dato

Gregory Hansen
(Navn)

Vedlegg 3 Redegjørelse for konveksjonstap

Varmeoverføring som følge av konveksjon kan uttrykkes ved "Newtons lov for kjøling"(Çengel 2006):

Formel 17 Newtons Lov for Kjøling

$$\frac{dQ}{dt} = h * A_s * (T(t) - T_{env}) = h * A_s * \Delta T(t)$$

Dersom en kjenner energimengden og temperaturdifferansen over en gitt tidsperiode kan dette uttrykkes som *temperaturspesifikk konveksjon*^f:

Formel 18 Temperaturspesifikk Konveksjon

$$\frac{dQ}{dt * \Delta T(t)} = h * A_s$$

Således fremgår det entydig at den temperaturspesifikke konveksjonen er direkte proporsjonalt med to parametere – *konveksjonskoeffisienten* "h" og overflatearealet "A_s". Konveksjonskoeffisienten "h" er igjen en funksjon av flere ulike parametere. Çengel fastsetter at "*Experience shows that the convection heat transfer strongly depends on the fluid properties dynamic viscosity (μ), thermal conductivity k , density (ρ), and specific heat (C_p), as well as the fluid velocity...*". I all hovedsak vil hastigheten på fluidet ha en innvirkning på koeffisienten "h", hvilket leder til at en kan anta at *vindhastighet* vil ha en innvirkning på et byggs konveksjon. Hvorvidt differansen mellom byggets temperatur og luftens temperatur er negativ eller positiv, vil definere varmetransportens retning, dvs. hvorvidt bygget tar til seg eller gir fra seg varme.

^f Likningen for temperaturspesifikk konveksjon har benevnningen [W/K], ikke ulikt benevnningen for et byggs *varmetapstall*. En kan si at konveksjonstapet utgjør en del av byggets totale *varmetapstall*

Vedlegg 4 Test av normalfordeling

```
> shapiro.test(middelvind)
Shapiro-Wilk normality test
data:  middelvind
W = 0.9235, p-value < 2.2e-16          # p << 0.05 - IKKE normalfordelt

> shapiro.test(gd17)
Shapiro-Wilk normality test
data:  gd17
W = 0.9914, p-value = 4.145e-05 # p << 0.05 - IKKE normalfordelt

> shapiro.test(middelsky)
Shapiro-Wilk normality test
data:  middelsky
W = 0.8839, p-value < 2.2e-16          # p << 0.05 - IKKE normalfordelt

> shapiro.test(sol)
Shapiro-Wilk normality test
data:  sol
W = 0.8099, p-value < 2.2e-16          # p << 0.05 - IKKE normalfordelt

> shapiro.test(middelp)
Shapiro-Wilk normality test
data:  middelp
W = 0.9936, p-value = 0.0006528 # p << 0.05 - IKKE normalfordelt

> shapiro.test(nedbor)
Shapiro-Wilk normality test
data:  nedbor
W = 0.6754, p-value < 2.2e-16          # p << 0.05 - IKKE normalfordelt

> shapiro.test(spesRH)
Shapiro-Wilk normality test
data:  spesRH
W = 0.9751, p-value = 2.836e-11 # p << 0.05 - IKKE normalfordelt

> shapiro.test(middelRH)
Shapiro-Wilk normality test
data:  middelRH
W = 0.9143, p-value < 2.2e-16          # p << 0.05 - IKKE normalfordelt
```

Vedlegg 5 Utskrift fra R – totalpopulasjonen & meteorologiske variable

```
> # TESTALGORITME #
>
> # Hente inn datasett:
> setwd("~/Dropbox/Universitetet/Høst 2014/Masteroppgave/Excel/SUM år/R")

> dataset=read.table('sumv2.0.txt',header=T,sep='\t',dec=',')

> # Feste og Sjekke datasett:
> attach(dataset)
The following objects are masked from dataset (position 3):

    gd17, helg, id1, middelp, middelRH, middelsky, middelvind, nedbor, season, sol,
    spesRH, totalsum,
    ukedag, vindgrader, vindretn, ysum, zsum

> str(dataset)
'data.frame':901 obs. of 17 variables:
 $ id1      : Factor w/ 901 levels "01.01.08","01.01.09",...: 1 31 62 94 126 158
190 222 254 286 ...
 $ season   : Factor w/ 3 levels "fall","spring",...: 3 3 3 3 3 3 3 3 3 ...
 $ ukedag   : Factor w/ 7 levels "fredag","lordag",...: 6 4 7 1 2 5 3 6 4 7 ...
 $ helg     : Factor w/ 2 levels "J","N": 2 2 2 2 1 1 2 2 2 2 ...
 $ ysum     : num 138180 136120 149430 150670 149670 ...
 $ zsum     : num 73992 84069 103598 97808 76128 ...
 $ totalsum : num 212172 220189 253028 248478 225798 ...
 $ vindgrader: int 25 55 59 89 58 15 71 216 158 183 ...
 $ vindretn : Factor w/ 4 levels "E","N","S","W": 2 1 1 1 1 2 1 3 3 3 ...
 $ middelvind: num 5 3.7 5.2 4.5 4 1.9 2.7 1.7 2.9 3.7 ...
 $ gd17     : num 17.7 16.6 19.3 19.3 19.1 18.4 16.4 16.3 15.1 14.1 ...
 $ middelsky: num 6.3 7.7 7.3 8 8 8 8 8 7 ...
 $ sol      : num 0 0 0.6 0 0 0 0 0 1.8 ...
 $ middelp  : num 1023 1026 1024 1019 1001 ...
 $ nedbor   : num NA 1.1 1.3 0.9 0 18.4 1 4.2 0.9 6.7 ...
 $ spesRH   : num 2.7 3.2 2.5 2.4 2.8 3.1 3.7 3.7 4 4.1 ...
 $ middelRH : int 75 84 80 75 85 91 92 93 92 86 ...

> # Bestemme resposvariabel (Y) altså hvilket av forbrukene som skal analyseres
> # Responsvariablen Y = 'totalsum', dvs summen av alle forbrukene.
> .... [TRUNCATED]

        Shapiro-Wilk normality test

data: totalsum
W = 0.9773, p-value = 1.277e-10

> # p << 0.05 - liten støtte for H0. totalsum er ulik en normalfordeling
>
> # Er forbruket avhengig av helg eller hverdag?
> weekend=subset(dataset, .... [TRUNCATED]

> work=subset(dataset, helg=='N')
```

```
> boxplot(totalsum~helg, main='Energiforbruk, hverdag vs. helg',
col='lightskyblue1', xlab='Helg? Ja/Nei', ylab='Observert energiforbruk,
totalpopulasjo .... [TRUNCATED]
```

```
> # g=c(weekend$totalsum)
> # h=c(work$totalsum)
>
> # summary(g)
> # summary(h)
>
> # normalfordelt?
> shapiro.test(weekend$totalsum) # p << 0.05 - .... [TRUNCATED]
```

Shapiro-Wilk normality test

```
data: weekend$totalsum
W = 0.9721, p-value = 6.321e-05
```

```
> shapiro.test(work$totalsum) # p << 0.05 - ikke normalfordelt
```

Shapiro-Wilk normality test

```
data: work$totalsum
W = 0.9789, p-value = 5.056e-08
```

```
> # Wilcoxon's test for differences:
> # H0: Totalforbruk varierer ikke med dagkategori
> # H1: Totalforbruk varierer med dagkategori
> wilcox.test(tot .... [TRUNCATED]
```

Wilcoxon rank sum test with continuity correction

```
data: totalsum by helg
W = 83499, p-value = 0.8328
alternative hypothesis: true location shift is not equal to 0
```

```
> # p = 0.8328 > 0.05 - Støtte for H0 - forkast H1. Totalforbruk varierer ikke med
dagkategori
```

```
>
> # Definere forklaringsvariable (X):
>
> # Korrela .... [TRUNCATED]
```

```
> dataset2=dataset[corvector1]
```

```
> #plot(dataset2)
> cor.mat1=cor(dataset2, method='spearman')
```

```
> cor.mat1
```

	totalsum	middelvind	gd17	middelsky sol	middelp nedbor
spesRH	0.8124361	0.2718745	0.1060838	0.04839351	NA
totalsum	1.00000000	-0.1714525	0.97146832	0.07597861	NA
middelvind	-0.17145254	1.00000000	-0.21966475	0.17166494	NA
gd17	0.97146832	-0.21966475	1.00000000	0.07597861	NA
middelsky sol	0.07597861	0.17166494	0.07597861	1.00000000	NA
middelp nedbor	0.04839351	NA	NA	NA	1.00000000

```

gd17      0.97146832 -0.2196648  1.00000000  0.03199432  NA  0.03867104  NA -
0.8494312 0.2441086
middelsky 0.07597861  0.1716649  0.03199432  1.00000000  NA -0.15129606  NA
0.2748654 0.5768274
sol      NA      NA      NA      NA      1      NA      NA
NA      NA
middelp   0.04839351 -0.3160165  0.03867104 -0.15129606  NA  1.00000000  NA -
0.1404909 -0.1093535
nedbor    NA      NA      NA      NA      NA      NA      1
NA      NA
spesRH    -0.81243615  0.1060838 -0.84943121  0.27486536  NA -0.14049085  NA
1.0000000 0.2368366
middelRH  0.27187451 -0.2752527  0.24410865  0.57682736  NA -0.10935346  NA
0.2368366 1.0000000

```

```

> # Korrelasjon for Soltimer:
> cor.test(totalsum,sol, method='spearman')

```

Spearman's rank correlation rho

```

data: totalsum and sol
S = 48316131, p-value = 2.654e-14
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
-0.3026469

```

```

> # Observerer ganske svak negativ korrelasjon med forbruket - dvs soltimer opp =
forbruk ned, som forventet.
> # p << 0.05 - korrelasjonskoeffisient .... [TRUNCATED]

```

Spearman's rank correlation rho

```

data: middelvind and sol
S = 42415921, p-value = 0.000392
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
-0.1435719

```

```

> cor.test(gd17, sol, method='spearman')

```

Spearman's rank correlation rho

```

data: gd17 and sol
S = 46349261, p-value = 4.632e-10
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
-0.2496183

```

```

> cor.test(middelsky, sol, method='spearman')

```

Spearman's rank correlation rho

```
data: middelsky and sol
S = 66508187, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.7931213
```

```
> cor.test(middelp, sol, method='spearman')
```

```
Spearman's rank correlation rho
```

```
data: middelp and sol
S = 31101176, p-value = 6.516e-05
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.161484
```

```
> cor.test(nedbor, sol, method='spearman')
```

```
Spearman's rank correlation rho
```

```
data: nedbor and sol
S = 8446723, p-value = 0.001217
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.171984
```

```
> cor.test(spesRH, sol, method='spearman')
```

```
Spearman's rank correlation rho
```

```
data: spesRH and sol
S = 42426351, p-value = 0.0003817
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.1438531
```

```
> cor.test(middelRH, sol, method='spearman')
```

```
Spearman's rank correlation rho
```

```
data: middelRH and sol
S = 61781640, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.6656893
```

```
> cor.test(totalsum,nedbor, method='spearman')
```

Spearman's rank correlation rho

```
data: totalsum and nedbor
S = 24244983, p-value = 0.0004268
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.1568091
```

```
> # Observerer svært svak negativ korrelasjon med forbruket - økt nedbørsmengde =
redusert forbruk.
> # p < 0.05 - korrelasjonskoeffisient er statist .... [TRUNCATED]
```

Spearman's rank correlation rho

```
data: middelvind and nedbor
S = 17768125, p-value = 0.0006293
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.1522235
```

```
> cor.test(gd17, nedbor, method='spearman')
```

Spearman's rank correlation rho

```
data: gd17 and nedbor
S = 24626280, p-value = 8.226e-05
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.175002
```

```
> cor.test(middelsky, nedbor, method='spearman')
```

Spearman's rank correlation rho

```
data: middelsky and nedbor
S = 17868967, p-value = 0.000935
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.1474119
```

```
> cor.test(middelp, nedbor, method='spearman')
```

Spearman's rank correlation rho

```
data: middelp and nedbor
S = 29019018, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.3845942
```

```
> cor.test(spesRH, nedbor, method='spearman')
```

Spearman's rank correlation rho

```
data: spesRH and nedbor
S = 14678338, p-value = 7.476e-12
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.2996475
```

```
> cor.test(middelRH, nedbor, method='spearman')
```

Spearman's rank correlation rho

```
data: middelRH and nedbor
S = 15327961, p-value = 9.906e-10
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.2686518
```

```
> # Sjekker sammenheng med forbruk og vindretning:
```

```
> north=subset(dataset, vindretn=='N')
```

```
> east=subset(dataset, vindretn=='E')
```

```
> south=subset(dataset, vindretn=='S')
```

```
> west=subset(dataset, vindretn=='W')
```

```
> boxplot(totalsum~vindretn, main='Energiorbruk fordelt på Vindretning',
col='darkgoldenrod1', xlab='Vindretning', ylab='Observert energiforbruk,
totalpopula .... [TRUNCATED]
```

```
> # Sjekke normalitet:
```

```
> shapiro.test(north$totalsum) #nei
```

Shapiro-Wilk normality test

```
data: north$totalsum
W = 0.9824, p-value = 0.002483
```

```
> shapiro.test(east$totalsum) #nei
```

Shapiro-Wilk normality test

```
data: east$totalsum
W = 0.9798, p-value = 0.01525
```

```
> shapiro.test(south$totalsum) #nei
```

```

        Shapiro-Wilk normality test

data:  south$totalsum
W = 0.9716, p-value = 2.098e-06

> shapiro.test(west$totalsum) #nei

        Shapiro-Wilk normality test

data:  west$totalsum
W = 0.9575, p-value = 0.0008466

> # Kruskal-Wallis test:
> # H0: Forbruk varierer ikke med vindretning
> # H1: Forbruk varierer med vindretning
>
> kruskal.test(totalsum~vindretn) # .... [TRUNCATED]

        Kruskal-Wallis rank sum test

data:  totalsum by vindretn
Kruskal-Wallis chi-squared = 89.1029, df = 3, p-value < 2.2e-16

> # Er det korrelasjon mellom vindretning og vindstyrke?
> boxplot(middelvind~vindretn, main='Vindstyrke fordelt på Vindretning',
xlab='Vindretning', .... [TRUNCATED])

> # Normalfordelt?
> shapiro.test(north$middelvind) #nei

        Shapiro-Wilk normality test

data:  north$middelvind
W = 0.9568, p-value = 4.889e-07

> shapiro.test(east$middelvind) #nei

        Shapiro-Wilk normality test

data:  east$middelvind
W = 0.9465, p-value = 5.989e-06

> shapiro.test(south$middelvind) #nei

        Shapiro-Wilk normality test

data:  south$middelvind
W = 0.9392, p-value = 7.765e-11

> shapiro.test(west$middelvind) #nei

        Shapiro-Wilk normality test

```

```
data: west$middelvind
W = 0.9425, p-value = 6.847e-05
```

```
> kruskal.test(middelvind~vindretn) # p << 0.05 Liten støtte for H0 - dvs
vindstyrken varierer med vindretning
```

Kruskal-Wallis rank sum test

```
data: middelvind by vindretn
Kruskal-Wallis chi-squared = 76.6583, df = 3, p-value < 2.2e-16
```

```
> # Observerer generelt høyere vindstyrke fra north
>
> windRose(dataset, ws='middelvind', wd='vindgrader', angle=5, main='Vindrose',
paddle=F)
```

```
> # Er det korrelasjon mellom vindretning og energigrader?
> boxplot(gd17~vindretn, main='Energigrader fordelt på Vindretning',
xlab='Vindretning', ylab=.... [TRUNCATED])
```

```
> # Normalfordelt?
> shapiro.test(north$gd17) #nei
```

Shapiro-Wilk normality test

```
data: north$gd17
W = 0.9871, p-value = 0.01849
```

```
> shapiro.test(east$gd17) #ja
```

Shapiro-Wilk normality test

```
data: east$gd17
W = 0.9926, p-value = 0.5561
```

```
> shapiro.test(south$gd17) #nei
```

Shapiro-Wilk normality test

```
data: south$gd17
W = 0.9833, p-value = 0.0004073
```

```
> shapiro.test(west$gd17) #nei
```

Shapiro-Wilk normality test

```
data: west$gd17
W = 0.9758, p-value = 0.03006
```

```
> kruskal.test(gd17~vindretn) # p << 0.05 Liten støtte for H0 - dvs Energigrader
varierer med vindretning.
```

Kruskal-Wallis rank sum test

data: gd17 by vindretn

Kruskal-Wallis chi-squared = 84.7994, df = 3, p-value < 2.2e-16

```
> # Observerer at temperaturfall sammenhengende med vindretning. Skyldes gd17
kraftig korrelert med totalsum.
```

```
>
```

```
> # Sjekke korrelasjon mellom sesong .... [TRUNCATED]
```

```
> vår=subset(dataset, season=='spring')
```

```
> høst=subset(dataset, season=='fall')
```

```
> shapiro.test(vinter$middelvind) # nei, p << 0.05
```

Shapiro-Wilk normality test

data: vinter\$middelvind

W = 0.9094, p-value = 8.316e-14

```
> shapiro.test(vår$middelvind) # nei, p << 0.05
```

Shapiro-Wilk normality test

data: vår\$middelvind

W = 0.9367, p-value = 1.545e-09

```
> shapiro.test(høst$middelvind) # nei, p << 0.05
```

Shapiro-Wilk normality test

data: høst\$middelvind

W = 0.9208, p-value = 1.109e-10

```
> boxplot(middelvind~season, main='Vindstyrke fordelt på årstid',
col='lightgoldenrod1', xlab='Årstid', ylab='Vindstyrke [m/s]', range=0)
```

```
> kruskal.test(middelvind~season) # p = 0.0369 Svak støtte for H0: vindstyrke
varierer med sesong.
```

Kruskal-Wallis rank sum test

data: middelvind by season

Kruskal-Wallis chi-squared = 6.6015, df = 2, p-value = 0.03686

```
> median(vinter$middelvind) # median = 2.0
```

```
[1] 2
```

```
> median(vår$middelvind) # median = 2.2
```

```
[1] 2.2
```

```

> median(høst$middelvind) # median 2.3
[1] 2.3

> # Med signifikante siffer = 1, er medianen lik for alle sesongene.
>
> cor.test(middelvind,gd17, method='spearman')

Spearman's rank correlation rho

data: middelvind and gd17
S = 148683597, p-value = 2.628e-11
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.2196648

> vind_tab=table(vindretn,season)

> chisq.test(vind_tab)

Pearson's Chi-squared test

data: vind_tab
X-squared = 25.3037, df = 6, p-value = 3e-04

> barplot(vind_tab, main='Frekvens av observert vindretning fordelt på årstidene',
xlab='vindretning',col=c('brown','lightgreen','cornflowerblue',"yel ..." ...
[TRUNCATED]

> #####
> # MULTIPPEL LINEÆRREGRESJON!
> ##### .... [TRUNCATED]

> summary(mlr1)

Call:
lm(formula = totalsum ~ middelvind + gd17 + sol + middelp + nedbor +
    middelRH)

Residuals:
    Min       1Q   Median       3Q      Max
-62210 -11990     521  11199  60373

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -325543.61   75977.35  -4.285 2.38e-05 ***
middelvind    2377.09     714.19    3.328 0.000968 ***
gd17          11702.57     173.09   67.609 < 2e-16 ***
sol           -1302.63     401.71   -3.243 0.001300 **
middelp         341.54      74.30    4.597 6.03e-06 ***
nedbor        -116.12     175.02   -0.663 0.507460
middelRH         53.33     123.10    0.433 0.665100
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16760 on 344 degrees of freedom

```

```

(550 observations deleted due to missingness)
Multiple R-squared: 0.939, Adjusted R-squared: 0.938
F-statistic: 882.8 on 6 and 344 DF, p-value: < 2.2e-16

```

```

> # nedbor ikke signifikant. fjerner fra regresjonen
> mlr2=lm(totalsum~middelvind+gd17+sol+middelp)

```

```

> summary(mlr2)

```

Call:

```
lm(formula = totalsum ~ middelvind + gd17 + sol + middelp)
```

Residuals:

Min	1Q	Median	3Q	Max
-69901	-10709	-173	11191	60776

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-284211.15	57232.38	-4.966	8.92e-07	***
middelvind	2692.12	554.92	4.851	1.56e-06	***
gd17	11584.80	127.31	90.995	< 2e-16	***
sol	-842.15	179.80	-4.684	3.49e-06	***
middelp	301.66	57.09	5.284	1.77e-07	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17100 on 601 degrees of freedom

(295 observations deleted due to missingness)

Multiple R-squared: 0.9416, Adjusted R-squared: 0.9412

F-statistic: 2421 on 4 and 601 DF, p-value: < 2.2e-16

```

> #setter inn koeffisienter og lager vektor for y-funksjon

```

```

> yfunkt=c(-284211.15+(11584.8*gd17)+(2692.12*mittelvind)-
(842.15*sol)+301.66*middelp)

```

```

> plot(yfunkt,totalsum, main='Observert energiforbruk vs. Q estimert',
ylab='Observert energiforbruk, totalpopula.... [TRUNCATED]

```

```

> lmfinal=lm(totalsum~yfunkt)

```

```

> summary(lmfinal)

```

Call:

```
lm(formula = totalsum ~ yfunkt)
```

Residuals:

Min	1Q	Median	3Q	Max
-69901	-10709	-173	11191	60776

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-0.13485	2106.44303	0.00	1	
yfunkt	1.00000	0.01014	98.66	<2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 17060 on 604 degrees of freedom
(295 observations deleted due to missingness)
Multiple R-squared: 0.9416, Adjusted R-squared: 0.9415
F-statistic: 9733 on 1 and 604 DF, p-value: < 2.2e-16

```

```
> abline(lmfinal, col='red',lwd=3)
```

```
> cor.test(totalsum,yfunk, method='spearman')
```

```
Spearman's rank correlation rho
```

```

data: totalsum and yfunk
S = 1144901, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.9691324

```

```
> mlr3=lm(totalsum~gd17)
```

```
> summary(mlr3)
```

```
Call:
```

```
lm(formula = totalsum ~ gd17)
```

```
Residuals:
```

```

      Min       1Q   Median       3Q      Max
-87710 -12631    510  12822  55208

```

```
Coefficients:
```

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept) 19408.31    1621.81   11.97  <2e-16 ***
gd17         11917.54     97.79  121.87  <2e-16 ***
---

```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```

Residual standard error: 18290 on 899 degrees of freedom
Multiple R-squared: 0.9429, Adjusted R-squared: 0.9429
F-statistic: 1.485e+04 on 1 and 899 DF, p-value: < 2.2e-16

```

```

> plot(gd17, totalsum, main='Energiforbruk vs. Energigrader', xlab='Energigrader
[°C]',ylab='Observert energiforbruk, totalpopulasjonen [kWh]')

```

```
> abline(mlr3, col='red', lwd=3)
```

```

> #####
> # ENKEL LINEÆRREGRESJON
> #####
>
> .... [TRUNCATED]

```

```

> plot(alt.yfunk,totalsum, main='Observert energiforbruk vs. Q estimert,total',
ylab='Observert energiforbruk, .... [TRUNCATED]

```

```
> abline(lmfinal, col='red', lwd=3)
```

```

> lmfinal2=lm(totalsum~alt.yfunk)

> abline(lmfinal2, col='green', lwd=3)

> cor.test(totalsum,alt.yfunk, method='spearman')

Spearman's rank correlation rho

data: totalsum and alt.yfunk
S = 3478164, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.9714683

> #####
> # DIGRESJON #
> #####
>
>
imr=lm(totalsum~gd17+sol+middelvind+I(gd17^2)+I(sol^2)+I(middelvind^2)+I(gd17^3)+I(
sol^3)+I(midde .... [TRUNCATED]

> summary(imr)

Call:
lm(formula = totalsum ~ gd17 + sol + middelvind + I(gd17^2) +
I(sol^2) + I(middelvind^2) + I(gd17^3) + I(sol^3) + I(middelvind^3))

Residuals:
    Min       1Q   Median       3Q      Max
-80528 -10380     -25   10799   58496

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   45812.021    7815.120   5.862 7.58e-09 ***
gd17           6827.312    1298.133   5.259 2.02e-07 ***
sol          -3162.531    1022.023  -3.094 0.00206 **
middelvind     2017.903    3865.920   0.522 0.60188
I(gd17^2)       261.547      80.656   3.243 0.00125 **
I(sol^2)        297.844     194.986   1.528 0.12716
I(middelvind^2) -360.528    1122.758  -0.321 0.74824
I(gd17^3)       -4.119       1.540  -2.674 0.00769 **
I(sol^3)        -7.770       9.316  -0.834 0.40459
I(middelvind^3)  53.096      95.065   0.559 0.57670
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17000 on 596 degrees of freedom
(295 observations deleted due to missingness)
Multiple R-squared:  0.9427,    Adjusted R-squared:  0.9419
F-statistic: 1090 on 9 and 596 DF,  p-value: < 2.2e-16

> imr2=lm(totalsum~gd17+sol+I(gd17^2)+I(gd17^3))

```

```

> summary(imr2)

Call:
lm(formula = totalsum ~ gd17 + sol + I(gd17^2) + I(gd17^3))

Residuals:
    Min       1Q   Median       3Q      Max
-75826 -10705    164   11164   56315

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  54167.928   6558.044    8.260 9.32e-16 ***
gd17          5887.435   1269.038    4.639 4.29e-06 ***
sol          -988.006    182.950   -5.400 9.59e-08 ***
I(gd17^2)     312.220     79.446    3.930 9.48e-05 ***
I(gd17^3)     -5.058      1.529   -3.309 0.000993 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17320 on 601 degrees of freedom
(295 observations deleted due to missingness)
Multiple R-squared:  0.9401,    Adjusted R-squared:  0.9397
F-statistic: 2357 on 4 and 601 DF,  p-value: < 2.2e-16

> qtest=c(gd17+sol+(gd17^2)+(gd17^3))

> plot(log(qtest),log(totalsum))

> kurve1=curve(0+1*x, add=T, col='red', lwd=3)

```

Vedlegg 6 Utskrift fra RStudio – boligpopulasjonen

```
> # TESTALGORITME #
>
> # Hente inn datasett:
> setwd("~/Dropbox/Universitetet/Høst 2014/Masteroppgave/Excel/SUM år/R")

> dataset=read.table('sumv2.0.txt',header=T,sep='\t',dec=',')

> # Feste og Sjekke datasett:
> attach(dataset)

> str(dataset)
'data.frame':901 obs. of 17 variables:
 $ id1      : Factor w/ 901 levels "01.01.08","01.01.09",...: 1 31 62 94 126 158
190 222 254 286 ...
 $ season   : Factor w/ 3 levels "fall","spring",...: 3 3 3 3 3 3 3 3 3 ...
 $ ukedag    : Factor w/ 7 levels "fredag","lordag",...: 6 4 7 1 2 5 3 6 4 7 ...
 $ helg     : Factor w/ 2 levels "J","N": 2 2 2 2 1 1 2 2 2 2 ...
 $ ysum      : num 138180 136120 149430 150670 149670 ...
 $ zsum      : num 73992 84069 103598 97808 76128 ...
 $ totalsum  : num 212172 220189 253028 248478 225798 ...
 $ vindgrader: int 25 55 59 89 58 15 71 216 158 183 ...
 $ vindretn  : Factor w/ 4 levels "E","N","S","W": 2 1 1 1 1 2 1 3 3 3 ...
 $ middelvind: num 5 3.7 5.2 4.5 4 1.9 2.7 1.7 2.9 3.7 ...
 $ gd17      : num 17.7 16.6 19.3 19.3 19.1 18.4 16.4 16.3 15.1 14.1 ...
 $ middelsky : num 6.3 7.7 7.3 8 8 8 8 8 8 7 ...
 $ sol       : num 0 0 0.6 0 0 0 0 0 0 1.8 ...
 $ middelp   : num 1023 1026 1024 1019 1001 ...
 $ nedbor    : num NA 1.1 1.3 0.9 0 18.4 1 4.2 0.9 6.7 ...
 $ spesRH    : num 2.7 3.2 2.5 2.4 2.8 3.1 3.7 3.7 4 4.1 ...
 $ middelRH  : int 75 84 80 75 85 91 92 93 92 86 ...

> # Bestemme resposvariabel (Y) altså hvilket av forbrukene som skal analyseres
> # Responsvariablen Y = 'ysum', dvs summen av alle forbrukene.
>
> # .... [TRUNCATED]

      Shapiro-Wilk normality test

data:  ysum
W = 0.991, p-value = 2.506e-05

> # p << 0.05 - liten støtte for H0. ysum er ulik en normalfordeling
>
> # Er forbruket avhengig av helg eller hverdag?
> weekend=subset(dataset, hel .... [TRUNCATED]

> work=subset(dataset, hel=='N')

> boxplot(ysum~helg, main='Energiforbruk, hverdag vs. helg', col='lightskyblue1',
xlab='Helg? Ja/Nei', ylab='Observert energiforbruk, bolig [kWh]',range ....
[TRUNCATED]

> # normalfordelt?
> shapiro.test(weekend$ysum) # p << 0.05 - ikke normalfordelt
```

Shapiro-Wilk normality test

```
data: weekend$ysum
W = 0.9878, p-value = 0.02787
```

```
> shapiro.test(work$ysum) # p << 0.05 - ikke normalfordelt
```

Shapiro-Wilk normality test

```
data: work$ysum
W = 0.9908, p-value = 0.0004742
```

```
> # Wilcoxon's test for differences:
> # H0: Totalforbruk varierer ikke med dagkategori
> # H1: Totalforbruk varierer med dagkategori
> wilcox.test(ysu .... [TRUNCATED]
```

Wilcoxon rank sum test with continuity correction

```
data: ysum by helg
W = 83658, p-value = 0.7978
alternative hypothesis: true location shift is not equal to 0
```

```
> # p = 0.7978 > 0.05 - Støtte for H0 - forkast H1. Totalforbruk varierer ikke med dagkategori
```

```
>
> # Sjekker sammenheng med forbruk og vindretning:
> .... [TRUNCATED]
```

```
> east=subset(dataset, vindretn=='E')
```

```
> south=subset(dataset, vindretn=='S')
```

```
> west=subset(dataset, vindretn=='W')
```

```
> boxplot(ysum~vindretn, main='Energiforbruk fordelt på Vindretning',
col='darkgoldenrod1', xlab='Vindretning', ylab='Observert energiforbruk, bolig [kW]
.... [TRUNCATED]
```

```
> # Sjekke normalitet:
> shapiro.test(north$ysum) #nei
```

Shapiro-Wilk normality test

```
data: north$ysum
W = 0.9933, p-value = 0.2947
```

```
> shapiro.test(east$ysum) #nei
```

Shapiro-Wilk normality test

```
data: east$ysum
W = 0.9904, p-value = 0.3186
```

```

> shapiro.test(south$ysum) #nei

      Shapiro-Wilk normality test

data:  south$ysum
W = 0.9847, p-value = 0.0008733

> shapiro.test(west$ysum) #nei

      Shapiro-Wilk normality test

data:  west$ysum
W = 0.9707, p-value = 0.01047

> # Kruskal-Wallis test:
> # H0: Forbruk varierer ikke med vindretning
> # H1: Forbruk varierer med vindretning
>
> kruskal.test(ysum~vindretn) # p < .... [TRUNCATED]

      Kruskal-Wallis rank sum test

data:  ysum by vindretn
Kruskal-Wallis chi-squared = 88.0362, df = 3, p-value < 2.2e-16

> #####
> # MULTIPPEL LINEÆRREGRESJON!
> ##### .... [TRUNCATED]

> summary(mlr1)

Call:
lm(formula = ysum ~ middelvind + gd17 + sol + middelp + nedbor +
    middelRH)

Residuals:
    Min       1Q   Median       3Q      Max
-18849.3  -3371.4   528.9   3553.4  20862.4

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -46281.92   26220.08  -1.765  0.07843 .
middelvind    982.55    246.47    3.986 8.19e-05 ***
gd17         5892.86     59.73   98.651 < 2e-16 ***
sol          -458.13    138.63   -3.305  0.00105 **
middelp       74.35     25.64    2.900  0.00398 **
nedbor        -45.56     60.40   -0.754  0.45120
middelRH       84.98     42.48    2.000  0.04625 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5785 on 344 degrees of freedom
(550 observations deleted due to missingness)
Multiple R-squared:  0.97,    Adjusted R-squared:  0.9695

```

F-statistic: 1856 on 6 and 344 DF, p-value: < 2.2e-16

```
> # nedbor ikke signifikant. fjerner fra regresjonen  
> mlr2=lm(ysum~middelvind+gd17+sol+middelp)
```

```
> summary(mlr2)
```

Call:

```
lm(formula = ysum ~ middelvind + gd17 + sol + middelp)
```

Residuals:

Min	1Q	Median	3Q	Max
-19367.3	-3932.0	-123.9	4016.7	19987.3

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-38835.11	20241.23	-1.919	0.055507 .
middelvind	1068.94	196.26	5.447	7.49e-08 ***
gd17	5954.65	45.03	132.248	< 2e-16 ***
sol	-327.19	63.59	-5.145	3.62e-07 ***
middelp	71.59	20.19	3.546	0.000422 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6048 on 601 degrees of freedom

(295 observations deleted due to missingness)

Multiple R-squared: 0.9712, Adjusted R-squared: 0.971

F-statistic: 5061 on 4 and 601 DF, p-value: < 2.2e-16

```
> #setter inn koeffisienter og lager vektor for y-funksjon
```

```
> yfunk=c(-38835.11+(5954.65*gd17)+(1068.94*middelvind)-(327.19*sol)+71.59*middelp)
```

```
> plot(yfunk,ysum, main='Observert energiforbruk vs. Q estimert', ylab='Observert  
energiforbruk, bolig [kWh]', xlab='Verdi av Q es.... [TRUNCATED]')
```

```
> lmfinal=lm(ysum~yfunk)
```

```
> summary(lmfinal)
```

Call:

```
lm(formula = ysum ~ yfunk)
```

Residuals:

Min	1Q	Median	3Q	Max
-19367	-3932	-124	4017	19987

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.680e+00	9.054e+02	0.003	0.998
yfunk	1.000e+00	7.011e-03	142.642	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6033 on 604 degrees of freedom

(295 observations deleted due to missingness)

Multiple R-squared: 0.9712, Adjusted R-squared: 0.9711

F-statistic: 2.035e+04 on 1 and 604 DF, p-value: < 2.2e-16

```
> abline(lmfinal, col='red',lwd=3)
```

```
> cor.test(ysum,yfunk, method='spearman')
```

Spearman's rank correlation rho

data: ysum and yfunk

S = 569712.1, p-value < 2.2e-16

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho

0.98464

```
> #####
```

```
> # ENKEL LINEÆRREGRESJON
```

```
> #####
```

```
>
```

```
> .... [TRUNCATED]
```

```
> summary(mlr3)
```

Call:

```
lm(formula = ysum ~ gd17)
```

Residuals:

	Min	1Q	Median	3Q	Max
	-20561.2	-4695.3	-103.6	4578.8	18043.5

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	33447.16	606.17	55.18	<2e-16 ***
gd17	6008.14	36.55	164.39	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6837 on 899 degrees of freedom

Multiple R-squared: 0.9678, Adjusted R-squared: 0.9678

F-statistic: 2.702e+04 on 1 and 899 DF, p-value: < 2.2e-16

```
> #plot(gd17, ysum, main='Forbruket vs. Energigrader', xlab='Energigrader  
[°C]',ylab='Sum av alle forbruk [10 Wh]')
```

```
> #abline(mlr3, col='red', lwd=3)
```

```
.... [TRUNCATED]
```

```
> plot(alt.yfunk,ysum, main='Observervert energiforbruk vs. Q estimert,bolig',  
ylab='Observervert energiforbruk, bolig [kWh]', x .... [TRUNCATED])
```

```
> lmfinal2=lm(ysum~alt.yfunk)
```

```
> abline(lmfinal2, col='green', lwd=3)
```

```
> cor.test(ysum,alt.yfunk, method='spearman')
```

Spearman's rank correlation rho

```
data: ysum and alt.yfunk
S = 2031944, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.9833318
```

```
> cor.test(ysum, yfunk, method='spearman')
```

Spearman's rank correlation rho

```
data: ysum and yfunk
S = 569712.1, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.98464
```

Vedlegg 7 Uskrift fra RStudio – næringspopulasjonen

```
> # TESTALGORITME #
>
> # Hente inn datasett:
> setwd("~/Dropbox/Universitetet/Høst 2014/Masteroppgave/Excel/SUM år/R")

> dataset=read.table('sumv2.0.txt',header=T,sep='\t',dec=',')

> # Feste og Sjekke datasett:
> attach(dataset)

> str(dataset)
'data.frame':901 obs. of 17 variables:
 $ id1      : Factor w/ 901 levels "01.01.08","01.01.09",...: 1 31 62 94 126 158
190 222 254 286 ...
 $ season   : Factor w/ 3 levels "fall","spring",...: 3 3 3 3 3 3 3 3 3 ...
 $ ukedag   : Factor w/ 7 levels "fredag","lordag",...: 6 4 7 1 2 5 3 6 4 7 ...
 $ helg     : Factor w/ 2 levels "J","N": 2 2 2 2 1 1 2 2 2 2 ...
 $ ysum     : num 138180 136120 149430 150670 149670 ...
 $ zsum     : num 73992 84069 103598 97808 76128 ...
 $ totalsum : num 212172 220189 253028 248478 225798 ...
 $ vindgrader: int 25 55 59 89 58 15 71 216 158 183 ...
 $ vindretn : Factor w/ 4 levels "E","N","S","W": 2 1 1 1 1 2 1 3 3 3 ...
 $ middelvind: num 5 3.7 5.2 4.5 4 1.9 2.7 1.7 2.9 3.7 ...
 $ gd17     : num 17.7 16.6 19.3 19.3 19.1 18.4 16.4 16.3 15.1 14.1 ...
 $ middelsky : num 6.3 7.7 7.3 8 8 8 8 8 8 7 ...
 $ sol      : num 0 0 0.6 0 0 0 0 0 0 1.8 ...
 $ middelp  : num 1023 1026 1024 1019 1001 ...
 $ nedbor   : num NA 1.1 1.3 0.9 0 18.4 1 4.2 0.9 6.7 ...
 $ spesRH   : num 2.7 3.2 2.5 2.4 2.8 3.1 3.7 3.7 4 4.1 ...
 $ middelRH : int 75 84 80 75 85 91 92 93 92 86 ...

> # Bestemme resposvariabel (Y) altså hvilket av forbrukene som skal analyseres
> # Responsvariablen Y = 'zsum', dvs summen av alle forbrukene.
>
> # .... [TRUNCATED]

> # H0: zsum er ikke ulik en normalfordeling
> # H1: zsum er ulik en normalfordeling
>
> shapiro.test(zsum)

      Shapiro-Wilk normality test

data:  zsum
W = 0.9561, p-value = 8.829e-16

> # p << 0.05 - liten støtte for H0. zsum er ulik en normalfordeling
>
> # Er forbruket avhengig av helg eller hverdag?
> weekend=subset(dataset, hel .... [TRUNCATED]

> work=subset(dataset, hel=='N')
```

```

> boxplot(zsum~helg, main='Energiforbruk, hverdag vs. helg', col='lightskyblue1',
xlab='Helg? Ja/Nei', ylab='Observert energiforbruk, næring [kWh]', ran ....
[TRUNCATED]

> # normalfordelt?
> shapiro.test(weekend$zsum) # p << 0.05 - ikke normalfordelt

      Shapiro-Wilk normality test

data:  weekend$zsum
W = 0.9472, p-value = 5.234e-08

> shapiro.test(work$zsum) # p << 0.05 - ikke normalfordelt

      Shapiro-Wilk normality test

data:  work$zsum
W = 0.9594, p-value = 2.446e-12

> # Wilcoxon's test for differences:
> # H0: Totalforbruk varierer ikke med dagkategori
> # H1: Totalforbruk varierer med dagkategori
> wilcox.test(zsu .... [TRUNCATED]

      Wilcoxon rank sum test with continuity correction

data:  zsum by helg
W = 83658, p-value = 0.7978
alternative hypothesis: true location shift is not equal to 0

> # p = 0.7978 > 0.05 - Støtte for H0 - forkast H1. Totalforbruk varierer ikke med
dagkategori
> par(mfrow=c(1,2))

> hist(weekend$zsum)

> hist(work$zsum)

> par(mfrow=c(1,1))

> cor.test(zsum,sol, method='spearman')

      Spearman's rank correlation rho

data:  zsum and sol
S = 48706821, p-value = 2.948e-15
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.3131803

> # Observerer ganske svak negativ korrelasjon med forbruket - dvs soltimer opp =
forbruk ned, som forventet.
> # p << 0.05 - korrelasjonskoeffisient .... [TRUNCATED]

```

Spearman's rank correlation rho

```
data: zsum and nedbor
S = 24520152, p-value = 0.0001323
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
-0.1699383
```

```
> # Observerer svært svak negativ korrelasjon med forbruket - økt nedbørmengde =
redusert forbruk.
> # p < 0.05 - korrelasjonskoeffisient er statist .... [TRUNCATED]

> east=subset(dataset, vindretn=='E')

> south=subset(dataset, vindretn=='S')

> west=subset(dataset, vindretn=='W')

> boxplot(zsum~vindretn, main='Energiforbruk fordelt på Vindretning',
col='darkgoldenrod1', xlab='Vindretning', ylab='Observert energiforbruk, næring [k
.... [TRUNCATED]

> # Sjekke normalitet:
> shapiro.test(north$zsum) #nei
```

Shapiro-Wilk normality test

```
data: north$zsum
W = 0.9648, p-value = 4.968e-06
```

```
> shapiro.test(east$zsum) #nei
```

Shapiro-Wilk normality test

```
data: east$zsum
W = 0.9637, p-value = 0.0002369
```

```
> shapiro.test(south$zsum) #nei
```

Shapiro-Wilk normality test

```
data: south$zsum
W = 0.9485, p-value = 9.433e-10
```

```
> shapiro.test(west$zsum) #nei
```

Shapiro-Wilk normality test

```
data: west$zsum
W = 0.9353, p-value = 2.244e-05
```

```
> kruskal.test(zsum~vindretn) # p << 0.05, Liten støtte for H0 - dvs forbruket
varierer med vindretning.
```

Kruskal-Wallis rank sum test

```
data: zsum by vindretn
Kruskal-Wallis chi-squared = 86.4627, df = 3, p-value < 2.2e-16
```

```
> #####
> # MULTIPPEL LINEÆRREGRESJON!
> ##### .... [TRUNCATED]
```

```
> summary(mlr1)
```

Call:

```
lm(formula = zsum ~ middelvind + gd17 + sol + middelp + nedbor +
    middelRH)
```

Residuals:

Min	1Q	Median	3Q	Max
-49822	-8547	-244	9801	39511

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-279261.69	64016.49	-4.362	1.70e-05	***
middelvind	1394.54	601.76	2.317	0.0211	*
gd17	5809.71	145.84	39.836	< 2e-16	***
sol	-844.50	338.47	-2.495	0.0131	*
middelp	267.19	62.60	4.268	2.55e-05	***
nedbor	-70.56	147.46	-0.479	0.6326	
middelRH	-31.65	103.72	-0.305	0.7605	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14120 on 344 degrees of freedom

(550 observations deleted due to missingness)

Multiple R-squared: 0.8445, Adjusted R-squared: 0.8418

F-statistic: 311.4 on 6 and 344 DF, p-value: < 2.2e-16

```
> # nedbor og middelRH ikke signifikant. fjerner fra regresjonen
```

```
> mlr2=lm(zsum~middelvind+gd17+sol+middelp)
```

```
> summary(mlr2)
```

Call:

```
lm(formula = zsum ~ middelvind + gd17 + sol + middelp)
```

Residuals:

Min	1Q	Median	3Q	Max
-54936	-8015	-130	9244	40789

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-245376.04	47772.45	-5.136	3.79e-07	***
middelvind	1623.18	463.19	3.504	0.000492	***
gd17	5630.15	106.27	52.980	< 2e-16	***

```
sol          -514.96      150.08  -3.431 0.000642 ***
middelp      230.07       47.65   4.828 1.75e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 14270 on 601 degrees of freedom
(295 observations deleted due to missingness)
Multiple R-squared:  0.8469,    Adjusted R-squared:  0.8459
F-statistic: 831.4 on 4 and 601 DF,  p-value: < 2.2e-16
```

```
> #setter inn koeffisienter og lager vektor for y-funksjon
> yfunk=c(-245376.04+(5630.15*gd17)+(1623.18*middelvind)-
(514.96*sol)+230.07*middelp)
```

```
> plot(yfunk,zsum, main='Observert energiforbruk vs. Q estimert', ylab='Observert
energiforbruk, næring [kWh]', xlab='Verdi av Q esti .... [TRUNCATED]
```

```
> lmfinal=lm(zsum~yfunk)
```

```
> summary(lmfinal)
```

```
Call:
lm(formula = zsum ~ yfunk)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-54936  -8015   -130    9244   40789
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.8379   1371.9778  -0.002    0.998
yfunk         1.0000     0.0173   57.810 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 14240 on 604 degrees of freedom
(295 observations deleted due to missingness)
Multiple R-squared:  0.8469,    Adjusted R-squared:  0.8467
F-statistic: 3342 on 1 and 604 DF,  p-value: < 2.2e-16
```

```
> abline(lmfinal, col='red',lwd=3)
```

```
> cor.test(zsum,yfunk, method='spearman')
```

```
Spearman's rank correlation rho
```

```
data:  zsum and yfunk
S = 3180866, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.914241
```

```
> #####
> # ENKEL LINEÆRREGRESJON
```

```

> #####
>
> .... [TRUNCATED]

> summary(mlr3)

Call:
lm(formula = zsum ~ gd17)

Residuals:
    Min       1Q   Median       3Q      Max
-68503  -9305   1166  10414  37164

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -14038.85    1325.68  -10.59  <2e-16 ***
gd17          5909.40      79.93   73.93  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14950 on 899 degrees of freedom
Multiple R-squared:  0.8588,    Adjusted R-squared:  0.8586
F-statistic: 5466 on 1 and 899 DF,  p-value: < 2.2e-16

> alt.yfunk=c(-14038.85+5909.4*gd17)

> plot(alt.yfunk,zsum, main='Observervert energiforbruk vs. Q estimert,næring',
xlab='Kalkulert verdi av Q estimert pr. o .... [TRUNCATED]

> abline(lm(zsum~alt.yfunk), col='green', lwd=3)

> cor.test(zsum,alt.yfunk, method='spearman')

Spearman's rank correlation rho

data:  zsum and alt.yfunk
S = 8993880, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
rho
0.9262224

> #####
> # ikke-LINEÆR REGRESJON
> #####
>
> .... [TRUNCATED]

> alt.yfunk2=lm(zsum~I(gd17)+I(gd17^2)+I(gd17^3))

> summary(alt.yfunk2)

Call:
lm(formula = zsum ~ I(gd17) + I(gd17^2) + I(gd17^3))

Residuals:

```

Min	1Q	Median	3Q	Max
-71513	-8436	788	10129	36086

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	13249.587	3902.683	3.395	0.000716 ***
I(gd17)	451.922	822.281	0.550	0.582734
I(gd17^2)	304.066	53.120	5.724	1.42e-08 ***
I(gd17^3)	-4.917	1.040	-4.729	2.62e-06 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14490 on 897 degrees of freedom
Multiple R-squared: 0.8677, Adjusted R-squared: 0.8672
F-statistic: 1961 on 3 and 897 DF, p-value: < 2.2e-16

```
> alt.yfunk3=lm(zsum~I(gd17^2)+I(gd17^3))
```

```
> summary(alt.yfunk3)
```

Call:

```
lm(formula = zsum ~ I(gd17^2) + I(gd17^3))
```

Residuals:

Min	1Q	Median	3Q	Max
-71772	-8462	871	10132	35905

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	15274.8016	1284.9226	11.89	<2e-16 ***
I(gd17^2)	332.5384	11.7391	28.33	<2e-16 ***
I(gd17^3)	-5.4460	0.3919	-13.89	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14480 on 898 degrees of freedom
Multiple R-squared: 0.8676, Adjusted R-squared: 0.8673
F-statistic: 2943 on 2 and 898 DF, p-value: < 2.2e-16

```
> kurve1=curve(15274.8016+322.5384*x^2-5.4460*x^3, add=T, col='red', lwd=3)
```

```
> abline(lm(zsum~gd17), col='green', lwd=3)
```

```
> cub.est=c(15274.8016+322.5384*gd17^2-5.4460*gd17^3)
```

```
> cor.test(zsum,cub.est, method='spearman')
```

Spearman's rank correlation rho

data: zsum and cub.est

S = 8993880, p-value < 2.2e-16

alternative hypothesis: true rho is not equal to 0

sample estimates:

rho
0.9262224

Vedlegg 7 Fotokopi av tabeller fra Kristensen, Martin S sin masteroppgave

TABELL 9: FJERNVARME BRUKT TIL OPPVARMING AV TAPPEVANN, NÆRINGSKUNDER (TV2)

År	Energi til tappevann, kWh	Total energibruk, kWh	Andel
2008	28,861,348	178,195,988	16.2 %
2009	29,543,569	208,542,107	14.2 %
2010	30,424,773	266,171,055	11.4 %
2011	29,529,815	187,942,162	15.7 %
2012	27,791,890	192,114,708	14.5 %
2013	26,399,712	188,959,664	14.0 %
Alle år	172,551,107	1,221,925,684	14.1 %

TABELL 7: FJERNVARME BRUKT TIL OPPVARMING AV TAPPEVANN, BOLIGKUNDER (TV2)

År	Energi til tappevann, kWh	Total energibruk, kWh	Andel
2008	19,583,279	54,866,408	35.7 %
2009	22,150,639	62,329,982	35.5 %
2010	22,070,502	70,139,749	31.5 %
2011	20,761,289	59,276,576	35.0 %
2012	20,500,513	60,555,853	33.9 %
2013	21,897,925	61,570,740	35.6 %
Alle år	126,964,147	368,739,309	34.4 %



Norges miljø- og
biovitenskapelige
universitet

Postboks 5003
NO-1432 Ås
67 23 00 00
www.nmbu.no