

Kjetil Sander presenterer:

**e**studie.no

# Dataanalyse



**Hva er en dataanalyse og hvordan  
gå frem for å gjennomføre en  
dataanalyse av det innsamlede  
datagrunnlaget fra en  
feltundersøkelse?**

Skrevet av:  
**Kjetil Sander**

Utgitt av:  
**eStudie.no**

Revisjon:  
**1.0 (Sept. 2017)**

# Innhold

<b>1 Dataanalyse</b>	<b>6</b>
<b>1.1 Hva er en dataanalyse?</b>	<b>6</b>
<b>1.2 Faktorer som avgjør analysearbeidet</b>	<b>6</b>
1.2.1 Datainnsamlingmetode	7
1.2.2 Utvalget	7
1.2.3 Målinger	8
1.2.4 Andeler og tellinger	8
1.2.5 Samvariasjon	8
1.2.6 Variablene	8
<b>1.3 Frekventativ statistikk</b>	<b>9</b>
1.3.1 Bayesisk statistikk	11
<b>1.4 Arbeidsmodell for dataanalysen</b>	<b>12</b>
1.4.1 Koding og registrering	13
1.4.2 Tabulering	15
1.4.3 Sammenhenger	15
1.4.4 Statistisk inferens	15
1.4.5 Tolkning og kontroll	15
1.4.6 Kontroll/feilkilder	16
<b>1.5 Den analytiske vitenskapelige metode</b>	<b>18</b>
<b>1.6 Kategorisering, koding og indekseringsspråk</b>	<b>19</b>
1.6.1 Slektskap (generiske forhold)	20
1.6.2 Partitive forhold	20
1.6.3 Kategorier	20
<b>1.7 Inndelingskriterier for dataanalysen</b>	<b>21</b>
<b>1.8 Antall kategorier</b>	<b>23</b>
<b>1.9 Protokoll, ordningsobservatorer og notasjon</b>	<b>24</b>
1.9.1 Ordnet rekkefølge	24
<b>1.10 Synkron- eller diakron analyse?</b>	<b>24</b>
1.10.1 Hva er forskjellen mellom en synkron- og diakron analyse?	25
1.10.2 Synkron analyse	25
1.10.3 Diakron analyse	25
1.10.4 Start med en diakron analyse	26
<b>1.11 Prinsipper for situasjonsanalysen</b>	<b>26</b>
1.11.1 Avslutt med en synkron analyse for å punktbelyse de kritiske suksessfaktorene	27
<b>1.12 Analyseteknikker for dataanalysen</b>	<b>28</b>
<b>2 Univariat analyse og deskriptiv statistikk</b>	<b>29</b>
<b>2.1 Populasjonens størrelse</b>	<b>30</b>
<b>2.2 Utvalgets størrelse</b>	<b>31</b>
<b>2.3 Fordelingsanalyse</b>	<b>31</b>
<b>2.4 Frekvensfordeling</b>	<b>32</b>
<b>2.5 Datamatiser</b>	<b>34</b>
<b>2.6 Grafiske fremstillinger</b>	<b>34</b>
<b>2.7 Utvalget/stikkprøvens midtpunkt (Median, gjennomsnitt og modus)</b>	<b>35</b>
2.7.1 Median	35
2.7.2 Gjennomsnitt	37

2.7.3	Modus (typetall) .....	39
2.7.4	Forholdet mellom median, snitt og typetall.....	41
<b>2.8</b>	<b>Variasjonsbredde i utvalget (stikkprøven) .....</b>	<b>42</b>
2.8.1	Avkortet spredning og interkvartil variasjonsbredde.....	44
2.8.2	Interkvartil variasjonsbredde.....	45
2.8.3	Modalprosent.....	45
<b>2.9</b>	<b>Prosentil og kvartil .....</b>	<b>45</b>
2.9.1	Nedre kvartil .....	47
2.9.2	Øvre kvartil .....	48
2.9.3	Kvartilbredde .....	49
2.9.4	Kvartilavvik.....	49
2.9.5	Videoforklaring.....	49
2.9.6	Eksempler fra Wikipedia.....	49
<b>2.10</b>	<b>Standardavvik og varians .....</b>	<b>50</b>
2.10.1	Formler .....	52
2.10.2	Utvalgsvariens .....	54
<b>2.11</b>	<b>Populasjonsvariens.....</b>	<b>54</b>
2.11.1	Standardavviket for utvalget/stikkprøven beregnes slik:.....	55
<b>2.12</b>	<b>Skjevhet og normalfordeling .....</b>	<b>56</b>
2.12.1	Høyreskjev fordeling .....	57
2.12.2	Symetrisk fordeling .....	58
2.12.3	Venstreskjev fordeling.....	58
2.12.4	Sentralgrenseteoremet – normalfordelt .....	59
2.12.5	Standard normal distribusjon .....	61
<b>3</b>	<b>Bivariat analyse.....</b>	<b>62</b>
<b>3.1</b>	<b>Krysstabulering og marginalfordeling.....</b>	<b>63</b>
<b>3.2</b>	<b>Epsilon - et mål for samvariasjonens styrke og retning .....</b>	<b>64</b>
<b>3.3</b>	<b>Parallell plott og trend.....</b>	<b>65</b>
3.3.1	Parallele trender .....	66
3.3.2	Analoge trender .....	66
3.3.3	Avhengige trender .....	66
3.3.4	Serie – trender .....	67
3.3.5	Kryssende trender.....	67
3.3.6	Motstridende eller sprikende trender .....	67
3.3.7	Resultattrender.....	67
3.3.8	Hierarkiske trender – Sumtrender .....	68
<b>3.4</b>	<b>Trender og trendforlengelse (kurvefremskrivning) .....</b>	<b>68</b>
3.4.1	Metode for trendforlengelse og kurvefremskrivning.....	69
3.4.2	Beregninger og trendforlengelse .....	73
<b>3.5</b>	<b>Gruppegjennomsnitt .....</b>	<b>74</b>
<b>3.6</b>	<b>Mediandifferanse .....</b>	<b>75</b>
<b>3.7</b>	<b>Mann - Whitney observator.....</b>	<b>76</b>
3.7.1	Sammenligning av to medianer .....	78
<b>3.8</b>	<b>Parvise sammenligninger .....</b>	<b>80</b>
3.8.1	Fortegnstest .....	81
3.8.2	Wilcoxon's tegnrangetest .....	83
<b>3.9</b>	<b>Korrelasjon og kovarians.....</b>	<b>87</b>
3.9.1	Kovarians .....	87
3.9.2	Teoretisk kovarians .....	87
3.9.3	Empirisk kovarians.....	88

3.9.4	Korrelasjon .....	89
<b>4</b>	<b>Multivariat analyse.....</b>	<b>92</b>
<b>4.1</b>	<b>Multivariate metoder.....</b>	<b>96</b>
4.1.1	Multippel regresjonsanalyse.....	96
4.1.2	Conjoint analyse .....	96
4.1.3	Faktoranalyse.....	96
4.1.4	Clusteranalyse.....	97
<b>4.2</b>	<b>Multippel regresjon.....</b>	<b>98</b>
<b>4.3</b>	<b>Diskriminantanalyse.....</b>	<b>101</b>
<b>4.4</b>	<b>Clusteranalyse .....</b>	<b>102</b>
<b>4.5</b>	<b>Faktoranalyse .....</b>	<b>103</b>
4.5.1	Eksplorerende og konfirmerende faktoranalyse.....	104
4.5.2	Egenverdi .....	105
4.5.3	Kommunalitet .....	105
4.5.4	Ortogonal og oblikk rotasjon.....	105
<b>5</b>	<b>Sannsynlighet .....</b>	<b>106</b>
5.1.1	Subjektiv sannsynlighet.....	107
5.1.2	Sannsynlighet som populasjonandel.....	107
<b>5.2</b>	<b>Utfall og utfallsrom.....</b>	<b>109</b>
<b>5.3</b>	<b>Stokastisk forsøk.....</b>	<b>109</b>
<b>5.4</b>	<b>Utfallsrom.....</b>	<b>109</b>
5.4.1	Enkeltutfall .....	109
5.4.2	Hendelse .....	109
5.4.3	Diskrete og kontinuerlig utfall.....	110
5.4.4	Stokastiske variabler.....	110
5.4.5	Begivenhet .....	111
<b>5.4.6</b>	<b>Gunstige- og antall mulige utfall.....</b>	<b>111</b>
<b>5.5</b>	<b>Sannsynlighet som relativ hyppighet (de store talls lov).....</b>	<b>114</b>
<b>5.6</b>	<b>Union .....</b>	<b>115</b>
<b>5.7</b>	<b>Uniform sannsynlighet.....</b>	<b>115</b>
<b>5.8</b>	<b>Sannsynligheten av et snitt .....</b>	<b>116</b>
<b>5.9</b>	<b>Venndiagram.....</b>	<b>117</b>
5.9.1	Krysstabell .....	118
5.9.2	Komplementære hendelser .....	118
5.9.3	Union .....	119
5.9.4	Snitt.....	120
5.9.5	Disjunkte hendelser .....	120
5.9.6	Oppsummering.....	121
<b>5.10</b>	<b>Betinget sannsynlighet .....</b>	<b>121</b>
<b>5.11</b>	<b>Kombinatorikk .....</b>	<b>126</b>
<b>5.12</b>	<b>Permutasjoner og kombinasjoner.....</b>	<b>126</b>
5.12.1	Permutasjoner med repetisjon (tilbakelegging) .....	128
5.12.2	Permutasjoner uten repetisjon (tilbakelegging) .....	128
5.12.3	Kombinasjoner uten repetisjon.....	129
5.12.4	Kombinasjoner med repetisjon .....	130
5.12.5	Enumerativ kombinatorikk .....	131
<b>5.13</b>	<b>Aksiomatisk definisjon av sannsynlighet.....</b>	<b>133</b>
<b>6</b>	<b>Hypoteseprøving.....</b>	<b>134</b>
<b>6.1</b>	<b>Teori/problemstilling .....</b>	<b>136</b>
<b>6.2</b>	<b>Operasjonalisering .....</b>	<b>136</b>

<b>6.3 Signifikansnivå .....</b>	<b>136</b>
<b>6.4 Metode for hypoteseprøving .....</b>	<b>137</b>
<b>7 Generalisering.....</b>	<b>139</b>
<b>7.1 Tilfeldig variasjon .....</b>	<b>140</b>
<b>7.2 Estimering .....</b>	<b>141</b>
<b>7.3 Konfidensintervall .....</b>	<b>142</b>
7.3.1 Feilsannsynlighet.....	144
7.3.2 Konfidensnivå .....	144
<b>7.4 Trangere konfidensintervall.....</b>	<b>145</b>
7.4.1 Tosidig konfidensintervall.....	145
7.4.2 Ensidige konfidensintervaller .....	149
7.4.3 Venstre grense intervall.....	150
7.4.4 Høyre grense intervall .....	150

# 1 Dataanalyse

Når dataene er samlet inn, kodet og organisert kan vi begynne dataanalysen som skal lede frem til at vi finner svaret på vår problemstilling.

## 1.1 Hva er en dataanalyse?

Dataanalyse kan defineres som:

**"Systematisk bearbeiding av innsamlet informasjon"**

Dataanalysens formål er å kategorisere de innsamlede dataene slik at vi kan beskrive og analysere det vi har funnet.

- **Hva karakteriserer dataene?**
- **Finnes det mønstre og/eller relasjoner?**
- **Kan sammenhengen mellom variabler skyldes mellomliggende årsaksforhold?**

Dataanalysens mål er å forenkle datamengden slik at det blir mulig å besvare undersøkelsens problemstilling ved å se på det innsamlede datamaterialets bakenforliggende mønstre og relasjoner.

## 1.2 Faktorer som avgjør analysearbeidet

Den viktigste enkeltfaktoren som avgjør analysevalget er **problemstillingen**. Dette fordi analysens formål er å løse problemstillingen forskningsprosjektet bygger på. Deretter vil faktorer som **tid, penger, kunnskap og ressurser** styre analysevalget i stor grad.

Etter at vi har tatt hensyn til overstående faktorer, vil de innsamlede dataene i seg selv avgjøre hvilke analyser vi kan gjøre. Når det gjelder det innsamlede datamaterialet vil **datainnsamlingmetoden, utvalget** og variabelens **egenskaper** og **målenivå** vil være de største begrensningen for hva slags analyse som kan foretas.