

2 Oversigt over kapitel 2

2.1 Begreber

Discussion questions 1, 2 og 3.

Bemærk, at for en **quantitative** (kvantitativ) variabel har det mening at udregne numeriske forskelle mellem værdier af den variable (fx som differenser, eller som forhold mellem to værdier), hvorimod en **categorical** (kategorisk) variabel blot udtrykker resultatet af en *klassifikation* (fx mand/kvinde), også selv om klasserne eventuelt er repræsenteret ved talværdier (mand = 0, kvinde = 1).

I datavinduet i SAS-insight viser rubrikken over variabelnavnet i vinduets øverste linie se, om den pågældende variabel fortolkes som kvantitativ (Int, dvs intervalskala), eller kategorisk (Nom, dvs nominalskala). Man kan ændre fortolkningen af den variable ved at klikke på rubrikken og markere den ønskede fortolkning. Hvis værdierne af en variabel er indlæst som tekststreng, bliver de automatisk fortolket som Nom.

2.2 Grafisk repræsentation af fordelinger af data

Discussion questions: 4 og 5.

bar chart: søjlediagram

Bemærk, på engelsk opfattes *frequency* sædvanligvis som *absolut hyppighed*, dvs et *antal*. Hvis man på engelsk mener *relativ hyppighed* (en *andel*) vil man eksplicit sige *relative frequency*. På dansk tager man det ikke altid så nøje, og det fremgår forhåbentlig af sammenhængen, om man mener absolut eller relativ hyppighed.

2.3 Sammenfatning af kvantitative data

Discussion question 6

mode *modus*

2.4 Outliere

Discussion questions 7 og 8

Begrebet five-number-summary af fordelinger af data er endnu ikke særligt udbredt i Danmark. Men Box-and-Whisker plottet benyttes en del.

2.5 Outlierresistente mål

Discussion questions 9, 10, 11, 12

resistant: resistant, eller outlier-resistant. Undertiden ser man en mere generel betegnelse *robust* (side 270) brugt i betydningen outlierresistens. Det anbefales dog at bruge ordet resistant, eller outlier-resistant.