

ANDRE FUNKTIONER TIL DATAANALYSE

FREQUENCY()/FREVENS()

FREKVENS() er en lidt speciel funktion, da den altid skal anvendes i form af en såkaldt matrixformel. Matrixformler er karakteriseret ved, de udfører beregninger i mere end en celle adgangen. Se nedenfor om oprettelsen af sådanne.

FREKVENS() anvendes til at "tælle" hvor mange værdier, der ligger inden for bestemte intervaller. Funktionen kræver dels at der et område med værdier, dels at der defineres yderligere et område med intervalgrænser. Antag eksempelvis, at man vil tælle hvor mange tal i P-kolonnen, der ligger i hvert af intervallerne 0-3, 4-6, 7-10, 11-20 og over 20. Dette kunne gøres ved hjælp af et antal TÆL.HVIS() formler, men det går noget nemmere med FREKVENS().

3	U5:U8
6	
10	
20	

Opret først et område, fx i en tom kolonne (her U5 til U8), hvor du kan indtaste intervallerne. Der skal være plads til at indsætte formler i kolonnen ved siden af. Det er, som det fremgår af illustrationen til venstre intervallerne øvre grænser, der skal indtastes. Marker dernæst de celler, der ligger umiddelbart til højre for intervallerne. Der skal markers én celle mere, end der er intervaller, se figuren til venstre.

Mens området er markeret, indtastes nu følgende formel: =FREKVENS(P2:P26929;U5:U8). Tryk nu Ctrl+Shift+Enter samtidigt*. Så oprettes matrixformlen. At der er tale om en matrixformel, kan ses ved at klikke i en enkelt af cellerne efter oprettelsen, så ses, at formelen er "indpakket" i tuborgparenteser: {=FREKVENS(P2:P26929;U5:U8)}. Denne formel indsættes i hver af de markerede celler. Det første område er det, der skal tælles, det andet indeholder intervalgrænserne.

3	14138
6	4842
10	3023
20	3182
	1743

I figuren her, ses resultatet af beregningerne. De skal læses som følger: 14.138 linier har værdier på 3 eller derunder. 4.842 har værdier mellem 4 og 6 og så fremdeles. De 1.743 i den sidste række repræsenterer de rækker, hvor værdierne er over 20.

* NB! Det er IKKE muligt at konstruere en matrixformel ved at rette tallene selv.

AVEDEV()(/MAD())

Det er nemt at beregne gennemsnittet i Excel. Imidlertid er gennemsnittet ikke altid et godt mål for resultatet af en prøve. Forestil jer, at en undersøgelse viser at værdierne er enten 1 (75% af forekomsterne) eller 21 (25% af forekomsterne). Dette vil give et gennemsnit på 6, men dette gennemsnit er ikke et godt mål for, hvad de reelle værdier er. Funktionen MAD() beregner den gennemsnitlige afvigelse fra gennemsnittet! I ovenstående eksempel vil MAD() blive beregnet til 7,5. Det betyder, at gennemsnittet er 6, men at de reelle værdier i gennemsnit ligger 7,5 over eller under dette gennemsnit. Da ingen værdier er under 0, er heller ikke dette mål særlig godt, men behøver det heller ikke at være, da formålet med det, alene er at fortælle om gennemsnittet er et godt mål. Jo højere resultatet af MAD() er, jo dårligere et mål er gennemsnittet.

QUARTILE()/KVARTIL()

KVARTIL() bruges til at opdele data i kvartiler. Syntaksen er KVARTIL(område;kvartil). Der findes i alt fem forskellige værdier, der kan anvendes for kvartil. 0, 1, 2, 3 og 4.

- 0 giver minimumsværdien,
- 1 giver 1. kvartilværdien,
- 2 giver medianværdien,
- 3 giver 3. kvartilværdien
- 4 giver maksimumsværdien.

Anvendes denne funktion kolonne P i det tænkte eksempel, giver det følgende resultat. Første, 3. og 5 værdi kunne også være fundet med henholdsvis MIN(), MEDIAN() og MAKS() funktionerne. De anvendte formler er =KVARTIL(P2:P26929;0) ud for Minimum, og så videre med de øvrige kvartilværdier.

Minimum	0
1. kvartil	1
Median	3
3. kvartil	8
Maksimum	232

MEDIAN()/MEDIAN()

Denne funktion hedder det samme, både på dansk og engelsk. Den returnerer det midterste tal i en talrække. Det er altså ikke det samme som gennemsnittet. Har man fx en opstilling med fem 1-taller, et "12-tal" og igen fem 1-taller, er gennemsnittet 2, mens medianværdien er 12. Er talrækken sorteret vil der være lige mange værdier, som er større og mindre end medianværdien.

RANK()/PLADS() og PERCENTRANK()/PROCENTPLADS()

PLADS() viser placering af et element på en liste. Syntaksen er PLADS(Tal;Område;Sortering). Har man fx værdierne 1, 1, 2, 4, 5, 7 i celleområdet A1:A6 vil =PLADS(A5;A1:A6;1) give 2, fordi 5 er det næsthøjeste tal. Ved at ændre 1 til 0 i sidste argument vendes rækkefølgen, og resultatet bliver så 5, fordi 5 er det 5.-mindste tal. Har flere tal samme værdi, får de også samme plads.

Vil man fx vide, hvilken "placering" celle P7 har blandt alle tal i kolonne P, kan man skrive =PLADS(P7;P2:P26929;0), og svaret vil være 3.575.

PROCENTPLADS() giver placeringen i procent af datasættet. Syntaksen er lidt anderledes end PLADS(), nemlig PROCENTPLADS(område;element). =PROCENTPLADS(P2:P26929;P7) vil fx give resultatet 0,853, hvilket indikerer at ca. 15% af tallene er større end tallet i P7, mens ca. 85% er mindre.

S	T
	p23
Procentplads	77%

Funktionen kan anvendes til at vurdere om et givent element er i den gode eller den dårlige del af en population. Man skal bare huske selv at bestemme, hvad der er godt og hvad, der er dårligt. Ønsker man at kunne teste for forskellige værdier, kan man ændre formlen til fx =PROCENTPLADS(P2:P26929;INDIREKTE(T1)). Så kan man i celle T1 indtaste adressen på den celle, man vil vide noget om, fx P7. Se eksemplet ovenfor.

TRIMMEAN()/TRIMMIDDELVÆRDI?()

Har man koncentreret sine data omkring gennemsnittet kan nogle få afvigelser i yderpunkterne af værdiserien få betydning for beregning af gennemsnittet. Dette kan TRIMMIDDELVÆRDI() rette op på. Forestil dig, at du har følgende data: 3,4,5,3,4,5,3,4,5. MIDDEL() af disse værdier vil være 4. Tilføjer vi nu -34 til datasættet vil gennemsnittet ændre sig til 0,2, og havde vi tilføjet 34 i stedet, ville snittet være steget til 7. Dette skyldes at en sådan ekstrem afvigelse "vejer meget tungt" i forhold til resten. Ved at skære noget af vægten fra, bliver gennemsnittet mere realistisk, og det er netop det, som TRIMMIDDELVÆRDI() gør.

TRIMMIDDELVÆRDI(A1:A10;0,2) betyder, at vi vil beregne gennemsnittet, men vi vil skære 20% af i top og bund. I ovenstående tilfælde (-34) betyder det, at snittet ville blive beregnet til 3,875 og ved +34 til 4,125, hvilket ligger betydeligt tættere på gennemsnittet uden "afvigelse". Skærer vi 80 % af, vil tallet i dette tilfælde blive 4 – som om 34 slet ikke regnes med.