

1. Inleiding. Het belangrijkste bij het gebruik van statistische software is, dat je weet wat een statistisch begrip of toets betekent en hoe hierbij de gegevens gebruikt worden. Via de menu's wijst de rest zich dan vanzelf.

Excel is een "spreadsheet"-programma met veel ruimere mogelijkheden dan de statistische het verwerking van gegevens. In dit overzichtje beperken we ons echter hiertoe. Als je statistische functies in Excel wilt gebruiken moet je in het **Tools**-menu het veld **Data Analysis** ... aanklikken (of eventueel via **Add-Ins** het **Analysis ToolPak** kiezen). Via dit veld kun je de gebruikelijk statistische toetsen selecteren. In het kader van de cursus "Begrippen van Kansrekening en Statistiek" zijn de volgende tools van belang:

Correlation & Covariance	Random Number Generation
Descriptive Statistics	Regression
F-Test Two-Sample for Variances	t-Test: Paired Two-Sample for Means
Histogram	t-Test: Two-Sample Assuming Equal Variances

2. Het invoeren van data. Bij het opstarten van Excel verschijnt er op het scherm een **werkblad (workbook of worksheet)** met cellen genummerd met een letter voor de kolom en een nummer voor de rij. Iedere cel is vanuit ieder andere cel uniek adresseerbaar met zijn kolomletter en rijnummer. Bij het tussenvoegen van kolommen of rijen worden alle referenties automatisch aangepast. In een cel kan tekst, een getal of een formule geplaatst worden, door de cel aan te klikken, de gewenste symbolen in te typen en af te sluiten met een **Return**. Tijdens het typen verschijnt een copie van de celinhoud in de bovenbalk; als je de inhoud wilt wijzigen, moet je eerst de goede plaats in deze tekst op de bovenbalk aanklikken, dan de wijzigingen intypen en afsluiten via return of via het aanklikken van het $\sqrt{\quad}$ -symbooltje naast deze verbeterde tekst. Via de **Cells** ...-optie in het **Format**-menu kan de vorm (aantal decimalen – centering – font – size etc.) naar behoeven worden aangepast.

FORMULES. Een formule in een cel begint altijd met het =-teken. Na het intypen van een correcte formule wordt door excel onmiddellijk het resultaat berekend en neergeschreven in de cel op je werkblad. Dat er in een cel een formule staat is alleen nog te zien door de cel aan te klikken en in de bovenbalk naar de werkelijke inhoud te kijken. Voorbeelden

=AVERAGE(A1:B5)	bereken het gemiddelde van de genoemde 10 cellen
=VAR(A1:B5)	bereken de steekproefvariantie van de genoemde 10 cellen
=SUMPRODUCT(A1:A5,B1:B5)	bereken het product van de overeenkomstige elementen en sommeer deze (= inproduct $\sum_{i=1}^5 A_i B_i$).

Een lijst van beschikbare functies is te vinden via de standaard **Excel Help**-procedure.

COPIËREN, WISSEN, VERPLAATSEN EN DUPLICEREN van celinhoud. In het **Edit**-menu zijn onder andere de volgende functies beschikbaar. Zij werken op een van te voren geselecteerde cel of reeks cellen:

naam	toetscombinatie	betekenis
delete	command-K	verwijder de geselecteerde cellen uit het werkblad
cut	command-X	zet selectie klaar voor verplaatsing naar elders in het werkblad
copy	command-C	zet selectie klaar voor copiëring naar elders in het werkblad
paste	command-C	voeg de selectie van cut of copy in op de gekozen plaats
fill	command-R	dupliceer de celinhoud naar geselecteerde cellen
	command-D	rechts (R), beneden (D), links of boven.

Bij het copiëren of dupliceren van een formule verschuiven ook de referenties naar andere

* T_EX-datum: May 15, 2001

cellen, tenzij de kolomletter en/of het rijnummer in de formule is vastgezet door het er een \$-teken voor te zetten. Als bijvoorbeeld in cel C1 de formule

$$=A1*B1/SUM(A1:A10)$$

staat en we doen een **fill-down** naar cel C2, dan komt hierin de formule

$$=A2*B2/SUM(A2:A11)$$

te staan. Alle referenties zijn één veld naar beneden opgeschoven. Als wel de referenties in de teller maar niet die in de noemer willen opschuiven, dus als we

$$=A_i*B_i/SUM(A1:A10) \text{ met } i = 1 \dots 10$$

in de cellen C1 \dots C10 willen hebben, moeten we in cel C1 de formule

$$=A1*B1/SUM(A\$1:A\$10)$$

intipen en vervolgens een fill-down doen naar de negen cellen eronder.

Analoog schuiven alle kolomreferenties in een formule op naar rechts bij een **fill-right**, tenzij er een \$-teken voor de kolomletter staat.

VRAAG. Stel dat je werkblad in de cellen A1 \dots A10 de klassemiddens en in B1 \dots B10 de frequenties van een serie metingen. Schrijf een Excel-functie die in een cel het klassengemiddelde uitrekent. Schrijf ook een functie in de cel C1, die zo is, dat als je een fill-down naar C2 \dots C10 doet, deze cellen de cumulatieve frequenties bevatten.

3. Beschrijvende Statistiek. Stel, dat de cellen A1:A65 je metingen bevatten. Selecteer in het **Tools/Data Analysis**-menu *descriptive statistics*, dan krijg je een menu, waarin je de plaats van je data (= input range) in het werkblad moet opgeven (hier dus A1:A65), en moet aangeven of er bovenaan je kolom een label staat, of je een betrouwbaarheidsinterval (of beter de halve lengte van het BTI) wilt hebben en met welke betrouwbaarheid (b.v. 95%) en waar de uitvoer moet worden neergeschreven (in het huidige werkblad, op een nieuwe pagina van datzelfde werkblad of in een nieuw werkblad). Het programma geeft dan de volgende grootheden:

Mean – Standard Error – Median – Mode – Standard Deviation

Sample Variance – Kurtosis – Skewness – Range

Minimum – Maximum – Sum – Count – Confidence Level(95.0%).

Om vervolgens een Boxplot te kunnen tekenen moet je in twee cellen met de functie

$$= \text{quartile}(< \text{range} >, 1) \text{ en } = \text{quartile}(< \text{range} >, 3),$$

het eerste resp. derde kwartiel uitrekenen.

HISTOGRAM: Selecteer in het **Tools/Data Analysis**-menu *histogram*. Excel vraagt dan een *input range* (hier dus A1:A65) en een facultatieve *bin range* en of de gespecificeerde datakolommen labels bevatten. Je kunt in je werkblad zelf een klassenindeling maken (in een kolom met de benedengrenzen van de klassen) en deze opgeven; anders verdeelt excel zelf het waardengebied in 10 klassen. Tenslotte, als je een *chart* vraagt, wordt naast de frequentietabel het histogram getekend.

4. F- en t-Toetsen. Selecteer in het **Tools/Data Analysis**-menu de gewenste toets en geef de input range op. Excel geeft dan de overschrijdskansen voor de een- en tweezijdige versies van de toets en de grens van de kritieke zone, behorende bij het gevraagde niveau (α). Een χ^2 -toets voor de variantie in één groep metingen (one group χ^2 -test) is niet beschikbaar.

5. Kruistabellen en de χ^2 -toets op een kansverdeling. Deze toetsen moet je doen met gebruik van de excel-functie *CHITEST(actual range, expected range)*. Deze functie berekent met formule (20.4) uit de syllabus de χ -waarde van je steekproef en de overschrijdskans die hierbij hoort. Je moet dus wel zelf de verwachte frequenties berekenen die behoren

bij de gepostuleerde theoretische verdeling. Als je b.v. de frequentietabel van je toetst tegen de normale verdeling, dan kun je uit de klassegrenzen met de functie *NORMDIST* de percentielen berekenen (zie Excel Help).

Je kunt *CHITEST* ook toepassen bij kruistabellen, als je zelf eerst de matrix van verwachte frequenties maakt. Voor een 2×2 kruistabel kun je b.v. het volgende maken

	A	B	C	D
1	$A1 = 300$	$B1 = 150$	$= (A1 + B1) * (A1 + A2) / SUM(A1 : B2)$??
2	$A2 = 80$	$B2 = 20$??	??

en dan in het veld A3 de functie “=*CHITEST*(??, ??)” plaatsen.

6. Het berekenen van statistische functies. Het DATA ANALYSIS pakket in Excel bevat een groot aantal statistische functies, te vinden via *Help* \rightarrow *index* \rightarrow *statistical analysis, functions*. Geloof echter niet voetstoots de beschrijving. Voor de functie *TDIST*(**x,n,k**) wordt opgegeven dat deze de kansen van de *t*-verdeling geeft: als $T \sim t_n$ zou dit voor $k=1$ de kans $P(T \leq x)$ geven. In feite geeft *TDIST*(**x,n,1**) de kans $P(T \geq x)$ en *TDIST*(**x,n,2**) de kans $P(|T| \geq x)$ voor $x \geq 0$. De derde variabele is het aantal staarten, $k = 1$ of $k = 2$.