Absolute und relative Häufigkeit

Beispiel 1 An einem Gelsenkirchener Gymnasium wählten in den vergangenen Jahren stets unterschiedlich viele Schüler den Leistungskurs Mathematik. Die Tabelle enthält die Daten einiger Jahre.

Jahr	2002	2003	2004	2005	2006	2007
Schüler in Stufe 12	104	96	110	127	116	98
Schüler mit Mathematik-Leistungskurs	32	33	43	50	58	62

a) Stellen Sie den Verlauf und einen möglicherweise erkennbaren Trend in der Entwicklung des Wahlverhaltens der Schüler graphisch dar.

b) Treffen Sie eine Vorhersage über die relative Häufigkeit der Schüleranzahl eines Mathematik-Leistungskurses in der Jahrgangsstufe 12 in den nächsten Jahren.

c) Wann wird die Anzahl der Schüler mit einem Mathematik-Leistungskurs in der Jahrgangsstufe 12 zum nächsten Mal auf 40 sinken? Diskutieren Sie die Glaubwürdigkeit dieser Vorhersage in der Praxis.

Beispiel 2 Es kann helfen, Häufigkeiten (meistens absolute Häufigkeiten) in Form von *Histogrammen* darzustellen.

Stellen Sie die Anzahl der Schüler der Jahrgangsstufe 12 aus Beispiel 1 in Abhängigkeit des Jahres dar.

Führen Sie diese graphische Auswertung unter Statistik durch.

Beispiel 3 Am 22. September 2013 fand die 19. hessische Landtagswahl statt. Es waren 4.392.213 Bürger wahlberechtigt. Davon gab es 1.176.007 Nichtwähler und 85.425 Wahlstimmen waren ungültig. Die folgende Tabelle gibt die Ergebnisse der Landesstimmenzählung an:

CDU	SPD	GRÜNE	DIE LINKE	FDP	AfD	PIRATEN
1.199.633	961.896	348.661	161.488	157.451	126.906	60.159

FREIE WÄHLER	NPD	Sonstige	Ungültig	Nichtwähler
38.433	33.433	42.721	85.425	1.176.007

a) Berechnen Sie von jeder Partei die Prozentsätze der wahlberechtigten Teilnehmer und der gültigen Wählerstimmen in Prozent. Stellen Sie dieses Ergebnis als Kreisdiagramm dar. Worauf ist zu achten?

> **Hinweis.** Die Kreisdiagramme können unter Tabellenkalkulation mithilfe von gezeichnet werden.

b) Sitze im Landtag werden nach der „Fünf-Prozent-Hürde“ erst bei 5 % oder mehr der gültigen Gesamtstimmenzahl vergeben. Daher hielten nur fünf Parteien Einzug in den Landtag mit 110 Plätzen. Die Berechnung der Sitzverteilung erfolgt prozentual auf der Basis der Gesamtstimmenzahl der in den Landtag ziehenden Parteien.

Berechnen Sie die hier beschriebene Platzverteilung und runden Sie die Ergebnisse ab. Summieren Sie die Platzzahlen. Was fällt auf? Diskutieren Sie über Ihre Ergebnisse und einen möglichen gerechten Ausgleich des Ergebnisses.

Erstmals in der Geschichte der Landtagswahlen in Hessen ergaben sich Überhang- und Ausgleichsmandate. Dadurch hat der Landtag nun 118 Mitglieder anstelle von bisher 110. Von den acht zusätzlichen Mandaten entfielen vier Überhangmandate auf die CDU, zwei Ausgleichsmandate auf die SPD und je ein Ausgleichsmandat auf FDP und Grüne.

Stellen Sie die Platzverteilung zusammen. Erstellen Sie ein Kreisdiagramm der Platzverteilung im Landtag.

Lösung von Beispiel 1. a) Öffnen Sie Statistik und geben Sie die Werte in eine Liste ein, vgl. Abbildung 1, links. (Beachten Sie die gewählten Bezeichnungen der Spalten. Diese müssen vor der Dateneingabe festgelegt werden.)

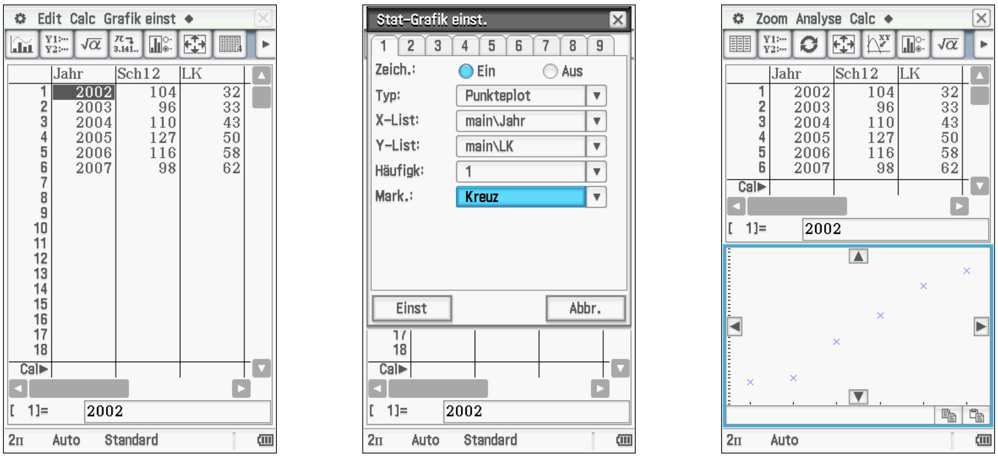

Abbildung 1: Erstellen eines Graphen zur absoluten Häufigkeit

Um die Daten der Tabelle graphisch auswerten zu können, wählen Sie , vgl. Mitte der Abbildung 1, und verwenden Sie sodann die sichtbaren Einstellungen. Es soll ein Graph aus Einzelpunkten gezeichnet werden, dessen Markierungen durch Kreuze dargestellt werden. Mit der Taste erhalten Sie einen Graphen, der dem rechten Graphen in Abbildung 1 ähnlich ist.

An dem Graphen ist erkennbar, dass die Anzahl der Schüler, die einen Mathematik-Leistungskurs gewählt haben, im Verlauf der Zeit gestiegen ist.

Andererseits hat sich die Gesamtzahl der Schülerinnen und Schüler der Jahrgangsstufen im Laufe der Zeit verändert. Interessant ist die Kenntnis über den Anteil der Schülerinnen und Schüler der Jahrgangsstufen, die einen Mathematik-Leistungskurs wählten. Hierfür sind die Werte der Spalte `LK` durch die Werte der Spalte `Sch12` zu dividieren. Diese Werte werden in der vierten Spalte eingetragen.

Aktivieren Sie zunächst das Listenfenster und wählen Sie sodann Resize, wodurch das Graphikfenster geschlossen wird. Definieren Sie die vierte Spalte durch `relHfg`) (vgl. Abbildung 2).

Um die Dividenden der Werte aus den Spalten `LK` und `Sch12` einzutragen, aktivieren Sie in der Fußzeile den Block `Cal` der Spalte `relHfg` und geben Sie `LK/Sch12` ein, s. Abbildung 2.

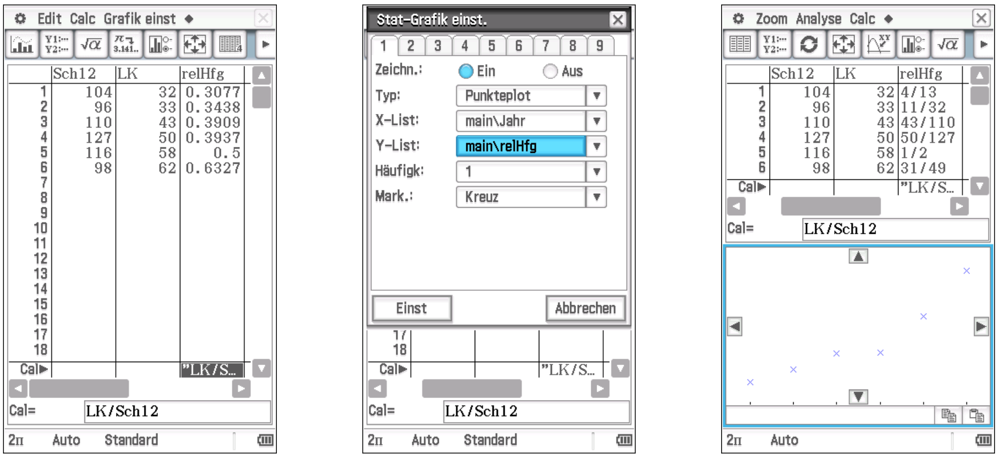

Abbildung 2: Berechnung und Regression der relativen Häufigkeit

Analog lässt sich ein Graph dieser Werte in Bezug auf das Jahr durchführen. Dies liefert den Graphen in Abbildung 2 (Skalierung auf der x-Achse: 1; auf der y-Achse: 0,1.

b) Die Vorhersage lässt sich mithilfe einer Regression machen. Um eine Regression durchzuführen, aktivieren Sie unter Graphik-Einstellungen neben `StatGraph1` auch `StatGraph2`, vgl. Abbildung 3, links.

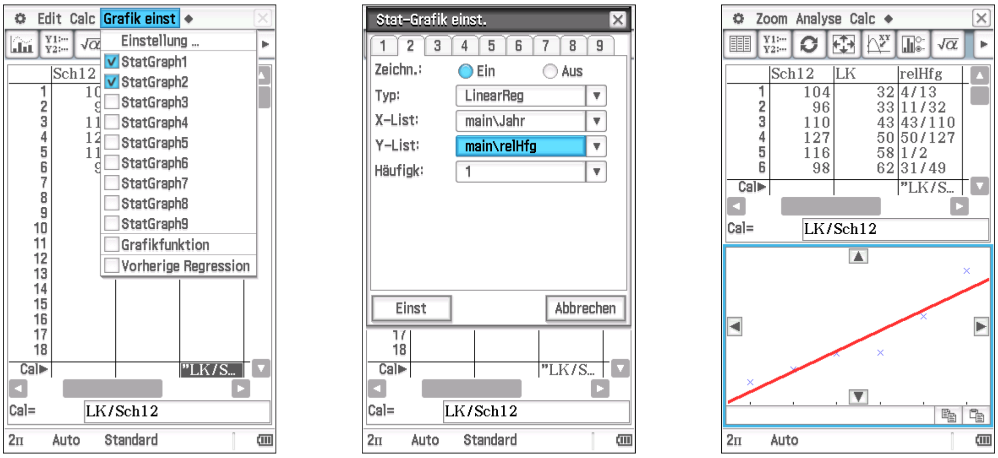

Abbildung 3: Darstellung einer linearen Regression

Die Einstellungen für den Graphen 2 wählen Sie unter

`Grafik einst` $\rightarrow$ `Einstellungen` ...

Geben Sie in `StatGraph2` die in der Mitte von Abbildung 3 sichtbaren Einstellungen ein und wählen eine lineare Regression aus. Das Ergebnis sehen Sie in Abbildung 3, rechts.

Die Abweichungen der Werte ist im Fall einer kubischen Regression geringer, s. Abbildung 4.

Nun soll die Funktionsvorschrift der Regressionskurve bestimmt werden (vgl. auch Abbildung 4, links) :

`Calc` $\rightarrow$ `Regression` $\rightarrow$ `Kubische Reg.` (Regression mit $y = ax^3 + bx^2 + cx + d$)

Die Funktionsvorschrift $f(x)$ kann gespeichert werden, vgl. Abbildung 5, links, für `y1`. Die Funktion ist näherungsweise durch die Vorschrift

$$f(x) \approx 4{,}81 \cdot 10^{-3} \cdot x^3 - 28{,}93 \cdot x^2 + 57974 \cdot x - 38719120$$

definiert, vgl. Mitte von Abbildung 5.

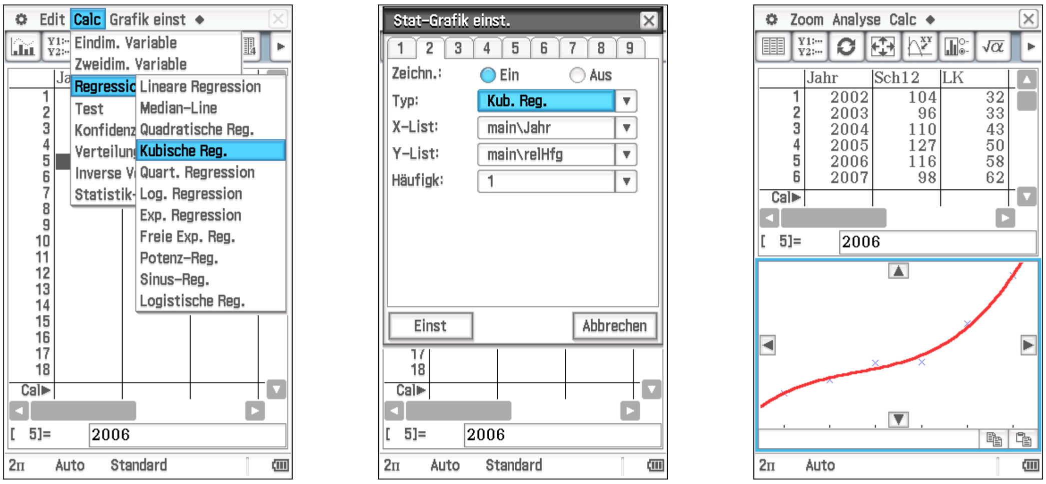

Abbildung 4: Darstellung einer kubischen Regression

Die übrigen Konstanten `r`^{2} und `MSe` haben die folgenden Bedeutungen:

`r`^{2}: Bestimmtheitsmaß (Quadrat des Korrelationskoeffizienten)

`MSe`: mittlerer quadratischer Fehler

Für die kubische Regression ist das Bestimmungsmaß r^2 und damit die Regression r besser als für die lineare Regression, wie in Abbildung 5, rechts, zu sehen ist. Es ist eine Entwicklung der relativen Häufigkeit der Leistungskurswähler zu erkennen.

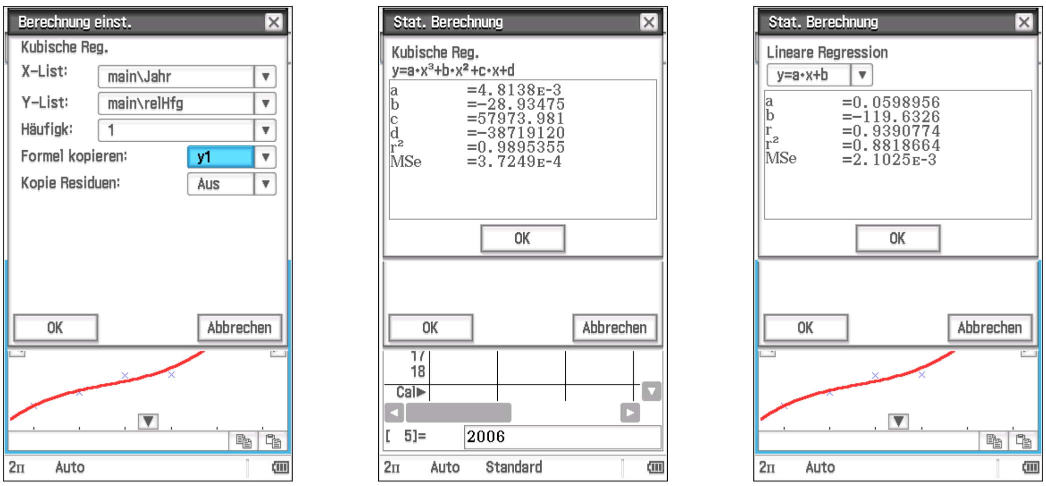

Abbildung 5: Berechnungen zur Regression

c) Gefragt ist die Berechnung der absoluten Häufigkeiten von Wahlen eines Mathematik-Leistungskurses in den nächsten Jahren. Stellen Sie die absoluten Häufigkeiten der Schüler in den Mathematik-Leistungskursen in Abhängigkeit der Jahre unter `StatGraph3` dar, siehe Abbildung 6. Die Entwicklung ähnelt einer Schwingung. Dies legt nahe, für die Regression die Sinusfunktion zu wählen. Dazu wählen Sie

`Calc` → `Regression` → `Sinus-Regr.` (Regression mit $y = a\sin(bx + c) + d$)

Die Formel der Funktion lässt sich in die Vorschrift y_1 kopieren, vgl. Abbildung 6, links.

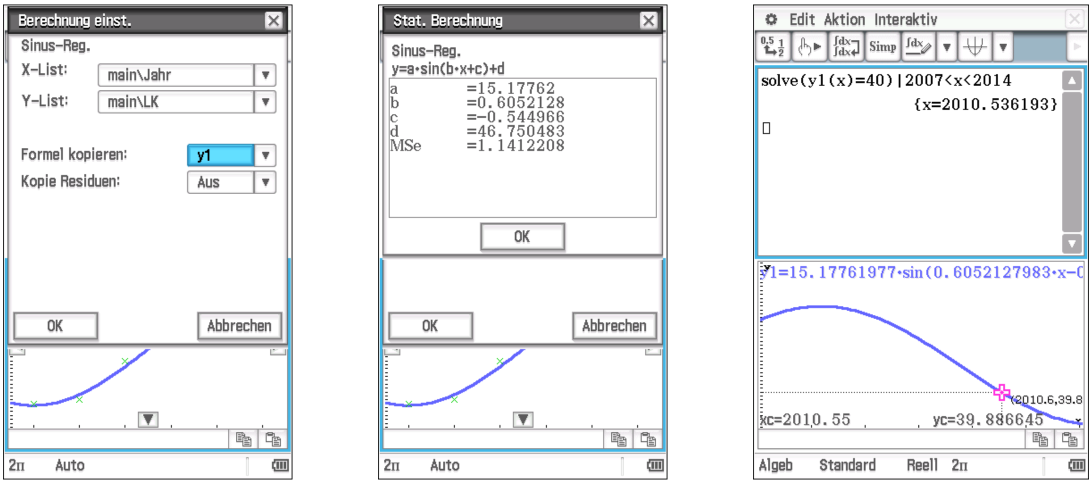

Abbildung 6: Vergleich des Datensatzes mit der Regression

Das Ergebnis sehen Sie in der Mitte der Abbildung 6. Die Regressionsfunktion f ist also näherungsweise gegeben durch

$$f(x) \approx 15{,}18 \cdot \sin(0{,}61 \cdot x - 0{,}54) + 46{,}75.$$

Der mittlere quadratische Fehler liegt bei 0,67. Ein Graph der Regressionsfunktion ist in der Mitte von Abbildung 6 zu sehen.

Zur Berechnung des Jahres, in dem die Anzahl der Wähler wieder 40 beträgt, ist die Gleichung

$$f(x) = 40$$

zu lösen. Dies wird unter Main mithilfe des `solve`-Befehls berechnet:

```
solve(y1(x)=40)|2007<x<2014
```

Durch `2007 < x < 2014` wird die Berechnung der Lösungen auf das Intervall $]2007; 2014[$ beschränkt. Mit dieser Berechnung erhält man das Ergebnis $x \approx 2010.536$. Das Ergebnis muss auf eine ganze Zahl abgerundet werden, da es im auf die berechnete Zahl folgenden ganzzahligen Jahr eintritt. Hierzu ist der Graph evtl. zu verschieben. Wie in Abbildung 6, rechts, sichtbar ist, ist dies voraussichtlich im Jahr 2010 der Fall.

Das Ergebnis ist allerdings nicht glaubwürdig, da sich vermutlich die Anzahl der Schüler in den kommenden Jahren nicht wie in den letzten Jahren entwickeln wird. Für die Vorhersage müsste man also zusätzlich die Geburtenrate berücksichtigen. Die relative Häufigkeit liefert ein glaubwürdigeres Ergebnis.

Lösung von Beispiel 2. Die Einstellungen der Graphik unter Statistik sind in Abbildung 7, links, dargestellt. Nachdem sie gespeichert wurden, sind die Einstellungen der x-Koordinaten des Histogramms zu wählen.

Der Startwert `H-Start` der Koordinaten ist auf das Jahr 2002 gesetzt.

Die Breite der Blöcke des Histogramms wird durch `H-Schr.` definiert; sie sollte eine Einheit betragen, damit die Balken jeweils für den Verlauf eines Jahres gelten und aneinander grenzen. Damit ergibt sich das in Abbildung 7, rechts, sichtbare Diagramm.

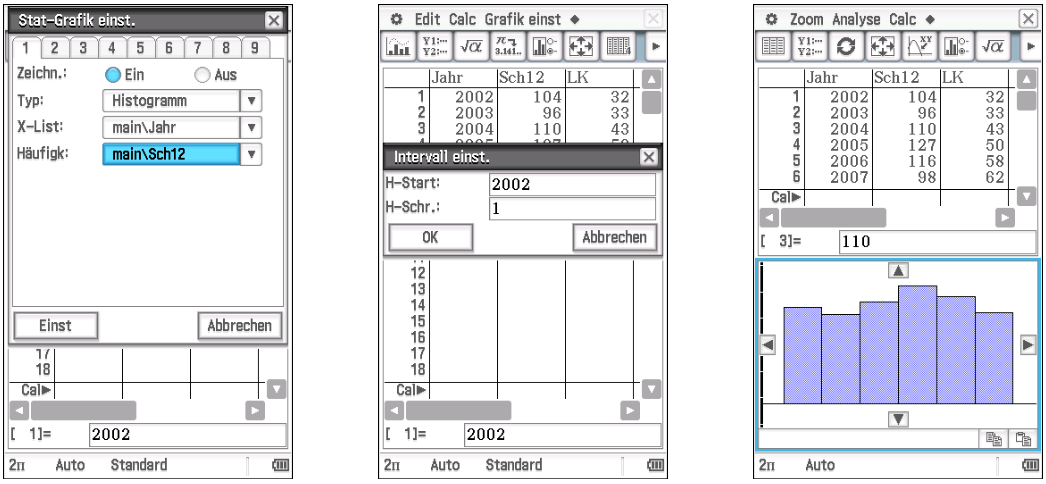

Abbildung 7: Erstellen eines Histogramms

Lösung von Beispiel 3. a) Zunächst wird die Tabellenkalkulation geöffnet. Die Wahlergebnisse tragen wir in die Zellen `A1` bis `A10` ein. In der Zelle `A12` berechnen wir mit `=sum(A1:A10)` (z.B. mit dem Befehl `sum(` unter `Calc` → `Liste-Berechnen`) die Gesamtzahl der Wählerstimmen. Mithilfe der ungültigen Stimmen ergibt sich die in Abbildung 8, links, sichtbare Anzahl 3216206 der Wahlteilnehmer.

Nun berechnen wir die Prozentsätze der Wahlergebnisse an der Wählergesamtzahl. In Spalte `B` werden mit den Quotienten der Wahlstimmen `A1` bis `A10` und `A12` die Anteile der Parteistimmen an der Gesamtwählerzahl notiert. Es ist darauf zu achten, dass die Berechnung unter Berücksichtigung der ungültigen Stimmen gemacht werden muss! Dies lässt sich mit

`Edit` → `Füllen` → `Mit Wert füllen` (Vorschrift für Zellenzeilen/-spalten definieren)

machen. Damit öffnet sich das Fenster in der Mitte von Abbildung 8, in das die dort sichtbaren Werte eingegeben werden sollen. In Spalte `B` werden damit die in Abbildung 8, rechts, gezeigten Anteile der Wahlstimmen von Wahlberechtigten angegeben. Zur Überprüfung des Ergebnisses wurde hierbei in Zelle `B12` die Summe der Einträge in `B1` bis `B10` berechnet, was die Summe 1 der relativen Häufigkeiten ergibt.

Die Anteile der Wahlstimmen sollen als Kreisdiagramm wiedergegeben werden. Sollen die Prozentsätze eingetragen werden, so sind die Werte aus Spalte `B` mit 100 zu multiplizieren und in Spalte `C` einzutragen. Dann kann nach der Aktivierung von Spalte `C` mit den Markierungen analog zur Mitte von Abbildung 9 ein Kreisdiagramm gezeichnet werden.

b) Die Sitzzahlen werden prozentual von der Gesamtzahl der Parteien mit mindestens 5 % der Stimmen berechnet. Diese Gesamtzahl erhalten wir mit `=sum(A1:A5)` zu 2829129. In Zelle `D1` definieren wir jetzt `=A1/2510190`. Analog berechnen wir die Einträge in der Zellen `D2` bis `D5`, vgl. Abbildung 9, rechts.

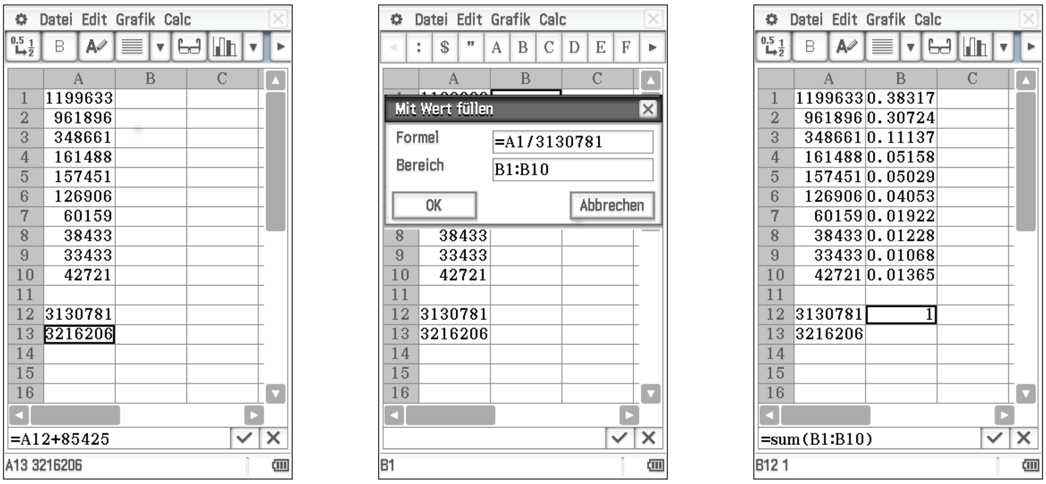

Abbildung 8: Wahlstimmen und ihre Anteile

Die Einträge in der Spalte `D` geben die Stimmenanteile an der Gesamtstimmenzahl der in den Landtag ziehenden Parteien an. Um die Anzahl der Sitze zu berechnen, sind diese Ergebnisse mit der Anzahl 110 der Sitze zu multiplizieren (z.B. `=A1/2510190×110`), vgl. Abbildung 10.

Partei	Anzahl an Sitzen (ungerundet)	Anzahl an Sitzen
CDU	47,791	47
SPD	38,320	38
GRÜNE	13,890	13
DIE LINKE	6,433	6
FDP	6,272	6

Analog zu Teil a) erstellen wir das Kreisdiagramm und kommen zu dem in Abbildung 10, rechts, gezeigten Ergebnis.

In diesem Jahr ging die Verteilung der Sitze auf. In der Wahl zuvor war dies nicht der Fall. Rundete man bei der Berechnung der Sitzplatzvergabe alle Zahlen ab, so ergab sich:

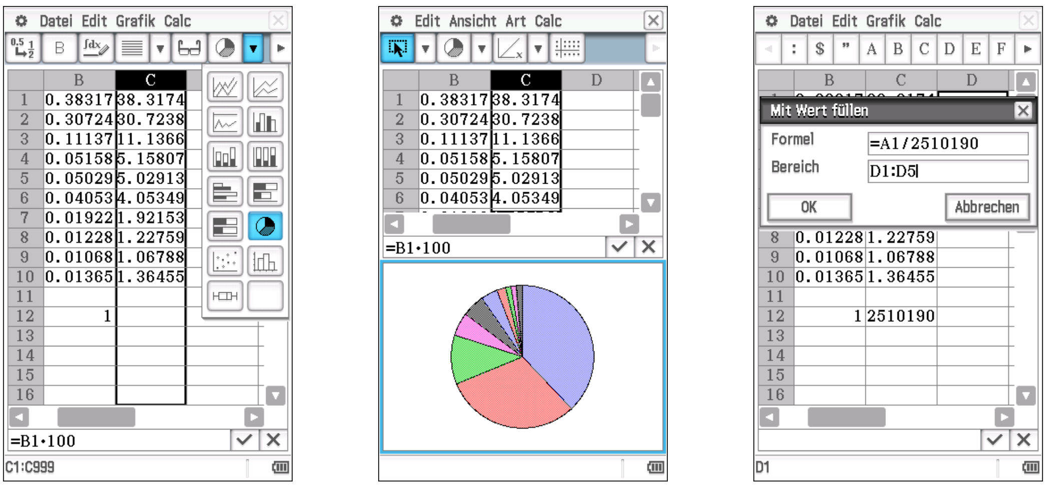

Abbildung 9: Prozentsätze graphisch und rechnerisch

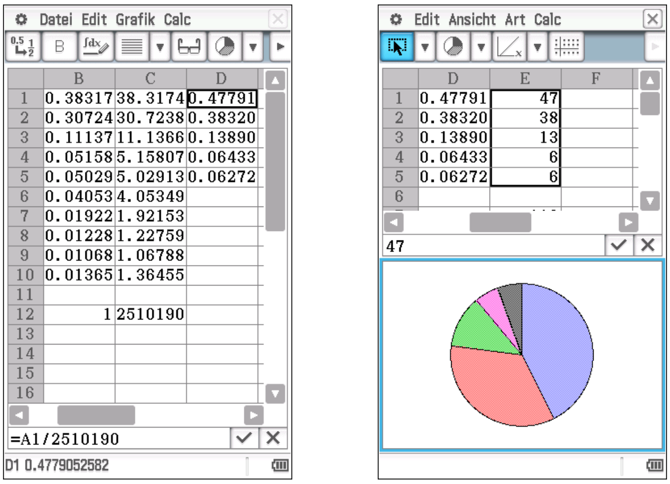

Abbildung 10: Sitzverteilung im Landtag

Partei	Anzahl an Sitzen (ungerundet)	Anzahl an Sitzen
CDU	42,51	42
SPD	27,11	27
FDP	18,54	18
GRÜNE	15,70	15
DIE LINKE	6,13	6

Es fällt auf, dass die Summe der Plätze $108 < 110$ beträgt, d. h., dass durch das Abrunden der Ergebnisse zwei Plätze nicht besetzt sind. Es scheint sinnvoll, je einen der Plätze an die Parteien mit den höchsten Nachkommazahlen in den Ergebnissen zu verteilen (d. h. je einen Platz an die FDP und die Grünen). Dies wurde so berechnet und liefert die korrigierten Ergebnisse der folgenden Tabelle. Mit den zusätzlichen Mandaten ergeben sich die Sitze der rechten Spalte dieser Tabelle.

Partei	Anzahl an Sitzen	korrigierte Anzahl der Sitze inkl. Überhang- und Ausgleichsmandate
CDU	42	46
SPD	27	29
FDP	19	20
GRÜNE	16	17
DIE LINKE	6	6
$\sum$	110	118