***Schluss von der Stichprobe auf die Gesamtheit – Konfidenzintervall***

*Schätzen von Parametern für binomialverteilte Zufallsvariable mit Hilfe der Normalverteilung.*

*Ermitteln eines Konfidenzintervalls für die Wahrscheinlichkeit p.*

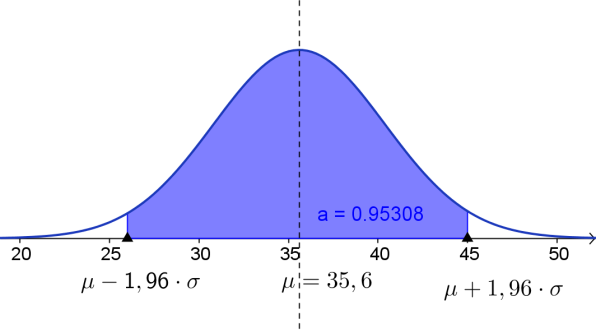
*Wir vergleichen folgende zwei Fragestellungen:*

Bei einer Wahl erhielt die Partei A in der Gesamtheit 35,6% der Stimmen. Wir wählen nach dem Auszählen der Stimmen zufällig 100 Stimmzettel aus. Auf wie vielen Stimmzetteln der Stichprobe wird die Partei A angekreuzt sein?

Bevor die Stimmzettel einer Wahl ausgezählt werden, zieht man 100 Stimmzettel zufällig aus den Wahlurnen. Auf 47 Stimmzetteln ist die Partei A angekreuzt. Welchen Anteil der Stimmen hat diese Partei in der Gesamtheit errungen?

*(1) (2)*

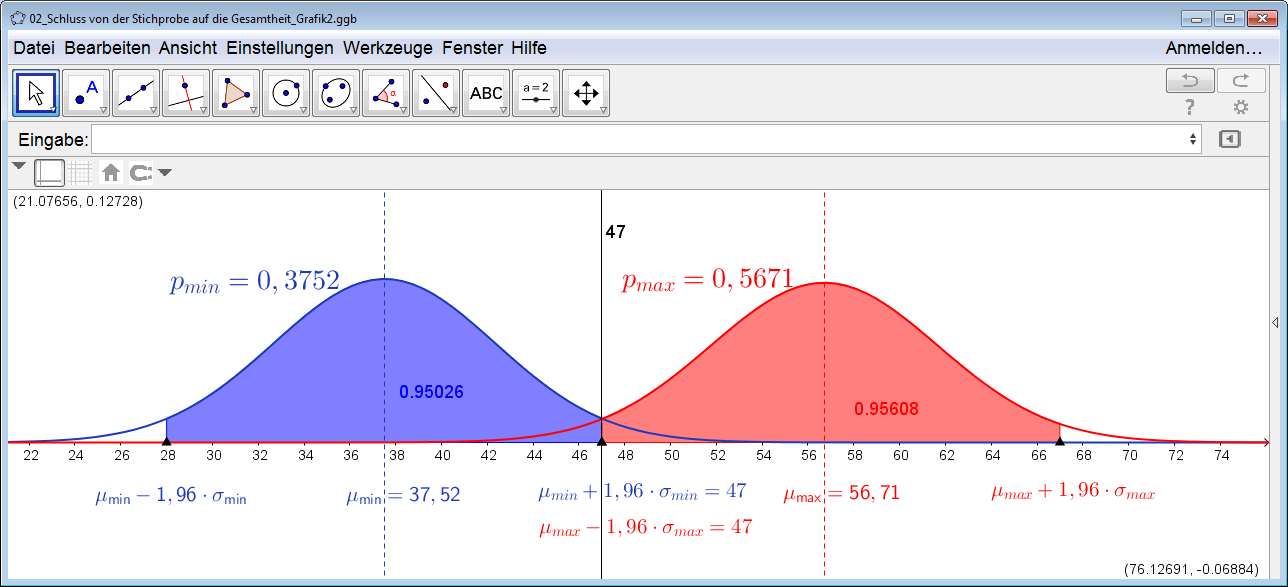
Quelle: Geretschläger Robert u.a., Elemente der Mathematik 8, Wien 2008, S.124ff

1. *****Schluss von der Gesamtheit auf die Stichprobe:*** *Die zugrunde liegende* ***Erfolgswahrscheinlichkeit*** *der Gesamtheit ist* ***bekannt.*** *Es soll ein γ-Schätzbereich ermittelt werden. ( n = 100 , p = 0,356 ⇒ und xmin = ? xmax = ? )*

*Für einen 95% ‑ Schätzbereich um μ gilt daher:*

1. ***Schluss von der Stichprobe auf die Gesamtheit:*** *Die zugrunde liegende* ***Erfolgswahrscheinlichkeit*** *der Gesamtheit ist* ***nicht******bekannt****. Sie muss aus der relativen Häufigkeit der Stichprobe geschätzt werden. Es soll ein Konfidenzintervall ermittelt werden. (n =100 , p = ? , xmin = 47 bzw. xmax = 47 soll möglich sein. )*

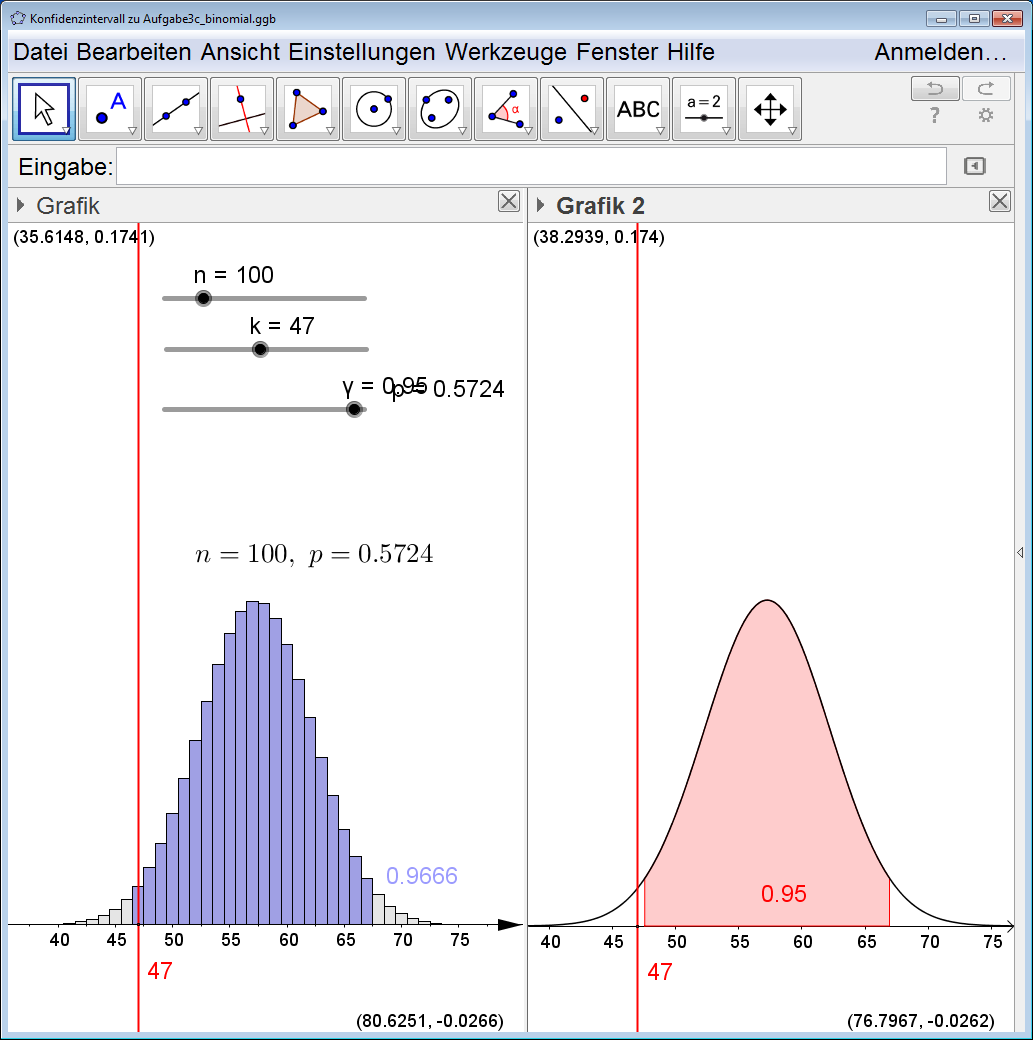
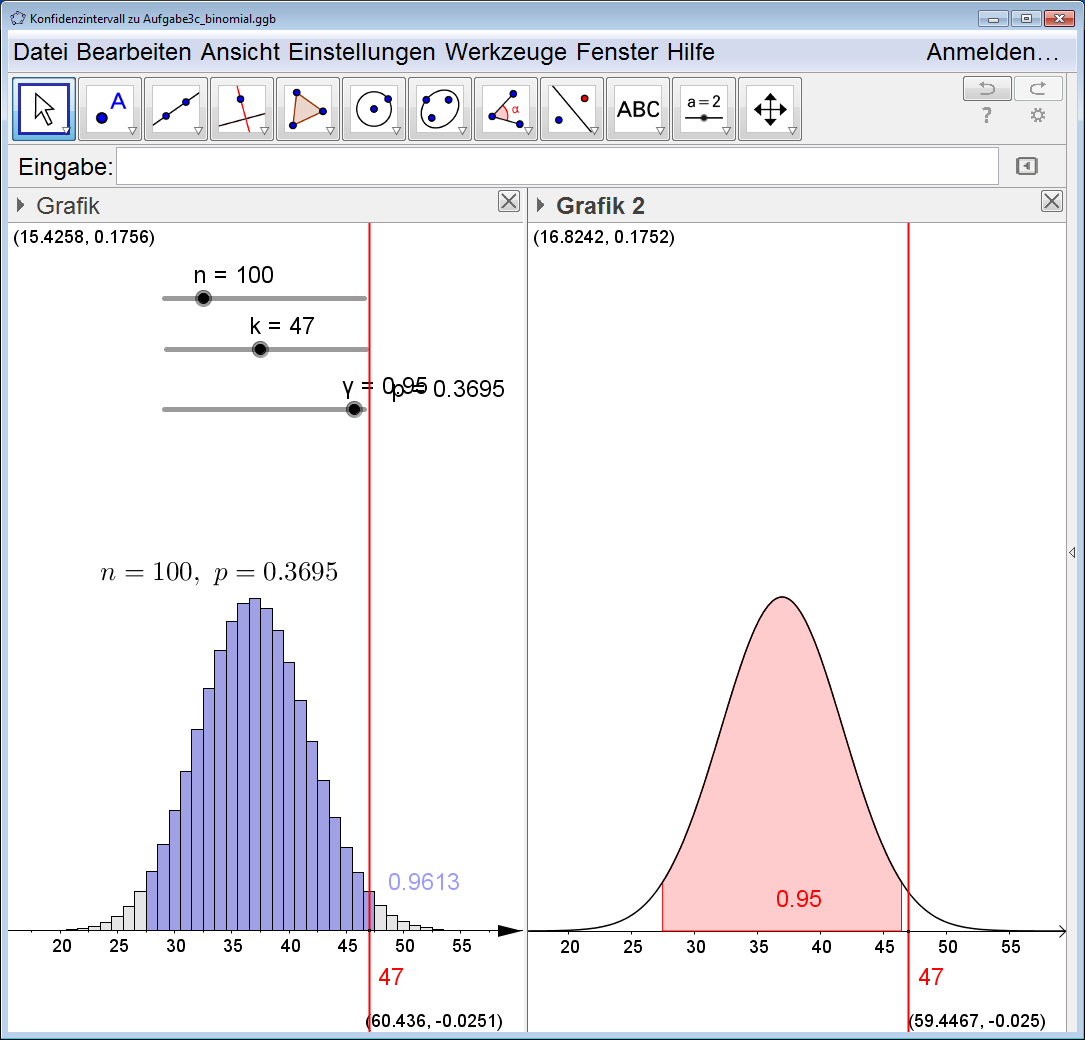
*Für einen 95% ‑ Schätzereich um μ muss daher für xmin = 47 bzw. xmax = 47 gelten:*



*durch numerisches Lösen der Gleichung nach p*

***Achtung!***  *Die obige Abbildung würde nahelegen, dass man die untere Grenze pmin aufrunden und die obere Grenze pmax des Konfidenzintervalls abrunden muss, weil beim Abrunden der unteren Grenze bzw. Aufrunden der oberen Grenze der 95%‑Bereich den Wert x=47 nicht mehr überdecken würde.*

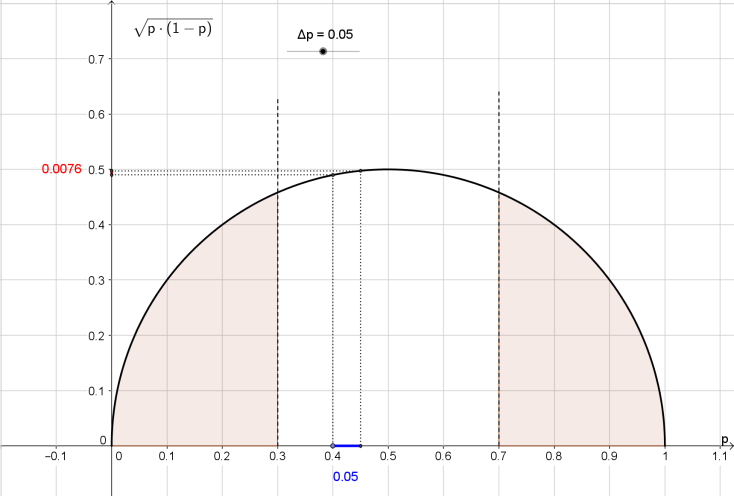
*Berücksichtigt man aber, dass man die real vorliegenden Binomialverteilung durch eine Normalverteilung approximiert hat, ist das* ***Abrunden der unteren Grenze*** *und das* ***Aufrunden der oberen Grenze*** *des Konfidenzintervalls gerechtfertigt. Siehe folgende Abbildung.*



*Das mit der Binomialverteilung exakt ermittelte Konfidenzintervall ist immer breiter als jenes, das mit der Normalapproximation berechnet wurde.*

*Näherungsweises Bestimmen des Konfidenzintervalls:*

Für 0,3 < p < 0,7 ergibt sich ein guter Näherungswert.



**Achtung!** Unterschiedliche Schreibweise für die relative Häufigkeit der Stichprobe :

*Thema-Mathematik 8: Formelhefte:*

***Näherungsweises (1-α) Konfidenzintervall nach Wald:***

Die allgemeine Berechnung des Konfidenzintervalls ergibt für und

***Konfidenzintervall nach Wilson:***

*Für große Werte von n kann man die Terme vernachlässigen und bekommt:*

*und für das*

***Näherungsweise (1-α) Konfidenzintervall:***

*Die halbe Breite des Konfidenzintervalls*  *nennt man auch die* ***Genauigkeit der Schätzung****. Für das näherungsweise Konfidenzintervall, das symmetrisch um liegt, gilt dann für ein 95%-Konfidenzintervall folgende Abschätzung:*

*weil bei maximal ist. (siehe Skizze oben) und .*

*Für die Frage wie groß der* ***Stichprobenumfang***  *sein muss, um eine vorgegebene Genauigkeit zu erreichen, bedeutet dies:*

*oder aus*

*Zusammenfassung:*

*Zur Schätzung des unbekannten relativen Anteils eines Merkmals in einer Grundgesamtheit wird eine Stichprobe vom Umfang erhoben. In dieser lässt sich der Wert für die relative Häufigkeit des untersuchten Merkmals in der Stichprobe ermitteln.*

*Die Menge aller Schätzwerte für , deren zugehörige -Schätzbereiche den in der Stichprobe beobachteten Wert überdecken, heißt* ***Konfidenzintervall*** *mit Sicherheit (‑Konfidenzintervall oder Vertrauensintervall zum Konfidenzniveau ) für den unbekannten relativen Anteil . nennt man auch Irrtumswahrscheinlichkeit.*

*Ein Konfidenzintervall wird nur auf der Grundlage einer konkreten Stichprobe ermittelt. Jedes derartige Konfidenzintervall hängt zufallsbedingt von der gewählten Stichprobe ab. Bei einer sehr großen Anzahl von Stichproben liegt der tatsächliche Wert für in z.B. 95% ( Konfidenzniveau   ) aller Fälle in dem jeweiligen Konfidenzintervall. In nur 5% ( Irrtumswahrscheinlichkeit   ) aller Fälle wird der Wert nicht im jeweils ermittelten Konfidenzintervall liegen.*

*Die* ***Breite des Konfidenzintervalls*** *wird vom Konfidenzniveau    und vom Umfang  der Stichprobe bestimmt.*

*Je höher das Konfidenzniveau  und je kleiner damit die Irrtumswahrscheinlichkeit    ist, desto breiter (größer) ist das Konfidenzintervall.*

*Je größer der Umfang  der Stichprobe ist, desto enger (kleiner) ist das Konfidenzintervall.*

*Wenn die Breite (Größe) des Konfidenzintervalls und damit die Genauigkeit (z.B. vom Auftraggeber einer Umfrage) vorgegeben ist, muss vor einer Befragung der entsprechende Umfang der Stichprobe ermittelt werden.*

Beispiel 1:

Vor einer Wahl werden 500 Wahlberechtigte befragt: 120 davon sprachen sich für die Partei A aus. Wie viel Prozent wird die Partei A daher bei der Wahl voraussichtlich erhalten? (n=500 , x=120 , p=?)

Umformulierung der Fragestellung:

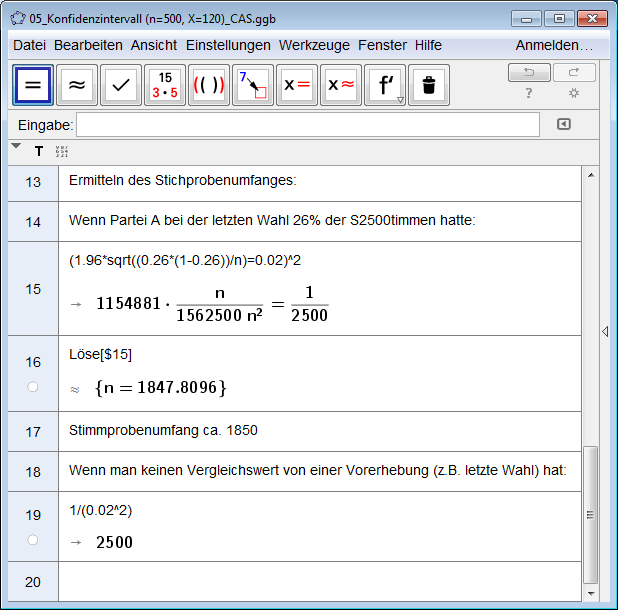
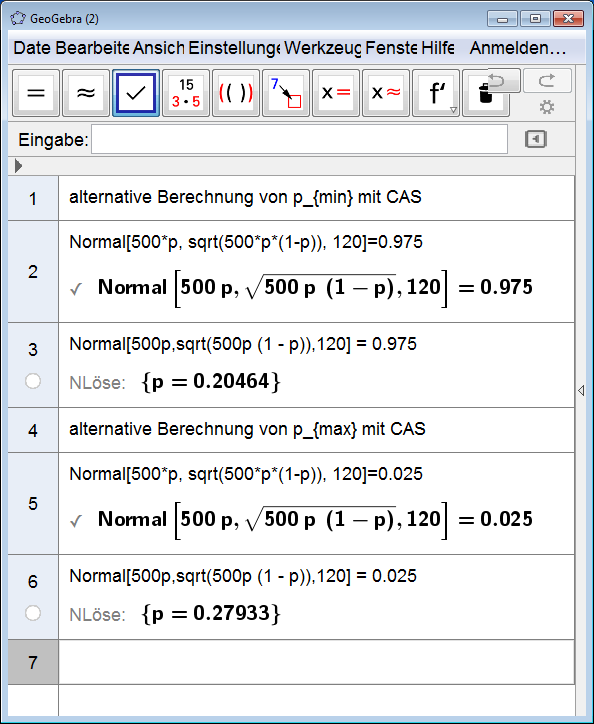
Gib ein 95% Vertrauensintervall für den Anteil der Wähler in der Menge der Wahlberechtigten an. Oder: Mit welchen Anteilen (Wahrscheinlichkeiten) p ist das Stichprobenergebnis x=120 verträglich?

Zusatzfragen:

Berechne ein näherungsweises 95%-Konfidenzintervall!

Wie viele Personen müssten befragt werden, damit mit einer Irrtumswahrscheinlichkeit von 5% und einer Genauigkeit von ±2% der Prozentsatz der Stimmen für die Partei A geschätzt werden kann?  
a) Wenn die Partei A bei der letzten Wahl 26% der Stimmen erhalten hat.  
b) Wenn man keinen Vergleichswert von der letzten Wahl hat.

Berechnung mit CAS:



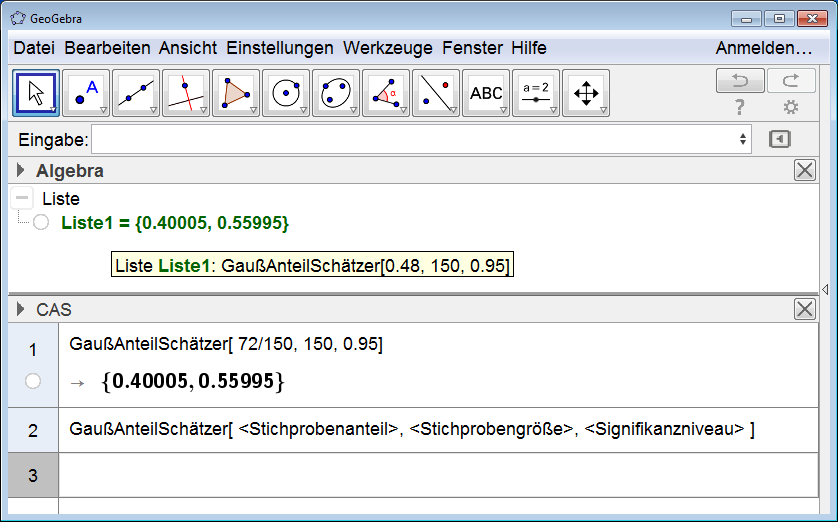
Beispiel 2: (Aufgabe 3c Haupttermin 2015)

In einer österreichischen Gemeinde, in der 1800 Einwohner/innen Blut spenden könnten, nahmen 150 Personen an einer freiwilligen Blutspendeaktion teil. Es wird angenommen, dass die Blutspender/innen eine Zufallsstichprobe darstellen. 72 Blutspender/innen hatten Blutgruppe A.

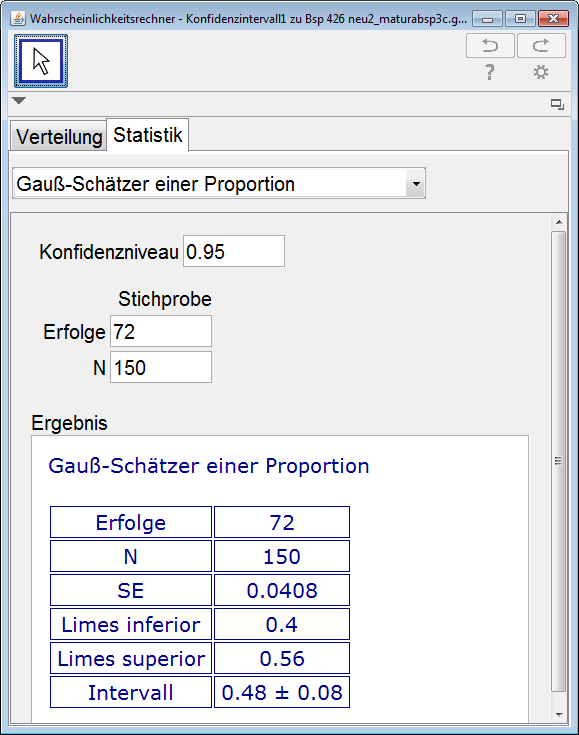
(1) Berechnen Sie aufgrund dieses Stichprobenergebnisses ein symmetrisches 95%‑Konfidenzintervall für den tatsächlichen (relativen) Anteil *p* der Einwohner/innen dieser Gemeinde mit Blutgruppe A, die Blut spenden könnten!

(2) Die Breite des Konfidenzintervalls wird vom Konfidenzniveau (Sicherheitsniveau) und vom Umfang der Stichprobe bestimmt. Geben Sie an, wie jeweils einer der beiden Parameter geändert werden müsste, um eine Verringerung der Breite des Konfidenzintervalls zu erreichen! Gehen Sie dabei von einem unveränderten (gleichbleibenden) Stichprobenergebnis aus.

(1) Ermitteln im Algebrafenster oder mit CAS:



(1) Ermitteln mit dem Wahrscheinlichkeitsrechner:



Quelle: <http://elearn.bgamstetten.ac.at/srp/doku.php?id=tech:ggb:beurteilende_statistik>

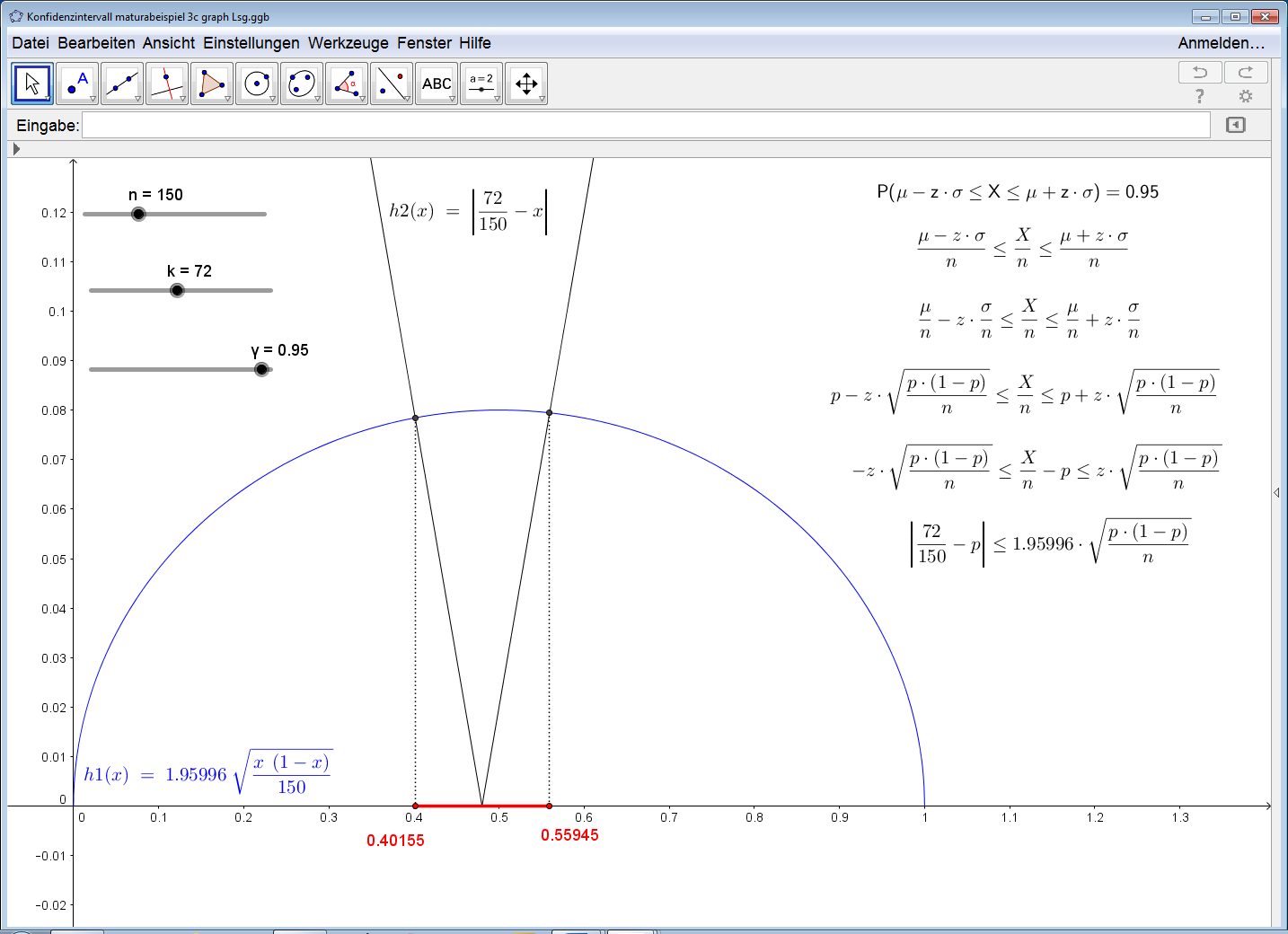
(2) Bei gleichem Stichprobenergebnis führen eine größere Stichprobe und / oder ein geringeres Konfidenzniveau zu einer Verringerung der Breite des Konfidenzintervalls

Konfidenzintervall - graphische Lösungen:

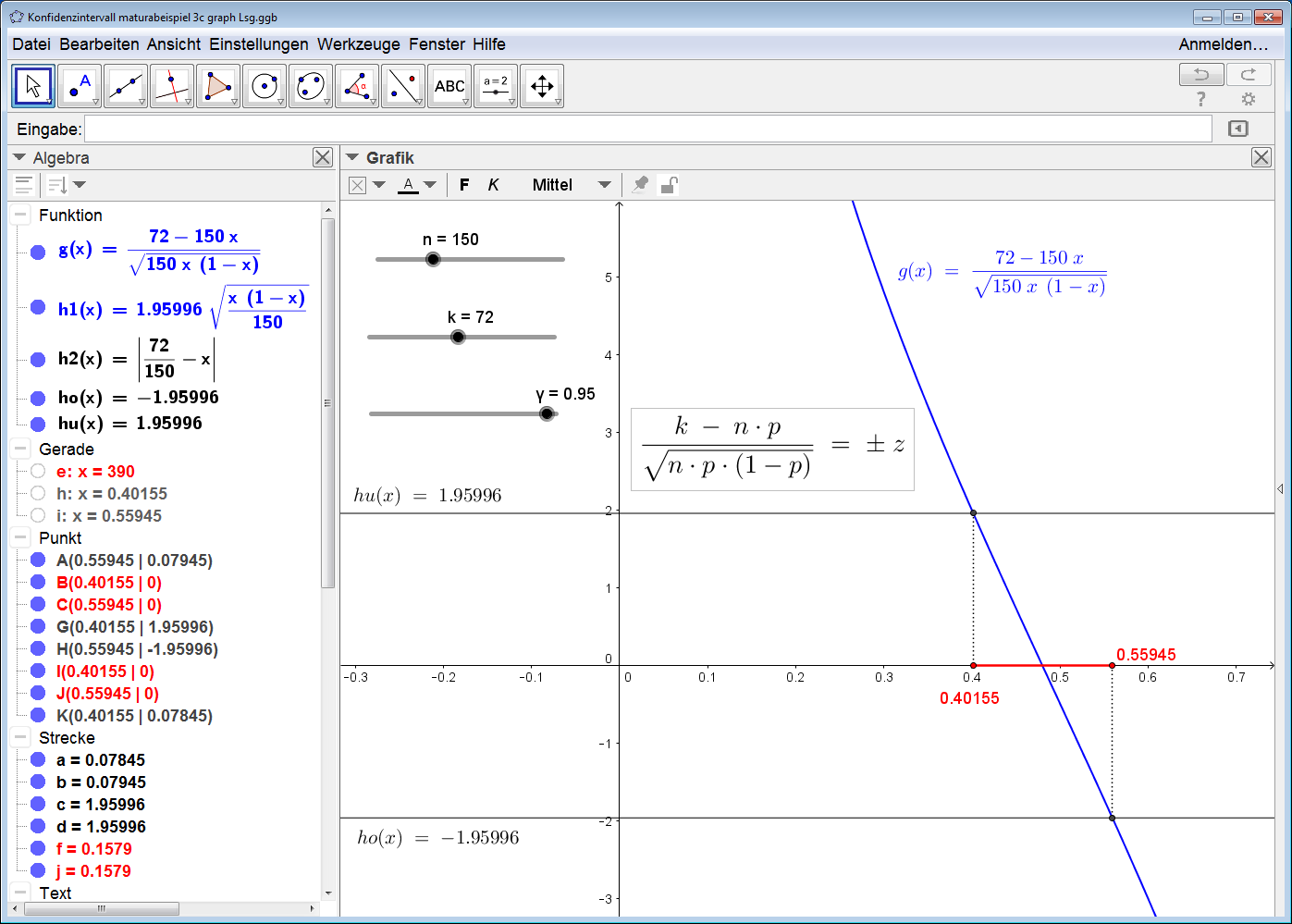
Graphisches Lösen einer quadratischen Ungleichung



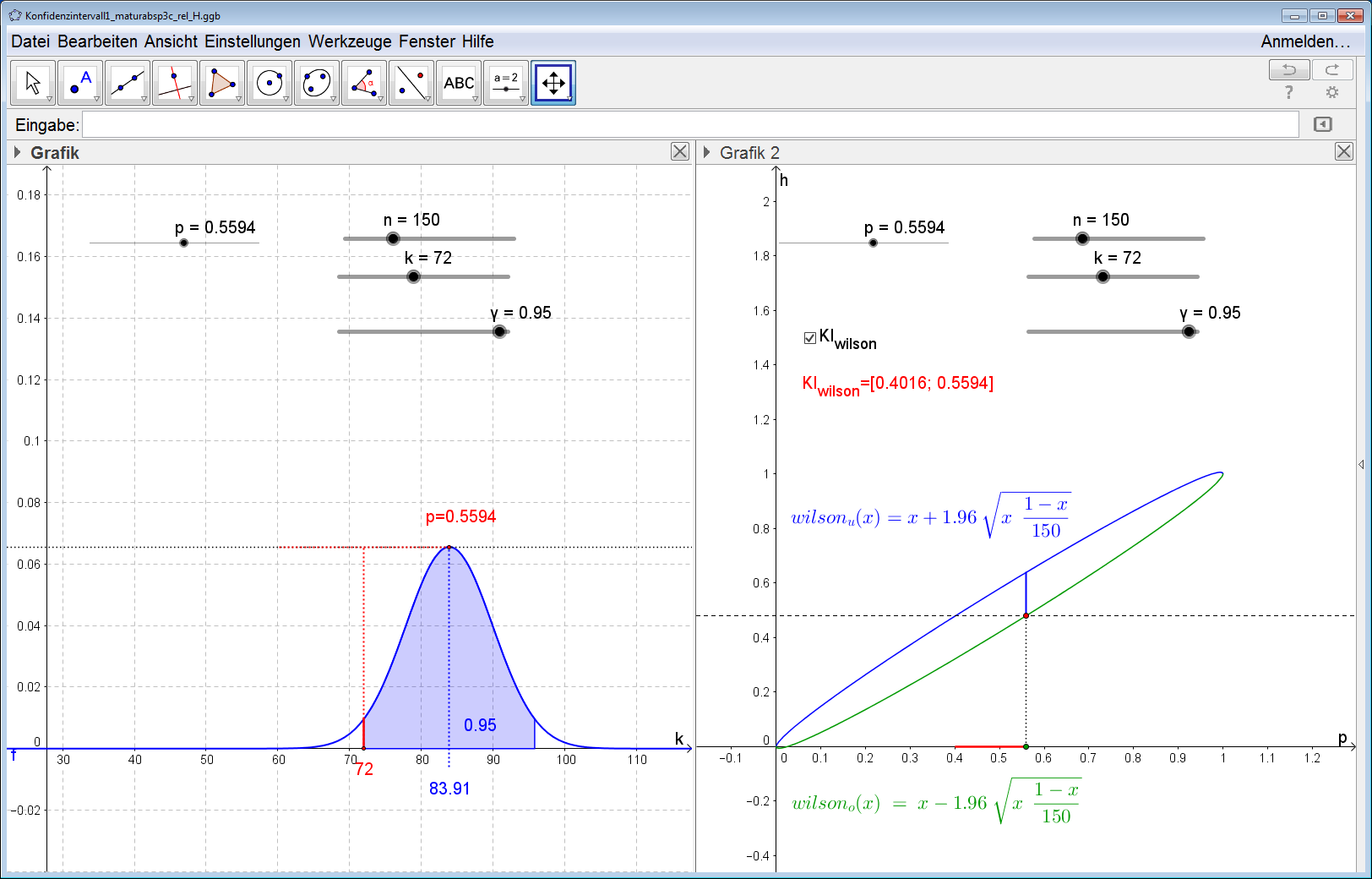
Graphisches Lösen einer Betragsungleichung



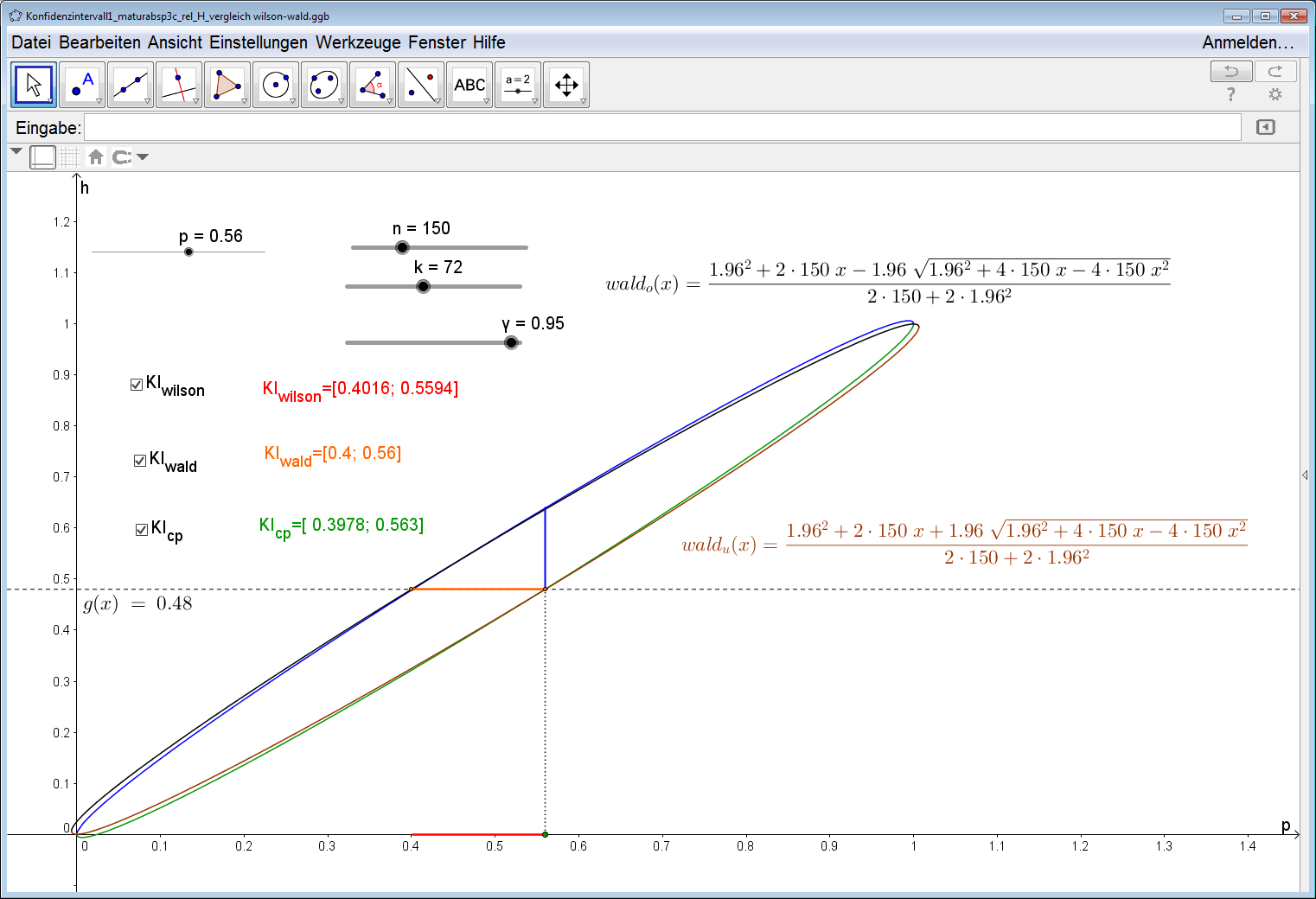
Graphisches Lösen 3



Graphische Lösung mittels Wilson-Konfidenzellipse



Vergleich zwischen KIWilson und KIWald mittels Konfidenzellipsen



Anhang 1:

**Berechnen der Grenzen des Clopper-Pearson-Konfidenzintervalls mit Geogebra**

Nach Ersetzen von durch

|  |  |
| --- | --- |
|  | und Umformen erhält man |

Quellen: Werner Timischl; [Biostatistik: Eine Einführung für Biologen und Mediziner](https://books.google.at/books?id=NNUiBgAAQBAJ&printsec=frontcover&dq=isbn:3709163137&hl=de&sa=X&ei=DJeeVcqSAsSBU_KCmLAL&ved=0CB8Q6AEwAA), S.75

<https://books.google.at/books?isbn=3709163137>

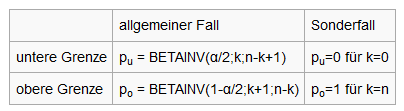
<http://www.mathepedia.de/F-Verteilung.aspx>

In der Befehlsstruktur von Geogebra:

cpu := k / (k + (n - k + 1) InversFVerteilung[2 (n - k + 1), 2k, (1 + γ) / 2])

cpo := (k + 1) InversFVerteilung[2 (k + 1), 2 (n - k), (1 + γ) / 2] / (n - k + (k + 1) InversFVerteilung[2 (k + 1), 2 (n - k), (1 + γ) / 2])

**Berechnen der Grenzen des Clopper-Pearson-Konfidenzintervalls mit Excel**



Quelle: <https://de.wikipedia.org/wiki/Konfidenzintervall_einer_unbekannten_Wahrscheinlichkeit>

Anhang 2:

**Literatur zum Thema Konfidenzintervalle**

Vehling Reimund, Konfidenzintervalle mit dem TI-NspireTM CAS, TI-Nachrichten 1/15, S.21-26. <http://www.ti-unterrichtsmaterialien.net/index.php?id=1&detail=1216>

Vehling Reimund, Ein zweiter Blick auf Konfidenzintervalle, DASU Didaktischer Arbeitskreis Schule Universität, Band 30, Hannover   
[http://www.unikik.uni-hannover.de/dasu\_archiv.html](file:///D:\Users\Haller\Documents\M\8BG_1415\8BG_1415_Beurteilende%20Statistik\Material%20von%20Raimund%20Vehling%20und%20anderen\dereferrer?redirectUrl=http%3A%2F%2Fwww.unikik.uni-hannover.de%2Fdasu_archiv.html)

Vehling Reimund, Mit Simulationen zum Konfidenzintervall, PM Praxis der Mathematik in der Schule,Heft 39, 53.Jahrgang, 201, S.25ff. <http://www.aulis.de/files/materials/downloads_links/PM_Jahresinhalt_2011.pdf>

Weiß Siegfried, Konfidenzintervalle verstehen, TI-Nachrichten 1/11, S.19-22.   
<http://www.ti-unterrichtsmaterialien.net/index.php?id=1&detail=993>

Bergmann Lars, Dynamische Entwicklung von Konfidenzintervallen, TI-Nachrichten 2/13, S.1-4. <http://www.ti-unterrichtsmaterialien.net/index.php?id=1&detail=1155>

Rolfs G., Konfidenzintervall

<http://nibis.ni.schule.de/~lbs-gym/Stochastikpdf/Konfidenzintervall.pdf>

<http://nibis.ni.schule.de/~lbs-gym/Stochastikpdf/Konfidenzintervall2.pdf>

<http://nibis.ni.schule.de/~lbs-gym/Verschiedenespdf/Ueberdeckung.pdf>

Wikipedia: Konfidenzintervall

<https://de.wikipedia.org/wiki/Konfidenzintervall>

<https://de.wikipedia.org/wiki/Konfidenzintervall_einer_unbekannten_Wahrscheinlichkeit>

Peschek Werner, Was kann man von „pflichtbewussten“ Stichproben erwarten? Wahrscheinlichkeitsrechnung im Dienste angewandter Statistik, Wien 2000 [http://www.oemg.ac.at/DK/Didaktikhefte/2000 Band 32/Peschek2000.pdf](http://www.oemg.ac.at/DK/Didaktikhefte/2000%20Band%2032/Peschek2000.pdf)

Peschek Werner u.a., Stochastik in der Schule: Globale Ideen, lokale Bedeutungen, zentrale Tätigkeiten, Klagenfurt 2000

Meyer Jörg, Schwierigkeiten mit Konfidenzintervallen, Stochastik in der Schule 33 (2013) 3, S. 10-17

Diepgen Raphael, Schwierigkeiten mit Konfidenzintervallen? In der Tat! Stochastik in der Schule 34 (2014) 2, S. 26-31

Geretschläger Robert u.a., Elemente der Mathematik 8, Wien 2008, S.124ff

Brand Clemens u.a., Thema Mathematik 8, 6. Auflage, Linz 2014, S. 106ff.