hausübung aus biostatistik - beispiel 1 & 2 vom 10ten okt. 2001

Berechne den alters-mittelwert der **österreichischen bevölkerung** unter annahme dass das mittlere alter der einzelnen klassen genau in der mitte der klassenbreite genau in der mitte der klassenbreite liegt; d.h. 2.5, 7.5, etc. jahre (fettgedruckte tabellen-einträge sind die angaben).

a) ist der somit berechnete mittelwert mit dem "wahren" mittelwert identisch? diskutiere mögliche abweichungen;
b) berechne die standard-abweichung
c) wo wird der median liegen?
d) stelle die häufigkeiten und verteilungen der österr. und ober-osterr. bevölkerung grafisch dar und diskutiere die unterschiede.

			absolute häufigkeit				**relative** häufigkeit	
sample	altersgr.	(x_i)	ober-österr.	österr. (f_i)	x_i*f_i	x_i^2*f	ober-österr.	österr.
1	**0 - 5**	2,5	**77746**	**429320**	1073300	2683250	5,64582072	5,30532037
2	**5 - 10**	7,5	**86089**	**473546**	3551595	26636962,5	6,25167931	5,85184301
3	**10 - 15**	12,5	**86197**	**467544**	5844300	73053750	6,25952214	5,77767332
4	**15 - 20**	17,5	**88210**	**484149**	8472607,5	148270631,3	6,40570377	5,98286954
5	**20 - 25**	22,5	**80051**	**475806**	10705635	240876787,5	5,81320703	5,87977095
6	**25 - 30**	27,5	**97841**	**600476**	16513090	454109975	7,10509537	7,42038003
7	**30 - 35**	32,5	**121185**	**715044**	23238930	755265225	8,80030848	8,83615368
8	**35 - 40**	37,5	**122214**	**700121**	26254537,5	984545156,3	8,87503322	8,65174277
9	**40 - 45**	42,5	**102879**	**593476**	25222730	1071966025	7,47094885	7,33387756
10	**45 - 50**	47,5	**87546**	**517074**	24561015	1166648213	6,35748489	6,38974011
11	**50 - 55**	52,5	**76811**	**484231**	25422127,5	1334661694	5,57792214	5,98388286
12	**55 - 60**	57,5	**81899**	**524227**	30143052,5	1733225519	5,94740657	6,47813329
13	**60 - 65**	62,5	**64315**	**376168**	23510500	1469406250	4,6704777	4,64849472
14	**65 - 70**	67,5	**60229**	**353099**	23834182,5	1608807319	4,37375731	4,36341964
15	**70 - 75**	72,5	**54606**	**334116**	24223410	1756197225	3,96542184	4,12883728
16	**75 - 80**	77,5	**47022**	**285649**	22137797,5	1715679306	3,41468091	3,5299065
17	**80 - 85**	82,5	**20987**	**130394**	10757505	887494162,5	1,52405062	1,61134339
18	**85 - 90**	87,5	**16263**	**108350**	9480625	829554687,5	1,18099944	1,33893474
19	**90 - 95**	92,5	**4118**	**32503**	3006527,5	278103793,8	0,29904419	0,40165571
20	**> 95**	97,5	**846**	**6961**	678697,5	66173006,25	0,0614355	0,08602053
Σ =			1377054	8092254	318632165	16603358938		
probe =							100	100

antwort auf frage "a" (mittelwert):

der mittelwert (x_{quer} = mean of frequency distribution) errechnet sich aus der summe der produkte (altersgruppe * österreicher-häufigkeit) dividiert durch deren anzahlsumme:

$x_{quer} = \Sigma(f_i * x_i) / \Sigma f = 318632165 / 8092254 =$

$=$ **39,37495845**

der alters-mittelwert der österreicher beträgt (x_{quer}) 39.4 jahre und ist mit dem "wahren" wert ??????????ß

antwort auf frage "b" (standard-abweichung):

die standardabweichung wird über die varianz (variance of frequency distribution) gerechnet:

σ^2 = [Σx^2*f - (Σxf)^2 / Σf)] / Σf-1 = [16603358938 - (318632165^2 / 8092254) / (8092254-1)] =

= 501

die standard abweichung ist letzlich die wurzel aus s^2 (varianz^2):

σ = wurzel [σ^2] =

= **22,39134205**

der standard-abweichung des mittelwertes (standard deviation) beträgt demnach: 22.4

antwort auf frage "c" (lage des medians = depth of the median; Johnson p67):

x_{tilde} = (anzahl der sample-klassen + 1) / 2 = (20 + 1) / 2 =

= **10,5**

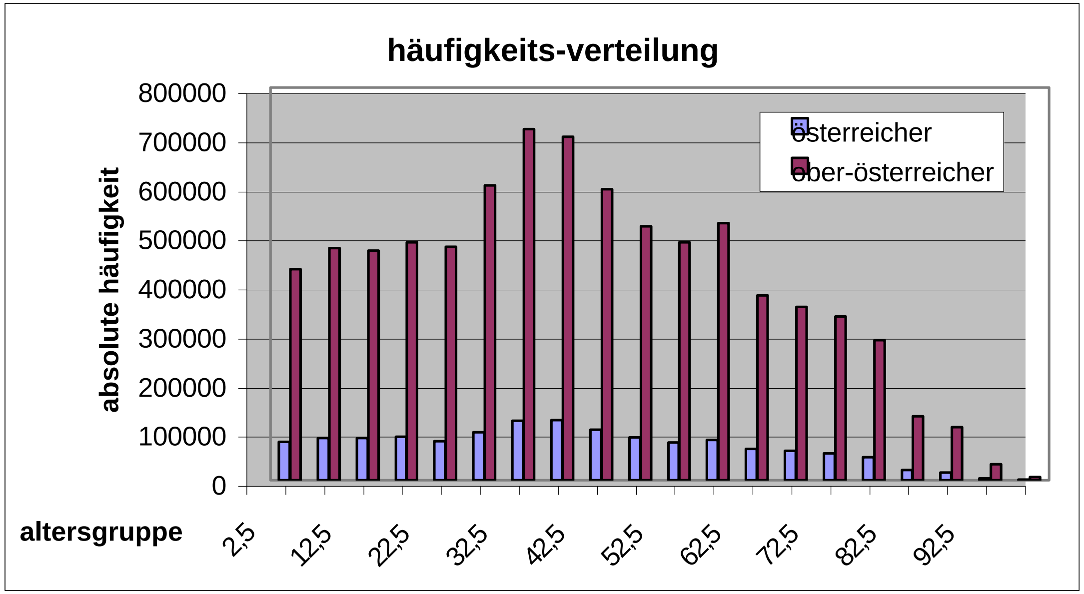
die lage des medians befindet sich zwischen der 10ten und 11ten österreicher-klasse; i.e 50

als zusatz lässt sich der modus-wert bestimmern = häufigste altersgruppe (datensatz):

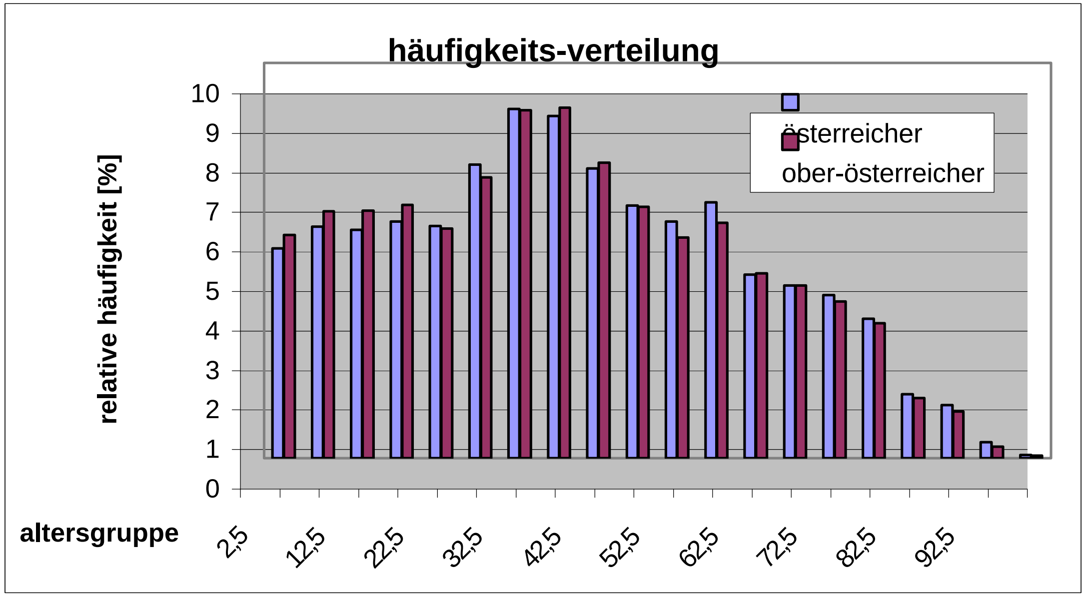
x_{mode} = **32.5**

antwort auf frage "d" (grafische dar-
stellung) mit diskussion:

absolute häufigkeits-verteilung der österreicher und ober-österreicher



relative häufigkeits-verteilung der österreiher und ober-österreicher



diskussion:

i) der modus-wert d. ober-österr. liegt bei 37.5 jahre gegenüber 32.5 bei den österreichern.
i) die oberösterr. bevölkerung ist etwas jünger als die österr.
i) generell zeigt sich ein einheitliches bild, in dem die "baby-boomer" generation (30-40jährige) dominiert.
i) die kein/einzelkind-famile bzw. singelhaushalte finden sich in der bevölkerungsstatistik wieder.

hausübung aus biostatistik - beispiel 3 vom 10ten okt. 2001

Die bestimmung von chromosomen-aberrationen an insgesamt 1000 zellen hat folgende werte ergeben:
bestimme für diesen datensatz folgende statischtische parameter (fettgedruckte tabelleneinträge sind die angaben):
a) den mittelwert (x_{quer})?
b) die standard-abweichung (s)?
c) den standardfehler des mittelwertes (Sx_{quer})?

sample	häufigktn d. aberrat. (x_i) abs.	rel.	anzahl der zellen (f_i) abs.(n_i)	rel.	abs.häufgkt. f_i*x_i	x_i^2*f
1	**0**	0	**360**	0	0	0
2	**1**	0,028	**372**	0,287	372	372
3	**2**	0,056	**185**	0,2855	370	740
4	**3**	0,083	**59**	0,1366	177	531
5	**4**	0,111	**15**	0,0463	60	240
6	**5**	0,139	**7**	0,9722	35	175
7	**6**	0,167	**1**	0,0046	6	36
8	**7**	0,194	**1**	0,0054	7	49
9	**8**	0,222	**0**	0	0	0
Σ =	36		1000		1027	2143
gegenprobe =		1		1,7377		

antwort zu frage "a" (mittelwert):

der mittelwert (x_{quer} = mean of frequency distribution) errechnet sich aus der summe der produkte (zellen * abberations-häufigkeit) dividiert durch deren anzahlsumme:

x_{quer} = Σ(f_i*x_i) / Σf = 1027/1000 =

= **1,027**

der mittelwert beträgt demnach 1.03

antwort zu frage "b" (standard-abweichung - Johnson p84):

die standardabweichung wird über die varianz (variance of frequency distribution) gerechnet:

σ^2 = [Σx^2*f - (Σxf)^2 / Σf)] / Σf-1 = [2143 - (1027^2 / 1000) / (1000-1)] =

= 1,089

die standard abweichung ist letzlich die wurzel aus s^2 (varianz^2):

σ = wurzel [σ^2] =

= **1,044**

die abweichung vom mittelwert (standard deviation of mean) beträgt demnach 1.04

antwort zu frage "c" (standardfehler, Johnson p325):

der standardfehler errechnet sich aus der standard-abweichung dividiert durch die wurzel aus n:

σx_{quer} = s / wurzel(n) =

= **0,033**

der standardfehler des mittelwertes (standard error of mean) beträgt demnach 0.03

hausübung aus biostatistik - beispiel 4 vom 10ten okt. 2001

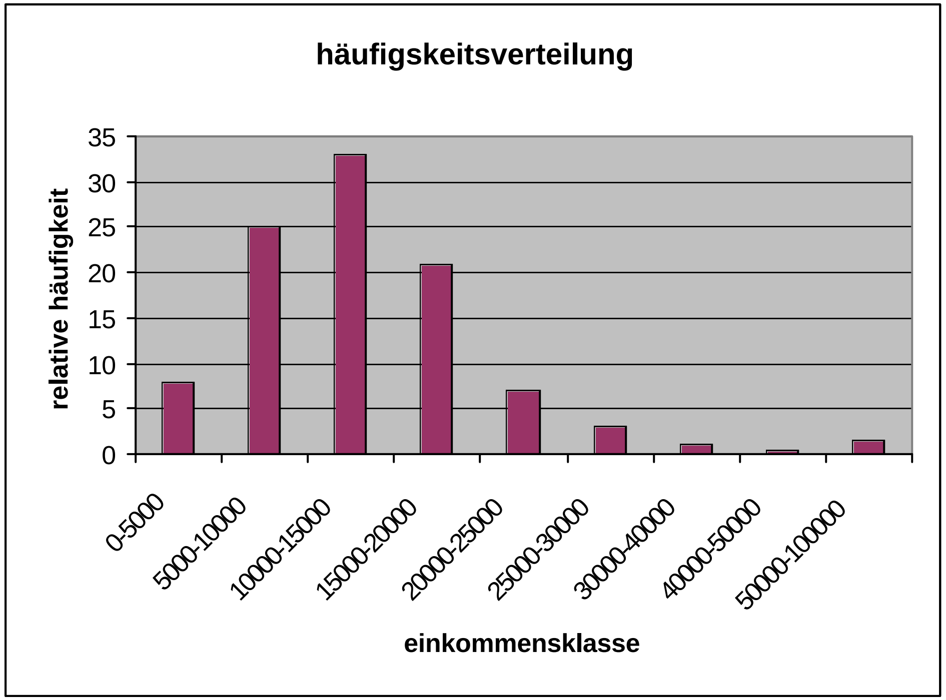
Die nachstehende tabelle enthält die einkommensverteilung in N$ der unselbständigen bevölkerung des vereinigten staatenbundes von nirvana.
a) wie hoch ist das durschnitts-einkommen in nirwana (fettgedruckte tabelleneinträge sind die angaben)?
b) wo liegt das medianeinkommen?
c) wo liegt das modaleinkommen?
d) in der bevölkerungsstatistiken und in den wirtschaftsberichten wird in der einkommensverteilung immer das durchschnitts-einkommen angeben.
Warum ist diese angabe irreführend, zumindest jedoch unvollständig?

einkommen (N$) klasse	kl.-mittel (x_i)	prozent [%]	rel.häufgkt. h_{rel}	abs. häufgkt. $h_{rel} * x_i$
0-5000	2500	**8**	0,08	200
5000-10000	7500	**25**	0,25	1875
10000-15000	12500	**33**	0,33	4125
15000-20000	17500	**21**	0,21	3675
20000-25000	22500	**7**	0,07	1575
25000-30000	27500	**3**	0,03	825
30000-40000	35000	**1**	0,01	350
40000-50000	45000	**0,5**	0,005	225
50000-100000	75000	**1,5**	0,015	1125
summe =		100	1	13975

anzahl = 9

$x_{quer} = \Sigma(x_{quer} * h_{rel}) =$

$=$ **13975** N$



antworten zu folgenden fragen:

a) wie hoch ist das durchschnitts-einkommen in Nirwana?	13975	N$
b) wo liegt das medianeinkommen (entspricht der 5ten klasse)?	20000-25000	N$
c) wo liegt das modaleinkommen (das häufigste)?	10000-15000	N$

d) warum ist die angabe des durchschnitts-einkommens irreführend?
durch die asymmetrische verteilung ist der MW nicht representativ, besser ist das modal-einkommen.

hausübung aus biostatistik - beispiel 5 von 17ten okt. 2001

Wie hoch ist die **wahrscheinlichkeit**, dass bei 2-maligem werfen eines 6-seitigen würefels die ztiffernsumme 7 ist?
Der würfel ist "gezinkt"; die ziffern 2 bis 5 erscheinen jeweils mit einer wahrscheinlichkeit von 0.15,
die ziffern 1 und 6 aber mit je 20%iger wahrscheinlichkeit.

da es sich hierum keine bedingte wahrscheinlichkeit handelt, muss regel #4 (subtraktionsregel)
in kombination mit regel #5 (multiplikationsregel) angewendet werden
P(A und B) = p(A) * p(B)
für die wahrscheinlichkeit dass die wurfergebnisse 1 und 6 7 ergeben gilt
p(7) = 0.2*0.2 = 0.04 (= 4%)
die wahrscheinlichkeit dass 2, 3, 4, und 5 eine 7 ergeben liegt bei:
p(7) = 0.15*0.15 = 0.0225 (=2.25%)
da allerdings 1 und 6 häufiger vorkommen muss man jetzt die subtraktionsregel (#4) anwenden:
p(7) = 0.04 **-** 0.0225 = 0.0175 (=**1.75%**)

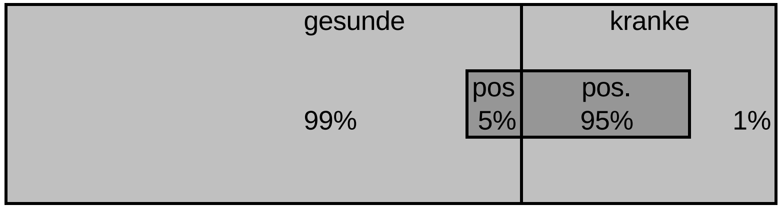
gegenannahme: bei "ungezinktem" würfel muss die wahrscheinlichkeit eine ziffernsumme 7 zu erreichen entsprechend höher sein:
p(7) = 1/36 = 0.0278 (=2.78%)

hausübung aus biostatistik - beispiel 6 vom 17ten okt. 2001

(erweitertes beispiel aus bährlocher S24): 1% der bevölkerung ist von einer bedrohlichen krankheit befallen (antrhax?)
Ein dafür entwickelter test liefert in 95% der infizierten personen ein positives ergebnis, diese werden korrekt identifiziert.
Bei gesunden personen liefert der test in 5% der fälle ein positives ergebnis, obwohl die personen gesund sind! Die bevölkerung wird jetzt durchgetestet.
Wie hoch ist die **wahrscheinlichkeit,** dass in den personen mit negativem testergebnis ein kranker ist?

Anthrax-test
1% derbevölkerung ist von anthrax befallen (krank)

Testergebnisse	neg. und gesund	95% der infizierten personen ein postives ergebnis,
	neg. und krank	5% nicht erkrankt



Frage: Vie hoch ist die wahrscheinlichkeit, dass in den personer
mit negativen testergebnis ein kranker ist?

P (neg./krank)	0,05
P (krank)	0,01
P (gesund)	0,99
P(neg/gesund)	0,95

da es sich um eine bedingte wahrscheinlihkeit handelt kommt die Bayessche formel zur anwendung:
(zu erwarten is ein sehr kleiner %-satz):

P(krank/neg.) = [P(neg./krank) * P(krank)] / [P(neg./krank) * P(krank) + P(neg./gesund) * P(gesund)]
= 0,00053135
= **0,053134963** %

hausübung aus biostatistik - beispiel 7 vom 24ten okt. 2001

In einer schweinezucht tritt eine krankheit auf, die von nur einem einzigen genlokus mit den allelen A und a gesteuert wird.
Die krankheit zeigt einen rezessiven vererbungsgang; d.h., dass nur die tiere mit dem gentypus "aa" die krankheit zeigen.
Wie hoch ist nun die **wahrscheinlichkeit,** dass bei einem wurf von 8 ferkeln aus einer paarung von zwei anlageträgern:
a) alle ferkel krank sind?
b) kein ferkel krank ist?
c) genau 2 ferkel krank sind?
d) mindestens 3 ferkel krank sind?

Da es sich um "indipendent trial" handelt, wendet man die "Binomialverteilung" an;
jeder versuch kann zwei verschiedene ereigniszustände annehmen (krank oder gesund).

genlokus	a	A
a	aa	aA
A	Aa	AA

p(krank) = 0,25
p(nicht krank) = q = 0,75

antwort auf frage "a" (alle ferkel krank):	P(k=8) =	n! / [k! * (n-k)!]	*	p^k * q^(n-k)		n =	8
	=	1	*	1,52588E-05		k =	8
	=	1,52588E-05	=	**0,001525879**	%		
antwort auf frage "b" (kein ferkel krank):	P(k=0) =	formel wie oben				n =	8
	=	1	*	0,100112915		k =	0
	=	0,100112915	=	**10,0112915**	%		
antwort auf frage "c" (2 ferkel krank):	P(k=2) =	formel wie oben				n =	8
	=	28	*	0,011123657		k =	2
	=	0,311462402	=	**31,14624023**	%		

antwort a. frage "d" (mind. 3 ferkel krank): mindestens 3 heisst 3, 4, 5, 6, 7 und 8 ferkel sind krank;
daher muss die summe der einzelberechnungen gemacht werden:

P(k=3) =	formel wie oben				n =	8
	=	56	*	0,003707886	k =	3
	= 0,207641602					
P(k=4) =	formel wie oben				n =	8
	=	70	*	0,001235962	k =	4
	= 0,086517334					
P(k=5) =	formel wie oben				n =	8
	=	56	*	0,000411987	k =	5
	= 0,023071289					
P(k=6) =	formel wie oben				n =	8
	=	28	*	0,000137329	k =	6
	= 0,003845215					
P(k=7) =	formel wie oben				n =	8
	=	8	*	4,57764E-05	k =	7
	= 0,000366211					
P(k=8) =	formel wie oben				n =	8
	=	1	*	1,52588E-05	k =	8
	= 1,52588E-05					

die wahrscheinlichkeit für die erkrankung von mind. 3 ferkel (Σ von k = 4 bis 8) liegt bei:

P(k>=3) = 0,321456909 = **32,14569092** %

hausübung aus biostatistik - beispiel 8 vom 24ten okt. 2001

Bleiben wir bei der scweinezucht;
a) wie viele männliche ferkel von 100 jungferkel werden in einer schweinezucht erwartet?
b) wie gross ist die **warscheinlichkeit,** dass mehr als 75 jungferkel männlichen geschechts sind?
c) wie viele weibliche ferkel sind am **wahrscheinlichsten** zu erwarten?

aufgrund der erfolgswahrscheinlichkeit nutzt man hier den durschnitt (μ) und die varianz (σ^2):

antwort auf frage "a" & "c" (männl./weibl. ferkel): p = 0,50
n = 100

E(x) = μ = n * p = 50 %

antwort auf frage "b" (75 männl. ferkel):

das zu erwartende wahrscheinlichkeit muss sehr klein sein da wir in diesem fall mehr als die hälfte (75) an männlichen individuen erhalten möchten (gegenüber nur 25 weiblichen individuen); um die wahrscheinlichkeit zu bestimmen muss man hier auf die binomialverteilung bezug nehmen und die daraus resultierenden wahrscheinlichkeiten [P(76) - P(100)] aufsummieren:

$$P(k>75) = \sum_{i=70}^{100}\{100! / [i! * (100-i)!] * p^i * q^{(100-i)}\}$$

männlich p = 0,50
nicht männlich q = 0,50

P(k=76) =	n! / [k! * (n-k)!]	*	p^k * q^(n-k)				
=	7,97761E+22	*	7,88861E-31			n =	100
=	6,29322E-08	=	6,29322E-06	%		k =	76
P(k=77) =	formel wie oben						
=	2,48653E+22	*	7,88861E-31			n =	100
=	1,96152E-08	=	1,96152E-06	%		k =	77
P(k=78) =	formel wie oben						
=	7,33207E+21	*	7,88861E-31			n =	100
=	5,78398E-09	=	5,78398E-07	%		k =	78
P(k=79) =	formel wie oben						
=	2,04184E+21	*	7,88861E-31			n =	100
=	1,61073E-09	=	1,61073E-07	%		k =	79
P(k=80) =	formel wie oben						
=	5,35983E+20	*	7,88861E-31			n =	100
=	4,22816E-10	=	4,22816E-08	%		k =	80

P(k=81) =	formel wie oben						
=	1,32342E+20	*	7,88861E-31			n =	100
=	1,04399E-10	=	1,04399E-08	%		k =	81
P(k=82) =	formel wie oben						
=	3,06645E+19	*	7,88861E-31			n =	100
=	2,419E-11	=	2,419E-09	%		k =	82
P(k=83) =	formel wie oben						
=	6,65013E+18	*	7,88861E-31			n =	100
=	5,24603E-12	=	5,24603E-10	%		k =	83
P(k=84) =	formel wie oben						
=	1,34586E+18	*	7,88861E-31			n =	100
=	1,0617E-12	=	1,0617E-10	%		k =	84
P(k=85) =	formel wie oben						
=	2,53338E+17	*	7,88861E-31			n =	100
=	1,99849E-13	=	1,99849E-11	%		k =	85
P(k=86) =	formel wie oben						
=	4,41869E+16	*	7,88861E-31			n =	100
=	3,48574E-14	=	3,48574E-12	%		k =	86
P(k=87) =	formel wie oben						
=	7,11054E+15	*	7,88861E-31			n =	100
=	5,60923E-15	=	5,60923E-13	%		k =	87
P(k=88) =	formel wie oben						
=	1,05042E+15	*	7,88861E-31			n =	100
=	8,28636E-16	=	8,28636E-14	%		k =	88
P(k=89) =	formel wie oben						
=	1,4163E+14	*	7,88861E-31			n =	100
=	1,11726E-16	=	1,11726E-14	%		k =	89
P(k=90) =	formel wie oben						
=	1,73103E+13	*	7,88861E-31			n =	100
=	1,36554E-17	=	1,36554E-15	%		k =	90
P(k=91) =	formel wie oben						
=	1,90223E+12	*	7,88861E-31			n =	100
=	1,5006E-18	=	1,5006E-16	%		k =	91
P(k=92) =	formel wie oben						
=	1,86088E+11	*	7,88861E-31			n =	100
=	1,46797E-19	=	1,46797E-17	%		k =	92

P(k=93) =	formel wie oben						
=	16007560800	*	7,88861E-31		n =	100
=	1,26277E-20	=	1,26277E-18	%	k =	93
P(k=94) =	formel wie oben					
=	1192052400	*	7,88861E-31		n =	100
=	9,40364E-22	=	9,40364E-20	%	k =	94
P(k=95) =	formel wie oben					
=	75287520	*	7,88861E-31		n =	100
=	5,93914E-23	=	5,93914E-21	%	k =	95
P(k=96) =	formel wie oben					
=	3921225	*	7,88861E-31		n =	100
=	3,0933E-24	=	3,0933E-22	%	k =	96
P(k=97) =	formel wie oben					
=	161700	*	7,88861E-31		n =	100
=	1,27559E-25	=	1,27559E-23	%	k =	97
P(k=98) =	formel wie oben					
=	4950	*	7,88861E-31		n =	100
=	3,90486E-27	=	3,90486E-25	%	k =	98
P(k=99) =	formel wie oben					
=	100	*	7,88861E-31		n =	100
=	7,88861E-29	=	7,88861E-27	%	k =	99
P(k=100) =	formel wie oben					
=	1	*	7,88861E-31		n =	100
=	7,88861E-31	=	7,88861E-29	%	k =	100

die wahrscheinlichkeit für 75:25 (männl.:weibl) ferkel zu erhalten (Σ von k = 76 bis 100) liegt bei:

P(k>75) = 9,05001E-06 = **0,000905001** %

da die wahrscheinlichkeiten ein männl. oder weibl. ferkel zu erhalten gleich gross sind, muss die wahrscheinlichkeit 75 männliche und nur 25 weibliche pro wurf von 100 zu erhalten sehr gering sein; daher ist dieses ergebnis sehr glaubwürdig.

hausübung aus biostatistik - beispiel 9 vom 24ten okt. 2001

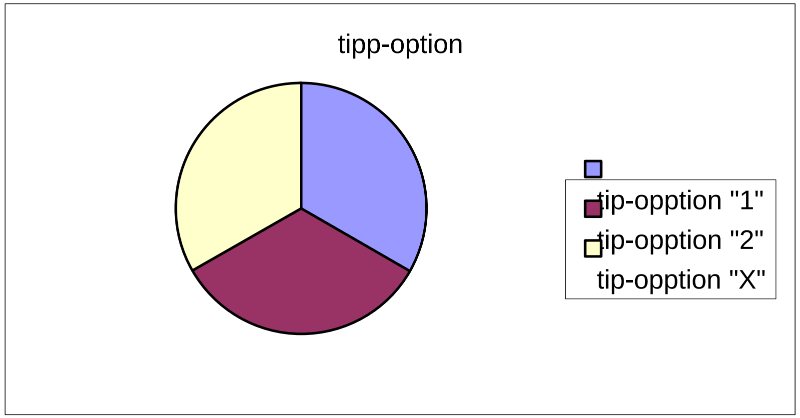
Wie gross ist die **wahrscheinlichkeit** beim toto-spielen 11 richtige tips zu erzielen?

hier liegt eine wahrscheinlichkeit nach der binomialverteilung zugrunde:
das totospiel setzt sich aus einer 3er-kombination von "X, 2, und 2"er tips zusammen; d.h. p=1/3 und q = 2/3

P(k=11) =	n! / [k! * (n-k)!]	*	p^k * q^(n-k)	
=	12	*	3,76335E-06	
=	4,51602E-05	=	**0,004516023**	%

richtiger tip p =	0,33
falscher tip q =	0,67
n =	12
k =	11
tipptypen =	3

alternativ: die gesamtmöglichen tippkombinationen sind:
n' = tipptyp^n = 3^12 =
= 531441
durch 24 falsche tipptypen liegt die wahrscheinlichkeit P(k=11) bei 24/n':
P(k=11) = 4,51602E-05 = **0,004516023** %



tip-opption "1"	=	0,33
tip-opption "2"	=	0,33
tip-opption "X"	=	0,33

hausübung aus biostatistik - beispiel 10 vom 24ten okt. 2001

Im tennisclub sind insgesmt 10 spieler; wie viele 4er-mannschaften können aus diesen 10 spielern ausgewählt werden?

hier liegt eine **wahrscheinlichkeit** nach dem kombinationsprinzip (selected and not created, Johnson p715) zugrunde:

p (k=4)	=	n! / [k! + (n-k)!]	n =	10,00
	=	**210**	k =	4,00

hausübung aus biostatistik - beispiel 11 vom 24ten okt. 2001

In einem kartenspiel mit 32 karten sind 32 karten mit 4 unterschiedlichen farben, die jeweil gleich verteilt sind (herz, karo, pick, kreuz).
Es gibt 4 könige, von jeder farbe einen; wie gross ist die **wahrscheinlichkeit**:
a) bei einmaligem ziehe den herz könig zu ziehen?
b) beim zweiten zug (ohne zurücklegen) den herz könig zu ziehen, wenn die erste karte ein herz war?

antwort auf frage "a" (herz könig): p = 1/32 = 0,03 = **3,13** % n = 32

die wahrscheinlichkeit einen herz-könig zu ziehen liegt bei 3.13%

antwort auf frage "b" (2maliger zug): es ist nur ein herz-könig als ereignisraum möglich; d.h. p = 1/(32-1)
p = 1/(32-1) = 0,032 = **3,23** %

die wahrscheinlichkeit nach 2-maligem ziehen einen herz-könig zu ziehen liegt bei 3.23%

hausübung aus biostatistik - beispiel 12 vom 24ten okt. 2001

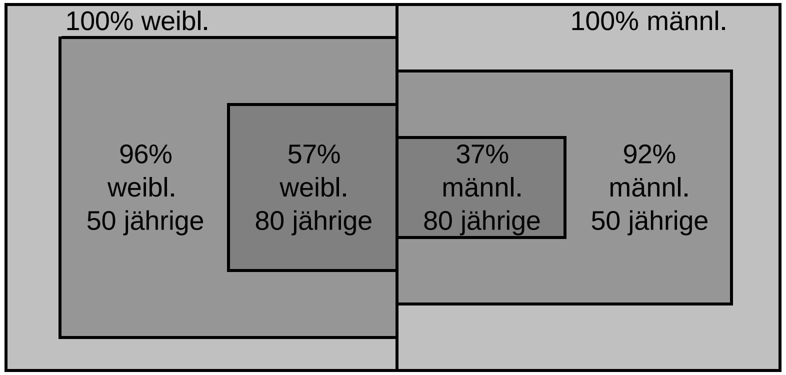
Die wahrscheinlichkeit 50 bzw 80 jahre alt zu werden ist für frauen P(Ew80) = 0.57 bzw. P(Ew50) = 0.96
und für männer P(Ew80) = 0.57 bzw. P(Ew50).
Wie gross ist die **wahrscheinlichkeit** für eine/n 50-jährige/n 80 zu werden?

Ein 50jährige/r gehört zu der gruppe der 96/92%igen die dieses alter schon erreicht haben;
daher lässt sich dieses beispiel am einfachsten mit einer schlussrechnung lösen;

alter	50jährig	80jährig
Pweibl.	0,96	0,57
Pmännl.	0,92	0,37

weibl. 0.92.....100%
0.37....x% = 0.37x100/0.92
männl. 0.96.....100%
0.57....x%

%satz der weiblichen: **59,375** %
%satz der männlichen: **40,2173913** %



hausübung aus biostatistik - beispiel 13 vom 24ten okt. 2001

Eva mit blutgruppe O hat eine tochter Elsa mit blutgruppe A. Elsa heiratet Tom mit Blutgruppe AB.
Wie gross ist die **wahrscheinlichkeit**, dass ein kind mit blutgruppe A geboren wird?

anmerkung: blutgruppe O (genotyp OO)
blutgruppe A (genotyp AA oder AO)
blutgruppe AB (genotyp AB)

genlocus	A	O
A	AA	AO
B	AB	BO

die wahrscheinlichkeit der einzeleregnisse liegt bei 1/4 (25%) da die blutgruppe A sowohl durch den genotyp AA alsauch AO hervor geht, lässt sich eine wahrscheinlichkeit von 2*1/4 (50%) erwarten;

$E(x) = \mu = n + p = 1 * 0.5$
$= 0,5$
$=$ **50** %

$p = 1/2 = 0,5$
$q = 1/2 = 0,5$
$n = 1$

da laut anmerkung sowohl AA alsauch AO vorkommt liegt eine **50%ige wahrscheinlichkeit** vor

hausübung aus biostatistik - beispiel 14 vom 30ten okt. 2001

Die erwachsenen männlichen individuen der population einer rinderrasse sei normalverteilt mit einem durschnitts-gewicht von 550kg
und einer standardabweichung von 32kg.
Wie gross ist der anteil von stieren, die schwerer als 600kg sind?

diesem problem liegt eine gauss'sche normalverteilung zugrunde; daher ist es angebracht mithilfe der zufallsvariable "z" es zu lösen.
P(x>600)?

$z = (x - \mu) / s$
$= (600-550)/32$
$= 1{,}5625$

x =	600	kg
μ =	550	kg
σ =	32	kg

laut tabelleneintrag (std-norm-verteilung, Johnson p724) resultiert eine wahrscheinlichkeit von:

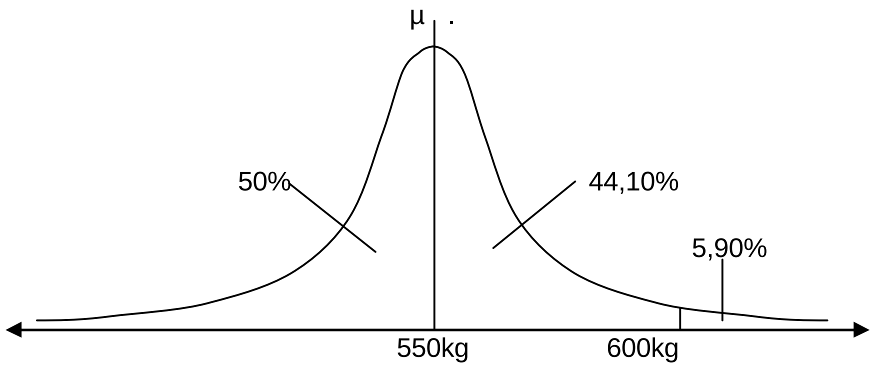
$P_l = 0{,}4406$

da sch der tabelleneintrag auf die halbe fläche der kurve bezieht muss noch 0.5 hinzugerechnet werden um jene wahrscheinlichkeit zu bestimmen in der die tiere zusammengefasst sind die leichter als 600kg sind;

$P_G = 0{,}9406$

um die wahrscheinlichkeit jener tiere zu erfassen die schwerer als 600kg sind, muss P_G von der relativierten norm-summenfläche abgezogen werden:

$P(x>600) = 1-P_G =$ 0,0594 oder = **5,94** %



hausübung aus biostatistik - beispiel 15 vom 30ten okt. 2001

Kühe sind leichter im schnitt; ihr durschnittliches gewicht (erwartugswert) ist gut bekannt und er liegt 40kg unter dem der stiere, die standardabwechung sei aber gleich gross. Wie gross is dann der anteil der kühe, die kleiner sind als der durchschnittswert der stiere?

Diesem problem liegt eine gauss'sche normalverteilung zugrunde; daher ist es angebracht mithilfe der zufallsvariable "z" es zu lösen.

$P(x<\mu)$?

aus dem vorigen beispiel sind die zur berechnung erforderlichen daten zu beziehen: $\Delta =$ 40 kg

somit kann die gesuchte wahrscheinlichkeit folgendermassen präzisiert werden: $\mu =$ 550 kg

$P(510 < x < 550)$ $\sigma =$ 32 kg

$z = (x - \mu) / s$

$= (510-550)/32$

$= -1{,}25$

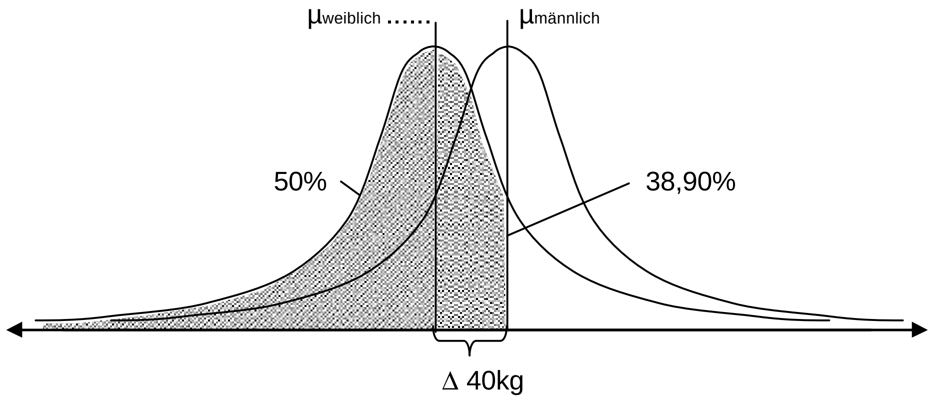
laut tabelleneintrag (std-norm-verteilung, Johnson p724) resultiert eine wahrscheinlichkeit von:

$P_l = 0{,}389$

da sch der tabelleneintrag auf die halbe fläche der kurve bezieht muss noch 0.5 hinzugerechnet werden um jene wahrscheinlichkeit zu bestimmen in der die tiere zusammengefasst sind die

$P(510 < x < 550) = 0{,}889$ oder = **88,9** %

d.h. 88.9% aller weiblichen kühe liegen unter dem durschnitsgewicht ihrer männlichen kollegen.



hausübung aus biostatistik - beispiel 16 vom 30ten okt. 2001

Wie gross ist unter den gegebenen annahmen die wahrscheinlichkeit, dass eine kuh zwichen 480 und 540 kg schwer ist?

diesem problem liegt eine gauss'sche normalverteilung zugrunde; daher ist es angebracht mithilfe der zufallsvariable "z" es zu lösen.
P(480 < x < 540)?

aus dem vorigen beispiel sind die zur berechnung erforderlichen daten zu beziehen:................

σ = 32 kg
μ = 510 kg
unteres limit = 480 kg
oberes limit = 540 kg

$z = (x - \mu) / \sigma$
$z_{480} = (480-510)/32$
$= -0,9375$
$z_{540} = (540-510)/32$
$= 0,9375$

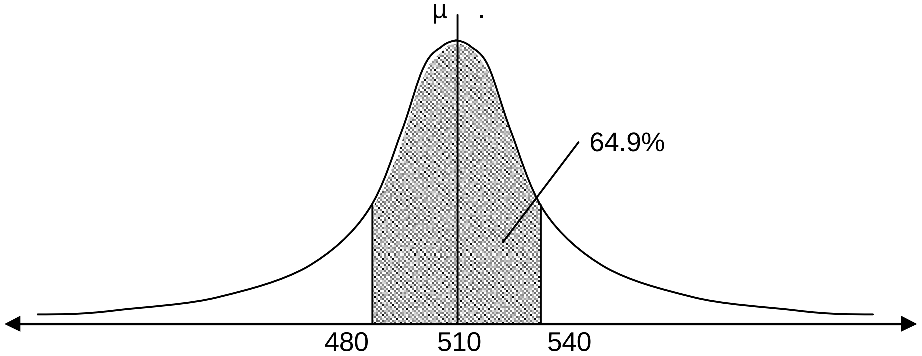
laut tabelleneintrag (std-norm-verteilung, Johnson p724) resultiert eine wahrscheinlichkeit von:

$P_{L=R} = 0,3245$

da sch der tabelleneintrag auf die halbe fläche der kurve bezieht und zudem die ergebnisse symmetrisch zum mittelwert liegen wird zur bestimmung der wahrscheinlichkeit P_G, $P_{L=R}$ doppelt gerechnet:

$P_G = 0,649$ oder = **64,9** %

d.h.64,9% aller weiblichen kühe liegen innerhalb dieses spektrums.



hausübung aus biostatistik - beispiel 17 vom 30ten okt. 2001

(Annahme!) die verkehrsüberwachung auf autobahnen in einem land, das bekannt ist für seine raser, hat gezeigt, dass die geschwindigkeit der KFZ eine normalverteilte ZV ist. 90% fahren langsamer als 160 km/h, aber nur 5% fahren langsamer als 120 km/h. Wie hoch ist die durchschnitts-geschwindigkeit und wie hoch ist der anteil der geschwindigkeitsverliebten, die über 170 km/h unterwegs sind?

diesem problem liegt eine gauss'sche normalverteilung zugrunde; daher ist es angebracht mithilfe der zufallsvariable "z" über die wahrscheinlichkeiten zurück zu rechnen; da sch der tabelleneintrag auf die halbe fläche der kurve bezieht, muss noch 0.5 von den 0.9 (90%) abgezogen werden:

a) antwort auf frage "a"; wie hoch ist die durchschnitts-geschwindigkeit (μ)?

P_{160} : 0.9 = 0.5 + 0.4 $\qquad$ P_{160} = 90 %

$P_{160'}$ = 0.4

laut tabelleneintrag (std-norm-verteilung, Johnson p724) resultiert eine zufallswert "z" von:

z_{160} = 1,29

P_{120} : 0.05 = 0.5 - 0.45 $\qquad$ P_{120} = 5 %

$P_{120'}$ = 0.45

laut tabelleneintrag (std-norm-verteilung, Johnson p724) resultiert eine zufallswert "z" von:

z_{120} = -1,645

z = (x - μ) / σ $\qquad$ durch diese zwei zufallswerte erhält man 2 gleichungen mit 2 unbekannten:

z_{160} : 1.29 = (160 - μ) / σ $\qquad$ umgeformt: μ = 160 - 129 * σ

z_{120} : -1.645 = (160 - μ) / s $\qquad$ und in folgende eingesetzt:

-1645 = [120 - (160 - 1,29 * σ) / σ]

umgeformt resultiert ein σ:

σ = 40 / 2.94 = **13,63** km/h

in μ eingesetzt:

μ = 160-1,29*s = **142,4** km/h

b) antwort auf frage "b"; wie hoch ist der anteil der geschwindigkeits-verliebten P(x>170)?

hausübung aus biostatistik - beispiel 18 vom 7ten nov. 2001

Nehme wir an, aus einer normalverteilung mit dem mittelwert 20 g (z.B. das gewicht kleiner mäuse in g) und der varianz 12g^2
wird eine (mäuse-) probe von n = 15 gezogen.
Berechne die wahrscheinlichkeit, dass der durchschnittswert zwischen 17 g und 23 g liegt.

ges.: P(17 < Z < 23)

N (20; 3.46)

σ^2 =	12	g^2
σ =	3,5	g^2
μ =	20	g
n =	15	-
lower limit =	17	g
upper limit =	23	g

der lösungsansatz dieses beispiels ist am besten mit der normalverteilung zu machen:

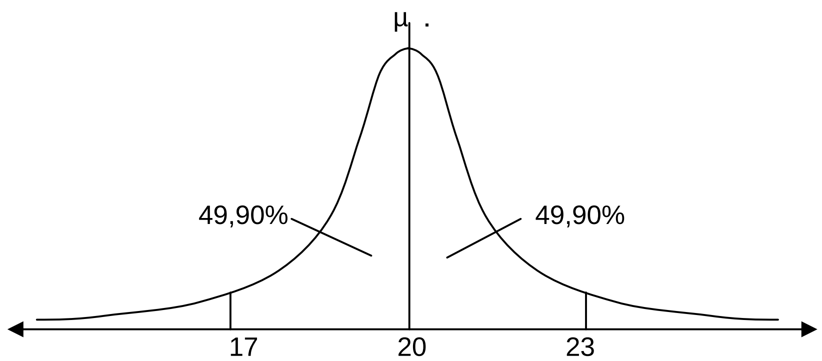
z = (x - μ) * wurzel(n) / σ

z_1 = (17 - 20) * wurzel (n) / 3.46 =
-3,35410197
laut tabelle (Johnson p724) resultiert eine wahrscheinlichkeit von:
P_1 = -0,4996

z_2 = (23 - 20) / 3.46 =
3,35410197
laut tabelle (Johnson p724) resultiert eine wahrscheinlichkeit von:
P_1 = 0,4996

P_G(17<Z<23)= P(0<Z<3.1) - P(-0.31<Z<0) =
= 0,9992 oder = **99,92** %

die wahrscheinlichkeit dass P(17<Z<23) dazwischen liegt beträgt 99.9%



d) P(x>170)? $z = (x - \mu) / s$ x = 170 km/h

$z_{170} = (170 - 142.4)/ 13.63 =$ 2,02494497

laut tabelleneintrag (std-norm-verteilung, Johnson p724) resultiert eine wahrscheinlichkeit von:

$P_R =$ 0,4785

da sch der tabelleneintrag auf die halbe fläche der kurve bezieht muss noch 0.5 hinzugerechnet

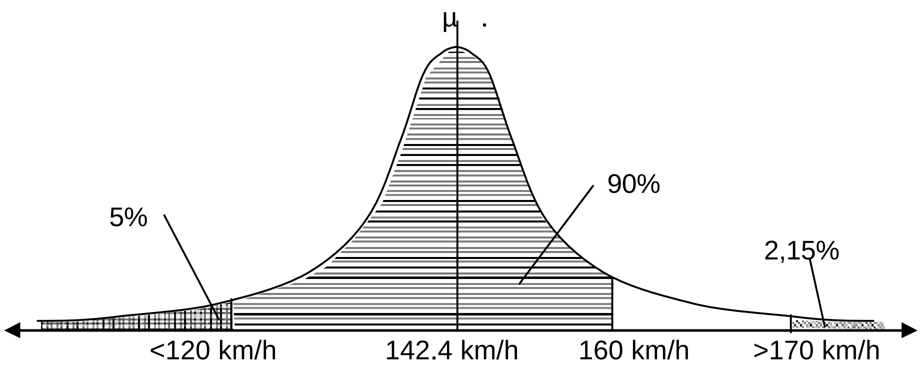
$P_G =$ 0,9785 oder = 97,85 %

demnach fahren 97.9% aller auofahrer langsamer als 170km/h;

jener anteil der schneller unterwegs ist errechnet sich aus:

$P_{170} =$ 0.9785 = 0,0215 oder = **2,15** %

d.h. 2.15% aller autofaherer sind schneller als 170km/h unterwegs.



hausübung aus biostatistik - beispiel 19 vom 7ten nov. 2001

Eine stichprobe vom umfang n= 10 wird gezogen aus einer grundgesamtheit mit μ = 180 und σ = 12. N(20; 3.46)
a) bestimme die wahrscheinkichkeit, dass der mittelwert der probe zwischen 178 und 182 liegt.
b) über welchen wert x liegt der durchschnittswert mit einer wahrscheinlichkeit von 0.01?

antwort auf frage "a" P(178 < Z 182): wiederum isdt der lösungsansatz am besten mit der normalverteilung zu machen:

σ =	12	g^2
μ =	180	g
n =	10	-
lower limit =	178	g
upper limit =	182	g

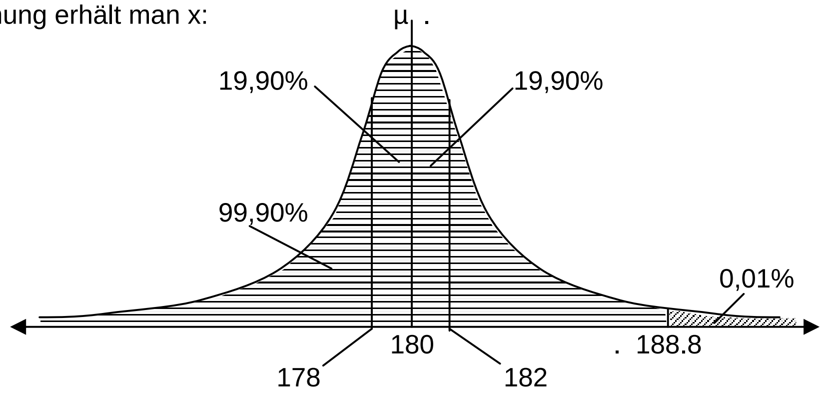
z = (x - μ)* wurzel(n) / σ
z_1 = (178 - 180) / 12 =
= -0,52704628 laut tabelle (Johnson p724) resultiert eine wahrscheinlichkeit von:
P_1 = -0,199

z_1 = (182 - 180) / 12 =
= 0,52704628 laut tabelle (Johnson p724) resultiert eine wahrscheinlichkeit von:
P_1 = 0,199

P(178 < Z < 182) = P(0<Z<0.167) - P(-0.167<Z<0) =
= 0,398 oder = **39,8** %
die wahrscheinlichkeit dass P(178 < Z < 182) dazwischen liegt beträgt rund **40%**

antwort auf frage "b" P(>0.99); i.e. dass die wahrscheinlichkeit bei >99.9% liegt: laut tabelle (Johnson p724) resultiert eine wahrscheinlichkeit von:

z = 2,325 durch umformen der obigen gleichung erhält man x:
x = z * σ/wurzel(n) + μ
= 0,0025 * 12/wurzel(10) + 180 =
= **188,822755**
bei einer wahrscheinlichkeit von 1% liegt der
wert mit 0.03 über dem durschnittswert



hausübung aus biostatistik - beispiel 20 vom 7ten nov. 2001

Von einer fischpopulation ist bekannt, dss sie normalverteilt ist; eine stichprobe vom umfang n=20 ergibt einen mittelwert x_{quer} = 1,5kg und einer varianz von σ^2 = 0,0225kg^2. Gib ein konfidenzintervall (KI) von 90% für den wahren mittelwert μ der verteilung an.

ges.: x_{quer} +/- ?g bei einem KI von 90%

σ^2 =	0,023	kg^2
σ =	150	g
μ =	1500	g
n =	20	-

die lösung ist hier mit der standard normalverteilung anzusetzen:

$z = (x_{quer}-\mu) / [\sigma/\text{wurzel}(n)]$

laut tabelle (Johnson p725) ergibt sich für $z(\alpha/2)$:

$z_{\alpha/2} = 1,65$

der maximal zu erwartende schätzfehler (max. error of estimate):

$E = \mu - x_{quer} = z_{\alpha/2} * \sigma / \text{wurzel}(n)$

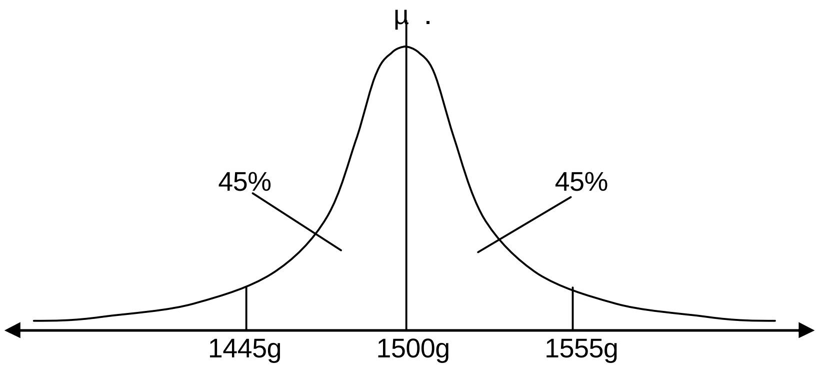
$= 1.65*0.15 / \text{wurzel}(20)$

$= 55,3426824$

für die oberen und unteren grenzwerte ergibt sich - nach formel-umwandlung $\mu = x_{quer} - z*\sigma/\text{wurzel}(n)$:

1.5 +/- 0.0553 = **1.4447** bis **1.5553**

bei 90%igem konfidenzintervall ergibt sich eine mittelwert-verteilung von **1.44 bis 1.56kg**.



hausübung aus biostatistik - beispiel 21 vom 7ten nov. 2001

Die fischpopulation in (20) sei normalverteilt mit σ^2 = S^2 = 0,0225 kg^2.
Wie viele fische muss eine stichprobe umfassen um den mittelwert auf +/-20g bei einem vertrauensniveau von 90% aneben zu können?

ges.: stichproben-groesse "n" (1500 +/-20g bei einem KI von 90%)?

warum ist n angegeben, nur des df's wegen, wenn doch ein neues n berechnet wird?
d.h. der der ansatz der t-verteilung ist aufgrund der angabe der standard-abweichung hier nicht angebracht;
es ist aus diesem grund besser mit der z-verteilung zu rechnen.
da die standardabweichung bekannt ist, findet man die lösung per z-verteilung (Johnson p.354):

σ^2 =	0,023	kg^2
σ =	150	g
x_{quer} - μ = +/-	20	g

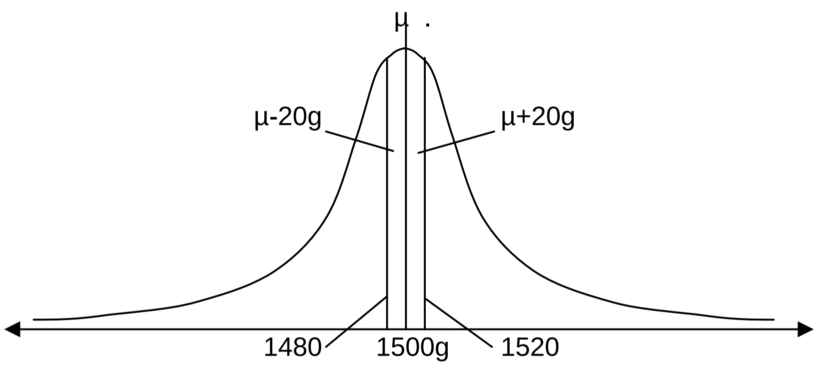
z = (x_{quer}-μ) / [σ/wurzel(n)]

laut tabelle (Johnson p724) ergibt sich für +/-z (bei α/2):
z = 1,645 bzw.
$z_{1/2}$ = +/-1645

nach umformung oben stehender gleichung ergibt sich:

n = [z*s / (x_{quer} - μ)]^2 = [1.33 * 150 / 20]^2 =
= **152,213906**

demnach muss die stichprobe mindestens **153 fische** umfassen um in einem konfidenzintervall von 90% einen gewichts-mittelwert von P(1480 < x_{quer} < 1520) = 1500+/-20g zu erzielen.



hausübung aus biostatistik - beispiel 22 vom 7ten nov. 2001

Nehmen wir an, das verhältnissv des gewichtes zur körpergrösse (v=Gev/Kgr) der individuen der population des *Homo sapiens* sei normalverteilt mit μ_o und σ, v~ N(μ_o,σ). Alle individuen, deren v innerhalb μ +/-σ liegt, gelten als normalgewichtig.
Die bewohner eines dürregebietes im süden sind schrecklich unterernährt. Bei ihnen hat sich das durchschnittliche Verhältniss v (µ1) so weit nach links verschoben, dass es bei μ_o-s zu liegen kommt, also: $\mu_1 = \mu_o$-s.
Die standardabweichung s dieser neuen verteilung bleibt llerdings gleich. Wie hoch ist
a) der anteil der bevölkerung, die noch als normalgewichtig zu betrachten sind?
b) der anteil der vom hunger extrem bedrohten personen, wenn diese grenze bei μ_1-σ liegt?

antwort auf frage "a" (%-ueller anteil der normalgewichtigen):

sowohl hunger-gesellschaft alsauch normalgewichtige gesellschaft sind normalverteilt; lediglich die hunger-gesellschaft ist um den betrag der standard-abweichung nach links verschoben (in bezug zur normalgewichtigen gesellschaft);
geg.: $\mu_1 = \mu_0 - \sigma$
ges.: $\mu_1 + \sigma$ in prozent;

da keine zahlenangaben vorhanden sind findet man den ansatz über die z-verteilung; der anteil der normalgewichtigen ist daher innerhalb der positiven abweichungen von μ_1 zu finden ($\mu_1 + 2\sigma$);

$(\mu_1 + 2\sigma) = \phi(Z) - \phi(0)$
$= 2 - 0$
$\phi(Z) = 2$
$\phi(0) = 0,5$ entspricht einer wahrscheinlichkeitswert von 50%; i.e. 0.5

laut tabelleneintrag (Johnson p724) ergibt sich ein wahrscheinlichkeitswert von:
$\phi(Z) = 0,9772$
$\phi(0) = 0,5$
$(\mu_1 + 2\sigma) =$ **0,4772** oder = 47,72 %

folglich sind nur fast 48% der hunger-gesellschaft als normalgewichtig zu bezeichnen;

hausübung aus biostatistik - beispiel 23 vom 15ten nov. 2001

Du befindest dich im casino und hast beim roulett-spiel beobachtet, dass während der letzten 10 kugeln 8 mal eine rote zahl "errolt" wurde. Nachdem es von insgesamt 37 möglichkeiten 18 möglichkeiten für "rot" gibt (18 für "schwarz" und 1 für "grün"), erscheint dir das beobachtete ergebnis sehr hoch. Du vermutest dass die wahrscheinlichkeit für "rot" grösser ist; begründe deine vermutung, definiere die relative null- und alternativ- hypothese und erstelle eine tabelle für die kritische region α (5%-bereich).

ges. H_0 und H_A;

H_0: μ_{rot} >gleich μ_0; H_A: $\mu_{rot} < \mu_0$;

da es sich bei diesem beispiel um eine diskrete verteilung handelt, benutzt man hier die binomialverteilung;

daher muss man zuerst dieses ergebniss auf seine wahrscheinlichkeit hin überprüfen:

						n =	10
P(k=8) =	n! / [k! * (n-k)!]	*	p^k * q^(n-k)			k =	8
=	45	*	0,00082731			p_{rot} = 18/37 =	0,48648649
=	0,037229098	=	**3,72290981**	%		$q_{n/\,rot}$ = (18+1)/37 =	0,51351351
						$p_{grün}$ = 1/37 =	0,03
						$p_{schw.}$ = 18/37 =	0,48648649

um diesen wert mit dem 5%igen signifikanz-niveau in relation zu stellen

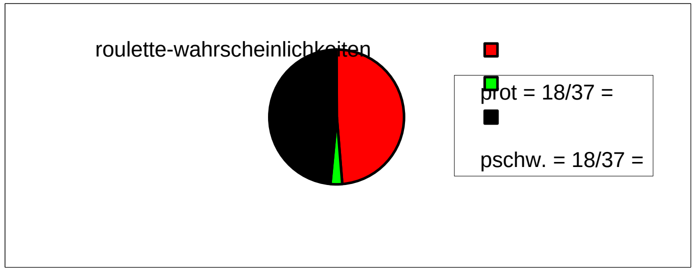
muss die aufsummierte wahrscheinlichkeit von k = 7, 8, 9, 10 berechnet werden:

					n =	10
P(k=7) =	n! / [k! * (n-k)!]	*	p^k * q^(n-k)		k =	7
=	120	*	0,00087328		p_{rot} =	0,48648649
=	0,104793017	=	**10,4793017**	%	$q_{n/\,rot}$ =	0,51351351
					n =	10
P(k=9) =	formel wie oben				k =	9
=	10	*	0,00078377		p_{rot} =	0,48648649
=	0,007837705	=	**0,78377049**	%	$q_{n/\,rot}$ =	0,51351351
					n =	10
P(k=10) =	formel wie oben				k =	10
=	1	*	0,00074252		p_{rot} =	0,48648649
=	0,000742519	=	**0,07425194**	%	$q_{n/\,rot}$ =	0,51351351

k	....	5	6	7	8	9	10
in [%]				10,4793017	3,72290981	0,78377049	0,07425194
$\Sigma_{(c>8)}$ =					4,58093224		
$\Sigma_{(c>7)}$ =				15,0602339			

5% margine

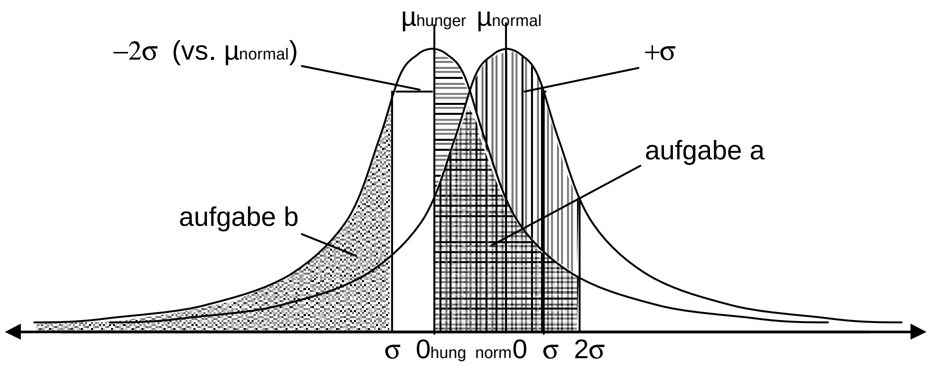
folglich liegt der wert von 3.7% innerhalb des kritischen 5%-bereichs; folglich liegt das beobachtete resultat doch signifikant über den zu erwartenden wert (bei gegebenem signifikanz-niveau von 5%).
Die nullhypothese (H_0) ist demnach beizubehalten (rot kommt öfter - tisch ist "getürkt")!



antwort auf frage "b" (%-ueller anteil der untergewichtigen):

geg.: $\mu_1 = \mu_0 - \sigma$
ges.: prozentueller anteil der unterhalb des schwellwertes von $\mu_1 - \sigma$ liegt
$(\mu_1 - \sigma) = \phi(-1)$
laut tabelleneintrag (Johnson p724) ergibt sich ein wahrscheinlichkeitswert von:
$\phi(-1) =$ **0,159** oder = 15,9 %

folglich sind rund 16% der hunger-gesellschaft als extrem untergewichtig zu bezeichnen;



hausübung aus biostatistik - beispiel 24 vom 15ten nov. 2001

Die populationsdichte einer seltenen tierart ist sehr niedrig. Man nimmt an, dass sie bei 20 individuen pro 100km2 liegt.
a) in einer beobachtungsfläche von 10km2 wurde kein exemplar dieser tierart gesichtet. Kann nan mit recht behaupten, dass in diesem gebiet diese tierart nicht vorkommt?
b) in einer anderen beobachtungsfläche derselben grösse wurden 5 tiere beobachtet. Ist die annahme, dass die tiere gehäuft in diesem gebiet vorkommen, aus statistischer sicht gut zu begründen?
stelle für "a" und "b" die entsprechenden hypothesen und kritischen bereiche auf.

da es sich um eine relativ kleine population handelt rechnet man mit der poisson-verteilung (formel lt. Johnson p. 329):

antwort auf frage "a" (kein tier gesichtet): H_0: $\mu_{n/tier}$ <gleich $\mu_{2-tiere}$; H_A: $\mu_{n/tier}$ > $\mu_{2-tiere}$;

P(k=0) = e^-μ * μ^k / k!				μ = 10 * 20/100 = 2 / 10 km^2
= 0,13533528	=	**13,5335283**	%	k = 0

bei einem signifikanz-niveau von 5% ist das ergebnis durchaus möglich; würde man das nicht tun würde man ein fehler erster ordnung begehen indem man 13.5% ignorieren würde; daher ist die null-hypothese beizubehalten!

antwort auf frage "b" (5 tiere gesichtet): H_0: $\mu_{5-tiere}$ > $\mu_{2-tiere}$; H_A: $\mu_{5-tiere}$ <gleich $\mu_{2-tiere}$;

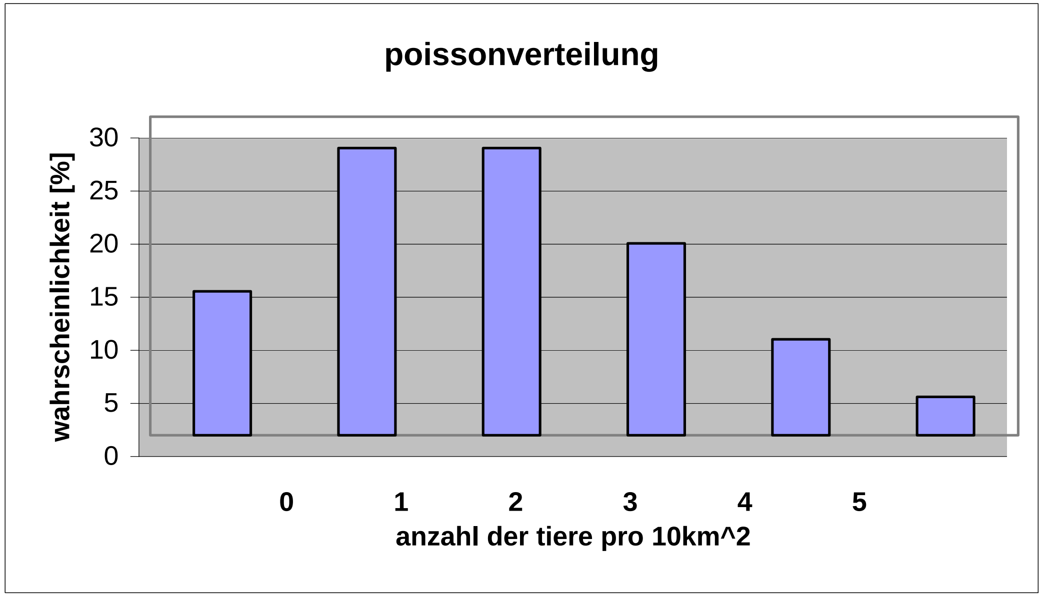
P(k=5) = e^-μ * μ^k / k!				μ = 2 / 10 km^2
= 0,03608941	=	**3,60894089**	%	k = 5

da nun allerdings ein durschnittswert von nur 2 tieren pro 10km2 zu erwarten sind muss man die fälle für k = 0,1,2,3,4 noch berücksichtigen P(x>5) = 1-p(x<5):

P(k=1) = e^-μ * μ^k / k!			μ = 2 / 10 km^2
= 0,27067057	=	**27,0670566**	k = 1
P(k=2) = e^-μ * μ^k / k!			μ = 2 / 10 km^2
= 0,27067057	=	**27,0670566**	k = 2
P(k=3) = e^-μ * μ^k / k!			μ = 2 / 10 km^2
= 0,18044704	=	**18,0447044**	k = 3
P(k=4) = e^-μ * μ^k / k!			μ = 2 / 10 km^2
= 0,09022352	=	**9,02235222**	k = 4

c	0	1	2	3	4	5
p(x=k) in [%]	13,5335283	27,0670566	27,0670566	18,0447044	9,02235222	3,60894089
$\Sigma_{(c<5)}$ =					94,7346983	
1-p(x<5) =	$100 - \Sigma_{(c<5)}$ =				5,26530173	

bei einem signifikanz-niveau von 5% fällt das ergebnis in den kritischen bereich;
in diesem fall ist die null-hypothese zu verwerfen!



hausübung aus biostatistik - beispiel 25 vom 15ten nov. 2001

Der blutdruck von erwachsenen personen in der altersklasse zwischen 50 und 60 jahren sei mit einem mittel auf 140mm Hg gesenkt worden. Nun testet ein arzt ein neues medikament und kommt zu folgenden messergebnissen:

135	140	148	132	137	139	141	136	136

a) der arzt will nun wissen, ob dieses neue medikament wirksamer ist als das bisher verwendete medikament.
Führe den test mit $\alpha_{eins} = 0.01$ und $\alpha_{eins} = 0.05$ aus.
b) ändert sich die interpretation wenn man das messergebnis von 148 als ausreisser betrachtet?

antwort auf frage "a" ($\alpha_{eins} = 0.01$ und $\alpha_{eins} = 0.05$):

da die standardabweichung nicht bekannt ist, rechnet man mit der t-verteilung (Johnson p.416);
H_0: μ_{neu} <gleich μ_0; H_A: $\mu_{neu} > \mu_0$; da es sich um einen grösser / kleiner vergleich handelt beruhft man sich auf eine **1-seitige t-verteilung**; dazu muss zuerst die der mittelwert ermitelt werden um die standard-abweichung s berechnen zu können:

	x_i	$x_{quer} - x_i$	$(x_{quer} - x_i)$^2
x_1	135,00	3,22	10,382716
x_2	140,00	-1,78	3,16049383
x_3	148,00	-9,78	95,6049383
x_4	132,00	6,22	38,7160494
x_5	137,00	1,22	1,49382716
x_6	139,00	-0,78	0,60493827
x_7	141,00	-2,78	7,71604938
x_8	136,00	2,22	4,9382716
x_9	136,00	2,22	4,9382716
x_{quer} =	138,22		167,555556

s^2 = 1 / (n-1) * $\Sigma(x_{quer}$ - xi)^2   n = 9
= 0,125 * 167,555556
= 20,9444444
s = wurzel (s^2)
= 4,57651007

eingesetzt in die t-verteilung ergibt sich:

$t = (x_{quer}-\mu_0) / [s/wurzel(n)]$
$= -1{,}16537127$

$\mu_0 = 140$ mm Hg
$s = 4{,}57651$ mm Hg
$x_{quer} = 138{,}22$ mm Hg
$n = 9$ -
$df = n-1 = 8$ -

laut tabelle (Johnson p.727) ergibt sich ein t-wert von

α	t(df;α)	t_k aus tabelle	t_p
0.01	t(8;0.01)	-2.9	-1.2
0.05	t(8;0.05)	-1.86	-1.2

sowohl bei einem signifikanz-niveau von 1 bzw. 5% liegt das resultat ausserhalb des kritischen bereiches (im linken, negativen Teil der T-verteilung); folglich ist die null-hypothese beizubebalten; womit besagt ist dass das medikament nicht blutdruck- -senkend wirkt (H_0: $\mu>\mu_0$; H_A: $\mu<\mu_0$)

antwort auf frage "b" (ohne den 148er ausreisser):

	x_i	$x_{quer} - x_i$	$(x_{quer} - x_i)$^2
x_1	135,00	2,00	4
x_2	140,00	-3,00	9
x_3	132,00	5,00	25
x_4	137,00	0,00	0
x_5	139,00	-2,00	4
x_6	141,00	-4,00	16
x_7	136,00	1,00	1
x_8	136,00	1,00	1
x_{quer} =	137,00		
Σ =			60

s^2 = $\Sigma(x_{quer}$ - xi)^2 / (n-1)
= 60 / 7
= 8,57142857
s = wurzel (s^2)
= 2,92770022

n = 8

eingesetzt in die t-verteilung ergibt sich:

$t = (x_{quer} - \mu_0) / [s/wurzel(n)]$
$= -2{,}89827535$

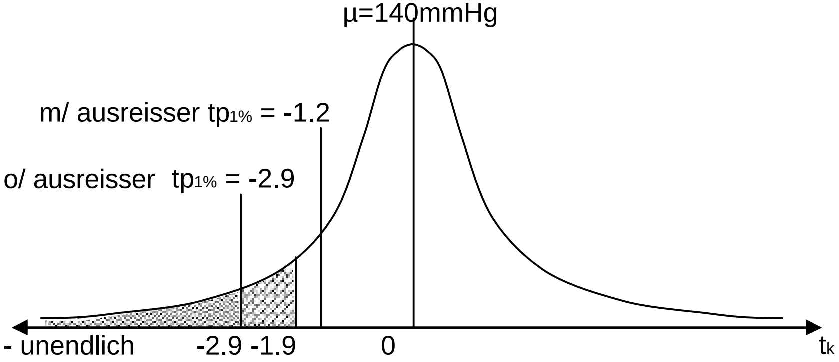
$\mu_0 =$	140	mm Hg
s =	2,9277	mm Hg
$x_{quer} =$	137,00	mm Hg
n =	8	-
df = n-1 =	7	-

laut tabelle (Johnson p.727) ergibt sich ein t-wert von

α	t(df;α)	$t_{krit.}$ tabelle	t_p	α für t-dist (laut excel)	
0.01	t(7;0.01)	-2.9	**-2.9**	0,01152127	1,20%
0.05	t(7;0.05)	-1.86	**-2.9**	0,01152127	1,20%

bei einem signifikanz-niveau von 1% liegt das resultat gerade noch unterhalb des kritischen bereiches (im linken, negativen Teil der T-verteilung); folglich ist die null-hypothese zu verwerfen; womit besagt ist dass das medikament **nicht** blutdrucksenkend wirkt;

bei einem signifikanz-niveau von 5% liegt das resultat innerhalb des kritischen bereiches (im linken, negativen Teil der T-verteilung); folglich ist die null-hypothese zu akzeptieren; womit besagt ist dass das medikament blutdrucksenkend wirkt;



hausübung aus biostatistik - beispiel 26 vom 15ten nov. 2001

Ein getreidefeld wird in 10 gleich grosse versuchsflächen aufgeteilt, auf denen ein düngeversuch mit einer neuen düngemethode durchgeführt wird. Der versuch liefert nach der ernte folgende ergebnisse [in kg]:

1350	1500	1250	1450	1440	1530	1300	1440	1580	1510

Aus erfahrung ist bekannt, dass die ertragserwartung für eine fläche bei 1350kg im langjährigen durchschnitt liegt. Unter annahme dass die stichprobe aus einer normalverteilung stammt, ist zu überprüfen ob die neue düngemethode höhere erträge bringt.
Ist das erbegnis signifikant auf dem 5% niveau und auf dem 1% niveau, bzw. auf welchem niveau ist der unterschied signifikant?

ges.: H_0: $\mu>\mu_0$; H_A: $\mu<\mu_0$; da es sich um einen grösser / kleiner vergleich handelt kommt die 1-seitige t-verteilung zum zug;
a) bei α (0.01) b) bei α (0.05)

da die standardabweichung nicht bekannt ist, rechnet man mit der t-verteilung (Johnson p.77):
dazu muss zuerst die der mittelwert ermitelt werden um die standard-abweichung s berechnen zu können:

	x_i	$x_{quer} - x_i$	$(x_{quer} - x_i)$^2
x_1	1350,00	85,00	7225
x_2	1500,00	-65,00	4225
x_3	1250,00	185,00	34225
x_4	1450,00	-15,00	225
x_5	1440,00	-5,00	25
x_6	1530,00	-95,00	9025
x_7	1300,00	135,00	18225
x_8	1440,00	-5,00	25
x_9	1580,00	-145,00	21025
x_{10}	1510,00	-75,00	5625
x_{quer} =	1435,00		
Σ =			99850

s^2 = $\Sigma(x_{quer}$ - xi)^2 / (n-1)   n = 10
= 99850 / 9
= 11094,4444
s = wurzel (s^2)
= 105,330169

eingesetzt in die t-verteilung ergibt sich:

$t = (x_{quer} - \mu_0) / [s/\text{wurzel}(n)]$
$= 2{,}55191465$

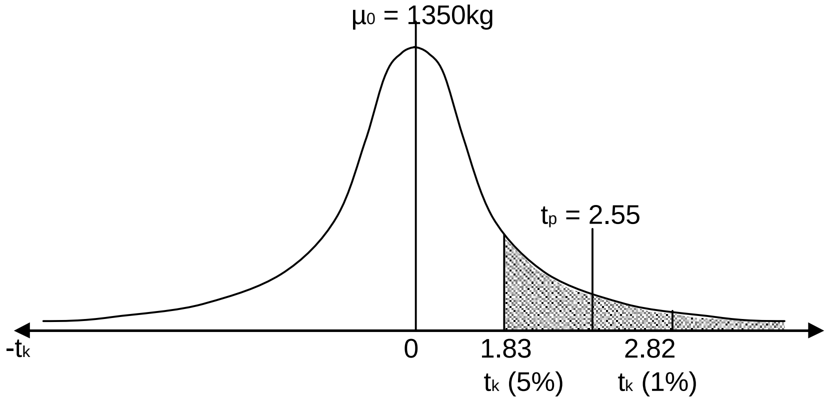
μ_0 =	1350	kg
s =	105,3302	kg
x_{quer} =	1435,00	kg
n =	10	-
df = n-1 =	9	-

laut tabelle (Johnson p.727) ergibt sich ein t-wert von

α	$t(df;\alpha)$	t_k aus tabelle	t_p	α für t-dist (laut excel)	
0.01	t(9;0.01)	2.82	**2.55**	0,01554845	1,60%
0.05	t(9;0.05)	1.83	**2.55**	0,01554845	1,60%

bei einem signifikanz-niveau von 1 % liegt das resultat ausserhalb des kritischen bereiches (im rechten, positiven Teil der T-verteilung); folglich ist die null-hypothese beizubebalten; womit besagt ist dass mit der neuen düngemethode ein höherer ernteertrag erzielbar ist (H_0: $\mu>\mu_0$; H_A: $\mu<\mu_0$)

bei einem signifikanz-niveau von 5 % liegt das resultat innerhalb des kritischen bereiches (im rechten, positiven Teil der T-verteilung); folglich ist die null-hypothese zu verwerfen; womit besagt ist dass mit der neuen düngemethode kein höherer ernteertrag erzielbar ist (H_0: $\mu<\mu_0$; H_A: $\mu>\mu_0$)



hausübung aus biostatistik - beispiel 27 vom 28ten nov. 2001

In einem see werden an 2 verschieden weit entfernten stellen jeweils 10 proben einer fischart gefangen und gewogen.
Gibt es irgendwelche hinweise darauf, dass sich die populationen in den beiden unterschiedlichen gebieten bezüglich ihres durchschnittlichen gewichtes unterscheiden?
Gewicht der fischproben [g]:
standort #1 458, 443, 502, 494, 478, 494, 510, 520, 481, 475;
standort #2 525, 488, 510, 456, 479, 465, 515, 490, 487, 473;

ges: H_0:$\mu = x_{1quer} = x_{2quer}$ H_A: μ ungleich x_{1quer} ungleich x_{2quer}
da es sich um eine (un-)gleichheit handelt bezieht man sich auf eine **2-seitige t-verteilung**;

	standort 1 - angaben in [g bzw g^2]			standort 2 - angaben in [g bzw g^2]		
n	x_i	$x_i - x_{quer}$	$(x_i - x_{quer})$^2	x_i	$x_i - x_{quer}$	$(x_i - x_{quer})$^2
1	458	-27,5	756,25	525	36,2	1310,44
2	443	-42,5	1806,25	488	-0,8	0,64
3	502	16,5	272,25	510	21,2	449,44
4	494	8,5	72,25	456	-32,8	1075,84
5	478	-7,5	56,25	479	-9,8	96,04
6	494	8,5	72,25	465	-23,8	566,44
7	510	24,5	600,25	515	26,2	686,44
8	520	34,5	1190,25	490	1,2	1,44
9	481	-4,5	20,25	487	-1,8	3,24
10	475	-10,5	110,25	473	-15,8	249,64
Σ	4855		4956,5	4888		4439,6
x_{quer} =	485,5			488,8		
*s =	23,4674716			22,2101078		

$n_1 = n_2 = 10$

*) die standard-abweichung berechnet sich laut folgender formel:
s = wurzel {[1/(n-1)] * $\Sigma(x_i - x_{quer})$^2 }

die prüfgrösse t_p (bei $n_1 = n_2$) errechnet sich demnach aus folgender formel:

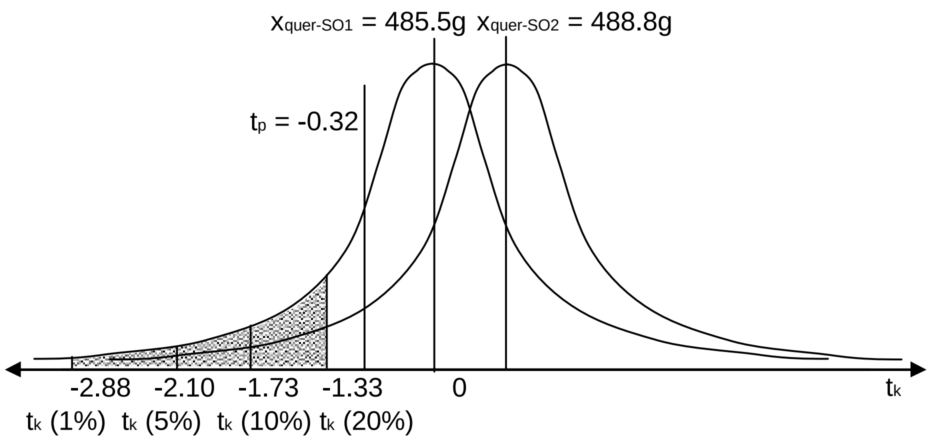
$t_p = (x_{1quer} - x_{2quer}) / wurzel[(s_1\text{^}2 + s_2\text{^}2)/n]$ df = 2n-2 = 18

= -0,3229694

die entsrpechenden einträge für die krit. bereiche holt man sich aus der tabelle (Johnson p.727)

α	t(df;α)	t_k aus tabelle	t_p	α für t-dist (laut excel)	
0.01	t(18;0.01)	-2.88	**-0.32**	0,37522003	37,50%
0.05	t(18;0.05)	-2.10	**-0.32**	0,37522003	37,50%
0,1	t(18;0.1)	-1.73	**-0.32**	0,37522003	37,50%
0,2	t(18;0.2)	-1.33	**-0.32**	0,37522003	37,50%

für alle signifikanz-niveaus (1-20%) it t_k stets grösser (negativer) als t_p; folglich besteht kein unterschied hinsichtlich ihres durschnittlichen gewichtes; die H_0 braucht nicht verworfen zu werden.



hausübung aus biostatistik - beispiel 28 vom 28ten nov. 2001

In einem produktionsbetrieb für farmazeutische produkte werden tabletten hergestellt, die im durchschnitt 100mg einer therapeutischen substanz enthalten. Nun wird in der qualitätssicherung eine stichprobe von 50 tabletten gezogen, die einen mittelwert von 98mg mit einer standard-abweichung von 5mg aufweisen.
Kann man mit einer irrtums-wahrscheinlichkeit von 0.01% die behauptung aufrecht erhalten, dass sich die konzentration umd damit die qualität des produktes verschlechtert hat?

ges: H_0:µ >gleich 100mg H_A: µ < 100mg

da es sich um eine grösser/kleiner beziehung handelt bezieht man sich auf eine 1**-seitige t-verteilung**;

die prüfgrösse t_p (bei $n_1 = n_2$) errechnet sich demnach aus folgender formel:

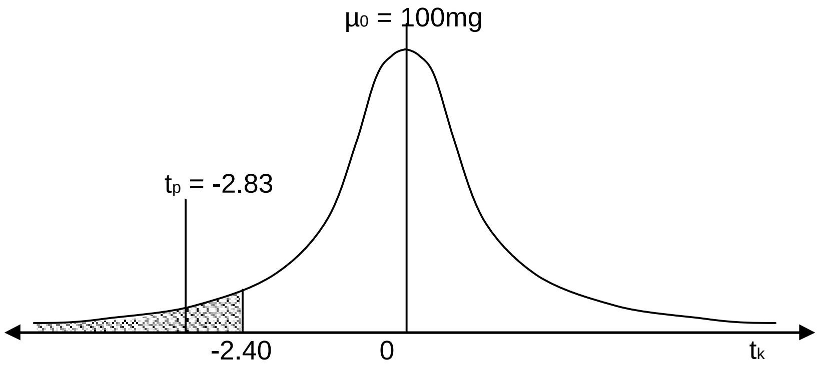
t = $(x_{quer}-\mu_0)$ / [s/wurzel(n)]
= -2,82842712

µ =	100	mg
x_{quer} =	98	mg
s_2 =	5	mg
n_2 =	50	-
df = n-1 =	49	-

die entsrpechenden einträge für die krit. bereiche holt man sich aus der tabelle (Johnson p.727)

α	t(df;α)	t_k aus tabelle	t_p	α für t-dist (laut excel)	
0.01	t(49;0.01)	-2.40	**-2.83**	0,00337927	0,30%

womit H_0 zu verwerfen ist;die tabletten entsprechen der H_A, d.h. die produktion hat **sich verschlechtert**.



hausübung aus biostatistik - beispiel 29 vom 28ten nov. 2001

Das mittlere gewicht von 40 männlichen studenten, die eine überdurchschnittliche beteiligung am hochschul-sport zeigten, war 68.2kg mit einer standrard-abweichung von 2.5kg. Im vergleich dazu hatten 40 männliche studenten, die an eier beteiligung kein interesse zeigten, ein mittleres körpergewicht von 67.5kg mit einer standard-abweichung von 2.8kg.
Zu testen ist die hypothese, dass männliche studenten die am hochschul-sport teilnemen, schwerer als andere männliche studenten sind.
Vergleiche die ergebnisse, wenn sie a) die t-verteilung heran ziehst oder
b) die normalverteilung verwendest.

ges: $H_0: \mu_1 > \mu_2$ H_A: μ_1 <gleich μ_2

da es sich um eine grösser/kleiner beziehung handelt bezieht man sich auf eine **1-seitige t-verteilung**;
die prüfgrösse t_p (bei $n_1 = n_2$) errechnet sich demnach aus folgender formel:

t_p = (x_{1quer} - x_{2quer}) / wurzel[(s_1^2 + s_2^2)/n]
= 1,179431

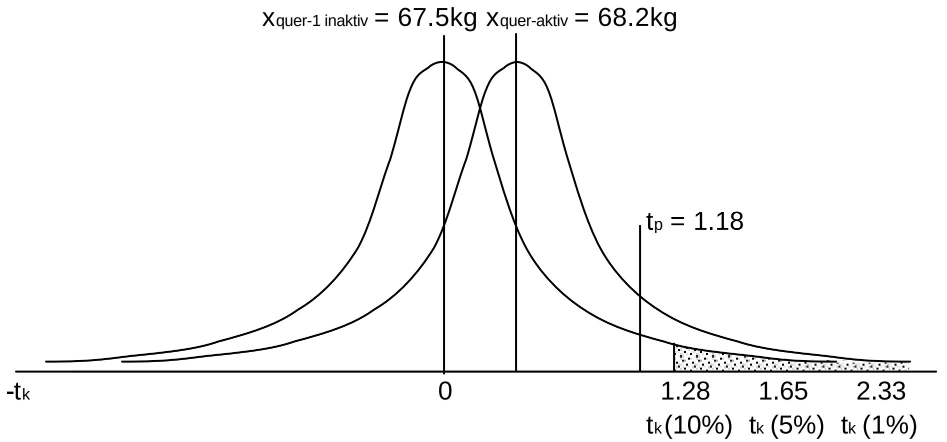
die entsrpechenden einträge für die krit. bereiche holt man sich aus der tabelle

x_{1quer} = 68,2	kg
s_1 = 2,5	kg
n_1 = 40	-
x_{2quer} = 67,5	kg
s_2 = 2,8	kg
n_2 = 40	-

		(John. p.727)				(John. p.725)
α	t(df;α)	$t_{krit.}$ tabelle	r t-dist (laut excel)		t_p	z aus tabelle
0.01	t(78;0.01)	2.375	0,12090577	12,10%	**1.18**	2.33
0.05	t(78;0.05)	1.668	0,12090577	12,10%	**1.18**	1.65
0,1	t(78;0.1)	1.29	0,12090577	12,10%	**1.18**	1.28

df = 2n-2 = 78 -

sowohl bei der t- alsauch bei der normal-verteilung ist die H_0 bei zu behalten da in keinem fall t_p im kritischen bereich liegt.



hausübung aus biostatistik - beispiel 30 vom 6ten dez. 2001

Auf einer versuchsfläche soll überprüft werden, welches von 2 verwendeten mitteln zur reduktion der aufnahme von radioaktivem Sr durch die pflanzen wirksamer ist. Zu diesem zweck werden je 9 flächen ausgewählt, von denen man annimmt, dass sie gleiche ausgangs-bedingungen gewährleisten. Nach der anwendung wird die aktivität in den pflanzen gemessen.
Ist eines der mittel wirksamer, bzw. auf welchem α-niveau kannst du das behaupten?
(Die verfteilungen sind nicht bekannt, die verfahren werden auf verschiedenen flächen durchgeführt).

	aktivität in [Bq/kg]								
fläche	1	2	3	4	5	6	7	8	9
mittel A	51	45	57	52	54	48	61	58	53
mittel B	40	38	39	42	44	50	49	45	43

ges.: H_0: $\mu_A = \mu_B$ oder mittel A entspricht mittel B; H_A: mittel A ungleich mittel B;
da es sich um stichproben handelt die **unabhängig** voneinenader sind (zwei verschiedene populationen) wendet man hier de Mann-Withney-U-test an; dazu müssen die aktivitätsergebnisse aufsteigend in einer tabelle angeordnet werden;

pflanze #	mittel A	mittel B	rang mA	rang mB
1		38		1
2		39		2
3		40		3
4		42		4
5		43		5
6		44		6
7	45		7,5	
8		45		7,5
9	48		9	
10		49		10
11		50		11
12	51		12	
13	52		13	
14	53		14	
15	54		15	
16	57		16	
17	58		17	
18	61		18	

$n_A = 9$
$n_B = 9$
$\Sigma_A = 121,5$
$\Sigma_B = 49,5$

$U_A = n_A{*}n_B + [\ n_A{*}(n_A+1)/2] - \Sigma R_A$
$= 81 + 45 - 121.5$
$= 4,5$

$U_B = n_A{*}n_B + [\ n_B{*}(n_B+1)/2] - \Sigma R_B$
$= 81 + 45 - 49.5$
$= 76,5$

probe: $n_A{*}n_B = U_A + U_B = 81$

in 121.5 vergleichen schneidet A besser als B ab; in 49.5 der vergleiche hingegen B; zur weiteren auswertung wird nur der

Σ	121,5	49,5	kleinere der zwei U-werte herangezogen;

aufgrund der populationen von A bzw. B ergiben sich laut MW-U-tabelle die entsprechenden p-werte (lt. kopie):

n_A; n_B	$\alpha_{(10\%)}$	$\alpha_{(5\%)}$	$\alpha_{(2\%)}$	$\alpha_{(1\%)}$	
9;9	21	17	14	11	
9;9	21	15	n.a.	n.a.	laut buch (Johnson p.739)

der score von U_A liegt unterhalb der 1% marke; d.h. bei verwerfen von H_0 wuerde man nur einen fehler von <1% machen.
denn nur in weniger als 1% der fälle könnte man behaupten dass mittel A dem mittel B überlegen ist;

hausübung aus biostatistik - beispiel 31 vom 6ten dez. 2001

Es wird der fettanteil von schweinen mit ausreichend bewegungsfreiheit in einem gehege mit jenem von schweinen einer masthatlung verglichen. Angegeben ist die fettzunahme der schweine innerhalb eines jahres.
Erstelle die null- und alternativhypothese und berechne mit hilfe des Mann-Withney-U-tests ob ein signifikanter unterschied besteht:

schwein	masthaltung	gehege	schwein	masthaltung	gehege
1	0,09	0,08	6	0,14	0,12
2	0,15	0,04	7	0,12	0,03
3	0,17	0,05	8	0,19	
4	0,12	0,07	9	0,13	
5	0,08	0,09	10	0,15	

ges.: H_0: $\mu_A = \mu_B$ oder mittel mast-tierhaltung entspricht gehegehatung; H_A: masttier ungleich gehegehaltung;

schwein #	mittel A	mittel B	rang mA	rang mB
1		0,03		1
2		0,04		2
3		0,05		3
4		0,07		4
5	0,08		5,5	
6		0,08		5,5
7	0,09		7,5	
8		0,09		7,5
9	0,12		10	
10		0,12		10
11	0,12		10	
12	0,13		12	
13	0,14		13	
14	0,15		14	
15	0,15		15	
16	0,17		16	
17	0,19		17	
Σ			120	33

$n_A = 10$
$n_B = 7$
$\Sigma_A = 120$
$\Sigma_B = 33$

$U_A = n_A * n_B + [\, n_A*(n_A+1)/2] - \Sigma R_A$
$= 70 + 55 - 120$
$= 5$

$U_B = n_A * n_B + [\, n_B*(n_B+1)/2] - \Sigma R_B$
$= 70 + 28 - 33$
$= 65$

probe: $n_A * n_B = U_A + U_B = 70$

in 5 vergleichen schneidet A besser als B ab; in 65 der vergleiche hingegen B; zur weiteren auswertung wird nur der kleinere der zwei U-werte herangezogen;

aufgrund der populationen von A bzw. B ergeben sich laut MW-U-tabelle die entsprechenden p-werte (lt. kopie):

n_A; n_B	$\alpha_{(10\%)}$	$\alpha_{(5\%)}$	$\alpha_{(2\%)}$	$\alpha_{(1\%)}$
10;7	26	22	17	15
10;7	17	14	n.a.	n.a.

2-seitig da auf gleichcheit getestet wird
laut buch (Johnson p.739)

der score von U_A liegt unterhalb der 1% marke; d.h. verwirft man H_0 wuerde man nur einen fehler von <1% machen.
denn nur in weniger als 1% der fälle könnte man behaupten dass mast-tierhaltung der gehegehaltung überlegen ist;

hausübung aus biostatistik - beispiel 32 vom 6ten dez. 2001

Wie wirken sich 2 tassen kaffee am abend aus?
In einem experiment wurden bei 10 personsn die zeit bis zum einschlafen [in min] gemessen.
Wie wirkt sich kaffee nun bei diesen 10 personen aus (α = 0.05)?

H_0: $\mu_{o/coffe} = \mu_{m/coffe}$; H_A: $\mu_{o/coffe}$ ungleich $\mu_{m/coffe}$;

da es sich um stichproben handelt die **abhängig** voneinenader sind (eine populationen an der zwei untersuchungen durch- -geführt werden) wendet man hier den Wilcoxon-Signed-R-test an; dazu müssen die aktivitätsergebnisse aufsteigend in einer tabelle angeordnet werden;

	einschlaf-zeit [min]		differenz		rang	
person	o/ kaffe	m/ kaffe	Δ (+/-)	Δ [min]	positiv	negativ
1	**67**	**60**	-	-7		7
2	**43**	**41**	-	-2		2,5
3	**94**	**93**	-	-1		1
4	**72**	**77**	+	5	6	
5	**30**	**22**	-	-8		8
6	**69**	**69**	0	0	0	
7	**33**	**35**	+	2	2,5	
8	**79**	**65**	-	-14		9
9	**58**	**62**	+	4	5	
10	**48**	**45**	-	-3		4
rang-Σ					13,5	31,5

n = 10

da allerdings ein ergebnis mit der zeit-differenz 0 aus- fällt, verbleiben nur 9 nutz- -bare ergebnisse; d.h.:

$n_{neu} = 9$

verbleiben maximal:

$R_n = n*(n+1)/2$
$= 45$

zur weiteren auswertung wird nur der kleinere der zwei R-werte herangezogen;

aufgrund der populationen von pos./neg. ergeben sich laut WS-R-tabelle die entsprechenden p-werte (lt. kopie):

n	$\alpha_{(10\%)}$	$\alpha_{(5\%)}$	$\alpha_{(2\%)}$	$\alpha_{(1\%)}$
9	8	5	3	1

2-seitig da auf gleichcheit getestet wird

der score von R liegt über der 10% marke; d.h. verwirft man H_0 wuerde man nur einen fehler von >10% machen.
denn in mehr als 10% der fälle trägt kaffe zur kürzeren einschalf-zeit ein als kein kaffe-genuss.

hausübung aus biostatistik - beispiel 33 vom 6ten dez. 2001

Versuchspersonsn machen für die dauer von 3 monaten eine vorgegebene diät. Es wird ihr gewicht vor der diät und eine jahr nach der diät erhoben. Ziel der erhebung ist es festzustellen, ob die versuchspersonen ihr gewicht [in kg] nach der diät halten konnten.
Erstelle die null- und alternativhypothese und berechne, ob ein signifikanter unterschied besteht (α = 0.05)!

H_0: $\mu_{vorher} = \mu_{nachher}$; H_A: μ_{vorher} ungleich $\mu_{nachher}$;

personen	vorher [kg]	1 jahr später [kg]	differenz Δ (+/-)	Δ [kg]	rang positiv	negativ
1	85	89	+	4	1	
2	76	68	-	-8		5
3	94	83	-	-11		6
4	89	89	0	0	0	
5	72	77	+	5	2	
6	105	99	-	-6		3,5
7	102	75	-	-27		7
8	91	85	-	-6		3,5
rang-Σ					3	25

n = 8

da allerdings ein ergebnis mit der zeit-differenz 0 aus- fällt, verbleiben nur 7 nutz- -bare ergebnisse; d.h.:

n_{neu} = 7

verbleiben maximal:

$R_n = n*(n+1)/2$
$= 28$

zur weiteren auswertung wird nur der kleinere der zwei R-werte herangezogen;

aufgrund der populationen von pos./neg. ergeben sich laut WS-R-tabelle die entsprechenden p-werte (lt. kopie):

n	$\alpha_{(10\%)}$	$\alpha_{(5\%)}$	$\alpha_{(2\%)}$	$\alpha_{(1\%)}$
7	3	2	0	-

2-seitig da auf gleichcheit getestet wird

der kleinere rangwert entspricht laut WS-R-tabelle einem signifikanz-niveau von 10% (2-seitig);
d.h. verwirft man H_0 wuerde man nur einen fehler von 10% machen.

hausübung aus biostatistik - beispiel 34 vom 6ten dez. 2001

In der tabelle ist die altersstruktur der österreichischen bevölkerung angegeben. Jemand meint, die wiener bevölkerung sei gegenüber der durchschnittlichen österreichischen bevölkerung ganz anders strukturiert. Hat er damit recht? Wie begründest du deine antwort?
fettgedruckte tabelleinträge stellen die ausgangsdaten dar;

H_0: μ_W ungleich μ_{AUT} (wiener sind anders struktiruert); H_A: $\mu_W = \mu_{AUT}$ (struktur der wiener entspricht jener der österreicher);

sample	altersgr. (x_i)		absolute häufigkeit wiener (b_i)	absolute häufigkeit österreicher	relative häufigkeit wiener	relative häufigkeit österreicher	öst-wien (e_i)	(e_i-b_i)	(e_i-b_i)^2	(e_i-b_i)^2/e_i
1	**0 - 5**	2,5	**78590**	**429320**	0,04903958	0,0530532	85022,1624	6432,162374	41372712,8	486,610922
2	**5 - 10**	7,5	**84696**	**473546**	0,05284968	0,05851843	93780,6413	9084,641255	82530706,7	880,039906
3	**10 - 15**	12,5	**77442**	**467544**	0,04832324	0,05777673	92592,0104	15150,01035	229522814	2478,86198
4	**15 - 20**	17,5	**77993**	**484149**	0,04866706	0,0598287	95880,4502	17887,45023	319960876	3337,08149
5	**20 - 25**	22,5	**88102**	**475806**	0,054975	0,05879771	94228,2097	6126,209705	37530445,3	398,293096
6	**25 - 30**	27,5	**126803**	**600476**	0,07912414	0,0742038	118917,749	-7885,251063	62177184,3	522,85874
7	**30 - 35**	32,5	**154649**	**715044**	0,09649984	0,08836154	141606,697	-13042,30319	170101673	1201,22619
8	**35 - 40**	37,5	**145036**	**700121**	0,0905014	0,08651743	138651,359	-6384,641239	40763643,7	294,001041
9	**40 - 45**	42,5	**117547**	**593476**	0,07334846	0,07333878	117531,475	-15,52502307	241,026341	0,00205074
10	**45 - 50**	47,5	**102909**	**517074**	0,06421446	0,0638974	102400,889	-508,1111819	258176,973	2,52123762
11	**50 - 55**	52,5	**107427**	**484231**	0,06703366	0,05983883	95896,6894	-11530,31056	132948062	1386,36758
12	**55 - 60**	57,5	**119405**	**524227**	0,07450784	0,06478133	103817,463	-15587,53723	242971317	2340,3704
13	**60 - 65**	62,5	**68142**	**376168**	0,04252011	0,04648495	74495,9862	6353,986155	40373140,1	541,950542
14	**65 - 70**	67,5	**62316**	**353099**	0,03888473	0,0436342	69927,4213	7611,421299	57933734,2	828,483778
15	**70 - 75**	72,5	**64751**	**334116**	0,04040415	0,04128837	66168,0444	1417,044358	2008014,71	30,3471975
16	**75 - 80**	77,5	**60181**	**285649**	0,0375525	0,03529907	56569,6815	-3611,318504	13041621,3	230,54083
17	**80 - 85**	82,5	**29038**	**130394**	0,0181195	0,01611343	25823,1153	-3214,884747	10335483,9	400,24156
18	**85 - 90**	87,5	**26206**	**108350**	0,01635235	0,01338935	21457,5405	-4748,459487	22547867,5	1050,81323
19	**90 - 95**	92,5	**9207**	**32503**	0,0057451	0,00401656	6436,86608	-2770,133924	7673641,96	1192,13945
20	**> 95**	97,5	**2143**	**6961**	0,00133722	0,00086021	1378,55043	-764,4495661	584383,139	423,911324
Σ =			1602583	8092254	1,00	1,00	1602583	0,00		**18026,6625**

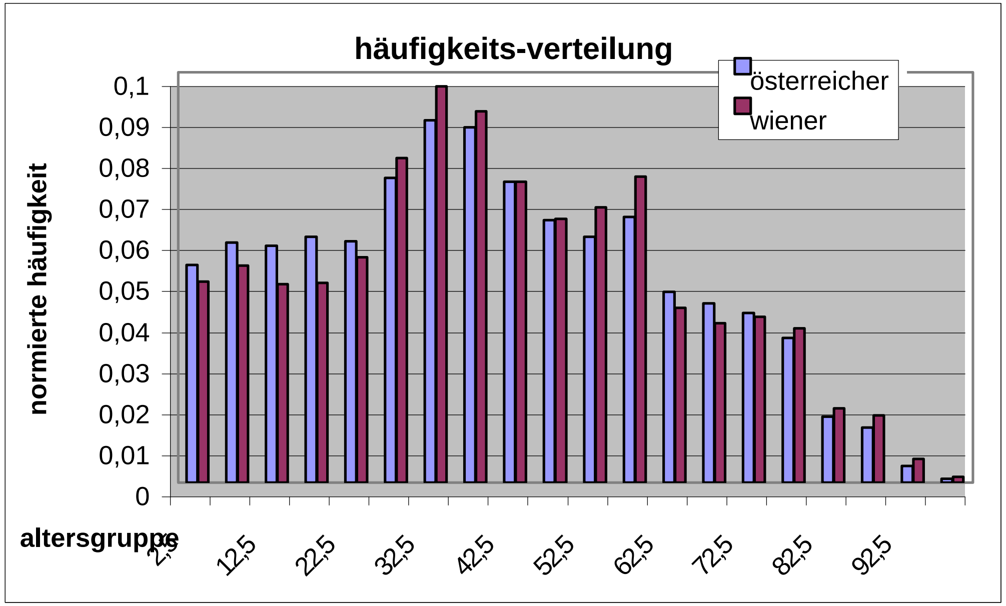
prüfgrösse χ^2 muss mit dem tabellenwert (Johnson p.729) verglichen werden .... P(χ^2 >18027; df = 19)

0

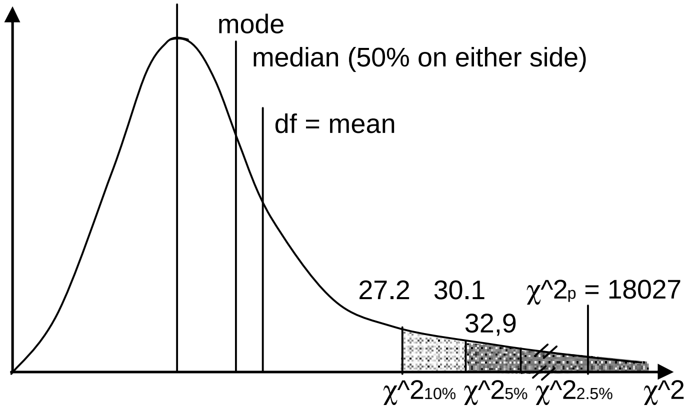
df = n-1 = 19

α (1,r)	0,1	0,05	0,025	0,01
χ^2 tabelle	27,2	30,1	32,9	36,2
χ^2 (excel)	27,2035648	30,1435051	32,852337	36,1907747

.... area in right hand tail (larger values disagree w/ H_0)



es zeigt sich allerdings dass der wert von 18000 weit über den tabelleneinträgen liegt (rechts aussen der poisson-verteilung), womit besagt ist, dass die wiener sich anders verteioen als die österreicher generell;
die wiener bevölkerung ist schlichtweg überaltert;
.... kritischer bereich wird bei weiten übertschritten -



hausübung aus biostatistik - beispiel 35 vom 13ten dez. 2001

Es ist zu überprüfen, ob die westlichen bundesländer in bezug auf die religions-zugehörigkeiten untereinander die gleiche struktur aufweisen; (fettgedruckte zahlenabagen in der tabelle sind die ausgangsdaten).

H_0: die religions-zugehörigkeit in den bundesländern verteilt sich gleich - keine bevorzugung
H_A: religions-zugehörigkeit verteilt sich je nach nach bundesland verschieden

klasse	oberösterreicher (j1)		salzburg (j2)		tirol (j3)		vorarlberg (j4)		
gruppe	b_i	e_i*	b_i	e_i*	b_i	e_i*	b_i	e_i*	Σ
röm. kath. (i1)	**1101640**	1105170,02	**380459**	399777,526	**549078**	523303,987	**271794**	274719,468	2302971
evangelisch (i2)	**62482**	51929,2483	**22614**	18784,5726	**15187**	24588,7802	**7928**	12908,3989	108211
sonstige (i3)	**48463**	65823,949	**26285**	23810,7577	**28138**	31167,9962	**34279**	16362,2972	137165
ohne bekent. (i4)	**84710**	70371,8552	**33395**	25455,8898	**18768**	33321,4545	**9769**	17492,8005	146642
ohna angabe (i5)	**36185**	40184,9294	**19612**	14536,2536	**20239**	19027,7816	**7702**	9989,03539	83738
Σ	1333480		482365		631410		331472		2778727

χ^2-berechnung	oberösterreicher (j1)		
	$e_i - b_i$	$(e_i - b_i)$^2	$(e_i - b_i)$^2/e_i
röm. kath. (i1)	-3530,018	12461028,3	11,2752138
evangelisch (i2)	10552,752	111360569	2144,46719
sonstige (i3)	-17360,95	301402549	4578,91927
ohne bekent. (i4)	14338,145	205582396	2921,3724
ohna angabe (i5)	-3999,929	15999435	398,145156
Σ			10054,1792

*) der erwartungswert errechnet sich aus folgender beziehung:
e_i = Σb_i einer klasse * Σi einer gruppenzeile über alle klassen / Σ$b_i$ges

j = 4
i = 5
ν = (j-1)*(i-1) = 12

χ^2_{single} = (Σe_i - Σb_i) / Σe_i
χ^2 = Σ(χ^2_{single}) = **53371,6799**

χ^2-berechnung	salzburg (j2)		
	$e_i - b_i$	$(e_i - b_i)$^2	
röm. kath. (i1)	-19318,53	373205460	933,532865
evangelisch (i2)	3829,4274	14664514,4	780,667981
sonstige (i3)	2474,2423	6121875,11	257,105431
ohne bekent. (i4)	7939,1102	63029470,6	2476,027
ohna angabe (i5)	5075,7464	25763201,5	1772,34122
Σ			6219,6745

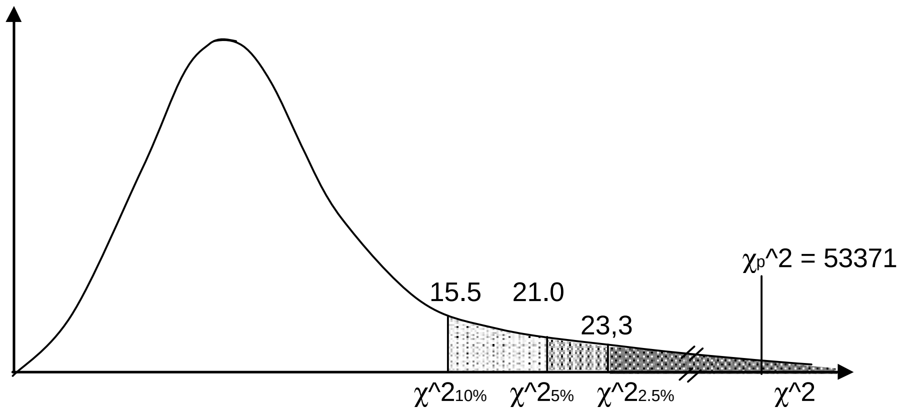
prüfgrösse χ^2 muss mit dem tabellenwert (Johnson p.729) verglichen werden .... P(c^2 >53371; df = 12)

	kritische werte für χ^2			
α(1,r)	0,1	0,05	0,025	0,01
χ^2 tabelle	15,5	21	23,3	26,2
χ^2 excell	18,5493402	21,0260554	23,3366602	26,21696368

χ^2-berechnung	tirol (j3)		
	$e_i - b_i$	$(e_i - b_i)$^2	
röm. kath. (i1)	25774,013	664299723	1269,43371
evangelisch (i2)	-9401,78	88393471,4	3594,87012
sonstige (i3)	-3029,996	9180877,13	294,561032
ohne bekent. (i4)	-14553,45	211803037	6356,3563
ohna angabe (i5)	1211,2184	1467049,93	77,1004185
Σ			11592,3216

χ^2-berechnung	vorarlberg (j4)		
	$e_i - b_i$	$(e_i - b_i)$^2	
röm. kath. (i1)	-2925,468	8558363,34	31,153101
evangelisch (i2)	-4980,399	24804373,4	1921,56855
sonstige (i3)	17916,703	321008241	19618,7759
ohne bekent. (i4)	-7723,8	59657094	3410,37983
ohna angabe (i5)	-2287,035	5230530,86	523,627223
Σ			25505,5046

folglich ist die religions-zugehörigkeit in den einzelnen bundelsändern signifikant verschieden;
die H_0 ist zu verwerfen;



hausübung aus biostatistik - beispiel 36 vom 13ten dez. 2001

Ein typisches verhalten be der jungen wachtel, wie im allgebeinen be nestflüchtern ist der isolationsruf auch "distress vocalization" gennant oder kurz "DV". Dieser tritt auf, wenn ein individuum aus der gruppe isoliert wird. Im versuch werden wachtelküken aus der gruppe isoliert und in eine versuchsbox gesetzt. Im verlauf von 10min werden 3 intervalle zu je 2min beobachtet, eines zu beginn, eines in der mitte une eines am schluss. Erstelle die H_0 und H_A sowie berechne den F-wert mittels varianzanalyse (ANOVA) und vergleich mit dem tabellierten wert (df?; KI 95%). Überprüfe welche intervalle signifikant verschieden sind, mittels "Scheffe test"; (fettgedruckte tabellenabagen sind die ausgangsdaten).

H_0: $\mu_1 = \mu_2 = \mu_3$ oder alle 3 DVs sind gleich; H_A: μ_1, μ_2, μ_3 sind ungleich bzw. alle 3 DVs sind verschieden;

tier	intervall 1		intervall 2		intervall 3		Σ_{row}
	i_1	i_1^2	i_2	i_2^2	i_3	i_3^2	
1	**139**	19321	**74**	5476	**32**	1024	245
2	**111**	12321	**59**	3481	**15**	225	185
3	**126**	15876	**93**	8649	**31**	961	250
4	**131**	17161	**51**	2601	**40**	1600	222
5	**68**	4624	**36**	1296	**19**	361	123
6	**151**	22801	**84**	7056	**54**	2916	289
7	**137**	18769	**100**	10000	**77**	5929	314
8	**101**	10201	**53**	2809	**45**	2025	199
9	**137**	18769	**98**	9604	**59**	3481	294
Σ =	1101		648		372		2121
n =	9		9		9		27
SS_{column} =		139843		50972		18522	

antwort auf frage "a" (F-test):

Σ of squares (SS) =	Σ(x^2)	-	(Σx)^2	/	n	.... according to Johnson p.578
SS_{total} =	209337		4498641		27	
=	42720,6667					

SS_{factor} =	(Σi_1^2/n_1	+	Σi_2^2/n_2	+	Σi_3^2/n_3)	-	(Σx)^2/n_{total}	.... according to Johnson p.579
SS_{DVs} =	134689		46656		15376		166616,333	
=	30104,6667							

$SS_{error} = \quad \Sigma(x\text{^}2) \quad - \quad (\Sigma i_1\text{^}2/n_1 \quad + \quad \Sigma i_2\text{^}2/n_2 \quad + \quad \Sigma i_3\text{^}2/n_3)$ .... according to Johnson p.578

$SS_{error} = \quad 209337 \quad 134689 \quad 46656 \quad 15376$

$= \quad 12616$

gegenprobe:

$SS_{total} = SS_{DVs} \quad + \quad SS_{error} =$

$= \quad 30104{,}6667 \quad 12616$

$42720{,}6667$ .... stimmt mit SS_{total} überein

$Meansquare_{factor}$ (MS) = $SS_{factor} / df_{factor}$ .... according to Johnson p.580

$MS_{DVs} = SS_{DVs} \quad / \quad df_{DVs}$

$= \quad 30104{,}6667 \quad 2$

$= \quad 15052{,}3333$

$MS_{error} = SS_{error} \quad / \quad df_{error}$

$= \quad 12616 \quad 24$

$= \quad 525{,}666667$

$df_{DVs} = \#_{coulmns} -1 = 2$

d_{total} = all events -1 = 26

df_{error} = all events - $\#_{columns}$ = 24

F-wert = $MS_{factor} \quad / \quad MS_{error} \quad = \quad MS_{DVs} \quad / \quad MS_{error}$

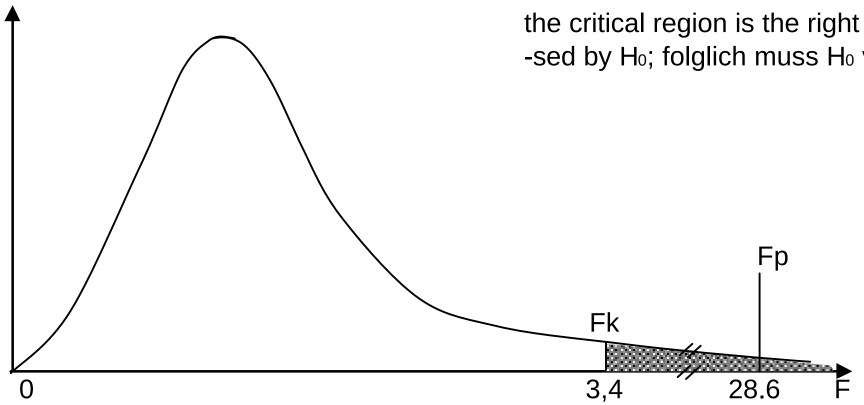
= **28,6347495**

---------- Johnson p.730 ----------

α	F(df$_n$;df$_d$;α)	nominator	denominat.	F$_{krit.}$ tabelle	F$_p$	F$_{krit.}$ excel
0,05	F(2;24;0.05)	2	24	3,4	28,6347495	

$df_{DVs} = df_{numerator} = 2$

$d_{error} = df_{denominator} = 24$



the critical region is the right hand tail; larger values of F indicate "not equal" as expres- -sed by H_0; folglich muss H_0 verworfen werden; stattdessen wird H_A herangezogen;

antwort auf frage "b" (Scheffe-test):

hausübung aus biostatistik - beispiel 37 vom 13ten dez. 2001

Hormone haben auswirkungen auf die ausbildung und expression des isolaionsrufes (DV, s.o.) des wachtelkükens. Dazu wurde die gesamtzahl der DVs über einen zaitraum von 10min bei folgenden gruppen aufgezeichnet. Erstelle die H0 und HA, berechne den F-wert un vergleiche diesen mit der tabelle (KI 95%), sowie führe den post-hoc test durch; (fettgedruckte tabellenabagen sind die ausgangsdaten).

H_0: $\mu_1 = \mu_2 = \mu_3$ oder alle 3 (inkl. kontrolle) sind gleich; H_A: μ_1, μ_2, μ_3 sind ungleich bzw. alle 3 sind verschieden;

	testosteron		estradiol		kontrolle		Σ_{row}
	i_1	i_1^2	i_2	i_2^2	i_3	i_3^2	
1	**347**	120409	**441**	194481	**410**	168100	1198
2	**346**	119716	**402**	161604	**289**	83521	1037
3	**155**	24025	**584**	341056	**338**	114244	1077
4	**194**	37636	**424**	179776	**254**	64516	872
5	**330**	108900	**444**	197136	**199**	39601	973
6	**367**	134689	**320**	102400	**417**	173889	1104
7	**410**	168100	**503**	253009	**485**	235225	1398
8	**284**	80656	**958**	917764	**432**	186624	1674
9	**266**	70756	**319**	101761	**437**	190969	1022
10	**237**	56169	**538**	289444		0	775
11	**254**	64516		0		0	254
Σ =	3190		4933		3261		11384
n =	11		10		9		30
SS_{column} =		985572		2738431		1256689	

antwort auf frage "a" (F-test):

Σ of squares (SS) = Σ(x^2) - (Σx)^2 / n .... according to Johnson p.578

SS_{total} = 4980692 129595456 30

= 660843,467

SS_{factor} = (Σi_1^2/n_1 + Σi_2^2/n_2 + Σi_3^2/n_3) - (Σx)^2/n_{total} .... according to Johnson p.579

$SS_{hormone}$ = 925100 2433448,9 1181569 4319848,53

= 220269,367

$SS_{error} = \Sigma(x^2) - (\Sigma i_1^2/n_1 + \Sigma i_2^2/n_2 + \Sigma i_3^2/n_3)$ .... according to Johnson p.578

$SS_{error} =$ 4980692 925100 2433448,9 1181569

= 440574,1

gegenprobe:

$SS_{total} = SS_{hormone} + SS_{error} =$

= 220269,367 440574,1

= 660843,467 .... stimmt mit SS_{total} überein

$Meansquare_{factor}$ (MS) = $SS_{factor} / df_{factor}$ .... according to Johnson p.580

$MS_{hormon} = SS_{hormon} / df_{hormon}$

= 220269,367 2

= 110134,683

$MS_{error} = SS_{error} / df_{error}$

= 440574,1 27

= 16317,5593

$df_{hormon} = \#_{coulmns} -1 = 2$

d_{total} = all events -1 = 29

df_{error} = all events - $\#_{columns}$ = 27

F-wert = $MS_{factor} / MS_{error} = MS_{hormon} / MS_{error}$

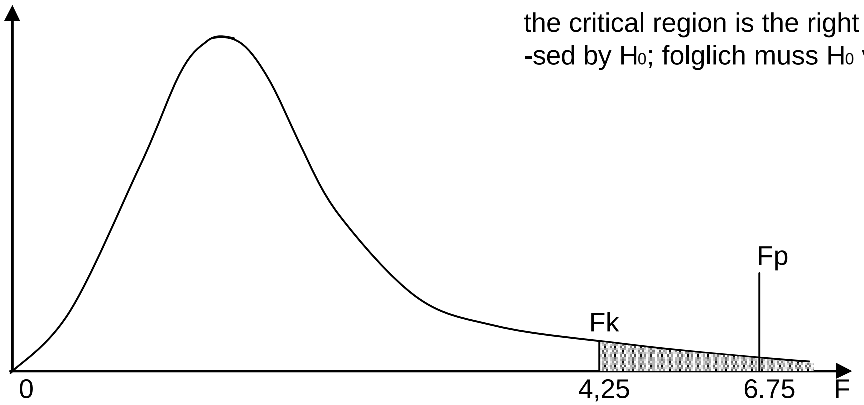
= **6,74945815**

---------- Johnson p.730 ----------

α	$F(df_n;df_d;\alpha)$	nominator	denominat.	$F_{krit.}$ tabelle	F_p	$F_{krit.}$ excel
0,05	F(2;27;0.05)	2	27	3,355	6,74945815	

$df_{DVs} = df_{numerator} = 2$

$d_{error} = df_{denominator} = 27$



the critical region is the right hand tail; larger values of F indicate "not equal" as expres- -sed by H_0; folglich muss H_0 verworfen werden; stattdessen wird H_A herangezogen;

antwort auf frage "b" (post hoc+A31-test):

hausübung aus biostatistik - beispiel 38 vom 13ten dez. 2001

Auf 3 benachbarten versuchsflächen mit je 10*10m wurden (jeweils gleichverteilt über die fläche) 9 proben genommen - von einer fläche fehlt eine probe - und in diesen die aktivitätskonzentrationen von $_{137}$CS bestimmt. Prüfe, ob die kontamination homogen verteilt ist, oder ob es kleinräumige unterschiede gibt. Erstelle die entsprechende H_0 und H_A, sowie führe den Post-Hoc test durch; (fettgedruckte tabellenabagen stellen die ausgangsdaten dar).

H_0: $\mu_1 = \mu_2 = \mu_3$ oder alle 3 aktivitätsgruppen sind gleich; H_A: μ_1, μ_2, μ_3 sind ungleich bzw. alle 3 sind verschieden;

	spezifische aktivität von $_{137}$Cs [Bq/kg]						Σ_{row}
	fläche 1		fläche 2		fläche 3		
	i_1	i_1^2	i_2	i_2^2	i_3	i_3^2	
1	**195,2**	38103,04	**166,3**	27655,69	**188,4**	35494,56	549,9
2	**270,2**	73008,04	**201,6**	40642,56	**127,7**	16307,29	599,5
3	**160,8**	25856,64	**162,8**	26503,84	**128,6**	16537,96	452,2
4	**223**	49729	**205,8**	42353,64	**121,9**	14859,61	550,7
5	**159,4**	25408,36	**220,6**	48664,36	**140,9**	19852,81	520,9
6	**136,6**	18659,56	**168,5**	28392,25	**119,8**	14352,04	424,9
7	**171,6**	29446,56	**152,9**	23378,41	**151,5**	22952,25	476
8	**95,6**	9139,36	**130,8**	17108,64	**184,8**	34151,04	411,2
9		0	**170**	28900	**195,2**	38103,04	365,2
Σ =	1412,4		1579,3		1358,8		4350,5
n =	8		9		9		26
SS_{column} =		269350,56		283599,39		212610,6	

antwort auf frage "a" (F-test):

Σ $_{of\ squares}$ (SS) =	Σ(x^2) -	(Σx)^2 /	n		.... according to Johnson p.578
SS_{total} =	765560,55	18926850,3	26		
=	37604,7712				
SS_{factor} =	(Σi_1^2/n_1 +	Σi_2^2/n_2 +	Σi_3^2/n_3) -	(Σx)^2/n_{total}	.... according to Johnson p.579
$SS_{aktivität}$ =	249359,22	277132,054	205148,604	727955,779	
=	3684,10004				

$SS_{error} = \Sigma(x^2) - (\Sigma i_1{}^2/n_1 + \Sigma i_2{}^2/n_2 + \Sigma i_3{}^2/n_3)$ .... according to Johnson p.578

$SS_{error} =$ 765560,55 249359,22 277132,054 205148,604

= 33920,6711

gegenprobe:

$SS_{total} = SS_{aktivität} + SS_{error} =$

= 3684,10004 33920,6711

= 37604,7712 .... stimmt mit SS_{total} überein

$Meansquare_{factor}$ (MS) = $SS_{factor} / df_{factor}$ .... according to Johnson p.580

$MS_{aktivität} = SS_{aktivität} / df_{aktivität}$

= 3684,10004 2

= 1842,05002

$MS_{error} = SS_{error} / df_{error}$

= 33920,6711 23

= 1474,81179

$df_{aktivität} = \#_{coulmns} -1 = 2$

d_{total} = all events -1 = 25

df_{error} = all events - $\#_{columns}$ = 23

F-wert = $MS_{factor} / MS_{error} = MS_{aktivität} / MS_{error}$

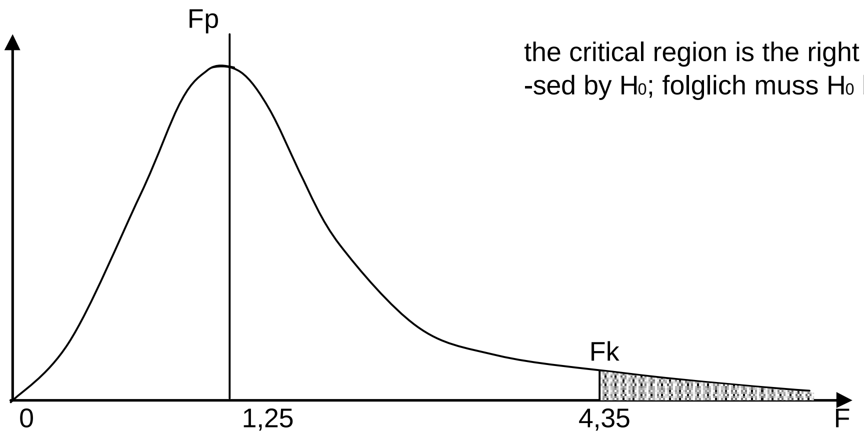
= **1,24900685**

---------- Johnson p.730 ----------

α	F(df$_n$;df$_d$;α)	nominator	denominat.	$F_{krit.}$ tabelle	F_p	$F_{krit.}$ excel
0,05	F(2;23;0.05)	2	23	3,42	1,24900685	

$df_{DVs} = df_{numerator} = 2$

$d_{error} = df_{denominator} = 23$



the critical region is the right hand tail; larger values of F indicate "not equal" as expres- -sed by H_0; folglich muss H_0 beibehalten werden; alle drei aktivitätsgruppen sind gleich;

hausübung aus biostatistik - beispiel 39 vom 17ten jan. 2002

Die angeführten punktepaare sind (fikitv) die empirisch gemessenen gewichte der weiblichen bevölkerung. Zeichne ein diagramm und versuche an die datenpunkte eine lineare regression anzupassen. Zu berechnen, bzw. anzugeben sind:

a) die parameter r, α-hood, β-hood, der linearen regression;

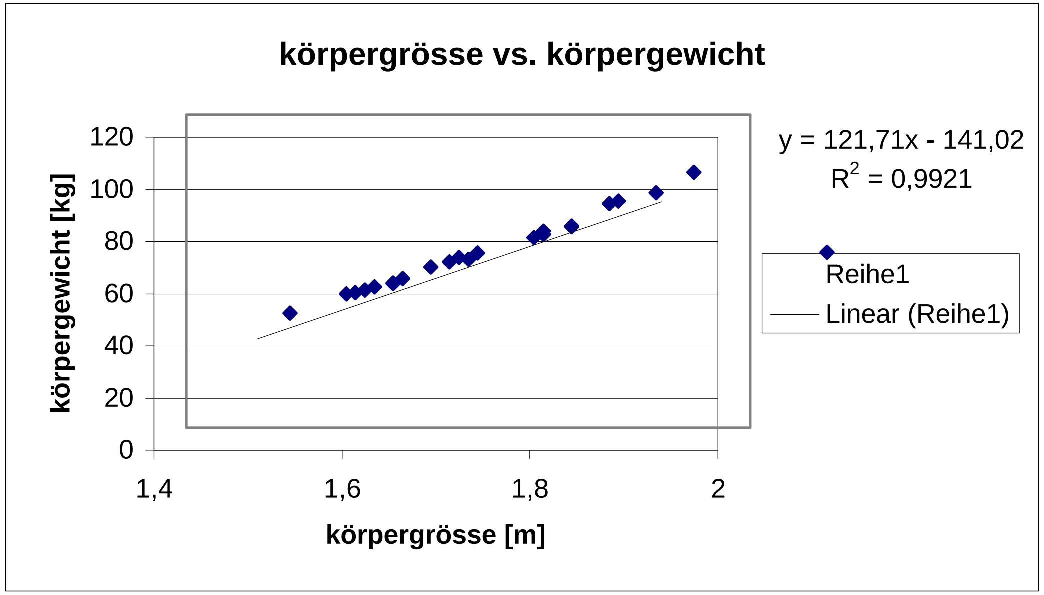
b) wie gross waere nach der regression das gewicht eines 0.5m grossen individuum?

c) welche bedeutung hat α-hood,?

d) liefert die regression ein adequates modell für den gesammten grössenbereich?

antwort auf frage "a"; die parameter r, α-hood, β-hood, der linearen regression;

	körpergrösse	
	[m]	[kg]
1	1,51	44
2	1,57	51,4
3	1,58	51,8
4	1,59	52,8
5	1,6	54
6	1,62	55,4
7	1,62	55,2
8	1,63	57,2
9	1,66	61,6
10	1,68	63,5
11	1,69	65,2
12	1,7	64,6
13	1,71	67,1
14	1,77	72,9
15	1,78	75,3
16	1,78	74,2
17	1,81	77
18	1,81	77,2
19	1,85	85,9
20	1,86	86,9
21	1,9	90
22	1,94	97,9



Linear Regression (Σ-STAT) Donnerstag, Jänner 17, 2002, 15:38:07

körpergewicht = -141,018 + (121,707 * körpergrösse)

N = 22,000
R = 0,996 Rsqr = 0,992 Adj Rsqr = 0,992

Standard Error of Estimate = 1,319

	Coefficient	Std. Error	t	P
Constant	-141,018	4,173	-33,794	<0,001
körpergrösse	121,707	2,432	50,042	<0,001

Analysis of Variance:

	DF	SS	MS	F	P
Regression	1	4353,846	4353,846	2504,168	<0,001
Residual	20	34,773	1,739		
Total	21	4388,619	208,982		

Normality Tes Passed (P = 0,717)
Constant Vari Passed (P = 0,051)
Power of performed test with alpha = 0,050: 1,000

r = 0,99604217
a **= -141,02**
b **= 121,71**

antwort auf die frage "b"; wie gross waere nach der regression das gewicht eines 1.5m grossen individuum?

Kgew. = -141,02 + (121,71*Kgrösse) = **-80,16 [kg]**
unsinn: linear modell kann für den gesamte grössen bereich nicht zutreffen!!!!

antwort auf frage "c"; welche bedeutung hat α-hood,?

α-hood stellt den schnittpunkt mit der x-achse dar (y=0) und ist in diesem fall negativ (-141.02)

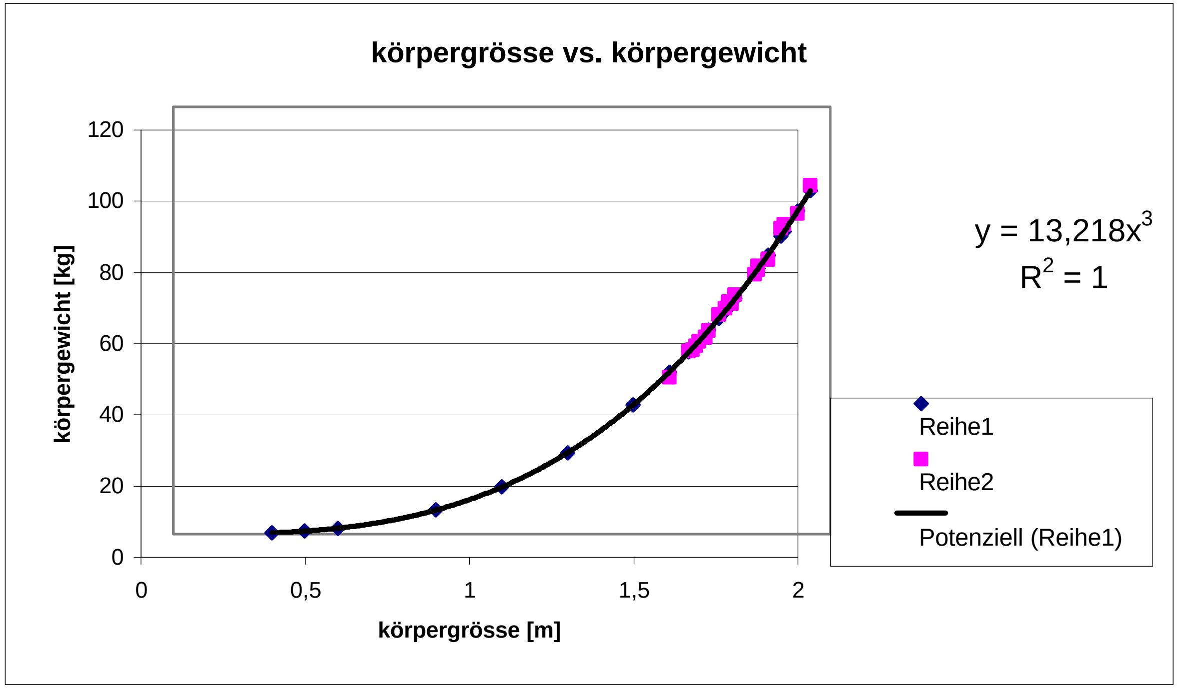
antwort auf frage "d"; liefert die regression ein adequates modell für den gesammten grössenbereich?

ja, da die Σ abweichungsquadrate des regressionsmodells (R^2) der einzelwerte x_i/y_i **nahezu 1** ist!

zusatzantwort: aufgrund der beschränkten anwendbarkeit des linearen modells wäre es besser ein n-lineares heranzuziehen;

-------- körpergrösse --------		
[m]	formel [kg]	messung [kg]
0,3	0,35687979	
0,4	0,84593728	
0,5	1,65222125	
0,8	6,76749824	
1	13,21777	
1,2	22,8403066	
1,4	36,2695609	
1,51	45,5081344	44
1,57	51,1513556	51,4
1,58	52,1350088	51,8
1,59	53,1311925	52,8
1,6	54,1399859	54
1,62	56,1957193	55,4
1,62	56,1957193	55,2
1,63	57,2428178	57,2
1,66	60,4619924	61,6
1,68	62,6738012	63,5
1,69	63,7996512	65,2
1,7	64,938904	64,6
1,71	66,0916389	67,1
1,77	73,2956144	72,9
1,78	74,5449448	75,3
1,78	74,5449448	74,2
1,81	78,3779527	77
1,81	78,3779527	77,2
1,85	83,689963	85,9
1,86	85,0544466	86,9
1,9	90,6606844	90
1,94	96,5080144	97,9

Gew = K*Kgr^3; y = K*Kgrösse^3
K= 13,21777 (laut Lettner)
körpergrösse zu körpergewicht ist eine funktion der 3ten potenz



hausübung aus biostatistik - beispiel 40 vom 17ten jan. 2002

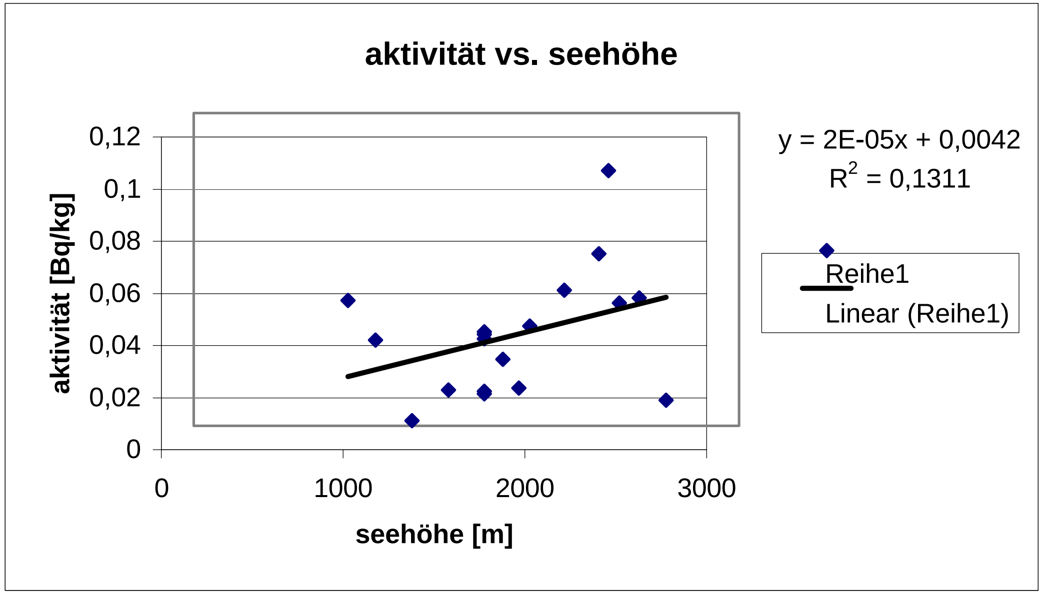
Die angeführten punktepaare sind die experimentell ermittelten aggregierten tranferfaktoren T_{agg} von 137Cs für unterscvhiedliche seehöhen (T_{agg} ist das verhältniss der spezifischen aktivität in dertrockensubstanz von pflanzen in Bq/kg zur flächenbezogenen aktivität des bodens in Bq/m^2). Überprüfe ob zwischen transferfaktor und seehöhe eine korrelation besteht. Falls dies zustrifft, gilt diese korrelation für alle messpunkte oder gibt es abweichungen (ausreisser)? Zeichne ein diagramm und gib an:

a) die parameter r, α-hood, β-hood der linearen regression

b) liefert die regression ein adäquates modell für den gesamten höhenbereich?

antwort auf frage "a", die parameter r, α-hood, β-hood der linearen regression (mit ausreisser):

	seehöhe [m]	t_{agg} 3q/kg ; Bq/m^2]
1	850	0,048
2	1000	0,033
3	1200	0,002
4	1400	0,0137
5	1600	0,0123
6	1600	0,0334
7	1600	0,0351
8	1600	0,036
9	1600	0,0132
10	1700	0,0255
11	1790	0,0146
12	1850	0,0384
13	2040	0,052
14	2230	0,066
15	2280	0,098
16	2340	0,047
17	2450	0,049
18	2600	0,00992



Linear Regression (Σ-STAT) Donnerstag, Jänner 17, 2002, 15:40:32

r = 0,36207734
a = 0,0042
b = 2,00E-05

seehöhe = 1500,227 + (7535,900 * T-agg)

N = 18,000
R = 0,362 Rsqr = 0,131 Adj Rsqr = 0,0768

Standard Error of Estimate = 470,536

	Coefficient	Std. Error	t	P
Constant	1500,227	202,098	7,423	<0,001
T-agg	7535,9	4849,263	1,554	0,14

Analysis of Variance:

	DF	SS	MS	F	P
Regression	1	534693,194	534693,194	2,415	0,14
Residual	16	3542467,92	221404,245		
Total	17	4077161,11	239833,007		

Normality Test: Passed (P = 0,206)
Constant Variance Test: Passed (P = 0,882)

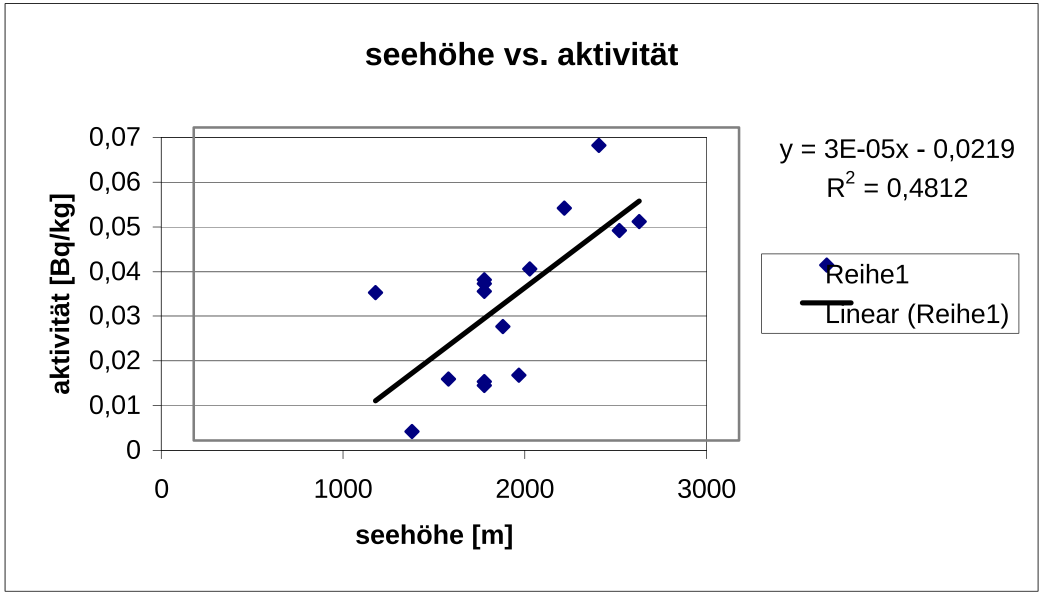
Power of performed test with alpha = 0,050: 0,312
The power of the performed test (0,312) is below the desired power of 0,800.
You should interpret the negative findings cautiously.

antwort auf frage "b" (mit ausreisser); liefert die regression ein adäquates modell für den gesamten höhenbereich?

das modell liefert eher eine unzureichende anpassung an die messwerte, da die Σ der abweichungsquadrate (R^2) <1 ist;
r-werte < 0,7 sind als nicht korrelieren zu betrachten das modell ist zu verwerfen.

antwort auf frage "a" die parameter r, α-hood, β-hood der linearen regression (ohne ausreisser):

	seehöhe [m]	t_{agg} Bq/kg ; Bq/m^2]
1		
2	1000	0,033
3	1200	0,002
4	1400	0,0137
5	1600	0,0123
6	1600	0,0334
7	1600	0,0351
8	1600	0,036
9	1600	0,0132
10	1700	0,0255
11	1790	0,0146
12	1850	0,0384
13	2040	0,052
14	2230	0,066
15		
16	2340	0,047
17	2450	0,049
18		



Linear Regression (Σ-STAT) Donnerstag, Jänner 17, 2002, 16:04:51

r =	**0,69368581**
a **=**	**0,0219**
b **=**	**3,00E-05**

aktivität = -0,0219 + (0,0000308 * seehöhe)

N = 15,000 Missing Observations = 2
R = 0,694 Rsqr = 0,481 Adj Rsqr = 0,441
Standard Error of Estimate = 0,013

	Coefficient	Std. Error	t	P
Constant	-0,0219	0,0157	-1,393	0,187
seehöhe	0,0000308	0,00000887	3,472	0,004

Analysis of Variance:

	DF	SS	MS	F	P
Regression	1	0,00214	0,00214	12,056	0,004
Residual	13	0,00231	0,000178		
Total	14	0,00446	0,000318		

Normality Test: Passed (P = 0,676)
Constant Variance Test: Passed (P = 0,154)
Power of performed test with alpha = 0,050: 0,842

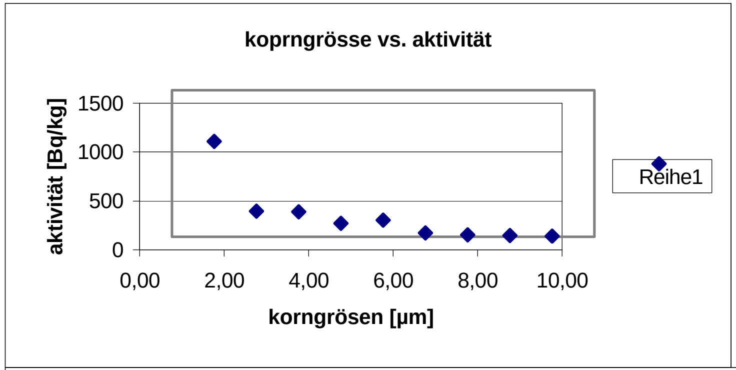
antwort auf frage "b" (ohne ausreisser); liefert die regression ein adäquates modell für den gesamten höhenbereich?

unter ausklammerung der ausreisser liefert das modell eine bessere anpassung an die messwerte, als das vorangehende welches die ausreisser berückichtigt; die Σ der abweichungsquadrate (R^2) liegt bei nahezu 0.7 ist;
Wichtig: dass streichen von messpunkten muss aber begründet werden in diesem fall läst sich das machen da eine cross- korrelation vs. seehöhe in oberen und unteren extrembereichen kritische note dastellen

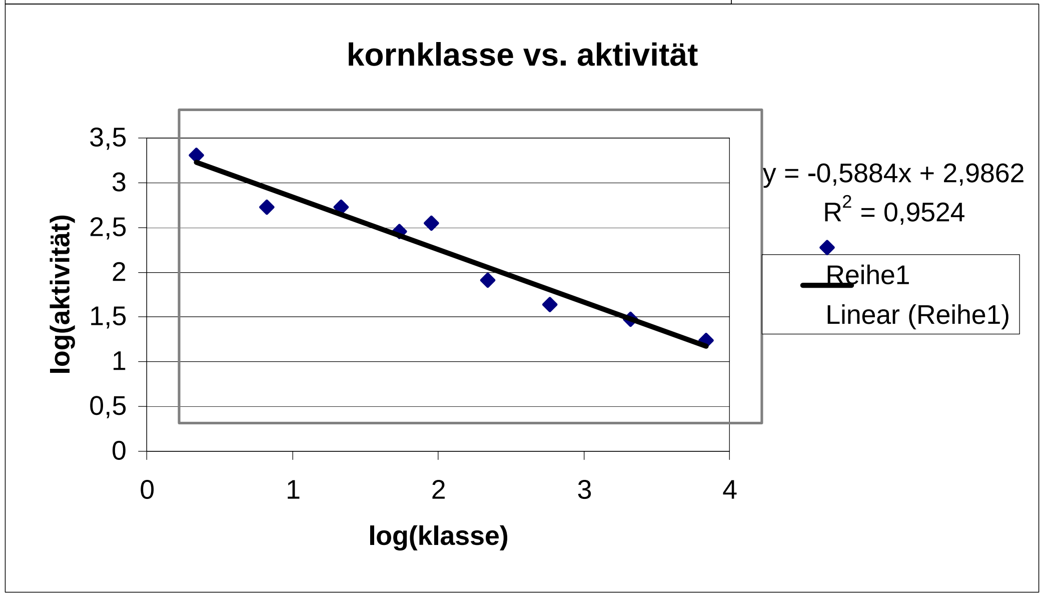
hausübung aus biostatistik - beispiel 41 vom 17ten jan. 2002

es soll untersucht werden, obein sedimenten zwischen der spezifischen aktivität von 137Cs [Bq/kg] un der korngrösse des substrates ein zusammenhang besteht. Für eine korreltation, falls diese existiert, sind die entsprechenden paramter anzugeben.

	korngrösse [µm]	spez. Aktivit. [Bq/kg]
1	<2	977
2	2 - 6	260
3	6 - 20	258
4	20 - 45	137
5	45 - 63	170
6	63 - 200	39,7
7	200 - 500	21,2
8	500 - 2000	14,4
9	>2000	8,4



	klasse	log[klasse]	log[aktiv.]
1	1,315	0,11892575	2,98989456
2	4	0,60205999	2,41497335
3	13	1,11394335	2,41161971
4	32,5	1,51188336	2,13672057
5	54	1,73239376	2,23044892
6	131,5	2,11892575	1,59879051
7	350	2,54406804	1,32633586
8	1250	3,09691001	1,15836249
9	4150	3,6180481	0,92427929



Linear Regression (Σ-STAT) Donnerstag, Jänner 17, 2002, 15:58:12

r = 0,97590983
a = **3**
b = **-5,88E-01**

aktivität = 2,9862 - (0,5884 * klasse)

N = 9,000

R = 0,974 Rsqr = 0,949 Adj Rsqr = 0,941

Standard Error of Estimate = 0,167

	Coefficient	Std. Error	t	P
Constant	2,9862	0,116	26,467	<0,001
klasse	-0,639	0,0561	-11,389	<0,001

Analysis of Variance:					
	DF	SS	MS	F	P
Regression	1	2,9862	2,9862	129,702	<0,001
Residual	7	0,195	0,0278		
Total	8	3,801	0,475		

Normality Test: Passed (P = 0,752)
Constant Variance Test: Passed (P = 0,520)
Power of performed test with alpha = 0,050: 1,000

die oberfläche nimmt mit zunehmenden korn-durchmesser ab; daher fällt die aktivität exponentiell)
das doppelt-logarithmische modell mit linearer regressionsanpassung liefert einen korrelationswert ® von nahezu 1;
das modell ist anwendbar

anmerkung:
$y = k_1*x\text{^}k_2$
$e\text{^}ln(y) = e\text{^}ln(k_1) * e\text{^}ln(x*k_2)$
$ln(y) = ln(k_1 + k_2*ln(x)$ .... geradengleichung
$y = k_1*1/x\text{^}2$

$y = k_1*x\text{^}3$
$ln(y) = ln(k_1) + 3*ln(x)$