Michael Hirschbrich

Statistik

Formel-Kompendium I
für SOWI

Impressum: MI: Hochschülerschaft an der Johannes Kepler Universität Linz,
Druck: Wirtschaftsabteilung der Universität, beide: Altenbergerstraße 69, 4040 Linz

öH-SKRIPT

öh linz SHOP

ersetzt nicht
das Studium der gängigen Literatur
und den Besuch der Lehrveranstaltungen

Denise Rudel
ÖH Shop-Referentin

Julia Sageder
ÖH Vorsitz-Team

Susi Aichinger
ÖH Vorsitz-Team

Liebe Kollegin, lieber Kollege!

Vor dir siehst du ein Skript des Open Courseware Projekts der ÖH Linz, welches allen Studierenden und Interessierten frei und kostenlos zur Verfügung steht.

Das OCW- Projekt der ÖH Linz

Im Jahr 2007 haben der Vorsitz der österreichischen HochschülerInnenschaft Linz und das Referat für Skripten, Lernbehelfe und OCW mit der Umsetzung von Open Courseware an der Johannes Kepler Universität begonnen. Alle Skripten sollten den Studierenden und Interessierten kostenlos zugänglich sein, zudem sollten die Unterlagen frei verändert und vervielfältigt werden dürfen um die Qualität und Aktualität der Unterlagen zu verbessern.

Zu diesem Zweck wurden alle Unterlagen, deren Lizenz bei der ÖH liegt, digitalisiert, mit einer Struktur und Suchfunktion versehen und über eine Homepage allen InternetnutzerInnen zugänglich gemacht. Darüber hinaus wurde den Lehrenden an der JKU die Möglichkeit gegeben jederzeit Verbesserungen und Ergänzungen bei den Unterlagen vorzunehmen.

Lizenz

Weitere und genauere Informationen über Creative Commons findest du unter ***http://www.creativecommons.at.***
Solltest du noch weitere Fragen zum OCW Projekt haben, oder dich beteiligen wollen, erreichst du uns unter ***oeh@oeh.jku.at*** **oder** ***+43 732 2468 8535.***

Wir wünschen dir viel Spaß mit den OCW Skripten und viel Erfolg bei deinen Kursen!

Statistik-Formel-Kompendium I
für SoWi-Studierende

INHALT

1. Eindimensionale Verteilungen

$h_i = h_1 + h_2 + \ldots h_n$

Absolute Häufigkeit

...absolute Häufigkeit der Klasse i
h = Häufigkeit; i = 1+2+3+...(i = Klasse)

$p_i = \frac{h_i}{h.}; p_i = \frac{hi}{n}$

Relative Häufigkeit

... relative Häufigkeit der Klasse i
p = proportion; (bei Stichproben n: n = h.)

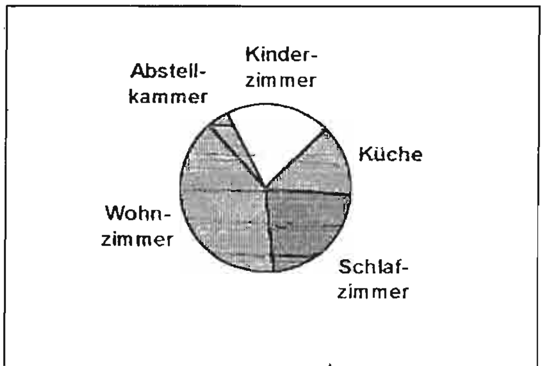
$P_i = p_i \times 100$

Relative Häufigkeit in Prozent

1.1.1 Darstellungsformen bei diskreten Merkmalen

$\alpha_i = 360°.p_i$

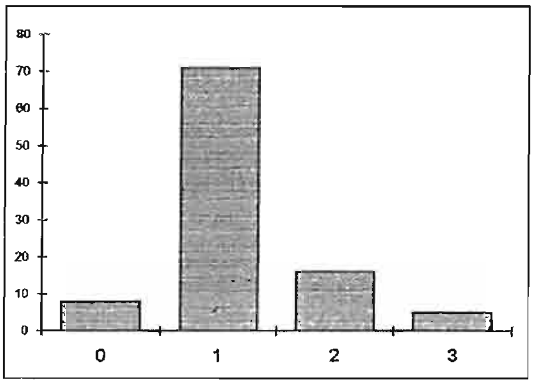
Kreisdiagramm



Eine Eigentumswohnung hat folgende Zimmeraufteilung bezogen auf die gesamte Wohnfläche:

Wohnzimmer: $P_0 = 40\%$
Schlafzimmer: $P_1 = 22\%$
Kinderzimmer: $P_2 = 20\%$
Küche: $P_3 = 14\%$
Abstellkammer: $P_4 = 4\%$

Stabdiagramm



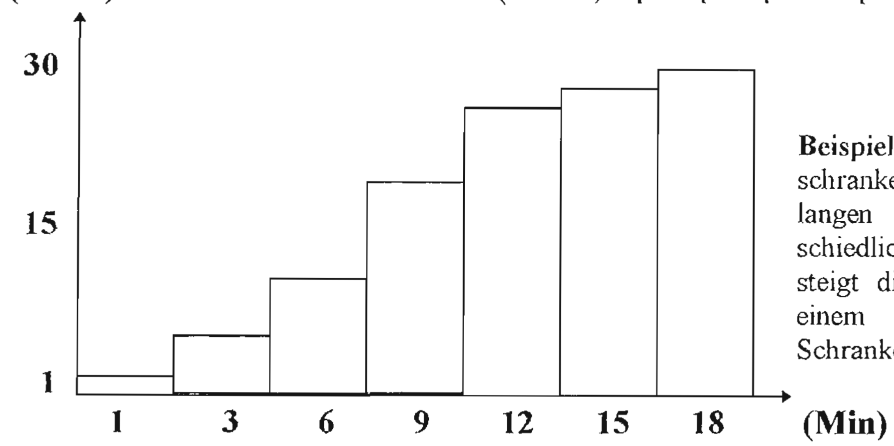
Anzahl von Fernsehgeräten in einigen untersuchten Haushalten:

0 Geräte: p = 8%
1 Gerät: P = 71%
2 Geräte: p = 16%
3 Geräte: p = 4%
(> 3 Geräte: p = 1%)

Summen(häufigkeits)funktion

Die Werte bei diskreten Merkmalen werden auch als Sprungstellen bezeichnet. Daraus ergibt sich die Treppenform der Summenkurve (gilt für absolute und relative Summenhäufigkeiten).

Fi (oder Pi) = p1 + p2 + p3 ... + pi



Beispiel: Bei einem Eisenbahnschranken warten bei verschieden langen Halte-Intervallen unterschiedlich viele Autos. Natürlich steigt die Anzahl der Autos mit einem länger geschlossenen Schranken. (in absoluten Werten)

1.1.2 Darstellungsformen bei stetigen Merkmalen

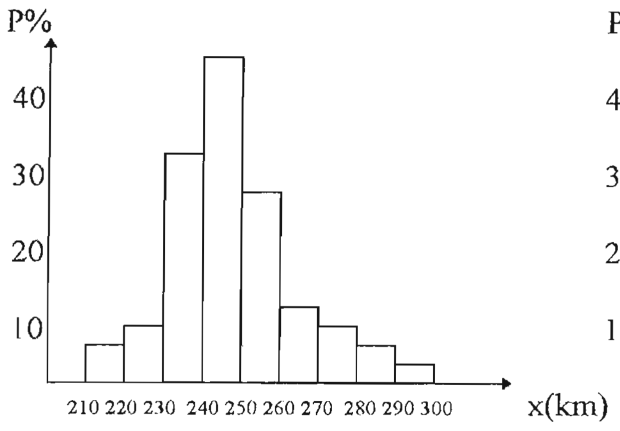
$$f_i = \frac{p_i}{d_i} \times 100$$

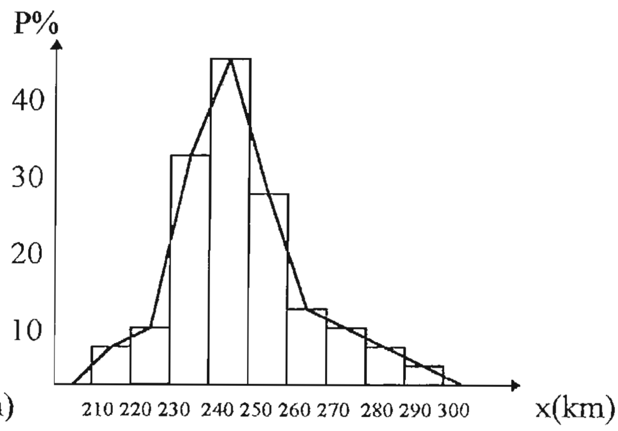
Histogramm, Häufigkeitspolygon

Beim Häufigkeitspolygon werden die Mittelpunkte der Oberkante der im Histogramm mit gleichen Intervallbreiten dargestellten Rechtecke verbunden. Die Flächeninhalte bleiben gleich.

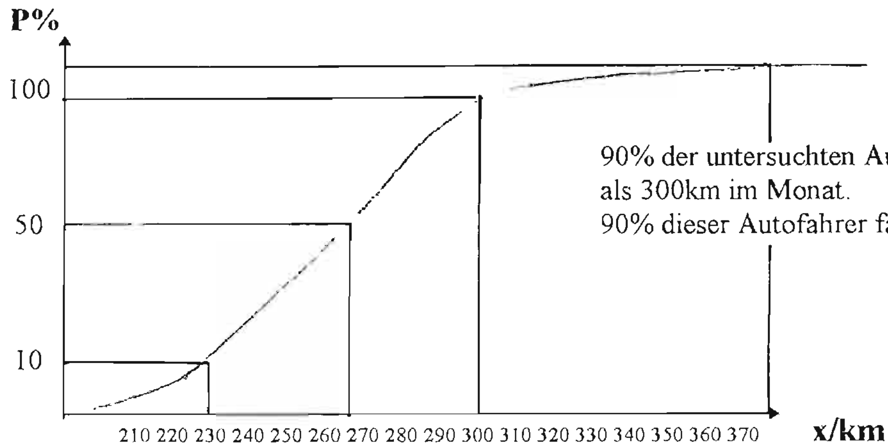
$d_i = e_i - e_{i-1}$ (Intervallbreite); $xi = 0,5 \times (e_{i-1}+e_i)$ (Intervallmitte); e_{i-1} (kleinster Wert des Iv); ei (gr.W.d.Iv)

($p_i = f_i \times d_i$ = Höhe x Breite)





$$F(x) = F(e_{i-1}) + \frac{pi}{di} \times (x - e_{i-1})$$ **Summen(häufigkeits)kurve**



1.2 Maßzahlen bei eindimensionalen Verteilungen

1.2.1 Positionsmaßzahlen (Lage)

Modus

Das Ergebnis, welches am Häufigsten vorkommt (oder auch das Intervall mit der größten Häufigkeit), wird als das Richtige angesehe. Als Ausgangspunkt gilt also der häufigste Wert.

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_N}{N}$$

Mittel

Das (arithmetische) Mittel (Mittelwert) errechnet sich aus der Summe der Ausprägungen durch deren Anzahl. Bei all diesen Beispielen ist zu beachten, ob es sich um eine N (Grundgesamtheit) oder lediglich um eine n (Stichprobe; h.) handelt (allerdings ist dies ein rein formaler Aspekt).

$\tilde{x} = x_{(\frac{N+1}{2})}$ (wenn N *ungerade* ist)

Median

$$\tilde{x} = \frac{x_{(\frac{N}{2})} + x_{(\frac{N}{2}+1)}}{2}$$ (wenn N *gerade* ist)

$$\tilde{x} = [0{,}5 - F(e_{i-1})] \times \frac{d_i}{p_i} + e_{i-1}$$ (Bei Intervalleinteilung)

Bei Berrechnung des Medians müssen die Werte erst nach deren Größe geordnet werden. Nur so kann man den *zentralen* Wert erkennen.

$$g = x1^{p1} \times x2^{p2} \times x3^{p3} \ldots$$

Geometrisches Mittel

Steigerungsrate oder Wachstumsfaktor ($\neq$ prozentuelle Steigerung) $= g_t = \frac{u_t}{u_{t-1}}$; Wachstumsrate $p_t = \frac{u_t - u_{t-1}}{u_{t-1}} \times 100$ (=Änderung in Prozent). Das geometrische Mittel ist niemals größer als das arithmetische Mittel)

1.2.2 Dispersionsmaßzahlen (Streuung)

$$R = x(\max.) - x(\min.)$$

Range

(Spannweite der Verteilung)

$$\Delta = \frac{2}{h.(h.-1)} \times \sum_i \sum_j |x_i - x_j|$$

Gini-Maß

Dieses Maß gilt als der mittlere absolute Abstand zwischen Objekt-Paaren.

$$S^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})^2$$

Varianz

Gibt den mittleren (quadratischen) Abstand zwischen allen Objekten an. (Bei n: 1/(n-1)-Korrektur)

$$s = \sqrt{s^2}$$

Standardabweichung

Die Standardabweichung entspricht dem Wert der positiven Wurzel der Varianz.

$$V = \frac{S}{\bar{x}}$$

Variationskoeffizient

Beim Variationskoeffizient wird die Standardabweichung durch das Mittel dividiert.

1.2.3 Schiefe und Wölbung

$$\alpha_3 = \frac{m_3(\overline{x})}{S^3 x}$$

Schiefe-Maß

$$\gamma = \frac{m_4(\overline{x})}{S^4 - 3}$$

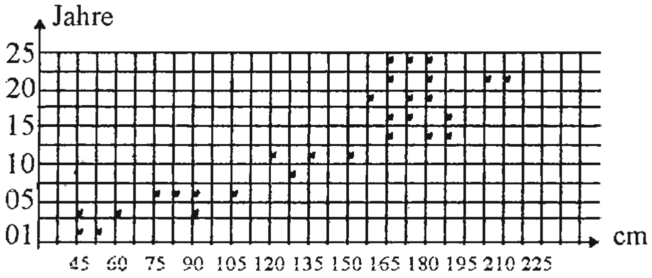
Koeffizient der Wölbung

2. Zwei-und mehrdimensionale Verteilungen

2.1.1 Darstellungsformen bei diskreten Merkmalen

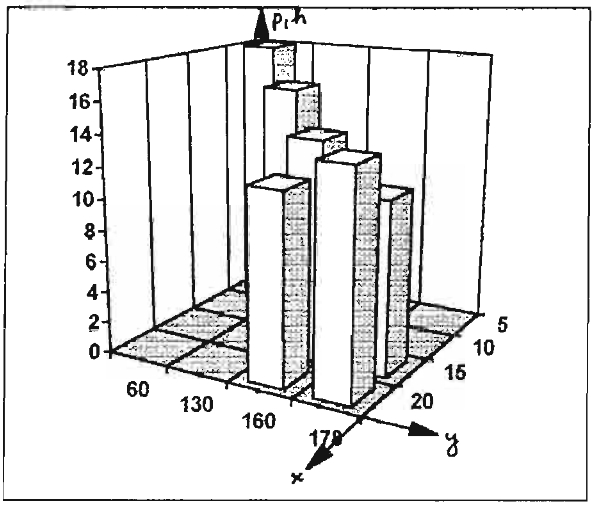
Beim Stabdiagramm wird zusätzlich eine Dimension aufgetragen. Das Prinzip ist gleich dem der eindim. Vert.

2.1.2 Darstellungsformen bei stetigen Merkmalen



Streudiagramm

Streudiagramm für:
Alter und Körpergröße
Die nach einer Untersuchung erhaltenen Meßwerte sind in der x,y-Ebene verstreut.



Histogramm

Als Basis für diese grafischen Darstellungen gelten Tabellen (mit vordefinierten Intervallen). Auf der z-Achse können sowohl absolute als auch relative H. aufgetragen werden.

2.2 Maßzahlen bei zwei- und mehrdimensionalen Verteilungen

$$\chi^2 = \sum_i \sum_j \frac{(hij^b - hij^e)^2}{hij^e}$$

Chi-Quadrat

b = beobachtete Werte, e = erwartete Werte

$$\chi^2 = n \times \frac{(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

„4-Felder"

(Hier werden verschiedenste Schreibformen verwendet. Diese scheint am Einfachsten).

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

$$\phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

Kontingenz- oder Phikoeffizient

$$k = \frac{ad - bc}{ad + bc}$$

Assoziationskoeffizient

$$T^2 = \frac{\chi^2}{n \times \sqrt{(r-1)(s-1)}}$$

Allgemeine Kontingenz

(T = Tschuprow; r,s=Ausprägungen)

$$Kov_{xy} = \frac{1}{N} \sum_{i=1}^{N} x_i y_i - \overline{x}\,\overline{y}$$

Kovarianz

$$\rho = \frac{Kov_{xy}}{S_x S_y}$$

Korrelations-Koeffizient

$$\rho_K = \frac{4 \times (Punktepaare - Anzahl)}{N(N-1)} - 1$$

Korrelationskoeffizient nach *Kendall*

$$\rho_S = 1 - \frac{6 \times \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

Korrelationskoeffizient nach *Spearman*

3. Wahrscheinlichkeitsverteilungen

3.1.1 Diskrete Verteilungen

$$P(X=K)=\frac{\binom{A}{K}\times\binom{N-A}{n-K}}{\binom{N}{n}}$$

Hypergeometrische Verteilung

OHNE ZURÜCKLEGEN ! (N=Grundgesamtheit, A=Anz. der Elementarereignisse , n=Stichprobe, K=Anz. der ausgezeichneten Einheiten in der Stichprobe).

$$P(X=K)=\binom{n}{k}\times\Pi^{k}\times(1-\Pi)^{n-k}$$

$$\Pi=\frac{A}{N}$$

$$(\mu=n\times\Pi)$$

Binomialverteilung

MIT ZURÜCKLEGEN !

$$P(X=K)=e^{-\gamma}\times\frac{\gamma^{k}}{k!}$$

$$\gamma=n\times\frac{A}{N}$$

Poissonverteilung

e (Eulersche Zahl) = 2,718282

APPROXIMATIONSREGELN

(H-Hypergeometrisch; N-Normalverteilung; B-Binomialverteilung; P-Poissonverteilung)

$H_{N,A,n}\approx B_{n,\frac{A}{N}}$ ... falls $n\leq\frac{N}{10}$

$B_{n,\Pi}\approx P_{\mu}$ ... falls $\Pi\leq\frac{1}{10}$

$H_{N,A,n}\approx P_{\mu}$ ... falls $n\leq\frac{N}{10}\,und\,\frac{A}{N}\leq\frac{1}{10}$

3.1.2 Stetige Verteilungen

$$f(x)=\frac{1}{b-a}$$

Gleichverteilung

Dichte (b=Endzeit, a=Anfangszeit)

$$z = \frac{x-\mu}{\sqrt{\delta^2}}$$

Normalverteilung

$$Dichte: f = \frac{1}{\delta \times \sqrt{2\Pi}} \times \ell^{-\frac{(x-\mu)^2}{2\delta^2}}$$

(μ=Mittelwert, δ^2=Varianz); Wird eine B zu einer N, so wird eine Korrektur um +0,5 an beiden Grenzen notwendig.

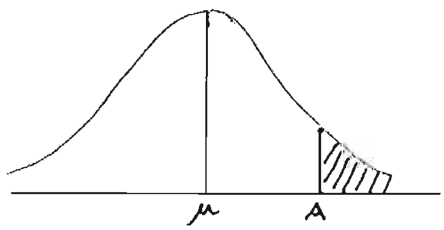
APPROXIMATIONSREGELN

der Normalverteilung von Binomial- und Poissonverteilung

$B_{n,\pi} \approx N(n\pi, n\pi(1-\pi))$ ... falls $\pi(1-\pi) \geq 10$

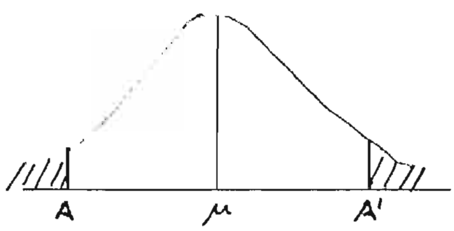
$P_\mu \approx N(\mu,\mu)$ ... falls $\mu \geq 10$

1. Sonderfall:



1 - Pr (x < A)

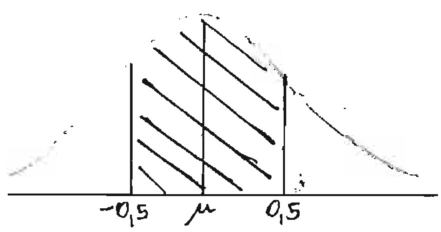
2. Sonderfall:



1 - Pr (x < A')

(Spiegelung)

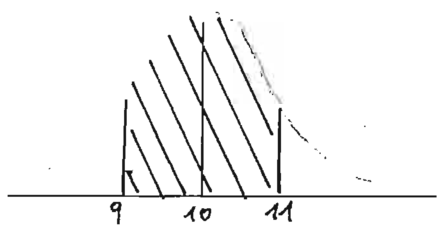
3.Sonderfall:



1 - 2.Pr (x < 0,5)

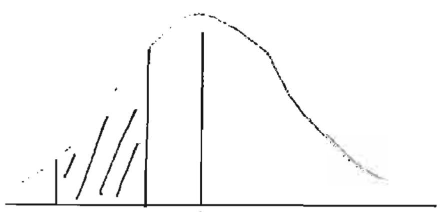
(gleiche Distanzen)

4. Sonderfall:



Pr (x ≤ 11) - Pr (x ≤ 9)

5. Sonderfall:



Pr (x ≤ X_2) - Pr (x ≤ X_1)

X_1 X_2

Allgemeine Vorgehensweise:

1. Standardisieren:

$$z = \frac{x-\mu}{\sigma}$$

2. Tabelle: Erhaltenen Wert in T. suchen

(siehe Anhang)

4. Exkurs: Hypothesen

Der Statistiker stellt eine Hypothese auf, die man üblicherweise **H_1-Hypothese** nennt, und will diese auch bestätigt wissen. Die **Gegenhypothese** dazu bezeichnet man als **H_0**-Hypothese.
Die Bestätigung der H_1 kann durch eine bestimmte Stichprobe erfolgen. Diesen Vorgang nennt man **statistischen Test**, bei dem selbstverständlich auch Fehler auftreten können.

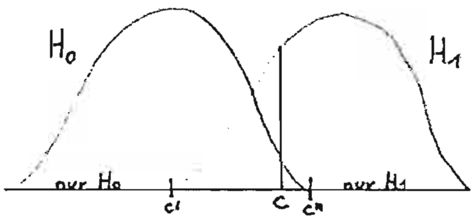
Man unterscheidet 2 Arten:
Fehler 1. Art: Der Statistiker behauptet H_1 sei richtig, in Wahrheit ist aber H_0 richtig.
Fehler 2. Art: Der Statistiker behauptet H_0 sei richtig, in Wahrheit ist aber H_1 richtig.

Die Wahrscheinlichkeit, daß ein Fehler 1. Art auftritt, wird durch das Signifikanzniveau dargestellt. Je niedriger dieses Niveau (α), desto besser der Test. (Fehler 2. Art: β)

Fehler 1. Art: $Pr(E_1/H_0) = Pr(x \geq c/H_0) = \alpha$
Fehler 2. Art: $Pr(E_0/H_1) = Pr(x < c/H_1) = \beta$

Die **einseitige** Fragestellung: H_0: $x \leq 0{,}5$
H_1: $x > 0{,}5$

Die **zweiseitige** Fragestellung: H_0: $x = 0{,}5$
H_1: $x \neq 0{,}5$



weitere Formeln und Anmerkungen

ANHANG

Tabelle 1:

z-Werte von 0 bis 9

		0	1	2	3	4	5	6	7	8	9
	0,50.	0,000	0,003	0,005	0,008	0,010	0,013	0,015	0,018	0,020	0,023
	0,51.	0,025	0,028	0,030	0,033	0,035	0,038	0,040	0,043	0,045	0,048
Ø (z)	0,52.	0,050	0,053	0,055	0,058	0,060	0,063	0,065	0,068	0,070	0,073
	0,53.	0,075	0,078	0,080	0,083	0,085	0,088	0,090	0,093	0,095	0,098
	0,54.	0,100	0,103	0,105	0,108	0,111	0,113	0,116	0,118	0,121	0,123
	0,55.	0,126	0,128	0,131	0,133	0,136	0,138	0,141	0,143	0,146	0,148
	0,56.	0,151	0,154	0,156	0,159	0,161	0,164	0,166	0,169	0,171	0,174
	0,57.	0,176	0,179	0,181	0,184	0,187	0,189	0,192	0,194	0,197	0,199
	0,58.	0,202	0,204	0,207	0,210	0,212	0,215	0,217	0,220	0,222	0,225
	0,59.	0,228	0,230	0,233	0,235	0,238	0,240	0,243	0,246	0,248	0,251
	0,60.	0,253	0,256	0,259	0,261	0,264	0,266	0,269	0,272	0,274	0,277
	0,61.	0,279	0,282	0,285	0,287	0,290	0,292	0,295	0,298	0,300	0,303
	0,62.	0,305	0,308	0,311	0,313	0,316	0,319	0,321	0,324	0,327	0,329
	0,63.	0,332	0,335	0,337	0,340	0,342	0,345	0,348	0,350	0,353	0,356
	0,64.	0,358	0,361	0,364	0,366	0,369	0,372	0,375	0,377	0,380	0,383
	0,65.	0,385	0,388	0,391	0,393	0,396	0,399	0,402	0,404	0,407	0,410
	0,66.	0,412	0,415	0,418	0,421	0,423	0,426	0,429	0,432	0,434	0,437
	0,67.	0,440	0,443	0,445	0,448	0,451	0,454	0,457	0,459	0,462	0,465
	0,68.	0,468	0,470	0,473	0,476	0,479	0,482	0,485	0,487	0,490	0,493
	0,69.	0,496	0,499	0,502	0,504	0,507	0,510	0,513	0,516	0,519	0,522
	0,70.	0,524	0,527	0,530	0,533	0,536	0,539	0,542	0,545	0,548	0,550
	0,71.	0,553	0,556	0,559	0,562	0,565	0,568	0,571	0,574	0,577	0,580
	0,72.	0,583	0,586	0,589	0,592	0,595	0,598	0,601	0,604	0,607	0,610
	0,73.	0,613	0,616	0,619	0,622	0,625	0,628	0,631	0,634	0,637	0,640
	0,74.	0,643	0,646	0,650	0,653	0,656	0,659	0,662	0,665	0,668	0,671
	0,75.	0,674	0,678	0,681	0,684	0,687	0,690	0,693	0,697	0,700	0,703
	0,76.	0,706	0,710	0,713	0,716	0,719	0,722	0,726	0,729	0,732	0,736
	0,77.	0,739	0,742	0,745	0,749	0,752	0,755	0,759	0,762	0,765	0,769
	0,78.	0,772	0,776	0,779	0,782	0,786	0,789	0,793	0,796	0,800	0,803
	0,79.	0,806	0,810	0,813	0,817	0,820	0,824	0,827	0,831	0,834	0,838
	0,80.	0,842	0,845	0,849	0,852	0,856	0,860	0,863	0,867	0,871	0,874
	0,81.	0,878	0,882	0,885	0,889	0,893	0,896	0,900	0,904	0,908	0,912
	0,82.	0,915	0,919	0,923	0,927	0,931	0,935	0,938	0,942	0,946	0,950
	0,83.	0,954	0,958	0,962	0,966	0,970	0,974	0,978	0,982	0,986	0,990
	0,84.	0,994	0,999	1,003	1,007	1,011	1,015	1,019	1,024	1,028	1,032
	0,85.	1,036	1,041	1,045	1,049	1,054	1,058	1,063	1,067	1,071	1,076
	0,86.	1,080	1,085	1,089	1,094	1,098	1,103	1,108	1,112	1,117	1,122
	0,87.	1,126	1,131	1,136	1,141	1,146	1,150	1,155	1,160	1,165	1,170
	0,88.	1,175	1,180	1,185	1,190	1,195	1,200	1,206	1,211	1,216	1,221
	0,89.	1,227	1,232	1,237	1,243	1,248	1,254	1,259	1,265	1,270	1,276
	0,90.	1,282	1,287	1,293	1,299	1,305	1,311	1,317	1,323	1,329	1,335
	0,91.	1,341	1,347	1,353	1,359	1,366	1,372	1,379	1,385	1,392	1,398
	0,92.	1,405	1,412	1,419	1,426	1,433	1,440	1,447	1,454	1,461	1,468
	0,93.	1,476	1,483	1,491	1,499	1,506	1,514	1,522	1,530	1,538	1,546
	0,94.	1,555	1,563	1,572	1,580	1,589	1,598	1,607	1,616	1,626	1,635
	0,95.	1,645	1,655	1,665	1,675	1,685	1,695	1,706	1,717	1,728	1,739
	0,96.	1,751	1,762	1,774	1,787	1,799	1,812	1,825	1,838	1,852	1,866
	0,97.	1,881	1,896	1,911	1,927	1,943	1,960	1,977	1,995	2,014	2,034
	0,98.	2,054	2,075	2,097	2,120	2,144	2,170	2,197	2,226	2,257	2,290
	0,99.	2,326	2,366	2,409	2,457	2,512	2,576	2,652	2,748	2,878	3,090
	0,999.	3,090	3,121	3,156	3,195	3,239	3,291	3,353	3,432	3,540	3,719
	0,9999.	3,719	3,746	3,775	3,808	3,846	3,891	3,944	4,013	4,107	4,265
	0,99999.	4,265	4,288	4,314	4,344	4,378	4,417	4,465	4,526	4,611	4,753
	0,999999.	4,753	4,775	4,798	4,825	4,856	4,892	4,935	4,991	5,069	5,199
	0,9999999.	5,199	5,219	5,241	5,265	5,293	5,327	5,367	5,419	5,491	5,612
	0,99999999.	5,612	5,630	5,650	5,673	5,700	5,731	5,768	5,817	5,884	5,998
	0,999999999.	5,998	6,015	6,034	6,055	6,080	6,109	6,145	6,190	6,254	6,361