Inferenzstatistik

Vorlesung SS 08

Kontinuierliche Prüfverteilungen

Prof. Dr. Rolf Steyer

19. Juni 2008

Kontinuierliche Zufallsvariablen und ihre Wahrscheinlichkeitsdichte

Während wir das Konzept der Verteilung für die Definition der Wahrscheinlichkeitsfunktion verwendet haben, nutzen wir das Konzept der Verteilungsfunktion für die Definition der *Wahrscheinlichkeitsdichte*.

Definition 1 (Kontinuierliche Zufallsvariablen und ihre Wahrscheinlichkeitsdichte)**.** Sei $X : \Omega \to \overline{\mathbb{R}}$ eine numerische Zufallsvariable auf dem Wahrscheinlichkeitsraum $(\Omega, \mathfrak{A}, P)$, und sei F die kumulative Verteilung von X. Wenn es eine Funktion $f : \mathbb{R} \to \mathbb{R}_+$ gibt, derart dass für jedes $x \in \mathbb{R}$

$$F(x) \equiv \int_{-\infty}^{x} f(x)\, dx, \quad x \in \mathbb{R} \tag{1}$$

gilt, dann nennen wir X *kontinuierlich*, und die Funktion f die *Wahrscheinlichkeitsdichte* von X.

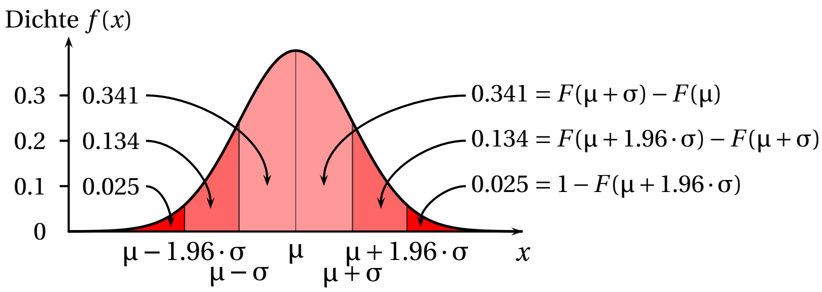
Bemerkung 1. Dieser Definition zufolge ist also die Fläche unter der Wahrscheinlichkeitsdichte $f(x)$ — genau diese wird mit dem o. g. Riemann-Integral berechnet — die Wahrscheinlichkeit dafür, dass X einen Wert kleiner oder gleich x annimmt.

Dichte der Normalverteilung

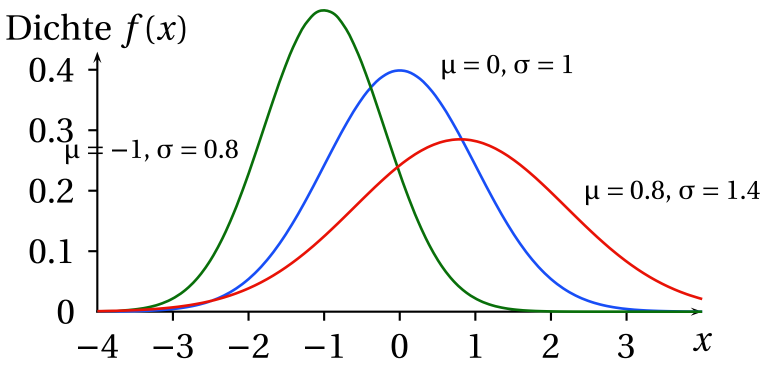
Definition 2 (Normalverteilung)**.** Die Dichte der *Normalverteilung* ist definiert durch

$$f(x) \equiv \frac{1}{\sqrt{2\pi\sigma^2}} \cdot exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right), \tag{2}$$

wobei μ der Erwartungswert und σ^2 die Varianz von X sind. Sind $\mu = 0$ und $\sigma^2 = 1$, dann ist $f(x)$ die Dichte der *Standardnormalverteilung*.



Normalverteilungen mit verschiedenen Erwartungswerten und Varianzen



R-Befehle zur Normalverteilung

Description

- Density, distribution function, quantile function and random generation for the normal distribution with mean equal to *mean* and standard deviation equal to *sd*.

Usage

- *dnorm*(*x*, *mean*=0, *sd*=1, *log* = FALSE) Dichtefunktion
- *pnorm*(*q*, *mean*=0, *sd*=1, lower.tail = TRUE, *log.p* = FALSE) Verteilungsfunktion
- *qnorm*(*p*, *mean*=0, *sd*=1, lower.tail = TRUE, *log.p* = FALSE)
- *rnorm*(*n*, *mean*=0, *sd*=1)

Arguments

- *x*, *q* vector of quantiles.
- *p* vector of probabilities.
- *n* number of observations. If *length*(*n*) > 1, the length is taken to be the number required.
- *mean* vector of means.
- *sd* vector of standard deviations.
- *log*, *log.p* logical; if TRUE, probabilities *p* are given as *log*(*p*).
- lower.tail logical; if TRUE (default), probabilities are $P(X \leq x)$, otherwise, $P(X > x)$.

Tabelle der Standardnormalverteilung

Table I

Cumulative normal probabilities

z	$F(z)$	z	$F(z)$	z	$F(z)$	z	$F(z)$
.00	.5000000	.21	.5831662	.42	.6627573	.63	.7356527
.01	.5039894	.22	.5870604	.43	.6664022	.64	.7389137
.02	.5079783	.23	.5909541	.44	.6700314	.65	.7421539
.03	.5119665	.24	.5948349	.45	.6736448	.66	.7453731
.04	.5159534	.25	.5987063	.46	.6772419	.67	.7485711
.05	.5199388	.26	.6025681	.47	.6808225	.68	.7517478
.06	.5239222	.27	.6064199	.48	.6843863	.69	.7549029
.07	.5279032	.28	.6102612	.49	.6879331	.70	.7580363
.08	.5318814	.29	.6140919	.50	.6914625	.71	.7611479
.09	.5358564	.30	.6179114	.51	.6949743	.72	.7642375
.10	.5398278	.31	.6217195	.52	.6984682	.73	.7673049
.11	.5437953	.32	.6255158	.53	.7019440	.74	.7703500
.12	.5477584	.33	.6293000	.54	.7054015	.75	.7733726
.13	.5517168	.34	.6330717	.55	.7088403	.76	.7763727
.14	.5556700	.35	.6368307	.56	.7122603	.77	.7793501
.15	.5596177	.36	.6405764	.57	.7156612	.78	.7823046
.16	.5635595	.37	.6443088	.58	.7190427	.79	.7852361
.17	.5674949	.38	.6480273	.59	.7224047	.80	.7881446
.18	.5714237	.39	.6517317	.60	.7257469	.81	.7910299
.19	.5753454	.40	.6554217	.61	.7290691	.82	.7938919
.20	.5792597	.41	.6590970	.62	.7323711	.83	.7967306

Aus: Hays, W. L. (1994). *Statistics.* Fort Worth, TX: Harcourt Brace.

Tabelle der Standardnormalverteilung (2a)

Table I (continued)

z	$F(z)$	z	$F(z)$	z	$F(z)$	z	$F(z)$
.84	.7995458	1.32	.9065825	1.79	.9632730	2.26	.9880894
.85	.8023375	1.33	.9082409	1.80	.9640697	2.27	.9883962
.86	.8051055	1.34	.9098773	1.81	.9648521	2.28	.9886962
.87	.8078498	1.35	.9114920	1.82	.9656205	2.29	.9889893
.88	.8105703	1.36	.9130850	1.83	.9663750	2.30	.9892759
.89	.8132671	1.37	.9146565	1.84	.9671159	2.31	.9895559
.90	.8159399	1.38	.9162067	1.85	.9678432	2.32	.9898296
.91	.8185887	1.39	.9177356	1.86	.9685572	2.33	.9900969
.92	.8212136	1.40	.9192433	1.87	.9692581	2.34	.9903581
.93	.8238145	1.41	.9207302	1.88	.9699460	2.35	.9906133
.94	.8263912	1.42	.9221962	1.89	.9706210	2.36	.9908625
.95	.8289439	1.43	.9236415	1.90	.9712834	2.37	.9911060
.96	.8314724	1.44	.9250663	1.91	.9719334	2.38	.9913437
.97	.8339768	1.45	.9264707	1.92	.9725711	2.39	.9915758
.98	.8364569	1.46	.9278550	1.93	.9731966	2.40	.9918025
.99	.8389129	1.47	.9292191	1.94	.9738102	2.41	.9920237
1.00	.8413447	1.48	.9305634	1.95	.9744119	2.42	.9922397
1.01	.8437524	1.49	.9318879	1.96	.9750021	2.43	.9924506
1.02	.8461358	1.50	.9331928	1.97	.9755808	2.44	.9926564
1.03	.8484950	1.51	.9344783	1.98	.9761482	2.45	.9928572
1.04	.8508300	1.52	.9357445	1.99	.9767045	2.46	.9930531
1.05	.8531409	1.53	.9369916	2.00	.9772499	2.47	.9932443

Aus: Hays, W. L. (1994). *Statistics.* Fort Worth, TX: Harcourt Brace.

Tabelle der Standardnormalverteilung (2b)

1.06	.8554277	1.54	.9382198	2.01	.9777844	2.48	.9934309
1.07	.8576903	1.55	.9394292	2.02	.9783083	2.49	.9936128
1.08	.8599289	1.56	.9406201	2.03	.9788217	2.50	.9937903
1.09	.8621434	1.57	.9417924	2.04	.9793248	2.51	.9939634
1.10	.8643339	1.58	.9429466	2.05	.9798178	2.52	.9941323
1.11	.8665005	1.59	.9440826	2.06	.9803007	2.53	.9942969
1.12	.8686431	1.60	.9452007	2.07	.9807738	2.54	.9944574
1.13	.8707619	1.61	.9463011	2.08	.9812372	2.55	.9946139
1.14	.8728568	1.62	.9473839	2.09	.9816911	2.56	.9947664
1.15	.8749281	1.63	.9484493	2.10	.9821356	2.57	.9949151
1.16	.8769756	1.64	.9494974	2.11	.9825708	2.58	.9950600
1.17	.8789995	1.65	.9505285	2.12	.9829970	2.59	.9952012
1.18	.8809999	1.66	.9515428	2.13	.9834142	2.60	.9953388
1.19	.8829768	1.67	.9525403	2.14	.9838226	2.70	.9965330
1.20	.8849303	1.68	.9535213	2.15	.9842224	2.80	.9974449
1.21	.8868606	1.69	.9544860	2.16	.9846137	2.90	.9981342
1.22	.8887676	1.70	.9554345	2.17	.9849966	3.00	.9986501
1.23	.8906514	1.71	.9563671	2.18	.9853713	3.20	.9993129
1.24	.8925123	1.72	.9572838	2.19	.9857379	3.40	.9996631
1.25	.8943502	1.73	.9581849	2.20	.9860966	3.60	.9998409
1.26	.8961653	1.74	.9590705	2.21	.9864474	3.80	.9999277
1.27	.8979577	1.75	.9599408	2.22	.9867906	4.00	.9999683
1.28	.8997274	1.76	.9607961	2.23	.9871263	4.50	.9999966
1.29	.9014747	1.77	.9616364	2.24	.9874545	5.00	.9999997
1.30	.9031995	1.78	.9624620	2.25	.9877755	5.50	.9999999
1.31	.9049021						

Aus: Hays, W. L. (1994). *Statistics.* Fort Worth, TX: Harcourt Brace.

Dichte der zentralen χ^2-Verteilung

Definition 3 (Zentrale χ^2-Verteilung)**.** Seien X_i unabhängige, normalverteilte Zufallsvariablen mit Erwartungswert $\mu = 0$ und Varianz $\sigma^2 = 1$. Dann hat

$$X \equiv \sum_{i=1}^{n} X_i^2 \tag{3}$$

eine *zentrale χ^2-Verteilung mit n Freiheitsgraden*, d. h. die Dichte von X ist

$$f(x) = \begin{cases} \dfrac{x^{(n/2)-1} \cdot exp\left(\dfrac{-x}{2}\right)}{2^{n/2} \cdot \Gamma\left(\dfrac{n}{2}\right)}\,, & \text{falls } x > 0 \\ 0\,, & \text{falls } x \leq 0. \end{cases} \tag{4}$$

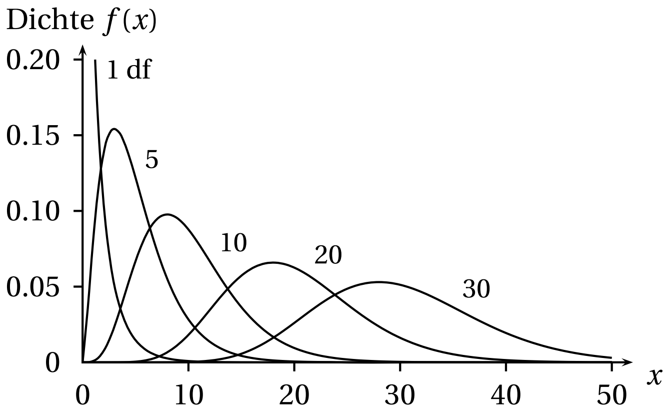
Dabei bezeichnet $\Gamma(n/2)$ die *Gamma-Funktion*

$$\Gamma(a) \equiv \int_0^\infty t^{a-1} exp(-t)\; dt\,, \qquad a > 0, \tag{5}$$

mit dem Argument $n/2$. Dabei ist $\Gamma(a) = (a-1)!$, falls a eine positive ganze Zahl ist und $\Gamma(a) = (a-1) \cdot \Gamma(a-1)$.

Bemerkung 2. Der Erwartungswert von X ist n und die Varianz $2n$.

χ^2 Verteilungen mit 1,5, 10, 20 und 30 Freiheitsgraden



R-Befehle zur χ^2-Verteilung

Description

- Density, distribution function, quantile function and random generation for the χ^2 distribution with *df* degrees of freedom and optional non-centrality parameter *ncp*.

Usage

- *dchisq*(x, *df*, *ncp*=0, *log* = *FALSE*) Dichtefunktion
- *pchisq*(q, *df*, *ncp*=0, *lower.tail* = *TRUE*, *log.p* = *FALSE*) Verteilungsfunktion
- *qchisq*(p, *df*, *ncp*=0, *lower.tail* = *TRUE*, *log.p* = *FALSE*)
- *rchisq*(n, *df*, *ncp*=0)

Arguments

- x, q vector of quantiles.
- p vector of probabilities.
- n number of observations. If *length*(n) > 1, the length is taken to be the number required.
- *df* degrees of freedom (non-negative, maybe non-integer).
- *ncp* non-centrality parameter (non-negative);
- *log*, *log.p* logical; if *TRUE*, probabilities p are given as $log(p)$.
- *lower.tail* logical; if *TRUE* (default), probabilities are $P(X \le x)$, otherwise, $P(X > x)$.

χ^2-Verteilung Tabelle

Tabelle B.2: P-Quantile $\chi^2(f;p)$ der χ^2-Verteilung (kritische Werte des χ^2-Tests) mit f Freiheitsgraden

f	P							
	0,01	0,025	0,05	0,1	0,9	0,95	0,975	0,99
1	$1571 \cdot 10^{-7}$	$9821 \cdot 10^{-7}$	$3932 \cdot 10^{-6}$	0,01579	2,706	3,841	5,024	6,635
2	0,02010	0,05064	0,1026	0,2107	4,605	5,991	7,378	9,210
3	0,1148	0,2158	0,3518	0,5844	6,251	7,815	9,348	11,34
4	0,2971	0,4844	0,7107	1,064	7,779	9,488	11,14	13,28
5	0,5543	0,8312	1,145	1,610	9,236	11,07	12,83	15,09
6	0,8721	1,237	1,635	2,204	10,64	12,59	14,45	16,81
7	1,239	1,690	2,167	2,833	12,02	14,07	16,01	18,48
8	1,646	2,180	2,733	3,490	13,36	15,51	17,53	20,09
9	2,088	2,700	3,325	4,168	14,68	16,92	19,02	21,67
10	2,558	3,247	3,940	4,865	15,99	18,21	20,48	23,21
11	3,053	3,816	4,575	5,578	17,28	19,68	21,92	24,72
12	3,571	4,404	5,226	6,304	18,55	21,03	23,34	26,22
13	4,107	5,009	5,892	7,042	19,81	22,36	24,74	27,69
14	4,660	5,629	6,571	7,790	21,06	23,68	26,12	29,14
15	5,229	6,262	7,261	8,547	22,31	25,00	27,49	30,58
16	5,812	6,908	7,962	9,312	23,54	26,30	28,85	32,00
17	6,408	7,564	8,672	10,09	24,77	27,59	30,19	33,41
18	7,015	8,231	9,390	10,86	25,99	28,87	31,53	34,81
19	7,633	8,907	10,12	11,65	27,20	30,14	32,85	36,19
20	8,260	9,591	10,85	12,44	28,41	31,41	34,17	37,57

Aus: Rasch, D. & Kubinger, K. D. (2006), *Statistik für das Psychologiestudium.* Heidelberg: Elsevier.

t-Verteilung

Definition 4 (Zentrale t-Verteilung)**.** Seien X_1 und X_2 zwei unabhängige Zufallsvariablen, wobei X_1 standardnormalverteilt ist und X_2 (zentral) χ^2-verteilt mit n Freiheitsgraden. Die Dichte von

$$X \equiv \frac{X_1}{\sqrt{X_2/n}} \tag{6}$$

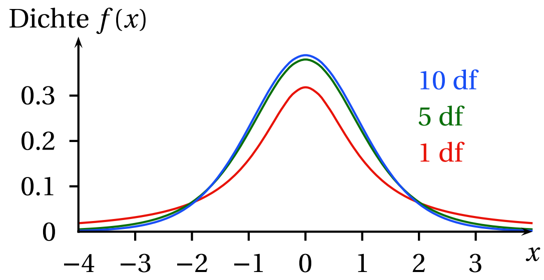
ist dann

$$f(x) = \frac{\Gamma\left[\frac{n+1}{2}\right]}{\sqrt{n\,\pi}\cdot\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-(n+1)/2}, \tag{7}$$

die *t-Verteilung mit n Freiheitsgraden.* [Siehe Gl. (5) zur Definition der Gamma-Funktion].

Bemerkung 3. Der Erwartungswert von X ist 0 und die Varianz $n/(n-2)$.

Zentrale t-Verteilung für $df = 1,5,10$



R-Befehle zur t-Verteilung

Description

- Density, distribution function, quantile function and random generation for the t distribution with *df* degrees of freedom (and optional non-centrality parameter *ncp*).

Usage

- *dt*(*x*, *df*, *ncp*, *log* = *FALSE*)   Dichtefunktion
- *pt*(*q*, *df*, *ncp*, *lower.tail* = *TRUE*, *log.p* = *FALSE*)   Verteilungsfunktion
- *qt*(*p*, *df*, *ncp*, *lower.tail* = *TRUE*, *log.p* = *FALSE*)
- *rt*(*n*, *df*, *ncp*)

Arguments

- x, q vector of quantiles.
- p vector of probabilities.
- n number of observations. If *length*(n) > 1, the length is taken to be the number required.
- df degrees of freedom (> 0, maybe non-integer). *df* = *Inf* is allowed. For *qt* only values of at least one are currently supported.
- *ncp* non-centrality parameter delta; currently except for *rt*( ), only for *abs*(*ncp*) ≤ 37.62. If omitted, use the central t distribution.
- *log*, *log.p* logical; if *TRUE*, probabilities p are given as $log(p)$.
- lower.tail logical; if *TRUE* (default), probabilities are $P(X \leq x)$, otherwise, $P(X > x)$.

t-Verteilung Tabelle

Tabelle B.1: P-Quantile der t-Verteilung mit f Freiheitsgraden (für $f = \infty$, P-Quantile der Standardnormalverteilung)

f	P								
	0,6	0,7	0,8	0,85	0,9	0,95	0,975	0,99	0,995
1	0,3249	0,7265	1,3764	1,9626	3,0777	6,3138	12,7062	31,8205	63,6567
2	0,2887	0,6172	1,0607	1,3862	1,8856	2,9200	4,3027	6,9646	9,9248
3	0,2767	0,5844	0,9785	1,2498	1,6377	2,3534	3,1824	4,5407	5,8409
4	0,2707	0,5686	0,9410	1,1896	1,5332	2,1318	2,7764	3,7469	4,6041
5	0,2672	0,5594	0,9195	1,1558	1,4759	2,0150	2,5706	3,3649	4,0321
6	0,2648	0,5534	0,9057	1,1342	1,4398	1,9432	2,4469	3,1427	3,7074
7	0,2632	0,5491	0,8960	1,1192	1,4149	1,8946	2,3646	2,9980	3,4995
8	0,2619	0,5459	0,8889	1,1081	1,3968	1,8595	2,3060	2,8965	3,3554
9	0,2610	0,5435	0,8834	1,0997	1,3830	1,8331	2,2622	2,8214	3,2498
10	0,2602	0,5415	0,8791	1,0931	1,3722	1,8125	2,2281	2,7638	3,1693
11	0,2596	0,5399	0,8755	1,0877	1,3634	1,7959	2,2010	2,7181	3,1058
12	0,2590	0,5386	0,8726	1,0832	1,3562	1,7823	2,1788	2,6810	3,0545
13	0,2586	0,5375	0,8702	1,0795	1,3502	1,7709	2,1604	2,6503	3,0123
14	0,2582	0,5366	0,8681	1,0763	1,3450	1,7613	2,1448	2,6245	2,9768
15	0,2579	0,5357	0,8662	1,0735	1,3406	1,7531	2,1314	2,6025	2,9467
16	0,2576	0,5350	0,8647	1,0711	1,3368	1,7459	2,1199	2,5835	2,9208
17	0,2573	0,5344	0,8633	1,0690	1,3334	1,7396	2,1098	2,5669	2,8982
18	0,2571	0,5338	0,8620	1,0672	1,3304	1,7341	2,1009	2,5524	2,8784
19	0,2569	0,5333	0,8610	1,0655	1,3277	1,7291	2,0930	2,5395	2,8609
20	0,2567	0,5329	0,8600	1,0640	1,3253	1,7247	2,0860	2,5280	2,8453

Aus: Rasch, D. & Kubinger, K. D. (2006), *Statistik für das Psychologiestudium*. Heidelberg: Elsevier.

F-Verteilung

Definition 5 (Zentrale F-Verteilung)**.** Seien X_1 und X_2 zwei unabhängige Zufallsvariablen mit zentraler χ^2-Verteilung mit n_1 bzw. n_2 Freiheitsgraden. Die Dichte von

$$X \equiv \frac{X_1/n_1}{X_2/n_2} \tag{8}$$

ist dann

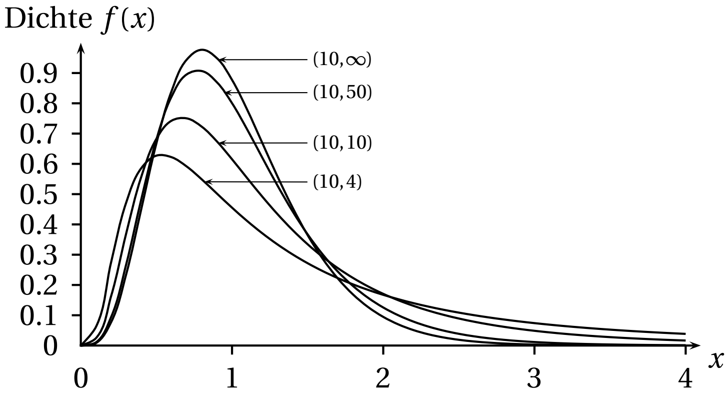
$$f(x) = \frac{\Gamma\left(\dfrac{n_1+n_2}{2}\right) \cdot n_1^{n_1/2} \cdot n_2^{n_2/2} \cdot x^{(n_1/2)-1}}{\Gamma\left(\dfrac{n_1}{2}\right) \cdot \Gamma\left(\dfrac{n_2}{2}\right) \cdot (n_2 + n_1 x)^{(n_1+n_2)/2}}\,, \qquad x > 0, \tag{9}$$

die *zentrale F-Verteilung mit n_1 und n_2 Freiheitsgraden.* Dabei bezeichnet $\Gamma(n_i/2)$ wieder die Gamma-Funktion mit dem Argument $n_i/2$, wobei $i = 1,2$.

Bemerkung 4. Der Erwartungswert von X ist $n_2/(n_2-2)$, für $n_2 > 2$, und die Varianz ist

$$\frac{2n_2^2(n_1+n_2-2)}{n_1 \cdot (n_2-2)^2 \cdot (n_2-4)}\,, \qquad \text{für } n_2 > 4\,. \tag{10}$$

F-Verteilungen



R-Befehle zur F-Verteilung

Description

- Density, distribution function, quantile function and random generation for the F distribution with df_1 and df_2 degrees of freedom (and optional non-centrality parameter *ncp*).

Usage

- $df(x, df_1, df_2, ncp, log = FALSE)$ Dichtefunktion
- $pf(q, df_1, df_2, ncp, lower.tail = TRUE, log.p = FALSE)$ Verteilungsfunktion
- $qf(p, df_1, df_2, ncp, lower.tail = TRUE, log.p = FALSE)$
- $rf(n, df_1, df_2, ncp)$

Arguments

- x, q vector of quantiles.
- p vector of probabilities.
- n number of observations. If $length(n) > 1$, the length is taken to be the number required.
- df_1, df_2 degrees of freedom. *Inf* is allowed.
- *ncp* non-centrality parameter. If omitted the central F is assumed.
- *log, log.p* logical; if *TRUE*, probabilities p are given as $log(p)$.
- *lower.tail* logical; if *TRUE* (default), probabilities are $P(X \leq x)$, otherwise, $P(X > x)$.

F-Verteilung Tabelle

Tabelle B.3: 95%-Quantile der F-Verteilung mit f_1 und f_2 Freiheitsgraden

f_2	f_1 = 1	2	3	4	5	6	7	8	9
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
12	4,75	3,89	3,49	3,27	3,11	3,00	2,91	2,85	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39

Aus: Rasch, D. & Kubinger, K. D. (2006), *Statistik für das Psychologiestudium.* Heidelberg: Elsevier.