TILASTOTIETEEN JATKOKURSSI / SYKSY
Ylimääräinen 2. välikoe 19.12.2007

1. a) Verenpainelääkkeen testauksessa samojen potilaiden verenpaine mitattiin ennen ja jälkeen lääkkeen nauttimisen. Verenpaineet (mm/Hg) on esitetty alla olevassa taulukossa. Testaa hypoteesia, että lääke ei vaikuta verenpaineeseen. (8 potilasta).

	1	2	3	4	5	6	7	8
Ennen	128	176	110	149	183	136	118	158
Jälkeen	134	174	118	152	187	136	125	168

b) Havainnollista a) –kohdan esimerkin avulla toisistaan riippuviin otoksiin/aineistoihin ja riippumattomiin otoksiin/aineistoihin perustuvien tilastollisten analyysien eroja ja yhtäläisyyksiä.

2. Muuttujien x ja Y havaitut arvot ovat:

x	1	3	4	6	8	9	11	14
Y	1	2	4	4	5	7	8	9

a) Määrää muuttujien x ja Y välinen korrelaatio.
Lisäksi: Määrättäessä yo. datan pohjalta regressiomallin $Y = \alpha + \beta x + \varepsilon$ regressiokertoimien PNS-estimaatteja saatiin:

$Y = 0.54545 + 0.63636*X$

b) Määrää estimoidun regressiomallin mukaiset Y-arvojen sovitteet/ennusteet sekä jäännökset/residuaalit.
c) Testaa tehtävän 1 mallin kerrointa β koskevaa nollahypoteesia H: $\beta = 0$ käyttäen t-testiä.

3. a) Hattivateilla on todettu sähköallergiaa, jota yritetään parantaa shokki- ja vierihoidolla. Sadan allergisen hattivatin
joukosta 70 sai shokkihoitoa ja loput vierihoitoa. Shokkihoitoa saaneista 49 parani allergiasta. Vierihoitoryhmässä
parantuneita hattivatteja oli 16. Testaa oliko hoitomuodolla merkitystä paranemiseen.

	parani	ei	Rivisumma
Shokkihoito	49	21	70
Vierihoito	16	14	30
Sarakesumma	65	35	100

b) Havainnollista testauksen yleisperiaatteita a)-kohdan esimerkin valossa, ts. nollahypoteesin muotoilua, kriittisen alueen määräämistä, testisuureen ja sen jakauman määräämistä .

4. a) Selitysasteen R^2 määritelmä ja käyttötarkoitus.
b) Havainnollista R^2 :n käyttökelpoisuutta hyvän mallin valintakriteerinä alla olevan havaintoaineiston tapauksessa, jossa muuttujaa S (= Sales) yritetään mallintaa muuttujien P (= Price) ja A (= Advertising) avulla. Käytä havainnollistuksessa estimoituja malleja

$$S = 4.26\,P - 1.48\,A + 176 + \text{error} \quad (R^2 = 0.955)$$

ja

$$S = 140 + 3.48\,P + \text{error} \quad (R^2 = 0.946)$$

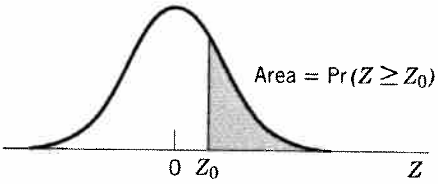
sekä mallin hyvyyden arviointia sisällöllisin perustein.
c) Miksi mallien pienestä R^2 -arvojen erosta huolimatta niiden ero sisällöllisin perustein arvioituna on suuri.

Table C.1 Sales, price and advertising data

	1979	1980	1981	1982	1983	1984	Average
Sales (in £ million), *S*	250	340	300	200	290	360	290
Price (in £), *P*	25	48	44	20	38	60	39
Advertising (in £'000), *A*	35	32	38	30	34	46	36

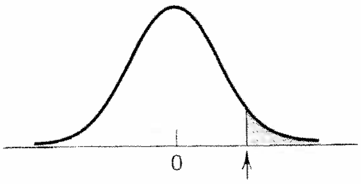
TABLE IV

Standard Normal, Cumulative Probability in Right-Hand Tail
(For Negative Values of Z, Areas Are Found by Symmetry)



Z_o	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
	Second Decimal Place of Z_o									
0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641
0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0722	.0708	.0694	.0681
1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
1.8	.0359	.0352	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
2.9	.0019	.0018	.0017	.0017	.0016	.0016	.0015	.0015	.0014	.0014
3.0	.00135									
3.5	.000 233									
4.0	.000 031 7									
4.5	.000 003 40									
5.0	.000 000 287									

t Critical Points

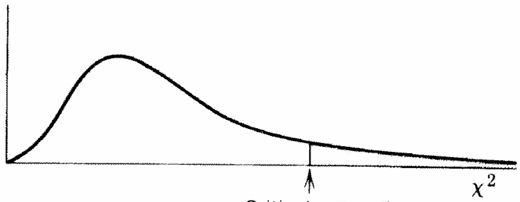


Critical point. For example:
$t_{.025}$ leaves .025 probability
in the tail.

$d.f.$	$t_{.25}$	$t_{.10}$	$t_{.05}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.0025}$	$t_{.0010}$	$t_{.0005}$
1	1.000	3.078	6.314	12.706	31.821	63.637	127.32	318.31	636.62
2	.816	1.886	2.920	4.303	6.965	9.925	14.089	22.326	31.598
3	.765	1.638	2.353	3.182	4.541	5.841	7.453	10.213	12.924
4	.741	1.533	2.132	2.776	3.747	4.604	5.598	7.173	8.610
5	.727	1.476	2.015	2.571	3.365	4.032	4.773	5.893	6.869
6	.718	1.440	1.943	2.447	3.143	3.707	4.317	5.208	5.959
7	.711	1.415	1.895	2.365	2.998	3.499	4.020	4.785	5.408
8	.706	1.397	1.860	2.306	2.896	3.355	3.833	4.501	5.041
9	.703	1.383	1.833	2.262	2.821	3.250	3.690	4.297	4.781
10	.700	1.372	1.812	2.228	2.764	3.169	3.581	4.144	4.537
11	.697	1.363	1.796	2.201	2.718	3.106	3.497	4.025	4.437
12	.695	1.356	1.782	2.179	2.681	3.055	3.428	3.930	4.318
13	.694	1.350	1.771	2.160	2.650	3.012	3.372	3.852	4.221
14	.692	1.345	1.761	2.145	2.624	2.977	3.326	3.787	4.140
15	.691	1.341	1.753	2.131	2.602	2.947	3.286	3.733	4.073
16	.690	1.337	1.746	2.120	2.583	2.921	3.252	3.686	4.015
17	.689	1.333	1.740	2.110	2.567	2.898	3.222	3.646	3.965
18	.688	1.330	1.734	2.101	2.552	2.878	3.197	3.610	3.922
19	.688	1.328	1.729	2.093	2.539	2.861	3.174	3.579	3.883
20	.687	1.325	1.725	2.086	2.528	2.845	3.153	3.552	3.850
21	.686	1.323	1.721	2.080	2.518	2.831	3.135	3.257	3.189
22	.686	1.321	1.717	2.074	2.508	2.819	3.119	3.505	3.792
23	.685	1.319	1.714	2.069	2.500	2.807	3.104	3.485	3.767
24	.685	1.318	1.711	2.064	2.492	2.797	3.091	3.467	3.745
25	.684	1.316	1.708	2.060	2.485	2.787	3.078	3.450	3.725
26	.684	1.315	1.706	2.056	2.479	2.779	3.067	3.435	3.707
27	.684	1.314	1.703	2.052	2.473	2.771	3.057	3.421	3.690
28	.683	1.313	1.701	2.048	2.467	2.763	3.047	3.408	3.674
29	.683	1.311	1.699	2.045	2.462	2.756	3.038	3.396	3.659
30	.683	1.310	1.697	2.042	2.457	2.750	3.030	3.385	3.646
40	.681	1.303	1.684	2.021	2.423	2.704	2.971	3.307	3.551
60	.679	1.296	.1671	2.000	2.390	2.660	2.915	3.232	3.460
120	.677	1.289	1.658	1.980	2.358	2.617	2.860	3.160	3.373
∞	.674 $= z_{.25}$	1.282 $= z_{.10}$	1.645 $= z_{.05}$	1.960 $= z_{.025}$	2.326 $= z_{.010}$	2.576 $= z_{.005}$	2.807 $= z_{.0025}$	3.090 $= z_{.0010}$	3.291 $= z_{.0005}$

TABLE VII

χ^2 Critical Points



Critical point. For example: $\chi^2_{.05}$ leaves 5% probability in the tail.

d.f.	$\chi^2_{.25}$	$\chi^2_{.10}$	$\chi^2_{.05}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$	$\chi^2_{.001}$
1	1.32	2.71	3.84	5.02	6.63	7.88	10.8
2	2.77	4.61	5.99	7.38	9.21	10.6	13.8
3	4.11	6.25	7.81	9.35	11.3	12.8	16.3
4	5.39	7.78	9.49	11.1	13.3	14.9	18.5
5	6.63	9.24	11.1	12.8	15.1	16.7	20.5
6	7.84	10.6	12.6	14.4	16.8	18.5	22.5
7	9.04	12.0	14.1	16.0	18.5	20.3	24.3
8	10.2	13.4	15.5	17.5	20.1	22.0	26.1
9	11.4	14.7	16.9	19.0	21.7	23.6	27.9
10	12.5	16.0	18.3	20.5	23.2	25.2	29.6
11	13.7	17.3	19.7	21.9	24.7	26.8	31.3
12	14.8	18.5	21.0	23.3	26.2	28.3	32.9
13	16.0	19.8	22.4	24.7	27.7	29.8	34.5
14	17.1	21.1	23.7	26.1	29.1	31.3	36.1
15	18.2	22.3	25.0	27.5	30.6	32.8	37.7
16	19.4	23.5	26.3	28.8	32.0	34.3	39.3
17	20.5	24.8	27.6	30.2	33.4	35.7	40.8
18	21.6	26.0	28.9	31.5	34.8	37.2	42.3
19	22.7	27.2	30.1	32.9	36.2	38.6	32.8
20	23.8	28.4	31.4	34.2	37.6	40.0	45.3
21	24.9	29.6	32.7	35.5	38.9	41.4	46.8
22	26.0	30.8	33.9	36.8	40.3	42.8	48.3
23	27.1	32.0	35.2	38.1	41.6	44.2	49.7
24	28.2	33.2	36.4	39.4	32.0	45.6	51.2
25	29.3	34.4	37.7	40.6	44.3	46.9	52.6
26	30.4	35.6	38.9	41.9	45.6	48.3	54.1
27	31.5	36.7	40.1	43.2	47.0	49.6	55.5
28	32.6	37.9	41.3	44.5	48.3	51.0	56.9
29	33.7	39.1	42.6	45.7	49.6	52.3	58.3
30	34.8	40.3	43.8	47.0	50.9	53.7	59.7
40	45.6	51.8	55.8	59.3	63.7	66.8	73.4
50	56.3	63.2	67.5	71.4	76.2	79.5	86.7
60	67.0	74.4	79.1	83.3	88.4	92.0	99.6
70	77.6	85.5	90.5	95.0	100	104	112
80	88.1	96.6	102	107	112	116	125
90	98.6	108	113	118	124	128	137
100	109	118	124	130	136	140	149

Tilastotieteen jatkokurssi kevät 2009 korvaava 2. välikoe 19.5.2009

1. Muuttujien x ja y havaitut arvot ovat:

```
x 1 3 4 6 8 9 11 14
y 1 2 4 4 5 7  8  9
```

Määrää regressiomallin $y = \alpha + \beta x + \varepsilon$ ...

a) β-kerroin

b) vakiotermi α.

c) määritä β-kertoimen luottamusväli (95 % luottamustaso).

2 . Tiedetään, että alueella A asukkaiden keskitulo on 1850 €, ja että tulojen keskihajonta on 150 €.

a) Millä välillä 200 alueelta satunnaisesti valitun asukkaan keskitulo vaihtelee 95 %:n todennäköisyydellä.

Alueella B tehtiin vastaava tutkimus ja keskituloksi saatiin 60 havainnon otoksessa 1900 €.

b) Millä välillä alueen A:n keskitulo voi 95 %:n todennäköisyydellä vaihdella, mikäli tulojen
hajonta oletetaan tuntemattomaksi ja käytetään otoksen perusteella arvioitua hajontaa s = 200 €.

c) Miten selittäisit ei-tilastotieteilijälle kohtien a) ja b) tulosten eron syyn?

3) a) Miten havainnollistaisit ei-tilastotieteilijälle tilastollisen riippumattomuuden käsitettä.

b) Luokitusten riippumattomuuden testaus esimerkin tapauksessa: Hiusten väri & silmien väri (ks. oheinen havaintoaineisto).

4. a) Eräässä 50 kunnan otoksessa suhteellisen rikollisuuden (rikoksia per 1000 asukasta) ja asukastiheyden (asukasta per neliökilometri) välinen korrelaatiokerroin oli r = 0.157 .

Testaa hypoteesia, että ko. muuttujat ovat korreloimattomia.

b) Miten vastaisit ei-tilastotieteilijän kysymykseen: "Mitä korrelaatiokerroin mittaa?"

Table B.3 Observed frequencies of people with a particular combination of hair and eye colour

	Hair colour			
Eye colour	Black	Brunette	Red	Blond
Brown	68	119	26	7
Blue	20	84	17	94
Hazel	15	54	14	10
Green	5	29	14	16

Tilastotieteen kaavakokoelma

$$P(A^c) = P(C(A)) = 1 - P(A) \qquad P(A - B) = P(A \bigcap B^c)$$

$$P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B) = P(A) + P(B) - P(AB)$$

$$P(A \bigcap B) = P(A) \cdot P(B|A) = P(B) \cdot P(A|B) \qquad P(A|B) = \frac{P(A \bigcap B)}{P(B)}$$

$$P(A \cdot B \cdot C \cdot D) = P(A) \cdot P(B|A) \cdot P(C|AB) \cdot P(D|ABC)$$

$$P(B) = \sum_1^k P(A_i) P(B|A_i) \qquad E = \bigcup_1^k A_i \qquad \bigcap A_i A_j = \emptyset \qquad i \neq j$$

$$P(B_j \mid A) = \frac{P(A \mid B_j) \cdot P(B_j)}{\sum_{i=1}^{k} P(A \mid B_i) \cdot P(B_i)}$$

$$P(x = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k} = \frac{n!}{k!(n-k)!} \pi^k (1 - \pi)^{n-k}$$

$$P(x = k | N, K, n) = \frac{\binom{K}{k} \binom{N-K}{n-k}}{\binom{N}{n}}$$

$$k! = 1 \cdot 2 \cdot 3 \ldots (k-1) \cdot k$$

$$E(X) = \sum x_i p_i = x_1 p_1 + x_2 p_2 + x_3 p_3 + \cdots + x_n p_n$$

$$D^2(X) = \sum p_i \cdot (x_i - E(X))^2 = E(X^2) - (E(X))^2$$

$$\hat{p} = \frac{x}{n} \qquad z = \frac{\hat{p} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim N(0,1) \qquad z = \frac{\hat{p} - \pi}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \sim N(0,1)$$

$$z = \frac{\overline{x} - \mu}{\sigma / \sqrt{n}} \sim N(0,1) \qquad t = \frac{\overline{x} - \mu}{s / \sqrt{n}} \sim t(n-1)$$

$$P(-|z_{\alpha/2}| \leq \frac{\overline{x} - \mu}{\sigma / \sqrt{n}} \leq |z_{\alpha/2}|) = \alpha \qquad P(-|z_{\alpha/2}| \leq \frac{\hat{p} - \pi}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}} \leq |z_{\alpha/2}|) = \alpha$$

$$\overline{x} - |z_{\alpha/2}| \cdot \sigma / \sqrt{n} \leq \mu \leq \overline{x} + |z_{\alpha/2}| \cdot \sigma / \sqrt{n}$$

$$\hat{p} - |z_{\alpha/2}| \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \leq \pi \leq \hat{p} + |z_{\alpha/2}| \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\overline{x} - |t_{\alpha/2}(n-1)| \cdot S/\sqrt{n} \leq \mu \leq \overline{x} + |t_{\alpha/2}(n-1)| \cdot S/\sqrt{n}$$

$$\tilde{p} = \tfrac{(k+2)}{(n+4)} \qquad P(-|z_{\alpha/2}| \leq \frac{\tilde{p}-\pi}{\sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}} \leq |z_{\alpha/2}|) = \alpha$$

$$\tilde{p} - |z_{\alpha/2}| \cdot \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}} \leq \pi \leq \tilde{p} + |z_{\alpha/2}| \cdot \sqrt{\frac{\tilde{p}(1-\tilde{p})}{n+4}}$$

$$P(-|t_{\alpha/2}(n-1)| \leq \frac{\overline{x}-\mu}{s/\sqrt{n}} \leq |t_{\alpha/2}(n-1))| = \alpha$$

$$z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1) \qquad z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim N(0,1) \qquad \hat{p} = \frac{n_1\hat{p}_1 + n_2\hat{p}_2}{n_1 + n_2}$$

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim_{appr} N(0,1) \text{ Kun } n_1 \text{ ja } n_2 \text{ suuria.}$$

$$t = \frac{\overline{x}_1 - \overline{x}_2}{S_d} = \frac{\overline{x}_1 - \overline{x}_2}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2) \qquad S = \sqrt{\frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2}}$$

$$X^2 = \sum \frac{(o_i - e_i)^2}{e_i} \sim \chi^2(k - r - 1)$$

$$n = \left(\frac{|z_{\alpha/2}| \cdot \sqrt{\pi \cdot (1-\pi)})}{d} \right)^2 \qquad n = \frac{(t_{1-\alpha/2}(n-1) \cdot S)^2}{d^2}$$

$$\overline{x} = \tfrac{1}{n} \textstyle\sum_{i=1}^{n} x_i \qquad \overline{x} = \tfrac{1}{n} \textstyle\sum_{i=1}^{k} n_i x_i \qquad \overline{x} = \textstyle\sum_{i=1}^{n} f_i x_i$$

$$S^2 = \tfrac{1}{(n-1)} \textstyle\sum_{i=1}^{n} (x_i - \overline{x})^2 = \frac{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}{(n-1)} = \frac{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}{(n-1)}$$

$$S^2 = \tfrac{1}{(n-1)} \textstyle\sum_{i=1}^{k} n_i (x_i - \overline{x})^2 = \frac{\sum_{i=1}^{k} n_i x_i^2 - \frac{\left(\sum_{i=1}^{k} n_i x_i\right)^2}{n}}{(n-1)}$$

$$S = \sqrt{S^2} = \sqrt{\tfrac{1}{(n-1)} \textstyle\sum_{i=1}^{n} (x_i - \overline{x})^2} \qquad r.m.s = \sqrt{\tfrac{1}{n} \textstyle\sum_{i=1}^{n} x_i^2} \qquad V = \tfrac{S}{\overline{x}}$$

$$P(p - z_{\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}} \leq \pi \leq p + z_{\alpha/2} \cdot \sqrt{\frac{p \cdot (1-p)}{n}}) = 1 - \alpha$$

$$P(\overline{x} - t_{(n-1)}^{\alpha/2} \cdot \sqrt{\frac{S^2}{n}} \leq \mu \leq \overline{x} + t_{(n-1)}^{\alpha/2} \cdot \sqrt{\frac{S^2}{n}}) = 1 - \alpha$$

$$Z = \frac{p - \pi_o}{\sqrt{\pi_o(1 - \pi_o)/n}} \qquad Z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$p = \frac{n_1 \cdot p_1 + n_2 \cdot p_2}{n_1 + n_2} \qquad z \sim N(0,1)$$

$$Z = \frac{\overline{x} - \mu_o}{\sqrt{\sigma^2/n}} \qquad z \sim N(0,1) \qquad t = \frac{\overline{x} - \mu_o}{\sqrt{S^2/n}} \qquad t \sim t(n-1)$$

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{S^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad S^2 = \frac{(n_1 - 1) \cdot S_1^2 + (n_2 - 1) \cdot S_2^2}{n_1 + n_2 - 2} \qquad t \sim t(n_1 + n_2 - 2)$$

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad t \approx t(\nu) \qquad min(n_1 - 1, n_2 - 1) \leq \nu \leq n_1 + n_2 - 2$$

$$Z = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \qquad z \sim N(0,1) \qquad F = \frac{max(S_1^2, S_2^2)}{min(S_1^2, S_2^2)} \quad F \sim F_{os,nim}$$

$$r_{xy} = \frac{1}{(n-1)} \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{S_x}\right)\left(\frac{y_i - \overline{y}}{S_y}\right) = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\overline{y}}{(n-1) S_x S_y} \qquad Z = \frac{r\sqrt{n-2}}{\sqrt{1 - r^2}} \sim N(0,1)$$

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\left(\sum_{i=1}^{n} (x_i - \overline{x})^2\right)\left(\sum_{i=1}^{n} (y_i - \overline{y})^2\right)}} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}}{\sqrt{\left(\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}\right)\left(\sum_{i=1}^{n} y_i^2 - \frac{\left(\sum_{i=1}^{n} y_i\right)^2}{n}\right)}}$$

$$r_s = \frac{\sum_{i=1}^{n} (R(x_i) - \overline{R(x)})(R(y_i) - \overline{R(y)})}{\sqrt{\left(\sum_{i=1}^{n} (R(x_i) - \overline{R(x)})^2\right)\left(\sum_{i=1}^{n} (R(y_i) - \overline{R(y)})^2\right)}}$$

$$r_s = 1 - \frac{6\left(\sum_{i=1}^{n} (R(x_i) - (R(y_i))^2\right)}{n(n^2 - 1)} = 1 - \frac{6\left(\sum_{i=1}^{n} d^2\right)}{n(n^2 - 1)} \qquad tau = \frac{C - D}{n(n-1)/2}$$

$$T_A = \sum_{i=1}^{n_A} R(X_{Ai}) \qquad T_B = \sum_{i=1}^{n_B} R(X_{Bi})$$

$$U_A = n_A \cdot n_B + \frac{n_A(n_A + 1)}{2} - T_A \qquad U_B = n_B \cdot n_A + \frac{n_B(n_B + 1)}{2} - T_B \qquad U = min(U_A, U_B)$$

$$U \sim N\left(\mu = \frac{n_A \cdot n_B}{2}, \sigma^2 = \frac{n_A \cdot n_B \cdot (n_A + n_B + 1)}{12}\right)$$

$$\hat{y} = a + bx \qquad a = \overline{y} - b\overline{x}$$

$$b = r_{xy}\frac{S_y}{S_x} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{\sum_{i=1}^{n} x_i y_i - \frac{\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{i=1}^{n} y_i\right)}{n}}{\sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}}$$

$$e_1 = y_i - \hat{y}_i = y_i - (a + bx_i) = y_i - a - bx_i$$

$$SST = \sum_{i=1}^{n}(y_i - \overline{y})^2 \qquad SSE = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n}(y_i - (a + bx_i))^2$$

$$R^2 = r^2 = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST} \qquad R^2_{Adj} = 1 - \frac{S_e^2}{S_y^2}$$

$$S_e^2 = \frac{1}{(n-2)}\sum_{i=1}^{n} e_i^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n-2} = \frac{SSE}{n-2}$$

$$\sum_{ijh}(x_{ijh} - \overline{\overline{x}})^2 = \sum_{ijh}(\overline{x}_i - \overline{\overline{x}})^2 + \sum_{ijh}(\overline{x}_j - \overline{\overline{x}})^2 + \sum_{ijh}(\overline{x}_{ij} - \overline{x}_i - \overline{x}_j + \overline{\overline{x}})^2 + \sum_{ijh}(x_{ijh} - \overline{x}_{ij})^2$$

$$SST = SSR_r + SSR_s + SSR_{rs} + SSE \qquad n \cdot k \cdot l - 1 = (k-1) + (l-1) + (k-1)(l-1) + kl(n-1)$$

$$S_r^2 = SSR_r/(k-1) \qquad S_s^2 = SSR_s/(l-1) \qquad S_{rs}^2 = SSR_{rs}/(k-1)(l-1)$$
$$S_e^2 = SSE/kl(n-1)$$

$$F_r = S_r^2/S_e^2 \qquad F_s = S_s^2/S_e^2 \qquad F_r = S_{rs}^2/S_e^2$$

$$X^2 = \sum_{i=1}^{k} \frac{(n_i - e_i)^2}{e_i} \quad X^2 \sim \chi^2(k - r - 1) \; ; \; X^2 = \sum_{i,j=1}^{k,l} \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad X^2 \sim \chi^2(k-1)(l-1)$$

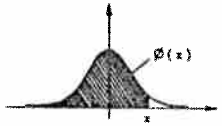
$$V = \sqrt{\frac{X^2/n}{min(\mathrm{k,l}) - 1}} \qquad C = \sqrt{\frac{X^2}{X^2 + n}}$$

TAULUKKO 1 : Standardisoitu normaalijakauma :

Kertymäfunktio $\Phi(x)$ argumentin x eri arvoilla.

$$\Phi(x) = \int_{-\infty}^{x} \varphi(u)\,du, \text{ jossa } \varphi(u) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{u^2}{2}}$$

$$\Phi(-x) = 1 - \Phi(x)$$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7703	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Esim. $\Phi(-2.18) = 1 - \Phi(2.18) = 1 - 0.9854 = 0.0146$

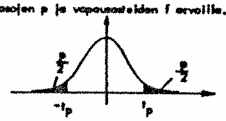
TAULUKKO 2 : Studentin t - jakauma :

Kriittisiä arvoja t_p joillakin merkitsevyystasojen p ja vapausasteiden f arvoilla.

Esim. Kaksisuuntainen testi

merkitsevyystasolla p

Jos $f=3$ ja $p=0.05$ on $P(|\underline{t}|>3.182)=0.05$



f	merkitsevyystaso p yksisuuntaisessa testissä					
	0.05	0.025	0.01	0.005	0.001	0.0005
	merkitsevyystaso p kaksisuuntaisessa testissä					
	0.10	0.05	0.02	0.01	0.002	0.001
1	6.314	12.706	31.821	63.657	318.311	636.619
2	2.920	4.303	6.965	9.925	22.326	31.598
3	2.353	3.182	4.541	5.841	10.213	12.941
4	2.132	2.776	3.747	4.604	7.173	8.610
5	2.015	2.571	3.365	4.032	5.893	6.859
6	1.943	2.447	3.143	3.707	5.208	5.959
7	1.895	2.365	2.998	3.499	4.785	5.405
8	1.860	2.306	2.896	3.355	4.501	5.041
9	1.833	2.262	2.821	3.250	4.297	4.781
10	1.812	2.228	2.764	3.169	4.144	4.587
11	1.796	2.201	2.718	3.106	4.025	4.437
12	1.782	2.179	2.681	3.055	3.930	4.318
13	1.771	2.160	2.650	3.012	3.852	4.221
14	1.761	2.145	2.624	2.977	3.787	4.140
15	1.753	2.131	2.602	2.947	3.733	4.073
16	1.746	2.120	2.583	2.921	3.686	4.015
17	1.740	2.110	2.567	2.898	3.646	3.965
18	1.734	2.101	2.552	2.878	3.611	3.922
19	1.729	2.093	2.539	2.861	3.579	3.883
20	1.725	2.086	2.528	2.845	3.552	3.850
21	1.721	2.080	2.518	2.831	3.527	3.819
22	1.717	2.074	2.508	2.819	3.505	3.792
23	1.714	2.069	2.500	2.807	3.485	3.767
24	1.711	2.064	2.492	2.797	3.467	3.745
25	1.708	2.060	2.485	2.787	3.450	3.725
26	1.706	2.056	2.479	2.779	3.435	3.707
27	1.703	2.052	2.473	2.771	3.421	3.690
28	1.701	2.048	2.467	2.763	3.408	3.674
29	1.699	2.045	2.462	2.756	3.396	3.659
30	1.697	2.042	2.457	2.750	3.385	3.646
40	1.684	2.021	2.423	2.704	3.307	3.551
50	1.676	2.009	2.403	2.678	3.262	3.495
60	1.671	2.000	2.390	2.660	3.232	3.460
80	1.664	1.990	2.374	2.639	3.195	3.415
100	1.660	1.984	2.365	2.626	3.174	3.389
200	1.653	1.972	2.345	2.601	3.131	3.339
500	1.648	1.965	2.334	2.586	3.106	3.310
∞	1.645	1.960	2.326	2.576	3.090	3.291

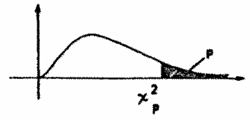
Huom.

Viimeisellä rivillä olevat luvut ovat normaalijakauman kriittisiä arvoja z_p.

TAULUKKO 3 : χ^2 - jakauma :

Yksisuuntaiseen testiin liittyviä kriittisiä arvoja χ^2_p joillakin merkitsevyystasojen p ja vapausasteiden f arvoilla.

Esim. Jos $f = 15$ ja $p = 0.05$ on $P(\underline{\chi}^2 > 24.996) = 0.05$

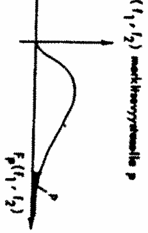


f \ p	0.99	0.95	0.10	0.05	0.01	0.001
1	0.000	0.004	2.706	3.841	6.635	10.828
2	0.020	0.103	4.605	5.991	9.210	13.816
3	0.115	0.352	6.251	7.815	11.345	16.266
4	0.297	0.711	7.779	9.488	13.277	18.467
5	0.554	1.145	9.236	11.070	15.086	20.515
6	0.872	1.635	10.645	12.592	16.812	22.458
7	1.239	2.167	12.017	14.067	18.475	24.322
8	1.646	2.733	13.362	15.507	20.090	26.125
9	2.088	3.325	14.684	16.919	21.666	27.877
10	2.558	3.940	15.987	18.307	23.209	29.588
11	3.053	4.575	17.275	19.675	24.725	31.264
12	3.571	5.226	18.549	21.026	26.217	32.909
13	4.107	5.892	19.812	22.362	27.688	34.528
14	4.660	6.571	21.064	23.685	29.141	36.123
15	5.229	7.261	22.307	24.996	30.578	37.697
16	5.812	7.962	23.542	26.296	32.000	39.252
17	6.408	8.672	24.769	27.587	33.409	40.790
18	7.015	9.390	25.989	28.869	34.805	42.312
19	7.633	10.117	27.204	30.144	36.191	43.820
20	8.260	10.851	28.412	31.410	37.566	45.315
21	8.897	11.591	29.615	32.671	38.932	46.797
22	9.542	12.338	30.813	33.924	40.289	48.268
23	10.196	13.091	32.007	35.172	41.638	49.728
24	10.856	13.848	33.196	36.415	42.980	51.179
25	11.524	14.611	34.382	37.652	44.314	52.620
26	12.198	15.379	35.563	38.885	45.642	54.052
27	12.879	16.151	36.741	40.113	46.963	55.476
28	13.565	16.928	37.916	41.337	48.278	56.892
29	14.256	17.708	39.087	42.557	49.588	58.302
30	14.953	18.493	40.256	43.773	50.892	59.703
40	22.164	26.509	51.805	55.758	63.691	73.402
50	29.707	34.764	63.167	67.505	76.154	86.661
60	37.485	43.188	74.397	79.082	88.379	99.607
70	45.442	51.739	85.527	90.531	100.425	112.317
80	53.540	60.391	96.578	101.879	112.329	124.839
90	61.754	69.126	107.565	113.145	124.116	137.208
100	70.065	77.930	118.498	124.342	135.806	149.448

TAULUKKO 4.1. : Fisherin F - jakauma :

Yksisuuntaiseen testiin liittyviä kriittisiä arvoja $F_p(f_1, f_2)$ merkitsevyystasolla p vapausasteiden f_1 ja f_2 eri kombinaatioilla.

$$F(f_1, f_2) = \frac{\chi_1^2/f_1}{\chi_2^2/f_2} = \frac{s_1^2}{s_2^2}$$



f_2 \ f_1	1	2	3	4	5	6	7	8	12	16	24	∞
	merkitsevyystaso p = 0.05											
1	161	200	216	225	230	234	237	239	244	246	249	254
2	18.5	19.0	19.2	19.2	19.3	19.3	19.3	19.4	19.4	19.4	19.5	19.5
3	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.74	8.69	8.64	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.91	5.84	5.77	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.68	4.60	4.53	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.00	3.92	3.84	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.57	3.49	3.41	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.28	3.20	3.12	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.07	2.99	2.90	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.91	2.83	2.74	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.79	2.70	2.61	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.69	2.60	2.51	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.60	2.51	2.42	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.53	2.44	2.35	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.48	2.38	2.29	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.42	2.33	2.24	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.38	2.29	2.19	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.34	2.25	2.15	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.31	2.21	2.11	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.28	2.18	2.08	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.25	2.16	2.05	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.23	2.13	2.03	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.20	2.11	2.00	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.18	2.09	1.98	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.16	2.07	1.96	1.71
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.15	2.05	1.95	1.69
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.13	2.04	1.93	1.67
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.12	2.02	1.91	1.65
29	4.18	3.33	2.93	2.70	2.55	2.43	2.35	2.28	2.10	2.01	1.90	1.64
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.09	1.99	1.89	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.00	1.90	1.79	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	1.92	1.82	1.70	1.39
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.85	1.75	1.63	1.28
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.75	1.64	1.52	1.00

Esim. $F_{0.05}(3, 15) = 3.29$ eli $P(\underline{F}(3, 15) \geq 3.29) = 0.05$