Tentamen

Linköpings universitet, Institutionen för datavetenskap, STIMA

Kurskod och namn:	732G45, Grundläggande statistik och dataanlys
Datum och tid:	2018-10-30, 8-12
Jourhavande lärare:	Isak Hietala
Tillåtna hjälpmedel:	Räknedosa av valfri modell, två stycken dubbelsidiga A4 med egna anteckningar, tabellsamling utan anteckningar
Betygsgränser:	Tentamen omfattar totalt 20 poäng, G från 12p, VG från 16p Siffrorna i uppgifterna är påhittade. Saknas någon siffra för att kunna lösa uppgiften? Skriv då tydligt ut att du saknar denna information, anta ett godtyckligt värde och lös uppgiften med detta antagande.

Redovisa, tolka och motivera tydligt alla dina lösningar!

Uppgift 1 (4p)

a) Vad skiljer vetenskap från icke-vetenskap enligt Nyquist? (2p)

b) Vad kännetecknas av den hypotetiskt-deduktiva metoden? Beskriv med hjälp av exempel. (2p)

Uppgift 2 (5p)

Ett slumpmässigt urval av 8 studenter på en idrottsutbildning har hoppat höjdhopp. Följande höjder i centimeter uppmättes: 179, 185, 182, 178, 194, 231, 183, 185.

a) Är variabeln "höjd" kvalitativ eller kvantitativ? Vilken variabelskala är variabeln mätt på? (1p)

b) Beräkna medelvärde, median och typvärde för höjd. Är det något/några av dessa tre lägesmått som bättre beskriver mitten på datamaterialet jämfört med något/några av de andra? Motivera. (2.5p)

c) Beräkna standardavvikelsen. (1.5p)

Uppgift 3 (4p)

I rollspelet Dungeons and Dragons, kastar spelare en jämviktad 20-sidig tärning varje gång man ska utföra någonting, t.ex. öppna en låst dörr eller klättra upp för ett träd, för att se huruvida utförandet lyckas eller misslyckas. Ett lyckat utfall brukar generellt definieras som värden på minst 12.

a) Vad är sannolikheten att man på 5 kast lyckas med det man ska utföra fler än 1 gång? (2p)

Om man lyckas kasta en 20:a på tärningen definieras det som ett perfekt utförande och man får någon form av bonus. Vi kan anta att man under ett spel kastar tärningen ca 250 gånger.

b) Vad är sannolikheten att man under ett spel kastar fler än 20 bonusgivande kast? (2p)

Uppgift 4 (3p)

Tabell 1 visar en föreslagen sannolikhetsfördelning som ska ha beräknats enligt funktionen $P(Y = y) = \frac{y}{40}$, där y är de enskilda utfallen från tabellen.

Tabell 1: Föreslagen sannolikhetsfördelning

y	p(y)
2	0.050
5	0.150
8	0.200
11	0.275
14	0.350

a) Stämmer det att tabell 1 är en sannolikhetsfördelning? Motivera. (1p)

Om du i a) bedömt att fördelningen i tabell 1 inte är en sannolikhetsfördelning, korrigera den innan du löser nästkommande uppgifter.

b) Beräkna väntevärdet och variansen av den givna sannolikhetsfördelningen. (2p)

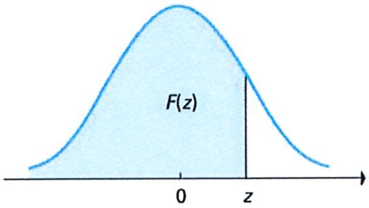
Uppgift 5 (2p)

Sannolikheten att det kommer regna under morgondagen är ca 20 procent. Sannolikheten att det kommer vara soligt är ca 40 procent. Sannolikheten att det antingen kommer regna eller vara soligt är ca 52 procent.

Bedöm ifall dessa händelser är oberoende samt ifall de är disjunkta. (2p)

Appendix Tables

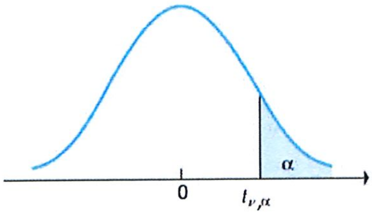
Table 1 Cumulative Distribution Function, $F(z)$, of the Standard Normal Distribution Table



z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997

Dr. William L. Carlson, prepared using Minitab 16.

Table 8 Upper Critical Values of Student's t Distribution with ν Degrees of Freedom

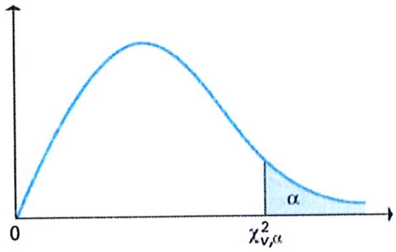


For selected probabilities, α, the table shows the values $t_{\nu,\alpha}$ such that $P(t_\nu > t_{\nu,\alpha}) = \alpha$, where t_ν is a Student's t random variable with ν degress of freedom. For example, the probability is .10 that a Student's t random variable with 10 degrees of freedom exceeds 1.372.

	Probability of Exceeding the Critical Value					
ν	0.10	0.05	0.025	0.01	0.005	0.001
1	3.078	6.314	12.706	31.821	63.657	318.313
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.215
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.782
8	1.397	1.860	2.306	2.896	3.355	4.499
9	1.383	1.833	2.262	2.821	3.250	4.296
10	1.372	1.812	2.228	2.764	3.169	4.143
11	1.363	1.796	2.201	2.718	3.106	4.024
12	1.356	1.782	2.179	2.681	3.055	3.929
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.624	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.583	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
21	1.323	1.721	2.080	2.518	2.831	3.527
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
25	1.316	1.708	2.060	2.485	2.787	3.450
26	1.315	1.706	2.056	2.479	2.779	3.435
27	1.314	1.703	2.052	2.473	2.771	3.421
28	1.313	1.701	2.048	2.467	2.763	3.408
29	1.311	1.699	2.045	2.462	2.756	3.396
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
60	1.296	1.671	2.000	2.390	2.660	3.232
100	1.290	1.660	1.984	2.364	2.626	3.174
∞	1.282	1.645	1.960	2.326	2.576	3.090

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, September 2011.

Table 7a Upper Critical Values of Chi-Square Distribution with ν Degrees of Freedom

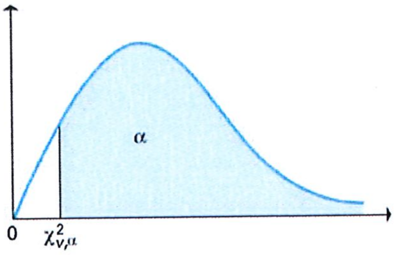


For selected probabilities α, the table shows the values $\chi^2_{v,\alpha}$ such that $P(\chi^2_v > \chi^2_{v,\alpha}) = \alpha$, where χ^2_v is a chi-square random variable with v degress of freedom. For example, the probability is .100 that a chi-square random variable with 10 degrees of freedom is greater than 15.987.

	Probability of Exceeding the Critical Value				
ν	0.10	0.05	0.025	0.01	0.001
1	2.706	3.841	5.024	6.635	10.828
2	4.605	5.991	7.378	9.210	13.816
3	6.251	7.815	9.348	11.345	16.266
4	7.779	9.488	11.143	13.277	18.467
5	9.236	11.070	12.833	15.086	20.515
6	10.645	12.592	14.449	16.812	22.458
7	12.017	14.067	16.013	18.475	24.322
8	13.362	15.507	17.535	20.090	26.125
9	14.684	16.919	19.023	21.666	27.877
10	15.987	18.307	20.483	23.209	29.588
11	17.275	19.675	21.920	24.725	31.264
12	18.549	21.026	23.337	26.217	32.910
13	19.812	22.362	24.736	27.688	34.528
14	21.064	23.685	26.119	29.141	36.123
15	22.307	24.996	27.488	30.578	37.697
16	23.542	26.296	28.845	32.000	39.252
17	24.769	27.587	30.191	33.409	40.790
18	25.989	28.869	31.526	34.805	42.312
19	27.204	30.144	32.852	36.191	43.820
20	28.412	31.410	34.170	37.566	45.315
21	29.615	32.671	35.479	38.932	46.797
22	30.813	33.924	36.781	40.289	48.268
23	32.007	35.172	38.076	41.638	49.728
24	33.196	36.415	39.364	42.980	51.179
25	34.382	37.652	40.646	44.314	52.620
26	35.563	38.885	41.923	45.642	54.052
27	36.741	40.113	43.195	46.963	55.476
28	37.916	41.337	44.461	48.278	56.892
29	39.087	42.557	45.722	49.588	58.301
30	40.256	43.773	46.979	50.892	59.703
40	51.805	55.758	59.342	63.691	73.402
50	63.167	67.505	71.420	76.154	86.661
60	74.397	79.082	83.298	88.379	99.607
70	85.527	90.531	95.023	100.425	112.317
80	96.578	101.879	106.629	112.329	124.839
90	107.565	113.145	118.136	124.116	137.208
100	118.498	124.342	129.561	135.807	149.449

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, September 2011.

Table 7b Lower Critical Values of Chi-Square Distribution with ν Degrees of Freedom



For selected probabilities α, the table shows the values $\chi^2_{v,\alpha}$ such that $P(\chi^2_v > \chi^2_{v,\alpha}) = \alpha$, where χ^2_v is a chi-square random variable with v degress of freedom. For example, the probability is 0.90 that a chi-square variable with 10 degrees of freedom is greater than 4.865.

	PROBABILITY OF EXCEEDING THE CRITICAL VALUE				
ν	0.90	0.95	0.975	0.99	0.999
1	.016	.004	.001	.000	.000
2	.211	.103	.051	.020	.002
3	.584	.352	.216	.115	.024
4	1.064	.711	.484	.297	.091
5	1.610	1.145	.831	.554	.210
6	2.204	1.635	1.237	.872	.381
7	2.833	2.167	1.690	1.239	.598
8	3.490	2.733	2.180	1.646	.857
9	4.168	3.325	2.700	2.088	1.152
10	4.865	3.940	3.247	2.558	1.479
11	5.578	4.575	3.816	3.053	1.834
12	6.304	5.226	4.404	3.571	2.214
13	7.042	5.892	5.009	4.107	2.617
14	7.790	6.571	5.629	4.660	3.041
15	8.547	7.261	6.262	5.229	3.483
16	9.312	7.962	6.908	5.812	3.942
17	10.085	8.672	7.564	6.408	4.416
18	10.865	9.390	8.231	7.015	4.905
19	11.651	10.117	8.907	7.633	5.407
20	12.443	10.851	9.591	8.260	5.921
21	13.240	11.591	10.283	8.897	6.447
22	14.041	12.338	10.982	9.542	6.983
23	14.848	13.091	11.689	10.196	7.529
24	15.659	13.848	12.401	10.856	8.085
25	16.473	14.611	13.120	11.524	8.649
26	17.292	15.379	13.844	12.198	9.222
27	18.114	16.151	14.573	12.879	9.803
28	18.939	16.928	15.308	13.565	10.391
29	19.768	17.708	16.047	14.256	10.986
30	20.599	18.493	16.791	14.953	11.588
40	29.051	26.509	24.433	22.164	17.916
50	37.689	34.764	32.357	29.707	24.674
60	46.459	43.188	40.482	37.485	31.738
70	55.329	51.739	48.758	45.442	39.036
80	64.278	60.391	57.153	53.540	46.520
90	73.291	69.126	65.647	61.754	54.155
100	82.358	77.929	74.222	70.065	61.918

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, September 2011.

Table 10 Cutoff Points for the Distribution of the Wilcoxon Test Statistic

For sample size n, the table shows, for selected probabilities α, the numbers T_α such that $P(T < T_\alpha) = \alpha$, where the distribution of the random variable T is that of the Wilcoxon test statistic under the null hypothesis.

n	α				
	.005	.010	.025	.050	.100
4	0	0	0	0	1
5	0	0	0	1	3
6	0	0	1	3	4
7	0	1	3	4	6
8	1	2	4	6	9
9	2	4	6	9	11
10	4	6	9	11	15
11	6	8	11	14	18
12	8	10	14	18	22
13	10	13	18	22	27
14	13	16	22	26	32
15	16	20	26	31	37
16	20	24	30	36	43
17	24	28	35	42	49
18	28	33	41	48	56
19	33	38	47	54	63
20	38	44	53	61	70

Reproduced with permission from R. L. McCormack, "Extended tables of the Wilcoxon matched pairs signed rank statistics," *Journal of the American Statistical Association* 60 (1965).

Table 11 Cutoff Points for the Distribution of Spearman Rank Correlation Coefficient

For sample size n, the table shows, for selected probabilities α, the numbers $r_{s,\alpha}$ such that $P(r_s > r_{s,\alpha}) = \alpha$, where the distribution of the random variable r_s is that of Spearman rank correlation coefficient under the null hypothesis of no association.

n	α			
	.050	.025	.010	.005
5	.900	—	—	—
6	.829	.886	.943	—
7	.714	.786	.893	—
8	.643	.738	.833	.881
9	.600	.683	.783	.833
10	.564	.648	.745	.794
11	.523	.623	.736	.818
12	.497	.591	.703	.780
13	.475	.566	.673	.745
14	.457	.545	.646	.716
15	.441	.525	.623	.689
16	.425	.507	.601	.666
17	.412	.490	.582	.645
18	.399	.476	.564	.625
19	.388	.462	.549	.608
20	.377	.450	.534	.591
21	.368	.438	.521	.576
22	.359	.428	.508	.562
23	.351	.418	.496	.549
24	.343	.409	.485	.537
25	.336	.400	.475	.526
26	.329	.392	.465	.515
27	.323	.385	.456	.505
28	.317	.377	.448	.496
29	.311	.370	.440	.487
30	.305	.364	.432	.478

Reproduced with permission from E. G. Olds, "Distribution of sums of squares of rank differences for small samples," *Annals of Mathematical Statistics* 9 (1938).