Department of Mathematics and Statistics, University of Helsinki
Johdatus tilastolliseen päättelyyn (Arjas), final exam 19.5.2009, English text

1. The weights of four students are 70 kg, 73 kg, 81 kg and 88 kg. Suppose you consider a random sub-sample of two of them. Let X be the absolute value of the difference of their weights. Determine the distribution of X, and its expected value and variance.

2. A society has in its statutes a paragraph according to which a motion made towards changing these statutes needs to be approved by at least ¾ of the total membership. The total membership of the society is several thousand. In order to have at a realistic idea about the support of the motion among the membership at large, the Board of the society sends questionnaires to 100 randomly selected members, asking them to present their position towards the motion. All 100 return the questionnaire, and 87 of them support the motion and 13 do not. Imagine that you are a member of the Board of this society, and the only one who has some knowledge of statistics. How would you explain the result of this simple survey to the other members of the Board, also properly accounting for the uncertainty of the conclusions that might be drawn from it? (Hint: You can use either a confidence interval based on the result, or a corresponding statistical test.)

3. Ville is interested in finding out the probability of a drawing pin 'landing on its back' when it is dropped on the floor. He decides to apply Bayesian inference for solving this problem, and then feels that the Uniform(0, 1) -distribution describes adequately his prior beliefs about the possible values of the parameter θ of the corresponding Binomial experiment. Ville then drops the drawing pin on the floor 100 times, finding that it lands 43 times on its back. Supposing that he would make one more try, what would be his (predictive) probability of the pin then landing on its back? Explain your reasoning.

4. The level of various substances in the blood of kidney dialysis patients is of concern because kidney failure and dialysis can lead to nutritional problems. A researcher performed blood tests on several dialysis patients on six consecutive clinic visits. One variable measured was the level of phosphate in the blood. Phosphate levels for a single person tend to vary normally over time. The data on one patient, in milligrams of phosphate per deciliter (mg/dl) of blood, are given below:

5.6 5.1 4.6 4.8 5.7 6.4

(a) Determine a 90% confidence interval for the patient's mean phosphate level.
(b) A friend of yours asks: Does this mean that the mean phosphate level of this patient lies in this interval with 90% probability? How do you respond? Explain your own reasoning.

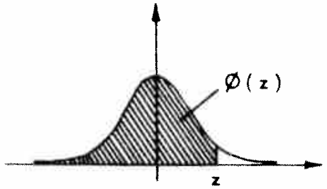
5. The normal range for blood phosphate levels is 2.6 to 4.8 mg/dl. The sample mean for the patient in problem 4 (above) falls above this range. Can this be considered to be sufficient evidence that this patient's mean level in fact falls above 4.8? State the corresponding statistical hypothesis H and its alternative A, and use these data to carry

TAULUKKO 1 : Standardisoitu normaalijakauma :

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

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

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



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

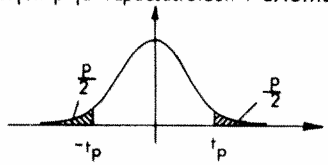
TAULUKKO 2 : Studentin t - jakauma :

Kriittisiä arvoja t_p joillakin merkitsevyystasojen p ja vapausasteiden f arvoilla.

Esim. Kaksisuuntainen testi

merkitsevyystasolla p

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



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.

Department of Mathematics and Statistics
Johdatus tilastolliseen päättelyyn (Arjas), final exam 11.6.2009 (English text)

1. Pekka and Paavo are playing in a golf tournament. Their scores vary as they play the course repeatedly. Pekka's score X has the N(110, 10^2) distribution and Paavo's score Y varies from round to round according to the N(100, 8^2) distribution. If they play independently, what is the probability that Pekka will score lower than Paavo and thus do better in the tournament?

2. Pekka has a number of white and black balls, which, apart from their colour, are identical. Hidden from you, he places one such ball into a box, having first determined its colour with a coin toss: white if the coin lands heads, and black if it lands tails. He then places another ball into the box. This time you see its colour, and it is white. After shaking the box, you are allowed to pick 'blindly' a ball from the box. What is the probability that it is white? Supposing it is, what is the probability that also the ball which is still left in the box is white? Explain your reasoning by applying Bayes' formula.

3. Consider an independent random sample $X_1, X_2, \ldots, X_n$ from a Poisson(λ) –distribution. Suppose you want to estimate the value of parameter λ on the basis of such an observed sample, by applying Bayesian inference and using the Gamma(α, β) –distribution, with density function given by $p(\lambda;\alpha, \beta) = \exp\{-\beta\lambda\}\beta^{\alpha}\lambda^{\alpha-1})/ \Gamma(\alpha)$, as the prior. Show that in this case also the posterior distribution is a Gamma distribution and find its parameters. Denoting by X_{n+1} "the next observation" in the sample, find its (posterior predictive) expectation $E(X_{n+1} \mid X_1, X_2, \ldots, X_n)$. *Hints:* The Poisson($\lambda$) –distribution is defined by $P(X = x) = \exp(- \lambda)\lambda^x /x!$, and the expectation of the Gamma(α, β) –distribution is α/β.

4. An man claims to be able to detect the presence of water with a forked stick. In a test of this claim, he is presented with five identical barrels, some containing water and some not.
   (a) Suppose the man has probability p of being correct. If he is just guessing, $p = ½$. State an appropriate null hypothesis H and its alternative A in terms of p for a test of whether he does better than guessing.
   (b) If the man is simply guessing, what is the distribution of X, the number of correct answers in the five tries?
   (c) The observed outcome turned out to be $X = 4$. What is the p-value of the test that takes large values of X to be evidence against H?

5. The table below gives the pre-test and post-test scores on listening test in English for 10 high school students who attended an intensive summer course in English:

Person no.	1	2	3	4	5	6	7	8	9	10
Pre-test score	30	28	31	26	20	30	34	15	28	20
Post-test score	29	30	32	30	16	25	31	18	33	25

   (a) We hope to show that attending the course improves listening skills. State an appropriate hypothesis H and its alternative A.
   (b) Carry out a corresponding significance test. Can you reject H at the 5 % significance level?
   (c) Give a 90 confidence interval for the mean increase in listening score due to attending the summer course.

THE STATISTICAL TABLES THAT ARE NEEDED WILL BE MADE AVAILABLE IN THE EXAM. "CLOSED BOOK" EXAM: NO OTHER SOURCE MATERIAL IS ALLOWED!

Taulukko 1. Standardinormaalijakauman kertymäfunktion Φ arvoja, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2}\, dt$.

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.500000	.503989	.507978	.511966	.515953	.519938	.523922	.527903	.531881	.535856
0.1	.539828	.543795	.547758	.551717	.555670	.559618	.563560	.567495	.571424	.575345
0.2	.579260	.583166	.587064	.590954	.594835	.598706	.602568	.606420	.610261	.614092
0.3	.617911	.621720	.625616	.629300	.633072	.636831	.640576	.644309	.648027	.651732
0.4	.655422	.659097	.662757	.666402	.670031	.673645	.677242	.680822	.684386	.687933
0.5	.691462	.694974	.698468	.702944	.705402	.708840	.712260	.715661	.719043	.722405
0.6	.725747	.729069	.732371	.735653	.738914	.742154	.745373	.748571	.751748	.754903
0.7	.758036	.761148	.764238	.767305	.770350	.773373	.776373	.779350	.782305	.785236
0.8	.788145	.791030	.793892	.796731	.799546	.802338	.805106	.807850	.810570	.813267
0.9	.815940	.818589	.821214	.823814	.826391	.828944	.831472	.833977	.836457	.838913
1.0	.841345	.843752	.846136	.848495	.850830	.853141	.855428	.857690	.859929	.862143
1.1	.864334	.866500	.868643	.870762	.872857	.874928	.876976	.879000	.881000	.882977
1.2	.884930	.886861	.888768	.890651	.892512	.894350	.896165	.897958	.899727	.901475
1.3	.903200	.904902	.906582	.908241	.909877	.911492	.913085	.914656	.916207	.917736
1.4	.919243	.920730	.922196	.923642	.925066	.926471	.927855	.929219	.930563	.931889
1.5	.933193	.934478	.935744	.936992	.938220	.939429	.940620	.941792	.942947	.944083
1.6	.945201	.946301	.947384	.948449	.949497	.950528	.951543	.952540	.953521	.954486
1.7	.955434	.956367	.957284	.958185	.959070	.959941	.960796	.961636	.962462	.963273
1.8	.964070	.964852	.965620	.966375	.967116	.967843	.968557	.969258	.969946	.970621
1.9	.971283	.971933	.972571	.973197	.973810	.974412	.975002	.975581	.976148	.976704
2.0	.977250	.977784	.978308	.978822	.979325	.979818	.980301	.980774	.981237	.981691
2.1	.982136	.982571	.982997	.983414	.983823	.984222	.984614	.984997	.985371	.985738
2.2	.986097	.986447	.986791	.987126	.987454	.987776	.988089	.988396	.988696	.988989
2.3	.989276	.989556	.989830	.990097	.990358	.990613	.990862	.991106	.991344	.991576
2.4	.991802	.992024	.992240	.992451	.992656	.992857	.993053	.993244	.993431	.993613
2.5	.993790	.993963	.994132	.994297	.994457	.994614	.994766	.994915	.995060	.995201
2.6	.995339	.995473	.995604	.995731	.995855	.995975	.996093	.996207	.996319	.996427
2.7	.996533	.996636	.996736	.996833	.996928	.997020	.997110	.997197	.997282	.997365
2.8	.997445	.997523	.997599	.997673	.997744	.997814	.997882	.997948	.998012	.998074
2.9	.998134	.998193	.998250	.998305	.998359	.998411	.998462	.998511	.998559	.998605
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3.0	.998650	.999032	.999313	.999517	.999663	.999767	.999841	.999892	.999928	.999952

t-jakauma vapausasteluvuin f.

TABLE IV. FRACTILES OF THE t DISTRIBUTION. $t_{1-P} = -t_P$.

P / f	Probability in per cent									
	60	70	80	90	95	97·5	99	99·5	99·9	99·95
1	·325	·727	1·376	3·078	6·314	12·71	31·82	63·66	318·3	636·6
2	·289	·617	1·061	1·886	2·920	4·303	6·965	9·925	22·33	31·60
3	·277	·584	·978	1·638	2·353	3·182	4·541	5·841	10·22	12·94
4	·271	·569	·941	1·533	2·132	2·776	3·747	4·604	7·173	8·610
5	·267	·559	·920	1·476	2·015	2·571	3·365	4·032	5·893	6·859
6	·265	·553	·906	1·440	1·943	2·447	3·143	3·707	5·208	5·959
7	·263	·549	·896	1·415	1·895	2·365	2·998	3·499	4·785	5·405
8	·262	·546	·889	1·397	1·860	2·306	2·896	3·355	4·501	5·041
9	·261	·543	·883	1·383	1·833	2·262	2·821	3·250	4·297	4·781
10	·260	·542	·879	1·372	1·812	2·228	2·764	3·169	4·144	4·587
11	·260	·540	·876	1·363	1·796	2·201	2·718	3·106	4·025	4·437
12	·259	·539	·873	1·356	1·782	2·179	2·681	3·055	3·930	4·318
13	·259	·538	·870	1·350	1·771	2·160	2·650	3·012	3·852	4·221
14	·258	·537	·868	1·345	1·761	2·145	2·624	2·977	3·787	4·140
15	·258	·536	·866	1·341	1·753	2·131	2·602	2·947	3·733	4·073
16	·258	·535	·865	1·337	1·746	2·120	2·583	2·921	3·686	4·015
17	·257	·534	·863	1·333	1·740	2·110	2·567	2·898	3·646	3·965
18	·257	·534	·862	1·330	1·734	2·101	2·552	2·878	3·611	3·922
19	·257	·533	·861	1·328	1·729	2·093	2·539	2·861	3·579	3·883
20	·257	·533	·860	1·325	1·725	2·086	2·528	2·845	3·552	3·850
21	·257	·532	·859	1·323	1·721	2·080	2·518	2·831	3·527	3·819
22	·256	·532	·858	1·321	1·717	2·074	2·508	2·819	3·505	3·792
23	·256	·532	·858	1·319	1·714	2·069	2·500	2·807	3·485	3·767
24	·256	·531	·857	1·318	1·711	2·064	2·492	2·797	3·467	3·745
25	·256	·531	·856	1·316	1·708	2·060	2·485	2·787	3·450	3·725
26	·256	·531	·856	1·315	1·706	2·056	2·479	2·779	3·435	3·707
27	·256	·531	·855	1·314	1·703	2·052	2·473	2·771	3·421	3·690
28	·256	·530	·855	1·313	1·701	2·048	2·467	2·763	3·408	3·674
29	·256	·530	·854	1·311	1·699	2·045	2·462	2·756	3·396	3·659
30	·256	·530	·854	1·310	1·697	2·042	2·457	2·750	3·385	3·646
40	·255	·529	·851	1·303	1·684	2·021	2·423	2·704	3·307	3·551
50	·255	·528	·849	1·298	1·676	2·009	2·403	2·678	3·262	3·495
60	·254	·527	·848	1·296	1·671	2·000	2·390	2·660	3·232	3·460
80	·254	·527	·846	1·292	1·664	1·990	2·374	2·639	3·195	3·415
100	·254	·526	·845	1·290	1·660	1·984	2·365	2·626	3·174	3·389
200	·254	·525	·843	1·286	1·653	1·972	2·345	2·601	3·131	3·339
500	·253	·525	·842	1·283	1·648	1·965	2·334	2·586	3·106	3·310
∞	·253	·524	·842	1·282	1·645	1·960	2·326	2·576	3·090	3·291
2 (1 − P)	80	60	40	20	10	5	2	1	0·2	0·1

Example: $P\{t < 2{\cdot}086\} = 97{\cdot}5\%$ for $f = 20$.

$P\{|t| > t_P\} = 2(1-P)$. $P\{|t| > 2{\cdot}086\} = 5\%$ for $f = 20$.