Exam Introduction to Statistics and Probability for CreaTe (191567030)
Tuesday 6th of November 2012, 8.45-11.45 h.

The use of a non-programmable calculator (not a "GR") is allowed and advised.
A list of formulas and probability tables for the z-, t-, binomial and Poisson-distribution are provided. *Motivate your solutions!*

<u>Part 1:</u> **The exercises 1, 2 and 3 are only for those who did not participate in the first partial test or did not pass it.**

- If you try to improve your score on the first partial test, the result of this test will be cancelled.
- If you want to use the result of part 1 (Oct. 2), your **score on the second part** should be **at least 4.5**

1. The following information about the events A and B is available:
$P(A) = 0.3$, $P(B) = 0.5$ and $P(A \cap B) = 0.1$

a. Are the events A and B **either independent or mutually exclusive** (disjoint)?
b. Compute $P(A \cup B)$.
c. Compute $P(B|A)$.

2. A student designed a new app that facilitates the choice between available bars and disco's in the region. Before investing in the content of the app he wants to evaluate the need for such an app in the region. Since students will be the primary group of users he chooses a sample of n students. The students in the sample were asked whether or not they are seriously interested in such an app (at a price of 1 Euro).
p is the probability that an arbitrary student is seriously interested and X is the number of seriously interested students in the sample of n students.

a. Argue which kind of probability distribution can be applied for X (n may be either large or small). Which assumptions are necessary?

Compute or approximate:

b. $P(X > 0)$, if $n = 10$ and $p = 0.2$
c. $P(X > 25)$, if $n = 100$ and $p = 0.2$
d. $P(X > 5)$, if $n = 250$ and $p = 0.01$

3. In the elevators of the Horst-building a signboard tells us: "Maximum 1200 kg or 16 persons". So, if 16 persons have an average **weight of at most 75 kg**, the elevator is not overloaded ($16 \times 75 = 1200$). Assume that the weights of the users of the elevator are independent and all normally distributed with mean 72 kg and standard deviation 6 kg.
$X_1, X_2, \ldots, X_{16}$ are the weights of 16 persons, who enter the elevator.

a. Compute $P(X_1 \leq 75)$, the probability that one person has a weight of at most 75 kg.
b. Compute the probability that all 16 persons have a weight of at most 75 kg.
(*Use 75% for the probability in a. if you did not solve a.*)
c. Compute the probability that the elevator will be overloaded if 16 (arbitrary) persons enter.

<u>Part 2:</u>

4. A tablet manufacturer developed a new design: not only the appearance but also the functionalities were modernized. As a part of the development process a users survey was conducted. At random 48 users of the old design tablet (produced by the same manufacturer) were chosen and they were invited to test the prototype of the new design (for free). At the end of the testing period the participants had to fill in a questionnaire.
This exercise and exercise 5 deal with the results of the survey.
One of the questions to be answered by the users was: "Do you prefer the new design?"
28 out of the 48 answered "Yes".
We will call p the proportion of preference for the new design in the population of all tablet users

(produced by this manufacturer).
X is the number of users preferring the new design in a random sample of 48 users.

a. Compute $E(X)$ and $var(X)$ if $p = 0.5$
b. Compute a 90% Confidence Interval for the proportion p, using the result of the sample.
c. Conduct an appropriate test at 5% significance level to show that the majority of the tablet users prefer the new design. Use the 5 steps of the test procedure (on the formula page) and draw your conclusion, using the **p-value**.

5. (Continuation of the tablet survey as introduced in exercise 4).
In the questionnaire the users were asked to grade (using grades 1, 2, ...,10) many aspects of the new design, such as "ease of use", design, transparency of the manual, price-quality ratio, etc.
Per user the final grade was computed. These final grades can be assumed to be approximately normal.
The researchers were interested in the differences in mean final grades of men and women:

	Sample size	Sample mean	Sample standard deviation
Women	25	6.85	1.38
Men	23	6.10	0.92

a. Compute a 99% confidence interval for the expected final grade of **female** users.
Give the proper interpretation of the computed interval.
b. Can we conclude from these observations that the (mean) appreciation of the new design by men and women is different?
The manufacturer told the researchers to use a significance level $\alpha = 10\%$.

6. 1000 rats were used to evaluate the preventive effect of a (homeopathic) medicine on the occurrence of a particular disease. Half of the rats were treated with the medicine and the other half was not. Both groups (untreated and treated rats) were observed during a year, under equal circumstances. After a year 140 of the 500 untreated rats got the disease and so did 100 of the 500 treated rats.

a. Compute a 95% Confidence Interval for the difference in disease proportions of the untreated and treated rats.
b. Use the result of part a. to decide whether or not the proportions are (statistically) different.
(If you did not find a solution in a. use the interval (0.02, 0.14))

Grades:

Part 1										Part 2							
1			2				3			4			5		6		Total
a	b	c	a	b	c	d	a	b	c	a	b	c	a	b	a	b	
3	2	2	2	2	4	3	2	2	3	2	3	5	4	5	4	2	50
25										25							

Final grade = 1 + 9×score / 50
+ bonus
(provided that you finished your SPSS assignment.)

Formulas Introduction to Statistics and Probability for CreaTe (191567030)

Probability: Bayes: $P(S_i \mid A) = \frac{P(S_i \cap A)}{P(A)} = \frac{P(A \mid S_i)P(S_i)}{P(A \mid S_1)P(S_1) + \dots + P(A \mid S_k)P(S_k)}$

Probability distributions: $E(X) = \sum x p(x)$ and $var(X) = \sum (x - \mu)^2 p(x)$

Name/parameters	Probability (density) function	E(X)	Var(X)
Binomial (n, p)	$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$	np	np(1-p)
Hypergeometric (N, r, n)	$P(X = x) = \binom{r}{x}\binom{N-r}{n-x} / \binom{N}{n}$	$n\frac{r}{N}$	$n\frac{r}{N}\left(1-\frac{r}{N}\right)\frac{N-n}{N-1}$
Poisson (λ)	$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$	λ	λ
Uniform (c, d)	$f(x) = 1/(d-c)$, $c \le x \le d$	(c+d)/2	$(d-c)^2 / 12$
Normal (μ, σ^2)	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	μ	σ^2

Statistics: $s^2 = \frac{1}{n-1}\sum (x_i - \overline{x})^2$

Testing procedure:

1. Define the appropriate **parameters** (statistical assumptions)
2. Formulate the null and alternative **hypotheses**.
3. Give the formula of **test statistic** and compute its observed value.
4. Determine whether the test is left-, right- or two-sided.
   Compute the **Rejection Region or the p-value** (use $\alpha = 0.05$, if α is not given)
5. **Conclusion**: reject H_0 or fail to reject H_0 and give the proper interpretation.

Qualitative variables:

	One sample	Two independent samples
Test	$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$, where $\hat{p} = \frac{X_1 + X_2}{n_1 + n_2}$
Confidence interval	$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$
Sample size	$n \ge p(1-p)\left(\frac{z_{\alpha/2}}{SE}\right)^2$	

Quantitative variables

One sample

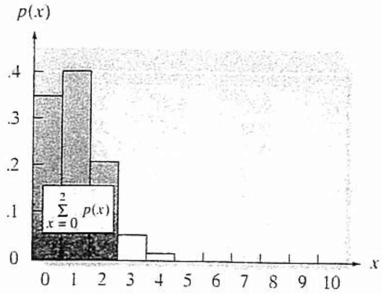
	One large sample (z-procedure) σ given: replace s by σ	One small sample (t-procedure)
Test	$z = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$	$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}}$, $df = n - 1$
Confidence interval	$\bar{x} \pm z_{\alpha/2}\frac{s}{\sqrt{n}}$	$\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}}$
Sample size	$n \geq \left(\frac{z_{\alpha/2} \times \sigma}{SE}\right)^2$	

Two pair wise dependent samples: one sample procedure for the differences

Two Independent Samples

	Two large independent samples	Two small independent samples
Test	$z = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}}$	$t = \frac{\bar{x}_1 - \bar{x}_2 - D_0}{\sqrt{s_p^2(\frac{1}{n_1}+\frac{1}{n_2})}}$, $s_p^2 = \frac{(n_1-1)s_1^2+(n_2-1)s_2^2}{n_1+n_2-2}$ and $df = n_1 + n_2 - 2$,
Confidence interval	$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}$	$\bar{x}_1 - \bar{x}_2 \pm t_{\alpha/2}\sqrt{s_p^2(\frac{1}{n_1}+\frac{1}{n_2})}$

Table II Binomial Probabilities



Tabulated values are $\sum_{x=0}^{k} p(x)$. *(Computations are rounded at the third decimal place.)*

a. $n = 5$

k \ p	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
0	.951	.774	.590	.328	.168	.078	.031	.010	.002	.000	.000	.000	.000
1	.999	.977	.919	.737	.528	.337	.188	.087	.031	.007	.000	.000	.000
2	1.000	.999	.991	.942	.837	.683	.500	.317	.163	.058	.009	.001	.000
3	1.000	1.000	1.000	.993	.969	.913	.812	.663	.472	.263	.081	.023	.001
4	1.000	1.000	1.000	1.000	.998	.990	.969	.922	.832	.672	.410	.226	.049

b. $n = 6$

k \ p	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
0	.941	.735	.531	.262	.118	.047	.016	.004	.001	.000	.000	.000	.000
1	.999	.967	.886	.655	.420	.233	.109	.041	.011	.002	.000	.000	.000
2	1.000	.998	.984	.901	.744	.544	.344	.179	.070	.017	.001	.000	.000
3	1.000	1.000	.999	.983	.930	.821	.656	.456	.256	.099	.016	.002	.000
4	1.000	1.000	1.000	.998	.989	.959	.891	.767	.580	.345	.114	.033	.001
5	1.000	1.000	1.000	1.000	.999	.996	.984	.953	.882	.738	.469	.265	.059

c. $n = 7$

k \ p	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
0	.932	.698	.478	.210	.082	.028	.008	.002	.000	.000	.000	.000	.000
1	.998	.956	.850	.577	.329	.159	.063	.019	.004	.000	.000	.000	.000
2	1.000	.996	.974	.852	.647	.420	.227	.096	.029	.005	.000	.000	.000
3	1.000	1.000	.997	.967	.874	.710	.500	.290	.126	.033	.003	.000	.000
4	1.000	1.000	1.000	.995	.971	.904	.773	.580	.353	.148	.026	.004	.000
5	1.000	1.000	1.000	1.000	.996	.981	.937	.841	.671	.423	.150	.044	.002
6	1.000	1.000	1.000	1.000	1.000	.998	.992	.972	.918	.790	.522	.302	.068

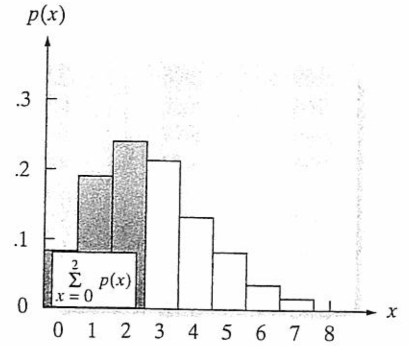
e. $n = 9$

k \ p	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
0	.914	.630	.387	.134	.040	.010	.002	.000	.000	.000	.000	.000	.000
1	.997	.929	.775	.436	.196	.071	.020	.004	.000	.000	.000	.000	.000
2	1.000	.992	.947	.738	.463	.232	.090	.025	.004	.000	.000	.000	.000
3	1.000	.999	.992	.914	.730	.483	.254	.099	.025	.003	.000	.000	.000
4	1.000	1.000	.999	.980	.901	.733	.500	.267	.099	.020	.001	.000	.000
5	1.000	1.000	1.000	.997	.975	.901	.746	.517	.270	.086	.008	.001	.000
6	1.000	1.000	1.000	1.000	.996	.975	.910	.768	.537	.262	.053	.008	.000
7	1.000	1.000	1.000	1.000	1.000	.996	.980	.929	.804	.564	.225	.071	.003
8	1.000	1.000	1.000	1.000	1.000	1.000	.998	.990	.960	.866	.613	.370	.086

f. $n = 10$

k \ p	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
0	.904	.599	.349	.107	.028	.006	.001	.000	.000	.000	.000	.000	.000
1	.996	.914	.736	.376	.149	.046	.011	.002	.000	.000	.000	.000	.000
2	1.000	.988	.930	.678	.383	.167	.055	.012	.002	.000	.000	.000	.000
3	1.000	.999	.987	.879	.650	.382	.172	.055	.011	.001	.000	.000	.000
4	1.000	1.000	.998	.967	.850	.633	.377	.166	.047	.006	.000	.000	.000
5	1.000	1.000	1.000	.994	.953	.834	.623	.367	.150	.033	.002	.000	.000
6	1.000	1.000	1.000	.999	.989	.945	.828	.618	.350	.121	.013	.001	.000
7	1.000	1.000	1.000	1.000	.998	.988	.945	.833	.617	.322	.070	.012	.000

Table III Poisson Probabilities

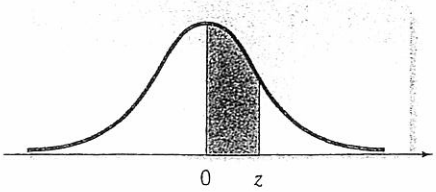


Tabulated values are $\sum_{x=0}^{k} p(x)$. *(Computations are rounded at the third decimal place.)*

λ \ k	0	1	2	3	4	5	6	7	8	9
.02	.980	1.000								
.04	.961	.999	1.000							
.06	.942	.998	1.000							
.08	.923	.997	1.000							
.10	.905	.995	1.000							
.15	.861	.990	.999	1.000						
.20	.819	.982	.999	1.000						
.25	.779	.974	.998	1.000						
.30	.741	.963	.996	1.000						
.35	.705	.951	.994	1.000						
.40	.670	.938	.992	.999	1.000					
.45	.638	.925	.989	.999	1.000					
.50	.607	.910	.986	.998	1.000					
.55	.577	.894	.982	.998	1.000					
.60	.549	.878	.977	.997	1.000					
.65	.522	.861	.972	.996	.999	1.000				
.70	.497	.844	.966	.994	.999	1.000				
.75	.472	.827	.959	.993	.999	1.000				
.80	.449	.809	.953	.991	.999	1.000				
.85	.427	.791	.945	.989	.998	1.000				
.90	.407	.772	.937	.987	.998	1.000				
.95	.387	.754	.929	.981	.997	1.000				
1.00	.368	.736	.920	.981	.996	.999	1.000			
1.1	.333	.699	.900	.974	.995	.999	1.000			
1.2	.301	.663	.879	.966	.992	.998	1.000			
1.3	.273	.627	.857	.957	.989	.998	1.000			
1.4	.247	.592	.833	.946	.986	.997	.999	1.000		
1.5	.223	.558	.809	.934	.981	.996	.999	1.000		
1.6	.202	.525	.783	.921	.976	.994	.999	1.000		
1.7	.183	.493	.757	.907	.970	.992	.998	1.000		
1.8	.165	.463	.731	.891	.964	.990	.997	.999	1.000	
1.9	.150	.434	.704	.875	.956	.987	.997	.999	1.000	
2.0	.135	.406	.677	.857	.947	.983	.995	.999	1.000	
2.2	.111	.355	.623	.819	.928	.975	.993	.998	1.000	
2.4	.091	.308	.570	.779	.904	.964	.988	.997	.999	1.000
2.6	.074	.267	.518	.736	.877	.951	.983	.995	.999	1.000
2.8	.061	.231	.469	.692	.848	.935	.976	.992	.998	.999
3.0	.050	.199	.423	.647	.815	.916	.966	.988	.996	.999
3.2	.041	.171	.380	.603	.781	.895	.955	.983	.994	.998
3.4	.033	.147	.340	.558	.744	.871	.942	.977	.992	.997

(continued)

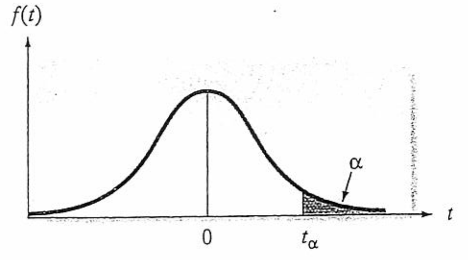
TABLE IV Normal Curve Areas



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Source: Abridged from Table I of A. Hald, *Statistical Tables and Formulas* (New York: Wiley), 1952. Reproduced by permission of A. Hald.

TABLE VI Critical Values of t



Degrees of Freedom	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$	$t_{.0005}$
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

Source: This table is reproduced with the kind permission of the Trustees of Biometrika from E. S. Pearson and H. O. Hartley (eds.), *The Biometrika Tables for Statisticians,* Vol. 1, 3d ed., Biometrika, 1966.