Department of Mathematics	Final exam – Statistics X_400004
VU University Amsterdam	19 December 2017

The use of a calculator is allowed. After correction exams can be inspected through the educational office. The exam consists of 3 exercises. The rating of each exercise can be found below with the exercise. Grade = total/5.

Please write your name, your student number and the name of your teaching assistant at the top of the front page.

Success!

Exercise 1 [16 points]

The manufacturer of a drug that lowers blood pressure has changed the composition of the drug with the purpose to improve its efficacy. The new and the old versions of the drug are now compared for efficacy. The old version has been prescribed to a group of 16 patients with hypertension (high blood pressure), and the new version to a group of 10 other patients with hypertension. The blood pressures $X_1, \ldots, X_{16}$ and $Y_1, \ldots, Y_{10}$ of the patients in the two groups have been measured after they have taken the respective versions of the drug. The observed measurements have the following summaries: $\overline{X} = 163$ and $\overline{Y} = 158$, $S_X = 7.8$ and $S_Y = 9.0$. Do these data give a reason to conclude that the new version of the drug is better than the old version? Carry out a suitable standard test at significance level 0.05. Report

(a) [2 points] the statistical model (all assumptions you make on the data and the parameters),

(b) [2 points] the null and alternative hypotheses,

(c) [4 points] the test statistic and its distribution when the new and the old versions of the drug have the same efficacy,
Hint: If $X_1, \ldots, X_m \sim N(\mu_A, \sigma_A^2)$ and $Y_1, \ldots, Y_n \sim N(\mu_B, \sigma_B^2)$, all independent, then $\dfrac{(\overline{X} - \overline{Y}) - (\mu_A - \mu_B)}{\sqrt{\sigma_A^2/m + \sigma_B^2/n}} \sim N(0,1)$ and $\dfrac{(m-1)S_X^2}{\sigma_A^2} + \dfrac{(n-1)S_Y^2}{\sigma_B^2}$ has a chi-square distribution.

(d) [2 points] the critical region,

(e) [4 points] the conclusion.

(f) [2 points] Are the data in this experiment paired or unpaired? Would you design the experiment differently in this regard and why (not)?

Exercise 2 [15 points]

Let $X_1, \ldots, X_8$ be a sample from the normal $N(0, \sigma^2)$ distribution, and $Y_1, \ldots, Y_8$ a sample from the normal $N(0, 9\sigma^2)$ distribution, where the parameter $\sigma^2 > 0$ is unknown. The two samples are independent and the random variables within each sample are independent.

(a) [4 points] Give the definition of a pivot. Construct a non-negative pivot out of the sample $X_1, \ldots, X_8$ and σ. What is the distribution of this pivot?

(b) [3 points] Construct a non-negative pivot out of both samples and σ. What is the distribution of this pivot?

(c) [3 points] Based on either of the pivots from *(a)* or *(b)* (only one of them), construct a confidence interval for σ^2 of confidence level 0.9.

(d) [2 points] Describe how you can test $H_0\colon \sigma^2 = \sigma_0^2$ VS $H_1\colon \sigma^2 \neq \sigma_0^2$ at significance level 0.1 using the confidence interval from *(c)*.

(e) [3 points] It has been observed that $\overline{X^2} = 1.3$, $\overline{Y^2} = 9.9$, $\overline{X} = 1.1$, and $\overline{Y} = 3.1$. Using *(d)*, test at significance level 0.1 whether σ^2 deviates from 1.6^2.
Hint: You may need to take into account only some of the summaries $\overline{X^2} = 1.3$, $\overline{Y^2} = 9.9$, $\overline{X} = 1.1$, and $\overline{Y} = 3.1$, not all of them. **Turn the page.**

Exercise 3 [19 points]

Let $X_1, \ldots, X_n$ be independent random variables from the Poisson distribution with unknown parameter $\lambda > 0$. Recall that the p.m.f. is given by $p_\lambda(x) = e^{-\lambda}\dfrac{\lambda^x}{x!}$, $x = 0, 1, 2, \ldots$, and $\mathbb{E}X_1 = \mathrm{Var}X_1 = \lambda$.

(a) [8 points] The maximum likelihood estimator for λ is $\widehat{\lambda}_{\mathrm{ML}} = \overline{X}$, and i_λ denotes the Fisher information (see Appendix 1). What is approximately the distribution of $\sqrt{n i_\lambda}(\overline{X} - \lambda)$ for large n? Based on this approximation, construct an approximate confidence interval for λ of confidence level $1 - \alpha$. What estimator for i_λ do you use?

(b) [3 points] Does the Cramer-Rao lower bound (see Appendix 1) imply that $\overline{X}$ is a UMVU estimator for λ?

(c) [4 points] Show that $\overline{X}$ is a sufficient and complete statistic (see Appendix 1).

(d) [4 points] Find a UMVU estimator for λ^2.

Appendix 1: Some relevant facts

Fisher information Let $X_1, \dots, X_n$ be i.i.d. with marginal p.d.f./p.m.f. $p_\theta(x)$ and $\ell_\theta(x) := \log p_\theta(x)$, $\dot\ell_\theta(x) := \frac{\partial}{\partial\theta}\ell_\theta(x)$, $\ddot\ell_\theta(x) := \frac{\partial^2}{\partial\theta^2}\ell_\theta(x)$. Then the Fisher information is given by

$$i_\theta := \mathbb{V}\ \dot\ell_\theta(X_1) = -\mathbb{E}\ \ddot\ell_\theta(X_1).$$

Cramer-Rao lower bound Let $X_1, \dots, X_n$ be i.i.d. random variables from a distribution parametrized by θ. Under certain conditions, every unbiased estimator $\widehat{g(\theta)}$ for $g(\theta)$ satisfies

$$\mathbb{V}\ \widehat{g(\theta)} \geq \frac{(g'(\theta))^2}{n i_\theta}.$$

Exponential family Statistical models parametrized by θ form an exponential family if the likelihood is of the form

$$L(\theta) = c(\theta) \cdot e^{\sum_{j=1}^k Q_j(\theta)\cdot V_j(X_1,\dots,X_n)} \cdot h(X_1, \dots, X_n).$$

.

Appendix 2: Table t-distribution

df	0.6	0.7	0.75	0.8	0.85	0.9	0.925	0.95	0.975	0.98	0.99	0.999
9	0.26	0.54	0.7	0.88	1.1	1.38	1.57	1.83	2.26	2.4	2.82	4.3
10	0.26	0.54	0.7	0.88	1.09	1.37	1.56	1.81	2.23	2.36	2.76	4.14
11	0.26	0.54	0.7	0.88	1.09	1.36	1.55	1.8	2.2	2.33	2.72	4.02
12	0.26	0.54	0.7	0.87	1.08	1.36	1.54	1.78	2.18	2.3	2.68	3.93
13	0.26	0.54	0.69	0.87	1.08	1.35	1.53	1.77	2.16	2.28	2.65	3.85
14	0.26	0.54	0.69	0.87	1.08	1.35	1.52	1.76	2.14	2.26	2.62	3.79
15	0.26	0.54	0.69	0.87	1.07	1.34	1.52	1.75	2.13	2.25	2.6	3.73
16	0.26	0.54	0.69	0.86	1.07	1.34	1.51	1.75	2.12	2.24	2.58	3.69
17	0.26	0.53	0.69	0.86	1.07	1.33	1.51	1.74	2.11	2.22	2.57	3.65
18	0.26	0.53	0.69	0.86	1.07	1.33	1.5	1.73	2.1	2.21	2.55	3.61
19	0.26	0.53	0.69	0.86	1.07	1.33	1.5	1.73	2.09	2.2	2.54	3.58
20	0.26	0.53	0.69	0.86	1.06	1.33	1.5	1.72	2.09	2.2	2.53	3.55
21	0.26	0.53	0.69	0.86	1.06	1.32	1.49	1.72	2.08	2.19	2.52	3.53
22	0.26	0.53	0.69	0.86	1.06	1.32	1.49	1.72	2.07	2.18	2.51	3.5
23	0.26	0.53	0.69	0.86	1.06	1.32	1.49	1.71	2.07	2.18	2.5	3.48
24	0.26	0.53	0.68	0.86	1.06	1.32	1.49	1.71	2.06	2.17	2.49	3.47
25	0.26	0.53	0.68	0.86	1.06	1.32	1.49	1.71	2.06	2.17	2.49	3.45
26	0.26	0.53	0.68	0.86	1.06	1.31	1.48	1.71	2.06	2.16	2.48	3.43
27	0.26	0.53	0.68	0.86	1.06	1.31	1.48	1.7	2.05	2.16	2.47	3.42
28	0.26	0.53	0.68	0.85	1.06	1.31	1.48	1.7	2.05	2.15	2.47	3.41

Table 1: Quantiles (columns) of the t-distribution with 9 to 28 degrees of freedom (rows).

Appendix 3: Table Chi-square distribution

df	0.001	0.005	0.01	0.025	0.05	0.1	0.125	0.2	0.25	0.333	0.5
1	0.000	0.000	0.000	0.001	0.004	0.016	0.025	0.064	0.102	0.186	0.455
2	0.002	0.010	0.020	0.051	0.103	0.211	0.267	0.446	0.575	0.811	1.386
3	0.024	0.072	0.115	0.216	0.352	0.584	0.692	1.005	1.213	1.568	2.366
4	0.091	0.207	0.297	0.484	0.711	1.064	1.219	1.649	1.923	2.378	3.357
5	0.210	0.412	0.554	0.831	1.145	1.610	1.808	2.343	2.675	3.216	4.351
6	0.381	0.676	0.872	1.237	1.635	2.204	2.441	3.070	3.455	4.074	5.348
7	0.598	0.989	1.239	1.690	2.167	2.833	3.106	3.822	4.255	4.945	6.346
8	0.857	1.344	1.646	2.180	2.733	3.490	3.797	4.594	5.071	5.826	7.344
9	1.152	1.735	2.088	2.700	3.325	4.168	4.507	5.380	5.899	6.716	8.343
10	1.479	2.156	2.558	3.247	3.940	4.865	5.234	6.179	6.737	7.612	9.342
11	1.834	2.603	3.053	3.816	4.575	5.578	5.975	6.989	7.584	8.514	10.341
12	2.214	3.074	3.571	4.404	5.226	6.304	6.729	7.807	8.438	9.420	11.340
13	2.617	3.565	4.107	5.009	5.892	7.042	7.493	8.634	9.299	10.331	12.340
14	3.041	4.075	4.660	5.629	6.571	7.790	8.266	9.467	10.165	11.245	13.339
15	3.483	4.601	5.229	6.262	7.261	8.547	9.048	10.307	11.037	12.163	14.339
16	3.942	5.142	5.812	6.908	7.962	9.312	9.837	11.152	11.912	13.083	15.338
17	4.416	5.697	6.408	7.564	8.672	10.085	10.633	12.002	12.792	14.006	16.338
18	4.905	6.265	7.015	8.231	9.390	10.865	11.435	12.857	13.675	14.931	17.338
19	5.407	6.844	7.633	8.907	10.117	11.651	12.242	13.716	14.562	15.859	18.338
20	5.921	7.434	8.260	9.591	10.851	12.443	13.055	14.578	15.452	16.788	19.337

df	0.6	0.667	0.75	0.8	0.87	0.9	0.95	0.975	0.99	0.995	0.999
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588
11	11.530	12.414	13.701	14.631	16.457	17.275	19.675	21.920	24.725	26.757	31.264
12	12.584	13.506	14.845	15.812	17.703	18.549	21.026	23.337	26.217	28.300	32.910
13	13.636	14.595	15.984	16.985	18.939	19.812	22.362	24.736	27.688	29.819	34.528
14	14.685	15.680	17.117	18.151	20.166	21.064	23.685	26.119	29.141	31.319	36.123
15	15.733	16.761	18.245	19.311	21.384	22.307	24.996	27.488	30.578	32.801	37.697
16	16.780	17.840	19.369	20.465	22.595	23.542	26.296	28.845	32.000	34.267	39.252
17	17.824	18.917	20.489	21.615	23.799	24.769	27.587	30.191	33.409	35.718	40.790
18	18.868	19.991	21.605	22.760	24.997	25.989	28.869	31.526	34.805	37.156	42.312
19	19.910	21.063	22.718	23.900	26.189	27.204	30.144	32.852	36.191	38.582	43.820
20	20.951	22.133	23.828	25.038	27.376	28.412	31.410	34.170	37.566	39.997	45.315

Table 2: Quantiles (columns) of the chi-square distribution with 1 to 20 degrees of freedom (rows).