Exercise Set 10

Goals

1) Practice ANOVA
2) Perform a first linear regression.
3) Practice data tests for the exam.

1 A new post-it chemical [basic]

Post-its are magical things. The glue used in them sticks quite well to most surfaces, yet they can be easily removed without residue. Even after a long time, the glue does not chemically react with the surfaces. In fact, this glue is a prime example of a serendipitous discovery: While the researchers initially looked for an extremely strong glue, they found a particularly removable one.

Over the years, the company continues to optimize the properties of the glue. Post-it has started an open competition to researchers, to challenge the quality of their glue. Two teams of scientists have made it into the finals. Each team prepares 40 post-its with their glues, and the sticking coefficients are measured by an independent laboratory. The group means and standard deviations are listed below (higher number means it sticks better). Is there statistical evidence on the $\alpha = 0.01$ level that the glues are not all the same?

The provided table does not include this level of significance, but you can use Python or https://fvaluecalculator.com/ to find the critical F-value.

	mean	sd
EPFL team	21.4	4.5
ETH team	16.9	5.5
post-it original	19.1	5.8

2 Is size a good predictor of weight? (a linear regression) [normal]

The weight and height of six female students are presented in the table below:

Height (cm)	170	167	171	181	158	166
Weight (kg)	71	58	56	85	45	53

a) Determine and plot the regression line for this table.

b) Establish the ANOVA table for a linear regression with one slope and run the F-test at a level of significance $\alpha = 0.05$.

c) Compute the "goodness of fit" coefficient R^2. What can you conclude from it?

3 Exam question from 2020

On the following page you will find an exam question from 2020. The exam back then lasted two hours.

The questions on the next page accounted for 17 out of a total of 59 points - with proportional time allocation, this would correspond to about 35 minutes.

In the exam tables for the z-test, t-test, χ^2 test and F-test were provided (see the StatTables pdf on the moodle from last week) .

Of course now that you do not have your condensed notes yet, you should use your course notes or favorite textbook, and can expect to take longer because of that.

4 A new milkshake recipe

The industrial production of milk drinks is surprisingly complex. The main reason is the difficulty to stabilize milk which is essentially a mixture of water, proteins and fat. In your PhD thesis, you develop a new food-safe stabilizer for milkshakes based on seaweed extract. The tests are promising, and you want to study how customers compare the consistency of your milkshake to the two leading brands, Emmi and Mueller. You invite 6 people into your laboratory, and give them a sample of each milkshake. Then they should grade the consistency on a scale between 10 (best) to 1 (worst). Here is the outcome of your test:

Your product (1)	6.2	7.0	4.0	10	9.4	4.0
Emmi (2)	3.9	8.9	5.8	9.0	3.1	4.8
Mueller (3)	4.6	1.4	7.1	2.6	2.2	1.1

Table 4: Taste of milkshakes graded by 6 randomly selected volunteers. Grades between 1 (worst) and 10 (best).

In a first attempt to analyze this data, you decide to do pairwise two-sided T-tests to see if the differences between the datasets are statistically significant. We denote the taste grade of the j-th tester in the group i by $X_{i,j}$. For example, the third tester in the Emmi (2) group rated that milkshake $X_{2,3} = 5.8$.

Your statistical software reports the following computations on your dataset:
$\sum_{j=1}^{6} X_{1,j} = 40.6, \sum_{j=1}^{6} X_{2,j} = 35.5, \sum_{j=1}^{6} X_{3,j} = 19,$
$\sum_{j=1}^{6}(X_{1,j})^2 = 307.8, \sum_{j=1}^{6}(X_{2,j})^2 = 241.71, \sum_{j=1}^{6}(X_{3,j})^2 = 86.34$

a) [1 pt] For your milkshake (group 1), compute the group mean $\bar{X}_1$ and the standard deviation s_1. For the other groups, the same calculation gives $\bar{X}_2 = 5.92$, $s_2 = 2.52$, $\bar{X}_3 = 3.17$, and $s_3 = 2.29$.

b) [1 pt] We want to check if these milkshakes are received significantly different by the customers. Formulate the null hypothesis H_0 and alternative hypothesis H_1 for a two-sided T-test between your milkshake (group 1) and the Mueller shake (group 3).

c) [2 pt] Next you plan to do two-sided T-tests between all pairs of groups. $T_{i,j}$ is the Welch statistic comparing groups i and j. Unfortunately your statistics software crashed and you lost some computations. Complete the following table of $T_{i,j}$ by manually computing $T_{1,3}$.

$T_{i,j}$	your shake (1)	Emmi (2)	Mueller (3)
your shake (1)	-	0.58	?
Emmi (2)	-	-	1.98
Mueller (3)	-	-	-

d) [5 pt] Perform two-sided Welch tests between all 3 pairs using the T-statistics computed in c). State the degrees of freedom used for each test, and the critical regions. Between which groups do you find statistically relevant differences at a level of significance $\alpha = 0.05$? Are these taste differences between the milkshakes statistically significant?

A different approach to investigate the differences between multiple groups would be a one-factor, three-level ANOVA analysis and a F-test. Let us reanalyze the same dataset in Table 4.

e) [1 pt] In the one factor model, your data is described by $X_{i,j} = \mu + \alpha_i + \epsilon_{i,j}$. Give the values for the α_i and μ that minimize the sum of least squares. *You do not need to derive them from the principle of least squares.*

f) [5 pt] Compute the one-factor ANOVA table for this model.

g) [2 pt] Based on your ANOVA analysis, perform an F-test. Can you state a statistically significant ($\alpha = 0.05$) difference between the groups?

Tableau de la fonction de répartition normale. La fonction $\Phi(z)$ est égale à la proportion des réalisations inférieures ou égales à z d'une varable aléatoire normale centrée et réduite.

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0,00	0,500	0,72	0,764	1,44	0,9251	2,16	0,9846	2,88	0,99801	3,80	0,9999277
0,02	0,508	0,74	0,770	1,46	0,9279	2,18	0,9854	2,90	0,99813	3,84	0,9999385
0,04	0,516	0,76	0,776	1,48	0,9306	2,20	0,9861	2,92	0,99825	3,88	0,9999478
0,06	0,524	0,78	0,782	1,50	0,9332	2,22	0,9868	2,94	0,99836	3,92	0,9999557
0,08	0,532	0,80	0,788	1,52	0,9357	2,24	0,9875	2,96	0,99846	3,96	0,9999625
0,10	0,540	0,82	0,794	1,54	0,9382	2,26	0,9881	2,98	0,99856	4,00	0,9999683
0,12	0,548	0,84	0,800	1,56	0,9406	2,28	0,9887	3,00	0,99865	4,04	0,9999733
0,14	0,556	0,86	0,805	1,58	0,9429	2,30	0,9893	3,02	0,99874	4,08	0,9999775
0,16	0,564	0,88	0,811	1,60	0,9452	2,32	0,9898	3,04	0,99882	4,12	0,9999811
0,18	0,571	0,90	0,816	1,62	0,9474	2,34	0,9904	3,06	0,99889	4,16	0,9999841
0,20	0,579	0,92	0,821	1,64	0,9495	2,36	0,9909	3,08	0,99996	4,20	0,9999867
0,22	0,587	0,94	0,826	1,66	0,9515	2,38	0,9913	3,10	0,99903	4,24	0,9999888
0,24	0,595	0,96	0,831	1,68	0,9535	2,40	0,9918	3,12	0,99910	4,28	0,9999907
0,26	0,603	0,98	0,836	1,70	0,9554	2,42	0,9922	3,14	0,99916	4,32	0,9999922
0,28	0,610	1,00	0,841	1,72	0,9573	2,44	0,9927	3,16	0,99921	4,36	0,9999935
0,30	0,618	1,02	0,846	1,74	0,9591	2,46	0,9931	3,18	0,99926	4,40	0,9999946
0,32	0,626	1,04	0,851	1,76	0,9608	2,48	0,9934	3,20	0,99931	4,44	0,9999955
0,34	0,633	1,06	0,855	1,78	0,9625	2,50	0,9938	3,22	0,99936	4,48	0,9999963
0,36	0,641	1,08	0,860	1,80	0,9641	2,52	0,9941	3,24	0,99940	4,52	0,9999969
0,38	0,648	1,10	0,864	1,82	0,9656	2,54	0,9945	3,26	0,99944	4,56	0,9999974
0,40	0,655	1,12	0,869	1,84	0,9671	2,56	0,9948	3,28	0,99948	4,60	0,9999979
0,42	0,663	1,14	0,873	1,86	0,9686	2,58	0,9951	3,30	0,99952	4,64	0,9999983
0,44	0,670	1,16	0,877	1,88	0,9799	2,60	0,9953	3,32	0,99955	4,68	0,9999986
0,46	0,677	1,18	0,881	1,90	0,9713	2,62	0,9956	3,34	0,99958	4,72	0,9999988
0,48	0,684	1,20	0,885	1,92	0,9726	2,64	0,9959	3,36	0,99961	4,76	0,9999990
0,50	0,691	1,22	0,889	1,94	0,9738	2,66	0,9961	3,38	0,99964	4,80	0,9999992
0,52	0,698	1,24	0,893	1,96	0,9750	2,68	0,9963	3,40	0,99966	4,84	0,9999994
0,54	0,705	1,26	0,896	1,98	0,9761	2,70	0,9965	3,42	0,99969	4,88	0,9999995
0,56	0,712	1,28	0,900	2,00	0,9772	2,72	0,9967	3,44	0,99971	4,92	0,9999996
0,58	0,719	1,30	0,903	2,02	0,9783	2,74	0,9969	3,46	0,99973	4,96	0,9999996
0,60	0,726	1,32	0,907	2,04	0,9793	2,76	0,9971	3,48	0,99975	5,00	0,9999997
0,62	0,732	1,34	0,910	2,06	0,9803	2,78	0,9973	3,50	0,99977	5,04	0,9999998
0,64	0,739	1,36	0,913	2,08	0,9812	2,80	0,9974	3,52	0,99978	5,08	0,9999998
0,66	0,745	1,38	0,916	2,10	0,9821	2,82	0,9976	3,54	0,99980	5,12	0,9999998
0,68	0,752	1,40	0,919	2,12	0,9830	2,84	0,9977	3,56	0,99981	5,16	0,9999999
0,70	0,758	1,42	0,922	2,14	0,9838	2,86	0,9979	3,58	0,99983	5,20	0,9999999

Quantiles de la loi t_ν de Student

ν	$qt_\nu(95\%)$	$qt_\nu(97{,}5\%)$	$qt_\nu(99\%)$	ν	$qt_\nu(95\%)$	$qt_\nu(97{,}5\%)$	$qt_\nu(99\%)$
1	6,314	12,71	31,82	21	1,721	2,080	2,518
2	2,920	4,303	6,965	22	1,717	2,074	2,508
3	2,353	3,182	4,541	23	1,714	2,069	2,500
4	2,132	2,776	3,747	24	1,711	2,064	2,492
5	2,015	2,571	3,365	25	1,708	2,060	2,485
6	1,943	2,447	3,143	26	1,706	2,056	2,479
7	1,895	2,365	2,998	27	1,703	2,052	2,473
8	1,860	2,306	2,896	28	1,701	2,048	2,467
9	1,833	2,262	2,821	29	1,699	2,045	2,462
10	1,812	2,228	2,764	30	1,697	2,042	2,457
11	1,796	2,201	2,718	32	1,694	2,037	2,449
12	1,782	2,179	2,681	34	1,691	2,032	2,441
13	1,771	2,160	2,650	36	1,688	2,028	2,434
14	1,761	2,145	2,624	38	1,686	2,024	2,429
15	1,753	2,131	2,602	40	1,684	2,021	2,423
16	1,746	2,120	2,583	50	1,676	2,009	2,403
17	1,740	2,110	2,567	60	1,671	2,000	2,390
18	1,734	2,101	2,552	120	1,658	1,980	2,358
19	1,729	2,093	2,539	∞	1,645	1,960	2,326
20	1,725	2,086	2,528				

Table 1 – Quantiles des distributions t_ν. Pour ν suffisamment grand, on peut substituer aux quantiles de la loi t_ν les quantiles de la loi normale.

Quantiles de la loi khi–deux

ν	$q\chi^2_\nu(1\%)$	$q\chi^2_\nu(2{,}5\%)$	$q\chi^2_\nu(5\%)$	$q\chi^2_\nu(95\%)$	$q\chi^2_\nu(97{,}5\%)$	$q\chi^2_\nu(99\%)$
1	$0,0^3 1571$	$0,0^3 9821$	0,003932	3,841	5,024	6,635
2	0,02010	0,05064	0,1026	5,991	7,378	9,210
3	0,1148	0,2158	0,3518	7,815	9,348	11,34
4	0,2971	0,4844	0,7107	9,488	11,14	13,28
5	0,5543	0,8312	1,145	11,07	12,83	15,09
6	0,8721	1,237	1,635	12,59	14,45	16,81
7	1,239	1,690	2,167	14,07	16,01	18,48
8	1,646	2,180	2,733	15,51	17,53	20,09
9	2,088	2,700	3,325	16,92	19,02	21,67
10	2,558	3,247	3,940	18,31	20,48	23,21
11	3,053	3,816	4,575	19,68	21,92	24,72
12	3,571	4,404	5,226	21,03	23,34	26,22
13	4,107	5,009	5,892	22,36	24,74	27,69
14	4,660	5,629	6,571	23,68	26,12	29,14
15	5,229	6,262	7,261	25,00	27,49	30,58
16	5,812	6,908	7,962	26,30	28,85	32,00
17	6,408	7,564	8,672	27,59	30,19	33,41
18	7,015	8,231	9,390	28,87	31,53	34,81
19	7,633	8,907	10,12	30,14	32,85	36,19
20	8,260	9,591	10,85	31,41	34,17	37,57

TABLE 5 – Quantiles des distributions χ^2_ν. Le chiffres pour $\nu = 21, \ldots, 100$ se trouvent à la page suivante.

ν	$q\chi^2_\nu(1\%)$	$q\chi^2_\nu(2{,}5\%)$	$q\chi^2_\nu(5\%)$	$q\chi^2_\nu(95\%)$	$q\chi^2_\nu(97{,}5\%)$	$q\chi^2_\nu(99\%)$
21	8,897	10,28	11,59	32,67	35,48	38,93
22	9,542	10,98	12,34	33,92	36,78	40,29
23	10,20	11,69	13,09	35,17	38,08	41,64
24	10,86	12,40	13,85	36,42	39,36	42,98
25	11,52	13,12	14,61	37,65	40,65	44,31
26	12,20	13,84	15,38	38,89	41,92	45,64
27	12,88	14,57	16,15	40,11	43,19	46,96
28	13,56	15,31	16,93	41,34	44,46	48,28
29	14,26	16,05	17,71	42,56	45,72	49,59
30	14,95	16,79	18,49	43,77	46,98	50,89
32	16,36	18,29	20,07	46,19	49,48	53,49
34	17,79	19,81	21,66	48,60	51,97	56,06
36	19,23	21,34	23,27	51,00	54,44	58,62
38	20,69	22,88	24,88	53,38	56,90	61,16
40	22,16	24,43	26,51	55,76	59,34	63,69
50	29,71	32,36	34,76	67,50	71,42	76,15
60	37,48	40,48	43,19	79,08	83,30	88,38
70	45,44	48,76	51,74	90,53	95,02	100,4
80	53,54	57,15	60,39	101,9	106,6	112,3
90	61,75	65,65	69,13	113,1	118,1	124,1
100	70,06	74,22	77,93	124,3	129,6	135,8

TABLE 6 – Quantiles des distributions χ^2_ν, dont la figure A.8 montre quelques densités.
L'approximation normale est $q\chi^2_\nu(p) \approx \nu + \sqrt{2\nu}\, q_{\text{normal}}(p)$, valable pour ν suffisamment grand. Pour $\nu = 20$, par exemple, cette formule donne $20 + 6,32 \times 1,64 = 30,37$ comme valeur approximative du 95%-quantile d'une loi χ^2_{20}.
Une meilleure approximation est celle de Wilson-Hilferty, définie par
$q\chi^2_\nu(p) \approx \nu \left(1 - 2/(9\nu) + \sqrt{2/(9\nu)}\, q_{\text{normal}}(p)\right)^3$. Dans l'exemple précédent cette formule donne la valeur approximative de $20(0,99 + 0,11 \times 1,64)^3 = 31,36$ pour $q\chi^2_{20}(95\%)$.

Quantiles de la loi F_{ν_1,ν_2} de Fisher

	$\nu_1 = 1$	2	3	4	5	6	7	8	10	12	24	∞
$\nu_2 = 1$	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	241,9	243,9	249,1	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,40	19,41	19,45	19,50
3	10,13	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,786	8,745	8,639	8,526
4	7,709	6,944	6,591	6,388	6,256	6,163	6,094	6,041	5,964	5,912	5,774	5,628
5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,735	4,678	4,527	4,365
6	5,987	5,143	4,757	4,534	4,387	4,284	4,207	4,147	4,060	4,000	3,841	3,669
7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,637	3,575	3,410	3,230
8	5,318	4,459	4,066	3,838	3,687	3,581	3,500	3,438	3,347	3,284	3,115	3,928
9	5,117	4,256	3,863	3,633	3,482	3,374	3,293	3,230	3,137	3,073	2,900	2,707
10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	2,978	2,913	2,737	2,538
11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,854	2,788	2,609	2,404
12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,753	2,687	2,505	2,296
13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,671	2,604	2,420	2,206
14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,602	2,534	2,349	2,131
15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,544	2,475	2,288	2,066
16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,494	2,425	2,235	2,010
17	4,451	3,592	3,197	2,965	2,810	2,699	2,614	2,548	2,450	2,381	2,190	1,960
18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,412	2,342	2,150	1,917
19	4,381	3,522	3,127	2,895	2,740	2,628	2,544	2,477	2,378	2,308	2,114	1,878
20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,348	2,278	2,082	1,843
21	4,325	3,467	3,072	2,840	2,685	2,573	2,488	2,420	2,321	2,250	2,054	1,812
22	4,301	3,443	3,049	2,817	2,661	2,549	2,464	2,397	2,297	2,226	2,028	1,783
23	4,279	3,422	3,028	2,796	2,640	2,528	2,442	2,375	2,275	2,204	2,005	1,757
24	4,260	3,403	3,009	2,776	2,621	2,508	2,423	2,355	2,255	2,183	1,984	1,733
25	4,242	3,385	2,991	2,759	2,603	2,490	2,405	2,337	2,236	2,165	1,964	1,711
26	4,225	3,369	2,975	2,743	2,587	2,474	2,388	2,321	2,220	2,148	1,946	1,691
27	4,210	3,354	2,960	2,728	2,572	2,459	2,373	2,305	2,204	2,132	1,930	1,672
28	4,196	3,340	2,947	2,714	2,558	2,445	2,359	2,291	2,190	2,118	1,915	1,654
29	4,183	3,328	2,934	2,701	2,545	2,432	2,346	2,278	2,177	2,104	1,901	1,638
30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,165	2,092	1,887	1,622
32	4,149	3,295	2,901	2,668	2,512	2,399	2,313	2,244	2,142	2,070	1,864	1,594
34	4,130	3,276	2,883	2,650	2,494	2,380	2,294	2,225	2,123	2,050	1,843	1,569
36	4,113	3,259	2,866	2,634	2,477	2,364	2,277	2,209	2,106	2,033	1,824	1,547
38	4,098	3,245	2,852	2,619	2,463	2,349	2,262	2,194	2,091	2,017	1,808	1,527
40	4,085	3,232	2,839	2,606	2,449	2,336	2,249	2,180	2,077	2,003	1,793	1,509
60	4,001	3,150	2,758	2,525	2,368	2,254	2,167	2,097	1,993	1,917	1,700	1,389
120	3,920	3,072	2,680	2,447	2,290	2,175	2,087	2,016	1,910	1,834	1,608	1,254
∞	3,841	2,996	2,605	2,372	2,214	2,099	2,010	1,938	1,831	1,752	1,517	1,000

Table 5 – Les 95%-quantiles, $qF_{\nu_1,\nu_2}(95\%)$, des distributions F_{ν_1,ν_2}.