Advanced Mathematics

Lab 11

Yu-Hao Chiang 3443130

Exercise 1 – bathymetric data

The date set *depths.txt* contains the bathymetric depths in meters for an ocean basin part. Find the mean depth, the standard deviation and the 95% confidence interval on the mean depth. What is the probability that a random depth measurement will be shallower than -4000m?

$$\mu = -4393.7\ [m]$$

$$\sigma = 909.53\ [m]$$

$$\alpha = 0.05$$

$$95\%\ confidence\ interval\ on\ the\ mean\ depth$$

$$[\mu - 1.96\sigma, \qquad \mu + 1.96\sigma] = [-6176.4, -2611.1]$$

$$x = -4000\ [m]$$

$$z = \frac{x-\mu}{\sigma} = \frac{\left(-4000 - (-4393.7)\right)}{909.53} = 0.43$$

$$Look\ at\ Table\ 1:$$

$$for\ z > 0\ use\ P(z) = 1 - P(-z)$$

$$z = 0.43, \qquad P(0.43) = 1 - P(-0.43) = 1 - 0.3336 = 0.6664$$

$$P(x > -4000) = 1 - 0.6664 = 0.3336$$

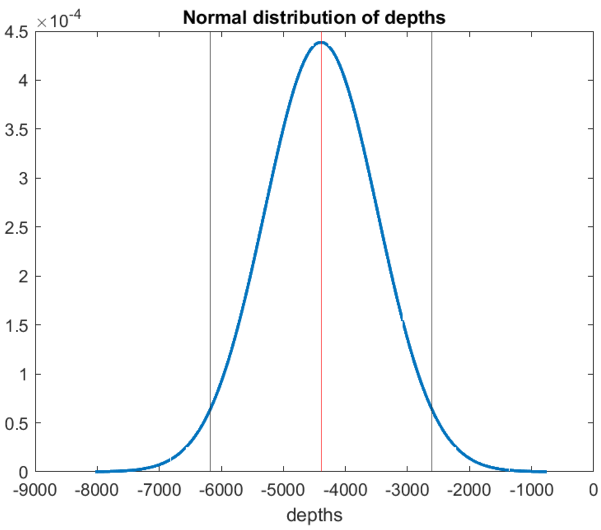


Figure 1 normal distribution of depths

z_P	**0.09**	**0.08**	**0.07**	**0.06**	**0.05**	**0.04**	**0.03**	**0.02**	**0.01**	**0.00**
-3.4	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003
-3.3	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005
-3.2	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007
-3.1	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010
-3.0	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013
-2.9	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019
-2.8	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026
-2.7	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035
-2.6	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047
-2.5	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062
-2.4	.0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080	.0082
-2.3	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107
-2.2	.0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139
-2.1	.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179
-2.0	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228
-1.9	.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287
-1.8	.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351	.0359
-1.7	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436	.0446
-1.6	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537	.0548
-1.5	.0559	.0571	.0582	.0594	.0606	.0618	.0630	.0643	.0655	.0668
-1.4	.0681	.0694	.0708	.0721	.0735	.0749	.0764	.0778	.0793	.0808
-1.3	.0823	.0838	.0853	.0869	.0885	.0901	.0918	.0934	.0951	.0968
-1.2	.0985	.1003	.1020	.1038	.1056	.1075	.1093	.1112	.1131	.1151
-1.1	.1170	.1190	.1210	.1230	.1251	.1271	.1292	.1314	.1335	.1357
-1.0	.1379	.1401	.1423	.1446	.1469	.1492	.1515	.1539	.1562	.1587
-0.9	.1611	.1635	.1660	.1685	.1711	.1736	.1762	.1788	.1814	.1841
-0.8	.1867	.1894	.1922	.1949	.1977	.2005	.2033	.2061	.2090	.2119
-0.7	.2148	.2177	.2206	.2236	.2266	.2296	.2327	.2358	.2389	.2420
-0.6	.2451	.2483	.2514	.2546	.2578	.2611	.2643	.2676	.2709	.2743
-0.5	.2776	.2810	.2843	.2877	.2912	.2946	.2981	.3015	.3050	.3085
-0.4	.3121	.3156	.3192	.3228	.3264	.3300	.3336	.3372	.3409	.3446
-0.3	.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783	.3821
-0.2	.3859	.3897	.3936	.3974	.4013	.4052	.4090	.4129	.4168	.4207
-0.1	.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562	.4602
-0.0	.4641	.4681	.4721	.4761	.4801	.4840	.4880	.4920	.4960	.5000

Table 1 Normal cumulative distribution function

Exercise 2 – plane booking

A small plane doing the Stuttgart-Frankfort link can accommodate 30 people every day. Statistics show 20% of customers who have booked do not come. Let X be the random variable: "number of customers who are present at the counter among 30 people who have reserved". What is the law of X (only the general form will be given)? What is its expectation and standard deviation? Give a confidence interval at the 95% threshold, making it possible to estimate the number of customers to expect.

$n = 30$

$P(x) = 0.8$

$E(x) = \mu = n * P = 24$

$\sigma = \sqrt{nP(1-P)} = 2.1909$

$95\%\ confidence\ interval\ threshold$

$[\mu - 1.96\sigma, \quad \mu + 1.96\sigma] = [19.7059, \quad 28.2941]$

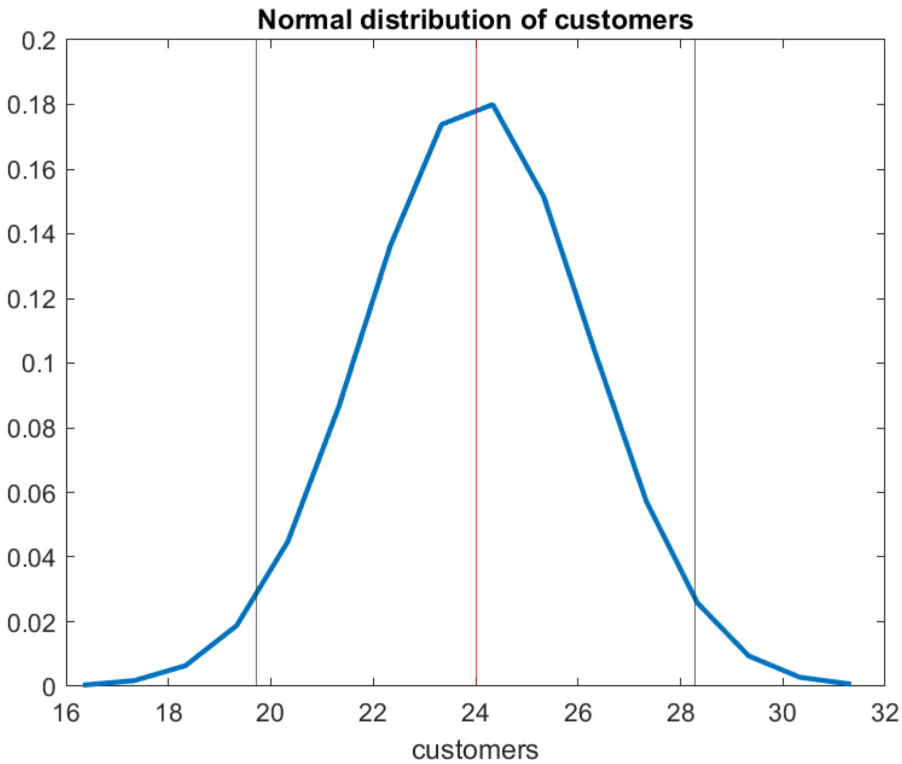


Figure 2 normal distribution of customers reserved the plane per day

$Look\ at\ figure\ 2, in\ 95\%\ confidence\ interval\ threshold:$

$we\ expect\ \mathbf{20} \sim \mathbf{28}\ people\ reserved\ the\ plane\ per\ day$

Exercise 3 – water contents of soils

The water contents of soils (in volume %) were measured at two sites A and B around Montpellier, France. There are reproduced in *soilwater.txt*. At the 99% level of confidence, do the soils at the two sites have different water content?

$\mu_A = 11.4153$

$\mu_B = 10.6513$

$\sigma_A = 2.0814$

$\sigma_B = 3.0275$

$n_A = 72$

$n_B = 80$

$\upsilon = n_A + n_B - 2 = 150\ (degree\ of\ freedom)$

$\alpha = 0.01\ (99\%\ level\ of\ confidence)$

For a two-tails test, we compute $1 - \alpha/2$, Look at the critical values for the student's distribution, using 99% confidence level, we can obtain the critical t value is 2.609.

$critical\ t\ value = 2.609$

We find the $t - statistic$ by evaluating.

$$t = \frac{\mu_A - \mu_B}{\sqrt{\frac{(n_A-1)\sigma_A^2 + (n_B-1)\sigma_B^2}{\nu}\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}} = 1.7934$$

The absolute value of the test statistic is less than the critical value (2.609), then we **accept** the null hypothesis (H_0).
The t distribution is symmetric so that:
$t_{1-\alpha/2,\nu} = -t_{\alpha/2,\nu}$

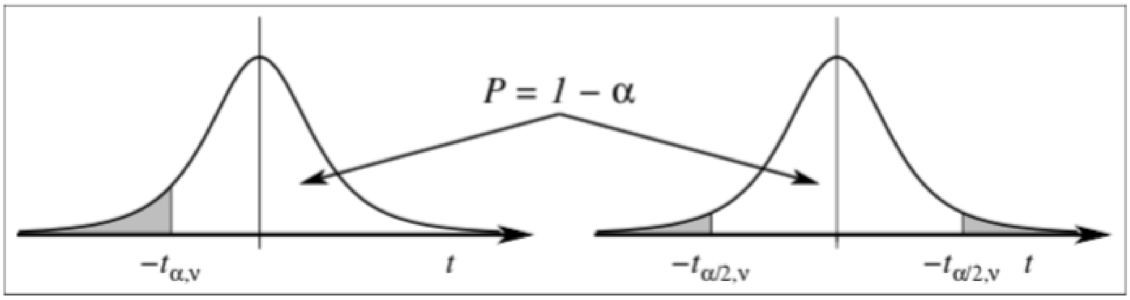


Figure 3 $t\ distribution$

Exercise 4 – toxic algae on beaches

We are interested in the problem of toxic algae that reaches certain beaches in France. After study, we note that 10% of beaches are affected by this type of algae and we want to test the influence of new chemical releases on the appearance of these algae. For that, 50 beaches close to the chemical rejection zones are observed. We then count 10 beaches affected by the harmful algae. Can you answer the question: “With the risk α = 0.05, have the chemical releases significantly changed the number of beaches affected?”.

$$n = 50$$

$$\bar{P} = \frac{10}{50} = 0.2$$

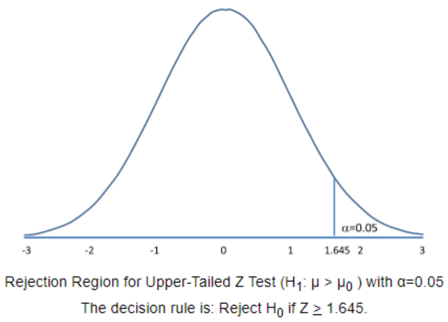
$$Hypotheses:$$

$$H_0: P \leq P_0\ (10\%)$$

$$H_a: P > P_0 (10\%)$$

$$Test\ statistic:$$

$$z = \frac{\bar{P} - P_0}{\sqrt{\frac{P_0(1 - P_0)}{n}}} = \frac{0.2 - 0.1}{\sqrt{\frac{0.1(1 - 0.1)}{50}}} = 2.36$$



Upper-Tailed Test	
α	Z
0.10	1.282
0.05	1.645
0.025	1.960
0.010	2.326
0.005	2.576
0.001	3.090
0.0001	3.719

Figure 4 Upper-Tailed Z test

$$\alpha = 0.05$$

$$z_{0.05} = 1.645$$

$$z = 2.36 > z_{0.05}$$

$$\boldsymbol{Reject}\ H_0$$

Exercise 5 – sulfur dioxide emissions

An environmental scientist measures the sulfur dioxide emissions from an industrial plant over an 80-day period. The amounts in tons per day are given in the file *sulfur.txt*. Bin the data using the categories less than 10, 10-15, 15-20, 20-25, 25 and above. Plot the histogram and indicate the counts. The scientist wonders if the emissions are well described by the expected normal distribution. What are the mean and standard deviation for the raw data? The scientist decides to use a χ^2test. What are the expected counts in each bin? Test whether or not the binned data are indistinguishable from a normal distribution at the 95% level of confidence. Should the scientist reject H_0?

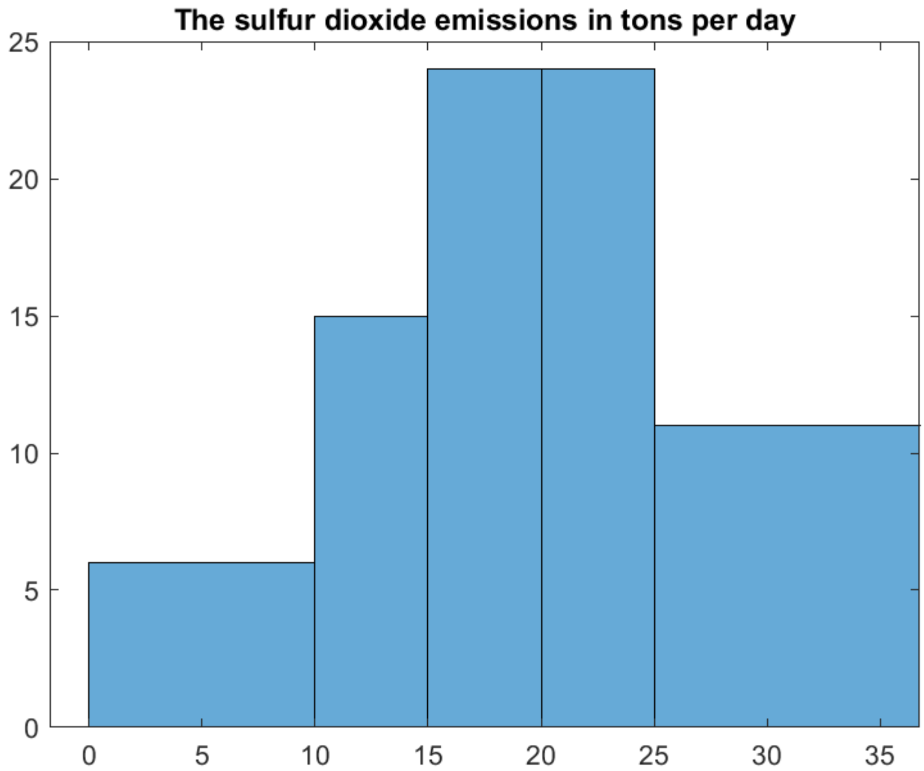


Figure 5 The sulfur dioxide emissions in tons per day.

$\mu = 18.8275$

$\sigma = 5.7195$

$normalize\ the\ data\ into\ normal\ scores:$

$$z_i = \frac{x_i - \mu}{\sigma}$$

$by\ desiring\ equal\ probabilities:$

$p_j = 0.2$

$the\ number\ of\ expected\ values\ in\ five\ bins.$

$\bar{x} = E_j = np_j = 16$

$the\ number\ of\ observed\ values\ in\ five\ bins: O_j$

$interval$	O_j
< -0.8416	18
$-0.8416 < z_i < -0.2533$	13
$-0.2533 < z_i < 0.2533$	17
$0.2533 < z_i < 0.8416$	14
> 0.8416	18

$normalize\ standard\ deviation$

$$s = \sqrt{np_j(1-p_j)} \approx \sqrt{np_j} = \sqrt{E_j}$$

$plugging\ in\ for\ \chi^2, we\ find$

$$\chi^2 = \sum_{j=1}^{5}\left(\frac{x_j - \bar{x}}{s}\right)^2 = \sum_{j=1}^{5}\left(\frac{O_j - E_j}{\sqrt{E_j}}\right)^2 = 1.375$$

The χ^2 distribution depends on υ, the degrees of freedom, which normally would be $\nu = k - 1\ (k = 5\ bins) = 4$ in our case (we lost one case since the bin counts must sum to n). However, we also used our observations to compute $\bar{x}$, then s, in order to determine the bin boundaries. These estimations further reduce ν by two, leaving just two degrees of freedom.

$\alpha = 0.05\ (95\%\ level\ of\ confidence)$

$\nu = 2$

$the\ critical\ value\ \chi^2 = 5.991$

Since it is much larger than our computed value, we conclude that we cannot at the 0.05 level of confidence reject the null hypothesis (**accept**)

Degrees of freedom, ν	Left tail (1 − α)					Right tail (α)				
	0.995	**0.99**	**0.975**	**0.95**	**0.90**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.60
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.34	12.84
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09	16.75
6	0.676	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81	18.55
7	0.989	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48	20.28
8	1.344	1.646	2.180	2.733	3.490	13.36	15.51	17.53	20.09	21.95
9	1.735	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67	23.59
10	2.156	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21	25.19
11	2.603	3.053	3.816	4.575	5.578	17.28	19.68	21.92	24.72	26.76
12	3.074	3.571	4.404	5.226	6.304	18.55	21.03	23.34	26.22	28.30
13	3.565	4.107	5.009	5.892	7.042	19.81	22.36	24.74	27.69	29.82
14	4.075	4.660	5.629	6.571	7.790	21.06	23.68	26.12	29.14	31.32
15	4.601	5.229	6.262	7.261	8.547	22.31	25.00	27.49	30.58	32.80
16	5.142	5.812	6.908	7.962	9.312	23.54	26.30	28.85	32.00	34.27
17	5.697	6.408	7.564	8.672	10.09	24.77	27.59	30.19	33.41	35.72
18	6.265	7.015	8.231	9.390	10.86	25.99	28.87	31.53	34.81	37.16
19	6.844	7.633	8.907	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.434	8.260	9.591	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.034	8.897	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.643	9.542	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.260	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.886	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.64
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.4	104.2
80	51.17	53.54	57.15	60.39	64.28	96.58	101.9	106.6	112.3	116.3
90	59.20	61.75	65.65	69.13	73.29	107.6	113.1	118.1	124.1	128.3
100	67.33	70.06	74.22	77.93	82.36	118.5	124.3	129.6	135.8	140.2

Table 2 Critical values for the χ^2 distribution