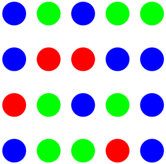
Statistics and Probability Theory

Exam Example



Please send your comments to: tijana.levajkovic@tuwien.ac.at

Exam - Before to start

- Please **bring**:
  a. a personal document/ID with your photo (mandatory)
  b. a pen with either blue or black ink (mandatory)
  c. a non-programmable calculator and a two-sided handwritten A4 sheet may be used during the exam (optional)
  d. water, fruits, snacks, etc. (optional)
- In the exam, you will **get** the following documents:
  1. Answer Form
  2. Question Form
  3. Statistical Tables
- Note that *computers, smartphones, tablets, notes, books*, etc., as well as *discussions and consultations* are **prohibited** during the exam.

1. Answer Form

- Blank Answer Form.

Form for answers

For automatic analysis

First name:

Surname:

Signature:

Invigilator

ID number

0	☐	☐	☐	☐	☐	☐	☐	☐	0
1	☐	☐	☐	☐	☐	☐	☐	☐	1
2	☐	☐	☐	☐	☐	☐	☐	☐	2
3	☐	☐	☐	☐	☐	☐	☐	☐	3
4	☐	☐	☐	☐	☐	☐	☐	☐	4
5	☐	☐	☐	☐	☐	☐	☐	☐	5
6	☐	☐	☐	☐	☐	☐	☐	☐	6
7	☐	☐	☐	☐	☐	☐	☐	☐	7
8	☐	☐	☐	☐	☐	☐	☐	☐	8
9	☐	☐	☐	☐	☐	☐	☐	☐	9

Group: A ☒ B ☐ C ☐ D ☐ E ☐ F ☐

This answer form will be scanned automatically. Please do not fold or spot. Use a black or blue pen to mark the fields:

☒

Only clear markings can be interpreted correctly! If you want to correct a marking, completely fill the box with color. This field will be interpreted like an empty box:

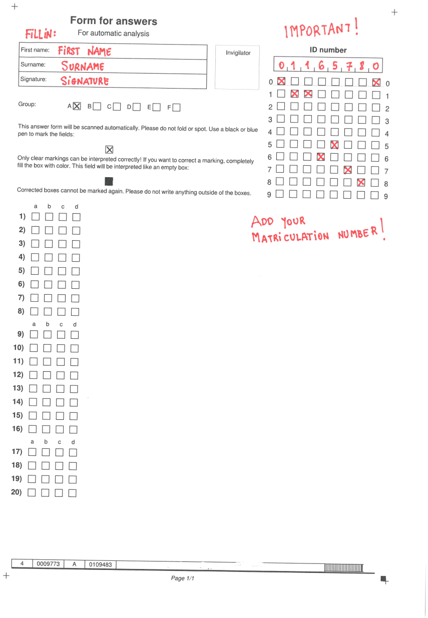
■

Corrected boxes cannot be marked again. Please do not write anything outside of the boxes.

	a	b	c	d
1)	☐	☐	☐	☐
2)	☐	☐	☐	☐
3)	☐	☐	☐	☐
4)	☐	☐	☐	☐
5)	☐	☐	☐	☐
6)	☐	☐	☐	☐
7)	☐	☐	☐	☐
8)	☐	☐	☐	☐
	a	b	c	d
9)	☐	☐	☐	☐
10)	☐	☐	☐	☐
11)	☐	☐	☐	☐
12)	☐	☐	☐	☐
13)	☐	☐	☐	☐
14)	☐	☐	☐	☐
15)	☐	☐	☐	☐
16)	☐	☐	☐	☐
	a	b	c	d
17)	☐	☐	☐	☐
18)	☐	☐	☐	☐
19)	☐	☐	☐	☐
20)	☐	☐	☐	☐

4	0009773	A	0109483

- Please, **fill in** your personal information: **Name**, **surname**, **signature**, matriculation number (student **ID number**) and its **coding**. A coding example is given below.



FILL IN:

Form for answers

For automatic analysis

First name: FIRST NAME

Surname: SURNAME

Signature: SIGNATURE

Invigilator

IMPORTANT!

ID number

0 1 1 6 5 7 8 0

Group: A ☒ B ☐ C ☐ D ☐ E ☐ F ☐

This answer form will be scanned automatically. Please do not fold or spot. Use a black or blue pen to mark the fields:

Only clear markings can be interpreted correctly! If you want to correct a marking, completely fill the box with color. This field will be interpreted like an empty box:

Corrected boxes cannot be marked again. Please do not write anything outside of the boxes.

ADD YOUR MATRICULATION NUMBER!

	a	b	c	d
1)	☐	☐	☐	☐
2)	☐	☐	☐	☐
3)	☐	☐	☐	☐
4)	☐	☐	☐	☐
5)	☐	☐	☐	☐
6)	☐	☐	☐	☐
7)	☐	☐	☐	☐
8)	☐	☐	☐	☐
9)	☐	☐	☐	☐
10)	☐	☐	☐	☐
11)	☐	☐	☐	☐
12)	☐	☐	☐	☐
13)	☐	☐	☐	☐
14)	☐	☐	☐	☐
15)	☐	☐	☐	☐
16)	☐	☐	☐	☐
17)	☐	☐	☐	☐
18)	☐	☐	☐	☐
19)	☐	☐	☐	☐
20)	☐	☐	☐	☐

4 | 0009773 | A | 0109483

Page 1/1

2. Question Form

- Blank Question Form.

107.254 Statistics and Probability Theory 2022W

Sixth examination term, Written Exam, Wednesday 17.01.2024, 12:00 PM

Group A

Name:

Matriculation number:

Signature:

This is a multiple-choice exam. It consists of 20 multiple-choice problems. All the problems carry 5 points each. The maximum score is 100.

For each problem, four possible answers (a, b, c, d) are offered and exactly one answer is correct. The answer that best completes the statement or answers the question should be chosen by ticking in the Answers form. There are no negative points for ticking a wrong answer. Ticking no answer or ticking more than one answer leads to the question being marked as incorrect. A pen with either blue or black ink has to be used.

A non-programmable calculator, a two-sided handwritten A4 sheet and statistical tables, which will be provided, may be used during the exam. The formulae sheet has to be submitted with the exam. Please note that a copy of a handwritten sheet is not a handwritten sheet and cannot be used in the exam. Computers, smartphones, tablets, notes, books, etc., as well as discussions and consultations are prohibited during the exam. It is mandatory to bring a personal document/ID with picture.

The examination time is 90 minutes.

Good luck!

[Group A]

- Please, **fill in** your **Name and surname, ID number** and **signature**.

107.254 Statistics and Probability Theory 2022W

Sixth examination term, Written Exam, Wednesday 17.01.2024, 12:00 PM

FILL IN

Group A

Name: NAME & SURNAME

Matriculation number: ID NUMBER

Signature: SIGNATURE

This is a multiple-choice exam. It consists of 20 multiple-choice problems. All the problems carry 5 points each. The maximum score is 100.

For each problem, four possible answers (a, b, c, d) are offered and exactly one answer is correct. The answer that best completes the statement or answers the question should be chosen by ticking in the Answers form. There are no negative points for ticking a wrong answer. Ticking no answer or ticking more than one answer leads to the question being marked as incorrect. A pen with either blue or black ink has to be used.

A non-programmable calculator, a two-sided handwritten A4 sheet and statistical tables, which will be provided, may be used during the exam. The formulae sheet has to be submitted with the exam. Please note that a copy of a handwritten sheet is not a handwritten sheet and cannot be used in the exam. Computers, smartphones, tablets, notes, books, etc., as well as discussions and consultations are prohibited during the exam. It is mandatory to bring a personal document/ID with picture.

The examination time is 90 minutes.

Good luck!

---

3. Statistical Tables

- Tables of Standard normal probabilities, Chi square distribution and t-distribution.

Table entry for z is the area under the standard normal curve to the left of z.

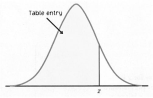


TABLE A Standard normal probabilities (*continued*)

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	.7704	.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

Table entry for p is the critical value x^* with probability p lying to its right.

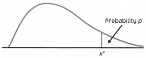


TABLE E Chi-square distribution critical values

	p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47	20.00
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51	22.11
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46	24.10
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32	26.02
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12	27.87
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88	29.67
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59	31.42
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26	33.14
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91	34.82
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53	36.48
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12	38.11
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70	39.72
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25	41.31
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79	42.88
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31	44.43
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82	45.97
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31	47.50
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80	49.01
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27	50.51
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73	52.00
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18	53.48
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62	54.95
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05	56.41
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48	57.86
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89	59.30
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30	60.73
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70	62.16
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40	76.09
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66	89.56
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61	102.7
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8	128.3
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4	153.2

Table entry for p and C is the critical value t^* with probability p lying to its right and probability C lying between $-t^*$ and t^*.

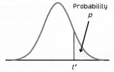


TABLE C t distribution critical values

	Upper tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	0.684	0.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	0.684	0.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	0.683	0.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	0.683	0.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	0.683	0.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	0.679	0.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	0.679	0.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	0.678	0.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	0.677	0.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	0.675	0.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
z^*	0.674	0.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level C											

Exam Example - Front page

This is a multiple-choice exam. It consists of 20 multiple-choice problems. All the questions carry 5 points each. The maximum score is 100.

For each problem, four possible answers (a, b, c, d) are offered and exactly one answer is correct. The answer that best completes the statement or answers the question should be chosen by ticking in the Answers form. There are no negative points for ticking a wrong answer. Ticking no answer or ticking more than one answer leads to the question being marked as incorrect. A pen with either blue or black ink has to be used.

The completion period is 90 minutes.

Good luck!

Examples

(1) A plumbing contractor obtains 60% of her boiler circulators from a company whose defect rate is 0.005, and the rest from a company whose defect rate is 0.01. What is the probability that a randomly chosen circulators is defective? If the chosen circulator is defective, what is the probability that it came from the first company?

a. 0.007 and 0.429
b. 0.007 and 0.571
c. 0.034 and 0.882
d. 0.034 and 0.118

Answer:
Let $C_i = \{\text{boiler circulators are obtained from i th company}\}$, $i = 1, 2$ and $D = \{\text{a circulator is defective}\}$. The following are given

$$P(D|C_1) = 0.005,\ P(D|C_2) = 0.01,\ P(C_1) = 0.6 \text{ and } P(C_2) = 0.4.$$

By the formula of the **total probability**, the probability of a defected circulator is

$$P(D) = P(D|C_1) \cdot P(C_1) + P(D|C_2) \cdot P(C_2) = 0.6 \cdot 0.005 + 0.4 \cdot 0.01 = 0.007.$$

From the **Bayes rule** we obtain

$$P(C_1|D) = \frac{P(C_1 \cap D)}{P(D)} = \frac{P(D|C_1) \cdot P(C_1)}{P(D)} = \frac{0.6 \cdot 0.005}{0.007} = 0.4285 \approx 0.429.$$

Answer: a

Examples

(2) Can the function

$$p(x) = \begin{cases} ax^2 + 2x - 1 & x = 1, 2, 3 \\ 0 & \text{else} \end{cases} \tag{1}$$

be the probability mass function for some discrete random variable? Here a is a real number.

a. Yes, only for a unique positive a.
b. No, because probabilities cannot be negative.
c. No, because probabilities cannot be greater than 1.
d. Yes, only for a unique negative a.

Answer:
The probabilities must be nonegative and to sum up to one.

x	1	2	3	
p	$a+1$	$4a+3$	$9a+5$	$14a+9$

Thus, from $1 = \sum_{x=1}^{3} P(X = x) = a + 1 + 4a + 3 + 9a + 5 = 14a + 9$ we get that only for $a = -4/7$ the probabilities sum up to one. In this case, $p(1) = P(X = 1) = 3/7$, $p(2) = P(X = 2) = 5/7$ but $p(3) = P(X = 3) = 9a + 5 = -\frac{36}{7} + 5 = -\frac{1}{7} < 0$. Thus, the function (1) cannot be the pdf of a discrete random variable.

Answer: b

Examples

(3) Pumpkins grown on a certain farm have normally distributed weights with a standard deviation of 2 kilograms. What is the expected weight if 85% of the pumpkins weigh less than 16 kg?

a. 14.30
b. 13.92
c. 14.88
d. 15.70

Answer:
Let X denotes the weight of a pampkin. Then $X \sim \mathcal{N}(\mu, 2^2)$. The goal is to find μ knowing that $P(X < 16) = 85\%$. Then, from

$$0.85 = P(X < 16) = P(\frac{X - \mu}{2} < \frac{16 - \mu}{2}) = P(Z < \frac{16 - \mu}{2}) = \Phi(\frac{16 - \mu}{2})$$

where Φ is the cdf of the standard normal distribution. From

$$\frac{16 - \mu}{2} = \Phi^{-1}(0.85)$$

we obtain

$$\mu = 16 - 2 \cdot \Phi^{-1}(0.85) \underset{\text{Tables}}{=} 16 - 2 \cdot 1.04 = 13.92.$$

Answer: b

Examples

(4) Let $X_1, \ldots, X_{64}$ be a random sample from a distribution with the expectation 1.2 and variance 4. Let

$$\bar{X} = \frac{1}{64}\sum_{i=1}^{64} X_i$$

be the sample mean. Determine the approximate value of $P(\bar{X} \leqslant 1.55)$ using the Central limit theorem and express it in terms of a a suitable R-function.

a. `pnorm(1.4)`
b. `pnorm(0.175)`
c. `pnorm(1.55, 1.2, 4)`
d. `pnorm(0.35, 0, 0.5)`

Answer:
For the sample mean we have $\mathbb{E}\bar{X} = \mathbb{E}X_i = 1.2$ and $\mathbb{V}ar\bar{X} = \frac{\sigma^2}{n} = \frac{1}{16} = 0.25^2$. By the CLT, for n large enough, the distribution of $\bar{X} \approx \mathcal{N}(\mu, \frac{\sigma^2}{n}) = \mathcal{N}(1.2, 0.25^2)$. Thus,

$$\frac{\bar{X} - \mathbb{E}\bar{X}}{\sqrt{\mathbb{V}ar\bar{X}}} = \frac{\bar{X} - 1.2}{\frac{1}{4}} = Z \approx \mathcal{N}(0, 1).$$

The unknown probability is approximately

$$P(\bar{X} \leqslant 1.55) \approx P(Z \leqslant \frac{1.55 - 1.2}{0.25}) = P(Z \leqslant 1.4) = \Phi(1.4) = \texttt{pnorm(1.4)}.$$

Answer: a

Examples

(5) Let X be a random variable with the probability density function

$$f_X(x) = \begin{cases} 1, & 0 \leqslant x \leqslant 1 \\ 0, & \text{else} \end{cases}.$$

Let $Z = -\ln X$ and let q be the first quartile of Z. Approximate q by using the table of approximate values of the natural logarithm.

x	1	1.5	2	4	8
$\approx \ln(x)$	0	0.4	0.7	1.4	2.1

a. q is smaller than 0.2
b. q is smaller than 0.45 but bigger than 0.2
c. q is bigger than 0.45
d. q is not unique

Answer:

For the unknown q it holds $F_Z(q) = 0.25$, i.e. $1 - e^{-q} = 0.25$. Namely,

$$0.25 = F_Z(q) = P(-\ln X \leqslant q) = P(\ln X \geqslant -q) = P(X \geqslant e^{-q}) = \int_{e^{-q}}^{1} dx = 1 - e^{-q}.$$

The lower quartile is $q = -\ln \frac{3}{4} = -\ln \frac{1.5}{2} = \ln 2 - \ln 1.5 \approx 0.3 \in (0.2, 0.45)$.

Answer: b

Examples

(6) Which one of the following holds **true** for two independent random variables X and Y whose variances are finite?

a. $\mathbb{V}ar(X+Y) \geqslant \mathbb{V}ar(X)$
b. $\mathbb{V}ar(X+Y) \leqslant \mathbb{V}ar(Y)$
c. $\mathbb{V}ar(2X-Y) = 2\mathbb{V}ar(X) - \mathbb{V}ar(Y)$
d. $\mathbb{V}ar(2X-Y) \leqslant 4\mathbb{V}ar(X)$

Answer:
In general, the variance of the sum of two random variables is

$$\mathbb{V}ar(X+Y) = \mathbb{V}ar(X) + \mathbb{V}ar(Y) + 2 \cdot \mathbb{C}ov(X,Y).$$

The random variables X and Y are independent, thus, their covariance equals zero. Then,

$$\mathbb{V}ar(X+Y) = \mathbb{V}ar(X) + \mathbb{V}ar(Y) \geqslant \mathbb{V}ar(X),$$

since $\mathbb{V}ar(Y) \geqslant 0$. Also,

$$\begin{aligned}\mathbb{V}ar(2X-Y) &= \mathbb{V}ar(2X) + \mathbb{V}ar(-Y) + 2\mathbb{C}ov(2X,-Y)\\ &= 2^2 \cdot \mathbb{V}ar(X) + (-1)^2 \cdot \mathbb{V}ar(Y) + 2 \cdot 2 \cdot (-1) \cdot \mathbb{C}ov(X,Y)\\ &= 4\mathbb{V}ar(X) + \mathbb{V}ar(Y) \geqslant 4\mathbb{V}ar(X).\end{aligned}$$

Answer: a

Examples

(7) In the situation of a left-sided one-sample t-test we find $\bar{x} = -8$, $s = 5$ and $n = 25$. For a given significance level we find the rejection region $R = (-\infty, -2.2]$. Then, for the null hypothesis $H_0 : \mu = -5$ it holds

a. we reject H_0, and we would also reject for any larger significance level.
b. we reject H_0, and we would also reject for any smaller significance level.
c. we do not reject H_0, but we would reject if only the significance level was chosen large enough.
d. we do not reject H_0, but we would reject if only the significance level was chosen small enough.

Answer:
We test $H_0 : \mu = -5$ against $H_A : \mu < -5$. We perform a t-test for the unknown population mean. The value of the test statistic $T = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$ for the given sample is

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{-8 - (-5)}{\frac{5}{5}} = -3 \in R.$$

We reject the null with the level of significance α. Moreover we would reject H_0 for any larger significance level as the rejection region $R \subseteq R_1$ for $\alpha \leqslant \alpha_1$.

Answer: a

Examples

(8) You perform a χ^2-test for independence in `R` using `chisq.test()`. From the output you can not read

a. the p-value
b. the χ^2-statistic
c. the degrees of freedom
d. the 95%-confidence interval for the expectation.

Answer: d

For example, let us look to the following output of `chisq.test(x,y)`

```
        Pearson's Chi-squared test

data:  x and y
X-squared = 6, df = 4, p-value = 0.1991
```

Examples

(9) In general, how does doubling the sample size change the confidence interval size?

a. Doubles the interval size
b. Halves the interval size
c. Multiplies the interval size by $\sqrt{2}$
d. Divides the interval size by $\sqrt{2}$

Answer:

The $(1-\alpha)\cdot 100\%$ CI for the unknown population mean is given by

$$(\bar{x} - q_{1-\frac{\alpha}{2}} \cdot SEM, \bar{x} + q_{1-\frac{\alpha}{2}} \cdot SEM),$$

where q_α is the αth quantile of the standard normal or of an appropriate t-distribution and $SEM = \frac{s}{\sqrt{n}}$. The size of the CI is $\text{size}_n = 2q_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$. For the doubled sample size $2n$ we obtain

$$\text{size}_{2n} = 2q_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{2n}} = 2q_{1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}} \cdot \frac{1}{\sqrt{2}} = \text{size}_n \cdot \frac{1}{\sqrt{2}}.$$

Answer: d

Examples

(10) What is the critical t-value for finding a 90% confidence interval estimate from a sample of 15 observations?

a. 1.341
b. 1.761
c. 1.350
d. 1.753

Answer:
With $df = 15 - 1 = 14$ and 5% in each tail, from the table we read that the critical t-value is 1.761.

TABLE C t distribution critical values

	Upper tail probability p											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015

Answer: b

Examples

(11) A study cross-classified 1500 randomly selected adults in three categories of education level (not a high school graduate, high school graduate, and college graduate) and five categories of major sports interest (baseball, basketball, football, hockey, and tennis). If the *p*-value of an appropriate χ^2- test for independence is 0.083, what can be said, Is there evidence of a relationship between education level and sports interest?

a. The data prove there is a relationship between education level and sports interest.
b. There is sufficient evidence at the 5% significance level of a relationship between education level and sports interest.
c. There is sufficient evidence at the 10% significance level, but not at the 5% significance level, of a relationship between education level and sports interest.
d. The *p*-value is greater than 0.10, so there is no evidence of a relationship between education level and sports interest.

Answer:
We use a χ^2- test for independence to test
H_0: education level and sports interest are independent against
H_A : education level and sports interest are not independent.

Examples

The p-value $= P_{H_0}(X^2 \geqslant x^2) = 0.083$.

If p-value $< \alpha$ then we reject H_0 and conclude that there is sufficient evidence at level α of a relationship between education level and interests in sports.

Since

$$0.05 < 0.083 < 0.10,$$

we reject H_0 for $\alpha = 10\%$ and do not reject H_0 for $\alpha = 5\%$.

Conclusion: there is sufficient evidence at the 10% significance level, but not at the 5% significance level, of a relationship between education level and sports interest.

Answer: c

Examples

(12) Which one of the following is a true statement?

a. The larger the sample, the larger the spread in the sampling distribution.
b. Sample parameters are used to make inferences about population statistics.
c. Provided that the population size is significantly greater than the sample size, the spread of the sampling distribution does not depend on the population size.
d. Statistics from smaller samples have less variability.

Answer:

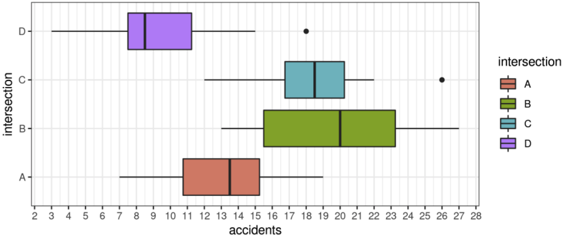
The sampling distribution, i.e. the distribution of $\bar{X}$, has expectation (mean) μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. This means that by increasing the sample size the variance of $\bar{X}$ is decreasing, i.e., the larger the sample, the smaller the spread in the sampling distribution.

Sample statistics are used to make inferences about population parameters. Statistics from smaller samples have more variability.

Answer: c

Examples

(13) Data on the number of yearly accidents were collected from four intersections (A-D) over a 20 year period. Which of the following statements is **false**?



a. During at least 75% of years, intersection D had fewer accidents than the lowest 25% of years at intersection A.
b. During at least 15 years, fewer than 12 accidents occurred at intersection D.
c. The maximum number accidents that occurred in a single intersection is 27.
d. All of the accidents totals at intersection D were lower than the median number of accidents at intersection B.

Answer:
First quartile A is lower than third quartile D.
15 is 3/4, at the third quartile of D is below 12.

Answer: a

Examples

(14) Suppose you do five independent right-sided tests for testing $H_0 : \mu = 38$, each at the $\alpha = 0.01$ significance level. What is the probability of committing a Type I error and incorrectly rejecting a true null hypothesis in at least one of the five tests?

a. 0.01
b. 0.049
c. 0.226
d. 0.951

Answer:

First note, Type I error = the level of significance = 0.01.
Denote by X the number of tests (out of five) in which a Type I error was committed. Then, $X \sim B(5, 0.01)$. We are looking for the probability

$$P(\text{at least one Type I error}) = 1 - P(\text{no Type I errors})$$
$$= 1 - P(X = 0) =$$
$$1 - (0.99)^5 = 0.049.$$

Answer: b

(15) Suppose the correlation is negative. Given two points from the scatterplot, which of the following is possible?

- I The first point has a larger x-value and a smaller y-value than the second point.
- II The first point has a larger x-value and a larger y-value than the second point.
- III The first point has a smaller x-value and a larger y-value than the second point.

a. I only
b. II only
c. I and III
d. I, II, and III

Answer:

Negative correlation shows a tendency for higher values of one variable to be associated with lower values of the other. However, given any two points, anything is possible.

Answer: d

Examples

(16) Suppose the average score on a certain test is 500 with a standard deviation of 100. If each score is increased by 25, what are the new mean and standard deviation?

a. 500, 100
b. 500, 125
c. 525, 100
d. 525, 125

Answer: It is given

$$\bar{x} = \frac{1}{n}\sum_i x_i = 500 \quad \text{and} \quad \sigma^2 = \frac{1}{n}\sum_i (x_i - \bar{x})^2 = 100^2.$$

Then, the sample mean and variance of the modified sequence are

$$\bar{x}_{mod} = \frac{1}{n}\sum_i (x_i + 25) = \frac{1}{n}(\sum_i x_i + 25n) = \frac{1}{n}\sum_i x_i + 25 = \bar{x} + 25 = 525$$

$$\sigma^2_{mod} = \frac{1}{n}\sum_i (x_i + 25 - \bar{x}_{mod})^2 = \frac{1}{n}\sum_i (x_i + 25 - \bar{x} - 25)^2 = \frac{1}{n}\sum_i (x_i - \bar{x})^2 = 100^2.$$

Adding the same constant to every value increases the mean by that same constant. The distances between the increased values and the increased mean stay the same, and so the standard deviation is unchanged.

Answer: c

(17) In a linear regression model ('y_i modeled as a linear function of x_i plus error') the parameters are estimated via least squares. For the mean and the empirical variance of the x and y values we obtain $\bar{x} = 5$, $s_x^2 = 4$, $\bar{y} = 7$ and $s_y^2 = 9$. It holds that

a. the regression line goes through $(5, 6)$.
b. the regression line goes through $(7, 7)$.
c. the slope of the regression line is smaller or equals 1.5.
d. the slope of the regression line is larger than 1.5.

Answer:

The regression line is $y = b_0 + b_1 x$, where the slope isgiven as

$$b_1 = \frac{r s_y}{s_x} = r \cdot \frac{3}{2}$$

and as $|r| \leqslant 1$, the slope is smaller or equals 1.5. Also, we know that the regression line passes through the center of the masses $(\bar{x}, \bar{y}) = (5, 7)$, so (a) and (b) are not correct. Answer: c

Examples

(18) The sampling distribution of the sample mean is close to the normal distribution

a. only if both the original population has a normal distribution and n is large.
b. if the standard deviation of the original population is known.
c. no matter what the value of n or what the distribution of the original population.
d. if n is large, no matter what the distribution of the original population.

Answer:

From the CLT, it follwos that for n large the distribution of the sample mean is

$$\bar{X} \approx \mathcal{N}(\mu, \frac{\sigma^2}{n})$$

no matter what is the distribution of the original population. In particular, if the sample comes from a normal distribution then the sample mean is also normally distributed.

Answer: d

Examples

(19) In the context of a statistical test at significance level α, the p-value is below α. Which of the following is **true**?

a. The test statistic lies in the rejection region.
b. We would also reject at level $\alpha/2$.
c. The value of the null hypothesis lies within the $(1-\alpha)100\%$ confidence interval computed from these data.
d. The rejection region would be bigger if we would have used level $\alpha/2$.

Answer:

As p-value $< \alpha$ we reject H_0 and the test statistic falls in the rejection region. As we reject on the level of significance α we would also reject for all significance levels bigger than α as in this case the rejection regions increase.

Answer: a

Examples

(20) A 95% confidence interval for a population mean based on a sample of size 400 was (20, 28). Which of the following is **true**?

a. If we were to collect a new sample, then the interval created using it would contain the true population mean with probability 95%.
b. Across many samples, 95% of sample means should lie within an interval made by this method.
c. The sample mean has a 95% chance of being in (20, 28).
d. There is a 95% probability that the true value of the population mean is 24.

Examples

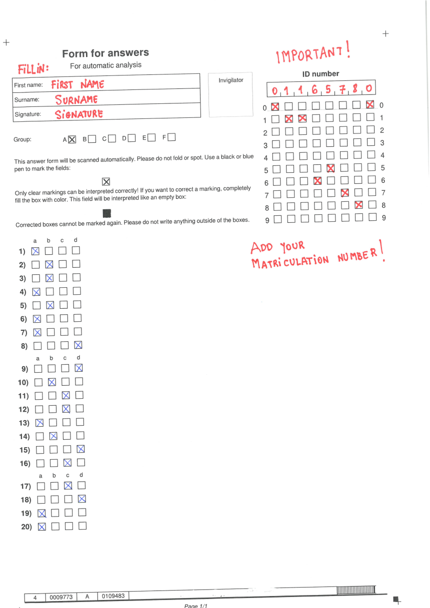
(20) A 95% confidence interval for a population mean based on a sample of size 400 was (20, 28). Which of the following is **true**?

a. If we were to collect a new sample, then the interval created using it would contain the true population mean with probability 95%.
b. Across many samples, 95% of sample means should lie within an interval made by this method.
c. The sample mean has a 95% chance of being in (20, 28).
d. There is a 95% probability that the true value of the population mean is 24.

Answer: a

Answer Form - Filled

- All questions answered. Anwer Form completed.



Form for answers

For automatic analysis

FILL IN:

First name: FIRST NAME

Surname: SURNAME

Signature: SIGNATURE

Invigilator

IMPORTANT!

ID number

0 1 1 6 5 7 8 0

Group: A B C D E F

This answer form will be scanned automatically. Please do not fold or spot. Use a black or blue pen to mark the fields:

Only clear markings can be interpreted correctly! If you want to correct a marking, completely fill the box with color. This field will be interpreted like an empty box:

Corrected boxes cannot be marked again. Please do not write anything outside of the boxes.

ADD YOUR MATRICULATION NUMBER!

	a	b	c	d
1)	X			
2)		X		
3)		X		
4)	X			
5)		X		
6)	X			
7)	X			
8)				X
9)				X
10)		X		
11)			X	
12)			X	
13)	X			
14)		X		
15)				X
16)			X	
17)			X	
18)				X
19)	X			
20)	X			

4 | 0009773 | A | 0109483

Page 1/1

Thank you for your attention and

good luck in the exam!

Course evaluation

Thank you for evaluating the course!