

Week 9

AP Statistics

Homework bundle for this week

本周作业合集

exercises.lol

Contents 目录

Topic 3 quiz	3
Exercise sheet 3.1	6
Past-paper questions 3.1	9
Exercise sheet 3.2	18
Past-paper questions 3.2	21
Exercise sheet 3.3	24
Past-paper questions 3.3	27
Exercise sheet 3.4	42
Past-paper questions 3.4	45
Exercise sheet 3.5	58
Past-paper questions 3.5	62
Exercise sheet 3.6	76

Past-paper questions 3.6	79
--------------------------------	----

Exercise sheet 3.7	90
--------------------------	----

Past-paper questions 3.7	93
--------------------------------	----

3. Collecting Data

Starter Quiz · 课前小测

AP Statistics

Name: _____ Score: _____

1. Collecting a much larger sample using the same flawed method will remove the bias.
☐ True ☐ False
2. Which type of study can establish a cause-and-effect relationship?
☐ a well-designed experiment with random assignment
☐ any observational study
☐ a large survey
☐ a call-in poll
3. A simple random sample (SRS) of size n gives an equal chance to...
☐ every possible group of n individuals
☐ only the first n people
☐ every other person
☐ the n most convenient people
4. A phone survey never reaches people without phones. This is...
☐ undercoverage
☐ response bias
☐ nonresponse
☐ a placebo effect
5. A fake treatment (like a sugar pill) that looks real is a...
☐ placebo
☐ factor
☐ block
☐ stratum
6. A matched-pairs design is a special case of a randomized block design with blocks of size...
☐ 2
☐ 1
☐ 10
☐ the whole sample

7. In an observational study, the researcher imposes a treatment on the subjects.

☐ True ☐ False

8. Random sampling is what justifies generalizing the sample result to the population.

☐ True ☐ False

9. Blocking groups similar experimental units to remove a known source of variability.

☐ True ☐ False

10. A difference too large to be reasonably explained by chance is called statistically ____.

AP Statistics

1. **False**

Size fights random error, not bias — a flawed method stays biased at any size.

2. **a well-designed experiment with random assignment**

Only random assignment in an experiment supports causation.

3. **every possible group of n individuals**

An SRS gives every group of n an equal chance.

4. **undercoverage**

Part of the population has no chance of selection — undercoverage.

5. **placebo**

A placebo controls for the placebo effect.

6. **2**

Matched pairs = blocks of size 2.

7. **False**

Observational = watch and measure without intervening; experiments impose treatments.

8. **True**

A bias-free random method supports generalization.

9. **True**

That's exactly the purpose of blocking.

10. **significant**

That's the meaning of statistically significant.

3.1 Introducing Statistics: Do the Data We Collected Tell the Truth?

Name: _____ Class: _____ Date: _____ **Total: 8 marks**

KEY WORDS bias 偏差 • population 总体 • generalization 推广 • sample 样本 • systematic 系统

Objective

Build the skills to answer exam questions on **whether the data tell the truth**.

You must be able to:

- explain how the **method** of data collection affects the conclusions
- recognise sources of **bias** 偏差 that make data unrepresentative
- distinguish data that support a **generalization** 推广 from data that do not

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ The method matters

How data are collected shapes what conclusions are valid. A flawed method can make even a large data set useless.

■ Bias

Bias systematically makes a sample **unrepresentative** of the population, so conclusions are distorted.

■ Generalizing

Only data from a well-chosen (ideally random) sample can be **generalized** to the whole population.

2 Practice

2.1 State how the method of data collection affects conclusions. [1]

2.2 Define bias in data collection. [1]

2.3 State what kind of sample allows generalization to a population. [1]

3 Exam-style questions

3.1 Biased data are [1]

- **A** always accurate
- **B** unrepresentative of the population
- **C** very large
- **D** perfectly random

3.2 The way data are collected affects [1]

- **A** the population size
- **B** the conclusions we can draw
- **C** the mean only
- **D** nothing at all

3.3 A poll surveys only people at a gym about their exercise habits.

(a) State whether the sample is representative of all adults. [1]

(b) Name the problem. [1]

(c) State whether the results can generalize to all adults. [1]

Solutions

2.1 a poor method can make the data unrepresentative, so the conclusions are invalid.

2.2 a systematic tendency for a sample to misrepresent the population.

2.3 a random (representative) sample.

3.1 B.

3.2 B.

3.3 (a) no. (b) bias (undercoverage — gym-goers are not typical). (c) no.

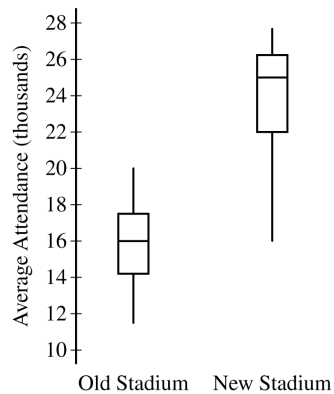
Part B

Question 6

Spend about 25 minutes on this part of the exam.

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

5. Attendance at games for a certain baseball team is being investigated by the team owner. The following boxplots summarize the attendance, measured as average number of attendees per game, for 47 years of the team's existence. The boxplots include the 30 years of games played in the old stadium and the 17 years played in the new stadium.

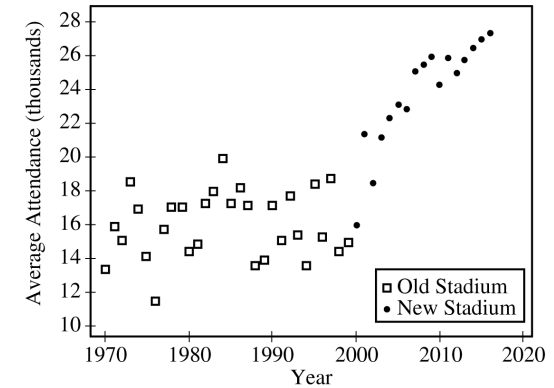


- (a) Compare the distributions of average attendance between the old and new stadiums.

GO ON TO THE NEXT PAGE.

Use a pencil or pen with black or dark blue ink only. Do NOT write your name. Do NOT write outside the box.

The following scatterplot shows average attendance versus year.

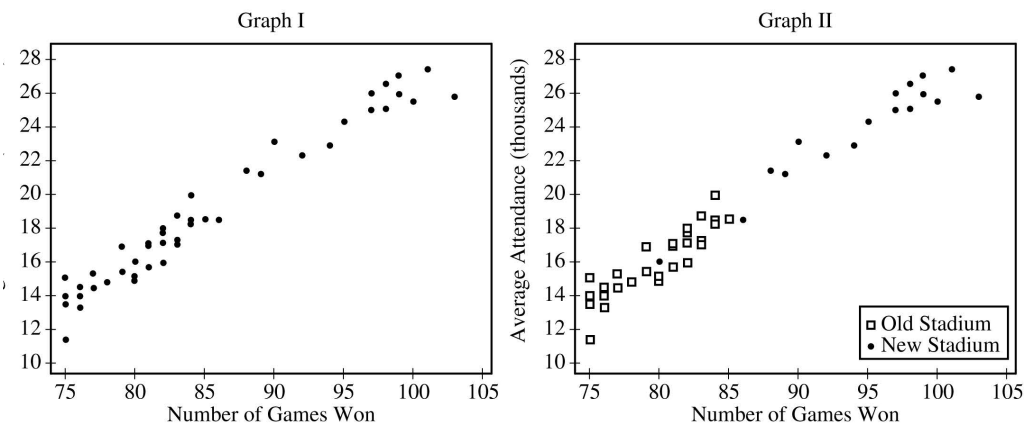


- (b) Compare the trends in average attendance over time between the old and new stadium.

GO ON TO THE NEXT PAGE.

Use a pencil or pen with black or dark blue ink only. Do NOT write your name. Do NOT write outside the box.

(c) Consider the following scatterplots.



(i) Graph I shows the average attendance versus number of games won for each year. Describe the relationship between the variables.

(ii) Graph II shows the same information as Graph I, but also indicates the old and new stadiums. Does Graph II suggest that the rate at which attendance changes as number of games won increases is different in the new stadium compared to the old stadium? Explain your reasoning.

GO ON TO THE NEXT PAGE.

Use a pencil or pen with black or dark blue ink only. Do NOT write your name. Do NOT write outside the box.

(d) Consider the three variables: number of games won, year, and stadium. Based on the graphs, explain how one of those variables could be a confounding variable in the relationship between average attendance and the other variables.

GO ON TO THE NEXT PAGE.

Use a pencil or pen with black or dark blue ink only. Do NOT write your name. Do NOT write outside the box.

STOP

END OF EXAM

Use a pencil or pen with black or dark blue ink only. Do NOT write your name. Do NOT write outside the box.

Table entry for z is the probability lying below z .

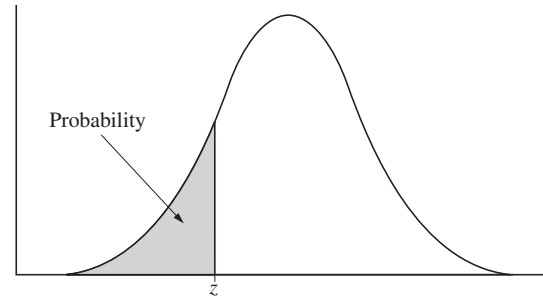


Table A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Name:

AP Statistics: 2021

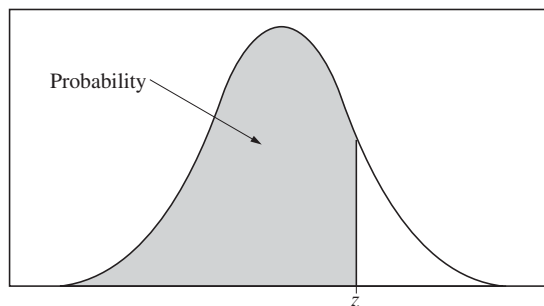


Table entry for z is the probability lying below z .

Table A (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

Name:

AP Statistics: 2021

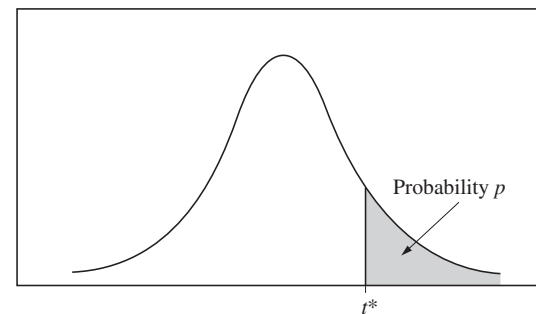


Table entry for p and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^* .

Table B t distribution critical values

df	Tail probability p									
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807
Confidence level C										
50% 60% 70% 80% 90% 95% 96% 98% 99% 99.5% 99.8% 99.9%										

Table entry for p is the point (χ^2) with probability p lying above it.

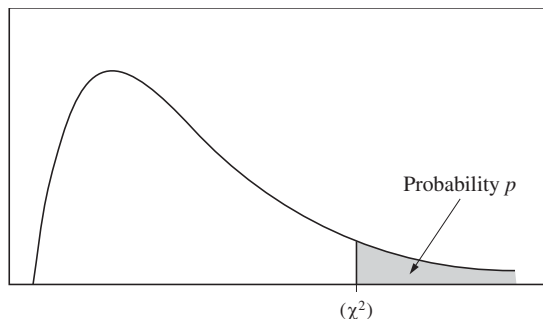


Table C χ^2 critical values

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

AP STATISTICS • EXERCISE SHEET

3.2 Introduction to Planning a Study

Name: _____ Class: _____ Date: _____ Total: 10 marks

KEY WORDS experiment 实验 • observational study 观察性研究 • population 总体
• sample 样本 • treatment 处理

Objective

Build the skills to answer exam questions on **planning a study**.

You must be able to:

- distinguish an **observational study** 观察性研究 from an **experiment** 实验
- identify the **population** 总体 and the **sample** 样本
- explain why only a well-designed experiment can show **cause and effect** 因果

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ Observational study vs experiment

- Observational study** — observe and record **without imposing** any treatment.
- Experiment** — deliberately **impose treatments** and compare the responses.

■ Population and sample

The **population** is the whole group of interest; the **sample** is the part actually studied.

■ Cause and effect

Only a **well-designed experiment** (with randomly assigned treatments) can establish a **cause-and-effect** relationship.

2 Practice

2.1 State the difference between an observational study and an experiment. [2]

2.2 State the difference between a population and a sample. [2]

2.3 State which type of study can establish cause and effect. [1]

3 Exam-style questions

3.1 Only a well-designed _____ can show cause and effect. [1]

- **A** observational study
- **B** experiment
- **C** survey
- **D** census

3.2 The group actually studied is the [1]

- **A** population
- **B** sample
- **C** treatment
- **D** response

3.3 Researchers give one group a new drug and another a placebo, then compare the results.

(a) Name this type of study. [1]

(b) Name the whole group of interest. [1]

(c) State whether a cause-and-effect conclusion is possible. [1]

Solutions

2.1 an observational study only observes; an experiment imposes treatments.

2.2 the population is the whole group of interest; the sample is the part actually studied.

2.3 a well-designed experiment.

3.1 B.

3.2 B.

3.3 (a) an experiment. (b) the population. (c) yes (with random assignment).

I. Descriptive Statistics

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{\sum x_i}{n}$$

$$\hat{y} = a + bx$$

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$$s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$\bar{y} = a + b\bar{x}$$

$$b = r \frac{s_y}{s_x}$$

II. Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 P(x_i)}$
If X has a binomial distribution with parameters n and p , then: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $x = 0, 1, 2, 3, \dots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1-p)}$
If X has a geometric distribution with parameter p , then: $P(X = x) = (1-p)^{x-1} p$ where $x = 1, 2, 3, \dots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \sqrt{\frac{1-p}{p^2}}$

III. Sampling Distributions and Inferential Statistics

$$\text{Standardized test statistic: } \frac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$$

$$\text{Confidence interval: } \text{statistic} \pm (\text{critical value})(\text{standard error of statistic})$$

$$\text{Chi-square statistic: } \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$ where $\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}$

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\bar{X}$	$\mu_{\bar{X}} = \mu$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{X}} = \frac{s}{\sqrt{n}}$
For two populations: $\bar{X}_1 - \bar{X}_2$	$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$	$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Sampling distributions for simple linear regression:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For slope: b	$\mu_b = \beta$	$\sigma_b = \frac{\sigma}{\sigma_x \sqrt{n}}$, where $\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$	$s_b = \frac{s}{s_x \sqrt{n-1}}$, where $s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$ and $s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

*Standard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the standard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.

3. Alzheimer’s disease results in a loss of cognitive ability beyond what is expected with typical aging. A local newspaper published an article with the following headline.

Study Finds Strong Association Between Smoking and Alzheimer’s

The article reported that a study tracked the medical histories of 21,123 men and women for 23 years. The article stated that, for those who smoked at least two packs of cigarettes a day, the risk of developing Alzheimer’s disease was 2.57 times the risk for those who did not smoke.

(a) Identify the explanatory and response variables in the study.

Explanatory variable:

Response variable:

(b) Is the study described in the article an observational study or an experiment? Explain.

(c) Exercise status (regular weekly exercise versus no regular weekly exercise) was mentioned in the article as a possible confounding variable. Explain how exercise status could be a confounding variable in the study.

AP STATISTICS • EXERCISE SHEET

3.3 Random Sampling and Data Collection

Name: _____ Class: _____ Date: _____

Total: 9 marks

KEY WORDS sample 样本 • cluster 整群 • random sampling 随机抽样 • systematic 系统
• bias 偏差

Objective

Build the skills to answer exam questions on **random sampling and data collection**.

You must be able to:

- explain why **random sampling** 随机抽样 helps produce a representative sample
- describe a **simple random sample (SRS)** 简单随机抽样
- distinguish **stratified** 分层, **cluster** 整群, and **systematic** 系统 random sampling

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ Why random?

Random sampling gives every individual a known chance of selection, which produces a **representative** sample and reduces bias.

■ Sampling methods

- **Simple random sample (SRS)** — every possible sample of the chosen size is equally likely.
- **Stratified** — split into groups (strata), then randomly sample **within each**.
- **Cluster** — randomly select whole **groups** and use everyone in them.
- **Systematic** — pick every *k*-th individual from a list.

2 Practice

2.1 State why random sampling helps. [1]

2.2 Describe a simple random sample. [1]

2.3 Name two other types of random sampling. [2]

3 Exam-style questions

3.1 Random sampling helps produce a _____ sample. [1]

- **A** biased
- **B** representative
- **C** small
- **D** voluntary

3.2 Splitting a population into groups and randomly sampling within each is _____ sampling. [1]

- **A** cluster
- **B** stratified
- **C** systematic
- **D** voluntary

3.3 A school selects every 10th student on an alphabetical list.

(a) Name this sampling method. [1]

(b) State one benefit of random sampling. [1]

(c) Name a method that samples whole groups. [1]

Solutions

2.1 it gives a representative sample and reduces bias.

2.2 a sample in which every possible group of that size is equally likely to be chosen.

2.3 any two of: stratified, cluster, systematic.

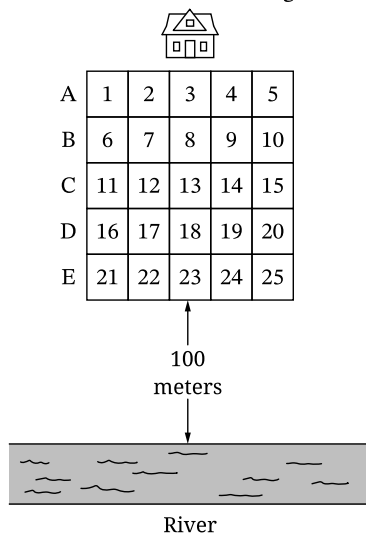
3.1 B.

3.2 B.

3.3 (a) systematic sampling. (b) it reduces bias / gives a representative sample. (c) cluster sampling.

2. Aphids are tiny insects that feed on plants such as cabbage plants. A farmer wants to reduce the number of aphids in a cabbage field. A river is located 100 meters south of the cabbage field. The farmer divides the field into 25 regions of equal size, as shown in the diagram. Each region has approximately the same number of cabbage plants.

Farmer's House and Cabbage Field



The farmer would like to estimate the proportion of cabbage plants in the field that are affected by aphids and believes that the extent of aphid damage is greater for the regions in the cabbage field closer to the river. To obtain the estimate, the farmer is considering three sampling methods.

- Sampling method I: Select region 3, which is closest to the farmer's house and farthest from the river. Examine every cabbage plant in the region for aphid damage.
- Sampling method II: Randomly select one row (A, B, C, D, or E). For every region in the selected row, examine every cabbage plant for aphid damage.
- Sampling method III: Randomly select one region from each of rows A, B, C, D, and E. For each selected region, examine every cabbage plant for aphid damage.

Descriptive Statistics

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{\sum x_i}{n}$$

$$\hat{y} = a + bx$$

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$$s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$\bar{y} = a + b\bar{x}$$

$$b = r \frac{s_y}{s_x}$$

Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 P(x_i)}$
If X has a binomial distribution with parameters n and p , then: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $x = 0, 1, 2, 3, \dots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1-p)}$
If X has a geometric distribution with parameter p , then: $P(X = x) = (1-p)^{x-1} p$ where $x = 1, 2, 3, \dots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1-p}}{p}$

Sampling Distributions and Inferential Statistics

$$\text{Standardized test statistic: } \frac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$$

$$\text{Confidence interval: } \text{statistic} \pm (\text{critical value})(\text{standard error of statistic})$$

$$\text{Chi-square statistic: } \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ where $\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}$

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\bar{X}$	$\mu_{\bar{X}} = \mu$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{X}} = \frac{s}{\sqrt{n}}$
For two populations: $\bar{X}_1 - \bar{X}_2$	$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$	$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Sampling distributions for simple linear regression:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For slope: b	$\mu_b = \beta$	$\sigma_b = \frac{\sigma}{\sigma_x \sqrt{n}}$ where $\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$	$s_b = \frac{s}{s_x \sqrt{n-1}}$ where $s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$ and $s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

Standard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the standard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.

STATISTICS
SECTION II
Part B
Question 6

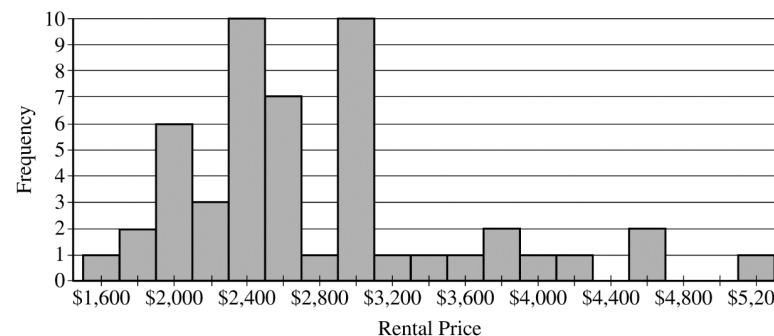
Spend about 25 minutes on this part of the exam.
Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. Emma is moving to a large city and is investigating typical monthly rental prices of available one-bedroom apartments. She obtained a random sample of rental prices for 50 one-bedroom apartments taken from a Web site where people voluntarily list available apartments.

(a) Describe the population for which it is appropriate for Emma to generalize the results from her sample.

The distribution of the 50 rental prices of the available apartments is shown in the following histogram.



- (b) Emma wants to estimate the typical rental price of a one-bedroom apartment in the city. Based on the distribution shown, what is a disadvantage of using the mean rather than the median as an estimate of the typical rental price?
- (c) Instead of using the sample median as the point estimate for the population median, Emma wants to use an interval estimate. However, computing an interval estimate requires knowing the sampling distribution of the sample median for samples of size 50. Emma has one point, her sample median, in that sampling distribution. Using information about rental prices that are available on the Web site, describe how someone could develop a theoretical sampling distribution of the sample median for samples of size 50.

Because Emma does not have the resources to develop the theoretical sampling distribution, she estimates the sampling distribution of the sample median using a process called bootstrapping. In the bootstrapping process, a computer program performs the following steps.

- Take a random sample, with replacement, of size 50 from the original sample.
- Calculate and record the median of the sample.
- Repeat the process to obtain a total of 15,000 medians.

Emma ran the bootstrap process, and the following frequency table is the bootstrap distribution showing her results of generating 15,000 medians.

Bootstrap Distribution of Medians					
Median	Frequency	Median	Frequency	Median	Frequency
2,345	1	2,585	1	2,825	247
2,390	13	2,587.5	171	2,837.5	7
2,395	18	2,600	22	2,847.5	1
2,400	56	2,612.5	1,190	2,872.5	317
2,445	4	2,625	174	2,885	10
2,447.5	56	2,672.5	5	2,950	700
2,450	55	2,675	1,924	2,962.5	93
2,475	3	2,687.5	1,341	2,972.5	6
2,495	66	2,700	2,825	2,975	65
2,497.5	136	2,735	35	2,985	12
2,500	1,899	2,747.5	619	2,987.5	1
2,522.5	2	2,750	2	2,995	6
2,525	945	2,795	278	3,000	2
2,550	1,673	2,812.5	16	3,062.5	3

The bootstrap distribution provides an approximation of the sampling distribution of the sample median. A confidence interval for the median can be constructed using a percentage of the values in the middle of the bootstrap distribution.

- (d) Use the frequency table to find the following.
- Value of the 5th percentile:
 - Value of the 95th percentile:
- (e) Find the percentage of bootstrap medians in the table that are equal to or between the values found in part (d).
- (f) Use your values from parts (d) and (e) to construct and interpret a confidence interval for the median rental price.

STOP

END OF EXAM

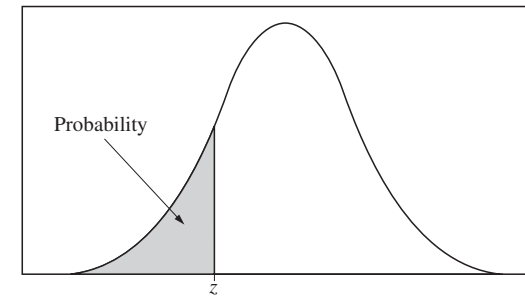


Table entry for z is the probability lying below z .

Table A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641



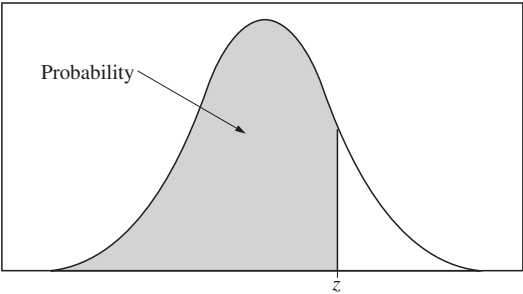


Table entry for z is the probability lying below z .

Table A (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 and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^* .

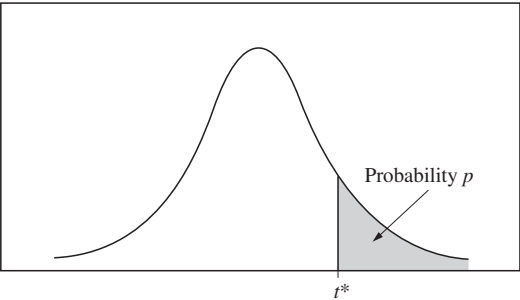


Table B t distribution critical values

df	Tail probability p									
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807
Confidence level C										
50% 60% 70% 80% 90% 95% 96% 98% 99% 99.5% 99.8% 99.9%										

Table entry for p is the point (χ^2) with probability p lying above it.

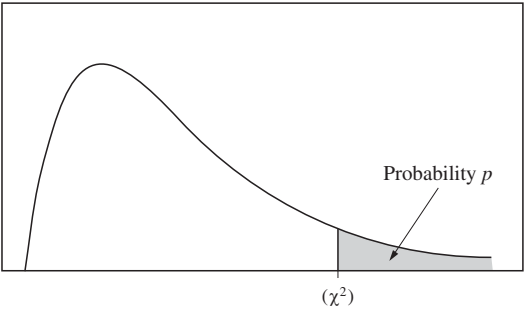


Table C χ^2 critical values

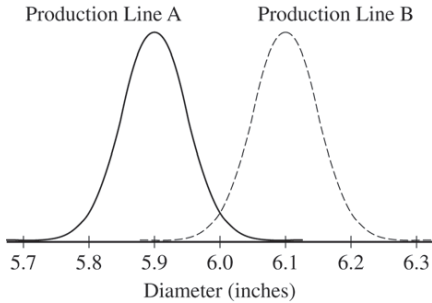
df	Tail probability p											
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83	12.12
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82	15.20
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27	17.73
4	5.39	5.99	6.74	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

STATISTICS
SECTION II
Part B
Question 6

Spend about 25 minutes on this part of the exam.
Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. Corn tortillas are made at a large facility that produces 100,000 tortillas per day on each of its two production lines. The distribution of the diameters of the tortillas produced on production line A is approximately normal with mean 5.9 inches, and the distribution of the diameters of the tortillas produced on production line B is approximately normal with mean 6.1 inches. The figure below shows the distributions of diameters for the two production lines.



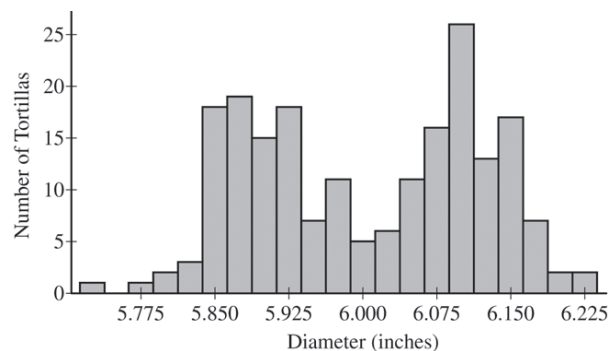
The tortillas produced at the factory are advertised as having a diameter of 6 inches. For the purpose of quality control, a sample of 200 tortillas is selected and the diameters are measured. From the sample of 200 tortillas, the manager of the facility wants to estimate the mean diameter, in inches, of the 200,000 tortillas produced on a given day. Two sampling methods have been proposed.

Method 1: Take a random sample of 200 tortillas from the 200,000 tortillas produced on a given day. Measure the diameter of each selected tortilla.

Method 2: Randomly select one of the two production lines on a given day. Take a random sample of 200 tortillas from the 100,000 tortillas produced by the selected production line. Measure the diameter of each selected tortilla.

- (a) Will a sample obtained using Method 2 be representative of the population of all tortillas made that day, with respect to the diameters of the tortillas? Explain why or why not.

- (b) The figure below is a histogram of 200 diameters obtained by using one of the two sampling methods described. Considering the shape of the histogram, explain which method, Method 1 or Method 2, was most likely used to obtain a such a sample.



- (c) Which of the two sampling methods, Method 1 or Method 2, will result in less variability in the diameters of the 200 tortillas in the sample on a given day? Explain.

Each day, the distribution of the 200,000 tortillas made that day has mean diameter 6 inches with standard deviation 0.11 inch.

- (d) For samples of size 200 taken from one day's production, describe the sampling distribution of the sample mean diameter for samples that are obtained using Method 1.
- (e) Suppose that one of the two sampling methods will be selected and used every day for one year (365 days). The sample mean of the 200 diameters will be recorded each day. Which of the two methods will result in less variability in the distribution of the 365 sample means? Explain.
- (f) A government inspector will visit the facility on June 22 to observe the sampling and to determine if the factory is in compliance with the advertised mean diameter of 6 inches. The manager knows that, with both sampling methods, the sample mean is an unbiased estimator of the population mean. However, the manager is unsure which method is more likely to produce a sample mean that is close to 6 inches on the day of sampling. Based on your previous answers, which of the two sampling methods, Method 1 or Method 2, is more likely to produce a sample mean close to 6 inches? Explain.

STOP

END OF EXAM

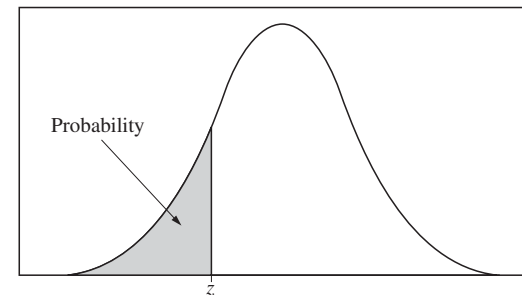


Table entry for z is the probability lying below z .

Table A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Name:

AP Statistics: 2015

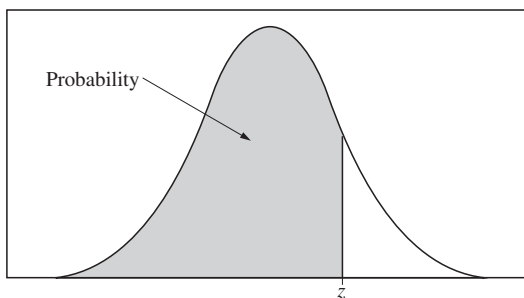


Table entry for z is the probability lying below z .

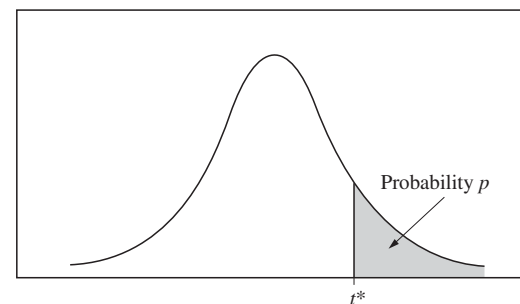
Table A (Continued)

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	.9995	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9997	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Name:

AP Statistics: 2015

Table entry for p and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^* .

Table B t distribution critical values

df	Tail probability p									
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807

50% 60% 70% 80% 90% 95% 96% 98% 99% 99.5% 99.8% 99.9%

Confidence level C

Table entry for p is the point (χ^2) with probability p lying above it.

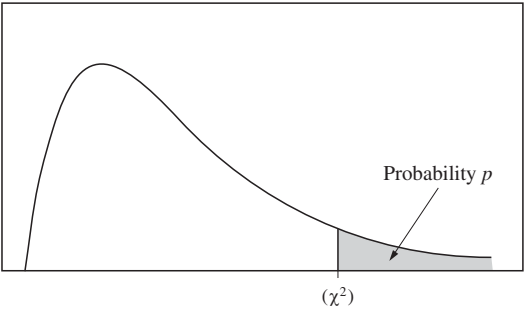


Table C χ^2 critical values

df	Tail probability p											
	.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

3.4 Potential Problems with Sampling

Name: _____ Class: _____ Date: _____ **Total: 8 marks**

KEY WORDS undercoverage 覆盖不足 • response bias 应答偏差 • voluntary response 自愿应答
• question wording bias 问题措辞偏差 • nonresponse 无应答

Objective

Build the skills to answer exam questions on **potential problems with sampling**.

You must be able to:

- identify **undercoverage** 覆盖不足 and **voluntary response** 自愿应答 as sources of bias
- describe how **nonresponse** 无应答 and **response bias** 应答偏差 distort results
- explain how poorly worded questions produce **question wording bias** 问题措辞偏差

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ Sources of sampling bias

- **Undercoverage** — part of the population is left out.
- **Voluntary response** — only those who choose to reply are counted (often the strongly opinionated).
- **Nonresponse** — selected people cannot be reached or refuse.
- **Response bias** — people give untrue answers (e.g. on sensitive topics).
- **Question wording bias** — a leading question pushes people toward an answer.

■ Spotting the bias

An online poll on a news site asks readers to click “agree” or “disagree.” This has **voluntary response** (only motivated readers click) and **undercoverage** (only that site’s visitors, not the whole population). Name the type from *who is left out* or *why the answers are distorted*.

2 Practice

2.1 Define undercoverage. [1]

2.2 Define nonresponse bias. [1]

2.3 State one source of bias that comes from how a question is asked. [1]

3 Exam-style questions

3.1 A survey in which people choose whether to respond suffers from [1]

- **A** undercoverage
 - **B** voluntary response bias
 - **C** randomization
 - **D** replication
-

3.2 When selected people cannot be reached or refuse to answer, it is [1]

- **A** nonresponse
 - **B** a placebo
 - **C** blocking
 - **D** an SRS
-

3.3 A survey uses a leading question that pushes people toward one answer.

(a) Name this type of bias. [1]

(b) Name a bias caused by leaving out part of the population. [1]

(c) Name a bias when only volunteers reply. [1]

Solutions

2.1 when part of the population is left out of the sampling process.

2.2 when selected individuals do not respond, distorting the results.

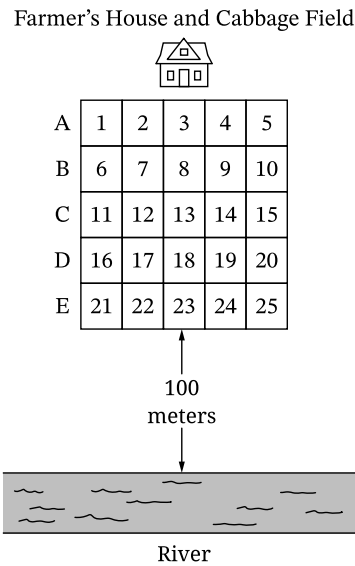
2.3 question wording bias (a leading or confusing question).

3.1 B.

3.2 A.

3.3 (a) question wording bias. (b) undercoverage. (c) voluntary response bias.

2. Aphids are tiny insects that feed on plants such as cabbage plants. A farmer wants to reduce the number of aphids in a cabbage field. A river is located 100 meters south of the cabbage field. The farmer divides the field into 25 regions of equal size, as shown in the diagram. Each region has approximately the same number of cabbage plants.



The farmer would like to estimate the proportion of cabbage plants in the field that are affected by aphids and believes that the extent of aphid damage is greater for the regions in the cabbage field closer to the river. To obtain the estimate, the farmer is considering three sampling methods.

- Sampling method I: Select region 3, which is closest to the farmer's house and farthest from the river. Examine every cabbage plant in the region for aphid damage.
- Sampling method II: Randomly select one row (A, B, C, D, or E). For every region in the selected row, examine every cabbage plant for aphid damage.
- Sampling method III: Randomly select one region from each of rows A, B, C, D, and E. For each selected region, examine every cabbage plant for aphid damage.

- A. Explain whether sampling method I is an appropriate sampling method for the farmer to use to estimate the proportion of cabbage plants in the field that are damaged by aphids.
- B. Using sampling method II, the farmer randomly selected row E and examined every cabbage plant in row E. If the farmer's belief is correct, determine whether the selection of row E is likely to provide an overestimate or an underestimate of the proportion of cabbage plants in the field that are damaged by aphids. Justify your answer.
- C. Using the information provided in the diagram of the cabbage field, describe how to implement sampling method III, which requires a random selection of one region from each of rows A, B, C, D, and E.

I. Descriptive Statistics

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{\sum x_i}{n}$$
$$\hat{y} = a + bx$$
$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$$s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$
$$\bar{y} = a + b\bar{x}$$
$$b = r \frac{s_y}{s_x}$$

II. Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 P(x_i)}$
If X has a binomial distribution with parameters n and p , then: $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$ where $x = 0, 1, 2, 3, \dots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1 - p)}$
If X has a geometric distribution with parameter p , then: $P(X = x) = (1 - p)^{x-1} p$ where $x = 1, 2, 3, \dots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1 - p}}{p}$

III. Sampling Distributions and Inferential Statistics

Standardized test statistic: $\frac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$
Confidence interval: $\text{statistic} \pm (\text{critical value})(\text{standard error of statistic})$
Chi-square statistic: $\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$

Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ where $\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}$

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\bar{X}$	$\mu_{\bar{X}} = \mu$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{X}} = \frac{s}{\sqrt{n}}$
For two populations: $\bar{X}_1 - \bar{X}_2$	$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$	$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Sampling distributions for simple linear regression:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For slope: b	$\mu_b = \beta$	$\sigma_b = \frac{\sigma}{\sigma_x \sqrt{n}}$, where $\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$	$s_b = \frac{s}{s_x \sqrt{n-1}}$, where $s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$ and $s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

*Standard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the standard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.



Descriptive Statistics

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{\sum x_i}{n}$$

$$\hat{y} = a + bx$$

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$$s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$\bar{y} = a + b\bar{x}$$

$$b = r \frac{s_y}{s_x}$$

Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 P(x_i)}$
If X has a binomial distribution with parameters n and p , then: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $x = 0, 1, 2, 3, \dots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1-p)}$
If X has a geometric distribution with parameter p , then: $P(X = x) = (1-p)^{x-1} p$ where $x = 1, 2, 3, \dots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1-p}}{p}$

Sampling Distributions and Inferential Statistics

Standardized test statistic: $\frac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$
Confidence interval: $\text{statistic} \pm (\text{critical value})(\text{standard error of statistic})$

Chi-square statistic: $\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$



Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ where $\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}$

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\bar{X}$	$\mu_{\bar{X}} = \mu$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{X}} = \frac{s}{\sqrt{n}}$
For two populations: $\bar{X}_1 - \bar{X}_2$	$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$	$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Sampling distributions for simple linear regression:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For slope: b	$\mu_b = \beta$	$\sigma_b = \frac{\sigma}{\sigma_x \sqrt{n}}$, where $\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$	$s_b = \frac{s}{s_x \sqrt{n-1}}$, where $s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$ and $s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

Standard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the standard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.



2. An environmental science teacher at a high school with a large population of students wanted to estimate the proportion of students at the school who regularly recycle plastic bottles. The teacher selected a random sample of students at the school to survey. Each selected student went into the teacher's office, one at a time, and was asked to respond yes or no to the following question.

Do you regularly recycle plastic bottles?

Based on the responses, a 95 percent confidence interval for the proportion of all students at the school who would respond yes to the question was calculated as (0.584, 0.816).

- How many students were in the sample selected by the environmental science teacher?
- Given the method used by the environmental science teacher to collect the responses, explain how bias might have been introduced and describe how the bias might affect the point estimate of the proportion of all students at the school who would respond yes to the question.
- The statistics teacher at the high school was concerned about the potential bias in the survey. To obtain a potentially less biased estimate of the proportion, the statistics teacher used an alternate method for collecting student responses. A random sample of 300 students was selected, and each student was given the following instructions on how to respond to the question.
 - In private, flip a fair coin.
 - If heads, you must respond no, regardless of whether you regularly recycle.
 - If tails, please truthfully respond yes or no.
- What is the expected number of students from the sample of 300 who would be required to respond no because the coin flip resulted in heads?
- The results of the sample showed that 213 of the 300 selected students responded no. Based on the results of the sample, give a point estimate for the proportion of all students at the high school who would respond yes to the question.

AP Statistics: 2016

3. Alzheimer's disease results in a loss of cognitive ability beyond what is expected with typical aging. A local newspaper published an article with the following headline.

Study Finds Strong Association Between Smoking and Alzheimer's

The article reported that a study tracked the medical histories of 21,123 men and women for 23 years. The article stated that, for those who smoked at least two packs of cigarettes a day, the risk of developing Alzheimer's disease was 2.57 times the risk for those who did not smoke.

- Identify the explanatory and response variables in the study.
 Explanatory variable:

 Response variable:
- Is the study described in the article an observational study or an experiment? Explain.
- Exercise status (regular weekly exercise versus no regular weekly exercise) was mentioned in the article as a possible confounding variable. Explain how exercise status could be a confounding variable in the study.



STATISTICS

SECTION II

Part B

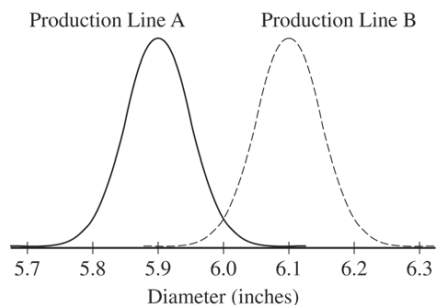
Question 6

Spend about 25 minutes on this part of the exam.

Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. Corn tortillas are made at a large facility that produces 100,000 tortillas per day on each of its two production lines. The distribution of the diameters of the tortillas produced on production line A is approximately normal with mean 5.9 inches, and the distribution of the diameters of the tortillas produced on production line B is approximately normal with mean 6.1 inches. The figure below shows the distributions of diameters for the two production lines.



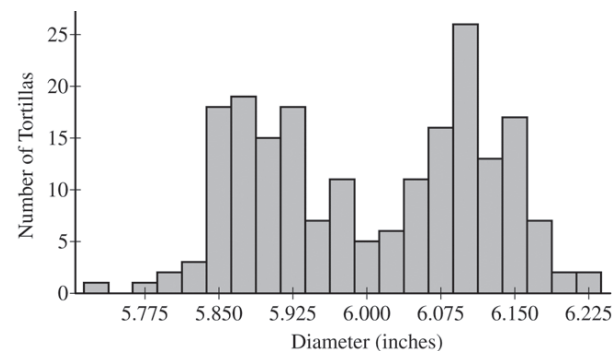
The tortillas produced at the factory are advertised as having a diameter of 6 inches. For the purpose of quality control, a sample of 200 tortillas is selected and the diameters are measured. From the sample of 200 tortillas, the manager of the facility wants to estimate the mean diameter, in inches, of the 200,000 tortillas produced on a given day. Two sampling methods have been proposed.

Method 1: Take a random sample of 200 tortillas from the 200,000 tortillas produced on a given day. Measure the diameter of each selected tortilla.

Method 2: Randomly select one of the two production lines on a given day. Take a random sample of 200 tortillas from the 100,000 tortillas produced by the selected production line. Measure the diameter of each selected tortilla.

- (a) Will a sample obtained using Method 2 be representative of the population of all tortillas made that day, with respect to the diameters of the tortillas? Explain why or why not.

- (b) The figure below is a histogram of 200 diameters obtained by using one of the two sampling methods described. Considering the shape of the histogram, explain which method, Method 1 or Method 2, was most likely used to obtain a such a sample.



- (c) Which of the two sampling methods, Method 1 or Method 2, will result in less variability in the diameters of the 200 tortillas in the sample on a given day? Explain.

Each day, the distribution of the 200,000 tortillas made that day has mean diameter 6 inches with standard deviation 0.11 inch.

- (d) For samples of size 200 taken from one day's production, describe the sampling distribution of the sample mean diameter for samples that are obtained using Method 1.
- (e) Suppose that one of the two sampling methods will be selected and used every day for one year (365 days). The sample mean of the 200 diameters will be recorded each day. Which of the two methods will result in less variability in the distribution of the 365 sample means? Explain.
- (f) A government inspector will visit the facility on June 22 to observe the sampling and to determine if the factory is in compliance with the advertised mean diameter of 6 inches. The manager knows that, with both sampling methods, the sample mean is an unbiased estimator of the population mean. However, the manager is unsure which method is more likely to produce a sample mean that is close to 6 inches on the day of sampling. Based on your previous answers, which of the two sampling methods, Method 1 or Method 2, is more likely to produce a sample mean close to 6 inches? Explain.

STOP

END OF EXAM

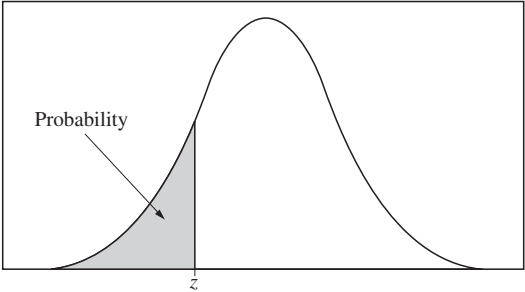


Table entry for z is the probability lying below z .

Table A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

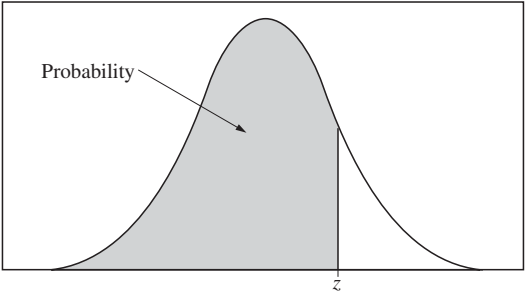


Table entry for z is the probability lying below z .

Table A (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 and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^* .

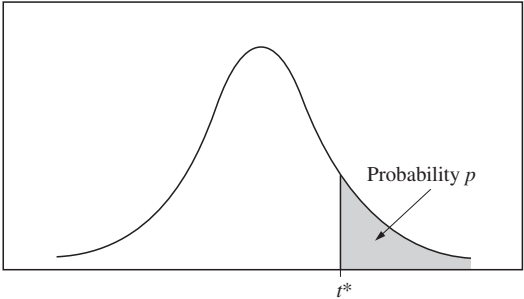


Table B t distribution critical values

df	Tail probability p											
	.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	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.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												

Table entry for p is the point (χ^2) with probability p lying above it.

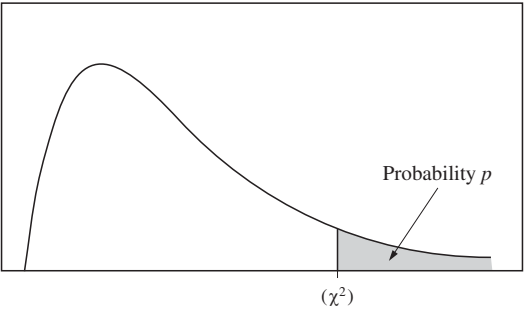


Table C χ^2 critical values

	Tail probability 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

4. As part of its twenty-fifth reunion celebration, the class of 1988 (students who graduated in 1988) at a state university held a reception on campus. In an informal survey, the director of alumni development asked 50 of the attendees about their incomes. The director computed the mean income of the 50 attendees to be \$189,952. In a news release, the director announced, "The members of our class of 1988 enjoyed resounding success. Last year's mean income of its members was \$189,952!"

- (a) What would be a statistical advantage of using the median of the reported incomes, rather than the mean, as the estimate of the typical income?

- (b) The director felt the members who attended the reception may be different from the class as a whole. A more detailed survey of the class was planned to find a better estimate of the income as well as other facts about the alumni. The staff developed two methods based on the available funds to carry out the survey.

Method 1: Send out an e-mail to all 6,826 members of the class asking them to complete an online form. The staff estimates that at least 600 members will respond.

Method 2: Select a simple random sample of members of the class and contact the selected members directly by phone. Follow up to ensure that all responses are obtained. Because method 2 will require more time than method 1, the staff estimates that only 100 members of the class could be contacted using method 2.

Which of the two methods would you select for estimating the average yearly income of all 6,826 members of the class of 1988? Explain your reasoning by comparing the two methods and the effect of each method on the estimate.

AP STATISTICS • EXERCISE SHEET

3.5 Introduction to Experimental Design

Name: _____ Class: _____ Date: _____ **Total: 9 marks**

KEY WORDS control 对照 • treatments 处理 • experimental units 实验单位
• control group 对照组 • placebo 安慰剂

Objective

Build the skills to answer exam questions on **experimental design**.

You must be able to:

- identify the **experimental units** 实验单位, **treatments** 处理, factors, and **response** variable
- explain the roles of **control** 对照, **randomization** 随机化, and **replication** 重复
- describe a **control group** 对照组 and a **placebo** 安慰剂

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ Parts of an experiment

- **Experimental units** — the who or what being treated.
- **Treatments** — the conditions applied (set by the **factor**, the explanatory variable).
- **Response** — the outcome measured.

■ Three principles

- **Control** — compare against a baseline group.
- **Randomization** — assign units to treatments by chance.
- **Replication** — use enough units to see a real effect.

A **control group** gets no treatment or a **placebo** (a fake treatment) for comparison.

2 Practice

2.1 Name the three principles of experimental design. [2]

2.2 State the purpose of a control group. [1]

2.3 Define a placebo. [1]

3 Exam-style questions

3.1 Assigning units to treatments by chance is [1]

- A control
- B randomization
- C replication
- D bias

3.2 A fake treatment used for comparison is a [1]

- A factor
- B placebo
- C block
- D response

3.3 A trial gives half the patients a real pill and half a sugar pill.

(a) Name the group given the sugar pill. [1]

(b) Name the sugar pill itself. [1]

(c) Name the principle of using enough patients. [1]

Solutions

2.1 control, randomization, replication.

2.2 to provide a baseline for comparison with the treated group(s).

2.3 a fake treatment given for comparison.

3.1 B.

3.2 B.

3.3 (a) the control group. (b) a placebo. (c) replication.

Name:

AP Statistics: 2026

2. Holly, a **botanist**, read an online report that stated that adding coffee grounds to the soil of rosebushes will produce more roses. Holly decides to conduct an experiment to investigate this claim. She grows 30 rosebushes in a greenhouse with a controlled environment so the conditions will be the same for all rosebushes.

When the rosebushes are a month old, she randomly assigns 15 rosebushes to have one-half cup of coffee grounds added to their soil weekly. The other 15 rosebushes will not have coffee grounds added to their soil.

After three months, she counts the number of roses on each rosebush.

botanist: a scientist who studies the biology of plants

Part A

Based on the information provided about Holly's experiment, identify each of the following.

- i. Treatments
- ii. Experimental units
- iii. Response variable

Part B

In the context of Holly's experiment, describe how the treatments can be randomly assigned to the experimental units so that each treatment has the same number of units.

Part C

After the experiment, Holly determines that her results were statistically significant at the $\alpha = 0.05$ level of significance. Explain the meaning of "statistically significant" in this context.

I. Descriptive Statistics

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{\sum x_i}{n}$$

$$\hat{y} = a + bx$$

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$$s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$\bar{y} = a + b\bar{x}$$

$$b = r \frac{s_y}{s_x}$$

II. Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 P(x_i)}$
If X has a binomial distribution with parameters n and p , then: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $x = 0, 1, 2, 3, \dots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1-p)}$
If X has a geometric distribution with parameter p , then: $P(X = x) = (1-p)^{x-1} p$ where $x = 1, 2, 3, \dots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \sqrt{\frac{1-p}{p^2}}$

III. Sampling Distributions and Inferential Statistics

$$\text{Standardized test statistic: } \frac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$$

$$\text{Confidence interval: } \text{statistic} \pm (\text{critical value})(\text{standard error of statistic})$$

$$\text{Chi-square statistic: } \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$ where $\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}$

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\bar{X}$	$\mu_{\bar{X}} = \mu$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{X}} = \frac{s}{\sqrt{n}}$
For two populations: $\bar{X}_1 - \bar{X}_2$	$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$	$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Sampling distributions for simple linear regression:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For slope: b	$\mu_b = \beta$	$\sigma_b = \frac{\sigma}{\sigma_x \sqrt{n}}$, where $\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$	$s_b = \frac{s}{s_x \sqrt{n-1}}$, where $s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$ and $s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

*Standard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the standard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.

Descriptive Statistics

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{\sum x_i}{n}$$
$$\hat{y} = a + bx$$
$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$$s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$
$$\bar{y} = a + b\bar{x}$$
$$b = r \frac{s_y}{s_x}$$

Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 P(x_i)}$
If X has a binomial distribution with parameters n and p , then: $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$ where $x = 0, 1, 2, 3, \dots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1 - p)}$
If X has a geometric distribution with parameter p , then: $P(X = x) = (1 - p)^{x-1} p$ where $x = 1, 2, 3, \dots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1 - p}}{p}$

Descriptive Statistics

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{\sum x_i}{n}$$
$$\hat{y} = a + bx$$
$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$$s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$
$$\bar{y} = a + b\bar{x}$$
$$b = r \frac{s_y}{s_x}$$

Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 P(x_i)}$
If X has a binomial distribution with parameters n and p , then: $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$ where $x = 0, 1, 2, 3, \dots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1 - p)}$
If X has a geometric distribution with parameter p , then: $P(X = x) = (1 - p)^{x-1} p$ where $x = 1, 2, 3, \dots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1 - p}}{p}$

Sampling Distributions and Inferential Statistics

Standardized test statistic: $\frac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$
Confidence interval: $\text{statistic} \pm (\text{critical value})(\text{standard error of statistic})$

$$\text{Chi-square statistic: } \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ where $\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}$

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\bar{X}$	$\mu_{\bar{X}} = \mu$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{X}} = \frac{s}{\sqrt{n}}$
For two populations: $\bar{X}_1 - \bar{X}_2$	$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$	$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Sampling distributions for simple linear regression:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For slope: b	$\mu_b = \beta$	$\sigma_b = \frac{\sigma}{\sigma_x \sqrt{n}}$, where $\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$	$s_b = \frac{s}{s_x \sqrt{n-1}}$, where $s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$ and $s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

Standard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the standard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.

2. Researchers are investigating the effectiveness of using a fungus to control the spread of an insect that destroys trees. The researchers will create four different concentrations of fungus mixtures: 0 milliliters per liter (ml/L), 1.25 ml/L, 2.5 ml/L, and 3.75 ml/L. An equal number of the insects will be placed into 20 individual containers. The group of insects in each container will be sprayed with one of the four mixtures, and the researchers will record the number of insects that are still alive in each container one week after spraying.

- (a) Identify the treatments, experimental units, and response variable of the experiment.

Treatments:

Experimental units:

Response variable:

- (b) Does the experiment have a control group? Explain your answer.
- (c) Describe how the treatments can be randomly assigned to the experimental units so that each treatment has the same number of units.

STATISTICS
SECTION II
Part B
Question 6

Spend about 25 minutes on this part of the exam.
Percent of Section II score—25

- Directions:** Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.
6. Consider an experiment in which two men and two women will be randomly assigned to either a treatment group or a control group in such a way that each group has two people. The people are identified as Man 1, Man 2, Woman 1, and Woman 2. The six possible arrangements are shown below.

Arrangement A		Arrangement B		Arrangement C	
Treatment	Control	Treatment	Control	Treatment	Control
Man 1	Woman 1	Man 1	Man 2	Man 1	Man 2
Man 2	Woman 2	Woman 1	Woman 2	Woman 2	Woman 1

Arrangement D		Arrangement E		Arrangement F	
Treatment	Control	Treatment	Control	Treatment	Control
Woman 1	Man 1	Man 2	Man 1	Man 2	Man 1
Woman 2	Man 2	Woman 2	Woman 1	Woman 1	Woman 2

- Two possible methods of assignment are being considered: the sequential coin flip method, as described in part (a), and the chip method, as described in part (b). For each method, the order of the assignment will be Man 1, Man 2, Woman 1, Woman 2.
- (a) For the sequential coin flip method, a fair coin is flipped until one group has two people. An outcome of tails assigns the person to the treatment group, and an outcome of heads assigns the person to the control group. As soon as one group has two people, the remaining people are automatically assigned to the other group.
- (i) Complete the table below by calculating the probability of each arrangement occurring if the sequential coin flip method is used.

Arrangement	A	B	C	D	E	F
Probability						

- (ii) For the sequential coin flip method, what is the probability that Man 1 and Man 2 are assigned to the same group?



The six arrangements are repeated below.

Arrangement A		Arrangement B		Arrangement C	
Treatment	Control	Treatment	Control	Treatment	Control
Man 1	Woman 1	Man 1	Man 2	Man 1	Man 2
Man 2	Woman 2	Woman 1	Woman 2	Woman 2	Woman 1

Arrangement D		Arrangement E		Arrangement F	
Treatment	Control	Treatment	Control	Treatment	Control
Woman 1	Man 1	Man 2	Man 1	Man 2	Man 1
Woman 2	Man 2	Woman 2	Woman 1	Woman 1	Woman 2

- (b) For the chip method, two chips are marked “treatment” and two chips are marked “control.” Each person selects one chip at random without replacement.
- (i) Complete the table below by calculating the probability of each arrangement occurring if the chip method is used.

Arrangement	A	B	C	D	E	F
Probability						

- (ii) For the chip method, what is the probability that Man 1 and Man 2 are assigned to the same group?
- (c) Sixteen participants consisting of 10 students and 6 teachers at an elementary school will be used for an experiment to determine lunch preference for the school population of students and teachers. As the participants enter the school cafeteria for lunch, they will be randomly assigned to receive one of two lunches so that 8 will receive a salad, and 8 will receive a grilled cheese sandwich. The students will enter the cafeteria first, and the teachers will enter next. Which method, the sequential coin flip method or the chip method, should be used to assign the treatments? Justify your choice.

STOP
END OF EXAM



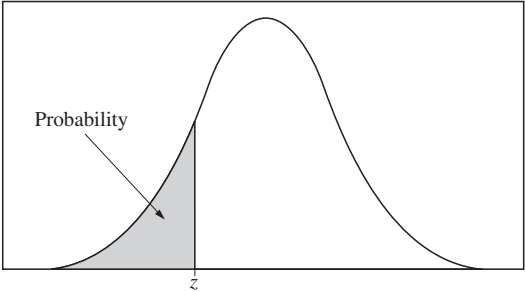


Table entry for z is the probability lying below z .

Table A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

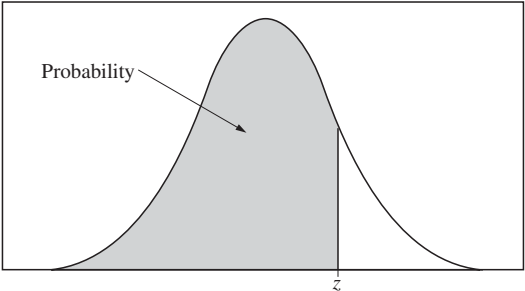
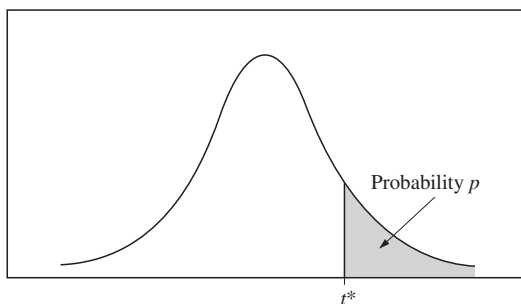


Table entry for z is the probability lying below z .

Table A (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 and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^* .

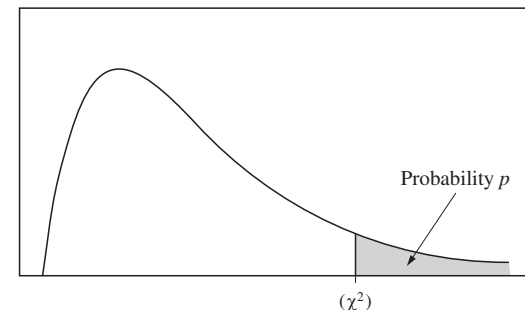
Table B t distribution critical values

	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	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.619
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.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%

50% 60% 70% 80% 90% 95% 96% 98% 99% 99.5% 99.8% 99.9%

Confidence level C

Table entry for p is the point (χ^2) with probability p lying above it.

Table C χ^2 critical values

	Tail probability 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

3. Alzheimer’s disease results in a loss of cognitive ability beyond what is expected with typical aging. A local newspaper published an article with the following headline.

Study Finds Strong Association Between Smoking and Alzheimer’s

The article reported that a study tracked the medical histories of 21,123 men and women for 23 years. The article stated that, for those who smoked at least two packs of cigarettes a day, the risk of developing Alzheimer’s disease was 2.57 times the risk for those who did not smoke.

(a) Identify the explanatory and response variables in the study.

Explanatory variable:

Response variable:

- (b) Is the study described in the article an observational study or an experiment? Explain.
- (c) Exercise status (regular weekly exercise versus no regular weekly exercise) was mentioned in the article as a possible confounding variable. Explain how exercise status could be a confounding variable in the study.

3.6 Selecting an Experimental Design

Name: _____ Class: _____ Date: _____

Total: 8 marks

KEY WORDS matched-pairs design 配对设计 • completely randomized design 完全随机设计
• blocking 区组 • experiment 实验 • treatment 处理

Objective

Build the skills to answer exam questions on **selecting an experimental design**.

You must be able to:

- describe a **completely randomized design** 完全随机设计
- explain how **blocking** 区组 reduces variability
- describe a **randomized block design** and a **matched-pairs design** 配对设计

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ **Completely randomized design**

All experimental units are assigned to the treatments **entirely at random**.

■ **Blocking**

Blocking groups similar units together (e.g. by age), then randomizes **within** each block. This **reduces variability** from that known factor, making treatment differences clearer.

■ **Matched pairs**

A **matched-pairs design** is a special case: two similar units per block (or the same unit measured twice, before and after).

2 Practice

2.1 Describe a completely randomized design. [1]

2.2 State the purpose of blocking. [1]

2.3 State what a matched-pairs design is. [1]

3 Exam-style questions

3.1 Blocking is used to [1]

- **A** add bias
- **B** reduce variability from a known factor
- **C** avoid randomization
- **D** increase error

3.2 A design that randomizes within groups of similar units is a [1]

- **A** completely randomized design
- **B** randomized block design
- **C** census
- **D** voluntary response survey

3.3 Researchers block by age group before assigning treatments.

(a) Name this design. [1]

(b) State why they block by age. [1]

(c) Name a design that pairs two similar units. [1]

Solutions

2.1 all units are assigned to the treatments entirely at random.

2.2 to reduce variability from a known factor, so treatment effects are clearer.

2.3 two similar units (or the same unit twice) form each pair, split between treatments.

3.1 B.

3.2 B.

3.3 (a) a randomized block design. (b) to reduce variability caused by age. (c) a matched-pairs design.

2. Holly, a **botanist**, read an online report that stated that adding coffee grounds to the soil of rosebushes will produce more roses. Holly decides to conduct an experiment to investigate this claim. She grows 30 rosebushes in a greenhouse with a controlled environment so the conditions will be the same for all rosebushes.

When the rosebushes are a month old, she randomly assigns 15 rosebushes to have one-half cup of coffee grounds added to their soil weekly. The other 15 rosebushes will not have coffee grounds added to their soil.

After three months, she counts the number of roses on each rosebush.

botanist: a scientist who studies the biology of plants

Part A

Based on the information provided about Holly’s experiment, identify each of the following.

- i. Treatments
- ii. Experimental units
- iii. Response variable

Part B

In the context of Holly’s experiment, describe how the treatments can be randomly assigned to the experimental units so that each treatment has the same number of units.

Part C

After the experiment, Holly determines that her results were statistically significant at the $\alpha = 0.05$ level of significance. Explain the meaning of “statistically significant” in this context.

Descriptive Statistics

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{\sum x_i}{n}$$

$$\hat{y} = a + bx$$

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$$s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$\bar{y} = a + b\bar{x}$$

$$b = r \frac{s_y}{s_x}$$

Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 P(x_i)}$
If X has a binomial distribution with parameters n and p , then: $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$ where $x = 0, 1, 2, 3, \dots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1 - p)}$
If X has a geometric distribution with parameter p , then: $P(X = x) = (1 - p)^{x-1} p$ where $x = 1, 2, 3, \dots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1 - p}}{p}$

Descriptive Statistics

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{\sum x_i}{n}$$

$$\hat{y} = a + bx$$

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$$s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$\bar{y} = a + b\bar{x}$$

$$b = r \frac{s_y}{s_x}$$

Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 P(x_i)}$
If X has a binomial distribution with parameters n and p , then: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $x = 0, 1, 2, 3, \dots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1-p)}$
If X has a geometric distribution with parameter p , then: $P(X = x) = (1-p)^{x-1} p$ where $x = 1, 2, 3, \dots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1-p}}{p}$

Sampling Distributions and Inferential Statistics

Standardized test statistic: $\frac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$
Confidence interval: $\text{statistic} \pm (\text{critical value})(\text{standard error of statistic})$

$$\text{Chi-square statistic: } \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$



Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$ where $\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}$

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\bar{X}$	$\mu_{\bar{X}} = \mu$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{X}} = \frac{s}{\sqrt{n}}$
For two populations: $\bar{X}_1 - \bar{X}_2$	$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$	$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Sampling distributions for simple linear regression:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For slope: b	$\mu_b = \beta$	$\sigma_b = \frac{\sigma}{s_x \sqrt{n}}$, where $\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$	$s_b = \frac{s}{s_x \sqrt{n-1}}$, where $s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$ and $s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

Standard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the standard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.



2. Researchers are investigating the effectiveness of using a fungus to control the spread of an insect that destroys trees. The researchers will create four different concentrations of fungus mixtures: 0 milliliters per liter (ml/L), 1.25 ml/L, 2.5 ml/L, and 3.75 ml/L. An equal number of the insects will be placed into 20 individual containers. The group of insects in each container will be sprayed with one of the four mixtures, and the researchers will record the number of insects that are still alive in each container one week after spraying.

(a) Identify the treatments, experimental units, and response variable of the experiment.

Treatments:

Experimental units:

Response variable:

(b) Does the experiment have a control group? Explain your answer.

(c) Describe how the treatments can be randomly assigned to the experimental units so that each treatment has the same number of units.



STATISTICS
SECTION II

Part B

Question 6

Spend about 25 minutes on this part of the exam.
Percent of Section II score—25

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

6. Consider an experiment in which two men and two women will be randomly assigned to either a treatment group or a control group in such a way that each group has two people. The people are identified as Man 1, Man 2, Woman 1, and Woman 2. The six possible arrangements are shown below.

Arrangement A	
Treatment	Control
Man 1	Woman 1
Man 2	Woman 2

Arrangement B	
Treatment	Control
Man 1	Man 2
Woman 1	Woman 2

Arrangement C	
Treatment	Control
Man 1	Man 2
Woman 2	Woman 1

Arrangement D	
Treatment	Control
Woman 1	Man 1
Woman 2	Man 2

Arrangement E	
Treatment	Control
Man 2	Man 1
Woman 2	Woman 1

Arrangement F	
Treatment	Control
Man 2	Man 1
Woman 1	Woman 2

Two possible methods of assignment are being considered: the sequential coin flip method, as described in part (a), and the chip method, as described in part (b). For each method, the order of the assignment will be Man 1, Man 2, Woman 1, Woman 2.

(a) For the sequential coin flip method, a fair coin is flipped until one group has two people. An outcome of tails assigns the person to the treatment group, and an outcome of heads assigns the person to the control group. As soon as one group has two people, the remaining people are automatically assigned to the other group.

(i) Complete the table below by calculating the probability of each arrangement occurring if the sequential coin flip method is used.

Arrangement	A	B	C	D	E	F
Probability						

(ii) For the sequential coin flip method, what is the probability that Man 1 and Man 2 are assigned to the same group?



The six arrangements are repeated below.

Arrangement A		Arrangement B		Arrangement C	
Treatment	Control	Treatment	Control	Treatment	Control
Man 1	Woman 1	Man 1	Man 2	Man 1	Man 2
Man 2	Woman 2	Woman 1	Woman 2	Woman 2	Woman 1

Arrangement D		Arrangement E		Arrangement F	
Treatment	Control	Treatment	Control	Treatment	Control
Woman 1	Man 1	Man 2	Man 1	Man 2	Man 1
Woman 2	Man 2	Woman 2	Woman 1	Woman 1	Woman 2

- (b) For the chip method, two chips are marked “treatment” and two chips are marked “control.” Each person selects one chip at random without replacement.
- (i) Complete the table below by calculating the probability of each arrangement occurring if the chip method is used.

Arrangement	A	B	C	D	E	F
Probability						

- (ii) For the chip method, what is the probability that Man 1 and Man 2 are assigned to the same group?
- (c) Sixteen participants consisting of 10 students and 6 teachers at an elementary school will be used for an experiment to determine lunch preference for the school population of students and teachers. As the participants enter the school cafeteria for lunch, they will be randomly assigned to receive one of two lunches so that 8 will receive a salad, and 8 will receive a grilled cheese sandwich. The students will enter the cafeteria first, and the teachers will enter next. Which method, the sequential coin flip method or the chip method, should be used to assign the treatments? Justify your choice.

STOP
END OF EXAM

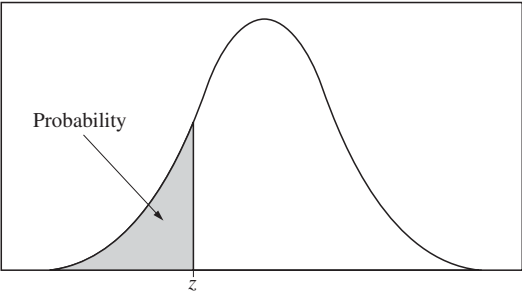


Table entry for z is the probability lying below z .

Table A Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

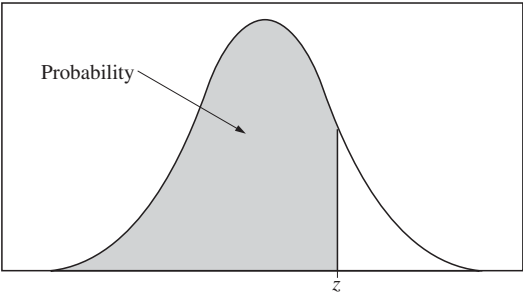


Table entry for z is the probability lying below z .

Table A (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	.9995	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9997	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Table entry for p and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^* .

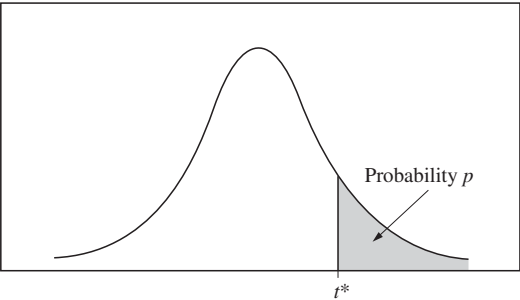


Table B t distribution critical values

	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	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
∞	.674	.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												

Table entry for p is the point (χ^2) with probability p lying above it.

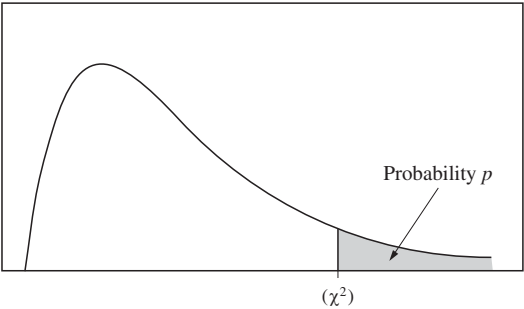


Table C χ^2 critical values

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

3.7 Inference and Experiments

Name: _____ Class: _____ Date: _____ **Total: 8 marks**

KEY WORDS statistically significant 统计显著 • random assignment 随机分配
• random selection 随机选取 • experiment 实验 • population 总体

Objective

Build the skills to answer exam questions on **inference and experiments**.

You must be able to:

- explain that **random assignment** 随机分配 allows a **cause-and-effect** conclusion
- explain that **random selection** 随机选取 allows **generalization** to a population
- describe when a difference is **statistically significant** 统计显著

1 Worked examples

Study these first. Each one shows the method for a question type used later.

■ Two kinds of randomness

- **Random assignment** of units to treatments lets you conclude **cause and effect**.
- **Random selection** of units from a population lets you **generalize** the results to that population.

■ Statistical significance

A difference between groups is **statistically significant** when it is **too large to be explained by chance alone**.

2 Practice

2.1 State what random assignment allows you to conclude. [1]

2.2 State what random selection allows. [1]

2.3 Define a statistically significant difference. [1]

3 Exam-style questions

3.1 Random assignment to treatments allows a conclusion about [1]

- **A** generalization
 - **B** cause and effect
 - **C** the population size
 - **D** bias
-

3.2 A result is statistically significant if it is [1]

- **A** small
 - **B** unlikely to be due to chance alone
 - **C** biased
 - **D** perfectly random
-

3.3 An experiment randomly assigns treatments and finds a large difference between the groups.

(a) State what random assignment allows. [1]

(b) State what random selection of units would allow. [1]

(c) State what “statistically significant” means. [1]

Solutions

2.1 a cause-and-effect relationship.

2.2 generalizing the results to the population.

2.3 a difference too large to be explained by chance alone.

3.1 B.

3.2 B.

3.3 (a) a cause-and-effect conclusion. (b) generalizing to the population. (c) the difference is unlikely to be due to chance alone.

2. Holly, a **botanist**, read an online report that stated that adding coffee grounds to the soil of rosebushes will produce more roses. Holly decides to conduct an experiment to investigate this claim. She grows 30 rosebushes in a greenhouse with a controlled environment so the conditions will be the same for all rosebushes.

When the rosebushes are a month old, she randomly assigns 15 rosebushes to have one-half cup of coffee grounds added to their soil weekly. The other 15 rosebushes will not have coffee grounds added to their soil.

After three months, she counts the number of roses on each rosebush.

botanist: a scientist who studies the biology of plants

Part A

Based on the information provided about Holly’s experiment, identify each of the following.

- i. Treatments
- ii. Experimental units
- iii. Response variable

Part B

In the context of Holly’s experiment, describe how the treatments can be randomly assigned to the experimental units so that each treatment has the same number of units.

Part C

After the experiment, Holly determines that her results were statistically significant at the $\alpha = 0.05$ level of significance. Explain the meaning of “statistically significant” in this context.

I. Descriptive Statistics

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{\sum x_i}{n}$$

$$\hat{y} = a + bx$$

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$$s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$\bar{y} = a + b\bar{x}$$

$$b = r \frac{s_y}{s_x}$$

II. Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 P(x_i)}$
If X has a binomial distribution with parameters n and p , then: $P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$ where $x = 0, 1, 2, 3, \dots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1 - p)}$
If X has a geometric distribution with parameter p , then: $P(X = x) = (1 - p)^{x-1} p$ where $x = 1, 2, 3, \dots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1 - p}}{p}$

III. Sampling Distributions and Inferential Statistics

Standardized test statistic: $\frac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$
Confidence interval: $\text{statistic} \pm (\text{critical value})(\text{standard error of statistic})$

Chi-square statistic: $\chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$

Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ where $\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}$

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\bar{X}$	$\mu_{\bar{X}} = \mu$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{X}} = \frac{s}{\sqrt{n}}$
For two populations: $\bar{X}_1 - \bar{X}_2$	$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$	$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Sampling distributions for simple linear regression:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For slope: b	$\mu_b = \beta$	$\sigma_b = \frac{\sigma}{\sigma_x \sqrt{n}}$, where $\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$	$s_b = \frac{s}{s_x \sqrt{n-1}}$, where $s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$ and $s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

*Standard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the standard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.

Descriptive Statistics

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{\sum x_i}{n}$$

$$\hat{y} = a + bx$$

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$$s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$\bar{y} = a + b\bar{x}$$

$$b = r \frac{s_y}{s_x}$$

Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 P(x_i)}$
If X has a binomial distribution with parameters n and p , then: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $x = 0, 1, 2, 3, \dots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1-p)}$
If X has a geometric distribution with parameter p , then: $P(X = x) = (1-p)^{x-1} p$ where $x = 1, 2, 3, \dots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1-p}}{p}$

STATISTICS

SECTION II, Part B

Suggested Time—25 minutes

1 Question

Directions: Show all your work. Indicate clearly the methods you use, because you will be scored on the correctness of your methods as well as on the accuracy and completeness of your results and explanations.

To compare success rates for treating allergies at two clinics that specialize in treating allergy sufferers, researchers selected random samples of patient records from the two clinics. The following table summarizes the data.

	Clinic A	Clinic B	Total
Unsuccessful treatment	51	33	84
Successful treatment	88	35	123
Total	139	68	207

- (a) (i) Complete the following table by recording the relative frequencies of successful and unsuccessful treatments at each clinic.

	Clinic A	Clinic B
Unsuccessful treatment		
Successful treatment		

- (ii) Based on the relative frequency table in part (a-i), which clinic is more successful in treating allergy sufferers? Justify your answer.

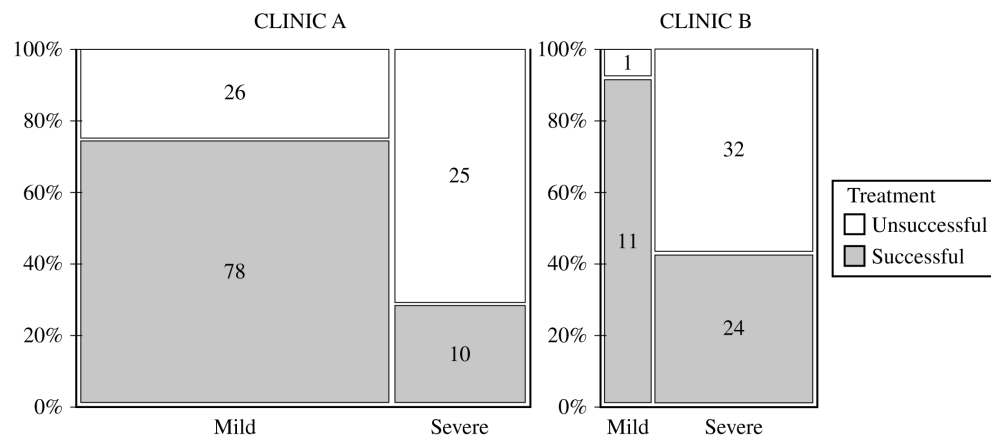
GO ON TO THE NEXT PAGE.

Continue your response to **QUESTION 6** on this page.

- (b) Based on the design of the study, would a statistically significant result allow the researchers to conclude that receiving treatments at the clinic you selected in part (a-ii) causes a higher percentage of successful treatments than at the other clinic? Explain your answer.

GO ON TO THE NEXT PAGE.

A physician who worked at both clinics believed that it was important to separate the patients in the study by severity of the patient's allergy (severe or mild). The physician constructed the following mosaic plot. The values in the mosaic plot represent the number of patients who were either successfully treated or unsuccessfully treated in each allergy severity group within each clinic. For example, the value 78 represents the number of patients successfully treated in the mild group within Clinic A.



Based on the mosaic plot, the physician concluded the following:

For mild allergy sufferers, **Clinic B** was more successful in treating allergies.

For severe allergy sufferers, **Clinic B** was more successful in treating allergies.

(c) (i) For each clinic, which allergy severity is treated more successfully? Justify your answer.

• Clinic A:

• Clinic B:

GO ON TO THE NEXT PAGE.

(ii) For each clinic, which allergy severity is more likely to be treated? Justify your answer.

• Clinic A:

• Clinic B:

(d) Using your answers from part (c), give a reasonable explanation of why the more successful clinic identified in part (a-ii) is the same as or different from the physician's conclusion that Clinic B is more successful in treating both severe and mild allergies.

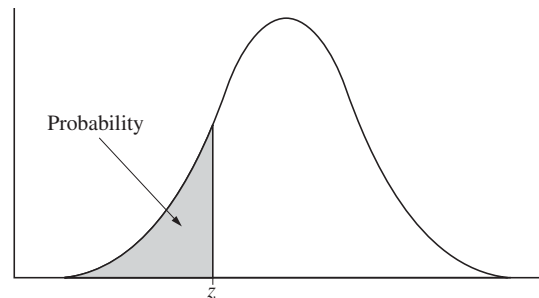
GO ON TO THE NEXT PAGE.

Name:

AP Statistics: 2022

STOP**END OF EXAM**Name:

AP Statistics: 2022

Table entry for z is the probability lying below z .**Table A** Standard normal probabilities

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
-3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
-3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
-3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
-3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
-3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
-2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
-2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
-2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
-2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
-2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
-2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
-2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
-2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
-2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
-2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
-1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
-1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
-1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
-1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
-1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
-1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
-1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
-1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
-1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
-1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
-0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
-0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
-0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
-0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
-0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
-0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
-0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
-0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
-0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
-0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641

Name:

AP Statistics: 2022

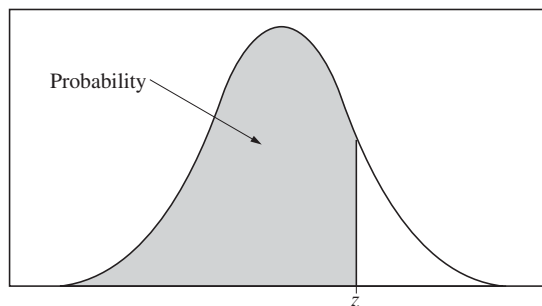


Table entry for z is the probability lying below z .

Table A (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

Name:

AP Statistics: 2022

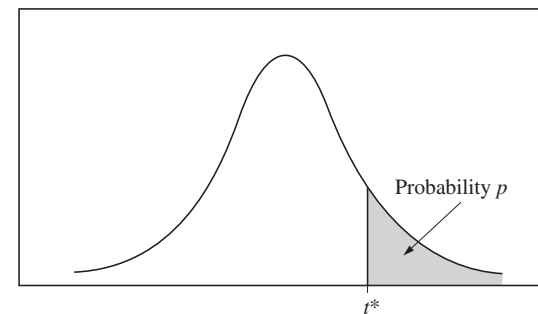
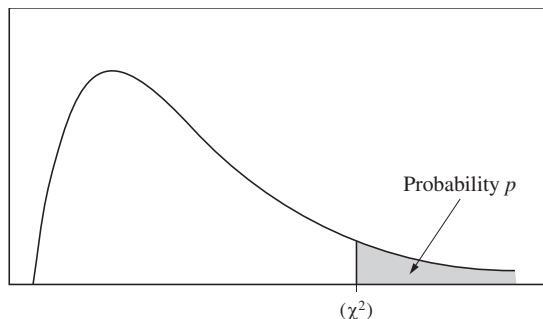


Table entry for p and C is the point t^* with probability p lying above it and probability C lying between $-t^*$ and t^* .

Table B t distribution critical values

df	Tail probability p									
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813
∞	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807
Confidence level C										
50% 60% 70% 80% 90% 95% 96% 98% 99% 99.5% 99.8% 99.9%										

Table entry for p is the point (χ^2) with probability p lying above it.

Table C χ^2 critical values

df	Tail probability p											
	.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

Descriptive Statistics

$$\bar{x} = \frac{1}{n} \sum x_i = \frac{\sum x_i}{n}$$

$$\hat{y} = a + bx$$

$$r = \frac{1}{n-1} \sum \left(\frac{x_i - \bar{x}}{s_x} \right) \left(\frac{y_i - \bar{y}}{s_y} \right)$$

$$s_x = \sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

$$\bar{y} = a + b\bar{x}$$

$$b = r \frac{s_y}{s_x}$$

Probability and Distributions

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Probability Distribution	Mean	Standard Deviation
Discrete random variable, X	$\mu_X = E(X) = \sum x_i P(x_i)$	$\sigma_X = \sqrt{\sum (x_i - \mu_X)^2 P(x_i)}$
If X has a binomial distribution with parameters n and p , then: $P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$ where $x = 0, 1, 2, 3, \dots, n$	$\mu_X = np$	$\sigma_X = \sqrt{np(1-p)}$
If X has a geometric distribution with parameter p , then: $P(X = x) = (1-p)^{x-1} p$ where $x = 1, 2, 3, \dots$	$\mu_X = \frac{1}{p}$	$\sigma_X = \frac{\sqrt{1-p}}{p}$

Sampling Distributions and Inferential Statistics

$$\text{Standardized test statistic: } \frac{\text{statistic} - \text{parameter}}{\text{standard error of the statistic}}$$

$$\text{Confidence interval: } \text{statistic} \pm (\text{critical value})(\text{standard error of statistic})$$

$$\text{Chi-square statistic: } \chi^2 = \sum \frac{(\text{observed} - \text{expected})^2}{\text{expected}}$$

Sampling distributions for proportions:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\hat{p}$	$\mu_{\hat{p}} = p$	$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$	$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
For two populations: $\hat{p}_1 - \hat{p}_2$	$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$	$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$	$s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ When $p_1 = p_2$ is assumed: $s_{\hat{p}_1 - \hat{p}_2} = \sqrt{\hat{p}_c(1-\hat{p}_c)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ where $\hat{p}_c = \frac{X_1 + X_2}{n_1 + n_2}$

Sampling distributions for means:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For one population: $\bar{X}$	$\mu_{\bar{X}} = \mu$	$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$	$s_{\bar{X}} = \frac{s}{\sqrt{n}}$
For two populations: $\bar{X}_1 - \bar{X}_2$	$\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$	$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	$s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Sampling distributions for simple linear regression:

Random Variable	Parameters of Sampling Distribution		Standard Error* of Sample Statistic
For slope: b	$\mu_b = \beta$	$\sigma_b = \frac{\sigma}{\sigma_x \sqrt{n}}$ where $\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$	$s_b = \frac{s}{s_x \sqrt{n-1}},$ where $s = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n-2}}$ and $s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$

Standard deviation is a measurement of variability from the theoretical population. Standard error is the estimate of the standard deviation. If the standard deviation of the statistic is assumed to be known, then the standard deviation should be used instead of the standard error.

4. The anterior cruciate ligament (ACL) is one of the ligaments that help stabilize the knee. Surgery is often recommended if the ACL is completely torn, and recovery time from the surgery can be lengthy. A medical center developed a new surgical procedure designed to reduce the average recovery time from the surgery. To test the effectiveness of the new procedure, a study was conducted in which 210 patients needing surgery to repair a torn ACL were randomly assigned to receive either the standard procedure or the new procedure.

- (a) Based on the design of the study, would a statistically significant result allow the medical center to conclude that the new procedure causes a reduction in recovery time compared to the standard procedure, for patients similar to those in the study? Explain your answer.
- (b) Summary statistics on the recovery times from the surgery are shown in the table.

Type of Procedure	Sample Size	Mean Recovery Time (days)	Standard Deviation Recovery Time (days)
Standard	110	217	34
New	100	186	29

Do the data provide convincing statistical evidence that those who receive the new procedure will have less recovery time from the surgery, on average, than those who receive the standard procedure, for patients similar to those in the study?