The normal and t-tables are attached at the end of the question paper. If the tables are not detailed enough, please use the average of the two nearest values.

Part A (For all sections A, B, and C)

1. (10 marks) Suppose that you are going to test a null hypothesis about the population mean μ based on a sample mean $\bar{x}$. For each of the following cases, decide the appropriate test statistic and state whether the statistic has the normal distribution, the t distribution or both:

   (a) (2 marks) a small sample ($n < 30$) from a **normal** population with **known** standard deviation.

   (b) (2 marks) a small sample ($n < 30$) from a **normal** population with **unknown** standard deviation.

   (c) (2 marks) a large sample ($n \geq 30$) from a **normal** population with **known** standard deviation.

   (d) (2 marks) a large sample ($n \geq 30$) from a **normal** population with **unknown** standard deviation.

   (e) (2 marks) a large sample ($n \geq 30$) from a **nonnormal** population with **unknown** standard deviation.

2. (10 marks) Suppose a public opinion is split 65% against and 35% in favour for increasing taxes to help balance the federal budget. 500 people from the population are selected randomly and interviewed.

   (a) (4 marks) Completely describe the distribution of the sample proportion $\bar{p}$ of people who are in favour of increasing taxes to help balance the federal budget, find the mean and standard deviation of the sampling proportion $\bar{p}$. Which classic theorem is fundamental to this assertion of the distribution of the sample proportion $\bar{p}$?

   (b) (3 marks) What is the probability that the sample proportion $\bar{p}$ favouring a tax increase is more than 30%?

   (c) (3 marks) Within what interval, 99% of the sample proportions $\bar{p}$ would lie?

3. (10 marks) The editor of a textbook publishing company is trying to decide whether or not to publish a proposed business statistics textbook. Information on such texts published in the past indicates that 10% were huge successes, 20% were modest successes, 40% break even and 30% are losers. Before the decision is made, the book will be reviewed. In the past 99% of the huge successes received favorable reviews, 70% of the moderate successes received favorable reviews, 40% of the break even books received favorable reviews and 20% of the losers received favorable reviews.

   (a) (4 marks) What is the probability of receiving a favorable review?

   (b) (2 marks) What is the probability of an unfavorable review?

   (c) (4 marks) If the textbook receives a favorable review, what is the probability that it will be a huge success? A moderate success? A break even? A loser?

Part B (For section A only)

4. (10 marks) Student A in our class ECON 2201 tried to estimate the average marks of the midterm held in November, 2013. Based on a random sample of 30 students' marks and with a 95% confidence level, the student arrived at an interval estimate for the average marks of between 50 and 80.

   (a) (5 marks) After receiving this result, student B in the same class claimed that there was a 95% chance that the true average marks of the midterm were between 50 and 80. How would you respond to this statement? Is it correct? Why or why not?

(b) (5 marks) Student C in the same class did not agree with student B and he claimed that there was a 95% chance that the true average marks of the **next** midterm were between 50 and 80. How would you respond to this statement? Is it correct? Why or why not?

5. (10 marks) The family income distribution in Ontario is unknown but has a mean of \$21500 and a standard deviation of \$1700.

   (a) (3 marks) A sample of size of 200 families was selected and the sample mean $\overline{x}$ calculated. Describe the distribution of the sample mean $\overline{x}$ in terms of both general shape and descriptive measures mean and variance.

   (b) (3 marks) If the sample size was actually 60 instead of 200, how would the distribution of the sample mean $\overline{x}$ be affected? Illustrate with a graph, indicating the mean and standard deviation.

   (c) (4 marks) What is the probability that a sample of $n = 60$ selected randomly from the population will have a mean equal or greater than \$21300?

6. (10 marks) A bank in Ottawa is considering a survey of its customers for the purpose of estimating the mean number of checks written per customer in each month. A sample of 360 customers was selected. The following sample values were recorded: $\overline{x} = 33.4$ and $s = 11.2$.

   (a) (3 marks) Provide a 90% confidence interval estimate for the mean number of checks written per customer in each month and interpret the estimate.

   (b) (3 marks) Suppose a mistake was made in counting the number of customers that was surveyed, and the actual sample size was 36 instead of 360. Recompute the 90% confidence interval estimate, and compare it to the estimate developed in part (a). Why are the estimates different even though the sample mean and standard deviation did not change?

   (c) (4 marks) The bank is interested in opening another branch in Ottawa. Its demographic studies indicate that the branch will attract about 2000 customers who will have checking accounts. The bank is attempting to prepare a system that would process the checks these customers would write. Would it be reasonable for them to install a system that would process 65000 checks per month? Explain your answer and the logic that led you to this answer.

7. (10 marks) 441 shoppers were selected from people within Ottawa and revealed that 76% made at least one purchase at a one-dollar store last month.

   (a) (5 marks) Based on this sample information, what is the 90% confidence interval estimate for population proportion of shoppers who made at least a one-dollar store purchase last month?

   (b) (5 marks) Determine a 90% confidence interval estimate for **the number of shoppers** who made at least one one-dollar store purchase last month.

8. (10 marks) Firm ABC in Ottawa very prides itself in its ability to fill customers' orders in six calendar days or less on the average. Periodically, the operations manager selects a random sample of customer orders and determines the number of days required to fill the orders. Based on this sample information, he decides if the desired standard is not being met. He will assume that the average number of days to fill customers' orders is six or less unless the data suggest strongly otherwise.

   (a) (2 marks) Establish the appropriate null and alternative hypotheses.

   (b) (5 marks) On one occasion when a sample of 20 customers was selected, the average number of days was 6.65, with a sample standard deviation of 1.5 days. Can the operations manager conclude that his business is achieving its goal? Do the test by using the critical value approach with a significance level of 0.025 and **specify** the assumption required to do the test.

(c) (3 marks) Redo the test by using p-value approach.

9. (10 marks) It is said that the average worker in Ottawa has no more than \$25000 of personal life insurance. To test whether this claim is still true or not, you randomly sample 100 workers in Ottawa. You find that this sample of workers has an average of \$26650 of personal life insurance and a standard derivation of \$12000.

   (a) (5 marks) Determine whether there is enough evidence to reject the claim. Assume the probability of committing a Type I error is 0.05.

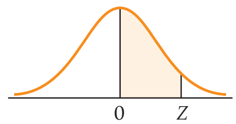
   (b) (5 marks) If the actual average for this population is \$30000, what is the probability of committing a Type II error?

10. (10 marks) A student in ECON 2201C is interested in estimating the mean of a population based on a random sample. She wants the confidence level to be 90% and the margin of error to be 0.30. She does not know what the population standard deviation is, so she has selected the following pilot sample:

8.80	4.89	10.98	15.11	14.79
16.93	1.27	9.06	14.38	5.65
7.24	3.24	2.61	6.09	6.91

Based on this pilot sample, how many more items must be sampled so that the student can make the desired confidence interval estimate?

Areas of the Standard Normal Distribution



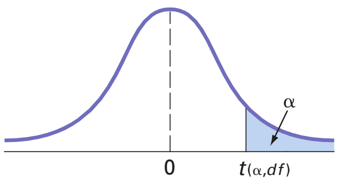
The entries in this table are the probabilities that a standard normal random variable is between 0 and Z (the shaded area).

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998
3.5	.4998									
4.0	.49997									
4.5	.499997									
5.0	.4999997									
6.0	.499999999									

TABLE E.3

Critical Values of *t*

For a particular number of degrees of freedom, entry represents the critical value of *t* corresponding to a specified upper-tail area (α)



Degrees of Freedom	Upper-Tail Areas 0.25	0.10	0.05	0.025	0.01	0.005
1	1.0000	3.0777	6.3138	12.7062	31.8207	63.6574
2	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248
3	0.7649	1.6377	2.3534	3.1824	4.5407	5.8409
4	0.7407	1.5332	2.1318	2.7764	3.7469	4.6041
5	0.7267	1.4759	2.0150	2.5706	3.3649	4.0322
6	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995
8	0.7064	1.3968	1.8595	2.2060	2.8965	3.3554
9	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693
11	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.6864	1.3232	1.7207	2.0796	2.5177	2.8314
22	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.6848	1.3178	1.7109	2.0639	2.4922	2.7969
25	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500
31	0.6825	1.3095	1.6955	2.0395	2.4528	2.7440
32	0.6822	1.3086	1.6939	2.0369	2.4487	2.7385
33	0.6820	1.3077	1.6924	2.0345	2.4448	2.7333
34	0.6818	1.3070	1.6909	2.0322	2.4411	2.7284
35	0.6816	1.3062	1.6896	2.0301	2.4377	2.7238
36	0.6814	1.3055	1.6883	2.0281	2.4345	2.7195
37	0.6812	1.3049	1.6871	2.0262	2.4314	2.7154
38	0.6810	1.3042	1.6860	2.0244	2.4286	2.7116
39	0.6808	1.3036	1.6849	2.0227	2.4258	2.7079
40	0.6807	1.3031	1.6839	2.0211	2.4233	2.7045
41	0.6805	1.3025	1.6829	2.0195	2.4208	2.7012
42	0.6804	1.3020	1.6820	2.0181	2.4185	2.6981
43	0.6802	1.3016	1.6811	2.0167	2.4163	2.6951
44	0.6801	1.3011	1.6802	2.0154	2.4141	2.6923
45	0.6800	1.3006	1.6794	2.0141	2.4121	2.6896
46	0.6799	1.3022	1.6787	2.0129	2.4102	2.6870
47	0.6797	1.2998	1.6779	2.0117	2.4083	2.6846
48	0.6796	1.2994	1.6772	2.0106	2.4066	2.6822

continued

TABLE E.3
Critical Values of *t* (*Continued*)

	Upper-Tail Areas					
Degrees of Freedom	**0.25**	**0.10**	**0.05**	**0.025**	**0.01**	**0.005**
49	0.6795	1.2991	1.6766	2.0096	2.4049	2.6800
50	0.6794	1.2987	1.6759	2.0086	2.4033	2.6778
51	0.6793	1.2984	1.6753	2.0076	2.4017	2.6757
52	0.6792	1.2980	1.6747	2.0066	2.4002	2.6737
53	0.6791	1.2977	1.6741	2.0057	2.3988	2.6718
54	0.6791	1.2974	1.6736	2.0049	2.3974	2.6700
55	0.6790	1.2971	1.6730	2.0040	2.3961	2.6682
56	0.6789	1.2969	1.6725	2.0032	2.3948	2.6665
57	0.6788	1.2966	1.6720	2.0025	2.3936	2.6649
58	0.6787	1.2963	1.6716	2.0017	2.3924	2.6633
59	0.6787	1.2961	1.6711	2.0010	2.3912	2.6618
60	0.6786	1.2958	1.6706	2.0003	2.3901	2.6603
61	0.6785	1.2956	1.6702	1.9996	2.3890	2.6589
62	0.6785	1.2954	1.6698	1.9990	2.3880	2.6575
63	0.6784	1.2951	1.6694	1.9983	2.3870	2.6561
64	0.6783	1.2949	1.6690	1.9977	2.3860	2.6549
65	0.6783	1.2947	1.6686	1.9971	2.3851	2.6536
66	0.6782	1.2945	1.6683	1.9966	2.3842	2.6524
67	0.6782	1.2943	1.6679	1.9960	2.3833	2.6512
68	0.6781	1.2941	1.6676	1.9955	2.3824	2.6501
69	0.6781	1.2939	1.6672	1.9949	2.3816	2.6490
70	0.6780	1.2938	1.6669	1.9944	2.3808	2.6479
71	0.6780	1.2936	1.6666	1.9939	2.3800	2.6469
72	0.6779	1.2934	1.6663	1.9935	2.3793	2.6459
73	0.6779	1.2933	1.6660	1.9930	2.3785	2.6449
74	0.6778	1.2931	1.6657	1.9925	2.3778	2.6439
75	0.6778	1.2929	1.6654	1.9921	2.3771	2.6430
76	0.6777	1.2928	1.6652	1.9917	2.3764	2.6421
77	0.6777	1.2926	1.6649	1.9913	2.3758	2.6412
78	0.6776	1.2925	1.6646	1.9908	2.3751	2.6403
79	0.6776	1.2924	1.6644	1.9905	2.3745	2.6395
80	0.6776	1.2922	1.6641	1.9901	2.3739	2.6387
81	0.6775	1.2921	1.6639	1.9897	2.3733	2.6379
82	0.6775	1.2920	1.6636	1.9893	2.3727	2.6371
83	0.6775	1.2918	1.6634	1.9890	2.3721	2.6364
84	0.6774	1.2917	1.6632	1.9886	2.3716	2.6356
85	0.6774	1.2916	1.6630	1.9883	2.3710	2.6349
86	0.6774	1.2915	1.6628	1.9879	2.3705	2.6342
87	0.6773	1.2914	1.6626	1.9876	2.3700	2.6335
88	0.6773	1.2912	1.6624	1.9873	2.3695	2.6329
89	0.6773	1.2911	1.6622	1.9870	2.3690	2.6322
90	0.6772	1.2910	1.6620	1.9867	2.3685	2.6316
91	0.6772	1.2909	1.6618	1.9864	2.3680	2.6309
92	0.6772	1.2908	1.6616	1.9861	2.3676	2.6303
93	0.6771	1.2907	1.6614	1.9858	2.3671	2.6297
94	0.6771	1.2906	1.6612	1.9855	2.3667	2.6291
95	0.6771	1.2905	1.6611	1.9853	2.3662	2.6286
96	0.6771	1.2904	1.6609	1.9850	2.3658	2.6280
97	0.6770	1.2903	1.6607	1.9847	2.3654	2.6275
98	0.6770	1.2902	1.6606	1.9845	2.3650	2.6269
99	0.6770	1.2902	1.6604	1.9842	2.3646	2.6264
100	0.6770	1.2901	1.6602	1.9840	2.3642	2.6259
110	0.6767	1.2893	1.6588	1.9818	2.3607	2.6213
120	0.6765	1.2886	1.6577	1.9799	2.3578	2.6174
∞	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758

Solution to the final exam

1. (a) The statistic is

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

and it has the (exact) standard normal distribution.

(b) The statistic is

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

and it has the $t(n-1)$ distribution.

(c) The statistic is

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

and it has the (exact) standard normal distribution.

(d) The statistic is

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

and it has $t(n-1)$ distribution. (Or the statistic is

$$z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

and its distribution is approximately standard normally distributed. As the sample size n goes to infinite, the $t(n-1)$ distribution converges to standard normal distribution.)

(e) The statistic is

$$z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

and its distribution is approximately standard normal.

2. Based on the information, the population proportion

$$p = 0.35$$

the sample size

$$n = 500.$$

Moreover,

$$\begin{aligned} \mu_{\bar{p}} &= p = 0.35 \\ \sigma_{\bar{p}} &= \sqrt{\frac{p\,(1-p)}{n}} = \sqrt{\frac{0.35\,(1-0.35)}{500}} = 0.02133 \end{aligned}$$

(a) Since $np = 175$ and $n(1-p) = 325$ are both greater than 5, based on the classic central limiting theorem, the distribution of the sample proportion $\bar{p}$ will be approximately normally distributed with a mean $\mu_{\bar{p}} = 0.35$ and a standard deviation $\sigma_{\bar{p}} = 0.02133$.

(b)

$$\begin{aligned} P(\bar{p} &> 0.30) = P\left(\frac{\bar{p} - \mu_{\bar{p}}}{\sigma_{\bar{p}}} > \frac{0.30 - \mu_{\bar{p}}}{\sigma_{\bar{p}}}\right) \\ &= P(z > -2.34) = P(z \le 2.34) \\ &= 0.50 + P(0 \le z \le 2.34) \\ &= 0.5 + 0.4904 = 0.9904 \end{aligned}$$

(c) We are looking for an interval (a, b) such that

$$P\left(a \leq \bar{p} \leq b\right) = 0.99.$$

To this end, based on the central limiting theorem, we have

$$\begin{aligned}
0.99 &= P\left(-z_{0.99/2} \leq \frac{\bar{p}-\mu_{\bar{p}}}{\sigma_{\bar{p}}} \leq z_{0.99/2}\right) = P\left(-z_{0.99/2} \leq z \leq z_{0.99/2}\right) \\
&= P\left(z \leq z_{0.99/2}\right) - P\left(z \leq -z_{0.99/2}\right) \\
&= P\left(z \leq z_{0.99/2}\right) - P\left(z \geq z_{0.99/2}\right) \\
&= 2P\left(z \leq z_{0.99/2}\right) - 1 \\
&= 2\left[0.5 + P\left(0 \leq z \leq z_{0.99/2}\right)\right] - 1 \\
&= 2P\left(0 \leq z \leq z_{0.99/2}\right)
\end{aligned}$$

That is,

$$P\left(0 \leq z \leq z_{0.99/2}\right) = \frac{0.99}{2} = 0.495$$

From the table, the critical value

$$z_{0.99/2} = \frac{2.57 + 2.58}{2} = 2.575.$$

(**If the students use** 2.57 **or** 2.58, **that is fine**.) Thus, the interval

$$\left(\mu_{\bar{p}} - z_{0.99/2} \times \sigma_{\bar{p}},\ \mu_{\bar{p}} + z_{0.99/2} \times \sigma_{\bar{p}}\right) = (0.29508, 0.40492)$$

will contain 99% of the measurements.

3. Let

$$\begin{aligned}
FR &= \text{favorable review} \\
HS &= \text{huge successes} \\
MS &= \text{modest successes} \\
BE &= \text{break even} \\
LO &= \text{losers}
\end{aligned}$$

Based on the given information,

$$\begin{aligned}
P\left(HS\right) &= 0.1 \text{ and } P\left(FR|HS\right) = 0.99 \\
P\left(MS\right) &= 0.2 \text{ and } P\left(FR|MS\right) = 0.70 \\
P\left(BE\right) &= 0.4 \text{ and } P\left(FR|BE\right) = 0.40 \\
P\left(LO\right) &= 0.3 \text{ and } P\left(FR|BE\right) = 0.20
\end{aligned}$$

(a) According to the conditional probability formula, we have

$$\begin{aligned}
P(\text{favorable review}) &= P\left(FR \cap \{HS, MS, BE, \text{ or } LO\}\right) \\
&= P\left(FR \cap HS\right) + P\left(FR \cap MS\right) \\
&\quad + P\left(FR \cap BE\right) + P\left(FR \cap LO\right) \\
&= P\left(HS\right)P\left(FR|HS\right) + P\left(MS\right)P\left(FR|MS\right) \\
&\quad + P\left(BE\right)P\left(FR|BE\right) + P\left(LO\right)P\left(FR|LO\right) \\
&= 0.1 \times 0.99 + 0.2 \times 0.7 + 0.4 \times 0.4 + 0.3 \times 0.3 = 0.459
\end{aligned}$$

(b)

$$P(\text{unfavorable review}) = 1.0 - P(\text{favorable review}) = 1 - 0.459 = 0.541$$

(c) Based on the Bayes formula,

$$\begin{aligned} P(\text{huge success}|\text{favorable review}) &= P(HS|FR) \\ &= \frac{P(HS \cap FR)}{P(FR)} = \frac{P(HS)P(FR|HS)}{P(FR)} \\ &= \frac{0.1 \times 0.99}{0.459} = 0.2157 \end{aligned}$$

$$\begin{aligned} P(\text{moderate success}|\text{favorable review}) &= P(MS|FR) \\ &= \frac{P(MS \cap FR)}{P(FR)} = \frac{P(MS)P(FR|MS)}{P(FR)} \\ &= \frac{0.2 \times 0.7}{0.459} = 0.3050 \end{aligned}$$

$$\begin{aligned} P(\text{break even}|\text{favorable review}) &= P(BE|FR) \\ &= \frac{P(BE \cap FR)}{P(FR)} = \frac{P(BE)P(FR|BE)}{P(FR)} \\ &= \frac{0.4 \times 0.4}{0.459} = 0.3486 \end{aligned}$$

and

$$\begin{aligned} P(\text{loser}|\text{favorable review}) &= P(LO|FR) = \frac{P(LO \cap FR)}{P(FR)} = \frac{P(LO)P(FR|BE)}{P(FR)} \\ &= \frac{0.3 \times 0.2}{0.459} = 0.1307 \end{aligned}$$

4. (a) This is not correct. The average marks of the midterm are a single value. Therefore it has no probability. What the confidence interval is telling you is that if you want to produce all the possible confidence intervals using each possible sample mean from the population, 95% of these intervals would contain the population mean.

   (b) This is also not correct. Although the next midterm has not happened yet, but the population mean (the true average marks of the next midterm) is deterministic rather than random. So, it has no any probability belonging to the interval (50, 80).

5. (a) Because this is a large sample, the distribution of the sample mean $\bar{x}$ will be approximately normally distributed regardless of the distribution of the parent population. The mean of the sampling mean $\bar{x}$ will be $\mu_{\bar{x}} = \mu = \$21,500$ and the standard deviation of the sampling mean $\bar{x}$ will be $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 1700/\sqrt{200} = \$120.21$

   (b) The distribution of the sample mean $\bar{x}$ would still be approximately normally distributed but the spread would be wider. The mean would be the same as before \$21,500, but the standard deviation would become $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 1700/\sqrt{60} = \$219.47$ which is bigger than $1700/\sqrt{200} = \$120.21$.

   (c)

$$\begin{aligned} P(\overline{x} &\geq 21300) = P\left(\frac{\overline{x} - \mu}{\sigma_{\bar{x}}} \geq \frac{21300 - 21500}{1700/\sqrt{60}}\right) \\ &= P(z \geq -0.91) = P(z \leq 0.91) \\ &= 0.5 + P(0 \leq z \leq 0.91) \\ &= 0.5 + 0.3186 = 0.8186 \end{aligned}$$

6. Given $\overline{x} = 33.4$, $s = 11.2$, and $n = 360$, though we don't know the population's distribution, the z-value

$$z = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

still follows normal distribution since the sample size $n \geq 30$.

To derive the interval estimation for the population mean for the confidence level $1 - \alpha = 0.9$, we have to derive the critical point $\bar{z}_{\alpha/2}$ first:

$$\begin{aligned}
1 - \alpha &= P\left(-\bar{z}_{\alpha/2} \leq z \leq \bar{z}_{\alpha/2}\right) \\
&= P\left(z \leq \bar{z}_{\alpha/2}\right) - P\left(z \leq -\bar{z}_{\alpha/2}\right) \\
&= P\left(z \leq \bar{z}_{\alpha/2}\right) - P\left(z \geq \bar{z}_{\alpha/2}\right) \\
&= P\left(z \leq \bar{z}_{\alpha/2}\right) - \left[1 - P\left(z \leq \bar{z}_{\alpha/2}\right)\right] \\
&= 2P\left(z \leq \bar{z}_{\alpha/2}\right) - 1 \\
&= 2\left[0.5 + P\left(0 \leq z \leq \bar{z}_{\alpha/2}\right)\right] - 1 \\
&= 2P\left(0 \leq z \leq \bar{z}_{\alpha/2}\right)
\end{aligned}$$

Thus, the critical value $\bar{z}_{\alpha/2}$ satisfies

$$P\left(0 \leq z \leq \bar{z}_{\alpha/2}\right) = \frac{1-\alpha}{2} = \frac{0.9}{2} = 0.45.$$

From the normal table,

$$\bar{z}_{\alpha/2} = \frac{1.64 + 1.65}{2} = 1.645.$$

(**If the students use** 1.64 or 1.65, **that is fine**.)

From

$$\begin{aligned}
-\bar{z}_{\alpha/2} &\leq z \leq \bar{z}_{\alpha/2} \Longleftrightarrow -\bar{z}_{\alpha/2} \leq \frac{\overline{x} - \mu}{s/\sqrt{n}} \leq \bar{z}_{\alpha/2} \Longleftrightarrow \\
-\bar{z}_{\alpha/2}s/\sqrt{n} &\leq \overline{x} - \mu \leq \bar{z}_{\alpha/2}s/\sqrt{n} \Longleftrightarrow \\
-\overline{x} - \bar{z}_{\alpha/2}s/\sqrt{n} &\leq -\mu \leq -\overline{x} + \bar{z}_{\alpha/2}s/\sqrt{n} \Longleftrightarrow \\
\overline{x} - \bar{z}_{\alpha/2}s/\sqrt{n} &\leq \mu \leq \overline{x} + \bar{z}_{\alpha/2}s/\sqrt{n}
\end{aligned}$$

That is, the interval estimation of μ is

$$\left(\overline{x} - \bar{z}_{\alpha/2} \times \frac{s}{\sqrt{n}}, \quad \overline{x} - \bar{z}_{\alpha/2} \times \frac{s}{\sqrt{n}}\right).$$

(a) If $n = 360$, then the interval estimate for μ is

$$\begin{aligned}
&\left(\overline{x} - \bar{z}_{\alpha/2} \times \tfrac{s}{\sqrt{n}}, \quad \overline{x} - \bar{z}_{\alpha/2} \times \tfrac{s}{\sqrt{n}}\right) \\
&= \left(33.4 - 1.645 \times \tfrac{11.2}{\sqrt{360}}, \quad 33.4 - 1.645 \times \tfrac{11.2}{\sqrt{360}}\right) \\
&= (32.4290, \quad 34.3710).
\end{aligned}$$

That is, each customer would write the number of checks per month between 32.4290 and 34.3710 with confidence level 90%.

(b) If $n = 36$, then the interval estimate for μ is

$$\begin{aligned}
&\left(\overline{x} - \bar{z}_{\alpha/2} \times \tfrac{s}{\sqrt{n}}, \quad \overline{x} - \bar{z}_{\alpha/2} \times \tfrac{s}{\sqrt{n}}\right) \\
&= \left(33.4 - 1.645 \times \tfrac{11.2}{\sqrt{36}}, \quad 33.4 - 1.645 \times \tfrac{11.2}{\sqrt{36}}\right) \\
&= (30.3293, \quad 36.4707)\,.
\end{aligned}$$

Decreasing the sample size will increase the margin of error which will increase the confidence interval.

(c) One approach would be to multiply the range for the sample of 36 which would be more conservative to determine if this system would handle them.

$$\begin{aligned}
&(30.3292 \times 2000, \quad 36.4707 \times 2000) \\
&= (60,658.4, \quad 72,941.40)
\end{aligned}$$

Since the upper limit of this interval is much greater than 65,000 students may say that this is not a good system.

7. (a)

$$\bar{p} = 0.76,$$

$n\bar{p} = 441 \times 0.76$ and $n(1-\bar{p}) = 441 \times (1 - 0.76)$ both are greater than 5, thus the z-value

$$z = \frac{\bar{p} - p}{\sigma_{\bar{p}}} = \frac{\bar{p} - p}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}}$$

follows normal distribution.

To derive the interval estimation for the population proportion p for the confidence level $1 - \alpha = 0.9$, we have to derive the critical point $\bar{z}_{\alpha/2}$ first:

$$\begin{aligned}
1 - \alpha &= P\left(-\bar{z}_{\alpha/2} \le z \le \bar{z}_{\alpha/2}\right) \\
&= P\left(z \le \bar{z}_{\alpha/2}\right) - P\left(z \le -\bar{z}_{\alpha/2}\right) \\
&= P\left(z \le \bar{z}_{\alpha/2}\right) - P\left(z \ge \bar{z}_{\alpha/2}\right) \\
&= P\left(z \le \bar{z}_{\alpha/2}\right) - \left[1 - P\left(z \le \bar{z}_{\alpha/2}\right)\right] \\
&= 2P\left(z \le \bar{z}_{\alpha/2}\right) - 1 \\
&= 2\left[0.5 + P\left(0 \le z \le \bar{z}_{\alpha/2}\right)\right] - 1 \\
&= 2P\left(0 \le z \le \bar{z}_{\alpha/2}\right)
\end{aligned}$$

Thus, the critical value $\bar{z}_{\alpha/2}$ satisfies

$$P\left(0 \le z \le \bar{z}_{\alpha/2}\right) = \frac{1-\alpha}{2} = \frac{0.9}{2} = 0.45.$$

From the normal table,

$$\bar{z}_{\alpha/2} = \frac{1.64 + 1.65}{2} = 1.645.$$

(**If the students use** 1.64 or 1.65, **that is fine**.)

From

$$
\begin{aligned}
-\bar{z}_{\alpha/2} &\le z \le \bar{z}_{\alpha/2} \Longleftrightarrow -\bar{z}_{\alpha/2} \le \frac{\bar{p}-p}{\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}} \le \bar{z}_{\alpha/2} \Longleftrightarrow \\
-\bar{z}_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} &\le \bar{p}-p \le \bar{z}_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \Longleftrightarrow \\
-\bar{p}-\bar{z}_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} &\le -p \le -\bar{p}+\bar{z}_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \Longleftrightarrow \\
\bar{p}-\bar{z}_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} &\le p \le \bar{p}+\bar{z}_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}
\end{aligned}
$$

That is, the interval estimation of p is

$$
\left(\bar{p}-\bar{z}_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}},\quad \bar{p}+\bar{z}_{\alpha/2}\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}\right).
$$

That is, The interval estimate of p is

$$
\begin{gathered}
\left(0.76-1.645\sqrt{\tfrac{0.76\times(1-0.76)}{441}},\quad 0.76+1.645\sqrt{\tfrac{0.76\times(1-0.76)}{441}}\right) \\
= \\
(0.7265,\ 0.7935)\,.
\end{gathered}
$$

(b) The 90% confidence interval estimate for the number of shoppers who made at least one one-dollar store purchase last month is

$$
\begin{gathered}
(0.7265,\ 0.7935)\times 750000 \\
= \\
(544875,\ \ 595125).
\end{gathered}
$$

8. (a)

H_0: $\mu \le 6$ days (status quo)
H_a: $\mu > 6$ days

(b) To do the test by using the critical value approach, we need to assume that the population is normally distributed.

Since the population standard deviation σ is unknown plus the above assumption, the statistics is

$$
t=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}.
$$

From the question, we have

$$
n=20,\ \bar{x}=6.65, s=1.5 \text{ and } \alpha=0.025
$$

and

$$
t=\frac{\bar{x}-\mu_0}{s/\sqrt{n}}=\frac{6.65-6}{1.5/\sqrt{20}}=1.9379.
$$

Based the definition, the critical value $\bar{t}_\alpha$ $(\bar{x}_\alpha)$ satisfies

$$
\begin{aligned}
\alpha &= P(\bar{x}>\bar{x}_\alpha) \\
&= P(t>\bar{t}_\alpha)
\end{aligned}
$$

For $\alpha = 0.025$ and $n-1 = 19$,

$$\bar{t}_\alpha = 2.0930$$

or

$$\bar{x}_\alpha = \mu_0 + \bar{t}_\alpha \times \frac{s}{\sqrt{n}} = 6 + 2.0930 \times \frac{1.5}{\sqrt{20}} = 6.7020$$

Since

$$\begin{aligned} t &= 1.9379 < \bar{t}_\alpha = 2.0930 \\ (\text{or } \bar{x} &= 6.65 < \bar{x}_\alpha = 6.7020) \end{aligned}$$

do not reject H_0 and conclude that the firm is achieving its goal.

(c) Based on the definition of p-value,

$$\begin{aligned} p\text{-value} &= P(td > t) = P(td > 1.9379) \\ &= \frac{0.05 + 0.025}{2} = 0.0375. \end{aligned}$$

(**If the students use** 0.05 or 0.025, **that is fine**.)
Since p-value$= 0.0375 > 0.025$, do not reject the null hypothesis H_0.

9.

$$\overline{x} = \$26650,\ n = 100,\ \text{and } s = \$12000.$$

(a) The hypothesis is

$$\begin{cases} H_0 : \mu \leq \$25000 \text{ (claim)} \\ H_a : \mu > \$25000 \end{cases}$$

and $\alpha = 0.05$.
Since the population standard deviation σ is unknown, but the sample size $n = 100 > 30$, the z-value

$$z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

approximately follows normal distribution.
From the definition, the critical value $\bar{z}_\alpha$ (or $\bar{x}_\alpha$) satisfies

$$\begin{aligned} \alpha &= P(\bar{x} > \bar{x}_\alpha) = P(z > \bar{z}_\alpha) \\ &= 1 - P(z \leq \bar{z}_\alpha) \\ &= 1 - [0.5 + P(0 \leq z \leq \bar{z}_\alpha)] \\ &= 0.5 - P(0 \leq z \leq \bar{z}_\alpha) \end{aligned}$$

That is,

$$P(0 \leq z \leq \bar{z}_\alpha) = 0.5 - \alpha = 0.5 - 0.05 = 0.45$$

and

$$\bar{z}_\alpha = \frac{1.64 + 1.65}{2} = 1.645.$$

(**If the students use** 1.64 or 1.65, **that is fine**.)
From

$$z = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{26650 - 25000}{12,000/\sqrt{100}} = 1.38 < 1.645 = \bar{z}_\alpha,$$

do not reject the null hypothesis.
In other words, there is no enough evidence to reject the claim.

(b) Let

$$\begin{aligned}
\mu_0 &= 25000 \\
\mu^* &= 30000 \\
z^{\mu^*} &= \frac{\overline{x}-\mu^*}{s/\sqrt{n}} \\
z^{\mu_0} &= \frac{\overline{x}-\mu_0}{s/\sqrt{n}} = \frac{\overline{x}-\mu^*+(\mu^*-\mu_0)}{s/\sqrt{n}} \\
&= z^{\mu^*} + \frac{\mu^*-\mu_0}{s/\sqrt{n}} \\
&= z^{\mu^*} + \frac{30000-25000}{12000/\sqrt{100}} \\
&= z^{\mu^*} + \frac{5000}{1200} = z^{\mu^*} + 4.17
\end{aligned}$$

then, by the definition,

$$\begin{aligned}
\beta^{30000} &= P\left(\text{Accept } H_0|\mu = 30000\right) \\
&= P\left(\bar{x} \le \bar{x}_\alpha|\mu = 30000\right) \\
&= P\left(z^{\mu_0} \le \bar{z}_\alpha|\mu = 30000\right) \\
&= P\left(z^{\mu^*} + 4.17 \le \bar{z}_\alpha|\mu = 30000\right) \\
&= P\left(z^{\mu^*} \le \bar{z}_\alpha - 4.17|\mu = 30000\right) \\
&= P\left(z^{\mu^*} \le 1.645 - 4.17|\mu = 30000\right) \\
&= P\left(z^{\mu^*} \le -2.522|\mu = 30000\right) \\
&= P\left(z^{\mu^*} \ge 2.522|\mu = 30000\right) \\
&= 1 - P\left(z^{\mu^*} \le 2.522|\mu = 30000\right) \\
&= 1 - \left[0.5 + P\left(0 \le z^{\mu^*} \le 2.522|\mu = 30000\right)\right] \\
&= 0.5 - P\left(0 \le z^{\mu^*} \le 2.522|\mu = 30000\right) \\
&= 0.5 - 0.4941 = 0.0059.
\end{aligned}$$

Thus, the probability of committing a Type II error is 0.0059 which is very small.

10. Based on the question, we have

$$\begin{aligned}
1-\alpha &= 0.90 \text{ (The confidence level)} \\
e &= 0.30 \text{ (The margin of error)}
\end{aligned}$$

Since the population's distribution and population standard deviation σ are not specified here, so, the t-distribution can not be used here for sure. To use the

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

statistic, we have to estimate σ from the given pilot sample.

$$\bar{x} = 8.53$$

and

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{m-1}} = 4.94.$$

By the definition, the critical value $\bar{z}_{\alpha/2}$ can be derived from

$$\begin{aligned}
\alpha &= P\left(-\bar{z}_{\alpha/2} \leq z \leq \bar{z}_{\alpha/2}\right) \\
&= P\left(z \leq \bar{z}_{\alpha/2}\right) - P\left(z \leq -\bar{z}_{\alpha/2}\right) \\
&= P\left(z \leq \bar{z}_{\alpha/2}\right) - P\left(z \geq \bar{z}_{\alpha/2}\right) \\
&= 2\left[0.5 + P\left(0 \leq z \leq \bar{z}_{\alpha/2}\right)\right] - 1 \\
&= 2P\left(0 \leq z \leq \bar{z}_{\alpha/2}\right)
\end{aligned}$$

That is,

$$P\left(0 \leq z \leq \bar{z}_{\alpha/2}\right) = \frac{\alpha}{2} = 0.45$$

and

$$\bar{z}_{\alpha/2} = \frac{1.64 + 1.65}{2} = 1.645.$$

So, we will use

$$n = \max\left\{\left(\frac{\bar{z}_{\alpha/2} \times s}{e}\right)^2, 30\right\} = \max\left\{\frac{\left(\bar{z}_{\alpha/2}\right)^2 s^2}{e^2}, 30\right\}$$

to determine the sample size.

Since

$$\left(\frac{\bar{z}_{\alpha/2} \times s}{e}\right)^2 = \left(\frac{1.645 \times 4.94}{0.30}\right)^2 = 733.74 \approx 734 > 30.$$

The required sample size is 734 items from the population. However, the 15 items in the pilot sample can be used so the net required sample is $734 - 15 = 719$.