STA220 Midterm 2, Mar. 14, 2023

(80 minutes; 4 questions; 8 pages; total points = 48)

Do not open this test until told to do so. Try to answer as many questions as you can. Clearly write down your name and student number on the front page. You can use a non-programmable calculator, and course notes (printed and handwritten). No devices are permitted. Write your answers in the space provided, only. You should explain all of your solutions clearly. A full mark will not be given for an answer without explanation.

Academic Integrity Acknowledgement Form

The University of Toronto's Code of Behaviour on Academic Matters outlines the behaviours that constitute academic misconduct, the processes for addressing academic offences, and the penalties that may be imposed. Potential offences include (but are not limited to):

- Looking at someone else's answers during the exam.
- Letting someone else look at your answers during the exam.
- Misrepresenting your identity or having someone else complete your exam.
- Sharing or posting the exam questions during or for 24 hours after the exam.

Prior to beginning this exam, you must affirm that you will follow the Code of Behaviour on Academic Matters in the completion of this exam, by completing the following statement:

I, ____ (print your full name), agree to fully abide by the Code of Behaviour on Academic Matters. I promise not to commit academic misconduct, and am aware that significant penalties may be imposed if I do.

Signature: ____

1. Suppose a random variable representing the internal body temperature of a randomly chosen mouse (measured in Celsius) follows a normal distribution with a mean of 36°C and a standard deviation of 2°C.

(a) [**2**] What is the probability that a randomly chosen mouse has a body temperature of **exactly** 38°C?

$X =$ body temperature (°C)

$X \sim N(36, 2^2)$

$P(X = 38) = \boxed{0}$ as X is continuous

(b) [**2**] What is the probability that a randomly chosen mouse has a body temperature **between** 34°C and 37°C?

$P(34 \le X \le 37) = P\left(\frac{34-36}{2} \le \frac{X-36}{2} \le \frac{37-36}{2}\right) =$

$P(-1 \le Z \le 0.5) = P(Z \le 0.5) - P(Z \le -1) =$ (table)

$= 1 - 0.3085 - 0.1587 = \boxed{0.5328}$

Here $Z \sim N(0,1)$

(c) [**2**] Find the value of a such that 25% of mice have body temperature **less** than a°C.

$P(Z \le -0.67) = 0.25$ (table)

$P\left(\frac{X-36}{2} \le -0.67\right) = P(X \le 36 - 0.67 \cdot 2) =$

$P(X \le 34.66) \Rightarrow \boxed{a = 34.66}$

(d) **[2]** Find the value of b such that 10% of mice have body temperature **more** than b°C. (It may be helpful to draw the distribution curve first)

$P(Z \le -1.28) = 0.1$ (table)

$P(Z \le -1.28) = P(Z \ge 1.28) =$

$P\left(\frac{X-36}{2} \ge 1.28\right) = P(X \ge 36 + 1.28 \cdot 2) =$

$P(X \ge 38.56) \Rightarrow \boxed{b = 38.56}$

(e) **[2]** If we take a sample of four mice, what is the distribution for the **average body temperature** of these mice? Explain why.

$X_1 \ldots X_4 \sim N(36, 2^2) \Rightarrow$

$\bar{X} = \frac{X_1 + \ldots + X_4}{4} \sim N\left(36, \frac{2^2}{4}\right) = \boxed{N(36, 1)}$

exactly, not approximately

(f) **[2]** To convert temperatures in degrees Celsius to Fahrenheit, multiply the value by 1.8 and add 32, i.e. °F = (°C × 1.8) + 32. What is the distribution of body temperature of a randomly chosen mouse measured in Fahrenheit? Explain why.

$Y = 1.8 \cdot X + 32 \Rightarrow$

$Y \sim N(1.8 \cdot 36 + 32, 1.8^2 \cdot 2^2) = \boxed{N(96.8, 3.6^2)}$

2. Suppose you roll a **fair four-sided die** (with values 1,2,3,4) and denote by X a random variable that represents the score that you got in one roll.

(a) [**2**] Write down the distribution table for X.

X	1	2	3	4
$P(x)$	1/4	1/4	1/4	1/4

(b) [**2**] Find the expectation $E(X)$.

$$\mu = E(X) = 1\cdot\frac{1}{4} + 2\cdot\frac{1}{4} + 3\cdot\frac{1}{4} + 4\cdot\frac{1}{4} = \boxed{2.5}$$

(c) [**2**] Find the variance $Var(X)$.

$$\sigma^2 = Var(X) = (1-2.5)^2\cdot\frac{1}{4} + (2-2.5)^2\cdot\frac{1}{4} + (3-2.5)^2\cdot\frac{1}{4} + (4-2.5)^2\cdot\frac{1}{4} =$$
$$\frac{1.5^2}{4} + \frac{0.5^2}{4} + \frac{0.5^2}{4} + \frac{1.5^2}{4} = \frac{1.5^2}{2} + \frac{0.5^2}{2} = \boxed{1.25}$$

(d) [**2**] Denote by Y the random variable that represents the **average score** in 50 rolls. What is the distribution of Y? Explain why.

By CLT $\bar{X} = \frac{X_1 + .. + X_{50}}{50} \overset{approx.}{\sim} N\left(\mu, \frac{\sigma^2}{50}\right) =$

$$N\left(2.5, \frac{1.25}{50}\right) = \boxed{N(2.5, 0.025)}$$

(e) [**2**] Find the probability that the **average score** in 50 rolls is less than 2.

$$P(\overline{X} < 2) = P\left(\frac{\overline{X} - 2.5}{\sqrt{0.025}} \leq \frac{2 - 2.5}{\sqrt{0.025}} \right) =$$

$$P(Z \leq -3.162) = \boxed{0.0008} \text{ (table)}$$

Here $Z \sim N(0, 1)$

(f) [**2**] You rolled the die 50 times and got a **sample of 50 scores**. In this sample: score 1 appeared 15 times; score 2 appeared 10 times; score 3 appeared 15 times; and score 4 appeared 10 times. Use this information to find the sample mean.

$$x_1, \dots x_n = (\underbrace{1 \dots 1}_{15}, \underbrace{2 \dots 2}_{10}, \underbrace{3 \dots 3}_{15}, \underbrace{4 \dots 4}_{10})$$

$$\overline{x} = \frac{1 \cdot 15 + 2 \cdot 10 + 3 \cdot 15 + 4 \cdot 10}{50} = \boxed{2.4}$$

(g) [**2**] Use the sample mean and the prior knowledge about the theoretical variance from part (c) to find the 95% confidence interval for the average score when rolling a four-sided die.

95% CI with <u>known</u> σ

$$\left[\overline{x} - 1.96 \cdot \frac{\sigma}{\sqrt{n}}, \ \overline{x} + 1.96 \cdot \frac{\sigma}{\sqrt{n}} \right] =$$

$$[2.4 - 1.96 \cdot \sqrt{0.025}, \ 2.4 + 1.96 \sqrt{0.025}] = \boxed{[2.09, \ 2.71]}$$

(h) [**2**] Does the confidence interval cover the theoretical value of the expectation from part (b)? How can you explain this result?

$\mu = 2.5 \Rightarrow$ CI does covers μ

95% of samples will result in CI that covers population mean μ. $\Rightarrow$ we are in these 95%

3. A professor is interested in estimating the average GRE score of admitted graduate students at U of T. To do so, they sampled the scores of 30 graduate students and computed the sample mean and sample standard deviation. The resulting 90% confidence interval is [275, 325].

(a) [**2**] From the provided information, can you recover the sample mean? If yes, what is the value?

$\bar{x}$ is the midpoint of CI $\Rightarrow \bar{x} = \frac{275+325}{2} = \boxed{300}$

(b) [**2**] From the provided information, can you recover the sample standard deviation? If yes, what is the value?

width of the interval $= 25 = t_{29}^{90\%} \cdot \frac{S}{\sqrt{n}}$

$\begin{cases} t_{29}^{90\%} = 1.7 \text{ (table)} \\ n = 30 \end{cases} \Rightarrow S = 25 \frac{\sqrt{30}}{1.7} = \boxed{80.547}$

(c) [**2**] From the provided information, can you calculate the 80% confidence interval? If yes, what is the value?

80% CI is $[\bar{x} - t_{29}^{80\%} \cdot \frac{S}{\sqrt{n}}, \bar{x} + t_{29}^{80\%} \cdot \frac{S}{\sqrt{n}}]$

$t_{29}^{80\%} = 1.31 \Rightarrow [300 - 1.31 \cdot \frac{80.547}{\sqrt{30}}, 300 + 1.31 \cdot \frac{80.547}{\sqrt{30}}] =$

$\boxed{[280.735, 319.265]}$

(d) [**2**] The curious professor took 1500 graduate students at U of T and split them into 50 groups of 30 students. For each group, Professor calculated a 90% confidence interval (50 intervals in total). How many of these intervals **will not** cover the average GRE score of admitted graduate students at U of T?

90% of CI should cover the population parameter $\mu \Rightarrow$ 10% will not

In total we have 50 CI $\Rightarrow \boxed{5}$ of them will not cover μ

4. The overall acceptance rate at U of T is 40%. The professor wants to test if there is a gender bias at U of T and if the acceptance rate for female applicants is different from the overall acceptance rate.

(a) [**2**] State the null and alternative hypotheses. Is the alternative one or two-sided?

μ = average acceptance rate for females

$H_0: \mu = 0.4$ $\quad H_a: \mu \neq 0.4$ (two-sided)

(b) [**2**] In 2023, 49 out of 100 female students who applied to U of T were accepted. Use this information to find the test statistic.

$\bar{x} = 49/100 = 0.49 \quad p = 0.4 \quad n = 100$

$$z_{obs} = \frac{\bar{x} - p}{\sqrt{p(1-p)/n}} = \frac{0.49 - 0.4}{\sqrt{0.4 \cdot 0.6/100}} = \boxed{1.837}$$

(c) [**2**] Use the test statistic from part (b) and the significance level $\alpha = 0.1$ to make a conclusion about the gender bias at U of T.

$$\text{p-value} = P(|Z| > |z_{obs}|) = 2 \cdot P(Z < -1.837) = 2 \cdot 0.0329 = \boxed{0.0658}$$

(table)

p-value $< 0.1 \Rightarrow$ reject H_0 in favor of $H_a \Rightarrow$ we are 90% confident that there is a gender bias

(d) [2] Can you make the same conclusion at significance level $\alpha = 0.05$?

$\boxed{\text{No}}$ as p-value $> 0.05 \Rightarrow$
we do not have enough evidence to conclude with 95% confidence that there is a gender bias

(e) [2] How the result in part (d) would change if the professor was testing that U of T **favors** female students (i.e. that the acceptance rate for female applicants is higher than the overall rate)?

$H_0: \mu = 0.4 \qquad H_a: \mu > 0.4$ (one-sided)

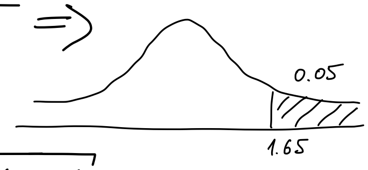
p-value $= P(Z > z_{obs}) = \boxed{0.0329}$

p-value $< 0.05 \Rightarrow$ reject H_0 in favor of $H_a \Rightarrow$
we are 95% confident that female students are favored

(f) [2] What is the **minimum number** of accepted female students (out of 100) required to reject the hypothesis in part (e)?

If H_0 is rejected $\Rightarrow$ p-value $\leq 0.05 \Rightarrow$

$P(Z \geq z_{obs}) \leq 0.05 \Rightarrow z_{obs} \geq 1.65$



$$\frac{m/100 - 0.4}{\sqrt{0.4 \cdot 0.6/100}} \geq 1.65 \Rightarrow m/100 \geq 1.65 \cdot \sqrt{\frac{0.4 \cdot 0.6}{100}} + 0.4 = 0.481$$

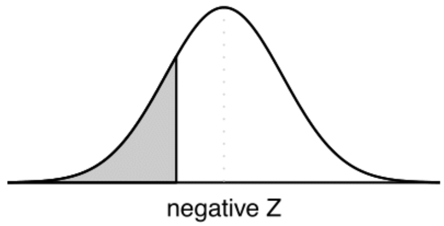
Thus $m > 48$, i.e $\boxed{m = 49}$ is the minimum number

Check: if $m = 48 \quad z_{obs} = \dfrac{0.48 - 0.4}{\sqrt{0.4 \cdot 0.6/100}} = 1.633 \Rightarrow$

p-value $= 0.0516 > 0.05$

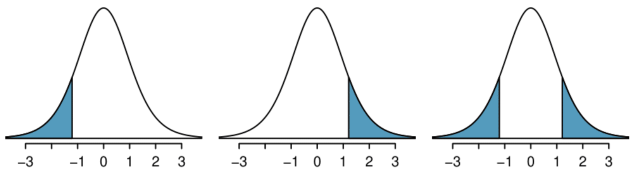
[END OF EXAMINATION; total points = 48]

Normal probability table



Second decimal place of Z										
0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01	0.00	Z
0.0002	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	-3.4
0.0003	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0005	0.0005	0.0005	-3.3
0.0005	0.0005	0.0005	0.0006	0.0006	0.0006	0.0006	0.0006	0.0007	0.0007	-3.2
0.0007	0.0007	0.0008	0.0008	0.0008	0.0008	0.0009	0.0009	0.0009	0.0010	-3.1
0.0010	0.0010	0.0011	0.0011	0.0011	0.0012	0.0012	0.0013	0.0013	0.0013	-3.0
0.0014	0.0014	0.0015	0.0015	0.0016	0.0016	0.0017	0.0018	0.0018	0.0019	-2.9
0.0019	0.0020	0.0021	0.0021	0.0022	0.0023	0.0023	0.0024	0.0025	0.0026	-2.8
0.0026	0.0027	0.0028	0.0029	0.0030	0.0031	0.0032	0.0033	0.0034	0.0035	-2.7
0.0036	0.0037	0.0038	0.0039	0.0040	0.0041	0.0043	0.0044	0.0045	0.0047	-2.6
0.0048	0.0049	0.0051	0.0052	0.0054	0.0055	0.0057	0.0059	0.0060	0.0062	-2.5
0.0064	0.0066	0.0068	0.0069	0.0071	0.0073	0.0075	0.0078	0.0080	0.0082	-2.4
0.0084	0.0087	0.0089	0.0091	0.0094	0.0096	0.0099	0.0102	0.0104	0.0107	-2.3
0.0110	0.0113	0.0116	0.0119	0.0122	0.0125	0.0129	0.0132	0.0136	0.0139	-2.2
0.0143	0.0146	0.0150	0.0154	0.0158	0.0162	0.0166	0.0170	0.0174	0.0179	-2.1
0.0183	0.0188	0.0192	0.0197	0.0202	0.0207	0.0212	0.0217	0.0222	0.0228	-2.0
0.0233	0.0239	0.0244	0.0250	0.0256	0.0262	0.0268	0.0274	0.0281	0.0287	-1.9
0.0294	0.0301	0.0307	0.0314	0.0322	0.0329	0.0336	0.0344	0.0351	0.0359	-1.8
0.0367	0.0375	0.0384	0.0392	0.0401	0.0409	0.0418	0.0427	0.0436	0.0446	-1.7
0.0455	0.0465	0.0475	0.0485	0.0495	0.0505	0.0516	0.0526	0.0537	0.0548	-1.6
0.0559	0.0571	0.0582	0.0594	0.0606	0.0618	0.0630	0.0643	0.0655	0.0668	-1.5
0.0681	0.0694	0.0708	0.0721	0.0735	0.0749	0.0764	0.0778	0.0793	0.0808	-1.4
0.0823	0.0838	0.0853	0.0869	0.0885	0.0901	0.0918	0.0934	0.0951	0.0968	-1.3
0.0985	0.1003	0.1020	0.1038	0.1056	0.1075	0.1093	0.1112	0.1131	0.1151	-1.2
0.1170	0.1190	0.1210	0.1230	0.1251	0.1271	0.1292	0.1314	0.1335	0.1357	-1.1
0.1379	0.1401	0.1423	0.1446	0.1469	0.1492	0.1515	0.1539	0.1562	0.1587	-1.0
0.1611	0.1635	0.1660	0.1685	0.1711	0.1736	0.1762	0.1788	0.1814	0.1841	-0.9
0.1867	0.1894	0.1922	0.1949	0.1977	0.2005	0.2033	0.2061	0.2090	0.2119	-0.8
0.2148	0.2177	0.2206	0.2236	0.2266	0.2296	0.2327	0.2358	0.2389	0.2420	-0.7
0.2451	0.2483	0.2514	0.2546	0.2578	0.2611	0.2643	0.2676	0.2709	0.2743	-0.6
0.2776	0.2810	0.2843	0.2877	0.2912	0.2946	0.2981	0.3015	0.3050	0.3085	-0.5
0.3121	0.3156	0.3192	0.3228	0.3264	0.3300	0.3336	0.3372	0.3409	0.3446	-0.4
0.3483	0.3520	0.3557	0.3594	0.3632	0.3669	0.3707	0.3745	0.3783	0.3821	-0.3
0.3859	0.3897	0.3936	0.3974	0.4013	0.4052	0.4090	0.4129	0.4168	0.4207	-0.2
0.4247	0.4286	0.4325	0.4364	0.4404	0.4443	0.4483	0.4522	0.4562	0.4602	-0.1
0.4641	0.4681	0.4721	0.4761	0.4801	0.4840	0.4880	0.4920	0.4960	0.5000	-0.0

t-Probability Table



one tail	0.100	0.050	0.025	0.010	0.005
two tails	0.200	0.100	0.050	0.020	0.010
df 1	3.08	6.31	12.71	31.82	63.66
2	1.89	2.92	4.30	6.96	9.92
3	1.64	2.35	3.18	4.54	5.84
4	1.53	2.13	2.78	3.75	4.60
5	1.48	2.02	2.57	3.36	4.03
6	1.44	1.94	2.45	3.14	3.71
7	1.41	1.89	2.36	3.00	3.50
8	1.40	1.86	2.31	2.90	3.36
9	1.38	1.83	2.26	2.82	3.25
10	1.37	1.81	2.23	2.76	3.17
11	1.36	1.80	2.20	2.72	3.11
12	1.36	1.78	2.18	2.68	3.05
13	1.35	1.77	2.16	2.65	3.01
14	1.35	1.76	2.14	2.62	2.98
15	1.34	1.75	2.13	2.60	2.95
16	1.34	1.75	2.12	2.58	2.92
17	1.33	1.74	2.11	2.57	2.90
18	1.33	1.73	2.10	2.55	2.88
19	1.33	1.73	2.09	2.54	2.86
20	1.33	1.72	2.09	2.53	2.85
21	1.32	1.72	2.08	2.52	2.83
22	1.32	1.72	2.07	2.51	2.82
23	1.32	1.71	2.07	2.50	2.81
24	1.32	1.71	2.06	2.49	2.80
25	1.32	1.71	2.06	2.49	2.79
26	1.31	1.71	2.06	2.48	2.78
27	1.31	1.70	2.05	2.47	2.77
28	1.31	1.70	2.05	2.47	2.76
29	1.31	1.70	2.05	2.46	2.76
30	1.31	1.70	2.04	2.46	2.75

one tail	0.100	0.050	0.025	0.010	0.005
two tails	0.200	0.100	0.050	0.020	0.010
df 31	1.31	1.70	2.04	2.45	2.74
32	1.31	1.69	2.04	2.45	2.74
33	1.31	1.69	2.03	2.44	2.73
34	1.31	1.69	2.03	2.44	2.73
35	1.31	1.69	2.03	2.44	2.72
36	1.31	1.69	2.03	2.43	2.72
37	1.30	1.69	2.03	2.43	2.72
38	1.30	1.69	2.02	2.43	2.71
39	1.30	1.68	2.02	2.43	2.71
40	1.30	1.68	2.02	2.42	2.70
41	1.30	1.68	2.02	2.42	2.70
42	1.30	1.68	2.02	2.42	2.70
43	1.30	1.68	2.02	2.42	2.70
44	1.30	1.68	2.02	2.41	2.69
45	1.30	1.68	2.01	2.41	2.69
46	1.30	1.68	2.01	2.41	2.69
47	1.30	1.68	2.01	2.41	2.68
48	1.30	1.68	2.01	2.41	2.68
49	1.30	1.68	2.01	2.40	2.68
50	1.30	1.68	2.01	2.40	2.68
60	1.30	1.67	2.00	2.39	2.66
70	1.29	1.67	1.99	2.38	2.65
80	1.29	1.66	1.99	2.37	2.64
90	1.29	1.66	1.99	2.37	2.63
100	1.29	1.66	1.98	2.36	2.63
150	1.29	1.66	1.98	2.35	2.61
200	1.29	1.65	1.97	2.35	2.60
300	1.28	1.65	1.97	2.34	2.59
400	1.28	1.65	1.97	2.34	2.59
500	1.28	1.65	1.96	2.33	2.59
∞	1.28	1.65	1.96	2.33	2.58