PROBABILITY AND STATISTICS, A24, FINAL EXAMINATION

Name: ________________________ Student number______________________

(1) (3 marks) Morin-Heights is a tourist village known for its outdoor activities. A sample of 417 visitors to Morin Heights during January showed that the average daily spending in the village was 87.40\$ with a sample standard deviation of 33.60\$. Construct a 90% confidence interval and also a 98% confidence interval for the average daily expenditure. Comment on the interaction between confidence and precision in interval estimates for the mean.

$n = 417 \quad \bar{x} = 87.40 \quad s = 33.60$

90% CI : $z_{\alpha/2} = 1.645$

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} = 87.40 \pm 1.645 \frac{33.60}{\sqrt{417}} = 87.40 \pm 2.707$$

$84.69 \leq \mu \leq 90.11$ \$ with 90% confidence. 1

98% CI : $z_{\alpha/2} = 2.325$

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}} = 87.40 \pm 2.325 \frac{33.60}{\sqrt{417}} = 87.40 \pm 3.826$$

$83.57 \leq \mu \leq 91.23$ \$ with 98% confidence. 1

Higher precision => Lower confidence.

Higher confidence => Lower precision. 1

(2) (3 marks) The management of Sommet Morin-Heights estimates that the average time skiers spend on the hill is 3 hours and 40 minutes with a standard deviation of 1 hour and 20 minutes. Hanah has collected a sample of 23 skiing times. The sample standard deviation for the time spent on the hill is 1 hour and 5 minutes. Assume the times on the hill are normally distributed. Test the management's claim on the population standard deviation with an alternative hypothesis that the standard deviation is less than what they assert. Be sure to include bounds for the p-value. Does this sample contradict the management's claim?

$H_0 : \sigma = 80$, $H_1 : \sigma < 80$ $\alpha = 0.05$

Sample : $n = 23$, $s = 65$

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{22 \cdot (65)^2}{80^2} = 14.52 \text{ with } 22 \text{ d.f.}$$ 1

$0.1 < \text{p-value} < 0.5$ Fail to reject H_0. 1

This sample does not provide enough evidence that the time spent by the skiers on the hill varies less than what the management claims. 1

(3) (3 marks) Morin-Heights is also the starting point for Corridor Aérobique which during the winter months is a $58km$ long cross-country skiing trail. Yvan & Co. ski on this trail frequently. The distance covered during these excursions is well-modelled by a random variable with a probability density function

$$f(x) = \begin{cases} (x-12)^2/1152 & 12 \le x < 24 \\ 9/(x-12)^2 & 24 \le x \le 48 \\ 0 & \text{otherwise} \end{cases}$$

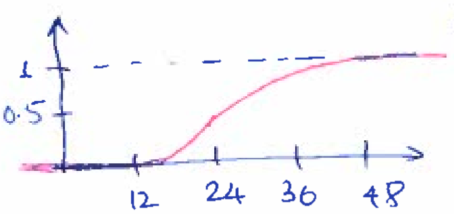
a) Find the cumulative distribution function of the distance covered.
b) Sketch the graph of the cumulative distribution function.
c) Find the ninety-eighth percentile, P_{98}, of the distance covered.
d) (Bonus 1 mark) Determine the average distance covered.

a) On $12 \le x < 24$: $F(x) = \int_{12}^{x} \frac{(t-12)^2}{1152} dt = \frac{(t-12)^3}{3456}\Big|_{t=12}^{t=x} = \frac{(x-12)^3}{3456}$ 0.5

$F(24) = \frac{(24-12)^3}{3456} = 0.5$

On $24 \le x \le 48$: $F(x) = 0.5 + \int_{24}^{x} \frac{9}{(t-12)^2} dt = 0.5 - \frac{9}{(t-12)}\Big|_{t=24}^{t=x} =$ 0.5

$= 0.5 - \frac{9}{x-12} + \frac{9}{24-12} = \frac{5}{4} - \frac{9}{x-12}$; $F(48) = \frac{5}{4} - \frac{1}{4} = 1$

b) 

0.5

$$F(x) = \begin{cases} 0 & x < 12 \\ (x-12)^3/3456 & 12 \le x < 24 \\ 5/4 - 9/(x-12) & 24 \le x \le 48 \\ 1 & x > 48 \end{cases}$$ 0.5

c) $F(x) = 0.98$; $5/4 - \frac{9}{x-12} = 0.98$; $\frac{9}{x-12} = 0.27$; $x - 12 = \frac{9}{0.27} = \frac{100}{3}$

$x = 12 + \frac{100}{3} = \frac{136}{3} = 45.33$ 1

d) $\mu = \int_{12}^{24} x \cdot \frac{(x-12)^2}{1152} dx + \int_{24}^{48} x \cdot \frac{9}{(x-12)^2} dx =$ $\quad y = x - 12$

$= \int_0^{12} (y+12) \frac{y^2}{1152} dy + \int_{12}^{36} (y+12) \frac{9}{y^2} dy$

$= \frac{1}{1152} \left[\frac{y^4}{4} + 12 \frac{y^3}{3} \right] \Big|_{y=0}^{12} + 9 \left[\ln y - \frac{12}{y} \right]_{y=12}^{36}$

$= \frac{1}{1152} \left[\frac{(12)^4}{4} + 4(12)^3 \right] + 9 \left[\ln 36 - \frac{12}{36} - \ln 12 + \frac{12}{12} \right] = \frac{21}{2} + 6 + 9\ln 3$

$= 10.5 + 15.89 = 26.39$ 1

(4) (3 marks) The air temperature at the start of the cross-country skiing for six outings last January was

$$-5.4C^\circ \quad -8.6C^\circ \quad -10.3C^\circ \quad -2.5C^\circ \quad -3.7C^\circ \quad -8.5C^\circ$$

Assume that the starting air temperature is normally distributed.

a) Construct a 95% confidence upper bound for the population mean. Does your result indicate that the population average starting temperature is negative with 95% confidence?

b) Construct a 95% confidence interval for the population standard deviation.

$\bar{x} = \frac{1}{6}(-5.4 - \dots - 8.5) = -6.5$

$s^2 = \frac{1}{5}\left[(-5.4+6.5)^2 + \dots + (-8.5+6.5)^2\right] = 9.58 \qquad s = 3.095$

a) $\mu \leq \bar{x} + t_\alpha \frac{s}{\sqrt{n}} = -6.5 + 2.015 \frac{3.095}{\sqrt{6}} = -3.95\,C^\circ$ with 95° confidence

1

Yes, since upper bound is negative we can claim that the population average starting temperature is negative with 95% confidence.

0.5

b) $$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

$$\frac{5\cdot(9.58)}{12.83} \leq \sigma^2 \leq \frac{5(9.58)}{0.83}$$

1

$3.733 \leq \sigma^2 \leq 57.628$ with 95% confidence

$1.932 \leq \sigma \leq 7.591\ C^\circ$ with 95% confidence.

0.5

(5) (3 marks) Somehow visits to Morin-Heights tend to end up in one of the local eateries. One highly rated choice is Brasserie Anorak. 459 recent diners at Anorak have filled out a survey and have obtained a tombola ticket in return. 48 of these diners have ordered Texas-style beef brisket. Consider the probability, P, that amongst the 12 randomly selected diners who hold winning tombola tickets, two or three have ordered the beef brisket.

i) Compute P.

ii) Estimate P using Poisson approximation.

iii) Estimate P using Gaussian approximation.

Hypergeometric: $N = 459$, $r = 48$, $n = 12$

i) $$p = p(X=2) + p(X=3) = \frac{{}_{48}C_2 \cdot {}_{411}C_{10}}{{}_{459}C_{12}} + \frac{{}_{48}C_3 \cdot {}_{411}C_9}{{}_{459}C_{12}} =$$

$$= 0.2424 + 0.0925 = 0.3349$$ 1

ii) Poisson approximation: $\lambda = n \frac{r}{N} = 12 \cdot \frac{48}{459} = 1.255$

$$p_P(2) + p_P(3) = \frac{e^{-1.255}(1.255)^2}{2!} + \frac{e^{-1.255}(1.255)^3}{3!}$$

$$= 0.2245 + 0.0939 = 0.3184$$ 1

iii) Gaussian approximation:

$$\mu = n\frac{r}{N} = 1.255, \quad \sigma^2 = n\frac{r}{N}\left(1 - \frac{r}{N}\right)\left(\frac{N-n}{N-1}\right) = 1.0967 \quad \sigma = 1.047$$

$$z_1 = \frac{(2-0.5) - 1.255}{1.047} = 0.234 \qquad z_2 = \frac{(3+0.5) - 1.255}{1.047} = 2.144$$

$$P \approx p(z < 2.144) - p(z < 0.234) = 0.9840 - 0.5925 = 0.3915$$ 1

(6) (3 marks) Up Route 329 from Morin-Heights is Saint-Adolphe-d'Howard which has its own ski hill - Mont Avalanche. The management of Mont Avalanche claims that on average skiers spend 4 hours on the hill with a standard deviation of 45 minutes. Diana has collected a sample of 42 skier times on the hill. The sample average is 3 hours and 40 minutes. Assume the times on the hill are normally distributed.

i) Using Diana's sample, test the claim for the population mean: $H_0 : \mu = 4$ versus $H_1 : \mu \neq 4$. Report a p-value and draw a conclusion in the context of the problem.

ii) What is the power of this test to discriminate a true population average time on Mont Avalanche of 3 hours and 30 minutes from the claimed time of 4 hours?

$H_0 : \mu = 4, \quad H_1 : \mu \neq 4$

$\bar{x} = 3.667, \quad n = 42, \quad s = 0.75$

i) $z = \dfrac{3.667 - 4}{0.75/\sqrt{42}} = -2.88$ 0.5

p-value $= 2 \times P(Z < -2.88) = 2 \times 0.0020 = 0.004$ 0.5

Reject H_0. Accept H_1.

The mean time spent by skiers on the hill is not 4h. 0.5

ii) $\mu_t = 3.5 \quad z_{\alpha/2} = 1.96$

$x_{1,cv} = 4 - 1.96 \dfrac{0.75}{\sqrt{42}} = 3.773 \qquad z_1 = \dfrac{3.773 - 3.5}{0.75/\sqrt{42}} = 2.36$

$x_{2,cv} = 4 + 1.96 \dfrac{0.75}{\sqrt{42}} = 4.227 \qquad z_2 = \dfrac{4.227 - 3.5}{0.75/\sqrt{42}} = 6.28$ 0.5

$\beta = P(2.36 < Z < 6.28) = 1 - 0.990863 = 0.009137$ 0.5

Power $= 1 - \beta = 0.990863$ 0.5

(7) (3 marks) The village Saint-Adolphe-d'Howard is built around Lac-Saint-Joseph. During the winter months, Lac Saint-Joseph is a popular destination for ice fishing. Many species of fish that can be caught through the ice, including yellow perch, walleye, pike, and trout. On a sunny winter day, a group of anglers on the lake ice are catching perch at the rate of 5 per hour. Another group is independently catching trout at the rate of 2 per hour.

a) What is the probability that the perch anglers will catch 12 perch in the next two hours?

b) What is the probability that the trout anglers will take less than half an hour to catch 3 trout?

c) What is the probability that either the perch anglers will catch three fish or that the trout anglers will catch two fish in less than half an hour?

a) Poisson: $\lambda = 2 \cdot 5 = 10$

$$P(12) = \frac{e^{-10}(10)^{12}}{12!} = 0.09478$$ 0.5

b) Erlang: $\lambda = 2(0.5) = 1 \quad r = 3$

$$P(X < 0.5) = 1 - \sum_{k=0}^{2} \frac{e^{-1}(1)^k}{k!} = 1 - \left(e^{-1} + e^{-1} + \frac{e^{-1}}{2}\right) = 0.0803$$ 1

c) Perch anglers: Erlang $(2.5, 3)$:

$$P_1 = 1 - \sum_{k=0}^{2} \frac{e^{-2.5}(2.5)^k}{k!} = 1 - e^{-2.5}\left(1 + (2.5)^1 + \frac{(2.5)^2}{2!}\right) = 0.4562$$ 0.5

Trout anglers: Erlang $(1, 2)$

$$P_2 = 1 - \sum_{k=0}^{1} \frac{e^{-1}(1)^k}{k!} = 1 - e^{-1}(1 + 1) = 0.2642$$ 0.5

$$P = P_1 + P_2 - P_1 P_2 = 0.4562 + 0.2642 - (0.4562)(0.2642) = 0.6$$ 0.5

(8) (3 marks) Just north of Saint-Adolphe-d'Howard is a luxurious chalet named Les Scandaleaux. Based on historical ratings the following table gives the probability distribution for Google ratings given by guests of Les Scandaleaux.

Rating: x	1	2	3	4	5
$P(x)$	0	0.02	0.13	0.32	0.53

i) Compute the average rating and the standard deviation.

ii) Stephane, the chalet keeper, gets a Christmas bonus if the average of the last 50 Google ratings exceeds 4. Estimate the probability that Stephane will get a Christmas bonus.

i) $E(X) = \sum x\, p(x) = 1(0) + \dots + 5(0.53) = 4.36$ 0.5

$Var(X) = \sum x^2 p(x) - E(X)^2 = 1^2(0) + \dots + 5^2(0.53) - (4.36)^2 = 0.6104$ 0.5

$\sigma(X) = 0.7813$ 0.5

ii) $z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \dfrac{4 - 4.36}{0.7813/\sqrt{50}} = -3.26$ 1

$p(\text{Christmas Bonus}) = p(z > -3.26) = 1 - 0.000557 = 0.999443$ 0.5

(9) (3 marks) Further north from Saint-Adolphe-d'Howard on Route 329 is the 'big' town of Sainte-Agathe-des-Monts. There are two restaurants on rue St Vincent in Sainte-Agathe-des-Monts. A sample of $n_1 = 32$ diners at restaurant Maison 1890 had an average bill of $\bar{x}_1 = 101\$$ with a standard deviation of $s_1 = 49\$$, and a sample of $n_2 = 64$ diners at restaurant Mikael had an average bill of $\bar{x}_2 = 32\$$ with a standard deviation of $s_2 = 11\$$. Assume the populations are normally distributed.

a) Implement a test of the hypothesis $H_0 : \sigma_1 = \sigma_2$ versus $H_1 : \sigma_1 > \sigma_2$ at $\alpha = 0.05$ level of significance. Make sure to report a range for the p-value and draw a conclusion in the context of the problem.

b) Implement a test of the hypothesis $H_0 : \mu_1 = \mu_2$ versus $H_1 : \mu_1 > \mu_2$ at $\alpha = 0.05$ level of significance. Make sure to report a range p-value and draw a conclusion in the context of the problem.

a) $f = \dfrac{s_1^2}{s_2^2} = \left(\dfrac{49}{11}\right)^2 = 19.84$ with $(31, 63)$ df 0.5

p-value $<< 0.01$ 0.5

Reject H_0. Accept H_1.

The two restaurants average bills have significantly different variances. 0.5

b) $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}} = \dfrac{101 - 32}{\sqrt{\dfrac{(49)^2}{32} + \dfrac{(11)^2}{64}}} = 7.87$ 0.5

$$\nu = \frac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{(s_1^2/n_1)^2}{n_1 - 1} + \dfrac{(s_2^2/n_2)^2}{n_2 - 1}} = \frac{\left(\dfrac{49^2}{32} + \dfrac{11^2}{64}\right)^2}{\dfrac{(49^2/32)^2}{31} + \dfrac{(11^2/64)^2}{63}} = 32.6 \longrightarrow 32 \text{ d.f.}$$

0.5

p-value $<< 0.0005$

Reject H_0. Accept H_1.

Maison 1890 is significantly more expensive than restaurant Mikael. 0.5

(10) (3 marks) Lac-des-Sables is the heart of Sainte-Agathe-des-Monts. During the winter months, the time spent skating on the frozen lake by visitors is well modelled by the random variable X with pdf $p(x) = 25/(48x^3)$, $0.5 \leq x \leq 2.5$, where x is in hours. Say, the amount of hot chocolate visitors buy at Couleur Café Signature after their skate is $Y = X^2$, where Y is in hundreds of milliliters. Check the 'Theorem of the Unconscious Statistian' on this example. Determine $E(X^2)$ in two ways:

a) Without finding the pdf of $Y = X^2$.

b) By first computing the pdf of $Y = X^2$.

a) $E(X^2) = \int_{0.5}^{2.5} x^2 \cdot \frac{25}{48x^3}\, dx = \frac{25}{48} \ln x \Big|_{0.5}^{2.5} = 0.8382$ (1)

b) On the interval $0.5 \leq x \leq 2.5$ the function $y = x^2$ is bijective.

$p(y) = p(x) \frac{dx}{dy} = \frac{25}{48} \frac{1}{y^{3/2}} \cdot \frac{d}{dy}(\sqrt{y}) = \frac{25}{96} \frac{1}{y^2}$ (1)

$E(Y) = \int_{(0.5)^2}^{(2.5)^2} y \cdot \frac{25}{96} \frac{1}{y^2}\, dy = \frac{25}{96} \ln y \Big|_{(0.5)^2}^{(2.5)^2} = 0.8382$ (1)

(11) (3 marks) Further north is the village of Sainte-Lucie-des-Laurentides. Sainte-Lucie has two depanneurs. For a randomly selected resident, let X be the number of weekly trips to the first depanneur and Y be the number of weekly trips to the second depanneur. Suppose that the joint pmf of X and Y is given by the accompanying table:

		Y			
	$p(x,y)$	0	1	2	$P(x)$
	0	0.06	0.03	0.01	0.10
X	1	0.20	0.18	0.12	0.50
	2	0.15	0.14	0.11	0.40
	$P(y)$	0.41	0.35	0.24	1

0.5

a) Compute the marginal probability distributions of X and Y.

b) Compute the conditional probability mass function of Y given that $X = 1$.

c) Compute the conditional mean of Y given $X = 1$. Write a sentence in English interpreting your findings.

d) Compute the correlation between the RV's X and Y. Interpret the value in the context of the problem.

b)

Y	0	1	2
$P(y\mid x=1)$	$\frac{0.2}{0.5}$	$\frac{0.18}{0.5}$	$\frac{0.12}{0.5}$

$\Rightarrow$

Y	0	1	2
$P(y\mid x=1)$	0.4	0.36	0.24

0.5

c) $E(Y \mid X=1) = 0(0.4) + 1(0.36) + 2(0.24) = 0.84$ 0.5

A Sainte-Lucie resident who makes one weekly trip to the 1st depanneur will make on average 0.84 trips per week to the 2nd depanneur.

d) $\mu_X = 0(0.10) + 1(0.50) + 2(0.40) = 1.3$

$\sigma_X^2 = 0^2(0.10) + 1^2(0.50) + 2^2(0.40) - (1.3)^2 = 0.41$

$\mu_Y = 0(0.41) + 1(0.35) + 2(0.24) = 0.83$

$\sigma_Y^2 = 0^2(0.41) + 1^2(0.35) + 2^2(0.24) - (0.83)^2 = 0.6211$ 0.5

$\sigma_{XY} = (1)(1)(0.18) + (1)(2)(0.12) + (2)(1)(0.14) + (2)(2)(0.11) - (1.3)(0.83) = 0.061$

$\rho_{XY} = \dfrac{0.061}{\sqrt{(0.41)(0.6211)}} = 0.121$ This is a weak positive correlation. 0.5

Residents who visit one of the depanneurs are slightly more likely to visit the other, but the relationship is not strong. 0.5

(12) (3 marks) Sainte-Lucie is a popular starting location for a moderately challenging hike to Mont Kaaikop. There are two trails from Sainte-Lucie to Mont Kaaikop: Western and Eastern. It has been disputed for decades in Sainte-Lucie which trail is longer. Finally, a statistical savant collected samples of times to track from the village to the top of the mountain and computed the following sample averages: Western trail: $\bar{x}_1 = 293min$ in $n_1 = 22$ tracks; Eastern trail: $\bar{x}_2 = 301min$ in $n_2 = 25$ tracks. Assuming the tracking times are normally distributed with population standard deviations of $\sigma_1 = \sigma_2 = 9$ minutes, test the hypothesis that the Eastern trail takes longer to traverse than the Western trail at the $\alpha = 0.01$ level of significance. Report the p-value. Conclude in the context of the problem.

$H_0: \mu_1 - \mu_2 = 0 \qquad H_1: \mu_1 - \mu_2 < 0$

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{293 - 301}{\sqrt{\frac{9^2}{22} + \frac{9^2}{25}}} = -3.04 \qquad 1$$

$$p\text{-value} = P(z < -3.04) = 0.001183 < \alpha = 0.01 \qquad 1$$

Reject H_0. Accept H_1.

The Eastern trail to Mont Kaaikop takes longer to traverse than the Western trail. 1