UNIVERSITY OF TORONTO SCARBOROUGH

Department of Computer and Mathematical Sciences

Midterm Test, Winter - 2017

STAB57H3: An Introduction to Statistics

Duration: Two hours (120 minutes)

LAST NAME: ______________________ FIRST NAME: ________________________

STUDENT NUMBER: _________________ SIGNATURE: ________________________

TUTORIAL: __________________________

Aids Allowed:

- A handwritten cheat-sheet, covering both sides of an A4/Letter sized paper. You need to submit your cheat-sheet with your answer-sheet after the test.
- A calculator (No phone calculators are allowed)

No other aids are allowed.

All your work must be presented clearly in order to get credit. Answer alone (even though correct) will only qualify for ZERO credit. Please show your work in the space provided; you may use the back of the pages, if necessary, but you MUST remain organized. Show your work and answer in the space provided, in ink. Pencil may be used, but then any re-grading will **NOT** be allowed.

There are 14 **pages including this page. Please check to see you have all the pages.**

Good Luck!

Question	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**Total**
Points	14	14	14	14	12	17	10	95
Score								

1 (a): An average of 0.8 accident occur per day in a particular large city. Let X denote the number of accidents that will occur in this city on a given day. Consider that the distribution of X is Poisson(λ), where the parameter $\lambda > 0$. So, what value would you record as a prediction of a future value of X?

[5 Points]

1 (b): Find the probability that the number of accident that will occur in this city on a given day is at most 2.

[3 Points]

1 (c): You are informed that $X > 2$. What value would you record as a prediction of a future value of X?

[6 Points]

2 (a): The following data give the number of orders received for a sample of 30 hours at the Timesaver Mail Order Company. Construct a five number summary table, and

34	44	31	52	41	47
38	35	32	39	28	24
46	41	49	53	57	33
27	37	30	27	45	38
34	46	36	30	47	50

produce a boxplot.

[4 + 3 = 7 Points]

2 (b): Construct a density histogram using the following intervals:

$$(20, 28], (28, 36], (36, 44], (44, 52], (52, 60].$$

[5 Points]

2 (c): Comment on the shape of the distribution from 2(a) and 2(b). [2 Points]

3 (a): Let $(x_1, x_2, \cdots, x_n)$ be a random sample from a Gamma(α_0, β) distribution, where α_0 is known and β is unknown. Using Factorization Theorem, find *sufficient statistic* for β.

[4 Points]

3 (b): Derive the maximum likelihood estimator (MLE) of β.

[6 Points]

3 (c): The following data give the times (in minutes) taken by 18 students to complete a statistics examination that was given a maximum time of 75 minutes to finish. Let

41	28	45	60	53	69
70	50	63	68	37	44
42	38	74	53	66	65

the random variable *times taken by the students* X follows a Gamma($\alpha_0 = 14.99, \beta$) distribution. Calculate the MLE of β.

[4 Points]

4 (a): Let $(X_1, X_2, \cdots, X_n)$ be a random sample of size n from a $N(\mu, \sigma^2)$ distribution, where $\mu \in R^1$ and $\sigma^2 > 0$ are unknown. Show that

$$E(\bar{X}) = \mu.$$

[3 Points]

4 (b): Show that

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}.$$

[4 Points]

4 (c): A random sample of 14 participants in a Zumba dance class had their heart rate measured before and after a moderate 10-minute workout. The following data correspond to the increase in each individual's heart rate (in beats per minute):

59	70	57	42	57	59	41
54	44	36	59	61	52	42

Construct a 0.95-confidence interval for the average increase in person's heart rate after a moderate 10-minute Zumba workout. Interpret your confidence interval.

[5 + 2 = 7 Points]

5 (a): A past study claimed that adults in America spent an average of 18 hours a week on leisure activities. A researcher wanted to test this claim. She took a sample of 12 adults and asked them about the time they spend per week on leisure activities. Their responses (in hours) are as follows. Assume that the times spent on leisure activities

13.6	14.0	24.5	24.6	22.9	37.7
14.6	14.5	21.5	21.0	17.8	21.4

by all American adults are normally distributed with $\sigma_0 = 4$. Calculate the sample mean of times spent on leisure activities.

[3 Points]

5 (b): Calculate p-value to test the claim of the study against alternative that the average amount of time spent by American adults on leisure activities has changed.

[7 Points]

5 (c): Write down your conclusion based on $5(a)$ and $5(b)$.

[2 Points]

6 (a): Suppose that $(x_1, x_2, \cdots, x_n)$ is a random sample from a $N(\mu_0, \sigma^2)$ distribution with μ_0 is known and σ^2 is unknown. If the prior distribution of $1/\sigma^2$ is Gamma(α_0, β_0), then derive the posterior of $1/\sigma^2$.

[7 Points]

6 (b): A researcher wanted to estimate the mean contributions made to charitable causes by major companies. A random sample of 18 companies produced the following data on contributions (in millions of dollars) made by them. It is assumed that the distribution

1.8	0.6	1.2	0.3	2.6	1.9
3.4	2.6	0.2	2.4	1.4	2.5
3.1	0.9	1.2	2.0	0.8	1.1

of contributions made to charitable causes X is $N(\mu_0 = 1.5, \sigma^2)$. If the prior distribution of $1/\sigma^2$ is Gamma$(\alpha_0 = 11, \beta_0 = 10)$, then determine the posterior distribution of $1/\sigma^2$. Calculate, posterior mean of $1/\sigma^2$.

[4 + 2 = 6 Points]

6 (c): Calculate 0.99-credible interval for $1/\sigma^2$. Interpret your interval. [2 + 2 = 4 Points]

[Please CHECK the correct answer in each question. An incorrect answer will PENALIZE 1 mark in each.]

7 (a): A *likelihood function* $L(\theta|x_1, \cdots, x_n)$ is

(a) a function of the sample $(x_1, \cdots, x_n)$ given the parameter θ.

(b) the maximum of the joint density function $f_\theta(x_1, \cdots, x_n)$.

(c) the product of the marginal density functions $f_\theta(x_i); \quad i = 1, \cdots, n$.

(d) the function of the parameter θ given the observed sample $(x_1, \cdots, x_n)$.

[2 Points]

7 (b): The most important use of *sufficient statistic* is

(a) the construction of confidence interval.

(b) to test the null hypothesis $H_0 : \theta = \theta_0$ against alternative $H_A : \theta \neq \theta_0$.

(c) data reduction.

(d) drawing a simple random sample from a population.

[2 Points]

7 (c): In *classical* and *bayesian* statistical methods, the parameter θ is

(a) non-random and fixed.

(b) a random variable.

(c) non-random and random, respectively.

(d) random and non-random, respectively.

[2 Points]

7 (d): A *confidence interval* is concerned with

(a) assigning a single value to a population parameter.

(b) assigning two values to a population parameter.

(c) assigning a range of values to a population parameter and attaching appropriate degrees of confidence to the interval of containing the true parameter.

(d) testing the null hypothesis H_0 that the parameter θ is equal to a particular value θ_0.

[2 Points]

7 (e): Everything else being fixed, the *width* of a confidence interval for the mean of a normal distribution decreases with the increase of sample size n. **TRUE / FALSE**

[2 Points]

Appendix

1. The probability mass function of a Poisson(λ) distribution is following:

$$P_\lambda(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad ; \quad x = 0, 1, 2, 3, \cdots$$

2. The probability density function of $X \sim$ Gamma(α, β) is

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x},$$

   where $\alpha > 0$, $\beta > 0$, and $x \in R^+$.

3. The 0.975th quantile of a t distribution with 13 degrees of freedom is

$$t_{0.975}(13) = 2.160369.$$

4. The cumulative probability of standard normal distribution evaluated 2.316618 is

$$\Phi(2.316618) = 0.9897377.$$

5. The probability density function of $X \sim N(\mu, \sigma^2)$ is

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\},$$

   where $\mu \in R^1$, $\sigma^2 \in R^+$, and $x \in R^1$.

6. The 0.005 and 0.995 quantiles of the Gamma(20, 18.02) distribution are 0.575 and 1.853, respectively.