



Université d'Ottawa • University of Ottawa

Faculté des sciences
Mathématiques et de statistique

Faculty of Science
Mathematics and Statistics

MAT 3377 Final Exam

July 26, 2019
Time: 3 hours

Professor M. Alvo

Student number : _____

Student Name:

| | | | | | | | | |
|----------|---|---|---|---|---|---|---|-------|
| Question | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Total |
| Marks | | | | | | | | |

This is an **open book** examination **out of 60**. Only official faculty calculators are permitted. Write your answers in the booklet(s) provided.

- Cellular phones, unauthorized electronic devices or course notes (unless an open-book exam) are not allowed during this exam. Phones and devices must be turned off and put away in your bag. Do not keep them in your possession, such as in your pockets. If caught with such a device or document, the following may occur: academic fraud allegations will be filed which may result in you obtaining a 0 (zero) for the exam. By signing below, you acknowledge that you have ensured that you are complying with the above statement.
- Signature :

1. (10 Points) Canada Post would like to estimate the average daily volume of letter mail received from the US for delivery in Canada. This forms the basis for the exchange of tariffs. The data below represent a sample of 10 containers received in Montreal which weigh x kg each and which contain a volume of y letters.

- Estimate the average volume per kg using a 95 confidence interval.
- If the total weight of letter mail received in Montreal on a given day is 1076 kg, estimate the total volume of mail using a 95% confidence interval.
- What is your estimate of the bias for estimation of the mean volume?

Postes Canada voudrait estimer le volume moyen par jour de courrier par lettres provenant des États Unis pour livraison au Canada. Ceci forme la base des échanges financiers. Les données suivantes représentent un échantillon de 10 contenus reçus à Montréal qui pèsent x kg et qui contiennent un volume de y lettres.

- Estimer le volume moyen par kg à l'aide d'un intervalle de confiance à 95%
- Si le poids total des lettres reçues à Montréal pour une certaine journée est de 1076 kg, estimez le volume total de lettres à l'aide d'un intervalle de confiance à 95%
- Quel serait votre mesure du biais pour le volume moyen?

| | | | | | | | | | | |
|------------|------|------|------|------|------|------|------|------|------|------|
| Volume y | 228 | 330 | 385 | 342 | 374 | 445 | 258 | 101 | 166 | 345 |
| Weight x | 3.22 | 4.91 | 5.94 | 6.06 | 6.02 | 6.61 | 5.21 | 2.25 | 2.78 | 6.30 |

$\bar{y} = 297.4$, $s_y^2 = 11428$, $\bar{x} = 4.93$, $s_x^2 = 2.555$, $\sum_{i=1}^{10} (x_i - \bar{x})(y_i - \bar{y}) = 1451.38$

2. (10 Points) We wish to verify the truthfulness of the label on a certain bottled beverage. We select a simple random sample of 4 boxes (clusters) containing 6 bottles each and we record the weight of each bottle. The data as well as the analysis of variance (ANOVA) table are displayed above.

- Estimate the average content of a bottle and provide the standard error of your estimate.
- How many cartons must we sample in the future if we wish to estimate the total content of a carton with an error of estimation not exceeding 0.10 ounces with a level of confidence equal to 95%?
- What is the efficiency of this method of estimation compared to a simple random sampling?

d) What is the value of the intra correlation coefficient? What are your conclusions?

On cherche à vérifier que le contenu des bouteilles d'un certain breuvage est tel qu'indiqué sur la bouteille. On pise 4 cartons (grappes) qui contiennent 6 bouteilles chaque et on observe les poids suivants en onces.

| Box | | | | | | | Average | Variance |
|-----|------|------|------|------|------|------|---------|----------|
| 1 | 16.1 | 15.9 | 16.1 | 16.2 | 15.9 | 15.8 | 16.000 | 0.0240 |
| 2 | 15.9 | 16.2 | 15.8 | 16 | 16.3 | 16.1 | 16.050 | 0.0350 |
| 3 | 16.2 | 16 | 15.7 | 16.3 | 15.8 | 16 | 16.000 | 0.0520 |
| 4 | 15.9 | 16.1 | 16.2 | 16.1 | 16.1 | 16.3 | 16.117 | 0.0177 |

The table below exhibits the analysis of variance results.

Le tableau ANOVA est le suivant

| Source | Sum of Squares | df | MS |
|------------------|----------------|----|--------|
| Between clusters | 0.0550 | 3 | 0.0183 |
| Within clusters | 0.6433 | 20 | 0.0322 |
| Total | 0.6983 | 23 | |

a) Calculez un estimé du contenu moyen d'une bouteille ainsi que son erreur standard.

b) Combien de cartons faut-il échantillonner à l'avenir si on veut estimer le contenu total d'un carton avec une erreur d'estimation d'au plus égale à 0.10 onces avec un niveau de confiance égale à 95%?

c) Quelle est l'efficacité de ce mode d'échantillonnage en comparaison avec un échantillon aléatoire simple?

d) Quelle est la valeur du coefficient de corrélation à l'intérieur des grappes? Quelle est votre conclusion?

3. (10 Points) The company Apple would like to estimate the demand for a new product before placing it on sale on an international scale. Products are usually sold to $N = 220$ wholesalers spread to North America, Europe and Asia as indicated in the table above. It is decided to place the product for sale at $n = 15$ wholesalers allocated proportionally as indicated. The sample means and variances of the sales are shown as well.

- a) Calculate an estimate of the total sales and provide its variance.
- b) Calculate an estimate of the average wholesale sales and provide its variance.
- c) What would have been the optimal sample sizes in the strata if the costs of sampling are identical and $n = 15$? What would be the optimal variance in part a) in that case?

La compagnie Apple veut estimer la demande pour un nouveau produit avant de le placer en vente sur le plan international. Le produit est généralement vendu à $N = 220$ grossistes réparties au monde selon le tableau ci-dessous. La compagnie décide de placer le produit chez $n = 15$ grossistes choisis approximativement selon une allocation proportionnelle. Les moyennes et variances des ventes en millions de \$ se trouvent dans le tableau ci dessous.

- a) Calculez un estimé des ventes en totale et calculez un estimé de la variance.
- b) Calculez un estimé des ventes moyennes par grossiste et calculez un estimé de la variance.
- c) Quelle serait le choix des tailles des échantillons pour une allocation optimale si $n = 15$ et les frais d'échantillonnage sont identiques pour chaque strate? Quelle serait la variance optimale dans la partie a) dans ce *càs*?

| | North America | Europe | Asia |
|---------------------|---------------|--------|-------|
| N_h | 112 | 68 | 40 |
| n_h | 7 | 5 | 3 |
| average \bar{y}_h | 199 | 100 | 98 |
| variance s_h^2 | 78.67 | 55.60 | 39.50 |
| | | | |

4. (10 Points) In a small college having a student population of 4500, we wish to estimate the weekly average amount spent on alcohol by sampling 30 students by choosing a systematic sample of every 150th student from an alphabetical list.

- a) If we observe for the sample $\{y_i\}$ the quantities $\sum_i y_i = 850$, $\sum_i y_i^2 = 33,904$, calculate an estimate of the average weekly amount spent on alcohol along with a corresponding 95% confidence interval.
- b) Suppose we take 4 separate systematic samples of the same size as in part a) and we observe the following weekly means: \$30, \$22, \$62, \$28. What then is your estimate of the average weekly amount spent on alcohol and what is the variance of your estimate?

c) What are the advantages of systematic sampling?

Dans un petit collège avec une population étudiante de 4500, on choisit systématiquement de 30 étudiant en prenant chaque 150^{ième} étudiant à partir d'une liste alphabétique afin d'estimer les dépenses moyennes par semaine pour la boisson.

a) Si pour l'échantillon on observe $\{y_i\}$, $\sum_i y_i = 850$, $\sum_i y_i^2 = 33,904$, calculez un estimé des dépenses moyennes hebdomadaires ainsi qu'un intervalle de confiance à 95%.

b) Supposons qu'on prend 4 échantillons systématiques de la même taille que dans a) et on obtient les moyennes suivantes: \$30, \$22, \$62, \$28. Quel serait l'estimé des dépenses moyennes par semaine dans ce cas et quelle serait sa variance?

c) Quels sont les avantages d'un échantillonnage systématique?

5. (10 Points) A forester would like to estimate the total number of infected trees in a certain forest which is split into $N = 10$ large regions. The regions are further sub-divided into smaller sub-regions for ease of management. Four regions are chosen at random from among the 10. Six sub-regions are chosen at random from each of the chosen regions and the number of infected trees $\{y_{ij}\}$ are counted. The data and some calculations along with output are given below.

a) Estimate the total number of infected trees by means of a 95% confidence interval.

b) Estimate the average number of infected trees by means of a 95% confidence interval.

On veut estimer le nombre total d'arbres infectés dans une certaine forêt que l'on divise en $N = 10$ grandes régions. Chaque région est divisée en sous-régions pour des raisons administratives. On choisit au hasard 4 régions parmi les 10. Ensuite, on choisit 6 sous-régions au hasard de chaque région et on compte le nombre $\{y_{ij}\}$ d'arbres infectés dans chacune. On retrouve ci-dessous les données

ainsi que le tableau ANOVA qui correspond.

| Region i | M_i | y_{ij} | \bar{y}_i | s_i^2 |
|------------|-------|-----------------------|-------------|---------|
| | | | | |
| 1 | 12 | 15, 14, 21, 13, 9, 10 | 13.667 | 18.27 |
| 2 | 15 | 4, 6, 10, 9, 8, 5 | 7 | 5.6 |
| 3 | 14 | 10, 11, 14, 10, 9, 15 | 11.5 | 5.9 |
| 4 | 21 | 8, 3, 4, 1, 2, 5 | 3.833 | 6.17 |
| | | | | |

Note that

$$\frac{1}{n} \sum M_i^2 \left(1 - \frac{6}{M_i}\right) \frac{s_i^2}{6} = 194.77143$$

ANOVA

| Source | SS | d.f. | MS |
|-----------------|--------|------|--------|
| | | | |
| Between regions | 352.33 | 3 | 117.44 |
| Within regions | 179.67 | 20 | 8.98 |
| Total | 532.0 | 23 | |
| | | | |

a) Estimez le nombre total d'arbres infectés avec l'aide d'un intervalle de confiance à 95%.

b) Estimez le nombre moyen d'arbres infectés avec l'aide d'un intervalle de confiance à 95%.

6. (5 Points) A survey is taken at the University of Ottawa in order to determine the proportion of university professors who believe in angels. A sample of 300 professors consisting of 210 men and 90 women is taken. It is found that 80 professors believe in angels.

a) Estimate by means of a 95% confidence interval the proportion of professors who believe in angels.

b) It is noticed that subsequent to the survey, there were 30 male and 50 female professors who believe in angels. However, at the university, 30% of the professors are female and 70% are male. Taking into account this additional information, estimate by means of a 95% confidence interval the fraction of professors who believe in angels.

On mène un sondage à l'Université d'Ottawa afin de déterminer la proportion de professeurs qui croient aux anges. Dans un échantillon de 300 professeurs dont 210 sont hommes et 90 femmes, on observe 80 qui croient aux anges.

a) Estimez avec un intervalle de confiance à 95% la proportion de professeurs qui croient aux anges.

b) On s'aperçoit que dans l'échantillon, il y avait 30 hommes et 50 femmes qui croient aux anges. Par contre, à l'université, il y a 30% de femmes et 70% d'hommes qui sont professeurs. Estimez avec un intervalle de confiance à 95% la proportion de professeurs qui croient aux anges tout en tenant compte de ces informations supplémentaires.

7. (5 Points) a) Describe sampling proportional to size in the case of two-stage cluster sampling. What are the advantages?

b) Give at least one concrete example (other than those presented in class) where this type of sampling would be advantageous.

a) Décrivez l'échantillonnage proportionnel à la grandeur (pps) dans le cas d'un échantillon par grappes en deux stages. Quels sont les avantages?

b) Donnez au moins un exemple concret (autre que ceux présentés au cours) où ce genre d'échantillonnage serait avantageux.