



Université d'Ottawa • University of Ottawa

Faculté des sciences Faculty of Science
Mathématiques et de statistique Mathematics and Statistics

MAT 2377 Final Exam

July 2016
Time: 3 hours

Professor M. Alvo

- This is an open book examination. Only official faculty calculators are permitted.
- Record your answers directly on this questionnaire in the boxes indicated. Marks will be given *only* for answers in the appropriate boxes; no rough work will be graded. Each correct answer is worth 3 marks. Several questions have been translated into French.

Question	Answer	Question	Answer	Question	Answer	Question	Answer
1	E D	6	E	11	E	16	D
2	B	7	C	12	D	17	D
3	B	8	E	13	C	18	A
4	D	9	C	14	C	19	F
5	A	10	B A	15	C	20	E

use estimating difference \bar{x}_1 & \bar{x}_2 w/ σ_1^2, σ_2^2 unknown,
 long formula for $v = (df)$
 (round to nearest int) ↗



1. Scores on a standardized test in statistics taken by students from large and small schools are normally distributed with unknown means and unknown common variance. A random sample of 9 students from a large school yielded $\bar{x} = 81.31$ and $s_X^2 = 60.76$ whereas a random sample of 15 students from a small school yielded $\bar{y} = 78.61$ and $s_Y^2 = 48.24$. Find a 95% confidence interval for the difference in the means $\mu_X - \mu_Y$. (The answer is up to 2 decimals)

Les notes sur un test standardisé en statistique pris par des étudiants de grandes et petites écoles sont distribuées normalement avec des moyennes inconnues et variance commune inconnue. Un échantillon aléatoire de 9 élèves d'une grande école a donné $\bar{x} = 81.31$ et $s_X^2 = 60.76$ alors qu'un échantillon aléatoire de 15 élèves d'une petite école a donné $\bar{y} = 78.61$, $s_Y^2 = 48.24$. Trouver un intervalle de confiance de 95 % pour la différence $\mu_X - \mu_Y$ entre les moyennes. (La réponse est à 2 décimales)

- a) (-3.2, 8.9) b) (-2.81, 8.23) c) (2.56, 7.96) **d) (-4.03, 9.43)**
 e) (-3.65, 9.05)

2. A printing office has 3 separate printers, each receiving a certain percentage of contracts and each having a certain percentage of delays in their outputs in accordance with the table below. A client notices that his document has incurred a delay of more than one month. What is the probability that the document was printed by printer #3?

Un bureau de publication a trois imprimantes pour ses services. Le bureau a recueilli les données suivantes sur les imprimantes. Un client remarque qu'un de leurs documents est plus d'un mois en retard. Quelle est la probabilité que ce document soit la responsabilité de l'imprimante #3?

Printer	Percentage of contracts	Percentage delayed
1	20%	10%
2	30%	40%
3	50%	20%

$0.5 \cdot 0.2$
 $0.5 \cdot 0.2 + 0.3 \cdot 0.4 + 0.2 \cdot 0.1$

- a) $\frac{1}{10}$ **b) $\frac{5}{12}$** c) $\frac{1}{2}$ d) $\frac{2}{7}$ e) $\frac{3}{5}$

3. The joint density function of X and Y is given below:

$$f(x, y) = \begin{cases} x + y, & 0 < x < 1, 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

$$\int_0^1 \int_0^1 x + y \, dx \, dy$$

Find the cumulative distribution function $F(u, v)$ at values $0 < u < 1, 0 < v < 1$.

La densité conjointe de X et de Y est donnée ci-dessus. Calculez la fonction de distribution cumulative $F(u, v)$ aux valeurs $0 < u < 1, 0 < v < 1$.

- a) $uv + v^2/2$ b) $uv(u + v)/2$ c) $(u^2v + uv^2)/4$ d) $2uv$ e) $u + v$

$$\begin{aligned} & \int_0^1 \left(\frac{x^2}{2} + xy \right) dy \\ &= \frac{x^2}{2} \cdot y + x \cdot \frac{y^2}{2} \\ &= xy(x + y)/2 \end{aligned}$$

4. A survey is to be conducted to estimate the proportion of individuals in a city who favor a particular political candidate. What is the maximum sample size required to ensure that the error in the estimate is at most 3% when the level of confidence is 90%?

(Round up your decimal answer to the next integer)

- a) 1000 b) 300 c) 267 d) 752 e) 1068

$$\begin{aligned} \max n &= \frac{1}{4} \left(\frac{Z_{\alpha/2}}{e} \right)^2 \\ &= \frac{1}{4} \left(\frac{1.645}{0.03} \right)^2 \end{aligned}$$

5. The percentage of males in 1986 who are 18-19 years old and married was 3.7%. To test whether or not this percentage increased in 2015, a random sample of 300 males aged 18-19 is taken and it is noted that 20 are married. Using a level of significance of 0.05 test the hypothesis H_0 that the percentage married is 3.7% against the alternative that it is larger. The observed value of the test statistic and the conclusion are:

Le pourcentage d'hommes en 1986 qui ont 18-19 ans et sont mariés a été de 3,7 %. Pour tester si ce pourcentage est plus grand en 2015, on prélève un échantillon aléatoire de 300 hommes âgés de 18-19, et il est à noter que 20 sont mariés. À l'aide d'un niveau de signification de 0,05 tester l'hypothèse H_0 que le pourcentage est de 3,7 % contre l'alternative qu'il est plus grand. La valeur de la statistique utilisée pour ce test ainsi que la conclusion sont:

- a) 2.722, reject H_0 b) 2.722, do not reject H_0
c) 3.292, reject H_0 d) 3.182, reject H_0

$$\begin{aligned} H_0: P &= 0.037 \quad \alpha = 0.05 \\ H_1: P &> 0.037 \quad n = 300 \\ \hat{p} &= \frac{20}{300} = \frac{1}{15} \\ Z &= \frac{\hat{p} - P_0}{\sqrt{\frac{P_0(1-P_0)}{n}}} \\ &= \frac{\frac{1}{15} - 0.037}{\sqrt{\frac{0.037(1-0.037)}{300}}} \\ &= 2.722 \\ Z &> Z_{0.05} \quad \leftarrow 1.645 \\ \Rightarrow & \text{reject } H_0 \end{aligned}$$

Any linear function of 2 normals is also normal

$X: \mu_x = 5, \sigma_x^2 = 9$
 $Y: \mu_y = 3, \sigma_y^2 = 16$
 X, Y are normal so $X - Y$ is normal:

$P(X > Y + 7) = P(X - Y > 7)$, $(X - Y): \mu_{(X-Y)} = 5 - 3 = 2, \sigma_{X-Y}^2 = 9 + 16 = 25$
never subtract σ^2
 So $P(X - Y > 7) = 1 - P(X - Y < 7) = 1 - P(Z < \frac{7-2}{\sqrt{25}}) = 1 - P(Z < \frac{5}{5}) = 1 - P(Z < 1)$
check normal
 $= 1 - 0.8413 = 0.1587 \approx 0.16$

6. X, Y are independent random variables distributed according to normal distributions. The mean and variance for X are 5 and 9 respectively whereas the mean and variance for Y are 3 and 16 respectively. Calculate to 2 decimals

$P(X > Y + 7)$

- a) 0.03 b) 0.97 c) 0.42 d) 0.84 **e) 0.16**

$H_0: \mu = 3315$
 $H_1: \mu > 3315$

$\alpha = 0.01$

Use table: $t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

3119, 2657, 3459, 3629, 3345, 3629, 3515, 3856, 3629, 3345, 3062

$t = \frac{3386 - 3315}{336/\sqrt{11}} = 0.7$, $t_\alpha = 2.764$, $\bar{x} = 3386, s = 336$

$v = 11 - 1 = 10$

check $t > t_\alpha$
 $0.7 < 2.764$

$\Rightarrow \therefore$ do not reject

Soit X le poids en gms d'un petit garçon né à la maison à Ottawa et supposez que X suit une distribution normale avec moyenne et variance inconnues. Tester l'hypothèse nulle que la moyenne est de 3315 gms. contre l'alternative qu'elle est plus grande avec un niveau de 1%. Utilisez les données ci-dessus et indiquez la valeur de votre statistique ainsi que votre décision.

- a) 0.70, reject H_0 b) 4.03, do not reject H_0 **c) 0.70 do not reject H_0** d) 0.211 reject H_0 e) 4.03 reject H_0

8. Consider the situation described in question 7, find the p -value of the test statistic.

En faisant référence au problème 7, calculez le seuil de signification empirique.

- a) $0.15 < p < 0.20$ b) $.30 < p < .40$ c) $.40 < p < .50$ d) $.50 < p$ **e) my p-value is $0.2 < p < 0.3$**

\rightarrow go in t -table @ $v = 10$, find our test statistic: 0.7

4

See it's inbetween columns 0.3 & 0.2

So P -value: $0.2 < p < 0.3$

Estimating variance: $\left(\frac{(n-1)s^2}{\chi^2_{n-1}(\alpha/2)}, \frac{(n-1)s^2}{\chi^2_{n-1}(1-\alpha/2)} \right) = (2.30, 30.73)$

$\alpha = 0.1, \alpha/2 = 0.05$
 $n = 5, 1 - \alpha/2 = 0.95$

$\chi^2_{n-1}(\alpha/2) = 9.488, \chi^2_{n-1}(1-\alpha/2) = 0.711$

9. Calculate a 90% confidence interval for the variance of a normal distribution based on the following sample mean and sample variance in a sample of 5.

note: \bar{x} is $\rightarrow \bar{x} = 33.427, s^2 = 5.462$

- not needed a) (2.13, 32.86) b) (1.96, 45.10) **c) (2.30, 30.73)**
 d) (2.47, 23.84)

$H_0: \beta_1 = \beta_0$
 $H_1: \beta_1 \neq \beta_0$

10. An engineer wants to study the relationship between the time (y) it takes a car to go from 0 to 60 and horsepower (x). A random sample of 14 cars was taken. The following results below were obtained. In the linear regression model $Y = \beta_0 + \beta_1x + \epsilon$, test the null hypothesis that the slope is 0 against the alternative that it is different from 0 at the 5% significance level. Give the value of the test statistic and state your decision.

$n = 14, \alpha = 0.05$
 $\alpha/2 = 0.025$

Use Hypothesis testing on slope

$t = \frac{b_1 - \beta_0}{\frac{s}{\sqrt{S_{xx}}}}$ use t w/ $v = n - 2$

Un ingénieur veut étudier la relation entre le temps (y) qu'il faut une voiture pour aller de 0 à 60 (y) et la puissance (x). Un échantillon aléatoire de 14 voitures a été pris avec les résultats suivants. Dans le modèle de régression linéaire $Y = \beta_0 + \beta_1x + \epsilon$, tester l'hypothèse nulle que la pente est 0 contre l'alternative qu'elle est différente de 0 au seuil de 5%. Indiquez la valeur de votre statistique ainsi que est votre décision.

need b_1 & S

$b_1 = \frac{S_{xy}}{S_{xx}} = \frac{-812.193}{54169.2} = -0.015$ $\bar{x} = 252.6$
 $S^2 = \frac{S_{yy} - b_1 S_{xy}}{n-2} = \frac{18.0836 - (-0.015)(-812.193)}{14-2} = 0.472$ $\bar{y} = 6.821$
 $\sqrt{S^2} = S = 0.7$ $S_{xx} = 54169.2$
 $S_{xy} = -812.193$ $S_{yy} = 18.0836$

- $t = \frac{-0.015 - 0}{\frac{0.7}{\sqrt{54169.2}}} = -5$ a) -5.0, reject H_0 **b) 5.0, Reject H_0** c) 2.14, do not reject H_0
 d) -2.14, do not reject H_0

11. Suppose that in a linear regression problem involving 12 observations, we find that

$\hat{\beta} = 0.742, \sum (y_i - \hat{y}_i)^2 = 217.709, \sum (x_i - \bar{x})^2 = 756.1$

Determine a 95% confidence interval for the slope. (Use 3 decimals)

$\alpha = 0.05$
 $\alpha/2 = 0.025$ 5

Use $\hat{\beta} - (t_{\alpha/2}) \frac{s}{\sqrt{S_{xx}}} < \hat{\beta} < \hat{\beta} + (t_{\alpha/2}) \frac{s}{\sqrt{S_{xx}}}$
 find $s: S^2 = \frac{S_{yy}}{n-2} = \frac{217.709}{12-2} = 21.7709$
 $S = \sqrt{S^2} = \sqrt{21.7709} \approx 4.67$

$\left(0.742 - 2.228 \frac{4.67}{\sqrt{756.1}}, 0.742 + 2.228 \frac{4.67}{\sqrt{756.1}} \right)$
 $\Rightarrow 0.742 \pm 0.378$

Plug in formula $\rightarrow t_{\alpha/2}(12-2) =$ (see table)
 $\rightarrow t_{0.025}(10) = 2.228$

$t_{\alpha/2} = 2.179$
 \hat{t} at $v = 14 - 2$
 $+ < -t_{\alpha/2}$ or $+ > t_{\alpha/2} \Rightarrow$ reject H_0
 $-5 < -2.179 \Rightarrow$ reject H_0
 make \hat{t} ?

On considère une régression linéaire impliquant 12 observations avec les données ci-dessus. Calculez un intervalle de confiance à 95% pour la pente (Utilisez 3 décimales).

- a) 0.742 ± 1.310 b) 0.742 ± 0.120 c) 0.742 ± 0.337 d) 0.742 ± 0.972
 e) 0.742 ± 0.378

$H_0: \mu_A - \mu_B = 0$
 $H_1: \mu_A - \mu_B \neq 0$

12. To test whether a golf ball of brand A can be hit a greater distance than a golf ball of brand B, each of 10 golfers hit a ball of each brand. $n=10$
 Test at the 5% significance level whether or not there is a difference between the brands of golf balls. Provide the value of the test statistic and specify whether or not brand A is better than brand B.

$\bar{d} = \frac{13-4+3+14-1+17+11+13+17+14}{10}$
 $= \frac{97}{10} = 9.7$

use $S_d^2 = \frac{\sum d_i^2 - \frac{(\sum d_i)^2}{n}}{n-1}$
 $= \frac{1455 - \frac{(97)^2}{10}}{10-1} = 57.122$

\hookrightarrow plug in $t = \frac{\sqrt{n}(\bar{d} - \mu_0)}{s}$
 $t = \frac{\sqrt{10} \cdot (9.7 - 0)}{\sqrt{57.122}} = 4.06$
 $t_{\alpha/2}(n-1) = 2.262$
 $t > t_{\alpha/2} \Rightarrow$ reject H_0
 since d uses A-B, $d > 0 \Rightarrow A > B$

Distance for Ball A	265	272	246	260	274	263	255	258	276	274
Distance for Ball B	252	276	243	246	275	246	244	245	259	260
Difference	13	-4	3	14	-1	17	11	13	17	14

- a) $t = 0$ A and B are equally good
 B c) $t = 1.28$, B is better than A
 B d) $t = 4.06$, A is better than B
 e) $t = 4.06$, B is better than A

13. There are 20 multiple choice questions in a final exam. Each question has 5 possible answers, only one of which is the correct one. A student decides to guess the answer for each question by choosing one at random. What is the probability of him failing?

$n=20$
 $P_{\text{right}} = \frac{1}{5} = 0.2$
 fail if less than 10
 $P(X < 10) = P(X \leq 9)$
 $n=9$
 binom table
 $\Rightarrow 0.9974 = c)$

Il y a 20 questions à choix multiples dans un examen final. Il y a un choix de 5 réponses pour chaque question. Si un étudiant choisit ses réponses au hasard, quelle est la probabilité qu'il échoue?

- a) 0.0026 b) 0.0006 c) 0.9974 d) 0.5000 e) 0.9994

14. Suppose that we wish to compare the weight of calcium in standard cement and cement doped with lead. Reduced levels of calcium would imply that water would attack various locations in the cement structure. Assuming normal distributions, test the hypothesis that $\mu_X = \mu_Y$ against the alternative that they are different using a 5% significance level and assuming a common variance for the two distributions. Specify the value of your statistic to two decimals.

$H_0: \mu_X = \mu_Y$
 $H_1: \mu_X \neq \mu_Y$
 $\rightarrow H_0: \mu_X - \mu_Y = 0, \sigma_X = \sigma_Y = \sigma$

\hookrightarrow 4th row in table of tests for calculating means

(continued on next page)

T 2377 D

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - d_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

$$\hookrightarrow (90 - 87) - 0$$

$$= \frac{(10-1)25 + (15-1)16}{10+15-2} = \frac{449}{23} \approx 19.52 \Rightarrow S_p = \sqrt{19.52} = 4.42$$

$$\frac{4.42 \sqrt{\frac{1}{10} + \frac{1}{15}}}{1} = 1.66 = t$$

compare to $t_{\alpha/2}(23) = 2.069 \Rightarrow t < t_{\alpha/2}$ and $t > -t_{\alpha/2}$
 $\hookrightarrow 0.025$ \therefore do not reject H_0

Section	sample mean	sample variance	sample size
Standard cement (X)	90	25	10
lead doped cement (Y)	87	16	15

$$v = 10 + 15 - 2 = 23$$

- a) 2.95, reject H_0 b) 0.70, do not reject H_0 c) 1.64, do not reject H_0
 d) 1.64 reject H_0 e) 0.70, reject H_0

15. Let X_1, X_2, \dots, X_{16} be independent and identically distributed random variables from a normal distribution with mean 5 and variance 16. Find the constant c such that $P(\bar{X} - 5 > Sc) = 0.05$, where \bar{X} is the sample mean and S is the sample standard deviation.

- a) 0.4365 b) 1.753 c) 0.438 d) 6.984 e) 7.012

$X = \bar{X} - 5, Z = Sc$
 $P(X > Z) = 0.05$
 $\Rightarrow P(X < Z) = 0.95$
 find in Z table
 $\Rightarrow Z$ is 1.645

16. Suppose we are given a sample of 100 observations from a distribution whose density is

$$E(X) = \int x \cdot f(x)$$

CLT $\Rightarrow \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

$$\mu = E(X) = \int_0^2 x \left(\frac{x^3}{4}\right) dx = 1.6, \quad E(X^2) = \int_0^2 x^2 \left(\frac{x^3}{4}\right) dx = \frac{x^4}{4} \Big|_0^2 = 4$$

$$V(X) = E(X^2) - (E(X))^2 = 4 - (1.6)^2 = 0.11 = \sigma^2$$

$$\rightarrow P\left(\frac{1.58 - 1.6}{0.11} < Z < \frac{1.63 - 1.6}{0.11}\right)$$

Find approximately to two decimals the probability that the sample mean lies in the interval (1.58, 1.63).

- a) 0.49 b) 0.91 c) 0.43 d) 0.55
 e) 0.38

$\bar{X} - 5 > 1.645$
 $\bar{X} - 5 > Sc$
 $Sc = 1.645$
 $c = \frac{1.645}{4} = 0.411$

17. The random variable X has a Poisson distribution with variance equal to 4. Poisson $\Rightarrow \mu^2 = \sigma^2 = 4$

$$P(X < 2 \cap X < 5) = P(X < 2)$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Calculate $P(X < 2 | X < 5)$.

- a) 0.0916 b) $e^{-(5-2)}$ c) 0.6288 d) 0.1457 e) 0.7851

$$P(X < 2 | X < 5) = \frac{P(X < 2) - P(X=1)}{P(X < 5) - P(X=5)}$$

18. In a certain communication system there is on average one error in a transmission lasting 10 seconds. If errors occur in accordance with a Poisson process, what is the probability that there will be at least one error in a transmission lasting 30 seconds?

$X = \# \text{ success}$
 $\lambda = \text{avg occurrence}$
 $t = \text{time interval}$
 $\mu = \sigma^2 = \lambda t = \frac{1}{10} \cdot 30 = 3$

check Poisson table
 $\frac{0.0916}{0.6288} = 0.1457$

Dans un système de communication, il y a en moyenne une erreur dans une transmission qui dure 10 secondes. Si les erreurs suivent une loi de Poisson, quelle est la probabilité qu'il y ait au moins une erreur dans une transmission qui dure 30 secondes?

$$P(X=0) = \frac{e^{-\lambda} \lambda^x}{x!} = \frac{e^{-3} \cdot 3^0}{0!} = e^{-3}$$

$$1 - P(X=0) = 1 - e^{-3}$$

a) $1 - e^{-3}$

b) $1 - e^{-1}$

c) $1 - 4e^{-3}$

d) $1 - 3e^{-3}$

e) $1 - 2e^{-1}$

p_0

$H_0: p_u - p_r = 0$

$H_1: p_u - p_r < 0$

$\hat{p}_u = \frac{50}{100} = \frac{1}{2}$ $\hat{p}_r = \frac{65}{100} = \frac{13}{20}$

$\alpha = 0.1$

$\hat{p} = \frac{50 + 65}{200} = \frac{115}{200} = \frac{23}{40}$

$z = \hat{p}_u - \hat{p}_r = \frac{1}{2} - \frac{13}{20} = \frac{10}{20} - \frac{13}{20} = -\frac{3}{20}$

$z = \frac{\hat{p}_u - \hat{p}_r}{\sqrt{\frac{\hat{p}\hat{q}}{n_1} + \frac{\hat{p}\hat{q}}{n_2}}} = \frac{-\frac{3}{20}}{\sqrt{\frac{\frac{23}{40} \cdot \frac{17}{40}}{100} \cdot 2}}$

$= -2.15, z_{\alpha} = 0.95$

$z < -z_{\alpha} \Rightarrow \text{reject } H_0$
make $z \oplus$!

19. A random sample of 100 urban residents reveals that 50 believe in angels whereas in a random sample of 100 rural residents it is found that 65 believe in angels. Test the null hypothesis that the percentage of urban and rural residents who believe in angels is the same against the alternative that it is higher for rural residents. Using a 10% significance level, specify the value of the test statistic and indicate the decision.

Un échantillon aléatoire de 100 résidents urbains révèle que 50 croient aux anges alors que dans un échantillon aléatoire de 100 résidents des régions rurales, on constate que 65 croient aux anges. Testez l'hypothèse nulle que le pourcentage des résidents urbains et ruraux qui croient aux anges est le même contre l'alternative qu'il est plus élevé pour ce dernier. Utiliser un niveau de signification de 10%, spécifier la valeur de la statistique de test et d'indiquer la décision.

a) 1.53, do not reject H_0

b) 4.70, reject H_0

c) 3.07, reject H_0

d) 0.47 do not reject H_0

e) 2.95, do not reject H_0

f) 2.17, reject H_0

20. In order to establish a control chart for the mean of a process, 25 = k samples each of size 4 = n are collected. We note that $\sum_{j=1}^{25} \bar{X}_j = 4000$ and $\sum_{j=1}^{25} \sigma_j = 50$. What is the value of the upper control limit of the chart for the mean.

Afin d'établir une carte de contrôle pour la moyenne d'un processus, 25 échantillons chacun de taille 4 sont collectées. Nous notons que $\sum_{j=1}^{25} \bar{X}_j = 4000$ et $\sum_{j=1}^{25} \sigma_j = 50$. Quelle est la valeur de la limite supérieure de contrôle pour la moyenne

a) 167.5

b) 161.3

c) 160.0

d) 254.0

e) 163.8

$UCL = \bar{\bar{x}} + 3 \frac{\bar{\sigma}}{c\sqrt{n}}$

$c = 0.7979$ (use table for c)

$= 160 + 3 \frac{2}{0.7979\sqrt{4}} = 163.76 \approx 163.8$