UNIVERSITY OF REGINA
Department of Mathematics and Statistics

Elementary Statistics For Applications
Stat 100

Final Exam, Fall 2013

Time: 3 hours

Pages: 10

Full Name:__________________________________

Student Number:____________________________

Instructor: (check one)

- ☐ Peter Douglas (100-001)
- ☐ Robert Petry (100-C01)
- ☐ Vijayaparvathy Agasthian (100-L01/L02)
- ☐ Patrick Maidorn (100-991)

INSTRUCTIONS

1. **All** work and answers are to be placed in this booklet. If you require more space for an answer, work on the **back** of the sheet and **indicate** where the work is to be found.
2. Place final answers in **blanks at right** when provided.
3. Each question will be graded for both the **correct answer** and **supporting work**.
4. The marks are placed **beside** each question.
5. Only **calculator**, **formula sheet**, and the provided **probability tables** and **scrap paper** are permitted in the exam.

For instructor use only:

Page	Marks	Score
2	15	
3	10	
4	15	
5	10	
6	11	
7	11	
8	10	
9	10	
10	8	
Total:	100	

Marks

(15) 1. *Joe's Nut Bar* sells snacks to office workers. A random selection of Nut Bar products had the following prices:

7.25, 7.50, 7.50, 7.50, 8.25, 11.00, 12.50, 13.00, 15.25 ($)

(a) Is the variable discrete or continuous? (Circle your answer at right.) **discrete/continuous**

Solution: Discrete as prices do not come in fractions of a cent.

(b) Determine the mode. **Answer:** ______

Solution:

$$\text{mode} = \$7.50$$

(c) Determine the mean. **Answer:** ______

Solution:

$$\bar{x} = \frac{\$89.75}{9} = \$9.9722222 \approx \$9.97$$

(d) Determine the standard deviation. **Answer:** ______

Solution:

$$s = \sqrt{\frac{968.1875 - \frac{(89.75)^2}{9}}{9-1}} = \sqrt{\frac{73.18055556}{8}} = \sqrt{9.147569445} = 3.024494907 \approx \$3.02$$

(e) Determine the first and third quartiles. **Answer:** Q_1=______

Q_3=______

Solution:

$$Q_1 = x_{\frac{1}{4}(9+1)} = x_{2.5} = \$7.50$$

since $x_2 = x_3 = \$7.50$

$$Q_3 = x_{\frac{3}{4}(9+1)} = x_{7.5} = \frac{x_7 + x_8}{2} = \frac{\$12.50 + \$13.00}{2} = \$12.75$$

(f) Is $15.25 an outlier in the above dataset? Circle your answer at right and show the relevant calculation used to determine your result below. **yes/no**

Solution: The interquartile range is $IQR = Q_3 - Q_1 = \$12.75 - \$7.50 = \$5.25$ The lower fence and upper fence are

$$\text{Lower Fence} = Q_1 - 1.5(IQR) = \$7.50 - 1.5(\$5.25) = -\$0.375$$

$$\text{Upper Fence} = Q_3 + 1.5(IQR) = \$12.75 + 1.5(\$5.25) = \$20.625$$

Since $15.25 lies within this range it is not an outlier.

(10) 2. A statistics instructor tracks the calls and texts that he receives on his cell phone over a month. His results are as follows.

	Contact Source				
Type of Message	Family Member	Opinion Survey	Scam Artist	Wrong Number	Total
Phone Call	7	10	13	6	36
Text	8	4	8	4	24
Total	15	14	21	10	60

Using this data, for a randomly selected message . . .

(a) What is the probability that it came from a family member? **Answer:** ____________

Solution:

$$P(F) = \frac{15}{60} = \frac{1}{4} = 0.25$$

(b) What is the probability it is a phone call that is a wrong number? **Answer:** ____________

Solution:

$$P(P \text{ and } W) = \frac{6}{60} = \frac{1}{10} = 0.10$$

(c) What is the probability that it is a text message or a scam artist? **Answer:** ____________

Solution:

$$P(T \text{ or } S) = P(T) + P(S) - P(T \text{ and } S) = \frac{24}{60} + \frac{21}{60} - \frac{8}{60} = \frac{37}{60} = 0.6167$$

(d) What is the probability that it is a family member contacting him, given that he received a text message? **Answer:** ____________

Solution:

$$P(F|T) = \frac{8}{24} = \frac{1}{3} = 0.3333$$

(e) Are the events "receives a text message" and "is contacted by a family member" independent? Why? (Circle your answer at right and show your work below.) **yes / no**

Solution: $\boxed{No}$ since (any of the following):

- $P(T \text{ and } F) = \frac{8}{60} = \frac{2}{15} = 0.1333 \neq P(T) \cdot P(F) = \frac{24}{60} \cdot \frac{15}{60} = \frac{1}{10} = 0.10$
- $P(T|F) = \frac{8}{15} = 0.5333 \neq P(T) = \frac{24}{60} = 0.4$
- $P(F|T) = \frac{8}{24} = 0.3333 \neq P(F) = \frac{15}{60} = 0.25$

(10) 3. Katniss Everdeen is an expert archer who can hit the bulls-eye on a target in 90% of her shots.

(a) In a particular training session Katniss takes 8 shots at the target. What is the probability that she hits the bulls-eye at least 6 times? **Answer:** ____________

Solution:

$$\begin{aligned} P(x=6) &= {}_8C_6(0.9)^6(0.1)^2 = 0.14880348 \approx 0.1488 \\ P(x=7) &= {}_8C_7(0.9)^7(0.1)^1 = 0.38263752 \approx 0.3826 \\ P(x=8) &= {}_8C_8(0.9)^8(0.1)^0 = 0.43046721 \approx 0.4305 \end{aligned}$$

So $P(6 \le x) = P(x=6) + P(x=7) + P(x=8) = 0.96190821 \approx 0.9619$

(b) Over a two-week period Katniss takes 240 shots at the target. Use the normal approximation to the binomial to estimate the probability that she hits the target at least 220 times. **Answer:** ____________

Solution: The mean for the binomial is $\mu = (240)(0.9) = 216$ with standard deviation $\sigma = \sqrt{(240)(0.9)(0.1)} = \sqrt{21.6} = 4.647580015$. The probability is

$$\begin{aligned} P(220 \le x_{\text{binomial}}) &= P(219.5 < x_{\text{normal}}) = P\left(\frac{219.5-216}{4.64\ldots} < z\right) \\ &= P(0.753080095 < z) = P(z < -0.75) = 0.2266 \end{aligned}$$

Note if no continuity correction is used the (incorrect) answer is 0.1949 .

---

(5) 4. A Toronto newspaper wishes to conduct a poll to estimate the current approval rating of the mayor. The question asked will be "Do you approve of the actions of the mayor?" with the only choice of answers being *Yes* or *No*. How large a sample should be selected if the newspaper wants its poll results to be accurate within 4% of the true population proportion, 19 out of 20 times? **Answer:** ____________

Solution: $E = 0.04$, $z = 1.960$ for $1-\alpha = 19/20 = 0.95$, and proportion estimate (since none given) is 0.50. Sample size is:

$$n = (0.50)(0.50)\left[\frac{1.960}{0.04}\right]^2 = 600.25 = 601 \text{ residents}$$

(10) 5. The weight of male reindeer of the *R. Santaclausus* subspecies is normally distributed with mean 102.4 kg and standard deviation 13.9 kg.

(a) What proportion of these reindeer would weigh more than 118.0 kg? **Answer:** ____________

Solution:

$$\begin{aligned} P(118\text{ kg} < x) &= P\left(\frac{118.0\text{ kg} - 102.4\text{ kg}}{13.9\text{ kg}} < z\right) \\ &= P(1.122302158 < z) \\ &= 1 - P(z < 1.12) \\ &= 1 - 0.8686 \\ &= 0.1314 \end{aligned}$$

(b) Rudolph is a fairly small reindeer. If exactly 10% of the reindeer weigh less than Rudolph, how much does Rudolph weigh? **Answer:** ____________

Solution: 10% of the probability in a normal distribution is less than $z = -1.28$. Hence

$$x = \mu + z\sigma = 102.4 + (-1.28)(13.9) = 84.608\text{ kg} \approx 84.6\text{ kg}$$

(c) If 36 reindeer are randomly selected and their average weight calculated, what is the probability that the mean weight is less than 100.0 kg? **Answer:** ____________

Solution: Standard error of the mean is $\sigma_{\bar{x}} = \sigma/\sqrt{n} = 13.9/\sqrt{36} = 2.31666666$ kg and so

$$\begin{aligned} P(\bar{x} < 100\text{ kg}) &= P\left(z < \frac{100.0\text{ kg} - 102.4\text{ kg}}{2.316\ldots\text{ kg}}\right) \\ &= P(z < -1.035971223) \\ &= P(z < -1.04) \\ &= 0.1492 \end{aligned}$$

(11) 6. Fish oil is a popular supplement taken to support cardiac and cognitive health. A pharmaceutical company is investigating a new process to extract raw fish oil from salmon. In a limited test run, the process was used on 18 one-kilogram batches of salmon. On average, the process was able to extract 50.7 grams of fish oil per batch, with sample standard deviation 5.9 grams.

Use this information to construct a 90% confidence interval for the true mean amount of fish oil that this process could extract per one-kilogram batch of salmon. As part of your answer state any necessary assumptions that have to be made.

Solution:

- Standard error = $\frac{5.9 \text{ g}}{\sqrt{18}} = 1.390643336$ g
- $t = 1.740$ for $df = 17$ and $\alpha = 0.90$
- Maximum error $E = 2.419719405 \text{ g} \approx 2.4\ g$

The 90% confidence interval is $50.7 \text{ g} - 2.4 \text{ g} < \mu < 50.7 \text{ g} + 2.4 \text{ g}$ or equivalently $48.3 \text{ g} < \mu < 53.1 \text{ g}$

Assumption: the amount of oil extracted per batch is approximately normally distributed.

(11) 7. Statistics Canada reports that half of all households have at least one pet, with pet ownership higher in Alberta, Manitoba, and Saskatchewan and lower in the Maritimes, British Columbia, Ontario, and Quebec.

(a) In a random sample of 300 households across Canada, 126 households said that they owned at least one pet. Does this data provide sufficient evidence to indicate that the proportion of households with at least one pet is different from that reported by the Statistics Canada? Test using a level of significance of $\alpha = 0.05$.

Solution: Critical z-values for the two-tailed test ("different") are ± 1.960 . Sample proportion $p = 126/300 = 0.42$. Standard error of proportion is

$$\sqrt{\frac{(0.5)(0.5)}{300}} = 0.028867513 \ .$$

Calculated z-value for the data is therefore

$$z = \frac{0.42 - 0.50}{0.0288\ldots} = -2.771281292 = -2.771$$

Therefore we reject H_0 at $\alpha = 0.05$ and accept H_a. The evidence supports that the proportion of households with at least one pet differs from the 50% reported.

(b) What is the P-value for the hypothesis test?

Solution: The area from the normal curve table for $z = -2.77$ is 0.0028 . Doubling for the two tails gives

$$P\text{-value} = 2(0.0028) = 0.0056$$

which is less than $\alpha = 0.05$ as expected for a test in which the null hypothesis was rejected.

Answer: ____________

(10) 8. Boston Bobby is a die-hard baseball fan and supports his hometown team the *Boston Red Sox*. Bobby has an ongoing argument with New York Nat over the defensive capabilities of the *Boston Red Sox* and the *New York Yankees* team. Bobby claims that the Red Sox have a superior defense and uses the following statistics to back up his claim.

		Runs Allowed per Game	
Team	Games Played	Mean (runs)	Standard Dev. (runs)
Boston Red Sox	35	3.96	1.07
New York Yankees	32	4.25	1.15

Bobby claims that Boston will allow fewer runs on average than New York. Test this hypothesis using a 5% level of significance.

Solution: Assuming a left-tailed test ($\mu_B - \mu_N < 0$) the critical z-value is -1.645. The standard error of the difference of means is $\sqrt{\frac{(1.07)^2}{35} + \frac{(1.15)^2}{32}} = 0.272102101$ runs. So the calculated z-value for the data is

$$z = \frac{3.96 - 4.25}{0.2721\ldots} = -1.065776407 \approx -1.066$$

Therefore we fail to reject H_0 at α=0.05 . The evidence does not support Bobby's claim. Alternatively the P-value for $z = -1.07$ is 0.1423 which is greater than $\alpha = 0.05$ resulting in the same conclusion. Note that a z-test was valid here due to the large sample sizes ($n \geq 30$).

(10) 9. A mathematics drop-in help centre keeps track of the number of students it helps on a given day with the following results:

Monday	Tuesday	Wednesday	Thursday	Friday
4	3	6	1	4
7	6	5	3	2
8	7	8	3	1
4	3	0	5	1
7	4	3	3	1
	6	0	2	3

Assuming that the underlying populations are normal with common standard deviation, can one conclude that the average number of students helped in the centre on a given day differs from any other at a level of significance of $\alpha = 0.01$?

Use the following *partially completed* ANOVA table in your analysis:

Source	*df*	Sum of Squares	Mean Squares	F
Treatments		55.8		
Error		99.0		
Total		154.8		

Solution: The exact ANOVA table is

Source	*df*	Sum of Squares	Mean Squares	F
Treatments	4	55.7586...	13.9396...	3.379...
Error	24	99.0000	4.1250	
Total	28	154.7586...		

For $df_1 = 4$ and $df_2 = 24$ at $\alpha = 0.01$ we have F_{critical}=4.218 . Since the calculated F=3.379 (3.382 for the rounded values given in the table) is less than this we fail to reject H_0 at $\alpha = 0.01$; the evidence does not support that the mean number on any particular day differs from any other. (Note that for the data the actual P-value=0.025 .)

(8) 10. Data from a study on the number of absences and the final grades of seven randomly selected students from a statistics class is given in the following table.

Number of absences (x)	Final Grade (y)
6	82
2	86
15	43
9	74
12	58
5	90
8	78

The following summations for the data have already been calculated:

$\Sigma x = 57,\quad \Sigma y = 511,\quad \Sigma x^2 = 579,\quad \Sigma y^2 = 38993,\quad \Sigma xy = 3745$

(a) The correlation coefficient for the relationship between number of absences and final grade is calculated to be $\boxed{r = -0.9442}$. Interpret the meaning of this value in plain English.

Solution: The negative sign indicates a negative correlation between the variables. (As absences x increase the grade y decreases.) The magnitude of r indicates that the correlation is extremely high. Therefore, in summary, there is an extremely high negative correlation.

(b) Find the equation of the best fit (least-squares) regression line.

Solution: Using the following (or other) standard formulas gives:

$$\begin{aligned}
a &= \frac{(\sum x^2)\cdot(\sum y) - (\sum x)\cdot(\sum xy)}{n\cdot(\sum x^2) - (\sum x)^2} \\
&= \frac{(579)\cdot(511) - (57)\cdot(3745)}{(7)\cdot(579) - (57)^2} = \frac{82404}{804} = 102.4925373 \approx 102.5\% \\
b &= \frac{n\cdot(\sum xy) - (\sum x)\cdot(\sum y)}{n\cdot(\sum x^2) - (\sum x)^2} \\
&= \frac{(7)\cdot(3745) - (57)\cdot(511)}{(7)\cdot(579) - (57)^2} = \frac{-2912}{804} = -3.621890547 \approx -3.62\ \%/\text{absence}
\end{aligned}$$

The linear regression line $\hat{y} = a + bx$ is therefore

$$\hat{y} = 102.5 - 3.62x \quad [\%]$$

Answer: ____________________