MATH1324 Applied Analytics

Online Test

Online Test date: 18 June 2020 **Starting Time:** 1400 hrs

MATH1324 Online Test Information

- The online test will be **OPEN-BOOK** and **No limitation for the number of pages.**
- Questions are more conceptual. No use/memorisation of R code needed. Simple calculations involved and show all your workings.
- The Test will cover Modules 1-9.
- Duration of test will be 2 hours 15 minutes but upload your file by **5pm, the same day** and marks will be deducted for late submissions.
- Convert you answers into **pdf** file and upload at the end of the test.
- Answer **ALL** questions (6 short questions worth 5 marks each) and use your statistics tables whenever there is need.
- Number your answers clearly.
- Weight: 15%
- You are free to type or use your handwriting, as long as it's legible and converted to a **pdf file (NOT images OR pictures).**

Please, the following declaration should be on the **first page** of your answer sheet/file:

I, Rashbir Singh Kohli , declare that the work I am submitting for this Online Test is entirely my own work. Nothing I submit is copied from another source – either from another student or person, or from another resource. All written work is in my own words.

I have not communicated with any other student during the period of this Assessment Task. I have not discussed the content of this Assessment with anyone else during the period of this Assessment Task.

Question 1.

Lily frequents one of two fast food restaurants, choosing McDonald 25% of the time and Burger King 75% of the time. Regardless of where she goes, she buys French Fries on 60% of her visits.

a) The next time Lily goes into a fast food restaurant, what is the probability that she goes to McDonald and orders a French Fries?

Answer:

$P(M) = 0.25$

$P(B) = 0.75$

$P(F/M) = P(F/B) = 0.60$

$P(M \cap F) = P(M) \,.\, P(F/M) = (0.25)(0.60) = 0.15$

b) Are the two events in the previous question independent? Explain.

Answer:

$P(F) = 0.60$ regardless of whether Lily visits McDonald or Burger King, the two events are independent

c) If Lily goes to a fast food restaurant and orders French Fries, what is the probability that she is at Burger King?

Answer:

$P(B/F) = P(B \cap F)/P(F) = P(B) \,.\, P(F/B)/P(F) = P(B) = 0.75.$

d) What is the probability that Lily goes to McDonald, or orders French Fries, or both?

Answer:

(0.25 + 0.6) - 0.15 = 0.70 (70%)

[5 marks]

Question 2.

A pathologist working in the haematology department of the local hospital determines that the mean weight of a pathological tumour within the intestinal track can be normally distributed with a mean of 2.5 kg and a standard deviation of 0.8 kg.

a) What is the probability that a randomly selected tumour will have a mean weight in excess of 3.6 kg?

Answer:

$P(X > 3.6) = 1 - P(X < 3.6)$

$= 1 - P(Z <)$

$= 1 - P(Z < 1.38)$

$= 0.08456572$ (from Z tables)

b) What is the probability that a randomly selected tumour will have a mean weight less than 4.3 kg?

Answer:

$P(X < 4.3) = P(Z <)$

$= 0.9877755$ (from Z tables)

c) What is the probability that a randomly selected tumour will have a mean weight between 3.6 kg and 4.5 kg?

Answer:

P(3.6 < X < 4.5)

= P(X < 4.5) – P(X < 3.6)

= 0.9937903 – 0.9154343

= 0.07835606

d) What weight would put a randomly selected tumour in the top 2.5% (in terms of weight) of all tumours the pathologist had to examine?

Answer:

$P(Z <) = 0.025$

Z for0.025 is +1.96.

$X = (+1.96 \times 0.8) + 2.5$

$X = 4.07$ kg

[5 marks]

Question 3.

a) A study was conducted to see how long medical center visitors had to wait before their appointment. A random sample of 36 patients showed the average waiting time was 20 minutes with a standard deviation of 15 minutes. Construct a 99% confidence interval for μ, the true mean waiting time.

Answer:

CI = 99% = 0.01 = $\alpha(Significance - Level)$

$$1 - \frac{\alpha}{2} = 0.995$$

$$Z_{0.995} = 2.575829$$

$$20 - \frac{2.575829 * 15}{\sqrt{36}} = 13.56043$$

b) A university administration is interested in knowing what proportion of applicants will be accepted to be enrolled into the statistics program. Of a random sample of 85 applicants, 22 were accepted to be enrolled in the statistics program.

i. Construct a 95% confidence interval for p, the true proportion of all applicants that will be accepted in this program.

Answer:

$$p \pm z_{1-(\alpha/2)}\sqrt{\frac{p(1-p)}{n}}$$

P = 0.25

$z_{1-(\alpha/2)} = 1.96$

N = 85

= Upper bound = 0.1698505

Lower Bound = 0.3652356

True proportion = 0.2588235

ii. If the administration claimed that the true proportion of the applicants accepted into the statistical program is 17%, would you agree or disagree? Explain.

Answer:

We would agree as 17% = 0.17 and it lies in our confidence interval of true proportion

iii. The university estimate the rate of enrolment to be 12 students per month. Calculate the 95% confidence interval for the annual enrolment rate.

Answer:

Upper bound = 6.200603

Lower Bound = 20.96156

[5 marks]

Question 4.

The public relations officer for a particular city claims the average monthly cost for childcare outside the home for a single child is \$700. A potential resident is interested in whether the claim is correct. She obtains a random sample of 64 records and computes the average monthly cost of this type of childcare to be \$689 with a standard deviation of \$40.

a) Perform the appropriate test of hypothesis for the potential resident, using α = 0.01, using the critical value approach.

   i. Formulate the appropriate null and alternative hypotheses

Answer:

Null Hypothesis(H_0) = Average monthly cost for childcare outside the home for a single child is \$700 ($\mu = 700$)
Alternative Hypothesis(H_A) = average monthly cost for childcare outside the home for a single child is not \$700 ($\mu \neq 700$)

   ii. Calculate the relevant test statistics

Answer:

$\bar{x} = 689$
$\mu = 700$
$\sigma = 40$
$n = 64$

$$\frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

= -2.2

   iii. Formulate the decision and conclusion (can you or you cannot reject the null hypotheses)

b) What effect, if any, would there be on the conclusion of the test of hypothesis in the first question if you changed α to 0.05?

Answer:

To calculate p-value $= Pr(t < -2.2 | t = 0) + p$-value $= Pr(t > 2.2 | t = 0)$

p = 0.03148254 < 0.05

so, we will reject our null hypothesis that average monthly cost for childcare outside the home for a single child is \$700.

[5 marks]

Question 5.

A soft drink distributor was interested in examining the relationship between the number of ads (x) for his product during prime time on a local television station and the number of sales per week (y) in 1000's of cases. He compiled the figures for 20 weeks and computed the following summary information:
$n = 20$, $\sum x_i = 92$, $\sum y_i = 177$, $\sum x_i\, y_i = 884$, $\sum x^2{}_i = 425$.

a) Find the slope and the y-intercept for the number of ads during prime time and weekly sales.

Answer:

$$L_{xy} = \sum_{i=1}^{n} x_i y_i - \frac{(\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{n}$$

$$L_{xx} = \sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}$$

$$b = \frac{L_{xy}}{L_{xx}}$$

= 16.55(slope)

$a = \bar{y} - b\bar{x}$
= -67.28(y-intercept)

b) Find the best-fitting line relating the number of ads during prime time and weekly sales.

Answer:
$y = 16.55x - 67.28$

c) If the soft drink distributor ran 21 TV ads per week for his product, what would you predict his sales to be?

Answer: 280.27 (280,270 cases)

[5 marks]

Question 6.

a) Give an example of a mutually exclusive event.

Answer:

(1) Turning left and turning right are Mutually Exclusive.

(2) Tossing a coin: Heads and Tails are Mutually Exclusive

(3) Cards: Kings and Aces are Mutually Exclusive

b) How do you minimise or reduce standard error?

Answer: By increasing sample size

c) In sampling distribution, when do you use the Student t-distribution?

Answer:

(1) When the sample size is small but greater than 30(for CLT to hold)

d) Explain how least of squares are used to determine the best linear regression model.

Answer:

The idea behind this method is to minimise the **sum of squared distances**,

S, for each (x_i, y_i) bivariate data point from a fitted regression line. The sum of squares is written as:

$$S = \sum_{i=1}^{n} d_i^2$$

e) Which test is appropriate to test association?

Answer:

The Chi-square test is appropriate to test association.

[5 marks]

- End of Test -