DUNDALK INSTITUTE OF TECHNOLOGY
INSTITIÚID TEICNEOLAÍOCHTA DHÚN DEALGAN

Continuous Assessment Cover Sheet

Student Name:		**Student Number:**
Programme:	**Stage:**	**Complete Student Checklist:** Re-read brief ☐ References and Bibliography ☐ Proofread ☐
Module:		
Due Date:	**No. Pages:**	
Lecturer(s) Name:		
Assignment No. and/or Description/Topic:		**Mode of Submission:** Softcopy ☐ Hardcopy ☐

DECLARATION: **I declare that:**

- This work is entirely my own, and no part of it has been copied from any other person's words or ideas, except as specifically acknowledged through the use of inverted commas and in-text references;
- No part of this assignment has been written for me by any other person except where such collaboration has been authorised by the lecturer(s) concerned;
- I have not used generative artificial intelligence (AI) (e.g. ChatGPT) unless it has been permitted by the lecturer(s) concerned;
- I understand that I am bound by DkIT Academic Integrity Policy. I understand that I may be penalised if I have violated the policy in any way;
- This assignment has not been submitted for any other module at DkIT or any other institution, unless authorised by the relevant Lecturer(s);
- I have read and abided by all of the requirements set down for this assignment.

SIGNATURE……..……
DATE……..……

Lecturer's Comments:

Provisional Mark : _____ **Lecturers Signature :** _____ **Date:** _____

Work submitted late will be subject to penalties in accordance with the DkIT Continuous Assessment Policy

Assignment 1 (10%):
Due at 23:00 on the 29th of October

For Section A please make sure you have joined the Khan academy classroom so that I can see your marks to assign them to you. Section B: Mostly answer using pen and paper, Scan/Photograph. Submit to moodle by zipping the folder. Two parts in Q3 are verify answers in python, please submit appropriate script with solutions (example .py/.ipynb file or html file). You must submit the cover sheet as part of the submission as well. Please see Academic Integrity Policy in relation to cover sheet below.

Academic Integrity:

a. PLEASE PAY SPECIAL ATTENTION TO THE ISSUE OF ACADEMIC INTEGRITY. The DkIT policies are available at https://www.dkit.ie/registrars-office/academic-policies/academic-integrity-policy-procedures
b. In summary, all work submitted by learners for assessment purposes, or for written or oral publication, must be their own work. Where this is informed by the work of others, the source must be properly referenced using the accepted norms and formats of the appropriate academic discipline.
c. Generative artificial intelligence (AI) tools are completed restricted for this assessment task. https://www.dkit.ie/about-dkit/policies-andguidelines/academic-policies.html).
d. Using generative artificial intelligence tools (e.g. ChatGPT) in this assignment unless explicitly permitted to do so and without proper acknowledgement of the use, is a form of plagiarism.
e. Student are not permitted to share any solution with any other individual in the class. Students may not distribute their solution to any student in any format (i.e. electronic, verbal, or hardcopy transmission).
f. Any plagiarism will be reported to the Head of Department and a report will be added to your permanent academic record.

Section A: Khan Academy [30 marks]

Khan Academy, go to Khan Academy account after joining the classroom link and should see that 3 unit tests have been assigned. Complete these 3 unit tests using Khan. (40-60 minutes). If you can not see the assignments please use the below links. Email me if there any questions.
https://www.khanacademy.org/math/statistics-probability/probability-library/conditional-probability-independence/test/probability-library-unit-test?referrer=upsell
https://www.khanacademy.org/math/statistics-probability/counting-permutations-and-combinations

https://www.khanacademy.org/math/statistics-probability/modeling-distributions-of-data/more-on-normal-distributions/test/modeling-distributions-of-data-unit-test?referrer=upsell

Section B: Probability Distributions and Hypothesis Testing

Question 1: **[20 marks]**

You are conducting a study to investigate the relationship between the use of a certain medication (exposure) and the occurrence of a specific side effect (outcome). The study duration is 5 years. You have the following data:
In a group of 1000 individuals who have taken the medication, 200 individuals developed the side effect.
In a group of 1000 individuals who did not take the medication, 50 individuals developed the side effect.

(a) Determine whether this study is prospective or retrospective and provide a brief explanation of your choice.

- **Prospective**: Looking forward
- **retrospective**: looking backwords

This study is a Prospective study because the study follows the patients from the very start till the end it is using new data and they are not looking at data thats been collected already by someone else.

(b) Create a table to show the data for the exposed and non-exposed groups.

	Side affects	No side affects	Total
Medication	200	800	1000
Placido	50	950	1000
Total	250	1750	2000

(c) Calculate the **odds ratio** of developing the **side effect** for those who were **exposed** to the medication (**verses those not exposed**). Show your calculations.

Medicated group
Odds ration = (Number of side affects with midication) / (Number of none side affects with medication)
Odds ration = (200) / (800) = 0.25

Placibo group
Odds ration = (Number of no affected) / (Number of none side affects)
Odds ration = (50) / (950) = 0.052631579

Medicated group = 0.25
Placibo group = 0.052631579

OR = (odds of medicated group) / (Odds of placibo group)
OR = 0.25 / 0.052631579 = 4.749999995

(d) Interpret the odds ratio you calculated in part (c) and compare it with the relative risk (if it is appropriate to calculate). Discuss any implications or conclusions that can be drawn from these measures of association.

RR = (A / (A+B)) / (C / (C+D))

RR = (200/(200 + 800)) / (50/(50+950))

RR = 4

The result is normal the relative risk is **4** and the odds ration is **4.749999995.**
The odds ration of a side affect is **4.75 times higher** then the medicated group.

Question 2: **[20 marks]**

Data from a case-control study on the relationship between male-pattern baldness and cardiovascular disease is presented. Cases were men less than 55 years of age who were hospitalized for an acute MI (heart attack). Controls were men in the same age range admitted to the same hospital for non-cardiac conditions. Baldness was an ordinal variable graded 1 for no baldness, 2 for moderate baldness, and 3 for extreme baldness. Data were:

Baldness	Cases	Controls
1	258	342
2	188	180
3	54	39

(a) Determine whether this study is prospective or retrospective and provide a brief explanation of your choice.

- **Prospective**: Looking forward
- **retrospective**: looking backwords

This study is a **retrospective** study because it is using data thats been collected in the past.

(b) Using the no baldness group as reference group, calculate the odds ratio associated with moderate baldness. Show your work.
OR = (Number no baldness group) / (Number of modern baldness group)
OR = (188 / 258) / (180 / 342) = 0.728682171 ÷ 0.526315789 = **1.384496126**
OR = 1.384496126

(c) Using the no baldness group as reference group, calculate the odds ratio associated with extreme baldness. Show your work

OR = (Number no baldness group) / (Number of extreme baldness group)
OR = (54 / 258) / (39 / 342) = 0.209302326÷0.114035088 = **1.835420393**
OR = **1.835420393**

(d) Interpret and compare both odds ratios calculated in (b) & (c).

OR group B = (No baldness) / (Moderate baldness) = **1.384496126**
OR group C = (Number no baldness group) / (Number of extreme baldness group) = **1.835420393**

Once both of the odds ratio was calculated we can see that men with extreme baldness have a higher odds of a heart attack.

OR group B = 1.384496126 < OR group C = 1.835420393

(e) Comment on the study design used to investigate the odds ratio associated with baldness and any observations you can gleam from it.

The study is a retrospective one and what has been observered is that men with sever boldness has a higher precentage of getting a heart attack.

Question 3: **[30 marks]**

(a) **[10 marks]**

A retail store accepts product returns, and they found that 10% of the items sold in a month are returned due to defects. If 25 items are randomly selected from a month's sales records, calculate the following:

(i) The expected number of items returned due to defects.

Probability of return = 0.10
randomly selected items = 25

Binomial distribution = (25 * 0.1) = 2.5

(ii) The probability that exactly three items are returned.

n = 25 , k = 3, p = 0.10 ,(1-p) = 0.90
$P(x = k) = (n/k)p^k (1-p)^{n-k}$

$P(X = 3) = (25/3)(0.10)^3(0.90)^{25 - 3}$

$P(X = 3) = (25/3)(0.10)^3(0.90)^{22}$

$(25/3) = 25/3 * 24/2 * 23/1 = 2300$

$(0.10)^3 = 0.001$

$(0.90)^{22} = 0.09847709$

$P(x = 3) = 2300 * 0.001 * 0.09847709$
$P(x = 3) = 2.3 * 0.09847709 = 0.226497307$

$p(x = 3) = 0.226497307$

(iii) The probability that four or more items are returned.

$n = 25$, $k = 4$, $p = 0.10$,$(1-p) = 0.90$

$P(x = k) = (n/k)p^k (1-p)^{n-k}$

$P(x \geq 4) = 1 – [P(X=0) + P(X = 1) + P(X = 2) + P(X = 3)]$

Already have $P(X = 3) = 0.226497307$

$P(x \geq 4) = 1 – [P(X=0) + P(X = 1) + P(X = 2) + 0.226497307]$

$P(X = 0) = (25/0)(0.10)^0(0.90)^{25 – 0} =$ **0.07178979877**
$P(X = 1) = (25/1)(0.10)^1(0.90)^{25 – 1} =$ **0.1994161077**
$P(X = 2) = (25/2)(0.10)^2(0.90)^{25 – 2} =$ **0.2658881436**
$P(X = 3) =$ **0.226497307** taken from part (ii)

$P(x \geq 4) = 1 – [P(X=0) + P(X = 1) + P(X = 2) + P(X = 3)]$
$P(x \geq 4) = 1−(0.07178979877 + 0.1994161077 + 0.2658881436 + 0.226497307)$
$P(x \geq 4) = 1 − 0.763591357$

$P(x \geq 4) =$ **0.23640864**

(iv) Using python, confirm your answers in questions (i)-(iii).

Part (i) the answer was the same.
Part (iii) the answer was the same.

```python
import math

# A retail store accepts product returns, and they found that 10% of the items sold in a month
# are returned due to defects.

# If 25 items are randomly selected from a month's sales records, calculate the following:

    #(i) The expected number of items returned due to defects.
        # Probability of return = 0.10
        # randomly selected items = 25
        # Binomial distribution = (25 * 0.1) = 2.5

    #(iii) The probability that four or more items are returned.
        # P(x > 4) = 0.236408643

    # (iv) Using python, confirm your answers in questions (i)-(iii).

Explain Code | Generate Tests | Generate Docstrings | Ask Sourcery
def binomial_pmf(n, k, p):
    # Probability Mass Function for Binomial Distribution
    # P(X = k) = C(n, k) * p^k * (1-p)^(n-k)
    return math.comb(n, k) * (p ** k) * ((1 - p) ** (n - k))

Explain Code | Generate Tests | Generate Docstrings | Ask Sourcery
def binomial_cdf(n, k_max, p):
    # Cumulative Distribution Function for Binomial Distribution
    # P(X <= k_max) = sum of P(X = k) for k = 0 to k_max
    return sum(binomial_pmf(n, k, p) for k in range(0, k_max + 1))

Explain Code | Generate Tests | Generate Docstrings | Ask Sourcery
def main():
    n = 25
    p = 0.10

    # (i) Expected number of returned items
    # Binomial distribution = (25 * 0.1) = 2.5
    expected = n * p
    print(f"Expected number returned (n*p): {expected}")

    # (iii) Probability 4 or more returned = 1 - P(X <= 3)
    part_3 = 1.0 - binomial_cdf(n, 3, p)
    print(f"P(X >= 4): {part_3:.9f}")


if __name__ == "__main__":
    main()
```

```
(.venv) ryanhabis@pop-os:~/Documents/College
e folder/semester-1/statistics/CA-1/ca_1.py"
Expected number returned (n*p): 2.5
P(X >= 4): 0.236408642
```

(b) **[10 marks]**

The length of time taken for faulty product to be returned after it is sold is approximately normally distributed with a mean of 10 days and a standard deviation of 3 days.

(i) Determine the proportion of products which take between 6 and 17 days to be returned to retailer.
Mean = 10, Standard diviation = 3 , x = (6,17)

$p(6 \leq x \leq 17) = p\ (z_1 \leq z_2)$

z = (x – mean)/ sigma

z= (x- 10) / 3

$z_1 = (6 – 10) / 3 = -4 / 3 = - 1.3333$

$z_2 = (17 – 10) / 3 = 7/3 = 2.3333$

$p = (z_1 \leq z_2)$

$p = (- 1.3333 \leq 2.3333)$

Checking table

$z_1 = - 1.3333 = 0.09176$

$z_2 = 2.3333 = 0.99010$

$p = (z_1 = 1.3333 \leq z_2 = 2.3333) = 0.99010 - 0.09176 =$ **0.89834**

(ii) Find the number of days within which 95% of faulty products are returned.

Looked up the normal distribution table for the closes number to 95% which in this case is **1.64 = 0.9495** and **1.65 = 0.9505**. I went with the middle of them both **1.645.**

z = (x – mean) / standard deviation
z = 1.645
mean = 10
standard deviation = 3

1.645 = (x – 10) / 3
1.645 * 3 = ((x – 10) / 3) * 3
4.935 = x – 10
4.935 + 10 = x – 10 + 10
x = 14.935

(iii) Using python, confirm your answers in questions (i)-(ii).

```python
import scipy.stats as stats

#  Analysis of a sample of 60 faulty products found that the average length of time taken for them to be returned was 15 days.
#  Using the population data in part (b), test the claim that the average length of time taken for faulty products to be returned is 10 days.
#  Conduct the test at a 0.05 level of significance.

mean = 10
std = 3

# (i)  Determine the proportion of products which take between 6 and 17 days to be returned to retailer.
p = stats.norm.cdf(17, mean, std) - stats.norm.cdf(6, mean, std)
print("Proportion between 6 and 17 days:", p)

# (ii)  Find the number of days within which 95% of faulty products are returned.
days = stats.norm.ppf(0.95, mean, std)
print("Number of days for 95% returns:", days)
```

```
(.venv) ryanhabis@pop-os:~/Documents/College folder$
/ca_1_part_2.py"
Proportion between 6 and 17 days: 0.8989734516454868
Number of days for 95% returns: 14.934560880854416
```

(c) **[10 marks]**

Analysis of a sample of 60 faulty products found that the average length of time taken for them to be returned was 15 days. Using the population data in part (b), test the claim that the average length of time taken for faulty products to be returned is 10 days. Conduct the test at a 0.05 level of significance.

(i) Clearly state the null and alternative hypothesis in symbolic form and in context.

Symbolic form:

- $H_0 : \mu = 10$

Context

The average time it take for a broken product to be returned is 10 days.

Symbolic form:

$H_1 : \mu > 10$

Context :

The average time it takes for a broken product to be returned is greater the 10 days.

(ii) Calculate the test statistic.

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

X = 15 , μ = 10 , Ω = 3, n = 60

$$z = \frac{15 - 10}{3\sqrt{60}}$$

$$z = \frac{5}{3/\sqrt{7.7459}}$$

$$z = \frac{5}{0.3873}$$

z = 12.91

(iii) Determine the rejection region(s).

The rejection region is 5% which is 1.645.

(iv) Clearly state your conclusions (in context).

Our test statistic is 12.91 which is greater then 1.645 therefore it is far beyond the rejection region.