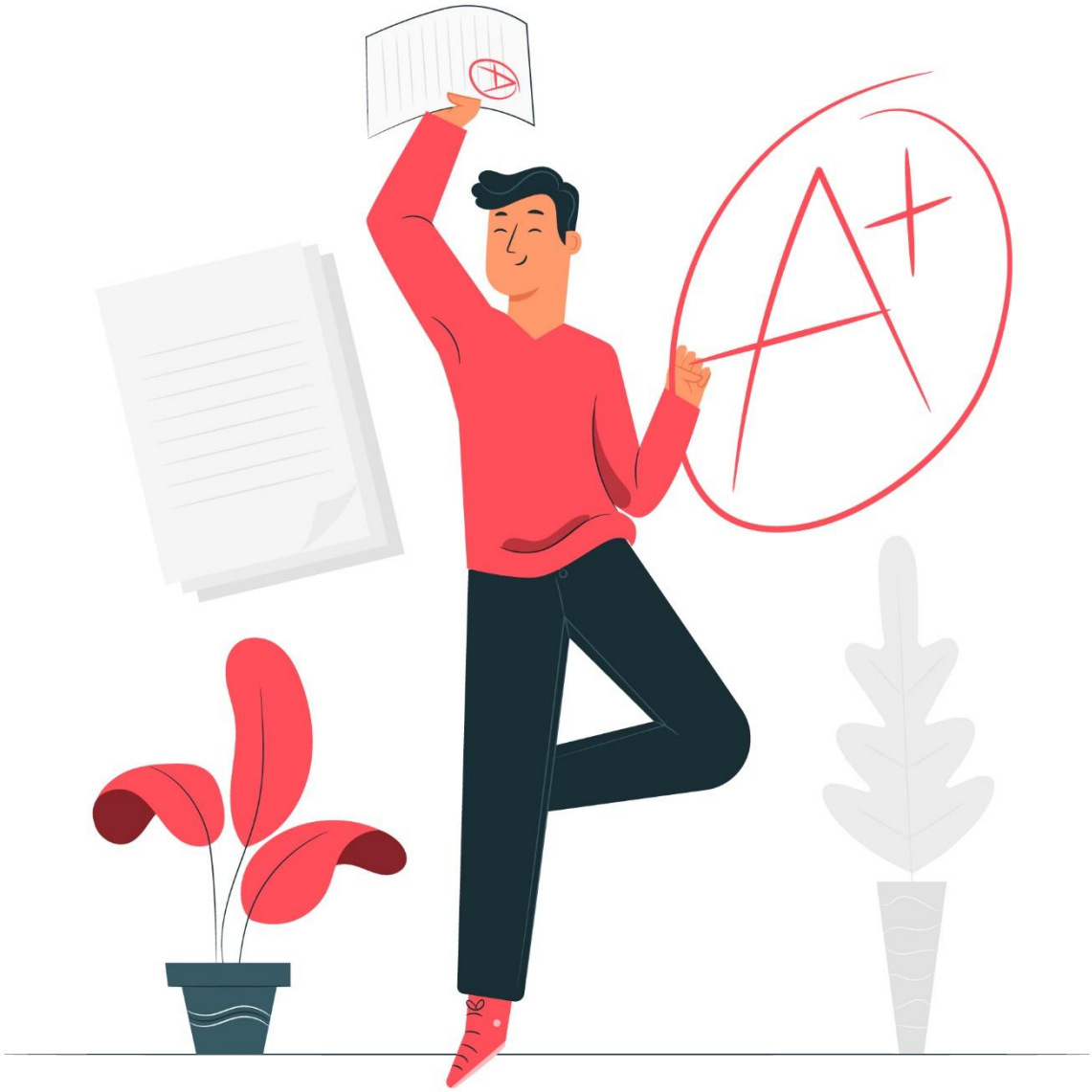
R Practical Test

Probability and Statistics – (IT2110)

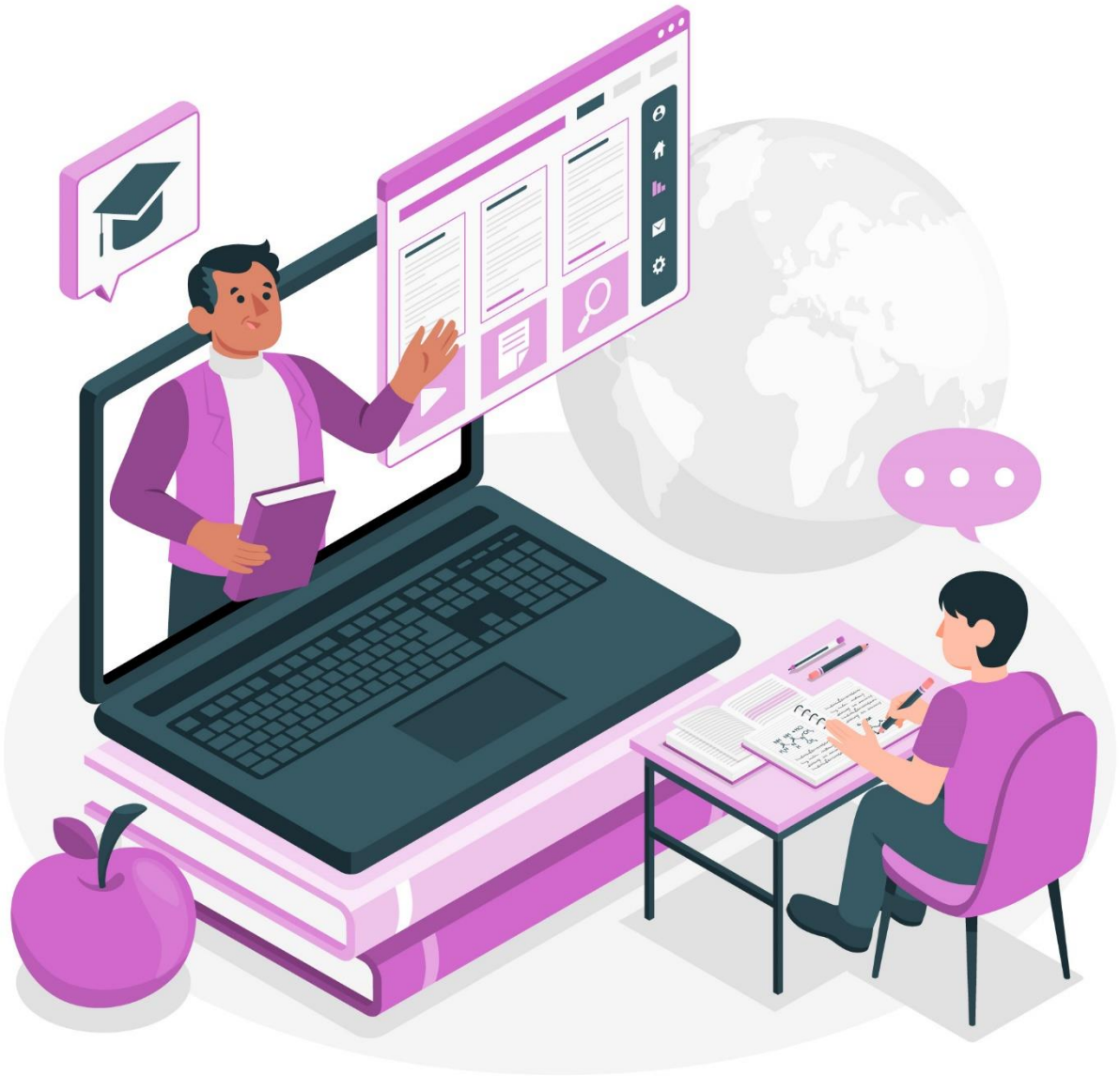


Sachi

R Test – PAS(IT2110)

Below you'll find

1. **A small synthetic dataset** (book_data.txt) with 25 book records and 8 variables.
2. **Thirty multiple-choice questions**:
    - **Q 01 – Q 15** use the dataset (you'll need to import it in R to answer).
    - **Q 16 – Q 25** are probability/stat-dist questions.
    - **Q 26 – Q 30** are short theory / syntax items about R.



Sachi

R Test – PAS(IT2110)

1 Dataset: book_data.txt

```
# book_data.txt
Title,Language,ISBN13,AvgRating,TextReviews,PageCount,PubYear,Genre
Hidden_Fires,eng,9.78e+12,3.79,312,340,2010,Romance
Silent_Code,en-US,9.79e+12,4.11,1250,410,2015,Thriller
Desert_Moon,spa,9.78e+12,4.28,540,290,2011,Adventure
Digital_Echo,eng,9.79e+12,3.64,685,370,2017,SciFi
Paper_Wings,fre,9.78e+12,3.92,810,280,2012,Drama
Crystal_Nights,en-US,9.78e+12,4.33,2500,450,2019,Fantasy
Broken_Symphony,eng,9.78e+12,3.48,230,330,2009,Romance
Quantum_Shift,eng,9.79e+12,4.57,1320,520,2020,SciFi
Whispered_Truths,spa,9.78e+12,3.96,910,305,2014,Drama
Amber_Dawn,en-US,9.79e+12,4.18,1770,398,2018,Adventure
Ion_Path,eng,9.78e+12,4.07,620,356,2016,SciFi
Velvet_Shadows,fre,9.79e+12,3.69,470,265,2013,Romance
Neon_Dreams,eng,9.79e+12,4.41,1999,510,2021,SciFi
Ironbound,eng,9.78e+12,3.54,410,380,2008,Fantasy
Lost_Horizon,spa,9.79e+12,4.26,890,420,2012,Adventure
Azure_Skies,eng,9.78e+12,3.83,350,310,2010,Drama
Clockwork_Heart,en-US,9.79e+12,4.35,2880,608,2019,Fantasy
Echoes_of_Time,fre,9.78e+12,3.71,560,275,2011,Romance
Binary_Storm,eng,9.79e+12,4.52,1450,495,2020,SciFi
Gold_Line,eng,9.78e+12,3.94,760,340,2015,Adventure
Scarlet_Code,spa,9.79e+12,4.10,980,370,2017,Thriller
Midnight_Tide,en-US,9.78e+12,3.67,430,288,2013,Drama
Silver_Gale,fre,9.79e+12,4.23,1110,415,2016,Fantasy
Polar_Flare,eng,9.78e+12,3.88,520,360,2009,Adventure
Fractal_Edge,eng,9.79e+12,4.47,1680,530,2021,SciFi
```

Sachi

2 Quiz-style questions

Dataset-driven (Q 01 – Q 15)

1. *Importing:*

Which command loads book_data.txt **as shown** so that character columns stay character (not factors)?

- A) read.csv("book_data.txt")
- B) read.csv("book_data.txt", stringsAsFactors=FALSE)
- C) read.table("book_data.txt")
- D) read.delim("book_data.txt", header=TRUE)

2. *Exporting:*

Which call correctly exports the data frame **books** to books_out.txt **including** the header?

- A) write.table(books, file="books_out.txt", sep=",", row.names=FALSE)
- B) write.txt(books, "books_out.txt", header=TRUE)
- C) read.table(books, file="books_out.txt")
- D) write.table(file="books_out.txt", books)

3. *Quartiles & IQR* – AvgRating:

Using default quantile(), the 25-th percentile is 3.79 and the 75-th is 4.28. What is the IQR?

- A) 0.49
- B) 0.25
- C) 0.74
- D) 0.85

4. *Outliers* – TextReviews:

How many outliers does boxplot.stats(books$TextReviews)$out flag?

- A) 1
- B) 2
- C) 3
- D) 4

5. *Variance* – PageCount:

What is the sample variance (rounded to nearest whole) of the variable PageCount?

- A) 8125
- B) 3755
- C) 8400
- D) 8369

6. *Grouped count:*

How many records have **Language = "en-US"** *and* ISBN13 =9.78e+12 ?

- A) 1
- B) 2
- C) 3
- D) 4

7. *Frequency interval:*

How many books have TextReviews between 500 and 3000 inclusive?

A) 17 B) 18 C) 19 D) 20

8. *Multiple bar chart (concept):*

Suppose you drew a bar chart crossing Language (x-axis) by ISBN13 (fill). Which language(s) have **more** books with ISBN 9.78e+12 than with 9.79e+12?

A) English (eng) only B) Spanish (spa) only C) French (fre) only D) English (eng) & Spanish (spa)

9. *Pie-chart sector:*

If you created a pie chart of ISBN13 counts and the slice for 9.78e+12 spans 14 observations, what is its central angle (degrees, 2-dec.)?

A) 334.29° B) 360.00° C) 201.60° D) 201.26°

10. *Mean by group:*

Using tapply() which expression gives mean PageCount by Genre?

A) tapply(PageCount, Genre, mean) B) tapply(books$PageCount, books$Genre, mean)

C) tapply(books, Genre, mean) D) Both A and B but not C

11. *sd with dplyr:*

Select the correct pipeline to compute the **SD** of AvgRating for each Genre.

A) books %>% group_by(Genre) %>% summarise(SD = sd(AvgRating))

B) books %>% summarise(SD = sd(AvgRating), by=Genre)

C) summarise(group_by(books, Genre), sd(AvgRating))

D) books %>% group_by(Genre) %>% sd(AvgRating)

12. *Histogram breaks:*

Which call forces **6 equal-width** bins from min to max of TextReviews?

A) hist(TextReviews, breaks=6)

B) hist(TextReviews, breaks=seq(min(TextReviews), max(TextReviews), length.out=7))

C) hist(TextReviews, breaks="FD")

D) hist(TextReviews, breaks=seq(min,max,6))

Sachi

13. *Frequency polygon concept:*

After binning with the breaks in Q 12, which statement best explains why class midpoints are used for the polygon's x-coords?

A) They mark the start of each class.

B) They represent the centre of mass of each class interval.

C) They avoid overlapping points.

D) They make the polygon area equal to the histogram area.

14. *attach dangers:*

If you execute attach(books) and later create a vector named PageCount, which of these is true?

A) The new vector masks the data-frame column until detach(books) is called.

B) R throws an error.

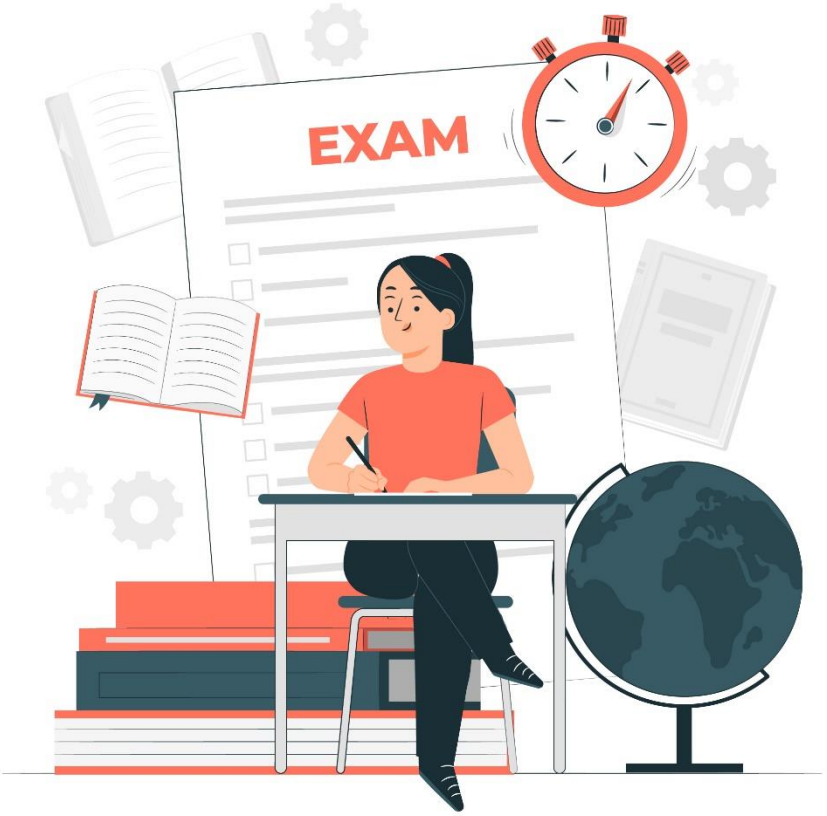
C) Both objects are merged.

D) The column masks your new vector.

15. *write vs read:*

write.table() by default separates columns with:

A) comma B) semicolon C) space/tab D) pipe (|)



Sachi

Probability & distributions (Q 16 – Q 25)

1. *Exponential:* The lifetime (months) of a bulb is Exp(λ = 0.6). What is P(T>9)?

A) 0.003 B) 0.011 C) 0.020 D) 0.027

2. *Exponential quantile:* For the same bulb, what lifetime t has P(T\le t)=0.95?

A) 4.99 mo B) 5.84 mo C) 4.32 mo D) 7.51 mo

3. *Poisson:* An insurer averages 6 claims/day. Probability of exactly 8 tomorrow?

A) 0.090 B) 0.103 C) 0.112 D) 0.122

4. *Poisson tail:* Probability of **8 or more** claims tomorrow?

A) 0.213 B) 0.254 C) 0.191 D) 0.302

5. *Binomial:* X\sim Bin(50,\,0.8). P(X=45) equals…

A) 0.022 B) 0.040 C) 0.048 D) 0.063

6. *Binomial CDF:* P(X\le40) for the same X.

A) 0.040 B) 0.081 C) 0.112 D) 0.145

7. *Normal tail:* Body temp N(36.8,0.4^2). P(T\ge37.5)?

A) 0.021 B) 0.035 C) 0.041 D) 0.056

8. *Normal quantile:* Find t with P(T\ge t)=0.01.

A) 37.46° B) 37.62° C) 37.73° D) 37.91°

9. *Uniform:* U\sim U(2,8). P(3<U<6) is…

A) 0.33 B) 0.40 C) 0.50 D) 0.60

10. *Gamma:* Y\sim\Gamma(k=3,\,θ=2). Mean of Y?

A) 3 B) 5 C) 6 D) 7

Theory / syntax (Q 26 – Q 30)

1. Which function lets you **temporarily** evaluate an expression with data-frame names visible *without* attaching?

A) within() B) with() C) transform() D) subset()

2. In base plotting, parameter las = 2 makes axis tick labels…

A) parallel to axis B) horizontal C) perpendicular to axis D) vertical

3. varwidth=TRUE inside boxplot() causes…

A) boxes coloured by group size B) whisker length $\propto n$

C) box width $\propto \sqrt{n}$ D) notches drawn around medians

4. The operator %>% is re-exported from package…

A) magrittr B) tidyr C) purrr D) ggplot2

5. rm(list = ls()) does what?

A) Removes attached data frames B) Deletes all objects in global env.

C) Refreshes the search path D) Clears plots but not variables

---