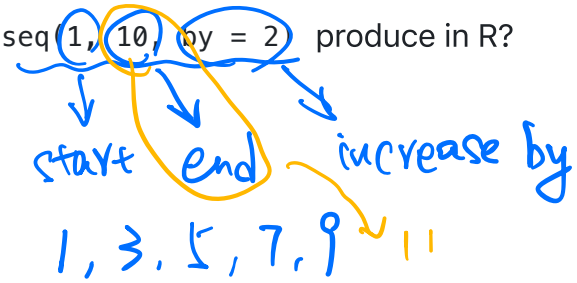
Multiple Choice about R grammar

1. What does the function `seq(1, 10, by = 2)` produce in R?

   - A) 1, 2, 3, 4, ..., 10
   - B) 1, 3, 5, 7, 9
   - C) 1, 4, 7, 10
   - D) 2, 4, 6, 8, 10



2. Which of the following is **not** a valid way to create a vector in R?

   - A) `c(2, 5, 7)`
   - B) `1:5` ⇒ 1, 2, 3, 4, 5
   - C) `vector(3)` error
   - D) `array(1, 3)` ⇒ c(1, 1, 1)

3. What is the result of `length(c(4, 9, 2, NA, 7))` ?

   - A) 4
   - B) 5
   - C) NA
   - D) Error

4. Given `x <- c(10, 20, 30)`, what is `x * 2` ?

   - A) `c(20, 40, 60)`

vector scalar ⇒ * : elementwise

- B) `c(10, 20, 30, 10, 20, 30)`
- C) `c(12, 24, 36)`
- D) Error

5. What does `NA` represent in R?

- A) The string `"NA"`
- B) Missing data / Not Available
- C) A logical FALSE
- D) Zero

6. What is the output type of `mean(c(1, 2, 3, 4))` ?

- A) integer
- B) numeric (double)
- C) character
- D) logical

7. Which operator in R is used for elementwise logical AND?

- A) `&`
- B) `&&`
- C) `|`
- D) `||`

8. What is the difference between `&` and `&&` ?

- A) `&` works only with scalars, `&&` is vectorized
- B) `&` does elementwise logical AND, `&&` only returns a single TRUE/FALSE (with short-circuit)
- C) They are identical
- D) `&&` works only on numeric vectors

9. What does the function `is.na()` do?

- A) Tests if a value is not numeric
- B) Tests if a value is `NA` (missing)
- C) Tests if a value is `NaN`
- D) Converts a value to NA

10. Suppose `x <- c(1, 2, NA, 4)`. What is the result of `sum(x)` ?

- A) 7
- B) NA
- C) 8
- D) Error

11. How can you instruct `sum()` to ignore `NA` values?

- A) `sum(x, remove = TRUE)`
- B) `sum(x, na.rm = TRUE)`
- C) `sum(x, na = TRUE)`
- D) `sum(x, ignore.na = TRUE)`

12. Which of the following is **not** a mode (or atomic type) in R?

- A) numeric
- B) logical
- C) factor
- D) character

a ← c("1", "2", "1", "2", "3")
factor(a) ⇒ levels: "1", "2", "3"
freq: 2, 2, 1

13. Suppose you have `y <- c("a", "b", "c")`. What is `y[2]`?

- A) `"a"`
- B) `"b"`
- C) `"c"`
- D) `NA`

14. What happens if you access an index out of bounds, e.g. `y[10]` for a vector of length 3?

- A) Error
- B) `NULL`
- C) `NA`
- D) The vector recycles

15. Which function gives the unique values of a vector?

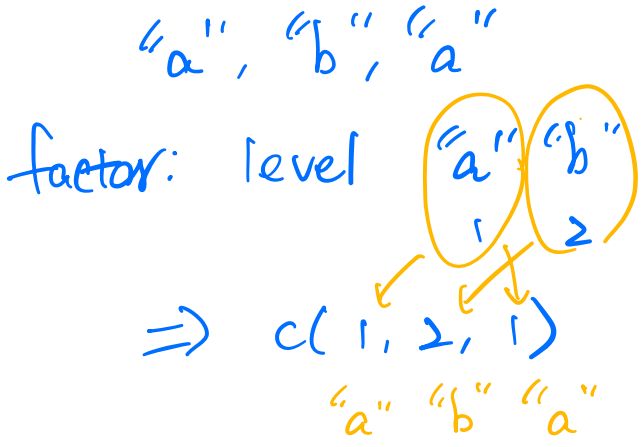
- A) `unique()`
- B) `distinct()`
- C) `uniq()`
- D) `levels()`

16. What does `factor()` do?

- A) Converts a numeric vector to binary
- B) Converts a vector into a factor (categorical) type
- C) Returns factorials
- D) Converts character to numeric

17. What is the output of `as.numeric(factor(c("a", "b", "a")))` ?

- A) `c(1, 2, 1)`
- B) `c("a", "b", "a")`
- C) `c(0, 1, 0)`
- D) `c("1", "2", "1")`



18. What does `data.frame()` create?

- A) A matrix
- B) A list
- C) A tabular structure with equal-length columns
- D) A vector

19. If `df <- data.frame(a = c(1,2), b = c("x","y"))` , what is `df$a` ?

- A) A data frame
- B) The column named “a” as a vector
- C) The entire data frame
- D) Error

20. What does the `str()` function do when given an R object?

- A) Prints only the names of the object
- B) Gives the structure (internal representation) of the object
- C) Summarizes statistical properties
- D) Converts object to string

Open Questions in R Grammar

21. **Given this R snippet, what is the output? Explain step by step.**

```
x <- c(5, NA, 10, 15)
mean_x <- mean(x)
```

create a vector of [5, NA, 10, 15]

calculate the mean of the vector

(NA ⇒ NA)

```
total <- sum(x, na.rm = TRUE)
c(mean_x, total)
```

22. **Identify the bug / error in the following R code, and suggest a fix.**

```
v <- c(2, 4, 6, 8)
w <- c(1, 2)
result <- v + w
print(result)
```

23. Write a short program in R that takes a numeric vector `v` and returns a vector of the same length where each entry is TRUE if the corresponding entry in `v`` is above the mean of `v` (ignoring NAs), and FALSE otherwise.

24. **Given the following code, describe what it does (in plain English):**

```
df <- data.frame(id = 1:5, score = c(10, 15, NA, 20, 18))
df$above_avg <- df$score > mean(df$score, na.rm = TRUE)
subset(df, above_avg == TRUE)
```

25. **What will be the result (or error) of this code? Explain.**

```
x <- factor(c("low","medium","high","low"))
as.numeric(x) + 1
levels(x)
```

Open Questions in Matrix Operation

26. Suppose

```
A <- matrix(c(1, 2, 3, 4), nrow = 2, byrow = TRUE)
B <- matrix(c(2, 0, 1, 2), nrow = 2, byrow = TRUE)
A %*% B
```

What is the result of A %*% B?

- A) `matrix(c(4, 4, 10, 8), nrow = 2, byrow = TRUE)`
- B) `matrix(c(2, 2, 6, 8), nrow = 2, byrow = TRUE)`

- C) matrix(c(4, 4, 8, 10), nrow = 2, byrow = TRUE)

- D) Error (dimensions do not match)

27/ Which operator is used for matrix multiplication in R?

- A) *

- B) %*%

- C) .


```
M <- matrix(1:6, nrow = 2, ncol = 3)
N <- matrix(1:6, nrow = 3, ncol = 2)
M %*% N
```

What is the dimension of the result?

- A) 2×2

- B) 3×3

- C) 2×3

- D) Error

29. Suppose:

```
X <- matrix(c(1, 2, 3, 4), nrow = 2)
Y <- matrix(c(5, 6), nrow = 2)
X %*% Y
```

What happens?

- A) 2×1 matrix result

- B) 1×2 matrix result

- C) Error (non-conformable arguments)

- D) A scalar

30. Write pseudo-code for multiplying two matrices A (m×n) and B (n×p) to produce matrix C (m×p). Clearly specify the loop structure you would use. You do not need to write actual R code, just outline the algorithm in clear steps.

Regresison Questions

31. In the linear regression model $y = X\beta + \epsilon$, which of the following is the closed-form solution for $\hat{\beta}$?

    - A) $X^\top y$
    - B) $(X^\top X)^{-1} X^\top y$
    - C) $(XX^\top)^{-1} y$
    - D) $X(X^\top y)^{-1}$

32. Under the standard linear regression assumptions, which of the following is true about bias and variance of $\hat{\beta}$?

    - A) $\hat{\beta}$ is unbiased, variance depends on $\sigma^2 (X^\top X)^{-1}$
    - B) $\hat{\beta}$ is biased, variance is always zero
    - C) $\hat{\beta}$ is unbiased, variance does not depend on X
    - D) $\hat{\beta}$ is biased, variance depends only on sample size

33. In R, which of the following correctly computes the OLS estimate?

    - A) beta_hat <- solve(t(X) %*% *X)* %*% t(X) %*% y
    - B) beta_hat <- X %*% *t(X)* %*% y
    - C) beta_hat <- lm(X, y)
    - D) beta_hat <- solve(X) %*% y

34. Explain in words what the bias-variance tradeoff means in the context of linear regression.

35. Suppose you fit a regression model in R using:

```r
model <- lm(y ~ x1 + x2, data = df)
summary(model)
```

Explain the following concepts:

The coefficient estimate of x1.

The p-value associated with the coefficient estimate of x1.

36. The following is the output of a summary function:

```
Call:
lm(formula = y ~ x1 + x2, data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-2.345  -0.876  -0.123   0.754   2.567

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.2350     0.4321   2.86   0.005**
x1            0.5678     0.0987   5.75  1.2e-07***
x2           -0.2345     0.1123  -2.09   0.039*

Residual standard error: 1.05 on 96 degrees of freedom
Multiple R-squared: 0.642, Adjusted R-squared: 0.631
F-statistic: 58.4 on 2 and 96 DF,  p-value: < 2.2e-16
```

Yes, statistically significant

Questions:

- How would you interpret the coefficient of x1?

increase x1 by 1, $\hat{y}$ ↑ by 0.5678

- Is x2 statistically significant at the 5% level? Why or why not?

- How do you interpret the F-test and its result?

β_1 or $\beta_2 \neq 0$

$H_0: y = \beta_0$ H_a: at least one of x1 and x2 is important in the model

Please revise the inclass practices as well.

v = c(1, 2, 3, 4)

w = c(1, 2)

v+w ⇒ no warning, v+w = c(2, 4, 4, 6)

---

v = c(1, 2, 3, 4)

w = c(1, 2, 3)

v+w ⇒ with warning, v+w = c(2, 4, 6, 5)

```
mult = function(A, B){
    m = the #row of A
    n = the #col of A
    p = #col B
    res = m x p matrix
    for i in 1, ..., m (rows of A)
        for j in 1, ..., p (cols of B)        B[:, j]
            res[i, j] : calculate the inner product of A[i, :] and
        end for
    end for
    return(res)
}
```