

# M5 Problem Set Distributions, Probability and z-score for TAs

Pablo E. Gutiérrez-Fonseca

2025-02-14 12:18:19

**R practice.**

```
library(BSDA)
```

```
## Warning: package 'BSDA' was built under R version 4.3.2
```

```
library(tidyverse)
```

```
## Warning: package 'tidyverse' was built under R version 4.3.3
```

```
## Warning: package 'ggplot2' was built under R version 4.3.3
```

```
## Warning: package 'tibble' was built under R version 4.3.1
```

```
## Warning: package 'tidyr' was built under R version 4.3.2
```

```
## Warning: package 'readr' was built under R version 4.3.2
```

```
## Warning: package 'purrr' was built under R version 4.3.1
```

```
## Warning: package 'dplyr' was built under R version 4.3.2
```

```
## Warning: package 'stringr' was built under R version 4.3.2
```

```
## Warning: package 'forcats' was built under R version 4.3.2
```

```
## Warning: package 'lubridate' was built under R version 4.3.3
```

```
library(lessR)
```

```
## Warning: package 'lessR' was built under R version 4.3.3
```

```
library(readxl)
```

```
#open up our data
sleeping <- read_xlsx("Sleeping_Mammals1.xlsx")
```

1. Based on the data, mammals spend most of their time in which type of sleep phase (dreaming or non-dreaming)?

```
#pivot gives us more summary statistics, for Non-Dreaming sleep:
pivot(sleeping, c(IQR, skew, kurtosis, mean, sd, var), NonDreaming)
```

```
##   n na IQR  skew   kurt  mean    sd   var
##  49 13 4.9 0.272 -0.289 8.541 3.744 14.018
```

```
#and dreaming sleep.
pivot(sleeping, c(IQR, skew, kurtosis, mean, sd, var), Dreaming)
```

```
##   n na IQR  skew   kurt  mean    sd   var
##  51 11 1.6 1.455 2.328 1.941 1.445 2.088
```

2. Which type of sleep phase has the highest variability across the species included here (dreaming or non-dreaming)?
3. What is the p-value for the goodness of fit test for the Dreaming variable (NOTE..just enter the number (no letters, symbols, equal signs etc).also be careful of decimal places.

```
shapiro.test(sleeping$Dreaming)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  sleeping$Dreaming
## W = 0.87556, p-value = 0.00007067
```

4. Based on this goodness of fit test, is the dreaming variable normally distributed? (You can check out this page if you are wondering. NO
5. What is the p-value for the goodness of fit test for the non-Dreaming variable (NOTE..just enter the number (no letters, symbols, equal signs etc).also be careful of decimal places.

```
shapiro.test(sleeping$NonDreaming)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  sleeping$NonDreaming
## W = 0.98257, p-value = 0.6763
```

6. Based on this goodness of fit test, is the non-Dreaming variable normally distributed? (You can check out this page if you are wondering. YES
7. From the table, enter the associated z-score for small class for 'baboon'.

```

# Filter small mammals
small_mammals <- subset(sleeping, Mammalian_Size_Group == "small")

# Extract Baboon's NonDreaming sleep value
baboon_value <- sleeping[sleeping$Species == "Baboon", "NonDreaming"]

# Compute mean and standard deviation for small mammals
mean_small <- mean(small_mammals$NonDreaming, na.rm = TRUE)
sd_small <- sd(small_mammals$NonDreaming, na.rm = TRUE)

# Compute z-score for Baboon's NonDreaming sleep within small mammals
z_score_baboon <- (baboon_value - mean_small) / sd_small

# Print the result
print(z_score_baboon)

```

```

##      NonDreaming
## 1  -0.2011717

```

13. Based on your knowledge that higher (absolute value) zscores are more unusual for a given group's distribution, which class would you assign baboon to (small or large)? (hint compare baboon's z scores for the small and large group..which group does the species z score fit "best" with...ie for which group is that species z score less extreme?

```

# Filter large mammals
large_mammals <- subset(sleeping, Mammalian_Size_Group == "large")

# Compute z-score for large mammals (using the same method as for small)
mean_large <- mean(large_mammals$NonDreaming, na.rm = TRUE)
sd_large <- sd(large_mammals$NonDreaming, na.rm = TRUE)

z_score_baboon_large <- (baboon_value - mean_large) / sd_large

# Calculate absolute values of the z-scores
abs_z_score_small <- abs(z_score_baboon)
abs_z_score_large <- abs(z_score_baboon_large)

# Compare which z-score is less extreme
if (abs_z_score_small < abs_z_score_large) {
  group_assigned <- "small"
} else {
  group_assigned <- "large"
}

# Print the result
print(paste("Baboon is assigned to the", group_assigned, "group."))

```

```

## [1] "Baboon is assigned to the small group."

```

16. Based on your knowledge that higher (absolute value) zscores are more unusual for a given group's distribution, which class would you assign roe deer to (small or large)? (hint compare roe deer's z scores for the small and large group..which group does the species z score fit "best" with...ie for which group is that species z score less extreme?

17. Identify the upper z-score that represents  $< 1\%$  TOTAL chance of occurring. This doesn't require any calculations, just the determination of the z-score from the normal probability table (or R function).

```
lower_z_score <- qnorm(1-0.01)
print(lower_z_score)
```

```
## [1] 2.326348
```

- 18 Identify the lower z-score that represents  $< 1\%$  TOTAL chance of occurring. This doesn't require any calculations, just the determination of the z-score from the normal probability table (or R function).

```
lower_z_score <- qnorm(0.01)
print(lower_z_score)
```

```
## [1] -2.326348
```

19. What is the actual nondreaming value associated with the upper z score identified in the question above?

```
# Calculate the mean and standard deviation, excluding NA values
mean_small <- mean(sleeping$NonDreaming, na.rm = TRUE)
sd_small <- sd(sleeping$NonDreaming, na.rm = TRUE)

# Calculate the NonDreaming value for the upper z-score (2.33)
z_score_upper <- 2.33
value_upper <- mean_small + z_score_upper * sd_small

# Print the result
print(value_upper)
```

```
## [1] 17.26444
```

20. How much non-dreaming sleep would a mammal need to sleep to be in the top 5% of sleepy mammals?

```
# Calculate the 95th percentile value for NonDreaming sleep
percentile_95 <- qnorm(0.95, mean = mean_small, sd = sd_small)
# Print the result
print(percentile_95)
```

```
## [1] 14.69922
```