# Problem set, group 3

## Alison, Laura, and Dylan

### 2024-12-02

```r
# keep this chunk in all your RMarkdown scripts
knitr::opts_chunk$set(echo = TRUE)
knitr::opts_chunk$set(tidy.opts = list(width.cutoff = 60), tidy = TRUE)
```

```r
# List required packages
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become error
```

```r
library(openxlsx)
library(lmtest)
```

```
## Loading required package: zoo
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```r
## Initial data loading, transforming
mammals <- read.xlsx("mammals_sleep.xlsx")
colnames(mammals) <- sub(pattern = "\\.\\((.*?)\\)", replacement = "",
    colnames(mammals))
colnames(mammals) <- sub(pattern = "\\((.*?)\\)", replacement = "",
    colnames(mammals))

mammals[, 3:12][mammals[, 3:12] == -999] <- NA
```

```
## analysis

summary_table <- mammals %>%
    group_by(order) %>%
    summarize(total = n(), mean_gest = mean(gestation.time, na.rm = TRUE),
        sd_gest = sd(gestation.time, na.rm = TRUE), median_gest = median(gestation.time,
            na.rm = TRUE), iqr = IQR(gestation.time, na.rm = TRUE),
        var = var(gestation.time, na.rm = TRUE), se = sd_gest/sqrt(n()),
        CI_low = mean_gest - 1.96 * se, CI_high = mean_gest +
            1.96 * se)
summary_table
```

```
## # A tibble: 18 x 10
##     order   total mean_gest sd_gest median_gest   iqr     var     se CI_low CI_high
##     <chr>   <int>     <dbl>   <dbl>       <dbl> <dbl>   <dbl>  <dbl>  <dbl>   <dbl>
## 1  Afroso~     1      60      NA          60      0     NA     NA     NA      NA
## 2  Artiod~     7     241.    134.        151    192.  17898.  50.6  142.    340.
## 3  Carniv~     8     102.     93.2        63     20    8682.  32.9   37.0   166.
## 4  Chirop~     2      42.5    10.6        42.5    7.5   112.   7.5   27.8    57.2
## 5  Cingul~     2     120      NA         120      0     NA     NA     NA      NA
## 6  Didelp~     2      13       1.41       13      1       2    1     11.0    15.0
## 7  Diprot~     2      25      11.3        25      8     128    8      9.32   40.7
## 8  Erinac~     2      42      NA          42      0     NA     NA     NA      NA
## 9  Eulipo~     1      21.5    NA          21.5    0     NA     NA     NA      NA
## 10 Hyraco~     3     217.     14.4       225     12.5   208.   8.33 200.    233
## 11 Lagomo~     1      31      NA          31      0     NA     NA     NA      NA
## 12 Monotr~     1      28      NA          28      0     NA     NA     NA      NA
## 13 Periss~     3     364.     28.0       365     28     784.  16.2  333.    396.
## 14 Primate    10     182.     57.5       175     79    3306.  18.2  147.    218.
## 15 Probos~     2     634.     14.8       634.    10.5   220.  10.5  614.    655.
## 16 Rodent~    11      40.4    28.0        30     20.5   785.   8.45  23.8    56.9
## 17 Scande~     1      46      NA          46      0     NA     NA     NA      NA
## 18 Sorico~     3      36       8.49       36      6      72    4.90  26.4    45.6
```

```
summary_table %>%
    filter(total > 1) %>%
    ggplot(aes(y = order, x = mean_gest, fill = order)) + geom_col(show.legend = FALSE) +
    geom_errorbar(aes(xmin = CI_low, xmax = CI_high), width = 0.3,
        color = "black")
```
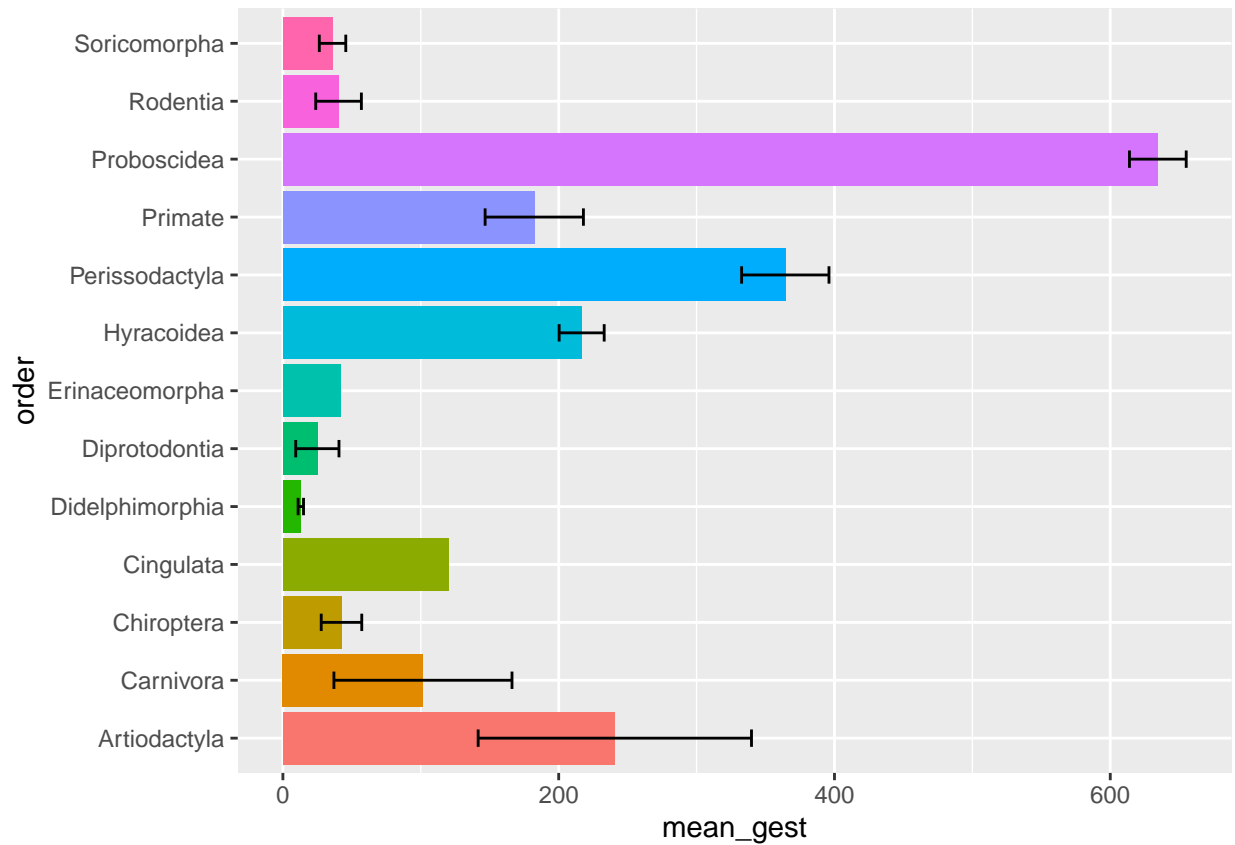
```
na.omit(mammals) %>%
    ggplot(aes(x = total.sleep, y = gestation.time)) + geom_point() +
    geom_smooth(aes(color = "linear"), se = FALSE, method = "lm") +
    geom_smooth(aes(color = "loess"), se = FALSE, span = 0.9)
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```
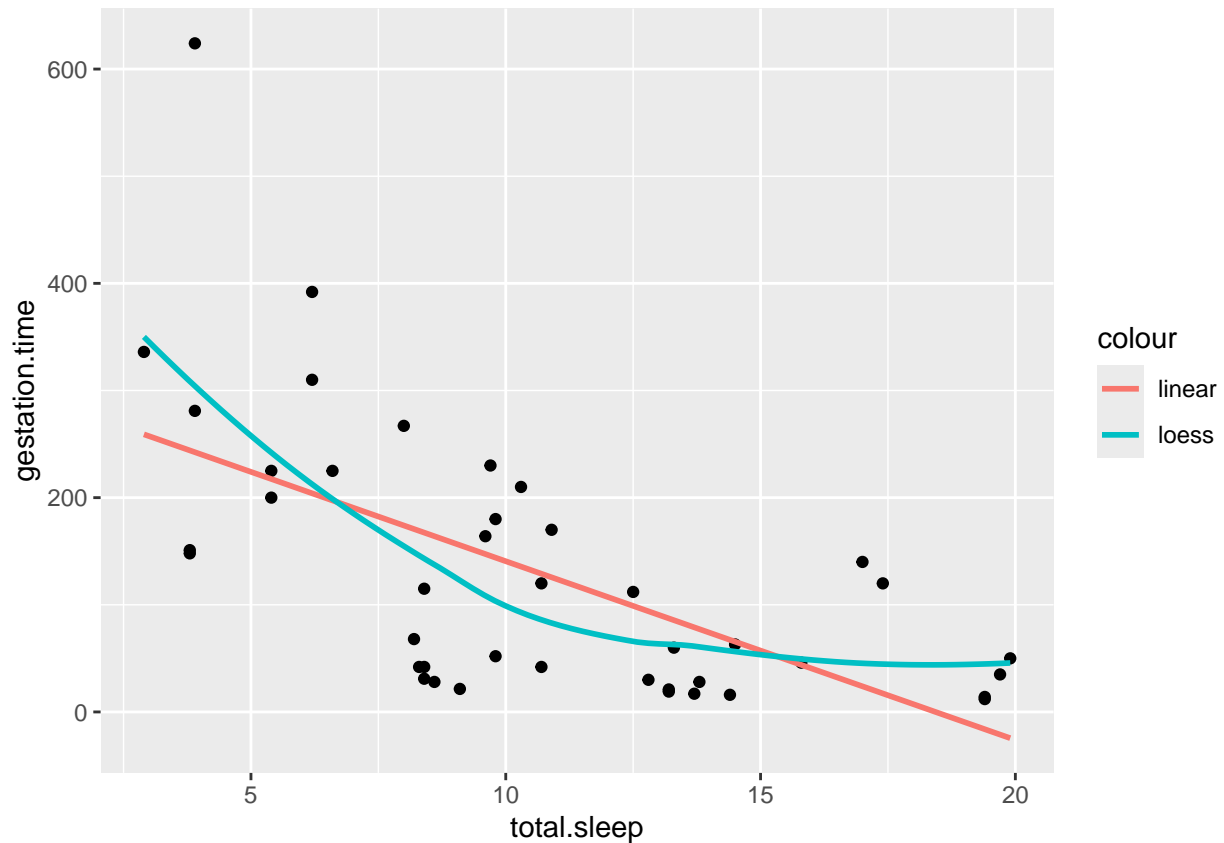
```r
cor.test(x = na.omit(mammals)$total.sleep, y = na.omit(mammals)$gestation.time)
```

```
##
##  Pearson's product-moment correlation
##
## data:  na.omit(mammals)$total.sleep and na.omit(mammals)$gestation.time
## t = -4.926, df = 40, p-value = 1.499e-05
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.7738778 -0.3818580
## sample estimates:
##        cor
## -0.6144743
```

```r
na.omit(mammals) %>%
    ggplot(aes(x = gestation.time)) + geom_histogram(binwidth = 30,
    na.rm = TRUE)
```
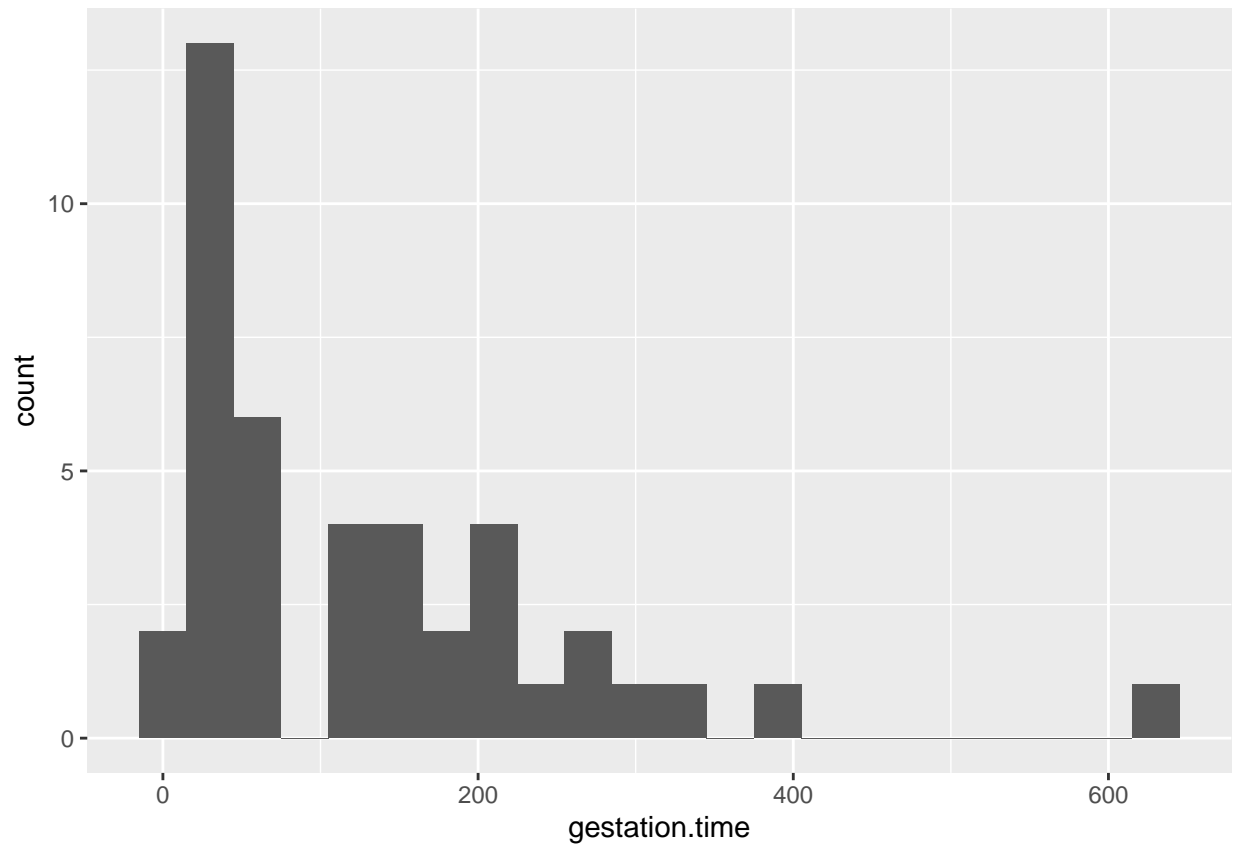
```
ggplot(na.omit(mammals), aes(sample = gestation.time)) + stat_qq(show.legend = FALSE) +
    geom_abline(aes(intercept = mean(x = gestation.time), slope = sd(x = gestation.time)))
```
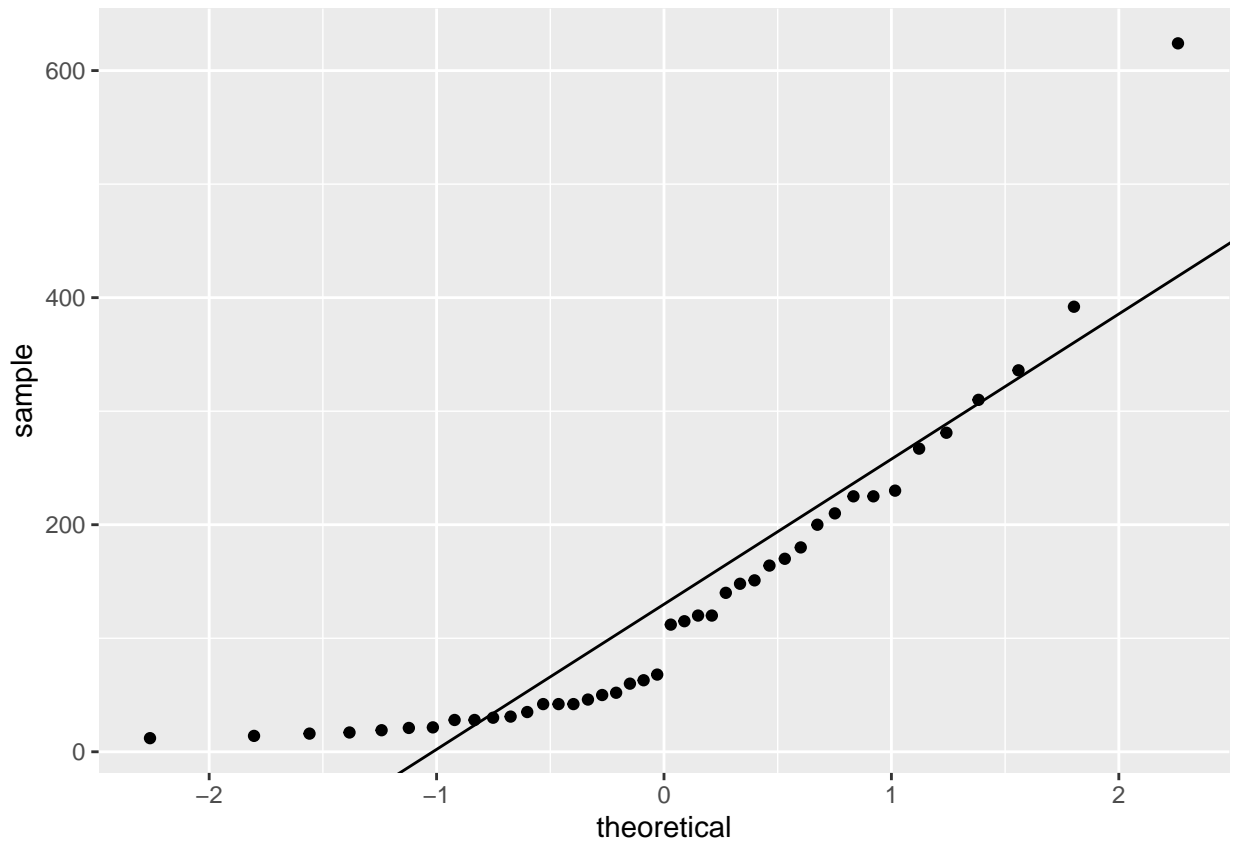
```
test_hom <- lmtest::bptest(na.omit(mammals)$total.sleep ~ na.omit(mammals)$gestation.time)
test_hom
```

```
##
##  studentized Breusch-Pagan test
##
## data:  na.omit(mammals)$total.sleep ~ na.omit(mammals)$gestation.time
## BP = 0.83182, df = 1, p-value = 0.3617
```

```
## 1. A rationale for your hypothesis, and a clear
## formulation of what you predict if your hypothesis is
## supported.

## Hypothesis: Gestation time correlates with sleep time
## within the an animal's order.  Prediction: Longer
## gestation time will relate to longer sleep time on
## average by order.  Null hypothesis: There is no
## relationship between sleep time and gestation tine.
```

```
## 2. A description of any potential changes to the dataset
## that were necessary to prepare it for analysis, such as
## recoding of variables, transformations, etc.

## We had to add a new column containing the order of each
## animal in order to group them.
```

```
## 3. A summary of your exploratory analyses, with key
## results reported in a standard format. As part of this
## section, please include some basic descriptive stats of
## your variables, incl groups

## 4. Results of your main analyses, reported in a standard
## format (can include a well formatted table, if more
## efficient than writing out all details). As part of this
## section, you need to clearly state what variables went
## into your analysis in what fashion (dependent,
## independent, original, transformed, recoded, whatever),
## and you need to report any special parameters in R
## functions (beyond the default ones) and justify why you
## selected them. This section also includes any post-hoc
## analyses that you conducted.

## 5. A graphical depiction of your results appropriately
## labeled and referenced in your writing. Each graph
## should have a short, descriptive caption that explains
## what is shown in the graph.

## 6. A brief discussion of the biological significance of
## your findings, i.e. an interpretation of your results
## based on the outcome of your hypothesis testing.
```