

# Homework 2 As Assigned

Ryo Iwata

24 February, 2024

## Homework 2

### Instructions:

0. Start early and leave time to review and revise your submission well in advance of the due date.
1. Read the problem in its entirety before starting. Often, we can ignore certain variables while working on early problems and then pay attention to them in later problems. Students can waste time trying to incorporate these effects too early and with limited guidance on how to do this.
2. Plan your approach to tidying the data and preparing it for figures and analysis.
3. Translate your plan into well structured comments (in your code chunks)
4. Translate your comments into code, leaving the comments in place (so we understand your intent)
5. Ensure your code, figure and text answers are neat and tidy, grammatically correct, clear, and concise. You may find the reindent lines (`ctrl-i`) or reformat code (`ctrl-A` i.e. `ctrl-shift-a`) useful. Break arguments to functions onto different lines and ensure your code and comments do not flow off the edge of the page.
6. Review your output (pdf) to ensure your outputs render well.
7. All tables, formulae, etc., should be carefully typeset. Do not just dump outputs from functions.

Note: We will grade both your answer and how you arrived at said answer (i.e., your code). Be sure to use sensible and sufficiently descriptive tibble, variable and model names. Make efficient and appropriate use of the tidyverse functionality.

## Problem #1 (60 pts):

Your laboratory is evaluating the diffusion of drug delivery vehicles (particles) injected into the brain. The lab is seeking to understand how the diameter *and* surface charge of the drug delivery vehicle affects the vehicle's diffusion.

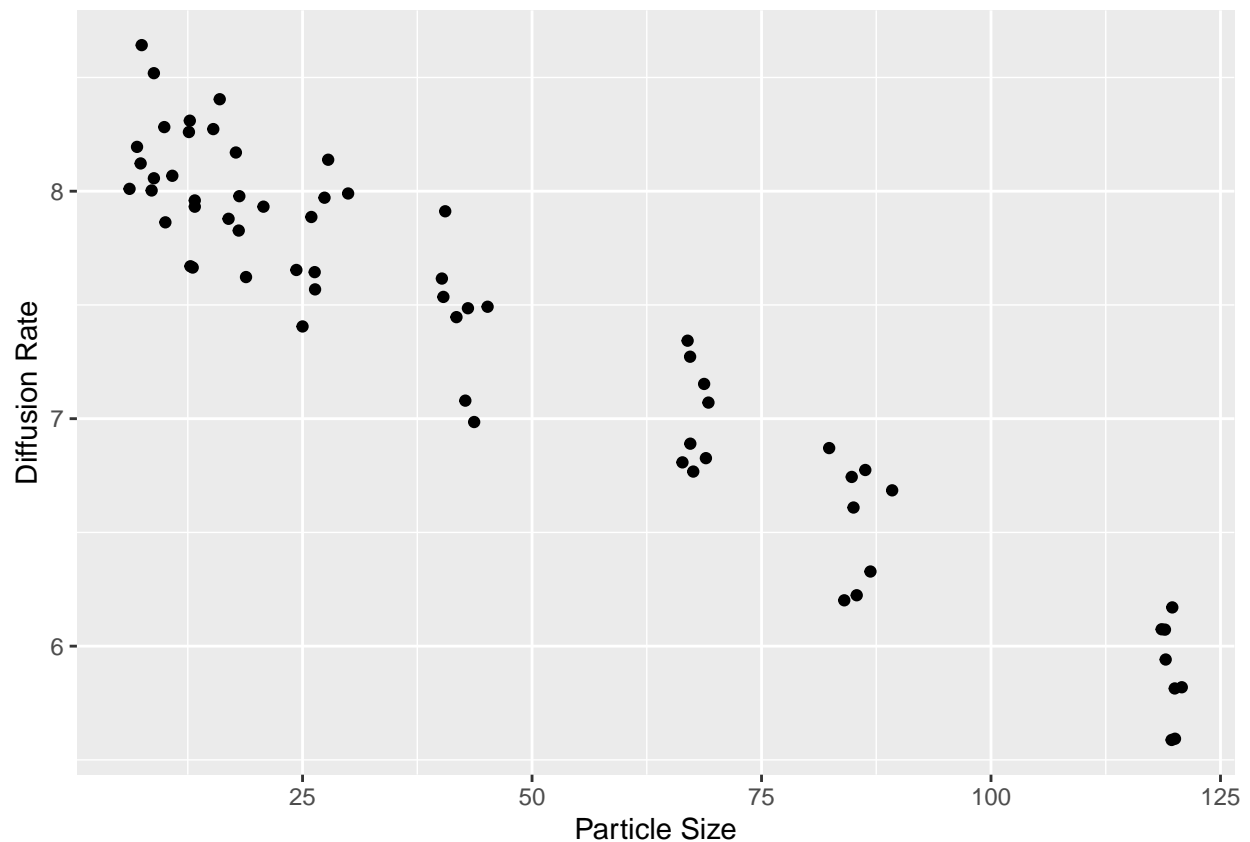
Load and evaluate the structure of data in HW2Data sheet DiffSizeCharge (only workbook/sheet in file)

```
drug_diffusion_raw <- readxl::read_excel("./data/HW2data.xlsx", sheet="DiffSizeCharge")
```

1) Investigate Particle Size and Diffusion Rate. At this point, ignore the effects of surface charge.

a) Create a plot to visualize the effect of particle diameter on particle diffusion. Think about the regression model you may be fitting, and place the appropriate variables on the x axis and y axis. Do not include any summary or smooth geometries (2 pts)

```
drug_diffusion_raw |> ggplot(aes(x=`Particle Size`, y=`Diffusion Rate`)) +  
  geom_point()
```



c) What are the model's estimated coefficients? Write out the generic formula for this model (i.e., with  $\beta$  terms ) and then again with the estimates correctly substituted (4 pts)

```
tidy_diffusion_model <- tidy(diffusion_model)

kable(tidy_diffusion_model)
```

term	estimate	std.error	statistic	p.value
(Intercept)	8.3279	0.0489	170.23	0
Particle Size	-0.0204	0.0008	-25.18	0

d) Calculate and present in a table the  $SS_x$ ,  $SS_y$ ,  $SS_{xy}$ ,  $SS_{\text{Regression}}$ , and  $SS_{\text{Error}}$  for this model. (10 pts)

e) What is  $R^2$  for this model (2 pts)?

f) Using the estimated slope term ( $\beta_1$ ) and its associated error, use a t-test to evaluate whether the slope term is zero or non-zero. What is the t-statistic and associated p-value (2 pts)?

g) Does particle size have a significant effect on particle diffusion with this model? (2 pts)

2) In the same experiment, these particles had different surface charges. You decide to investigate the effect of surface charge on particle diffusion.

a) Create a plot to visualize the effect of surface charge on particle diffusion in whole blood. At this point, ignore the effect of particle diameter (2 pts).

b) Fit a univariate (1-variable) regression model to these data. (4 pts)

c) What are the model's estimated coefficients? Write out the formula for the model with generic terms as well as substituting the estimated coefficients (2 pts)

d) Calculate and present in a table the  $SS_x$ ,  $SS_y$ ,  $SS_{xy}$ ,  $SS_{\text{Regression}}$ , and  $SS_{\text{Error}}$  for this model. (10 pts)

e) What is  $R^2$  for this model (2 pts)?

f) Using the estimated slope term ( $\beta_1$ ) and its associated error, use a t-test to evaluate whether the slope term is zero or non-zero. What is the t-statistic and associated p-value (2 pts)?

g) Does particle size have a significant effect on particle diffusion in this model? (2 pts)

### 3) Multiple Regression

- a) Create a 2-D scatterplot to visualize combined effect of surface charge and particle diameter on particle diffusion in whole blood. Think about the regression model you may be fitting, and use appropriate aesthetic mappings. Do not include summary or smooth geoms/stats (3 pts).
- b) Fit a 2-variable regression model to these data. Do not include any interactions or additional terms in your model. (4 pts)
- c) What are the model coefficients and how do they compare to the prior models' coefficients? Write out all three model equations and evaluate. (6 pts)
- d) What are  $SS_{\text{ParticleSize}}$ ,  $SS_{\text{SurfaceCharge}}$ ,  $SS_{\text{ParticleSizeY}}$ ,  $SS_{\text{SurfaceChargeY}}$ ,  $SS_Y$ ,  $SS_{\text{Regression}}$ , and  $SS_{\text{Error}}$  (10 pts)?
- e) What is  $R^2$  for this model (2 pts)?
- f) How does this model's  $R^2$  compare to the prior two models, and why is it so? (4 pts)?
- g) Use a t-test to evaluate whether each term is zero or non-zero. (4 pts)?
- h) Does size have a significant effect on particle diffusion and, if so, what is that effect? (2 pt)
- i) Does surface charge have a significant effect on particle diffusion and, if so, what is that effect? (2 pt)
- j) Extend your figures (use your code for your figures from part a of this problem as a base) to show both the raw data and the model using `geom_abline`. Do not use `stat_smooth` or other related functions. Think very carefully about where your model's plane is and where/how to represent it in your figures. (8 pt)
- k) If you ran the experiment that resulted in this data, where would you put your focus for optimization of diffusion rate and why? (5 pts)