

---

## Portfolio

An exploration of linear mixed models and common misconceptions in statistics

Erica Zhou

## Contents

|   |           |
|---|-----------|
| <b>Introduction</b>   | <b>3</b>  |
| <b>Statistical skills sample</b>  | <b>4</b>  |
| Task 1: Setting up libraries and seed value . . . . .   | 4         |
| Task 2a: Return to Statdew Valley: exploring sources of variance in a balanced<br>experimental design (teaching and learning world) . . . . . | 4         |
| Task 2b: Applying linear mixed models for the strawberry data (practical world) . . .   | 8         |
| Task 3a: Building a confidence interval interpreter . . . . .   | 11        |
| Task 3b: Building a p value interpreter . . . . .   | 12        |
| Task 3c: User instructions and disclaimer . . . . .   | 14        |
| Task 4: Creating a reproducible example (reprex) . . . . .  | 15        |
| Task 5: Simulating p-values . . . . .   | 16        |
| <b>Writing sample</b>   | <b>21</b> |
| References . . . . .  | 22        |
| <b>Reflection</b>   | <b>23</b> |

## List of Figures

|   |   |    |
|---|---|----|
| 1 | Task 2 Fig.1 Yield with different treatments by patch . . . . . | 5  |
| 2 | Task 5 Fig.1 The histogram of simulated values . . . . .        | 17 |
| 3 | Task 5 Fig.2 The histogram of p-values . . . . .                | 18 |
| 4 | Task 5 Fig.3 The QQ plot of p-value . . . . .                   | 19 |

## Introduction

My name is Erica. I am currently a third-year BSc student at University of Toronto. I am especially interested in the development of public facilities and services in urban areas, and I am currently focusing on the study of data analysis. This portfolio demonstrates my coding, critical thinking, statistical thinking, and communication skills. It shows my comprehensive application of statistical knowledge to problem solving.

In this portfolio, I will introduce some statistical skills via various samples and present my understanding of the statistical methods and inferences. The portfolio will include techniques of data cleaning, summaries, visualizations, and interpretations. The portfolio will also contain the construction of mixed linear model and detailed comparison between the models. In Task 3, I will explain confidence interval and p-value, and give some instructions. In the writing sample, I will go over the article *Common misconceptions about data analysis and statistics* (Motulsky, 2014), which focuses on the misunderstanding of some statistical concepts and some ethical issues.

The portfolio will be divided into 4 sections including Statistical skill sample, Writing sample, Reflection and this Introduction. All the codes and comments in section Statistical skill sample will be displayed and is open for any reproduction.

## Statistical skills sample

### Task 1: Setting up libraries and seed value

```
library(tidyverse)
library(lme4)
last3digplus <- 100 + 678
```

### Task 2a: Return to Statdew Valley: exploring sources of variance in a balanced experimental design (teaching and learning world)

#### Growing your (grandmother's) strawberry patch

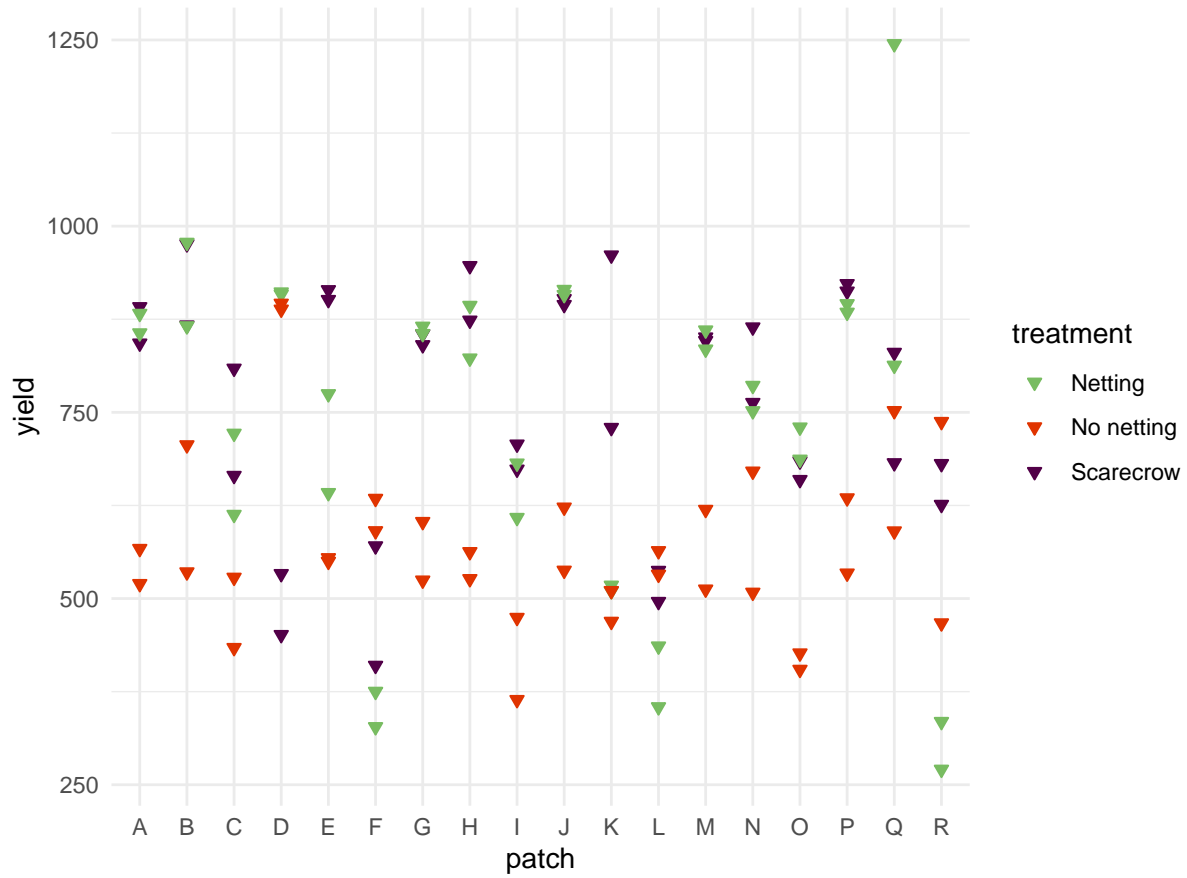
```
# Sourcing it makes a function available
source("grow_my_strawberries.R")
```

```
#create data
my_patch <- grow_my_strawberries(seed = last3digplus)

#alter the data
my_patch$treatment <- as.factor(my_patch$treatment)
```

#### Plotting the strawberry patch

```
# visualize the data
my_patch %>%
  ggplot(aes(x = patch, y = yield, fill = treatment, color = treatment)) +
  geom_point(pch = 25) +
  theme_minimal() +
# set the colors
  scale_fill_manual(values=c("#78BC61", "#E03400", "#520048")) +
  scale_color_manual(values=c("#78BC61", "#E03400", "#520048")) +
  labs(caption = "Created by Erica Zhou in STA303/1002, Winter 2022")
```



**Figure 1:** Task 2 Fig.1 Yield with different treatments by patch

### Demonstrating calculation of sources of variance in a least-squares modelling context

#### Model formula

$$y = \beta * Treatment + b_1 * Patch + b_2 * Patch * Treatment + \epsilon$$

$$b_1 \sim N(0, \phi_\theta), \quad b_2 \sim N(0, \omega_\theta), \quad \epsilon \sim N(0, \Lambda_\theta)$$

where:

- $\beta$  is the fixed effect of treatment on the mean of the outcome,
- $b_1$  is the random effect of patch and assumed to be normally distributed with mean 0,
- $b_2$  is the random effect of the interaction of treatment and patch and assumed to be normally distributed with mean 0, and

- $\epsilon$  is the residual and assumed to be normally distributed with mean 0.

```
# set up
n = 108
n_treatment = 3
n_patch = 18

# create aggregate data
agg_patch <- my_patch %>% group_by(patch) %>% summarise(yield_avg_patch = mean(yield))
agg_int <- my_patch %>% group_by(patch, treatment) %>% summarise(yield_avg_int =
  ↪ mean(yield), .groups = "drop")

# create the models
#an interaction model including main effects
int_mod <- lm(yield ~ patch*treatment, data = my_patch)

# an intercept only model
patch_mod <- lm(yield_avg_patch ~ 1, data = agg_patch)

# a main effects model
agg_mod <- lm(yield_avg_int ~ patch + treatment, data = agg_int)

# calculate the variances

# the residual variance unexplained by the interaction
var_int <- summary(int_mod)$sigma^2

# the portion of variability explained by the interaction of treatment and patch
var_ab <- summary(agg_mod)$sigma^2 - var_int/(n/(n_treatment*n_patch))

# the patch-to-patch variance
var_patch <- summary(patch_mod)$sigma^2 - summary(agg_mod)$sigma^2/n_treatment
```

```
# Source table
total <- var_ab + var_patch + var_int
tibble(`Source of variation` = c("Patch:Treatment",
                                "Patch",
                                "Residual"),
       Variance = c(var_ab, var_patch, var_int),
       Proportion = c(round(var_ab/total, 2),
                       round(var_patch/total, 2),
                       round(var_int/total, 2)) ) %>%
  knitr::kable(caption = "Sources of variation in yield not explained by treatment")
```

**Table 1:** Sources of variation in yield not explained by treatment

| Source of variation | Variance  | Proportion |
|---------------------|-----------|------------|
| Patch:Treatment     | 16645.246 | 0.57       |
| Patch               | 6742.934  | 0.23       |
| Residual            | 5941.151  | 0.20       |

**Task 2b: Applying linear mixed models for the strawberry data (practical world)**

```
# mod0 with only treatment
mod0 <- lm(yield ~ treatment, data = my_patch)

# mod1 with treatment and patch, where patch is random and treatment is fixed
mod1 <- lmer(yield ~ treatment + (1|patch), data = my_patch)

# mod2 with treatment, patch, and the interaction of treatment and patch
mod2 <- lmer(yield ~ treatment + (1|patch) + (1|patch:treatment), data = my_patch)

# Likelihood ratio test
lmtest::lrtest(mod0, mod1)
```

```
## Warning in modelUpdate(objects[[i - 1]], objects[[i]]): original model was of
## class "lm", updated model is of class "lmerMod"
```

```
## Likelihood ratio test
##
## Model 1: yield ~ treatment
## Model 2: yield ~ treatment + (1 | patch)
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    4 -705.94
## 2    5 -683.33  1 45.219  1.762e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
lmtest::lrtest(mod1, mod2)
```

```
## Likelihood ratio test
##
## Model 1: yield ~ treatment + (1 | patch)
## Model 2: yield ~ treatment + (1 | patch) + (1 | patch:treatment)
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1    5 -683.33
## 2    6 -664.73  1 37.21  1.061e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



```
lmtest::lrtest(mod0, mod2)
```

```
## Warning in modelUpdate(objects[[i - 1]], objects[[i]]): original model was of
## class "lm", updated model is of class "lmerMod"
```

```
## Likelihood ratio test
```

```
##
```

```
## Model 1: yield ~ treatment
```

```
## Model 2: yield ~ treatment + (1 | patch) + (1 | patch:treatment)
```

```
##   #Df  LogLik Df  Chisq Pr(>Chisq)
```

```
## 1    4 -705.94
```

```
## 2    6 -664.73  2 82.429  < 2.2e-16 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In the model construction, we used REML instead of ML because the ML variance estimator tends to be biased due to the use of an unknown mean estimator. Our goal here is to estimate the model parameters. REML will be more appropriate in this case as it takes into account the degrees of freedom lost when estimating the fixed effects, and it computes unbiased estimate for the variance.

### Justification and interpretation

Recall:

$$mod0 : yield \sim treatment$$

$$mod1 : yield \sim treatment + (1|patch)$$

$$mod2 : yield \sim treatment + (1|patch) + (1|patch : treatment)$$

According to the results of the Likelihood ratio test between mod0 and mod1, we found a small p-value. It means that we have a strong evidence to reject the hypothesis that the nested model(mod0) is as good as the full model(mod1). Therefore, we conclude that mod1 is better than mod0. Similarly, we also found a small p-value in the Likelihood ratio test between mod1 and mod2. Therefore, we have strong evidence to reject the hypothesis that the nested model(mod1) is as good as the full model(mod2), and we conclude that mod2 is better than mod1. We can also check again by applying the Likelihood ratio test on mod0 and mod2. Based on all the test results, mod2 is the best model.

```
summary(mod2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: yield ~ treatment + (1 | patch) + (1 | patch:treatment)
## Data: my_patch
##
## REML criterion at convergence: 1329.5
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.3444 -0.4127  0.0697  0.3808  3.2563
##
## Random effects:
## Groups           Name          Variance Std.Dev.
## patch:treatment (Intercept) 16645     129.02
## patch           (Intercept)  6743      82.12
## Residual                        5941      77.08
## Number of obs: 108, groups: patch:treatment, 54; patch, 18
##
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)      731.01      38.27  19.103
## treatmentNo netting -160.08      46.69  -3.429
## treatmentScarecrow   34.86      46.69   0.747
##
## Correlation of Fixed Effects:
##              (Intr) trtmNn
## trtmntNnttn -0.610
## trtmntScrcr -0.610  0.500
```

In mod2 summary, we see that treatment “Netting” is a reference, so the coefficients of the other two treatments shows the estimate changes when the treatment is “No netting” and “Scarecrow”. Based on the coefficients and the t-values, we can conclude that the yield will decrease if there is no netting comparing to the one with netting. It means that netting is helpful for the yield. When there is a scarecrow, the yield will be higher than the one with netting. It represents that a scarecrow is more efficient and useful than netting. However, the result for scarecrow is not statistically significant (assume  $\alpha = .05$ ). Among the random effect, about 23% of the

variance in yield can be explained by the difference in patches. About 57% of the variance can be explained by the interaction of patch and treatment. It means that the treatments have different levels of effect on different patches. About 20% of the variance cannot be explained by the given predictors.

### Task 3a: Building a confidence interval interpreter

```
#Construct the confidence interval interpreter

interpret_ci <- function(lower, upper, ci_level, stat){
  if(!is.character(stat)) {
    # produce a warning if the statement of the parameter isn't a character string
    # the spacing is a little weird looking so that it prints nicely in your pdf
    warning("
    Warning:
    stat should be a character string that describes the statistics of
    interest.")
  } else if(!is.numeric(lower)) {
    # produce a warning if lower isn't numeric
    warning("Warning: The lower bound of a confidence interval should be a number.")
  } else if(!is.numeric(upper)) {
    # produce a warning if upper isn't numeric
  } else if(!is.numeric(ci_level) | ci_level < 0 | ci_level > 100) {
    # produce a warning if ci_level isn't appropriate
    warning("Warning: The confidence level is not appropriate. The confidence level
    ↪ should be between 0 and 100 exclusively.")
  } else{
    # print interpretation
    # this is the main skill I want to see, writing a good CI interpretation.
    str_c("You can be ", ci_level,
          "% confident that the interval (", lower,",", upper,") includes the true ",
          ↪ stat, ".")
  }
}

# Test 1
ci_test1 <- interpret_ci(10, 20, 99, "mean number of shoes owned by students")

# Test 2
ci_test2 <- interpret_ci(10, 20, -1, "mean number of shoes owned by students")
```

```
# Test 3
ci_test3 <- interpret_ci(10, 20, -1, tibble(stat = 3))
```

**CI function test 1:** You can be 99% confident that the interval (10,20) includes the true mean number of shoes owned by students.

**CI function test 2:** Warning: The confidence level is not appropriate. The confidence level should be between 0 and 100 exclusively.

**CI function test 3:** Warning: stat should be a character string that describes the statistics of interest.

### Task 3b: Building a p value interpreter

```
# Construct the p-value interpreter
interpret_pval <- function(pval, nullhyp){
  if(!is.character(nullhyp)) {
    warning("
      Warning: stat should be a character string that describes the statistics
↪ of
      interest.")

  } else if(!is.numeric(pval)) {
    warning("The p value should be a number.")
  } else if(pval > 1) {
    warning("
      Warning: The P-value as a probability should not exceed 1. ")
  } else if(pval < 0){
    warning("
      Warning: The P-value as a probability should be at least 0. ")
  } else if (pval < 0.001){
    str_c("The p value is <.001, thus, there is very strong evidence against the null
↪ hypothesis that ", nullhyp, ".")
  } else if (pval > 0.1){
    str_c("The p value is ", round(pval, 3), ", thus, there is no evidence against the
↪ null hypothesis that ", nullhyp, ".")
  } else if(pval > 0.05){
```

```
    str_c("The p value is ", round(pval, 3),
          ", thus, there is weak evidence against the null hypothesis that ",
          ↪ nullhyp, ".")
  } else if(pval > 0.01){
    str_c("The p value is ", round(pval, 3),
          ", thus, there is moderate evidence against the null hypothesis that
          ↪ ", nullhyp, ".")
  } else if(pval > 0.001){
    str_c("The p value is ", round(pval, 3),
          ", thus, there is strong evidence against the null hypothesis that ",
          ↪ nullhyp, ".")

  }
}

pval_test1 <- interpret_pval(0.000000003,
                             "the mean grade for statistics students is the same as
                             ↪ for non-stats students")

pval_test2 <- interpret_pval(0.0499999,
                             "the mean grade for statistics students is the same as
                             ↪ for non-stats students")

pval_test3 <- interpret_pval(0.050001,
                             "the mean grade for statistics students is the same as
                             ↪ for non-stats students")

pval_test4 <- interpret_pval("0.05", 7)
```

**p value function test 1:** The p value is  $<.001$ , thus, there is very strong evidence against the null hypothesis that the mean grade for statistics students is the same as for non-stats students.

**p value function test 2:** The p value is 0.05, thus, there is moderate evidence against the null hypothesis that the mean grade for statistics students is the same as for non-stats students.

**p value function test 3:** The p value is 0.05, thus, there is weak evidence against the null hypothesis that the mean grade for statistics students is the same as for non-stats students.

**p value function test 4:** Warning: stat should be a character string that describes the statistics of interest.

### Task 3c: User instructions and disclaimer

#### Instructions

The confidence interval interpreter helps interpret the frequentist confidence interval with an easy and understandable language. Basically, the confidence interval is a range of plausible values for the population parameters. A population parameter is a **number** gives the information on the entire population rather than the sample. There are several ways to express the interpretation of a confidence interval. With an  $x(\text{ci\_level})\%$  confidence level, we can interpret that we are  $x\%$  **confident** that the true population parameter is included in this interval. Alternatively, we can say that there is  $x\%$  **confidence** that this interval captures the population parameter. However, we should be careful about the wording. It is **incorrect** to say that there is an  $x\%$  chance/probability that this interval captures the population parameter. By the definition of confidence interval, we know that the upper and lower bounds are both constants. As we have mentioned before, our interested population parameter is also a constant. Therefore, the chance/probability of

$$\text{lower bound}(\text{constant}) < \text{population parameter}(\text{constant}) < \text{upper bound}(\text{constant})$$

is either 0 or 1. Base on these facts, we should be very careful when interpreting the confidence interval.

The p-value interpreter helps interpret the p-value with an easy and understandable language. The p-value is the probability of observing an outcome at least as extreme as the outcome actually observed under the null hypothesis. A null hypothesis is the case that the observed difference is only caused by chance. In other words, a null hypothesis is a hypothesis that we want to reject to prove that what we have observed is not only caused by chance. The p-value shows the strength of evidence against the null hypothesis. This p-value interpreter gives some thresholds strictly on the p-value. Nevertheless, these thresholds are only for convenience, and we need not follow the guideline strictly. For example, when  $p\text{-value} = 0.05$ , we may either say it has a weak or moderate evidence. According to the thresholds,  $p\text{-value } 0.050001$  and  $p\text{-value } 0.049998$  give very different levels of evidence, but there is no big difference in practice. Therefore, we may refer to the results of the interpreter but we should not completely rely on them. Another reason that we should not completely rely on the p-value and the thresholds is that the significance level 0.05 (which is mostly used in hypothesis tests) was selected arbitrarily by Ronald A. Fisher(1890-1962), and arbitrary will leads to problems like P-Hacking and Replication Crisis.

## Disclaimer

As mentioned, the p-value interpreter is not completely reliable because of the fuzzy thresholds and the ambiguity of the significance level. People may refer to the interpreter but should not completely rely on it. People should also avoid P-Hacking and Replication Crisis in the analysis.

## Task 4: Creating a reproducible example (reprex)

A reproducible example (reprex) is an example that would allow other people (colleagues, group mates, etc.) to understand the contents and the codes. For example, a reprex allows others to reproduce the situation and to help solve the problems on their devices when we cannot solve the problems by ourselves. Below is a reprex.

```
my_data <- tibble(group = rep(1:10, each=10),
                  value = c(16, 18, 19, 15, 15, 23, 16, 8, 18, 18, 16, 17, 17,
                            16, 37, 23, 22, 13, 8, 35, 20, 19, 21, 18, 18, 18,
                            17, 14, 18, 22, 15, 27, 20, 15, 12, 18, 15, 24, 18,
                            21, 28, 22, 15, 18, 21, 18, 24, 21, 12, 20, 15, 21,
                            33, 15, 15, 22, 23, 27, 20, 23, 14, 20, 21, 19, 20,
                            18, 16, 8, 7, 23, 24, 30, 19, 21, 25, 15, 22, 12,
                            18, 18, 24, 23, 32, 22, 11, 24, 11, 23, 22, 26, 5,
                            16, 23, 26, 20, 25, 34, 27, 22, 28))
#> Error in tibble(group = rep(1:10, each = 10), value = c(16, 18, 19, 15, : could not find

my_summary <- my_data %>%
  summarize(group_by = group, mean_val = mean(value))
#> Error in my_data %>% summarize(group_by = group, mean_val = mean(value)): could not find

glimpse(my_summary)
#> Error in glimpse(my_summary): could not find function "glimpse"
```

## Task 5: Simulating p-values

### Setting up simulated data

```
set.seed(last3digplus)
# set simulation dataset

sim1 <- tibble(group = rep(1:1000, each = 100),
               val = rnorm(100000, 0, 1))

sim2 <- tibble(group = rep(1:1000, each = 100),
               val = rnorm(100000, 0.2, 1))

sim3 <- tibble(group = rep(1:1000, each = 100),
               val = rnorm(100000, 1, 1))

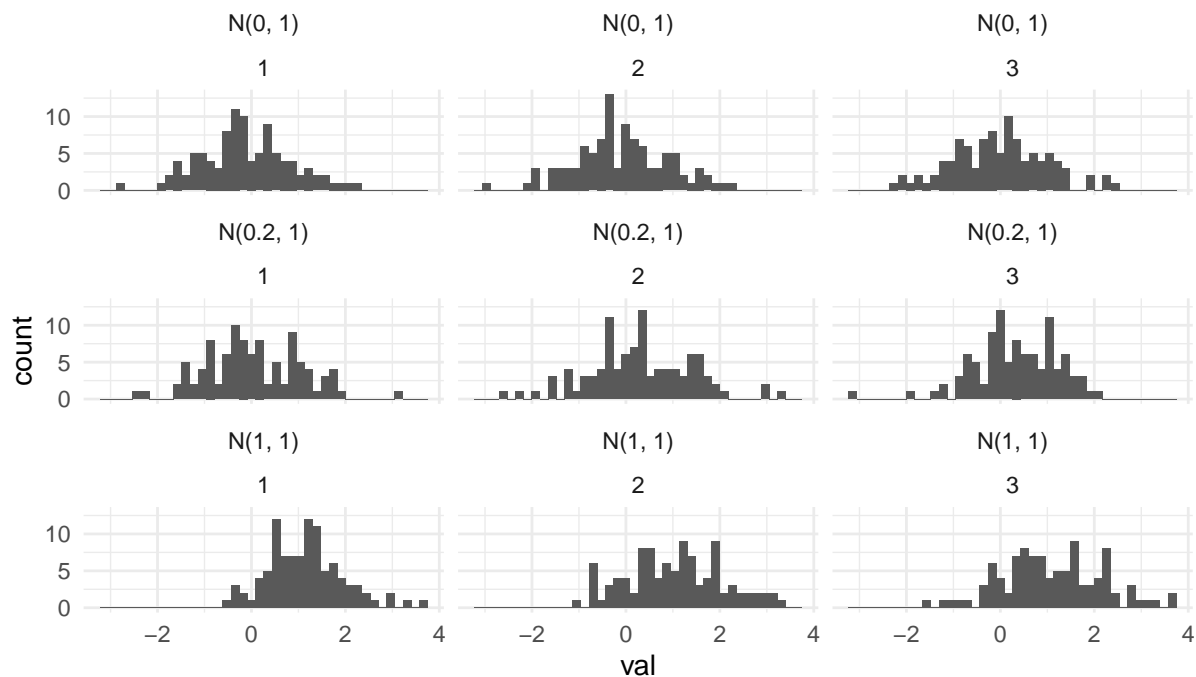
# combine the tibbles
all_sim <- bind_rows(sim1, sim2, sim3, .id = "sim")

# simulation description
sim_description <- tibble(sim = 1:3,
                          desc = c("N(0, 1)",
                                    "N(0.2, 1)",
                                    "N(1, 1)"))

# merge the dataset
all_sim <- merge(all_sim, sim_description, by = "sim")

# visualize the simulation
all_sim %>%
  filter(group <= 3) %>%
  ggplot(aes(x = val)) +
  geom_histogram(bins = 40) +
  facet_wrap(desc~group, nrow = 3) +
  theme_minimal() +
  labs(caption = "Created by Erica Zhou in STA303/1002, Winter 2022")
```





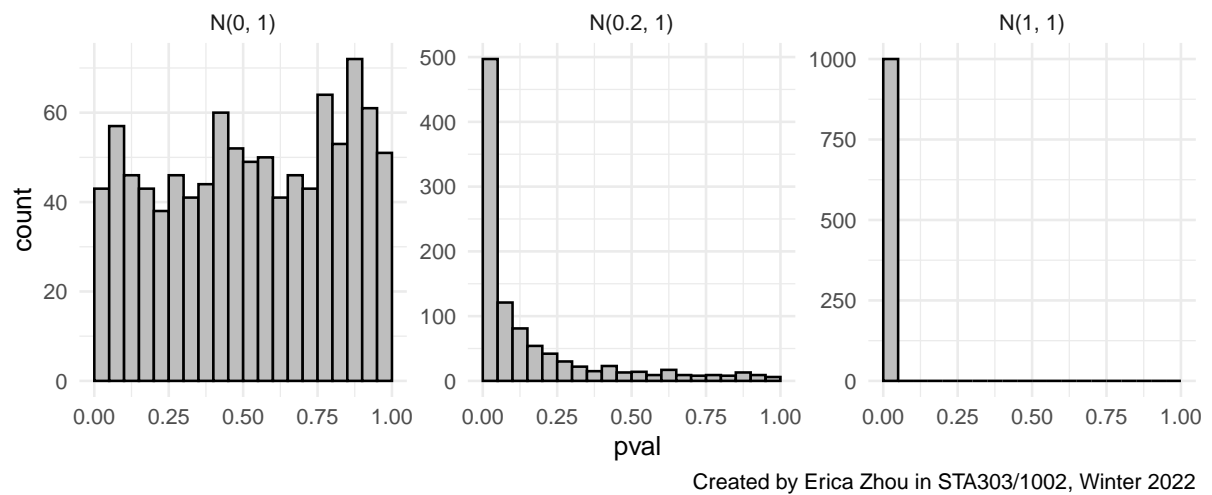
Created by Erica Zhou in STA303/1002, Winter 2022

**Figure 2:** Task 5 Fig.1 The histogram of simulated values

### Calculating p-values

```
# calculate p-value
pvals <- all_sim %>% group_by(desc, group) %>% summarise(pval = t.test(val, mu =
  ↪ 0)$p.value, .groups = "drop" )
```

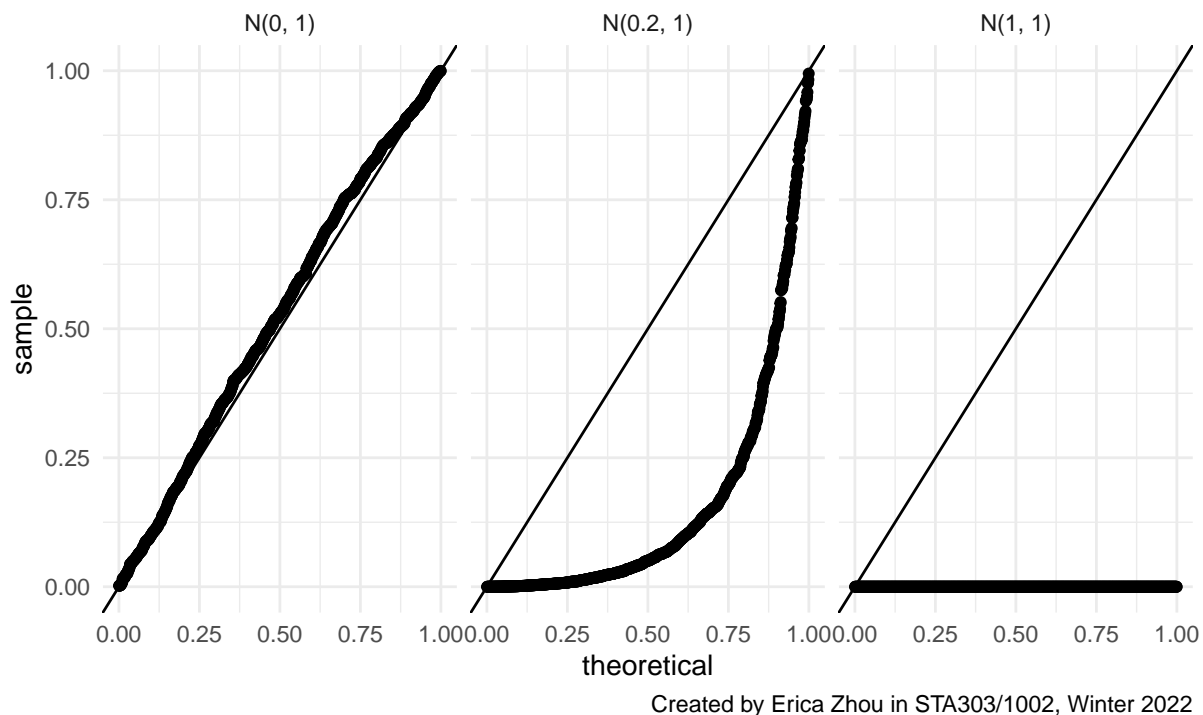
```
# visualize the distribution of the p-values
pvals %>% ggplot(aes(x = pval)) +
  geom_histogram(boundary = 0, binwidth = 0.05, fill = "grey", color = "black") +
  xlim(0, 1) +
  facet_wrap(desc~. , scales = "free_y") +
  theme_minimal() +
  labs(caption = "Created by Erica Zhou in STA303/1002, Winter 2022")
```



**Figure 3:** Task 5 Fig.2 The histogram of p-values

### Drawing Q-Q plots

```
pvals %>%
  ggplot(aes(sample = pval)) +
  geom_qq(distribution = qunif) +
  geom_abline(intercept = 0, slope = 1) +
  facet_wrap(~desc) +
  theme_minimal() +
  labs(caption = "Created by Erica Zhou in STA303/1002, Winter 2022")
```



**Figure 4:** Task 5 Fig.3 The Q-Q plot of p-value

All coding is done with R version 4.0.5 (2021-03-31).

### Conclusion and summary

Recall that the p-value is the probability of observing an outcome at least as extreme as the outcome actually observed under the null hypothesis. In Fig.1, we see that the data are sampled from normal distribution with different mean values. In Fig.2, we see that approximately 10% of the p-values are between 0.9 and 1. We also recognize that the distribution of the p-values simulated from a sample of a standard normal distribution is a uniform distribution. The Q-Q plot (Fig.3) also confirms this result. Therefore, the p-value does not follow a normal distribution as the data. Also, as we see in Fig.2 & 3, we notice that the p-values deviate away from a uniform distribution. This makes sense intuitively, because the data simulated from  $N(0.2, 1)$  and  $N(1, 1)$  already “reject” the null hypothesis  $\mu=0$ . As the population parameter  $\mu$  deviates away from 0, the evidence against the null hypothesis gets stronger. This explains the right skewness in the second and third histograms in Fig.2 and the non-uniform distribution of the p-values in the second and third Q-Q plots in Fig.3.

We see on Fig.2 that when the data has an  $N(0,1)$  distribution, there is still a small chance

that the p-value indicates some (weak) evidence against the null hypothesis  $\mu=0$  ( $p < \alpha = 0.1$ ), though the population parameter as we have set equals 0. In this case, there exists a type I error, which indicates some mistaken rejection of a true null hypothesis. Based on this rule, as  $\alpha$  gets smaller, the probability of the type I error will be reduced, and the evidence against the null becomes stronger. This is the reason why we want p-value to be small. However, a small p-value does not represent the importance of the observed effect, that is, a small p-value can be observed for some meaningless or unimportant effects. There are also other problems of the p-value such as P-Hacking. In conclusion, p-value can represent the strength of evidence against the null hypothesis and help people determine the significance of the effects that people are interested in but people ought to interpret p-value very carefully and avoid the misuse of p-value.

## Writing sample

According to the introductions we have made in previous sections, we understand that there exists misuse of the concept of statistical inference especially the p-value. However, the misunderstanding on p-value is now a common problem in science research and published findings. The article *Common misconceptions about data analysis and statistics* (Motulsky, 2014) explains the common misconceptions such as the misunderstanding on p-value and standard error and gives useful suggestions to analysts on how to avoid the related statistical and ethical problems as possible. As Statistics students, we always make statistical inferences. Like the scientists, we may have also misused the statistical concepts. It is important for us to understand the correct concept and to analyze the data appropriately and rigorously. From this article, we can find some suggestions.

In the introduction, Motulsky(2014) points out that many published findings are found not reproducible because many scientists misunderstand the statistical concepts especially p-value and standard error. Then he explains the underlying theories of p-value and standard error. Under the section of Misconception 1, Motulsky(2014) explains HARKing, which is one type of P-Hacking. HARKing is the abbreviation of “hypothesis after the results are known”, and it is a common misuse among Statistics students. In assignments and projects, due to lack of time and knowledge, we usually make HARKing during the analysis. We always want our results at least to seem “correct” so that we can avoid any further problems. HARKing is a problem especially for undergraduate students, who may not be proficient at data analysis. Fortunately, Motulsky(2014) has given his suggestions that we can make statements about the sample size and the use of each step for clarification. Also, we may label our conclusion as “preliminary” if we have ever used P-Hacking in the process.

Motulsky(2014) has given “a general solution” to the misuse of statistical inference. Under the section Misconception 5, he gives detailed suggestions on how to report the results honestly in detail. The reports should include the methods, the threshold of p-value, the interpretation of outliers, and so forth. Writing the reports honestly in detail can avoid the misuse of statistical methods and make the work reproducible.

In conclusion, although we as an undergraduate student does not always publish our works, we should pay attention to the use of statistical tools and interpretation. Any result needs to be fully understood before interpreting. Any misuse of the methods in the future may cause unreliability and distrust of science. We need to improve our understanding of statistical concepts and give details of our analysis honestly. The data is not always perfect, and our job is to reveal the truth including the imperfection.

**Word count:** 440 words

## References

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Motulsky, H. J. (2014). Common misconceptions about data analysis and statistics. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 387(11), 1017–1023.

## Reflection

### **What is something specific that I am proud of in this portfolio?**

In this portfolio, I have improved my understanding of a linear mixed model and the interaction effect, which I felt confused before. I have also understood the difference between an interaction model, a main effect model and an intercept only model, and how to use them to calculate the variance from different sources. I have also dug into the concept of p-value from the thresholds to the underlying theories. From the reading, I have found helpful suggestions to reduce the effect of P-Hacking, especially HARKing, which I used to do. With the improvement, I can make more accurate and reliable statistical inferences in the future.

### **How might I apply what I've learned and demonstrated in this portfolio in future work and study, after STA303/1002?**

In future projects and assignments, I can use linear mixed model to study the fixed and random effects. With linear mixed model, I can estimate the purer treatment effect that I am interested in. On the other hand, the confidence interval interpreter and the p-value interpreter can be used for analysis as tools. The instructions and disclaimer in Task 3 and the writing sample will help me interpreting the statistical inferences as a reference. Finally, with the help of reprex, I can ask coding questions in a more reproducible way. The improvement in understanding and coding skills will make me more efficient.

### **What is something I'd do differently next time?**

Next time, I will dig deeper into the interpretation of the linear model summary and write more description about the methods I used. I may use “ugly” data and to get familiar with the interpretation of the models, the treatment effect, the confidence intervals, the p-values and so forth. As mentioned in the writing sample, I will also follow the suggestions from Motulsky to label the graphs in detail, to clarify the P-Hacking problems, and to give more details of the analysis. Finally, I will write more about research backgrounds and data collection in future projects.