
STA303/1002 Portfolio

An exploration of linear mixed models and common misconceptions in statistics

Shiqi Li

2022-02-17

Contents

Introduction	3
Statistical skills sample	4
Task 1: Setting up libraries and seed value	4
Task 2a: Return to Statdew Valley: exploring sources of variance in a balanced experimental design (teaching and learning world)	4
Task 2b: Applying linear mixed models for the strawberry data (practical world) . . .	7
Task 3a: Building a confidence interval interpreter	9
Task 3b: Building a p value interpreter	11
Task 3c: User instructions and disclaimer	12
Task 4: Creating a reproducible example (reprex)	13
Task 5: Simulating p-values	15
Writing sample	19
References	20
Reflection	21

List of Figures

1	Strawberry Patch Yields Figure	5
2	Histogram of Each Simulated Data Set on First Three Groups	16
3	P-Value Histograms in Three Simulated Datasets	17
4	Q-Q Plots for p-values of Simulated Datasets	18

Introduction

This portfolio is written to exercise and demonstrate the skills I learned from the course STA303 which includes both statistical skills and presentation skills.

The first section demonstrates basic understanding of important statistical concepts and proficiency in linear mixed model. In the first subpart, I will present how to set up libraries for a data analysis project. The second subpart is where I fit linear mixed models using a simulated data set with both least squares method and a R package to discuss, interpret, and compare the results. The third part supplies two R functions to interpret confidence intervals and p-values in plain languages to demonstrate my understanding of these two concepts and error handling in R. The fourth part is simply a reproducible example displayed as a code chunk to illustrate my understanding of reprec. The final part is another simulation study that compares the distribution of p-values of a one sample t-test from different population distributions.

The second section presents my professional writing skills by paraphrasing an article and summarize takeaways as suggestions for the future myself regarding statistical analysis in scientific research. Specifically, I discussed how I would avoid common misconceptions such as p-hacking, p-values interpretations, and lack of reproducibility and exercise ethical practices.

At the end, I also presented how I view this portfolio and how can I improve upon the skills demonstrated. I also discuss what are the possible next steps for future portfolio writing.

Statistical skills sample

Task 1: Setting up libraries and seed value

```
library(tidyverse) # load tidyverse library
last3digplus = 512 # student ID
```

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

```
source("~/Desktop/rstudio-export/grow_my_strawberries.R")
my_patch = grow_my_strawberries(seed = last3digplus) # run the source the function
# change the order of treatment column
my_patch = my_patch %>%
  arrange(factor(treatment, levels = c("No netting", "Netting", "Scarecrow")))
```

Plotting the strawberry patch

```
# ggplot to produce strawberry yield versus patch
ggplot(my_patch, aes(x = patch, y = yield)) +
  geom_point(aes(colour = factor(treatment),
                    fill=factor(treatment)), pch = 25) + # set triangle shape
  scale_colour_manual(values = c("#E03400", "#78BC61", "#520048")) +
  scale_fill_manual(values = c("#E03400", "#78BC61", "#520048")) + # set colors
  theme_minimal() +
  labs(
    caption = "Created by Shiqi Li in STA303/1002, Winter 2022" # set caption
  )
```

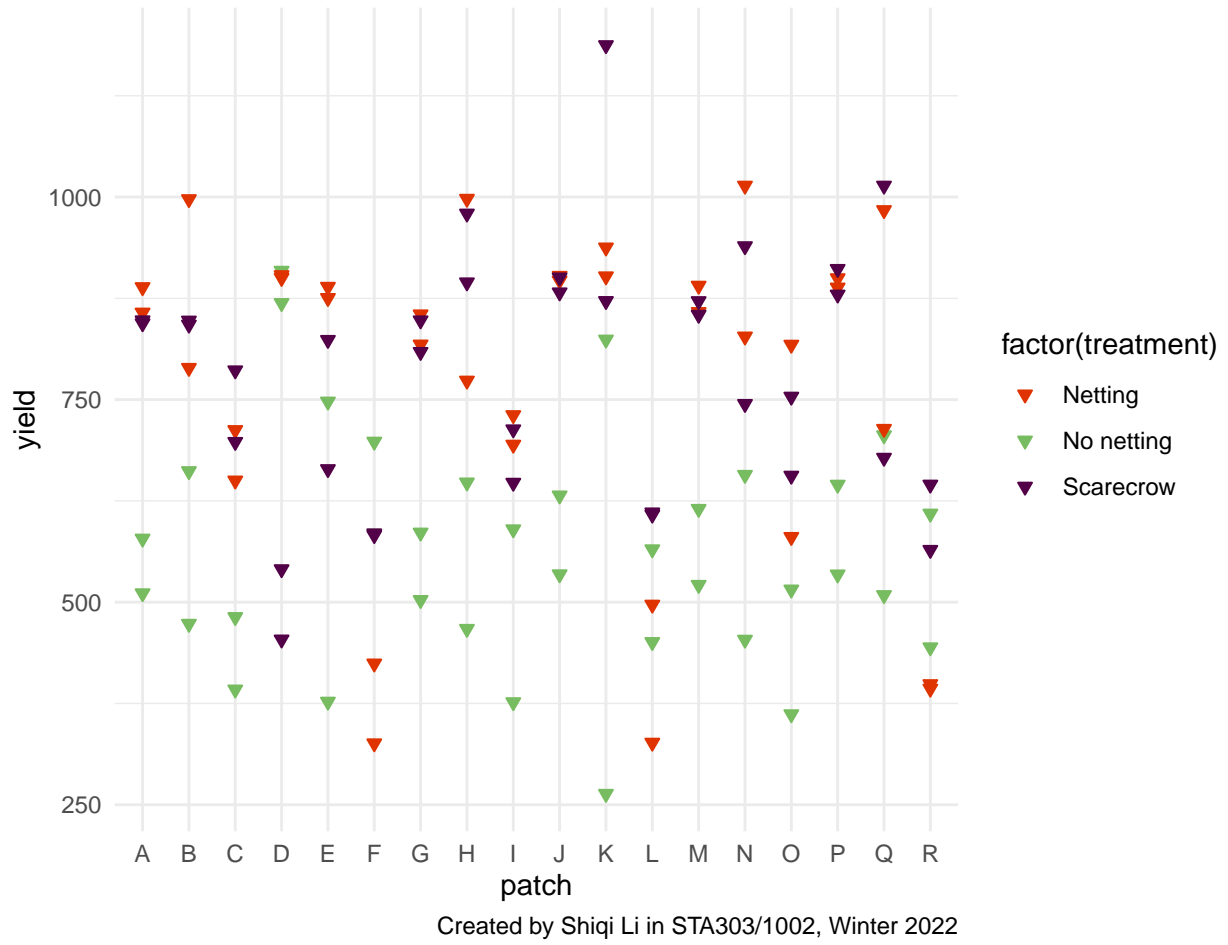


Figure 1: Strawberry Patch Yields Figure

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

Model formula

$$y_{ijk} = \mu + \alpha_i + b_j + (\alpha b)_{ij} + \epsilon_{ijk}$$

where:

- y_{ijk} : the amount of strawberry yield (in kg) under the i^{th} treatment for the j^{th} patch in the k^{th} harvest
- μ : the grand mean of strawberry yield
- α_i : the $I = 3$ fixed effects for treatments
- b_j : the random effects for patch j , which has a total of $J = 18$ patches, $b_k \sim N(0, \sigma_b^2)$
- $(\alpha b)_{ij}$: the $IJ = 54$ interaction terms for the interaction between the treatment and the patch, $ab_{ij} \sim N(0, \sigma_{ab}^2)$

- ϵ_{ijk} : random errors, $\epsilon_{ijk} \sim N(0, \sigma^2)$

```
agg_patch = my_patch %>%
  group_by(patch) %>%
  summarise(yield_avg_patch = mean(yield)) # group by patch to compute average
agg_int = my_patch %>%
  group_by(patch, treatment) %>% # group by interaction combination to compute average
  summarise(yield_avg_int = mean(yield), .groups="drop")

int_mod = lm(yield~patch*treatment, data=my_patch) # fit interaction model
patch_mod = lm(yield_avg_patch~1, data=agg_patch) # fit intercept only model with
  ↪ patch
agg_mod = lm(yield_avg_int~patch+treatment, data=agg_int) # fit the main effect models

# Calculation:
## for var_ab, K = n/I*J = 108/(3*18) = 2
## for var_patch, I = 18
var_int = summary(int_mod)$sigma^2 # residual variance
# variance in yield explained interaction after accounting for fixed effects and other
  ↪ sources
var_ab = summary(agg_mod)$sigma^2 - var_int/(108/(3*18))
var_patch = summary(patch_mod)$sigma^2 - (summary(agg_mod)$sigma^2)/18 #
  ↪ patch-to-patch

# create the ICC table
tibble(`Source of variation` = c("patch",
                                "treatment:patch",
                                "residual"),
       Variance = c(var_patch, var_ab, var_int),
       Proportion = c(round(var_patch / (var_ab + var_int + var_patch), 2),
                      round(var_ab / (var_ab + var_int + var_patch), 2),
                      round(var_int / (var_ab + var_int + var_patch), 2))) %>%
  knitr::kable(caption = "Intraclass Correlation Coefficient Table")
```

Table 1: Intraclass Correlation Coefficient Table

Source of variation	Variance	Proportion
patch	10343.802	0.32
treatment:patch	9114.408	0.28

Source of variation	Variance	Proportion
residual	12984.653	0.40

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

```
library(lme4)
library(lmtest)
mod0 = lm(yield~treatment, data=my_patch) # linear model with only treatment
mod1 = lmer(yield~treatment + (1 | patch), data=my_patch) # lmm with random effect
  ↪ patch
# lmm with random effect for patch plus interaction of patch and treatment
mod2 = lmer(yield~treatment + (1 | patch) + (1 | patch:treatment), data=my_patch)

lmtest::lrtest(mod0, mod1) # two likelihood ratio test to compare models
```

```
## 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 -704.05
## 2    5 -684.67  1 38.771  4.767e-10 ***
## ---
## 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 -684.67
## 2    6 -680.51  1 8.3134  0.003935 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In above we fit the three models, and conducted two likelihood ratio tests. First, we compare the fixed effects only model with the model that includes patch random effect. The p-value is -4.77×10^{-10} which is much lower than 0.05. This means we have evidence in the data suggesting that the linear mixed model with patch as a random effect fits the data better than the fixed effect only model. Likewise, the LR test to compare the two mixed model with and without interaction between patch and treatment also output a p-value of 0.0039, this means that including the interactions also improve the model fit as suggested by the data.

Here all three models have the same fixed effects, so we are not comparing nested models with the same random effects. In this situation, REML would work better, and the `lmer` function from the `lme4` package also uses REML by default which we can see when running `summary(mod1)`.

Justification and interpretation

From the above likelihood ratio test result, we see that the random effect model with interactions between patch and treatment (`mod2`) because it fits the data most well among all three models. Hence we will use this model for interpretation.

```
summary(mod2)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: yield ~ treatment + (1 | patch) + (1 | patch:treatment)
##    Data: my_patch
##
## REML criterion at convergence: 1361
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.7804 -0.5218  0.1169  0.3996  2.1392
##
## Random effects:
##   Groups             Name             Variance Std.Dev.
## patch:treatment (Intercept)    9114      95.47
```



```
## patch (Intercept) 6009 77.52
## Residual 12985 113.95
## Number of obs: 108, groups: patch:treatment, 54; patch, 18
##
## Fixed effects:
## Estimate Std. Error t value
## (Intercept) 772.58 34.65 22.295
## treatmentNo netting -208.82 41.64 -5.015
## treatmentScarecrow 4.66 41.64 0.112
##
## Correlation of Fixed Effects:
## (Intr) trtmNn
## trtmntNnttn -0.601
## trtmntScrcr -0.601 0.500
```

First, we consider the fixed effect estimates. First, the grand total of strawberry yields in all these 108 harvests are 772.58 kilograms. In terms of different treatment, we see that on average a patch with no netting will yield 208.82 less kilogram than a patch that has netting, and that a patch with scarecrow will yield 4.66 more kilogram than a patch that has netting in one harvest.

In terms of the variance partitioning for random effects, we see that there are 39.7121349 percent of variation left unexplained in the model, 33.2717641 percent of variation explained by the interaction random effect between treatment and patch after accounting for the fixed effects and other sources, and 27.0161009 percent of variation captured by the main random effect of patches. Note that if we square the above standard deviation, we would get exactly the same quantity of variance as we got in Task 2a.

Task 3a: Building a 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)) {
```

```

    warning("
    Warning:
    lower should be a numeric value that provides the lower confidence interval
↪   bound.")
  } else if(!is.numeric(upper)) {
    warning("
    Warning:
    upper should be a numeric value that provides the upper confidence interval
↪   bound.")
  } else if(!is.numeric(ci_level) | ci_level < 0 | ci_level > 100) {
    warning("Warning:
    ci_level should be a numeric value between 0 and 100 that specifies a level of
↪   confidence of
    the provided confidence interval.")
  } else{
    # print interpretation
    # this is the main skill I want to see, writing a good CI interpretation.
    str_c("We are ", ci_level, "% confident that the true value of the population ",
          stat, " lies between the value ", lower, " and ", upper, ".")
  }
}

# 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")

```

```

## Warning in interpret_ci(10, 20, -1, "mean number of shoes owned by students"): Warning:
##   ci_level should be a numeric value between 0 and 100 that specifies a level of confid
##   the provided confidence interval.

```

```

# Test 3
ci_test3 <- interpret_ci(10, 20, 95, 99)

```

```

## Warning in interpret_ci(10, 20, 95, 99):
##   Warning:
##   stat should be a character string that describes the statistics of
##   interest.

```

CI function test 1: We are 99% confident that the true value of the population mean number of shoes owned by students lies between the value 10 and 20.

CI function test 2: Warning: ci_level should be a numeric value between 0 and 100 that specifies a level of confidence of the provided confidence interval.

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

Task 3b: Building a p value interpreter

```
interpret_pval <- function(pval, nullhyp){
  if(!is.character(nullhyp)) {
    warning("
      Warning: nullhyp should be a character string that describes the null
      hypothesis of the test corresponds to the p-value pval.")
  } else if(!is.numeric(pval)) {
    warning("
      Warning: pval should be a numeric value.")
  } else if(pval > 1) {
    warning("
      Warning: pval should not be greater than 1.")
  } else if(pval < 0){
    warning("
      Warning: pval should not be less than 0.")
  } else if(pval >= 0.1){
    str_c("The p-value is ", round(pval, 3),
      ", so we have no evidence in the data against the claim that ", nullhyp, ".")
  } else if(pval >= 0.05 && pval < 0.1){
    str_c("The p-value is ", round(pval, 3),
      ", so we have week evidence in the data against the claim that ", nullhyp, ".")
  } else if(pval >= 0.01 && pval < 0.05){
    str_c("The p-value is ", round(pval, 3),
      ", so we have moderate evidence in the data against the claim that ", nullhyp,
      "↪ ".")
  } else if(pval >= 0.001 && pval < 0.01){
    str_c("The p-value is ", round(pval, 3),
      ", so we have strong evidence in the data against the claim that ", nullhyp, ".")
  } else if(pval < 0.001){
    str_c("The p-value is < .001, so we have very strong evidence in the data against
      ↪ the claim 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)

## Warning in interpret_pval("0.05", 7):
##           Warning: nullhyp should be a character string that describes the null
##           hypothesis of the test corresponds to the p-value pval.
```

p value function test 1: The p-value is $< .001$, so we have very strong evidence in the data against the claim 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, so we have moderate evidence in the data against the claim 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, so we have week evidence in the data against the claim that the mean grade for statistics students is the same as for non-stats students.

p value function test 4: Warning: nullhyp should be a character string that describes the null hypothesis of the test corresponds to the p-value pval.

Task 3c: User instructions and disclaimer

Instructions

The provided interpreter for confidence intervals and p-values can be used following the syntax `interpret_ci(lower, upper, ci_level, stat)` and `interpret_pval(pval, nullhyp)` correspondingly. The CI interpreter takes in the confidence level, the lower and upper bound, and the target statistic as argument. The argument `stat` is the population parameter that we are

estimating, which represents the true value of this quantity in the whole population. Note that the confidence level is not a probability measure, we cannot say that with this probability the confidence interval captures the true population parameter value.

The p-value interpreter interprets a well-worded null hypothesis in plain language according to the supplied p-value. The null hypothesis is a statement in which a statistical test is aiming to find evidence for or against based on the data. In order to use this interpreter, we should change the mathematical formulation of null hypothesis in to a statement about its targeting quantities based on the symbol of comparison. For example, for $H_0 : \mu_a = \mu_b$ we can translate to “the true mean of group a and group b are the same”.

Disclaimer

These two interpreters only output the plain language interpretation based on the argument supplied, and does not contain any calculations. The strength of evidence for different p-values are based on the most widely recognized standards but are not deterministic. Also, the underlying assumptions of the statistical test that the p-values come from are not considered in the function. Before using it, please make sure these are checked so the output interpretation can be valid and used for conclusions.

Task 4: Creating a reproducible example (reprex)

A reproducible example should be the simplest example that others would be able to reproduce right when they get it, that is to say, the example should be exactly the same and without any other dependencies. For presentation purpose, a good reprex should be completely reproducible (i.e. do not rely on other libraries not included in the example, and is completely self-contained) and explicitly show (by code and associated output) the problems in its body.

```
library(tidyverse)
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))
my_summary <- my_data %>%
```

```
summarize(group_by = group, mean_val = mean(value))

glimpse(my_summary)
#> Rows: 100
#> Columns: 2
#> $ group_by <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3...
#> $ mean_val <dbl> 19.67, 19.67, 19.67, 19.67, 19.67, 19.67, 19.67, 19.67, 19.67, 19.67...
```

Task 5: Simulating p-values

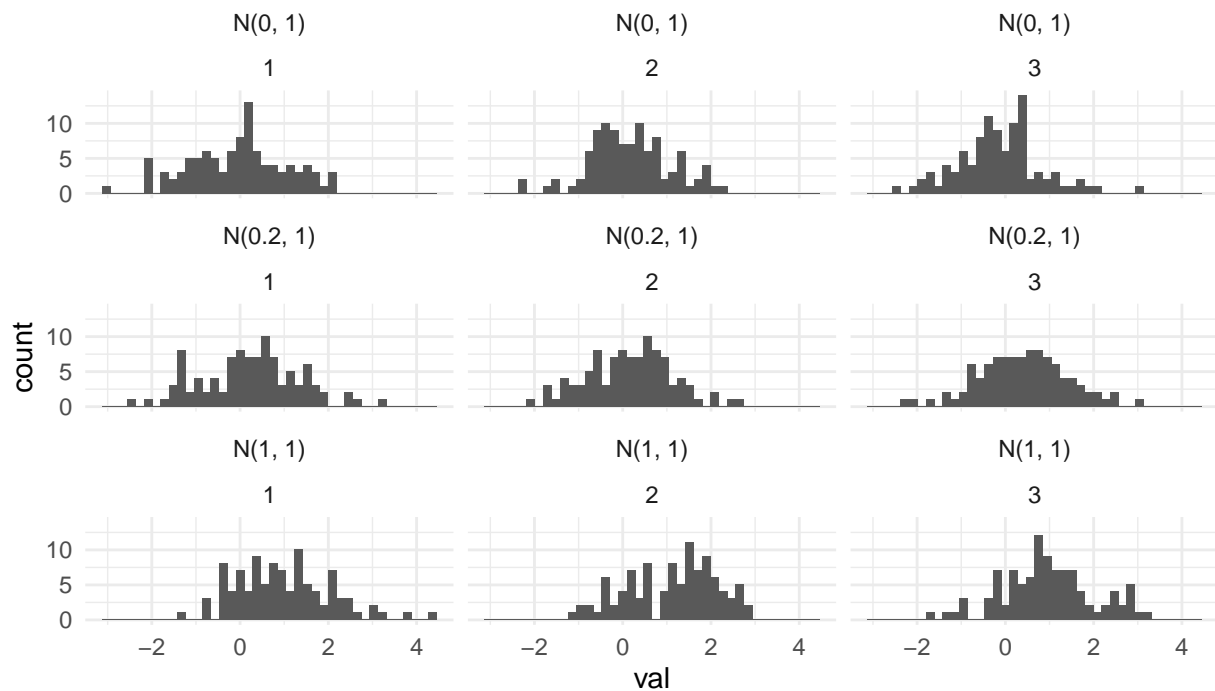
Setting up simulated data

```
set.seed(last3digplus) # set seed for reproducibility
sim1 = tibble(group=rep(1:1000, each=100),
              val=rnorm(100000, 0, 1)) # N(0, 1), 1000 samples with 100 observations
              ↪ each
sim2 = tibble(group=rep(1:1000, each=100),
              val=rnorm(100000, 0.2, 1)) # N(0.2, 1)
sim3 = tibble(group=rep(1:1000, each=100),
              val=rnorm(100000, 1, 1)) # N(1, 1)
all_sim = bind_rows(sim1, sim2, sim3, .id="sim") # stack all datasets
all_sim = all_sim %>%
  mutate(sim = as.numeric(sim)) # change sim type to numeric to match with
  ↪ sim_description

sim_description <- tibble(sim = 1:4,
                        desc = c("N(0, 1)",
                                "N(0.2, 1)",
                                "N(1, 1)",
                                "Pois(5)"))

all_sim = all_sim %>%
  inner_join(sim_description, by="sim") # join with sim_description

# visualize the data set using ggplot
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 Shiqi Li in STA303/1002, Winter 2022")
```



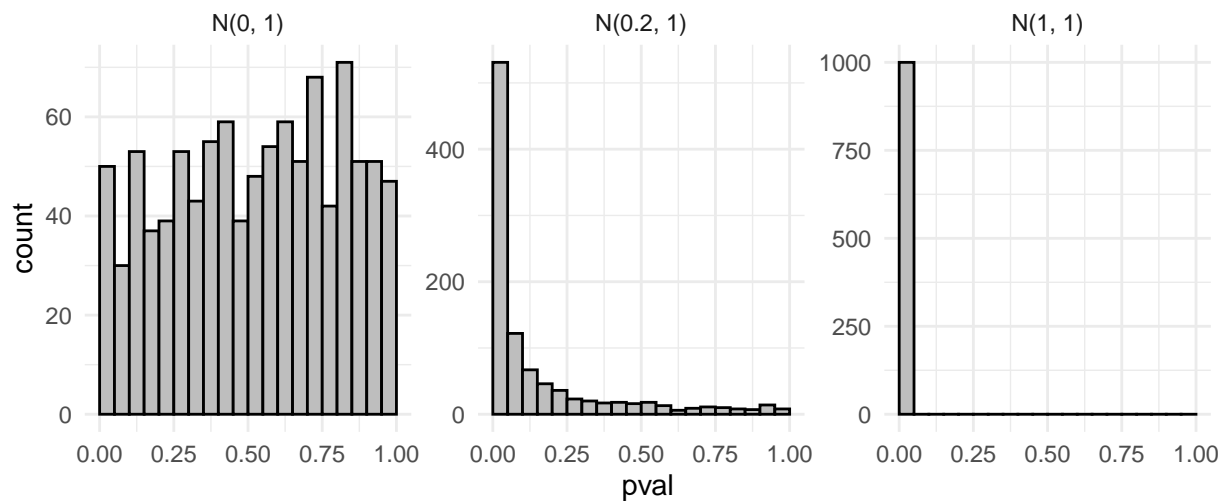
Created by Shiqi Li in STA303/1002, Winter 2022

Figure 2: Histogram of Each Simulated Data Set on First Three Groups

Calculating p values

```
# get p-value for each group/desc combination
pvals <- all_sim %>%
  group_by(desc, group) %>%
  summarize(pval = t.test(val, mu = 0)$p.value, .groups="drop")

# visualize p-values in all three datasets
ggplot(pvals, 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 Shiqi Li in STA303/1002, Winter 2022")
```

Created by Shiqi Li in STA303/1002, Winter 2022

Figure 3: P-Value Histograms in Three Simulated Datasets

Drawing Q-Q plots

```
# p-value QQ plot against a Unif(0, 1) distribution
pvals %>%
  ggplot(aes(sample = pval)) +
  geom_qq(distribution = stats::qunif) +
  geom_abline(intercept = 0, slope = 1) +
  facet_wrap(~desc) +
  theme_minimal() +
  labs(caption = "Created by Shiqi Li in STA303/1002, Winter 2022")
```

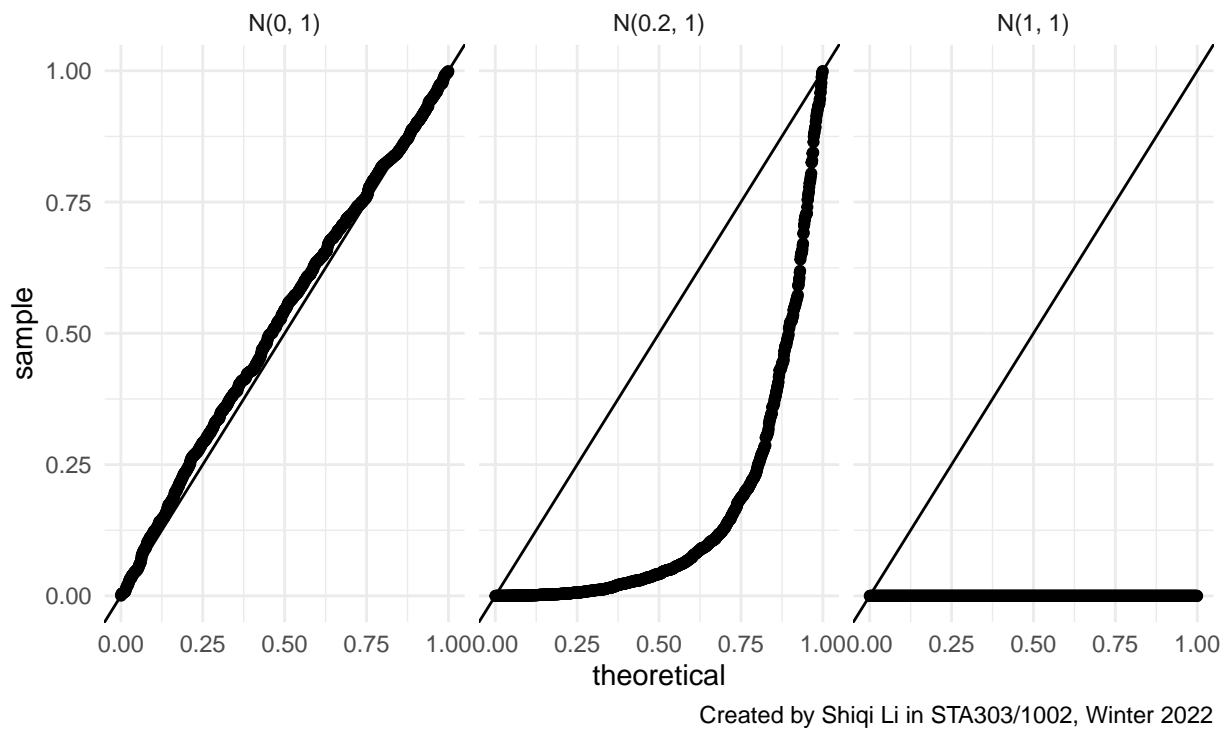


Figure 4: Q-Q Plots for p-values of Simulated Datasets

Conclusion and summary

The p-value is interpreted as the probability of getting a test statistic at least as extreme as what we have observed under null hypothesis. This makes sense in our simulations as with $N(0, 1)$ samples, the distribution of p-value should be uniform in $[0, 1]$ due to random variation of $sd = 1$, and with $N(1, 1)$ samples, we see the p-values are all close to zero.

This question corresponds to our $N(0, 1)$ samples, and as we know that $p \sim Unif(0, 1)$, the answer will be clearly D as the range $[0.9, 1]$ occupies 10% of the area bounded by the CDF of a $Unif(0, 1)$ random variable.

Writing sample

In my previous studies, the problem of p-hacking and reproducibility crisis was consistently discussed, and I was introduced to the harm and danger of this phenomenon. However, after reading this article, I was surprised there are still many other misconceptions of statistics that exist in scientific research. Even the act of p-hacking itself can have so many different forms. As a statistician, there is much more attention to pay when doing our jobs to be ethical.

Motulsky introduced four common misconceptions in data analysis in scientific research, and the very first one points undoubtedly to the problem of p-hacking. There are many different possibilities of p-hacking besides HARKing, and Motulsky also emphasizes the importance of reporting the intended sample size before the data collection process. Before I only consider sample size when trying to reduce variance of estimators, and plan to count on large sample asymptotic distribution. I did not realize that sample size plays such an important role in p-hacking as people would possibly stop at a point where desirable results are achieved. I think it is important that we should always exercise “planned ‘adaptive’ sample size methods” (Motulsky, 2014.)

Besides p-hacking, the correct interpretation and usage of p-values are also a crucial concern. The correct interpretation of p-value is the probability of getting a test statistic at least as extreme as what we observed. But according to Motulsky, linking small p-values with large effect sizes seem like a natural thinking for many researchers. I think it is important that we should always deem p-values and effect sizes as two separate things and avoid reporting p-values can also avoid letting readers misinterpret them.

There are also other common pitfalls in reporting such as overusing the word “statistically significant” and using standard error of the mean as a measure to quantifies variation in the data. As statistician, we use the word “significant” to describe the strength of evidence based on a p-value and a significance level, but this word can somehow be misleading to other people lack training. To avoid misconception, we should avoid using this word whenever possible, and report numerical values of test statistic and p-values directly. On the other hand, we know that the SE of the mean describes only the variation in an estimator, and it does not reveal information about the variability in the data. I think it’s good to distinguish the uncertainty in an estimator and the variation in the data set. The former can be reported as confidence intervals, and the latter should refer to the raw data.

As I become more mature in theories and skills, I discover that there is much more to consider to be a reliable data analyst. Holding a fair attitude towards any statistical tool and always remember to communicate the details of an analysis are essential. I would like myself to not only gain knowledge but also know the responsibilities as I move into a professional career.

Word count: 492 words

References

Motulsky, H. J. (2014). Common misconceptions about data analysis and statistics. *Naunyn-Schmiedeberg's Archives of Pharmacology*, 387(11), 1017–1023. <https://doi.org/10.1007/s00210-014-1037-6>

Reflection

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

I am proud in this portfolio that I can quickly reflect back on the course materials and come to a solution in R regarding linear mixed model, and I also found myself much confident in interpreting confidence intervals and p-values comparing to when I did the prerequisite knowledge check. I am also able to paraphrase a paper without referencing too much exact text by the author, and I can use this paper as a ethical guideline for my future works. This portfolio improves my confidence in the use of the `tidyverse` library and `ggplot2` for visualization, the proficiency in fitting linear mixed models in R, as well as paraphrasing ability and communicating ideas using a professional tone. The final typeset output file also looks nice and professional, which I think I am comfortable to showcase to other people.

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

This portfolio mainly focuses on simulation studies, demonstration of understanding of important statistics concepts, these are the essential skills to present, reproduce, and communicate concepts in any statistical study. These would be very helpful as I move on to future portfolios and projects. Also, the model fitting skills with linear mixed models would also be a powerful tool when I encounter similar data sets in the future. I will continue to improve these skills further in this course along with proficiency in using `tidyverse` and `ggplot`. In the writing sample, my paraphrasing skills also get improved, and that the ethical guidelines will also be an important asset in future works.

What is something I'd do differently next time?

I think overall this portfolio demonstrated different skills in pieces, so for the next time, I would like to add more structures to the write-up so that the demonstration can appear more logical for readers to follow. For examples, the individual sections of the statistical skill sample part are independent of each other, and are presented without an entirety. What I will do next time is, based the whole study on some simulated data sets, do inferences with them and use the p-value and confidence interval interpreters in this context, plot the distribution of p-values, fit and interpret linear mixed models using the data sets.

Also for the writing sample part, next time I would like to write it in a more professional rather than personal manner (for example, avoid using first person pronouns), and I can also discuss some technical or mathematical details presented in the paper beyond simple paraphrasing.