

Example Analysis: Using the SGBA-5 with Simulated Data

A Putman S Dogra

2025-07-26

Contents

1	About	5
1.1	Further Resources	5
2	Data Structure	7
2.1	Collected SGBA-5 Responses	7
2.2	Simulated Data	8
3	Descriptive Analysis	9
3.1	Visualize Distribution of SGBA-5 Responses	9
3.2	Descriptive Table of SGBA-5 Responses	11
4	SGBA of a Continuous Variable	13
4.1	Biological Sex	13
4.2	Gendered Aspects of Health: Assessing One Item	16
4.3	Gendered Aspects of Health: Assessing Multiple Items	21
5	Appendix	27
5.1	Simulated Data Creation	27
5.2	Ch. 3 Descriptive Analysis	33
5.3	Ch. 4 SGBA of a Continuous Variable	35

Chapter 1

About

The following example analysis demonstrates a way in which sex-and gender-based analysis (SGBA) can be conducted using results from the Sex- and Gender-Based Analysis Tool – 5 item (SGBA-5) by Putman and Dogra which can be downloaded [here](#). The SGBA-5 is free to use and is licensed under CC BY 4.0.

The examples shown in the following sections cover a descriptive analysis of simulated SGBA-5 responses (e.g., what SGBA-5 results may look like in a Table 1). Then an approach is demonstrated where a continuous variable (i.e., the primary outcome of a study) is assessed by looking for potential interactions between that continuous variable and the biological sex item or the gendered aspects of health items from the SGBA-5. The examples provided do not cover all potential analyses that could be done using the SGBA-5 but do provide a solid foundation from which researchers can select from and build upon.

The R code to reproduce the plots and statistical tests used in the examples is included in the Appendix.

1.1 Further Resources

More information on the SGBA-5, instructions for its use, and rationale for are included in the SGBA-5's documentation. Initial reliability and validity testing of the SGBA-5 are reported in the paper by Putman, Cole, & Dogra (2025) in PLOS ONE and in A Putman's thesis work.

Chapter 2

Data Structure

2.1 Collected SGBA-5 Responses

After collecting responses from the SGBA-5 you will have a dataset that looks something like this:

Where:

pt_id is the participant identifier.

sex is the SGBA-5 categorical Biological Sex item.

- Response options of “male”, “female”, and “intersex”.

gen_id is the SGBA-5 Gender Identity gendered aspect of health item.

- Responses are recorded as ordinal values between 0 to 100 on a feminine to masculine scale (measured in mm if completed on paper).

gen_exp is the SGBA-5 Gender Expression gendered aspect of health item.

Table 2.1: Example Data Structure

pt_id	sex	gen_id	gen_exp	gen_role	gen_rel
1	Female	48	99	70	25
2	Male	100	17	99	2
3	Female	64	48	88	40
4	Female	24	46	36	88
5	Male	73	23	19	26

The values in this table are placeholders for the example, not real data

- Responses are recorded as ordinal values between 0 to 100 on a feminine to masculine scale (measured in mm if completed on paper).

`gen_role` is the SGBA-5 Gender Role gendered aspect of health item.

- Responses are recorded as ordinal values between 0 to 100 on a feminine to masculine scale (measured in mm if completed on paper).

`gen_rel` is the SGBA-5 Gender Relations gendered aspect of health item.

- Responses are recorded as ordinal values between 0 to 100 on a feminine to masculine scale (measured in mm if completed on paper).

2.2 Simulated Data

This example analysis uses simulated data, if you wish to replicate the simulate data the code to do so is included in the Appendix or a .csv file of the simulated data can be downloaded from this example's github page.

Chapter 3

Descriptive Analysis

See Appendix section 5.2 for the R code used in this chapter.

Note: for conciseness, the following examples will only show results for two of the four gendered aspects of health items from the SGBA-5 (gender identity, and gender roles)

3.1 Visualize Distribution of SGBA-5 Responses

3.1.1 Biological Sex Item

In plot 3.1 we see that there are more participants who report their **biological sex** as assigned as female at birth (n=18) than males (n=12).

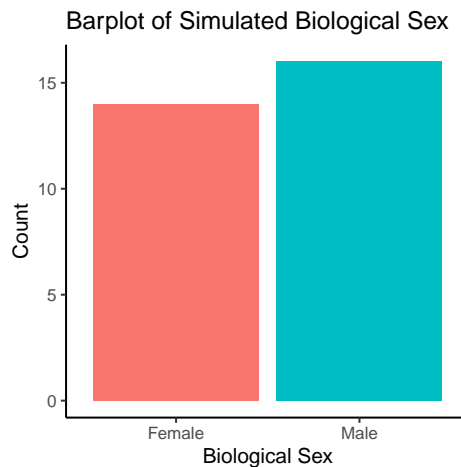


Figure 3.1: Barplot of Biological Sex Responses

3.1.2 Gendered Aspect of Health Items

Looking at the density plots for **gender identity** and **gender role** (figure 3.2), we see that while both variables are bimodal, the **gender identity** responses is more strongly bimodal with one peak closer to the feminine side of the feminine-masculine continuum and one peak closer to masculine end of that continuum. Further, we can also see that in general, participants reported their **gender identity** and **roles** as being more feminine, again with the **gender identity** responses showing this trend more strongly than the **gender role** responses.

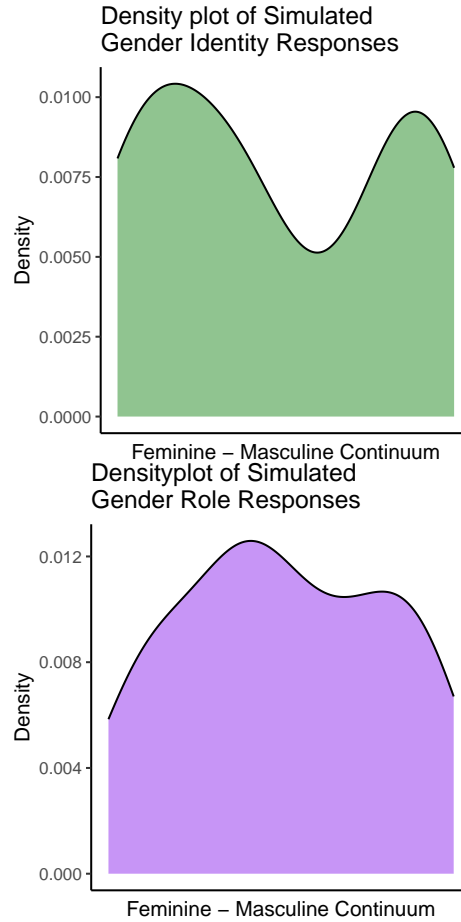


Figure 3.2: Density plots of Gender Identity and Roles

Presently, there is *no clear consensus on how to present descriptive statistics of bimodal variables in health research*. Reporting the mean(sd) or median(IQR) of a variable does not clearly communicate that there is more than one peak in a bimodal variable's distribution.

Table 3.1: Potential Interpretation of Sample Means for Gendered Aspect of Health Items.

Mean	Interpretation
>70	"Skews masculine"
55 to 70	"More masculine than feminine"
45 to 55	"Not strongly skewed"
30 to 45	"More feminine than masculine"
<30	"Skews feminine"
This table assumes you have recorded the gendered aspects of health items as 0 being the most feminine score and 100 being the most masculine score.	

When taken alongside the SGBA-5's assumption that the feminine-masculine continuum doesn't have a true 0 value, it is our suggestion that:

*If researchers decide to report a descriptive statistic for a bimodally distributed variable, they should report a nominal description of the variable's distribution **skew**.*

For example, if a gendered aspect of health item from the SGBA-5 demonstrates a bimodal distribution, the variable's **skew** can be described along the feminine-masculine continuum instead of reporting the numerical average alone.

To be able to describe the **skew** of one of the gender variables, the authors suggest calculating the variable's mean score along the feminine-masculine continuum and then reporting the skew using the suggested nominal classification guide seen in Table 3.1.

Please note that these suggested classification guidelines are arbitrary and may not be appropriate in all circumstances.

For the simulated dataset represented in the density plots above, the mean score for the **gender identity** item was 50.2 and 46.8 for the **gender role** item. This means that when reporting descriptive statistics on the simulated sample we could report that: *"The simulated sample was not strongly skewed on a feminine to masculine continuum for either the gender identity or gender role measures from the SGBA-5".*

3.2 Descriptive Table of SGBA-5 Responses

Taking all these together, an example of a sample characteristics table of the SGBA-5 items in the simulated dataset could be presented as has been displayed in Table 3.2

Table 3.2: Simulated sample characteristics.

SGBA Item	Sample (n = 30)
Biological Sex (n(%))	
<i>Female</i>	14(47%)
<i>Intersex</i>	NA
<i>Male</i>	16(53%)
Gendered Aspect of Health (skew)	
<i>Gender Identity</i>	Not strongly skewed
<i>Gendered Roles</i>	Not strongly skewed

See Appendix section 5.2 for the R code used in this chapter.

Chapter 4

SGBA of a Continuous Variable

See Appendix section 5.3 for the R code used in this chapter.

Note: for the biological sex and single gendered aspect of health item examples, an example outcome variable that demonstrates a relationship between that variable and the SGBA-5 item being assessed - the interaction example - and an example that does not demonstrate a relationship between itself and the SGBA-5 item - the no interaction example - is shown.

4.1 Biological Sex

A good idea is to start by visualizing the continuous variable's distribution disaggregated by sex like the density plot in Figure 4.1. Then calculate disaggregated summary statistics for the continuous variable disaggregated by sex (Tables 4.1 and 4.2), and conduct a statistical test of difference in means (Welch's t-test for this example).

4.1.1 Density Plots by Biological Sex

Interpretation: From the above density plots (Figure 4.1) we can see a distinct overlap in the **no interaction example** which suggests that in that example's sample does not have a meaningful difference in the continuous outcome by sex. Conversely, the **interaction example** density plot has two distinct peaks which suggests that its sample's continuous outcome scores are associated with a participant's sex.

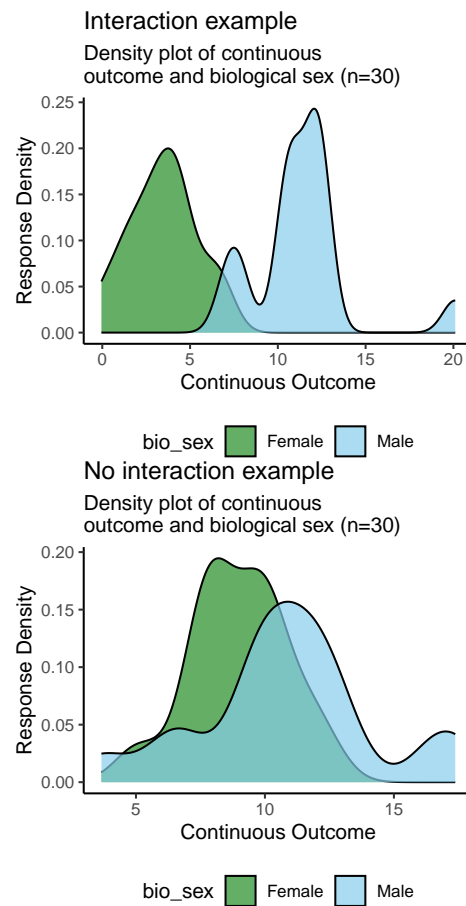


Figure 4.1: Density Plot of Continuous Variable by Biological Sex Examples

Table 4.1: Interaction example summary statistics: Continuous outcome and biological sex

biological sex	n	mean continuous	SD continuous	median continuous	IQR continuous
Female	14	3.5	1.94	4	2
Male	16	11.3	2.95	11	2

Table 4.2: No interaction example summary statistics: Continuous outcome and biological sex

biological sex	n	mean continuous	SD continuous	median continuous	IQR continuous
Female	14	9.0	1.84	9	2
Male	16	10.8	3.47	11	3

4.1.2 Summary Statistics by Biological Sex

Interpretation: As with the density plots, we see that the standard deviations of the continuous variable for both males and females overlap in the **no interaction example** (Table 4.1) - indicating a lack of significant difference by sex. The standard deviations of the continuous variable for both males and females do not overlap in the **interaction example** (Table 4.2) - indicating a potential association between that continuous outcome variable and sex.

4.1.3 Statistical Testing of Biological Sex: t-test

Next we will test the null hypothesis that biological sex does not have an impact on the continuous outcome variable being evaluated. Both the interaction example and no interaction example will be tested using a Welch's t-test at an alpha of .05.

Note: Using a Welch's t-test to test for statistically significant difference in these examples is a way, but by no means the only way in which this could be tested.

Interpretation: Similarly to the descriptive table and density plots, we see that the **no interaction example** (Table 4.3) does not show a significant

Table 4.3: Statistical test of difference: Continuous outcome and biological sex.

Example	Test	T-score	95% CI	df	p-value
Interaction	Welch's t-test	-8.73	(-9.72, -6.02)	26.1	0.000
No interaction	Welch's t-test	-1.80	(-3.85, 0.26)	23.4	0.084

difference by sex ($T=-1.80$, $95\%CI=(-3.85,0.26)$, $df=23.4$, $p\text{-value}>0.08$). The **interaction example** (Table 4.3) shows a potential association between that continuous outcome variable and sex ($T=-8.73$, $95\%CI=(-9.82,-6.02)$, $df=23.4$, $p\text{-value}<0.001$), meriting further research into this potential interaction. This means that when reporting whether there was an association found between a continuous outcome and biological sex in the no interaction sample we can report that: *“The no interaction sample did not show evidence that biological sex was associated with the continuous outcome in this study”*. Whereas the interaction sample, which did find a potential association, could report: *“The interaction sample found statistically significant evidence that biological sex was associated with the continuous outcome in this study”*.

4.1.4 Biological Sex: Interpretation Reporting Template

Below are example templates for reporting whether an outcome variable is associated with biological sex as categorized by the SGBA-5. Replace the words in the square brackets to complete.

If an association was found:

In this study’s sample we found that a person’s self-reported biological sex at birth had a statistically significant association with [continuous variable name]. More detailed investigation of this relationship is required to directly interpret the potential effects this interaction.

If an association was not found:

In this study’s sample we found that a person’s self-reported biological sex at birth did not show a statistically significant association with [continuous variable name].

4.2 Gendered Aspects of Health: Assessing One Item

In this section we will analyze a single of the gendered aspects of health results from the SGBA-5 with a continuous variable of interest. Using the gender identity item as an example. First we will generate scatter (Figure 4.2) and 2D-density plots (Figure 4.3) of the continuous variable by the feminine-masculine continuum used in the SGBA-5. Then we shall calculate a Pearson correlation coefficient for that gendered aspect of health item and the continuous variable.

4.2.1 Scatter & 2-D Density Plots for One Item

Interpretation: From the above scatter and 2-D density plots (Figures 4.2 and 4.3) we can see no clear pattern difference in the continuous outcome value across the feminine-masculine continuum in the **no interaction example** (i.e., the continuous outcome values of participants who are closer to the feminine end of

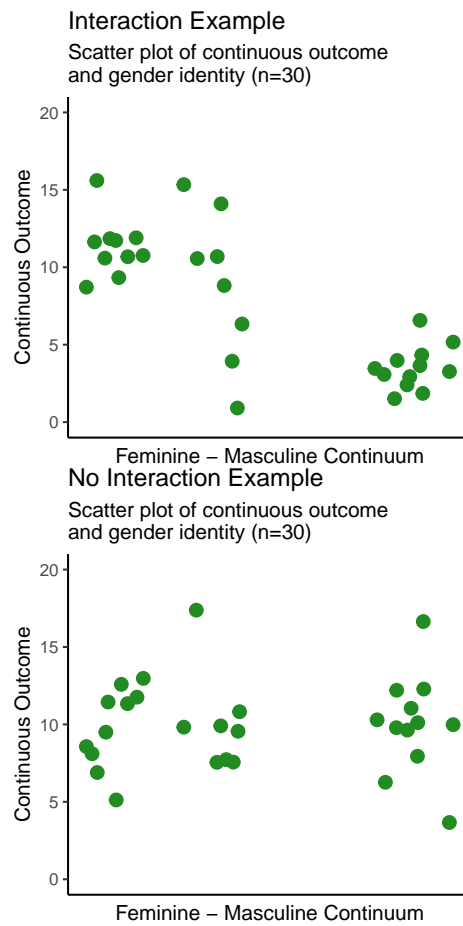


Figure 4.2: Scatter Plot of Continuous Variable by Gender Identity

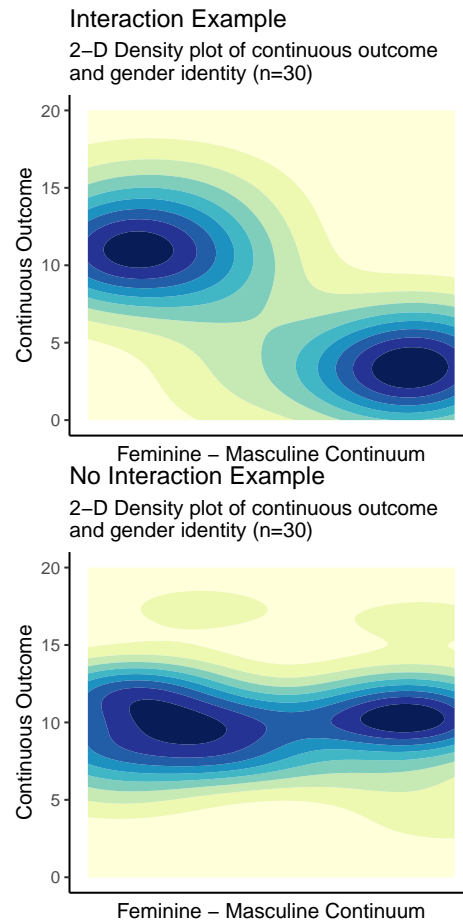


Figure 4.3: 2D Density Plots of Continuous Variable by Gender Identity

the continuum are not noticeably different than participants who are not closer to the feminine end of the continuum). This suggests that the **no interaction example** does not show a meaningful difference in the continuous outcome by gender identity within this sample. Conversely, the **interaction example** scatter and 2-D density plots show two distinct clusters with the cluster closer to the feminine end of the continuum having higher continuous outcome values than the cluster closer to the masculine end of the continuum. This suggests that the **interaction example** sample's continuous outcome scores are associated with participants' self-reported place on the SGBA-5's feminine-masculine gender identity continuum.

4.2.2 Summary Statistics for One Gendered Aspect of Health Item

We do not recommend creating disaggregated summary statistics by any of the gendered aspects of health items from the SGBA-5. Analysis for effects associated with these variables can be conducted visually using plots (section 4.2.1) or through statistical testing (section 4.2.3).

4.2.3 Statistical Testing for One Item: Pearson's Correlation Coefficient (r)

Next we will test the null hypothesis that a participant's self-reported gender identity on a feminine-masculine continuum does not have an impact on the continuous outcome variable being evaluated. Both the interaction example and no interaction example will be tested by calculating Pearson's Correlation Coefficient and describing that coefficient's strength and direction (*but not the exact coefficient value*) as well as test for statistical significance of the correlation at an alpha of .05.

To be able to describe the strength of correlation between a continuous outcome and one of the gendered aspect of health variables, the authors suggest using a nominal classification of correlation strength. Using a nominal classification to interpret the strength of a correlation is inherently somewhat arbitrary, however, this trade-off can reduce potential misinterpretation that reporting a numerical coefficient value could cause. There maybe context- or discipline-specific guidelines for the interpretation of correlation strength which apply to your analysis. For the examples presented here, we will use the guidelines recommended in *Statistical Power Analysis for the Behavioural Sciences* (2013) by Jacob Cohen, which can be seen in Table 4.4.

Please note that these suggested classification guidelines are arbitrary and may not be appropriate in all circumstances.

Note: Using Pearson's correlation coefficient to test for statistically significant difference in these examples is a way, but by no means

Table 4.4: Nominal Interpretations of Correlation Strength.

Strength of Correlation	Positive Coefficient (r)	Negative Coefficient (r)
None	0.0 to 0.1	0.0 to -0.1
Weak	0.1 to 0.3	-0.1 to -0.3
Moderate	0.3 to 0.5	-0.3 to -0.5
Strong	0.5 to 1.0	-0.5 to -1.0

Adapted from [__Statistical Power Analysis for the Behavioural Sciences__ (2013) by Jacob Cohen](<https://doi.org/10.4324/9780203771587>)

Table 4.5: Statistical test of difference: Continuous outcome and Gender Identity Item from the SGBA-5.

Example	Correlation Type	r	95% CI	df	Significance Test	T-score	p-value
Interaction	Pearson's r	-0.80	(-0.9, -0.61)	28	t-test	-6.96	0.000
No interaction	Pearson's r	0.02	(-0.34, 0.38)	28	t-test	0.10	0.922

the only way in which this could be tested.

Interpretation: The correlation coefficient we see for the **no interaction example** (Table 4.5) was 0.02. Using the classification criteria outlined in Table 4.4, we see that the strength of correlation is “None”. Thus, we can report that “*The no interaction sample did not show evidence of correlation between the gender identity item and the continuous outcome*”. Similarly, we see that the **no interaction example**’s correlation does not meet the statistical significance threshold of 0.05 from the null hypothesis (T=0.10, df=28, p-value>0.9).

The **interaction example** shows a correlation coefficient of -0.80 (Table 4.5), which falls under the “Strong” category of correlation strength from Table 4.4. Therefore, we could report that “*The interaction sample shows evidence of a strong correlation between the gender identity item and the continuous outcome*”. Similarly, we see that the **interaction example**’s correlation exceeds the statistical significance threshold of 0.05 from the null hypothesis (T=-6.96, df=28, p-value<0.001).

4.2.4 Gendered Aspects of Health: One Item Interpretation Reporting Template

Below are example templates for reporting whether an outcome variable is associated with a gendered aspect of health as measured using the SGBA-5. Replace the words in the square brackets to complete.

If an association was found:

In this study’s sample we found that a person’s self-reported [gen-

dered aspect of health item] had a statistically significant association with [continuous variable name]. More detailed investigation of this relationship is required to directly interpret the potential effects this interaction.

If an association was not found:

In this study's sample we found that a person's self-reported [gendered aspect of health item] did not show a statistically significant association with [continuous variable name].

4.3 Gendered Aspects of Health: Assessing Multiple Items

In this section we will show how an example of how you can analyze multiple gendered aspects of health results from the SGBA-5 with a continuous variable of interest. (Specifically, this example will demonstrate a SGBA using the gender identity and gender roles items from the SGBA-5. However, the process for assessing 3 of or all 4 items with a continuous variable can be extended from this example).

Like when assessing only one of the gendered aspect of health items, we will start by generating scatter (Figure 4.4) and 2D-density plots (Figure 4.5) of the continuous variable by the feminine-masculine continuum used in the SGBA-5. Then we shall calculate Pearson correlation coefficients between each gendered aspect of health item and the continuous variable.

4.3.1 Scatter & 2-D Density Plots for Two Items

Interpretation: From the above scatter and 2-D density plots (Figures 4.4 and 4.5) we can see a relatively clear pattern difference in the continuous outcome value across the feminine-masculine continuum in the **gender identity** item. Whereas, the **gendered roles** item scatter and 2-D density plots show less distinct clustering with a cluster closer to the masculine end of the continuum having lower continuous outcome values and the continuous outcome values closer to the feminine end of the continuum showing both high and low values. This suggests that in this example the **gender identity** SGBA-5 item is associated with the continuous outcome variable in this example and the **gendered roles** item is not clear, but may suggest the potential of a more complicated relationship between the continuous outcome and **gendered roles**.

4.3.2 Summary Statistics for Two Items

As with assessing only one of the gendered aspects of health items from the SGBA-5, we do not recommend creating disaggregated summary statistics by any of the gendered aspects of health items from the SGBA-5. Analysis for effects associated with these variables can

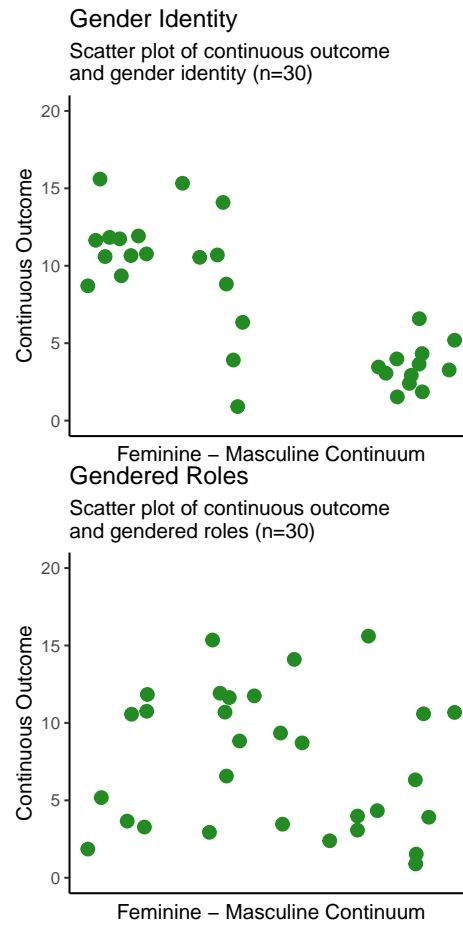


Figure 4.4: Scatter Plot of Continuous Variable by Gender Identity and Gendered Roles

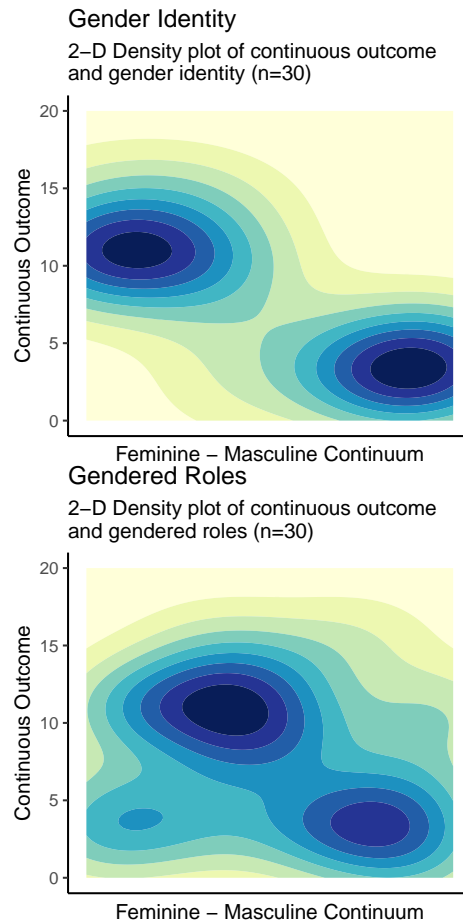


Figure 4.5: 2D Density Plots of Continuous Variable by Gender Identity and Gendered Roles

Table 4.6: Statistical test of difference: Continuous outcome by Gender Identity and Gendered Roles Items from the SGBA-5.

Example	Correlation Type	r	95% CI	df	Significance Test	T-score	p-value
Gender Identity	Pearson's r	-0.8	(-0.9, -0.61)	28	t-test	-6.96	0.000
Gendered Roles	Pearson's r	-0.1	(-0.45, 0.27)	28	t-test	-0.56	0.582

be conducted visually using plots (section 4.3.1) or through statistical testing (section 4.3.3).

4.3.3 Statistical Testing for Two Items: Pearson's Correlation Coefficient (r)

Next we will test the null hypothesis that a participant's self-reported gender identity or gendered roles on a feminine-masculine continuum do not impact the continuous outcome variable being evaluated.

*Note: typically when conducting multiple comparisons we adjust the threshold for statistical significance to account for those multiple comparisons (e.g., using a Bonferroni Correction for calculating p-value thresholds). In the the example below we do not use a multiple comparison correction. This is because the calculated correlation values are assumed to be approximately indicative of an underlying association rather than interpretable as exact values. However, **if you intend to look at SGBA-5 items in comparison to multiple primary variables in your analysis we highly recommend adjusting your significance thresholds for multiple comparisons.***

Both the simulated gender identity and gendered roles SGBA-5 responses shown in the plots above will be tested for association with the continuous outcome variable by calculating Pearson's Correlation Coefficient and describing that coefficient's strength and direction (*but not the exact coefficient value*) as well as test for statistical significance of the correlation at an alpha of .05.

As described previously in section 4.2.3, the authors suggest using a nominal classification of correlation strength. For the examples presented here, we will use the guidelines recommended in *Statistical Power Analysis for the Behavioural Sciences* (2013) by Jacob Cohen, which is summarized in Table 4.4.

Note: Using Pearson's correlation coefficient to test for statistically significant difference in these examples is a way, but by no means the only way in which this could be tested.

Interpretation: The **gender identity** item shows a correlation coefficient of -0.80 (Table 4.6), which falls under the "Strong" category of correlation strength

from Table 4.4. Therefore, we could report that “*Our testing found evidence of a strong correlation between the gender identity item and the continuous outcome*”. As we would expect given the strength of correlation, the **gender identity** item’s correlation exceeds the statistical significance threshold of 0.05 from the null hypothesis ($T=-6.96$, $df=28$, $p\text{-value}<0.001$).

The correlation coefficient we see for the **gendered roles** item (Table 4.6) was -0.1. Using the classification criteria outlined in Table 4.4, we see that the strength of correlation is on the border between “None” and “Weak”. Airing on the side of caution, we can report that “*Our testing did not find clear evidence of correlation between the gendered roles item and the continuous outcome*”. Similarly, we see that the **gendered roles** item’s correlation does not meet the statistical significance threshold of 0.05 from the null hypothesis ($T=-0.56$, $df=28$, $p\text{-value}>0.5$).

4.3.4 Gendered Aspects of Health: Multiple Item Interpretation Reporting Template

Below are example templates for reporting whether an outcome variable is associated with gendered aspects of health as measured using the SGBA-5. Replace the words in the square brackets to complete.

If associations were found for all of the gendered aspect of health items assessed:

In this study’s sample we found that a person’s self-reported [gendered aspect of health items] demonstrated statistically significant associations with [continuous variable name]. More detailed investigation of these relationships is required to directly interpret the potential effects these interactions.

If associations were found for at least one of the gendered aspect of health items assessed:

In this study’s sample we found that a person’s self-reported [gendered aspect of health items with association] demonstrated [a] statistically significant association[s] with [continuous variable name], whereas a person’s self-reported [gendered aspect of health items with no association] did not. More detailed investigation of these relationships is required to directly interpret the potential effects these interactions.

If no association was found:

In this study’s sample we found that a person’s self-reported [gendered aspect of health items] did not show a statistically significant association with [continuous variable name].

See Appendix section 5.3 for the R code used in this chapter.

Chapter 5

Appendix

5.1 Simulated Data Creation

Below is the code used to create the simulated data that will be used to create the example SGBA seen throughout the rest of this example analysis.

```
# load libraries -----
library(tidyverse)

## -- Attaching core tidyverse packages ----- tidyverse 2.
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr    1.5.1
## v ggplot2    3.5.2      v tibble     3.3.0
## v lubridate  1.9.4      v tidyr      1.3.1
## v purrr      1.0.4
## -- Conflicts ----- tidyverse_conflict
## 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
# set seed for predictability
set.seed(42, kind = "Mersenne-Twister")

# write simulation functions -----

# function for simulating bimodal gender variables
binom_bounded <- function(
  n, prop, mean1, sd1, mean2, sd2, lim_low = -Inf, lim_up = Inf, round
){
  # error checking
  if(prop > 1 || prop < 0){stop("binomial proportion should be between 0 and 1")}
}
```

```

if(n < 1){stop("n must be greater than or equal to 1")}
if(sd1 < 0 || sd2 < 0){stop("standard deviations must be non-negative")}
# create index
index <- rbinom(n, size = 1, prob = prop)
# simulate each peak using a bounded `rnorm()`
binom_sim1 <- index * qnorm(
  runif(
    n, pnorm(lim_low, mean1, sd1), pnorm(lim_up, mean1, sd1)
  ),
  mean1, sd1)
binom_sim0 <- (1 - index) * qnorm(
  runif(
    n, pnorm(lim_low, mean2, sd2), pnorm(lim_up, mean2, sd2)
  ),
  mean2, sd2)
binom_sim <- binom_sim0 + binom_sim1
binom_sim <- round(binom_sim, round)
binom_sim
}

# function for simulating Likert/ordinal outcome variable
ordinal_sim <- function(
  n, mean, sd, lim_low = -Inf, lim_up = Inf
){
  ord_sim <- qnorm(
    runif(
      n, pnorm(lim_low, mean, sd), pnorm(lim_up, mean, sd)
    ),
    mean, sd)
  ord_sim <- round(ord_sim, 0)
  ord_sim
}

# simulate SGBA-5 data with n of 30 -----

# biological sex: categorical (female, intersex, male)
bio_sex <- factor(
  sample(c('Female', 'Male'), 30, replace=TRUE, prob=c(0.6, 0.4)),
  )

# gender identity: ordered (0 [feminine] to 100 [masculine])
gen_id <- binom_bounded(
  n = 30, prop = 0.6, mean1 = 20, sd1 = 20, mean2 = 85, sd2 = 15, lim_low = 0,

```

```

    lim_up = 100, round = 0
  )

  # gender role: ordered (0 [feminine] to 100 [masculine])
  gen_role <- binom_bounded(
    n = 30, prop = 0.6, mean1 = 25, sd1 = 20, mean2 = 80, sd2 = 20, lim_low = 0,
    lim_up = 100, round = 0
  )

  # simulate example positive outcome data with n of 30 -----

  ## simulated numerical outcome: -----

  # by sex: males (mean = 12, SD = 3), females (mean = 4, sd = 2)
  outcome_num_pos_m <- rnorm(n = 16, mean = 12, sd = 3)
  outcome_num_pos_f <- rnorm(n = 14, mean = 4, sd = 2)
  outcome_num_pos_sex <- c(outcome_num_pos_f, outcome_num_pos_m)

  # by g_id: high (mean = 12, SD = 3), low (mean = 4, sd = 2)
  outcome_num_pos_gid_high <- rnorm(n = 15, mean = 12, sd = 3)
  outcome_num_pos_gid_low <- rnorm(n = 15, mean = 4, sd = 2)
  outcome_num_pos_gid <- c(outcome_num_pos_gid_high, outcome_num_pos_gid_low)

  # by g_role: high (mean = 12, SD = 3), low (mean = 4, sd = 2)
  outcome_num_pos_grol_high <- rnorm(n = 12, mean = 12, sd = 3)
  outcome_num_pos_grol_low <- rnorm(n = 18, mean = 7, sd = 2)
  outcome_num_pos_grol <- c(outcome_num_pos_grol_high, outcome_num_pos_grol_low)

  ## simulated Likert outcome: -----

  # by sex males (mean = 2, SD = 1), females (mean = 5, sd = 2)
  outcome_ord_pos_m <- ordinal_sim(
    n = 16, mean = 2, sd = 1, lim_low = 1, lim_up = 7
  )
  outcome_ord_pos_f <- ordinal_sim(
    n = 14, mean = 5, sd = 1, lim_low = 1, lim_up = 7
  )
  outcome_ord_pos_sex <- c(outcome_ord_pos_f, outcome_ord_pos_m) %>%
    factor(., levels = c(1,2,3,4,5,6,7), ordered = TRUE)

  # by g_id: high (mean = 5, SD = 2), low (mean = 1, sd = 2)
  outcome_ord_pos_high <- ordinal_sim(

```

```

  n = 15, mean = 5, sd = 2, lim_low = 1, lim_up = 7
)
outcome_ord_pos_low <- ordinal_sim(
  n = 15, mean = 1, sd = 2, lim_low = 1, lim_up = 7
)
outcome_ord_pos_gid <- c(outcome_ord_pos_high, outcome_ord_pos_low) %>%
  factor(., levels = c(1,2,3,4,5,6,7), ordered = TRUE)

# by g_role: high (mean = 5, SD = 2), low (mean = 1, sd = 2)
outcome_ord_pos_high <- ordinal_sim(
  n = 12, mean = 5, sd = 2, lim_low = 1, lim_up = 7
)
outcome_ord_pos_low <- ordinal_sim(
  n = 18, mean = 1, sd = 2, lim_low = 1, lim_up = 7
)
outcome_ord_pos_grol <- c(outcome_ord_pos_high, outcome_ord_pos_low) %>%
  factor(., levels = c(1,2,3,4,5,6,7), ordered = TRUE)

## simulated binary outcome: -----

# by sex males (yes = .2, no = .8), females (yes = .8, no = .2)
outcome_bin_pos_m <- sample(
  c('yes', 'no'), 16, replace = TRUE, prob = c(.2, .8)
)
outcome_bin_pos_f <- sample(
  c('yes', 'no'), 14, replace = TRUE, prob = c(.8, .2)
)
outcome_bin_pos_sex <- append(outcome_bin_pos_f, outcome_bin_pos_m) %>%
  factor()

# by g_id: high (yes = .4, no = .6), low (yes = .8, no = .2)
outcome_bin_pos_high <- sample(
  c('yes', 'no'), 15, replace = TRUE, prob = c(.4, .6)
)
outcome_bin_pos_low <- sample(
  c('yes', 'no'), 15, replace = TRUE, prob = c(.8, .2)
)
outcome_bin_pos_gid <- append(outcome_bin_pos_high, outcome_bin_pos_low) %>%
  factor()

# by g_rol: high (yes = .3, no = .5), low (yes = .8, no = .2)

```

```

outcome_bin_pos_high <- sample(
  c('yes','no'), 12, replace = TRUE, prob = c(.3, .7)
)
outcome_bin_pos_low <- sample(
  c('yes','no'), 18, replace = TRUE, prob = c(.8, .2)
)
outcome_bin_pos_grol <- append(outcome_bin_pos_high, outcome_bin_pos_low) %>%
  factor()

# simulate example negative outcome data with n of 30 -----

# simulated numerical outcome: continuous (mean = 10, SD = 3)
outcome_num_neg <- rnorm(n = 30, mean = 10, sd = 3)

# simulated Likert outcome: 7-point Likert scale (mean = 4, SD = 2)
outcome_ord_neg <- ordinal_sim(
  n = 30, mean = 4, sd = 2, lim_low = 1, lim_up = 7
)

# simulated binary outcome: categorical (yes, no)
outcome_bin_neg <- factor(
  sample(c('yes','no'), 30, replace = TRUE, prob = c(0.67, 0.33))
)

# create example data frame -----

# combine into dataframe
sim_data <- tibble(
  bio_sex, gen_id, gen_role, #outcome_bin_pos,
  outcome_num_neg, outcome_ord_neg, outcome_bin_neg
) %>% arrange(., bio_sex) %>%
  cbind(., outcome_num_pos_sex) %>%
  cbind(., outcome_ord_pos_sex) %>%
  cbind(., outcome_bin_pos_sex) %>%
  arrange(., gen_id) %>%
  cbind(., outcome_num_pos_gid) %>%
  cbind(., outcome_ord_pos_gid) %>%
  cbind(., outcome_bin_pos_gid) %>%
  arrange(., gen_role) %>%
  cbind(., outcome_num_pos_grol) %>%
  cbind(., outcome_ord_pos_grol) %>%
  cbind(., outcome_bin_pos_grol)

```

```
# save simulated df
write_csv(sim_data, file = "sim-data.csv")
```

5.1.1 Session Info for Data Creation

```
sessioninfo::session_info(pkgs = "loaded")
```

```
## - Session info -----
## setting value
## version R version 4.5.0 (2025-04-11)
## os macOS Sequoia 15.5
## system aarch64, darwin20
## ui X11
## language (EN)
## collate en_US.UTF-8
## ctype en_US.UTF-8
## tz America/Toronto
## date 2025-07-26
## pandoc 3.4 @ /Applications/RStudio.app/Contents/Resources/app/quarto/bin/tools/a
## quarto 1.6.42 @ /Applications/RStudio.app/Contents/Resources/app/quarto/bin/quar
##
## - Packages -----
## package * version date (UTC) lib source
## bit 4.6.0 2025-03-06 [1] CRAN (R 4.5.0)
## bit64 4.6.0-1 2025-01-16 [1] CRAN (R 4.5.0)
## bookdown 0.43 2025-04-15 [1] CRAN (R 4.5.0)
## cli 3.6.5 2025-04-23 [1] CRAN (R 4.5.0)
## crayon 1.5.3 2024-06-20 [1] CRAN (R 4.5.0)
## digest 0.6.37 2024-08-19 [1] CRAN (R 4.5.0)
## dplyr * 1.1.4 2023-11-17 [1] CRAN (R 4.5.0)
## evaluate 1.0.4 2025-06-18 [1] CRAN (R 4.5.0)
## farver 2.1.2 2024-05-13 [1] CRAN (R 4.5.0)
## fastmap 1.2.0 2024-05-15 [1] CRAN (R 4.5.0)
## forcats * 1.0.0 2023-01-29 [1] CRAN (R 4.5.0)
## generics 0.1.4 2025-05-09 [1] CRAN (R 4.5.0)
## ggplot2 * 3.5.2 2025-04-09 [1] CRAN (R 4.5.0)
## glue 1.8.0 2024-09-30 [1] CRAN (R 4.5.0)
## gtable 0.3.6 2024-10-25 [1] CRAN (R 4.5.0)
## hms 1.1.3 2023-03-21 [1] CRAN (R 4.5.0)
## htmltools 0.5.8.1 2024-04-04 [1] CRAN (R 4.5.0)
## knitr 1.50 2025-03-16 [1] CRAN (R 4.5.0)
## lifecycle 1.0.4 2023-11-07 [1] CRAN (R 4.5.0)
## lubridate * 1.9.4 2024-12-08 [1] CRAN (R 4.5.0)
## magrittr 2.0.3 2022-03-30 [1] CRAN (R 4.5.0)
```



```
## pillar          1.11.0 2025-07-04 [1] CRAN (R 4.5.0)
## pkgconfig       2.0.3 2019-09-22 [1] CRAN (R 4.5.0)
## purrr           * 1.0.4 2025-02-05 [1] CRAN (R 4.5.0)
## R6              2.6.1 2025-02-15 [1] CRAN (R 4.5.0)
## RColorBrewer    1.1-3 2022-04-03 [1] CRAN (R 4.5.0)
## readr           * 2.1.5 2024-01-10 [1] CRAN (R 4.5.0)
## rlang           1.1.6 2025-04-11 [1] CRAN (R 4.5.0)
## rmarkdown       2.29 2024-11-04 [1] CRAN (R 4.5.0)
## rstudioapi      0.17.1 2024-10-22 [1] CRAN (R 4.5.0)
## scales          1.4.0 2025-04-24 [1] CRAN (R 4.5.0)
## sessioninfo     1.2.3 2025-02-05 [1] CRAN (R 4.5.0)
## stringi         1.8.7 2025-03-27 [1] CRAN (R 4.5.0)
## stringr         * 1.5.1 2023-11-14 [1] CRAN (R 4.5.0)
## tibble          * 3.3.0 2025-06-08 [1] CRAN (R 4.5.0)
## tidyr           * 1.3.1 2024-01-24 [1] CRAN (R 4.5.0)
## tidyselect      1.2.1 2024-03-11 [1] CRAN (R 4.5.0)
## tidyverse       * 2.0.0 2023-02-22 [1] CRAN (R 4.5.0)
## timechange       0.3.0 2024-01-18 [1] CRAN (R 4.5.0)
## tzdb            0.5.0 2025-03-15 [1] CRAN (R 4.5.0)
## vctrs           0.6.5 2023-12-01 [1] CRAN (R 4.5.0)
## vroom           1.6.5 2023-12-05 [1] CRAN (R 4.5.0)
## withr           3.0.2 2024-10-28 [1] CRAN (R 4.5.0)
## xfun            0.52 2025-04-02 [1] CRAN (R 4.5.0)
## yaml           2.3.10 2024-07-26 [1] CRAN (R 4.5.0)
##
## [1] /Library/Frameworks/R.framework/Versions/4.5-arm64/Resources/library
## * -- Packages attached to the search path.
##
## -----
```

5.2 Ch. 3 Descriptive Analysis

5.2.1 Ch. 3 Set Up

```
# load libraries and simulated data -----
library(tidyverse)
library(ggpubr)

# set seed for reproducibility
set.seed(42)

# load data set (from RDS if using previous script to simulate the data)
sim_data <- read_csv(file = "sim-data.csv")

# default density plots and scale function -----
```

```

plot_dens_2D <- function(data, x, y){
  plot <- ggplot(data = data, aes(x = x, y = y)) +
    geom_density_2d_filled(aes(fill = after_stat(level)),
                          bins = 9, contour_var = 'ndensity') +
    scale_fill_brewer(palette = 16, direction = 1,
                     labels = c('Low', '', '', '', '', '', '', 'High')) +
    guides(fill = guide_legend(title = 'Density of Occurrences',
                              direction = 'horizontal', nrow = 1,
                              label.position = 'bottom'))

  plot
}

```

5.2.2 Figure 3.1

```

# biological sex bar plot
ggplot(sim_data, aes(x = bio_sex)) +
  geom_bar(aes(fill = bio_sex), show.legend = F) +
  labs(title = 'Barplot of Simulated Biological Sex') +
  ylab('Count') + xlab('Biological Sex') +
  theme_classic()

```

5.2.3 Figure 3.2

```

# gender identity density plot
ggplot(sim_data, aes(x = gen_id)) +
  geom_density(alpha=.5, fill = 'forestgreen', colour = 'black') +
  labs(title = 'Density plot of Simulated\nGender Identity Responses') +
  ylab('Density') + xlab('Feminine - Masculine Continuum') +
  theme_classic() +
  theme(
    legend.position = 'none',
    axis.ticks.x = element_blank(),
    axis.text.x = element_blank()
  )

# gender role density plot
ggplot(sim_data, aes(x = gen_role)) +
  geom_density(alpha=.5, fill = 'purple2', colour = 'black') +
  labs(title = 'Densityplot of Simulated\nGender Role Responses') +
  ylab('Density') + xlab('Feminine - Masculine Continuum') +
  theme_classic() +
  theme(
    legend.position = 'none',
    axis.ticks.x = element_blank(),

```

```
axis.text.x = element_blank()
)
```

5.3 Ch. 4 SGBA of a Continuous Variable

5.3.1 Ch. 4 Set Up

```
# load libraries and simulated data -----
library(tidyverse)
library(ggpubr)

# set seed for reproducibility
set.seed(42)

# load data set (from RDS if using previous script to simulate the data)
sim_data <- read_csv(file = "sim-data.csv")

# default density plots and scale function -----
plot_dens_2D <- function(data, x, y){
  plot <- ggplot(data = data, aes(x = x, y = y)) +
    geom_density_2d_filled(aes(fill = after_stat(level)),
                          bins = 9, contour_var = 'ndensity') +
    scale_fill_brewer(palette = 16, direction = 1,
                      labels = c('Low', '', '', '', '', '', '', 'High')) +
    guides(fill = guide_legend(title = 'Density of Occurences',
                               direction = 'horizontal', nrow = 1,
                               label.position = 'bottom'))

  plot
}
```

5.3.2 Figure 4.1

```
# biological sex bar plot - interaction
ggplot(sim_data, aes(x = outcome_num_pos_sex, fill = bio_sex)) +
  geom_density(alpha = 0.7) +
  scale_fill_manual(values = c('forestgreen', 'skyblue')) +
  labs(
    title = 'Interaction example',
    subtitle = 'Density plot of continuous \noutcome and biological sex (n=30)'
  ) +
  xlab('Continuous Outcome') +
  ylab('Response Density') +
  theme_classic() +
  theme(legend.position = 'bottom')
```

```
# biological sex bar plot - no interaction
ggplot(sim_data, aes(x = outcome_num_neg, fill = bio_sex)) +
  geom_density(alpha = 0.7) +
  scale_fill_manual(values = c('forestgreen', 'skyblue')) +
  labs(
    title = 'No interaction example',
    subtitle = 'Density plot of continuous \noutcome and biological sex (n=30)'
  ) +
  xlab('Continuous Outcome') +
  ylab('Response Density') +
  theme_classic() +
  theme(legend.position = 'bottom')
```

5.3.3 Table 4.1

```
# summary stats
pos_cont_sex_sum <- sim_data %>% group_by(bio_sex) %>%
  summarise(
    n = n(),
    `mean continuous` = round(mean(outcome_num_pos_sex),1),
    `SD continuous` = round(sd(outcome_num_pos_sex),2),
    `median continuous` = round(median(outcome_num_pos_sex),0),
    `IQR continuous` = round(IQR(outcome_num_pos_sex),)
  ) %>%
  rename(`biological sex` = bio_sex)

knitr::kable(pos_cont_sex_sum, booktabs = TRUE, caption = "Interaction example summary")
```

5.3.4 Table 4.2

```
neg_cont_sex_sum <- sim_data %>% group_by(bio_sex) %>%
  summarise(
    n = n(),
    `mean continuous` = round(mean(outcome_num_neg),1),
    `SD continuous` = round(sd(outcome_num_neg),2),
    `median continuous` = round(median(outcome_num_neg),0),
    `IQR continuous` = round(IQR(outcome_num_neg),)
  ) %>%
  rename(`biological sex` = bio_sex)

knitr::kable(neg_cont_sex_sum, booktabs = TRUE, caption = "No interaction example summary")
```

5.3.5 Table 4.3

```

# t-tests
con_sex_ttest_neg <- t.test(outcome_num_neg ~ bio_sex, data = sim_data)
con_sex_ttest_pos <- t.test(outcome_num_pos_sex ~ bio_sex, data = sim_data)

# example of collected data
con_sex_tab <- tibble(
  "Example" = c("Interaction", "No interaction"),
  "Test" = c("Welch's t-test", "Welch's t-test"),
  "T-score" = c(
    round(con_sex_ttest_pos$statistic,2),
    round(con_sex_ttest_neg$statistic,2)
  ),
  "95% CI" = c(
    paste0(
      "(",
      round(con_sex_ttest_pos$conf.int[[1]],2),
      ", ",
      round(con_sex_ttest_pos$conf.int[[2]],2),
      ")", sep = ""
    ),
    paste0(
      "(",
      round(con_sex_ttest_neg$conf.int[[1]],2),
      ", ",
      round(con_sex_ttest_neg$conf.int[[2]],2),
      ")", sep = ""
    )
  ),
  "df" = c(
    round(con_sex_ttest_pos$parameter,1),
    round(con_sex_ttest_neg$parameter,1)
  ),
  "p-value" = c(
    round(con_sex_ttest_pos$p.value, 3),
    round(con_sex_ttest_neg$p.value,3)
  )
)

knitr::kable(con_sex_tab, booktabs = TRUE, caption = "Statistical test of difference: Continuous

```

5.3.6 Figure 4.2

```
# scatter plot - interaction
ggplot(data = sim_data, aes(x = gen_id, y = outcome_num_pos_gid)) +
  coord_cartesian(ylim = c(0, 20)) + # ensures that the same y-axis length is used for
  geom_jitter(size = 3, colour = "forestgreen") +
  labs(
    title = 'Interaction Example',
    subtitle = 'Scatter plot of continuous outcome \nand gender identity (n=30)'
  ) +
  ylab("Continuous Outcome") + xlab("Feminine - Masculine Continuum") +
  theme_classic() +
  theme(
    legend.position = 'none',
    axis.ticks.x = element_blank(),
    axis.text.x = element_blank()
  )

# scatter plot - no interaction
ggplot(data = sim_data, aes(x = gen_id, y = outcome_num_neg)) +
  coord_cartesian(ylim = c(0, 20)) + # ensures that the same y-axis length is used for
  geom_jitter(size = 3, colour = "forestgreen") +
  labs(
    title = 'No Interaction Example',
    subtitle = 'Scatter plot of continuous outcome \nand gender identity (n=30)'
  ) +
  ylab("Continuous Outcome") + xlab("Feminine - Masculine Continuum") +
  theme_classic() +
  theme(
    legend.position = 'none',
    axis.ticks.x = element_blank(),
    axis.text.x = element_blank()
  )
)
```

5.3.7 Figure 4.3

```
# 2d density plot - interaction
ggplot(data = sim_data, aes(x = gen_id, y = outcome_num_pos_gid)) +
  geom_density_2d_filled(
    aes(fill = after_stat(level)), bins = 9, contour_var = 'ndensity'
  ) +
  scale_fill_brewer(
    palette = 16, direction = 1,
    labels = c('Low', '', '', '', '', '', '', '', 'High')
  ) +
```

```

guides(fill = guide_legend(
  title = 'Density of Occurences',
  direction = 'horizontal', nrow = 1,
  label.position = 'bottom')
) +
labs(
  title = 'Interaction Example',
  subtitle = '2-D Density plot of continuous outcome \nand gender identity (n=30)'
) +
ylab("Continuous Outcome") + xlab("Feminine - Masculine Continuum") +
theme_classic() +
theme(
  legend.position = 'none',
  axis.ticks.x = element_blank(),
  axis.text.x = element_blank()
) +
geom_blank(aes(y = 20)) + # extends y axis to ensure plots are comparable
geom_blank(aes(y = 0))

# 2d density plot - no interaction
ggplot(data = sim_data, aes(x = gen_id, y = outcome_num_neg)) +
  geom_density_2d_filled(
    aes(fill = after_stat(level)), bins = 9, contour_var = 'ndensity'
  ) +
  scale_fill_brewer(
    palette = 16, direction = 1,
    labels = c('Low', '', '', '', '', '', '', '', 'High')
  ) +
  guides(fill = guide_legend(
    title = 'Density of Occurences',
    direction = 'horizontal', nrow = 1,
    label.position = 'bottom')
  ) +
  labs(
    title = 'No Interaction Example',
    subtitle = '2-D Density plot of continuous outcome \nand gender identity (n=30)'
  ) +
  ylab("Continuous Outcome") + xlab("Feminine - Masculine Continuum") +
  theme_classic() +
  theme(
    legend.position = 'none',
    axis.ticks.x = element_blank(),
    axis.text.x = element_blank()
  ) +
  geom_blank(aes(y = 20)) + # extends y axis to ensure plots are comparable

```

```
geom_blank(aes(y = 0))
```

5.3.8 Table 4.5

```
# calculate correlations
con_gi_r_neg <- cor.test(sim_data$outcome_num_neg, sim_data$gen_id)
con_gi_r_pos <- cor.test(sim_data$outcome_num_pos_gid, sim_data$gen_id)

# output table of results
con_gi_tab <- tibble(
  "Example" = c("Interaction", "No interaction"),
  "Correlation Type" = c("Pearson's r", "Pearson's r"),
  "r" = c(
    round(con_gi_r_pos$estimate,2),
    round(con_gi_r_neg$estimate,2)
  ),
  "95% CI" = c(
    paste0(
      "(",
      round(con_gi_r_pos$conf.int[[1]],2),
      ", ",
      round(con_gi_r_pos$conf.int[[2]],2),
      ")", sep = ""
    ),
    paste0(
      "(",
      round(con_gi_r_neg$conf.int[[1]],2),
      ", ",
      round(con_gi_r_neg$conf.int[[2]],2),
      ")", sep = ""
    )
  ),
  "df" = c(
    round(con_gi_r_pos$parameter,1),
    round(con_gi_r_neg$parameter,1)
  ),
  "Significance Test" = c("t-test", "t-test"),
  "T-score" = c(
    round(con_gi_r_pos$statistic,2),
    round(con_gi_r_neg$statistic,2)
  ),
  "p-value" = c(
    round(con_gi_r_pos$p.value, 3),
    round(con_gi_r_neg$p.value,3)
  )
)
```



```
)
)

knitr::kable(con_gi_tab, booktabs = TRUE, caption = "Statistical test of difference: Continuous c
```

5.3.9 Figure 4.4

```
# scatter plot - interaction
ggplot(data = sim_data, aes(x = gen_id, y = outcome_num_pos_gid)) +
  coord_cartesian(ylim = c(0, 20)) + # ensures that the same y-axis length is used for both plots
  geom_jitter(size = 3, colour = "forestgreen") +
  labs(
    title = 'Gender Identity',
    subtitle = 'Scatter plot of continuous outcome \nand gender identity (n=30)'
  ) +
  ylab("Continuous Outcome") + xlab("Feminine - Masculine Continuum") +
  theme_classic() +
  theme(
    legend.position = 'none',
    axis.ticks.x = element_blank(),
    axis.text.x = element_blank()
  )

# scatter plot - no interaction
ggplot(data = sim_data, aes(x = gen_role, y = outcome_num_pos_gid)) +
  coord_cartesian(ylim = c(0, 20)) + # ensures that the same y-axis length is used for both plots
  geom_jitter(size = 3, colour = "forestgreen") +
  labs(
    title = 'Gendered Roles',
    subtitle = 'Scatter plot of continuous outcome \nand gendered roles (n=30)'
  ) +
  ylab("Continuous Outcome") + xlab("Feminine - Masculine Continuum") +
  theme_classic() +
  theme(
    legend.position = 'none',
    axis.ticks.x = element_blank(),
    axis.text.x = element_blank()
  )
)
```

5.3.10 Figure 4.5

```
# 2d density plot - interaction
ggplot(data = sim_data, aes(x = gen_id, y = outcome_num_pos_gid)) +
  geom_density_2d_filled(
```

```

    aes(fill = after_stat(level)), bins = 9, contour_var = 'ndensity'
  ) +
  scale_fill_brewer(
    palette = 16, direction = 1,
    labels = c('Low', '', '', '', '', '', '', '', 'High')
  ) +
  guides(fill = guide_legend(
    title = 'Density of Occurences',
    direction = 'horizontal', nrow = 1,
    label.position = 'bottom')
  ) +
  labs(
    title = 'Gender Identity',
    subtitle = '2-D Density plot of continuous outcome \nand gender identity (n=30)'
  ) +
  ylab("Continuous Outcome") + xlab("Feminine - Masculine Continuum") +
  theme_classic() +
  theme(
    legend.position = 'none',
    axis.ticks.x = element_blank(),
    axis.text.x = element_blank()
  ) +
  geom_blank(aes(y = 20)) + # extends y axis to ensure plots are comparable
  geom_blank(aes(y = 0))

# 2d density plot - no interaction
ggplot(data = sim_data, aes(x = gen_role, y = outcome_num_pos_gid)) +
  geom_density_2d_filled(
    aes(fill = after_stat(level)), bins = 9, contour_var = 'ndensity'
  ) +
  scale_fill_brewer(
    palette = 16, direction = 1,
    labels = c('Low', '', '', '', '', '', '', '', 'High')
  ) +
  guides(fill = guide_legend(
    title = 'Density of Occurences',
    direction = 'horizontal', nrow = 1,
    label.position = 'bottom')
  ) +
  labs(
    title = 'Gendered Roles',
    subtitle = '2-D Density plot of continuous outcome \nand gendered roles (n=30)'
  ) +
  ylab("Continuous Outcome") + xlab("Feminine - Masculine Continuum") +
  theme_classic() +

```

```

theme(
  legend.position = 'none',
  axis.ticks.x = element_blank(),
  axis.text.x = element_blank()
) +
geom_blank(aes(y = 20)) + # extends y axis to ensure plots are comparable
geom_blank(aes(y = 0))

```

5.3.11 Table 4.6

```

# calculate correlations
con_gi_r_pos <- cor.test(sim_data$outcome_num_pos_gid, sim_data$gen_id)
con_gr_r_pos <- cor.test(sim_data$outcome_num_pos_gid, sim_data$gen_role)

# output table of results
con_2gah_tab <- tibble(
  "Example" = c("Gender Identity", "Gendered Roles"),
  "Correlation Type" = c("Pearson's r", "Pearson's r"),
  "r" = c(
    round(con_gi_r_pos$estimate,2),
    round(con_gr_r_pos$estimate,2)
  ),
  "95% CI" = c(
    paste0(
      "(",
      round(con_gi_r_pos$conf.int[[1]],2),
      ", ",
      round(con_gi_r_pos$conf.int[[2]],2),
      ")", sep = ""
    ),
    paste0(
      "(",
      round(con_gr_r_pos$conf.int[[1]],2),
      ", ",
      round(con_gr_r_pos$conf.int[[2]],2),
      ")", sep = ""
    )
  ),
  "df" = c(
    round(con_gi_r_pos$parameter,1),
    round(con_gr_r_pos$parameter,1)
  ),
  "Significance Test" = c("t-test", "t-test"),
  "T-score" = c(

```

```
    round(con_gi_r_pos$statistic,2),
    round(con_gr_r_pos$statistic,2)
  ),
  "p-value" = c(
    round(con_gi_r_pos$p.value, 3),
    round(con_gr_r_pos$p.value,3)
  )
)

knitr::kable(con_2gah_tab, booktabs = TRUE, caption = "Statistical test of difference:
```