Norming Data

Bethany Gardner

3/30/2022

Contents

For the norming study, 51 participants on MTurk rated 92 first names on a scale from 1 being "very masculine" to 7 being "very feminine." The masculine and feminine names were selected from the top 100 names according to US census data:

> United States Social Security Administration. (2019). *Top names over the last 100 years* [Data Set]. United States Social Security Administration. https://www.ssa.gov/oact/babynames/decades/century.html

The androgynous names were selected from a list of names that were given at least one-third of the time to AFAB children in the US and also at least one-third of the time to AMAB children.

> Flowers, A. (2015). *Unisex names data* [Data Set]. FiveThirtyEight. https://github.com/fivethirtyeight/data/tree/master/unisex-names

```r
all_ratings <- read.csv("../data/exp0_data_norming.csv", stringsAsFactors=TRUE) %>%
  #don't need participant demographics right now
  select(-gender, -age, -race, -education, -english) %>%
  #pivot to have one row per name, not one column per name
  pivot_longer(cols = c(-ResponseId),
               names_to = "Name",
               values_to = "GenderRating")
```

Mean and SD of gender ratings for each name, sorted from most feminine to most masculine.

```r
mean_ratings <- all_ratings %>% group_by(Name) %>%
  summarise(MeanGenderRating=mean(GenderRating),
            SD=sd(GenderRating)) %>%
  arrange(desc(MeanGenderRating))

kable(mean_ratings)
```

Name	MeanGenderRating	SD
Emily	6.823529	0.7129062
Lisa	6.823529	0.7129062
Sarah	6.823529	0.5901146

Name	MeanGenderRating	SD
Amanda	6.803922	0.8250965
Barbara	6.803922	0.4906978
Elizabeth	6.803922	0.6639159
Laura	6.803922	0.8250965
Melissa	6.803922	0.7216539
Nancy	6.803922	0.7216539
Sandra	6.803922	0.6639159
Rebecca	6.784314	0.8321953
Susan	6.784314	0.6727176
Deborah	6.764706	0.7896388
Jennifer	6.764706	0.8387666
Donna	6.745098	0.7167465
Mary	6.745098	0.8448228
Stephanie	6.745098	0.8681737
Jessica	6.725490	0.8019584
Patricia	6.725490	0.6656856
Sharon	6.725490	0.8019584
Dorothy	6.705882	0.8073195
Kimberly	6.705882	0.7292220
Betty	6.686274	0.8364256
Margaret	6.686274	0.7871517
Michelle	6.686274	0.8829540
Cynthia	6.666667	0.9309493
Carol	6.607843	1.0784884
Karen	6.549020	1.2379616
Linda	6.509804	1.1553796
Ashley	6.274510	1.1327565
Elisha	5.882353	1.7960742
Jody	5.588235	1.2029376
Jackie	5.274510	1.1327565
Blair	5.215686	1.5008494
Kerry	4.725490	1.2661506
Amari	4.666667	1.4094916
Skyler	4.666667	1.4375906
Sage	4.647059	1.4943029
Carey	4.588235	1.5514699
Kendall	4.529412	1.8585257
Jessie	4.372549	1.2483715
Justice	4.352941	1.4117157
Riley	4.352941	1.3239868
Jaime	4.333333	1.1944315
Taylor	4.235294	1.1240682
Casey	4.196078	0.9801961
Harley	4.098039	1.6401817
Emery	4.058823	1.6298683
Kris	4.058823	1.2870395
Avery	4.000000	1.8867962
Reese	3.862745	1.6494800
Pat	3.784314	1.3610838
Quinn	3.725490	1.5757973
Peyton	3.509804	1.5016331
Stevie	3.176471	1.5060662

Name	MeanGenderRating	SD
Frankie	2.960784	1.3558877
Rowan	2.705882	1.5400535
Emerson	2.607843	1.4153223
Tommie	2.450980	1.6408988
Robbie	2.254902	1.4260875
Ollie	2.196078	1.3859236
Chris	2.156863	1.2549510
Ryan	1.647059	1.1103788
Michael	1.411765	0.8526774
Daniel	1.392157	0.9397538
Paul	1.352941	1.1458365
Thomas	1.333333	0.8640988
Donald	1.294118	0.8554325
James	1.294118	0.6097251
William	1.294118	0.8317239
Charles	1.274510	0.7766191
Jason	1.274510	0.8735773
Joseph	1.274510	0.7504247
Steven	1.274510	0.8265212
Christopher	1.254902	0.7705358
Jeffery	1.254902	0.7705358
Kenneth	1.254902	0.8208078
Robert	1.254902	0.8448228
Timothy	1.254902	0.8448228
Brian	1.235294	0.7372445
David	1.235294	0.6808299
Joshua	1.235294	0.8622815
Andrew	1.215686	0.7297596
Anthony	1.215686	0.8078051
George	1.215686	0.8321953
John	1.215686	0.5766706
Kevin	1.215686	0.8321953
Mark	1.215686	0.6727176
Edward	1.196078	0.8004900
Matthew	1.196078	0.7216539
Richard	1.196078	0.6330753
Ronald	1.196078	0.7751028

Selected 21 names from these results, with 3 names around each of the 7 intervals.

```
names_used <- mean_ratings %>% filter(str_detect(Name,
  "Matthew|Brian|James|Chris|Tommie|Emerson|Stevie|Quinn|Reese|Taylor|Riley|Jessie|Kerry|Blair|Jackie|J
  filter(Name!="Christopher")

kable(names_used, digits=2)
```

Name	MeanGenderRating	SD
Emily	6.82	0.71
Rebecca	6.78	0.83
Mary	6.75	0.84

Name	MeanGenderRating	SD
Ashley	6.27	1.13
Elisha	5.88	1.80
Jody	5.59	1.20
Jackie	5.27	1.13
Blair	5.22	1.50
Kerry	4.73	1.27
Jessie	4.37	1.25
Riley	4.35	1.32
Taylor	4.24	1.12
Reese	3.86	1.65
Quinn	3.73	1.58
Stevie	3.18	1.51
Emerson	2.61	1.42
Tommie	2.45	1.64
Chris	2.16	1.25
James	1.29	0.61
Brian	1.24	0.74
Matthew	1.20	0.72

To check to see if the norming data were biased to call names more masculine, I compared them to the US census data for gender assigned at birth.

United States Social Security Administration. (2020). *Beyond the top 1000 names* [Data Set]. United States Social Security Administration. https://www.ssa.gov/oact/babynames/limits.html

The norming study is on a scale from 1-7, and the census scale is probability 0-1. To try to compare this, I first subtracted 1 from the norming data, to put it on a scale from 0-6. Then, I divided by six, to put it on a scale from 0-1.

```r
census <- read.csv("../data/exp0_data_census.csv")

names_used <- left_join(names_used, census, by="Name") %>%
  mutate(MeanGenderRating06=MeanGenderRating-1,
  Norming_ProbFemale = MeanGenderRating06 / 6,
  Diff_ProbFemale = Census_ProbFemale - Norming_ProbFemale)
```

A few of the androgynous names have bigger discrepancies, likely because their gender associations have been changing over time. Overall, though, the mean difference is close to 0, and not all of the differences involve the norming data over-estimating the masculinity of a name.

```r
summary(names_used$Diff_ProbFemale)
```

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## -0.302257 -0.054896  0.027306 -0.003562  0.094850  0.202474
```

Calculate the correlation:

```r
cor.test(names_used$Norming_ProbFemale, names_used$Census_ProbFemale)
```

```
## 
##  Pearson's product-moment correlation
## 
## data:  names_used$Norming_ProbFemale and names_used$Census_ProbFemale
## t = 10.118, df = 19, p-value = 4.359e-09
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.8064177 0.9667888
## sample estimates:
##       cor
## 0.9183935
```

And visualize it:

```
plot_correlation <- ggplot(names_used, aes(x=Norming_ProbFemale, y=Census_ProbFemale,
                       color=Name, label=Name)) +
  geom_point(size=2.5, show.legend=FALSE) +
  geom_smooth(method=lm, color="darkgrey", fill="darkgrey",
              se=FALSE, show.legend=FALSE) +
  geom_text_repel(show.legend=FALSE) +
  coord_cartesian(xlim=c(-.05,1.05), ylim=c(-.05, 1.05)) +
  theme_classic() +
  theme(text=element_text(size=16)) +
  labs(title="Norming Study",
       x="Proportion Feminine in Norming Data",
       y="Proportion AFAB in Census Data")
plot_correlation
```

```
## `geom_smooth()` using formula 'y ~ x'
```

Norming Study