

Using Beta Regression to Better Model Norms in Political Psychology

Mark H. White II / markhwhiteii@gmail.com

2017-12-11

Louis, Mavor, & Terry (2003) note how typical correlational studies of ideology, politics, and prejudice often underestimate or overlook the role of norms. Their basic argument is that researchers analyze *variance* in common statistical procedures. However, strong norms lead to *invariance*—people doing the same thing (or, in our case, responding to survey items in the same way). Norms can lead to skewed distributions: floor effects (strong norm in disagreement with a survey item), ceiling effects (strong norm for agreement), and/or strongly leptokurtic distributions (strong norm for answering the midpoint). Simply reporting a correlation may misrepresent the results to readers, as I will show here.

A few years later, Smithson & Verkuilen (2006) discussed dealing with these issues using beta regression. The beta distribution is bounded at 0 and 1, so beta regression has often been used for dealing with proportions, rates, etc. However, they note that many (if not most) of the instruments used in psychology (and other fields that rely on self-reports on Likert scales) are inherently bounded between two values. For instance, a typical 1 (*Strongly Disagree*) to 7 (*Strongly Agree*) scale has bounds at 1 and 7. Typical ordinary least squares regression assumes that values can go outside of this 1 to 7 range. Most of the time this is fine, but when there are strong non-normalities like ceiling and floor effects in the distribution of a response variable, it can be helpful to use a flexible model that can handle those irregularities.

I want to focus on the extra insights we learn from using beta regression, so I will pay particularly close attention in this post to Louis et al.'s focus on norms. Smithson & Verkuilen mention this paper in passing, but I want to more fully dive into an example here.

First, I will introduce beta regression. Second, I will introduce data and show what ordinary least squares might tell us—and how this could be misleading. Third, I will demonstrate how to use beta regression on these data in a frequentist paradigm. Lastly, I'll show how to model these data in a Bayesian paradigm using Stan, focusing on what beta regression can tell us about norms.

Beta Regression

The Beta Distribution

The beta distribution can take many shapes. For example, see below Figure 1 from Cribari-Neto & Zeileis (2010).

The beta distribution has two parameters that determine its shape. One of the parameters pulls the density toward zero, and the other pulls it toward one. Let's call them p and q . Ferrari & Cribari-Neto (2004) showed a different parameterization of the beta density. In particular, they set a parameter called μ to be equal to $\frac{p}{(p+q)}$. This is the expected value (in this case, the mean) of the variable we are modeling—let's call that variable y . This is known as the “location parameter,” because it tells us where we expect the location of the variable to be.

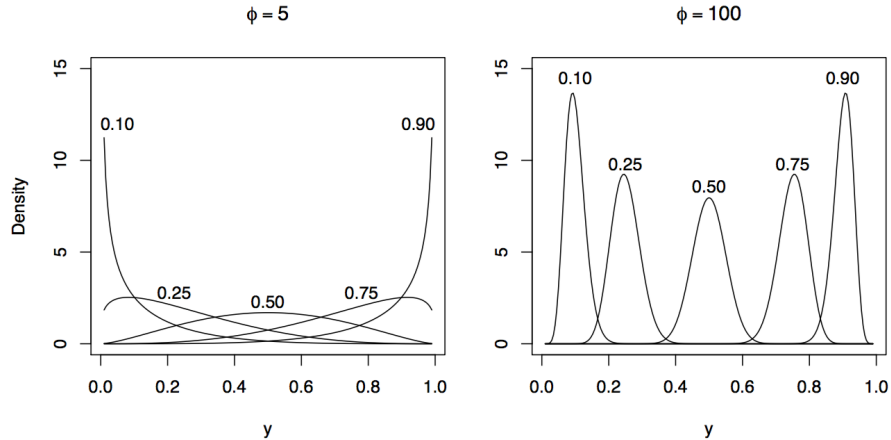


Figure 1: Probability density functions for beta distributions with varying parameters $\mu = 0.10, 0.25, 0.50, 0.75, 0.90$ and $\phi = 5$ (left) and $\phi = 100$ (right).

They set another parameter called ϕ to be equal to $p + q$. The variance of y is defined as $\frac{\mu(1+\mu)}{(1+\phi)}$. There are two important things to note here. First, ϕ is on the bottom of this equation. That means that the *larger* that ϕ gets, the *smaller* the variance of y is. For this reason, it is known as the “precision parameter.” Second, μ is included in this equation. This will mean that, when we get to beta regression, the model will be naturally *heteroskedastic*. That is, the variance will depend on the mean. Contrast this against ordinary least squares regression, where we assume homogeneity of variance.

Note that the beta density can then be “reparameterized” in terms of μ and ϕ by doing some algebra on the equations above and solving for p and q : $p = \mu\phi$ and $q = \phi - \mu\phi$. We will see in the Stan code how this is helpful for us.

The Beta Regression Model

The goal of beta regression is then to predict parameters μ and ϕ from some predictors we have. For this case, let’s just say we have one predictor named x . We can write the equations (for now) just like in ordinary least squares regression: $\mu_i = \beta_0 + \beta_1 X_i$ and $\phi_i = \beta_2 + \beta_3 X_i$.

However, there are two issues with these equations. First, μ has to be between 0 and 1, since it is assumed to be a beta-distributed outcome variable in beta regression. Just using that equation above doesn’t restrict μ to that range. What we can do, though, is apply a link function that smushes μ into that range. Many can be used, but I will use the logit (i.e., log-odds) link function. Thus, we have: $\mu_i = \frac{\exp(\beta_0 + \beta_1 X_i)}{1 + \exp(\beta_0 + \beta_1 X_i)}$. (This should look familiar to those who know logistic regression, since it also uses the logit link function). Second, ϕ has to be greater than zero since a variance cannot be negative. To do this, we can use the log link function. Thus, we have: $\phi_i = \exp(\beta_2 + \beta_3 X_i)$.

This creates two “submodels,” using the terminology of Smithson & Verkuilen. The first, for *mu*, is the “location submodel.” Predictors in this model tell us about where the expected values of y are (i.e., the location) for given predictor values (in this case, just x). The second, for *phi*, is the “dispersion submodel.” Predictors in this model tell us about what the variance of the dependent variable is (i.e., the dispersion) for given predictor values.

The Current Data: Conservatism and Social Dominance Orientation

So why does all of this matter? Let’s revisit Louis, Mavor, & Terry (2003) real quick. They examine the bivariate relationship between right-wing authoritarianism and how much people agree with the item “The White race is the best race.” The bounded nature of 7-point scales comes out here. 72% of the sample

selected “1, *Strongly Disagree*” on this item. However, what often gets reported is just the correlation: $r = .34$, $p < .001$. We learn that right-wing authoritarianism predicts prejudice, but that’s about it. What are we missing when this is all we see in the results section?

Louis et al. note that, for 72% of the sample, a correlation could not even be calculated. This was because there was no variance in the scores: Everyone adhered to the egalitarian norm that it is not OK to openly say the White race is the best race. For the rest of the sample, the correlation was .19. The authors worry about two things: First, that researchers could decide to exclude items like this, because they aren’t normally distributed; and second, researchers only report the correlation, and readers interpret it symmetrically. A “symmetric” interpretation is that people at *both* the high and low ends of the authoritarianism spectrum are driving the relationship, when we know that is not the case. I argue that beta regression can better describe relationships like these.

The rest of this post is aimed to be both (a) an example of when beta regression is useful and (b) how to do it using R and Stan.

The data analyzed here are taken from 175 participants from MTurk. The data come from Study 1 of White & Crandall (2017). The independent variable is a combination of two variables: identifying as conservative and Republican. The dependent variable is social dominance orientation, or how much people willingly endorse hierarchy and inequality in society (example item: “Inferior groups should stay in their place”). In the context of the study—which was about racist incidents on college campuses—people could see this as being about racial groups. There is a highly-replicated positive relationship between right-wing political identification (RWID) and social dominance orientation (SDO). What else can we learn about this relationship from these data?

The Ordinary Least Squares Approach

First, let’s prepare the data and look at its structure:

```
library(ggplot2)
library(betareg)
library(rstan)
library(bayesplot)
dat <- read.csv("betareg_politics.csv")
str(dat)

## 'data.frame': 175 obs. of 2 variables:
## $ sdo : num -0.9053 0.0089 1.1844 -0.9053 -0.1217 ...
## $ rw_pol_id: num 0.127 0.127 0.127 0.127 0.127 ...
```

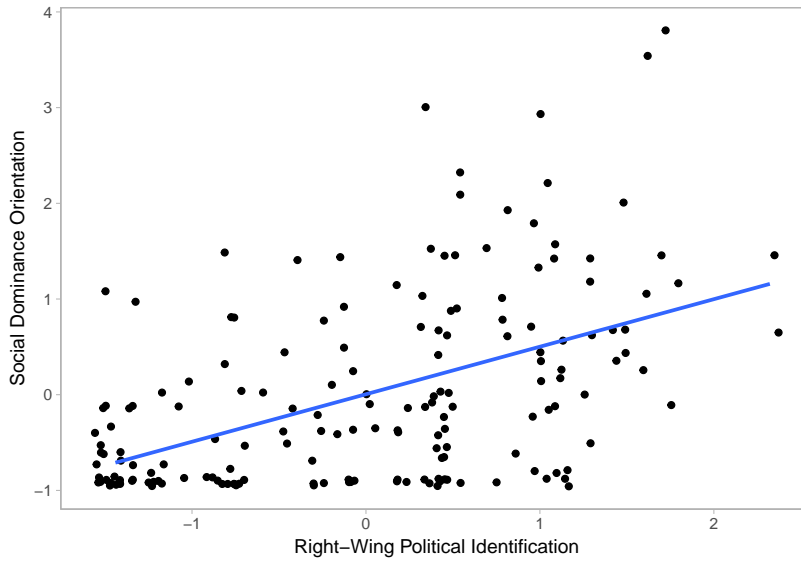
There are two variables, both z -scored. What is the correlation between the two?

```
cor(dat)[[2]]

## [1] 0.4962892
```

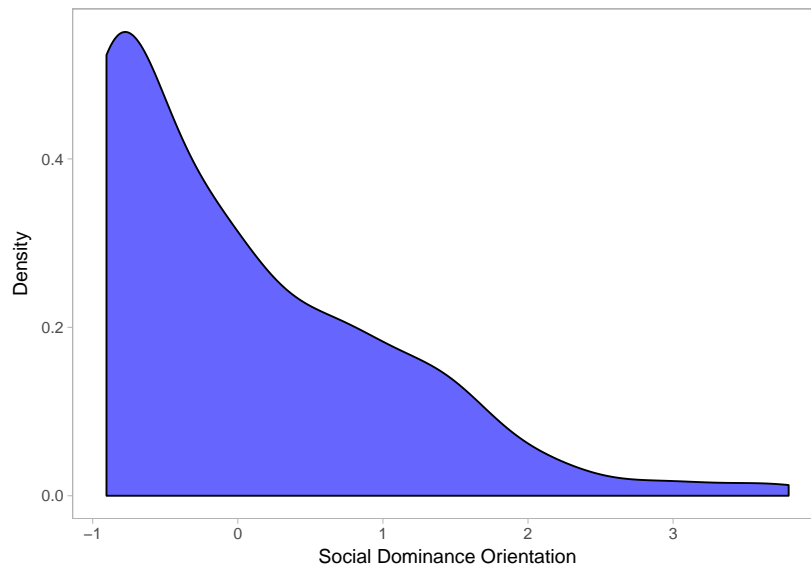
So, we have replicated past research. RWID predicts SDO, $r = .50$, $p < .001$. But what else is going on? Let’s take a look at a visual of the linear relationship:

```
ggplot(dat, aes(x = rw_pol_id, y = sdo)) +
  geom_jitter() +
  geom_smooth(method = "lm", se = FALSE) +
  theme_light() +
  labs(x = "Right-Wing Political Identification",
       y = "Social Dominance Orientation")
```



There's a linear relationship, but look at all of those values hitting the floor of SDO. What does the density of this variable look like?

```
ggplot(dat, aes(x = sdo)) +
  geom_density(fill = "blue", alpha = .6) +
  theme_light() +
  labs(x = "Social Dominance Orientation",
       y = "Density")
```



There you have it: A huge floor effect. Let's move on to beta regression to see if it can tell us something different.

The Frequentist Beta Regression Approach

With beta regression, or dependent variable has to be *between* 0 and 1. Thankfully, Smithson & Verkuilen (2006) show a way to transform our variable so that the highest value is *just* below 1 and the lowest value is *just* above 0. Here is a function that implements the transformation:

```

beta_normalize <- function(x) {
  x_ <- ((x - min(x)) / (max(x) - min(x)))
  (x_ * (length(x_) - 1) + 0.5) / length(x_)
}
dat$sdo <- beta_normalize(dat$sdo)
range(dat$sdo)

```

```
## [1] 0.002857143 0.997142857
```

Perfect. Beta regression is implemented easily using the `betareg` package. In a `lm` or `glm`, you would specify $y \sim x$ (i.e., your dependent variable is predicted by your independent variable). The only change here is that there is a bar in the formula: `|`. Whatever comes before the bar are variables predicting the μ part (i.e., location); whatever comes after are variables predicting the ϕ part (i.e., dispersion). These variables can be all the same, all different, or anywhere in between. I will use RWID to predict SDO in both submodels:

```

mod1 <- betareg(sdo ~ rw_pol_id | rw_pol_id,
               data = dat, link = "logit",
               link.phi = "log")

```

Note that I specify the link functions that I mentioned above. The results:

```
summary(mod1)
```

```

##
## Call:
## betareg(formula = sdo ~ rw_pol_id | rw_pol_id, data = dat, link = "logit",
##        link.phi = "log")
##
## Standardized weighted residuals 2:
##      Min      1Q  Median      3Q      Max
## -2.2562 -0.5157  0.1728  0.6876  3.1761
##
## Coefficients (mean model with logit link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.58477    0.10001 -15.847 < 2e-16 ***
## rw_pol_id    0.71500    0.09891   7.229 4.87e-13 ***
##
## Phi coefficients (precision model with log link):
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.1544    0.1128  10.231 < 2e-16 ***
## rw_pol_id   -0.3524    0.1092  -3.226 0.00126 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Type of estimator: ML (maximum likelihood)
## Log-likelihood: 177.1 on 4 Df
## Pseudo R-squared: 0.2303
## Number of iterations: 10 (BFGS) + 2 (Fisher scoring)

```

First, let's look at the **mean model with logit link**. This is what is predicting the location (μ). We can see that, replicating the zero-order correlation above, the more someone identifies as conservative and Republican, the more they endorse inequality, dominance, and hierarchy in society.

What I think is really cool is the second part, the **precision model with log link**. This tells about ϕ . We can see that precision of SDO decreases as RWID increases. In other words, the variance of SDO increases as RWID increases. Before I turn to interpretation, let's take a look at how to do this in a Bayesian framework,

using Stan. I like it because it requires us to be a little bit more explicit and hands-on in running the model. Plus, Bayesian inferences are more intuitive.

The Bayesian Beta Regression Approach

I am interfacing with Stan using `rstan`. I define my model:

```
stan_code <- "  
data {  
  int n;  
  vector[n] x;  
  vector<lower=0, upper=1>[n] y;  
}  
parameters {  
  vector[4] coef;  
}  
transformed parameters {  
  vector<lower=0, upper=1>[n] mu;  
  vector<lower=0>[n] phi;  
  vector[n] p;  
  vector[n] q;  
  for (i in 1:n) {  
    mu[i] = inv_logit(coef[1] + coef[2] * x[i]);  
    phi[i] = exp(coef[3] + coef[4] * x[i]);  
    p[i] = mu[i] * phi[i];  
    q[i] = phi[i] - mu[i] * phi[i];  
  }  
}  
model {  
  y ~ beta(p, q);  
  coef ~ normal(0, 2);  
}  
"
```

In the first block, I read in the data: my total number of participants n as well as my independent x and dependent y variables. In the second, I tell Stan that I am going to be estimating four coefficients (called `coef`)—these are two intercepts and two slopes (for both the location and precision submodels). Then, I tell Stan that I will be defining `mu` and `phi`. As you can see, these are predicted by intercepts (`coef[1]` and `coef[3]`) and the slopes of x (`coef[2]` and `coef[4]`). I use the logit and log link functions by calling `inv_logit` and `exp`, respectively. I also tell Stan that I'm going to define `p` and `q`, which are the shape parameters for the beta density. Note that I use the equations for p and q that I defined above—I'm simply taking the equations from the papers and plopping them in here.

Lastly, I define my model. I tell Stan that y is distributed `beta` with the shape parameters `p` and `q`. I then set somewhat informative priors on my coefficients. Before the fact, I think that my coefficients probably lie in a normal distribution with a mean of 0 and a standard deviation of 2.

Let's run this model:

```
stan_dat <- list(n = nrow(dat), x = dat$rw_pol_id, y = dat$sdo)  
set.seed(1839)  
mod2 <- stan(model_code = stan_code, data = stan_dat,  
            iter = 1000, chains = 4, cores = 4)
```

And look at the results:

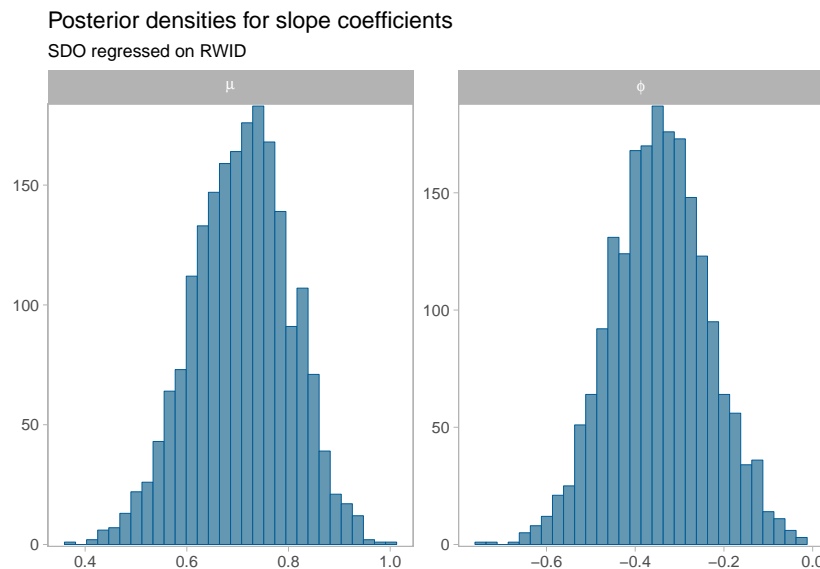
```
round(summary(mod2)$summary[1:4, ], 2)
```

```
##          mean se_mean  sd 2.5% 25% 50% 75% 97.5% n_eff Rhat
## coef[1] -1.57      0 0.10 -1.77 -1.64 -1.57 -1.51 -1.37 1008.51  1
## coef[2]  0.71      0 0.10  0.51  0.64  0.71  0.77  0.89 1267.63  1
## coef[3]  1.13      0 0.11  0.91  1.06  1.13  1.21  1.34 1147.24  1
## coef[4] -0.34      0 0.11 -0.56 -0.42 -0.35 -0.27 -0.13 1318.23  1
```

`coef[2]` agrees with the frequentist estimates using `betareg`: RWID positively predicts SDO. The coefficient here is .71, while it was .72 using `betareg`. However, we get an intuitive inference: It is 95% probable that our estimate lies somewhere between 0.51 and 0.89. `coef[4]` tells the same story, too: RWID predicts greater variance in SDO.

But if we want to think Bayesian, we don't want to just look at the point estimates. Let's look at the entire *distribution* of the posterior for these coefficients:

```
draws <- as.matrix(mod2)
draws <- draws[, c("coef[2]", "coef[4]")]
colnames(draws) <- c("mu", "phi")
mcmc_hist(draws, facet_args = list(labeler = label_parsed)) +
  ggtitle("Posterior densities for slope coefficients",
          "SDO regressed on RWID") +
  theme_light()
```

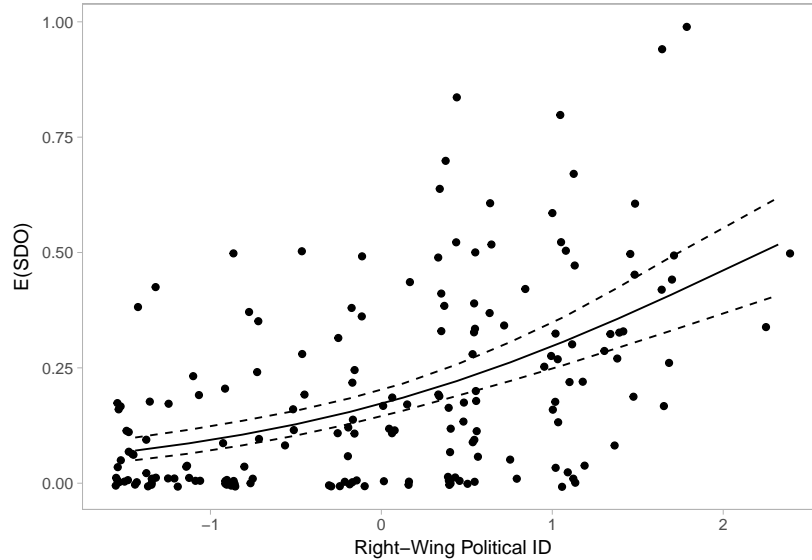


This shows histograms for draws from the posterior distributions for the slope coefficients predicting μ and ϕ from RWID, respectively (that is, `coef[2]` and `coef[4]`). Note that neither of these distributions cross zero, so we can infer that it is greater than 99% probable that `coef[2]` is larger than zero and 99% probable that `coef[4]` is smaller than zero.

Now, let's look at the prediction line for the μ , the expected value of SDO (i.e., $E(SDO)$). The dotted line represents the 95% credible interval for each of the points on the predicted line.

```
mod2_summary <- summary(mod2)$summary
mus <- mod2_summary[grepl("mu", rownames(mod2_summary)), "mean"]
mus_lo <- mod2_summary[grepl("mu", row.names(mod2_summary)), 4]
mus_hi <- mod2_summary[grepl("mu", row.names(mod2_summary)), 8]
ggplot() +
  geom_jitter(aes(x = dat$rw_pol_id, y = dat$sdo)) +
```

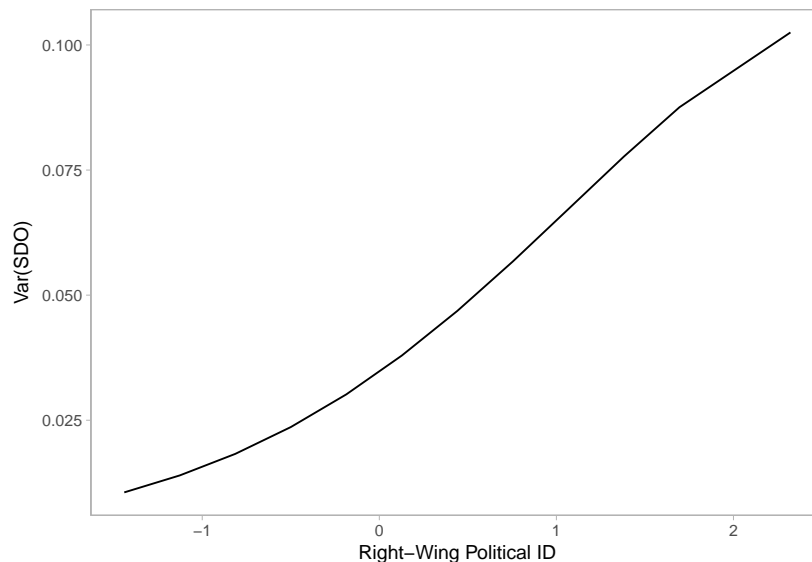
```
geom_line(aes(x = dat$rw_pol_id, y = mus)) +
geom_line(aes(x = dat$rw_pol_id, y = mus_lo), linetype = 2) +
geom_line(aes(x = dat$rw_pol_id, y = mus_hi), linetype = 2) +
theme_light() +
labs(x = "Right-Wing Political ID", y = "E(SDO)")
```



It has a slight bend to it, showing the trend isn't quite linear. We can also see that the uncertainty of our estimates increases as the predicted value increases.

Again, what I think is really cool is that we can also plot how the *variance* of SDO increases as RWID increases. I solve for the variance using the equation for $Var(Y)$ above:

```
phis <- mod2_summary[grep("phi", rownames(mod2_summary)), "mean"]
vars <- (mus * (1 - mus)) / (phis + 1)
ggplot() +
  geom_line(aes(x = dat$rw_pol_id, y = vars)) +
  theme_light() +
  labs(x = "Right-Wing Political ID", y = "Var(SDO)")
```



What It Means For Norms

Norms produce invariance. It is not seen as OK to admit that some groups are inferior to others, that those groups should just stay in their place, and that equality should *not* be a big priority in society. Therefore, the participants displayed a floor effect: About half of the sample responded to the *entire* scale by answering all “1, *Strongly Disagree*” (or “7, *Strongly Agree*,” if the items were reverse-scored).

In typical personality psychology studies, this normativity in responding is either ignored or items are thrown out because they aren’t normally-distributed. But using beta regression, we can actually quantify this change in variance by looking at the dispersion submodel and the ϕ parameter. It isn’t just that “conservatives endorse social inequality more,” it is also a *normative* phenomenon: Since (a) norms produce invariance, and (b) there is greater variance with increasing conservatism, it appears that (c) as conservatism increases, people do not adhere to the norms about forbidding what is socially acceptable to express. People at the low end of RWID stuck to the norm, producing invariance around the floor; people at high end of RWID did not stick to the norm and displayed more variance.

Louis, Mavor, & Terry (2003) lamented that there is “no conventional analysis of invariance” (p. 192), and they gave no concrete suggestions for how else to analyze data to better get at the issues they were raising. I hope that this demonstration shows that beta regression, introduced the year after their paper was published, can address the problems they pointed out.

References

- Cribari-Neto, F. & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, 34.
- Ferrari, S. L. P. & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics*, 31.
- Louis, W. R., Mavor, K. I., & Terry, D. J. (2003). Reflections on the statistical analysis of personality and norms in war, peace, and prejudice: Are deviant minorities the problem? *Analyses of Social Issues and Public Policy*, 3.
- Smithson, M. & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods*, 11.
- White, M. H., II., & Crandall, C. S. (2017). Freedom of racist speech: Ego and expressive threats. *Journal of Personality and Social Psychology*, 113.