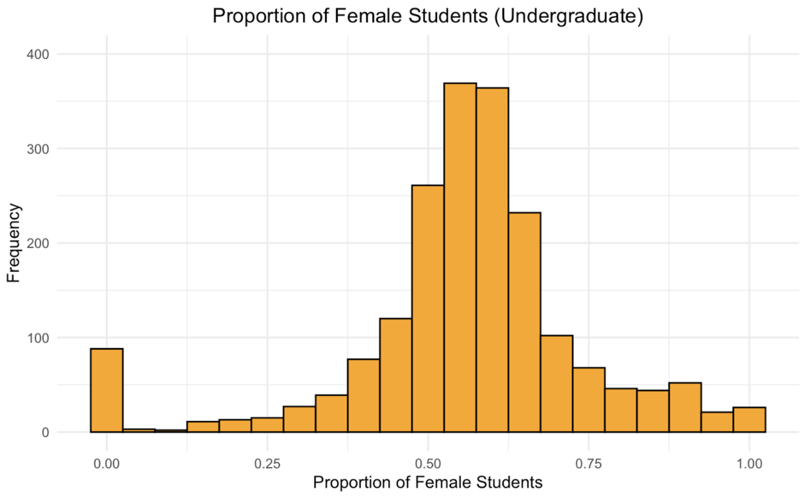
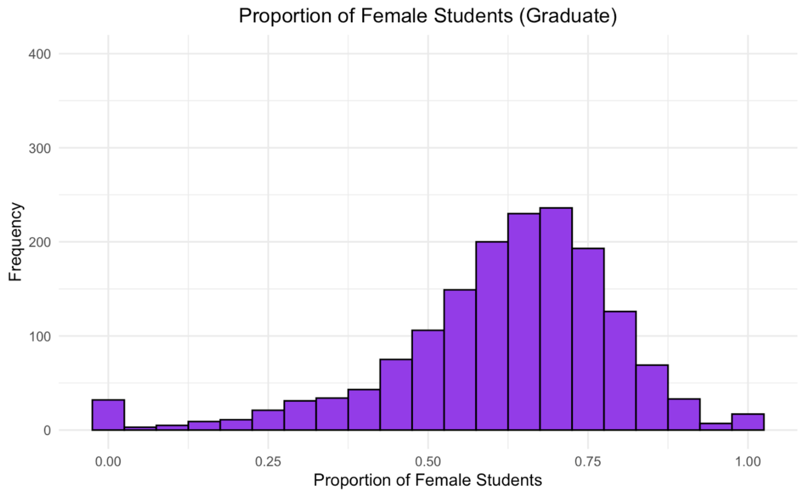
Beta regression

When to use

- Slider scales (opinions, confidence ratings)
- Proportional data (% of women in academia)

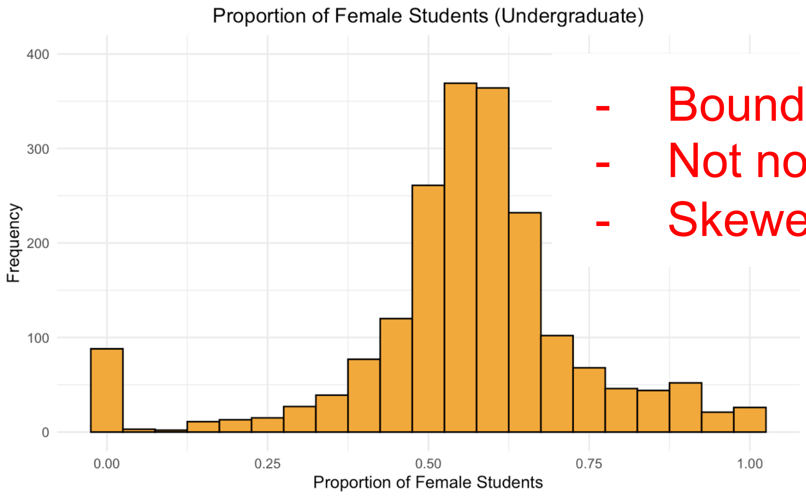




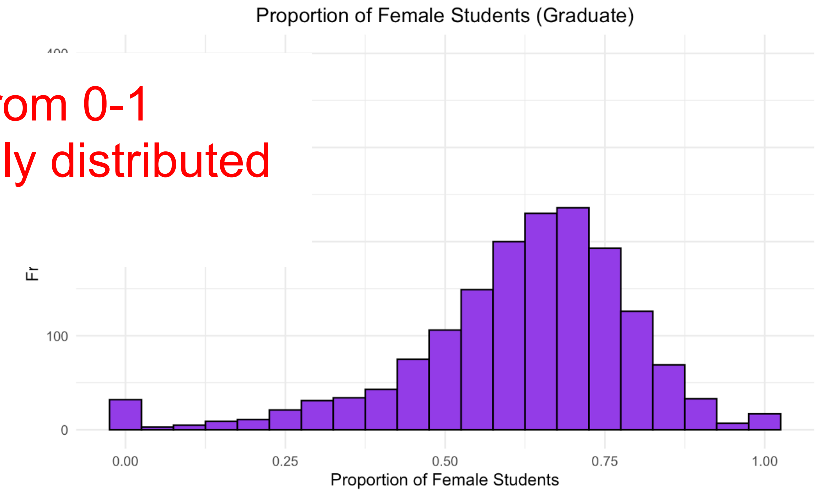
Data from the National Center for Education Statistics

When to use

- Slider scales (opinions, confidence ratings)
- Proportional data (% of women in academia)



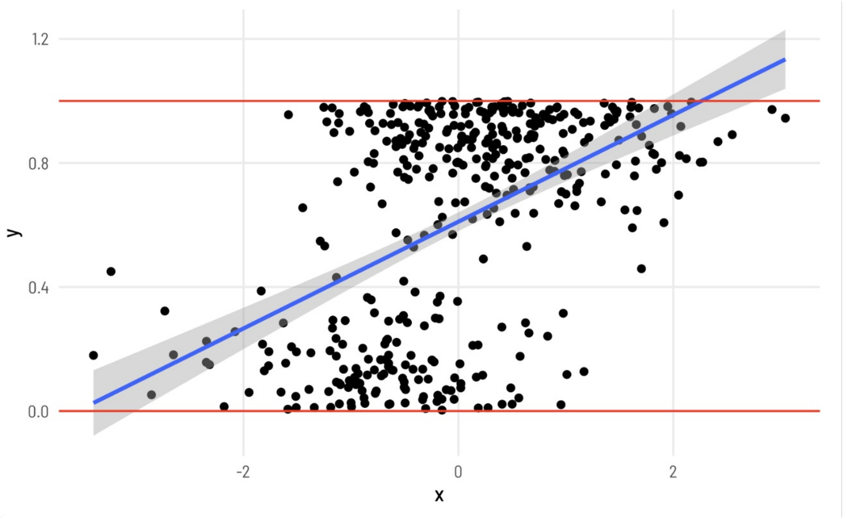
- Bounded from 0-1
- Not normally distributed
- Skewed



Data from the National Center for Education Statistics

Solutions

- Linear probability model
  - Just ordinary least squares (OLS) applied to a proportional outcome
  - Commonly used, but estimates can go out of the boundaries.

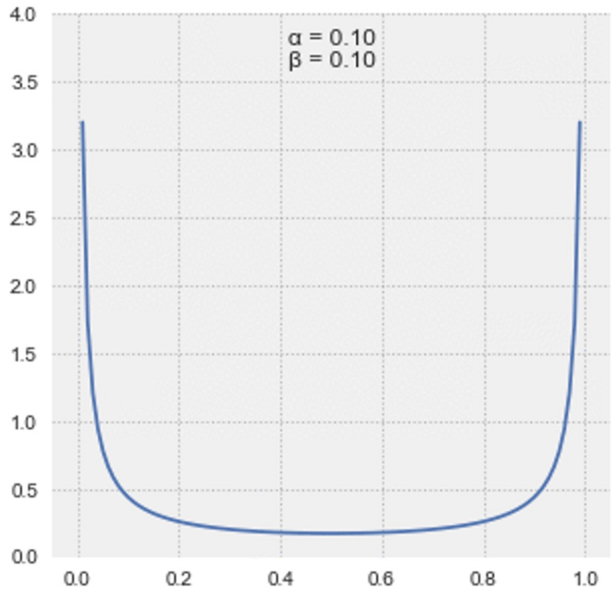


- Fractional logistic model
  - `glm(..., family = binomial(link = "logit"))`
  - Warning: using family = binomial() with a non-binary outcome variable

Beta regression

Beta - the use of the **Beta distribution** for the response variable

- Naturally limited to numbers between 0 and 1 (but doesn't include 0 or 1).
- Extremely flexible distribution and can take all sorts of different shapes and forms
- A probability distribution of probabilities

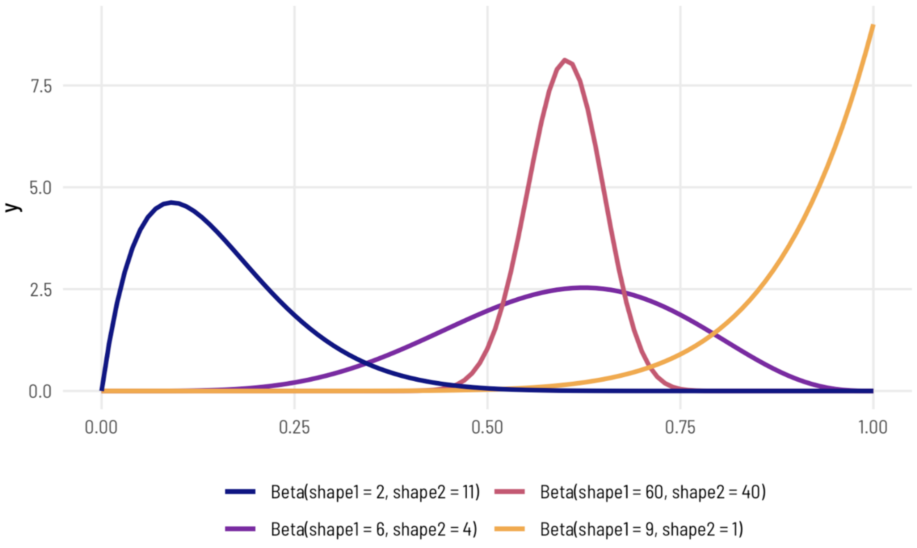


Beta distribution

- Two parameters: a/shape1, b/ shape 2
  - E.g., a = number of women, b = number of men

- Mean $$\mu = \frac{a}{a + b}$$

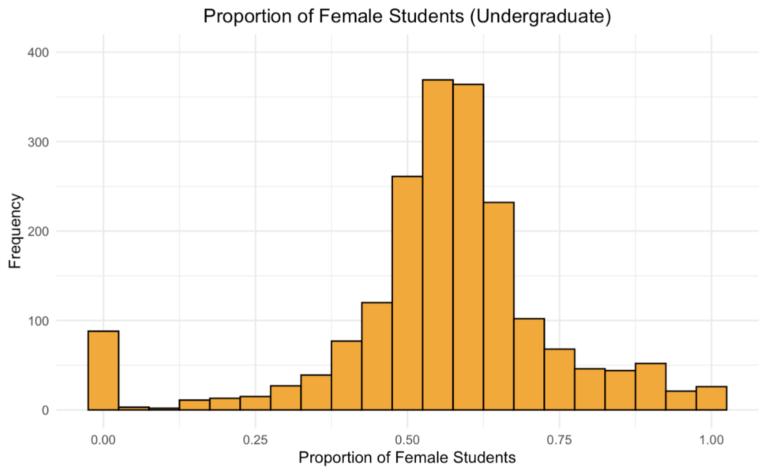
- Precision $$\phi = a + b$$

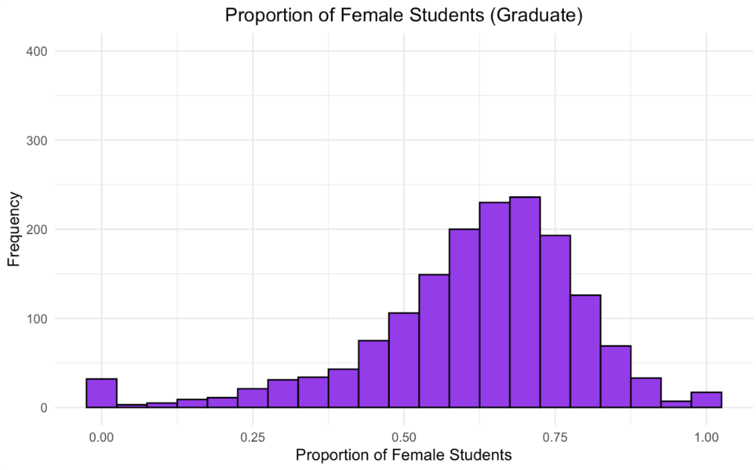


Beta regression

Distributional regression - modeling μ and Φ, instead of slope and intercept

- $0 \rightarrow 0.0001; \quad 1 \rightarrow 0.9999$
- `model_beta <- betareg(prop.female ~ level | level, data = df, link = "logit")`

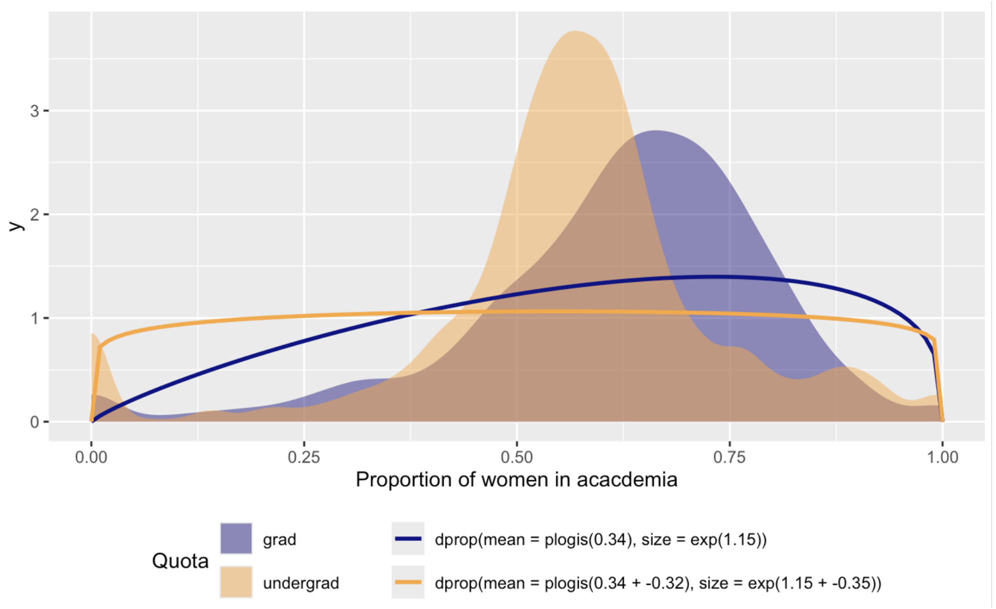




Beta regression

Distributional regression - modeling μ and Φ, instead of slope and intercept

- $0 \rightarrow 0.0001; \quad 1 \rightarrow 0.9999$
- `model_beta <- betareg(prop.female ~ level | level, data = df, link = "logit")`

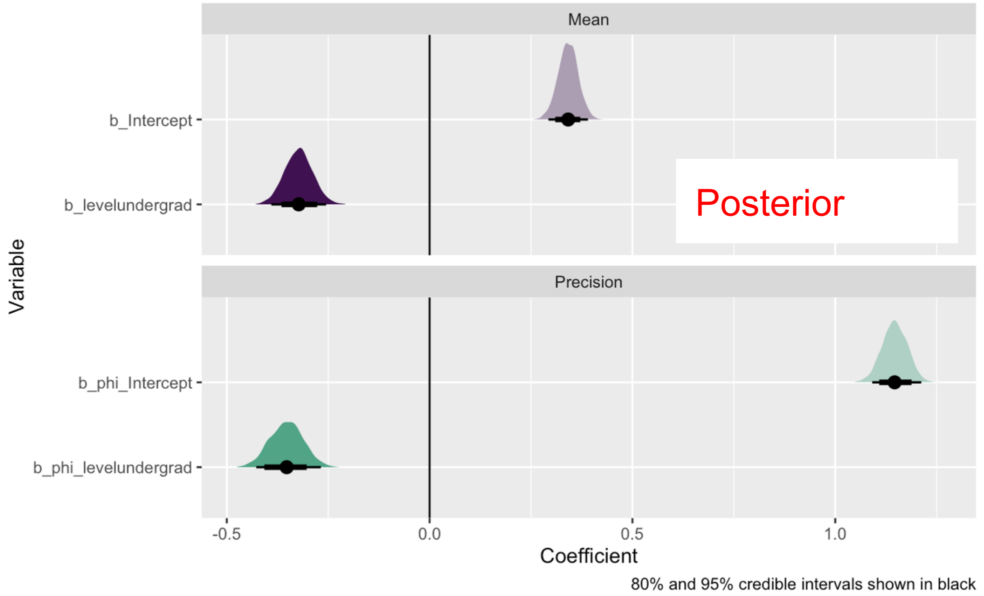



```
> tidy(model_beta)
# A tibble: 4 × 6
  component term            estimate std.error statistic  p.value
  <chr>     <chr>              <dbl>     <dbl>     <dbl>    <dbl>
1 mean      (Intercept)        0.376    0.0226     16.6  4.10e-62
2 mean      levelundergrad    -0.297    0.0320     -9.28 1.69e-20
3 precision (Intercept)        1.34     0.0316     42.4  0
4 precision levelundergrad    -0.334    0.0419     -7.97 1.57e-15
```

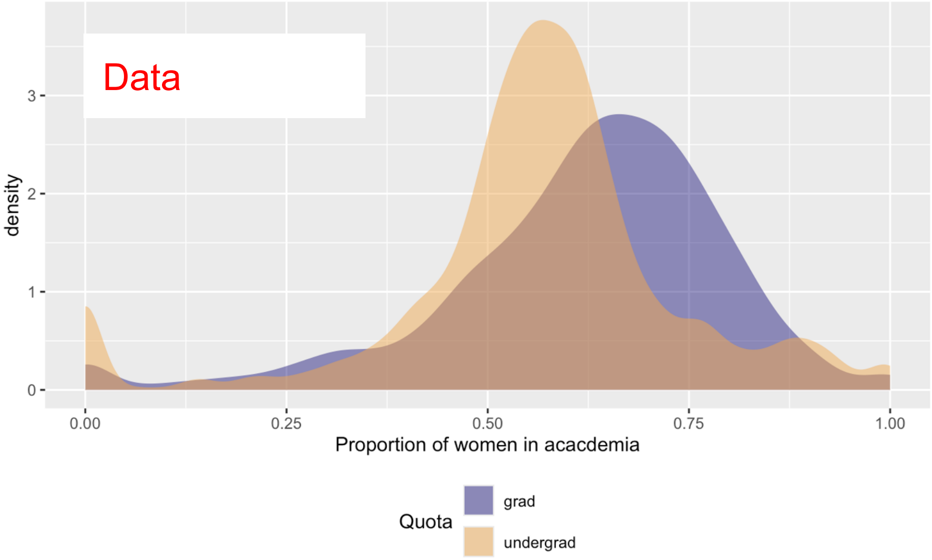
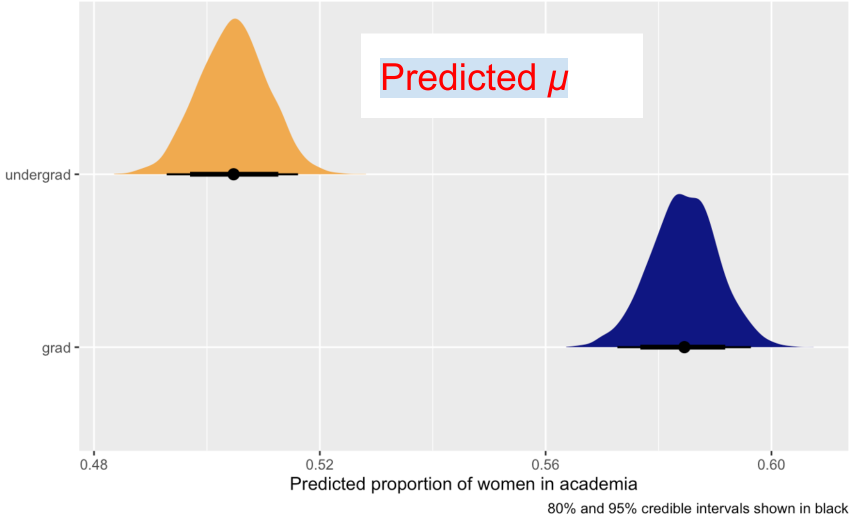
Beta regression - Bayesian

Treat μ and Φ as distributions to model uncertainty

- ```
  model_beta_bayes <- brm(bf(prop.female.posi ~ level, phi ~ level), data = df, family = Beta(), chains = 4, iter = 2000, warmup = 1000, cores = 4)
  ```



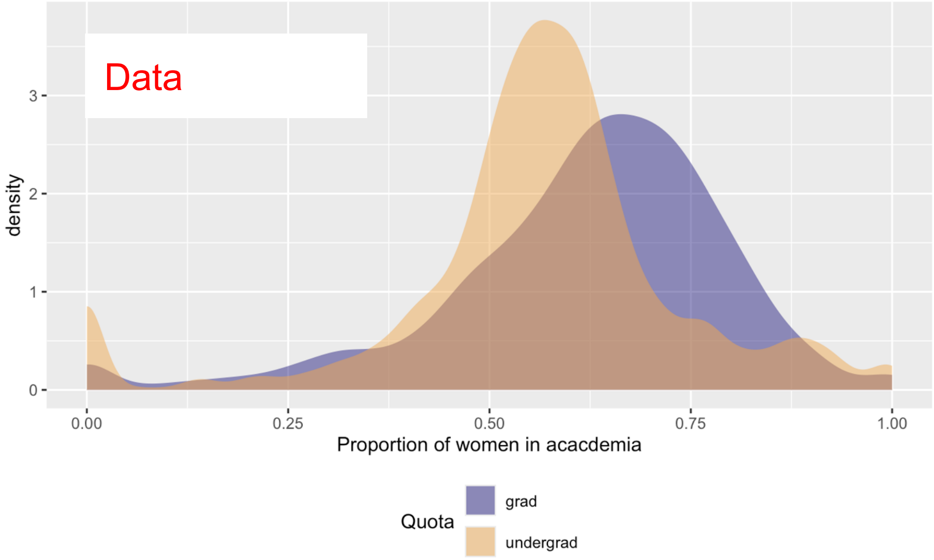
Population-Level Effects:

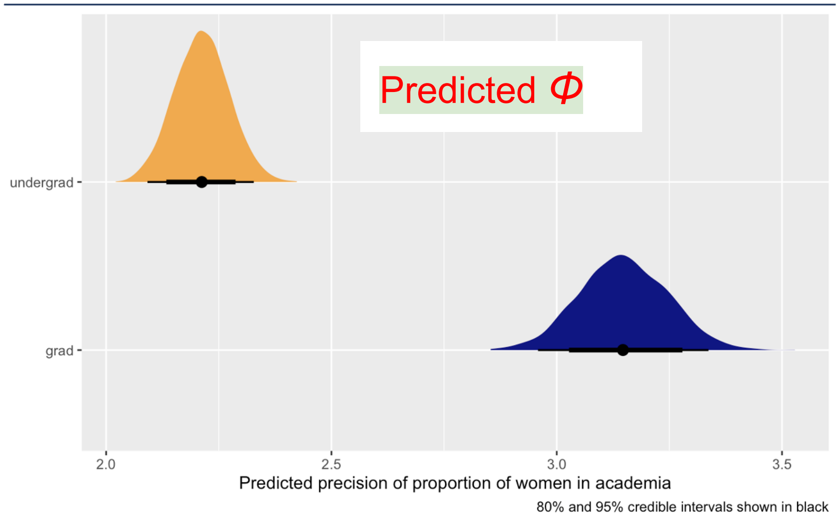
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	0.34	0.02	0.29	0.39	1.00	4432	3088
phi_Intercept	1.15	0.03	1.09	1.21	1.00	4025	3017
levelundergrad	-0.32	0.03	-0.39	-0.25	1.00	3939	2676
phi_levelundergrad	-0.35	0.04	-0.43	-0.27	1.00	4384	3092
  ```

Beta regression - Bayesian

Treat μ and Φ as distributions to model uncertainty

- ```
  model_beta_bayes <- brm(bf(prop.female.posi ~ level, phi ~ level), data = df, family = Beta(), chains = 4, iter = 2000, warmup = 1000, cores = 4)
  ```




  ```

Beta regression - Bayesian

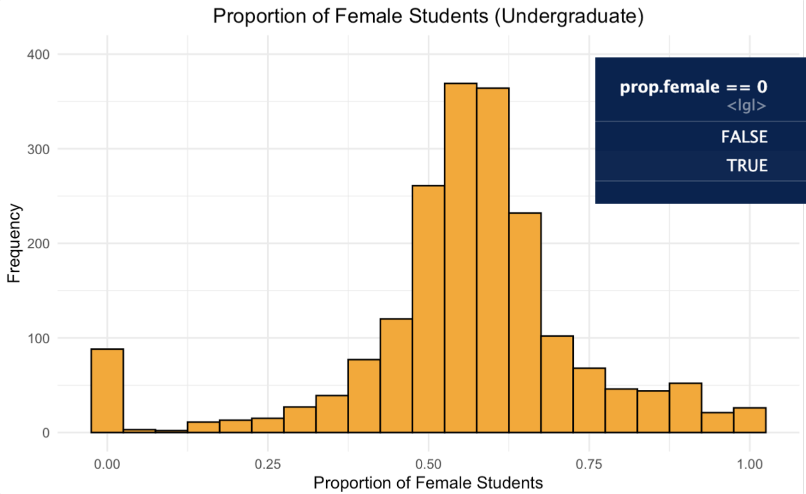
Treat μ and Φ as distributions to model uncertainty

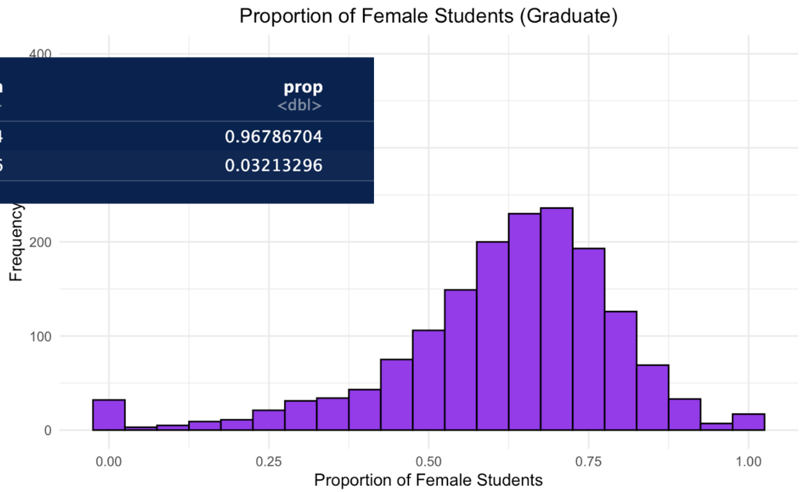
-
```
model_beta_bayes <- brm(bf(prop.female.posi ~ level, phi ~ level), data = df, family = Beta(), chains = 4, iter = 2000, warmup = 1000, cores = 4)
```





Zero-inflated beta regression



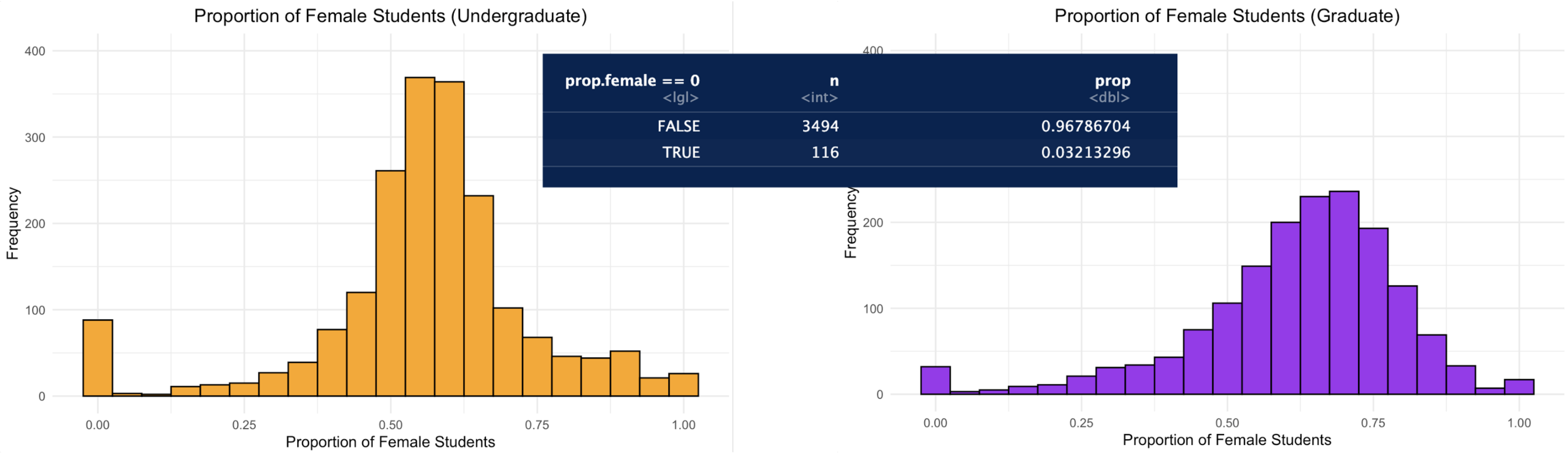


prop.female == 0 <lgl>	n <int>	prop <dbl>
FALSE	3494	0.96786704
TRUE	116	0.03213296

Zero-inflated beta regression

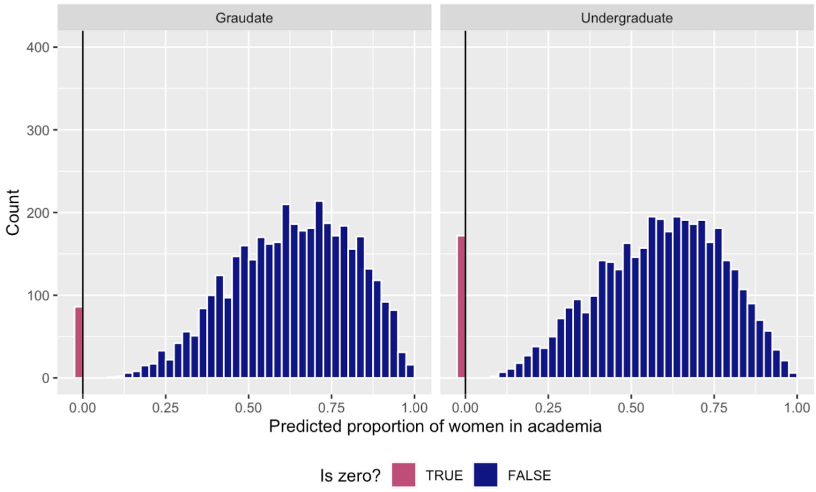
Modelling a mixture of data-generating processes:

1. A **logistic regression model** that predicts if an outcome is 0 or not, defined by **α**
2. A **beta regression model** that predicts if an outcome is between 0 and 1 if it's not zero (***μ*** and ***Φ***)



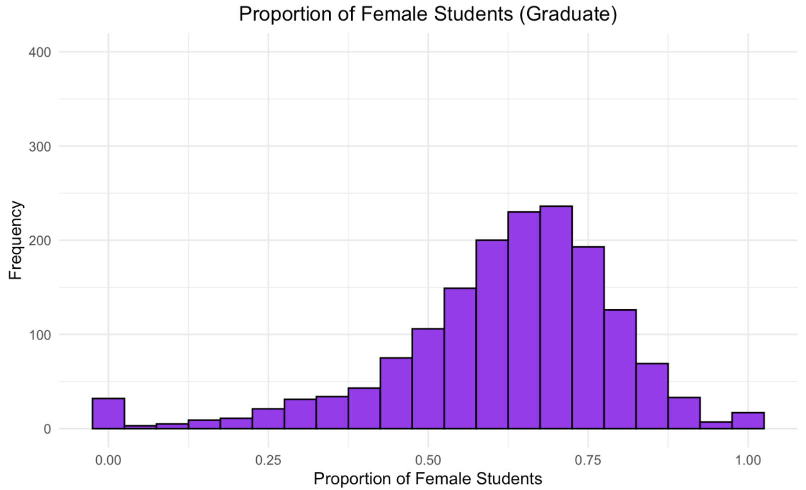
prop.female == 0 <lgl>	n <int>	prop <dbl>
FALSE	3494	0.96786704
TRUE	116	0.03213296

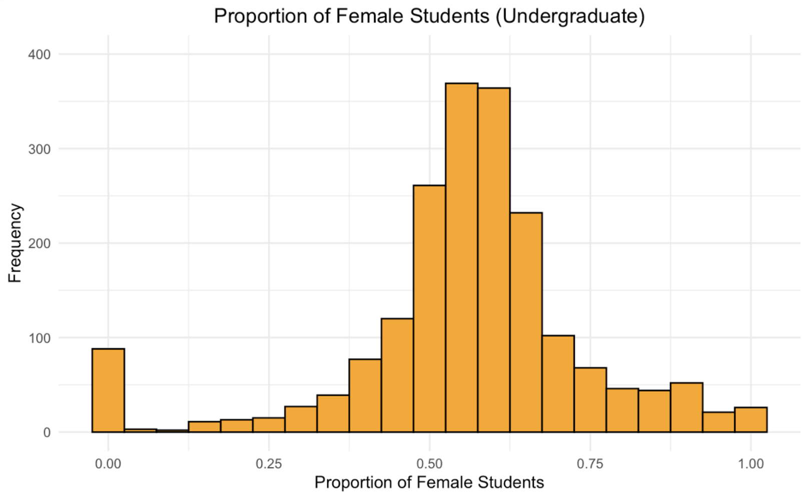
Zero-inflated beta regression



Modeling **α**, ***μ*** and ***Φ***

- 0 → 0; 1 → 0.9999
-
```
model_beta_zi <- brm(bf(prop.female ~
level, phi ~ level, zi ~ level), data = df,
family = zero_inflated_beta(),chains = 4,
iter = 2000, warmup = 1000, cores = 4)
```

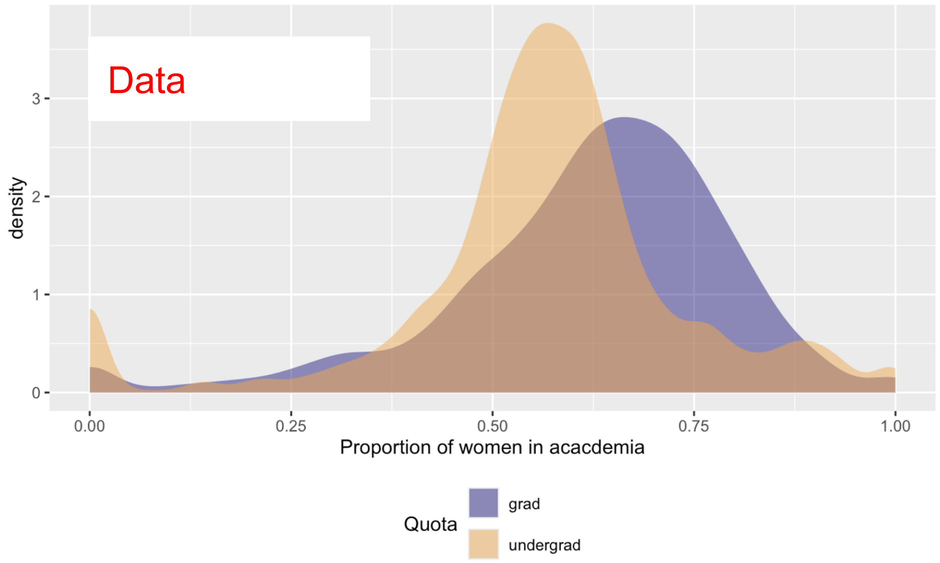


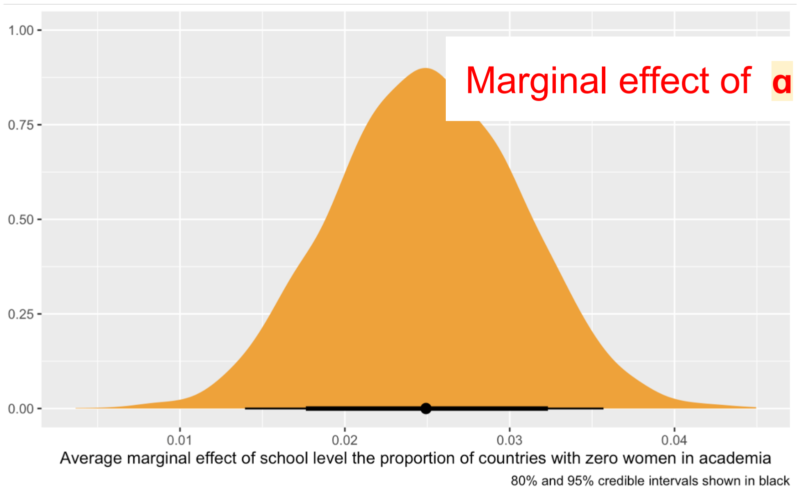


Zero-inflated beta regression

Modeling α, μ and Φ

- `model_beta_zi <- brm(bf(prop.female ~ level, phi ~ level, zi ~ level), data = df, family = zero_inflated_beta(),chains = 4, iter = 2000, warmup = 1000, cores = 4)`





Zero-**one**-inflated beta regression

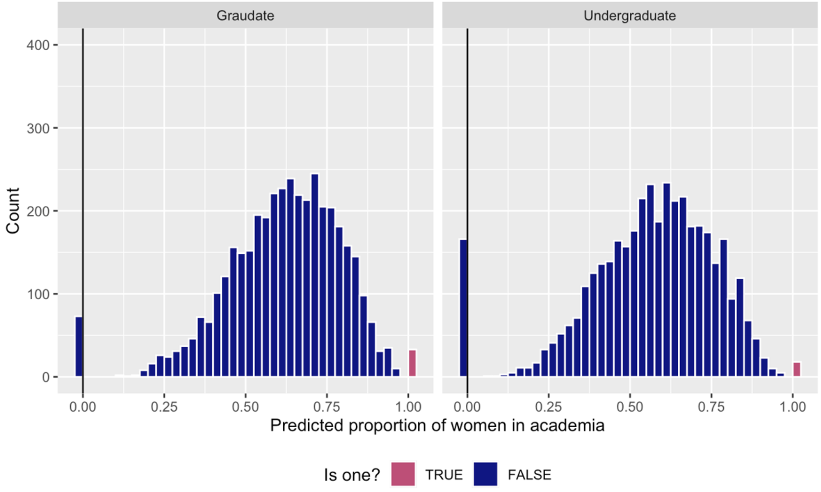
Zero-inflated beta regression:

1. A **logistic regression model** that predicts if an outcome is 0 or not, defined by **α**
2. A **beta regression model** that predicts if an outcome is between 0 and 1 if it's not zero (***μ*** and ***Φ***)

Zero-one-inflated beta regression:

1. A **logistic regression model** that predicts if an outcome is extreme (0 or 1) or not, defined by **α**
2. A **logistic regression model** that predicts if the extreme outcome is 1, defined by **γ**
3. A **beta regression model** that predicts if an outcome is between 0 and 1 if it's not zero (***μ*** and ***Φ***)

Zero-one-inflated beta regression



Modeling **α**, **γ**, **μ** and **Φ**

- $0 \rightarrow 0; \quad 1 \rightarrow 1$
-
```
model_beta_zi <- brm(bf(prop.female ~ level, phi ~ level, zoi
~ level, coi ~ level), data = df, family =
zero_inflated_beta(),chains = 4, iter = 2000, warmup = 1000,
cores = 4)
```