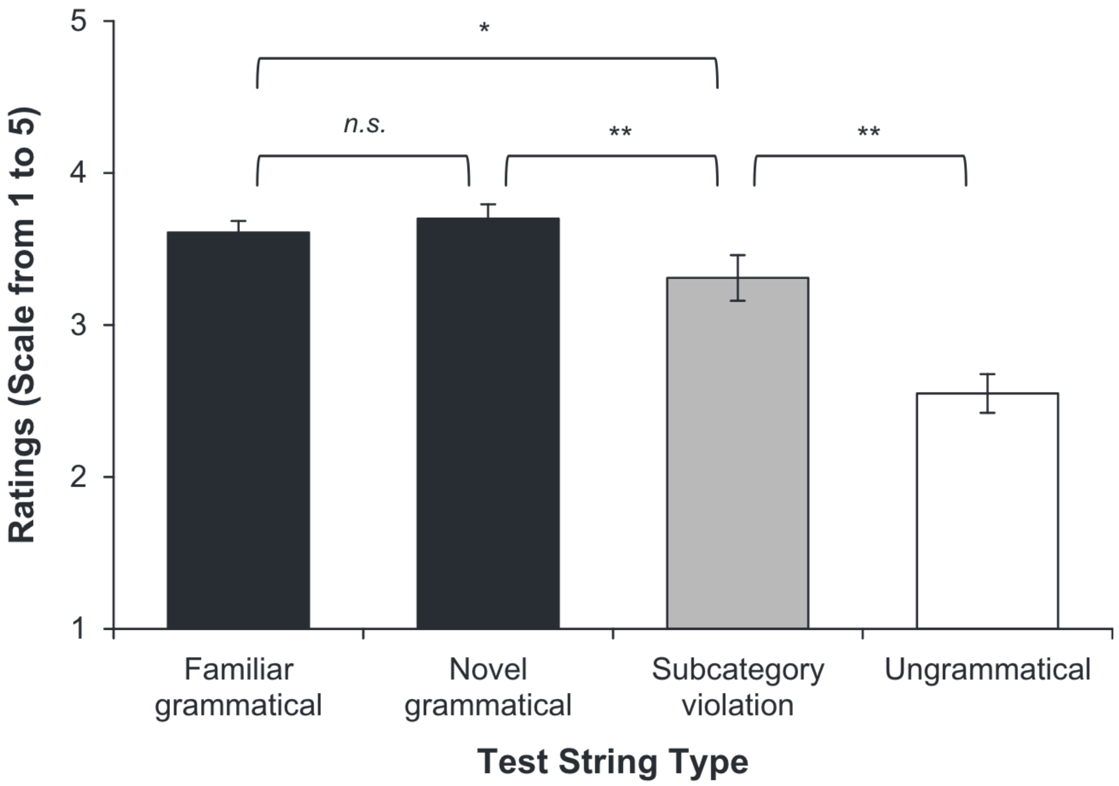
Three common mistakes in statistics and how to avoid them

Elizabeth Pankratz

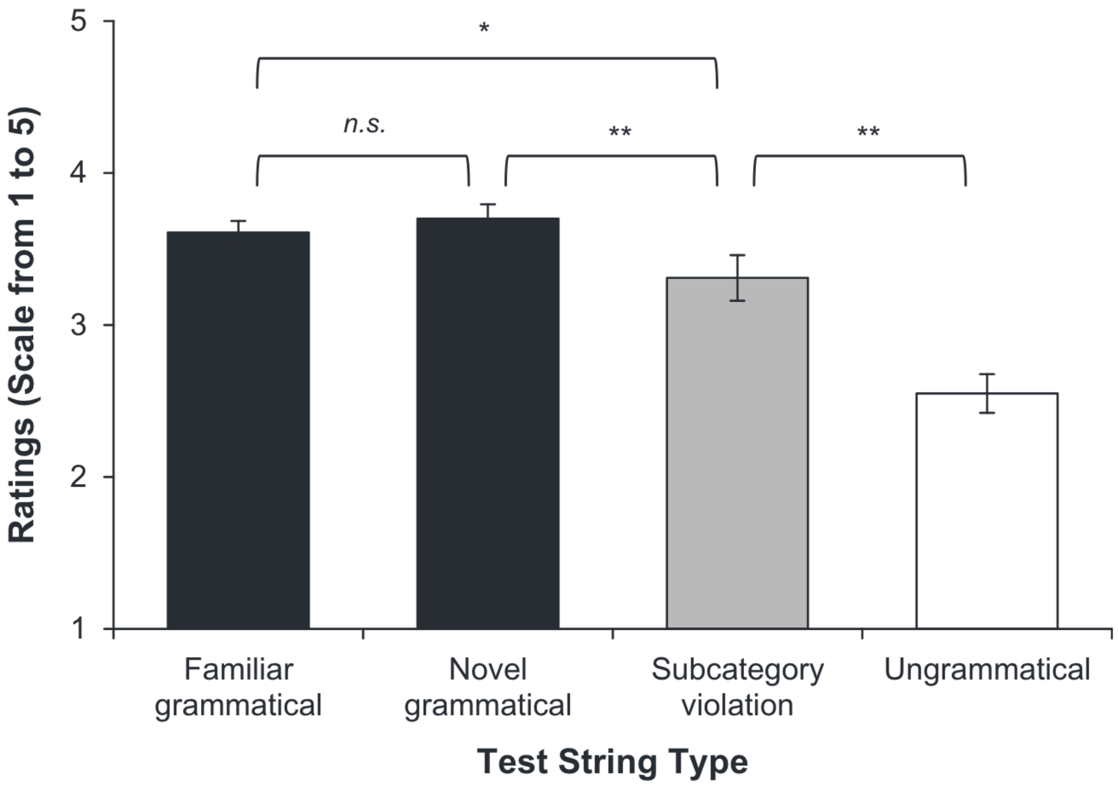
Department of Psychology
The University of Edinburgh

Something you won't be able to unsee

Something you won't be able to unsee



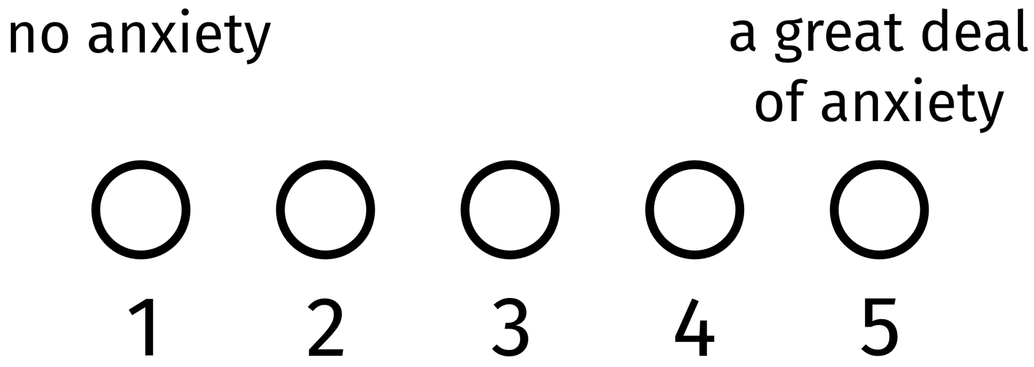
Something you won’t be able to unsee



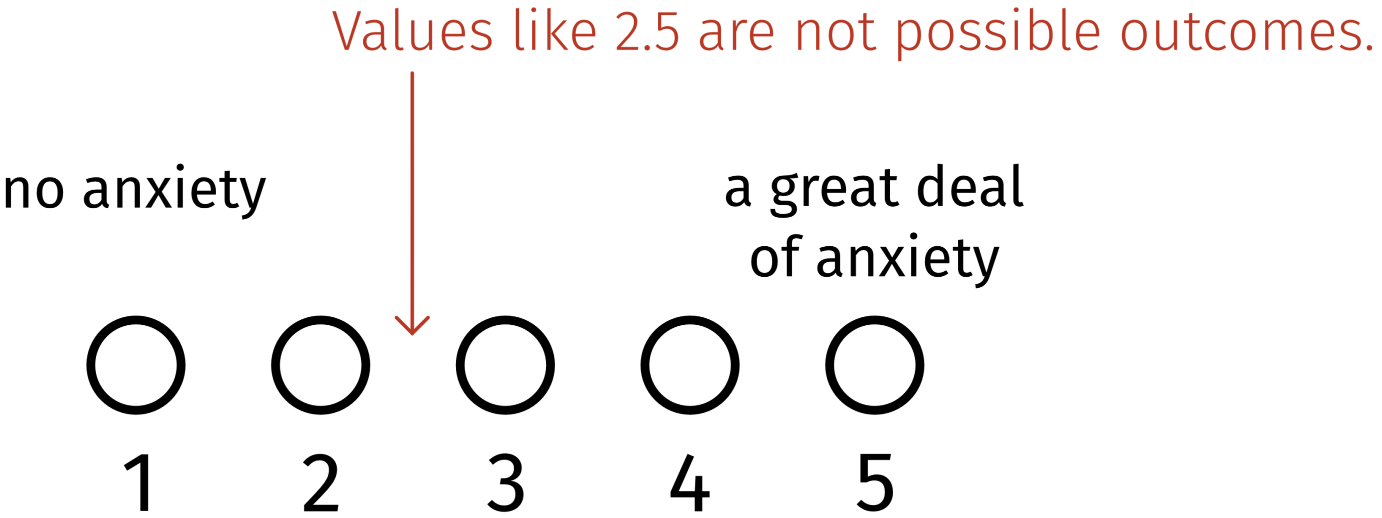
Taking the means of discrete ratings is very common—but a little strange!

Why Likert scale ratings aren't continuous numeric

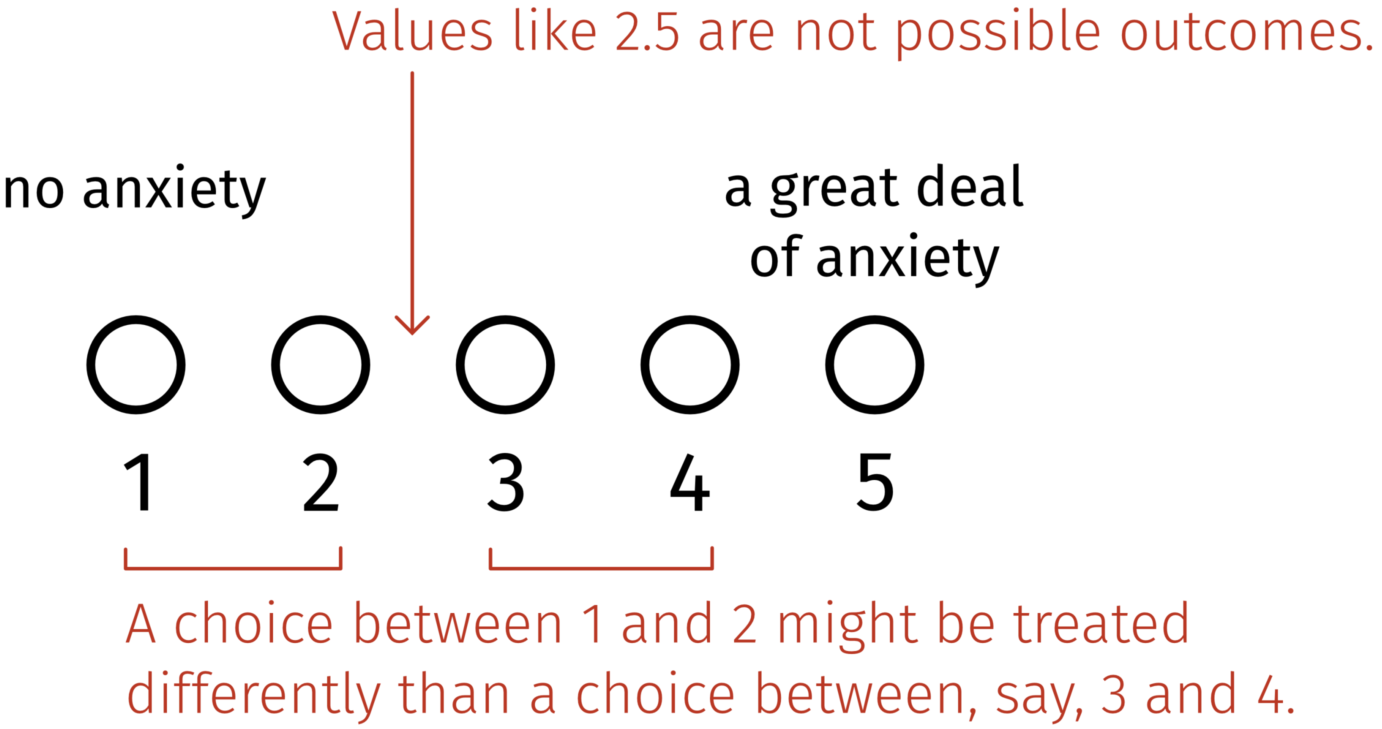
Why Likert scale ratings aren't continuous numeric



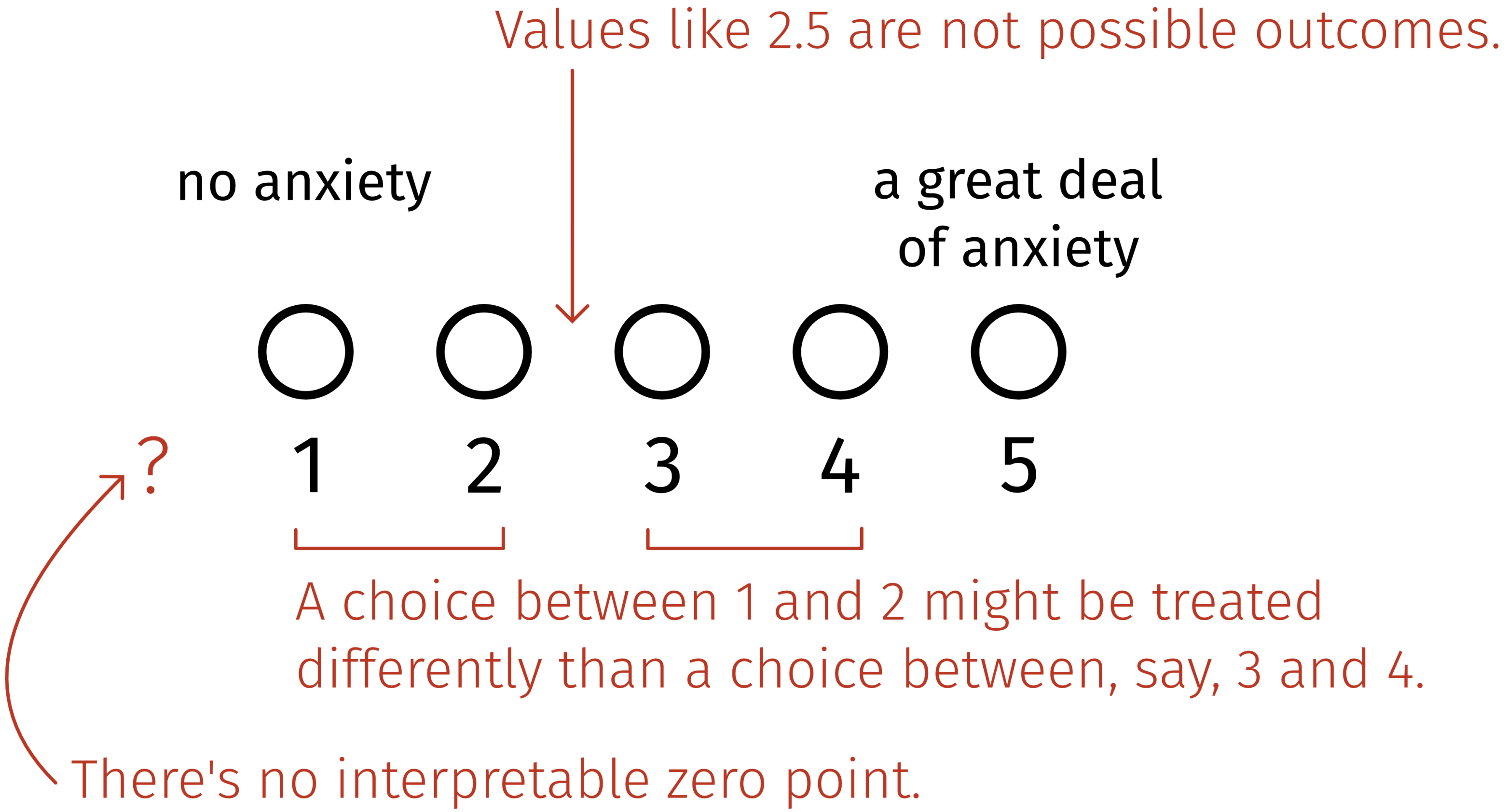
Why Likert scale ratings aren't continuous numeric



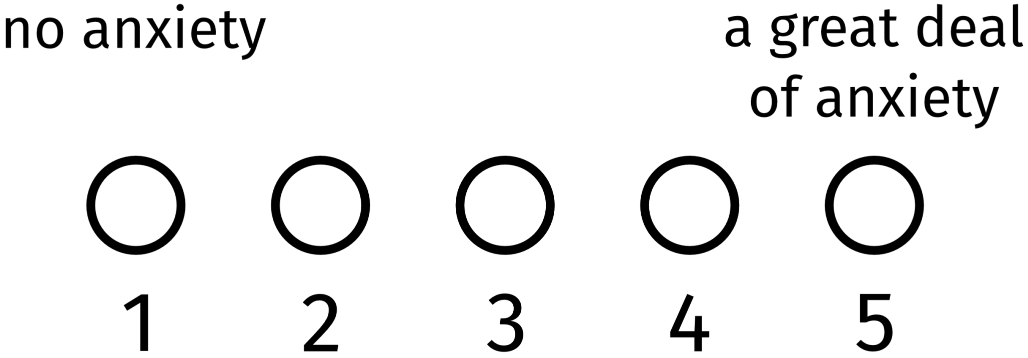
Why Likert scale ratings aren’t continuous numeric



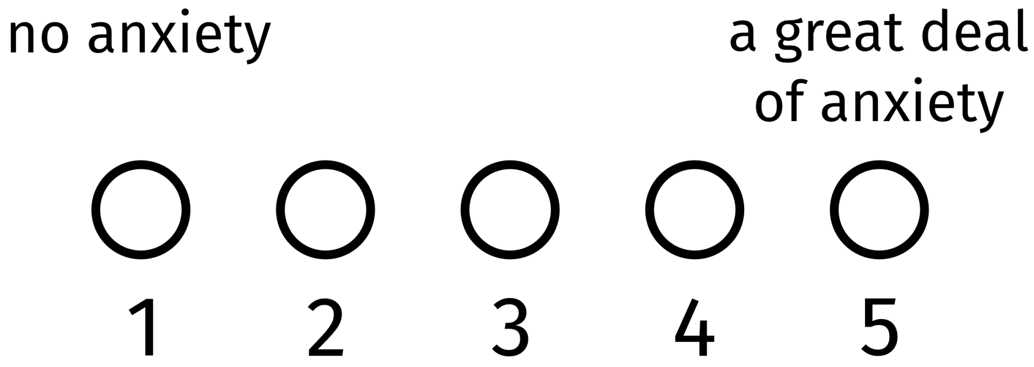
Why Likert scale ratings aren’t continuous numeric



Why Likert scale ratings aren't continuous numeric



Why Likert scale ratings aren't continuous numeric

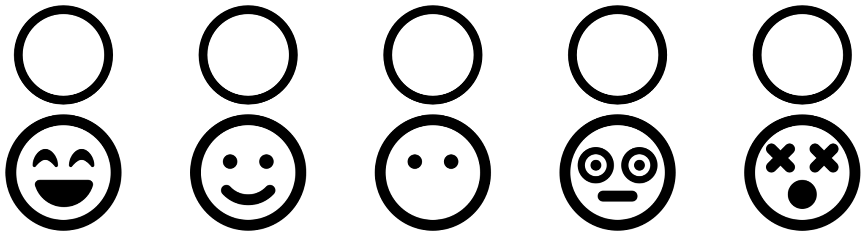


Numbers on a Likert scale are just labels.

Why Likert scale ratings aren’t continuous numeric

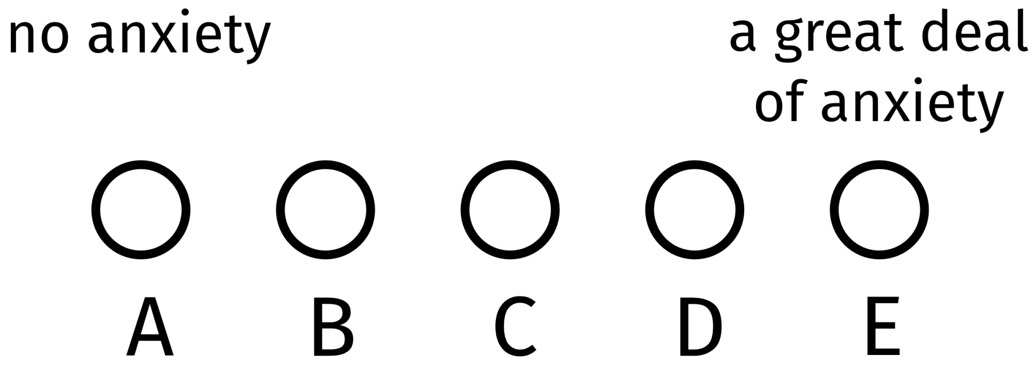
no anxiety

a great deal of anxiety



Numbers on a Likert scale are just labels.

Why Likert scale ratings aren't continuous numeric



Numbers on a Likert scale are just labels.

The mistake

How you'll avoid it

The mistake

</> **A common R mistake:** Letting R treat all variables that look like numbers as continuous numeric.

How you'll avoid it

The mistake

A common R mistake: Letting R treat all variables that look like numbers as continuous numeric.

An advanced stats mistake: Modelling categorical, ordinal data as if it were numeric.

How you’ll avoid it

The mistake

</> **A common R mistake:** Letting R treat all variables that look like numbers as continuous numeric.

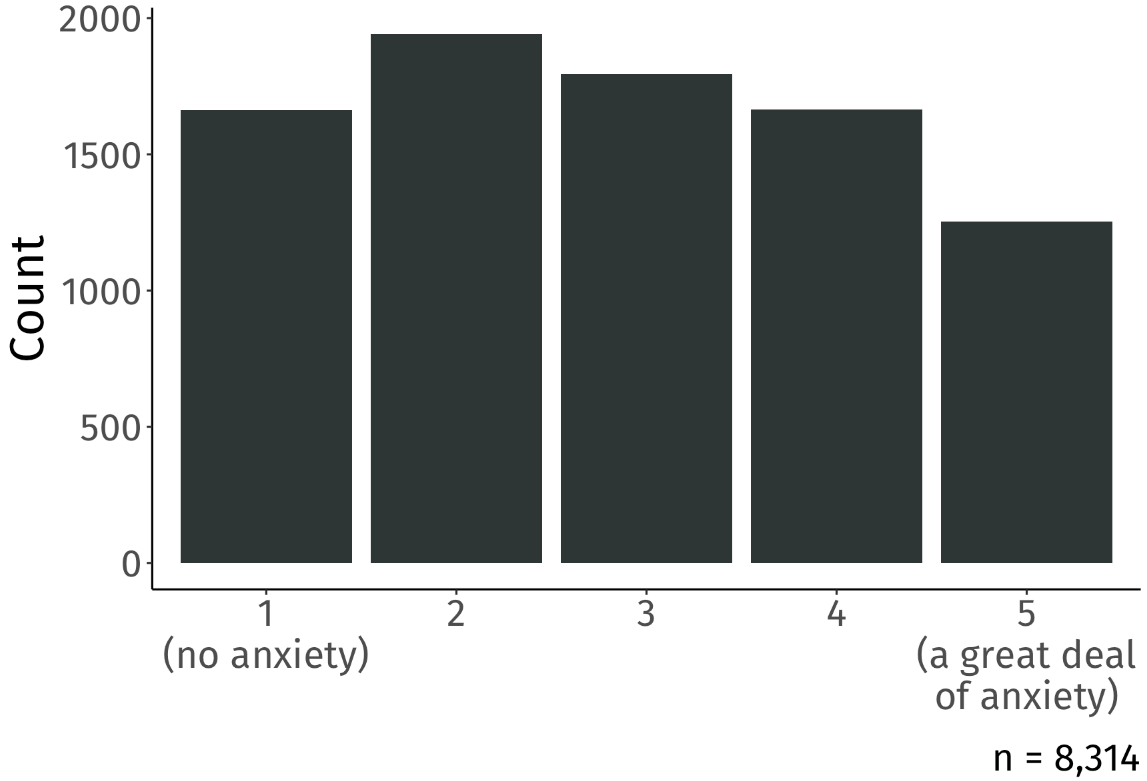
An advanced stats mistake: Modelling categorical, ordinal data as if it were numeric.

∗ **A foundational stats mistake:** Interpreting a significant *p*-value as evidence that an effect exists in the real world.

How you'll avoid it

The data: Students' anxiety ratings for "Going to ask my statistics teacher for individual help with material I am having difficulty understanding".

The data: Students' anxiety ratings for "Going to ask my statistics teacher for individual help with material I am having difficulty understanding".



The data: Students' anxiety ratings for "Going to ask my statistics teacher for individual help with material I am having difficulty understanding".

```r
slice(anx, 45:50)
```

```
## # A tibble: 6 × 3
##   unique_id gender          rating
##   <chr>     <chr>            <dbl>
## 1 7d28c303  Female/Woman         4
## 2 7d55383a  Another Gender       4
## 3 8116550a  Female/Woman         1
## 4 83491ff9  Female/Woman         4
## 5 8450f8ad  Male/Man             2
## 6 876547d6  Female/Woman         3
```

The data: Students' anxiety ratings for "Going to ask my statistics teacher for individual help with material I am having difficulty understanding".

```r
slice(anx, 45:50)
```

```
## # A tibble: 6 × 3
##   unique_id gender          rating
##   <chr>     <chr>            <dbl>
## 1 7d28c303  Female/Woman         4
## 2 7d55383a  Another Gender       4
## 3 8116550a  Female/Woman         1
## 4 83491ff9  Female/Woman         4
## 5 8450f8ad  Male/Man             2
## 6 876547d6  Female/Woman         3
```

`rating` looks like numbers, and R treats it like numbers, as `dbl`.

So it's tempting to manipulate it like numbers.

```r
mean(anx$rating)
```

```
## [1] 2.868054
```

Remember: We are smarter than R is

Remember: We are smarter than R is

Store categorical variables as factors.

```
anx <- anx |>
  mutate(rating = factor(rating))
```

Remember: We are smarter than R is

Store categorical variables as factors.

```
anx <- anx |>
  mutate(rating = factor(rating))
```

Now it’s impossible to incorrectly treat them as if they're numeric!

```
mean(anx$rating)
```

```
## [1] NA
```

The mistake

A common R mistake: Letting R treat all variables that look like numbers as continuous numeric.

An advanced stats mistake: Modelling categorical, ordinal data as if it were numeric.

A foundational stats mistake: Interpreting a significant *p*-value as evidence that an effect exists in the real world.

How you'll avoid it

The mistake

</> **A common R mistake:** Letting R treat all variables that look like numbers as continuous numeric.

An advanced stats mistake: Modelling categorical, ordinal data as if it were numeric.

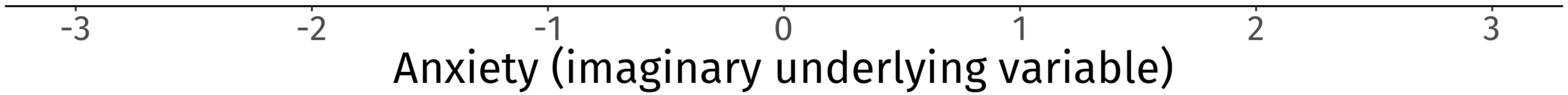
∗ **A foundational stats mistake:** Interpreting a significant *p*-value as evidence that an effect exists in the real world.

How you'll avoid it

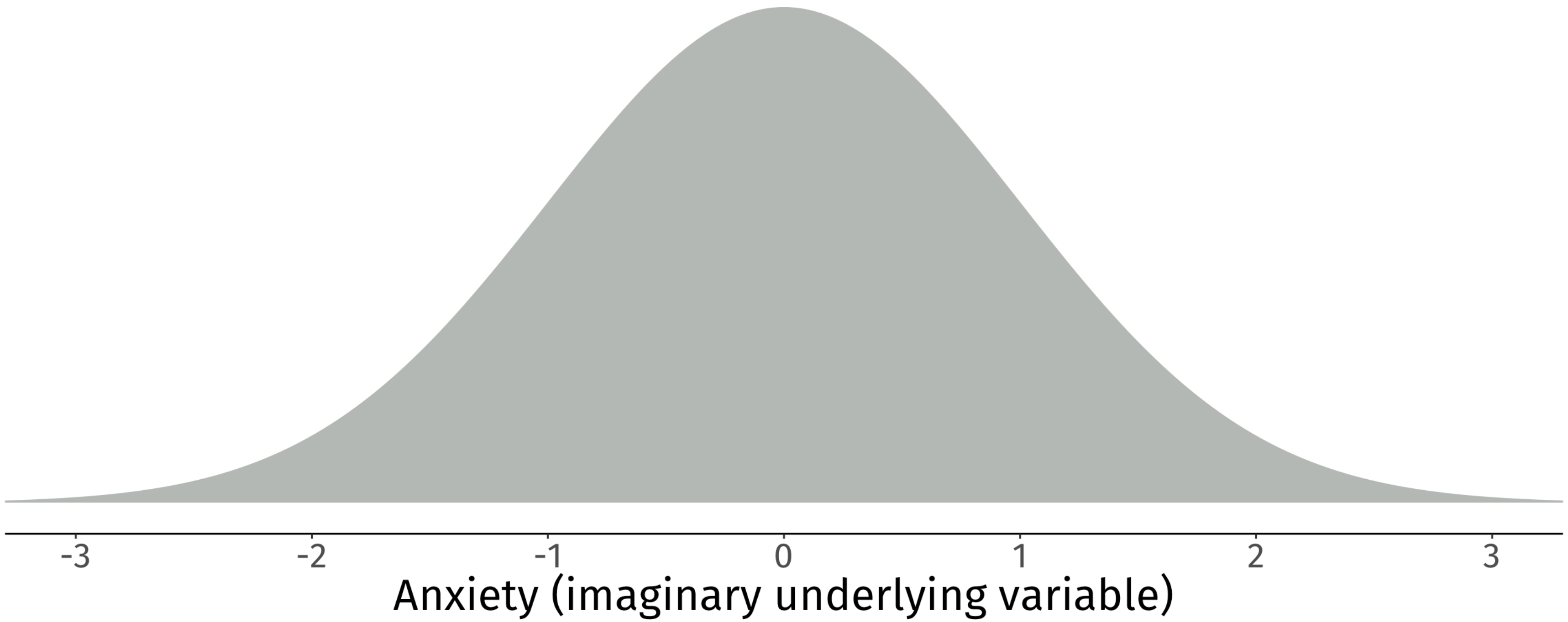
When a variable comes from a Likert scale, tell R it's categorical using `factor()`.

What ordinal regression models do

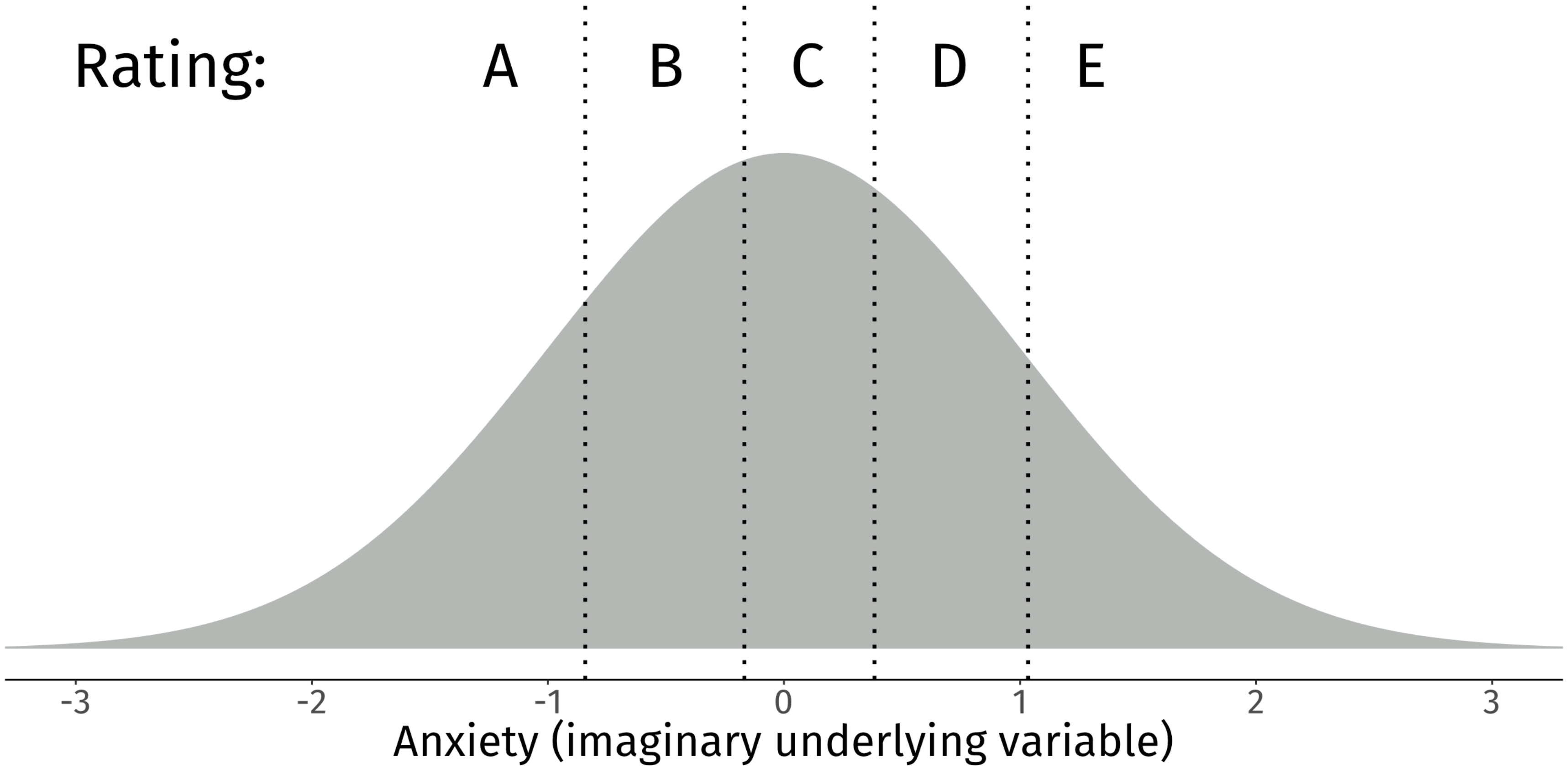
What ordinal regression models do



What ordinal regression models do



What ordinal regression models do



Fit ordinal regression models with `polr()`

Fit ordinal regression models with `polr()`

```
library(MASS)         # MASS contains the polr() function

anx_fit1 <- polr(
  rating ~ 1,         # intercept-only model, to start
  data = anx,
  Hess = TRUE, method = 'probit'   # ask me in the Q+A!
)
```

Fit ordinal regression models with `polr()`

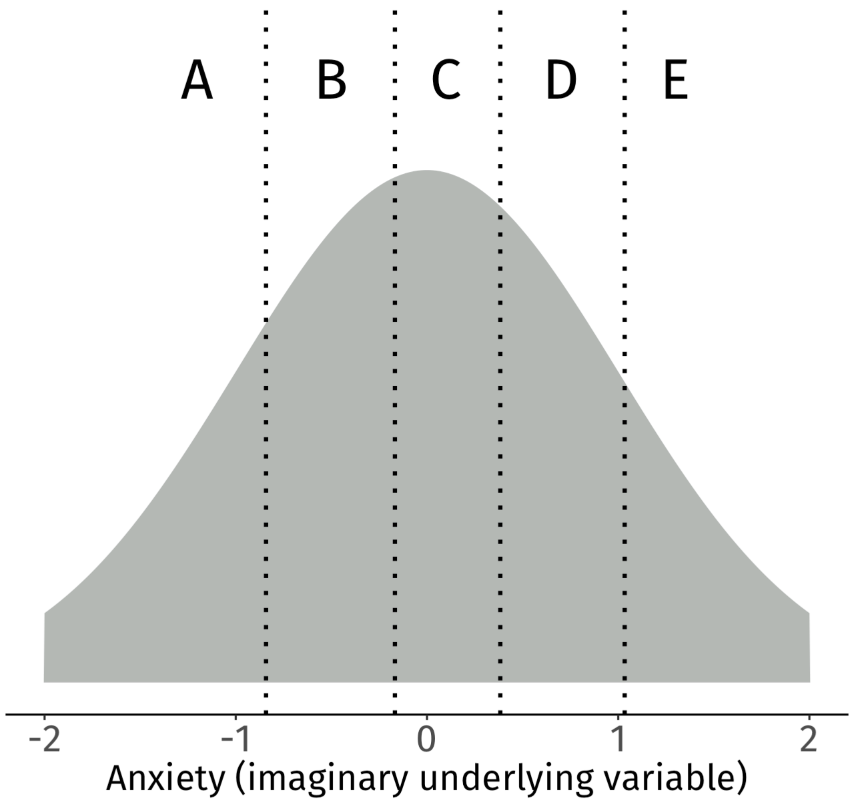
```
summary(anx_fit1)
```

```
## Intercepts:
##      Value    Std. Error t value
## A|B  -0.8420   0.0157   -53.7268
## B|C  -0.1678   0.0138   -12.1462
## C|D   0.3833   0.0141    27.1512
## D|E   1.0339   0.0168    61.6193
```

Fit ordinal regression models with `polr()`

```
summary(anx_fit1)
```

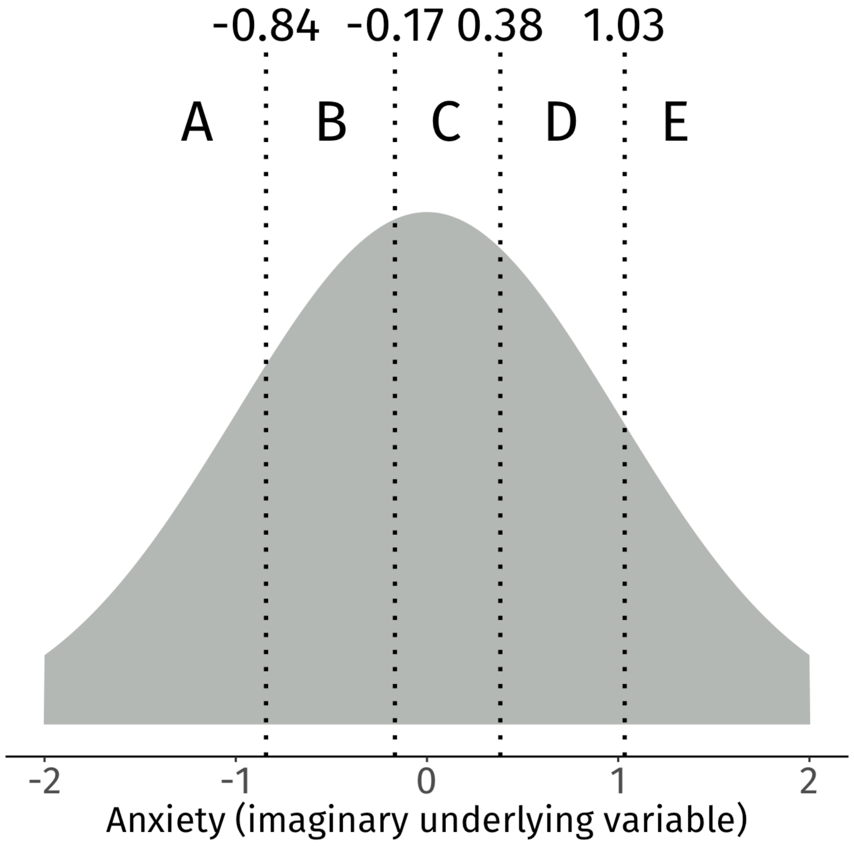
```
## Intercepts:
##      Value    Std. Error t value
## A|B  -0.8420   0.0157   -53.7268
## B|C  -0.1678   0.0138   -12.1462
## C|D   0.3833   0.0141    27.1512
## D|E   1.0339   0.0168    61.6193
```



Fit ordinal regression models with `polr()`

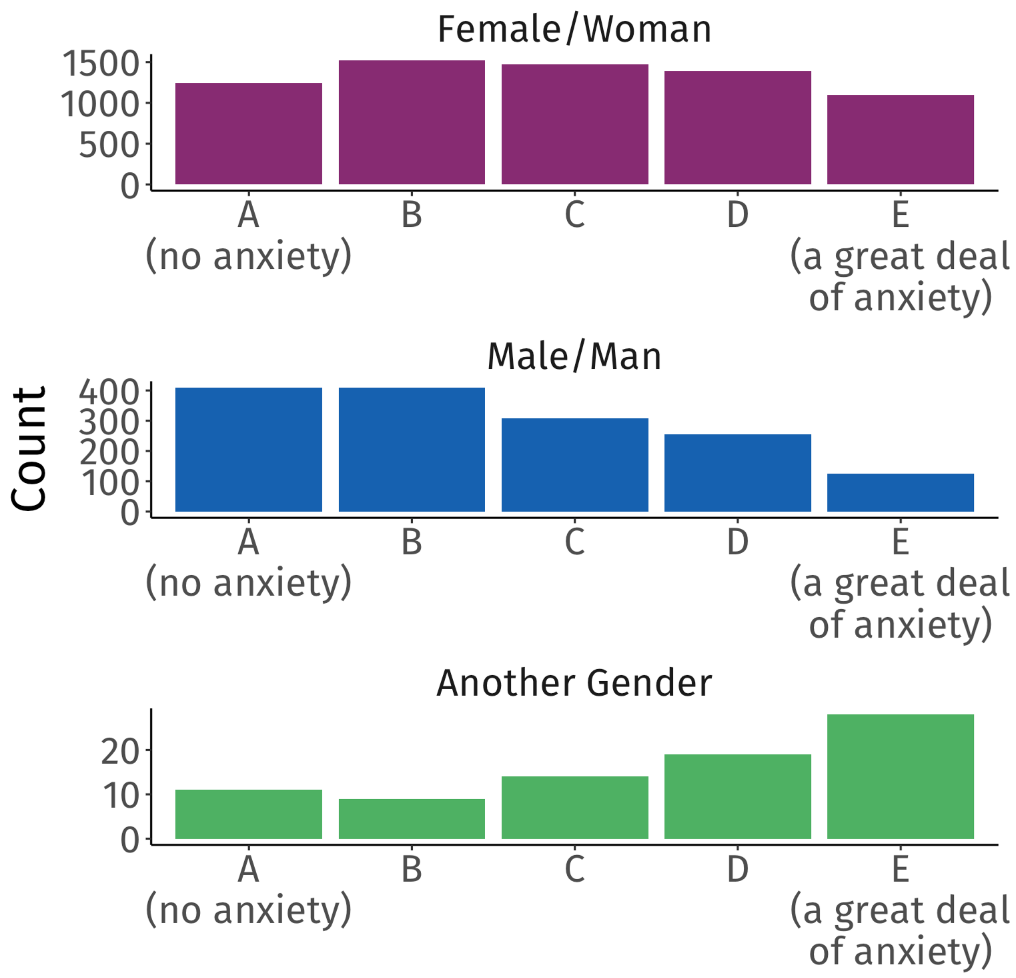
```
summary(anx_fit1)
```

```
## Intercepts:
##      Value    Std. Error t value
## A|B  -0.8420   0.0157   -53.7268
## B|C  -0.1678   0.0138   -12.1462
## C|D   0.3833   0.0141    27.1512
## D|E   1.0339   0.0168    61.6193
```

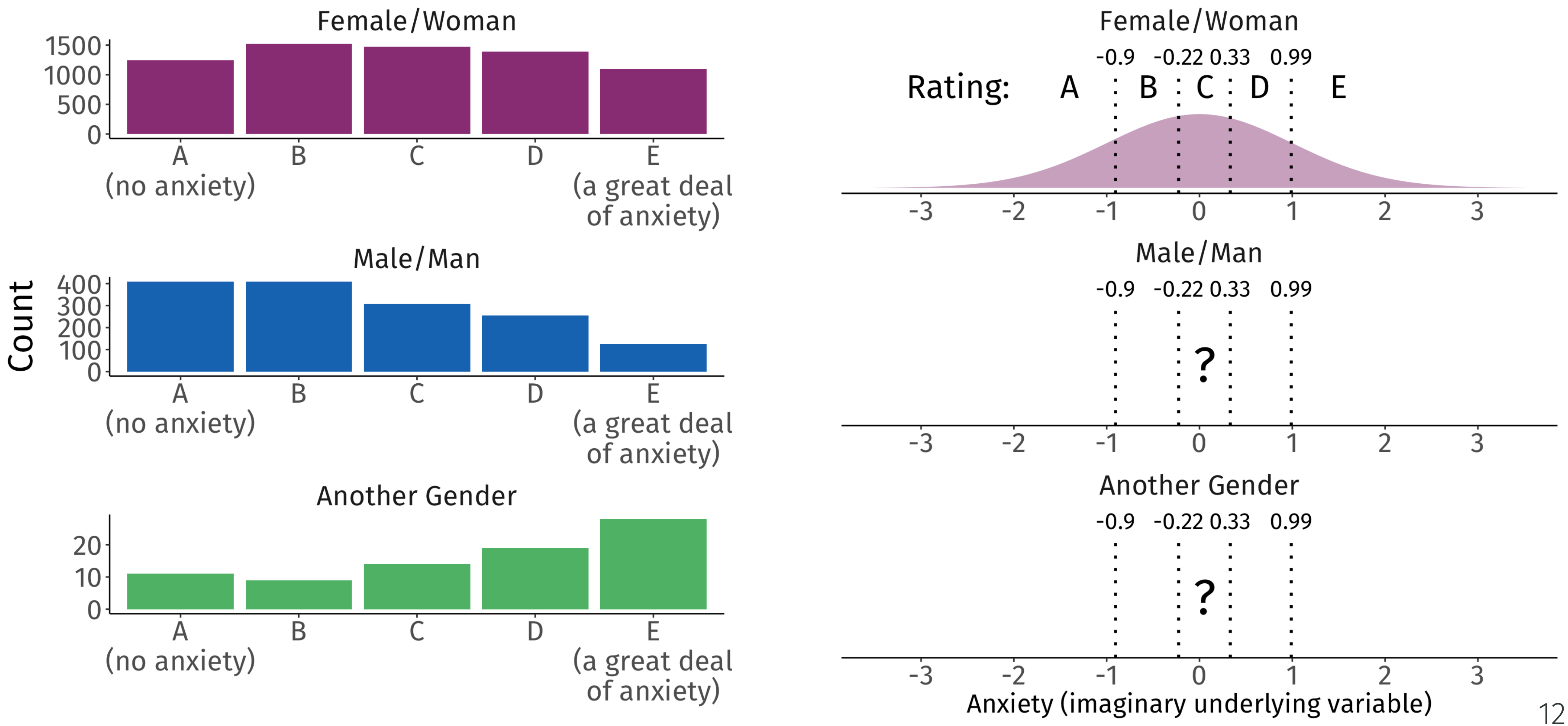


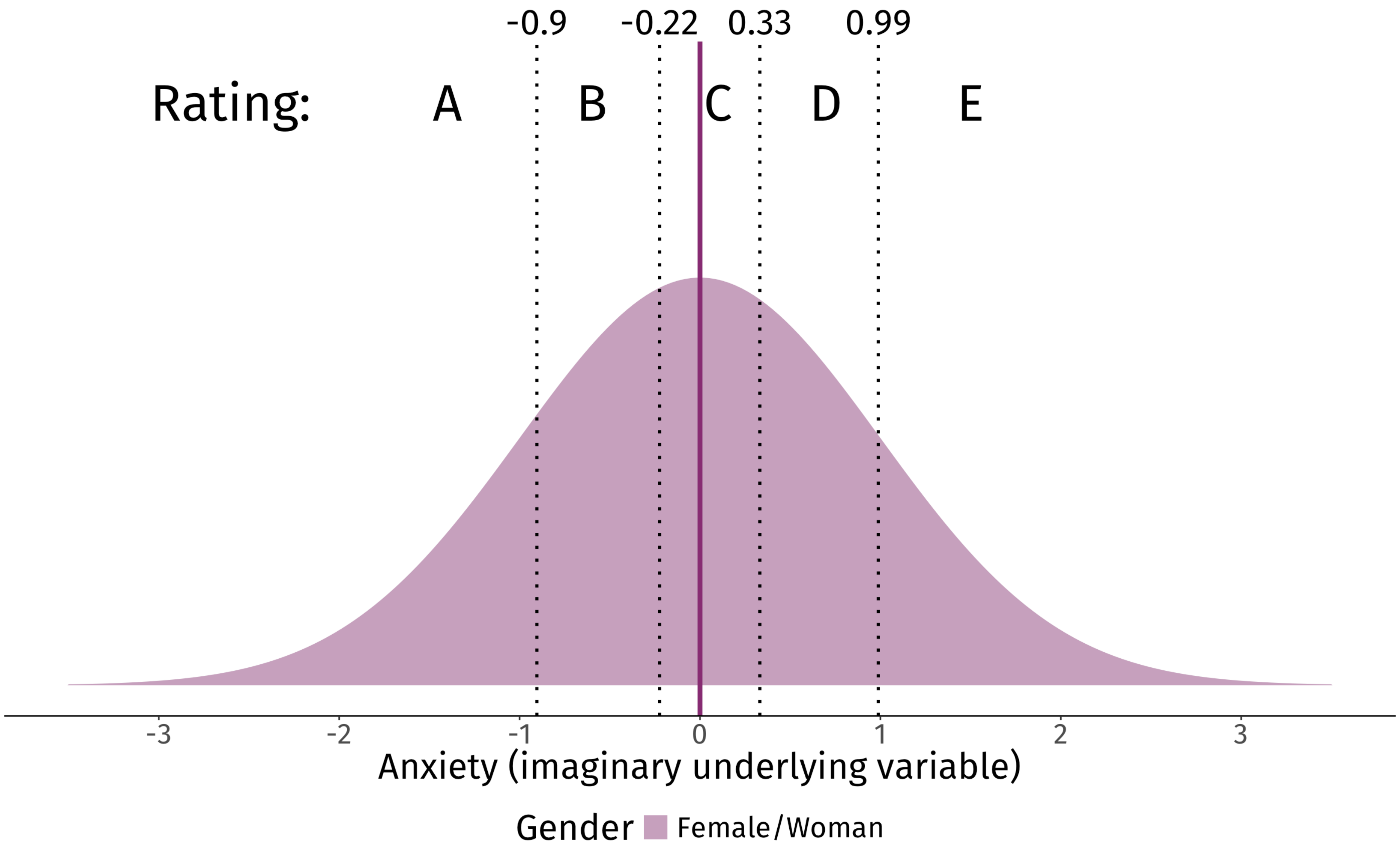
How does a student’s gender affect ratings for “Going to ask my statistics teacher for individual help with material I am having difficulty understanding”?

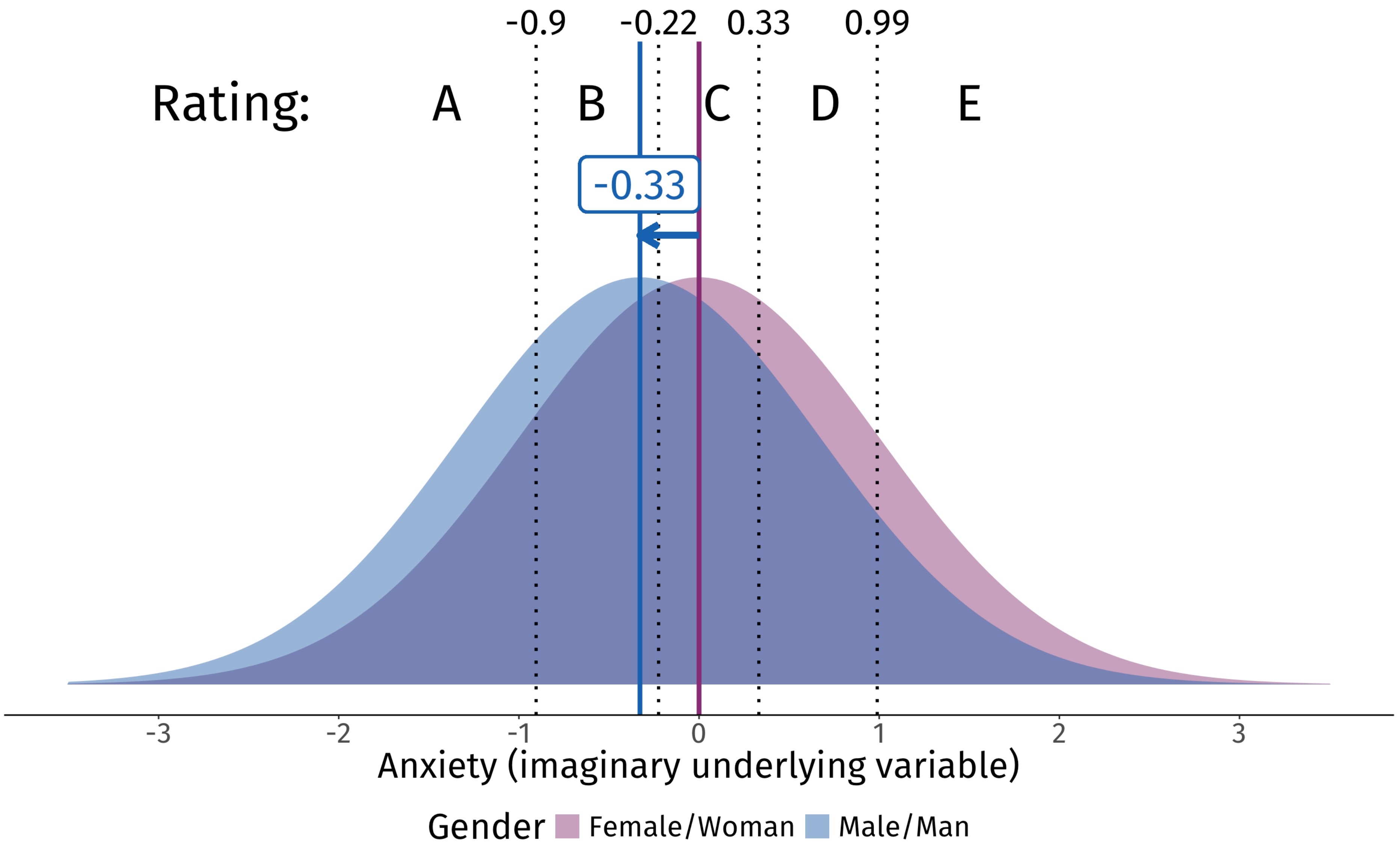
How does a student's gender affect ratings for "Going to ask my statistics teacher for individual help with material I am having difficulty understanding"?

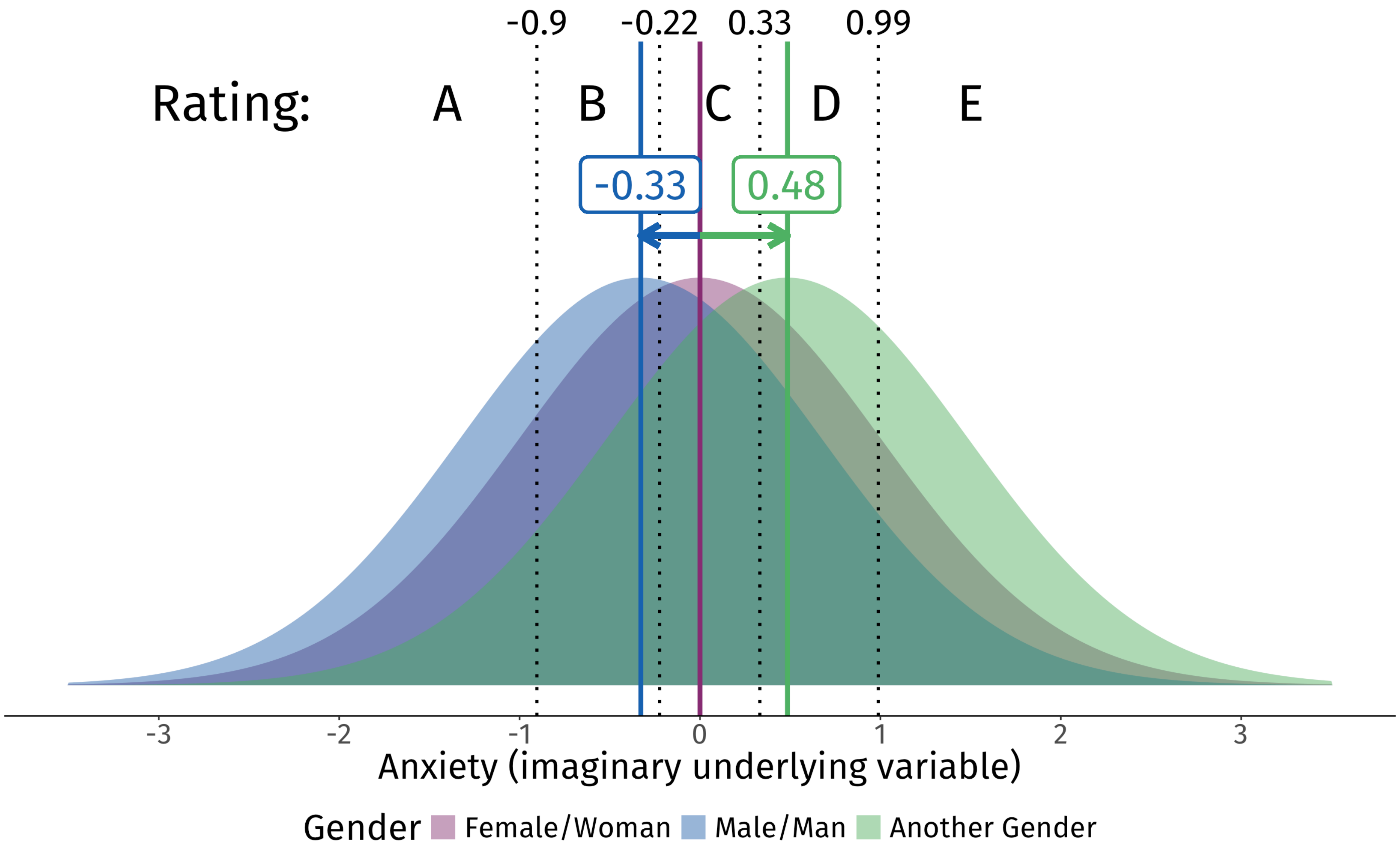


How does a student's gender affect ratings for "Going to ask my statistics teacher for individual help with material I am having difficulty understanding"?




-0.9
-0.22
0.33
0.99
Rating:
A
B
C
D
E
-3
-2
-1
0
1
2
3
Anxiety (imaginary underlying variable)
Gender
Female/Woman


-0.9
-0.22
0.33
0.99
Rating:
A
B
C
D
E
-0.33
-3
-2
-1
0
1
2
3
Anxiety (imaginary underlying variable)
Gender
Female/Woman
Male/Man


-0.9
-0.22
0.33
0.99
Rating:
A
B
C
D
E
-0.33
0.48
-3
-2
-1
0
1
2
3
Anxiety (imaginary underlying variable)
Gender
Female/Woman
Male/Man
Another Gender

The mistake

</> **A common R mistake:** Letting R treat all variables that look like numbers as continuous numeric.

An advanced stats mistake: Modelling categorical, ordinal data as if it were numeric.

∗ **A foundational stats mistake:** Interpreting a significant *p*-value as evidence that an effect exists in the real world.

How you'll avoid it

When a variable comes from a Likert scale, tell R it's categorical using `factor()`.

The mistake

</> **A common R mistake:** Letting R treat all variables that look like numbers as continuous numeric.

An advanced stats mistake: Modelling categorical, ordinal data as if it were numeric.

∗ **A foundational stats mistake:** Interpreting a significant *p*-value as evidence that an effect exists in the real world.

How you'll avoid it

When a variable comes from a Likert scale, tell R it's categorical using `factor()`.

Apply and interpret ordinal regression models (e.g., `polr()` from `MASS`).

Are the effects of `gender` significant?

```
## Coefficients:
##                         Value Std. Error t value
## genderMale/Man        -0.3280    0.03015 -10.880
## genderAnother Gender  0.4846    0.11992   4.041
```

No *p*-values in the model summary.

Are the effects of `gender` significant?

```
## Coefficients:
##                          Value Std. Error t value
## genderMale/Man         -0.3280    0.03015 -10.880
## genderAnother Gender  0.4846    0.11992   4.041
```

No *p*-values in the model summary.

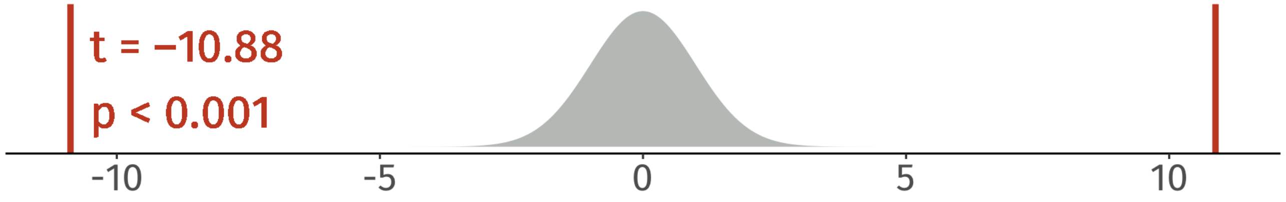
But it's common practice to compare these *t*-values to a standard normal distribution.

Are the effects of `gender` significant?

```
## Coefficients:
##                           Value Std. Error t value
## genderMale/Man          -0.3280    0.03015 -10.880
## genderAnother Gender     0.4846    0.11992   4.041
```

No *p*-values in the model summary.

But it's common practice to compare these *t*-values to a standard normal distribution.

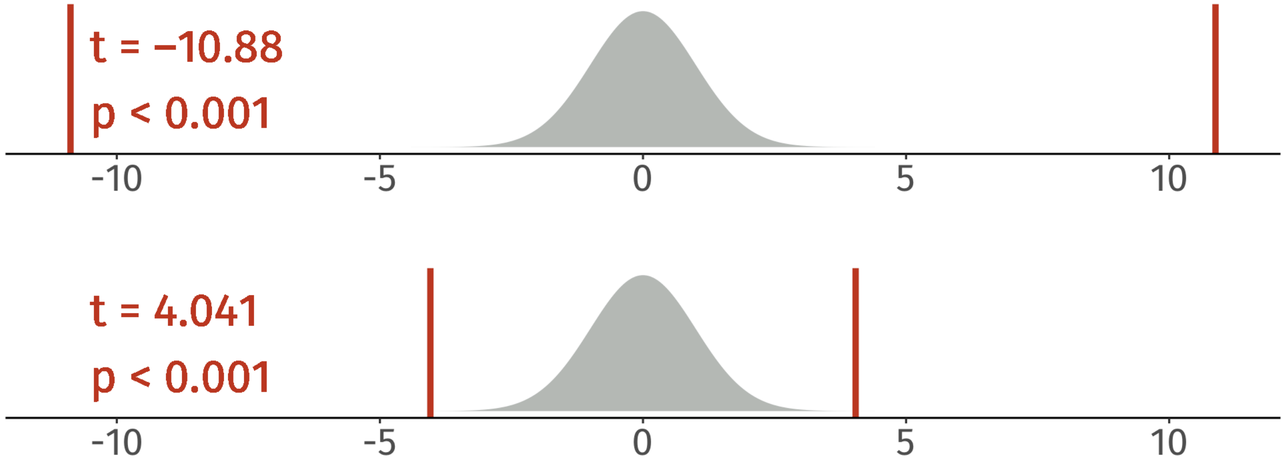


Are the effects of `gender` significant?

```
## Coefficients:
##                              Value Std. Error t value
## genderMale/Man             -0.3280    0.03015 -10.880
## genderAnother Gender        0.4846    0.11992   4.041
```

No *p*-values in the model summary.

But it's common practice to compare these *t*-values to a standard normal distribution.



Why don't significant p-values mean an effect exists?

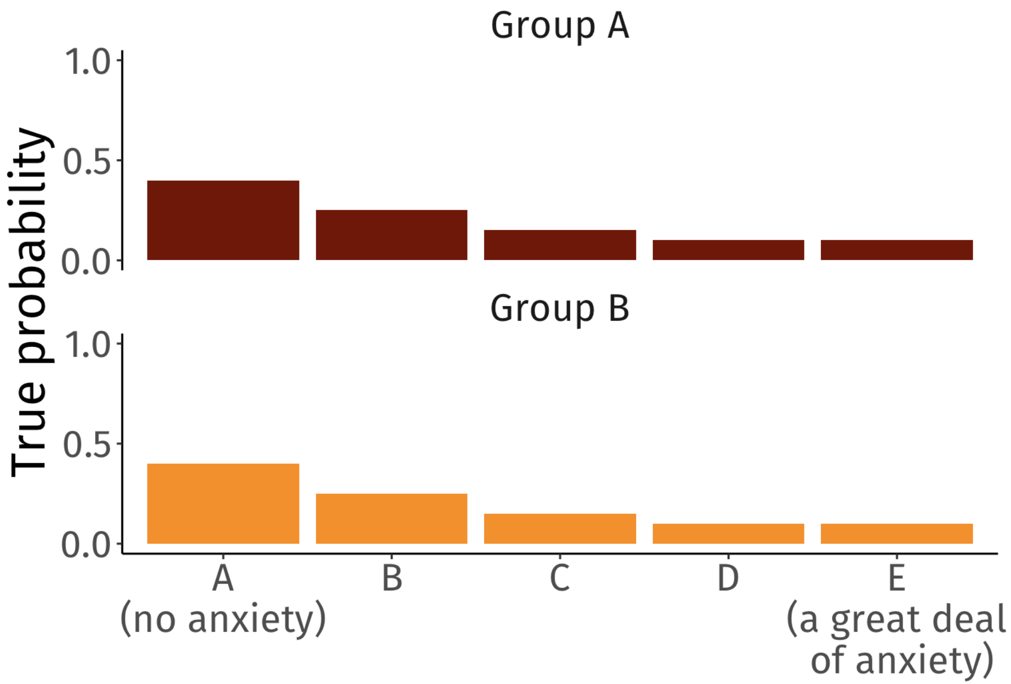
Why don't significant *p*-values mean an effect exists?

Because we can also get significant *p*-values when there really is *no* effect.

Why don't significant *p*-values mean an effect exists?

Because we can also get significant *p*-values when there really is *no* effect.

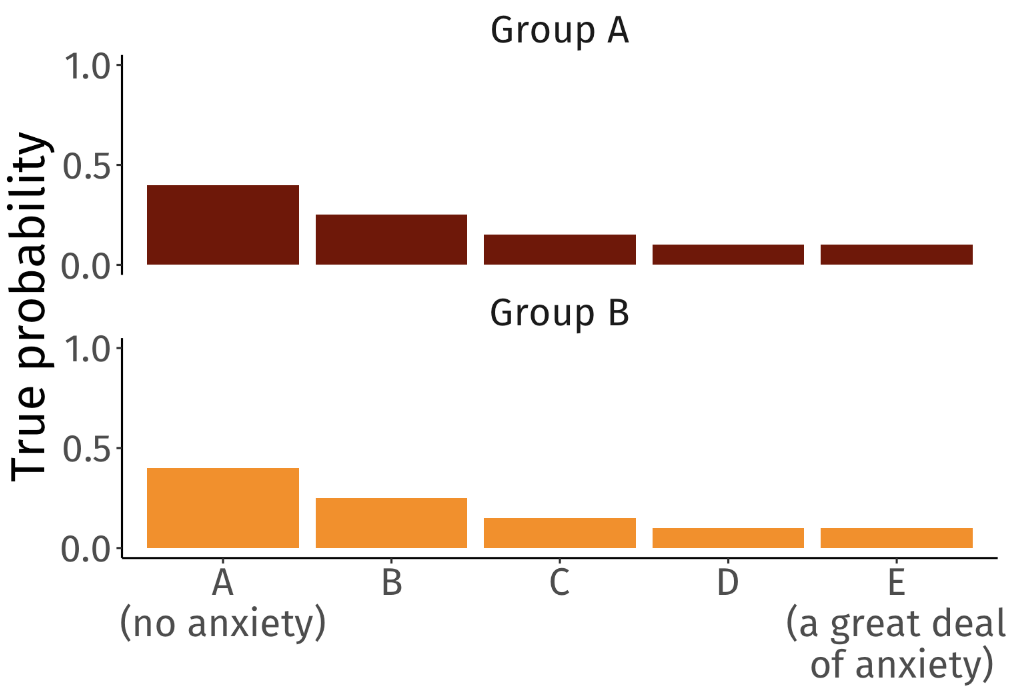
No difference in the true population:



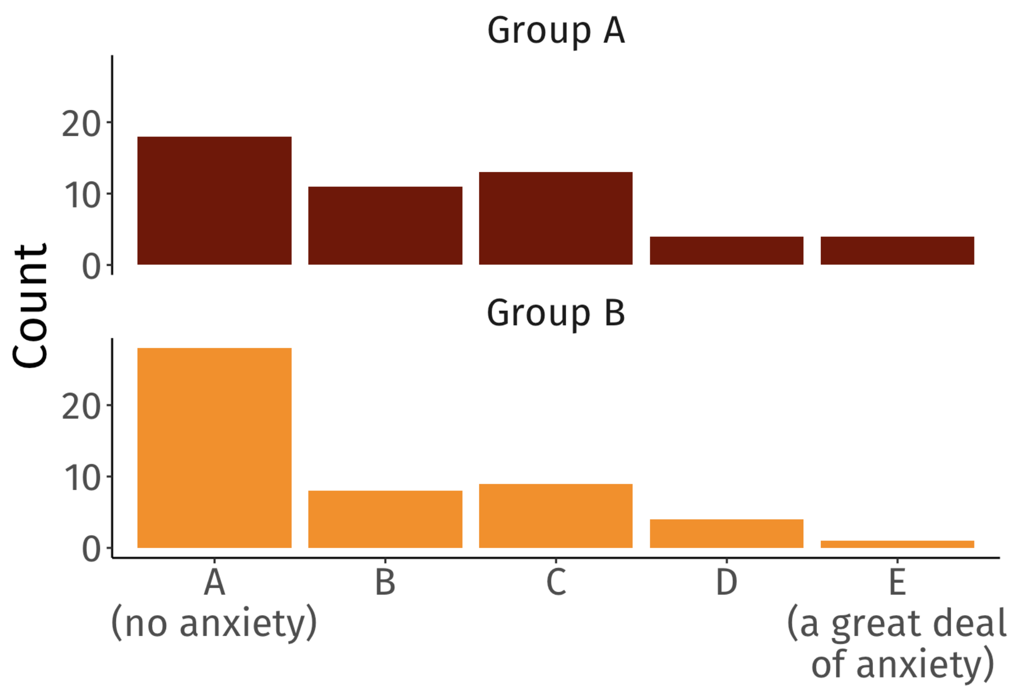
Why don't significant *p*-values mean an effect exists?

Because we can also get significant *p*-values when there really is *no* effect.

No difference in the true population:



A possible random sample (n = 50 per group):



Why don’t significant *p*-values mean an effect exists?

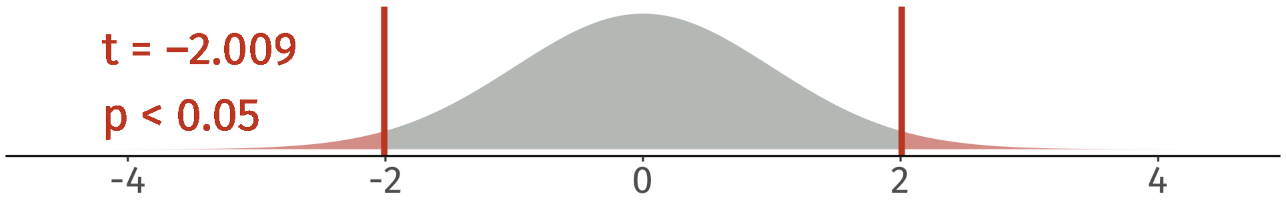
```
sim_fit <- polr(rating ~ group, data = simdat, method = 'probit', Hess = TRUE)
summary(sim_fit)
```

```
## Coefficients:
##                 Value Std. Error t value
## groupGroup B -0.4479     0.2229  -2.009
```

Why don't significant *p*-values mean an effect exists?

```
sim_fit <- polr(rating ~ group, data = simdat, method = 'probit', Hess = TRUE)
summary(sim_fit)
```

```
## Coefficients:
##                 Value Std. Error t value
## groupGroup B -0.4479     0.2229  -2.009
```



So *p* is significant, but in the true population, Group A and Group B were identical!

The mistake

</> **A common R mistake:** Letting R treat all variables that look like numbers as continuous numeric.

An advanced stats mistake: Modelling categorical, ordinal data as if it were numeric.

∗ **A foundational stats mistake:** Interpreting a significant *p*-value as evidence that an effect exists in the real world.

How you'll avoid it

When a variable comes from a Likert scale, tell R it's categorical using `factor()`.

Apply and interpret ordinal regression models (e.g., `polr()` from `MASS`).

The mistake

</> **A common R mistake:** Letting R treat all variables that look like numbers as continuous numeric.

An advanced stats mistake: Modelling categorical, ordinal data as if it were numeric.

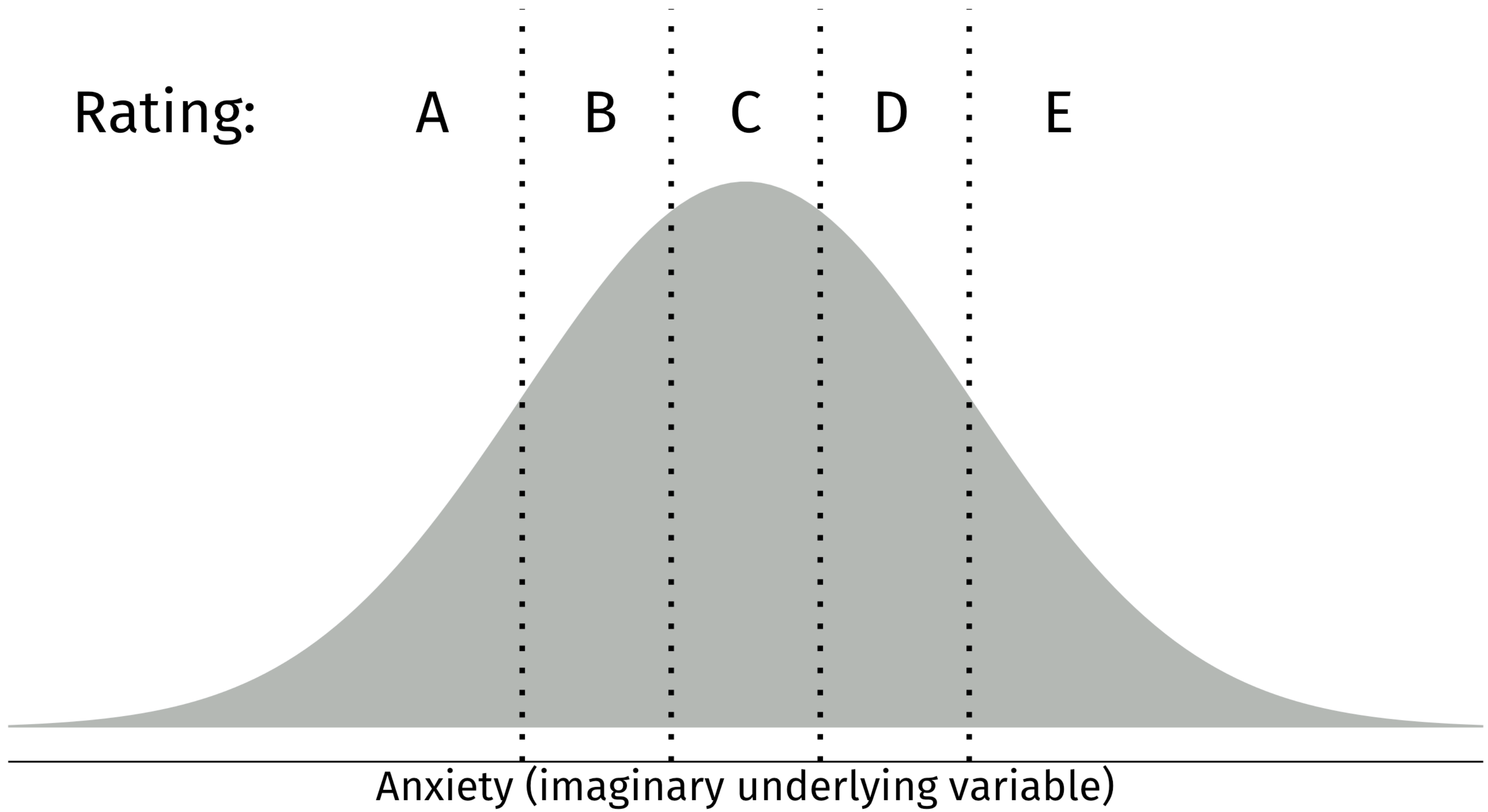
∗ **A foundational stats mistake:** Interpreting a significant *p*-value as evidence that an effect exists in the real world.

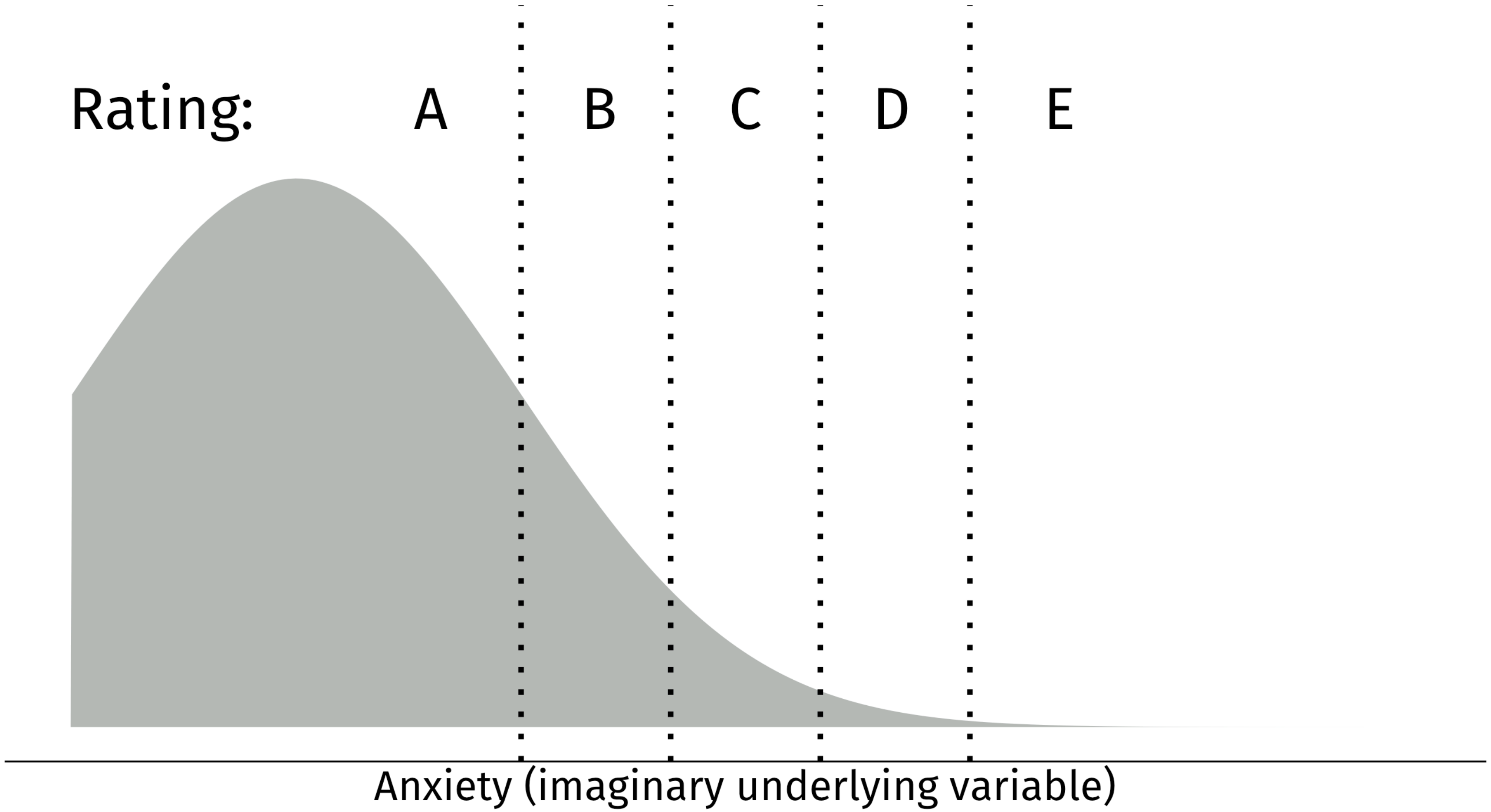
How you'll avoid it

When a variable comes from a Likert scale, tell R it's categorical using `factor()`.

Apply and interpret ordinal regression models (e.g., `polr()` from `MASS`).

Understand that significant *p*-values can arise even if no effect exists in the real world.


Rating:
A
B
C
D
E
Anxiety (imaginary underlying variable)



Rating:
A
B
C
D
E
Anxiety (imaginary underlying variable)

The mistake

</> A common R mistake: Letting R treat all variables that look like numbers as continuous numeric.

An advanced stats mistake: Modelling categorical, ordinal data as if it were numeric.

∗ A foundational stats mistake: Interpreting a significant *p*-value as evidence that an effect exists in the real world.

How you'll avoid it

When a variable comes from a Likert scale, tell R it's categorical using `factor()`.

Apply and interpret ordinal regression models (e.g., `polr()` from `MASS`).

Understand that significant *p*-values can arise even if no effect exists in the real world.

The mistake

</> **A common R mistake:** Letting R treat all variables that look like numbers as continuous numeric.

An advanced stats mistake: Modelling categorical, ordinal data as if it were numeric.

∗ **A foundational stats mistake:** Interpreting a significant *p*-value as evidence that an effect exists in the real world.

How you'll avoid it

When a variable comes from a Likert scale, tell R it's categorical using `factor()`.

Apply and interpret ordinal regression models (e.g., `polr()` from `MASS`).

Understand that significant *p*-values can arise even if no effect exists in the real world.

Thank you! 😊 Time for questions!

Some really nice resources

- Jamieson's (2004) paper **Likert scales: How to (ab)use them.**
- UCLA Statistical Methods and Data Analytics's web page **Ordinal Logistic Regression.**
- Kurz' (2021) blog post **Notes on the Bayesian cumulative probit.**
- Vasishth and Nicenboim's (2016) paper **Statistical Methods for Linguistic Research: Foundational Ideas – Part I.**
- Gelman and Hill's (2007) book **Data Analysis Using Regression and Multilevel/Hierarchical Models.**

Plot on Slide 2 from

Reeder, P. A., Newport, E. L., & Aslin, R. N. (2017). Distributional learning of subcategories in an artificial grammar: Category generalization and subcategory restrictions. *Journal of Memory and Language*, 97, 17–29.

Data from

Terry, J., Ross, R. M., Nagy, T., Salgado, M., Garrido-Vásquez, P., Sarfo, J. O., Cooper, S., Buttner, A. C., Lima, T. J. S., Öztürk, İ., Akay, N., Santos, F. H., Artemenko, C., Copping, L. T., Elsherif, M. M., Milovanović, I., Cribbie, R. A., Drushlyak, M. G., Swainston, K., … Field, A. P. (2023). Data from an International Multi-Centre Study of Statistics and Mathematics Anxieties and Related Variables in University Students (the SMARVUS Dataset). *Journal of Open Psychology Data*, 11(1), 8.