

ANALYSIS OF MULTIPLE GROUPS WITH ANOVA

FIGHTING JET LAG



THE KNEE'S THE THING!

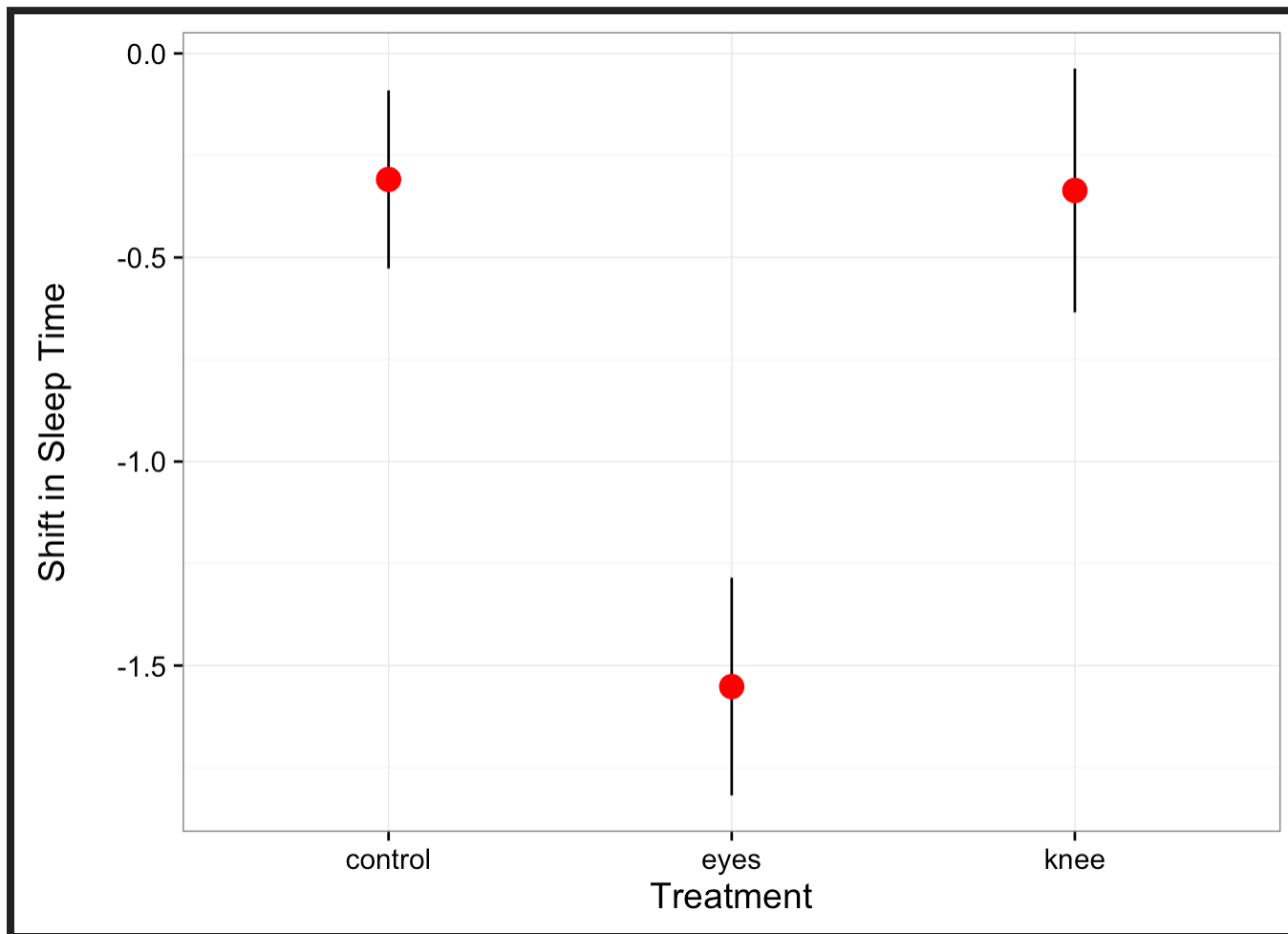


- Test if shining light behind knees or on eyes helped jet lag
- Also a control group with no light
- $n=7$

THE STEPS OF STATISTICAL MODELING

1. What is your question?
2. What model of the world matches your question?
3. Build a test
4. Evaluate test assumptions
5. Evaluate test results
6. Visualize

WHAT QUESTION WOULD YOU ASK OF THIS DATA?



MANY QUESTIONS

1. Are groups different from 0?
2. Are groups different from each other?
3. Are groups a meaningful explanatory variable?

INTRODUCING ANOVA: COMPARING VARIATION

Central Question: Is the data explained by the data generating process greater than that explained by the error generating process?

Test: Is a ratio of variability from data generating process v. error generating process large?

Ratio of two normal distributions = F Distribution

HOW DOES THE WORLD WORK?

$$y_i = T_{control}\mu_{control} + T_{eye}\mu_{eye} + T_{knee}\mu_{knee} + \epsilon_i$$

$$T = \{0, 1\}$$

$$\epsilon_i \sim N(0, \sigma)$$

LINKING YOUR MODEL TO YOUR QUESTION

Data Generating Process:

$$T_{control}\mu_{control} + T_{eye}\mu_{eye} + T_{knee}\mu_{knee}$$

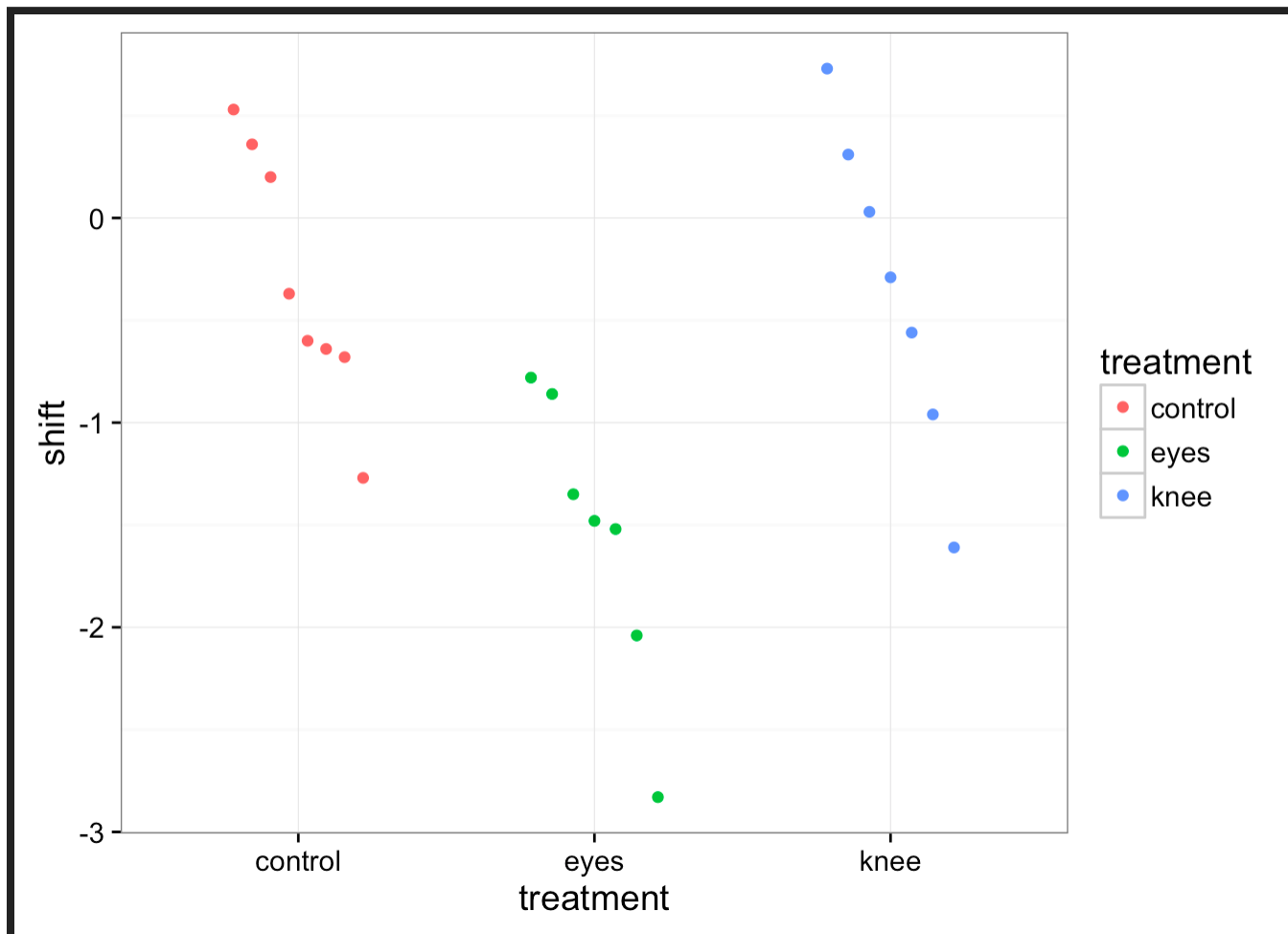
VERSUS

Error Generating Process

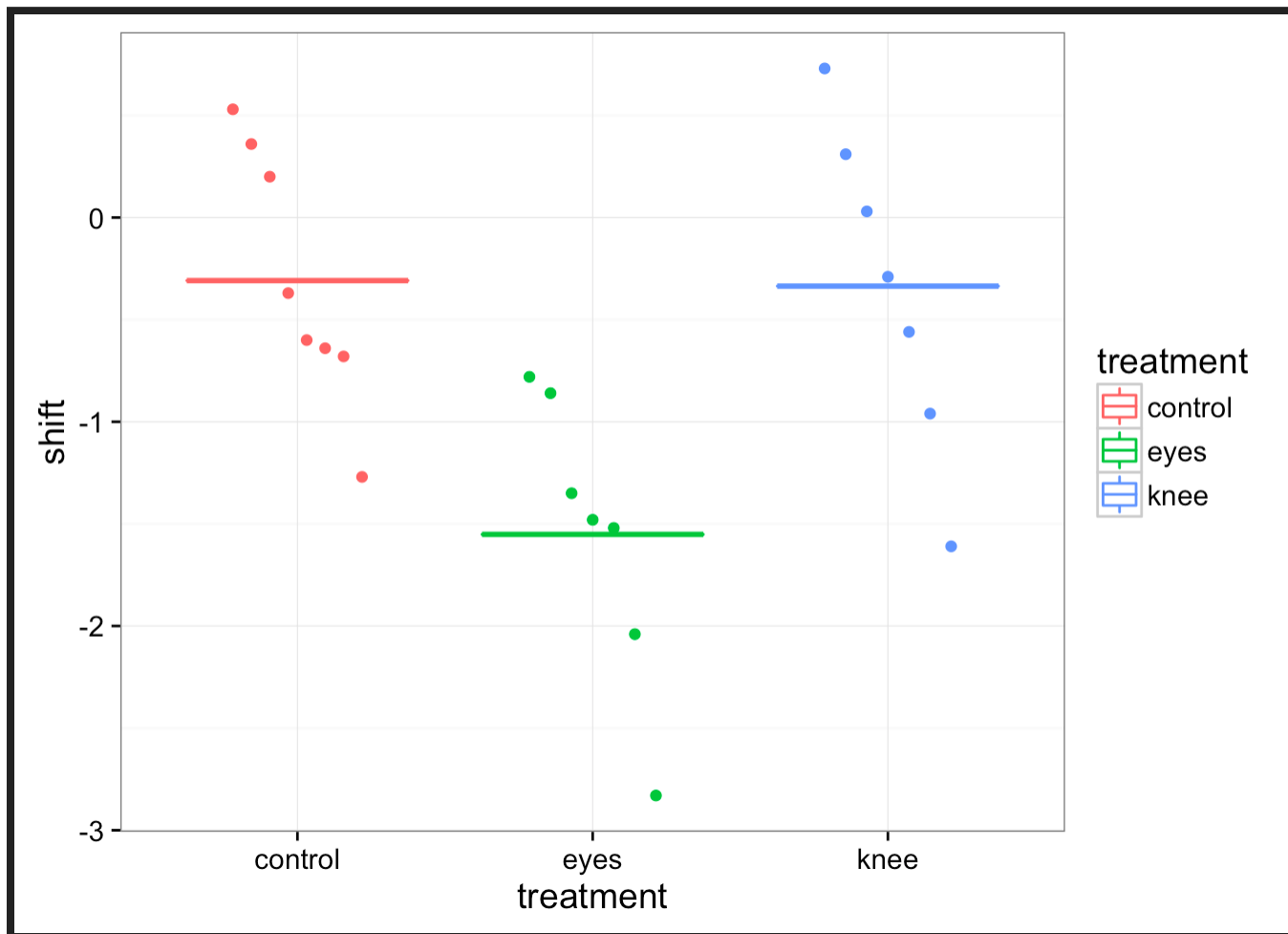
$$\epsilon_i \sim N(0, \sigma)$$

If groups are a meaningful explanatory variable, what does that imply about variability in the data?

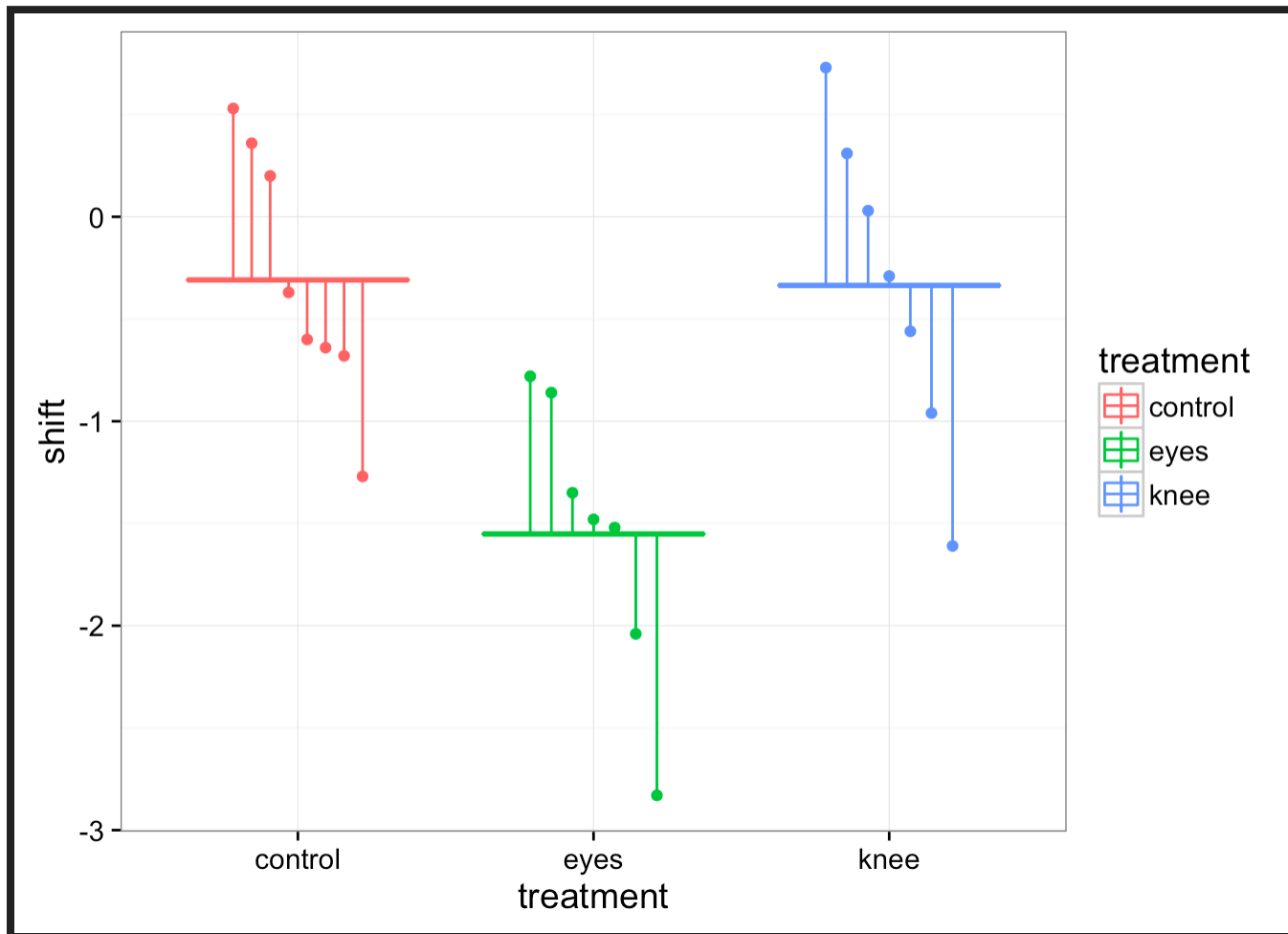
VARIABILITY DUE TO DGP VERSUS EGP



VARIABILITY DUE TO DGP VERSUS EGP



VARIABILITY DUE TO DGP VERSUS EGP



THE STEPS OF STATISTICAL MODELING

1. What is your question?
2. What model of the world matches your question?
3. **Build a test**
4. Evaluate test assumptions
5. Evaluate test results
6. Visualize

BUILD A LINEAR MODEL

- Our model states that shift is determined by treatment
- This relationship is linear with a normal error distribution

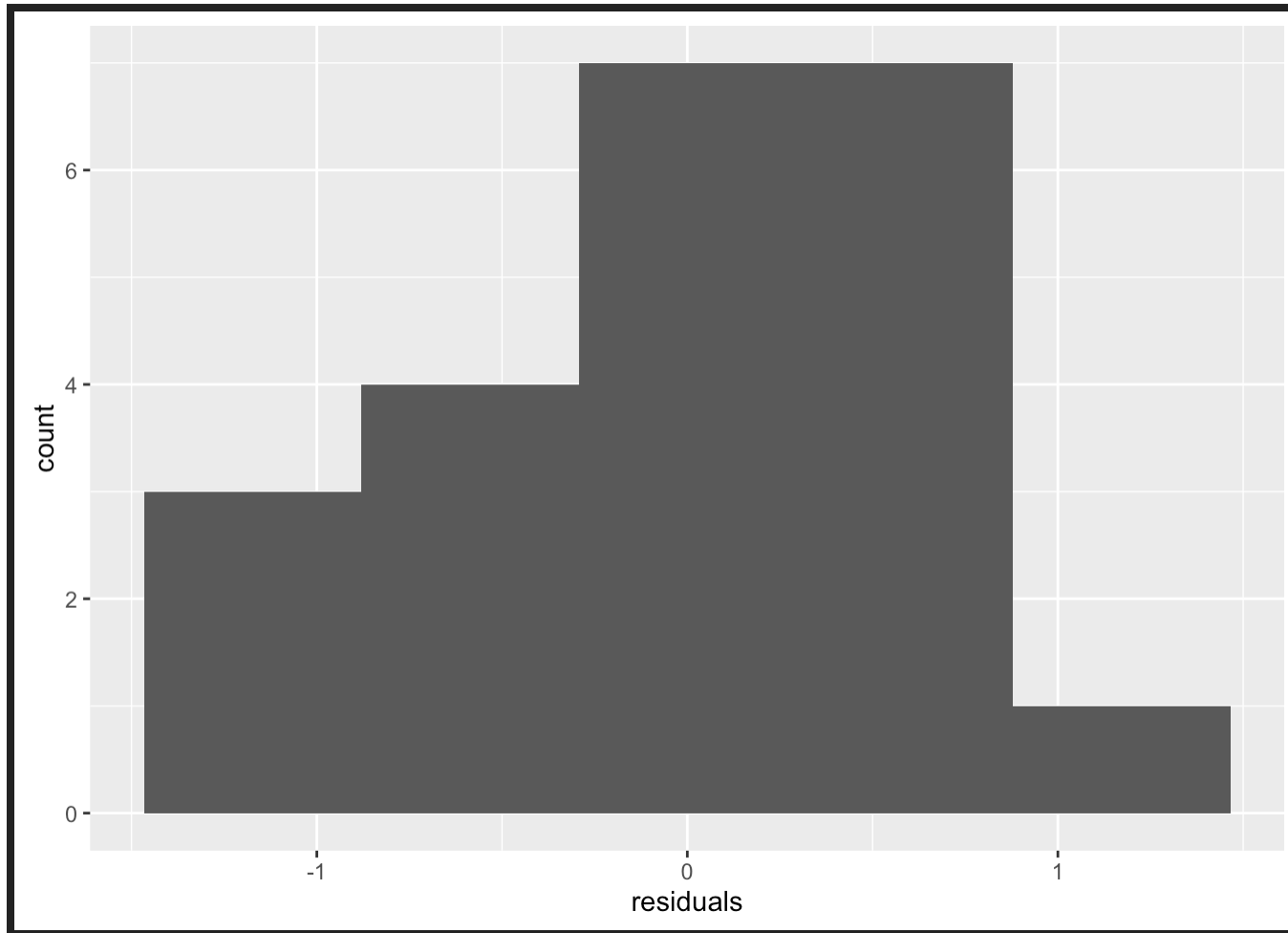
```
mod <- lm(shift ~ treatment, data=jetlag)
```

Note: The ~ syntax can also be used for `t.test`

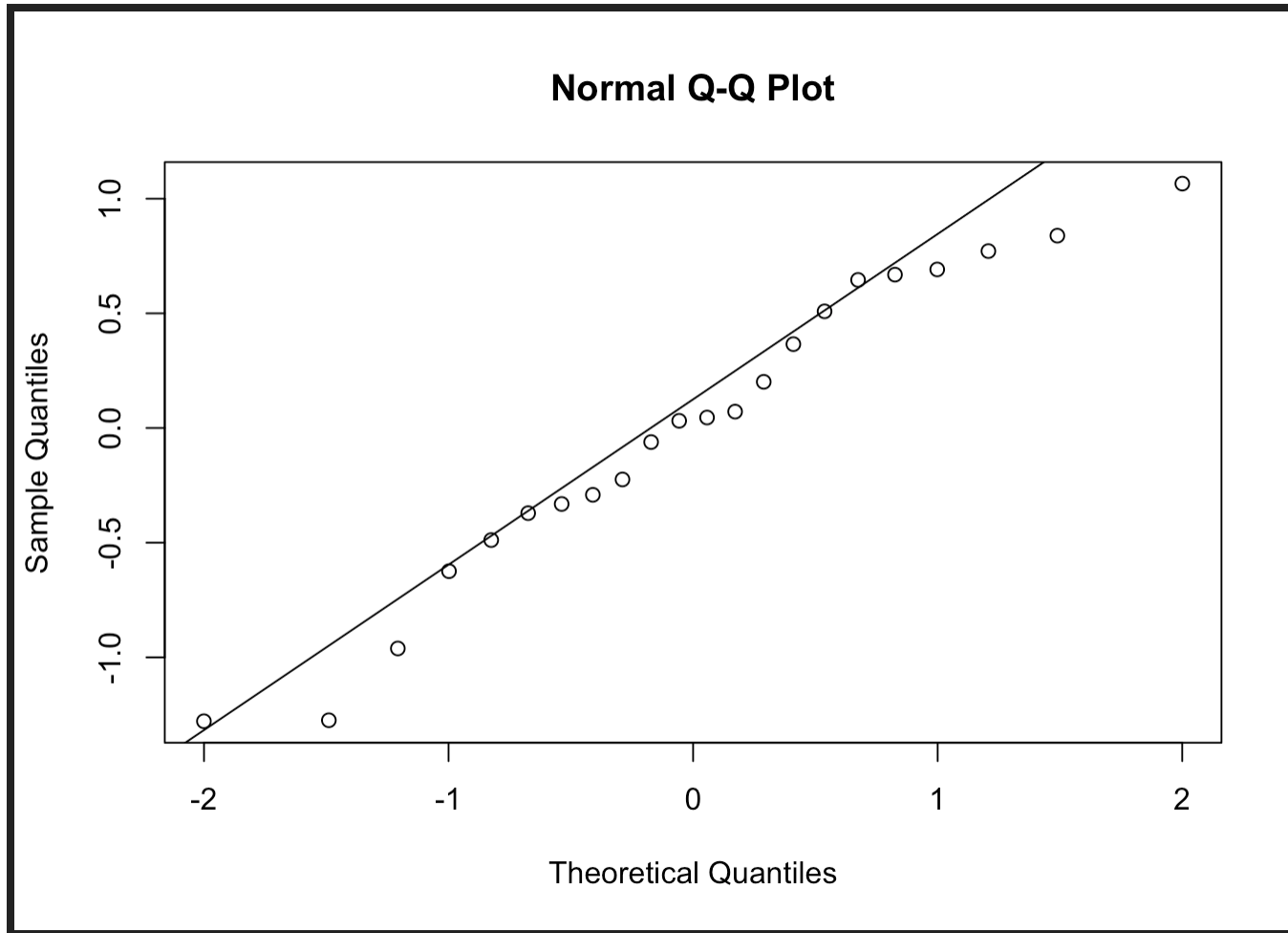
THE STEPS OF STATISTICAL MODELING

1. What is your question?
2. What model of the world matches your question?
3. Build a test
4. Evaluate test assumptions
5. Evaluate test results
6. Visualize

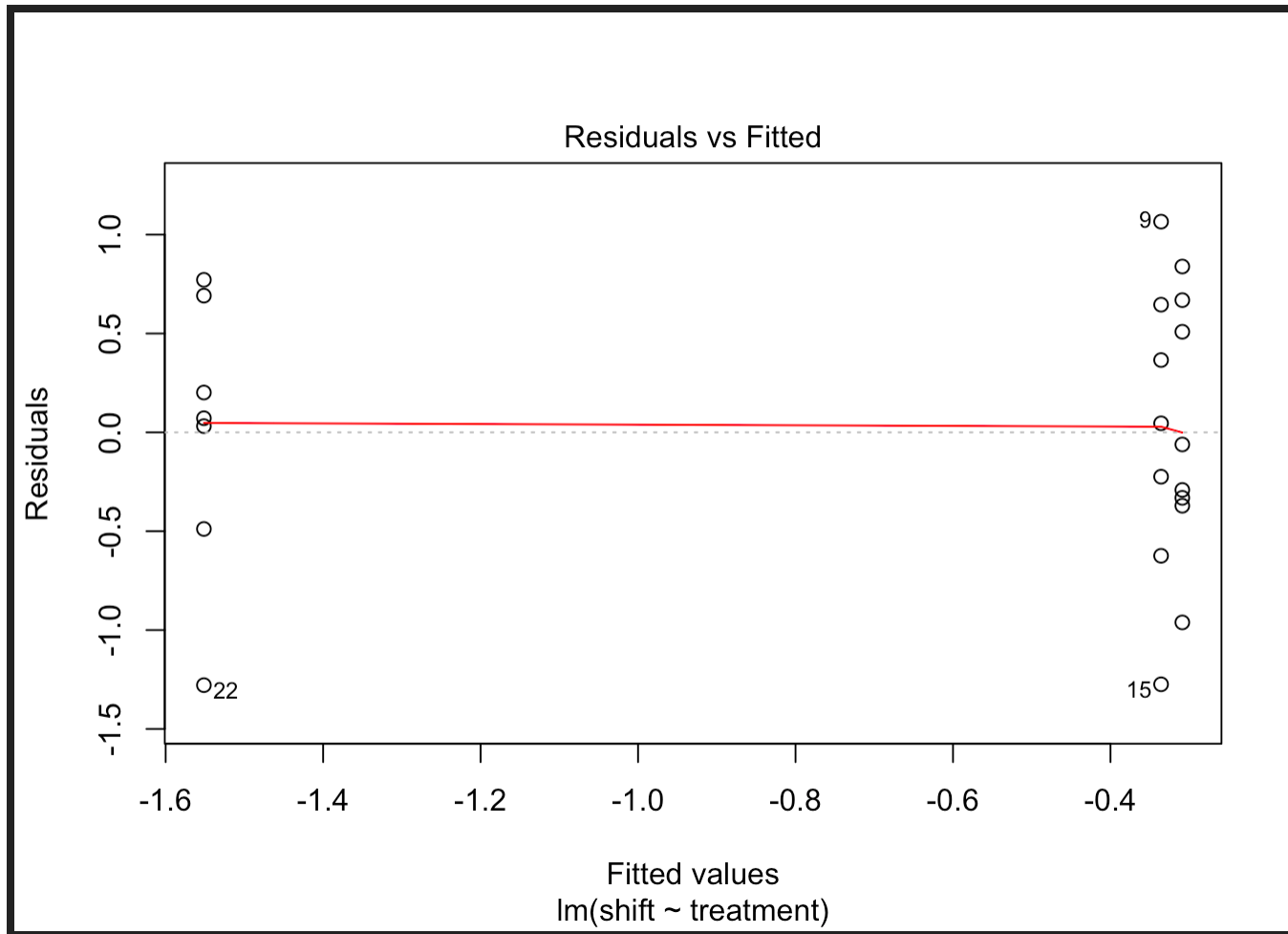
HISTOGRAM OF RESIDUALS



QQ PLOT



RESIDUALS VERSUS FITTED



THE STEPS OF STATISTICAL MODELING

1. What is your question?
2. What model of the world matches your question?
3. Build a test
4. Evaluate test assumptions
5. Evaluate test results
6. Visualize

ANOVA AND F-TESTS

F-Tests	T-Tests
Tests if data generating process different than error	Tests if parameter is different from 0

Essentially comparing a variation explained by a model with versus without a data generating process included.

F-TESTS

$F = \text{Mean Square Variability Explained by Model} / \text{Mean Square Error}$

DF for Numerator = $k-1$ DF for Denominator = $n-k$ k = number of groups, n = sample size

F-TABLE

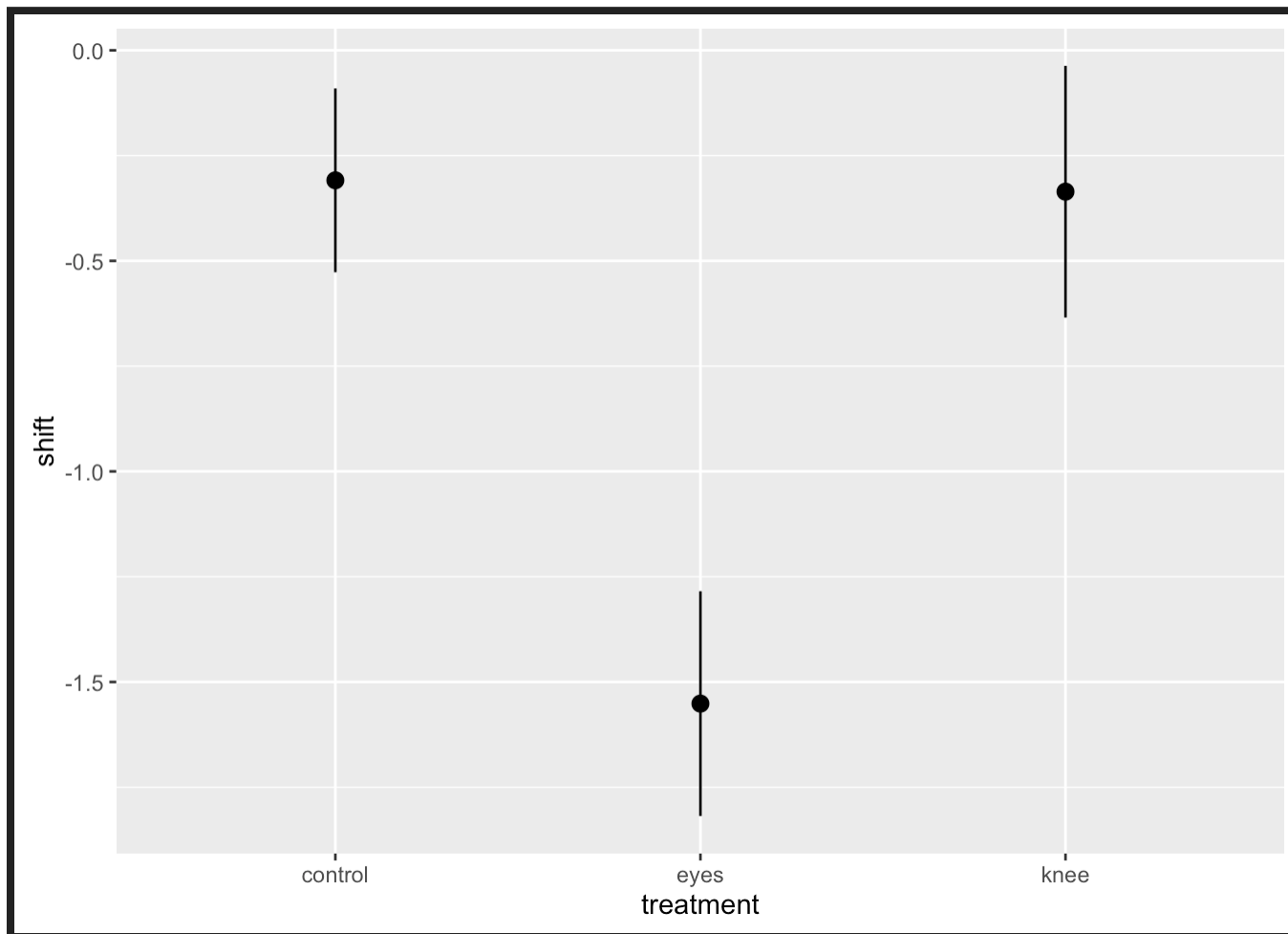
anova (mod)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
treatment	2	7.224492	3.6122459	7.289449	0.0044723
Residuals	19	9.415345	0.4955445	NA	NA

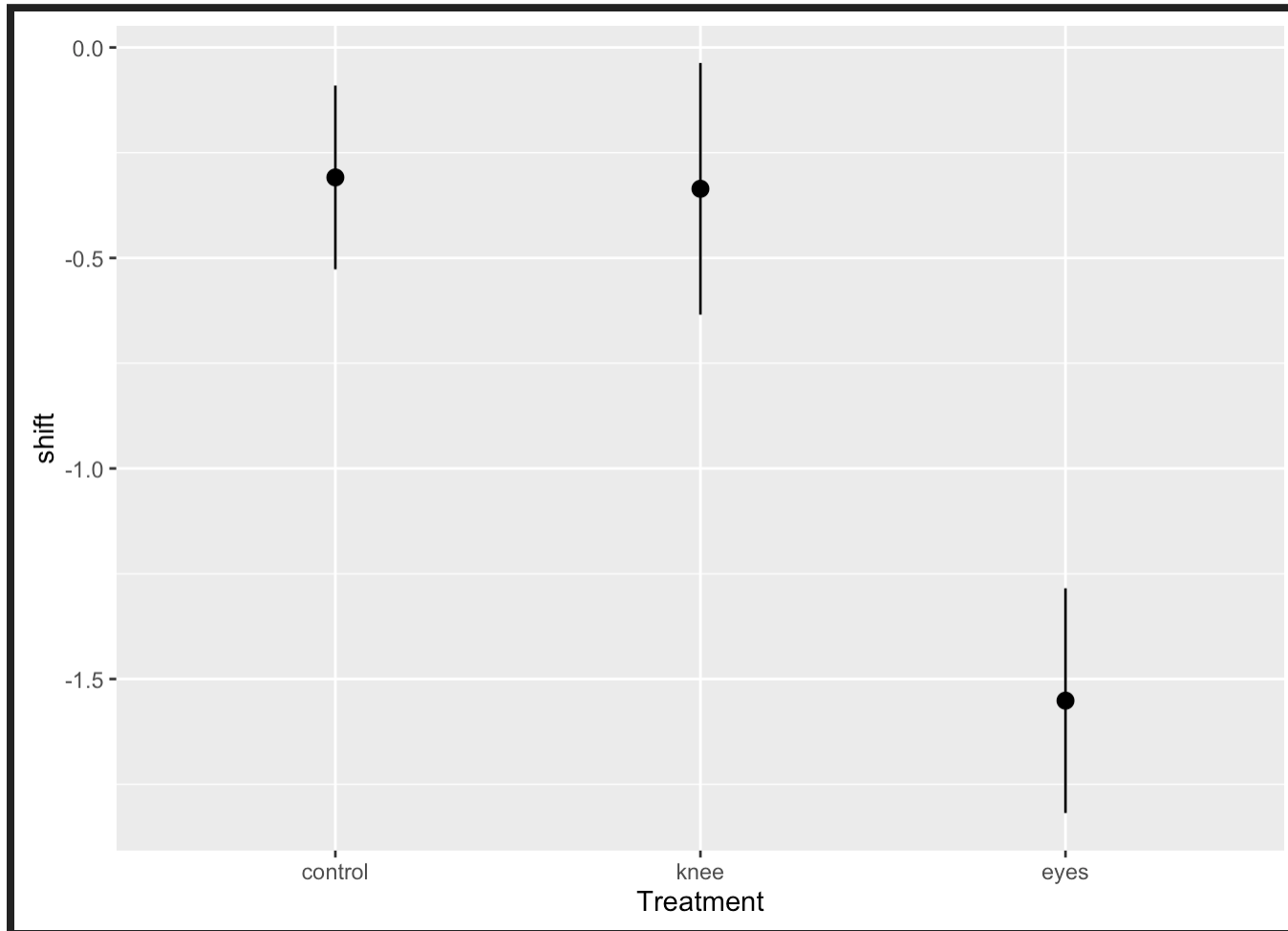
THE STEPS OF STATISTICAL MODELING

1. What is your question?
2. What model of the world matches your question?
3. Build a test
4. Evaluate test assumptions
5. Evaluate test results
6. Visualize

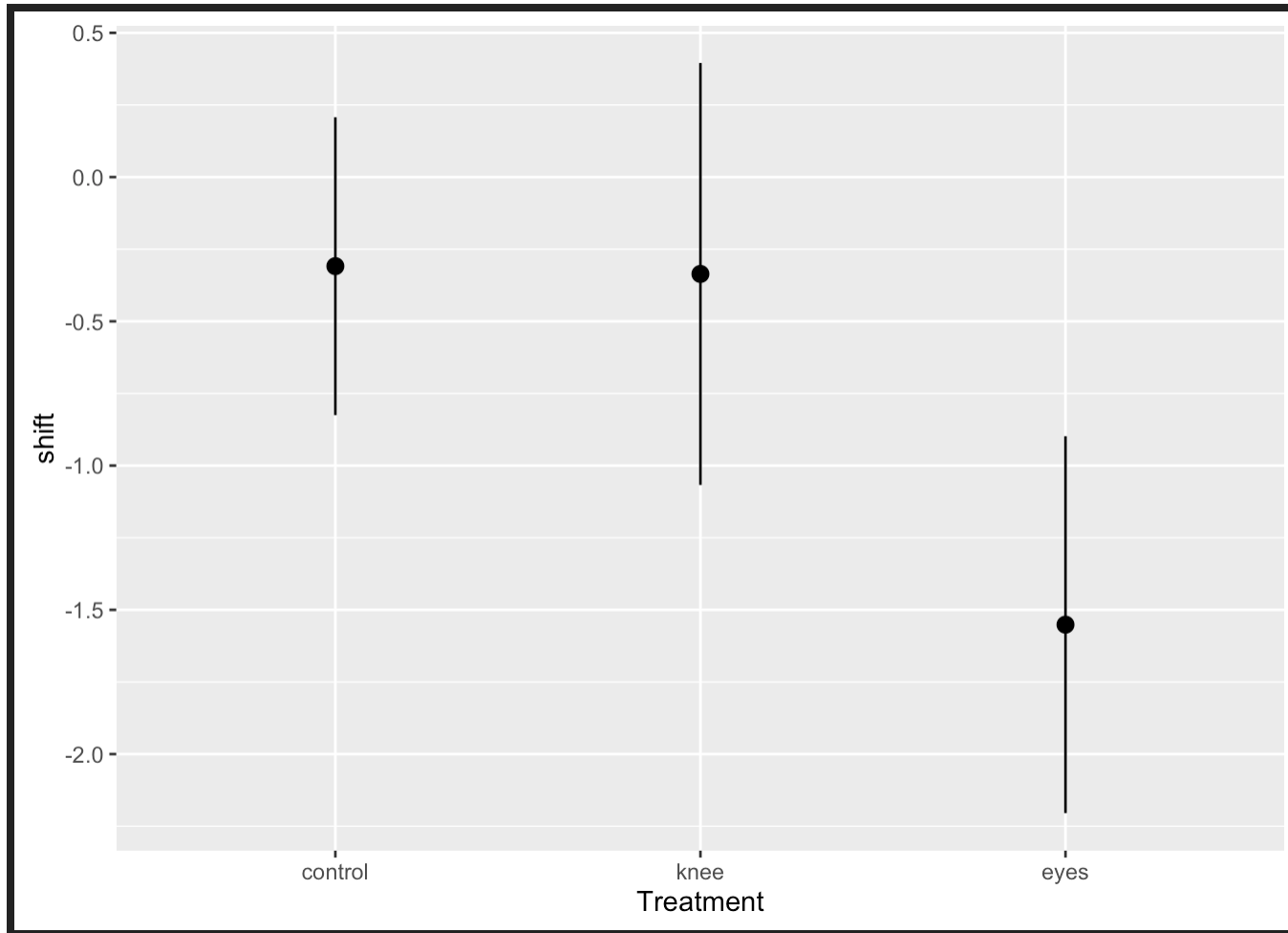
VISUALIZE THE MODEL MEANS AND SE



MUCK WITH ORDER, COLORS, ETC.



SHOW CI INSTEAD OF SE



FADED EXAMPLES

- Let's walk through an example
- Then I'll give you an example to work through, but with some code edited out

THIS ANALYSIS

```
jetlag <- read.csv("./anova/15e1KneesWhoSayNight.csv")

#fit
jetlag_mod <- lm(shift ~ treatment, data=jetlag)

#Assumptions
plot(jetlag_mod, which=1)
plot(jetlag_mod, which=2)

#Results
anova(jetlag_mod)

#Visualize
ggplot(data = jetlag, mapping=aes(x=treatment, y=shift)) +
  stat_summary(fun.data=mean_se)
```

NEMATODE LIFESPANS!

```
nemetodes <- read.csv("./anova/15q19NematodeLifespan.csv")

#fit
nem_mod <- lm(_____ ~ _____, data=nemetodes)

#Assumptions
plot(nem_mod, which=1)
plot(nem_mod, which=2)

#Results
anova(_____)

#Visualize
ggplot(data = nemetodes, mapping=aes(x=_____, y=_____)) +
  stat_summary(fun.data=mean_se)
```

EELGRASS GENOTYPES

```
eelgrass <- read.csv("../anova/EelgrassGenotypes.csv")
```

```
#fit
```

```
eelgrass_mod <- lm(_____, _____)
```

```
#Assumptions
```

```
plot(_____, which=1)
```

```
plot(_____, which=2)
```

```
#Results
```

```
_____ (_____)
```

```
#Visualize
```

```
ggplot(data = _____, mapping=aes(x=_____, y=_____)) +
```

```
_____ (_____)
```
