0:00 / 3:57

Comparing means adjusted for other predictors (analysis of covariance)

Professor Andy Field

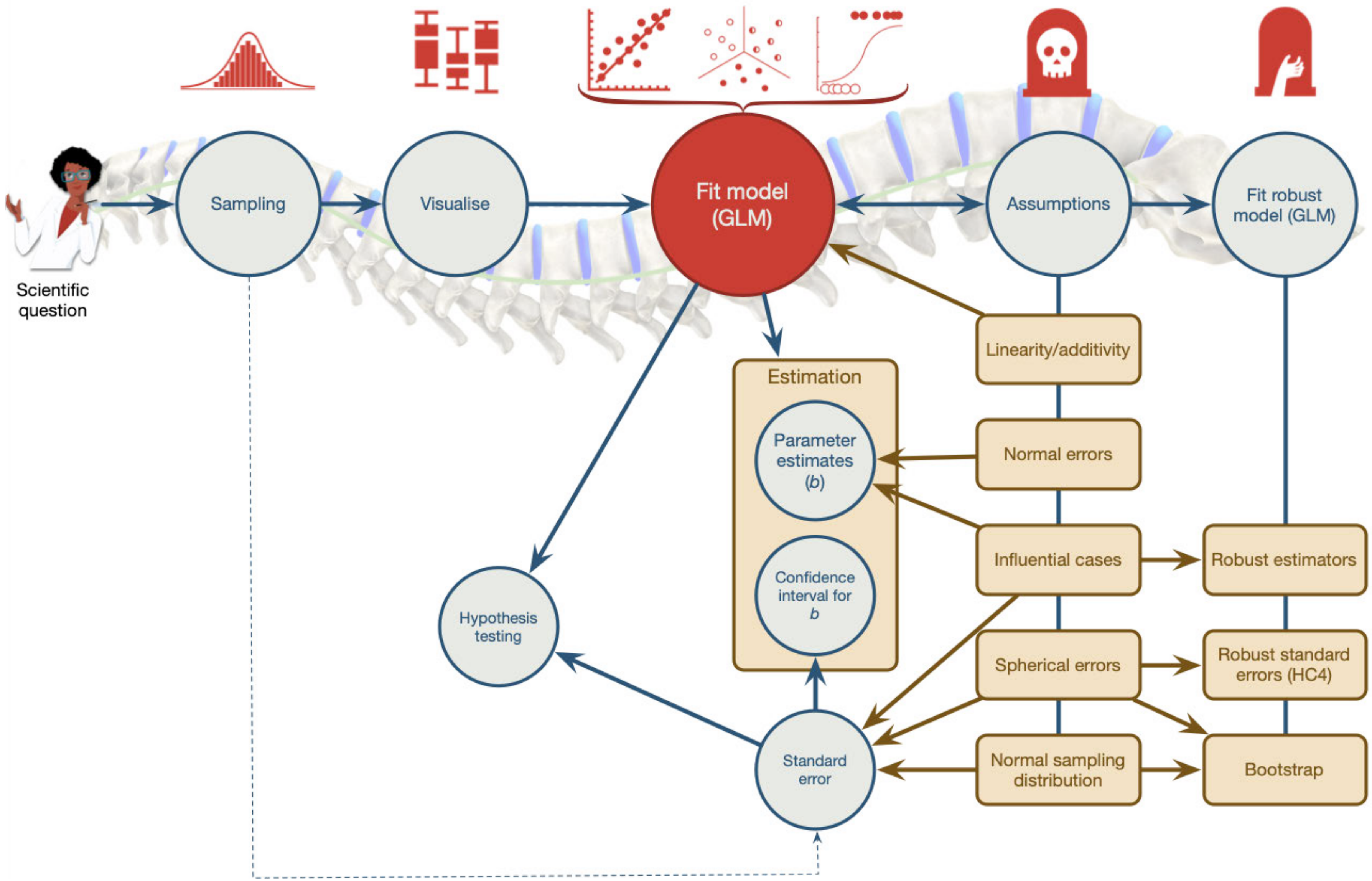
@profandyfield

www.youtube.com/user/ProfAndyField/

www.discoveringstatistics.com

www.milton-the-cat.rocks

www.discovr.rocks


Scientific question
Sampling
Visualise
Fit model (GLM)
Assumptions
Fit robust model (GLM)
Estimation
Parameter estimates (b)
Confidence interval for b
Hypothesis testing
Standard error
Linearity/additivity
Normal errors
Influential cases
Spherical errors
Normal sampling distribution
Robust estimators
Robust standard errors (HC4)
Bootstrap

Learning outcomes

- Explain how to compare means adjusting for other predictors using a linear model
    - Linear model with a categorical and continuous predictor
    - a.k.a. analysis of covariance (ANCOVA)
- Type I vs. Type III sums of squares
- Interpreting the model
    - Main effects
    - Covariates

When and why

Generally

- To test for differences between group means when we know that an extraneous variable affects the outcome variable
- Used to adjust the means for extraneous and confounding variables

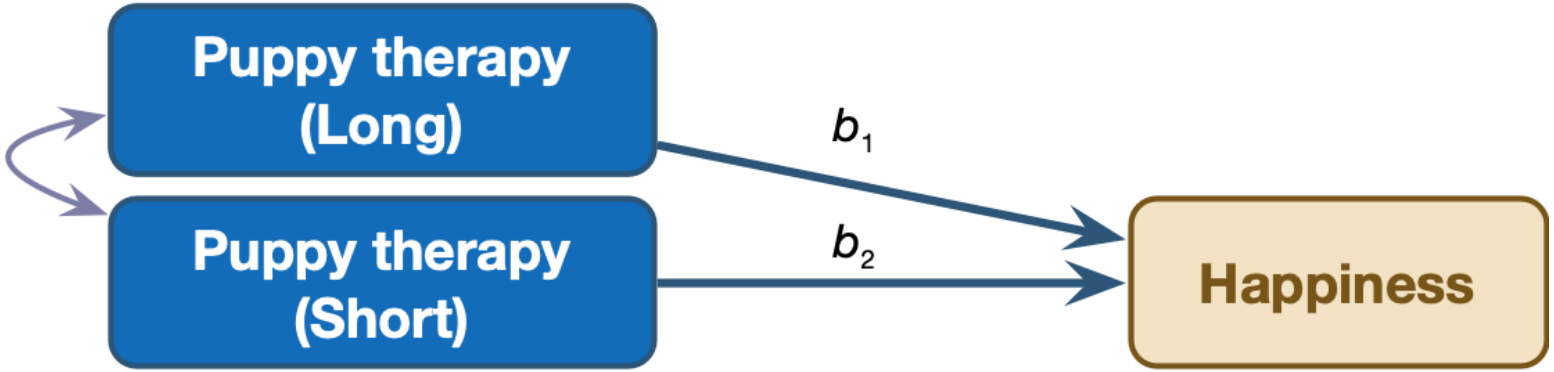
In experimental research

- Reduce error variance (sometimes)
  - By explaining some of the unexplained variance (SS_R) the error variance in the model can be reduced
- Greater experimental control
  - By adjusting for known confounds, we can gain greater insight into the effect of the predictor variable(s)

Extending the puppy example

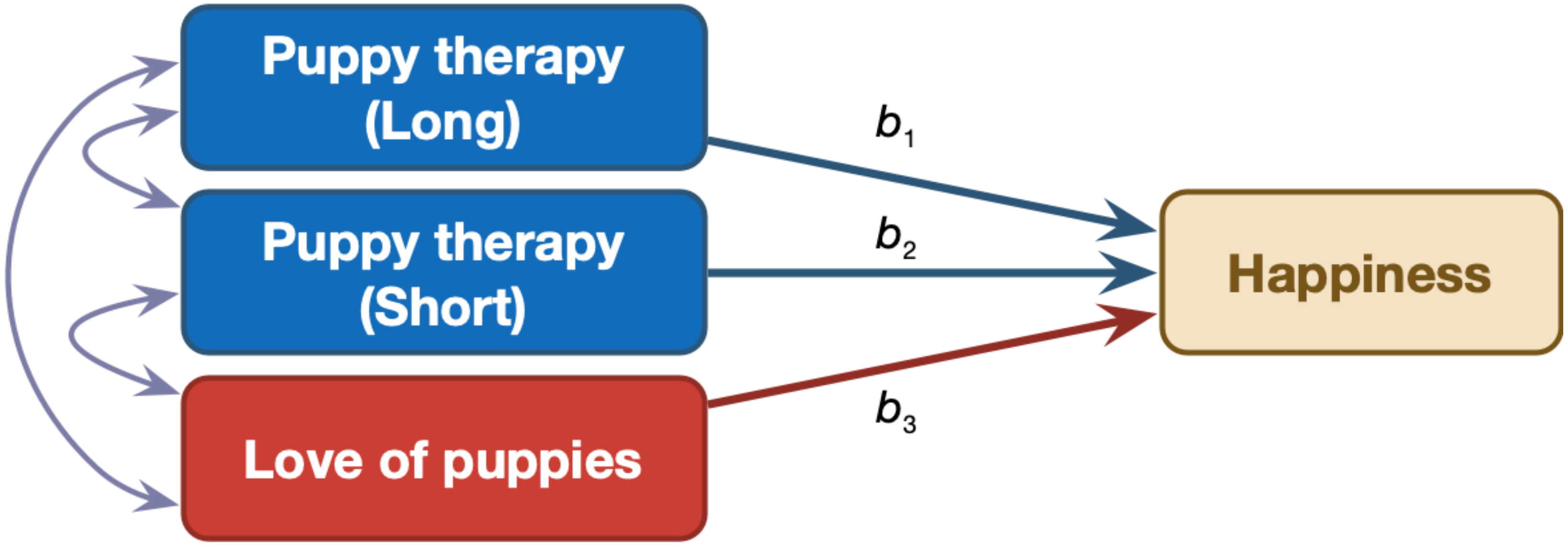
- A puppy therapy RCT
  - A no puppies control group
  - 15 minutes of puppy therapy
  - 30 minutes of puppy therapy
- Outcome variable
  - Happiness (0 = unhappy to 10 = happy).
- Covariate
  - Love of puppies (0 = no love, 7 = all the love)

The statistical model



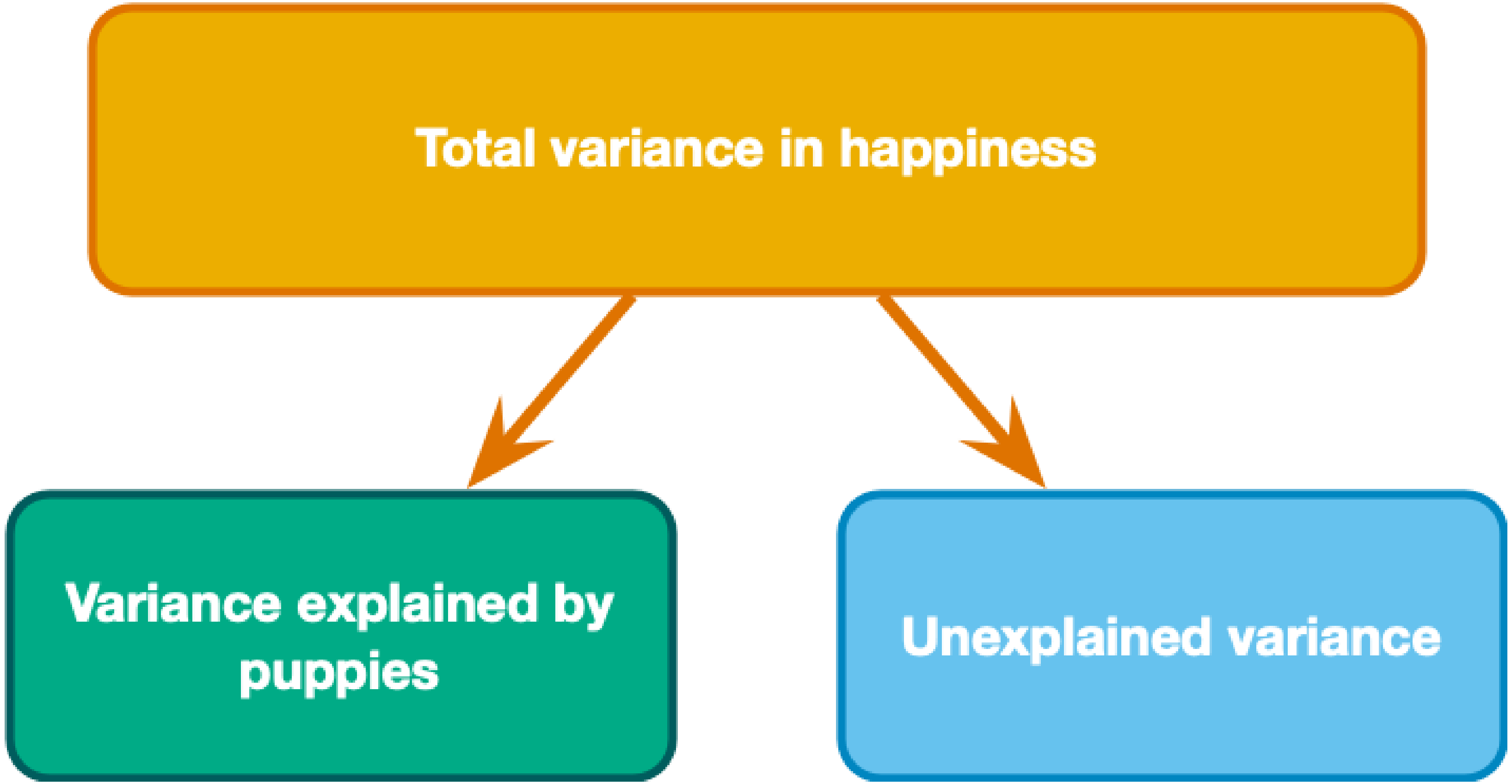
$$\widehat{\text{Happiness}}_i = \hat{b}_0 + \hat{b}_1\text{Long}_i + \hat{b}_2\text{Short}_i + e_i$$

The statistical model

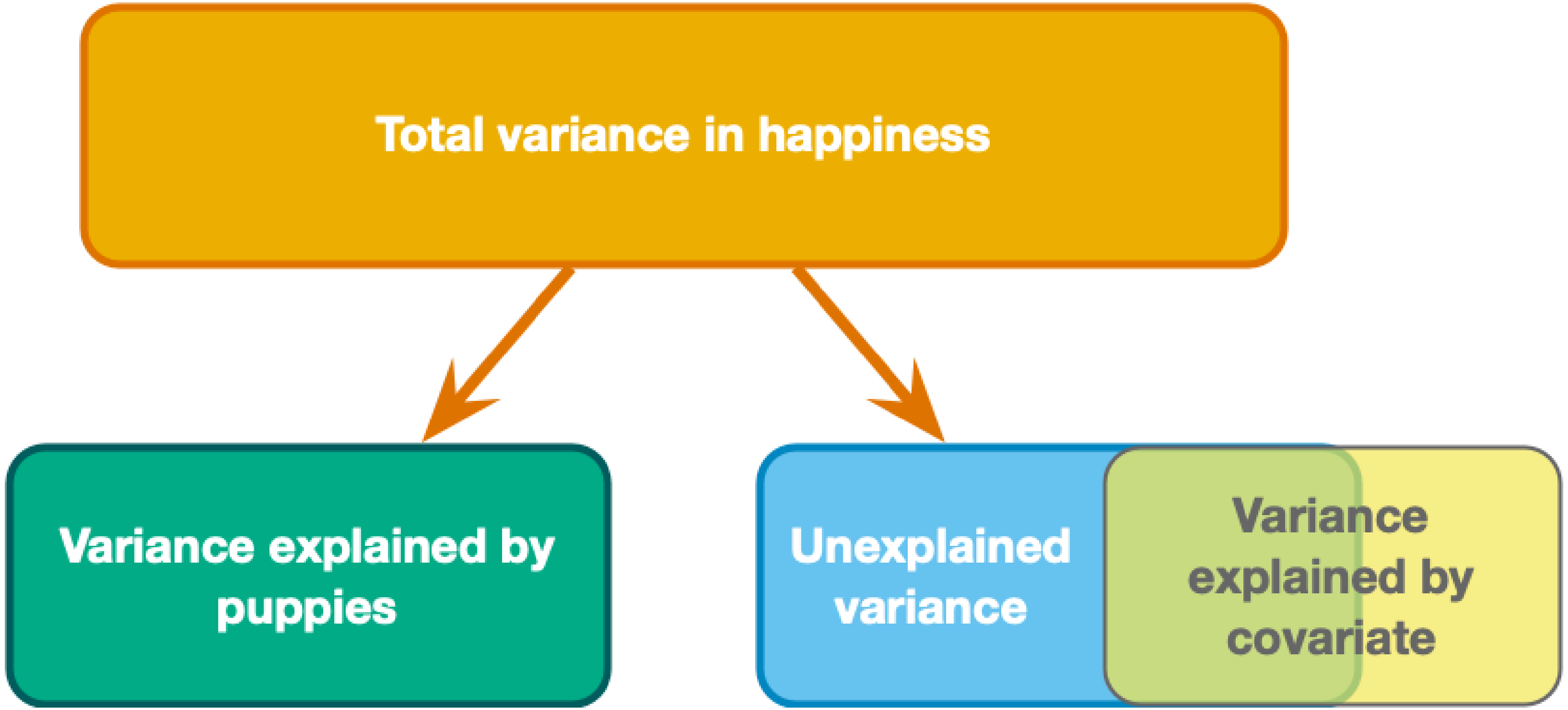


$$\widehat{\text{Happiness}}_i = \hat{b}_0 + \hat{b}_1\text{Long}_i + \hat{b}_2\text{Short}_i + \hat{b}_3\text{Puppy love}_i + e_i$$

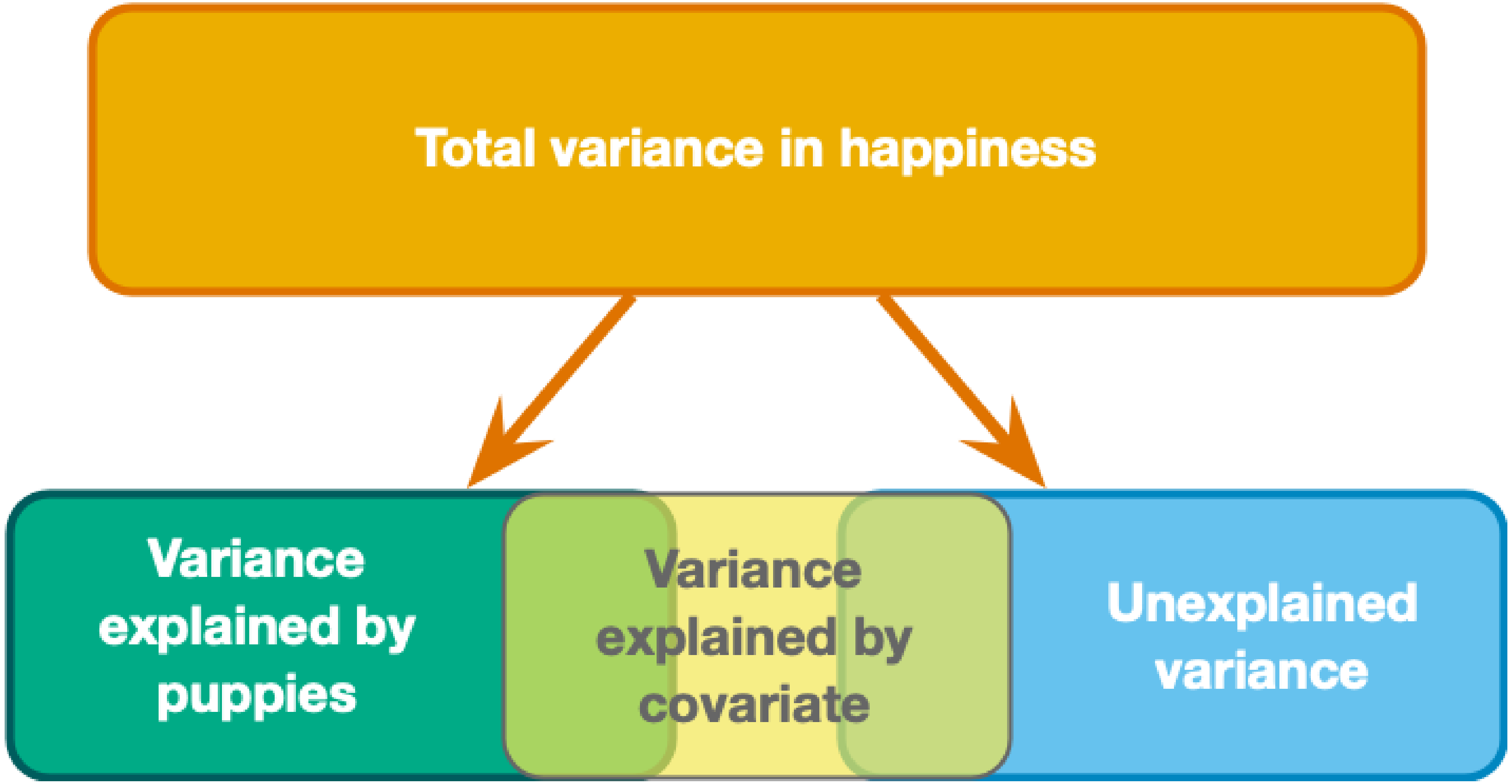
Partitioning variance



Partitioning variance



Partitioning variance



The data

Table 1: Data for the puppy therapy example

	id	dose	happiness	puppy_love
1	72kfu0	No puppies	3	4
2	2l47nl	No puppies	2	1
3	uev2x2	No puppies	5	5
4	fbnym1	No puppies	2	1
5	i3d8d1	No puppies	2	2
6	lg6i36	No puppies	2	2
7	22o7c9	No puppies	7	7
8	5223ie	No puppies	2	4
9	v596m4	No puppies	4	5
10	w36156	15 mins	7	5

Previous 1 2 3 Next

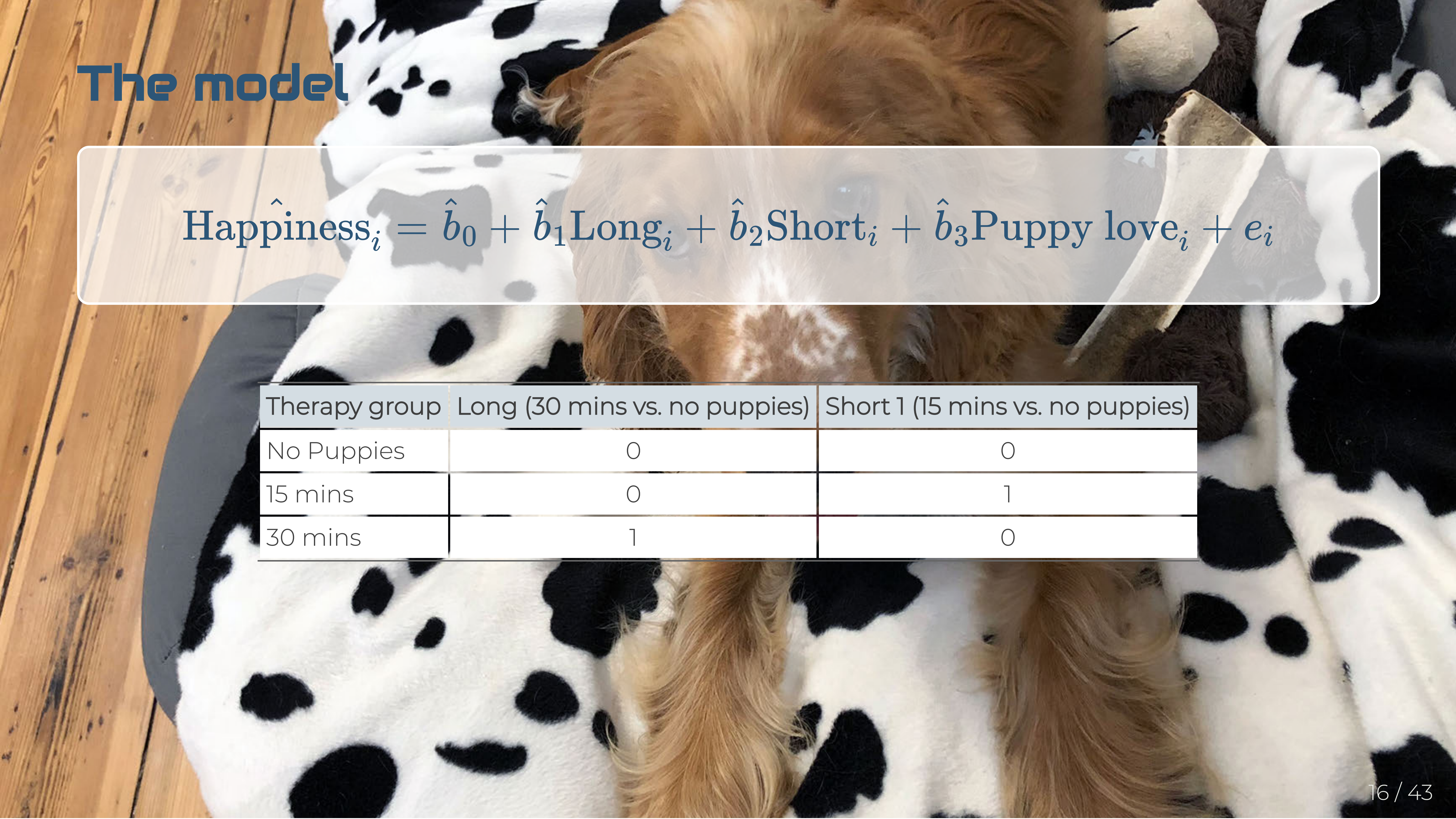
Data summary

By group

Therapy group	Mean (happiness)	SD (happiness)	Mean (puppy love)	SD (puppy love)
No puppies	3.22	1.79	3.44	2.07
15 mins	4.88	1.46	3.12	1.73
30 mins	4.85	2.12	2.00	1.63

Overall

Mean (happiness)	SD (happiness)	Mean (puppy love)	SD (puppy love)
4.37	1.96	2.73	1.86



The model

$$\hat{\text{Happiness}}_i = \hat{b}_0 + \hat{b}_1\text{Long}_i + \hat{b}_2\text{Short}_i + \hat{b}_3\text{Puppy love}_i + e_i$$

Therapy group	Long (30 mins vs. no puppies)	Short 1 (15 mins vs. no puppies)
No Puppies	0	0
15 mins	0	1
30 mins	1	0

What you'd expect the dummy variables to represent

Therapy group	Mean (happiness)
No puppies	3.22
15 mins	4.88
30 mins	4.85

$$
\begin{aligned}
\hat{b}_0 &= \bar{X}_{\text{No puppies}} = 3.22 \\
\hat{b}_1 &= 4.85 - 3.22 = 1.63 \\
\hat{b}_2 &= 4.88 - 3.22 = 1.66
\end{aligned}
$$

What you actually get

```
pupluv_lm <- lm(happiness ~ puppy_love + dose, data = pupluv_tib)
broom::tidy(pupluv_lm, conf.int = TRUE)
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	1.789	0.867	2.063	0.049	0.007	3.572
puppy_love	0.416	0.187	2.227	0.035	0.032	0.800
dose15 mins	1.786	0.849	2.102	0.045	0.040	3.532
dose30 mins	2.225	0.803	2.771	0.010	0.575	3.875

$$\widehat{\text{Happiness}}_i = 1.789 + 2.225\ \text{Long}_i + 1.786\ \text{Short}_i + 0.416\ \text{Puppy love}_i + e_i$$

Adjusting means

$$\hat{\text{Happiness}}_i = 1.789 + 2.225\ \text{Long}_i + 1.786\ \text{Short}_i + 0.416\ \text{Puppy love}_i + e_i$$

Mean (happiness)	SD (happiness)	Mean (puppy love)	SD (puppy love)
4.37	1.96	2.73	1.86

Control group

$$\begin{aligned}
\hat{\text{Happiness}}_i &= 1.789 + 2.225\ \text{Long}_i + 1.786\ \text{Short}_i + 0.416\ \text{Puppy love}_i + e_i \\
&= 1.789 + (2.225 \times 0) + (1.786 \times 0) + (0.416\ \text{Puppy love}_i) + e_i \\
&= 1.789 + (0.416 \times \bar{X}_{\text{Puppy love}}) + e_i \\
&= 1.789 + (0.416 \times 2.73) + e_i \\
&= 2.925
\end{aligned}$$

Adjusting means

15 minute group

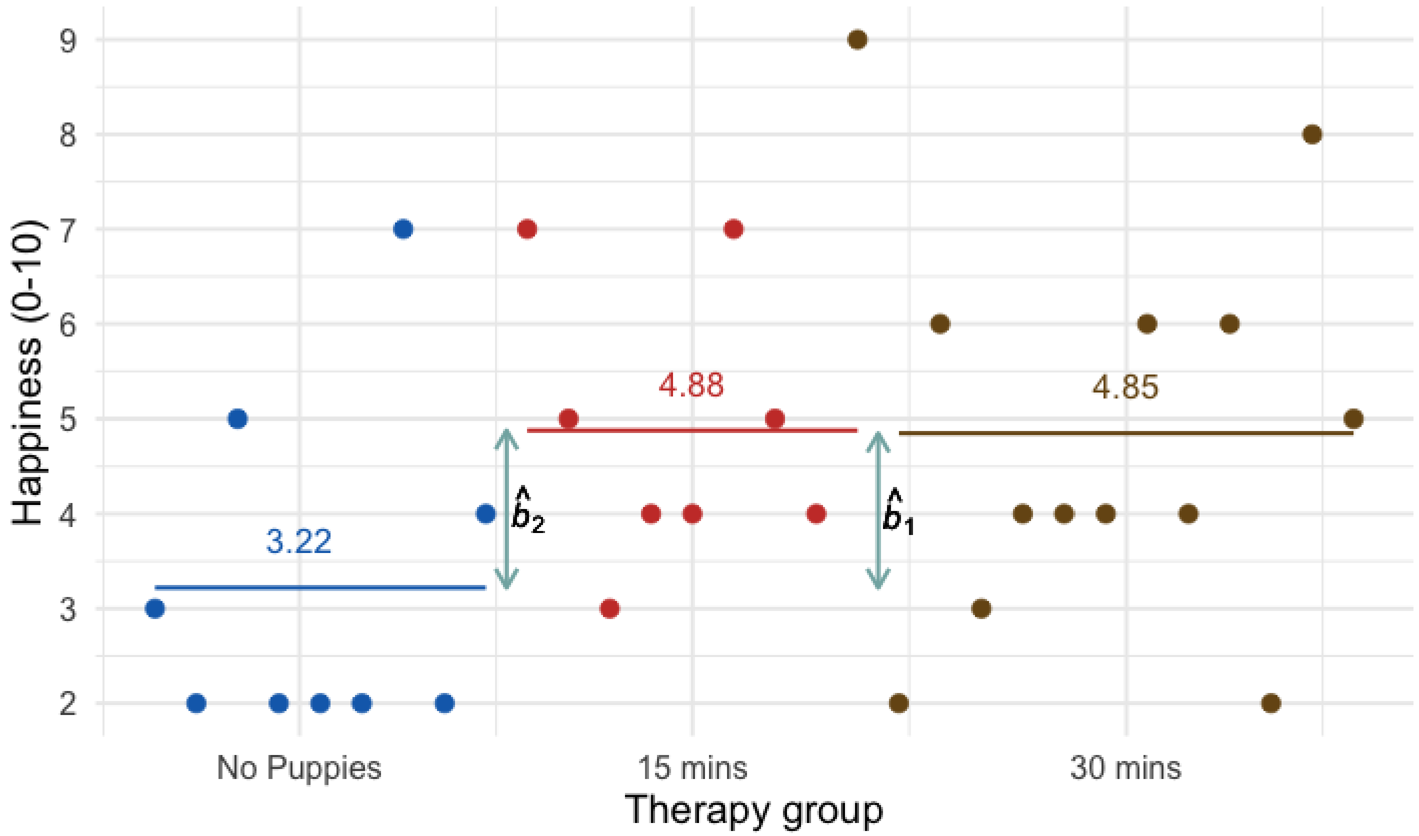
$$\begin{aligned}\widehat{\text{Happiness}}_i &= 1.789 + 2.225\ \text{Long}_i + 1.786\ \text{Short}_i + 0.416\ \text{Puppy love}_i + e_i \\ &= 1.789 + (2.225 \times 0) + (1.786 \times 1) + (0.416\ \text{Puppy love}_i) + e_i \\ &= 1.789 + 1.786 + (0.416 \times 2.73) + e_i \\ &= 4.71\end{aligned}$$

30 minute group

$$\begin{aligned}\widehat{\text{Happiness}}_i &= 1.789 + 2.225\ \text{Long}_i + 1.786\ \text{Short}_i + 0.416\ \text{Puppy love}_i + e_i \\ &= 1.789 + (2.225 \times 1) + (1.786 \times 0) + (0.416\ \text{Puppy love}_i) + e_i \\ &= 1.789 + 2.225 + (0.416 \times 2.73) + e_i \\ &= 5.15\end{aligned}$$

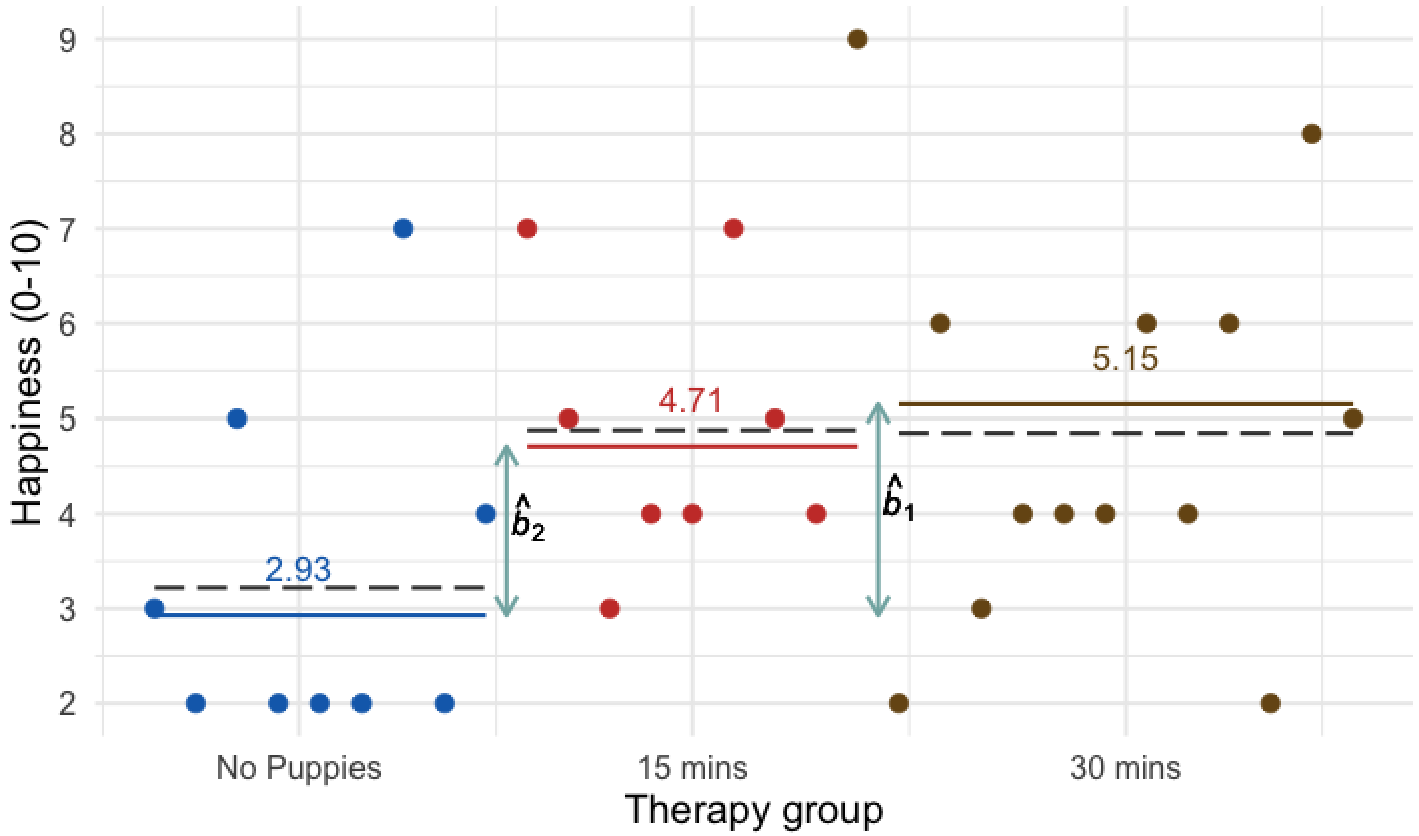
The unadjusted model

$$\hat{\text{Happiness}}_i = \hat{b}_0 + \hat{b}_1\text{Long}_i + \hat{b}_2\text{Short}_i + e_i$$



The adjusted model

$$\widehat{\text{Happiness}}_i = \hat{b}_0 + \hat{b}_1\text{Long}_i + \hat{b}_2\text{Short}_i + \hat{b}_3\text{Puppy love}_i + e_i$$



Model parameters

Therapy group	Mean (happiness)	Adjusted Mean (happiness)
No puppies	3.22	2.93
15 mins	4.88	4.71
30 mins	4.85	5.15

$$\hat{b}_1 = 5.15 - 2.93 = 2.22$$
$$\hat{b}_2 = 4.71 - 2.93 = 1.78$$

term	estimate	std.error	statistic	p.value
(Intercept)	1.789	0.867	2.063	0.049
puppy_love	0.416	0.187	2.227	0.035
dose15 mins	1.786	0.849	2.102	0.045
dose30 mins	2.225	0.803	2.771	0.010

The *F*-statistic with multiple predictors

- The *F*-statistic is calculated using sums of squares
- Type I (sequential)
    - The default in R
    - Each predictor is evaluated taking account of previous predictors
    - The order of predictors matters!
- Type III
    - Each predictor is evaluated taking account of all other predictors
    - The order of predictors doesn’t matter
- Type II and IV
    - These exist too but let's not confuse things …

Type I sums of squares

```r
lm(happiness ~ puppy_love + dose,
   data = pupluv_tib) %>%
  anova()
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
puppy_love	1	6.73	6.73	2.22	0.15
dose	2	25.19	12.59	4.14	0.03
Residuals	26	79.05	3.04	NA	NA

```r
lm(happiness ~ dose + puppy_love,
   data = pupluv_tib) %>%
  anova()
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
dose	2	16.84	8.42	2.77	0.08
puppy_love	1	15.08	15.08	4.96	0.03
Residuals	26	79.05	3.04	NA	NA

WARNING! Type I sums of squares: variable order matters!

Type III sums of squares

```
lm(happiness ~ puppy_love + dose,
   data = pupluv_tib) %>%
car::Anova(., type = 3)
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	12.94	1	4.26	0.05
puppy_love	15.08	1	4.96	0.03
dose	25.19	2	4.14	0.03
Residuals	79.05	26	NA	NA

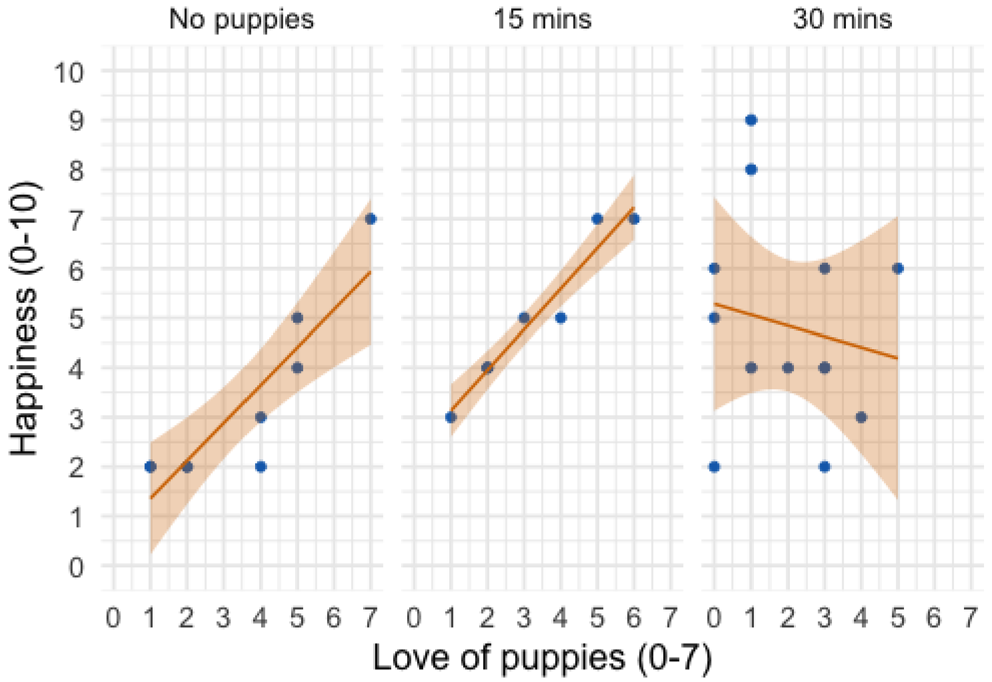
```
lm(happiness ~ dose + puppy_love,
   data = pupluv_tib) %>%
car::Anova(., type = 3)
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	12.94	1	4.26	0.05
dose	25.19	2	4.14	0.03
puppy_love	15.08	1	4.96	0.03
Residuals	79.05	26	NA	NA

TIP! If you want *F*-statistics and have several predictors, use Type III sums of squares

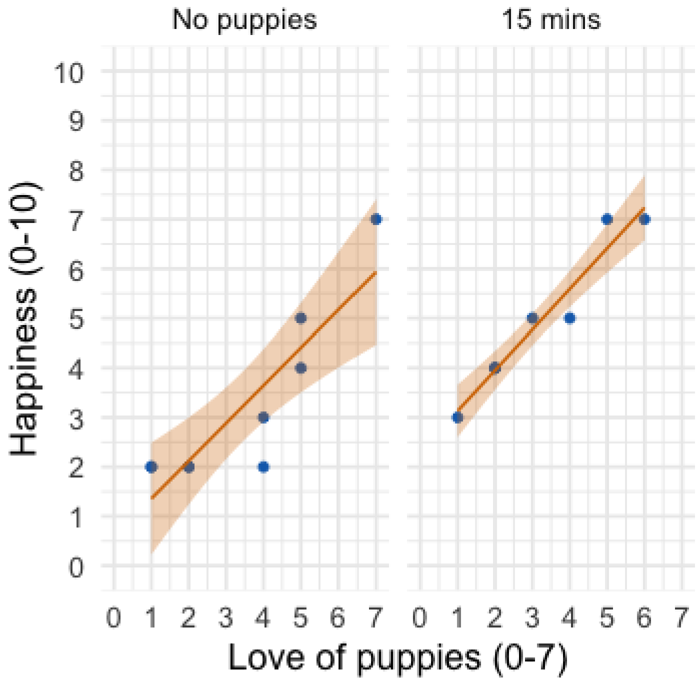
Bias in the *F*-statistic

- For the significance of *F*-statistics to be accurate we assume that the relationship between the covariate and outcome is similar across groups.
- Known as homogeneity of regression slopes
- When the assumption is met the resulting *F*-statistic can be assumed to follow the *F*-distribution and the corresponding *p*-value is accurate.
- When the assumption is not met the *F*-statistic might not follow the *F*-distribution and the corresponding *p*-value is inaccurate
- Only relevant for the *F*-statistic



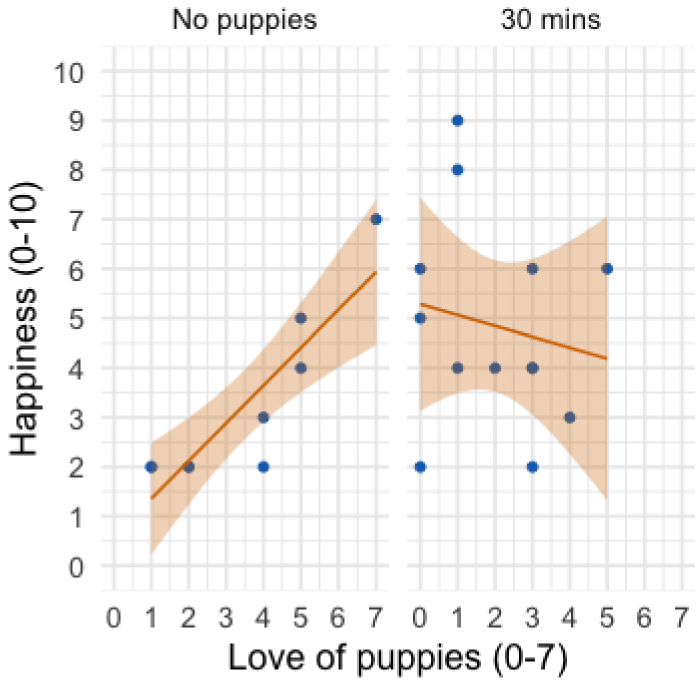
Homogeneity of regression slopes

- For the significance of *F*-statistics to be accurate we assume that the relationship between the covariate and outcome is similar across groups.
- Known as homogeneity of regression slopes
- When the assumption is met the resulting *F*-statistic can be assumed to follow the *F*-distribution and the corresponding *p*-value is accurate.
- When the assumption is not met the *F*-statistic might not follow the *F*-distribution and the corresponding *p*-value is inaccurate
- Only relevant for the *F*-statistic



Heterogeneity of regression slopes

- For the significance of *F*-statistics to be accurate we assume that the relationship between the covariate and outcome is similar across groups.
- Known as homogeneity of regression slopes
- When the assumption is met the resulting *F*-statistic can be assumed to follow the *F*-distribution and the corresponding *p*-value is accurate.
- When the assumption is not met the *F*-statistic might not follow the *F*-distribution and the corresponding *p*-value is inaccurate
- Only relevant for the *F*-statistic



Fitting the model

Overall fit of each predictor

```
puplub_lm <- lm(happiness ~ dose + puppy_love, data = pupluv_tib)
car::Anova(pupluv_lm, type = 3)
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	12.943	1	4.257	0.049
puppy_love	15.076	1	4.959	0.035
dose	25.185	2	4.142	0.027
Residuals	79.047	26	NA	NA

✎ The dose of puppy therapy had a significant effect on happiness, $F(1, 26) = 4.96$, $p = 0.035$.

✎ Love of puppies significantly predicted happiness, $F(2, 26) = 4.14$, $p = 0.027$.

Breaking down the overall effects (parameter estimates)

```
broom::tidy(pupluv_lm, conf.int = TRUE)
```

term	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	1.789	0.867	2.063	0.049	0.007	3.572
puppy_love	0.416	0.187	2.227	0.035	0.032	0.800
dose15 mins	1.786	0.849	2.102	0.045	0.040	3.532
dose30 mins	2.225	0.803	2.771	0.010	0.575	3.875

Love of puppies significantly predicted happiness, b = 0.42 [0.03, 0.80], t = 2.23, p = 0.035. For every unit increase in puppy love, predicted happiness increased by 0.42 units.

The dose of puppy therapy also significantly predicted happiness. Compared to no puppy controls, happiness was significantly higher after both 15 minutes, b = 1.79 [0.04, 3.53], t = 2.10, p = 0.045, and 30 minutes, b = 2.22 [0.57, 3.88], t = 2.77, p = 0.010, of therapy.

Adjusted means

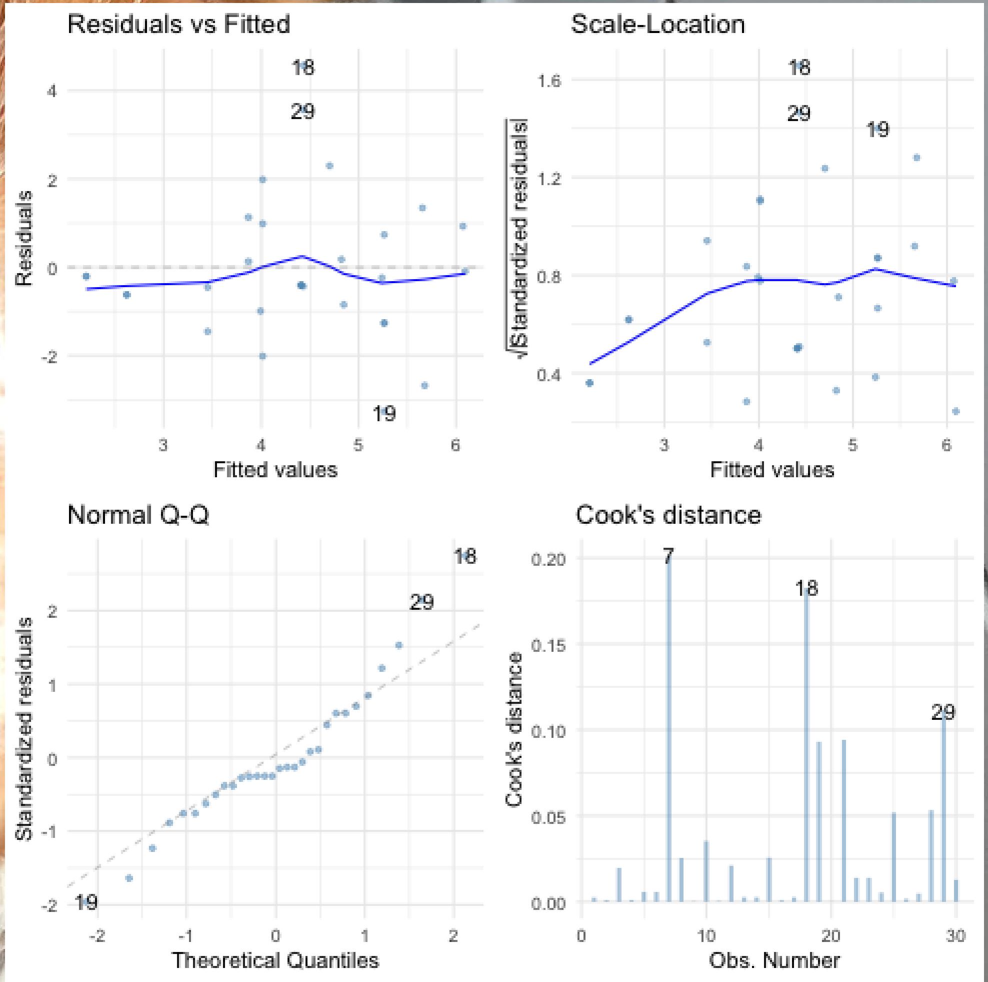
```r
modelbased::estimate_means(pupluv_lm, fixed = "puppy_love")
```

dose	puppy_love	Mean	SE	CI_low	CI_high
No puppies	2.73	2.93	0.60	1.70	4.15
15 mins	2.73	4.71	0.62	3.44	5.99
30 mins	2.73	5.15	0.50	4.12	6.18

✎ At average levels of love of puppies, the mean happiness in the no puppy control group was *M* = 2.93 [1.70, 4.15] compared to *M* = 4.71 [3.44, 5.99] in the 15-minute group and *M* = 5.15 [4.12, 6.18] in the 30-minute group.

Testing assumptions

```r
library(ggfortify)
ggplot2::autoplot(pupluv_lm,
                  which = c(1, 3, 2, 4),
                  colour = "#5c97bf",
                  alpha = 0.5,
                  size = 1) +
  theme_minimal()
```



Homogeneity of regression slopes

```
hors_lm <- lm(happiness ~ puppy_love*dose, data = pupluv_tib)
car::Anova(hors_lm, type = 3)
```

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	0.771	1	0.316	0.579
puppy_love	19.922	1	8.157	0.009
dose	36.558	2	7.484	0.003
puppy_love:dose	20.427	2	4.181	0.028
Residuals	58.621	24	NA	NA

Homogeneity of regression slopes cannot be assumed because:

1. The dose*puppy_love interaction effect is significant
2. The plots (see earlier) show that the relationship between puppy love and happiness is different in the 30 minute group to the other two groups.

Robust parameter estimates

```
pupluv_rob <- robust::lmRob(happiness ~ puppy_love + dose, data = pupluv_tib)
summary(pupluv_rob)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	1.041	2.103	0.495	0.625
puppy_love	0.633	0.556	1.139	0.265
dose15 mins	1.855	2.642	0.702	0.489
dose30 mins	1.400	1.750	0.800	0.431

✎ Robust estimates showed that love of puppies did not have a significant effect on happiness, b = 0.63, t = 1.14, p = 0.265. For every unit increase in puppy love (on the 0-7 scale), predicted happiness (on the 0-10 scale) increased by 0.63 units.

✎ The dose of puppy therapy also did not significantly predicted happiness. Robust estimates showed that compared to no puppy controls, happiness was not significantly higher after 15 minutes, b = 1.86, t = 0.70, p = 0.489, or 30 minutes, b = 1.40, t = 0.80, p = 0.431, of therapy.

Heteroscedasticity consistent standard errors

```
parameters::model_parameters(pupluv_lm, robust = TRUE, vcov.type = "HC4")
```

Parameter	Coefficient	SE	CI_low	CI_high	t	df_error	p
(Intercept)	1.789	0.581	0.596	2.983	3.081	26	0.005
puppy_love	0.416	0.199	0.008	0.824	2.095	26	0.046
dose15 mins	1.786	0.517	0.724	2.848	3.456	26	0.002
dose30 mins	2.225	0.734	0.716	3.734	3.031	26	0.005

✎ Love of puppies significantly predicted happiness, b = 0.42 [0.01, 0.82], t = 2.10, p = 0.046. For every unit increase in puppy love, predicted happiness increased by 0.42 units.

✎ The dose of puppy therapy also significantly predicted happiness. Compared to no puppy controls, happiness was significantly higher after both 15 minutes, b = 1.79 [0.72, 2.85], t = 3.46, p = 0.002, and 30 minutes, b = 1.79 [0.72, 2.85], t = 3.46, p = 0.002, of therapy.

Summary

- When we include both a categorical and continuous predictor, the categorical predictor compares means adjusted for the effect of the continuous predictor.
    - The effect of the categorical variable at average levels of the continuous predictor
- Test the overall effect of categorical predictors using the *F*-statistic
    - Use Type III sums of squares (other things being equal)
    - Test for homogeneity of regression slopes
- Break down the effects of categorical predictors using parameter estimates and their associated tests
    - Interpret in the same way as in previous lectures
- Test the usual assumptions for the linear model
- Apply a robust test if necessary