0:00 / 3:48

Model fit and multiple predictors

Professor Andy Field

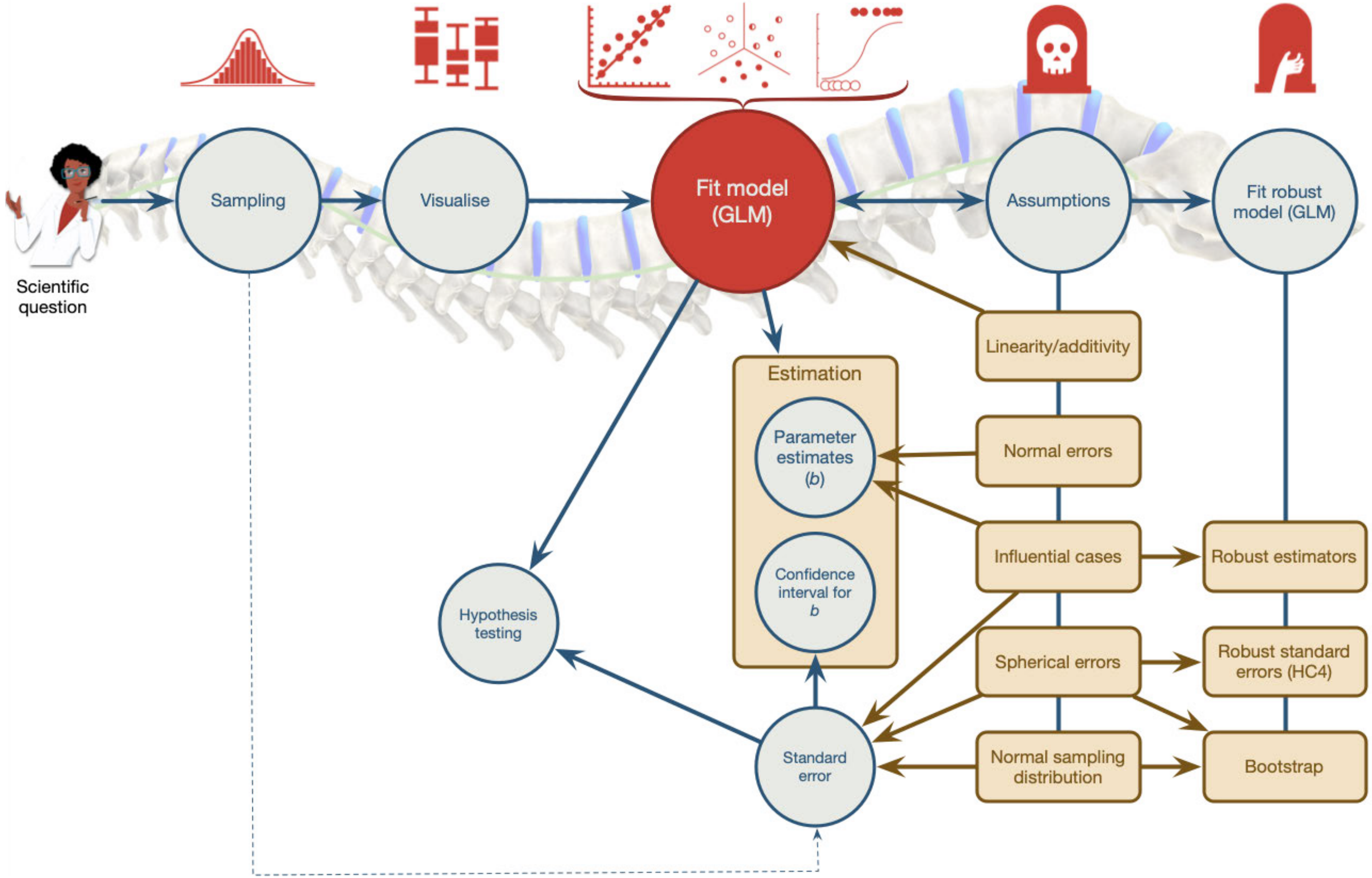
@profandyfield

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

- Understand how we establish the fit of a general linear model to the
    - Sums of squares
    - Mean squares
    - The *F*-statistic
    - R^2
- Understanding how to incorporate multiple predictors in the general linear model
    - The mathematical model
    - Visualizing the model
    - Methods for entering predictors to the model
    - Interpreting parameter estimates

Recap: The General Linear Model (GLM)

$$\text{outcome}_i = (\text{model}_i) + \text{error}_i$$
$$\widehat{\text{outcome}}_i = \hat{b}_0 + \hat{b}_1\text{predictor}_i + \text{error}_i$$

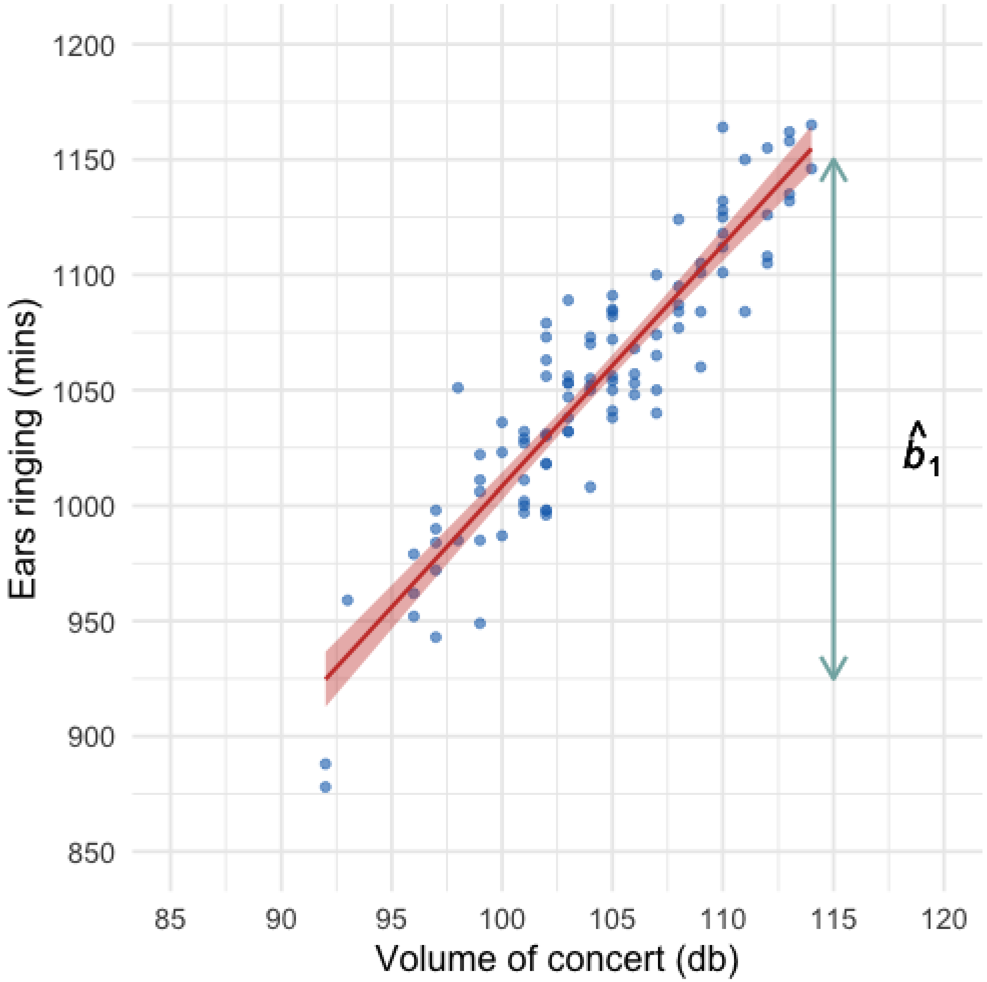
$\hat{b}_1$

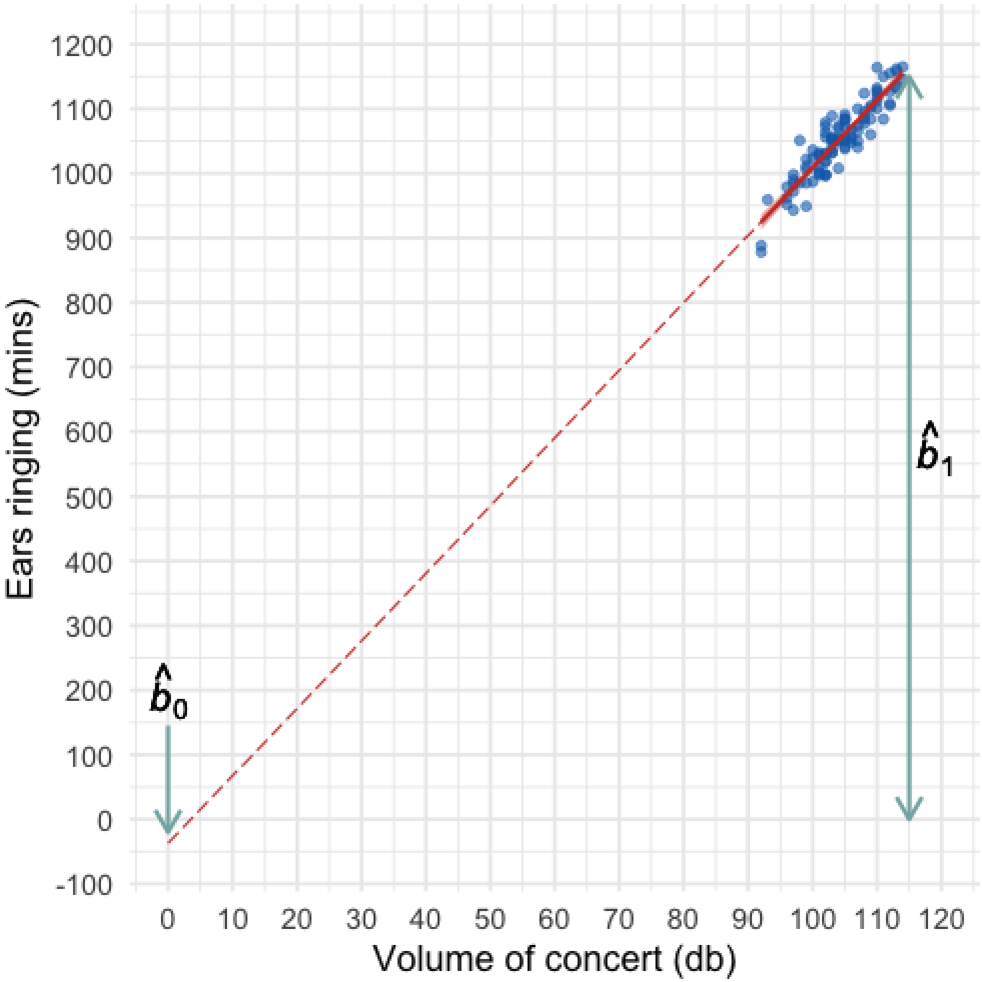
- Estimate of parameter for a predictor
  - Direction/strength of relationship/effect
  - Difference in means

$\hat{b}_0$

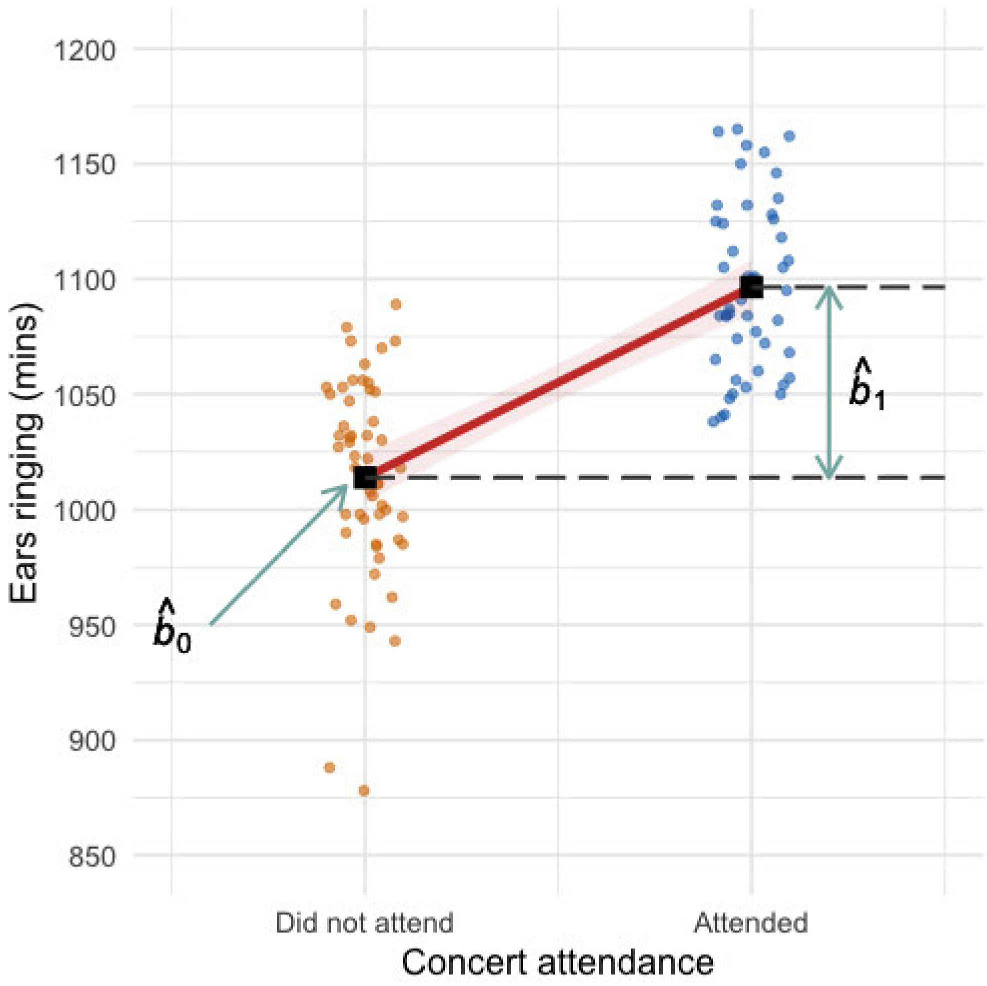
- Estimate of the value of the outcome when predictor(s) = 0 (intercept)

$$\hat{\text{ringing}}_i = -37.12 + 10.45\text{volume}_i + e_i$$





$$\hat{\text{ringing}}_i = \hat{b}_0 + \hat{b}_1 \text{attendance}_i + e_i$$



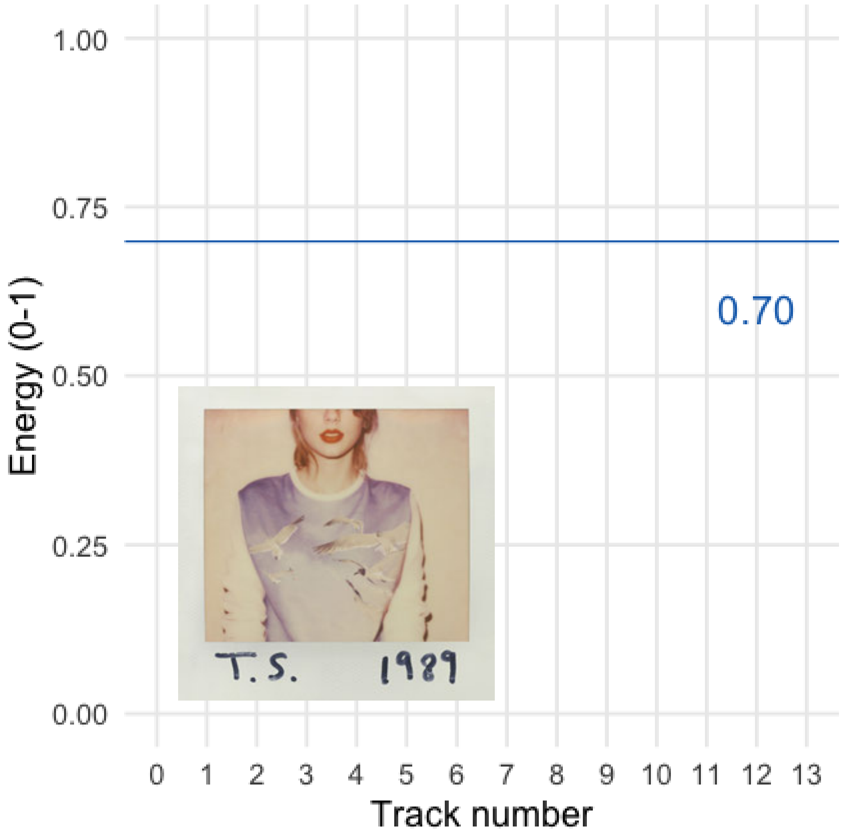
How do we tell if a model is a good fit?

- Let's look at a simple model: the mean
- How do we tell if it's a good fit?
- With some help from Taylor Swift
  - The album 1989
- produce measures of song content
  - Energy
  - Valence
  - Danceability

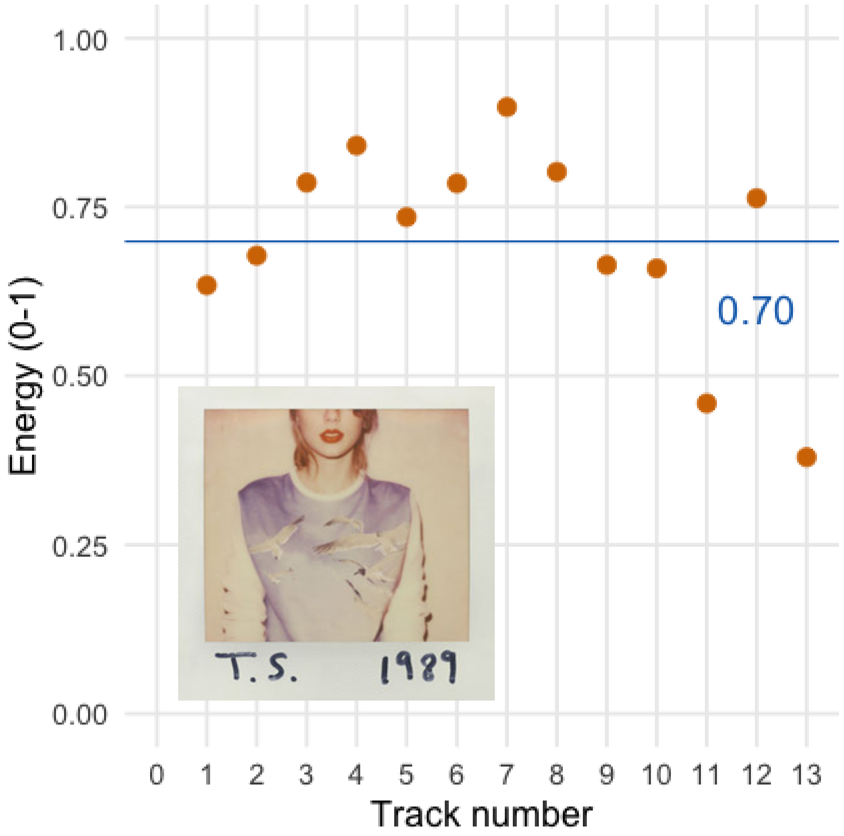
> "Energy is a measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity." (Spotify, API)

- The `spotifyr` package scrapes this data!

Is the average energy score a good fit?

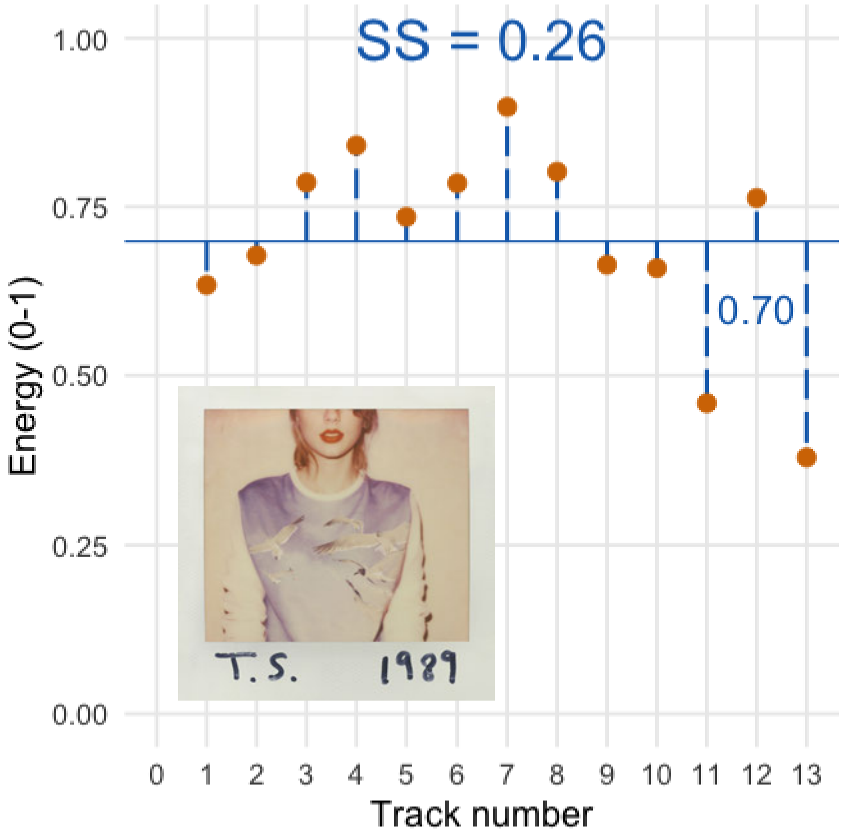


Is the average energy score a good fit?

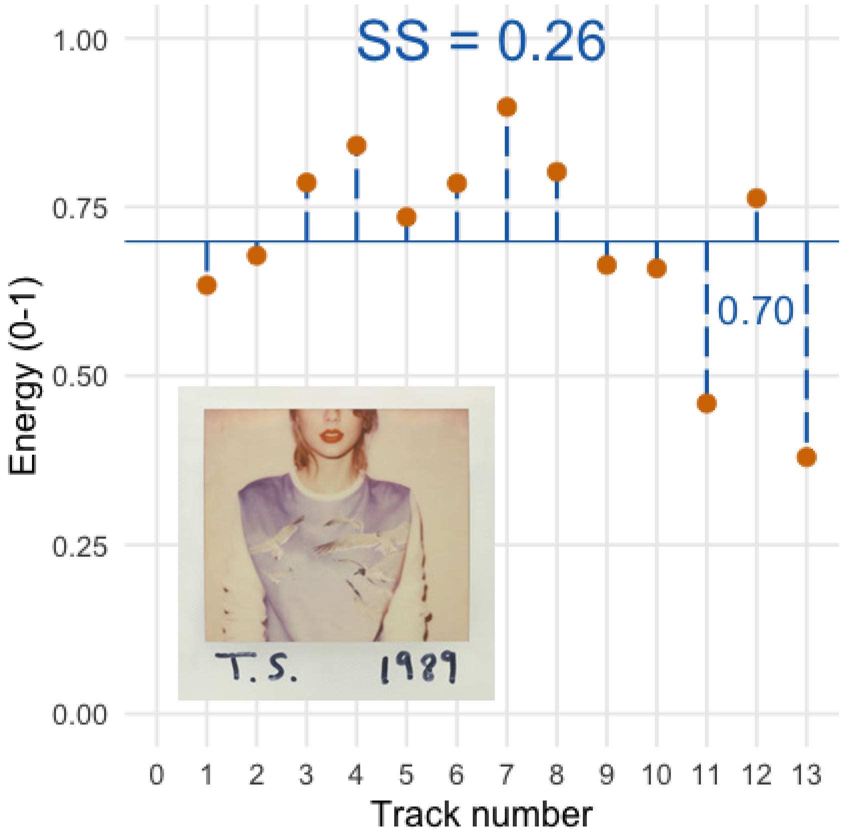


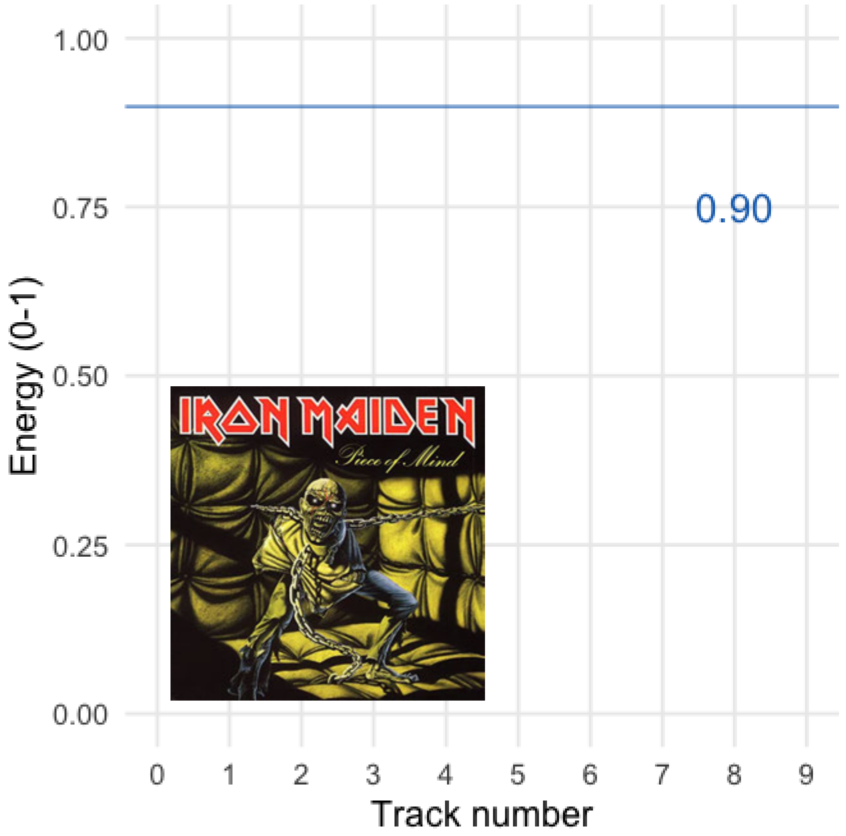
0:00 / 0:15

Is the average energy score a good fit?

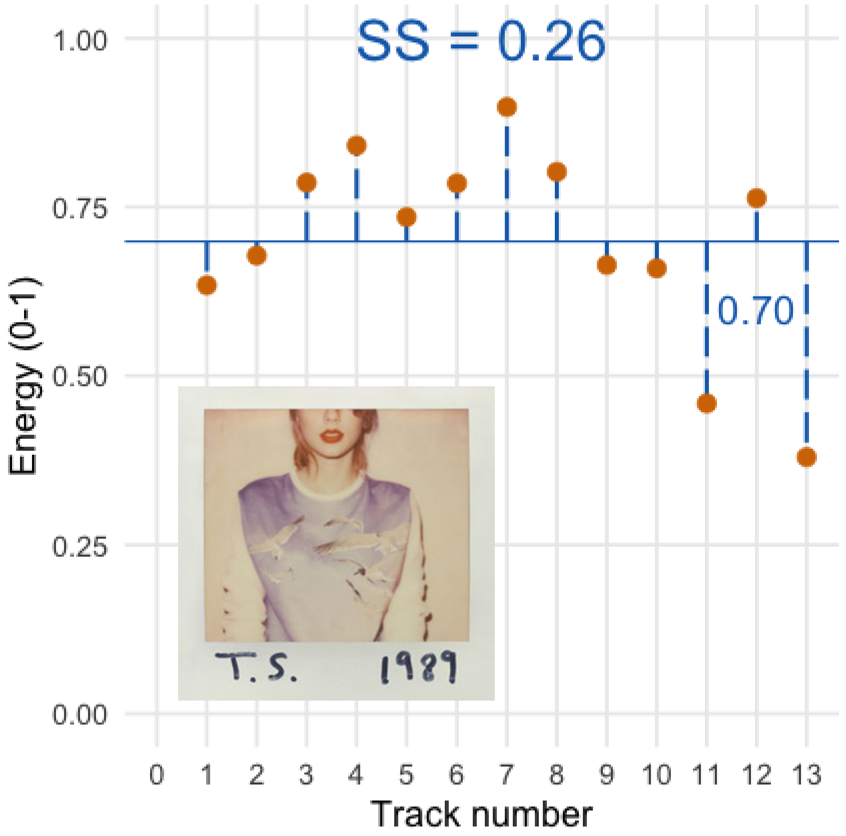


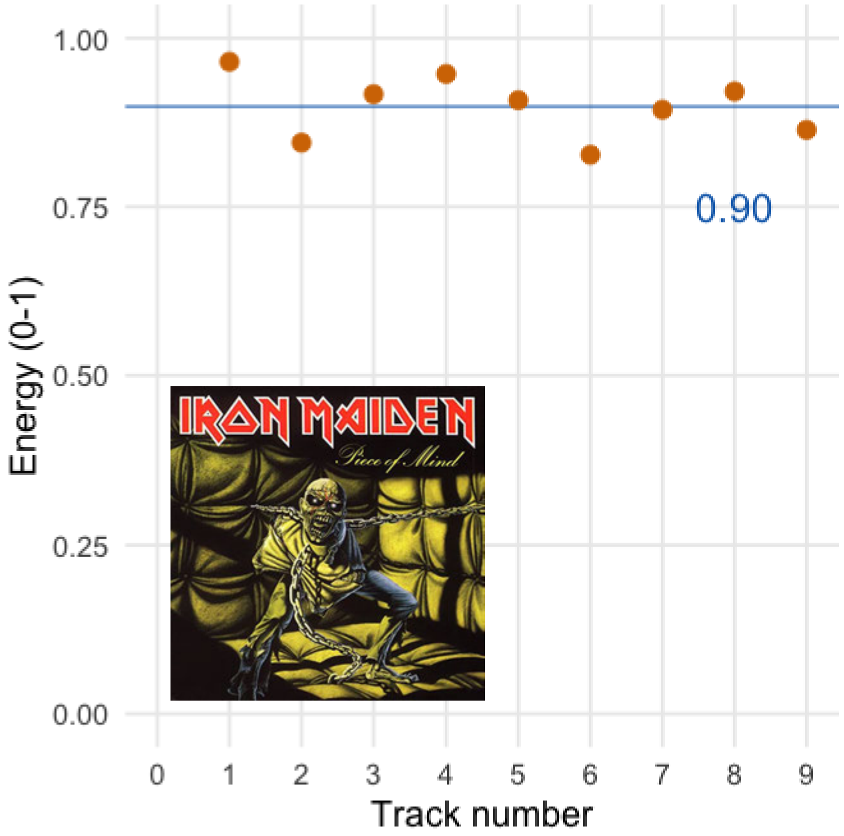
Is the average energy score a good fit?



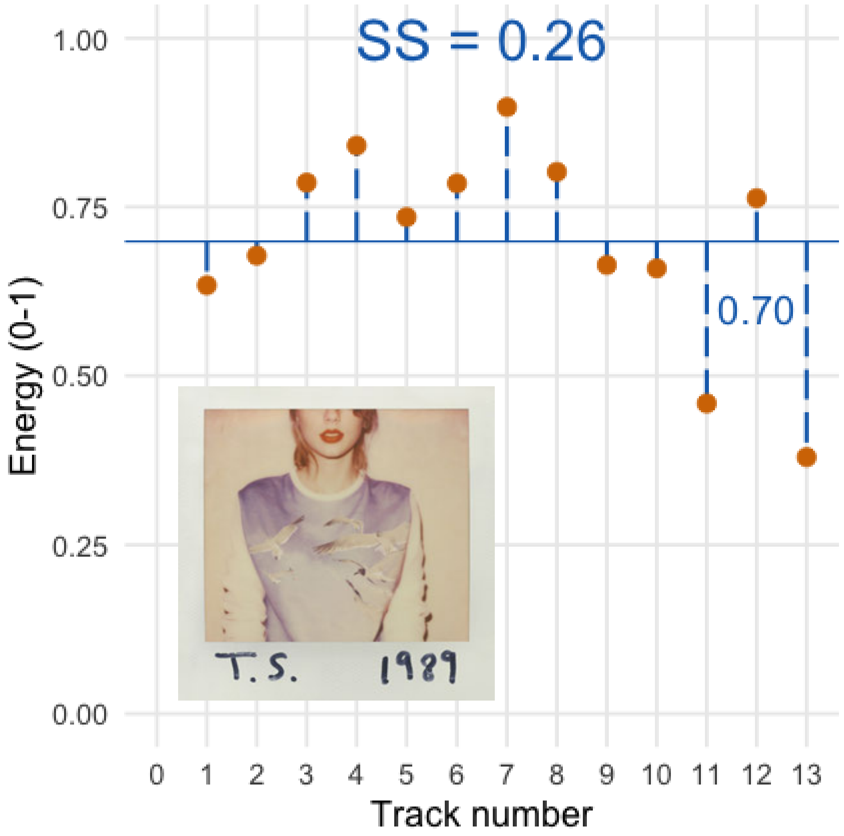


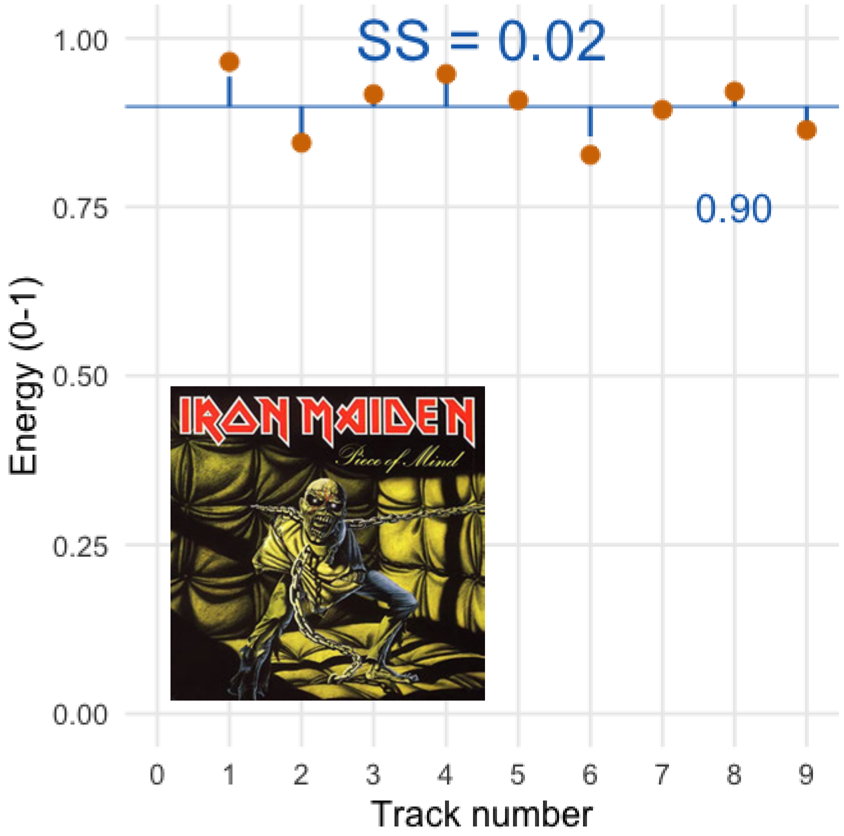
Is the average energy score a good fit?





Is the average energy score a good fit?



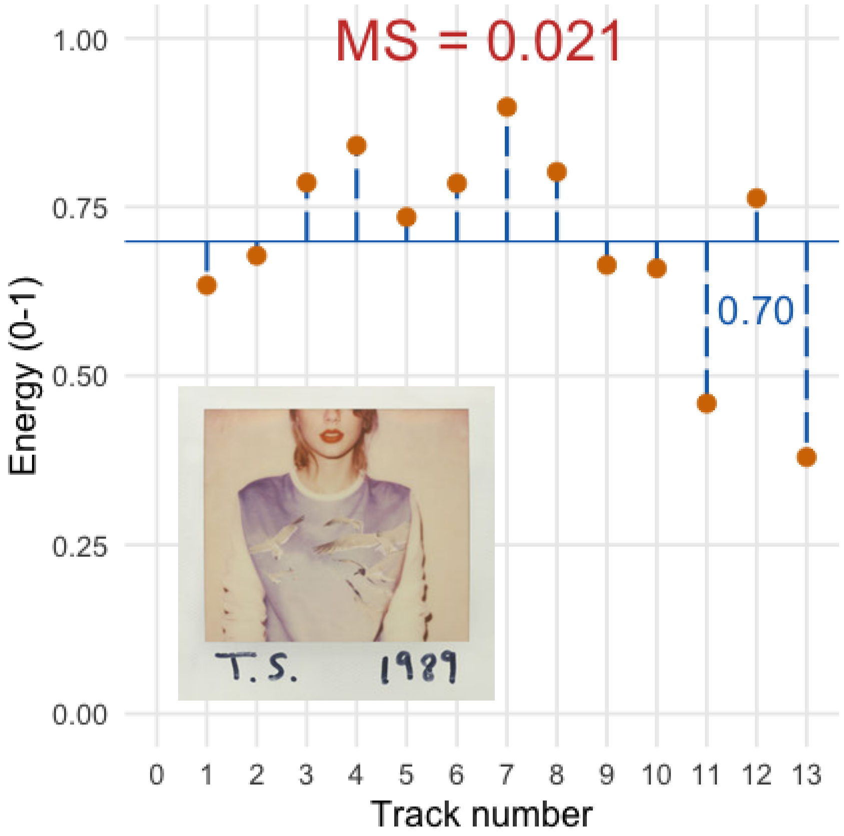


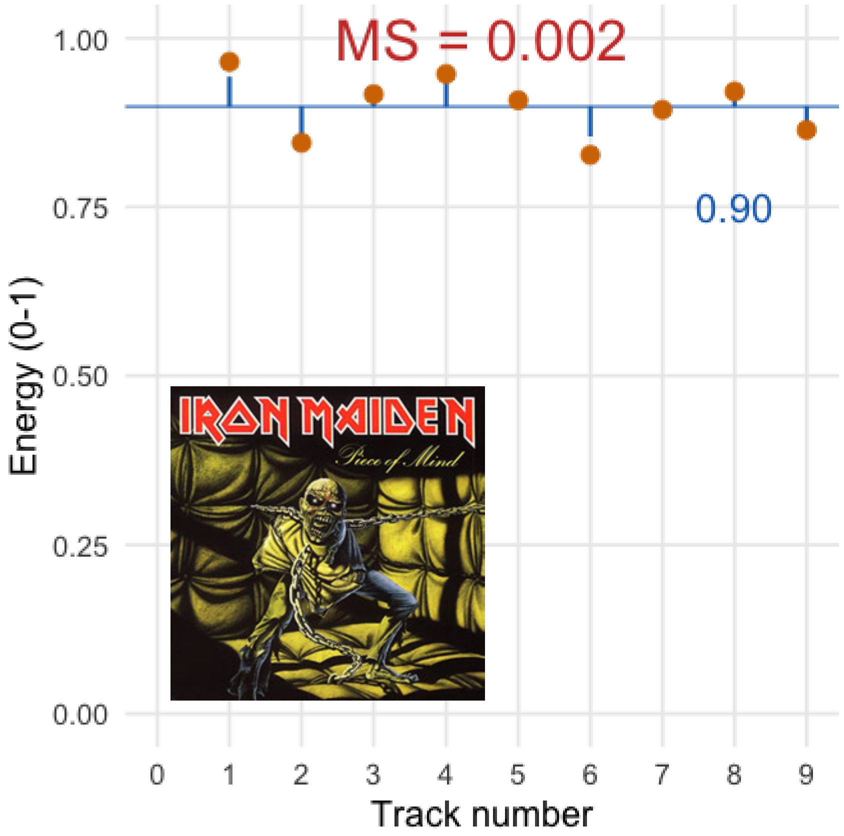
Comparing sums of squares

- Sums of squares represent total error
- Because sums of squares are totals we can compare them only when they are based on the same number of scores.
- Alternatively, we factor in the number of scores
- We can get the average error by divide by a function of the number of scores
  - The degrees of freedom (the number of independent pieces of information)
  - The number of scores minus the number of parameters
  - $\text{df} = N - p$

$$
\begin{aligned}
\text{MS} &= \frac{\text{SS}}{N - p} \\
&= \frac{\text{SS}}{N - 1}
\end{aligned}
$$

Is the average energy score a good fit?





Illusory Truth Effect (ITE)

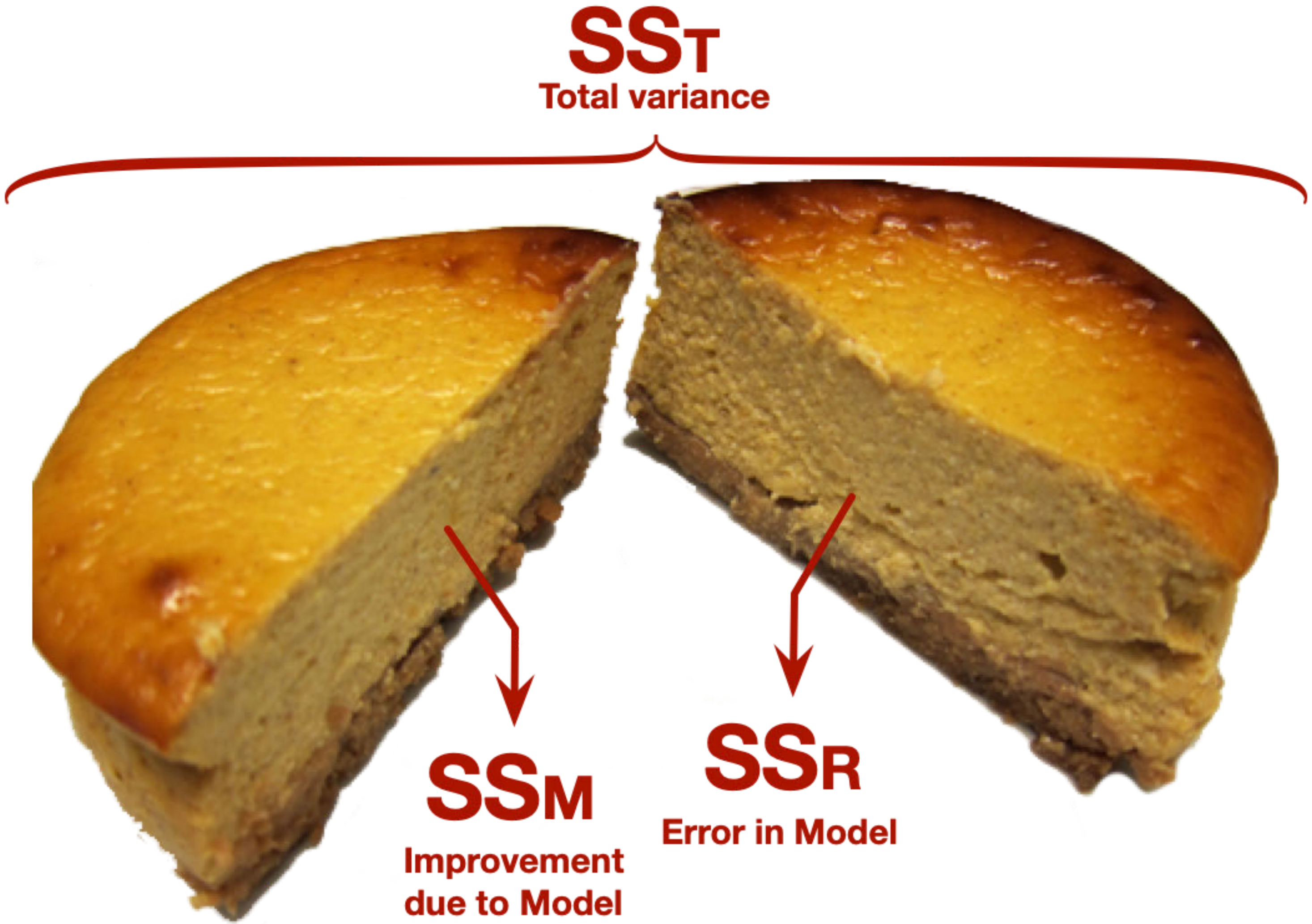
- Repetition increases perceived truthfulness (Hasher et al., 1977)
- This is equally true for plausible and implausible statements (Fazio et al., 2019)
- Worryingly, the effect is true for false political statements regardless of political ideology
    - Of 105 statements made by Donald Trump between 02/11/2016 and 9/10/2019 77 (73%) were only half true or worse.
    - In experiments, people exposed repeatedly to Trump statements rated them as more truthful than those who were not on a 6-point scale [1]
- Imagine a model that predicts ratings of truth of fake statements from number of exposures
    - Predictor: number of exposures
    - outcome: ratings of truth (0 = definitely false, 5 = definitely true)

Murray et al. (2020). https://doi.org/10.31234/osf.io/9evzc

Testing the fit of the general linear model

To see whether the model is a reasonable 'fit' of the observed data we use the sum of squared errors (SS):

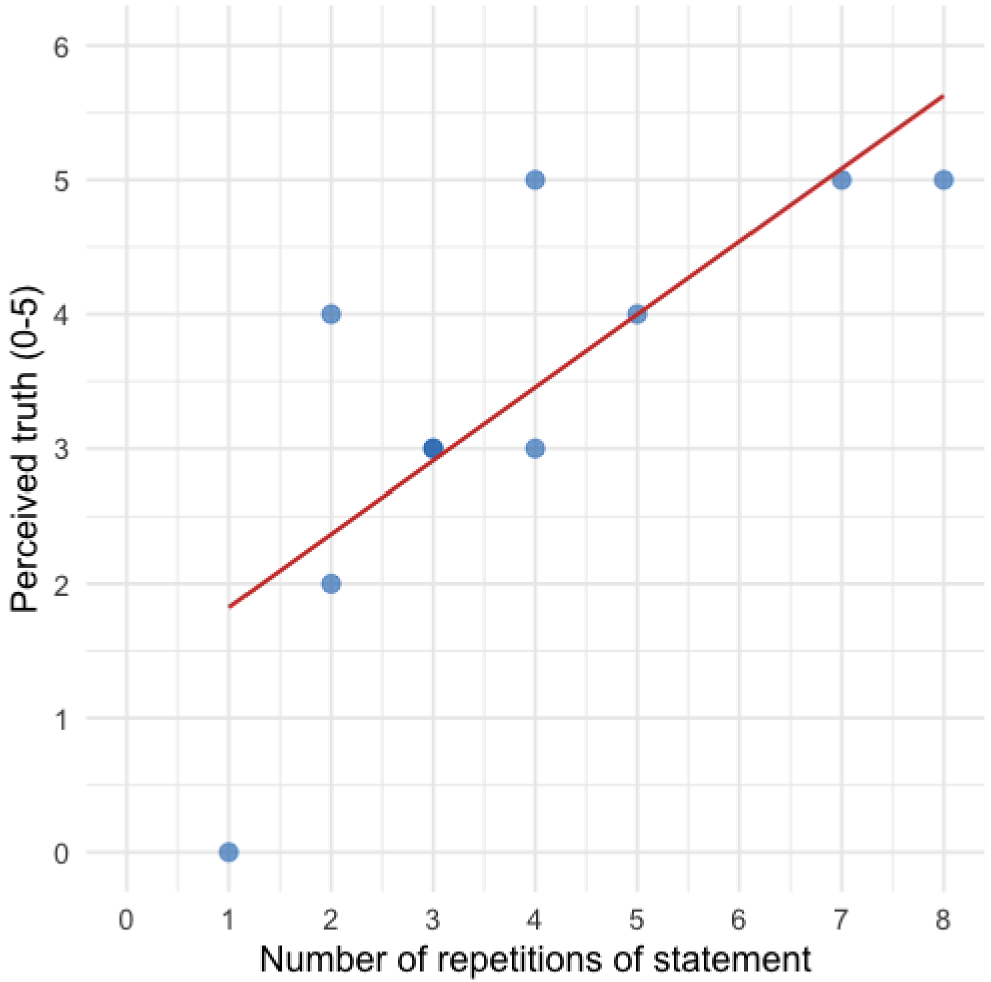
- SS_T
  - Total variability (variability between scores and the mean)
- SS_R
  - Total residual/error variability (variability between the model and the observed data)
  - How badly the model fits (in total)
- SS_M
  - Total model variability (difference in variability between the model and the grand mean)
  - How much better the model is at predicting *Y* than the mean
  - How well the model fits (in total)



SS_T
Total variance
SS_M
Improvement due to Model
SS_R
Error in Model

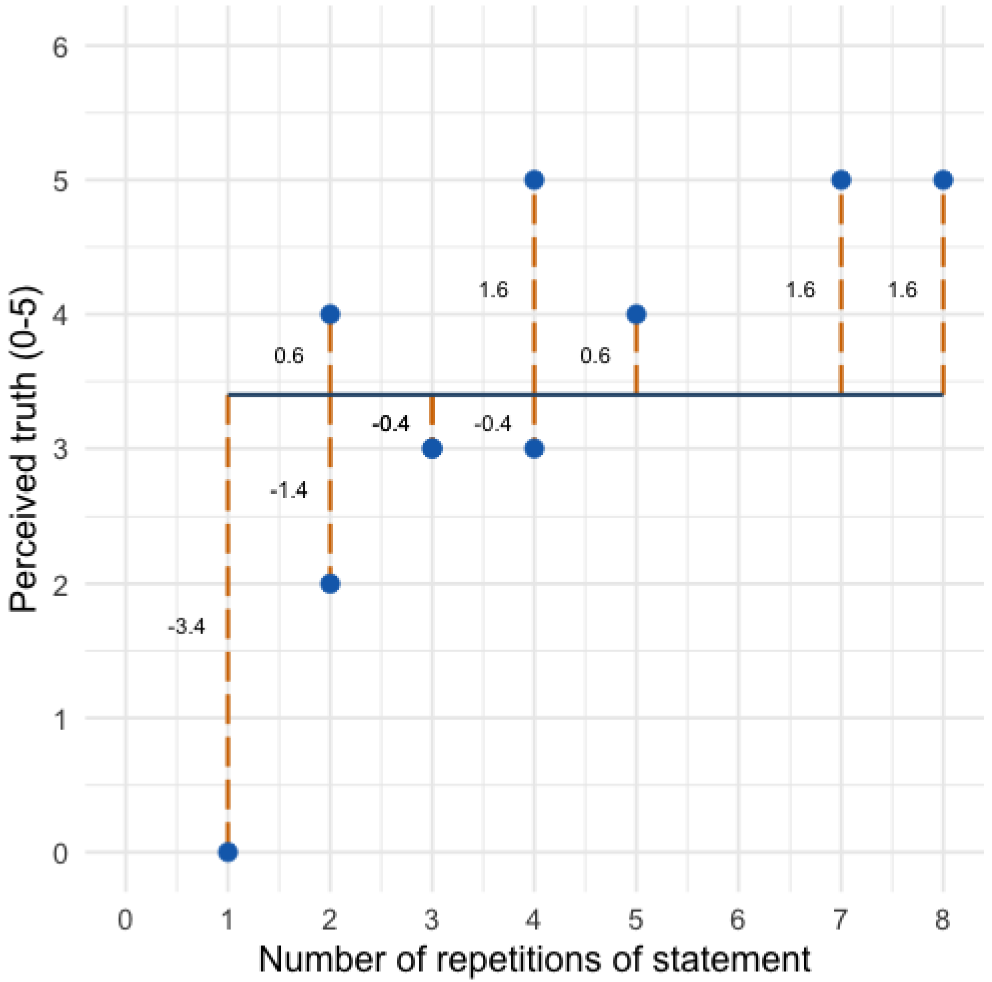
$$\hat{\text{perceived truth}}_i = \hat{b}_0 + \hat{b}_1\text{repetition}_i + e_i$$

$$\hat{\text{perceived truth}}_i = 1.28 + 0.54\ \text{repetition}_i + e_i$$



Total sum of squared error, SS_T

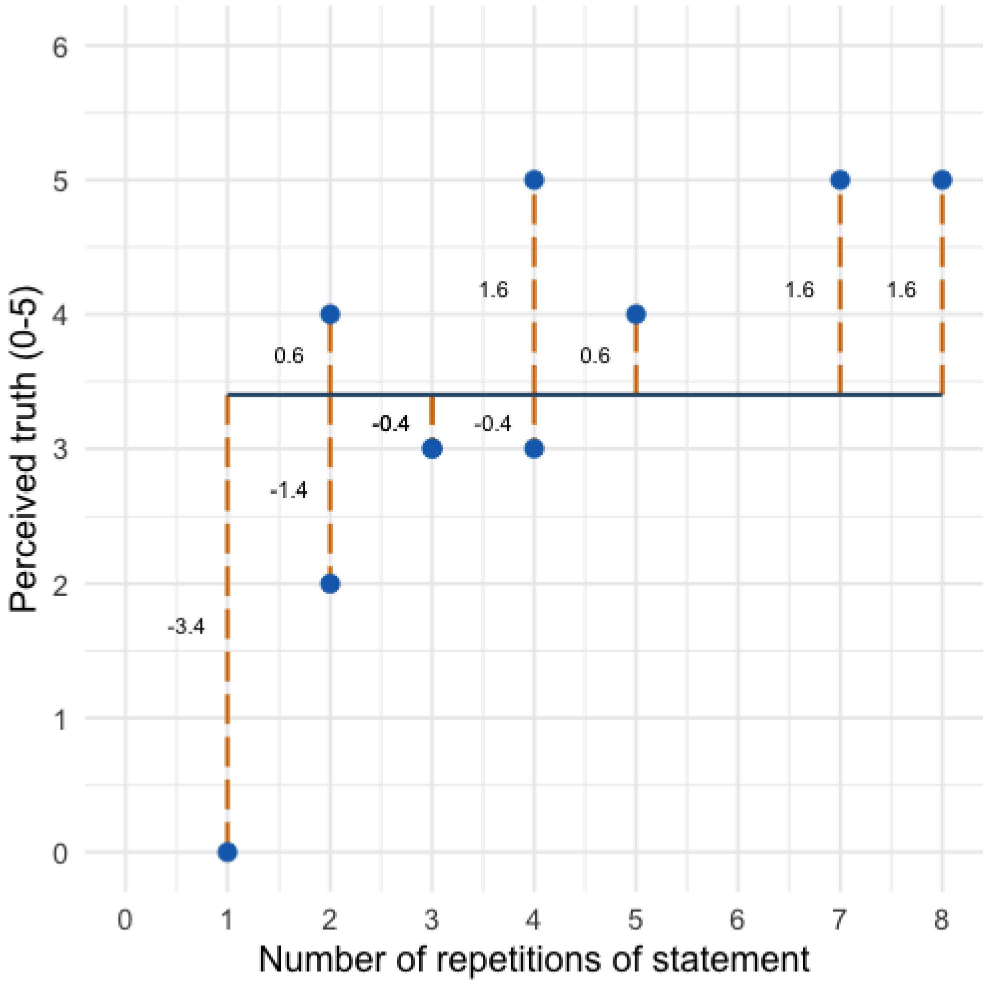
	Reps	Truth	Y_{Pred}	Error	$Error^2$
	1	0	3.4	-3.4	11.56
	2	2	3.4	-1.4	1.96
	2	4	3.4	0.6	0.36
	3	3	3.4	-0.4	0.16
	3	3	3.4	-0.4	0.16
	4	3	3.4	-0.4	0.16
	4	5	3.4	1.6	2.56
	5	4	3.4	0.6	0.36
	7	5	3.4	1.6	2.56
	8	5	3.4	1.6	2.56
SS_T	—	—	—	—	22.40



Total sum of squared errors, SS_T

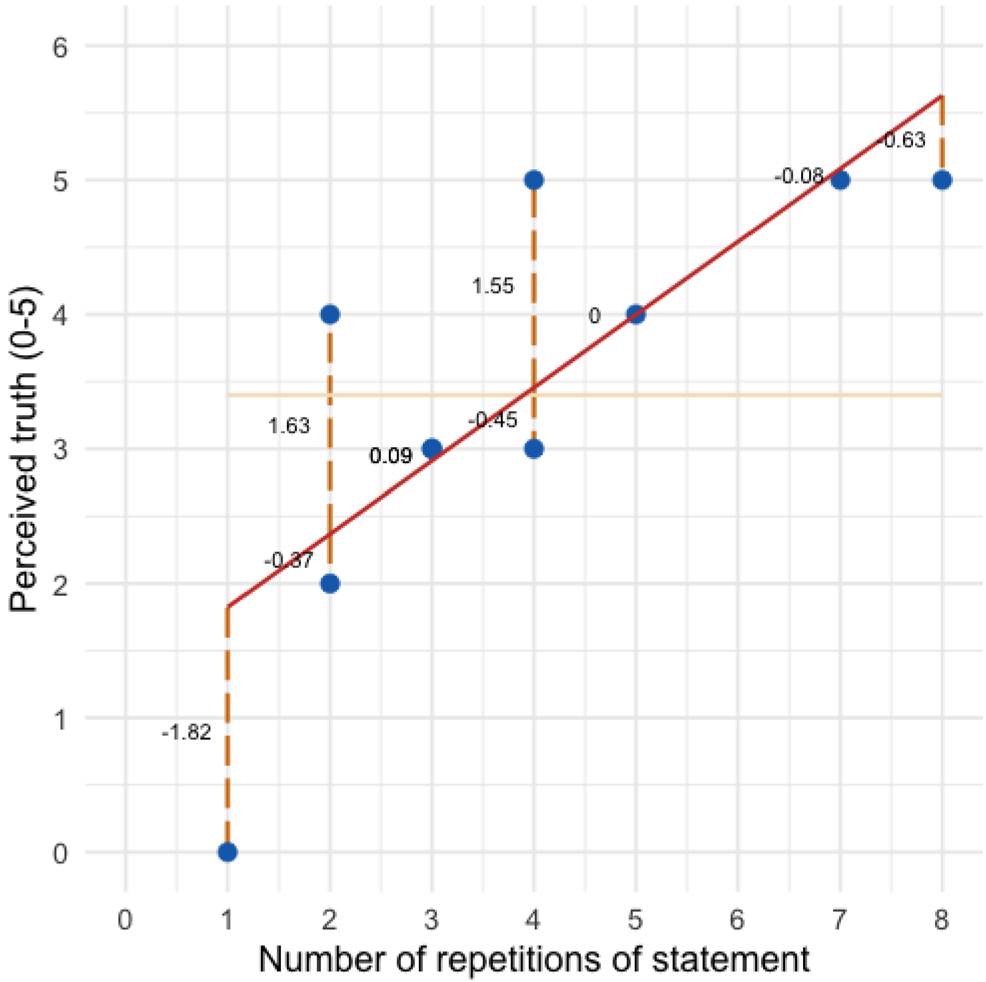
- Each SS has associated degrees of freedom (df)
- The *df* is the amount of *independent information* available to compute SS
- To begin with we have *N* pieces of independent information
- For every parameter (*p*) estimated we lose 1 piece of independent information
- To get SS_T we estimate 1 parameter (the overall mean):

$$\begin{aligned} \mathrm{df_T} &= N - p \\ &= 10 - 1 \\ &= 9 \end{aligned}$$



Residual sum of squared errors, SS_R

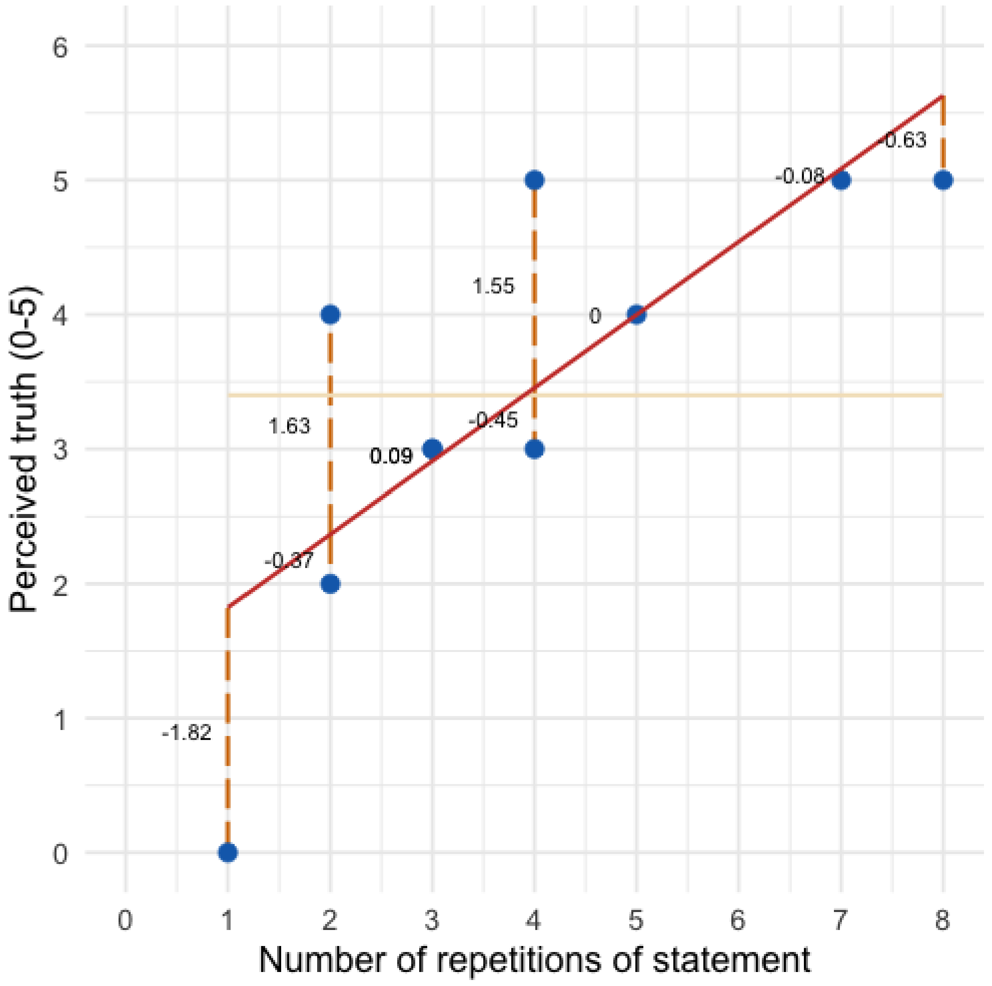
	Reps	Truth	Y_{Pred}	Error	$Error^2$
	1	0	1.82	-1.82	3.31
	2	2	2.37	-0.37	0.14
	2	4	2.37	1.63	2.66
	3	3	2.91	0.09	0.01
	3	3	2.91	0.09	0.01
	4	3	3.45	-0.45	0.20
	4	5	3.45	1.55	2.40
	5	4	4.00	0.00	0.00
	7	5	5.08	-0.08	0.01
	8	5	5.63	-0.63	0.40
SS_R	—	—	—	—	9.14



Residual sum of squared errors, SS_R

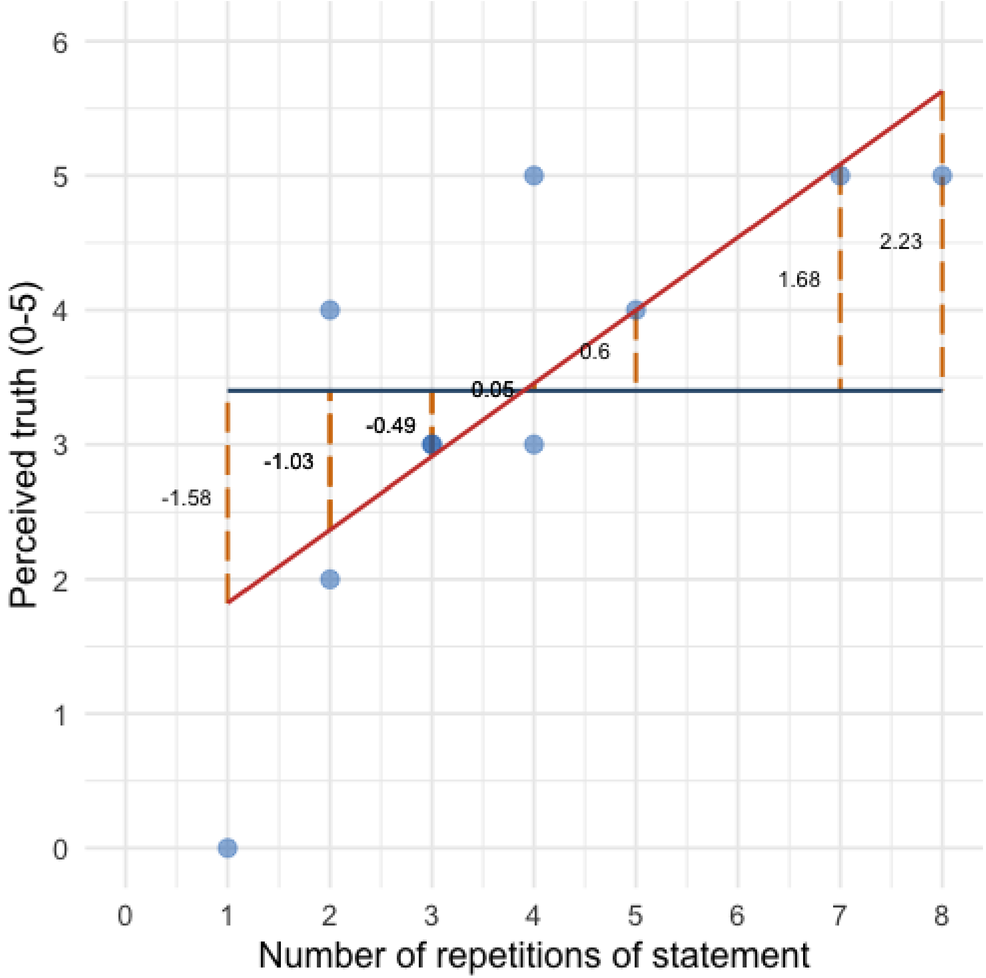
- To begin with we have *N* pieces of independent information
- To get SS_R we estimate two parameters (b_0 and b_1):

$$\begin{aligned} df_R &= N - p \\ &= 10 - 2 \\ &= 8 \end{aligned}$$



Model sum of squared errors, SS_M

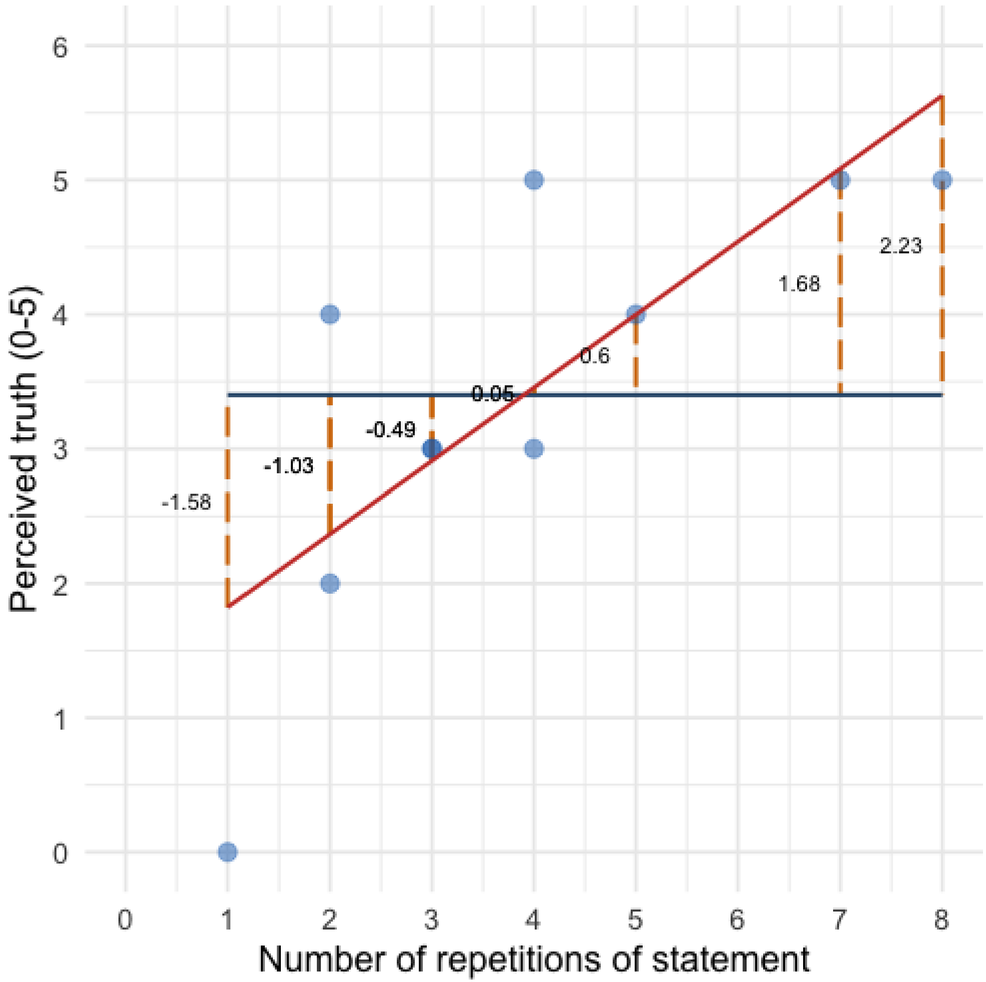
	Reps	Truth	Y_{Pred}	Mean	Error	$Error^2$
	1	0	1.82	3.4	-1.58	2.50
	2	2	2.37	3.4	-1.03	1.06
	2	4	2.37	3.4	-1.03	1.06
	3	3	2.91	3.4	-0.49	0.24
	3	3	2.91	3.4	-0.49	0.24
	4	3	3.45	3.4	0.05	0.00
	4	5	3.45	3.4	0.05	0.00
	5	4	4.00	3.4	0.60	0.36
	7	5	5.08	3.4	1.68	2.82
	8	5	5.63	3.4	2.23	4.97
SS_M	—	—	—	—	—	13.25



Model sum of squared errors, SS_M

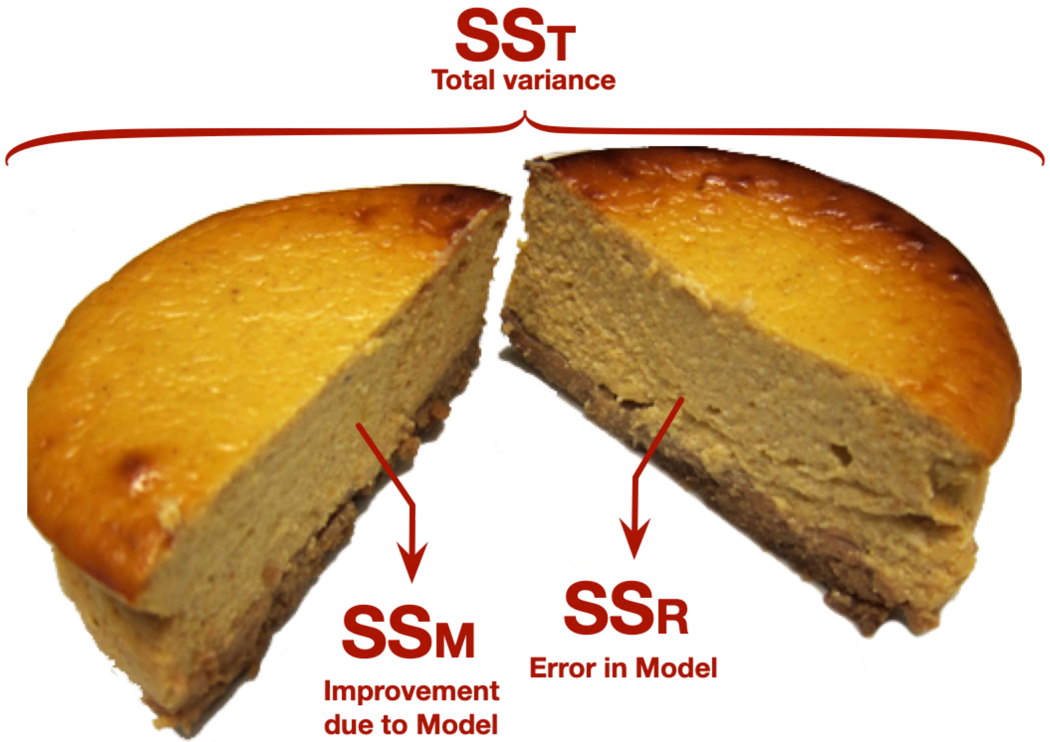
- The model is a rotation of the null model (the grand mean)
- Therefore, the null model and the estimated model are distinguished by 1 piece of independent information: the slope, b_1
  - (Note, the intercept, b_0, co-depends on the slope - it is not an independent piece of information)

$$
\begin{aligned}
df_M &= df_T - df_R \\
&= 9 - 8 \\
&= 1
\end{aligned}
$$



$$SS_T = SS_M + SS_R$$
$$22.40 = 13.25 + 9.14$$

(Within rounding error)



Mean squared error (MS)

- A sum/total of squared errors depends on the amount of information used to compute it
    - (If you add more squared errors, the sum increases)
- We can't compare sums of squared errors based on different amounts of information
- We can compute the average or mean squared error by dividing a SS by the amount of information used to compute it
- The df quantifies the amount of information used to compute a sum of squared errors

$$\text{MS} = \frac{\text{SS}}{\text{df}}$$

Mean squared error (MS)

- MS_R
  - Average residual/error variability (variability between the model and the observed data)
  - How badly the model fits (on average)

$$\mathrm{MS_R} = \frac{\mathrm{SS}}{\mathrm{df}} = \frac{9.14}{8} = 1.14$$

- MS_M
  - Average model variability (difference in variability between the model and the grand mean)
  - How much better the model is at predicting *Y* than the mean
  - How well the model fits (on average)

$$\mathrm{MS_M} = \frac{\mathrm{SS}}{\mathrm{df}} = \frac{13.25}{1} = 13.25$$

Testing the model fit: the *F*-statistic

- If the model results in better prediction than using the mean, then MS_M should be greater than MS_R
- The *F*-statistic is the ratio of MS_M to MS_R
  - It's the good-to-shit ratio

$$F = \frac{\text{MS}_\text{M}}{\text{MS}_\text{R}} = \frac{13.25}{1.14} = 11.62$$

```
ite_lm <- lm(belief ~ repetition, data = ite_tib)
anova(ite_lm)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
repetition	1	13.26	13.26	11.61	0.01
Residuals	8	9.14	1.14	NA	NA

Testing the model fit: R^2

- R^2
  - The proportion of variance accounted for by the model
  - The Pearson correlation coefficient between observed and predicted scores squared

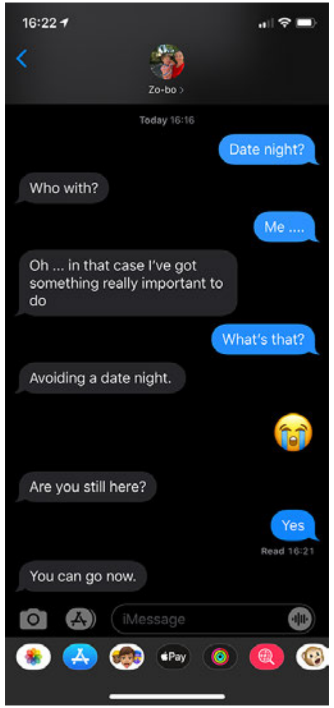
$$R^2 = \frac{SS_M}{SS_T} = \frac{13.25}{22.40} = 0.59$$

- Adjusted R^2
  - An estimate of R^2 in the population (shrinkage)

```
ite_lm <- lm(belief ~ repetition, data = ite_tib)
broom::glance(ite_lm)
```

r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.residual	nobs
0.59	0.54	1.07	11.61	0.01	1	-13.74	33.48	34.39	9.14	8	10

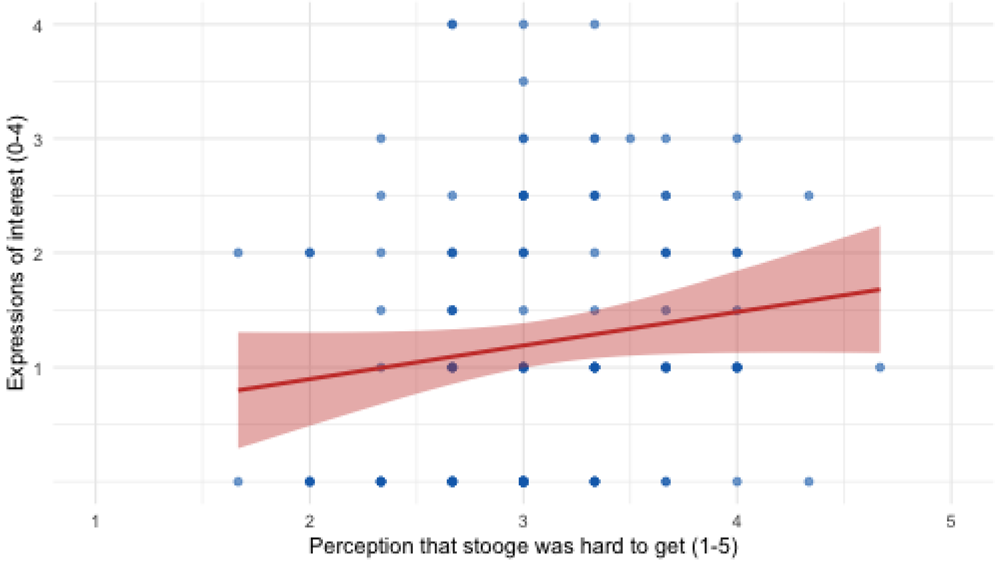
An example of measuring fit



- Does playing hard to get work?[1]
- Heterosexual participants conversed with an opposite-sex confederate over Instant Messenger for 8 mins
- Interest: Final message coded for the number of expressions of romantic interest (range 0 to 4)
- Hard to get: 3 items rated 1 (not at all) and 5 (very much so)
  - *The other participant is hard to get*
- Mate value: 4 items rated 1 (not at all) and 5 (very much so)
  - *I perceive the other participant as a valued mate*

[1] Birnbaum et al. (2020). *Journal of Social and Personal Relationships*. Study 3.

$$\hat{\text{interest}}_i = \hat{b}_0 + \hat{b}_1 \text{hard to get}_i + e_i$$



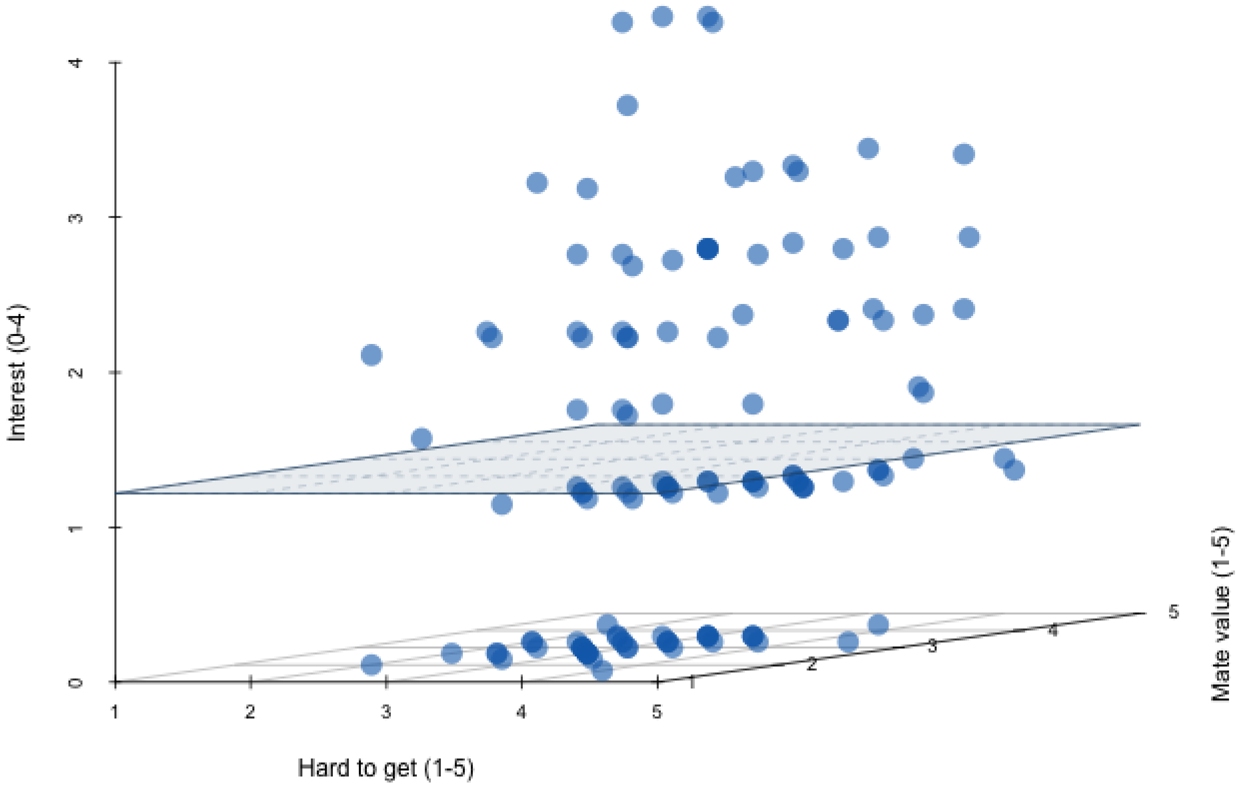
Effect	df	SS	MS	F	p
hard_to_get	1	3.844	3.844	3.115	0.08
Residuals	126	155.531	1.234	NA	NA

- The model is not a significant fit of the data $F(1, 126) = 3.11$, $p = 0.08$.

r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.residual	nobs
0.02	0.02	1.11	3.11	0.08	1	-194.09	394.18	402.74	155.53	126	128

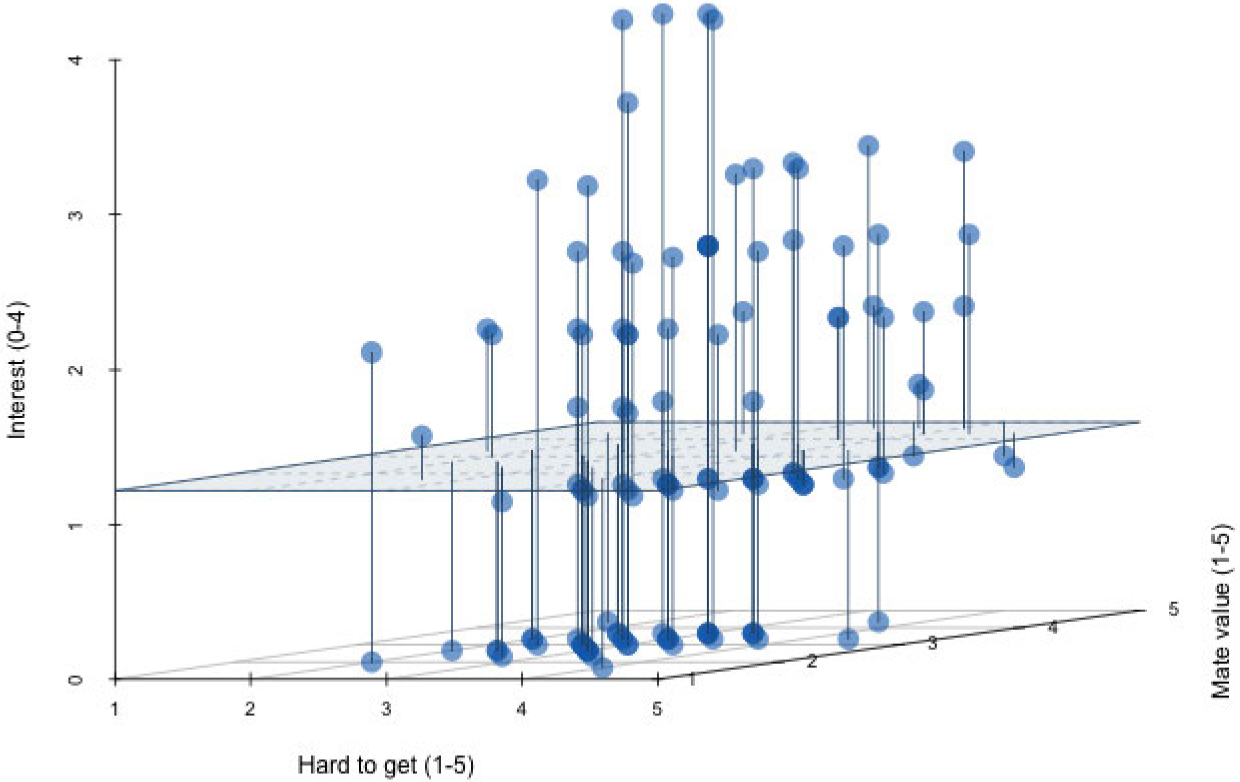
Extending the linear model

$$\hat{\text{interest}}_i = \hat{b}_0 + \hat{b}_1 \text{hard to get}_i + \hat{b}_2 \text{mate value}_i + e_i$$



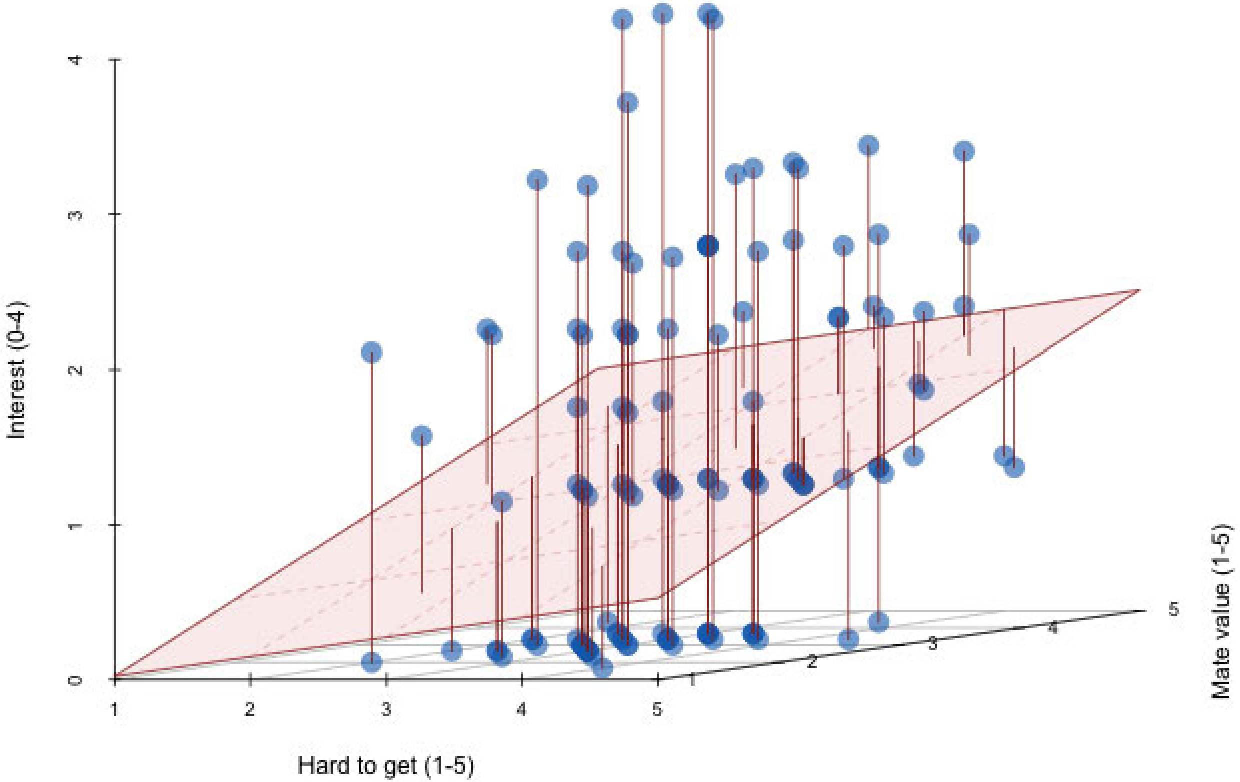
Extending the linear model (SS_T)

$$\hat{\text{interest}}_i = \hat{b}_0 + \hat{b}_1 \text{hard to get}_i + \hat{b}_2 \text{mate value}_i + e_i$$



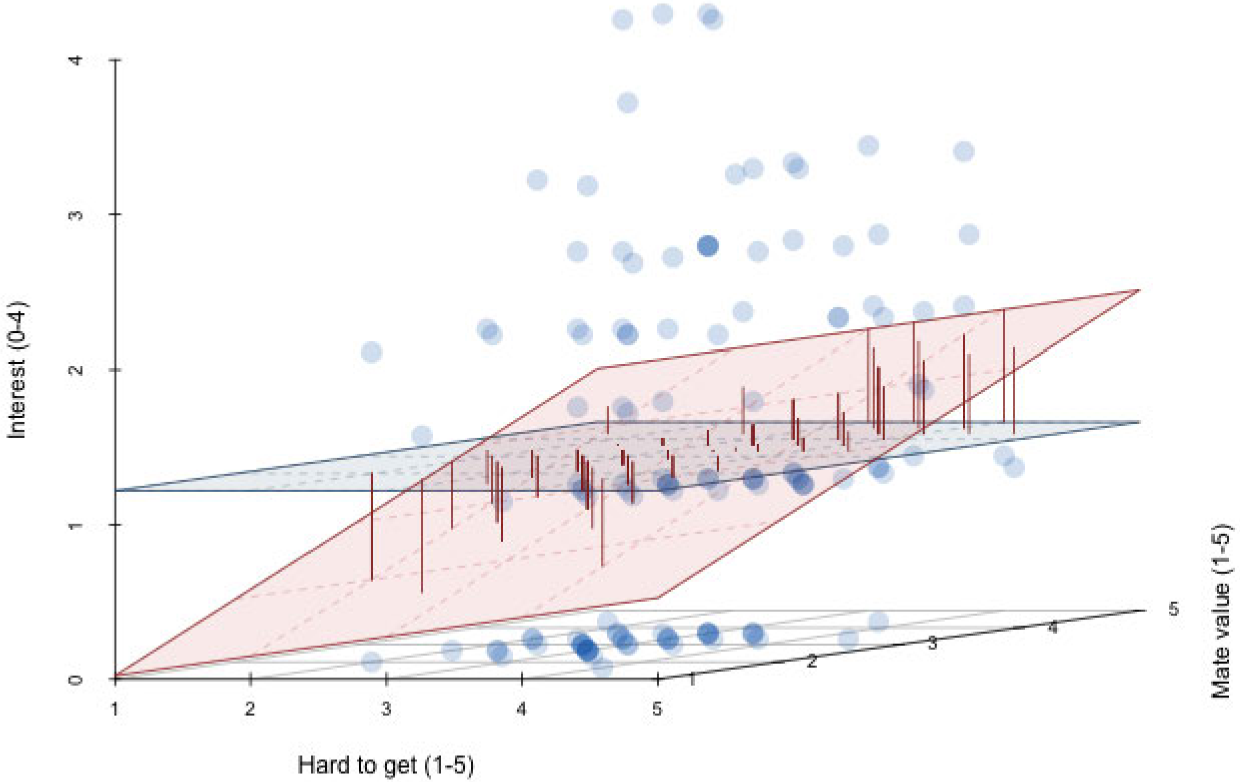
Extending the linear model (SS_R)

$$\hat{\text{interest}}_i = \hat{b}_0 + \hat{b}_1 \text{hard to get}_i + \hat{b}_2 \text{mate value}_i + e_i$$



Extending the linear model (SS_M)

$$\widehat{\text{interest}}_i = \hat{b}_0 + \hat{b}_1 \text{hard to get}_i + \hat{b}_2 \text{mate value}_i + e_i$$



Overall fit of the model

$$\widehat{\text{interest}}_i = \hat{b}_0 + \hat{b}_1 \text{hard to get}_i + \hat{b}_2 \text{mate value}_i + e_i$$

r.squared	adj.r.squared	statistic	df	df.residual	p.value
0.064	0.049	4.274	2	125	0.016

- The model is a significant fit of the data $F(2, 125) = 4.27$, $p = 0.016$.

Parameter estimates

$$\widehat{\text{interest}}_i = \hat{b}_0 + \hat{b}_1 \text{hard to get}_i + \hat{b}_2 \text{mate value}_i + e_i$$

Parameter estimates

$$\widehat{\text{interest}}_i = -0.490 + 0.126\ \text{hard to get}_i + 0.386\ \text{mate value}_i + e_i$$

term	estimate	std.error	statistic	p.value
(Intercept)	-0.490	0.621	-0.788	0.432
hard_to_get	0.126	0.179	0.702	0.484
mate_value	0.386	0.167	2.308	0.023

- As the perception that the other person was hard to get increased by 1 (on a scale from 1-5), 0.126 more expressions of interest were made (when mate value is constant)
    - This effect is not significant, t = 0.7, p = 0.484
    - This is the effect of 'hard to get' on interest adjusted for the effect of 'mate value'
- As the perception of mate value increased by 1 (on a scale from 1-5), 0.386 more expressions of interest were made (when perceptions of being hard to get are constant)
    - This effect is significant, t = 2.31, p = 0.023
    - This is the effect of 'mate value' on interest adjusted for the effect of 'hard to get'

How to enter predictors

- Hierarchical
  - Experimenter decides the order in which variables are entered into the model
  - Best for theory testing
- Forced entry
  - All predictors are entered simultaneously
- Stepwise
  - Predictors are selected using their semi-partial correlation with the outcome
  - Can produce spurious results
  - Use only for exploratory analysis

Summary

- We evaluate fit of a general linear model using Sums of Squared Errors (SS)
  - SS_T = the total variance/error in observed scores
  - SS_R = the total variance/error in predicted scores
  - SS_M = the total reduction in variance/error due to the model
- It can be useful to convert totals to averages or Mean Squared Errors (MS)
  - MS_R = the average variance/error in predicted scores
  - MS_M = the average reduction in variance/error due to the model
- R^2 is the proportion of variance in observed scores accounted for by the model
- F is the average variance accounted for by the model compared to the model's error in prediction
- Multiple predictors can be added to a linear model
  - Fit is evaluated in the same way as with a single predictor
  - *b*s are the change in the outcome associated with a unit change in the predictor when other predictors are held constant
  - Other things being equal, predictors are entered based on theory.