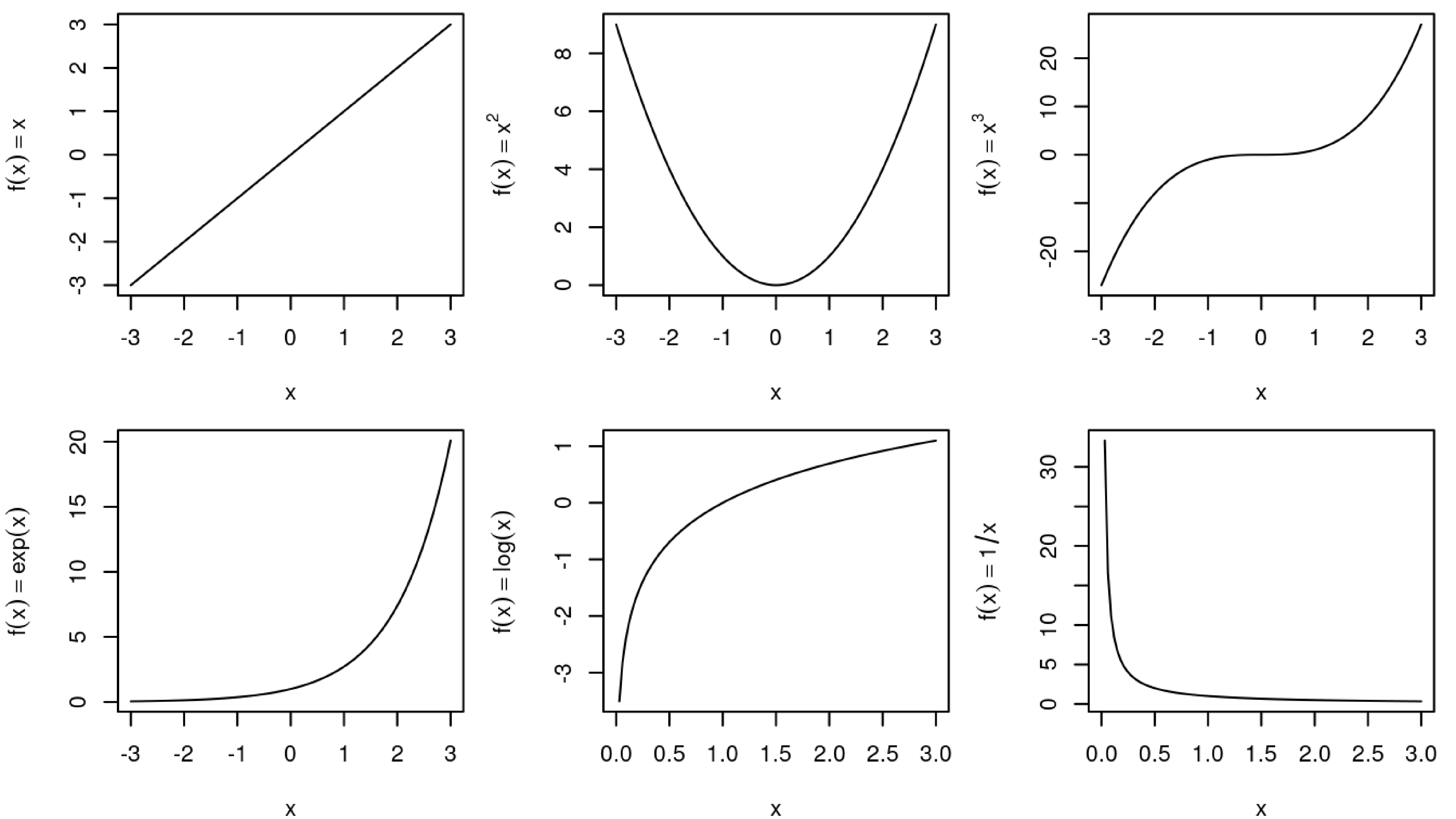
Functional Form
&
Transformation

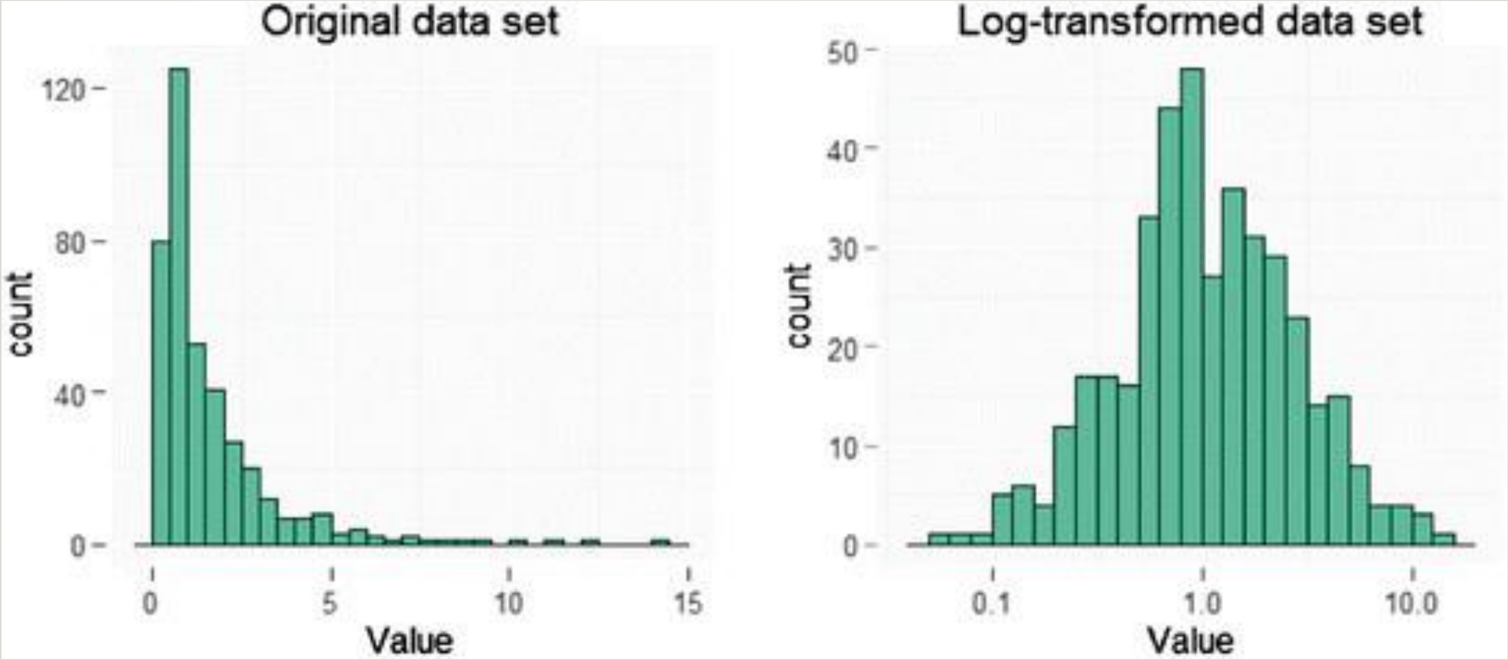


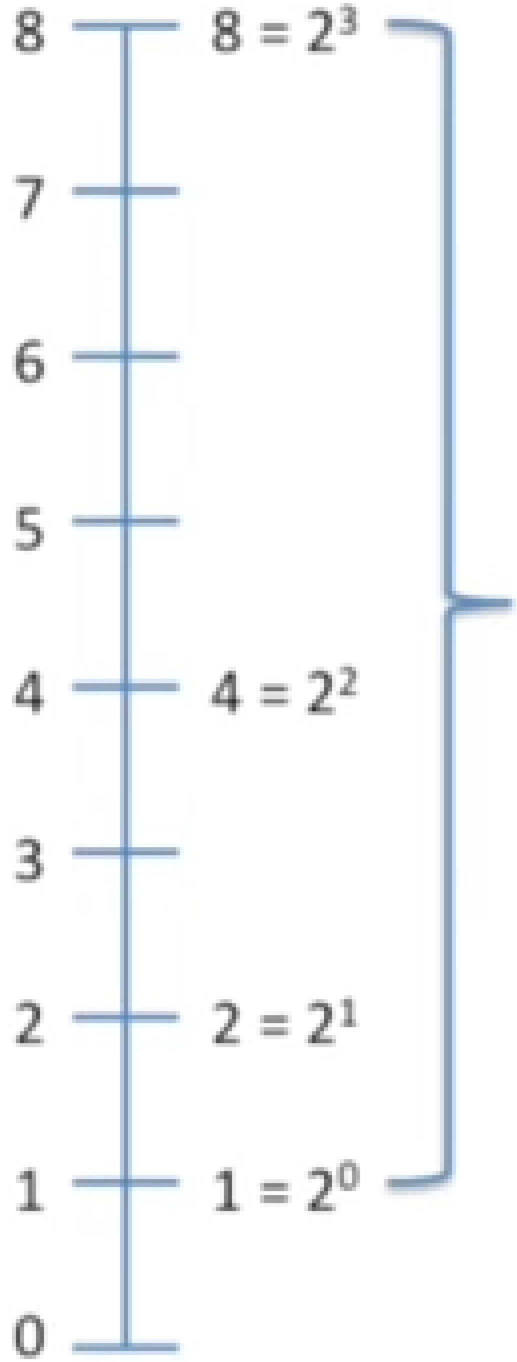
f(x) = x
f(x) = x^2
f(x) = x^3
f(x) = exp(x)
f(x) = log(x)
f(x) = 1/x
x

Log Transformation

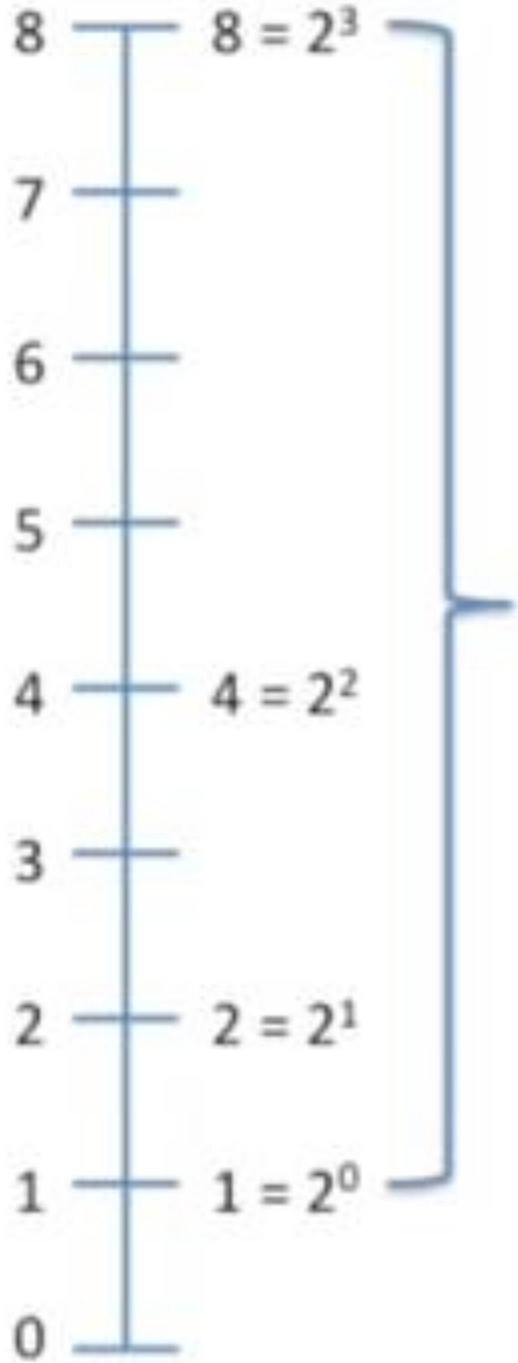
- Used mostly because of skewed distribution. Logarithm naturally reduces the dynamic range of a variable so that the differences are preserved while the scale is not that dramatically skewed.

- Imagine some people got 100,000,000 loan and some got 10000 and some 0. Any feature scaling will probably put 0 and 10000 so close to each other as the biggest number anyway pushes the boundary. Logarithm solves the issue.





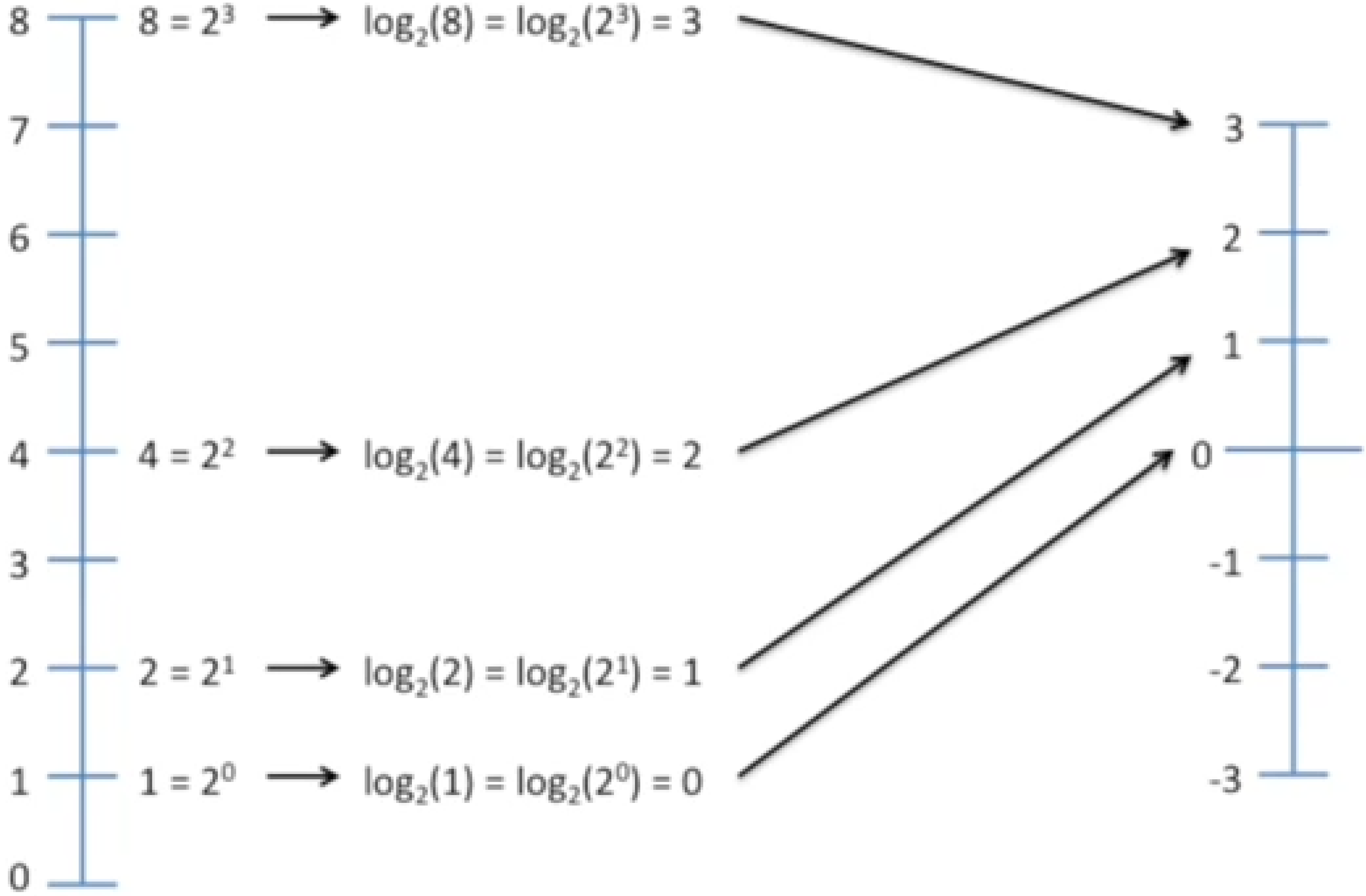
8, 4, 2, and 1 are easily rewritten as powers of 2.

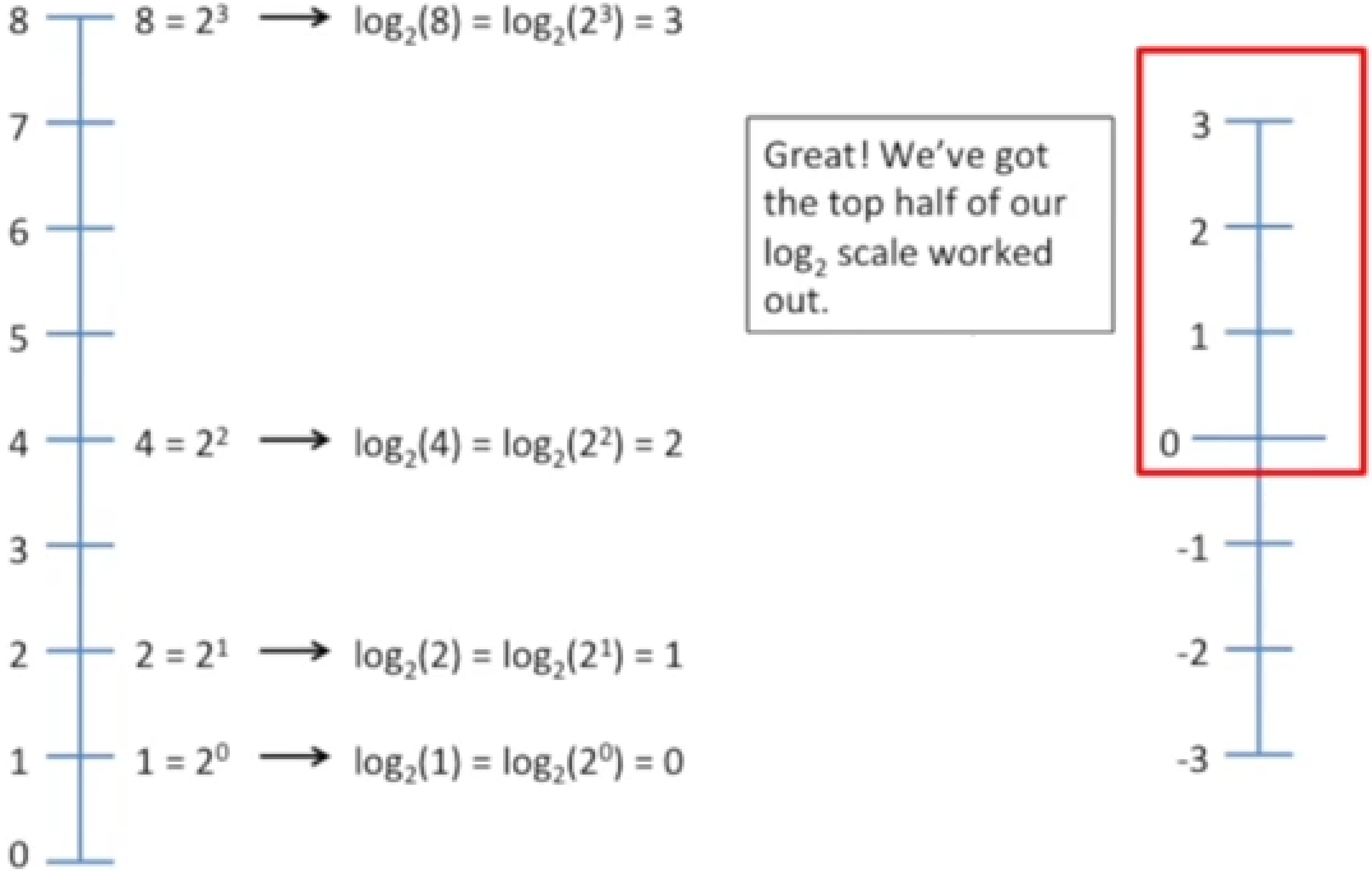


8, 4, 2, and 1 are easily rewritten as powers of 2.

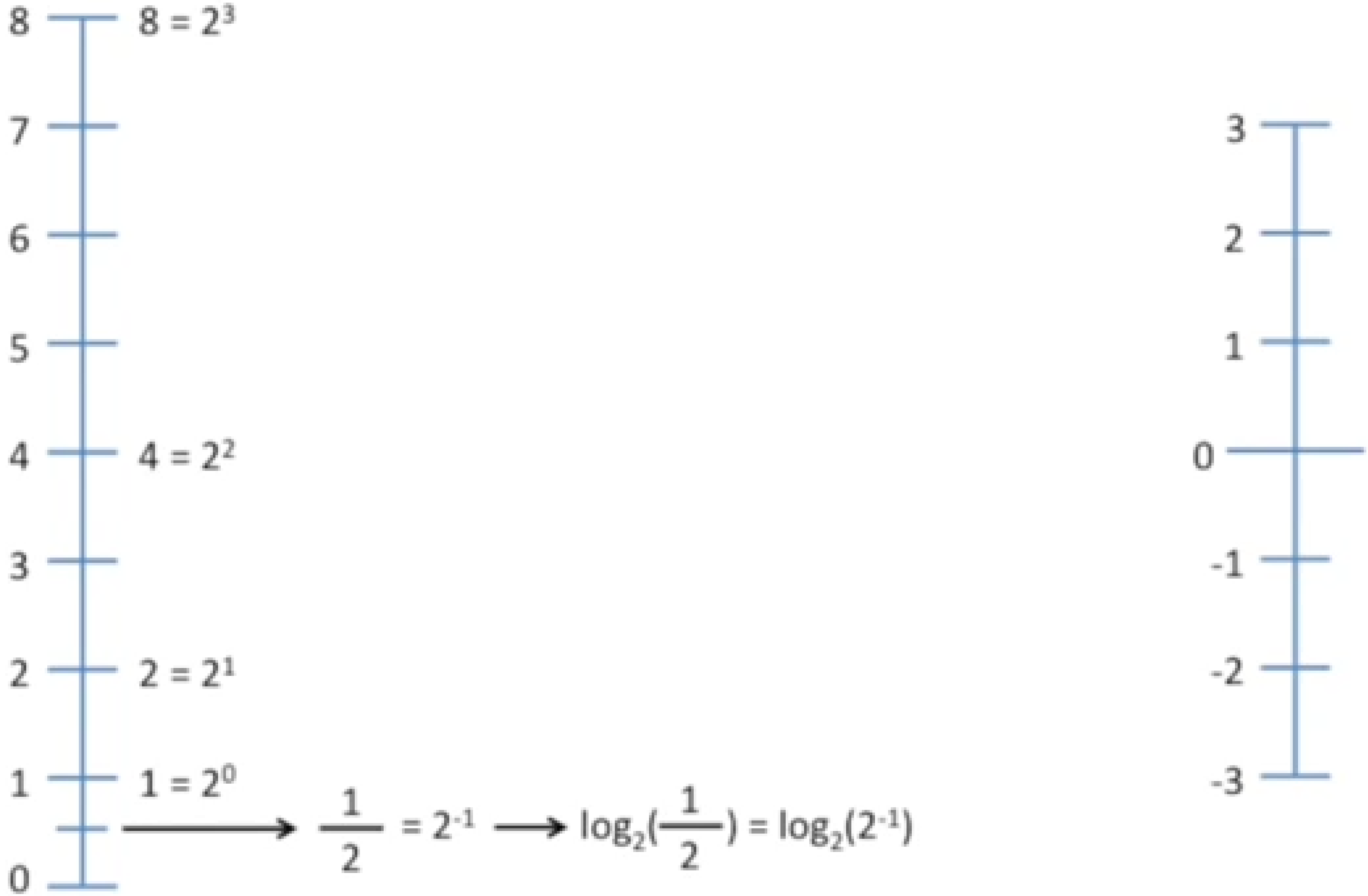
$7 = 2^{2.8}$
$6 = 2^{2.6}$
$5 = 2^{2.3}$

Other numbers can be written as powers of 2, it just isn't as neat and tidy.

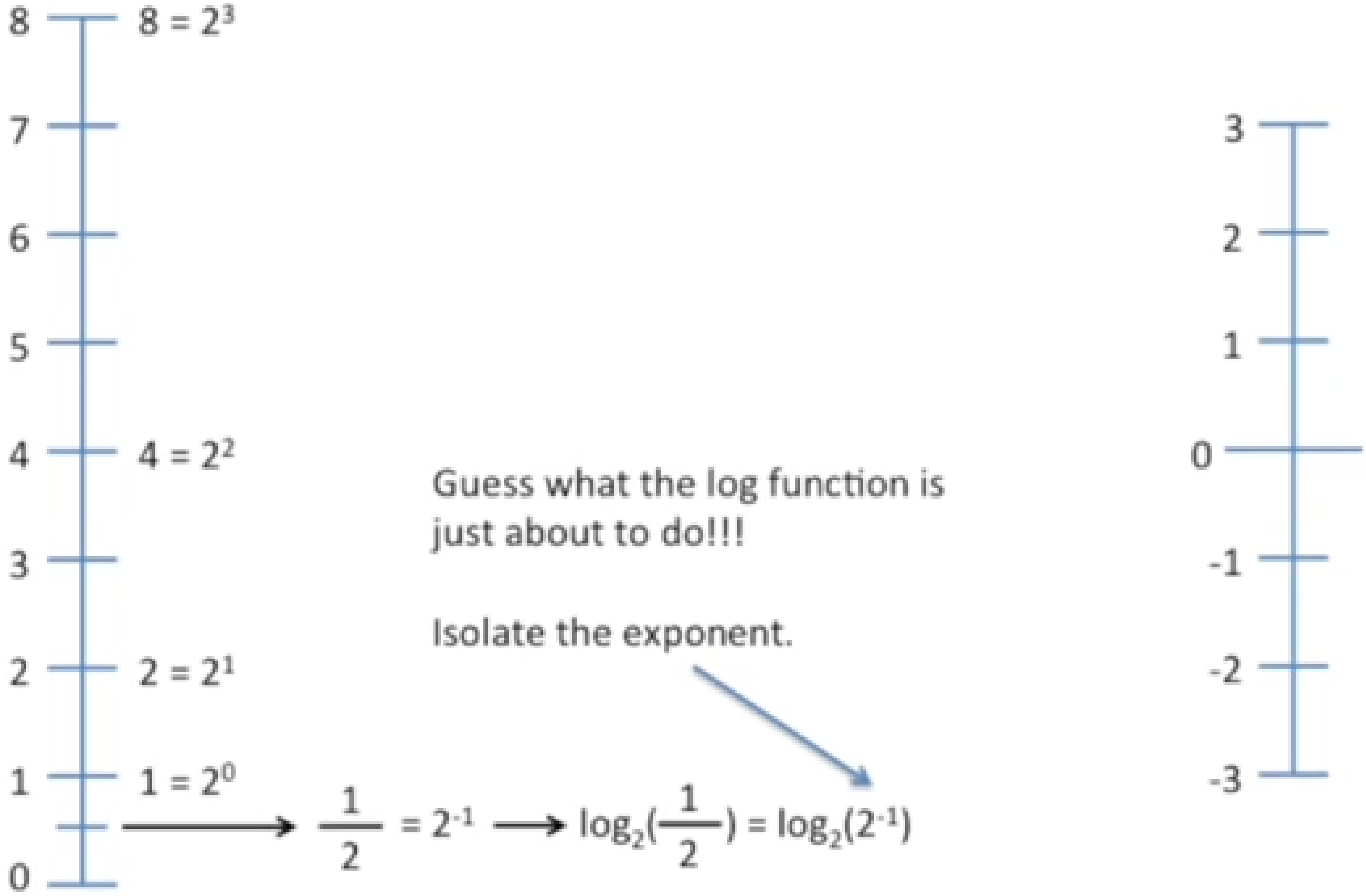

8
7
6
5
4
3
2
1
0
8 = 2^3
log_2(8) = log_2(2^3) = 3
4 = 2^2
log_2(4) = log_2(2^2) = 2
2 = 2^1
log_2(2) = log_2(2^1) = 1
1 = 2^0
log_2(1) = log_2(2^0) = 0
3
2
1
0
-1
-2
-3



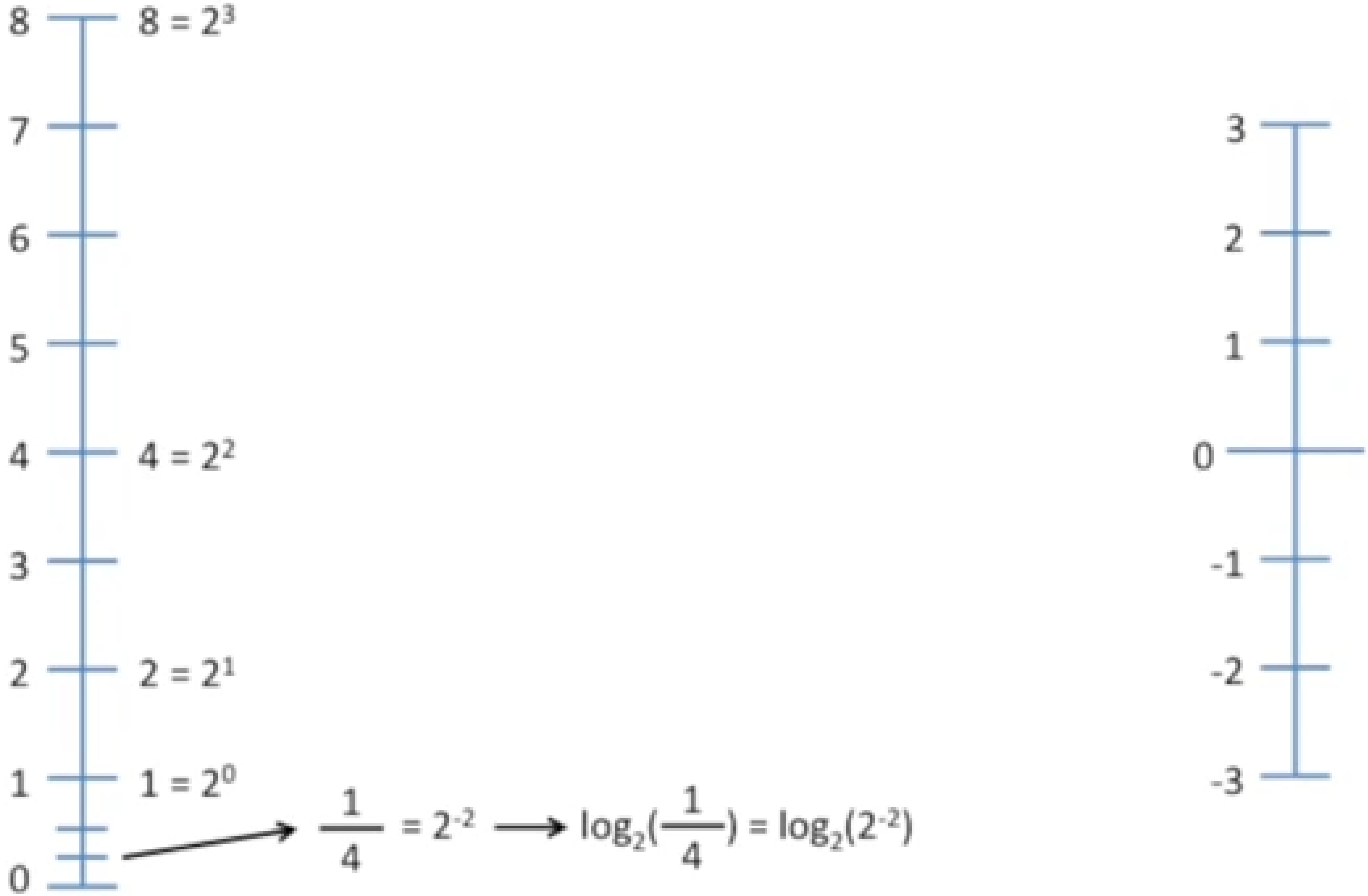
8
7
6
5
4
3
2
1
0
8 = 2^3 → log_2(8) = log_2(2^3) = 3
4 = 2^2 → log_2(4) = log_2(2^2) = 2
2 = 2^1 → log_2(2) = log_2(2^1) = 1
1 = 2^0 → log_2(1) = log_2(2^0) = 0
Great! We've got the top half of our log_2 scale worked out.
3
2
1
0
-1
-2
-3

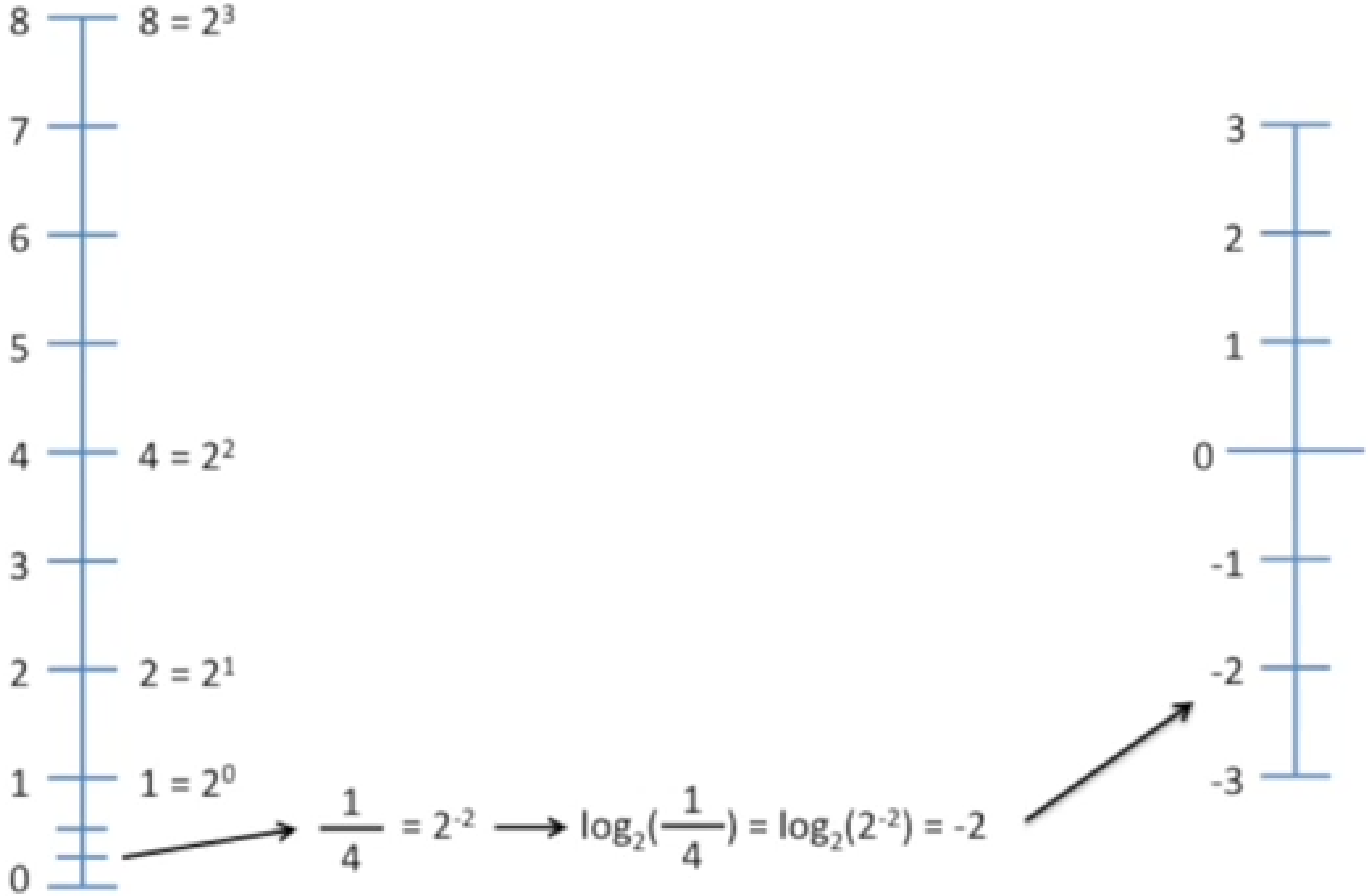


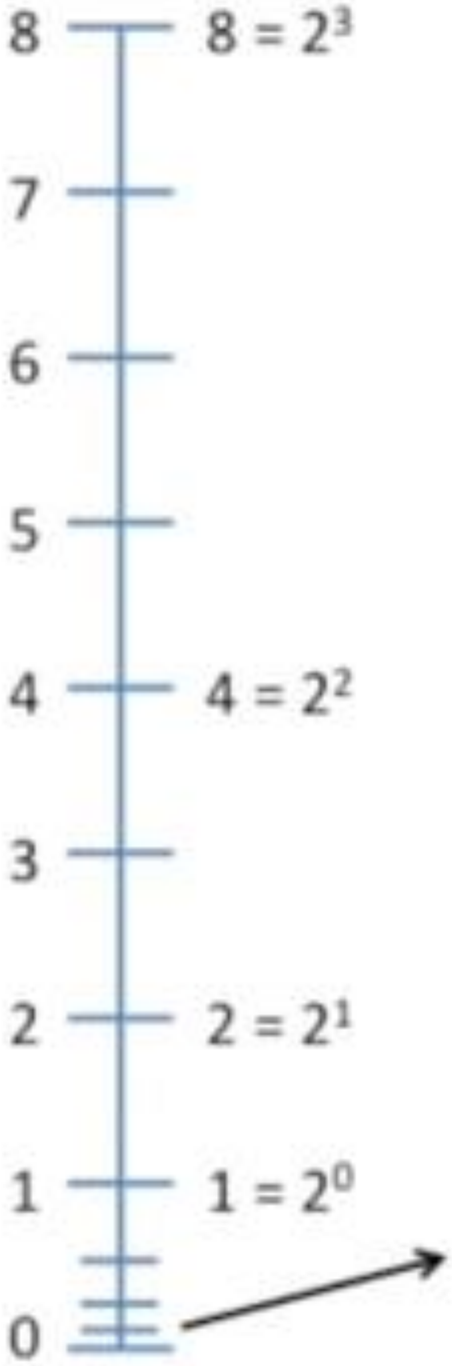
8
7
6
5
4
3
2
1
0
8 = 2^3
4 = 2^2
2 = 2^1
1 = 2^0
1/2 = 2^-1
log_2(1/2) = log_2(2^-1)
3
2
1
0
-1
-2
-3



8
7
6
5
4
3
2
1
0
$8 = 2^3$
$4 = 2^2$
$2 = 2^1$
$1 = 2^0$
$\frac{1}{2} = 2^{-1}$
$\log_2(\frac{1}{2}) = \log_2(2^{-1})$
Guess what the log function is just about to do!!!
Isolate the exponent.
3
2
1
0
-1
-2
-3


8
7
6
5
4
3
2
1
0
$8 = 2^3$
$4 = 2^2$
$2 = 2^1$
$1 = 2^0$
$\frac{1}{4} = 2^{-2}$
$\log_2(\frac{1}{4}) = \log_2(2^{-2})$
3
2
1
0
-1
-2
-3

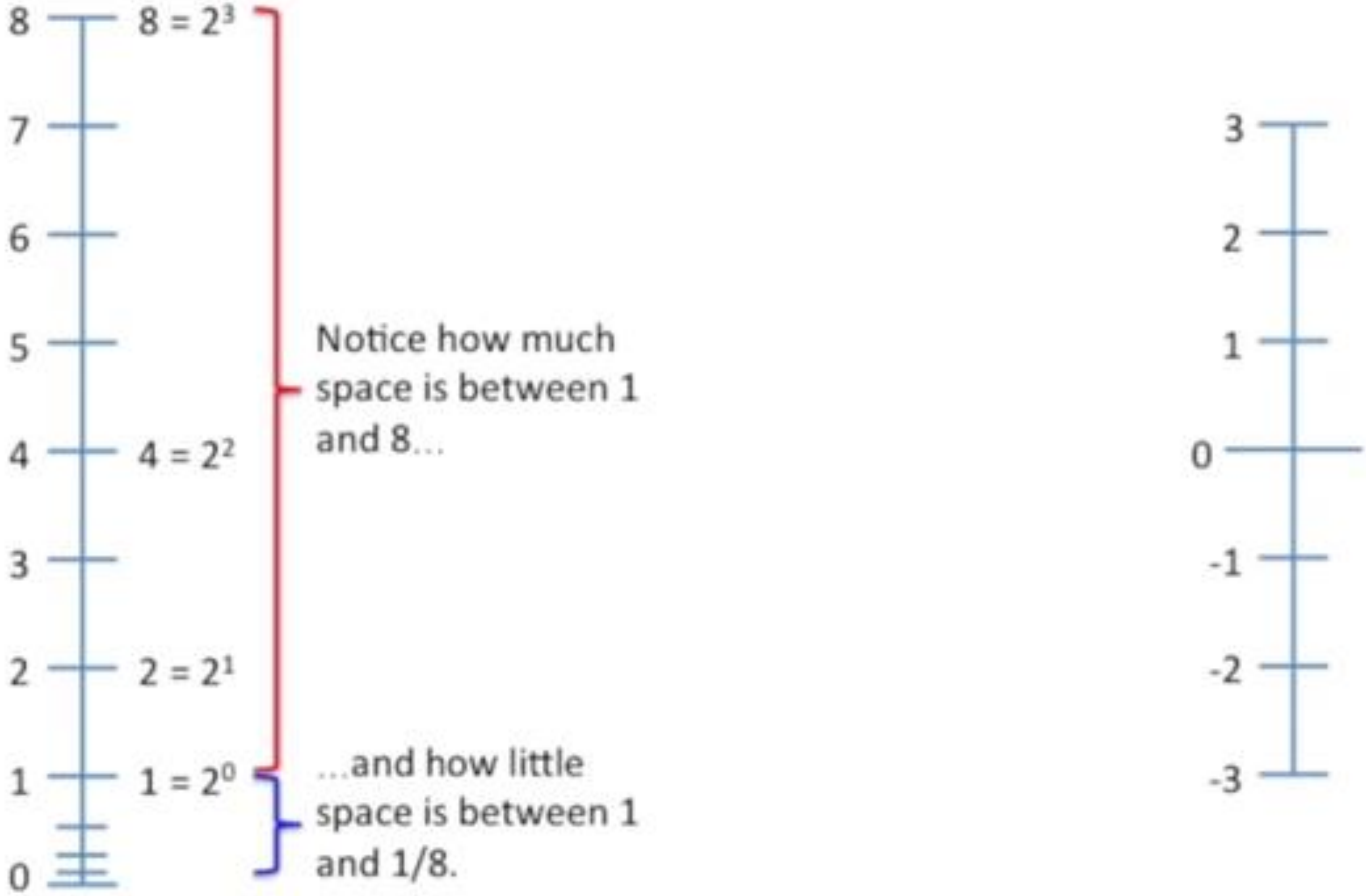

8
7
6
5
4
3
2
1
0
8 = 2^3
4 = 2^2
2 = 2^1
1 = 2^0
1/4 = 2^-2
log_2(1/4) = log_2(2^-2) = -2
3
2
1
0
-1
-2
-3

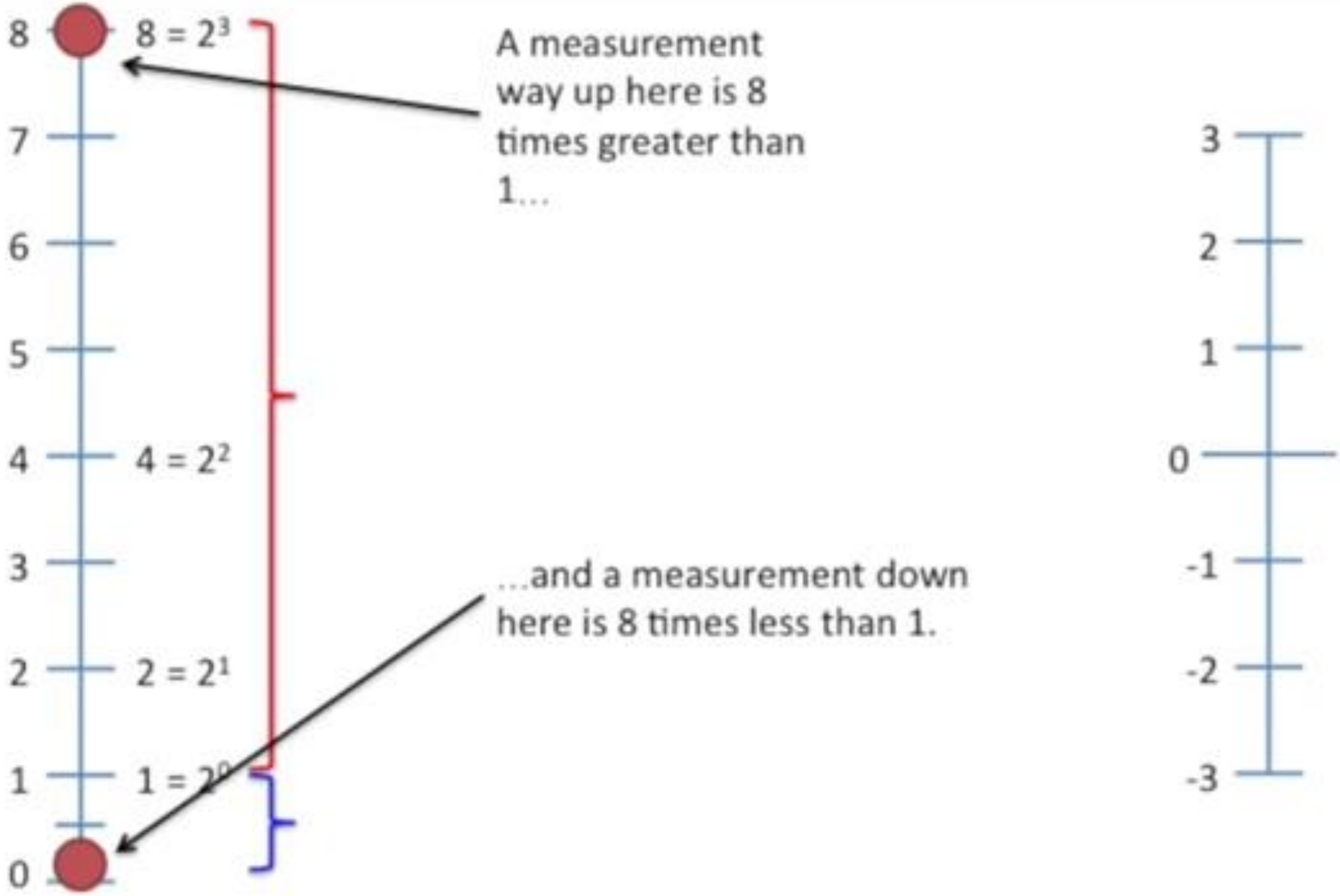


You get the idea…

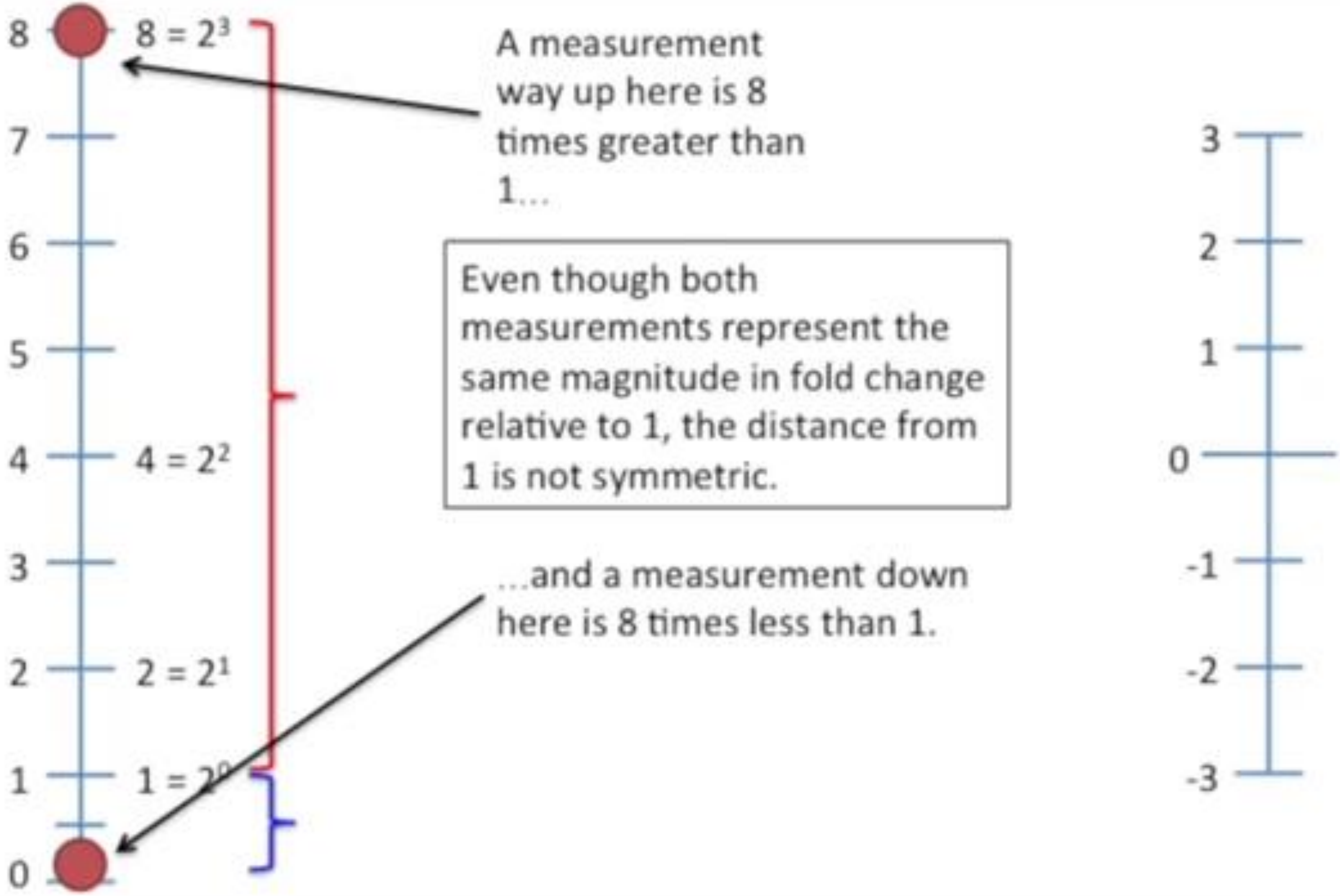
The log function isolates the exponent.

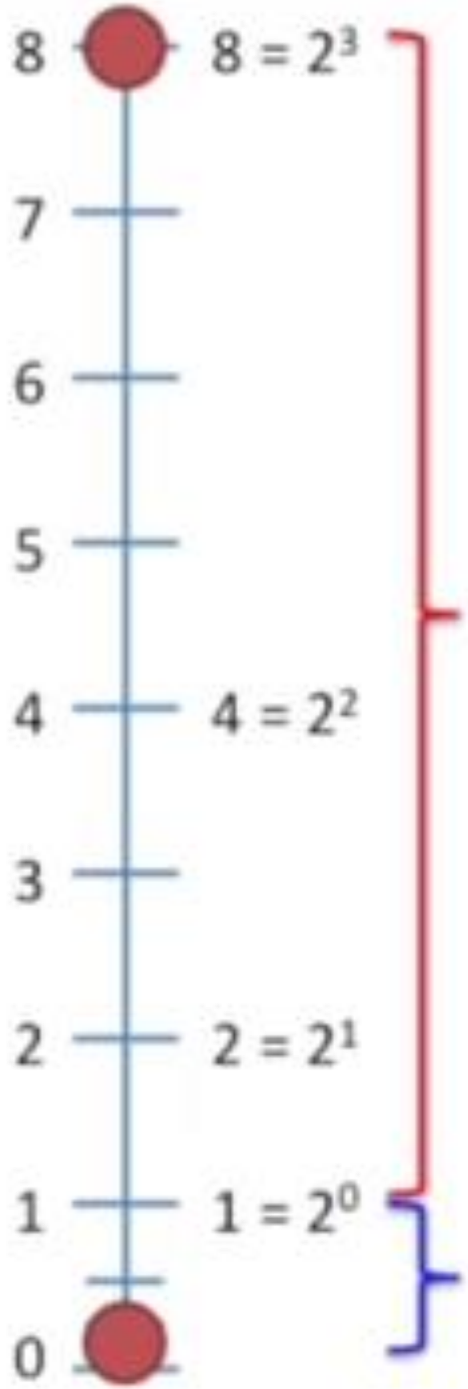
$$\frac{1}{8} = 2^{-3} \longrightarrow \log_2\left(\frac{1}{8}\right) = \log_2(2^{-3}) = -3$$


8
7
6
5
4
3
2
1
0
8 = 2³
4 = 2²
2 = 2¹
1 = 2⁰
Notice how much space is between 1 and 8…
…and how little space is between 1 and 1/8.
3
2
1
0
-1
-2
-3



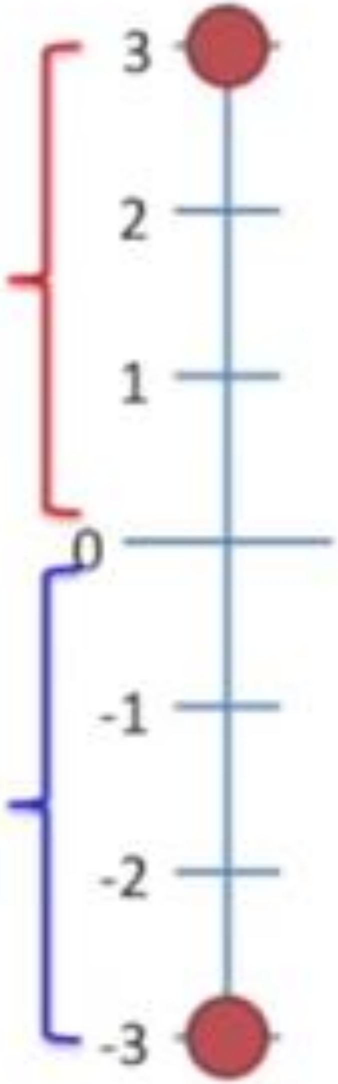
8
7
6
5
4
3
2
1
0
8 = 2³
4 = 2²
2 = 2¹
1 = 2⁰
A measurement way up here is 8 times greater than 1…
…and a measurement down here is 8 times less than 1.
3
2
1
0
-1
-2
-3

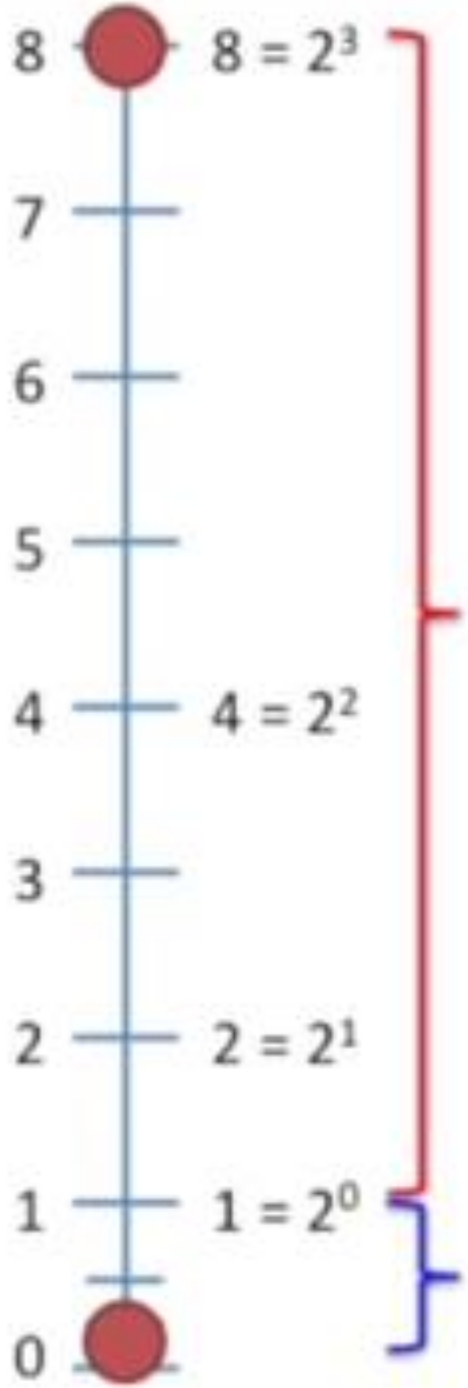

8
7
6
5
4
3
2
1
0
8 = 2^3
4 = 2^2
2 = 2^1
1 = 2^0
A measurement way up here is 8 times greater than 1…
Even though both measurements represent the same magnitude in fold change relative to 1, the distance from 1 is not symmetric.
…and a measurement down here is 8 times less than 1.
3
2
1
0
-1
-2
-3



In contrast, the magnitude is equidistant on the log axis.

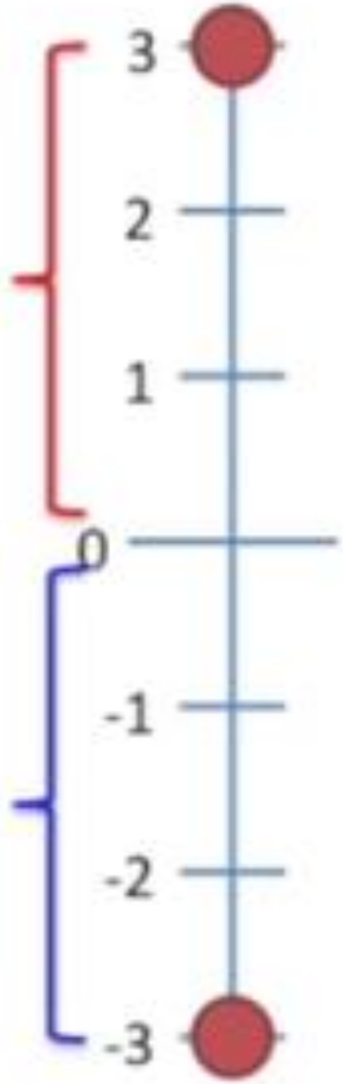
This is why fold changes should always be plotted on log axes.

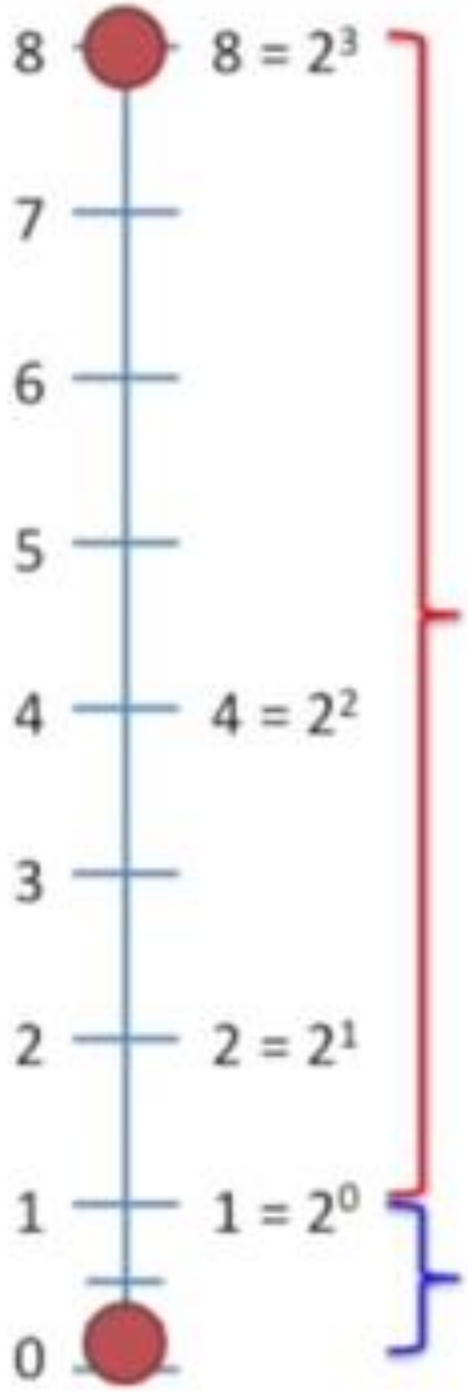




Take home message so far…

1) “logs” isolate exponents.

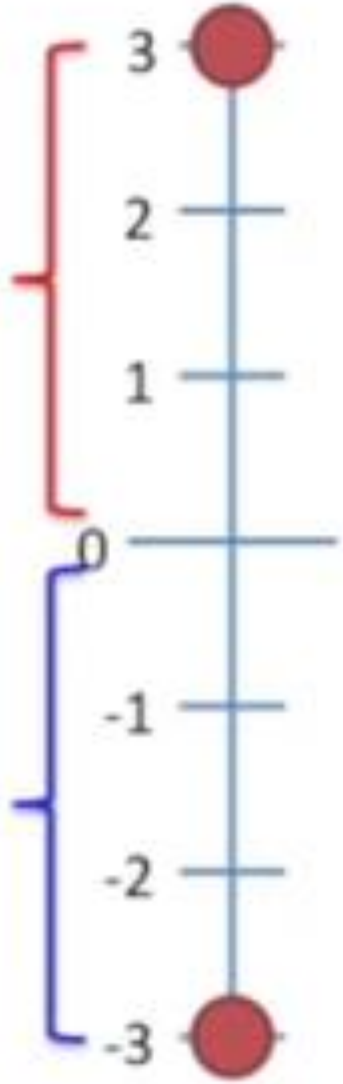




Take home message so far…

1) "logs" isolate exponents.

$$\log_2(8) = \log_2(2^3) = 3$$

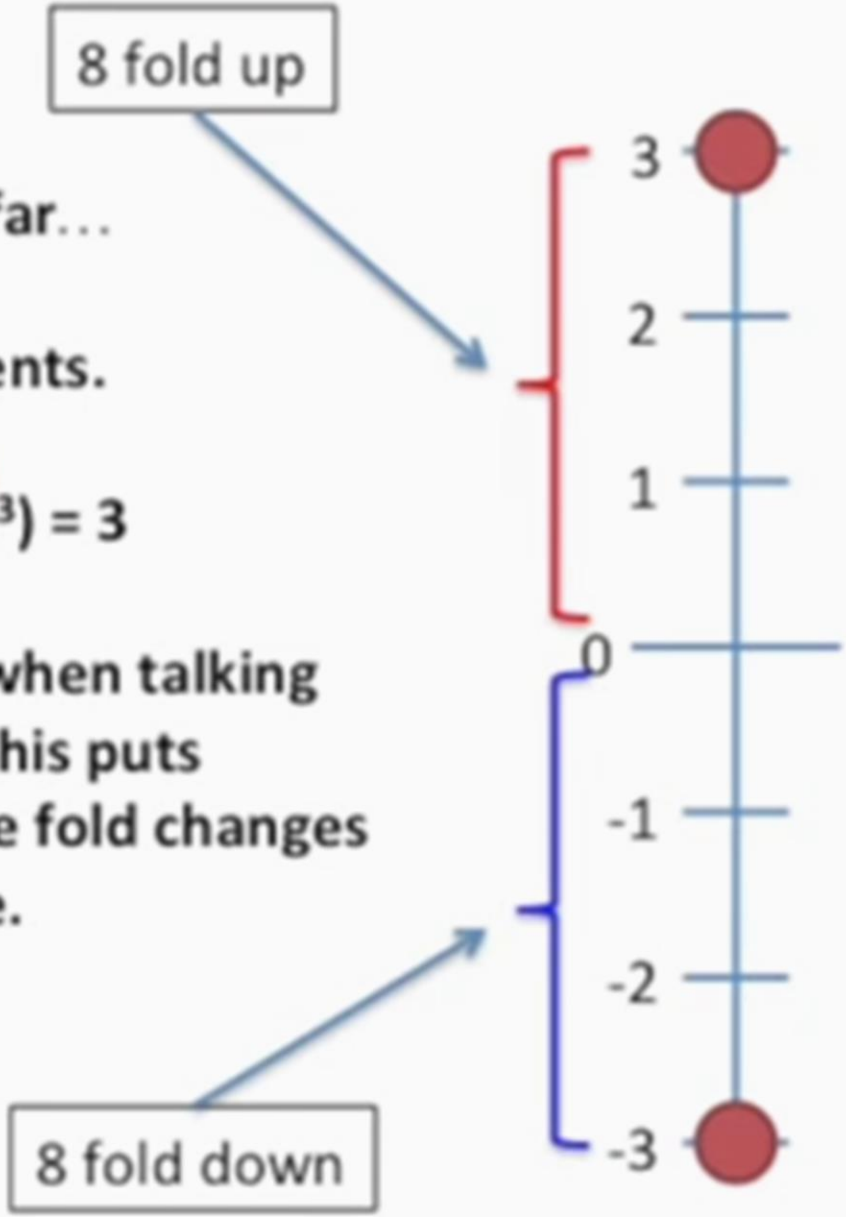


Take home message so far…

1) "logs" isolate exponents.

$$\log_2(8) = \log_2(2^3) = 3$$

2) Use a log scale/axis when talking about fold change. This puts positive and negative fold changes on a symmetric scale.



Regression

Functional Form & Transformations

Part a) Dealing with numerical variables

* Non-linear relationships (squared, inverse etc)

* Logarithms

Regression

Functional Form & Transformations

Part a) Dealing with numerical variables

* Non-linear relationships (squared, inverse etc)

* Logarithms

Part b) Dealing with categorical X variables

* Dummy variables

* Interaction variables

Regression

Functional Form & Transformations

Part a) Dealing with numerical variables

* Non-linear relationships (squared, inverse etc)

* Logarithms

Part b) Dealing with categorical X variables

* Dummy variables

* Interaction variables

Part c) Dealing with categorical Y variables

* Logit models

Regression

Dataset:

Jaybob's Used Car Sales (jaybob.csv)

Variables:

"Price" - advertised sale price ($AUD)

"Age" - model age (yrs)

"Odometer" - odometer reading ('000 kms)

"Pink slip" - presence of RWC (1= yes, 0=no)

"Sold" - whether car sold (1=yes, 0=no)

Regression

Dataset:

Jaybob's Used Car Sales (jaybob.csv)

	A	B	C	D	E	F
1	Car ID	Price	Age	Odometer	Pink slip	Sold?
2	1	$ 1,000	28	30.298	1	1
3	2	$ 9,000	40	19.647	1	0
4	3	$ 500	58	170.270	0	1
5	4	$ 3,000	12	68.394	1	1
6	5	$ 9,500	3	11.662	0	0
7	6	$ 1,500	23	87.973	0	0
8	7	$ 4,000	4	3.496	1	0
9	8	$ 2,000	13	40.986	1	1
10	9	$ 2,500	5	21.098	1	1

Numerical variables

Model 1

$$Price_i = \beta_0 + \beta_1 Age_i + \beta_2 Odometer_i + \varepsilon_i$$

Numerical variables

Model 1

$$Price_i = \beta_0 + \beta_1 Age_i + \beta_2 Odometer_i + \varepsilon_i$$

DV: Price	Coef	SE	t	P-value
Intercept	4615.901	792.153	5.83	0.0000
Age	98.922	29.974	3.30	0.0014
Odometer	-23.029	6.284	-3.66	0.0004

$R^2 = 0.171$

Numerical variables

Model 1

$$Price_i = \beta_0 + \beta_1 Age_i + \beta_2 Odometer_i + \varepsilon_i$$

DV: Price	Coef	SE	t	P-value
Intercept	4615.901	792.153	5.83	0.0000
Age	98.922	29.974	3.30	0.0014
Odometer	-23.029	6.284	-3.66	0.0004

$$R^2 = 0.171$$

$$\widehat{Price}_i = 4615.9 + 98.9(Age_i) - 23.0(Odometer_i)$$

Numerical variables

Model 1

$$Price_i = \beta_0 + \beta_1 Age_i + \beta_2 Odometer_i + \varepsilon_i$$

DV: Price	Coef	SE	t	P-value
Intercept	4615.901	792.153	5.83	0.0000
Age	98.922	29.974	3.30	0.0014
Odometer	-23.029	6.284	-3.66	0.0004

$R^2 = 0.171$

$$\widehat{Price}_i = 4615.9 + 98.9(Age_i) - 23.0(Odometer_i)$$

For every additional year in age, the car can be expected to increase in price by $98.92, on average, holding odometer constant.

Numerical variables

Model 1

$$Price_i = \beta_0 + \beta_1 Age_i + \beta_2 Odometer_i + \varepsilon_i$$

DV: Price	Coef	SE	t	P-value
Intercept	4615.901	792.153	5.83	0.0000
Age	98.922	29.974	3.30	0.0014
Odometer	-23.029	6.284	-3.66	0.0004

$R^2 = 0.171$

$$\widehat{Price}_i = 4615.9 + 98.9(Age_i) - 23.0(Odometer_i)$$

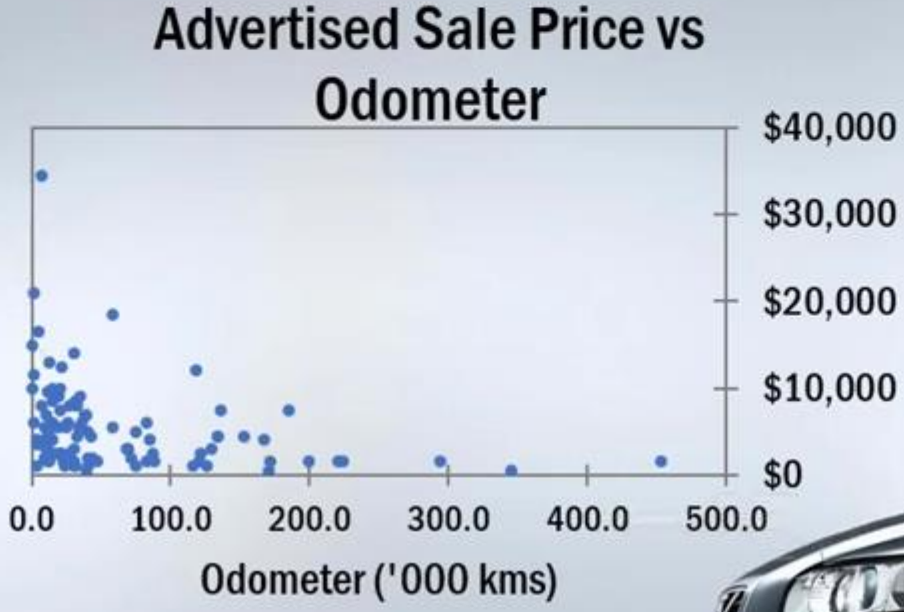
For every additional thousand km on the odometer, the price is expected to decrease by $23.03, on average, holding age constant.

Numerical variables

<u>Check scatter plots!</u>

Numerical variables

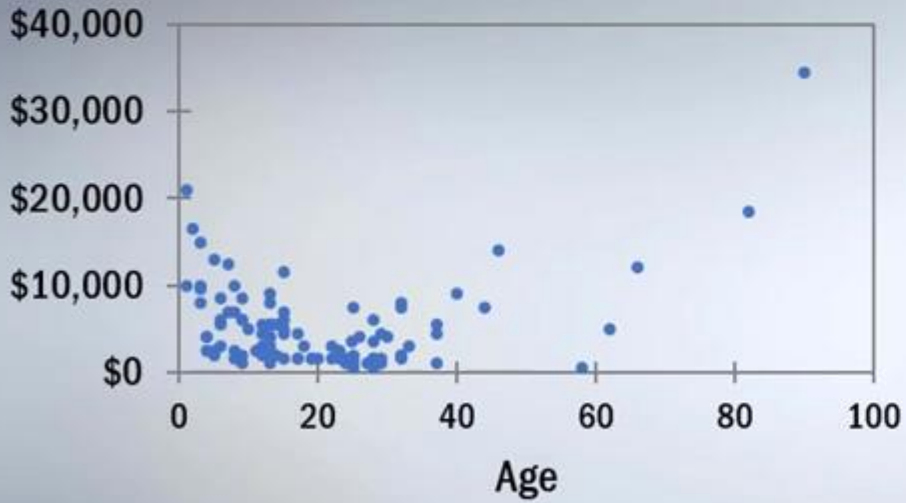
Check scatter plots!



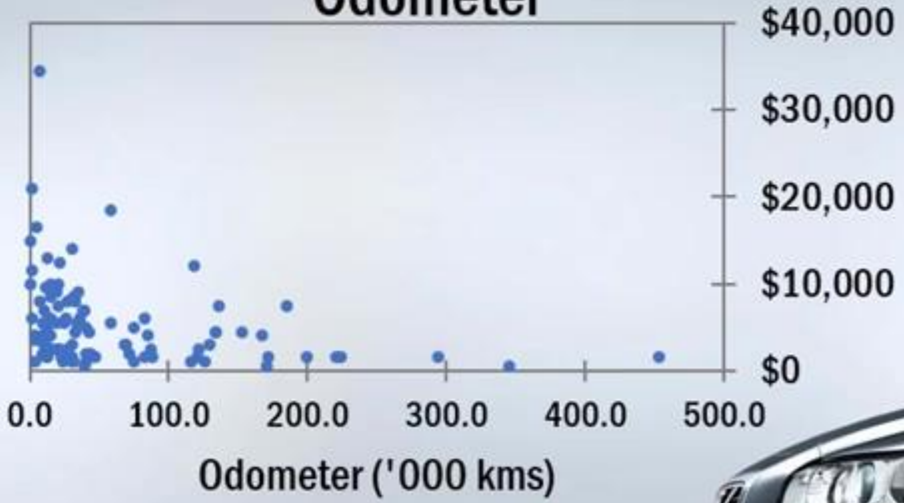
Numerical variables

Check scatter plots!

Advertised Sale Price vs Age



Advertised Sale Price vs Odometer



$$Price_i = \beta_0 + \beta_1 Age_i + \beta_2 Age_i^2 + \beta_3 \left(\frac{1}{Odometer_i}\right) + \varepsilon_i$$

Numerical variables

Model 2

$$Price_i = \beta_0 + \beta_1(Age_i) + \beta_2(Age_i)^2 + \beta_2\left(\frac{1}{Odometer_i}\right) + \varepsilon_i$$

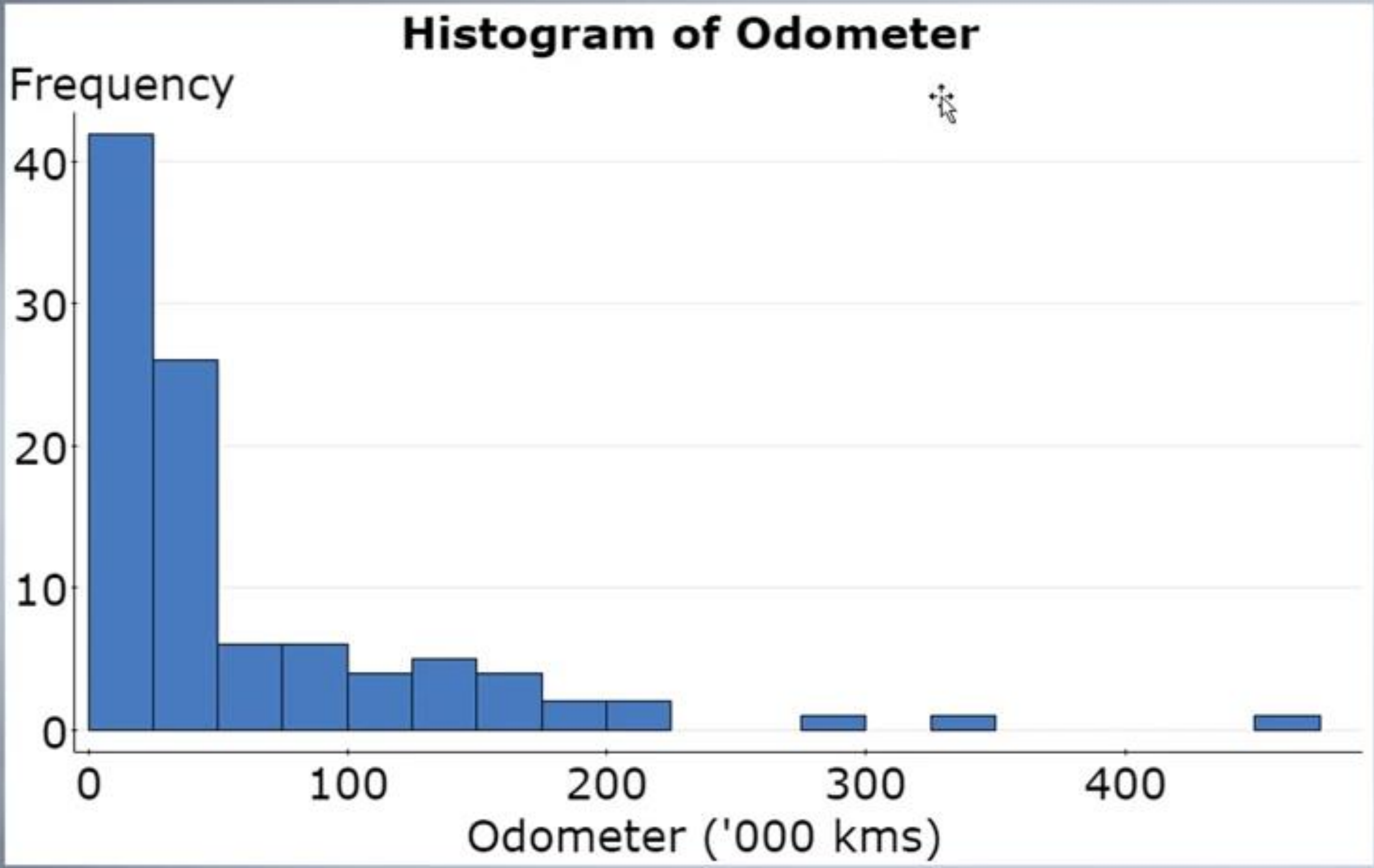
DV: Price	Coef	SE	t	P-value
Intercept	8809.879	806.803	10.92	0.0000
Age	-429.664	56.393	-7.62	0.0000
Age2	7.318	0.735	9.96	0.0000
1/Odometer	1942.243	676.217	2.87	0.0050

$$R^2 = 0.585$$

$$\widehat{Price}_i = 8809.9 - 429.7(Age_i) + 7.3(Age_i)^2 + 1942.2\left({}^{1}/_{Odometer_i}\right)$$

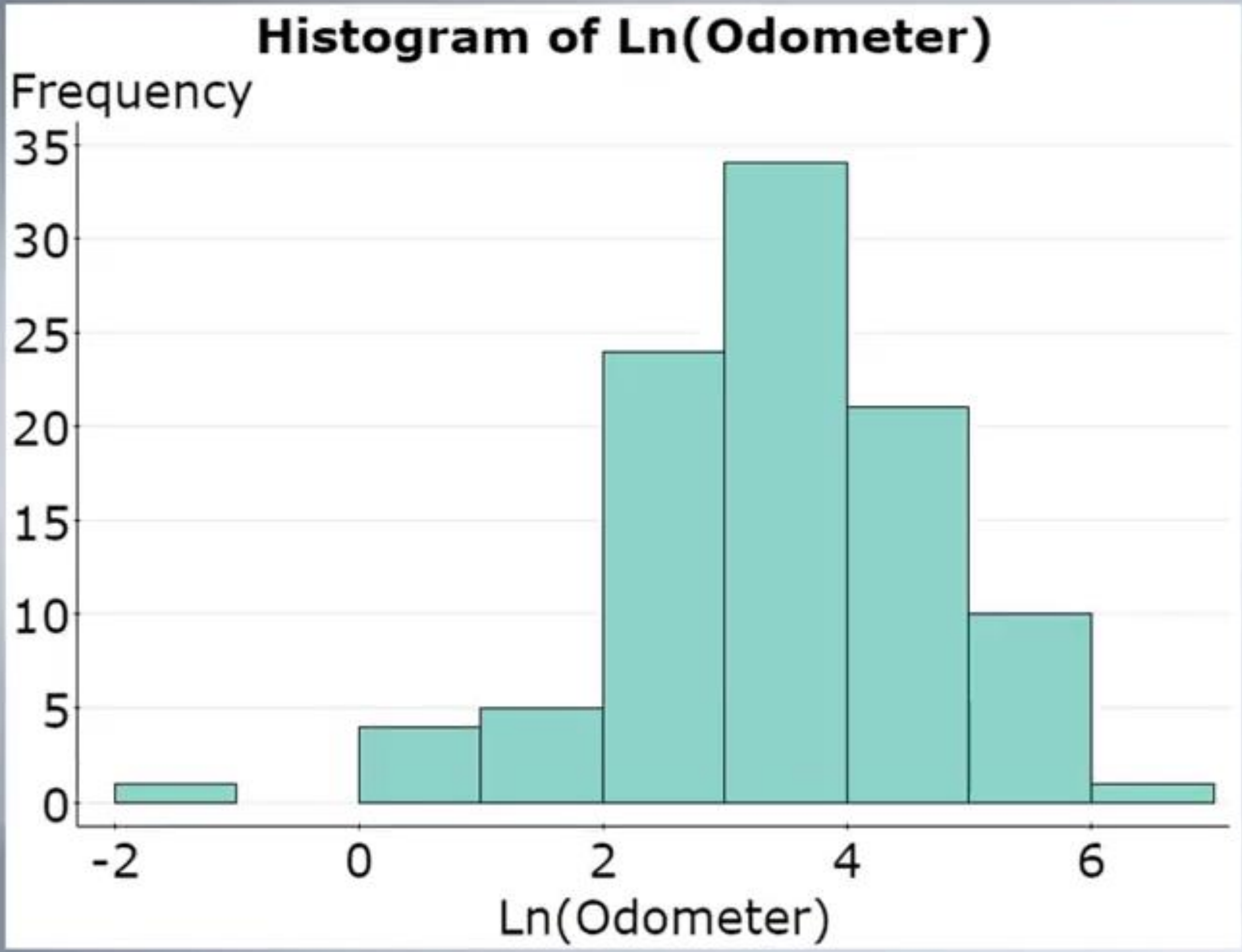
Numerical variables

Logarithms



Numerical variables

Logarithms



Numerical variables

Model 3

$$Price_i = \beta_0 + \beta_1(Age_i) + \beta_2(Age_i)^2 + \beta_3 Ln(Odometer_i) + \varepsilon_i$$

Numerical variables

Model 3

$$Price_i = \beta_0 + \beta_1(Age_i) + \beta_2(Age_i)^2 + \beta_3 Ln(Odometer_i) + \varepsilon_i$$

DV: Price	*Coef*	*SE*	*t*	*P-value*
Intercept	11863.069	940.941	12.61	0.0000
Age	-365.576	58.864	-6.21	0.0000
Age2	6.628	0.749	8.85	0.0000
Ln(Odometer)	-1079.375	272.050	-3.97	0.0001

$$R^2 = 0.613$$

$$\widehat{Price}_i = 11863.1 - 365.6\,(Age_i) + 6.63\,(Age_i)^2 - 1079.4\,Ln(Odometer_i)$$

Numerical variables

Model 3

$$\widehat{Price}_i = 11863.1 - 365.6\,(Age_i) + 6.63\,(Age_i)^2 - 1079.4\,Ln(Odometer_i)$$

Numerical variables

Model 3

$$\widehat{Price}_i = 11863.1 - 365.6\,(Age_i) + 6.63\,(Age_i)^2 - 1079.4\,Ln(Odometer_i)$$

A 1 unit increase in the natural log of the odometer reading decreases the price by $1079.40, on average, holding age constant

Numerical variables

Model 3

$$\widehat{Price}_i = 11863.1 - 365.6\,(Age_i) + 6.63\,(Age_i)^2 - 1079.4\,Ln(Odometer_i)$$

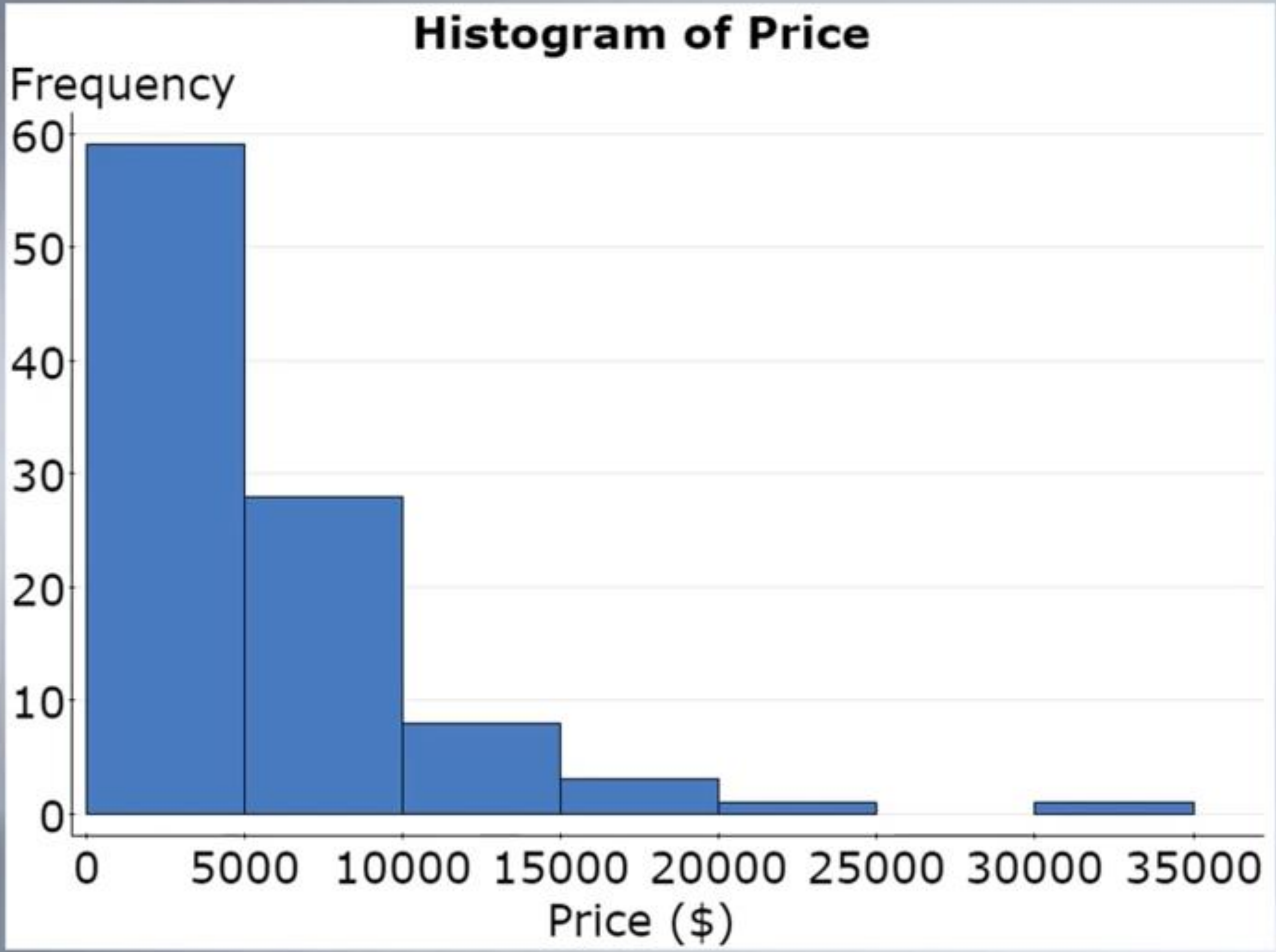
A 1% increase in the odometer reading decreases the price by:

1079.4/100 = $10.79

...on average, holding age constant.

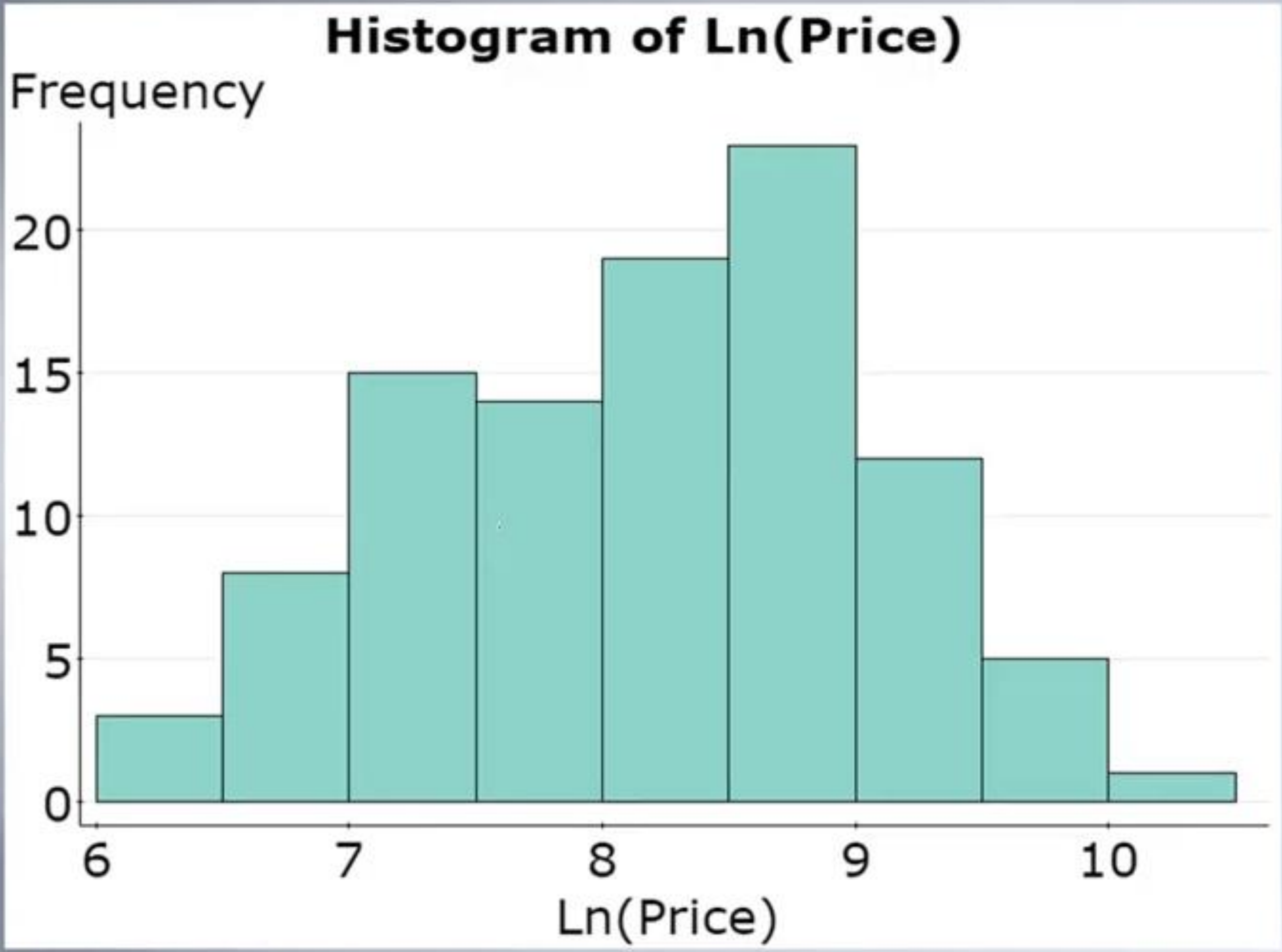
Numerical variables

Logarithms



Numerical variables

Logarithms



Numerical variables

Model 4

$$Ln(Price_i) = \beta_0 + \beta_1(Age_i) + \beta_2(Age_i)^2 + \beta_2 Ln(Odometer_i) + \varepsilon_i$$

DV: Ln(Price)	*Coef*	*SE*	*t*	*P-value*
Intercept	9.392	0.211	44.47	0.0000
Age	-0.054	0.013	-4.12	0.0001
Age2	0.001	0.000	5.00	0.0000
Ln(Odometer)	-0.197	0.061	-3.23	0.0017

$$R^2 = 0.362$$

$$Ln(Price_i) = 9.392 - 0.054\,(Age_i) + 0.001\,(Age_i)^2 - 0.197\,Ln(Odometer_i)$$

Numerical variables

Model 4

$$\widehat{Ln(Price_i)} = 9.392 - 0.054\,(Age_i) + 0.001\,(Age_i)^2 - 0.197\,Ln(Odometer_i)$$

Numerical variables

Model 4

$$\widehat{Ln(Price_i)} = 9.392 - 0.054\,(Age_i) + 0.001\,(Age_i)^2 - 0.197\,Ln(Odometer_i)$$

A 1% increase in the odometer reading decreases the price by 0.197%, on average, holding age constant

Part (b) - Categorical X variables

Categorical X variables

<u>Binary variables</u>

Pink Slip = 1   if car has roadworthy certificate

= 0   otherwise

Categorical X variables

<u>Binary variables</u>

Pink Slip = 1 if car has roadworthy certificate
= 0 otherwise

$$Price_i = \beta_0 + \beta_1(Pink\ Slip_i) + \varepsilon_i$$

Categorical X variables

Binary variables

Pink Slip = 1 if car has roadworthy certificate

= 0 otherwise

$$Price_i = \beta_0 + \beta_1(Pink\ Slip_i) + \varepsilon_i$$

DV: Price	Coef	SE	t	P-value
Intercept	3978.3	1056.29	3.76625	0.00028
Pink slip	1625.6	1203.76	1.35047	0.17998

$$\widehat{Price}_i = 3978 + 1626\ (Pink\ Slip_i)$$

Categorical X variables

Binary variables

Pink Slip = 1 if car has roadworthy certificate

= 0 otherwise

$$\widehat{Price}_i = 3978 + 1626\,(Pink\ Slip_i)$$

A car with a pink slip would command a sale price $1,626 more than a car without a pink slip, on average.

Categorical X variables

Model 5

$$\widehat{Ln(Price_i)} = 9.237 - 0.052(Age_i) + 0.001(Age_i)^2$$
$$-0.198\,Ln(Odometer_i) + 0.156(Pink\ Slip_i)$$

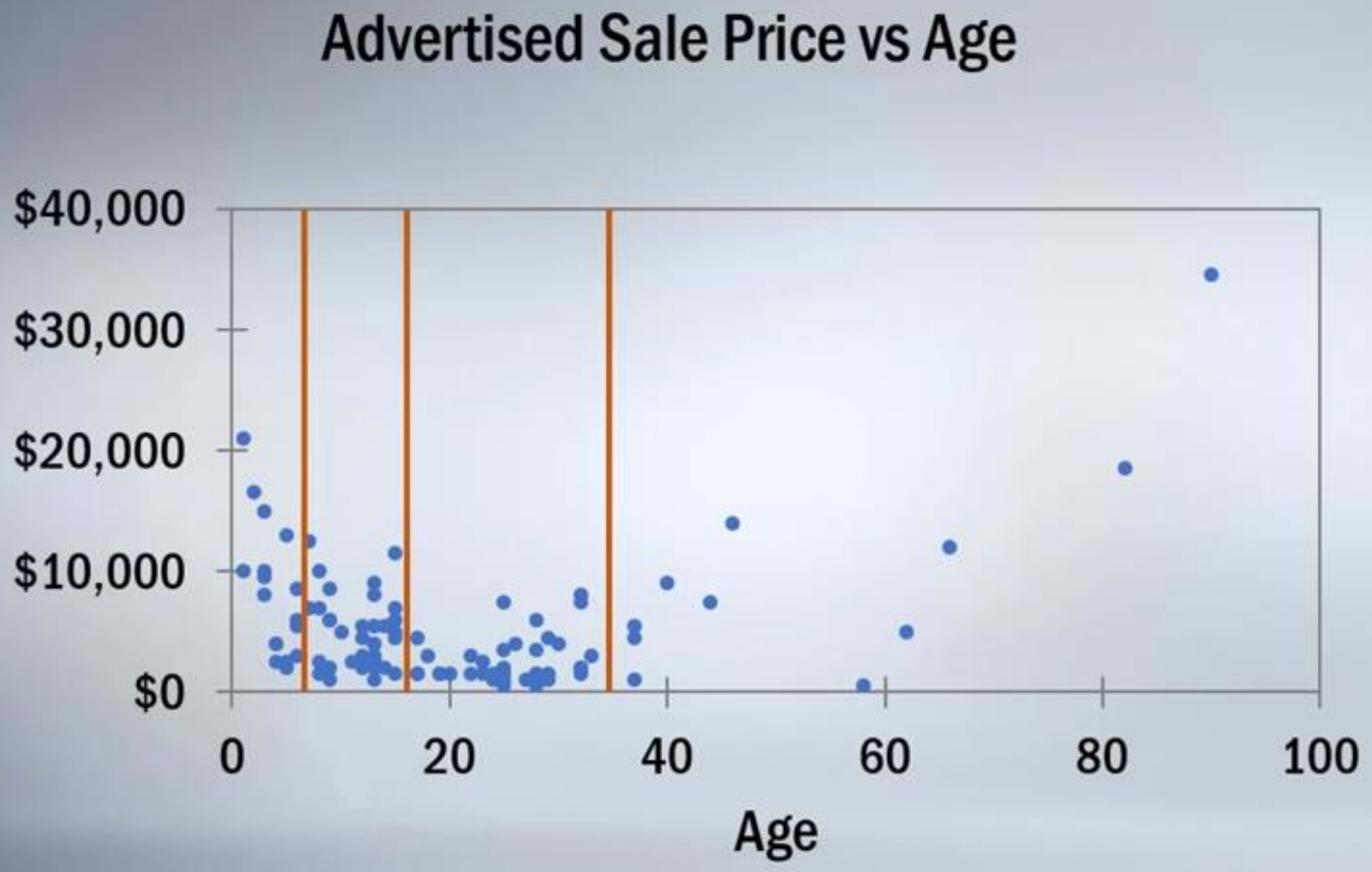
A car with a pink slip would command a sale price 15.6% higher than a car without a pink slip, on average, holding all other variables constant.

Categorical X variables

Multi-level categorical variables

Categorical X variables

Multi-level categorical variables



Categorical X variables

Multi-level categorical variables

AgeCat = 1 if age <= 5

= 2 if 5 < age <= 15

= 3 if 15 < age <= 35

= 4 if age > 35

Categorical X variables

Multi-level categorical variables

AgeCat = 1 if age <= 5
= 2 if 5 < age <= 15
= 3 if 15 < age <= 35
= 4 if age > 35

$$Ln(Price_i) = \beta_0 + \beta_1(Age_i) + \beta_2(Age_i)^2$$
$$+ \beta_3 Ln(Odometer_i) + \beta_4(Pink\ Slip_i) + \varepsilon_i$$

Categorical X variables

Multi-level categorical variables

AgeCat1 = 1 if age <= 5
= 0 otherwise

AgeCat2 = 1 if 5 < age <= 15
= 0 otherwise

AgeCat3 = 1 if 15 < age <= 35
= 0 otherwise

AgeCat4 = 1 if age > 35
= 0 otherwise

$$Ln(Price_i) = \beta_0 + \beta_1(AgeCat1_i) + \beta_2(AgeCat2_i) + \beta_3(AgeCat3_i) + \beta_4(AgeCat4_i) + \beta_5 Ln(Odometer_i) + \beta_6(Pink\ Slip_i) + \varepsilon_i$$

Categorical X variables

Multi-level categorical variables

$$
\begin{aligned}
Ln(Price_i) = \beta_0 &+ \beta_1(AgeCat1_i) + \beta_2(AgeCat2_i) \\
&+ \beta_3(AgeCat3_i) + \beta_4(AgeCat4_i) \\
&+ \beta_5 Ln(Odometer_i) + \beta_6(Pink\ Slip_i) + \varepsilon_i
\end{aligned}
$$

BUT!

$$AgeCat1_i = 1 - AgeCat2_i - AgeCat3_i - AgeCat4_i$$

Categorical X variables

Multi-level categorical variables

$$Ln(Price_i) = \beta_0 + \beta_1(AgeCat1_i) + \beta_2(AgeCat2_i) + \beta_3(AgeCat3_i) + \beta_4(AgeCat4_i) + \beta_5 Ln(Odometer_i) + \beta_6(Pink\ Slip_i) + \varepsilon_i$$

BUT!

$$AgeCat1_i = 1 - AgeCat2_i - AgeCat3_i - AgeCat4_i$$

Dummy variable TRAP

Categorical X variables

Model 6

$$Ln(Price_i) = \beta_0 + \beta_1(AgeCat2_i) + \beta_2(AgeCat3_i) + \beta_3(AgeCat4_i) + \beta_4 Ln(Odometer_i) + \beta_5(Pink\ Slip_i) + \varepsilon_i$$

Categorical X variables

Model 6

$$\widehat{Ln(Price_i)} = 8.948 - 0.129(AgeCat2_i) - 0.733(AgeCat3_i) + 0.474(AgeCat4_i) - 0.225Ln(Odometer_i) + 0.344(Pink\ Slip_i)$$

On average, holding all other variables constant, a car in age category 2 will command a price 12.9% lower than a car in age category 1.

Categorical X variables

Model 6

$$\widehat{Ln(Price_i)} = 8.948 - 0.129(AgeCat2_i) - 0.733(AgeCat3_i) + 0.474(AgeCat4_i) - 0.225Ln(Odometer_i) + 0.344(Pink\ Slip_i)$$

On average, holding all other variables constant, a car in age category 3 will command a price 73.3% lower than a car in age category 1.

Interaction terms

Interaction terms

Intuition

Build a model to explain the salary of all of Google's employees

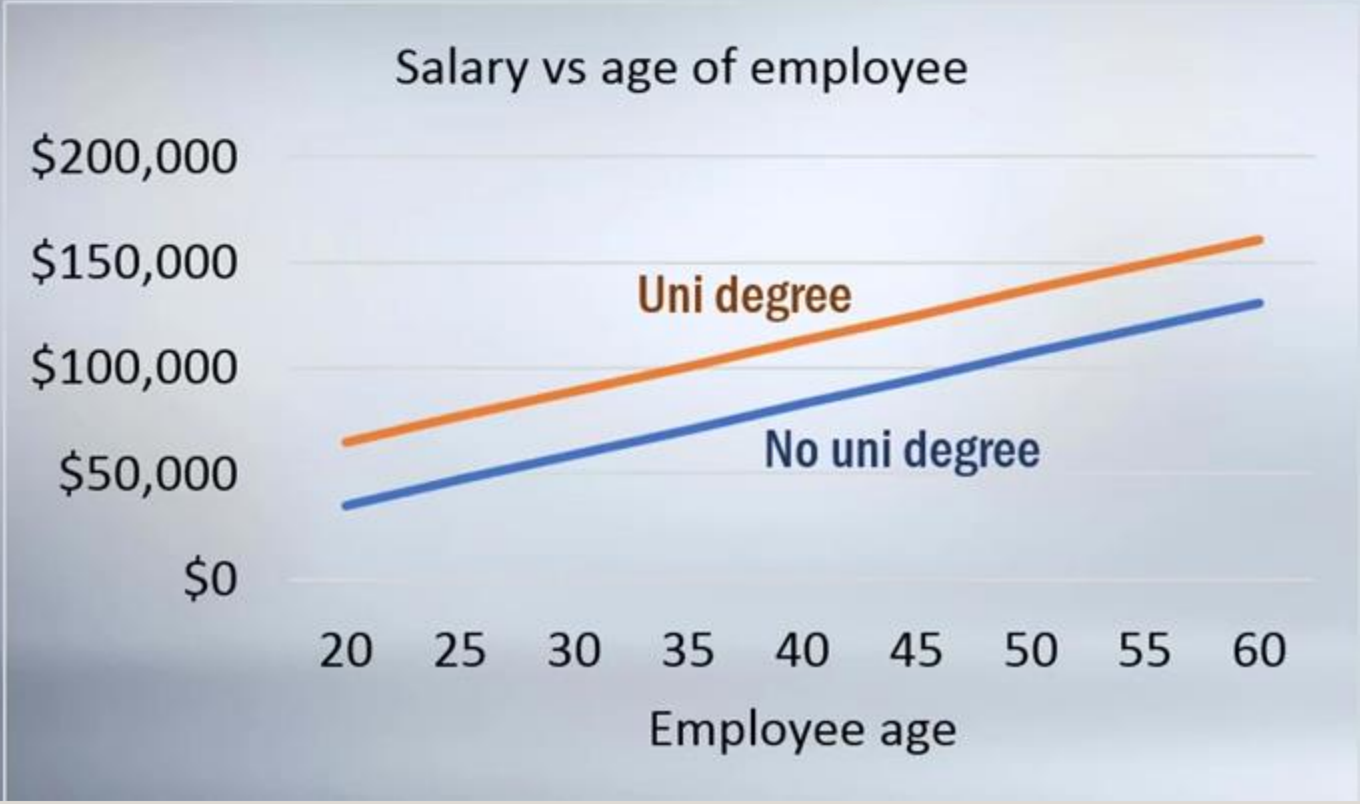
$$Salary_i = \beta_0 + \beta_1(Employee\ Age_i) + \beta_2(Uni\ degree_i) + \varepsilon_i$$

Interaction terms

Intuition

Build a model to explain the salary of all of Google's employees

$$Salary_i = \beta_0 + \beta_1(Employee\ Age_i) + \beta_2(Uni\ degree_i) + \varepsilon_i$$

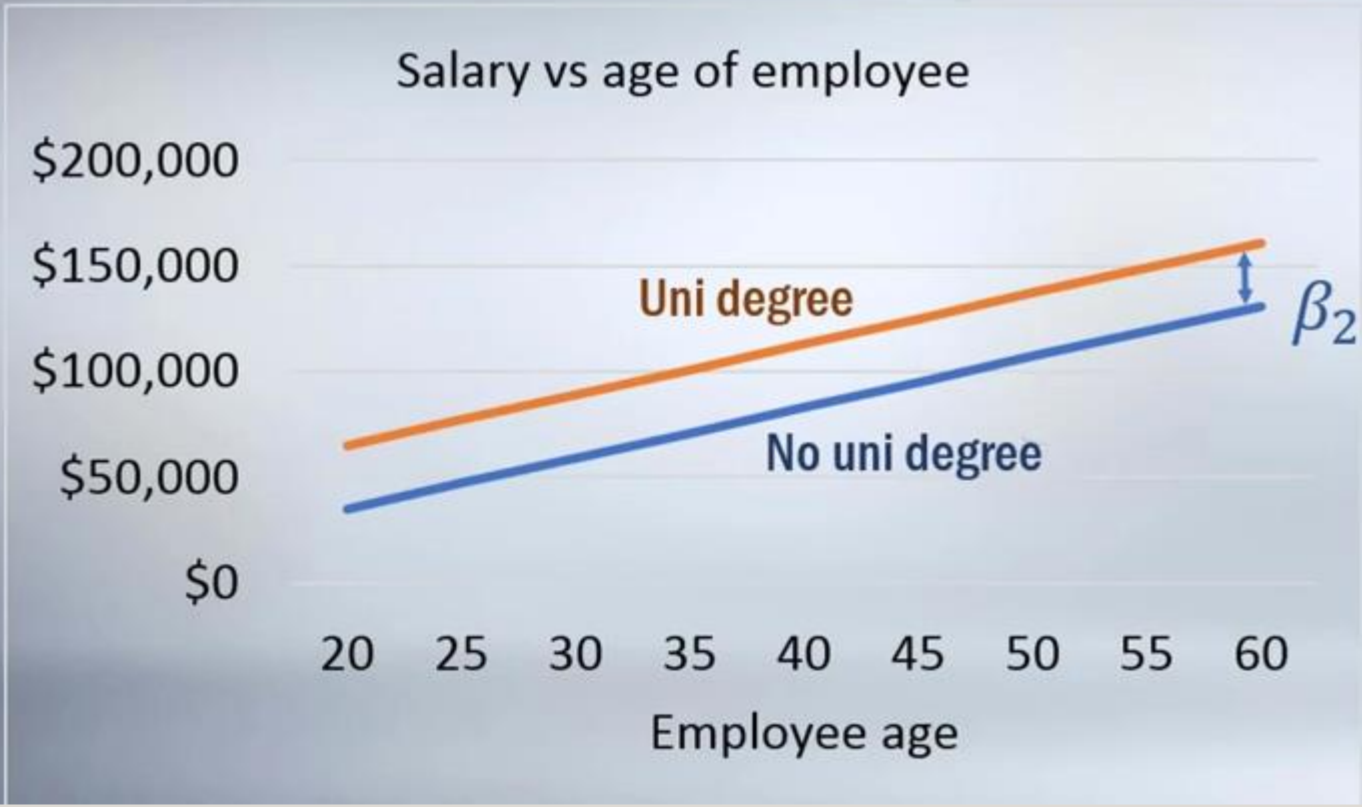


Interaction terms

Intuition

Build a model to explain the salary of all of Google's employees

$$Salary_i = \beta_0 + \beta_1(Employee\ Age_i) + \beta_2(Uni\ degree_i) + \varepsilon_i$$

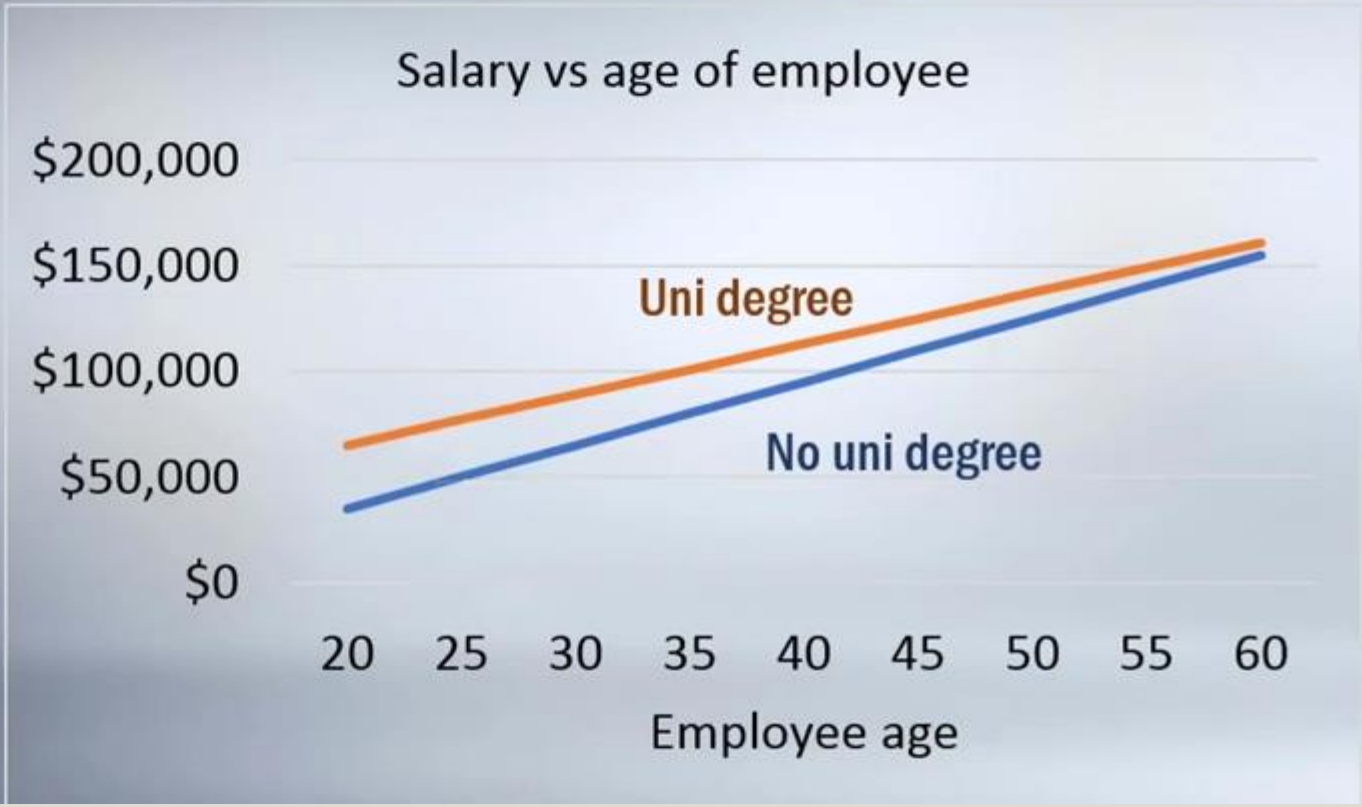


Interaction terms

Intuition

Build a model to explain the salary of all of Google's employees

$$Salary_i = \beta_0 + \beta_1(Employee\ Age_i) + \beta_2(Uni\ degree_i) + \varepsilon_i$$

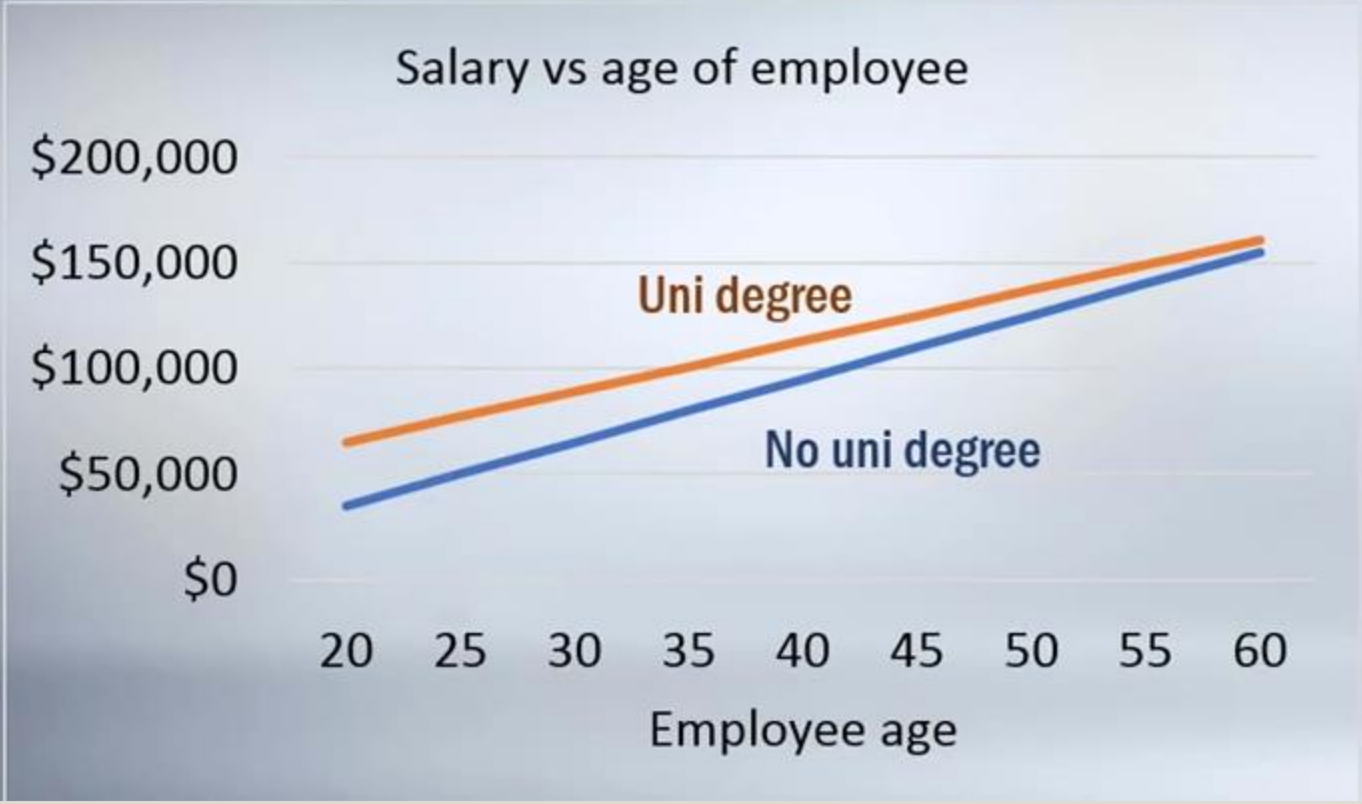


Interaction terms

Intuition

Build a model to explain the salary of all of Google's employees

$$Salary_i = \beta_0 + \beta_1(Employee\ Age_i) + \beta_2(Uni\ degree_i) + \beta_3(Employee\ Age_i) \times (Uni\ degree_i) + \varepsilon_i$$



Interaction terms

Required when:

X1 affects the relationship between X2 and Y

(eg. "Age of employee" affects the relationship between "Attainment of university degree" and "Salary")

Common misconception

"An interaction term is required when X1 and X2 are correlated"

Interaction terms

Model 7

$$Ln(Price_i) = \beta_0 + \beta_1(AgeCat2_i) + \beta_2(AgeCat3_i) + \beta_3(AgeCat4_i) + \beta_4 Ln(Odometer) + \beta_5(Pink\ Slip_i) + \beta_6(Pink\ Slip_i) \times (AgeCat4_i)$$

DV: Ln(Price)	Coef	SE	t	P-value
Intercept	9.125	0.274	33.28	0.0000
AgeCat2	-0.181	0.238	-0.76	0.4495
AgeCat3	-0.800	0.252	-3.18	0.0020
AgeCat4	-0.390	0.424	-0.92	0.3595
Ln(Odometer)	-0.209	0.059	-3.53	0.0007
Pink slip	0.123	0.182	0.68	0.5005
Pink slip X AgeCat4	1.371	0.453	3.02	0.0032

Non significant?

Interaction terms

Model 7

$$\widehat{Ln(Price_i)} = 9.125 - 0.181(AgeCat2_i) - 0.800(AgeCat3_i)$$
$$-0.390 \times (AgeCat4_i) - 0.209Ln(Odometer) + 0.123(Pink\ Slip_i)$$
$$+1.371(Pink\ Slip_i) \times (AgeCat4_i)$$

Interpret the coefficient of Pink slip:

Interaction terms

Model 7

$$\widehat{Ln(Price_i)} = 9.125 - 0.181(AgeCat2_i) - 0.800(AgeCat3_i)$$
$$-0.390 \times (AgeCat4_i) - 0.209Ln(Odometer) + 0.123(Pink\ Slip_i)$$
$$+1.371(Pink\ Slip_i) \times (AgeCat4_i)$$

Interpret the coefficient of Pink slip:

For models less than (or equal to) 35 years old, attaining a pink slip increases the price by an average of 12.3% , holding all else constant...

Interaction terms

Model 7

$$\widehat{Ln(Price_i)} = 9.125 - 0.181(AgeCat2_i) - 0.800(AgeCat3_i)$$
$$-0.390 \times (AgeCat4_i) - 0.209Ln(Odometer) + 0.123(Pink\ Slip_i)$$
$$+1.371(Pink\ Slip_i) \times (AgeCat4_i)$$

Interpret the coefficient of Pink slip:

... BUT for models older than 35 years, attaining a pink slip increases the price by an average of 149.4% , holding all else constant.

Interaction terms

REVISION QUESTION

Using model 7, find the expected sale price of my 1974 Datsun 120Y Coupe with 290,000km on the odometer and a road worthy certificate.

Interaction terms

REVISION QUESTION

Using model 7, find the expected sale price of my 1974 Datsun 120Y Coupe with 290,000km on the odometer and a road worthy certificate.

Interaction terms

REVISION QUESTION

Using model 7, find the expected sale price of my 1974 Datsun 120Y Coupe with 290,000km on the odometer and a road worthy certificate.

DV: Ln(Price)	*Coef*	*SE*	*t*	*P-value*
Intercept	9.125	0.274	33.28	0.0000
AgeCat2	-0.181	0.238	-0.76	0.4495
AgeCat3	-0.800	0.252	-3.18	0.0020
AgeCat4	-0.390	0.424	-0.92	0.3595
Ln(Odometer)	-0.209	0.059	-3.53	0.0007
Pink slip	0.123	0.182	0.68	0.5005
Pink slip X AgeCat4	1.371	0.453	3.02	0.0032

Interaction terms

$$\widehat{Ln(Price_i)} = 9.125 - 0.181(AgeCat2_i) - 0.800(AgeCat3_i)$$
$$-0.390\,(AgeCat4_i) - 0.209Ln(Odometer) + 0.123(Pink\ Slip_i)$$
$$+1.371(Pink\ Slip_i) \times (AgeCat4_i)$$

Interaction terms

$$Ln(\widehat{Price_i}) = 9.125 - 0.181(AgeCat2_i) - 0.800(AgeCat3_i)$$
$$-0.390\,(AgeCat4_i) - 0.209Ln(Odometer) + 0.123(Pink\ Slip_i)$$
$$+1.371(Pink\ Slip_i) \times (AgeCat4_i)$$

$$Ln(\widehat{Price_i}) = 9.125 - 0.390 \ - 0.209Ln(290) + 0.123 + 1.371$$

Interaction terms

$$Ln(\widehat{Price_i}) = 9.125 - 0.181(AgeCat2_i) - 0.800(AgeCat3_i)$$
$$-0.390\,(AgeCat4_i) - 0.209Ln(Odometer) + 0.123(Pink\ Slip_i)$$
$$+1.371(Pink\ Slip_i) \times (AgeCat4_i)$$

$$Ln(\widehat{Price_i}) = 9.125 - 0.390 - 0.209Ln(290) + 0.123 + 1.371$$

$$Ln(\widehat{Price_i}) = 9.044$$

Interaction terms

$$\widehat{Ln(Price_i)} = 9.125 - 0.181(AgeCat2_i) - 0.800(AgeCat3_i)$$
$$-0.390\,(AgeCat4_i) - 0.209Ln(Odometer) + 0.123(Pink\ Slip_i)$$
$$+1.371(Pink\ Slip_i) \times (AgeCat4_i)$$

$$\widehat{Ln(Price_i)} = 9.125 - 0.390 - 0.209Ln(290) + 0.123 + 1.371$$

$$\widehat{Ln(Price_i)} = 9.044$$

$$\widehat{Price_i} = e^{9.044}$$

$$\widehat{Price_i} = \$8,468$$

THANK YOU!