LOGS IN ANALYTICS

Matt Risley
University of Cincinnati, OBAIS

What is a log transformation?

- The logarithm is the inverse of the exponential function.

Let $f(x)$ be defined as the following exponential function.

$$f(x) = 10^x \tag{1}$$

Let $g(x)$ be defined as its inverse function.

$$g(x) = f^{-1}(x) \tag{2}$$

The inverse function $g(x)$ can therefore be written as follows.

$$g(x) = \log_{10} f(x) \tag{3}$$

The inverse function can be simplified to:

$$g(x) = x \tag{4}$$

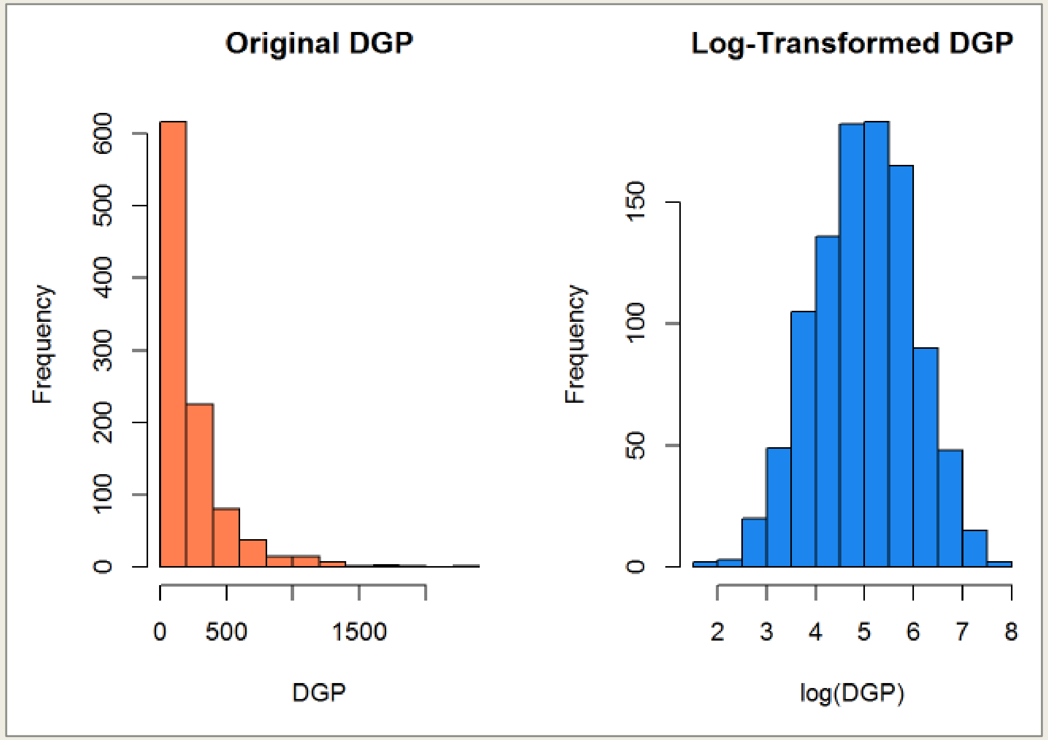
The logarithm (with its appropriate base) returns the exponent from the exponential function.

x	$f(x)$	$f^{-1}(x)$
0	1	0
1	10	1
2	100	2
3	1000	3
...	...	...
n	10^n	n

Why is it used in analytics?

- Most modeling techniques and statistical testing rely on the normal distribution.
- A normal distribution often does not describe observed data.
- The logarithmic transformation normalizes some skewed data.
- http://people.duke.edu/~rnau/411log.htm

$$\mathrm{DGP} \sim \mathrm{Lognormal}(5, 1)$$



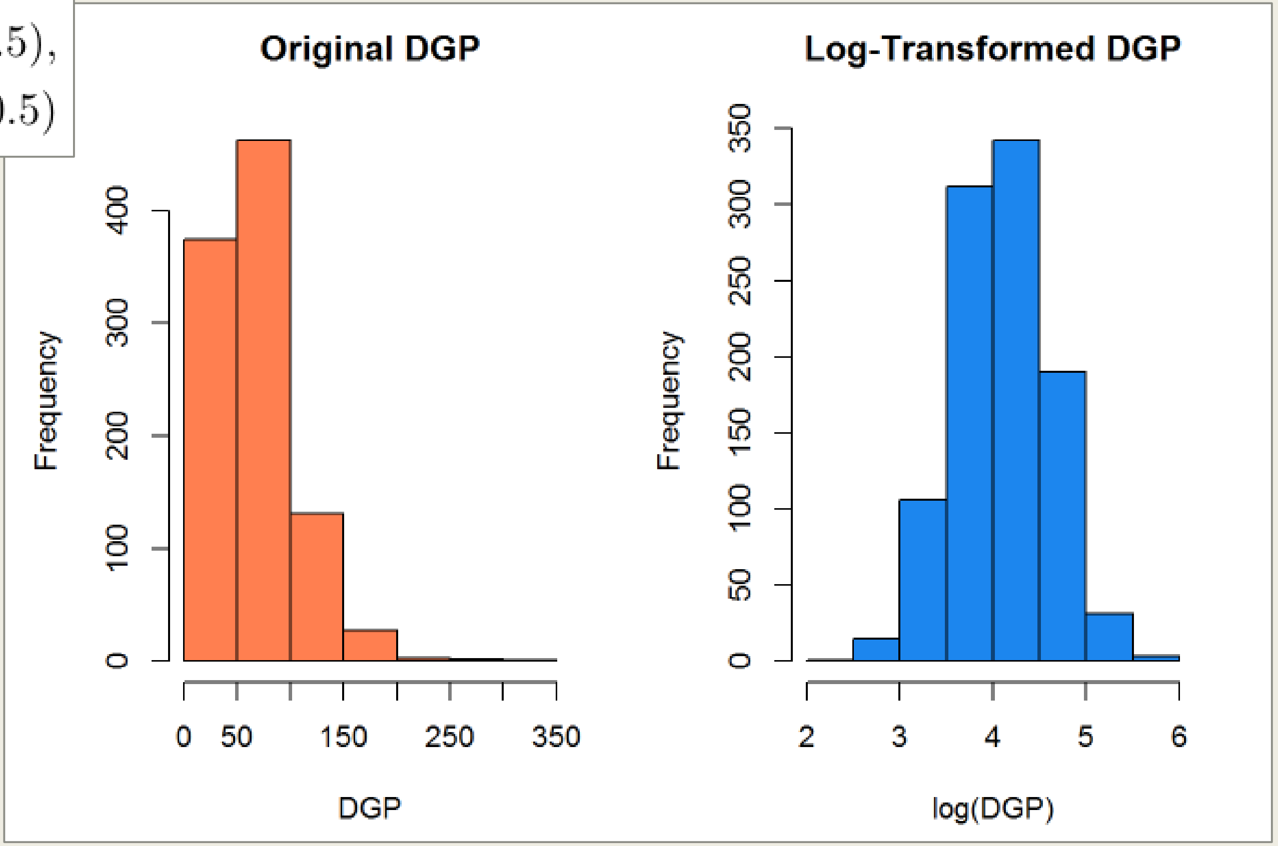

```r
# create data
dgp <- rlnorm(1000, mean=5, sd=1)
l_dgp <- log(dgp)

# plot data
par(mfrow=c(1, 2))

hist(dgp,
     main="Original DGP",
     col='coral',
     xlab="DGP")

hist(l_dgp,
     main="Log-Transformed DGP",
     col='dodgerblue2',
     xlab="log(DGP)")
```

Why is it used in analytics?

- Transforms multiplicative relationships into additive ones.
- Most economic and financial data include impacts from multipliers.
  - *i.e., the multiplier effect*

$$\log xy = \log x + \log y$$

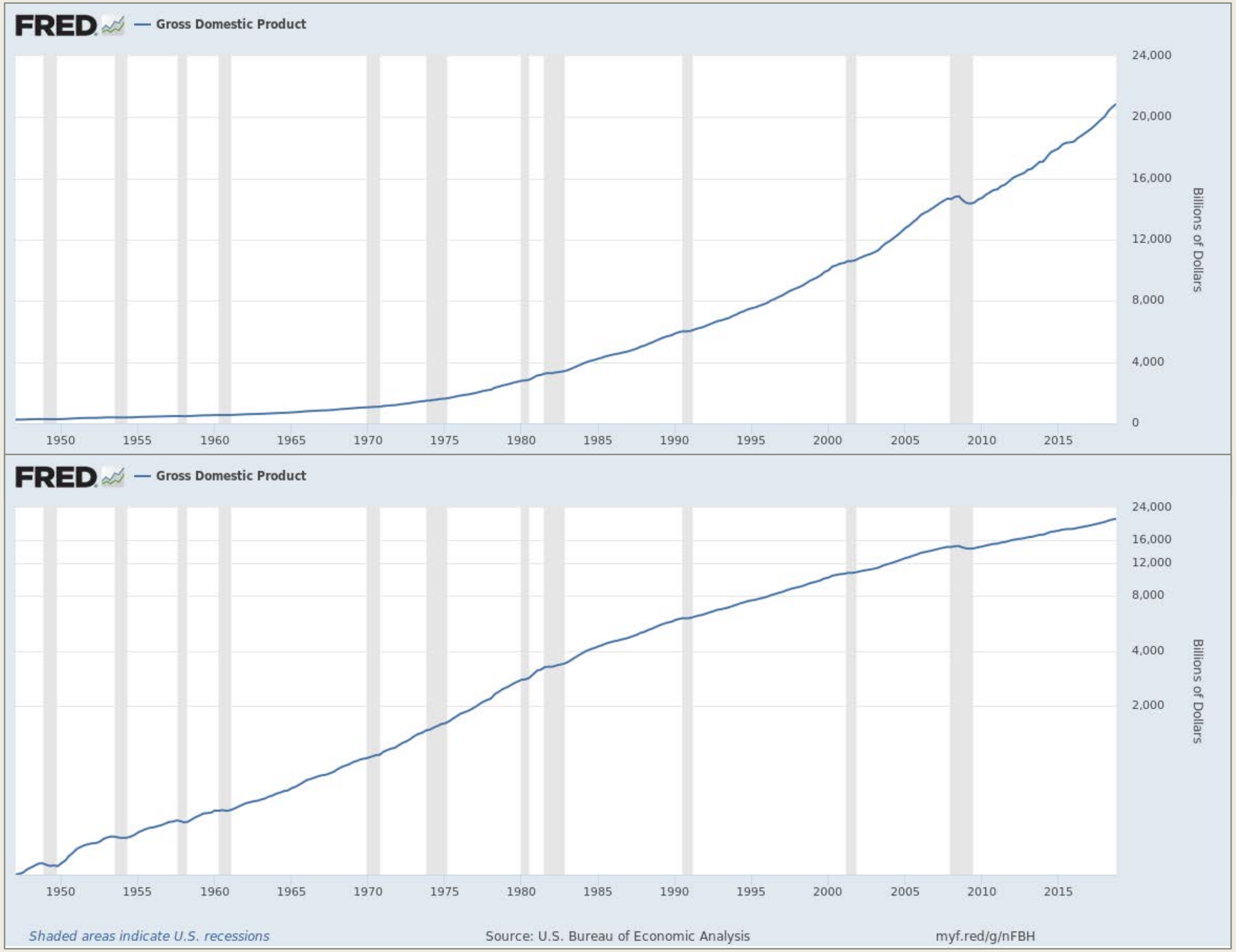
$$\text{DGP} = xy$$
$$\text{where } x \sim N(3, 0.5),$$
$$y \sim \text{Lognormal}(3, 0.5)$$




```r
# create data
x <- rnorm(1000, mean=3, sd=0.5)
y <- rlnorm(1000, mean=3, sd=0.5)
dgp <- x*y
l_dgp <- log(dgp)

# plot data
par(mfrow=c(1, 2))

hist(dgp,
     main="Original DGP",
     col='coral',
     xlab="DGP")

hist(l_dgp,
     main="Log-Transformed DGP",
     col='dodgerblue2',
     xlab="log(DGP)")
```

Why is it used in analytics?

- Transforms exponential growth trends into linear trends.
- Exponential growth = compounding growth.
- Most financial and economic series have compounding as part of their data generating processes.
  - *inflation, interest, expansion cycles*
- Albert Einstein supposedly once said that compounding [interest] is the most powerful force in the universe.



Why is it used in analytics?

- Data with these properties fit linear models well under the log transformation.

Why is the natural logarithm the default?

- natural logarithm uses e (Euler's number) as the base.

$$\log_e x = \ln(x) = \log x$$
$$e = \lim_{n\to\infty}(1+1/n)^n$$
$$e = 2.71828\ldots$$

n	$(1+1/n)^n$	interpretation
1	2	100% interest on \$1.00 compounded once
10	2.59	100% interest on \$1.00 compounded ten times
100	2.70	100% interest on \$1.00 compounded one hundred times
1000	2.717	100% interest on \$1.00 compounded one thousand times
…	…	…
∞	e	100% interest on \$1.00 compounded continuously

```r
fn <- function(x) {(1+1/x)^x}

fn(1); fn(10); fn(100); fn(1000);
```

```
## [1] 2
```

```
## [1] 2.593742
```

```
## [1] 2.704814
```

```
## [1] 2.716924
```

Why is the natural logarithm the default?

- Consistent estimator of growth, but not unbiased.
  - *Bias is a function of % change (Appendix 1)*
- Aids in interpretation.

$$\Delta \log x \approx \% \text{ change in } x$$

$$\%\Delta\ x = \frac{\Delta\ x}{x}$$

x_1	x_2	$\%\Delta\ x$	$\Delta \log x$
1	0.50	-50%	-69.3%
1	0.75	-25%	-28.8%
1	0.90	-10%	-10.5%
1	0.95	-5%	-5.1%
1	0.99	-1%	-1.0%
1	1.00	0%	0%
1	1.01	1%	1.0%
1	1.05	5%	5.1%
1	1.10	-0%	10.5%
1	1.25	25%	28.8%
1	1.50	50%	69.3%

```r
fn <- function(x2, x1=1) {
  grwth <- (x2-x1)/x1
  l_grwth <- log(x2/x1)

  msg1 <- paste0("Growth is ",
                 round(grwth*100, 3), "%")

  msg2 <- paste0("Logarithmic growth is ",
                 round(l_grwth*100, 3), "%")

  print(msg1); print(msg2)
}

fn(1.01); fn(800.5, 837.7); fn(3)
```

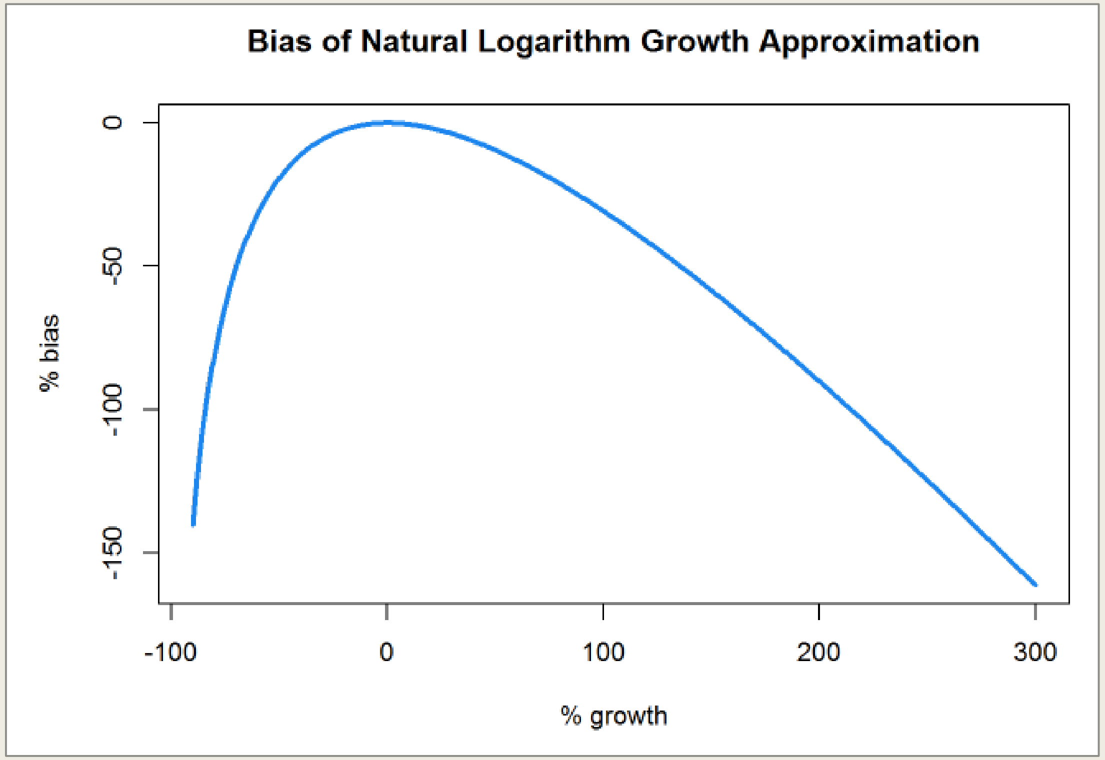
```
## [1] "Growth is 1%"
## [1] "Logarithmic growth is 0.995%"
```

```
## [1] "Growth is -4.441%"
## [1] "Logarithmic growth is -4.542%"
```

```
## [1] "Growth is 200%"
## [1] "Logarithmic growth is 109.861%"
```

Appendix 1

- As an estimator of growth, the natural logarithm has bias that is a function of the growth.




```
fn <- function(x2, x1=1) {
  grwth <- (x2-x1)/x1
  l_grwth <- log(x2/x1)

  bias <- l_grwth-grwth

  out <- cbind(grwth, bias)

  return(out)
}

x <- seq(from=1/10, to=4, length=1000)

gimme <- fn(x)*100

plot(gimme,
     type='l',
     main="Bias of Natural Logarithm Growth Approximation",
     xlab="% growth",
     ylab="% bias",
     col='dodgerblue2',
     lwd=3)
```