Analysis of oil

2019-04-10

Data Prep

1. Dropped records with missing oil values at t1 and t2. Call the resulting data set `df`.
2. Separated `df` into two subsets:
   - `df_tiny`: oil < 1000 at t1 or t2.
   - `df_main`: oil >= 1000 at t1 and t2
3. Created long-format version:
   - `df_long`: long-format version of the full set `df`
   - `df_tiny_long`: long-format version of the subset `df_tiny`
   - `df_main_long`: long-format version of the subset `df_main`
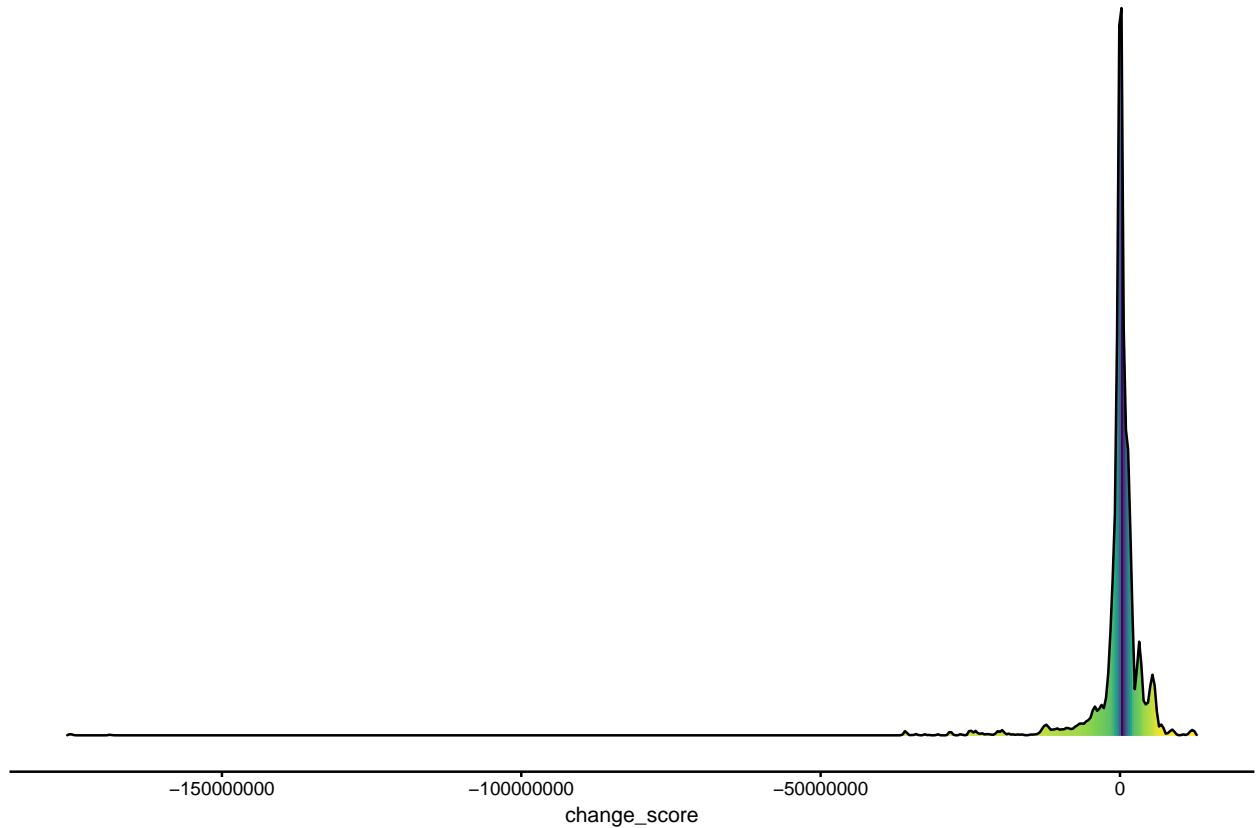
Analyze the subset `df_main`

Table 1: Sample Summary Statistics of oil

time	n_tribes	n	mean	median	std	skewness	kurtosis	SEM
t1	58	10057	1590954	253967	5998575	18.277535	492.721417	59815.52
t2	58	10057	1321923	405714	1921539	2.457108	7.561256	19160.86

Q1. Is there a difference between t1 and t2?

Descriptive Analysis

oil

change_score

NAs were dropped

Empirical Distribution of change_score

mean: −269030.74   median: 124963

Is change_score normally distributed?

Detrended Q–Q normal plot (n = 10057)

n = 10057

n = 10057
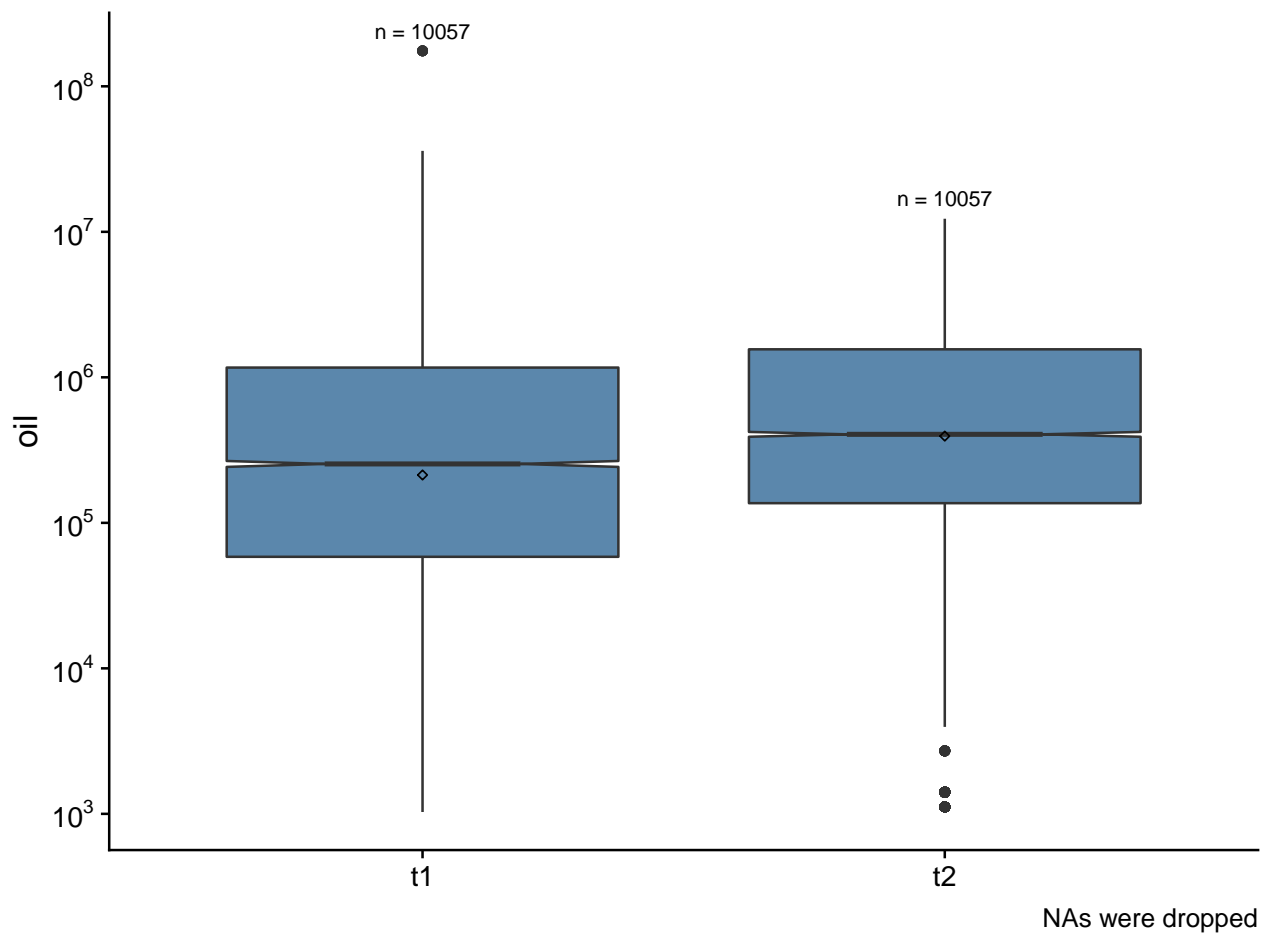
NAs were dropped
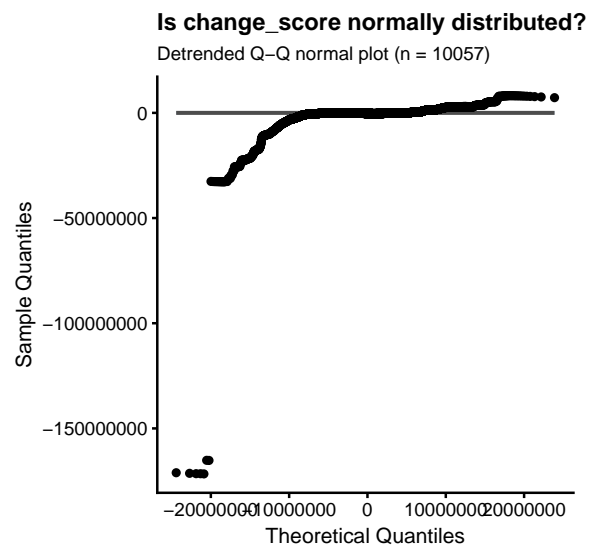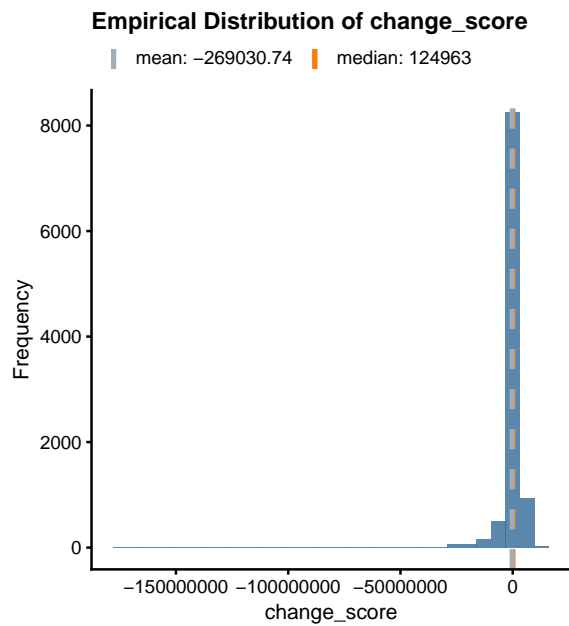
Statistical Analysis

```
1-way Repeated Measure ANOVA Output:

Error: tribe
          Df             Sum Sq          Mean Sq F value Pr(>F)
Residuals 57 29542191066690640 518284053801590

Error: tribe:time
          Df            Sum Sq          Mean Sq F value Pr(>F)
time       1    363950444378438 363950444378438   1.158  0.286
Residuals 57 17910727365974884 314223287122366

Error: Within
             Df             Sum Sq          Mean Sq F value Pr(>F)
Residuals 19998 351520995867542016 17577807574135

Linear Mixed Model Output:
            numDF denDF  F-value p-value
(Intercept)     1 20055 19.98251  <.0001
time            1 20055 19.75646  <.0001
```

Q2. How are t1 and t2 related?

oil

n = 58

$y = 3.77 + 0.315\,x,\ R^2 = 0.1$

$p-\text{value} = 0.014$

Mean value at t2 (y-axis), Mean value at t1 (x-axis)

Each dot is a tribe

oil

n = 58

$y = 1.97 + 0.661\,x,\ R^2 = 0.3$

$p-\text{value} < 0.001$

Median value at t2 (y-axis), Median value at t1 (x-axis)

Each dot is a tribe