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`
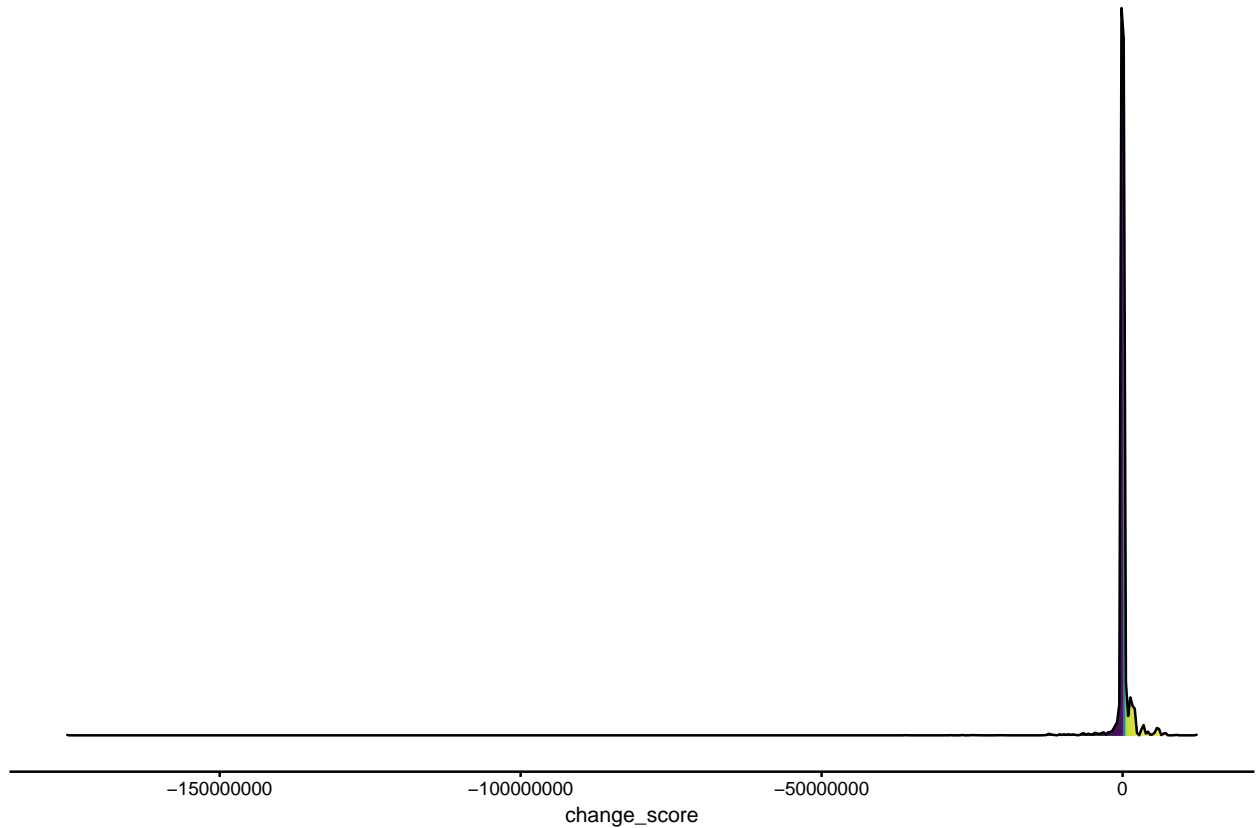
Analyze the full set `df_tiny`

Table 1: Sample Summary Statistics of oil

time	n_tribes	n	mean	median	std	skewness	kurtosis	SEM
t1	179	49914	213568.6	0	2662339.2	52.680938	3383.4668	11916.596
t2	179	49914	255272.3	0	891295.9	5.415651	36.6874	3989.429

Q1. Is there a difference between t1 and t2?

Descriptive Analysis

oil

change_score

NAs were dropped

Empirical Distribution of change_score

mean: 41703.65   median: 0

Is change_score normally distributed?
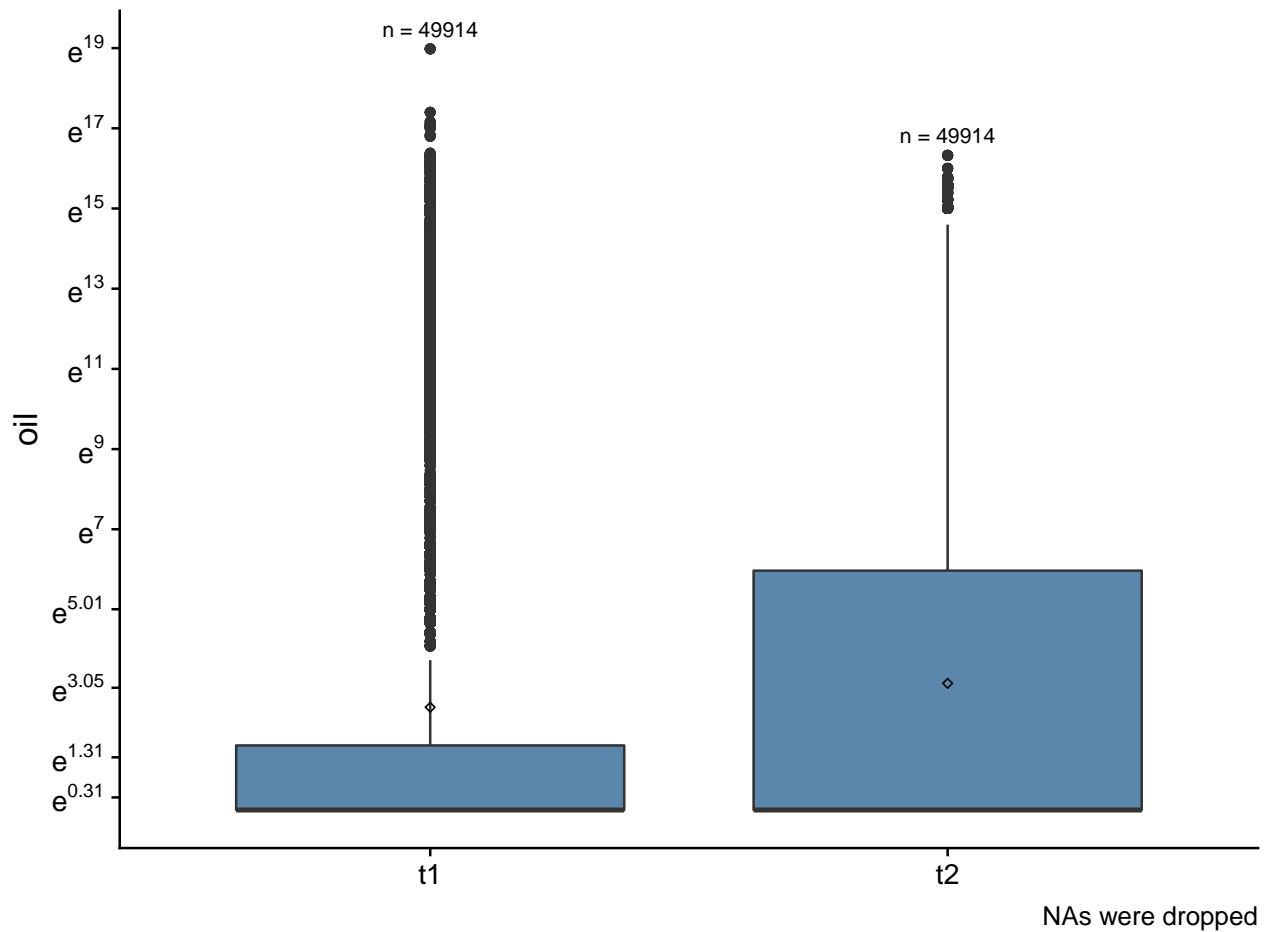
Detrended Q–Q normal plot (n = 49914)

NAs were dropped
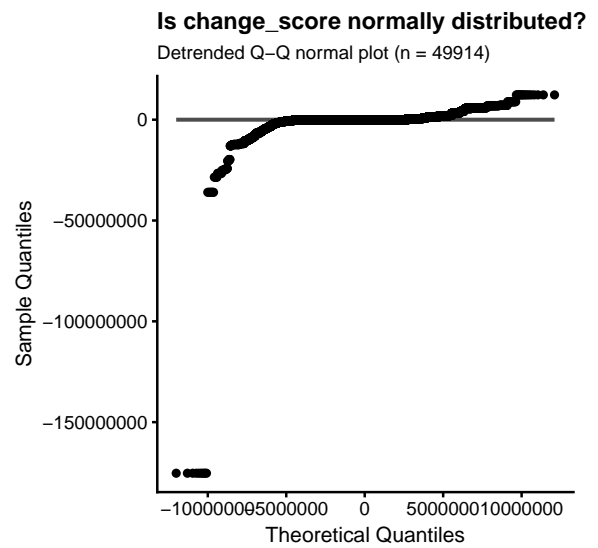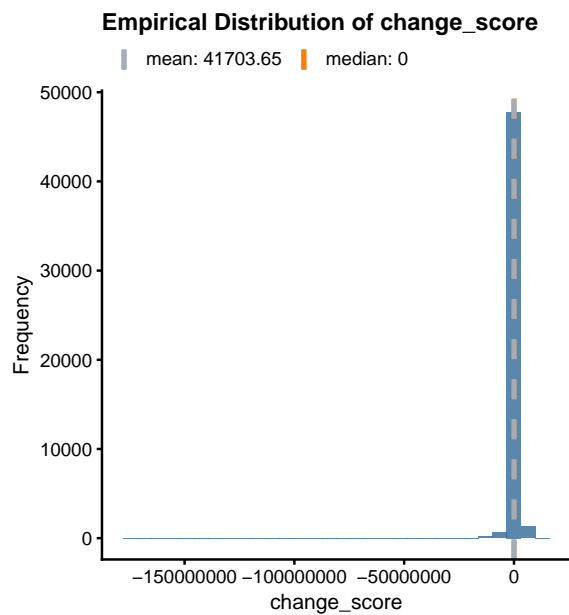
Statistical Analysis

1-way Repeated Measure ANOVA Output:

```
Error: tribe
          Df             Sum Sq          Mean Sq F value Pr(>F)
Residuals 178 14925793296631376 83852771329390

Error: tribe:time
          Df             Sum Sq          Mean Sq F value Pr(>F)
time        1    43405073008370 43405073008370   0.542  0.463
Residuals 178 14266730602669750 80150171925111

Error: Within
            Df              Sum Sq          Mean Sq F value Pr(>F)
Residuals 99470 364244627779095808 3661854104545
```

Linear Mixed Model Output:
```
            numDF denDF  F-value p-value
(Intercept)     1 99648 21.12690  <.0001
time            1 99648 11.42906  0.0007
```

Q2. How are t1 and t2 related?

oil

n = 179

$y = 2.25 + 0.269\,x,\ \ R^2 = 0.078$

$p-\text{value} < 0.001$

Mean value at t2

Mean value at t1

Each dot is a tribe

oil

n = 179

$y = 2.36 - 0.182\,x,\ \ R^2 = 0.011$

$p-\text{value} = 0.159$

Median value at t2

Median value at t1

Each dot is a tribe