

Can I use an interaction term without a main effect?

Create two populations that start at the same place, but have different responses to `time`.

```
popA <- c(rnorm(200, 0),
          rnorm(200, 5),
          rnorm(200, 10))

popB <- c(rnorm(200, 0),
          rnorm(200, 10),
          rnorm(200, 20))

fake_data <- data.frame(
  population = rep(c('A', 'B'), each = 600),
  response = c(popA, popB),
  time = rep(
    rep(c(0, 1, 2), each = 200),
    times = 2
  )
)
```

Full model

population term is not significant.

```
m_full <- lm(response ~ population*time,
             data = fake_data)
summary(m_full)
```

```
##
## Call:
## lm(formula = response ~ population * time, data = fake_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2324 -0.7023 -0.0501  0.6603  3.7427
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.04969    0.06348  -0.783   0.434
## populationB     0.13087    0.08977   1.458   0.145
## time           4.99291    0.04917 101.549 <2e-16 ***
## populationB:time 4.95916    0.06953  71.320 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.9834 on 1196 degrees of freedom
## Multiple R-squared:  0.9802, Adjusted R-squared:  0.9802
## F-statistic: 1.977e+04 on 3 and 1196 DF,  p-value: < 2.2e-16
```

Model without the main effect of population

In this instance, similar estimates (possibly not the case with more-complicated models?).

```
m_nomain <- lm(response ~ time + population:time,
               data = fake_data)
summary(m_nomain)
```

```
##
## Call:
## lm(formula = response ~ time + population:time, data = fake_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2978 -0.7069 -0.0623  0.6725  3.7165
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.01575    0.04490   0.351   0.726
## time           4.95365    0.04116 120.363 <2e-16 ***
## time:populationB 5.03768    0.04400 114.499 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9838 on 1197 degrees of freedom
## Multiple R-squared:  0.9802, Adjusted R-squared:  0.9802
## F-statistic: 2.963e+04 on 2 and 1197 DF,  p-value: < 2.2e-16
```

```
anova(m_full, m_nomain, test = 'LRT')
```

```
## Analysis of Variance Table
##
## Model 1: response ~ population * time
## Model 2: response ~ time + population:time
##   Res.Df    RSS Df Sum of Sq Pr(>Chi)
## 1     1196 1156.5
## 2     1197 1158.6 -1    -2.0551   0.1449
```