

Workshop 10 Solutions

Veerasak

null

We'll begin by loading some packages.

```
library(MASS)
library(plyr)
```

Interaction terms in regression

```
# Building up the familiar birthwt data...

# Rename the columns to have more descriptive names
colnames(birthwt) <- c("birthwt.below.2500", "mother.age", "mother.weight",
  "race", "mother.smokes", "previous.prem.labor", "hypertension", "uterine.irr",
  "physician.visits", "birthwt.grams")

# Transform variables to factors with descriptive levels
birthwt <- transform(birthwt,
  race = as.factor(mapvalues(race, c(1, 2, 3),
    c("white", "black", "other"))),
  mother.smokes = as.factor(mapvalues(mother.smokes,
    c(0,1), c("no", "yes"))),
  hypertension = as.factor(mapvalues(hypertension,
    c(0,1), c("no", "yes"))),
  uterine.irr = as.factor(mapvalues(uterine.irr,
    c(0,1), c("no", "yes")))
)
```

(a) Run a linear regression to better understand how birthweight varies with the mother's age and smoking status (do not include interaction terms).

```
# Run regression model
birthwt.lm <- lm(birthwt.grams ~ mother.age + mother.smokes, data = birthwt)
# Output coefficients table
summary(birthwt.lm)
```

```
##
## Call:
## lm(formula = birthwt.grams ~ mother.age + mother.smokes, data = birthwt)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2119.98  -442.66   52.92   532.38  1690.74
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2791.224    240.950   11.584  <2e-16 ***
## mother.age       11.290     9.881    1.143   0.255
## mother.smokesyes -278.356    106.987  -2.602   0.010 *
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 717.2 on 186 degrees of freedom
## Multiple R-squared:  0.04299,    Adjusted R-squared:  0.0327
## F-statistic: 4.177 on 2 and 186 DF,  p-value: 0.0168
```

(b) What is the coefficient of mother.age in your regression? How do you interpret this coefficient?

```
coef(birthwt.lm)["mother.age"]
```

```
## mother.age
##      11.28961
```

```
age.coef <- round(coef(birthwt.lm)["mother.age"], 1)
```

Note: This solution uses inline code chunks. The coefficient is 11.3. This means that among mothers with the same smoking status, each additional year of age is on average associated with a 11.3g increase in birthweight.

(c) How many coefficients are estimated for the mother's smoking status variable? How do you interpret these coefficients?

```
coef(birthwt.lm)["mother.smokesyes"]
```

```
## mother.smokesyes
##      -278.3561
```

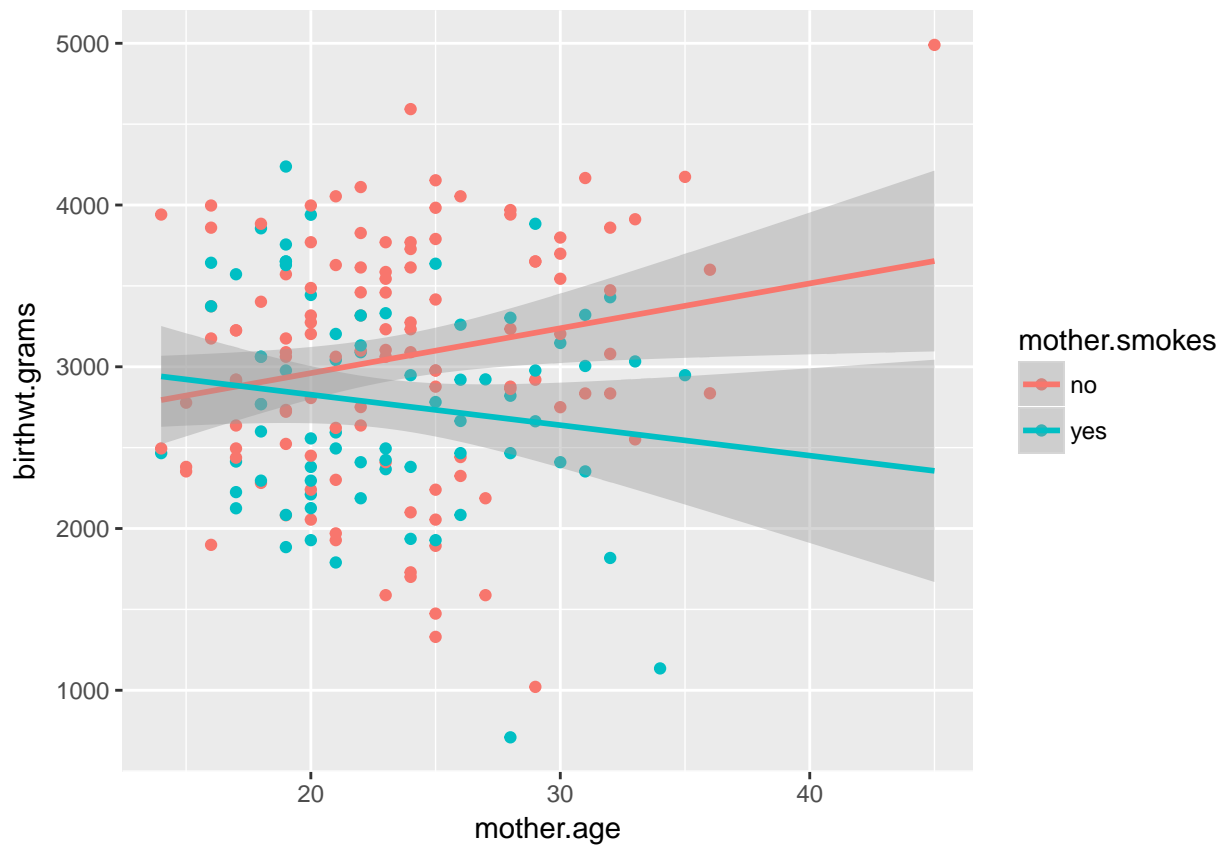
```
smoke.coef <- abs(round(coef(birthwt.lm)["mother.smokesyes"], 1))
```

Note: This solution uses inline code chunks. There is just one coefficient estimated. This coefficient gives us the average difference in birthweight between mothers that smoke and mother's that don't, in a model that adjusts for the effect of mother's age. That is, after we adjust for the effect of age, smoking leads to an average 278.4 decrease in birthweight.

(d) Using ggplot, construct a scatterplot with birthweight on the y-axis and mother's age on the x-axis. Color the points by mother's smoking status, and add smoking status-specific linear regression lines using the `stat_smooth` layer.

```
library(ggplot2)
```

```
# Note fullrange = TRUE is used here to extend the 'mother.smokes = yes' line beyond the maximum age (30)
ggplot(data = birthwt, x = mother.age, y = birthwt.grams, colour = mother.smokes) + stat_smooth(method =
```



```
birthwt.lm$model
```

```
##      birthwt.grams mother.age mother.smokes
## 85          2523         19         no
## 86          2551         33         no
## 87          2557         20         yes
## 88          2594         21         yes
## 89          2600         18         yes
## 91          2622         21         no
## 92          2637         22         no
## 93          2637         17         no
## 94          2663         29         yes
## 95          2665         26         yes
## 96          2722         19         no
## 97          2733         19         no
## 98          2751         22         no
## 99          2750         30         no
## 100         2769         18         yes
## 101         2769         18         yes
## 102         2778         15         no
## 103         2782         25         yes
## 104         2807         20         no
## 105         2821         28         yes
## 106         2835         32         no
## 107         2835         31         no
## 108         2836         36         no
## 109         2863         28         no
```

## 111	2877	25	no
## 112	2877	28	no
## 113	2906	17	yes
## 114	2920	29	no
## 115	2920	26	yes
## 116	2920	17	no
## 117	2920	17	no
## 118	2948	24	yes
## 119	2948	35	yes
## 120	2977	25	no
## 121	2977	25	no
## 123	2977	29	yes
## 124	2977	19	yes
## 125	2922	27	yes
## 126	3005	31	yes
## 127	3033	33	yes
## 128	3042	21	yes
## 129	3062	19	no
## 130	3062	23	no
## 131	3062	21	no
## 132	3062	18	yes
## 133	3062	18	yes
## 134	3080	32	no
## 135	3090	19	no
## 136	3090	24	no
## 137	3090	22	yes
## 138	3100	22	no
## 139	3104	23	no
## 140	3132	22	yes
## 141	3147	30	yes
## 142	3175	19	no
## 143	3175	16	no
## 144	3203	21	yes
## 145	3203	30	no
## 146	3203	20	no
## 147	3225	17	no
## 148	3225	17	no
## 149	3232	23	no
## 150	3232	24	no
## 151	3234	28	no
## 154	3260	26	yes
## 155	3274	20	no
## 156	3274	24	no
## 159	3303	28	yes
## 160	3317	20	no
## 161	3317	22	no
## 162	3317	22	yes
## 163	3321	31	yes
## 164	3331	23	yes
## 166	3374	16	no
## 167	3374	16	yes
## 168	3402	18	no
## 169	3416	25	no
## 170	3430	32	yes

## 172	3444	20	yes
## 173	3459	23	no
## 174	3460	22	no
## 175	3473	32	no
## 176	3544	30	no
## 177	3487	20	no
## 179	3544	23	no
## 180	3572	17	yes
## 181	3572	19	no
## 182	3586	23	no
## 183	3600	36	no
## 184	3614	22	no
## 185	3614	24	no
## 186	3629	21	no
## 187	3629	19	yes
## 188	3637	25	yes
## 189	3643	16	yes
## 190	3651	29	no
## 191	3651	29	no
## 192	3651	19	yes
## 193	3651	19	yes
## 195	3699	30	no
## 196	3728	24	no
## 197	3756	19	yes
## 199	3770	24	no
## 200	3770	23	no
## 201	3770	20	no
## 202	3790	25	no
## 203	3799	30	no
## 204	3827	22	no
## 205	3856	18	yes
## 206	3860	16	no
## 207	3860	32	no
## 208	3884	18	no
## 209	3884	29	yes
## 210	3912	33	no
## 211	3940	20	yes
## 212	3941	28	no
## 213	3941	14	no
## 214	3969	28	no
## 215	3983	25	no
## 216	3997	16	no
## 217	3997	20	no
## 218	4054	26	no
## 219	4054	21	no
## 220	4111	22	no
## 221	4153	25	no
## 222	4167	31	no
## 223	4174	35	no
## 224	4238	19	yes
## 225	4593	24	no
## 226	4990	45	no
## 4	709	28	yes
## 10	1021	29	no

## 11	1135	34	yes
## 13	1330	25	no
## 15	1474	25	no
## 16	1588	27	no
## 17	1588	23	no
## 18	1701	24	no
## 19	1729	24	no
## 20	1790	21	yes
## 22	1818	32	yes
## 23	1885	19	yes
## 24	1893	25	no
## 25	1899	16	no
## 26	1928	25	yes
## 27	1928	20	yes
## 28	1928	21	no
## 29	1936	24	yes
## 30	1970	21	no
## 31	2055	20	no
## 32	2055	25	no
## 33	2082	19	no
## 34	2084	19	yes
## 35	2084	26	yes
## 36	2100	24	no
## 37	2125	17	yes
## 40	2126	20	yes
## 42	2187	22	yes
## 43	2187	27	no
## 44	2211	20	yes
## 45	2225	17	yes
## 46	2240	25	no
## 47	2240	20	no
## 49	2282	18	no
## 50	2296	18	yes
## 51	2296	20	yes
## 52	2301	21	no
## 54	2325	26	no
## 56	2353	31	yes
## 57	2353	15	no
## 59	2367	23	yes
## 60	2381	20	yes
## 61	2381	24	yes
## 62	2381	15	no
## 63	2410	23	no
## 65	2410	30	yes
## 67	2410	22	yes
## 68	2414	17	yes
## 69	2424	23	yes
## 71	2438	17	no
## 75	2442	26	no
## 76	2450	20	no
## 77	2466	26	yes
## 78	2466	14	yes
## 79	2466	28	yes
## 81	2495	14	no

```
## 82      2495      23      yes
## 83      2495      17      no
## 84      2495      21      yes
```

Predict record # 20

```
pd <- predict(birthwt.lm, birthwt[20,], level = 0.99)
pd
```

```
##      105
## 2828.977
```

Actual

```
birthwt[20,]
```

```
##      birthwt.below.2500 mother.age mother.weight  race mother.smokes
## 105                0         28         120 white          yes
##      previous.prem.labor hypertension uterine.irr physician.visits
## 105                0             no          no             1
##      birthwt.grams
## 105          2821
```