> life <-read.csv(file.choose(), header=TRUE)

> # file name is "TermLife"

> life1 = subset(life, subset=FACE>0)

> attach(life1)

> lnFACE = log(FACE)

> lnINCOME = log(INCOME)

>

> model1 = lm(lnFACE ~ lnINCOME + EDUCATION + NUMHH + MARSTAT)

> summary(model1)

Call:

lm(formula = lnFACE ~ lnINCOME + EDUCATION + NUMHH + MARSTAT)

Residuals:

    Min      1Q  Median      3Q     Max

-5.7974 -0.8807  0.0958  0.8976  4.6550

Coefficients:

            Estimate Std. Error t value Pr(>|t|)

(Intercept)  2.55559    0.84663   3.019  0.00278 \*\*

lnINCOME     0.48307    0.07814   6.182 2.33e-09 \*\*\*

EDUCATION    0.20881    0.03889   5.370 1.70e-07 \*\*\*

NUMHH        0.27745    0.06874   4.036 7.07e-05 \*\*\*

MARSTAT      0.23523    0.22023   1.068  0.28643

---

Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.525 on 270 degrees of freedom

Multiple R-squared:  0.3453,

Adjusted R-squared:  0.3356

F-statistic:  35.6 on 4 and 270 DF,  p-value: < 2.2e-16

> # This model is incorrect since MARSTAT is not a quantitative variable, we don not want to interpret its coefficient here as 'for every 1 unit to MARSTAT added, lnFACE would increase by 0.23523' since that does not make sence for a categorical variable.

> model2 = lm(lnFACE ~ lnINCOME + EDUCATION + NUMHH + factor(MARSTAT))

> # this will tell R that MARSTAT is a categorical variable.

> # if these were word entries, you wouldn't have to do this since R would recognize them as categorical variables.

> summary(model2)

Call:

lm(formula = lnFACE ~ lnINCOME + EDUCATION + NUMHH + factor(MARSTAT))

Residuals:

    Min      1Q  Median      3Q     Max

-5.8875 -0.8505  0.1124  0.8468  4.5173

Coefficients:

                 Estimate Std. Error t value Pr(>|t|)

(Intercept)       2.83770    0.84882   3.343 0.000946 \*\*\*

lnINCOME          0.45151    0.07872   5.736 2.61e-08 \*\*\*

EDUCATION         0.20467    0.03862   5.299 2.42e-07 \*\*\*

NUMHH             0.24770    0.06940   3.569 0.000424 \*\*\*

factor(MARSTAT)1  0.55707    0.25929   2.148 0.032574 \*

factor(MARSTAT)2 -0.23234    0.53283  -0.436 0.663155

---

Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.513 on 269 degrees of freedom

Multiple R-squared:  0.358,

Adjusted R-squared:  0.3461

F-statistic:    30 on 5 and 269 DF,  p-value: < 2.2e-16

> # 0 is incorportated in the intercept. factor 1 slope is the increase in slope from being in the first category. 2 is also compared to 0. NOT 1 compared to 2. R will pick whatever is the baseline.

> # The category of 0 is incorporated as the baseline into the intercept

> # how we interpret factor(MARSTAT)1: If someone was married, they would purchase 0.55707 more log dollars of the lnFACE than someone who is 'other' (single), holding all other variables constant.

> # the p-value of 0.0325 means that at an alpha=0.05 level, this category of 1 is significantly different from the category of 0 in the amount of lnFACE they purchase.

>

> If you do not like the baseline category that R has chosen, you have to make vinary variables like below to choose them yourself.

> MAR0 = (1\*(MARSTAT==0))

> MAR1 = (1\*(MARSTAT==1))

> MAR2 = (1\*(MARSTAT==2))

> check = data.frame(MARSTAT, MAR0, MAR1, MAR2)

> fix(check)

> #do this to make sure they assigned the binary variables correctly

> # give nice little table with tallys and assignments.

>

> model3 = lm(lnFACE ~ lnINCOME + EDUCATION + NUMHH + MAR0 + MAR1)

> # if we choose this, mar2 will be the baseline. Omit which one you want as baseline.

> summary(model3)

Call:

lm(formula = lnFACE ~ lnINCOME + EDUCATION + NUMHH + MAR0 + MAR1)

Residuals:

    Min      1Q  Median      3Q     Max

-5.8875 -0.8505  0.1124  0.8468  4.5173

Coefficients:

            Estimate Std. Error t value Pr(>|t|)

(Intercept)  2.60536    0.95218   2.736 0.006629 \*\*

lnINCOME     0.45151    0.07872   5.736 2.61e-08 \*\*\*

EDUCATION    0.20467    0.03862   5.299 2.42e-07 \*\*\*

NUMHH        0.24770    0.06940   3.569 0.000424 \*\*\*

MAR0         0.23234    0.53283   0.436 0.663155

MAR1         0.78941    0.49532   1.594 0.112169

---

Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.513 on 269 degrees of freedom

Multiple R-squared:  0.358,

Adjusted R-squared:  0.3461

F-statistic:    30 on 5 and 269 DF,  p-value: < 2.2e-16

> # intercept will change every time

> # These slopes are the comparison from these categories to a baseline of category 2.

> model4 = lm(lnFACE ~ lnINCOME + EDUCATION + NUMHH + MAR0 + MAR1 + MAR2)

> summary(model4)

Call:

lm(formula = lnFACE ~ lnINCOME + EDUCATION + NUMHH + MAR0 + MAR1 +

    MAR2)

Residuals:

    Min      1Q  Median      3Q     Max

-5.8875 -0.8505  0.1124  0.8468  4.5173

Coefficients: (1 not defined because of singularities)

            Estimate Std. Error t value Pr(>|t|)

(Intercept)  2.60536    0.95218   2.736 0.006629 \*\*

lnINCOME     0.45151    0.07872   5.736 2.61e-08 \*\*\*

EDUCATION    0.20467    0.03862   5.299 2.42e-07 \*\*\*

NUMHH        0.24770    0.06940   3.569 0.000424 \*\*\*

MAR0         0.23234    0.53283   0.436 0.663155

MAR1         0.78941    0.49532   1.594 0.112169

MAR2              NA         NA      NA       NA

---

Signif. codes:  0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.513 on 269 degrees of freedom

Multiple R-squared:  0.358,

Adjusted R-squared:  0.3461

F-statistic:    30 on 5 and 269 DF,  p-value: < 2.2e-16

> #It will pick one of these to not include. YOU CANNOT DO THIS. will get one row of NAs in your output. The row that doesn't get values is automatically incorporated into the baseline intercept.

> boxplot(lnFACE ~ factor(MARSTAT), ylab="lnFACE", xlab="MARSTAT")

>

> #to interpret the intercept in the model above, you must also keep in mind which category is the reference level basebline. If someone has 0 in  lnINCOME, 0 years of EDUCATION, 0 number of people in their household,and if someone who is 'living with a partner' has a lnFACE value of 2.60536.

> boxplot(lnFACE ~ factor(MARSTAT), ylab="lnFACE", xlab="MARSTAT")