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

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

> attach(life1)

> lnFACE = log(FACE)

> lnINCOME = log(INCOME)

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

> summary(model1)

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

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

> # s of full model = 1.513

> anova(model1)

Analysis of Variance Table

Response: lnFACE

Df Sum Sq Mean Sq F value Pr(>F)

lnINCOME 1 222.63 222.629 97.280 < 2.2e-16 \*\*\*

EDUCATION 1 51.50 51.502 22.504 3.407e-06 \*\*\*

NUMHH 1 54.34 54.336 23.743 1.883e-06 \*\*\*

factor(MARSTAT) 2 14.81 7.406 3.236 0.04085 \*

Residuals 269 615.62 2.289

---

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

> #eRROR Sum of squares of full model = 615.62

> anova(model2)

Analysis of Variance Table

Response: lnFACE

Df Sum Sq Mean Sq F value Pr(>F)

lnINCOME 1 222.63 222.629 95.701 < 2.2e-16 \*\*\*

EDUCATION 1 51.50 51.502 22.139 4.048e-06 \*\*\*

NUMHH 1 54.34 54.336 23.357 2.256e-06 \*\*\*

Residuals 271 630.43 2.326

---

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

> #ERROR Sum of squares of reduced model = 630.43

> #p = # term being dropped =2

> summary(model2)

Call:

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

Residuals:

Min 1Q Median 3Q Max

-5.7420 -0.8681 0.0549 0.9093 4.7187

Coefficients:

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

(Intercept) 2.58408 0.84643 3.053 0.00249 \*\*

lnINCOME 0.49353 0.07754 6.365 8.32e-10 \*\*\*

EDUCATION 0.20641 0.03883 5.316 2.22e-07 \*\*\*

NUMHH 0.30605 0.06333 4.833 2.26e-06 \*\*\*

---

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

Residual standard error: 1.525 on 271 degrees of freedom

Multiple R-squared: 0.3425, Adjusted R-squared: 0.3353

F-statistic: 47.07 on 3 and 271 DF, p-value: < 2.2e-16

> dim(life1)

[1] 275 18

> #critical F value if Alpha = .05

> qf(.95, df1=2,df2=269, lower.tail=TRUE)

[1] 3.029343

> qf(.05, df1=2,df2=269, lower.tail=FALSE)

[1] 3.029343

> # Our F test stat = 3.236 > this F critical value = 3.0293,

> # so we reject Ho and believe the H1.This means factor(MARSTAT)

> #terms are jointly important and should not be dropped from the model

http://stat.ethz.ch/R-manual/R-patched/library/stats/html/Fdist.html

|  |  |
| --- | --- |
| lower.tail | logical; if TRUE (default), probabilities are *P[X ≤ x]*, otherwise, *P[X > x]*. |

> pf(3.236, df1=2, df2=269, lower.tail=FALSE)

[1] 0.04085667

> #area to the right of our 'high' F test stat

> Ho = 0;(They r all equal)

>H1!=0;

> #since this p-value=.04086 < alpha =.05,we should reject Ho

>They are jointly significant.

> anova(model2,model1)

Analysis of Variance Table

Model 1: lnFACE ~ lnINCOME + EDUCATION + NUMHH

Model 2: lnFACE ~ lnINCOME + EDUCATION + NUMHH + factor(MARSTAT)

Res.Df RSS Df Sum of Sq F Pr(>F)

1 271 630.43

2 269 615.62 2 14.812 3.236 0.04085 \*

---

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

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

> attach(auto)

The following objects are masked from life1:

AGE, GENDER

> hist(PAID)

> # strongly right skewed,so lets take the log

> lnPAID = log(PAID)

> hist(lnPAID)

> model1= lm(lnPAID ~ CLASS)

> # this is us making a 1 factor ANNOVA model,basically regression with

> # no numerica, variable,we don't need the factor(CLASS) AS OUR

> #entry because R automat defaults letter entried varibale as catagorical

> summary(model1)

Call:

lm(formula = lnPAID ~ CLASS)

Residuals:

Min 1Q Median 3Q Max

-4.7002 -0.6912 -0.0437 0.7128 4.1009

Coefficients:

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

(Intercept) 6.940938 0.039699 174.838 <2e-16 \*\*\*

CLASSC11 0.010564 0.050696 0.208 0.8349

CLASSC1A -0.075432 0.128202 -0.588 0.5563

CLASSC1B 0.057420 0.065380 0.878 0.3798

CLASSC1C -0.154707 0.178007 -0.869 0.3848

CLASSC2 -0.139442 0.142595 -0.978 0.3282

CLASSC6 -0.015057 0.053217 -0.283 0.7772

CLASSC7 -0.039775 0.053191 -0.748 0.4546

CLASSC71 0.012730 0.050887 0.250 0.8025

CLASSC72 0.241716 0.122626 1.971 0.0487 \*

CLASSC7A 0.122755 0.108174 1.135 0.2565

CLASSC7B 0.131512 0.056956 2.309 0.0210 \*

CLASSC7C 0.302596 0.125307 2.415 0.0158 \*

CLASSF1 0.062962 0.202561 0.311 0.7559

CLASSF11 -0.136891 0.173726 -0.788 0.4307

CLASSF6 -0.030546 0.094148 -0.324 0.7456

CLASSF7 -0.363874 0.144807 -2.513 0.0120 \*

CLASSF71 -0.005476 0.117810 -0.046 0.9629

---

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

Residual standard error: 1.07 on 6755 degrees of freedom

Multiple R-squared: 0.005048, Adjusted R-squared: 0.002544

F-statistic: 2.016 on 17 and 6755 DF, p-value: 0.00786

> #All of these coefficients are the change u would get in ur lnPAID

> #VALUE from being in that catagory,over being in catagory C1

> #avg lnPAID value for those in Catagory C1 = 6.9409

> #avg lnPAID value for those in Catagory C11 = 6.9409 + .0105 = 6.9514

> boxplot(lnPAID~CLASS)

> model2 = lm(lnPAID~CLASS +STATE +AGE +GENDER)

> summary(model2)

Call:

lm(formula = lnPAID ~ CLASS + STATE + AGE + GENDER)

Residuals:

Min 1Q Median 3Q Max

-4.7266 -0.6802 -0.0433 0.7072 4.1809

Coefficients:

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

(Intercept) 6.974923 0.140144 49.770 < 2e-16 \*\*\*

CLASSC11 0.055573 0.051695 1.075 0.282410

CLASSC1A -0.110591 0.128238 -0.862 0.388506

CLASSC1B 0.022671 0.066388 0.341 0.732740

CLASSC1C -0.160355 0.178056 -0.901 0.367842

CLASSC2 -0.183935 0.142784 -1.288 0.197719

CLASSC6 0.057560 0.058767 0.979 0.327395

CLASSC7 -0.021854 0.054531 -0.401 0.688605

CLASSC71 0.024622 0.053049 0.464 0.642569

CLASSC72 0.260057 0.123451 2.107 0.035192 \*

CLASSC7A 0.119315 0.108879 1.096 0.273183

CLASSC7B 0.124196 0.059098 2.102 0.035634 \*

CLASSC7C 0.299023 0.126300 2.368 0.017934 \*

CLASSF1 0.130054 0.202380 0.643 0.520491

CLASSF11 -0.058684 0.174394 -0.337 0.736503

CLASSF6 0.068612 0.098270 0.698 0.485079

CLASSF7 -0.309974 0.145271 -2.134 0.032898 \*

CLASSF71 0.029459 0.118848 0.248 0.804239

STATESTATE 02 0.098538 0.089280 1.104 0.269765

STATESTATE 03 0.006302 0.101050 0.062 0.950272

STATESTATE 04 0.019266 0.093712 0.206 0.837118

STATESTATE 06 0.283853 0.094105 3.016 0.002568 \*\*

STATESTATE 07 0.076593 0.106361 0.720 0.471475

STATESTATE 10 0.181802 0.105448 1.724 0.084739 .

STATESTATE 11 0.172272 0.365574 0.471 0.637487

STATESTATE 12 0.388006 0.108333 3.582 0.000344 \*\*\*

STATESTATE 13 0.188491 0.111980 1.683 0.092371 .

STATESTATE 14 0.068141 0.116720 0.584 0.559373

STATESTATE 15 0.065570 0.086624 0.757 0.449110

STATESTATE 17 0.199398 0.096914 2.057 0.039680 \*

AGE -0.003021 0.001690 -1.787 0.073914 .

GENDERM 0.038953 0.026907 1.448 0.147747

---

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

Residual standard error: 1.066 on 6741 degrees of freedom

Multiple R-squared: 0.01365, Adjusted R-squared: 0.009113

F-statistic: 3.009 on 31 and 6741 DF, p-value: 4.354e-08

> #The baseline value of 6.9749 of the interpret represent someone

> # who is in catagpry C1,State 1 Age of 0 and Female

> # This above is not what we call a one factor ANNOva model

> # Above the F p-value = .00786 told us how the model as a whole

> # is doing and the Class var was the whole model.So even though

> # individual terms may not have been significant,we know that the

> #Class var as a whole is significant since p-value = .00786

> #alpha = .05