Jiayuan Shi

MA684 HW5

1A. The slope of the ‘Repub’ variable:

Model A: the mean political awareness score for republican is 0.34 unit lower than democrat, controlling for age, sex and income.

Model B: the mean political awareness score for republican is 2.31 unit lower than independent, controlling for age, sex and income.

The slope of the ‘RepubPA’ variable:

Model C: the mean political awareness score for republicans is 1 unit lower than the average mean across all three categories, controlling for age, sex and income.

These slopes differ across these three models because in each model we pick up different indicator variable and reference variable, so the difference in mean political awareness score for the indicator variable compared to the reference variable is also different.

Repub is significant in Model A but not in Model B because in Model A, democrat is a reference variable, but in Model B, independent is the reference variable.

1B.

Model A: score = 1.93+0.06\*30-0.08\*1-0.01\*50+0=3.15

Model B: score = 4.24+0.06\*30-0.08\*1-0.01\*50-2.31\*1=3.15

Model C: score = 2.58+0.06\*30-0.08\*1-0.01\*50+1-1.65=3.15

2.

Partial R2 = 0.825-0.764 = 0.061 = 6.1%

Partial F:

p-value=0.36>0.05

H0: R2Reduced = R2Full, H1: H0 is not true

So we cannot reject the null hypothesis, which is that the Full model explains no more variability than the Reduced model.

Hence, we can conclude housing prices do not significantly differ across neighborhoods, after controlling for house size and the age of the house.

3A.

> depression <- read.csv("depression2012.csv",header=T)

> attach(depression)

> lm1 <- lm(dep\_symptoms~age+sexmale+yrsed+drinks)

> summary(lm1)

Call:

lm(formula = dep\_symptoms ~ age + sexmale + yrsed + drinks)

Residuals:

Min 1Q Median 3Q Max

-7.2010 -2.2581 0.1919 2.1905 7.3774

Coefficients:

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

(Intercept) 13.30492 1.19020 11.179 < 2e-16 \*\*\*

age -0.06996 0.01400 -4.996 1.01e-06 \*\*\*

sexmale -0.39773 0.34860 -1.141 0.25482

yrsed -0.22190 0.07154 -3.102 0.00211 \*\*

drinks 0.14179 0.05410 2.621 0.00922 \*\*

---

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

Residual standard error: 2.979 on 295 degrees of freedom

Multiple R-squared: 0.1404, Adjusted R-squared: 0.1288

F-statistic: 12.05 on 4 and 295 DF, p-value: 4.413e-09

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Slope | SE | p-value |
| Intercept  Age  Sex (1-males 0-females)  Years Education  Drinks per week | 13.30492  -0.06996  -0.39773  -0.22190  0.14179 | 1.19020  0.01400  0.34860  0.07154  0.05410 | < 2e-16  1.01e-06  0.25482  0.00211  0.00922 |

Model R2 = 0.1404

The proportion of variability in the depression explained by the model is 14.04%.

F-statistic = 12.05, df = (4, 295), p-value = 4.413e-09< 0.01

We can reject H0, which is and we can conclude that there is significant association between the depression and any of the independent variables.

Also, based on this analysis, there is a significant association between alcohol consumption and depressive symptoms, because the p-value is 0.00922, less than 0.01, so we can reject the null hypothesis which is there is no association between alcohol consumption and depressive symptoms.

Interpret the slope for alcohol consumption: If all other variables in the model are held constant, for each one increase in the alcoholic drinks per week, on average, the depressive symptoms expected to increase 0.14179.

3B.

> lighter <- NULL

> lighter[drinkcat==1] <- 1

> lighter[drinkcat==0 | drinkcat==2] <- 0

> heavier <- NULL

> heavier[drinkcat==2] <- 1

> heavier[drinkcat==0 | drinkcat==1] <- 0

> lm2 <- lm(dep\_symptoms~age+sexmale+yrsed+lighter+heavier)

> summary(lm2)

Call:

lm(formula = dep\_symptoms ~ age + sexmale + yrsed + lighter +

heavier)

Residuals:

Min 1Q Median 3Q Max

-7.398 -2.071 0.144 2.238 7.790

Coefficients:

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

(Intercept) 13.44591 1.19814 11.222 < 2e-16 \*\*\*

age -0.06223 0.01488 -4.181 3.83e-05 \*\*\*

sexmale -0.57524 0.36416 -1.580 0.11527

yrsed -0.21045 0.07163 -2.938 0.00356 \*\*

lighter -0.28337 0.40656 -0.697 0.48637

heavier 1.64234 0.69892 2.350 0.01944 \*

---

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

Residual standard error: 2.979 on 294 degrees of freedom

Multiple R-squared: 0.1431, Adjusted R-squared: 0.1285

F-statistic: 9.819 on 5 and 294 DF, p-value: 1.079e-08

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Slope | SE | p-value |
| Intercept  Age  Sex (1-males 0-females)  Years Education  Lighter Drinkers  Heavier Drinkers | 13.44591  -0.06223  -0.57524  -0.21045  -0.28337  1.64234 | 1.19814  0.01488  0.36416  0.07163  0.40656  0.69892 | < 2e-16  3.83e-05  0.11527  0.00356  0.48637  0.01944 |

Model R2 = 0.1431

The proportion of variability in the depression explained by the model is 14.31%.

The slope for Lighter Drinkers would be interpreted as the difference in mean depressive symptoms for lighter drinkers compared to non-drinkers is -0.28337, controlling for age, sex and years of education.

The slope for Heavier Drinkers would be interpreted as the difference in mean depressive symptoms for heavier drinkers compared to non-drinkers is 1.64234, controlling for age, sex and years of education.

3C.

> lm.full <- lm(dep\_symptoms~age+sexmale+yrsed+lighter+heavier)

> summary(lm.full)

Call:

lm(formula = dep\_symptoms ~ age + sexmale + yrsed + lighter +

heavier)

Residuals:

Min 1Q Median 3Q Max

-7.398 -2.071 0.144 2.238 7.790

Coefficients:

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

(Intercept) 13.44591 1.19814 11.222 < 2e-16 \*\*\*

age -0.06223 0.01488 -4.181 3.83e-05 \*\*\*

sexmale -0.57524 0.36416 -1.580 0.11527

yrsed -0.21045 0.07163 -2.938 0.00356 \*\*

lighter -0.28337 0.40656 -0.697 0.48637

heavier 1.64234 0.69892 2.350 0.01944 \*

---

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

Residual standard error: 2.979 on 294 degrees of freedom

Multiple R-squared: 0.1431, Adjusted R-squared: 0.1285

F-statistic: 9.819 on 5 and 294 DF, p-value: 1.079e-08

> summary.aov(lm.full)

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

age 1 278.4 278.35 31.364 4.91e-08 \*\*\*

sexmale 1 4.5 4.48 0.504 0.47814

yrsed 1 83.8 83.77 9.438 0.00232 \*\*

lighter 1 20.1 20.14 2.269 0.13303

heavier 1 49.0 49.01 5.522 0.01944 \*

Residuals 294 2609.3 8.88

---

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

> lm.reduced <- lm(dep\_symptoms~age+sexmale+yrsed)

> summary(lm.reduced)

Call:

lm(formula = dep\_symptoms ~ age + sexmale + yrsed)

Residuals:

Min 1Q Median 3Q Max

-7.4415 -2.0739 0.1715 2.2111 7.5953

Coefficients:

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

(Intercept) 14.11557 1.16064 12.162 < 2e-16 \*\*\*

age -0.07880 0.01373 -5.741 2.33e-08 \*\*\*

sexmale -0.25186 0.34752 -0.725 0.46919

yrsed -0.21980 0.07224 -3.043 0.00256 \*\*

---

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

Residual standard error: 3.008 on 296 degrees of freedom

Multiple R-squared: 0.1204, Adjusted R-squared: 0.1115

F-statistic: 13.5 on 3 and 296 DF, p-value: 2.78e-08

> summary.aov(lm.reduced)

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

age 1 278.4 278.35 30.762 6.47e-08 \*\*\*

sexmale 1 4.5 4.48 0.495 0.48238

yrsed 1 83.8 83.77 9.257 0.00256 \*\*

Residuals 296 2678.4 9.05

---

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

Partial R2 = 0.1431-0.1204=0.0227

> anova(lm.full,lm.reduced)

Analysis of Variance Table

Model 1: dep\_symptoms ~ age + sexmale + yrsed + lighter + heavier

Model 2: dep\_symptoms ~ age + sexmale + yrsed

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

1 294 2609.3

2 296 2678.4 -2 -69.146 3.8956 0.02139 \*

---

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

Partial F = 3.8956, df=2, 294

P-value = 0.02139 <0.05. We can reject H0, which is the Full model explains no more variability than the Reduced model. Hence, the association between alcohol consumption and depressive symptoms is significant in this model.

3D.

> lm3 <- lm(dep\_symptoms~age+sexmale+yrsed+relevel(factor(drinkcat),'0'))

> summary(lm3)

Call:

lm(formula = dep\_symptoms ~ age + sexmale + yrsed + relevel(factor(drinkcat),

"0"))

Residuals:

Min 1Q Median 3Q Max

-7.398 -2.071 0.144 2.238 7.790

Coefficients:

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

(Intercept) 13.44591 1.19814 11.222 < 2e-16 \*\*\*

age -0.06223 0.01488 -4.181 3.83e-05 \*\*\*

sexmale -0.57524 0.36416 -1.580 0.11527

yrsed -0.21045 0.07163 -2.938 0.00356 \*\*

relevel(factor(drinkcat), "0")1 -0.28337 0.40656 -0.697 0.48637

relevel(factor(drinkcat), "0")2 1.64234 0.69892 2.350 0.01944 \*

---

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

Residual standard error: 2.979 on 294 degrees of freedom

Multiple R-squared: 0.1431, Adjusted R-squared: 0.1285

F-statistic: 9.819 on 5 and 294 DF, p-value: 1.079e-08

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Slope | SE | p-value |
| Intercept  Age  Sex (1-males 0-females)  Years Education  Lighter Drinkers  Heavier Drinkers | 13.44591  -0.06223  -0.57524  -0.21045  -0.28337  1.64234 | 1.19814  0.01488  0.36416  0.07163  0.40656  0.69892 | < 2e-16  3.83e-05  0.11527  0.00356  0.48637  0.01944 |

Model R2 = 0.1431

The results of this analysis are the same as the results from 3B.

3E. Type II partial F-test:

> Anova(lm3,type="II")

Anova Table (Type II tests)

Response: dep\_symptoms

Sum Sq Df F value Pr(>F)

age 155.17 1 17.4844 3.827e-05 \*\*\*

sexmale 22.15 1 2.4952 0.115266

yrsed 76.61 1 8.6326 0.003563 \*\*

relevel(factor(drinkcat), "0") 69.15 2 3.8956 0.021390 \*

Residuals 2609.26 294

---

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

Type II partial F=3.8956 = the partial F in 3C

So these results of the F-test are the same as results in 3C.

3F.

I think treating alcohol consumption as a categorical variable (abstainers, light, heavy drinkers in 3D) is more appropriate, because even they both show that there is an association between alcohol consumption and depressive symptoms, different levels(none, lighter, heavier) of drinking have different influence on depressive symptoms. We should also discuss those situations on by one.