Name: Khoa Tran

**Assignment 9**

1.

Text, letter

Description automatically generated

2.

(1)

> sum(with(data, data$abvavg==1))

[1] 383

> sum(with(data, data$abvavg==1))/nrow(data)

[1] 0.3039683

(2)

> ols <- lm(abvavg~female,data = data)

> summary(ols)

Call:

lm(formula = abvavg ~ female, data = data)

Residuals:

Min 1Q Median 3Q Max

-0.3303 -0.2900 -0.2900 0.6697 0.7099

Coefficients:

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

(Intercept) 0.29005 0.01602 18.102 <2e-16 \*\*\*

female 0.04023 0.02724 1.477 0.14

---

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

Residual standard error: 0.4599 on 1258 degrees of freedom

Multiple R-squared: 0.001731, Adjusted R-squared: 0.0009373

F-statistic: 2.181 on 1 and 1258 DF, p-value: 0.14

* We can observe from the regression above that female is statistically insignificant at 5% level. Thus, the population fractions of above-average-looking women and men are likely to be the same at the 5% level (i.e. being female or male won’t increase the probability of being above average looking)

(3)

> ols2 <-lm(married~belavg+abvavg+bigcity+educ,data=data)

> summary(ols2)

Call:

lm(formula = married ~ belavg + abvavg + bigcity + educ, data = data)

Residuals:

Min 1Q Median 3Q Max

-0.7710 -0.6466 0.2692 0.3242 0.4259

Coefficients:

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

(Intercept) 0.799675 0.064445 12.409 <2e-16 \*\*\*

belavg -0.017005 0.040860 -0.416 0.6774

abvavg -0.072971 0.029322 -2.489 0.0130 \*

bigcity -0.055020 0.031673 -1.737 0.0826 .

educ -0.005737 0.005046 -1.137 0.2557

---

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

Residual standard error: 0.4607 on 1255 degrees of freedom

Multiple R-squared: 0.009414, Adjusted R-squared: 0.006257

F-statistic: 2.982 on 4 and 1255 DF, p-value: 0.01826

* Interpretation of belavg estimate: having a below average appearance decrease the chance of getting married by 1.7%. Yet it is not statistically significant at 5% level, so it is likely that having below average appearance won’t affect chance of getting married in the population.
* Interpretation of abvavg estimate: having an above average appearance decrease the chance of getting married by 7% (interesting!) and the estimate is statistically significant at 5% level.

(4)

> data$LPM.fitted <- ols2$fitted.values

> data$LPM.fitted <- ols2$fitted.values

>

> data$fit.married[data$LPM.fitted>=0.5]<-1

> data$fit.married[data$LPM.fitted<0.5]<-0

>

> sum(with(data, fit.married==married))

[1] 871

> sum(with(data, fit.married==married))/nrow(data)

[1] 0.6912698

(5)

> summary(data$LPM.fitted)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.5742 0.6579 0.7022 0.6913 0.7308 0.7710

* Discussion: we observe that no value of LPM.fitted is outside of the range [0,1]. Thus, the LPM model can be a relatively good model to facilitate our discussion. However, we need to note that there are still 2 issues. The first one is the issue of heteroskedasticity. The second issue is that the model fitted 1 for all observations (min of fitted > 0.5). Thus, it is likely that this model is not valid for the sample that we have.

(6)

> #Chow test to test whether we have to specify the model (A) differently for men and women

> # at the 5% level.

> #Overall model is ols2

> SSRA2 <- sum(ols2$residuals^2)

> #Find SSR for male only

> ols2M <-lm(married~belavg+abvavg+bigcity+educ,data=data,subset=(female==0))

> SSRA2M <- sum(ols2M$residuals^2)

> #find SSR for female only

> ols2N <-lm(married~belavg+abvavg+bigcity+educ,data=data,subset=(female==1))

> SSRA2N <- sum(ols2N$residuals^2)

> #Calculating Chow statistic

> kplus1<-nrow(summary(ols2)$coef)

> obsn<-ols2$df.residual+kplus1

> ChowS <-((SSRA2 - (SSRA2M+SSRA2N))/(kplus1))/((SSRA2M+SSRA2N)/(obsn-2\*(kplus1)))

> FC<-qf(.95, df1=kplus1, df2=(obsn-2\*(kplus1)))

> ChowP<-pf(ChowS,df1=kplus1, df2=(obsn-2\*(kplus1)),lower.tail=FALSE)

> ChowP

[1] 5.335843e-28

* We reject the null hypothesis at 5% level. Thus, we need to specify model A differently for men and women.