Econometrics Assignment 2

Eric Tu

(a) (i) Regression Coefficients

	Estimate	*Std.Error*	$Pr(>\|t\|)$
(*Intercept*)	2.442	0.155	0.000
SATV	0.063	0.028	0.023

(ii) 95 Percent Confidence Interval [0.007759 0.118413]

(b) Multiple Regression Coefficients

		Estimate	*Std.Error*	$Pr(>\|t\|)$
(i)	(*Intercept*)	1.557	0.216	0.000
	SATV	0.014	0.028	0.612
	SATM	0.173	0.032	0.000
	FEM	0.200	0.037	0.000

(ii) 95% Confidence Interval [-0.041692 0.070016]

(c) Correlation Matrix

	FGPA	*SATM*	*SATV*	*FEM*
FGPA	1.000	0.195	0.092	0.176
SATM	0.195	1.000	0.288	−0.162
SATV	0.092	0.288	1.000	0.034
FEM	0.176	−0.163	0.033	1.0000

The correlation matrix shows that the linear relationship found between FGPA to the SATV coefficient in part a is primarily due to the partial effects of SATM. Thus the SATV coefficient in part b is much less than in part a, as the contribution to FGPA is captured in the SATM coefficient.

(i) Unrestricted Model: $R_1^2 = 0.08296, n = 609, k = 3$ on Multiple Regression Model)
Restricted Model: $R_0^2 = 0.082574, g = 1$ on FGPA regressed against SATM and FEM
$F = \frac{\frac{(R_1^2 - R_0^2)}{g}}{\frac{1-R_1^2}{n-k}} = 0.257580$ F is less than critical value 3.9
Cannot reject the null that SATV is insignificant, SATV is most likely insignificant based on 5 percent p-value of F test.

(ii) Unrestricted Model, t = 0.507
$t^2 = 0.257 = F$

R Code attached

Package dependencies library(”xlsx”)

Read Data data ¡- read.xlsx(”TestExer2-GPA-round2.xls”, 1)

Data Summary # Observations = 609 # FGPA: Freshman grade point average (scale 0-4) # SATV: Score on SAT Verbal test (scale 0-10) # SATM: Score on SAT Mathematics test (scale 0-10) # FEM: Gender dummy (1 for females, 0 for males)

part a # (i) # Regress FGPA on a constant and SATV fit ¡- lm(FGPA SATV, data=data) # Report Coefficient of SATV and its standard error and p-value # Within 3 decimals coef ¡- summary(fit)$coef SATV ¡- coef[”SATV”, colnames(coef[, c(1, 2, 4)])] print(”SATV Coefficients:”) print(round(SATV, 3)) # Sanity check plot(data$SATV, data$FGPA, main=”FGPA vs SATV”, xlab=”SATV”, ylab=”FGPA”) abline(fit)

(ii) # Determine a 95% confidence interval (with 3 decimals) for # the effect on FGPA of an increase by 1 point in SATV min95 ¡- round(SATV[1] - 2 * SATV[2],6) max95 ¡- round(SATV[1] + 2 * SATV[2],6) print("95% Confidence Interval of effect on FGPA by 1 point increase in SATV:") print(c(min95, max95))

part b # Regress FGPA on a constant, SATV, SATM, and FEM fit_multi ¡- lm(FGPA   SATV + SATM + FEM, data=data) # Report Coefficient of SATV and its standard error and p-value # Within 3 decimals coef_multi ¡- summary(fit_multi)$coef coef_ans ¡- coef_multi[c("(Intercept)","SATV", "SATM", "FEM"), colnames(coef_multi[, c(1, 2, 4)])] print("Coefficients:") print(round(coef_ans, 6))

(ii) # Determine a 95% confidence interval (with 3 decimals) for # the effect on FGPA of an increase by 1 point in SATV min95_multi ¡- sum(round(coef_ans[1] - 2 * coef_ans[2],6)) max95_multi ¡- sum(round(coef_ans[1] + 2 * coef_ans[2],6)) print("95% Confidence Interval of effect on FGPA by 1 point increase in SATV:") print(c(min95_multi, max95_multi))

part c # Correlation Matrix print("Correlation Matrix") corr ¡- round(cor(data[,c(2,3,4,5)]),6) print(corr)

part d # Perform an F-test on the significance of the effect of SATV on FGPA, based on the # regression in part b and another regression

Unrestricted model R$\hat{2}$ r1_sq ¡- summary(fit_multi)$r.squared n ¡- nrow(data)

Restricted model R$\hat{2}$ fit_restr ¡- lm(FGPA   SATM + FEM, data=data) r0_sq ¡- summary(fit_restr)$r.squared g ¡- 1 k ¡- 3

F_score ¡- ((r1_sq - r0_sq)/g)/((1-r1_sq)/(n-k)) print("F Score") print(F_score)

if(F_score ¿ 3.9){ print("SATV is significant") }else{ print("SATV is not significant") }