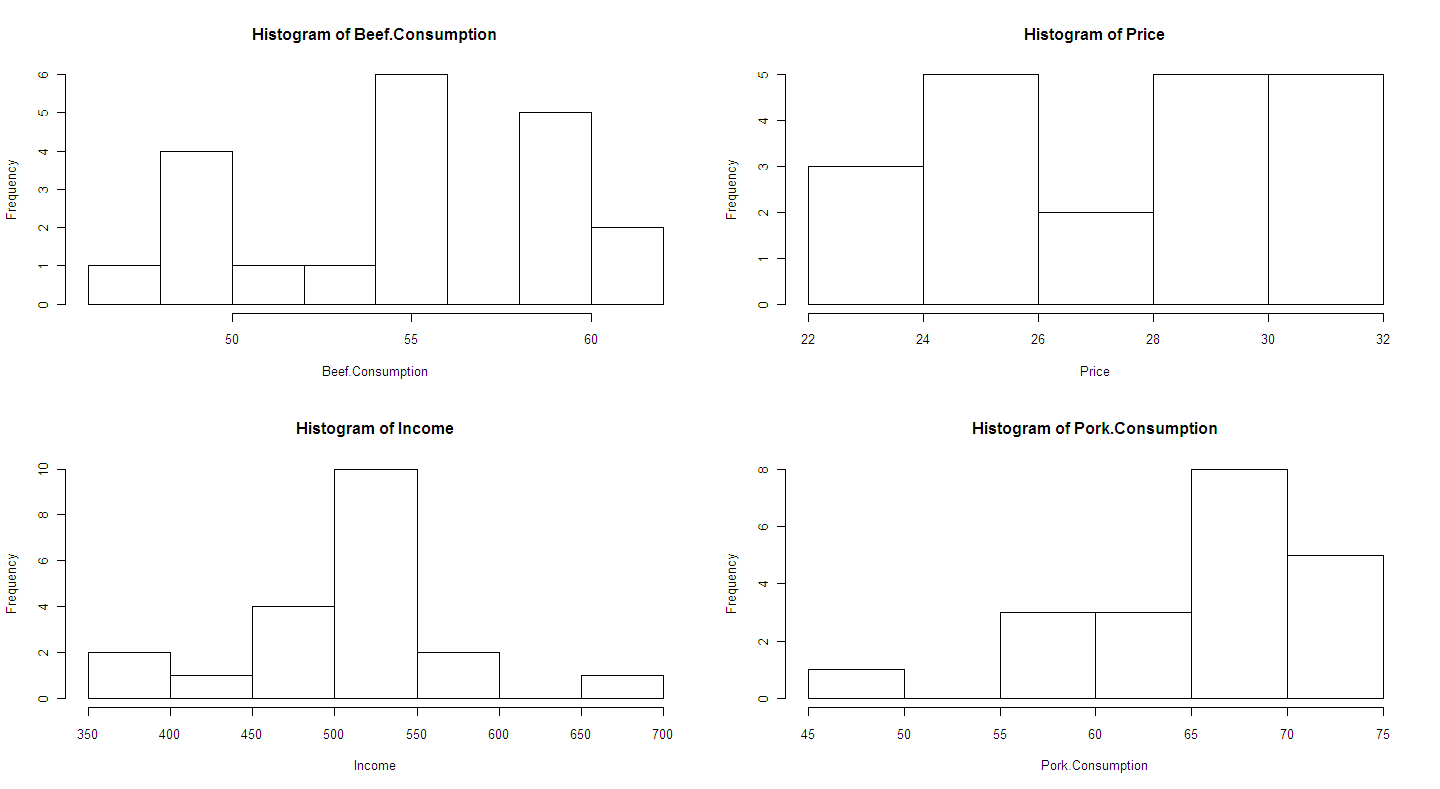
Sketch Solutions for Assignment 4

1. Examine the data with some preliminary graphical displays





Some evidence of a quadratic temporal trend in beef consumption; however, we are not asked to investigate “Year” so this will be ignored. From the plots, there is no apparent relationship between beef consumption and price (which seems counter-intuitive!); an increasing trend with income (which may as well be linear); and an apparently non-linear relationship with pork consumption. Start by considering a linear regression model with all possible interactions and a quadratic term for pork consumption. This model has 11 parameters.

Fit a sequence of models until a minimum adequate model is obtained:

> fit1<-lm(Beef.Consumption~Price\*Income\*Pork.Consumption + I(Pork.Consumption^2))

>fit2<-update(fit1, .~.-I(Pork.Consumption^2))

>fit3<-update(fit2,.~.- Price:Income:Pork.Consumption)

>fit4<-update(fit3,.~.-Income:Pork.Consumption)

>fit5<-update(fit4, .~. - Price:Income)

>fit6<-update(fit5, .~. - Price:Pork.Consumption)

> summary(fit6)

*Call:*

*lm(formula = Beef.Consumption ~ Price + Income + Pork.Consumption)*

*Residuals:*

*Min 1Q Median 3Q Max*

*-2.44996 -0.87212 0.04715 0.73206 2.47242*

*Coefficients:*

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

*(Intercept) 90.813646 5.266047 17.245 9.28e-12 \*\*\**

*Price -1.849850 0.145990 -12.671 9.32e-10 \*\*\**

*Income 0.083190 0.006868 12.113 1.80e-09 \*\*\**

*Pork.Consumption -0.415085 0.053945 -7.695 9.15e-07 \*\*\**

*---*

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

*Residual standard error: 1.372 on 16 degrees of freedom*

*Multiple R-squared: 0.9212, Adjusted R-squared: 0.9064*

*F-statistic: 62.33 on 3 and 16 DF, p-value: 4.799e-09*

The final model has all three main effects as statistically significant, and has an R-squared of 0.92.

Examine some model diagnostics:



No evidence of heteroscedasticity, no excessively large outliers, and errors look normal. Observation 20 may be influential. Remove this observation and refit:

fit6b<-update(fit6,.~.,subset=(1:length(Beef.Consumption)!=20))

> summary(fit6b)

Call:

lm(formula = Beef.Consumption ~ Price + Income + Pork.Consumption,

subset = (1:length(Beef.Consumption) != 20))

Residuals:

Min 1Q Median 3Q Max

-2.6827 -0.7044 0.3086 0.8855 2.2941

Coefficients:

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

(Intercept) 89.400665 5.344763 16.727 4.13e-11 \*\*\*

Price -1.894812 0.149369 -12.685 2.02e-09 \*\*\*

Income 0.089333 0.008584 10.406 2.95e-08 \*\*\*

Pork.Consumption -0.420366 0.053526 -7.854 1.08e-06 \*\*\*

---

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

Residual standard error: 1.356 on 15 degrees of freedom

Multiple R-squared: 0.9191, Adjusted R-squared: 0.903

F-statistic: 56.84 on 3 and 15 DF, p-value: 2.007e-08

Estimates seem stable to the removal of this observation. Finish off by writing down the final model and interpreting each term carefully…

2. Fit the model specified in the question:

> fit<-glm(Fall~Int + Sex+BI+SI,family=poisson)

> summary(fit)

Call:

glm(formula = Fall ~ Int + Sex + BI + SI, family = poisson)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.1854 -0.7819 -0.2564 0.5449 2.3625

Coefficients:

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

(Intercept) 0.489467 0.336869 1.453 0.14623

Int -1.069403 0.133154 -8.031 9.64e-16 \*\*\*

Sex -0.046606 0.119970 -0.388 0.69766

BI 0.009470 0.002953 3.207 0.00134 \*\*

SI 0.008566 0.004312 1.986 0.04698 \*

---

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 199.19 on 99 degrees of freedom

Residual deviance: 108.79 on 95 degrees of freedom

AIC: 377.29

Number of Fisher Scoring iterations: 5

Approximate 95% confidence intervals based on the normal distribution:

> beta<-c(0.489467,-1.069403,-0.046606,0.009470,0.008566)

> beta.se<-c(0.336869,0.133154,0.119970,0.002953,0.004312)

>

> cbind(beta-1.96\*beta.se,beta+1.96\*beta.se)

[,1] [,2]

[1,] -0.17079624 1.14973024

[2,] -1.33038484 -0.80842116

[3,] -0.28174720 0.18853520

[4,] 0.00368212 0.01525788

[5,] 0.00011448 0.01701752

b) The p-value for a goodness-of-fit test based on the deviance is:

> 1-pchisq(108.79,df=95)

[1] 0.1577902

Not sufficient evidence to reject the adequacy of this model.

c) plot(seq(1,100,1),resid(fit))



No outliers are apparent.

d)

> fit2<- glm(Fall~Int+BI+SI,family=poisson)

> anova(fit2,fit,test="Chisq")

Analysis of Deviance Table

Model 1: Fall ~ Int + BI + SI

Model 2: Fall ~ Int + Sex + BI + SI

Resid. Df Resid. Dev Df Deviance P(>|Chi|)

1 96 108.941

2 95 108.790 1 0.151 0.698

Not sufficient evidence against the hypothesis that the coefficient of gender is =0 and thus gender can be removed from the model.

e) From fit2 the 95% CI is

> c(-1.077770 -1.96\*0.131415,-1.077770 +1.96\*0.131415)

[1] -1.3353434 -0.8201966

Under repeated sampling, the proportion of intervals constructed in this manner would contain the true value of beta1 is 0.95.

f) From fit2 YES, aerobic exercise is associated with a decrease in the frequency of falls. The estimated relative risk is

> exp(-1.077770)

[1] 0.3403537