Recap: Chapter 6-8

STAT 3240

Michael McIsaac

UPEI

Learning Objectives for Chapter 6

You should be able to

- Understand the concept and utility of multiple linear regression
- Interpret general linear regression coefficients

- Be aware of qualitative predictors, polynomial regression, and interactions

- Express model, estimation, fitted values, residuals, and ANOVA in matrix form
- Conduct and interpret a general linear regression ANOVA F test
- Calculate and interpret multiple R^2 and r

- Conduct and interpret inference and joint inference on specific parameters

- Compute and interpret independent and simultatneous CIs for $E[Y_h]$ and PIs for new observations
- Apply regression diagnostics to the multiple regression setting.

Learning Objectives for Chapter 7

You should be able to

- Understand the concept of the extra sums of squares principle
- Conduct and interpret tests concerning regression coefficients using ESS principle
- Compute and interpret coefficients of partial determination
- Understand multicollinearity and its effects

Learning Objectives for Chapter 8

You should be able to

- Understand the utility and disadvantages of polynomial regression
- Understand the need for centering
- Understand the danger of overfitting
- Compute and interpret parameters in a polynomial regression model
- Understand the utility and disadvantages of interactions in regression
- Compute and interpret parameters in regression models with interactions

- Compute and interpret parameters in curvilinear regression models with interactions

- Implement and interpret regression using indicator (dummy) variables

- Implement and interpret regression involving interactions between indicator and quantitative variables
- Implement and interpret regression involving interactions between multiple indicator variables
- Implement and interpret tests for differences among regression functions

Test 2: Chapter 6-8 example

```
spending_subset = spending_subset %>% select(clothing_expenditure, income, sex, food_expe
clothing_model = lm(clothing_expenditure~income+sex+food_expenditure, data=spending_subse
msummary(clothing_model)
```

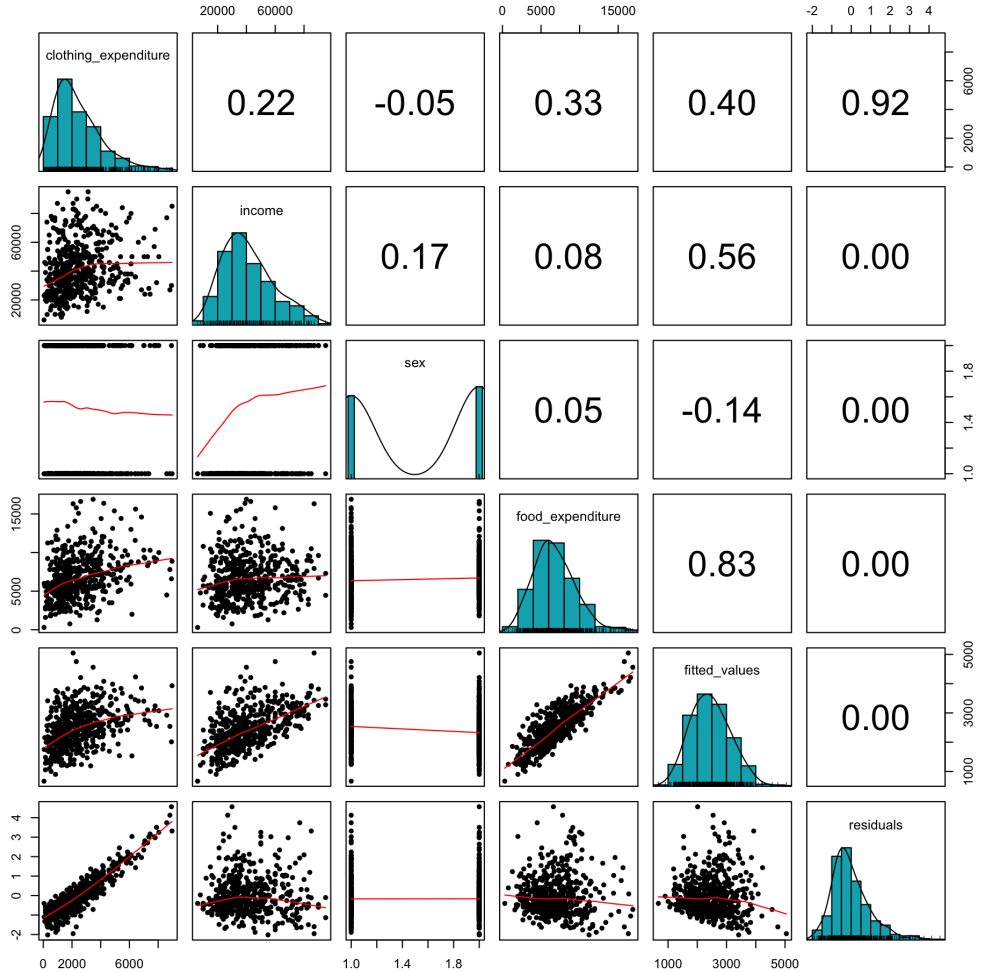
```
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       4.86e+02   2.39e+02    2.04    0.042
## income            2.04e-02   3.96e-03    5.15  3.8e-07
## sexmale          -3.58e+02   1.41e+02   -2.54    0.011
## food_expenditure  1.93e-01   2.49e-02    7.74  5.7e-14
##
## Residual standard error: 1550 on 496 degrees of freedom
## Multiple R-squared:  0.16,    Adjusted R-squared:  0.155
## F-statistic: 31.6 on 3 and 496 DF,  p-value: <2e-16
```

```
anova(clothing_model)
```

```
## Analysis of Variance Table
##
## Response: clothing_expenditure
##                   Df   Sum Sq  Mean Sq F value  Pr(>F)
## income             1 7.13e+07 7.13e+07   29.64 8.2e-08
## sex                1 1.25e+07 1.25e+07    5.21   0.023
## food_expenditure   1 1.44e+08 1.44e+08   59.89 5.7e-14
## Residuals        496 1.19e+09 2.40e+06
```

```
round(confint(clothing_model, level=.95), 4)
```

```
##                     2.5 %   97.5 %
## (Intercept)       17.3180 955.1598
## income             0.0126   0.0281
## sexmale         -634.7807 -80.8111
## food_expenditure    0.1439   0.2418
```


clothing_expenditure
0.22
-0.05
0.33
0.40
0.92
income
0.17
0.08
0.56
0.00
sex
0.05
-0.14
0.00
food_expenditure
0.83
0.00
fitted_values
0.00
residuals

```
msummary(lm(clothing_expenditure~sex, data=spending_subset))
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)     2554        110   23.32   <2e-16
## sexmale         -184        151   -1.22     0.22
##
## Residual standard error: 1690 on 498 degrees of freedom
## Multiple R-squared:  0.00298,    Adjusted R-squared:  0.000982
## F-statistic: 1.49 on 1 and 498 DF,  p-value: 0.223
```

```
anova(lm(clothing_expenditure~sex, data=spending_subset))
```

```
## Analysis of Variance Table
##
## Response: clothing_expenditure
##            Df   Sum Sq Mean Sq F value Pr(>F)
## sex         1 4.24e+06 4237682    1.49   0.22
## Residuals 498 1.42e+09 2843114
```

```
msummary(lm(clothing_expenditure~sex+(income + I(income^2))*sex, data=spending_subset))
```

```
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)          8.30e+02   5.40e+02    1.54    0.125
## sexmale             -2.95e+02   8.33e+02   -0.35    0.723
## income               6.46e-02   2.63e-02    2.45    0.014
## I(income^2)         -4.24e-07   2.85e-07   -1.49    0.138
## sexmale:income       3.39e-04   3.79e-02    0.01    0.993
## sexmale:I(income^2) -3.55e-08   3.91e-07   -0.09    0.928
##
## Residual standard error: 1640 on 494 degrees of freedom
## Multiple R-squared:  0.0701,    Adjusted R-squared:  0.0607
## F-statistic: 7.45 on 5 and 494 DF,  p-value: 9.48e-07
```

```
anova(lm(clothing_expenditure~sex+(income + I(income^2))*sex, data=spending_subset))
```

```
## Analysis of Variance Table
##
## Response: clothing_expenditure
##                  Df   Sum Sq  Mean Sq F value  Pr(>F)
## sex               1 4.24e+06  4237682    1.59   0.209
## income            1 7.95e+07 79540922   29.75 7.8e-08
## I(income^2)       1 1.54e+07 15399374    5.76   0.017
## sex:income        1 3.32e+05   332336    0.12   0.725
## sex:I(income^2)   1 2.20e+04    21995    0.01   0.928
## Residuals       494 1.32e+09  2673231
```