ME EN 2550

Ryan Dalby u0848407

Homework 8- Part A

8.2: a) Predicted birth weight = 120.07 - 1.93(parity)

b) If a child is not the first born then there is a predicted 1.93 decrease in predicted birth weight.

Predicted birth weight of firstborns = 120.07 oz.

Predicted birth weight of non-firstborns = 118.14 oz.

c) No, we fail to reject the null hypothesis that is, holding all other variables constant, parity is unpredictive of birth weight. This is because the P-value = 0.1052 which is greater than alpha = 0.05.

8.3: Predicted birth weight = -80.41 + 0.44(gestation) - 3.33(parity) - 0.01(age) + 1.15(height) +0.05(weight) - 8.40(smoke)

b) For a unit increase in gestation length, there is a predicted 0.44 oz. increase in birth weight all else held constant. For a unit increase in the age of the mother, there is a predicted 0.01oz. Decrease in birth weight all else held constant.

c) There is a difference because we are dealing with a completely different model that has many more predictors. There is also the chance that parity might be correlated with another variable in the model.

d) Actual: 120 Predicted: 120.58 Residual: -0.58

e) R^2 = 0.2504 Adjusted R^2 = 0.2468

8.5: a) 95% CI (-.3212 < coefficient of gender < 0.1612) There is a 95 % chance that the calculated interval contains the true coefficient of gender which predicts GPA when all other predictors are held constant.

b) Yes because for all other predictors the P-values are greater than alpha = 0.05.

8.7 Age would be the predictor to be removed first because without it the model has the highest R^2 adjusted of any other model that had one predictor removed.

8.13: Nearly Normal Residuals: The probability plot appears linear and thus overall normally distributed.

Constant Variability: The scatterplots show no major patterns other than from the discrete domains of some of the predictors.

Independent: The residuals all appear randomly distributed.

Linearly Related: It appears that there are no major patterns in the residual plots thus a linear model appears to work decently well for this data.

Overall it appears that the assumptions for regression are met.

8.14: Nearly Normal Residuals: The probability plot appears approximately linear and it is reasonable to conclude the data is overall normally distributed, although possibly with some skewness.

Constant Variability: The scatterplots show no major patterns other than from the discrete domains of some of the predictors.

Independent: The residuals all appear randomly distributed.

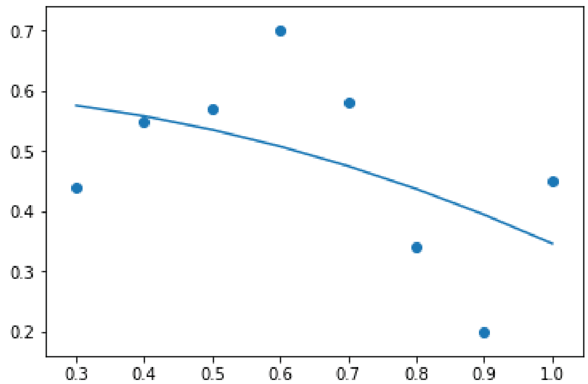
Linearly Related: It appears that there are no major patterns in the residual plots thus a linear model appears to work decently well for this data.

Overall it appears that the assumptions for regression are met.

```
Python 3.6.5 |Anaconda, Inc.| (default, Mar 29 2018, 13:32:41) [MSC v.1900 64 bit (AMD64)]
Type "copyright", "credits" or "license" for more information.

IPython 6.4.0 -- An enhanced Interactive Python.

In [1]: runfile('C:/Users/hoops/OneDrive/Documents/School/ME EN 2550 Statistics and
Probability/HW8/HW8.py', wdir='C:/Users/hoops/OneDrive/Documents/School/ME EN 2550
Statistics and Probability/HW8')
B1
b)
```




```
a)
                            OLS Regression Results
==============================================================================
Dep. Variable:              viscosity   R-squared:                       0.269
Model:                            OLS   Adj. R-squared:                  0.147
Method:                 Least Squares   F-statistic:                     2.210
Date:                Tue, 16 Apr 2019   Prob (F-statistic):              0.188
Time:                        15:39:10   Log-Likelihood:                 5.2642
No. Observations:                   8   AIC:                            -6.528
Df Residuals:                       6   BIC:                            -6.370
Df Model:                           1
Covariance Type:            nonrobust
=========================================================================================
                            coef    std err          t      P>|t|      [0.025      0.975]
-----------------------------------------------------------------------------------------
Intercept                 0.5986      0.095      6.269      0.001       0.365       0.832
np.power(ratio, 2)       -0.2523      0.170     -1.487      0.188      -0.668       0.163
==============================================================================
Omnibus:                        0.731   Durbin-Watson:                   1.510
Prob(Omnibus):                  0.694   Jarque-Bera (JB):                0.528
Skew:                          -0.085   Prob(JB):                        0.768
Kurtosis:                       1.753   Cond. No.                         4.13
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.



B2
a)
```

```
                            OLS Regression Results
==============================================================================
Dep. Variable:           satisfaction   R-squared:                       0.904
Model:                            OLS   Adj. R-squared:                  0.884
Method:                 Least Squares   F-statistic:                     46.87
Date:                Tue, 16 Apr 2019   Prob (F-statistic):           6.95e-10
Time:                        15:39:10   Log-Likelihood:                -82.062
No. Observations:                  25   AIC:                             174.1
Df Residuals:                      20   BIC:                             180.2
Df Model:                           4
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    143.8672      6.044     23.804      0.000     131.260     156.474
age           -1.1172      0.138     -8.075      0.000      -1.406      -0.829
severity      -0.5862      0.136     -4.324      0.000      -0.869      -0.303
surgmed        0.4149      3.008      0.138      0.892      -5.859       6.689
anxiety        1.3064      1.084      1.205      0.242      -0.955       3.568
==============================================================================
Omnibus:                        4.082   Durbin-Watson:                   2.102
Prob(Omnibus):                  0.130   Jarque-Bera (JB):                2.355
Skew:                          -0.665   Prob(JB):                        0.308
Kurtosis:                       3.701   Cond. No.                         297.
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
```

b)

```
Standard error of regression coefficents:
Intercept    6.043698
age          0.138342
severity     0.135556
surgmed      3.007787
anxiety      1.084055
dtype: float64
```

c) Not all the model parameters are estimated with the same precision. This is because all parameters are fit to a single model and thus the predictors will have varying standard error values/precision

B3

a)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      y   R-squared:                       0.852
Model:                            OLS   Adj. R-squared:                  0.768
Method:                 Least Squares   F-statistic:                     10.08
Date:                Tue, 16 Apr 2019   Prob (F-statistic):            0.00496
Time:                        15:39:10   Log-Likelihood:                -43.397
No. Observations:                  12   AIC:                             96.79
Df Residuals:                       7   BIC:                             99.22
Df Model:                           4
Covariance Type:            nonrobust
```

```
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   -123.1312    157.256     -0.783      0.459    -494.983     248.720
x1             0.7573      0.279      2.713      0.030       0.097       1.417
x2             7.5188      4.010      1.875      0.103      -1.964      17.001
x3             2.4831      1.809      1.372      0.212      -1.795       6.762
x4            -0.4811      0.555     -0.867      0.415      -1.794       0.832
==============================================================================
Omnibus:                        2.436   Durbin-Watson:                   1.808
Prob(Omnibus):                  0.296   Jarque-Bera (JB):                1.069
Skew:                          -0.288   Prob(JB):                        0.586
Kurtosis:                       1.656   Cond. No.                     6.82e+03
==============================================================================

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly
specified.
[2] The condition number is large, 6.82e+03. This might indicate that there are
strong multicollinearity or other numerical problems.
b)
Standard error of regression coefficents:
Intercept    157.256058
x1             0.279090
x2             4.010121
x3             1.809386
x4             0.555174
dtype: float64
Not all the model parameters are estimated with the same precision.  This is because all
parameters are fit to a single model and thus the predictors will have varying standard
error values/precision
c)
The predicted power consumption for a month with the given values is : 290.442068
dtype: float6






C:\Users\hoops\Anaconda3\lib\site-packages\scipy\stats\stats.py:1394: UserWarning:
kurtosistest only valid for n>=20 ... continuing anyway, n=8
  "anyway, n=%i" % int(n))
C:\Users\hoops\Anaconda3\lib\site-packages\scipy\stats\stats.py:1394: UserWarning:
kurtosistest only valid for n>=20 ... continuing anyway, n=12
  "anyway, n=%i" % int(n))
```

In [2]:

```python
# -*- coding: utf-8 -*-
"""
Created on Mon Apr 15 19:49:07 2019
ME EN 2550
Homework 8
@author: Ryan Dalby
"""

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import statsmodels.api as sm

#B1
print("B1")
b1data = pd.DataFrame({"viscosity": [0.45, 0.20, 0.34, 0.58, 0.70, 0.57, 0.55, 0.44], "ratio":
plt.scatter(b1data['ratio'], b1data['viscosity'])
print("b)")
b1model = sm.OLS.from_formula(formula="viscosity ~ np.power(ratio,2)", data=b1data).fit()
b1predictions = b1model.predict(b1data['ratio'])
plt.plot(b1data['ratio'], b1predictions)
plt.show()
print("a)")
print(b1model.summary())
print()

#B2
print("\n\nB2")
age = [55,46,30,35,59,61,74,38,27,51,53,41,37,24,42,50,58,60,62,68,70,79,63,39,49]
severity = [50,24,46,48,58,60,65,42,42,50,38,30,31,34,30,48,61,71,62,38,41,66,31,42,40]
surg = [0,1,1,1,0,0,1,1,0,1,1,0,0,0,0,1,1,1,0,0,1,1,1,0,1]
anx = [2.1,2.8,3.3,4.5,2.0,5.1,5.5,3.2,3.1,2.4,2.2,2.1,1.9,3.1,3.0,4.2,4.6,5.3,7.2,7.8,7.0,6.2
sat = [68,77,96,80,43,44,26,88,75,57,56,88,88,102,88,70,52,43,46,56,59,26,52,83,75]
b2data = pd.DataFrame({"age": age, "severity":severity, "surgmed":surg, "anxiety":anx, "satisf
b2model = sm.OLS.from_formula(formula="satisfaction ~ age + severity + surgmed + anxiety", dat
print("a)")
print(b2model.summary())
print("b)")
print("Standard error of regression coefficents: \n{}".format(b2model.bse))
print("c) Not all the model parameters are estimated with the same precision.  This is because


#B3
print("\n\nB3")
y = [240,236,270,274,301,316,300,296,267,276,288,261]
x1 = [25,31,45,60,65,72,80,84,75,60,50,38]
x2 = [24,21,24,25,25,26,25,25,24,25,25,23]
x3 = [91,90,88,87,91,94,87,86,88,91,90,89]
x4 = [100,95,110,88,94,99,97,96,110,105,100,98]
b3data = pd.DataFrame({"y":y,"x1":x1,"x2":x2,"x3":x3,"x4":x4})
b3model = sm.OLS.from_formula(formula="y ~ x1 + x2 + x3 + x4", data=b3data).fit()
print("a)")
print(b3model.summary())
print("b)")
print("Standard error of regression coefficents: \n{}".format(b3model.bse))
print("Not all the model parameters are estimated with the same precision.  This is because al
```

```
print("c)")
b3predict = b3model.predict(exog = dict(x1 = 75, x2 = 24, x3 = 90, x4 = 98))
print("The predicted power consumption for a month with the given values is : {}\n\n\n\n\n\n".
```