Example Runs

Loading dataset and displaying the first ten rows

In [2]:
```
# specifying file name
file = './birthweight_low.xlsx'

# setting pandas print options
pd.set_option('display.max_rows', 500)
pd.set_option('display.max_columns', 500)
pd.set_option('display.width', 1000)

# reading the file into Python
birthweight = pd.read_excel (io = file)

# outputing the first ten rows of the dataset
birthweight.head(n = 10)
```

Out[2]:

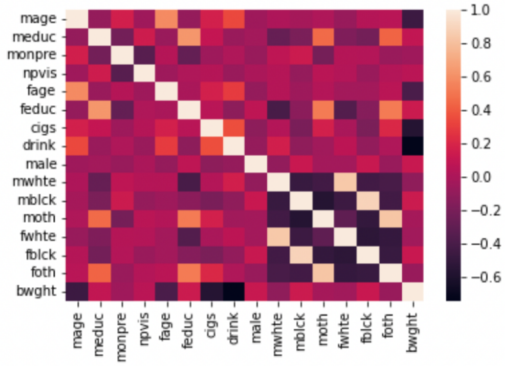
	mage	meduc	monpre	npvis	fage	feduc	omaps	fmaps	cigs	drink	male	mwhte	mblck	moth	fwhte	fblck	foth	bwght
0	69	NaN	5	2.0	62	NaN	4	7	23	9	1	0	1	0	0	1	0	697
1	68	12.0	3	10.0	61	11.0	4	6	25	11	1	1	0	0	1	0	0	1290
2	71	12.0	3	6.0	46	12.0	2	7	21	12	1	0	1	0	0	1	0	1490
3	59	16.0	1	8.0	48	16.0	7	8	21	10	0	0	0	1	0	0	1	1720
4	48	12.0	4	6.0	39	12.0	2	9	17	13	0	1	0	0	1	0	0	1956
5	67	11.0	4	8.0	40	8.0	4	9	16	14	0	1	0	0	1	0	0	1984
6	54	12.0	2	12.0	46	12.0	9	9	17	12	1	0	1	0	0	1	0	2050
7	71	14.0	4	7.0	51	11.0	9	8	15	13	0	1	0	0	1	0	0	2068
8	56	12.0	1	9.0	53	14.0	8	9	14	9	1	1	0	0	1	0	0	2148
9	58	12.0	2	12.0	61	16.0	9	9	13	6	0	0	1	0	0	1	0	2180

Pearson correlation and instantiating a heatmap

In [9]:
```
# linear (Pearson) correlation
birthweight_corr = birthweight.corr(method = 'pearson').round(decimals = 2)

# instantiating a heatmap
sns.heatmap(birthweight_corr)


# displaying the plot
plt.show()
```



Performing OLS Regression

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  bwght   R-squared:                       0.706
Model:                            OLS   Adj. R-squared:                  0.695
Method:                 Least Squares   F-statistic:                     64.57
Date:                Mon, 12 Dec 2022   Prob (F-statistic):           1.12e-46
Time:                        14:45:04   Log-Likelihood:                -1426.0
No. Observations:                 196   AIC:                             2868.
Df Residuals:                     188   BIC:                             2894.
Df Model:                           7
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept   2970.7332     66.826     44.455      0.000    2838.908    3102.559
mage         -14.4535      2.688     -5.376      0.000     -19.757      -9.150
cigs         -36.4272      4.610     -7.901      0.000     -45.522     -27.333
drink       -117.3954      9.729    -12.066      0.000    -136.588     -98.203
moth        1017.7862     69.182     14.712      0.000     881.313    1154.259
foth         979.6780     73.415     13.344      0.000     834.854    1124.502
mwhte       1026.4093     80.694     12.720      0.000     867.227    1185.591
fwhte        937.0330     66.384     14.115      0.000     806.080    1067.986
fblck       1054.0221     86.440     12.194      0.000     883.505    1224.540
mblck        926.5377     77.920     11.891      0.000     772.828    1080.248
==============================================================================
Omnibus:                        5.189   Durbin-Watson:                   1.278
Prob(Omnibus):                  0.075   Jarque-Bera (JB):                6.684
Skew:                          -0.139   Prob(JB):                       0.0354
Kurtosis:                       3.861   Cond. No.                     4.53e+17
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The smallest eigenvalue is 1.79e-30. This might indicate that there are
strong multicollinearity problems or that the design matrix is singular.
```

Generating training and testing scores for OLS Regression

```
OLS Training Score : 0.7238
OLS Testing Score  : 0.6667
OLS Train-Test Gap : 0.0571
```

Generating training and testing scores for Lasso Regression

```
Lasso Training Score : 0.7228
Lasso Testing Score  : 0.6758
Lasso Train-Test Gap : 0.047
```

Generating training and testing scores for ARD Regression

```
ARD Training Score: 0.7233
ARD Testing Score : 0.6749
ARD Train-Test Gap : 0.0484
```