

Syndicate 5 Statistical Learning Problem Set #1

May 19, 2021

Table 1: Correlation Matrix (numeric data only)

| | Income | Limit | Rating | Cards | Age | Balance |
|---------|--------|-------|--------|--------|-------|---------|
| Income | 1 | 0.792 | 0.791 | -0.018 | 0.175 | 0.464 |
| Limit | 0.792 | 1 | 0.997 | 0.010 | 0.101 | 0.862 |
| Rating | 0.791 | 0.997 | 1 | 0.053 | 0.103 | 0.864 |
| Cards | -0.018 | 0.010 | 0.053 | 1 | 0.043 | 0.086 |
| Age | 0.175 | 0.101 | 0.103 | 0.043 | 1 | 0.002 |
| Balance | 0.464 | 0.862 | 0.864 | 0.086 | 0.002 | 1 |

Table 2:

| Wald stat | p-value |
|-----------|---------|
| 3.487 | 0.062 |

Table 3:

| Wald stat | p-value |
|-----------|---------|
| 259.208 | 0 |

Table 4: Model Selection

| | <i>Dependent variable:</i> | | |
|-------------------------|----------------------------|---------------------------|----------------------------|
| | Balance | | |
| | (1) | (2) | (3) |
| Income | −7.816*** (0.235) | −6.312*** (0.486) | −6.238*** (0.486) |
| Limit | 0.189*** (0.033) | | |
| I(Income^2) | | −0.021*** (0.003) | −0.021*** (0.003) |
| Rating | 1.166** (0.491) | 2.482*** (0.136) | 2.471*** (0.136) |
| Cards | 17.760*** (4.346) | | |
| I(Rating^2) | | 0.002*** (0.0002) | 0.002*** (0.0002) |
| Age | −0.598** (0.295) | −0.702*** (0.262) | −0.729*** (0.261) |
| Edu_BinsBachelors | 6.884 (11.901) | 5.168 (10.579) | |
| Edu_BinsPost-Grad | −5.565 (12.329) | −10.154 (10.976) | |
| GenderFemale | −10.894 (9.925) | −9.313 (8.839) | |
| StudentYes | 426.109*** (16.774) | 428.894*** (14.914) | 428.341*** (14.755) |
| MarriedYes | −8.535 (10.361) | −12.452 (9.192) | |
| EthnicityAsian | 15.965 (14.157) | 20.495 (12.588) | |
| EthnicityCaucasian | 10.027 (12.217) | 13.910 (10.898) | |
| Constant | −497.360*** (29.040) | −330.930*** (29.526) | −329.576*** (26.542) |
| Observations | 400 | 400 | 400 |
| R ² | 0.955 | 0.964 | 0.964 |
| Adjusted R ² | 0.954 | 0.963 | 0.963 |
| Residual Std. Error | 98.853 (df = 387) | 88.052 (df = 387) | 88.218 (df = 393) |
| F Statistic | 686.991*** (df = 12; 387) | 874.252*** (df = 12; 387) | 1,740.723*** (df = 6; 393) |

Note:

*p<0.1; **p<0.05; ***p<0.01