Multivariate Assignment 3

* 1. PCA model built upon the first 100 rows of data.
  2. The scores shown indicate that the first component provides a massive amount of information.
  3. The outliers were investigated through the Hoteling’s T^2 and SPE graphs. The number of components were changed to find the most ideal way for the data to be explained by the features.
  4. Outliers were very evident in the raw data. With every iteration, some more outliers were evident. These were identified to be:  
       
     - Outlier 21 – G3, G6, G9 were lower, and had more variation than the general trend of the data. G8 was super high, which was also what made the point an outlier  
     - Outlier 23 – G7 was very high, and G3 and G8 was very low in the Hoteling’s T^2 plot. It was above the 99% line.  
     - Outlier 39 – This was the most evident outlier because everything was much higher compared to the rest of the dataset. This point definitely sticks out, but since the SPE is within the limit, it is very well explained by the features.  
     - Outlier 40 – G8 was much larger than the general trend of the data, all the other values are high too, but the Hoteling’s T^2 was very large.  
     - Outlier 49 – It is well explained by the model, but the Hoteling’s T^2 was unusually large, probably caused by the low value of G8  
     - Outlier 66 – G1, G3, G5, and G7 were much lower than the trend of data. This led the Hoteling’s to be above 95%.
  5. After all the points were excluded, and with the no. of features selected to 3, there are no more noticeable outliers in the SPE and the Hoteling’s T^2 graphs.   
       
      A screenshot of a cell phone

     Description automatically generated

A screenshot of a cell phone

Description automatically generated

2.)

|  |  |  |  |
| --- | --- | --- | --- |
| X-Space | W[1] | w (Regression) | Difference |
| Tin\_LDPE.csv | 0.082864 | 0.0777 | 0.005164122 |
| Tmax1\_LDPE.csv | 0.355936 | 0.336 | 0.019936106 |
| Tout1\_LDPE.csv | 0.159201 | 0.1492 | 0.010001277 |
| Tmax2\_LDPE.csv | 0.474321 | 0.4446 | 0.029721391 |
| Tout2\_LDPE.csv | 0.239899 | 0.2248 | 0.015098786 |
| Tcin1\_LDPE.csv | 0.025606 | 0.024 | 0.001606232 |
| Tcin2\_LDPE.csv | -0.14826 | -0.139 | -0.00926213 |
| z1\_LDPE.csv | -0.29534 | -0.2768 | -0.01853579 |
| z2\_LDPE.csv | -0.43161 | -0.4045 | -0.02710829 |
| Fi1\_LDPE.csv | 0.281443 | 0.2638 | 0.01764279 |
| Fi2\_LDPE.csv | 0.398787 | 0.3738 | 0.024987477 |
| Fs1\_LDPE.csv | -0.01473 | -0.0138 | -0.0009319 |
| Fs2\_LDPE.csv | -0.05317 | -0.0498 | -0.00336739 |
| Press\_LDPE.csv | 0.145376 | 0.1363 | 0.009076044 |

Since the difference is really small we can say the w1 regresses into U1 very well by the NIPALS algorithm.

3.)

|  |  |  |  |
| --- | --- | --- | --- |
| Training | calculated T1 | promvT1 | Difference |
| 1 | -0.152705846 | -0.14352654 | 0.009179307 |
| 2 | 2.952679908 | 2.769720844 | -0.182959064 |
| 3 | 0.150993148 | 0.141719075 | -0.009274072 |
| 4 | -1.386871176 | -1.29894897 | 0.087922211 |
| 5 | -0.95401734 | -0.89383602 | 0.060181319 |
| 6 | 1.314084339 | 1.235179444 | -0.078904895 |
| 7 | 2.653716881 | 2.48780435 | -0.165912531 |
| 8 | 1.05209734 | 0.982462462 | -0.069634878 |
| 9 | -1.515295718 | -1.42098028 | 0.094315441 |
| 10 | 0.151254312 | 0.145997724 | -0.005256587 |
| 11 | 1.482791136 | 1.389826309 | -0.092964826 |
| 12 | 0.900041233 | 0.842540062 | -0.057501171 |
| 13 | 2.479761143 | 2.325300575 | -0.154460568 |
| 14 | -0.568715131 | -0.53104902 | 0.037666111 |
| 15 | -0.983523819 | -0.92340601 | 0.060117814 |
| 16 | 0.624098461 | 0.588390183 | -0.035708278 |
| 17 | 2.987343196 | 2.802022618 | -0.185320578 |
| 18 | 0.275814528 | 0.258510982 | -0.017303545 |
| 19 | 3.310406076 | 3.104714899 | -0.205691178 |
| 20 | 2.827435116 | 2.651379642 | -0.176055475 |
| 21 | -2.640420798 | -2.47578104 | 0.164639755 |
| 22 | -0.826933347 | -0.77470422 | 0.052229125 |
| 23 | 0.829298159 | 0.776655935 | -0.052642223 |
| 24 | -1.744924268 | -1.63822785 | 0.106696413 |
| 25 | 1.669681271 | 1.566014651 | -0.10366662 |
| 26 | -0.243443323 | -0.22408951 | 0.019353817 |
| 27 | -1.746859639 | -1.64145714 | 0.105402504 |
| 28 | -0.986945516 | -0.92340031 | 0.063545205 |
| 29 | 0.436131288 | 0.409176667 | -0.026954621 |
| 30 | 2.294175814 | 2.152117965 | -0.142057849 |
| 31 | -0.836383496 | -0.78243621 | 0.053947286 |
| 32 | -1.467494695 | -1.37622408 | 0.091270612 |
| 33 | -3.849869526 | -3.61093601 | 0.238933518 |
| 34 | 1.430667044 | 1.341450328 | -0.089216717 |
| 35 | -1.541513581 | -1.44253004 | 0.098983538 |
| 36 | 3.962403704 | 3.716844825 | -0.245558878 |
| 37 | 2.170068117 | 2.03445844 | -0.135609677 |
| 38 | 2.835570209 | 2.661468347 | -0.174101861 |
| 39 | 0.204402715 | 0.192414212 | -0.011988503 |
| 40 | -1.2039221 | -1.1308836 | 0.073038502 |
| 41 | -1.763995028 | -1.65644801 | 0.107547013 |
| 42 | -0.200110943 | -0.18624321 | 0.013867735 |
| 43 | 0.051969446 | 0.047157731 | -0.004811715 |
| 44 | -2.727483114 | -2.55761185 | 0.169871262 |
| 45 | -0.089357658 | -0.08291053 | 0.006447125 |
| 46 | 0.603316017 | 0.565773985 | -0.037542033 |
| 47 | -1.411697595 | -1.32412769 | 0.087569908 |
| 48 | -1.756393767 | -1.64832523 | 0.108068532 |
| 49 | 3.804179036 | 3.56829784 | -0.235881197 |
| 50 | -0.679634766 | -0.63992071 | 0.039714057 |
| 51 | -1.68457478 | -1.5808531 | 0.103721676 |
| 52 | -2.462540506 | -2.31137792 | 0.15116259 |
| 53 | -3.406191898 | -3.19748437 | 0.20870753 |
| 54 | -4.622560265 | -4.33968063 | 0.282879634 |

We can see that the difference in the values is because the calculated weight is not unit vector.

4.)

|  |  |  |  |
| --- | --- | --- | --- |
| Y-Space | c [1] | c [1] (calculated) | Difference |
| Conv\_LDPE.csv | 0.454061 | 0.4541 | -3.93668E-05 |
| Mn\_LDPE.csv | -0.5054 | -0.5054 | -3.60141E-06 |
| Mw\_LDPE.csv | 0.132621 | 0.1326 | 2.05734E-05 |
| LCB\_LDPE.csv | 0.460872 | 0.4609 | -2.79763E-05 |
| SCB\_LDPE.csv | 0.520439 | 0.5204 | 3.86438E-05 |

In the Y – Space the difference is very close to zero, and therefore negligible.

5.)

|  |  |  |  |
| --- | --- | --- | --- |
| training | u [1] | u [1](calculated) | Difference |
| 1 | -0.52375801 | -0.523767322 | 9.31134E-06 |
| 2 | 2.846368647 | 2.846380751 | -1.2105E-05 |
| 3 | 0.589733966 | 0.589729023 | 4.94249E-06 |
| 4 | -2.00233455 | -2.002325821 | -8.7307E-06 |
| 5 | -0.93003561 | -0.930051729 | 1.61171E-05 |
| 6 | 1.538741007 | 1.538721957 | 1.90496E-05 |
| 7 | 2.187347234 | 2.18732316 | 2.40731E-05 |
| 8 | 0.352127446 | 0.352083969 | 4.34764E-05 |
| 9 | -0.90861079 | -0.908593924 | -1.6865E-05 |
| 10 | -0.39166446 | -0.391672437 | 7.97331E-06 |
| 11 | 1.380900662 | 1.380902109 | -1.4472E-06 |
| 12 | 0.99486667 | 0.99485266 | 1.40105E-05 |
| 13 | 2.953504788 | 2.953517274 | -1.2486E-05 |
| 14 | -0.09499928 | -0.094982547 | -1.6728E-05 |
| 15 | -1.24800443 | -1.248024996 | 2.05639E-05 |
| 16 | 1.809483536 | 1.809527574 | -4.4038E-05 |
| 17 | 2.294476938 | 2.294432015 | 4.4923E-05 |
| 18 | 0.431110153 | 0.431110644 | -4.9037E-07 |
| 19 | 2.782188212 | 2.782175886 | 1.23259E-05 |
| 20 | 3.172397923 | 3.172400786 | -2.8638E-06 |
| 21 | -2.13589873 | -2.135884592 | -1.414E-05 |
| 22 | -0.78537978 | -0.78535633 | -2.3453E-05 |
| 23 | 1.678113895 | 1.678108675 | 5.21991E-06 |
| 24 | -0.88077074 | -0.880716666 | -5.4076E-05 |
| 25 | 1.444865035 | 1.444825994 | 3.90411E-05 |
| 26 | 0.472235718 | 0.472275387 | -3.9669E-05 |
| 27 | -1.62797332 | -1.627971222 | -2.1015E-06 |
| 28 | -1.09578501 | -1.095802883 | 1.78733E-05 |
| 29 | 1.073767247 | 1.073785821 | -1.8574E-05 |
| 30 | 1.545649609 | 1.545659591 | -9.9815E-06 |
| 31 | -1.35434234 | -1.354342842 | 5.02805E-07 |
| 32 | -1.47435714 | -1.474365472 | 8.33203E-06 |
| 33 | -3.67437489 | -3.674388789 | 1.39009E-05 |
| 34 | 1.449579513 | 1.449563368 | 1.61445E-05 |
| 35 | -0.40265259 | -0.402626723 | -2.5867E-05 |
| 36 | 3.528866404 | 3.528820314 | 4.60895E-05 |
| 37 | 1.867038769 | 1.867007713 | 3.1056E-05 |
| 38 | 2.549485106 | 2.549478538 | 6.56754E-06 |
| 39 | -0.04899668 | -0.048999481 | 2.79597E-06 |
| 40 | -1.21085622 | -1.210856471 | 2.52866E-07 |
| 41 | -1.38122541 | -1.381214184 | -1.1224E-05 |
| 42 | -0.19940676 | -0.199394455 | -1.23E-05 |
| 43 | -0.21263273 | -0.212629504 | -3.2273E-06 |
| 44 | -2.80947853 | -2.80948604 | 7.51101E-06 |
| 45 | 0.229799909 | 0.22981057 | -1.0661E-05 |
| 46 | -0.1017418 | -0.101784095 | 4.22906E-05 |
| 47 | -0.44487417 | -0.44482931 | -4.4861E-05 |
| 48 | -1.62097294 | -1.620982657 | 9.71627E-06 |
| 49 | 3.065277384 | 3.065231478 | 4.59064E-05 |
| 50 | -1.0613426 | -1.061333054 | -9.5471E-06 |
| 51 | -1.99278957 | -1.99277238 | -1.7195E-05 |
| 52 | -2.8474671 | -2.847441112 | -2.5985E-05 |
| 53 | -3.80553216 | -3.805500263 | -3.1897E-05 |
| 54 | -4.96966741 | -4.969627957 | -3.9455E-05 |

Again, the difference is negligible, and shows how the NIPALS algorithm is pretty effective.

6.)

|  |  |  |  |
| --- | --- | --- | --- |
| Training | T[1] | u[1] | difference(t-u) |
| 1 | -0.14353 | -0.52376 | 0.380231 |
| 2 | 2.769721 | 2.846369 | -0.07665 |
| 3 | 0.141719 | 0.589734 | -0.44801 |
| 4 | -1.29895 | -2.00233 | 0.703386 |
| 5 | -0.89384 | -0.93004 | 0.0362 |
| 6 | 1.235179 | 1.538741 | -0.30356 |
| 7 | 2.487804 | 2.187347 | 0.300457 |
| 8 | 0.982462 | 0.352127 | 0.630335 |
| 9 | -1.42098 | -0.90861 | -0.51237 |
| 10 | 0.145998 | -0.39166 | 0.537662 |
| 11 | 1.389826 | 1.380901 | 0.008926 |
| 12 | 0.84254 | 0.994867 | -0.15233 |
| 13 | 2.325301 | 2.953505 | -0.6282 |
| 14 | -0.53105 | -0.095 | -0.43605 |
| 15 | -0.92341 | -1.248 | 0.324598 |
| 16 | 0.58839 | 1.809484 | -1.22109 |
| 17 | 2.802023 | 2.294477 | 0.507546 |
| 18 | 0.258511 | 0.43111 | -0.1726 |
| 19 | 3.104715 | 2.782188 | 0.322527 |
| 20 | 2.65138 | 3.172398 | -0.52102 |
| 21 | -2.47578 | -2.1359 | -0.33988 |
| 22 | -0.7747 | -0.78538 | 0.010676 |
| 23 | 0.776656 | 1.678114 | -0.90146 |
| 24 | -1.63823 | -0.88077 | -0.75746 |
| 25 | 1.566015 | 1.444865 | 0.12115 |
| 26 | -0.22409 | 0.472236 | -0.69633 |
| 27 | -1.64146 | -1.62797 | -0.01348 |
| 28 | -0.9234 | -1.09579 | 0.172385 |
| 29 | 0.409177 | 1.073767 | -0.66459 |
| 30 | 2.152118 | 1.54565 | 0.606468 |
| 31 | -0.78244 | -1.35434 | 0.571906 |
| 32 | -1.37622 | -1.47436 | 0.098133 |
| 33 | -3.61094 | -3.67437 | 0.063439 |
| 34 | 1.34145 | 1.44958 | -0.10813 |
| 35 | -1.44253 | -0.40265 | -1.03988 |
| 36 | 3.716845 | 3.528866 | 0.187978 |
| 37 | 2.034458 | 1.867039 | 0.16742 |
| 38 | 2.661468 | 2.549485 | 0.111983 |
| 39 | 0.192414 | -0.049 | 0.241411 |
| 40 | -1.13088 | -1.21086 | 0.079973 |
| 41 | -1.65645 | -1.38123 | -0.27522 |
| 42 | -0.18624 | -0.19941 | 0.013164 |
| 43 | 0.047158 | -0.21263 | 0.25979 |
| 44 | -2.55761 | -2.80948 | 0.251867 |
| 45 | -0.08291 | 0.2298 | -0.31271 |
| 46 | 0.565774 | -0.10174 | 0.667516 |
| 47 | -1.32413 | -0.44487 | -0.87925 |
| 48 | -1.64833 | -1.62097 | -0.02735 |
| 49 | 3.568298 | 3.065277 | 0.50302 |
| 50 | -0.63992 | -1.06134 | 0.421422 |
| 51 | -1.58085 | -1.99279 | 0.411936 |
| 52 | -2.31138 | -2.84747 | 0.536089 |
| 53 | -3.19748 | -3.80553 | 0.608048 |
| 54 | -4.33968 | -4.96967 | 0.629987 |

The difference here is quite huge, and therefore it shows that the correlation is low, and therefore will build a very poor prediction model.