Initial Data:

ID Race Age Menopause Smoke Grade ER

1 x5055 White [46,55) Post Never Smoked 1 1

2 x5110 Black [46,55) Pre Never Smoked 3 0

3 x5539 White [46,55) Pre Never Smoked 3 0

4 x5550 Black [27,46) Pre Current Smoker 1 1

5 x5561 White [46,55) Post Current Smoker 2 0

6 x5737 White [27,46) Post Former Smoker 2 1

ERorPR HER2 IHC Regimen NumCycles ToxCall

1 Y Borderline <NA> 2nd [1, 5) Y

2 N Positive Her2/ER- 2nd [1, 5) N

3 N Positive Her2/ER- 2nd [1, 5) Y

4 Y Positive LuminalB 2nd [5,18] Y

5 N Positive Her2/ER- 2nd [5,18] Y

6 Y Positive LuminalB 2nd [1, 5) N

Neutropenia Myalgia Neuropathy DoseInterval

1 N N Y Q3wks

2 N N N Q3wks

3 N N N Q3wks

4 N N Y Weekly

5 N N N Weekly

6 N N N Weekly

TotalWeeks FollowupStatus StagePre StageFinal

1 [ 8.29,12.29) LTFU IIIA IIB

2 [ 8.29,12.29) NED IIIB IIA

3 [ 8.29,12.29) Dead IV IV

4 [12.29,23.00] Dead IIIA I

5 [12.29,23.00] Dead IIIB IIIA

6 [ 1.00, 8.29) Dead IV IV

RespRegimens ResponseTaxane ResponseNonTaxane

1 PR PR PR

2 CR CR PR

3 PR PR PR

4 PR SD PR

5 CR CR SD

6 CR UE CR

Manova using all of the dose response variables:

(see fig)

Multivariate Tests: Smoke

Df test stat approx F num Df den Df Pr(>F)

Pillai 2 0.3472308 1.722741 20 164 0.034215 \*

Wilks 2 0.6810383 1.715202 20 162 0.035507 \*

Hotelling-Lawley 2 0.4268373 1.707349 20 160 0.036895 \*

Roy 2 0.2769690 2.271146 10 82 0.021130 \*

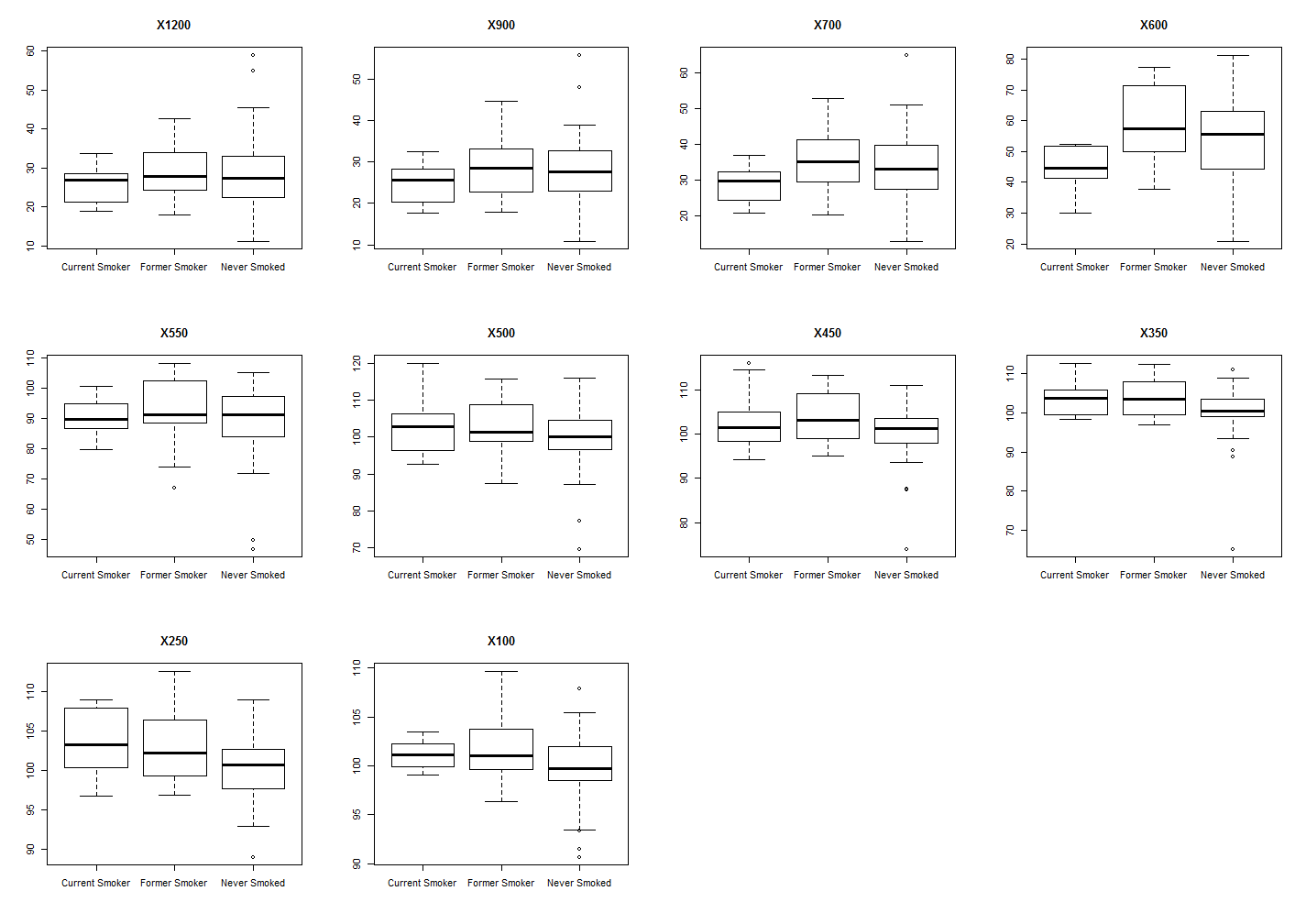
---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Pairwise contrasts:

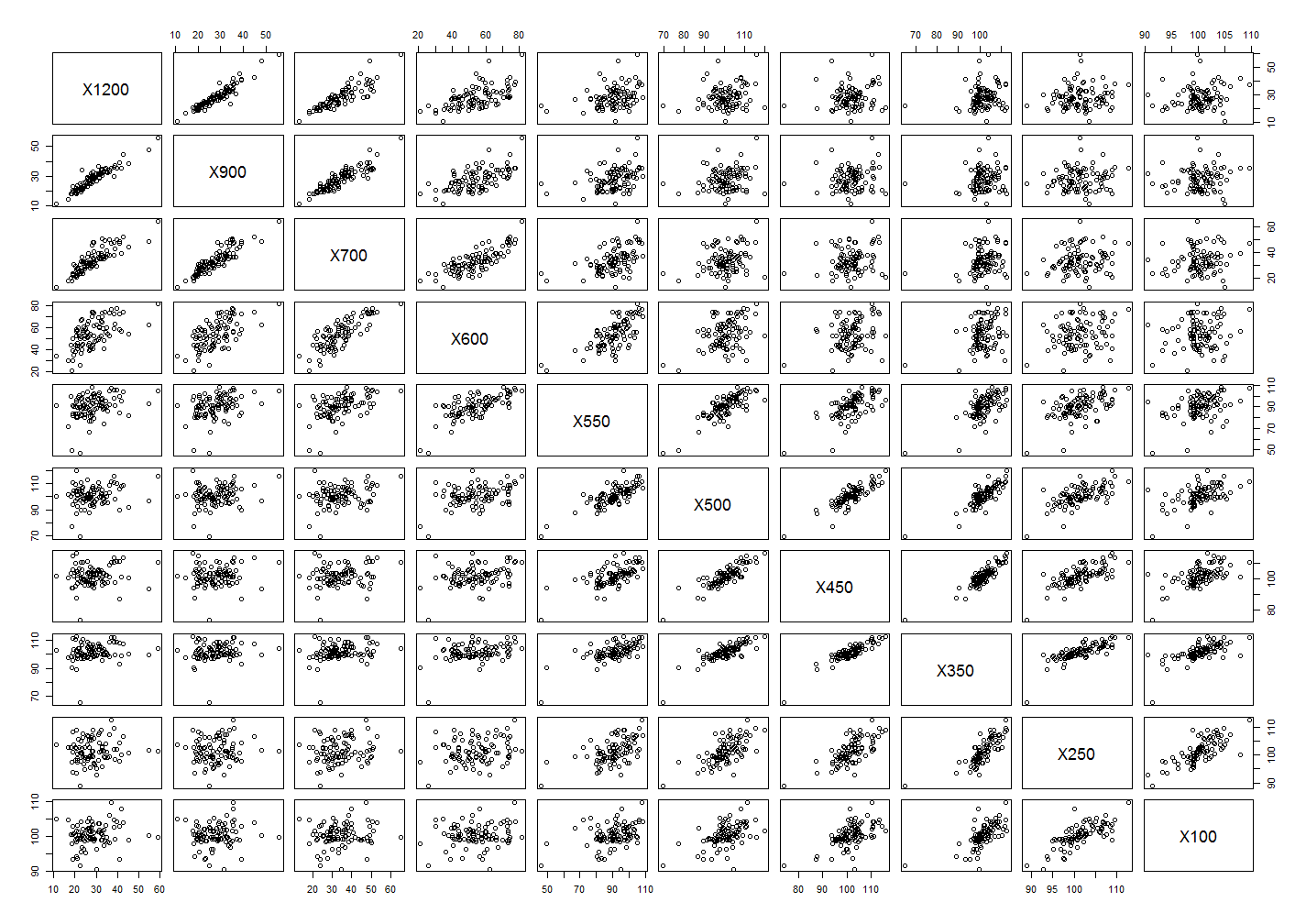
Used Bonferroni correction for multiple testing results in a significance level of .05/3 = .016



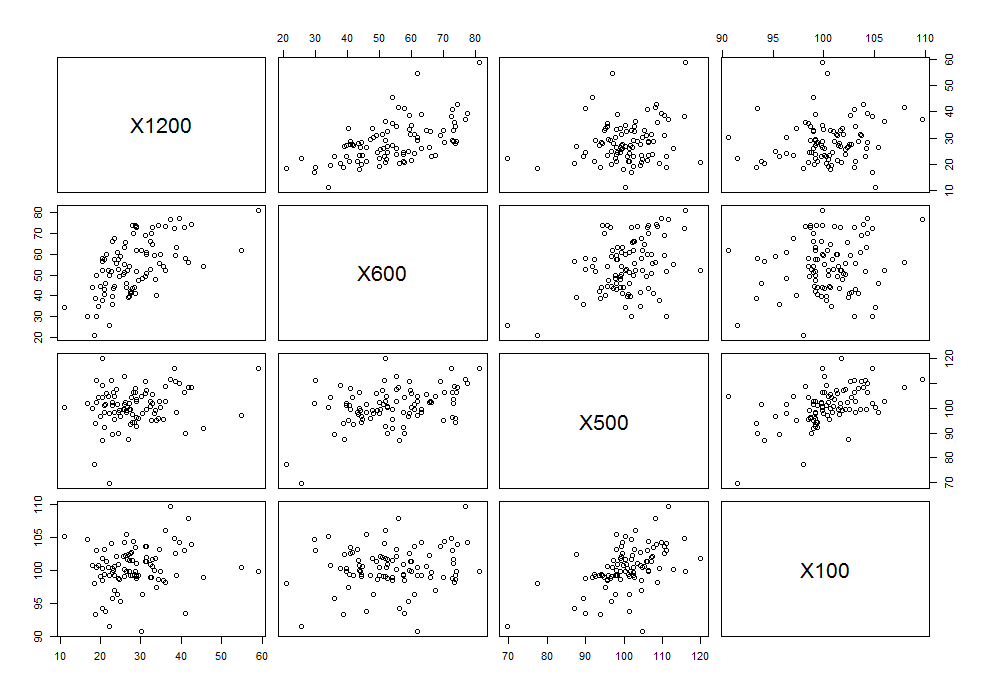




Manova using only the most uncorrelated dose response variables:



After distilled down to 4 of the least correlated variables:



New Correlation matrix:

X1200 X600 X500 X100

X1200 1.0000000 0.5735777 0.2050327 0.1588358

X600 0.5735777 1.0000000 0.4065645 0.1411192

X500 0.2050327 0.4065645 1.0000000 0.5181742

X100 0.1588358 0.1411192 0.5181742 1.0000000

(See fig)

Multivariate Tests: Smoke

Df test stat approx F num Df den Df Pr(>F)

Pillai 2 0.2308702 2.870984 8 176 0.0050083 \*\*

Wilks 2 0.7819378 2.846490 8 174 0.0053729 \*\*

Hotelling-Lawley 2 0.2624942 2.821812 8 172 0.0057659 \*\*

Roy 2 0.1603322 3.527309 4 88 0.0101953 \*

Bonferroni correction for multiple testing results in a significance level of .05/20 = .0025. This probably way too conservative though, because there are many variables that probably should not have been included in the test. And its Bonferroni.

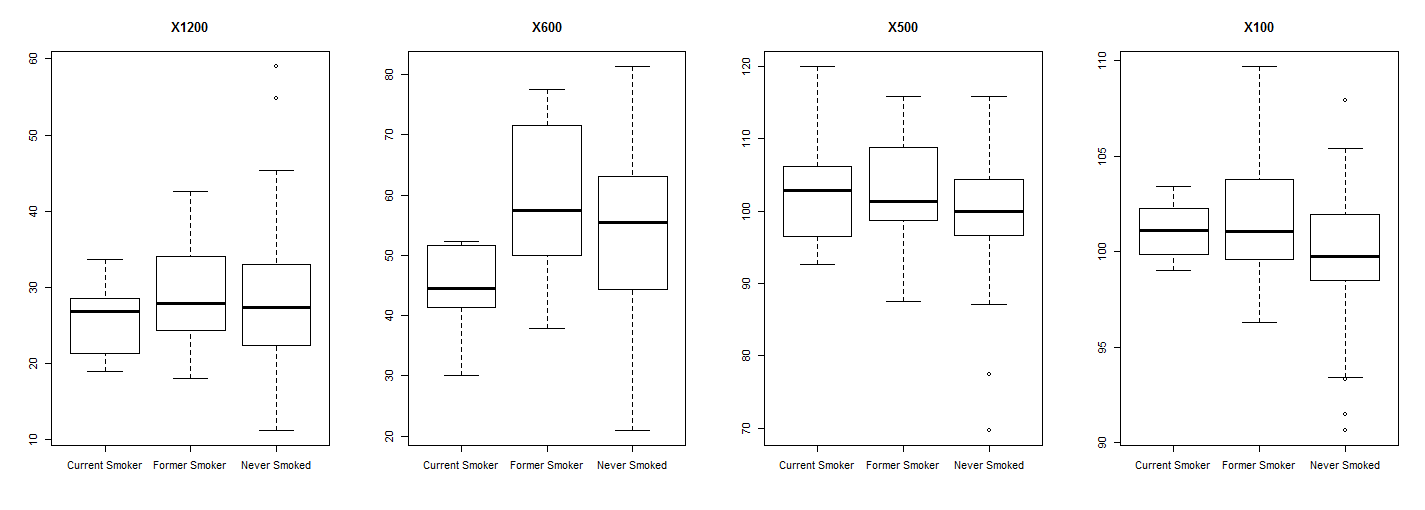
Contrasts:

Used Bonferroni correction for multiple testing results in a significance level of .05/3 = .016

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Contrast | X1200 | X600 | X500 | X100 | p-value |
| Former Smoker - Never Smoked | 0.17 | 4.15 | 3.53 | 1.9 | 0.072 |
| Former Smoker - Current Smoker | 3.5 | 13.78 | 0.23 | 0.41 | 0.072 |
| Never Smoked - Current Smoker | 3.33 | 9.62 | -3.3 | -1.48 | 0.013 |

~





Focus on X600:

> summary(aov(resp2[,2]~Smoke))

Df Sum Sq Mean Sq F value Pr(>F)

Smoke 2 1577 788.5 5.186 0.00739 \*\*

Residuals 90 13685 152.1

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> TukeyHSD(aov(resp2[,2]~Smoke))

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = resp2[, 2] ~ Smoke)

$Smoke

diff lwr upr p adj

Former Smoker-Current Smoker 13.776835 3.5816127 23.972057 0.0050232

Never Smoked-Current Smoker 9.624913 0.2466723 19.003153 0.0429193

Never Smoked-Former Smoker -4.151922 -11.0782118 2.774367 0.3306578

Only significant result from John Jack’s previous experiment:

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = nocurves$EMAX\_Score ~ nocurves$Smoke)

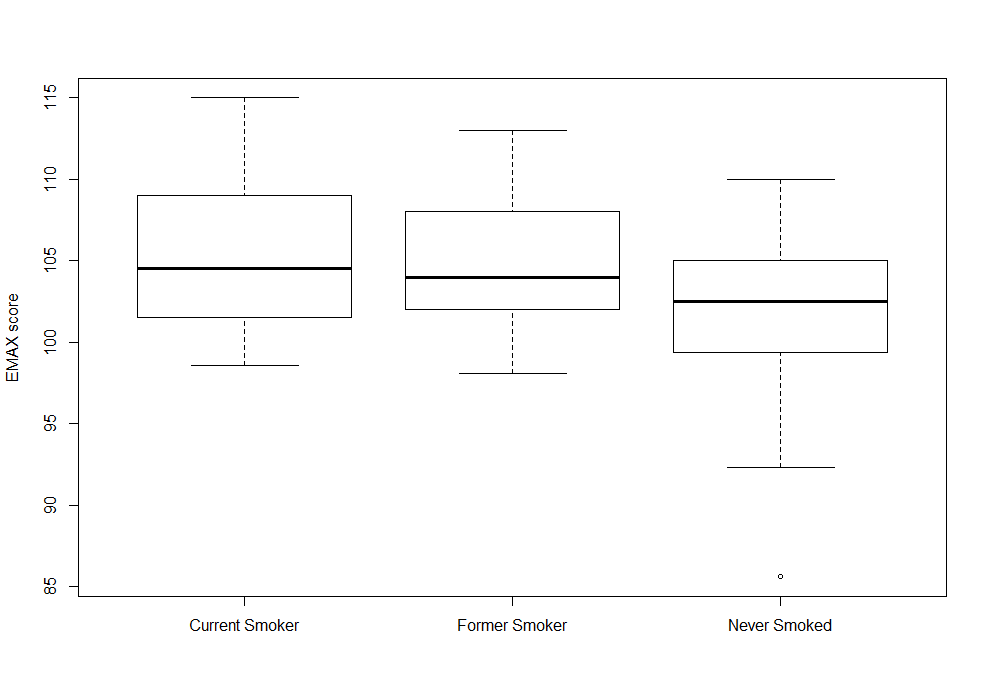
$`nocurves$Smoke`

diff lwr upr p adj

Former Smoker-Current Smoker -0.7222222 -4.458943 3.01449871 0.8897986

Never Smoked-Current Smoker -3.5092593 -6.946543 -0.07197593 0.0442884

Never Smoked-Former Smoker -2.7870370 -5.325639 -0.24843506 0.0278635



Conclusions:

1. The differences between the dose response profiles for people with different smoking status may not be significant if we correct for multiple testing. Pvalue: 0.034
2. When only one dose response variable is taken per group of highly correlated response variables, differences between dose response profiles will probably be significant after correction for multiple testing. Pvalue: 0.005.
3. For higher concentrations Non Smokers have higher viability than Current Smokers, while for lower concentrations, Non Smokers have lower viability. Former Smokers always have higher viability than Current Smokers, but may not have a significant difference from Non Smokers.
4. If we only look at the X600 response variable, Former Smokers and Non Smokers have higher viability than Current Smokers.

Future Directions

1. Accounting for repeated measurements in MANOVA
2. Include many grouping variables and do model selection (lasso for MANOVA)
3. Check joint normality of responses (small sample size so this is a concern)
4. Do other diagnostics on MANOVA
5. Perform Multidimensional Scaling on distance matrix that accounts for all variables used in Linear Regression to motivate machine learning approach
6. Machine learning (some tree base method that still accounts for variable importance)