MAT 443: HW 7

Name: Eric Agyemang

**Question 2**

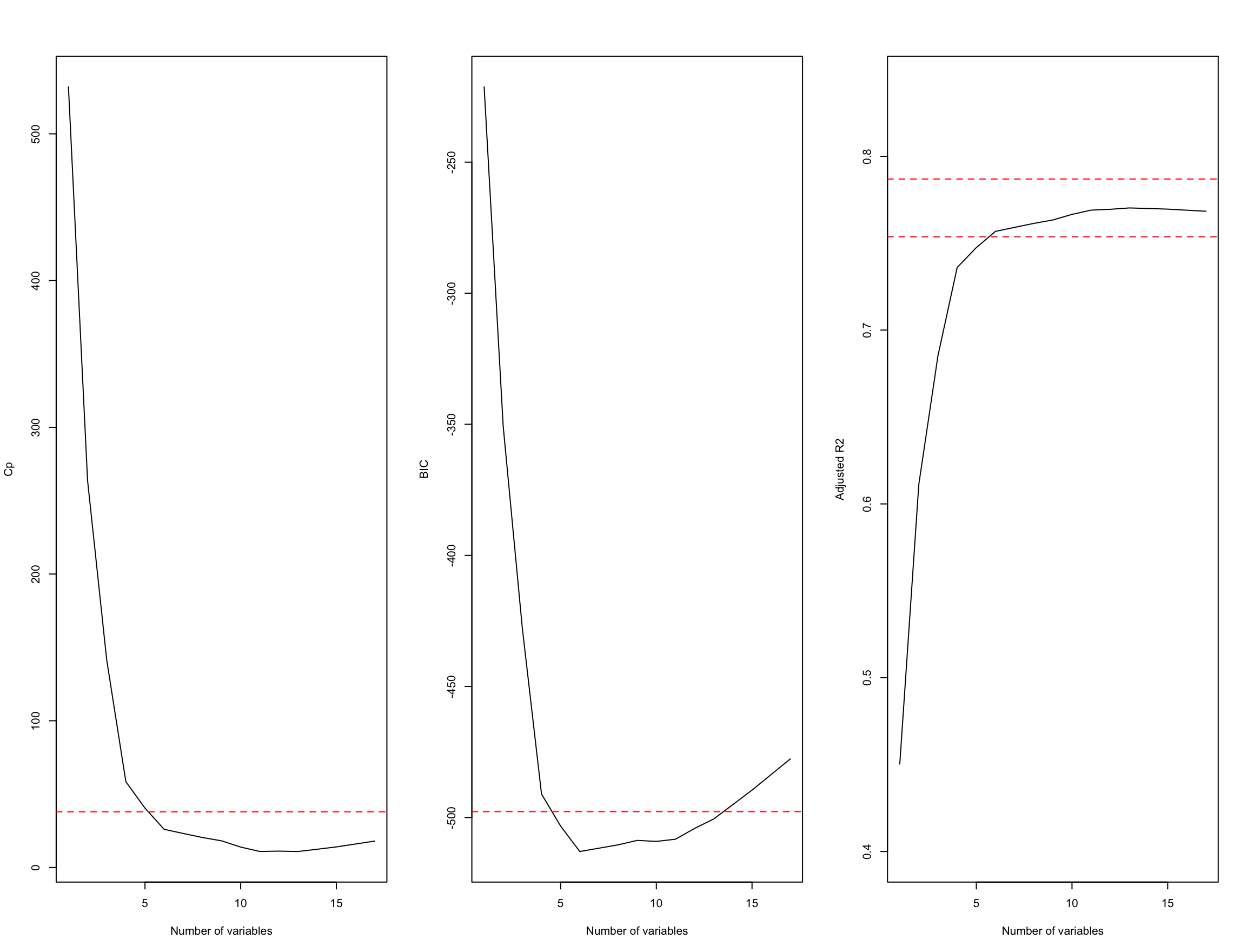
1. In this scenario because a large smoothing parameter
2. In this scenario because a large smoothing parameter
3. In this scenario because a large smoothing parameter
4. In this scenario because a large smoothing parameter
5. The penalty term doesn’t play any role, so in this case g is the interpolating spline

**Question 5**

1. The smoothing spline will probably have the smaller training RSS because it will be higher order polynomial due to the order the penalty term (it will be more flexible).
2. As mentioned above we expect to be more flexible, so it may overfit the data. It will probably be that have smaller test RSS.
3. If , we , so they will have the same training and test RSS.

**Question 10**

(a)



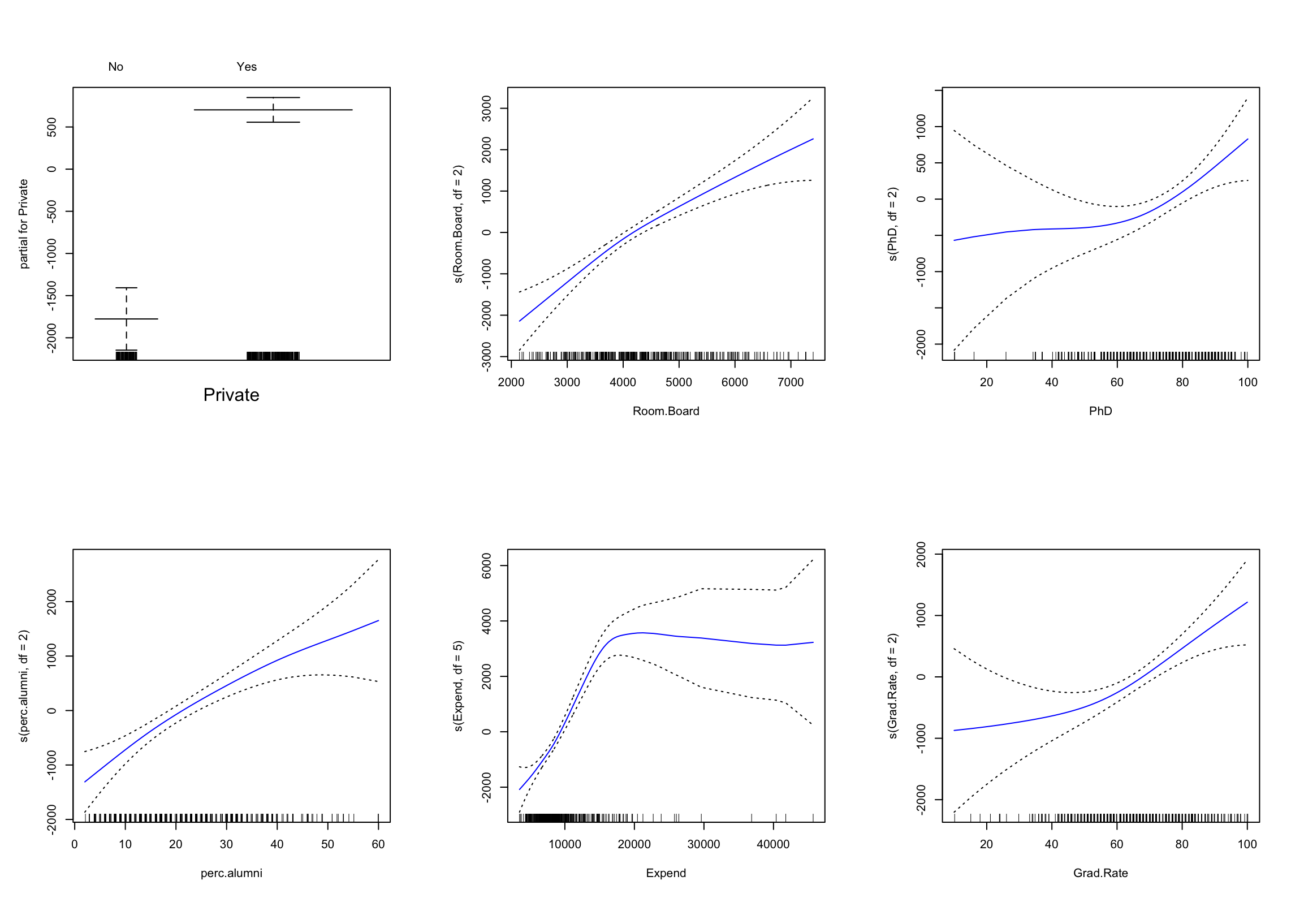
show that size 6 is the minimum size for the subset for which the scores are within 0.2 standard deviations of optimum.

[1] "(Intercept)" "PrivateYes" "Room.Board"

[4] "PhD" "perc.alumni" "Expend"

[7] "Grad.Rate"

(b)



(c)

1. [1] 3745460
2. [1] 0.7696916

(d)

Call: gam(formula = Outstate ~ Private + s(Room.Board, df = 2) + s(PhD,

df = 2) + s(perc.alumni, df = 2) + s(Expend, df = 5) + s(Grad.Rate,

df = 2), data = College.train)

Deviance Residuals:

Min 1Q Median 3Q Max

-4977.74 -1184.52 58.33 1220.04 7688.30

(Dispersion Parameter for Gaussian family taken to be 3300711)

Null Deviance: 6221998532 on 387 degrees of freedom

Residual Deviance: 1231165118 on 373 degrees of freedom

AIC: 6941.542

Number of Local Scoring Iterations: 2

Anova for Parametric Effects

Df Sum Sq

Private 1 1779433688

s(Room.Board, df = 2) 1 1221825562

s(PhD, df = 2) 1 382472137

s(perc.alumni, df = 2) 1 328493313

s(Expend, df = 5) 1 416585875

s(Grad.Rate, df = 2) 1 55284580

Residuals 373 1231165118

Mean Sq F value

Private 1779433688 539.106

s(Room.Board, df = 2) 1221825562 370.171

s(PhD, df = 2) 382472137 115.876

s(perc.alumni, df = 2) 328493313 99.522

s(Expend, df = 5) 416585875 126.211

s(Grad.Rate, df = 2) 55284580 16.749

Residuals 3300711

Pr(>F)

Private < 2.2e-16 \*\*\*

s(Room.Board, df = 2) < 2.2e-16 \*\*\*

s(PhD, df = 2) < 2.2e-16 \*\*\*

s(perc.alumni, df = 2) < 2.2e-16 \*\*\*

s(Expend, df = 5) < 2.2e-16 \*\*\*

s(Grad.Rate, df = 2) 5.232e-05 \*\*\*

Residuals

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Anova for Nonparametric Effects

Npar Df Npar F

(Intercept)

Private

s(Room.Board, df = 2) 1 3.5562

s(PhD, df = 2) 1 4.3421

s(perc.alumni, df = 2) 1 1.9158

s(Expend, df = 5) 4 16.8636

s(Grad.Rate, df = 2) 1 3.7208

Pr(F)

(Intercept)

Private

s(Room.Board, df = 2) 0.06010 .

s(PhD, df = 2) 0.03786 \*

s(perc.alumni, df = 2) 0.16715

s(Expend, df = 5) 1.016e-12 \*\*\*

s(Grad.Rate, df = 2) 0.05450 .

---

Signif. codes:

0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

ANOVA shows a strong evidence of non-linear relationship “Outstate” and “Expend”, and a moderately strong non-linear relationship (using p-value of 0.05) between “outstate” and “Grad.Rate” or “PHD”