\_solution-3.R

wickhamc

Thu Jan 29 12:34:52 2015

library(Sleuth3)
library(ggplot2)

# Q1

## 1

# head(ex0923)
qplot(AFQT, Income2005, data=ex0923)



Income and intelligence score appear to be positively related. The spread of Income appears to increase with increasing average income, an indication the constant spread assumption my be violated. A log transform of income may alleviate this problem. In general, income is very right skewed (more evidence to do a log transform) resulting in a few outliers with very high income.

qplot(Educ, Income2005, data=ex0923)



Education and income also appear to be positively related. It's really hard to evaluate spread because of the drastically different numbers of people in the education categories. I probably would do boxplots:

qplot(factor(Educ), Income2005, data=ex0923, geom = "boxplot")



Now it's obvious spread is also increasing with increasing Education, validating our suggestion to log transform income.

# qplot(AFQT, log(Income2005), data=ex0923) # things improve!
# qplot(factor(Educ), log(Income2005), data=ex0923, geom = "boxplot")

qplot(AFQT,Educ, data=ex0923)



There appears to be a positive relationship between education and intelligence score, although there is a lot of spread. For example, at very 8 or less years of education, AFQT scores are all below 25, but at 12 years of education we see scores over the entire range of AFQT. Again, the picture might be clearer with boxplots

qplot(factor(Educ), AFQT, data=ex0923, geom = "boxplot")



There seems to be a positive non-linear relationship between average intelligence score and years of education, the spread in intelligence score also seems smaller at the extremes of Education. Since, both education and intelligence score are entering our model as explantories we don't require an distributional assumptions and no transfomations are neccessary for them.

There are a few outlying people who are very highly educated ( > 20 years) but with low intelligence score

## 2

fit <- lm(log(Income2005)~Gender + Educ + AFQT, data=ex0923)
summary(fit)

##
## Call:
## lm(formula = log(Income2005) ~ Gender + Educ + AFQT, data = ex0923)
##
## Residuals:
## Min 1Q Median 3Q Max
## -7.091 -0.330 0.140 0.509 2.545
##
## Coefficients:
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.731212 0.102629 85.08 < 2e-16 \*\*\*
## Gendermale 0.624509 0.034175 18.27 < 2e-16 \*\*\*
## Educ 0.076951 0.008489 9.06 < 2e-16 \*\*\*
## AFQT 0.005914 0.000766 7.72 1.6e-14 \*\*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.866 on 2580 degrees of freedom
## Multiple R-squared: 0.21, Adjusted R-squared: 0.209
## F-statistic: 229 on 3 and 2580 DF, p-value: <2e-16

confint(fit)

## 2.5 % 97.5 %
## (Intercept) 8.529969 8.932454
## Gendermale 0.557496 0.691522
## Educ 0.060305 0.093596
## AFQT 0.004413 0.007415

## 3

exp(0.6245)

## [1] 1.867

exp(c(0.557, 0.691))

## [1] 1.745 1.996

There is convincing evidence that the median income in 2005 is not the same for males and females after accounting for education and intelligence score (p-value < 0.0001, t-test for different intercepts). It is estimated that the median income for males is 1.86 times that for females with the same education and intelligence score. With 95% confidence, the median income for males is between 1.74 and 1.99 times higher than for females with the same education and intelligence score. (Or, equivalently, With 95% confidence, the median income for males is between 74% and 99% higher than for females with the same education and intelligence score.)

*(Also acceptable)* There is convincing evidence that the mean log Income in 2005 is not the same for males and females after accounting for education and intelligence score (p-value < 0.0001, t-test for different intercepts). It is estimated that the mean log income for males is 0.62 units higher than females with the same education and intelligence score. With 95% confidence, the mean log income for males is between 0.56 and 0.69 units higher than females with the same education and intelligence score.

###Q2

qplot(log(Height),log(Force),data = ex0722,colour = Species)



fit <- lm(log(Force) ~ log(Height) \* Species, data = ex0722)
summary(fit)

##
## Call:
## lm(formula = log(Force) ~ log(Height) \* Species, data = ex0722)
##
## Residuals:
## Min 1Q Median 3Q Max
## -0.7668 -0.2851 -0.0231 0.2425 0.8882
##
## Coefficients:
## Estimate Std. Error t value
## (Intercept) -1.967 1.449 -1.36
## log(Height) 2.068 0.621 3.33
## SpeciesHemigrapsus nudus 2.486 1.761 1.41
## SpeciesLophopanopeus bellus -1.813 1.853 -0.98
## log(Height):SpeciesHemigrapsus nudus -1.660 0.789 -2.10
## log(Height):SpeciesLophopanopeus bellus 0.905 0.830 1.09
## Pr(>|t|)
## (Intercept) 0.1840
## log(Height) 0.0022 \*\*
## SpeciesHemigrapsus nudus 0.1675
## SpeciesLophopanopeus bellus 0.3354
## log(Height):SpeciesHemigrapsus nudus 0.0433 \*
## log(Height):SpeciesLophopanopeus bellus 0.2837
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.433 on 32 degrees of freedom
## Multiple R-squared: 0.795, Adjusted R-squared: 0.762
## F-statistic: 24.8 on 5 and 32 DF, p-value: 3.94e-10

unique(ex0722$Species)

## [1] Hemigrapsus nudus Lophopanopeus bellus Cancer productus
## Levels: Cancer productus Hemigrapsus nudus Lophopanopeus bellus

confint(fit)

## 2.5 % 97.5 %
## (Intercept) -4.9187 0.98411
## log(Height) 0.8039 3.33299
## SpeciesHemigrapsus nudus -1.0997 6.07255
## SpeciesLophopanopeus bellus -5.5879 1.96236
## log(Height):SpeciesHemigrapsus nudus -3.2671 -0.05313
## log(Height):SpeciesLophopanopeus bellus -0.7858 2.59617

ex0722$Species1 <- relevel(ex0722$Species, ref = "Hemigrapsus nudus")
fit <- lm(log(Force) ~ log(Height) \* Species1, data = ex0722)
summary(fit)

##
## Call:
## lm(formula = log(Force) ~ log(Height) \* Species1, data = ex0722)
##
## Residuals:
## Min 1Q Median 3Q Max
## -0.7668 -0.2851 -0.0231 0.2425 0.8882
##
## Coefficients:
## Estimate Std. Error t value
## (Intercept) 0.519 1.000 0.52
## log(Height) 0.408 0.487 0.84
## Species1Cancer productus -2.486 1.761 -1.41
## Species1Lophopanopeus bellus -4.299 1.528 -2.81
## log(Height):Species1Cancer productus 1.660 0.789 2.10
## log(Height):Species1Lophopanopeus bellus 2.565 0.735 3.49
## Pr(>|t|)
## (Intercept) 0.6073
## log(Height) 0.4079
## Species1Cancer productus 0.1675
## Species1Lophopanopeus bellus 0.0083 \*\*
## log(Height):Species1Cancer productus 0.0433 \*
## log(Height):Species1Lophopanopeus bellus 0.0014 \*\*
## ---
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.433 on 32 degrees of freedom
## Multiple R-squared: 0.795, Adjusted R-squared: 0.762
## F-statistic: 24.8 on 5 and 32 DF, p-value: 3.94e-10

confint(fit)

## 2.5 % 97.5 %
## (Intercept) -1.51793 2.556
## log(Height) -0.58337 1.400
## Species1Cancer productus -6.07255 1.100
## Species1Lophopanopeus bellus -7.41212 -1.186
## log(Height):Species1Cancer productus 0.05313 3.267
## log(Height):Species1Lophopanopeus bellus 1.06743 4.063

There is moderate evidence that the relationship between mean log force and log height is different for Hemigrapsus nudus compared to Cancer productus (t-test for equal slope, p-value = 0.04). It is estimated that a one unit increase in mean log force for Hemigrapsus nudus is associated with a 1.66 unit increase in log height *less* than Cancer productus. With 95% confidence, a one unit increase in mean log force for Hemigrapsus nudus is associated with between a 0.05 and 3.27 unit increase in log height *less* than Cancer productus.

There is no evidence that the relationship between mean log force and log height is different for Lophopanopeus bellus compared to Cancer productus (t-test for equal slope, p-value = 0.28). It is estimated that a one unit increase in mean log force for Lophopanopeus bellus is associated with a 0.9 unit increase in log height *more* than Cancer productus. With 95% confidence a one unit increase in mean log force for Lophopanopeus bellus is associated with between a 2.59 unit increase in log height *more* than Cancer productus.and a 0.79 unit increase in log height *less* than Cancer productus.

There is convincing evidence that the relationship between mean log force and log height is different for Lophopanopeus bellus compared to Hemigrapsus nudus (t-test for equal slope, p-value = 0.001). It is estimated that a one unit increase in mean log force for Lophopanopeus bellus is associated with a 2.57 unit increase in log height more than Hemigrapsus nudus. With 95% confidence a one unit increase in mean log force for Lophopanopeus bellus is associated with between a 1.06 and 4.06 unit increase in log height more than Hemigrapsus nudus.