STAT 512 Homework 3

Kathleen Wendt

02/18/2020

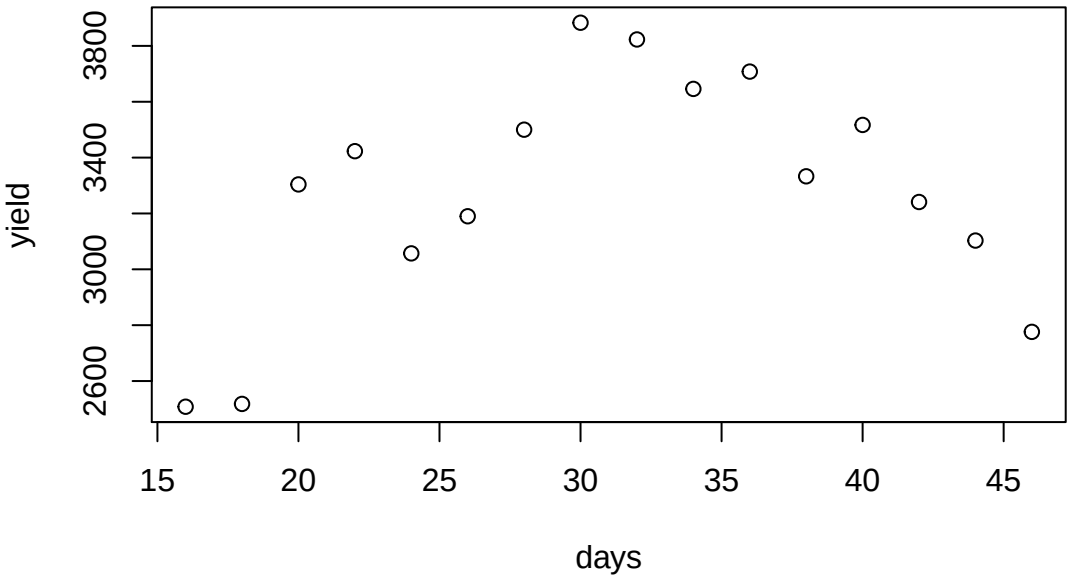
Questions 1 through 6 (Grain Yield): Data relating grain Yield (Y) to the number of Days (X) after flowering that harvesting took place was examined in "Determination of Biological Maturity and Effect of Harvesting and Drying Conditions on Milling Quality of Paddy" (J of Ag Engr. Research (1975):353-361.) The data is available from Canvas as "Grain.csv".

Notes:

- For consistency, please use the `I()` or `poly( , raw = TRUE)` functions for fitting the quadratic and cubic models.
- For questions 2-4, you do NOT need to include the diagnostic plots in your assignment. Just discuss your findings.

Question 1: Scatterplot

Create a scatterplot of Yield vs Days. Include this plot in your assignment.



Question 2: Simple linear regression

Fit a linear regression model of Yield on Days. Include the parameter estimate information ("Coefficients" table) in your assignment. Examine a plot of the residuals versus predicted values. What does the residual plot suggest? (4 pts)

```
##
## Call:
## lm(formula = yield ~ days, data = grain_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -691.07 -217.65   45.85  271.77  612.14
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2902.96     364.67   7.961 1.45e-06 ***
## days           12.26      11.28   1.088    0.295
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 415.8 on 14 degrees of freedom
## Multiple R-squared:  0.07791,    Adjusted R-squared:  0.01205
## F-statistic: 1.183 on 1 and 14 DF,  p-value: 0.2951
```

The plot of the residuals vs. predicted values for the simple linear regression of yield (Y) on days (X) revealed curvature, indicating a violation of the assumption of linearity.

Question 3: Quadratic regression model

Fit a quadratic regression model (including both linear and quadratic terms). Include the parameter estimate information ("Coefficients" table) in your assignment. Examine a plot of the residuals versus predicted values and comment. (4 pts)

```
##
## Call:
## lm(formula = yield ~ days + I(days^2), data = grain_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -303.96 -118.11   13.86  115.67  319.06
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1070.3977   617.2527  -1.734    0.107
## days          293.4829    42.1776   6.958 9.94e-06 ***
## I(days^2)      -4.5358     0.6744  -6.726 1.41e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 203.9 on 13 degrees of freedom
## Multiple R-squared:  0.7942, Adjusted R-squared:  0.7625
## F-statistic: 25.08 on 2 and 13 DF,  p-value: 3.452e-05
```

The plot of the residuals vs. predicted values for the quadratic regression of yield (Y) on days (X) with a quadratic term of days (X^2) showed an improvement in scatter, although non-constant variance might be an issue.

Appendix

```r
# load packages
library(tidyverse)
library(janitor)
library(broom)
library(car)
# set global options
knitr::opts_chunk$set(fig.width = 6,
                      fig.height = 4,
                      fig.path = "figs/",
                      echo = FALSE,
                      warning = FALSE,
                      message = FALSE)
# read grain data
grain_data <- readr::read_csv("data/Grain.csv") %>% janitor::clean_names()
# 1. pairwise plot (base) for grain data
plot(grain_data)
# 2. simple linear regression with grain data
grain_lm <- lm(yield ~ days, data = grain_data)
summary(grain_lm)
# 2. grain lin reg res vs. fitted plot
plot(grain_lm, which = 1)
# 3. grain quad reg model
grain_quadlm <- lm(yield ~ days + I(days^2), data = grain_data)
summary(grain_quadlm)
# 3. res vs fitted plot of grain quadratic model
plot(grain_quadlm, which = 1)
```