Stat 21 Homework 6

Problem 7 Solution

The data shown below present the average number of surviving bacteria in a canned food product and the minutes of exposure to 300 degree Fahrenheit heat. Use this data to answer Problems 6-7.

```
bacteria_data <- tibble(bacteria_count = c(175, 108, 95, 82, 71, 50, 49, 31, 28, 17, 16, 11),
                        minutes_exposure = c(1,2,3,4,5,6,7,8,9,10,11,12))
```

Problem 7

Identify an appropriate transformed model for these data. (Hint: Transform the response variable by taking the logarithm or taking the inverse and using this transformed data as the new response variable. Create residual plots for each transformation you try to select which method you want to use.)

(a) Fit a SLR model to the transformed data and display the residual plot and calculate the coefficient of determination to comment on the adequacy of this model.

(b) What is the average effect on the bacterial growth per each additional minute of exposure?

Solution Problem 7:

(a) The coefficient of determination is 0.9804 (or 0.9822 is also acceptable). The residual plot is shown below. With a coefficient of determination so close to one and with a residual plot that shows evenly scattered residuals, without any discernible pattern or funneling, we conclude that the linear model is appropriate for this transformed data.

For full credit (5 points) for part (a) the student must have a clearly labeled residual plot and must state the coefficient of determination and must provide a statement asserting the appropriateness of the model based on both of these factors.

(b) The average effect on the *logarithm* of bacterial growth per each additional minute of exposure is -0.236. This means that the average effect on bacterial growth is $e^{-0.236} = 0.790$ since $ln(a) = b$ is equivalent to $a = e^b$, where e is Euler's number and ln is the natural logarithm (i.e. log base e). Therefore, if -0.236 is the amount that the natural logarithm of the counts changes for an increase in time, then $e^{-0.236} = 0.790$ is the amount that the counts of bacteria change for an increase in time.

For full credit (5 points) for part (b) the student must have the correct estimate of 0.79 (or close with rounding) and must state that this is the amount the counts of the bacteria are expect to change by and must show some algebra explaining how they found this value (in R or in words). If the student only reports -0.236 but clearly states that this is the amount that the logarithm of the counts changes for an increase in time, then then should lose 2 points.

```
bacteria_data2 <- bacteria_data %>% mutate(log_count = log(bacteria_count))
mod <- lm(log_count ~ minutes_exposure, bacteria_data2)
summary(mod)
```

```
##
## Call:
## lm(formula = log_count ~ minutes_exposure, data = bacteria_data2)
##
## Residuals:
```

```
##       Min        1Q    Median        3Q       Max
## -0.184303 -0.083994  0.001453  0.072825  0.206246
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        5.33878    0.07409   72.05 6.47e-15 ***
## minutes_exposure -0.23617    0.01007  -23.46 4.49e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1204 on 10 degrees of freedom
## Multiple R-squared:  0.9822, Adjusted R-squared:  0.9804
## F-statistic: 550.3 on 1 and 10 DF,  p-value: 4.489e-10
```

```
mydata_all <- bacteria_data2 %>% mutate(mod_residuals = mod$residuals,
                                        mod_fitted = mod$fitted.values)

ggplot(mydata_all, aes(x=mod_fitted, y=mod_residuals)) +
  geom_point() +
  labs(title="Residual plot",
       y='Residuals',
       x='Fitted Values')
```