Andrew Goldberg

IS606 Ch 8 HW

Graded: 8.2, 8.4, 8.8, 8.16, 8.18

8.2 Baby weights, Part II. Exercise 8.1 introduces a data set on birth weight of babies. Another variable we consider is parity, which is 0 if the child is the first born, and 1 otherwise. The summary table below shows the results of a linear regression model for predicting the average birth weight of babies, measured in ounces, from parity.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 120.07 | 0.60 | 199.94 | 0.000 |
| parity | -1.93 | 1.19 | -1.62 | 0.1052 |

(a) Write the equation of the regression line.

**Birth weight = 120.07 – 1.93 (parity)**

(b) Interpret the slope in this context, and calculate the predicted birth weight of first borns and

others.

**The birth weight of the baby is estimated to be 1.93 lb’s lower if the baby is not first born.**

First born weight = **120.07**

Non-first born weight = 120.07 – 1.93 = **118.14**

(c) Is there a statistically significant relationship between the average birth weight and parity?

Ho: b1 = 0

Ha: b1 ≠ 0

**Because p-value is over .05 (or even .025), we fail to reject the null hypothesis**

8.4 Absenteeism, Part I. Researchers interested in the relationship between absenteeism from school and certain demographic characteristics of children collected data from 146 randomly sampled students in rural New South Wales, Australia, in a particular school year. Below are three observations from this data set.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Eth | Sex | Lrn | days |
| 1 | 0 | 1 | 1 | 2 |
| 2 | 0 | 1 | 1 | 11 |
| … | … | … | … | … |
| 146 | 1 | 0 | 0 | 37 |

The summary table below shows the results of a linear regression model for predicting the average number of days absent based on ethnic background (eth: 0 - aboriginal, 1 - not aboriginal), sex (sex: 0 - female, 1 - male), and learner status (lrn: 0 - average learner, 1 - slow learner).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Std. Error | t value | Pr(>|t|) |
| (Intercept) | 18.93 | 2.57 | 7.37 | 0.000 |
| eth | -9.11 | 2.60 | -3.51 | 0.000 |
| sex | 3.10 | 2.64 | 1.18 | .2411 |
| lrn | 2.15 | 2.65 | 0.81 | .4177 |

(a) Write the equation of the regression line.

**Average number of days absent = 18.93 - 9.11(ethnicity) + 3.10(sex) + 2.15(learn)**

(b) Interpret each one of the slopes in this context.

**All else held constant:**

**βeth: The model predicts a non-aborignal student will be absent 9 less days than an aboriginal student**

**βsex: The model predicts a student will be absent 3 more days if they are male than if they are female**

**βlrn: The model predicts a slower learn will be absent 2 more days than an average learner**

(c) Calculate the residual for the first observation in the data set: a student who is aboriginal, male, a slow learner, and missed 2 days of school.

Y 1  = 18.93 – 9.11(1) + 3.10(1) + 2.15(1) = **15.07**

Residual = 15.07 – 2days = **13.07**

(d) The variance of the residuals is 240.57, and the variance of the number of absent days for all students in the data set is 264.17. Calculate the R2 and the adjusted R2. Note that there are 146 observations in the data set.

R2 = 1 – (240.57/264.17) = **.089**

Adjusted R2 = 1 – (240.57/264.17)(145/142) = **.070**

8.8 Absenteeism, Part II. Exercise 8.4 considers a model that predicts the number of days absent using three predictors: ethnic background (eth), gender (sex), and learner status (lrn). The table below shows the adjusted R-squared for the model as well as adjusted R-squared values for all models we evaluate in the \_rst step of the backwards elimination process.

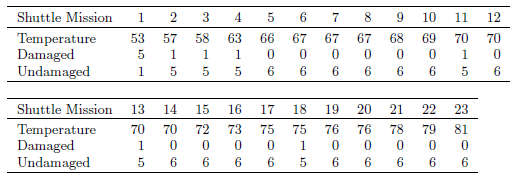
|  |  |  |
| --- | --- | --- |
|  | Model | Adjusted R2 |
| 1 | Full model | .0701 |
| 2 | No ethnicity | -.0033 |
| 3 | No sex | .0676 |
| 4 | No learn status | .0723 |

Which, if any, variable should be removed from the model first?

**Remove no learn status, because it has highest adjusted R2**

8.16 Challenger disaster, Part I. On January 28, 1986, a routine launch was anticipated for the Challenger space shuttle. Seventy-three seconds into the flight, disaster happened: the shuttle broke apart, killing all seven crew members on board. An investigation into the cause of the disaster focused on a critical seal called an O-ring, and it is believed that damage to these O-rings during a shuttle launch may be related to the ambient temperature during the launch. The table below summarizes observational data on O-rings for 23 shuttle missions, where the mission order

is based on the temperature at the time of the launch. Temp gives the temperature in Fahrenheit, Damaged represents the number of damaged O-rings, and Undamaged represents the number of O-rings that were not damaged.



(a) Each column of the table above represents a different shuttle mission. Examine these data and describe what you observe with respect to the relationship between temperatures and damaged O-rings.

**The lower temperature appears to create an environment where o-rings are more likely to be damaged, especially under roughly 65 degrees. Under 57ish appears to be very damaging.**

(b) Failures have been coded as 1 for a damaged O-ring and 0 for an undamaged O-ring, and

a logistic regression model was fit to these data. A summary of this model is given below.

Describe the key components of this summary table in words.



**The intercept is 11.66, which .2162 is subtracted from for each additional degree of temperature.**

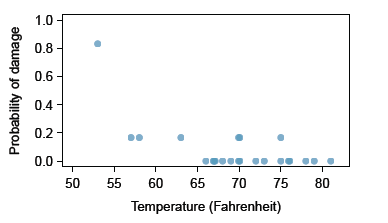
(c) Write out the logistic model using the point estimates of the model parameters.

**Chance of O-Ring damage = Log(pi/1-pi) = 11.6630 – .2162(temperature)**

(d) Based on the model, do you think concerns regarding O-rings are justified? Explain.

**Yes, at 53 degrees, there is a predicted damage rate of 55%, at 60 degrees it is 21%.** Those are very high numbers when lives are at stake.

8.18 Challenger disaster, Part II. Exercise 8.16 introduced us to O-rings that were identified as a plausible explanation for the breakup of the Challenger space shuttle 73 seconds into takeoff in 1986. The investigation found that the ambient temperature at the time of the shuttle launch was closely related to the damage of O-rings, which are a critical component of the shuttle. See this earlier exercise if you would like to browse the original data.



(a) The data provided in the previous exercise are shown in the plot. The logistic model fit to these data may be written as

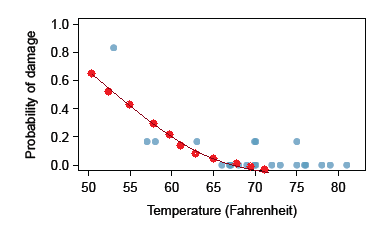


where ^p is the model-estimated probability that an O-ring will become damaged. Use the model to calculate the probability that an O-ring will become damaged at each of the following ambient temperatures: 51, 53, and 55 degrees Fahrenheit. The model-estimated probabilities for several additional ambient temperatures are provided below, where subscripts indicate the temperature:



@51: **.65**, @53: **.55**, @55: **.44**

(b) Add the model-estimated probabilities from part (a) on the plot, then connect these dots using a smooth curve to represent the model-estimated probabilities.



(c) Describe any concerns you may have regarding applying logistic regression in this application, and note any assumptions that are required to accept the model's validity.

**The actual data doesn’t appear linear, so the model is likely not optimal. There’s a curve – possibly linear – up until about 65 degrees, at which point it becomes flat. So each predictor isn’t linearly related.**

**Similarly, each outcome has to be independent of the other outcomes, so the residuals need to be evenly spaced out by order of collection. This would probably be fine.**