Comprehension Check: Logistic Regression

Q1

1/1 point (graded)

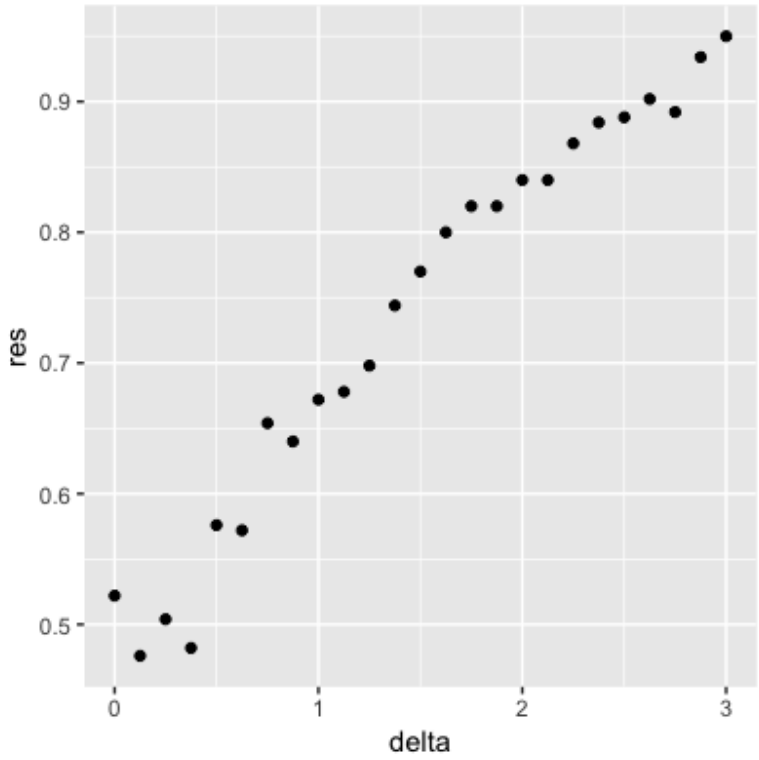
Define a dataset using the following code:

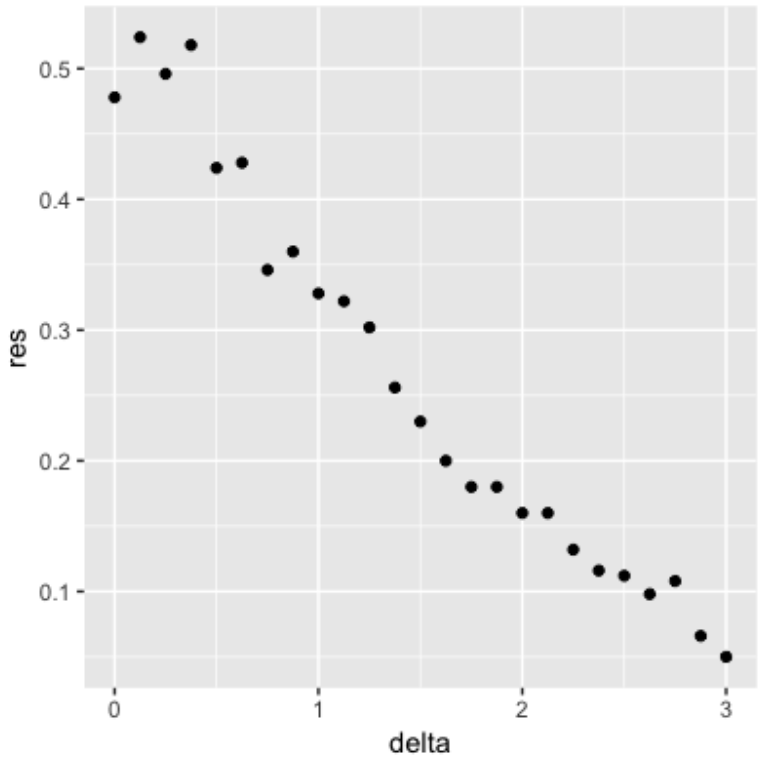
```
set.seed(2) #if you are using R 3.5 or earlier
set.seed(2, sample.kind="Rounding") #if you are using R 3.6 or later
make_data <- function(n = 1000, p = 0.5,
                                mu_0 = 0, mu_1 = 2,
                                sigma_0 = 1,  sigma_1 = 1){

y <- rbinom(n, 1, p)
f_0 <- rnorm(n, mu_0, sigma_0)
f_1 <- rnorm(n, mu_1, sigma_1)
x <- ifelse(y == 1, f_1, f_0)

test_index <- createDataPartition(y, times = 1, p = 0.5, list = FALSE)

list(train = data.frame(x = x, y = as.factor(y)) %>% slice(-test_index),
        test = data.frame(x = x, y = as.factor(y)) %>% slice(test_index))
}
dat <- make_data()
```

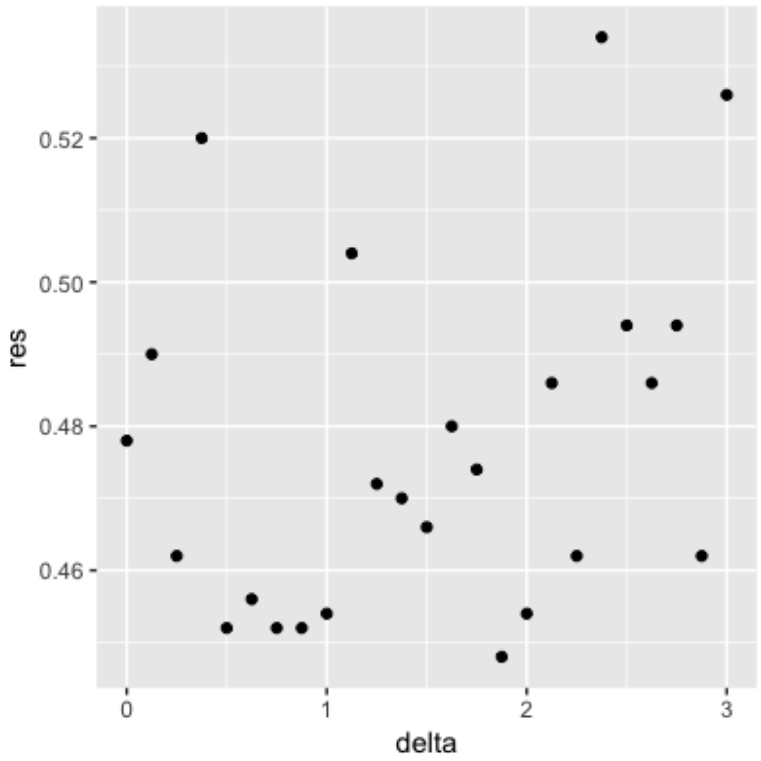
Note that we have defined a variable `x` that is predictive of a binary outcome `y`: `dat$train %>% ggplot(aes(x, color = y)) + geom_density()`.

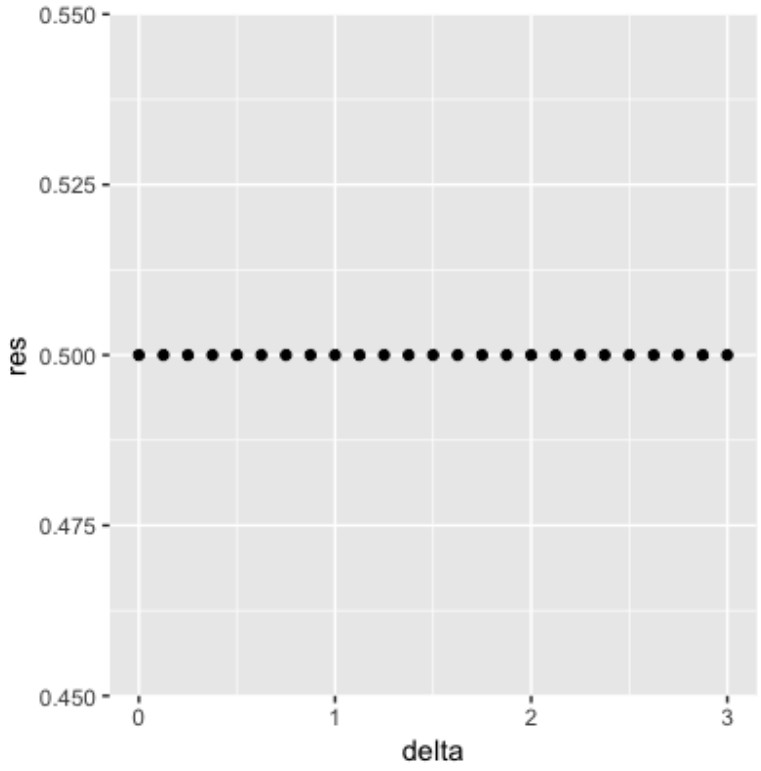
Set the seed to 1, then use the `make_data` function defined above to generate 25 different datasets with `mu_1 <- seq(0, 3, len=25)`. Perform logistic regression on each of the 25 different datasets (predict 1 if p>0.5) and plot accuracy (`res` in the figures) vs mu_1 (`delta` in the figures).”

Which is the correct plot?









✔

Explanation

The correct plot can be generated using the following code:

```
set.seed(1) #if you are using R 3.5 or earlier
set.seed(1, sample.kind="Rounding") #if you are using R 3.6 or later
delta <- seq(0, 3, len = 25)
res <- sapply(delta, function(d){
        dat <- make_data(mu_1 = d)
        fit_glm <- dat$train %>% glm(y ~ x, family = "binomial", data = .)
        y_hat_glm <- ifelse(predict(fit_glm, dat$test) > 0.5, 1, 0) %>% factor(levels =
c(0, 1))
        mean(y_hat_glm == dat$test$y)
})
qplot(delta, res)
```

Submit You have used 1 of 2 attempts

Answers are displayed within the problem

Ask your questions or make your comments about Logistic Regression here! **Remember, one of the best ways to reinforce your own learning is by explaining something to someone else, so we encourage you to answer each other's questions (without giving away the answers, of course).**

Some reminders:

- Search the discussion board before posting to see if someone else has asked the same thing before asking a new question
- Please be specific in the title and body of your post regarding which question you're asking about to facilitate answering your question.
- Posting snippets of code is okay, but posting full code solutions is not.
- If you do post snippets of code, please format it as code for readability. If you're not sure how to do this, there are instructions in a pinned post in the "general" discussion forum.

Discussion: Logistic Regression

Show Discussion

Topic: Section 3: Linear Regression for Prediction, Smoothing, and Working with Matrices / 3.1.2: Logistic Regression