565_HW6

Quetion 1

a

```
rem <- read.table("Data/remission.txt", header = TRUE)

tab <- with(rem, table(trt, rem))
chisq.test(tab)
```

```
## 
##  Pearson's Chi-squared test with Yates' continuity correction
## 
## data:  tab
## X-squared = 2.3233, df = 1, p-value = 0.1274
```

There is no significant difference at alpha $= .05$ without controlling for covariates.

b

```
library(ggplot2)

m1 <- glm(rem ~ trt*prog, family = binomial, data = rem)

m2 <- glm(rem ~ trt + prog, family = binomial, data = rem)
anova(m1, m2, test = "LRT")
```

```
## Analysis of Deviance Table
## 
## Model 1: rem ~ trt * prog
## Model 2: rem ~ trt + prog
##   Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
## 1       216     271.10
## 2       217     281.94 -1  -10.838 0.0009946 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The model with interaction is significantly better than the model without so we will use that going forward

1

```
newdat <- expand.grid(
  trt  = c("A", "B"),
  prog = seq(min(rem$prog), max(rem$prog), length.out = 100)
)

newdat$pr <- predict(m1, newdata = newdat, type = "response")

ggplot(rem, aes(prog, rem, colour = trt)) +
  geom_jitter(height = .03, width = 0) +
  geom_line(data = newdat, aes(prog, pr, colour = trt), size = 1) +
  labs(x = "Baseline cancer stage",
       y = "Pr(remission)",
       colour = "Treatment") +
  theme_classic()
```
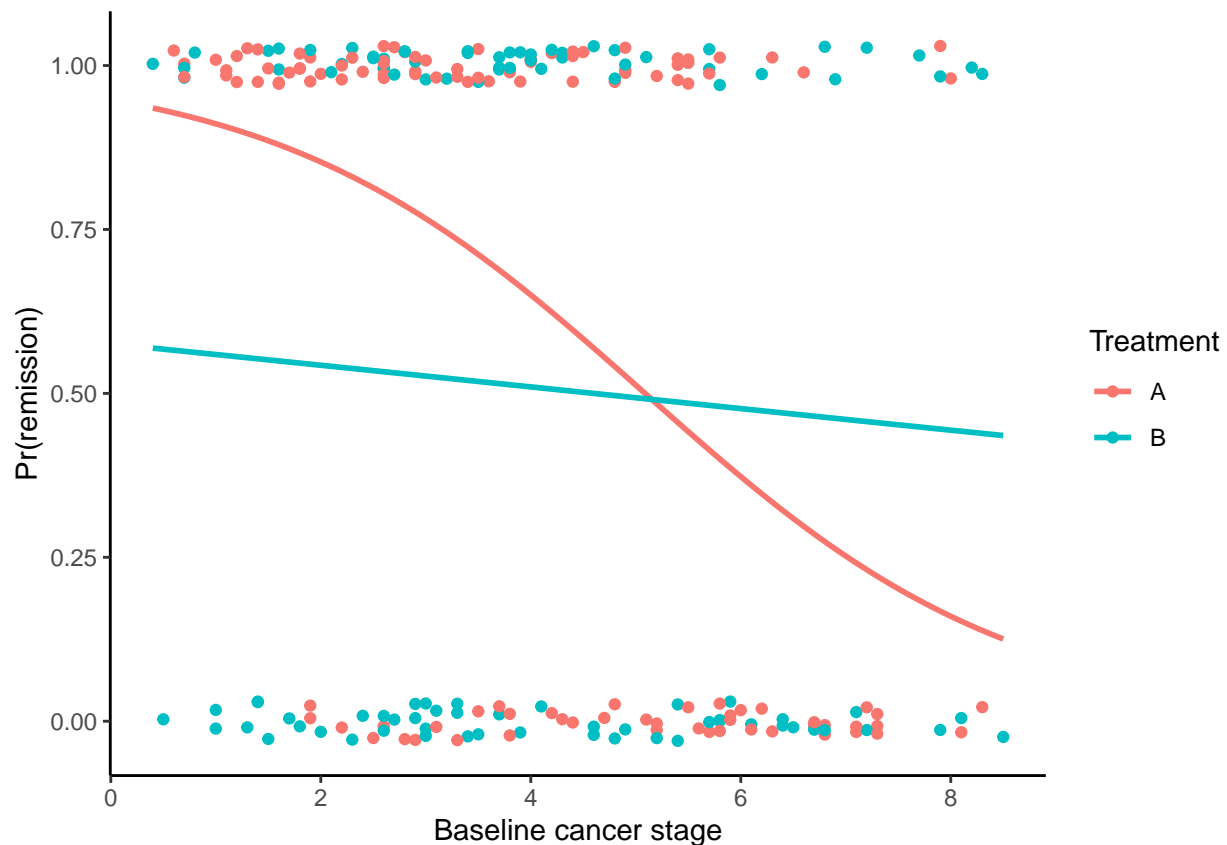
```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```


It appears that there is a strong interaction meaning that the treatment effect depends on the baseline stage of the cancer. For cancer stage 1-4, treatment A has a significantly higher remission probability than treatment b. Then after level 4, treatment b performs better than tretament A.

c

In this case, the covariate was added for the third reason, because it interacts with the main predictor of interest and must be added so that we can properly measure the effect.

Question 2

a

People who choose the Mediterranean diet may drink more red wine, and moderate wine is thought to lower cardiac risk. If we compare diets without adjusting for wine, the Mediterranean plan may look better just because its followers drink more wine.

b

The benefit of a Mediterranean diet might be substantial for non-drinkers but negligible (or even harmful) for heavy drinkers because high alcohol could offset diet gains. As such, the benfit of the diet changes based on the wine consumption indicating interaction.

Question 3

a

If Marketing rolls out the cartoon only in college-town papers where a large percentage of residents are students, and students already order more late-night pizza then higher sales might be due to student density, not the cartoon. We would need to control for this confounder

b

The cartoon mascot may resonate with students but annoy older customers. As such, in cities where students are a low percent of the population, ads with cartoons may decrease sales, but when there are many studetns, the same add may increase sales. The cartoon's effectiveness depends on the proportion of students in the market.