Reproducing Class Example

Shannon Houser

10/15/2021

```r
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.1 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.4
## v tibble  3.1.3     v dplyr   1.0.7
## v tidyr   1.1.3     v stringr 1.4.0
## v readr   2.0.0     v forcats 0.5.1
```

```
## -- Conflicts ------------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```r
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```r
ddh <- read_csv("ddh.csv")
```

```
## Rows: 572 Columns: 24
```

```
## -- Column specification ----------------------------------------------------
## Delimiter: ","
## dbl (24): ID, cold_ischemia, r_height, r_weight, r_bmi, r_age, r_caucasian, ...
```

```
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
# make categorical twa variable
ddh <- ddh %>%
  mutate(factor_glutwa = case_when(
    glutwa < 141.8 ~ "<142",
    glutwa < 172.1 ~ "142-172",
    glutwa < 209.4 ~ "172-209",
    TRUE ~ ">209"
  ))

ddh %>%
  count(factor_glutwa)
```

```
## # A tibble: 4 x 2
##   factor_glutwa     n
##   <chr>         <int>
## 1 <142            143
## 2 >209            143
## 3 142-172         143
## 4 172-209         143
```

```r
# used weird numbers as cutoffs but got the right number in each group?
```

```r
ddh %>%
  group_by(factor_glutwa) %>%
  summarise(med = round(median(r_bmi), 0),
            q1 = round(quantile(r_bmi, .25), 0),
            q3 = round(quantile(r_bmi, .75), 0))
```

```
## # A tibble: 4 x 4
##   factor_glutwa   med    q1    q3
##   <chr>         <dbl> <dbl> <dbl>
## 1 <142             28    24    32
## 2 >209             29    24    34
## 3 142-172          27    24    32
## 4 172-209          29    25    33
```

```r
ddh %>%
  ggplot(aes(y = glutwa, x = factor(delayed_fn), group = delayed_fn)) +
  stat_boxplot(geom ='errorbar', coef = 4) +
  geom_boxplot(coef = 4) +
  scale_y_continuous(trans = "log10",
                     limits = c(50,500),
                     labels = c(50, "", "", "", "", 100, 200, 300, 400, 500),
                     breaks = c(50, 60, 70, 80, 90, 100, 200, 300, 400, 500)) +
  scale_x_discrete(labels = c("IGF", "LGD")) +
  theme_minimal() +
  labs(x = "TWA",
       y = NULL)
```
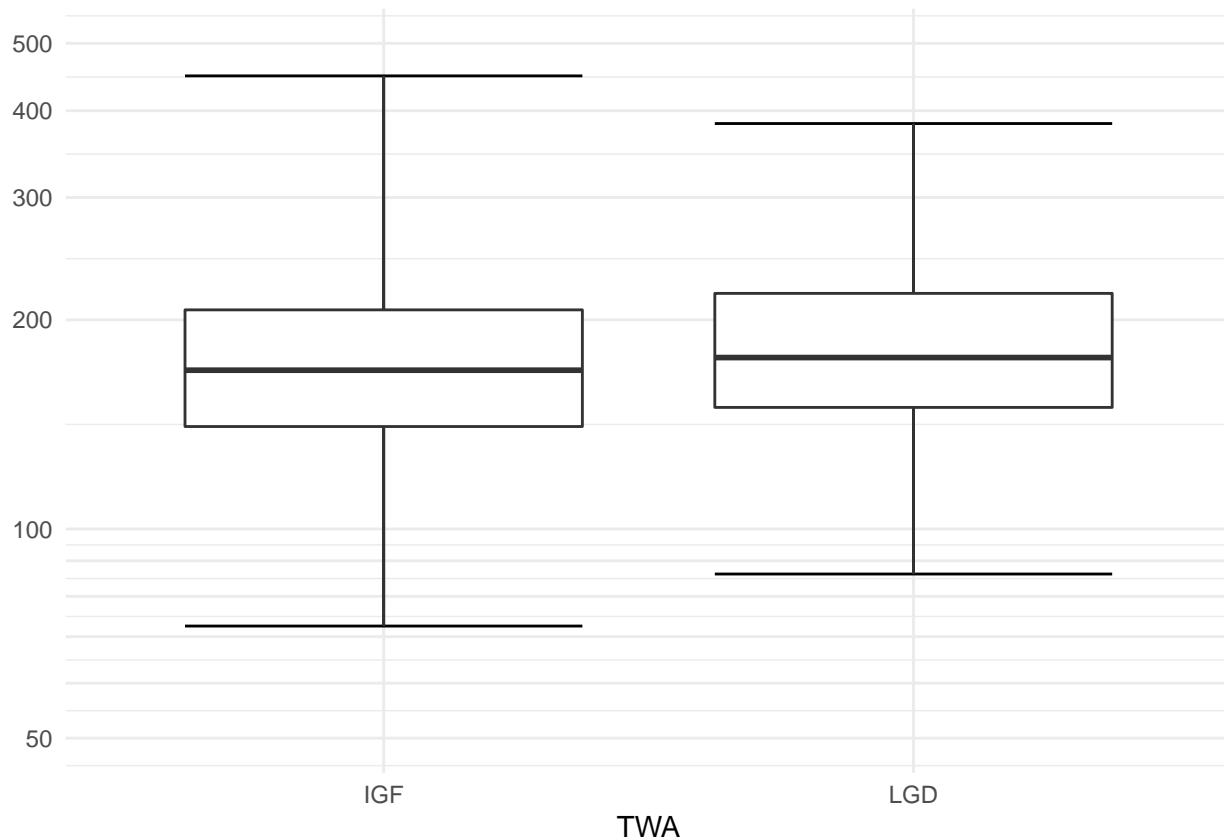
```
# boxplot without log scale and with outliers still
```

```
unadjusted <- glm(delayed_fn ~ glutwa, data = ddh, family = "binomial")
adjusted <- glm(delayed_fn ~ glutwa + r_age + d_cod + hemo_instability, data = ddh, family = "binomial")

summary(unadjusted)
```

```
##
## Call:
## glm(formula = delayed_fn ~ glutwa, family = "binomial", data = ddh)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0587  -0.7732  -0.7330   1.4102   1.7850
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.622352   0.330433  -4.910 9.12e-07 ***
## glutwa       0.002978   0.001716   1.735   0.0828 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 647.67  on 571  degrees of freedom
```

```
## Residual deviance: 644.71  on 570  degrees of freedom
## AIC: 648.71
##
## Number of Fisher Scoring iterations: 4
```

```r
summary(adjusted)
```

```
##
## Call:
## glm(formula = delayed_fn ~ glutwa + r_age + d_cod + hemo_instability,
##     family = "binomial", data = ddh)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1376  -0.7806  -0.7231   1.4226   1.9283
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)      -0.702449   0.613002  -1.146   0.2518
## glutwa            0.003079   0.001754   1.755   0.0792 .
## r_age            -0.013030   0.007961  -1.637   0.1017
## d_cod            -0.093601   0.117555  -0.796   0.4259
## hemo_instability -0.075574   0.259324  -0.291   0.7707
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 626.14  on 549  degrees of freedom
## Residual deviance: 620.26  on 545  degrees of freedom
##   (22 observations deleted due to missingness)
## AIC: 630.26
##
## Number of Fisher Scoring iterations: 4
```

```r
exp(cbind(coef(unadjusted), confint(unadjusted)))
```

```
## Waiting for profiling to be done...
```

```
##                            2.5 %     97.5 %
## (Intercept) 0.1974338 0.1027375 0.3764315
## glutwa      1.0029820 0.9995819 1.0063549
```

```r
exp(cbind(coef(adjusted), confint(adjusted)))
```

```
## Waiting for profiling to be done...
```

```
##                              2.5 %    97.5 %
## (Intercept)      0.4953708 0.1470238 1.635928
## glutwa           1.0030840 0.9996073 1.006533
## r_age            0.9870545 0.9718598 1.002807
## d_cod            0.9106460 0.7206511 1.143515
## hemo_instability 0.9272108 0.5637113 1.563730
```

```
# Got different p-values and odds ratios
# Not sure how to account for the odds ratios being for a relatice doubling of twa glucose
```