PHStatsII - HW4b

Marty Ross

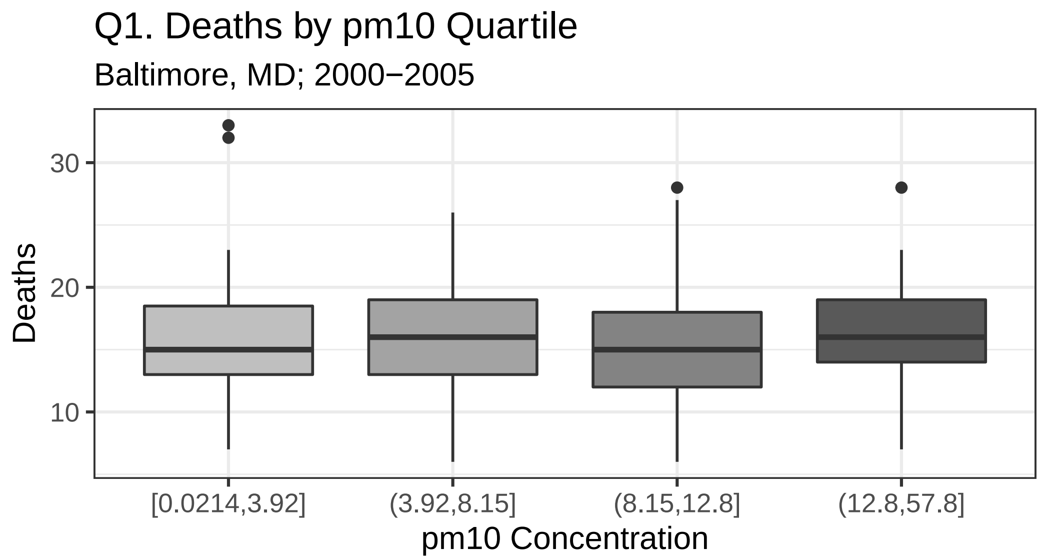
2022-10-16

Part A: What is the unadjusted relationship between daily mortality and PM 10 concentration?

```
library(tidyverse)
library(gtools)
library(kableExtra)
```

Q1. Side-by-side boxplots of the daily number of deaths in Baltimore during 2000-2005, comparing the quartiles of PM 10 concentration [Hint: see full code for how do split pm10 into quartiles in order to do this!]

```
baltPM <- read.csv('baltPM.csv')
baltPM$pm10Q <- quantcut(baltPM$pm10, q = c(0, 0.25, 0.5, 0.75, 1))
baltPM <- baltPM %>%
  group_by(pm10Q) %>%
  mutate(med_pm10Q = median(pm10)) #color by median on a gradient

box_Q1 <- ggplot(baltPM, aes(x = pm10Q, y = death, fill = med_pm10Q)) +
  geom_boxplot() +
  scale_fill_gradient(low = 'gray75', high = 'gray35') +
  theme_bw(base_size = 12) +
  theme(legend.position = 'none') +
  labs(title = 'Q1. Deaths by pm10 Quartile',
       subtitle = 'Baltimore, MD; 2000-2005',
       x = 'pm10 Concentration',
       y = 'Deaths')
box_Q1
```



After running a simple Poisson regression to relate the log incidence rate of mortality to PM10 concentration, report: Q2. The estimated incidence rate ratio of mortality for days that are 10 units higher PM10 concentration, compared to those that are 10 units lower

```
model1 <- glm(death ~ pm10 + offset(log(pop/100000)), family = poisson, data = baltPM)
# summary(model1)
co1 <- coefficients(model1)
conf1 <- confint(model1)
cat(paste0('Q2. The estimated incidence rate ratio of mortality for days that are 10 units higher\n',
           'PM10 concentration, compared to those that are 10 units lower:\n  ',
           'IRR = ', round(exp(co1[2]*10), 3),'\n'))
```

```
## Q2. The estimated incidence rate ratio of mortality for days that are 10 units higher
## PM10 concentration, compared to those that are 10 units lower:
##   IRR = 1.014
```

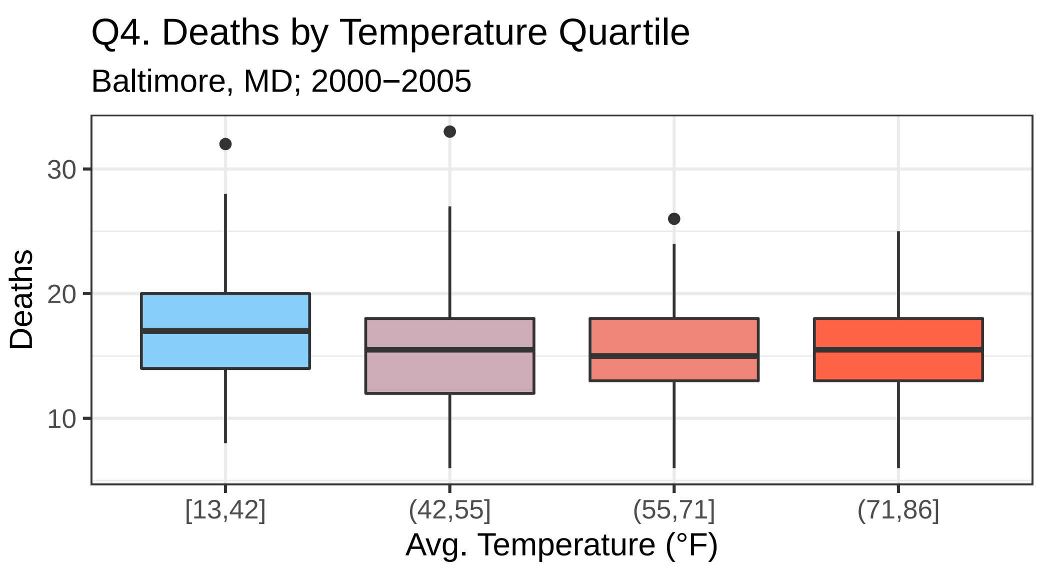
```
cat(paste0('Q3. A 95% confidence interval for the incidence rate ratio from Q2:\n  ',
           '95%CI: (',paste(round(exp(conf1[2,]*10), 3), collapse = ', '),')'))
```

```
## Q3. A 95% confidence interval for the incidence rate ratio from Q2:
##   95%CI: (0.981, 1.048)
```

Part B: What is the relationship between daily mortality and PM 10, when adjusted for temperature?

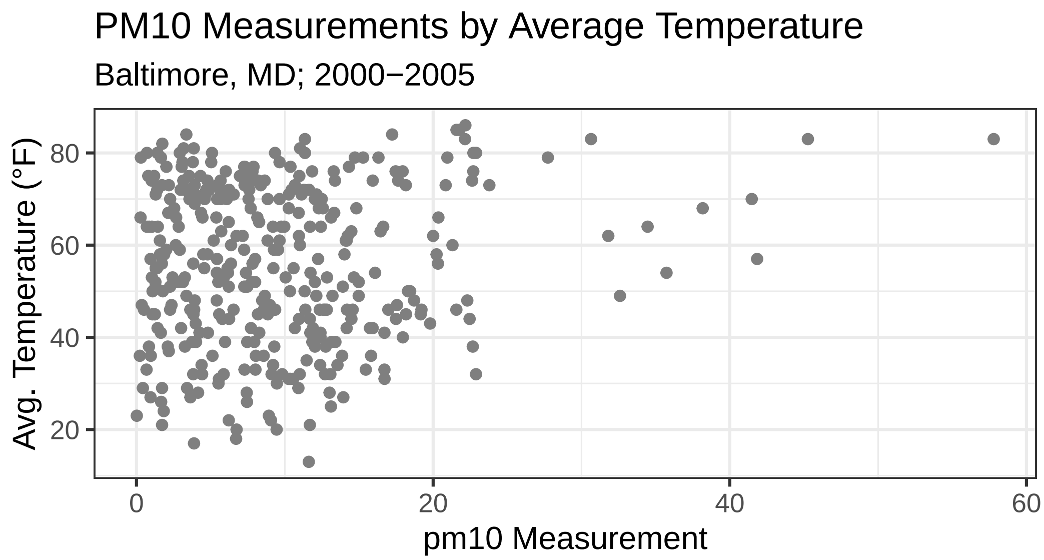
Q4. Side-by-side boxplots of the daily number of deaths in Baltimore during 2000-2005, comparing the quartiles of temperature [Hint: see full code for how do split tempF into quartiles in order to do this!]

```
baltPM$tempfQ <- quantcut(baltPM$tempf, q = c(0, 0.25, 0.5, 0.75, 1))
baltPM <- baltPM %>%
  group_by(tempfQ) %>%
  mutate(med_tempfQ = median(tempf)) #color by median on a gradient

box_Q2 <- ggplot(baltPM, aes(x = tempfQ, y = death, fill = med_tempfQ)) +
  geom_boxplot() +
  scale_fill_gradient(low = 'lightskyblue', high = 'tomato1') +
  theme_bw(base_size = 12) +
  theme(legend.position = 'none') +
  labs(title = 'Q4. Deaths by Temperature Quartile',
       subtitle = 'Baltimore, MD; 2000-2005',
       x = 'Avg. Temperature (°F)',
       y = 'Deaths')
box_Q2
```



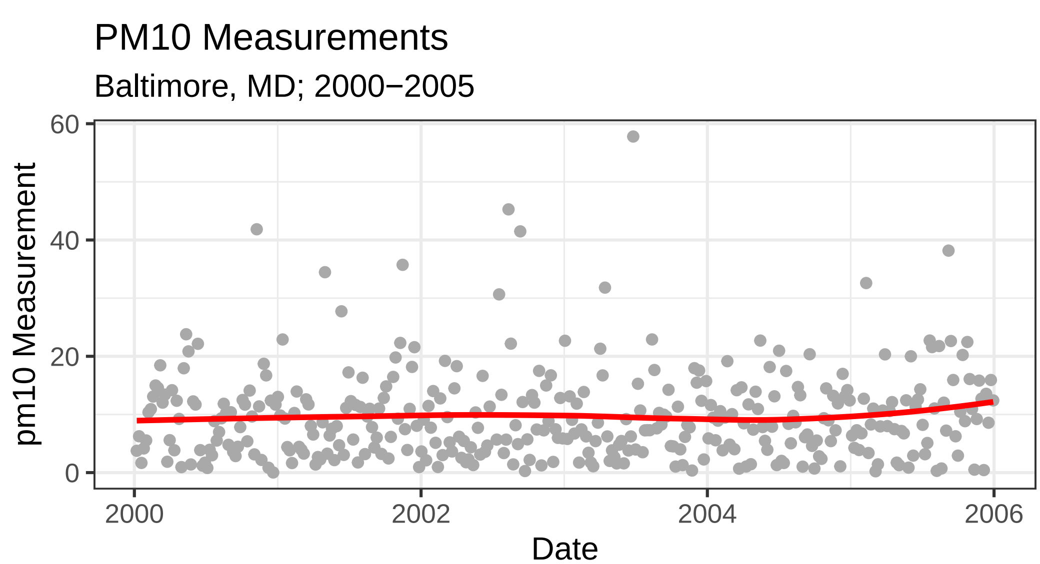
Q5. A scatterplot showing the relationship between temperature and PM 10 concentration

```
sct_temp <- ggplot(baltPM, aes(x = pm10, y = tempf)) +
  geom_point(color = 'gray50') +
  theme_bw(base_size = 12) +
  theme(legend.position = 'none') +
  labs(title = 'PM10 Measurements by Average Temperature',
       subtitle = 'Baltimore, MD; 2000-2005',
       x = 'pm10 Measurement',
       y = 'Avg. Temperature (°F)')

sct_temp
```



Bonus Plot! (because I'm curious) Not much of a trend as a function of time. A lot of stochasticity.

```
baltPM$dateDT <- as.Date(baltPM$date, "%m/%d/%y")
baltPM$test <- as.numeric(baltPM$dateDT)
sct_timepm10 <- ggplot(baltPM, aes(x = dateDT, y = pm10)) +
  geom_point(color = 'darkgray') +
  geom_smooth(method = loess, formula = y ~ x, span = 18, color = 'red', se = F) +
  theme_bw(base_size = 12) +
  theme(legend.position = 'none') +
  labs(title = 'PM10 Measurements',
       subtitle = 'Baltimore, MD; 2000-2005',
       x = 'Date',
       y = 'pm10 Measurement')

sct_timepm10
```



After running a multiple Poisson regression to relate the log incidence rate of mortality to PM 10 concentration and temperature, report:

```r
model2 <- glm(death ~ pm10 + tempf + offset(log(pop/100000)), family = poisson, data = baltPM)
# summary(model2)
co2 <- coefficients(model2)
conf2 <- confint(model2)

cat(paste0('Q6. The estimated incidence rate of mortality for days with an average temperature\n',
           'of 30 degrees F and a PM 10 concentration of 0 (expressed as deaths per 100,000\n',
           'person days)\n  Est. IR: ', round(exp(co2[1] + co2[2]*0 + co2[3]*30),3), '\n'))
```

```
## Q6. The estimated incidence rate of mortality for days with an average temperature
## of 30 degrees F and a PM 10 concentration of 0 (expressed as deaths per 100,000
## person days)
##   Est. IR: 2.597
```

```r
cat(paste0('Q7. The estimated incidence rate of mortality for days with an average temperature\n',
           'of 30 degrees F and a PM 10 concentration of 10 (expressed as deaths per 100,000\n',
           'person days)\n  Est IR: ', round(exp(co2[1] + co2[2]*10 + co2[3]*30),3), '\n'))
```

```
## Q7. The estimated incidence rate of mortality for days with an average temperature
## of 30 degrees F and a PM 10 concentration of 10 (expressed as deaths per 100,000
## person days)
##   Est IR: 2.651
```

```r
cat(paste0('Q8. The estimated incidence rate ratio of mortality for days that are 10 units higher\n',
           'PM10 concentration, compared to those that are 10 units lower but with the same average\n',
           'temperature)\n  Est IRR: ', round(exp(co2[2]*10),3), '\n'))
```

```
## Q8. The estimated incidence rate ratio of mortality for days that are 10 units higher
## PM10 concentration, compared to those that are 10 units lower but with the same average
## temperature)
##   Est IRR: 1.021
```

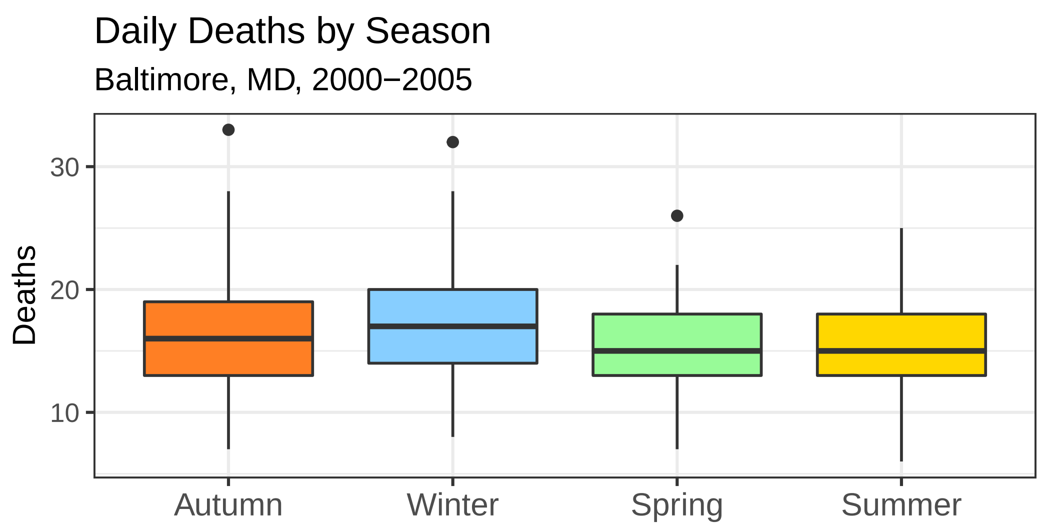
```r
cat(paste0('Q9. A 95% confidence interval for incidence rate ratio of mortality from Q8:\n  95%CI: (',
           paste(round(exp(conf2[2,]*10),3), collapse = ', '), ')\n'))
```

```
## Q9. A 95% confidence interval for incidence rate ratio of mortality from Q8:
##   95%CI: (0.987, 1.055)
```

Part C: What is the relationship between daily mortality and PM 10, when adjusted for season?

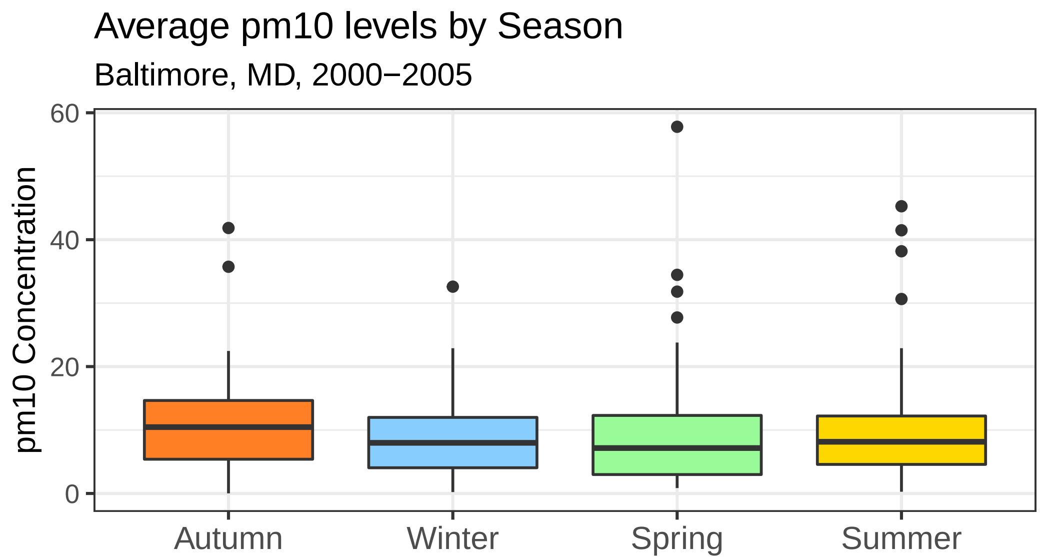
Q10. Side-by-side boxplots showing the difference in the daily number of deaths across the four seasons

```
baltPM$season <- factor(baltPM$season, levels = c('Autumn', 'Winter','Spring', 'Summer'))
bx_dthseas <- ggplot(data = baltPM, aes(x = season, y = death, fill = season)) +
  geom_boxplot() +
  labs(x = '', y = 'Deaths', title = 'Daily Deaths by Season', subtitle = 'Baltimore, MD, 2000-2005') +
  scale_fill_manual(values = c('Autumn' = 'chocolate1', 'Spring'='palegreen',
                    'Summer'='gold1', 'Winter'='skyblue1')) +
  theme_bw(base_size = 12) +
  theme(legend.position = 'none',
        axis.text.x = element_text(size=12))
bx_dthseas
```



Q11. Side-by-side boxplots showing the difference in PM 10 concentrations across the four seasons

```
bx_pm10seas <- ggplot(data = baltPM, aes(x = season, y = pm10, fill = season)) +
  geom_boxplot() +
  labs(title = 'Average pm10 levels by Season',
       subtitle = 'Baltimore, MD, 2000-2005',
       y = 'pm10 Concentration') +
  scale_fill_manual(values = c('Autumn' = 'chocolate1', 'Spring'='palegreen',
                    'Summer'='gold1', 'Winter'='skyblue1')) +
  theme_bw(base_size = 12) +
  theme(legend.position = 'none',
        axis.text.x = element_text(size=12),
        axis.title.x = element_blank())
bx_pm10seas
```



Q12. The estimated incidence rate of mortality for days with a PM 10 concentration of 0 for each of the four seasons (expressed as deaths per 100,000 person days).

```
model3 <- glm(death ~ pm10 + season + offset(log(pop/100000)), family = poisson, data = baltPM)
# summary(model3)
co3 <- coefficients(model3)
conf3 <- confint(model3)
```

```
irseas_df <- data.frame(Season = c('Autumn', 'Winter', 'Spring', 'Summer'),
                        `IR` = c(round(exp(co3[1]),2), round(exp(co3[1] + co3[3]*1),2),
                               round(exp(co3[1] + co3[4]*1),2), round(exp(co3[1] + co3[5]*1),2))
colnames(irseas_df) <- c('Season','Est. IR')

knitr::kable(irseas_df, format = 'latex', align = 'c',
             caption = 'Q12. IR of Mortality by Season') %>%
  kable_styling(latex_options = "hold_position")
```

Table 1: Q12. IR of Mortality by Season

Season	Est. IR
Autumn	2.48
Winter	2.64
Spring	2.33
Summer	2.33

```
cat(paste0('Q13. The estimated incidence rate ratio (and 95% confidence interval) of mortality for\n',
           'days in winter, compared to days in the autumn with the same PM 10 concentration:\n  ',
           'IR & 95%CI:\n  ', round(exp(co3[3]), 3), ' (',
           paste(round(exp(conf3[3, ]),3),collapse = ', '), ')\n'))
```

```
## Q13. The estimated incidence rate ratio (and 95% confidence interval) of mortality for
## days in winter, compared to days in the autumn with the same PM 10 concentration:
```

```
##   IR & 95%CI:
##   1.067 (0.993, 1.146)
```

```
cat(paste0('Q14. The estimated incidence rate ratio of mortality for days that are 10 units higher PM10'
           'concentration, compared to those that are 10 units lower but in the same season:\n  IRR: ',
           round(exp(co3[2]*10),3), '\n'))
```

```
## Q14. The estimated incidence rate ratio of mortality for days that are 10 units higher PM10
## concentration, compared to those that are 10 units lower but in the same season:
##   IRR: 1.017
```

```
cat(paste0('Q15. A 95% confidence interval for incidence rate ratio of mortality from Q14:\n  95%CI: ('
           paste(round(exp(conf3[2,]*10),3), collapse = ', '), ')'))
```

```
## Q15. A 95% confidence interval for incidence rate ratio of mortality from Q14:
##   95%CI: (0.984, 1.051)
```

Q16. Based on data gathered in the National Morbidity and Mortality Air Pollution Study (NMMAPS) between 2000-2005, it was observed that there was an estimated 6.7% increase in mortality (95%CI: -0.007%, 14.6%) in the city of Baltimore, MD in Winter as compared to Autumn when adjusted for PM10 concentration. As the 95% CI range contains 0, it was not determined to be significant at the 0.05 level.

Q17. See above