HS616 Data simulation Project

Rashmi Manjunath

May 9, 2017

References:

1. Joo, Young Min et al. "Impact Of Timely Antibiotic Administration On Outcomes In Patients With Severe Sepsis And Septic Shock In The Emergency Department". Clinical and Experimental Emergency Medicine 1.1 (2014): 35-40. Web. 9 May 2017.
2. Wan, Xiang et al. "Estimating The Sample Mean And Standard Deviation From The Sample Size, Median, Range And/Or Interquartile Range". BMC Medical Research Methodology 14.1 (2014): n. pag. Web. 9 May 2017.
3. "Logistic Regression - Simulation For A Power Calculation.". R-bloggers. N.p., 2010. Web. 8 May 2017.

Data simulation to show the impact of timely antibiotic administration on outcomes in patients with severe sepsis and septic shock in the emergency department

Background:

Severe sepsis and septic shock are severe illnesses caused by infection that can lead to death through progression of systemic inflammatory response syndrome and multiple organ failure. Mortality rates are reported to range from 20% to 50%, and with the recent steady rise in incidence rates, there is increasing emphasis on early and active treatment of sepsis. Proper treatment of the infection source is vital for effective and prompt treatment of patients with severe sepsis and septic shock. Early antibiotic treatment is a particularly crucial element. In patients with septic shock or severe sepsis, a delay in antibiotic administration from the time when hypotension first appears is significantly associated with increased mortality rates.

Introduction:

A retrospective study - *Impact of timely antibiotic administration on outcomes in patients with severe sepsis and septic shock in the emergency department* from Young Min Joo et. al aims to investigate the effect of timely antibiotic administration on outcomes in patients with severe sepsis and septic shock. They use Multivariable logistic and linear regression analyses to assess associations between timely administration of antibiotics and outcomes, including hospital mortality, 48-hour change in Sequential Organ Failure Assessment (SOFA) score (delta SOFA), and hospital length of stay (LOS). In the study, early antibiotic use is defined as administration of a broad-spectrum antibiotic within three hours from the time of ED arrival and patients are classified into early and delayed groups based on this criteria.

Conforming to the aforementioned study, data is synthesized for the following features.

- Age

- Gender
- Comorbidities
  - Malignancy

  - Hypertension

  - Diabetes 132 (22.3)

- Infection sites
  - Intra-abdominal infection

    - Pneumonia
    - Urinary tract infection
    - Others
- Initial serum lactate (mmol/L)
- APACHE II score
- Targets of initial resuscitation
    - CVP greater than or equal to 8 mmHg achieved
    - ScvO2 greater than or equal to 70% achieved

Method:

For the continuous variables - age and Targets of initial resuscitation (cvp and scvo2), sample size, median and interquartile ranges are used to simulate the data. MOre specifically, the standard deviation is calculated using sample size and interquartile ranges using the formula derived by * Wan, Xiang, Wenqian Wang, Jiming Liu, and Tiejun Tong. 2014.* Further, median itself is used as mean since sample size N ~ 600. Dichotomus and Categorical variables are simulated using the respective probabilities provided in the paper. Data is simulated separately for patients with early antibiotic administration and those with a delay. The below code shows the simulation process followed for the early group.

```
set.seed(101)
N1 = 377      # number of patients in the early group/ derived from the reference paper
age <- round(rnorm(N1, 65, 13.3))
gender <-
        sample(c("Male", "Female"),
               N1,
               replace = TRUE,
               prob = c(.56, .44))#Female and Male
Chr_renal_dis <-
        sample(c(1, 0), N1, replace = TRUE, prob = c(.04, .96))
Chr_hep_dis <-
        sample(c(1, 0), N1, replace = TRUE, prob = c(.09, .91))
Malignancy <- sample(c(1, 0), N1, replace = TRUE, prob = c(.59, .4))
Inf_site <-
        as.factor(sample(
                c("abdomen", "Pneumonia", "uti", "others"),
                N1,
                replace = TRUE,
                prob = c(.37, .29, .17, .18)
        ))
isl <- rnorm(N1, 4.8, 1.8)
apache_score <- rnorm(N1, 15, 6.7)
cvp <- sample(c(1, 0), N1, replace = TRUE, prob = c(.55, .45))
cvo2 <- sample(c(1, 0), N1, replace = TRUE, prob = c(.55, .45))
response_time <- rep(0, N1)

early_group <-
        data.frame(
                age,
                gender,
                Chr_renal_dis,
                Chr_hep_dis,
                Malignancy,
                Inf_site,
                isl,
```

```
                 apache_score,
                 cvp,
                 cvo2,
                 response_time
        )
```

Similar to the above, data is simulated for the delayed group.

The data so generated are combined and the first few rows are printed.

```
final.data <- rbind(early_group, delayed_group)
head(final.data)
```

```
##   age gender Chr_renal_dis Chr_hep_dis Malignancy  Inf_site      isl
## 1  61   Male             0           0          1 Pneumonia 6.550523
## 2  72 Female             0           0          0       uti 3.753876
## 3  56   Male             0           0          1   abdomen 2.374042
## 4  68   Male             0           0          1   abdomen 5.373013
## 5  69 Female             0           0          0 Pneumonia 5.952448
## 6  81   Male             0           0          0 Pneumonia 4.381339
##   apache_score cvp cvo2 response_time
## 1     7.118873   0    0             0
## 2    28.158328   0    0             0
## 3     9.200079   0    1             0
## 4     8.009377   1    1             0
## 5    10.857541   1    1             0
## 6     7.829702   0    0             0
```

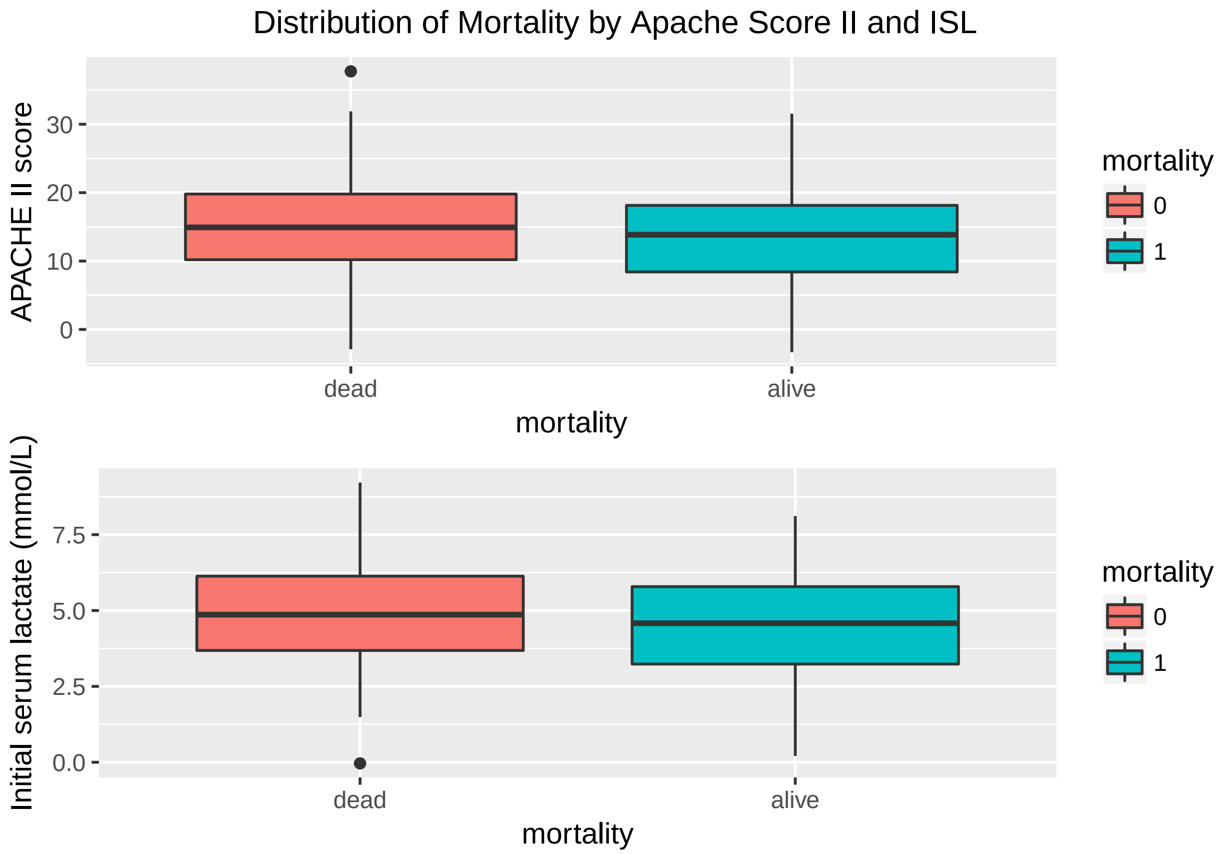
Using the odds ratios provided in the paper for significant predictor variables, corresponding coefficients are calculated which are further used to simulate the in hospital mortality rate, our first outcome.

```
# get the coeffs
beta1 <- log(0.54)
beta2 <- log(1.06)
beta3 <- log(1.2)

# using a linear combination of significant predictors calculate the overall probability
logistic <- function(t)
        1 / (1 + exp(-t))
linpred <-
        beta1 * final.data$response_time + beta2 * final.data$apache_score +
        beta3 * final.data$isl

prob <- logistic(linpred)
final.data$mortality <-
        as.factor(ifelse(runif(length(prob)) < prob, 0, 1))
```

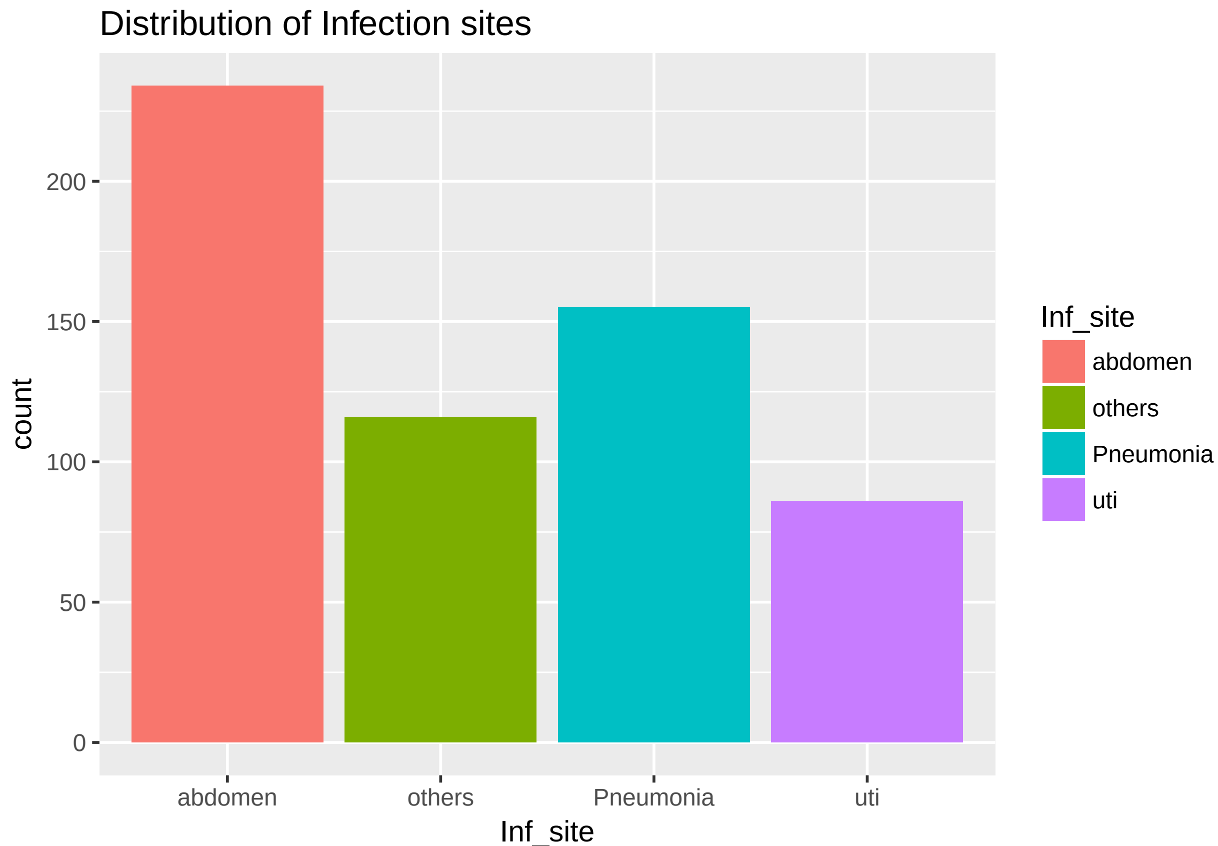
The plots below clearly indicate the differences between the two patient groups.

Higher levels of serum lactate and APACHE II score both indicative of severe sepsis/ septic shock are associated with patient mortality.

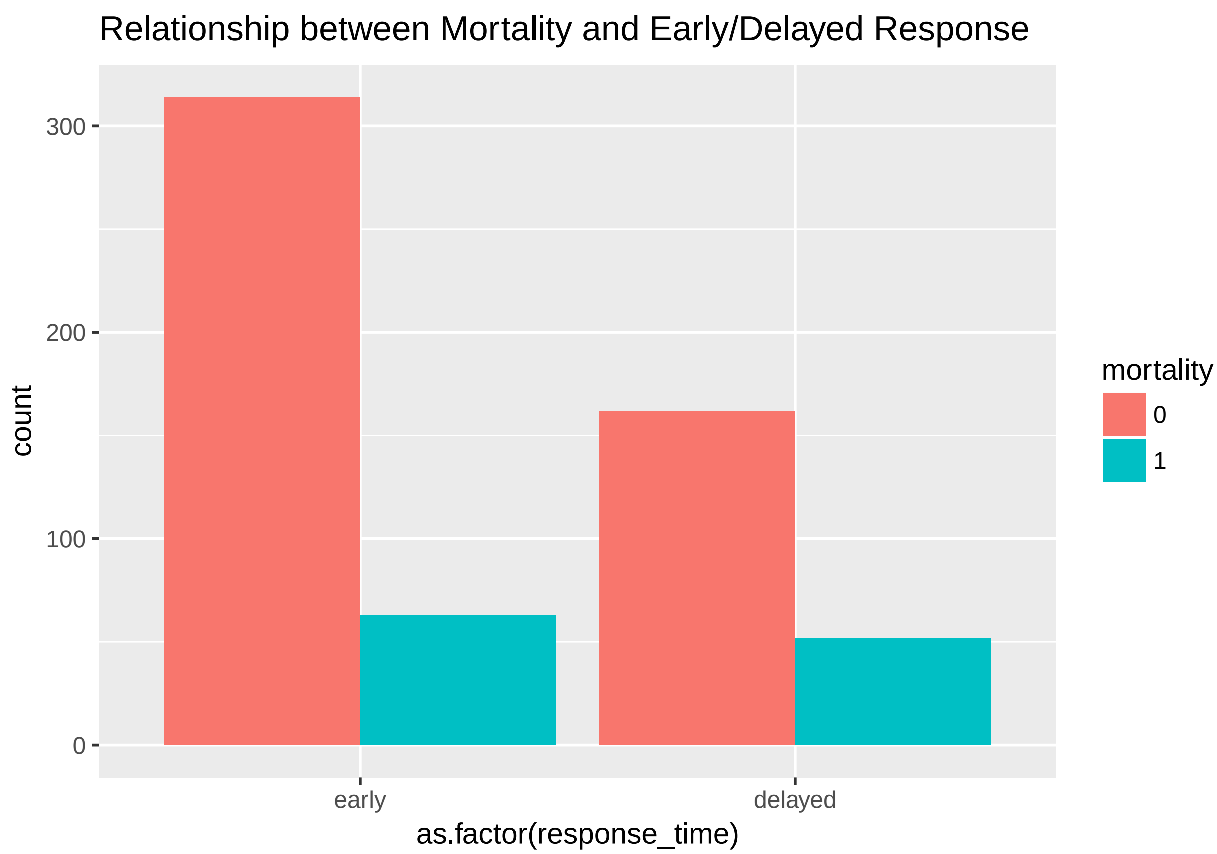
```
## Warning: package 'ggplot2' was built under R version 3.3.3
```



Most of the patients had Intra adbdomenal infections, followed by Pneumonia and UTI and others.



The significant effect of timely antibiotic administration on Mortality is indicated in the below plot in accordance with the study. Out of 377 patients who received early antibiotic administration, we see an in-hospital mortality rate of nearly 16 percent and for the delayed group comprising of 214 patients, it is 24 percent.



Results and Conclusion:

Fitting a logisic regression model on the simulated data, we see that the coefficients and p-values generated for the significant predictors are very similar to that obtained in the study stressing the fact that early antibiotic administration reduces the likelihood of patient mortaility.

```
mod <-
        glm(mortality ~ response_time + apache_score + isl,
            data = final.data,
            family = "binomial")
summary(mod)
```

```
##
## Call:
## glm(formula = mortality ~ response_time + apache_score + isl,
##     family = "binomial", data = final.data)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.0477  -0.7075  -0.5881  -0.4551   2.3787
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -0.20736    0.39896  -0.520  0.60324
## response_time  0.53939    0.21512   2.507  0.01216 *
## apache_score  -0.04429    0.01652  -2.681  0.00734 **
```

```
## isl           -0.16933    0.06675  -2.537  0.01119 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 582.49  on 590  degrees of freedom
## Residual deviance: 563.69  on 587  degrees of freedom
## AIC: 571.69
##
## Number of Fisher Scoring iterations: 4
```