Taylor Swift's Influence on the NFL Project

Amy Lai

December 2023

1 Background

The Kansas City Chiefs are a professional (American) football team in the National Football League (NFL). One of their tight ends (a football position) is Travis Kelce, who is rumored to be romantically involved with Taylor Swift. Taylor Swift needs no introduction. She is arguably the most famous singer-songwriter of all time. Her current Eras Tour is expected to generate $5 billion in consumer spending, an economic impact close to that of a small country.
https://time.com/6307420/taylor-swift-eras-tour-money-economy/

Taylor Swift's attendance at the two most recent Kansas City Chiefs games, has garnered much publicity. Her presence impacted ticket sales, Travis Kelce jersey sales, and broadcast ratings. In this question (and Question 3), you will assess the impact of Taylor Swift on broadcast ratings through an analysis of the weekly viewership for Kansas City regular-season games over the 2021-2023 seasons.

2 Data

Data was curated from www.sportsmediawatch.com. It includes the viewership information of games between September 12, 2021 and October 1, 2023.

3 Analysis

3.1 Data Exploration

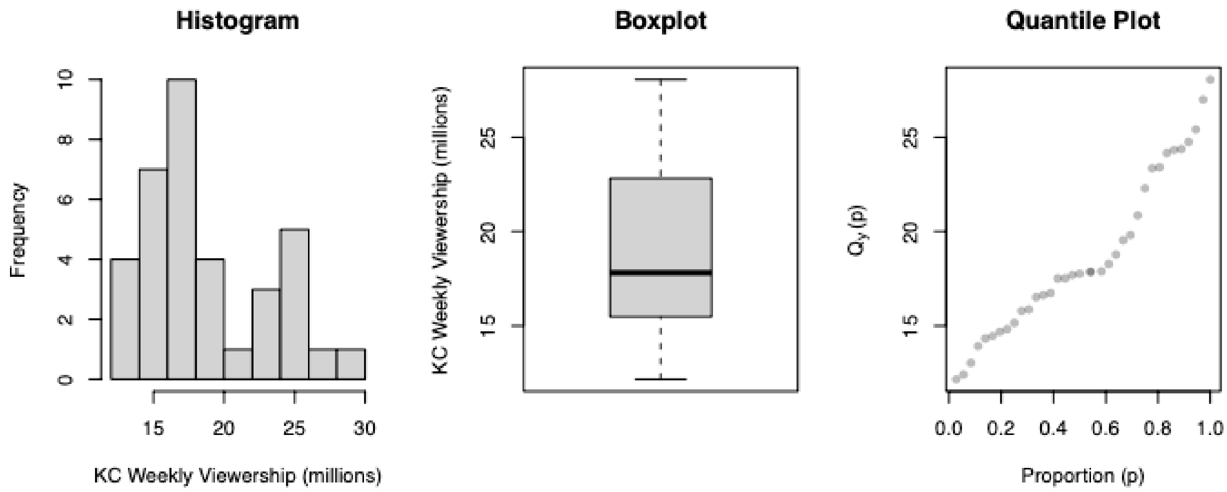
Before evaluating influence, we'll first take a simple look into the distribution of the data.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   12.16   15.63   17.81   18.75   22.55   28.06
```

```
##     Opponent       x
## 1       ARIZ 16.6200
## 2        BAL 19.8100
## 3        BUF 21.4650
## 4        CHI 24.3200
## 5        CIN 21.0900
## 6        CLE 19.5400
## 7        DAL 28.0600
## 8        DEN 17.5100
## 9        DET 24.7500
## 10        GB 24.3700
## 11       HOU 14.8300
## 12       IND 14.6800
## 13       JAX 15.4850
## 14       LAC 15.7775
## 15       LAR 23.3500
## 16        LV 16.0620
## 17       NYG 13.9200
## 18       NYJ 27.0000
## 19       PHI 18.2800
## 20       PIT 24.1600
## 21       SEA 15.8600
## 22        SF 22.2900
## 23        TB 20.8500
## 24       TEN 16.4200
## 25       WFT 12.4100
```

The statistics above summarizes the viewership data. We observe that the mean viewership is 18.75 million people per game. Games against DAL have the largest average viewership (29.06 million viewers. Games against WFT have the lowest average viewership (12.41 million viewers).

Below, we will examine three plots that describe the shape of the distribution.



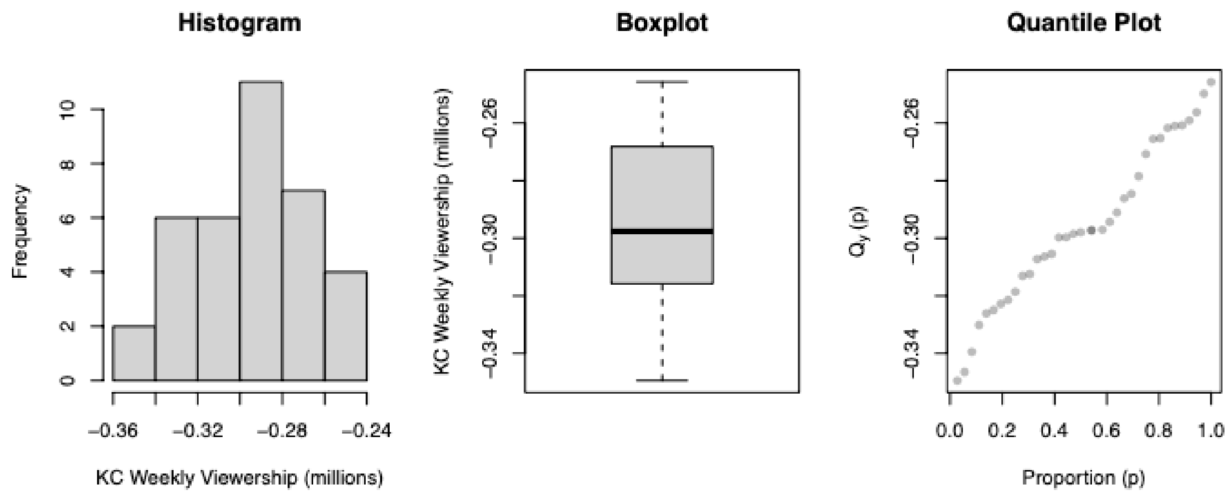
We observe that the histogram is right skewed with a slight bimodal shape, the data is centered around [14,25]. Based on the boxplot, we observe that the median is around 17 and the lower whisker is shorter, meaning that lower viewership numbers are more tightly clustered. The box covers viewer values around [16,24]. Based on the quantile, we notice that many points gather around value 17 and 24. Altogether, a weekly viewership between 14 millions to 25 millions is considered "typical".

To adjust the skewness and optimize, we apply power transformation to the above data. The Pearson's skewness coefficient quantifies this skewness. Since we want to account for both positive and negative skewness, the skewness coefficient is:

$$\rho(\alpha; \mathcal{P}) = \left| \frac{\frac{1}{N} \sum_{u \in \mathcal{P}} \left(t_u - \bar{t}\right)^3}{[SD(t)]^3} \right|$$

where t_u is the power-transformed observation for unit u. The argmin of this function can be found with the optim() function in R.

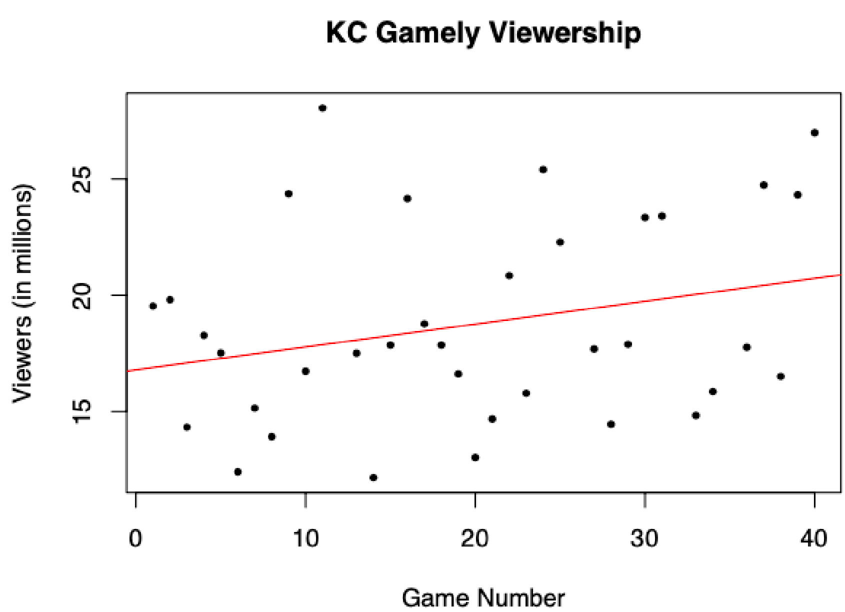
Applying this function can identify the optimal alpha = -0.42 for power transformation. Now we redraw the three diagrams with the transformed data.



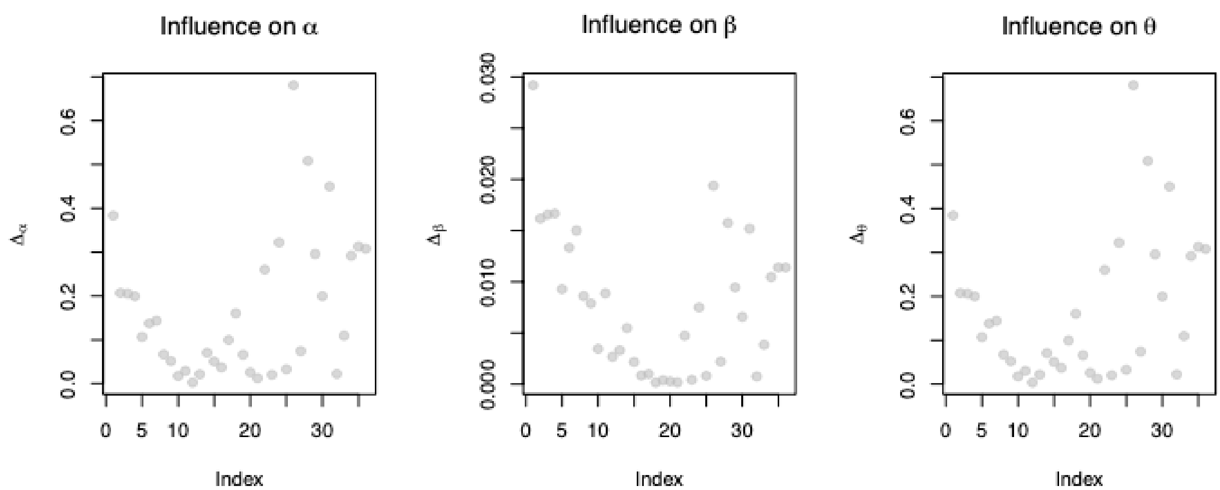
While the distribution for the transformed data looks more symmetric, this optimal transformation (alpha = -0.42) is hard to interpret. The ladder rule could give a more easily interpreted tranformation, though the trade-off would be that it's less optimal.

4 Evaluating Influence

We'll first fit a least square regression line to the viewership data.



For each game, we'll calculate it's influence on the fitted regression line. Let y = Viewers, x = Game number and $y_u = \alpha + \beta x_u + r_u, u \in P, \alpha, \beta \in R$. The influence for each game u is given by:
$\Delta(\alpha, u) = ||\hat{\alpha} - \hat{\alpha}_{[-u]}||_2$
$\Delta(\beta, u) = ||\hat{\beta} - \hat{\beta}_{[-u]}||_2$
$\Delta(\theta, u) = ||\hat{\theta} - \hat{\theta}_{[-u]}||_2$
where $\hat{\theta}_{[-u]}$ is the regression coefficients estimated from all of the data excluding unit u and $||x||_2$ is the Euclidean norm.

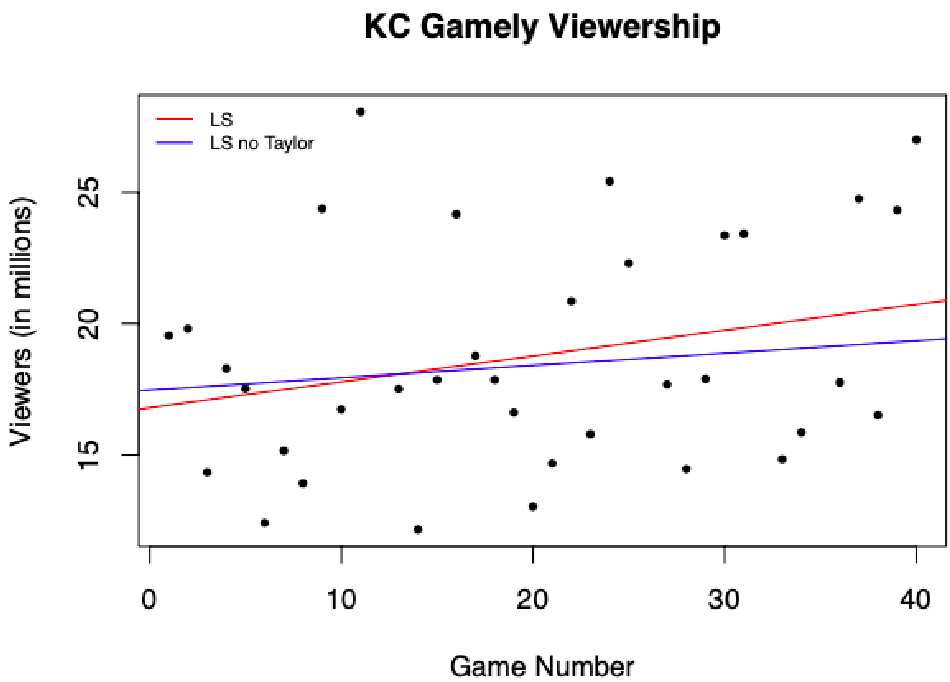


Since Δ_θ examines the influence on the slope and intercept together, it best demonstrates which games most influence the line of best fit. Although the left most and right most graph look identical, their values actually differ. The reason why they look similar is because the influence of the games on the slope is insignificant. The four most influential games are listed below:

```
##    Season Week       Date Opponent Viewers Game_Num
## 1    2023    4 2023-10-01      NYJ   27.00       40
## 30   2021   11 2021-11-21      DAL   28.06       11
## 32   2021    9 2021-11-07       GB   24.37        9
## 35   2021    6 2021-10-17      WFT   12.41        6
```

There are two approaches to evaluate the influence of Taylor Swift. Firstly, we can remove the games attended by Taylor Swift and compare the new regression line with the regression line we have. Alternatively, we can use robust regression, which mitigates the impact of outliers. In this project, we will explore both.

4.1 Approach 1



4.2 Approach 2

An objective function that facilitates robust regression is the Tukey Bisquare objective function:

$$\rho(\boldsymbol{\theta}; \mathcal{P}) = \sum_{u \in \mathcal{P}} \rho_k(r_u)$$

where $\boldsymbol{\theta} = (\alpha, \beta)^T$, $r_u = y_u - \alpha - \beta x_u$ and

$$\rho_k(r) = \begin{cases} \frac{r^2}{2} - \frac{r^4}{2k^2} + \frac{r^6}{6k^4} & \text{for } |r| \leq k \\ \frac{k^2}{6} & \text{for } |r| > k \end{cases}$$

We'll construct the appropriate function in R by first taking the derivative:

$$\begin{aligned} \boldsymbol{g} &= \nabla \rho(\boldsymbol{\theta}; \mathcal{P}) \\ &= \frac{d}{d\theta} \sum_{u \in \mathcal{P}} \rho_k(r_u) \\ &= \sum_{u \in \mathcal{P}} \frac{d}{d\theta} \rho_k(r_u) \\ &= \sum_{u \in \mathcal{P}} \frac{d\rho_k(r_u)}{dr_u} \frac{dr_u}{d\theta} \end{aligned}$$

$$\frac{d\rho_k(r_u)}{dr_u} = \begin{cases} r - 2\frac{r^3}{k^2} + \frac{r^5}{k^4} & \text{for } |r| \leq k \\ 0 & \text{for } |r| > k \end{cases}$$

$$\begin{aligned} \frac{r_u}{d\theta} &= \begin{bmatrix} \frac{dr_u}{d\alpha} \\ \frac{dr_u}{d\beta} \end{bmatrix} \\ &= -\begin{bmatrix} 1 \\ x_u \end{bmatrix} \end{aligned}$$

Therefore,

$$\boldsymbol{g} = -\begin{bmatrix} \sum_{u \in \mathcal{P}} \frac{\partial \rho_k(r_u)}{\partial r_u} \\ 0 \end{bmatrix}$$

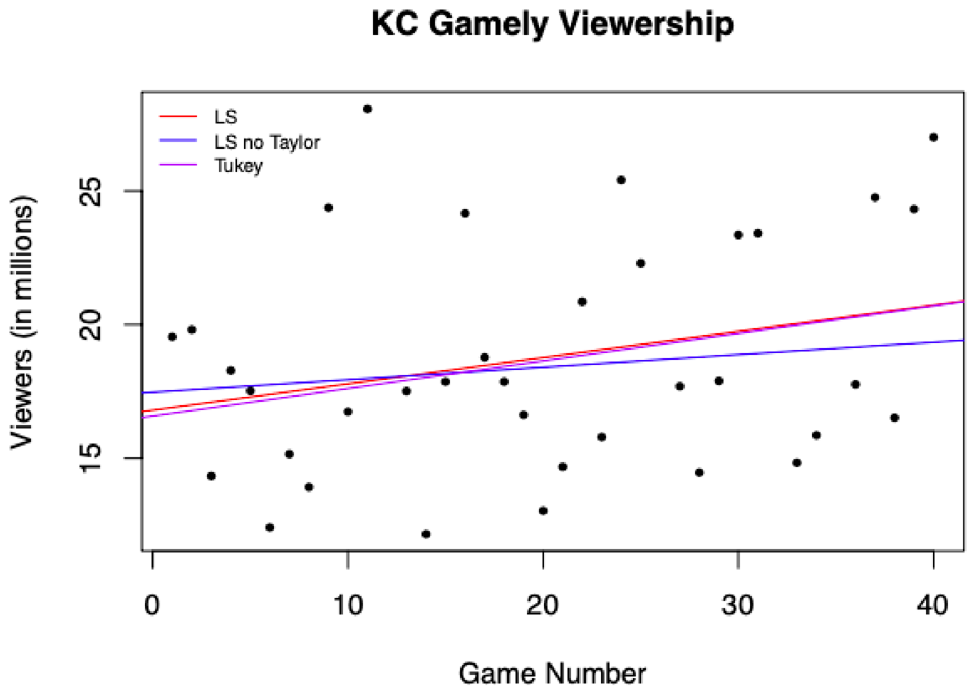
```r
tukey.fn <- function(r, k) {
    val = (r^2)/2 - (r^4)/(2 * k^2) + (r^6)/(6 * k^4)
    subr = abs(r) > k
    val[subr] = (k^2)/6
    return(val)
}
tukey.fn.prime <- function(r, k) {
    val = r - (2 * r^3)/(k^2) + (r^5)/(k^4)
    subr = abs(r) > k
    val[subr] = 0
    return(val)
}

createRobustTukeyRho <- function(x, y, kval) {
    ## Return this function
    function(theta) {
        alpha <- theta[1]
        beta <- theta[2]
        sum(tukey.fn(y - alpha - beta * x, k = kval))
    }
}

createRobustTukeyGradient <- function(x, y, kval) {
    function(theta) {
        alpha <- theta[1]
        beta <- theta[2]
        ru = y - alpha - beta * x
        rhok = tukey.fn.prime(ru, k = kval)
        -1 * c(sum(rhok * 1), sum(rhok * (x)))
    }
}
```

```r
rho <- createRobustTukeyRho(x = views$Game_Num, y = views$Viewers, kval = 4.685*5)
gradient <- createRobustTukeyGradient(x = views$Game_Num, y = views$Viewers, kval = 4.685*5)
gradientDescent(theta = m1$coef,
                rhoFn = rho, gradientFn = gradient,
                lineSearchFn = gridLineSearch,
                testConvergenceFn = testConvergence, maxIterations = 1000, relative = TRUE,
                lambdaStepsize = 0.0001, tolerance = 1E-10)
```

The final scatterplot with the new fitted line is shown below:



The blue line representing viers growth without the games Taylor Swift attend may not be suitable because it's not appropriate to simply eliminate data. The red and purple lines are better representations of the viewer growth over the years. Notice that these two lines are very similar, demonstrating that the outliers do not have a strong impact on the viewership. In other words, Taylor Swift's influence on Kansas City Chief's viewership is not significant (at least based on the data given).