Hurricane Femininity and Deaths

Sam Kinghorn

2023-01-13

1 Introduction

According to the paper published by Jung et al. (2014), hurricanes with female names are deadlier than ones with male names because people unconsciously think the storm will be less dangerous. I will explore this claim through an exploratory data analysis and model deaths based on femininity to see if there is any significance.

2 Data

The data is found at in the rethinking package in R (McElreath 2021), where the original data comes from Jung et al. (2014). The data set consists of hurricane name, year of hurricane, number of deaths, hurricane category, minimum pressure, normalized damage in dollars, whether the name is female or not, and femininity score. Minimum pressure is a measure of storm strength, where lower values indicate a stronger storm. The femininity score is on a scale from 1-11 where 1 is masculine and 11 is feminine. The score for each name was determined by an average of nine scores from nine raters that weren't aware they were hurricane names.

3 EDA

First, I load the data and investigate the femininity of hurricanes. 62 hurricanes in this data set have female names and 30 have male names. The femininity score is skewed left with most observations having a relatively high score.

```
library(rethinking); library(tidyverse); library(lme4); library(kableExtra); library(MASS)
```

```
data("Hurricanes")
df = Hurricanes
```

```
table(df$female)
```

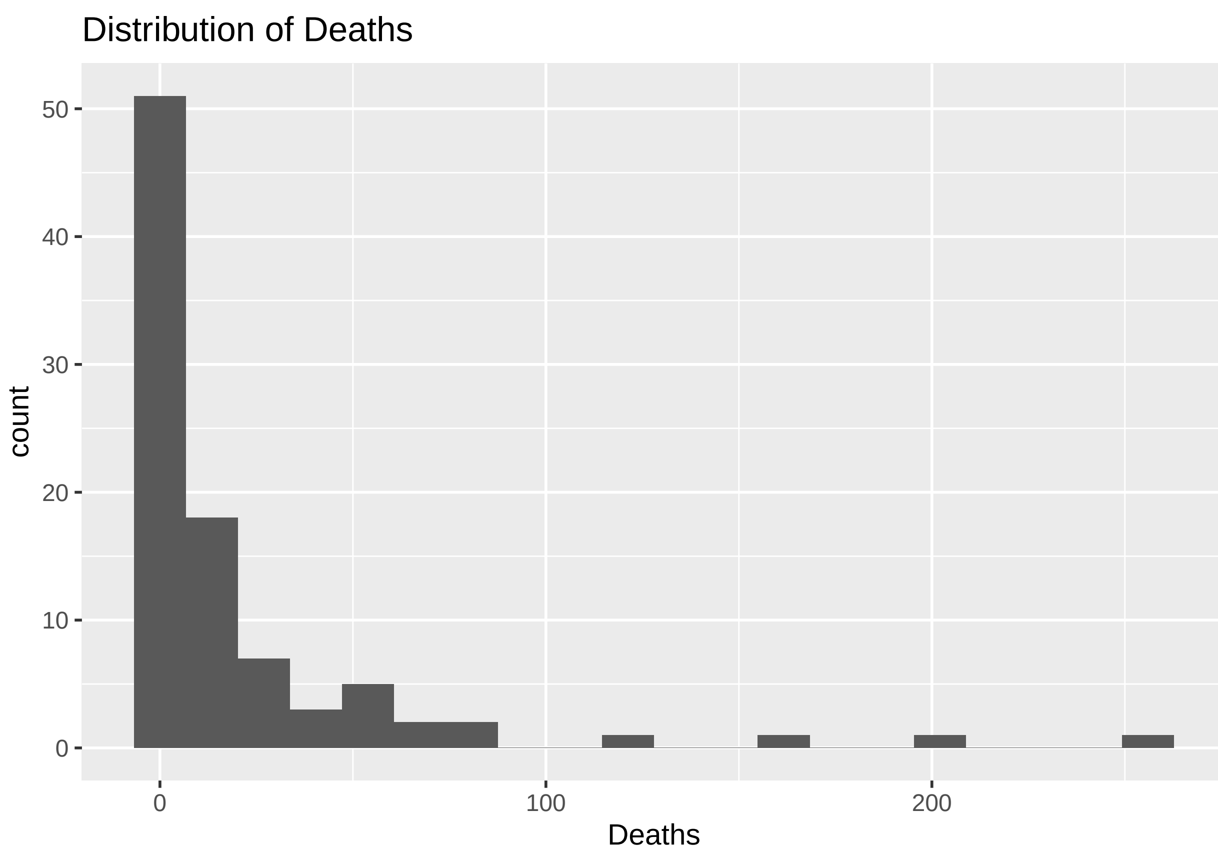
```
##
##  0  1
## 30 62
```

```r
summary(df$femininity)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.056   2.667   8.500   6.781   9.389  10.444
```

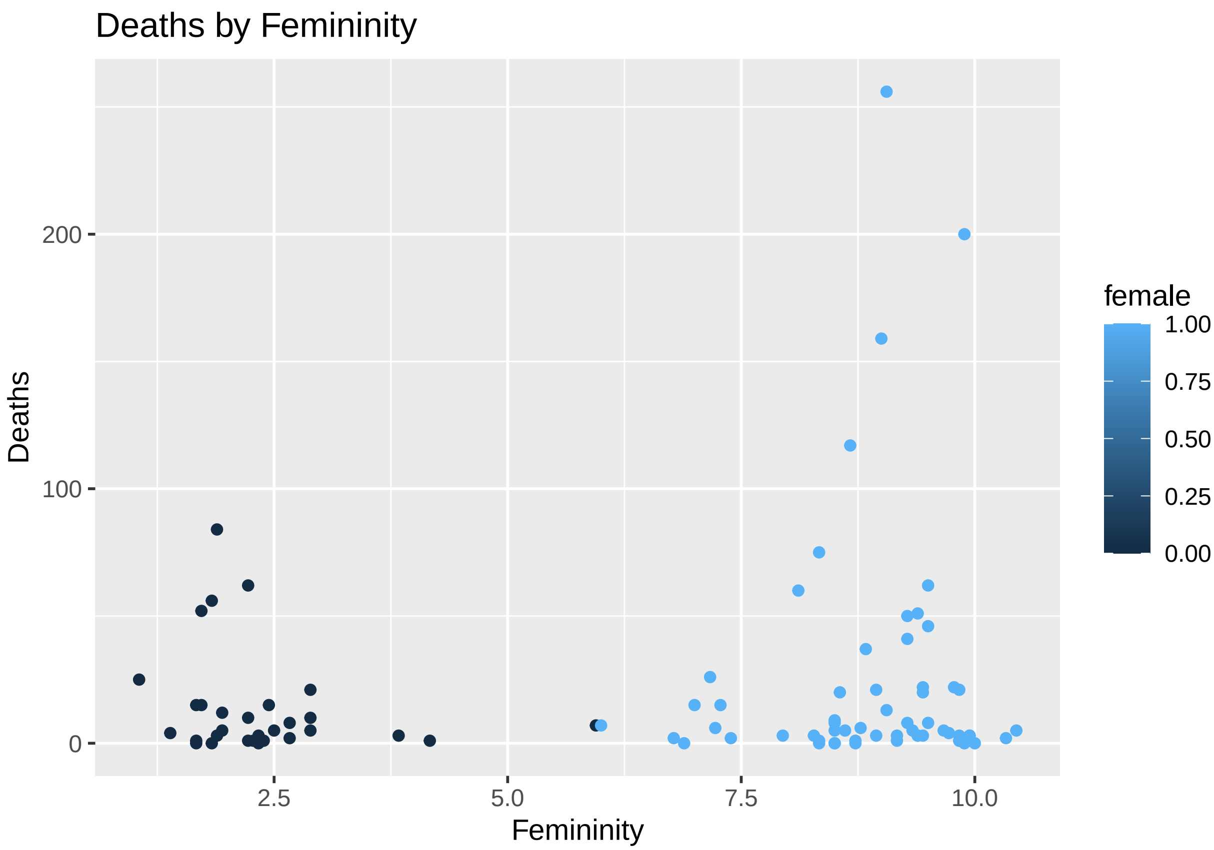
The distribution of deaths is skewed right with most death counts occurring between 0 and 10. There are extreme observations where hurricanes had death counts over 200.

```r
ggplot( df, aes( x=deaths ) ) + geom_histogram( bins=20 ) + labs(title='Distribution of Deaths
```

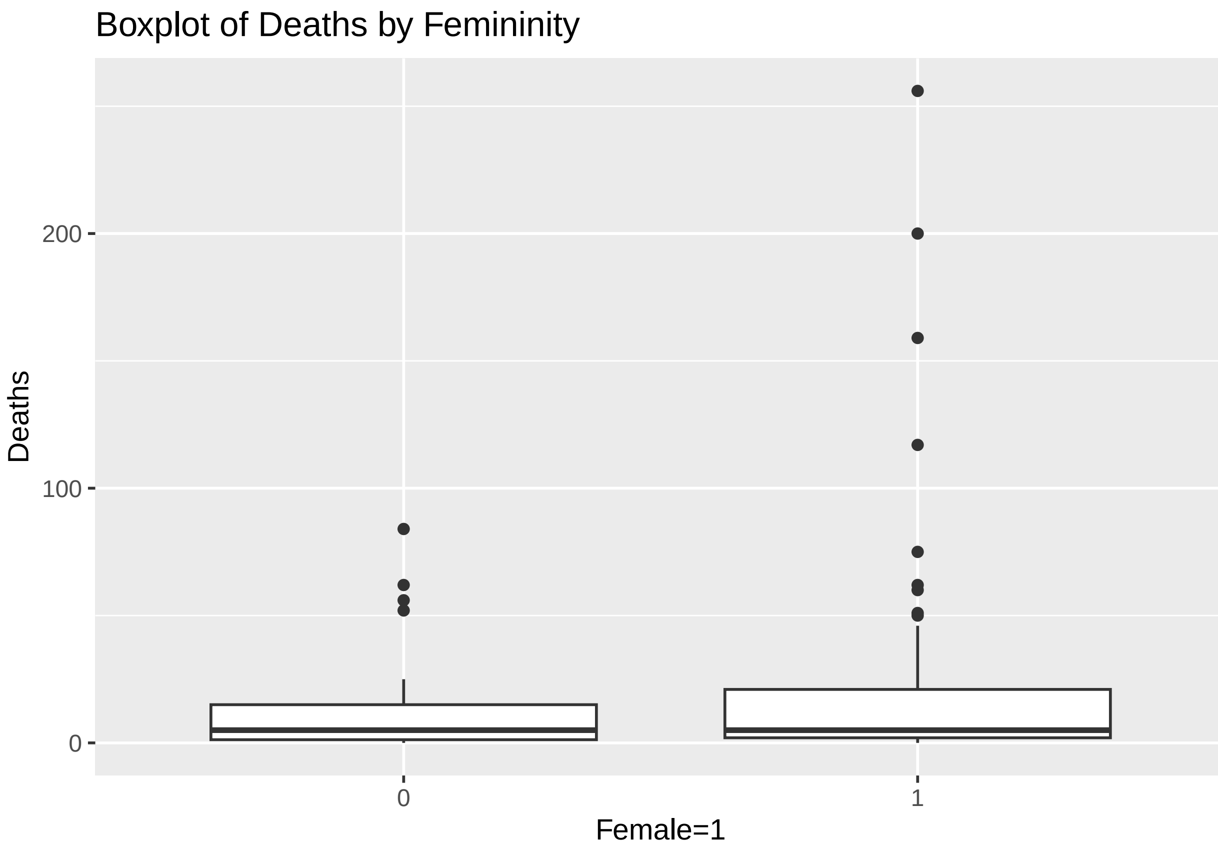


The next two graphs show that some female hurricanes have a high number of deaths compared to male hurricanes. It's difficult to say if this is significant since there are more than double the number of female hurricanes.

```
ggplot(data=df, mapping = aes(x=femininity, y=deaths, color=female)) +
  geom_point() + labs(title="Deaths by Femininity", y="Deaths", x="Femininity")
```

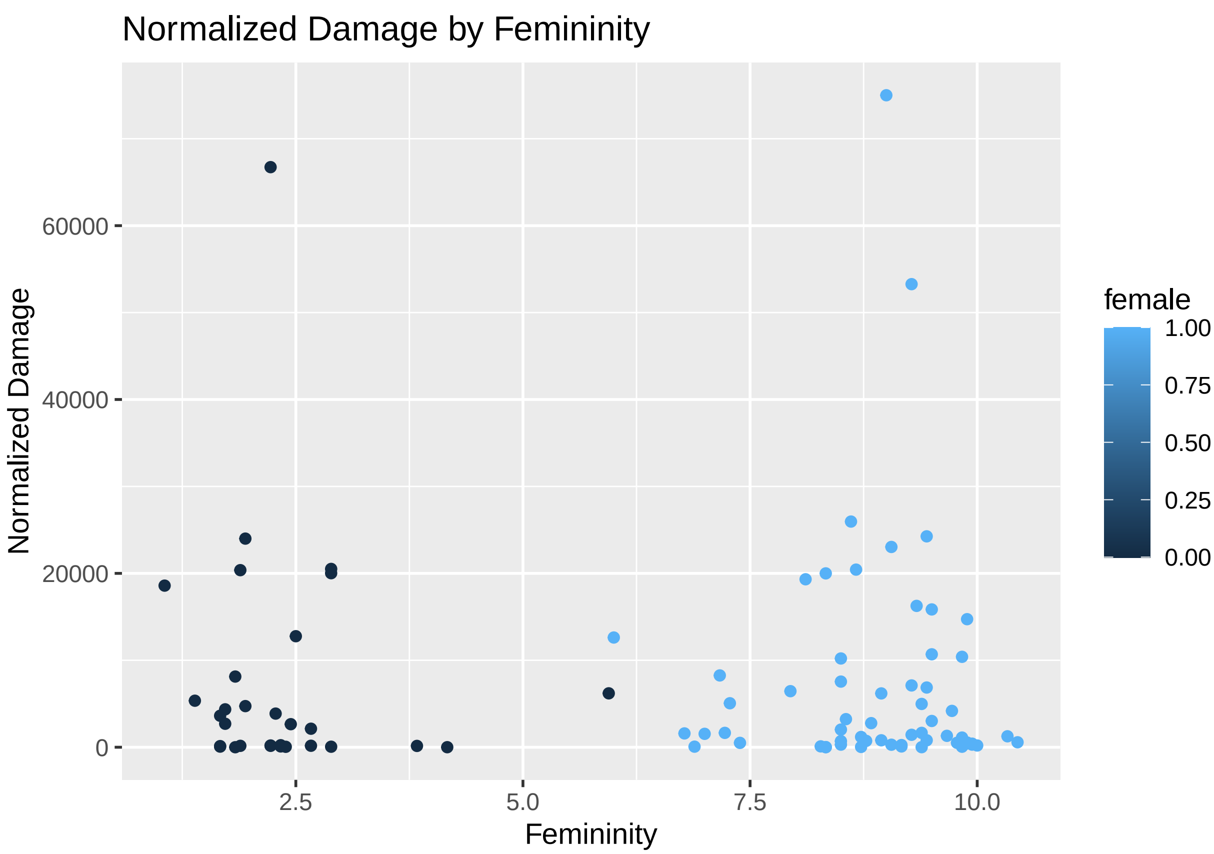


```
ggplot(data=df, mapping = aes(x=factor(female), y=deaths)) +
  geom_boxplot() + labs(title="Boxplot of Deaths by Femininity", y="Deaths", x="Female=1")
```

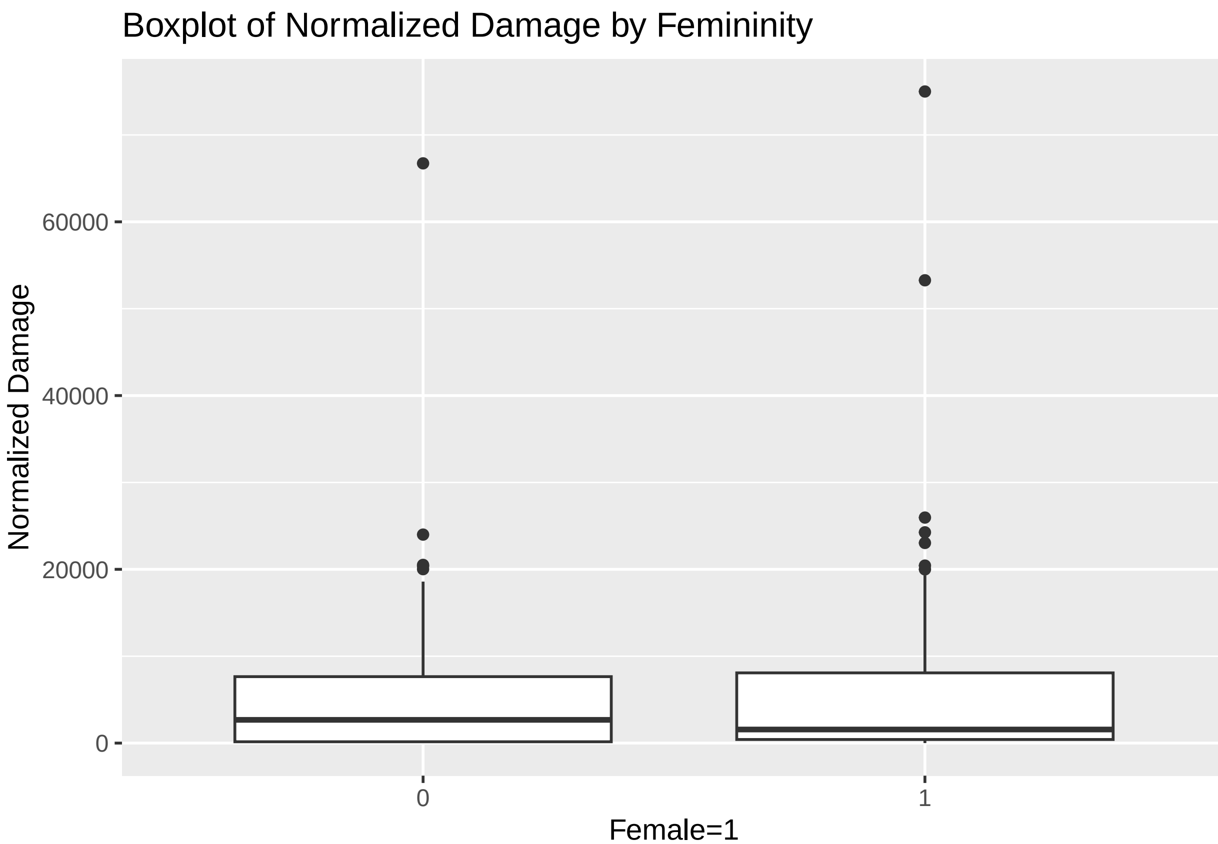


The next three graphs display the femininity in relation to damage and hurricane category. The first graph shows that low femininity has just as much damage as high femininity. The second graph reinforces this point through a boxplot. The third graph shows boxplots of femininity for hurricane categories. If female hurricanes were actually more dangerous, we might expect to see more male hurricanes in the first few categories and more female hurricanes in the last few. This doesn't appear to be the case.

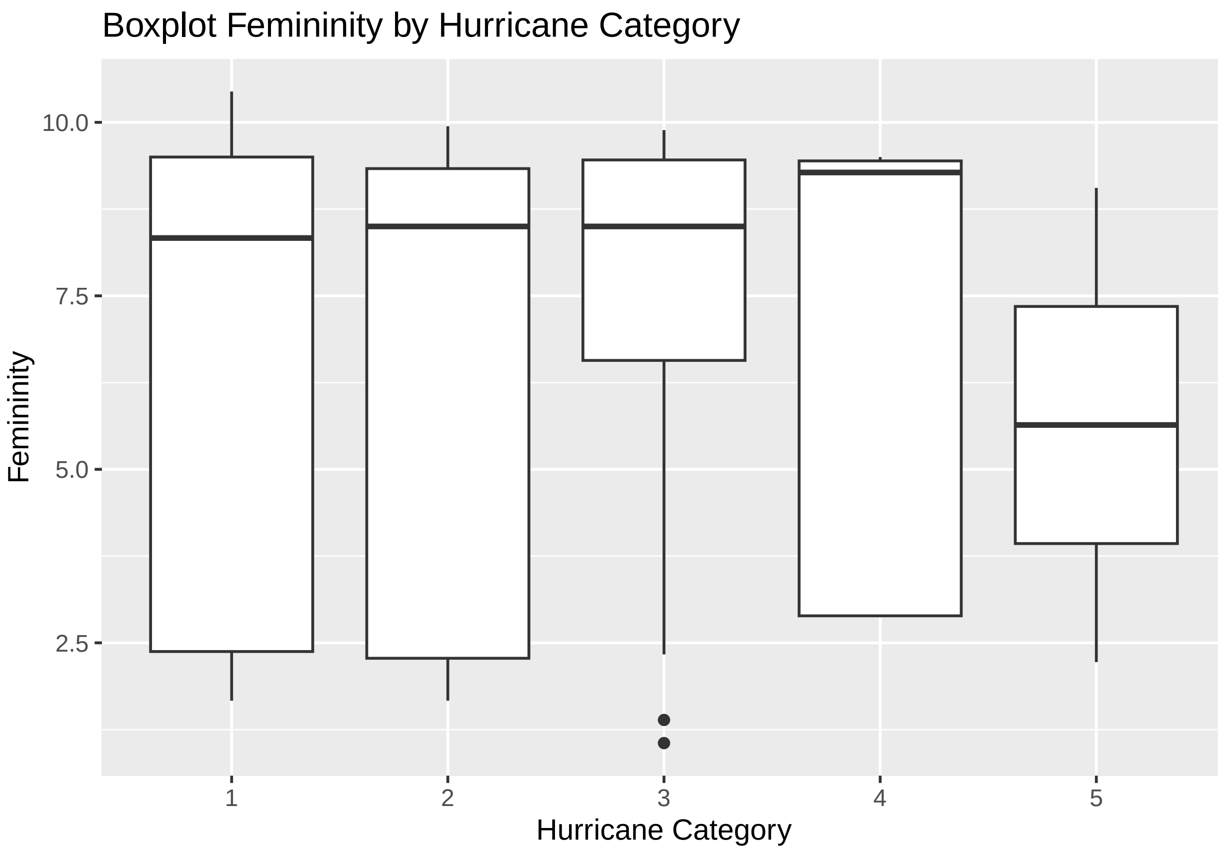
```
ggplot(data=df, mapping = aes(x=femininity, y=damage_norm, color=female)) +
  geom_point() +
  labs(title="Normalized Damage by Femininity", y="Normalized Damage", x="Femininity")
```



```
ggplot(data=df, mapping = aes(x=factor(female), y=damage_norm)) +
  geom_boxplot() +
  labs(title="Boxplot of Normalized Damage by Femininity", y="Normalized Damage", x="Female=1")
```



```r
ggplot(data=df, mapping = aes(x=factor(category), y=femininity)) +
  geom_boxplot() +
  labs(title="Boxplot Femininity by Hurricane Category", y="Femininity", x="Hurricane Category"
```



The EDA suggests that female hurricanes may result in more deaths but they don't appear to be more dangerous in terms of damage and category. This corroborates the paper by Jung et al. (2014), that people may perceive female hurricanes as less dangerous and not take necessary precautions, despite female hurricanes being just as dangerous as male hurricanes. Next, I will fit a model to determine if femininity is significant in explaining deaths.

4 Count Models

The response variable is the number of deaths, therefore we are modeling a count variable. Poisson regression is a popular method for modeling count data, but there is a restrictive assumption on the model: the mean and variance of the response variable are equal. This comes from the fact that the mean and variance of the Poisson distribution are equal. In the case of our data, the mean and variance are 20.65 and 1673.15, respectively.

```
sprintf("mean = %1.2f, var = %1.2f", mean(df$deaths), var(df$deaths))
```

```
## [1] "mean = 20.65, var = 1673.15"
```

Since the variance is much larger than the mean, the count data is over-dispersed. When count data is over-dispersed or under-dispersed, negative binomial regression can be used. Negative binomial regression is a generalization of Poisson regression and doesn't have the assumption of equidispersion. Negative binomial regression has an extra parameter, θ that can accommodate over-dispersion. The mean of the count is $E[y_i|\boldsymbol{x}_i] = \lambda_i$ and the variance of the count is $Var[y_i|\boldsymbol{x}_i] = \lambda_i + \frac{\lambda^2}{\theta}$ (Greene 2008). Depending on the severity of over-dispersion, Poisson regression might have inaccurate results, i.e. smaller standard errors and confidence intervals, because the assumptions of the model are not met.

I will use both models and compare the results. The covariates include femininity score, minimum pressure of the storm, and normalized estimated damage of the storm in dollars. Minimum pressure is a measure of storm strength where lower values indicate a stronger storm. The main research question is on the significance of femininity score, but the other variables will be included to account for their effects on deaths. I did not include interaction terms in the model because it's not clear what the interpretation would be. An interaction between femininity and minimum pressure or damage would suggest that one would depend on the other, or vice versa. Since the names of the hurricanes are picked before the hurricane starts, it doesn't make sense for these variables to depend on one another.

5 Poisson Regression

We can model deaths through Poisson regression. The research hypothesis and conceptual model are:

5.1 Hypothesis

The number of deaths increases with femininity.

5.2 Model

$$deaths_i \sim Poisson(\mu_i)$$

$$log(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}$$

where x_{1i} is the level of femininity, x_{2i} is the minimum pressure measurement, and x_{3i} is the normalized estimate of damages.

5.3 Results

The R printout for the model results can be seen in Appendix section 1. The printout shows all variables as highly significant, i.e. all variables have a p-value below 0.01. This means femininity score offers some explanatory power for hurricane deaths. We can interpret the coefficients by exponentiating them

```
model_pois = glm(deaths ~ femininity + min_pressure + damage_norm, family="poisson", data=df)
```

```
exp(cbind(Coefficient = coef(model_pois), confint(model_pois)))
```

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

##               Coefficient        2.5 %       97.5 %
## (Intercept)  1.847418e+10 1.355398e+09 2.498235e+11
## femininity   1.085669e+00 1.069062e+00 1.102831e+00
## min_pressure 9.777788e-01 9.751343e-01 9.804375e-01
## damage_norm  1.000025e+00 1.000023e+00 1.000028e+00
```

The interpretation for the coefficient associated with femininity is: for every unit increase in femininity score, there is an approximate 8% increase in hurricane deaths.

6 Negative Binomial Regression

We can also model deaths through negative binomial regression. The research hypothesis and conceptual model are the same as for Poisson regression:

6.1 Hypothesis

The number of deaths increases with femininity.

6.2 Model

$$log(\mu_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i}$$

where x_{1i} is the level of femininity, x_{2i} is the minimum pressure measurement, and x_{3i} is the normalized estimate of damages.

6.3 Results

The R printout for the model results can be seen in Appendix section 2. The printout shows that normalized damage and minimum pressure are significant in explaining deaths, i.e. the variables have a p-value below 0.05. The coefficient associated with femininity score has a p-value above 0.10, indicating it isn't significant at any level. Thus, we fail to reject the null hypothesis that the coefficient of femininity score is 0. In other words, this model provides evidence that there is not an association between femininity score and hurricane deaths. The additional dispersion parameter, θ, is estimated from the data and is shown in the output to be 0.717. The coefficients have the same interpretation as Poisson regression.

A likelihood ratio test is performed to determine the significance of femininity. The chi-square test has a p-value of 0.38 indicating femininity is not a statistically significant predictor of hurricane deaths.

```
model_nb = glm.nb(deaths ~ femininity + min_pressure + damage_norm, data=df)
model_nb2 = glm.nb(deaths ~ min_pressure + damage_norm, data=df)
```

```
anova(model_nb, model_nb2)
```

```
## Likelihood ratio tests of Negative Binomial Models
##
## Response: deaths
##                                     Model     theta Resid. df    2 x log-lik.
## 1              min_pressure + damage_norm 0.7108020        89       -656.5366
## 2 femininity + min_pressure + damage_norm 0.7174634        88       -655.7669
##     Test    df  LR stat.   Pr(Chi)
## 1
## 2 1 vs 2     1 0.7696128 0.3803369
```

7 Comparison

The biggest difference in the two models is the level of significance for the regression coefficients. All of the coefficients are highly significant in the Poisson model but femininity is insignificant in the negative binomial model. As expected, the confidence intervals in the negative binomial model are wider to account for over-dispersion. To determine which model is more appropriate, a likelihood ratio test is performed

```
pchisq(2*(logLik(model_nb) - logLik(model_pois)), df=1, lower.tail=FALSE)
```

```
## 'log Lik.' 0 (df=5)
```

The likelihood ratio test shows a p-value of 0, indicating the addition of the dispersion parameter in the negative binomial model is more appropriate than the Poisson model. Also, the AIC for the negative binomial regression is much smaller than the Poisson regression, which suggests a better fit.

In terms of the other parameters, the models are similar. An increase in minimum pressure is estimated to decrease deaths and an increase in damage is estimated to increase deaths.

8 Conclusion

The goal of this analysis was to determine if there was evidence of higher hurricane death counts for more feminine names. Since the response variable is death counts, the results of Poisson regression and negative binomial regression were compared. Negative binomial regression proved to be a better fit because of over-dispersion within death count. Based on the model results, femininity score is an insignificant predictor of hurricane death count. This provides evidence that hurricane deaths and the femininity score of the hurricane name are not associated, and that female hurricanes are not deadlier than male hurricanes.

9 References

Greene, William. 2008. "Functional Forms for the Negative Binomial Model for Count Data." *Economics Letters* 99 (3): 585–90. https://doi.org/https://doi.org/10.1016/j.econlet.2007.10.015.

Jung, Kiju, Sharon Shavitt, Madhu Viswanathan, and Joseph M. Hilbe. 2014. "Female Hurricanes Are Deadlier Than Male Hurricanes." *Proceedings of the National Academy of Sciences* 111 (24): 8782–87. https://doi.org/10.1073/pnas.1402786111.

McElreath, Richard. 2021. *Rethinking: Statistical Rethinking Book Package.*

10 Appendix

10.1 Poisson Model

```r
model_pois = glm(deaths ~ femininity + min_pressure + damage_norm, family="poisson", data=df)
summary(model_pois)
```

```
##
## Call:
## glm(formula = deaths ~ femininity + min_pressure + damage_norm,
##     family = "poisson", data = df)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -7.8790  -4.0675  -2.3891  -0.0269  26.2522
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   2.364e+01  1.331e+00   17.76   <2e-16 ***
## femininity    8.220e-02  7.932e-03   10.36   <2e-16 ***
## min_pressure -2.247e-02  1.384e-03  -16.24   <2e-16 ***
## damage_norm   2.532e-05  1.125e-06   22.51   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 4031.9  on 91  degrees of freedom
## Residual deviance: 2472.6  on 88  degrees of freedom
## AIC: 2805.5
##
## Number of Fisher Scoring iterations: 6
```

10.2 Negative Binomial Model

```
model_nb = glm.nb(deaths ~ femininity + min_pressure + damage_norm, data=df)
summary(model_nb)
```

```
##
## Call:
## glm.nb(formula = deaths ~ femininity + min_pressure + damage_norm,
##     data = df, init.theta = 0.7174633564, link = log)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0670  -1.0152  -0.5627   0.3224   3.2265
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept)   1.734e+01  7.948e+00   2.182   0.0291 *
## femininity    3.480e-02  4.014e-02   0.867   0.3860
## min_pressure -1.622e-02  8.167e-03  -1.986   0.0470 *
## damage_norm   8.433e-05  1.170e-05   7.206 5.75e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(0.7175) family taken to be 1)
##
##     Null deviance: 166.16  on 91  degrees of freedom
## Residual deviance: 104.05  on 88  degrees of freedom
## AIC: 665.77
##
## Number of Fisher Scoring iterations: 1
##
##
##               Theta:  0.717
##           Std. Err.:  0.107
##
##  2 x log-likelihood:  -655.767
```