STAT 021 S22 HW 8

Shikha Shrestha

TOTAL POINTS

10 / 10

QUESTION 1

1 Description 3 / 3

✓ **+ 3 pts** **Correct**

QUESTION 2

2 Identify possible mistakes 3 / 3

✓ **+ 3 pts** **Complete**

+ 0 pts Incomplete

QUESTION 3

3 Connection to ASA guidelines 4 / 4

✓ **+ 4 pts** **Correct**

I. Describe the statistical analysis and research question

The study titled "Female Hurricanes is deadlier than male hurricanes " concluded that the higher number of fatalities associated with feminine-sounding hurricanes compared to names associated with males is due to gender bias. The study consisted of an observational study that involved categorizing the names of hurricanes into a male-sounding name or a female-sounding name and then comparing the number of deaths across the two groups using negative binomial regression analysis. The second part of the study involved an experiment where participants were asked to predict the intensity of a hurricane(1-5) based on their names.

Describe the statistical analysis and research question A. Include the source where you found this analysis referenced as "bad statistics" B. Also, include the original source whenever possible.

The study used negative binomial regression analysis and ANOVA test with the gender of the hurricane name and participants' sex as the categorical variables used to explain the number of deaths.

The source that referenced bad statistics: https://gizmodo.com/no-female-hurricanes-are-not-deadlier-than-male-hurric-1585563258

Original Source:
https://www.pnas.org/doi/epdf/10.1073/pnas.1402786111

II. Identify where you think mistakes occurred and why you think they occurred A. Think broadly about the steps of modeling: choose, fit, assess, use

The study included a major outlier, Hurricane Katrina, which had a disproportionately higher death rate. This could have been obvious if a residual plot was assessed at the assessing stage of the modeling. Likewise, the study used hurricane data from 1950 to 1973 when the death tolls were higher and all storms had female names. Likewise based on the article I've referenced the study included indirect deaths in their fatality counts, which included people who were for instance killed by fallen electrical lines in the clean-up after a storm. Likewise, it counted hurricane Sandy which had the highest fatalities, as strongly feminine, even though it is often used for both men and women. Similarly, it also avoided including deaths from outside of the US. All these errors seem to have occurred during data collection and choosing the model.

Likewise, this statement from the study misrepresents correlation as causation: "*Our model suggests that changing a severe hurricane's name from Charley … to Eloise … could nearly triple its death toll*,". This seems to be a common mistake in the realm of statistical use.

B. Also reflect on study design and data collection

It seems to me that majority of the flaw in this study came during the process of design and data collection. For the observational part of this study, the primary source of the data came from The National Hurricane Center's paper which included data for 159 years. However, the 2014 study only looked at 60 years of data. The hurricanes from the later years were much less deadly, and the male and female names are distributed more equally.

III. Connect the analysis to the ASA's guidelines for Ethical Statistical Practice

A. Which principles do you think were violated?

I think the following principle from principle A was violated:

2. Uses methodology and data that are valid, relevant, and appropriate, without favoritism or prejudice, and in a manner intended to produce valid, interpretable, and reproducible results.

Likewise, the following principle from principle B was violated:

1 Description **3 / 3**

✓ **+ 3 pts** **Correct**

I. Describe the statistical analysis and research question

The study titled "Female Hurricanes is deadlier than male hurricanes " concluded that the higher number of fatalities associated with feminine-sounding hurricanes compared to names associated with males is due to gender bias. The study consisted of an observational study that involved categorizing the names of hurricanes into a male-sounding name or a female-sounding name and then comparing the number of deaths across the two groups using negative binomial regression analysis. The second part of the study involved an experiment where participants were asked to predict the intensity of a hurricane(1-5) based on their names.

Describe the statistical analysis and research question A. Include the source where you found this analysis referenced as "bad statistics" B. Also, include the original source whenever possible.

The study used negative binomial regression analysis and ANOVA test with the gender of the hurricane name and participants' sex as the categorical variables used to explain the number of deaths.

The source that referenced bad statistics: https://gizmodo.com/no-female-hurricanes-are-not-deadlier-than-male-hurric-1585563258

Original Source:
https://www.pnas.org/doi/epdf/10.1073/pnas.1402786111

II. Identify where you think mistakes occurred and why you think they occurred A. Think broadly about the steps of modeling: choose, fit, assess, use

The study included a major outlier, Hurricane Katrina, which had a disproportionately higher death rate. This could have been obvious if a residual plot was assessed at the assessing stage of the modeling. Likewise, the study used hurricane data from 1950 to 1973 when the death tolls were higher and all storms had female names. Likewise based on the article I've referenced the study included indirect deaths in their fatality counts, which included people who were for instance killed by fallen electrical lines in the clean-up after a storm. Likewise, it counted hurricane Sandy which had the highest fatalities, as strongly feminine, even though it is often used for both men and women. Similarly, it also avoided including deaths from outside of the US. All these errors seem to have occurred during data collection and choosing the model.

Likewise, this statement from the study misrepresents correlation as causation: "*Our model suggests that changing a severe hurricane's name from Charley … to Eloise … could nearly triple its death toll*,". This seems to be a common mistake in the realm of statistical use.

B. Also reflect on study design and data collection

It seems to me that majority of the flaw in this study came during the process of design and data collection. For the observational part of this study, the primary source of the data came from The National Hurricane Center's paper which included data for 159 years. However, the 2014 study only looked at 60 years of data. The hurricanes from the later years were much less deadly, and the male and female names are distributed more equally.

III. Connect the analysis to the ASA's guidelines for Ethical Statistical Practice

A. Which principles do you think were violated?

I think the following principle from principle A was violated:

2. Uses methodology and data that are valid, relevant, and appropriate, without favoritism or prejudice, and in a manner intended to produce valid, interpretable, and reproducible results.

Likewise, the following principle from principle B was violated:

3. Communicates the stated purpose and the intended use of statistical practices. Is transparent regarding a priori versus post hoc objectives and planned versus unplanned statistical practices. Discloses when multiple comparisons are conducted and any relevant adjustments.

B. Were these violations intentional? How could you tell?

Since the authors have conveniently excluded data that might negatively impact the hypothesis of the study I would say the authors have intentionally violated the principles. However, since the authors were marketing researchers and not statisticians some of the mistakes such as the inclusion of the outlier could have been unintentional.

2 Identify possible mistakes **3 / 3**

✓ **+ 3 pts** **Complete**

+ 0 pts Incomplete

3. Communicates the stated purpose and the intended use of statistical practices. Is transparent regarding a priori versus post hoc objectives and planned versus unplanned statistical practices. Discloses when multiple comparisons are conducted and any relevant adjustments.

B. Were these violations intentional? How could you tell?

Since the authors have conveniently excluded data that might negatively impact the hypothesis of the study I would say the authors have intentionally violated the principles. However, since the authors were marketing researchers and not statisticians some of the mistakes such as the inclusion of the outlier could have been unintentional.

3 Connection to ASA guidelines 4 / 4

✓ + 4 pts Correct