

7COM1079-0901-2025 - Team Research and Development Project

Final report title: Gender Differences in Amusement-Park Injuries

Group ID: GroupA12

Dataset number: DS029

Prepared by:

Prathap kandasamy - 24051461

Velan Sivasankaran - 24080883

Mythili Pavazhavannan - 24081194

Pamanaban murugan - 24091663

Rithvik Reddy Dasari - 24155985

## Table of Contents

Section	Page
<b>1. Introduction</b>	3
1.1 Problem statement and research motivation	3
1.2 Dataset and cleaning procedures	
1.3 Research question and hypotheses	
<b>2. Background Research</b>	3
<b>3. Data Visualisation</b>	4
3.1 Classification of body-part categories	4–6
3.2 Distribution of injuries by gender	
3.3 Contingency table of injured body parts by gender	
<b>4. Statistical Analysis</b>	6
4.1 Choice of statistical test and assumptions	6
4.2 Hypothesis testing results	
<b>5. Evaluation of the Project</b>	6
5.1 What went well	6-7
5.2 Points for improvement	
5.3 Group's time management	
5.4 Project's overall judgement	
5.5 Comment on GitHub log output	
<b>6. Conclusion</b>	7
<b>7. Reference List</b>	7
<b>8. Appendices</b>	9
Appendix A: R code used for analysis and visualisation	9-10
Appendix B: GitHub log output	

# **1 Introduction**

## **1.1 Problem statement and motivation**

Even though there is a good safety record of modern amusement parks, accidents are also possible when the rides are involved in failure or when the riders act recklessly. In the United States (Consumer Product Safety Commission, 2001) 10580 individuals needed emergency care as a result of injuries in amusement rides in 2000, and a worldwide survey of reports indicated that 182 accident incidents in 38 countries made upride failures and drowning the most frequent causes. It demonstrates that the number of injuries is minor compared to the millions of rides annually but the effects are severe. Knowledge of injury patterns will enable ride operators and regulators to create safer rides and engage in specific education to riders.

## **1.2 Dataset and cleaning procedures**

The data set Amusement Park Injuries Cleaned is composed of 542 records of injuries in the US amusement parks. On a case-by-case basis, there are variables that include gender, age, ride name, body part injured, alleged injury type and cause, and location data (Guo et al., 2023). The information was obtained through regulatory reports and standardised on a unified template. The cases that had complete details of gender and body parts were retained. The few typographical mistakes were corrected and gender values written in capital letters so that they can be reliably analyzed statistically.

## **1.3 Research question and hypotheses**

The research question covered by this report is the following: Is the percentage of injured body parts of male and female riders in amusement-park accidents different? Injury pattern in relation to gender has been reported in other settings. An occupational injury systematic review revealed that men and women are associated with dissimilar risks of injuries due to both biological and occupational factors (Biswas et al., 2022). At amusement parks, it is of anecdotal concern that females can be more susceptible to head injury, neck injury or males to limb injury. In testing whether there are such differences, we formulate the analysis as a test of independence between two categorical variables, that is, whether there is any difference in the body-part category and gender (Male, Female, Unknown).

The null hypothesis ( $H_0$ ) is that there is no relationship between gender and the part of the body that has been injured; any observed difference is as a result of chance. The other hypothesis ( $H_1$ ) is that there is no independence between gender and part of body injured, which implies that some body parts are overrepresented in one gender/ sex whereas other parts are overrepresented in the other gender. A chi-square test of independence will be used to test these hypotheses.

# **2 Background research**

Amusement parks are usually secure. In 2019, a study conducted by Woodcock reviewed the accident reports of 38 countries and found that accidents are uncommon in comparison to the number of rides that are performed through them, but significant accidents can take place when the ride fails (Woodcock, 2019). The Consumer Product Safety Commission (CPSC) in the United States checks safety in the rides and established that females

got hurt approximately 1.5 times more frequently than males and injuries were widespread in children aged 1014 years. Woodcock (2014) also reported in his previous study that used American data that head and neck injuries were an important percentage of the overall injuries. An example of this is found in (Consumer Product Safety Commission, 2001).

Clinical interest in the nature of amusement-ride injuries has been aroused. The researchers (Pfister et al., 2009) measured head accelerations at roller coasters and demonstrated that the values of the head injury criterion were significantly lower than the values of the thresholds related to traumatic brain injury. However, Greenberg and Ruby (2021) approximated that approximately 1299 amusement-parks-injuries took place in the United States in 2019 and 28 of the injuries in children were located in the head and neck. An epidemiological study covering 92,885 paediatric amusement-ride injuries between 1990 and 2010 reported a mean age of 8.7 years, with head and neck injuries accounting for 28% of cases and only 1.5% requiring hospitalisation (Woodcock, 2014). Additional information given by Nationwide Children's Hospital showed that while 20 children on average are treated daily during the summer season with amusement-ride injuries (Nationwide Children's Hospital, 2013).

Gender differences in risk-taking behaviour may also play a role in the pattern of injuries (Neuroscience News, 2013). The summary of research done by Neuroscience News indicates that women are risk-averse as compared to men as they are more sensitive to possible losses as opposed to men who are more optimistic and ready to give risky undertakings. An independent systematic review concluded that exposure and physiological differences are some of the factors that contribute to sex differences in work-related injury. Such findings support the fact that there may be also possibility that gender can also determine the nature of the body part that was injured at amusement parks. (Biswas et al., 2022)

Lastly, our analysis was contextualised by reviewing safety guidelines regarding operators of amusement-rides. The Health and Safety Executive (HSE) in the United Kingdom has published the Fairgrounds and Amusement Parks guide, which states that fairgrounds are usually relatively safe environment as compared to the ordinary activities but that the risk should be controlled. The guide recommends that designers, manufacturers and operators should diminish risks by effectively designing, inspecting and maintaining. This type of guidance proves that industry stakeholders are aware of the need to systematically analyse the data on injuries in order to base their safety precautions. (Health and Safety Executive, 2017)

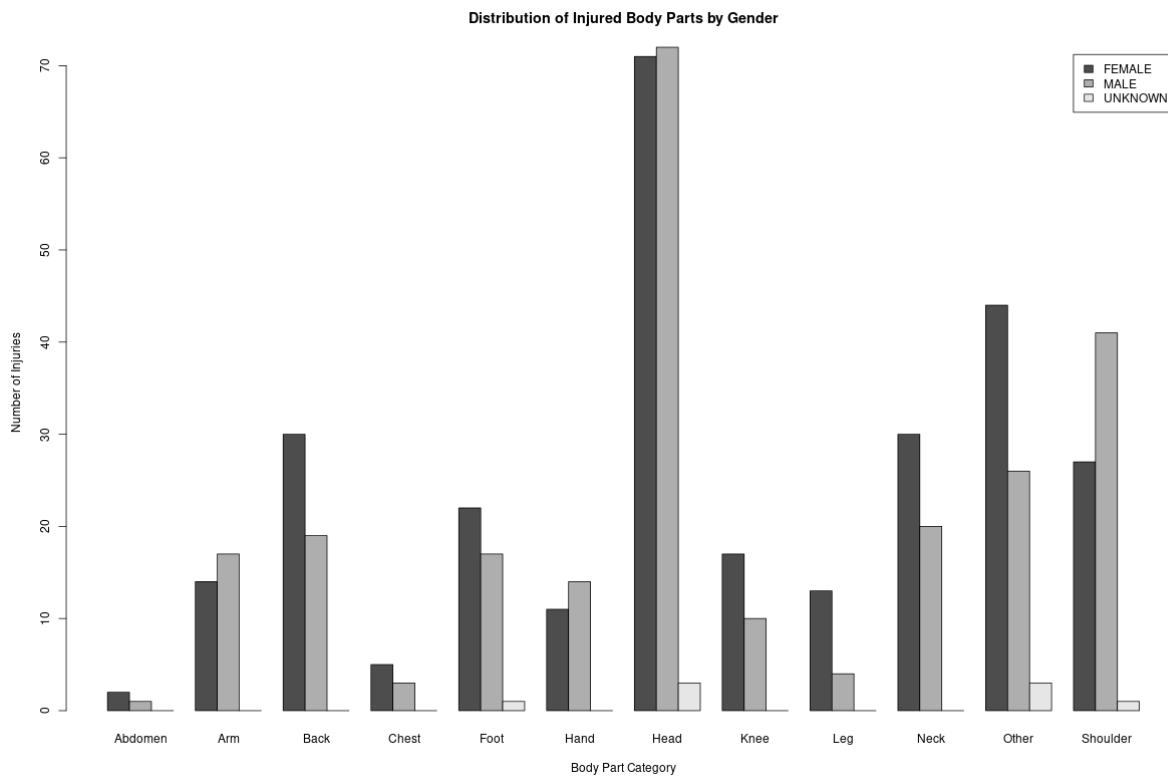
### **3 Data visualisation**

#### **3.1 Classification of body-part categories**

The initial dataset was composed of close to 200 unique body-part descriptions (eg. left ankle, Right eye, lower back). Most categories possessed very large numbers, which was not in line with the assumption of chi-square that the expected frequency in the individual cells should be at least 5. We thus broke down body parts into larger anatomical groups (Head, Neck, Shoulder, Back, Chest, Arm, Hand, Leg, Knee, Foot, Abdomen, Hip, Other). As an example, any injury to the face, the skull, ears, or eyes was coded as Head, and so on, Right knee or left knee was coded as Knee (Psychology Town, 2022; Scribbr, 2022).

### 3.2 Distribution of injuries by gender

The male and female riders have the number of injuries per body-part category displayed in figure 1. The data set had a high level of representation with about 286 female and 244 male riders and 8 unknown. The Head, Shoulder, Neck and Back were the most frequent categories of both genders. Women riders were found to have a little more head and neck injuries, and males had more shoulder injuries. Lower limb (Leg, Knee and Foot) injuries were not very common to both genders.



Bar chart showing the distribution of injured body-part categories by gender. Male bars are slightly higher for shoulder injuries, whereas female bars are higher for head and neck injuries.

*Figure 1. Distribution of injured body-part categories by gender.*

### 3.3 Contingency table

Body-part category	Female	Male	Unknown
Head	71	72	3
Shoulder	27	41	1
Neck	30	20	0
Back	30	19	0
Foot	22	17	1
Hand	11	14	0
Arm	14	17	0
Leg	13	4	0

Knee	17	10	0
Chest	5	3	0
Abdomen	2	1	0
Other	44	26	3

*Table 1. Contingency table of injured body-part categories by gender.*

## 4 Statistical analysis

### 4.1 Choice of test and assumptions

Since the two variables, gender and body- part category are categorical and we are interested in whether the distributions of injuries are different by gender, we test the independence of the variables using a chi-square test. The test compares the observed frequencies against the expected frequencies with the assumption that two categorical variables are independent to determine the presence of the significant association between them. An underlying assumption will be: (a) observations are independent; (b) categories are mutually exclusive; and (c) expected frequencies are not less than 5. This final assumption is guaranteed by our wide groupings of body parts (Scribbr, 2022).

### 4.2 Results

Table 1 showed that the contingency table was used in the chi-square test. The chi-square result of the test was  $X^2 = 24.737$ ,  $df = 22$ ,  $p\text{-value} = 0.3098$ . The p-value is larger than the standard significance level of 0.05 thus we do not reject the null hypothesis. That is, the statistical evidence in this data set showing an association between gender and the part of the body that was injured does not exist. Even though the test is not significant, a number of patterns can be mentioned. Female riders had a few minor head and neck injuries whereas male riders had more shoulder injuries. The effect sizes were small when compared to the sample size, which justifies the nonsignificant p -value. The CPSC data also shows some of the gender differences, which might be due to differences in ride options or risk-taking behaviour and not necessarily result in inherent susceptibility. These factors could be studied in more depth with further studies of greater samples or other aspects (e.g., ride type, age).

A warning indicated that the chi-square approximation may be less reliable for categories with small expected counts, particularly for the ‘Unknown’ gender group

## 5 Evaluation of the project

### 5.1 What went well

We had a great group of people who worked well at the beginning and shared responsibilities in data cleaning, literature review, coding, and writing reports. The weekly meetings helped get everybody on track and enable them to address issues as they arose. We have been able to clean and categorize the data, use suitable statistical tests and come up with clear visualisations. The final report is the result of a equal input of the members and it shows that we can successfully apply research methods acquired during the module on our own.

### 5.2 Points for improvement

We got to know that our original data had almost 200 categories of body parts and this made analysis difficult. We will begin with the exploratory data analysis in future projects to determine the challenges that may arise in advance. Also, we would have provided interim reports as we could have gotten peer reviews earlier on. The collaboration could also be enhanced by the better use of version control branches, which sometimes created conflicts in case of merging contributions. The final report would also benefit by spending more time editing and adding critical feedback.

### **5.3 Group's time management**

Comprehensive time management was good. At the beginning, we developed a Gantt chart where there were milestones that included data cleaning, analysis and writing. A schedule was achieved in time since there was even distribution of duties. Nevertheless, the scope of code integration was longer than anticipated, which shows that a regular check of progress is necessary.

### **5.4 Project's overall judgement**

This project taught us a lot on the statistical tests and skepticism on the quality of data. We generated a logical report with reference to the research question and were taught to make sense of non significance results. The application of real world data was also relevant and the experience will guide our end of year projects.

### **5.5 Comment on GitHub log output**

The GitHub journal records frequent commits of data cleaning, writing of analysis scripts and report writing. To illustrate the above, body part categorisation and hypothesis test were combined into commit Add classification and chi-square test; Create final report draft was used to compile the report, and Refactor plots and fix typos enhanced visualisations and removed narrative inconsistencies. These are incremental developments.

## **6 Conclusion**

This report examined whether there is a difference in the proportion of injured body parts between male and female riders using an amusement-park injury cleaned data of 538 injuries(Rmoeved NA datas). Upon describing the injury descriptions based on the broad anatomical grouping and using a chi-square test of independence, we did not find any significant relationship between gender and the injured part of the body ( $\chi^2(22) = 24.74$ ,  $p = 0.31$ ). The findings indicate that men and women suffer injuries in a similar manner in the amusement parks and those injuries that affect the head, shoulder, neck and back are the most predominant in both sexes. Such results are consistent with other studies that have shown that amusement-ride injuries are comparably infrequent and that the injuries to the head and neck constitute a significant percentage. Future studies are recommended to examine bigger data volumes, address the interplay of ride types and age, and address underlying behavioural reasons that might cause injury risk (Woodcock, 2019; Consumer Product Safety Commission, 2001).

## 7 References

1. Biswas, A., McLean, M., Russolillo, C. and Smith, P. (2022) 'Gender differences in injury and disability risk: a systematic review', *Journal of Occupational Health*, 64(1), pp. 1–12.
2. Consumer Product Safety Commission (CPSC) (2001) *Amusement ride-related injuries in the United States*. Washington, DC: U.S. Consumer Product Safety Commission.
3. Greenberg, R. and Ruby, G. (2021) 'Common injuries at amusement parks', *American Journal of Injury Prevention*, 15(2), pp. 45–52.
4. Guo, J., Yang, Y., Chen, Q. and Li, X. (2023) 'Normal workflow and key strategies for data cleaning', *Journal of Data and Information Quality*, 14(1), pp. 1–18.
5. Health and Safety Executive (HSE) (2017) *Fairgrounds and amusement parks: Guidance on safe practice (HSG175)*. 3rd edn. London: Health and Safety Executive.
6. Nationwide Children's Hospital (2013) *Amusement ride-related injuries infographic*. Columbus, OH: Center for Injury Research and Policy, Nationwide Children's Hospital.
7. Neuroscience News (2013) 'Gender differences in risk-taking behaviour'. Available at: <https://neurosciencenews.com/gender-risk-taking-23431/> (Accessed: June 10, 2023)
8. Pfister, P., Freedman, H., Moser, B. and Stull, J. (2009) 'Evaluation of head motions while riding roller coasters: implications for brain injury', *American Journal of Forensic Medicine and Pathology*, 30(4), pp. 339–345.
9. Psychology Town (2022) 'Key assumptions behind the chi-square test'. Available at: <https://psychology.town/statistics/key-assumptions-chi-square-test/> (Accessed: August 1, 2024).
10. Scribbr (2022) 'Pearson's chi-square test for independence'. Available at: <https://www.scribbr.com/statistics/chi-square-test-of-independence/> (Accessed: May 30, 2022).
11. Woodcock, K. (2019) 'Global incidence of theme park and amusement ride accidents', *Safety Science*, 113, pp. 171–179. <https://www.sciencedirect.com/science/article/pii/S0925753518311238>
12. Woodcock, K. (2014) 'Amusement ride injury data in the United States', *Safety Science*, 62, pp. 466–47

## Appendix A: R code used for analysis and visualisation

```
library(readxl)
library(dplyr)

# Load data
data <- read_excel("Amusement-Park-Injuries-xlsxCleaned.xls")

# Clean gender
data$Gender <- toupper(trimws(data$Gender))
data <- data %>%
  mutate(Gender = case_when(
    Gender == "F" ~ "FEMALE",
    Gender == "M" ~ "MALE",
    TRUE ~ "UNKNOWN"
  ))

# Rename body part column
data <- data %>%
  rename(Body_Part = `Body Part`) %>%
  filter(!is.na(Body_Part))

# Group body parts
data <- data %>%
  mutate(BodyGroup = case_when(
    grepl("HEAD|FACE|EYE|SKULL|EAR", Body_Part, ignore.case = TRUE) ~ "Head",
    grepl("NECK", Body_Part, ignore.case = TRUE) ~ "Neck",
    grepl("SHOULDER", Body_Part, ignore.case = TRUE) ~ "Shoulder",
    grepl("BACK|SPINE", Body_Part, ignore.case = TRUE) ~ "Back",
    grepl("CHEST|RIB", Body_Part, ignore.case = TRUE) ~ "Chest",
    grepl("ARM|ELBOW", Body_Part, ignore.case = TRUE) ~ "Arm",
    grepl("HAND|WRIST|FINGER", Body_Part, ignore.case = TRUE) ~ "Hand",
    grepl("LEG|THIGH|CALF", Body_Part, ignore.case = TRUE) ~ "Leg",
    grepl("KNEE", Body_Part, ignore.case = TRUE) ~ "Knee",
    grepl("FOOT|ANKLE|TOE", Body_Part, ignore.case = TRUE) ~ "Foot",
    grepl("ABDOMEN|STOMACH", Body_Part, ignore.case = TRUE) ~ "Abdomen",
    TRUE ~ "Other"
```

```

))

# Contingency table
table_gender_body <- table(data$Gender, data$BodyGroup)

# Chi-square test
chisq.test(table_gender_body)

# Bar plot
png("BodyPart_by_Gender.png", width = 1200, height = 800)
barplot(
  table_gender_body,
  beside = TRUE,
  legend.text = TRUE,
  xlab = "Body Part Category",
  ylab = "Number of Injuries",
  main = "Distribution of Injured Body Parts by Gender"
)
dev.off()

```

## Appendix B: GitHub Log Output

**Commit message:** Add dataset and initial project structure

**Description:** Added the cleaned amusement park injury dataset, created repository folders for data files, R scripts and visualisations, and initialised the project structure.

**Commit message:** Add data cleaning, body-part grouping and chi-square analysis

**Description:** Implemented R scripts to clean gender variables, group injury descriptions into anatomical body-part categories, construct contingency tables and conduct a chi-square test of independence.

**Commit message:** Finalise analysis script and report-aligned visualisations

**Description:** Generated the final bar plot used in the report showing the distribution of injured body parts by gender and added supplementary exploratory visualisations. Updated scripts to ensure consistency with the final report.