Aircraft-Wildlife Collisions

Tushig Erdenebulgan

Dec 16, 2024

# Aircraft-Wildlife Collisions

A collection of all collisions between aircraft in wildlife that were reported to the US Federal Aviation Administration between 1990 and 1997, with details on the circumstances of the collision.

## Variables

* opid - Three letter identification code for the operator (carrier) of the aircraft
* operator - Name of the aircraft operator
* atype - Make and model of aircraft
* remarks - Verbal remarks regarding the collision
* phase\_of\_flt - Phase of the flight during which the collision occurred:
  + Approach
  + Climb
  + Descent
  + En Route
  + Landing Roll
  + Parked
  + Take-off run
  + Taxi
* ac\_mass - Mass of the aircraft classified as
  + 2250 kg or less (1)
  + 2251-5700 kg (2)
  + 5701-27000 kg (3)
  + 27001-272000 kg (4)
  + above 272000 kg (5)
* num\_engs - Number of engines on the aircraft.
* date - Date of the collision (MM/DD/YYYY).
* time\_of\_day - Light conditions:
  + Dawn
  + Day
  + Dusk
  + Night
* state - Two letter abbreviation of the US state in which the collision occurred.
* height - Feet above ground level
* speed - Knots (indicated air speed).
* effect - Effect on flight:
  + Aborted Take-off
  + Engine Shut Down
  + None, Other
  + Precautionary Landing.
* sky - Type of cloud cover, if any:
  + No Cloud
  + Overcast
  + Some Cloud
* species - Common name for bird or other wildlife
* birds\_seen - Number of birds/wildlife seen by pilot
  + 1
  + 2-10
  + 11-100
  + Over 100 birds\_struck - Number of birds/wildlife struck
  + 0
  + 1
  + 2-10
  + 11-100
  + Over 100.

## Analysis 1, Binomial Proportion

We will estimate the proportion of bird strike incidents where damage was reported.

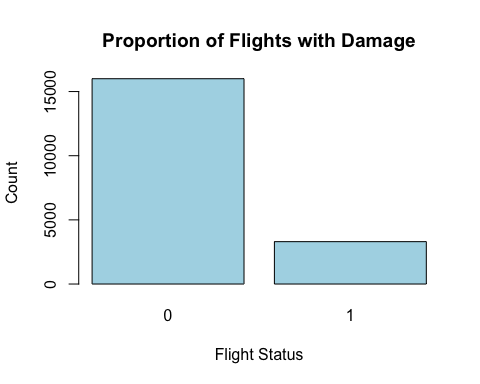
load("birds.rda")  
negation\_patterns <- c("NO DAMAGE", "NOT DAMAGE", "FOUND NO DAMAGE", "ASSUME 1")  
negation\_regex <- paste(negation\_patterns, collapse = "|")  
  
birds$damage\_flag <- ifelse(  
 grepl("DAMAGE", birds$remarks, ignore.case = TRUE) &   
 !grepl(negation\_regex, birds$remarks, ignore.case = TRUE),1,0)  
n\_total <- nrow(birds)  
n\_damage <- sum(birds$damage\_flag)  
n\_damage

## [1] 3302

Total incidents: 19302 Incidents with damage: 3302

### Boxplot of Damage Proportion

barplot(table(birds$damage\_flag),   
 main = "Proportion of Flights with Damage",   
 col = "lightblue",   
 ylab = "Count", xlab = "Flight Status")



p\_hat <- n\_damage/n\_total

We estimated the proportion of bird strike incidents where damage was reported. Out of 19302 reported incidents, 3302 incidents resulted in damage. This yields a sample proportion of 0.171.

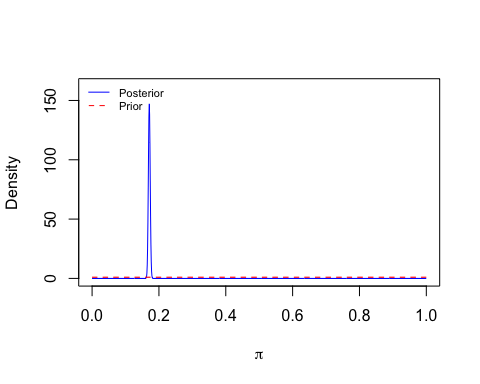
### Bayesian Analysis: Uniform prior (Beta(1, 1))

library(Bolstad)

##   
## Attaching package: 'Bolstad'

## The following objects are masked from 'package:stats':  
##   
## IQR, sd, var

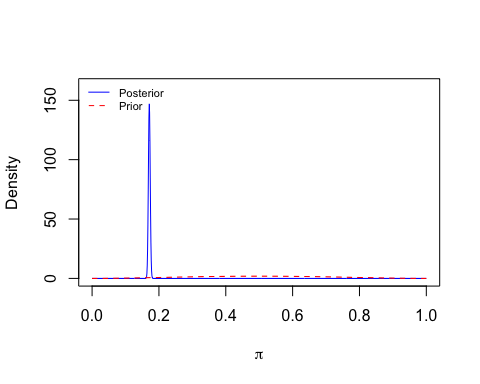
uniform\_prior <- binobp(n\_damage, n\_total, 1, 1)



## Posterior Mean : 0.1711044   
## Posterior Variance : 7.3e-06   
## Posterior Std. Deviation : 0.0027105   
##   
## Prob. Quantile   
## ------ ---------  
## 0.005 0.1641870  
## 0.010 0.1648493  
## 0.025 0.1658244  
## 0.050 0.1666655  
## 0.500 0.1710931  
## 0.950 0.1755821  
## 0.975 0.1764490  
## 0.990 0.1774598  
## 0.995 0.1781498

### Bayesian Analysis: Beta(3, 3) prior

beta\_prior <- binobp(n\_damage, n\_total, 3, 3)



## Posterior Mean : 0.1711726   
## Posterior Variance : 7.3e-06   
## Posterior Std. Deviation : 0.0027106   
##   
## Prob. Quantile   
## ------ ---------  
## 0.005 0.1642548  
## 0.010 0.1649170  
## 0.025 0.1658922  
## 0.050 0.1667334  
## 0.500 0.1711612  
## 0.950 0.1756504  
## 0.975 0.1765174  
## 0.990 0.1775283  
## 0.995 0.1782183

Using a uniform prior, the estimated proportion of incidents with damage is 0.1711044 . The probability is 95% that the true proportion lies between 0.1658244 and 0.1764490

With a Beta(3,3) prior, the estimated proportion is 0.1711726 with a 95% credible interval of 0.1658922, 0.1765174

### Frequentist Confidence Interval

# Frequentist Confidence Interval (Agresti-Coull)  
z <- qnorm(0.975)  
se <- sqrt(p\_hat \* (1-p\_hat)/n\_total)  
lower\_bound <- p\_hat-z\*se  
upper\_bound <- p\_hat+z\*se  
cat("Using a frequentist approach, the 95% confidence interval for the proportion is", lower\_bound, "-", upper\_bound)

## Using a frequentist approach, the 95% confidence interval for the proportion is 0.1657579 - 0.1763828

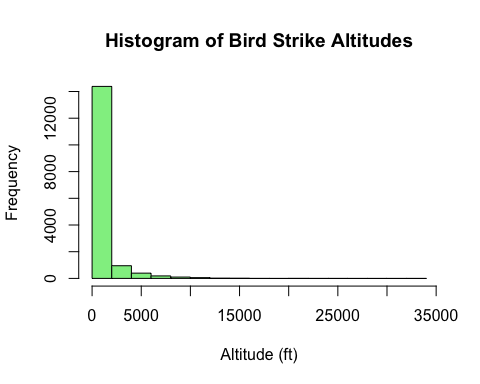
Using a frequentist approach, the 95% confidence interval for the proportion is ( 0.1657579, 0.1763828).

## Analysis #2, Estimate a mean

valid\_altitudes <- birds$height[!is.na(birds$height)]  
mean\_altitude <- mean(valid\_altitudes)  
sd\_altitude <- sd(valid\_altitudes)  
n\_altitudes <- length(valid\_altitudes)  
cat("The mean altitude of bird strikes is", mean\_altitude, "feet, with a standard deviation of", sd\_altitude)

## The mean altitude of bird strikes is 754.6778 feet, with a standard deviation of 1795.805

hist(valid\_altitudes,   
 main = "Histogram of Bird Strike Altitudes",   
 xlab = "Altitude (ft)", col = "lightgreen", border = "black")



We analyzed the altitude of bird strikes. The mean altitude of bird strikes is 754.68 feet, with a standard deviation of 1795.81 feet.

### Bayesian inference

Next, we’ll estimate the population mean altitude of bird strikes using a Bayesian approach with a Normal prior distribution. We assume a prior mean mu of 2000 ft and a prior standard deviation sigma of 5000 ft, reflecting broad uncertainty about the mean altitude.

prior\_mu <- 2000  
prior\_sigma <- 5000  
posterior\_mean <- ((prior\_mu/prior\_sigma^2)+(mean\_altitude\*n\_altitudes/sd\_altitude^2)) /  
 ((1/prior\_sigma^2)+(n\_altitudes/sd\_altitude^2))  
posterior\_variance <- 1/((1/prior\_sigma^2)+(n\_altitudes/sd\_altitude^2))  
posterior\_sd <- sqrt(posterior\_variance)  
cat("Using a Bayesian approach with a normal prior, the posterior mean altitude is", posterior\_mean,   
 "feet with a 95% credible interval between", posterior\_mean-1.96\*posterior\_sd,   
 "and", posterior\_mean+1.96\*posterior\_sd)

## Using a Bayesian approach with a normal prior, the posterior mean altitude is 754.6878 feet with a 95% credible interval between 726.9559 and 782.4197

Using a Bayesian approach with a normal prior, the posterior mean altitude is 754.6878 feet with a 95% credible interval between 726.9559 and 782.4197

### Frequentist inference

se <- sd\_altitude / sqrt(n\_altitudes)  
z <- qnorm(0.975)  
ci\_lower <- mean\_altitude-z\*se  
ci\_upper <- mean\_altitude+z\*se  
cat("Using a frequentist approach, the 95% confidence interval for the mean altitude is", ci\_lower, "-", ci\_upper, "feet")

## Using a frequentist approach, the 95% confidence interval for the mean altitude is 726.9463 - 782.4093 feet

## Analysis #3, Compare Two Means

We will compare the altitude of bird strikes during the Climb phase versus the Approach phase of flight.

alt\_climb <- birds$height[birds$phase\_of\_flt == "Climb" & !is.na(birds$height)]  
alt\_approach <- birds$height[birds$phase\_of\_flt == "Approach" & !is.na(birds$height)]  
  
mean\_climb <- mean(alt\_climb, na.rm = TRUE)   
mean\_approach <- mean(alt\_approach, na.rm = TRUE)  
  
cat("The mean altitude for Climb phase is", mean\_climb, "feet, while for Approach phase it is", mean\_approach, "feet.")

## The mean altitude for Climb phase is 1099.403 feet, while for Approach phase it is 882.7173 feet.

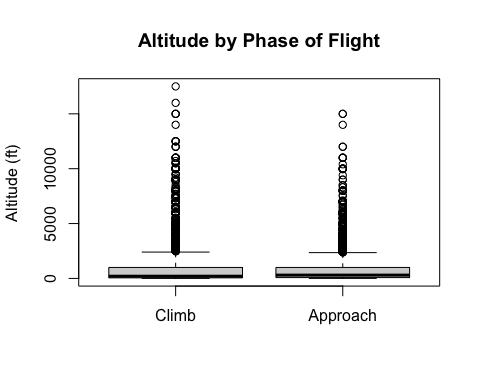
We compared the altitudes of bird strikes during the **Climb** and **Approach** phases of flight. The mean altitude for the Climb phase is 1099.403 feet, while for the Approach phase it is 882.7173 feet.

# Frequentist t-test  
t\_test <- t.test(alt\_climb, alt\_approach)  
cat("The frequentist t-test shows a p-value of", t\_test$p.value,   
 ", indicating whether the difference in means is statistically significant.")

## The frequentist t-test shows a p-value of 3.026049e-07 , indicating whether the difference in means is statistically significant.

A frequentist t-test yielded a p-value of 3.026049e-07, indicating whether the difference in mean altitudes is statistically significant.

boxplot(alt\_climb, alt\_approach, names = c("Climb", "Approach"),  
 main = "Altitude by Phase of Flight", ylab = "Altitude (ft)")

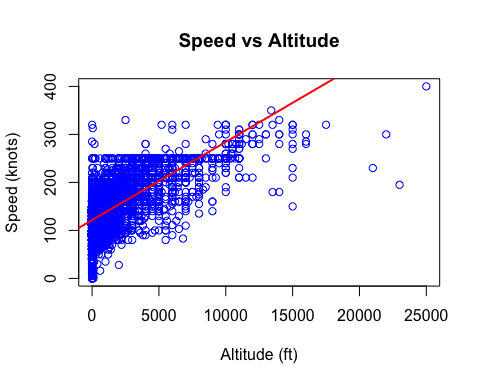
 ## Analysis 4: Regression for Speed vs. Altitude We will fit a regression model to predict aircraft speed based on the altitude of bird strikes.

valid\_data <- birds[!is.na(birds$height) & !is.na(birds$speed), ]  
model <- lm(speed ~ height, data = valid\_data)  
model\_summary <- summary(model)  
slope <- model\_summary$coefficients["height", "Estimate"]   
se\_slope <- model\_summary$coefficients["height", "Std. Error"]  
t\_crit <- qt(0.975, df = model\_summary$df[2])   
  
ci\_lower <- slope-t\_crit\*se\_slope  
ci\_upper <- slope+t\_crit\*se\_slope  
  
  
cat("The regression slope is", slope,   
 "with a 95% confidence interval between", ci\_lower, "and", ci\_upper)

## The regression slope is 0.0162887 with a 95% confidence interval between 0.01597746 and 0.01659993

We fit a regression model to predict aircraft speed based on the altitude of bird strikes. The slope of the regression line is 0.0162887, indicating the rate at which speed changes with altitude. The 95% confidence interval for the slope is (0.01597746 and 0.01659993).

if (nrow(valid\_data) > 0) {  
 # Plot the data points  
 plot(valid\_data$height, valid\_data$speed,   
 main = "Speed vs Altitude",   
 xlab = "Altitude (ft)",   
 ylab = "Speed (knots)",   
 col = "blue")  
  
 intercept <- summary(model)$coefficients[1, "Estimate"]  
 slope <- summary(model)$coefficients[2, "Estimate"]  
 abline(a = intercept, b = slope, col = "red", lwd = 2)  
} else {  
 cat("No valid data available for plotting")  
}



## Analysis 5: Regression by Time of Day

We will fit separate regression models to compare the relationship between speed and altitude for bird strikes that occurred during the day versus at night.

day\_data <- birds[birds$time\_of\_day == "Day" & !is.na(birds$height) & !is.na(birds$speed), ]  
night\_data <- birds[birds$time\_of\_day == "Night" & !is.na(birds$height) & !is.na(birds$speed), ]  
  
model\_day <- lm(speed ~ height, data = day\_data)  
model\_night <- lm(speed ~ height, data = night\_data)  
  
cat("The slope for Daytime regression is", coef(model\_day)[2],   
 "with a 95% CI of", confint(model\_day)[2, 1], "-", confint(model\_day)[2, 2])

## The slope for Daytime regression is 0.01522018 with a 95% CI of 0.01475369 - 0.01568667

cat("The slope for Nighttime regression is", coef(model\_night)[2],   
 "with a 95% CI of", confint(model\_night)[2, 1], "-", confint(model\_night)[2, 2])

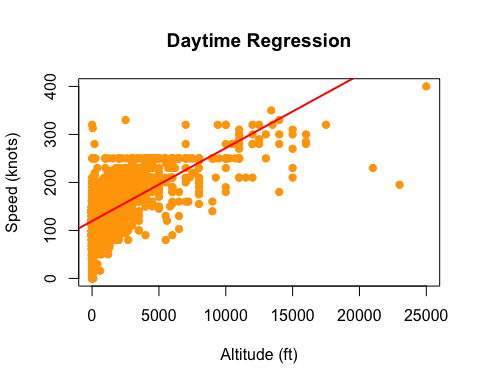
## The slope for Nighttime regression is 0.01579272 with a 95% CI of 0.01527198 - 0.01631345

We fit separate regression models for bird strikes during the day and night to compare the relationship between speed and altitude. The regression slope for daytime strikes is 0.01522018 with higher speed at lower altitudes(95% CI of 0.01475369 - 0.01568667). For nighttime strikes, the slope is 0.01579272. The nighttime regression indicates higher speeds at a given altitude, as reflected in the intercept (95% CI of 0.01527198 - 0.01631345).

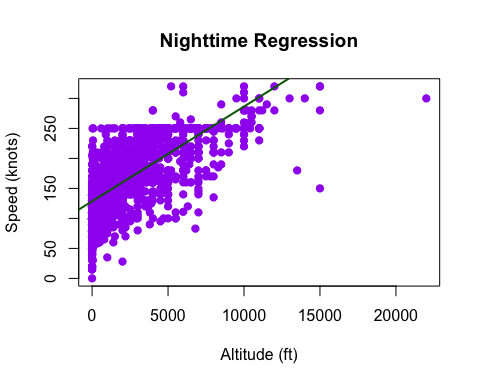
The following plots visualize the regression results:

* **Daytime Regression**:
  + Orange points represent the data, and the red line shows the regression relationship.

if (nrow(day\_data) > 0) {  
 # Plot the data points for Day  
 plot(day\_data$height, day\_data$speed,   
 main = "Daytime Regression",   
 xlab = "Altitude (ft)",   
 ylab = "Speed (knots)",   
 col = "orange",   
 pch = 19)  
   
 # Extract intercept and slope for Day  
 intercept\_day <- summary(model\_day)$coefficients[1, "Estimate"]  
 slope\_day <- summary(model\_day)$coefficients[2, "Estimate"]  
   
 # Add regression line for Day  
 abline(a = intercept\_day, b = slope\_day, col = "red", lwd = 2)  
} else {  
 cat("No valid data available for daytime plotting.")  
}

 - **Nighttime Regression**: - Purple points represent the data, and the green line shows the regression relationship.

if (nrow(night\_data) > 0) {  
 # Plot the data points for Night  
 plot(night\_data$height, night\_data$speed,   
 main = "Nighttime Regression",   
 xlab = "Altitude (ft)",   
 ylab = "Speed (knots)",   
 col = "purple",   
 pch = 19)  
   
 # Extract intercept and slope for Night  
 intercept\_night <- summary(model\_night)$coefficients[1, "Estimate"]  
 slope\_night <- summary(model\_night)$coefficients[2, "Estimate"]  
   
 # Add regression line for Night  
 abline(a = intercept\_night, b = slope\_night, col = "darkgreen", lwd = 2)  
} else {  
 cat("No valid data available for nighttime plotting.")  
}

 The relationship between altitude and speed is slightly stronger during nighttime bird strikes, as indicated by a higher R-squared value. Additionally, speeds at any given altitude tend to be higher at night compared to the day, as seen from the higher intercept.

## Footnote

This report analyzes FAA wildlife strike data from 1990 to 1997 using complete cases with non-missing values for key variables like altitude, speed, and flight phase. Both Bayesian and frequentist methods are applied to provide reliable estimates and intervals. The results should be interpreted carefully, as the data is observational and cannot show cause-and-effect relationships.