

# PGA and LPGA Tour 2008 Driving Accuracy

2416963E

## 1 Introduction

## 2 Exploratory Data Analysis

Summary statistics of the PGA and LPGA driving accuracies.

Table 1: Summary statistics on driving accuracy by Gender of 354 professional golfers.

Gender	n	Mean	St.Dev	Min	Q1	Median	Q3	Max
Female	157	67.6	5.8	49.3	64.5	68.3	71.3	79.8
Male	197	63.4	5.5	49.0	59.5	63.1	66.9	80.4

Table 1 shows that there are slightly more Men in the sample (197 compared to 157) and that the summaries of the of the male golfers were consistently similar ot that of the female golfers. These similarities can be seen in the following boxplots which summarise the distributions of the drive accuracies of the male and female professional golfers in the 2008 PGA and LPGA tours.

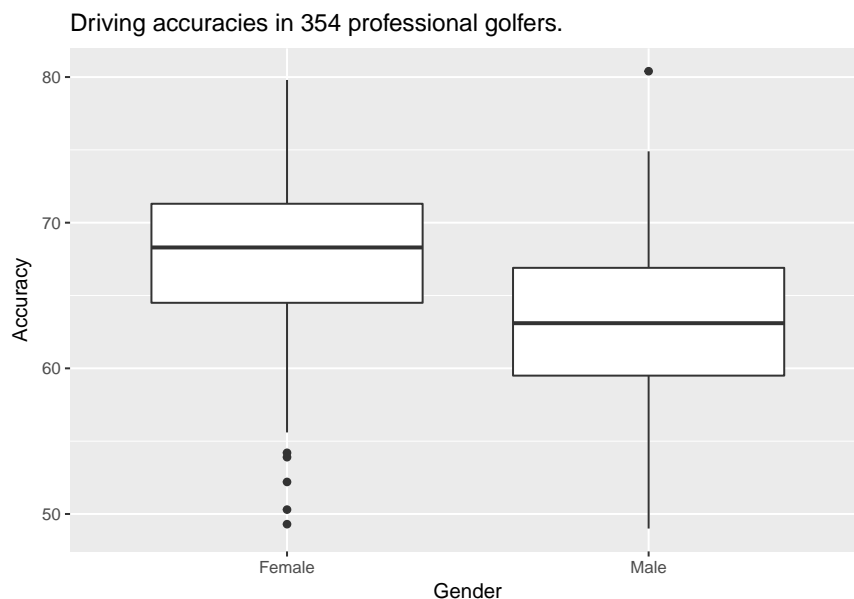


Figure 1: Driving accuracy by gender.

The boxplot shows that female golfers have a higher drive accuracies on average compared to male drive accuracies. The drive accuracies of both genders are about evenly distributed, which is also reflected in the standard deviation of female and male drive accuracies (5.8% compared to 5.5% respectively). There are also a number of potential outliers (5 female and 1 male) as shown by the points beyond the “whiskers” of the boxplots.

### 3 Formal Data Analysis

To analyse the PGA and LPGA drive accuracy data formally, a linear model is fit to the data.

$$\widehat{\text{Accuracy}} = \hat{\alpha} + \hat{\beta}_{\text{Female}} \cdot \mathbb{I}_{\text{Female}}(x)$$

where,

- the intercept  $\hat{\alpha}$  is the mean drive accuracy for the baseline category of Males;
- $\hat{\beta}_{\text{Female}}$  is the difference in the mean heart weight of a Females relative to the baseline category Males;
- $\mathbb{I}_{\text{Female}}(x)$  is an indicator function such that

$$\mathbb{I}_{\text{Female}}(x) = \begin{cases} 1 & \text{if Sex of } x\text{th observation is Female,} \\ 0 & \text{Otherwise.} \end{cases}$$

When this model is fitted to the data, the following estimates of  $\alpha$  (**intercept**) and  $\beta_{\text{Female}}$  (**GenderFemale**) are returned:

Table 2: Estimates of the param  
model.

---

term

intercept

GenderFemale

Hence the model estimates the average drive accuracy of male golfers is grams (which agrees with the sample mean report

---

Before we can proceed to use the fitted model (for example to perform statistical inference) we must check the assumptions of the model. These are best considered in light of the residual plots in Figure 2.

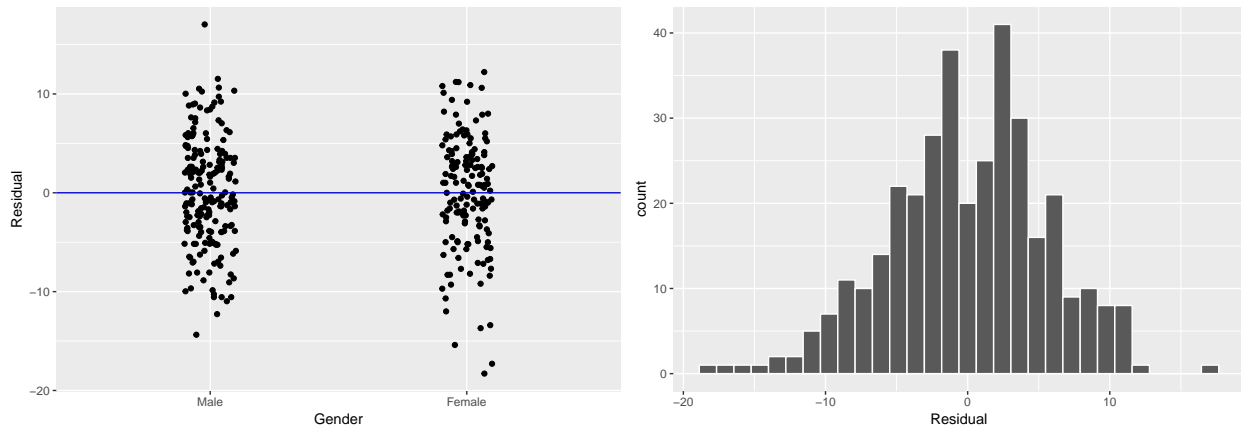


Figure 2: Scatterplot of the residuals by Gender (left) and a histogram of the residuals (right).

The scatterplots show an approximately even spread of the residuals above and below the zero line for each gender, and hence the assumption that the residuals have mean zero appears valid. The assumption of constant variance within the two genders is not supported, however, as the spread of the residuals in the vertical scatter of the male cats is considerably more than that of the females (as was noted above when the standard deviations were considered). The histogram supports the assumption of normally distributed errors in the model, with the exception of a potential outlier.

## 4 Conclusions