

# PGA and LPGA Tour 2008 Driving Accuracy

2416963E

## 1 Introduction

The opening stroke (or “drive”) in golf is an important metric in determining how well a golfer performs. For a golfer to be competitive it is important for the drive to accurately land on the fairway, minimizing the number strokes on the hole. Data on 354 golfers (male and female) from the 2008 PGA and LPGA tours are analysed through their drive accuracy (Accuracy) in percent and Gender (Gender).

In particular, this report presents numerical and graphical summaries of the drive accuracies of the professional golfers and fits a linear model to estimate the difference, on average, between the drive accuracies of professional female and male golfers.

## 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.

The boxplot in Figure 1 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.

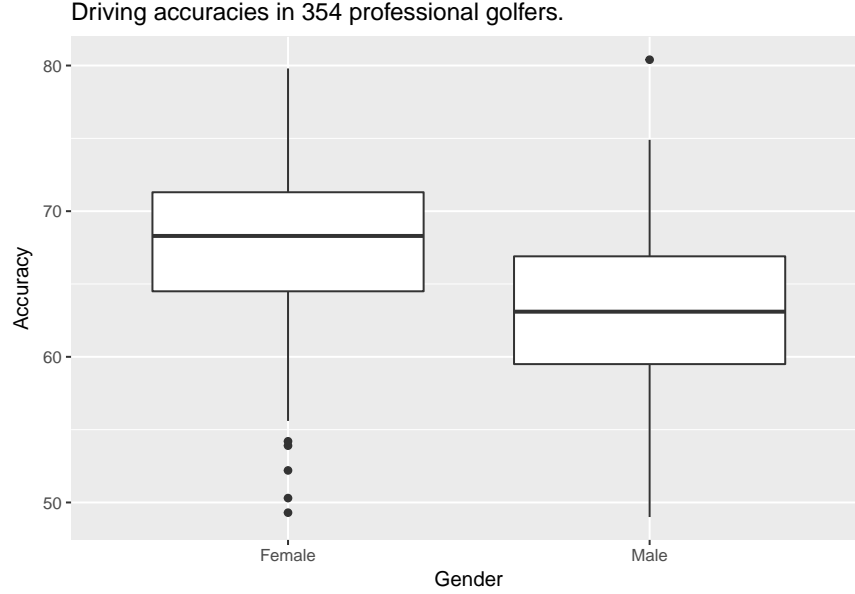


Figure 1: Driving accuracy by gender.

### 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 parameters from the fitted linear regression model.

term	estimate
intercept	63.365
GenderFemale	4.226

Hence the model estimates the average drive accuracy of male golfers is grams (which agrees with the sample mean reported in Table 1) and that the female golfer drive accuracy are, on average, percent higher than the male golfers' drive accuracy.

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.

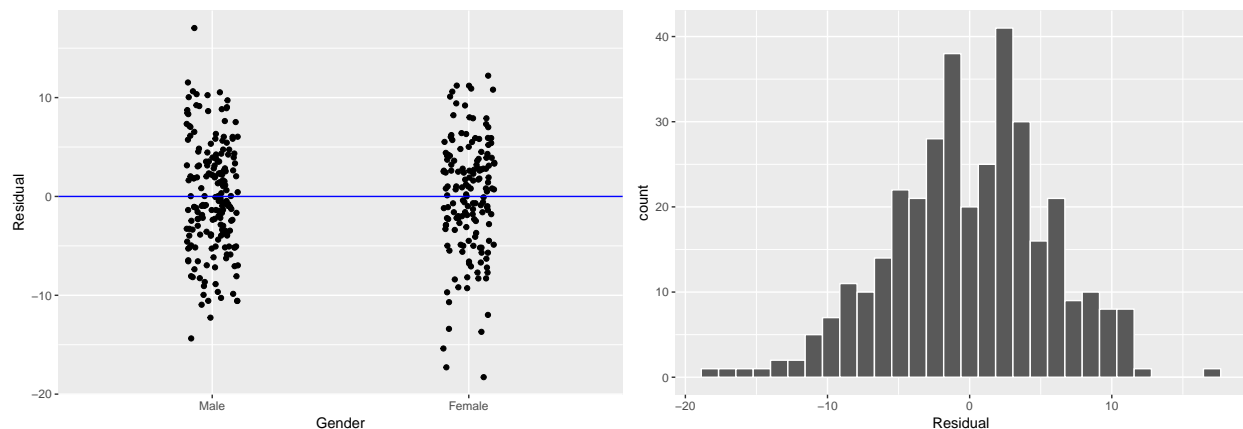


Figure 2: Scatterplot of the residuals by Gender (left) and a histogram of the residuals (right).

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 supported, as the spread of the residuals in the vertical scatter of the male golfers is similar to that of the female golfers (as was noted above when the standard deviations were considered). The histogram supports the assumption of normally distributed errors about mean zero in the model.

The assumptions of the Normal model appear to hold and we can now perform statistical inference in the form of a confidence interval. A confidence interval  $(-5.41, -3.04)$  indicates that the true difference in means is not equal to zero and we conclude that there is a difference in drive accuracies between professional female and professional male golfers in the 2008 LPGA and PGA tours, respectively.

## 4 Conclusions

In summary, we have estimated that, on average, professional female golfers in the 2008 LPGA Tour have drive accuracies 4.23 percent more than professional male golfers in the 2008 PGA Tour. In particular, we estimate the average drive accuracy of professional male golfers is 63.36 percent and the average drive accuracy of professional female golfers is 67.59 percent with a confidence interval between the two means  $(-5.41, -3.04)$  indicating there is a true difference in means.