

# Regression Model

W241: Fox, George, McCleary

7/27/2022

```
d <- data.table(readxl::read_xlsx("Final_Data_clean.xlsx"))
d[is.na(Value), account_disabled := T]
d[!is.na(Value), account_disabled := F]
d
```

```
##      Due Date      Researcher Sample ID      GMail Address
##  1: 2022-06-12  Hannah George    001 W241.HG.alex.doe.001@gmail.com
##  2: 2022-06-12  Hannah George    001 W241.HG.alex.doe.001@gmail.com
##  3: 2022-06-12  Hannah George    001 W241.HG.alex.doe.001@gmail.com
##  4: 2022-06-12  Hannah George    001 W241.HG.alex.doe.001@gmail.com
##  5: 2022-06-12  Hannah George    002 W241.HG.alex.doe.002@gmail.com
## ---
## 500: 2022-07-24  Carlie McCleary    041 W241.CM.alex.doe.041@gmail.com
## 501: 2022-07-24  Carlie McCleary    042 W241.CM.alex.doe.042@gmail.com
## 502: 2022-07-24  Carlie McCleary    042 W241.CM.alex.doe.042@gmail.com
## 503: 2022-07-24  Carlie McCleary    042 W241.CM.alex.doe.042@gmail.com
## 504: 2022-07-24  Carlie McCleary    042 W241.CM.alex.doe.042@gmail.com
##      Age Group  Birthday      Gender  Website      Metric
##  1:    Young 1999-06-20      Male  YouTube    Average Bias
##  2:    Young 1999-06-20      Male  YouTube  Average Reliability
##  3:    Young 1999-06-20      Male  Facebook    Average Bias
##  4:    Young 1999-06-20      Male  Facebook  Average Reliability
##  5:    Young 1992-02-06    Female  YouTube    Average Bias
## ---
## 500:      Old 1954-03-19      Female  Facebook  Average Reliability
## 501:      Old 1953-03-06  Prefer Not To Answer  YouTube    Average Bias
## 502:      Old 1953-03-06  Prefer Not To Answer  YouTube  Average Reliability
## 503:      Old 1953-03-06  Prefer Not To Answer  Facebook    Average Bias
## 504:      Old 1953-03-06  Prefer Not To Answer  Facebook  Average Reliability
##      Value account_disabled
##  1:  1.771667          FALSE
##  2: 43.026667          FALSE
##  3: -6.256667          FALSE
##  4: 44.001667          FALSE
##  5: -3.399167          FALSE
## ---
## 500:      NA          TRUE
## 501: -4.609167          FALSE
## 502: 45.238333          FALSE
## 503:      NA          TRUE
## 504:      NA          TRUE
```

```
# Is there differential attrition between Facebook and Google?
website_diff_attrition_model <- d[, lm(account_disabled ~ as.factor(Website))]
website_diff_attrition_coeftest <- coeftest(website_diff_attrition_model,
                                           vcovHC(website_diff_attrition_model))
```

```
# Stargazer output.
stargazer(website_diff_attrition_model, header=F)
```

```
##
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \hline
## \hline \hline
## & \multicolumn{1}{c}{\textit{Dependent variable:}} \hline
## \cline{2-2}
## \hline \hline
## as.factor(Website)YouTube &  $-\$0.413^{***}$  \hline
## & (0.031) \hline
## & \hline
## Constant &  $0.413^{***}$  \hline
## & (0.022) \hline
## & \hline
## \hline \hline
## Observations & 504 \hline
##  $R^2$  & 0.260 \hline
## Adjusted  $R^2$  & 0.259 \hline
## Residual Std. Error & 0.349 (df = 502) \hline
## F Statistic &  $176.378^{***}$  (df = 1; 502) \hline
## \hline
## \hline \hline
## \textit{Note:} & \multicolumn{1}{r}{ $^*p < 0.1$ ;  $^{**}p < 0.05$ ;  $^{***}p < 0.01$ } \hline
## \end{tabular}
## \end{table}
```

```
# Does demographic information influence Facebook's decision to ban an account?
facebook_diff_attrition_model <- d[Website=="Facebook", lm(account_disabled ~ as.factor(Gender) + as.factor(Website))]
facebook_diff_attrition_coeftest <- coeftest(facebook_diff_attrition_model,
                                           vcovHC(facebook_diff_attrition_model))
```

```
# Stargazer output.
stargazer(facebook_diff_attrition_model, header=F)
```

```
##
## \begin{table}[!htbp] \centering
##   \caption{}
##   \label{}
## \begin{tabular}{@{\extracolsep{5pt}}lc}
## \hline
## \hline \hline
## & \multicolumn{1}{c}{\textit{Dependent variable:}} \hline
## \cline{2-2}
```

Table 1: Differential Attrition between YouTube and Facebook

	<i>Dependent variable:</i>
	Account Disabled
Website: YouTube	-0.413*** (0.031)
Constant	0.413*** (0.022)
Observations	504
R <sup>2</sup>	0.260
Adjusted R <sup>2</sup>	0.259
Residual Std. Error	0.349 (df = 502)
F Statistic	176.378*** (df = 1; 502)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01

```
## \[-1.8ex] & account\_disabled \\
## \hline \[-1.8ex]
## as.factor(Gender)Male & $-0.048 \\
## & (0.076) \\
## & \\
## as.factor(Gender)Prefer Not To Answer & 0.071 \\
## & (0.076) \\
## & \\
## as.factor('Age Group')Young & 0.063 \\
## & (0.062) \\
## & \\
## Constant & 0.373$^{***}$ \\
## & (0.062) \\
## & \\
## \hline \[-1.8ex]
## Observations & 252 \\
## R$^{2}$ & 0.014 \\
## Adjusted R$^{2}$ & 0.002 \\
## Residual Std. Error & 0.493 (df = 248) \\
## F Statistic & 1.177 (df = 3; 248) \\
## \hline
## \hline \[-1.8ex]
## \textit{Note:} & \multicolumn{1}{r}{\textit{$^{*}$p<$0.1$; $^{**}$p<$0.05$; $^{***}$p<$0.01$}} \\
## \end{tabular}
## \end{table}
```

Summary: Attrition is significantly different between YouTube and Facebook, however, Facebook attrition is not significantly different due to any of the collected demographic variables.

Table 2: Differential Attrition on Facebook between Demographics

	<i>Dependent variable:</i>
	Account Disabled
Gender: Male	-0.048 (0.076)
Gender: Prefer Not To Answer	0.071 (0.076)
Age Group: Young	0.063 (0.062)
Constant	0.373*** (0.062)
Observations	252
R <sup>2</sup>	0.014
Adjusted R <sup>2</sup>	0.002
Residual Std. Error	0.493 (df = 248)
F Statistic	1.177 (df = 3; 248)
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01