45871 - Assignment 4 solutions

2025-12-01 17:34:26.582442

First we prepare data for our analysis.

```
rm(list = ls())
setwd("C:/Users/pxue/Box/TA/TIS/2025 mini 2/hw4")

# packages for data analysis
library(tidyverse)
library(ggplot2)

# load the data from first-price auctions
library(readxl)
landing_data <- read_excel("landing_game_from 2022To2025.xlsx")

# rename columns
colnames(landing_data)[1] = 'year'
colnames(landing_data)[3] = 'session'
colnames(landing_data)[4] = 'subject'
colnames(landing_data)[5] = 'subject_id'
colnames(landing_data)[6] = 'valuation' # (a)
colnames(landing_data)[7] = 'n_gates_start' # (b)
colnames(landing_data)[8] = 'n_orders' # (c)
colnames(landing_data)[9] = 'n_orders_market' # (d)
colnames(landing_data)[10] = 'n_transactions' # (e)
colnames(landing_data)[11] = 'n_gates_end'
colnames(landing_data)[12] = 'last_session_completed' # (g)
colnames(landing_data)[13] = 'wealth_start' # (h)
colnames(landing_data)[14] = 'wealth_end' # (i)
colnames(landing_data)[15] = 'cash_end' # (j)

# dropping unused columns
landing_data = landing_data[ , c(1, 3:15)]

# use column 'n_gates_end' to create (f) whether they owned more than two gates at the end of
the session
landing_data$more_than_two = (landing_data$n_gates_end > 2)
# modify column 'last_session_completed' to make it take value 0 instead of NA
landing_data$last_session_completed[is.na(landing_data$last_session_completed)] = 0
```

Question 1

Only the coefficient of n-gates-start (b) is significant. Its quantitative impact is substantial: having one more gate at the start decreases profit by 188,938 (since the monetary unit here is hundreds of thousands of dollars), all else equal.

The coefficient associated with last-session-completed (g) is not statistically significant, so there is no evidence

that trading more improves traders' performance.

R-squared is 0.06, so 6-percent of the variation in profits is explained by this regression.

```
# create 'profit' column
landing_data$profit = landing_data$wealth_end - landing_data$wealth_start

# regress profit on (b)-(g), omit (f)
reg1 <- lm(profit ~ n_gates_start + n_orders + n_orders_market + n_transactions + last_sessio
n_completed, data = landing_data)
summary(reg1)
```

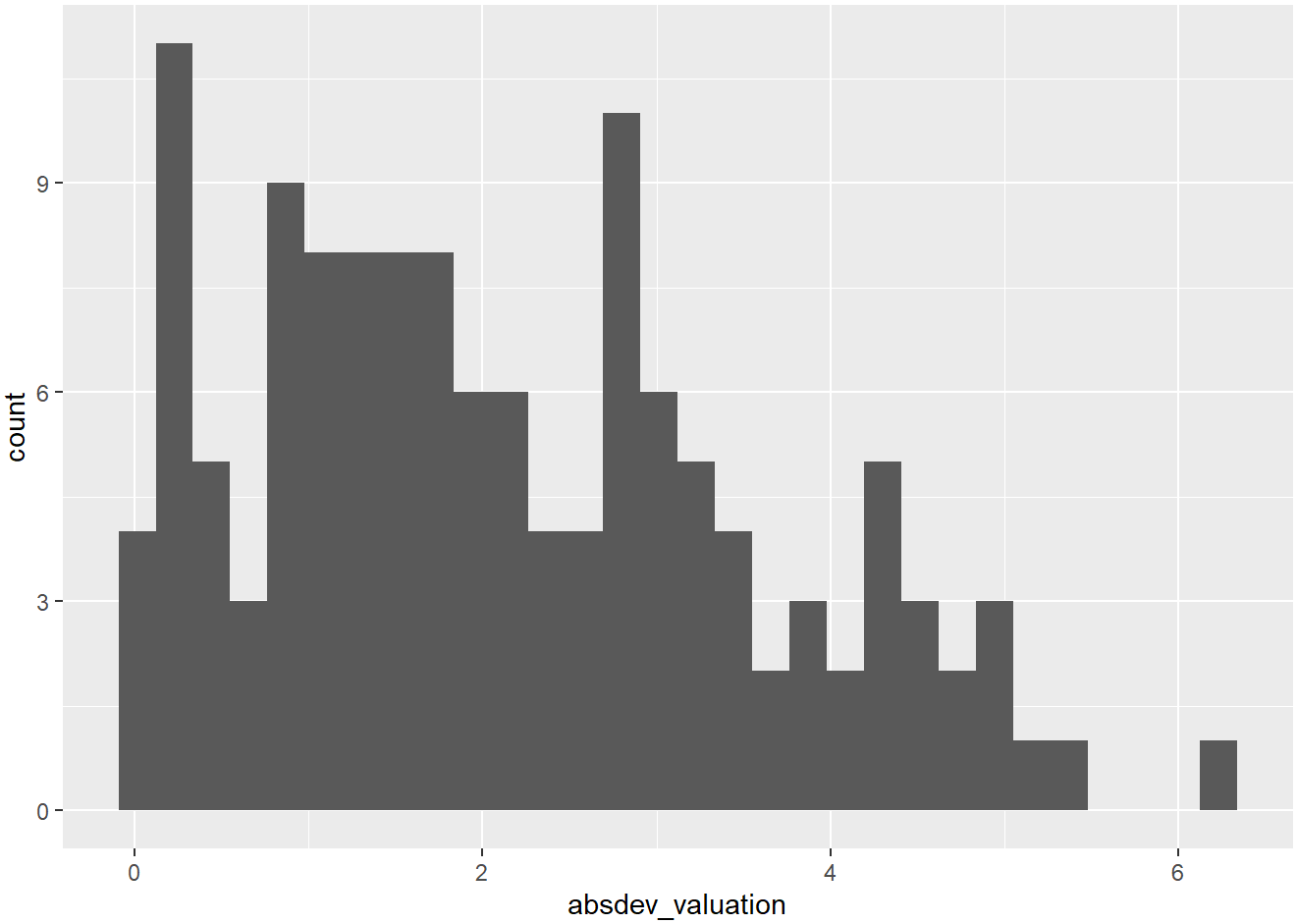
```
##
## Call:
## lm(formula = profit ~ n_gates_start + n_orders + n_orders_market +
##     n_transactions + last_session_completed, data = landing_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -53.880  -1.478   0.878   2.761  12.975
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              4.10911    1.55058   2.650  0.00908 **
## n_gates_start           -1.88938    0.85113  -2.220  0.02822 *
## n_orders                -0.04973    0.26161  -0.190  0.84955
## n_orders_market         -1.22903    1.11474  -1.103  0.27234
## n_transactions           0.04221    0.99545   0.042  0.96625
## last_session_completed -1.13359    1.31385  -0.863  0.38989
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.353 on 126 degrees of freedom
## Multiple R-squared:  0.0645, Adjusted R-squared:  0.02737
## F-statistic: 1.737 on 5 and 126 DF,  p-value: 0.1308
```

Question 2

```
# calculate mean valuation for each session
landing_data = landing_data %>%
  group_by(session) %>%
  mutate(mean_valuation = mean(valuation))

# calculate each player's absolute deviation from session mean valuation
landing_data$absdev_valuation = abs(landing_data$valuation - landing_data$mean_valuation)

# plot the distribution of this second variable over all the sessions
ggplot(data = landing_data, aes(x = absdev_valuation)) + geom_histogram(bins=30)
```




```
# total number of gates in each session
landing_data = landing_data %>%
  group_by(session) %>%
  mutate(tot_gates = sum(n_gates_start))
```

Question 3

Now all the coefficients are insignificant. The coefficient of absdev-valuation is statistically insignificant (and the estimate is negative), so there is no evidence in favor of the hypothesis that those traders with very low (or high) valuations can profit a lot while those with valuations around the mean valuation cannot typically benefit much.

```
# add two variables created in Q2 to the regression in Q1
reg2 <- lm(profit ~ n_gates_start + n_orders + n_orders_market + n_transactions + last_sessio
n_completed + absdev_valuation + tot_gates, data = landing_data)
summary(reg2)
```

```
##
## Call:
## lm(formula = profit ~ n_gates_start + n_orders + n_orders_market +
##     n_transactions + last_session_completed + absdev_valuation +
##     tot_gates, data = landing_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -53.481  -1.665   0.889   2.652  12.706
##
## Coefficients:
##                         Estimate Std. Error t value Pr(>|t|)
## (Intercept)              5.09933    2.19098   2.327   0.0216 *
## n_gates_start           -1.67419    0.97870  -1.711   0.0897 .
## n_orders                -0.07584    0.26693  -0.284   0.7768
## n_orders_market         -1.13257    1.14433  -0.990   0.3242
## n_transactions           0.05943    1.01238   0.059   0.9533
## last_session_completed -1.17858    1.32433  -0.890   0.3752
## absdev_valuation        -0.21713    0.46090  -0.471   0.6384
## tot_gates               -0.15606    0.36036  -0.433   0.6657
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.399 on 124 degrees of freedom
## Multiple R-squared:  0.06764,    Adjusted R-squared:  0.01501
## F-statistic: 1.285 on 7 and 124 DF,  p-value: 0.263
```