Statistical Analysis of Player Behaviour in Online Poker

Roman Prokhorov, Maksym Holovin, Nazar Pasichnyk

1. Research Topic

We aim to analyze patterns in online poker player behavior using real transaction and gameplay data. Our study will examine relationships between player demographics, betting patterns, and financial outcomes to understand different player types and their gambling strategies.

2. Data Source

Source: The Transparency Project (www.thetransparencyproject.org)

Provider: Division on Addiction, Cambridge Health Alliance, Harvard Medical School

Dataset: "Second Session at the Virtual Poker Table"

Sample Size: 5,028 online poker players who registered in February 2015

Reference: Tom, M. A., et al. (2022). Second Session at the Virtual Poker Table: A Contemporary Study of Actual Online Poker Activity. *Journal of Gambling Studies.* https://doi.org/10.1007/s10899-022-10147-1

3. Data Description

The dataset consists of five interconnected CSV files:

- **Demographics** (5,028 players): Age, Gender, Country of residence (encoded using ISO 3166-1 country codes)

- **Cash Games** (51,763 daily records): Stakes wagered, winnings, number of sessions

- **Tournaments** (82,831 daily records): Entry fees, prizes, number of tournaments

- **Deposits** (295,119 transactions): Amount, payment method, transaction status

- **Withdrawals** (32,307 transactions): Amount, payment method, transaction status

Key variables include player demographics, daily gambling activity (stakes and winnings), and financial transactions (deposits and withdrawals in Euros). Country information is mapped from numeric ISO 3166-1 codes to country names for analysis.

Sample Data Structure

Table 1: Demographics Sample (with Country Names)

User ID	Age	Gender	Country
11	32	M	Germany
14	39	M	Germany
22	21	M	Belgium
37	25	M	Germany
38	60	F	Germany

Table 2: Deposits Sample

User ID	Amount (€)	Status	Payment Method
11	10.00000	S	PayPal
14	48.74999	S	MAESTRO
167	20.07214	S	VISA
70	61.09888	F	MASTERCARD
208	102.00000	S	VISA

Table 3: Cash Games Sample

User ID	Date	Sessions	Stakes (€)	Winnings (€)
11	2015-02-09	1	0.69	0.73
14	2015-02-01	6	111.78	68.73
14	2015-02-02	3	114.11	71.13
14	2015-02-03	2	15.00	4.38
37	2015-02-01	4	18.15	8.95

Table 4: Top 10 Countries by Player Count

Country	Number of Players
Germany	1416
France	1252
Spain	336
Belgium	300
United Kingdom of Great Britain and Northern Ireland	231
Austria	216
Czechia	190
Switzerland	185
Hungary	184
Netherlands, Kingdom of the	140

4. Preliminary Hypotheses

Hypothesis 1 (Linear Regression)

There is a positive linear relationship between the number of bets a player places and their total stakes wagered in cash games. We hypothesize that players who place more bets will wager larger total amounts,

as betting volume directly contributes to overall gambling activity. We will use linear regression analysis to test this relationship, calculate the correlation coefficient, and assess whether the slope is significantly different from zero.

Hypothesis 2 (Logistic Regression/Classification)

Player demographic characteristics (age and gender) can predict whether a player primarily engages in cash games versus tournaments. We hypothesize that younger players and male players are more likely to prefer cash games over tournaments. We will develop a logistic regression model, train it on 70% of the data, and test its classification accuracy on the remaining 30%.

5. Initial Insights

Descriptive Statistics

Table 5: Demographics Summary

N Players	Mean Age	SD Age	% Male
5028	28.89	8.84	90.65

Table 6: Financial Activity Overview

Metric	Value
Total Deposits	295088
Successful Deposits	223601
Total Withdrawals	32307
Successful Withdrawals	17182
Players with Withdrawals	1739

Key Observations

Preliminary exploration reveals several important patterns:

- Only **1739** of **5028** players (**34.6%**) made successful withdrawals, suggesting many players experience net losses
- **4232** players engaged in cash games and **3448** played tournaments, with some overlap
- The dataset shows high variability in betting amounts and player engagement levels
- Rich temporal structure (daily records from February 2015) allows for analysis of betting patterns over time

Exploratory Visualizations
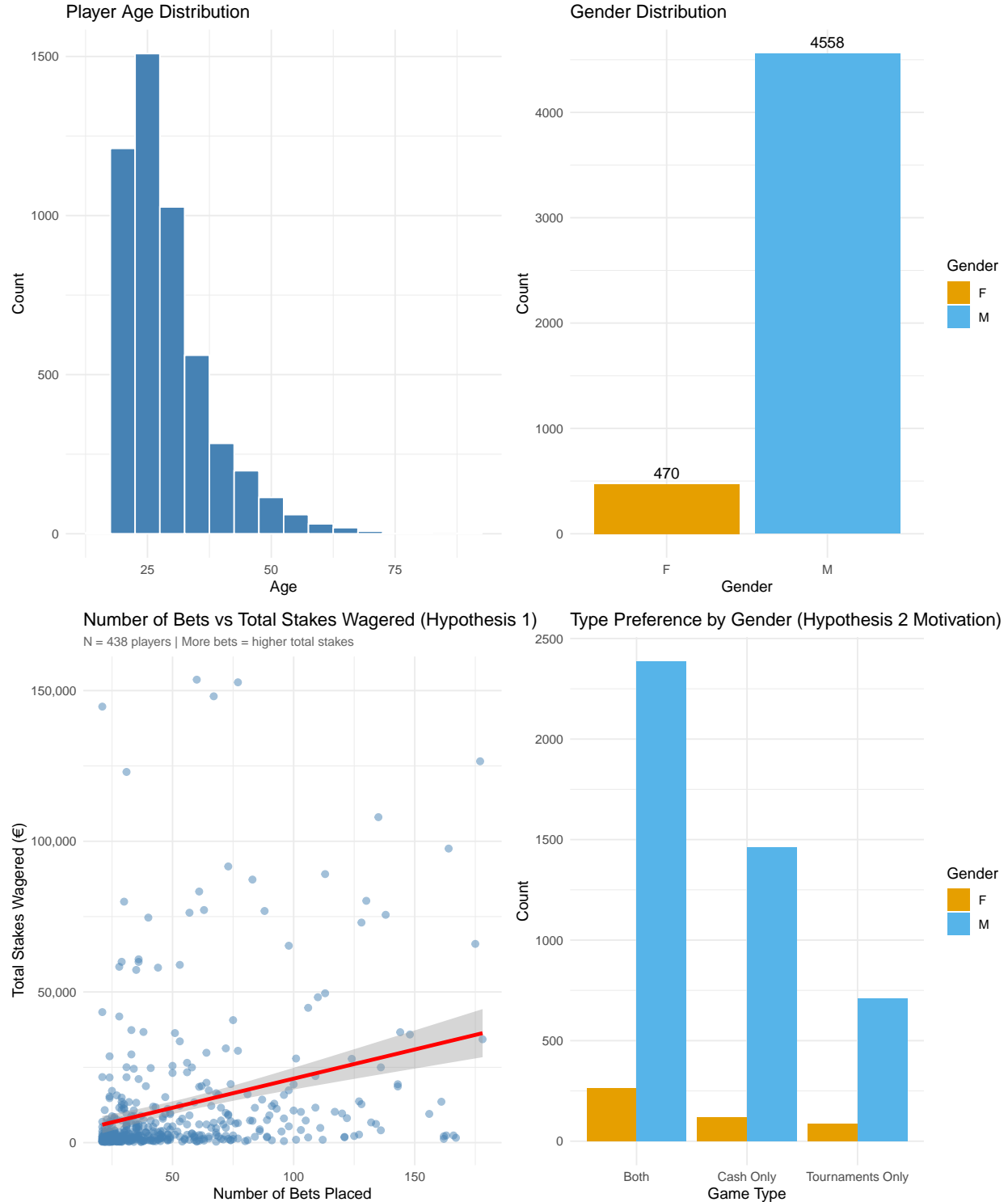
These visualizations motivate our hypotheses and demonstrate interesting patterns in the data that warrant statistical testing. The relationship between number of bets and total stakes will quantify how betting volume contributes to overall gambling activity, testing our first hypothesis. The game type preferences show potential demographic differences worth exploring in our second hypothesis through logistic regression.