Esports athletes: How do they make so much money?

Exploring factors behind the income from playing games at a professional level

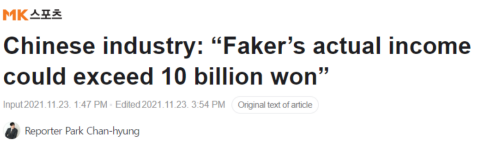
Duc Quan Nguyen, Vinh Hoang Nguyen, Nhu Thai Minh Nguyen

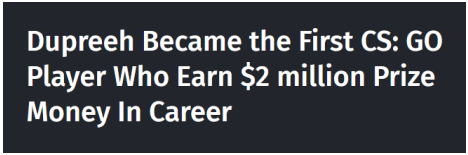
ENAC, Toulouse

January 17, 2024,
Empirical Project

Table of Contents


LIST: NOTAIL IS THE WEALTHIEST ESPORTS ATHLETE WITH $7M IN EARNINGS
by Phoenix 2 years ago


MK 스포츠
Chinese industry: “Faker’s actual income could exceed 10 billion won”
Input 2021.11.23. 1:47 PM · Edited 2021.11.23. 3:54 PM
Original text of article
Reporter Park Chan-hyung


Dupreeh Became the First CS: GO Player Who Earn $2 million Prize Money In Career


16-year old Kyle "Bugha" Giersdorf wins $3 million at Fortnite World Cup
Updated on: July 29, 2019 / 2:38 PM EDT / CBS News

Data source

Most of the data is collected from esportsearnings.com and escharts.com, based on freely available public information.

Sample size

After collecting data from top 100 male and top 100 female, the sample size is N=200. They competed in 43 different game titles and franchises.

First model

Variables

Name	Variable	Type	Explanation	Expected sign
Sex	SEX	category	1=man 2=woman	?
Region	REG	category	1=North America 2=South America 3=Europe 4=Asia 5=Oceania 6=Africa	?
Earning duration (years)	LNE	numeric	/	+
Champion (times)	LNC	numeric	/	+
Runner-up (times)	LNR	numeric	/	+
3rd place/Semi-finals (times)	LNS	numeric	/	+
Competed in more than one game	CMP	category	1=no 2=yes	?
Game with most earning	GWM	category	1 to 43	?
Game popularity (Twitter follows)	GPOP	category	1=more than 6m 2=more than 3m 3=more than 1m 4=more than 500k 5=more than 100k 6= more than 50k 7= more than 5k 8=less than 5k	+
Tournament by developer	TDEV	category	1=no 2=yes	+
3rd party tournament	T3P	category	1=no 2=yes	+

First model

Variables

Gender (SEX)

The woman with the highest earning from tournaments is at rank 518 in the highest overall earnings ranking.

Game with most earning (GWM)

Only 7 out of 43 games titles and franchises became the biggest source of income for the players: DotA, Fortnite, League of Legends, Call of Duty, Counter Strike, Arena of Valor, PUBG Mobile.

Tournaments hosted by the game developer or a 3rd party (TDEV and T3P)

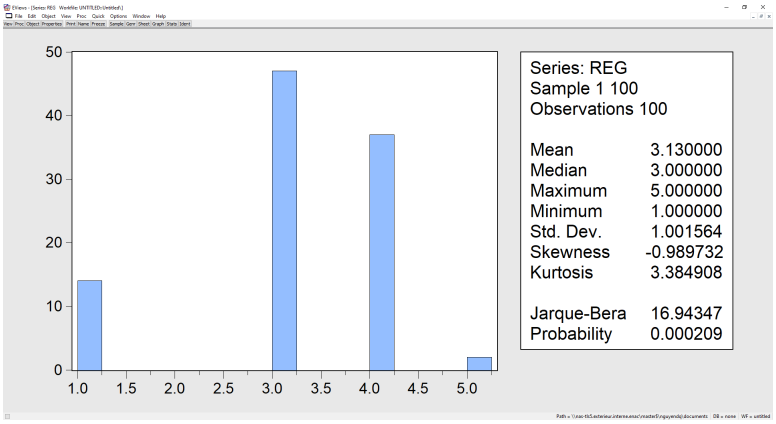
There are variations of these 2 variables for 43 game titles, but for the 7 game titles mentioned above, all of them have both kinds of tournament.

Changes to data

- SEX removed. N = 100
- TDEV , T3P removed
- GWM: ~~1 to 43~~ 1 to 7:
  - 1: DotA
  - 2: Fortnite
  - 3: Counter Strike
  - 4: League of Legends
  - 5: Call of Duty
  - 6: PUBG Mobile
  - 7: Arena of Valor
- GPOP: ~~8 categories~~ Number of followers of 7 game titles mentioned

First model

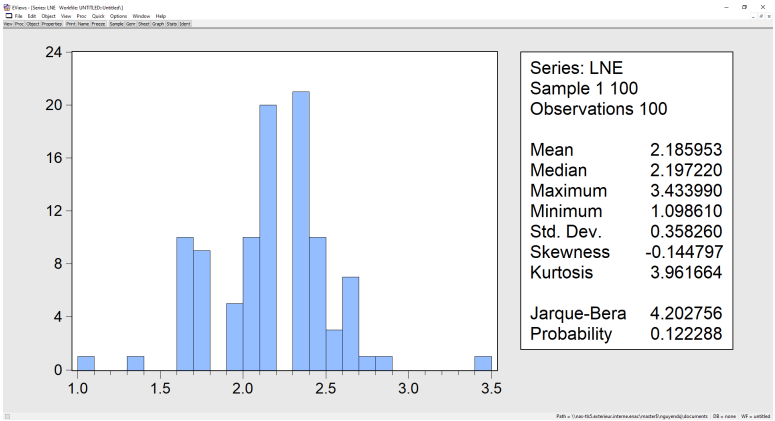
Variables: Region



- Players from Asia and Europe tends to earn more.
- No player from South America and Africa.

First model

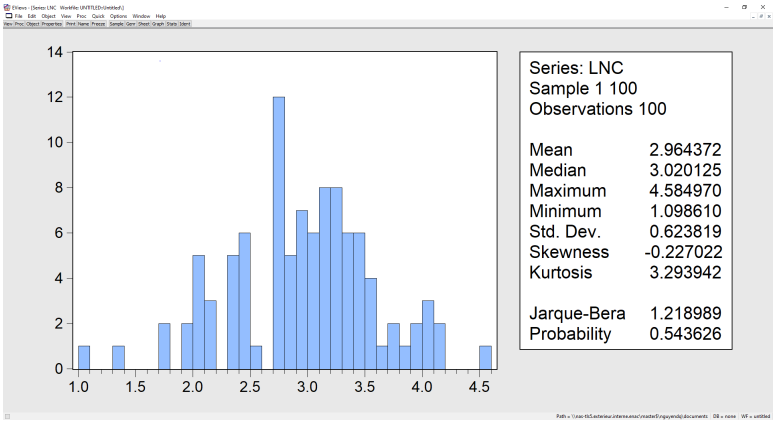
Variables: Earning duration



- High earning players tend to have already played from 5 to 12 years

First model

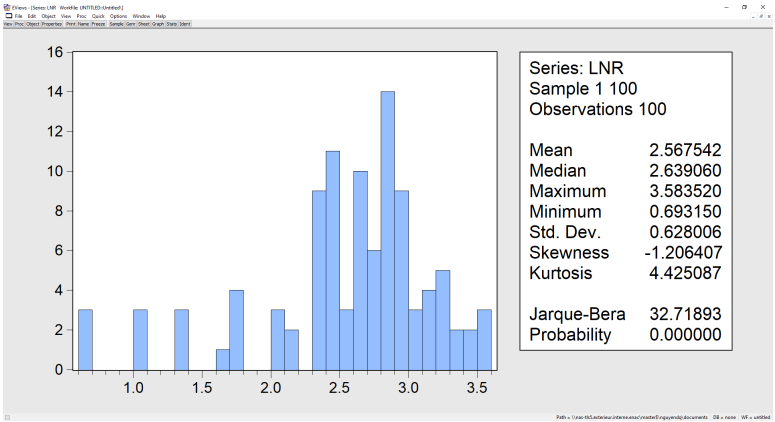
Variables: Champion



- Players with high earnings tend to win a tournament more times than the median value.

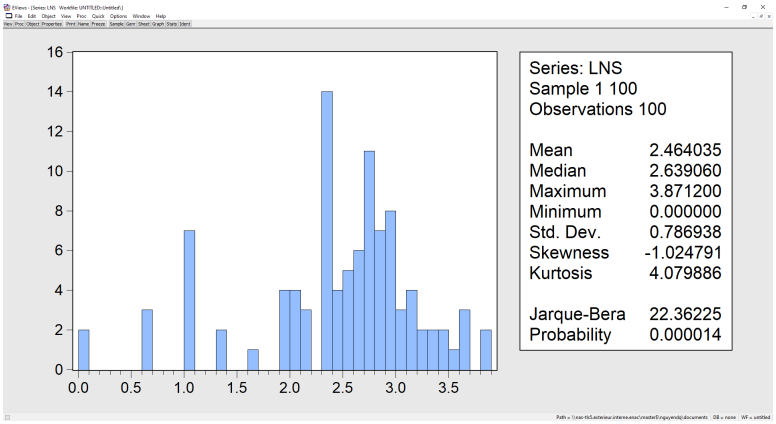
First model

Variables: Runner-up



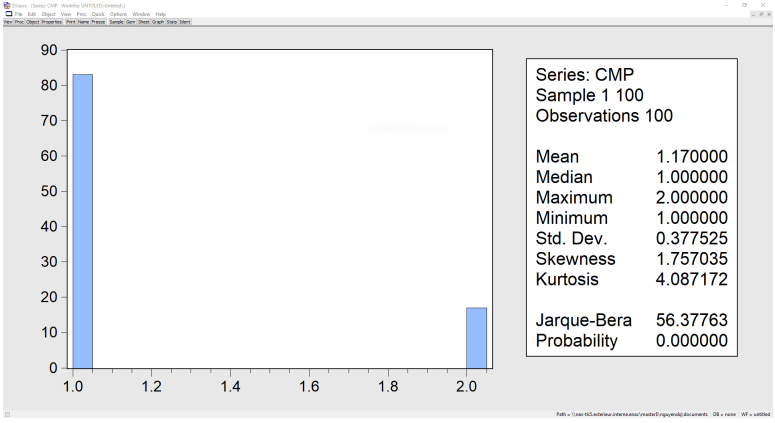
First model

Variables: 3rd place/Semi-finals



First model

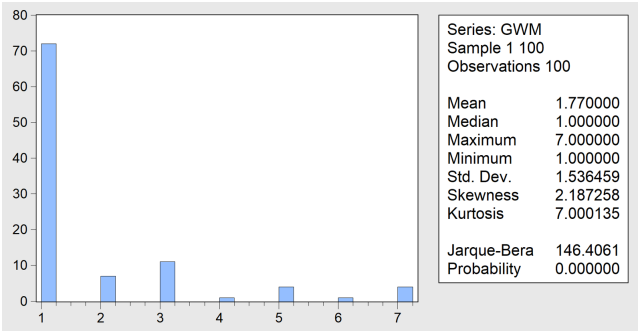
Variables: Competed in more than one game



- Most competed in only one game.

First model

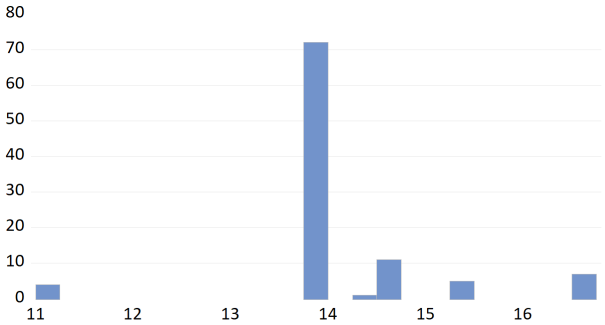
Variables: Game with most earning



- Players earned most from DotA.

First model

Variables: Game popularity



Series: GPOP	
Sample 1 100	
Observations 100	
Mean	14.15772
Median	13.93108
Maximum	16.68710
Minimum	11.20960
Std. Dev.	0.967285
Skewness	0.229325
Kurtosis	7.012005
Jarque-Bera	67.94393
Probability	0.000000

First model

First regression

Dependent Variable: LNT
Method: Least Squares
Date: 01/16/24 Time: 17:28
Sample: 1 100
Included observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	16.20066	1.067672	15.17381	0.0000
CMP	0.075114	0.129097	0.581838	0.5621
GPOP	-0.088899	0.058605	-1.516925	0.1328
GWM	-0.148966	0.031539	-4.723218	0.0000
LNC	0.131904	0.120672	1.093076	0.2772
LNE	0.032393	0.205179	0.157878	0.8749
LNR	-0.069646	0.138973	-0.501148	0.6175
LNS	-0.093568	0.126497	-0.739684	0.4614
REG	-0.052620	0.053815	-0.977807	0.3308

R-squared	0.214281	Mean dependent var	14.65401
Adjusted R-squared	0.145207	S.D. dependent var	0.458820
S.E. of regression	0.424202	Akaike info criterion	1.208475
Sum squared resid	16.37520	Schwarz criterion	1.442940
Log likelihood	-51.42375	Hannan-Quinn criter.	1.303367
F-statistic	3.102183	Durbin-Watson stat	0.387165
Prob(F-statistic)	0.003847		

Variables	Expected sign	First model
SEX	?	X
REG	?	-
LNE	+	+
LNC	+	+
LNR	+	-
LNS	+	-
CMP	?	+
GWM	?	-
GPOP	+	-
TDEV	+	X
T3P	+	X

Result of the first regression

- Dependent variable is total earning.
- Model only captures $R^2 = 21.4\%$; adjusted $R^2 = 14.5\%$ of the variability of the total earning.

Variables

- Significance at 99%: GWM
- Significance $> 70\%$: GPOP, LNC
- Significance $> 50\%$: REG, LNS
- Not significant: CMP, LNE, LNR

Wald Test:
Equation: REGRESSION1

Test Statistic	Value	df	Probability
F-statistic	0.251149	(1, 91)	0.6175
Chi-square	0.251149	1	0.6163

Null Hypothesis Summary:

Normalized Restriction (= 0)	Value	Std. Err.
C(7)	-0.069646	0.138973

Restrictions are linear in coefficients.

Result

- Wald test on LNR gives p-value > 0.05.
- We can remove LNR.

Second model

Second regression

Dependent Variable: LNT
Method: Least Squares
Date: 01/16/24 Time: 17:24
Sample: 1 100
Included observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	16.28846	0.828495	19.66030	0.0000
GPOP	-0.089546	0.050338	-1.778882	0.0785
GWM	-0.145990	0.030324	-4.814401	0.0000
LNC	0.121279	0.107677	1.126326	0.2629
LNS	-0.115860	0.089058	-1.300960	0.1965
REG	-0.058247	0.049378	-1.179600	0.2411

R-squared	0.208299	Mean dependent var	14.65401
Adjusted R-squared	0.166187	S.D. dependent var	0.458820
S.E. of regression	0.418964	Akaike info criterion	1.156060
Sum squared resid	16.49988	Schwarz criterion	1.312370
Log likelihood	-51.80299	Hannan-Quinn criter.	1.219321
F-statistic	4.946326	Durbin-Watson stat	0.386855
Prob(F-statistic)	0.000460		

Second regression result

- All variables' significance is over 70%
- R^2 lowered to 20.8%

Increase R^2

- Introduces 2 new variables based on publicly available information:

Name	Variables	Type	Expected sign
Total prize pool (USD)	TPP	Numeric	+
Total tournaments held	TTS	Numeric	+

Third model

Third regression

Dependent Variable: LNT
Method: Least Squares
Date: 01/16/24 Time: 17:26
Sample: 1 100
Included observations: 100

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	13.58467	4.289247	3.167146	0.0021
GPOP	-0.052133	0.074977	-0.695320	0.4886
GWM	-0.108618	0.070182	-1.547657	0.1251
LNC	0.141334	0.112530	1.255963	0.2123
LNS	-0.105626	0.095151	-1.110088	0.2699
REG	-0.061136	0.052340	-1.168049	0.2458
TPP	0.125164	0.191275	0.654366	0.5145
TTS	-0.051486	0.094798	-0.543109	0.5884

R-squared	0.212235	Mean dependent var	14.65401
Adjusted R-squared	0.152296	S.D. dependent var	0.458820
S.E. of regression	0.422439	Akaike info criterion	1.191076
Sum squared resid	16.41785	Schwarz criterion	1.399489
Log likelihood	-51.55379	Hannan-Quinn criter.	1.275425
F-statistic	3.540864	Durbin-Watson stat	0.376062
Prob(F-statistic)	0.002060		

Conclusion

Variables

- Most significant: GWM
- Significance $> 70\%$: LNC, LNS, REG
- Not significance: GPOP, TPP, TTS, CMP, LNE, LNR

Statistical values

- $R^2 = 0.21$: Explain 21% of the changes in the dependent variable.
- Adjusted $R^2 = 0.15$: Some of the variables introduced are not significant.
- F-statistic $= 3.5$ and p-value $= 0.002$: At least one independent variable is contributing to explaining the variation.

Conclusion

What led to a low explanation capability?

- Many information are not publicly available.
- Differences in how different games host their tournaments.
- Contain large amount of players from one game.

How to improve?

- Focus on one game.
- Different approach to selecting factors.


Thank
you!