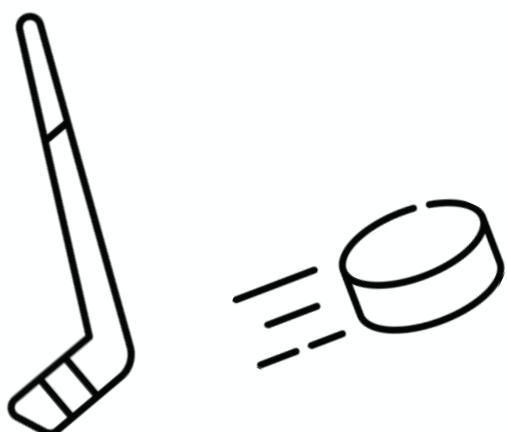


Ice Hockey

Presentation 2023



Group 27

B104020009 鄒仁迪

B104020011 危湘妤

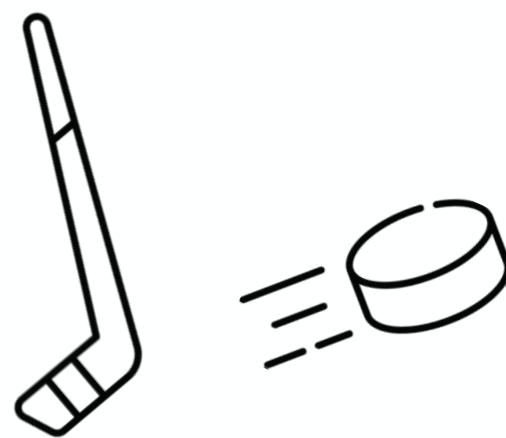
B104020021 程愷元

B104020032 翁武麟

B104020038 林敬原

Agenda

1. Research Background
2. Data Source and Preprocessing
3. Data Analysis
4. Conclusion



2



Research Motivation

受到「冰球壞小子」這部電影啟發
感受到了冰上曲棍球優雅華麗的進球與「執行者」在場上的保護與威嚇，對冰上曲棍球產生了極大的興趣

Research Target

1. 探討球員數據對於薪水的影響



文獻回顧



Salary Determination in the National
Hockey League: The Effects of Skills,
Franchise Characteristics, and
Discrimination

J. C. H. Jones, William D. Walsh

技能似乎是NHL球員薪水的主要決定因素
有限的支持在壟斷性的位置（前鋒）有一定正向的影響
沒有證據支持對法裔加拿大人存在普遍的薪酬歧視的假設

Salary disclosure and individual
effort: Evidence from the National
Hockey League
James Flynn

描述當球員發現自己得薪水相對較低的時候會把精力從防
守轉移至進攻



Data Source

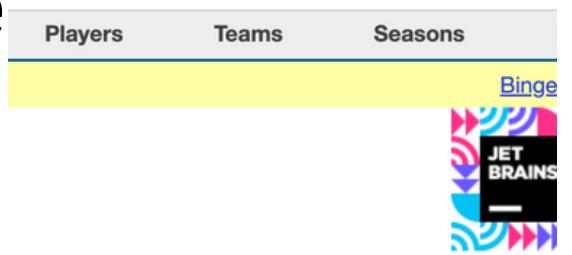


資料來源：
爬蟲自 IcyData、Hockey Reference
資料型態：
NHL 2022-2023 球員賽季資料
NHL 2023-2024 球員賽季薪水
資料總時間：2023年～2024年

ICYDATA

Player	Team	GP	G	A	Pts	+/-	PN	PIM	S	SB	MS	H	GV	TK	BS	FW	FL	F%
Alex DeBrincat	DET	6	8	4	12	7	0	0	18	7	10	3	6	4	2	0	2	0
Dylan Larkin	DET	6	3	8	11	5	3	9	27	6	5	2	3	5	1	51	50	50.5
Jack Hughes	N.J	4	4	6	10	-2	2	4	23	15	7	1	6	6	0	17	37	31.48
Elias Pettersson	VAN	5	2	8	10	5	0	0	12	6	5	5	2	1	3	23	29	44.23
Leon Draisaitl	EDM	5	4	5	9	0	4	8	18	7	9	3	7	4	0	52	39	57.14
William Nylander	TOR	5	4	5	9	-1	1	2	21	11	9	5	1	3	1	0	4	0

HOCKEY
REFERENCE

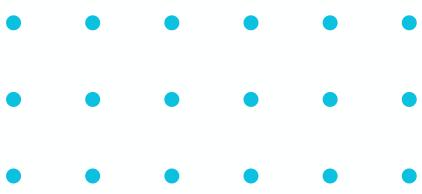


NHL Salaries for 2023-24

Salaries

Player	Tm	Salary	Cap Hit
Kirill Kaprizov	MIN	12,500,000	9,000,000
Adam Fox	NYR	12,000,000	9,500,000
Drew Doughty	LAK	11,000,000	11,000,000
Cale Makar	COL	11,000,000	9,000,000
Dougie Hamilton	NJD	10,600,000	9,000,000
Brady Tkachuk	OTT	10,500,000	8,205,714
Darnell Nurse	EDM	10,400,000	9,250,000
Elias Pettersson	VAN	10,250,000	7,350,000
Nick Suzuki	MTL	10,000,000	7,875,000
Miro Heiskanen	DAL	10,000,000	8,450,000
Filip Forsberg	NSH	10,000,000	8,500,000

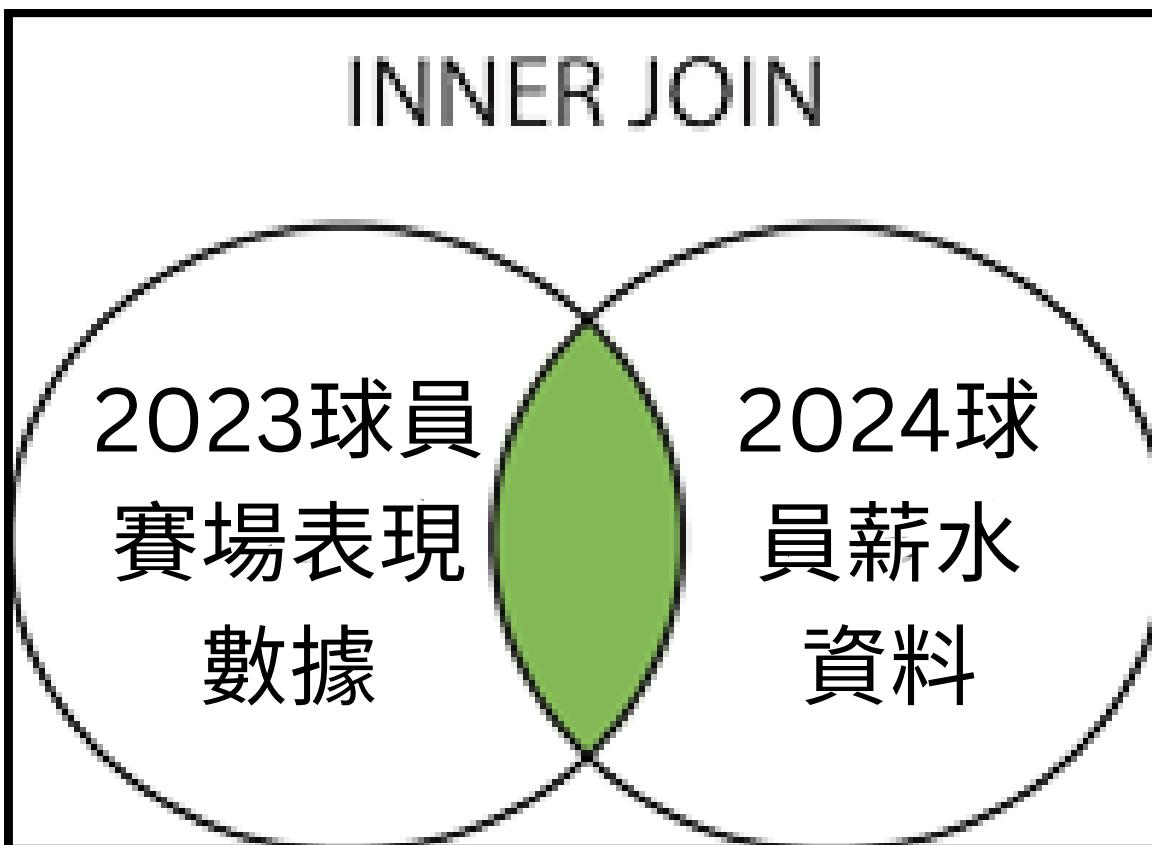




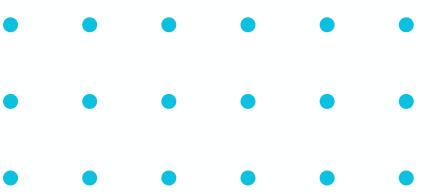
資料預處理



Data Preprocessing



2022-2023球員賽季資料
inner join with
2023-2024球員賽季薪水
最終資料筆數：351筆



連續變數描述



連續變數描述

Variable	說明	Min	Q1	Median	Q3	Max
GP	參加比賽數量	1	56	76	81	82
G	進球數	0	4	10	19	65
A	助攻數	0	9	19	32	90
Pts	分數	0	15	30	50	155

連續變數描述

Variable	說明	Min	Q1	Median	Q3	Max
+/-	球場影響力	-41	-8	1	10	45
PN	處罰	0	8	13	20	42
PIM	罰球犯規分鐘	0	16	30	44	136
S	射門次數	0	73	121	175	367

連續變數描述

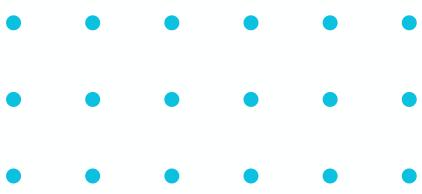
Variable	說明	Min	Q1	Median	Q3	Max
SB	射門被擋球數	0	31	54	78	189
MS	射門偏出數	0	26	46	67	140
H	故意使對方失球控	0	40	67	107	305
GV	失誤導致轉移球權	0	17	31	46	105

連續變數描述

Variable	說明	Min	Q1	Median	Q3	Max
TK	抄截數	0	15	28	42	92
BS	防守成功數	0	24	40	72	196
FW	對峙成功數	0	0	5	105	977
FL	對峙失敗數	0	0	9	144	869

連續變數描述

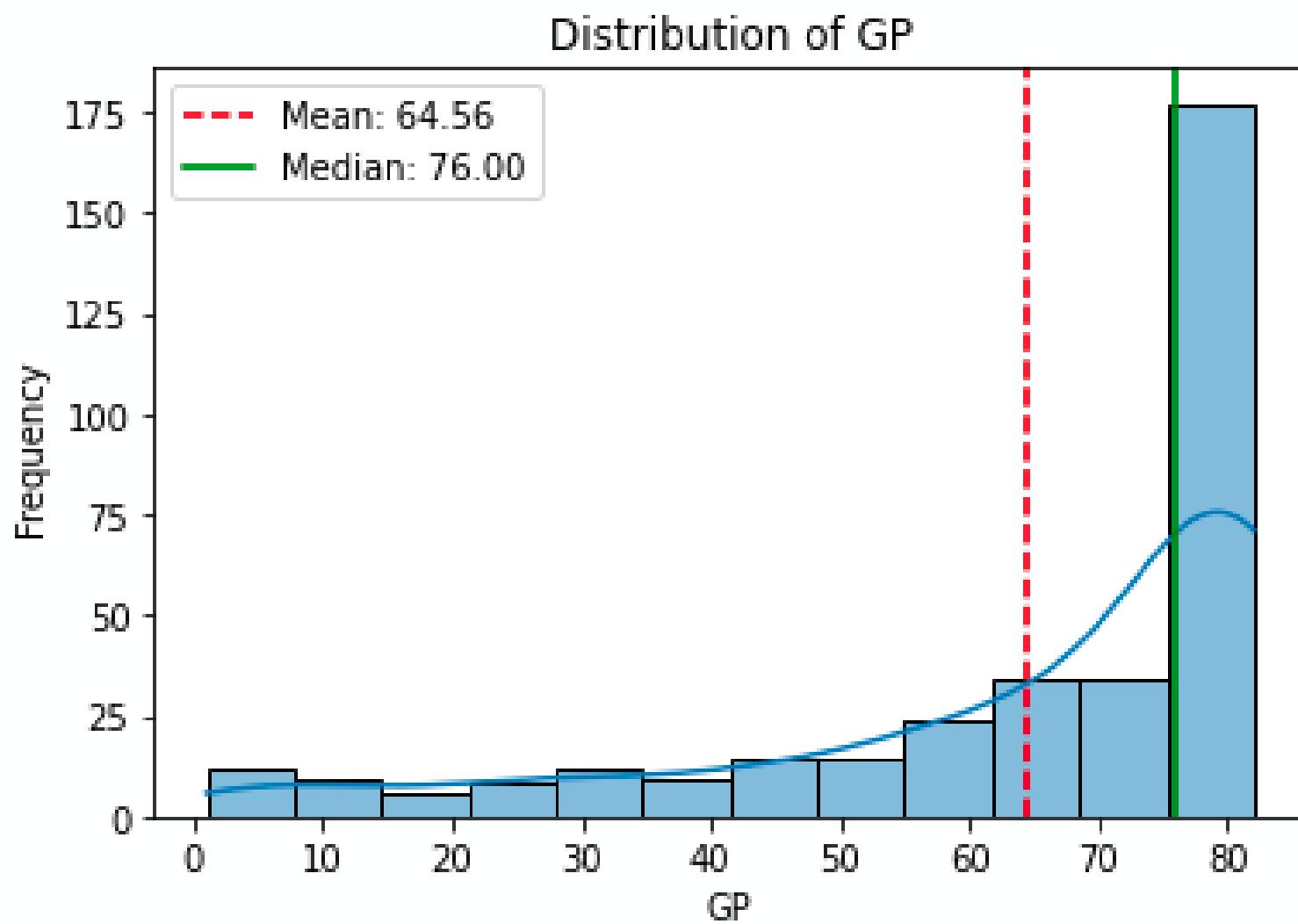
Variable	說明	Min	Q1	Median	Q3	Max
Height	身高	1.727200	1.828800	1.854200	1.905000	2.032000
Weight	體重	74.842680	74.842680	86.182480	95.254320	116.573144
F%	對峙成功率	0	0	33.33	48.3	100
Salary	薪水	775000	850000	2950000	5100000	12500000
Cap.Hit	頂薪	758333	925000	3500000	5950000	12600000



單變量分析



GP(參加比賽數量)



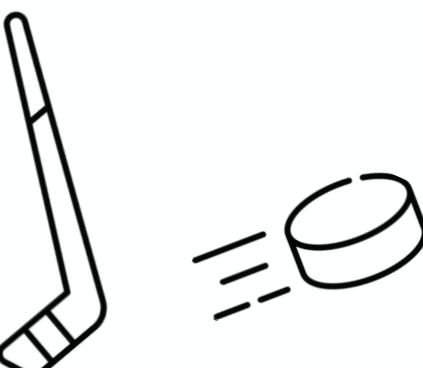
Shapiro Test
p-value = 0.0
拒絕 H₀
不符合正態分佈

例行賽

82場
分區內的24場
不同分區40場
18場跨聯盟比賽
冰球比分積分制
贏的球隊+2分，
延長賽輸的隊伍+1

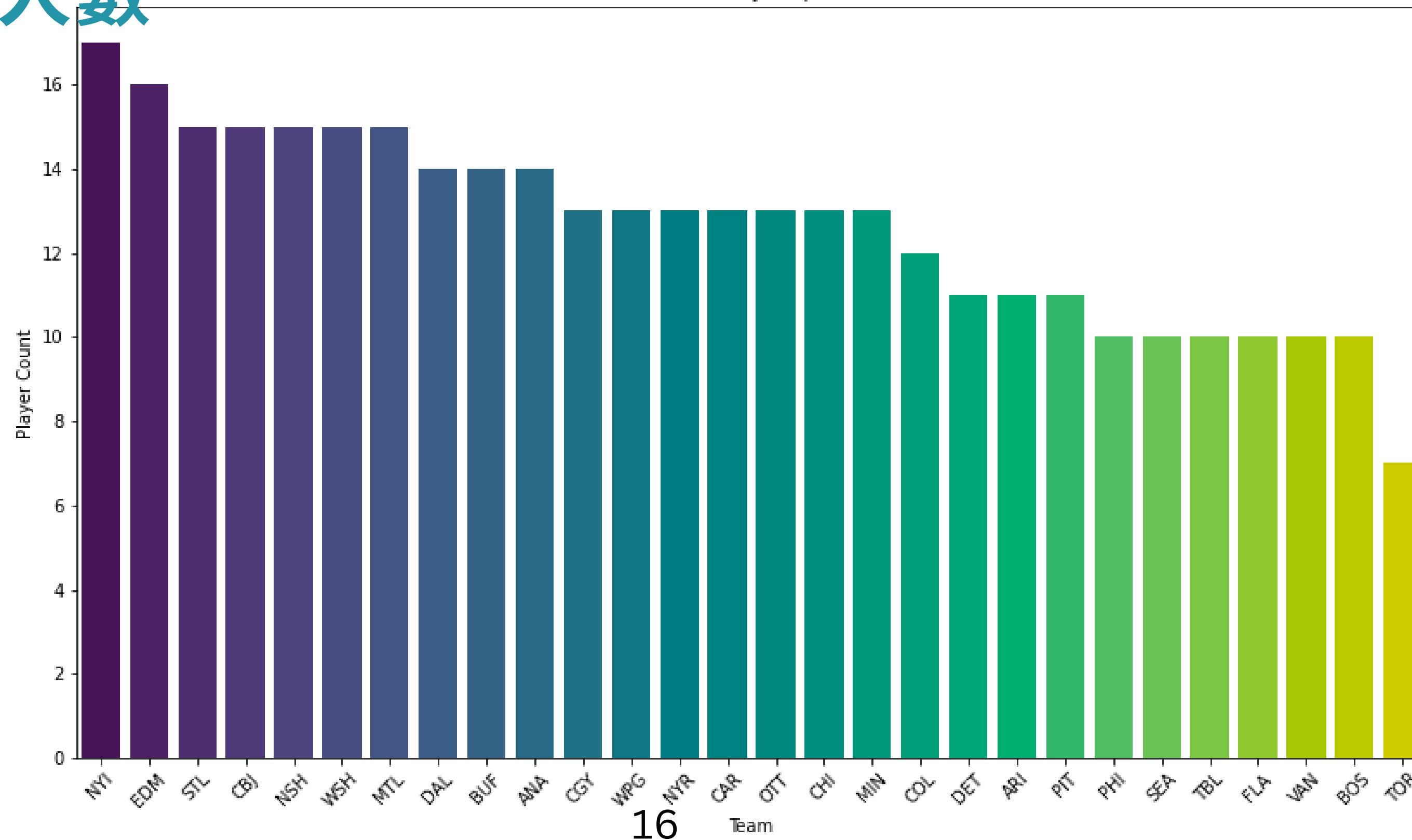
季後賽

16支隊伍將會晉級季後賽
採7戰4勝淘汰賽制

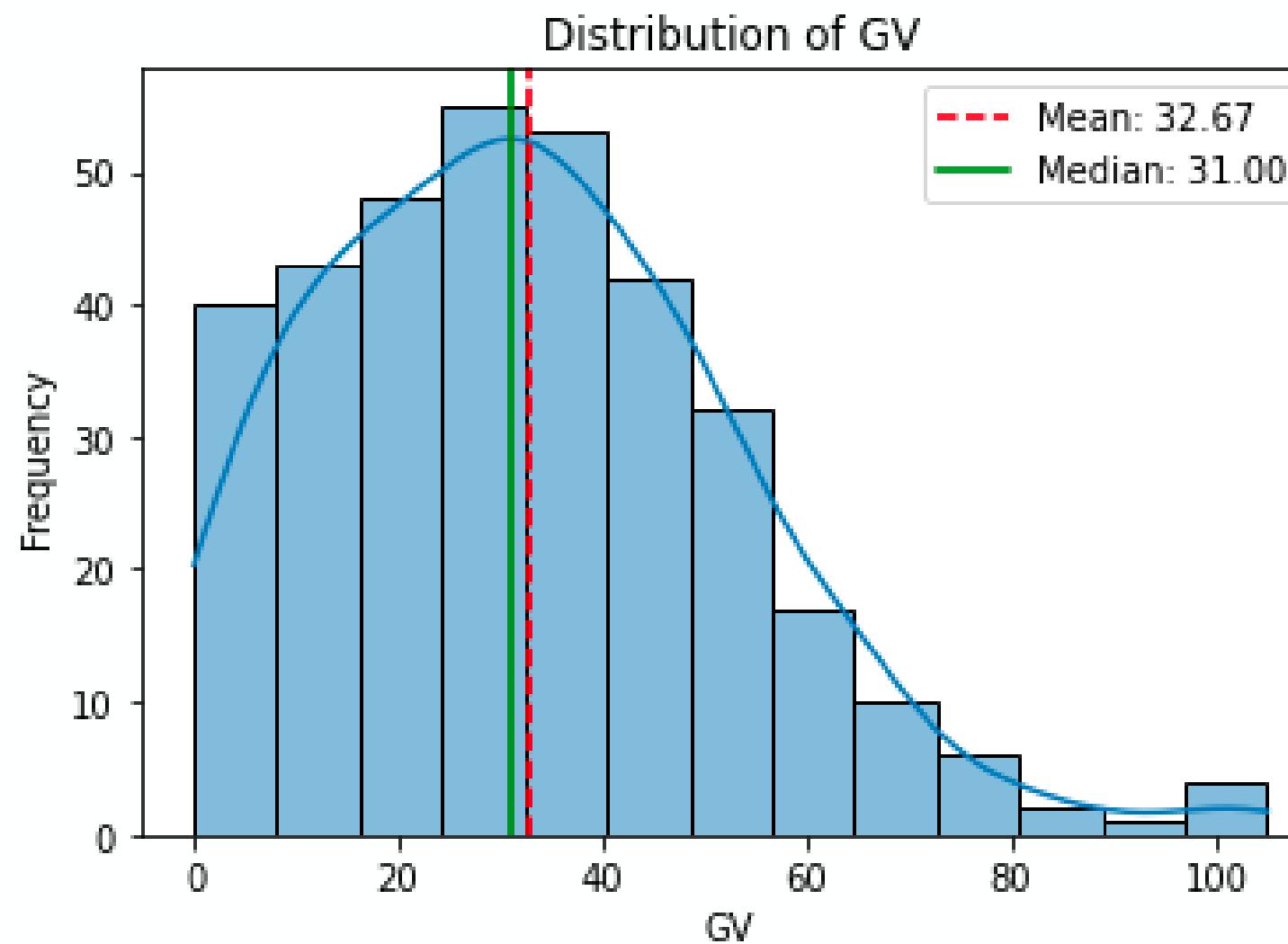


每隊人數

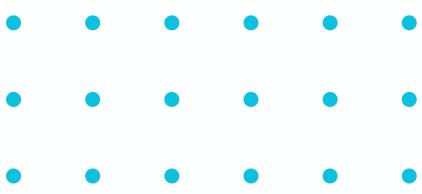
Number of Players per Team



GV(失誤導致轉移球權)



Shapiro Test
 $p\text{-value} = 0.0$
拒絕 H_0
不符合正態分佈

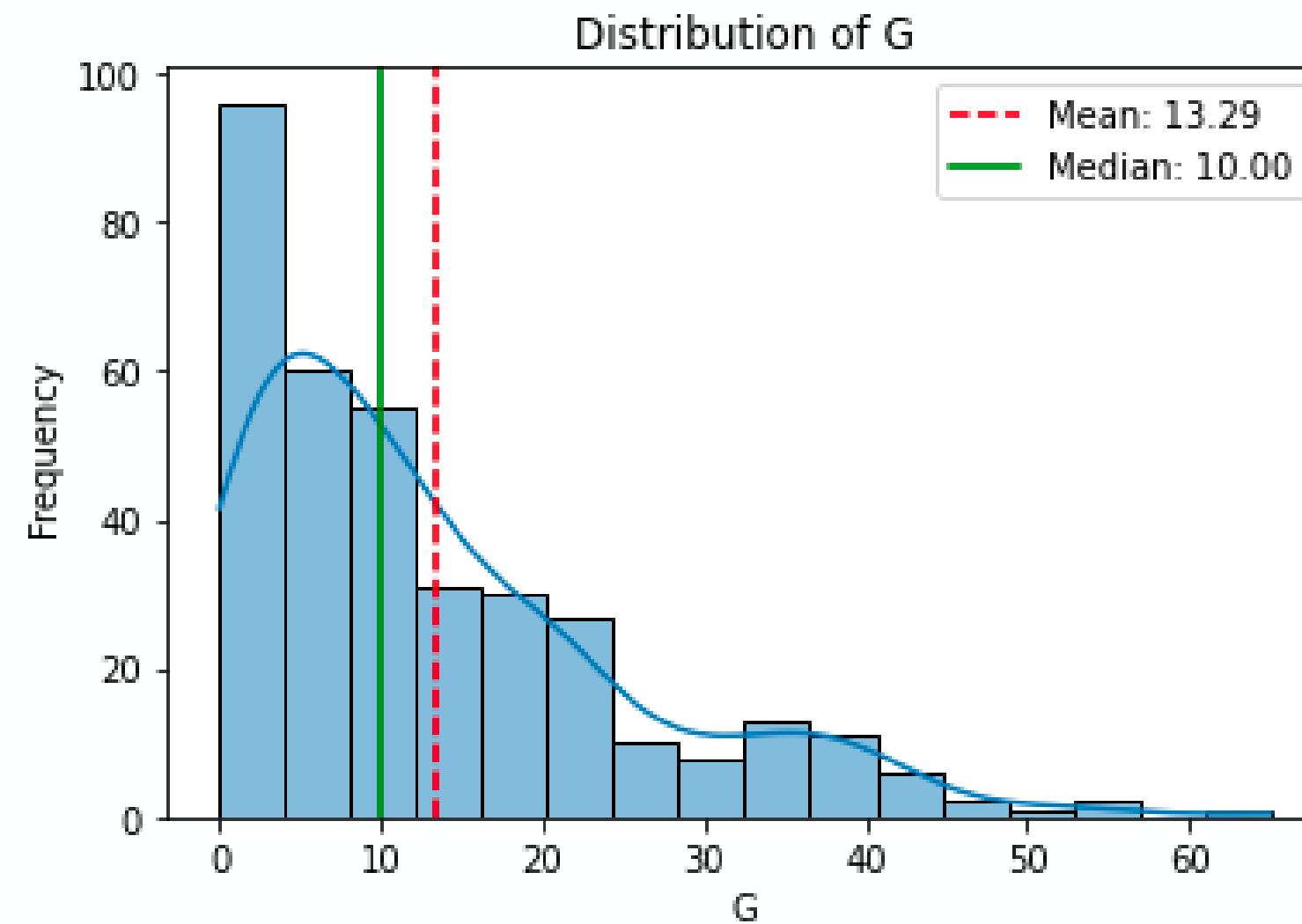


進攻數據

- 1.G(進球數)
- 2.A(助攻數)
- 3.Pts(分數)
- 4.+/- (球場影響力)
- 5.S(射門次數)
- 6.SB(射門被阻擋數)
- 7.MS(射門偏出數)

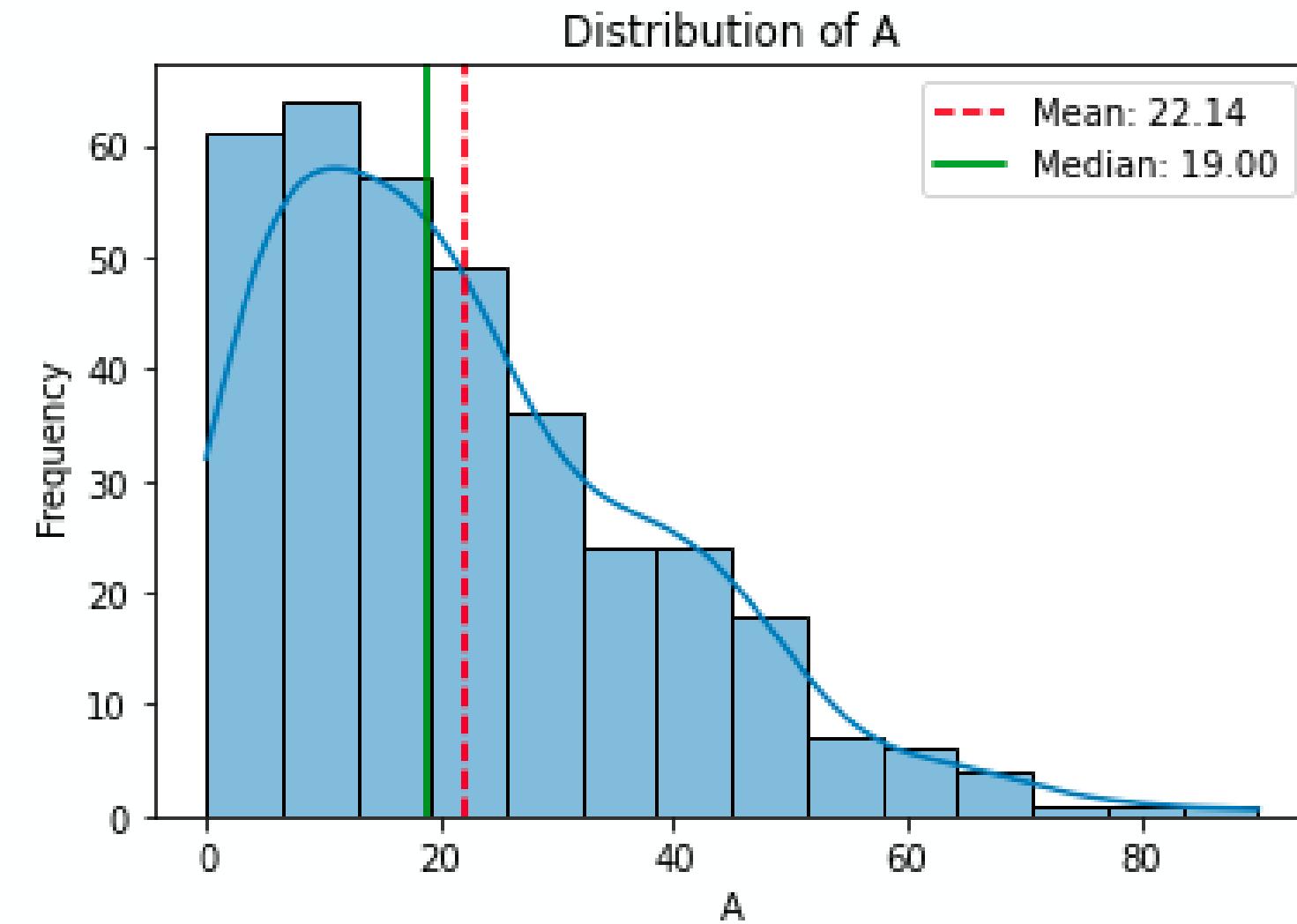


G(進球數)

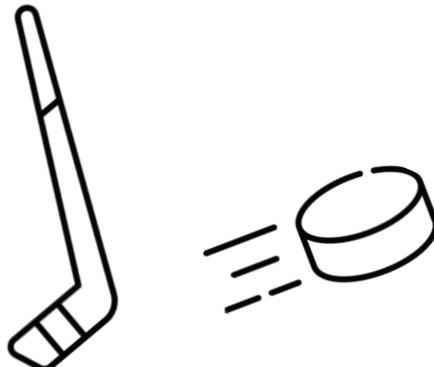


Shapiro Test
 $p\text{-value} = 0.0$
拒絕 H_0
不符合正態分佈

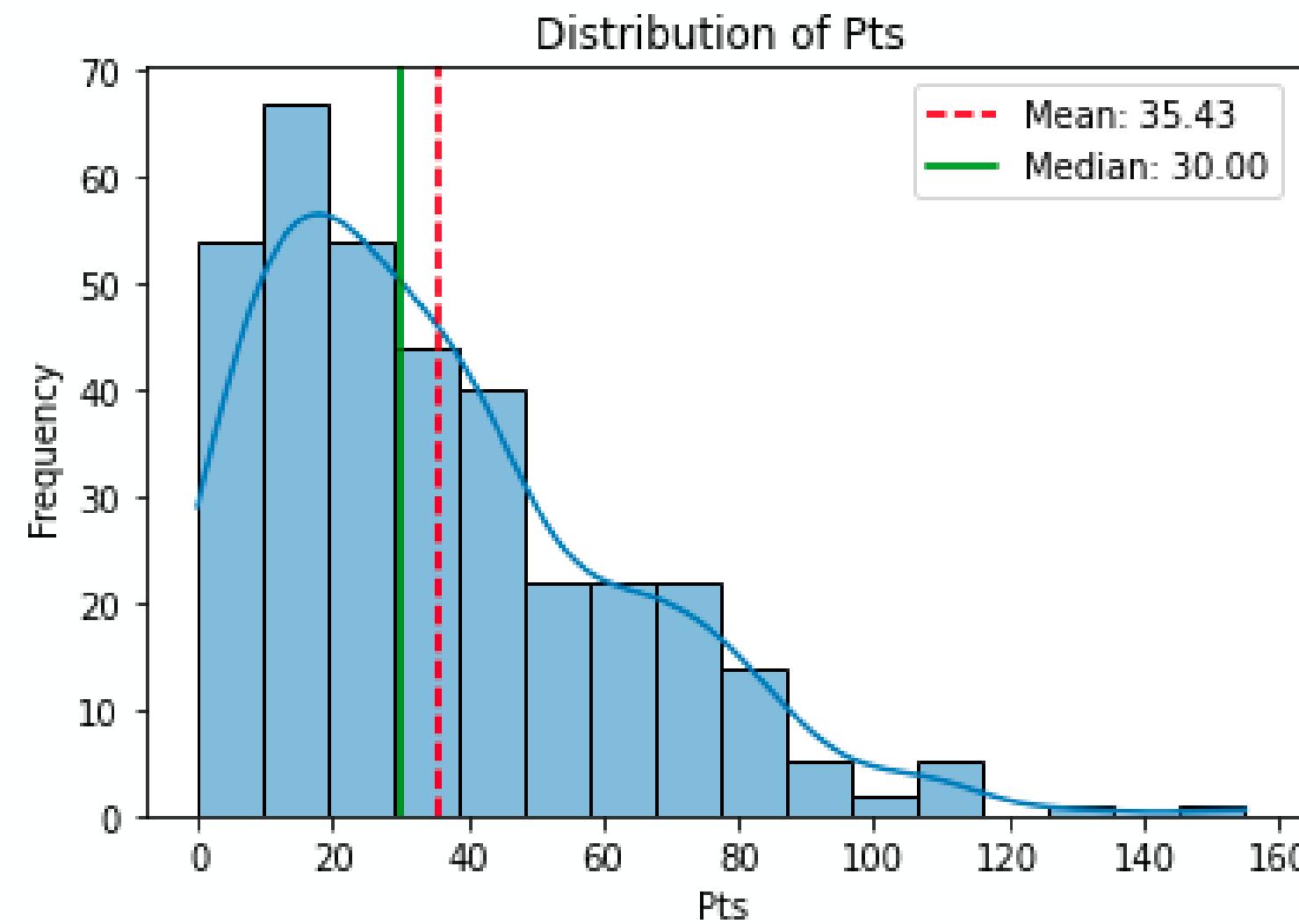
A(助攻數)



Shapiro Test
p-value = 0.0
拒絕 H_0
不符合正態分佈

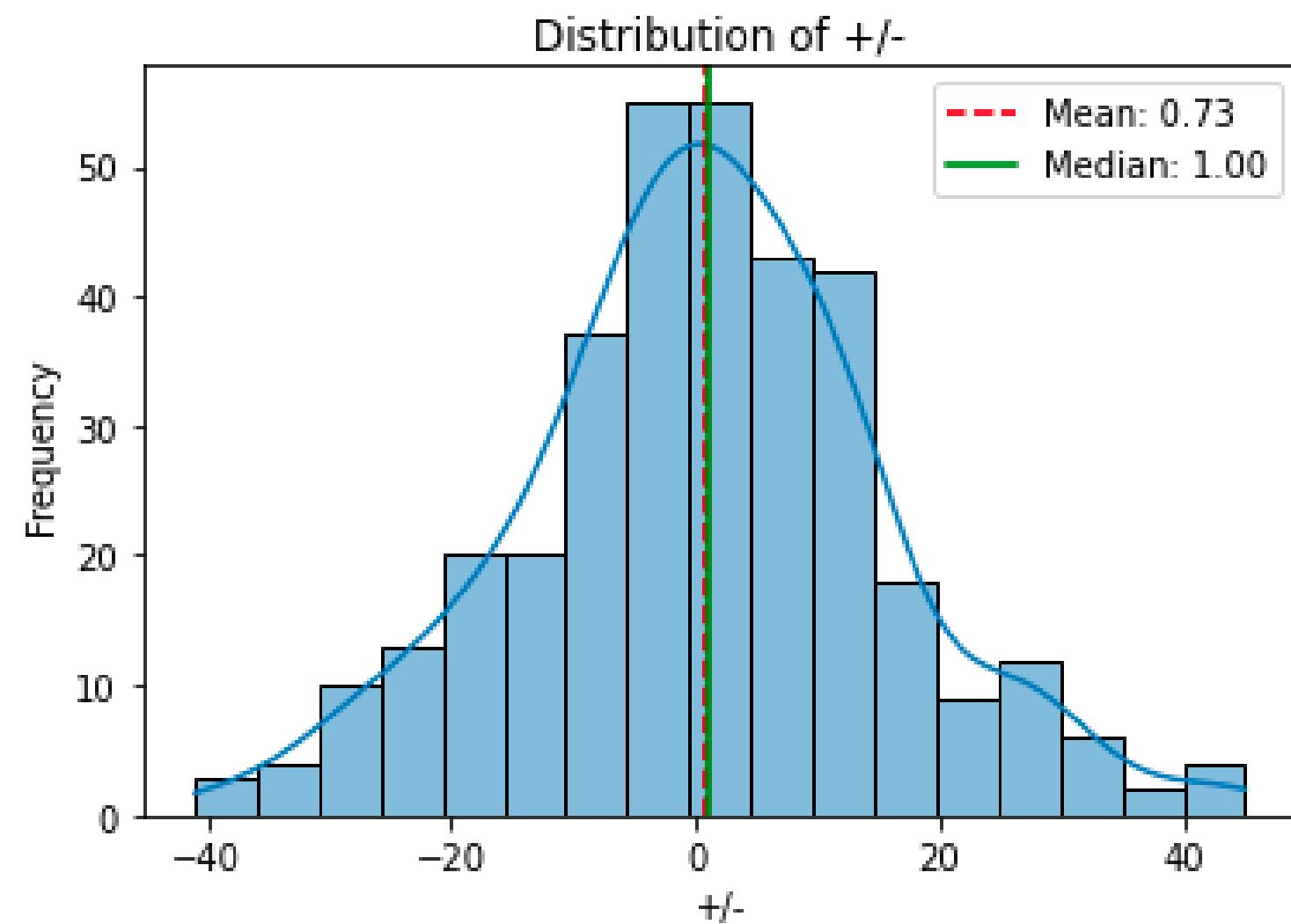


Pts(分數)



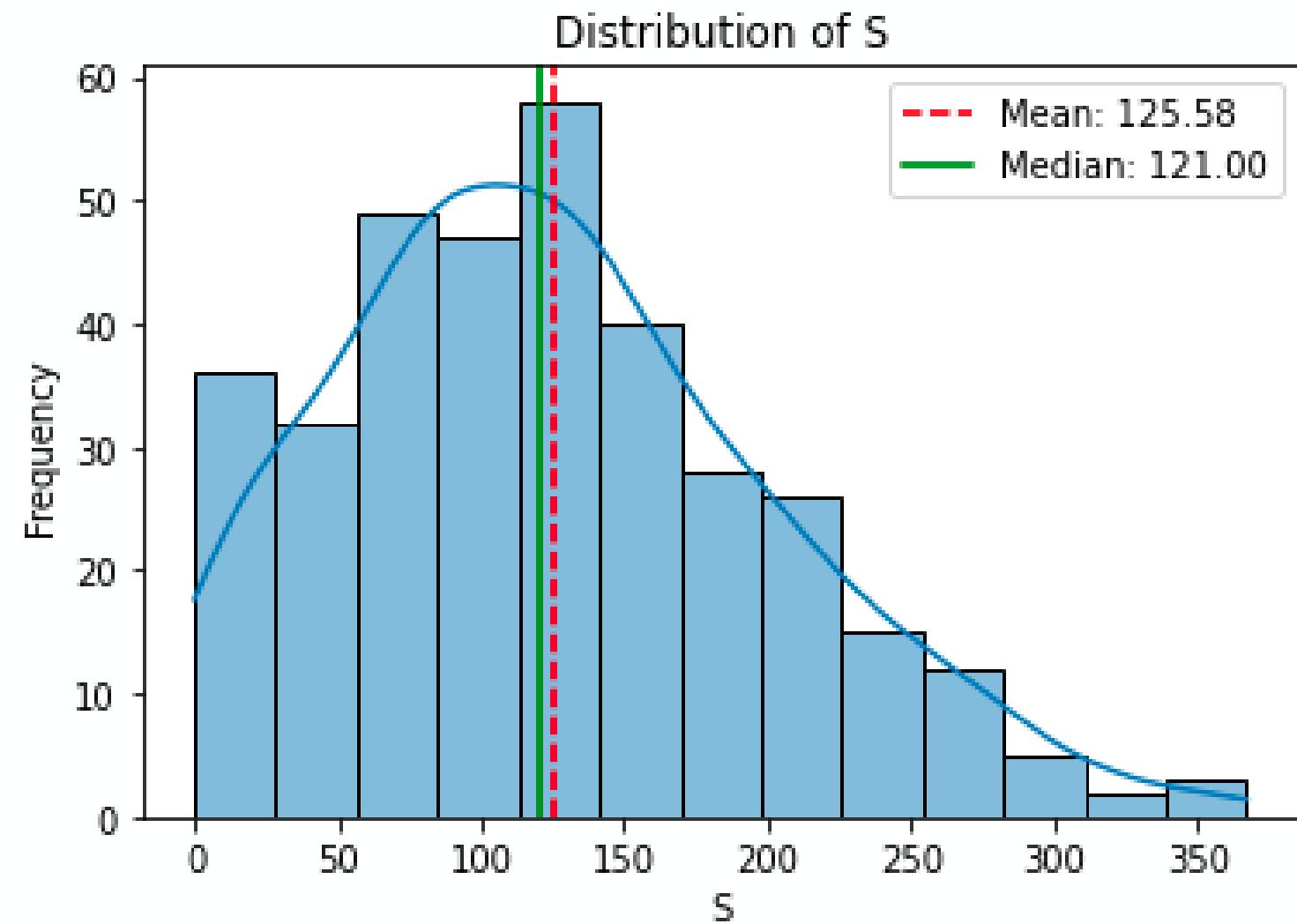
Shapiro Test
p-value = 0.0
拒絕 H_0
不符合正態分佈

+/- (球場影響力)

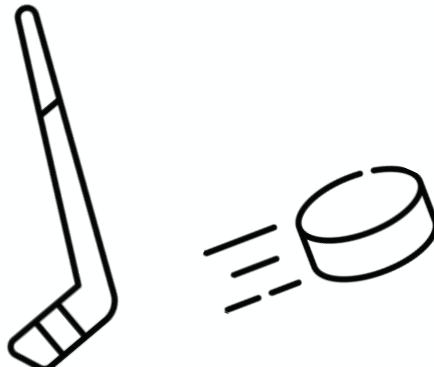


Shapiro Test
p-value = 0.038692
拒絕 H_0
不符合正態分佈

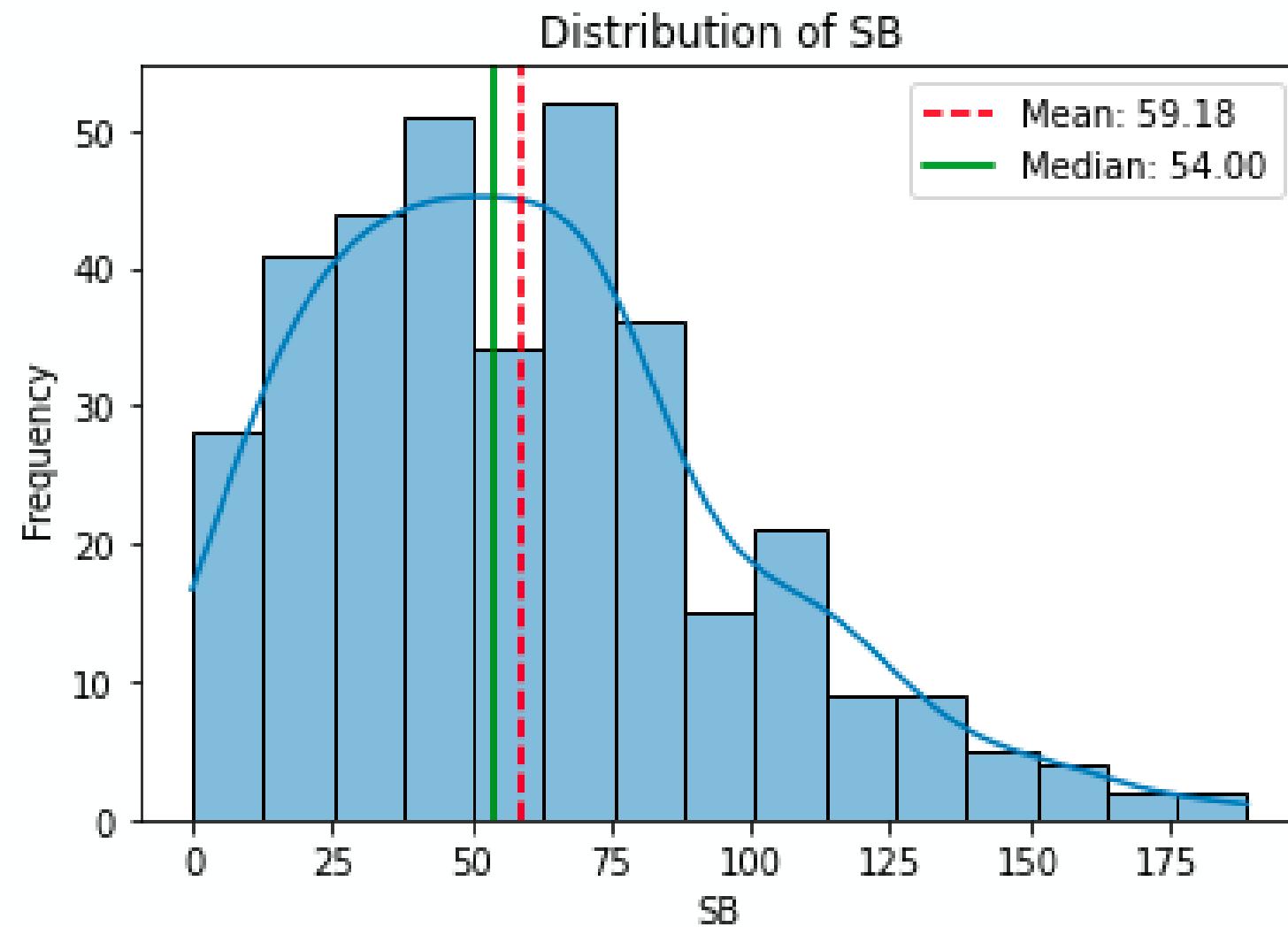
S(射門次數)



Shapiro Test
p-value = 0.000003
拒絕 H_0
不符合正態分佈

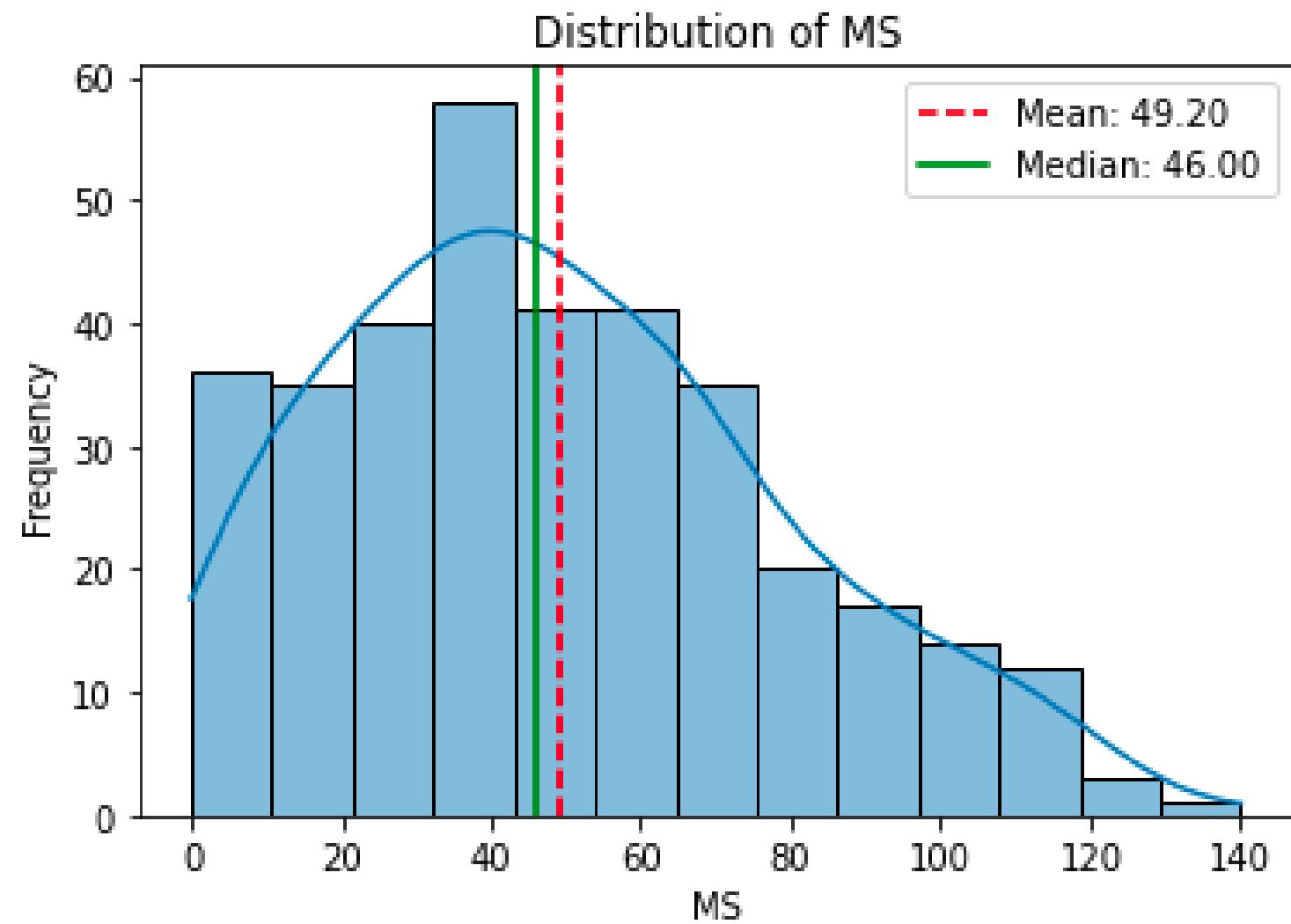


SB(被阻擋球數)

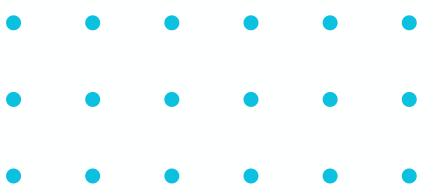


Shapiro Test
p-value = 0.0
拒絕 H_0
不符合正態分佈

MS(射門偏出)



Shapiro Test
p-value = 0.000002
拒絕 H_0
不符合正態分佈

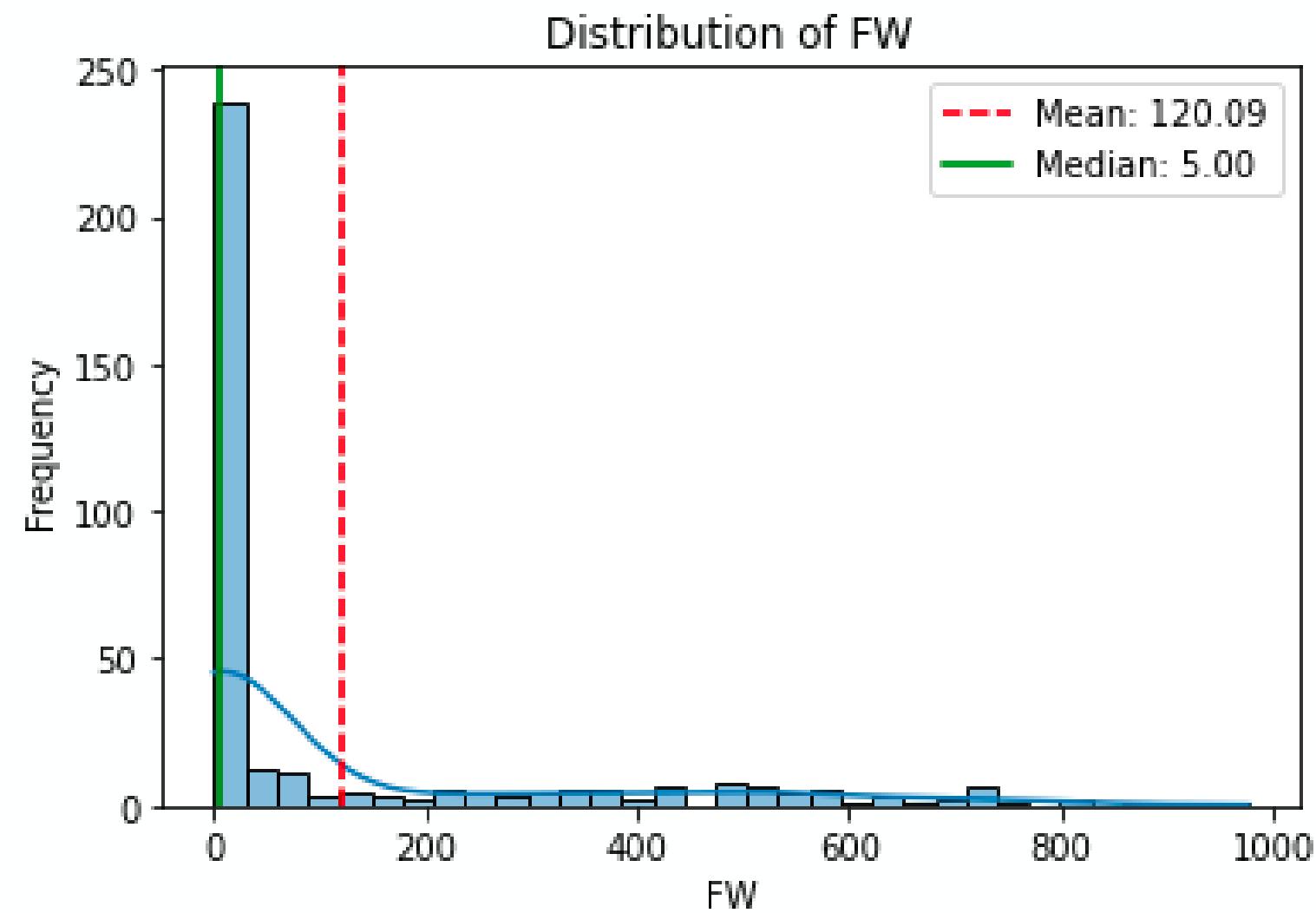


對峙分析

1. FW(對峙勝利)
2. FL(對峙失敗)
3. F%(對峙成功率)

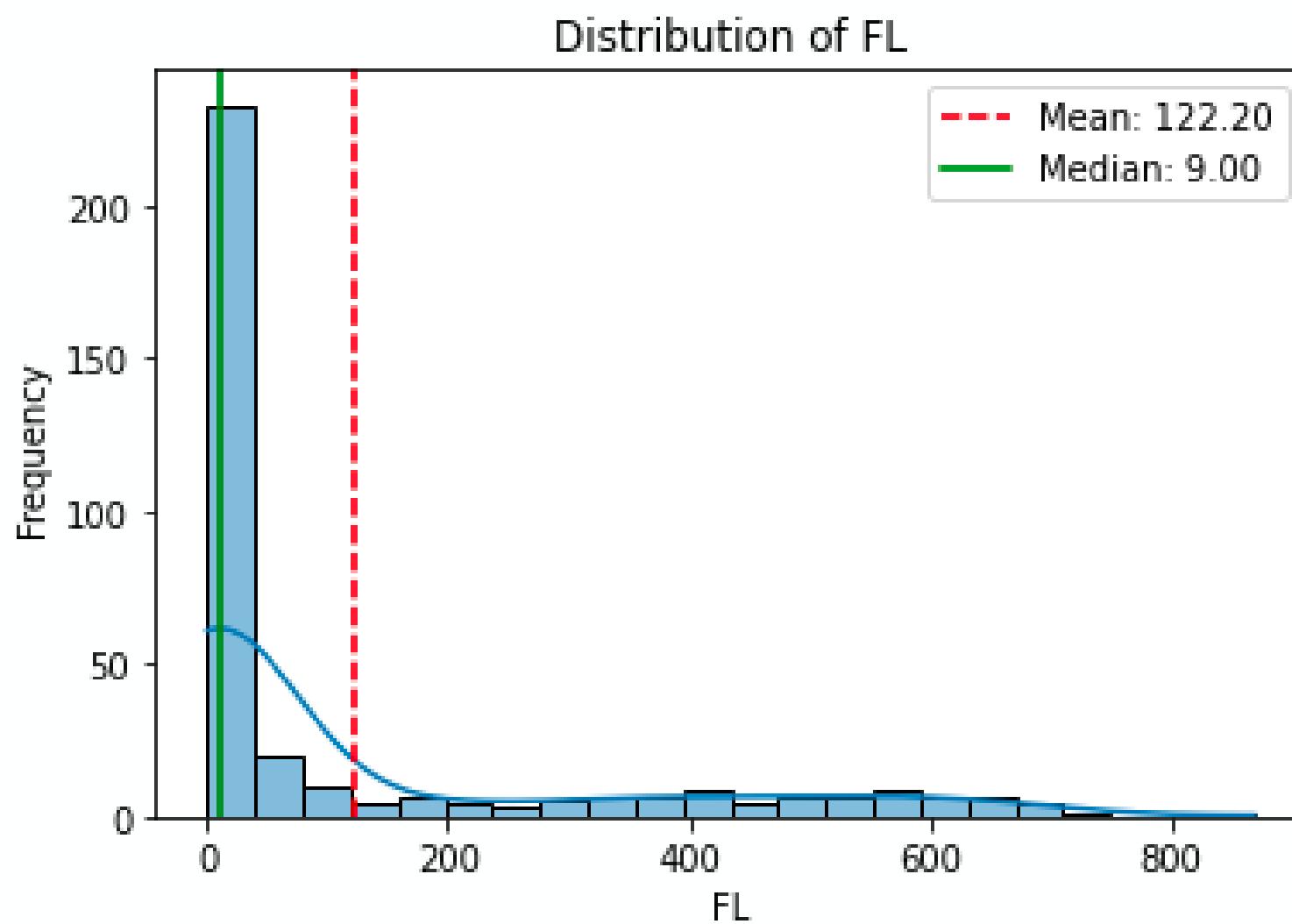


FW(對峙勝利)



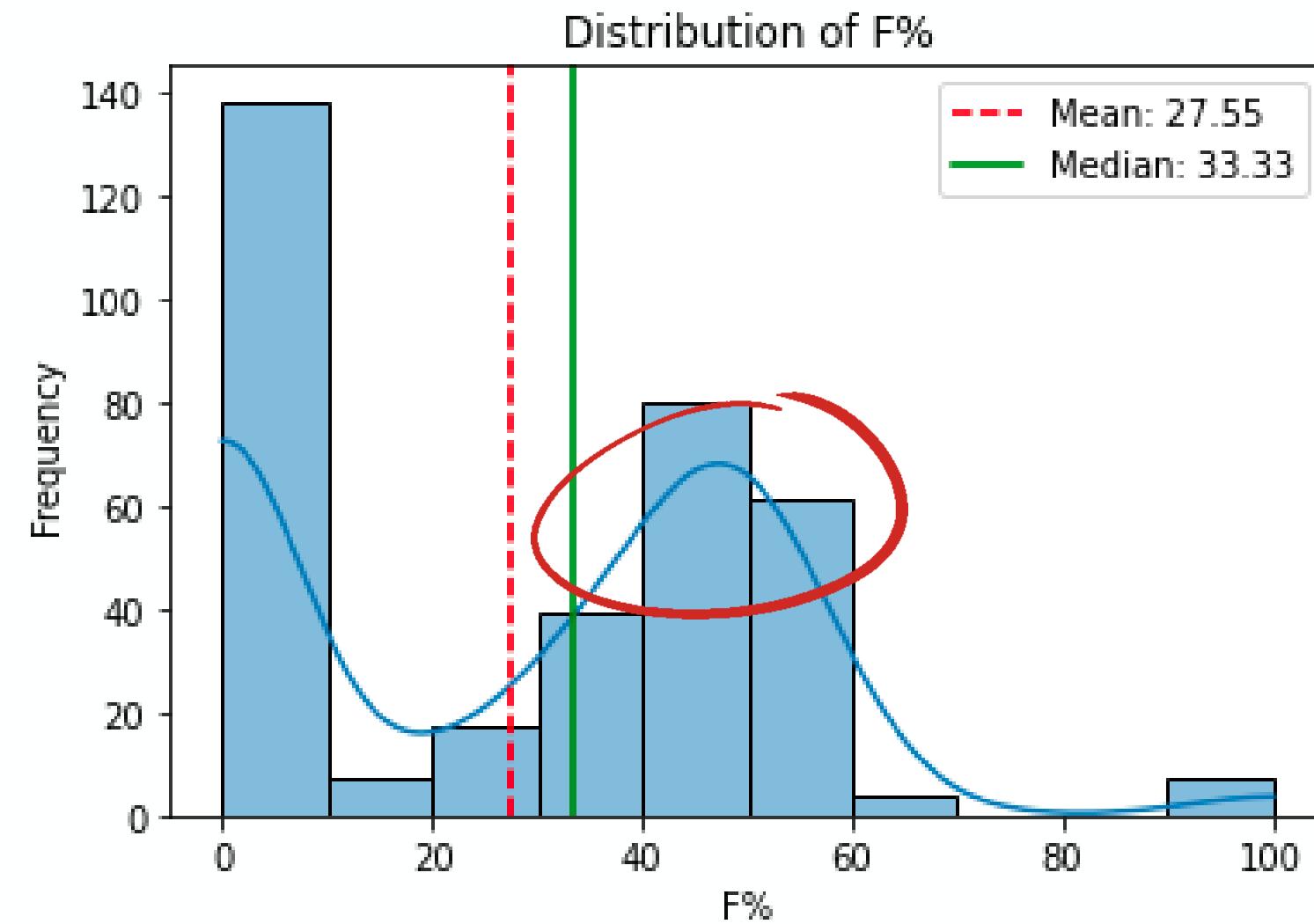
Shapiro Test
p-value = 0.0
拒絕 H_0
不符合正態分佈

FL(對峙失敗)

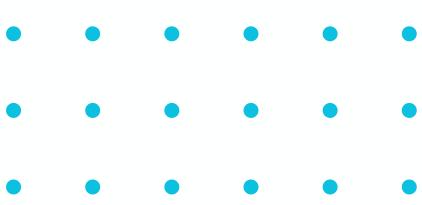


Shapiro Test
p-value = 0.0
拒絕 H_0
不符合正態分佈

F%(對峙成功率)



Shapiro Test
 $p\text{-value} = 0.0$
拒絕 H_0
不符合正態分佈

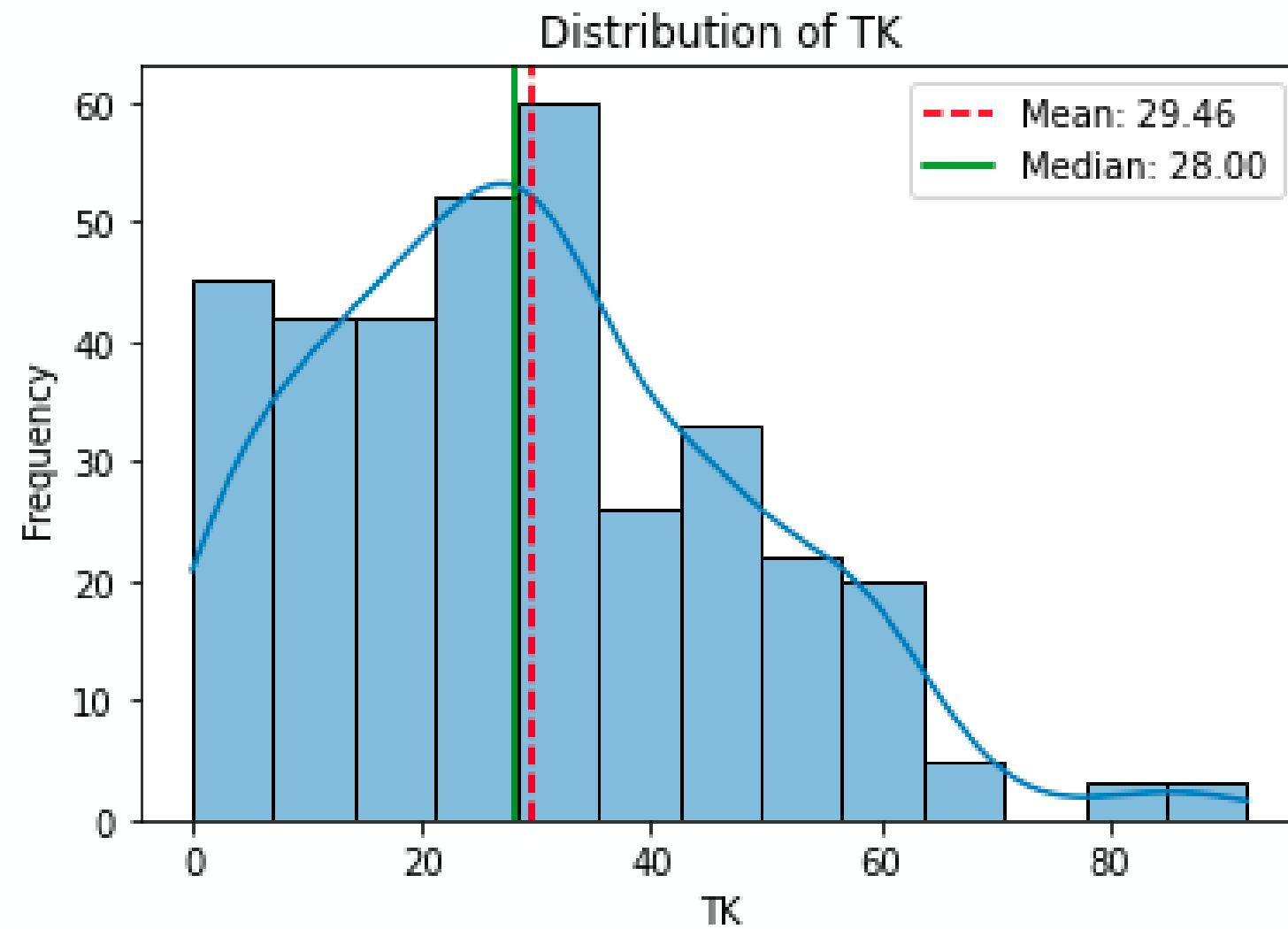


防守分析

- 1.TK(抄截)
- 2.BS(防守數)

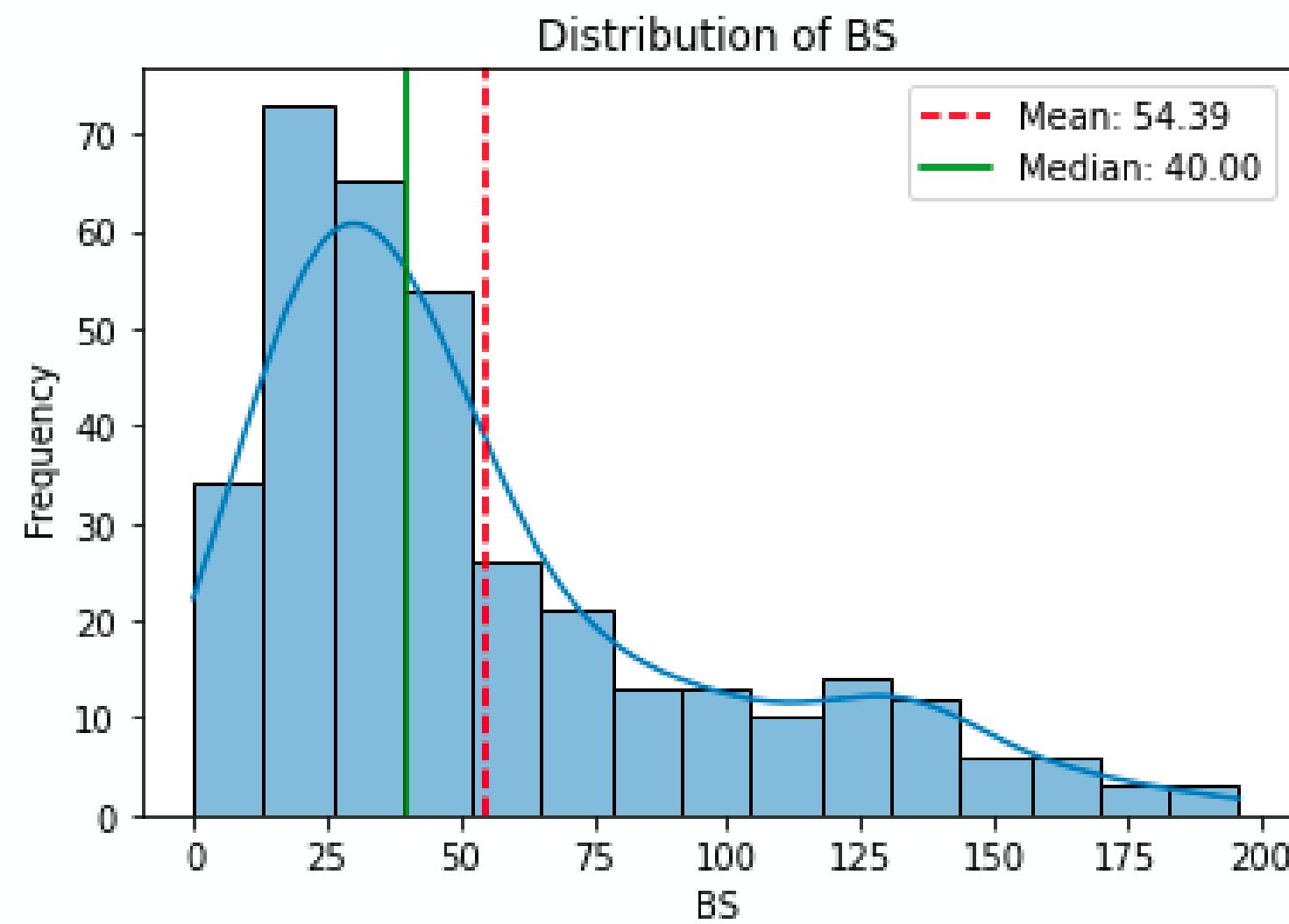


TK(抄截)

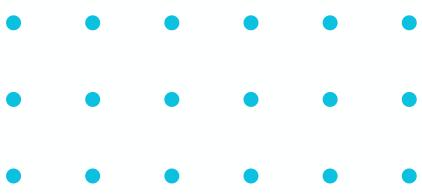


Shapiro Test
p-value = 0.0
拒絕 H_0
不符合正態分佈

BS(防守數)



Shapiro Test
 $p\text{-value} = 0.0$
拒絕 H_0
不符合正態分佈

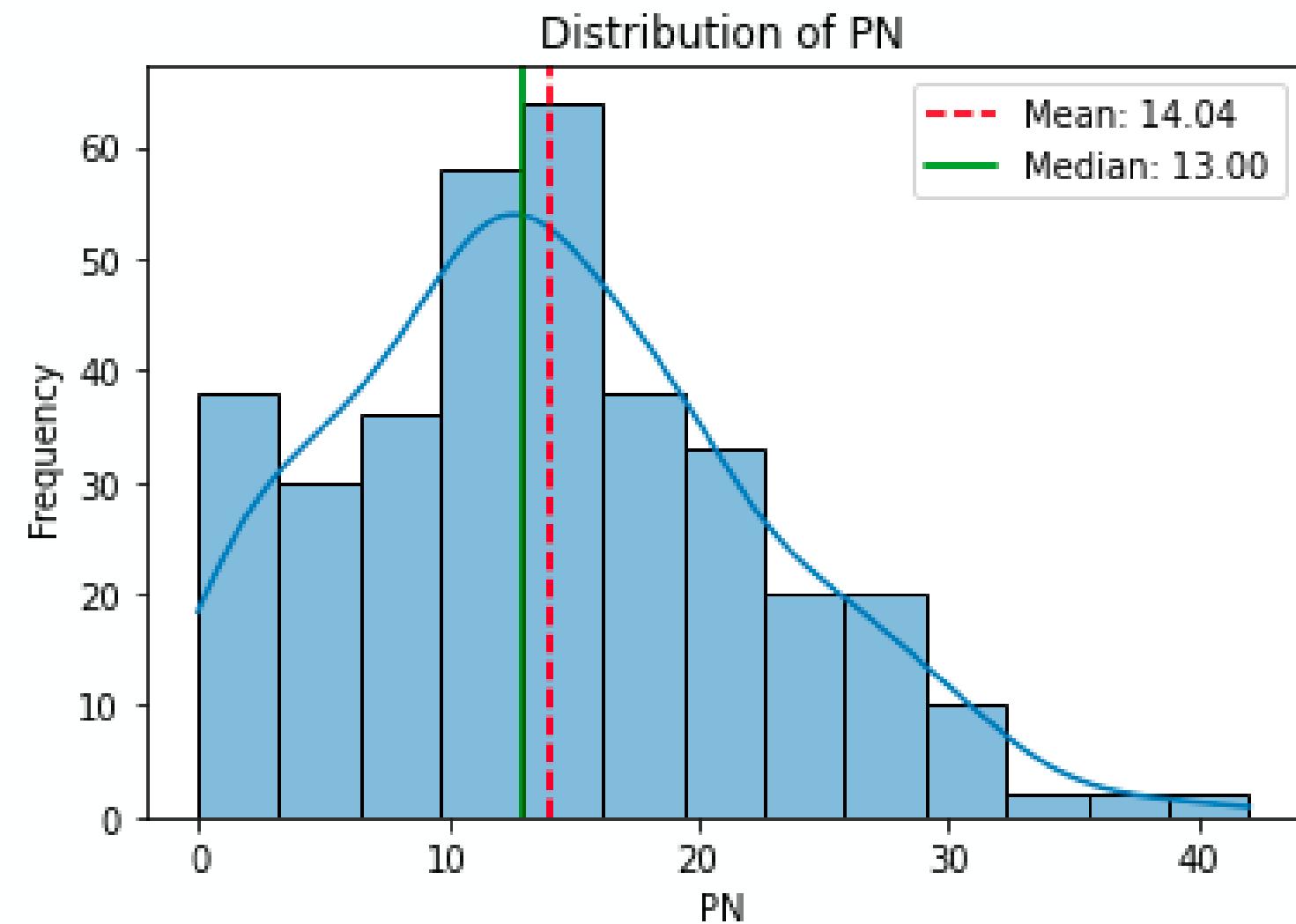


犯規分析

1. PN(處罰)
2. PIM(罰球犯規分鐘數)
3. H(故意與持有冰球的球員發生接觸，使其失去冰球)

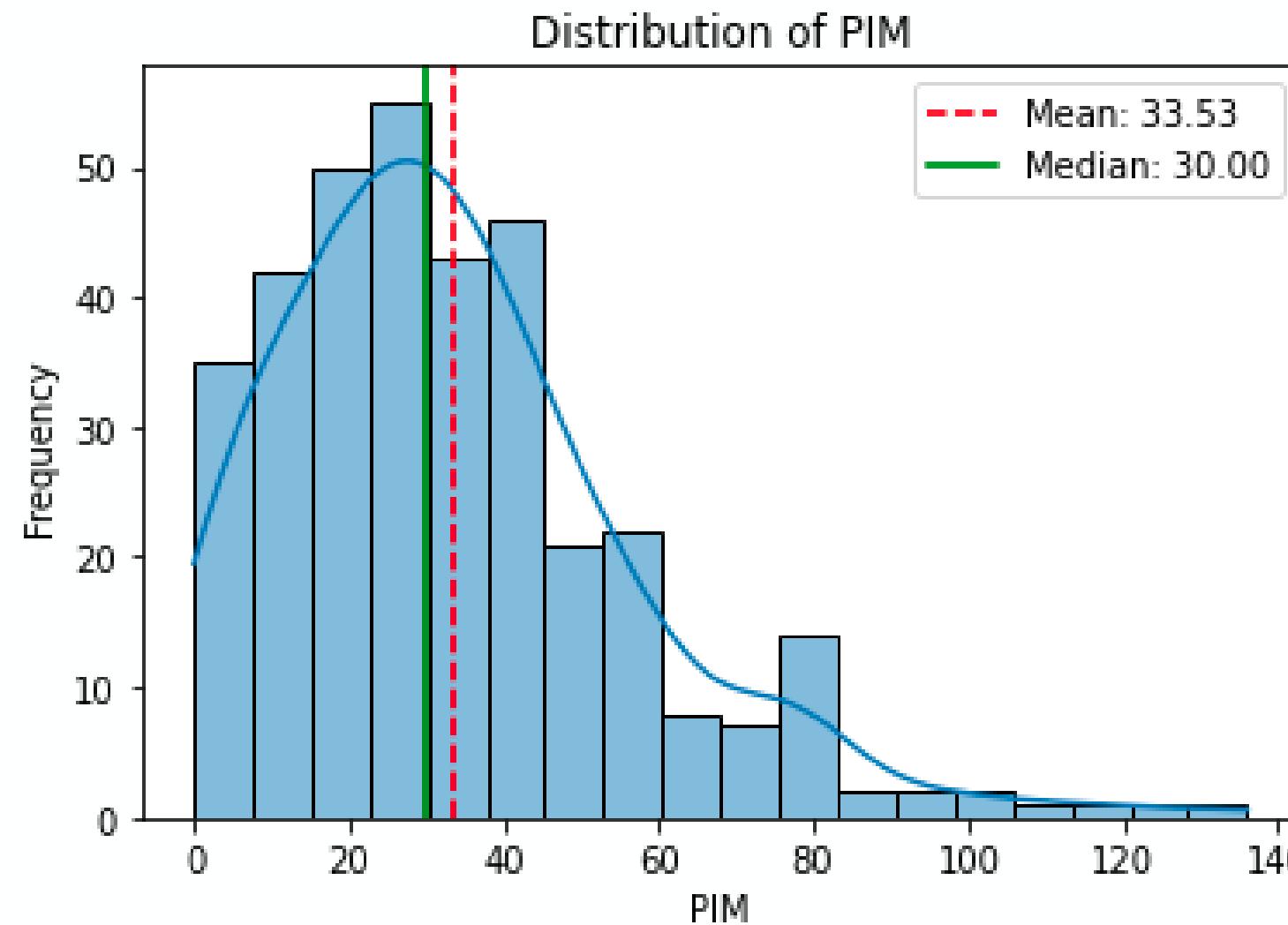


PN(處罰)



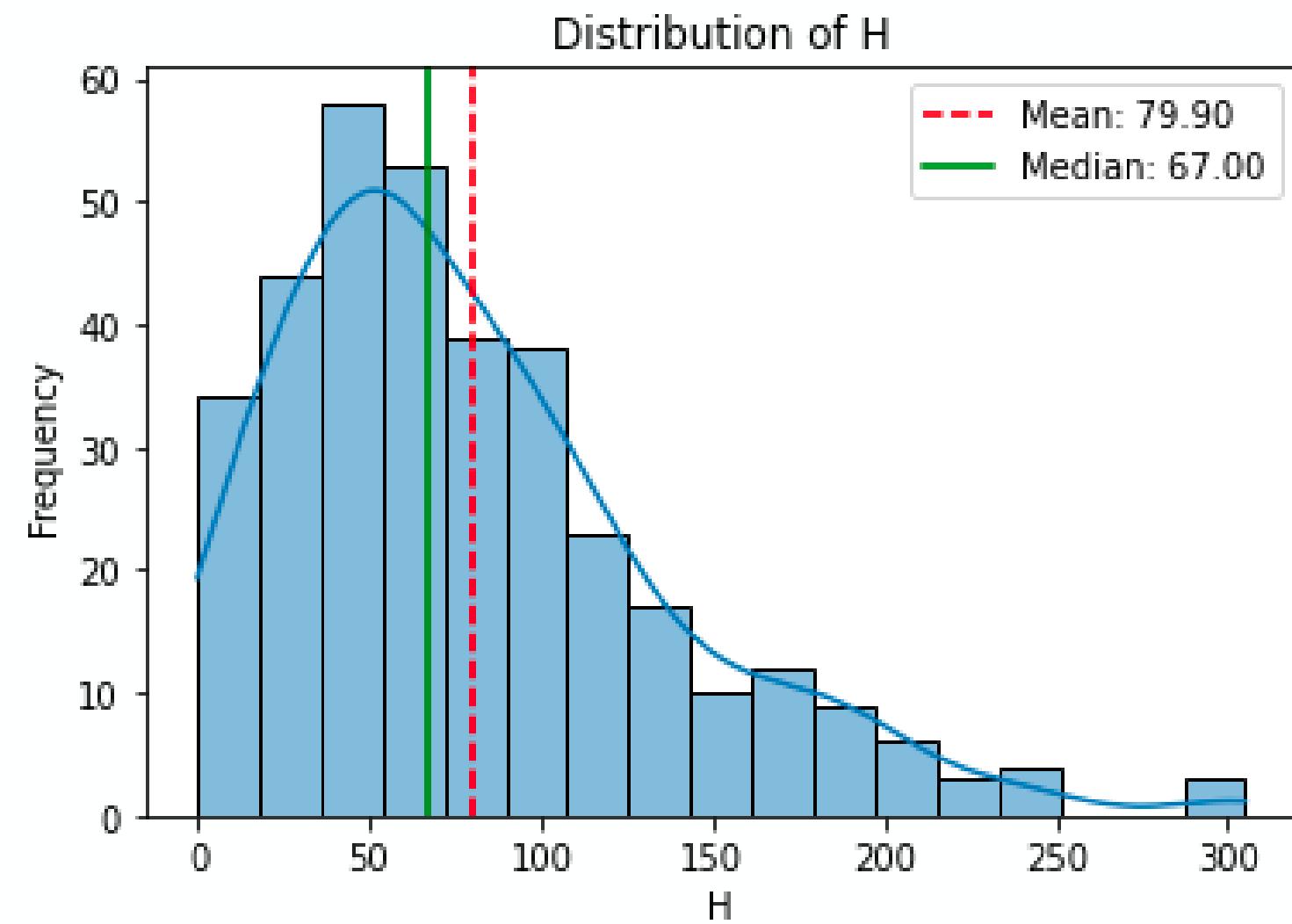
Shapiro Test
p-value = 0.000023
拒絕 H_0
不符合正態分佈

PIM(罰球犯規分鐘數)

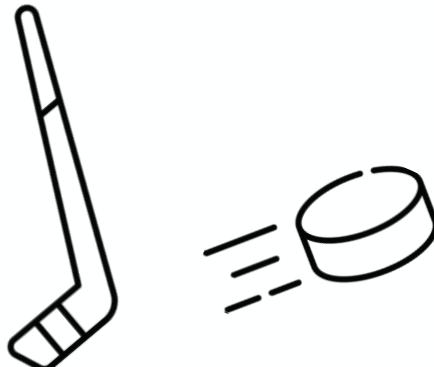


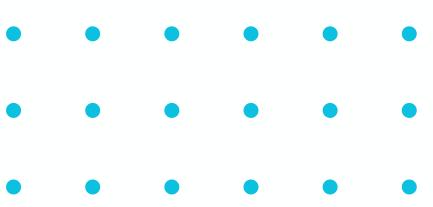
Shapiro Test
 $p\text{-value} = 0.0$
拒絕 H_0
不符合正態分佈

H(故意與持有冰球的球員發生接觸，使其失去冰球)



Shapiro Test
 $p\text{-value} = 0.0$
拒絕 H_0
不符合正態分佈



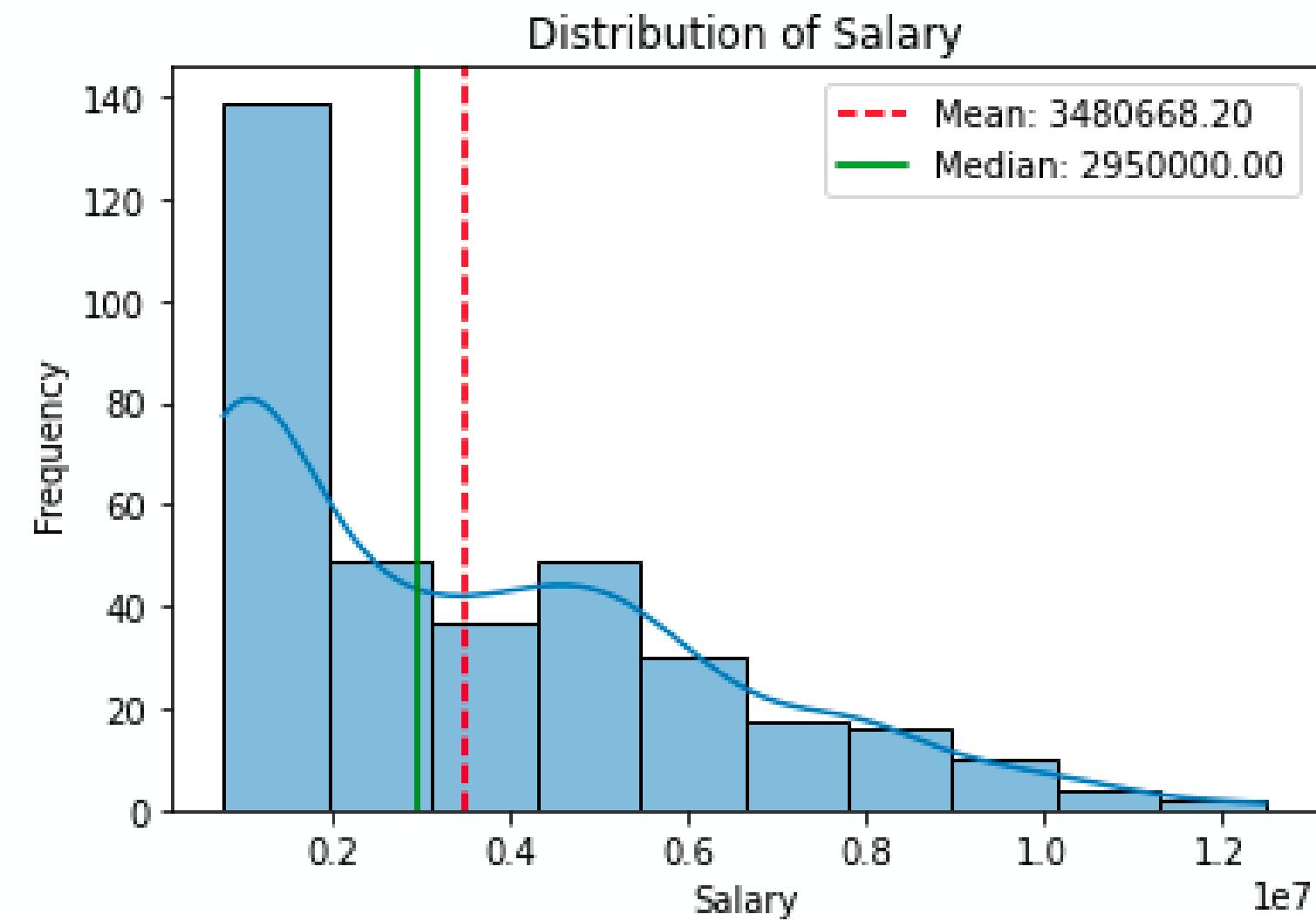


薪水

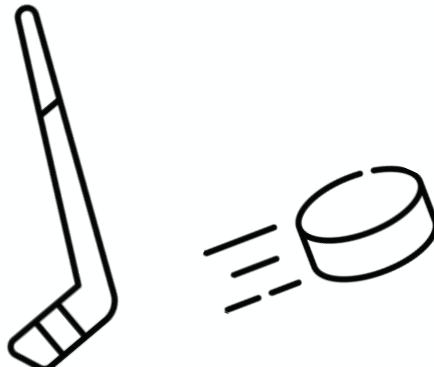
1. Cap-hit(頂薪)
2. Salary(薪水)



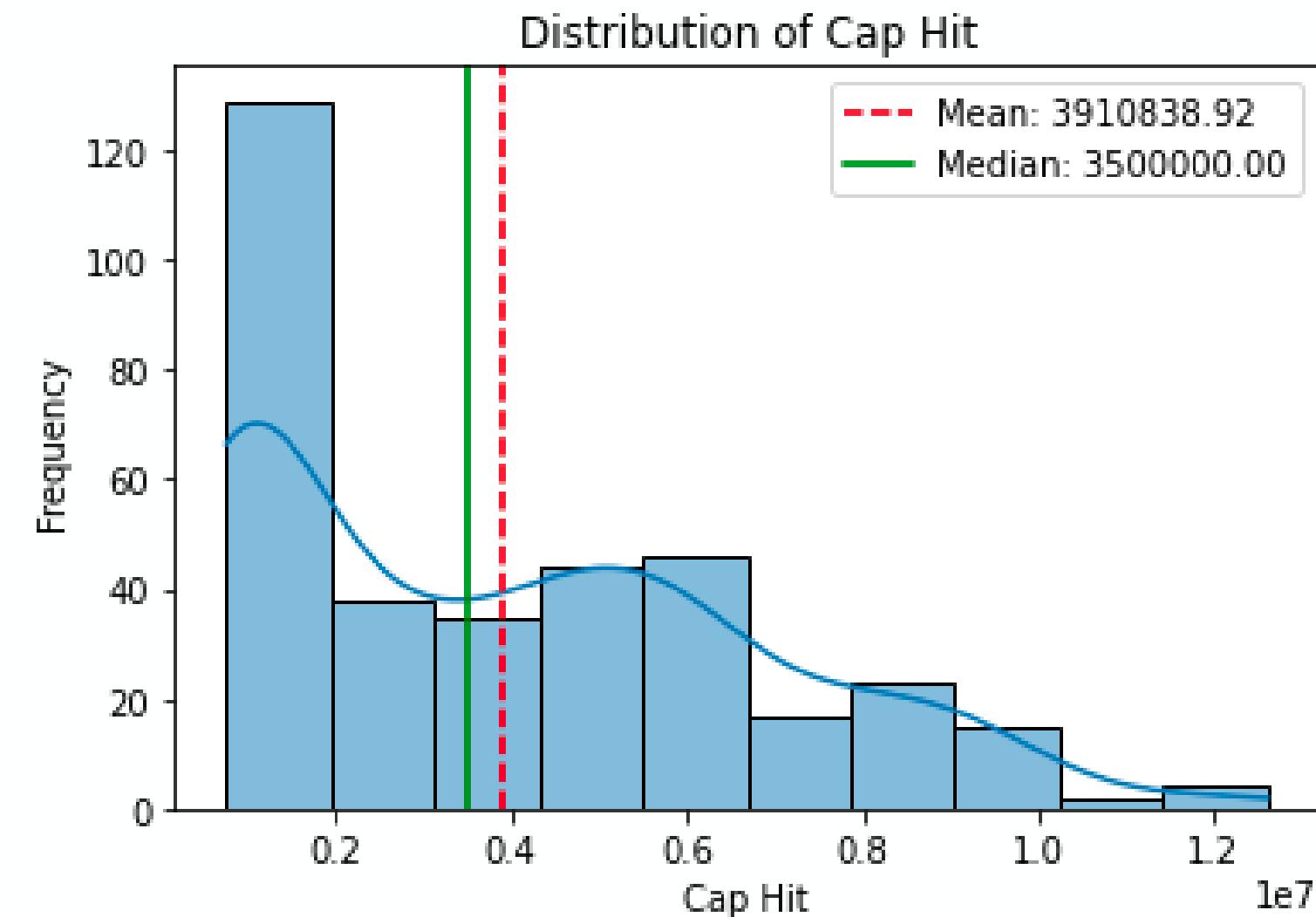
Salary(薪水)



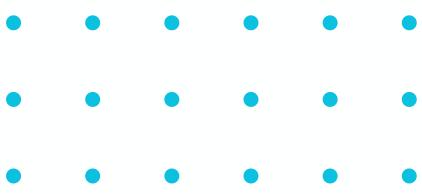
Shapiro Test
p-value = 0.0
拒絕 H_0
不符合正態分佈



Cap-hit(頂薪)



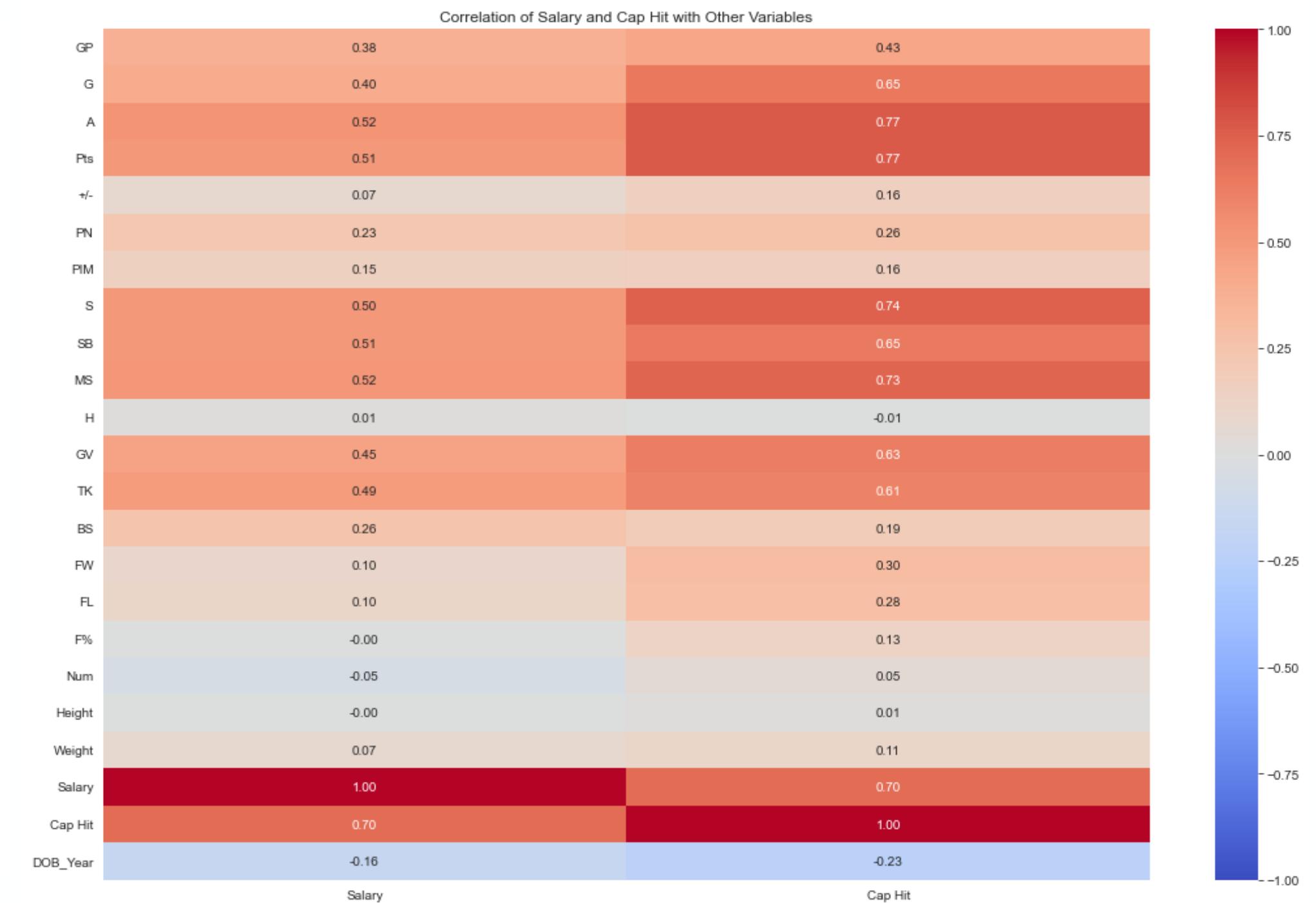
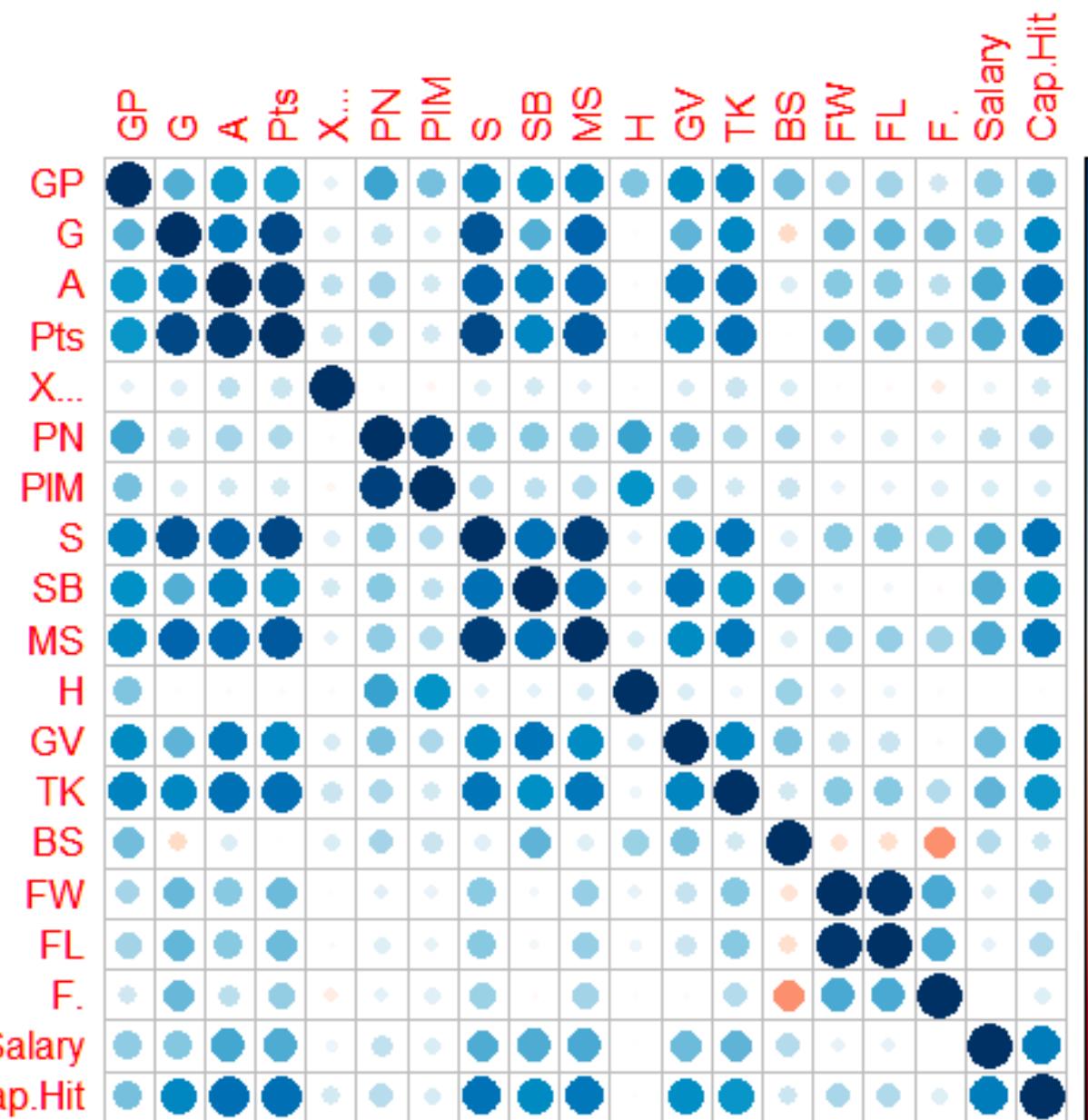
Shapiro Test
p-value = 0.0
拒絕 H_0
不符合正態分佈



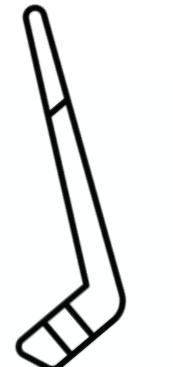
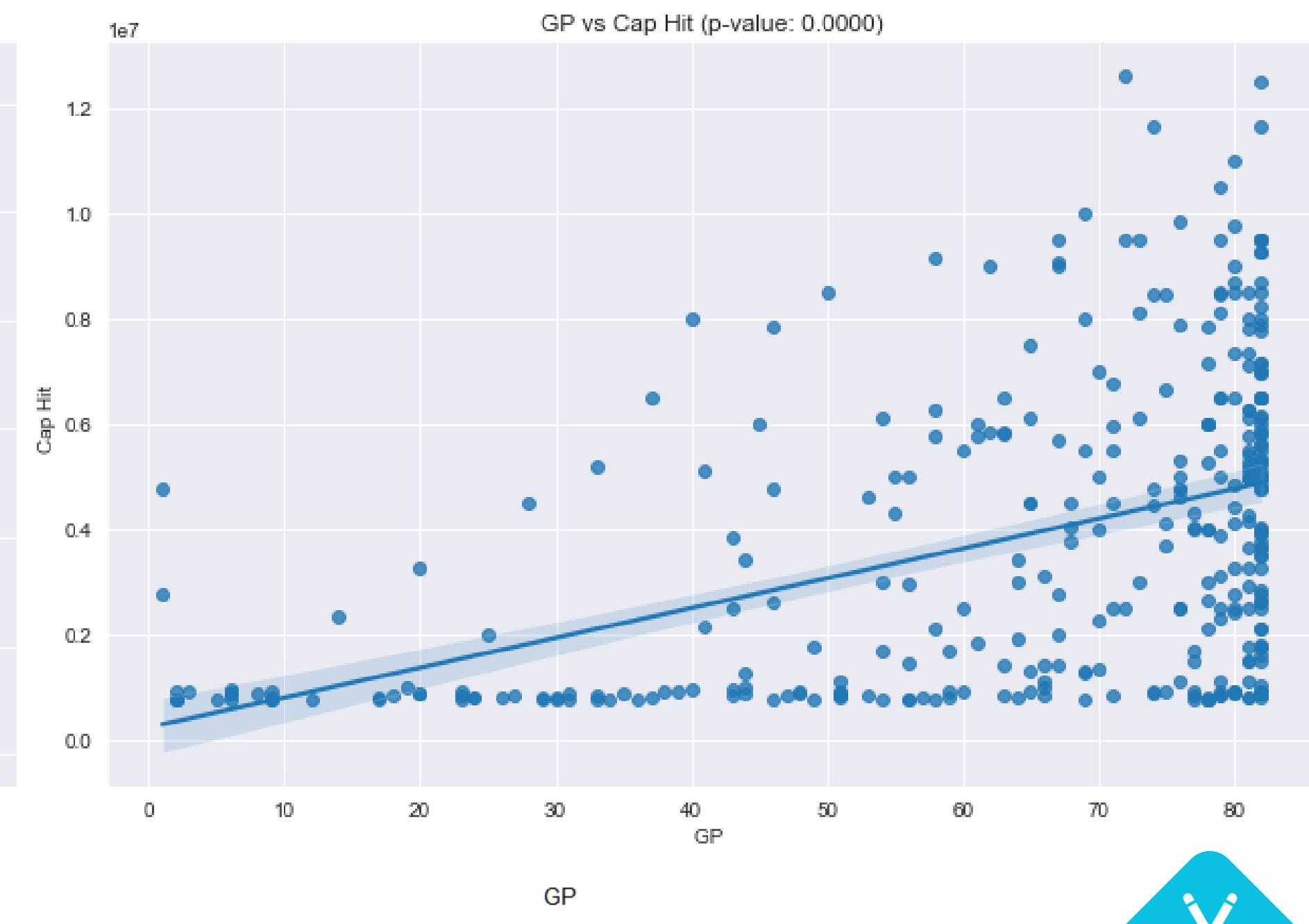
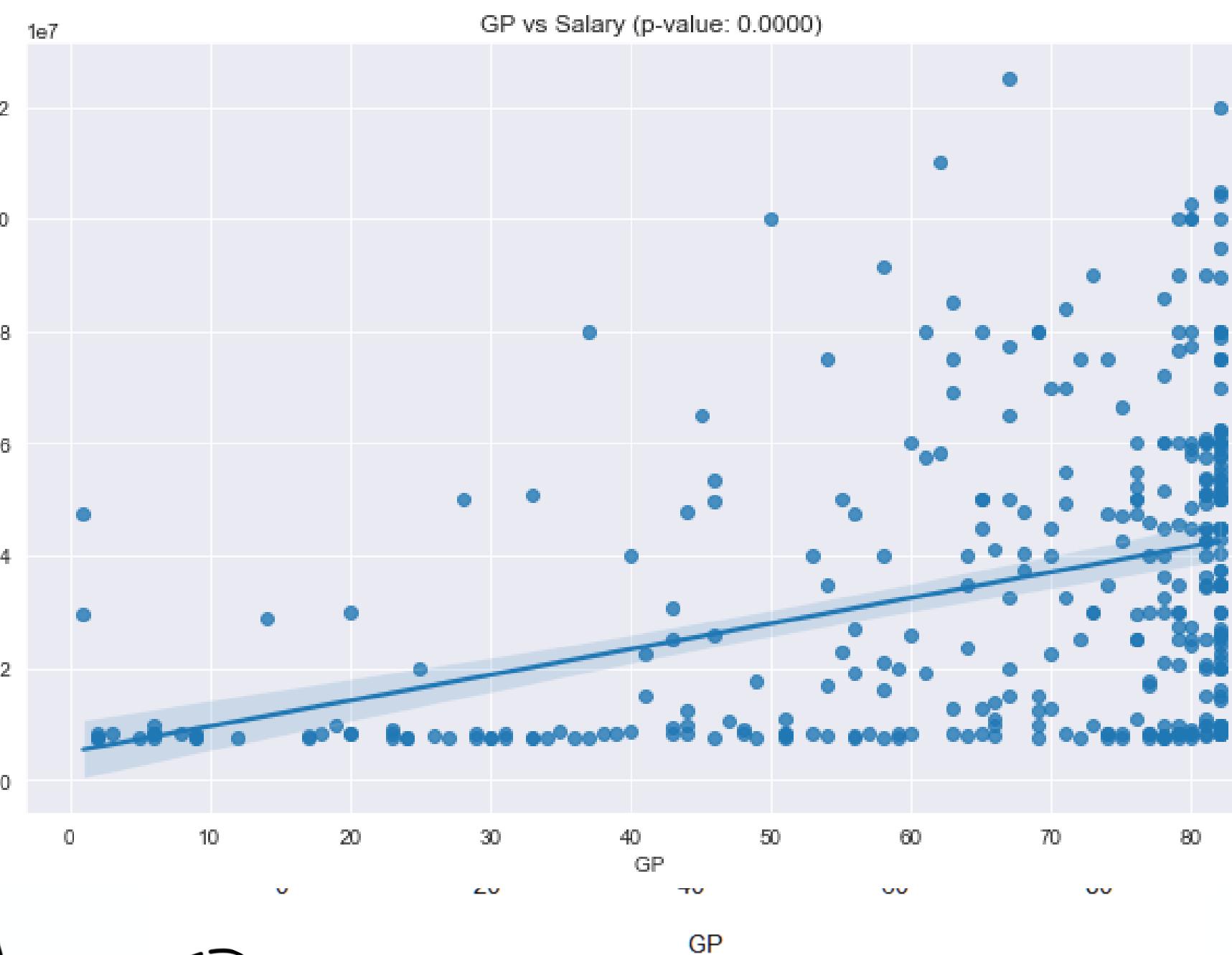
雙變量分析



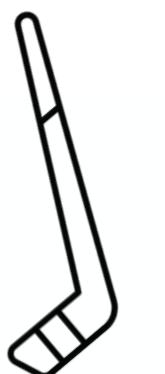
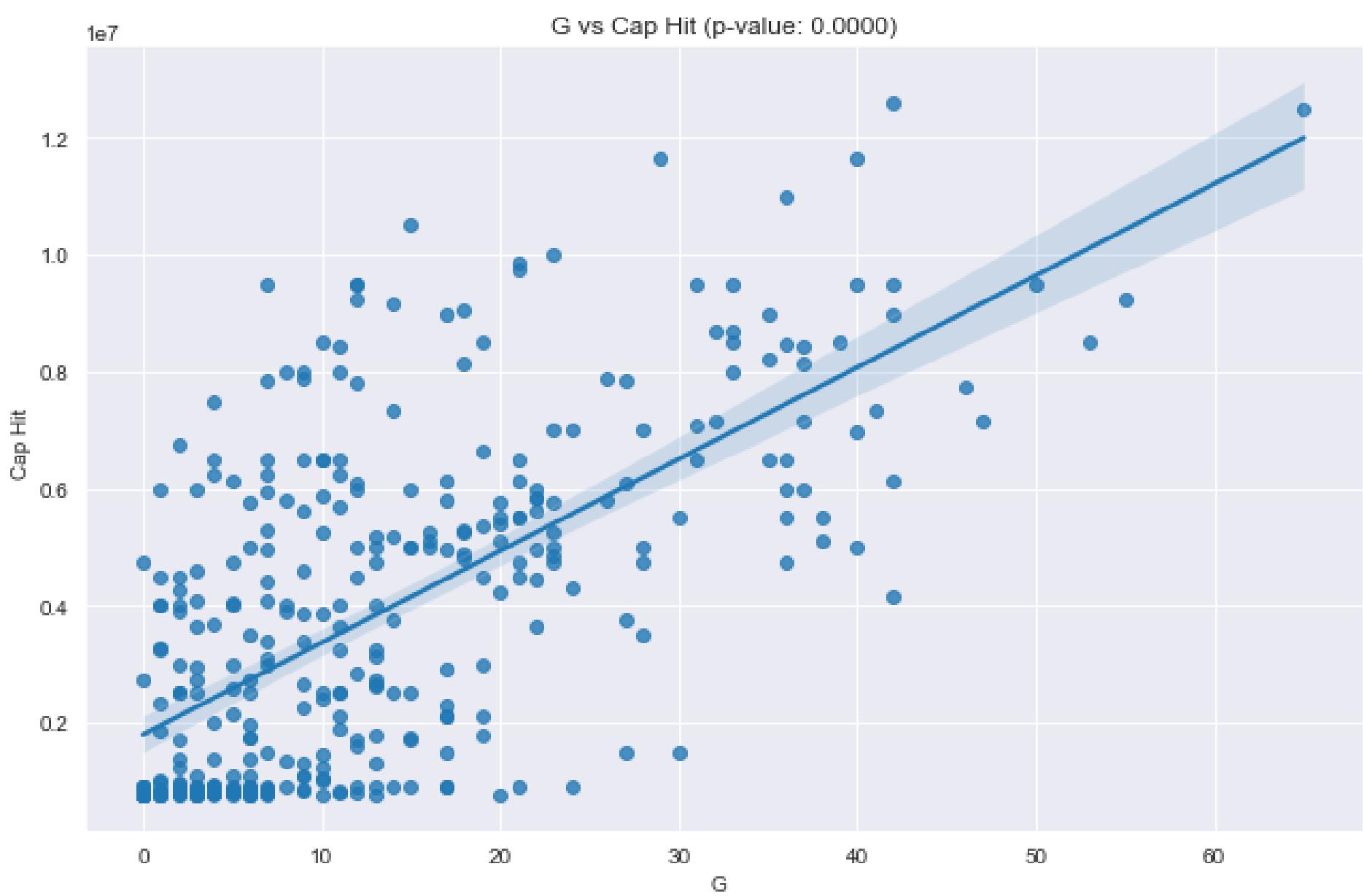
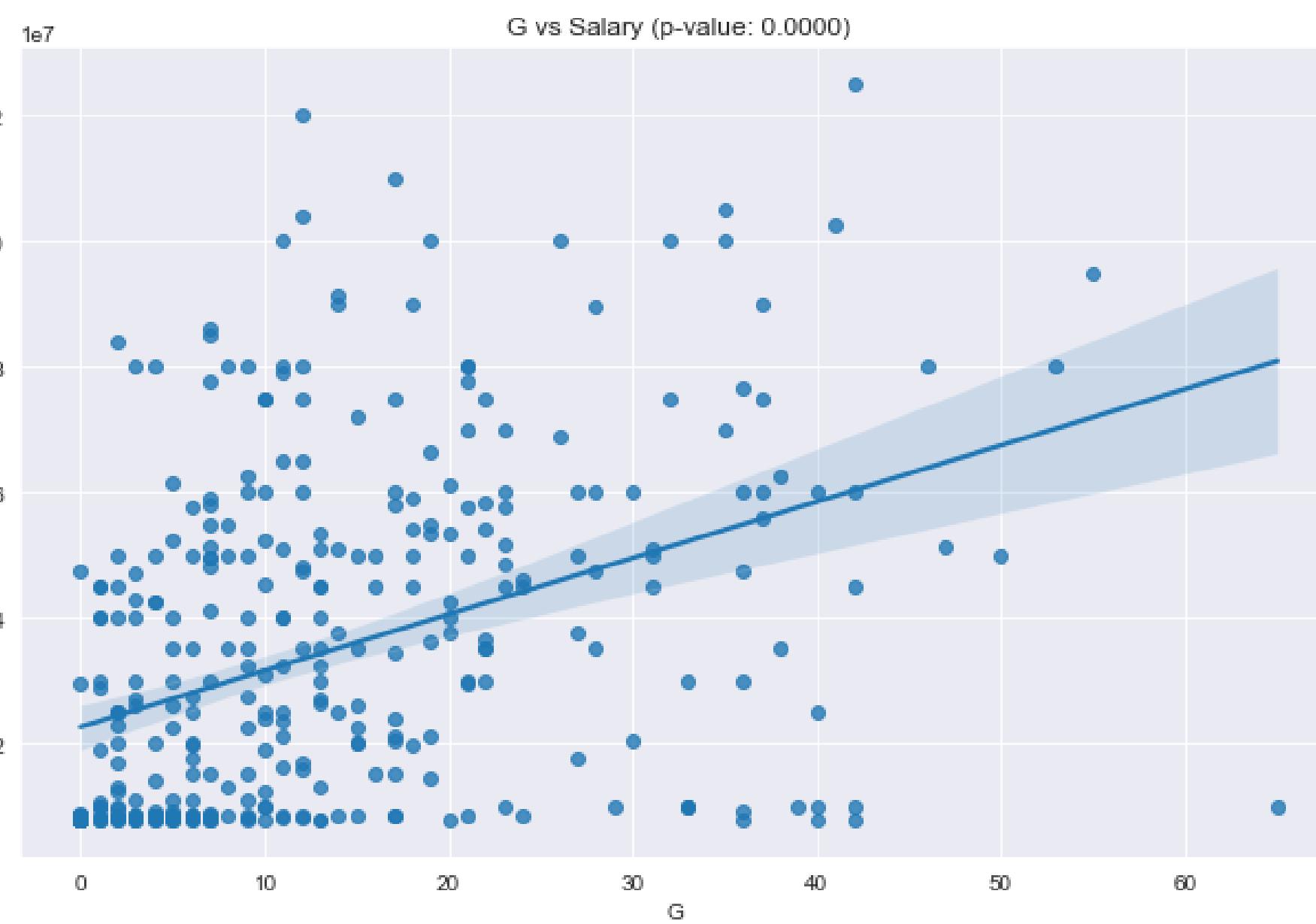
相關矩陣



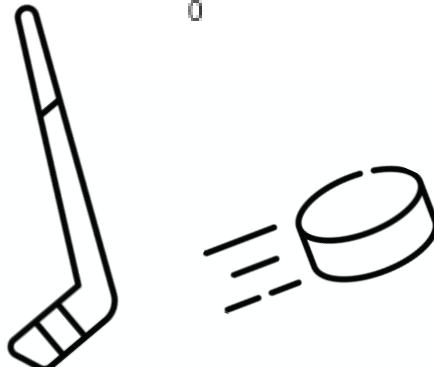
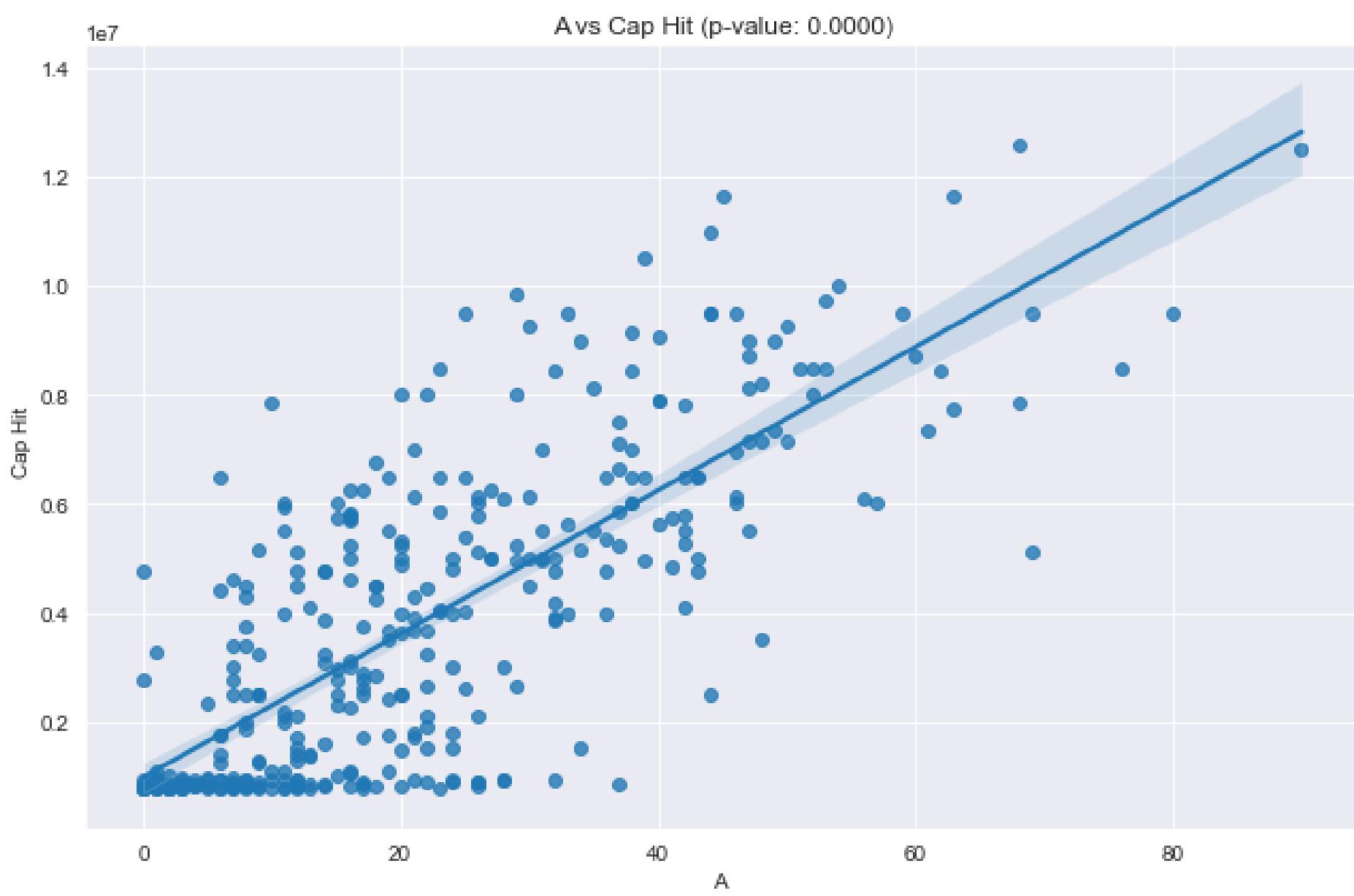
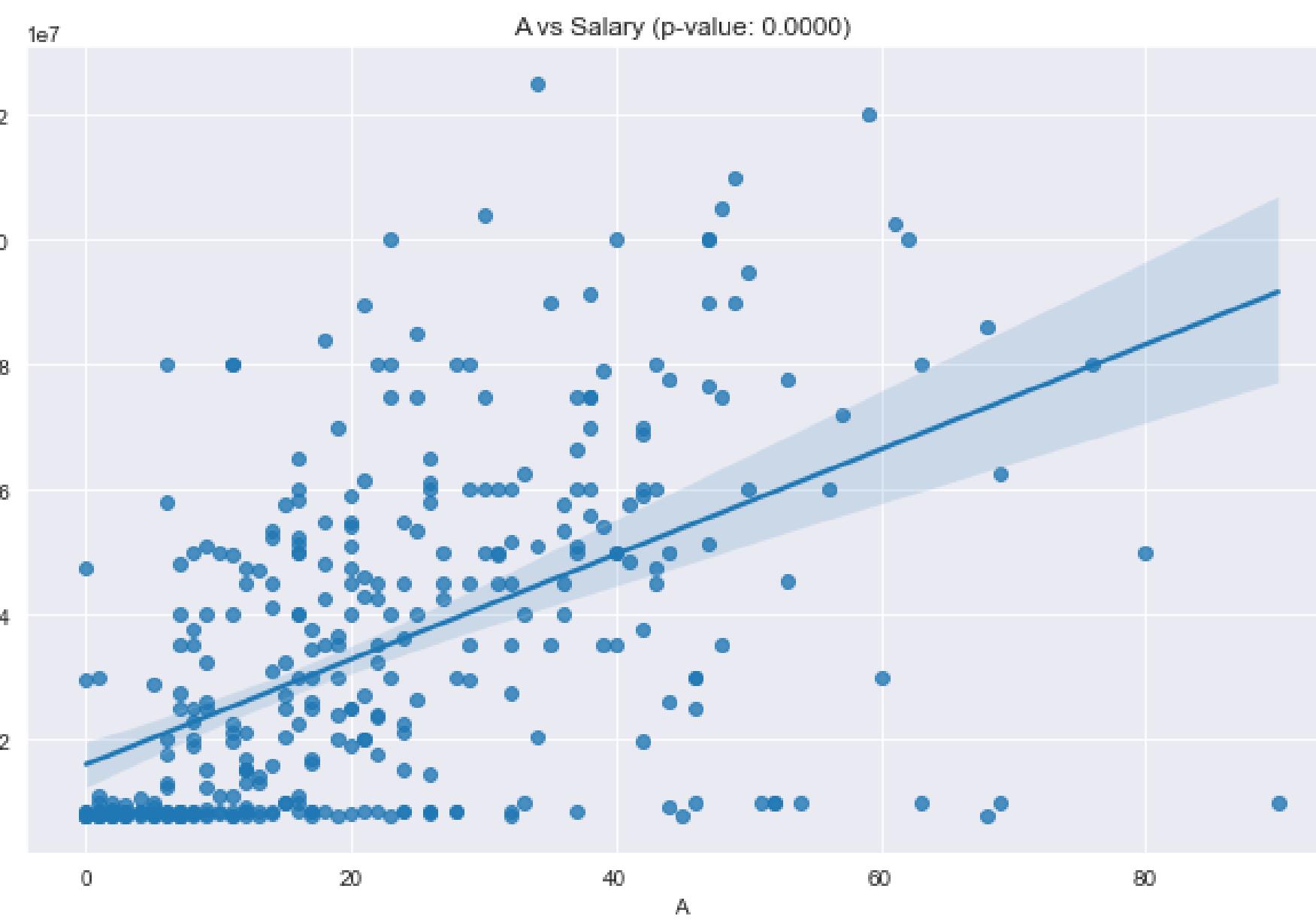
GP(參加比賽數量)



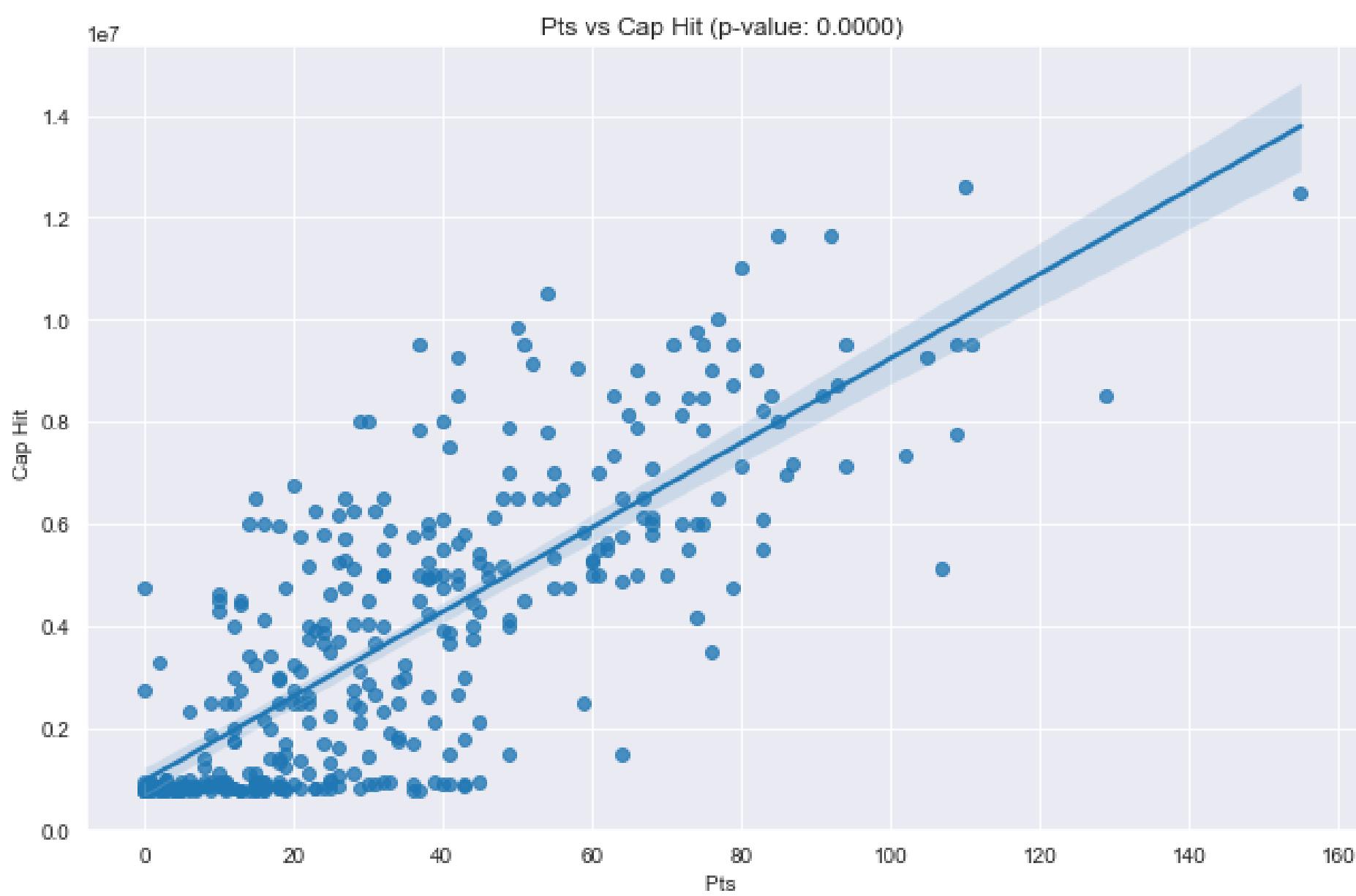
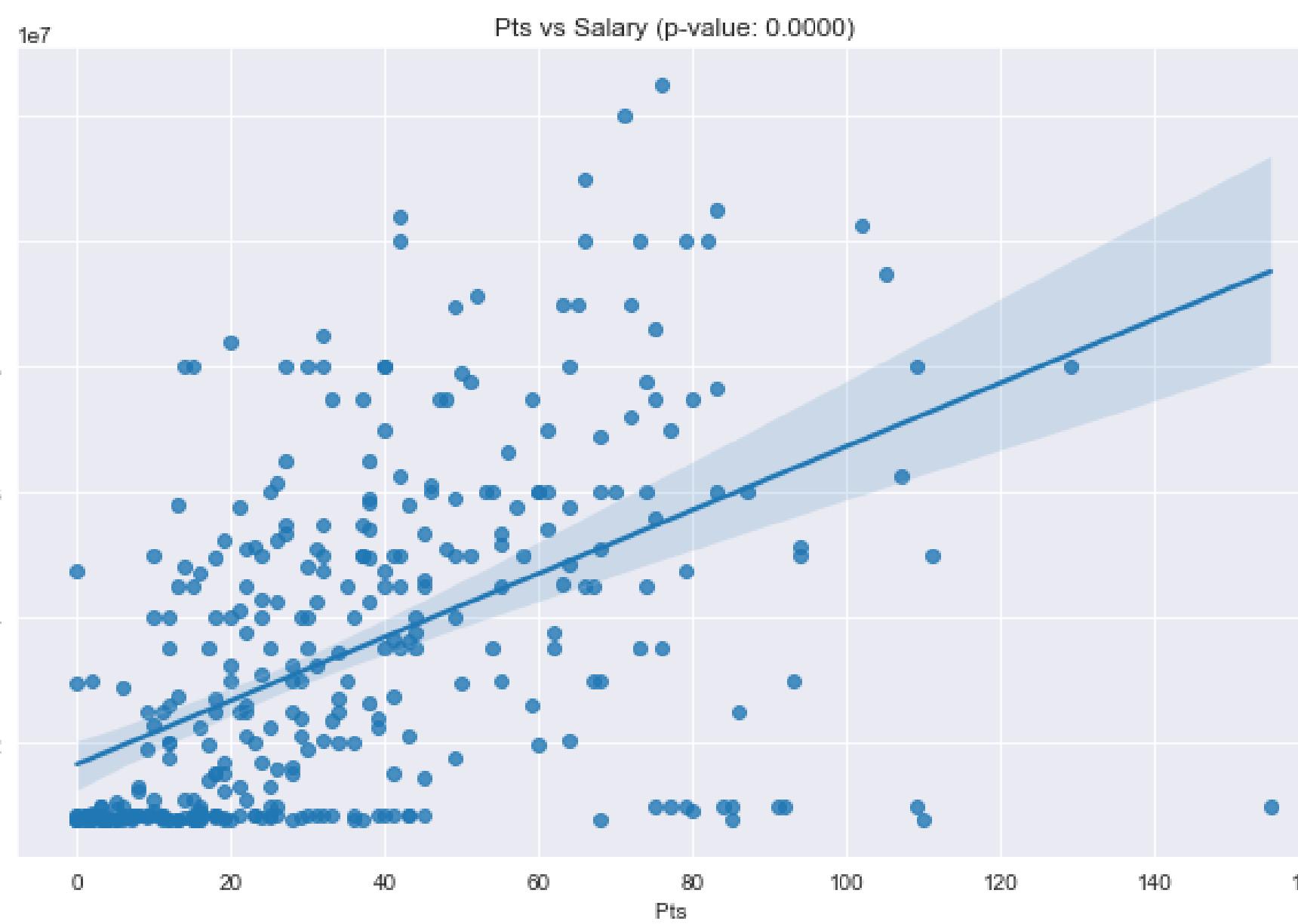
G(進球數)



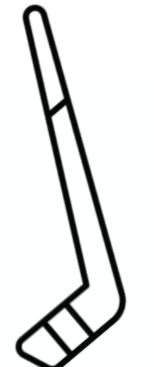
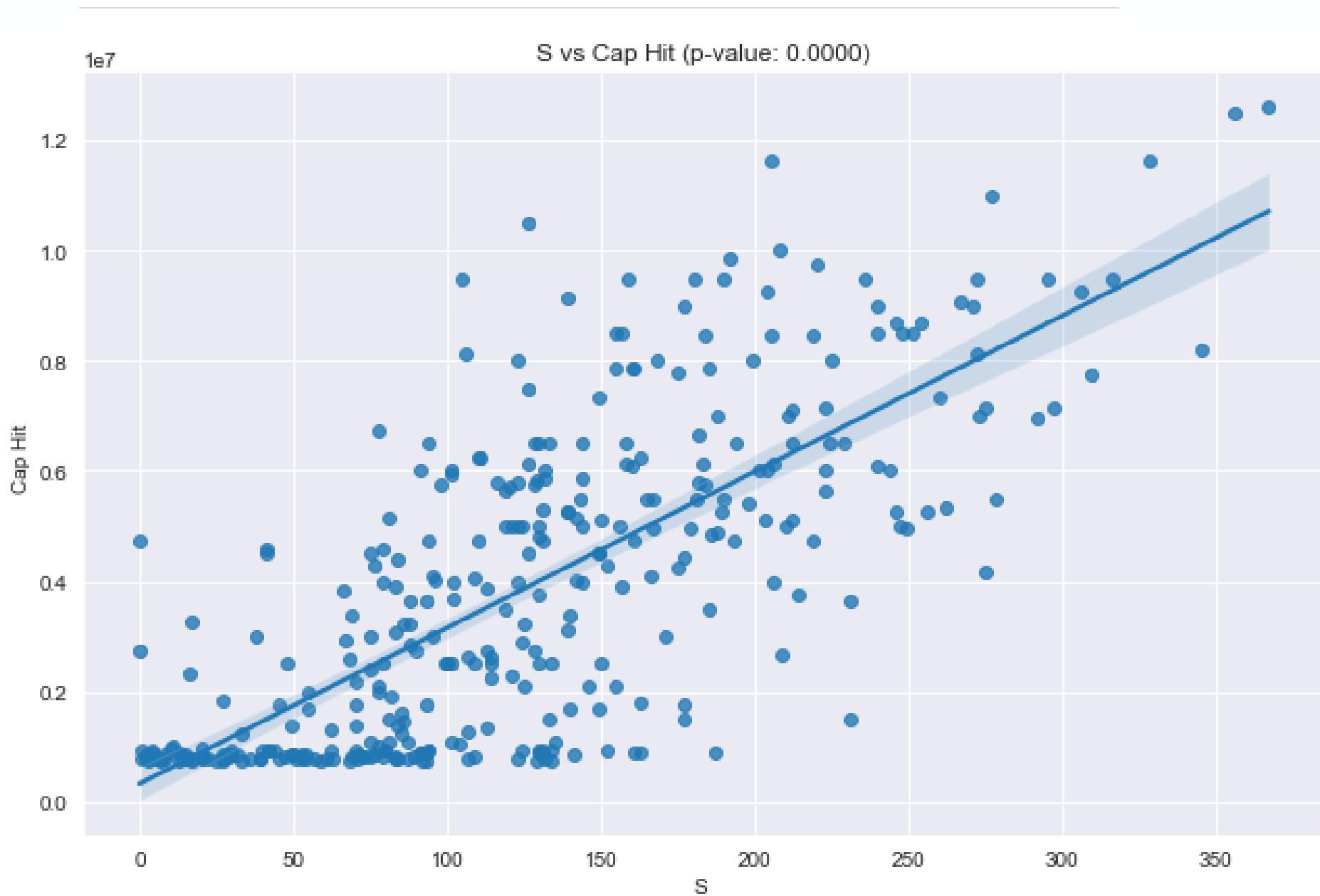
A(助攻數)



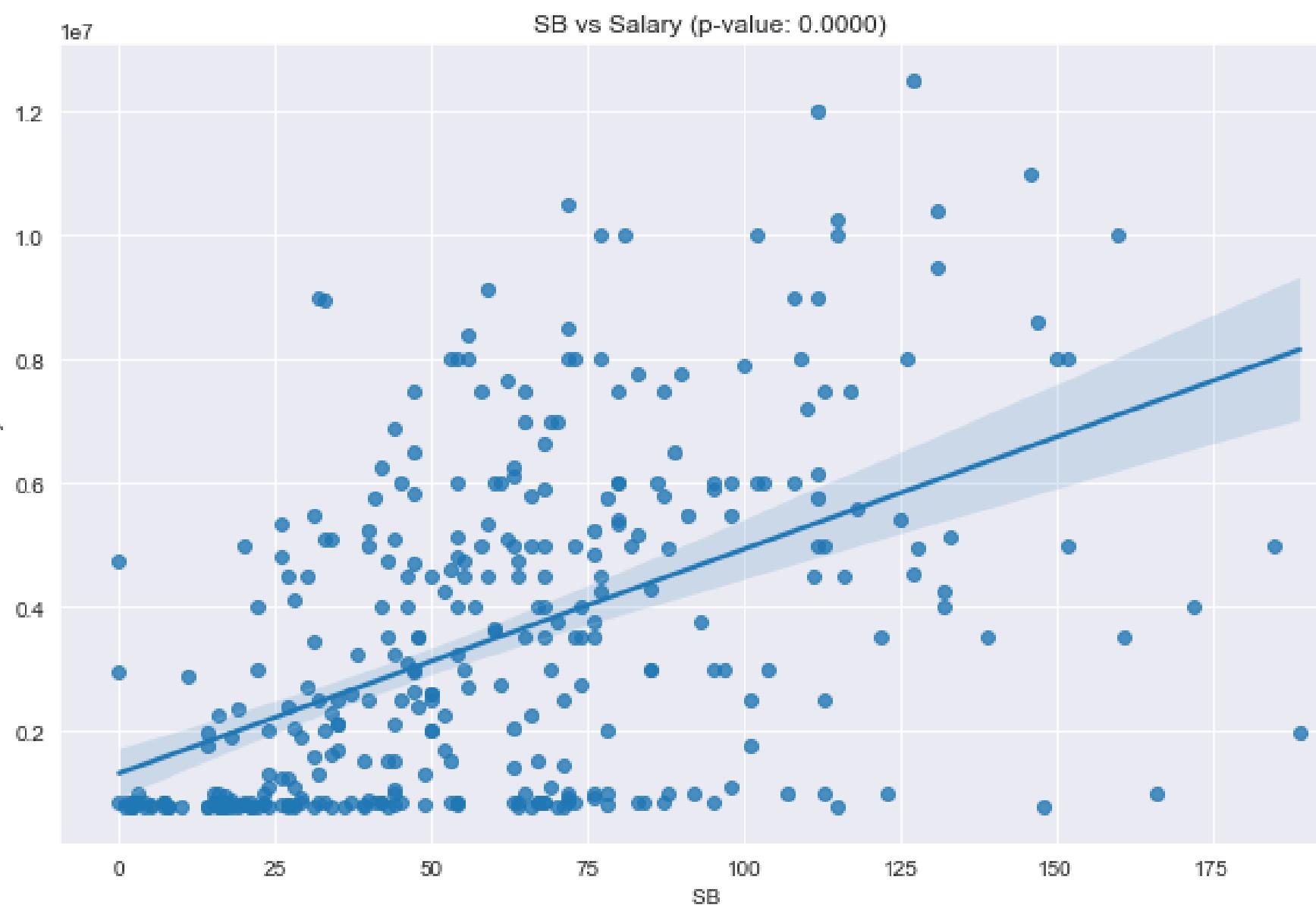
Pts(分數)



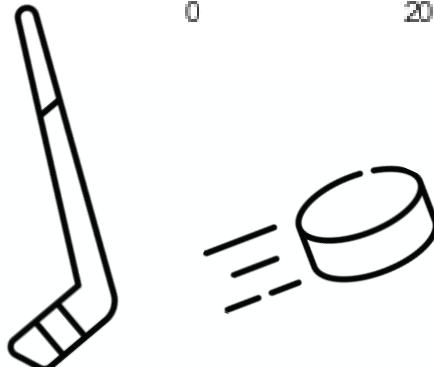
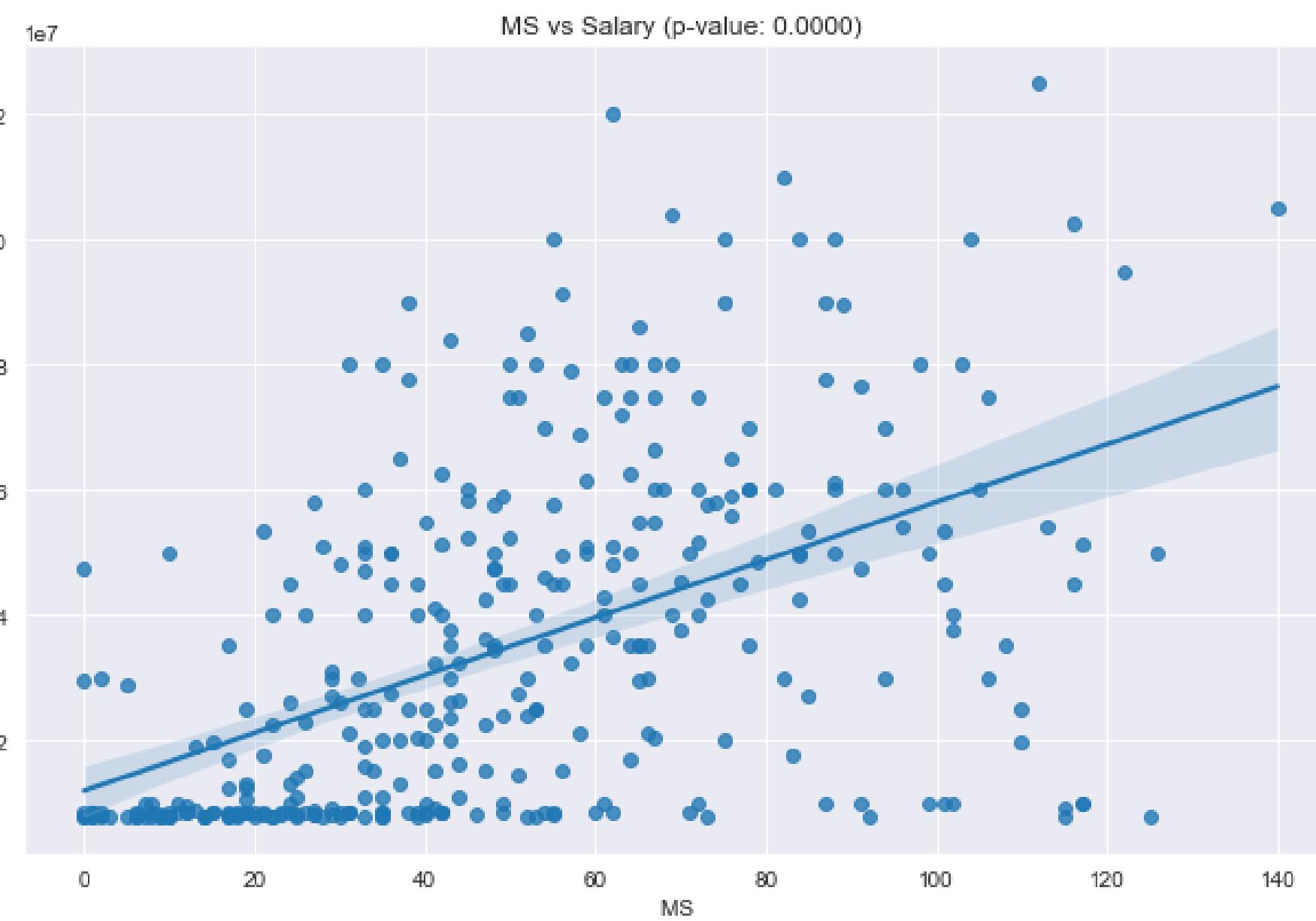
S(射門次數)



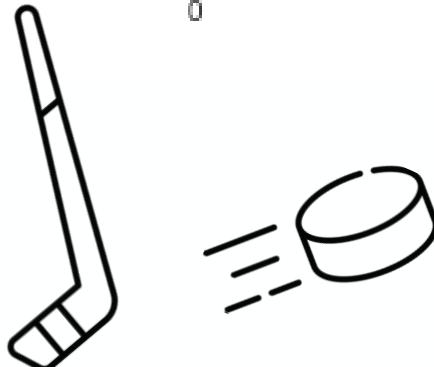
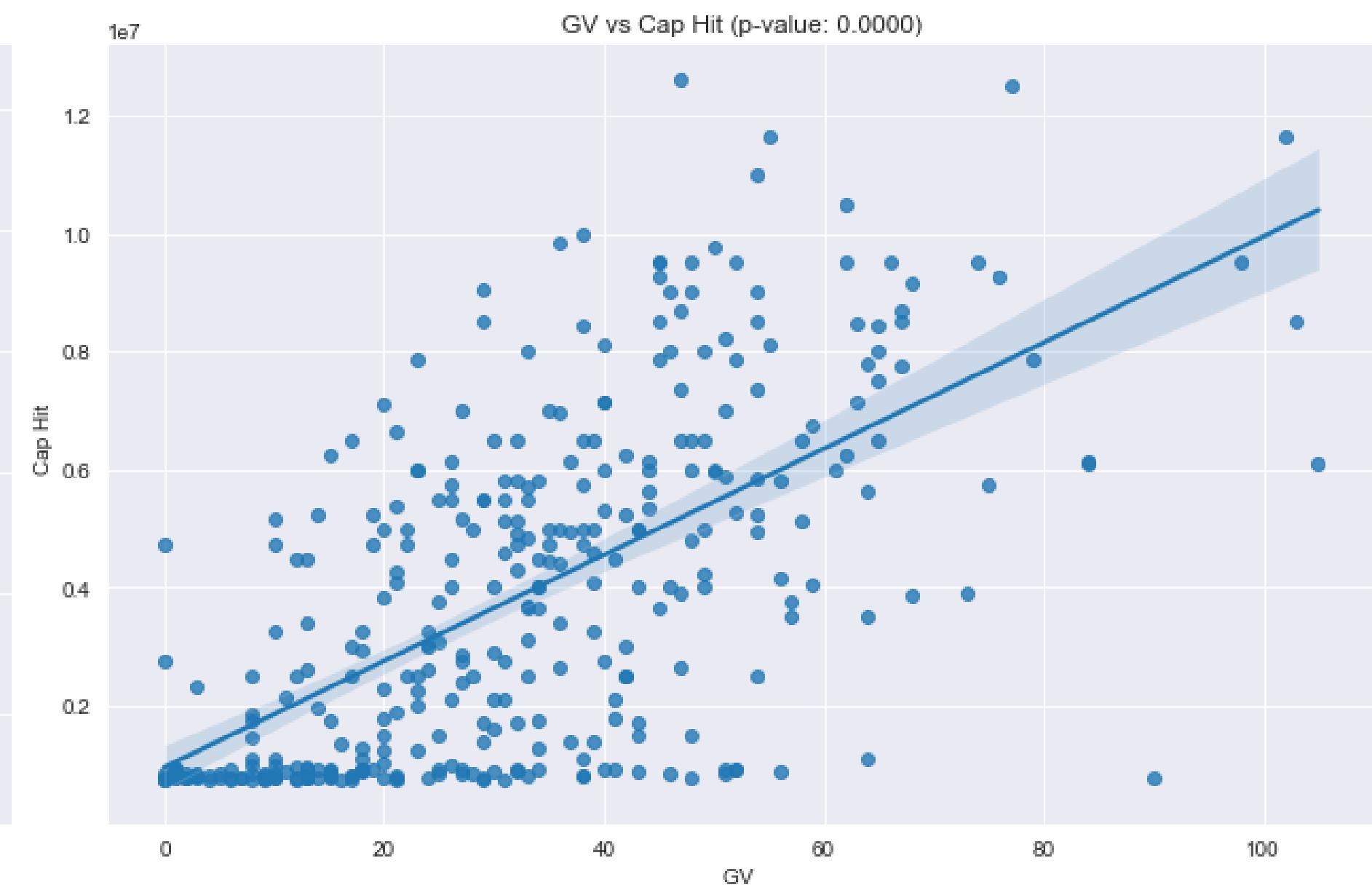
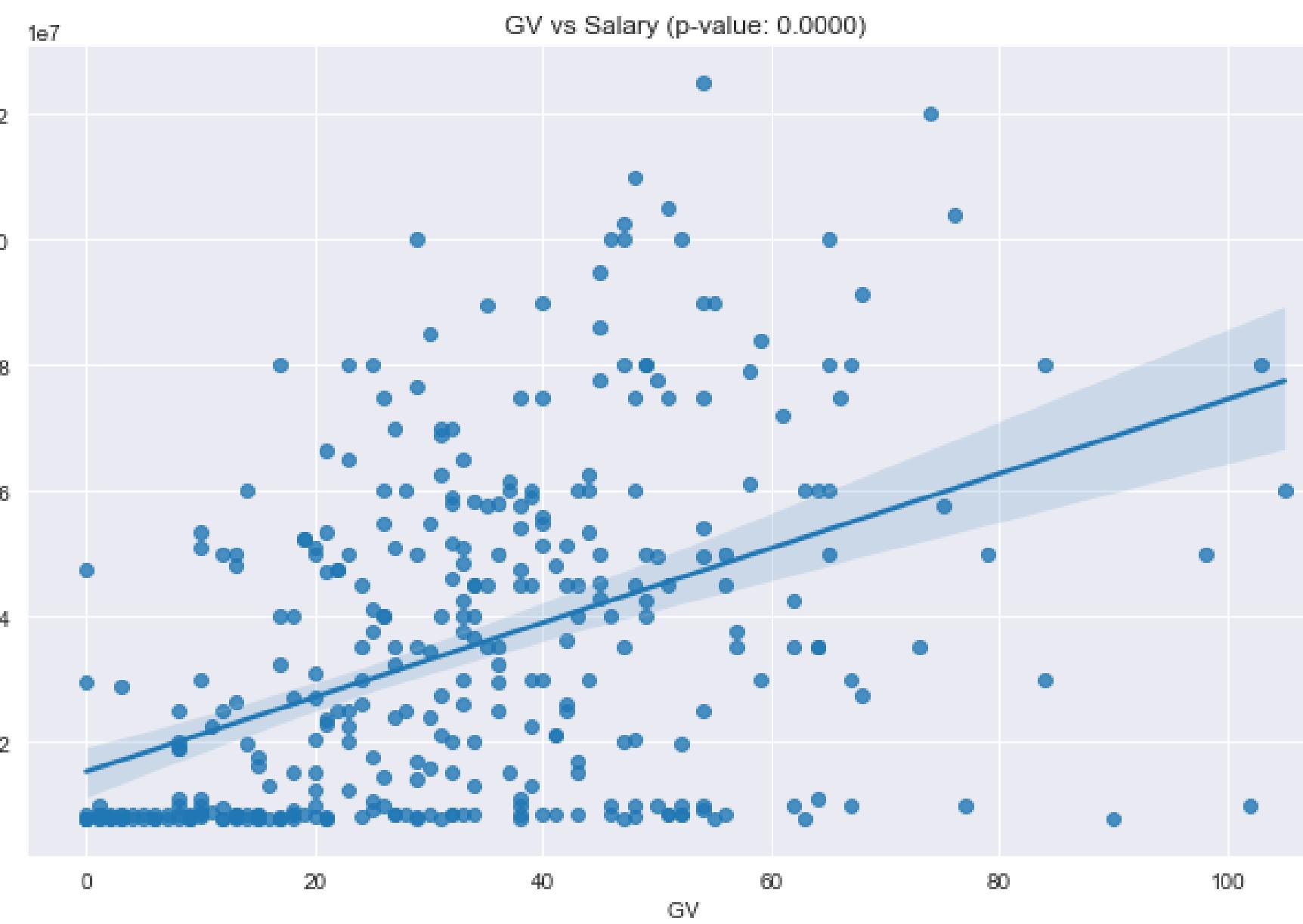
SB(被阻擋球數)



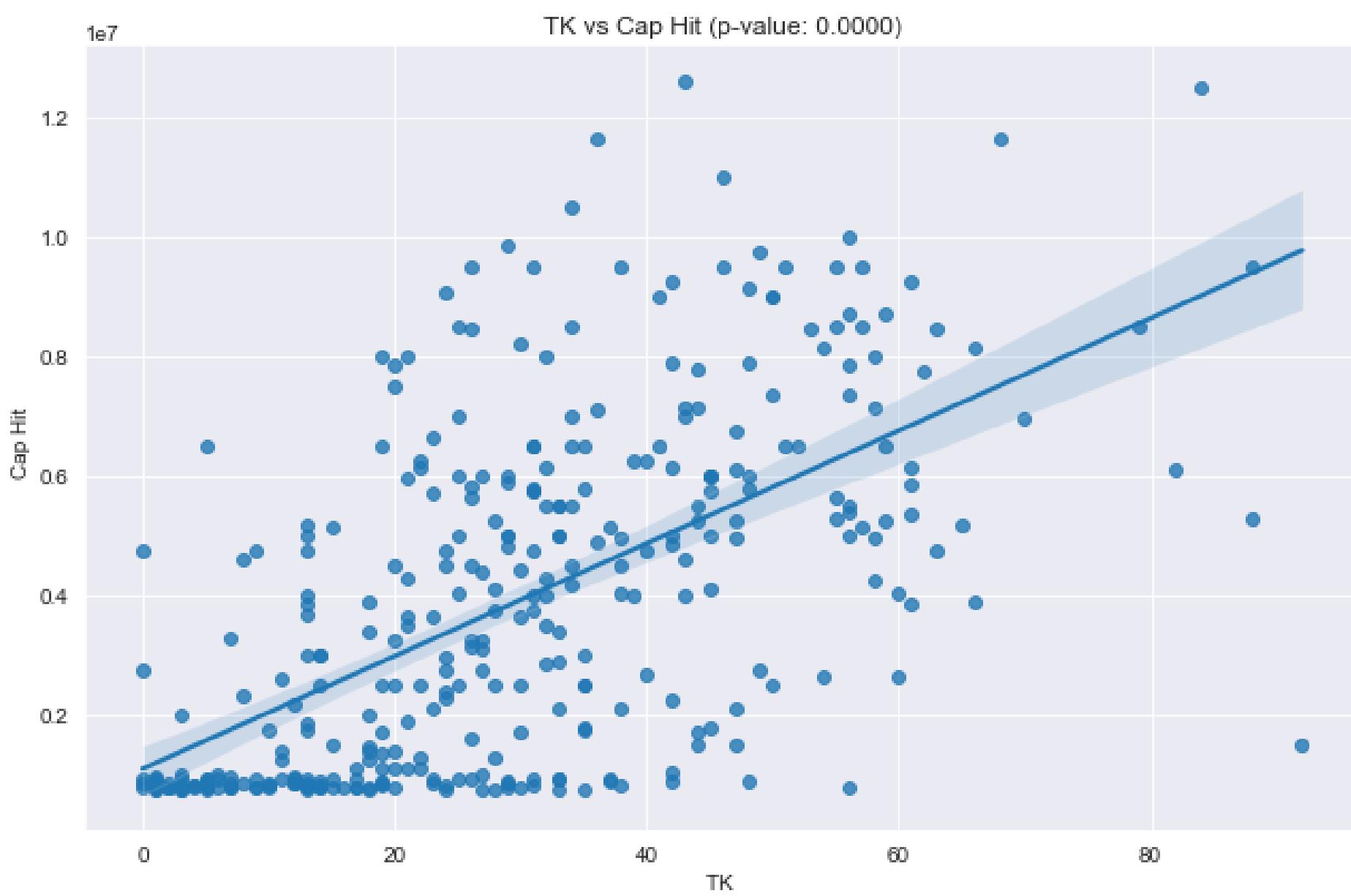
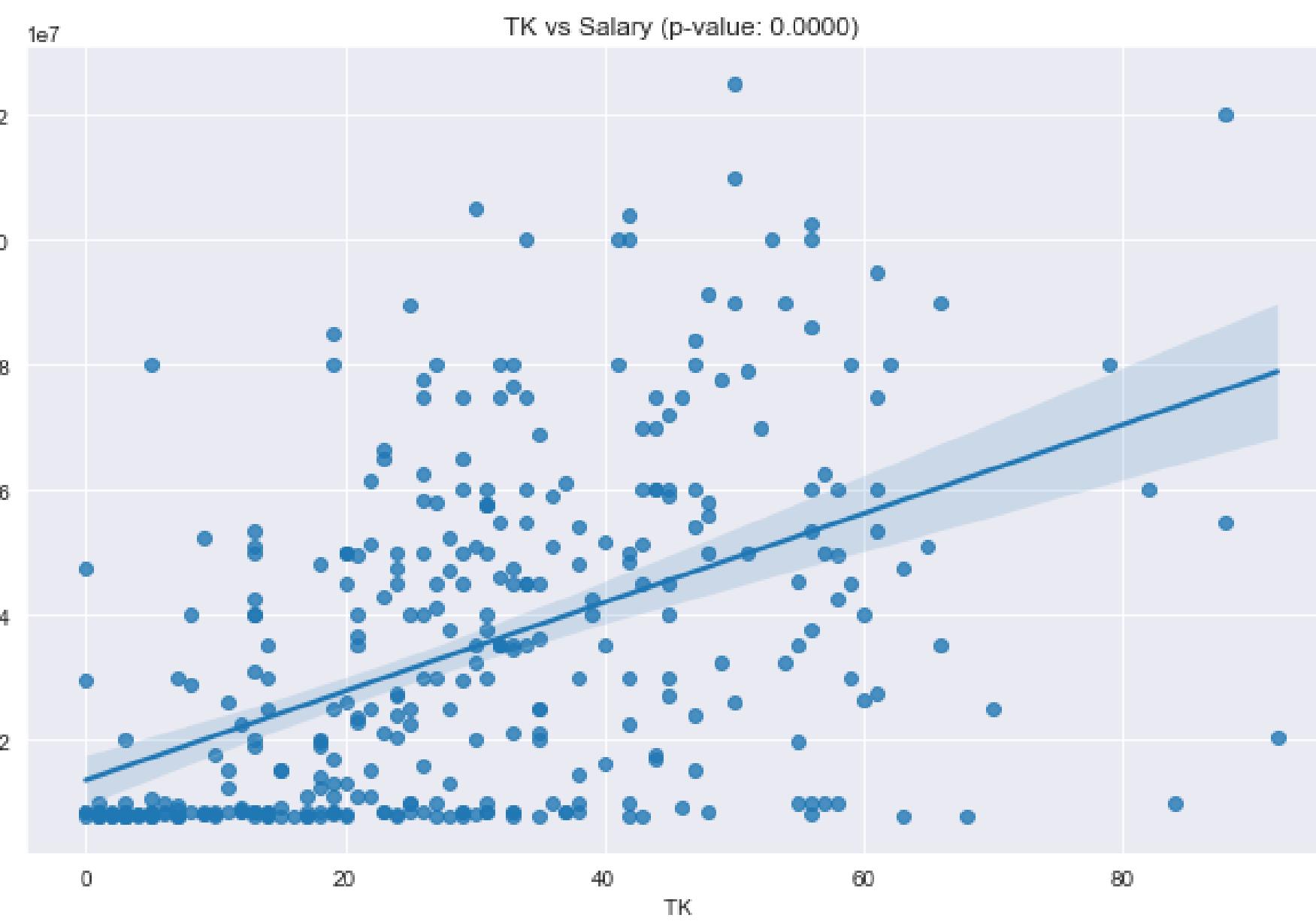
MS(射門偏出)



GV(失誤導致轉移球權)



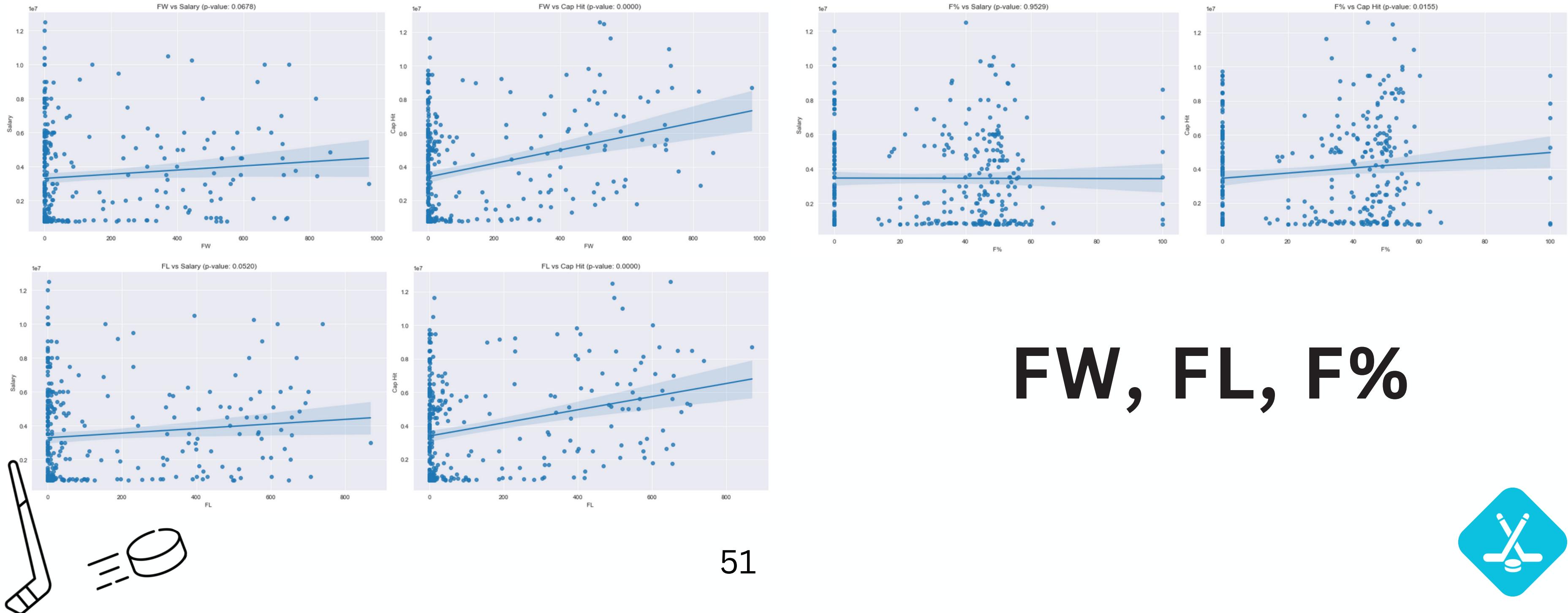
TK(抄截)



50

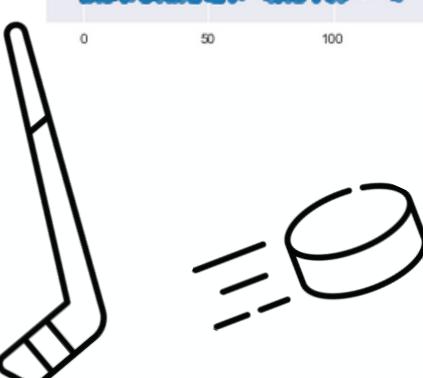
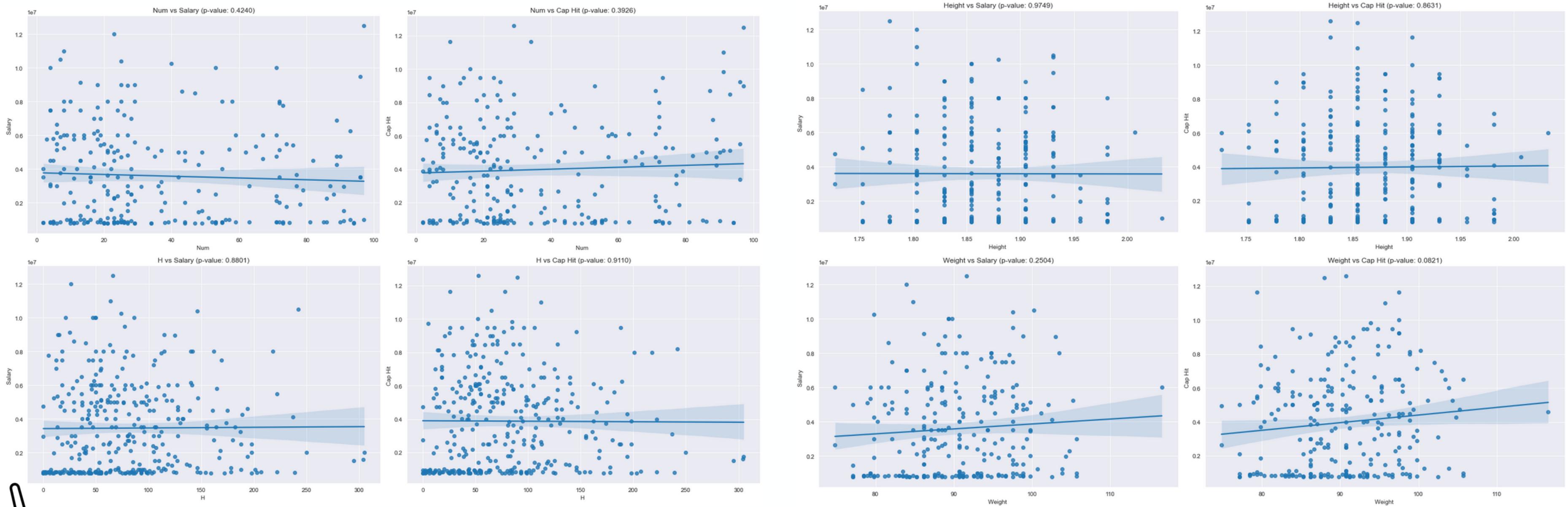


只在 Cap Hit 相關的

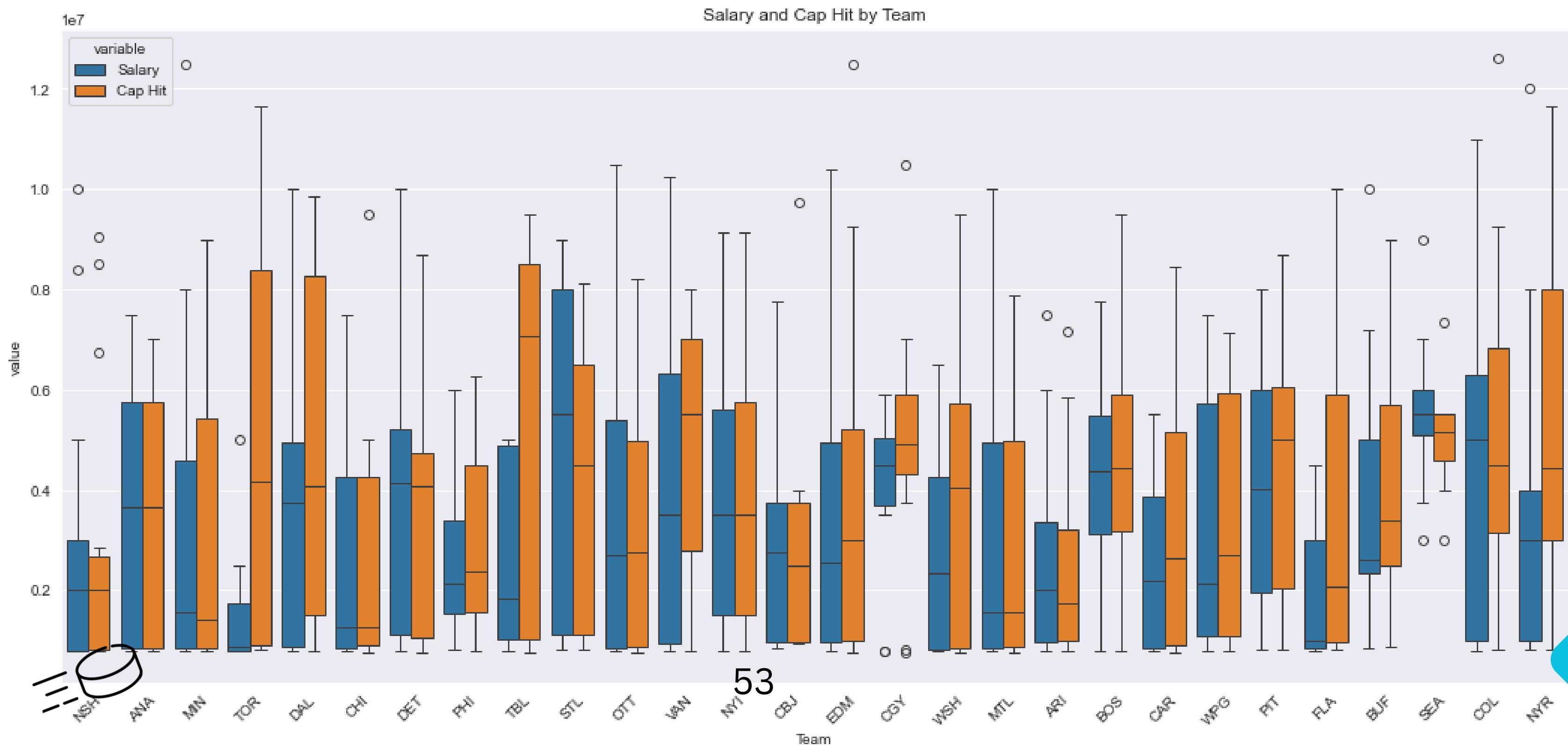


其餘不太相關的

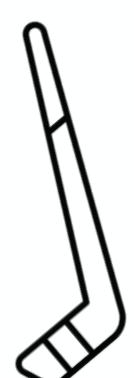
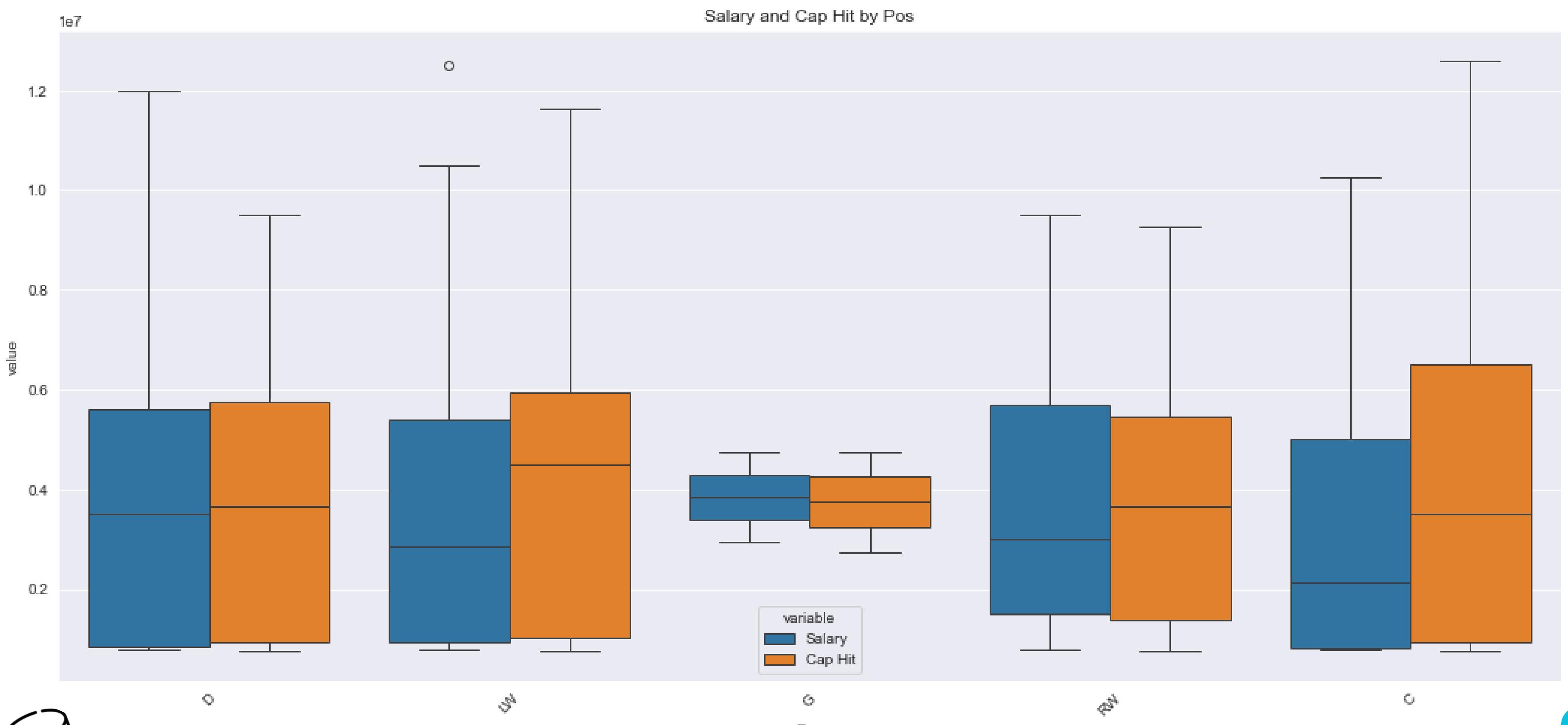
Num, H, Height, Weight



各隊差異

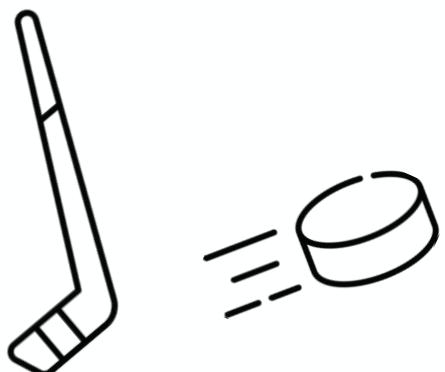


各站位差異



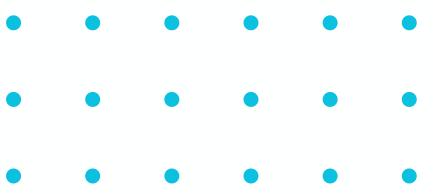
ANOVA

	Cap hit	Salary
不同定位	0.619872	0.701700
不同隊伍	0.270558	0.231114

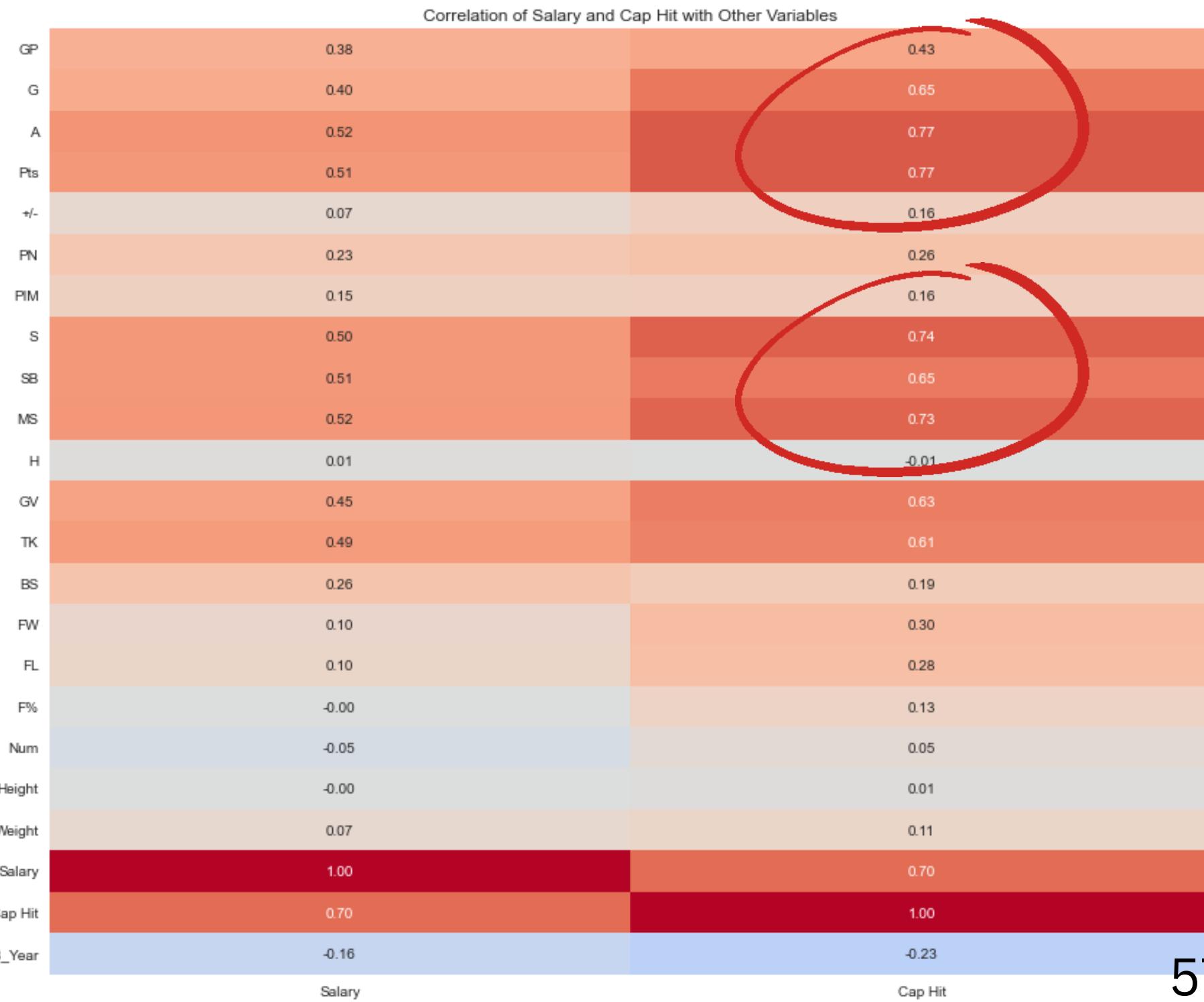




總結



CONCLUSION



57



S (射門數), MS (射偏), A (助攻), PTS (分數) 等相關性較高，說明比賽參與活躍相關。此外，可以發現，其他變數對 **Cap hit** (右方) 的 **Correlation** 比 **Salary** 的 **Correlation** 還要高，因為 **Cap hit** 包含 **Signing Bonuses**，較貼近總薪資。



8

NEXT STEP

1. 建模預測球員薪水
2. 擴增資料集

