퀴즈

1. 1루. 타자가 타율을 1푼 올리면 연봉이 얼마나 올라가나?
2. 지명타자가 타율을 1푼 올리면 연봉이 얼마나 올라가나?
3. 1루, 지명타자 중 어느 포지션이 타율이 연봉을 잘 표현하는가?
4. team 승수. 팀명. salary 연봉. 팀명 2016년 groupby(팀명) 30개 팀 평균 데이터를 구하시오.

```
import pandas as pd
df = pd.read_csv('http://wolfpack.hnu.ac.kr/Stat_Notes/example_data/baseball.csv')
```

```
ct = pd.DataFrame(df.Position.value_counts())
ct.reset_index(inplace=True)
ct.columns=['Position', 'count']
df0 = pd.merge(df,ct,on='Position',how='inner')
#df1 = df0[df0['count']>=16]
df1 = df0[df0.Salary!='.']
df1['Salary']=pd.to_numeric(df1.Salary)
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:7: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_guide/i
  import sys
```

```
df2=df1[df1.Position == '1B']
```

```
hit = df2['Hits']/df2['TimesatBat']
```

```
df2["hit"] = hit
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_guide/i
  """Entry point for launching an IPython kernel.
```

1. 1루 타자가 타율을 1푼 올리면 연봉이 얼마나 올라가나?

```
import statsmodels.api as sm
y=df2['Salary']
X=sm.add_constant(df2['hit'])
```

```
model=sm.OLS(y, X).fit()
model.summary()
```

```
/usr/local/lib/python3.6/dist-packages/numpy/core/fromnumeric.py:2495: FutureWarning: Method .
  return ptp(axis=axis, out=out, **kwargs)
```

OLS Regression Results

Dep. Variable:	Salary	**R-squared:**	0.201
Model:	OLS	**Adj. R-squared:**	0.164
Method:	Least Squares	**F-statistic:**	5.526
Date:	Thu, 21 Nov 2019	**Prob (F-statistic):**	0.0281
Time:	05:41:05	**Log-Likelihood:**	-187.18
No. Observations:	24	**AIC:**	378.4
Df Residuals:	22	**BIC:**	380.7
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1707.9789	1068.687	-1.598	0.124	-3924.299	508.342
hit	9228.0140	3925.714	2.351	0.028	1086.582	1.74e+04

Omnibus:	1.431	**Durbin-Watson:**	1.504
Prob(Omnibus):	0.489	**Jarque-Bera (JB):**	1.198
Skew:	0.368	**Prob(JB):**	0.549
Kurtosis:	2.190	**Cond. No.**	33.5

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

타율을 1푼 올리면 연봉이 92.280140천불만큼 올라간다.

2. 지명타자가 타율을 1푼 올리면 연봉이 얼마나 올라가나?

```
df3=df1[df1.Position == '3B']
```

```
hit0 = df3['Hits']/df3['TimesatBat']
df3["hit"] = hit0
```

```
/usr/local/lib/python3.6/dist-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/user_guide/i
```

```
import statsmodels.api as sm
y=df3['Salary']
X=sm.add_constant(df3['hit'])
model=sm.OLS(y, X).fit()
model.summary()
```

```
/usr/local/lib/python3.6/dist-packages/numpy/core/fromnumeric.py:2495: FutureWarning: Method .
  return ptp(axis=axis, out=out, **kwargs)
```

OLS Regression Results

Dep. Variable:	Salary	**R-squared:**	0.266
Model:	OLS	**Adj. R-squared:**	0.239
Method:	Least Squares	**F-statistic:**	10.13
Date:	Thu, 21 Nov 2019	**Prob (F-statistic):**	0.00355
Time:	05:42:59	**Log-Likelihood:**	-225.04
No. Observations:	30	**AIC:**	454.1
Df Residuals:	28	**BIC:**	456.9
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	-1940.7085	808.102	-2.402	0.023	-3596.030	-285.387
hit	9411.9454	2956.951	3.183	0.004	3354.906	1.55e+04

Omnibus:	6.537	**Durbin-Watson:**	2.376
Prob(Omnibus):	0.038	**Jarque-Bera (JB):**	4.770
Skew:	0.818	**Prob(JB):**	0.0921
Kurtosis:	4.069	**Cond. No.**	38.4

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

타율을 1푼 올리면 연봉이 94.119454천불만큼 올라간다.

3. 1루, 3루 중 어느 포지션이 타율이 연봉을 잘 표현하는가?

```
df0=pd.merge(df2,df3,how='outer')
import numpy as np
df0 = df0.replace('.',np.nan)
```

```
import statsmodels.api as sm
y=df2['hit']
X=sm.add_constant(df2['Salary'])
model=sm.OLS(y, X).fit()
model.summary()
```

```
/usr/local/lib/python3.6/dist-packages/numpy/core/fromnumeric.py:2495: FutureWarning: Method .
  return ptp(axis=axis, out=out, **kwargs)
```

OLS Regression Results

Dep. Variable:	hit	**R-squared:**	0.201
Model:	OLS	**Adj. R-squared:**	0.164
Method:	Least Squares	**F-statistic:**	5.526
Date:	Thu, 21 Nov 2019	**Prob (F-statistic):**	0.0281
Time:	05:48:49	**Log-Likelihood:**	51.204
No. Observations:	24	**AIC:**	-98.41
Df Residuals:	22	**BIC:**	-96.05
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.2532	0.010	26.644	0.000	0.234	0.273
Salary	2.175e-05	9.25e-06	2.351	0.028	2.56e-06	4.09e-05

Omnibus:	0.553	**Durbin-Watson:**	2.584
Prob(Omnibus):	0.759	**Jarque-Bera (JB):**	0.654
Skew:	0.253	**Prob(JB):**	0.721
Kurtosis:	2.369	**Cond. No.**	1.60e+03

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.6e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

```
1 import statsmodels.api as sm
2 y=df3['hit']
3 X=sm.add_constant(df3['Salary'])
4 model=sm.OLS(y, X).fit()
5 model.summary()
```

```
/usr/local/lib/python3.6/dist-packages/numpy/core/fromnumeric.py:2495: FutureWarning: Method .
  return ptp(axis=axis, out=out, **kwargs)
```

OLS Regression Results

Dep. Variable:	hit	**R-squared:**	0.266
Model:	OLS	**Adj. R-squared:**	0.239
Method:	Least Squares	**F-statistic:**	10.13
Date:	Thu, 21 Nov 2019	**Prob (F-statistic):**	0.00355
Time:	05:49:04	**Log-Likelihood:**	69.329
No. Observations:	30	**AIC:**	-134.7
Df Residuals:	28	**BIC:**	-131.9
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.2544	0.007	35.765	0.000	0.240	0.269
Salary	2.823e-05	8.87e-06	3.183	0.004	1.01e-05	4.64e-05

Omnibus:	3.187	**Durbin-Watson:**	2.282
Prob(Omnibus):	0.203	**Jarque-Bera (JB):**	1.998
Skew:	0.610	**Prob(JB):**	0.368
Kurtosis:	3.334	**Cond. No.**	1.26e+03

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.26e+03. This might indicate that there are
strong multicollinearity or other numerical problems.

1루는 20%

3루는 26%

영향을 받으므로 3루이다

1루는 20%

3루는 26%

영향을 받으므로 3루이다

교수님 사랑합니다.♥