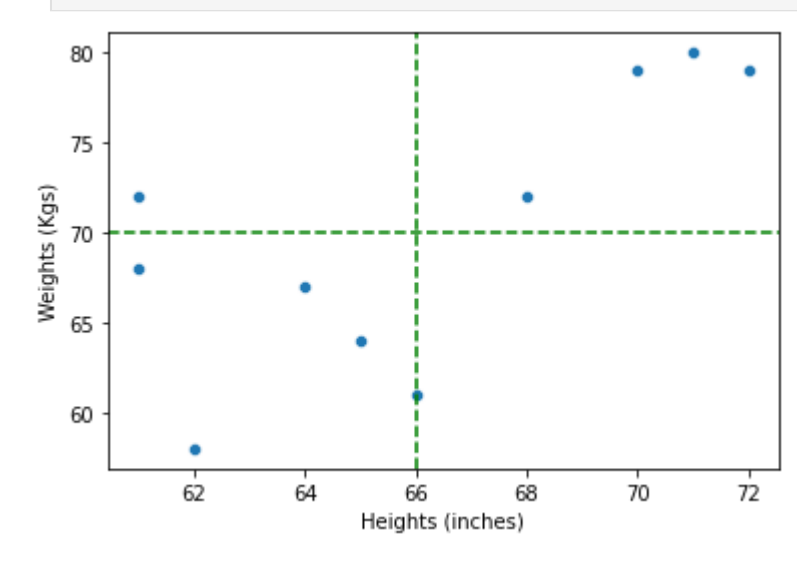
In [46]:
```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats import pearsonr, spearmanr
```

In [29]:
```python
# Lalit
heights = np.array([68, 62, 64, 61, 70, 66, 61, 65, 71, 72]) # inches
weights = np.array([72, 58, 67, 72, 79, 61, 68, 64, 80, 79]) # kgs

sns.scatterplot(x=heights, y=weights)
plt.axvline(heights.mean(),color="green",linestyle="--")
plt.axhline(weights.mean(),color="green",linestyle="--")

plt.xlabel("Heights (inches)")
plt.ylabel("Weights (Kgs)")
plt.show()
```



In [30]:
```python
covariance_lalit=((heights-heights.mean())*(weights-weights.mean())).sum()/len(heights)
covariance_lalit
```

Out[30]: 20.4

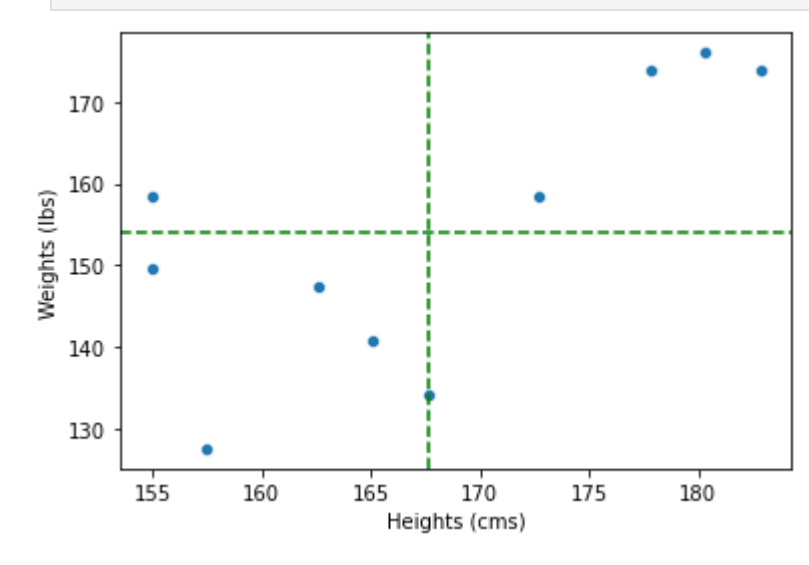
In [38]:
```python
correlation_lalit= covariance_lalit/(heights.std()*weights.std())
correlation_lalit
```

Out[38]: 0.7094289771951879

In [31]:
```python
# Nakul
heights1 = np.array([68, 62, 64, 61, 70, 66, 61, 65, 71, 72])*2.54 # cms
weights1 = np.array([72, 58, 67, 72, 79, 61, 68, 64, 80, 79])*2.2 # lbs

sns.scatterplot(x=heights1, y=weights1)
plt.axvline(heights1.mean(),color="green",linestyle="--")
plt.axhline(weights1.mean(),color="green",linestyle="--")

plt.xlabel("Heights (cms)")
plt.ylabel("Weights (lbs)")
plt.show()
```



In [32]:
```python
covariance_nakul=((heights1-heights1.mean())*(weights1-weights1.mean())).sum()/len(heights1)
covariance_nakul
```

Out[32]: 113.9952

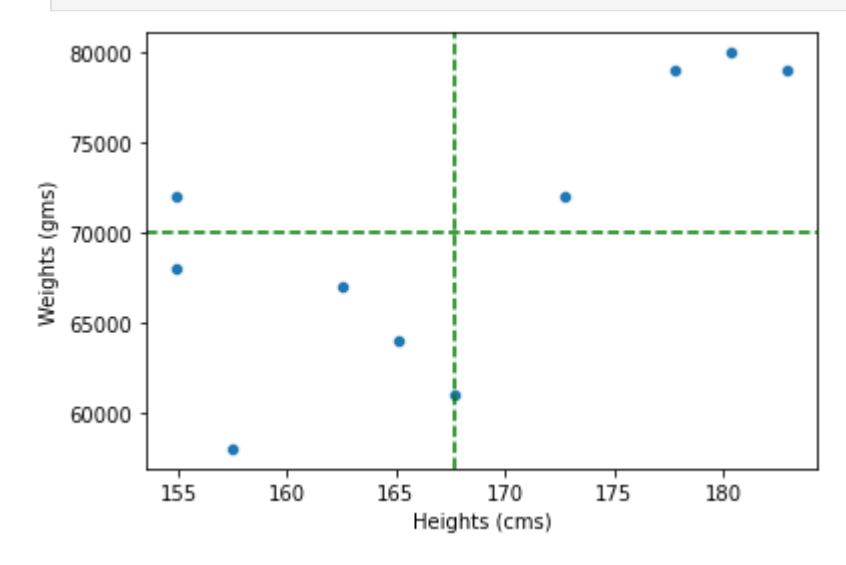
In [40]:
```python
correlation_nakul= covariance_nakul/(heights1.std()*weights1.std())
correlation_nakul
```

Out[40]: 0.7094289771951878

In [36]:
```python
# Darsh
heights2 = np.array([68, 62, 64, 61, 70, 66, 61, 65, 71, 72])*2.54 # cms
weights2 = np.array([72, 58, 67, 72, 79, 61, 68, 64, 80, 79])*1000 # gms

sns.scatterplot(x=heights2, y=weights2)
plt.axvline(heights2.mean(),color="green",linestyle="--")
plt.axhline(weights2.mean(),color="green",linestyle="--")

plt.xlabel("Heights (cms)")
plt.ylabel("Weights (gms)")
plt.show()
```



In [37]:
```python
covariance_darsh=((heights2-heights2.mean())*(weights2-weights2.mean())).sum()/len(heights2)
covariance_darsh
```

Out[37]: 51816.00000000001

In [41]:
```python
correlation_darsh= covariance_darsh/(heights2.std()*weights2.std())
correlation_darsh
```

Out[41]: 0.7094289771951878

In [43]:
```python
np.corrcoef(weights,heights)[0,1]
```

Out[43]: 0.7094289771951878

In [47]:
```python
pearsonr(weights,heights)
```

Out[47]: (0.7094289771951878, 0.021575164034828447)

In [52]:
```python
spearmanr(weights,heights)
```

Out[52]: SpearmanrResult(correlation=0.6024493002399199, pvalue=0.06529313205986141)

In [50]:
```python
df=pd.DataFrame([heights,weights]).transpose()
df.rename(columns={0:"Heights",1:"Weights"},inplace=True)
df
```

Out[50]:

	Heights	Weights
0	68	72
1	62	58
2	64	67
3	61	72
4	70	79
5	66	61
6	61	68
7	65	64
8	71	80
9	72	79

In [56]:
```python
df_rank=df.rank()
df_rank.rename(columns={"Heights":"Heights_rank","Weights":"Weights_rank"},inplace=True)
df_rank
```

Out[56]:

	Heights_rank	Weights_rank
0	7.0	6.5
1	3.0	1.0
2	4.0	4.0
3	1.5	6.5
4	8.0	8.5
5	6.0	2.0
6	1.5	5.0
7	5.0	3.0
8	9.0	10.0
9	10.0	8.5

In [57]:
```python
pearsonr(df_rank["Heights_rank"],df_rank["Weights_rank"])
```

Out[57]: (0.6024493002399198, 0.06529313205986152)

In [58]:
```python
df.corr()
```

Out[58]:

	Heights	Weights
Heights	1.000000	0.709429
Weights	0.709429	1.000000

In [59]:
```python
df.corr(method="spearman")
```

Out[59]:

	Heights	Weights
Heights	1.000000	0.602449
Weights	0.602449	1.000000

In [60]:
```python
!ls
```

```
08_Correlation_Tests_Notebook.ipynb problem_solving.csv
Sachin_ODI.csv                      waiting_time.csv
aerofit.csv                         weight-height.csv
iq_two_schools.csv
```

In [62]:
```python
df=pd.read_csv("Sachin_ODI.csv")
df
```

Out[62]:

	runs	NotOut	mins	bf	fours	sixes	sr	Inns	Opp	Ground	Date	Winner	Won	century
0	13	0	30	15	3	0	86.66	1	New Zealand	Napier	1995-02-16	New Zealand	False	False
1	37	0	75	51	3	1	72.54	2	South Africa	Hamilton	1995-02-18	South Africa	False	False
2	47	0	65	40	7	0	117.50	2	Australia	Dunedin	1995-02-22	India	True	False
3	48	0	37	30	9	1	160.00	2	Bangladesh	Sharjah	1995-04-05	India	True	False
4	4	0	13	9	1	0	44.44	2	Pakistan	Sharjah	1995-04-07	Pakistan	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
355	14	0	34	15	2	0	93.33	2	Australia	Sydney	2012-02-26	Australia	False	False
356	39	0	45	30	5	0	130.00	2	Sri Lanka	Hobart	2012-02-28	India	True	False
357	6	0	25	19	1	0	31.57	1	Sri Lanka	Dhaka	2012-03-13	India	True	False
358	114	0	205	147	12	1	77.55	1	Bangladesh	Dhaka	2012-03-16	Bangladesh	False	True
359	52	0	93	48	5	1	108.33	2	Pakistan	Dhaka	2012-03-18	India	True	False

360 rows × 14 columns

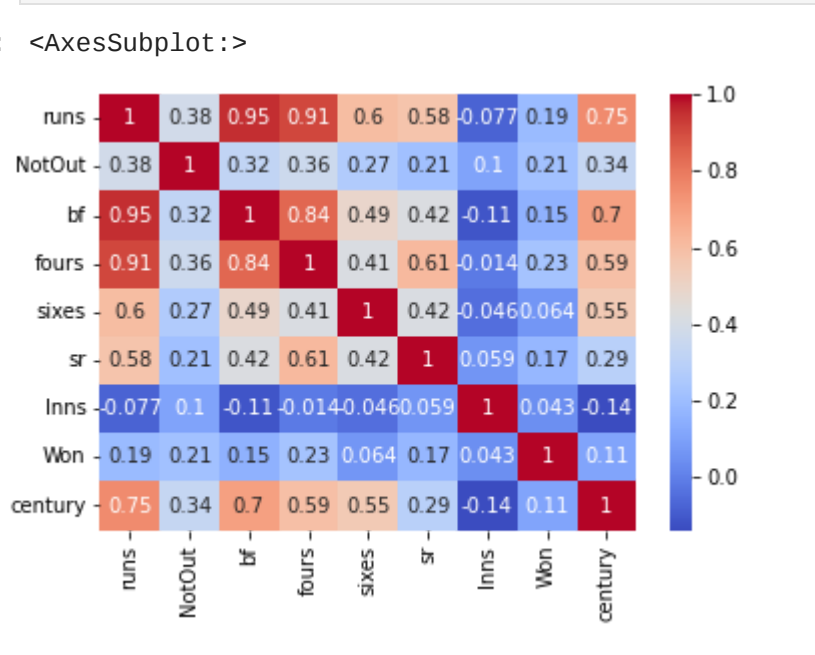
In [63]:
```python
df.corr()
```

Out[63]:

	runs	NotOut	bf	fours	sixes	sr	Inns	Won	century
runs	1.000000	0.377869	0.953559	0.907561	0.604070	0.576354	-0.076998	0.180319	0.751176
NotOut	0.377869	1.000000	0.324253	0.360667	0.260437	0.210590	0.104015	0.214461	0.336174
bf	0.953559	0.324253	1.000000	0.841240	0.467939	0.419307	-0.114803	0.153801	0.696394
fours	0.907561	0.360667	0.841240	1.000000	0.414766	0.615875	-0.014255	0.225120	0.591924
sixes	0.604070	0.260437	0.467939	0.414766	1.000000	0.415239	-0.045657	0.063837	0.547777
sr	0.576354	0.210590	0.419307	0.615875	0.415239	1.000000	0.058539	0.174703	0.292648
Inns	-0.076998	0.104015	-0.114803	-0.014255	-0.045657	0.058539	1.000000	0.043207	-0.137966
Won	0.180319	0.214461	0.153801	0.225120	0.063837	0.174703	0.043207	1.000000	0.108010
century	0.751176	0.336174	0.696394	0.591924	0.547777	0.292648	-0.137966	0.108010	1.000000

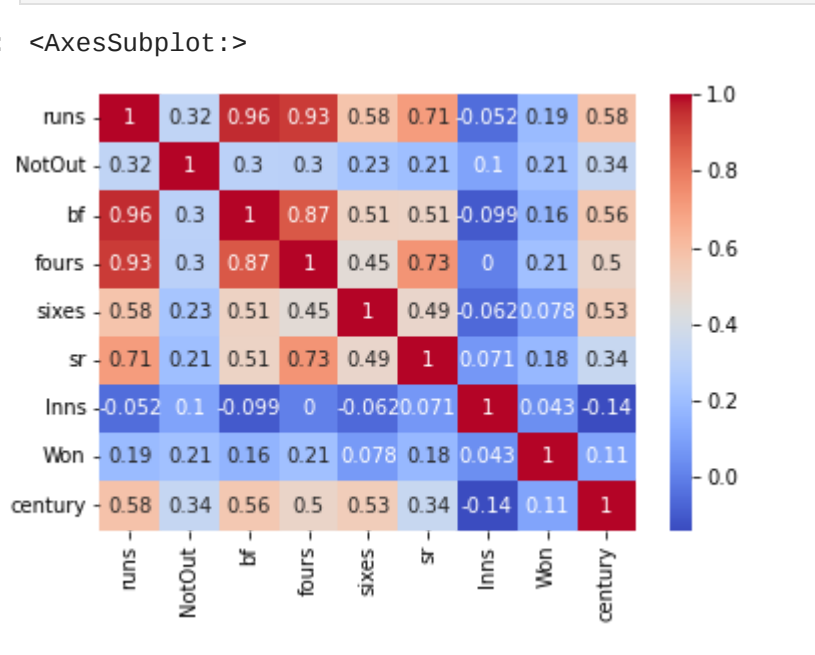
In [64]:
```python
sns.heatmap(df.corr(),annot=True,cmap="coolwarm")
```

Out[64]: <AxesSubplot:>



In [65]:
```python
sns.heatmap(df.corr(method="spearman"),annot=True,cmap="coolwarm")
```

Out[65]: <AxesSubplot:>



In [66]:
```python
df
```

Out[66]:

	runs	NotOut	mins	bf	fours	sixes	sr	Inns	Opp	Ground	Date	Winner	Won	century
0	13	0	30	15	3	0	86.66	1	New Zealand	Napier	1995-02-16	New Zealand	False	False
1	37	0	75	51	3	1	72.54	2	South Africa	Hamilton	1995-02-18	South Africa	False	False
2	47	0	65	40	7	0	117.50	2	Australia	Dunedin	1995-02-22	India	True	False
3	48	0	37	30	9	1	160.00	2	Bangladesh	Sharjah	1995-04-05	India	True	False
4	4	0	13	9	1	0	44.44	2	Pakistan	Sharjah	1995-04-07	Pakistan	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
355	14	0	34	15	2	0	93.33	2	Australia	Sydney	2012-02-26	Australia	False	False
356	39	0	45	30	5	0	130.00	2	Sri Lanka	Hobart	2012-02-28	India	True	False
357	6	0	25	19	1	0	31.57	1	Sri Lanka	Dhaka	2012-03-13	India	True	False
358	114	0	205	147	12	1	77.55	1	Bangladesh	Dhaka	2012-03-16	Bangladesh	False	True
359	52	0	93	48	5	1	108.33	2	Pakistan	Dhaka	2012-03-18	India	True	False

360 rows × 14 columns

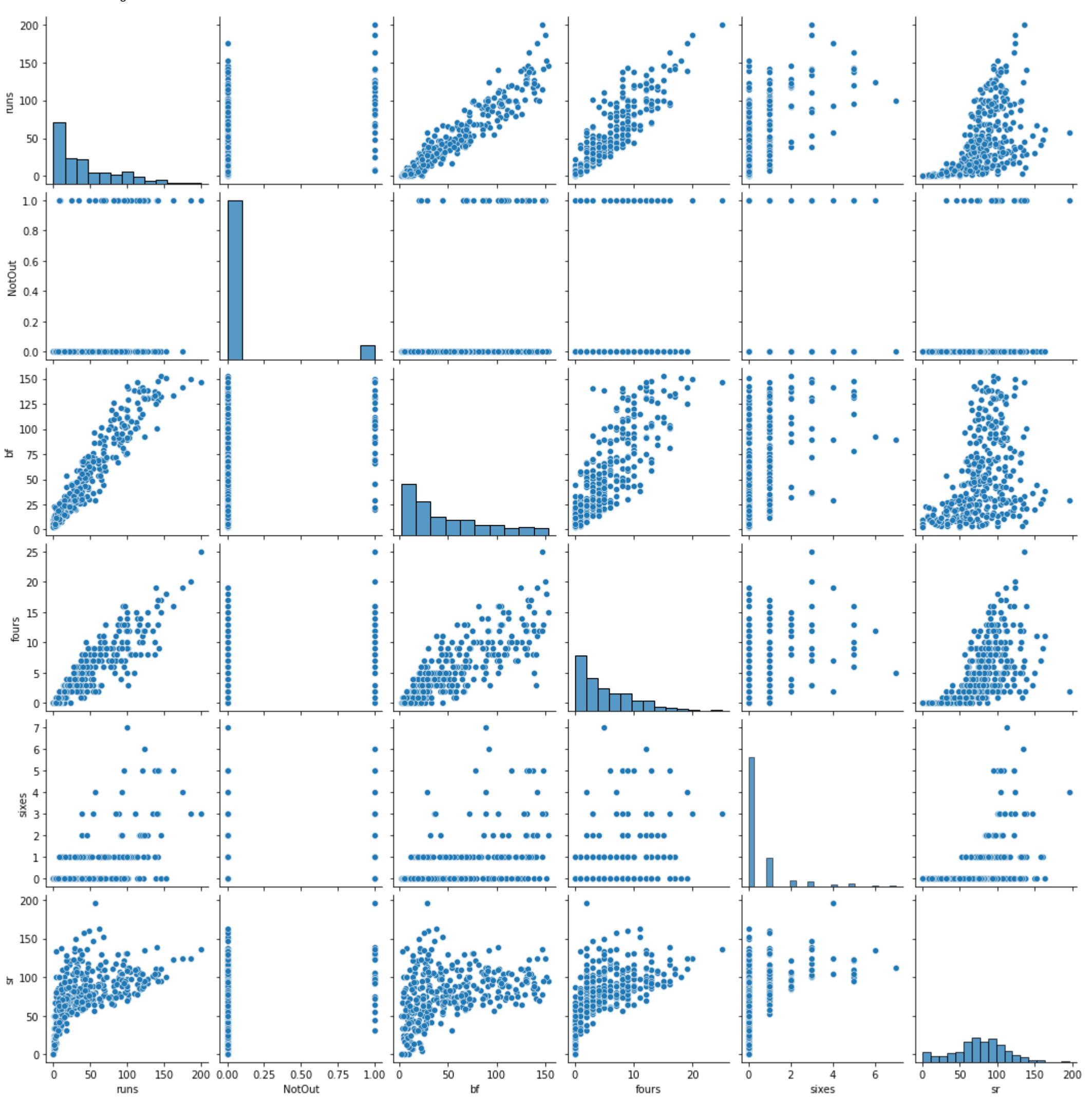
In [68]:
```python
df.columns
```

Out[68]:

```
Index(['runs', 'NotOut', 'mins', 'bf', 'fours', 'sixes', 'sr', 'Inns', 'Opp',
       'Ground', 'Date', 'Winner', 'Won', 'century'],
      dtype='object')
```

In [70]:
```python
sns.pairplot(df[['runs', 'NotOut', 'mins', 'bf', 'fours', 'sixes', 'sr']])
```

Out[70]: <seaborn.axisgrid.PairGrid at 0x7fe3bbdda490>



In [71]:
```python
spearmanr(df["sr"],df["fours"])
```

Out[71]: SpearmanrResult(correlation=0.7325696280969253, pvalue=9.01929054275424e-62)

In [72]:
```python
pearsonr(df["sr"],df["fours"])
```

Out[72]: (0.6159732925596058, 1.1277291712772218e-38)

In [ ]:

In [ ]: