In [33]:
```
import numpy as np
import pandas as pd
pd.set_option('display.float_format', lambda x: '%.5f' % x)
import matplotlib.pyplot as plt
import seaborn as sns
plt.style.use('fivethirtyeight')
import warnings
from scipy import stats
from scipy.stats import skew, kurtosis
warnings.filterwarnings('ignore')
%matplotlib inline

if os.name == 'posix': # Mac 환경 폰트 설정
    plt.rc('font', family='AppleGothic')
elif os.name == 'nt': # Windows 환경 폰트 설정
    plt.rc('font', family='Malgun Gothic')

plt.rc('axes', unicode_minus=False) # 마이너스 폰트 설정

# 글씨 선명하게 출력하는 설정
%config InlineBackend.figure_format = 'retina'
```

In [34]:
```
def z_score_method(df, variable_name):
    columns = df.columns
    z = np.abs(stats.zscore(df))
    threshold = 3
    outlier = []
    index=0
    for item in range(len(columns)):
        if columns[item] == variable_name:
            index = item
    print("index :",index)
    for i, v in enumerate(z[:, index]):
        if v > threshold:
            outlier.append(i)
        else:
            continue
    return outlier
```

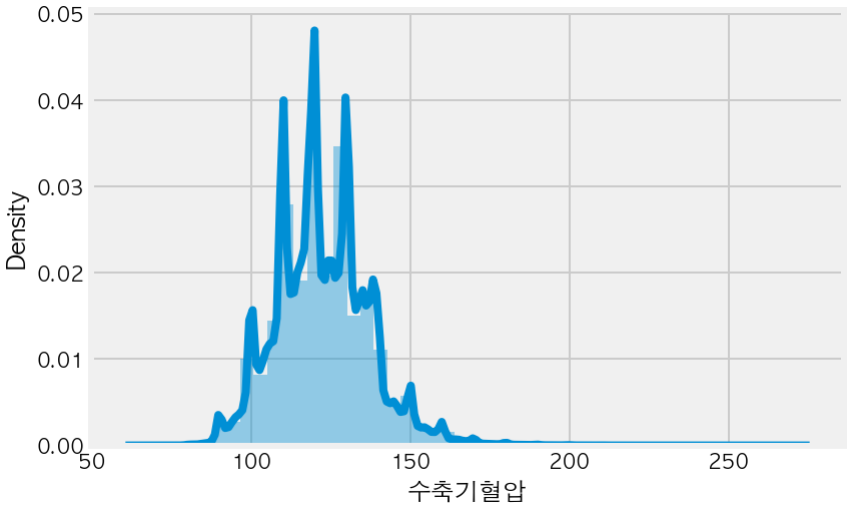
In [35]:
```
cleandata = pd.read_csv('cleandata.csv')
```

In [36]:
```
data_M = cleandata['수축기혈압'].copy()
data_N = cleandata['이완기혈압'].copy()
data_V = cleandata['혈청크레아티닌'].copy()
print("수축기혈압 왜도 :",skew(data_M))
print("이완기혈압 왜도 :",skew(data_N))
print("혈청크레아티닌 왜도 :",skew(data_V))
print("수축기혈압 첨도 :",kurtosis(data_M, fisher=True))
print("이완기혈압 첨도 :",kurtosis(data_N, fisher=True))
print("혈청크레아티닌 첨도 :",kurtosis(data_V, fisher=True))
```

```
수축기혈압 왜도 : 0.47695611875857297
이완기혈압 왜도 : 0.39537647951689325
혈청크레아티닌 왜도 : 110.17456302546702
수축기혈압 첨도 : 0.9638096142038237
이완기혈압 첨도 : 0.8654945497011899
혈청크레아티닌 첨도 : 18697.09874979088
```

In [44]:
```
sns.distplot(data_M)
```

Out[44]:
```
<AxesSubplot:xlabel='수축기혈압', ylabel='Density'>
```



In [37]:
```
for_M = cleandata[['수축기혈압','식전혈당(공복혈당)','당뇨여부']].copy()
for_N = cleandata[['이완기혈압','식전혈당(공복혈당)','당뇨여부']].copy()
for_V = cleandata[['혈청크레아티닌','식전혈당(공복혈당)','당뇨여부']].copy()
```

In [38]:
```
z_outlier = z_score_method(for_M, '수축기혈압')
# sample = for_M.loc[z_outlier]
# sample = sample['수축기혈압'].copy()
# sample.sort_values(inplace=True)
# for i in sample:
#     print(i, end=" ")
# 수축기혈압 Lower_bound : 78
# 수축기혈압 Upper_bound : 167
# 총 개수 : 12425
for_M.drop(z_outlier, inplace=True)
for_M.reset_index(drop=True, inplace=True)
for_M.corr(method="kendall")
```

```
index : 0
```

Out[38]:

	수축기혈압	식전혈당(공복혈당)	당뇨여부
수축기혈압	1.00000	0.16333	0.09301
식전혈당(공복혈당)	0.16333	1.00000	0.34051
당뇨여부	0.09301	0.34051	1.00000

In [39]:
```
z_outlier = z_score_method(for_N, '이완기혈압')
# sample = for_N.loc[z_outlier]
# sample = sample['이완기혈압'].copy()
# sample.sort_values(inplace=True)
# for i in sample:
#     print(i, end=" ")
# 수축기혈압 Lower_bound : 46
# 수축기혈압 Upper_bound : 105
# 총 개수 : 9347
for_N.drop(z_outlier, inplace=True)
for_N.reset_index(drop=True, inplace=True)
for_N.corr(method="kendall")
```

```
index : 0
```

Out[39]:

	이완기혈압	식전혈당(공복혈당)	당뇨여부
이완기혈압	1.00000	0.12830	0.05649
식전혈당(공복혈당)	0.12830	1.00000	0.34119
당뇨여부	0.05649	0.34119	1.00000

In [40]:
```
z_outlier = z_score_method(for_V, '혈청크레아티닌')
# print(len(z_outlier))
# sample = for_V.loc[z_outlier]
# sample = sample['혈청크레아티닌'].copy()
# sample.sort_values(inplace=True)
# for i in sample:
#     print(i, end=" ")
# 수축기혈압 Lower_bound : 0
# 수축기혈압 Upper_bound : 2.4
# 총 개수 : 2664
for_V.drop(z_outlier, inplace=True)
for_V.reset_index(drop=True, inplace=True)
for_V.corr(method="kendall")
```

```
index : 0
```

Out[40]:

	혈청크레아티닌	식전혈당(공복혈당)	당뇨여부
혈청크레아티닌	1.00000	0.09531	0.03916
식전혈당(공복혈당)	0.09531	1.00000	0.34158
당뇨여부	0.03916	0.34158	1.00000

In [41]:
```
stats.pointbiserialr(for_M['수축기혈압'], for_M['당뇨여부'])
```

Out[41]:
```
PointbiserialrResult(correlation=0.11367785030396452, pvalue=0.0)
```

In [42]:
```
stats.pointbiserialr(for_N['이완기혈압'], for_N['당뇨여부'])
```

Out[42]:
```
PointbiserialrResult(correlation=0.06780320762302249, pvalue=0.0)
```

In [43]:
```
stats.pointbiserialr(for_V['혈청크레아티닌'], for_V['당뇨여부'])
```

Out[43]:
```
PointbiserialrResult(correlation=0.052130688513511436, pvalue=0.0)
```

In [45]:
```
for_M.corr()
```

Out[45]:

	수축기혈압	식전혈당(공복혈당)	당뇨여부
수축기혈압	1.00000	0.22049	0.11368
식전혈당(공복혈당)	0.22049	1.00000	0.69996
당뇨여부	0.11368	0.69996	1.00000

In [46]:
```
for_N.corr()
```

Out[46]:

	이완기혈압	식전혈당(공복혈당)	당뇨여부
이완기혈압	1.00000	0.16251	0.06780
식전혈당(공복혈당)	0.16251	1.00000	0.70066
당뇨여부	0.06780	0.70066	1.00000

In [47]:
```
for_V.corr()
```

Out[47]:

	혈청크레아티닌	식전혈당(공복혈당)	당뇨여부
혈청크레아티닌	1.00000	0.11708	0.05213

	혈청크레아티닌	식전혈당(공복혈당)	당뇨여부
식전혈당(공복혈당)	0.11708	1.00000	0.70066
당뇨여부	0.05213	0.70066	1.00000

In [ ]: