数据：[13,15,16,16,19,20,20,21,22,22,25,25,25,25,30,33,33,35,35,35,35,36,40,45,46,52,70]

数据已排序，首先计算Q1，Q3分位数，

data = [13,15,16,16,19,20,20,21,22,22,25,25,25,25,30,33,33,35,35,35,35,36,40,45,46,52,70]

data = np.array(data)

# 计算Q1,Q3 分位数

Q1 = np.percentile(data, 25)

Q3 = np.percentile(data, 75)

计算IQR，使用1.5倍IQR计算数据箱线上下限

# 求IQR

IQR = Q3 - Q1

#箱线下限

low\_limit = Q1 - 1.5\*IQR

#箱线上限

upper\_limit = Q3 + 1.5\*IQR

判断是否是离群点，只需要看数据是否超出上下限

直接判断是否超过箱线

outliers = list(data[data < low\_limit]) + list(data[data > upper\_limit])

print('直接法离群点:',outliers)

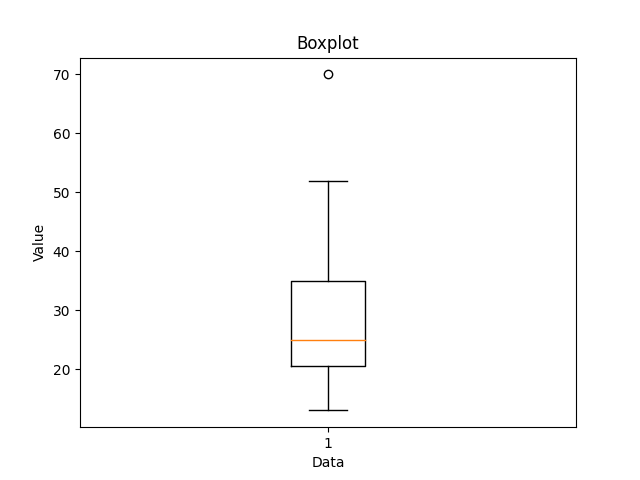
还可以使用统计方法，用z值法判断，计算标准分数z值。

zscore = stats.zscore(data)

outliers = data[np.abs(zscore) > 3]

print('Z值法离群点:',outliers)

最后画出箱线图直观展示数据状态。



检验数据是否符合正态分布，此时使用ks检验

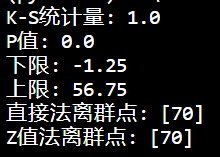
from scipy import stats

# 使用K-S检验

ks\_result = stats.kstest(data, 'norm')

print('K-S统计量:', ks\_result[0])

print('P值:', ks\_result[1])



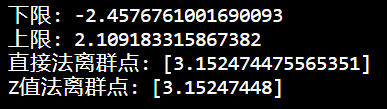
看来是完全不符合，使用standardscaler处理数据。

data = (data - data.mean()) / data.std()

再次进行ks检验



符合假设，是正态了，再次进行箱线图计算



看来还是离群点。

