

```

# -*- coding: utf-8 -*-
"""
Created on Fri Feb 22 11:29:51 2019
HW4 ME EN 2550
@author: Ryan Dalby
"""

import scipy.stats as stat
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sb
import numpy as np
import statsmodels.stats.api as sm

excelFilename = "hw4-spring2019.csv"
#extract data from csv file storing it in a dataframe
data = pd.read_csv(excelFilename)

#Problem B1
print("B-1:")
b1data = data['Question B-1']
b1datafiltered = b1data[b1data < 6000] #This removes outlier
plt.title("Shear Strengths of spot welds in a ti-alloy")
b1datafiltered.hist()
plt.show()
b195Percentile = b1datafiltered.quantile(q=.95)
print("The histogram appears to be almost normally distributed but overall does have left skew(Once
print("The 95th percentile of strength is {}".format(b195Percentile))

#Problem B2

def CIForSD(SD, n, CL):
    """
    Gets confidence interval for standard deviation
    """
    print(SD)
    alpha = 1-CL
    lower = SD * np.sqrt((n-1)/stat.distributions.chi2.ppf((1- alpha/2),df=(n-1)))
    upper = SD * np.sqrt((n-1)/stat.distributions.chi2.ppf((alpha/2),df=(n-1)))
    return(lower,upper)

print()
print("B-2:")
b2data = data['Question B-2'].dropna() #gets all actual values for B2
plt.title("Coefficient of restitution of baseballs")
sb.distplot(b2data, fit=stat.norm, kde = False, axlabel = False)
plt.figure()
stat.probplot(b2data, plot=plt.subplot(111))
plt.show()
print("a) Yes there is evidence that the coefficient of restitution is normally distributed since a
mean95ci = sm.DescrStatsW(b2data).zconfint_mean(alpha = .05) #gets 95% 2sided confidence interval f
print("b) A 95% confidence interval for the mean coefficient of restitution is {}".format(mean95ci)
sd99ci = CIForSD(b2data.std(), b2data.count(), .99)
print("c) A 99% confidence interval for the standard deviation of the coefficient of restitution is

```

```

#Problem B3
print()
print("B-3:")
b3data = data['Question B-3'].dropna() #gets all values for B3
(_,b3Pval) = sm.ztest(b3data, value = 130)
print("a) Using the sample data to conduct a two sided z-test for means against the null hypothesis
and since {0:.3} > .05 at the 95% confidence level we cannot reject the null hypothesis".format(b3Pval))
plt.title("Sodium content of boxes of cornflakes in mg")
sb.distplot(b3data, fit=stat.norm, kde = False, axlabel = False)
plt.figure()
stat.probplot(b3data, plot=plt.subplot(111))
plt.show()
print("b) We can see that the data is approximately normal by fitting a normal curve to the data at
b3power = sm.tt_solve_power(effect_size=((b3data.mean()-130.5)/b3data.std()), nobs=b3data.count(),
print("c) The power of the test if the true mean is 130.5 mg is {}".format(b3power))
b3numSamp = sm.tt_solve_power(effect_size=((b3data.mean()-130.1)/b3data.std()), power=.75, alpha =
print("d) The sample size to detect a true mean of 130.1 mg and have power of at least .75 is: {}")

```