Title: Statistical Analysis of building energy use intensity (EUI).

Description:

This study aims to compare the energy use intensity (EUI) between commercial and residential building among five cities in the United States. The dataset is collected from BPD website [1]. Building EUI data of five cities were collected and then we have processed the data. Missing data are excluded from the dataset and the data are factorized to nominal, ordinal, or interval/ratio from characterized forms. The dataset is created in **dataset.R** script and correct factorizations of the data were made. The final dataset is named as **building_data.csv**. The final dataset contains 37,937 observations and 7 variables. The description of the dataset is given in table 1.

Part 1: z- test between sample and population

Our observation is based on the building made after the year 1900. In total, 36,920 observations are used. Then dataset is tested for normality and homogeneity using Shapiro test and Levene test, respectively. To support the Shapiro test, dataset is down sampled to 4,999 observation and z-test are performed. The z-test (z = 0.46, p = 0.65) for the site_eui confirms that the sample (M = 76.11, SD = 64.11) represents the population (M = 75.71, SD = 60.98). All the analysis are made in analysis.R script.

N.B: Because of the random sampling of the data, the values may slightly differ than the one presented here in the doc file. Value should be checked after running the R script.

Table 1: Description of the dataset

Variable	Explanation	Mean	SD
city	Five different cities	-	-
building_class	Two different building class	-	-
floor_area	Building floor areas in sq.ft	197,414	306,052
year_built	The year in which building was built	1975	30
electric_eui	Electrical energy use intensity	37.92	43.77
fuel_eui	Fuel energy use intensity	38.19	44.45
site_eui	Building total energy use intensity	76.11	64.11

Part 2: Comparing the residential and commercial building EUI (site_eui) using independent two sample t-test.

Step 1: H_0 : mu1 = mu2. There is no mean difference between commercial and residential buildings in terms of energy use intensity (EUI).

H_1: mu1 ≠ mu2. There are mean differences between commercial and residential buildings in terms of energy use intensity (EUI).

Step 2: alpha = 0.05, df = 4,999 -2 = 4,997

Step 3: Calculation and Assumption

Assumptions: 1. Normality, 2. Independence, 3. HOV (equal variance)

Null hypothesis is that there are no departures from the given assumption.

Alternative hypothesis is that there are departures from the given assumption

Check for normality assumption: (Shapiro test)

For both commercial and residential buildings, $p < 0.001$.

Reject Null Hypothesis, the dataset is not normally distributed

Check for HOV: (levene test)

$p < 0.001$. Departure from HOV

Reject Null Hypothesis, the dataset does not have equal variance.

Calculations:

Calculations are done in R studio.

From independent sample t-test, $p < 0.001$

From cohens'd test, $d = 0.54$. Medium effect size.

Step 4:

p < 0.001. Reject H_0 : mu1 = mu2. There are mean difference between commercial and residential buildings in terms of energy use intensity (EUI).

Summary

Step 1	H_0 *: mu1 = mu2*	H_1*: mu1 ≠ mu2*
Step 2	alpha = 0.05	Welch Two Sample t-test, df = 4,672
Step 3	Independent sample t-test,Two tailed, Normality and HOV violated	t_{calc} = 13.53, $p < 0.001$
Step 3	$p < 0.001$	Reject: H_0 *: mu1 = mu2*

APA writeup:

We assessed differences in EUI between the two building types with an independent sample t-test. Prior to conducting the test, assumptions for normality and homogeneity of variance were assessed. The Shapiro Wilks test found that both building types deviate from a normal curve ($p < 0.001$) and Levene's test confirms that the dataset deviates homogeneity of variances (HOV) assumption, $p < 0.001$. As such, interpretations of results should be done with caution.

The result of our t-test, t (4,672) = 13.53, $p < 0.001$, $d = 0.54$, suggests that the commercial buildings (M = 85.06, SD = 75.52) had higher EUI than the residential buildings (M = 63.01, SD = 38.61) with a medium overall effect for EUI.

Part 3: Comparing the commercial building EUI of five cities using one way ANOVA test.

For this part, the dataset containing 36,920 observations is divided into commercial (22,351) and residential (14,548) dataset. We assessed the difference in commercial building EUI separately using ANOVA test.

Step 1: H0: mu(Atlanta) = mu(Austin) = mu(Chicago) = mu(Denver) = mu(Minneapolis). There is no mean difference among cities in terms of energy use intensity (EUI).

H_1: at least one mean (EUI of city) is different

Step 2: alpha = 0.05.

df_M (between groups) = 5-1 = 4, df_R (within groups) = 22,351-5 = 22,346.

Step 3: Calculation and Assumption

Assumptions: 1. Normality, 2. Independence, 3. HOV (equal variance)

Null hypothesis is that there are no departures from the given assumption.

Alternative hypothesis is that there are departures from the given assumption

We checked the assumptions of our dataset in Part 1 and skip in this section to avoid down sampling for Shapiro test. Our dataset deviates from the normality and HOV assumptions.

Calculations:

Calculations are done in R studio.

Variance	*SS*	*df*	*MS*	*F-ratio*	*p-value*
Between Group (M)	5,113,526	4	1,278,381	254	2e-16
Within Group (R)	112,286,805	22,346	5,025		
Total (T)	117,400,331	22,350			

From omega square test, $\omega^2 = 0.04$, small effect.

From post hoc test,

Combination	diff	lwr bound	upr bound	p adjusted
Austin-Atlanta	-16.34	-22.30	-10.38	0.0000
Chicago-Atlanta	25.80	20.29	31.31	0.0000
Denver-Atlanta	7.94	2.38	13.50	0.0009
Minneapolis-Atlanta	22.79	16.05	29.52	0.0000
Chicago-Austin	42.14	38.33	45.94	0.0000
Denver-Austin	24.28	20.40	28.16	0.0000
Minneapolis-Austin	39.12	33.69	44.56	0.0000
Denver-Chicago	-17.86	-21.00	-14.72	0.0000
Minneapolis-Chicago	-3.01	-7.95	1.92	0.4559
Minneapolis-Denver	14.85	9.86	19.84	0.0000

Step 4:

$p < 0.001$. Reject H_0: mu(Atlanta) = mu(Austin) = mu(Chicago) = mu(Denver) = mu(Minneapolis). There are mean differences among cities in terms of energy use intensity (EUI).

Step 1	H_0: mu(Atlanta) = mu(Austin) = mu(Chicago) = mu(Denver) = mu(Minneapolis).	H_1: at least one mean (EUI of city) is different
Step 2	alpha = 0.05	$df_M = 4$, $df_R = 22{,}346$
Step 3	One way ANOVA, normality and HOV violated	$F_{calc} = 254$, $p < 0.001$
Step 3	$p < 0.001$	Reject: H_0. Null hypothesis.

APA writeup:

A one-way ANOVA indicated that cities were significantly different with regards to building EUI; F (4, 22,346) = 254, $p < 0.001$, with an overall small effect, $\omega^2 = 0.04$. Post-Hoc procedures using Tukey adjustments revealed that all the cities varied significantly except Minneapolis (M = 96.60, SD = 52.81) and Chicago (M = 99.61, SD = 74.10). Homogeneity of Variance and normality assumptions could not be assumed for the dataset. As such, interpretations of results should be done with caution.

Reference

1. https://bpd.lbl.gov/explore