HW#2

FICO scores are a means by which banks evaluate whether customers applying for various types of loans – home mortgage, car purchase, etc – are credit worthy and can expect to have the loan advanced by the bank, repaid and on time. FICO scores range from 300 to 850 (worst to best). An individual with a FICO score of 800 or above has an exceptional credit history. FICO scores in the mid- to upper 700s are good scores. The most common scores sit somewhere between 650 and 750. A a score of 599 or lower is considered a poor credit score.

Attached find a copy of FICO scores in the US by State, from 2017-19. Decide whether to examine these data by state or region, whether the experimental unit chosen is generally credit worthy, and whether FICO scores change for that experimental unit over the three-year period.

Using the appropriate SAS programs, answer the following questions:

1. *Do FICO scores change over the years in this dataset?*
2. *Are some experimental units (states or regions) better credit risks than others?*
3. *Although other variables that might affect FICO scores are not available, what factors might have produced the results obtained in Q1 and Q2?*
4. *Based on your results, how would you advise banks that might be considering making loans in these states or regions?*

To perform the analysis, I created a new Variable – Region. For each state, I categorized it as belonging to one of the nine regions recognized by the census bureau: *Pacific, Mountain, West North Central, West South Central, East North Central, East South Central, New England, Middle Atlantic, South Atlantic.*

Answers

1. Yes, year on year there is a change in the FICO score. From the GLM procedure we see that the within subject effects of year is significant even after correction. Using the GLM Procedure we get the following output:

The GLM Procedure
Repeated Measures Analysis of Variance
Univariate Tests of Hypotheses for Within Subject Effects

| Source | DF | Type III SS | Mean Square | F Value | Pr > F | Adj Pr > F | |
						G - G	H-F-L
year	2	19408.68500	9704.34250	1659.77	<.0001	<.0001	<.0001
year*Region	16	133.00138	8.31259	1.42	0.1522	0.2068	0.2061
Error(year)	82	479.43862	5.84681				

2. Some regions are absolutely better credit risks than others. If we look at the boxplots for each of the years, we can see that the East South Central, West South Central, and South Atlantic Regions have generally lower credit scores than the New England & West North Central Regions which rank as the two best regions year after year.

Distribution of _2018

F	9.59
Prob > F	<.0001

Distribution of _2019

F	9.31
Prob > F	<.0001
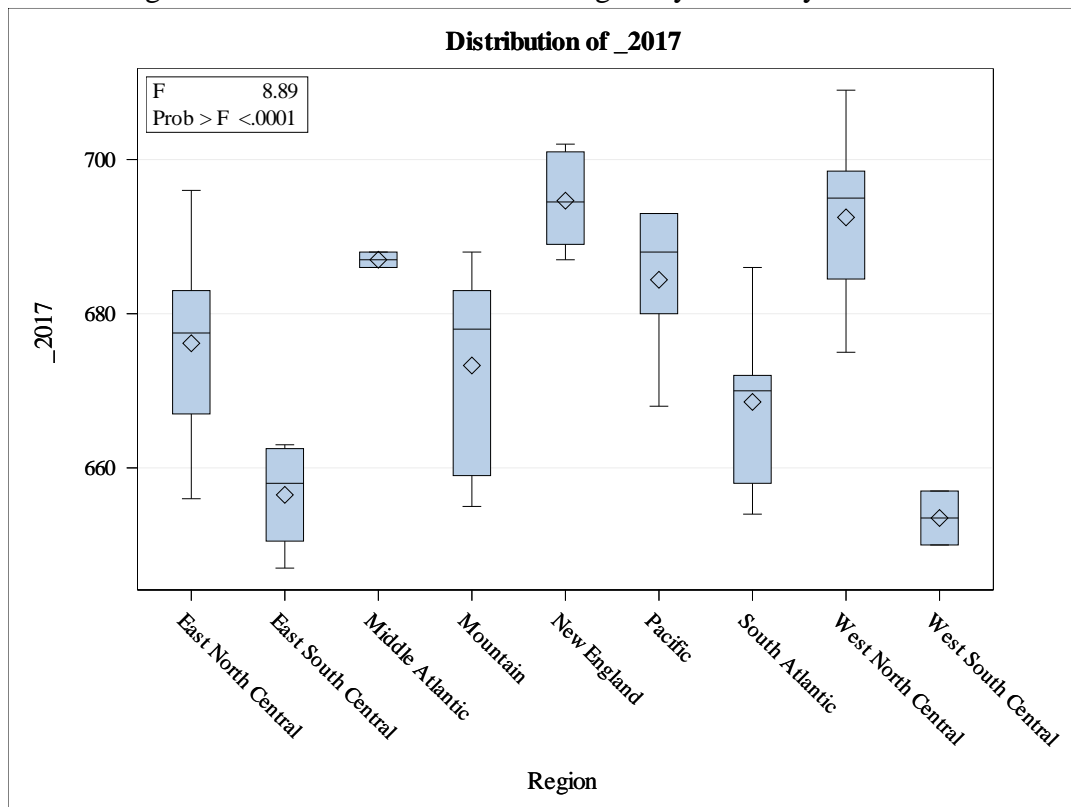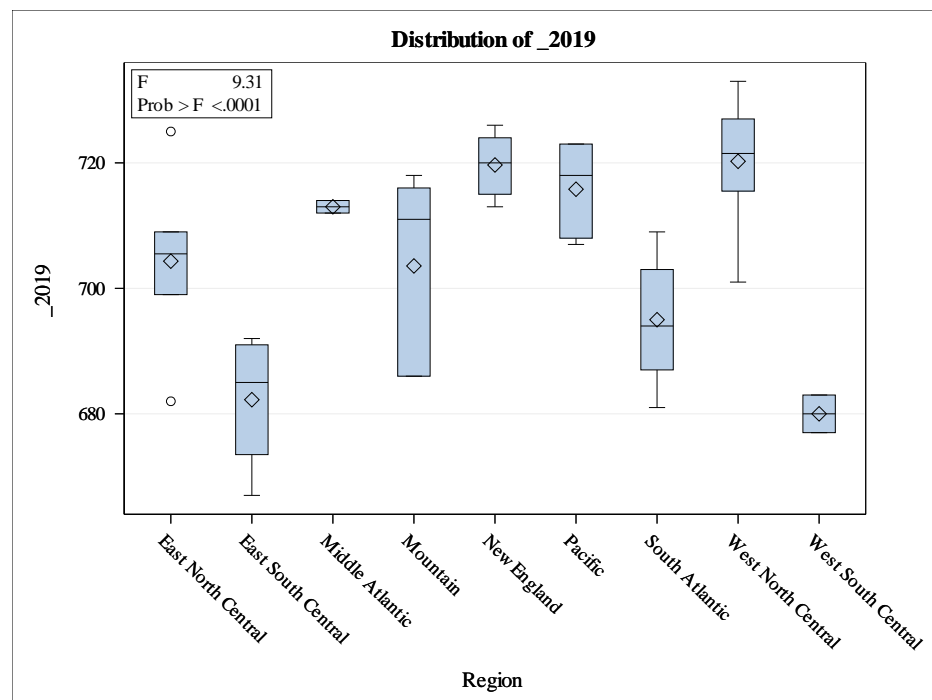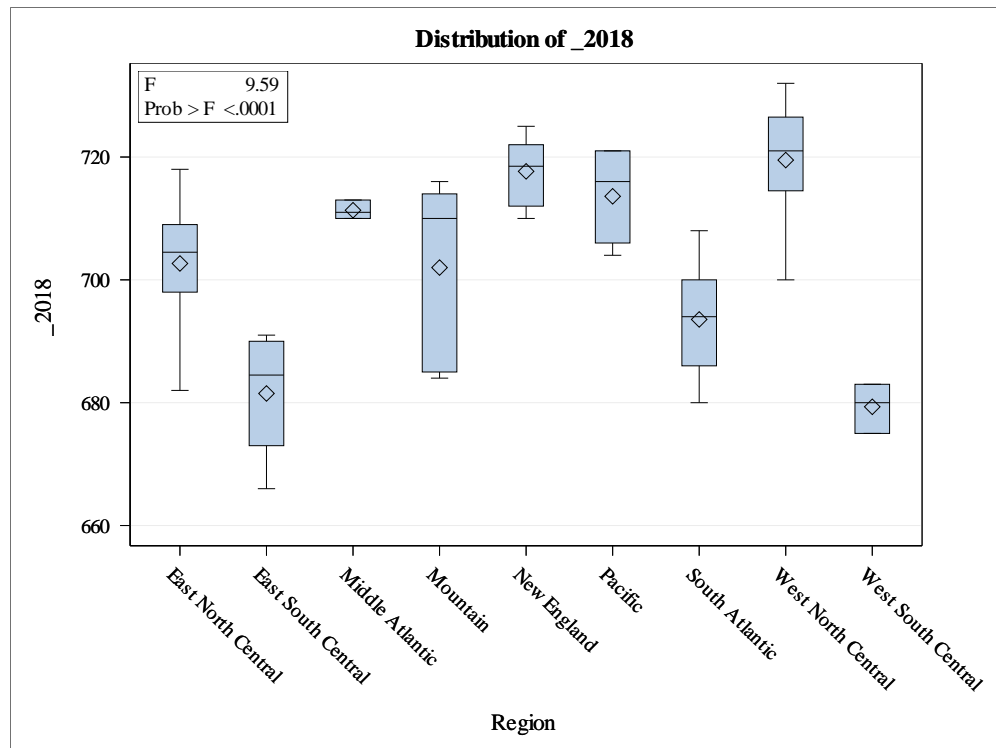
3. There are a number of factors that could account for the year-on-year improvement in FICO scores:

a. A decrease in unemployment would mean people are earning money again and can pay their debts
b. Changing of tax laws that make it easier to open businesses in certain regions, spurring job creation.
c. People not taking on new debt – increasing the average age of their accounts
d. People making more on time payments during the time period sampled.
e. Due to c & d people are now using less of their available credit and so, on average, scores go up.

4. The analysis shows that there is a significant between subjects effect. So, we know that there are differences between the regions and if we're going to make a recommendation, it's clear based on a visual analysis of the distributions of FICO scores, we would recommend higher interest rates in the East South Central, West South Central, Mountain & South Atlantic Regions to compensate for the increased credit risk while offering lower interest rates in the other regions as they display less of a credit risk.

The GLM Procedure
Repeated Measures Analysis of Variance
Test of Hypotheses for Between Subjects Effects

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Region	8	23767.89466	2970.98683	9.55	<.0001
Error	41	12750.26534	310.98208		

```sas
data _null_;
    rc=dlgcdir("Z:\OneDrive - Smart City Real Estate\Personal\Baruch\S5\STA
9797\HW\HW2");
    put rc=;
  run;
options msglevel =I;
run;

proc import file = "fico.xlsx" dbms = xlsx out = fico replace;
run;
ods rtf file = "STA 9797 HW 2.rtf";
ods graphics on;
proc glm data = fico plots=diagnostics;
    title "glm repeated by region";
    class Region;
    model _2017 - _2019 = Region;
    repeated year 3 (1 2 3) profile / printe;
run;
ods graphics off;
ods rtf close;
```