PROBLEM 1

This Data set had to be prepared before use. This is the Metada.

Data Set Name	WORK.STOCKS_AGE_A2	Observations	366
Member Type	DATA	Variables	5
Engine	V9	Indexes	0
Created	03/06/2021 22:19:17	Observation Length	32
Last Modified	03/06/2021 22:19:17	Deleted Observations	0
Protection		Compressed	NO
Data Set Type		Sorted	NO
Label			
Data Representation	SOLARIS_X86_64, LINUX_X86_64, ALPHA_TRU64, LINUX_IA64		
Encoding	utf-8 Unicode (UTF-8)		

Engine/Host Dependent Information	
Data Set Page Size	131072
Number of Data Set Pages	1
First Data Page	1
Max Obs per Page	4078
Obs in First Data Page	366
Number of Data Set Repairs	0
Filename	/saswork/SAS_work2C6000000B05_odaws01-usw2.oda.sas.com/SAS_workC86300000B05_odaws01-usw2.oda.sas.com/stocks_age_a2.sas7bdat
Release Created	9.0401M6
Host Created	Linux
Inode Number	537010501
Access Permission	rw-r--r--
Owner Name	u54770142
File Size	256KB
File Size (bytes)	262144

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
2	Age	Char	16	$16.	$16.	Age
3	C	Char	1	$1.	$1.	C
4	D	Char	1	$1.	$1.	D
5	E	Char	1	$1.	$1.	E
1	Stock_Inv	Num	8	COMMA15.1		Stock_Inv

```
*Problem 1 [10 marks] Use 5% as a significance level
In the last decade stockbrokers have drastically changed the way they do business.
Internet trading has become quite common and online trades can cost as little as $7.
It is now easier and cheaper to invest in the stock market than ever before.
What are the effects of these changes? To help answer this question,
a financial analyst randomly sampled 366 American households and asked each to report
the age of the head of the household and the proportion of their financial assets that
are invested in the stock market. The age categories are:
            Young (under 35)
            Early middle age (35 to 40)
            Late middle age (50 to 65)
            Senior (over 65)
The analyst was particularly interested in determining whether the ownership of stocks
varied by age. Do these data allow the analyst to determine that there are differences
in stock ownership between the four age groups? Check the required conditions.;


*I'm working with the following hypothesis:

Null hypothesis: There is no difference in the mean of Stock Investments by Age Group:
H0 : μ1= μ2= μ3= μ4

Alternative hypothesis: At least one of the age groups differs in their mean of Stock Investments:
H1: Not all the means are equal;


ods graphics on;
Title "One-way Anova Analysis to determine whether ownership of stocks varied by Age";
Proc glm data=work.stocks_age_a2 plots= diagnostics;
 Class Age;
 Model Stock_Inv = Age /ss3;
 Means Age / hovtest;
Run;
Quit;
Ods graphics off;
```

One-way Anova Analysis to determine whether ownership of stocks varied by Age

The GLM Procedure

Class Level Information		
Class	**Levels**	**Values**
Age	4	Early_Middle_Age Late_Middle_Age Senior Young

Number of Observations Read	366
Number of Observations Used	366

One-way Anova Analysis to determine whether ownership of stocks varied by Age

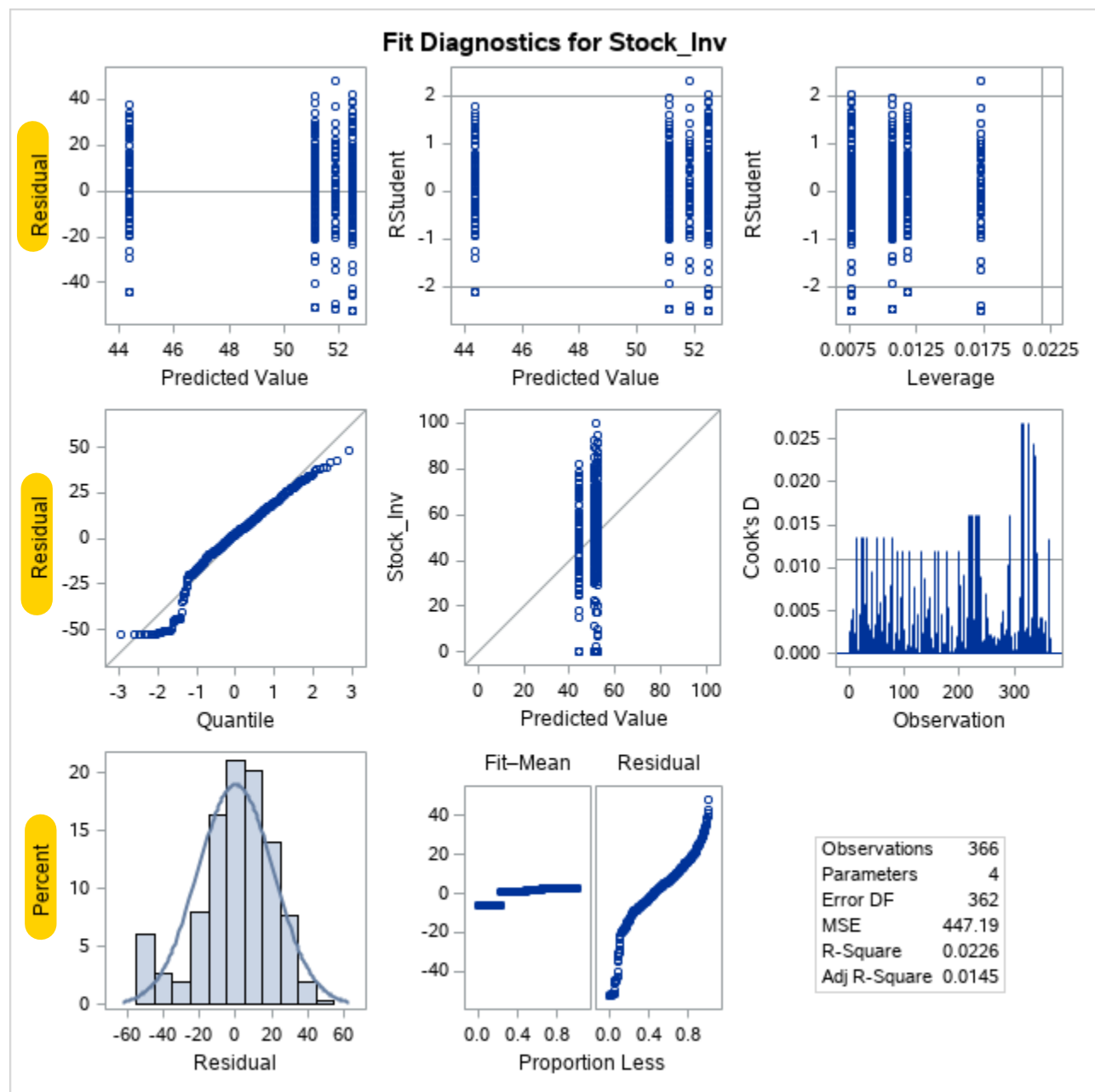### The GLM Procedure

Dependent Variable: Stock_Inv Stock_Inv

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	3745.9993	1248.6664	2.79	0.0403
Error	362	161881.8423	447.1874		
Corrected Total	365	165627.8416			

The P value is less than 0.05. Thus we reject the Null Hypothesis. The Ownership of Stock Investments varied by Age group

The fit of the model is not very good -> 0.0226

R-Square	Coeff Var	Root MSE	Stock_Inv Mean
0.022617	42.13956	21.14681	50.18279

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Age	3	3745.999289	1248.666430	2.79	0.0403

Fit Diagnostics for Stock_Inv

Observations	366
Parameters	4
Error DF	362
MSE	447.19
R-Square	0.0226
Adj R-Square	0.0145

Plot 1 gives us an idea of the variances within each Age Group (it shows residuals for each value of Age).
Plot 4 we can find out that the residuals are normally distributed because the residuals fall mostly on the diagonal line.
Plot 7, histogram of residuals confirms the assumption that the data points are normally distributed.
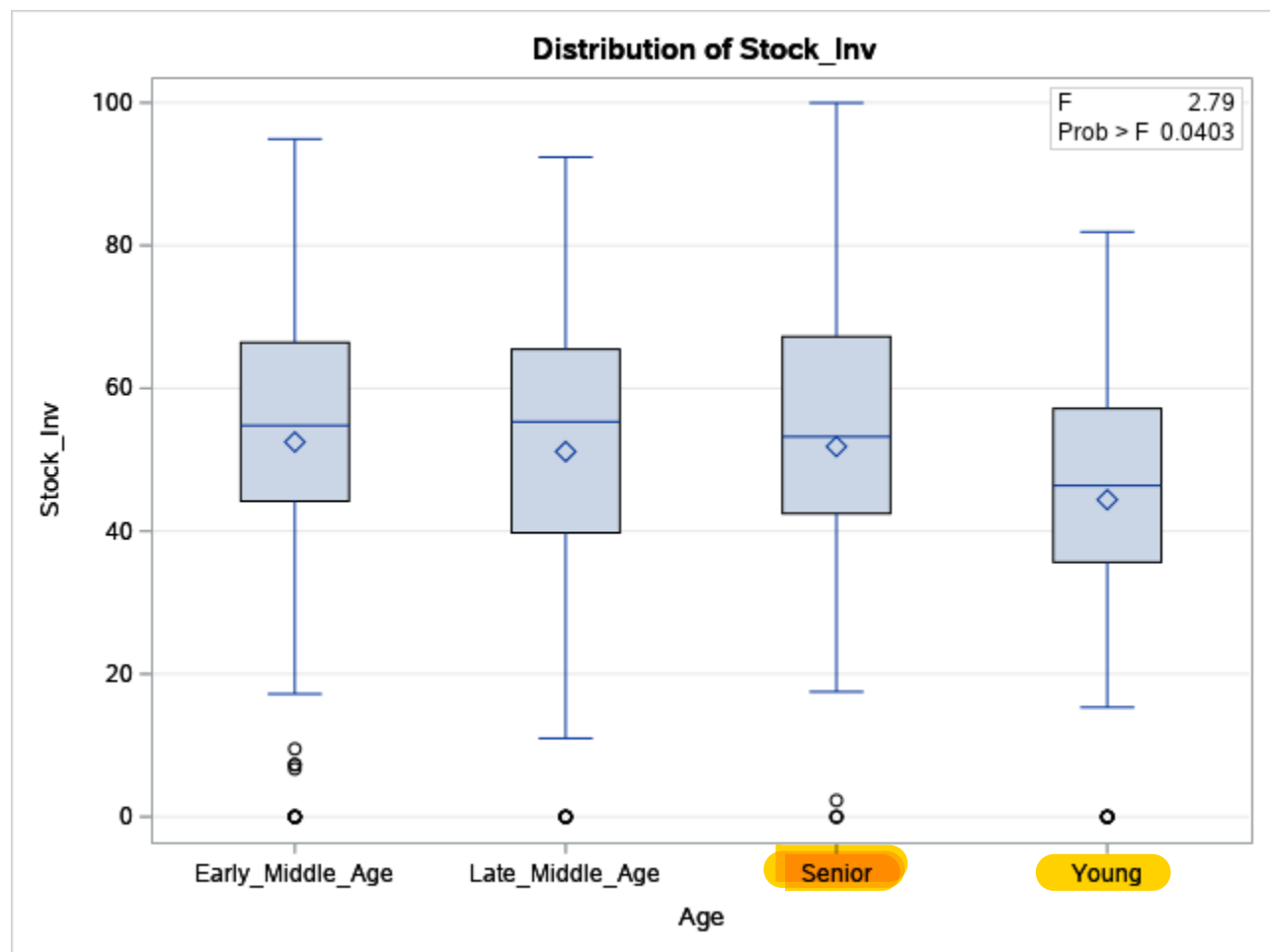
Distribution of Stock_Inv

| F | 2.79 |
| Prob > F | 0.0403 |

### One-way Anova Analysis to determine whether ownership of stocks varied by Age

The GLM Procedure

Levene's Test for Homogeneity of Stock_Inv Variance ANOVA of Squared Deviations from Group Means					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Age	3	436334	145445	0.30	0.8249
Error	362	1.7511E8	483723		

I ran the Levene's test as an additional test. Because the P value is higher than 0.05, we fail to reject the hypothesis of equal variance (homogeneity of variance). If he had obtained a value less than 0,05, we would have had to alternative methods to analyse the data, which is not the case

---

One-way Anova Analysis to determine whether ownership of stocks varied by Age

The GLM Procedure

Distribution of Stock_Inv
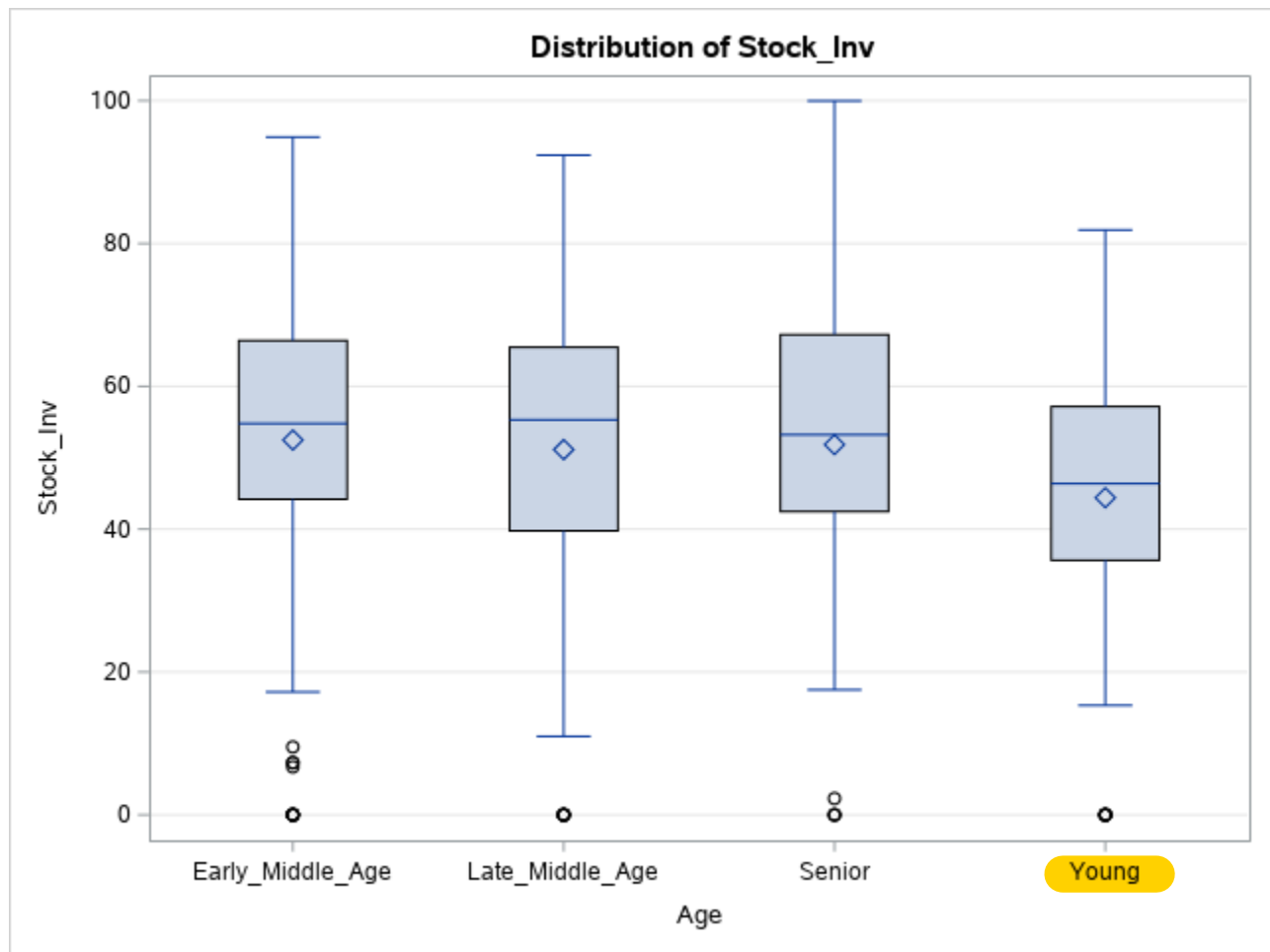
Level of Age	N	Stock_Inv	
		Mean	Std Dev
Early_Middle_Age	131	52.4801527	21.6684259
Late_Middle_Age	93	51.1390323	21.7215074
Senior	58	51.8381034	21.0900334
Young	84	44.3983333	19.6607843

Here we can see that the stock investment is lower for the Young Age group, increasing considerably for the Early_Middle_Age and staying at comparable levels for the other older groups.