STAT430 Project 2

Name: Ali Arshad Paracha

UID: 114689573

2. I will explore the effects of age, gender and employment type on the hours an individual spends playing video games per week. I want to see if there is a correlation between hours spent playing video games (per week) and age, gender and employment. Hence, the thing that I am trying to predict are hours and the predictors variable(s) that I will use may consist of either age, gender or employment type.

There are 5 variables in my study gender, age, employment type (student, part time, full time), preferred gaming platform and House spent playing video games per week. The data was collected by conducting surveys over multiple subreddits and through the STST430 class.

The results from project 1 show that both age and hours are skewed right and have outliers, so the best measure of center and spread for these quantitative variables was the five-number summary, which consists of the median and IQR. For the categorical data, project 1 shows that we have more males than females and more students than part/full time employees. This fact can be used to help our analysis later.

Note: Outliers were removed to get a better dataset, and to get better regression results.

3.

I. **Refer to Figure 1**:

The results show us that Age and hours spent gaming are correlated but negatively, they have a negative correlation value (-0.66) which means that as one of them increases the other decreases.

The table also tells us that, hours spent gaming and gender are correlated as well. Their correlation value is negative (-0.65) but this is because we encoded males to be the value 0 and females to be the value 1. This tells us that as the encoding increases the hours spent playing video games decreases, which in our context means that females spend lesser time playing video games than males do (which was also the conclusion we reached in project 1).

We also see that time spent playing video games is correlated to employment as well, though this one is a weaker correlation (-0.41). The correlation is negative because of our encodings. In our context and the way, we encoded the different employment types, this negative correlation tells us that students spend the most time playing video games, and the time spent playing video games decreases when individuals are part-time employees and then full-time employees.

Lastly, we see that Age and employment are heavily correlated as well (0.77). This means that as age increases the employment type goes from Student to Part time Employee to Full time employee, which is what we expect.

II.

i. **Refer to Figure 2:**

The reason we use Adjusted R squared values is because, it takes cares of more predictors being added and increases only if adding new predictors would increase its value by that which is expected by chance otherwise it decreases. It takes the number of variables into account and this gives us a much better idea of whether we should use the variables for regression.

Looking at the results of Figure 2 we can see that we have a high adjusted r-squared value for Gender and Age (0.819), this means that these two together have an effect on the hours spent gaming, provided they are not correlated. The same can be said about Gender and Employment, since the adjusted r squared value is high (0.708), we can assume that these two independent variables together effect the hours spent gaming, again with the assumption that they are not correlated. The same can be said about gender, employment and age all three together, their adjusted r square value is 0.8199. However, we can disregard this one because we saw above that both age and employment are heavily correlated (backed up by the regression results in **Figure 3**), and we are already considering age and gender together, so we can disregard this.

ii. **A) Refer to Figure 3 (Regression with Age, Gender and Employment as predictor variables):** Immediately we can see that the employment codes have a p value significantly greater than 0.05 (0.1444 and 0.9959), which tell us that employment is not a meaningful addition to the model and is not statistically significant. (This is most likely because age and employment are correlated, see results of **Figure 2**) This means that changes in its value are not related to the value of the response (Predicted) variable, so we can and should disregard it. The next regression we do (Figure 4) makes use of this fact and drops employment as one of the predictor variables.

B) Refer to Figure 4 (Regression with Age and Gender as predictor variables):

ANOVA Results: Looking at the results of Figure 4 we can see that the p - value is less than 0.05 (<0.001). This is significant because it means that our all our independent variables are most likely significant and that the y-variable (hours) and the x-variables (gender and age) are related. Furthermore, we have a small value of the MSE (8.775), which backs up the fact mentioned above and also tell us that we probably have a good model.

Regression Results:

1) The regression equation is:
Hours_Spent_Gaming = - 8.795 * Gender - 1.306 * Age + 44.673

2) Both gender and age have p-value less than 0.05 (< 0.001 and < 0.001 respectively), this means that both these variables are meaningful additions to

the model, because changes in their values are related to changes in the response variable (variable we are predicting).

3) The residual plot for age shows that the residuals are evenly (normally) distributed, they are evenly spread out across the zero line. Furthermore, they are also independent. The residual plot for Gender shows the same, the residuals are normally distributed across the zero line and are independent. (Note, the reason all points are zero and one is because that is how we encoded the two genders Male and Female, so those two are the only values they can take). Finally, we also look at the residual plot for the Predicted value and see that all points are normally distributed, this tell us that our model is a good one and the regression results are meaningful.

III. Assumptions:

a. Our dependent Variable is continuous, since it deals with time (Hours spent gaming). Hence, we are trying to predict a quantitative variable so multiple linear regression can be used. Our independent variable gender is categorical, and age is quantitative.

b. We see that our predictor terms are independent from one another. Age is independent of Gender, and also, we saw in the residual plots of Figure 4 that both age and gender are evenly distributed across the zero line, which fulfills another assumption.

c. Lastly, the residual plots in Figure 4 showed us that age, gender and hours were normally distributed, so our model fulfills all the necessary assumptions for Multiple Linear Regression.

Since, all these assumptions are met, we can be sure that our results are meaningful as these assumptions guarantee a useful model, one that we can use to accurately predict our dependent variable.

IV. Good Sample?

I believe the sample is good because, as mentioned above the it has a high adjusted r squared value (0.8193). Most predictor variables are independent of one another and have normally distributed residuals. Also, the sample has a small Mean Squared Error, compared to other models (8.77), which makes our sample a good one. That being said, there is also a slight chance the sample may not be good because it contains only 60 observations and it may not be completely representative of the entire population.

```
/* NOTE: Whenever I mention hours spent gaming below, I am referring to hours spent
playing video games PER WEEK, not mentioning it again and again to remove redundancy  */

/* Reading the dataset */
Data survey;
    INFILE '/home/aliparacha19960/EPG194/Survey.csv' delimiter=',';
    Length Platform $20;
    Length Employment $20;
    Input Gender $ Age Platform $ Hours_Spent_Gaming Employment $;
run;

/* Creating a new dataset that has encodings for all the catrgorical variables
(Gender, Platform and Employment type)  */
data survey2;
    set survey;

    if Gender='Male' then
        do;
            gender_code=0;
        end;
    else
        do;
            gender_code=1;
        end;

    if Platform='PC' then
        do;
            platform_code=0;
        end;
    else if Platform="XBOX" then
        do;
            platform_code=1;
        end;
    else
        do;
            platform_code=2;
        end;

    if Employment='Student' then
        do;
            employment_code1=0;
            employment_code2=0;
        end;
    else if Employment="Part - time Employee" then
        do;
            employment_code1=0;
            employment_code2=1;
        end;
    else
        do;
            employment_code1=1;
            employment_code2=0;
        end;
run;

/* Part 1: Find Correlation Cooefficient  */
/* Finding Pearson's coorelation coefficient, this will help us see which variables might
be correlated with one another  */
proc corr data=survey2;
Title "Figure 1";
run;
/* The results show us that Age and hours spent gaming are correlated but negatively,
they have a negative coorelation value (-0.66) which means that as one of them increases
the other decreases.
```

```
 The table also tells us that, hours spent gaming and gender are coorelated as well. Their
 coorelation value is negative as well (-0.65) but this is because we encoded males to be
 the value 0 and females to be the value 1. This tells us that as the encoding increases
 the hours spent playing video games decreases, which in our context means that females spend
 lesser time playing video games than males do (which was also the conclusion we reached in
 project 1).

 We also see that time spent playing video games is correlated to employment as well,
 though this one is a weaker coorelation (-0.41). The coorelation is negative because of our
 encodings. In our context and the way we encoded the different emplloyment types, this
 negative coorelation tell us that students spend the most time playing video games, and the time
 spent playing video games decreases when individuals are part-time employees and then full-time
 employees.

 Lastly, we see that Age and amployment are heavil coorelated as well (0.77). This means that
 as age increases the employment type goes from Student to Part time Employee to Full time
 employee, which is what we expect.
 */

/* Part 2: Perform Regression  */

/* I will now plot the Adjusted R-squared values to find the best model*/
proc reg data=survey2;
    model hours_spent_gaming = gender_code employment_code1 employment_code2 age / selection = adjrsq;
    Title "Figure 2";
run;
/* The reason we use Adjusted R squared values is because, it takes care of more predictors
being added and increases only if adding new predictors would increase its value by that which
is expected by chance, otherwise it decreases. This gives us a much better idea of whether we
should use the variables for regression.

Looking at the results of Figure 2 we can see that we have a really high adjusted r square
value for Gender and Age (0.819), this means that most likely these two together have an effect
on the hours spent gaming. The same can be said about Gender and Employment, since the adjusted
r squared value is high (0.708), we can assume that these two independent variables together
effect the hours spent gaming. The same can be said about gender, employment and age all three
together, their adjusted r square value is 0.8199 which means all three together might also effect the
time spent playing video games provided there isn't a correlation between the variables. */


/* Performing Regression by using gender, age and employment_type as predictor variables.
Hours_Spent_Gaming is the predicted variable.  */
proc reg data=survey2;
     model hours_spent_gaming = gender_code employment_code1 employment_code2 age;
     Title "Figure 3";
run;
/* Immediately we can see that the employment codes have a p value siginificantly greater than
0.05 (0.1444 and 0.9959), which tell us that employment is not a meaningful addition to the model
and is not statistically significant. This means that changes in its value are not related to
the value of the response (Predicted) vairable, so we can and should disregard it. */


/* Performing Regression by using gender and age as the predictor variables.
Hours_spent_gaming is the predicted variable */
proc reg data=survey2;
     model hours_spent_gaming = gender_code age;
     Title "Figure 4";
     plot hours_spent_gaming*age;
run;
/* Looking at the ANOVA results of Figure 3 we can see that the p-value is less than 0.05 (<0.001).
This is significant because it means that our indpendent variables are significant and
that the y-variable (hours) and the x-variables (gender and age) are related. Furthermore,
we have a small value of the MSE (8.775), which backs up the fact mentioned above and also
tell us that we probably have a good model.
```

```
Looking at the REGRESSION results of Figure 4 we see that

1) The regression equation is: Hours_Spent_gaming = -8.795 * Gender -1.306 * Age + 44.673
2) Both gender and age have p-value less than 0.05 (< 0.001 and < 0.001 respectively), this means
that both these variables are meaningful additions to the model, because  changes in their
values are related to changes in the response variable (variable we are predicting).
3) The reisudal plot for age shows that the residuals are evenly (normally) distributed, they
are evenly spread out across the zero line. Furthermore, they are also independent.
The residual plot for Gender shows the same, the residuals are normally distrbiuted across
the zero line and are independent. (Note, the reason all points are zero and one is because
that is how we encoded the two genders Male and Female, so those two are the only values they
can take). Finally, we also look at the residual plot for the Predicted value and see that all
points are normally distributed, this tell us that our model is a good one and the regression
results are meaningful.
*/

/* Part 3: Assumptions

i) Our dependent Variable is continuous, since it deals with time (Hours spent gaming). Hence,
we are trying to predict a quantitative variable so multiple linear regression can be used.
Our independent variable gender is categorical and age is quantitative.
ii) We see that our predictor terms are independent from one another. Age is independent of Gender,
and also we saw in the residual plots of Figure 4 that both age and gender are evenly distributed
across the zero line, which fulfills another assumption.
iii) Lastly, the residual plots in Figure 4 showed us that age,gender and hours were
normally distributed, so our model fulfills all the necessary assumptions for
Multiple Linear Regression.

Since, all these assunptions are met, we can be sure that our results are meaningful as these
assumptions gurantee a useful model, one that we can use to accurately predict our
dependent variable.

I believe the sample is good because, as mentioned above the it has a high
adjusted r squared value (0.8193). Most predictor variables are independent of one another
and have normally distributed residuals. Also, the sample has a small Mean Squared Error,
compared to other models (8.77), which makes our sampple a good one.
That being said, there is also a slight chance the sample may not be good because it
contains only 60 observations and it may not be completely representative of the entire
population.
 */
```

Figure 1

The CORR Procedure

6 Variables:	Age Hours_Spent_Gaming gender_code platform_code employment_code1 employment_code2

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Age	60	21.73333	3.37923	1304	16.00000	32.00000
Hours_Spent_Gaming	60	12.76667	6.96813	766.00000	1.00000	28.00000
gender_code	60	0.40000	0.49403	24.00000	0	1.00000
platform_code	60	0.73333	0.97192	44.00000	0	2.00000
employment_code1	60	0.20000	0.40338	12.00000	0	1.00000
employment_code2	60	0.15000	0.36008	9.00000	0	1.00000

Pearson Correlation Coefficients, N = 60
Prob > |r| under H0: Rho=0

	Age	Hours_Spent_Gaming	gender_code	platform_code	employment_code1	employment_code2
Age	1.00000	-0.66131 <.0001	0.04467 0.7347	-0.00138 0.9917	0.81071 <.0001	-0.06407 0.6267
Hours_Spent_Gaming	-0.66131 <.0001	1.00000	-0.65188 <.0001	-0.01435 0.9134	-0.46552 0.0002	0.10200 0.4380
gender_code	0.04467 0.7347	-0.65188 <.0001	1.00000	0.15532 0.2360	-0.15309 0.2429	-0.05717 0.6644
platform_code	-0.00138 0.9917	-0.01435 0.9134	0.15532 0.2360	1.00000	-0.20751 0.1116	0.06780 0.6067
employment_code1	0.81071 <.0001	-0.46552 0.0002	-0.15309 0.2429	-0.20751 0.1116	1.00000	-0.21004 0.1072
employment_code2	-0.06407 0.6267	0.10200 0.4380	-0.05717 0.6644	0.06780 0.6067	-0.21004 0.1072	1.00000

Figure 2

The REG Procedure
Model: MODEL1
Dependent Variable: Hours_Spent_Gaming

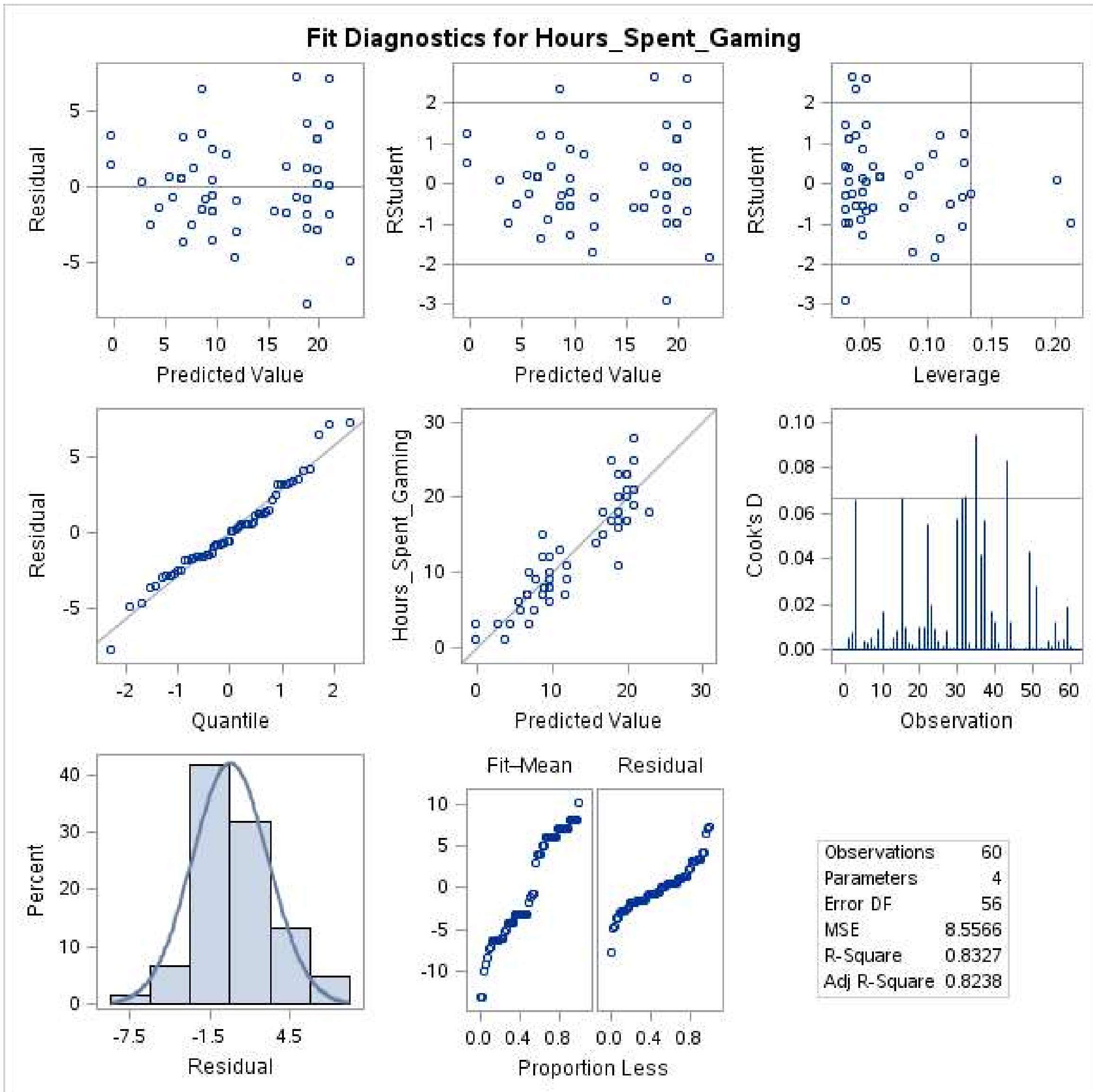
Adjusted R-Square Selection Method

Number of Observations Read	60
Number of Observations Used	60

Number in Model	Adjusted R-Square	R-Square	Variables in Model
3	0.8238	0.8327	gender_code employment_code1 Age
4	0.8206	0.8327	gender_code employment_code1 employment_code2 Age
2	0.8193	0.8254	gender_code Age
3	0.8168	0.8261	gender_code employment_code2 Age
2	0.7435	0.7522	gender_code employment_code1
3	0.7432	0.7562	gender_code employment_code1 employment_code2
2	0.4326	0.4519	employment_code1 Age
3	0.4322	0.4611	employment_code1 employment_code2 Age
1	0.4276	0.4373	Age
2	0.4213	0.4409	employment_code2 Age
1	0.4150	0.4249	gender_code
2	0.4091	0.4291	gender_code employment_code2
1	0.2032	0.2167	employment_code1
2	0.1892	0.2167	employment_code1 employment_code2
1	-.0067	0.0104	employment_code2

Figure 2

The REG Procedure
Model: MODEL1
Dependent Variable: Hours_Spent_Gaming



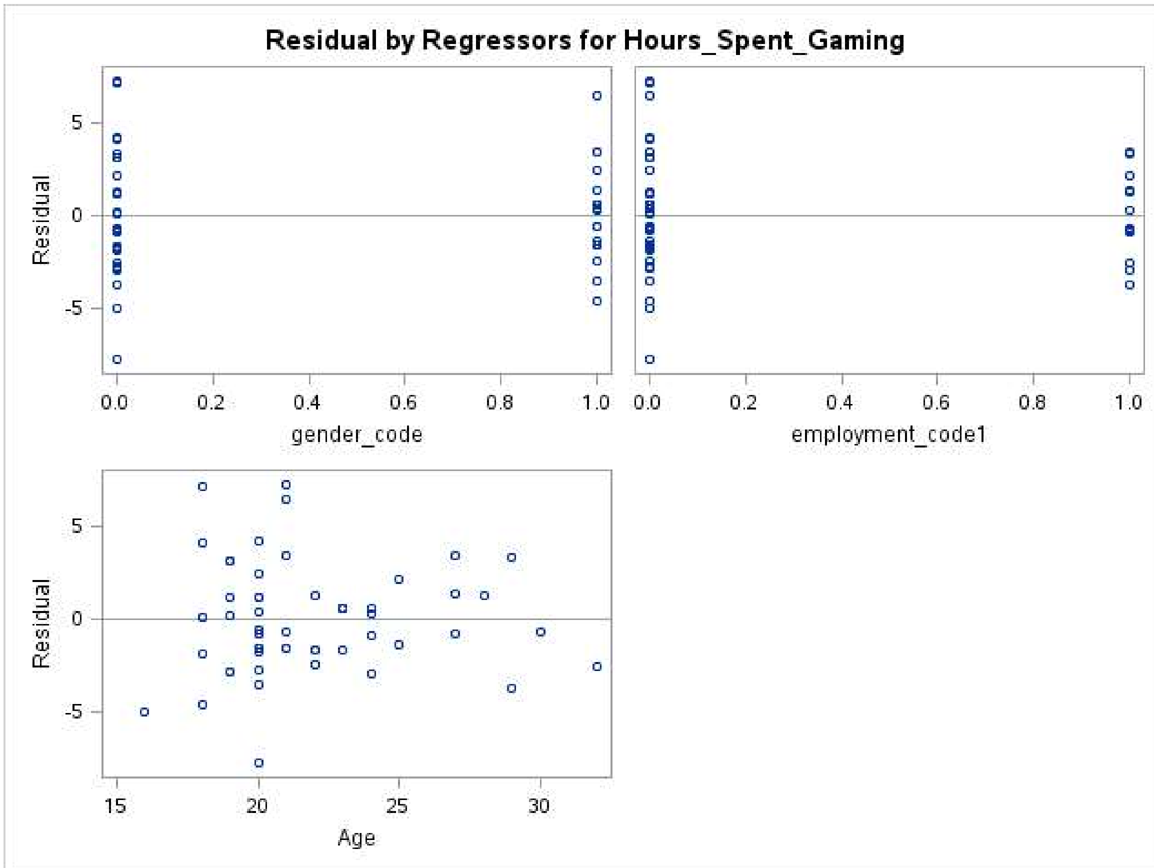


Figure 3

The REG Procedure
Model: MODEL1
Dependent Variable: Hours_Spent_Gaming

Number of Observations Read	60
Number of Observations Used	60

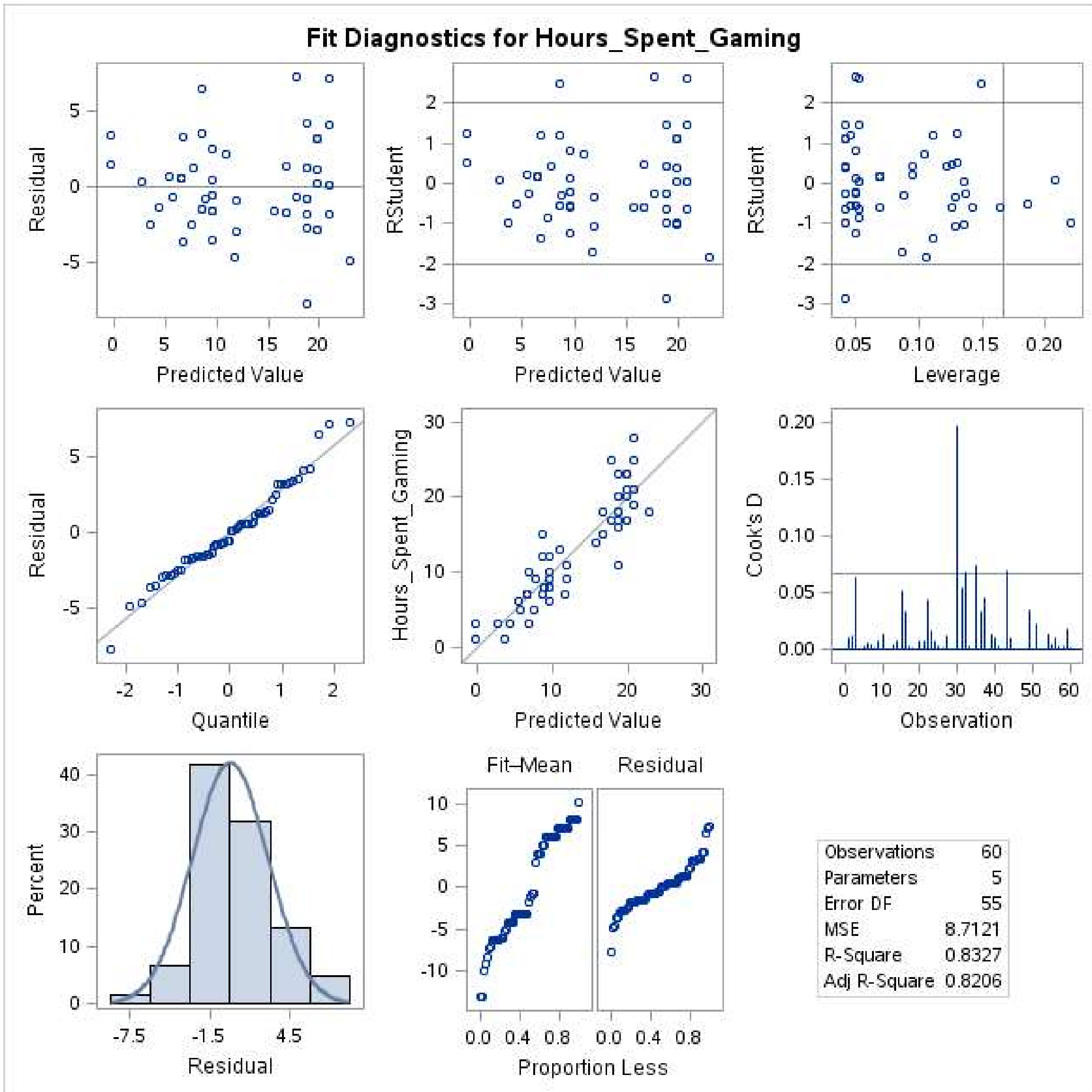
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	2385.56660	596.39165	68.46	<.0001
Error	55	479.16673	8.71212		
Corrected Total	59	2864.73333			

Root MSE	2.95163	R-Square	0.8327
Dependent Mean	12.76667	Adj R-Sq	0.8206
Coeff Var	23.11982		

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	39.69314	4.12594	9.62	<.0001
gender_code	1	-9.20964	0.83310	-11.05	<.0001
employment_code1	1	-2.67362	1.80589	-1.48	0.1444
employment_code2	1	-0.00580	1.12457	-0.01	0.9959
Age	1	-1.04480	0.20833	-5.02	<.0001

Figure 3

The REG Procedure
Model: MODEL1
Dependent Variable: Hours_Spent_Gaming



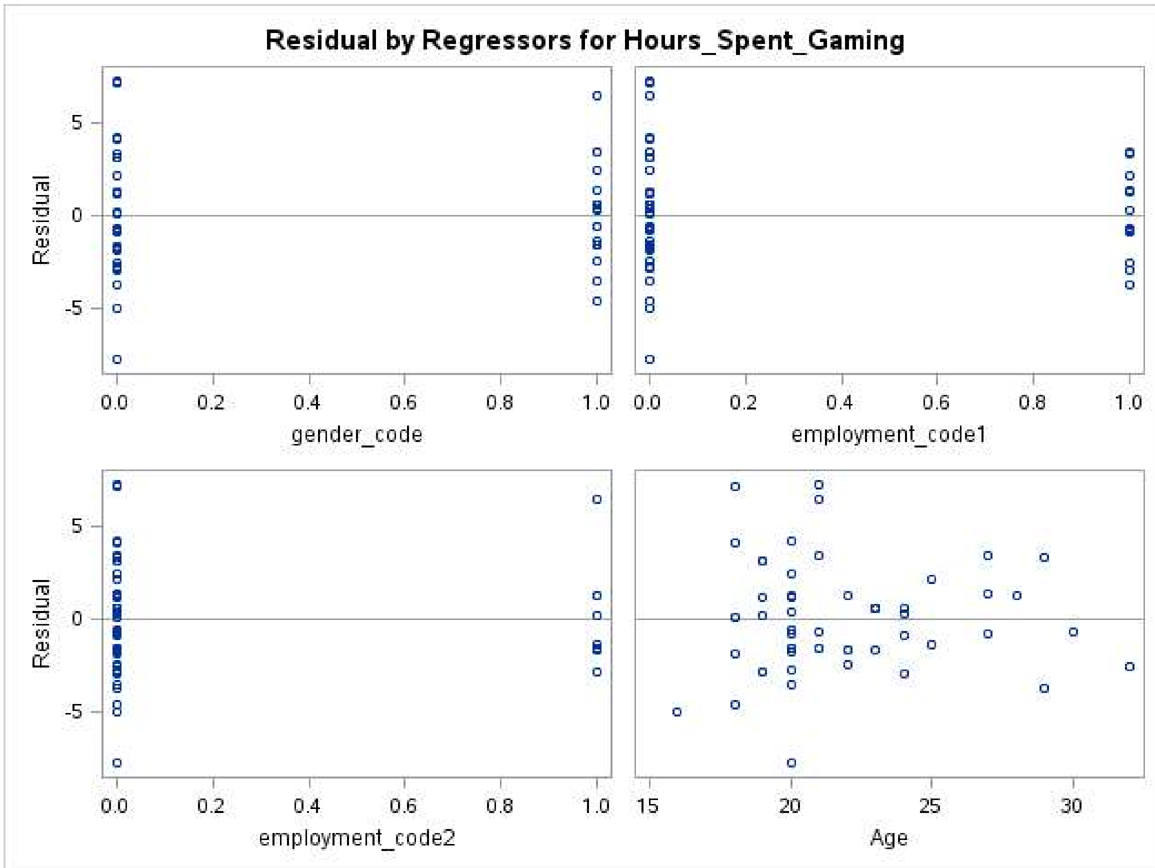


Figure 4

The REG Procedure
Model: MODEL1
Dependent Variable: Hours_Spent_Gaming

Number of Observations Read	60
Number of Observations Used	60

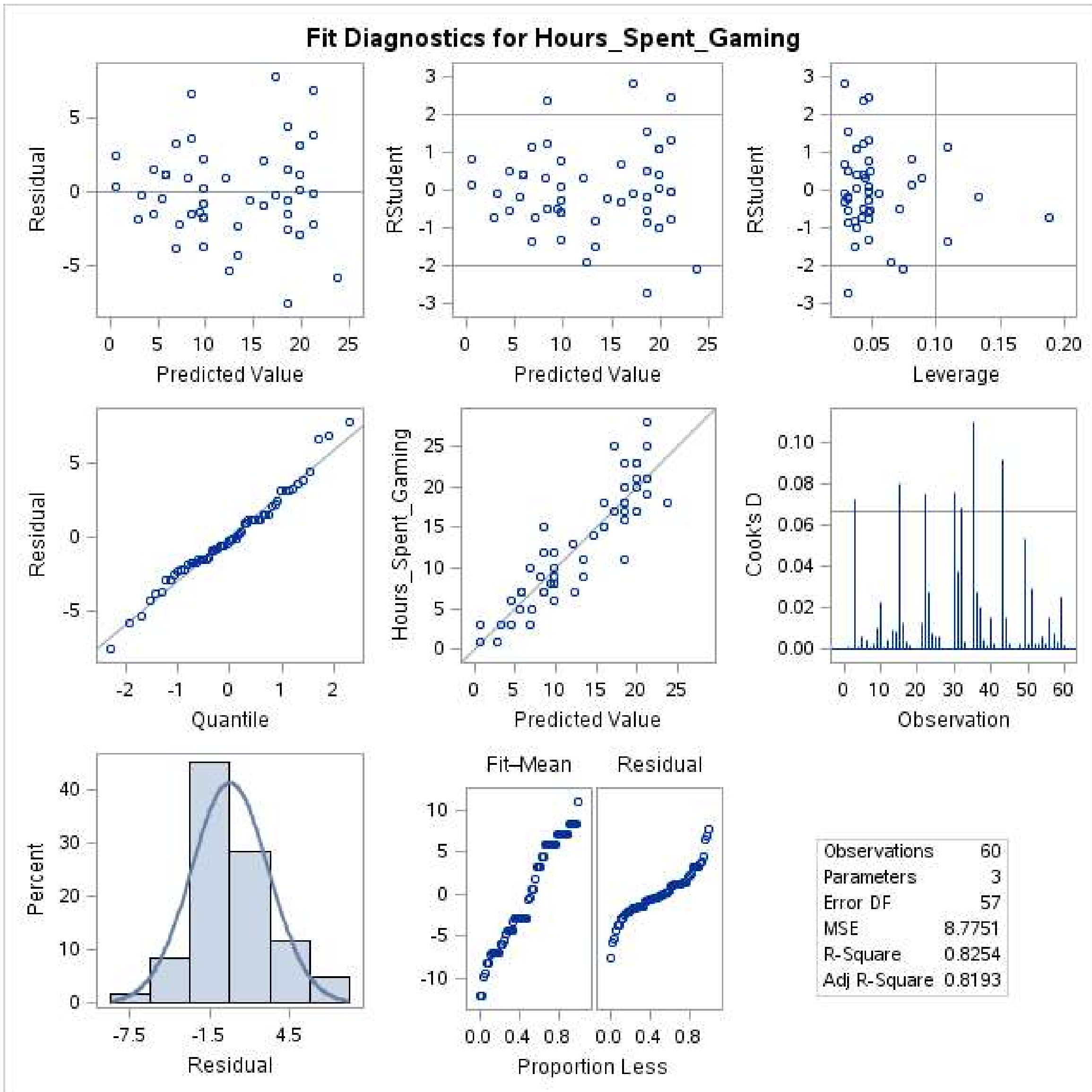
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	2	2364.55533	1182.27767	134.73	<.0001
Error	57	500.17800	8.77505		
Corrected Total	59	2864.73333			

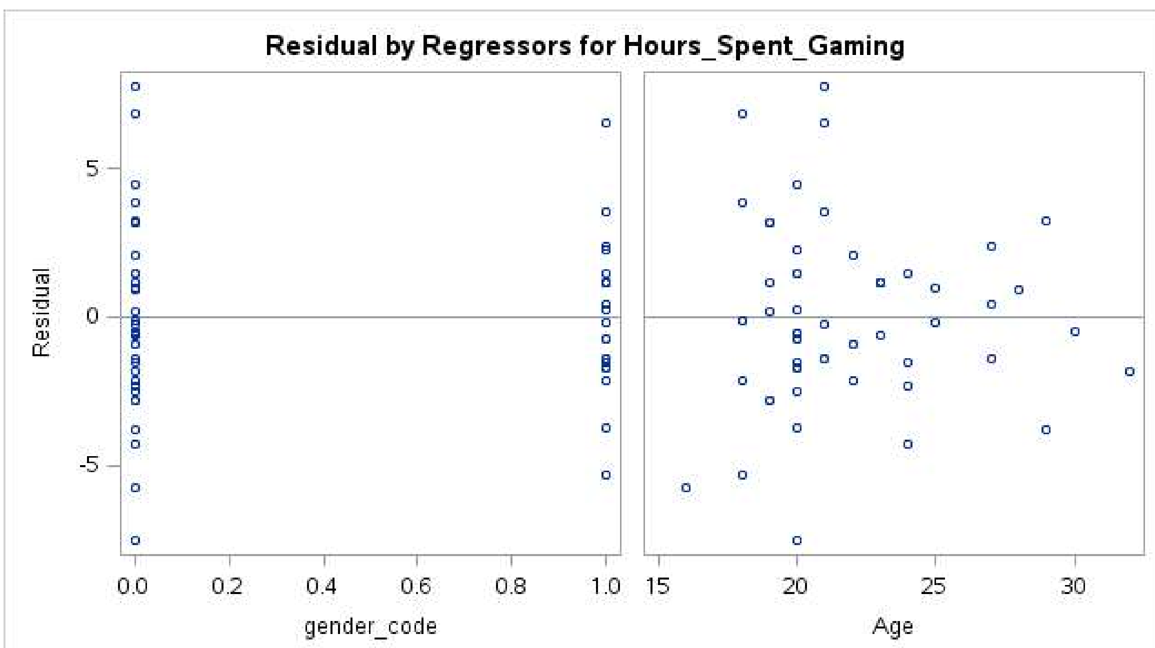
Root MSE	2.96227	**R-Square**	0.8254
Dependent Mean	12.76667	**Adj R-Sq**	0.8193
Coeff Var	23.20317		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	44.67299	2.51771	17.74	<.0001
gender_code	1	-8.79533	0.78141	-11.26	<.0001
Age	1	-1.30621	0.11424	-11.43	<.0001

Figure 4

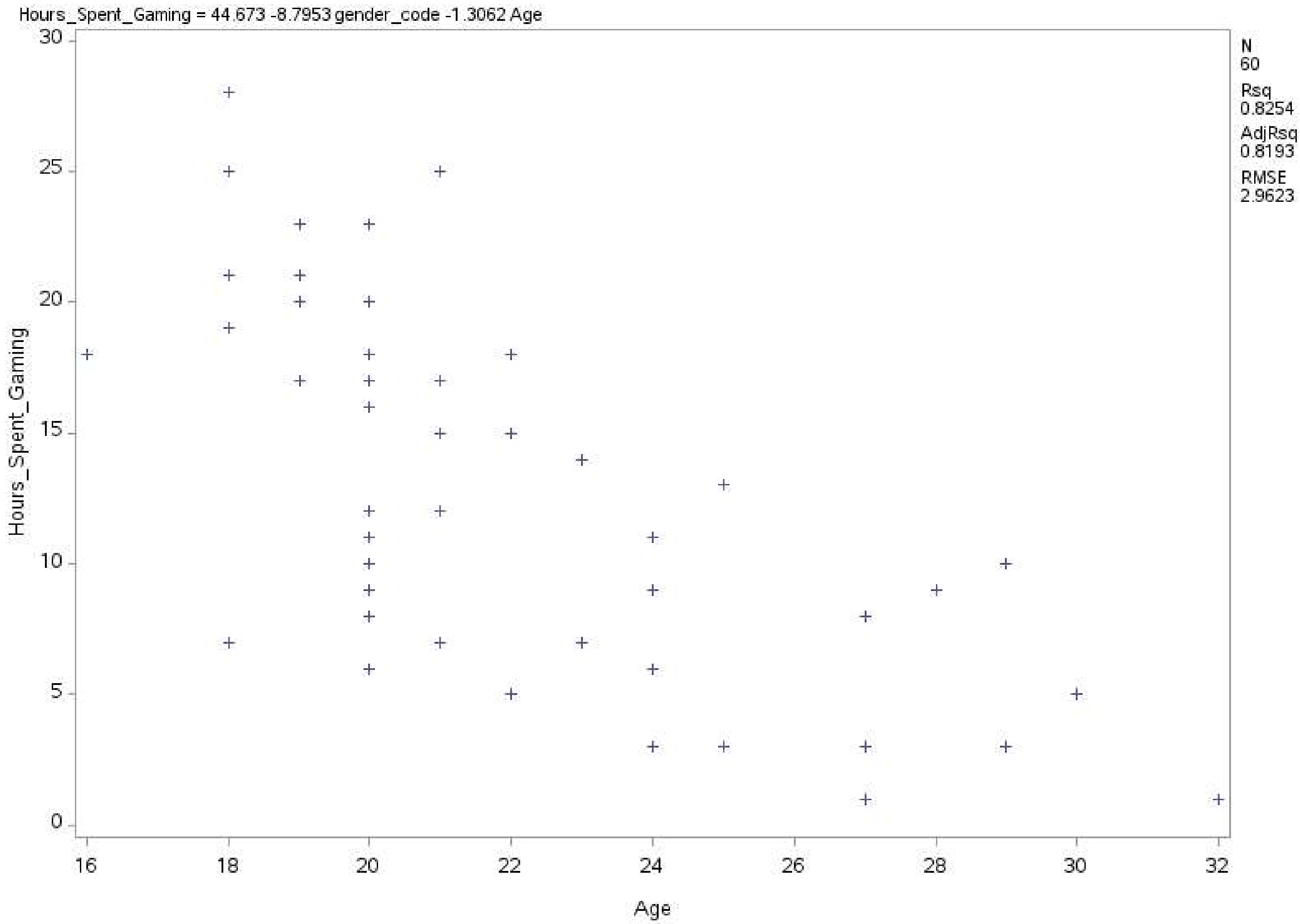
The REG Procedure
Model: MODEL1
Dependent Variable: Hours_Spent_Gaming





The REG Procedure

Figure 4



```
/* NOTE: Whenever I mention hours spent gaming below, I am referring to hours spent
playing video games PER WEEK, not mentioning it again and again to remove redundancy  */

/* Reading the dataset */
Data survey;
    INFILE '/home/aliparacha19960/EPG194/Survey.csv' delimiter=',';
    Length Platform $20;
    Length Employment $20;
    Input Gender $ Age Platform $ Hours_Spent_Gaming Employment $;
run;

/* Creating a new dataset that has encodings for all the catrgorical variables
(Gender, Platform and Employment type)  */
data survey2;
    set survey;

    if Gender='Male' then
        do;
            gender_code=0;
        end;
    else
        do;
            gender_code=1;
        end;

    if Platform='PC' then
        do;
            platform_code=0;
        end;
    else if Platform="XBOX" then
        do;
            platform_code=1;
        end;
    else
        do;
            platform_code=2;
        end;

    if Employment='Student' then
        do;
            employment_code1=0;
            employment_code2=0;
        end;
    else if Employment="Part - time Employee" then
        do;
            employment_code1=0;
            employment_code2=1;
        end;
    else
        do;
            employment_code1=1;
            employment_code2=0;
        end;
run;

/* Part 1: Find Correlation Cooefficient  */
/* Finding Pearson's coorelation coefficient, this will help us see which variables might
be correlated with one another  */
proc corr data=survey2;
Title "Figure 1";
run;
/* The results show us that Age and hours spent gaming are correlated but negatively,
they have a negative coorelation value (-0.66) which means that as one of them increases
the other decreases.
```

```
 The table also tells us that, hours spent gaming and gender are coorelated as well. Their
 coorelation value is negative as well (-0.65) but this is because we encoded males to be
 the value 0 and females to be the value 1. This tells us that as the encoding increases
 the hours spent playing video games decreases, which in our context means that females spend
 lesser time playing video games than males do (which was also the conclusion we reached in
 project 1).

 We also see that time spent playing video games is correlated to employment as well,
 though this one is a weaker coorelation (-0.41). The coorelation is negative because of our
 encodings. In our context and the way we encoded the different emplloyment types, this
 negative coorelation tell us that students spend the most time playing video games, and the time
 spent playing video games decreases when individuals are part-time employees and then full-time
 employees.

 Lastly, we see that Age and amployment are heavil coorelated as well (0.77). This means that
 as age increases the employment type goes from Student to Part time Employee to Full time
 employee, which is what we expect.
 */

/* Part 2: Perform Regression  */

/* I will now plot the Adjusted R-squared values to find the best model*/
proc reg data=survey2;
    model hours_spent_gaming = gender_code employment_code1 employment_code2 age / selection = adjrsq;
    Title "Figure 2";
run;
/* The reason we use Adjusted R squared values is because, it takes care of more predictors
being added and increases only if adding new predictors would increase its value by that which
is expected by chance, otherwise it decreases. This gives us a much better idea of whether we
should use the variables for regression.

Looking at the results of Figure 2 we can see that we have a really high adjusted r square
value for Gender and Age (0.819), this means that most likely these two together have an effect
on the hours spent gaming. The same can be said about Gender and Employment, since the adjusted
r squared value is high (0.708), we can assume that these two independent variables together
effect the hours spent gaming. The same can be said about gender, employment and age all three
together, their adjusted r square value is 0.8199 which means all three together might also effect the
time spent playing video games provided there isn't a correlation between the variables. */


/* Performing Regression by using gender, age and employment_type as predictor variables.
Hours_Spent_Gaming is the predicted variable.  */
proc reg data=survey2;
     model hours_spent_gaming = gender_code employment_code1 employment_code2 age;
     Title "Figure 3";
run;
/* Immediately we can see that the employment codes have a p value siginificantly greater than
0.05 (0.1444 and 0.9959), which tell us that employment is not a meaningful addition to the model
and is not statistically significant. This means that changes in its value are not related to
the value of the response (Predicted) vairable, so we can and should disregard it. */


/* Performing Regression by using gender and age as the predictor variables.
Hours_spent_gaming is the predicted variable */
proc reg data=survey2;
     model hours_spent_gaming = gender_code age;
     Title "Figure 4";
     plot hours_spent_gaming*age;
run;
/* Looking at the ANOVA results of Figure 4 we can see that the p-value is less than 0.05 (<0.001).
This is significant because it means that our indpendent variables are significant and
that the y-variable (hours) and the x-variables (gender and age) are related. Furthermore,
we have a small value of the MSE (8.775), which backs up the fact mentioned above and also
tell us that we probably have a good model.
```

```
Looking at the REGRESSION results of Figure 4 we see that

1) The regression equation is: Hours_Spent_gaming = -8.795 * Gender -1.306 * Age + 44.673
2) Both gender and age have p-value less than 0.05 (< 0.001 and < 0.001 respectively), this means
that both these variables are meaningful additions to the model, because  changes in their
values are related to changes in the response variable (variable we are predicting).
3) The reisudal plot for age shows that the residuals are evenly (normally) distributed, they
are evenly spread out across the zero line. Furthermore, they are also independent.
The residual plot for Gender shows the same, the residuals are normally distrbiuted across
the zero line and are independent. (Note, the reason all points are zero and one is because
that is how we encoded the two genders Male and Female, so those two are the only values they
can take). Finally, we also look at the residual plot for the Predicted value and see that all
points are normally distributed, this tell us that our model is a good one and the regression
results are meaningful. However, we still must compare this model to the one in which we use gender
and employment as predictor variables, because since age and employment are correlated we can not
be sure which model is a better one.
*/

/* Performing Regression that uses gender and employment as predictor variabes */
proc reg data=survey2;
     model hours_spent_gaming = gender_code employment_code1 employment_code2;
     Title "Figure 5";
run;
/* The results of figure 5 show us that the model had a adjusted r-squared value of
0.7435, which is less than the Adjusted R-squared value for our previous model (Figure 4).
Also, the MSE for this model is higher(12.45) compared to our previous model, hence
we can reach the conclusion that this model is not a better one than the one that contains
age as a predictor variable. */


/* Part 3: Assumptions

i) Our dependent Variable is continuous, since it deals with time (Hours spent gaming). Hence,
we are trying to predict a quantitative variable so multiple linear regression can be used.
Our independent variable gender is categorical and age is quantitative.
ii) We see that our predictor terms are independent from one another. Age is independent of Gender,
and also we saw in the residual plots of Figure 4 that both age and gender are evenly distributed
across the zero line, which fulfills another assumption.
iii) Lastly, the residual plots in Figure 4 showed us that age,gender and hours were
normally distributed, so our model fulfills all the necessary assumptions for
Multiple Linear Regression.

Since, all these assunptions are met, we can be sure that our results are meaningful as these
assumptions gurantee a useful model, one that we can use to accurately predict our
dependent variable.

I believe the sample is good because, as mentioned above the it has a high
adjusted r squared value (0.8193). Most predictor variables are independent of one another
and have normally distributed residuals. Also, the sample has a small Mean Squared Error,
compared to other models (8.77), which makes our sampple a good one.
That being said, there is also a slight chance the sample may not be good because it
contains only 60 observations and it may not be completely representative of the entire
population.
 */
```

Figure 1

The CORR Procedure

6 Variables:	Age Hours_Spent_Gaming gender_code platform_code employment_code1 employment_code2

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Age	60	21.73333	3.37923	1304	16.00000	32.00000
Hours_Spent_Gaming	60	12.76667	6.96813	766.00000	1.00000	28.00000
gender_code	60	0.40000	0.49403	24.00000	0	1.00000
platform_code	60	0.73333	0.97192	44.00000	0	2.00000
employment_code1	60	0.20000	0.40338	12.00000	0	1.00000
employment_code2	60	0.15000	0.36008	9.00000	0	1.00000

Pearson Correlation Coefficients, N = 60
Prob > |r| under H0: Rho=0

	Age	Hours_Spent_Gaming	gender_code	platform_code	employment_code1	employment_code2
Age	1.00000	-0.66131 <.0001	0.04467 0.7347	-0.00138 0.9917	0.81071 <.0001	-0.06407 0.6267
Hours_Spent_Gaming	-0.66131 <.0001	1.00000	-0.65188 <.0001	-0.01435 0.9134	-0.46552 0.0002	0.10200 0.4380
gender_code	0.04467 0.7347	-0.65188 <.0001	1.00000	0.15532 0.2360	-0.15309 0.2429	-0.05717 0.6644
platform_code	-0.00138 0.9917	-0.01435 0.9134	0.15532 0.2360	1.00000	-0.20751 0.1116	0.06780 0.6067
employment_code1	0.81071 <.0001	-0.46552 0.0002	-0.15309 0.2429	-0.20751 0.1116	1.00000	-0.21004 0.1072
employment_code2	-0.06407 0.6267	0.10200 0.4380	-0.05717 0.6644	0.06780 0.6067	-0.21004 0.1072	1.00000

Figure 2

The REG Procedure
Model: MODEL1
Dependent Variable: Hours_Spent_Gaming

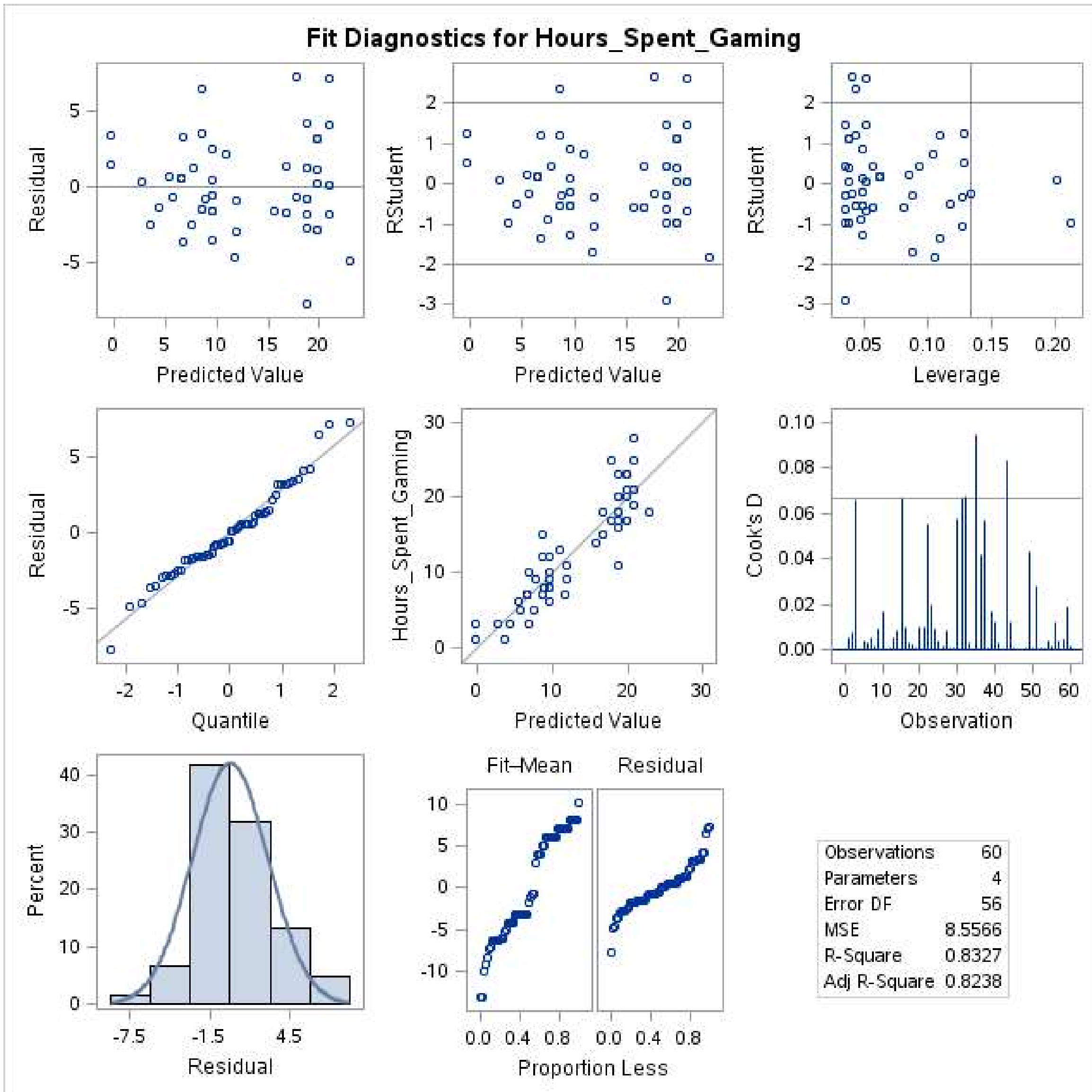
Adjusted R-Square Selection Method

Number of Observations Read	60
Number of Observations Used	60

Number in Model	Adjusted R-Square	R-Square	Variables in Model
3	0.8238	0.8327	gender_code employment_code1 Age
4	0.8206	0.8327	gender_code employment_code1 employment_code2 Age
2	0.8193	0.8254	gender_code Age
3	0.8168	0.8261	gender_code employment_code2 Age
2	0.7435	0.7522	gender_code employment_code1
3	0.7432	0.7562	gender_code employment_code1 employment_code2
2	0.4326	0.4519	employment_code1 Age
3	0.4322	0.4611	employment_code1 employment_code2 Age
1	0.4276	0.4373	Age
2	0.4213	0.4409	employment_code2 Age
1	0.4150	0.4249	gender_code
2	0.4091	0.4291	gender_code employment_code2
1	0.2032	0.2167	employment_code1
2	0.1892	0.2167	employment_code1 employment_code2
1	-.0067	0.0104	employment_code2

Figure 2

The REG Procedure
Model: MODEL1
Dependent Variable: Hours_Spent_Gaming



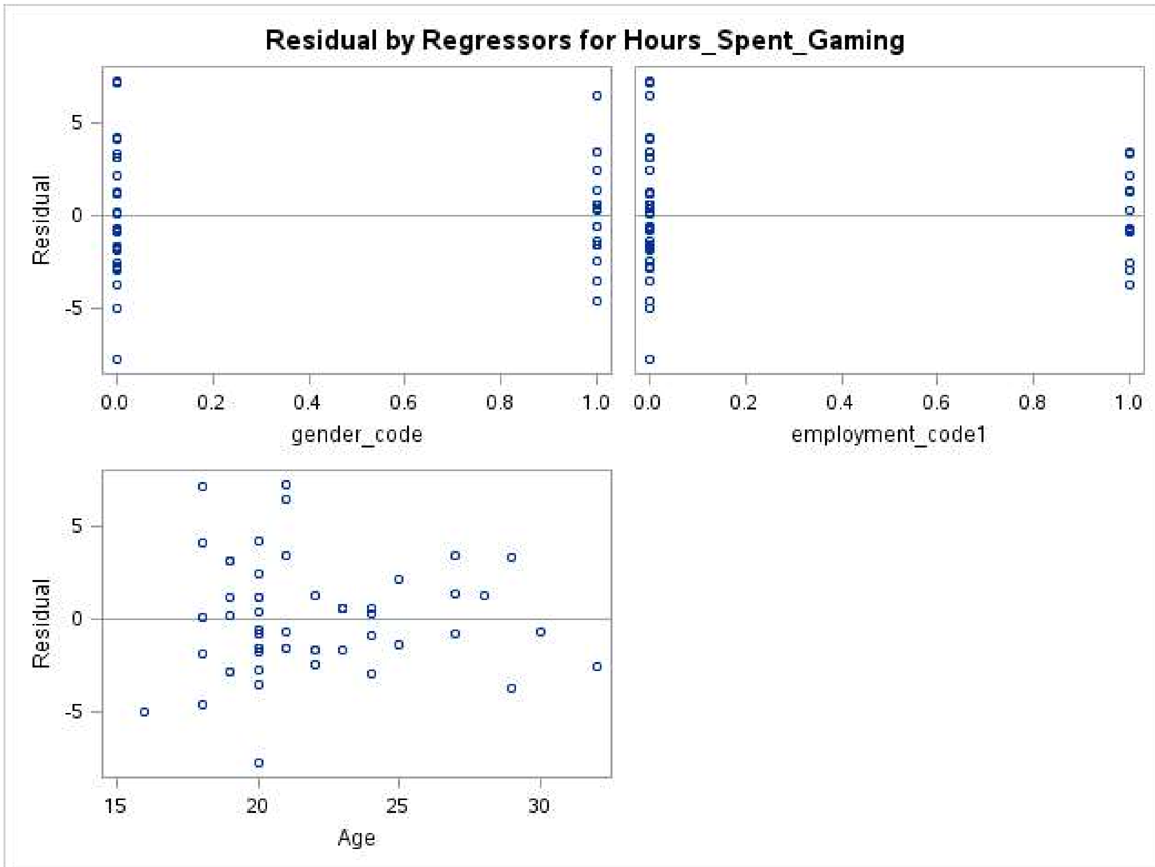


Figure 3

The REG Procedure
Model: MODEL1
Dependent Variable: Hours_Spent_Gaming

Number of Observations Read	60
Number of Observations Used	60

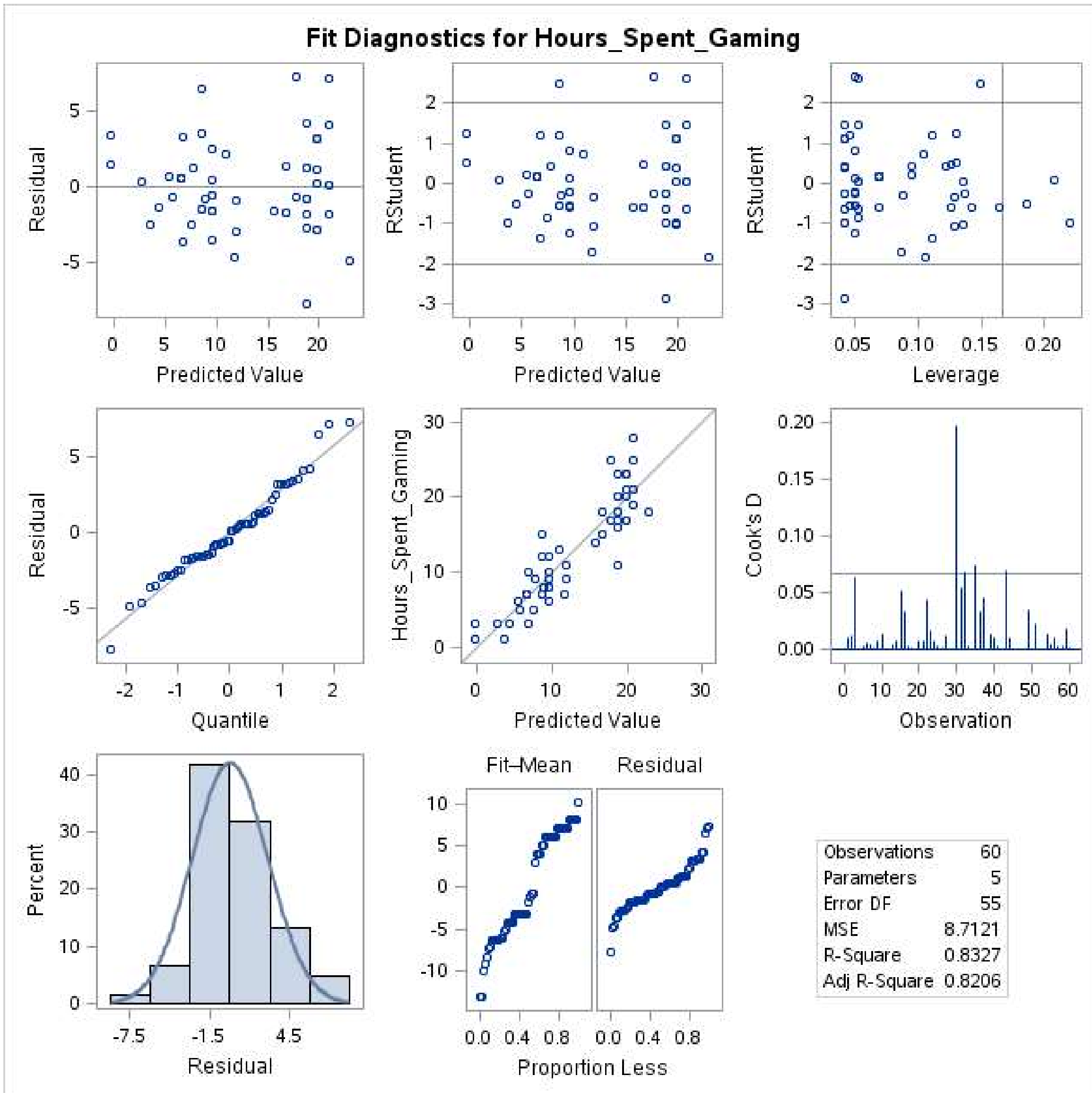
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	4	2385.56660	596.39165	68.46	<.0001
Error	55	479.16673	8.71212		
Corrected Total	59	2864.73333			

Root MSE	2.95163	**R-Square**	0.8327
Dependent Mean	12.76667	**Adj R-Sq**	0.8206
Coeff Var	23.11982		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	39.69314	4.12594	9.62	<.0001
gender_code	1	-9.20964	0.83310	-11.05	<.0001
employment_code1	1	-2.67362	1.80589	-1.48	0.1444
employment_code2	1	-0.00580	1.12457	-0.01	0.9959
Age	1	-1.04480	0.20833	-5.02	<.0001

Figure 3

The REG Procedure
Model: MODEL1
Dependent Variable: Hours_Spent_Gaming



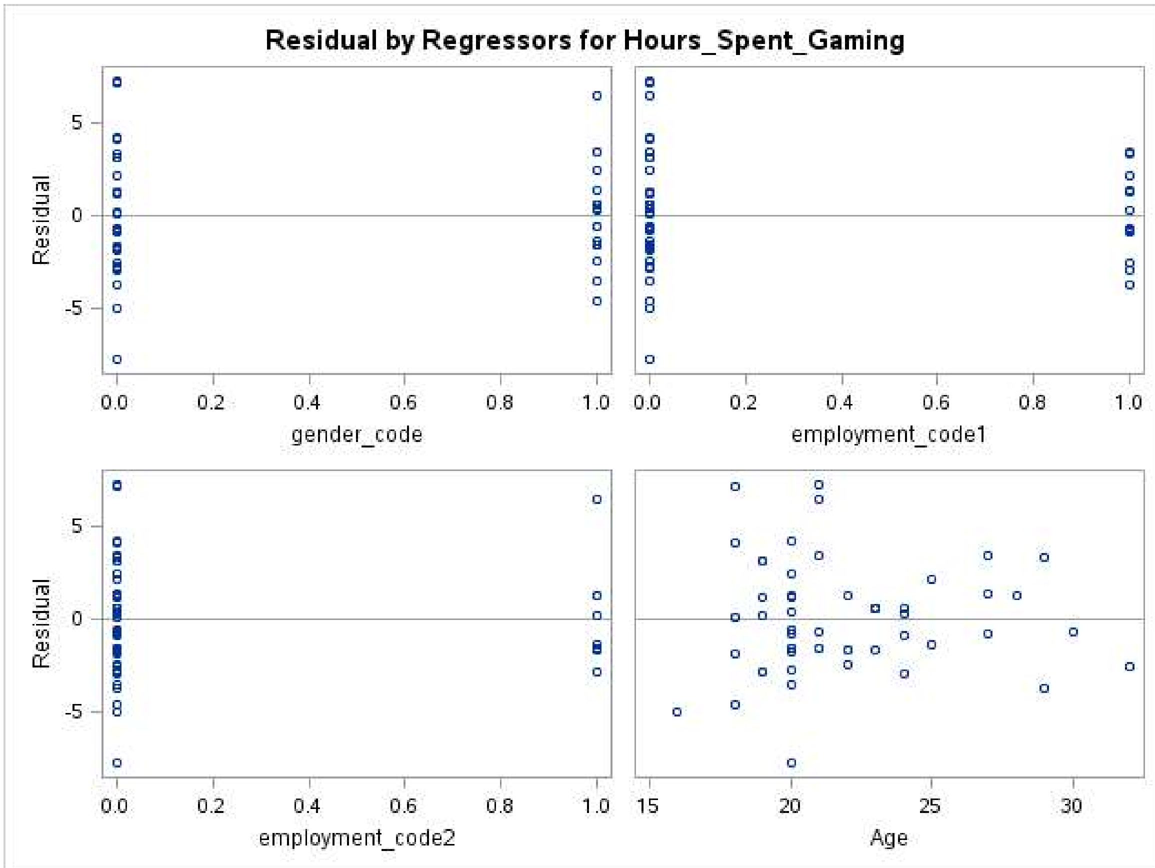


Figure 4

The REG Procedure
Model: MODEL1
Dependent Variable: Hours_Spent_Gaming

Number of Observations Read	60
Number of Observations Used	60

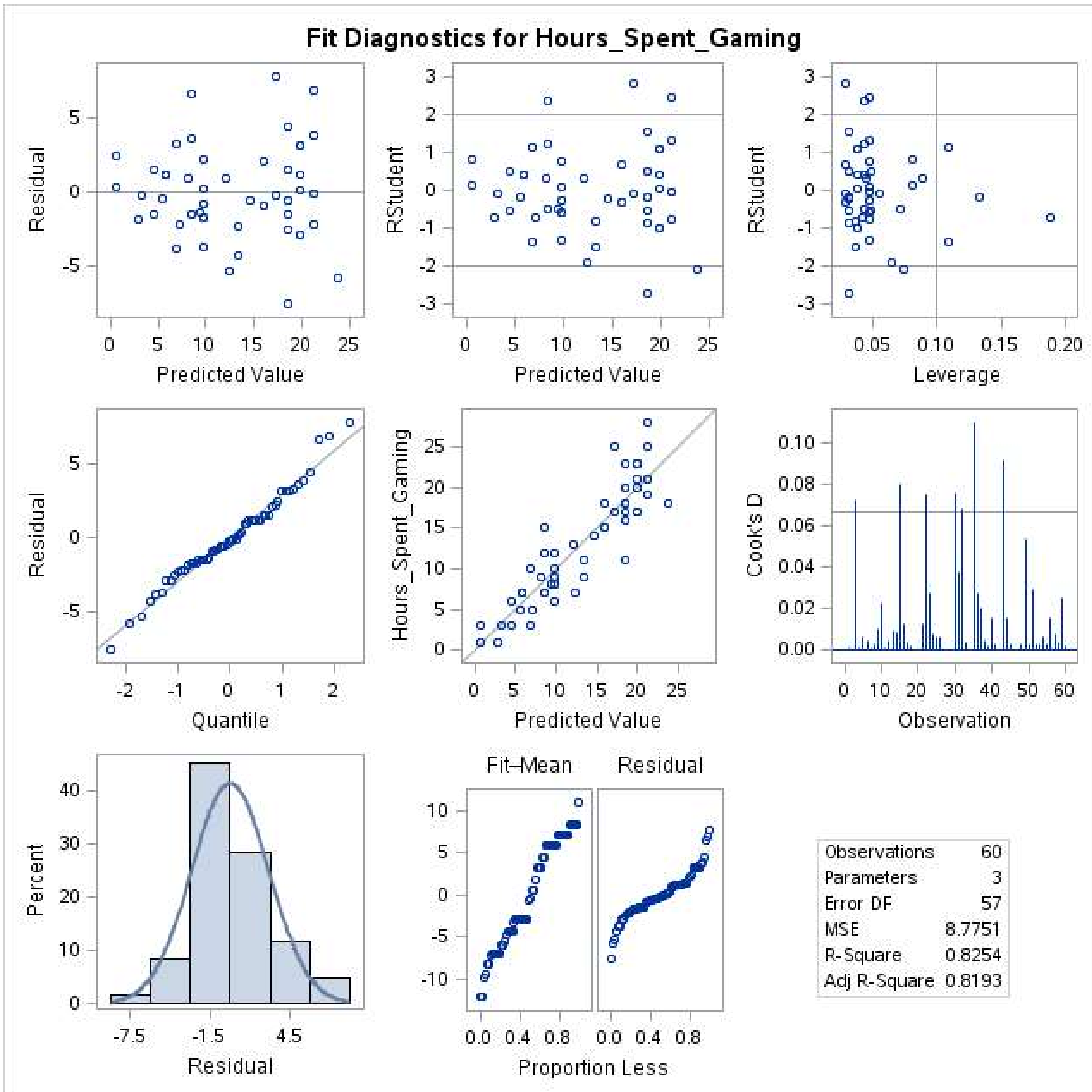
Analysis of Variance					
Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Pr > F**
Model	2	2364.55533	1182.27767	134.73	<.0001
Error	57	500.17800	8.77505		
Corrected Total	59	2864.73333			

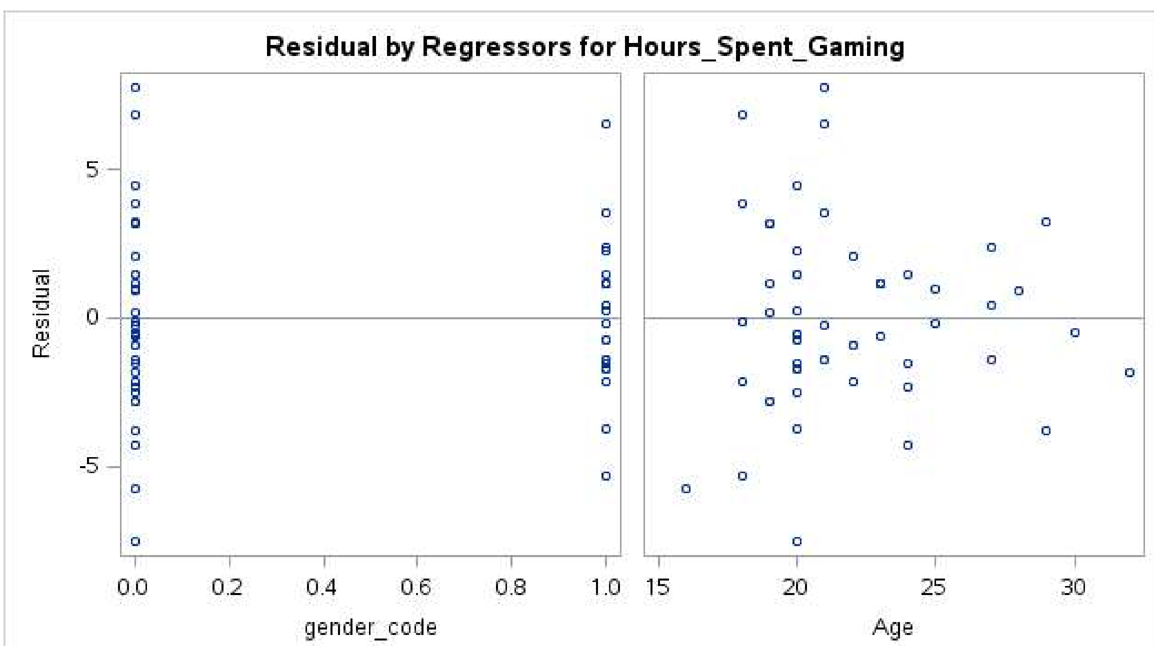
Root MSE	2.96227	**R-Square**	0.8254
Dependent Mean	12.76667	**Adj R-Sq**	0.8193
Coeff Var	23.20317		

Parameter Estimates					
Variable	**DF**	**Parameter Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	44.67299	2.51771	17.74	<.0001
gender_code	1	-8.79533	0.78141	-11.26	<.0001
Age	1	-1.30621	0.11424	-11.43	<.0001

Figure 4

The REG Procedure
Model: MODEL1
Dependent Variable: Hours_Spent_Gaming





The REG Procedure

Figure 4

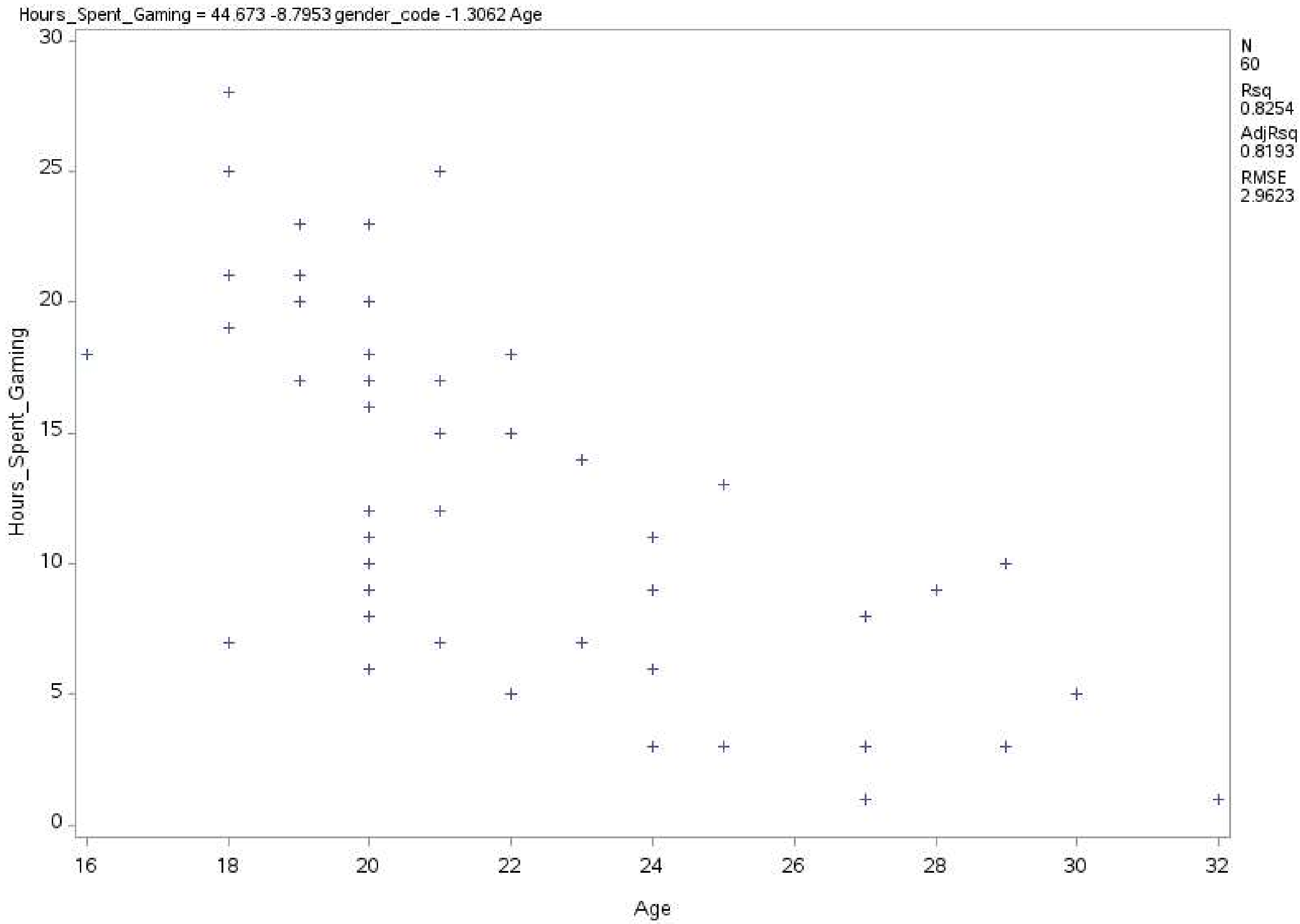


Figure 5

The REG Procedure
Model: MODEL1
Dependent Variable: Hours_Spent_Gaming

Number of Observations Read	60
Number of Observations Used	60

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	2166.43448	722.14483	57.91	<.0001
Error	56	698.29885	12.46962		
Corrected Total	59	2864.73333			

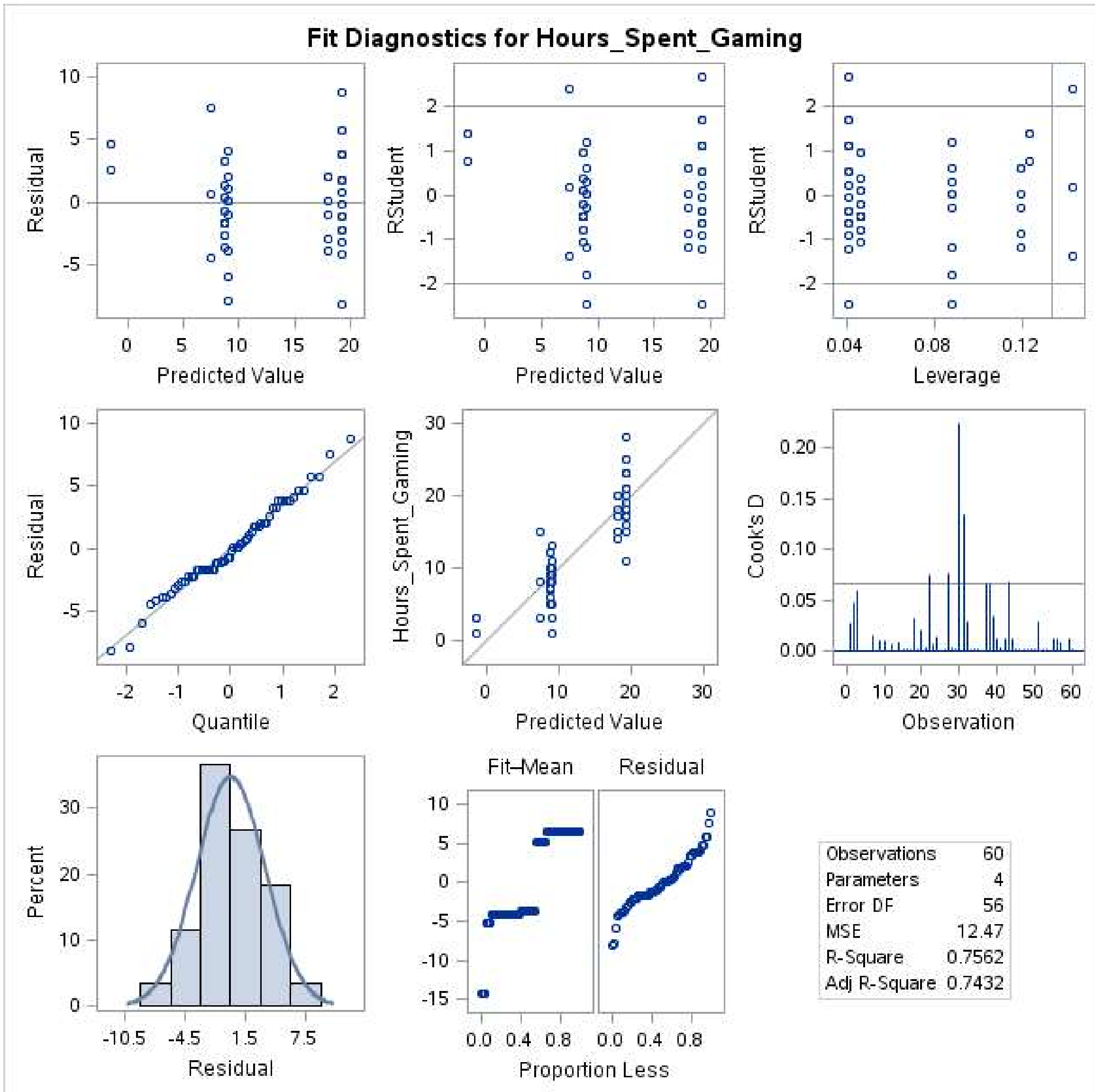
Root MSE	3.53124	**R-Square**	0.7562
Dependent Mean	12.76667	**Adj R-Sq**	0.7432
Coeff Var	27.65981		

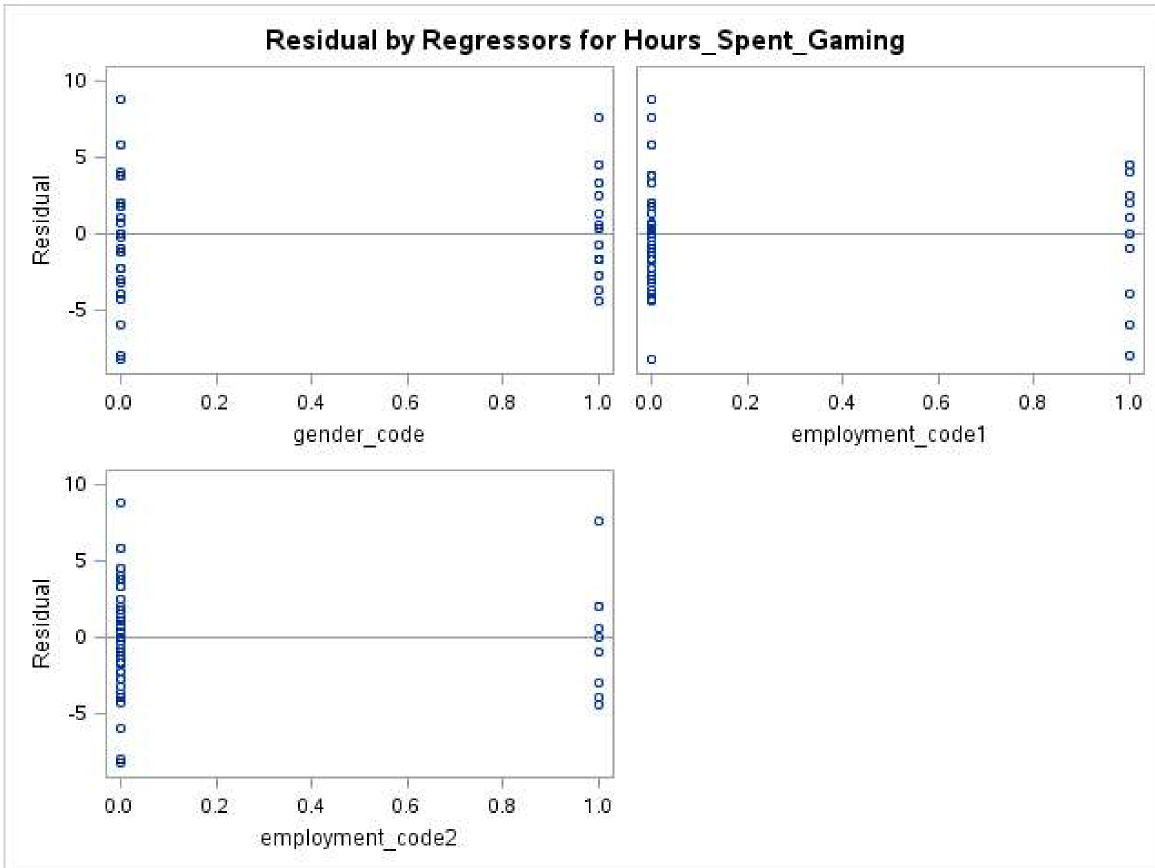
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	19.21839	0.71432	26.90	<.0001
gender_code	1	-10.52874	0.94571	-11.13	<.0001
employment_code1	1	-10.25287	1.18275	-8.67	<.0001
employment_code2	1	-1.26437	1.31147	-0.96	0.3391

Figure 5

The REG Procedure
Model: MODEL1
Dependent Variable: Hours_Spent_Gaming




Residual by Regressors for Hours_Spent_Gaming
Residual
gender_code
employment_code1
employment_code2
10
5
0
-5
0.0
0.2
0.4
0.6
0.8
1.0