Deliverable 8:

Testing the Centers of Body Fat Percentage as Predicted by Gender and Age

Rebekah Sander

Research Question

Does Age, Sex or combinations of Age and Sex Predict body fat percentage?

Research Variables

WHO	WHAT measurement is made on each subject?		TYPE OF MEASURE
	Name of Variable	Question Asked	
One Participant	gender	What is the sex of the participant?	Categorical Variable levels: M=Male, F=Female
	age	what age group does the participant belong to?	Categorical Variable levels: Under 25, 25 to 31, 32 to 47, 48 and older
	Body Fat %	What percentage of body fat does the participant have?	Quantitative Variable Unit: Percent
Two categorical variables predicting one quantitative variable.			
1. Two-way ANOVA on Data (Homogeneous and Normal) 2. Two-way ANOVA on Ranks (Heterogeneous OR Not Normal) 3. Transformations (To try making data normal)			

SAS Code: Examining the Data

```
/*Keeping what we want*/
data work.body_performance;
    set work.body_performance (keep = age gender 'body fat_%'n);
    rename 'body fat_%'n='Body Fat %'n;
run;
```

(1)

gender

gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
F	4926	36.78	4926	36.78
M	8467	63.22	13393	100.00

```
/* Check for and fix miscoding/missing values */
PROC FREQ DATA=WORK.body_performance;
TABLE age gender 'Body Fat %'n;
run;
```

(2)

Variables in Creation Order

#	Variable	Type	Len	Format	Informat	Label
1	age	Num	8	BEST.		age
2	gender	Char	1	$1.	$1.	gender
3	Body Fat %	Num	8	BEST.		body fat_%

```
Proc Contents data=work.body_performance varnum;
run;
```

(3)

```
Proc Means data = work.body_performance MAXDEC=2 n mean stddev median Qrange RANGE min Q1 Q3 max;
    var age 'Body Fat %'n; /*Quantitative variable*/
run;
```

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Median	Quartile Range	Range	Minimum	Lower Quartile	Upper Quartile	Maximum
age	age	13393	36.78	13.63	32.00	23.00	43.00	21.00	25.00	48.00	64.00
Body Fat %	body fat_%	13393	23.24	7.26	22.80	10.00	75.40	3.00	18.00	28.00	78.40

SAS Code: Examining the Data

```
/*Splitting age into quartiles*/
/*Getting the values*/
proc univariate data=work.body_performance;
    var age;
    output out=work.age_quartiles
    pctlpts = 25, 50, 75
    pctlpre = P_;
run;

title "Table 1: Age Quartiles";
proc print data=work.age_quartiles noobs;
run;
/*25, 32, 48*/
```

(1)

Table 1: Age Quartiles

P_25	P_50	P_75
25	32	48

```
data work.body_performance;
set work.body_performance;
length age_cat $20;
    if age <25 then age_cat='A';
    else if 25<= age <32 then age_cat='B';
    else if 32<= age <48 then age_cat='C';
    else if age >=48 then age_cat='D';
RUN;
```

```
Proc format;
Value $ageformat
'A'="Under 25"
'B'="25 to 31"
'C'="32 to 47"
'D'="48 and Older";
run;

data work.body_performance;
set work.body_performance;
format age_cat ageformat.;
drop age;
rename age_cat = age;
run;
```

SAS Code: Examining the Data

(1)
```
PROC FREQ data=work.body_performance;
tables age;
run;
```

(1)

Table 1: Age Quartiles

The FREQ Procedure

age	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Under 25	3038	22.68	3038	22.68
25 to 31	3465	25.87	6503	48.56
32 to 47	3457	25.81	9960	74.37
48 and Older	3433	25.63	13393	100.00

(2)
```
Proc Contents data=work.body_performance varnum;
run;
```

(2)

Variables in Creation Order

#	Variable	Type	Len	Format	Informat	Label
1	gender	Char	1	$1.	$1.	gender
2	Body Fat %	Num	8	BEST.		body fat_%
3	age	Char	20	$AGEFORMAT.		

(3)
```
Proc Means data = work.body_performance MAXDEC=2 n mean stddev median Qrange RANGE min Q1 Q3 max;
    var 'Body Fat %'n;
run;
```

(3)

The MEANS Procedure

Analysis Variable : Body Fat %% body fat_%%									
N	Mean	Std Dev	Median	Quartile Range	Range	Minimum	Lower Quartile	Upper Quartile	Maximum
13393	23.24	7.26	22.80	10.00	75.40	3.00	18.00	28.00	78.40

SAS Code: Assessing Normality and Homogeneity

```
/*Bivariate Analysis*/
PROC FREQ data=work.body_performance;
tables age*gender;
run;

/*normality and homogeneity*/
proc means data=work.body_performance;
class gender age;
var 'Body Fat %'n;
run;
```

(1)

(2)

The FREQ Procedure

Frequency Percent Row Pct Col Pct	Table of age by gender		
	gender(gender)		
age	F	M	Total
Under 25	1254 9.36 41.28 25.46	1784 13.32 58.72 21.07	3038 22.68
25 to 31	1048 7.82 30.25 21.27	2417 18.05 69.75 28.55	3465 25.87
32 to 47	1086 8.11 31.41 22.05	2371 17.70 68.59 28.00	3457 25.81
48 and Older	1538 11.48 44.80 31.22	1895 14.15 55.20 22.38	3433 25.63
Total	4926 36.78	8467 63.22	13393 100.00

(1)

The MEANS Procedure

Analysis Variable : Body Fat %% body fat_%%							
gender	age	N Obs	N	Mean	Std Dev	Minimum	Maximum
F	Under 25	1254	1254	27.3645055	5.5717396	12.2000000	48.2000000
	25 to 31	1048	1048	26.9477559	5.6601723	12.3000000	53.5000000
	32 to 47	1086	1086	27.7567383	6.4076617	9.4000000	50.2000000
	48 and Older	1538	1538	30.9637841	6.2354369	3.5000000	50.6000000
M	Under 25	1784	1784	18.4035875	6.1981017	3.0000000	78.4000000
	25 to 31	2417	2417	19.1878079	5.8473959	3.0000000	54.9000000
	32 to 47	2371	2371	20.8975006	5.7604142	5.8000000	46.2000000
	48 and Older	1895	1895	22.2565528	5.2820503	3.5000000	47.4000000

(2)

Assessing Normality

- H_0: The data came from a population where the body fat percentages are normally distributed
- H_A: The data came from a population where the body fat percentages are not normally distributed.
- $\alpha = 0.01$

The MEANS Procedure

Analysis Variable : Body Fat %% body fat_%%							
gender	age	N Obs	N	Mean	Std Dev	Minimum	Maximum
F	Under 25	1254	1254	27.3645055	5.5717396	12.2000000	48.2000000
	25 to 31	1048	1048	26.9477559	5.6601723	12.3000000	53.5000000
	32 to 47	1086	1086	27.7567383	6.4076617	9.4000000	50.2000000
	48 and Older	1538	1538	30.9637841	6.2354369	3.5000000	50.6000000
M	Under 25	1784	1784	18.4035875	6.1981017	3.0000000	78.4000000
	25 to 31	2417	2417	19.1878079	5.8473959	3.0000000	54.9000000
	32 to 47	2371	2371	20.8975006	5.7604142	5.8000000	46.2000000
	48 and Older	1895	1895	22.2565528	5.2820503	3.5000000	47.4000000

- The sample sizes for all categories are greater than 30 observations. Therefore, normality is supported.

Assessing Homogeneity

- H_0: $\sigma^2_{\text{F,Under 25}} = \sigma^2_{\text{F,25 to 31}} = \sigma^2_{\text{F,32 to 47}} = \sigma^2_{\text{F,48 and older}}$
  $= \sigma^2_{\text{M,Under 25}} = \sigma^2_{\text{M,25 to 31}} = \sigma^2_{\text{M,32 to 47}} = \sigma^2_{\text{M,48 and older}}$

H_A: At least one variance is different than the rest.

The MEANS Procedure

Analysis Variable : Body Fat %% body fat_%%							
gender	age	N Obs	N	Mean	Std Dev	Minimum	Maximum
F	Under 25	1254	1254	27.3645055	5.5717396	12.2000000	48.2000000
	25 to 31	1048	1048	26.9477559	5.6601723	12.3000000	53.5000000
	32 to 47	1086	1086	27.7567383	6.4076617	9.4000000	50.2000000
	48 and Older	1538	1538	30.9637841	6.2354369	3.5000000	50.6000000
M	Under 25	1784	1784	18.4035875	6.1981017	3.0000000	78.4000000
	25 to 31	2417	2417	19.1878079	5.8473959	3.0000000	54.9000000
	32 to 47	2371	2371	20.8975006	5.7604142	5.8000000	46.2000000
	48 and Older	1895	1895	22.2565528	5.2820503	3.5000000	47.4000000

Ratio of standard deviations:

$$= \frac{\text{SD}_{\text{F,32 to 47}}}{\text{SD}_{\text{M,48 and older}}}$$

$$= \frac{6.4077}{5.2821}$$

$$= 1.2131 < 2$$

Choosing Hypothesis Test: Two-Way ANOVA on Data

- Since the data is normal and homogeneous, we will perform the two-way ANOVA on Data.
- To perform the two-way ANOVA, we must
  1.) Perform the Global F Test
  2.) Test the Interaction
- The level of significance, $\alpha = 0.01$

SAS Code: Two-Way ANOVA on Data

```
/*Performing Two-Way ANOVA on data*/
proc GLM data=work.body_performance PLOTS(MAXPOINTS= none);
title1 "Table 1: Does a combination of participant gender and age group result in a larger body fat percentage?";
class gender age;
model 'Body Fat %'n= gender age gender*age;
run;
quit;
```

The GLM Procedure

Dependent Variable: Body Fat % body fat_%

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	245907.7260	35129.6751	1023.67	<.0001
Error	13385	459338.9121	34.3174		
Corrected Total	13392	705246.6381			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
gender	1	198702.5895	198702.5895	5790.13	<.0001
age	3	28469.1254	9489.7085	276.53	<.0001
gender*age	3	2089.8614	696.6205	20.30	<.0001

1. Performing the Global F Test

- The null hypothesis is,

  H_0: Gender, age group, and a combination of the two do not predict the average body fat percentage.

- The alternative hypothesis is,

  H_A: At least one of the terms, gender, age group, or a combination of the two, predicts the average body fat percentage.

- The level of significance, $\alpha = 0.01$, tells us that 1% of the time the analysis will conclude that at least one term in the model predicts body fat percentage when none of the terms predict body fat percentage.

1. Performing the Global F Test

- F Statistic: The variance for the model using gender, age group, and the interaction(35,129.7%^{2}) is 1023.67 times the within combination variance (34.3%^{2})
- P-value: There is a less than 0.01% chance of getting an F-value of 1023.67 or more when gender, age group, and the interaction do not predict the average body fat percentage.
- Conclusion: Since less than 0.01% is less than 1%, we reject H_0. We are 99% confident that at least one of these terms(gender, age group, or interaction) predicts average body fat percentage.

The GLM Procedure

Dependent Variable: Body Fat % body fat_%

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	7	245907.7280	35129.6751	1023.67	<.0001
Error	13385	459338.9121	34.3174		
Corrected Total	13392	705246.6381			

2. Testing the Interaction

- The null hypothesis is,

  H_0: There is no interaction between gender and age group.

- The alternative hypothesis is,

  H_A: There is interaction between gender and age group.

- The level of significance, $\alpha = 0.01$, tells us that 1% of the time the analysis will conclude that there is an interaction when there is not an interaction between gender and age group

2. Testing the Interaction

- F Statistic: The variance between the combinations of gender and age group ($696.6\%^2$) is 20.30 times the variance within the combinations ($34.3\%^2$)
- P-value: There is a less than 0.01% chance of getting an F-value of 20.30 or more when there is no interaction between gender and age group.
- Conclusion: Since less than 0.01% is less than 1%, we reject H_0. We are 99% confident that there is an interaction.

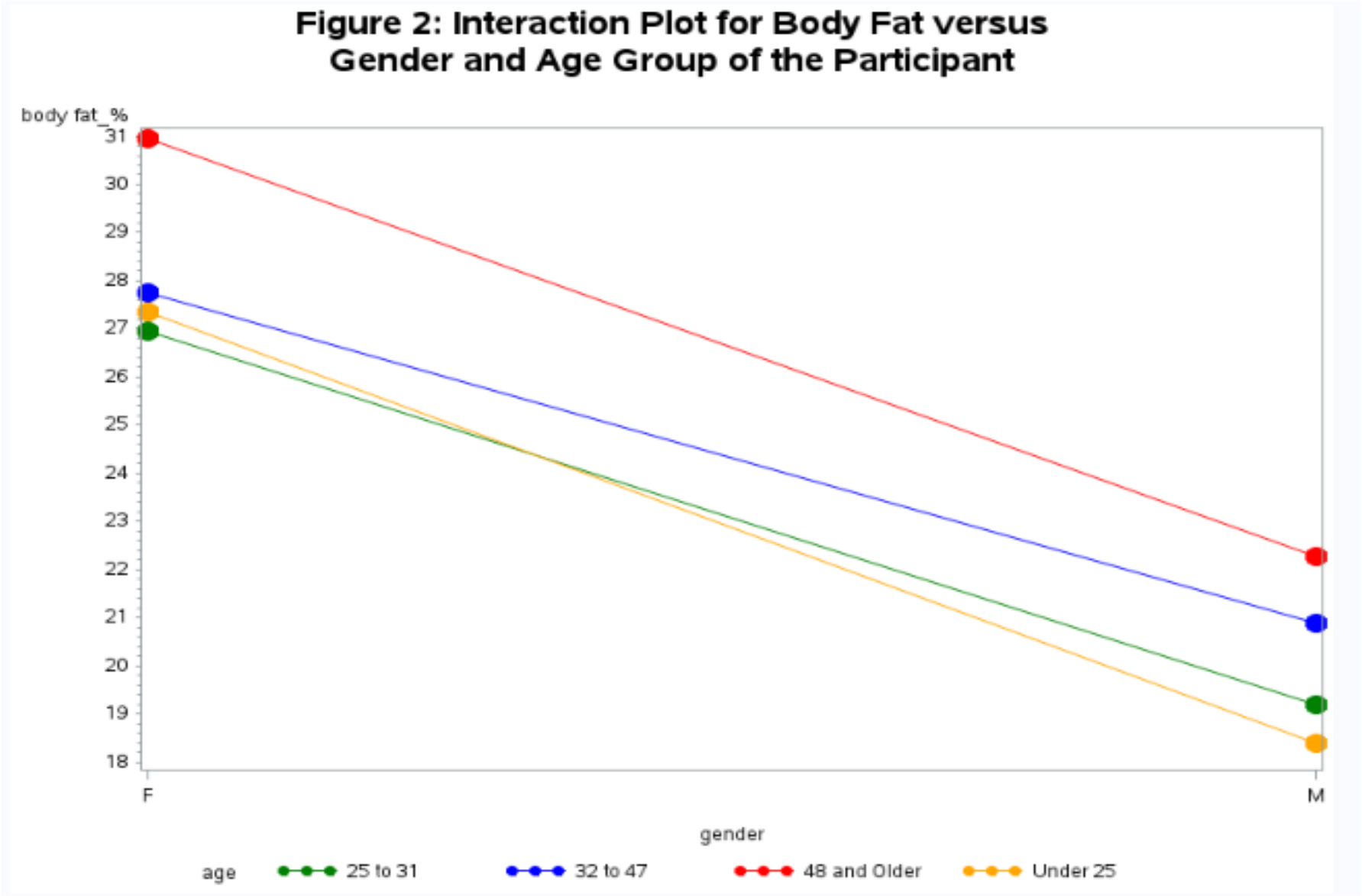
Source	DF	Type III SS	Mean Square	F Value	Pr > F
gender	1	198702.5895	198702.5895	5790.13	<.0001
age	3	28469.1254	9489.7085	276.53	<.0001
gender*age	3	2089.8614	696.6205	20.30	<.0001

SAS Code: Interaction Plot

```
/*Interaction Plot*/
proc summary data=work.body_performance NWAY;
class gender age;
var 'Body Fat %'n;
OUTPUT OUT=outfile MEAN = 'Body Fat'n;
run;

PROC GPLOT DATA = outfile;
PLOT 'Body Fat'n * gender = age;
SYMBOL1 V=dot H=2 I=join COLOR=green;
SYMBOL2 V=dot H=2 I=join COLOR=blue;
SYMBOL3 V=dot H=2 I=join COLOR=red;
SYMBOL4 V=dot H=2 I=join COLOR=orange;
TITLE1 H=2 'Figure 2: Interaction Plot for Body Fat versus';
TITLE2 H=2 'Gender and Age Group of the Participant';
run;
quit;
```

Interaction Plot



Taking Action

- This analysis provides useful insight into the health of individuals. A personal trainer may find this analysis useful as they work with a variety of clients. Knowing the age and sex of their client may help the understanding of what body percentage should be for them and will help guide the planning.

SAS Code: Screen Recording

SAS® Studio — SAS Programmer

SANDER Deliverable 8 SAS CODE.sas | SANDER Deliverable 9 SAS CODE.sas | marketing.ctl

CODE | LOG | RESULTS

```
/*Research Question: Does Age, Sex or combinations of Age and Sex Predict body fat percentage?
 Use alpha = 0.01 */
/*Unit of observation: One Participant*/
/*Categorical: gender; What is the sex of the participant?;
               levels= Male or Female*/
/*Quantitative: age; What is the age of the participant?; units: age
                body fat_%; What percentage of body fat does the participant have?; units: percent */


/* Determine hypothesis testing options for answering the question.*/
/*
1. Two-Way ANOVA on Data
2. Two-Way ANOVA on Ranks
3. Transformations
*/


/* Import the data set.*/
%web_drop_table(WORK.body_performance);

FILENAME REFFILE '/home/u62685438/sasuser.v94/stat3130/data/bodyPerformance.xlsx';

PROC IMPORT DATAFILE=REFFILE
    DBMS=XLSX
    OUT=WORK.body_performance;
```

/home/u62685438/sasuser.v94/stat3130/deliverables/SANDER Deliverable 8 SAS CODE.sas