BODY IMAGE WITH OF PATIENTS WITH CANCER

MODEL APPLIED: ONE-WAY ANOVA, TUKEY TEST,

KRUSKAL-WALLIS H-TEST, and WILCOXON RANK SUM TEST

STAT495 PROJECT

Professor: Dr. Olga

Chi Nguyen

4 May 2024

Table of Contents

1. INTRODUCTION

Cancer has become a leading global health concern in recent years, with approximately 19.3 million new cases recorded in 2020 and almost 10 million deaths within the same year.

Cancer impacts strongly not only the physical body of the patients but also the mental mindset of the patients and their families or friends. It creates physical changes and directly affects the satisfaction with patients' physical appearance which would influence the quality of life of the cancer patients.

2. BACKGROUND

The study's goal is to have more information on cancer's patients regarding their body images. It applies the German version of the Image Scale to analyze the sale image scale between cancer patients with the population and to determine the factors associate with the body image satisfaction of cancer patients.

The study was published on the PLOS ONE website, in November 2021, under the article "Body image in patients with different types of cancer", by Jan Brederecke, Anja Heise, and Taja Zimmermann.

3. DATA DESCRIPTION

The data was collected from 2002 to 2016, recruited cancer patients in both inpatient (from rehabilitation clinic Oberharz, municipal hospital Braunschweig and university hospital Munich) and outpatient settings (from oncology practices in Braunschweig). There were 536 cancer patients who participated in this survey, who were full age with German language skills and without recognizable mental disorders.

The data used for this report have been modified including,

- Computing the means of the indicator scale for the variables.
- Convert AGE variables into different AGE GROUP, which are below 50 years old (labeled as "Other"), 50s, 60s, and above 7-year-olds.
- Convert CANCER DURATION into different group, which are 1-3 years, 3-5 years, 5-10 years, and above 10 years.
- Convert RELATIONSHIP DURATION into different groups, which are below 3 years, 3-10 years, 10-49 years, and above 50 years.

- Ignore the data for the CANCER and CANER TYPE variables.

Some abbreviations apply within the data and report:

- SIS = Self-Image Scale (SIS) measure for body image satisfaction
- QMI = Relationship satisfaction - Quality of Marriage Index
- PHQ = Depression Scale
- PAF = Fear of Progression Questionnaire Short Form
- GAD = General Anxiety Disorder Questionnaire
- HADS= Hospital Anxiety and Depression Scale.

4. RESULTS

The data is examined with the One-Way ANOVA test, Tukey test, Kruskal-Wallis H-test, and Wilcoxon Rank Sum test.

o ONE-WAY ANOVA TEST

The data was examined with One-way ANOVA first to compare the means of different groups of 1 variable with one indicator German Scale, to identify if there are statistically significant differences between them. (One-way ANOVA test can help to the simultaneous comparison of all groups, reducing the overall Type I error rate.)

After running the comparisons for different variables groups, below is the table with some examples with p-values after the One-way ANOVA test, at a 5% significant level (alpha = 0.05)

P-VALUE < alpha	Age Group	Work Status	Education Years
SIS Self-Image Scale for body image satisfaction	0.0145	0.0242	0.0336
GAD General Anxiety Disorder	<0.0001	0.0001	0.3481
QMI Quality of Marriage Relationship Satisfaction	0.0435	0.0120	0.2048

For example:

1. The p-value between SIS and WORK STATUS, p-value= 0.0242 < 0.05, so we **reject *H*0** and can conclude that the mean of SIS is **not the same** among the different groups of WORK STATUS at the 5% significance level.

2. The p-value between QMI and EDUCATION YEARS with p-value= 0.3481 > 0.05, so we **cannot reject *H*0** and can conclude that the mean of QMI is the **same** among the different groups of EDUCATION YEARS at the 5% significance level.

To examine further with Tukey and Kruskal Wallis H-test, between WORK STATUS and SIS mean

o TUKEY TEST

Tukey's HSD test is a post hoc test used to determine which specific groups differ from each other when the one-way ANOVA indicates that there are significant differences between at least two groups, to identify where the differences lie.

The result for the Tukey test between WORK STATUS and SIS mean is:

- There is a statistically significant **difference** between the "Employed" and "Retired" groups (the confidence interval (of 0.01671 to 0.36091) does not include zero.
- There are **no** statistically significant differences between other pairs of WORK STATUS groups.

Employed Other Unemployed Retired

Comparisons significant at the 0.05 level are indicated by ***.				
work_status Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
employed - other	0.01194	-0.35138	0.37526	
employed - unemployed	0.17964	-0.06918	0.42847	
employed - retired	0.18881	0.01671	0.36091	***
other - employed	-0.01194	-0.37526	0.35138	
other - unemployed	0.16771	-0.24289	0.57830	
other - retired	0.17687	-0.19230	0.54605	
unemployed - employed	-0.17964	-0.42847	0.06918	
unemployed - other	-0.16771	-0.57830	0.24289	
unemployed - retired	0.00917	-0.24814	0.26647	
retired - employed	-0.18881	-0.36091	-0.01671	***
retired - other	-0.17687	-0.54605	0.19230	
retired - unemployed	-0.00917	-0.26647	0.24814	

o KRUSKAL WALLIS H-TEST & WILCOXON TEST

The Kruskal-Wallis H-test can be used to determine if there are statistically significant differences between three or more groups based on ranks rather than means.

The result for the Kruskal Wallis H-test between WORK STATUS and SIS mean is:

At a 5% significant level, "Employed" & "Retired" have a **different mean** SIS, whereas "Other" with other 3 groups "Unemployed", "Employed" and "Retired", and "Unemployed" & "Retired" have the **same mean** SIS.

Employed Other Unemployed Retired

Work Status pairs	**P-value vs α = 0.05**
Employed – Unemployed	0.0908 > alpha
Employed - Retired	0.0049 < alpha
Employed - Other	0.7640 > alpha
Unemployed - Retired	0.7823 > alpha
Unemployed - Other	0.4481 > alpha
Retired - Other	0.2664 > alpha

5. CODE

o SAS CODE

```
proc import out=mydata
datafile="//vdi-fileshare02/UEMprofiles/028631185/Desktop/Stat
495/Cancee.csv"
dbms=csv replace;
data mydata;
    set mydata;
    sis_mean = mean(of sis1 sis2 sis3 sis4 sis5 sis6 sis7 sis8
sis9   sis10    sis11);
    paf_mean = mean(of paf1 paf2 paf3 paf4 paf5 paf6 paf7 paf8
paf9   paf10    paf11    paf12);
  qsc_mean = mean(of qsc1  qsc2 qsc3 qsc4 qsc5 qsc6 qsc7 qsc8
qsc9   qsc10    qsc11    qsc12    qsc13    qsc14
```

```
qsc15   qsc16     qsc17     qsc18     qsc19     qsc20
qsc21   qsc22     qsc23);
   qmi_mean = mean(of qmi1  qmi2 qmi3 qmi4 qmi5 qmi6);
   hads_mean = mean(of hads1     hads2     hads3     hads4
hads5   hads6     hads7     hads8     hads9     hads10
hads11  hads12    hads13    hads14);
   phq_mean = mean(of phq1  phq2 phq3 phq4 phq5 phq6 phq7 phq8
phq9);
   gad_mean = mean(of gad1  gad2 gad3 gad4 gad5 gad6 gad7);
select;
        when (age >= 50 and age < 60) age_group = '50s';
        when (age >= 60 and age < 70) age_group = '60s';
        when (age >= 70) age_group = '70s';
        otherwise age_group = 'Other';
    end;
select;
        when (cancer_duration_months < 12)
cancer_duration_group = '> 1 year';
        when (cancer_duration_months >= 12 and
cancer_duration_months < 36) cancer_duration_group = '1-3
years';
        when (cancer_duration_months >= 36 and
cancer_duration_months < 60) cancer_duration_group = '3-5
years';
        when (cancer_duration_months >= 60 and
cancer_duration_months < 120) cancer_duration_group = '5-10
years';
        otherwise cancer_duration_group = 'More than 10
years';
    end;
select;
        when (relationship_duration_years < 3)
relationship_duration_group = 'Below 3 years';
        when (relationship_duration_years >= 3 and
relationship_duration_years < 10) relationship_duration_group
= '3-10 years';
        when (relationship_duration_years >= 10 and
relationship_duration_years < 50) relationship_duration_group
= '10-49 years';
        otherwise relationship_duration_group = 'More than 50
years';
```

```
    end;
run;
/*ONE-WAY ANOVA TEST*/
proc anova;
class work_status ;
model sis_mean=work_status;
/*TUKEY TEST*/
means work_status/tukey;
run;
/*KRUSKAL WALLIS H-TEST*/
proc npar1way data=mydata;
  class work_status;
  var sis_mean;
  where work_status in ("employed","unemployed");
run;
```

- **R CODE:**

```
install.packages("readr")
install.packages("car")
install.packages("dplyr")
library(dplyr)
library(car)
library(readr)
# Read the CSV file
mydata <-
read_csv("//vdi-fileshare02/UEMprofiles/028631185/Desktop/Stat495/Cancee.csv")

# Check summary statistics of age variable
summary(mydata$age)

# Calculate means for each set of variables
mydata$sis_mean <- rowMeans(mydata[, c("sis1", "sis2", "sis3", "sis4", "sis5", "sis6", "sis7", "sis8", "sis9", "sis10", "sis11")])
```

```
mydata$paf_mean <- rowMeans(mydata[, c("paf1", "paf2", "paf3",
"paf4", "paf5", "paf6", "paf7", "paf8", "paf9", "paf10",
"paf11", "paf12")])
mydata$qsc_mean <- rowMeans(mydata[, c("qsc1", "qsc2", "qsc3",
"qsc4", "qsc5", "qsc6", "qsc7", "qsc8", "qsc9", "qsc10",
"qsc11", "qsc12", "qsc13", "qsc14", "qsc15", "qsc16", "qsc17",
"qsc18", "qsc19", "qsc20", "qsc21", "qsc22", "qsc23")])
mydata$qmi_mean <- rowMeans(mydata[, c("qmi1", "qmi2", "qmi3",
"qmi4", "qmi5", "qmi6")])
mydata$hads_mean <- rowMeans(mydata[, c("hads1", "hads2",
"hads3", "hads4", "hads5", "hads6", "hads7", "hads8", "hads9",
"hads10", "hads11", "hads12", "hads13", "hads14")])
mydata$phq_mean <- rowMeans(mydata[, c("phq1", "phq2", "phq3",
"phq4", "phq5", "phq6", "phq7", "phq8", "phq9")])
mydata$gad_mean <- rowMeans(mydata[, c("gad1", "gad2", "gad3",
"gad4", "gad5", "gad6", "gad7")])

# Create age group variable
mydata$age_group <- cut(mydata$age, breaks = c(0, 50, 60, 70,
max(mydata$age)), include.lowest = TRUE)

# Create labels for the age group variable
mydata$age_group <- factor(mydata$age_group, labels = c("0-49",
"50s", "60s", "70s"))

# Create cancer duration group variable
mydata$cancer_duration_group <-
cut(mydata$cancer_duration_months, breaks = c(0, 12, 36, 60,
120, Inf), labels = c("> 1 year", "1-3 years", "3-5 years",
"5-10 years", "More than 10 years"), include.lowest = TRUE)

# Create relationship duration group variable
```

```
mydata$relationship_duration_group <-
cut(mydata$relationship_duration_years, breaks = c(0, 3, 10, 50,
Inf), labels = c("Below 3 years", "3-10 years", "10-49 years",
"More than 50 years"), include.lowest = TRUE)

# Perform ANOVA
anova_result <- aov(sis_mean ~ work_status, data = mydata)
tukey_result <- TukeyHSD(anova_result)

# Perform Kruskal-Wallis test
kruskal_result <- kruskal.test(sis_mean ~ work_status, data =
filter(mydata, work_status %in% c("employed", "unemployed")))

filtered_data <- mydata[mydata$work_status %in% c("employed",
"unemployed"), ]

# Perform the Wilcoxon test
wilcox_result <- wilcox.test(sis_mean ~ work_status, data =
filtered_data)

# Display results
summary(anova_result)
print(tukey_result)
print(kruskal_result)
print(wilcox_result)
```

NOTE: For each test, One-way ANOVA, Tukey, or Kruskal Wallis H-test and Wilcoxon Rank Sum test, the variables are substituted repeatedly to have the p-value required to compare with alpha at a 5% siginificant level.

6. APPENDIX

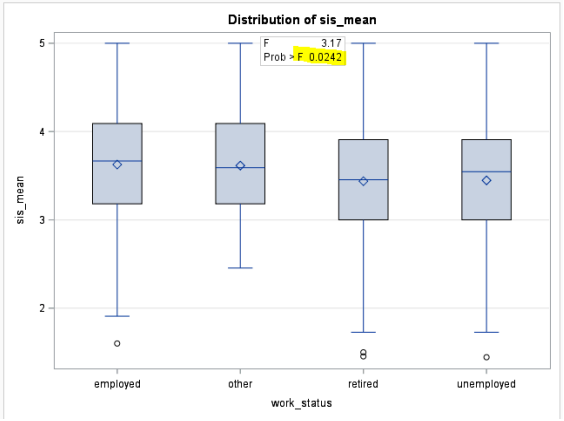
i. DATA SET

	education_ye	cancer_dura	age	sex	family_status	relationship_	relationship_	work_status	dataset	cancer_grou	cancer	sis1	sis2	sis3	sis4	sis5	sis6	sis7	sis8	sis9	sis10	si
2	less_nine	13	58	1	in_a_relation	1	38	unemployed	1	other	bone_marrov	1	3	3	2	1	1	1	1	1	3	
3	less_nine	3	56	1	divorced	0		employed	1	prostate_anc	prostate	1	5		4	2				2		
4	ten	16	58	1	in_a_relation	1	36	employed	1	visceral	colon	4	4	5	5	4	4	5	5	4	5	
5	ten	14	67	1	in_a_relation	1	41	retired	1	prostate_anc	prostate	3	4	4	4	3	3	2	3	3	4	
6	more_ten	9	61	1	in_a_relation	1	35	employed	1	visceral	gallbladder	5	5	5	5	3	5	4	4	5	5	
7	less_nine	10	62	1	in_a_relation	1		employed	1	other	bladder	4	4	4	5	4	3	3	4	5	4	
8	ten	5	64	1	in_a_relation	1		retired	1	visceral	esophagus	2	2	4	4	2	1	1	2	4	2	
9	more_ten	19	62	1	in_a_relation	1	38	employed	1	prostate_anc	prostate	3	3	4	5	3	4	4	4	4	4	
10	less_nine	69	58	1	in_a_relation	1	30	employed	1	prostate_anc	prostate	2	4	4	5	3	2	2	2	4	2	
11	less_nine	81	79	1	in_a_relation	1	60	retired	1	prostate_anc	prostate	4	4	4	5	4		3	3	3	3	
12	less_nine	12	61	1	divorced	1	33	employed	1	visceral	esophagus	3	5	5	5	4	5	4	4	5	5	
13	less_nine	20	54	1	in_a_relation	1	26	employed	1	other	kidney	2	3	4	5	3	4	3	4	4	4	
14	less_nine	4	23	1	single	1		employed	1	prostate_anc	testicle	4	4	1	5	4	4	4	4	5	4	
15	less_nine	20	44	1	in_a_relation	1	24	employed	1	other	leukemia	3	4	4	4	2	3	2	3	4	3	
16	ten	91	55	1	in_a_relation	1	36	employed	1	other	lymphatic_gl	5	5	5	5	3	4	4	4	5	4	
17	less_nine	87	60	1	in_a_relation	1	44	employed	1	other	leukemia	1				4	5	4			5	
18	more_ten	30	69	1	in_a_relation	1	30	retired	1	visceral	colon	3	2	4	5	3	4	4	4	4	4	
19	ten	5	30	1	single	0		employed	1	prostate_anc	testicle	4	4		5	4		4		5		
20	ten	1	40	1	in_a_relation	1	10	employed	1	other	kidney	3	4	4	5	4	4	3	4	5	4	
21	less_nine	8	46	1	divorced	1	1.5	unemployed	1	visceral	esophagus	3	3	4	3	3	4	3	5	3	5	
22	ten	12	57	1	in_a_relation	1		employed	1	prostate_anc	prostate	4	4	4	5	3	1	3	3	5	4	
23	less_nine	2	66	1	in_a_relation	1	46	retired	1	prostate_anc	prostate	5	5	5	5	4	5	4	5	5	5	
24	more_ten	3	72	1	in_a_relation	1	49	retired	1	other	leukemia	3	3	4	4	3	3	3	3	4	4	
25	less_nine	58	52	1	single	1		employed	1	visceral	esophagus	3	3	5	3	3	5	3	5	3	5	
26	less_nine	2	57	1	in_a_relation	1	30	employed	1	other	kidney	4	4	4	2	4	4	4	4	5	4	
27	ten	8	53	1	in_a_relation	1	12	employed	1			4	4	4	4	3	4	3	4	5	4	
28	ten	4	46	1	in_a_relation	1	27	employed	1	other	skin	4	4	4	4	3	4	4	4	4	4	
29	ten	7	48	1	in_a_relation	1	28	employed	1	visceral	stomach	2	4	4	4	3	5	4	4	5	4	
30	more_ten	35	23	1	single	1	5	employed	1	prostate_anc	testicle	3	3	4	4	3	4	3	4	4	3	
31	ten	11	41	1	in_a_relation	1	17	employed	1	visceral	colon	1	1	3	5	3	1	2	1	2	3	
32	ten	10	60	1	in_a_relation	1	39	employed	1	prostate_anc	prostate	3	2	5	5	2	2	3	4	4	5	

ii. EXAMPLES OF THE RESULTS OF 3 TESTS

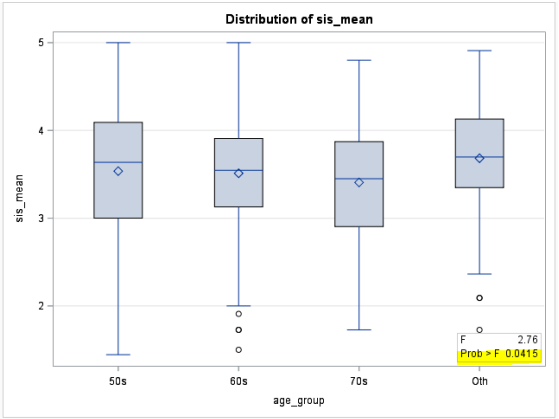
a) ONE-WAY ANOVA

SIS with Work Status



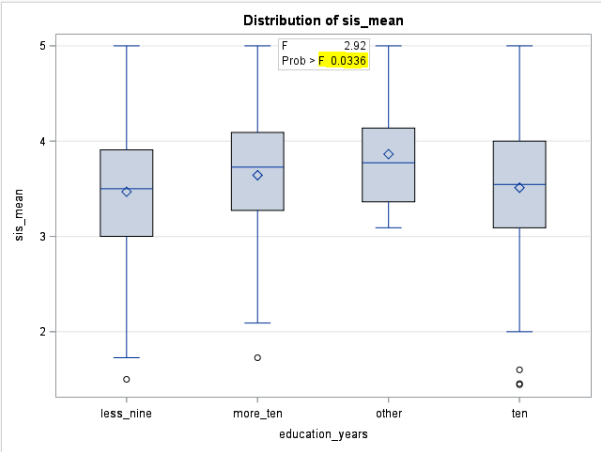

```
> summary(anova_result)
             Df Sum Sq Mean Sq F value Pr(>F)
work_status   3    4.4  1.4665   3.291 0.0206 *
Residuals   414  184.5  0.4455
```

SIS with Age Group



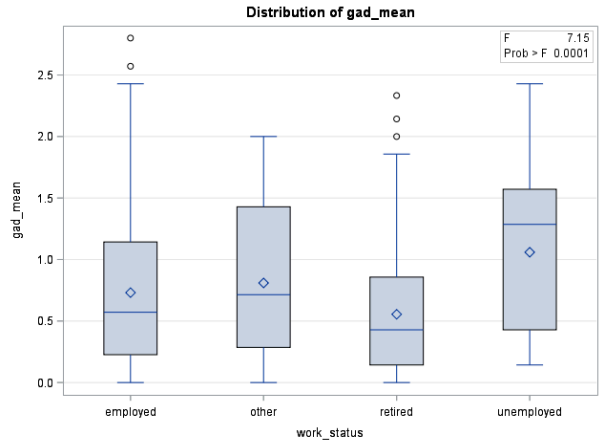

```
             Df Sum Sq Mean Sq F value Pr(>F)
age_group     3   2.37  0.7901   1.762  0.154
Residuals   418 187.46  0.4485
109 observations deleted due to missingness
```

SIS with Education Year



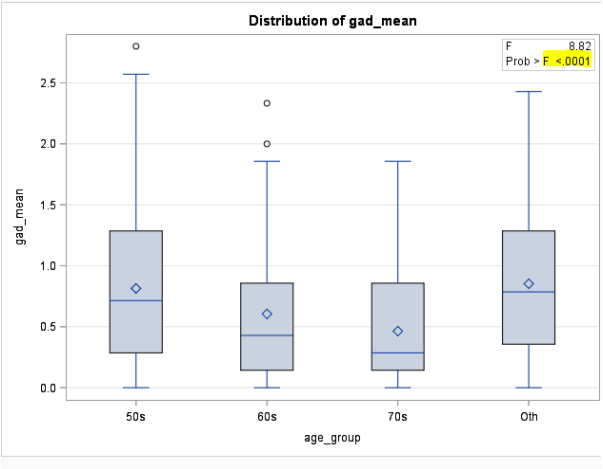

```
> summary(anova_result)
                  Df Sum Sq Mean Sq F value Pr(>F)
education_years    3   2.88  0.9609   2.147 0.0937 .
Residuals        417 186.61  0.4475
```

GAD with Work Status



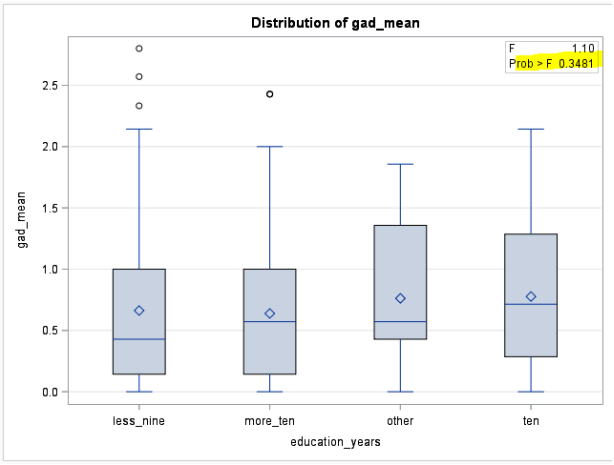

```
> summary(anova_result)
             Df Sum Sq Mean Sq F value   Pr(>F)
work_status   3   8.05  2.6819    8.31 2.28e-05 ***
Residuals   385 124.26  0.3227
---
```

GAD with Age Group



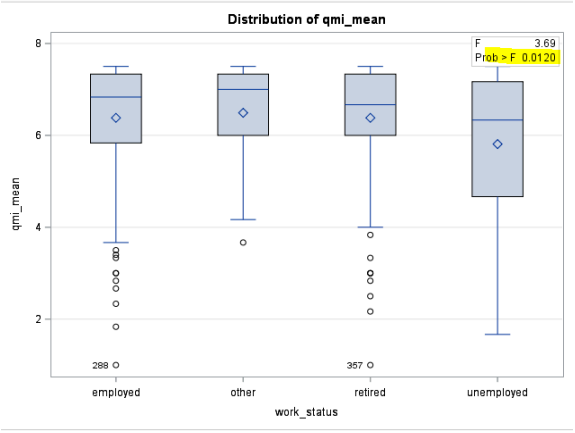

```
> summary(anova_result)
             Df Sum Sq Mean Sq F value   Pr(>F)
age_group     3  11.41   3.804   12.14 1.31e-07 ***
Residuals   388 121.60   0.313
---
```

GAD with Education Years



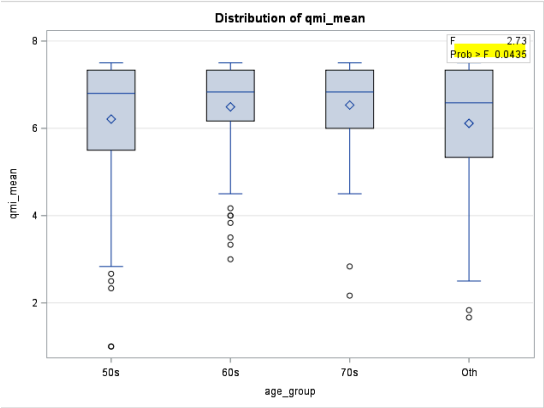

```
> summary(anova_result)
                Df Sum Sq Mean Sq F value Pr(>F)
education_years  3   1.45  0.4849   1.426  0.235
Residuals      387 131.56  0.3399
140 observations deleted due to missingness
```

QMI with Work Status



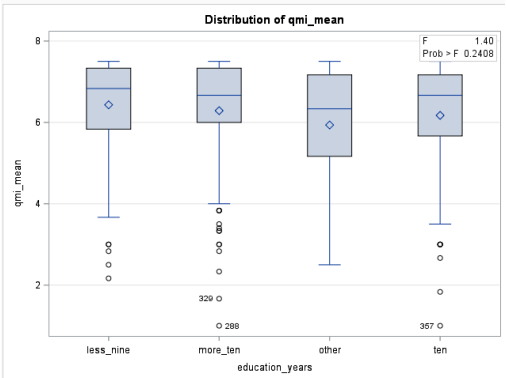

```
> summary(anova_result)
             Df Sum Sq Mean Sq F value Pr(>F)
work_status   3   15.7   5.229   3.187 0.0236 *
Residuals   451  740.0   1.641
---
```

QMI with Age Group




```
> summary(anova_result)
             Df Sum Sq Mean Sq F value Pr(>F)
age_group     3   11.2   3.732   2.273 0.0794 .
Residuals   455  747.1   1.642
```

QMI with Education Years




```
> summary(anova_result)
                Df Sum Sq Mean Sq F value Pr(>F)
education_years  3    6.0   2.012   1.215  0.304
Residuals      454  751.8   1.656
```

b) TUKEY TEST (WORK STATUS versus SIS MEAN)

Comparisons significant at the 0.05 level are indicated by ***.

work_status Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
employed - other	0.01194	-0.35138	0.37526	
employed - unemployed	0.17964	-0.06918	0.42847	
employed - retired	0.18881	0.01671	0.36091	***
other - employed	-0.01194	-0.37526	0.35138	
other - unemployed	0.16771	-0.24289	0.57830	
other - retired	0.17687	-0.19230	0.54605	
unemployed - employed	-0.17964	-0.42847	0.06918	
unemployed - other	-0.16771	-0.57830	0.24289	
unemployed - retired	0.00917	-0.24814	0.26647	
retired - employed	-0.18881	-0.36091	-0.01671	***
retired - other	-0.17687	-0.54605	0.19230	
retired - unemployed	-0.00917	-0.26647	0.24814	

```
Fit: aov(formula = sis_mean ~ work_status, data = mydata)

$work_status
                          diff        lwr         upr     p adj
other-employed      -0.04112230 -0.4199332  0.33768856 0.9923292
retired-employed    -0.21714338 -0.4071067 -0.02718005 0.0176835
unemployed-employed -0.18582293 -0.4494572  0.07781131 0.2661943
retired-other       -0.17602108 -0.5638048  0.21176268 0.6455964
unemployed-other    -0.14470063 -0.5734117  0.28401045 0.8200371
unemployed-retired   0.03132045 -0.2450517  0.30769263 0.9912958
```

c) KRUSKAL-WALLIS H-TEST (WORK STATUS versus SIS MEAN)

The NPAR1WAY Procedure

Wilcoxon Scores (Rank Sums) for Variable sis_mean Classified by Variable work_status

work_status	N	Sum of Scores	Expected Under H0	Std Dev Under H0	Mean Score
unemployed	63	15301.50	16317.0	1110.54840	242.880952
employed	245	68197.50	63455.0	1695.14473	278.357143
retired	183	43379.00	47397.0	1623.44682	237.043716
other	26	7025.00	6734.0	741.93645	270.192308

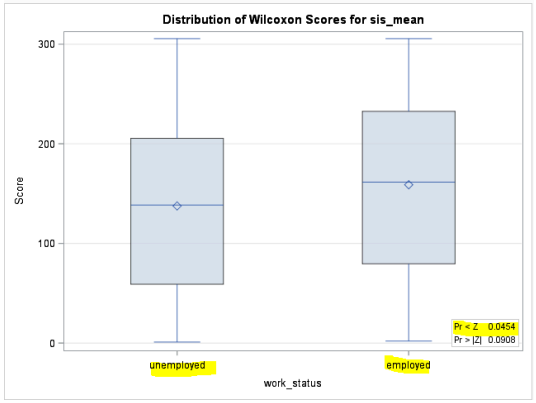
Average scores were used for ties.

Kruskal-Wallis Test	
Chi-Square	8.9556
DF	3
Pr > Chi-Square	0.0299

```
        Kruskal-Wallis rank sum test

data:  sis_mean by work_status
Kruskal-Wallis chi-squared = 9.5575, df = 3, p-value = 0.02273
```

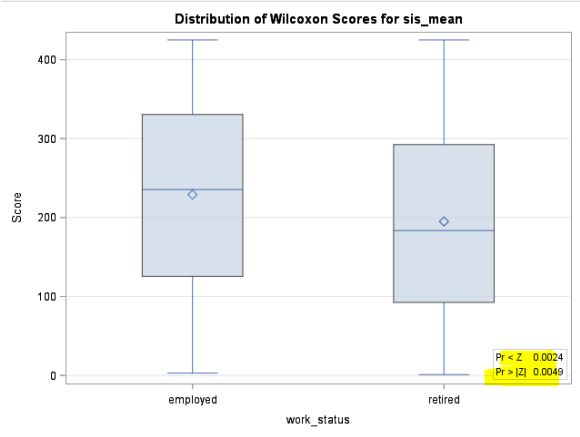
Employed – Unemployed



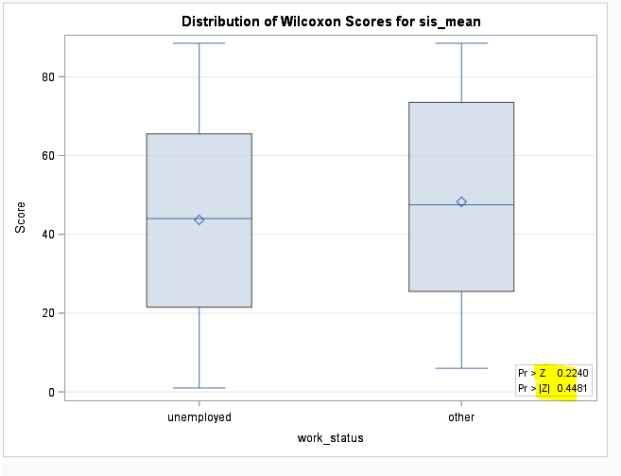
Employed – Other

Chi-Square	0.0909
DF	1
Pr > Chi-Square	0.7630

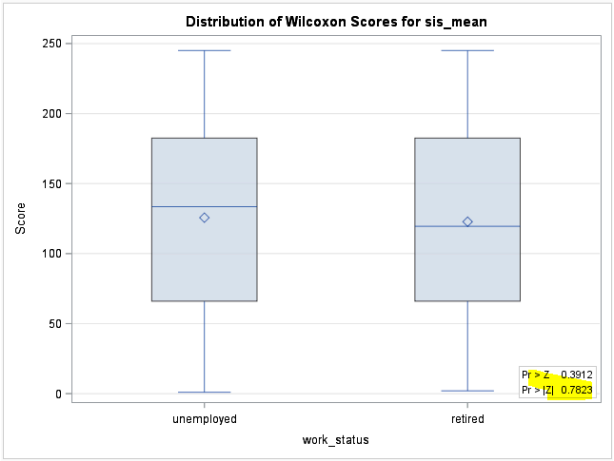
Employed – Retired



Unemployed – Other



Unemployed – Retired



Other – Retired

7. REFFERENCE

Brederecke, J., Heise, A., & Zimmermann, T. (2021). Body image in patients with different types of cancer. *PLOS ONE*, 16(11), e0260602. https://doi.org/10.1371/journal.pone.026060