Body image in Patients with Cancers

Model applied: One-way ANOVA, Turkey Test, Kruskal-Wallis H-test, Wilcoxon test

Chi Nguyen

Background

With approximately 19.3 millions new cancer diagnoses and an estimated 9.9 millions of deaths from cancer in 2020, cancer is currently a leading to the global health concern.

Cancer can cause physical changes and affect the satisfaction with patient's physical appearance, which would strongly impact the quality of life in general.

Study compares **satisfaction** with **body image** of patients with different types of cancer with the general population and across sexes and identifies risk factors for diminished body image.

Dataset

- acquired from PLOS ONE, published in 2021.
- dataset: n= 531, using German Image Scale, with patients voluntarily recruited from 2002 to 2016

Data adjustment:

Compute the **means** for the indicator variables.

Convert Age, Cancer duration, Relationship duration

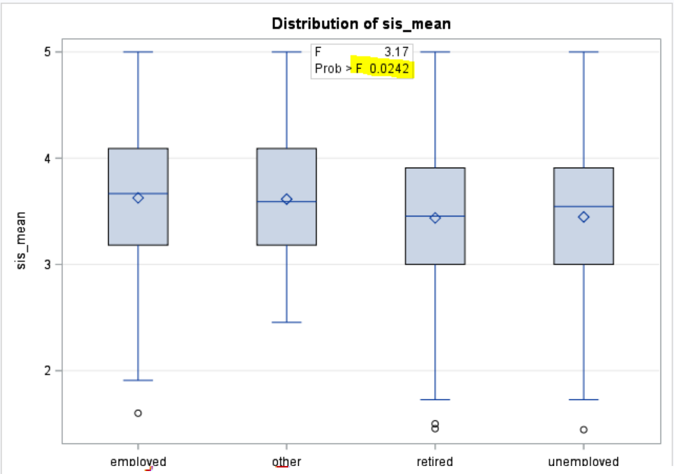
Delete value of Cancer and Cancer Type Variables

Variables:

Age group. Sex, Family Status. Relationship duration group. Cancer duration group. Work Status, Education Year
& German Scale (means) to for measurement, including SIS (self image scale for body image satisfaction), GAD (General Anxiety Disorder), PHQ (Depression Scale), etc.
This project focus on compare different groups of **Work Status** with the **SIS Mean** (the self image scale for body image satisfaction) of patients with cancers.

1	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	

One-Way ANOVA & Turkey Test




```
> 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
```

One-way ANOVA

Conclusion :

At 5% significant level, p-value < alpha, there are significant differences between the Work Status groups being compared with the mean of SIS

<u>Turkey Test:</u>

There is a statistically **significant difference** between the **"Employed"** and **"Retired"** group (the confidence interval (0.01671 to 0.36091) does not include zero)

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	

```
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
```

Kruskal Wallis H-test & Wilcoxon Rank Sum Test

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
```

Work group	P-value vs $\alpha = 0.05$
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

Kruskal-Wallis H-test

P-value = 0.0299 < alpha

At 5% significant level, reject H0, and can conclude that all for groups of WORK STATUS (Employed, Unemployed, Retired and Other) have a **different** SIS mean.

Wilcoxon Rank Sum Test

At 5% significant level, "Employed" & "Retired" have a **different mean** SIS, whereas the rest of compared pairs have the **same** mean SIS.

SAS code

```
proc import out=mydata datafile="//vdi-fileshare02/UEMprofiles/028631185/Desktop/Stat495/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
    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;
```

```
proc anova;
class work_status ;
model sis_mean=work_status;
/*means work_status/tukey;*/
run;
```

ANOVA & Turkey Test

```
/*KRUSKAL WALLIS H-TEST*/
proc npar1way data=mydata Wilcoxon;
  class work_status;
  var sis_mean;
  exact;
  where work_status in ("employed","unemployed");
run;
```

Kruskal-Wallis H-test & Wilcoxon test

R code

```r

# 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$qsc_mean <- rowMeans(mydata[, c("qsc1", "qsc2", "qsc3", "qsc4", "qsc5", "qsc6", "qsc7", "qsc8", "qsc9", "qsc10", "qsc11", "qsc12", "qsc13", "qsc14", "qsc15", "q
mydata$qmi_mean <- rowMeans(mydata[, c("qmi1", "qmi2", "qmi3", "qmi4", "qmi5", "qmi6")])
mydata$hads_mean <- rowMeans(mydata[, c("hads1", "hads2", "hads3", "hads4", "hads5", "hads6", "hads7", "hads8", "hads9", "hads10", "hads11", "hads12", "hads13", "hads1
mydata$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",

# 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", "

```

```r
# 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")))

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

Reference

- 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.0260602

THANK YOU

DR.OLGA & STAT495 CLASSMATES