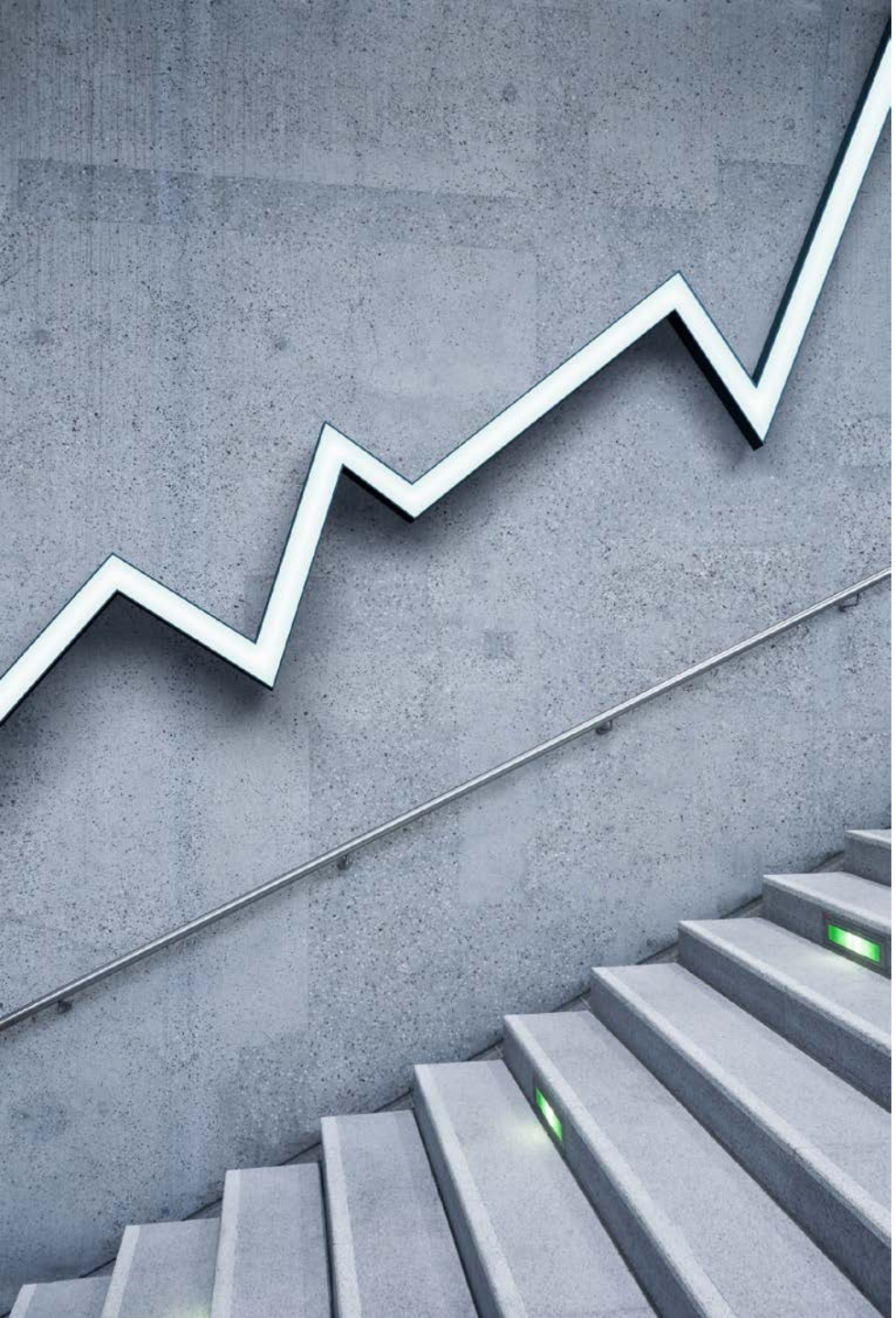


Design and Analysis of Experiments for Cancer Patients:

Esophageal Cancer in Different Lifestyle Groups

KASIA KRUEGER

MAT 4720

SUMMER 2021

Introduction

Data from a case-control study of esophageal cancer in Ille-et-Vilaine, France

Data set in R (esoph)

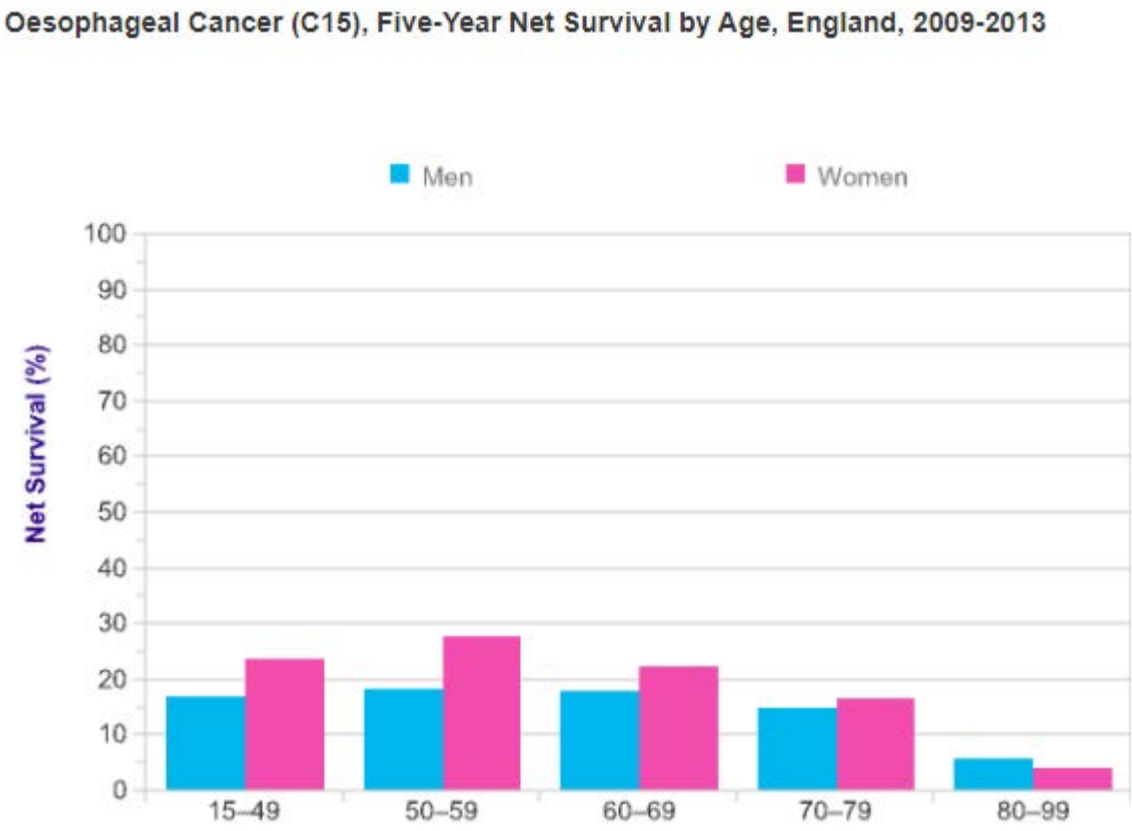
4x4x6 Factorial Experiment

- 3 Levels: Daily Alcohol Use, Daily Tobacco Use, and Age group
- 88 total records
- 200 cancer cases
- 975 controls (no cases)

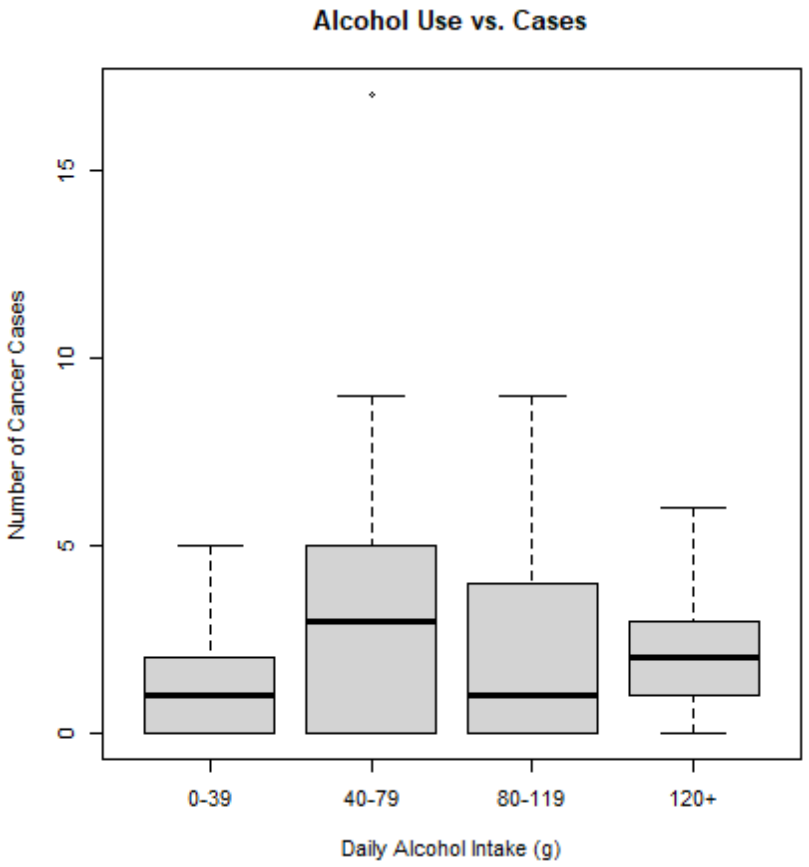
Response variable is number of esophageal (throat) cancer cases based on 3 factors of lifestyle: daily alcohol use, daily tobacco use, and age of patient

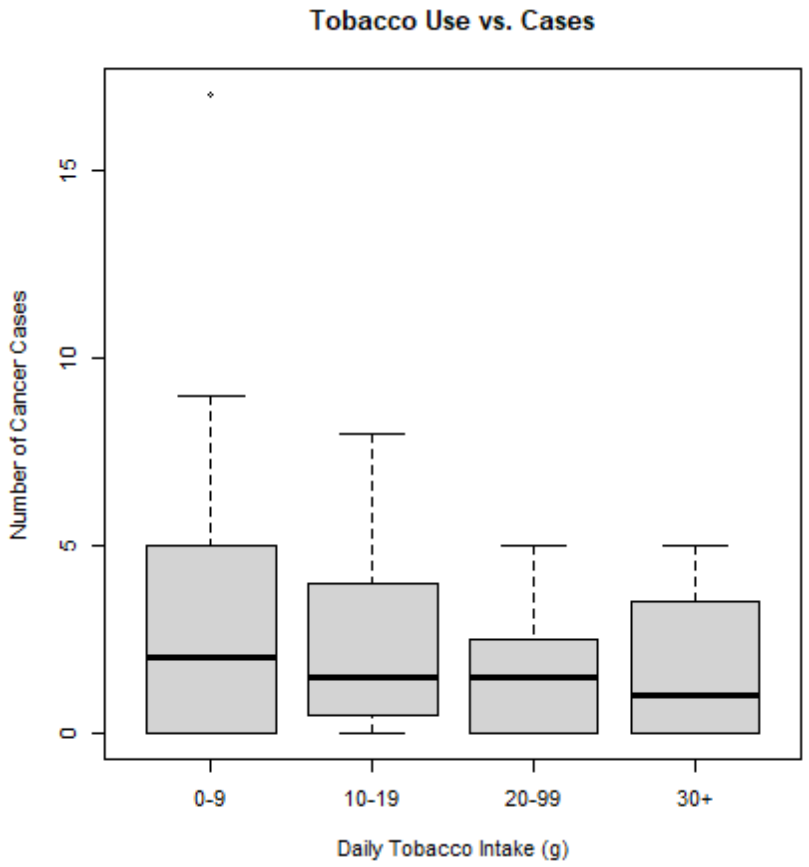
Experiment with three crossed factors

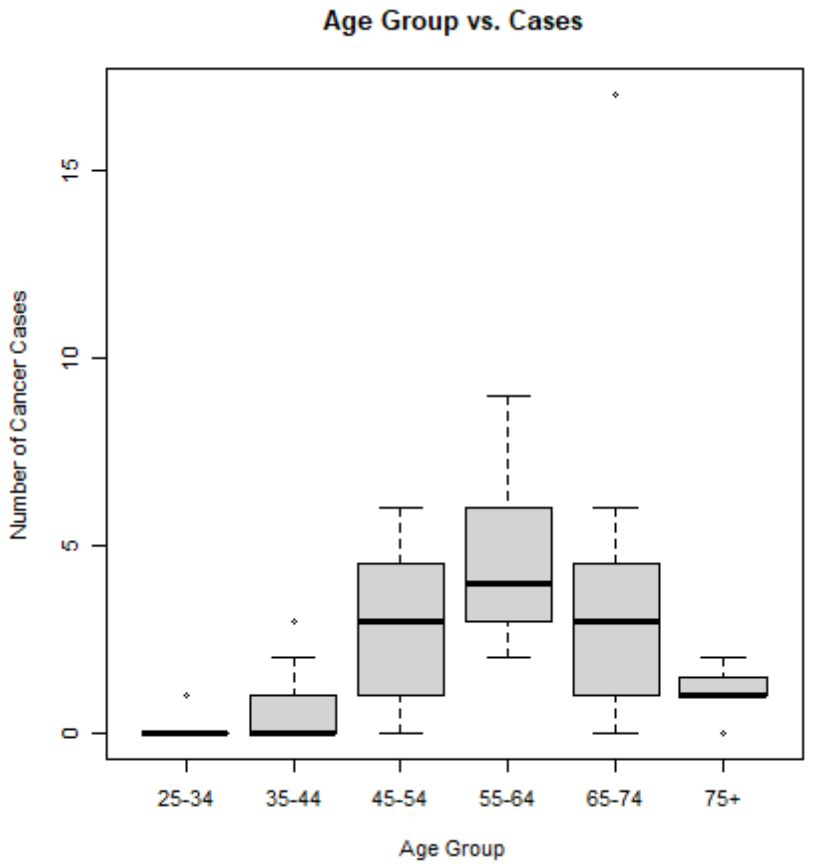
All analyses have been conducted using RStudio, using an overall significance level of 0.05 (95% confidence).



(Cancer Research UK.)





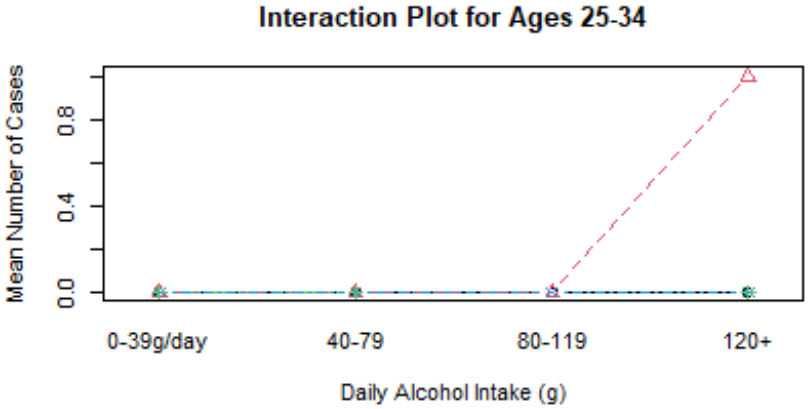


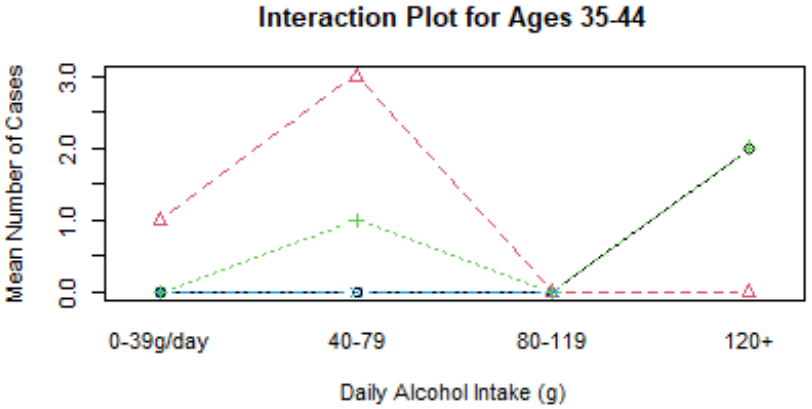
Box Plots

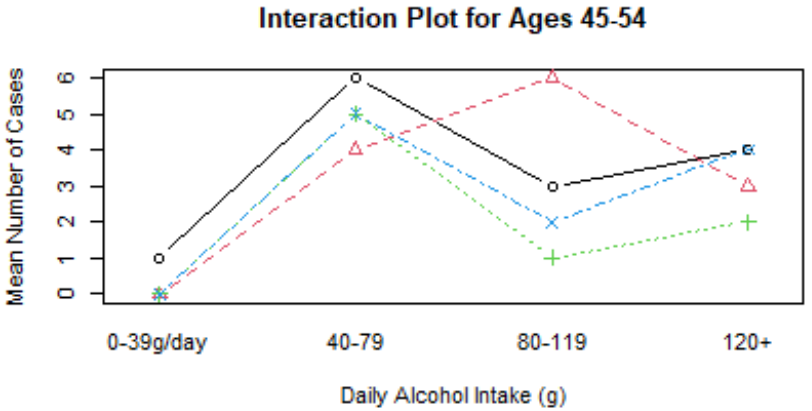
Show the number of esophageal cancer cases for each factor

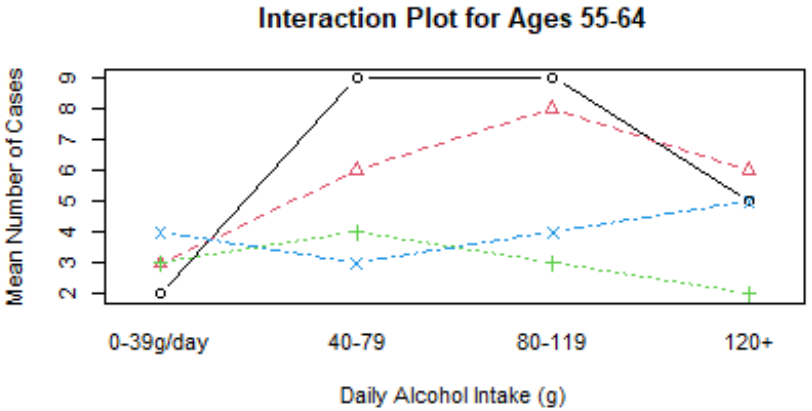
Interactions

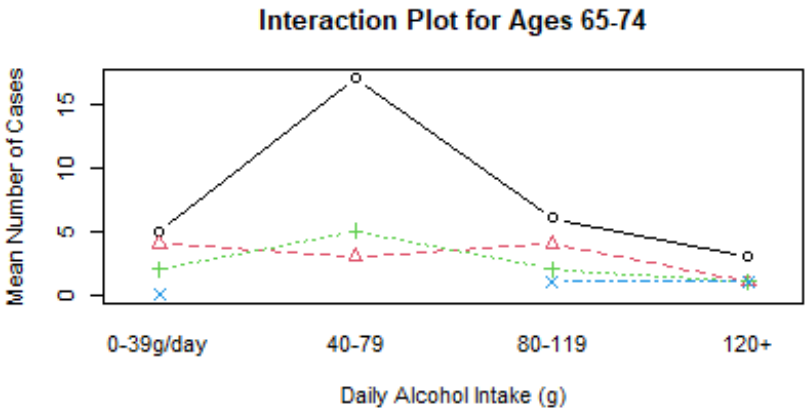
Plots for three-way interaction, age as the third factor

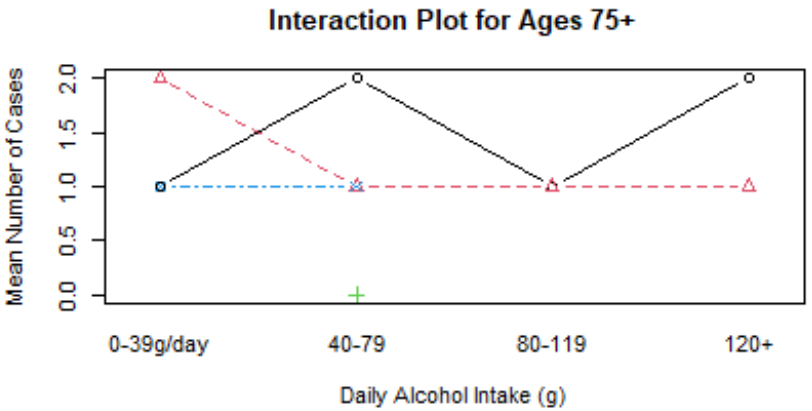


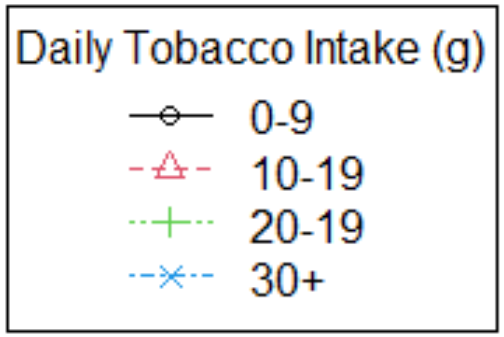












Two-way interaction model

Nonnegligible effects:

- Alcohol use (A)
- Tobacco use (B)
- Age (C)
- Interaction of Alcohol use and Age (AC)

Negligible interactions:

- Interaction of Alcohol and Tobacco uses (AB)
- Interaction of Tobacco use and Age (BC)

Model used for experiment:

Number of cases = Alcohol + Tobacco + Age + Interaction of Alcohol use and Age

Two-way Analysis of Variance Table

Factor	Degrees of Freedom	F value	Pr(>F)
Alcohol Use (fA)	3	6.5317	0.001173
Tobacco Use (fB)	3	5.2316	0.004120
Age group (fC)	5	20.1972	1.13e-09
fA:fB	9	1.3714	0.236083
fA:fC	15	2.3900	0.015714
fB:fC	15	1.8950	0.057106
Residuals	37		

Contrasts for Alcohol (Factor A)

Pairwise Contrasts using Tukey

- Two contrasts are significant at confidence level 95%, (0-39g/day) - (40-79) and (40-79g/day) - (120+).
- Number of cases decreases between (0-39g/day) - (40-79)
- Number of cases increases between (40-79g/day) - (120+).

contrast	lower.CL	upper.CL	t.ratio	p.value
(0-39g/day) - (40-79)	-3.6141	-0.839	-4.238	0.0004
(0-39g/day) - (80-119)	-2.2728	0.610	-1.524	0.4297
(0-39g/day) - (120+)	-2.1406	0.747	-1.275	0.5821
(40-79) - (80-119)	-0.0487	2.839	2.552	0.0619
(40-79) - (120+)	0.0883	2.971	2.803	0.0334
(80-119) - (120+)	-1.3582	1.627	0.238	0.9952

Contrasts for Tobacco (Factor B)

Pairwise Contrasts using Tukey

- Only two contrasts are significant at confidence level 95%, (0-9g/day) - (20-29) and (0-9g/day) - (30+).
- In both cases, number of cancer cases increases when patients who use 0-9g/day and 20-29 and 30+ g/day).

contrast	lower.CL	upper.CL	t.ratio	p.value
(0-9g/day) - (10-19)	-0.515	2.18	1.633	0.3683
(0-9g/day) - (20-29)	0.495	3.38	3.550	0.0041
(0-9g/day) - (30+)	0.168	3.05	2.949	0.0228
(10-19) - (20-29)	-0.338	2.54	2.021	0.1915
(10-19) - (30+)	-0.666	2.21	1.420	0.4917
(20-29) - (30+)	-1.830	1.17	-0.576	0.9389

Contrasts for Age (Factor C)

Pairwise Contrasts using Tukey

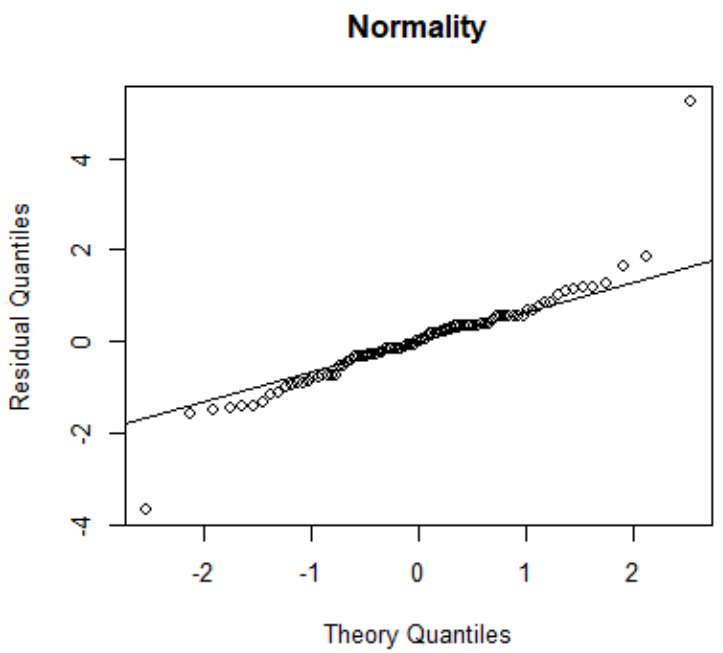
- Nine contrasts are significant at confidence level 95%, mostly between younger age groups and older:
  - (25-34) - (45-54)
  - (25-34) - (55-64)
  - (25-34) - (65-74)
  - (35-44) - (45-54)
  - (35-44) - (55-64)
  - (35-44) - (65-74)
  - (45-54) and (55-64)
  - (55-64) - (75+)
  - (65-74) and (75+)

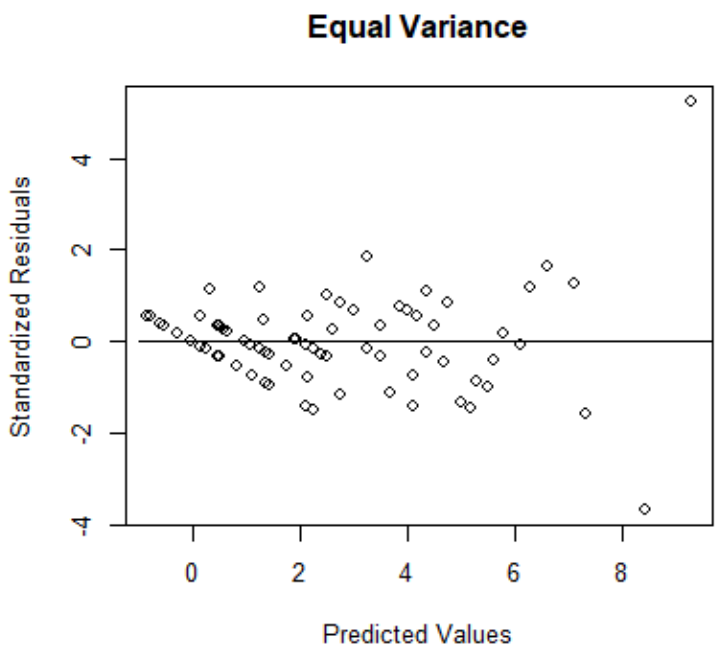
contrast		lower.CL	upper.CL	t.ratio	p.value
(25-34)	- (35-44)	-2.5300	1.3088	-0.936	0.9355
(25-34)	- (45-54)	-4.7617	-1.0036	-4.514	0.0004
(25-34)	- (55-64)	-6.6367	-2.8786	-7.449	<.0001
(25-34)	- (65-74)	-5.8425	-2.0037	-6.014	<.0001
(25-34)	- (75+)	-2.9501	1.3358	-1.108	0.8762
(35-44)	- (45-54)	-4.1510	-0.3929	-3.557	0.0091
(35-44)	- (55-64)	-6.0260	-2.2679	-6.493	<.0001
(35-44)	- (65-74)	-5.2268	-1.3982	-5.091	0.0001
(35-44)	- (75+)	-2.3440	1.9510	-0.269	0.9998
(45-54)	- (55-64)	-3.7142	-0.0358	-2.999	0.0432
(45-54)	- (65-74)	-2.9196	0.8385	-1.629	0.5828
(45-54)	- (75+)	-0.0425	4.1934	2.883	0.0580
(55-64)	- (65-74)	-1.0446	2.7135	1.307	0.7803
(55-64)	- (75+)	1.8325	6.0684	5.488	<.0001
(65-74)	- (75+)	0.9685	5.2635	4.269	0.0009

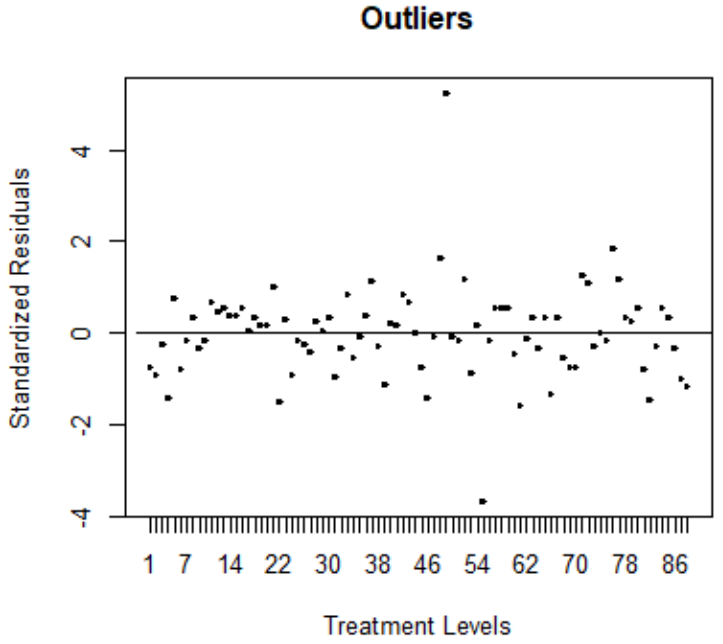
Checking Assumptions of Model

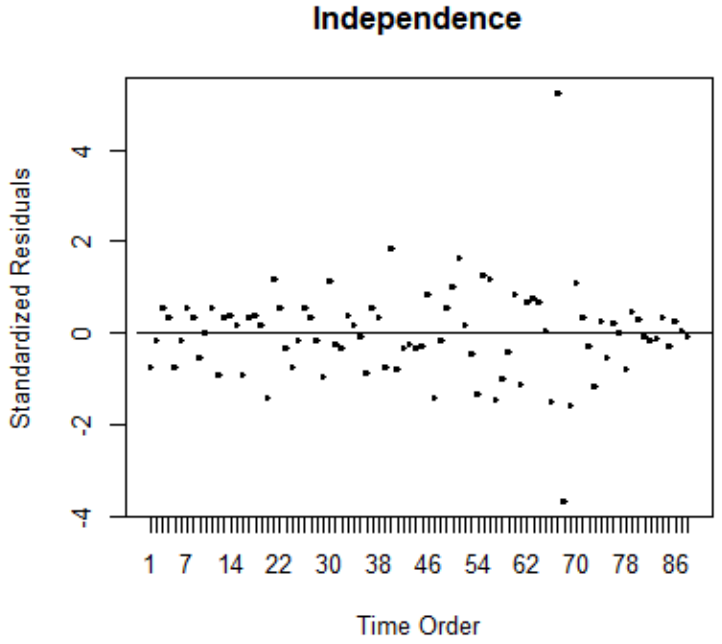
Checking residuals in model are

- Independent and identically distributed
- Have equal variance
- Normally distributed with a mean of zero
- Have no outliers > 3



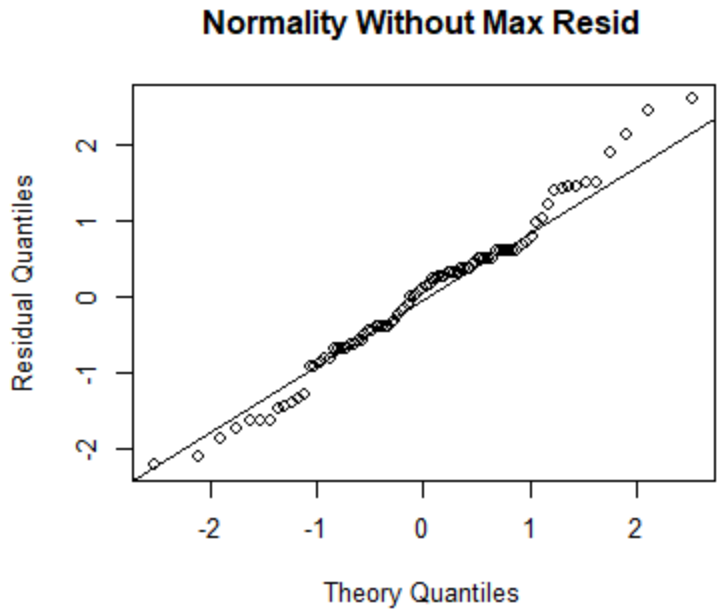


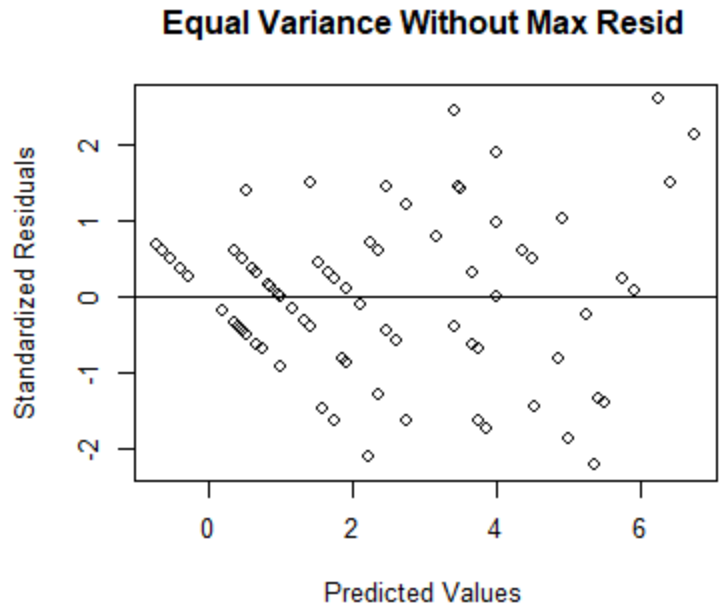


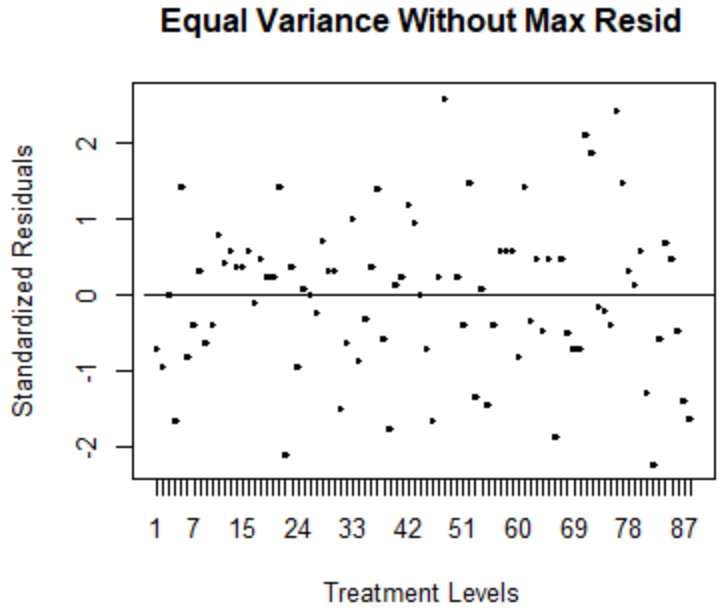


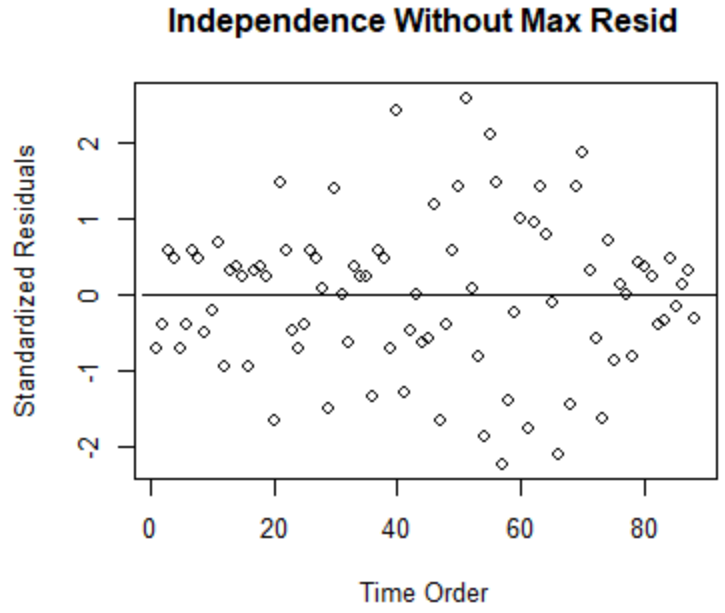
Removing Maximum Residual

Maximum standardized residual was found to be: 6.406573









Conclusion

The analysis of this experiment found there is sufficient evidence to support there is a statistically significant difference on the response from the different factors and interactions, and that there is sufficient evidence to support there is a statistically significant difference on the response from the different levels of each factor.

The motivations for the experiment are met and show any amount of alcohol or tobacco, as well as increased age, increases the risk of esophageal cancer.

References

Esophageal Cancer - Statistics. Cancer.Net. (2021, May 27). https://www.cancer.net/cancer-types/esophageal-cancer/statistics.

Esophageal cancer survival statistics. Cancer Research UK. (2020, April 7). https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/oesophageal-cancer/survival.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Russell V. Lenth (2016). Least-Squares Means: The R Package lsmeans. Journal of Statistical Software, 69(1), 1-33. doi:10.18637/jss.v069.i01