

**Introduction to Contraceptive Pills and Mortality**

The objective of the project is to view the risk of mortality among women whom consume oral contraceptives, versus women who never used them. The Study: Royal College of General Practitioners’ Oral Contraception RCGP study started in 1968. Data Supplied: The data for the study was supplied by the practitioners whom participated in the study, and the National Health Service central services. Amount: There were 1,200 general practices within the United Kingdom. Subjects: There were 46, 112 women whom underwent the 39 years in the study. Results: There were 1,747 women that died and never took the pills. There were 2,864 whom used the pills and died.

**Hypothesis/Statistical Question**

My hypothesis questions were what was the association of women whom died and were taking the pills? Were the women whom consumed the contraceptive pills already at risk of dying before taking the pill? What are other factors that may have caused death while the women took the pills?

**EDA Outcome**

The EDA was very, very challenging because the data set and structure was not looking good. I cleaned the data and it started to look better compared to its raw status. I was considering changing my project to a different project, but I decided to stick with this topic. After the analysis, the outliers were noticeable in the graphs. The graphs showed that the subjects had some correlation between them. There wasn’t much of an incline with the scatter plots, but there was visual evidence that there was a stable relationship among the data that was closest to the x-axis.

Missed in the Analysis

I missed some of the outliers and my plan is to catch them before I continue on with the data observations. Some other observations that could have been added are women whom have children versus women that do not have any. Did complications from the pregnancy or birth, have anything to their supposed risk of dying if they take the pill.

Some of the obstacles I faced were dealing with cleaning the dataset. The data on the website had data tables with ridiculous structure. Then when I read the file on Python, it looked like this:

Input:

#DSC530

#Final Project

#LaChandra Ash

#2.29.2020

import csv

with open('pills.csv') as csvfile:

readCSV = csv.reader(csvfile, delimiter = ",")

for row in readCSV:

print(row)

output:

ï»¿Cause of death', 'ICD-8 codes', 'No Pill Observed Rate', 'No Pill Standard Rate', 'Used Pill Observed Rate', 'Used Pill Standard Rate', 'Adjusted Relative Risk']

['All causes', '000-999, all E codes', '462.16 (1747)', '417.45', '349.62 (2864)', '365.51', '0.88 (0.82 to 0.93)']

['All cancers', '140-209', '205.29 (776)', '194.55', '160.16 (1312)', '165.45', '0.85 (0.78 to 0.93)']

['â€ƒLarge bowel and rectum', '153-154', '21.16 (80)', '20.05', '11.84 (97)', '12.41', '0.62 (0.46 to 0.83)']

['â€ƒGallbladder/liver', '155-156', '3.17 (12)', '3.12', '1.83 (15)', '2.03', '0.65 (0.30 to 1.39)']

['â€ƒLung', '162', '26.45 (100)', '26.08', '31.49 (258)', '31.7', '1.22 (0.96 to 1.53)']

['â€ƒMelanoma', '172', '2.65 (10)', '2.67', '1.95 (16)', '1.95', '0.73 (0.33 to 1.61)']

['â€ƒBreast', '174', '44.44 (168)', '43.91', '38.09 (312)', '39.41', '0.90 (0.74 to 1.08)']

['â€ƒInvasive cervix', '180', '3.70 (14)', '4.02', '5.62 (46)', '5.38', '1.34 (0.74 to 2.44)']

['â€ƒUterine body', '182', '5.03 (19)', '4.47', '1.59 (13)', '1.94', '0.43 (0.21 to 0.88)']

['â€ƒOvary', '183', '19.84 (75)', '18.04', '9.16 (75)', '9.47', '0.53 (0.38 to 0.72)']

['â€ƒMain gynaecological', '180, 182, 183', '28.57 (108)', '26.51', '16.36 (134)', '16.8', '0.63 (0.49 to 0.82)']

['â€ƒCNS-pituitary', '191, 1943', '5.03 (19)', '4.47', '3.42 (28)', '3.74', '0.84 (0.47 to 1.50)']

['â€ƒSite unknown', '199', '22.22 (84)', '20.5', '17.21 (141)', '18.02', '0.88 (0.67 to 1.15)']

['â€ƒOther cancers', '140-209, except above', '51.59 (195)', '47.19', '37.96 (311)', '39.39', '0.83 (0.70 to 1.00)']

['All circulatory diseases', '390-458', '132.54 (501)', '115.18', '93.14 (763)', '99.15', '0.86 (0.77 to 0.96)']

['â€ƒIschaemic heart disease', '410-414', '64.02 (242)', '57.41', '41.02 (336)', '42.85', '0.75 (0.63 to 0.88)']

['â€ƒOther heart', '420-429', '15.34 (58)', '11.9', '9.03 (74)', '10.12', '0.85 (0.60 to 1.20)']

['â€ƒCerebrovascular disease', '430-438', '32.54 (123)', '27.86', '27.71(227)', '29.19', '1.05 (0.84 to 1.30)']

['â€ƒOther circulatory', '390-409, 440-458', '20.63 (78)', '18.02', '15.38 (126)', '16.98', '0.94 (0.71 to 1.25)']

['All digestive disease', '520-577', '18.25 (69)', '16.53', '15.38 (126)', '15.67', '0.95 (0.71 to 1.27)']

['â€ƒLiver disease', '570-573', '5.56 (21)', '5.48', '7.20 (59)', '7.2', '1.32 (0.80 to 2.16)']

['Violence', '800-999, E800-999', '13.49 (51)', '12.86', '19.04 (156)', '19.2', '1.49 (1.09 to 2.05)']

['â€ƒSuicide', 'E950-959', '4.50 (17)', '4.79', '6.10 (50)', '6.03', '1.26 (0.73 to 2.18)']

['All other diseases', 'All codes, except above', '92.06 (348)', '77.8', '61.4 (503)', '65.59', '0.84 (0.74 to 0.97)']

Traceback (most recent call last):

File "<ipython-input-1-e897524ca10a>", line 1, in <module>

runfile('C:/Users/LAsh8/Desktop/Ash\_LaChandra\_DSC530.py', wdir='C:/Users/LAsh8/Desktop')

File "C:\Users\LAsh8\Anaconda3\lib\site-packages\spyder\_kernels\customize\spydercustomize.py", line 827, in runfile

execfile(filename, namespace)

File "C:\Users\LAsh8\Anaconda3\lib\site-packages\spyder\_kernels\customize\spydercustomize.py", line 110, in execfile

exec(compile(f.read(), filename, 'exec'), namespace)

runfile('C:/Users/LAsh8/Desktop/Ash\_LaChandra\_DSC530.py', wdir='C:/Users/LAsh8/Desktop').

Reference

The Bmj (2010). Mortality Among Contraceptive Pill Users: Cohort Evidence from Royal College of General Practitioners’ Oral Contraception Study. Retrieved from <https://www.bmj.com/content/340/bmj.c927>, on Feb 29, 2020.