Assignment-02
On
Lifetime Data Analysis
AST 405

Submitted By: *Submitted To:*
Shafayet Khan Shafee   Dr. Mahbub Latif
FH-033-011   Professor
4th Year   ISRT

January 21, 2022

Contents

Answers to Questions

Question 01

Table 1: Descriptive Statistics

	level	Overall
n		3154
age (mean (SD))		46.35 (5.56)
bmi (mean (SD))		24.48 (2.55)
arcus (%)	0	2219 (70.4)
	1	934 (29.6)
	NA	1 (0.0)
behpat (%)	A1	275 (8.7)
	A2	1290 (40.9)
	B3	1236 (39.2)
	B4	353 (11.2)
chd69 (%)	No	2900 (91.9)
	Yes	254 (8.1)

Question 02

Table 2: Correlation Test (age, bmi) and T-test (arcus) with chol

term	estimate	statistic	p.value
age	0.081	4.550	<.001
bmi	0.060	3.385	<.001
arcus	-12.672	-7.604	<.001

Table 3: One Way Analysis of Variance for chol on behpat

term	df	sumsq	meansq	statistic	p.value
behpat	3	30741.67	10247.223	5.475	<.001
Residuals	3135	5867632.50	1871.653	NA	NA

Question 03

Table 4: 2x2 table with Chi-square test result

	CHD	
behpat	No	Yes
A1	246	29
A2	1158	132
B3	1164	72
B4	332	21

Pearson's Chi-squared test	
statistic	p.value
20.978	<.001

Question 04

Table 5: Estimate of Model Parameters

term	estimate	std.error	statistic	p.value
(Intercept)	196.971	6.483	30.381	<.001
age	0.632	0.139	4.550	<.001

$$\widehat{\text{chol}} = 196.97 + 0.63(\text{age}) \tag{1}$$

Question 05

Table 6: Estimate of Model Parameters

term	estimate	std.error	statistic	p.value
(Intercept)	228.352	1.097	208.102	$<.001$
dibpatType B	-4.147	1.546	-2.682	0.007

$$\widehat{\text{chol}} = 228.35 - 4.15(\text{dibpat}_{\text{Type B}}) \tag{2}$$

Question 06

Table 7: Estimate of Model Parameters

term	estimate	std.error	statistic	p.value
(Intercept)	235.073	2.618	89.778	$<.001$
behpatA2	-8.150	2.883	-2.827	0.005
behpatB3	-10.226	2.894	-3.533	$<.001$
behpatB4	-13.122	3.491	-3.759	$<.001$

$$\widehat{\text{chol}} = 235.07 - 8.15(\text{behpat}_{\text{A2}}) - 10.23(\text{behpat}_{\text{B3}}) - 13.12(\text{behpat}_{\text{B4}}) \tag{3}$$

Question 07

Table 8: Estimate of Model Parameters

term	estimate	std.error	statistic	p.value
(Intercept)	207.388	7.132	29.080	<.001
age	0.582	0.140	4.172	<.001
behpatA2	-7.699	2.878	-2.675	0.008
behpatB3	-9.268	2.896	-3.200	0.001
behpatB4	-11.841	3.496	-3.387	<.001

$$\widehat{\text{chol}} = 207.39 + 0.58(\text{age}) - 7.7(\text{behpat}_{A2}) - 9.27(\text{behpat}_{B3}) - \\ 11.84(\text{behpat}_{B4}) \tag{4}$$

Question 08

Table 9: Estimate of Model Parameters

term	estimate	std.error	statistic	p.value
(Intercept)	222.252	1.171	189.755	<.001
age40	0.632	0.139	4.550	<.001

$$\widehat{\text{chol}} = 222.25 + 0.63(\text{age40}) \tag{5}$$

Table 10: Estimate of Model Parameters

term	estimate	std.error	statistic	p.value
(Intercept)	223.702	1.597	140.043	<.001
dibpatType B	-3.644	2.175	-1.675	0.094
smokeYes	8.728	2.189	3.988	<.001
dibpatType B:smokeYes	0.265	3.085	0.086	0.932

$$\widehat{\text{chol}} = 223.7 - 3.64(\text{dibpat}_{\text{Type B}}) + 8.73(\text{smoke}_{\text{Yes}}) + 0.27(\text{dibpat}_{\text{Type B}} \times \text{smoke}_{\text{Yes}}) \tag{6}$$

R-code

```
knitr::opts_chunk$set(echo = FALSE,
                      message = FALSE,
                      warning = FALSE
                      )
## ------- package setup -----------------------

library(dplyr)
library(purrr)
library(knitr)
library(broom)
library(tableone)
library(kableExtra)
library(equatiomatic)




## ------- data setup -----------------------

load(here::here("data", "wcgs.Rdata"))
sid <- 011
set.seed(sid)
mydat <- sample_n(wcgs, size = n(), replace = TRUE)




## ------- utility functions -----------------------

kab_tab <- function(tab, ...) {
  knitr::kable(
    tab,
    format = "latex", booktabs = TRUE,
    digits = 3, ...
```

```r
  )
}


p_format <- function(pval) {
  ifelse(pval < .001, " <.001 ", as.character(round(pval, 3)))
}


mod_tab <- function(mod, ...) {
  mod %>% tidy() %>%
    mutate(p.value = p_format(p.value)) %>%
    kab_tab(
      align = "lrrrr",
      caption = "Estimate of Model Parameters", ...
    ) %>%
    kable_styling(latex_options = "HOLD_position")
}


reg_eq <- function(mod, ...) {
  extract_eq(mod, use_coefs = TRUE, intercept = "beta", wrap = TRUE, ...)
}



## ------- Code for Question-01 -------------------------
tab <- CreateTableOne(
    data = mydat,
    vars = c("age", "bmi", "arcus","behpat", "chd69"),
    factorVars = "arcus",
    includeNA = TRUE,
    addOverall = TRUE
  )
```

```r
tab_p <- print(tab, showAllLevels = TRUE, printToggle = FALSE)

kab_tab(tab_p, caption = "Descriptive Statistics") %>%
  kable_styling(latex_options = "HOLD_position")



## ------- Code for Question-02 ------------------------
arcus <- t.test(chol ~ arcus, data = mydat) %>%
  tidy() %>%
  mutate(term = "arcus") %>%
  select(term, estimate, statistic, p.value)

mydat %>%
  select(age, bmi) %>%
  map(~ cor.test(x = .x, y = mydat$chol)) %>%
  map_dfr(broom::tidy, .id = "term") %>%
  select(term:p.value) %>%
  bind_rows(arcus) %>%
  mutate(p.value = p_format(p.value)) %>%
  kab_tab(
    align = "lrrr",
    caption = "Correlation Test (age, bmi) and T-test (arcus) with chol"
  ) %>%
  kable_styling(latex_options = "HOLD_position")


anova(lm(chol ~ behpat, data = mydat)) %>%
  tidy() %>%
  mutate(p.value = p_format(p.value)) %>%
  kab_tab(
    align = "lrrrrr",
    caption = "One Way Analysis of Variance for chol on behpat"
```

```
  ) %>%
  kable_styling(latex_options = "HOLD_position")



## ------- Code for Question-03 ------------------------

tab_chi <- mydat %>% janitor::tabyl(behpat, chd69)

tab_chi %>%
  kab_tab(
    align = "lrr",
    caption = "2x2 table with Chi-square test result"
    ) %>%
  add_header_above(header = c(" " = 1, "CHD" = 2)) %>%
  kable_styling(position = "center", latex_options = "HOLD_position")

tab_chi %>% janitor::chisq.test() %>%
  tidy() %>%
  select(statistic, p.value) %>%
  mutate(p.value = p_format(p.value)) %>%
  kab_tab(align = "lrr") %>%
  add_header_above(header = c("Pearson's Chi-squared test" = 2)) %>%
  kable_styling(position = "center", latex_options = "HOLD_position")



## ------- Code for Question-04 ------------------------

m1 <- lm(chol ~ age, data = mydat)
m1 %>% mod_tab()
reg_eq(m1)
```

```
## ------- Code for Question-05 ------------------------

m2 <- lm(chol ~ dibpat, data = mydat)
m2 %>% mod_tab()

reg_eq(m2)

## ------- Code for Question-05 ------------------------

m3 <- lm(chol ~ behpat, data = mydat)
m3 %>% mod_tab()

reg_eq(m3)

## ------- Code for Question-08 ------------------------

m4 <- lm(chol ~ age + behpat, data = mydat)
m4 %>% mod_tab()

reg_eq(m4)

## ------- Code for Question-08 ------------------------

mydat %>%
  mutate(age40 = age - 40) %>%
  lm(chol ~ age40, data = .) -> m5

m5 %>% mod_tab()

reg_eq(m5)
## ------- Code for Question-08 ------------------------
```

```
m6 <- lm(chol ~ dibpat * smoke, data = mydat)
m6 %>% mod_tab()
reg_eq(m6)
```