R version 3.4.1 (2017-06-30) -- "Single Candle"

Copyright (C) 2017 The R Foundation for Statistical Computing

Platform: x86\_64-apple-darwin15.6.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.

You are welcome to redistribute it under certain conditions.

Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.

Type 'contributors()' for more information and

'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or

'help.start()' for an HTML browser interface to help.

Type 'q()' to quit R.

> ##### R Lab 8: Multivariate Regression -- Hospital Infection Data ####

> install.packages("tidyverse")

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/tidyverse\_1.2.1.tgz'

Content type 'application/x-gzip' length 77756 bytes (75 KB)

==================================================

downloaded 75 KB

The downloaded binary packages are in

/var/folders/79/tx9tjz8j0hl99904bkw5dy740000gn/T//Rtmp6xzF78/downloaded\_packages

> install.packages("GGally")

also installing the dependencies ‘prettyunits’, ‘progress’

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/prettyunits\_1.0.2.tgz'

Content type 'application/x-gzip' length 23970 bytes (23 KB)

==================================================

downloaded 23 KB

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/progress\_1.2.0.tgz'

Content type 'application/x-gzip' length 40085 bytes (39 KB)

==================================================

downloaded 39 KB

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/GGally\_1.4.0.tgz'

Content type 'application/x-gzip' length 1260324 bytes (1.2 MB)

==================================================

downloaded 1.2 MB

The downloaded binary packages are in

/var/folders/79/tx9tjz8j0hl99904bkw5dy740000gn/T//Rtmp6xzF78/downloaded\_packages

> install.packages("psych")

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/psych\_1.8.10.tgz'

Content type 'application/x-gzip' length 5809056 bytes (5.5 MB)

==================================================

downloaded 5.5 MB

The downloaded binary packages are in

/var/folders/79/tx9tjz8j0hl99904bkw5dy740000gn/T//Rtmp6xzF78/downloaded\_packages

> install.packages("car")

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/car\_3.0-2.tgz'

Content type 'application/x-gzip' length 1482243 bytes (1.4 MB)

==================================================

downloaded 1.4 MB

The downloaded binary packages are in

/var/folders/79/tx9tjz8j0hl99904bkw5dy740000gn/T//Rtmp6xzF78/downloaded\_packages

> install.packages("QuantPsyc")

trying URL 'https://cran.rstudio.com/bin/macosx/el-capitan/contrib/3.4/QuantPsyc\_1.5.tgz'

Content type 'application/x-gzip' length 98671 bytes (96 KB)

==================================================

downloaded 96 KB

The downloaded binary packages are in

/var/folders/79/tx9tjz8j0hl99904bkw5dy740000gn/T//Rtmp6xzF78/downloaded\_packages

> library(GGally)

Loading required package: ggplot2

Warning messages:

1: package ‘GGally’ was built under R version 3.4.4

2: package ‘ggplot2’ was built under R version 3.4.4

> library(psych)

Attaching package: ‘psych’

The following objects are masked from ‘package:ggplot2’:

%+%, alpha

Warning message:

package ‘psych’ was built under R version 3.4.4

> library(QuantPsyc)

Loading required package: boot

Attaching package: ‘boot’

The following object is masked from ‘package:psych’:

logit

Loading required package: MASS

Attaching package: ‘QuantPsyc’

The following object is masked from ‘package:base’:

norm

> library(car)

Loading required package: carData

Attaching package: ‘car’

The following object is masked from ‘package:boot’:

logit

The following object is masked from ‘package:psych’:

logit

Warning messages:

1: package ‘car’ was built under R version 3.4.4

2: package ‘carData’ was built under R version 3.4.4

> library(tidyverse)

── Attaching packages ─────────────────────────────────────────────────────────── tidyverse 1.2.1 ──

✔ tibble 1.4.2 ✔ purrr 0.2.5

✔ tidyr 0.8.1 ✔ dplyr 0.7.6

✔ readr 1.1.1 ✔ stringr 1.3.1

✔ tibble 1.4.2 ✔ forcats 0.3.0

── Conflicts ────────────────────────────────────────────────────────────── tidyverse\_conflicts() ──

✖ psych::%+%() masks ggplot2::%+%()

✖ psych::alpha() masks ggplot2::alpha()

✖ dplyr::filter() masks stats::filter()

✖ dplyr::lag() masks stats::lag()

✖ dplyr::recode() masks car::recode()

✖ dplyr::select() masks MASS::select()

✖ purrr::some() masks car::some()

Warning messages:

1: package ‘tidyverse’ was built under R version 3.4.2

2: package ‘tibble’ was built under R version 3.4.3

3: package ‘tidyr’ was built under R version 3.4.4

4: package ‘purrr’ was built under R version 3.4.4

5: package ‘dplyr’ was built under R version 3.4.4

6: package ‘stringr’ was built under R version 3.4.4

7: package ‘forcats’ was built under R version 3.4.3

> hospinf <- read\_csv("hospital\_infct.csv")

Error: 'hospital\_infct.csv' does not exist in current working directory ('/Users/aaditirokade/Desktop/Quant/R\_Lab8').

> hospinf <- read\_csv("hospital\_infct.csv")

Parsed with column specification:

cols(

ID = col\_integer(),

Stay = col\_double(),

Age = col\_double(),

InfctRsk = col\_double(),

Culture = col\_double(),

Xray = col\_double(),

Beds = col\_integer(),

MedSchool = col\_integer(),

Region = col\_integer(),

Census = col\_integer(),

Nurses = col\_integer(),

Facilities = col\_double()

)

> # Basic Statistis of our variables

> describe(hospinf[2:12]) # we don't care about 'ID'

vars n mean sd median trimmed mad min max range skew kurtosis se

Stay 1 113 9.65 1.91 9.42 9.47 1.60 6.7 19.56 12.86 2.01 7.48 0.18

Age 2 113 53.23 4.46 53.20 53.28 3.85 38.8 65.90 27.10 -0.10 0.90 0.42

InfctRsk 3 113 4.35 1.34 4.40 4.38 1.19 1.3 7.80 6.50 -0.12 0.07 0.13

Culture 4 113 15.79 10.23 14.10 14.56 8.60 1.6 60.50 58.90 1.57 3.62 0.96

Xray 5 113 81.63 19.36 82.30 81.48 18.98 39.6 133.50 93.90 0.01 -0.33 1.82

Beds 6 113 252.17 192.84 186.00 221.59 139.36 29.0 835.00 806.00 1.34 1.10 18.14

MedSchool 7 113 1.85 0.36 2.00 1.93 0.00 1.0 2.00 1.00 -1.93 1.74 0.03

Region 8 113 2.36 1.01 2.00 2.33 1.48 1.0 4.00 3.00 0.06 -1.14 0.09

Census 9 113 191.37 153.76 143.00 168.30 124.54 20.0 791.00 771.00 1.34 1.52 14.46

Nurses 10 113 173.25 139.27 132.00 151.96 103.78 14.0 656.00 642.00 1.34 1.35 13.10

Facilities 11 113 43.16 15.20 42.90 42.95 16.90 5.7 80.00 74.30 0.07 -0.50 1.43

> # Correlation Matrix

> round(cor(hospinf[2:12]),2)

Stay Age InfctRsk Culture Xray Beds MedSchool Region Census Nurses Facilities

Stay 1.00 0.19 0.53 0.33 0.38 0.41 -0.30 -0.49 0.47 0.34 0.36

Age 0.19 1.00 0.00 -0.23 -0.02 -0.06 0.15 -0.02 -0.05 -0.08 -0.04

InfctRsk 0.53 0.00 1.00 0.56 0.45 0.36 -0.23 -0.19 0.38 0.39 0.41

Culture 0.33 -0.23 0.56 1.00 0.42 0.14 -0.24 -0.31 0.14 0.20 0.19

Xray 0.38 -0.02 0.45 0.42 1.00 0.05 -0.09 -0.30 0.06 0.08 0.11

Beds 0.41 -0.06 0.36 0.14 0.05 1.00 -0.59 -0.11 0.98 0.92 0.79

MedSchool -0.30 0.15 -0.23 -0.24 -0.09 -0.59 1.00 0.10 -0.61 -0.59 -0.52

Region -0.49 -0.02 -0.19 -0.31 -0.30 -0.11 0.10 1.00 -0.15 -0.11 -0.21

Census 0.47 -0.05 0.38 0.14 0.06 0.98 -0.61 -0.15 1.00 0.91 0.78

Nurses 0.34 -0.08 0.39 0.20 0.08 0.92 -0.59 -0.11 0.91 1.00 0.78

Facilities 0.36 -0.04 0.41 0.19 0.11 0.79 -0.52 -0.21 0.78 0.78 1.00

> # Correlation and Scatterplot

> ggpairs(hospinf, columns = 2:12)

> model\_guess <- lm(InfctRsk ~ Stay + Age + Culture + Xray +

+ Beds + MedSchool + Region + Census +

+ Nurses + Facilities, data = hospinf)

> summary(model\_guess)

Call:

lm(formula = InfctRsk ~ Stay + Age + Culture + Xray + Beds +

MedSchool + Region + Census + Nurses + Facilities, data = hospinf)

Residuals:

Min 1Q Median 3Q Max

-1.86974 -0.56269 -0.02893 0.51925 2.32390

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -3.239992 1.413934 -2.291 0.023992 \*

Stay 0.242240 0.070668 3.428 0.000879 \*\*\*

Age 0.010077 0.021500 0.469 0.640280

Culture 0.053444 0.010545 5.068 1.8e-06 \*\*\*

Xray 0.012632 0.005280 2.392 0.018568 \*

Beds -0.003173 0.002660 -1.193 0.235605

MedSchool 0.562186 0.319765 1.758 0.081727 .

Region 0.297558 0.104486 2.848 0.005324 \*\*

Census 0.002829 0.003430 0.825 0.411415

Nurses 0.002064 0.001688 1.223 0.224125

Facilities 0.023272 0.010028 2.321 0.022291 \*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9217 on 102 degrees of freedom

Multiple R-squared: 0.5697, Adjusted R-squared: 0.5276

F-statistic: 13.51 on 10 and 102 DF, p-value: 7.974e-15

> # Create your model (here, Xray & Hospital info)

> model1 <- lm(InfctRsk ~ Xray, data = hospinf)

> summary(model1)

Call:

lm(formula = InfctRsk ~ Xray, data = hospinf)

Residuals:

Min 1Q Median 3Q Max

-2.98805 -0.85492 -0.04643 0.91416 2.73426

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.792024 0.491363 3.647 0.000406 \*\*\* ---   
# Y intercept for impact of Xray on Infection risk

Xray 0.031396 0.005858 5.359 4.58e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.201 on 111 degrees of freedom

Multiple R-squared: 0.2056, Adjusted R-squared: 0.1984

F-statistic: 28.72 on 1 and 111 DF, p-value: 4.585e-07

# Add one more variable to Xray Hosp info > culture

> model2 <- lm(InfctRsk ~ Xray + Culture, data = hospinf)

> summary(model2)

Call:

lm(formula = InfctRsk ~ Xray + Culture, data = hospinf)

Residuals:

Min 1Q Median 3Q Max

-2.5275 -0.8468 0.0826 0.8536 2.9011

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 1.940969 0.440614 4.405 2.46e-05 \*\*\*

Xray 0.018235 0.005792 3.148 0.00211 \*\*

Culture 0.058598 0.010958 5.348 4.89e-07 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.074 on 110 degrees of freedom

Multiple R-squared: 0.3695, Adjusted R-squared: 0.358

F-statistic: 32.23 on 2 and 110 DF, p-value: 9.629e-12

# Add one more variable Hosp info > culture

> model3 <- lm(InfctRsk ~ Xray + Culture + Facilities, data = hospinf)

> summary(model3)

Call:

lm(formula = InfctRsk ~ Xray + Culture + Facilities, data = hospinf)

Residuals:

Min 1Q Median 3Q Max

-2.51974 -0.69713 0.06382 0.59712 2.69452

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 0.934951 0.469224 1.993 0.04881 \*

Xray 0.017360 0.005373 3.231 0.00163 \*\*

Culture 0.051748 0.010280 5.034 1.91e-06 \*\*\*

Facilities 0.027471 0.006305 4.357 2.99e-05 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.996 on 109 degrees of freedom

Multiple R-squared: 0.463, Adjusted R-squared: 0.4482

F-statistic: 31.33 on 3 and 109 DF, p-value: 1.107e-14

> model4 <- lm(InfctRsk ~ Xray + Culture + Facilities + Stay, data = hospinf)

> summary(model4)

Call:

lm(formula = InfctRsk ~ Xray + Culture + Facilities + Stay, data = hospinf)

Residuals:

Min 1Q Median 3Q Max

-2.22907 -0.65305 -0.00399 0.64169 2.51097

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -0.063581 0.533207 -0.119 0.905305

Xray 0.012052 0.005351 2.252 0.026316 \*

Culture 0.046446 0.009923 4.680 8.35e-06 \*\*\*

Facilities 0.020465 0.006347 3.224 0.001671 \*\*

Stay 0.188411 0.054714 3.444 0.000818 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9499 on 108 degrees of freedom

Multiple R-squared: 0.5161, Adjusted R-squared: 0.4982

F-statistic: 28.8 on 4 and 108 DF, p-value: 2.728e-16

> # Final Model

> model5 <- lm(InfctRsk ~ Xray + Culture + Facilities + Stay + Region, data = hospinf)

> summary(model5)

Call:

lm(formula = InfctRsk ~ Xray + Culture + Facilities + Stay +

Region, data = hospinf)

Residuals:

Min 1Q Median 3Q Max

-1.88387 -0.57390 0.07203 0.56280 2.20716

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.451284 0.730097 -1.988 0.04939 \*

Xray 0.013191 0.005219 2.528 0.01294 \*

Culture 0.050106 0.009741 5.144 1.23e-06 \*\*\*

Facilities 0.021071 0.006175 3.413 0.00091 \*\*\*

Stay 0.246947 0.057441 4.299 3.80e-05 \*\*\*

Region 0.273426 0.101308 2.699 0.00809 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9234 on 107 degrees of freedom

Multiple R-squared: 0.547, Adjusted R-squared: 0.5258

F-statistic: 25.84 on 5 and 107 DF, p-value: < 2.2e-16

# Conclusion: Model pass the VIF multicollinearity test as well

> # VIF for multicollinearity

> vif(model5)

Xray Culture Facilities Stay Region

1.341427 1.305754 1.157190 1.583533 1.373755

# Conclusion: Model passes the autocorrelation test as well  
> # Durbin - Watson autocorrelation test

> durbinWatsonTest(model5)

lag Autocorrelation D-W Statistic p-value

1 -0.004544764 1.987763 0.892

Alternative hypothesis: rho != 0

> Stay and culture have the greatest amount of impact in determining

> # Beta's

> lm.beta(model5)

Xray Culture Facilities Stay Region

0.1904893 0.3824423 0.2388617 0.3520207 0.2058351

> # AIC (AKK… info criteria test)

> # AIC

> AIC(model\_guess)

[1] 314.6735.   
#We see that with additional variables, AIC values are decreasing.

> AIC(model1)

[1] 365.9682

> AIC(model2)

[1] 341.856

> AIC(model3)

[1] 325.7127

> AIC(model4)

[1] 315.9408

> AIC(model5)

[1] 310.4985

> AIC(model6)

[1] 326.3338

> # Additional Model

> model6 <- lm(InfctRsk ~ Culture + Stay , data = hospinf)

> summary(model5)

Call:

lm(formula = InfctRsk ~ Xray + Culture + Facilities + Stay +

Region, data = hospinf)

Residuals:

Min 1Q Median 3Q Max

-1.88387 -0.57390 0.07203 0.56280 2.20716

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.451284 0.730097 -1.988 0.04939 \*

Xray 0.013191 0.005219 2.528 0.01294 \*

Culture 0.050106 0.009741 5.144 1.23e-06 \*\*\*

Facilities 0.021071 0.006175 3.413 0.00091 \*\*\*

Stay 0.246947 0.057441 4.299 3.80e-05 \*\*\*

Region 0.273426 0.101308 2.699 0.00809 \*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9234 on 107 degrees of freedom

Multiple R-squared: 0.547, Adjusted R-squared: 0.5258

F-statistic: 25.84 on 5 and 107 DF, p-value: < 2.2e-16