Booklet of Figures
for
STAC33 Final Exam

List of Figures in this document by page:

List of Figures

```r
library(MASS)
library(tidyverse)
```

```
## -- Attaching packages -------
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.1.5     v dplyr   1.0.8
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   2.0.1     v forcats 0.5.1
## -- Conflicts ----------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x dplyr::select() masks MASS::select()
```

```r
library(smmr)
library(PMCMRplus)
library(broom)
library(cmdstanr)
```

```
## This is cmdstanr version 0.4.0
## - Online documentation and vignettes at mc-stan.org/cmdstanr
## - CmdStan path set to: /home/ken/.cmdstanr/cmdstan-2.26.1
## - Use set_cmdstan_path() to change the path
##
## A newer version of CmdStan is available. See
?install_cmdstan() to install it.
## To disable this check set option or environment variable
CMDSTANR_NO_VER_CHECK=TRUE.
```

Figure 1: Packages

```
Ni:Current
19.1:0.095
38.2:0.174
57.3:0.256
76.2:0.348
95:0.429
114:0.500
131:0.580
150:0.651
170:0.722
```

Figure 2: Nickel data

```
##     state sex       diag      death status T.categ age
## 1    NSW   M 1989-11-09 1990-05-04      D      hs  35
## 2    NSW   M 1990-03-13 1990-05-19      D      hs  53
## 3    NSW   M 1986-02-24 1987-05-02      D      hs  42
## 4    NSW   M 1986-03-22 1986-06-07      D    haem  44
## 5    NSW   M 1987-06-03 1988-03-04      D      hs  39
## 6    NSW   M 1987-04-20 1988-04-27      D      hs  36
## 7    NSW   M 1989-06-03 1990-06-27      D   other  36
## 8    NSW   M 1987-06-30 1990-04-22      D      hs  31
## 9    NSW   M 1988-08-25 1989-12-30      D      hs  26
## 10   NSW   M 1988-07-31 1989-10-08      D    hsid  27
## 11   NSW   M 1988-07-08 1988-07-24      D      hs  45
## 12   NSW   M 1987-12-21 1988-10-24      D      hs  36
## 13   NSW   M 1988-06-07 1988-09-07      D      hs  27
## 14   NSW   M 1988-05-19 1989-02-08      D      hs  35
## 15   NSW   M 1988-08-13 1991-07-01      A      hs  30
## 16   NSW   M 1988-11-21 1989-04-02      D      hs  39
## 17   NSW   M 1989-03-22 1990-08-31      D      hs  30
## 18   NSW   M 1989-11-27 1991-07-01      A    haem  21
## 19   NSW   M 1990-02-05 1991-07-01      A      hs  56
## 20   NSW   M 1990-03-04 1990-08-02      D      hs  41
```

Figure 3: Australian AIDS data, some

```
river_water

## # A tibble: 5 x 2
##   upstream downstream
##      <dbl>      <dbl>
## 1      4.8        5
## 2      5.2        4.7
## 3      5          4.9
## 4      4.9        4.8
## 5      5.1        4.9
```

Figure 4: River water data

```r
with(river_water, t.test(upstream, downstream, alternative = "less",
                         paired = TRUE))

##
##  Paired t-test
##
## data:  upstream and downstream
## t = 1.2472, df = 4, p-value = 0.8598
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf 0.3792992
## sample estimates:
## mean of the differences
##                    0.14
```

Figure 5: River water data analysis 1

```r
t.test(oxygen~where, alternative = "less", data = river2)

##
##  Welch Two Sample t-test
##
## data:  oxygen by where
## t = -1.6059, df = 7.2746, p-value = 0.07536
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##        -Inf 0.02423248
## sample estimates:
## mean in group downstream   mean in group upstream
##                     4.86                     5.00
```

Figure 6: River water data analysis 2

```
t.test(oxygen~where, alternative = "less", var.equal = TRUE,
       data = river2)

##
##  Two Sample t-test
##
## data:  oxygen by where
## t = -1.6059, df = 8, p-value = 0.07348
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##        -Inf 0.02211164
## sample estimates:
## mean in group downstream   mean in group upstream
##                     4.86                     5.00
```

Figure 7: River water data analysis 3

```
median_test(river2, oxygen, where)

## $table
##             above
## group        above below
##   downstream     1     2
##   upstream       3     1
##
## $test
##        what     value
## 1 statistic 1.2152778
## 2        df 1.0000000
## 3   P-value 0.2702894

0.2703/2

## [1] 0.13515
```

Figure 8: River water data analysis 4

```
d1

## # A tibble: 2 x 3
##       x     y     z
##   <dbl> <dbl> <dbl>
## 1    10    11    13
## 2     9    12    14
```

Figure 9: Dataframe `d1`

```
d2

## # A tibble: 2 x 4
##   x     y_first y_second y_third
##   <chr>   <dbl>    <dbl>   <dbl>
## 1 a          10       11      13
## 2 b          12       15      16
```

Figure 10: Dataframe `d2`

```
d3

## # A tibble: 4 x 3
##      id x         y
##   <dbl> <chr> <dbl>
## 1     1 a        10
## 2     2 a        11
## 3     1 b        12
## 4     2 b        13
```

Figure 11: Dataframe `d3`

```
d4

## # A tibble: 4 x 3
##   x     y         z
##   <chr> <chr> <dbl>
## 1 a     b         9
## 2 c     d        10
## 3 a     d        11
## 4 c     b        12
```

Figure 12: Dataframe `d4`

```
## Rows: 414 Columns: 8-- Column specification -----
## Delimiter: ","
## dbl (8): No, X1 transaction date, X2 house age, X3 distance to
the nearest M...
## i Use `spec()` to retrieve the full column specification for
this data.
## i Specify the column types or set `show_col_types = FALSE` to
quiet this message.

## # A tibble: 414 x 5
##    sale_date   age    mrt  conv price
##        <dbl> <dbl>  <dbl> <dbl> <dbl>
##  1     2013.  32     84.9    10  37.9
##  2     2013.  19.5  307.      9  42.2
##  3     2014.  13.3  562.      5  47.3
##  4     2014.  13.3  562.      5  54.8
##  5     2013.   5    391.      5  43.1
##  6     2013.   7.1 2175.      3  32.1
##  7     2013.  34.5  623.      7  40.3
##  8     2013.  20.3  288.      6  46.7
##  9     2014.  31.7 5512.      1  18.8
## 10     2013.  17.9 1783.      3  22.1
## # ... with 404 more rows
```

Figure 13: Taiwan house data (some)

```r
boxcox(price ~ sale_date + age + mrt + conv, data = houses)
```

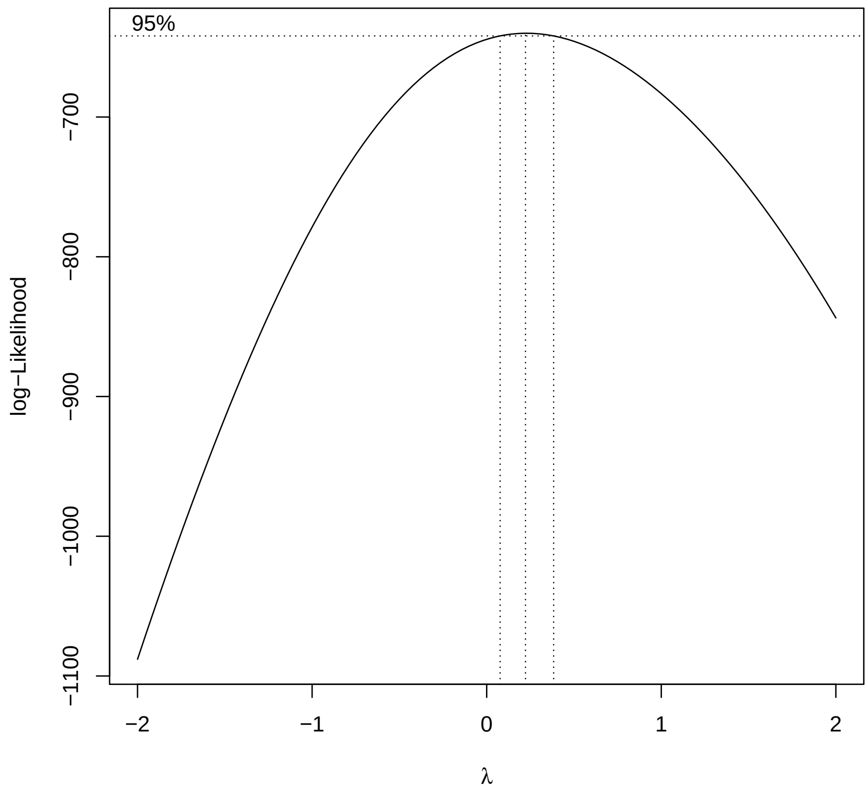

Figure 14: Taiwan houses Box-Cox analysis

```r
houses.1 <- lm(log(price) ~ sale_date + age + mrt + conv, data = houses)
summary(houses.1)
```

```
##
## Call:
## lm(formula = log(price) ~ sale_date + age + mrt + conv, data = houses)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.77830 -0.11950 -0.00312  0.12172  1.06692
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -3.146e+02  8.268e+01  -3.806 0.000163 ***
## sale_date    1.582e-01  4.107e-02   3.851 0.000136 ***
## age         -6.459e-03  1.017e-03  -6.353 5.62e-10 ***
## mrt         -1.898e-04  1.152e-05 -16.470  < 2e-16 ***
## conv         3.189e-02  4.933e-03   6.465 2.88e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2345 on 409 degrees of freedom
## Multiple R-squared:  0.6465,Adjusted R-squared:  0.643
## F-statistic:   187 on 4 and 409 DF,  p-value: < 2.2e-16
```

Figure 15: Taiwan houses regression analysis

```r
ggplot(houses.1, aes(x = .fitted, y = .resid)) + geom_point()
```

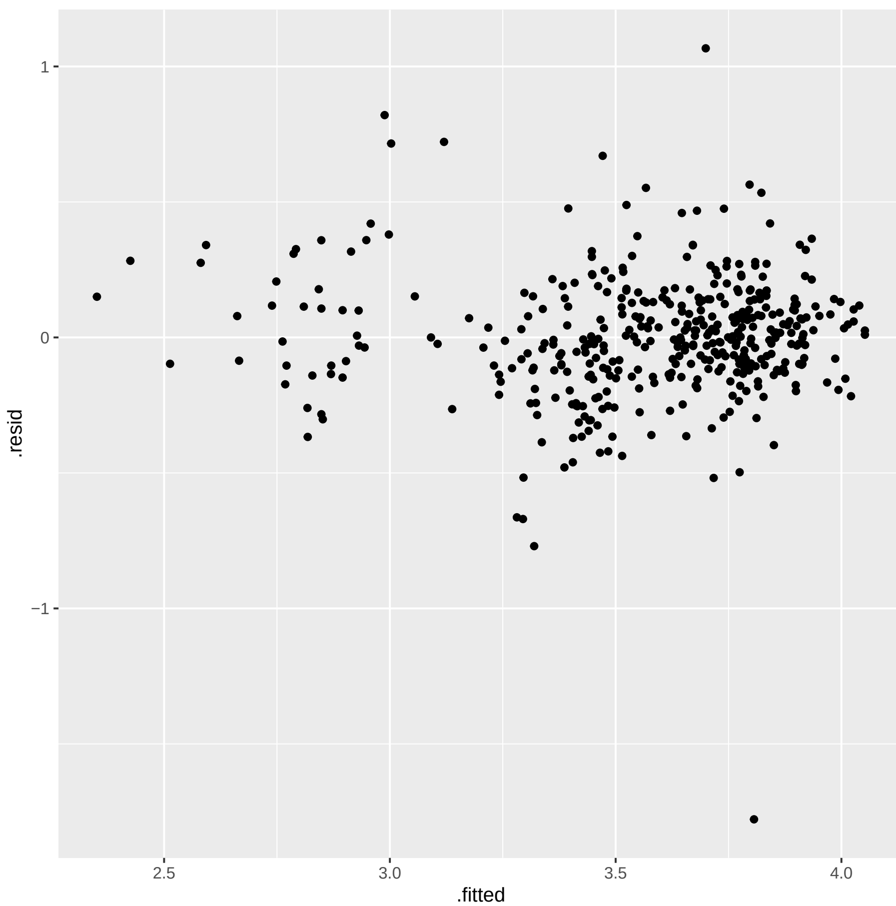

Figure 16: Taiwan houses residual plot 1

```r
ggplot(houses.1, aes(sample = .resid)) + stat_qq() + stat_qq_line()
```

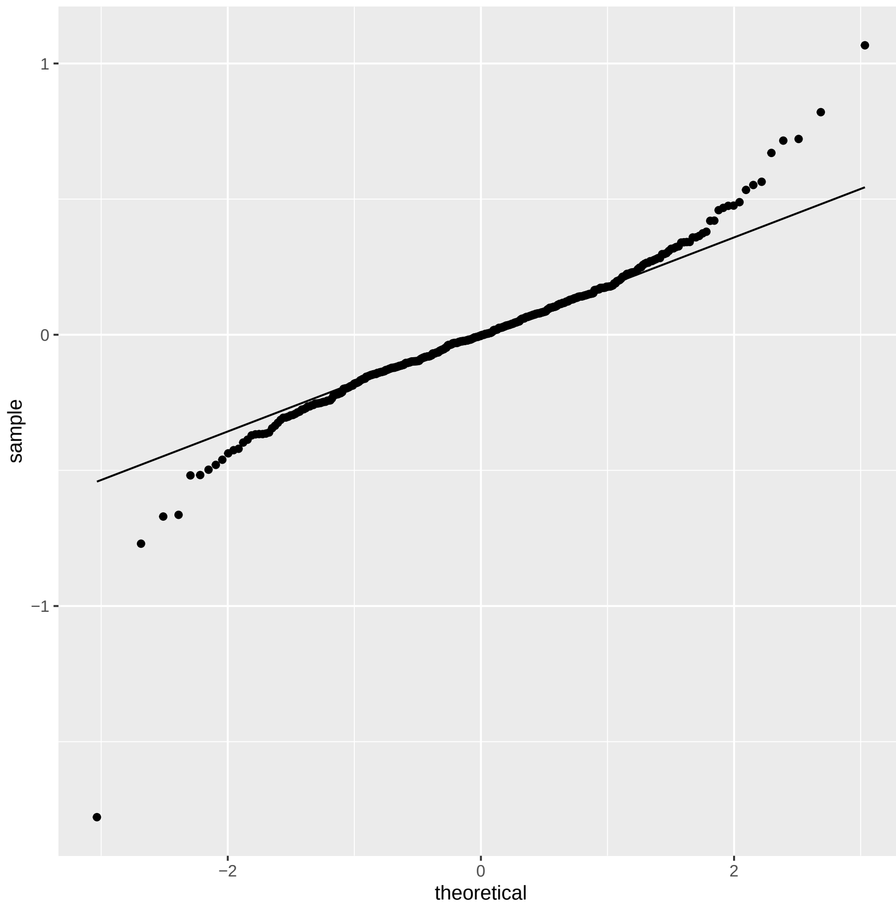

Figure 17: Taiwan houses residual plot 2

```
houses.1 %>% augment(houses) %>%
  pivot_longer(sale_date:conv) %>%
  ggplot(aes(x = value, y = .resid)) + geom_point() +
  facet_wrap(~name, scales = "free")
```

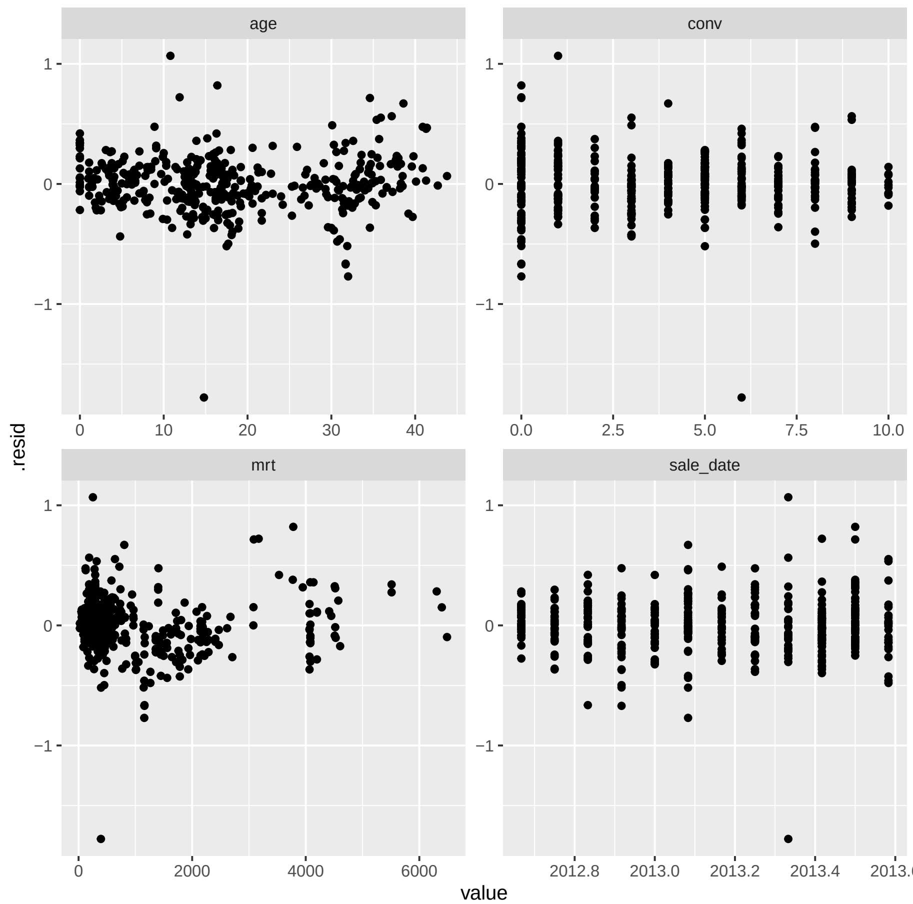

Figure 18: Taiwan houses residual plot 3

```
tibble(x = seq(0, 1, 0.01)) %>%
  mutate(dens = dgamma(x, 4.6, 17.2)) %>%
  ggplot(aes(x = x, y = dens)) + geom_line()
```

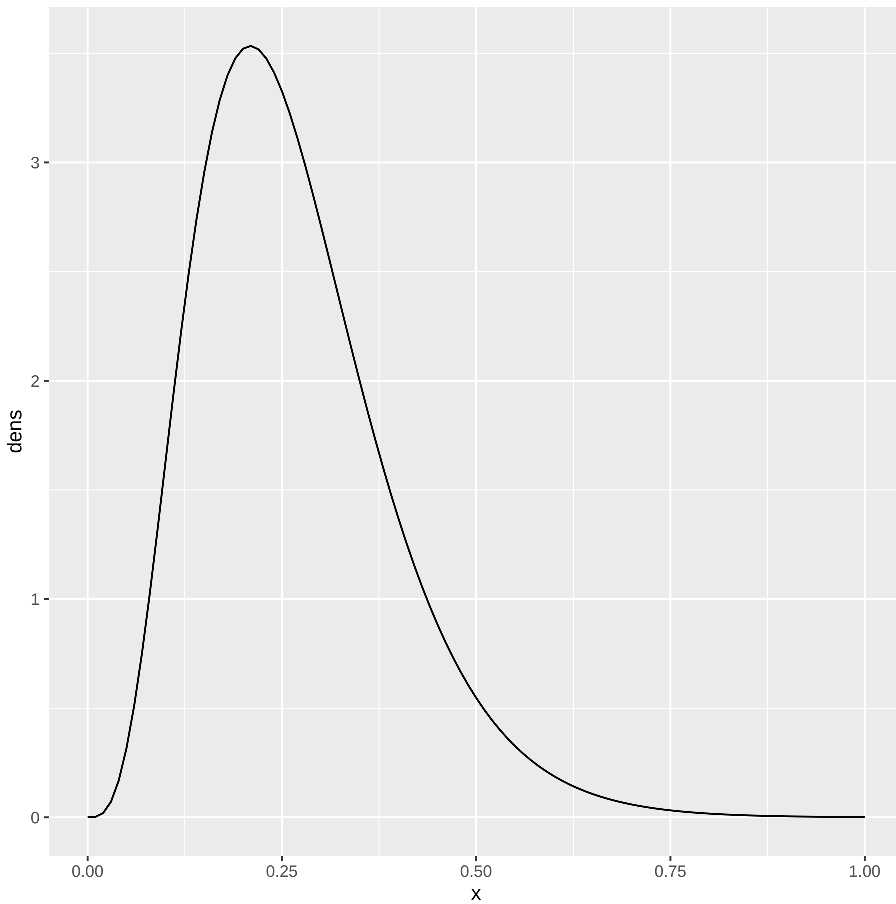

Figure 19: Prior distribution for β

```
my_y <- c(0.5, 0.7, 1.0, 2.5, 7)
```

Figure 20: Data for estimation of β

```
1 / mean(my_y)

## [1] 0.4273504
```

Figure 21: Maximum likelihood estimate of β

```
expo_fit

##  variable   mean median   sd  mad     q5    q95 rhat ess_bulk ess_tail
##      lp__ -20.70 -20.43 0.70 0.35 -22.23 -20.18 1.00     1834     2319
##      beta   0.33   0.32 0.11 0.11   0.18   0.52 1.00     1384     1701
```

Figure 22: Posterior distribution summary