Booklet of Figures
for
STAD29/STA 1007 Midterm Exam

List of Figures in this document by page:

List of Figures

1

```
library(MASS)
library(lubridate)
library(tidyverse)
library(broom)
library(survival)
library(survminer)
library(nnet)
library(car)
```

Figure 1: Packages

```
leuk

## # A tibble: 27 x 7
##    remiss  cell smear infil    li blast  temp
##     <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1       1  0.8   0.83  0.66   1.9  1.1   1
## 2       1  0.9   0.36  0.32   1.4  0.74  0.99
## 3       0  0.8   0.88  0.7    0.8  0.18  0.98
## 4       0  1     0.87  0.87   0.7  1.05  0.99
## 5       1  0.9   0.75  0.68   1.3  0.52  0.98
## 6       0  1     0.65  0.65   0.6  0.52  0.98
## 7       1  0.95  0.97  0.92   1    1.23  0.99
## 8       0  0.95  0.87  0.83   1.9  1.35  1.02
## 9       0  1     0.45  0.45   0.8  0.32  1
## 10      0  0.95  0.36  0.34   0.5  0     1.04
## # ... with 17 more rows
```

Figure 2: Leukemia remission data (some)

2

```
leuk.1 <- glm(remiss ~ cell+smear+infil+li+blast+temp, family = "binomial", data = leuk)
summary(leuk.1)

##
## Call:
## glm(formula = remiss ~ cell + smear + infil + li + blast + temp,
##     family = "binomial", data = leuk)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -1.95404  -0.66259  -0.02516   0.78184   1.57465
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)   64.25808   74.96480   0.857    0.391
## cell          30.83006   52.13520   0.591    0.554
## smear         24.68632   61.52601   0.401    0.688
## infil        -24.97447   65.28088  -0.383    0.702
## li             4.36045    2.65798   1.641    0.101
## blast         -0.01153    2.26634  -0.005    0.996
## temp        -100.17340   77.75289  -1.288    0.198
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 34.372  on 26  degrees of freedom
## Residual deviance: 21.594  on 20  degrees of freedom
## AIC: 35.594
##
## Number of Fisher Scoring iterations: 8
```

Figure 3: Leukemia logistic regression 1

```
leuk.2 <- glm(remiss ~ li, family = "binomial", data = leuk)
summary(leuk.2)

##
## Call:
## glm(formula = remiss ~ li, family = "binomial", data = leuk)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9448  -0.6465  -0.4947   0.6571   1.6971
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -3.777      1.379  -2.740  0.00615 **
## li             2.897      1.187   2.441  0.01464 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 34.372  on 26  degrees of freedom
## Residual deviance: 26.073  on 25  degrees of freedom
## AIC: 30.073
##
## Number of Fisher Scoring iterations: 4
```

Figure 4: Leukemia logistic regression 2

```
anova(leuk.2, leuk.1, test = "Chisq")

## Analysis of Deviance Table
##
## Model 1: remiss ~ li
## Model 2: remiss ~ cell + smear + infil + li + blast + temp
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1        25     26.073
## 2        20     21.594  5   4.4791   0.4827
```

Figure 5: Leukemia test

4

```
new <- tibble(li = c(0.5, 1, 1.5))
p <- predict(leuk.2, new, type = "response")
cbind(new, p)

##    li          p
## 1 0.5 0.08878928
## 2 1.0 0.29320341
## 3 1.5 0.63847599
```

Figure 6: Leukemia predictions

```
subject impairment ses life_events
1 Well high 1
2 Well high 9
3 Well high 4
4 Well high 3
5 Well low 2
6 Well high 0
7 Well low 1
8 Well high 3
9 Well high 3
10 Well high 7
11 Well low 1
12 Well low 2
13 Mild high 5
14 Mild low 6
15 Mild high 3
16 Mild low 1
17 Mild high 8
18 Mild high 2
19 Mild low 5
20 Mild high 5
21 Mild high 9
22 Mild low 3
23 Mild high 3
24 Mild high 1
25 Moderate low 0
26 Moderate high 4
27 Moderate low 3
28 Moderate low 9
29 Moderate high 6
30 Moderate low 4
31 Moderate low 3
32 Impaired high 8
33 Impaired high 2
34 Impaired high 7
35 Impaired low 5
36 Impaired low 4
37 Impaired low 4
38 Impaired high 8
39 Impaired low 8
40 Impaired low 9
```

Figure 7: Mental health data (all)

```
mh.1 <- multinom(impairment ~ ses + life_events, data = mh)
mh.2 <- polr(fct_inorder(impairment) ~ ses + life_events, data = mh)
mh.3 <- polr(impairment ~ ses + life_events, data = mh)
```

Figure 8: Possible models for mental health data

```
drop1(mh.2, test = "Chisq")

## Single term deletions
##
## Model:
## fct_inorder(impairment) ~ ses + life_events
##             Df    AIC    LRT Pr(>Chi)
## <none>         109.10
## ses          1 110.53 3.4292 0.064054 .
## life_events  1 114.87 7.7765 0.005293 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 9: More output for mental health data

```
mh.4 <- update(mh.2, .~. - ses)
drop1(mh.4, test = "Chisq")

## Single term deletions
##
## Model:
## fct_inorder(impairment) ~ life_events
##             Df    AIC   LRT Pr(>Chi)
## <none>         110.53
## life_events  1 115.04 6.515   0.0107 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 10: Another model for mental health data

```
new <- tibble(life_events = c(2, 4, 6))
p <- predict(mh.4, new, type = "probs")
cbind(new, p)

##   life_events      Well      Mild  Moderate  Impaired
## 1           2 0.4220346 0.3245379 0.1354544 0.1179730
## 2           4 0.2910507 0.3324785 0.1842907 0.1921801
## 3           6 0.1875290 0.2946523 0.2204890 0.2973297
```

Figure 11: Predictions for mental health data

	A	B	C
1	treatment	died	time
2	linoleic_acid	no	1
3	linoleic_acid	no	5
4	linoleic_acid	yes	6
5	linoleic_acid	yes	6
6	linoleic_acid	no	9
7	linoleic_acid	yes	10
8	linoleic_acid	yes	10
9	linoleic_acid	no	10
10	linoleic_acid	yes	12
11	linoleic_acid	yes	12
12	linoleic_acid	yes	12
13	linoleic_acid	yes	12
14	linoleic_acid	no	12
15	linoleic_acid	no	13
16	linoleic_acid	no	15
17	linoleic_acid	no	16
18	linoleic_acid	no	20
19	linoleic_acid	yes	24
20	linoleic_acid	no	24
21	linoleic_acid	no	27
22	linoleic_acid	yes	32
23	linoleic_acid	no	34
24	linoleic_acid	no	36
25	linoleic_acid	no	36
26	linoleic_acid	no	44
27	control	no	3
28	control	yes	6

Figure 12: Screenshot of spreadsheet with cancer data (some)

```
cancer %>% mutate(y = Surv(time, died == "yes")) -> cancer
```

Figure 13: Creating a new column in cancer dataframe

```
cancer.1 <- coxph(y ~ treatment, data = cancer)
summary(cancer.1)

## Call:
## coxph(formula = y ~ treatment, data = cancer)
##
##   n= 49, number of events= 22
##
##                          coef exp(coef) se(coef)      z Pr(>|z|)
## treatmentlinoleic_acid -0.2487    0.7798   0.4303 -0.578    0.563
##
##                        exp(coef) exp(-coef) lower .95 upper .95
## treatmentlinoleic_acid    0.7798      1.282    0.3355     1.812
##
## Concordance= 0.527  (se = 0.062 )
## Likelihood ratio test= 0.34  on 1 df,   p=0.6
## Wald test            = 0.33  on 1 df,   p=0.6
## Score (logrank) test = 0.34  on 1 df,   p=0.6

drop1(cancer.1, test = "Chisq")

## Single term deletions
##
## Model:
## y ~ treatment
##           Df    AIC     LRT Pr(>Chi)
## <none>       143.93
## treatment  1 142.26 0.33579   0.5623
```

Figure 14: Cox model for cancer data

```
plants

## # A tibble: 15 x 4
##    Treatment Height Width Weight
##    <chr>      <dbl> <dbl>  <dbl>
##  1 A           15.8   3.9   29.4
##  2 A           15.1   3.8   29.9
##  3 A           14.8   4.1   30.2
##  4 A           14.4   4.7   30.1
##  5 A           15.1   3.7   30.9
##  6 B           15.7   4.8   31.2
##  7 B           15.9   4.3   31.4
##  8 B           15.4   4.5   31.8
##  9 B           16.7   5.4   32.4
## 10 B           16.9   5.8   32.6
## 11 C           16.5   5.4   33.2
## 12 C           16.9   5.8   33.4
## 13 C           17.1   6.4   33.1
## 14 C           17.3   4.2   32.9
## 15 C           17.5   6.7   33.1
```

Figure 15: Plants data

```
plants.1 <- manova(y~Treatment, data = plants)
summary(plants.1)

##            Df  Pillai approx F num Df den Df  Pr(>F)
## Treatment   2 0.91674    3.103      6     22 0.02343 *
## Residuals  12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 16: MANOVA for plants data

11

```
plants.2 <- lda(Treatment ~ Weight + Width + Height, data = plants)
plants.2

## Call:
## lda(Treatment ~ Weight + Width + Height, data = plants)
##
## Prior probabilities of groups:
##         A         B         C
## 0.3333333 0.3333333 0.3333333
##
## Group means:
##    Weight Width Height
## A  30.10  4.04  15.04
## B  31.88  4.96  16.12
## C  33.14  5.70  17.06
##
## Coefficients of linear discriminants:
##                 LD1         LD2
## Weight  1.79365123 -1.3000104
## Width  -0.09653954  0.3631087
## Height  0.75212731  1.6673815
##
## Proportion of trace:
##   LD1   LD2
## 0.999 0.001
```

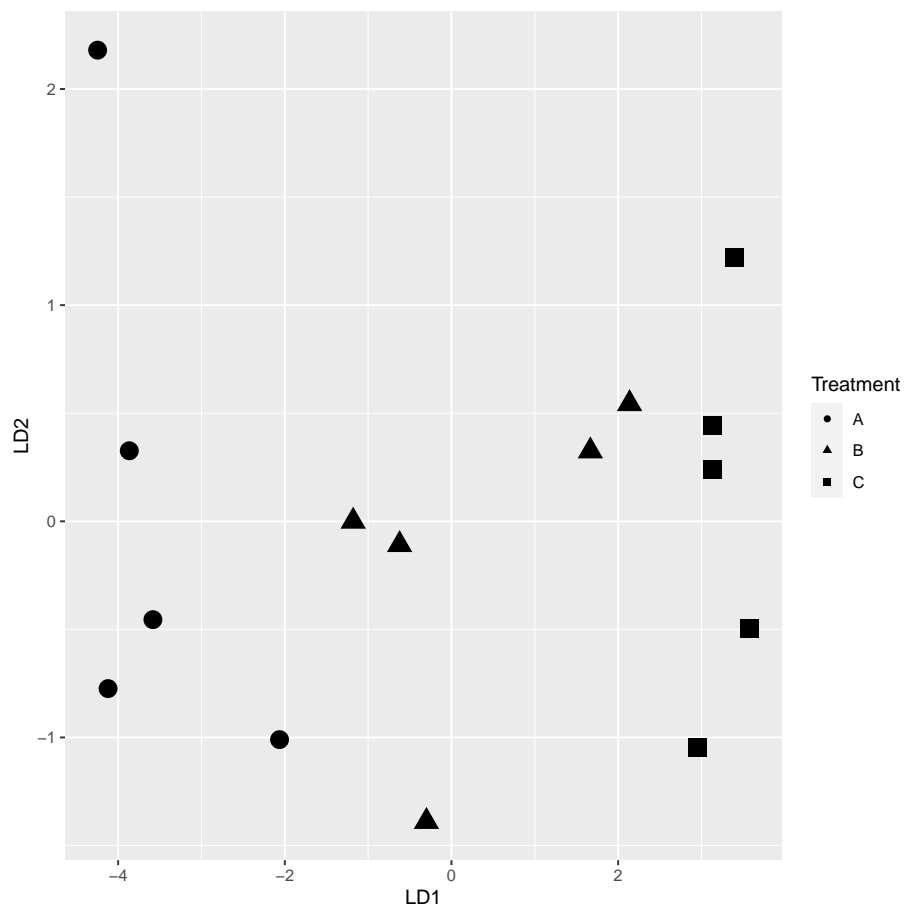Figure 17: Discriminant analysis for plants data

Figure 18: Plot of discriminant scores for plants data. Treatments are shown by shape (circle, triangle, square).

```
growth

## # A tibble: 27 x 6
##      sub sex     d08    d10    d12    d14
##    <dbl> <chr> <dbl> <dbl> <dbl> <dbl>
## 1      1 F      21     20    21.5  23
## 2      2 F      21     21.5  24    25.5
## 3      3 F      20.5   24    24.5  26
## 4      4 F      23.5   24.5  25    26.5
## 5      5 F      21.5   23    22.5  23.5
## 6      6 F      20     21    21    22.5
## 7      7 F      21.5   22.5  23    25
## 8      8 F      23     23    23.5  24
## 9      9 F      20     21    22    21.5
## 10    10 F      16.5   19    19    19.5
## # ... with 17 more rows
```

Figure 19: Growth of children data (some)

```
growth %>%
  pivot_longer(starts_with("d"), names_to = "age", values_to = "distance") %>%
  group_by(age, sex) %>%
  summarize(mean_distance = mean(distance)) %>%
  ggplot(aes(x = age, y = mean_distance, linetype = sex, group = sex)) +
    geom_point() + geom_line()
```
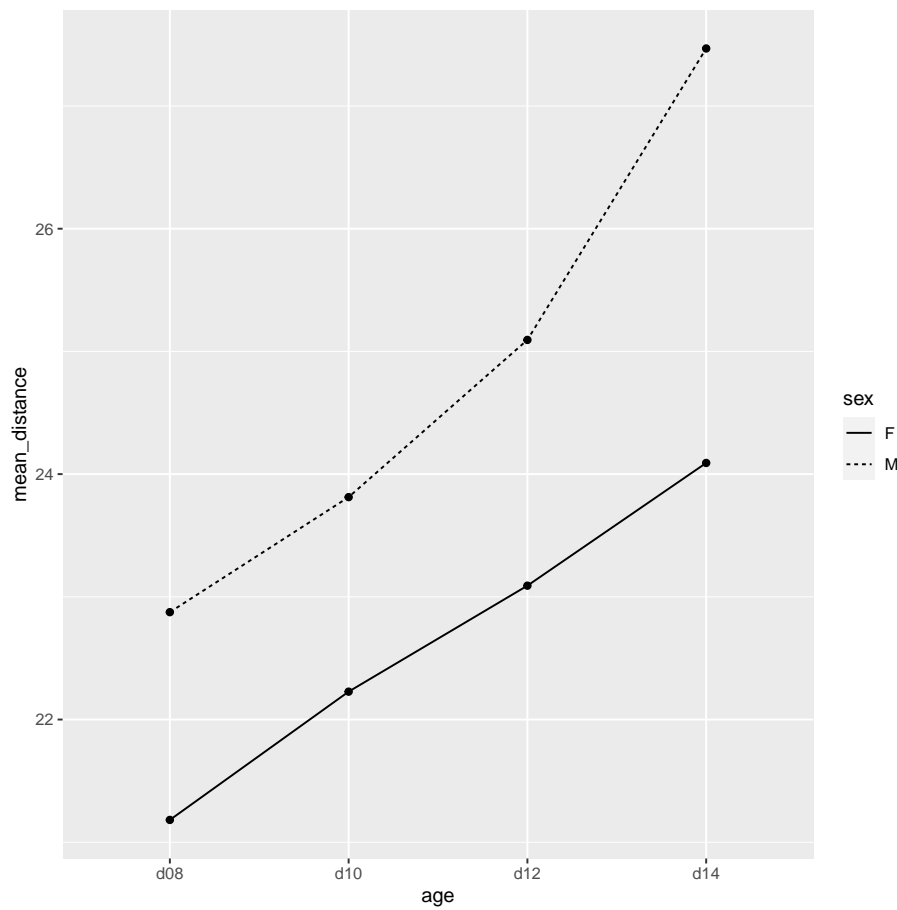

Figure 20: Graph of mean distance by gender and age

```
y <- with(growth, cbind(d08, d10, d12, d14))
ages <- colnames(y)
ages.df <- data.frame(ages = factor(ages))
growth.1 <- lm(y~sex, data = growth)
growth.2 <- Manova(growth.1, idata = ages.df, idesign = ~ages)
growth.2

##
## Type II Repeated Measures MANOVA Tests: Pillai test statistic
##             Df test stat approx F num Df den Df    Pr(>F)
## (Intercept)  1   0.99360   3882.7      1     25 < 2.2e-16 ***
## sex          1   0.23479      7.7      1     25   0.01043 *
## ages         1   0.82602     36.4      3     23 6.673e-09 ***
## sex:ages     1   0.28936      3.1      3     23   0.04561 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 21: Repeated measures analysis of children's growth data

```
cancer %>% count(treatment) -> new
new

## # A tibble: 2 x 2
##   treatment         n
##   <chr>         <int>
## 1 control          24
## 2 linoleic_acid    25

s <- survfit(cancer.1, newdata = new, data = cancer)
ggsurvplot(s, conf.int = FALSE)
```
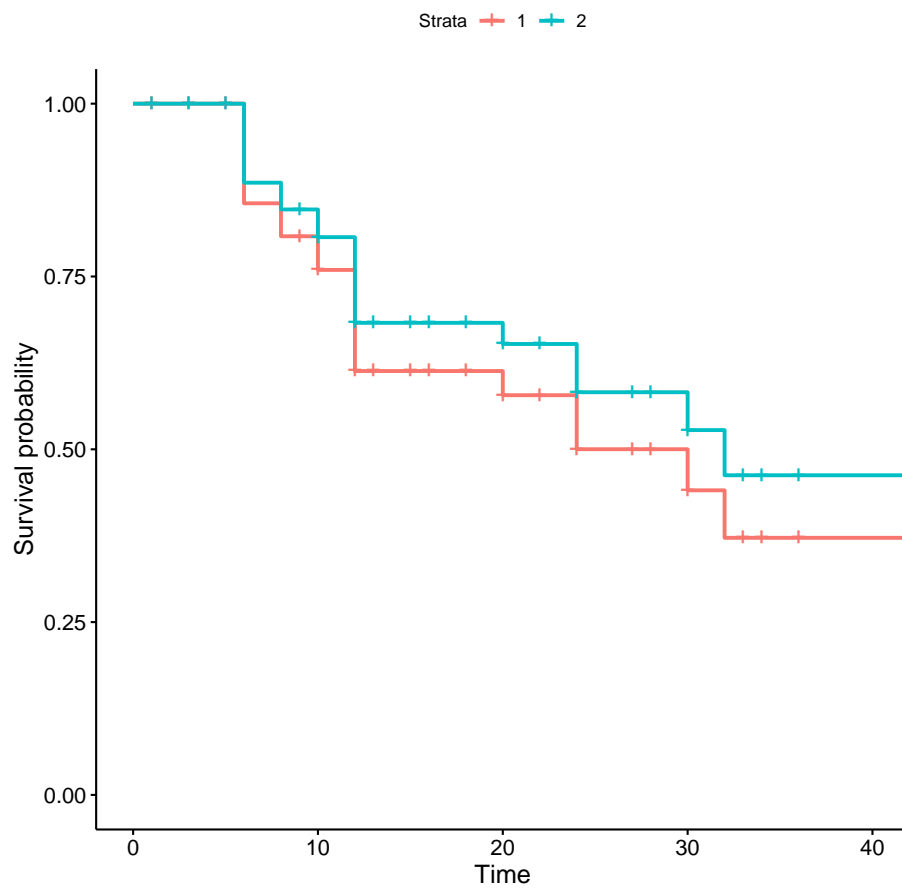

Figure 22: Estimated survival curves