C240424_11

YOUSEF IBRAHIM GOMAA MAHMOUD MABROUK

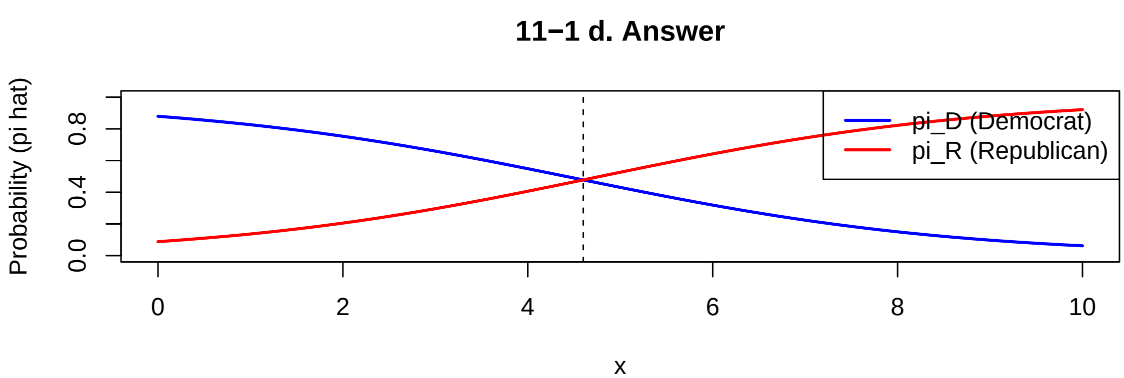
2025-01-23

Exercise 11-1

a. The "Independent" group, denoted by I.
b. $log(\frac{\hat{\pi}_R}{\hat{\pi}_D}) = log(\frac{\hat{\pi}_R}{\hat{\pi}_I}) - log(\frac{\hat{\pi}_D}{\hat{\pi}_I}) = (1.0 - 3.3) + [0.3 - (-0.2)] = -2.3 + 0.5x$
Which means that the odds of being Republican (R) to Democrat (D) are multiplied by $e^{0.5} \approx 1.65$ for each $10,000 increase in annual income. This means that as income increases, the relative preference for Republicans over Democrats increases by a factor of 1.65.
c. $\hat{\pi}_I = \frac{1}{1+exp(1.0+0.3x)+exp(3.3-0.2x)}$
d. The range be calculated by solving $log(\frac{\hat{\pi}_R}{\hat{\pi}_D}) > 0$:

- -2.3 + 0.5x > 0
- 0.5x > 2.3
- x > 4.6 and when plotted:

```
x <- seq(0, 10, by = 0.1)
pi_D <- exp(3.3 - 0.2 * x) / (1 + exp(3.3 - 0.2 * x) + exp(1.0 + 0.3 * x))
pi_R <- exp(1.0 + 0.3 * x) / (1 + exp(3.3 - 0.2 * x) + exp(1.0 + 0.3 * x))
plot(x, pi_D, type = "l", col = "blue", lwd = 2, ylim = c(0, 1),
     ylab = "Probability (pi hat)", main = "11-1 d. Answer")
lines(x, pi_R, col = "red", lwd = 2)
abline(v=4.6, col = "black", lty = 2)
legend("topright", legend = c("pi_D (Democrat)", "pi_R (Republican)"),
       col = c("blue", "red"), lty = 1, lwd = 2)
```



From the plot, the probability of preferring Republicans (red curve) and the probability of preferring Democrats (blue curve) intersect at x=4.6, which means for the range x > 4.6, individuals are relatively likely to prefer republicans over democrats.

Exercise 11-2

The estimated prediction equation is $logit[\hat{P}(Y <= j)] = \hat{\alpha}_j - 0.54x_1 + 0.60x_2 + 1.19x_3$. a. (i) As x1 increases, the odds of being in a lower job satisfaction (i.e., Y<=j) category than a higher category ~(i.e., Y>j) decrease by a factor of $e^{-0.54} \approx 0.583$. Meaning that higher earnings are associated with higher job satisfaction. (ii) As x2 increases, the odds of being in a lower job satisfaction category (vs. a higher category) times the estimated odds for someone with a one-unit lower value of x2. ($e^{0.60}$) As freedom about how you work decreases (x2 increasing), job satisfaction tends to decrease. (iii) As the work environment becomes less conducive to productivity (x3 increasing), the odds of being in a lower job satisfaction category against a higher category increase by a factor of 3.287. ($e^{1.19}$) As the work environment becomes less conducive to productivity, job satisfaction decreases.

b. In order to maximize job satisfaction P(Y=j), P(Y<=j-1) should be minimized. $\because P(Y = c) = 1 - P(Y <= c - 1)$ As such, value of x1 should be as maximized (x1=4), while x2 and x3 should be minimized. (x2=x3=1) $\therefore P(Y = c) = 1 - \hat{\alpha}_j - 0.54(4) + 0.60(1) + 1.19(1) = \hat{\alpha}_j - 0.37$

Exercise 11-3

a.
```r
library(VGAM)
```

```
## Loading required package: stats4

## Loading required package: splines
```

```r
happy <- read.csv("happy.csv")

happy.fit <- vglm(happiness ~ income,
                  family = multinomial(refLevel = "very"),
                  data = happy)

summary(happy.fit)
```

```
## Call:
## vglm(formula = happiness ~ income, family = multinomial(refLevel = "very"),
##     data = happy)
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept):1 -2.55518    0.72559  -3.522 0.000429 ***
## (Intercept):2 -0.35129    0.26837  -1.309 0.190554
## income:1      -0.22751    0.34119  -0.667 0.504903
## income:2      -0.09615    0.12202  -0.788 0.430694
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
##
## Residual deviance: 926.2112 on 1198 degrees of freedom
##
## Log-likelihood: -463.1056 on 1198 degrees of freedom
```

```
##
## Number of Fisher scoring iterations: 6
##
## Warning: Hauck-Donner effect detected in the following estimate(s):
## '(Intercept):1'
##
##
## Reference group is level  3  of the response
```

b.
- $\hat{\pi}_{not} = \frac{exp(-2.56-0.23income)}{1+exp(-2.56-0.23income)+exp(-0.35-0.10income)}$
- $\hat{\pi}_{pretty} = \frac{exp(-0.35-0.10income)}{1+exp(-2.56-0.23income)+exp(-0.35-0.10income)}$
- $\hat{\pi}_{very} = \frac{1}{1+exp(-2.56-0.23income)+exp(-0.35-0.10income)}$

c. First equation: $logit(\frac{\hat{\pi}_{not}}{\hat{\pi}_{very}}) = -2.56 - 0.23income$

- As income increases, the log-odds of being 'not happy' (vs 'very happy') decrease by a factor of $e^{-0.23} \approx 0.811$. Which means that the probability of being unhappy decreases as income decreases.

d.
```r
happy_1.fit <- vglm(happiness ~ 1,
                    family = multinomial(refLevel = "very"),
                    data = happy)
lrtest(happy.fit, happy_1.fit)
```

```
## Likelihood ratio test
##
## Model 1: happiness ~ income
## Model 2: happiness ~ 1
##    #Df  LogLik Df  Chisq Pr(>Chisq)
## 1 1198 -463.11
## 2 1200 -463.58  2 0.9439     0.6238
```

- p_value = $0.6238 > 0.05$, as such, income has no significant effect on marital happiness, and we fail to reject the null hypothesis and conclude that marital happiness is independent of family income.

e. Average income is denoted by 2.

```r
predict(happy.fit,
        newdata = data.frame(income = 2),
        type = "response")
```

```
##         not    pretty      very
## 1 0.0302356 0.3562457 0.6135187
```

Exercise 11-4

a.
```r
happyordinal.fit <- vglm(happiness ~ income,
                         family = cumulative(parallel = TRUE),
                         data = happy)
summary(happyordinal.fit)
```

```
## Call:
## vglm(formula = happiness ~ income, family = cumulative(parallel = TRUE),
##     data = happy)
##
## Coefficients:
##                Estimate Std. Error z value Pr(>|z|)
## (Intercept):1  -3.2466     0.3404  -9.537   <2e-16 ***
## (Intercept):2  -0.2378     0.2592  -0.917    0.359
## income         -0.1117     0.1179  -0.948    0.343
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Names of linear predictors: logitlink(P[Y<=1]), logitlink(P[Y<=2])
##
## Residual deviance: 926.2674 on 1199 degrees of freedom
##
## Log-likelihood: -463.1337 on 1199 degrees of freedom
##
## Number of Fisher scoring iterations: 3
##
## No Hauck-Donner effect found in any of the estimates
##
##
## Exponentiated coefficients:
##    income
## 0.8942748
```

b. There are two intercepts because the outcome variable happiness has three ordered categories (c-1), while there is only one income effect as it is the same across all cumulative logits with only coefficient.

c. The coefficient for income is interpreted as follows: For $1 <= j < 3$, the estimated odds that a family is unhappy about their marriage (i.e, Y<=j) rather than being happy (i.e., Y>j) are multiplied by $e^{-0.112} \approx 0.894$ per 1 unit increase of j. (income)

d.
```r
happyordinal_1.fit <- vglm(happiness ~ 1,
                         family = cumulative(parallel = TRUE),
                         data = happy)
lrtest(happyordinal.fit, happyordinal_1.fit)
```

```
## Likelihood ratio test
##
## Model 1: happiness ~ income
## Model 2: happiness ~ 1
##    #Df  LogLik Df  Chisq Pr(>Chisq)
## 1 1199 -463.13
## 2 1200 -463.58  1 0.8876     0.3461
```

- p_value = 0.3461 > 0.05, similar to the result of the nominal analysis, income has no significant effect on marital happiness, and we fail to reject the null hypothesis and conclude that marital happiness is independent of family income.

e. Average income is denoted by 2.

```r
predict(happyordinal.fit,
        newdata = data.frame(income = 2),
        type = "response")
```

```
##          not    pretty      very
## 1 0.03017419 0.3565176 0.6133082
```