ST5202_Tut5

Name:Zhu Xu
User ID:E0337988
Matriculation ID:A0191344H

2019/4/17

Question 1:

a)

```
job = read.table(
'/Users/xuzhu/Desktop/Notes/Sem2/ST5202-Applied_Regression_Analysis/Tut/job_proficiency.txt')
colnames(job) = c('y', 'x1', 'x2', 'x3', 'x4')
fm = lm(log(y)~., job)
fmfull = lm(log(y)~1, job)
anova(fmfull, fm)
```

```
## Analysis of Variance Table
##
## Model 1: log(y) ~ 1
## Model 2: log(y) ~ x1 + x2 + x3 + x4
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     24 1.11873
## 2     20 0.04977  4     1.069 107.38 3.208e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$
$$H_a : \beta_1 \neq 0 \quad or \quad \beta_2 \neq 0 \quad or \quad \beta_3 \neq 0 \quad or \quad \beta_4 \neq 0$$
$$F^* = \frac{SSE(R) - SSE(F)}{df_R - df_F} / \frac{SSE(F)}{df_F} \sim F(4, 20)$$

Thus the rejection region is

$$(F(0.95; 4, 20), +\infty) \Rightarrow (2.866, +\infty)$$

$F^* = 107.38 > 2.866$, therefore we reject H_0

c)

```
fm13 = lm(log(y)~x1+x2+I(x1^2)+I(x2^2)+x1:x2, job)
step(fm13,
scope=list(lower=~1,upper=~x1+x2+I(x1^2)+I(x2^2)+x1:x2), direction='backward')
```

```
## Start:  AIC=-85.63
## log(y) ~ x1 + x2 + I(x1^2) + I(x2^2) + x1:x2
##
##            Df Sum of Sq     RSS     AIC
## - I(x2^2)  1   0.032705 0.53619 -86.053
## <none>                   0.50349 -85.627
## - I(x1^2)  1   0.054906 0.55839 -85.039
## - x1:x2    1   0.075114 0.57860 -84.150
##
## Step:  AIC=-86.05
## log(y) ~ x1 + x2 + I(x1^2) + x1:x2
##
##            Df Sum of Sq     RSS     AIC
## <none>                   0.53619 -86.053
## - I(x1^2)  1   0.044659 0.58085 -86.053
## - x1:x2    1   0.079901 0.61609 -84.581
##
## Call:
## lm(formula = log(y) ~ x1 + x2 + I(x1^2) + x1:x2, data = job)
##
## Coefficients:
## (Intercept)           x1           x2      I(x1^2)        x1:x2
##   5.096e+00   -2.228e-03   -1.889e-02   -9.111e-05    2.405e-04
```

Question 2:

a)

```
data = read.table('/Users/xuzhu/Desktop/Notes/Sem2/ST5202-Applied_Regression_Analysis/Tut/annual_dues.t:
colnames(data)=c('y', 'x')
logit_fm = glm(y~x, data, family=binomial('logit'))
summary(logit_fm)
```

```
##
## Call:
## glm(formula = y ~ x, family = binomial("logit"), data = data)
##
## Deviance Residuals:
```

```
##       Min       1Q   Median       3Q      Max
## -1.7651  -1.0012   0.6502   0.9828   1.6455
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.80751    2.65576  -1.810   0.0703 .
## x            0.12508    0.06676   1.874   0.0610 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 41.455  on 29  degrees of freedom
## Residual deviance: 37.465  on 28  degrees of freedom
## AIC: 41.465
##
## Number of Fisher Scoring iterations: 4
```

$$logit(\hat{\pi}) = log(\frac{\hat{\pi}}{1-\hat{\pi}}) = -4.8075 + 0.1251x$$

b)

$e^{b_1} = e^{0.1251} = 1.333$

c)

$x = 40, log(\frac{\hat{\pi}}{1-\hat{\pi}}) = 0.1957 \Rightarrow \quad \hat{\pi} = \frac{e^{0.1957}}{1+e^{0.1957}} = 0.5488$

d)

$\pi = 0.75 \quad \Rightarrow log(\frac{\pi}{1-\pi}) = 1.0986 \quad \Rightarrow \hat{x} = 47.2188$