Homework 4, MATH455: Due Monday, 03/05/2018

Your Name: (replace this)

March 6, 2018

Instructions: The homework assignment editing this LaTeX document. Download the LaTeX source from the class web page and study it to learn more about LaTeX. Replace the text with appropriate information. Run "pdflatex" on this document.

You will submit this assignment in two parts:

1. Print out the PDF file and bring it to class, and
2. Send an e-mail to:

   gang@math.binghamton.edu

   before class on the due date with two attachments:

   - The LaTeX source file, and
   - The generated PDF document.

Please complete the following:

1. Read chapter 3 and finish questions 3.2, 3.4 (on pages 49-50) in this chapter.

```
data(cheddar)
cheeseMod = lm(taste~Acetic+H2S+Lactic,cheddar)
summary(cheeseMod)
Call:
lm(formula = taste ~ Acetic + H2S + Lactic, data = cheddar)

Residuals:
Min      1Q  Median      3Q     Max
-17.390  -6.612  -1.009   4.908  25.449

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -28.8768    19.7354  -1.463  0.15540
Acetic        0.3277     4.4598   0.073  0.94198
H2S           3.9118     1.2484   3.133  0.00425 **
Lactic       19.6705     8.6291   2.280  0.03108 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 10.13 on 26 degrees of freedom
Multiple R-squared:  0.6518,Adjusted R-squared:  0.6116
F-statistic: 16.22 on 3 and 26 DF,  p-value: 3.81e-06
```

Accordingly, H2S and Lactic are the two parameters significant at the 5 percent level.
After applying the exponential function to both Acetic and H2S, we get the following results

```
> cheeseModP = lm(taste~exp(Acetic)+exp(H2S)+Lactic,cheddar)
> summary(cheeseModP)

Call:
lm(formula = taste ~ exp(Acetic) + exp(H2S) + Lactic, data = cheddar)

Residuals:
Min      1Q  Median      3Q     Max
-16.209  -7.266  -1.651   7.385  26.335

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.897e+01  1.127e+01  -1.684   0.1042
exp(Acetic)  1.891e-02  1.562e-02   1.210   0.2371
exp(H2S)     7.668e-04  4.188e-04   1.831   0.0786 .
```

```
Lactic       2.501e+01  9.062e+00   2.760   0.0105 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 11.19 on 26 degrees of freedom
Multiple R-squared:  0.5754,Adjusted R-squared:  0.5264
F-statistic: 11.75 on 3 and 26 DF,  p-value: 4.746e-05
```

Thus, only Lactic remains statistically significant.
We can not operate the f-test on these data sets

```
anova(cheeseMod,cheeseModP)Analysis of Variance Table

Model 1: taste ~ Acetic + H2S + Lactic
Model 2: taste ~ exp(Acetic) + exp(H2S) + Lactic
Res.Df    RSS Df Sum of Sq F Pr(>F)
1     26 2668.4
2     26 3253.6  0    -585.2
```

This is because our degrees of freedom are the same, and thus we are dividing by zero and will be unable to compute anything.
According to our summary, H2S=3.9118, thus for every increase of .01, we increase taste by .039 approximately.

```
> log(10)
[1] 2.302585
> log(10.01)
[1] 2.303585
> log(10.01)/log(10)
[1] 1.000434
```

So about a .04 percent increase given an additive of .01 on the log scale.

```
> scores = lm(total~expend+ratio+salary,sat)
> scoresSZ = lm(total~expend+ratio,sat)
> scoresNull=lm(total~1,sat)
> anova(scores,scoresSZ)
Analysis of Variance Table

Model 1: total ~ expend + ratio + salary
Model 2: total ~ expend + ratio
Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     46 216812
2     47 233443 -1    -16631 3.5285 0.06667 .
---
```

```
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
> anova(scores,scoresNull)
Analysis of Variance Table

Model 1: total ~ expend + ratio + salary
Model 2: total ~ 1
Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     46 216812
2     49 274308 -3    -57496 4.0662 0.01209 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Accordingly, it appears that with $H_0 : \beta_{salary} = 0$, we can not reject that null hypothesis and salary may not be indicative. Meanwhile, all three parameters do seem to have some indication on total score.

```
> anova(tscores,scores)
Analysis of Variance Table

Model 1: total ~ expend + ratio + salary + takers
Model 2: total ~ expend + ratio + salary
Res.Df    RSS Df Sum of Sq      F    Pr(>F)
1     45  48124
2     46 216812 -1   -168688 157.74 2.607e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
> summary(tscores)

Call:
lm(formula = total ~ expend + ratio + salary + takers, data = sat)

Residuals:
Min      1Q  Median      3Q     Max
-90.531 -20.855  -1.746  15.979  66.571

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 1045.9715    52.8698  19.784  < 2e-16 ***
expend         4.4626    10.5465   0.423    0.674
ratio         -3.6242     3.2154  -1.127    0.266
salary         1.6379     2.3872   0.686    0.496
takers        -2.9045     0.2313 -12.559 2.61e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

```
Residual standard error: 32.7 on 45 degrees of freedom
Multiple R-squared:  0.8246,Adjusted R-squared:  0.809
F-statistic: 52.88 on 4 and 45 DF,  p-value: < 2.2e-16
```

as we can see, the t-value demonstrated in summary is the same as the F value provided by anova, which demonstrates their equivalence

2. Read chapter 4 and finish questions 4.1, 4.5 (on pages 56-58) in this chapter.
3. Read chapter 7.3 and finish question 7.8 (on page 111) in this chapter.