Lecture 8

Example 1 Sample size $n = 100$. $x_i, e_i \sim N(0,1)$

$y_i = x_i + e_i$ ($E[y|x] = \beta_0 + \beta_1 x$ with $\beta_0 = 0$, $\beta_1 = 1$)

```
> x <- rnorm(100)
>  e <- rnorm(100)
>  y <- x+e
>  Model1 <- lm(y~x)
>  Model2 <- lm(y~x + I(x^2))
>  Model3 <- lm(y~x + I(x^2) + I(x^3))
>
> summary(Model1)

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max
-3.3898 -0.7832  0.1945  0.7032  3.0001

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.01262    0.10737   0.118    0.907
x            1.07479    0.10654  10.088   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.056 on 98 degrees of freedom
Multiple R-squared:  0.5094,  Adjusted R-squared:  0.5044
F-statistic: 101.8 on 1 and 98 DF,  p-value: < 2.2e-16
```

$H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$ (same)

$H_0: E[y|x] = \beta_0$ vs $H_1: E[y|x] = \beta_0 + \beta_1 x$ (same)

half is explained by x, half is by e

$0.5094 = 0.088^2$; $2-1$; $100-(1+1)$

(Reject)

```
> summary(Model2)

Call:
lm(formula = y ~ x + I(x^2))

Residuals:
    Min      1Q  Median      3Q     Max
-3.3976 -0.7857  0.1962  0.6969  3.0042

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.006729   0.136642   0.049    0.961
x           1.071566   0.116515   9.197 7.28e-15 ***
I(x^2)      0.006380   0.090755   0.070    0.944
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.061 on 97 degrees of freedom
Multiple R-squared:  0.5095,  Adjusted R-squared:  0.4993
F-statistic: 50.37 on 2 and 97 DF,  p-value: 9.954e-16
```

$E[y|x] = \beta_0 + \beta_2 x^2$; $E[y|x] = \beta_0 + \beta_1 x + \beta_2 x^2$

$H_0: \beta_1 = 0$ (Reject) vs $H_1: \beta_1 \neq 0$

$H_0: \beta_2 = 0$ (Accept) vs $H_1: \beta_2 \neq 0$

$H_0: E[y|x] = \beta_0$ vs $H_1: E[y|x] = \beta_0 + \beta_1 x + \beta_2 x^2$ (Reject)

$3-1$; $100-(2+1)$

```
> summary(Model3)

Call:
lm(formula = y ~ x + I(x^2) + I(x^3))
```

```
Residuals:
    Min      1Q  Median      3Q     Max
-3.4012 -0.7753  0.1866  0.6933  2.9917

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.001833   0.141756   0.013    0.990
x            1.092728   0.191718   5.700 1.32e-07 ***
I(x^2)       0.013926   0.106066   0.131    0.896
I(x^3)      -0.009714   0.069672  -0.139    0.889
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.067 on 96 degrees of freedom
Multiple R-squared:  0.5096,  Adjusted R-squared:  0.4942
F-statistic: 33.25 on 3 and 96 DF,  p-value: 7.993e-15
```

Example 2

```
> x <- rnorm(100)
> y <- 2*x+rnorm(100)
>
> Model1 <- lm(y ~ x)
> Model2 <- lm(y ~ x + I(x^2))
> Model3 <- lm(y ~ x + I(x^2) + I(x^3))
>
> summary(Model2)

Call:
lm(formula = y ~ x + I(x^2))

Residuals:
     Min       1Q   Median       3Q      Max
-2.16954 -0.71778 -0.09017  0.71397  2.56037

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.04546    0.12146   0.374    0.709
x            1.98096    0.09620  20.593   <2e-16 ***
I(x^2)       0.10624    0.06872   1.546    0.125
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9518 on 97 degrees of freedom
Multiple R-squared:  0.8258,    Adjusted R-squared:  0.8223
F-statistic:   230 on 2 and 97 DF,  p-value: < 2.2e-16

>
> anova(Model1, Model2)
Analysis of Variance Table

Model 1: y ~ x
Model 2: y ~ x + I(x^2)
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     98 90.043
```

$y_i = 2x_i + \epsilon_i \quad (\beta_0 = 0, \beta_1 = 2)$

$H_0 : \beta_2 = 0 \quad vs \quad H_1 : \beta_2 \neq 0$

H_0 : Model 1 : $E[Y|x] = \beta_0 + \beta_1 x$

H_1 : Model 2 : $E[Y|x] = \beta_0 + \beta_1 x + \beta_2 x^2$

Should be the same

```
2     97 87.878  1    2.1653 2.3901 0.1254
>

> summary(Model3)

Call:
lm(formula = y ~ x + I(x^2) + I(x^3))

Residuals:
     Min       1Q   Median       3Q      Max
-1.82899 -0.75960 -0.02243  0.74387  2.49038

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.09454    0.12562   0.753    0.454
x            2.17362    0.16557  13.128   <2e-16 ***
I(x^2)       0.05192    0.07826   0.663    0.509
I(x^3)      -0.07269    0.05098  -1.426    0.157
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9468 on 96 degrees of freedom
Multiple R-squared:  0.8295,    Adjusted R-squared:  0.8241
F-statistic: 155.6 on 3 and 96 DF,  p-value: < 2.2e-16

>
> anova(Model1, Model3)
Analysis of Variance Table

Model 1: y ~ x
Model 2: y ~ x + I(x^2) + I(x^3)
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     98 90.043
2     96 86.055  2    3.9879 2.2244 0.1137
>
>
> anova(Model3)
Analysis of Variance Table
Response: y
          Df  Sum Sq Mean Sq  F value Pr(>F)
x          1 414.54  414.54 462.4504 <2e-16 ***
I(x^2)     1   2.17    2.17   2.4155 0.1234
I(x^3)     1   1.82    1.82   2.0332 0.1571
Residuals 96  86.06    0.90
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
> library(car)
> Anova(Model3)
Anova Table (Type II tests)

Response: y
            Sum Sq Df F value  Pr(>F)
x          154.501  1 172.3558 <2e-16 ***
I(x^2)       0.395  1   0.4401 0.5086
I(x^3)       1.823  1   2.0332 0.1571
Residuals   86.055 96
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```