

R Handout - Indicator Variable Regression

Mar 2014, Vermont CFWRU Workshop

Dr. Derek Ogle

Northland College

Preliminaries

```
> library(FSA)      # for Subset(), fitPlot()
```

Lake Trout Data

Curtis (1990) examined the population dynamics related to the recovery of an offshore lake trout population near Stannard Rock, Lake Superior. Relative abundance of lake trout greater than 43.2 cm long was recorded as the CPE (fish caught per 50,000 m of 114.3-mm-mesh gill net) of each age group in each year.

Year	Age-Group								
	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
1959	64	219	241	121	33	9	1	0.5	1
1963	129	339	331	192	70	16	0.5	0.5	0.5
1964	149	524	515	201	63	18	2	0.5	0.5
1965	75	379	501	328	133	39	11	1	0.5
1966	149	488	459	172	64	22	5	0.5	0.5
1967	63	368	287	130	55	19	6	0.5	0.5
1968	50	215	259	141	55	18	5	1	0.5
1969	45	150	153	76	23	6	0.5	0.5	0.5
1973	101	759	1268	1116	491	141	40	4	0.5
1974	151	733	1114	1092	571	163	50	9	5
1975	109	901	1517	1606	1076	342	117	12	7
1976	53	604	1204	1560	1146	396	156	18	10
1977	157	867	1343	1410	1031	417	192	17	7
1978	89	735	1307	1623	1150	445	198	18	14
1979	29	299	718	1268	1195	585	300	36	14

```
> ages <- 9:12
> yc67 <- c(1560,1031,445,300)
> yc64 <- c(1116,571,342,156)
> yc57 <- c(172,55,18,0.5)
> yc54 <- c(192,63,39,5)
> d <- data.frame(yc=factor(rep(c(1967,1964,1957,1954),each=4)),
                  age=rep(ages,times=4),
                  cpe=c(yc67,yc64,yc57,yc54))
> d <- within(d,logcpe <- log(cpe))
> d
```

	yc	age	cpe	logcpe
1	1967	9	1560.0	7.3524
2	1967	10	1031.0	6.9383
3	1967	11	445.0	6.0981

4	1967	12	300.0	5.7038
5	1964	9	1116.0	7.0175
6	1964	10	571.0	6.3474
7	1964	11	342.0	5.8348
8	1964	12	156.0	5.0499
9	1957	9	172.0	5.1475
10	1957	10	55.0	4.0073
11	1957	11	18.0	2.8904
12	1957	12	0.5	-0.6931
13	1954	9	192.0	5.2575
14	1954	10	63.0	4.1431
15	1954	11	39.0	3.6636
16	1954	12	5.0	1.6094

Model Fitting

```
> # Compare 1957 and 1967 year-classes
> lm1 <- lm(logcpe~age*yc,data=Subset(d,yc %in% c(1957,1967)))
> anova(lm1)
```

Analysis of Variance Table

Response: logcpe

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
age	1	14.91	14.91	32.41	0.0047
yc	1	27.16	27.16	59.03	0.0015
age:yc	1	4.13	4.13	8.98	0.0401
Residuals	4	1.84	0.46		

```
> summary(lm1)
```

Call:

```
lm(formula = logcpe ~ age * yc, data = Subset(d, yc %in% c(1957,
1967)))
```

Residuals:

1	2	3	4	5	6	7	8
-0.0386	0.1258	-0.1358	0.0486	-0.4864	0.2374	0.9843	-0.7353

Coefficients:

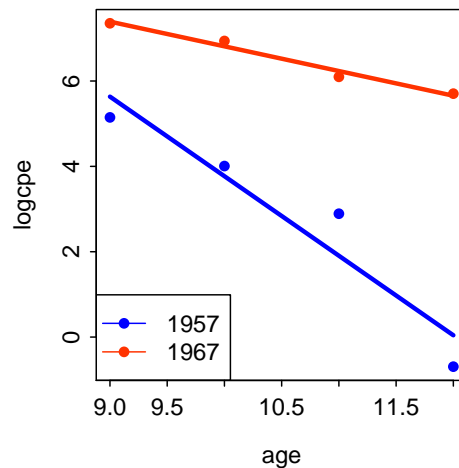
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	22.409	3.203	7.00	0.0022
age	-1.864	0.303	-6.14	0.0036
yc1967	-9.810	4.530	-2.17	0.0963
age:yc1967	1.285	0.429	3.00	0.0401

Residual standard error: 0.678 on 4 degrees of freedom
Multiple R-squared: 0.962, Adjusted R-squared: 0.933
F-statistic: 33.5 on 3 and 4 DF, p-value: 0.00272

```
> confint(lm1)
```

	2.5 %	97.5 %
(Intercept)	13.51506	31.303
age	-2.70615	-1.022
yc1967	-22.38791	2.768
age:yc1967	0.09413	2.476

```
> fitPlot(lm1, legend="bottomleft")
```



```
> # Compare 1964 and 1967 year-classes
> lm2 <- lm(logcpe~age*yc,data=Subset(d,yc %in% c(1964,1967)))
> anova(lm2)
```

Analysis of Variance Table

Response: logcpe

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
age	1	3.72	3.72	293.92	6.8e-05
yc	1	0.42	0.42	33.53	0.0044
age:yc	1	0.01	0.01	0.78	0.4265
Residuals	4	0.05	0.01		

```
> summary(lm2)
```

Call:

```
lm(formula = logcpe ~ age * yc, data = Subset(d, yc %in% c(1964,
1967)))
```

Residuals:

1	2	3	4	5	6	7	8
-0.03863	0.12583	-0.13576	0.04856	-0.00721	-0.03578	0.09320	-0.05021

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	12.7987	0.5314	24.08	1.8e-05
age	-0.6416	0.0503	-12.75	0.00022
yc1967	-0.2001	0.7515	-0.27	0.80324
age:yc1967	0.0629	0.0712	0.88	0.42650

Residual standard error: 0.113 on 4 degrees of freedom

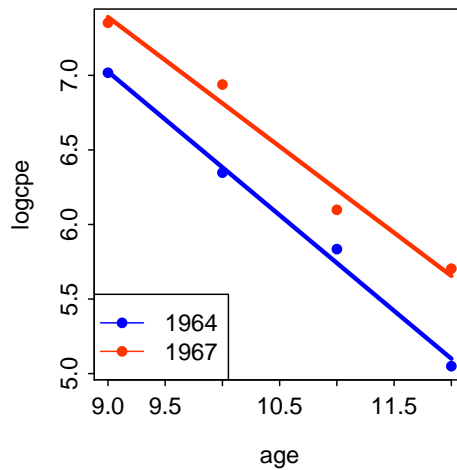
Multiple R-squared: 0.988, Adjusted R-squared: 0.979

F-statistic: 109 on 3 and 4 DF, p-value: 0.000271

```
> confint(lm2)
```

	2.5 %	97.5 %
(Intercept)	11.3233	14.2741
age	-0.7813	-0.5018
yc1967	-2.2866	1.8865
age:yc1967	-0.1347	0.2605

```
> fitPlot(lm2, legend="bottomleft")
```



```
> # Fit without the insignificant interaction term as a demonstration
> lm2a <- lm(logcpe~age+yc,data=Subset(d,yc %in% c(1964,1967)))
> anova(lm2a)
```

Analysis of Variance Table

Response: logcpe

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
age	1	3.72	3.72	307.3	1.1e-05
yc	1	0.42	0.42	35.1	0.002
Residuals	5	0.06	0.01		

```
> summary(lm2a)
```

Call:

```
lm(formula = logcpe ~ age + yc, data = Subset(d, yc %in% c(1964,
1967)))
```

Residuals:

1	2	3	4	5	6	7	8
-0.0858	0.1101	-0.1200	0.0958	0.0400	-0.0200	0.0775	-0.0974

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	12.4683	0.3695	33.74	4.3e-07
age	-0.6101	0.0348	-17.53	1.1e-05
yc1967	0.4608	0.0778	5.92	0.002

Residual standard error: 0.11 on 5 degrees of freedom

Multiple R-squared: 0.986, Adjusted R-squared: 0.98

F-statistic: 171 on 2 and 5 DF, p-value: 2.49e-05

```
> confint(lm2a)
```

	2.5 %	97.5 %
(Intercept)	11.5184	13.4182
age	-0.6995	-0.5206
yc1967	0.2607	0.6608