Interpreting Linear Models

Jim Robison-Cox

September 4, 2013

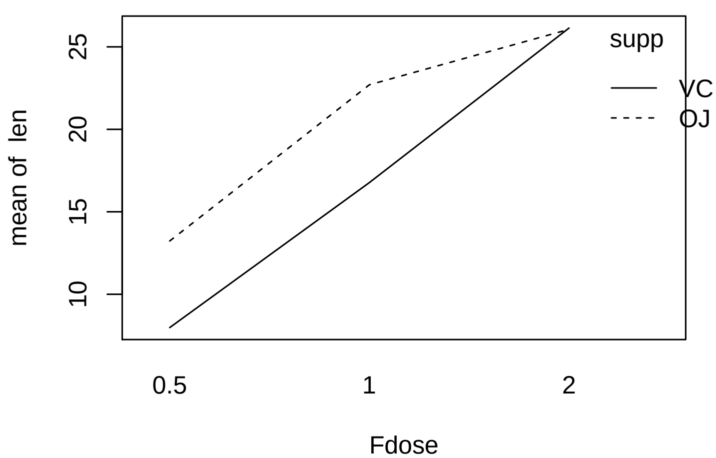
The `ToothGrowth` data is a builtin R dataset, so we don't have to load it. I'll use it to illustrate several types of models.

2-way ANOVA, analysis of variance

```
ToothGrowth$Fdose <- factor(ToothGrowth$dose)
anova.fit <- lm(len ~ supp * Fdose, ToothGrowth)
require(xtable)
xtable(summary(anova.fit))
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	13.2300	1.1484	11.52	0.0000
suppVC	-5.2500	1.6240	-3.23	0.0021
Fdose1	9.4700	1.6240	5.83	0.0000
Fdose2	12.8300	1.6240	7.90	0.0000
suppVC:Fdose1	-0.6800	2.2967	-0.30	0.7683
suppVC:Fdose2	5.3300	2.2967	2.32	0.0241

```
with(ToothGrowth, interaction.plot(Fdose, supp, len))
```



```r
unique(round(fitted(anova.fit), 2))
```

[1] 7.98 16.77 26.14 13.23 22.70 26.06

```r
xtable(with(ToothGrowth, tapply(len, list(supp, Fdose), mean)))
```

	0.5	1	2
OJ	13.23	22.70	26.06
VC	7.98	16.77	26.14

$$y_{ijk} = \mu + \tau_i + \beta_j + \gamma_{ij} + \epsilon_{ijk} \; i = 1, 2; \; j = 1, 2, 3; k = 1, \ldots, 10$$

I am using $i = 1, 2$ to index supplements (OJ, VC), j for dose, and k for replicate within a dose-supp combination.

1. How many parameters does the model have? Compare to the number of unique fitted values.
   There are only 6 uniques fitted values, but the model has 12 parameters (one μ, two τ's, three β's, six γ's).

2. What combination of Greek letters is the `Intercept` line in the output estimating?

   $$\mu + \tau_1 + \beta_1 + \gamma_{11}$$

   The so called intercept is estimating the mean growth for the first supplement (OJ) at the first dose (0.5).

3. Similarly, show what each other line in the coefficient table is estimating.

Label	estimate	parameters
(Intercept)	13.23	$E(\bar{y}_{11}) = \mu + \tau_1 + \beta_1 + \gamma_{11}$
suppVC	-5.25	$E(\bar{y}_{21} - \bar{y}_{11}) = \tau_2 - \tau_1 + \gamma_{21} - \gamma_{11}$
Fdose1	9.47	$E(\bar{y}_{12} - \bar{y}_{11}) = \beta_2 - \beta_1 + \gamma_{12} - \gamma_{11}$
Fdose2	12.83	$E(\bar{y}_{13} - \bar{y}_{11}) = \beta_3 - \beta_1 + \gamma_{13} - \gamma_{11}$
suppVC:Fdose1	-0.68	$E(\bar{y}_{22} - \bar{y}_{21} - \bar{y}_{12} + \bar{y}_{11}) = \gamma_{22} - \gamma_{21} - \gamma_{21} + \gamma_{11}$
suppVC:Fdose2	5.33	$E(\bar{y}_{23} - \bar{y}_{21} - \bar{y}_{13} + \bar{y}_{11}) = \gamma_{23} - \gamma_{13} - \gamma_{21} + \gamma_{11}$

   The 2nd line is the adjustment to the mean when going from OJ to VC at low dose. 3rd line: adjustment to mean in going from 0.5mg to 1 mg of OJ. 4th line: adjustment to mean in going from 0.5mg to 2 mg of OJ. 5th: the change in effect of dose (going from .5 to 1) when looking at VC instead of the baseline OJ. Alternatively, it's also the change in the OJ versus VC effect at 1 mg over the VC to OJ comparison for .5 mg. 6th: the change in effect of dose (going from .5 to 2) when looking at VC instead of the baseline OJ. Alternatively, it's also the change in the OJ versus VC effect at 2 mg over the VC to OJ comparison for .5 mg.

ANCOVA, analysis of covariance

$$y_i = \beta_0 + \beta_1 I_{VCi} + \beta_2 x_i + \beta_3 I_{VCi} x_i + \epsilon_i \quad i = 1, \ldots, 60$$

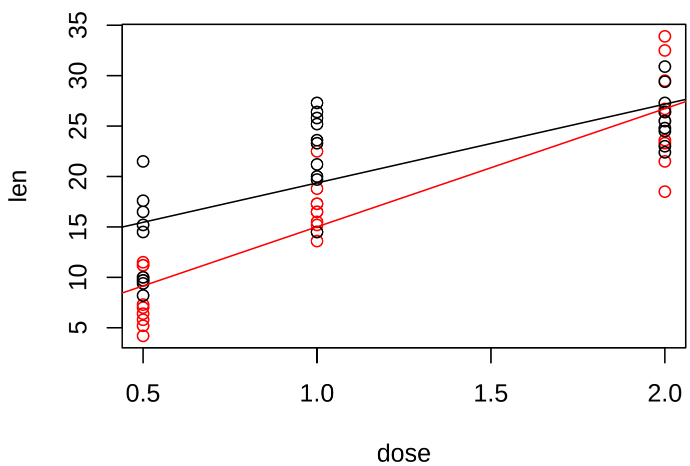
```r
ancova.fit <- lm(len ~ supp * dose, ToothGrowth)
xtable(summary(ancova.fit))
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	11.5500	1.5814	7.30	0.0000
suppVC	-8.2550	2.2364	-3.69	0.0005
dose	7.8114	1.1954	6.53	0.0000
suppVC:dose	3.9043	1.6906	2.31	0.0246

```r
unique(round(fitted(ancova.fit), 2))
```

[1] 9.15 15.01 26.73 15.46 19.36 27.17

```r
plot(len ~ dose, data = ToothGrowth, col = as.numeric(supp))
abline(coef(ancova.fit)[c(1, 3)])
abline(11.55 - 8.255, 7.811 + 3.904, col = "red")
```



Interpret each coefficient.

The estimated intercept (β_0) is 11.55 which estimates tooth growth with no OJ. The intercept for VC is estimated to be $\widehat{\beta}_1 = -8.255$ mm lower than that of OJ. Increasing dose by 1 mg for the OJ group is estimated to increase tooth growth by $\widehat{\beta}_2 = 7.811$ mm. For the VC group, the slope is even steeper, as a 1 unit increase in vitamin C is estimated to increase growth by $\widehat{\beta}_3 = 3.90$ mm in addition to the 7.81 mm for OJ.

Sequential F tests

ANOVA fit:

```r
xtable(anova(anova.fit))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
supp	1	205.35	205.35	15.57	0.0002
Fdose	2	2426.43	1213.22	92.00	0.0000
supp:Fdose	2	108.32	54.16	4.11	0.0219
Residuals	54	712.11	13.19		

What does each line test?

1. H_0: *supplements have the same mean effect versus* H_a *supplement effects differ.*

2. H_0: *doses have the same mean effect versus* H_a *dose effects differ* **given that supplements are in the model**.

3. H_0: *the effect of supplement (OJ vs VC) is the same at all doses versus* H_a *supplement effects differ by dosage* **given that supplements and dosages are in the model**.

Conclusion: there are strong differences between supplements and dosages, and the effect of dosage does differ from OJ to VC. ANCOVA fit:

```r
xtable(anova(ancova.fit))
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
supp	1	205.35	205.35	12.32	0.0009
dose	1	2224.30	2224.30	133.42	0.0000
supp:dose	1	88.92	88.92	5.33	0.0246
Residuals	56	933.63	16.67		

What does each line test?

1. H_0: *supplements have the same mean effect versus* H_a *supplement effects differ.*

2. H_0: *the slope on doses is zero* H_a *non zero slope on dose* **given that supplements are in the model**.

3. H_0: *the slope on dose is the same for OJ and VC versus* H_a *slopes on dose differ by supplement* **given that supplements and dose are in the model**.

Conclusion: there are strong differences between supplements and a there is a linear trend over dosages. Also, the effect of dosage does differ from OJ to VC.

Model Comparison:

```r
xtable(anova(ancova.fit, anova.fit))
```

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	56	933.63				
2	54	712.11	2	221.53	8.40	0.0007

What does this command test?

Conclusion: There is strong evidence ($F_{2,45} = 8.4$, p-value $< .001$) that the six means model fits the data better than a model which forces the means to lie on two straight lines. If tooth growth is inhibited by lack of vitamin C, then it seems sensible that there is a ceiling on the amount of improvement possible by increasing vitamin C intake. At some point, vitamin C will no longer be the limiting factor. It appears that with OJ we've started to get close to the ceiling, as the improvement slows down.