Lecture 14 Monday February 13

Notes pp. 72 – 78

Topics:

One-way ANOVA – sums of squares

Mean squares, expected mean squares (new formula)

F test for equality of factor level means

Announcements

Quiz 2 is due Tuesday Feb. 14

HW04 is due Wednesday Feb. 15

Analysis of variance Basis is decomposition of observations, sums of squares and df, like in regression.

Decomposition of observations: Add and subtract $\bar{Y}_{i\cdot}$:

$$Y_{ij} - \overline{Y}_{..} = (\overline{Y}_{i.} - \overline{Y}_{..}) + (Y_{ij} - \overline{Y}_{i.}) \tag{1}$$

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$$

Decomposition of sums of squares and degrees of freedom:

total sum of squares — Treatment SS — Error SS

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{..})^2 = \sum_{i=1}^{r} n_i(\overline{Y}_{i.} - \overline{Y}_{..})^2 + \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i.})^2$$

$$df = \sum_{i=1}^{r}(n_i - 1) = \sum_{i=1}^{r} n_i - r = n_T - r$$

$$n_T - 1 = (r - 1) + (n_T - r)$$

The decomposition of sums of squares holds because when you square the right-hand side of Equation (1), the cross-product term is zero.

Explanation of why cross-product term is zero:

Decomposition of observations:

$$Y_{ij} - \overline{Y}_{..} = (\overline{Y}_{i.} - \overline{Y}_{..}) + (Y_{ij} - \overline{Y}_{i.})$$

When you square the right-hand side of the above equation, the cross-product term is:

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i} 2(\overline{Y}_{i.} - \overline{Y}_{..})(Y_{ij} - \overline{Y}_{i.}) = \sum_{i=1}^{r} 2(\overline{Y}_{i.} - \overline{Y}_{..}) \sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i.}) \quad = 0$$

0

Decomposition of sum of squares:

$$\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{..})^2 = \sum_{i=1}^{r} n_i(\overline{Y}_{i.} - \overline{Y}_{..})^2 + \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y}_{i.})^2$$

SSTO = SSTR + SSE

Ex. Kenton Food Company

Package Design	Number of Stores	Mean	SD
i	n_i	$\bar{Y}_{i.}$	s_i
1	5	14.6	2.302
2	5	13.4	3.647
3	4	19.5	2.646
4	5	27.2	3.962

Grand mean: $\bar{Y}_{..} = (5(14.6) + 5(13.4) + 4(19.5) + 5(27.2))/19 = 18.632$
[Correct two typos below:]

$$\text{SSTR} = 5(14.6 - 18.632)^2 + 5(13.4 - \overset{18.632}{\cancel{18.682}})^2 + 4(19.5 - 18.632)^2 + 5(\overset{27.2}{\cancel{19.5}} - 18.632)^2 = 588.2211$$

$$\text{SSE} = 4(2.302^2) + 4(3.647^2) + 3(2.646^2) + 4(3.962^2) = 158.2$$

$$SSE = \sum_{i=1}^{r} \sum_{j=1}^{n_i} e_i^2 = \sum_{i=1}^{r} (n_i - 1) s_i^2$$

$$= \sum \sum (Y_{ij} - \bar{Y}_{i.})^2$$

Mean Square: a sum of squares, divided by its degrees of freedom. Note that a *mean square* is a statistic and has a sampling distribution.

Mean square for treatment $$\text{MSTR} = \frac{\text{SSTR}}{r-1}$$

Mean square for error $$\text{MSE} = \frac{\text{SSE}}{n_T - r}$$

Define $\mu_{\cdot}$ to be the weighted mean of the $\mu_i, i = 1, \ldots, r$ in the cell-means model:

$$\mu_{\cdot} = \frac{\sum_{i=1}^{r} n_i \mu_i}{n_T}$$

Expected Mean Square: $\underbrace{\text{the mean of the sampling distribution}}_{\text{expected value}}$ of a mean square.

Recall: $\text{MSTR} = \frac{\sum_{i=1}^{r} n_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2}{r-1}$

$$\text{E(MSTR)} = \sigma^2 + \frac{\sum n_i (\mu_i - \mu_{\cdot})^2}{r-1}$$

$$\text{E(MSE)} = \sigma^2$$

EX. (p. 64) Hypothetical true parameter values

$\mu_1 = 15, \quad \mu_2 = 16, \quad \mu_3 = 20, \quad \mu_4 = 28, \quad \sigma = 1.5$

Take the sample size to be $n_T = 19$ w/ $n_1 = 5, n_2 = 5, n_3 = 4, n_4 = 5$

Find $E(MSTR)$ and $E(MSE)$.

First get $\mu_{\cdot} = \dfrac{\sum_{i=1}^{r} n_i \mu_i}{n_T} = \dfrac{5(15) + 5(16) + 4(20) + 5(28)}{19}$

$$= \frac{375}{19} \doteq 19.74$$

Next get $\dfrac{\sum^{r} n_i (\mu_i - \mu_{\cdot})^2}{r-1} = \dfrac{5(15-19.7)^2 + 5(16-19.7)^2 + 4(20-19.7)^2 + 5(28-19.7)^2}{3}$

$$= 174.56$$

Then, $E(MSTR) = \sigma^2 + 174.56 = 2.25 + 174.56 = 176.8$

And $E(MSE) = \sigma^2 = 2.25$

ANOVA Table

Outline of "Theoretical" ANOVA table.

Source of Variation	Sum of Squares	df	Mean Square	Expected Mean Square
Treatment	SSTR	$r-1$	$\frac{SSTR}{r-1}$	
Error	SSE	$n_T - r$	$\frac{SSE}{n_T - r}$	
Total	SSTO	$n_T - 1$		

Kenton Food Company

Practical ANOVA table

Source of Variation	Sum of Squares	df	Mean Square	F	P
Treatment	588.22	3	196.07		
Error	158.2	15	10.55		
Total	746.42	18			

F Test for Equality of Factor Level Means

Hypotheses to be tested:

H_0: $\mu_1 = \mu_2 = \cdots = \mu_r$ vs.

H_a: at least two of the means are not equal.

For some i, j $\mu_i \neq \mu_j$

The F Statistic (The F Ratio)

$$F_{\text{obs}} = \frac{MSTR}{MSE}$$

For fixed r and $n_1, n_2, \ldots, n_r$, if H_0 is true then the distribution of F is completely determined.

Give proof based on the invariance argument:
Changing the common mean and common variance is equivalent to using $Y'_{ij} = aY_{ij} + b$. Show that the numerical value of F_{obs} would not change.

$$\bar{Y}_{i\cdot}' = a\bar{Y}_{i\cdot} + b, \qquad \bar{Y}_{\cdot\cdot}' = a\bar{Y}_{\cdot\cdot} + b$$

$$\bar{Y}_{i\cdot}' - \bar{Y}_{\cdot\cdot}' = a(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})$$

$$SSTR' = \sum_{i=1}^{r} n_i \left[a(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})\right]^2$$

$$= a^2 \sum_{i=1}^{r} n_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$$

$$= a^2 \, SSTR$$

<u>Show</u> $SSE' = a^2 \, SSE$

$$F_{\text{obs}} = \frac{\frac{\text{SSTR}}{r-1}}{\frac{\text{SSE}}{n_T-r}} \qquad \frac{\text{MSTR}}{\text{MSE}}$$

Theorem

If H_0 is true, the distribution of F_{obs} is F_{r-1,n_T-r}. The two parameters are the numerator degrees of freedom and denominator degrees of freedom $(r-1, n_T-r)$. [3]

If H_0 is not true, the numerator MSTR will tend to be larger than what would be expected if H_0 was true. To see this, compare E(MSTR) when H_0 is true, to E(MSTR) when H_0 is false.

So, under H_a, F_{obs} will also be larger than expected under H_0. So the test consists of comparing the ratio with the 95% point of the F distribution with $r-1$ and n_T-r degrees of freedom.

[3] The numerator is a χ^2 r.v. divided by its degrees of freedom, and the denominator is another χ^2 r.v. divided by its degrees of freedom. The numerator and denominator are independent of each other by Cochran's Thm.

Recall $MSTR = \dfrac{\sum_{i=1}^{r} n_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2}{r-1}$

and $E(MSTR) = \sigma^2 + \dfrac{\sum_{i=1}^{r} n_i (\mu_i - \mu_\cdot)^2}{r-1}$

(Valid under both H_0 and H_a)

Two Cases:

1) H_0 is true. $H_0: \mu_1 = \mu_2 = \dots = \mu_r$

   In this case, $E(MSTR) = \sigma^2$ (why?)

2) H_a is true. At least two means are not equal.

   In this case, $E(MSTR) > \sigma^2$ (why?)

Recall: $E(MSE) = \sigma^2$ under both H_0, and H