Chapter 9: The One-Way Model

The One-Way Means Model

Example

A study was conducted to assess the effects of feedback in a repetitive industrial task. The task was to grind a metal piece to a specified size and shape. Eighteen male workers were divided randomly into three groups. All subjects were given the same introduction to the task.

After beginning the experimental period, the subjects in one group received no feedback about the task, those in the second group were given vague and intermittent feedback, and subjects in the third group were given accurate and continuous feedback.

The response consisted of a measure of the value, in dollars, added to production by each subject during the experimental period. This measure was a function of the number of pieces produced, the accuracy of the grinding operation and the amount of reworking necessary in the remaining stages of production. One worker became ill during the study, and his data were dropped. The data, found in SASDATA.FEEDBACK, are:

Type of Feedback		
None	Vague	Accurate
40.85	38.32	48.59
35.21	40.26	40.71
38.17	47.47	45.33
43.96	44.10	43.76
34.88	40.09	46.41
	42.67	44.19

Let's explore the data visually:

A model appropriate for data of this type is the **one-way means model**:

$$Y_{ij} = \mu_i + \epsilon_{ij},\ j = 1, \ldots, n_i,\ i = 1, \ldots, k,$$

where the random errors ϵ_{ij} are assumed to be independent random variables from the same zero-mean distribution having variance σ^2. Usually, it is assumed $\epsilon_{ij} \sim N(0, \sigma^2)$.
For the present data, $i = 3$, $n_1 = 5$, and $n_2 = n_3 = 6$; μ_1 is the population mean for the no feedback group, and μ_2 and μ_3 are the population means for the vague and accurate feedback groups, respectively.

The one-way means model should look familiar. If $k = 2$ (i.e., there are two populations), it is just the two population C+E model introduced in chapters 5 and 6, with the assumption that both population variances are the same value: σ^2.

Fitting the Model

The least squares estimator of μ_i is just the mean of the observations from population i:

$$\hat{\mu}_i = \overline{Y}_{i\cdot} = \frac{1}{n_i} \sum_{j=1}^{n_i} Y_{ij}$$

The corresponding estimator of σ^2 is the pooled variance estimator

$$S_p^2 = \frac{1}{n-k} \sum_{i=1}^{k} (n_i - 1) S_i^2,$$

where S_i^2 is the sample variance of the observations from population i.

Let's check out the fit of the model to the feedback data.

Checking the Fit

The residuals

$$e_i = Y_{ij} - \hat{\mu}_i = Y_{ij} - \overline{Y}_{i\cdot}$$

are used to check the fit.

Here's how to check the fit of the model to the feedback data.

Testing the Equality of Means

The question researchers most often ask concerning the means model is "Are the population means all equal?"
Formally, the hypotheses are

$$
\begin{aligned}
H_0 : \quad & \mu_1 = \mu_2 = \ldots = \mu_k \\
H_a : \quad & \text{Not all the population means } \mu_i \text{ are equal.}
\end{aligned}
$$

These hypotheses are tested using the F statistic. To see what the F statistic is all about, we first need to learn about partitioning the variation in the response into different components, in what is known as the Analysis of Variance (aka ANOVA).

The Analysis of Variance

The total variation in the responses is measured by the **total sum of squares**, SSTO:

$$\text{SSTO} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{..})^2,$$

where

$$\overline{Y}_{..} = \frac{1}{n} \sum_{i=1}^{k} \sum_{j=1}^{n_i} Y_{ij},$$

is the mean of all the observations (called the grand mean). Here, $n = \sum_{i=1}^{k} n_i$ is the total number of observations.

SSTO can be broken into two components: the variation explained by the model, the **model sum of squares**

$$\text{SSM} = \sum_{i=1}^{k} n_i(\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2,$$

and the variation left unexplained by the model, the **error sum of squares**

$$\text{SSE} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y}_{i\cdot})^2.$$

Further, these components add: SSTO=SSM+SSE.

Each SS has associated with it a number, called its **degrees of freedom**, which counts the number of independent pieces of data going into the SS. The df for SSTO, SSM and SSE are $n-1$, $k-1$ and $n-k$. These df add exactly like their SS.

The **mean square** is the SS divided by its df. Thus MSM=SSM/$(k-1)$, and MSE=SSE/$(n-k)$.

The test statistic for testing equality of population means is the F statistic $F =$MSM/MSE. It is compared with its distribution under H_0, which is an $F_{k-1,n-k}$ distribution. Large values of F support H_a over H_0.

The information about SS, df, MS and the F test is summarized in an ANOVA table. Let's have a look...

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Stat	Prob > F
Model	$k-1$	SSM	MSM	F =MSM/MSE	p-value
Error	$n-k$	SSE	MSE		
C Total	$n-1$	SSTO			

Table: *ANOVA table for the one-way model*

TYU 19

If the null hypothesis of equality of population means is rejected, the next question is usually "which means differ"?
Pairs of population means can be compared using t tests or confidence intervals.

- To test $H_0 : \mu_i = \mu_j$ versus $H_a : \mu_i \neq \mu_j$, perform a two-sided t test using the t_{n-k} distribution and the test statistic

$$t_{ij0} = \frac{\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}}{\hat{\sigma}(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot})},$$

  where

$$\hat{\sigma}(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}) = \sqrt{\text{MSE}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)}.$$

For the feedback example, let's test for differences between the population means of the accurate and no feedback groups. The sample means are 44.832 and 38.614, and MSE is 10.684. The observed value of the test statistic is

$$t^*_{ij0} = \frac{44.832 - 38.614}{\sqrt{10.684\left(\frac{1}{6} + \frac{1}{5}\right)}} = 3.141.$$

The p-value is $p\pm = 2min(p_-, p^+)$, where $p^+ = P(t_{14} \geq 3.141) = 0.0036$, and $p_- = 1 - p^+ = 0.9964$. So, $p\pm = 2 \times 0.0036 = 0.0072$.

- A level L confidence interval for $\mu_i - \mu_j$ is

$$\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot} \pm \hat{\sigma}(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot})t_{n-k,\frac{1+L}{2}}.$$

Using the fact that $t_{14,0.975} = 2.1448$, a level 0.95 interval for the difference between the population means of the accurate and no feedback groups is

$$44.832 - 38.614 \pm 2.1448 \times \sqrt{10.684\left(\frac{1}{6} + \frac{1}{5}\right)} = (1.973, 10.463).$$

Consider the confidence interval we've just shown for comparing two population means. This type of comparison is called a **pairwise comparison**, since when we set the confidence level we are only concerned with that particular comparison. If we are doing a lot of these comparisons, we can run into problems interpreting the confidence levels. These problems have to do with the fact that if you do a lot of tests, you will often get a lot of false positives. Let's see how this works:

The problems with doing a lot of comparisons also have to do with

- Exploratory and confirmatory studies.
- Formal and informal inference, and the necessity of each.
- Data snooping and its relation to formal and informal inference.

One possible solution is to use **multiple comparison procedures**.

Multiple Comparison Procedures

Two multiple comparison procedures which control the overall error rate for all comparisons made are the **Bonferroni** and **Tukey** procedures.

- The Tukey procedure considers all pairwise comparisons and gives an overall level L error rate. It is based on the distribution of the difference between the largest and smallest mean of a set of sample means.

When studentized by dividing by estimated standard error, this distribution is called the **studentized range distribution**. It is suitable for use in data snooping. A level L Tukey interval for $\mu_i - \mu_j$ is

$$\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot} \pm \hat{\sigma}(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot})\frac{q_{L,k,n-k}}{\sqrt{2}},$$

where $q_{L,k,n-k}$ is the L^{th} quantile of the studentized range distribution. Note how the Tukey procedure controls the overall error rate in the simulation we ran:

- The Bonferroni is a very general procedure which can be used for any set of comparisons, not just pairwise comparisons. However, it can be used for data snooping only if all comparisons of the type of interest-for example, all pairwise comparisons-are included among the possible comparisons. If used for all pairwise comparisons, the Bonferroni interval for $\mu_i - \mu_j$ is

$$\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot} \pm \hat{\sigma}(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot})t_{n-k,1-\frac{(1-L)}{2N}},$$

  where $N = k(k-1)/2$ is the number of pairwise comparisons possible.

Here's an example...

TYU 20

What Happens When Model Assumptions Are Violated?

- Nonnormality (F test is robust. Individual and multiple comparisons robust unless n_i are small)
- Heteroscedasticity (F test is robust for equal sample sizes. Individual and multiple comparisons are not.)
- Nonindependence (Can present serious difficulties.)

The One-Way Effects Model

The one-way effects model is the one-way means model parametrized to emphasize the deviations of population means from an overall mean. It is written

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij},\ j = 1 \ldots, n_i,\ \ i = 1, \ldots, k,$$

where $\mu = \sum_{i=1}^{k} \mu_i / k$ is an overall mean for all populations, and $\tau_i = \mu_i - \mu$ is the effect due to the i^{th} population.

Blocking in the One-Way Model

Example

Four types of highway surface are being tested for durability. Engineers obtained 10 different sites on existing highways to test these surfaces. Since the sites are on different types of highways, the engineers decided to divide each site into four equal sections and randomly assign one surface to each section in such a way that all four surface types appear at each site. In reality, the test sites were monitored periodically and a number of measures of wear were taken on each occasion.

The response we will consider is an index of severity of wear, coded on a scale of 0 (no wear) to 100 (severe wear). The data are found in SASDATA.ASPHALT.
Let's look at the data.

A Short Review of the CRD and the RCBD

In chapter 3 you studied something about experimental designs. The two most basic designs presented there are the **Completely Randomized Design (CRD)** and the **Randomized Complete Block Design (RCBD)**.

In a CRD treatments are assigned to experimental units completely at random. The design of the asphalt experiment would be a CRD if the treatments (asphalt types) were assigned to experimental units (sections) completely at random. One possible result might be that some sites might have more than one section with the same asphalt type, which would defeat the purpose of dividing each site into four sections.

In the kind of RCBD we are studying, the experimental units are first divided into homogeneous blocks. Then, treatments are randomly assigned to the experimental units within a block in such a way that each unit gets a different treatment. This is what was done in the asphalt experiment. Why is this a better design for the asphalt experiment than a CRD?

The Randomized Complete Block Model

One useful model for data of this type is the randomized complete block model:

$$Y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij},\ i = 1, \ldots, k,\ j = 1 \ldots, b$$

Notice that this is an effects model with two factors: blocks, represented by the β effects and treatments, represented by the τ effects. Notice also the model is additive: the effects add.

TYU 21

Fitting the RCB Model

The least squares estimators of the parameters are:

$$\hat{\mu} = \overline{Y}_{..}, \ \hat{\tau}_i = \overline{Y}_{i.} - \overline{Y}_{..}, \ \text{ and } \ \hat{\beta}_j = \overline{Y}_{.j} - \overline{Y}_{..}$$

The fitted values are

$$\begin{aligned} \hat{Y}_{ij} &= \hat{\mu} + \hat{\tau}_i + \hat{\beta}_j \\ &= \overline{Y}_{..} + (\overline{Y}_{i.} - \overline{Y}_{..}) + (\overline{Y}_{.j} - \overline{Y}_{..}) \\ &= (\overline{Y}_{i.} + \overline{Y}_{.j} - \overline{Y}_{..}) \end{aligned}$$

The residuals are, as usual, the observed minus the fitted values,

$$e_{ij} = Y_{ij} - \hat{Y}_{ij} = Y_{ij} - \overline{Y}_{i.} - \overline{Y}_{.j} + \overline{Y}_{..}$$

Checking the Fit

The fit is checked by

- Plotting residuals versus predicted, block and treatment.
- Plotting studentized residuals versus $t_{(k-1)(b-1)}$ quantiles.
- Looking at interaction plots.
- Testing for interaction (Tukey).

Let's check out the fit for the asphalt data ourselves.

The Analysis of Variance

We test

$$\begin{aligned} H_{0\tau}: &\quad \tau_1 = \tau_2 = \cdots = \tau_k = 0 \\ H_{a\tau}: &\quad \text{Not all the population effects } \tau_i \text{ are } 0. \end{aligned}$$

As for the one-way model, the ANOVA table shows sums of squares, degrees of freedom and mean squares for the RCB model.

Let's look at the analysis for the asphalt data.

Individual Comparisons

- To test

$$\begin{array}{rcl} H_0: & \tau_i & = & \tau_j \\ H_a: & \tau_i & \neq & \tau_j. \end{array}$$

we use the test statistic

$$t_{ij0} = \frac{\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}}{\hat{\sigma}(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot})}.$$

Under H_0, t_{ij0} has a $t_{(k-1)(b-1)}$ distribution.

- A level L confidence interval for $\tau_i - \tau_j$ has endpoints

$$\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot} \pm \hat{\sigma}(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot})t_{(k-1)(b-1),\frac{1+L}{2}}$$

Multiple Comparisons

As for the one-way model, we may use either the Bonferroni or the Tukey procedure to compare more than one pair of means.

- A set of Bonferroni confidence intervals for comparing N pairs of population effects with overall confidence level L, computes the endpoints of the interval for $\tau_i - \tau_j$ as

$$\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot} \pm \hat{\sigma}(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot})t_{(k-1)(b-1),1-\frac{2(1-L)}{N}}$$

  When doing all $k(k-1)/2$ pairwise comparisons for k populations, take $N = k(k-1)/2$.

- A set of Tukey confidence intervals for all pairwise comparisons of k population effects with overall confidence level L, computes the endpoints of the interval for $\tau_i - \tau_j$ as

$$\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot} \pm \hat{\sigma}(\overline{Y}_{i\cdot} - \overline{Y}_{j\cdot}) \frac{q_{L,k,(k-1)(b-1)}}{\sqrt{2}}$$

The confidence level is exact for equal sample sizes from all populations and is conservative if the sample sizes are not all equal.

The Benefits of Blocking

Consider the ANOVA table for the asphalt data considered as a RCBD:

and as a one-way model:

Now consider what this does for estimation. Here is a level 0.95 confidence interval for the difference in mean wear between asphalt types 2 and 3 computed from the one-way model:

$$(4.915, 20.085)$$

And here is a level 0.95 confidence interval for the difference in mean wear between asphalt types 2 and 3 computed from the RCB model:

$$(8.963, 16.037).$$

TYU 22