Analysis of Factor Level Means

- **Inferences for Single Factor Level Mean:**

  An unbiased point estimator of the factor level mean μ_i is $\hat{\mu}_i = \bar{Y}_{i\cdot}$.

  1. $E\{\bar{Y}_{i\cdot}\} = \mu_i$

  2. $\sigma^2\{\bar{Y}_{i\cdot}\} = \dfrac{\sigma^2}{n_i}$

  3. $s^2\{\bar{Y}_{i\cdot}\} = \dfrac{MSE}{n_i}$

  4. $\dfrac{\bar{Y}_{i\cdot} - \mu_i}{s\{\bar{Y}_{i\cdot}\}} \sim t_{df=(n_T - r)}$

Example. Kenton Food Company wished to test four different package designs for a new breakfast cereal. Twenty stores, with approximately equal sales volumes, were selected as the experimental units. Each store was randomly assigned one of the package designs, with each package design assigned to five stores. A fire occurred in one store during the study period, so this store had to be dropped from the study. The data are stored in the file ‘KentonFood.csv’. Construct a 95% confidence interval for μ_1.

```
sales=c(11,17,16,14,15,12,10,15,19,11,23,20,18,17,27,33,22,26,28)
design=c(rep(1,5),rep(2,5),rep(3,4),rep(4,5))
cbind(sales,design)

result=aov(sales~factor(design))
summary(result)
#                Df Sum Sq Mean Sq F value   Pr(>F)
# factor(design)  3  588.2  196.07   18.59 2.58e-05 ***
# Residuals      15  158.2   10.55

plot(result$fit,result$res)
qqnorm(result$res)
qqline(result$res)

par(mfrow=c(2,2))
plot(result)

tapply(sales,design,mean)
#    1    2    3    4
# 14.6 13.4 19.5 27.2

tapply(sales,design,sd)
       1        2        3        4
2.302173 3.646917 2.645751 3.962323

# 95% C.I. for mean 1 (Kenton Food Example).
lwr=14.6-qt(1-.05/2,df=15)*sqrt(10.547/5)   # lwr=11.50433
upr=14.6+qt(1-.05/2,df=15)*sqrt(10.547/5)   # upr=17.69567

# Or you can use the built-in R function ‘predict’
predict(result,interval="confidence",level=0.95)
```

- **Inferences for Difference between Two Factor Level Mean:**

  Frequently two treatments or factor levels are to be compared by estimating the difference D between two factor level means, say, μ_i and $\mu_{i'}$:

$$D = \mu_i - \mu_{i'}$$

  1. $\hat{D} = \bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}$ $\qquad \Rightarrow E\{\hat{D}\} = \mu_i - \mu_{i'}$

  2. $\sigma^2\{\hat{D}\} = \sigma^2\{\bar{Y}_{i\cdot}\} + \sigma^2\{\bar{Y}_{i'\cdot}\} = \sigma^2(\frac{1}{n_i} + \frac{1}{n_{i'}})$

  3. $s^2\{\hat{D}\} = MSE(\frac{1}{n_i} + \frac{1}{n_{i'}})$

  4. $\frac{\hat{D} - D}{s\{\hat{D}\}} \sim t_{df=(n_T - r)}$ $\qquad \Rightarrow$ The $(1-\alpha)100\%$ C.I. for D is given by: $\hat{D} \pm t_{\frac{\alpha}{2}} s\{\hat{D}\}$

  Example. In the Kenton Food Company example, package designs 1 and 2 used 3-color printing and designs 3 and 4 used 5-color printing. Also, package designs 1 and 3 utilized cartoons while no cartoons were utilized in designs 2 and 4. Construct a 95% C.I. for $\mu_3 - \mu_1$.

- **Inferences for Contrast of Factor Level Means:**

  A *contrast* is a comparison involving two or more factor level means and includes that previous case of a pairwise difference between two factor level means. A contrast will be denoted by L, and is defined as a linear combination of the factor level means μ_i where the coefficients c_i sum to zero:

$$L = \sum_{i=1}^{r} c_i \mu_i \qquad \text{where } \sum_{i=1}^{r} c_i = 0$$

  Example. In the Kenton Food Company example, the following contrasts may be of interest:

  1. Comparison of the mean sales for the two 3-color designs:

  2. Comparison of the mean sales for the 3-color and 5-color designs:

  3. Comparison of the mean sales for designs with and without cartoons:

  4. Comparison of the mean sales for design 1 with average sales for all 4 designs:

  1. $\hat{L} = \sum_{i=1}^{r} c_i \bar{Y}_i$ $\qquad \Rightarrow E\{\hat{L}\} = \sum_{i=1}^{r} c_i \mu_i$

  2. $s^2\{\hat{L}\} = MSE \sum_{i=1}^{r} \frac{c_i^2}{n_i}$ $\qquad \Rightarrow \sigma^2\{\hat{L}\} = \sigma^2 \sum_{i=1}^{r} \frac{c_i^2}{n_i}$

  3. $\frac{\hat{L} - L}{s\{\hat{L}\}} \sim t_{df=(n_T - r)}$ $\qquad \Rightarrow$ The $(1-\alpha)100\%$ C.I. for L is given by: $\hat{L} \pm t_{\frac{\alpha}{2}} s\{\hat{L}\}$

- **Homework.** Do 17.3, 17.8(a,b,c), and 17.9(a,b,c). See pages 736 and 739 for examples of line plot and bar-interval graph.