ANOVA Table

Source of Variation	Sum of Squares	df	Mean Square	Expected Mean Square
Treatment	SSTR	$r-1$	$\frac{\text{SSTR}}{r-1}$	$\sigma^2 + \frac{\sum n_i(\mu_i - \mu_.)^2}{r-1}$
Error	SSE	$n_T - r$	$\frac{\text{SSE}}{n_T-r}$	σ^2
Total	SSTO	$n_T - 1$		

Kenton Food Company

Source of Variation	Sum of Squares	df	Mean Square	F	P
Treatment	588.22	3	196.07	18.591	2.58×10^{-5}
Error	158.2	15	10.55		
Total	746.42	18			

$$F_{obs} = \frac{MSTR}{MSE} = \frac{196.07}{10.55} \approx 18.58$$

a little off due to rounding

P-value of the F test is area to the right of the observed statistic F_{obs}.

Ex: Kenton Food Company P-value $= P(F_{3,15} > 18.591)$

Null distr

$F_{r-1,\, n_T - r}$

```
> 1 - pf(18.591,df1=3,df2=15)
[1] 2.585007e-05
```

Since $P < .05$, reject $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$ at level $\alpha = .05$. Conclude that at least two of the package designs differ in number of packages sold on average.

F Test with pre-set Type I error probability α

For $\alpha = .05$ first find the $.95$ quantile of the appropriate F distribution.

```
> qf(.95,df1=3,df2=15)
[1] 3.287382
```

Decision Rule: For a level $\alpha = .05$ test, we will reject H_0 if $\underbrace{F_{\text{obs}} > 3.288}_{\text{critical region}}$

Factor Effects Model (Section 16.7)

Same one-way model as cell means model. Different formulation.

μ_i

$$Y_{ij} = \boxed{\mu_. + \tau_i} + \epsilon_{ij},\ \text{i} = 1, \ldots, r;\ \text{j} = 1, \ldots, n_i$$

where: $\mu_i = \mu_. + \tau_i \Rightarrow \tau_i = \mu_i - \mu_.$

- $\mu_.$ is the "overall mean," more accurately described as a "constant component common to all observations"
- τ_i is the effect of the i^{th} factor level
- the errors, ϵ_{ij}, are independent $\mathcal{N}(0, \sigma^2)$

Usually take $\mu_.$ to be the unweighted mean of the factor level means:

$$\mu_. = \frac{\sum_{i=1}^{r} \mu_i}{r}$$

This leads to the constraint on the τ_i's:

$$\sum_{i=1}^{r} \tau_i = 0$$

✓ because $\sum \tau_i = \sum_{i=1}^{r} (\mu_i - \mu_.) = 0$

The *factor-effects model* is a different formulation of the one-way ANOVA model, that is, it looks different from the cell-means model, but it is the **same model**.

We will use the unweighted mean for the parameter $\mu_{\cdot}$.

Weighted mean is sometimes appropriate:

1. A car rental company is interested in average fuel consumption for a population of vehicles which has 80 percent compacts, 15 percent mid-sized and 5 percent vans.

   Let $\mu_1 =$ mean fuel consumption for the population of compacts, $\mu_2 =$ mean fuel consumption for the population of mid-sized vehicles, and $\mu_3 =$ mean fuel consumption for the population of vans.

   The relevant parameter is $.8\mu_1 + .15\mu_2 + .05\mu_3$.

2. If random samples were used, weights proportional to sample size would make sense.

CAUTION: Recall the formula for E(MSTR): from lec 14

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

where $\mu_.$ is the weighted mean with weights proportional to the number of observations in the group (n_i).
This is a different definition of $\mu_{\cdot}$.

Keep straight which $\mu_{\cdot}$ to use from the situation.

Show $\sum_{i=1}^{r} \tau_i = 0$.

TAKE 1 - won't work in the end

Linear model approach to factor effects model *Example* $r = 4$, $n = 2$ observations per group

Set up ingredients for equation $E\underset{\sim}{Y} = \underset{\sim}{X}\underset{\sim}{\beta}$

Parameter vector:

$$\boldsymbol{\beta} = \begin{pmatrix} \mu_{.} \\ \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \end{pmatrix}$$

Design matrix:

dummy var., Group 1, 2, 3, 4

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Let $\underset{\sim}{c}_i$ = i^{th} column, vector of length 8

Note:

$\underset{\sim}{c}_1 = \underset{\sim}{c}_2 + \underset{\sim}{c}_3 + \underset{\sim}{c}_4 + \underset{\sim}{c}_5$

The design matrix is **not full rank**

The design matrix is **not full rank**

Recall, this means that one of the columns can be expressed as a linear combination of the other columns; for example,

TAKE 2 - WILL WORK

Linear model approach to factor effects model

We will take the approach of solving for one of the τ's in terms of the others. In the example, take $\tau_4 = -\tau_1 - \tau_2 - \tau_3$

$$\sum_{i=1}^{r} \tau_i = 0$$

Parameter vector:

$$\boldsymbol{\beta} = \begin{pmatrix} \mu_{\cdot} \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix}$$

$$\underset{\sim}{X}\,\underset{\sim}{\beta}$$

Design matrix:

$$X\beta = \underset{\sim}{\mathbf{X}} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & -1 \end{pmatrix} \begin{pmatrix} \mu_{\cdot} \\ \tau_1 \\ \tau_2 \\ \tau_3 \end{pmatrix} = \begin{matrix} \mu_{\cdot} + \tau_1 \\ \mu_{\cdot} + \tau_1 \\ \mu_{\cdot} + \tau_2 \\ \mu_{\cdot} + \tau_2 \\ \mu_{\cdot} + \tau_3 \\ \mu_{\cdot} + \tau_3 \\ \mu_{\cdot} - \tau_1 - \tau_2 - \tau_3 \\ \mu_{\cdot} - \tau_1 - \tau_2 - \tau_3 \end{matrix}$$

Exercise (1) Cell means model for a case w/ $r=3$ $n=2$

$\hat{\beta} = (X'X)^{-1} X'Y$ is the L.S. estimate.

Show $\hat{\beta} = \begin{pmatrix} \bar{Y}_{1.} \\ \bar{Y}_{2.} \\ \bar{Y}_{3.} \end{pmatrix}$

where $\bar{Y}_{i.} = \dfrac{Y_{i1} + Y_{i2}}{2}$.

(2) Do same thing for factor-effects model.