13.2 ANOVA

Spse we would like to compare means between two independent pops with equal variances.

i.e. $Y_1 \sim N(\mu_1, \sigma_1^2)$ $\quad Y_2 \sim N(\mu_2, \sigma_2^2)$ $\quad \sigma_1^2 = \sigma_2^2 = \sigma^2$

Under these conditions we developed the <u>Pooled T-test</u>

$$\boxed{t = \frac{(\bar{Y}_1 - \bar{Y}_2) - 0}{S_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim T_{(n_1+n_2-2)}} \;,\; \boxed{S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$$

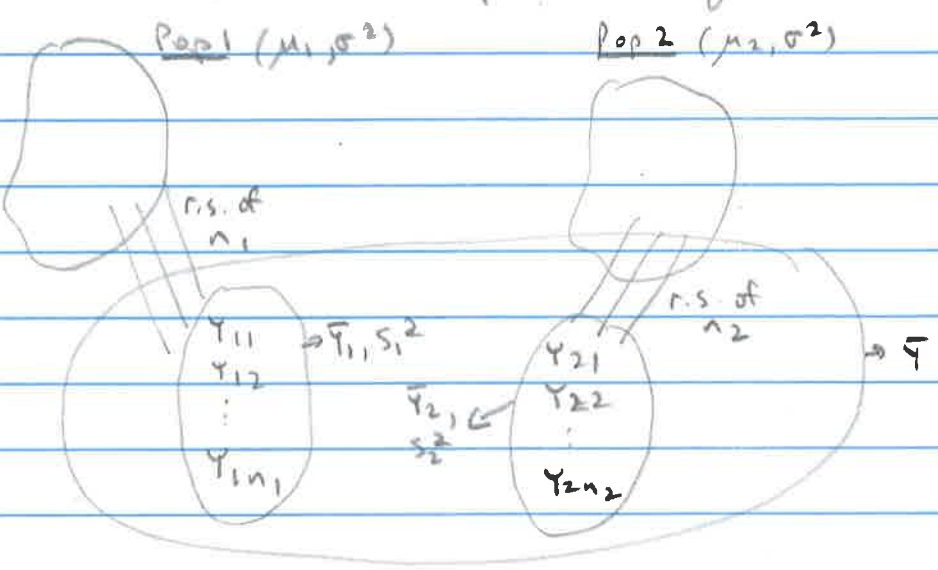
<u>note:</u> we developed the T-dist from

$$z = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}}} \sim N(0,1) \qquad V = \frac{(n_1+n_2-2)S_p^2}{\sigma^2} \sim \chi^2_{(n_1+n_2-2)}$$

<u>note:</u> assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$t = \frac{z}{\sqrt{V/(n_1+n_2-2)}}$$

We will analyze the same problem using an ANOVA procedure



Y_{ij} = j^{th} element from i^{th} pop. $(i=1,2 \quad j=1 \text{ to } n_i)$
$\bar{Y}_i$ = sample mean for i^{th} pop. $(i=1,2)$
$\bar{Y}$ = grand mean

$$\boxed{\text{Total SS} = \sum_{i=1}^{2} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2}$$ **Total variation in Y**

→ This quantifies all the variability in Y if we put all sample data into one group. The total variability can be partitioned into variability due to treatments and variability due to error.

$$\text{Total SS} = \sum_{i=1}^{2} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i + \bar{Y}_i - \bar{Y})^2$$

$$= \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left[(Y_{ij} - \bar{Y}_i) + (\bar{Y}_i - \bar{Y}) \right]^2$$

$$= \sum_{i=1}^{2} \sum_{j=1}^{n_i} \left[(Y_{ij} - \bar{Y}_i)^2 + 2(Y_{ij} - \bar{Y}_i)(\bar{Y}_i - \bar{Y}) + (\bar{Y}_i - \bar{Y})^2 \right]$$

$$= \sum_{i=1}^{2} \left[\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2 + 2(\bar{Y}_i - \bar{Y}) \underbrace{\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)}_{0} + n_i (\bar{Y}_i - \bar{Y})^2 \right]$$

$$= \underbrace{\sum_{i=1}^{2} \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_i)^2}_{SSE} + \underbrace{\sum_{i=1}^{2} n_i (\bar{Y}_i - \bar{Y})^2}_{SSTrT}$$

SSE → variability amongst EUs "treated alike" i.e. error

SSTrT → variability amongst EUs treated differently

Another way to write SSE is:

$$\sum_{i=1}^{2} (n_i - 1) s_i^2 = (n_1 - 1) s_1^2 + (n_2 - 1) s_2^2$$

$$= (n_1 + n_2 - 2) s_p^2$$

$$\Rightarrow \frac{SSE}{\sigma^2} = \frac{(n_1 + n_2 - 2) s_p^2}{\sigma^2} \sim \chi^2_{(n_1 + n_2 - 2)}$$

note: if $n_1 = n_2 = n$ then

$$\frac{SSE}{\sigma^2} = \frac{(2n-2) s_p^2}{\sigma^2} \sim \chi^2_{(2n-2)}$$

Now lets look at SS_{Trt} for balanced design $(n_1 = n_2 = n)$

$$SS_{Trt} = \sum_{i=1}^{2} n_i (\bar{Y}_i - \bar{Y})^2 = n \cdot \sum_{i=1}^{2} (\bar{Y}_i - \bar{Y})^2$$

$$= n \left[\left(\bar{Y}_1 - \left[\frac{\bar{Y}_1 + \bar{Y}_2}{2}\right]\right)^2 + \left(\bar{Y}_2 - \left[\frac{\bar{Y}_1 + \bar{Y}_2}{2}\right]\right)^2 \right]$$

$$= n \left[\left(\frac{\bar{Y}_1 - \bar{Y}_2}{2}\right)^2 + \left(\frac{\bar{Y}_2 - \bar{Y}_1}{2}\right)^2 \right]$$

$$= \frac{n (\bar{Y}_1 - \bar{Y}_2)^2}{2}$$

Now, for $n_1 = n_2 = n$ and $\sigma_1^2 = \sigma_2^2 = \sigma^2$ . . .

$H_0: \mu_1 - \mu_2 = 0$

$$z = \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{n}}} \stackrel{\downarrow}{=} \frac{(\bar{Y}_1 - \bar{Y}_2)\cdot\sqrt{n}}{\sqrt{2}\,\sigma} \sim N(0,1)$$

$$\Rightarrow \frac{n(\bar{Y}_1 - \bar{Y}_2)^2}{2\sigma^2} \sim \chi^2_{(1)} \quad \Rightarrow \quad \frac{SS_{TRT}}{\sigma^2} \sim \chi^2_{(1)}$$

For an ANOVA, we develop the test statistic as ...

$$F = \frac{\left(\frac{SS_{TRT}}{\sigma^2}\right)/1}{\left(\frac{SSE}{\sigma^2}\right)/(2n-2)} = \frac{SS_{TRT}/1}{SSE/(2n-2)} = \frac{MS_{TRT}}{MSE} \sim F_{1,2n-2}$$

note:

Source	df	SS	MS	F
Trt	1	SS_{TRT}	$MS_{TRT} = SS_{TRT}/1$	$F = \frac{MS_{TRT}}{MSE}$
Error	$2n-2$	SSE	$MSE = SSE/2n-2$	
Total	$2n-1$	SS_{total}		

$$F = \frac{\left(\frac{n(\bar{Y}_1 - \bar{Y}_2)^2}{2\sigma^2}\right)/1}{\left(\frac{(2n-2)S_p^2}{\sigma^2}\right)/(2n-2)} = \frac{n(\bar{Y}_1 - \bar{Y}_2)^2}{2S_p^2} = \frac{(\bar{Y}_1 - \bar{Y}_2)^2}{S_p^2\left(\frac{1}{n} + \frac{1}{n}\right)}$$

$$= \left(\frac{\bar{Y}_1 - \bar{Y}_2}{S_p\sqrt{\frac{1}{n} + \frac{1}{n}}}\right)^2 = t^2$$

i.e. $F = t^2$ for $k = 2$ grps

note: This result holds for $n_i \neq n_j$ but algebra gets messy!

e.g. Back to 10.71 body temps

Men	Women	
$n_1 = 9$	$n_2 = 9$	· assume $\sigma_1^2 = \sigma_2^2 = \sigma^2$
$\bar{y}_1 = 97.856$	$\bar{y}_2 = 98.4889$	· normal dists.
$s_1^2 = .340$	$s_2^2 = .301$	

$H_0: \mu_1 - \mu_2 = 0$

$H_a: \mu_1 - \mu_2 \neq 0$

we calculated $T = \dfrac{97.856 - 98.4884}{\sqrt{(.3207)\left(\frac{1}{9} + \frac{1}{9}\right)}} = -2.37$

Now,

$$\bar{y} = \frac{97.856 + 98.4889}{2} = 98.1725$$

$$SS_{TRT} = 9\left[(97.856 - 98.1725)^2 + (98.4889 - 98.1725)^2\right]$$

$$= 1.8025$$

$$SSE = 8(.340) + 8(.301) = 5.128$$

$$F = \frac{1.8025}{5.128/16} = 5.62 \qquad (-2.37)^2 = 5.62$$

13.3 Comparison of more than 2 means: ANOVA for one-way layout

Spse we are interested in testing $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$.
Even though we are testing means, we use a procedure called Analysis of Variance (ANOVA).

The types of designs we will examine in this section are completely randomized designs or one-way ANOVA.

e.g. We want to compare 5 types of animal diet
(1 factor: diet at 5 levels or 5 trts)
$H_0: \mu_1 = \mu_2 = \cdots = \mu_5$

---

e.g. Compare 5 types of corn seed and two types of soil
Factor 1: corn seed (5 levels)
Factor 2: soil type (2 levels)
We could run this as a one-way ANOVA with $5 \times 2 = 10$ trts, but we could estimate the separate effects of seed and soil with a 2-way ANOVA. If we had multiple EUs in each seed/soil combination, we could also estimate interaction i.e. best type of seed changes from one soil type to the next.

Back to one-way ANOVA ...

Data layout:

	Trt 1	2	...	k	
	y_{11}	y_{21}		y_{k1}	→ overall avg = $\bar{y}_{\cdot\cdot}$
	$\vdots$	$\vdots$		$\vdots$	
	y_{1n_1}	y_{2n_2}		y_{kn_k}	
	$\bar{y}_{1\cdot}$	$\bar{y}_{2\cdot}$		$\bar{y}_{k\cdot}$	

y_{ij} : j^{th} observation in i^{th} trt

$\bar{y}_{i\cdot}$ = i^{th} trt mean

$$= \frac{\sum_{j=1}^{n_i} y_{ij}}{n_i}$$

$$\bar{y}_{\cdot\cdot} = \text{overall mean} = \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij} \Big/ (n_1+\cdots+n_k)$$

Model:

means model $\quad y_{ij} = \mu_i + \varepsilon_{ij}$

μ_i = true mean of i^{th} trt

ε_{ij} = random error w/ j^{th} meas. in i^{th} trt , $\varepsilon_{ij} \sim iid\ N(0,\sigma^2)$

Sums of Squares

$$\boxed{\text{Total SS} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{\cdot\cdot})^2} = \boxed{\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}\right)^2}{n}}$$

→ represents the total variability amongst all EUs when they are combined in one group.

$$\boxed{SS_{TRT} = \sum_{i=1}^{k} n_i(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2} = \boxed{\sum_{i=1}^{k}\frac{Y_{i\cdot}^2}{n_i} - \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}\right)^2}{n}}$$

→ represents variability between trts ; i.e. variability amongst EUs treated differently

note: $Y_{i\cdot} = \sum_{j=1}^{n_i} y_{ij}$ = i^{th} trt total

$$\boxed{SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2} = \boxed{\sum_{i=1}^{k} (n_i - 1) s_i^2}$$

This is the error sum of squares and it quantifies the variability among EUs treated alike.

note: $\frac{SSE}{n-k} = \frac{(n_1-1)s_1^2 + \cdots + (n_k-1)s_k^2}{n-k}$ = "pooled" sample variance

As we know, the sums of squares are additive. We showed this for $k=2$ in the last section, but we will now verify that it is a general result.

$$y_{ij} - \bar{y}_{\cdot\cdot} = (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) + (y_{ij} - \bar{y}_{i\cdot})$$

$$\Rightarrow (y_{ij} - \bar{y}_{\cdot\cdot})^2 = (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 + 2(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})(y_{ij} - \bar{y}_{i\cdot}) + (y_{ij} - \bar{y}_{i\cdot})^2$$

Now sum over i and j . . .

$$\sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{\cdot\cdot})^2 = \sum_{i=1}^{k} \cdot \sum_{j=1}^{n_i} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2$$

$$+ 2 \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})(y_{ij} - \bar{y}_{i\cdot})$$

$$+ \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})^2$$

$$\Rightarrow \text{Total SS} = \sum_{i=1}^{k} n_i (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 + 2 \cdot \sum_{i=1}^{k} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})$$

$$+ SSE$$

$\Rightarrow$ Total SS = SS_{TRT} + SSE $+ \cancel{2 \cdot \sum_{i=1}^{k} (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) \cdot \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\cdot})} \to 0$

$\Rightarrow$ $\boxed{\text{Total SS} = SS_{TRT} + SSE}$

One-way ANOVA Table:

Source	df	SS	MS	F
Trts	$k-1$	SS_{TRT}	$MS_{TRT} = \frac{SS_{TRT}}{k-1}$	$F = \frac{MS_{TRT}}{MSE}$
Error	$n-k$	SSE	$MSE = SSE/(n-k)$	$\sim F_{k-1,\,n-k}$
Total	$n-1$	Total SS		

Assumptions:

(1) Sampling from normal pops. (robust)

(2) $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$ or $n_1 = n_2 = \cdots = n_k$

note: MSE is the pooled estimate of the common variance σ^2

Example 13.2

Four grps of students subjected to different individualized tutoring techniques.

$\boxed{\text{Table 13.2 p.671}}$

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_a: at least one μ_i is different

- n_i are different, how about σ_i^2? $\sigma_1^2 = \cdots = \sigma_4^2$ looks reasonable

$$SS_{TRT} = \left(\frac{454^2}{6} + \frac{549^2}{7} + \frac{425^2}{6} + \frac{351^2}{4} \right) - \frac{(1779)^2}{23}$$

$$= 712.6$$

$$SS_{TOT} = (65^2 + 87^2 + \cdots + 88^2) - \frac{1779^2}{23} = 139{,}511 - 137{,}602$$

$$= 1{,}909.2$$

Source	df	SS	MS	F
Trt	3	712.6	237.5	3.77
Error	19	1196.6	62.98	
Total	22	1,909.2		

p-value = .02808

$\boxed{\text{Boxplots}}$

13.5

$W = U + V$, U, V are ind. RVs

$W \sim \chi^2_{(r)}$, $V \sim \chi^2_{(s)}$

$$E[e^{tW}] = E[e^{t(U+V)}] \overset{\text{ind.}}{=} E[e^{tU}] \cdot E[e^{tV}]$$

$$M_W(t) = M_U(t) \cdot M_V(t)$$

$$\frac{1}{(1-2t)^{r/2}} = M_U(t) \cdot \frac{1}{(1-2t)^{s/2}} \Rightarrow$$

(b)

$$m_u(t) = \frac{1}{(1-2t)^{r/2}} \cdot \frac{1}{(1-2t)^{s/2}} = \frac{1}{(1-2t)^{(r-s)/2}}$$

$$\Rightarrow \boxed{u \sim \chi^2_{(r-s)}}$$

13.6

Trt 1	2	⋯	k
y_{11}	y_{21}		y_{k1}
⋮	⋮		⋮
y_{1n_1}	y_{2n_2}		y_{kn_k}

Recall the ANOVA assumptions of normality / equal variances. If H_0 is true, then the data above are k different ind. R.S. from identical dists ($Y \sim N(\mu, \sigma^2)$)

$$\frac{SSE}{\sigma^2} = \frac{\sum_{i=1}^{k} (n_i - 1) \cdot S_i^2}{\sigma^2} = \underbrace{\frac{(n_1-1)S_1^2}{\sigma^2} + \frac{(n_2-1)S_2^2}{\sigma^2} + \cdots + \frac{(n_k-1)S_k^2}{\sigma^2}}_{\text{ind. samples}}$$

$$= \chi^2_{(n_1-1)} + \chi^2_{(n_2-1)} + \cdots + \chi^2_{(n_k-1)}$$

$$= \chi^2_{(n_1+\cdots+n_k-k)} = \chi^2_{(n-k)}$$

$$\Rightarrow \boxed{\frac{SSE}{\sigma^2} \sim \chi^2_{(n-k)}}$$

Now, if H_0 is true Y_{ij} is a r.s. of size $n = n_1 + \cdots + n_k$ from $Y \sim N(\mu, \sigma^2)$

Thm 7.3

$$\Rightarrow \frac{\sum_i \sum_j (Y_{ij} - \bar{Y}_{..})^2}{\sigma^2} = \frac{\text{Total SS}}{\sigma^2} \sim \chi^2_{(n-1)}$$

Let $W = \frac{\text{Total SS}}{\sigma^2}$ and $V = \frac{SSE}{\sigma^2}$

$W \sim \chi^2_{(n-1)}$ and $V \sim \chi^2_{(n-k)}$

$W = U + V$, $U = \frac{SS_{TRT}}{\sigma^2}$

note: Since U is a fct of sample means from normal dists.
and V is a fct. of sample variances from normal dists.
$\Rightarrow$ U, V are ind. (Thm. 7.3)

So by (13.5) result

$$W = U + V \Rightarrow U \sim \chi^2_{(n-1-(n-k))} \sim \chi^2_{(k-1)}$$

$$\text{or } \frac{SS_{TRT}}{\sigma^2} \sim \chi^2_{(k-1)}$$

$$\Rightarrow F = \frac{\frac{SS_{TRT}}{\sigma^2} / (k-1)}{\frac{SSE}{\sigma^2} / (n-k)} = \frac{SS_{TRT}/(k-1)}{SSE/(n-k)} = \frac{MS_{TRT}}{MSE} \sim F_{k-1, n-k}$$

i.e. $F = \frac{MS_{TRT}}{MSE}$ is ratio of ind. χ^2 RVs divided by their degrees of freedom, $k-1$ and $n-k$ for the numerator and denominator respectively

13.5/13.6 A Statistical Model for the One-Way Layout / Expected MS

Means model from before:

$$Y_{ij} = \mu_i + \varepsilon_{ij} \quad , \quad \varepsilon_{ij} \sim iid\ N(0, \sigma^2)$$

Effects model:

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim iid\ N(0, \sigma^2)$$

μ = overall mean

τ_i = non-random effect of i^{th} trt

note: we generally assume $\sum_{i=1}^{k} \tau_i = 0$ i.e. the trt effects sum to 0

ε_{ij} = random error associated w/ j^{th} subject in i^{th} trt

For the means model $H_0: \mu_1 = \cdots = \mu_k = \mu$

Equivalent

For the effects model $H_0: \tau_1 = \cdots = \tau_k = 0$

also tests for no significant treatment differences.

The entire ANOVA table (SS, MS, df, F) are identical in both models. The effects model is used to compute expected mean squares, $E(MS_{Trt})$ and $E(MSE)$

Expected Mean Squares

Back to the effects model:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad , \quad \varepsilon_{ij} \sim iid\ N(0, \sigma^2)$$

To make the algebra easier we will assume a balanced design ($n_1 = n_2 = \cdots = n_k = n$)

$$\bar{y}_{i\cdot} = \frac{\sum_{i=1}^{n}(\mu+\tau_i+\varepsilon_{ij})}{n} = \frac{n\mu+n\tau_i+\sum_{i=1}^{n}\varepsilon_{ij}}{n} = \mu+\tau_i+\bar{\varepsilon}_{i\cdot}$$

Since $\varepsilon_{ij} \sim iid\ N(0,\sigma^2) \Rightarrow \bar{\varepsilon}_{i\cdot} \sim N(0, \sigma^2/n)$

note: this also means $E(\bar{\varepsilon}_{i\cdot}^2) = \frac{\sigma^2}{n} + 0 = \sigma^2/n$

$$\bar{y}_{\cdot\cdot} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n} y_{ij}}{k \cdot n} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n}(\mu+\tau_i+\varepsilon_{ij})}{k \cdot n} = \mu + \frac{\sum_{i=1}^{k}\tau_i}{k} + \bar{\varepsilon}_{\cdot\cdot} = \mu+\bar{\varepsilon}_{\cdot\cdot}$$

$$\bar{\varepsilon}_{\cdot\cdot} \sim N(0, \sigma^2/kn) \Rightarrow E[\bar{\varepsilon}_{\cdot\cdot}^2] = \frac{\sigma^2}{kn} + 0^2 = \frac{\sigma^2}{kn}$$

Now, $MS_{TRT} = \dfrac{\sum_{i=1}^{k} n_i(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2}{k-1} = \dfrac{n\sum_{i=1}^{k}(\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2}{k-1}$

$$\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot} = (\mu+\tau_i+\bar{\varepsilon}_{i\cdot})-(\mu+\bar{\varepsilon}_{\cdot\cdot})^2 = \tau_i + (\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon}_{\cdot\cdot})$$

$$\Rightarrow (\bar{y}_{i\cdot}-\bar{y}_{\cdot\cdot})^2 = \tau_i^2 + 2\tau_i(\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon}_{\cdot\cdot}) + (\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon}_{\cdot\cdot})^2$$

$$\Rightarrow MS_{TRT} = \frac{n}{k-1}\left[\sum_{i=1}^{k}\tau_i^2 + 2\sum_{i=1}^{k}\tau_i(\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon}_{\cdot\cdot}) + \sum_{i=1}^{k}(\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon}_{\cdot\cdot})^2\right]$$

$$\Rightarrow E(MS_{TRT}) = \frac{n}{k-1}\left[\sum_{i=1}^{k}\tau_i^2 + 2\cdot\cancelto{0}{\sum_{i=1}^{k}\tau_i E[\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon}_{\cdot\cdot}]} + \sum_{i=1}^{k}E[(\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon}_{\cdot\cdot})^2]\right]$$

now look at $\sum_{i=1}^{k} E[(\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon}_{\cdot\cdot})^2] = E\left[\sum_{i=1}^{k}(\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon}_{\cdot\cdot})^2\right]$

$$\sum_{i=1}^{k}(\bar{\varepsilon}_{i\cdot}-\bar{\varepsilon}_{\cdot\cdot})^2 = \sum_{i=1}^{k}(\bar{\varepsilon}_{i\cdot}^2 - 2\bar{\varepsilon}_{i\cdot}\cdot\bar{\varepsilon}_{\cdot\cdot} + \bar{\varepsilon}_{\cdot\cdot}^2)$$

$$= \sum_{i=1}^{k} \bar{\varepsilon}_{i\cdot}^2 - 2\bar{\varepsilon}_{\cdot\cdot} \sum_{i=1}^{k} \bar{\varepsilon}_{i\cdot} + k \cdot \bar{\varepsilon}_{\cdot\cdot}^2 = \sum_{i=1}^{k} \bar{\varepsilon}_{i\cdot}^2 - 2k\bar{\varepsilon}_{\cdot\cdot}^2 + k\bar{\varepsilon}_{\cdot\cdot}^2$$

$$= \sum_{i=1}^{k} \bar{\varepsilon}_{i\cdot}^2 - k \sum_{i=1}^{k} \bar{\varepsilon}_{\cdot\cdot}^2$$

So,

$$E\left[\sum_{i=1}^{k} \bar{\varepsilon}_{i\cdot}^2 - k\bar{\varepsilon}_{\cdot\cdot}^2\right] = \sum_{i=1}^{k} E(\bar{\varepsilon}_{i\cdot}^2) - kE\left[\bar{\varepsilon}_{\cdot\cdot}^2\right]$$

$$= \frac{k\sigma^2}{n} - k \cdot \left(\frac{\sigma^2}{kn}\right) = \frac{(k-1)\sigma^2}{n}$$

$$\Rightarrow E(MS_{TRT}) = \frac{n}{k-1} \sum_{i=1}^{k} \tau_i^2 + \frac{n}{k-1} \cdot \frac{(k-1)}{n}\sigma^2$$

$$= \boxed{\frac{n}{k-1} \sum_{i=1}^{k} \tau_i^2 + \sigma^2}$$

Under $H_0: \tau_i = 0 \Rightarrow E(MS_{TRT}) = \sigma^2$

<u>note</u>: If H_0 is false MS_{TRT} should get large!

<u>Expected Mean Square Error</u> (MSE)

For a balanced design:

$$MSE = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i\cdot})^2}{n-k}$$

$$y_{ij} - \bar{y}_{i\cdot} = (\mu + \tau_i + \varepsilon_{ij}) - (\mu + \tau_i + \bar{\varepsilon}_{i\cdot}) = \varepsilon_{ij} - \bar{\varepsilon}_{i\cdot}$$

$$\sum_{i=1}^{k} \sum_{j=1}^{n} (\varepsilon_{ij} - \bar{\varepsilon}_{i\cdot})^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} (\varepsilon_{ij}^2 - 2\varepsilon_{ij}\bar{\varepsilon}_{i\cdot} + \bar{\varepsilon}_{i\cdot}^2)$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{n} \varepsilon_{ij}^2 - 2 \sum_{i=1}^{k} \bar{\varepsilon}_{i\cdot} \cdot \sum_{j=1}^{n} \varepsilon_{ij} + \sum_{i=1}^{k} n \cdot \bar{\varepsilon}_{i\cdot}^2$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{n} \varepsilon_{ij}^2 - 2n \sum_{i=1}^{k} \bar{\varepsilon}_{i\cdot}^2 + n \sum_{i=1}^{k} \bar{\varepsilon}_{i\cdot}^2$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{n} \varepsilon_{ij}^2 - n \sum_{i=1}^{k} \bar{\varepsilon}_{i\cdot}^2$$

$$\Rightarrow E\left[\sum_{i=1}^{k} \sum_{j=1}^{n} \varepsilon_{ij}^2 - n \sum_{i=1}^{k} \bar{\varepsilon}_{i\cdot}^2 \right]$$

$$= nk(\sigma^2 + 0^2) - nk\left(\frac{\sigma^2}{n} + 0^2 \right) = nk\sigma^2 - k\sigma^2$$

$$= (nk - k)\sigma^2$$

$$\Rightarrow E(MSE) = \frac{(nk-k)\sigma^2}{nk-k} = \boxed{\sigma^2}$$

So to Summarize...

- when H_0 is true, we "expect" MS_{TRT} and MSE to be near 0

  so $F = \dfrac{MS_{TRT}}{MSE} \approx 1$

- when H_0 is false, we "expect" MS_{TRT} to be "large" relative to MSE

$$F \approx \frac{\frac{n}{k-1} \sum_{i=1}^{k} \tau_i^2 + \sigma^2}{\sigma^2}$$

13.7 Estimation in the One-Way Layout

If we reject $H_0: \mu_1 = \mu_2 = \cdots = \mu_k$ then we conclude at least one trt mean is different. This leads us to estimating what the actual trt means are.

Estimate μ_i

note: We still have the normality assumption and equal variances

i.e. $Y_{ij} \sim N(\mu_i, \sigma^2)$

$\Rightarrow \bar{Y}_{i\cdot} \sim N(\mu_i, \sigma^2/n_i)$

$\Rightarrow$ $100(1-\alpha)\%$ CI for μ (σ unknown)

$$\boxed{\bar{Y}_{i\cdot} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n_i}}}$$

s = "pooled" sample std. dev. $= \sqrt{MSE}$

d.f. $= (n_1 + \cdots + n_k) - k = n - k$

Example 13.3

Find 95% CI for mean score of tutoring technique 1

$$75.67 \pm \underset{\text{d.f}=14}{(2.093)} \frac{\sqrt{62.98}}{\sqrt{6}} = \boxed{75.67 \pm 6.78}$$

Estimate $\mu_i - \mu_\ell$:

$Y_{ij} \sim N(\mu_i, \sigma^2)$ $\qquad$ $Y_{\ell j} \sim N(\mu_\ell, \sigma^2)$

$\bar{Y}_{i\cdot} \sim N(\mu_i, \frac{\sigma^2}{n_i})$ $\qquad$ $\bar{Y}_{\ell\cdot} \sim N(\mu_\ell, \sigma^2/n_\ell)$

$\Rightarrow$ $100(1-\alpha)\%$ CI for $\mu_i - \mu_\ell$:

$$(\bar{Y}_{i\cdot} - \bar{Y}_{\ell\cdot}) \pm t_{\alpha/2} \cdot s \sqrt{\frac{1}{n_i} + \frac{1}{n_\ell}}$$

again, $s = \sqrt{MSE}$, d.f. $= (n_1 + \cdots + n_k) - k = n - k$

Example 13.4

Find 95% CI for $\mu_1 - \mu_4$:

$$(75.67 - 87.75) \pm 2.093 \sqrt{62.98} \sqrt{\frac{1}{6} + \frac{1}{4}}$$

$$-12.08 \pm 10.73 \Rightarrow \boxed{(-22.81, -1.35)}$$

$\Rightarrow$ We believe tutoring technique 4 produces significantly higher scores than tutoring technique 1

note:

1. These CIs are fine if we only compute 1. If we make several pairwise comparisons, we need to control inflated type I error. (Bonferroni, Tukey, Dunnett, etc.)

2. Best practice: Planned Orthogonal Contrasts