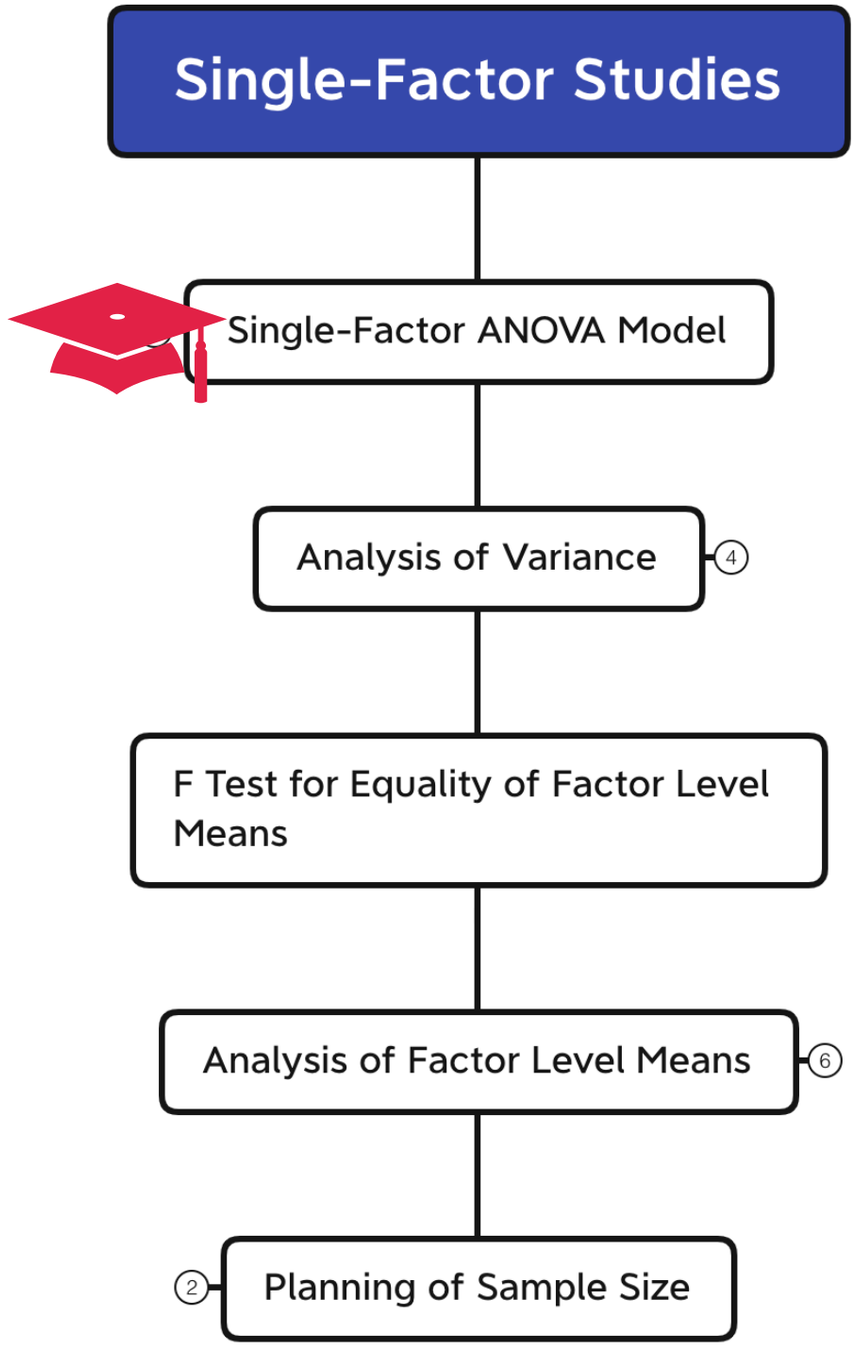
Lecture 2:
Single-Factor Studies

STA 106: Analysis of Variance

Suggested reading: ALSM Chapter 16 & 17

Xiner Zhou
Department of Statistics, University of California, Davis


Single-Factor Studies
Single-Factor ANOVA Model
Analysis of Variance
4
F Test for Equality of Factor Level Means
Analysis of Factor Level Means
6
2
Planning of Sample Size

Example

(The Rehab Study)

The objective of the study:
A rehabilitation center researcher was interested in the relationship between
physical fitness prior to surgery of corrective knee surgery
Time required in physical therapy until successful rehabilitation

The study setup:
Patient records during the past year were examined: 24 male patients with age 18-30

		j									
i		**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	Below Average	29	42	38	40	43	40	30	42	•	
2	Average	30	35	39	28	31	31	29	35	29	33
3	Above Average	26	32	21	20	23	22				

Substantive Research Questions of Interest:

Whether the factor levels or treatments differ in terms of response?
If the factor levels differ in terms of response, in what way do they differ or how do they differ?

These research questions lead to statistical questions usually performed in two steps, correspondingly.

Single-Factor ANOVA Model

r : number of factor levels

Y_{ij} : observed value of the outcome or response variable for the jth unit in ith factor level

Subscript $i = 1 \ldots r$: ith factor level

n_i : number of units in ith factor level

Subscript $j = 1 \ldots n_i$: jth unit in a given factor level

$n_T = \sum_{i=1}^{r} n_i$: total number of units in the study

Single-Factor ANOVA Model

Assume: observed value of response variable is the sum of two components

- Treatment mean = factor level mean μ_i — fixed term
- Error term ϵ_{ij} — random term

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

- ϵ_{ij} error term

But not technically errors (mistakes), it reflects all other extraneous factors that influence the response but are not measured or not considered in current study

We assume $\epsilon_{ij}'s$ are independently and identically distributed as $N(0,\sigma^2)$, for all$i$, j

$$\epsilon_{ij} \sim N(0,\sigma^2)$$

Single-Factor ANOVA Model

Assume: observed value of response variable is the sum of two components

- Treatment mean = factor level mean μ_i — fixed term
- Error term ϵ_{ij} — random term

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

- μ_i

$$E\left(Y_{ij}\right) = E\left(\mu_i + \varepsilon_{ij}\right) = \mu_i + E\left(\varepsilon_{ij}\right) = \mu_i$$

Mean response of ith factor level or treatment

Interpretation:

Experimental study

μ_i is the mean response that would be obtained if the ith treatment were applied to all units in the population under study

Observational study

μ_i is the mean response for ith factor level subpopulation

Single-Factor ANOVA Model

Assume: observed value of response variable is the sum of two components

- Treatment mean = factor level mean μ_i — fixed term
- Error term ϵ_{ij} — random term

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

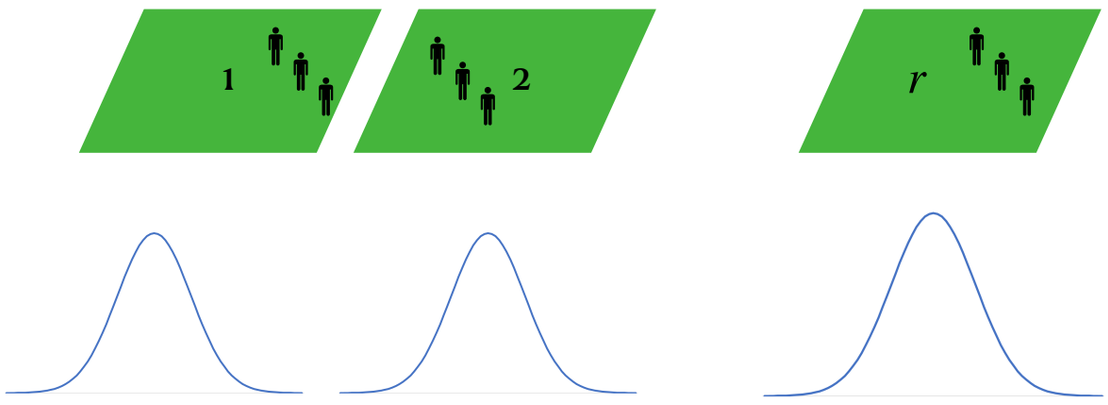
- What is the distribution of Y_{ij} ?

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

Different treatment means depend on which treatment this unit is from

same variance of error term: homogeneity of error variance

Y_{ij} is a random draw from a normal population with mean μ_i and variance σ^2



r sub-populations corresponding to r treatments, each one follows a normal distribution with different means, but same variance

Single-Factor ANOVA Model

Assume: observed value of response variable is the sum of two components

- Treatment mean = factor level mean μ_i — fixed term
- Error term ϵ_{ij} — random term

$$Y_{ij} = \mu_i + \epsilon_{ij}$$

- Unknown parameters

  Treatment mean = factor level mean $\mu_1 \ldots \mu_r$

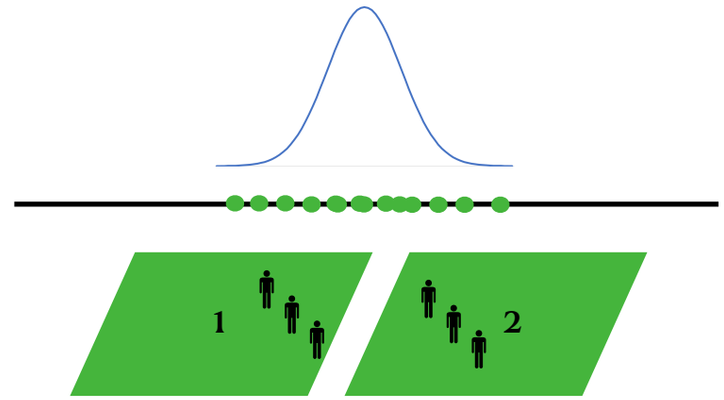
  Error variance σ^2

→ How do we estimate them?

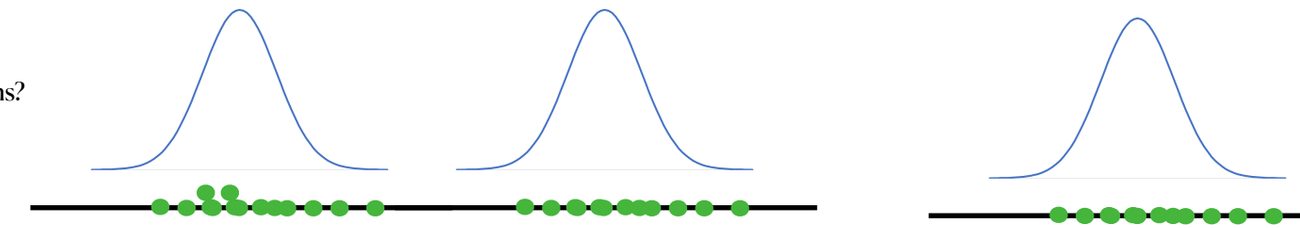
Fitting the ANOVA Model (Estimate Model Parameters $\mu_1 \ldots \mu_r$)

Treatment mean = factor level mean $\mu_1 \ldots \mu_r$?

Single population?



Many subpopulations?



Fitting the ANOVA Model (Estimate Model Parameters $\mu_1 \ldots \mu_r$)

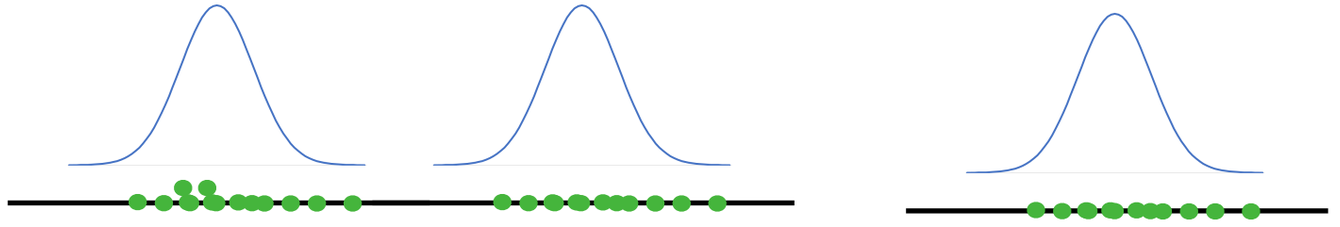
Treatment mean = factor level mean $\mu_1 \ldots \mu_r$?

Least Squares Method:

The estimates for the centers of populations $\hat{\mu}_1 \ldots \hat{\mu}_r$ should minimize the dispersion in the data so that each observation Y_{ij} is as close as possible to its corresponding mean μ_i

$$Q(\mu_1 \ldots \mu_r) = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \mu_i)^2 = \sum_{j=1}^{n_1} (Y_{ij} - \mu_1)^2 + \sum_{j=2}^{n_i} (Y_{ij} - \mu_2)^2 + \ldots + \sum_{j=1}^{n_r} (Y_{ij} - \mu_r)^2$$

How to measure dispersion in the data: sum of squared deviations of observations away from its population means



$$\hat{\mu}_1 = \arg\min_{\mu_1} Q(\mu_1 \ldots \mu_r) \quad \hat{\mu}_2 = \arg\min_{\mu_2} Q(\mu_1 \ldots \mu_r) \quad \hat{\mu}_r = \arg\min_{\mu_r} Q(\mu_1 \ldots \mu_r)$$

Fitting the ANOVA Model (Estimate Model Parameters $\mu_1 \ldots \mu_r$)

Treatment mean = factor level mean $\mu_1 \ldots \mu_r$?

Least Squares Method:

$\hat{\mu}_1 = \arg\min_{\mu_1} Q(\mu_1 \ldots \mu_r)$ $\hat{\mu}_2 = \arg\min_{\mu_2} Q(\mu_1 \ldots \mu_r)$ $\hat{\mu}_r = \arg\min_{\mu_r} Q(\mu_1 \ldots \mu_r)$

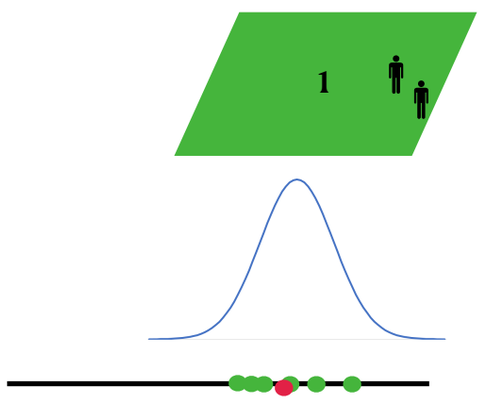
Why it's a simpler than it looks?

When minimizes with respect to μ_1, only the first term matters, other terms are constants

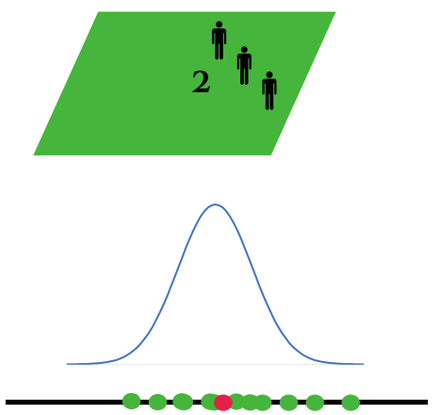
$$\frac{dQ}{d\mu_1} = \frac{d}{d\mu_1} \sum_{j=1}^{n_1} (Y_{ij} - \mu_1)^2 = -2 \sum_{j=1}^{n_1} (Y_{ij} - \mu_1) = 0$$

Least squares estimates (LSE):

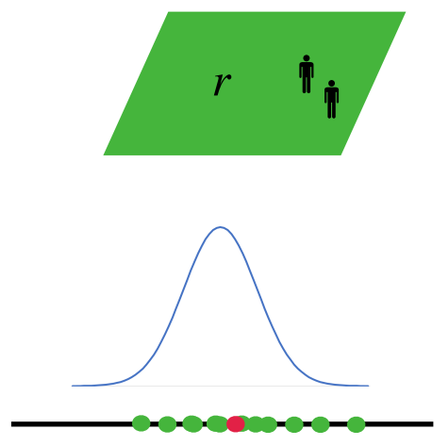
$$\hat{\mu}_1 = \frac{\sum_{i=1}^{n_1} Y_{1j}}{n_1} = \overline{Y_{1\cdot}}$$



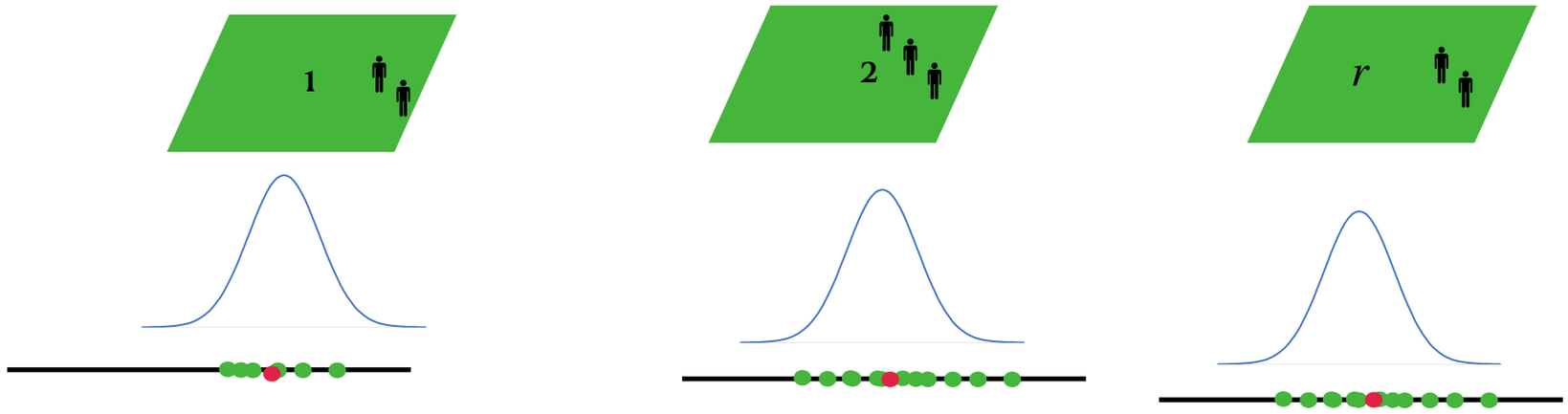
$$\hat{\mu}_2 = \frac{\sum_{i=1}^{n_2} Y_{2j}}{n_2} = \overline{Y_{2\cdot}}$$



$$\hat{\mu}_r = \frac{\sum_{i=1}^{n_r} Y_{rj}}{n_r} = \overline{Y_{r\cdot}}$$



Fitting the ANOVA Model (Estimate Model Parameters $\mu_1 \ldots \mu_r$)



- fitted value for an observation Y_{ij}

  ANOVA model's "best guess" or "best prediction" for Y_{ij}

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

- residual e_{ij} corresponds to observation Y_{ij} is

$$e_{ij} = \underbrace{Y_{ij} - \hat{Y}_{ij}} = Y_{ij} - \bar{Y}_{i\cdot}$$

Difference between observed value and fitted value which is estimated factor level mean

Residuals are approximations or estimation for the error term

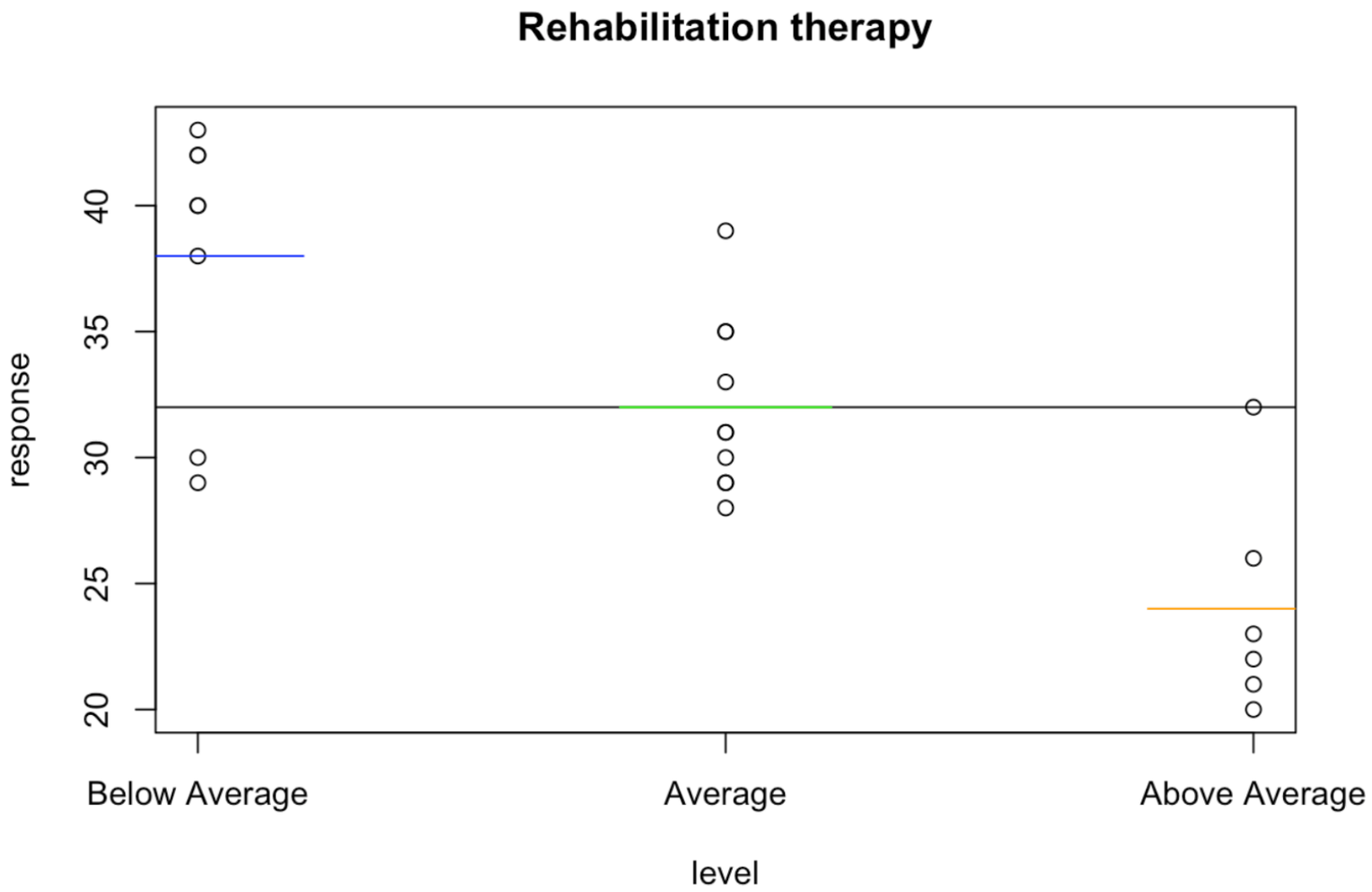
$$\hat{\epsilon}_{ij} = e_{ij}$$

Residuals are highly useful for checking whether assumptions of ANOVA Model is appropriate for the data at hand

Residuals sum to 0 for each treatment :

For ith treatment: $\sum_{j=1}^{n_i} e_i = \sum_{j=1}^{n_i} Y_{ij} - \overline{Y_{i\cdot}} = 0$

Example



Do the factor level means appear to differ?

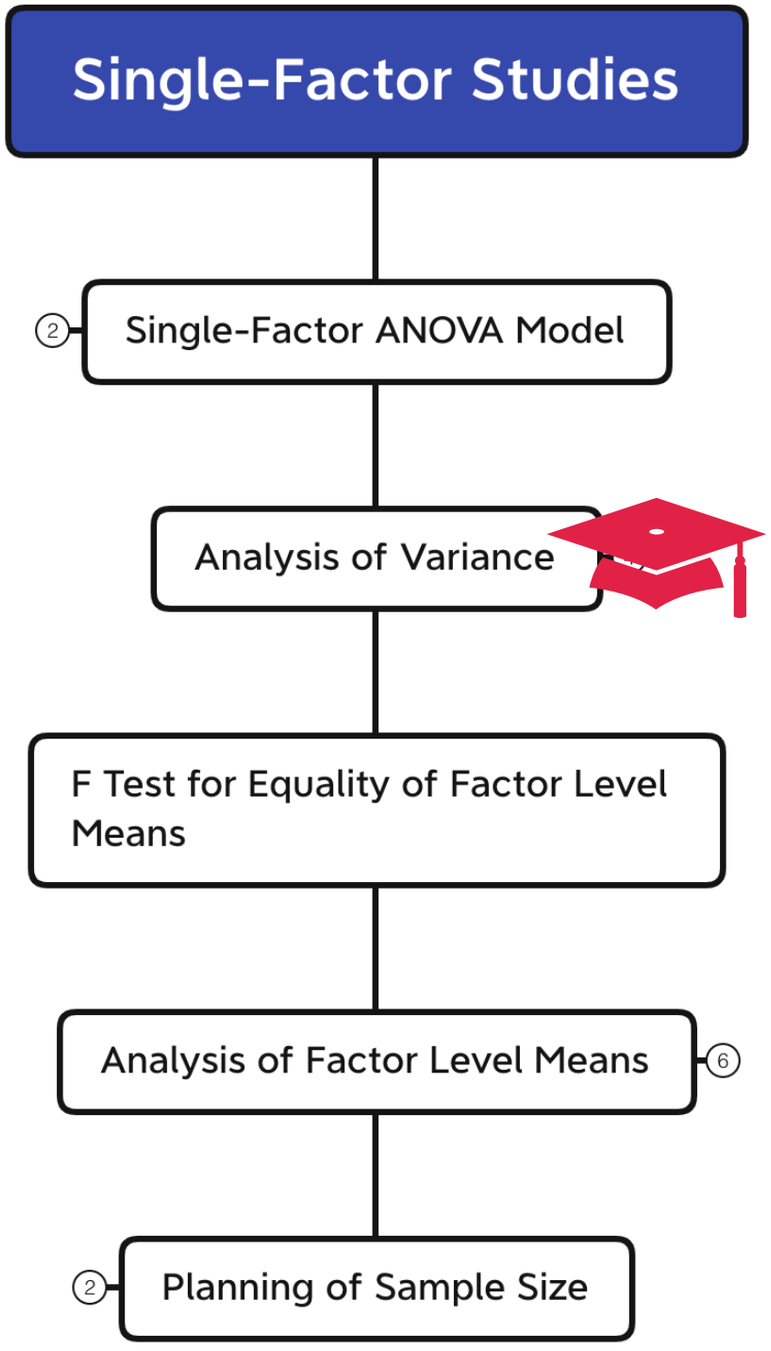
Does the variability of the observations within each factor level appear to be approximately the same for all factor levels?

Example

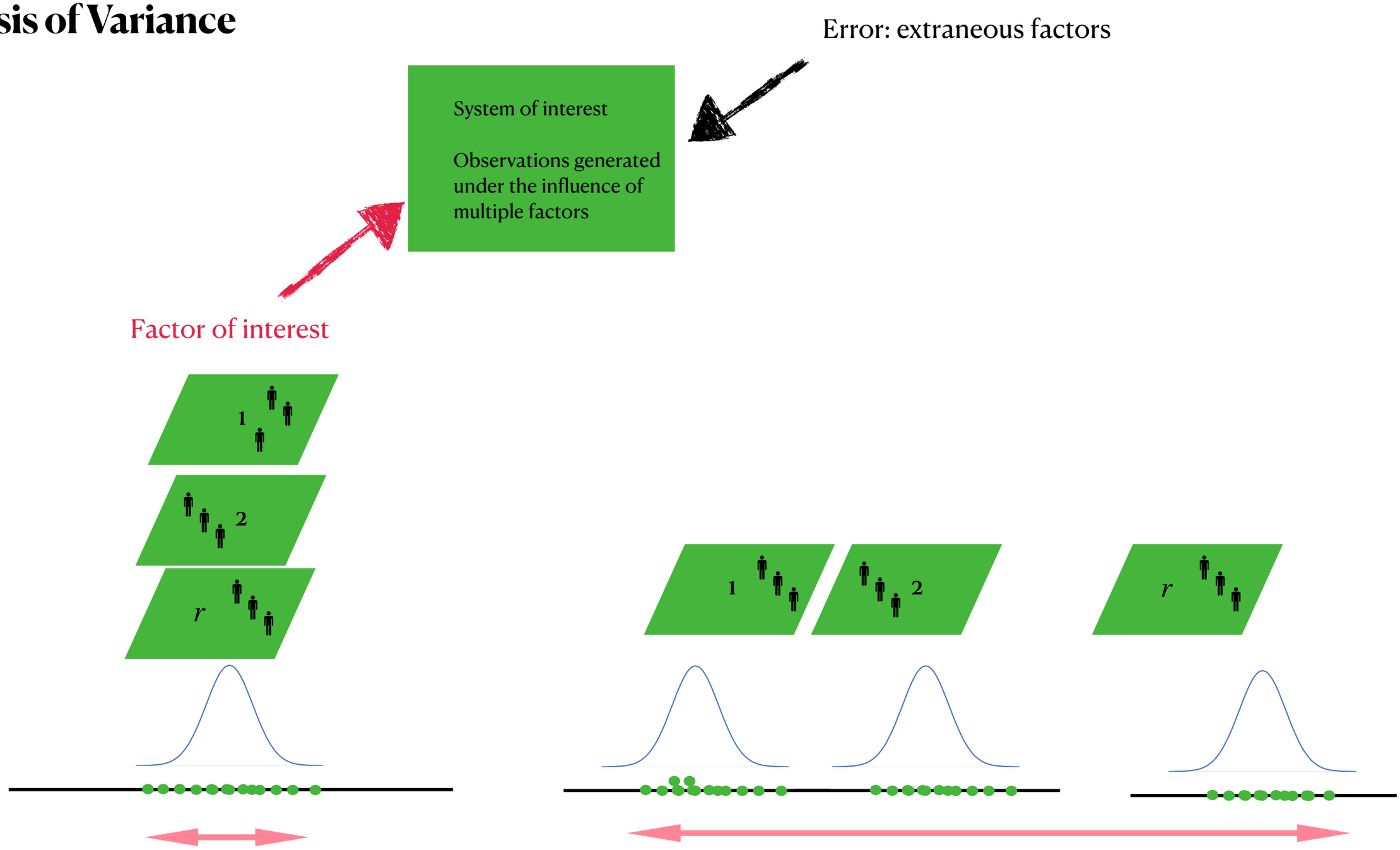
Fitted values and Residuals

Treatment levels	Response|fitted value|residual
1	29 | 38 | -9
1	42 | 38 | 4
1	38 | 38 | 0
1	40 | 38 | 2
1	43 | 38 | 5
1	40 | 38 | 2
1	30 | 38 | -8
1	42 | 38 | 4
2	30 | 32 | -2
2	35 | 32 | 3
2	39 | 32 | 7
2	28 | 32 | -4
2	31 | 32 | -1
2	31 | 32 | -1
2	29 | 32 | -3
2	35 | 32 | 3
2	29 | 32 | -3
2	33 | 32 | 1
3	26 | 24 | 2
3	32 | 24 | 8
3	21 | 24 | -3
3	20 | 24 | -4
3	23 | 24 | -1
3	22 | 24 | -2

Do residuals sum to zero within each treatment ?


Single-Factor Studies
2
Single-Factor ANOVA Model
Analysis of Variance
F Test for Equality of Factor Level Means
Analysis of Factor Level Means
6
2
Planning of Sample Size

Analysis of Variance



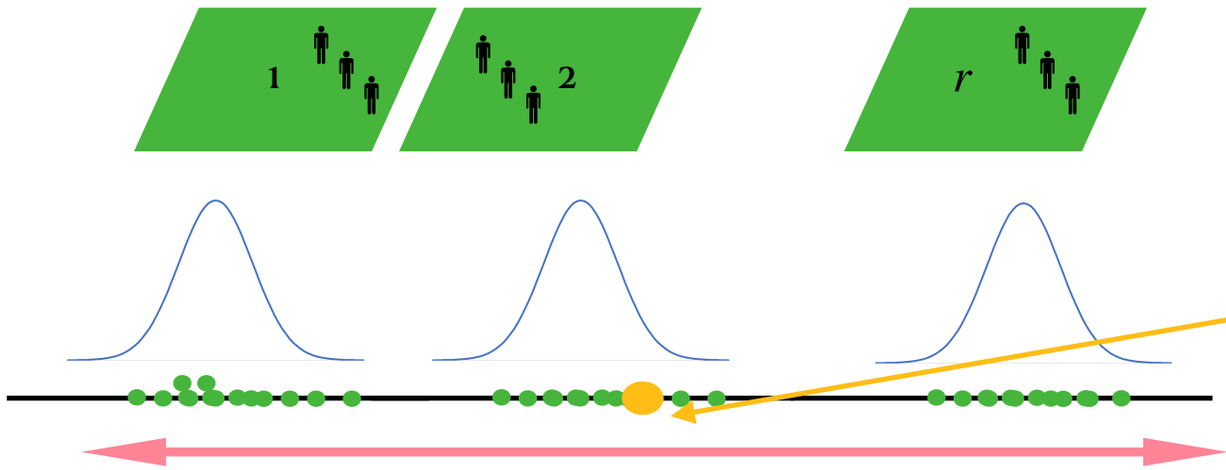
Without factor of interest, the observations have some natural variation due to other extraneous factors, i.e. “error variance”

If the factor of interest indeed has some effects on the system, then we would expect more volatility than a system without the factor

Analysis of Variance

Idea : Partition sums of squares (= variation) associated with responses $Y'_{ij}s$

Notion of "Total Variation":



$$\overline{Y}_{..} = \frac{1}{n_T}\sum_{i=1}^{r}\sum_{j=1}^{n_i} Y_{ij}$$

Each observation Y_{ij} deviates from overall sample mean by $Y_{ij} - \overline{Y}_{..}$

→ Measure of total variation is sum of squared deviations

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

SSTO: Total Sum of Squares

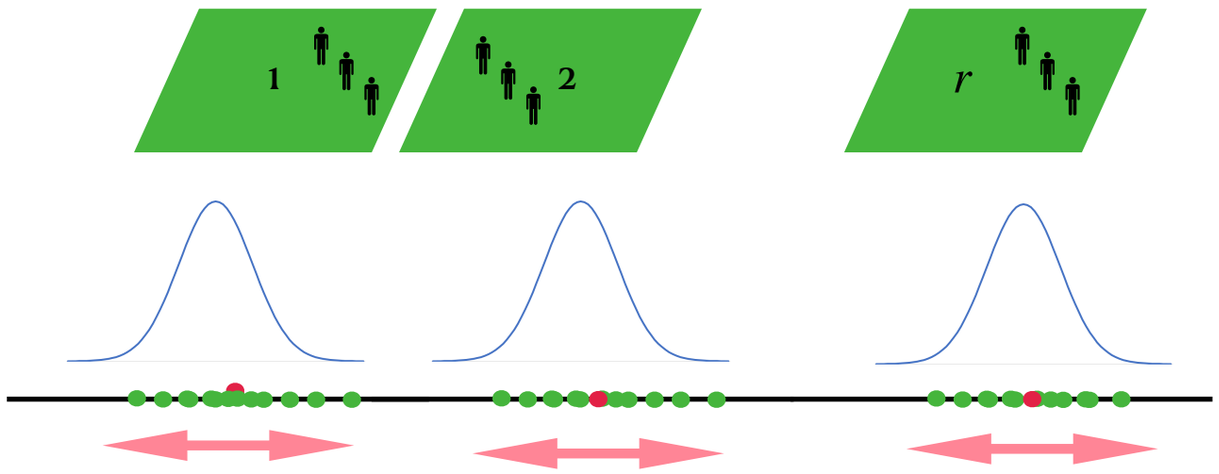
- If all $Y'_{ij}s$ are the same -> $SSTO = 0$
- If there is more variation among $Y'_{ij}s$ —> SSTO increases

→ SSTO: measures total variation or uncertainty observed in data, while these variation can be due to many different factors (reasons)

Analysis of Variance

Idea : Partition sums of squares (= variation) associated with responses $Y'_{ij}s$

Notion of "Variation due to error":



When we consider the factor under study, the variation within treatment is due to extraneous factors that we collectively call "error".

Each observation Y_{ij} deviates from treatment-specific sample mean by $Y_{ij} - \overline{Y_{i\cdot}}$

Measure of variation due to error is sum of squared deviations

$$SSE = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y_{i\cdot}})^2 = \sum_{i=1}^{r} \sum_{j=1}^{n_i} e_{ij}^2$$

SSE: Error Sum of Squares or Residual Sum of Squares

- If there is no extraneous factors influencing the response, then we would expect all $Y'_{ij}s$ within each treatment to be exactly the same
  $Y_{ij} = \overline{Y_{i\cdot}}$ —> SSE = 0

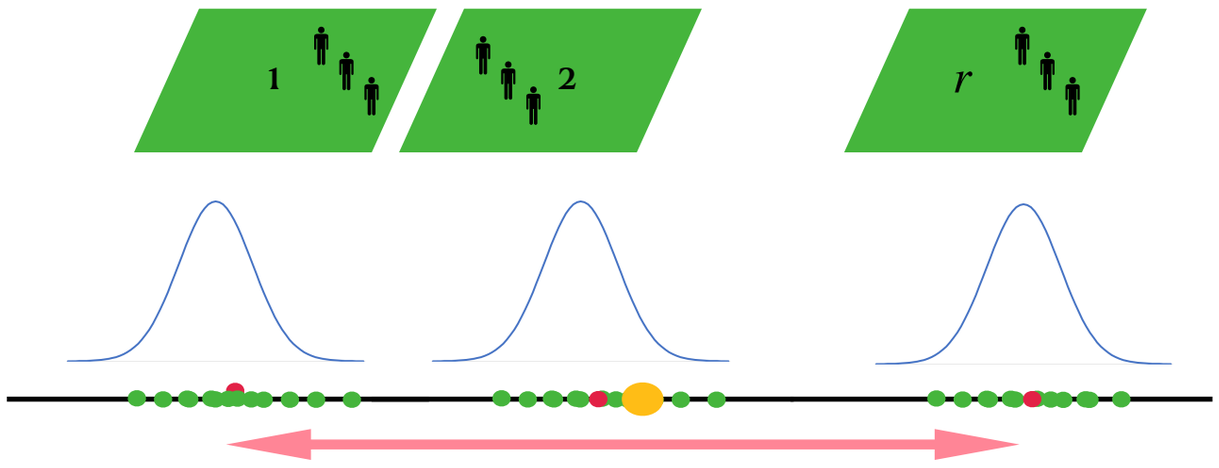
- If there are many extraneous factors influencing the response, then we would expect $Y'_{ij}s$ vary dramatically within each treatment
  Y_{ij} and $\overline{Y_{i\cdot}}$ very different —> SSE large

SSE: measures residual variation or remaining variation observed in data, that is not due to treatment, but due to extraneous factors

Analysis of Variance

Idea : Partition sums of squares (= variation) associated with responses $Y_{ij}'s$

Notion of "Variation due to treatment":



For each observation Y_{ij}, we expect some variation for the fact that they belong to different treatment groups that have different treatment means

So we expect to see some variation in each observation Y_{ij} by an amount $\overline{Y_{i.}} - \overline{Y_{..}}$

→ Measure of variation due to treatment is sum of squared deviations

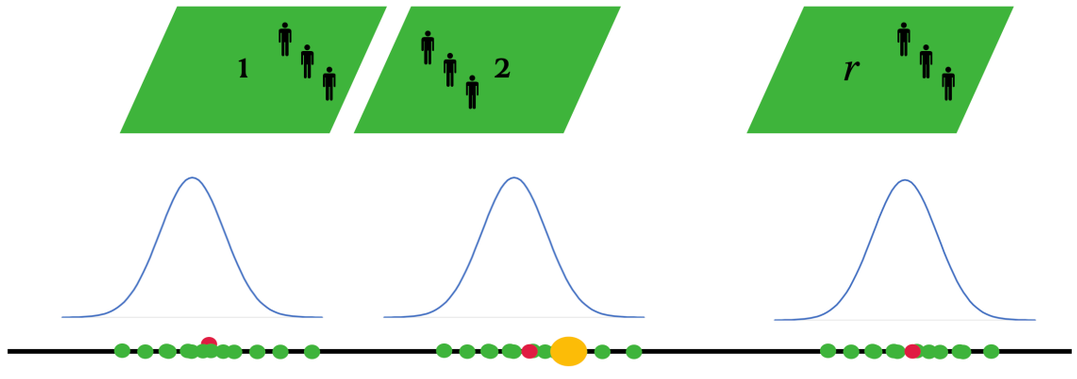
$$SSTR = \sum_{i=1}^{r} \sum_{j=1}^{n_i} (\overline{Y_{i.}} - \overline{Y_{..}})^2 = \sum_{i=1}^{r} n_i (\overline{Y_{i.}} - \overline{Y_{..}})^2$$

SSTR: Treatment Sum of Squares

- If there is no difference whatsoever in the treatment means, that is, the factor has no treatment effect at all, then we would expect SSTR to be small
- If there is huge difference in the treatment means, that is, the factor under study do make a difference, then we would expect SSTR to be large

→ SSTR: measures variation due to the factor under study

Analysis of Variance



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

Total deviation | Deviation around estimated treatment mean | Deviation of estimated treatment mean around overall mean

$$\rightarrow \quad (Y_{ij} - \overline{Y_{..}})^2 = (Y_{ij} - \overline{Y_{i.}})^2 + (\overline{Y_{i.}} - \overline{Y_{..}})^2 + 2(Y_{ij} - \overline{Y_{i.}})(\overline{Y_{i.}} - \overline{Y_{..}})$$

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

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

Observed that: $\sum_{j=1}^{n_i}\left(Y_{ij} - \overline{Y_{i.}}\right) = 0$ for all i

Convince yourself if not clear

$$= 2\sum_{i=1}^{r}\left(\overline{Y_{i.}} - \overline{Y_{..}}\right)0 = 0$$

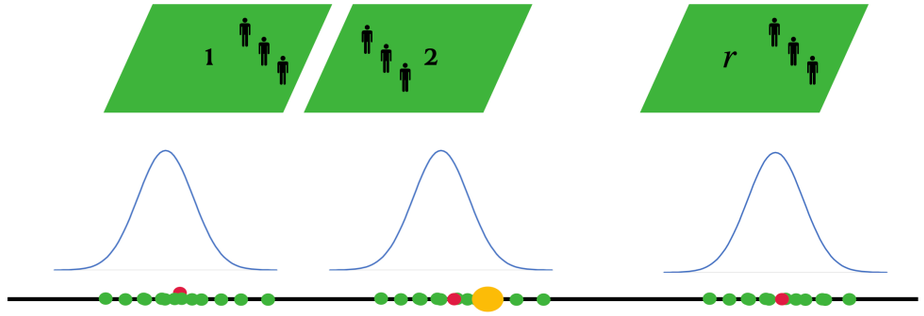
$$\rightarrow \quad \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y_{..}})^2 = \sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y_{i.}})^2 + \sum_{i=1}^{r} n_i(\overline{Y_{i.}} - \overline{Y_{..}})^2$$

Total variation | Variation due to extraneous factors/ error | Variation due to factor or treatment under study

SSTO: Total sum of squares | SSE: Error sum of squares | SSTR: Treatment sum of squares

Analysis of Variance

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



Total variation

Variation due to extraneous factors/ error

Variation due to factor or treatment under study

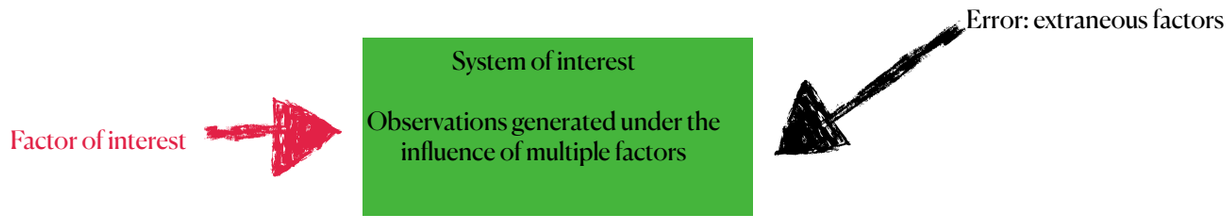
SSTO: Total sum of squares

SSE: Error sum of squares

SSTR: Treatment sum of squares

Partition of SSTO:

SSTO	=	SSTR	+	SSE



Can we separate the force that comes from the factor under study and the extraneous forces not of interest?

Yes! by analyzing and decomposing variation

- If there is no difference whatsoever in the treatment means, that is, the factor has no treatment effect at all, then we would expect SSTR to be small
- If there is huge difference in the treatment means, that is, the factor under study do make a difference, then we would expect SSTR to be large

This will be our intuition for detecting (testing) if the factor under study indeed has effect versus not!

Degrees of Freedom

Estimates of parameters can be based on different amount of “independent” information.
The number of independence pieces of information that go into the estimate of a parameter is called:
degree of freedom (d.o.f.)

Think of: dimensions of the space where an estimator lives in and allows to run free

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

Total variation — Variation due to extraneous factors/ error — Variation due to factor or treatment under study

SSTO: Total sum of squares — SSE: Error sum of squares — SSTR: Treatment sum of squares

How many independence pieces of information go into each quantity?

$Y_{ij}-\overline{Y_{..}}$

n_T pieces

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

$df(SSTO)=n_T-1$

$Y_{ij}-\overline{Y_{i.}}$

n_T pieces

But $\sum_{j=1}^{n_i}(Y_{ij}-\overline{Y_{i.}})=0$ for $i=1\ldots r$

$df(SSE)=n_T-r$

$\overline{Y_{i.}}-\overline{Y_{..}}$

r pieces

But $\sum_{i=1}^{r} n_i((\overline{Y_{i.}}-\overline{Y_{..}})=0$

$df(SSTR)=r-1$

$$df(SSTO)=df(SSE)+df(SSTR)$$

Mean Squares

$$\text{Mean Squares} = \frac{Sum\ of\ Squares}{d.f.}$$

Variation due to difference sources, but in equal comparison footing (in unit d.f.)

Treatment Mean Square:

$$\text{MSTR} = \frac{SSTR}{d.f.(SSTR)} = \frac{SSTR}{r-1}$$

Error Mean Square:

$$\text{MSE} = \frac{SSE}{d.f.(SSE)} = \frac{SSE}{n_T - r}$$

What's expected values of Mean Squares?

$$MSE = \frac{1}{n_T - r} \sum_{i=1}^{r} \sum_{j=1}^{n_i} (Y_{ij} - \overline{Y_{i.}})^2$$

$$= \frac{1}{n_T - r} \sum_{i=1}^{r} (n_i - 1) \underbrace{\frac{\sum_{j=1}^{n_i} (Y_{ij} - \overline{Y_{i.}})^2}{n_i - 1}}_{S_i^2 \text{ : sample variance}}$$

$$= \frac{1}{n_T - r} \sum_{i=1}^{r} (n_i - 1) S_i^2$$

$$E[MSE] = \frac{1}{n_T - r} \sum_{i=1}^{r} (n_i - 1) \underbrace{E[S_i^2]}_{\text{Sample variance is unbiased estimate of population variance which is error variance in this case}}$$

$$= \sigma^2$$

What's expected values of Mean Squares?

$$MSTR = \frac{1}{r-1}\sum_{i=1}^{r} n_i(\overline{Y_{i\cdot}} - \overline{Y_{\cdot\cdot}})^2$$

$$(\overline{Y_{i\cdot}} - \overline{Y_{\cdot\cdot}})^2 = (\overline{Y_{i\cdot}} - \frac{\sum_{i=1}^{r} n_i\overline{Y_{i\cdot}}}{n_T})^2$$

$$= ((1 - \frac{n_i}{n_T})\overline{Y_{i\cdot}} - \frac{\sum_{k\neq i} n_i\overline{Y_{k\cdot}}}{n_T})^2$$

$$= ((1 - \frac{n_i}{n_T})(\overline{Y_{i\cdot}} - \mu_i + \mu_i) - \sum_{k\neq i}\frac{n_i}{n_T}(\overline{Y_{k\cdot}} - \mu_k + \mu_k))^2$$

$$= ((1 - \frac{n_i}{n_T})(\overline{Y_{i\cdot}} - \mu_i) + (1 - \frac{n_i}{n_T})\mu_i - \sum_{k\neq i}\frac{n_i}{n_T}(\overline{Y_{k\cdot}} - \mu_k) - \sum_{k\neq i}\frac{n_i}{n_T}\mu_k)^2$$

$$= ((1 - \frac{n_i}{n_T})(\overline{Y_{i\cdot}} - \mu_i) - \sum_{k\neq i}\frac{n_i}{n_T}(\overline{Y_{k\cdot}} - \mu_k))^2 + ((1 - \frac{n_i}{n_T})\mu_i - \sum_{k\neq i}\frac{n_i}{n_T}\mu_k)^2 + 2((1 - \frac{n_i}{n_T})(\overline{Y_{i\cdot}} - \mu_i) - \sum_{k\neq i}\frac{n_i}{n_T}(\overline{Y_{k\cdot}} - \mu_k))((1 - \frac{n_i}{n_T})\mu_i - \sum_{k\neq i}\frac{n_i}{n_T}\mu_k)$$

$$E[*] = (1 - \frac{n_i}{n_T})^2\frac{\sigma^2}{n_i} + \sum_{k\neq i}(\frac{n_k}{n_T})^2\frac{\sigma^2}{n_k}$$

$$= \frac{\sigma^2}{n_T^2}\{\frac{(n_T - n_i)^2}{n_i} + \sum_{k\neq i} n_k\}$$

$$E[*] = (\mu_i - \overline{\mu_{\cdot}})^2$$

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

$$E[*] = 0$$

What's expected values of Mean Squares?

$$E[MSTR] = \frac{1}{r-1}\sum_{i=1}^{r} n_i \left(\frac{\sigma^2}{n_T^2}\{\frac{(n_T - n_i)^2}{n_i} + \sum_{k\neq i} n_k\} + \left(\mu_i - \overline{\mu}_.\right)^2 \right)$$

$$= \frac{1}{r-1}\sum_{i=1}^{r} \frac{\sigma^2}{n_T^2}\left((n_T - n_i)^2 + n_i(n_T - n_i)\right) + \frac{1}{r-1}\sum_{i=1}^{r} n_i(\mu_i - \overline{\mu}_.)^2$$

$$= \frac{1}{r-1}\sum_{i=1}^{r} \frac{\sigma^2}{n_T}\left(n_T - n_i\right) + \frac{1}{r-1}\sum_{i=1}^{r} n_i(\mu_i - \overline{\mu}_.)^2$$

$$= \frac{r}{r-1}\sigma^2 - \frac{1}{r-1}\sum_{i=1}^{r} \frac{n_i}{n_T}\sigma^2 + \frac{1}{r-1}\sum_{i=1}^{r} n_i(\mu_i - \overline{\mu}_.)^2$$

$$= \sigma^2 + \frac{1}{r-1}\sum_{i=1}^{r} n_i(\mu_i - \overline{\mu}_.)^2$$

$$\overline{\mu}_. = \frac{1}{n_T}\sum_{i=1}^{r} n_i\mu_i$$

What's the distributions of Mean Squares?

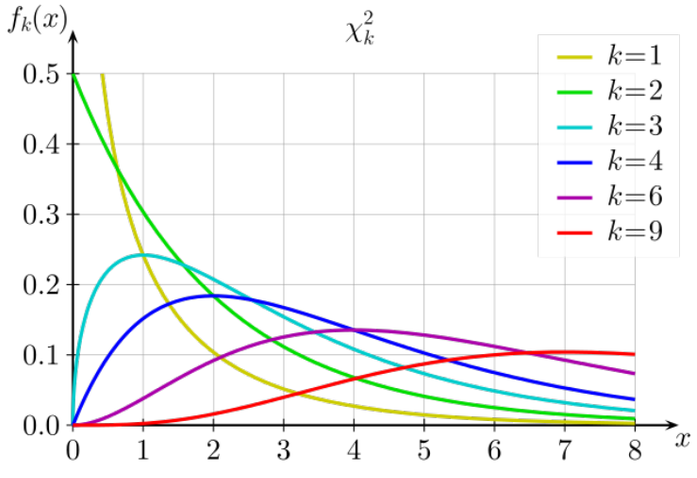
$$\frac{n_T - r}{\sigma^2} MSE = \frac{SSE}{\sigma^2} \sim \chi^2_{n_T - r}$$

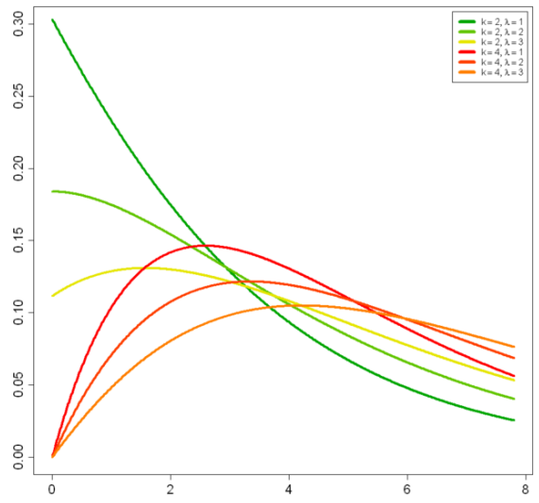
Chi-square distribution with degree of freedom $n_T - r$

$$\frac{r-1}{\sigma^2} MSTR = \frac{SSTR}{\sigma^2} \sim \chi^2_{r-1} \left(\frac{1}{\sigma^2} \sum_{i=1}^{r} n_i \left(\mu_i - \overline{\mu}_. \right)^2 \right)$$

Non-central Chi-square distribution with degree of freedom $r - 1$, and non-central parameter $\frac{1}{\sigma^2} \sum_{i=1}^{r} n_i \left(\mu_i - \overline{\mu}_. \right)^2$

MSE and MSTR are independent random variables





ANOVA Table for Single-factor Studies

Source of Variation	*SS*	*df*	*MS*	*E{MS}*
Between treatments	$SSTR = \sum n_i(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$	$r - 1$	$MSTR = \frac{SSTR}{r-1}$	$\sigma^2 + \frac{\sum n_i(\mu_i - \mu_\cdot)^2}{r-1}$
Error (within treatments)	$SSE = \sum\sum(Y_{ij} - \bar{Y}_{i\cdot})^2$	$n_T - r$	$MSE = \frac{SSE}{n_T - r}$	σ^2
Total	$SSTO = \sum\sum(Y_{ij} - \bar{Y}_{\cdot\cdot})^2$	$n_T - 1$		

$$E[MSE] = \sigma^2$$

→ The expected value of MSE is the error variance

MSE is an unbiased estimate for σ^2

Thus, we use MSE as the estimate for the parameter σ^2

$$\hat{\sigma}^2 = MSE = \frac{1}{n_T - r}\sum_{i=1}^{r}\sum_{j=1}^{n_i}(Y_{ij} - \overline{Y_{i\cdot}})^2$$

ANOVA Table for Single-factor Studies

Intuition confirmed by mathematical derivation from analyzing variance:

$$E[MSTR] = \sigma^2 + \frac{1}{r-1}\sum_{i=1}^{r} n_i(\mu_i - \overline{\mu}_\cdot)^2$$

$$\vee\!\!\wedge$$

$$E[MSE] = \sigma^2$$

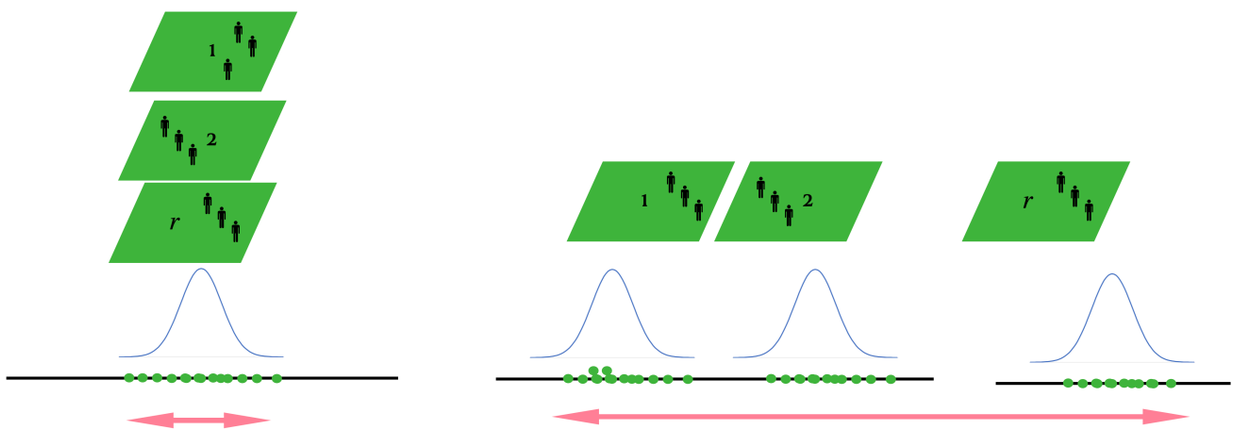
= holds if and only if when $\mu_i = \overline{\mu}_\cdot$ for all i

→ When all μ_i are equal: $E[MSTR] = E[MSE]$

MSTR will tend to be very close to MSE

When not all μ_i are equal:

MSTR tends to be larger than MSE,
how much larger is determined by how much different among population treatment means $\frac{1}{r-1}\sum_{i=1}^{r} n_i(\mu_i - \overline{\mu}_\cdot)^2$



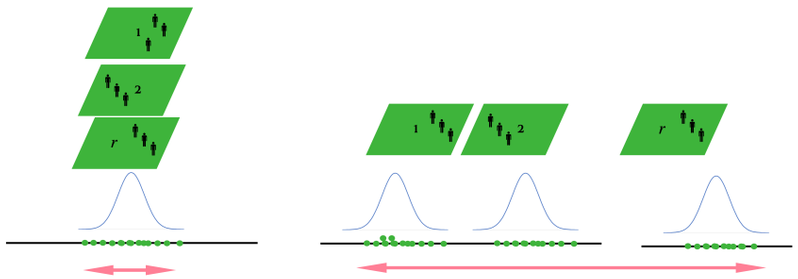
ANOVA Table for Single-factor Studies

Intuition confirmed by mathematical derivation from analyzing variance:

$$E[MSTR] = \sigma^2 + \frac{1}{r-1}\sum_{i=1}^{r} n_i(\mu_i - \overline{\mu}_\cdot)^2$$

$$\Vdash$$

$$E[MSE] = \sigma^2$$



This unique property of MSE and MSTR (two sources of variance) gives us tool to construct test that can signal which situation is more likely to be true:

If we observe $MSTR \approx MSE$

Then it signals that population treatment means are more likely to be the same, that is, the factor does not have any effect on the response.

If we observe $MSTR >> MSE$

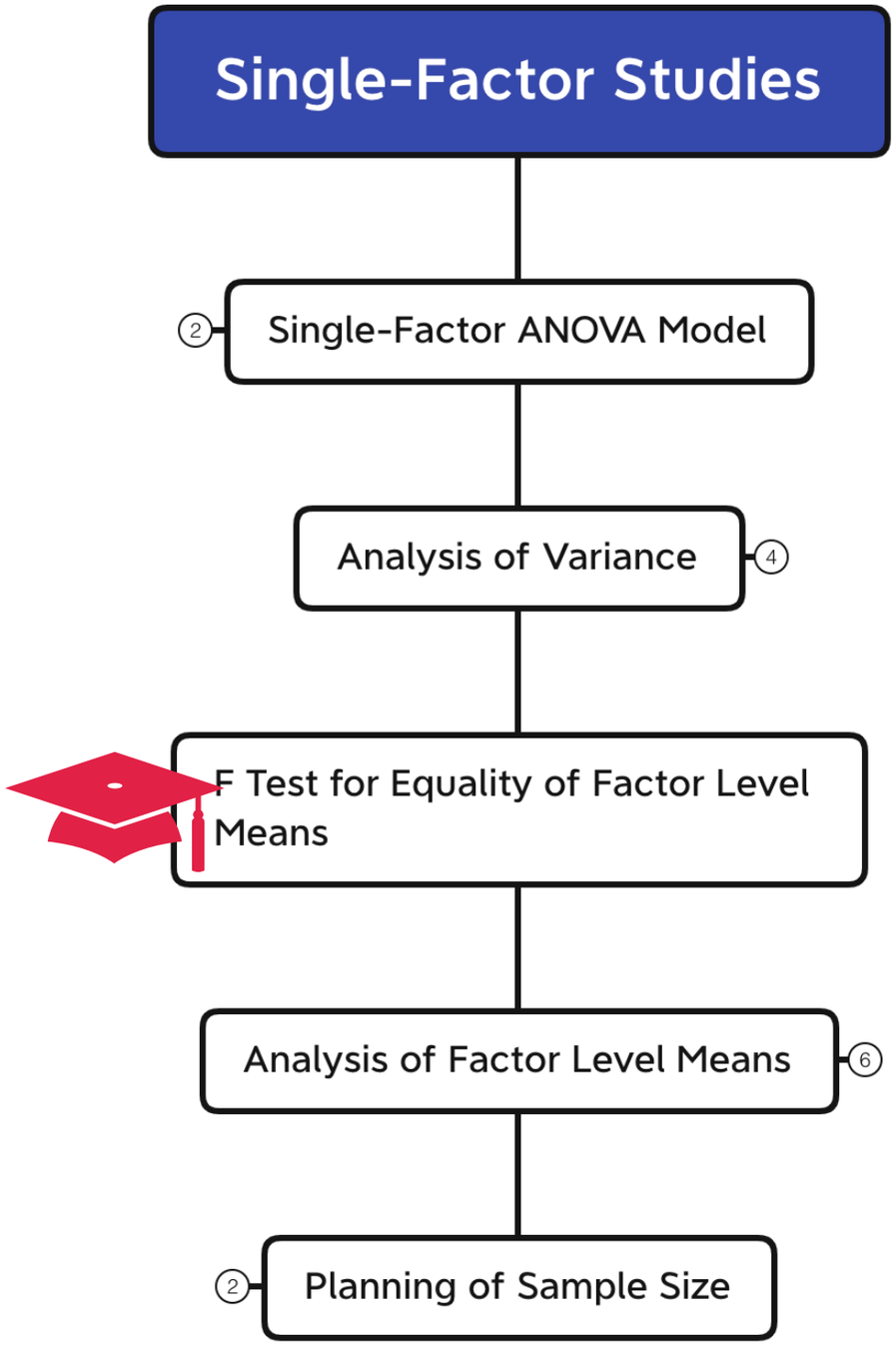
Then it signals that population treatment means are more likely to be not the same, that is, the factor does not have some effect on the response.

Geometry of Decomposition of Variance:

Example

ANOVA Table

	SS	df	MS
Between treatments	672	2	336.0
Error (within treatments)	416	21	19.8
Total	1088	23	0.0


Single-Factor Studies
2
Single-Factor ANOVA Model
Analysis of Variance
4
F Test for Equality of Factor Level Means
Analysis of Factor Level Means
6
2
Planning of Sample Size

F Test for Equality of Factor Level Means

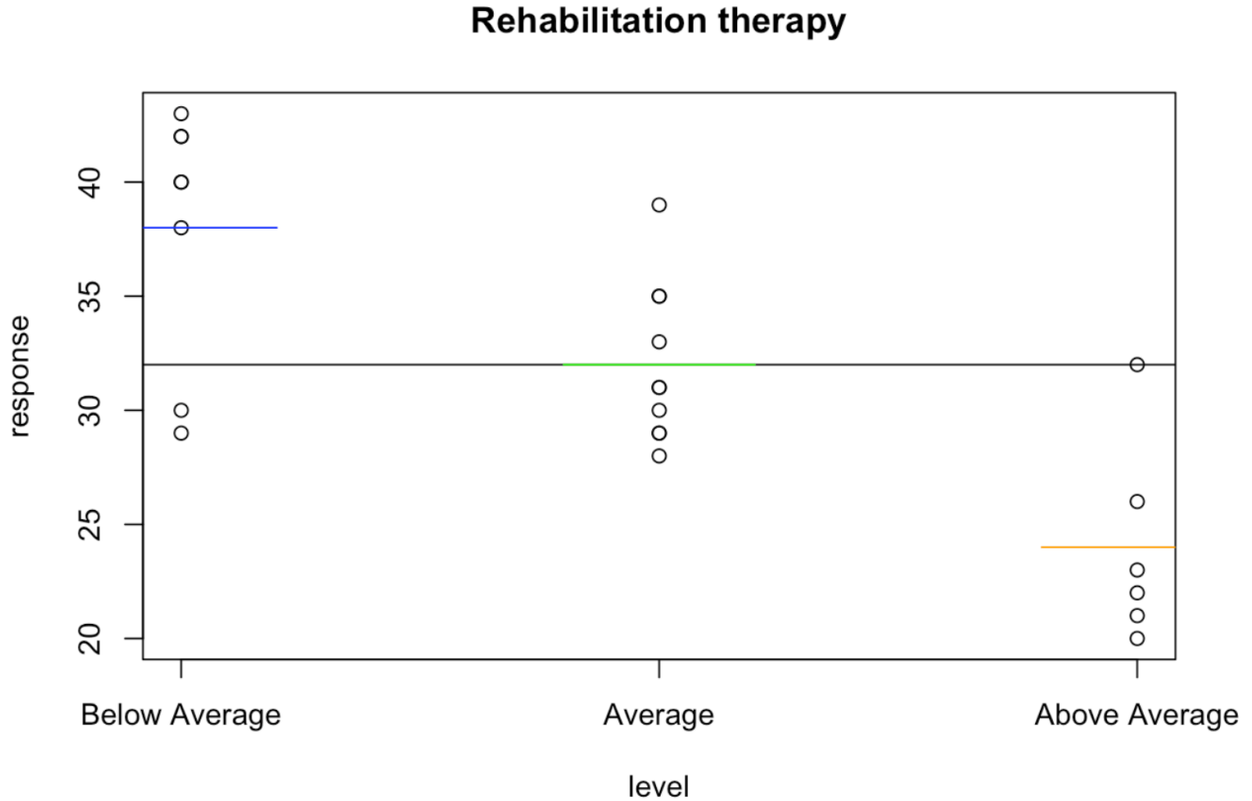
Substantive Research Questions of Interest:

Whether the factor levels or treatments differ in terms of response?
If the factor levels differ in terms of response, in what way do they differ or how do they differ?

These research questions lead to statistical questions usually performed in two steps, correspondingly.

Statistical Questions of Interest:

Whether the factor level means $\mu_i's$ are all equal or not?



F Test for Equality of Factor Level Means

To test whether or not the factor level means are the same:

$H_0 : \mu_1 = u_2 = \ldots = \mu_r$

H_a : not all μ_r are equal

Test statistic: $F^* = \dfrac{MSTR}{MSE}$

Large value of F^* support H_a

Small value, when $F^* \approx 1$ support H_0

—> We reject H_0 for large value of F^*, i.e. $F^* \geq c$

How do we decide what is “large” and what is “small”?

		Decision	
		H_0	H_a
Truth	H_0	✓	Type I error
	H_a	Type II error	✓

F Test for Equality of Factor Level Means

We want to control type I error at significance level α (usually .05):

If H_0 is true: $\mu_1 = \ldots = \mu_r$

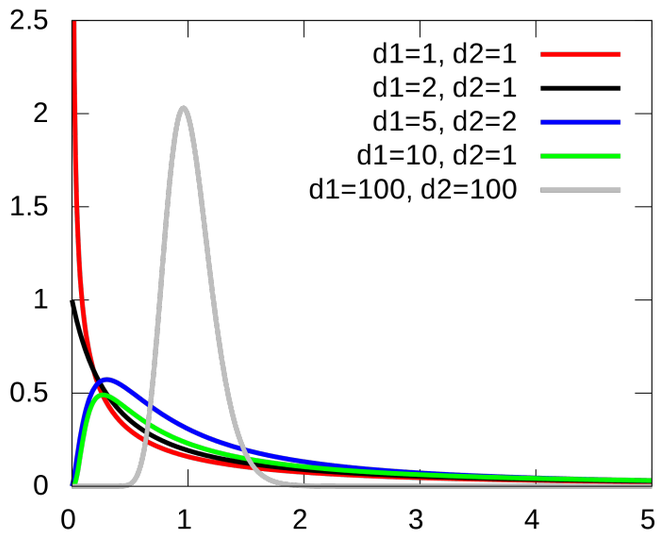
$$\frac{n_T - r}{\sigma^2} MSE = \frac{SSE}{\sigma^2} \sim \chi^2_{n_T - r}$$

$$\frac{r-1}{\sigma^2} MSTR = \frac{SSTR}{\sigma^2} \sim \chi^2_{r-1} \left(\frac{1}{\sigma^2} \left(\mu_i - \overline{\mu}_. \right)^2 \right) = \chi^2_{r-1}$$

MSE and MSTR are independent random variables

Re-write the test statistic:

$$\rightarrow \quad F^* = \frac{MSTR}{MSE} = \frac{\frac{(r-1)MSTR}{\sigma^2}}{\frac{(n_T - r)MSE}{\sigma^2}} \sim \frac{\frac{\chi^2_{df=r-1}}{r-1}}{\frac{\chi^2_{df=n_T - r}}{n_T - r}} \sim F(r-1, n_T - r)$$

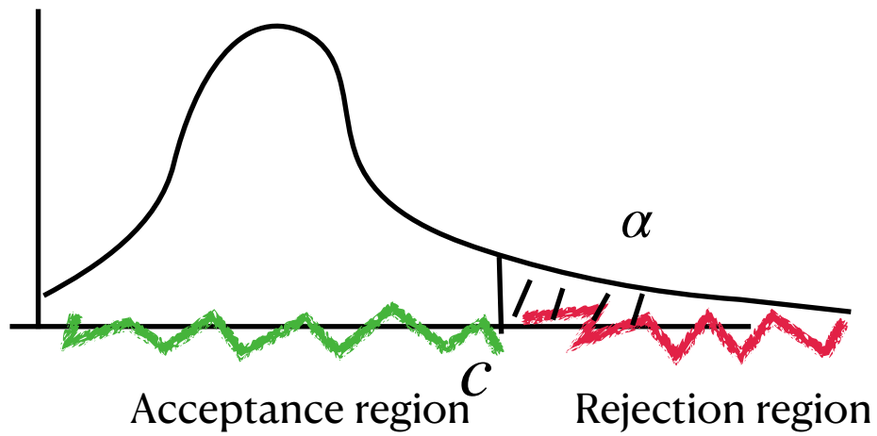


F Test for Equality of Factor Level Means

The probability of making Type I error = $P(H_0$ is true, but we reject $H_0)$

$$= P(F(r-1, n_T - r) \geq c)$$

We want to control the risk of Type I error to be α



Critical value $c = F_{1-\alpha}(r-1, n_T - r)$

$(1-\alpha)100$ percentile of the F distribution

Decision rule:

If $F^* \leq F_{1-\alpha}(r-1, n_T - r)$, then conclude H_0

If $F^* > F_{1-\alpha}(r-1, n_T - r)$, then conclude H_a

Example

Is the mean number of days required for successful rehabilitation the same for the three fitness groups? Control the significance level at .01

To test whether or not the factor level means are the same:

$$H_0 : \mu_1 = u_2 = \ldots = \mu_3$$

H_a : not all μ_i are equal

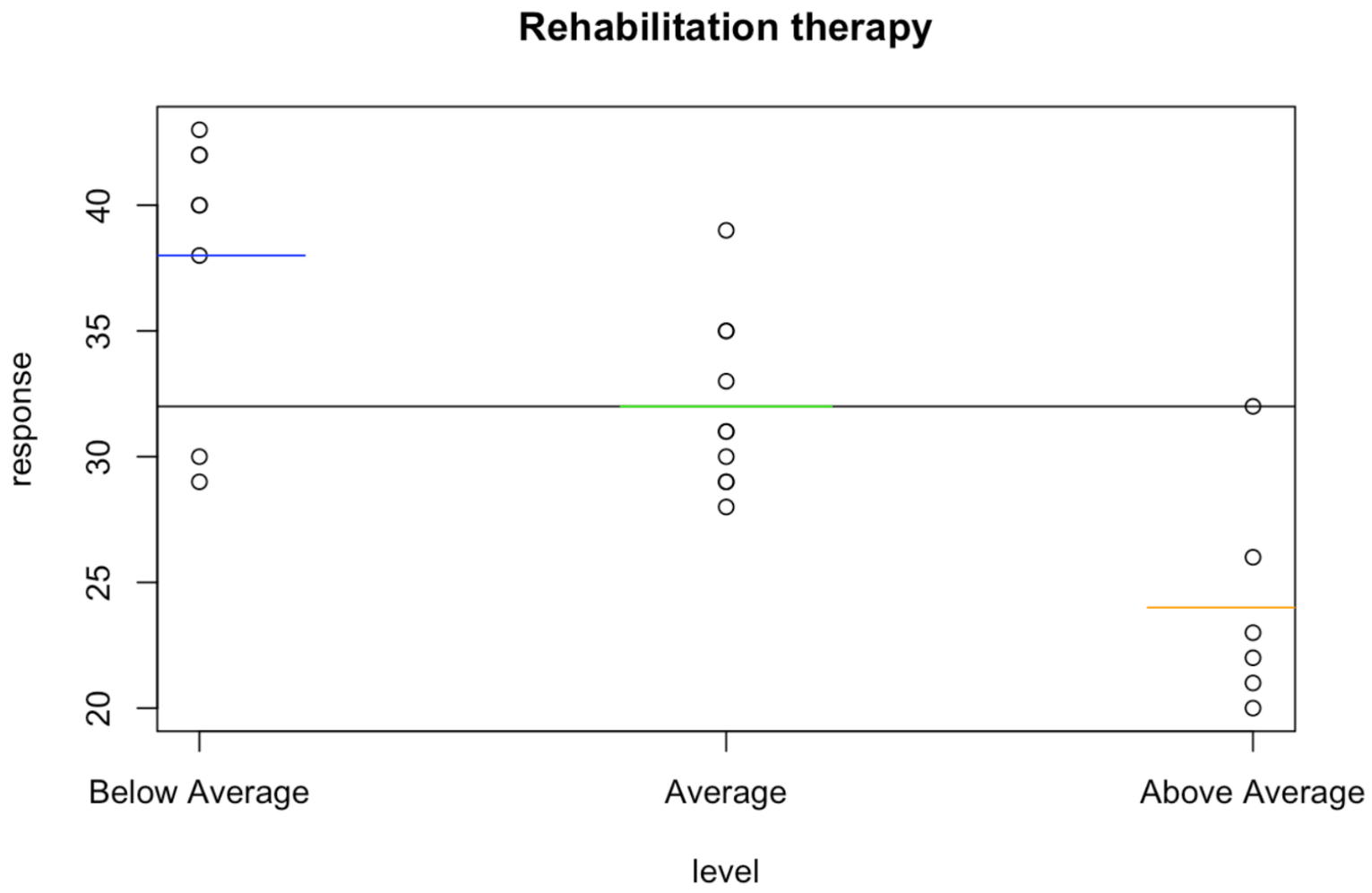
$$F^* = \frac{MSTR}{MSE} = \frac{336}{19.8} = 16.96$$

$$F\left(0.99; r-1, n_T - r\right) = 5.78$$

Since $F^* > F\left(0.99; r-1, n_T - r\right)$, we reject H_0 and conclude that, the mean number of days required for successful rehabilitation is different for the three fitness groups, which suggests that there exists some relation, further detailed analysis of the nature of the relation is required.

Example

What appears to be the nature of the relationship between physical fitness status and time required for physical therapy?



The sample means of days required for successful rehabilitation is the longest when prior fitness is below average, and it's the shortest when prior fitness is above average.

It suggests that the better prior fitness level is, the quicker the rehabilitation can be.

Single-Factor Studies

2 Single-Factor ANOVA Model

Analysis of Variance 4

F Test for Equality of Factor Level Means

Analysis of Factor Level Means

2 Planning of Sample Size

Substantive Research Questions of Interest:

Whether the factor levels or treatments differ in terms of response?
If the factor levels differ in terms of response, in what way do they differ or how do they differ?

These research questions lead to statistical questions usually performed in two steps, correspondingly.

Statistical Questions of Interest:

F Test for Equality of Factor Level Means

Whether the factor level means $\mu_i's$ are all equal or not?

To test whether: $\mu_1 = \mu_2 = \ldots = \mu_r$

If factor level means $\mu_i's$ are equal

No relation between the factor and the response variable is present.

No further analysis is needed

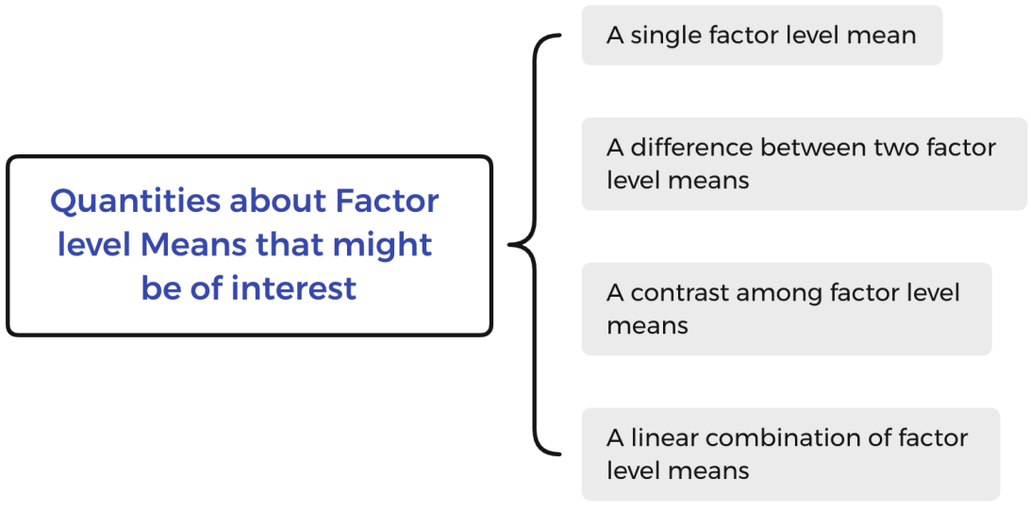
If factor level means $\mu_i's$ differ

A relation between the factor and the response variable is present.

Thorough analysis of the nature of the factor level means, how do they differ?

Analysis of Factor Level Means

Quantities about Factor Level Means



μ_1 :

Estimate the average number of days required for below average fitness patients?
Is the average number of days required for below average fitness patients > 40 days?

$\mu_1 - \mu_3$:

Estimate the difference in average number of days required between below average fitness and above fitness patients?
Is the difference > 30 days?

$\mu_1 - \frac{\mu_2 + u_3}{2}$:

Suppose we consider average and above average fitness patients as “ideal patients”, while below average fitness patients as “not ideal”.
Estimate the difference in average number of days required between the “not ideal” and “ideal” patients?
Is the difference > 10 days?

$0.3\mu_1 + 0.4\mu_2 + 0.3\mu_3$

Suppose we know the population composed of 30%, 40%, 30% of below average, average and above average patients.
Estimate the average number of days required for a randomly selected patient from the population?
Is it < 30 days?

Inference (Estimation & Testing) of a Single Quantity

A single factor level mean μ_i :

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

CLT : $\overline{Y}_{i\cdot} \sim N(\mu_i, \frac{\sigma^2}{n_i})$ (unknown)

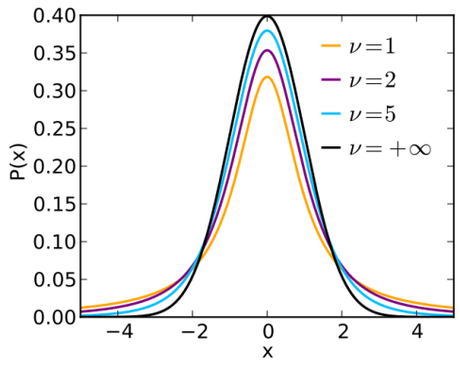
Sampling variance of $\overline{Y}_{i\cdot}$: $S^2(\overline{Y}_{i\cdot}) = \frac{MSE}{n_i}$

$$\frac{\overline{Y}_{i\cdot} - \mu_i}{S(\overline{Y}_{i\cdot})} = \frac{(\overline{Y}_{i\cdot} - \mu_i)/\sigma/\sqrt{n_i}}{\sqrt{MSE/n_i}/\sigma/\sqrt{n_i}} = \frac{(\overline{Y}_{i\cdot} - \mu_i)/\sigma/\sqrt{n_i}}{\sqrt{\frac{MSE}{\sigma^2}}}$$

- $\overline{Y}_{i\cdot} \perp\!\!\!\perp MSE$
- $(\overline{Y}_{i\cdot} - \mu_i)/\sigma/\sqrt{n_i} \sim N(0,1)$
- $(n_T - r)\frac{MSE}{\sigma^2} \sim \chi^2_{n_T - r}$

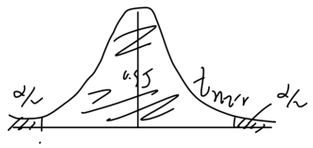
General t-distn with $df = r$ is : $\frac{Z}{\sqrt{\frac{\chi^2_r}{r}}} \sim t_{df=r}$

$$\Rightarrow \frac{\overline{Y}_{i\cdot} - \mu_i}{S(\overline{Y}_{i\cdot})} \sim t_{n_T - r}$$



Inference (Estimation & Testing) of a Single Quantity

2. $(1-\alpha)100\,\%$ confidence interval:

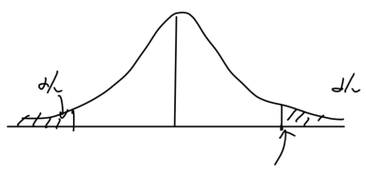


$$\rightarrow \quad \hat{\mu}_i \pm t\left(1-\frac{\alpha}{2}; n_T - r\right) s\left(\hat{\mu}_i\right)$$

3. Hypothesis testing (t-test)

$H_0 : \mu_i = c \;\; H_a : \mu_i \neq c$

$t^* = \dfrac{\hat{\mu}_i - c}{s(\hat{\mu}_i)} \sim t_{n_T - r}$ if H_0 is true



If $|t^*| \leq t(1-\alpha/2; n_T - r)$, conclude H_0

If $|t^*| > t(1-\alpha/2; n_T - r)$, conclude H_a

Inference (Estimation & Testing) of a Single Quantity

Difference between two factor level means $D = \mu_i - \mu_j$: (a pairwise comparison)

1. Point estimate: $\hat{D} = \bar{Y}_{i\cdot} - \bar{Y}_{j\cdot}$

CLT : $\bar{Y}_{i\cdot} \sim N(\mu_i, \frac{\sigma^2}{n_i})$

$\bar{Y}_{i'\cdot} \sim N(\mu_{i'}, \frac{\sigma^2}{n_{i'}})$

$\bar{Y}_{i\cdot} \perp\!\!\!\perp \bar{Y}_{i'\cdot} \Rightarrow \hat{D} \sim N(\mu_i - \mu_{i'} = D, \sigma^2(\frac{1}{n_i} + \frac{1}{n_{i'}}))$

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

$$\frac{\hat{D} - D}{s(\hat{D})} = \frac{(\hat{D} - D) / \sigma\sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}}}{\sqrt{MSE(\frac{1}{n_i} + \frac{1}{n_{i'}}) / \sigma\sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}}}}$$

$$= \frac{(\hat{D} - D) / \sigma\sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}}}{\sqrt{MSE/\sigma^2}}$$

- $\hat{D} \perp\!\!\!\perp MSE$
- $(\hat{D} - D) / \sigma\sqrt{\frac{1}{n_i} + \frac{1}{n_{i'}}} \sim N(0,1)$
- $(n_T - r)\frac{MSE}{\sigma^2} \sim \chi^2_{n_T - r}$

$$\Rightarrow \frac{\hat{D} - D}{s(\hat{D})} \overset{d}{=} \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n_T - r}}{n_T - r}}} \sim t_{df = n_T - r}$$

Inference (Estimation & Testing) of a Single Quantity

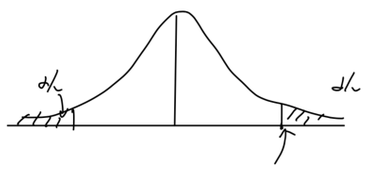
2. $(1-\alpha)100\,\%$ confidence interval:

$$\hat{D} \pm t\left(1-\frac{\alpha}{2}; n_T - r\right) s\left(\hat{D}\right)$$

3. Hypothesis testing (t-test)

$H_0 : D = 0\ H_a : D \neq 0$

$t^* = \dfrac{\hat{D}}{s(\hat{D})} \sim t_{n_T - r}$ if H_0 is true



If $|t^*| \leq t(1-\alpha/2; n_T - r)$, conclude H_0

If $|t^*| > t(1-\alpha/2; n_T - r)$, conclude H_a

Inference (Estimation & Testing) of a Single Quantity

A contrast of several factor level means $L = \sum_{i=1}^{r} c_i \mu_i$ where $\sum_{i=1}^{r} c_i = 0$

1. Point estimate: $\hat{L} = \sum_{i=1}^{r} c_i \bar{Y}_{i\cdot}$

CLT : $\bar{Y}_{i\cdot} \sim N(\mu_i, \frac{\sigma^2}{n_i})$

$\bar{Y}_{i\cdot}$'s indep

$\sigma^2(\hat{L}) = \sigma^2(c_1\bar{Y}_{1\cdot}) + \cdots + \sigma^2(c_r\bar{Y}_{r\cdot})$

$= \sum_{i=1}^{r} c_i^2 \frac{\sigma^2}{n_i} = \sigma^2 \sum \frac{c_i^2}{n_i}$

$\Rightarrow \hat{L} \sim N(L, \sigma^2 \sum \frac{c_i^2}{n_i})$

$s^2(\hat{L}) = MSE(\sum \frac{c_i^2}{n_i})$

$$\frac{\hat{L} - L}{s(\hat{L})} = \frac{(\hat{L} - L) / \sigma\sqrt{\sum \frac{c_i^2}{n_i}}}{\sqrt{MSE(\sum \frac{c_i^2}{n_i})} / \sigma\sqrt{\sum \frac{c_i^2}{n_i}}} = \frac{(\hat{L} - L) / \sigma\sqrt{\sum \frac{c_i^2}{n_i}}}{\sqrt{MSE/\sigma^2}}$$

- $\hat{L} \perp MSE$
- $(\hat{L} - L) / \sigma\sqrt{\sum \frac{c_i^2}{n_i}} \sim N(0,1)$
- $(n_T - r)\frac{MSE}{\sigma^2} \sim \chi^2_{n_T - r}$

$$\Rightarrow \frac{\hat{L} - L}{s(\hat{L})} \overset{d}{=} \frac{N(0,1)}{\sqrt{\frac{\chi^2_{n_T - r}}{n_T - r}}} \sim t_{df = n_T - r}$$

Inference (Estimation & Testing) of a Single Quantity

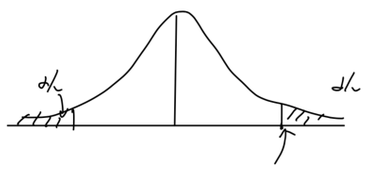
2. $(1-\alpha)100\,\%$ confidence interval:

$$\hat{L} \pm t\left(1-\frac{\alpha}{2}; n_T - r\right) s\left(\hat{L}\right)$$

3. Hypothesis testing (t-test)

$H_0 : L = 0\ H_a : L \neq 0$

$t^* = \dfrac{\hat{L}}{s(\hat{L})} \sim t_{n_T - r}$ if H_0 is true



If $|t^*| \leq t(1-\alpha/2; n_T - r)$, conclude H_0

If $|t^*| > t(1-\alpha/2; n_T - r)$, conclude H_a

Inference (Estimation & Testing) of a Single Quantity

A linear combination of several level means $L = \sum_{i=1}^{r} c_i \mu_i$

- Example of linear combination?

  Overall mean $\bar{\mu}_{\cdot} = \sum_{i=1}^{r} \frac{n_i}{n_T} \mu_i$

- All previous quantities are special cases of “linear combination”

- Same inference as “contrast”

Example

The researcher is interested in estimating the mean number of days required for patients with average physical fitness, report the 99 percent confidence interval and give an interpretation.

$$\overline{Y_{2\cdot}} \pm t\left(0.995; n_T - r\right) s\left(\overline{Y_{2\cdot}}\right) = \overline{Y_{2\cdot}} \pm t\left(0.995; n_T - r\right) \sqrt{\frac{MSE}{n_2}}$$

$$\bar{Y}_2 \cdot = 32, t\left(0.995; n_T - r\right) = 2.83, MSE = 19.8, n_2 = 10$$

→ [28.01,35.99]

Meaning of a confidence interval:

if repeated the same experiment many times, and each time we construct a confidence interval as above, then we would expect that 99% of times, the confidence intervals contracted this way will include the true average number of days required for patients with average fitness.

Interpret in the context:

the mean number of days required for patients with average fitness is estimated to be somewhere between 28.01 and 35.99 days, with 99% confidence.

Simultaneous Inference Procedures for Multiple Quantities about Factor Level Means

we aim to conclude something like:

" The below average fitness group has the longer time required for rehabilitation than the average fitness group,
The average fitness group has the longer time required for rehabilitation than the above average fitness group"

That is equivalent to test the following comparisons:

$$\mu_1 > \mu_2$$
$$\mu_2 > \mu_3$$
$$\mu_1 > \mu_3$$

It is tempting to just conduct 3 separate tests as above section, then piece them together:

$$\begin{cases} H_0^1 : \mu_1 = \mu_2 \text{ vs } H_a^1 \mu_1 > \mu_2 \\ H_0^2 : \mu_2 = \mu_3 \text{ vs } H_a^2 \mu_2 > \mu_3 \\ H_0^3 : \mu_1 = \mu_3 \text{ vs } H_a^3 \mu_1 > \mu_3 \end{cases}$$

Q: What might go wrong?

Simultaneous Inference Procedures for Multiple Quantities about Factor Level Means

What might go wrong?

What is considered “mistake” is different
when we make a composite statement that includes results of multiple tests or multiple C.I.s:

> In a composite statement, if one sub-hypothesis test has a wrong conclusion, then it makes the whole statement wrong.
> That is, we have much more ways to make mistakes
>
> The confidence level of C.I.s and significance level of tests apply only to each quantity considered individually,
> it will fail to compensate for multiple comparisons when it’s desired to have a statement
> which include multiple tests or multiple C.I.s.
>
> E.g. Suppose we consider the efficacy of a drug in terms of reduction of any one of the disease symptoms. If we just consider more symptoms, it’s almost guaranteed that the drug will appear to be an improvement in terms of at least one symptom.

Data snooping

Simultaneous Inference Procedures for Multiple Quantities about Factor Level Means

"Family-wise" confidence coefficient $1 - \alpha$:

The proportion of correct families, when repeated sets of samples are selected and simultaneous tests or C.I.s are calculated each time.

> A family including simultaneous inference for multiple quantities is considered correct, if everyone single quantity inference is correct.

Thus, a family-wise confidence coefficient indicates that all conclusions in this family will be correct in $(1 - \alpha)100$ percent of repetitions.

Multiple Comparison Procedure: Bonferroni

Suppose we're interested in making inferenece about multiple quantities,
that are linear combinations of factor level means,
i.e., a family containing g linear combinations of factor level means

$$\mathscr{L} = \{L_1 = \sum_{i=1}^{r} c_{1i}\mu_i, \ldots, L_g = \sum_{i=1}^{r} c_{gi}\mu_i\}$$

family-wise error rate = P(Making Type I error for any of the hypothesis in the family)

$= P(H_0^i$ falsely rejected for some $i)$

$= p(\cup_{i=1}^{g} \{H_0^i$ falsely rejected $\})$

$\leq \sum_{i=1}^{g} P(H_0^i$ falsely rejected $)$

If we control individual test at significance level α_0
$= g\alpha_0$

Bonferroni's idea:
One very easy and conservative way to control family-wise error rate at α is to control individual test's significance level at $\alpha_0 = \frac{\alpha}{g}$

Multiple Comparison Procedure: Bonferroni

$(1-\alpha)100\,\%$ confidence interval for individual quantity in this family:

$$\hat{L}_i \pm Bs\left(\hat{L}_i\right) \text{ for } i = 1 \ldots g$$

$$B = t\left(1 - \frac{\alpha}{2g}; n_T - r\right)$$

Guarantee:
family-wise confidence coefficient is at least $(1-\alpha)100\,\%$

Meaning:
in at least $(1-\alpha)100\,\%$ of repetition of experiments, all the intervals in the family cover the true corresponding $L_i's$ $\alpha\,\%$ of repetition of experiments, some tests in the family made false discovery when the null hypothesis was true.

Hypothesis testing (t-test) for individual quantity in this family:

$$H_0^i : L_i = 0 \; H_a^i : L_i \neq 0$$

$$t^* = \frac{\hat{L}_i}{s(\hat{L}_i)} \sim t_{n_T - r} \text{if } H_0 \text{ is true}$$

If $|t^*| \leq B$, conclude H_0

If $|t^*| > B$, conclude H_a

Guarantee:
family-wise Type I error is at most α

Meaning:
in at most $\alpha\,\%$ of repetition of experiments, some tests in the family made false discovery when the null hypothesis was true.

Multiple Comparison Procedure: Sheffe

Suppose we're interested in making inference about <u>all possible contracts of factor level means</u>

i.e., a family containing all possible contracts of factor level means

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

Infinitely many claims or quantities

$\forall L \in \mathscr{L}$:

$$\hat{L} - L = \sum_{i=1}^{r} c_i (\bar{Y}_{i\cdot} - \mu_{i\cdot}) = \sum_{i=1}^{r} c_i \left[(\bar{Y}_{i\cdot} - \mu_i) + (\bar{Y}_{\cdot\cdot} - \mu) \right]$$

$$= \sum_{i=1}^{r} \frac{c_i}{\sqrt{n_i}} \sqrt{n_i} \left[(\bar{Y}_{i\cdot} - \mu_i) + (\bar{Y}_{\cdot\cdot} - \mu) \right]$$

$$\leq \sqrt{\sum \frac{c_i^2}{n_i}} \sqrt{\sum n_i \left[(\bar{Y}_{i\cdot} - \mu_i) + (\bar{Y}_{\cdot\cdot} - \mu) \right]^2}$$

$$\overset{\text{let}}{=} \sum n_i (\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$$

$$= \sqrt{SSTR}.$$

$$S(\hat{L}) = S\left(\sum c_i \bar{Y}_{i\cdot}\right) = \sqrt{\sum c_i^2 S^2(\bar{Y}_{i\cdot})}$$

$$= \sqrt{\sum c_i^2 \frac{MSE}{n_i}}$$

$$= \sqrt{MSE} \sqrt{\sum \frac{c_i^2}{n_i}}$$

$$\Rightarrow \frac{\hat{L} - L}{S(\hat{L})} \leq \sqrt{\frac{SSTR}{MSE}} = \sqrt{(r-1) \frac{MSTR}{MSE}}$$

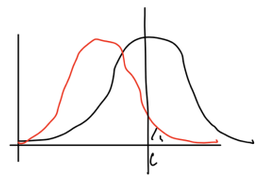
Under H_0,

$$\frac{r-1}{\sigma^2} MSTR \sim \chi^2_{r-1}$$

$$\frac{n_T - r}{\sigma^2} MSE \sim \chi^2_{n_T - r}$$

$$= \sqrt{\frac{\sigma^2 \chi^2_{r-1} / r-1}{\sigma^2 \chi^2_{n_T-r} / n_T - r}} \sqrt{r-1}$$

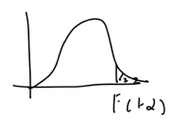
$$\overset{d}{=} \sqrt{(r-1) F_{(r-1, n_T-1)}}$$



$$\Rightarrow P\left(\left| \frac{\hat{L} - L}{S(\hat{L})} \right| \geq c \right) \quad \text{for all } L$$

$$\leq P\left(\sqrt{\frac{MSTR}{MSE}} \geq c \right)$$

$\frac{MSTR}{MSE} \geq \frac{c^2}{r-1}$



$$\frac{c^2}{r-1} = F(1-\alpha)$$

$$\Rightarrow c = \sqrt{(r-1) F(1-\alpha)}$$

$$\Rightarrow \text{critical value } S = \sqrt{(r-1) F(1-\alpha; r-1, n_T - r)}$$

Multiple Comparison Procedure: Sheffe

$(1-\alpha)100\,\%$ confidence interval for individual quantity in this family:

$$\hat{L}_i \pm Ss\left(\hat{L}_i\right)$$

$$S = \sqrt{(r-1)F(1-\alpha; r-1, n_T - r)}$$

Guarantee:
family-wise confidence coefficient is at least $(1-\alpha)100\,\%$

Meaning:
in at least $(1-\alpha)100\,\%$ of repetition of experiments, all the intervals in the family cover the true corresponding $L_i's$ $\alpha\,\%$ of repetition of experiments, some tests in the family made false discovery when the null hypothesis was true.

Hypothesis testing (t-test) for individual quantity in this family:

$$H_0^i : L_i = 0 \; H_a^i : L_i \neq 0$$

$$t^* = \frac{\hat{L}_i}{s(\hat{L}_i)}$$

If $|t^*| \leq S$, conclude H_0

If $|t^*| > S$, conclude H_a

Guarantee:
family-wise Type I error is at most α

Meaning:
in at most $\alpha\,\%$ of repetition of experiments, some tests in the family made false discovery when the null hypothesis was true.

Multiple Comparison Procedure: Tukey

Suppose we're interested in making inference about all pairwise comparisons of factor level means
i.e., a family containing <u>all pairwise comparisons of factor level means</u>

$$\mathscr{L} = \{D_{ii'} = \mu_i - \mu_{i'} \text{ for } i \neq i'\}$$

$\frac{r(r-1)}{2}$ Pairwise comparisons

$\bar{Y}_{i\cdot} - \mu_i \overset{iid}{\sim} N(0, \frac{\sigma^2}{n})$

MSE is unbiased estimator of σ^2. $\frac{MSE}{\sigma^2} \sim \chi^2_{n_T - r}$

Studentized range $= \frac{\text{range of data}}{\text{estimated s.d.}} = \frac{W}{S}$

$$= \frac{\max\{\bar{Y}_{i\cdot} - \mu_i\} - \min\{\bar{Y}_{i\cdot} - \mu_i\}}{\sqrt{\frac{MSE}{n}}}$$

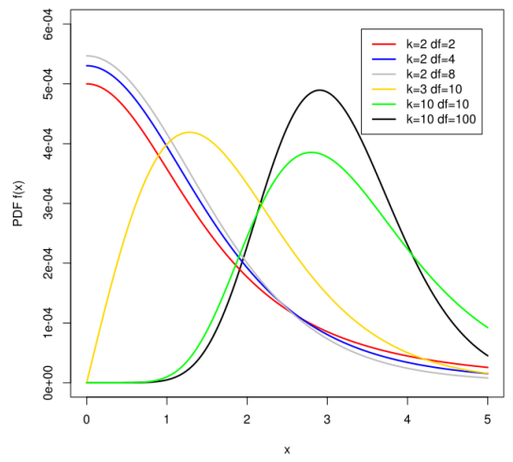
$\sim q(r, n_T - r)$

Studentized range distribution

$$|\hat{D}_{ii'} - D_{ii'}| = |(\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot}) - (\mu_i - \mu_{i'})|$$
$$= |(\bar{Y}_{i\cdot} - \mu_i) - (\bar{Y}_{i'\cdot} - \mu_{i'})|$$
$$\leq \max\{\quad\} - \min\{\quad\}$$

$$S(\hat{D}_{ii'}) = S(\bar{Y}_{i\cdot} - \bar{Y}_{i'\cdot})$$
$$= \sqrt{S^2(\bar{Y}_{i\cdot}) + S^2(\bar{Y}_{i'\cdot})}$$
$$= \sqrt{MSE(\tfrac{1}{n_i} + \tfrac{1}{n_{i'}})}$$

$$\Rightarrow \frac{|\hat{D}_{ii'} - D_{ii'}|}{S(\hat{D}_{ii'})} \leq \frac{1}{\sqrt{2}} \text{ Studentized range for all } i \neq i'$$



Multiple Comparison Procedure: Tukey

$(1-\alpha)100\,\%$ confidence interval for individual quantity in this family:

$$\hat{D}_{ii'} \pm Ts\left(\hat{D}_{ii'}\right)$$

$$T = \frac{1}{\sqrt{2}} q(1-\alpha; r, n_T - r)$$

Guarantee:
family-wise confidence coefficient is at least $(1-\alpha)100\,\%$

Meaning:
in at least $(1-\alpha)100\,\%$ of repetition of experiments, all the intervals in the family cover the true corresponding $L_i's$ $\alpha\,\%$ of repetition of experiments, some tests in the family made false discovery when the null hypothesis was true.

Hypothesis testing (t-test) for individual quantity in this family:

$$H_0^i : D_{ii'} = 0 \; H_a^i : D_{ii'} = \neq 0$$

$$q^* = \frac{\hat{D}_{ii'}}{s(\hat{D}_{ii'})}$$

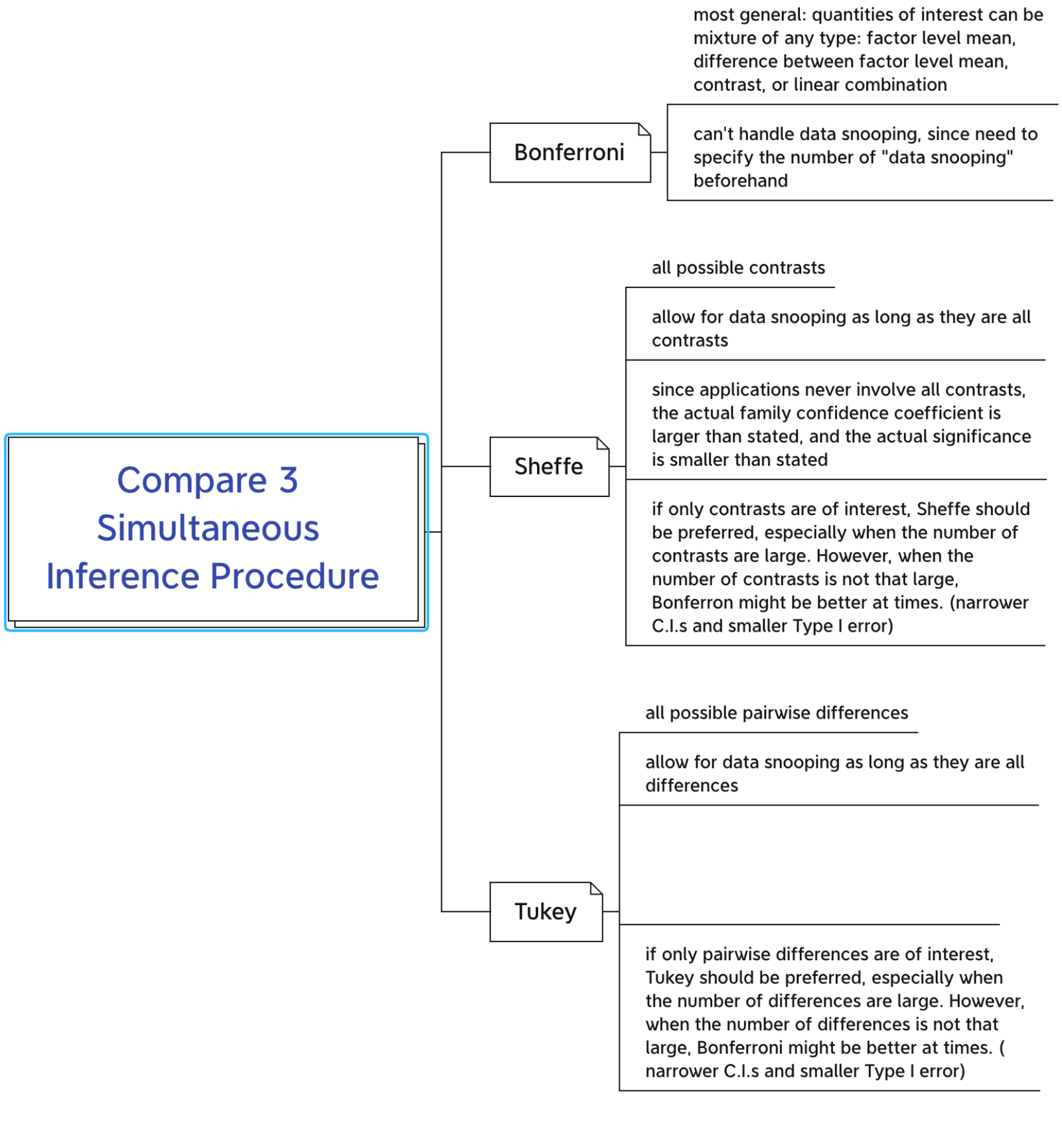
If $|q^*| \leq T$, conclude H_0

If $|q^*| > T$, conclude H_a

Guarantee:
family-wise Type I error is at most α

Meaning:
in at most α % of repetition of experiments, some tests in the family made false discovery when the null hypothesis was true.

Compare: Bonferroni vs Scheffe vs Tukey



Compare: Bonferroni vs Scheffe vs Tukey

All three procedures are of the form "estimator ± multiplier ×SE."

The only difference among the three procedures is the multiplier.

In any given problem, one may compute the Bonferroni multiple as well as the Scheffé multiple and, when appropriate, the Tukey multiple, and select the one that is smallest.

Smaller multiple means: more efficient
- narrower C.I.s —> more precise estimate
- Smaller actual Type I error, larger power to detect true difference in hypothesis testing

Example

Suppose before seeing the data (a priori), the researcher wants to estimate the confidence intervals for

$$D_1 = \mu_2 - \mu_3, D_2 = \mu_1 - \mu_2$$

With a 95 percent family confidence coefficient.

Since the quantities of interest are pairwise comparisons (and contrasts of course), all three methods apply.

$$\hat{D}_1 = \overline{Y_{2.}} - \overline{Y_{3.}} = 8 \qquad \hat{D}_2 = \bar{Y}_{1.} - \bar{Y}_{2.} = 6$$

$$s\left(\hat{D}_1\right) = \sqrt{\text{MSE}\left(\frac{1}{n_2} + \frac{1}{n_3}\right)} = 2.298378 \qquad s\left(\hat{D}_2\right) = \sqrt{\text{MSE}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} = 2.111195$$

Bonferroni:

$B = t(1 - \frac{\alpha}{2\times 2}; n_T - r) = 2.41384$

$$D_1 : \hat{D_1} \pm Bs(\hat{D_1}) = 8 \pm 2.41384 \times 2.298378 = [2.452083, 13.54792]$$

$$D_2 : \hat{D_2} \pm Bs(\hat{D_2}) = 6 \pm 2.41384 \times 2.111195 = [0.9039131, 11.09609]$$

Sheffe: $\sqrt{(r-1)F(1-\alpha; r-1, n_T - r)} = 2.633173$

$$D_1 : \hat{D_1} \pm Ss(\hat{D_1}) = 8 \pm 2.63 \times 2.298378 = [1.955266, 4.04473]$$

$$D_2 : \hat{D_2} \pm Ss(\hat{D_2}) = 6 \pm 2.63 \times 2.111195 = [0.4475572, 11.55244]$$

Tukey: $T = \frac{1}{\sqrt{2}} q(1 - \alpha; r, n_T - r) = 2.52$

$$D_1 : \hat{D_1} \pm Ts(\hat{D_1}) = 8 \pm 2.52 \times 2.298378 = [2.2, 13.79]$$

$$D_2 : \hat{D_2} \pm Ts(\hat{D_2}) = 6 \pm 2.52 \times 2.111195 = [0.6797886, 11.32021]$$

Since the Bonferroni Multiple is the smallest, Bonferroni is the most efficient, it leads to tightest confidence intervals at the same level of family confidence coefficient.

Example

If the researcher also wants to estimate $D_3 = \mu_1 - \mu_3$, still with a 95 percent family confidence coefficient.
Would the multiples change?
What procedure(s) is appropriate for “data snooping” if the researcher wants to look at many such comparisons not a priori determined?

The Bonferroni multiple will change, but Sheffe and Tukey won't.

Tukey and Sheffe are applicable for data snooping as they are developed to include all pairwise comparisons and all contrasts, correspondingly.

But Bonferroni is not open for data snooping.

Example

The researcher is interested in comparing all pairs of factor level means, test for all pairs of factor level means whether or not they differ.

Use the procedure that is most efficient with $\alpha = 0.05$.
Set up groups by whether or not their factor levels means differ.

There are 3 pairwise comparisons.

$$B = t\left(1 - \frac{\alpha}{2 \times 3}; n_T - r\right) = 2.60135$$

$$S = \sqrt{(r-1)F\left(1-\alpha; r-1, n_T - r\right)} = 2.633173$$

$$T = \frac{1}{\sqrt{2}} q\left(1-\alpha; r, n_T - r\right) = 2.52$$

Therefore Tukey's procedure is most efficient.

$$\begin{cases} H_0^1 : \mu_1 = \mu_2 \text{ vs } H_a^1 : \mu_1 \neq \mu_2 \\ H_0^2 : \mu_2 = \mu_3 \text{ vs } H_a^2 : \mu_2 \neq \mu_3 \\ H_0^3 : \mu_1 = \mu_3 \text{ vs } H_a^3 : \mu_1 \neq \mu_3 \end{cases}$$

Compare with critical value T, we reject all null hypotheses.

We conclude that, the average number of days required for different physical fitness group are all different from each other, with familywise confidence coefficient of 0.95.

We can set up groups within which factor level means do not differ:

Group 1: Below Average
Group 2: Average
Group 3: Above Average

Example

If before seeing the data, the researcher wants to estimate the difference in difference of mean days required between adjacent factor levels with a 99 percent confidence interval.

That is, to estimate the contrast $L = (\mu_1 - \mu_2) - (\mu_2 - \mu_3)$

Interpret the interval estimate.

This is a contrast with $c_1 = 1, c_2 = -2, c_3 = 1$

$\hat{L} = (\bar{Y}_1 - \bar{Y}_2.) - (\bar{Y}_2. - \bar{Y}_3.) = -2$

$$s(\hat{L}) = \sqrt{\text{MSE}\left(\sum \frac{c_i^2}{n_i}\right)}$$

A 99 percent confidence interval is

$$\hat{L} \pm t\left(1 - \frac{\alpha}{2}; n_T - r\right) s(\hat{L}) = [-12.48, 8.48]$$

Example

Estimate the following comparisons using all appropriate procedures with a 95 percent family confidence coefficient:

$$D_1 = \mu_1 - \mu_2, D_2 = \mu_1 - \mu_3, D_3 = \mu_2 - \mu_3, L_1 = D_1 - D_3,$$

Which procedure is more efficient?
Interpret results and describe findings using the one that is most efficient.

Both Bonferroni and Sheffe are appropriate.

$$B = t\left(1 - \frac{\alpha}{2 \times 4}; n_T - r\right) = 2.731632$$

$$S = \sqrt{(r-1)F\left(1-\alpha; r-1, n_T - r\right)} = 2.633173$$

The Sheffe multiple is smaller, therefore Sheffe is most efficient.

$$D_1 : \hat{D}_1 \pm Ss\left(\hat{D}_1\right) = 6 \pm 2.63 \times 2.1112 = [0.45, 11.55]$$

$$D_2 : \hat{D}_2 \pm S_s\left(\hat{D}_2\right) = 14 \pm 2.63 \times 2.4037 = [7.68, 20.32]$$

$$D_3 : \hat{D}_3 \pm S_s\left(\hat{D}_3\right) = 8 \pm 2.63 \times 2.2984 = [1.96, 14.04]$$

$$L_1 : \hat{L}_1 \pm Ss\left(\hat{L}_1\right) = -2 \pm 2.63 \times 3.7 = [-11.73, 7.73]$$

Since the intervals for D_1, D_2, D_3 do not include zero, while the one for L does, we can conclude the statement that:

> the average days required for different physical fitness are all different with each other, and the difference between "below average" and "average" is the same as the difference between "average" and "above average".

Single-Factor Studies

2 Single-Factor ANOVA Model

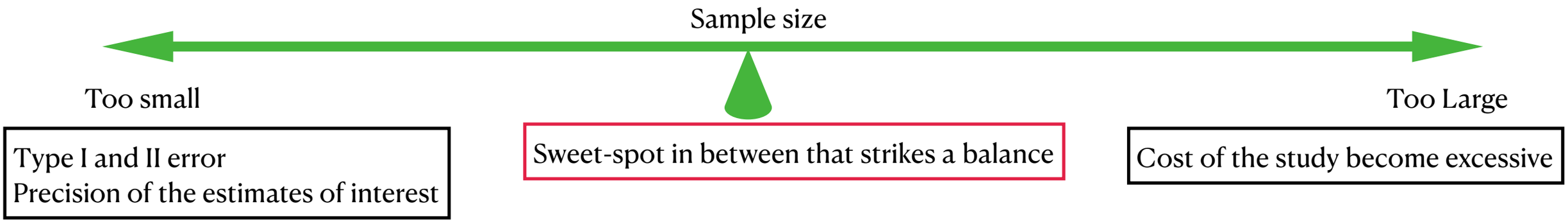
Analysis of Variance 4

F Test for Equality of Factor Level Means

Analysis of Factor Level Means 6

Planning of Sample Size

Planning of Sample Size



Power approach

Control risks of making Type I and II errors for the overall F test

$H_0 : \mu_1 = \ldots = \mu_r$

Estimation approach

Control the precision (width of C.I.) of the estimates of interest,

Specify the major comparisons of interest and determine the expected widths of C.I.s which decreases as sample size increases

Review: F Test for Equality of Factor Level Means

To test whether or not the factor level means are the same:

$H_0 : \mu_1 = u_2 = \ldots = \mu_r$

H_a : not all μ_r are equal

Test statistic: $F^* = \dfrac{MSTR}{MSE}$

Large value of F^* support H_a

Small value, when $F^* \approx 1$ support H_0

—> We reject H_0 for large value of F^*, i.e. $F^* \geq c$

How do we decide what is “large” and what is “small”?

		Decision	
		H_0	H_a
Truth	H_0	✓	Type I error
	H_a	Type II error	✓

Review: F Test for Equality of Factor Level Means

We want to control type I error at significance level α (usually .05):

If H_0 is true: $\mu_1 = \ldots = \mu_r$

$$\frac{n_T - r}{\sigma^2} MSE = \frac{SSE}{\sigma^2} \sim \chi^2_{n_T - r}$$

$$\frac{r-1}{\sigma^2} MSTR = \frac{SSTR}{\sigma^2} \sim \chi^2_{r-1} \left(\frac{1}{\sigma^2} \sum_{i=1}^{r} n_i \left(\mu_i - \overline{\mu}_. \right)^2 \right) = \chi^2_{r-1}$$

MSE and MSTR are independent random variables

Re-write the test statistic:

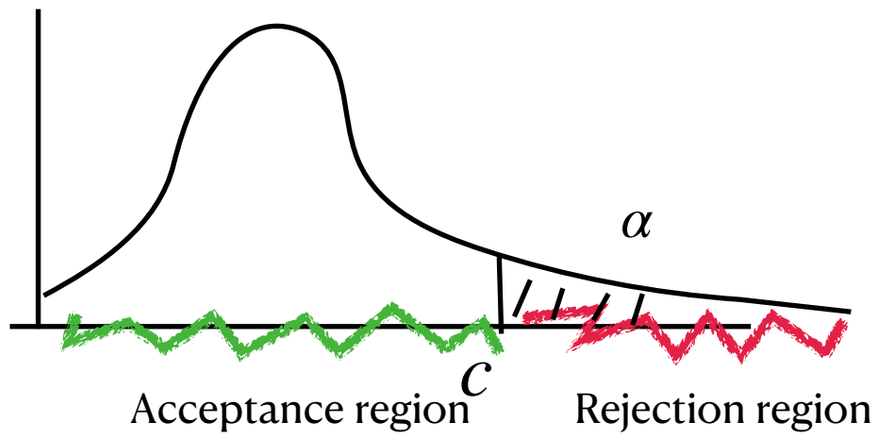
$$\rightarrow \quad F^* = \frac{MSTR}{MSE} = \frac{\frac{(n-1)MSTR}{\sigma^2}}{\frac{(n_T - r)MSE}{\sigma^2}} \sim \frac{\frac{\chi^2_{df=r-1}}{r-1}}{\frac{\chi^2_{df=n_T-r}}{n_T - r}} \sim F(n_T - r, r - 1)$$

Review: F Test for Equality of Factor Level Means

The probability of making Type I error = $P(H_0$ is true, but we reject $H_0)$

$$= P(F(n_T - r, r - 1) \geq c)$$

We want to control the risk of Type I error to be smaller than α



Critical value $c = F_{1-\alpha}(n_t - r, r - 1)$

$(1 - \alpha)100$ percentile of the F distribution

Decision rule:

If $F^* \leq F_{1-\alpha}(n_T - r, r - 1)$, then conclude H_0

If $F^* > F_{1-\alpha}(n_T - r, r - 1)$, then conclude H_a

Power Approach

The power of a test is probability that decision rule will lead to H_a (statistically significant result)

$$power = P(\text{ conclude } H_a \text{ when } H_a \text{ is true }) = 1 - P(\text{Type II error}) = 1 - \beta$$

Power Approach

$$power = P(\text{ conclude } H_a \text{ when } H_a \text{ is true })$$

$$= P(F(r-1, n_T - r, \phi) > F(1-\alpha; r-1, n_T - r))$$

Non centrality parameter $\phi = \frac{1}{\sigma^2}\sum_{i=1}^{r} n_i(\mu_i - \overline{\mu}_\cdot)^2$

Consider balanced design, $n_1 = \ldots = n_r = n$

$$\phi = \frac{n}{\sigma^2}\sum_{i=1}^{r}(\mu_i - \overline{\mu}_\cdot)^2$$

How different the true treatment means are!

We don't know the truth about treatment means and error variance, but they are fixed and hidden

How does power change with n?

As n increases —> ϕ increases —> power increases, Type II error decreases

Power Approach

Many single factor studies are undertaken because of reasons and the expectation that factor level means differ,

So we are more likely in situations where there indeed exists differences, our job is to detect the difference.

Type II error, which is we falsely conclude there is no difference when there is difference, is critical for the study, though we still want to control Type I error.

That is, we want high power under the alternative hypothesis:

> "We have x% chance that the decision rule will lead to detection of difference, when the differences truly exist."

Power Approach

To plan sample size:

- $\sum (\mu_i - \mu)^2$: treatment means variation: strength of treatment signal
- σ^2 : error variance: strength of noise

- Direct specification
- Effect size: standardized index measures the strength or how different the treatment means are

Important / meaningful difference:

As sample size n increases, high power of detecting any difference.

Focus on tests that have high power of detecting important and meaningful difference, not waste sample size (money) to detect unimportant difference

Minimum range of treatment means $\Delta = max(\mu_i) - min(\mu_i)$

$$\sum_{i=1}^{r} (\mu_i - \bar{\mu}_\cdot \geq \frac{\Delta^2}{2}$$

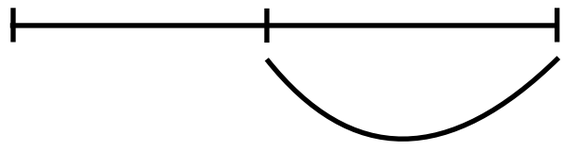
$$\phi = \frac{1}{\sigma^2} \sum_{i=1}^{r} n_i(\mu_i - \overline{\mu}_\cdot)^2 \geq \frac{n\Delta^2}{2\sigma^2} = \frac{n}{2}\left(\frac{\Delta}{\sigma}\right)^2$$

Effect size $\frac{\Delta}{\sigma}$

So, if for specified effect size, we can make sure that the power satisfies minimal requirement, then the true power would be even greater.

Estimation Approach

Precision of an estimate:



Margin of error

Example

Suppose that the sample sizes have not yet been determined but it has been decided to use the same number of patients for each physical fitness group.

Assume that:

- a reasonable planning value for the error standard deviation is $\sigma = 4.5$ days.
- the range of the treatment means is 5.63 days
- the α risk is to be controlled at .01?

What would be the required sample sizes if we want the differences in the mean times for the three physical fitness categories are to detected with probability .80

Hide

```
# search the smallest n value that satisfies above condition
n=2
delta=5.63
sigma=4.5
alpha=0.01
flag=FALSE
while(flag==FALSE){
  temp=pf(qf(1-alpha,2,3*(n-1)),df1=2,df2=3*(n-1),ncp=(n/2)*(delta/sigma)^2,lower.tail = FALSE)
  if(temp>=0.8){flag=TRUE;print(paste("The minimum sample size for each treatment:",n))}
  n=n+1
}
```

```
## [1] "The minimum sample size for each treatment: 20"
```

Example

If the sample sizes were employed, what would be the power of the test for treatment mean differences when $\mu_1 = 37, \mu_2 = 32, \mu_3 = 28$?

$$\phi = \frac{n\sum_{i=1}^{r}\left(\mu_i - \mu_{\cdot}\right)^2}{\sigma^2} = \frac{20\sum_{i=1}^{3}\left(\mu_i - \mu_{\cdot}\right)^2}{4.5^2} = 40.16461$$

$$\text{power} = P(F(2, 3*20-3; 40.16461) > F(0.99; 2, 3*20-3)) = 0.9992062$$

Example

Suppose primary interest is in estimating the two pairwise comparisons:

$$L_1 = \mu_1 - \mu_2 \quad L_2 = \mu_3 - \mu_2$$

What would be the required sample sizes be if the precision of each comparison is to be ± 3.0 days, using the most efficient multiple comparison procedure with a 95 percent family confidence coefficient?

There are 2 pairwise comparisons, Bonferroni is most efficient.

$$\sigma\left(\hat{L}_1\right) = \sqrt{\sigma^2(1/n + 1/n)} \quad \sigma\left(\hat{L}_2\right) = \sqrt{\sigma^2(1/n + 1/n)}$$

Therefore, we want $t\left(1 - \frac{\alpha}{2 \times 2}; n_T - r\right) \times \sqrt{\sigma^2(1/n + 1/n)} \leq 3$

Code

```
## [1] "The minimum sample size for each treatment: 24"
```

Example

Suppose that primary interest is in comparing the below-average and above-average physical fitness groups. respectively. with the average physical fitness group. Thus. two comparisons are of interest:
$L_1 = \mu_1 - \mu_2 \quad L_2 = \mu_3 - \mu_2$

Assume that a reasonable planning value for the error standard deviation is $\sigma = 4.5$days.

If below-average and above-average groups have equal sample sizes n, the average physical fitness group has 2 n, what would be the required sample sizes if the precision of each pairwise comparison is to be ± 2.5 days, using the Bonferroni procedure and a 90 percent family confidence coefficient?

$$\sigma(\hat{L_1}) = \sqrt{\sigma^2(1/n + 1/(2 * n))} \quad \sigma(\hat{L_2}) = \sqrt{\sigma^2(1/n + 1/(2 * n))}$$

Therefore, we want $t(1 - \frac{\alpha}{2\times 2}; 4n - 3) \times \sqrt{\sigma^2(1/n + 1/(2n))} \leq 2.5$

Hide

```
# search the smallest n value that satisfies above condition
n=2
sigma=4.5
alpha=0.1
flag=FALSE
while(flag==FALSE){
  temp=qt(1-alpha/4,4*n-3)*sqrt(sigma^2*(1/n+1/(2*n)))
  if(temp<=2.5){flag=TRUE;print(paste("The minimum sample size n=:",n))}
  n=n+1
}
```

```
## [1] "The minimum sample size n=: 20"
```

Example

if the sample size for the average physical fitness group is to be: (1) n and (2) 3 n, all other specifications remaining the same.

if the sample size for the average physical fitness group is to be:
(1) n

$$\sigma\left(\hat{L}_1\right)=\sqrt{\sigma^2(1/n+1/n)} \quad \sigma\left(\hat{L}_2\right)=\sqrt{\sigma^2(1/n+1/n)}$$

Therefore, we want $t\left(1-\frac{\alpha}{2\times 2};3n-3\right)\times\sqrt{\sigma^2(1/n+1/(n))}\leq 2.5$

Hide

```
# search the smallest n value that satisfies above condition
n=2
sigma=4.5
alpha=0.1
flag=FALSE
while(flag==FALSE){
  temp=qt(1-alpha/4,3*n-3)*sqrt(sigma^2*(1/n+1/(n)))
  if(temp<=2.5){flag=TRUE;print(paste("The minimum sample size n=:",n))}
  n=n+1
}
```

```
## [1] "The minimum sample size n=: 26"
```

Example

if the sample size for the average physical fitness group is to be: (1) n and (2) 3 n, all other specifications remaining the same.

if the sample size for the average physical fitness group is to be:
(1) 3n

$$\sigma\left(\hat{L}_1\right) = \sqrt{\sigma^2(1/n + 1/(3n))} \quad \sigma\left(\hat{L}_2\right) = \sqrt{\sigma^2(1/n + 1/(3n))}$$

Therefore, we want $t\left(1 - \frac{\alpha}{2 \times 2}; 5n - 3\right) \times \sqrt{\sigma^2(1/n + 1/(3n))} \leq 2.5$

Hide

```
# search the smallest n value that satisfies above condition
n=2
sigma=4.5
alpha=0.1
flag=FALSE
while(flag==FALSE){
  temp=qt(1-alpha/4,5*n-3)*sqrt(sigma^2*(1/n+1/(3*n)))
  if(temp<=2.5){flag=TRUE;print(paste("The minimum sample size n=:",n))}
  n=n+1
}
```

```
## [1] "The minimum sample size n=: 18"
```

Example

Compare previous results, which design leads to the smallest total sample size here?

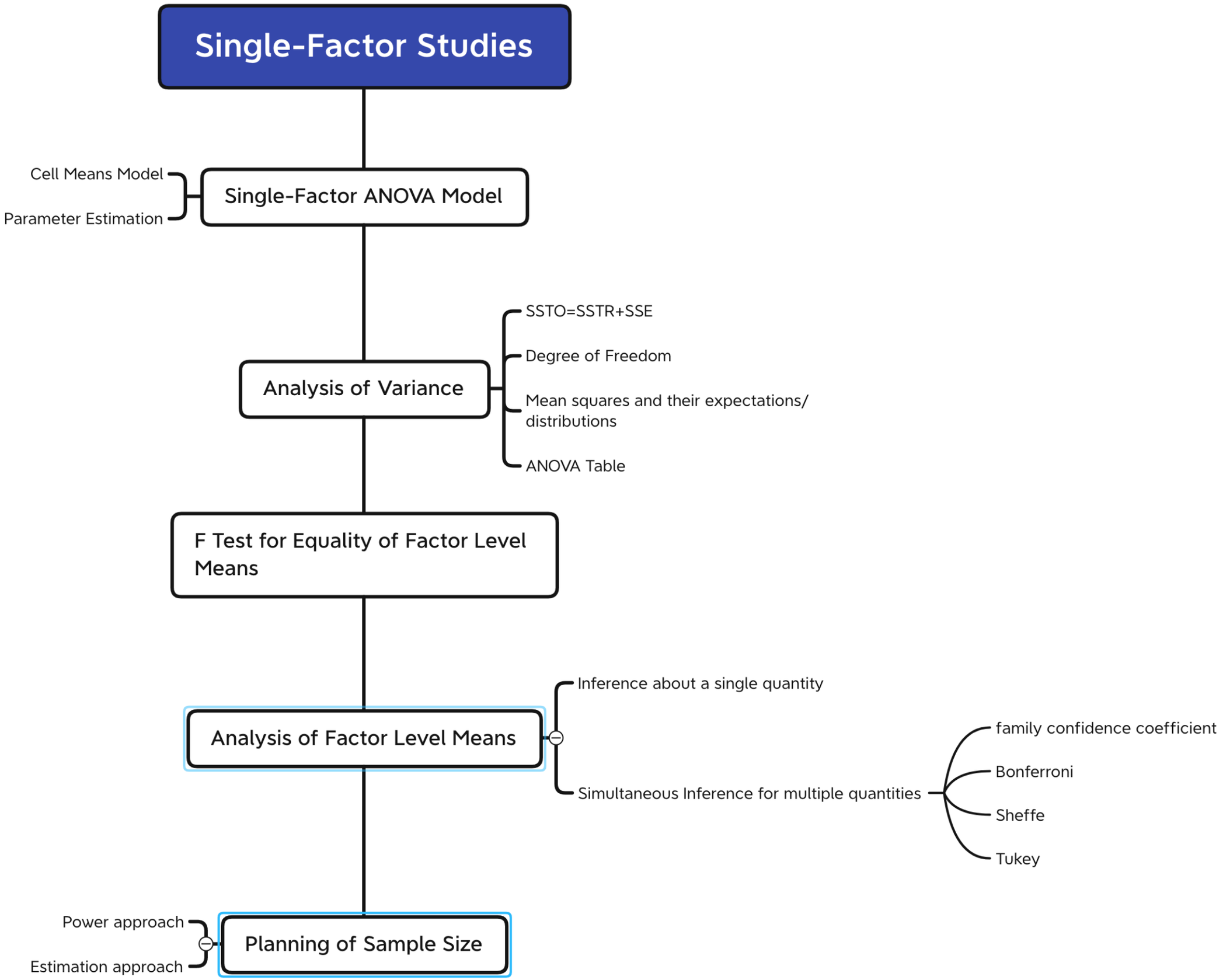
if the sample size for the average physical fitness group is to be: (1) $2n$, total sample size=$4n = 80$.

if the sample size for the average physical fitness group is to be: (1) n, total sample size=$3n = 78$.

if the sample size for the average physical fitness group is to be: (1) $3n$, total sample size=$5n = 90$.

Therefore, equal sample size leads to the smallest total sample size required.

Summary


Single-Factor Studies
Cell Means Model
Parameter Estimation
Single-Factor ANOVA Model
Analysis of Variance
SSTO=SSTR+SSE
Degree of Freedom
Mean squares and their expectations/ distributions
ANOVA Table
F Test for Equality of Factor Level Means
Analysis of Factor Level Means
Inference about a single quantity
Simultaneous Inference for multiple quantities
family confidence coefficient
Bonferroni
Sheffe
Tukey
Power approach
Estimation approach
Planning of Sample Size