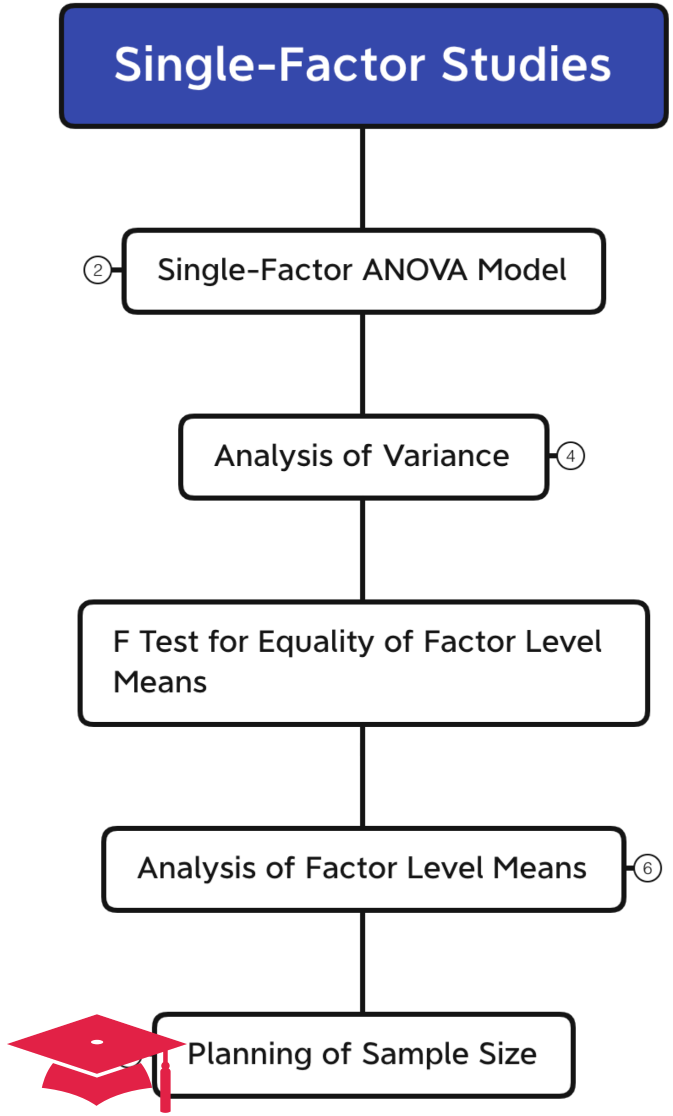
Lecture 2: Single-Factor Studies

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
2
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

Sample size

Too small

Type I and II error
Precision of the estimates of interest

Sweet-spot in between that strikes a balance

Too Large

Cost of the study become excessive

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: $\mu_1 = \ldots = \mu_r$

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

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

MSE and MSTR are independent random variables

Re-write the test statistic:

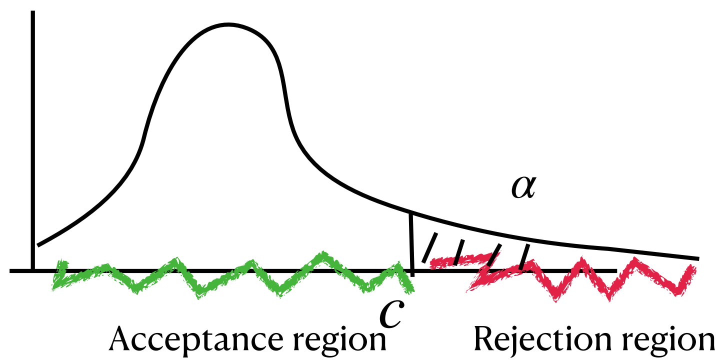
$$\longrightarrow \quad F^* = \frac{MSTR}{MSE} = \frac{\dfrac{\frac{(n-1)MSTR}{\sigma^2}}{n-1}}{\dfrac{\frac{(n_T - r)MSE}{\sigma^2}}{n_T - r}} \sim \frac{\dfrac{\chi^2_{df=r-1}}{r-1}}{\dfrac{\chi^2_{df=n_T-r}}{n_T - r}} \sim F(n_T - r, r-1)$$

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, $n_1 = \ldots = n_r = n$

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

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 $\Delta = max(\mu_i) - min(\mu_i)$

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

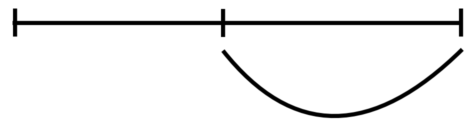
$$\phi = \frac{1}{\sigma^2} \sum_{i=1}^{r} n_i(\mu_i - \overline{\mu_.})^2 \geq \frac{n\Delta^2}{2\sigma^2} = \frac{n}{2}\left(\frac{\Delta}{\sigma}\right)^2$$

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

$\longrightarrow$ As sample size n $\nearrow$ —> width $\searrow$ —> precision $\nearrow$

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 $\mu_1 = 37, \mu_2 = 32, \mu_3 = 28$?

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

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

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

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?

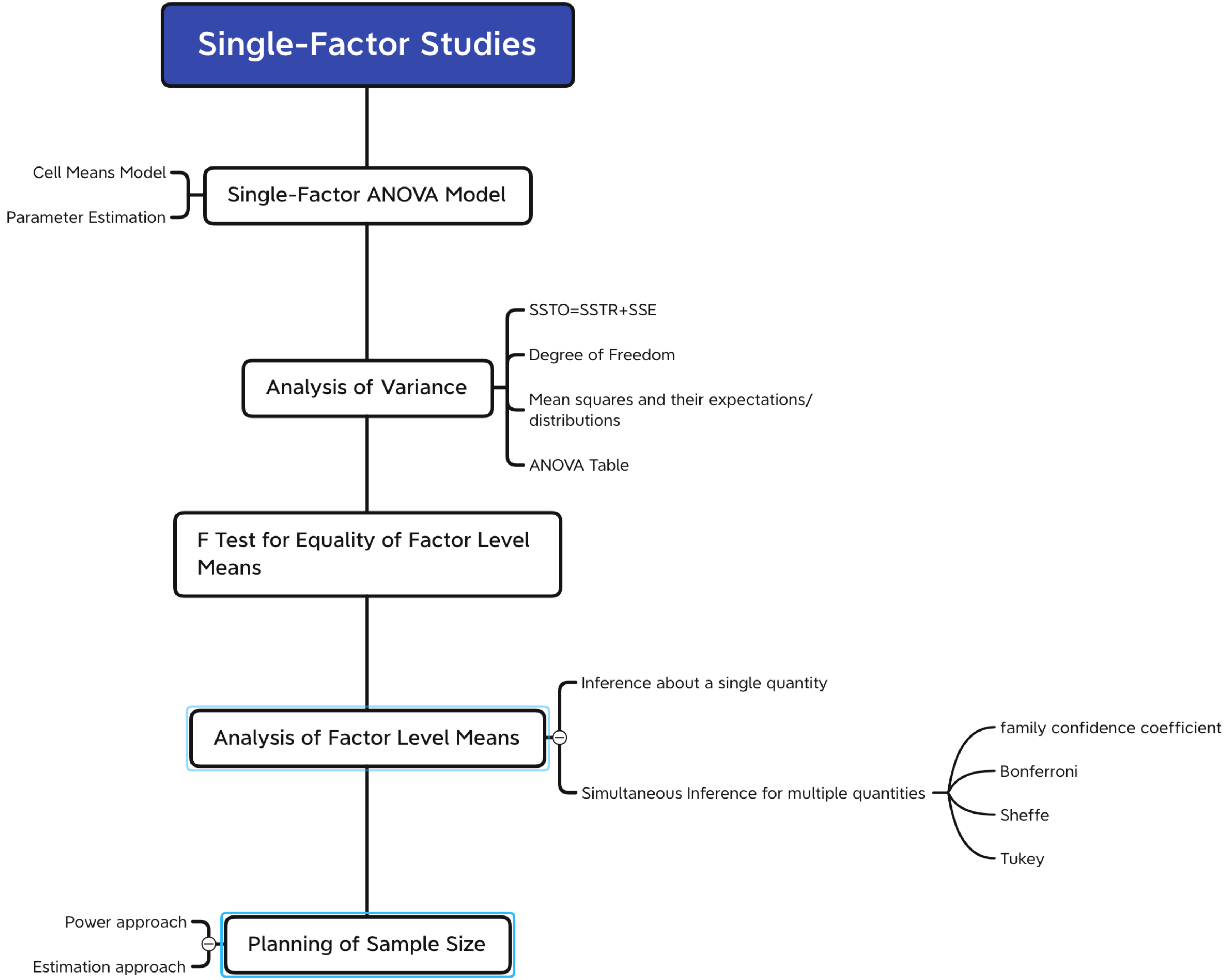
if the sample size for the average physical fitness group is to be: (1) $2n$, total sample size=$4n = 80$.

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
Single-Factor ANOVA Model
Cell Means Model
Parameter Estimation
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
Planning of Sample Size
Power approach
Estimation approach