Single-Factor Studies

- **Factor Effects Model:** (Alternative Formulation of ANOVA Model)

$$Y_{ij} = \mu. + \tau_i + \epsilon_{ij}$$

where:

1. Y_{ij} is the value of the response variable in the jth trial for the ith treatment.
2. $\mu.$ is a constant component common to all observations.
3. τ_i is the effect of the ith factor level (a constant for each factor level).
4. ϵ_{ij} are independent $N(0, \sigma^2)$.
5. $i = 1, \dots, r$; $j = 1, \dots, n_i$.

- **Null and Alternative Hypotheses.**

H_0 : *The factor has NO effect.*	vs.	H_1 : *The factor has an effect.*
$H_0 : \mu_1 = \mu_2 = \cdots = \mu_r$	vs.	H_1 : Not all μ_i are equal.
$H_0 : \tau_1 = \tau_2 = \cdots = \tau_r = 0$	vs.	H_1 : Not all τ_i equal zero.

- **Definition of $\mu.$**

1. ***Unweighted Mean***: $\mu. = \frac{1}{r}\sum_{i=1}^{r} \mu_i$ $\Rightarrow$ $\sum_{i=1}^{r} \tau_i = 0.$

   Example. In a study of length of hospital stay (in number of days) of persons in four income groups, the parameters are as follows: $\mu_1 = 5.1, \mu_2 = 6.3, \mu_3 = 7.9, \mu_4 = 9.5, \sigma = 2.8$. Assume that ANOVA model is appropriate.

2. ***Weighted Mean***: $\mu. = \sum_{i=1}^{r} w_i\mu_i$ where $\sum_{i=1}^{r} w_i = 1$ $\Rightarrow$ $\sum_{i=1}^{r} w_i\tau_i = 0.$

   Example. A car rental firm wanted to estimate the average fuel consumption (in miles per gallon) for its large fleet of cars, which consists of 50% compacts, 30% sedans, and 20% station wagons. In this example, a meaningful measure of $\mu.$ would be in terms of the overall mean fuel consumption for the company. If μ_1, μ_2, and μ_3 are the mean fuel consumptions for the three types of cars in the fleet, define $\mu.$ and propose a way to estimate it.

   Example. When exact weights are unknown, the subgroup sample sizes may be useful as weights of relative importance. For instance, the proportions of households in a city with no children, one child, and more than one child are not known. A random sample of n_T households was selected, which contained n_1 households with no child, n_2 households with one child, and n_3 households with more than one child. If μ_1, μ_2, and μ_3 are the mean entertainment expenditures for the three types of households, define $\mu.$ and propose a way to estimate it.

- **Power of F Test.** The power of the F test for a single-factor study is the probability that the decision rule will lead to conclusion H_a, that the treatment means differ, when in fact H_a is true. That is,

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

where ϕ is the *noncentrality parameter*, that is, a measure of how unequal the treatment means μ_i are:

$$\phi = \frac{1}{\sigma}\sqrt{\frac{\sum n_i(\mu_i - \mu.)^2}{r}}$$

and

$$\mu. = \frac{1}{n_T}\sum n_i\mu_i.$$

When all factor level samples are of equal size n, the parameter ϕ becomes:

$$\phi = \frac{1}{\sigma}\sqrt{\frac{n}{r}\sum(\mu_i - \mu.)^2} \quad \text{where, } \mu. = \frac{1}{r}\sum \mu_i.$$

- **Examples.**

  1. Consider the case where $r = 3$, $n_T = 11$, $\phi = 3$, and $\alpha = 0.05$. Determine the power of the F test.

  2. Suppose an analyst wishes to consider the case when $\mu_1 = 12.5, \mu_2 = 13, \mu_3 = 18, \mu_4 = 21$ with $n_1 = 5, n_2 = 5, n_3 = 4, n_4 = 5$ and $\sigma = 3.5$. Find ϕ and then calculate the power of the F test.

- **Alternatively, you can use the following R functions:**

```
power.anova.test(groups = 4, n = 5, between.var = 1, within.var = 3)       # Calculates the power.
[1] Power = 0.3535594

power.anova.test(groups = 4, between.var = 1, within.var = 3, power = .80) # Calculates the sample size.
[1] n = 11.92613

# Assuming we have prior knowledge of the group means:
groupmeans = c(120, 130, 140, 150)
power.anova.test(groups = length(groupmeans),between.var = var(groupmeans),within.var = 500, power = .90)
[1] n = 15.18834

# Using the R package ‘pwr’:
library(help=pwr)
pwr.anova.test(f=0.28,k=4,n=20,sig.level=0.05)       # Calculates the power.
pwr.anova.test(f=0.28,k=4,power=0.80,sig.level=0.05) # Calculates the sample size.
```

- **Homework.**

  1. Do 16.24 and 16.25, on page 728.

  2. Calculate the power of a balanced ANOVA test involving 4 treatment arms, 10 responses per treatment, $\sigma^2 = 4$, and variance between treatment means of 2. Use a 0.01 level of significance.

  3. Using these two R functions, determine the effects of n, k, f, α, σ^2, and variance between treatment means to the power of the ANOVA test.