9 Factorial Experiments

Choose several ____________ of each of several ____________ (or ____________), and then test each possible combination. A combination is called a ____________.

9.1 One-Factor Experiments

9.2 Pairwise Comparisons in One-Factor Experiments

9.3 Two-Factor Experiments

9.4 Randomized Complete Block Designs

9.5 2^p Factorial Experiments

9.1 One-Factor Experiments

One-way analysis of variance (ANOVA)

A *factor* is ______________________________ in an experiment; its values are ____________.

A *one-way analysis of variance* (ANOVA) tests "$H_0 : \mu_1 = \cdots = \mu_I$", where I is the ____________ for one factor, against "H_1 : at least two μ_i's are different" by comparing two estimates of σ^2. One estimate is from variation ______________________, and the other is from variation ______________________. Their ratio is compared to an ______ distribution.

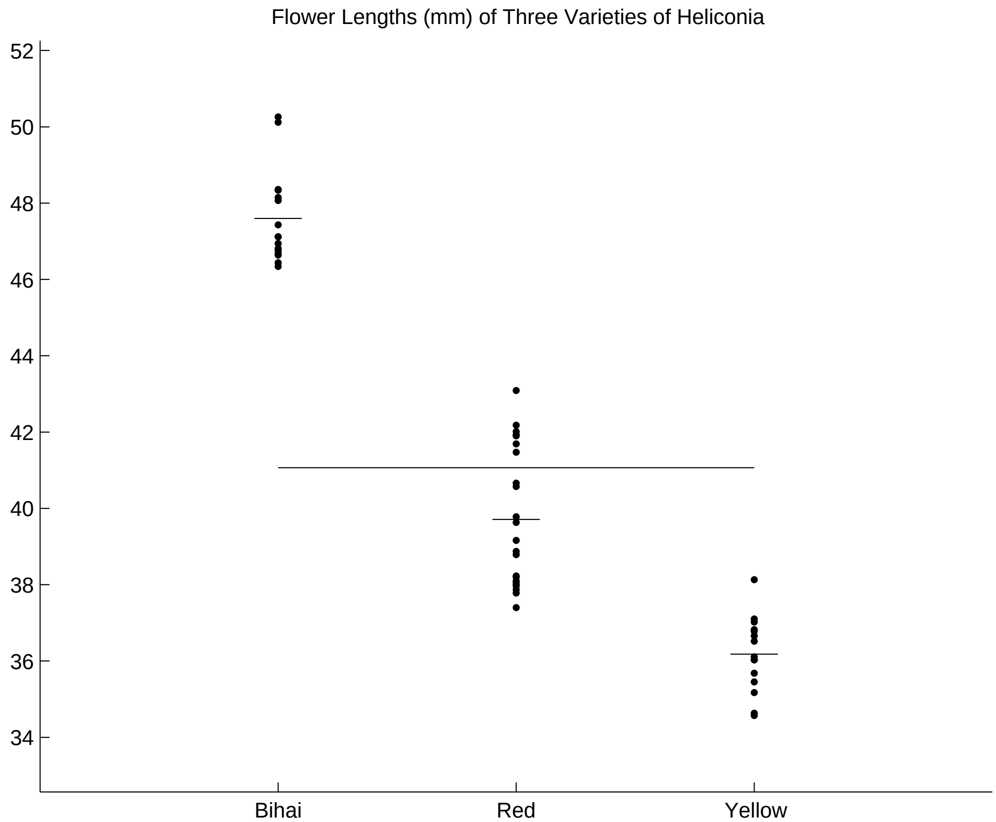
e.g. Here are lengths (mm) for samples of three varieties of Heliconia flowers on Dominica, each fertilized by a different species of hummingbird. The flowers' forms and birds' beaks seem to have co-evolved to match each other (draw).

								J_i	$\bar{x}_{i.}$	s_i
Heliconia bihai										
47.12	46.75	46.81	47.12	46.67	47.43	46.44	46.64			
48.07	48.34	48.15	50.26	50.12	46.34	46.94	48.36	16	47.60	1.21
Heliconia caribaea red										
41.90	42.01	41.93	43.09	41.47	41.69	39.78	40.57			
39.63	42.18	40.66	37.87	39.16	37.40	38.20	38.07			
38.10	37.97	38.79	38.23	38.87	37.78	38.01		23	39.71	1.80
Heliconia caribaea yellow										
36.78	37.02	36.52	36.11	36.03	35.45	38.13	37.10			
35.17	36.82	36.66	35.68	36.03	34.57	34.63		15	36.18	.98

Overall: $N = 54, \bar{x}_{..} = 41.07$

e.g. The factor in this study is ______________________. It has three levels: ______________________

Here is a plot of the data:



Notation

- $I = \#$______________
- $J_i = i^{\text{th}}$ sample ______________ $(i = 1 \text{ to } I)$; $N = \sum_{i=1}^{I} J_i$ = sum of all sample ______________
- $X_{ij} = i^{\text{th}}$ sample's j^{th} observation (______________ in plot)
- $\bar{X}_{i.} = \frac{1}{J_i} \sum_{j=1}^{J_i} X_{ij} = i^{\text{th}}$ sample's mean (_____________________ in plot)
- $\bar{X}_{..} = \frac{1}{N} \sum_{i=1}^{I} \sum_{j=1}^{J_i} X_{ij} = \frac{1}{N} \sum_{i=1}^{I} J_i \bar{X}_{i.}$ = *grand sample mean* (_________________ in plot)
- $X_{ij} - \bar{X}_{i.}$ = *residual* of i^{th} sample's j^{th} observation with respect to i^{th} sample's mean (the difference from ____________________________ in plot)

Is the visually apparent difference in sample means $\{\bar{X}_{1.}, \bar{X}_{2.}, \bar{X}_{3.}\}$ statistically significant? It matters only _____________________ the spread of individual observations. Is it due to ______________, or statistically significant?
(Could we run a difference-of-two-means test for each pair of means? _____________________)

Sums of Squares

Partition the sum of squared deviations from __________________________ into the sum of squared deviations from ____________________ plus the sum of squared deviations of sample means from ____________________.

- The *total sum of squares* is

$$\text{SST} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(X_{ij} - \bar{X}_{..})^2$$

- Variation _____________ samples is measured by the _____________ *sum of squares* (SSE)

$$\text{SSE} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(X_{ij} - \bar{X}_{i.})^2 = \sum_{i=1}^{I}(J_i - 1)s_i^2$$

  Its average is the ____________________ *square* (MSE)

$$\text{MSE} = \frac{SSE}{N - I}$$

  where $N - I$ is the degrees of freedom for SSE. (It's "_______", not "_______".)
  s^2 = MSE is an estimate of _______.
  e.g. MSE is the variance of the residuals (_____________ minus respective _________________):
  SSE =

  MSE =

- Variation _____________ samples is measured by the _____________ *sum of squares*,

$$\text{SSTr} = \sum_{i=1}^{I} J_i(\bar{X}_{i.} - \bar{X}_{..})^2$$

  Its average is the ____________________ square,

$$\text{MSTr} = \frac{SSTr}{I - 1}$$

  where _____________ is the degrees of freedom for SSTr.
  A _________ MSTr is evidence __________ H_0, while a _________ MSTr leaves H_0 __________.
  e.g. MSTr measures variation of the __________________________ about the _________________:
  SSTr =

  MSTr =

Here the analysis of variance identity is SST = _______ + _______. SST is used in finding the *coefficient of determination*: $R^2 = \frac{\text{SSTr}}{\text{SST}}$.

Assumptions

- The treatment populations are ______________
- The treatment populations all have the same _____________________
- All observations are ______________

Checks include:

- A probability plot of residuals $\{X_{ij} - \bar{X}_{i.}\}$ against $N($_______, _______$)$ should be $\approx$ _____________________
- The ______________ sample standard deviation shouldn't be more than about _____________________ as the ______________ sample standard deviation.
- A residual plot of residuals $\{X_{ij} - \bar{X}_{i.}\}$ against their respective sample mean $\bar{X}_{i.}$ should show reasonbly ______________ spread across samples and no ______________

The F-test of $H_0 : \mu_1 = \cdots = \mu_I$ for One-Way ANOVA

Under H_0, MSTr and MSE are both estimates of σ^2, the common variance of the populations.

- MSTr ______________ on the truth of H_0
  - H_0 true $\implies \mu_{\text{MSTr}} = \sigma^2$
  - H_0 false $\implies \mu_{\text{MSTr}}$_______$\sigma^2$
- MSE ____________________________ on the truth of H_0. Either way, $\mu_{\text{MSE}} = \sigma^2$.
- Use the test statistic $F = \dfrac{\text{MSTr}}{\text{MSE}} = \dfrac{\text{average variation among sample means}}{\text{average variation among individuals in the same sample}}$
  - H_0 true $\implies F$ should be near ______________
  - H_0 false $\implies F$ should be ______________

  Under H_0, $F \sim F_{I-1,N-I}$. (It's "_______", not "$N-1$".)

e.g. For the flower data, find F and its P-value, and draw a conclusion.

$F =$ $\implies$ P-value =

$\implies$

The preceding work can be summarized in this ANOVA table:

Source	DF	SS	MS	F	P
Treatment	2	1084.00	541.50	259.09	< .001
Error	51	106.58	2.09		
Total	53	1190.58			

Rejecting $H_0 : \mu_1 = \cdots = \mu_I$, doesn't tell ____: §9.2.

(A *balanced* experiment has equal sample sizes and is more robust against __________________ than an unbalanced one. $I = 2 \implies F$-test is __________ to t-test of $H_0 : \mu_X - \mu_Y = 0$, with $F =$ _____.)