Chapter 10 Summary

§10.1 Analysis of Variance

X is divided by "factor" values $X^{(1)}, X^{(2)}, \dots, X^{(k)}$

Test hypothesis

H_0: All $X^{(f)}$ have same distribution

(assuming all variances are equal $\sigma = Var[X^{(f)}]$)

Plan: Compare variances

$$\frac{\text{variance between factors}}{\text{variance within factors}} \sim F$$

Notation: For factor value f $\rightsquigarrow$ $X^{(f)}$ = random variable

n_f = # samples and $\bar{x}^{(f)}$ = sample mean

	Sum of Square (SS)	Degrees of Freedom (df)	Mean Square (MS)
Factor	$SS_F = \sum_f n_f \cdot (\bar{x}^{(f)} - \bar{x})^2$	# factors − 1	$MS_F = SS_F / \#\text{factors} - 1$
Residual	$SS_R = \sum_{f,i} (x_i^{(f)} - \bar{x}^{(f)})^2$	# samples − # factors	$MS_R = SS_R / \#\text{samples} - \#\text{factors}$
Total	$SS_T = \sum_{f,i} (x_i^{(f)} - \bar{x})^2$	# samples − 1	$MS_T = SS_T / \#\text{samples} - 1$

[sometimes MS_T will be omitted]

Note: $SS_F + SS_R = SS_T$

Note: $df_F + df_R = df_T$

Warning: $MS_F + MS_R \neq MS_T$

(fractions don't work this way...)

$$F = MS_F / MS_E \sim F(\#\text{factors} - 1, \#\text{samples} - \#\text{factors})$$

- If factor doesn't matter (H_0 True) then $F \approx 1$
- If factor does matter (H_0 False) then $F \gg 1$

Effect Size $\eta^2 = \dfrac{SS_F}{SS_T} = \dfrac{SS_F}{SS_F + SS_E}$

$0 < \eta^2 < 1$

$$\eta^2 \approx \begin{cases} 0 & \text{if factor isn't important} \\ 1 & \text{if factor is important} \end{cases}$$