

8) Analysis of Variance (ANOVA)

Vitor Kamada

January 2018

Tables, Graphics, and Figures from
**Introductory Statistics with
Randomization and Simulation**

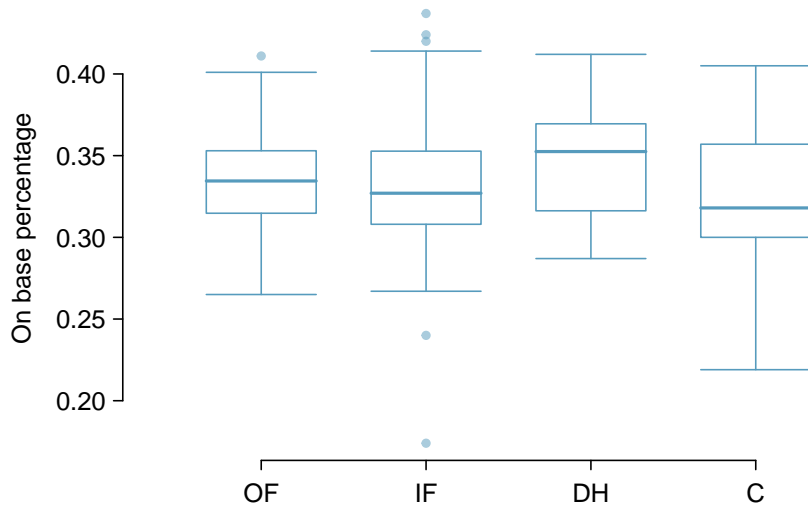
Diez et al. (2014): Chapter 4 - Inference for
Numerical Data

Major League Baseball Data

	name	team	position	AB	H	HR	RBI	AVG	OBP
1	I Suzuki	SEA	OF	680	214	6	43	0.315	0.359
2	D Jeter	NYN	IF	663	179	10	67	0.270	0.340
3	M Young	TEX	IF	656	186	21	91	0.284	0.330
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮		
325	B Molina	SF	C	202	52	3	17	0.257	0.312
326	J Thole	NYM	C	202	56	3	17	0.277	0.357
327	C Heisey	CIN	OF	201	51	8	21	0.254	0.324

variable	description
position	The player's primary field position (OF, IF, DH, C)
AB	Number of opportunities at bat
H	Number of hits
HR	Number of home runs
RBI	Number of runs batted in
AVG	Batting average, which is equal to H/AB
OBP	On-base percentage, which is roughly equal to the fraction of times a player gets on base or hits a home run

Is Batting Performance related to Player Position?



Summary Statistics of On-base Percentage

$$H_0 : \mu_{OF} = \mu_{IF} = \mu_{DH} = \mu_C$$

H_A : The average on-base percentage (μ_i) varies across some (or all) groups.

	OF	IF	DH	C
Sample size (n_i)	120	154	14	39
Sample mean (\bar{x}_i)	0.334	0.332	0.348	0.323
Sample SD (s_i)	0.029	0.037	0.036	0.045

Mean Square between Groups:

$$MSG = \frac{SSG}{df_G} = \frac{1}{k-1} \sum_{i=1}^k n_i (\bar{x}_i - \bar{x})^2$$

Mean Square Error:

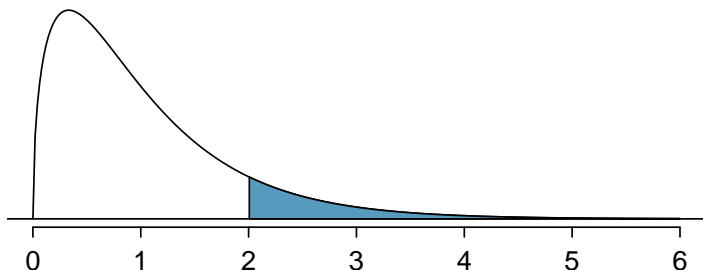
$$MSE = \frac{SSE}{df_E} = \frac{1}{n-k} \sum_{i=1}^k (n_i - 1) s_i^2$$

$$F = \frac{MSG}{MSE}$$

Analysis of Variance (ANOVA)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
position	3	0.0076	0.0025	1.9943	0.1147
Residuals	323	0.4080	0.0013		

$$s_{pooled} = 0.036 \text{ on } df = 323$$



Normal Probability Plot of On-base Percentage

