

Crafting with Data

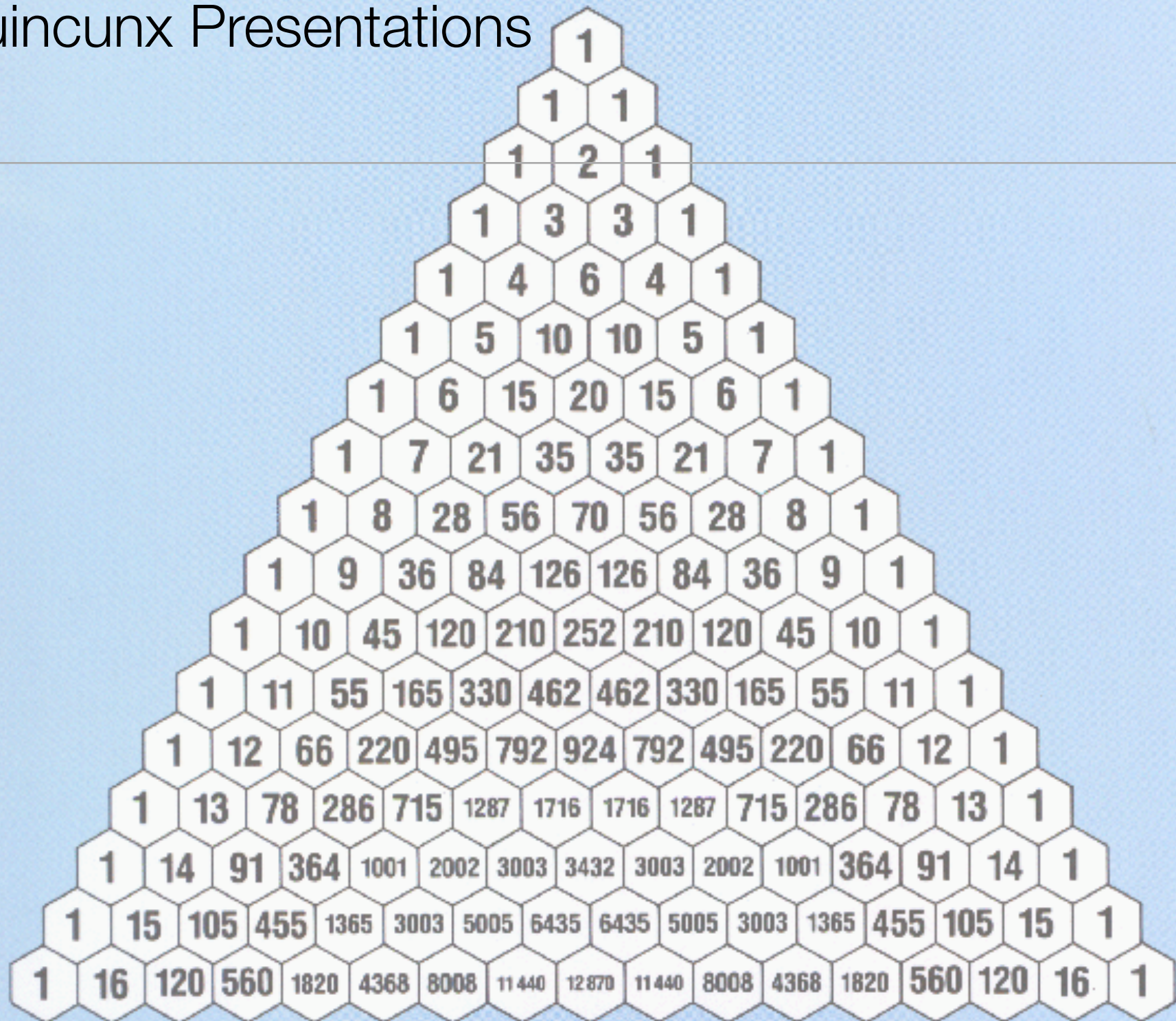
Reality, Illusions, Truth & the Future

Instructor: Rob Faludi

Plan for Today

- Confirmatory Statistics, review and completion
- Quincunx presentations
- Overview of statistical packages
- Readings & Assignments

Quincunx Presentations



Binomial Distribution



Binomial Distribution

- The probability that a random variable X with binomial distribution $B(n,p)$ is equal to the value k , where $k = 0, 1, \dots, n$. p is the probability of success

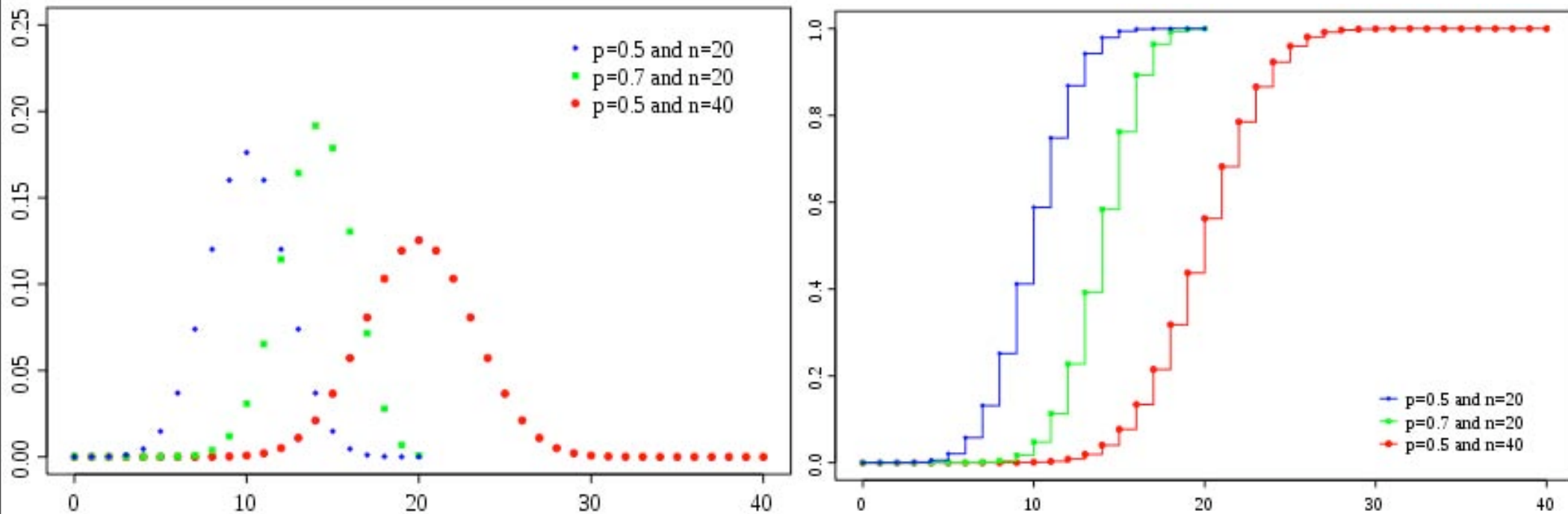
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- n choose k (number of ways to choose k successes from n variations, like five heads in ten coin tosses)

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Binomial Distribution

- Mass and Cumulative



<http://www.stat.tamu.edu/~west/applets/binomialdemo.html>

Chi-square



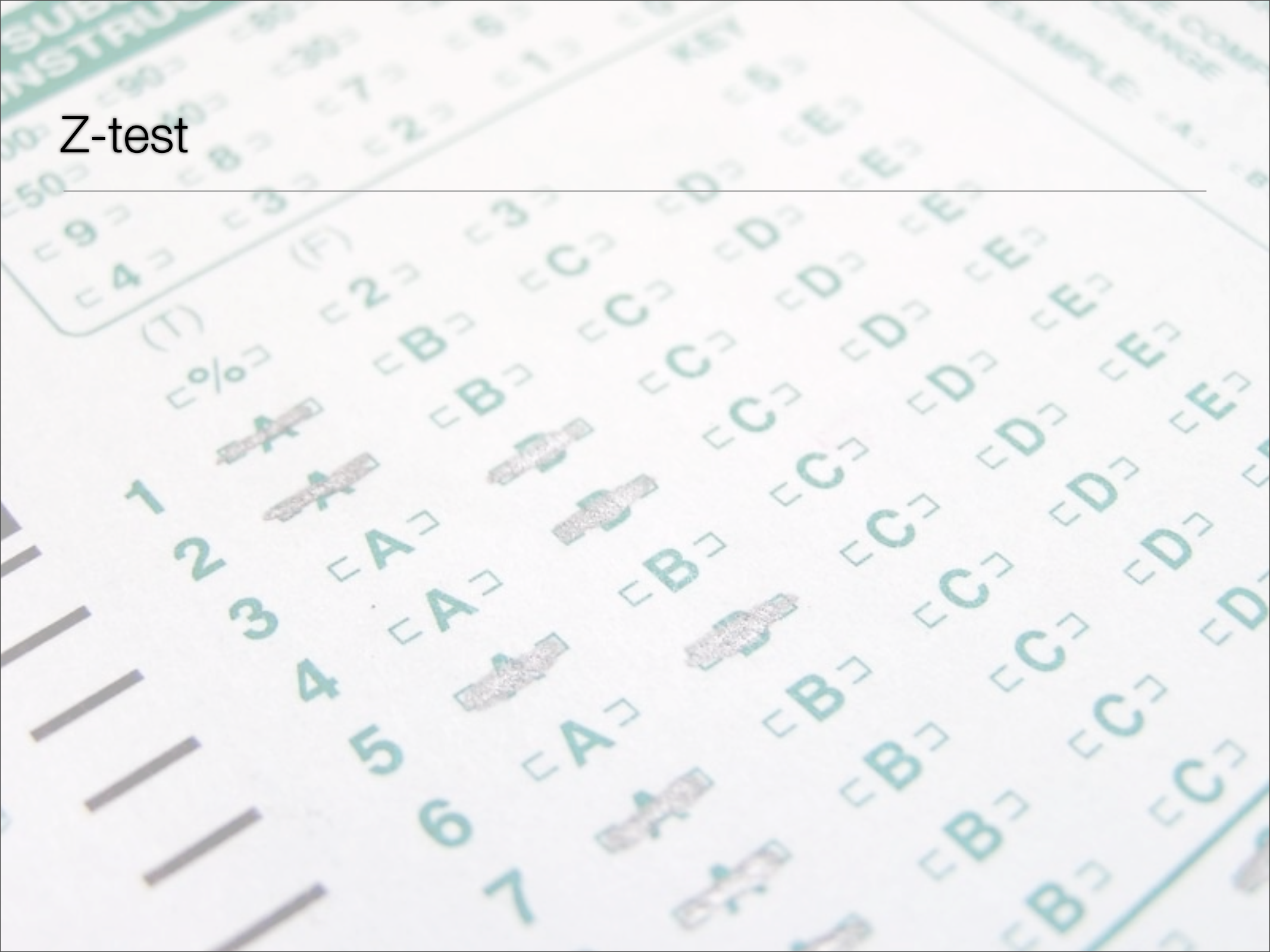
Chi-square

- Where O is the observed frequency and E is the expected frequency

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

- look up result in a table, degrees of freedom are $k-1$
- breaks down if expected frequencies are too low

Z-test



Z-test

- Used in standardized testing
- Requires a random sample
- Requires normal distribution
- Need to know sigma, the standard deviation of the population

$$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$$

- lookup result in table

T-test



T-test

- Requires a random sample
- Requires normal distribution
- Sigma not needed, yay!
- difference between the means
- variability of the samples
- lookup result in table

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

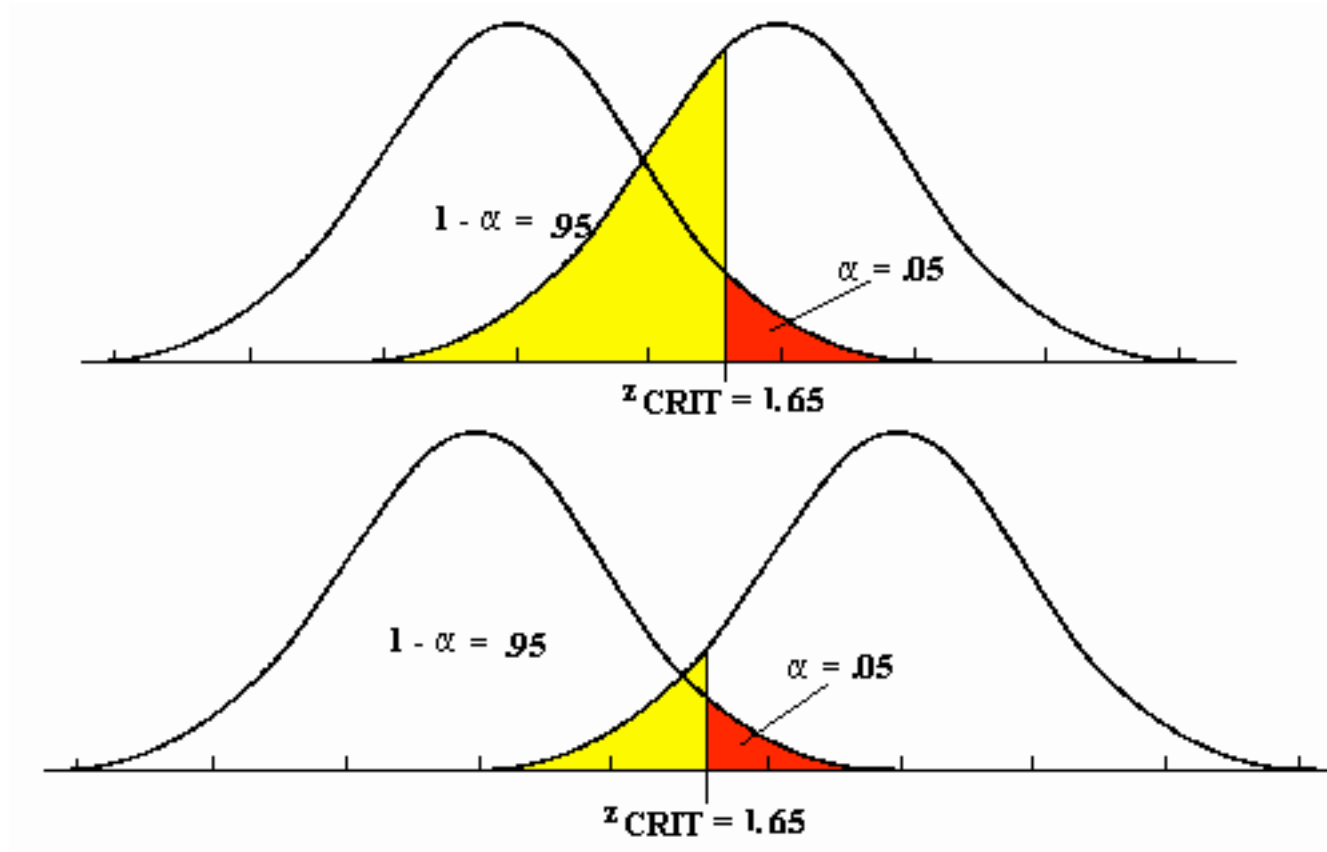


- nice walkthrough: <http://www.biologyforlife.com/ib%20bio/Internal%20Assessment/stats.htm#T-Test>

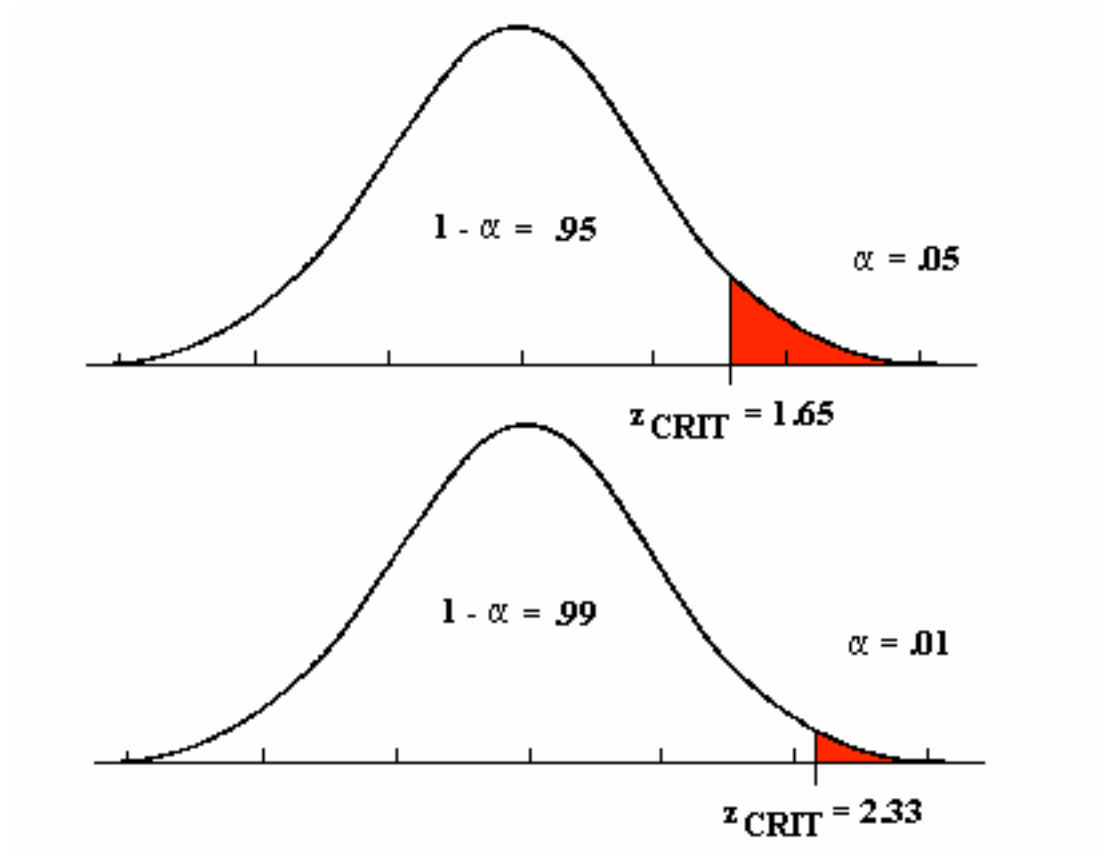
Some Errors



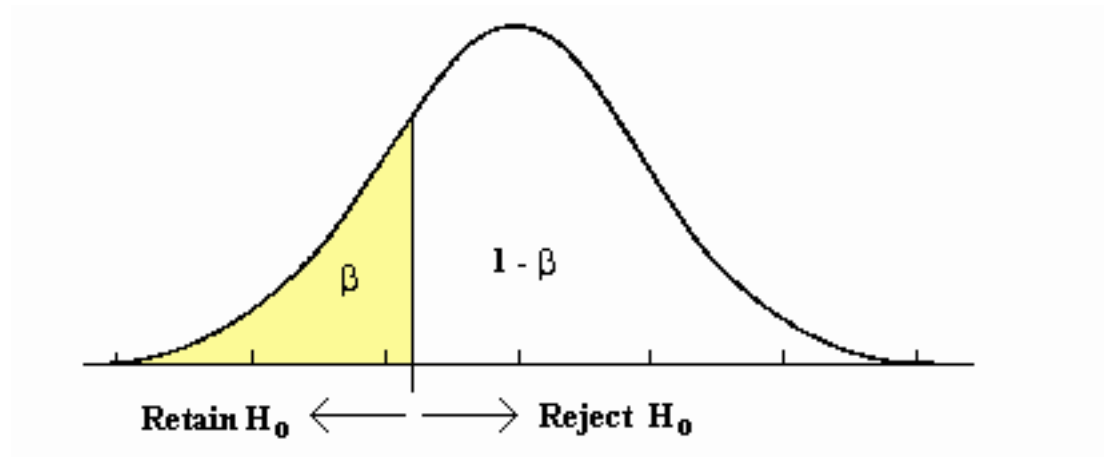
Type 1 and Type 2 Errors



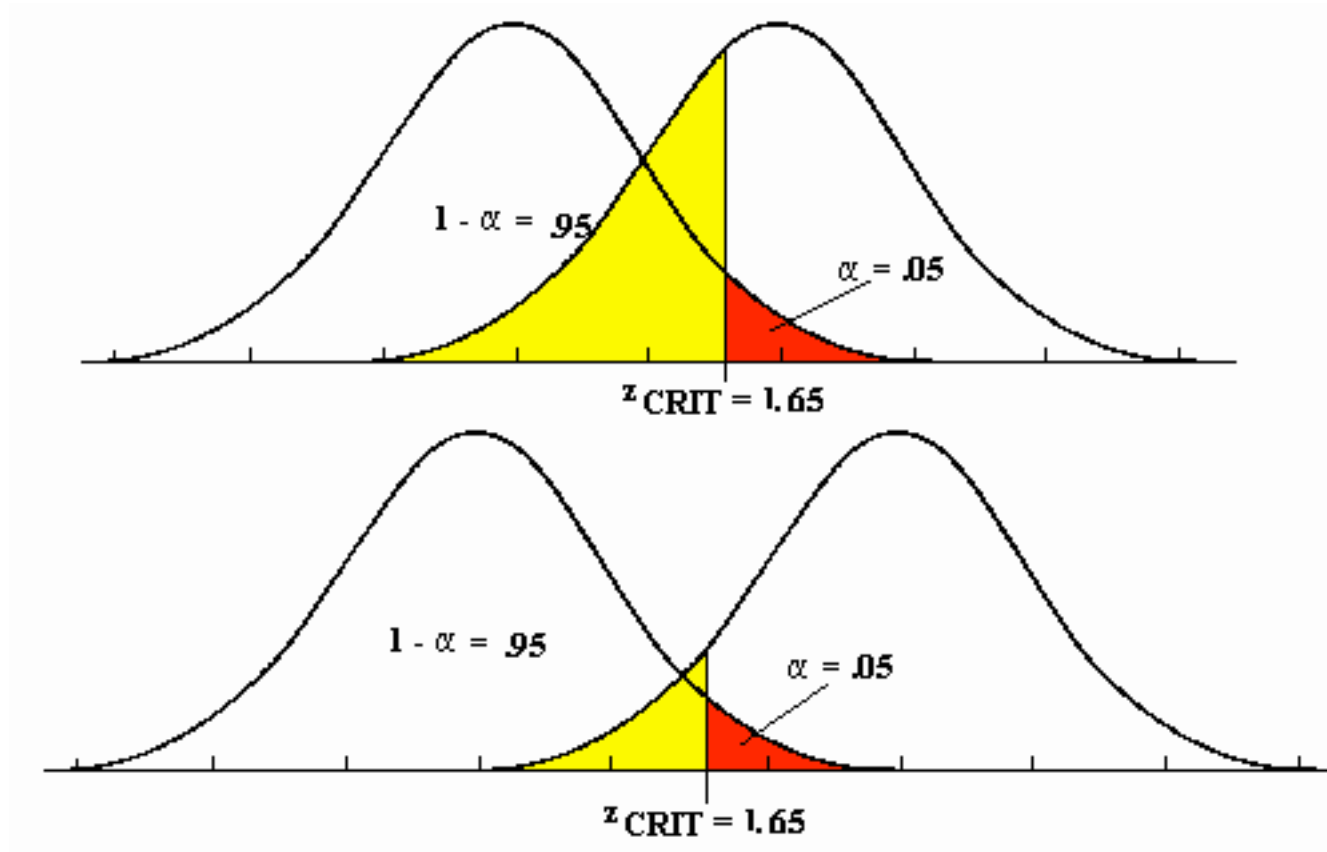
Type 1 Error



Type 2 Error



Type 1 and Type 2 Errors

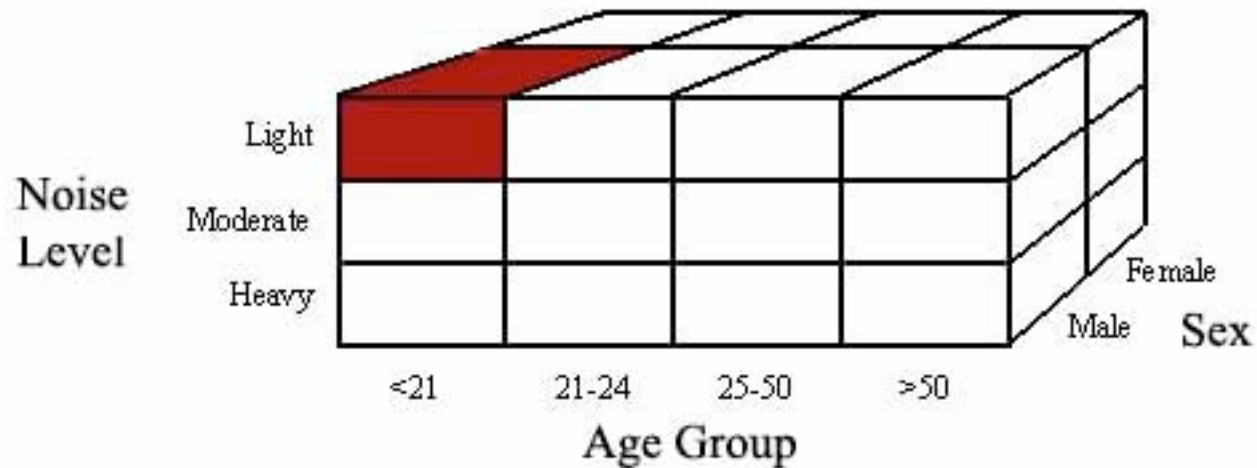


ANOVA: Analysis of Variance



ANOVA: Analysis of Variance

Between-Groups Design



ANOVA: Analysis of Variance

- When you want to compare many factors

$$\mathbf{F} = \frac{\text{Var Between Means}}{\text{Var Within Groups}} = \frac{\text{MS}_{\text{Bet}}}{\text{MS}_{\text{Within}}}$$

- It's just a ratio, then look the value up in a table
- <http://www.psych.utah.edu/stat/introstats/anovafash.html>

- Multivariate analysis of variance: MANOVA

$$SS_T = \sum x^2 - \frac{(\sum x_T)^2}{N}$$

$$SS_b = \sum \frac{(\sum x.)^2}{n} - \frac{(\sum x_T)^2}{N}$$

$$SS_w = SS_T - SS_b$$

$$df_b = (\text{number of groups} - 1)$$

$$df_T = (\text{number of subjects} - 1)$$

$$df_w = df_T - df_b$$

$$MS_b = \frac{SS_b}{df_b}$$

$$MS_w = \frac{SS_w}{df_w}$$

$$F = \frac{MS_b}{MS_w}$$

SPSS, SAS, Excel, Docs, Java, R and free stuff



The Predictive Analytics Company



Free Statistics

www.freestatics.info

<http://www.r-project.org/>

<http://www.freestatics.info/stat.php>

Readings and Assignments

- Readings

- How to Lie with Maps, *selected chapters for Wednesday*

