

Crafting with Data

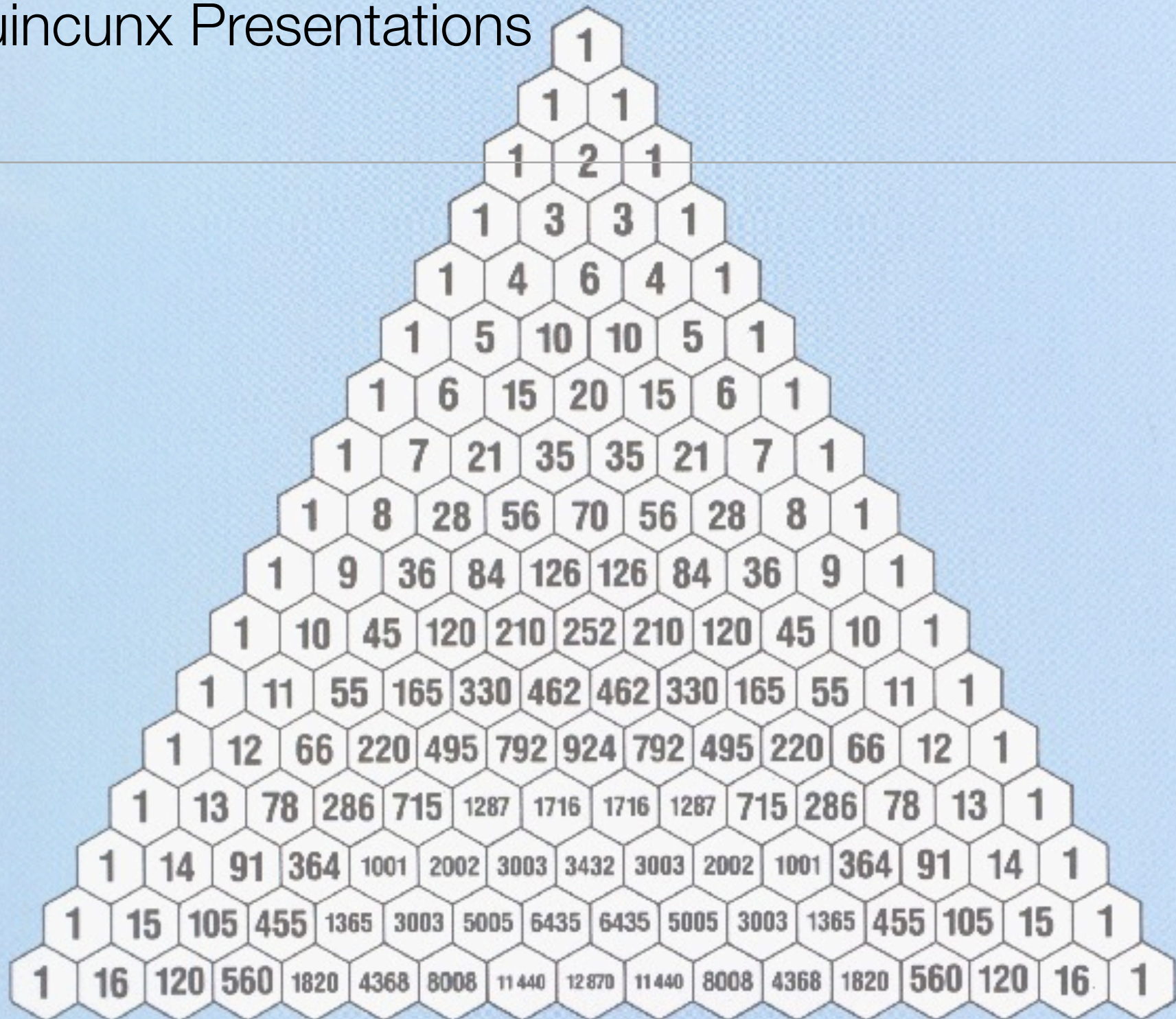
Reality, Illusions, Truth & the Future

Instructor: Rob Faludi

Plan for Today

- Present Quincunx results
- Statistics Intro
- Confirmatory Statistics
 - Binomial
 - Chi-square
 - Z-test
 - T-test
- Readings & Assignments

Quincunx Presentations



Statisticians

- Aristotle (statements based upon thought)
- Galileo (statements based upon observation)
- Newton & Copernicus (clockwork universe)
- Pascal & Fermat (probability)
- Bernoullis (many elements of modern statistics)

Statistics is New



Thomas Bayes



Blaise Pascal



Jakob Bernoulli



Daniel Bernoulli



Pierre Fermat



Francis Galton

...really new



Ronald Fisher



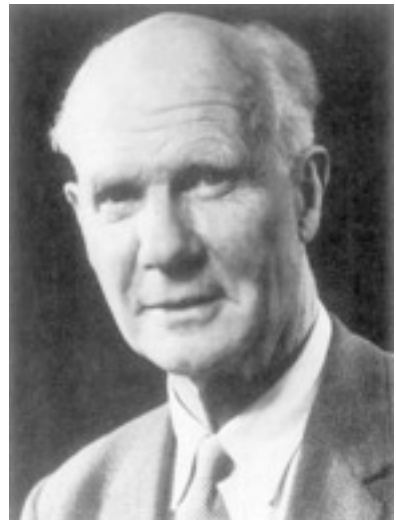
Stanislaw Ulam



Karl Pearson



A. N. Kolmogorov



Egon Pearson



John von Neumann



John W. Tukey



William Gossett
(‘Student’)

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

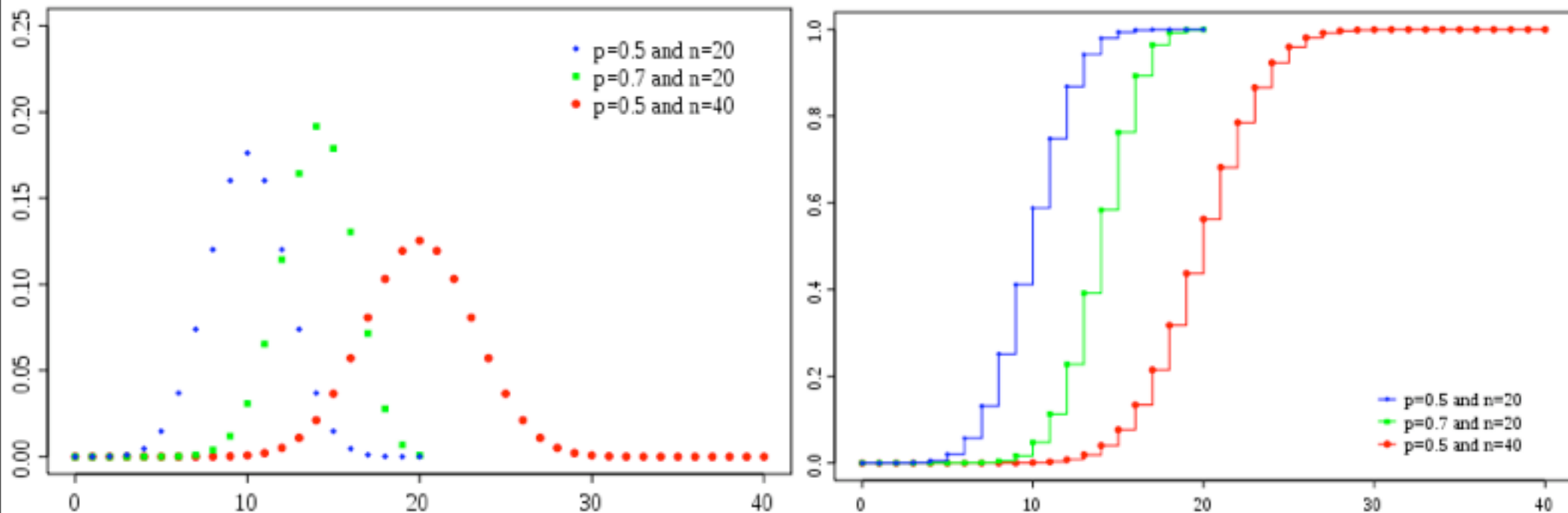
$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

- 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

Chi-test calculation

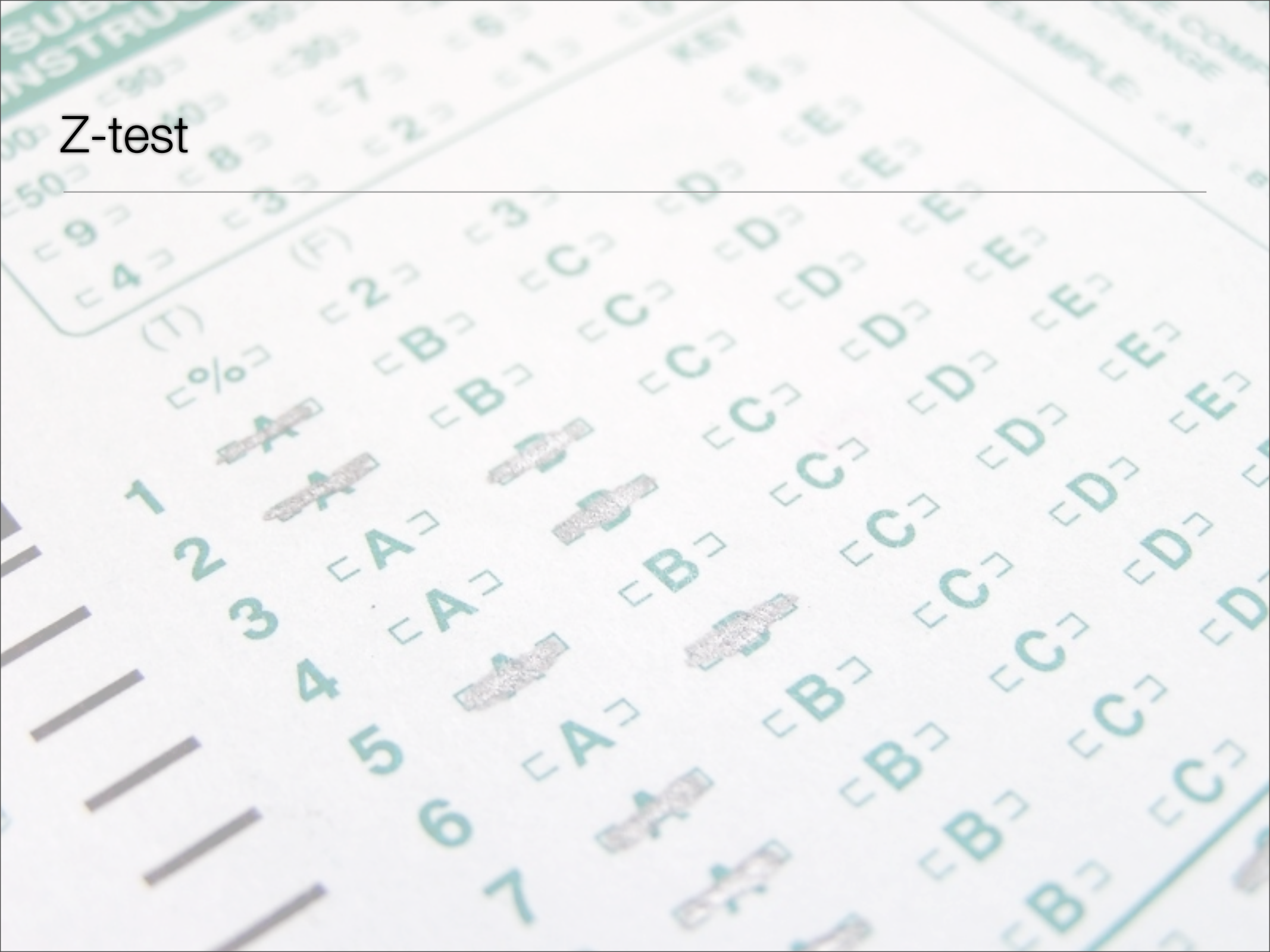
- in Excel: `CHITEST(actual_range, expected_range)`
- in Google spreadsheets it has to be done by hand. Bleh.
- in R it will be `chisq.test`

Spinning Coins

- Results
- Chi-square



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>

T-test calculation

- in Excel: `TTEST(array1,array2,tails,type)`
- in Google spreadsheets it has to be done by hand, but this is instructive
- in R it will be `t.test`

Readings and Assignments

- Readings
 - The Lady Tasting Tea, by David Salsburg — Chapters 3, 4 & 5
- Assignments
 - Finish your documentation
 - IRB test

Exercise One

- Individually, measure how tall your classmates are
 - Try to use a unique method
 - Do not share answers
 - Write down your results
- When you are done, we'll go around the room and add each measurement to our database
- Histograms in spreadsheets: BINS, FREQUENCY, CHARTS