

Cancer Biostatistics Workshop



2008 Schedule



Series One:

What You Probably Don't Know About Biostatistics

January 18th - Yu Shyr, PhD

"The Science of Doing Science – Biostatistics"

February 15th – Dan Ayers, MS

"Skittles, An Ounce of Measurement, and the Intent of
Science"

March 21st – Ayumi Shintani, MPH, PhD

"Introduction to Power and Sample Size Estimation"



Series Two:

Essentials of Efficient Experimental Design

April 18th - Tatsuki Koyama, PhD

“Phase II Clinical Trials”

May 16th - Leena Choi, PhD

“Adaptive Designs in Dose-Finding (Phase I) Studies: A Bayesian Approach”

June 20th - Fei Ye, PhD

“Phase III Clinical Trials”



Series Three: High-Dimensional Data

August 15th - Chun Li, PhD

“Genome-wide Association Analysis for the
Shanghai Breast Cancer Study”

September 19th - Irene Feurer, PhD

“An Introduction to Principal Component Analysis
and Data Reduction Methodologies”

October 17th - Ming Li, PhD

“Statistical Analysis Strategies for MALDI-TOF and
Shotgun Proteomics Data”



Special Topics:

Observational Data and Epidemiology

November 21st - Ayumi Shintani, MPH, PhD

“Introduction to Observational Study Design”

~ ~ - - - - - ~ ~ ~ ~ ~

Course schedule information and .pdf of past talks can be found at:

<http://www.vicc.org/biostatistics/workshop2008.php>

Please call 6-2502 or e-mail shaun.m.haskins@vanderbilt.edu if you have any questions or feedback.



Skittles,
An Ounce of Measurement, and
the Intent of Science

Dan Ayers

Department of Biostatistics

Cancer Biostatistics Center

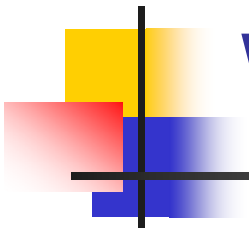
Vanderbilt-Ingram Cancer Center



Audience

- How many basic scientists?
- How many clinicians?
- How many submitted research for publication or grant?
- How many consulted with a statistician?
- For how many was this consultation beneficial?

A Simple Study: What can happen?



Truth

Study Says

	No Effect	Effect
No Effect	correct	Type II error
Effect	Type I error	correct



A Simple Study

- Primary outcome is a response
- Uninteresting response rate = 20%
- Interesting response rate = 40%
- Two-sided Type I error rate = 5%
- Sample size of 40 individuals gives:
 - 80% power to detect a $\pm 20\%$ difference in response rate)
 - If ≤ 2 or ≥ 14 reject the hypothesis that the true response rate is 20%



A Skittles Experiment (or the power of randomness)

- 10 , 1lb bags of skittles
- 42 skittles weigh 2.17 oz
- Population size = 3097 skittles
- Randomly select 30 samples of 40 skittles each
- Count the number of purple skittles (number of responses)

The Truth

History

1979

SKITTLES® candies roll out across the US in force in the original flavor combination of Orange, Lemon, Lime, Grape and Strawberry. At first they're imported, but from 1982 SKITTLES® candies are manufactured in the US.

| [Next date >](#)



Facts

Divided we stand

Each flavor makes up approximately 20 percent of each bag.

[Next fact >](#)



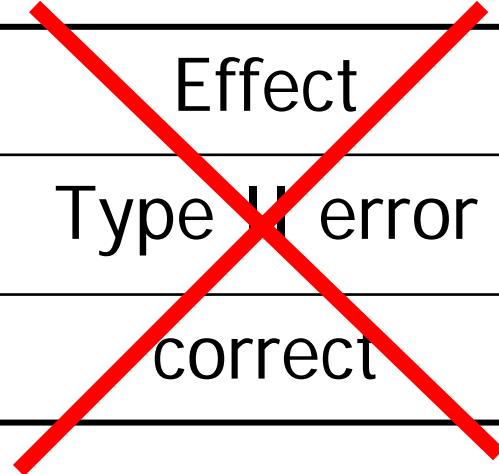
Expectations and decisions

- $20\% \times 40$ skittles = 8 purple skittles
- 2 or fewer purple skittles and we conclude the true proportion $< 20\%$
- 14 or more purple skittles and we conclude the true proportion $> 20\%$.



What can happen?

		Truth	
		No Effect	Effect
Study Says	No Effect	correct	Type II error
	Effect	Type I error	correct



Unlike biological experiments, we know the truth in statistical experiments!



What did happen?

- Everyone with a numbered bag of Skittles please stand up
 - 30 people
- Everyone with 3 or more or 13 or fewer purple skittles please sit down
 - 28 people
- 2 people left (6.7%)



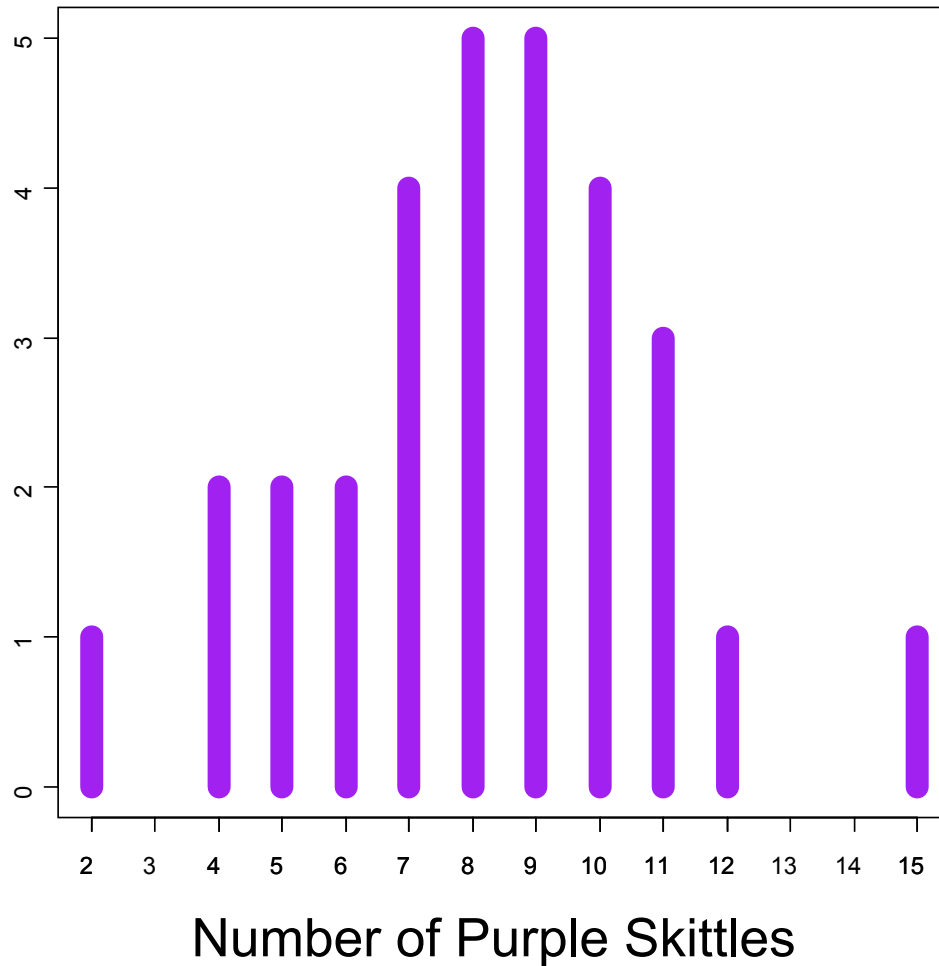
What did happen?

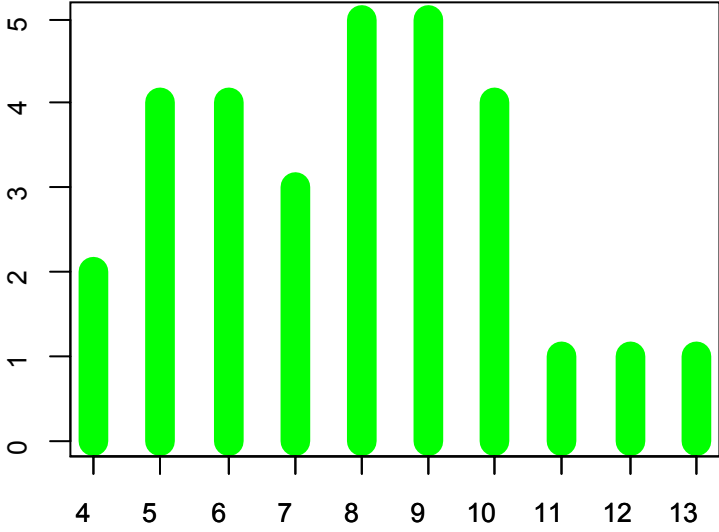
- Random variation happened.
- Randomness elicits surprising results as the rule, not the exception.
- Which experiments of the 30 conducted are more likely to be published?

The Studies That Led Us
to the Wrong Conclusion!

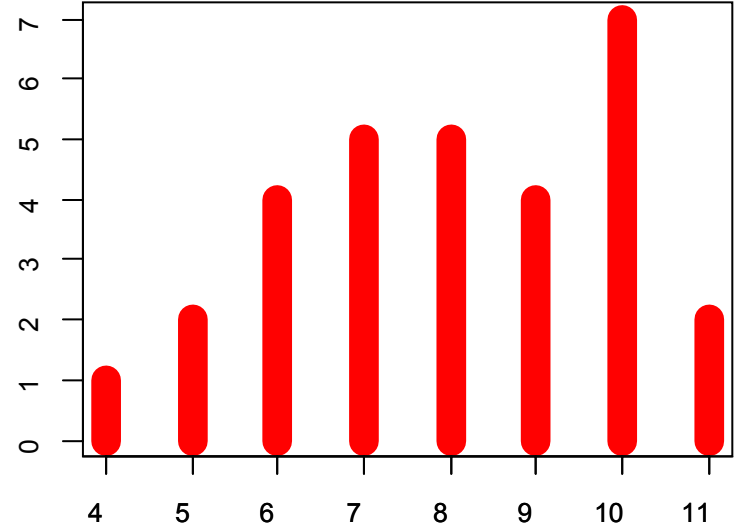
What did happen?

Number of Bags

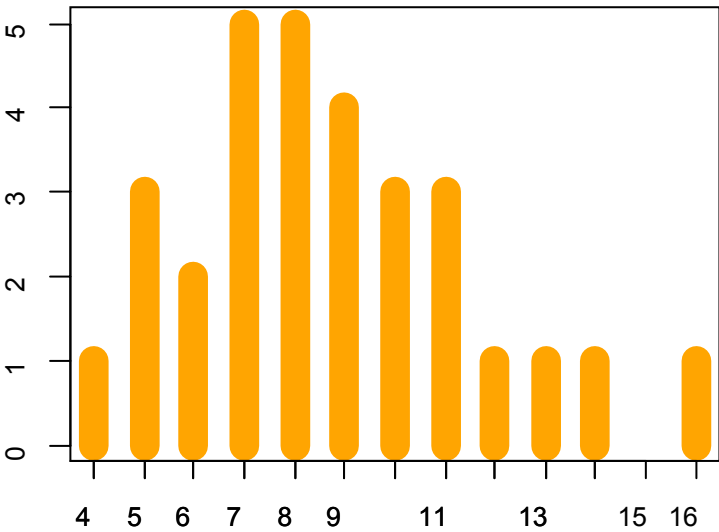




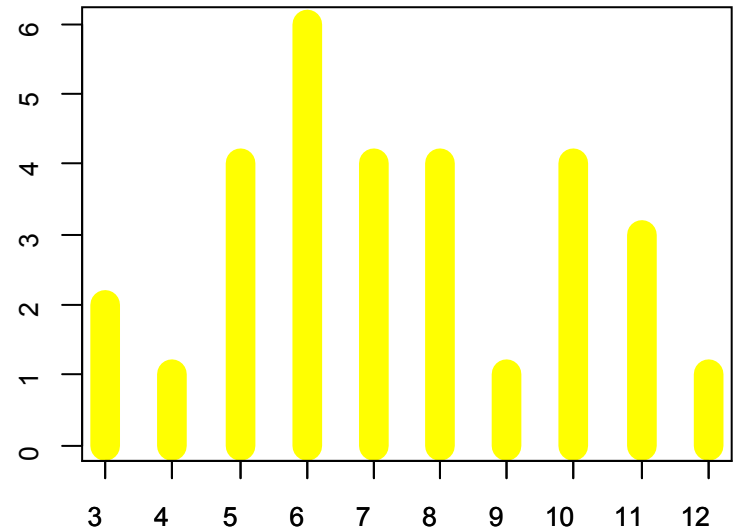
Number of Green Skittles



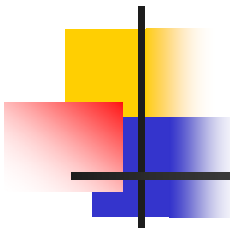
Number of Red Skittles



Number of Orange Skittles



Number of Yellow Skittles

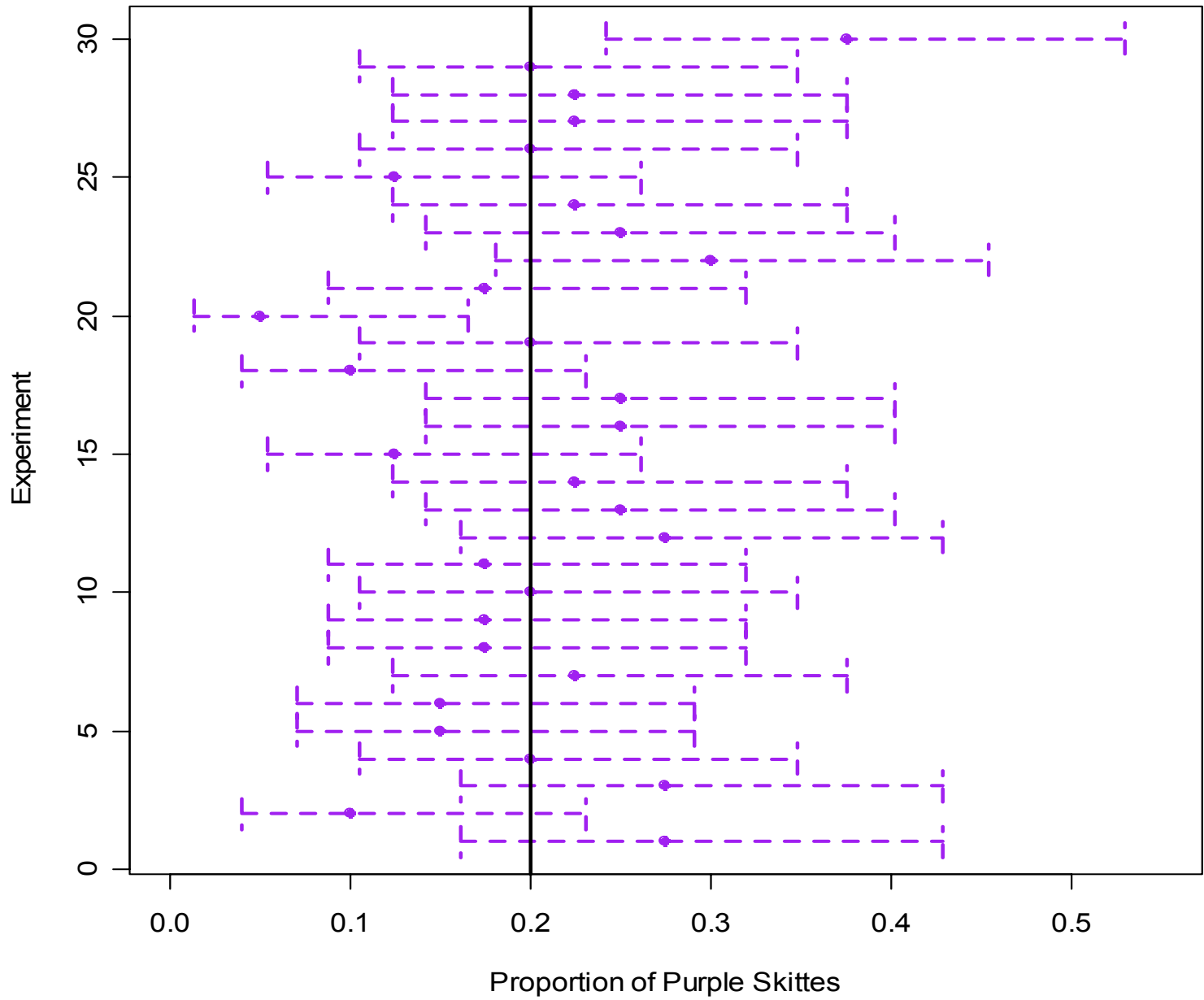


What is the probability that the true proportion is within any given 95% confidence interval?

■ It's in there or it isn't, I don't know!

■ 95%.

■ What are you talking about??





Confidence Interval

- The confidence interval, like the mean and standard deviation of a set of measurements, is a sample statistic.
- The population mean (μ) is fixed, it is the confidence interval that varies from sample to sample.



An ounce of measurement

- Nominal
 - Eye color
- Ordinal
 - Staining intensity
- Interval
 - Degrees C, F
- Ratio
 - Weight

C
O
N
T
I
N
U
O
U
S

**Amount
Of
Information**





Don't move backwards in information

- Dichotomizing a normally distributed predictor is equivalent to losing a third of the data.
- Increases the probability of false positive results
- Impossible to detect non-linearity (nonlinearity is very common)



Backwards

- Can seriously underestimate the extent of variation in risk by concentrating it in two groups
- Residual confounding in a dichotomized variable when used to adjust for the confounding of a continuous variable
- Nonsensical. 2 patients closely, but on either side of the cutoff are characterized as having very different risk when they actually have the same risk.



Show and tell

www.bolderstats.com/gallery/corrReg/medianSplitPlain.html

-Many thanks to Gary McClelland, Department of Psych, University of Colorado, for developing this instructional aid and letting us use it.

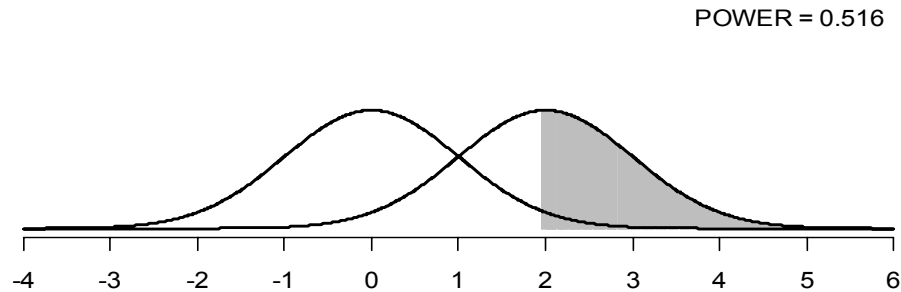


Get more power

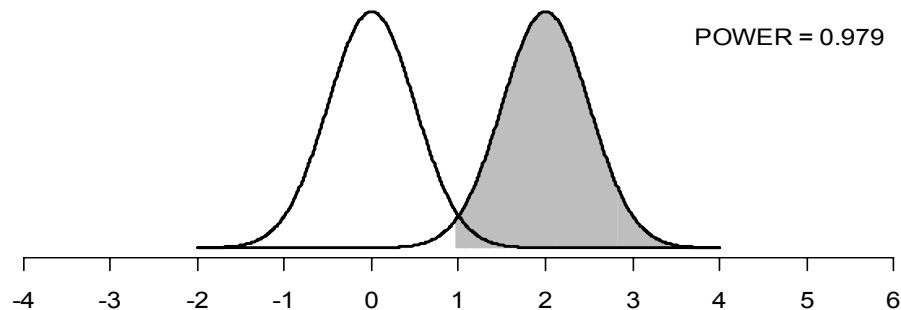
- Assuming a fixed effect size, get more power by,
 - Increasing precision of lab procedures
 - Decrease standard deviation of measurements directly
 - Increasing sample size
 - Decrease standard error of the mean
 - By Design
 - Partition nuisance sources of variation

Get more power by increasing precision

standard deviation = 1



standard deviation = 0.5



Get more power by increasing sample size

