

## Ch. 2 - Measurement & Stats

- Why numbers?
- Precision vs. Accuracy
- Logic and Logical Fallacies
- Descriptive vs. Inferential Statistics
- Measurement Scales
- Central Tendency
- In-class exercise
- Population vs. Sample
- Norms

Psychology 402 - Fall 2010 - Dr. Michael Dohr

52

## Basic Statistics

- Why use numbers?
- Pros:
  - convenient, succinct
  - universal
  - well-defined, repeatable
- Cons:
  - precision vs. accuracy
  - numerical fallacy

Psychology 402 - Fall 2010 - Dr. Michael Dohr

53

## Precision vs. Accuracy

- Precision : the level of detail a measurement is made with, often specified with an error-range
  - “about 6 feet plus or minus 1 foot” vs. “6 foot 11 inches plus or minus 1 inch”
- Accuracy: how close the measured value is to the actual value, does it “hit the target”
  - Think arrow vs. shotgun
- A number can be precise and accurate, precise but inaccurate, or accurate but imprecise.

Psychology 402 - Fall 2010 - Dr. Michael Dohr

54

## Logical Arguments I

- Logical arguments or inferences generally have several components:
  - Premises
  - Conclusions
- Example:
  - Premise: All English people are musicians
  - Premise: John Lennon was English
  - Conclusion: John Lennon was a musician

Psychology 402 - Fall 2010 - Dr. Michael Dohr

55

## Logical Arguments 2

- An Inference can be either Valid or Invalid -- this refers to the Structure of the argument (not the Facts themselves)
  - All A are B  
All C are A  
All C are B
- A Valid inference can still come to a false conclusion, and vice-versa

Psychology 402 - Fall 2010 - Dr. Michael Dohr

56

## Logical Fallacies

- A Logical Fallacy generally means that your inference is Invalid to begin with. In addition, your facts may or may not be true, but the flaw in reasoning has occurred before you even apply facts.
- Example: Affirming the consequent
  - If P, then Q      bank owners are rich
  - Q is true      Bill Gates is rich
  - Therefore P      Bill Gates works at a bank

•

Psychology 402 - Fall 2010 - Dr. Michael Dohr

57

## Nominal and Numeric

- **Nominal Fallacy:** The tendency to believe that something has a name or identification, it exists or has special meaning.  
“I am sleepy” vs. “I am suffering from activity-induced-rest-reduction-performance-impairment syndrome”
- **Numerical Fallacy:** belief that something has been measured and assigned a number, it actually exists.  
“I’m really sad” vs. “I scored a 32 on the Beck Depression Inventory”

Psychology 402 - Fall 2010 - Dr. Michael Dahr

58

## Biased Sample

- Every individual  $x$  that we have seen from sample  $X$  has characteristic  $Z$   
Therefore ALL  $X$  have characteristic  $Z$
- Every student I talk to in this class is interested in Psychology  
Therefore, ALL students are interested in Psychology

Psychology 402 - Fall 2010 - Dr. Michael Dahr

59

## Reification Fallacy

- To Reify - to make something more concrete or real
- Examples:
  - “An A student”
  - “High IQ”
  - “Top of the class”
  - “An F Grade”

Psychology 402 - Fall 2010 - Dr. Michael Dahr

60

## Ranking Fallacy

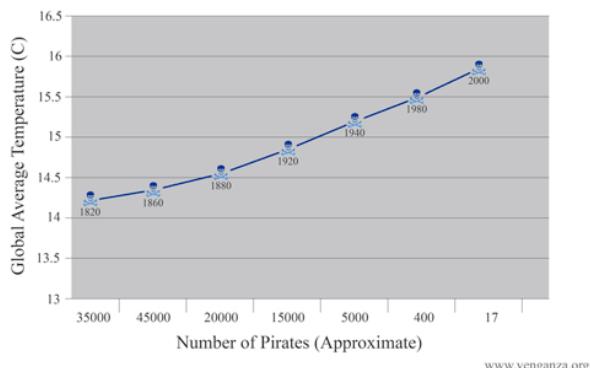
- Reducing a complex phenomenon (e.g. intelligence), giving it a single number (reification) and then ordering based on that number
- Examples:
  - A IQ of 93 is better than an IQ of 90
  - An income of \$50,000 is better than \$45,000

Psychology 402 - Fall 2010 - Dr. Michael Dahr

61

## Correlation = Causation

Global Average Temperature Vs. Number of Pirates

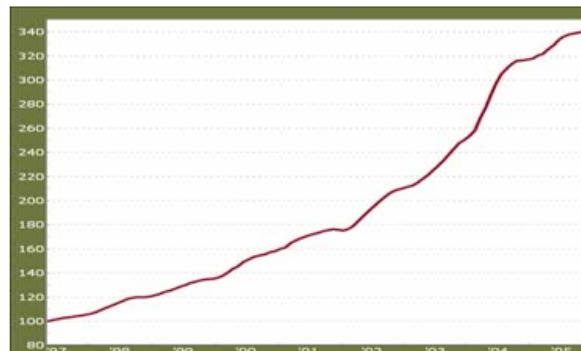


Psychology 402 - Fall 2010 - Dr. Michael Dahr

62

## Hasty Generalization I

S&P/Case-Shiller® San Diego Home Price Index



Psychology 402 - Fall 2010 - Dr. Michael Dahr

63

# Hasty Generalization 2

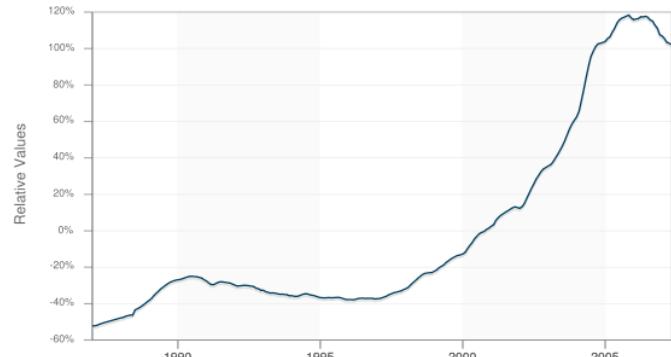
S&P/Case-Shiller® San Diego Home Price Index



Psychology 402 - Fall 2010 - Dr. Michael Dahr

64

# Hasty Generalization 3



<http://www.swivel.com/graphs/show/22856911/graph%255Bscale>

65

## Other Fallacies

- Begging the question -- circular argument
- Correlation implies causation
- Post hoc ergo propter hoc (*after this, therefore because of this*)
- Appeal to Authority
- Ad-hominem
- Straw Man
- False Dilemma

Psychology 402 - Fall 2010 - Dr. Michael Dahr

66

## Types of Statistics

- Descriptive:
  - Goal: help you describe the data
  - reduce the amount of data necessary for understanding
  - don't draw conclusions -- "just the facts"
- Inferential:
  - Goal: draw conclusions from your sample to the larger data set (population)

Psychology 402 - Fall 2010 - Dr. Michael Dahr

67

## Measurement Scales

- Nominal
- Ordinal
- Interval
- Ratio

Psychology 402 - Fall 2010 - Dr. Michael Dahr

68

## Nominal Scale

- Nominal: Name or ID only
  - red, blue, green....
  - john, tony, fred...
  - Sci2-243, Sci2-245...
- does not signify Ordering, Ranking, or More/Less
- Gotcha: even if used with Numbers it may be still a Nominal.
- Example: colors, names, room numbers, ID numbers

Psychology 402 - Fall 2010 - Dr. Michael Dahr

69

## Ordinal Scale

- Ordinal : ordering
- first, second, third.... 1, 2, 3... A, B, C...
- signifies Order, but can't assume distance between items is the same, e.g. the difference between an A and a B may be much different than a B and a C
- Example: Class Rank, Assignment Grade, Product Ratings

## Interval Scale

- Interval: specifies orders AND inter-item distance
- -3, -2, -1, 0, 1, 2, 3.... 100, 105, 115
- the difference between two numbers IS the same, e.g. 100 to 105 should be the same amount as 105 to 110
- Does NOT have an absolute zero.
- Example: temperature in Degrees Farenheit

## Ratio Scale

- Ratio: specifies orders AND inter-item distance and has absolute zero
- 0, 1, 2, 3.... 100, 105, 115
- the difference between two numbers IS the same, e.g. 100 to 105 should be the same amount as 105 to 110
- Does have an absolute zero.
- Example: temperature in Degrees Kelvin

## Measurement Scales

	Magnitude	Equal Intervals	Absolute Zero
Nominal			
Ordinal	✓		
Interval	✓	✓	
Ratio	✓	✓	✓

## Scales: Practical Info

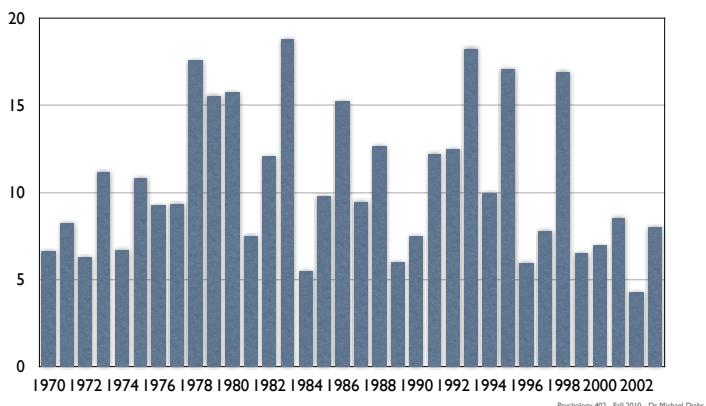
- Nominal Scale: common
  - usable stats: Count & Frequency only
- Ordinal Scale: less common
  - stats: specialized “nonparametric” techniques required
- Ratio and Interval: common
  - Often can be treated identically with same statistical techniques

## Descriptive Statistics

- Count (N)
- Range (minimum, maximum)
- Frequency Distribution (histogram)
- Rank order, percentile (%ile)
- Central Tendency
  - Mean
  - Median
  - Mode
- Variation / Dispersion (Variance, Standard Deviation)

# Data Distributions

■ San Diego Rainfall (inches)



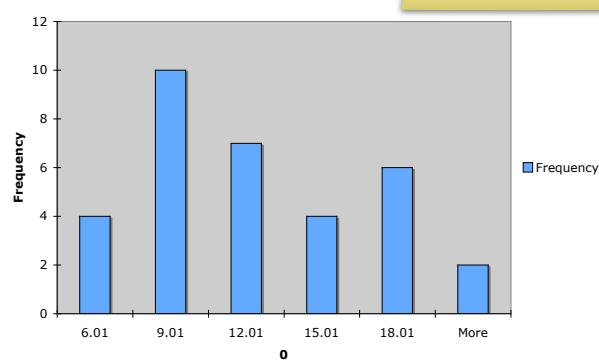
Psychology 402 - Fall 2010 - Dr. Michael Dohr

76

# Data Distributions

Histogram

Note: Textbook version of this graph is wrong!



Psychology 402 - Fall 2010 - Dr. Michael Dohr

77

## Ranks, Percentiles

- Given a distribution of scores, and a single score
- Rank = the item # of the single score when sorted high to low
- Percentile Rank = the % of scores which are lower than the given score
- Percentile = the score at which a given percent of scores are below a given score

Psychology 402 - Fall 2010 - Dr. Michael Dohr

78

## Rank & Percentile

- Infant mortality per 1000 live births
- Sorted low to high

Country	Score
Sweden	2.4
Japan	3.4
France	4.5
USA	7.5
Colombia	20.4
China	37.9
Bolivia	66.4
Ethiopia	142.6
Mozambique	148.6
Zambia	168.1

Psychology 402 - Fall 2010 - Dr. Michael Dohr

79

## Rank & Percentile

- Determine Rank #

Country	Score	Rank
Sweden	2.4	1
Japan	3.4	2
France	4.5	3
USA	7.5	4
Colombia	20.4	5
China	37.9	6
Bolivia	66.4	7
Ethiopia	142.6	8
Mozambique	148.6	9
Zambia	168.1	10

Psychology 402 - Fall 2010 - Dr. Michael Dohr

80

## Percentile Rank

Percentile Rank = # of cases with worse value divided by # of cases  
e.g. France is third of 10 (it has 7 cases with worse values)  
 $7 / 10 = 70\%$

there are different definitions of percentile rank that handle end cases differently

Country	Score	Rank	%ile Rank
Sweden	2.4	1	100
Japan	3.4	2	90
France	4.5	3	80
USA	7.5	4	70
Colombia	20.4	5	60
China	37.9	6	50
Bolivia	66.4	7	40
Ethiopia	142.6	8	30
Mozambique	148.6	9	20
Zambia	168.1	10	10

Psychology 402 - Fall 2010 - Dr. Michael Dohr

81

## Describing Distributions

- Why do this? Large sets of numbers are hard to work with. Easier to reduce dozens, hundreds or thousands of data points down to a few numbers.
- Issue: Any time you reduce the number of data (called “Degrees of freedom” you are throwing away data).
- In essence we are Modeling our data using a simplification.
- “All models are wrong, some models are useful”

Psychology 402 - Fall 2010 - Dr. Michael Dohr

82

## Descriptive stats. problems

- Descriptive statistics rely on certain assumptions. When those assumptions are not met, weird things happen.
- Example: Joe Smith is 6 feet tall, his child is 1 foot tall. Thus, the average height in the Smith household is 3.5 feet.
- If you are sitting in bar, and Bill Gates walks in, suddenly you are (on average), a multi-millionaire.

Psychology 402 - Fall 2010 - Dr. Michael Dohr

83

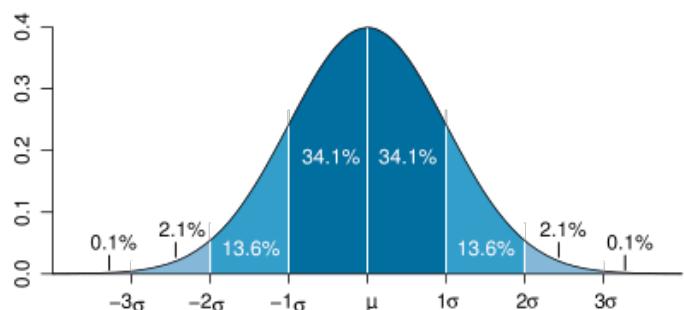
## Central Tendencies

- Assuming the values tend to cluster around a point, what is the point?
- Mean : most common statistic, commonly referred to as the “average”. Formula  $\sum X / N$
- Mode: the most common value in a set
  - rare to use in statistics
- Median: the middle-most value in a set
  - the value at which half are above and half are below. Aka the 50th percentile.

Psychology 402 - Fall 2010 - Dr. Michael Dohr

84

## Normal Distribution

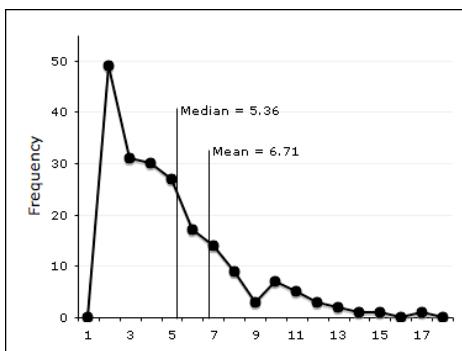


In a normal distribution, the mean, mode, and median are all the same

Psychology 402 - Fall 2010 - Dr. Michael Dohr

85

## Skewed Distribution



In a skewed distribution, the mean, mode, and median are all often different

Psychology 402 - Fall 2010 - Dr. Michael Dohr

86

## Measures of Central Tendency I

	Description	Algorithm	Formula
Mean	the “average”	sum values, divide by N	$\frac{\sum X}{N}$
Median	the “middle-most value”	order values, report middle	50th percentile
Mode	the “most common” value	find most frequent value	...

Psychology 402 - Fall 2010 - Dr. Michael Dohr

87

## Measures of Central Tendency 2

Behavior:	Normal Distribution	Skewed Distribution
Mean	same	overly affected by outliers
Median	same	fairly resistant to outliers
Mode	same	resistant to outliers

Psychology 402 - Fall 2010 - Dr. Michael Dohr

88

## Measures of Dispersion 1

- In addition to measuring the “average” value, we’d like to know how widely dispersed the values are around the average value
- Answers the question “for a typical value, how far away is it from the mean”
- “Difference score” or “residual” can be calculated as the difference between the actual score and the mean. In other words,  $x = X - \bar{X}$
- Then, we can just take the average (mean) of the difference scores.
- Average difference score =  $\text{Sum}(x) / N$

Psychology 402 - Fall 2010 - Dr. Michael Dohr

89

## Variance

- Oops. This “average dispersion” score will **ALWAYS** be equal to zero
- We need another measure.
- It turns out that if you square the residuals before adding them, it removes the negative values. The average of the squared residuals is called Variance ( $S^2$ )
- Variance is nice, but units are now squared. E.g. if you were measuring height in meters, your variance is now in meters<sup>2</sup>
- Desire a more convenient measure...

Psychology 402 - Fall 2010 - Dr. Michael Dohr

90

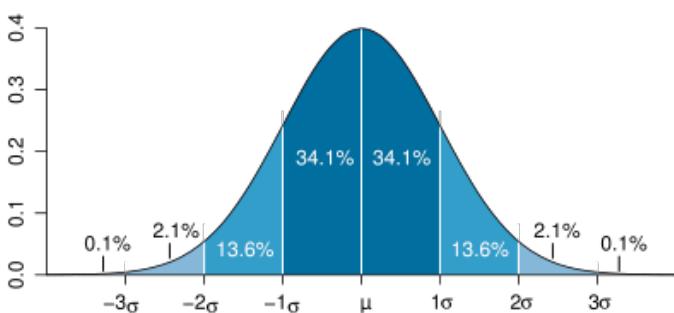
## Standard Deviation

- If you take the square root of Variance ( $S^2$ ) you get  $S$ , which is called “Standard Deviation”.
- Also abbreviated SD, StdDev or  $\sigma$  (Greek sigma)
- SD is often easier to understand because it’s in the same units as your measurement.
- SD also is a unique property of the normal distribution -- given a mean and a SD you have uniquely specified the distribution

Psychology 402 - Fall 2010 - Dr. Michael Dohr

91

## Normal Distribution



In a normal distribution, about 68.2% of values fall within  $\pm 1$  SD

Psychology 402 - Fall 2010 - Dr. Michael Dohr

92

## Exercise: normal distribution

- Everyone flip a coin 10 times and record the # of heads
- Prediction -- on average, one half of should be heads, e.g. 50%
- But will everyone get exactly 5 of 10 heads?
- hint: What is  $N$ ? 10 coin flips, or # of students?
- List and Graph results
- Does the distribution look normal?

Psychology 402 - Fall 2010 - Dr. Michael Dohr

93

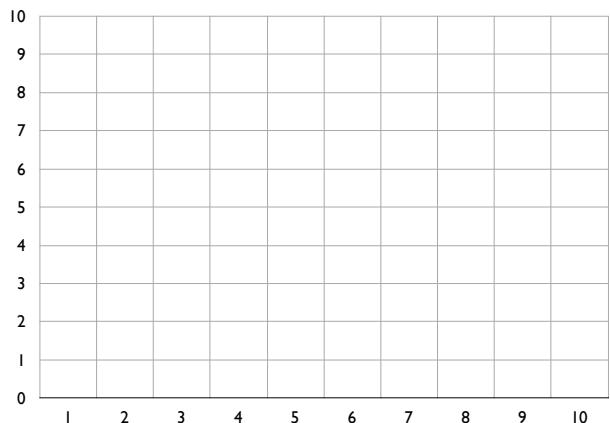
### 10 Coin Flips: Heads=1 Tails=0

X	X - $\bar{X}$	(residual) <sup>2</sup>
$\bar{X}$	$\Sigma$ Residuals	$\Sigma$ (Residuals <sup>2</sup> )/N

Psychology 402 - Fall 2010 - Dr. Michael Dohr

94

### Exercise: Class distribution

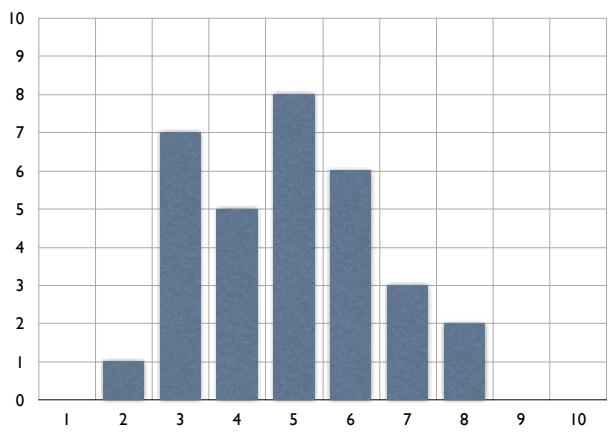


Psychology 402 - Fall 2010 - Dr. Michael Dohr

95

### Exercise: Class distribution

# of Heads out of 10 coin flips



Psychology 402 - Fall 2010 - Dr. Michael Dohr

96

### Exercise: normal distribution 2

- Compute Mean ( $\bar{X}$ ) - is it near 0.5?
- Compute residuals
- Compute sum of residuals -- do they add to zero?
- Compute squared residuals
- Compute Sum of squared residuals (SSR)
- Divide SSR by (N-1) - this is Variance or ( $S^2$ )
- Take square root of variance - this is S or Standard Deviation
- Check -- did 68.2% of results fall within 1 SD?

Psychology 402 - Fall 2010 - Dr. Michael Dohr

97

### Exercise: key points

- Events in the real world usually are distributed in a (nearly) normal distribution
- Assuming a normal distribution, the easiest way to describe the data is by two factors: Mean and SD.

Psychology 402 - Fall 2010 - Dr. Michael Dohr

98

### Population vs. Sample

- In a perfect world, you could take measurements from everybody and there is no need to make any statistical inferences.
- Real world we can't do this, thus we take samples.
- Key to understanding difference between the "true" values in the world and the smaller set of sampled values that you have collected.
- Some statistics are slightly different depending on which group you are using.
- The "Law of Large Numbers", "Sampling Theory", "Central Limit Theorem" makes life easier

Psychology 402 - Fall 2010 - Dr. Michael Dohr

99

## Law of Large Numbers

- Basic gist : if you take enough\* samples, in the long run the sample mean will approach the population mean
- Example: a 6-sided die has possible results 1-6. The average expected result is therefore 3.5.
- We know that if you just roll a die a few times, you are very unlikely to get exactly 3.5 on average.
- However, in the long run, LLN states that you will.

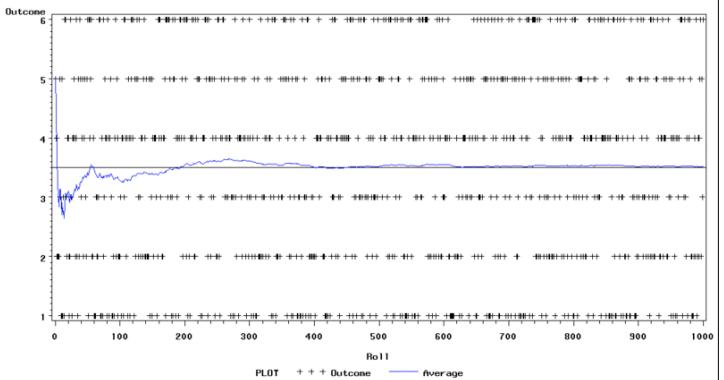
\* what is “enough”? Rule of thumb : 100.

Psychology 402 - Fall 2010 - Dr. Michael Dierh

100

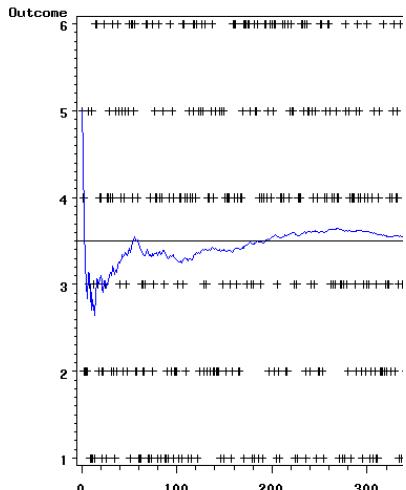
### LAW OF LARGE NUMBERS IN AVERAGE OF DIE ROLLS

AVERAGE CONVERGES TO EXPECTED VALUE OF 3.5



Psychology 402 - Fall 2010 - Dr. Michael Dierh

101



Psychology 402 - Fall 2010 - Dr. Michael Dierh

102

## 9 Heads in a row

- You are flipping a coin, and get 9 heads in a row  
H H H H H H H H H
- What is the % chance the next flip will be a H ?
- Three plausible answers:
  - 50/50
  - more likely Heads
  - more likely Tails

Psychology 402 - Fall 2010 - Dr. Michael Dierh

103

## 9 Heads: Classical Inference

- Coin flips are independent 50/50 events, therefore 50% : Logical/Statistical
- This is the \*correct\* answer for a fair coin

Psychology 402 - Fall 2010 - Dr. Michael Dierh

104

## 9 Heads: Gambler's Fallacy

- Coin flips are independent 50/50 events, but I've seen 9/10 heads, therefore a Tail is "due"
- This is the "Gambler's Fallacy" and the reason Casinos make tons of money. The reasoning is false.
- Note: when dealing with draws w/o replacement, this logic is \*correct\*. For example, a single-card blackjack deck -- if no face cards have come up after 30 cards, then face cards are due

Psychology 402 - Fall 2010 - Dr. Michael Dierh

105

## 9 Heads: Bayesian Statistics

- Coin flips are supposed to be 50/50 events, but I've seen 9/10 heads, therefore the data is telling us that perhaps this is not a fair coin.
- Bayes' theorem suggests you evaluate the prior probabilities in determining future behavior
- In this case, you'd conclude that Head is more likely

## Central Limit Theorem

- Basic gist : no matter what the Population distribution looks like, if you take enough (\*) samples of the mean, the distribution of your samples of the mean will have a Normal distribution
- Central Limit Theorem Exercise (Java)
- Choose a Skewed distribution to show
- This fact makes our life easy: Many statistics assume a normal distribution. The CLT provides us a normal distribution in most cases, even when the population data is skewed

## Population v. Sample Terms

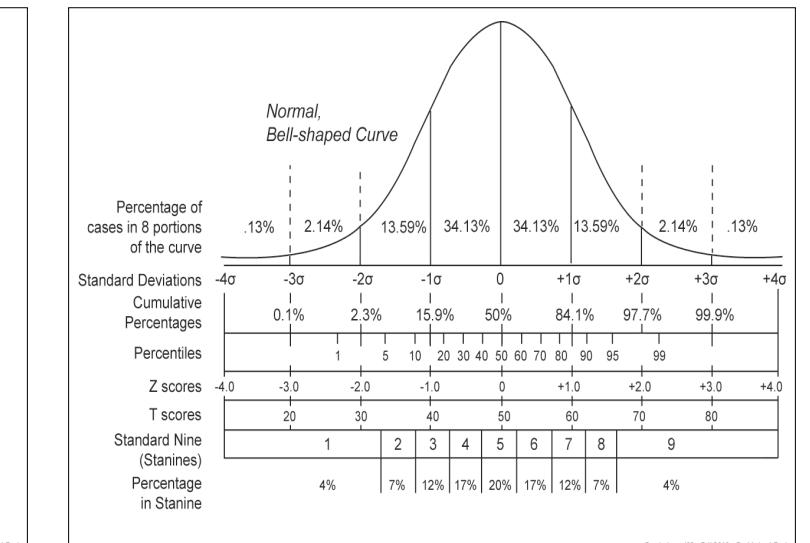
	Population	Sample
Definition	the entire set of items	the actual subset you measured
Descriptives	“Parameters”	“Statistics”
Symbols	Greek	Roman
Mean	$\mu$	$\bar{x}$
Std. Deviation	$\sigma$	$s$
Variance	$\sigma^2$	$s^2$

## Standard Scores

- How do we describe an individual score in comparison to the group?
- We already discussed one method: Ranks & Percentiles
- Another way: raw score differences
- Problem using raw score units: Is a difference of “3” big or little? On a 100 point test it's not very large, but on a 10 point test it's the difference between an A and a C
- Desire a system that is independent of the raw score units, just like letter grades are.

## Standard Scores 2

- Since we already know the mean and standard deviation, we will use those as anchor points
- Express a score as the distance from the mean, but scaled by the standard deviation so that it's independent of raw score unit.
- All standard scores are basically the same, they just differ in terms of the scale applied. Similar to how 1.0 and 100% mean the same thing, one is just expressed in units 100x the other.
- Note: there are many Standard Score systems, some very similar to others (e.g. WIQ vs. SB IQ)



## Standard Scores

	Z scores	IQ scores	T scores	Scaled Scores	
Mean	0.0	100	50	10	
SD	1.0	15	10	3	
Example: top 3%		top 3% is 97 percentile, a Z score of approximately 1.9, IQ score of 128, T score of 19, and scaled score of 15.7			
Example: top 1%		top 1% is about z = 2.4, an IQ score of 136, t score of 74, and scaled score of 17.2 Use Appendix I page 634 for %ile to Z score table			

Psychology 402 - Fall 2010 - Dr. Michael Dohr

112

Z-Score ( $X - \bar{X})/S$	T-Score $10z + 50$	Wechsler IQ ( $15z + 100$ )	Stanford IQ ( $16z + 100$ )	Scaled Score ( $3z + 10$ )	Percentile Rank
3.00	80	145	148	19	99.9
2.90	79	144	146	--	99.8
2.80	78	142	145	--	99.7
2.70	77	141	143	18.1	99.6
2.60	76	139	142	--	99.5
2.50	75	138	140	--	99.4
2.40	74	136	138	--	99.2
2.30	73	135	137	17.2	98.9
2.20	72	133	135	--	98.6
2.10	71	132	134	--	98.2
2.00	70	130	132	16	97.7
1.90	69	129	130	--	97.1

Psychology 402 - Fall 2010 - Dr. Michael Dohr

113

## Norms 1

- Standard Scores provide us with a way of describing how a particular score relates to others in the population.
- In essence, we are describing how a given score relates to other scores, which we presume are “normal”. Hence the term “normative data” and “norms”.
- Key questions: What is the normative group? What features or factors of the group may affect scores?

Psychology 402 - Fall 2010 - Dr. Michael Dohr

114

## Norms 2

- Remember distinction between “norm-referenced” tests and “criterion-referenced” tests.  
Example: an 85 year old in excellent shape could be in the top 5% of his class for firefighting ability, but this may still be a “failing” grade.
- Common factors considered include:
  - in tests of mental ability: Gender, Age, Education, Ethnicity/Race, Language
  - in test of physical ability: (above, plus) handedness, height, weight

Psychology 402 - Fall 2010 - Dr. Michael Dohr

115

## Norms 3

- In ideal world, we'd sample data from people having every possible combination of the factors believed to influence performance.
- Practically this is impossible.
- The influences of these factors should be independent in many cases.
- We measure these influences using “fancy” statistics, creating a model that predicts what a person's theoretical score should be. We can then compare their actual score with their predicted score... See chapter 3.

Psychology 402 - Fall 2010 - Dr. Michael Dohr

116

## Morton's Data as printed

Race	N	Cranial Volume Mean
Caucasian	52	87
Mongolian	10	83
American	144	82
Malay	18	81
Ethiopian	29	78

Psychology 402 - Fall 2010 - Dr. Michael Dohr

117

## Data, corrected

Race	Mean (Morton)	Mean (corrected)
Caucasian	87	87
Mongolian	83	87
American	82	86
Malay	81	85
Ethiopian	78	83

Psychology 402 - Fall 2010 - Dr. Michael Diehr

118

## Seed vs. Shot

Race	Difference (seed - shot)
Caucasian	1.8
Mongolian	n/a
American	2.2
Malay	n/a
Ethiopian	5.4

Psychology 402 - Fall 2010 - Dr. Michael Diehr

119

## Morton's errors

- Fundamental arithmetic errors
- Data selection errors
- Failure to measure or control for external variables (gender, stature, etc.)
- Basic Statistical errors (averaging measurements from unequal size subgroups)
- The racist thumb press?
- Is he a liar? Conscious or subconscious?

Psychology 402 - Fall 2010 - Dr. Michael Diehr

120

## Is skull size related to IQ?

- Skull size is related to IQ, but the relation is not terribly strong
- R is perhaps 0.20 to 0.40
- Thus, the observed 3-4 cubic inch difference between the races would account for, *at most*, a 2-3 point IQ difference\*

\* Measured IQ differences between ethnic groups are actually much higher than 2-3 points -- reasons discussed later.

\* 3 IQ points = about the Flynn effect per decade

Psychology 402 - Fall 2010 - Dr. Michael Diehr

121