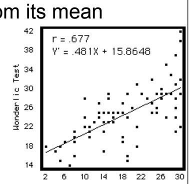
# Correlational Studies of Differences between Means


# Review

- Correlation: relation between variables
  - Focus on relations between two score variables
- Prediction: predict the value of one variable (predicted variable) from the value of another variable (predictor variable)
  - Predict how far a value on one variable differs from the mean of that variable based on how far the value on the other variable differs from its mean
    - Pearson coefficient
  - Prediction based on regression line
    - Regression coefficient
    - Regression constant




# Clicker Question

For the correlation between the average speed a person drives and gas mileage, r = -.80. The correlation accounts for

A. -80% of the variance
B. 80% of the variance
C. 64% of the variance
D. Cannot tell from the information given

## Clicker Question

Which of the following is true if the regression line relating math ability and happiness score is defined by

happiness = 32 – .8 math ability

A. r = -0.8
B. r = 0.64
C. r = 32
D. r is less than 0

## Do Humans Have Abnormally Large Brains?

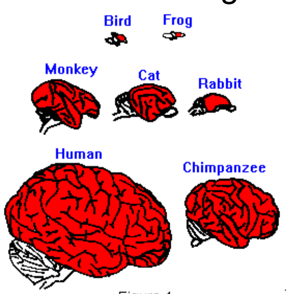
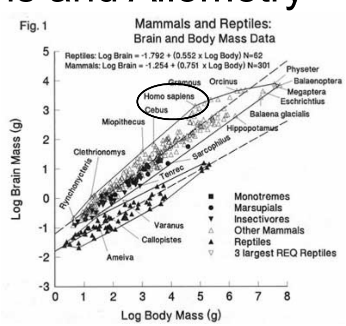


Figure 1

- Comparing the brains of a number of species, humans do seem to have larger brains
  - But hardly the largest!
- But humans also have larger bodies
  - How do brain sizes correlate with body sizes across species?

## Correlations and Allometry

- Allometry correlates the size of parts of organisms (brains) with overall size
- Useful for determining whether the part is unusually larger in a given species
- Human brains only slightly larger than expected



## Correlations in samples and populations

- The interest in correlations typically goes beyond the sample studied—investigators want to know about the broader population.
- Two approaches
  - Estimating correlation in population (ρ) from correlation in sample (r)
    - Confidence interval
  - Determining whether there is a correlation in a given direction in the real population from correlation in sample
    - Statistical significance



## Statistical significance and p-values

- Fundamental question: **How likely is it that the result (correlation in the sample) is due to chance rather than a real correlation in the population?**
- Formally: **How statistically significant is the correlation?**
  - How likely is a given correlation in the sample if there were **no correlation** (or a correlation in the other direction) in the population?
  - This is specified by the **p-value**
    - A p-value < .05 means there is less than a 1 chance in 20 of a correlation in the sample without a correlation in the real population
    - That is, more than 19 times out of 20 the correlation found in the sample is due to a correlation in the real population`



## Statistical significance and p-values

- p-values typically reported as less than some value
  - <.05 is the most commonly used significance level
    - If a study reports that the results are statistically significant with no p value, usually p<.05 is the intended meaning
  - <.01 is a higher, more demanding significance level
    - Less than 1 chance in 100 of getting the result by chance
- For some purposes, lower p values are useful to know
  - Prediction with reliably of only .10 or .25 could be important to know
    - Chemical exposure and cancer, etc.

## Clicker Question

A study reports a negative correlation between cell phone use and age at death with $p<.15$. From this you should conclude

A. There is no correlation between cell phone use and age at death since p is not less than .05
B. There is less than a 15% chance that the correlation is due to chance
C. There is less than a 15% chance of a correlation in the actual population
D. There is at least a 15% chance that the correlation is due to chance

## Significance vs. Importance

- A statistically significant finding may or may not be important.
  - All statistical significance means is that the finding is statistically reliable—not likely to have occurred by chance
    - where the p-value specifies what we count as likely
- Whether it is important—worth knowing—depends on the finding



## Correlations are hard to detect

- Humans are terrible at recognizing intuitively whether two variables are correlated
  - We see correlations where none exist
  - We fail to see correlations that do exist
- Must actually look at the evidence, not rely on our impressions
  - Perform statistical analyses!

# Fallacies of Prediction

1. Seeing correlations that don't exist
2. Failing to recognize regression to the mean
3. Explaining streaks that are to be expected
4. Failing to consider base rates



# Fallacy of Prediction 1: Seeing correlations that don't exist

- "When I'm waiting for the bus, the one going in the other direction always comes first!"
- Evelyn Marie Adams won the New Jersey lottery twice, a 1 in 17 trillion likelihood—seem unlikely?
  - Given the millions of people who buy state lottery tickets, it was practically a sure thing that someone, someday, somewhere would win twice.



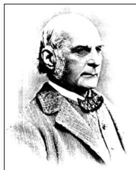
# Coincidences happen

- Adams, Jefferson, and Monroe, three of the first five presidents of the US, died on the same date—July 4!
- Charles Schulz died of a heart attack on the day his last published Peanuts cartoon!

## Fallacy of Prediction 2: Failing to recognize regression to the mean

- Last month you took the SAT/GRE and scored 750 out of a possible 800 on the quantitative part
  - For kicks, you decide to take the test again
    - different questions, but of the same difficulty
    - assume that there was no learning or practice effect from the first test
  - What score should you/we predict for you on the second test?
- The surprising answer is that you are more likely to score **below** 750 than **above** 750
  - the best guess is that you would score about 725

16



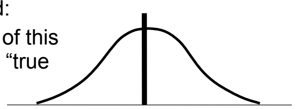
## Regression to the Mean

- Phenomenon discovered by Francis Galton, half cousin of Charles Darwin
- Developed a regression analysis of height between human children and their parents
- Found that "It appeared from these experiments that the offspring did not tend to resemble their parents in size, but always to be more mediocre than they - to be smaller than the parents, if the parents were large; to be larger than the parents, if the parents were small."
  - In fact, this applies only to extreme values

17

## A way to understand regression to the mean

- A given test is really a sample from a distribution. Assume that there is a large number, say 1,000 forms of a test and that
  - you take all 1,000 tests
  - there are no learning, practice, or fatigue effects.
- Scores will be distributed:
  - Identify the mean of this distribution as the "true score"



18

## A way to understand regression to the mean - 2

- Differences in the scores on these tests are due to *chance* factors:
  - guessing
  - knowing more of the answers on some tests than on others.



## A way to understand regression to the mean - 3

- How could a first score of 750 have arisen:
  - It reflected the true score (all chance factors balanced out)
  - Your true score was <750 and you scored above it due to chance factors pushing you up
  - Your true score was >750 and you only scored 750 due to chance factors dragging you down
- Which is more likely?
  - There are very few people with "true" scores above 750 (roughly 6 in 1,000)
  - There are many more people with true scores between 700 and 750 (roughly 17 in 1,000).
  - Thus, it is more likely that you are from the latter group



## A way to understand regression to the mean - 4

Same principle applies to anyone at an edge of the normal distribution

More likely their true score is less different from the mean than the score obtained on a particular occasion when they obtained a very high score

- Baseball player who has a great or horrible batting average one year
- Sales representative who had a spectacular or horrible year

## Clicker Question

Why is it that most players who win “rookie of the year” honors perform less well their second year?

A. By chance, the player performed above his/her natural level in the first year
B. By chance, the player performed below his/her natural level in the second year
C. Opposing players try harder against them
D. The award winners don’t try as hard the next year

## Regression to the mean and punishment

- Makes it seem like punishment works:
  - When someone is doing particularly poorly (for them), chastising them seems to result in better performance
    - But in fact it is only a case of regression
- But praising someone does not seem to work:
  - When someone is doing particularly well (for them), praise is usually followed by poorer results
    - Just another instance of regression!
- “Nature operates in such a way that we often feel punished for rewarding others and rewarded for punishing them” (David Myers, Intuition, p. 148).



## Watch out for pseudo explanations

- A program proposes to help those who score at the very bottom end of a standardized test
  - For example, intervenes with those scoring less than 300 on the SAT
- After the intervention, the individuals are tested again
  - A larger proportion of this group exhibits improved scores than decreased scores
- The program claims success BUT
  - It may have contributed nothing!
  - The results might totally be due to regression to the mean

## Fallacy of Prediction 3: Explaining expected streaks

- 3.1415926535
- THTTTHHTTT

- 3.1415926535 8979323846 2643383279 5028841971
- THTTTHHTTT HTTTTHTHHH HHHTTHTHTT THHHHHHTTTT
- 6939937510 5820974944 5923078164 0628620899
- HTTTTTTTTH THHHTTHTHH TTHTHTHTHH HHHHHHHHTT
  8628034825 3421170679
- HHHHHTHHHT THHTTTHHTT



## Hot hand?

If someone just hit three shots in a row, is it a good idea to pass to them? What if they had missed three in a row?

Philadelphia 76ers' game data from the 1980-81 season (using all shots from the field)—success on next shot

| | |
|---|---|
| **Three Straight Hits** | **.46** |
| **Two Straight Hits** | **.50** |
| **One Hit** | **.51** |
| **One Miss** | **.54** |
| **Two Straight Misses** | **.53** |
| **Three Straight Misses** | **.56** |

***Source: Gilovich, Vallone, and Tversky (1985, Cognitive Psychology, Table 1)***



## Fallacy of Prediction 4: Neglecting base rates

- In trying to make predictions, we very often ignore the most important variable for making a prediction

- Frank was drawn at random from a group of thirty lawyers and seventy engineers. He spends most of his free time hanging around his country club. At the bar he often talks about his regrets at having tried to follow in his esteemed father’s footsteps. The long hours he spent slaving in school could have been better spent learning to be less quarrelsome in his relationships with other people.
  - Is Frank a lawyer or an engineer?

# What to base predictions on?

- Would you answer this one any differently?

- Frank was drawn at random from a group of thirty engineers and seventy lawyers. He spends most of his free time hanging around his country club. At the bar he often talks about his regrets at having tried to follow in his esteemed father's footsteps. The long hours he spent slaving in school could have been better spent learning to be less quarrelsome in his relationships with other people.
  - Is Frank a lawyer or an engineer?

# Clicker Question

In a city in which two cab companies, Blue and Green, operate, a taxicab was involved in a nighttime hit and run accident

- 85% of the cabs in the city are Green, 15% Blue
- A eyewitness identified the cab as Blue
- The Court tested the ability of the witness to identify cab colors under appropriate visibility conditions, and he/she made the correct identification 80% of the time
- What is the probability that the cab involved was Blue?

A. ≈80%
B. ≈60%
C. ≈40%
D. ≈15%

# What to base legal decisions on?

| | Said Blue | Said Green | Totals |
|---|---|---|---|
| **Blue** | 12 | 3 | 15 |
| **Green** | 17 | 68 | 85 |
| **Totals** | 29 | 71 | 100 |

- Of the times he/she said it was Blue, it was blue 12/29 or 41%
- Is <50% accuracy good enough to convict?

# Base Rates

- Base rates are often the best predictor
  - It matters greatly whether the population was 70/30 lawyers or 70/30 engineers
  - It matters greatly that 85% of the cabs were Green
    - This trumps the witness's 80% accuracy!
- But humans almost universally ignore base rates if there is ***anything*** else on which to base the decision
  - Police, lawyers, scientists, doctors . . .
  - Even philosophy professors

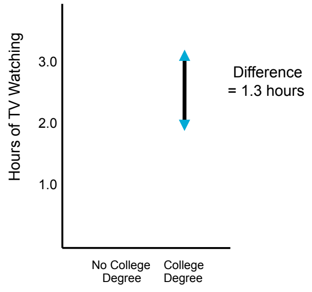
# Comparing two populations

- Populations defined in terms of nominal variables
  - Men/women
  - Gay/straight
  - Taking Phil 12/not taking Phil 12
- Compare the two populations on another variable. If this variable is a score variable, ask:
  - Do the distributions differ?
    - Do the means differ?
    - Do the variances differ? (asked much less often)



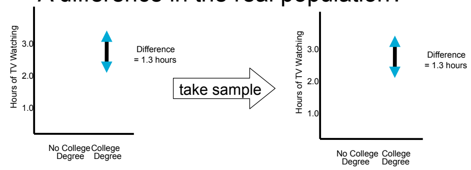
# Diagramming differences between means

- Use bar graph
- Difference between heights of columns reflects differences in means
- When the whole population is tabulated—very straightforward



## Using samples to assess differences between means

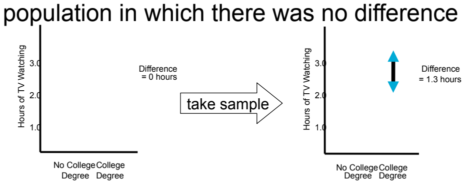
- You take a sample and there is a difference in means
- Where did this difference come from?
  - A difference in the real population?




Real population Sample

## Using samples to assess differences between means - 2

- But it could also arise from a real population in which there was no difference




Real population Sample

- In this case, the result in the sample is due to who happened to get chosen for the sample

## How to tell whether a sample difference is real?

- What is the probability that the difference in the sample could have resulted by chance had there been no difference in the population?
- The hypothesis that there is no difference between the means of the two groups is known as the *null hypothesis*.
  - Strategy: try to reject the null hypothesis
- Conclude that there is a difference in the real population when the sample you get would be very unlikely were the null hypothesis true

## Clicker Question

A null hypothesis

A. Is the claim that there is a difference in the means in the two actual population
B. Is the claim that there is no difference in the means in the two actual populations
C. Is the claim that there is no difference in the means in the two samples
D. Is the claim that the difference in means in the samples is the same as that between the actual populations

## Testing ESP

- Your friend claims to have extrasensory perception—ESP
  - Being a good skeptic, you want to put him to the test
- You use a set of five cards, each randomly presented twice
- You look at and think about the symbol on the card
- Your friend tries to figure out the symbol on the card you are looking at
- You do this ten times, and your friend gets
  - 2 right
  - 3 right
  - 4 right
  - 5 right
- How many does your friend have to get right before you are impressed?

## Testing ESP - 2

Two correct out of 10 trials is the most likely result if the null hypothesis were true

But results of 0, 1, 2, 3, 4 are all quite likely even if the null hypothesis were true

How unlikely a result should we demand?

- How important is it to be right about ESP?

| Number of correct answers | Probabilit y |
|---|---|
| 10 | .00000+ |
| 9 | .00000+ |
| 8 | .00007 |
| 7 | .00079 |
| 6 | .00551 |
| 5 | .02642 |
| 4 | .08808 |
| 3 | .20133 |
| 2 | .30199 |
| 1 | .26844 |
| 0 | .10737 |

## Statistical significance again

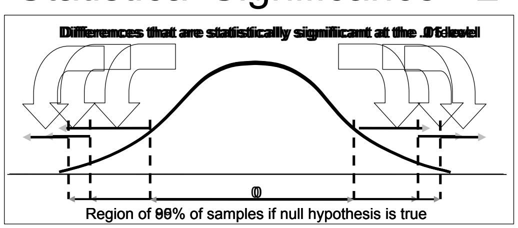
- Just as with correlations between score variables, we use the notion of statistical significance to evaluate results
- A difference in a sample is said to be ***statistically significant*** when it has a very low probability of occurring if the means in the population are equal
  - How low a probability is very low?
  - Investigators have to specify how high a probability they are willing to accept of being wrong
  - For many purposes, scientists accept a 1/20 risk of being wrong—5% ($p < .05$)

## Clicker Question

If it is extremely important not to claim a difference between populations when there isn't one, one should

A. Insist that the difference in the means of the samples be large
B. Not worry about p-values since they aren't important
C. Insist on a p-value <.01 rather than <.05
D. Insist on a  p-value <.1 rather than <.05

## Statistical Significance - 2




If not being wrong when you claim there is a difference  is extremely important, you might require a higher p value ($p < .01$)
If not missing a difference that really exists is really important, you might take note of an even lower significance level ($p < .20$)—although you would want further study

## Testing for Statistical Significance

- There are a number of statistical tests that are employed (depending upon the specifics of the study) to determine whether a difference is statistically significant
- The t-test

$$t = \frac{\text{difference between group means}}{\text{within-group variability}}$$

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{s_1^2/N_1} + \sqrt{s_2^2/N_2}}$$

The t-statistic thus obtained must be compared with a distribution derived from the null hypothesis

If it exceeds that value, the result is significant (at the specified level).

## What has beer taught science?


William Sealey Gosset:

So that future statistics s
(who would surely com in
for his test) couldn't find
published under the nam *nt*

- Trained as a chemist ked at the Guinness brewery in Dublin
  - How to determine, from small samples, which ingredients gave the best results?
- Published anonymously to avoid being accused of giving away trade secrets

## A biological example

Biomass produced by two strains of bacteria

| Bacterium A | Bacterium B |
|---|---|
| 520 | 230 |
| 460 | 270 |
| 500 | 250 |
| 470 | 280 |

Are these differences reliable? t-statistic = 13.01

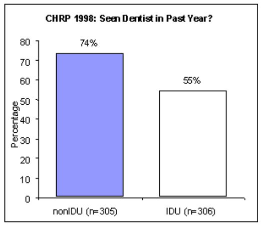
Criterion value for $p<.05$ is 2.45

Criterion value for $p<.001$ is 5.96

Result is significant at $p<.001$

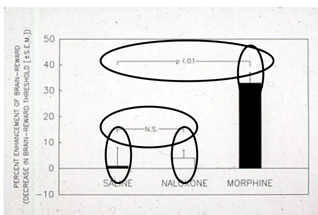
# A Social Science Example

- A sample of intravenous drug users is compared with a sample of non-intravenous drug users
  - How many see a dentist within a year?
  - In this case, p<.001
- It is extremely likely that there is a difference in the actual population
  - although not necessarily exactly the same as the difference in the sample



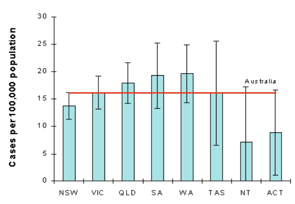

# Showing Statistically Significant Differences with Error Bars

- Error bars can be used to identify 1 or more standard deviations above and below the mean
- If the error bars overlap, the difference is not statistically significant
- If they do not, the difference may be statistically significant




# Showing Statistical (non)-Significance with Error Bars

- The bar graph to the right shows suicide rates of people between 15 and 24 in the different States and territories of Australia
- Error bars show 95% confidence interval
- No differences are statistically significant



## Non-significant Difference versus No Difference

- If the difference in your sample is not significant, you conclude that you cannot tell whether there is actually a difference in the real population
  - There may be one, but the power of your test was too weak to find it
- It is important to keep in mind that we impose a high standard on significance
  - If we use $p<.05$, the result is not likely to happen more than 1 in 20 times by chance
  - If p is only <.1, then the result is typically termed non-significant, but 9 times out of 10 there is a difference in the actual population