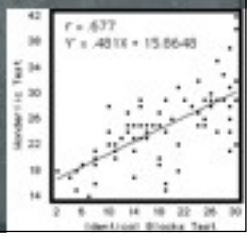
Correlational Studies of Differences between Means

Review - 1

- 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 < 0

Do Humans Have Abnormally Large Brains?

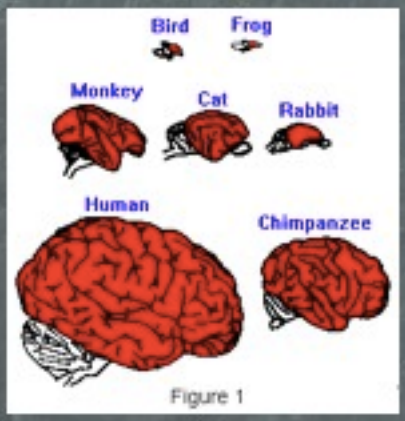
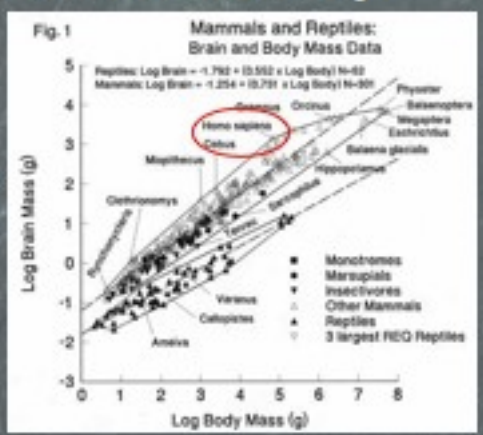


Figure 1

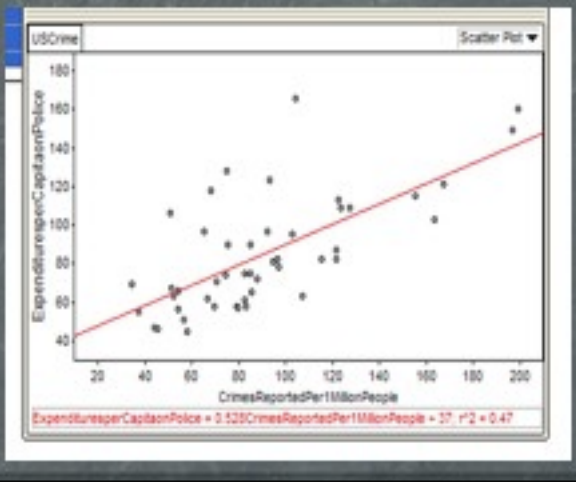
- Comparing the brains of a number of species, humans do seem to have larger brains
  - But hardly the largest!
- But humans also have larger bodies
  - How do brain sizes correlate with body sizes across species?

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



Interpret the Correlation



Review – 2

- You have found a correlation in a sample
  - How likely is it that there is a correlation in that direction in the actual population?
  - What level of statistical significance is enough?
    - $p < .05$: less than 1 chance in 20 that the result is due to chance factors in drawing the sample
    - $p < .01$: less than 1 chance in 100 that the result is due to chance factors in drawing the sample
  - Choose significance level (p value) you require according to how important it is not to be wrong

Clicker Question

For given study that shows a positive correlation between age and interest in the Beatles, $p < .01$. This means

A. There is little correlation between age and interest in the Beatles
B. The correlation is highly important
C. The correlation has low significance
D. There is less than 1 in 100 chance that the correlation in the sample is due to chance

Review – 3

- Statistical significance and importance are very different things
  - Statistical significance simply tells you whether the sample is informative about the real population
- Some surprising results are simply sampling phenomena
  - and so don't require special explanation
- Regression to the mean
  - Extreme scores are likely to be in large part due to chance factors in sampling
    - There is no sophomore slump
- Streaks in samples are expected: TTTHTHHHHTTT

Clicker Question

After taking a SAT review course, those who scored below 250 improved their scores by an average of 30 points. This improvement

A. Must be attributed to the quality of the review course
B. Probably represents the hot hand phenomenon
C. Is likely an instance of regression to the mean
D. Must be due to the students trying harder

Do streaks require explanation?

- 3.1415926535
- THTTTHHTTT

- 3.1415926535 8979323846 2643383279 5028841971
- THTTTHHTTT HTTTTHTHHH HHHTTHTHTT THHHHHTTTT
- 6939937510 5820974944 5923078164 0628620899
- HTTTTTTTH 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)

What to base predictions on?

- 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

Clicker Question

Fred is a 40 something college professor active in social causes and very concerned about the environment. Which of the first two is more likely?

A. Fred is a Republican
B. Fred is a Republican who recycles his trash
C. 1 and 2 are equally likely
D. You cannot tell which is more likely from the information provided

Combining probabilities

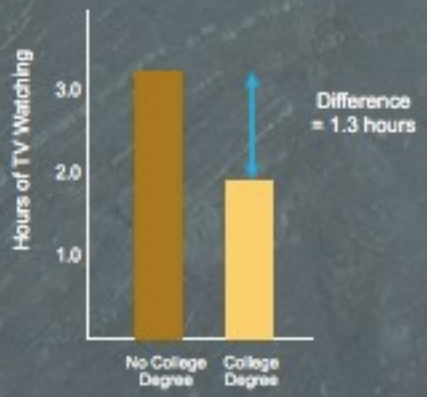
- If two events are independent (the probability of one does not depend on the probability of the other)
  - you determine their joint probability by multiplying
- Since probabilities have values in the range 0 to 1, the product will never be larger than the probability of the more likely event
- If the events are not independent, one has to adjust the formula, but multiplication is still the central operation and the above point still holds

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)

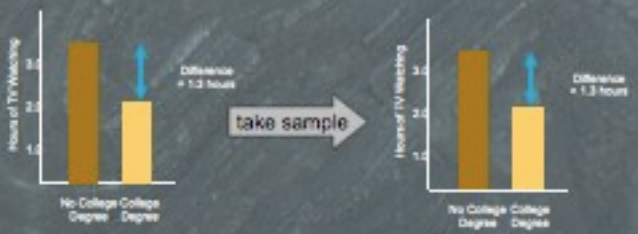
Diagramming differences between means

- 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?

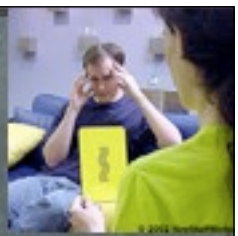
- What is the probability that the difference in the sample could have resulted by chance had there been no difference in the population?
- The hypothesis that there is no difference between the means of the two groups is known as the ***null hypothesis***.
  - **Strategy: try to reject the null hypothesis**
- 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	Probability
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

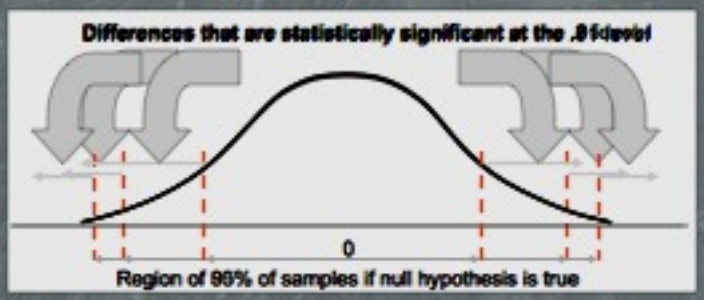
- Just as with correlations between score variables, we use the notion of statistical significance to evaluate results
- A difference in a sample is said to be ***statistically significant*** when it has a very low probability of occurring if the means in the population are equal
  - How low a probability is very low?
  - Investigators have to specify how low 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

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

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

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 stu
(who would surely come
for his test) couldn't find h
published under the name

- Trained as a chemist and t 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

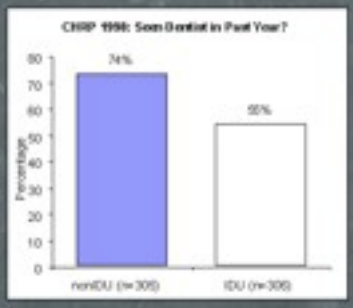
Biomass produced by two strains of bacteria

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

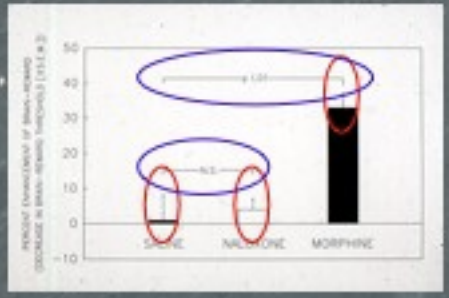
A Social Science Example

- 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



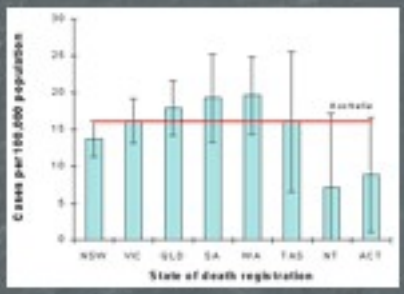
Showing Statistically Significant Differences with Error Bars

- 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, than the result is typically termed *non-significant*, but 9 times out of 10 there is a difference in the actual population

Two dangers

- Type I error (*over confidence*): Thinking there is a difference between means when there is none
  - Use higher significance levels: instead of requiring only $p<.05$, require $p<.01$ or even $p<.001$
- Type II error (*humility*): Thinking there is *no* difference between means when there is one
  - Use a larger sample, which has a greater chance of finding a significant difference if one is to be found

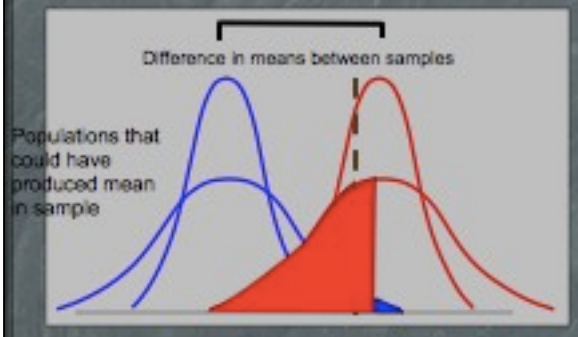
Two dangers - 2

	H_o is true	H_o is false
Did not reject H_o	**Correct failure to reject**	**Type II error (β)**
Did reject H_o	**Type I error (α)**	**Correct rejection**

α and β levels

- α-level is the probability of rejecting the null hypothesis when it is true
  - Statistical significance and p-value
- β-level is the probability of failing to reject the null hypothesis when it is false
  - (1- β) is probability that the researcher will correctly reject the null when the null is indeed false
  - The statistical *power* of the test

Two types of error - 2



Reduce type I error by increasing p-value

Increase Type II error

Increase sample size to reduce Type II error

Clicker Question

Under what conditions should one focus more on reducing type II errors than type I errors?

A. When it is critical not to claim a difference when there isn't one
B. One should always be more concerned with type I errors
C. When it is critical not to miss a difference when there is one
D. When there is little worry about being wrong

Science without Error?

- One can reduce the risk of type I and type II errors to whatever level one desires
  - If one is willing to use a large enough sample
- But one cannot eliminate the risk of error
  - It is always possible that the claim that there is a difference in means is false despite obtaining a significant result in one's sample
  - It is always possible that there is a real difference in means, but the difference in the sample is not significant
- This is one more example of how scientific knowledge remains fallible!