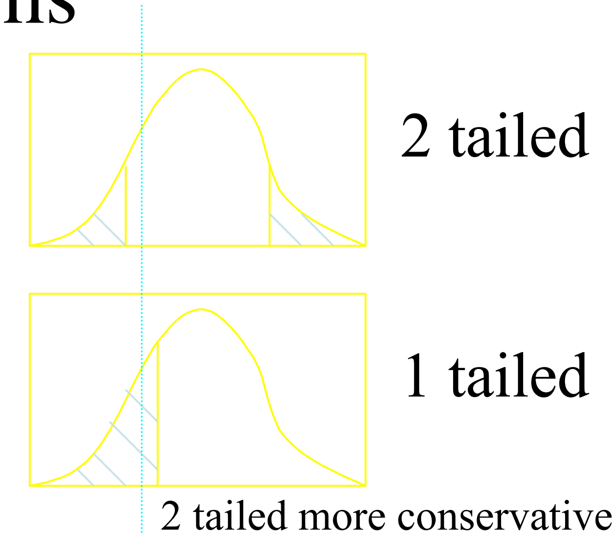
Statistical Notions for Cognitive Science

- Testing causal relations
  - Is X related to Y?
  - Claims of causation, rather than just correlation, come from empirical methods
- Null hypothesis - there is no effect of X on Y
  - Statistical significance as "rejecting null hypothesis"
- Statistical significance
  - p values: probability of result, assuming null hypothesis
  - ($p<.05$) = probability of getting this result if null hypothesis is correct
  - For our class, $p<.2$ counts as significant, and should be interpreted

More Statistical Notions

- One- and two-tailed significance tests
  - One-tailed if Group 1 can only be ≤ Group 2
  - Two-tailed if no apriori requirement about ordering
  - We will always use two-tailed significance tests



- Variability - the heart of statistics
  - Variability in data can be due to variables or noise
  - Reducing variability due to noise increases statistical power - your ability to find relations involving variables
  - Set 1: 2, 5, 5, 5, 5, 5, 8 Set 2: 2, 2, 2, 5, 8, 8, 8 - both have same range
  - Set 2 is more variable

$$Variance = \frac{\sum_{i=1}^{N} \left(X_i - MeanX \right)^2}{N - 1}$$

$$\text{Standard Deviation} = \sqrt{\text{variance}}$$

Set 1 variance = ((2-5)^2+(8-5)^2)/(7-1)= 3

Set 2 variance = (3*(2-5)^2+3*(8-5)^2)/(7-1)=9

More Statistical Notions

- Statistical Power
  - Validity = P(difference | say "significant")
  - Power = P(say "significant" | difference) =probability of rejecting null hypotheses when the null hypothesis should be rejected.
  - Power increases as function SQRT (N), N= sample size
- Continuous and discrete variables
  - Continuous = levels of factor are ordered, quantitatively different
  - Discrete = levels of factor are qualitatively different
  - Continuous: Temperature, GPA, Salary
  - Discrete: Political party, Gender, Class (?), Major
  - Often times, we will turn continuous variables into classes
    - {small, medium, large} {poor, rich} {simple, complex} {abstract, concrete}
  - Kind of variable determines statistical test

Statistical Tests

- T-test
  - Do Groups X and Y have different means?
  - One continuous dependent variable, one discrete independent variable
- Correlation
  - Are values on measures X and Y related to each other?
  - X and Y are both continuous variables
- ANOVA
  - Generalization of T-test
    - More than two levels of an independent variable
    - More than one independent variable
  - Discrete independent variables
- Regression
  - Generalization of correlation
  - Predict Y based on multiple variables X1, X2, X3 …
- All four tests are really same thing: Generalized Linear Model
  - T-test:Anova::Correlation:Regression

T-Test

- Do two groups differ from each other?
  - Dependent variable = continuous
  - Independent variable = discrete. Two categories
  - Paired - can match up data from two groups
  - Unpaired - cannot match up data

Unpaired T-test

$$Tvalue = \frac{MeanX_1 - MeanX_2}{\sqrt{\left(\frac{(N_1 - 1)S_1^2 + (N_2 - 1)S_2^2}{(N_1 + N_2 - 2)}\right)\left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

Difference between groups

Variability within groups

N_1 =number of data in Group 1

S_1^2 =variance of data in Group 1

Unpaired T-Test

IQs at I.U.	IQs at Purdue
122	101
116	97
123	107
125	105
106	107
112	114

Variable	# of Cases	Mean	SD	SE of Mean
IU IQ 1	6	117.3333	7.367	3.007
Purdue IQ 2	6	105.1667	5.8109	2.3723

Mean Difference = 12.1667

T-test for Equality of Means

Variances	t-value	df	2-Tail Sig	SE of Diff	95% CI for Diff
Equal	3.176	10	.010	3.83	(3.63, 20.70)
Unequal	3.176	9.49	.030	3.83	(3.57, 20.76)

In paper: "There was a significance difference between IQs at Purdue (mean = 105) and IU (mean = 117), two-tailed, unpaired T-test, $t(10)=3.176$, $\underline{p} < .05$"

Paired T-Test

- Use if you can match scores across two groups
  - Scores come from the same subject
  - Related subjects: married, twins, same SES
  - Advantages: reduced variability, increase power

Paired T-test

$$Tvalue = \frac{\sum_{i=1}^{N}\left(X_{1,i} - X_{2,i}\right)}{\sqrt{\dfrac{N\sum_{i=1}^{N}\left(X_{1,i} - X_{2,i}\right)^2 - \left(\sum_{i=1}^{N}\left(X_{1,i} - X_{2,i}\right)\right)^2}{N-1}}}$$

Difference between groups

Variability of this difference

N =number of pairs of data

$X_{1,3}$ =The Group 1 score for Pair 3

Paired T-Test

Subject	Morning RT	Evening RT
Ed	430 msec	320
Mary	480	340
Kevin	410	390
Fred	520	510
Barbara	540	480
Susan	480	450

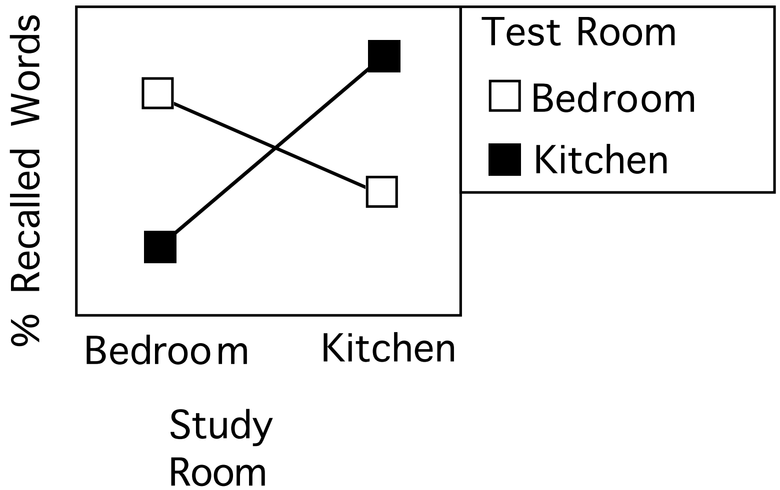
```
T-tests for Paired Samples
               Number of               2-tail
Variable          pairs       Corr      Sig        Mean          SD   SE of Mean
--------------------------------------------------------------------------
Morning                                          476.6667      50.067   20.440
                    6         .735     .096
Evening                                          415.0000      77.136   31.491
--------------------------------------------------------------------------

         Paired Differences              |
   Mean            SD     SE of Mean     |      t-value         df    2-tail Sig
-----------------------------------------|--------------------------------
61.6667      52.694         21.512       |        2.87          5        .035
95% CI (6.368, 116.966)                  |
```

In paper: “Morning RTs (mean=477 msec.) were significantly longer than evening RTs (mean=415 msec.), two-tailed, paired T-test, $t(5)=2.87$, $\underline{p} < .05$

ANOVA (Repeated measures)

More than 1 discrete independent variable

More than 2 levels of an independent variable



Can observe main effects and interactions between variables

Within—groups (hurts finding significant effects) and between-groups variance (helps finding significant effects)

ANOVA

Subject	Study- Kitchen Test-Kitchen	Study – Kitchen Test-Bedroom	Study-Bedroom Test-Kitchen	Study-Bedroom Test-Bedroom
Janice	7	5	4	6
John	8	6	5	7
David	9	7	6	8
Clarice	7	5	4	6
Ben	8	6	5	7
Carol	9	7	6	7
MEAN	**8**	**6**	**5**	**6.83**

Study main effect= (8+6)/2=7 compared to (5+6.83)/2=5.92
Test main effect = (8+5)/2=6.5 compared to (6+6.83)/2 = 6.42
Study-Test interaction = (8+6.83)/2 = 7.42 compared to (6+5)/2 = 5.5
Within-group variance = residual= variability within one column

Repeated measures ANOVA

- Treat the different levels and combinations of all within-subject independent variables as different measures taken from a subject
- Analogous to a paired T-Test: pairing up data by subject. Reduced variability because eliminate individual differences

ANOVA

Subject	Study- Kitchen Test-Kitchen	Study – Kitchen Test-Bedroom	Study-Bedroom Test-Kitchen	Study-Bedroom Test-Bedroom
Janice	7	5	4	6
John	8	6	5	7
David	9	7	6	8
Clarice	7	5	4	6
Ben	8	6	5	7
Carol	9	7	6	7
MEAN	**8**	**6**	**5**	**6.83**

Source of Variation	Sum of Squares	DF	Mean Square	F	Sig of F
Main Effects	6.000	2	3.000	3.750	.041
STUDY	6.000	1	6.000	7.500	.013
TEST	.000	1	.000	.000	1.00
2-Way Interactions	24.000	1	24.000	30.000	.000
STUDY TEST	24.000	1	24.000	30.000	.000
Explained	30.000	3	10.000	12.500	.000
Residual	16.000	20	.800		
Total	46.000	23	2.000		

"There was a significant interaction between study and test context on words recalled, $F(1, 20) = 30.0$, $p < .01$."

Univariate ANOVA (one dependent measure)

Recall	Study	Test
7	kitchen	kitchen
8	kitchen	kitchen
9	kitchen	kitchen
7	kitchen	kitchen
8	kitchen	kitchen
9	kitchen	kitchen
5	kitchen	bedroom
6	kitchen	bedroom
7	kitchen	bedroom
5	kitchen	bedroom
6	kitchen	bedroom
7	kitchen	bedroom
4	bedroom	kitchen
5	bedroom	kitchen
…	…	…

Data are not longer paired by subject

Every datum is independent

Used if between-subject designs, or all data from a single subject

Tests of between-subjects effects

Source	Sum of Squares	DF	Mean Square	F	Sig of F
Intercept	1001.04	1	1001.04	1349.7	.0000
STUDY	7.04	1	7.04	9.49	.006
TEST	.042	1	.042	.056	.815
STUDY * TEST	22.04	1	22.04	29.72	.000
Error	14.83	20	.742		
Total	1045.00	24			

Sum of squares = total variability due to a variable
Error DF = total DF-main effects and interactions, Total DF = number of data, Main effect DFs = # levels – 1
This statistic is less likely to produce significant effects than a repeated measures ANOVA because it does not eliminate individual differences

Correlation

- Is there a relation between X and Y?
  - X and Y are both continuous variables
  - Correlation ranges between -1 and +1. 0 = no relation

$$Correlation = r_{X,Y} = \frac{\left(\dfrac{\sum_{i=1}^{N} X_i Y_i}{N} \right) - MeanX\ MeanY}{S_X S_Y}$$

Relation between groups

Variability of groups

N =number of pairs of data

S_X =variance of Group X data

Correlation

Subject	Quantitative GRE	Verbal GRE
Ed	430	520
Mary	480	540
Kevin	410	590
Fred	520	610
Barbara	540	680
Susan	480	550

```
- -  Correlation Coefficients  - -
             Quantitative   Verbal
Quantitative 1.0000       .6515
              P= .          P= .161

Verbal              .6515    1.0000
              P= .161      P= .
```

Regression: Find best-fitting linear relation between X and Y

$Y = M X + b$, where M is related to $r_{x,y}$

There is a significant relation between quantitative and verbal GRE, Pearson's r = .652, p<.2.

Multiple Regression

- More than 1 continuous independent variable
  - Find best fitting equation: $Y = M_1 X_1 + M_2 X_2 + M_3 X_3 + \ldots + b$
  - Can observe main effects and interactions between variables
- Useful for finding a predicted value for a variable when you know other variables
- ANOVA:T-test::Multiple regression:Correlation

Multiple Regression

Subject	Shoe Size	Height	Weight	IQ
Janice	6.00	155.00 (cm)	108.00	116.00
John	10.00	176.00	160.00	121.00
David	9.00	184.00	181.00	100.00
Clarice	7.00	164.00	125.00	104.00
Ben	11.00	178.00	169.00	112.00
Carol	8.00	168.00	133.00	124.00
Frank	13.00	182.00	190.00	113.00

Variable	B	SE B	Beta	T	Sig T
HEIGHT	-.257632	.172527	-1.124200	-1.493	.2322
IQ	.077414	.042913	.276558	1.804	.1690
WEIGHT	.162732	.058863	2.074275	2.765	.0699
(Constant)	20.0473	22.119504		.906	.4316

← Significance of each variable

Multiple R	.96701	Overall R with all 3 variables
R Square	.93511	Proportion of shoe size variance accounted for
Adjusted R Square	.87023	
Standard Error	.86828	

Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	3	32.59542	10.86514
Residual	3	2.26173	.75391

F = 14.41174 Signif F = .0275 Overall significance with all 3 variables

Shoe size = 20.05 - .258*height + .078*IQ + .163*weight

Chi-square Test of Contingency

- Is there a significant relation between two discrete variables?
  - Are females more likely than males to carry their books like babies?
  - Are Republicans more likely than Democrats to be pro-life?
  - Are religions distributed equally above and below the Mason-Dixon line?

Subject	Gender	Carrying style
Ed	Male	Side
Mary	Female	Baby
Kevin	Male	Side
Fred	Male	Side
Barbara	Female	Baby
Susan	Female	Side
John	Male	Baby
Charlotte	Female	Baby
Karen	Female	Baby

Chi-square Test of Contingency

Summary table of frequencies

	Male	Female
Baby	1 a	4 b
Side	3 c	1 d

$$\chi^2 = \frac{N(ad - bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

$$\chi^2 = \frac{9(1*1 - 3*4)^2}{(1+4)(3+1)(1+3)(4+1)} = 2.72$$

Degrees of Freedom = (Rows - 1)(Columns – 1)

Chi-square Tests

	Value	df	Asymp. Sig
Pearson Chi-Square	2.723	1	.099
Continuity correction	.951	1	.330
Likelihood ratio	2.863	1	.091
N of valid cases	9		

There is a significant relation between gender and book carrying strategy, $\chi^2(1) = 2.72$, $\underline{p}<.2$

Which test would you use?

- T-test
  - Paired
  - Unpaired
- Anova
  - Univariate
  - Repeated Measure
- Correlation
- Multiple Regression
- Chi-Square test of contingency

Which test would you use?

Peggy wants to know whether (20) people are faster with geometric shapes presented on the left side, and words presented on the right side.

- Dependent variable
  - Response Time
  - Continuous
- Two independent variables
  - Presentation side: left or right (discrete)
  - Materials: shapes or words (discrete)
- ANOVA
  - Repeated measures if all people get all conditions
  - One row of data per subject
  - Four columns per row: left shape, left word, right shape, right word
  - Critical test: Is there a Presentation X Material interaction?

Which test would you use?

- David wants to know if people who get more sleep show more improvement at a discrimination task from Day 1 to Day 2?
  - Dependent variable
    - Amount of improvement
    - % Correct Day 2 - % Correct Day 1
    - Continuous
  - Independent variable
    - Amount of sleep
    - Continuous
  - Correlation
    - Critical test: Is the correlation significantly positive?

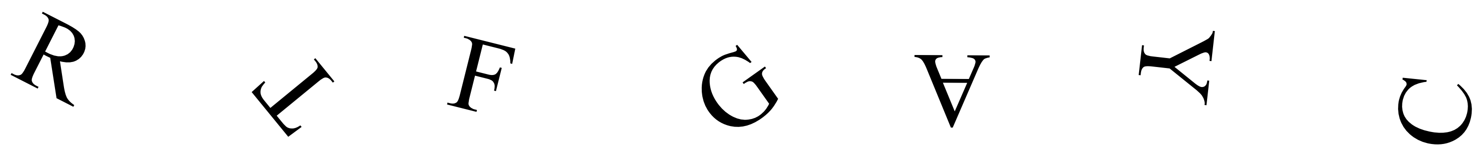
Which test would you use?

- Mozart effect? Is it true that children who listen to Mozart in their cribs learn their first words sooner than those who listen to Nsynch?
  - Dependent variable
    - Age when first word is learned
    - Continuous
  - Independent variable
    - Kind of music exposed to
    - Discrete
  - T-test
    - Can each child be their own control? No
    - Unpaired T-test

Which test would you use?

- Men are from Mars, women are from Venus. Is it true that men are less likely to ask for directions than are women? Observe 80 lost, single drivers at rest areas.
  - Dependent variable
    - Ask for directions? Yes or No
    - Discrete
  - Independent variable
    - Sex of driver
    - Discrete
  - Chi-square test of contingency

Which test would you use?

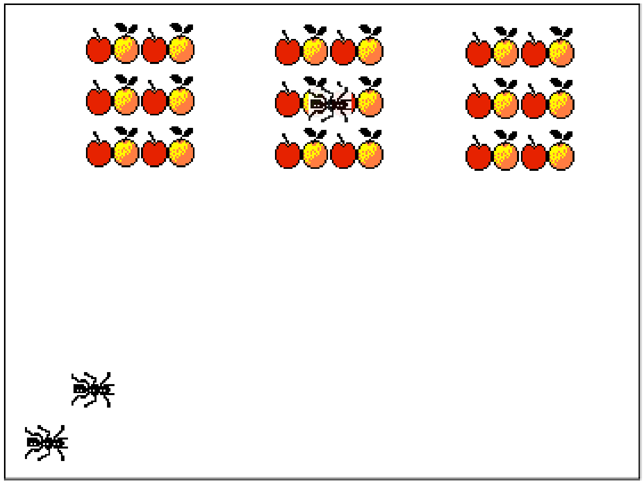


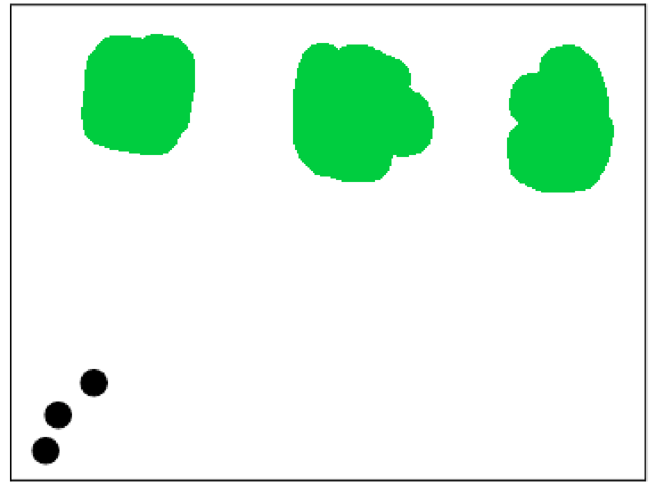
- Does the time required to identify a letter vary as a linear function of the amount of rotation of the letter? If so, what is the best linear model of this relation?
  - Dependent variable
    - Response time to identify letter
    - Continuous
  - Independent variable
    - Amount of rotation of letter
    - Continuous
  - Regression
    - Correlation could also be used, but regression will provide the best fitting linear equation.
    - Critical statistics: What are M and B in the equation: RT = M*rotation+B

Which test would you use?

- Dual code? Is memory better for concrete words than abstract words? Give each of 15 subjects 20 words of each kind to remember.
  - Dependent variable
    - Number of words remembered
    - Continuous
  - Independent variable
    - Kind of word: Abstract or Concrete
    - Discrete
  - T-test - comparing two groups
    - Paired T-test because each subject gives us two numbers.
    - Pair up data by subject. 15 rows of data, 2 columns

Which test would you use?





- Learning styles. Do relatively poor comprehenders show better understanding from concrete or abstract materials?
  - Poor understanders need scaffolding supplied by concrete materials?
  - Poor understanders are distracted by concrete materials because tend to interpret things too literally?
    - This is what Goldstone & Sakamoto (in press) found

Which test would you use?

- Learning styles. Do relatively poor comprehenders show better understanding from concrete or abstract materials?
  - Dependent variable
    - Quiz score: continuous
  - Independent variables
    - Comprehension level: Good or Poor (discrete, but could be continuous)
    - Materials: Abstract or concrete (discrete)
  - ANOVA: Two independent variables
    - Critical test: Is there an interaction between comprehension level and materials on quiz score?
    - Can data be paired up (if so, repeated measures)
      - No, subject is either a good or poor performer, not both
      - Can only give subject one kind of materials
      - So, Univariate ANOVA

Virtual Reality: Realer is better


H
C
F
½
E
MPH
km/h
MI
RPM
X1000
FUEL DOOR
UNLEADED FUEL ONLY

Is realer always better?

Is realer always better?