Overview of Common Statistical Tests

October 11, 2019

Ninet Sinaii, PhD, MPH
Biostatistics & Clinical Epidemiology Service
NIH Clinical Center
sinaiin@cc.nih.gov

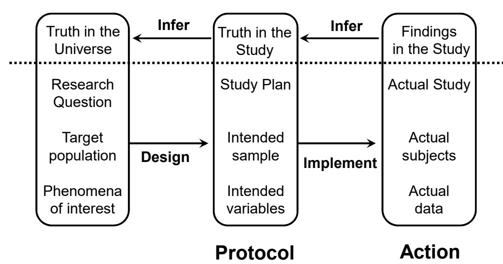
Lecture Overview

- Considerations for the choice of statistical tests
  - Study design and hypothesis, type of data and their distributions
  - Brief review of important statistical features
- Describe basic concepts for common statistical tests
  - Chi-square, paired and two-sample t-tests, ANOVA, correlations, simple and multiple regression, logistic regression, and some non-parametric tests

Please Note

- This lecture covers general concepts behind common statistical procedures to help you better:
  - understand and prepare your data
  - interpret your results and findings
  - design and prepare your studies
  - make sense of results in published literature
- However, consult a statistician:
  - During the planning/design stage of a study
  - For data analysis and interpretation

Our Purpose ;) Sample vs Target Population

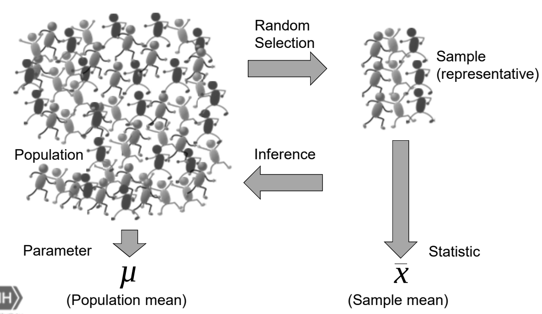


Goal is simple: make the strongest possible conclusions from limited amounts of data. Statistics help us extrapolate from a set of data (sample) to make more general conclusions (population)

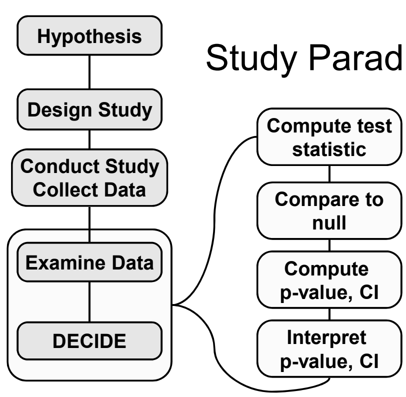
Sources: SG Hilsenbeck, Baylor College of Medicine; Motulsky "Intuitive Biostatistics"

Sample vs Target Population

We want to know about this — We have to work with this



Study Paradigm



Source: SG Hilsenbeck, Baylor College of Medicine

What do medical studies look for?

- Does a specific factor increase or decrease the likelihood of an outcome?
  - Interested in the etiology (cause or causes) of a specific outcome
  - However, it is the association between the exposure and the outcome that is assessed

- To understand
- To compare
- To predict
- To intervene

Measure of Association

- Quantity that expresses the relationship between two (or more) variables
  - Strength: strong, weak, none
    - *How much* association is there?
    - Effect size or the *degree* of association
    - Depends on magnitude and sample size
  - Direction: positive, negative, none

Measure of Association

- Commonly used measures of association include:
  - Differences between means, proportions, rates
  - RR or OR
  - Correlation coefficient
  - Regression coefficient
  - And many others

- Choice depend on the study design and hypothesis, and type of data, and their distributions and assumptions!

Lecture Overview

- What contributes to choice of statistical test?
  - Study design and hypothesis
  - Type of data and their distributions

…and some reviews

Types of Studies

Hypothesis Generating: **Descriptive**
- Case reports
- Case series
- Ecological

Hypothesis Testing: **Analytic**
- Observational
  - Cohort
  - Case-control
  - Cross-sectional
- Experimental
  - Clinical Trials

Hypothesis Testing

- Process of translating a research question to a hypothesis that can be tested
- Involves conducting a test of statistical significance and quantifying the degree that variability may account for the results observed in the study (so we can make a conclusion about the population based on our study sample)

Hypothesis Testing

- Statistical hypothesis testing automates decision making – the primary goal of analyzing data
  - – Examine the evidence provided by the data

Step 1: Requires making an explicit statement of the hypothesis to be tested

- Null hypothesis (H_0):
  there is no association between IV and DV
  eg, $I_E = I_{\bar{E}}$ or $r_p = 0$ or $\mu_x = \mu_y$
- Alternative (H_a):
  there is an association between IV and DV
  eg, $I_E \neq I_{\bar{E}}$ or $r_p \neq 0$ or $\mu_x \neq \mu_y$
  (one-sided/directional: $I_E > I_{\bar{E}}$ or $r_p > 0$ or $\mu_x > \mu_y$)

Question: Is there enough evidence to reject H_0?
We expect (hope) to reject H_0 in favor of H_a.

Step 2: Once H_0 and H_a specified, test of statistical significance can be performed

- Choice of test depends on the hypothesis and type of data → construct a test statistic from our data
- Tests lead to a probability statement or *p*-value
  - – P-value = the probability of obtaining a result as extreme or more extreme than the one observed, *if* H_0 is actually true

P-values

- How do we use p-values in relation to our hypothesis?
  - – if p-value ≤ α (alpha), reject H_0 and conclude statistical compatibility
  - – if p-value > α, cannot reject H_0 and conclude NO statistical compatibility

- Commonly used α values: 0.05, 0.01, or even 0.1 depending on study purpose

P-values: Cautions

- if p-value ≤ α (alpha), reject H_0 and conclude statistical compatibility
  - – means results are surprising and would not commonly occur if H_0 were true
  - – means the number (p-value) we calculated from data is smaller than a threshold we had previously set → that's it!

P-values: Cautions

- if p-value > α, cannot reject H_0 and conclude NO statistical compatibility

Important notes:

- – High p-value does not prove the H_0
- – Deciding not to reject the H_0 is not the same as believing that the H_0 is definitely true: absence of evidence is not evidence of absence
- – "Not statistically compatible" does not mean "no difference"!

P-values

- Important reminders:
  - P-values do not measure the probability that the study hypothesis is *true*
  - Decisions should not be only on whether a p-value passes a specific threshold
  - A p-value or statistical "significance" does not measure the size of an effect or importance of a result
  - By itself, a p-value is not sufficient evidence regarding a study, methodology, or hypothesis
  - Statistical "significance" does not mean clinical importance

There is More…

- P-values must be interpreted in context
  - How firm are the data? Based on many studies?
- How many p-values were calculated?
  - Collected multiple times per subject? Compared many times?
- How large is the discrepancy or difference?
  - Is it clinically relevant?
- What is the sample size?

Sample Size and P-value

- As sample size increases, so does the power of the significance test
  - Larger sample sizes narrow the distribution of the test statistic (hypothesized and true hypothesis become more distinct from one another)
  - Is the observed difference meaningful?
- P-values are not enough to describe a result!
  - Must always also assess the size of the observed difference (effect size)

Statistical Hypothesis Testing & Confidence Intervals

- Hypothesis testing computes a range (95% sure if $\alpha=0.05$) that would contain experimental results if H_0 is true (so any result in this range is not statistically compatible, outside of it is)
- Confidence intervals compute a range (eg, 95% sure) that contains the population value

Based on same statistical theory and assumptions

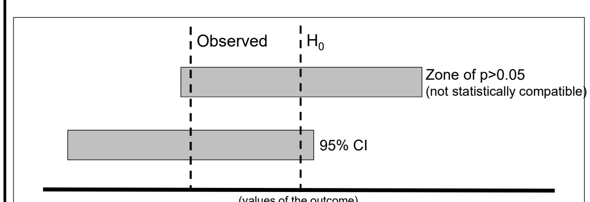
Source: Motulsky "Intuitive Biostatistics"

Statistical Hypothesis Testing & Confidence Intervals

- If 95% CI does not contain the value of the H_0
  → statistically compatible ($p<0.05$)
- If 95% CI contains value of the H_0
  → not statistically compatible ($p>0.05$)

Source: Motulsky "Intuitive Biostatistics"

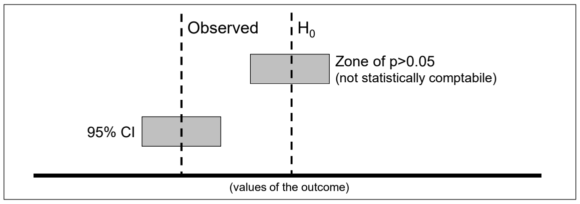
Statistical Hypothesis Testing & Confidence Intervals



What is the conclusion based on this?

Source: Motulsky "Intuitive Biostatistics"

Statistical Hypothesis Testing & Confidence Intervals



What is the conclusion based on this?

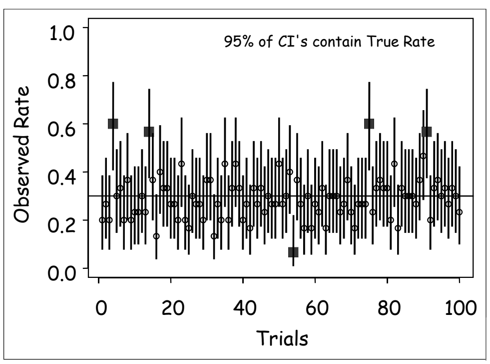
Source: Motulsky "Intuitive Biostatistics"

Thought Experiment:
Catching the real response rate

- Suppose the real response rate for a new therapy is 0.3 (30%)
- Suppose we run a small safety and efficacy clinical trial, and calculate the response rate and a 95% confidence interval for the response rate… over and over and over
- How often will the interval capture the real value?

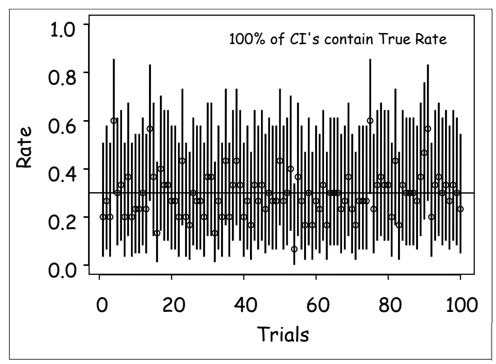
Source: SG Hilsenbeck, Baylor College of Medicine

True Rate=0.3, n=30, Confidence=95%



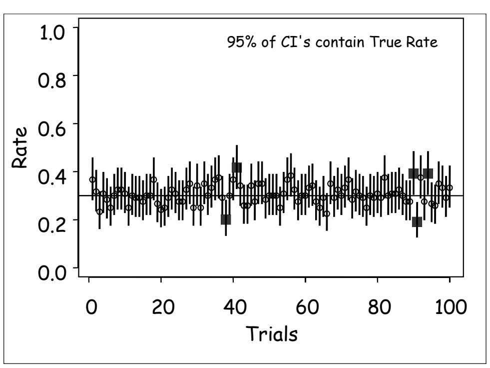
Source: SG Hilsenbeck, Baylor College of Medicine

True Rate=0.3, n=30, Confidence=99.9%



Source: SG Hilsenbeck, Baylor College of Medicine

True Rate=0.3, n=120, Confidence=95%



Source: SG Hilsenbeck, Baylor College of Medicine

Combining Magnitude & Strength of Association

- CIs combine magnitude and strength
- Where applicable, reporting results with 95% CI preferable than p-values alone because they reveal strength, direction, and plausible range of an effect

Lecture Overview

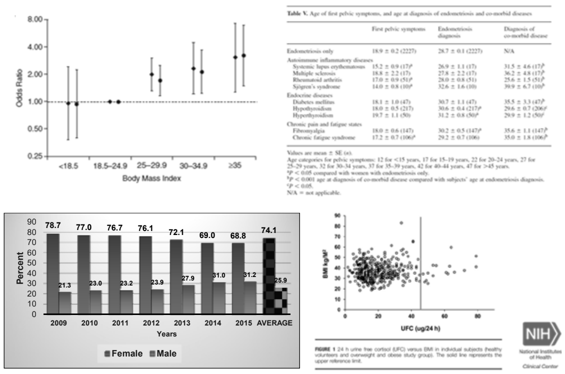
- What contributes to choice of statistical test?
  - – Study design and hypothesis
  - – Type of data and their distributions

…and some reviews

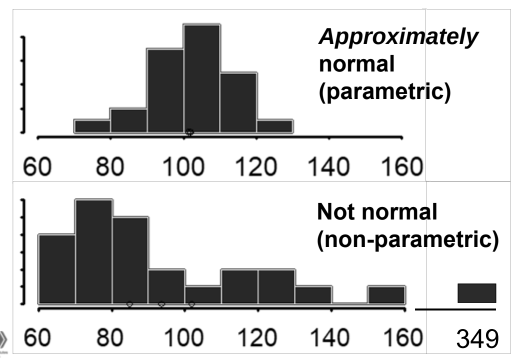
Types of Data

Scale	Examples	Summary Statistics
Continuous (continuum, scale)	age, VAS, HDL, BMI, anxiety score	mean, median, SD, etc
Nominal (Binary or 2+ categories, no ordering, discrete)	gender, race, group, response, treatment	frequency count & percentage, response rate
Ordinal (2+ categories, clear ordering, discrete)	stage, duration, severity, VAS, performance	frequency count & percentage, median

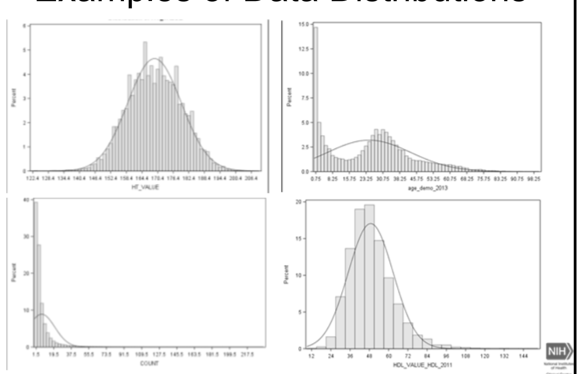
Examples of Types of Reported Data



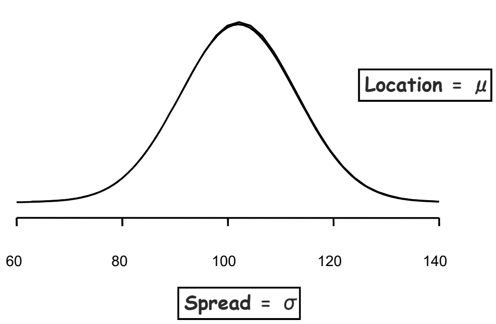
Distribution of Data



Examples of Data Distributions



Normal (Gaussian) Distribution Has Two Parameters

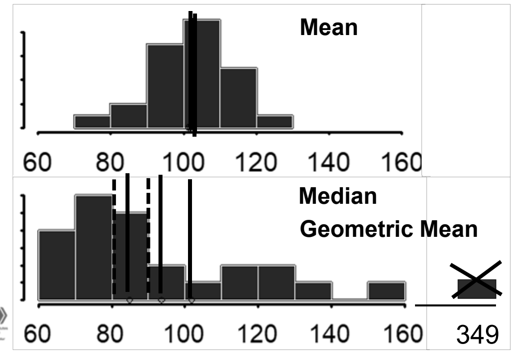


Source: SG Hilsenbeck, Baylor College of Medicine

Summarizing Descriptive Statistics

- Measures of location (in a sample):
  - Arithmetic mean = average (sum of all observations / number of observations)
  - Median = middle value
  - Mode = most frequently occurring value among all observations
  - Geometric mean = average of log transformed values, then taking the antilog (positive integers only)

Features to Summarize Summary Statistics: Middle

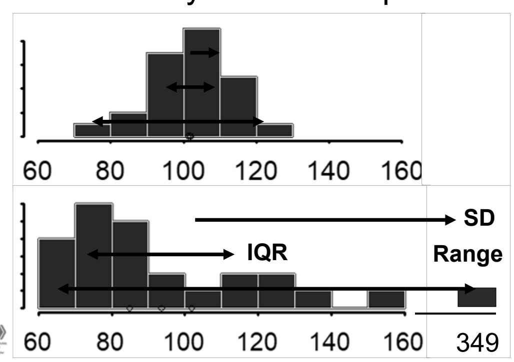


Source: SG Hilsenbeck, Baylor College of Medicine

Summarizing Descriptive Statistics

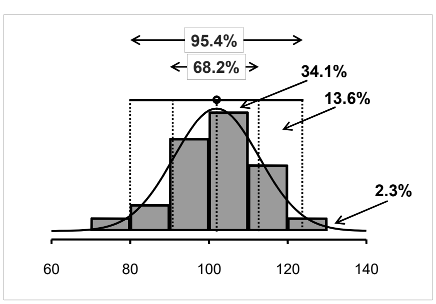
- Measures of spread (in a sample):
  - Standard deviation = most common way to quantify variation
  - Variance = square of standard deviation
  - Quantiles = percentiles (quartiles, quintiles, deciles, etc; 50th percentile = median)
  - Range = minimum and maximum values

Features to Summarize Summary Statistics: Spread



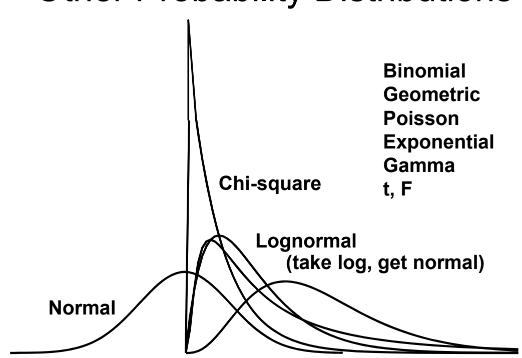
Source: SG Hilsenbeck, Baylor College of Medicine

Sample Data Approximately Normally Distributed



Source: SG Hilsenbeck, Baylor College of Medicine

Other Probability Distributions



Source: SG Hilsenbeck, Baylor College of Medicine

Degrees of Freedom

- Simply, they are the number of independent pieces of information that go into calculating an estimate
  - "the number of values that are free to vary"
  - Not the same as number of items [and vary by test, eg, df=n–1, or $(N_1+N_2)–2$]

Parametric vs Non-parametric

- Parametric tests have certain conditions about the population parameters from which the sample is drawn:
  - Independent observations
  - Drawn from normally distributed populations
  - Populations have the same variances
  - Continuous data, focus on mean difference
- Non-parametric tests do not have conditions about the population parameters:
  - No stringent assumptions about parameters (distribution-free)
  - Apply to data in an ordinal or nominal scale; continuous data changed to orders/ranks/signs
  - Focus on difference between medians

Lecture Overview

- Describe basic concepts for common statistical tests:
  - Chi-square
  - Paired and two-sample t-tests
  - ANOVA
  - Correlations
  - Simple and multiple regression
  - Logistic regression

(includes non-parametric tests)

Scenario

- Study compared the proportion of new breast cancers in Tamoxifen treated and placebo treated women over 5 years.
  - What is the outcome (dependent variable)?
  - What is the independent variable?
  - What types of data are these?

Classification of Common Tests

General Guide

Independent Variable	Dependent (Outcome) Variable: Dichotomous	Nominal (>2)	Continuous (not ~normal), or ordinal (>2)	Continuous (~normal)
Dichotomous	Chi-square			
Nominal (>2)				
Continuous (not ~normal), or ordinal (>2)				
Continuous (~normal)				

Adapted and modified from Hulley and Cummings, 1988

Chi-Square (χ^2) test

- Statistical test commonly applied to sets of categorical data
  - Tests H_0 that frequency distribution of observed events is consistent with a certain theoretical distribution (expected); tests whether unpaired observations are independent of each other
- Common method for 2x2 (rxc) table
  - Caveat: small numbers (expected <5 per cell), better to use Fisher's exact test instead

2x2 Table for Categorical Results

Dependent Variable

		Yes	No	
Independent Variable	Yes	a	b	a + b
	No	c	d	c + d
		a + c	b + d	N

$$\chi^2_c = \sum \frac{(O_i - E_i)^2}{E_i}$$

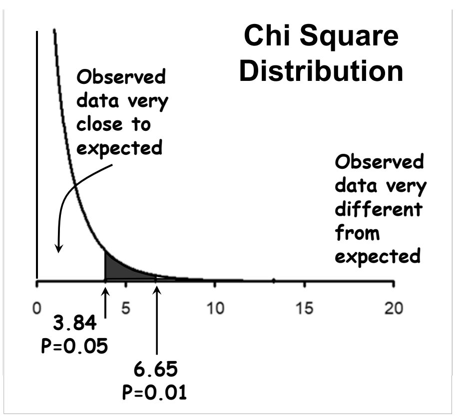
Chi-Square Example

Compare the proportion of new breast cancers in Tamoxifen treated and placebo treated subjects over 5 years?

```
Frequency|
Expected |
Row %    |
Col %    |BRCA    |Dis Free|  Total
---------+--------+--------+
TAM      |      16|     984|   1000
         |      25|     975|
         |    1.6%|   98.4%|
         |   32.0%|   50.5%|
---------+--------+--------+
Placebo  |      34|     966|   1000
         |      25|     975|
         |    3.4%|   96.6%|
         |   68.0%|   49.5%|
---------+--------+--------+
Total          50     1950     2000
```

Test Statistic	DF	Value	P-value
Chi-Square	1	6.65	0.01

Hypothetical data representative of Fisher et al, 1998, JNCI 90:1371-1388



Adapted from: SG Hilsenbeck, Baylor College of Medicine

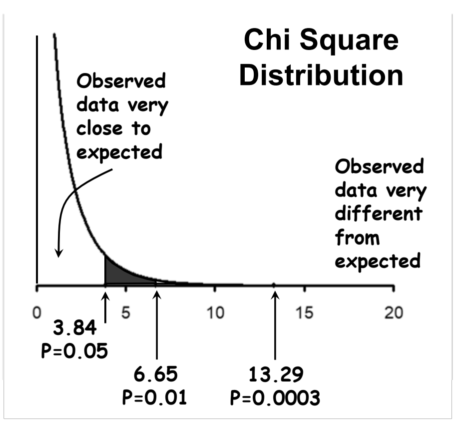
What if we double the sample?

Compare the proportion of new breast cancers in Tamoxifen treated and placebo treated subjects over 5 years?

```
Frequency|
Expected |
Row %    |
Col %    |BRCA    |Dis Free|  Total
---------+--------+--------+
TAM      |      32|    1968|   2000
         |      50|    1950|
         |    1.6%|   98.4%|
         |   32.0%|   50.5%|
---------+--------+--------+
Placebo  |      68|    1932|   2000
         |      50|    1950|
         |    3.4%|   96.6%|
         |   68.0%|   49.5%|
---------+--------+--------+
Total         100     3900     4000
```

Test Statistic	DF	Value	P-value
Chi-Square	1	13.29	0.0003

Hypothetical data representative of Fisher et al, 1998, JNCI 90:1371-1388



Adapted from: SG Hilsenbeck, Baylor College of Medicine

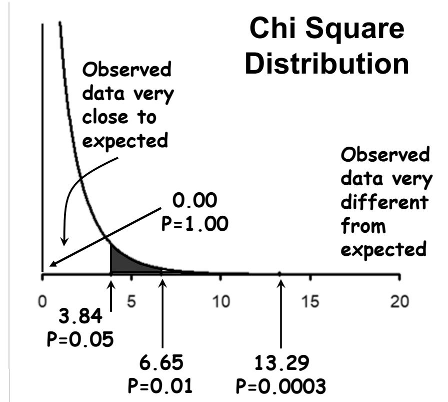
Chi-Square Example

- Smaller sample:
  - BRCA in 1.6% of those on TAM
  - 95% CI: 0.8%, 2.4%
- Larger sample:
  - BRCA in 1.6% of those on TAM
  - 95% CI: 1.1%, 2.1%
- What is the H_0? Are these results statistically compatible at α=0.05?

Chi-Square Example

```
Frequency|
Expected |
Row %    |
Col %    |BRCA    |Dis Free|  Total
---------+--------+--------+
TAM      |      25|     975|   1000
         |      25|     975|
         |    2.5%|   97.5%|
         |   50.0%|   50.0%|
---------+--------+--------+
Placebo  |      25|     975|   1000
         |      25|     975|
         |    2.5%|   97.5%|
         |   50.0%|   50.0%|
---------+--------+--------+
Total          50      1950    2000
```

Test Statistic	DF	Value	P-value
Chi-Square	1	0.00	1.00



Adapted from: SG Hilsenbeck, Baylor College of Medicine

Chi-Square Example

- Smaller sample:
  - BRCA in 1.6% of those on TAM
  - 95% CI: 0.8%, 2.4%
- Larger sample:
  - BRCA in 1.6% of those on TAM
  - 95% CI: 1.1%, 2.1%
- What is the H_0? Are these results statistically compatible at α=0.05?

Classification of Common Tests

General Guide

Independent Variable	Dependent (Outcome) Variable: Dichotomous	Nominal (>2)	Continuous (not ~normal), or ordinal (>2)	Continuous (~normal)
Dichotomous	Chi-square	Chi-square		
Nominal (>2)	Chi-square	Chi-square		
Continuous (not ~normal), or ordinal (>2)				
Continuous (~normal)				

*Fisher's exact test is used anywhere chi-square used when expected cell counts <5

Adapted and modified from Hulley and Cummings, 1988

Paired Nominal Data

- Paired:
  - same subjects, different measures or intervals, pre-post, etc
  - matched pairs
- Use McNemar's test for paired data:
  - Tests H_0 of "marginal homogeneity" – that the two marginal probabilities for each outcome are the same (eg, no treatment effect)
  - Tests change in binary data
- Is a non-parametric test
- Has a chi-square distribution
- For 2x2 table only

2x2 Table for Paired Results

Test 1 \ Test 2	Positive	Negative	
Positive	a	b	a + b
Negative	c	d	c + d
	a + c	b + d	N

$$\chi^2 = \frac{(b-c)^2}{b+c}$$

H_0: $p_b = p_c$ versus H_a: $p_b \neq p_c$

Scenario

- Study compared golf scores for males and females in a PE class.
  - What is the outcome (dependent variable)?
  - What is the independent variable?
  - What types of data are these?

Classification of Common Tests

General Guide

Independent Variable	Dependent (Outcome) Variable: Dichotomous	Nominal (>2)	Continuous (not ~normal), or ordinal (>2)	Continuous (~normal)
Dichotomous				t-test
Nominal (>2)				
Continuous (not ~normal), or ordinal (>2)				
Continuous (~normal)				

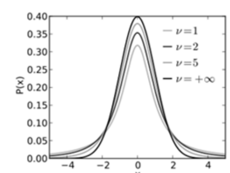
Adapted and modified from Hulley and Cummings, 1988

Student's *t*-test

- Used to test hypotheses about equality of means:
  - One-sample: tests if mean of study sample has a specified value in the null hypothesis (H_0: μ_x=0)
  - Two-sample: tests if means of two study samples are equal (H_0: μ_x=μ_y)
  - Paired: tests if difference between two responses in the same subject (or matched-pairs) has a mean of 0 (H_0: Δ=0)

Student's *t*-test

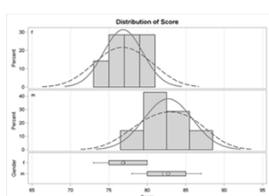
- Follows the t-distribution estimating the mean of normally distributed population
- Assumptions:
  - Independent random samples from two approximately normally distributed populations
  - Variances of the two populations are equal



- Formulas vary by sample and variances

Two Sample *t*-test Example

Compares values from two different groups (if groups are independent, and data are normally/lognormally distributed).
Example: Study compared golf scores for males and females in a PE class.
N = 7, equal in each group
H_0: $\mu_f = \mu_m$ vs H_a: $\mu_f \neq \mu_m$



The TTEST Procedure
Variable: Score

Gender	N	Mean	Std Dev	Std Err	Minimum	Maximum
f	7	76.8571	2.5448	0.9619	73.0000	80.0000
m	7	82.7143	3.1472	1.1895	78.0000	87.0000
Diff (1-2)		-5.8571	2.8619	1.5298		

Source: SAS Institute, Inc

Two Sample *t*-test Example

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
f		76.8571	74.5036	79.2107	2.5448	1.6399	5.6039
m		82.7143	79.8036	85.6249	3.1472	2.0280	6.9303
Diff (1-2)	Pooled	-5.8571	-9.1902	-2.5241	2.8619	2.0522	4.7242
Diff (1-2)	Satterthwaite	-5.8571	-9.2064	-2.5078			

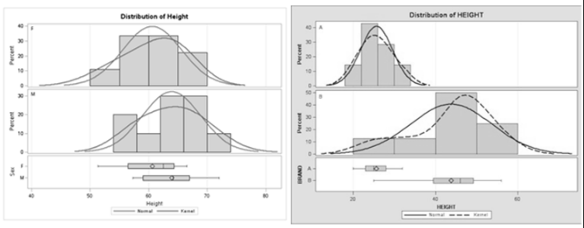
Equality of Variances

Method	Num DF	Den DF	F Value	Pr > F
Folded F	6	6	1.53	0.6189

Note: H_0: variances are equal

Source: SAS Institute, Inc

Two Sample *t*-test Example



Two Sample *t*-test Example

Gender	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
f		76.8571	74.5036	79.2107	2.5448	1.6399	5.6039
m		82.7143	79.8036	85.6249	3.1472	2.0280	6.9303
Diff (1-2)	Pooled	-5.8571	-9.1902	-2.5241	2.8619	2.0522	4.7242
Diff (1-2)	Satterthwaite	-5.8571	-9.2064	-2.5078			

Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	12	-3.83	0.0024
Satterthwaite	Unequal	11.496	-3.83	0.0026

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	6	6	1.53	0.6189

Note: H_0: variances are equal

What is the H_0?
H_0: $\mu_f = \mu_m$ vs H_a: $\mu_f \neq \mu_m$
What about the 95% CIs? Are the results statistically compatible based on them?

Source: SAS Institute, Inc

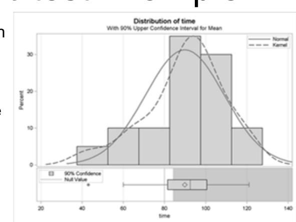
Student's *t*-test

- Used to test hypotheses about equality of means:
  - One-sample: tests if mean of study sample has a specified value in the null hypothesis (H_0: μ_x=0)
  - Two-sample: tests if means of two study samples are equal (H_0: $\mu_x=\mu_y$)
  - Paired: tests if difference between two responses in the same subject (or matched-pairs) has a mean of 0 (H_0: Δ=0)

One Sample *t*-test Example

Compares a sample mean to a given value (not always 0).

Example: Study tested if the mean length of a certain type of court case was more than 80 days.
N = 20 randomly selected cases
H_0: mean=80d vs H_a: mean>80d



The TTEST Procedure
Variable: time

N	Mean	Std Dev	Std Err	Minimum	Maximum
20	89.8500	19.1456	4.2811	43.0000	121.0

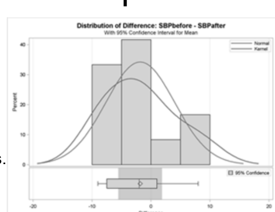
Mean	90% CL Mean		Std Dev	90% CL Std Dev	
89.8500	84.1659	Infty	19.1456	15.2002	26.2374

DF	t Value	Pr > t
19	2.30	0.0164

Source: SAS Institute, Inc
Ref: Huntsberger and Billingsley, 1989

Paired *t*-test Example

Detects differences between means of paired observations.
Natural pairing of data exists, and cannot assume two groups of data are independent. Deltas are assumed to be normally distributed.
Example: Study aimed to determine effect on SBP before and after stimulus.
N = 12; H_0: Δ=0 vs H_1: Δ≠0



The TTEST Procedure
Variable: SBPbefore-SBPafter

N	Mean	Std Dev	Std Err	Minimum	Maximum
12	-1.8333	5.8284	1.6825	-9.0000	8.0000

Mean	95% CL Mean		Std Dev	95% CL Std Dev	
-1.8333	-5.5365	1.8698	5.8284	4.1288	9.8958

DF	t Value	Pr > |t|
11	-1.09	0.2992

Source: SAS Institute, Inc

Scenario

- Study examined the effect of 6 levels of bacteria on the nitrogen content of red clover plants.
  - What is the outcome (dependent variable)?
  - What is the independent variable?
  - What types of data are these?

Classification of Common Tests

General Guide

Independent Variable	Dependent (Outcome) Variable: Dichotomous	Nominal (>2)	Continuous (not ~normal), or ordinal (>2)	Continuous (~normal)
Dichotomous				
Nominal (>2)				ANOVA
Continuous (not ~normal), or ordinal (>2)				
Continuous (~normal)				

Adapted and modified from Hulley and Cummings, 1988

Analysis of Variance (ANOVA)

- Statistical test to determine if the means of several groups are equal (thus, generalizes the *t*-test to 2+ groups)
  - More power, less type I error than multiple two-sample *t*-tests
  - Many consideration terms and classes of models
- Assumptions: independence of observations, normal distribution of residuals, equality of variances

ANOVA Example

Considers one treatment factor with 2+ treatment levels. Goal is to test for differences among the means of the levels and to quantify these differences.

Example: Study examined the effect of bacteria on the nitrogen content of red clover plants. Treatment factor is bacteria strain (6 levels). Red clover plants are inoculated with the treatments. Nitrogen content is measured at the end of the study.

Ref: Erdman (1946); Steel and Torrie (1980).

```
title1 'Nitrogen Content of Red Clover Plants';
data Clover;
   input Strain $ Nitrogen @@;
   datalines;
3DOK1  19.4 3DOK1  32.6 3DOK1  27.0 3DOK1  32.1 3DOK1  33.0
3DOK5  17.7 3DOK5  24.8 3DOK5  27.9 3DOK5  25.2 3DOK5  24.3
3DOK4  17.0 3DOK4  19.4 3DOK4   9.1 3DOK4  11.9 3DOK4  15.8
3DOK7  20.7 3DOK7  21.0 3DOK7  20.5 3DOK7  18.8 3DOK7  18.6
3DOK13 14.3 3DOK13 14.4 3DOK13 11.8 3DOK13 11.6 3DOK13 14.2
COMPOS 17.3 COMPOS 19.4 COMPOS 19.1 COMPOS 16.9 COMPOS 20.8
;
```

Class Level Information

Class	Levels	Values
Strain	6	3DOK1 3DOK13 3DOK4 3DOK5 3DOK7 COMPOS

Number of Observations Read	30
Number of Observations Used	30

Source: SAS Institute, Inc

ANOVA Example

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	847.046667	169.409333	14.37	<.0001
Error	24	282.928000	11.788667		
Corrected Total	29	1129.974667			

Model as a whole accounts for a significant amount of variation in the dependent variable

R-Square	Coeff Var	Root MSE	Nitrogen Mean
0.749616	17.26515	3.433463	19.88667

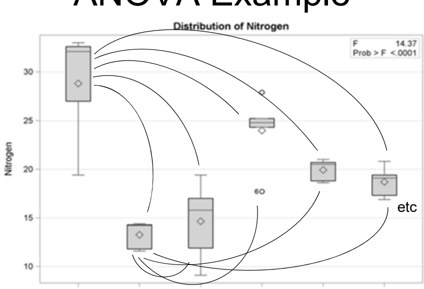
Source	DF	Anova SS	Mean Square	F Value	Pr > F
Strain	5	847.0466667	169.4093333	14.37	<.0001

Note:
F-test is used for comparing the factors of total deviation.
H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

Some contrast between the means for the different strains is different from zero

Source: SAS Institute, Inc

ANOVA Example



Post-hoc tests

What is the H_0?

The F test suggests there are differences among the bacterial strains, but it does not reveal any information about the nature of the differences

What about the 95% CI?

Source: SAS Institute, Inc

Analysis of Variance (ANOVA)

- Terms explained:
  - Balanced design: cells have similar number of observations
  - Blocked design: schedule for conduct of experiment restricted to blocks
  - Factor: independent variables
    - Single vs Multiple:
    - -Example: two-way ANOVA has factors x and y as the main effects and their interaction (xy)
    - -Example: Three-way ANOVA has factors x, y, z as the main effects and their interactions (xy, xz, yz, xyz)
    - -Termed *factorial* when all combinations of levels of each factor are involved.
    - -Note: interactions are complicated and caution is advised!

Scenario

- Study examined the effects of codeine and acupuncture on post-op dental pain.
  - What is the outcome (dependent variable)?
  - What is/are the independent variable(s)?
  - What types of data are these?

ANOVA Example

Example: Study used randomized blocks with two treatment factors occurring in a factorial structure (4 distinct treatment combinations). Examined the effects of codeine and acupuncture on post-op dental pain. The 32 subjects were assigned to eight blocks of four subjects each, based on pain tolerance. Both treatment factors had two levels (codeine: capsule or sugar capsule; acupuncture: two active or two inactive points).

Ref: Neter et al (1990)

blocking variable

two treatment factors

Pain relief (outcome)

```
title1 'Randomized Complete Block With Two Factors';
data PainRelief;
   input PainLevel Codeine Acupuncture Relief @@;
   datalines;
1 1 1 0.0  1 2 1 0.5  1 1 2 0.6  1 2 2 1.2
2 1 1 0.3  2 2 1 0.6  2 1 2 0.7  2 2 2 1.3
3 1 1 0.4  3 2 1 0.8  3 1 2 0.8  3 2 2 1.6
4 1 1 0.4  4 2 1 0.7  4 1 2 0.9  4 2 2 1.5
5 1 1 0.6  5 2 1 1.0  5 1 2 1.5  5 2 2 1.9
6 1 1 0.9  6 2 1 1.4  6 1 2 1.6  6 2 2 2.3
7 1 1 1.0  7 2 1 1.8  7 1 2 1.7  7 2 2 2.1
8 1 1 1.2  8 2 1 1.7  8 1 2 1.6  8 2 2 2.4
```

Source: SAS Institute, Inc

ANOVA Example

Randomized Complete Block With Two Factors

Class Level Information		
Class	Levels	Values
PainLevel	8	1 2 3 4 5 6 7 8
Codeine	2	1 2
Acupuncture	2	1 2

Number of Observations Read	32
Number of Observations Used	32

Model as a whole accounts for a significant amount of variation in the dependent variable

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	10	11.33500000	1.13350000	78.37	<.0001
Error	21	0.30375000	0.01446429		
Corrected Total	31	11.63875000			

R-Square	Coeff Var	Root MSE	Relief Mean
0.973902	10.40152	0.120268	1.156250

Note:
F-test is used for comparing the factors of total deviation.
H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

Source: SAS Institute, Inc

ANOVA Example

Dependent Variable: Rlief

Source	DF	Anova SS	Mean Square	F Value	Pr > F
PainLevel	7	5.59875000	0.79982143	55.30	<.0001
Codeine	1	2.31125000	2.31125000	159.79	<.0001
Acupuncture	1	3.38000000	3.38000000	233.68	<.0001
Codeine*Acupuncture	1	0.04500000	0.04500000	3.11	0.0923

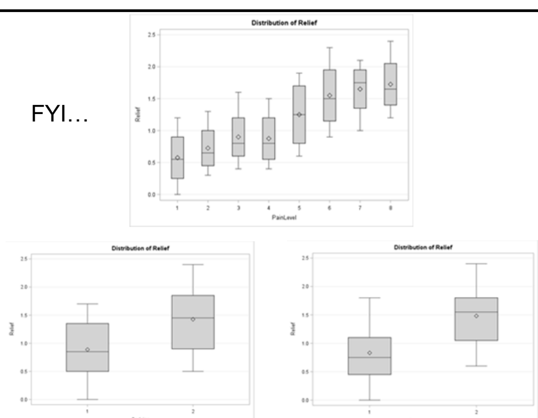
- What is the H_0?

Note:
F-test is used for comparing the factors of total deviation.
H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

- What about 95% CIs

Source: SAS Institute, Inc

FYI…



Source: SAS Institute, Inc

Scenario

- Study aimed to determine how well a child's height and weight are related.
  - What is the outcome (dependent variable)?
  - What is the independent variable?
  - What types of data are these?

Classification of Common Tests

General Guide

Independent Variable	Dependent (Outcome) Variable: Dichotomous	Nominal (>2)	Continuous (not ~normal), or ordinal (>2)	Continuous (~normal)
Dichotomous				
Nominal (>2)				
Continuous (not ~normal), or ordinal (>2)				
Continuous (~normal)				Correlation, Linear regression

Adapted and modified from Hulley and Cummings, 1988

Correlation

- Used to test hypotheses about how two continuous variables are related
  - How much of the *variability* in one is explained by the other
  - Sample correlation coefficient (r) is estimated, ranging between -1 and 1, that quantifies the *direction* and *strength* of the linear association
  - Assumes normality
  - Pearson's correlation coefficient between random variables x and y:

$$\rho_{X,Y} = \mathrm{corr}(X,Y) = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}$$

Classification of Common Tests

General Guide

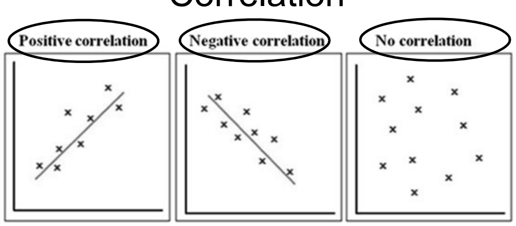
Independent Variable	Dependent (Outcome) Variable: Dichotomous	Nominal (>2)	Continuous (not ~normal), or ordinal (>2)	Continuous (~normal)
Dichotomous				
Nominal (>2)				
Continuous (not ~normal), or ordinal (>2)			Spearman rank correlation	Spearman rank correlation
Continuous (~normal)			Spearman rank correlation	Correlation, Linear regression

Adapted and modified from Hulley and Cummings, 1988

Non-parametric Correlation

- Tests:
  - Spearman's rank correlation coefficient
  - Kendall tau rank correlation coefficient
- No linear relationship requirement
- Measures how one variable changes as the other changes
  - If one increases and so does the other, rank is positive
  - If one increases, and the other decreases, rank is negative
- Interpretation is essentially the same

Correlation



The points lie close to a straight line, which has a positive gradient.

This shows that as one variable **increases** the other **increases**.

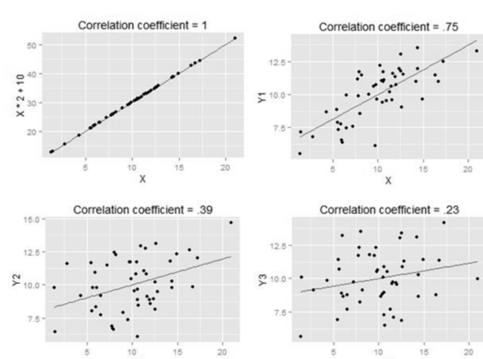
The points lie close to a straight line, which has a negative gradient.

This shows that as one variable **increases**, the other **decreases**.

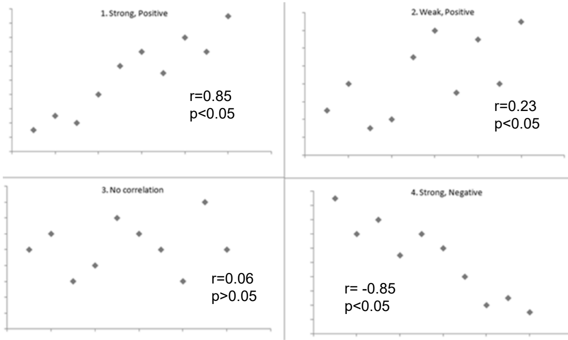
There is no pattern to the points.

This shows that there is **no connection** between the two variables.

Correlation Example



Correlation Example

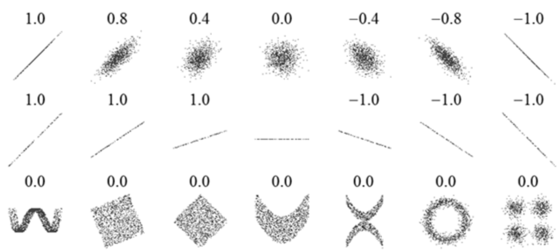


Correlation Example

- What is the H_0?
- What about 95% CIs?

- CIs are very helpful!
- For example:
  - r_p=0.87, 95% CI: 0.76, 0.94
    - p<0.0001
  - r_p=0.08, 95% CI: -0.12, 0.27
    - p=0.43

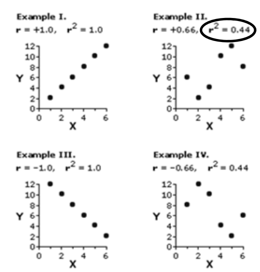
Correlation: Caution



Pictures are worth a thousand words ;)

Correlation and Regression

- Correlation analysis is related to regression analysis (which assesses the relation between an outcome variable and *one or more* covariates)



Regression

- Type depends on the number and types of dependent and independent variables
  - Continuous DV: Simple, or multiple (2+ predictors/explanatory/independent variables)
  - Categorical DV: Logistic

Linear Regression

- A quantitative dependent variable y and one or more explanatory/independent variables x

$$y_i = \beta_0 1 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^\top \boldsymbol\beta + \varepsilon_i, \qquad i = 1, \ldots, n,$$

- Assumptions:
  - Linearity
  - Constant variance
  - Independent errors
  - Lack of multicollinearity

Linear Regression

- Other important features explained:
  - Multiple regression is also known as multivariable linear regression
    Note: *Multivariate* linear regression refers to models where multiple correlated DVs are predicted
  - Generalized linear models used for dependent variables that are bounded or discrete (eg, skewed, Poisson distribution/count data, ordinal data)

Scenario

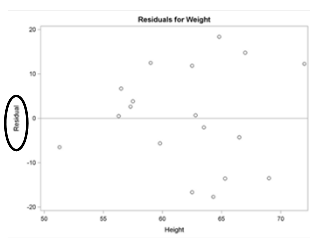
- Study aimed to determine how well a child's weight can be predicted if the child's height is known.
  - What is the outcome (dependent variable)?
  - What is the independent variable?
  - What types of data are these?

Linear Regression Example

Example: Use regression analysis to determine how well a child's weight can be predicted if child's height is known.

N = 19 children; height and weight were measured.

Equation of interest is:
Weight = $\beta_0 + \beta_1$*height + ε



Where weight is the response/dependent variable
β are unknown parameters
Height is the regression/independent variable
ε is unknown error

Source: SAS Institute, Inc

Linear Regression Example

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	7193.24912	7193.24912	57.08	<.0001
Error	17	2142.48772	126.02869		
Corrected Total	18	9335.73684			

Root MSE	11.22625	R-Square	0.7705
Dependent Mean	100.02632	Adj R-Sq	0.7570
Coeff Var	11.22330		

r^2 — 77% of variability in Y explained by variability in X

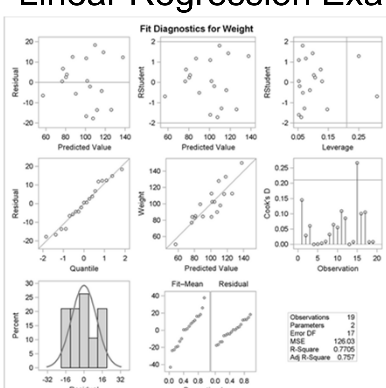
Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-143.02692	32.27459	-4.43	0.0004
Height	1	3.89903	0.51609	7.55	<.0001

From the parameter estimates, the fitted model is:
Weight = -143.0 + 3.9*height

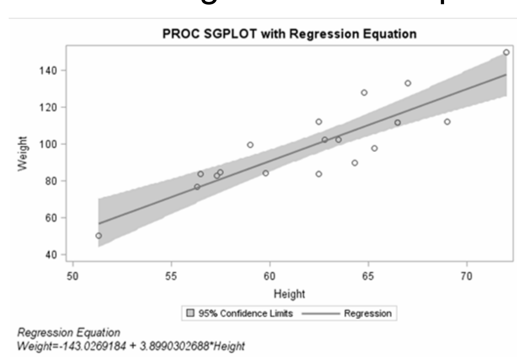
Source: SAS Institute, Inc

Linear Regression Example



- What is the H_0?
- What about 95% CIs?

Linear Regression Example



Source: SAS Institute, Inc

Scenario

- Study investigated the role of confusion and use of certain medications to predict falls in the elderly.
  - What is the outcome (dependent variable)?
  - What is/are the independent variable(s)?
  - What types of data are these?

Classification of Common Tests

General Guide

Independent Variable	Dependent (Outcome) Variable: Dichotomous	Nominal (>2)	Continuous (not ~normal), or ordinal (>2)	Continuous (~normal)
Dichotomous				
Nominal (>2)				
Continuous (not ~normal), or ordinal (>2)				
Continuous (~normal)	Logistic regression			

Adapted and modified from Hulley and Cummings, 1988

Logistic Regression

- Regression model where the dependent variable is categorical (outcome achieved vs not: treatment successful vs failed; alive vs dead; recovered vs not recovered)
  - Multinomial logistic regression (more than two outcome variables, or ordinal (ordered) outcomes
  - Conditional logistic regression (matched studies)
  - 1 or more predictor variables (mixed types)

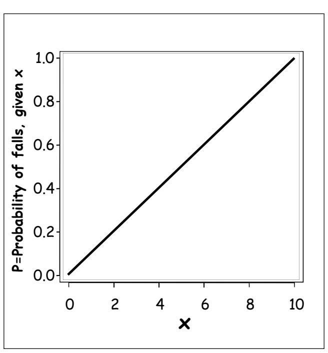
Classification of Common Tests

General Guide

Independent Variable	Dependent (Outcome) Variable: Dichotomous	Nominal (>2)	Continuous (not ~normal), or ordinal (>2)	Continuous (~normal)
Dichotomous				
Nominal (>2)		Chi-square, Logistic Regression		
Continuous (not ~normal), or ordinal (>2)				
Continuous (~normal)		(Multinomial) Logistic regression		

Adapted and modified from Hulley and Cummings, 1988

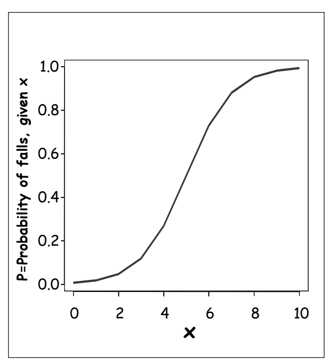
Logistic Regression



- Used to predict whether a subject has a given outcome (ie, disease), based on observed characteristics of the subject (eg, age, gender, clinical characteristics, etc)
- Assume linear regression

Source: SG Hilsenbeck, Baylor College of Medicine

Logistic Regression

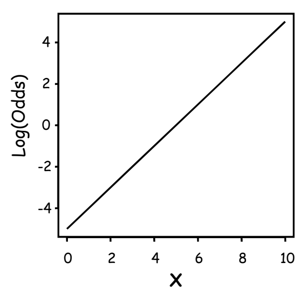


- Assumptions violated (residuals not normally distributed)
- Logit transformation:
  - Takes odds of event happening at different levels of each IV
  - Then takes ratio of those odds
  - Then takes the log of those ratios
- Creates continuous criterion as a transformed version of the DV

Source: SG Hilsenbeck, Baylor College of Medicine

Logistic Regression

Logit of outcome (success) then fitted to predictors using linear regression analysis



$$p = \frac{e^{a+bx}}{1+e^{a+bx}}$$

$$\frac{p}{1-p} = e^{a+b*x}$$

$$\log(\frac{p}{1-p}) = a + b*x$$

$$\log(Odds) = a + b*x$$

Source: SG Hilsenbeck, Baylor College of Medicine

Logistic Regression Example

- Study investigating:
  - outcome of falls (yes vs no)
  - using predictors of confusion and use of certain medications (and other considerations)
    - Note: Examples of univariable results are shown here

Chi Square Test for Independence

```
         Obs Freq |
         Exp      |
         Diff     |          Group
         Row Pct  |
         Col Pct  |Cases    |Controls|  Total
         ---------+---------+--------+
          No      |     96  |    170 |    266
                  |    114  |    152 |
Confusion         |  27-18  |  48 18 |  76.00
         ---------+---------+--------+
         Yes      |     54  |     30 |     84
                  |     36  |     48 |
                  |  15.18  |   8-18 |  24.00
         ---------+---------+--------+
         Total         150       200      350
                     42.86     57.14   100.00

         Statistic        DF      Value      Prob
         ------------------------------------------
         Chi-Square        1     20.7237    <.0001
```

Source: SG Hilsenbeck, Baylor College of Medicine

Logistic Regression

$$\log(Odds) = a + b*(confusion) + c*(drugs) + \ldots$$

$$Odds = e^{(a+b*(confusion)+c*(drugs)+\ldots)}$$

$$OR = \frac{Odds(confusion = 1)}{Odds(confusion = 0)}$$

$$= \frac{e^{a+b*1+c*drugs+\ldots}}{e^{a+b*0+c*drugs+\ldots}}$$

$$= e^{(a+b+c*drugs+\ldots)-(a+c*drugs+\ldots)}$$

$$= e^{b}$$

Source: SG Hilsenbeck, Baylor College of Medicine

```
       Model Fitting Information and Testing Global Null Hypothesis BETA=0
                                     Intercept
                         Intercept      and
       Criterion           Only      Covariates    Chi-Square for Covariates
       AIC               480.036       461.389           .
       SC                483.894       469.105           .
       -2 LOG L          478.036       457.389        20.647 with 1 DF (p=0.0001)
       Score                 .             .          20.724 with 1 DF (p=0.0001)
                          Analysis of Maximum Likelihood Estimates
                    Parameter   Standard       Wald          Pr >      Standardized
 Variable    DF      Estimate      Error    Chi-Square   Chi-Square      Estimate
 INTERCPT    1        -0.5715     0.1277       20.0353       0.0001              .
 CONFUSN     1         1.1592     0.2611       19.7185       0.0001       0.273348
                    Conditional Odds Ratios and 95% Confidence Intervals
                                                          Wald
                                                    Confidence Limits
                                           Odds
                    Variable        Unit        Ratio        Lower       Upper
                    CONFUSN       1.0000        3.187        1.911       5.317
```

- What is the H_0?
- What about 95% CI?

Source: SG Hilsenbeck, Baylor College of Medicine

Chi Square Test for Independence

```
Confusion      Group

Obs Freq |
Exp      |
Diff     |
Row Pct  |
Col Pct  |Cases    |Controls|  Total
---------+---------+--------+
 No      |     96  |    170 |    266
         |    114  |    152 |
         |  27-18  |  48 18 |  76.00
         |  36.09  |  63.91 |
         |  64.00  |  85.00 |
---------+---------+--------+
Yes      |     54  |     30 |     84
         |     36  |     48 |
         |  15.18  |   8-18 |  24.00
         |  64.29  |  35.71 |
         |  36.00  |  15.00 |
---------+---------+--------+
Total         150       200      350
            42.86     57.14   100.00
         Statistic        DF      Value      Prob
         ------------------------------------------
         Chi-Square        1     20.7237    <.0001
```

OR = 3.19
95% CI: 1.91, 5.32

Is this consistent with chi-square test results? Is it statistically compatible at α=0.05?

Source: SG Hilsenbeck, Baylor College of Medicine

Classification of Common Tests

General Guide

Independent Variable	Dependent (Outcome) Variable: Dichotomous	Nominal (>2)	Continuous (not ~normal), or ordinal (>2)	Continuous (~normal)
Dichotomous			Wilcoxon rank sum	
Nominal (>2)			Kruskal-Wallis	
Continuous (not ~normal), or ordinal (>2)	Wilcoxon rank sum	Kruskal-Wallis	Spearman rank correlation	Spearman rank correlation
Continuous (~normal)			Spearman rank correlation	

Adapted and modified from Hulley and Cummings, 1988

Other Non-Parametric Tests

- Wilcoxon rank-sum tests (same as Mann-Whitney *U* test)→ two-sample *t*-test
- Wilcoxon signed-rank test → paired *t*-test
- Kruskal-Wallis → ANOVA (or singly-ordered contingency table)
  - Jonckheere-Terpstra for doubly-ordered data
- Spearman's correlation/Kendall's tau → Pearson's correlation

Classification of Common Tests

General Guide

Independent Variable	Dependent (Outcome) Variable: Dichotomous	Nominal (>2)	Continuous (not ~normal), or ordinal (>2)	Continuous (~normal)
Dichotomous	Chi-square	Chi-square	Wilcoxon rank sum	t-test
Nominal (>2)	Chi-square	Chi-square	Kruskal-Wallis	ANOVA
Continuous (not ~normal), or ordinal (>2)	Wilcoxon rank sum	Kruskal-Wallis	Spearman rank correlation	Spearman rank correlation
Continuous (~normal)	Logistic regression	Logistic regression	Spearman rank correlation	Correlation, Linear regression

Adapted and modified from Hulley and Cummings, 1988

Summary

- General concepts were covered in this lecture to help you understand your/published data and results a little better, and help interpret findings.
- However, consult a statistician:
  - During the planning/design stage of a study
  - For data analysis and interpretation

Take Home Message

- "Focusing on whether or not a result is statistically compatible can get in the way of good science"
- When reading the scientific literature, don't let a conclusion about statistical significance stop you from trying to understand what the data really show: *the point of statistics is to quantify scientific evidence and uncertainty*

Source: Motulsky "Intuitive Biostatistics"

Questions?

LET'S HOPE LIFE ON MARS IS MORE INTELLIGENT THAN LIFE ON EARTH.