Normality, Hypothesis Tests, and Inference

D. Alex Hughes

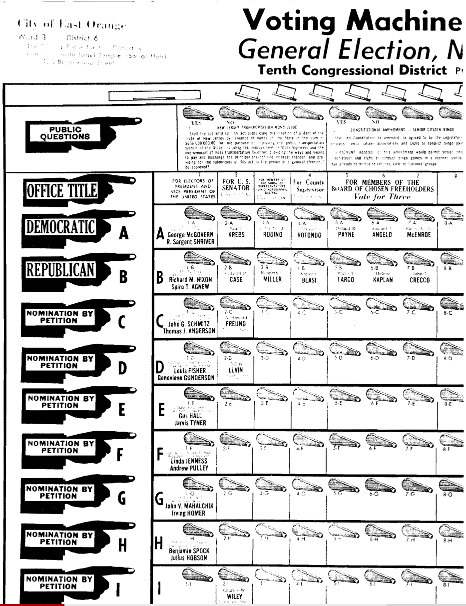
October 15, 2014

1. Review
    - Caputo
    - Normal Distribution
    - Chi-Squared & F Distributions
2. Hypothesis Testing
    - Theory
    - The Null and Alternative Hypotheses
    - Critical Regions
    - Test Statistics
    - The p-value
    - Core Questions/Issues
3. Hypothesis Tests: Practically
    - Fisher Hypothesis Test Setup
    - Test Statistics
4. Power

1. Review
    - Caputo
    - Normal Distribution
    - Chi-Squared & F Distributions
2. Hypothesis Testing
    - Theory
    - The Null and Alternative Hypotheses
    - Critical Regions
    - Test Statistics
    - The p-value
    - Core Questions/Issues
3. Hypothesis Tests: Practically
    - Fisher Hypothesis Test Setup
    - Test Statistics
4. Power

Nicholas Caputo

- Essex County Commissioner
- Sets ballot order



City of East Orange
Ward 3 District 6

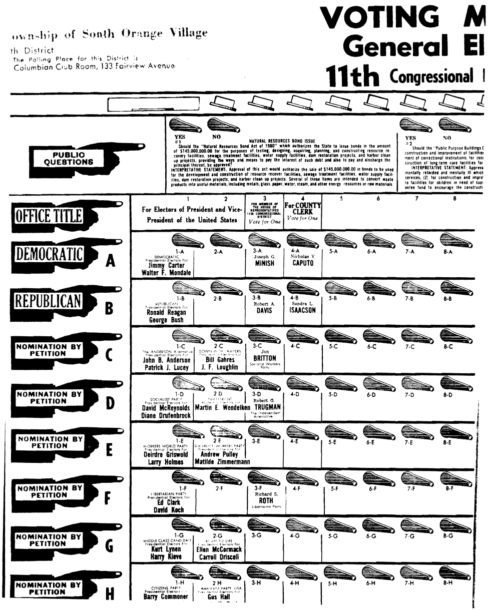
Voting Machine
General Election, N
Tenth Congressional District

PUBLIC QUESTIONS

NEW JERSEY TRANSPORTATION BOND ISSUE

CONSTITUTIONAL AMENDMENT — SENIOR CITIZEN BINGO

OFFICE TITLE		1 FOR ELECTORS OF PRESIDENT AND VICE PRESIDENT OF THE UNITED STATES	2 FOR U. S. SENATOR	3 FOR MEMBER OF THE HOUSE OF REPRESENTATIVES	4 For County Supervisor	5 FOR MEMBERS OF THE BOARD OF CHOSEN FREEHOLDERS Vote for Three	6	7	8
DEMOCRATIC	A	1-A George McGOVERN R. Sargent SHRIVER	2-A Paul J. KREBS	3-A Peter W. Jr. RODINO	4-A ROTONDO	5-A Donald M. PAYNE	6-A Samuel ANGELO	7-A McENROE	8-A
REPUBLICAN	B	1-B Richard M. NIXON Spiro T. AGNEW	2-B Clifford P. CASE	3-B Kenneth MILLER	4-B Walter BLASI	5-B FARCO	6-B Helene KAPLAN	7-B John CRECCO	8-B
NOMINATION BY PETITION	C	1-C John G. SCHMITZ Thomas J. ANDERSON	2-C A. Howard FREUND	3-C	4-C	5-C	6-C	7-C	8-C
NOMINATION BY PETITION	D	1-D Louis FISHER Genevieve GUNDERSON	2-D LEVIN	3-D	4-D	5-D	6-D	7-D	8-D
NOMINATION BY PETITION	E	1-E Gus HALL Jarvis TYNER	2-E	3-E	4-E	5-E	6-E	7-E	8-E
NOMINATION BY PETITION	F	1-F Linda JENNESS Andrew PULLEY	2-F	3-F	4-F	5-F	6-F	7-F	8-F
NOMINATION BY PETITION	G	1-G John V. MAHALCHIK Irving HOMER	2-G	3-G	4-G	5-G	6-G	7-G	8-G
NOMINATION BY PETITION	H	1-H Benjamin SPOCK Julius HOBSON	2-H	3-H	4-H	5-H	6-H	7-H	8-H
NOMINATION BY PETITION	I	1-I	2-I Charles W. WILEY	3-I	4-I	5-I	6-I	7-I	8-I



ownship of South Orange Village
th District
The Polling Place for this District is
Columbian Club Room, 133 Fairview Avenue

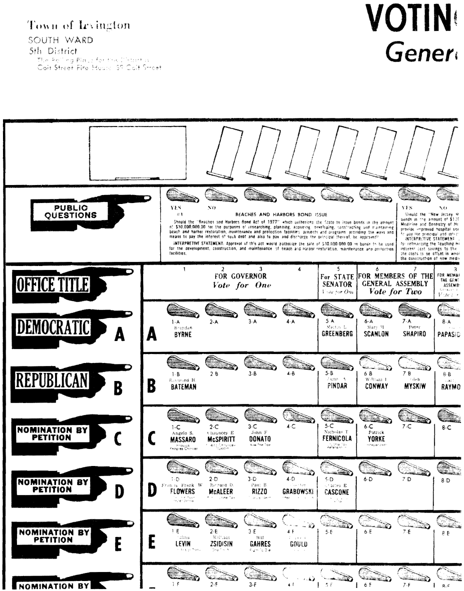
VOTING M
General El
11th Congressional

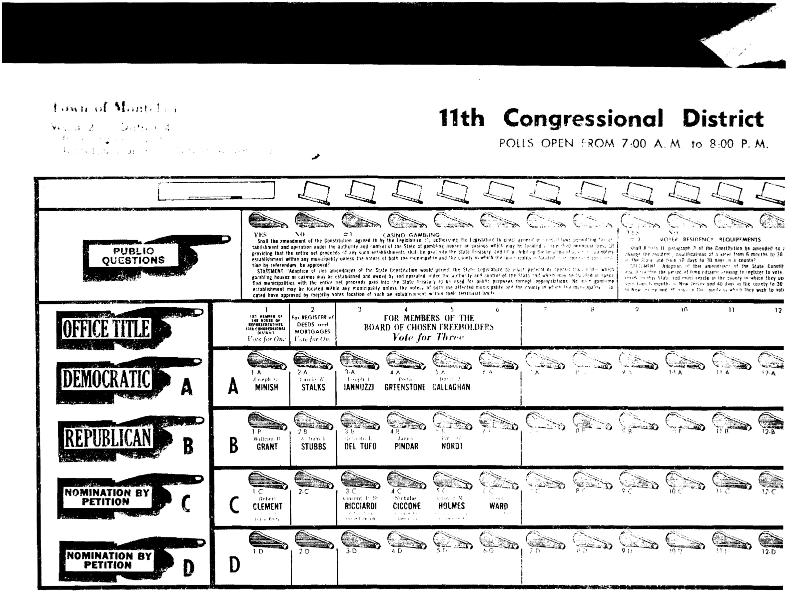
PUBLIC QUESTIONS

YES NO
#1
NATURAL RESOURCES BOND ISSUE
Should the "Natural Resources Bond Act of 1980" which authorizes the State to issue bonds in the amount of $145,000,000.00 for the purposes of testing, designing, acquiring, planning, and constructing resource recovery facilities, sewage treatment facilities, water supply facilities, dam restoration projects, and harbor clean up projects, providing the ways and means to pay the interest of such debt and also to pay and discharge the principal thereof, be approved?
INTERPRETATIVE STATEMENT: Approval of this act would authorize the sale of $145,000,000.00 in bonds to be used for the development and construction of resource recover facilities, sewage treatment facilities, water supply facilities, dam restoration projects, and harbor clean up projects. Several of these items are intended to convert waste products into useful materials, including metals, glass, paper, water, steam, and other energy resources or raw materials

YES NO
#2
Should the "Public Purpose Buildings C
construction and improvement of facilities
ment of correctional institutions; for cons
struction of long-term care facilities for
INTERPRETATIVE STATEMENT: Approva
mentally retarded and mentally ill which
services; (2) for construction and impro
to facilities for children in need of sup
antee fund to encourage the constructi

OFFICE TITLE	1 For Electors of President and Vice-President of the United States	2	3 For Member of the House of Representatives 11th Congressional District Vote for One	4 For COUNTY CLERK Vote for One	5	6	7	8
DEMOCRATIC A	1-A DEMOCRATIC Presidential Electors For Jimmy Carter Walter F. Mondale	2-A	3-A Joseph G. MINISH	4-A Nicholas V. CAPUTO	5-A	6-A	7-A	8-A
REPUBLICAN B	1-B REPUBLICAN Presidential Electors For Ronald Reagan George Bush	2-B	3-B Robert A. DAVIS	4-B Sandra L. ISAACSON	5-B	6-B	7-B	8-B
NOMINATION BY PETITION C	1-C The ANDERSON Alternative Presidential Electors For John B. Anderson Patrick J. Lucey	2-C Bill Gahres J. F. Loughlin	3-C Jon BRITTON Socialist Workers Party	4-C	5-C	6-C	7-C	8-C
NOMINATION BY PETITION D	1-D SOCIALIST PARTY Presidential Electors for David McReynolds Diane Drufenbrock	2-D Martin E. Wendelken	3-D Robert G. TRUGMAN The Independent Alternative	4-D	5-D	6-D	7-D	8-D
NOMINATION BY PETITION E	1-E WORKERS WORLD PARTY Presidential Electors For Deirdre Griswold Larry Holmes	2-E Presidential Electors For Andrew Pulley Matilde Zimmermann	3-E	4-E	5-E	6-E	7-E	8-E
NOMINATION BY PETITION F	1-F LIBERTARIAN PARTY Presidential Electors For Ed Clark David Koch	2-F	3-F Richard S. ROTH Libertarian Party	4-F	5-F	6-F	7-F	8-F
NOMINATION BY PETITION G	1-G MIDDLE CLASS CANDIDATE Presidential Electors For Kurt Lynen Harry Kieve	2-G RIGHT TO LIFE Presidential Electors For Ellen McCormack Carroll Driscoll	3-G	4-G	5-G	6-G	7-G	8-G
NOMINATION BY PETITION H	1-H CITIZENS PARTY Presidential Electors For Barry Commoner	2-H COMMUNIST PARTY USA Presidential Electors For Gus Hall	3-H	4-H	5-H	6-H	7-H	8-H


Town of Irvington
SOUTH WARD
5th District
VOTIN
Gener
PUBLIC QUESTIONS
BEACHES AND HARBORS BOND ISSUE
OFFICE TITLE
FOR GOVERNOR Vote for One
For STATE SENATOR
FOR MEMBERS OF THE GENERAL ASSEMBLY Vote for Two
DEMOCRATIC A
BYRNE
GREENBERG
SCANLON
SHAPIRO
REPUBLICAN B
BATEMAN
PINDAR
CONWAY
MYSKIW
NOMINATION BY PETITION C
MASSARO
McSPIRITT
DONATO
FERNICOLA
YORKE
NOMINATION BY PETITION D
FLOWERS
McALEER
RIZZO
GRABOWSKI
CASCONE
NOMINATION BY PETITION E
LEVIN
ZSIDISIN
GAHRES
GOULD


11th Congressional District
POLLS OPEN FROM 7:00 A. M. to 8:00 P. M.
PUBLIC QUESTIONS
YES
NO
CASINO GAMBLING
VOTER RESIDENCY REQUIREMENTS
OFFICE TITLE
FOR MEMBERS OF THE BOARD OF CHOSEN FREEHOLDERS Vote for Three
For REGISTER of DEEDS and MORTGAGES
DEMOCRATIC A
MINISH
STALKS
IANNUZZI
GREENSTONE
CALLAGHAN
REPUBLICAN B
GRANT
STUBBS
DEL TUFO
PINDAR
NORDT
NOMINATION BY PETITION C
CLEMENT
RICCIARDI
CICCONE
HOLMES
WARD
NOMINATION BY PETITION D

Nicholas Caputo

- 40 of 41 times, picks a Democrat for the top of the ballot. Hmmmm.
- Lucky or cheating? We'll never know.
- But we can examine reasonable doubt. If the process were fair,
  - What is the probability of seeing a result like this one;
  - Is this beyond a reasonable doubt?

Nicholas Caputo

Our framework for hypothesis testing is strikingly similar to a jury trial.

- Innocent until proven guilty.
  We have a null hypothesis, usually that we are wrong
    - The coin is fair
    - The factor does not matter
    - The two groups aren't different
- The burden of proof is on the data (and therefore us) is to show that this is incorrect.
- We examine the data (evidence).

Nicholas Caputo

- We have to show "guilt" beyond a reasonable doubt. In this case, reasonable doubt is quantified precisely as:
  - The probability of seeing a result this far - or further - from the null hypothesis, if in fact the null is true.
  - Or, the "likelihood" of seeing all this evidence if in fact the suspect is guilty.
- So - we need to know how to calculate probabilities.
- Aside: note that whatever we conclude, we could be right or wrong - and we'll never know. Just like in a real jury trial...

Aside? A little function to toss coins

```r
meetbinom <- function(n,p){
  ## This little function will toss coins
  ## of number n with probability of
  ## success p and make a barchart.
  res <- rep(NA, n)
  for(k in 1:n){
    res[k] <- dbinom(k,n,p)
    }
  barplot(res, names.arg = 1:n,
    main = 'Probability of "k" successes',
    ylab = 'Probability',
    xlab = 'k')
  }

meetbinom(3,.5)
meetbinom(10,.2)
```

../../code/meetbinom.R

What about Caputo?

- One solution - add up disjoint events. Example: P(40 or more successes in 41 trials, $p = .5$):

$$P(40Ds) = 1 - P(1D) = \binom{41}{1}(.5)^{.40}(.5)^{1} = 1.8e^{-11}$$

- Another solution - the binomial kind of looks like a normal curve. Maybe we could pretend that it is normal and use "the normal approximation"? We'll learn more about this later, but for now:

$$P(K \geq 40) = 1 - \psi(Z)$$

$$Z = \frac{40 - 20}{\sqrt{\frac{.5*(1-.5)}{41}}} =$$

$$P(K \geq 40) = 1 - \psi$$

- Another solution - toss a coin 41 times. Count the number of heads. Repeat m times. What fraction of the m data points are 40 or greater?

What about Caputo?

```r
caput <- function(nSims = 100, prob = 0.5){
  ## Let's simulate some Caputo trials.
  res <- rep(NA, nSims)
  for(i in 1:nSims){
    res[i] <- rbinom(n = 1, size = 41, p = 0.5)
  }
  print(table(res)/nSims)
}
```

../../code/caputoSim.R

1. Review
   - Caputo
   - Normal Distribution
   - Chi-Squared & F Distributions
2. Hypothesis Testing
   - Theory
   - The Null and Alternative Hypotheses
   - Critical Regions
   - Test Statistics
   - The p-value
   - Core Questions/Issues
3. Hypothesis Tests: Practically
   - Fisher Hypothesis Test Setup
   - Test Statistics
4. Power

Why be normal?

- A strong desire to conform
- Fear of deviance
- Parental Guidance
- Religious institutions
- The Central Limit Theorem

The Central Limit Theorem in Action

The mean of a sample of independent draws from the same distribution with expected value μ and finite variance σ^2 will be normally distributed with an expected value of μ and a variance of $\frac{\sigma^2}{n}$, or standard deviation of $\frac{\sigma}{\sqrt{n}}$.

Take a sample. Calculate the mean of the sample. Do the same thing again, and again, and again... The distribution of sample means will be normal.

So.... Whatever the distribution we are working with looks like, if our *statistic* is something like a mean, we can use the normal curve!!!

The Central Limit Theorem in Action

```r
cltbinom<-function(m,n,N,p){
  allsamplemeans<-NA
  for(i in 1:m){
    sample<-rbinom(n,N,p)
    allsamplemeans[i] <- mean(sample)
  }
  par(mfrow=c(1,2))
  hist(rbinom(m,N,p), main=paste("Sample of",m,"draws"),
      xlim=c(0,1))
  hist(allsamplemeans, main=paste("Sample Means, sample size
      = ",n))
}

cltbinom(10000,10,10,.1)
```

../../code/cltBinom.R

More Normalcy

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}$$

$$E(x) = \mu$$

$$V(x) = \sigma^2$$

$$\Phi(x)$$

Transforming to a Standard Normal: $N(0, 1)$

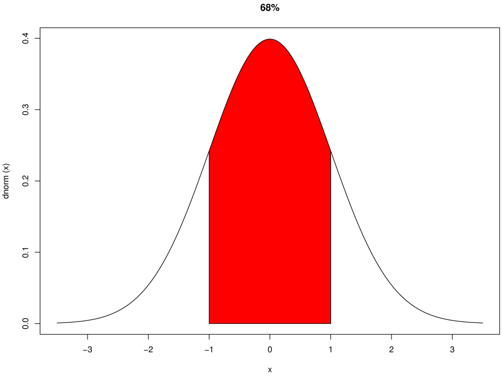
- Desert Island
- Teaching
- Standardizes test statistics

$$X \sim N(\mu, \sigma^2)$$
$$Z = \frac{X - \mu}{\sigma}$$
$$Z \sim N(0, 1)$$

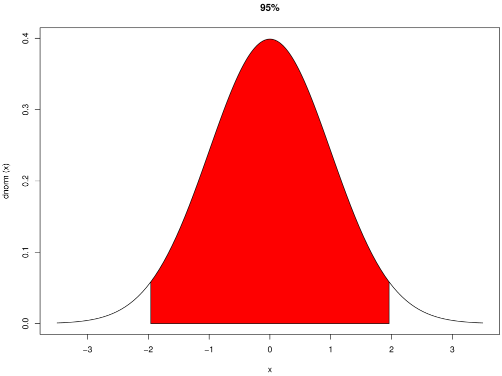
Getting probabilities from Normals

- Standard Normal Tables
- Facts: 1,2,3, 1.96
- R

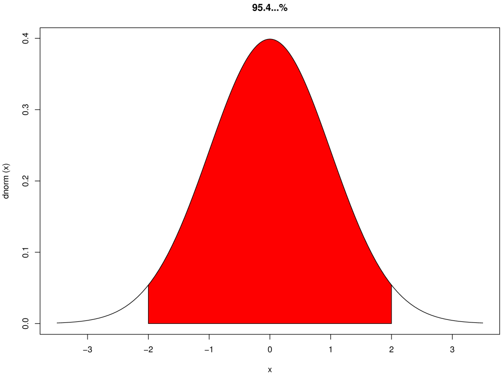
Normal Probabilities: 1 SD



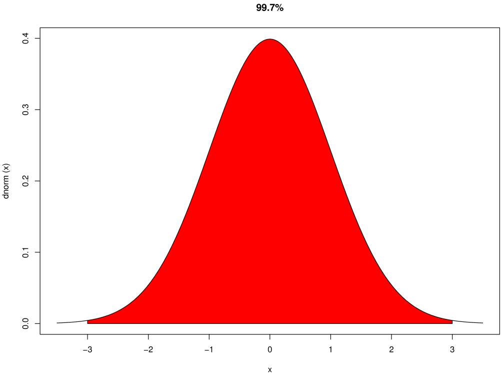
Normal Probabilities: 95%



Normal Probabilities: 2SD



Normal Probabilities



Chi-Square

$$X_i \sim N(0,1) \forall i$$

$$Y = \sum_{i=1}^{n} X_i^2$$

$$Y \sim \chi^2_n$$

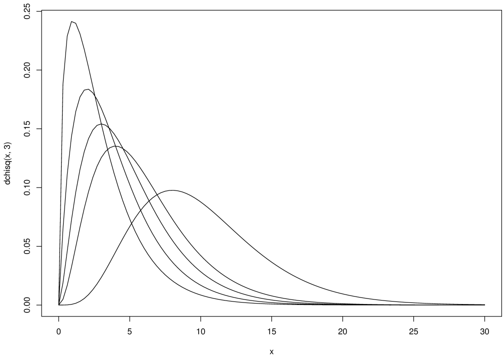
$$f(Y;n) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} y^{n/2-1} e^{-y/2}, y \geq 0$$

$$E(Y) = n; V(Y) = 2n$$

Chi Squared Distribution

```
curve(dchisq(x,3),from=0,to=30,ylab="Density")
curve(dchisq(x,4),from=0,to=30,add=TRUE)
curve(dchisq(x,5),from=0,to=30,add=TRUE)
curve(dchisq(x,6),from=0,to=30,add=TRUE)
curve(dchisq(x,10),from=0,to=30,add=TRUE)
```

Chi Squared Distribution



F-Distribution

$$K_1 \sim \chi^2_{d_1}$$

$$K_2 \sim \chi^2_{d_2}$$

$$G = \frac{K_1/d_1}{K_2/d_2}$$

$$G \sim F_{d_1,d_2}$$

1 Review
Caputo
Normal Distribution
Chi-Squared & F Distributions

2 Hypothesis Testing
Theory
The Null and Alternative Hypotheses
Critical Regions
Test Statistics
The p-value
Core Questions/Issues

3 Hypothesis Tests: Practically
Fisher Hypothesis Test Setup
Test Statistics

4 Power

From Probability to Inference

- We want to know if the coin is fair - without using a physicist.
- Two related questions:
  1. What do we think P(H) is? And how certain are we about this?
  2. Is the coin fair, or not?
- The first is about *Confidence Intervals*, the second is about *Hypothesis Testing*

A Simple Framework

There is an unobserved fact about the world that we care about:

- Is the coin fair?
- Do negative advertisements reduce turnout?
- Do women in Latin America have different opinions about policy than men?
- What percentage of California men will vote for Clinton?
- Are Republican Presidents worst at balancing budgets?

A Simple Framework

- In each case there is some unobserved *parameter* of interest
- We never get to see it. Too bad for us.
- Instead, realizations of a random sampling process of driven by our parameter of interest.
- Use these samples to produce *estimates* (also random variables).
- We use information contained in that sample to try and say something about the true value, the parameter.
- Note that in some cases we rely on the Star Trek Theorem of Sampling.

A Simple Framework

- Similar to trial by jury.
- Null hypothesis: innocent until proven guilty. Doesn't work like Men in Black.
- Alternative hypothesis: guilty beyond a reasonable doubt.
  - So much evidence against the suspect that although s/he *could* be innocent...
  - it would require such a ridiculous combination of events that we'll just call her/him guilty.
- Note that we accept the possibility that we are wrong with each of our verdicts.

1 Review
Caputo
Normal Distribution
Chi-Squared & F Distributions

2 Hypothesis Testing
Theory
The Null and Alternative Hypotheses
Critical Regions
Test Statistics
The p-value
Core Questions/Issues

3 Hypothesis Tests: Practically
Fisher Hypothesis Test Setup
Test Statistics

4 Power

A Simple Framework

- **The Key Question**: What is the probability of seeing a sample like the one we have, if the null hypothesis were true? (Or, assume the suspect is innocent. Could their fingerprints have just randomly appeared at the crime scene?)
- Two Possible (Ambiguous) Answers:
  - Guilty. *Reject the null hypothesis*. The data we have is so inconsistent with what we would expect under the null hypothesis, that we don't think the null hypothesis is correct.
  - Not Guilty. The data isn't so different from what we would expect under the null - so we *fail to reject the null.*

The Null and Alternative Hypothesis

- We're interested in estimating some unknown parameter θ whose value is unknown but that we know must lie in some sample space Ω.
- Suppose that Ω can be seperated into two disjoint subsets Ω_0 and Ω_1.
- We're interested in determining whether θ lies in Ω_0 or Ω_1.
- Let H_0 demote the hypotehsis that $\theta \in \Omega_0$ and $\mathrm{H}_1 \in \Omega_1$.
- Since Ω_0 and Ω_1 partition the space, we know that θ must lie somewhere on the space

H_0 is the **null hypothesis** and H_1 is the **alternative hypothesis**.

The Null and Alternative Hypothesis

- If θ lies in Ω_1 we are said to *reject* H_0.
- If we decide that θ lies in Ω_0 we *fail to reject* the null hypothesis.
- **Note that this is different than accepting the null hypothesis as correct.**

Example: Egyptian Skulls

Example

We've found some skulls (or more appropriately, Thomson and Randall-Maciver (1905) have found some skulls). We're interested in knowing how the size of these skulls compares to the size of present day-skulls.

- What are some ways that you could set up a set of hypotheses?
- What ways seem best?
- I propose the following setup: $\mathrm{H}_0 : \mu \geq 140$; H_1: $\mu < 140$.
- This is a *one-sided* hypothesis test.
- In a two-sided hypothesis test, of the form:
    1. $H_0 : \mu = k$;
    2. $H_1 : \mu \neq \mu_0$

1 Review
Caputo
Normal Distribution
Chi-Squared & F Distributions

2 Hypothesis Testing
Theory
The Null and Alternative Hypotheses
Critical Regions
Test Statistics
The p-value
Core Questions/Issues

3 Hypothesis Tests: Practically
Fisher Hypothesis Test Setup
Test Statistics

4 Power

Critical Regions

Example

Suppose that $\mathbf{X} = (X_1, \ldots, X_n)$ is a random sample from the normal distribution with unknown mean μ and known variance σ^2. We then might want to test the hypothesis:

1. $\mathrm{H}_0 : \mu = \mu_0$,
2. $\mathrm{H}_1 : \mu \neq \mu_0$.

- It might be reasonable to reject H_0 if $\bar{X}_n$ is **pretty far** from μ_0.
- This would just mean picking some number c (for critical value), or maybe d (for distance).
- If the distance between μ_0 and $\bar{X}_n$ is grater than this value, *we would reject the null hypothesis*.

Critical Regions

Answer

$S_0 = \{x : -c \leq \bar{X}_n - \mu_0 \leq c\}$, and
$S_1 = S_0^C$

- Under this framework, S_1 is the **critical region** of the test, the area in which we would reject the null hypothesis if we observed a value in that range.

Example

1. What if we are interested in finding the high-limit on a uniform distribution?
2. What if we're testing the hypothesis that the high-limit $b \in [3, 4]$?

1 Review
Caputo
Normal Distribution
Chi-Squared & F Distributions

2 Hypothesis Testing
Theory
The Null and Alternative Hypotheses
Critical Regions
Test Statistics
The p-value
Core Questions/Issues

3 Hypothesis Tests: Practically
Fisher Hypothesis Test Setup
Test Statistics

4 Power

From Critical Regions to Test Statistics

- In practice, we will use a related concept known as a *test statistic* for more of our tests.

Definition

- Let $\mathbf{X}$ be a random sample from a distribution that depents on a parameter θ.
- Let $T = r(\mathbf{X})$ be a *test-statistic*, and let R be a subset of the real line.
- If we partition the Ω to form hypotheses, creating a hypothesis of the form "reject H_0 if $\mathrm{T} \in \mathrm{R}$",
- Then T is the *test-statistic* and R defines the bounds of the *rejection region* of the test.

Test Statistics

- We use test statistics because is povides a common framework for nearly *all* types of tests presented.
- Once we move to a test-statistic framework, we will just present results in terms of the test statistic (and p-value), not rejection regions

Major Points

- Because test-statistics are functional transformations of a set of data, these test-statistics *themselves* have distributions.
- Distributions of distributions ?!? C'mon...

1. What is Z associated with a 95% Confidence Interval?
2. Also gives a principled way of calculating *lots and lots and lots* of things.

Test Statistics

- Single number that characterizes some attribute of a data sample
- Distribution of a test-statistic is known under the null hypothesis (either exactly or approximately)

Test Statistic Example

Example

Geiger Counters measure alpha, beta, and gamma rays.

- Portrayed in movies, they scale radioactivity from 0 – 10 and produce audible clicks.
- The [0 – 10] might be thought of as a test statistic, characterizing how, um, hot things are.
- 0 zero risk of radioactive harm; 10 extreme risk.

1. Review
   - Caputo
   - Normal Distribution
   - Chi-Squared & F Distributions
2. Hypothesis Testing
   - Theory
   - The Null and Alternative Hypotheses
   - Critical Regions
   - Test Statistics
   - The p-value
   - Core Questions/Issues
3. Hypothesis Tests: Practically
   - Fisher Hypothesis Test Setup
   - Test Statistics
4. Power

The p-value

Definition

The **p-value** is the smallest level α_0 such that we would reject the null-hypothesis at a level α_0 with the observed data.

- Statistics is data-reduction: finding the mean and variance of a sample of IQ's reduces the data to just three parameters: $\Phi(\mu, \sigma^2, \epsilon)$
- For example, what if we choose a critical level of $\alpha_0 = 0.05$. Then, if we generate a test statistic of $Z = 1.97$ or $Z = 100.1$, we would report the same fact: **reject the null**.
- What could we do to provide a little more information about the test? Provide, in addition to the test statistic, the probability of observing a test statistic that large or larger under the null hypothesis.

Core Questions

1. What level of α_0 should one choose? What considerations shape this choice?
2. When we find a result, we frequently pat ourselves on the back, maybe crack open a beer, and conclude that we have a statistically significant result at level α_0. Crucially, *this does not mean that we should behave as if H_0 is false*.
3. This is a particularly important distinction to make when working with large amounts of data, where one is likely to observe statistically significant differences between hypotheses and observed values, even when those differences are very small.

Fisher Hypothesis Test Setup

1. State the Null and Alternative Hypothesis
2. Consider the statistical and distributional assumptions about the sample for the test
3. Identify the appropriate test-statistic
4. Derive the distribution of that test statistic under the null hypothesis
5. State the confidence level (α) required to reject the null in your test, and identify the critical values of the test statistic associated with that α
6. Compute the test statistic from your observed sample
7. Compare calculated test statistic and critical values, accept or reject hypothesis

Hypothesis Test: Short-Form

1. State null and alternative hypothesis
2. State α
3. Identify test statistic
4. Calculate test-statistic, check p-value
5. Conclude

Finishing the Caputo Example

Example

Recall, Nicholas Caputo: 41 elections 40 listed the democrats first. *Set up the hypothesis test using our framework.*

- $H_0 : P(H) = 0.50$
- $H_A : P(H) \neq 0.50$
- $\alpha = 0.05$
- Identify T and $P(T)$: `2*sum(pbinom(c(0,1), 41, .5))`
- Reject H_0, `p = 3.910827e-11`

Test Statistics

1. One sample Proportion
2. Two sample Proportion
3. Sample Mean
4. Two Sample Means
5. Inference for Counts
6. Variance of a Sample
7. Difference in Variance Between Two Samples

For each of these, we will introduce the test statistic, try an example, and (maybe) talk about Margin of Error and Confidence Intervals.

One-sample Proportion

$$z = \frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$$

- $\hat{p}$ is the calculated proportion
- p_0 is the probability under the null hypothesis

One-sample Proportion

Example

Do we have a fair coin?

- $n = 100$, 55 heads
- $n = 200$, 90 heads

Answer

- H_0: The coin is fair; H_A: The coin is unfair.
- $\alpha = 0.05$
- $$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{.55 - .50}{\sqrt{\frac{.5(1-.5)}{100}}} = \frac{.05}{\frac{.5}{10}} = 1$$
- p-val ~ 0.32 `(2*(1-pnorm(1)))`
- Conclude: *Fail to reject the null hypothesis*, not enough evidence.

More Examples

Example

Let the null hypothesis be that $p = .35$. Can we reject the possibility that $P(Heads) = .35$ for our coin if we observe: 1000 tosses with 385 heads?

Example

More Realistic Example We know that 55% of Californians approved of Schwarztenegger in 2006. We take a sample of 1,000 and find that only 485 respondents “approved” of his administration. Can we reject the null hypothesis that overall support for the Governator is unchanged? *Note - this requires that we know that 55% approved of the governor in 2006. How would we know that?*

Confidence Intervals

What if we want to give a sense for the range that we might expect $\hat{p}$ to fall on? *See aside on interpreting confidence intervals...*

Definition

The **Margin of Error** of a sample proportion is the standard error multiplied by the z-score (from α).

Definition

The **Confidence Interval** of a sample proportion is the estimate $\pm$ the Margin of Error.

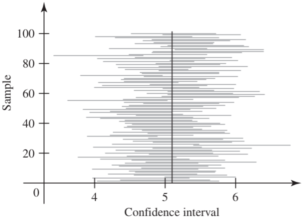
$$95\% CI = \hat{p} \pm 1.96 * \sqrt{\frac{\hat{p} * (1 - \hat{p})}{n}}$$

Interpreting Confidence Intervals

Interpreting confidence intervals requires a somewhat delicate hand.

- If we were to take the proportions of heads from many sets of n tosses, 95% of the confidence intervals would cover the true mean.

Figure 8.5 A sample of one hundred observed 95% confidence intervals based on samples of size 26 from the normal distribution with mean $\mu = 5.1$ and standard deviation $\sigma = 1.6$. In this figure, 94% of the intervals contain the value of μ.



Confidence Interval Example

Example

What is the 90% confidence interval for the probability of heads given 200 tosses and 90 heads?

Two-sample Proportion

$$z = \frac{(\hat{p}_1 - \hat{p}_2)}{\sqrt{\frac{\hat{p}_p(1-\hat{p}_p)}{n_1} + \frac{\hat{p}_p(1-\hat{p}_p)}{n_2}}}$$

- $\hat{p}_p = \frac{k_1+k_2}{n_1+n_2}$
- k_i is "successes" in sample i
- n_i is number of trials in sample i

Two-sample Proportions

1. State null and alternative hypothesis
2. State α
3. Identify test statistic
4. Calculate test-statistic, check p-value
5. Conclude

Hypotheses are possibly of two forms:

- $p_1 \neq p_2$; or,
- $p_1 \leq p_2$

Two-sample Proportions

Example

1. $n_1 = 13, k_1 = 3; n_2 = 20, k_2 = 16$
2. $n_1 = 111, k_1 = 85; n_2 = 124, k_2 = 102$
3. $n_1 = 147, k_1 = 19; n_2 = 154, k_2 = 55$

Answer

1. $H_0 : p_1 = p_2;\ H_A : p_1 \neq p_2$
2. $\alpha = 0.05$
3. z on last slide
4. $z = \frac{(3/13-16/20)}{\sqrt{(19/33)(14/33)/13+(19/33)(14/33)/20}} = 3.2$
5. Reject the null hypothesis

Two-sample Proportions

Example

1. $n_1 = 111, k_1 = 85; n_2 = 124, k_2 = 102$
2. $n_1 = 147, k_1 = 19; n_2 = 154, k_2 = 55$

Two-sample Proportions: Example 2

Example

- Given two independent samples, each from a different population is there a difference between p_1 and p_2 in the two populations?
- p might be the proportion of subjects that want the Governator to be re-elected, and one population might be Southern California residents, the others from Northern California.

For example:

- Southern California Survey (#1): $n_1 = 111, k_1 = 85$
- Northern California Survey (#2): $n_2 = 124, k_2 = 102$

Two-sample Proportions: Example 2

Answer

1. Hypotheses: Null: $p_1 = p_2$, or $p_1 - p_2 = 0$ Alternative: $p_1 \neq p_2$, or $p_1 - p_2 \neq 0$.
2. Set $\alpha = .05$
3. Calculate Test Statistic

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_p(1-p_p)}{n_1} + \frac{p_p(1-p_p)}{n_2}}}$$

where $\hat{p}_1 = \frac{k_1}{n_1} = \frac{85}{111} \sim .77$ and $\hat{p}_2 = \frac{k_2}{n_2} = \frac{102}{124} \sim .82$ and $p_p = \frac{k_1+k_2}{n_1+n_2} = \frac{85+102}{111+124} \sim .80$

Two sample proportions: Example 2

Answer

3.

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_p(1-p_p)}{n_1} + \frac{p_p(1-p_p)}{n_2}}}$$

$$= \frac{.77 - .82}{\sqrt{\frac{.80(1-.80)}{111} + \frac{.80(1-.80)}{124}}} \sim -1.0785$$

Two sample proportions: Example 2

Answer

4. Get p-value: `2*(pnorm(-1.0785)=.2808107`
5. Conclude: Because our p-value (.2808107) is greater than our α (.05), we *fail to reject the null hypothesis that* $p_1 = p_2$. Under the null hypothesis, it is fairly likely that we would observe such a large difference between $\hat{p}_1$ and $\hat{p}_2$ just due to random sampling error.

Two sample proportions: Example 3

Example

Let's try another example: $n_1 = 1000$, $n_2 = 2500$, $k_1 = 575$, $k_2 = 1450$. Test whether $p_1 = p_2$.

Confidence Interval for Difference of Two Proportions

$$CI = \hat{p}_1 - \hat{p}_2 \pm ME$$

$$ME = 1.96 * \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$CI = \hat{p}_1 - \hat{p}_2 \pm 1.96 * \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

Again, notice the subtle difference in calculation between the margin of error and the test statistic.

Sample Mean

Recall that the sample mean is defined by $\bar{x} = \frac{1}{n}\sum_{i=1}^{N} x_i$. Now, we define two related, but (importantly) distinct concepts.

Definition

The **sample standard deviation** typically denoted s is, quite simply the square-root of the variance of the sample.

$$s = SE(x) = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n-1}}$$

Definition

The **standard error of the mean** typically denoted $SE(\bar{x})$ or $SE_{\bar{x}}$, is the "sample variance" divided by the root of n.

$$SE_{\bar{x}} = \sqrt{\frac{s^2}{n}}$$

Sample Mean

If we knew what the "*true*" σ were, we would just obtain a maximally *efficient* estimate of $\bar{x}$ as $N(\mu, \sigma/n)$. Buut... we don't know σ.

Degrees of Freedom Fries?

How would the following samples of $n = 2$ affect our estimates of μ and σ?

- Imagine unknown μ, we draw two values, -2 and $+2.5$
- Imagine known $\mu = 0$, we draw two values, -2 and $+2.5$
- Imagine known $\mu = 0$, we draw two values, 2 and 2.5
- Imagine unknown μ, we draw two values, 2 and 2.5
- Jointly calculating μ and σ puts us at risk of making a mistake.
- Problem goes away as n gets large
- This is where "t-tests" and "t-statistics" come from, and is also why we divide by $n - 1$ when calculating sample standard deviation.

Normal and T-Distributions

```
curve(dnorm,
  from=-5, to=5,
  lty = 1, lwd = 3, col = "blue",
  main = "Plot of Normal (blue) and t (red) Distributions")
curve(dt(x,df=1),
  from=-5,to=5,
  lty = 2, lwd = 3, col = "red",
  add=TRUE)
pnorm(-2)
pt(-2,df=1)
```

Test Statistic of Sample Mean

With that overhead out of the way...

$$t_{n-1} = \frac{\bar{x} - \mu_0}{SE_{\bar{x}}}$$

And the confidence interval is a straightforward translation:

$$CI = \bar{x} \pm t_{\alpha,df} * SE_{\bar{x}}$$

Europe

Example

- We know that Europeans have an average trust level of 5.32 (assume we did a census and asked them *all*).
- We want to know about Paraguayans, so we take a random sample of $n = 50$ and from our sample calculate a sample mean of $\bar{x} = 4.93$ and a sample standard deviation of $s = 3.5$.
- Are Paraguayan trust levels different from those in Europe?

Europe

Answer

1. Hypotheses: **Null**: Paraguayans have the same mean interpersonal trust as Europeans, $\bar{x}_p = \mu_e = 5.32$ **Alternative** $\bar{x}_p \neq 5.32$
2. Require $\alpha = 0.05$
3. Calculate test statistic: t_{n-1}

$$t_{n-1} = \frac{\bar{x} - \mu_0}{SE(\bar{x})}$$

where $SE(\bar{x}) = \sqrt{\frac{s^2}{n}}$, and $n = 50$. so:

$$t_{49} = \frac{4.93 - 5.32}{\sqrt{\frac{3.5^2}{50}}} = \frac{-.39}{\sqrt{.245}} \sim -.79$$

4. Obtain p-value: `2*pt(-0.79, df=49)` $= 0.4345$.
5. Conclude: Fail to reject the null. There is a 43% chance this would

Two Sample Means

$$t_{df} = \frac{\bar{x_1} - \bar{x_2}}{SE_{\bar{x_1} - \bar{x_2}}}$$

$$CI = \bar{x_1} - \bar{x_2} \pm SE_{\bar{x_1} - \bar{x_2}} * t_{\alpha, n-1}$$

$$SE_{\bar{x_1} - \bar{x_2}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Key issues:

1. Do you assume that both populations have the same standard deviation?
2. How do you calculate the degrees of freedom when working with two samples?

df for two sample means

- Equal sample sizes, equal variance

$$df = 2n - 2$$

- Unequal sample sizes, equal variance

$$df = n1 + n2 - 2$$

- Unequal sample sizes, unequal variance

$$df = \frac{(s_1^2/n_1 + s_2^2)^2}{\frac{(s_1^2/n_1)^2}{(n_1-1)} + \frac{(s_2^2/n_2)^2}{(n_2-1)}}$$

Review of assumptions

1. Populations follow the Normal Distribution (test via Kolmogorov-Smirnov test?)
2. Same variance (classic t-test), or use appropriate adjustment if different
3. The data is drawn from an independent sample of the data

But what if these don't hold? Or, what if you want to *know* that your data is normal?

There's an app for that.

Kolmogorov-Smirnov Test

Definition

The **Kolmogorov-Smirnov Test** (KS-test) is a nonparametric test for the equality of a distribution either (a) with another sample distribution, or (b) a reference distribution (i.e. Gaussian).

- Maps a distance between the CDF of suspect distribution and reference distribution
- Has *known* distribution under null hypothesis.

Demonstrating the distribution and its properties are beyond the scope of this class, but are straightforward.

Kolmogorov-Smirnov Test

The process of a hypothesis test is *exactly* the same, in spite of the fact that we have a slightly more baroque test-statistic.

1. Hypotheses: **Null**: same hypothesis; **Alternative**: different hypotheses
2. Set $\alpha = 0.05$
3. Calculate test-statistic (`R`, not by hand...)
4. Conclude

Inference for Counts – χ^2

Definition

The $\chi^{\mathbf{2}}$ tests the relationship between categorical variables. Are two variables independent?

$$\chi^2 = \sum \frac{(Obs - Exp)^2}{Exp}$$

Where

- *Obs* is the observed value in a cell,
- *Exp* is the expected value under the assumption of independence.
- And, for cell i, j in the table,

  $$Exp_{ij} = \frac{c_j * r_i}{n}$$

  where c_j is the column total, r_i is the row total, and n is the total of all observations.
- The chi-square statistic is distributed chi-square with (rows-1)*(columns-1) *df*

Example, χ^2

Example

What is the relationship between occupation and party preference? Are students more likely to vote for liberals? Are retirees more likely to vote for conservatives? Are the employed more likely to be confused?

	PAN	PRI	PRD
Employed	320	245	288
Student	98	24	17
Unemployed	18	19	5
Retired	17	2	2

Test whether there is a relationship between occupation and party preference.

Let's first just look at the data. Does it look like there is any sort of relationship? What should we calculate?

Example, χ^2

	PAN	PRI	PRD	
Employed	320	245	288	853
Student	98	24	17	139
Unemployed	18	19	5	42
Retired	17	2	2	21
	453	290	312	

	PAN	PRI	PRD	
Employed	38%	29%	34%	100
Student	71%	17%	12%	100
Unemployed	43%	45%	12%	100
Retired	81%	10%	10%	100

Example, χ^2

Answer

1. Hypotheses:
    - **Null**: There is no relationship;
    - **Alternative**: There is a relationship between occ. and

   party preference in Mexico.
2. Set $\alpha = 0.05$
3. Calculate a test statistic...

Example, χ^2

Table : Observed

	PAN	PRI	PRD
Employed	320	245	288
Student	98	24	17
Unemployed	18	19	5
Retired	17	2	2

Table : Expected

	PAN	PRI	PRD
Employed	366.26	234.47	252.26
Student	59.68	38.21	41.11
Unemployed	18.03	11.55	12.42
Retired	9.02	5.77	6.21

Example, χ^2

3 Calculate Test Statistic

$$\chi^2 = \sum \frac{(Obs - Exp)^2}{Exp} = 77.03323$$

4 Obtain p-value.

$$\chi_{3,2} = 77.03323$$

```
1-pchisq(77.03323,df=6) = 0
```

5 Conclude: Reject the null hypothesis that there is no relationship between party preference and employment status.

More Generally...

Though it is a blunt test, it *broadly* applicable.

- The case we examined, often called "Pearson's Chi Square", is most often applied to the relationship between two categorical variables.
- Can be generalized to test whether the sample comes from a particular underlying distribution.
- Consider 120 rolls of a die. We expect 20 1's, 20 2'2, and so on. We can calculate a chi-square statistic to see if our die is fair.

Again, we use:

$$\chi^2_{K-1} = \sum_{i=1}^{K} \frac{(O_k - E_k)^2}{E_k}$$

where K is the number of categories.

Play along at home...

Example

1	2	3	4	5	6
28	13	25	20	15	19
20	20	20	20	20	20

What do they win, Johnny...

Answer

$$\chi^2 = \frac{(28-20)^2}{20} + \frac{(13-20)^2}{20} + \frac{(25-20)^2}{20} + \frac{(20-20)^2}{20} + \frac{(15-20)^2}{20} + \frac{(19-20)^2}{20}$$
$$= 8$$
$$p - val = .15$$

Variance of a Sample

Do we need to only concern ourselves with the first central moment or independence? *of course not!*

- Assume sample drawn from a normal distribution.
- Null: $\sigma = \sigma_0$
- Alternative: $\sigma \neq \sigma_0$
- Test statistic: $\frac{(n-1)s^2}{\sigma_0} \sim \chi^2_{n-1}$

Difference Between Two Variances

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_A : \sigma_1^2 \neq \sigma_2^2$$
$$\frac{s_1^2}{s_2^2} \sim F_{n_1,n_2}$$

So, how do you think *this one* would proceed?

```
v1 <- 10
v2 <- 20
x1<-rnorm(100, mean = 0, sd=sqrt(v1))
x2<-rnorm(121, mean = 0, sd=sqrt(v2))
var.test(x1,x2)
```

Variance Test Example

Example

Is the Republican or Democratic party more cohesive ideologically? Instead of using the cohesion score, we can try using the test for differences in variances. Ideological scores are assigned to legislatures using Poole's WNOMINATE, and we obtain the following for a hypothetical Senate: 47 Republicans, $s_R = .15$, 52 Democrats, $s_D = .20$.

Variance Test Example

Answer

1. State null and alternative hypotheses:
   - H_0: The two populations have the same underlying variance in ideology: $\frac{\sigma_D^2}{\sigma_R^2} = 1$.
   - H_A: The two populations have different variances in ideology: $\frac{\sigma_D^2}{\sigma_R^2} \neq 1$.
2. Set $\alpha = .05$.
3. Calculate Test Statistic (express as a ratio ¿ 1):

$$F = \frac{s_D^2}{s_R^2} = \frac{.20^2}{.15^2} \sim 1.78$$

4. How likely? `2*(1-pf(df1=46,df2=52,1.78)) = .04510496`
5. Conclude: Reject the null...

Many,Many Others

We can test for:

- Medians
- Skew
- Kurtosis
- Variances
- Nonparametric sign test
- Range
- Same distribution
- Whatever you like...

1. Don't memorize any particular test.
2. Instead, understand the framework, check the assumptions, calculate a test-statistic, and identify the right test.

1. Review
    - Caputo
    - Normal Distribution
    - Chi-Squared & F Distributions
2. Hypothesis Testing
    - Theory
    - The Null and Alternative Hypotheses
    - Critical Regions
    - Test Statistics
    - The p-value
    - Core Questions/Issues
3. Hypothesis Tests: Practically
    - Fisher Hypothesis Test Setup
    - Test Statistics
4. Power

Power

We have constructed *all* of our tests in terms of α. Remind me what we're accomplishing here?

But, this is only *one* of the ways that our test could (with probability) be wrong.

Definition

The **Power**, frequently denoted $(1 - \beta)$ of a test is the probability that we correctly reject the null, *given the (unknown) true state of the world should reject the null.*

Canonically, this is shown with the following table:

Power

	Truth:	
Action	H_0	H_A
DN Reject	Correct	Type II Error "False Negative" β
Reject	Type I Error "False Positive" α	Correct

Power

Draw on Board...

1. Review
    - Caputo
    - Normal Distribution
    - Chi-Squared & F Distributions
2. Hypothesis Testing
    - Theory
    - The Null and Alternative Hypotheses
    - Critical Regions
    - Test Statistics
    - The p-value
    - Core Questions/Issues
3. Hypothesis Tests: Practically
    - Fisher Hypothesis Test Setup
    - Test Statistics
4. Power

Properties of Estimators

Many ways to skin a cat (...or to measure a parameter).

- **Central Moment**:
  - mean
  - median
  - mean(max,min)
- **Second Central Moment**
  - sd
  - median of $x_i - x_j$
  - sd(any two values)

Which is the "best" estimator?

Properties of Estimators

What do we want in an estimate? **1. Unbiasedness:**

- On average, it is right!
- $E(\hat{\theta} - \theta) = 0$
- max(x) versus mean(x)

This is for fixed n. There is also asymptotic unbiasedness, as n becomes large.

Properties of Estimators

2. Efficiency

- How close do we think we are?
- Mean-squared error: $MSE = E[(\hat{\theta} - \theta)^2]$
- More efficient = smaller MSE

Properties of Estimators

3. Consistency: As sample size gets bigger and bigger, the estimate approaches the true value.

- With an infinite sample size, we know everything.
- Limit

Properties of Estimators

4. Sufficiency: The estimate contain all the information about the parameter that is in the data.

- n or sample $\frac{n}{2}$?
- Mean versus median
- Learn more: Bickel and Doksum, *Mathematical Statistics: Basic Ideas and Selected Topics*

Compare Estimators

Consider Four Estimators of μ

- Mean
- Median
- Average of max and min
- Maximum

Compare Estimators When we actually use these estimators, can we talk intelligently about how *"good"* they are?