METIS

Introduction to **Hypothesis Testing**

Estimation vs. Inference

Estimation vs. Inference

- **Estimation** is the application of an algorithm

Estimation vs. Inference

- **Estimation** is the application of an algorithm
  - E.g. Taking an average: $\bar{X} = \sum_{i=1}^{N} x_i / n$

Estimation vs. Inference

- **Estimation** is the application of an algorithm
  - E.g. Taking an average: $\bar{X} = \sum_{i=1}^{N} x_i / n$

- **Inference** is putting an accuracy on our estimate

Estimation vs. Inference

- **Estimation** is the application of an algorithm
  - E.g. Taking an average: $\bar{X} = \sum_{i=1}^{N} x_i / n$
- **Inference** is putting an accuracy on our estimate
  - E.g. Standard error around average: $\left[\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n(n-1)}\right]^{1/2}$

Estimation vs. Inference

- **Estimation** is the application of an algorithm
  - E.g. Taking an average: $\bar{X} = \sum_{i=1}^{N} x_i / n$
- **Inference** is putting an accuracy on our estimate
  - E.g. Standard error around average: $\left[\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n(n-1)}\right]^{1/2}$
- (Possibly surprising) fact about inference: we use the same data we used for estimation to do inference!

Estimation vs. Inference

- **Estimation** is the application of an algorithm
  - E.g. Taking an average: $\bar{X} = \sum_{i=1}^{N} x_i / n$

- **Inference** is putting an accuracy on our estimate
  - E.g. Standard error around average: $\left[\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n(n-1)} \right]^{1/2}$

- Check for understanding: Can we name some other types of estimation or inference we've seen before?

Quick aside on philosophy of statistics (review)

Frequentist Statistics

- A **frequentist** is concerned with repeated observations in the limit

Frequentist Statistics

- A **frequentist** is concerned with repeated observations in the limit
- Processes may have true frequencies, but we're interested in **modeling probabilities as many, many (many!) repeats of an experiment**

Frequentist Statistics

- A **frequentist** is concerned with repeated observations in the limit
- Processes may have true frequencies, but we're interested in **modeling probabilities as many, many (many!) repeats of an experiment**

1. Derive the probabilistic property of a procedure
2. Apply the probability directly to our observed data

Bayesian Statistics

- A **Bayesian** describes parameters by probability distributions

Bayesian Statistics

- A **Bayesian** describes parameters by probability distributions
- Before seeing any data, a **prior distribution** (based on the experimenters' belief) is formulated

Bayesian Statistics

- A **Bayesian** describes parameters by probability distributions
- Before seeing any data, a **prior distribution** (based on the experimenters' belief) is formulated
- This prior distribution is then updated after seeing data (a sample from the distribution)

Bayesian Statistics

- A **Bayesian** describes parameters by probability distributions
- Before seeing any data, a **prior distribution** (based on the experimenters' belief) is formulated
- This prior distribution is then updated after seeing data (a sample from the distribution)
- After updating, the distribution is called the **posterior distribution**

Frequentist vs. Bayesian

- We use much of the same math and the same formulas in both frequentist and Bayesian statistics
- What differs is the interpretation
- We will point out the difference in interpretation where appropriate

Back to our regularly scheduled hypothesis testing lecture…

Hypothesis Testing

- A hypothesis is a statement about a population parameter

Hypothesis Testing

- A hypothesis is a statement about a population parameter

- In **hypothesis testing**, we create two hypotheses
  - The **null hypothesis (H_0)**
  - And the **alternative hypothesis (H_1 or H_A)**

Hypothesis Testing

- A hypothesis is a statement about a population parameter

- In **hypothesis testing**, we create two hypotheses
    - The **null hypothesis (H_0)**
    - And the **alternative hypothesis (H_1 or H_A)**

- We decide which one to call the null depending on how our problem is set up

Hypothesis Testing

- A hypothesis is a statement about a population parameter

- In **hypothesis testing**, we create two hypotheses
  - The **null hypothesis (H_0)**
  - And the **alternative hypothesis (H_1 or H_A)**

- Check for understanding: Can we give some examples of a null and alternative hypothesis from OLS?

Decision Rules: Frequentist Interpretation

- A hypothesis testing procedure gives us a rule to decide:
  - For which values of our test statistic do we accept H_0
  - For which values of our test statistic do we reject H_0 and accept H_1

Decision Rules: Frequentist Interpretation

- A hypothesis testing procedure gives us a rule to decide:
  - For which values of our test statistic do we accept H_0
  - For which values of our test statistic do we reject H_0 and accept H_1
- You may hear some people say that you can reject H_0 but that you **never accept H_1** — for our purposes, this doesn't matter so much: we're using hypothesis testing in order to decide which of two paths to take in our project

Decision Rules: Bayesian Interpretation

- In the **Bayesian interpretation** (example to follow), we don't get a decision boundary
- Instead, we get updated (**posterior**) probabilities

Likelihood Ratio Test

Coin Tossing Example

- You have two coins
  - Coin 1 has a .7 probability of coming up heads
  - Coin 2 has a .5 probability of coming up heads
- Pick one coin without looking
- Toss the coin 10 times and record # heads
- **Question:** Given the number of heads you see, which of the two coins did you toss?

Coin Tossing Example: Likelihood Ratio

- Given what we know about coins 1 and 2, we can make a table of the probability of seeing *x* heads out of 10 tosses

x	0	1	2	3	4	5	6	7	8	9	10
Coin 1 **P(Head) = .5**	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.001
Coin 2 **P(Head) = .7**	0.000	0.0001	0.001	0.009	0.037	0.103	0.200	0.267	0.236	0.121	0.028

- We can now calculate a **likelihood ratio**, based on the number of heads we saw when tossing the unidentified coin

Coin Tossing Example: Likelihood Ratio

x	0	1	2	3	4	5	6	7	8	9	10
Coin 1 P(Head) = .5	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.001
Coin 2 P(Head) = .7	0.000	0.0001	0.001	0.009	0.037	0.103	0.200	0.267	0.236	0.121	0.028

- Let’s say we saw three heads

Coin Tossing Example: Likelihood Ratio

x	0	1	2	3	4	5	6	7	8	9	10
Coin 1 P(Head) = .5	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.001
Coin 2 P(Head) = .7	0.000	0.0001	0.001	0.009	0.037	0.103	0.200	0.267	0.236	0.121	0.028

- Let's say we saw three heads
  - $P_1(3)/P_2(3) = 0.117/0.009 = 13$
  - Coin 1 was 13 times more likely to give us the output 3 heads than coin 2 was

Coin Tossing Example: Likelihood Ratio

x	0	1	2	3	4	5	6	7	8	9	10
Coin 1 P(Head) = .5	0.001	0.010	0.044	0.117	0.205	0.246	0.205	0.117	0.044	0.010	0.001
Coin 2 P(Head) = .7	0.000	0.0001	0.001	0.009	0.037	0.103	0.200	0.267	0.236	0.121	0.028

- Let's say we saw three heads
  - $P_1(3)/P_2(3) = 0.117/0.009 = 13$
  - Coin 1 was 13 times more likely to give us the output 3 heads than coin 2 was
  - We call this the **likelihood ratio**

Hypothesis Testing: Bayesian Interpretation

- In the Bayesian interpretation of hypothesis testing, we need to have priors for each hypothesis
    - In this case, we randomly chose the coin to flip
    - $P(H_1 = \text{we chose coin 1}) = 1/2$ and
    - $P(H_2 = \text{we chose coin 2}) = 1/2$
    - … we have no way, before seeing the data, to determine the coin that was chosen, so just assign 1/2 to each

Hypothesis Testing: Bayesian Interpretation

- Priors: $P(H_1) = 1/2 = P(H_2) = 1/2$
- Updating priors after seeing the data (e.g. $x = 3$ heads)
    - $$P(H_1 | x) = \frac{P(x | H_1)P(H_1)}{P(x)}$$
    - The priors ($P(H_0)/P(H_1)$) are multiplied by the likelihood ratio, which does not depend on the priors
    - The likelihood ratio tells us how we should update the priors in reaction to seeing a given set of data!

Hypothesis Testing: Bayesian Interpretation

- Priors: $P(H_1) = 1/2 = P(H_2) = 1/2$
- Updating priors after seeing the data (e.g. $x = 3$ heads)
  - $$P(H_1 | x) = \frac{P(x | H_1)P(H_1)}{P(x)}$$
  - We can write out the ratio: $$\frac{P(H_1 | x)}{P(H_2 | x)} = \frac{P(H_1)P(x | H_1)}{P(H_2)P(x | H_2)}$$
  - The priors ($P(H_0)/P(H_1)$) are multiplied by the likelihood ratio, which does not depend on the priors
  - The likelihood ratio tells us how we should update the priors in reaction to seeing a given set of data!

Hypothesis Testing: Bayesian Interpretation

- Priors: $P(H_1) = 1/2 = P(H_2) = 1/2$
- Updating priors after seeing the data (e.g. $x = 3$ heads)
  - $$P(H_1 | x) = \frac{P(x | H_1)P(H_1)}{P(x)}$$
  - We can write out the ratio: $\dfrac{P(H_1 | x)}{P(H_2 | x)} = \dfrac{P(H_1)P(x | H_1)}{P(H_2)P(x | H_2)}$
  - The priors ($P(H_1)/P(H_2)$) are multiplied by the likelihood ratio, which does not depend on the priors
  - The likelihood ratio tells us how we should update the priors in reaction to seeing a given set of data!

Neyman-Pearson Interpretation

- Neyman-Pearson paradigm (1933) is non-Bayesian
- Gives up or down vote on H_0 vs H_1

Neyman-Pearson Interpretation

- Neyman-Pearson paradigm (1933) is non-Bayesian
- Gives up or down vote on H_0 vs H_1
- Terminology:

		Decision	
		Accept H_0	**Reject H_0**
Truth	**H_0**	Correct	Type I error
	H_1	Type II error (β)	Correct

  - Power of a test = 1 - P(type II error)

Neyman-Pearson Interpretation

- The likelihood ratio is called a **test statistic**: we use it to decide whether to accept/reject H_0

Neyman-Pearson Interpretation

- The likelihood ratio is called a **test statistic**: we use it to decide whether to accept/reject H_0
  - The set of values of the test statistic that lead to rejection of H_0 is called the **rejection region**
  - The set of values of the test statistic that lead to acceptance of H_0 is called the **acceptance region**

Neyman-Pearson Interpretation

- The likelihood ratio is called a **test statistic**: we use it to decide whether to accept/reject H_0
  - The set of values of the test statistic that lead to rejection of H_0 is called the **rejection region**
  - The set of values of the test statistic that lead to acceptance of H_0 is called the **acceptance region**
- The test statistic's distribution when the null is true is called the **null distribution**

Neyman-Pearson Interpretation of Coin Tossing Example

- In the coin tossing example:
    - H_0: the coin is fair and $P(H) = .5$
    - H_1: the coin is unfair and $P(H) > .7$
- Check for understanding: How can we test the null hypothesis? Take a minute and write it out!

Neyman-Pearson Interpretation of Coin Tossing Example

- In the coin tossing example:
  - H_0: the coin is fair and P(H) = .5
  - H_1: the coin is unfair and P(H) > .7
- Test the null hypothesis
  - We know H_0 is distributed binom(10, .5)
  - Choose a p-value cutoff (more on p-values soon), say .05
  - Calculate the CDF of 3 positives from a binom(10, .5)
  - = 82%
  - This is > 5%, so we don't reject H_0

Significance Level & P-Values

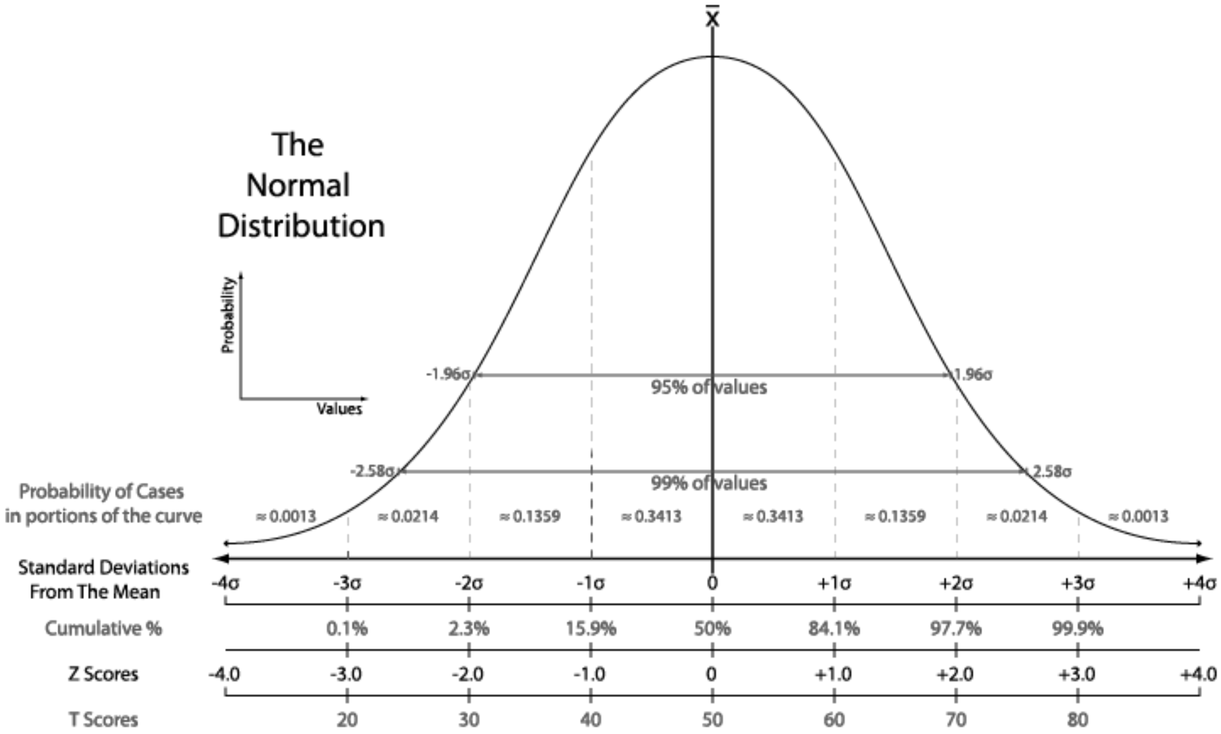
Significance Level and P-Values

- We know the distribution of our null hypothesis
- To get a rejection region, we calculate our test statistic
- We will choose, before testing our data, the level at which we will reject our null hypothesis

Significance Level and P-Values

- A significance level (α) is a probability threshold below which the null hypothesis will be rejected
- We must choose an α **before** computing our test statistic! If we don't, we might be accused of p-hacking
- Choosing α is somewhat arbitrary, but often .01 or .05
- The **p-value** is the smallest significance level at which the null hypothesis would be rejected
- Fisher interpretation of p-value: the probability under the null of a result as or more extreme than actually observed
- The **confidence interval**: the values of our statistic for which we accept the null

Significance Level and P-Values



F-Statistic

F-Statistic

- H_0: the data can be modeled by setting all betas to zero
- Reject the null if the p-value is small enough

OLS Regression Results

Dep. Variable:	Y	**R-squared:**	0.733
Model:	OLS	**Adj. R-squared:**	0.663
Method:	Least Squares	**F-statistic:**	10.50
Date:	Mon, 08 Oct 2018	**Prob (F-statistic):**	1.24e-05
Time:	20:16:45	**Log-Likelihood:**	-97.250
No. Observations:	30	**AIC:**	208.5
Df Residuals:	23	**BIC:**	218.3
Df Model:	6		
Covariance Type:	nonrobust		

QUESTIONS?