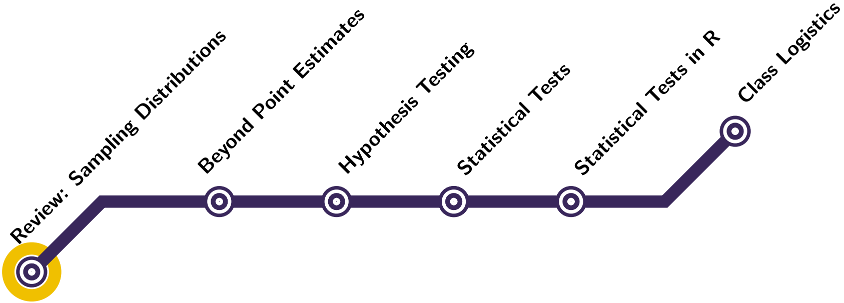
Today's Roadmap



Foundations for Statistical Inference

- We carry out an experiment and get a random sample of the underlying population
- Data are the values in the sample
- Our aim is to infer the population probability distribution (parameters) from the data we observe in the sample
- Our assumption is that that samples behave approximately as the population

Foundations for Statistical Inference

- We carry out an experiment and get a random sample of the underlying population
- Data are the values in the sample
- Our aim is to infer the population probability distribution (parameters) from the data we observe in the sample
- Our assumption is that that samples behave approximately as the population

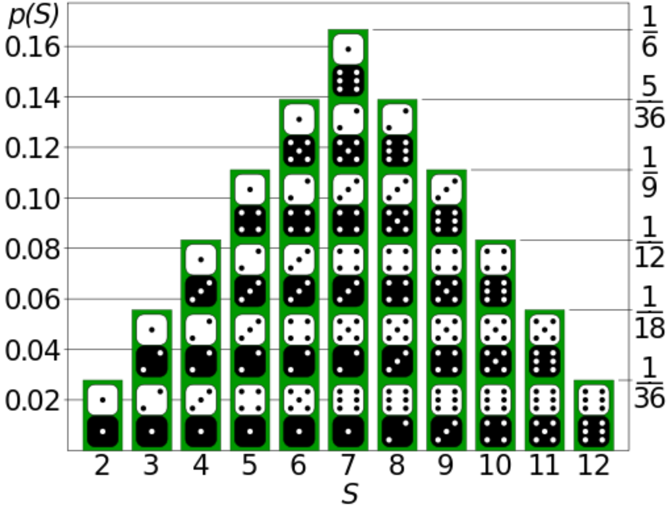
This is statistical inference!

Introduction to Random Variables

- Consider some experiment for which the sample space is S
- A real-valued function that is defined on S is called a random variable
- A random variable X is a random process with a numerical outcome.

Random Variables

Probability distribution for the sum of two dice:



Allows for computation of probabilities of events.

Types of Random Variables

- Discrete: finite or countable list of possible values; probability mass function (pmf) assigns a probability to each value

$$f_X(x) = Pr(X = x)$$

- Continuous: taking any numerical value in an interval; probability density function (pdf) assigns a probabilities to intervals

$$Pr(a \leq X \leq b) = \int_a^b f_X(x)dx$$

Statistical Inference: Point Estimates

- In many situations we want to estimated the population mean based on a sample.
- What should we do? Use the sample mean!
- The sample mean $\overline{x}$ is called a point estimate of the population mean.

Point Estimates

- Point estimates from a sample may be used to estimate population parameters
- Point estimates are not exact: they vary from one sample to another
- We can quantify the standard error for the point estimate

Point Estimates

- Point estimates from a sample may be used to estimate population parameters
- Point estimates are not exact: they vary from one sample to another
- We can quantify the standard error for the point estimate

The sampling distribution represents the distribution of the point estimates based on samples of a fixed size from a certain population.

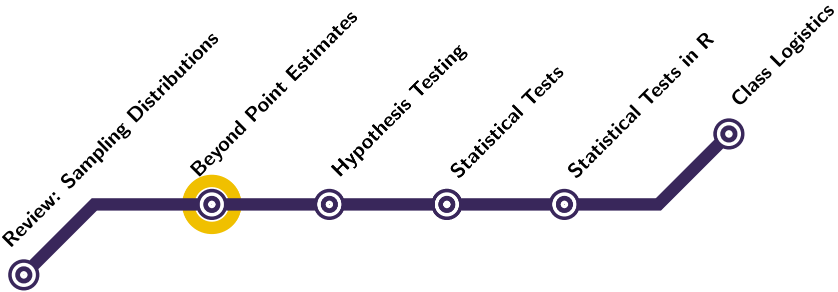
The Central Limit Theorem

Given certain conditions, the mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution.

The Central Limit Theorem

Given certain conditions, the mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution.

Today's Roadmap



Beyond Point Estimates

- A point estimate provides a single best guess for the parameter

Beyond Point Estimates

- A point estimate provides a single best guess for the parameter
- Nest step: provide a plausible range of values for the parameter

Beyond Point Estimates

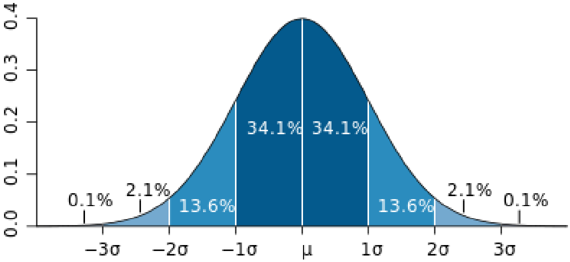
- A point estimate provides a single best guess for the parameter
- Nest step: provide a plausible range of values for the parameter
- A plausible range of values for the population parameter is called a confidence interval

Confidence Intervals for μ

- The Central Limit Thm tells us that the sampling distribution of our estimate of the mean, $\overline{x}$ is approximately normal.

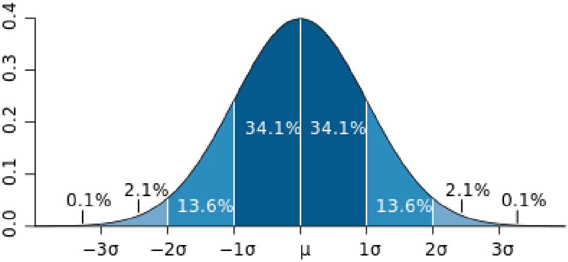
Confidence Intervals for μ

- The Central Limit Thm tells us that the sampling distribution of our estimate of the mean, $\overline{x}$ is approximately normal.
- 95% of the time the estimate will be within 2 standard errors of the parameter



Confidence Intervals for μ

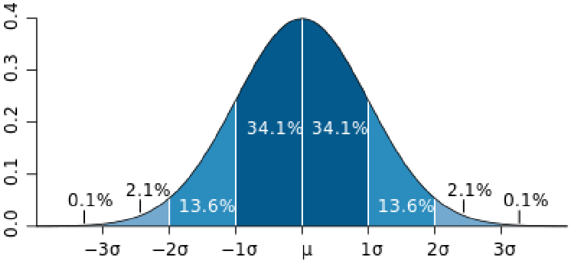
- The Central Limit Thm tells us that the sampling distribution of our estimate of the mean, $\overline{x}$ is approximately normal.
- 95% of the time the estimate will be within 2 standard errors of the parameter



$$\text{point estimate} \pm 2 \times \text{SE}$$

Confidence Intervals for μ

- The Central Limit Thm tells us that the sampling distribution of our estimate of the mean, $\overline{x}$ is approximately normal.
- 95% of the time the estimate will be within 2 standard errors of the parameter



$$\text{point estimate} \pm 1.96 \times \text{SE}$$

Confidence Intervals for μ

$$\Pr(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = \Phi(2) - \Phi(-2)$$
$$\approx 0.9545$$

Confidence Intervals for μ

- We can construct a confidence interval for the mean for any confidence level
- If the point estimate follows the normal model with standard error SE, then a confidence interval for the population parameter is:

  point estimate $\pm$ z $\times$ SE

  where z corresponds to the confidence level selected.

Generalized Confidence Intervals

- We choose and interval (A, B) that has a high probability of containing θ.

Generalized Confidence Intervals

- We choose and interval (A, B) that has a high probability of containing θ.
- A $1 - \alpha$ confidence interval for a parameter θ is an interval $C_n = (a, b)$ where $a = a(X_1, \dots, X_n)$ and $b = b(X_1, \dots, X_n)$ are functions of the data such that

$$Pr_\theta(\theta \in C_n) \geq 1 - \alpha$$

for all $\theta \in \Theta$.

Interpreting Confidence Intervals

- Commonly, people use 95% confidence intervals, which corresponds to $\alpha = 0.05$.

Interpreting Confidence Intervals

- Commonly, people use 95% confidence intervals, which corresponds to $\alpha = 0.05$.
- It is **not** correct to say that θ lies in the interval (a, b) with probability $1 - \alpha$.

Interpreting Confidence Intervals

- On day 1 you collect data and construct a 95 percent confidence interval for some parameter θ_1.

Interpreting Confidence Intervals

- On day 1 you collect data and construct a 95 percent confidence interval for some parameter θ_1.
- On day 2 you collect data and construct a 95 percent confidence interval for some unrelated parameter θ_2.

Interpreting Confidence Intervals

- On day 1 you collect data and construct a 95 percent confidence interval for some parameter θ_1.
- On day 2 you collect data and construct a 95 percent confidence interval for some unrelated parameter θ_2.
- On day 3 you collect data and construct a 95 percent confidence interval for some unrelated parameter θ_3.

Interpreting Confidence Intervals

- On day 1 you collect data and construct a 95 percent confidence interval for some parameter θ_1.
- On day 2 you collect data and construct a 95 percent confidence interval for some unrelated parameter θ_2.
- On day 3 you collect data and construct a 95 percent confidence interval for some unrelated parameter θ_3.
- And so on for a sequence of unrelated parameters $\theta_1, \theta_, \ldots$, then 95 percent of your confidence intervals will trap the true parameter value.

Confidence Intervals Example

- Frequently we see opinion polls reported in the newspaper.

Confidence Intervals Example

- Frequently we see opinion polls reported in the newspaper.
- For example, "83 percent of the population support X".

Confidence Intervals Example

- Frequently we see opinion polls reported in the newspaper.
- For example, "83 percent of the population support X".
- Hopefully you also see a statement such as "this poll is accurate with within 4 points 95% of the time.

Confidence Intervals Example

- Frequently we see opinion polls reported in the newspaper.
- For example, "83 percent of the population support X".
- Hopefully you also see a statement such as "this poll is accurate with within 4 points 95% of the time.
- This is a confidence interval on their estimate, 83 ± 4 is the 95% CI.

Normal-based Confidence Intervals

- In other words $\hat{\theta}_n \approx N(\theta, \hat{se}^2)$

Normal-based Confidence Intervals

- In other words $\hat{\theta}_n \approx N(\theta, \hat{se}^2)$
- Then we know that

$$Z_n = \frac{\hat{\theta}_n - \theta}{\hat{se}} \approx N(0,1)$$

Normal-based Confidence Intervals

- Consider constructing a 95% CI for θ.

Normal-based Confidence Intervals

- Consider constructing a 95% CI for θ.
- For a standard normal random variable, we can construct an interval such that

Normal-based Confidence Intervals

- Consider constructing a 95% CI for θ.
- For a standard normal random variable, we can construct an interval such that

$$1 - \alpha = Pr(-z_{\alpha/2} < Z < z_{\alpha/2})$$

- where $z_{\alpha/2} = \Phi^{-1}(1 - (\alpha/2))$

Normal-based Confidence Intervals

- Consider constructing a 95% CI for θ.
- For a standard normal random variable, we can construct an interval such that

$$1 - \alpha = Pr(-z_{\alpha/2} < Z < z_{\alpha/2})$$

- where $z_{\alpha/2} = \Phi^{-1}(1 - (\alpha/2))$
- Then,

$$\begin{aligned} 1 - \alpha = & \quad Pr(-z_{\alpha/2} < Z < z_{\alpha/2}) \\ = & \quad Pr(-z_{\alpha/2} < \tfrac{\hat{\theta}_n - \theta}{\hat{se}} < z_{\alpha/2}) \\ = & \quad Pr(\hat{\theta}_n - z_{\alpha/2}\hat{se} < \theta < \hat{\theta}_n + z_{\alpha/2}\hat{se}) \end{aligned}$$

Normal-based Confidence Intervals

- For a 95% CI and $z_{\alpha/2} = 1.96$ we have a CI for θ that is

$$C_n = (\hat{\theta}_n - 1.96\hat{se},\ \hat{\theta}_n + 1.96\hat{se})$$

Normal-based Confidence Intervals Example

- Let $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$
- Let $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. $\hat{\text{se}} = \sqrt{\hat{p}_n(1-\hat{p}_n)/n}$

Normal-based Confidence Intervals Example

- Let $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$
- Let $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. $\hat{\text{se}} = \sqrt{\hat{p}_n(1-\hat{p}_n)/n}$
- Then by the CLT $\hat{p}_n \approx N(p, \hat{\text{se}})$.

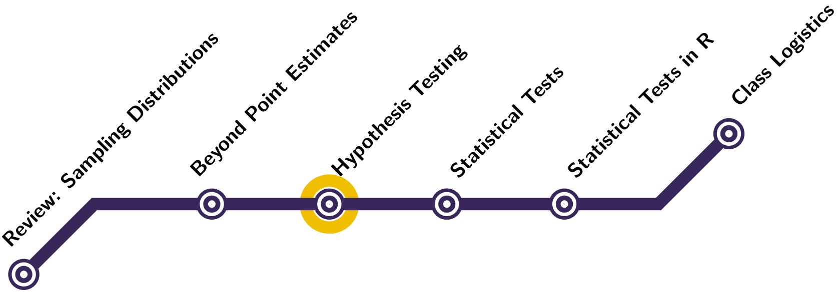
Normal-based Confidence Intervals Example

- Let $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$
- Let $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. $\hat{\text{se}} = \sqrt{\hat{p}_n(1-\hat{p}_n)/n}$
- Then by the CLT $\hat{p}_n \approx N(p, \hat{\text{se}})$.
- So an approximate $1-\alpha$ CI for p is

$$\hat{p_n} \pm z_{\alpha/2}\sqrt{\frac{\hat{p_n}(1-\hat{p_n})}{n}}$$

Break

Today's Roadmap



Hypothesis Testing

- Let Θ be the parameter space, all possible values of θ

Hypothesis Testing

- Let Θ be the parameter space, all possible values of θ
- Let Θ_0 and Θ_1 partition the space, $\Theta_0 \cup \Theta_1 = \Theta$

Hypothesis Testing

- Let Θ be the parameter space, all possible values of θ
- Let Θ_0 and Θ_1 partition the space, $\Theta_0 \cup \Theta_1 = \Theta$
- So we with to test

$$H_0 : \theta \in \Theta_0$$

versus

$$H_1 : \theta \in \Theta_1$$

Hypothesis Testing

- Let Θ be the parameter space, all possible values of θ
- Let Θ_0 and Θ_1 partition the space, $\Theta_0 \cup \Theta_1 = \Theta$
- So we with to test

$$H_0 : \theta \in \Theta_0$$

versus

$$H_1 : \theta \in \Theta_1$$

- We call H_0 the null hypotheses and H_1 the alternative hypothesis

Hypothesis Testing Intuition

- Let $X_1, \ldots, X_n \sim \mathrm{Bernoulli}(p)$ be n independent coin flips.

Hypothesis Testing Intuition

- Let $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$ be n independent coin flips.
- Suppose we want to test if the coin is fair, H_0: fair.
- Let H_1 denote the hypothesis that this coin in not fair.

Hypothesis Testing Intuition

- Let $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$ be n independent coin flips.
- Suppose we want to test if the coin is fair, H_0: fair.
- Let H_1 denote the hypothesis that this coin in not fair.
- We have

$$H_0 : p = 1/2$$
$$H_1 : p \neq 1/2$$

Hypothesis Testing Intuition

- Let $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$ be n independent coin flips.
- Suppose we want to test if the coin is fair, H_0: fair.
- Let H_1 denote the hypothesis that this coin in not fair.
- We have

$$H_0 : p = 1/2$$

$$H_1 : p \neq 1/2$$

- It seems reasonable to reject H_0 if $T = |\hat{p}_n - (1/2)|$ is large
- How large T should be to reject H_0?

Hypothesis Testing

- A procedure for deciding whether to reject H_0 is called a test procedure or simply test

Hypothesis Testing

- A procedure for deciding whether to reject H_0 is called a test procedure or simply test
- Hypothesis testing is like a legal trial

Hypothesis Testing

- A procedure for deciding whether to reject H_0 is called a test procedure or simply test
- Hypothesis testing is like a legal trial
- We assume someone is innocent unless proven guilty

Hypothesis Testing

- A procedure for deciding whether to reject H_0 is called a test procedure or simply test
- Hypothesis testing is like a legal trial
- We assume someone is innocent unless proven guilty
- Similarly we retain H_0 unless there is strong evidence to reject H_0.

Hypothesis Testing

- In such a framework there are two types of errors we can make:
  1. We can reject H_0 when H_0 is true - this is called type I error
  2. We can fail to reject H_0 when H_1 is true - this is called type II error

	H_0 True	H_0 False
Reject H_0	Type I Error	Correct
Retain H_0	Correct	Type II Error

Hypothesis Testing

- Suppose we are trying to determine whether or not to reject the null hypothesis:

$$H_0 : \theta \in \Theta_0$$
$$H_1 : \theta \in \Theta_1$$

Hypothesis Testing

- Suppose we are trying to determine whether or not to reject the null hypothesis:

$$H_0 : \theta \in \Theta_0$$
$$H_1 : \theta \in \Theta_1$$

- We observe a random sample $X_1, \ldots, X_n$ drawn from some distribution that is characterized by an unknown parameter θ.

Hypothesis Testing

- Suppose we are trying to determine whether or not to reject the null hypothesis:

$$H_0 : \theta \in \Theta_0$$
$$H_1 : \theta \in \Theta_1$$

- We observe a random sample $X_1, \ldots, X_n$ drawn from some distribution that is characterized by an unknown parameter θ.
- Partitioning the sample space of all possible outcome of the random sample into two subsets

Hypothesis Testing

- Suppose we are trying to determine whether or not to reject the null hypothesis:

$$H_0 : \theta \in \Theta_0$$
$$H_1 : \theta \in \Theta_1$$

- We observe a random sample $X_1, \ldots, X_n$ drawn from some distribution that is characterized by an unknown parameter θ.
- Partitioning the sample space of all possible outcome of the random sample into two subsets
- One subset contains all the values of the observed data for which we reject H_0 and the other contains all the values of the observed data for which we retain H_0.

Hypothesis Testing

- Suppose we are trying to determine whether or not to reject the null hypothesis:

$$H_0 : \theta \in \Theta_0$$
$$H_1 : \theta \in \Theta_1$$

- We observe a random sample $X_1, \ldots, X_n$ drawn from some distribution that is characterized by an unknown parameter θ.
- Partitioning the sample space of all possible outcome of the random sample into two subsets
- One subset contains all the values of the observed data for which we reject H_0 and the other contains all the values of the observed data for which we retain H_0.
- The subset for which H_0 will be rejected is called the critical or rejection region, R.

Hypothesis Testing

- In most hypotheses testing problems the critical region is defined by a test statistic, $T = r(\mathbf{X})$.

Hypothesis Testing

- In most hypotheses testing problems the critical region is defined by a test statistic, $T = r(\mathbf{X})$.
- Usually the rejection region has the form

$$R = \{x : T(\mathbf{X}) > c\}$$

Hypothesis Testing

- In most hypotheses testing problems the critical region is defined by a test statistic, $T = r(\mathbf{X})$.
- Usually the rejection region has the form

$$R = \{x : T(\mathbf{X}) > c\}$$

- where T is the test statistic and c is called the critical value.

Hypothesis Testing

- In most hypotheses testing problems the critical region is defined by a test statistic, $T = r(\mathbf{X})$.
- Usually the rejection region has the form

$$R = \{x : T(\mathbf{X}) > c\}$$

- where T is the test statistic and c is called the critical value.
- So most tests are of a form "Reject H_0 if T $\geq$ c"

Hypothesis Testing

- In most hypotheses testing problems the critical region is defined by a test statistic, $T = r(\mathbf{X})$.
- Usually the rejection region has the form

$$R = \{x : T(\mathbf{X}) > c\}$$

- where T is the test statistic and c is called the critical value.
- So most tests are of a form "Reject H_0 if T $\geq$ c"
- The problem of hypothesis testing is then to find the right test and right value of the c.

Hypothesis Testing

- A hypothesis test of the form

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

is called a two-sided test.

Hypothesis Testing

- A hypothesis test of the form

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta \neq \theta_0$$

is called a two-sided test.

- A hypothesis test of the form

$$H_0 : \theta \leq \theta_0$$
$$H_1 : \theta > \theta_0$$

or

$$H_0 : \theta \geq \theta_0$$
$$H_1 : \theta < \theta_0$$

is called a two-sided test.

p-values

- Reporting that we have rejected the null hypothesis is not super informative

p-values

- Reporting that we have rejected the null hypothesis is not super informative
- Instead, we could ask for every level of α whether the test rejects at that level

p-values

- Reporting that we have rejected the null hypothesis is not super informative
- Instead, we could ask for every level of α whether the test rejects at that level
- If we reject the null at some level α then we will also reject for another $\alpha' > \alpha$

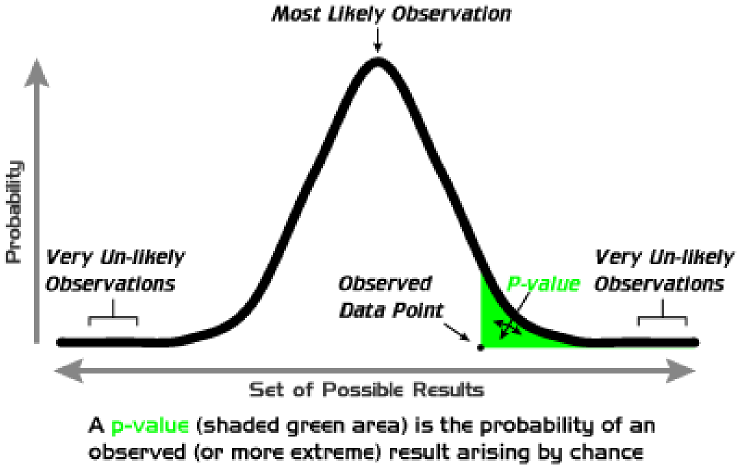
p-values

- Reporting that we have rejected the null hypothesis is not super informative
- Instead, we could ask for every level of α whether the test rejects at that level
- If we reject the null at some level α then we will also reject for another $\alpha' > \alpha$
- So there is a smallest α at which the test rejects, we call this number the p-value

p-values

The p-value is the probability (under H_0) of observing a value of the test statistic the same as or more extreme than what was actually observed.

p-values

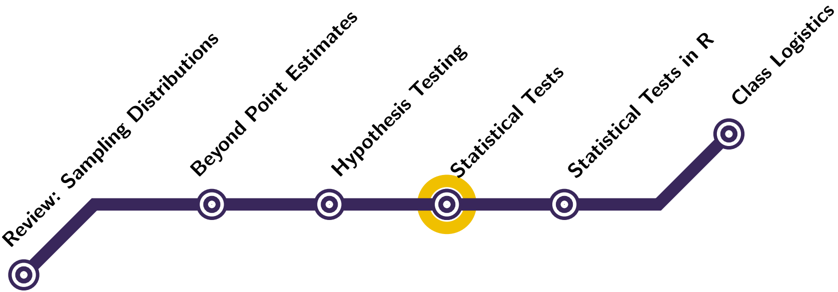


Most Likely Observation
Probability
Very Un-likely Observations
Observed Data Point
P-value
Very Un-likely Observations
Set of Possible Results
A p-value (shaded green area) is the probability of an observed (or more extreme) result arising by chance

p-values

- We often say something is statistically significant, but note that something may be significant even if the actual effect is small.
- In this way it may be that a "significant result" has little real world effect, scientific significance, etc.

Today's Roadmap



Statistical Tests: Numerical Data

1. Determine which point estimate or test statistic is useful.
2. Identify an appropriate sampling distribution for the point estimate or test statistic.
3. Apply the previous ideas - CIs and hypothesis testing

t Tests

- Suppose $X_1, \ldots, X_n$ form a random sample from a normal distribution with mean μ and variance σ^2 both unknown.

t Tests

- Suppose $X_1, \dots, X_n$ form a random sample from a normal distribution with mean μ and variance σ^2 both unknown.
- Consider the following one-sided hypothesis test:

$$H_0 : \mu \leq \mu_0$$
$$H_1 : \mu > \mu_0$$

t Tests

- Suppose $X_1, \ldots, X_n$ form a random sample from a normal distribution with mean μ and variance σ^2 both unknown.
- Consider the following one-sided hypothesis test:

$$H_0 : \mu \leq \mu_0$$
$$H_1 : \mu > \mu_0$$

- Consider the statistic,

$$T = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S}$$

where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X}_n)^2$.

t Tests

- Suppose $X_1, \ldots, X_n$ form a random sample from a normal distribution with mean μ and variance σ^2 both unknown.
- Consider the following one-sided hypothesis test:

$$H_0 : \mu \leq \mu_0$$
$$H_1 : \mu > \mu_0$$

- Consider the statistic,

$$T = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S}$$

where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$.
- When $\mu = \mu_0$ the distribution of T will be a t distribution with $n-1$ degrees of freedom.

t Tests

- Suppose $X_1, \dots, X_n$ form a random sample from a normal distribution with mean μ and variance σ^2 both unknown.
- Consider the following one-sided hypothesis test:

$$H_0 : \mu \leq \mu_0$$
$$H_1 : \mu > \mu_0$$

- Consider the statistic,

$$T = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S}$$

where $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$.

- When $\mu = \mu_0$ the distribution of T will be a t distribution with $n-1$ degrees of freedom.
- Then we reject H_0 when $T \geq t_{n-1}^{-1}(1-\alpha)$, the $1-\alpha$ quantile of the t distribution with $n-1$ degrees of freedom

t Tests Activity

- A random sample of students at State University were asked: "How many hours TV do you watch a day in average?"

t Tests Activity

- A random sample of students at State University were asked: "How many hours TV do you watch a day in average?"
- The sample data resulted in the following summary statistics: $\overline{X}_n = 2.09$, $S = 1.644$ and $n = 175$.

t Tests Activity

- A random sample of students at State University were asked: "How many hours TV do you watch a day in average?"
- The sample data resulted in the following summary statistics: $\overline{X}_n = 2.09$, $S = 1.644$ and $n = 175$.
- Is this sufficient evidence (significance 0.05) that to conclude that students at State University watch in average more than 2 hours TV per day?

t Tests Activity

- A random sample of students at State University were asked: "How many hours TV do you watch a day in average?"
- The sample data resulted in the following summary statistics: $\overline{X}_n = 2.09$, $S = 1.644$ and $n = 175$.
- Is this sufficient evidence (significance 0.05) that to conclude that students at State University watch in average more than 2 hours TV per day?
- Assume the number of hours of TV watched is normally distributed.

[Example from Dr. Karen Buro]

t Tests Activity

- First let's construct the hypothesis that this question is asking us to evaluate:

$$H_0 : \mu \leq 2$$
$$H_1 : \mu > 2$$

- The one-sample t test statistic is:

$$t_0 = \frac{2.09 - 2}{1.644/\sqrt{175}} = 0.7242$$

which we compare to the quantiles of a t_{174} distribution.

- To compute the p-value we have

$$\text{p} - \text{value} = Pr(t > t_0) = Pr(t > 0.7242) > 0.1$$

- So since the p-value is greater than the α level we are going to use, we retain H_0

t Tests Continued

- Suppose $X_1, \ldots, X_n$ form a random sample from a normal distribution with mean μ and variance σ^2 both unknown.
- Consider the following two-sided hypothesis test:

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

- Again we have the test statistic

$$T = \frac{\sqrt{n}(\overline{X}_n - \mu_0)}{S}$$

- Now, we have a test that rejects H_0 if $|T| \geq t_{n-1}^{-1}(1 - \alpha/2)$.

Two Sample t Tests

- Suppose $X_1, \ldots, X_m$ form a random sample from a normal distribution with mean μ_1 and variance σ^2 both unknown.
- And suppose $Y_1, \ldots, Y_n$ form a random sample from a normal distribution with mean μ_2 and variance σ^2 both unknown.

Two Sample t Tests

- Suppose $X_1, \ldots, X_m$ form a random sample from a normal distribution with mean μ_1 and variance σ^2 both unknown.
- And suppose $Y_1, \ldots, Y_n$ form a random sample from a normal distribution with mean μ_2 and variance σ^2 both unknown.
- Consider the following hypothesis test, at a significance level of α

$$H_0 : \mu_1 \leq \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

Two Sample t Tests

- Suppose $X_1, \ldots, X_m$ form a random sample from a normal distribution with mean μ_1 and variance σ^2 both unknown.
- And suppose $Y_1, \ldots, Y_n$ form a random sample from a normal distribution with mean μ_2 and variance σ^2 both unknown.
- Consider the following hypothesis test, at a significance level of α

$$H_0 : \mu_1 \leq \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

- Define $S_X^2 = \sum_{i=1}^{m}(X_i - \overline{X}_m)^2$ and $S_Y^2 = \sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2$

Two Sample t Tests

- Suppose $X_1, \ldots, X_m$ form a random sample from a normal distribution with mean μ_1 and variance σ^2 both unknown.
- And suppose $Y_1, \ldots, Y_n$ form a random sample from a normal distribution with mean μ_2 and variance σ^2 both unknown.
- Consider the following hypothesis test, at a significance level of α

$$H_0 : \mu_1 \leq \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

- Define $S_X^2 = \sum_{i=1}^{m}(X_i - \overline{X}_m)^2$ and $S_Y^2 = \sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2$
- Our test statistic is then,

$$T = \frac{\sqrt{m+n-2}(\overline{X}_m - \overline{Y}_n)}{\sqrt{(\frac{1}{m} + \frac{1}{n})}\sqrt{(S_X^2 + S_Y^2)}}$$

Two Sample t Tests

- Suppose $X_1, \ldots, X_m$ form a random sample from a normal distribution with mean μ_1 and variance σ^2 both unknown.
- And suppose $Y_1, \ldots, Y_n$ form a random sample from a normal distribution with mean μ_2 and variance σ^2 both unknown.
- Consider the following hypothesis test, at a significance level of α

$$H_0 : \mu_1 \leq \mu_2$$
$$H_1 : \mu_1 > \mu_2$$

- Define $S_X^2 = \sum_{i=1}^{m}(X_i - \overline{X}_m)^2$ and $S_Y^2 = \sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2$
- Our test statistic is then,

$$T = \frac{\sqrt{m+n-2}(\overline{X}_m - \overline{Y}_n)}{\sqrt{(\frac{1}{m} + \frac{1}{n})}\sqrt{(S_X^2 + S_Y^2)}}$$

- Under the null hypothesis T will have a t distribution with $m+n-2$ degrees of freedom.

Two Sample t Tests

- If we instead assume that the variance is not the same for both distributions we have the following test statistic:

$$T = \frac{(\overline{X}_m - \overline{Y}_n)}{\sqrt{(\frac{S_X^2}{m} + \frac{S_Y^2}{n})}}$$

Two Sample t Test Example

- Suppose 6 subjects were given a drug (treatment group) and an additional 6 subjects a placebo (control group).

Two Sample t Test Example

- Suppose 6 subjects were given a drug (treatment group) and an additional 6 subjects a placebo (control group).
- Their reaction time to a stimulus was measured (in ms).

Two Sample t Test Example

- Suppose 6 subjects were given a drug (treatment group) and an additional 6 subjects a placebo (control group).
- Their reaction time to a stimulus was measured (in ms).
- We want to perform a two-sample t-test for comparing the means of the treatment and control groups.

Two Sample t Test Example

- Suppose 6 subjects were given a drug (treatment group) and an additional 6 subjects a placebo (control group).
- Their reaction time to a stimulus was measured (in ms).
- We want to perform a two-sample t-test for comparing the means of the treatment and control groups.
- Let μ_1 be the mean of the population taking medicine and μ_2 the mean of the untreated population.

Two Sample t Test Example

- Suppose 6 subjects were given a drug (treatment group) and an additional 6 subjects a placebo (control group).
- Their reaction time to a stimulus was measured (in ms).
- We want to perform a two-sample t-test for comparing the means of the treatment and control groups.
- Let μ_1 be the mean of the population taking medicine and μ_2 the mean of the untreated population.
- The hypothesis of interest can be expressed as:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 < 0$$

Two Sample t Test Example

- Suppose 6 subjects were given a drug (treatment group) and an additional 6 subjects a placebo (control group).
- Their reaction time to a stimulus was measured (in ms).
- We want to perform a two-sample t-test for comparing the means of the treatment and control groups.
- Let μ_1 be the mean of the population taking medicine and μ_2 the mean of the untreated population.
- The hypothesis of interest can be expressed as:

$$H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 - \mu_2 < 0$$

- Suppose we assume equal variance for both groups.

Two Sample t Test Example

Let's think about performing this test in **R**:
Control = c(91, 87, 99, 77, 88, 91)
Treat = c(101, 110, 103, 93, 99, 104)

t.test(Control,Treat,alternative="less", var.equal=TRUE)

Two Sample t-test

data: Control and Treat t = -3.4456, df = 10,
p-value = 0.003136
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval: -Inf -6.082744
sample estimates:
mean of x mean of y
88.83333 101.66667

Paired t-test

- The two-sample t-test was used when we computed statistics from two independent random samples and we wanted to make decisions about how much they differed.
- Other times we may want to know the difference between pairs of things that are linked by some known relationship.
- A paired t-test follows the same pattern as with the one sample t-test, except that we perform the hypothesis test on a statistic computed from the differences between each pair of observations.

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i = \frac{\sum_{i=1}^{n} y_{i1}}{n} - \frac{\sum_{i=1}^{n} y_{i2}}{n}$$

Paired t-test

- For example, we might test whether or not the mean of the two populations is identical - the true difference in means is 0:

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d \neq 0$$

Paired t-test

- For example, we might test whether or not the mean of the two populations is identical - the true difference in means is 0:

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d \neq 0$$

- Then,

$$t_d = \frac{\bar{d} - 0}{\hat{se}_{\bar{d}}}$$

Paired t-test

- For example, we might test whether or not the mean of the two populations is identical - the true difference in means is 0:

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d \neq 0$$

- Then,

$$t_d = \frac{\bar{d} - 0}{\hat{se}_{\bar{d}}}$$

- As before, under H_0 this test statistic is t_{n-1} distributed.

Paired t-test

- For example, we might test whether or not the mean of the two populations is identical - the true difference in means is 0:

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d \neq 0$$

- Then,

$$t_d = \frac{\bar{d} - 0}{\hat{se}_{\bar{d}}}$$

- As before, under H_0 this test statistic is t_{n-1} distributed.
- So the confidence interval for the difference $\mu_d = \mu_2 - \mu_1$ is

$$\bar{d} \pm t_{n-1,\alpha/2} \hat{se}_{\bar{d}}$$

Paired t-test Example

- Suppose we want to test two different prediction algorithm on the same test set of size n.

Paired t-test Example

- Suppose we want to test two different prediction algorithm on the same test set of size n.
- Let $X_i = 1$ if algorithm 1 is correct on test case i and let $X_i = 0$ otherwise.
- Let $Y_i = 1$ if algorithm 2 is correct on test case i and let $Y_i = 0$ otherwise.

Paired t-test Example

- Suppose we want to test two different prediction algorithm on the same test set of size n.
- Let $X_i = 1$ if algorithm 1 is correct on test case i and let $X_i = 0$ otherwise.
- Let $Y_i = 1$ if algorithm 2 is correct on test case i and let $Y_i = 0$ otherwise.
- Define $D_i = X_i - Y_i$, then a typical data set would look like:

Test Case	X_i	Y_i	D_i
1	1	0	1
2	1	1	0
3	1	1	0
4	0	1	-1
5	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$
n	0	1	-1

Paired t-test Example

- Let $\delta = \mathrm{E}(D_i)$ then $\hat{\delta} = \overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$ and $\hat{\mathrm{se}}(\hat{\delta}) = S/\sqrt{n}$ where $S^2 = \frac{1}{n}\sum_{i=1}^{n}(D_i - \overline{D})^2$.

Paired t-test Example

- Let $\delta = \mathrm{E}(D_i)$ then $\hat{\delta} = \overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$ and $\hat{\mathrm{se}}(\hat{\delta}) = S/\sqrt{n}$ where $S^2 = \frac{1}{n}\sum_{i=1}^{n}(D_i - \overline{D})^2$.
- To test

$$H_0 : \delta = 0$$
$$H_1 : \delta \neq 0$$

Paired t-test Example

- Let $\delta = \mathrm{E}(D_i)$ then $\hat{\delta} = \overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$ and $\widehat{\mathrm{se}}(\hat{\delta}) = S/\sqrt{n}$ where $S^2 = \frac{1}{n}\sum_{i=1}^{n}(D_i - \overline{D})^2$.
- To test

$$H_0 : \delta = 0$$

$$H_1 : \delta \neq 0$$

$$t_\delta = \frac{\bar{\delta}}{\hat{s}e_{\bar{\delta}}}$$

Paired t-test Example

- Let $\delta = \mathrm{E}(D_i)$ then $\hat{\delta} = \overline{D} = \frac{1}{n}\sum_{i=1}^{n} D_i$ and $\hat{\mathrm{se}}(\hat{\delta}) = S/\sqrt{n}$ where $S^2 = \frac{1}{n}\sum_{i=1}^{n}(D_i - \overline{D})^2$.
- To test

$$H_0 : \delta = 0$$

$$H_1 : \delta \neq 0$$

$$t_\delta = \frac{\bar{\delta}}{\hat{se}_{\bar{\delta}}}$$

- Reject H_0 if $|t_\delta| > t_{n-1}^{-1}(\alpha/2)$

Multiple Comparisons

- The scenario of testing many pairs of groups is called multiple comparisons.

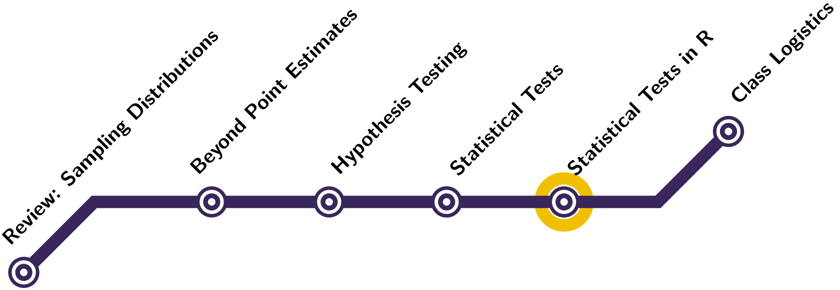
Multiple Comparisons

- The scenario of testing many pairs of groups is called multiple comparisons.
- The Bonferroni correction suggests that a more stringent significance level is more appropriate for these tests:

$$\alpha^* = \frac{\alpha}{K\text{Îś}}$$

  where K is the number of comparisons being considered (formally or informally).
- If there are k groups, then usually all possible pairs are compared and $K = \frac{k(k\text{âĹŠ}1)}{2}$.

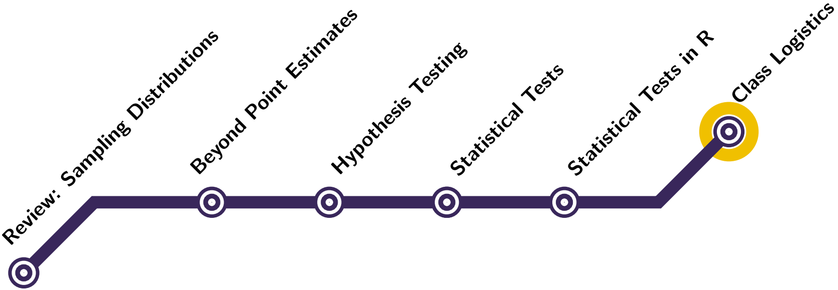
Today's Roadmap



Statistical Tests in R

http://www.ats.ucla.edu/stat/r/whatstat/whatstat.htm

Today's Roadmap



Problem Set 1
Due Friday 11:59 am

Questions?