Hypothesis tests for multiple proportions

Overview

Review of parametric interference for two means

Inference for the difference of two proportions

Hypothesis tests for more than two proportions

- Randomization test using the chi-square statistic
- Parametric chi-square statistic test for goodness-of-fit

Announcement: homework 10

Homework 10 has been posted!

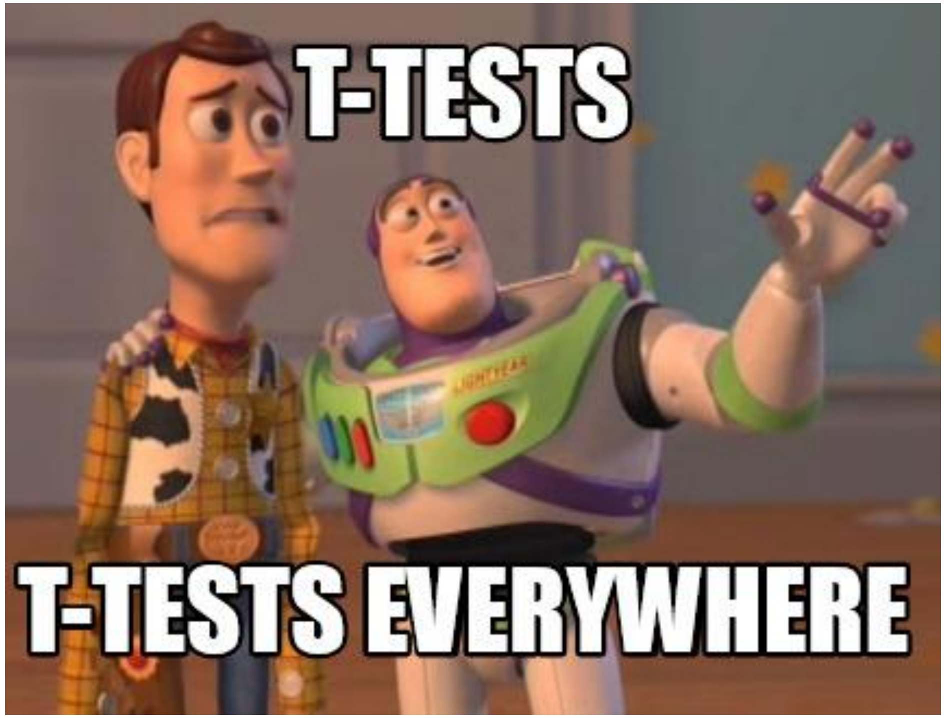
- The last homework ☹

It is due on Gradescope on Monday December 1st at 11pm

Also, please work on your class final project

- It is due on Gradescope on Sunday December 7th

Review: parametric inference for means

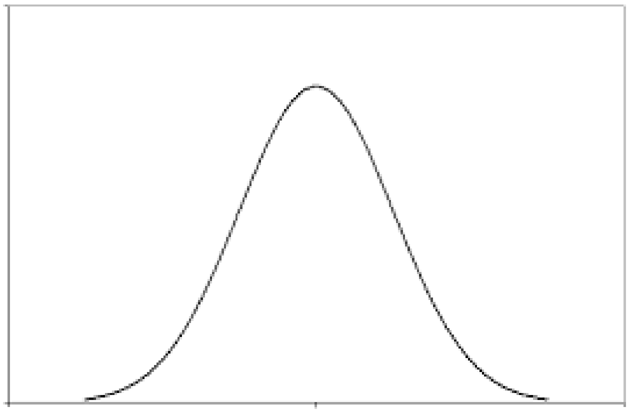


Review: Using parametric methods for inference

When using "parametric" methods for statistical inference, we have mathematical functions for our distributions of statistics

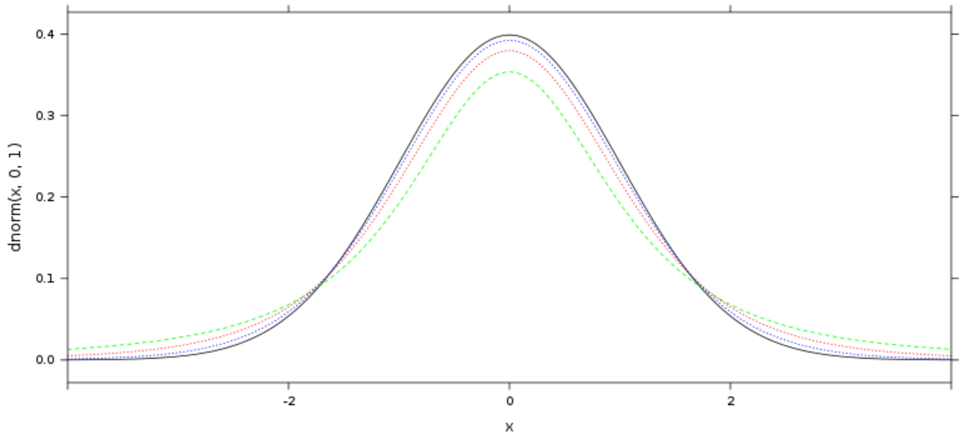
- e.g., we have mathematical functions for the null distribution rather than needing to use randomization methods such as shuffling our data

Normal distribution N(μ,σ)



Inference on a single proportion π

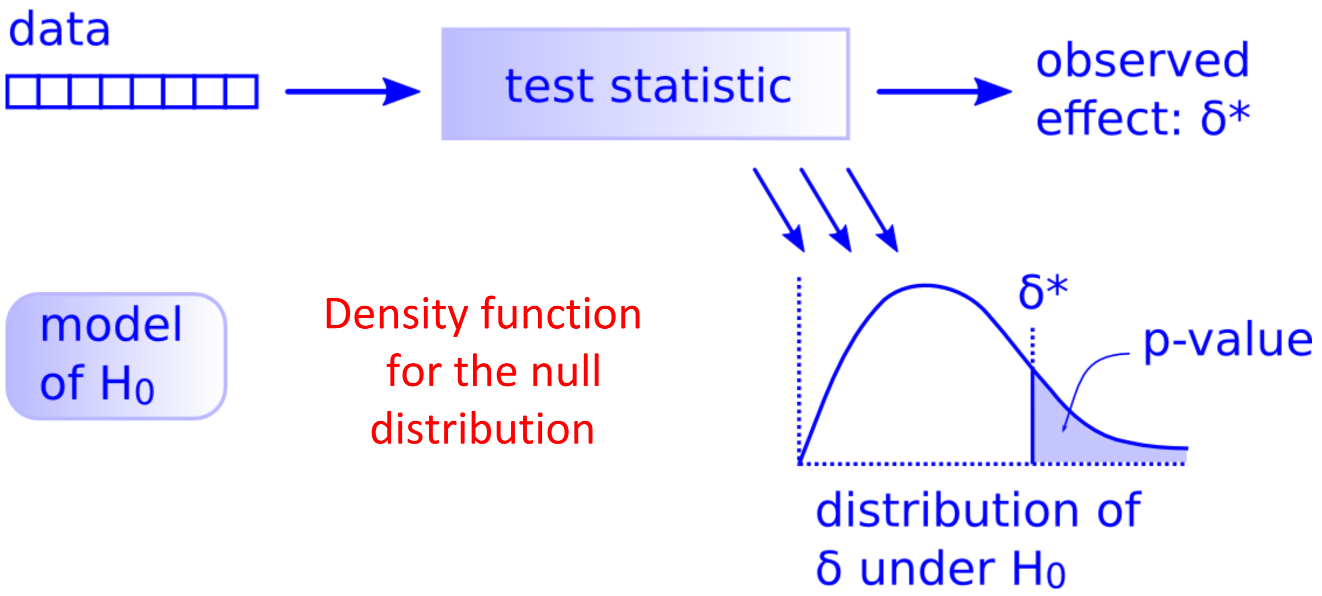
T-distribution



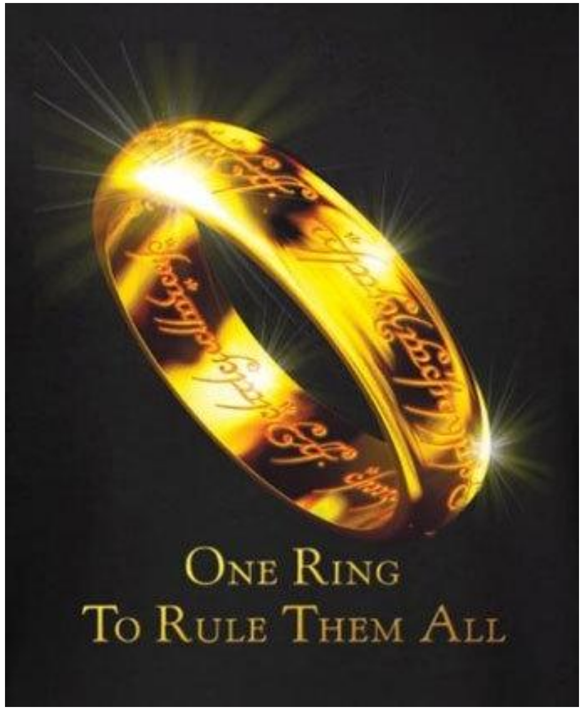
Inference on means μ (when σ is unknown)

One test to rule them all

There is only one hypothesis test!



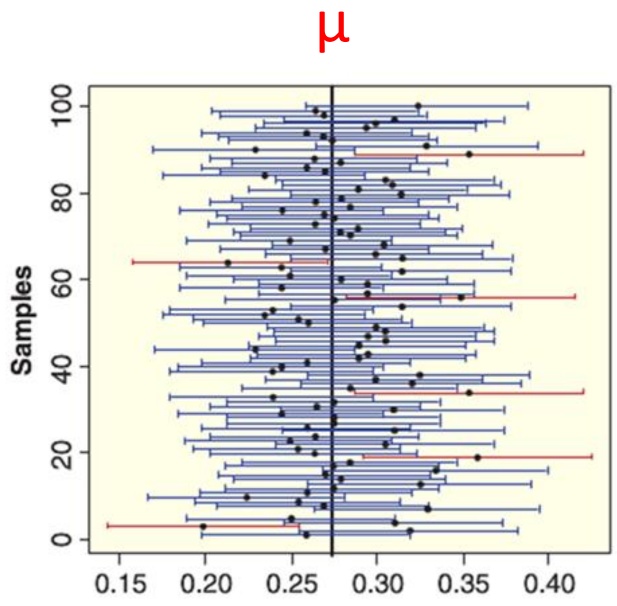
Density function for the null distribution



Just follow the 5 hypothesis tests steps!

CI: stat $\pm$ $q^* \cdot SE$

Confidence intervals for means



Single mean μ: $\bar{x} \pm t^* \cdot \frac{s}{\sqrt{n}}$

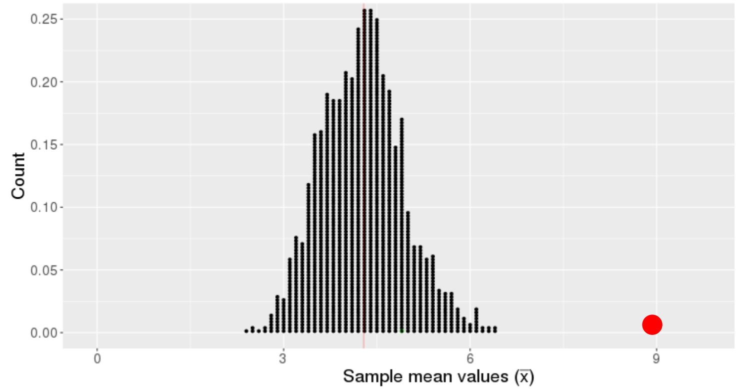
Difference of two means $\mu_1 - \mu_2$: $(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Mean of differences for paired data μ_d: $\bar{x}_d \pm t^* \cdot \frac{s_d}{\sqrt{n_d}}$

(the t-distribution is appropriate provided n is large (≥ 30) or the data is not too skewed)

Hypothesis tests means (t-tests)

$H_0: \cancel{\mu} = \mu_0$



Single mean μ: $$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$ df = n - 1

- H_0: $\mu = \mu_0$

Comparing two means $\mu_1 - \mu_2$: $$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$ df = $\min(n_1, n_2) - 1$

- H_0: $\mu_1 = \mu_2$

Comparing mean of differences for paired data μ_d: $$t = \frac{\bar{x}_d}{\frac{s_d}{\sqrt{n}}}$$ df = n - 1

- H_0: $\mu_d = 0$

(the t-distribution is appropriate provided n is large (≥ 30) or the data is not too skewed)

Warm-up problem 1: Are you getting enough sleep?

It is recommended that adults sleep at least 8 hours a night

A Statistics professor asked 12 undergraduate students how much sleep they were getting and found the average was 6.2 hours with a standard deviation of 1.7 hours

Assuming this is representative of all students in a Statistics class, does this provide evidence that students in the class are not getting enough sleep on average?

Run a hypothesis test to examine this question!

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Warm-up problem 2: Who eats more fiber, males or females?

Nierenberg et al (1989) collected data on eating habits. Let's use this data to find a 95% confidence interval for the differences in the number of grams of fiber eaten in a day between males and females

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Let's try it in R!

Inference for the differences of two proportions

The plan: comparing proportions

For inference on the difference of two proportions, π_1 and π_2, we will use a **normal distribution**

For inference on more than two proportions, π_1, π_2, …, π_k, we will use a **χ^2 distribution**

Inference for the differences of two proportions

Inference on the differences of two proportions $\pi_1 - \pi_2$ is similar to inference on a single proportion π

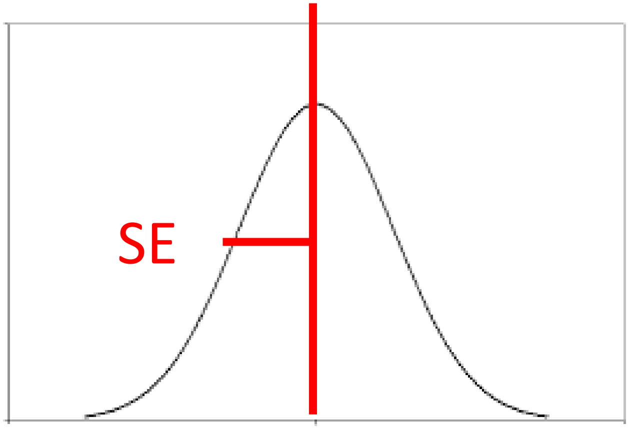
$$\hat{p_1} - \hat{p_2} = N\left(\pi_1 - \pi_2, \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}\right)$$

Suppose we have two samples of categorical data of sizes n_1 and n_2

The statistic $\hat{p}_1 - \hat{p}_2$ comes from a **normal distribution** that is:

- Centered at $\pi_1 - \pi_2$
- Has a standard error $SE = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$

$\pi_1 - \pi_2$



$\hat{p}_1 - \hat{p}_2$

To ensure the samples sizes are large enough, check that $n\pi_i \geq 10$ and $n(1 - \pi_i) \geq 10$ for i = 1 and 2

Confidence intervals for differences in proportions

If we have large samples n_1 and n_2 from two different groups, we can construct a confidence interval for the difference in proportions $\pi_1 - \pi_2$ as:

$$\hat{p_1} - \hat{p_2} \pm z^* \sqrt{\frac{\hat{p_1}(1 - \hat{p_1})}{n_1} + \frac{\hat{p_2}(1 - \hat{p_2})}{n_2}}$$

Note we are substituting p̂ for π

Test for difference of proportions

To test for $H_0: \pi_1 = \pi_2$ vs $H_A: \pi_1 \neq \pi_2$ (or the one-tail alternative), based on sample sizes of n_1 and n_2 we use the standardized test statistic:

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

Combine data from both groups and calculate the proportion from all the data

Where p̂ is the pooled overall proportion from combining the data from both groups

Flight delays

The table below gives the flight arrival numbers from a random sample of flights for two airlines

	Early	On-time	Late	Total
Airline A	133	416	151	700
Airline B	58	355	87	500
	191	771	238	1200

Let’s test whether there is a difference between the two airlines in the percent of flights that arrive late

Let’s try it in R!

Randomization test to compare multiple proportions

Testing more than two categories

A Yale professor was interesting in examining whether Yale students were equally likely to be born in each month

Q_1: What is the null and alternative hypotheses he could use to test this hypothesis?

Chi-square goodness-of-fit test

If we want to test proportions for $k > 2$ categories, we can use:

1. A randomization test
2. A parametric "*chi-square goodness-of-fit*" test

These test:

H_0: $\pi_1 = a$, $\pi_2 = b$, … $\pi_k = z$

H_A: Some π_i is not as specified in H_0

The tests don't specify which proportion differs from what is specified in the null hypothesis, just that at least one proportion does differ

Chi-square statistic

The **chi-square statistic**, denoted χ^2, is found by comparing the **observed counts** from a sample with the **expected counts** derived from a null hypothesis and is computed as:

$$\chi^2 = \sum_{i=1}^{k} \frac{(\text{Observed}_i - \text{Expected}_i)^2}{\text{Expected}_i}$$

Note this is a Greek symbol even though it is a statistic ☹

Birth months from 198 Yale students

Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	**Total**
20	17	21	24	11	18	12	17	13	14	17	14	**198**

Q: What is the expected value for each month?

A: 198/12 = 16.5

$$\chi^2 = \sum_{i=1}^{k} \frac{(\text{Observed}_i - \text{Expected}_i)^2}{\text{Expected}_i} = 10.12$$

Q: Is the observed statistic beyond what we would expect if H_0 was true?

Creating a null distribution

Any ideas how we could create a null distribution using randomization methods?

We could use randomization method:

- Roll k-sided weighted die
- Probability of getting each side is equal to the proportions specified in the null hypothesis
- Roll the die n times to simulate one experiment
  - Calculate χ^2 statistic based on these rolls
- Repeat many times to get a null distribution

Creating a null distribution

For the Yale birth month example…

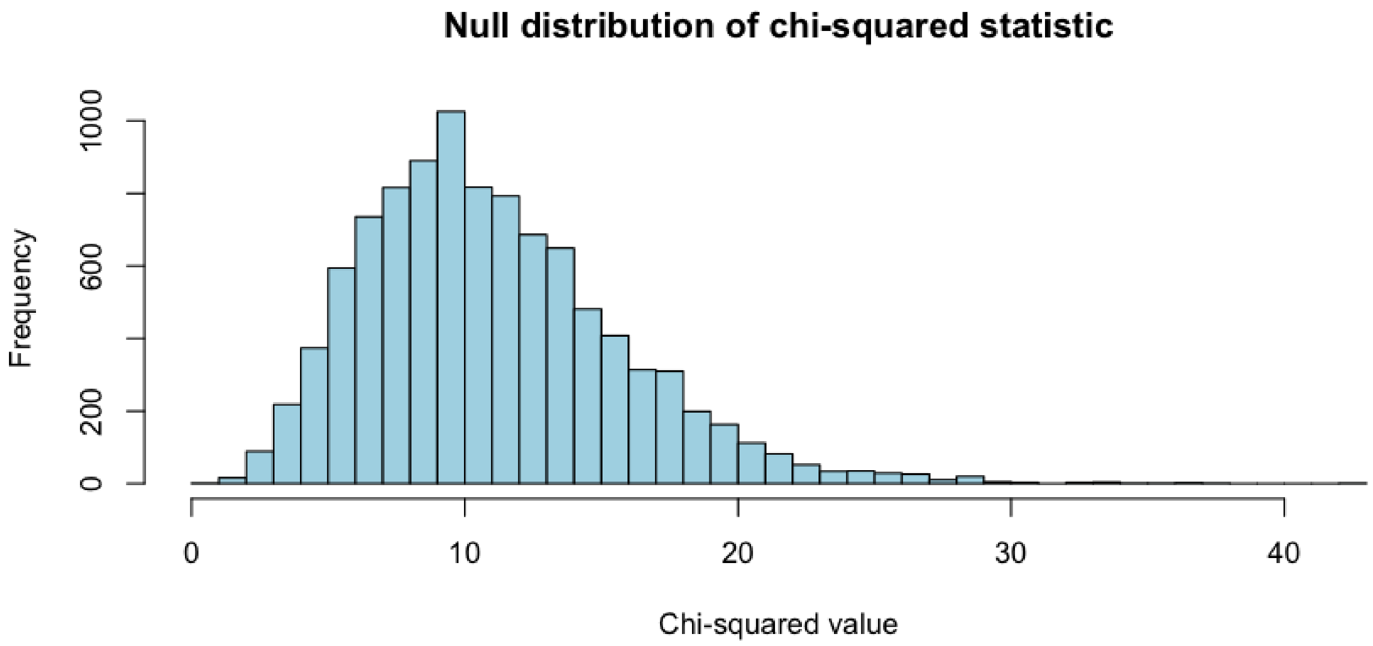
How many sides would the die have?

What would the probability be of getting each side?

How many times would we roll the die to simulate one data set?

How many times would we repeat this process?

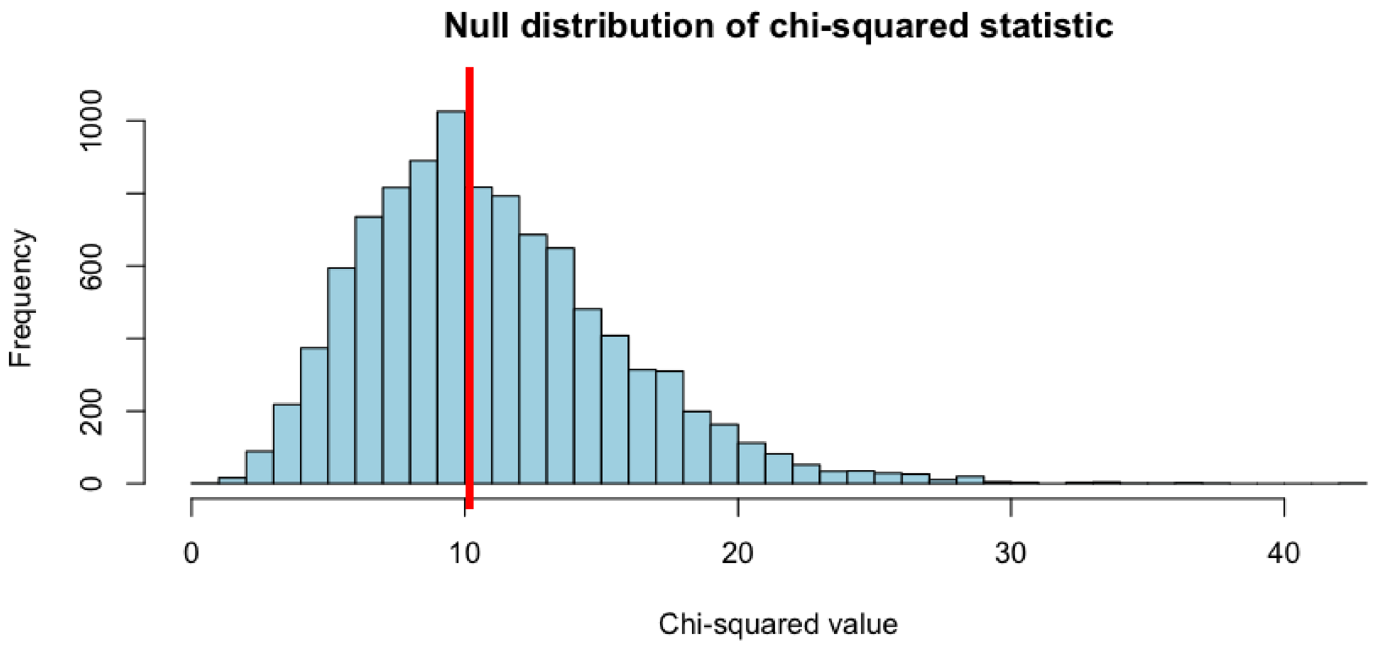
Randomization null distribution




```
simulated_counts <- rmultinom(1,  n,  expected_proportions)
simulated_counts <- rmultinom(1,  198,  rep(1/12, 12))
```

Calculate statistic χ^2 and repeat 10,000 times

Randomization null distribution



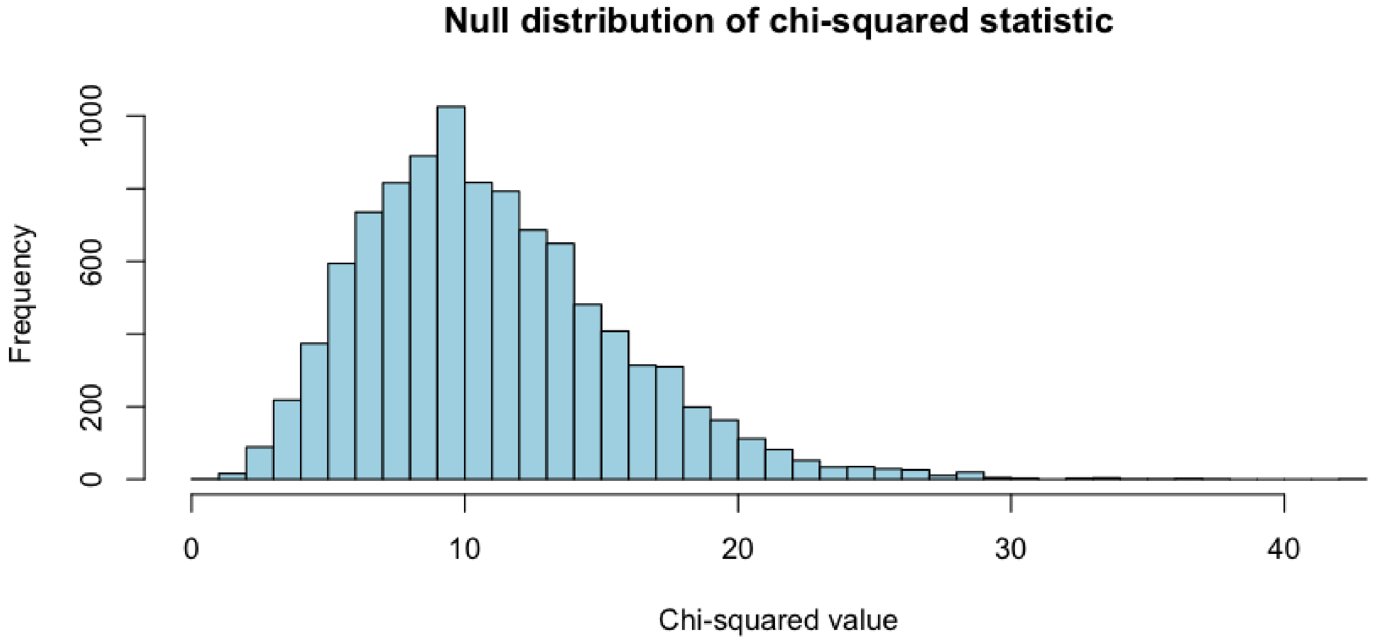
$\chi^2 = 10.12$

p-value = …?

Let’s try it in R!

χ^2 test for goodness-of-fit to compare multiple proportions

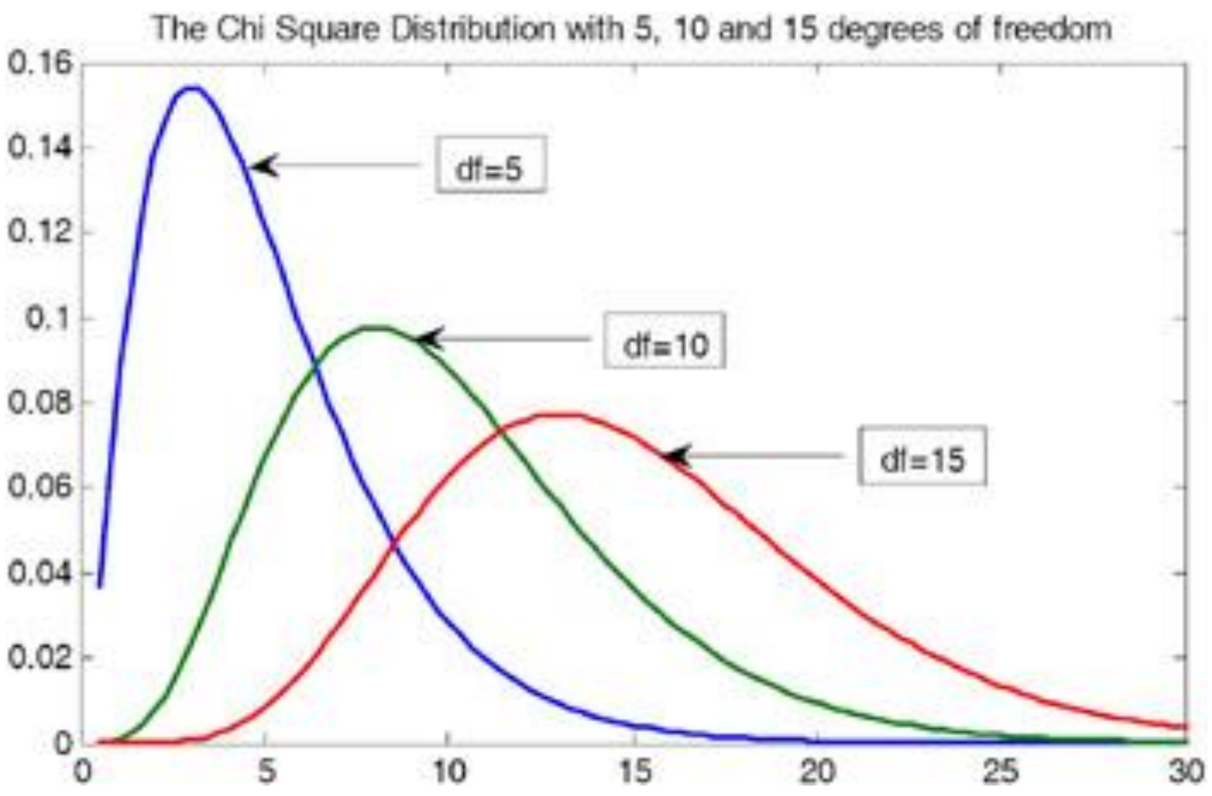
Parametric null distribution



Is there a parametric null distribution we could use instead of simulations?

Yes! The χ^2 distribution!

Chi-square distribution

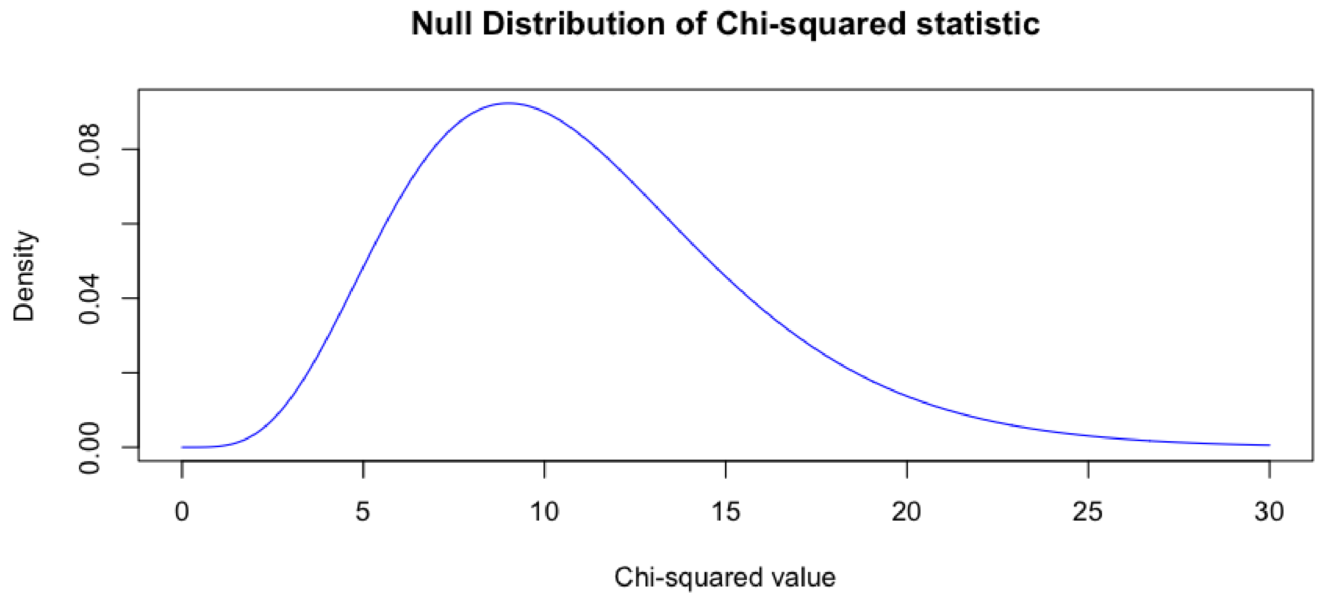


k is the number of groups

The χ^2 has one parameter called 'degrees of freedom', which is equal to k - 1

χ^2 distribution can be used as a null distribution as long as there are at least 5 expected counts in every condition

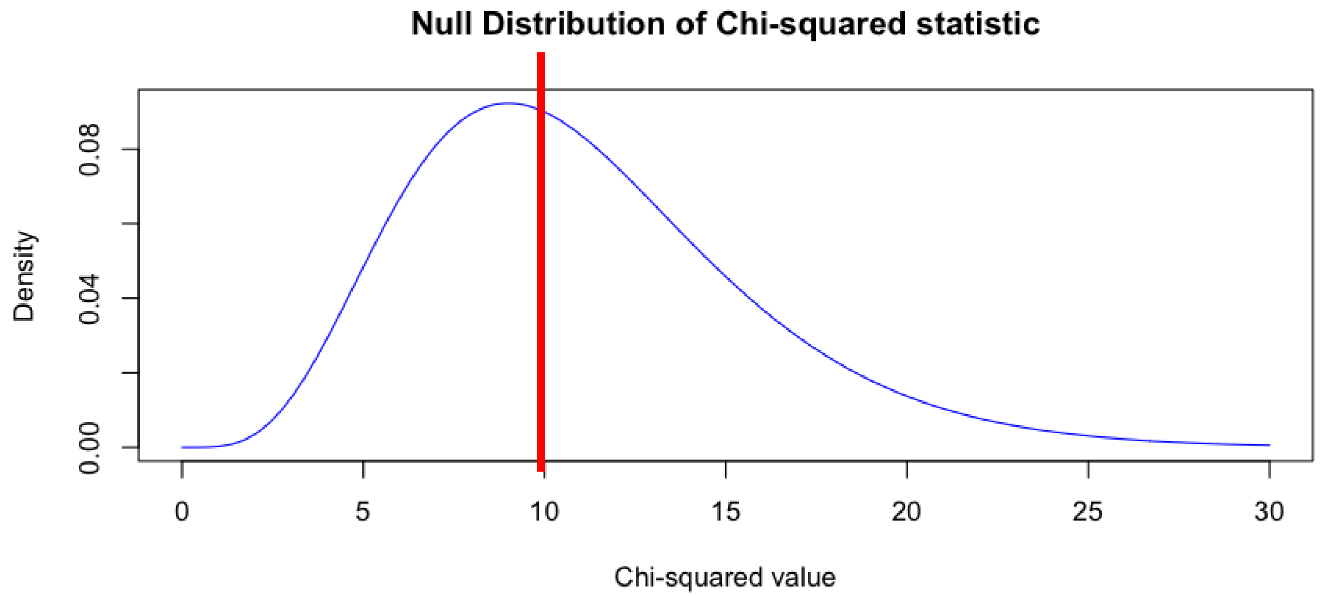
Chi-square distribution



To plot the chi-squared density we can use: y_vals <- dchisq(x_vals, df)

To get a p-value we can use: pchisq(chi_stat, df, lower.tail = FALSE)

Chi-square distribution



For Yale birth month example, since there 12 months there are 11 degrees of freedom (df = 11)

P-value based on the chi-squared distribution = …?

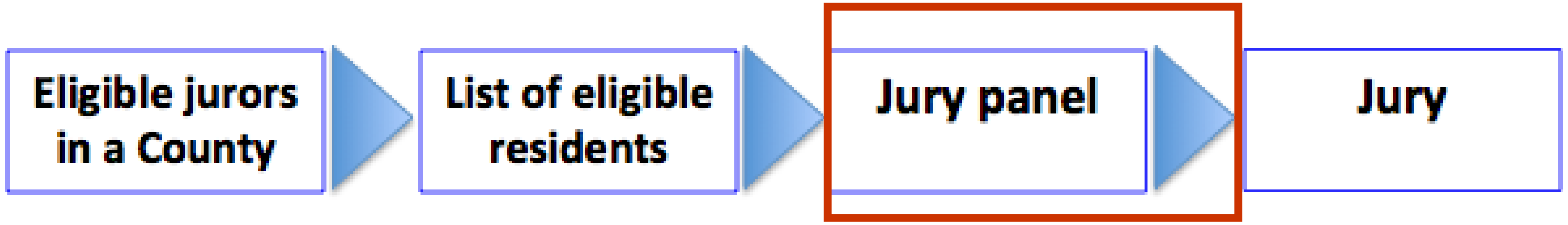
Let’s try it in R!

Another example

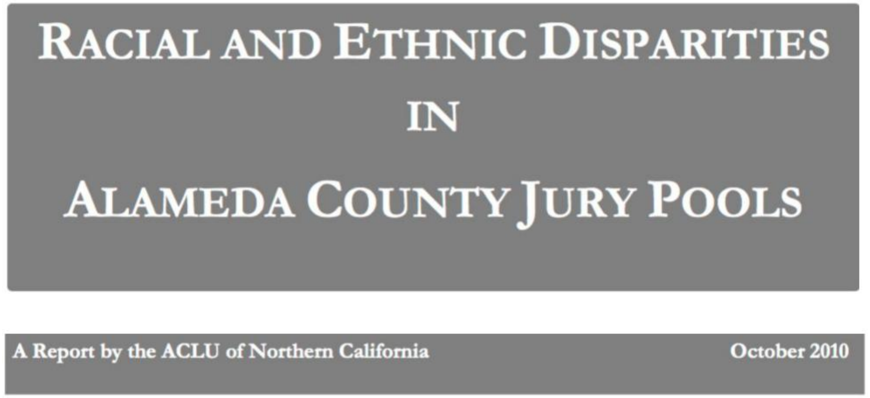
Jury selection in Alameda county

Section 197 of California's Code of Civil Procedure says:

" All persons selected for jury service shall be selected at random, from a source or sources inclusive of a representative cross section of the population of the area served by the court."



In 2010, the American Civil Liberties Union (ACLU) of Northern California presented a report that concluded that certain racial and ethnic groups are underrepresented among jury panelists in Alameda County.



RACIAL AND ETHNIC DISPARITIES IN ALAMEDA COUNTY JURY POOLS

A Report by the ACLU of Northern California October 2010

Jury selection in Alameda county data

The ACLU compiled data on the composition of 1453 people who were on jury panels from in the years 2009 and 2010.

The demographics and jury panel proportions were:

	Asian	Black	Latino	White	Other
Eligible	0.15	0.18	0.12	0.54	0.01
Panel	0.26	0.08	0.08	0.54	0.04

Question: were ethnicities selected to be on the panel representative of what would be expected from the underlying population?

Jury selection in Alameda county data

The ACLU compiled data on the composition of 1453 people who were on jury panels from in the years 2009 and 2010.

The demographics and jury panel proportions were:

	Asian	Black	Latino	White	Other
Eligible	0.15	0.18	0.12	0.54	0.01
Panel	0.26	0.08	0.08	0.54	0.04

Question: What are the null and alternative hypotheses?

H_0: $\pi_A = .15$, $\pi_B = .18$, $\pi_L = .12$, $\pi_W = .54$, $\pi_o = .01$

H_A: Some π_i is not as specified in H_0

Let's try it in R!