Tests for multiple proportions

Overview

Review of parametric interference for two means

Hypothesis tests for two proportions

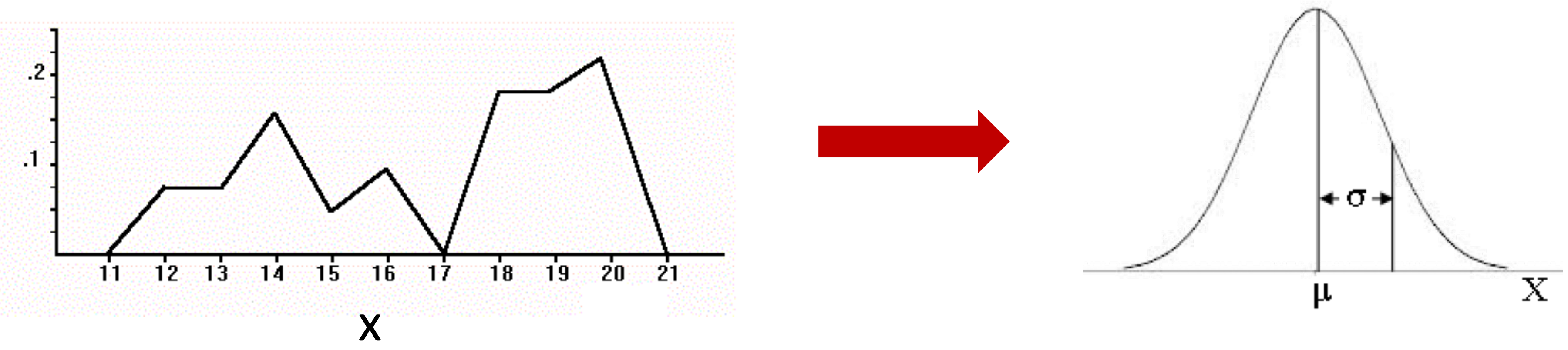
Hypothesis tests for more than two proportions

- Testing goodness-of-fit for a single categorical variables

Review: parametric inference on two mean

Review: Central Limit Theorem for Sample means

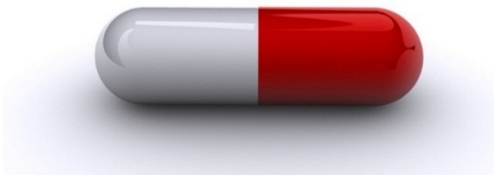
The sampling distribution of sample means ($\bar{x}$) *from **any population distribution*** will be normal, provided that the sample size is large enough



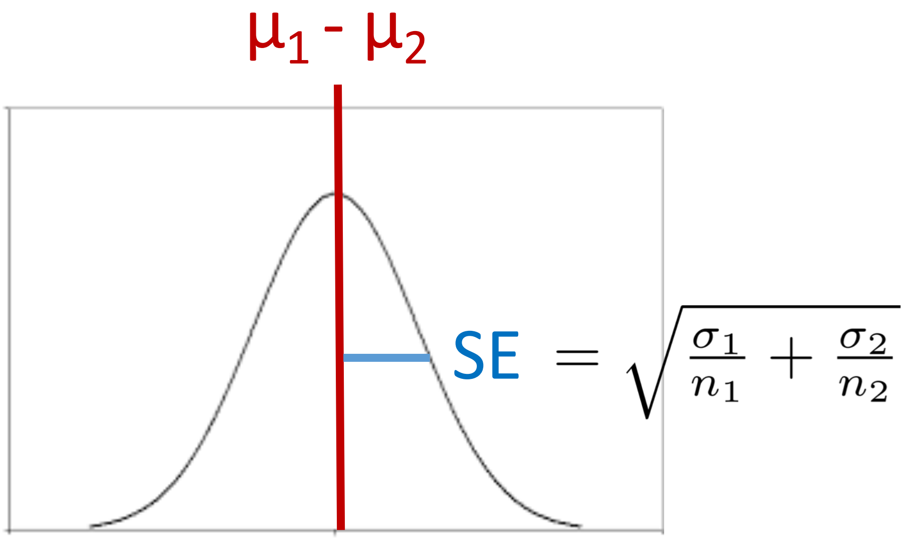
The more skewed the distribution, the larger sample size we will need for the normal approximate to be good (usually 30 is sufficient)

Recall also that all normal distributions density models have two parameters N(**μ, σ**)

Distribution of differences in means



$$\overline{x}_1 - \overline{x}_2 \sim N\left(\mu_1 - \mu_2, \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$



The standard error of differences of means

Similar to the standard error for means from a single sample we do not know σ.

We can substitute s for σ

$$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \qquad SE = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Our sample statistic $(\bar{x}_1 - \bar{x}_2)/SE$ comes from a t-distribution

(provided n is large or the data is not too skewed)

We will use the minimum of $n_1 - 1$, or $n_2 - 1$ as a conservative estimate of the df

Confidence intervals and two sample t-tests

Confidence intervals can then be calculated as:

$$(\overline{x_1} - \overline{x_2}) \pm t^* \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

And we can create t-statistics, which we can find the probability distribution for using the t-distribution:

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Use the smaller of n_1 -1 and n_2 - 1 to give the degrees of freedom

Inference for a difference in means with paired data

To estimate the difference in means based on paired data, we first compute the difference for each data pair

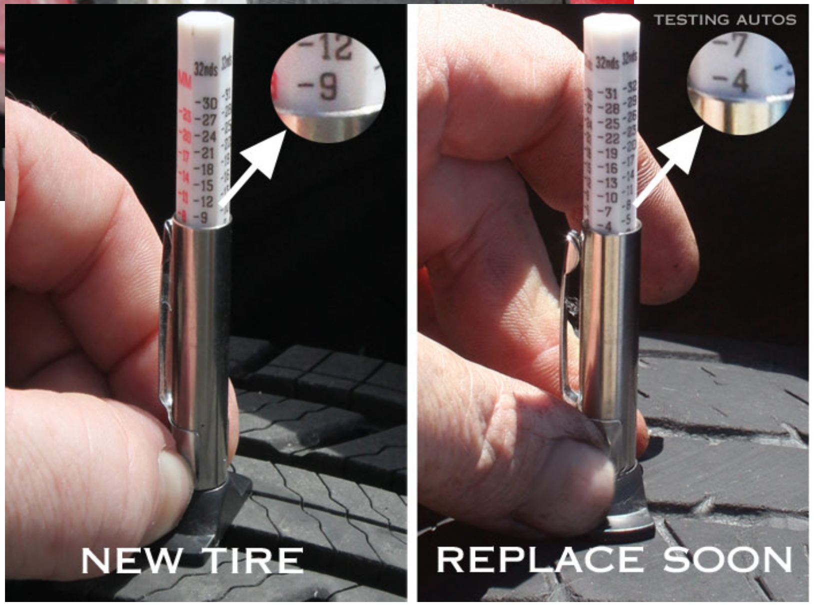
We can then compute the mean $\overline{x}_d$ the standard deviation $\overline{s}_d$, and the sample size n_d for the sample difference to test…

H_0: $\mu_d = 0$

H_A: $\mu_d \neq 0$

we use the t-statistic: $$t = \frac{\overline{x}_d}{s_d / \sqrt{n_d}}$$

Do left or right tires wear out quicker?



Car #	Right tire	Left tire
1	42	54
2	75	73
3	24	22
4	56	59
5	52	51
6	56	45
7	23	29
8	55	58
9	46	49
10	52	58
11	47	49
12	62	67
13	55	58
14	62	64

Do left or right tires wear out quicker?

Independent samples t-test

H_0: $\mu_{left} - \mu_{right} = 0$

H_A: $\mu_{left} - \mu_{right} \neq 0$

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

We will use the minimum of $n_1 - 1$, or $n_2 - 1$ as a conservative estimate of the df

Car #	Right tire	Left tire
1	42	54
2	75	73
3	24	22
4	56	59
5	52	51
6	56	45
7	23	29
8	55	58
9	46	49
10	52	58
11	47	49
12	62	67
13	55	58
14	62	64

Do left or right tires wear out quicker?

Paired samples t-test

H_0: $\mu_{diff} = 0$

H_A: $\mu_{diff} \neq 0$

$$t = \frac{\overline{x}_d}{s_d / \sqrt{n_d}}$$

The t-statistic comes from a t-distribution with n – 1 degrees of freedom

Car #	Right tire	Left tire
1	42	54
2	75	73
3	24	22
4	56	59
5	52	51
6	56	45
7	23	29
8	55	58
9	46	49
10	52	58
11	47	49
12	62	67
13	55	58
14	62	64

Do left or right tires wear out quicker?

Let’s try both of these t-tests in R!

Brief note on inference for the differences in two proportions

Comparing multiple proportions

What symbol do we use for the population parameter for proportions?

- A: π

What symbol do we use for the sample statistic for proportions?

- A: $\hat{p}$

When comparing multiple proportions, we will ask questions about multiple parameters: $\pi_1, \pi_2, \ldots, \pi_k$

When doing inference on two proportions, π_1 and $\pi_{2,}$ we can use a normal distribution

When doing inference on more than two proportions we will use a χ^2 distribution

Briefly: Inference for the differences in two proportions

Computing confidence intervals and p-values for differences of proportions $\pi_1 - \pi_2$ is very similar to computing them for a single proportion π

The statistic here is $\hat{p}_1 - \hat{p}_2$ and we just have a different formula is used the SE

Note: by difference of proportions, we mean proportions computed from two different samples

- E.g., proportion of left-handed people who believe in climate change vs. the proportion of right-handed who do

Central Limit Theorem for Differences in Proportions

When choosing random samples of size n_1 and n_2 from a population with a proportions π_1 and π_2, the distribution of the difference in sample proportions $\hat{p}_1 – \hat{p}_2$ has the following characteristics:

Center: The mean is equal to the population proportion $\pi_1 - \pi_2$

Spread: The standard error is: $$SE = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$$

Shape: As both sample sizes get large, the distribution of $\pi_1 - \pi_2$ is reasonably normal

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

We can write the sampling distribution of differences of proportion as:

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

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 compute p-values when the null distribution is normal we use:

$$z = \frac{Sample\ \ Statistic\ -\ Null\ \ Parameter}{SE}$$

In the context of proportions we usually state H_0: $\pi_1 = \pi_2$ or equivalently $\pi_1 - \pi_2 = 0$ so our z statistic becomes:

$$z = \frac{(\hat{p_1} - \hat{p_2}) - 0}{SE} \qquad = \frac{\hat{p_1} - \hat{p_2}}{SE}$$

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:

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

We can get a p-value using: pnorm(z, 0, 1)

Example: try this at home

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

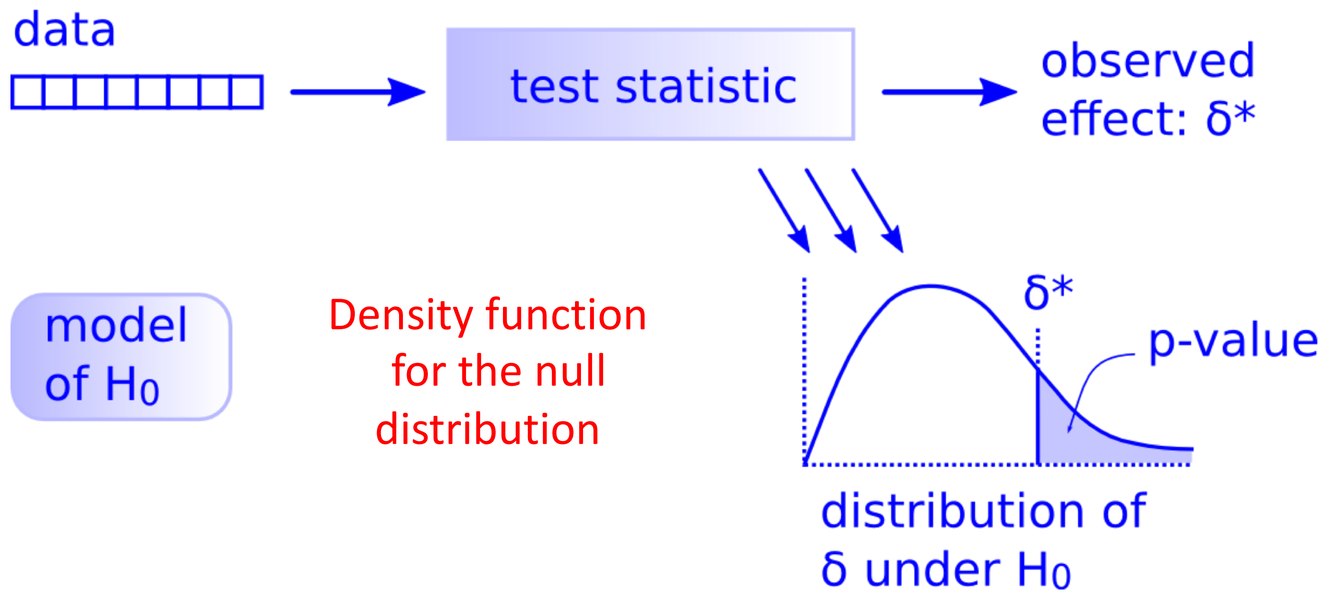
	Early	On-time	Late	Total
Airline A	133	416	151	700
Airline B	58	355	87	500
	191	771	238	1200

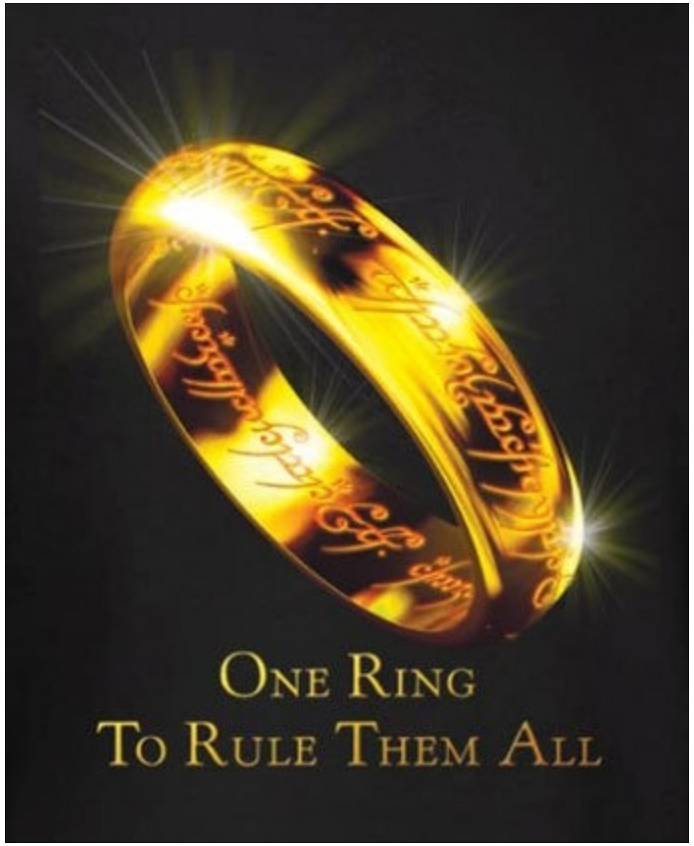
Test whether there is a difference between the two airlines in the percent of flights that arrive late

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

One test to rule them all

There is only one hypothesis test!





Just follow the 5 hypothesis tests steps!

The sprinkle business (fictional)

ACME
CORPORATION

PERFECT
Corporation

ACME corporation believes that if they had the correct proportions of colors of sprinkles that PERFECT corporation uses, their sales will increase

Testing more than 2 categories

Suppose ACME believed that the proportion of all sprinkles colors in a population was the same

Q_1: What is the null and alternative hypotheses?

(recall the sprinkle colors were: red, pink, orange, yellow, green, white)

Q_2: Any ideas how we could test this?

Testing more than 2 categories

Suppose ACME believed that the proportion of all sprinkles colors in a population was the same

Q_1: What is the null and alternative hypotheses?

(recall the sprinkle colors were: red, pink, orange, yellow, green, white)

There are 6 sprinkle colors, so the null hypothesis would be:

H_0: $\pi_{red} = \pi_{pink} = \pi_{orange} = \pi_{yellow} = \pi_{green} = \pi_{white} = 1/6$

H_A: One of the proportions π_i is different

Testing more than 2 categories

Suppose ACME believed that the proportion of all sprinkles colors in a population was the same

Q_2: Any ideas how we could test this?

One solution: we could do 6 hypothesis tests for each $\pi_i = 1/6$

Problem: multiple comparisons would lead to higher type I error rate

Chi-square goodness-of-fit test

If we want to test proportions for $k > 2$ categories we can use a *chi-square goodness-of-fit* test

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

H_A: Some π_i is not as specified in H_0

Doesn't specify which proportion differs from the null hypothesis, just that one of them does

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:

$$\chi^2 = \sum_{i=1}^{n} \frac{(Observed_i - Expected_i)^2}{Expected_i}$$

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

ACME scientists counted 726 sprinkles

	Red	Pink	Orange	Yellow	Green	White		Total
Observed	138	99	106	115	104	164		726
Expected	121	121	121	121	121	121		726

$$\chi^2 = \sum_{i=1}^{n} \frac{(Observed_i - Expected_i)^2}{Expected_i} = 26.21$$

Is the observed statistic beyond what we would expect by chance?

Creating a null distribution

Any ideas how we could create a null distribution?

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 sprinkle example…

How many sides would the die have?

- 6 sided die

What would the probability be of getting each side?

- 1/6

How many times would we roll the die?

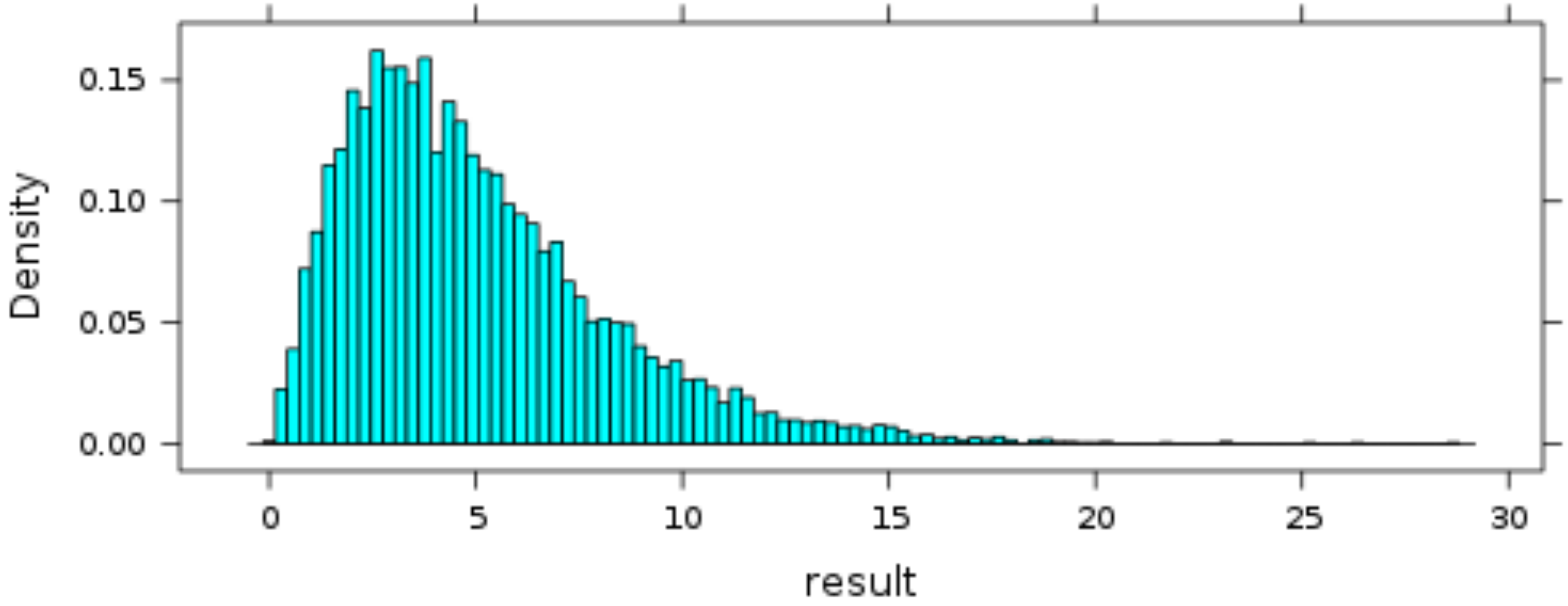
- n = 726
- Calculate χ^2 statistic based on these 726 rolls

$$\chi^2 = \sum_{i=1}^{n} \frac{(Observed_i - Expected_i)^2}{Expected_i}$$

How many times would we repeat this process?

- 10,000 times

Randomization test to determine significance of a chi-square statistic

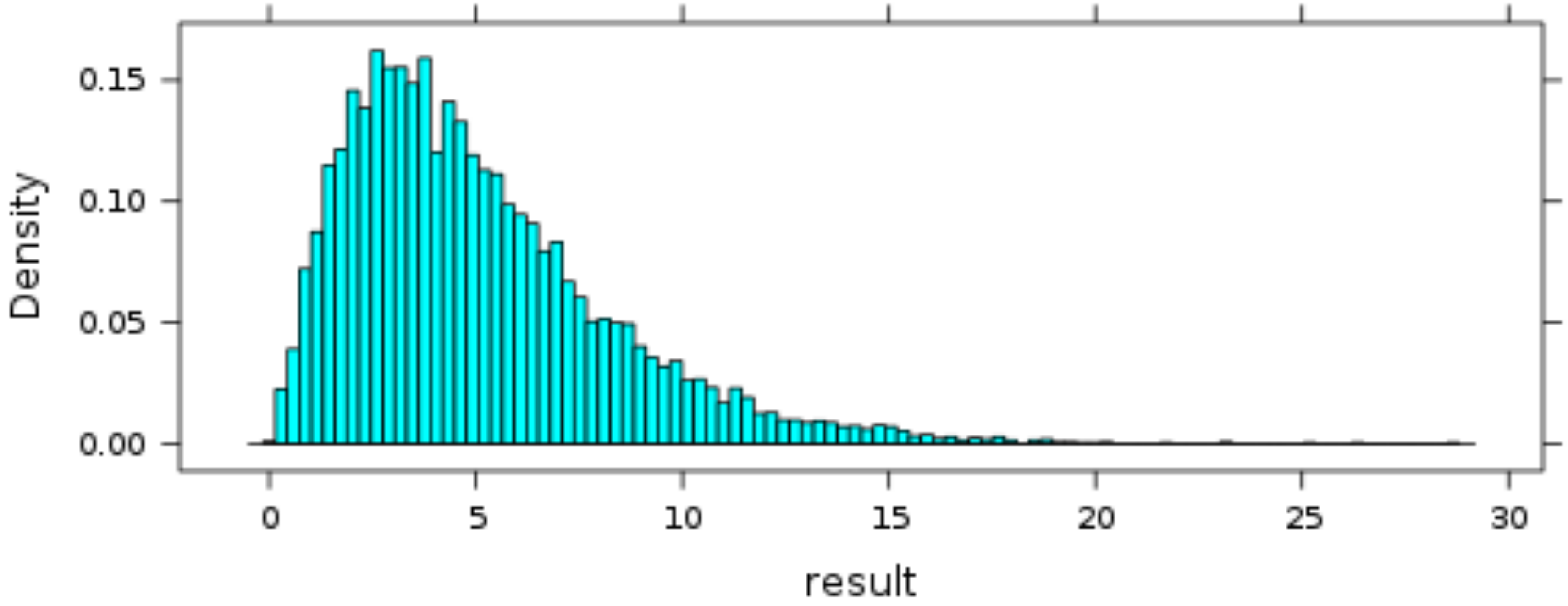


simulated_counts <- rmultinom(1, n, expected_proportions)

simulated_counts <- rmultinom(1, 726, c(1/6, 1/6, 1/6, 1/6, 1/6, 1/6))

Calculate statistic χ^2 and repeat 10,000 times

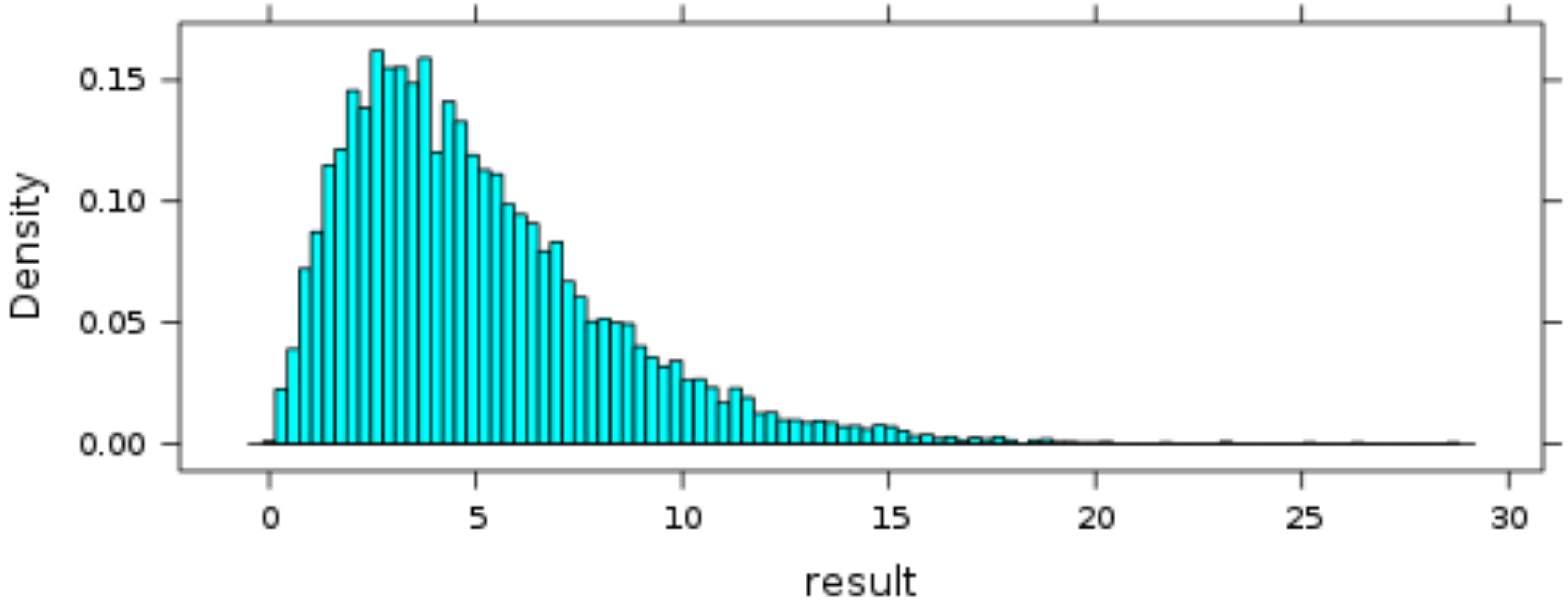
Randomization test to determine significance of a chi-square statistic



$\chi^2 = 26.21$

P-value = 0

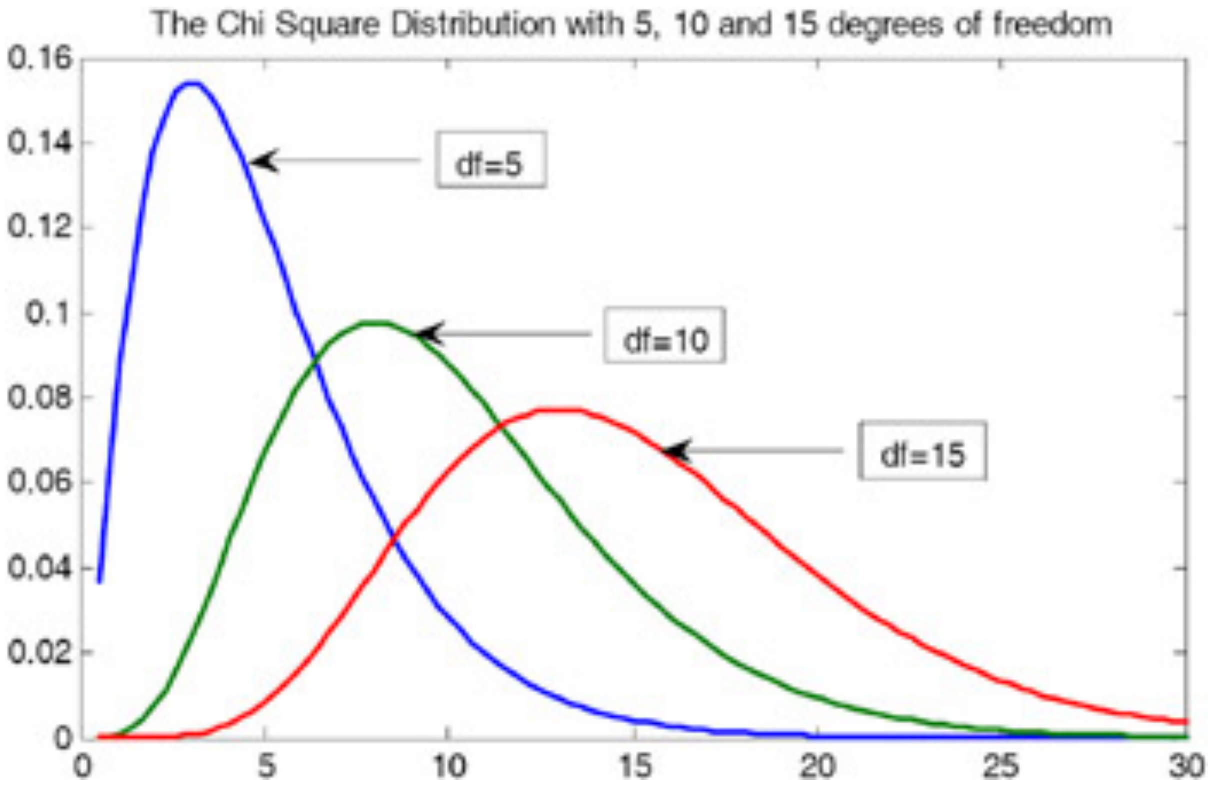
Randomization test to determine significance of a chi-square statistic



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

Yes! The χ^2 distribution!

Chi-square distribution

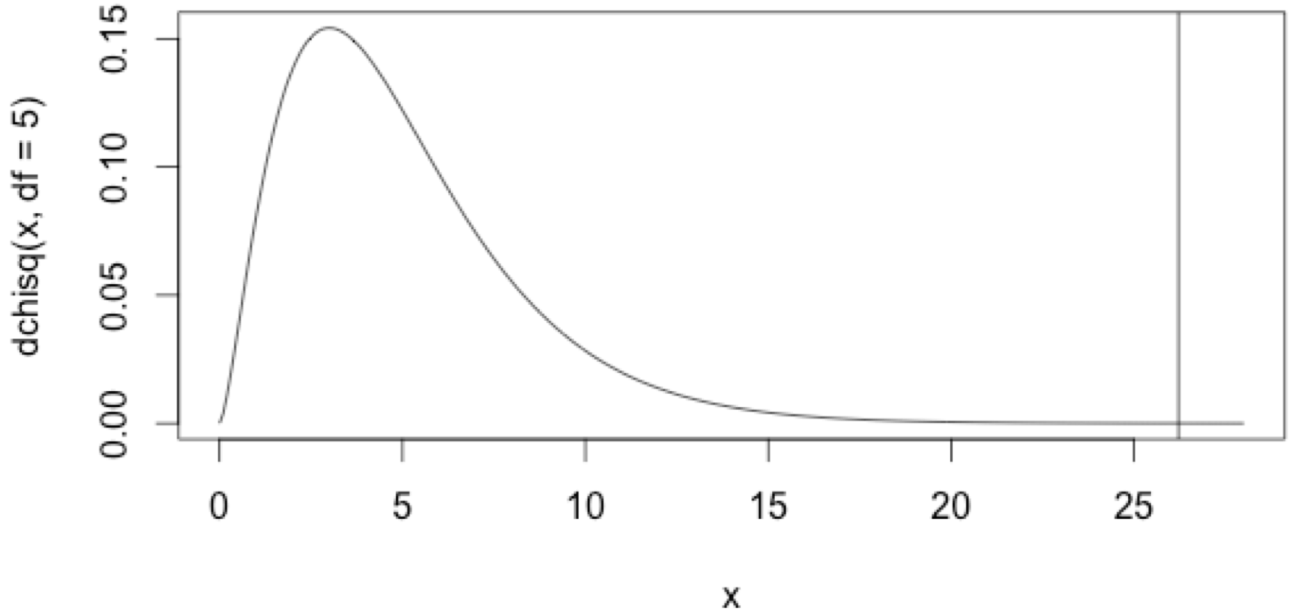


k is the number of groups

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

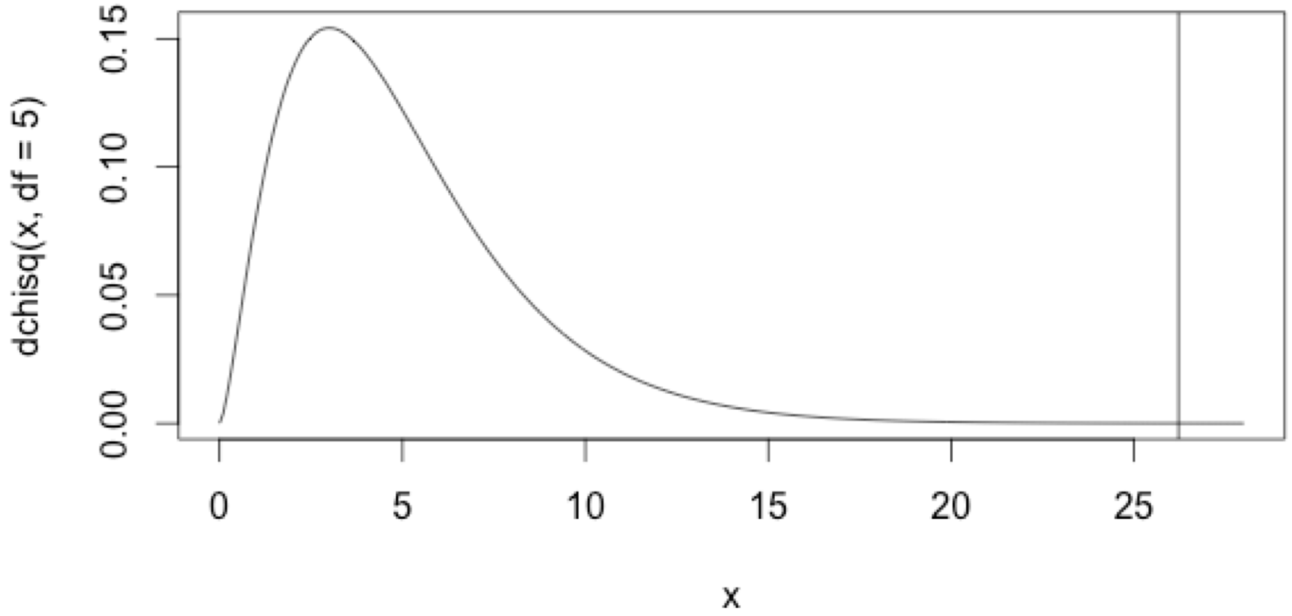
Chi-square distribution



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 the sprinkle example, since there was 6 sprinkle colors there are 5 degrees of freedom (df = 5).

P-value based on the chi-squared distribution = 0.00008

Let’s try it in R

But first…

Homework 10