

Goodness of Fit
hypothesis testing:

Pearson's chi-square test

Did you know that M&M's[®] Milk Chocolate Candies are supposed to come in the following percentages: 24% blue, 20% orange, 16% green, 14% yellow, 13% red, 13% brown?

<http://www.scientificameriken.com/candy5.asp>

“To our surprise M&Ms met our demand to review their procedures in determining candy ratios. It is, however, noted that the figures presented in their email differ from the information provided from their website (<http://us.mms.com/us/about/products/milkchocolate/>). An email was sent back informing them of this fact. To which M&Ms corrected themselves with one last email:

In response to your email regarding M&M'S CHOCOLATE CANDIES

Thank you for your email.

On average, our new mix of colors for M&M'S[®] Chocolate Candies is:

M&M'S[®] Milk Chocolate: 24% blue, 20% orange, 16% green, 14% yellow, 13% red, 13% brown.

M&M'S[®] Peanut: 23% blue, 23% orange, 15% green, 15% yellow, 12% red, 12% brown.

M&M'S[®] Kids MINIS[®]: 25% blue, 25% orange, 12% green, 13% yellow, 12% red, 13% brown.

M&M'S[®] Crispy: 17% blue, 16% orange, 16% green, 17% yellow, 17% red, 17% brown.

M&M'S[®] Peanut Butter and Almond: 20% blue, 20% orange, 20% green, 20% yellow, 10% red, 10% brown.

Have a great day!

Your Friends at Masterfoods USA
A Division of Mars, Incorporated



How to accept or reject the null hypothesis that these probabilities are correct from a finite sample?

Pearson χ^2 Goodness of Fit Test

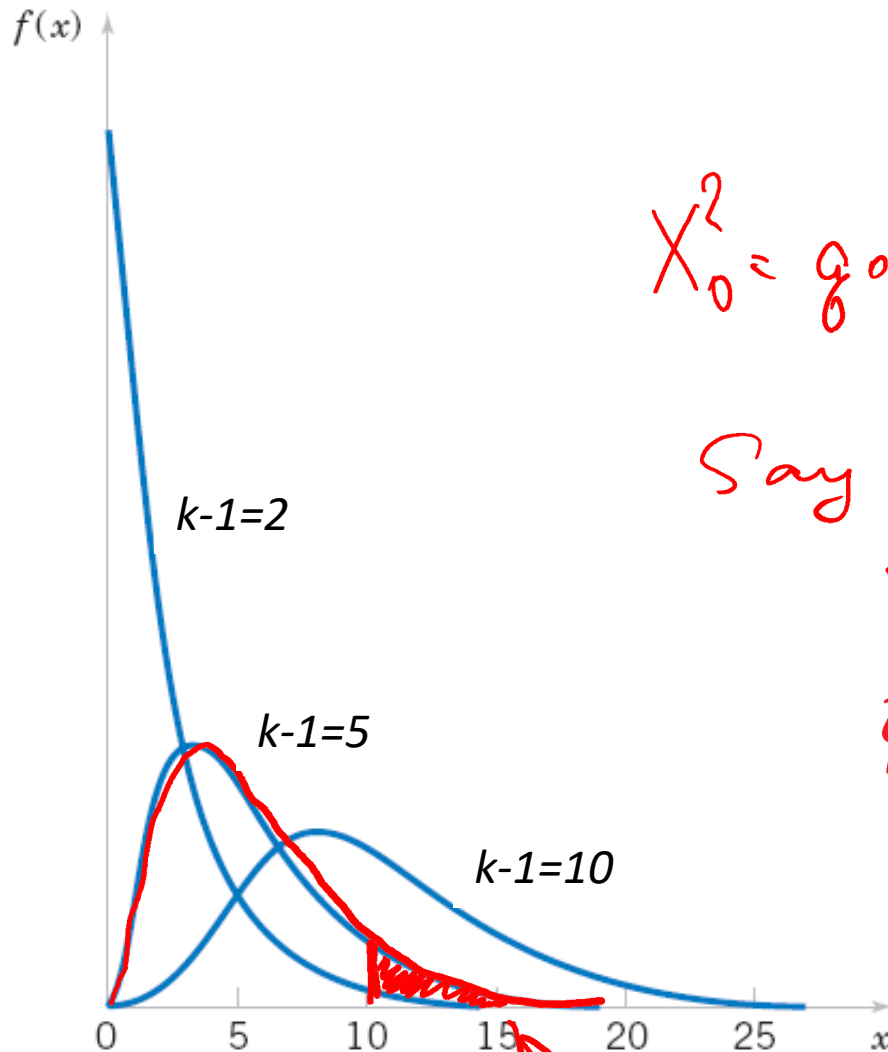
- Assume there is a **sample of size n** from a population with **k classes** (e.g. 6 M&M colors)
- **Null hypothesis** H_0 : class i has frequency f_i in the population
- **Alternative hypothesis** H_1 : some population frequencies are inconsistent with f_i
- Let O_i be the **observed number** of sample elements in the i th class and $E_i = n f_i$ be the **expected number** of sample elements in the i th class.
- **Group any bin** with $E_i < 3$ with
 - a) if numerical value of i is important, group it with its neighbor ($k=i-1$ or $k=i+1$) which has the smallest E_k until $E_{group} \geq 3$;
 - b) If numerical value of i is irrelevant, group together all $E_i < 3$ bins until $E_{group} \geq 3$
- The **test statistic** is

$$X_0^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (9-47)$$

P-value is calculated based on the **chi-square distribution** with **$k-1$ degrees of freedom**:

$$\text{P-value} = \text{Prob}(H_0 \text{ is correct}) = 1 - \text{CDF_chi-squared}(X_0^2, k-1)$$

chi² Goodness of Fit Test is a one-sided hypothesis



$$\chi^2_{\text{gof}} = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

Say $\chi^2_{\text{gof}} = 10$

For M&M

$$k = 6 \rightarrow k-1 = 5$$

χ^2_{gof}

p-value

that null hypothesis
is correct

M&M group exercise

- **DO NOT EAT CANDY BEFORE COUNTING IS FINISHED!**
THEN, PLEASE, DO.
- We will be testing three null hypotheses one after another:
 - M&M official data: 24% blue, 20% orange, 16% green, 14% yellow, 13% red, 13% brown
 - Website (fan collected) data from <http://joshmadison.com/2007/12/02/mms-color-distribution-analysis>:
18.36% blue, 20.76% orange, 18.44% green, 14.08% yellow, 14.20% red, 14.16% brown
 - Uniform distribution: 1/6~16.67% of each candy color
- You will estimate P-values for each one of these null hypotheses
- Hints: O_i – is the observed # of candies of color i ;
calculate the expected # $E_i = (\# \text{ candies in your sample}) * f_i$

Use **1-chi2cdf(X0squared, 5)** for P-value

$$X_0^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

M&M matlab exercise

- `observed=mm_table(group,:); group % use when analyzing one group`
- `f_mm=[0.24,0.2,0.16, 0.14, 0.13,0.13];`
- `f_u=1./6.*ones(1,6);`
- `f_website=[18,21,18,14,14,14,14];`
- `f_website=f_website./sum(f_website);`
- `%p_website=[0.1836, 0.2076, 0.1844, 0.1408, 0.1420, 0.1416]`
- `%p_u=[0.1500, 0.2200, 0.2100, 0.1200, 0.1600, 0.1500];`
- `n=sum(observed)`
- `expected_u=n.*f_u;`
- `expected_mm=n.*f_mm;`
- `expected_website=n.*f_website;`
- `gf_mm=0; gf_u=0; gf_website=0;`
- `for m=1:6;`
- `gf_mm=gf_mm+(observed(m)...`
- `-expected_mm(m)).^2./expected_mm(m);`
- `gf_u=gf_u+(observed(m)-expected_u(m)).^2./expected_u(m);`
- `gf_website=gf_website+(observed(m)...`
- `-expected_website(m)).^2./expected_website(m);`
- `end;`
- `disp('goodness of fit of MM ='); disp(num2str(gf_mm));`
- `disp('p-value of MM ='); disp(num2str(1-chi2cdf(gf_mm,5))); disp(' ');`
- `disp('goodness of fit of website ='); disp(num2str(gf_website));`
- `disp('p-value of MM ='); disp(num2str(1-chi2cdf(gf_website,5))); disp(' ');`
- `disp('goodness of fit of uniform ='); disp(num2str(gf_u));`
- `disp('p-value of uniform='); disp(num2str(1-chi2cdf(gf_u,5)));`

Statistical tests of independence

How to test the hypothesis if multiple samples are drawn from the same population?

- Table: **samples (Student groups) – rows**, **classes (M&M colors) – columns**
- Test if color fractions are independent from group
- **$P(\text{Group 1 and Color = green}) = P(\text{Group 1}) * P(\text{Color green})$**
- Compute for all groups/colors $6 * 4 = 24$ in our case

$$E_{\text{green}}(\text{group 1}) = n_{\text{tot}} * (\text{group 1} / n_{\text{tot}}) * (\text{green} / n_{\text{tot}})$$

- $$\chi^2 = \sum_{\text{groups \& colors}}^{n_{\text{tot}}} \frac{(O_{\text{color}}(\text{group}) - E_{\text{color}}(\text{group}))^2}{E_{\text{color}}(\text{group})}$$
- # degrees of freedom = **(colors-1) * (groups-1)**

- M&M exercise Spring 2024
- Was the M&M box from Costco well mixed?
Let's compare the first two groups' data

Title	Blue	Orange	Green	Yellow	Red	Brown	Sample	Origin
group 1	29	22	34	45	41	14	185	Costco
group 2	30	28	25	43	44	27	197	Costco
all Costco	59	50	59	88	85	41	382	
							0	
group 3	44	30	52	10	50	27	213	Schnuck
group 4	53	31	58	17	41	30	230	Schnuck
all Schnuck	97	61	110	27	91	57	443	

- Using $\chi^2 = \sum_{groups \& \ colors}^{24} \frac{(O_{color}(group) - E_{color}(group))^2}{E_{color}(group)}$

with # degrees of freedom $(colors-1) * (groups-1)$

Find P-value of null hypothesis H_0 that
samples are independent from each other

Was the Costco box well mixed?

- `clear mm_table`
- `mm_table=mm_table_all(1:2,:);`
- `ngroups=2;`
- `ncolors=6;`
- `sumt=sum(sum(mm_table))`
- `sum_color=sum(mm_table, 1)`
- `sum_group=sum(mm_table, 2)`
- `mm_exp=kron(sum_group,sum_color)./sumt`
- `gof=sum(sum((mm_table-mm_exp).^2./mm_exp))`
- `P_value_gof=1-chi2cdf(gof, (ngroups-1)*(ncolors-1))`
- **`%gof = 6.0121; P_value_gof = 0.3050`**
- **The null model that samples are independent is not rejected → The Costco box was well mixed!**

Batch effect

Does color composition vary between Costco and Schnucks

- Costco: 59 50 59 88 85 41
- Schnucks: 97 61 110 27 91 57
- Test if they are significantly different from each other:
- Same statistical independence test:
ngroups=2; ncolors=6;
- Results:
Goodness of Fit = 56.7101
P-value = 5.8028e-11
- Batch effect is **highly statistically significant!**
Costco and Schnucks do not represent the same population

Do Costco (groups 1 and 2) and Schnucks (groups 3 and 4) data come from the same population (factory?)

- `clear mm_table`
- `mm_table(1,:)=sum(mm_table_all(1:2,:));`
- `mm_table(2,:)=sum(mm_table_all(3:4,:));`
- `ngroups=2;`
- `ncolors=6;`
- `sumt=sum(sum(mm_table))`
- `sum_color=sum(mm_table, 1)`
- `sum_group=sum(mm_table, 2)`
- `mm_exp=kron(sum_group,sum_color)./sumt`
- `gof=sum(sum((mm_table-mm_exp).^2./mm_exp))`
- `P_value_gof=1-chi2cdf(gof, (ngroups-1)*(ncolors-1))`
- **% Goodness of Fit = 56.7101**
- **% P-value = 5.8028e-11**
- The null model that samples are independent is **rejected**
- **Costco and Schnucks get candy from different factories**

Credit: XKCD
comics

WHY ARE THERE SLAVES IN THE BIBLE

WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS
WHY IS HTTPS CROSSED OUT IN RED
WHY IS THERE A LINE THROUGH HTTPS
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK
WHY IS HTTPS IMPORTANT

QUESTIONS FOUND IN GOOGLE AUTOCOMLETE



WHY ARE THERE WEEKS
WHY DO I FEEL DIZZY
WHY ARE THERE ZIPPERS
ON ALL ZIPPERS

WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS
WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS
WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY
WHY ARE THERE SO MANY SVCHOST.EXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT
WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES
WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS
WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD
WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS
WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO
WHY IS OHIO WEATHER SO WEIRD

WHY AREN'T ECONOMISTS RICH
WHY DO AMERICANS CALL IT SOCCER
WHY ARE MY EARS RINGING
WHY ARE THERE SO MANY AVENGERS
WHY ARE THE AVENGERS FIGHTING THE X MEN
WHY IS WOLVERINE NOT IN THE AVENGERS

WHY ARE THERE SWARMS OF GNATS
WHY IS THERE PHLEGM
WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN
WHY IS PSYCHIC WEAK TO BUG
WHY DO CHILDREN GET CANCER
WHY IS POSEIDON ANGRY WITH ODYSSEUS
WHY IS THERE ICE IN SPACE

WHY ARE THERE ANTS IN MY LAPTOP

WHY ARE THERE BRIDESMAIDS
WHY DO DYING PEOPLE REACH UP
WHY AREN'T THERE VARICOSE ARTERIES
WHY ARE OLD KLINGONS DIFFERENT



WHY ARE THERE TINY SPIDERS IN MY HOUSE
WHY DO SPIDERS COME INSIDE
WHY ARE THERE HUGE SPIDERS IN MY HOUSE
WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE
WHY ARE THERE SPIDERS IN MY ROOM
WHY ARE THERE SO MANY SPIDERS IN MY ROOM
WHY DO SPIDER BITES ITCH
WHY IS DYING SO SCARY



WHY IS THERE AN OWL IN MY BACKYARD
WHY IS THERE AN OWL OUTSIDE MY WINDOW
WHY IS THERE AN OWL ON THE DOLLAR BILL
WHY DO OWLS ATTACK PEOPLE
WHY ARE AK 47s SO EXPENSIVE
WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE
WHY ARE THERE GODS
WHY ARE THERE TWO SPOCKS

WHY IS PROGRAMMING SO HARD
WHY IS THERE A 0 OHM RESISTOR
WHY DO AMERICANS HATE SOCCER
WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON CNN
WHY AREN'T POKEMON REAL
WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL

WHY IS THERE NO GPS IN LAPTOPS
WHY DO KNEES CLICK
WHY AREN'T THERE E GRADES
WHY IS ISOLATION BAD
WHY DO BOYS LIKE ME
WHY DON'T BOYS LIKE ME
WHY IS THERE ALWAYS A JAVA UPDATE
WHY ARE THERE RED DOTS ON MY THIGHS
WHY IS LYING GOOD



WHY IS MT VESUVIUS THERE
WHY DO THEY SAY T MINUS
WHY ARE THERE OBELISKS
WHY ARE WRESTLERS ALWAYS WET
WHY ARE OCEANS BECOMING MORE ACIDIC
WHY IS ARWEN DYING
WHY AREN'T MY QUAIL LAYING EGGS
WHY AREN'T MY QUAIL EGGS HATCHING
WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA

WHY ARE CIGARETTES LEGAL
WHY ARE THERE DUCKS IN MY POOL
WHY IS JESUS WHITE
WHY IS THERE LIQUID IN MY EAR
WHY DO Q TIPS FEEL GOOD
WHY DO GOOD PEOPLE DIE



WHY ARE ULTRASOUNDS IMPORTANT
WHY ARE ULTRASOUND MACHINES EXPENSIVE
WHY IS STEALING WRONG

WHY ARE DOGS AFRAID OF FIREWORKS
WHY IS THERE NO KING IN ENGLAND

Goodness of fit with a PDF defined by **m** parameters

- As before: **k** classes (e.g. M&M colors)
- Use **parameter estimators** to find **the best parameters** for the fit
 - Method of moments
 - MLE: method of maximum likelihood
- Use chi-squared distribution with **k-1-m** degrees of freedom
- As before: if $E_i < 3$, group it together with another group and reduce **k** by 1

$$X_0^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad (9-47)$$

9-7 Testing for Goodness of Fit

Example 9-12

EXAMPLE 9-12 Printed Circuit Board Defects Poisson Distribution

The number of defects in printed circuit boards is hypothesized to follow a Poisson distribution. A random sample of $n = 60$ printed boards has been collected, and the following number of defects observed.

Number of Defects	Observed Frequency
0	32
1	15
2	9
3	4

9-7 Testing for Goodness of Fit

Example 9-12

The mean of the assumed Poisson distribution in this example is unknown and must be estimated from the sample data. The estimate of the mean number of defects per board is the sample average, that is, $(32 \cdot 0 + 15 \cdot 1 + 9 \cdot 2 + 4 \cdot 3) / 60 = 0.75$. From the Poisson distribution with parameter 0.75, we may compute p_i , the theoretical, hypothesized probability associated with the i th class interval. Since each class interval corresponds to a particular number of defects, we may find the p_i as follows:

$$p_1 = P(X = 0) = \frac{e^{-0.75}(0.75)^0}{0!} = 0.472$$

$$p_2 = P(X = 1) = \frac{e^{-0.75}(0.75)^1}{1!} = 0.354$$

$$p_3 = P(X = 2) = \frac{e^{-0.75}(0.75)^2}{2!} = 0.133$$

$$p_4 = P(X \geq 3) = 1 - (p_1 + p_2 + p_3) = 0.041$$

9-7 Testing for Goodness of Fit

Example 9-12

The expected frequencies are computed by multiplying the sample size $n = 60$ times the probabilities p_i . That is, $E_i = np_i$. The expected frequencies follow:

Number of Defects	Probability	Expected Frequency
0	0.472	28.32
1	0.354	21.24
2	0.133	7.98
3 (or more)	0.041	2.46

9-7 Testing for Goodness of Fit

Example 9-12

Since the expected frequency in the last cell is less than 3, we combine the last two cells:

Number of Defects	Observed Frequency	Expected Frequency
0	32	28.32
1	15	21.24
2 (or more)	13	10.44

The chi-square test statistic in Equation 9-47 will have $k - p - 1 = 3 - 1 - 1 = 1$ degree of freedom, because the mean of the Poisson distribution was estimated from the data.

9-7 Testing for Goodness of Fit

Example 9-12

The seven-step hypothesis-testing procedure may now be applied, using $\alpha = 0.05$, as follows:

1. **Parameter of interest:** The variable of interest is the form of the distribution of defects in printed circuit boards.
2. **Null hypothesis:** H_0 : The form of the distribution of defects is Poisson.
3. **Alternative hypothesis:** H_1 : The form of the distribution of defects is not Poisson.
4. **Test statistic:** The test statistic is

$$\chi_0^2 = \sum_{i=1}^k \frac{(o_i - E_i)^2}{E_i}$$

9-7 Testing for Goodness of Fit

Example 9-12

5. **Reject H_0 if:** Reject H_0 if the P -value is less than 0.05.

6. **Computations:**

$$\chi_0^2 = \frac{(32 - 28.32)^2}{28.32} + \frac{(15 - 21.24)^2}{21.24} + \frac{(13 - 10.44)^2}{10.44} = 2.94$$

7. **Conclusions:** We find from Appendix Table III that $\chi_{0.10,1}^2 = 2.71$ and $\chi_{0.05,1}^2 = 3.84$. Because $\chi_0^2 = 2.94$ lies between these values, we conclude that the P -value is between 0.05 and 0.10. Therefore, since the P -value exceeds 0.05 we are unable to reject the null hypothesis that the distribution of defects in printed circuit boards is Poisson. The exact P -value computed from Minitab is 0.0864.

Credit: XKCD
comics

WHY ARE THERE SLAVES IN THE BIBLE

WHY DO TWINS HAVE DIFFERENT FINGERPRINTS
WHY ARE AMERICANS AFRAID OF DRAGONS

WHY IS HTTPS CROSSED OUT IN RED
WHY IS THERE A LINE THROUGH HTTPS
WHY IS THERE A RED LINE THROUGH HTTPS ON FACEBOOK
WHY IS HTTPS IMPORTANT

QUESTIONS

FOUND IN GOOGLE AUTOCOMPLETE



WHY ARE THERE WEEKS
WHY DO I FEEL DIZZY

WHY AREN'T ECONOMISTS RICH

WHY ARE THERE SO MANY CROWS IN ROCHESTER, MN
WHY IS THERE PHLEGM

WHY DO AMERICANS CALL IT SOCCER

WHY IS PSYCHIC WEAK TO BUG

WHY ARE MY EARS RINGING

WHY DO CHILDREN GET CANCER

WHY ARE THERE SO MANY AVENGERS

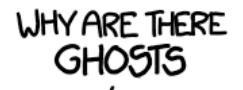
WHY IS POSEIDON ANGRY WITH ODYSSEUS

WHY ARE THE AVENGERS FIGHTING THE X MEN

WHY IS THERE ICE IN SPACE

WHY ARE THERE ANTS IN MY LAPTOP

WHY IS EARTH TILTED



WHY IS THERE AN OWL IN MY BACKYARD

WHY IS SPACE BLACK



WHY IS THERE AN OWL OUTSIDE MY WINDOW

WHY IS OUTER SPACE SO COLD

WHY IS THERE AN OWL ON THE DOLLAR BILL

WHY ARE THERE PYRAMIDS ON THE MOON

WHY DO OWLS ATTACK PEOPLE

WHY IS NASA SHUTTING DOWN

WHY ARE AK 47s SO EXPENSIVE

WHY ARE THERE MALE AND FEMALE BIKES

WHY ARE THERE HELICOPTERS CIRCLING MY HOUSE

WHY ARE THERE TINY SPIDERS IN MY HOUSE

WHY ARE THERE GODS

WHY DO SPIDERS COME INSIDE

WHY ARE THERE TWO SPOCKS

WHY ARE THERE HUGE SPIDERS IN MY HOUSE

WHY IS LIFE SO BORING

WHY ARE THERE LOTS OF SPIDERS IN MY HOUSE

WHY ARE CIGARETTES LEGAL

WHY ARE THERE SPIDERS IN MY ROOM

WHY ARE THERE DUCKS IN MY POOL

WHY ARE THERE SO MANY SPIDERS IN MY ROOM

WHY IS JESUS WHITE

WHY DO SPIDER BITES ITCH

WHY IS THERE LIQUID IN MY EAR

WHY IS DYING SO SCARY

WHY DO Q TIPS FEEL GOOD



WHY ARE DOGS AFRAID OF FIREWORKS
WHY IS THERE NO KING IN ENGLAND

WHY DO WHALES JUMP
WHY ARE WITCHES GREEN
WHY ARE THERE MIRRORS ABOVE BEDS

WHY DO I SAY UH
WHY IS SEA SALT BETTER
WHY ARE THERE TREES IN THE MIDDLE OF FIELDS

WHY IS THERE NOT A POKEMON MMO
WHY IS THERE LAUGHING IN TV SHOWS
WHY ARE THERE DOORS ON THE FREEWAY

WHY ARE THERE SO MANY SVCHOST.EXE RUNNING
WHY AREN'T THERE ANY COUNTRIES IN ANTARCTICA
WHY ARE THERE SCARY SOUNDS IN MINECRAFT

WHY IS THERE KICKING IN MY STOMACH
WHY ARE THERE TWO SLASHES AFTER HTTP
WHY ARE THERE CELEBRITIES

WHY DO SNAKES EXIST
WHY DO OYSTERS HAVE PEARLS
WHY ARE DUCKS CALLED DUCKS

WHY DO THEY CALL IT THE CLAP
WHY ARE KYLE AND CARTMAN FRIENDS
WHY IS THERE AN ARROW ON AANG'S HEAD

WHY ARE TEXT MESSAGES BLUE
WHY ARE THERE MUSTACHES ON CLOTHES
WHY ARE THERE MUSTACHES ON CARS

WHY ARE THERE MUSTACHES EVERYWHERE
WHY ARE THERE SO MANY BIRDS IN OHIO
WHY IS THERE SO MUCH RAIN IN OHIO

WHY IS OHIO WEATHER SO WEIRD
WHY ARE THERE BRIDESMAIDS
WHY DO DYING PEOPLE REACH UP

WHY AREN'T THERE VARICOSE ARTERIES
WHY ARE OLD KUNGONS DIFFERENT

WHY IS PROGRAMMING SO HARD
WHY IS THERE A 0 OHM RESISTOR
WHY DO AMERICANS HATE SOCCER

WHY DO RHYMES SOUND GOOD
WHY DO TREES DIE
WHY IS THERE NO SOUND ON CNN

WHY AREN'T THERE DINOSAUR GHOSTS

WHY ARE THERE FEMALE MR NIMES

WHY IS SEX SO IMPORTANT



WHY IS GPS FREE

WHY IS THERE HELL IF GOD FORGIVES



WHY AREN'T BULLETS SHARP
WHY DO DREAMS SEEM SO REAL

WHY IS THERE ALWAYS A JAVA UPDATE

WHY ARE THERE RED DOTS ON MY THIGHS

WHY AREN'T MY QUAIL LAYING EGGS

WHY AREN'T MY QUAIL EGGS HATCHING

WHY AREN'T THERE ANY FOREIGN MILITARY BASES IN AMERICA