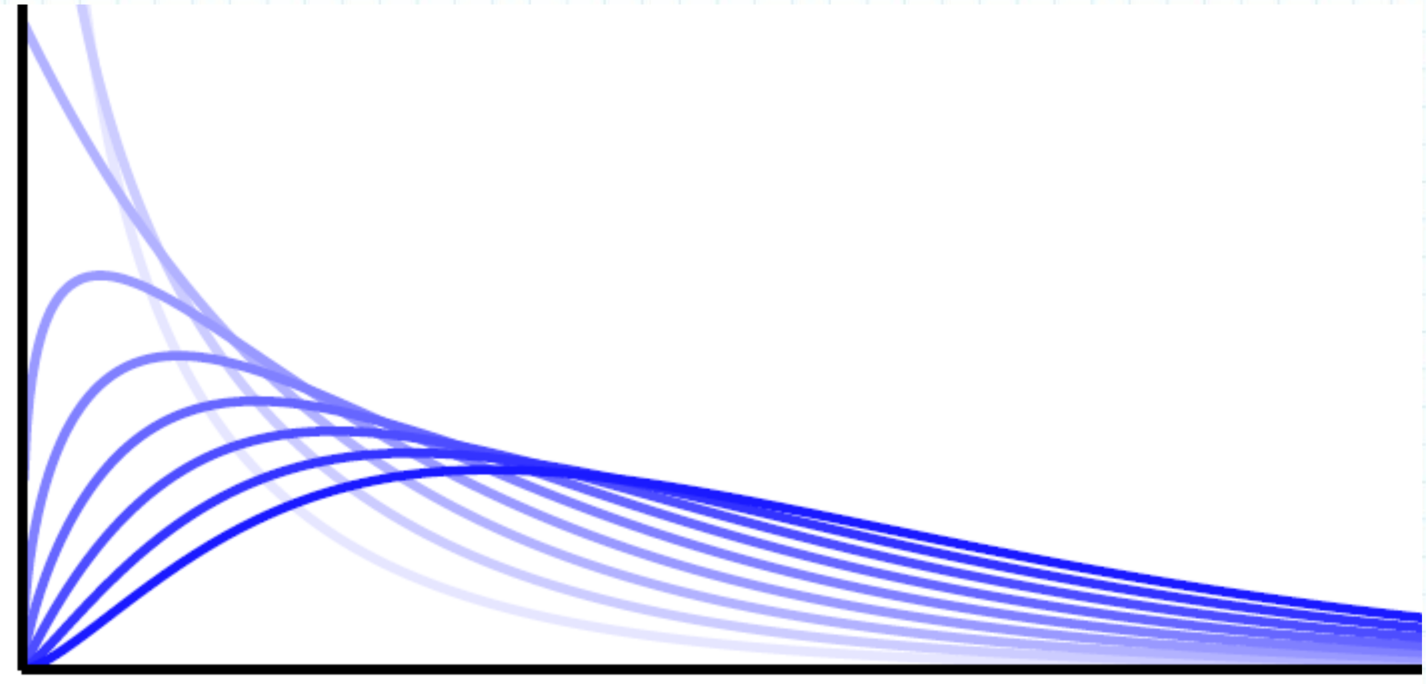
The χ^2 Tests

Math 122

The symbol χ is the Greek letter chi

Some χ^2 Distributions



We consider observed frequencies (counts) and how they match frequencies expected for an assumed distribution.

Goodness of Fit

- Goodness of Fit tests test claims that observations either do or do not match some claimed distribution.

- H_0: The observed frequencies match the claimed distribution.
- H_1: The observed frequencies do not match the claimed distribution.

Theoretical

Calculating Expected Frequencies

- Suppose that all pieces of a certain type of candy are red, blue, or green and that the colors are **uniformly distributed.**
- What are the expected numbers of each color among 200 pieces of candy?

For each color, expect $\frac{200}{3}$

Calculating Expected Frequencies

- Suppose that all pieces of a certain type of candy are red, blue, green or orange and that the colors match the distribution below.
- What are the expected numbers of each color among 200 pieces of candy?

Color	Relative Frequency
Red	0.1
Blue	0.4
Green	0.4
Orange	0.1

Color	Expected in 200
Red	200 x .1
Blue	200 x .4
Green	200 x .4
Orange	200 x .1

Test Statistic For Goodness of Fit

- O=observed frequency
- E=expected frequency

- $\chi^2 = \sum \frac{(O-E)^2}{E}$

Goodness of Fit Test

- If the observed frequencies O are **very different** from the expected frequencies E, then χ^2 will be **large**.
- In this case, we **reject H_0**.
- The observations **do not match** the expectations.

$(O-E)^2$

Goodness of Fit Test

- If the observed frequencies O are **not very different** from the expected frequencies E, then χ^2 will be **small**.
- In this case, we **do not reject H_0**.
- The observations **seem to match** the expectations.

$(O-E)^2$

- We suspect that the weights reported by a group of people are not actual measurements. If they were, the final digits in the weights would be uniformly distributed (all 10 digits equally likely).
- Use the frequency counts below to test the claim that **the final digits in these weights are not uniformly distributed.**

Digit	0	1	2	3	4	5	6	7	8	9
Count	7	14	0	16	8	4	5	6	12	8

Observed	Expected
7	8
14	8
0	8
16	8
8	8
4	8
5	8
6	8
12	8
8	8

+

$80 \div 10 = 8$

~~H0: The observed counts match the expected frequencies.~~

H1: The observed counts do not match the expected frequencies

claim

Df= $10-1=9$

P-value: 0.0019

Formal Conclusion:

Reject H_0 / Support H_1

Final conclusion:

The data support the claim that these #s are not uniformly distributed

Are the final digits approximately uniformly distributed?

NO

World Series

If teams in the World Series are equally matched and if each team is equally likely to win each game, then we can calculate the probability of a team winning the Series in a fixed number of games.

Number of Games	Probability
4	2/16
5	4/16
6	5/16
7	5/16

World Series

- Below is a table listing the number of games needed to win the World Series in recent span of 91 years.
- Test the claim that the actual frequencies match the distribution expected if the teams are equally matched.

These numbers are taken from baseball-almanac.com and only looks at seasons after 1921 (which was the last best-of-9 year).

Number of Games	Actual Count	Theoretical Probability
4	18	2/16
5	19	4/16
6	20	5/16
7	34	5/16

Observed	Expected
18	91×2/16
19	91x 4/16
20	91x 5/16
34	91 x 5/16

H0: The observed counts match the expected frequencies. Claim

H1: The observed counts do not match the expected frequencies

Df= 4 - 1 = 3

P-value: .0446

Formal Conclusion:

Reject H_0 / Support H_1

Final conclusion:

There is enough evidence to reject

the claim that the # of games required

to win the series matches the theoretical dist.

Does the number of game required to win the series match the distribution expected if the teams are evenly matched?

NO

Does the distribution match the distribution expected if one team is better than the other?

NO

Birthdays

Below are the number of students in this class born on each day of the week.

not

Test the claim that people are born on each day of the week with equal frequency.

Day	Count
Sunday	0
Monday	5
Tuesday	8
Wednesday	4
Thursday	4
Friday	10
Saturday	3

Observed	Expected
0	34÷7
5	34÷7
8	34÷7
4	34÷7
4	34÷7
10	34÷7
3	34÷7

+

34

H0: The observed counts match the expected frequencies.

H1: The observed counts do not match the expected frequencies

claim

Df= 7 - 1 = 6

P-value: .0378 ~~.0639~~

Formal Conclusion:

Reject Ho / Support H_1

Final conclusion:

The data supports the claim that ~~the~~ People are not born on each day of the week w/ equal Frequency

Does it look like people are born on each day of the week with equal frequency?

NO