Chi-Squared Examples

ECE 376: Embedded Systems

Lecture #15b

note: All lecture notes, homework sets, and solutions are posted on www.BisonAcademy.com

Chi-Squared Test

- Is the data consistent with an assumed distribution?

Procedure

- Collect Data
- Split into N bins
- Compare the expected frequency (np) for each bin vs. observed frequency (N)

$$\chi^2 = \sum \left(\frac{(np_i - N_i)^2}{np_i} \right)$$

- Use a chi-squred table to convert the chi-sqared score to a probability

This Lecture:

- Are world temperatures changing?
- Does the gain of a transistor have a uniform distribution?
- Does the gain of a transistor have a normal distribution?
- Am I psychic?

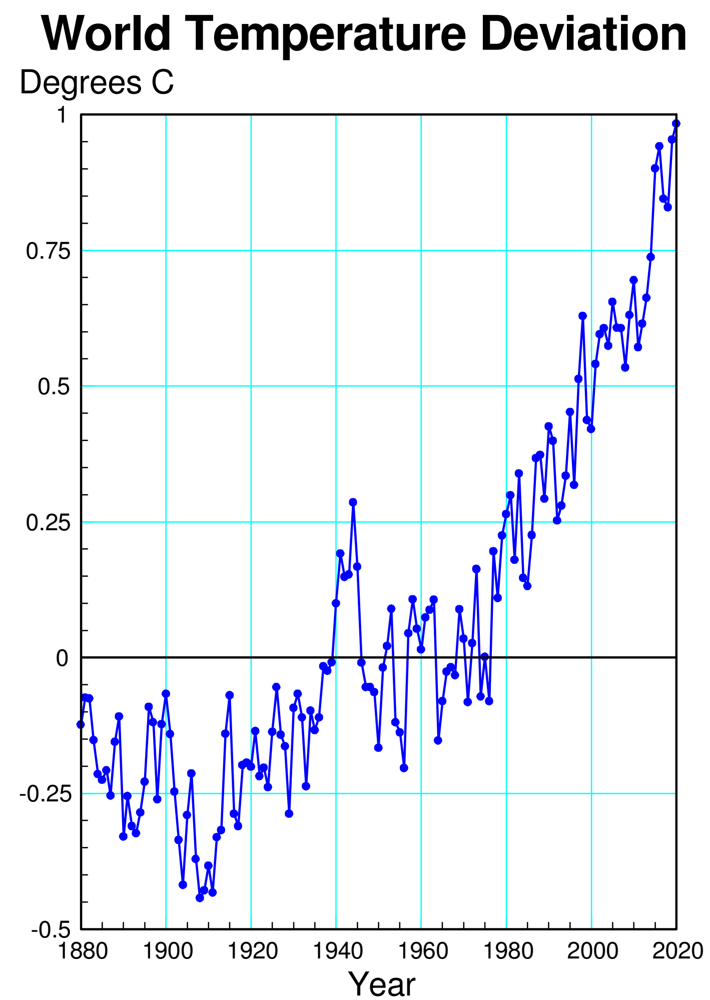
Are world temperatures changing?

https://www.ncdc.noaa.gov/cag/global/time-series/globe/land_ocean/p12/12/1880-2020.csv

NASA Goddard has been monitoring world temperatures since 1880.

- 8 of the past 10 years have been the hottest on record. Is this random?
- Is there a pattern?

These are actually chi-squared tests



8 of the Past 10 Years have been in the top-10 hottest years...

H0: Assume all years have equal probability of being in the top 10 hottest years

- p(hottest) = 10/141
- p(other) = 131/141

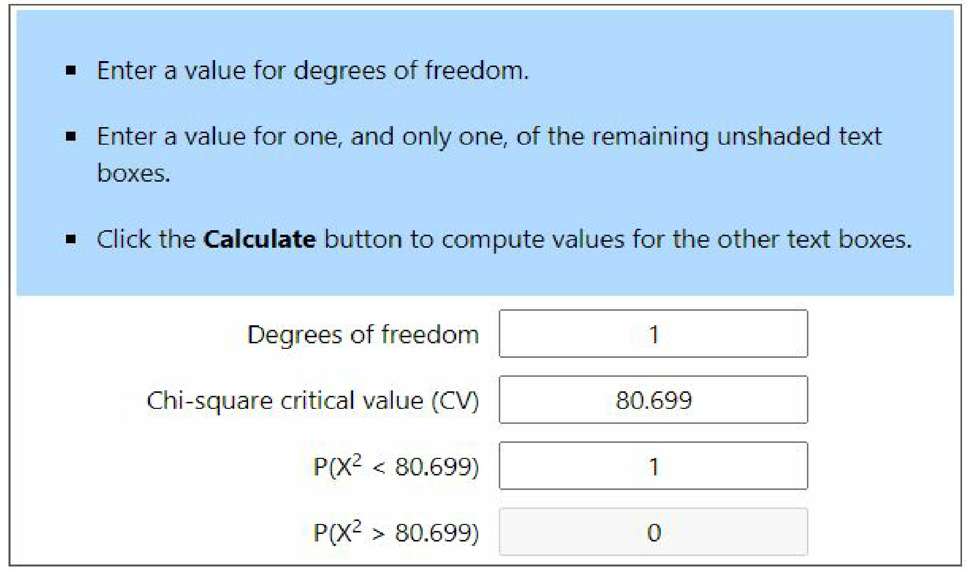
Set up a chi-squared table

	p	np	N	chi-squared
hottest 10	10/141	0.709	8	74.977
other	131/141	9.291	2	7.72
			Sum	80.699

Use a chi-squred table to convert 80.699 to a probability

p(reject) = 1.0000

- actually p(reject) > 0.99995
- (rounding)
- (nothing is 100% certain)
- There is at least a 99.995% chance that all years are *not* equally likely



You can calculate the odds

binomial distribution (coin toss)

$$p(m) = \binom{n}{m} p^m (1-p)^{n-m}$$

$$p(8) = \binom{10}{8} \left(\frac{10}{141}\right)^8 \left(\frac{131}{141}\right)^2$$

$$p(8) = 0.0000002486$$

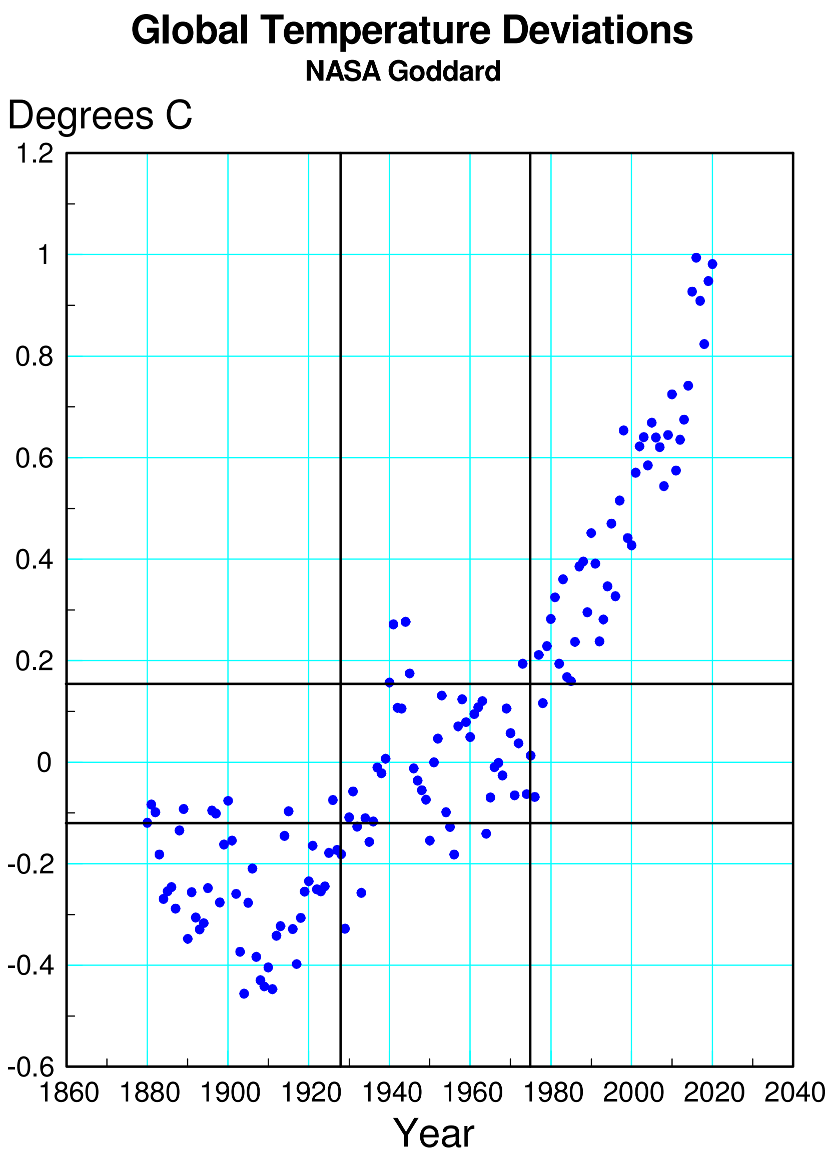
Is there a pattern?

A little harder to analyze the data for this question.

Split the data into 9 regions

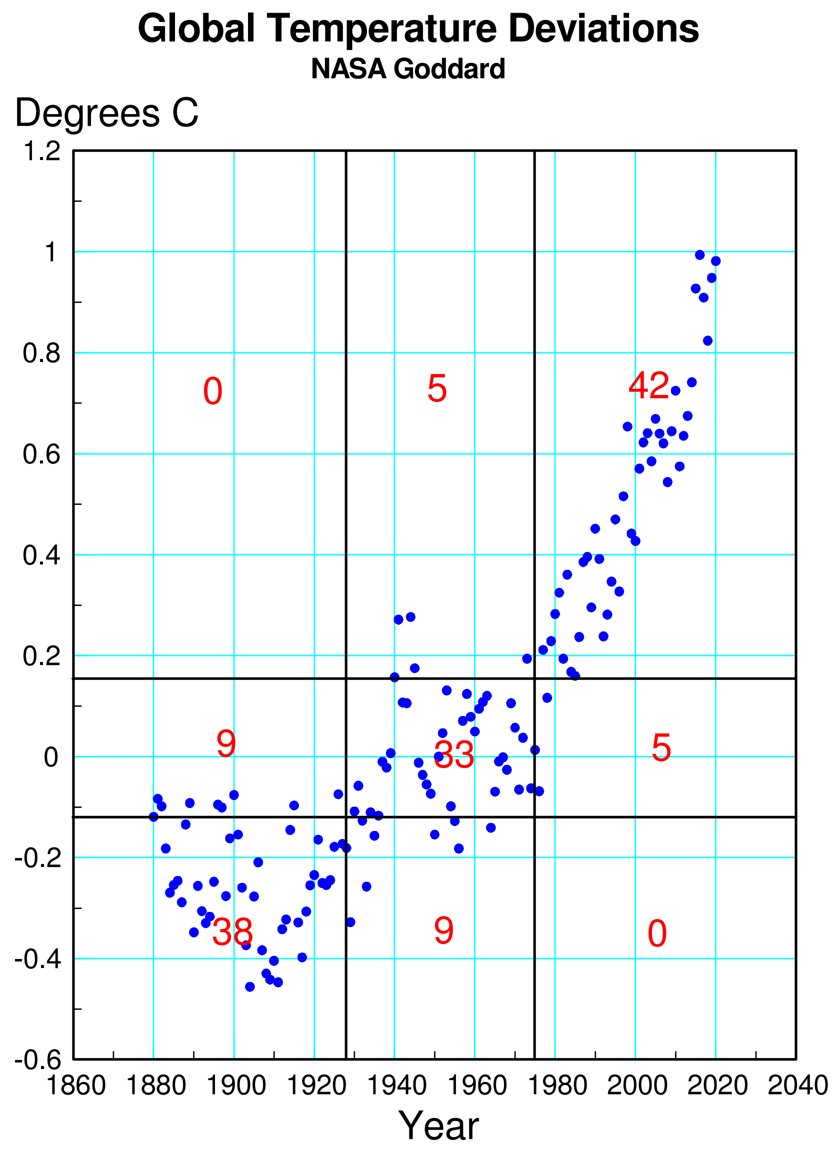
- First 47 years, middle 47 years, last 47 years
- Hottest 47 years, middle 47 years coldest 47 years

If there is no pattern, each region should contain 1/9th of the data



Chi-Squared Test

Region	p	np	N	chi-squared
(1,1)	1/9	15.67	0	15.67
(1,2)	1/9	15.67	5	7.27
(1,3)	1/9	15.67	42	44.24
(2,1)	1/9	15.67	9	2.84
(2,2)	1/9	15.67	33	19.17
(2,3)	1/9	15.67	5	7.27
(3,1)	1/9	15.67	38	31.82
(3,2)	1/9	15.67	9	2.84
(3,3)	1/9	15.67	0	15.67
			Sum	146.78

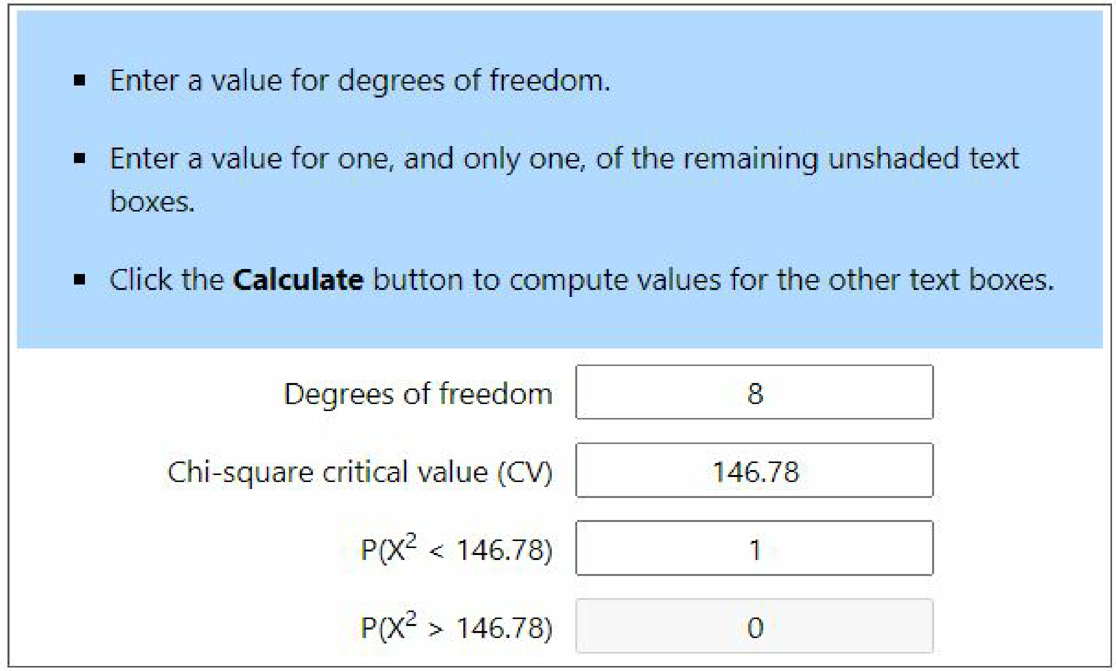


Use a chi-squared table to convert this to a probability

- StatTrek

Again, this is *way* off the chart

- 8 degrees of freedom
- chi-squared score of 146.78
- p(reject) > 0.99995
- Rounded to 1.0000
- (nothing is 100% certain)



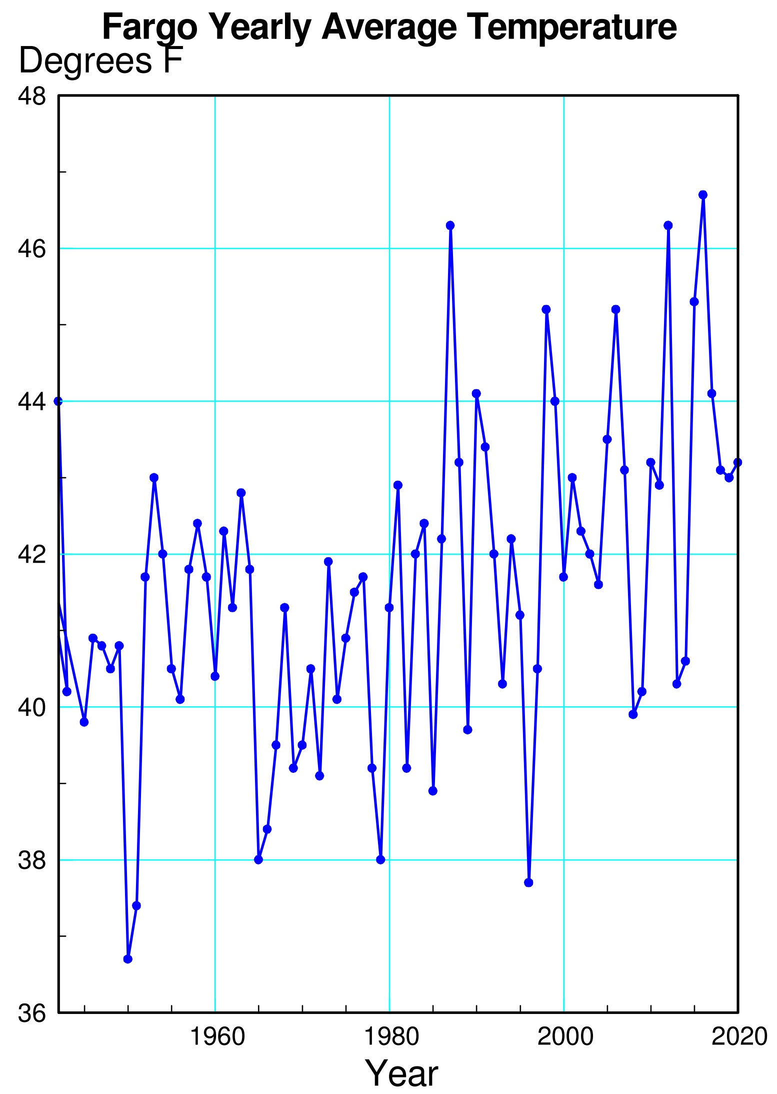
The data is almost certainly not random

Is Fargo Getting Warmer?

Data:

- Hector Airport has been measuring the temperature in Fargo since 1942
- High / average / low for each month and year
- https://www.wunderground.com/history/monthly/us/nd/fargo/KFAR/date/2020-7
- http://www.bisonacademy.com/ECE111/Code/Fargo_Weather_Monthly_Avg.txt

Use the yearly average since 1942

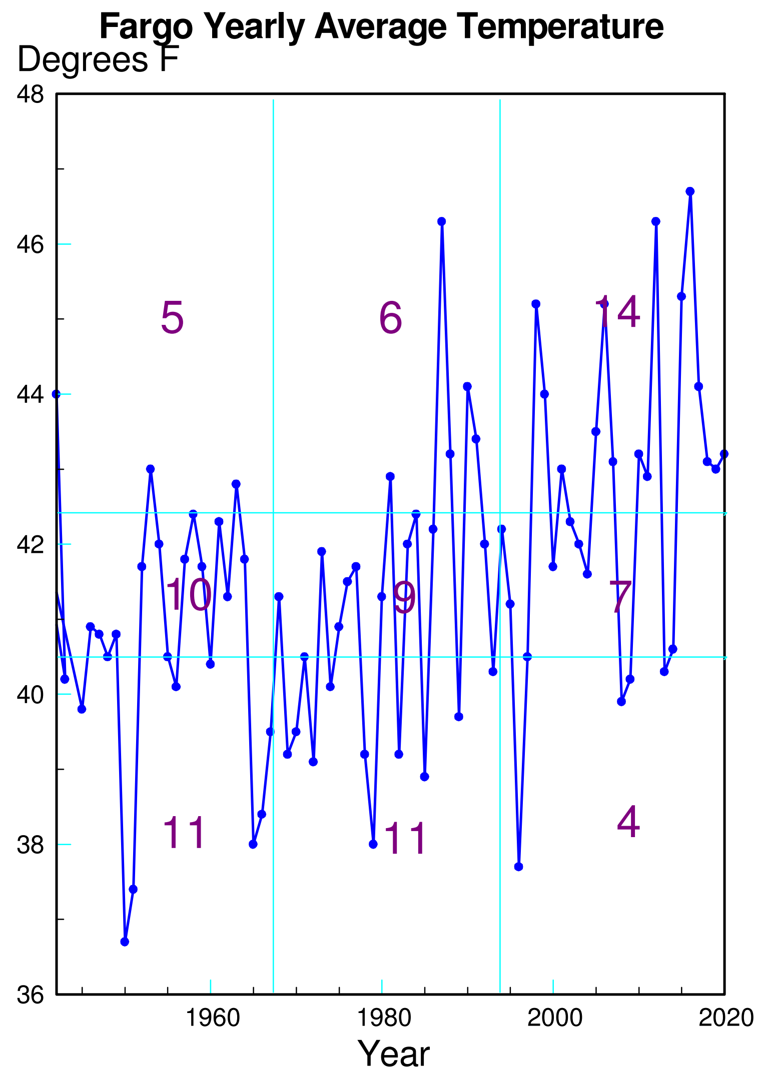


Procedure

There isn't a lot of data (79 data points).

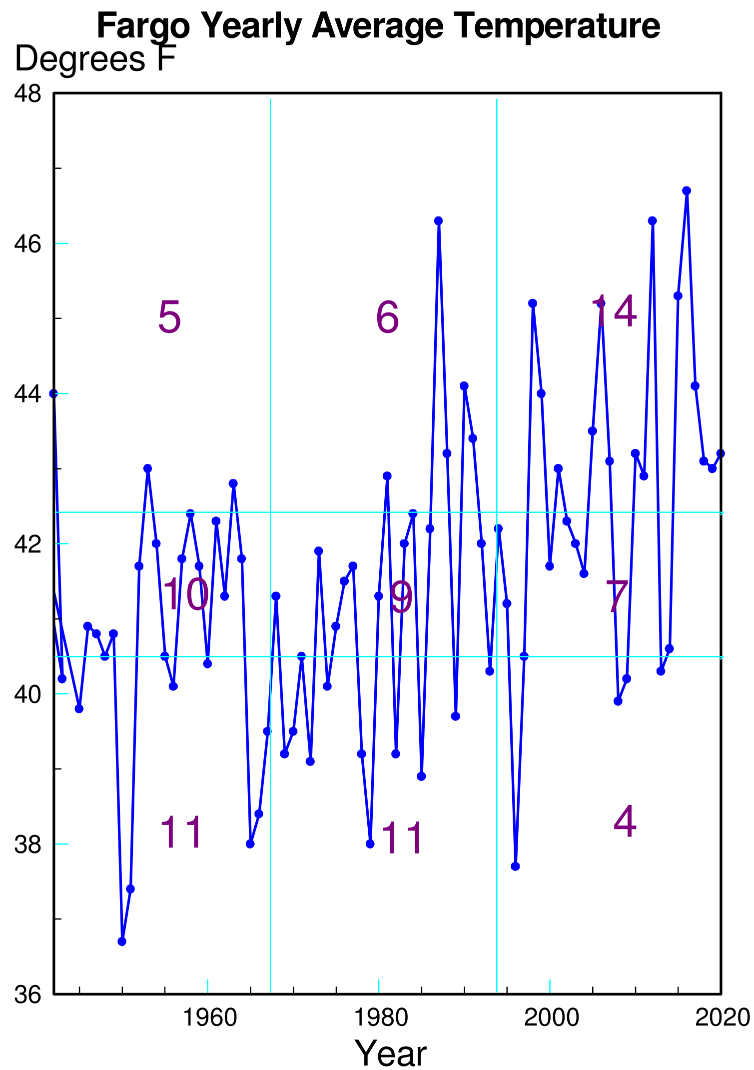
- Split into 9 bins (should get 8.77 events per bin)
- Split years into 3 intervals
- Split temperature into 3 tiers

Count how many times a given year falls into each bin



Chi-Squared Test

Years	Tier	np	Actual	Chi-Squared
1942 - 1967	hot	8.56	5	1.4806
	middle	8.56	10	0.2422
	cold	8.56	11	0.6955
1968 - 1993	hot	8.56	6	0.7656
	middle	8.56	9	0.0226
	cold	8.56	11	0.6955
1994 - 2020	hot	8.56	14	3.4572
	middle	8.56	7	0.2843
	cold	8.56	4	2.4292
Total				**10.0727**

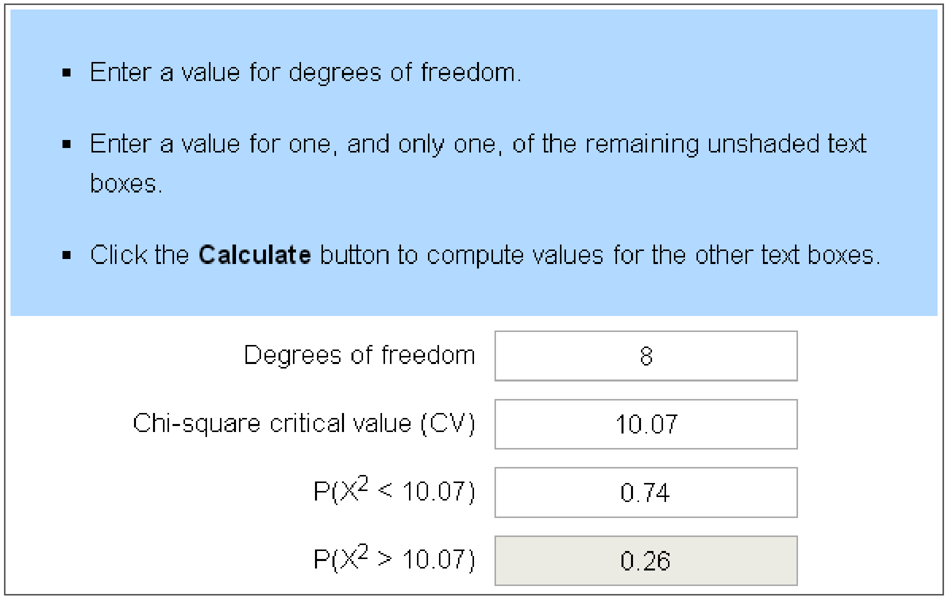


Interpreting the Result

Convert the chi-squared score to a probability

- Chi-squared table
- StatTrek

With 8 degrees of freedom (9 bins), a chi-squared score of 10.07 corresponds to a probability of at least 0.74



I'm 74% certain that the temperature in Fargo is changing

Does the gain of a transistor have a uniform distribution?

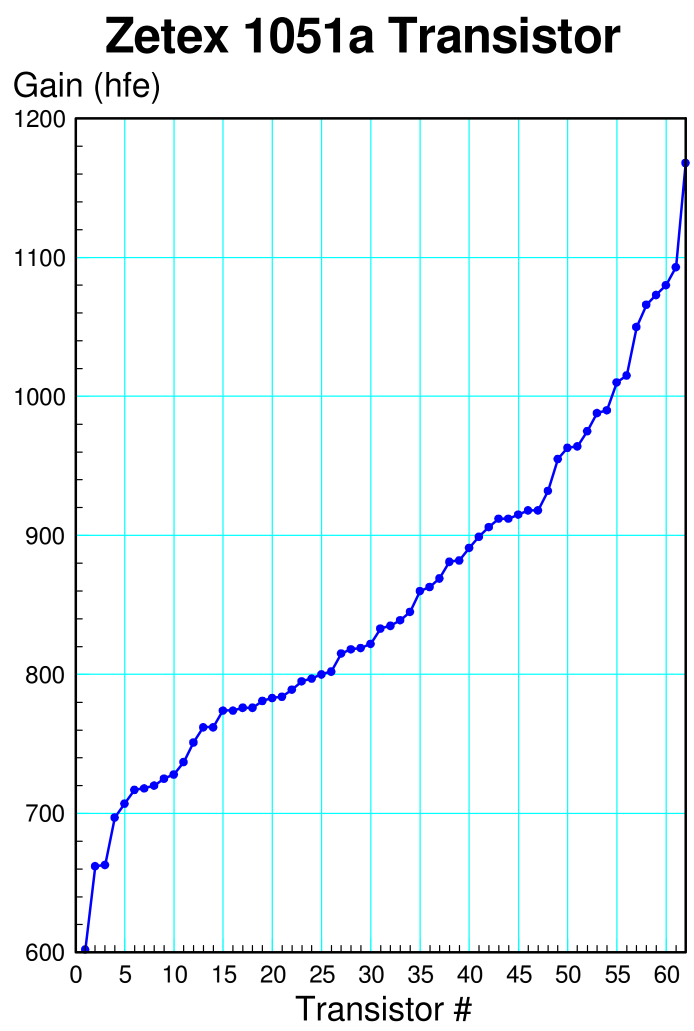
Each transistor's gain is slightly different.

Does a uniform distribution describe the variability in a transistor's gain?

Is the gain measured consistent with a uniform distribution?

Data:

- Measure the gain of 62 Zetex 1051a transistors
- Sort the gains and plot



Data Analysis

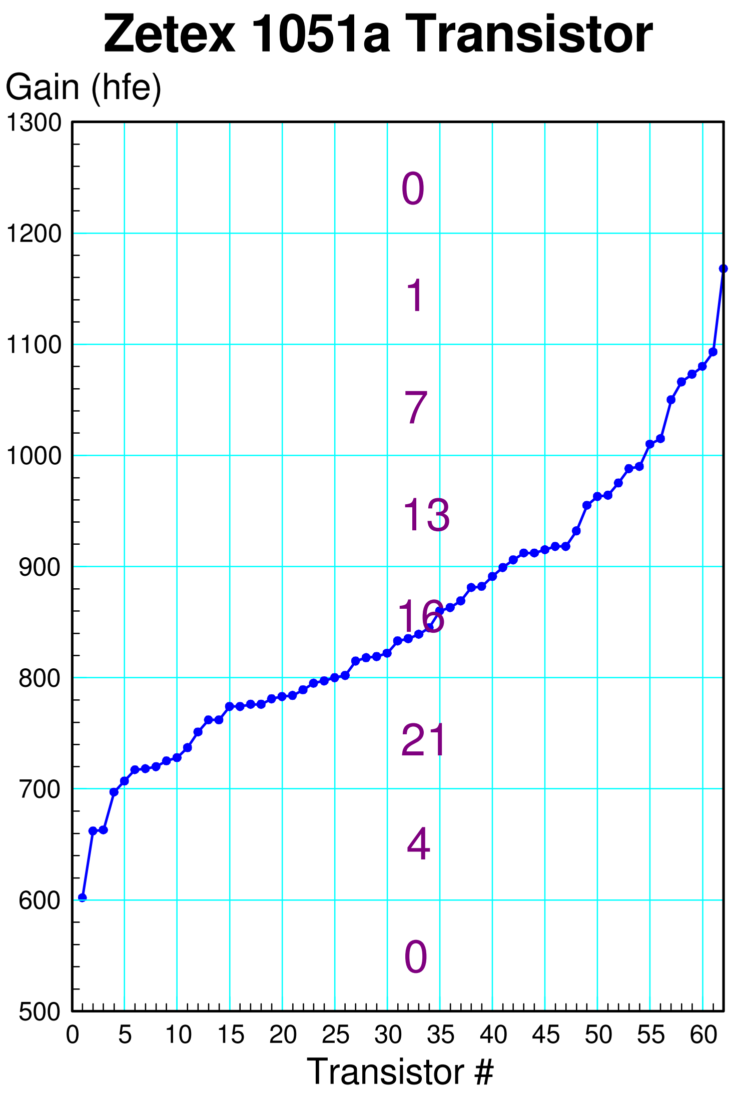
Null Hypothesis:

- The gain of a Zetex 1051a transistor has a uniform distribution over the range of (600, 1200)

Split this into N regions

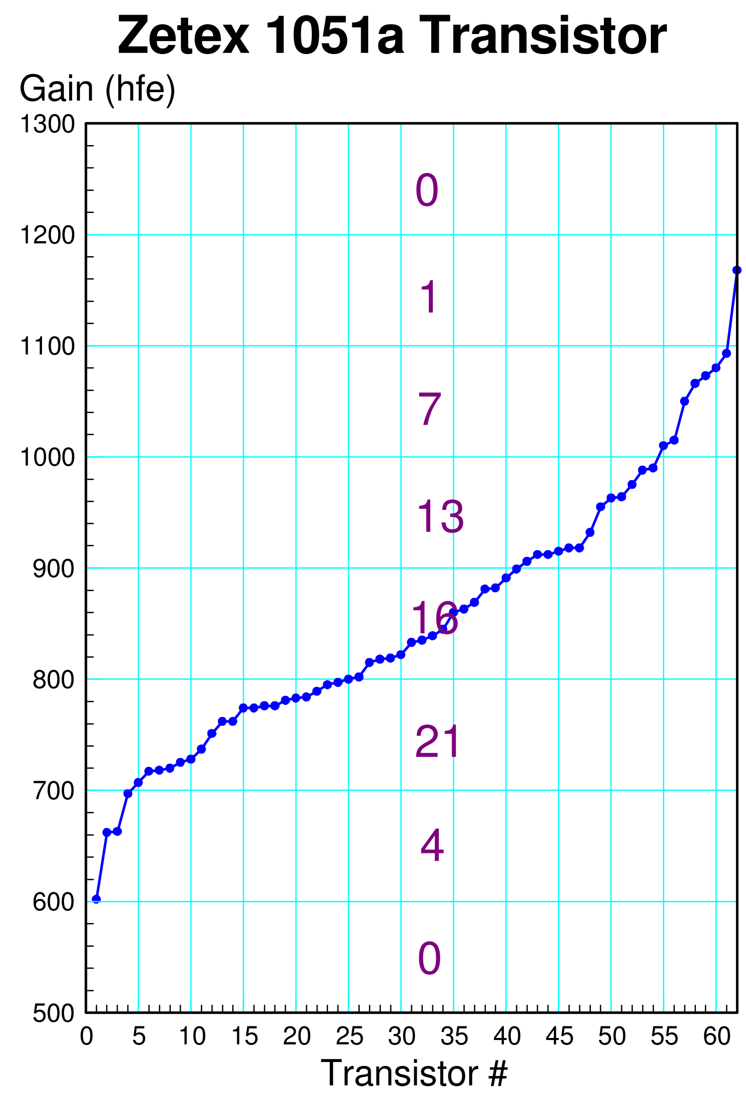
- (0, 600)
- (600, 700)
- :
- (1100, 1200),
- (1200, infinity)

Count the number of occurrences in each bin



Chi-Squred Test

gain	np	Actual	Chi-Squared
>1200	0	0	0
1100 - 1199	10.33	1	8.4268
1000 - 1099	10.33	7	1.0735
900 - 999	10.33	13	0.6901
800 - 899	10.33	16	3.1122
700 - 799	10.33	21	11.0212
600 - 699	10.33	4	3.8789
0 - 599	0	0	0
Total			28.2027



Interpreting the Results

Convert the chi-squared score to a probability

- Chi-squared table
- StatTrek

With 7 degrees of freedom (8 bins), a chi-squared score of 28.2 corresponds to a probability of at least 0.9998

- Enter a value for degrees of freedom.
- Enter a value for one, and only one, of the remaining unshaded text boxes.
- Click the **Calculate** button to compute values for the other text boxes.

Degrees of freedom	7
Chi-square critical value (CV)	28.2
$P(X^2 < 28.2)$	0.9998
$P(X^2 > 28.2)$	0.0002

I'm 99.98% certain that the gain of a Zetex 1051a transistor does not have a uniform distribution

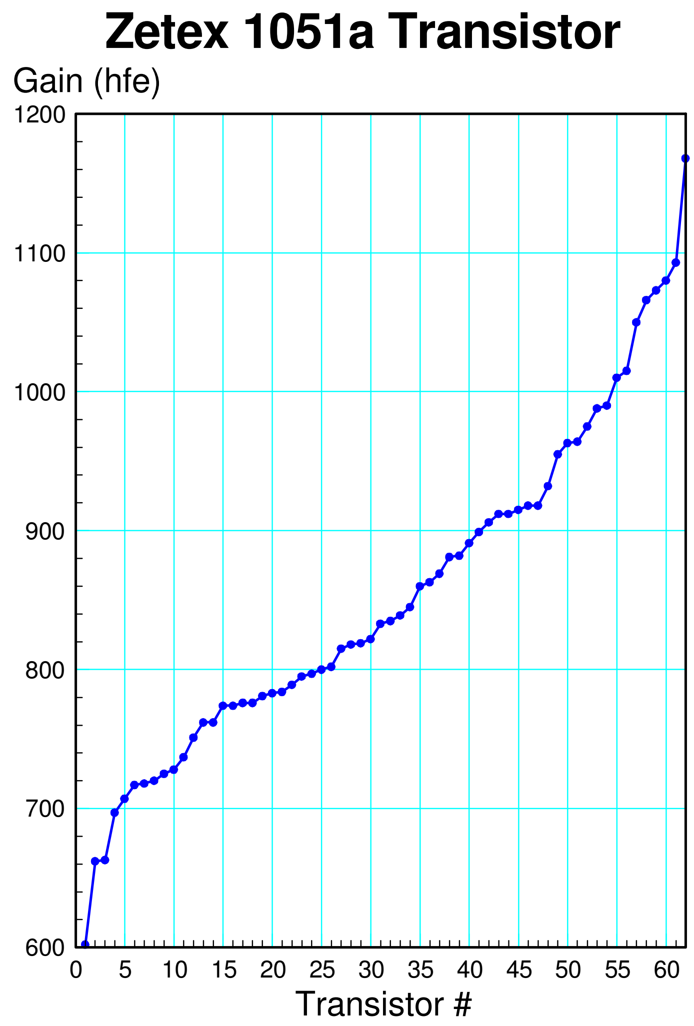
- The data is inconsistent with a uniform distribution

Does the gain of a transistor have a Normal distribution?

- mean = 854.1290
- standard deviation = 120.2034

Same procedure as before but the probabilities change

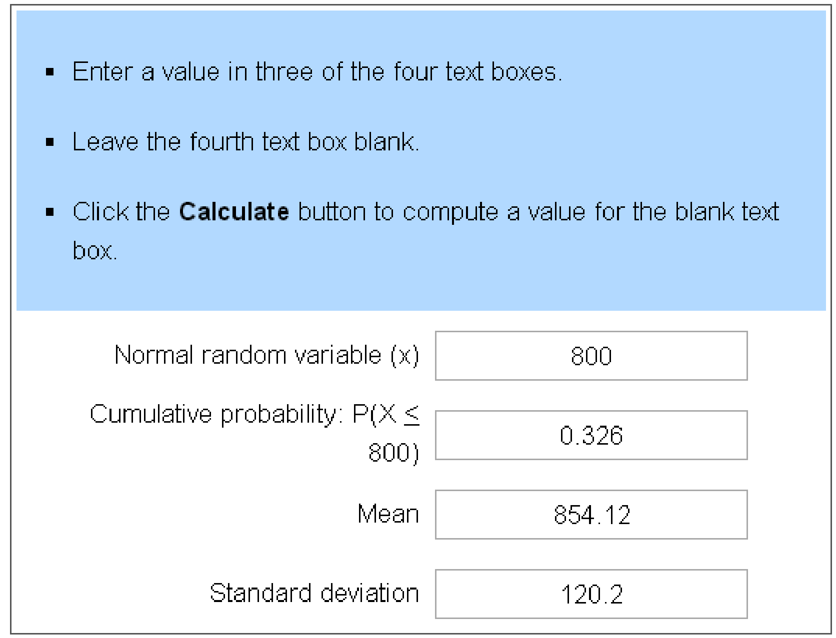
- Use a normal distribution and a z-score to determine the probability of each region



Probabilities of Each Region

- Use StatTrek to find the cdf
- From that, find the probability of each region

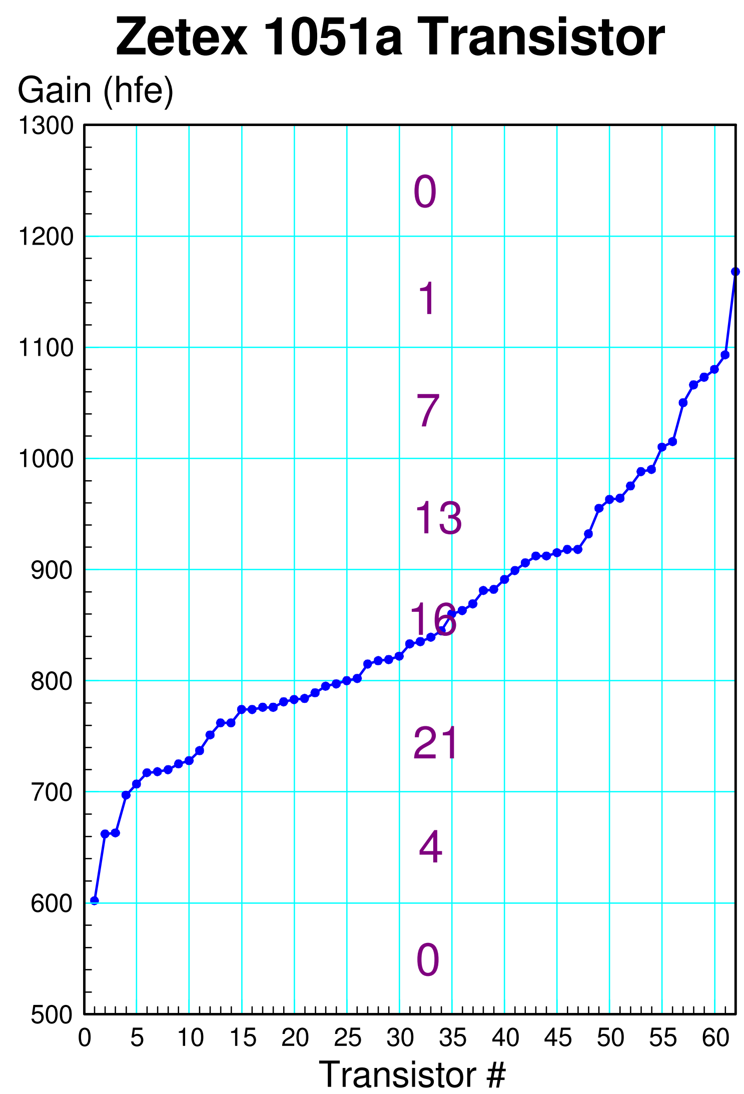
region	cdf	p(region)
1,200	0.998	0.018
1,100	0.98	0.092
1,000	0.888	0.239
900	0.649	0.323
800	0.326	0.226
700	0.1	0.083
600	0.017	0.017



Chi-Squred Calculations

Use the probabilities from the previous slide

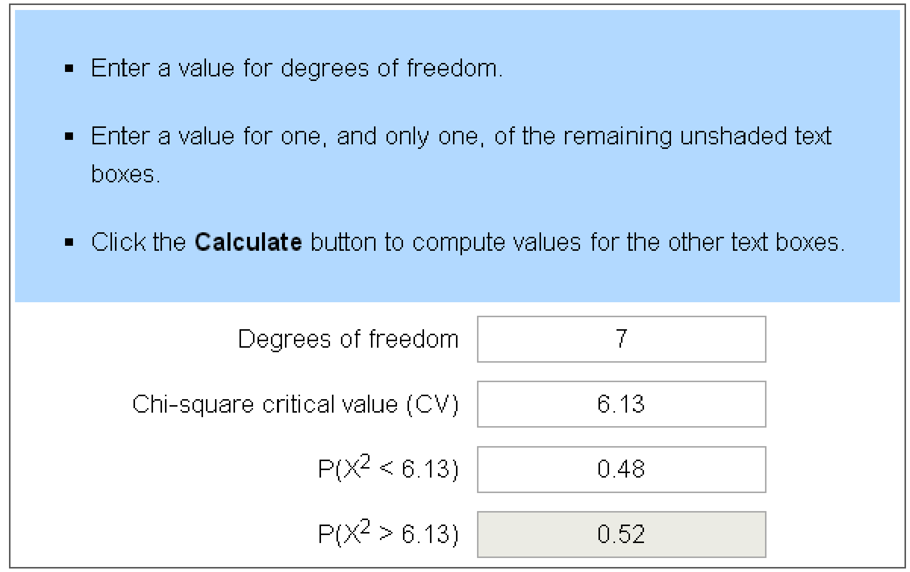
gain	p	np	Actual	Chi-Squared
>1200	0.002	0.124	0	0
1100 - 1199	0.018	1.116	1	0.0121
1000 - 1099	0.092	5.704	7	0.2945
900 - 999	0.239	14.818	13	0.223
800 - 899	0.323	20.026	16	0.8094
700 - 799	0.226	14.012	21	3.485
600 - 699	0.083	5.146	4	0.2552
0 - 599	0.017	1.054	0	1.054
Total				**6.1332**



Interpreting the Results

A chi-squared score of 6.13 corresponds to a probability of 0.48

- There is a 48% chance of rejecting the null hypothesis (this is a normal distribution)

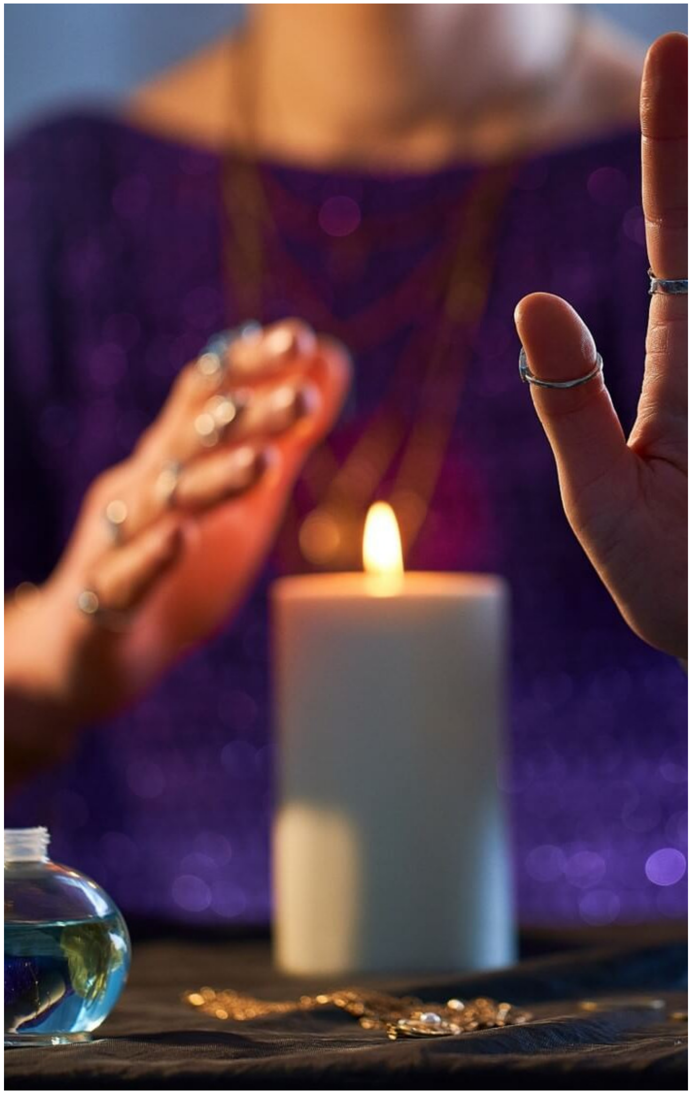


Midrange numbers like this mean "no conclusion"

- The data is consistent with a normal distribution
  - *the chi-squred score is not too large*
- It does not appear that the data was fudged
  - *The chi-squared score is not too small*

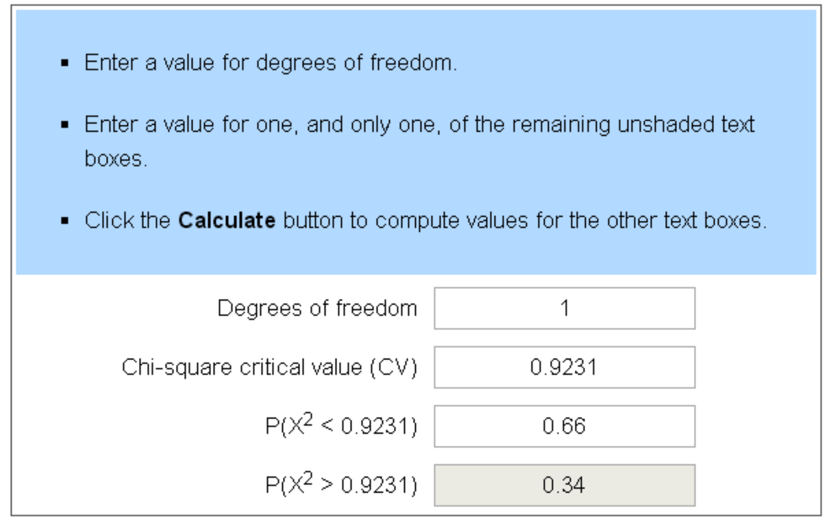
Am I Psychic?

- Take a deck of playing cards
- Shuffle them
- Predict the suit for the top card
- Flip it up and place in one pile if correct, another pile if incorrect
- Count how many times I'm right
- Use a chi-squared test to see if I'm able to foresee the suit with odds that pure chance cannot explain



Data

- Predicted Correctly: 10 times
- Predicted Incorrectly: 42 times



Chi-Squred Test

case	np	Actual	Chi-Squared
Correct	13	10	0.6923
Incorrect	39	42	0.2308
Total			0.9231

Result:

- probability = 66%
- There is a 66% chance of rejecting the null hypothesis
  - *66% chance I'm not just guessing randomly*
  - *66% chance I'm worse than the monkey score*

Summary:

A chi-squared test is a test of a distribution

- Is your data consistent with the assumed distribution.

With it, you can

- Determine if global temperatures are random
- If Fargo is getting warmer,
- If the gain of a transistor has a uniform or normal distribution, and
- If you're psychic