Chi-Squared Examples

ECE 341: Random Processes

Lecture #30

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?

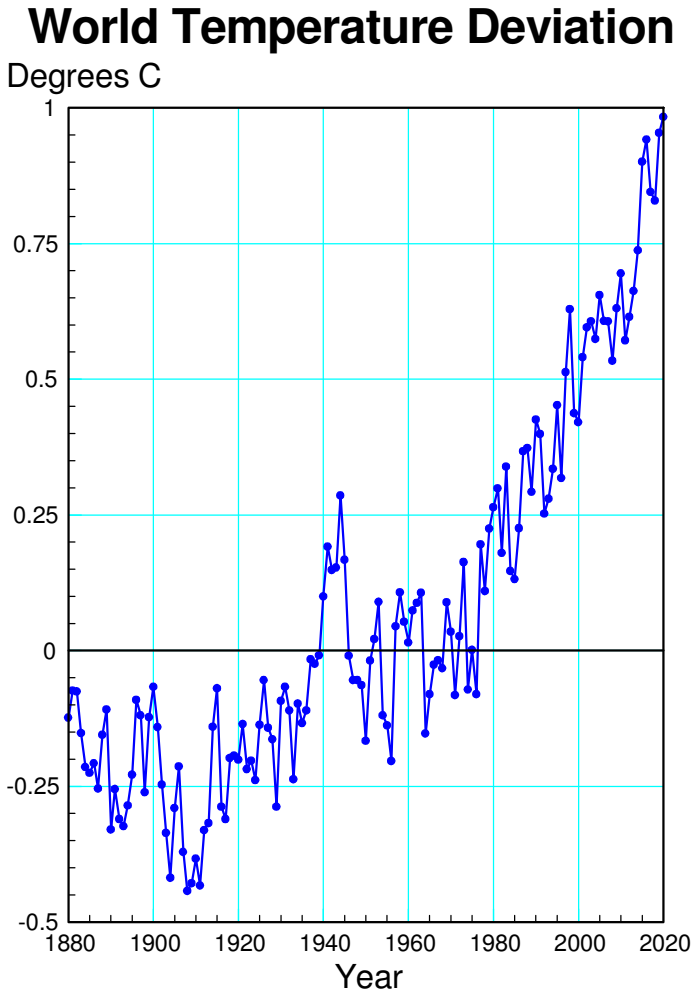
Data:

- The National Oceanic and Atmospheric Administration has been monitoring world temperatures since 1880 (141 years of data)

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

Null Hypothesis:

- Temperatures are not changing



Procedure:

Split the years into 4 bins

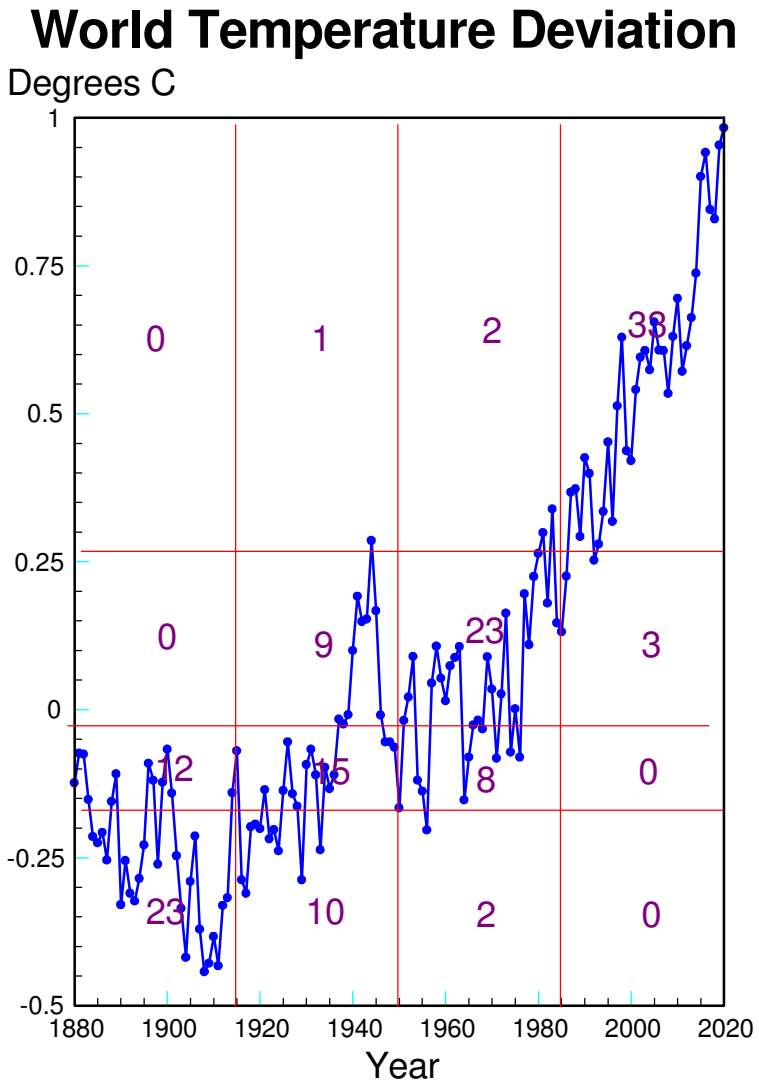
- Every 35 years

Split the temperatures into 4 bins

- Hottest 25% to coldest 25%

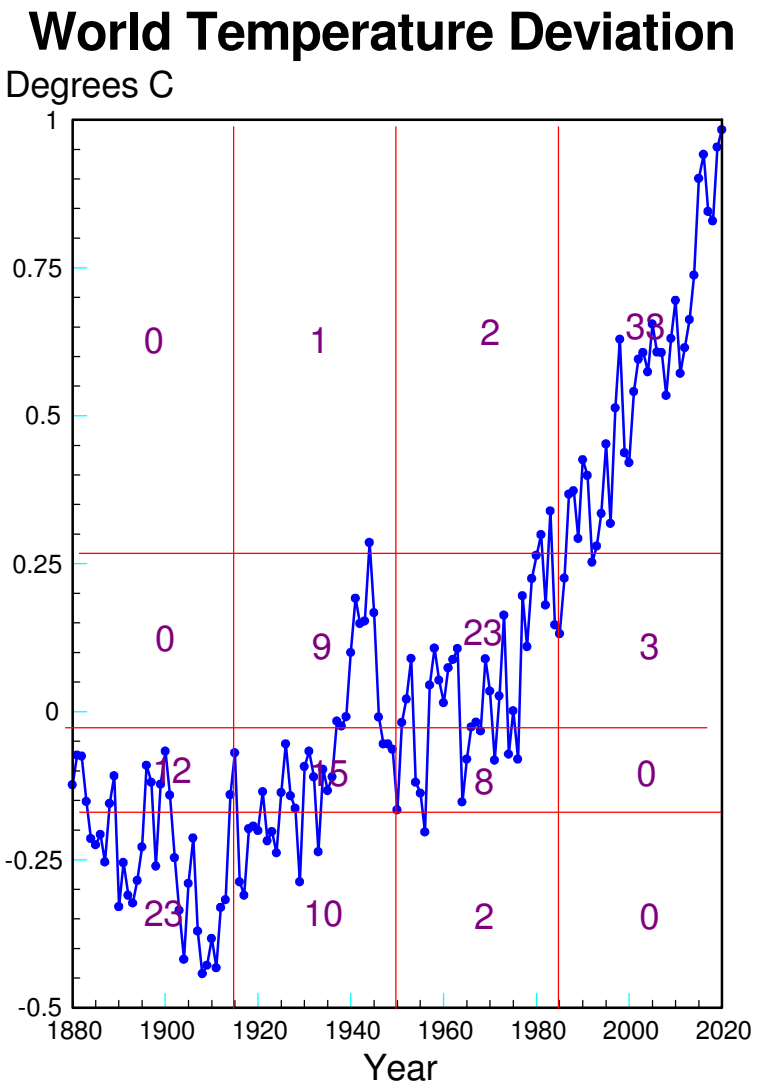
Null Hypothesis:

- This is a uniform distribution (equal probability)
- Data should be evenly distributed among each of the 16 regions



Chi-Squared Test

Years	Tier	np	Actual	Chi-Squared
1880 - 1915	1	8.81	0	8.81
	2	8.81	0	8.81
	3	8.81	12	1.1551
	4	8.81	23	22.8554
1916 - 1950	1	8.81	1	6.9235
	2	8.81	9	0.0041
	3	8.81	15	4.3492
	4	8.81	10	0.1607
1951 - 1985	1	8.81	2	5.264
	2	8.81	23	22.8554
	3	8.81	8	0.0745
	4	8.81	2	5.264
1986 - 2020	1	8.81	33	66.4195
	2	8.81	3	3.8316
	3	8.81	0	8.81
	4	8.81	0	8.81
Total				**174.397**



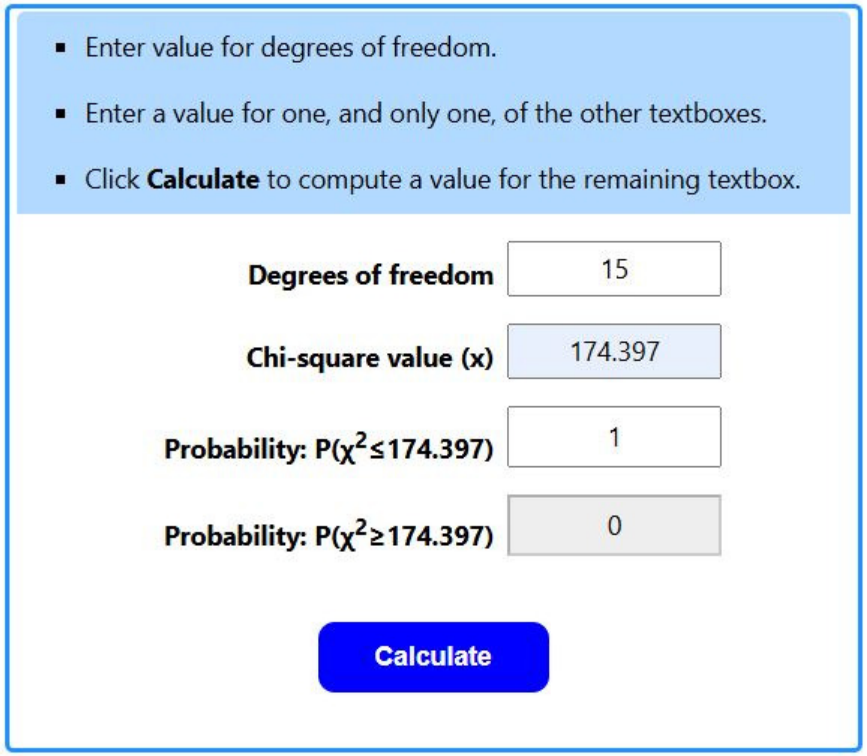
Interpreting the Result

Convert the chi-squared score to a probability

- Chi-squared table
- StatTrek also works

$p > 0.99995$

- 15 degrees of freedom
- rounded to 1 on StatTrek

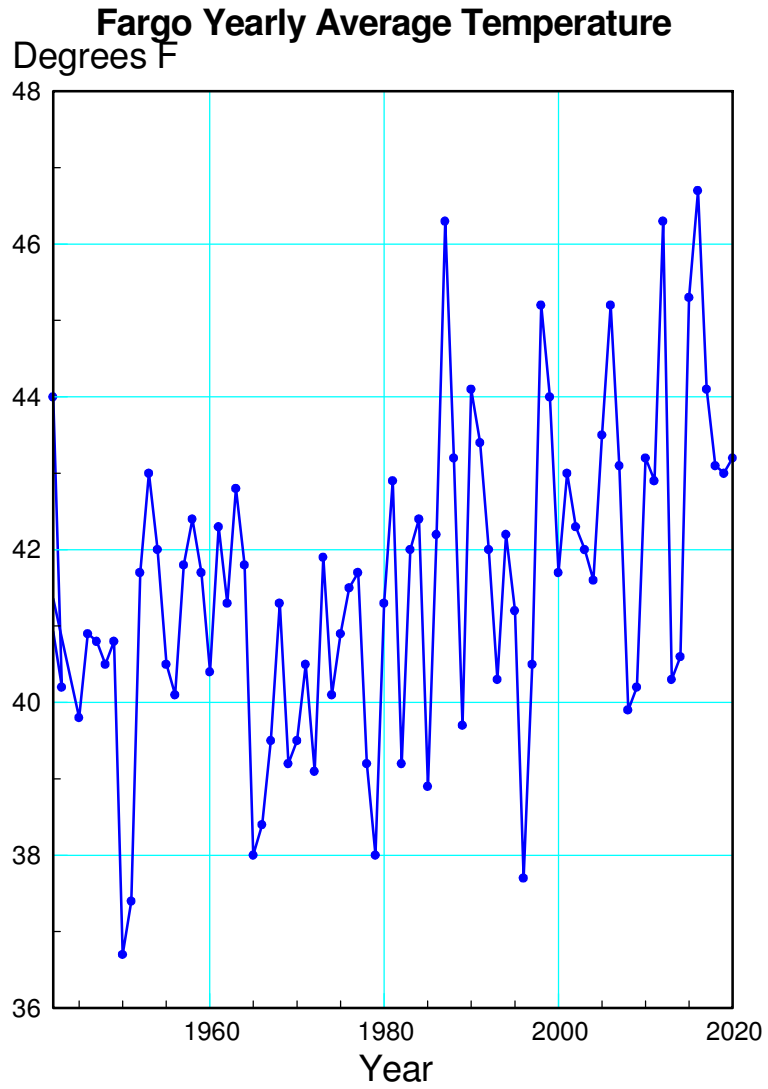


I'm >99.995% certain that the global temperatures are not uniformly distributed year by year (i.e. the temperatures are changing).

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



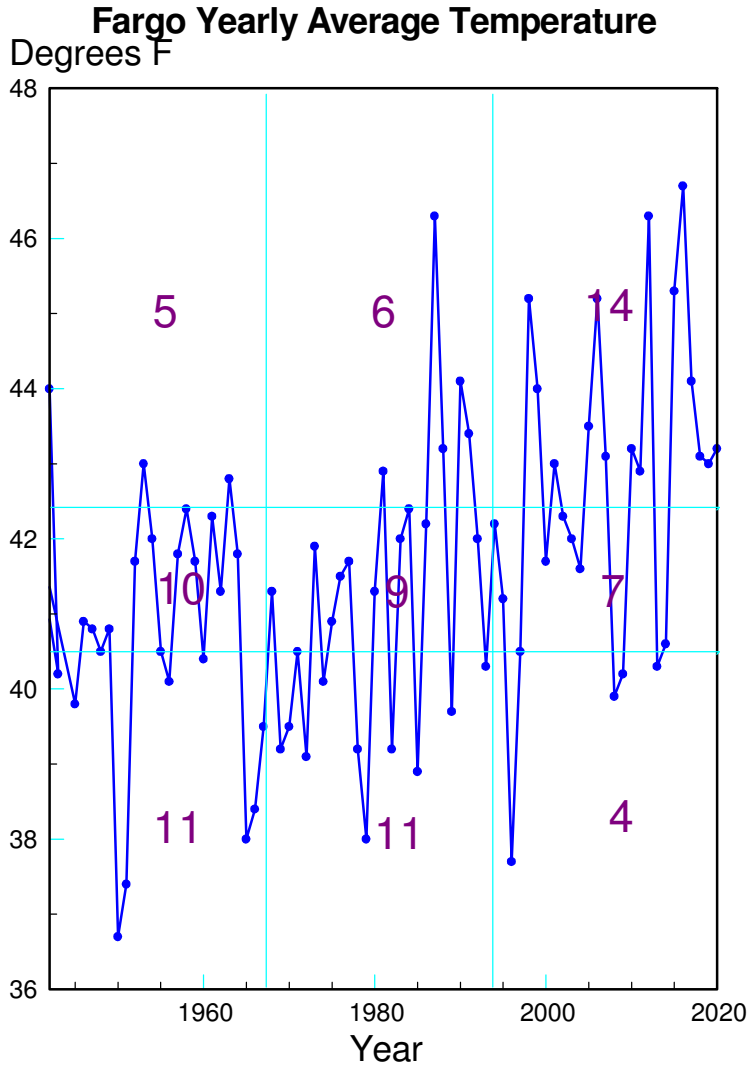
Procedure

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*

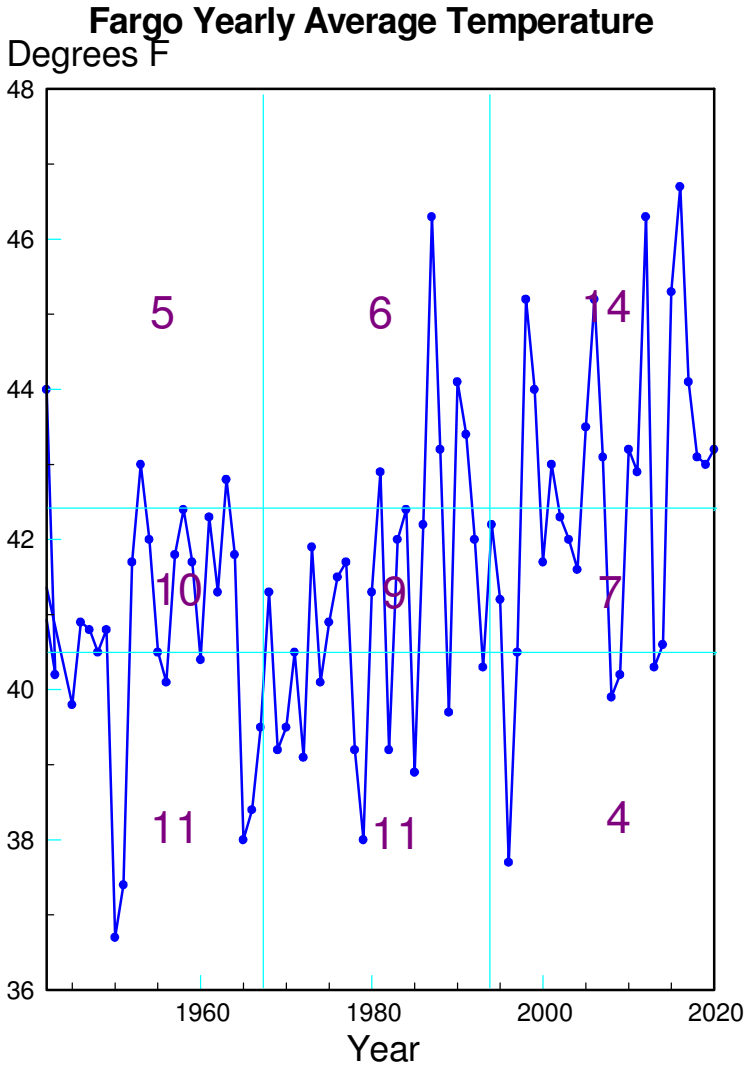
Count how many times a given year falls into each bin

- Expect 1/9th of the data to fall within each region



Chi-Squared Test

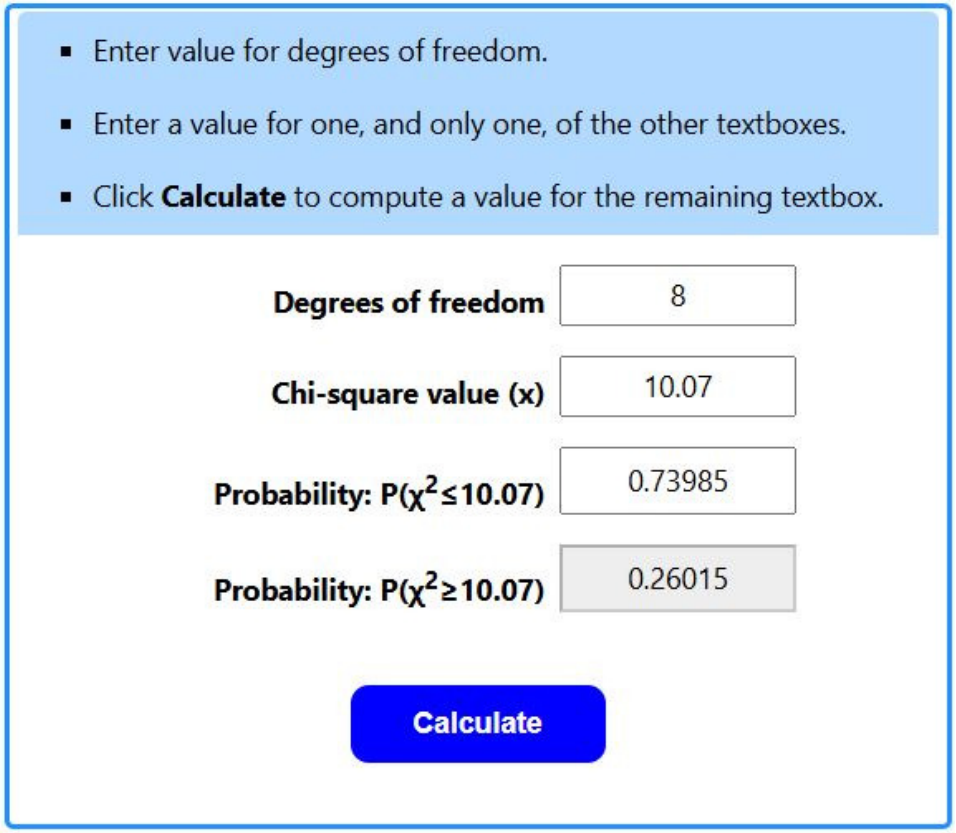
Years	Tier	np	Actual	Chi-Squared
1942 - 1967	hot	8.56	5	1.4806
	middle	8.56	10	0.2422
	cold	8.56	11	0.6955
1968 - 1993	hot	8.56	6	0.7656
	middle	8.56	9	0.0226
	cold	8.56	11	0.6955
1994 - 2020	hot	8.56	14	3.4572
	middle	8.56	7	0.2843
	cold	8.56	4	2.4292
Total				**10.0727**



Interpreting the Result

Convert the chi-squared score to a probability

- Chi-squared table
- StatTrek
- p = 0.73985



I'm 73.985% certain that the temperature in Fargo is changing

- Chi-squared doesn't tell you why
- Chi-squared doesn't tell you direction
- It just says it's 73.985% probable that it's *not* a uniform distribution.

Does the gain of a transistor have a uniform distribution?

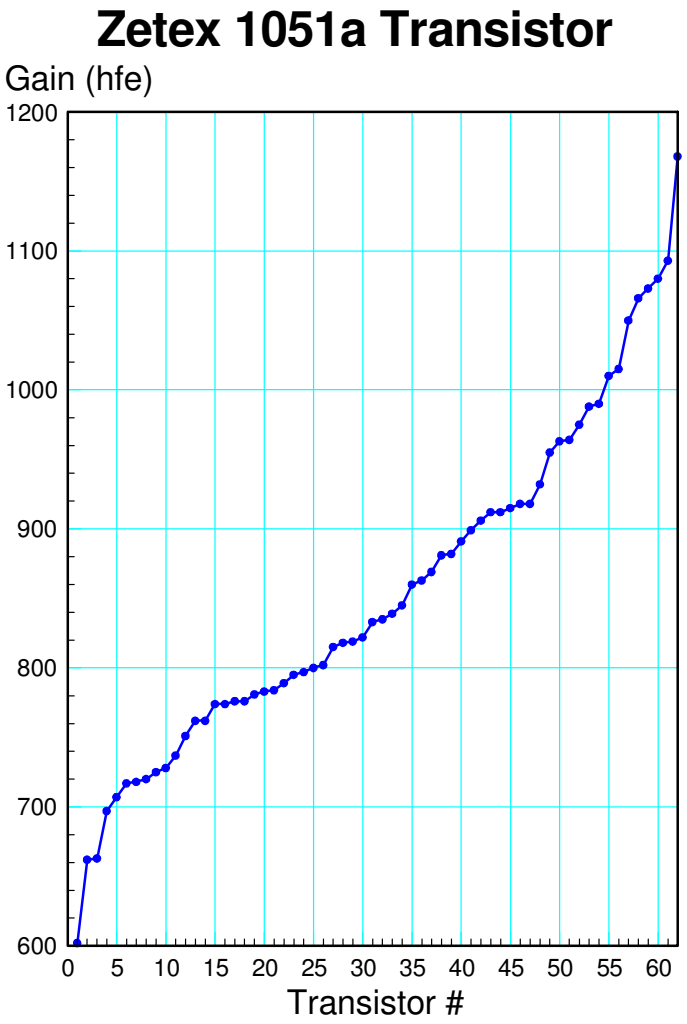
Each transistor's gain is slightly different.

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

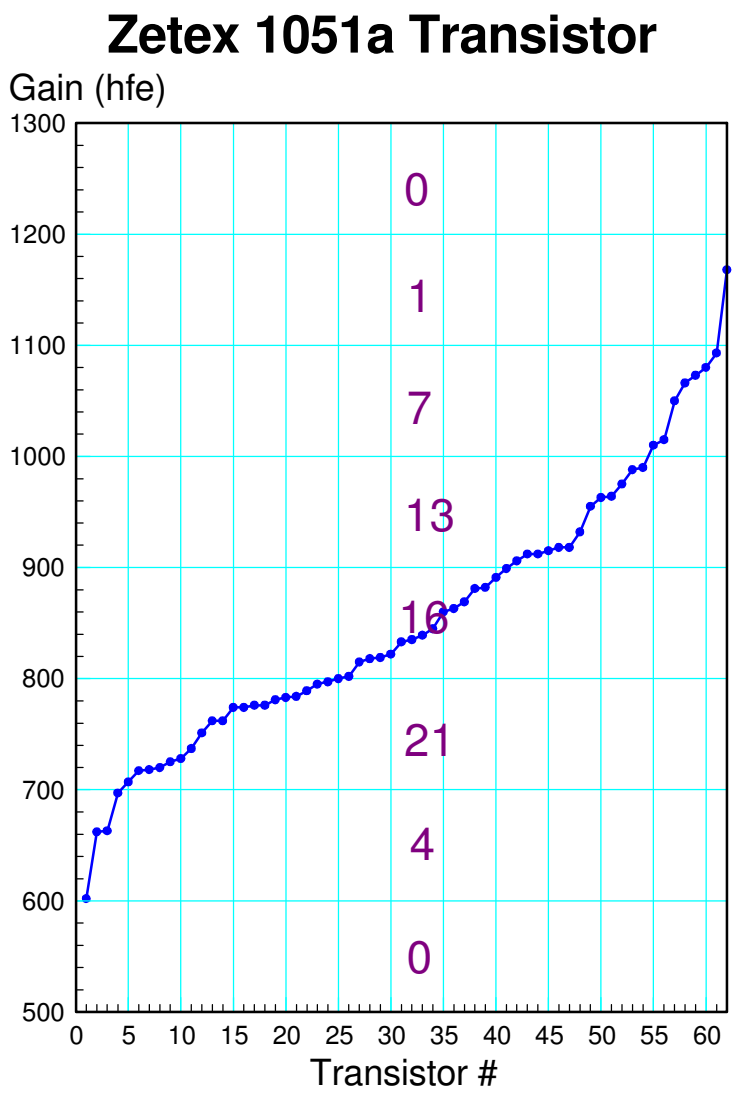
Null Hypothesis:

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

Split this into M regions

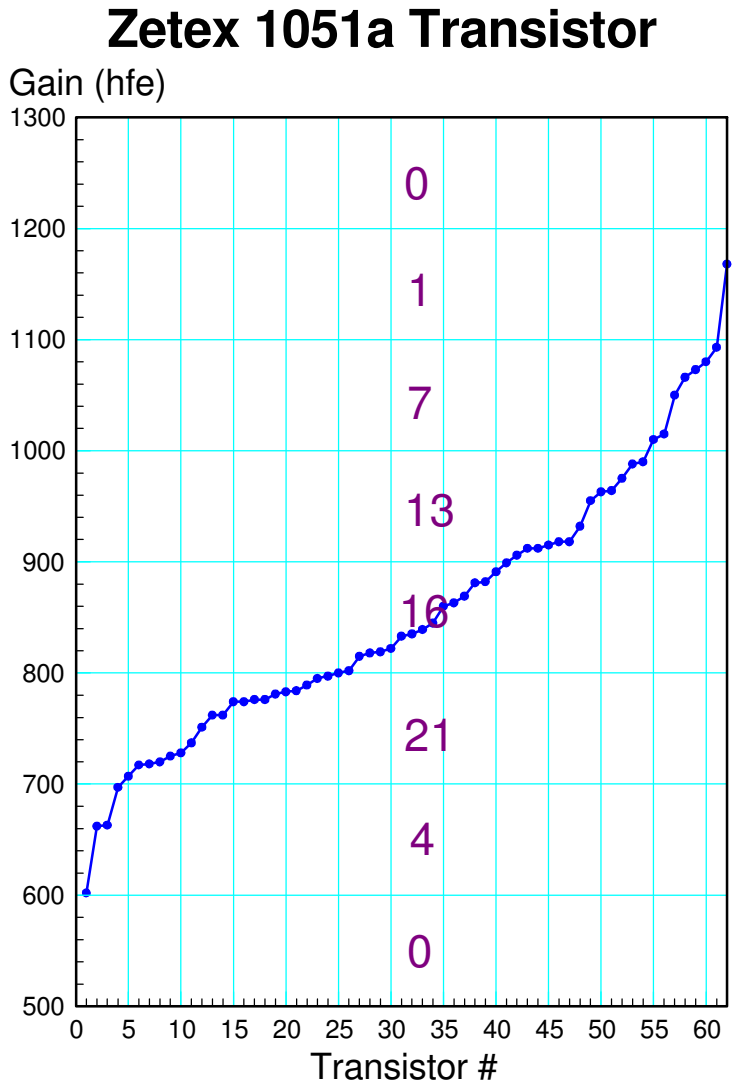
- (0, 600)
- (600, 700)
- :
- (1100, 1200),
- (1200, infinity)

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



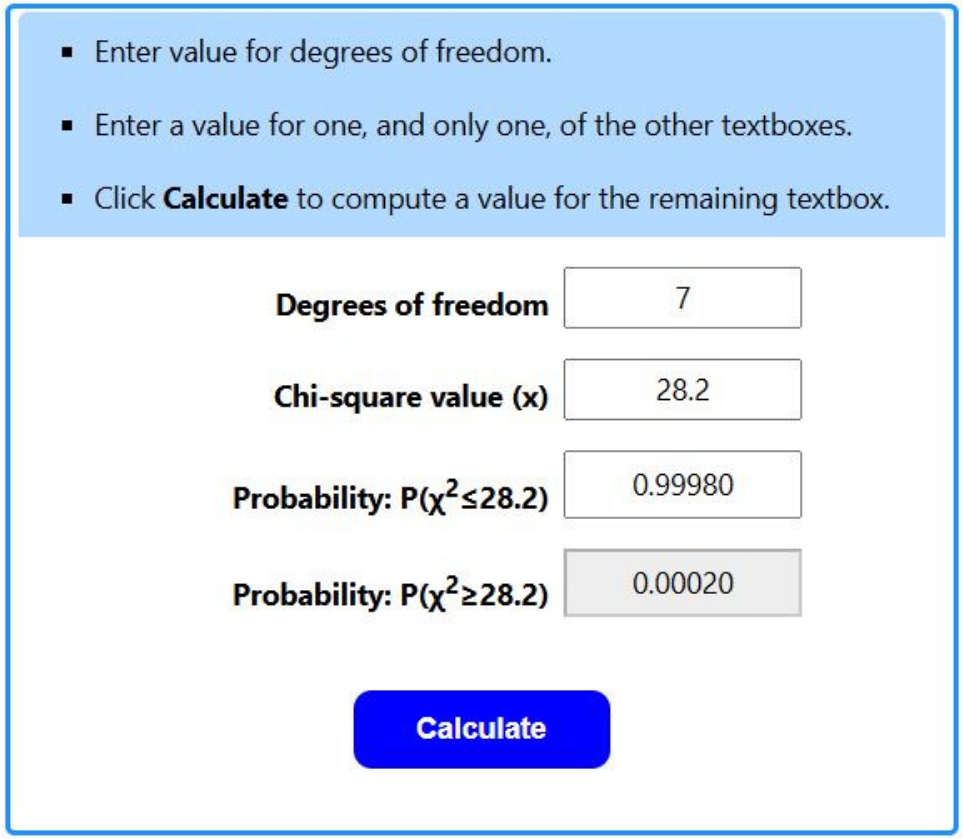
Interpreting the Results

Convert the chi-squared score to a probability

- Chi-squared table
- StatTrek
- p = 0.99980

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

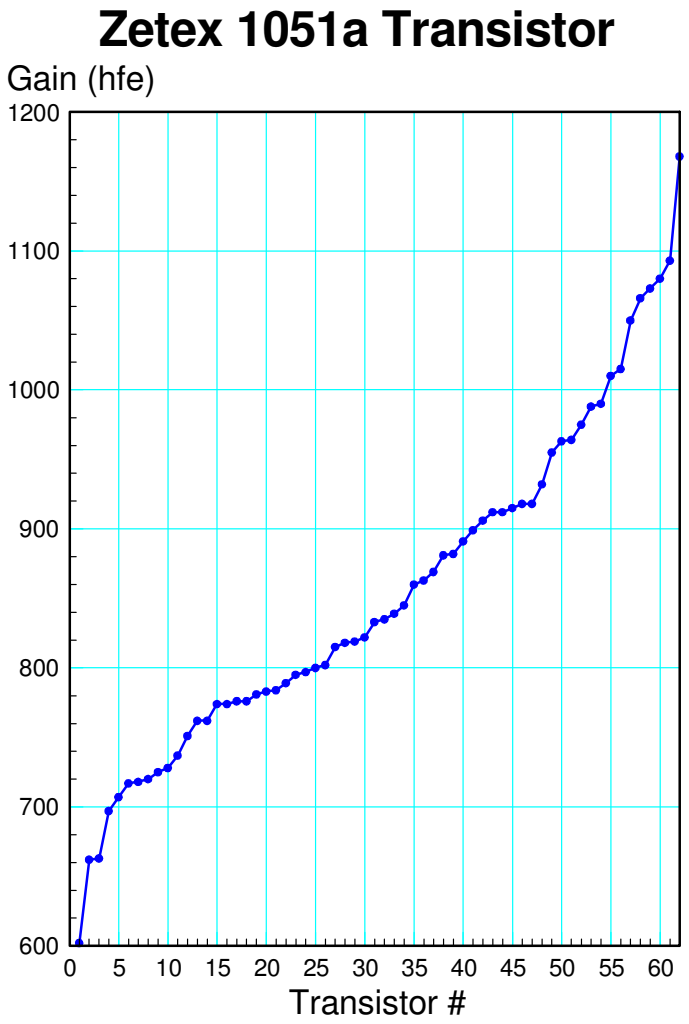


Does the gain of a transistor have a Normal distribution?

- mean = 854.1290
- standard deviation = 120.2034

Same procedure as before but the probabilities change

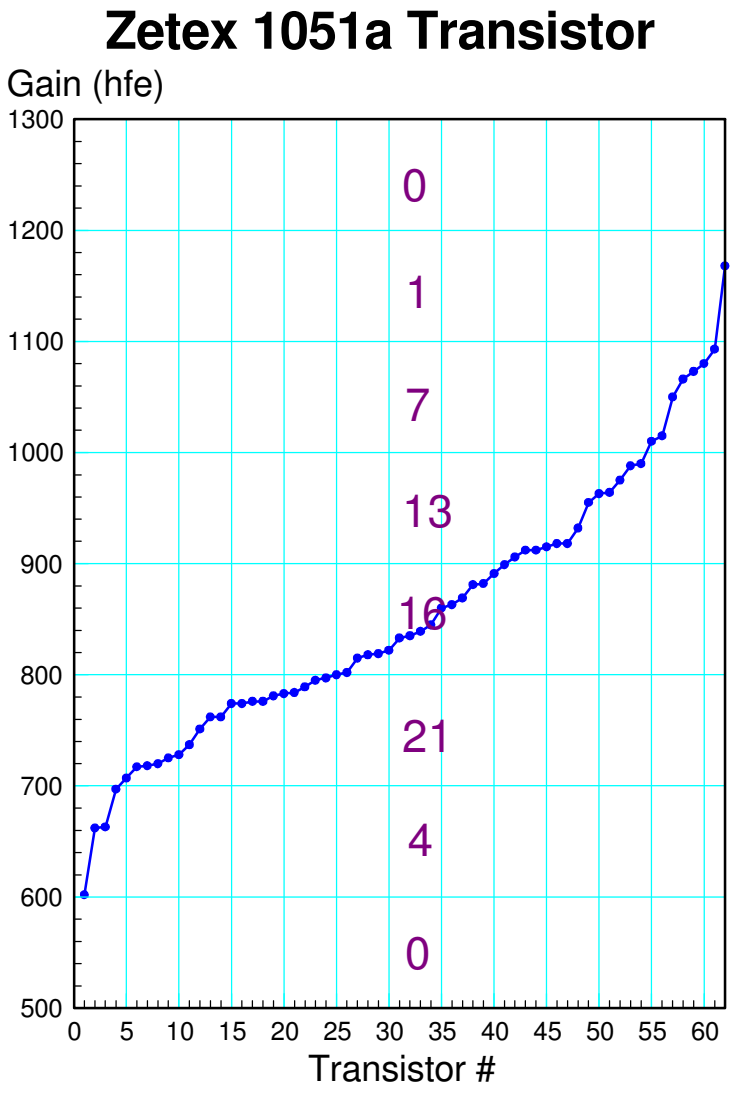
- Use a normal distribution and a z-score to determine the probability of each region



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

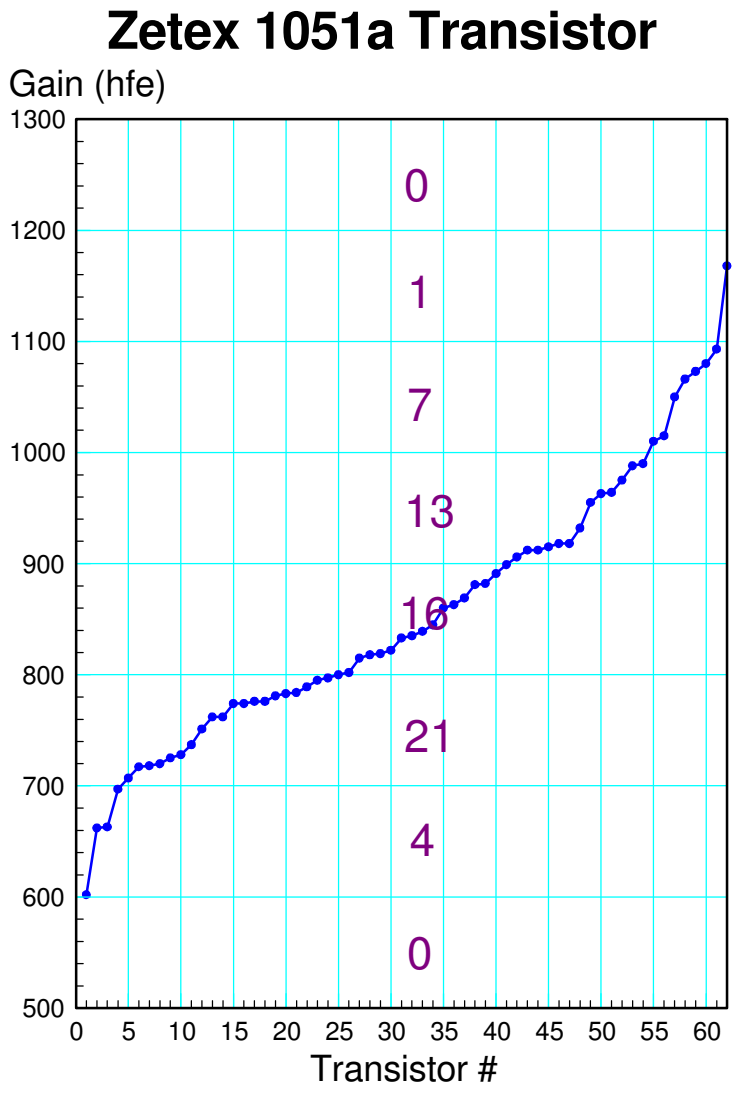


Chi-Squred Calculations

Use the probabilities from the previous slide

- Compute the chi-squared score

gain	p	np	N	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**



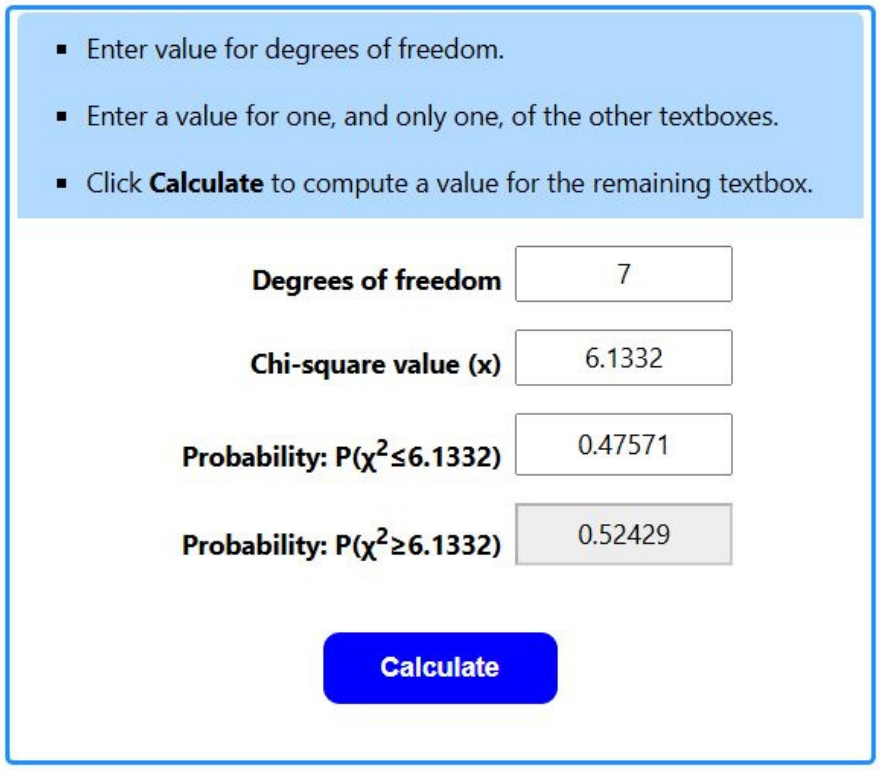
Interpreting the Results

A chi-squared score of 6.13 corresponds to a probability of 0.47571

- There is a 47.571% chance of rejecting the null hypothesis (this is a normal distribution)

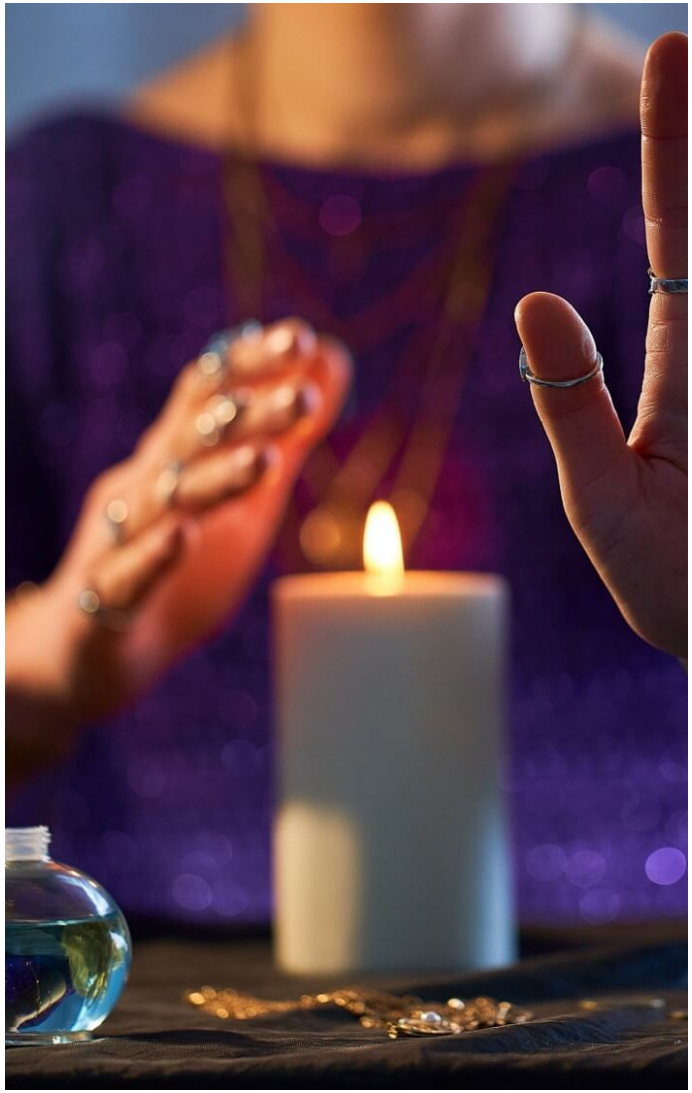
Midrange numbers like this mean "no conclusion"

- 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

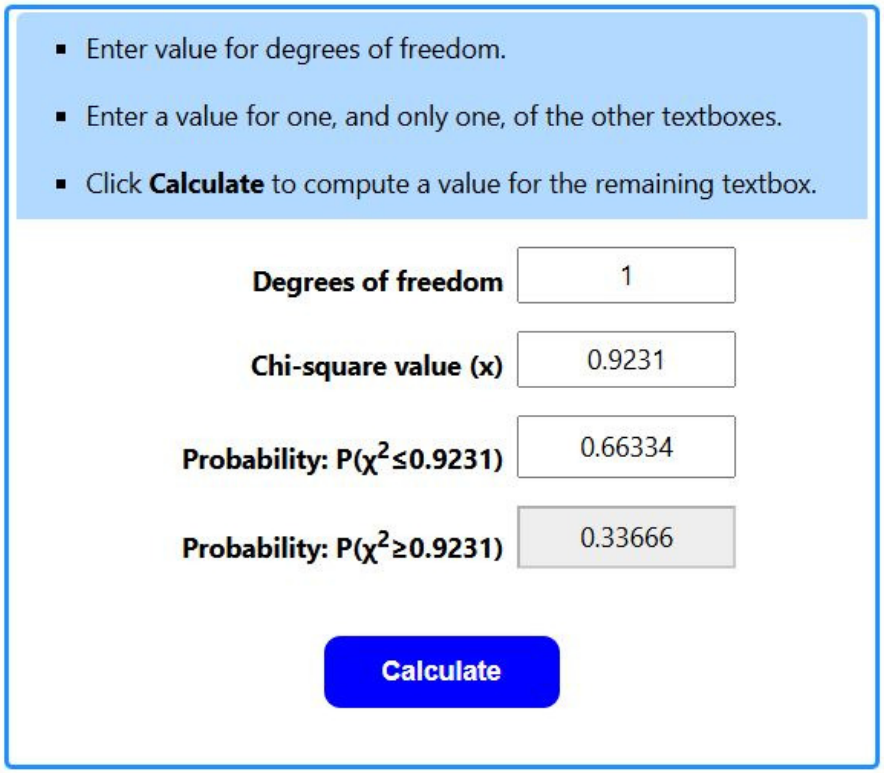


Data

- 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-squred test is a test of a distribution

- Assume a distribution
- Split into M bins
- Collect data
- Compare the expected frequency to the observed frequency

With a chi-squared test, you can tell

- Global temperatures are changing
- Transistor gains do not have a uniform distribution,
- Transistor gains do appear to have a normal distribution, and
- Sadly, I'm not psychic