

## Basic Statistics Definitions:

**Statistics** – Practice or science of collecting and analyzing numerical data

**Data** – Values collected by direct or indirect observation

**Population** – Complete set of all observations in existence

**Sample** – Slice of population meant to represent, as accurately as possible, that population

**Measure** – Measurement of population/sample, an example would be some “score” (a.k.a. an observation)

**Hypothesis** – Educated guess about what’s going on

**Skew** – Not symmetrical, crooked or uneven

**Impute** – To fill in missing values

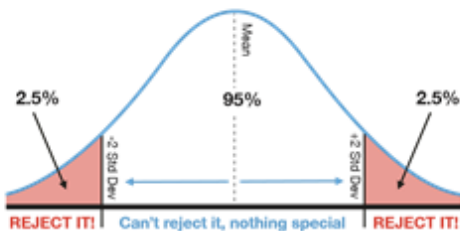
**Type I Error (false positive)** – In hypothesis testing, when you incorrectly reject Null Hypothesis

**Type II Error (false negative)** – In hypothesis testing, when you incorrectly fail to reject Null Hypothesis

## Is My Data Special?

### Null Hypothesis in Layman's Terms:

There is nothing different, or special, about this data



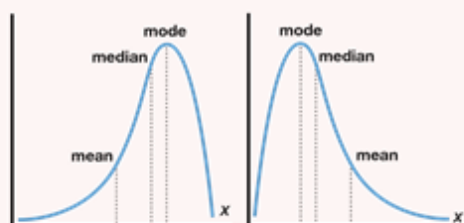
- Best used when you need to know if your data is different or somehow special
- Always start out assuming Null Hypothesis is TRUE
- Goal is to either “reject” or “fail to reject” Null Hypothesis
- If **FAIL TO REJECT** Null Hypothesis then there is nothing really different about the data
- If **REJECT** Null Hypothesis then we are confident that what we see is different or special
- On curve above, can only say that an observation is different/special if it falls in either of shaded regions (called “tails”)
- The tails are 2 Standard Deviations away from (either above or below) the Mean
- Assumes dealing with a normal distribution!

See ⚠ Hazards!

**Big takeaway:** If your data falls within +/- 2 Standard Deviations of Mean then its probably not all that different. If your data falls outside those boundaries then it is most likely something to take note of.

### ⚠ Caution Hazard ⚠

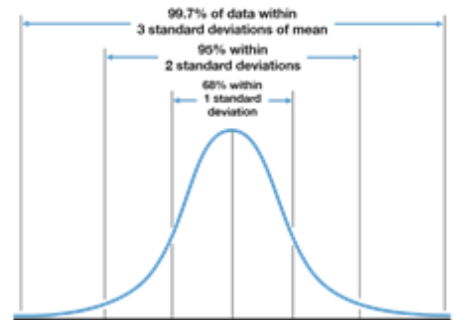
## Skewed Distributions...



Not all data is normally distributed... and when your data is not normally distributed, all those helpful characteristics of a normal distribution no longer apply! For instance Hypothesis testing limits will change, Mean & Median will shift, and most statistical models (think regression) rely heavily on assumption that your data is normally distributed!

## A Normal Distribution:

- A.K.A. “Bell Curve”
- Way to visualize how volume of a population is distributed based on some measurement
- Largest volume is packed around middle
- Volume curves down towards zero to left and right
- Symmetrical around middle
- Interesting Fact: The Mean, Median, and Mode are all the same and at the exact center



## How We Describe Things... (Measures of Central Tendency)

**Mean** – Also called “Average”, probably the most popular statistic, calculated as sum of all values divided by number of values

**Median** – Value at center

**Mode** – Value that occurs most

**Standard Deviation** – Measurement relative to mean, so a measure of how far a value is away from the mean. The further a value is from the mean the more unique... and perhaps interesting... it becomes.

Make sure to review ⚠ Hazards! section regarding skewed distributions

## Sampling

### Good Sampling Rule of Thumb:

- Consider sampling when population you’re working with is too big to handle
- Aim is to get a good representative for actual population
- Generally the bigger the sample the better, but a simple tip is:
  - At minimum your sample size should be 100
  - At maximum your sample size should be 10% or 1000, whichever is smaller
- Keep bias out of it by ensuring a **RANDOM** sample!

### Random Numbers

Are an excellent way to create a Simple Random Sample. Most analytical tools (including Excel & Google Sheets) have a random number generator you can use. Just apply a random number to each row, sort in ascending order by the random number then select the top however-many rows.

### ⚠ Caution Hazard ⚠

### Beware of... **BIAS**

**Bias** can effect both how samples are selected, and also what conclusions you draw from them (i.e. interpretation).

**Selection Bias** – when an individual or observation is more likely to be picked for sampling (in other words, NOT random)

**Observer Bias** – when you subconsciously let your preconceptions influence how you perform your analysis

**Detection Bias** – when something is more likely to be detected in a specific set of observations (e.g. measuring website traffic on Black Friday)

**Funding Bias** – when selection or interpretation favors a financial sponsor

**Extrapolation Bias** – when you assume results of a study describe a larger population than what you originally started with (e.g. assuming a study of college students is a good proxy for entire country)

**Reporting Bias** – when availability of data favors a certain subgroup within true population

**Confirmation Bias** – tend to listen only to information that confirms hypothesis, assumption, or opinion

## Imputing Missing Values...

Missing values are a part of real-life data analysis. But, resist temptation to just fill them in with Mean or Median.

Sometimes this is an OK option, but remember that missing values can be trying to send you a message about some process that you are unaware of (i.e. telling a story).

Also, there are a number of imputation methods out there, be sure to review them thoroughly to see if there are any that better fit your needs/data.

## Confusing Confidence Intervals...

...with probability. 95% confidence just means that 95% of the time the true (population) value will be within the limits.

### Multiple Inference...Faking it ‘till you’re making it

Running a hypothesis test over and over, the same way on the same data, until you get a “significant” result greatly increases chances you will get a false positive (Type I Error) result because... there is always the chance of getting a randomly significant result.

# Statistics Cheat Sheet

## Population

The entire group one desires information about

## Sample

A subset of the population taken because the entire population is usually too large to analyze  
Its characteristics are taken to be representative of the population

## Mean

Also called the arithmetic mean or average

The sum of all the values in the sample divided by the number of values in the sample/population

$\mu$  is the mean of the population;  $\bar{x}$  is the mean of the sample

## Median

The value separating the higher half of a sample/population from the lower half

Found by arranging all the values from lowest to highest and taking the middle one (or the mean of the middle two if there are an even number of values)

## Variance

Measures dispersion around the mean

Determined by averaging the squared differences of all the values from the mean

Variance of a population is  $\sigma^2$

Can be calculated by subtracting the square of the mean from the average of the squared scores:

$$\sigma^2 = \frac{\sum (x - \mu)^2}{n}$$

$$\sigma^2 = \frac{\sum x^2}{n} - \mu^2$$

Variance of a sample is  $s^2$ ; note the  $n-1$

Can be calculated by:

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1}$$

$$s^2 = \frac{\sum x^2 - \frac{(\sum x)^2}{n}}{n - 1}$$

## Standard Deviation

Square root of the variance

Also measures dispersion around the mean but in the same units as the values (instead of square units with variance)

$\sigma$  is the standard deviation of the population and  $s$  is the standard deviation of the sample

## Standard Error

An estimate of the standard deviation of the sampling distribution—the set of all samples of size  $n$  that can be taken from a population

Reflects the extent to which a statistic changes from sample to sample

For a mean,  $\frac{s}{\sqrt{n}}$

For the difference between two means,

Assuming equal variances  $\sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$ ; unequal variances  $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

## T-test

### One-Sample

Tests whether the mean of a normally distributed population is different from a specified value

Null Hypothesis ( $H_0$ ): states that the population mean is equal to some value ( $\mu_0$ )

Alternative Hypothesis ( $H_a$ ): states that the mean does not equal/is greater than/is less than  $\mu_0$

t-statistic: standardizes the difference between  $\bar{x}$  and  $\mu_0$

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} \quad \text{Degrees of freedom (df)} = n-1$$

Read the table of t-distribution critical values for the p-value (probability that the sample mean was obtained by chance given  $\mu_0$  is the population mean) using the calculated t-statistic and degrees of freedom.

$H_a: \mu > \mu_0 \rightarrow$  the t-statistic is likely positive; read table as given

$H_a: \mu < \mu_0 \rightarrow$  the t-statistic is likely negative; the t-distribution is symmetrical so read the probability as if the t-statistic were positive

Note: if the t-statistic is of the 'wrong' sign, the p-value is 1 minus the  $p$  given in the chart

$H_a: \mu \neq \mu_0 \rightarrow$  read the p-value as if the t-statistic were positive and double it (to consider both less than and greater than)

If the p-value is less than the predetermined value for significance (called  $\alpha$  and is usually 0.05), reject the null hypothesis and accept the alternative hypothesis.

### Example:

You are experiencing hair loss and skin discoloration and think it might be because of selenium toxicity. You decide to measure the selenium levels in your tap water once a day for one week. Your results are given below. The EPA maximum contaminant level for safe drinking water is 0.05 mg/L. Does the selenium level in your tap water exceed the legal limit (assume  $\alpha=0.05$ )?

Day	Selenium mg/L
1	0.051
2	0.0505
3	0.049
4	0.0516
5	0.052
6	0.0508
7	0.0506

$H_0: \mu=0.05$ ;  $H_a: \mu>0.05$

Calculate the mean and standard deviation of your sample:

$$\bar{x} = 0.0508$$

$$s^2 = \frac{\sum (x - \bar{x})^2}{n-1} = \frac{(0.051 - 0.0508)^2 + (0.0505 - 0.0508)^2 + \text{etc...}}{6} = 9.15 \times 10^{-7}$$

$$s = \sqrt{s^2} = 9.56 \times 10^{-4}$$

The t-statistic is:  $t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{0.0508 - 0.05}{\frac{9.56 \times 10^{-4}}{\sqrt{7}}} = 2.17$  and the degrees of freedom are  $n-1 = 7-1 = 6$

Looking at the t-distribution of critical values table, 2.17 with 6 degrees of freedom is between  $p=0.05$  and  $p=0.025$ . This means that the p-value is less than 0.05, so you can reject  $H_0$  and conclude that the selenium level in your tap water exceeds the legal limit.

## T-test

### Two-Sample

Tests whether the means of two populations are significantly different from one another

#### Paired

Each value of one group corresponds directly to a value in the other group; ie: before and after values after drug treatment for each individual patient

Subtract the two values for each individual to get one set of values (the differences) and use

$\mu_0 = 0$  to perform a one-sample t-test

#### Unpaired

The two populations are independent

$H_0$ : states that the means of the two populations are equal ( $\mu_1=\mu_2$ )

$H_a$ : states that the means of the two populations are unequal or one is greater than the other ( $\mu_1 \neq \mu_2$ ,  $\mu_1 > \mu_2$ ,  $\mu_1 < \mu_2$ )

t-statistic:

$$\text{assuming equal variances: } t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}} \quad \text{assuming unequal variances: } t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left( \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)}}$$

degrees of freedom =  $(n_1-1) + (n_2-1)$

Read the table of t-distribution critical values for the p-value using the calculated t-statistic and degrees of freedom. Remember to keep the sign of the t-statistic clear (order of subtracting the sample means) and to double the p-value for an  $H_a$  of  $\mu_1 \neq \mu_2$ .

**Example:**

Consider the lifespan of 18 rats. 12 were fed a restricted calorie diet and lived an average of 700 days (standard deviation=21 days). The other 6 had unrestricted access to food and lived an average of 668 days (standard deviation=30 days). Does a restricted calorie diet increase the lifespan of rats (assume  $\alpha=0.05$ )?

$\mu_1=700, s_1=21, n_1=12; \mu_2=668, s_2=30, n_2=6$

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 > \mu_2$  (because we are only asking if a restricted calorie diet increases lifespan)

We cannot assume that the variances of the two populations are equal because the different diets could also affect the variability in lifespan.

The t-statistic is:  $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left( \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)}} = \frac{700 - 668}{\sqrt{\frac{21^2}{12} + \frac{30^2}{6}}} = 2.342$

Degrees of freedom =  $(n_1-1) + (n_2-1) = (12-1) + (6-1) = 16$

From the t-distribution table, the p-value falls between 0.01 and 0.02, so we do reject  $H_0$ . The restricted calorie diet does increase the lifespan of rats.

## Chi-Square Test

### For Goodness of Fit

Checks whether or not an observed pattern of data fits some given distribution

$H_0$ : the observed pattern fits the given distribution

$H_a$ : the observed pattern does not fit the given distribution

The chi-square statistic is:  $\chi^2 = \sum \frac{(O - E)^2}{E}$  ( $O$  is the observed value and  $E$  is the expected value)

Degrees of freedom = number of categories in the distribution - 1

Get the p-value from the table of  $\chi^2$  critical values using the calculated  $\chi^2$  and df values. If the p-value is less than  $\alpha$ , the observed data does not fit the expected distribution. If  $p > \alpha$ , the data likely fits the expected distribution

**Example 1:**

You breed puffskeins and would like to determine the pattern of inheritance for coat color and purring ability.

Puffskeins come in either pink or purple and can either purr or hiss. You breed a purebred, pink purring male with a purebred, purple hissing female. All individuals of the  $F_1$  generation are pink and purring. The  $F_2$  offspring are shown below. Do the alleles for coat color and purring ability assort independently (assume  $\alpha=0.05$ )?

Pink and Purring	Pink and Hissing	Purple and Purring	Purple and Hissing
143	60	55	18

Independent assortment means a phenotypic ratio of 9:3:3:1, so:

$H_0$ : the observed distribution of  $F_2$  offspring fits a 9:3:3:1 distribution

$H_a$ : the observed distribution of  $F_2$  offspring does not fit a 9:3:3:1 distribution

The expected values are:

Pink and Purring	Pink and Hissing	Purple and Purring	Purple and Hissing
155.25	51.75	51.75	17.25

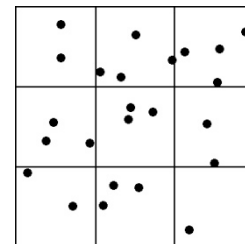
$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(143 - 155.25)^2}{155.25} + \frac{(60 - 51.75)^2}{51.75} + \frac{(55 - 51.75)^2}{51.75} + \frac{(18 - 17.25)^2}{17.25} = 2.519$$

df=4-1=3

From the table of  $\chi^2$  critical values, the p-value is greater than 0.25, so the alleles for coat color and purring ability do assort independently in puffskeins.

### Example 2:

You are studying the pattern of dispersion of king penguins and the diagram on the right represents an area you sampled. Each dot is a penguin. Do the penguins display a uniform distribution (assume  $\alpha=0.05$ )?



$H_0$ : there is a uniform distribution of penguins

$H_a$ : there is not a uniform distribution of penguins

There are a total of 25 penguins, so if there is a uniform distribution, there should be 2.778 penguins per square. There actual observed values are 2, 4, 4, 3, 3, 3, 2, 3, 1, so the  $\chi^2$  statistic is:

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(1 - 2.778)^2}{2.778} + 2 \left( \frac{(2 - 2.778)^2}{2.778} \right) + 4 \left( \frac{(3 - 2.778)^2}{2.778} \right) + 2 \left( \frac{(4 - 2.778)^2}{2.778} \right) = 2.72$$

df=9-1=8

From the table of  $\chi^2$  critical values, the p-value is greater than 0.25, so we do not reject  $H_0$ . The penguins do display a uniform distribution.

## Chi-Square Test

### For Independence

Checks whether two categorical variables are related or not (independence)

$H_0$ : the two variables are independent

$H_a$ : the two variables are not independent

Does not make any assumptions about an expected distribution

The observed values ( $\#_1$ ,  $\#_2$ ,  $\#_3$ , and  $\#_4$ ) are usually presented as a table. Each row is a category of variable 1 and each column is a category of variable 2.

		Variable 1		Totals
		Category x	Category y	
Variable 2	Category a	$\#_1$	$\#_2$	$\#_1 + \#_2$
	Category b	$\#_3$	$\#_4$	$\#_3 + \#_4$
Totals		$\#_1 + \#_3$	$\#_2 + \#_4$	$\#_1 + \#_2 + \#_3 + \#_4$

The proportion of category x of variable 1 is the number of individuals in category x divided by the total number of individuals  $\left( \frac{\#_1 + \#_3}{\#_1 + \#_2 + \#_3 + \#_4} \right)$ . Assuming independence, the expected number of individuals that fall within category

a of variable 2 is the proportion of category x multiplied by the number of individuals in category a

$\left( \frac{\#_1 + \#_3}{\#_1 + \#_2 + \#_3 + \#_4} \right) (\#_1 + \#_2)$ . Thus, the expected value is:

$$E = \frac{(\#_1 + \#_3)(\#_1 + \#_2)}{\#_1 + \#_2 + \#_3 + \#_4} = \frac{(\text{row total})(\text{column total})}{\text{grand total}}$$

Degrees of freedom =  $(r-1)(c-1)$  where r is the number of rows and c is the number of columns

The chi-square statistic is still  $\chi^2 = \sum \frac{(O - E)^2}{E}$

Read the p-values from the table of  $\chi^2$  critical values.

### Example:

Given the data below, is there a relationship between fitness level and smoking habits (assume  $\alpha=0.05$ )?

	Fitness Level				
	Low	Medium-Low	Medium-High	High	
Never smoked	113	113	110	159	495
Former smokers	119	135	172	190	616
1 to 9 cigarettes daily	77	91	86	65	319
$\geq 10$ cigarettes daily	181	152	124	73	530
	490	491	492	487	1960

$H_0$ : fitness level and smoking habits are independent

$H_a$ : fitness level and smoking habits are not independent

First, we calculate the expected counts. For the first cell, the expected count is:

$$E = \frac{(\text{row total})(\text{column total})}{\text{grand total}} = \frac{(495)(490)}{1960} = 123.75$$

	Fitness Level			
	Low	Medium-Low	Medium-High	High
Never smoked	123.75	124	124.26	122.99
Former smokers	154	154.31	154.63	153.06
1 to 9 cigarettes daily	79.75	79.91	80.08	79.26
$\geq 10$ cigarettes daily	132.5	132.77	133.04	131.69

$$\chi^2 = \sum \frac{(O - E)^2}{E} = \frac{(113 - 123.75)^2}{123.75} + \frac{(113 - 124)^2}{124} + \frac{(110 - 124.26)^2}{124.26} + \text{etc...} = 91.73$$

$$df = (r-1)(c-1) = (4-1)(4-1) = 9$$

From the table of  $\chi^2$  critical values, the p-value is less than 0.001, so we reject  $H_0$  and conclude that there is a relationship between fitness level and smoking habits.

### Type I error

The probability of rejecting a true null hypothesis

Equals  $\alpha$

### Type II error

The probability of failing to reject a false null hypothesis

### Probability

#### Joint Probability

The probability of events A and B occurring

$P(A \text{ and } B) = P(A) \times P(B)$  when events A and B are independent

#### Union of Events

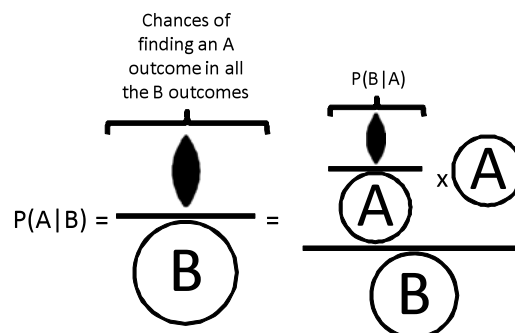
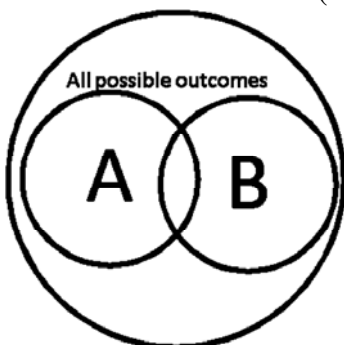
The probability of either event A or event B occurring

$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

#### Conditional Probability

The probability of event A occurring given that event B has occurred

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} \quad \text{or} \quad P(A|B) = \frac{P(B|A) \times P(A)}{P(B)}$$



#### Example 1:

Assume that eye color is an autosomally inherited trait controlled by one gene with two alleles. Brown is dominant to blue. A brown-eyed man with genotype Bb and a blue-eyed woman have three children. The first has blue eyes. What is the probability that all three children have blue eyes?

Without considering the first child, the probability that the couple has three children with blue eyes is  $0.5 \times 0.5 \times 0.5 = 0.125 = P(A \text{ and } B) = P(2 \text{ children} = bb \text{ and } 1\text{st child } bb)$

With his parents, the probability that the 1<sup>st</sup> child is bb is:  $P(B) = P(1\text{st child} = bb) = 0.5$

Therefore,  $P(2 \text{ children} = bb \mid 1\text{st child } bb) = P(A \mid B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{0.125}{0.5} = 0.25$

### **Example 2:**

Based on an analysis of her pedigree, it is determined that a woman has a 70% chance of being Zz and a 30% chance of being ZZ for a sex-linked trait, where Z is dominant to z. If she now has a son with the Z phenotype, what is the probability of her being Zz?

We're looking for:  $P(W=Zz \mid S=Z)$

But it's hard to find  $P(W=Zz \text{ and } S=Z)$  because the two events are not independent. Instead, let us use:

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}$$

$P(S = Z \mid W = Zz) = 0.5$  (50% chance of passing on the Z allele)

$P(W = Zz) = 0.7$  (given)

$P(S = Z) = (0.7 \times 0.5) + (0.3 \times 1) = 0.65$  (son can be Z from the woman being either Zz or ZZ)

$$P(W = Zz \mid S = Z) = \frac{0.5 \times 0.7}{0.65} = 0.538$$

## **Multiple Experiments**

### **Binomial distribution**

For when you are not concerned about the order of the events, only that they occur

$$P(X = m) = \frac{n! \times p^m \times (1 - p)^{(n-m)}}{m! \times (n - m)!}$$

for  $m$  outcomes of event X in  $n$  total trials with  $p$ =probability of X occurring once

### **Example:**

What is the probability that a couple has one boy out of five children?

$$P(1 \text{ boy of } 5 \text{ children}) = \frac{5! \times 0.5^1 \times 0.5^4}{1! \times (4)!} = 0.15625$$

### **Poisson distribution**

The binomial distribution works for a small number of trials but as  $n$  gets too large, the factorials become unwieldy.

The Poisson distribution is an estimate of the binomial distribution for large  $n$ .

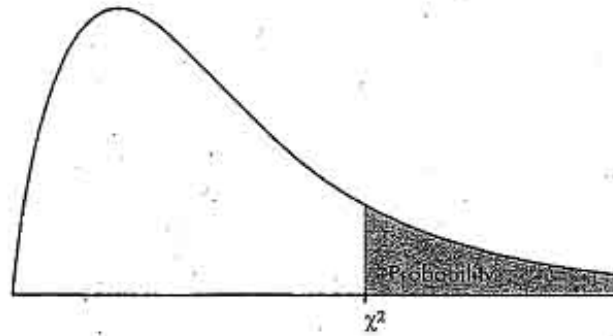
$$P(X = m) = \frac{e^{-np} \times (np)^m}{m!}$$

Note:  $np$  is also known as the number of expected outcomes for event X



	Tail probability $p$											
df	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001	.0005
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.92
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.611	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.768
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646
40	.681	.851	1.050	1.303	1.684	2.021	2.123	2.423	2.704	2.971	3.307	3.551
50	.679	.849	1.047	1.299	1.676	2.009	2.109	2.403	2.678	2.937	3.261	3.496
60	.679	.848	1.045	1.296	1.671	2.000	2.099	2.390	2.660	2.915	3.232	3.460
80	.678	.846	1.043	1.292	1.664	1.990	2.088	2.374	2.639	2.887	3.195	3.416
100	.677	.845	1.042	1.290	1.660	1.984	2.081	2.364	2.626	2.871	3.174	3.390
1000	.675	.842	1.037	1.282	1.646	1.962	2.056	2.330	2.581	2.813	3.098	3.300
$\infty$	.674	.841	1.036	1.282	1.645	1.960	2.054	2.326	2.576	2.807	3.091	3.291
	50%	60%	70%	80%	90%	95%	96%	98%	99%	99.5%	99.8%	99.9%
	Confidence level $C$											



$\chi^2$  CRITICAL VALUESTABLE C:  $\chi^2$  CRITICAL VALUES

df	Tail probability $p$										
	.25	.20	.15	.10	.05	.025	.02	.01	.005	.0025	.001
1	1.32	1.64	2.07	2.71	3.84	5.02	5.41	6.63	7.88	9.14	10.83
2	2.77	3.22	3.79	4.61	5.99	7.38	7.82	9.21	10.60	11.98	13.82
3	4.11	4.64	5.32	6.25	7.81	9.35	9.84	11.34	12.84	14.32	16.27
4	5.39	5.99	6.74	7.78	9.49	11.14	11.67	13.28	14.86	16.42	18.47
5	6.63	7.29	8.12	9.24	11.07	12.83	13.39	15.09	16.75	18.39	20.51
6	7.84	8.56	9.45	10.64	12.59	14.45	15.03	16.81	18.55	20.25	22.46
7	9.04	9.80	10.75	12.02	14.07	16.01	16.62	18.48	20.28	22.04	24.32
8	10.22	11.03	12.03	13.36	15.51	17.53	18.17	20.09	21.95	23.77	26.12
9	11.39	12.24	13.29	14.68	16.92	19.02	19.68	21.67	23.59	25.46	27.88
10	12.55	13.44	14.53	15.99	18.31	20.48	21.16	23.21	25.19	27.11	29.59
11	13.70	14.63	15.77	17.28	19.68	21.92	22.62	24.72	26.76	28.73	31.26
12	14.85	15.81	16.99	18.55	21.03	23.34	24.05	26.22	28.30	30.32	32.91
13	15.98	16.98	18.20	19.81	22.36	24.74	25.47	27.69	29.82	31.88	34.53
14	17.12	18.15	19.41	21.06	23.68	26.12	26.87	29.14	31.32	33.43	36.12
15	18.25	19.31	20.60	22.31	25.00	27.49	28.26	30.58	32.80	34.95	37.70
16	19.37	20.47	21.79	23.54	26.30	28.85	29.63	32.00	34.27	36.46	39.25
17	20.49	21.61	22.98	24.77	27.59	30.19	31.00	33.41	35.72	37.95	40.79
18	21.60	22.76	24.16	25.99	28.87	31.53	32.35	34.81	37.16	39.42	42.31
19	22.72	23.90	25.33	27.20	30.14	32.85	33.69	36.19	38.58	40.88	43.82
20	23.83	25.04	26.50	28.41	31.41	34.17	35.02	37.57	40.00	42.34	45.31
21	24.93	26.17	27.66	29.62	32.67	35.48	36.34	38.93	41.40	43.78	46.80
22	26.04	27.30	28.82	30.81	33.92	36.78	37.66	40.29	42.80	45.20	48.27
23	27.14	28.43	29.98	32.01	35.17	38.08	38.97	41.64	44.18	46.62	49.73
24	28.24	29.55	31.13	33.20	36.42	39.36	40.27	42.98	45.56	48.03	51.18
25	29.34	30.68	32.28	34.38	37.65	40.65	41.57	44.31	46.93	49.44	52.62
26	30.43	31.79	33.43	35.56	38.89	41.92	42.86	45.64	48.29	50.83	54.05
27	31.53	32.91	34.57	36.74	40.11	43.19	44.14	46.96	49.64	52.22	55.48
28	32.62	34.03	35.71	37.92	41.34	44.46	45.42	48.28	50.99	53.59	56.89
29	33.71	35.14	36.85	39.09	42.56	45.72	46.69	49.59	52.34	54.97	58.30
30	34.80	36.25	37.99	40.26	43.77	46.98	47.96	50.89	53.67	56.33	59.70
40	45.62	47.27	49.24	51.81	55.76	59.34	60.44	63.69	66.77	69.70	73.40
50	56.33	58.16	60.35	63.17	67.50	71.42	72.61	76.15	79.49	82.66	86.66
60	66.98	68.97	71.34	74.40	79.08	83.30	84.58	88.38	91.95	95.34	99.61
80	88.13	90.41	93.11	96.58	101.9	106.6	108.1	112.3	116.3	120.1	124.8
100	109.1	111.7	114.7	118.5	124.3	129.6	131.1	135.8	140.2	144.3	149.4