Hypothesis testing

Null & Alternative Hypothesis: The null hypothesis, H_0 is the commonly accepted fact; it is the opposite of the alternate hypothesis (H_1). Researchers work to reject, nullify or disprove the null hypothesis.

Why is it Called the "Null"?
The word "null" in this context means that it's a commonly accepted fact that researchers work to nullify. It doesn't mean that the statement is null itself! (Perhaps the term should be called the "**nullifiable hypothesis**" as that might cause less confusion).

Example: A researcher is studying the effects of radical exercise program on knee surgery patients. There is a good chance that the therapy will improve recovery time, but there's also the possibility it will make it worse. Average recovery times for knee surgery is 8.2 weeks.

H_0 (Null Hypothesis) : Average recovery times for knee surgery is 8.2 weeks.
H_1 (Alternative Hypothesis) : Average recovery times for knee surgery is NOT 8.2 weeks (could be more or less).

P-Value: When you perform a hypothesis test in statistics, a p-value helps you determine the significance of your results. The p-value is a number between 0 and 1 and interpreted in the following way: A small p-value (typically ≤ 0.05) indicates strong evidence against the null hypothesis, so you reject the null hypothesis.

A neurologist is testing the effect of a drug on response time by injecting 100 rats with a unit dose of the drug, subjecting each to neurological stimulus, and recording its response time. The neurologist knows that the mean response time for rats not injected with the drug is 1.2 seconds. The mean of the 100 drug injected rats' response time is 1.05 seconds with a sample standard deviation of 0.5 seconds. Do you think that the drug has an effect on response time?

H_0 : Drug has no effect on response time. (i.e., the mean response time is 1.2 seconds)
H_1 : Drug has an effect on response time. (i.e., the mean response time is not1.2 seconds)

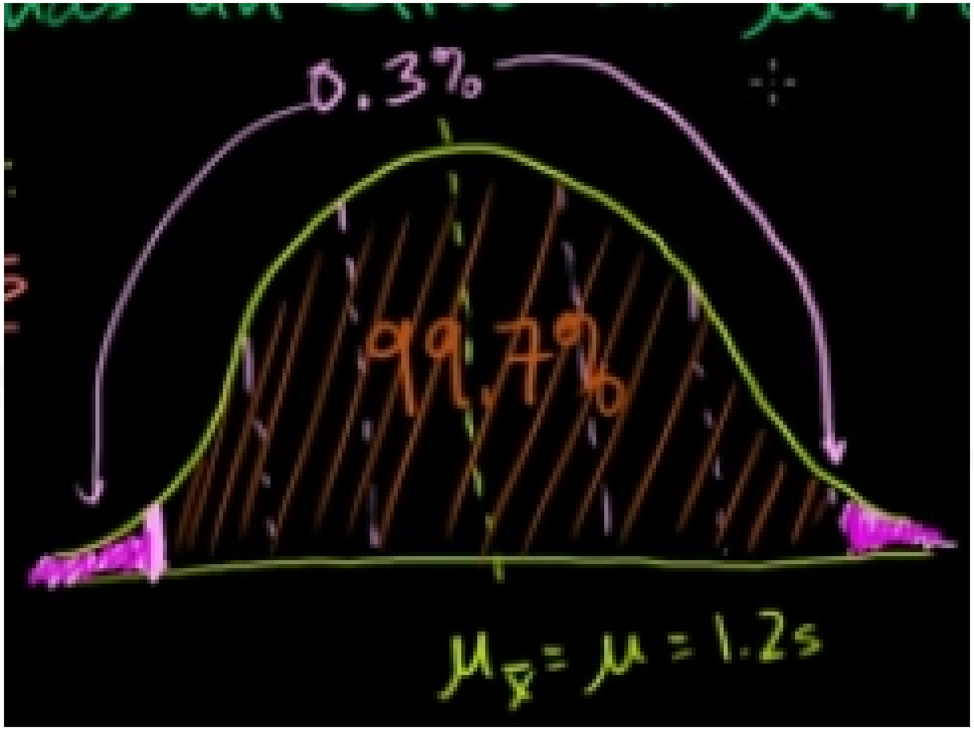
* Assume H_0 is true → From central limit theorem the mean of the sampling distribution is 1.2 seconds (population mean).
* Now we have to find how likely(probability) that the response time is 1.05, when drug has no effect (null hypothesis)
* That is how many standard deviations away 1.05 is from 1.2. This gives the area under the curve of a normal distribution.
* Need to find standard deviation of the sampling distribution. The formula for this is = population SD/sqrt(# of samples)
* We don't know the population standard deviation, hence we take best approximation as 0.5 (sample SD) as population SD.

$$\begin{aligned}\text{Sampling distribution SD} &= 0.5/\text{sqrt}(100)\\ &= 0.5/10\\ &= 0.05\end{aligned}$$

$$\begin{aligned}\text{Z-Score (how may SD away 1.05 from 1.2)} &= (1.2 - 1.05)/0.05\\ &= 3\end{aligned}$$

* **P-value:** 3 standard deviations away (99.7%). Only 0.3% (P-value is 0.003) chance getting 1.05 seconds response when drug has no effect.

Hence H_0 (Null Hypothesis) is rejected.



One-tailed and Two-tailed Tests

The above example is a Two-tailed test. This depends on the alternative hypothesis we chose. Above example H_1 is looking for has an effect on response time, but not for more or less.

H_0 : Drug has no effect on response time (mean = 1.2 seconds)
H_1 : Drug has an effect on response time. (mean != 1.2 seconds)

One-Tailed test: If the H_1 is mentioned as below then it would have been One-Tailed test. The P-Value would have been 0.0015.

H_1 : Drug lowers the response time (< 1.2 seconds)

Z-Statistic vs T-Statistic

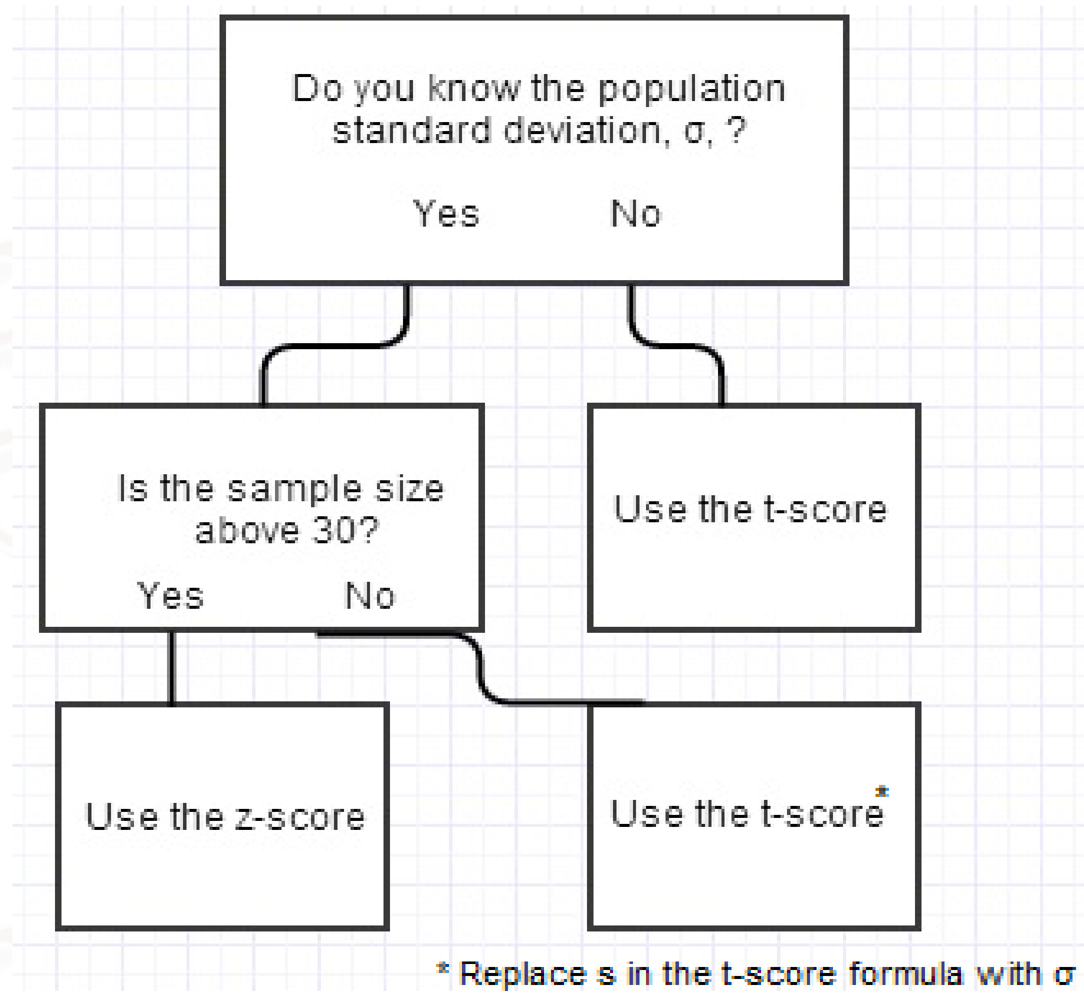
A z-score and a t score are both used in hypothesis testing. When sample size is less than 30 then use t-distribution, or else use z-distribution.

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

$$z = (X - \mu)/\sigma$$

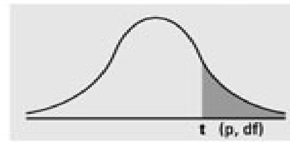
$$s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$$

$$t = (X - \bar{X})/s$$



T-Distribution Table

Numbers in each row of the table are values on a *t*-distribution with
(*df*) degrees of freedom for selected right-tail (greater-than) probabilities (*p*).



df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	43178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193
22	0.256432	0.685805	1.321237	1.717144	2.07387	2.50832	2.81876	3.7921
23	0.256297	0.685306	1.319460	1.713872	2.06866	2.49987	2.80734	3.7676
24	0.256173	0.684850	1.317836	1.710882	2.06390	2.49216	2.79694	3.7454
25	0.256060	0.684430	1.316345	1.708141	2.05954	2.48511	2.78744	3.7251
26	0.255955	0.684043	1.314972	1.705618	2.05553	2.47863	2.77871	3.7066
27	0.255858	0.683685	1.313703	1.703288	2.05183	2.47266	2.77068	3.6896
28	0.255768	0.683353	1.312527	1.701131	2.04841	2.46714	2.76326	3.6739
29	0.255684	0.683044	1.311434	1.699127	2.04523	2.46202	2.75639	3.6594
30	0.255605	0.682756	1.310415	1.697261	2.04227	2.45726	2.75000	3.6460
z	0.253347	0.674490	1.281552	1.644854	1.95996	2.32635	2.57583	3.2905
CI	———	———	80%	90%	95%	98%	99%	99.9%

Chi-Square Goodness of Fit Test

When an analyst attempts to fit a statistical model to observed data, he or she may wonder how well the model actually reflects the data. How "close" are the observed values to those which would be expected under the fitted model? One statistical test that addresses this issue is the chi-square goodness of fit test.

In general, the chi-square test statistic is of the form

$$X^2 = \sum \frac{(\text{observed - expected})^2}{\text{expected}}$$

Example: A new casino game involves rolling 3 dice. The winnings are directly proportional to the total number of sixes rolled. Suppose a gambler plays the game 100 times, with the following observed counts:

Number of Sixes	Number of Rolls
0	48
1	35
2	15
3	3

The casino becomes suspicious of the gambler and wishes to determine whether the dice are fair. What do they conclude?

* Probability to rolling a 6 on any given toss – 1/6. Assuming the 3 dice are independent (the roll of one die should not affect the roll of the others), we might assume that the number of sixes in three rolls is distributed Binomial.

* To determine whether the gambler's dice are fair, we may compare his results with the results expected under this distribution. On 3 dies

$$P_p(n \mid N) = \binom{N}{n} p^n q^{N-n} = \frac{N!}{n!\,(N-n)!}\, p^n (1-p)^{N-n},$$

Null Hypothesis:

→ The expected values for 0 sixes P(X=0) = 3^C0 x $(1/6)^0$ x $(1-1/6)^3$ = 1 x 1 x $(5/6)^3$ = 0.578
→ The expected values for 1 sixes P(X=1) = 3^C1 x $(1/6)^1$ x $(1-1/6)^2$ = 3 x (1/6) x $(5/6)^2$ = 0.347
→ The expected values for 2 sixes P(X=2) = 3^C2 x $(1/6)^2$ x $(1-1/6)^1$ = 3 x $(1/6)^2$ x $(5/6)^3$ = 0.069
→ The expected values for 3 sixes P(X=3) = 3^C3 x $(1/6)^3$ x $(1-1/6)^0$ = 1 x $(1/6)^3$ x 1 = 0.005

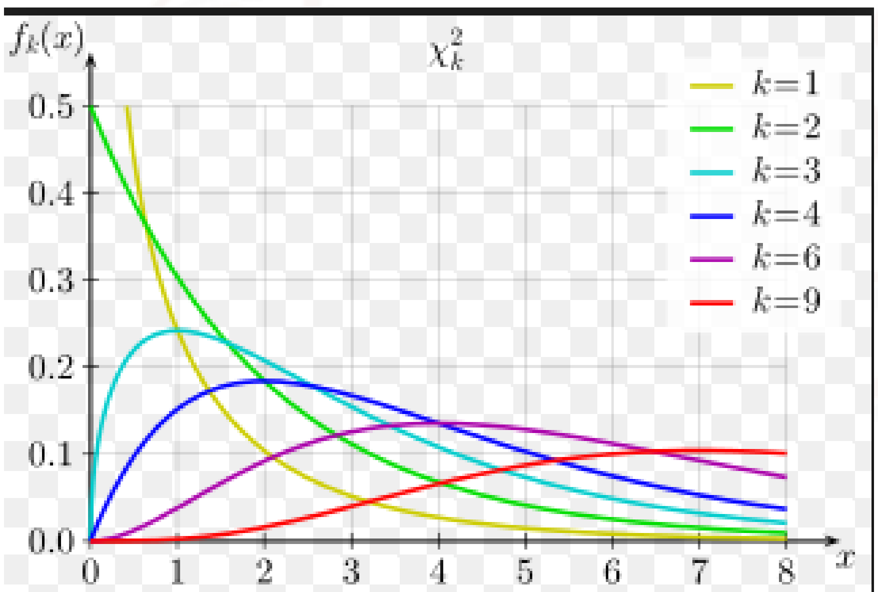
Since the gambler plays 100 times, the expected counts are the following:

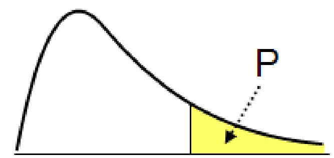
Number of Sixes	Expected Counts	Observed Counts
0	58.0	48
1	34.5	35
2	7.0	15
3	0.5	3

The chi-square statistic is = $(48-58)^2/58 + (35-34.5)^2/58 + (15-7)^2/7 + (3-0.5)^2/0.5$
= 1.72 + 0.007 + 9.14 + 12.5
= 23.367.

* We need to look into Chi-Square distribution table and curve. We have four random variables, hence the degrees of freedom is 3.

* If we are interested in a significance level of 0.05 we may reject the **null hypothesis (that the dice are fair)** if > 7.815, the value corresponding to the 0.05 significance level (,3 df) for the distribution. Since 23.367 is clearly greater than 7.815, we may reject the null hypothesis that "**the dice are fair at the 0.05 significance level**".





	P										
DF	**0.995**	**0.975**	**0.20**	**0.10**	**0.05**	**0.025**	**0.02**	**0.01**	**0.005**	**0.002**	**0.001**
1	0.0000393	0.000982	1.642	2.706	3.841	5.024	5.412	6.635	7.879	9.550	10.828
2	0.0100	0.0506	3.219	4.605	5.991	7.378	7.824	9.210	10.597	12.429	13.816
3	0.0717	0.216	4.642	6.251	7.815	9.348	9.837	11.345	12.838	14.796	16.266
4	0.207	0.484	5.989	7.779	9.488	11.143	11.668	13.277	14.860	16.924	18.467
5	0.412	0.831	7.289	9.236	11.070	12.833	13.388	15.086	16.750	18.907	20.515
6	0.676	1.237	8.558	10.645	12.592	14.449	15.033	16.812	18.548	20.791	22.458
7	0.989	1.690	9.803	12.017	14.067	16.013	16.622	18.475	20.278	22.601	24.322
8	1.344	2.180	11.030	13.362	15.507	17.535	18.168	20.090	21.955	24.352	26.124
9	1.735	2.700	12.242	14.684	16.919	19.023	19.679	21.666	23.589	26.056	27.877
10	2.156	3.247	13.442	15.987	18.307	20.483	21.161	23.209	25.188	27.722	29.588
11	2.603	3.816	14.631	17.275	19.675	21.920	22.618	24.725	26.757	29.354	31.264
12	3.074	4.404	15.812	18.549	21.026	23.337	24.054	26.217	28.300	30.957	32.909
13	3.565	5.009	16.985	19.812	22.362	24.736	25.472	27.688	29.819	32.535	34.528
14	4.075	5.629	18.151	21.064	23.685	26.119	26.873	29.141	31.319	34.091	36.123
15	4.601	6.262	19.311	22.307	24.996	27.488	28.259	30.578	32.801	35.628	37.697
16	5.142	6.908	20.465	23.542	26.296	28.845	29.633	32.000	34.267	37.146	39.252
17	5.697	7.564	21.615	24.769	27.587	30.191	30.995	33.409	35.718	38.648	40.790
18	6.265	8.231	22.760	25.989	28.869	31.526	32.346	34.805	37.156	40.136	42.312
19	6.844	8.907	23.900	27.204	30.144	32.852	33.687	36.191	38.582	41.610	43.820
20	7.434	9.591	25.038	28.412	31.410	34.170	35.020	37.566	39.997	43.072	45.315
21	8.034	10.283	26.171	29.615	32.671	35.479	36.343	38.932	41.401	44.522	46.797
22	8.643	10.982	27.301	30.813	33.924	36.781	37.659	40.289	42.796	45.962	48.268

F- Statistic (ANOVA)

Figure out how much of the total variance comes from:

The variance *between* the groups

The variance *within* the groups

Calculate the ratio:

$$F = \frac{between\ groups}{within\ groups}$$

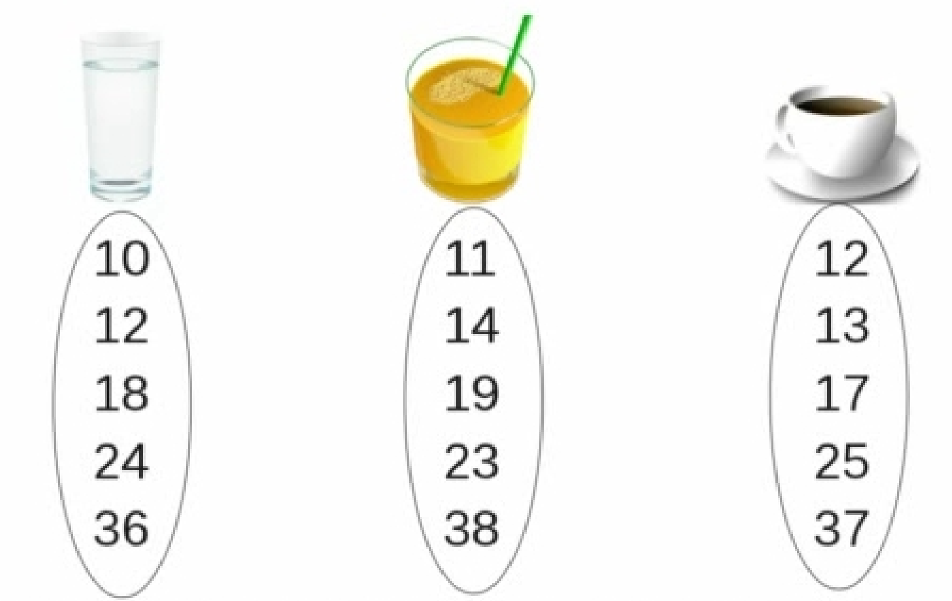
The larger the ratio, the more likely it is that the groups have different means (reject H_0).

Suppose that three groups have given water, fruit juice, coffee. Now we want to test the response between three groups. We have to use F-test as we have more than two groups (otherwise we would have used t-test)

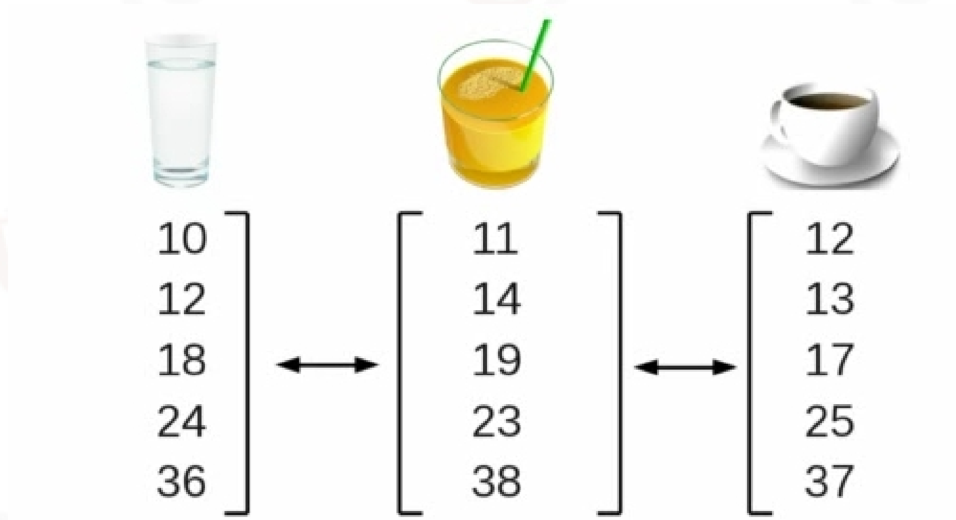
Null Hypothesis: Drink didn't make much difference

Example 1: Lot of variation with in group, but a little between groups.

This case we will **accept** Null Hypothesis, variation between groups is not much.



There's a lot of variation in each group...



...but each group looks pretty much the same.

Example 2: A little variation with in group, but a lot of variation between groups.

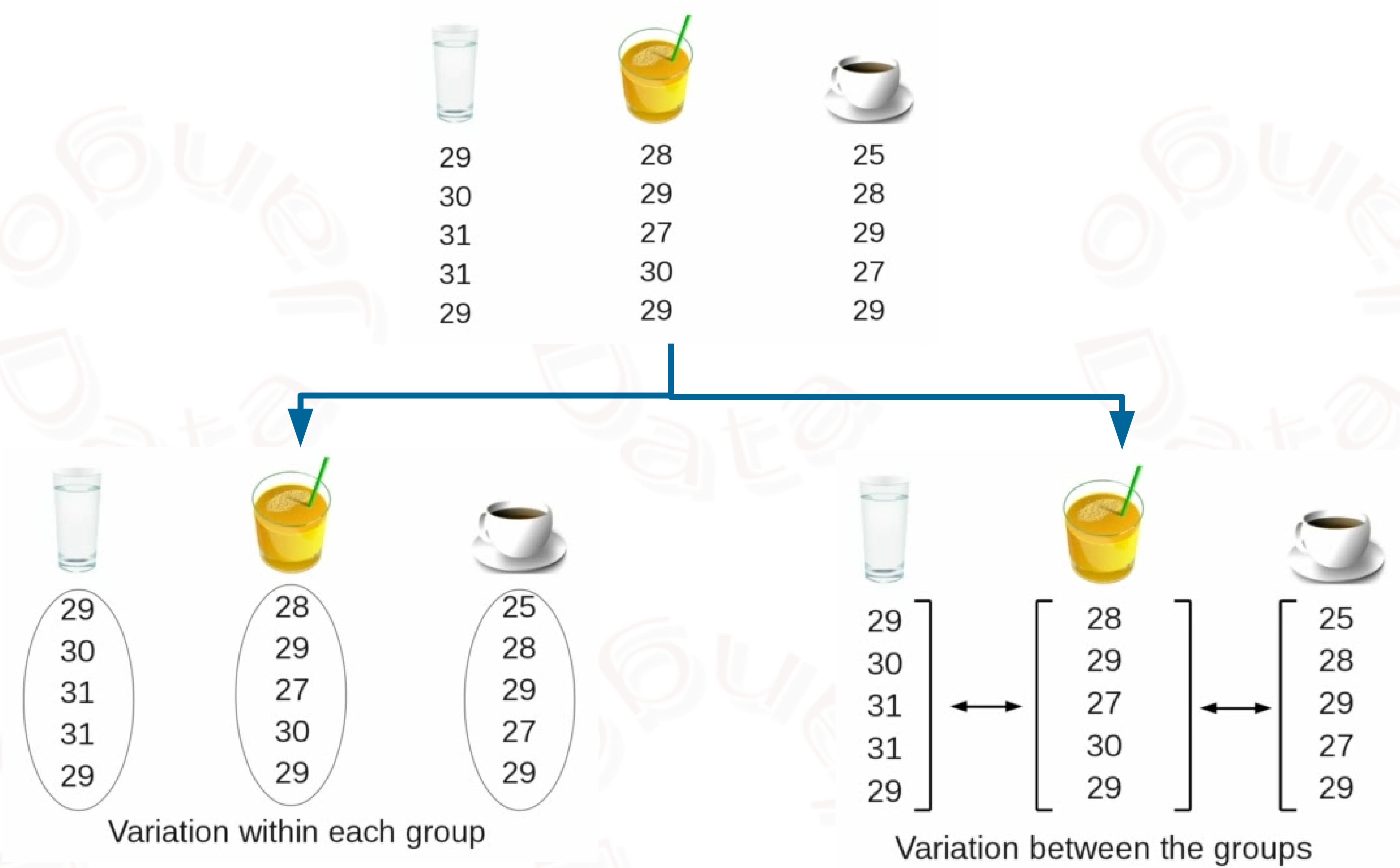
This case we will **reject** the Null Hypothesis

29	17	10
29	18	11
30	19	12
31	19	12
31	20	13

Conclusion: it’s the drink that makes the difference, not the people.

Example 3: There is not much variation with in OR between groups.

The result of ANOVA calculation is F(2, 12) = 4.7, the p-value = 0.04 (from F-distribution table) . We can **reject** the Null Hypothesis.



Calculate the degrees of freedom as follows:

$$F(b, w)$$

b is the degrees of freedom for variance between groups.

w is the degrees of freedom for variance within groups.

b = number of groups – 1
w = total number of observations – number of groups

ANOVA Calculation example

Sample 1 : Stress under normal condition
Sample 2 : Stress after announced layoffs
Sample 3 : Stress during layoffs
Need to measure the impact of announced layoffs.

Null Hypothesis : No impact of announced layoffs on employee stress

Analysis of Variance

Sum of Squares Within Groups

sample

2	- 4 =	-2^2	4
3	- 4 =	-1^2	1
7	- 4 =	3^2	9
2	- 4 =	-2^2	4
6	- 4 =	2^2	4
			22

sample

10	- 8 =	2^2	4
8	- 8 =	0^2	0
7	- 8 =	-1^2	1
5	- 8 =	-3^2	9
10	- 8 =	2^2	4
			18

sample

10	- 13 =	-3^2	9
13	- 13 =	0^2	0
14	- 13 =	1^2	1
13	- 13 =	0^2	0
15	- 13 =	2^2	4
			14

Sum of **S**quares **W**ithin Groups = 22 + 18 + 14 = 54

Total Sum of Squares = Sum of Squares Between Groups + Sum of Squares Within Groups

54

observation		mean	observation - mean	(observation - mean) ^{2}
2	-	8.3	= -6.3	40.1
3	-	8.3	= -5.3	28.4
7	-	8.3	= -1.3	1.8
2	-	8.3	= -6.3	40.1
6	-	8.3	= -2.3	5.4
10	-	8.3	= 1.7	2.7
8	-	8.3	= -0.3	0.1
7	-	8.3	= -1.3	1.8
5	-	8.3	= -3.3	11.1
10	-	8.3	= 1.7	2.8
10	-	8.3	= 1.7	2.8
13	-	8.3	= 4.7	21.8
14	-	8.3	= 5.7	32.1
13	-	8.3	= 4.7	21.8
15	-	8.3	= 6.7	44.4

Total **S**um of **S**quares

SST = 257.3

Total Sum of Squares = Sum of Squares Between Groups + Sum of Squares Within Groups

257.3 | 54

Analysis of Variance

Sum of Squares Between Groups

2	2	10	10
3	3	8	13
7	7	7	14
2	2	5	13
6	6	10	15
10	mean	mean	mean
8			
7			
5			
10			
10			
13			
14			
13			
15			
mean			

1. mean - mean   mean - mean   mean - mean

2. (mean - mean)^2 (mean - mean)^2 (mean - mean)^2

3. (mean - mean)^2 + (mean - mean)^2 + (mean - mean)^2

4. (mean - mean)^2 + (mean - mean)^2 + (mean - mean)^2 x 5

Analysis of Variance

Sum of Squares Between Groups

2
3
7
2
6
10
8
7
5
10
10
13
14
13
15

mean 8.3

	2	10	10
	3	8	13
	7	7	14
	2	5	13
	6	10	15
mean	$4 - 8.3 = (-4.3)^2$	$8 - 8.3 = (-.3)^2$	$13 - 8.3 = (4.7)^2$

$18.8 + .1 + 21.8 = 40.7$

$40.7 \times 5 = 203.3$

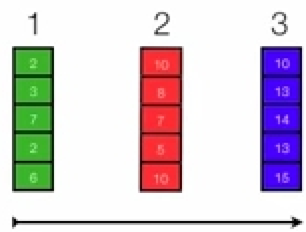
Total Sum of Squares	=	Sum of Squares Between Groups	+	Sum of Squares Within Groups
257.3	=	203.3	+	54

Final Calculations

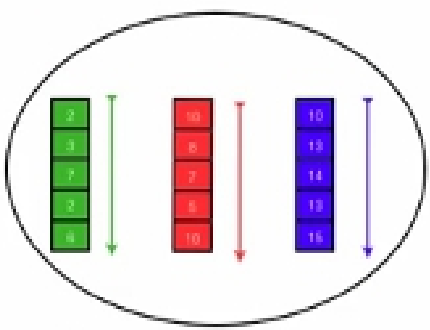
$$\frac{\text{Sum of Squares Between Groups}}{\text{degrees of freedom}} = \frac{203.3}{}$$

Groups - 1

3 - 1 = 2



$$\frac{\text{Sum of Squares Between Groups}}{\text{degrees of freedom}} = \frac{203.3}{2} = 101.667$$



15 - 3 = 12

Observations - Groups

$$\frac{\text{Sum of Squares Within Groups}}{\text{degrees of freedom}} = \frac{54}{12} = 4.5$$

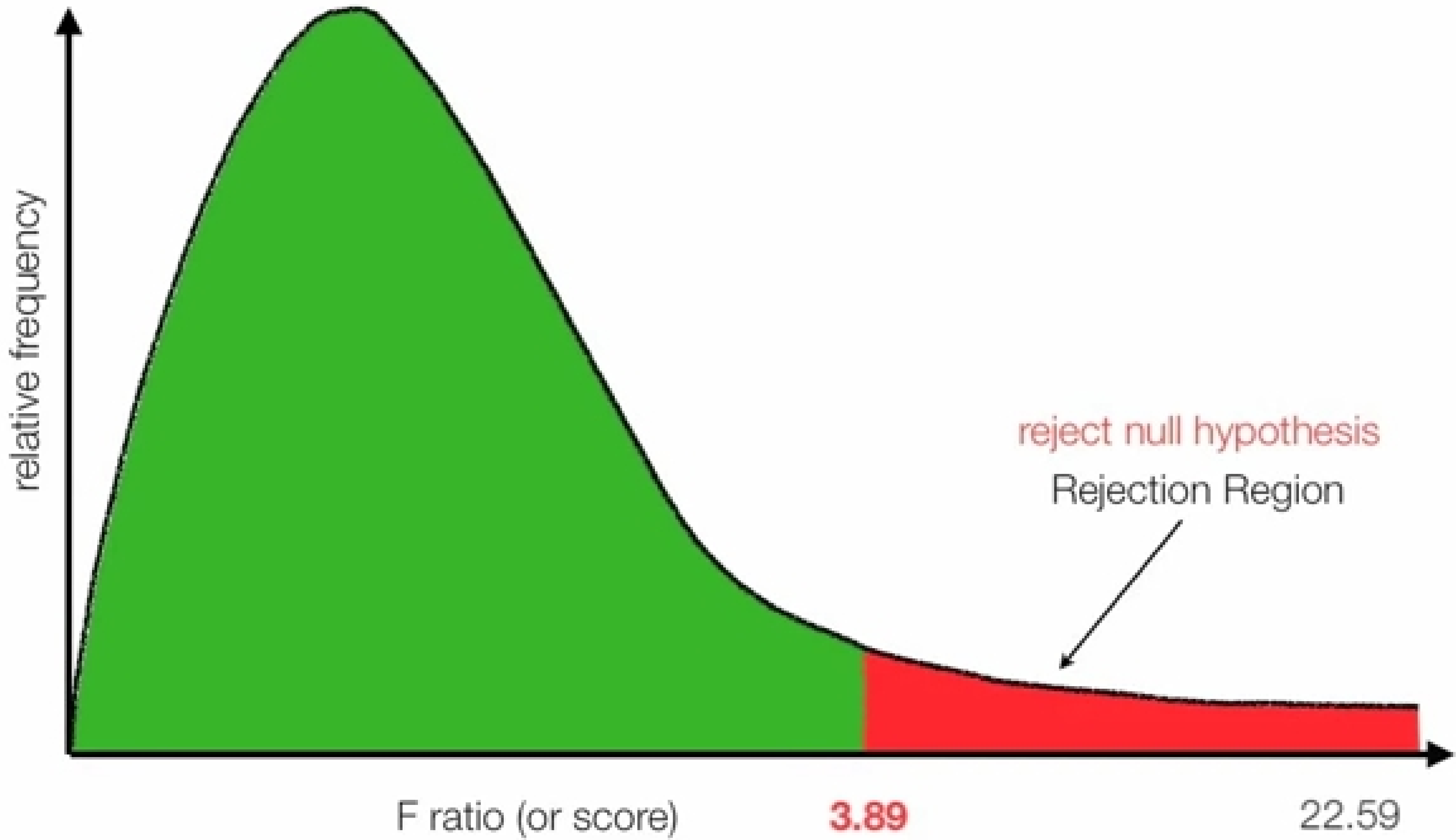
$$\boxed{F = \frac{101.667}{4.5} = 22.59}$$

F Distribution $F(2, 12) = 22.59$, $p<.05$

degrees of freedom numerator

degrees of freedom denominator

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	161.5	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9	246.0	248.0	249.1	250.1
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.45	19.46
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25

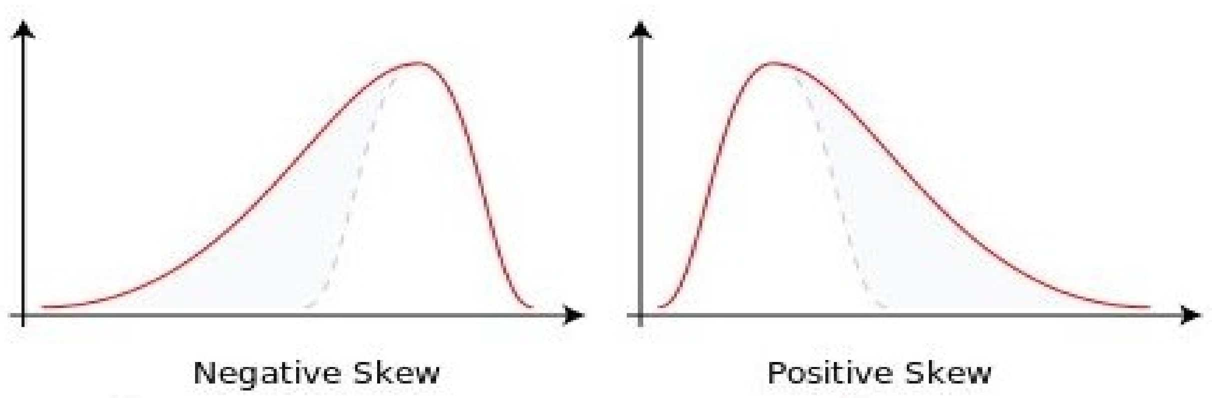

relative frequency
reject null hypothesis
Rejection Region
F ratio (or score)
3.89
22.59

Skewness

Consider the two distributions in the figure just below. Within each graph, the values on the right side of the distribution taper differently from the values on the left side. These tapering sides are called tails, and they provide a visual means to determine which of the two kinds of skewness a distribution has:

negative skew: The left tail is longer; the mass of the distribution is concentrated on the right of the figure. The distribution is said to be left-skewed, left-tailed, or skewed to the left, despite the fact that the curve itself appears to be skewed or leaning to the right; left instead refers to the left tail being drawn out and, often, the mean being skewed to the left of a typical center of the data. A left-skewed distribution usually appears as a right-leaning curve.

positive skew: The right tail is longer; the mass of the distribution is concentrated on the left of the figure. The distribution is said to be right-skewed, right-tailed, or skewed to the right, despite the fact that the curve itself appears to be skewed or leaning to the left; right instead refers to the right tail being drawn out and, often, the mean being skewed to the right of a typical center of the data. A right-skewed distribution usually appears as a left-leaning curve.[1]



Skewness in a data series may sometimes be observed not only graphically but by simple inspection of the values. For instance, consider the numeric sequence (49, 50, 51), whose values are evenly distributed around a central value of 50. We can transform this sequence into a negatively skewed distribution by adding a value far below the mean, e.g. (40, 49, 50, 51). Similarly, we can make the sequence positively skewed by adding a value far above the mean, e.g. (49, 50, 51, 60).

The Pearson mode skewness, or first skewness coefficient, is defined as

$$\frac{\text{mean} - \text{mode}}{\text{standard deviation}}.$$

The Pearson median skewness, or second skewness coefficient, is defined as

$$\frac{3\,(\text{mean} - \text{median})}{\text{standard deviation}}.$$

Person MODE vs MEDIAN skewness: Pearson’s first coefficient of skewness uses the mode. Therefore, if the mode is made up of too few pieces of data it won’t be a stable measure of central tendency. For example, the mode in both these sets of data is 9:
1 2 3 4 5 6 7 8 9 9.
1 2 3 4 5 6 7 8 9 9 9 9 9 9 9 9 9 9 9 9 10 12 12 13.
In the first set of data, the mode only appears twice. This isn’t a good measure of central tendency so you would be cautioned not to use Pearson’s coefficient of skewness. The second set of data has a more stable set (the mode appears 12 times). Therefore, Pearson’s coefficient of skewness will likely give you a reasonable result.

Interpretation (Skewness):

- The direction of skewness is given by the sign.
- The coefficient compares the sample distribution with a normal distribution. The larger the value, the larger the distribution differs from a normal distribution.
- A value of zero means no skewness at all.
- A large negative value means the distribution is negatively skewed.
- A large positive value means the distribution is positively skewed.

Kurtosis

Population Kurtosis Formula

$$K = n \frac{\sum_{i=1}^{n}(X_i - X_{avg})^4}{(\sum_{i=1}^{n}(X_i - X_{avg})^2)^2}$$

Sample Kurtosis Formula

$$K = \frac{n(n+1)(n-1)}{(n-2)(n-3)} \frac{\sum_{i=1}^{n}(X_i - X_{avg})^4}{(\sum_{i=1}^{n}(X_i - X_{avg})^2)^2}$$