A with Honors Projects | Honors Program

Spring 2023

Calculating P-values

Kylie Fuoss

Kylie Fuoss
Instructor: Lesile Smith

Calculating P-values

Part 1:

EX: A Marine Biologist claims that the mean production of bubbles blown by the population of sunfish is less than 25 bubbles per hour. (Assume the standard deviation of the population of fish bubbles blown per hour is 3.3). A random sample of 35 fish are collected and observed to find that they have a mean of 22 bubbles blown per hour.
Test the Marine Biologists claim at a 0.05 level of significance.

Step 1: H_0: μ= 25 H_1: μ< 25

Step 2: α= 0.05

Step 3:
Using Ztest

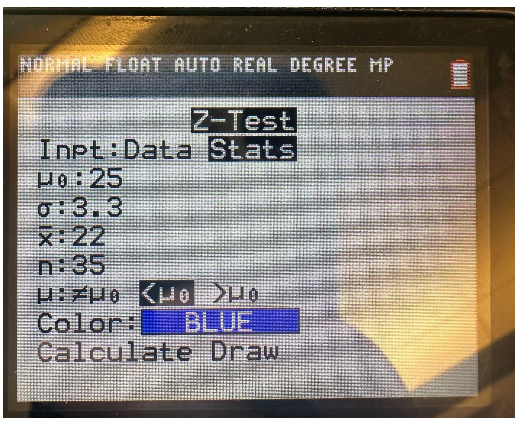
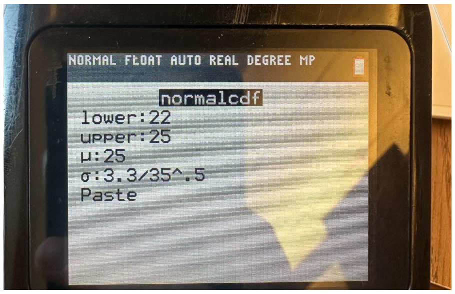


Figure 1

We are able to utilize a Ztest because we are given the population standard deviation, if we were not given a population standard deviation and instead a sample standard deviation we would then utilize a Ttest if n≥30. μ_0 is our population mean σ is our population standard deviation “x bar” is the mean of our sample and n is the number in our sample. We select < μ_0 because we are trying to prove if the mean number of bubbles blown per hour is less than 25. This in turn becomes the null hypothesis μ= 25 and our alternative hypothesis μ< 25. Once you press “Calculate” the p-value= 3.77×10^{-8}.

\

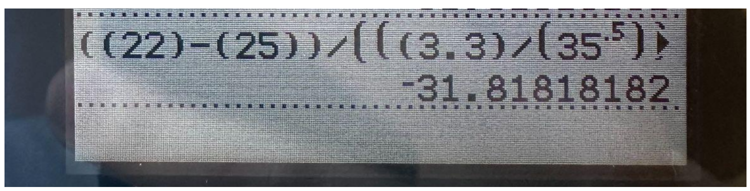
Using Normal CDF



To utilize normal CDF we must have a normal population meaning n≥30. We are trying to find the p-value which will be the area of the curve to the left of 22 bubbles per hour. Noramalcdf calculates the area under the curve between an interval a and b. Normalcdf syntax is normalcdf(a,b,μ,σ/(n)$^{.5}$) μ being population mean and σ being population standard deviation and n being number in sample. Because we are utilizing a normal curve we must add or subtract from 0.5 or 1 to achieve the probability, here we will subtract from 0.5 because 22 is less than the population mean 25 and on that half of the curve the total area would be 0.5. Therefore to calculate less than 22 we will use 22 as “a” and 25 as “b” for our interval and subtract it from the total 0.5 to achieve the probability we want. 0.5- normalcdf(22,25,25,3.3/35$^{.5}$) our p-value is 3.77x10^{-8}.

Using Table and calculating Z value

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$



z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414
-0.1	.46017	.45620	.45224	.44828	.44433	.44034	.43640	.43251	.42858	.42465
-0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
-0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
-0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
-0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
-0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
-0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
-0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
-0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
-1	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08692	.08534	.08379	.08226
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-2	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-3	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-4	.00003	.00003	.00003	.00003	.00003	.00003	.00002	.00002	.00002	.00002

To calculate the Z score we use the above equation where “x bar” is sample mean and μ is population mean; σ is calculated by $\sigma/(n)^{.5}$ and then divided by square root of sample size. The calculated Z value is -31.81. You then utilize the table, on the side labeled Z scroll to -31.8 and then go to the row containing 0.01.

Step 4:
P-value < α
Reject null hypothesis.

Step 5:
There is significant evidence at the 0.05 level of significance to conclude that the mean number of bubbles sunfish blow per hour is less than 25.

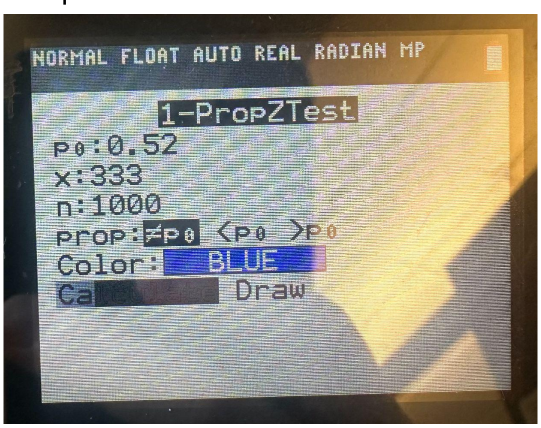
Part 2:

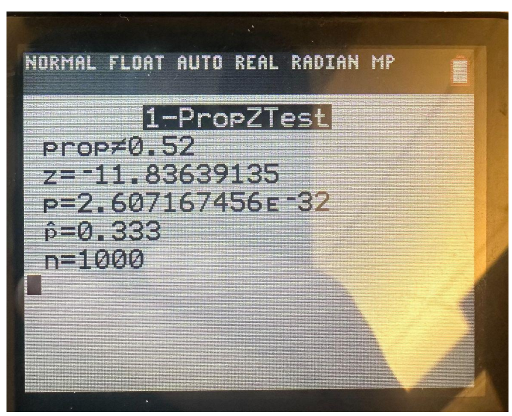
EX: A study was done on leaves testing the time the leaves turned brown. 1000 leaves were sampled and showed that 333 of them turned brown by the beginning of September. Is there significant evidence to conclude that the rate of leaves turning brown by the beginning of September differs from 0.52, the proportion of leaves that had turned brown last year by the beginning of September? Test with a 0.01 level of significance.

Step 1:
H_0: p=0.52 H_1: p≠ 0.52

Step 2: α=0.01

Step 3:
1PropZTest





Here we use 1PropZTest because this problem deals with proportions. We are given a proportion to test p_0=0.52, x is the number out of the total sample 333 and n is the sample 1000. We choose p≠ because we are just testing if it is different, not if it is above or below. Press “Calculate” to find that p= 2.61×10^{-32}

Using Normal CDF

$$Z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1 - p_o)}{n}}}$$

To utilize normalcdf we will need a normal distribution. To create this we find a Z score using the above equation. P□ is our population proportion 0.52 and “p hat” is our sample proportion with 333/1000=0.333. In a normal distribution the mean is 0 and the standard deviation is 1 which will be utilized in normalcdf. The calculated Z value is -11.84. Using this as our interval number with 0 we calculate normalcdf.

Using Table and calculating Z value

z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414
-0.1	.46017	.45620	.45224	.44828	.44433	.44034	.43640	.43251	.42858	.42465
-0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
-0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
-0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
-0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
-0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
-0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
-0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
-0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
-1	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
-1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
-1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
-1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08692	.08534	.08379	.08226
-1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
-1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
-1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
-1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
-1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
-1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
-2	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
-2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
-2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
-2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
-2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
-2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
-2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
-2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
-2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
-2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
-3	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
-3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
-3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
-3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
-3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
-3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
-3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
-3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
-3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
-3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003
-4	.00003	.00003	.00003	.00003	.00003	.00003	.00002	.00002	.00002	.00002

The z value has already been calculated above to be -11.84. Using the above table we can use the column to find -11.8 and the row to find .04 for our p-value.

Step 4:
p-value < α
Reject null hypothesis

Step 5:
There is evidence at the 0.01 significance level that the number of leaves turned brown at the beginning of September is not 0.52.

Part 3:

EX: In a balloon museum, attendants have been polled to determine if a relationship exists between age and enjoyment of said balloon museum. The results are present in the table below. Test at a level of 0.05 significance.

Museum Enjoyment Level	**Age**		Total
	Child (<18)	Adult (>18)	
Enjoyed:	100	25	125
Did not Enjoy	45	20	65
Total	145	45	190

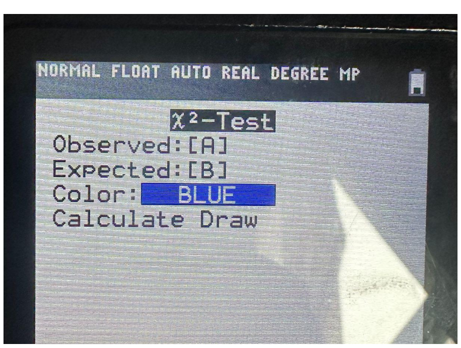
Step1:
H_0: Enjoyment of museum and age are independent
H_1: Enjoyment of museum and age are dependent

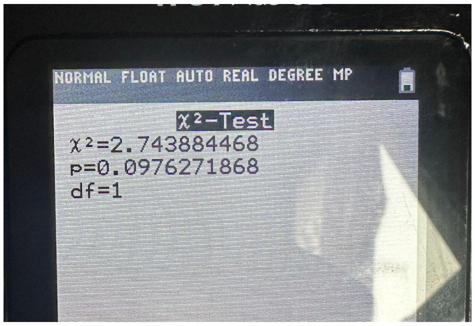
Step 2:
α= 0.05

Step 3:
Using X^2- test for independence (calculator)

A= 100 25
45 20
E= (RxC)/n

B= 95.39 29.61
49.61 15.39



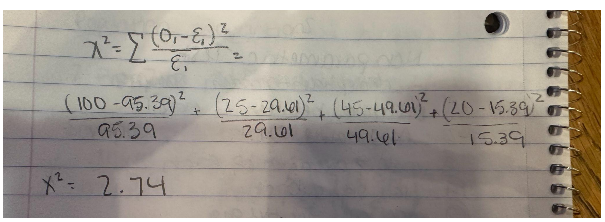


p-value= 0.0976

Here we just simply calculate the matrix values from matrix A into matrix B using E= (RxC)/n where R is the sum of the row, C is the sum of the column and n is the total number. We then utilize our calculator Chi Square Test for Independence to find the p value 0.0976.

Using X^2cdf

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$



Utilizing this X^2cdf we can take our X^2 value from the formula above calculating 2.744 and utilize it in X^2 cdf. With the formula p-value= 1- X^2cdf(0,2.744,n-1) n-1 being the degree of freedom which is 2-1=1.

Using Chi- Squared Table

degrees of freedom	Area to the right of the Critical Value									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.452	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.144	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.196	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.772	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.43	104.21
80	51.172	53.540	57.153	60.391	64.278	96.578	101.88	106.63	112.33	116.32
90	59.196	61.754	65.647	69.126	73.291	107.57	113.15	118.14	124.12	128.30
100	67.328	70.065	74.222	77.929	82.358	118.50	124.34	129.56	135.81	140.17

To utilize this table we calculate our degrees of freedom (n-1) in our case 189 and then we look at our significance level and then we will find our p-value.

Step 4:
P-value > α
Fail to reject the null hypothesis

Step 5:
There is not enough evidence at the 0.05 significance level to conclude that the enjoyment of the museum and age are dependent.

Part 4

EX:
A sewer is interested in the preference of 3 different sewing machines, Singer, Brother and Juki.The last 900 sales show that 350 customers bought Singer, 323 bought Brother and 227 bought Juki. Test the hypothesis using 0.01 significance level that the 3 sewing brands are equally popular.

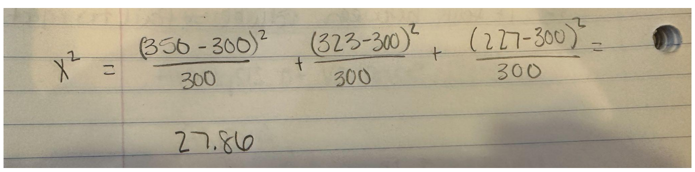
Step 1:
H_0: $p_1=p_2=p_3$
H_1: proportions are different from what was specified in H_0

Step 2:
α= 0.01

Step 3:
X^2 Goodness of Fit Test (formula)

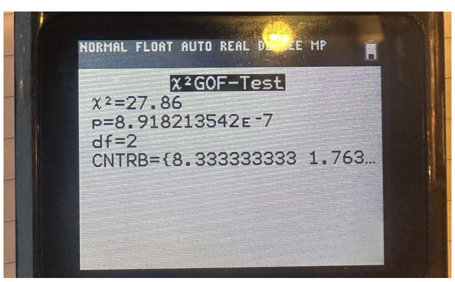
Observed Frequencies	350	323	227
Expected Frequencies	300	300	300

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$



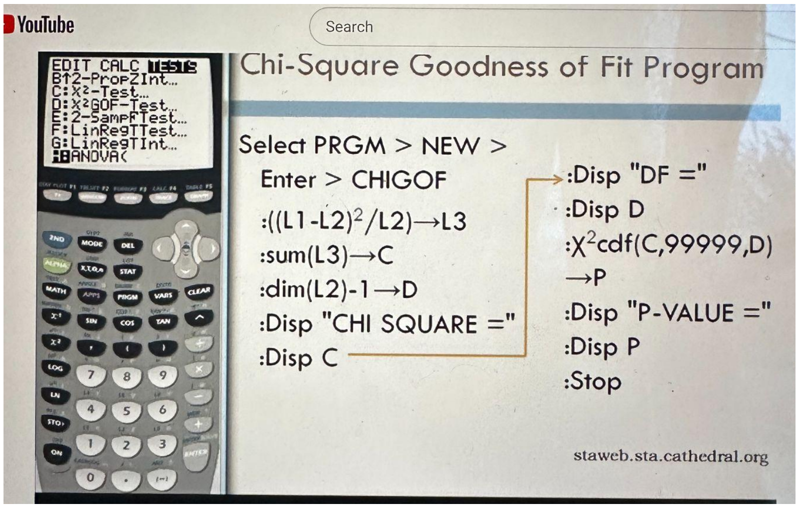
Using this formula where O is observed, E is expected, we can find our X^2 value. We can then use X^2 Cdf with formula p-value= 1- X^2Cdf(0, x^2, degrees of freedom) to find our p-value.

Using X^2 Goodness of Fit (calculator)



For this method we just type in our data into L_1 and L_2 then choose the GOF test and it will calculate the p-value for use, p-value= 8.92 x 10^{-7}

Using GOF program



These are the steps to take to add the GOF program to your calculator if your calculator does not already contain it. Afterward you can enter your L_1 and L_2 lists then select programs and select CHIGOF and it will calculate the p-value.

Using Chi-Squared Table

degrees of freedom	Area to the right of the Critical Value									
	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.452	10.982	12.338	14.042	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.144	36.741	40.113	43.194	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.196	41.337	44.461	48.278	50.993
29	13.121	14.257	16.047	17.708	19.768	39.087	42.557	45.772	49.588	52.336
30	13.787	14.954	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.43	104.21
80	51.172	53.540	57.153	60.391	64.278	96.578	101.88	106.63	112.33	116.32
90	59.196	61.754	65.647	69.126	73.291	107.57	113.15	118.14	124.12	128.30
100	67.328	70.065	74.222	77.929	82.358	118.50	124.34	129.56	135.81	140.17

To utilize this table we calculate our degrees of freedom (n-1) in our case 2 and then we look at our significance level and then we will find our p-value.

Step 4:
p-value<α
Reject null hypothesis

Step 5: There is significant evidence at the 0.01 level that there is a preference for a certain brand of sewing machine.