APL103 Experiment - 11
Chi-Square Test

1) **Aditya Agrawal (2021AM10198)**
2) **Swapnil Kashyap (2021AM10782)**
Worked on DATASET10

OBJECTIVE:
To test whether experimental data are well described by a Normal (Gaussian) Distribution using the χ^2 test.

APPARATUS/INSTRUMENTS:
1 Wind Tunnel 2. Instrumented chimney model
3. Strain Data Acquisition System 4. Computer.

PROCEDURE:

1. Ensure that the strain-gauge bridges on the chimney model are properly aligned with respect to the flow.
2. Start the data acquisition program and set all the bridges to zero (6 in all).
3. Turn on the wind tunnel and set the desired wind speed.
4 Acquire 1000 data points (bending strain) from each bridge over a period of approximately two minutes.
5. Store the data in a file for analysis.

ANALYSIS:
1. Calculate the mean and the standard deviation for bending strain.

Mean = 5.22; Standard Deviation = 12.312

2. Calculate the standard error and express the best estimate of the bending strain as μ = $\bar{\mu}$ ± standard error.

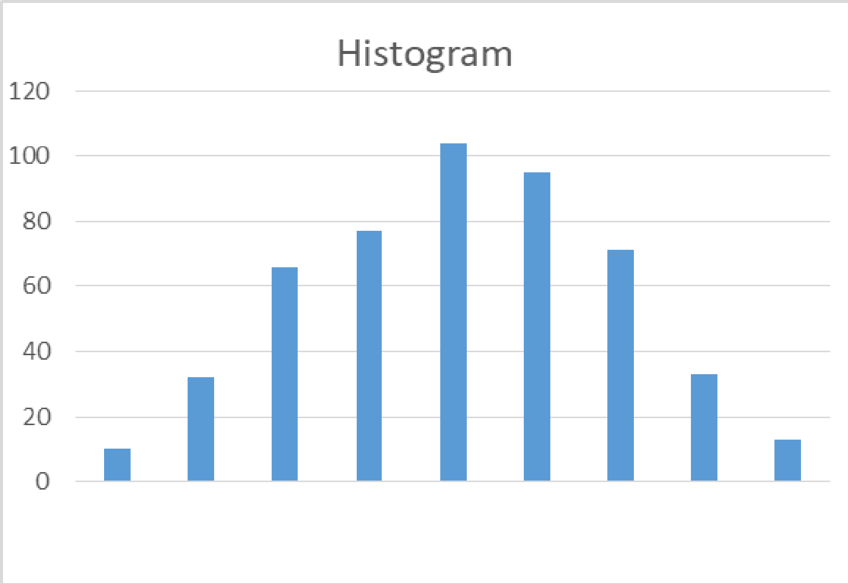
Standard error = 0.551

∴ μ = 5.22 ± 0.551

3. Divide the whole data into "n" groups; n is generally between 8 and 10. Ensure that each group has a frequency of at least ten.

The value of n chosen is 9.

4. Draw the normalized histogram, compare it with a normal curve, and note down your observation of their closeness. Can you say that the data follows a normal distribution? If yes, what is your confidence level? (Guess)



5. Use χ^2 test to test the data for a normal distribution.

$$\chi^2 = \sum R_i \quad where,\ R_i = \frac{(f_{oi} - f_{ei})^2}{f_{ei}}$$

Determine f_e by calculating the standardized Gaussian variable Z, as shown in the table shown below.
If, $\sum f_e$, does not equal 1000, then scale it up so that it does.

S.No.	Class Interval	f0	Z1	Z2	P(Z1)	P(Z2)	$\triangle$P(Z)	f_e	$f_{e(corrected)}$	Ri
1	-35 -- -19.5	10	-3.27	-2.01	0.00054	0.02222	0.02168	10.86168	11	0.090909
2	-19.5 -- -11	32	-2.01	-1.32	0.02222	0.09342	0.0712	35.6712	36	0.444444
3	-11 -- -5	66	-1.32	-0.83	0.09342	0.20327	0.10985	55.03485	55	2.2
4	-5 -- 1	77	-0.83	-0.34	0.20327	0.36693	0.16366	81.99366	82	0.304878
5	1 -- 8	104	-0.34	0.23	0.36693	0.591	0.22407	112.2591	112	0.571429
6	8 - 14	95	0.23	0.71	0.591	0.7611	0.1701	85.2201	85	1.176471
7	14 - 21	71	0.71	1.28	0.7611	0.8997	0.1386	69.4386	69	0.057971
8	21 - 30	33	1.28	2.01	0.8997	0.9778	0.0781	39.1281	39	0.923077
9	30 -- 48	13	2.01	3.47	0.9778	0.9997	0.0219	10.9719	11	0.363636
Sum		501							501	6.132815

6. For 10% probability, find the χ^2 value from χ^2 – F table. Does the experimental data follow the normal distribution at this probability?

χ^2 = 6.132815

Degree of Freedom = 9 - 3 = 6.

For P(x) < 0.6, which is less than the χ^2 value from the χ^2 – F table. Therefore, the experimental data given does follow normal distribution.

For 10% probability, χ^2 value is 2.204.

DISCUSSION:

1. **How many constraints are there for comparison against Normal distribution? Give reasons.**

   The number of constraints m is equal to the number of quantities obtained from the observation used in the calculation of expected frequencies. And m=3 because there are three quantities which have been used, i.e. total number of observations, the mean value and the standard deviation of the data.

   So degree of freedom = 9-3 = 6.

2. **At what probability can you accept the hypothesis that the given data follows Normal distribution?**

   When the critical value obtained from the χ^2-F table is more than the χ^2 value obtained from the data, one can accept the hypothesis that the given data follows Normal distribution for all probabilities.

3. **Instead of correcting the expected frequencies so that they add up to 1000, what other method could you use to ensure a proper comparison between the observed and expected frequencies?**

   One alternative to ensure proper comparison between observed and expected frequencies is using standardised residuals. It involves calculating the difference between observed and expected frequency, then dividing by the square root of the expected frequency, and comparing the resulting value to a threshold of 2 or -2.

4. **Why should the value of f_e not fall below 10?**

   If f_e value is below 10, then the impact of that particular class is higher than the others which causes error in χ^2 test.