Statistical Tests, Correlation and Regression

Math & Stats Tutorial
Day 6

Prabesh Dhakal

20 May 2019

REVIEW

1. Reviewed previous topics
2. Density Distribution
3. Process of Hypothesis Testing
4. p-Value
5. Chi-square Test

PLAN FOR TODAY

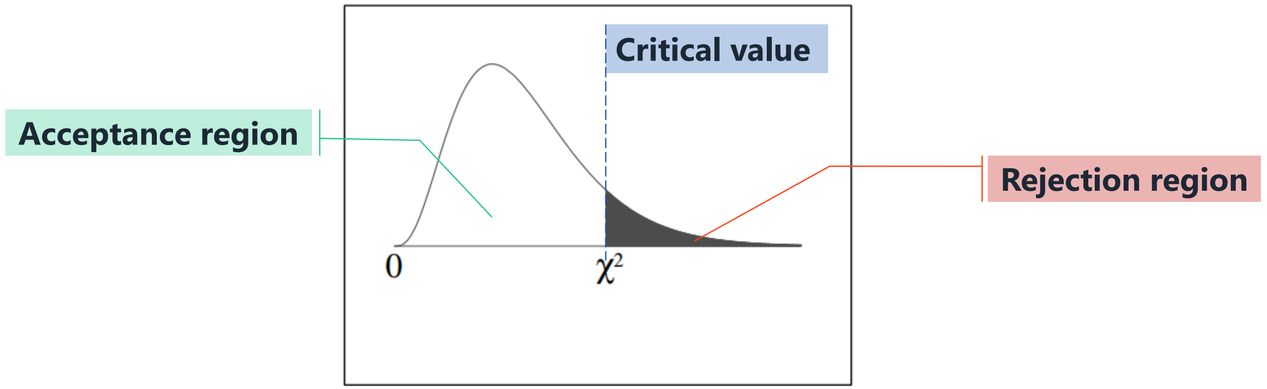
1. Critical Value/Rejection Region
2. Tails of tests
3. Central Limit Theorem
4. t-Test & F-test
5. Correlation

PROCESS FOR HYPOTHESIS TESTING

Step 1	Specify H_0, H_1, and an acceptable level of α (significance level)
Step 2	Define a sample-based test statistic and the rejection region for the specified H_0
Step 3	Collect the sample data and calculate the test statistic
Step 4	Decide to either reject or fail to reject H_0
Step 5	Interpret the results/make recommendation for action

Rejection Region/Critical Value

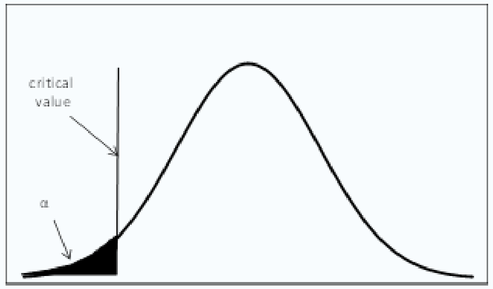
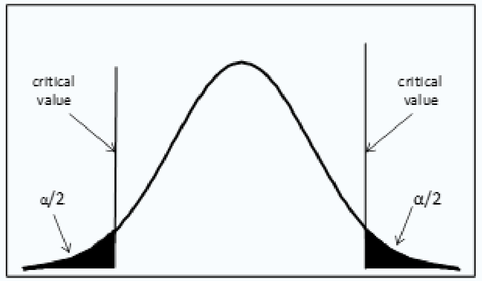
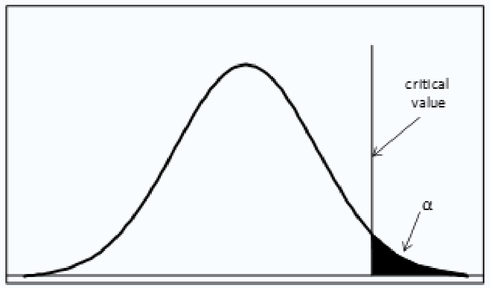
Chi-Square Distribution Table



The shaded area is equal to α for $\chi^2 = \chi^2_\alpha$.

df	$\chi^2_{.995}$	$\chi^2_{.990}$	$\chi^2_{.975}$	$\chi^2_{.950}$	$\chi^2_{.900}$	$\chi^2_{.100}$	$\chi^2_{.050}$	$\chi^2_{.025}$	$\chi^2_{.010}$	$\chi^2_{.005}$
1	0.000	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278

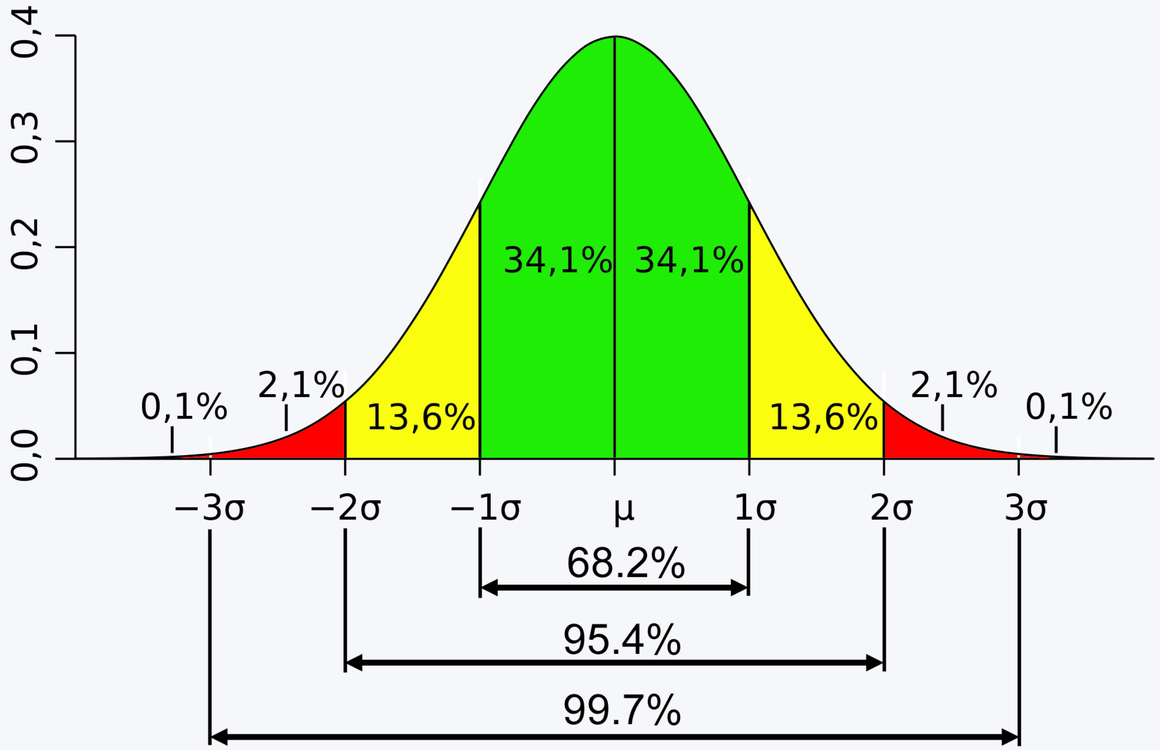
Test: 1 tailed vs 2 tailed

One Tailed Test (Left Tail)	Two-Tailed Test	One Tailed Test (Right Tail)
$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$
When population parameter is believed to be **lower** than the assumed one.	It determines whether the sample tested falls **within or outside** a certain range of values.	When population parameter is believed to be **higher** than the assumed one.
Reject H_0 if **test statistic < critical value**	Reject H_0 if **test statistic < critical value or test statistic > critical value**	Reject H_0 if **test statistic > critical value**
		

CENTRAL LIMIT THEOREM

Core Idea

Sample mean follows a normal distribution



$F - Test$

F-Test is used to **compare variance** of two groups and check if they are different from each other.

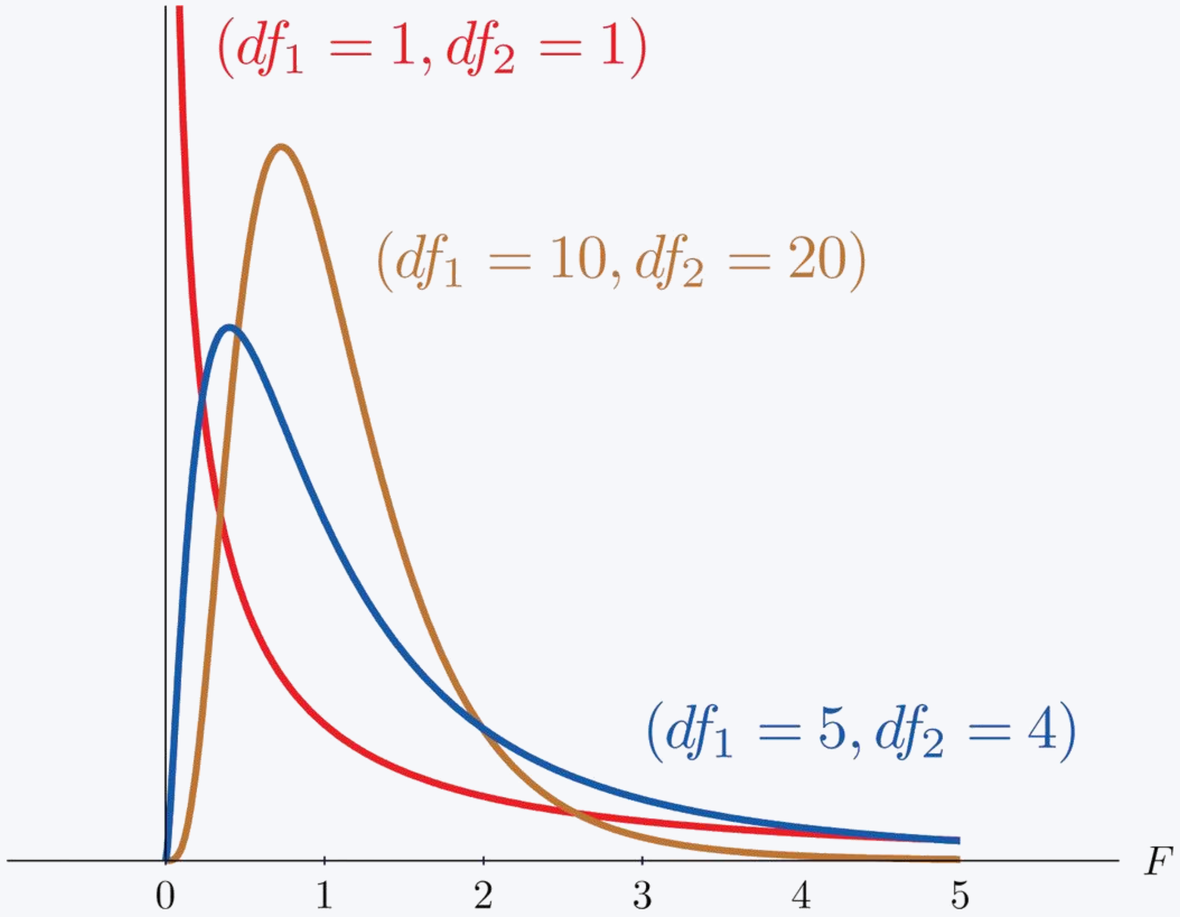
$$H_0: ratio\ of\ variance = 1$$
$$H_1: ratio\ of\ variance\ \neq 1$$

Assumption: data is normally distributed

Problem: very weak if data deviates from a normal distribution

Recommendation: use non-parametric tests (eg. Levine's Test)

F-Distribution



F Table for diff. alphas: http://www.socr.ucla.edu/Applets.dir/F_Table.html

$t - Test$

t-Test is used to **compare means** of two groups and check if they are different from each other.

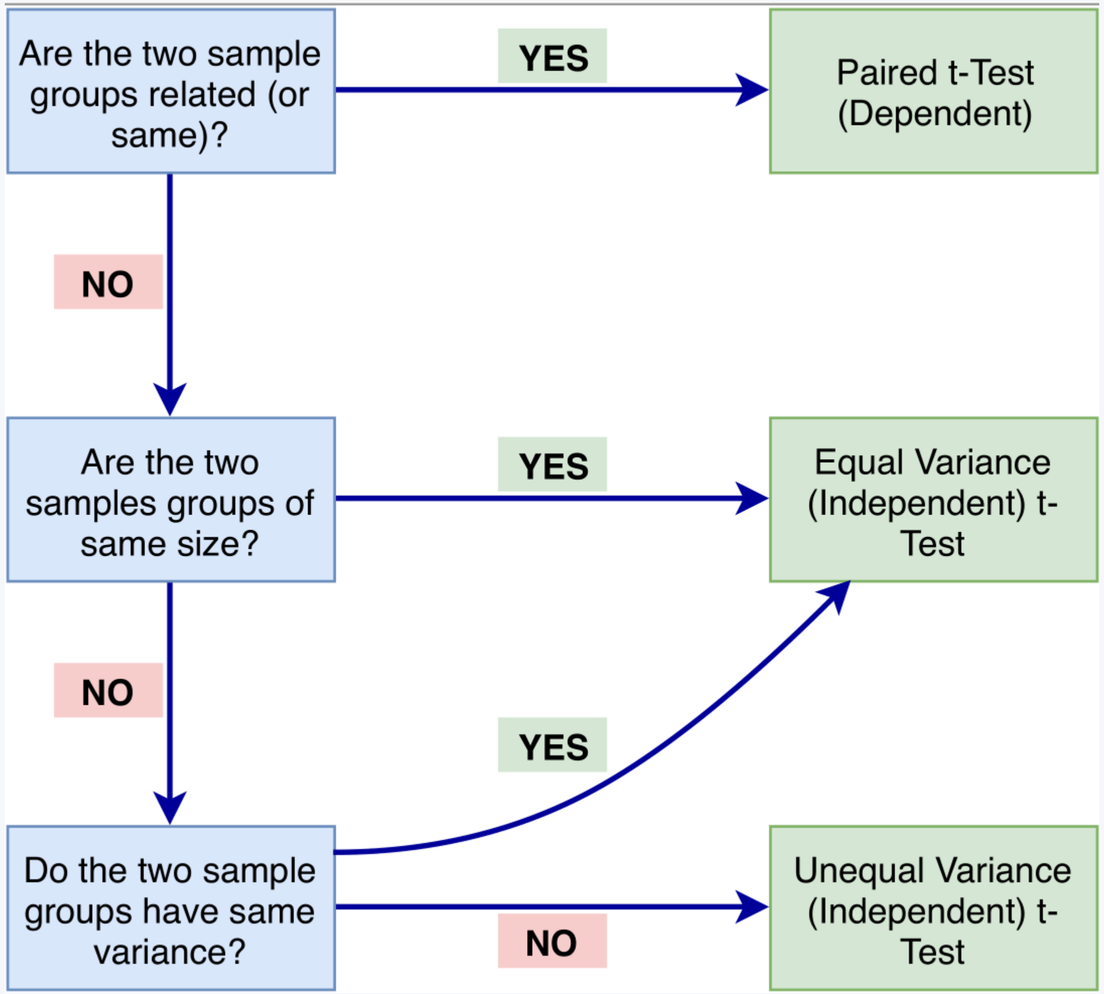
Assumption: data from both groups are sampled randomly from a normal distribution and have similar variance

Works well with smaller sample size (sweet spot: n = 20 to 30)

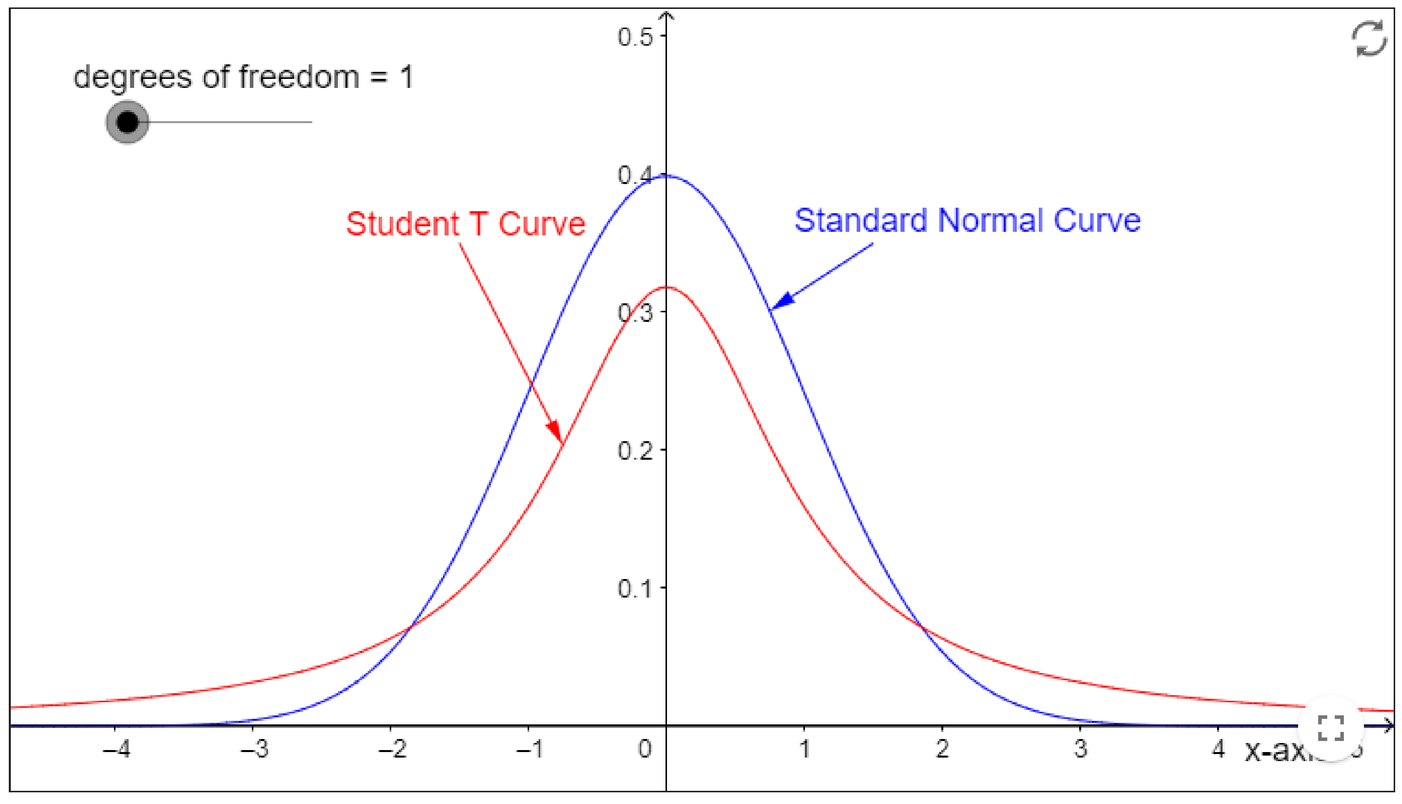
Test statistic: $t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$

More on t-Test: https://www.youtube.com/watch?v=pTmLQvMM-1M

$t - Test$ Decisions



t Distribution



t vs z distribution: https://www.geogebra.org/m/zxUzwgkH

R

1. Download the R file and Data File uploaded today
2. Save the files into one folder.
3. Open the R file.

PLAN FOR NEXT WEEK

That’s it for today! :-)

Next week, we are going to discuss:

1. A few more tests (ANOVA …)
2. Correlation and Regression

If you want to reach me, mail me at:

`prabesh.dhakal@stud.leuphana.de`