

# Statistics Cheat-Sheet

Abdullah Al Shamim

## T-test

Data:

- ✓ A numeric variable: এই ধরনের ডেটা যেগুলো আমরা সংখ্যা দ্বারা প্রকাশ করতে পারি, যেমন ওজন।
- ✓ A categorical variable: এই ধরনের ডেটা যেগুলো আমরা শ্রেণীবিভাগ দ্বারা প্রকাশ করতে পারি, যেমন লিঙ্গ (পুরুষ এবং নারী)।

Observation from the data:

- ✓ তুলনাঃ পুরুষের গড় ওজন নারীদের গড় ওজন অপেক্ষা বেশী।

Research question:

- ✓ এই ধরনের ব্যবধানের কারণ শুধুই কি বাহ্যিক পরিবর্তন (সংখ্যাগত পরিবর্তন)-এর জন্যে নাকি তাদের মধ্যে আসলেই তাৎপর্যপূর্ণ ব্যবধান আছে?

Start with a null hypothesis:

- ✓ ধরি, এই গড় ওজন প্রতিটি লিঙ্গের জন্যে সমান।

Perform a Student's T-test

- ✓ যদি p value 0.05 এর চেয়ে কম হয়, তাহলে আমরা null hypothesis কে অগ্রাহ্য করতে পারি। এবং এই সিদ্ধান্তে উপনীত হতে পারি, এই গড় ওজন প্রতিটি লিঙ্গের জন্যে আসলেই তাৎপর্যপূর্ণ।



## Chi Squared test

Data:

- ✓ Two categorical variables: এই ধরনের ডেটা যেগুলো আমরা শ্রেণীবিভাগ দ্বারা প্রকাশ করতে পারি, যেমন লিঙ্গ (পুরুষ এবং নারী) এবং উচ্চতার শ্রেণীবিভাগ (খাটো, মধ্যম এবং লম্বা)।

Observation from the data:

- ✓ আমরা যদি উচ্চতার ভিত্তিতে শ্রেণীবিভাগ করি তাহলে দেখবো বিভিন্ন শ্রেণীর নারী এবং পুরুষ। আর এই গ্রুপের মধ্যেই যদি আমরা আলাদাভাবে নারী এবং পুরুষের শ্রেণীবিভাগ করি তাহলে খাটো, মধ্যম এবং লম্বা নারী-পুরুষের ভিত্তিতে পুরোপুরি ভিন্ন শ্রেণীবিভাগ পাবো।

Research question:



- ✓ এই ধরনের ব্যবধানের কারণ শুধুই কি বাহ্যিক পরিবর্তন (সংখ্যাগত পরিবর্তন)-এর জন্যে নাকি তাদের মধ্যে আসলেই তাৎপর্যপূর্ণ ব্যবধান আছে?

Start with a null hypothesis:

- ✓ ধরি, লিঙ্গ এবং উচ্চতার শ্রেণীবিভাগের মাঝে কোন সম্পর্ক নেই।

Perform a Chi-test

- ✓ যদি p value 0.05 এর চেয়ে কম হয়, তাহলে আমরা null hypothesis কে অগ্রাহ্য করতে পারি। এবং এই সিদ্ধান্তে উপনীত হতে পারি, লিঙ্গ এবং উচ্চতার শ্রেণীবিভাগের মাঝে সম্পর্ক আসলেই তাৎপর্যপূর্ণ।

	 Female	 Male
Without graduation	6	7
College	13	16
Bachelor's degree	16	15
Master's degree	8	11
Total	43	49

Is there a relationship between gender and the highest level of education?

↓

Chi<sup>2</sup> - Test

## ANOVA

Data:

- ✓ A numeric variable: ঐ ধরনের ডেটা যেগুলো আমরা সংখ্যা দ্বারা প্রকাশ করতে পারি, যেমন ওজন।
- ✓ A categorical variable with 3 or more categories: ঐ ধরনের ডেটা যেগুলো আমরা শ্রেণীবিভাগ দ্বারা প্রকাশ করতে পারি, যেমন জাতীয়তা (আরবীয়, বাংলাদেশী, রোমান)

Observation from the data:

- ✓ এই গড় ওজন প্রতিটি জাতীয়তার জন্য সমান নয়।

Research question:

- ✓ এই ধরনের ব্যবধানের কারণ শুধুই কি বাহ্যিক পরিবর্তন (সংখ্যাগত পরিবর্তন)-এর জন্যে নাকি তাদের মধ্যে আসলেই তাৎপর্যপূর্ণ ব্যবধান আছে?

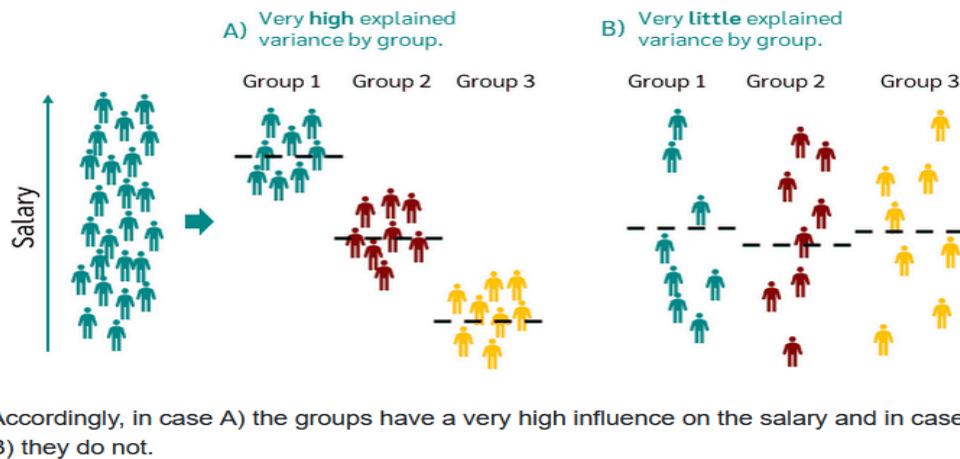
Start with a null hypothesis:

- ✓ ধরি, এই গড় ওজন প্রতিটি জাতীয়তার জন্য সমান।

Perform an ANOVA

- ✓ যদি p value 0.05 এর চেয়ে কম হয়, তাহলে আমরা null hypothesis কে অগ্রাহ্য করতে পারি। এবং এই সিদ্ধান্তে উপনীত হতে পারি, গড় ওজন প্রতিটি জাতীয়তার জন্য ভিন্ন।

- ✓ এই পরীক্ষা দ্বারা পুরোপুরি বলা যায় না কোন জাতীয়তার মাঝে গড় ওজনের পার্থক্য আছে, এই পরীক্ষা শুধু জানায় তাদের গড় ওজন একই নয়। এর জন্যে পরবর্তীতে TukeyHSD test করতে হয়।



## Correlation test

Data:

- ✓ Two numeric variables: ঐ ধরনের ডেটা যেগুলো আমরা সংখ্যা দ্বারা প্রকাশ করতে পারি, যেমন ওজন এবং বয়স।

Observation from the data:

- ✓ তাদের মধ্যে (ওজন এবং বয়স) তাৎপর্যপূর্ণ সম্পর্ক রয়েছে। অর্থাৎ যেকোন একটি variable এর মান পরিবর্তন করা হলে অন্য variable টির মান স্বতস্ফূর্তভাবে পরিবর্তিত হয় (বাড়ে অথবা কমে)।

Research question:

- ✓ এই দুইটি variable এর মাঝে সম্পর্কের কারণ শুধুই কি বাহ্যিক পরিবর্তন (সংখ্যাগত পরিবর্তন)-এর জন্যে নাকি তাদের মধ্যে আসলেই তাৎপর্যপূর্ণ সম্পর্ক আছে?

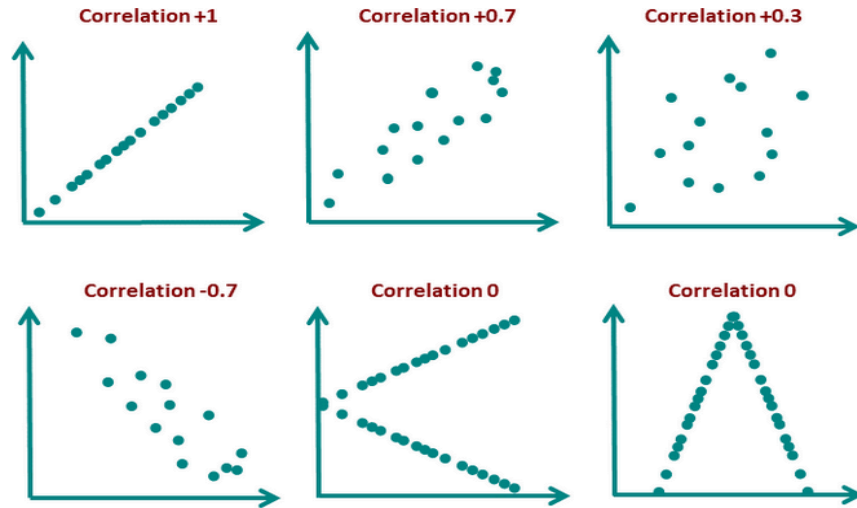
Start with a null hypothesis:

- ✓ ধরি এই দুইটি variable এর মাঝে কোন ধরনের সম্পর্ক নেই।

Perform a correlation-test

- ✓ যদি p value 0.05 এর চেয়ে কম হয়, তাহলে আমরা null hypothesis কে অগ্রাহ্য করতে পারি। এবং এই সিদ্ধান্তে উপনীত হতে পারি, এই দুইটি variable এর মাঝে সম্পর্ক বিদ্যমান।

- ✓ এই টেস্ট আপনাকে correlation coefficient ভ্যালুর মান জানাবে, যা কিনা এই দুইটি variable এর মাঝে কোন ধরনের সম্পর্ক (সম্পর্কের প্রকৃতি) বিদ্যমান তা জানায়।



## **Social Media Links:**

Website: [www.premieranalytics.com.bd](http://www.premieranalytics.com.bd)

YouTube: <https://www.youtube.com/@ShamimTheAnalyst>

Facebook: <https://www.facebook.com/ShamimTheAnalystOfficial>

Facebook Group:

<https://www.facebook.com/groups/datascienceclubbangladesh>

Instagram: <https://www.instagram.com/shamimtheanalyst/>

LinkedIn: <https://www.linkedin.com/in/abdullah-al-shamim-863149328/>

GitHub: <https://github.com/ShamimtheAnalyst/YouTube-Repo>

X: <https://x.com/ShamimTheAnalyst>

Kaggle: <https://www.kaggle.com/abdullahalshamim>