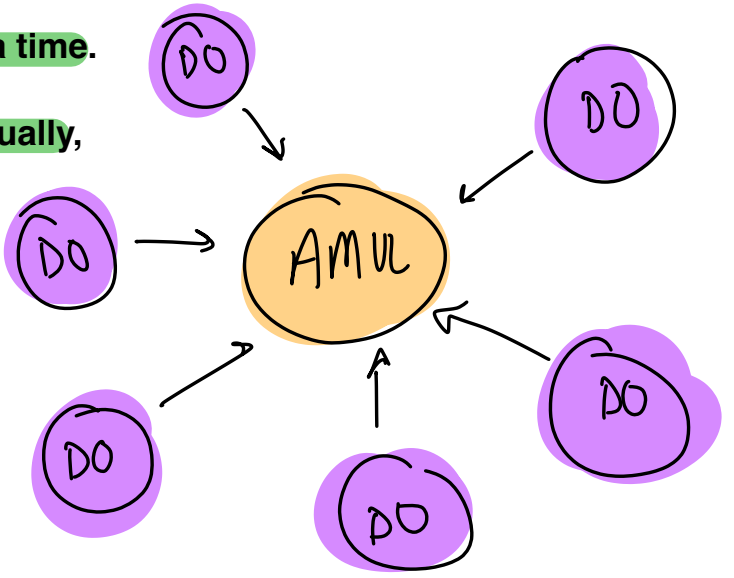
Pooled Milk test

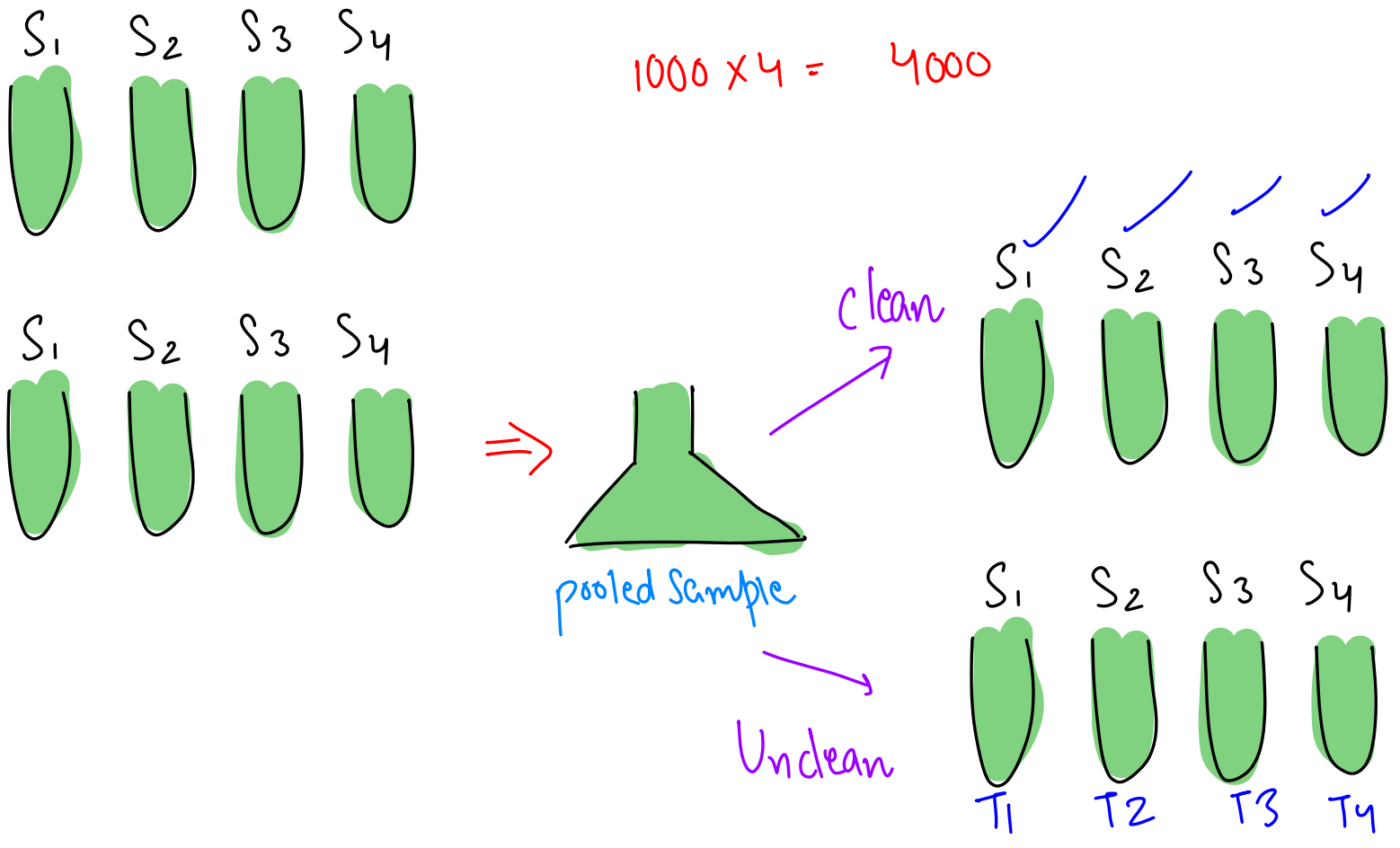
A milk bank tests pooled sample tests of 4 people at a time.
If clean, the bank stores all of them.
if Unacceptable, then all 4 samples are tested individually,
The probability of any sample being dirty is 0.1.
Find the expected number of tests.

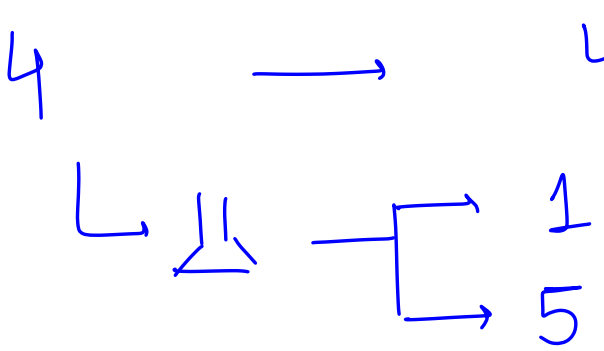


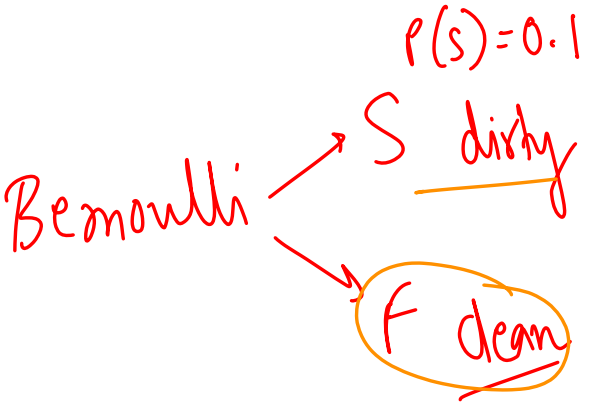
Rs 1000

40000 DOs → 10 Samples → 40000 Samples

Per day Amul is spending ⇒ 40000 × 1000 = 4Cr


S1 S2 S3 S4
S1 S2 S3 S4
1000 x 4 = 4000
clean
S1 S2 S3 S4
pooled Sample
Unclean
S1 S2 S3 S4
T1 T2 T3 T4





X = No. of tests I need to perform.

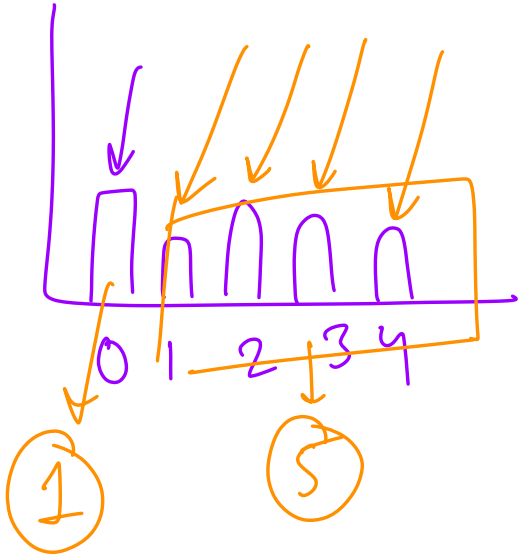
X	P(X)	P(X)
1	Binom.pmf(n=4, k=0, p=0.1)	0.6561
5	1 − Binom.pmf(n=4, k=0, p=0.1)	0.3439

Binom.pmf $(n=4, k=1, p=0.1)$ + Binom.pmf $(n=4, k=2, p=0.1)$
+ Binom $(n=4, k=3, p=0.1)$ + Binom.pmf $(n=4, k=4, p=0.1)$

$$E[X] = \sum X \, P(X)$$

$$= 1 \times 0.6561 + 5 \times 0.3439$$

$$= 2.3756 \text{ tests per 4 samples}$$



4 Samples → 4 tests → Rs 4000

4 Samples → 2.375 tests → Rs 2375

In one day 40000 test → 4 Cr

2375 test → 2.375 Cr. } 1.62 Cr

1 year Amul is saving ≈ 591 Crs per year.

Bernoulli → Success Clean 0.9
Failure dirty 0.1

X	P(X)
1	Binom. pmf $(n=4, k=4, p=0.9)$
5	Binom. pmf $(n=4, k=0, p=0.9)$ + Binom pmf $(n=4, k=1, p=0.9)$ + Binom.pmf $(n=4, k=2, p=0.9)$ + Binom.pmf $(n=4, k=3, p=0.9)$

$\Rightarrow$ Binom. cdf $(n=4, k=3, p=0.9)$

$\Rightarrow$ $1 -$ Binom. pmf $(n=4, k=4, p=0.9)$

Module owners of current (Fundamentals) module wants to know in the initial class that the learners have a good grasp of previous module.

6 Learners were chosen and were given P&S test. MO wants the class to score above 70 on the test for a successful module completion.

The 6 learners get the following scores : 62,92,75,68,83,95.

Can current MO have 90% C.I. that the mean score for the class on the test would be above 70?

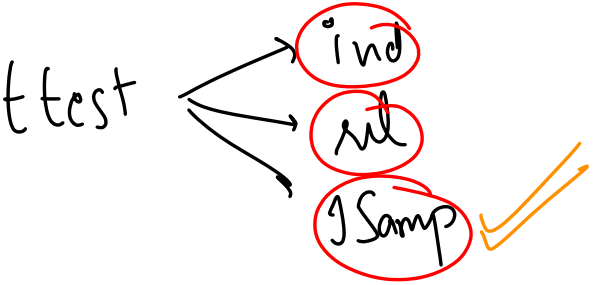
$H_0: \mu = 70$

$H_a: \mu > 70$

Confidence = 90%

$\alpha = 0.1$

$p = 0.07$



"Right tailed test"

$p < \alpha$

"Reject H_0"

Avg. IQ of the population is 100 with std. of 15. A team of scientists want to test the new medication to see if it has either a positive or negative effect or no affect at all. A sample of 30 participants who have taken the medicines were found to have avg. IQ 140. Did the medicine affect intelligence?

H_0: Not effect

H_a: Affect intelligence

"Right tailed"

$\mu = 100$

$\sigma = 15$

$n = 30$

$\bar{x} = 140$

Z test

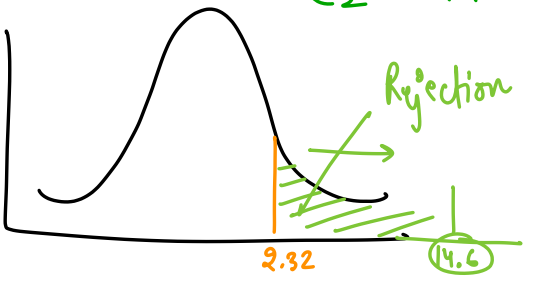
pvalue = 0

$Z = \frac{140 - 100}{15/\sqrt{30}} = 14.6$

"Reject H_0"

$\alpha = 0.01$

CI = 99%



Amazon Fulfillment centre

Amazon is trying to experiment 2 different machines that will pack the deliveries. It is supposed that the new machine will package on an average quicker in comparison to the old one. To test the hypothesis, the time it takes to pack the deliveries are recorded.
Do the data provide sufficient evidence to conclude that , on average the new machine is better than the old one?

```
new_machine = np.array([42.1,41,41.3,41.8,42.4,42.8,43.2,42.3,41.8,42.7])
old_machine = np.array([42.7,43.6,43.8,43.3,42.5,43.5,43.1,41.7,44,44.1])
```

LEFT

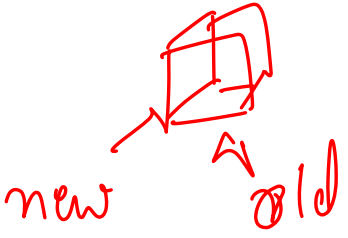
TEST

H_0 new machine = old machine | $\mu_N = \mu_0$ | ttest_rel.

H_a : new > old | $\mu_N < \mu_0$ |

CI = 99%

$\alpha = 0.01$



pvalue = 0.006

pvalue < α => "Reject H_0"

A gym is claiming to help you lose weight and they take help from an independent consultant to prove their claim. The consultant takes in sample of 15 people, weighing each of them before the gym begins and 3 months later. Test their Claim.

H_0 : $\mu_A = \mu_B$

H_a : $\mu_A < \mu_B$

Confidence = 95%

$\alpha = 0.05$

Valid

ttest_rel (After , Before, alternative = "less")

pvalue $< \alpha$ "Reject H_0"

A study claims that children learns most effectively with a constant background sound in comparison to unpredictable sound or no sound at all. Scientists divided 24 children into groups of 8 each. All the children are asked to learn a passage for 3 days under 3 conditions. After 3 days, they were asked to take a test and their scores are calculated.

$H_0: \mu_1 = \mu_2 = \mu_3$

$H_a: \mu_1 \neq \mu_2 \neq \mu_3$ (At least one of them is different)

$p = 0.04 \qquad \alpha = 0.05$

$p < 0.05 \rightarrow$ "Reject"

A random sample of 500 Adult Indians are questioned about their opinions and afflictions to political parties on the new bill passed in Rajya Sabha. We need to test whether their affiliation and their opinion are dependant at 5% Alpha.

	Favour	Indifferent	Opposed	Total
BJP	138	83	64	285
INC	64	67	84	215
Total	200	150	148	500

H_0 : Independant

H_a : dependent

pvalue: 1.54×10^{-5}

$\alpha = 0.05$

$p < \alpha \rightarrow$ "Reject H_0"

dependent.

Twenty-five men between the ages of 25 and 30, who were participating in a well-known heart study carried out in Framingham, Massachusetts, were randomly selected. Of these, 11 were smokers and 14 were not. The following data refer to readings of their systolic blood pressure.
Use these data to test the hypothesis that the mean blood pressures of smokers and nonsmokers are the same.

```
non_smokers= [130, 122, 128, 129, 118, 122, 116, 127, 135, 120, 122, 120,
smokers= [124, 134, 136, 125, 133, 127, 135, 131, 133, 125, 118]
```

$H_0: \mu_S = \mu_{NS}$ 2 tailed test

$H_a: \mu_S \neq \mu_{NS}$ T-test-ind.

$p = 0.018$ $\alpha = 0.05$

$p < \alpha$ "Reject H_0" →

In a test of the ability of a certain polymer to remove toxic wastes from water, experiments were conducted at three different temperatures. The data below give the percentages of the impurities that were removed by the polymer in 21 independent attempts.

```
low_temp=[42,41,37,29,35,40,32]
mid_temp=[36,35,32,38,39,42,34]
high_temp=[33,44,40,36,44,37,45]
```

$H_0: \quad \mu_1 = \mu_2 = \mu_3$

$H_a: \quad \mu_1 \neq \mu_2 \neq \mu_3$ (At least one of them is diff.)

Q) Recommender System

When a customer buys a T-Shirt , a recommender algorithm suggests a few related items. The recommender system in production(legacy) that has a success rate of 10%. You and your team have developed a new DL algorithm for recommendation. It is tested before deploying. Of the next 500 customers, 72 bought items recommended by the new model. Is the improvement brought by the new model is statistically significant at 95% confidence?

H_0 : old = new

H_a : old < new

"Right"

Binomial

Bernoulli => S 0.1, F 0.9

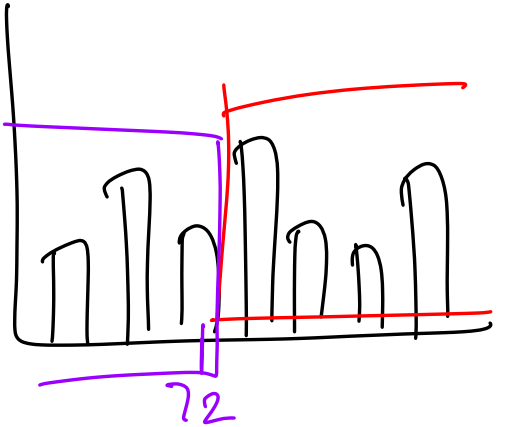
n = 500

k = 72

p = 0.1

$P[X \geq 72]$ pvalue

1 - binom.cdf(n=500, k=71, p=0.1)



Q) A sample of 300 cars having cellular phones and 400 cars without phones were tracked for 1 year. The following table gives the number of these cars involved in accidents over that year.
Use the above to test the hypothesis that having a cellular phone in your car and being involved in an accident are independent. Use the 5 percent level of significance.

	Accident	No accident
Cellular phone	22	278
No phone	26	374

H_0 : ind

H_a : dep

$\alpha = 0.05$

$p = 0.77$

$\therefore$ fail to reject H_0

Numerical → 2 Categories → tTest → rel
→ ind
→ 1 samp

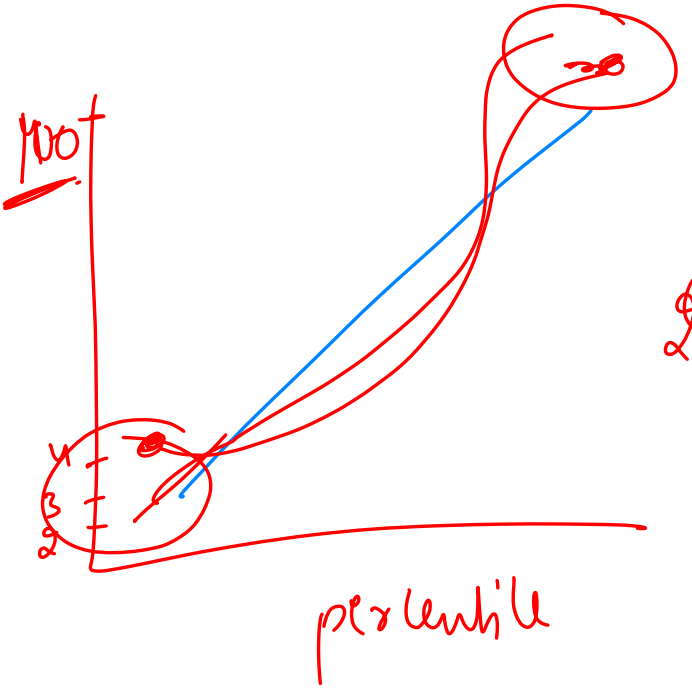
Numerical - Numerical → Correlation

Categorical - Categorical → Chisquare

↳ Goodness of fit

dependant
Indipend

2 Categorical → Test of Independance -



2, 4, 3, 100

100

25TH 50TH 75TH 100

2 3 4 100

Ttest 2nd

(2) Numerical 2 Categories

Numerical value } μ_1, μ_2 } Non Smoker / Smoker.

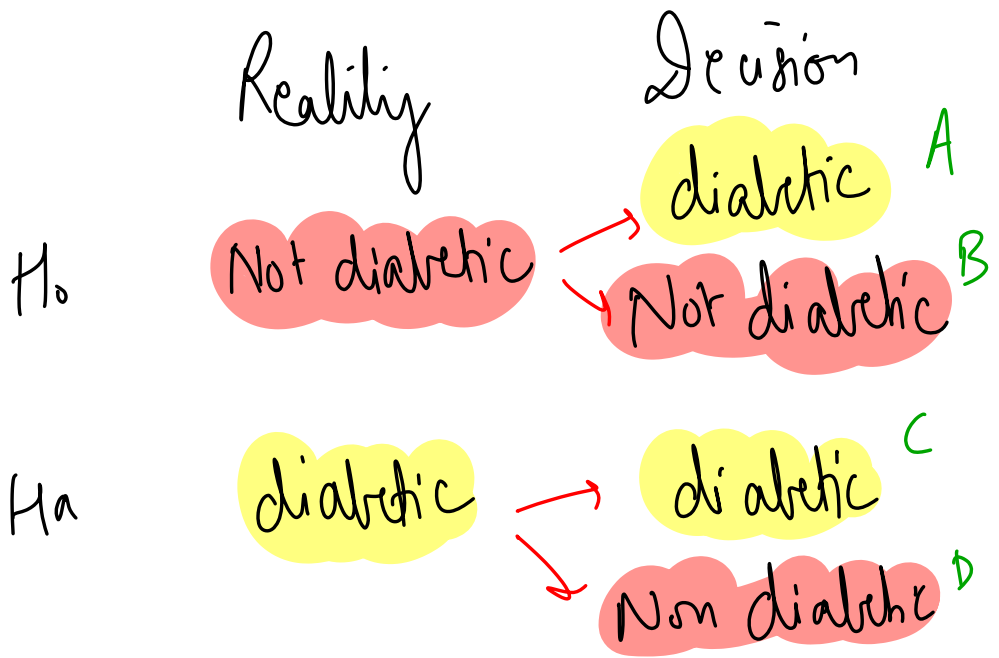
Chisquare Test of Independa

(2) Categorical Column mentioning

freq

Numerical	Categorical
Mean	Mode
Median	"Unknown"
-1	"Other"
0	

① Diabetic (+ve)
Non diabetic (-ve)



(α) Type I error

FALSE POSITIVE

TRUE NEGATIVE

TRUE POSITIVE

FALSE NEGATIVE

(β) Type II error.