Hypothesis Testing

2

- Z-test V/s t-test ; T-distribution
- 2 sample t-test (AB testing)
- Python - coding implementation of 2 sample t-test

Test statistic for Sample mean:

- Mean weight of packet → 450 gm → μ
- std. dev. → 3 gm → σ

} Population mean / std. dev. (from historical data)

⇒ test statistic: Z score

- n : Sample Size
- $\bar{x}$: sample mean

$\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ $\quad Z = \left(\frac{\bar{x} - \mu}{\sigma/\sqrt{n}}\right)$ Normal distribution

- In hypothesis testing, we often don't have population standard deviation → σ.
- What to do then?

- from 'n' samples

$\bar{x}$ → sample mean

s → sample std. dev.

- If n is large enough; usually $n > 30$

we can safely assume

$$Z = \left(\frac{\bar{x} - \mu}{s/\sqrt{n}} \right)$$

$Z \sim$ Normal distribution

• If $n < 30$;

test statistic $\rightarrow \left(\dfrac{\bar{x} - \mu}{s/\sqrt{n}} \right)$ follows T-distribution

$$\Rightarrow \quad t = \left(\frac{\bar{x} - \mu}{s/\sqrt{n}} \right) \Rightarrow t \sim \text{T-distribution}$$

$\downarrow$

We use T-distribution CDF to find out p-value.

Summary: When to use what.

σ : population std. deviation
s : sample std. dev.

- 2 possibilities (σ)
  - we know σ → z-test
  - we don't know σ
    - $n < 30$ → (t-test)
    - $n > 30$ → (t-test / z-test)

Only one parameter → t-distribution. degree of freedom (dof)

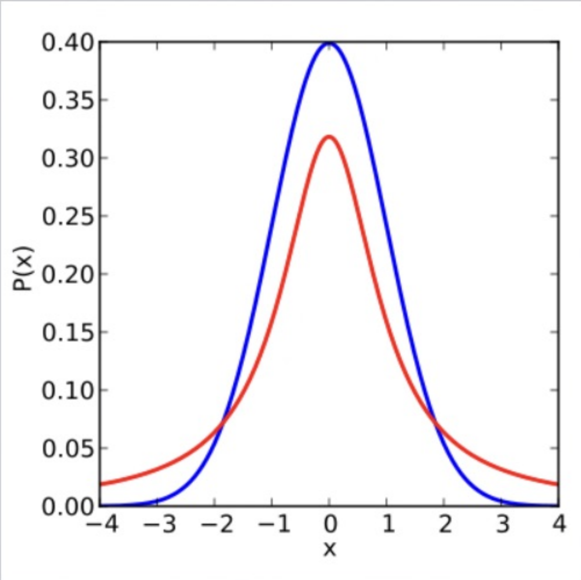
⇒ sample size = n

↳ test statistic

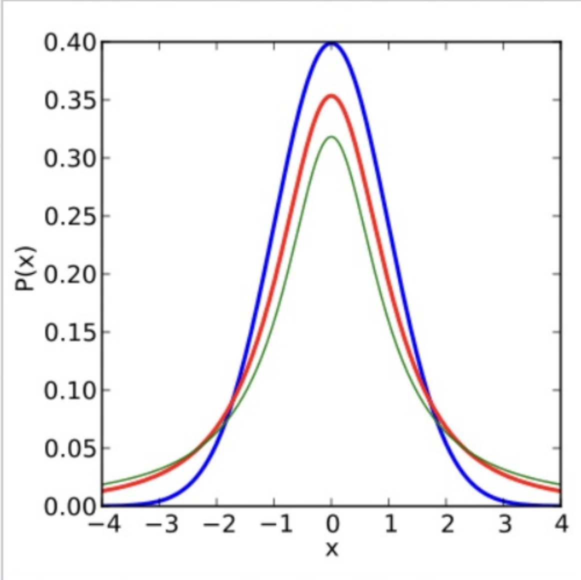
$t \sim T[\text{dof} = (n-1)]$

⇒ as $n \uparrow$ T-distribution becomes more and more Gaussian (Normal)

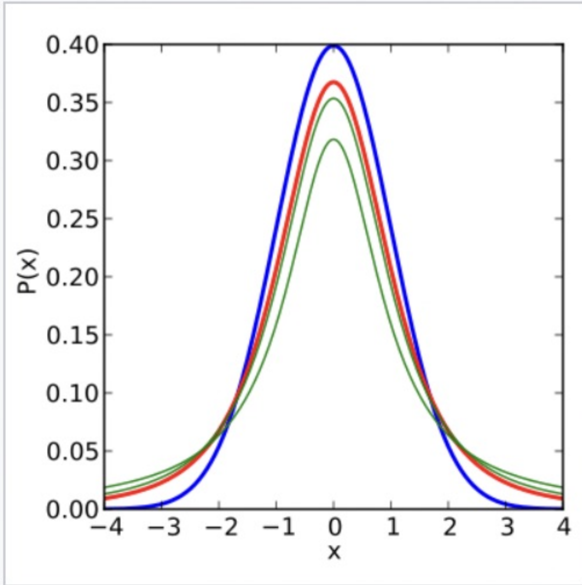
Density of the *t*-distribution (red) for 1, 2, 3, 5, 10, and 30 degrees of freedom compared to the standard normal distribution (blue).
Previous plots shown in green.



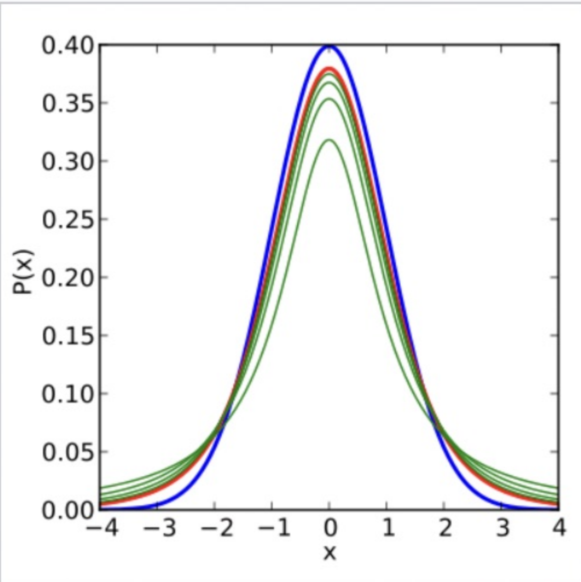
1 degree of freedom



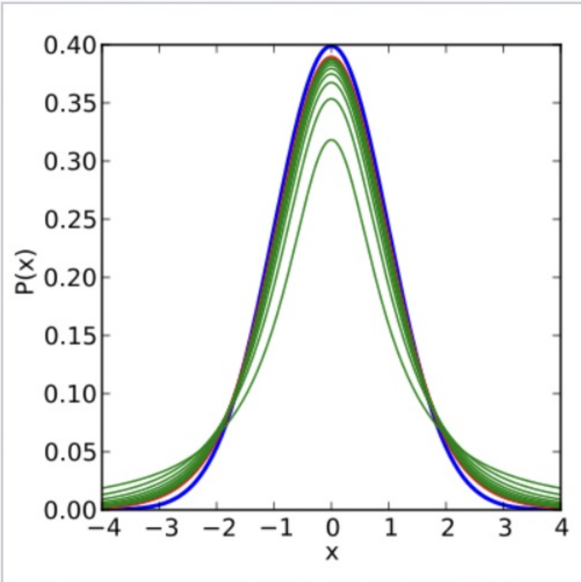
2 degrees of freedom



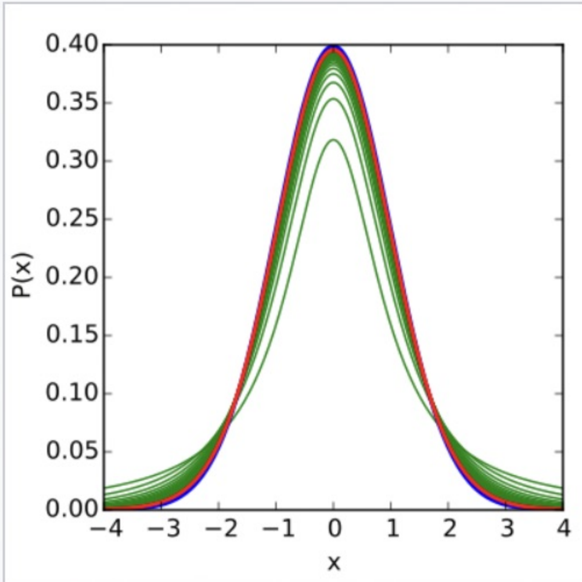
3 degrees of freedom



5 degrees of freedom



10 degrees of freedom



30 degrees of freedom

One Sample t-test

A public health official claims that the mean home water use is 350 gallons a day.
To verify this claim, a study of 20 randomly selected homes was instigated with the result that the average daily water uses of these 20 homes were as follows:

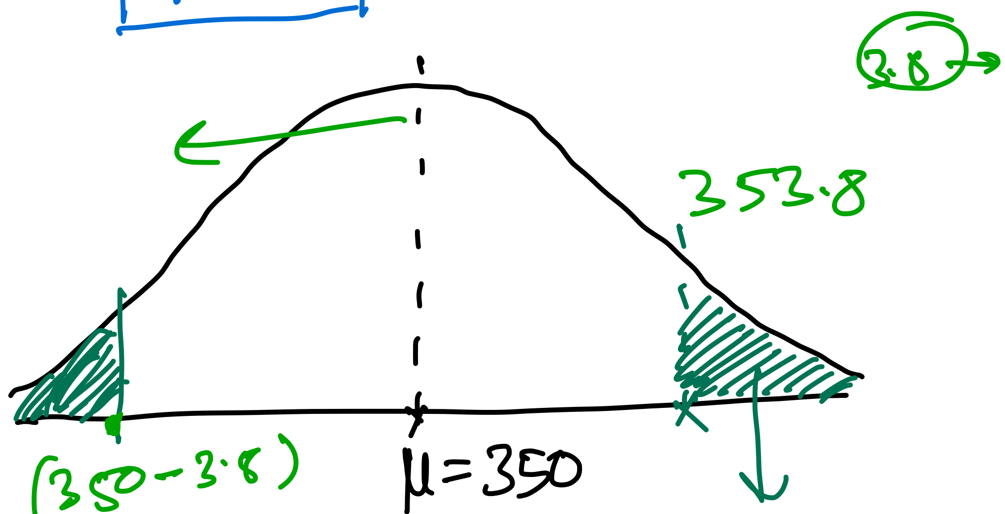
usage = [340, 344, 362, 375, 356, 386, 354, 364, 332, 402, 340, 355, 362, 322, 372, 324, 318, 360, 338, 370]

- $H_0 : \mu = 350$
  $H_A : \mu \neq 350$ → Two tailed

$\boxed{n = 20}$ $\alpha = 0.05$

m → sample mean ; s → sample std. dev.

$m = 353.8$



- test statistic

$$t = \left(\frac{m - \mu}{s/\sqrt{n}} \right)$$

$$\Rightarrow \boxed{t = 0.798}$$

Since 2 tailed

```
from scipy.stats import t
p_value = 2*(1 - t.cdf(t_stat, df = n-1))
p_value
```

t_stat → 0.798
n-1 → df = 19

p_value = 0.4347

∴ p_value > α ⇒ failed to reject H_0

Two Samples

Drug Recovery

Suppose two companies develop a drug for a disease.

Drug 1 was tested on 100 people, and the recovery days look like this

[8, 5, 9, 10, ……….., 16]

The mean recovery time was 7.1 days

Drug 2 was tested on 120 people, and the recovery days look like this

[12, 4, 7, 13, …….., 8]

The mean recovery time was 8.07 days

Can we say one drug was better than the other?

Or was this small difference a coincidence?

For such cases we use the **two-sample Z-test or two-sample T-test**

* Comparing 2 versions (A & B) $\Rightarrow$ A/B TESTING ✓

Let mu_1 be the average recovery for drug 1, and mu_2 be the average for drug 2

$H_0 : mu_1 = mu_2$

$H_a : mu1 \neq mu2$

- from data of $drug_1$ and $drug_2$

$n_1 = 100$ $\quad$ $n_2 = 120$ — number of samples

$m_1 = 7.1$ $\quad$ $m_2 = 8.07$ — sample mean

s_1 $\quad$ s_2 — sample std. dev^n.

Test statistic :

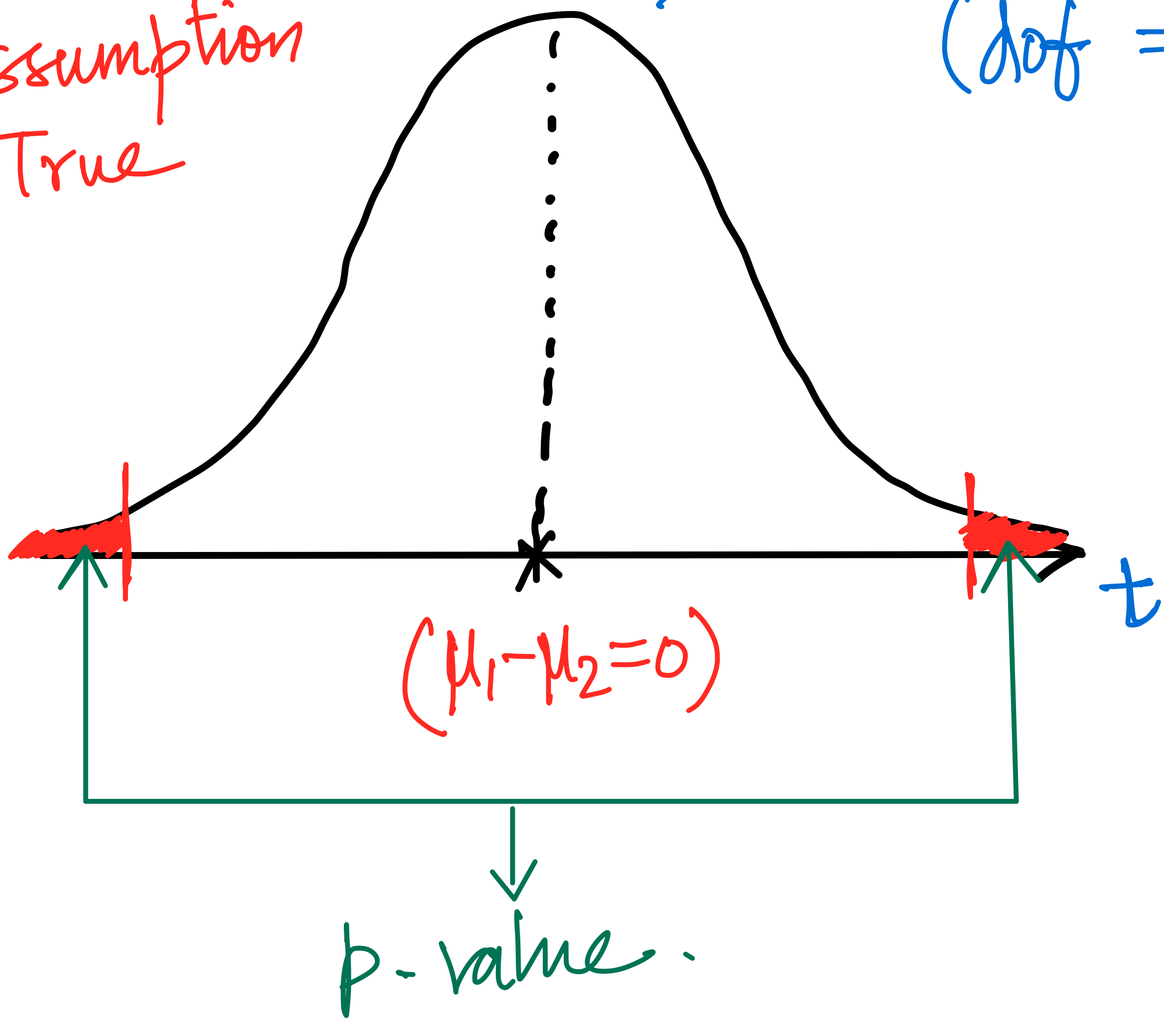
$$t = \left(\frac{(m_1 - m_2)}{\sqrt{\frac{s_1^2}{\sqrt{n_1}} + \frac{s_2^2}{\sqrt{n_2}}}} \right)$$

→ if $n_1, n_2 < 30$ use 2 sample t-test

→ if $n_1, n_2 > 30$ 2 sample t test / 2 sample z-test

Two Tailed

p_value



· under the assumption that H_0 is True $(\mu_1 = \mu_2)$

If p-value < α (let's say $\alpha = 0.05$).

we reject H_0 (No statistically significant diff. b/w mean recovery time of D_1 & D_2).

And conclude that there is statistically significant diff. b/w mean recovery time of D_1 & D_2

Else; failed to Reject H_0.

Code . .. _ . . ...