Biostatistics
Week III

Ege Ülgen, M.D.

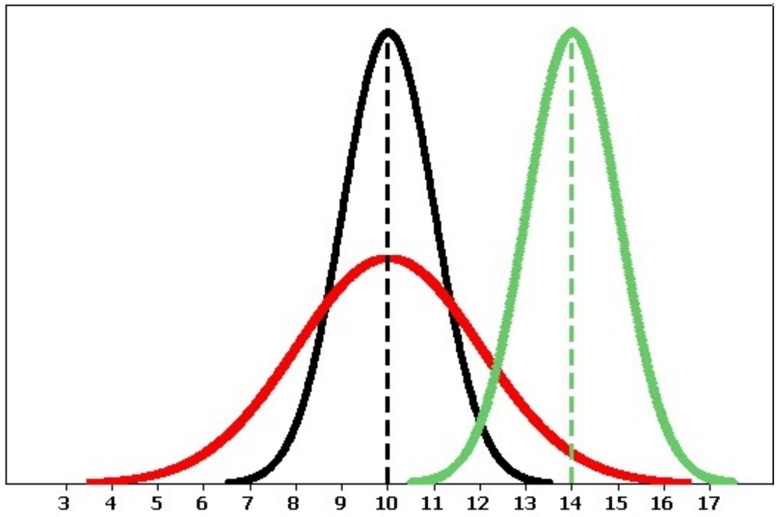
21 October 2021

ACIBADEM
MEHMET ALİ AYDINLAR
ÜNİVERSİTESİ

Normal Distribution

- The distributions of many variables follow a “normal distribution”

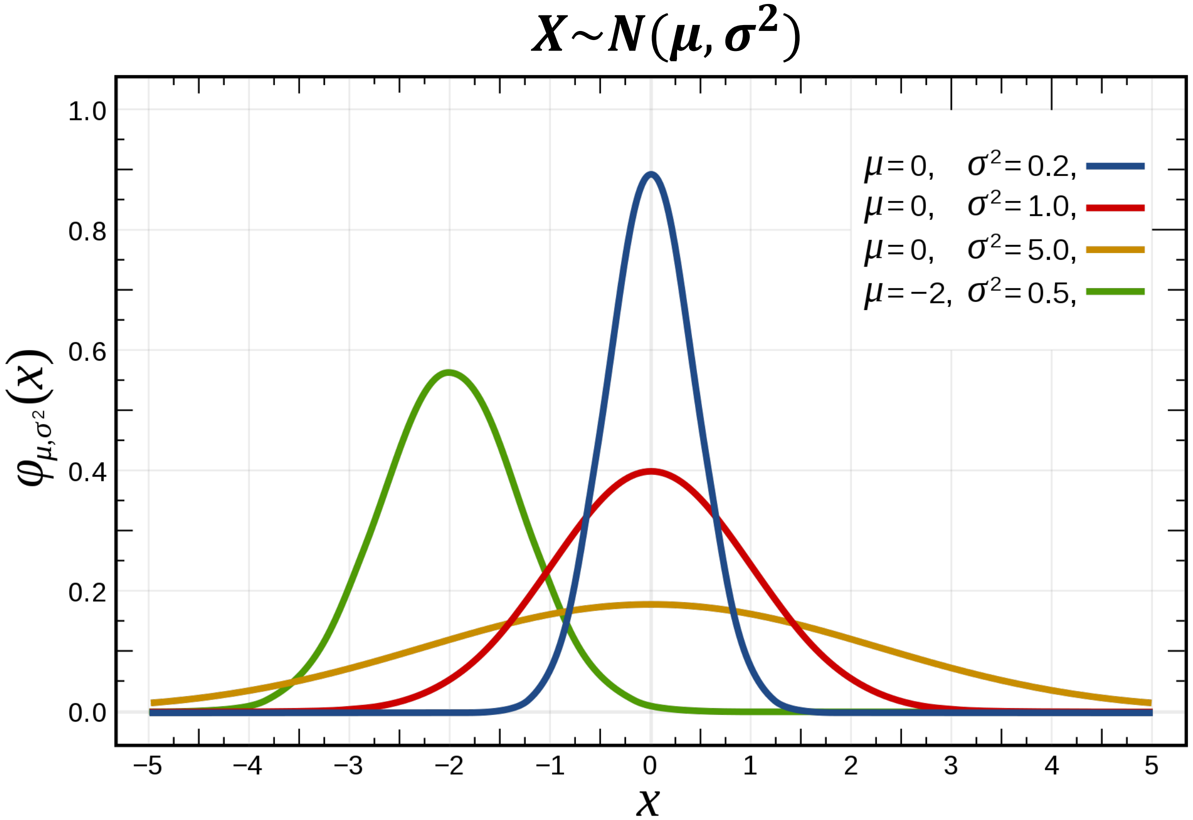
- The **bell-shape** indicates that values closer to the mean are more likely, and it becomes increasingly unlikely to take values far from the mean in either direction

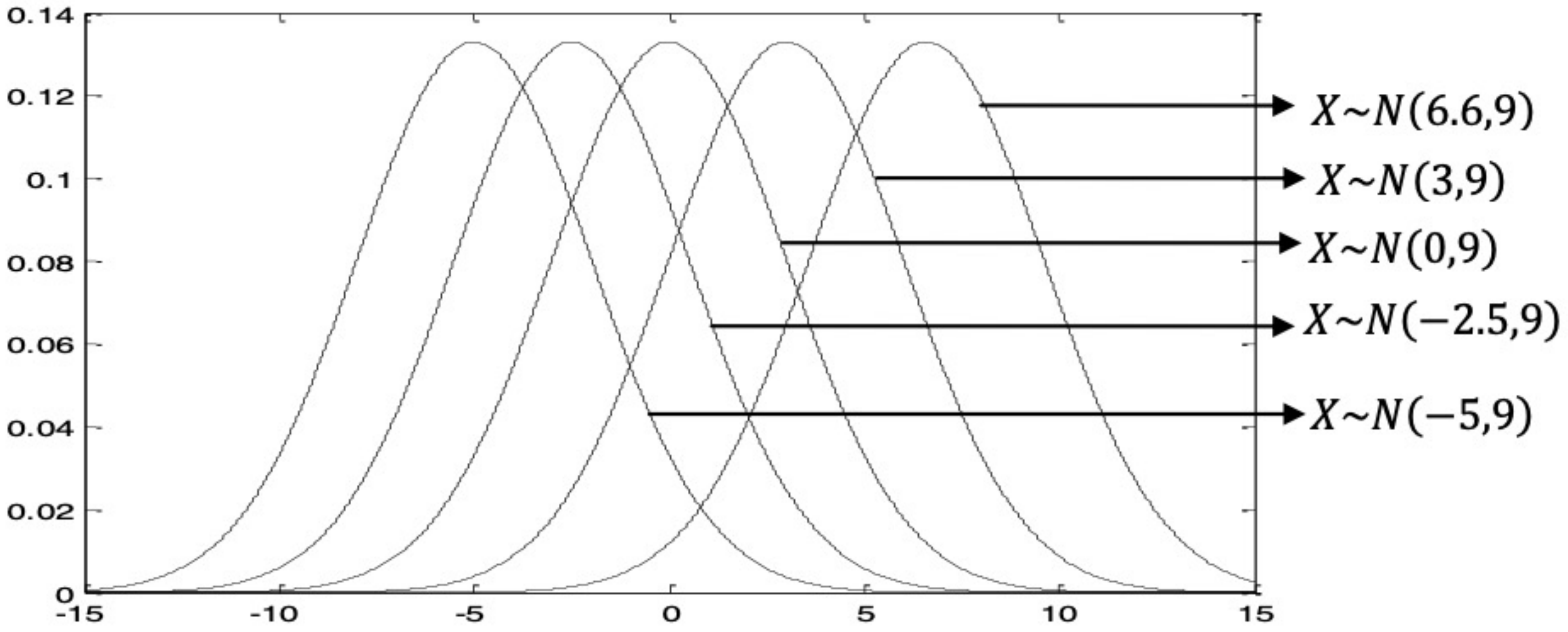


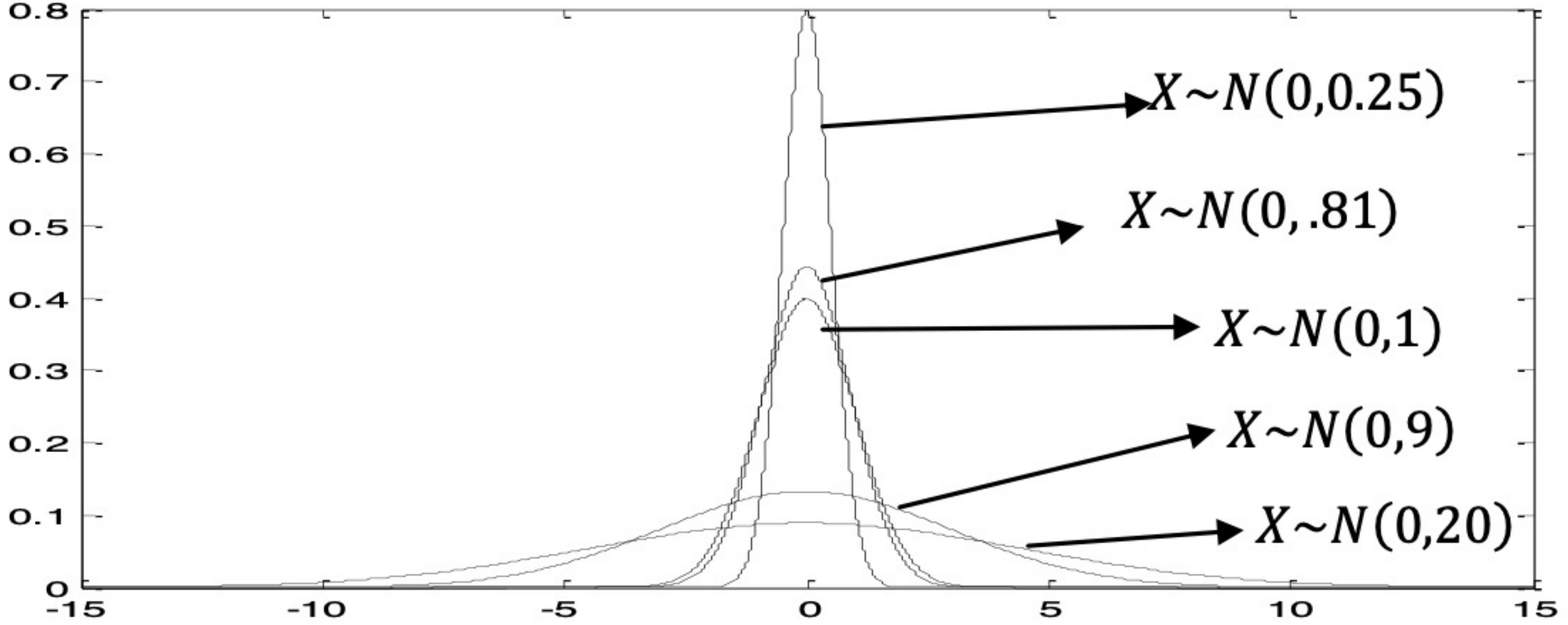
Normal Distribution

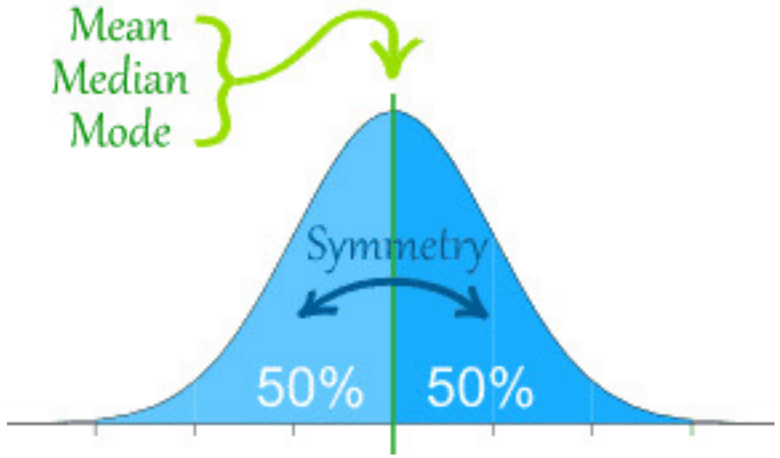
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma^2 > 0$$

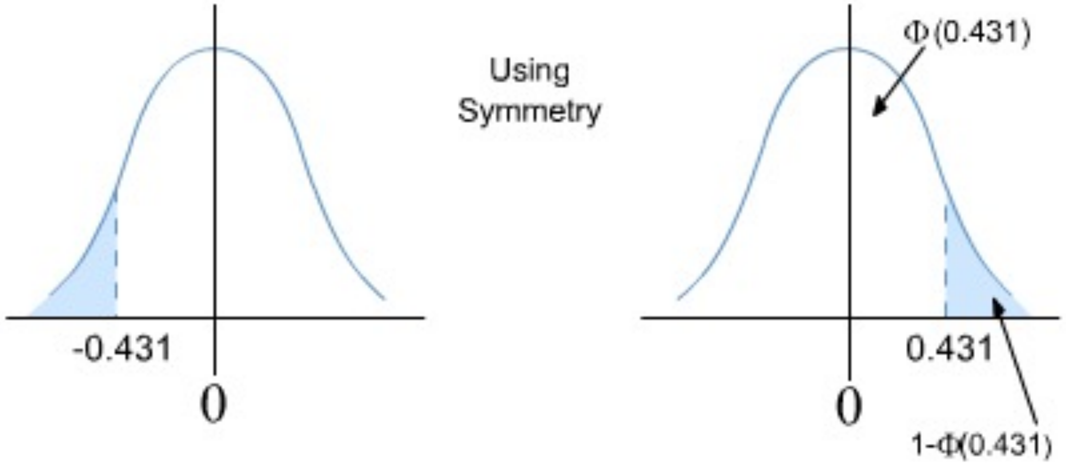
- Mean = Median = Mode= μ
- Variance = σ^2




0.14
0.12
0.1
0.08
0.06
0.04
0.02
0
-15
-10
-5
0
5
10
15
$X \sim N(6.6,9)$
$X \sim N(3,9)$
$X \sim N(0,9)$
$X \sim N(-2.5,9)$
$X \sim N(-5,9)$

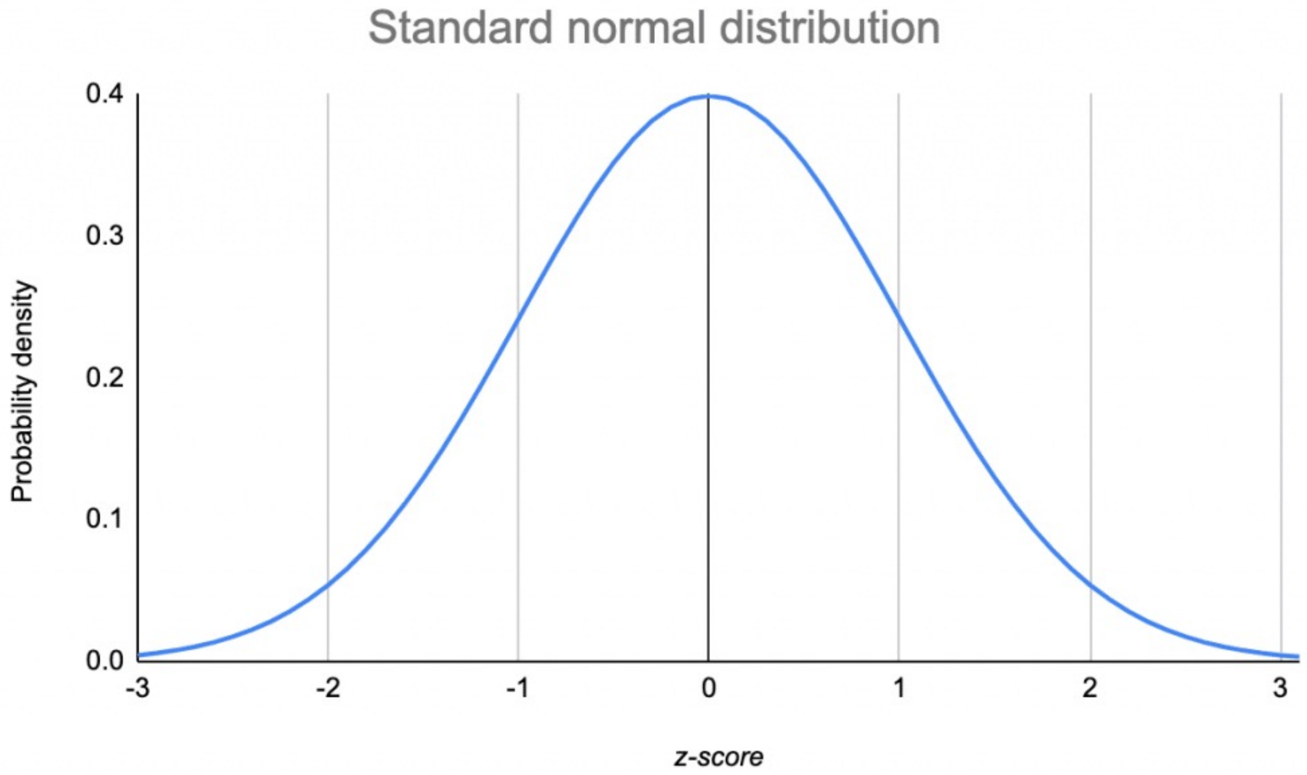

0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-15
-10
-5
0
5
10
15
$X \sim N(0,0.25)$
$X \sim N(0,.81)$
$X \sim N(0,1)$
$X \sim N(0,9)$
$X \sim N(0,20)$

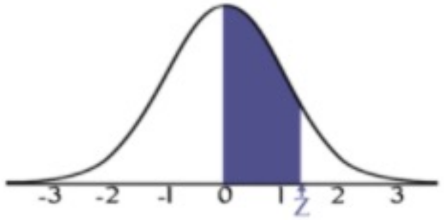

Mean
Median
Mode
Symmetry
50%
50%


-0.431
0
Using
Symmetry
Φ(0.431)
0.431
0
1-Φ(0.431)

Standard Normal Distribution

- Normal distribution for which $\mu = 0$ and $\sigma^2 = 1$
- Usually denoted with Z





STANDARD NORMAL TABLE (Z)

Entries in the table give the area under the curve between the mean and z standard deviations above the mean. For example, for $z = 1.25$ the area under the curve between the mean (0) and z is 0.3944.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0190	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2969	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3513	0.3554	0.3577	0.3529	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993
3.2	0.4993	0.4993	0.4994	0.4994	0.4994	0.4994	0.4994	0.4995	0.4995	0.4995
3.3	0.4995	0.4995	0.4995	0.4996	0.4996	0.4996	0.4996	0.4996	0.4996	0.4997
3.4	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4997	0.4998

https://access-excel.tips/z-score-probability-spss-excel/

Standardization

$$X \sim N(\mu, \sigma^2) \implies \boldsymbol{Z} = \frac{\boldsymbol{X} - \boldsymbol{\mu}}{\boldsymbol{\sigma}} \sim N(0,1)$$

Normal distribution [Internet]. [cited 2021 Oct 4]. Available from: https://www.mathsisfun.com/data/standard-normal-distribution.html

Normal Distribution - Example

- In a hospital, the systolic blood pressure of patients follow a normal distribution with mean = 15, variance = 9

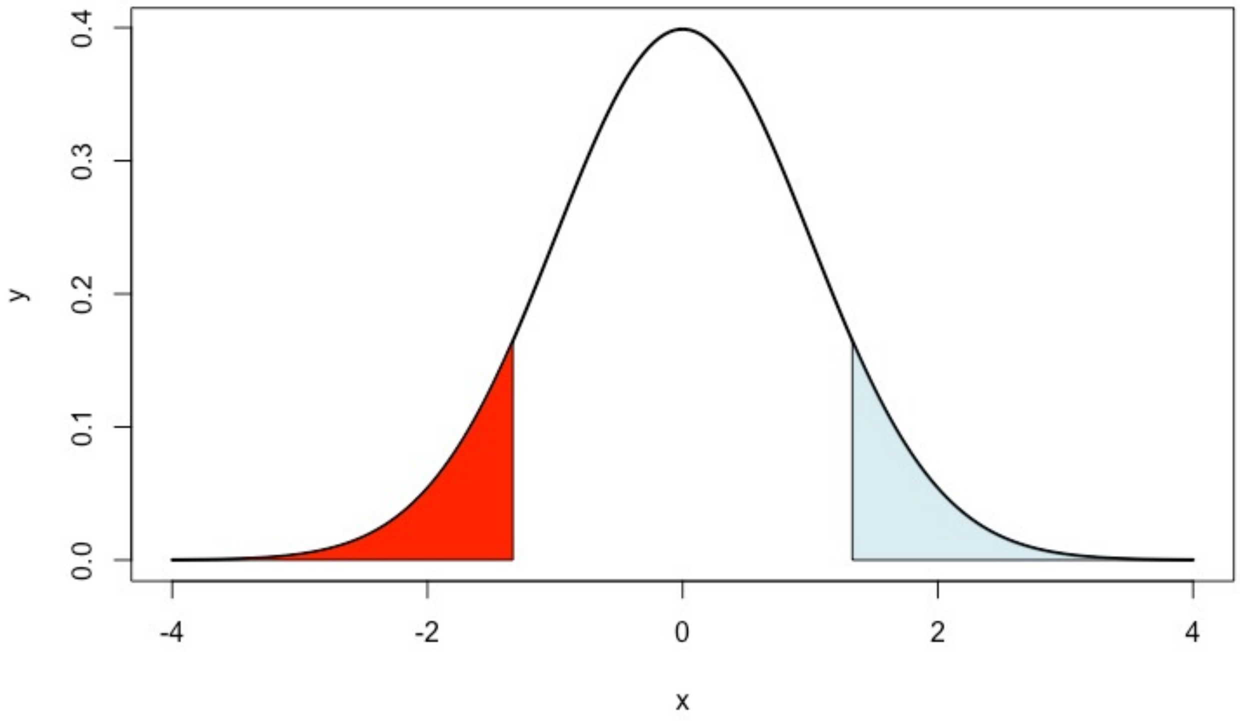
$$X \sim N(15,9)$$

- For a randomly selected patient, what is the probability that their SBP is:
  - a) Smaller than 11?
  - b) Larger than 12?
  - c) Between 9 and 16?

$X \sim N(15,9)$

a) < 11

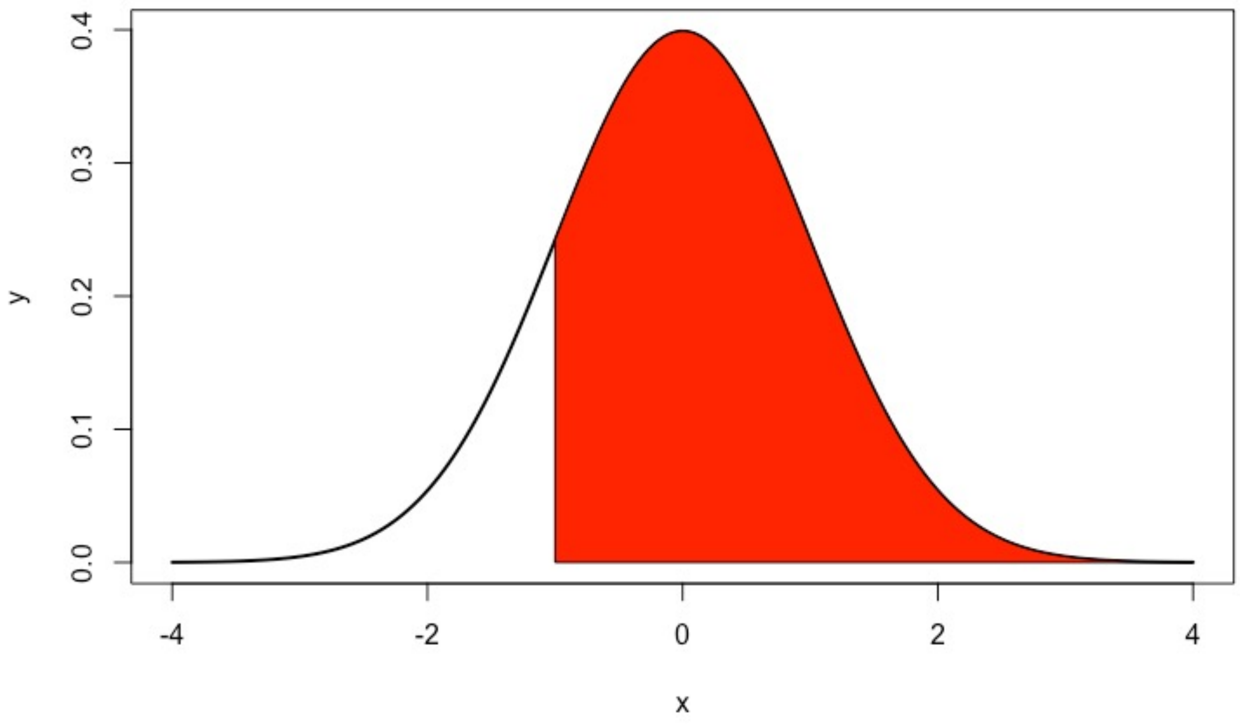
$$P(X \leq x) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = P\left(Z \leq \frac{11-15}{3}\right) = P(Z \leq -1.33) = P(Z \geq 1.33) = 0.0918$$



$X \sim N(15,9)$

b) > 12

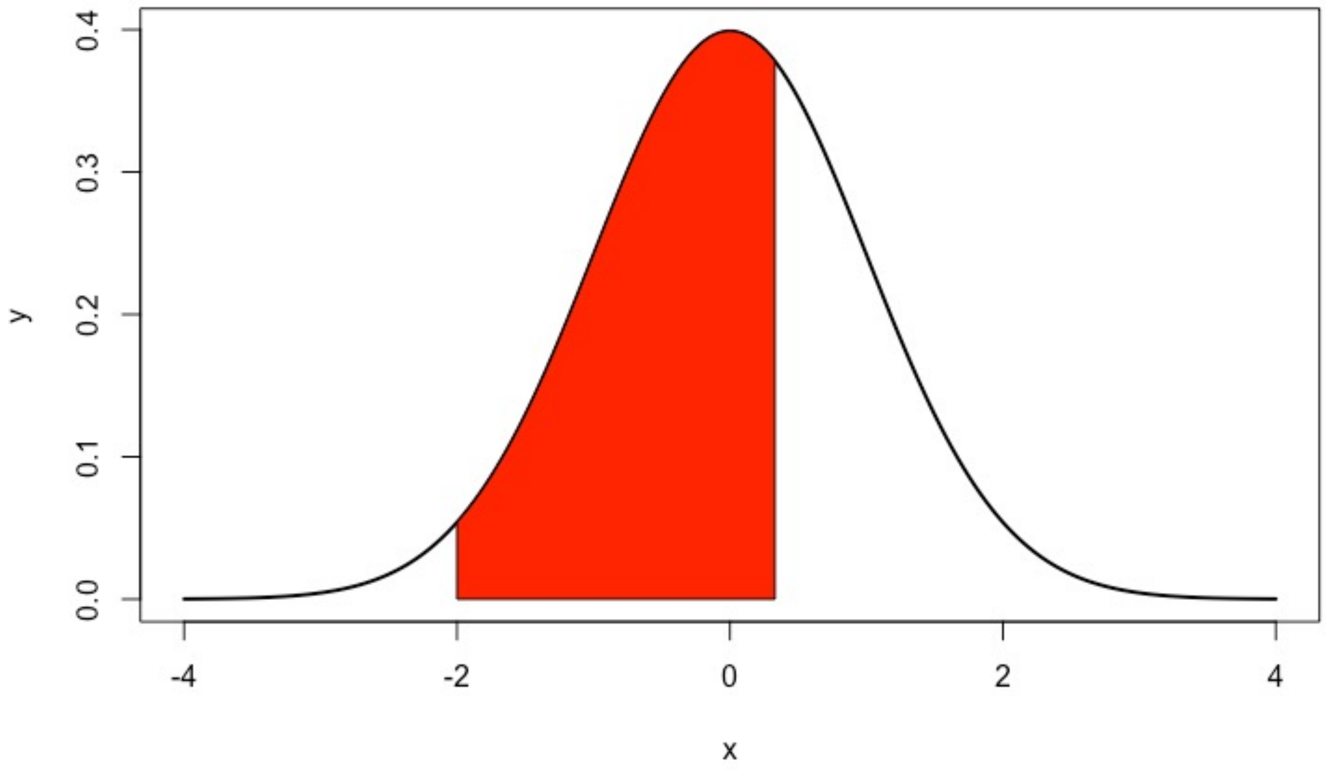
$$P(X > x) = P\left(Z > \frac{x-\mu}{\sigma}\right) = P\left(Z > \frac{12-15}{3}\right) = P(Z > -1) = 0.8413$$



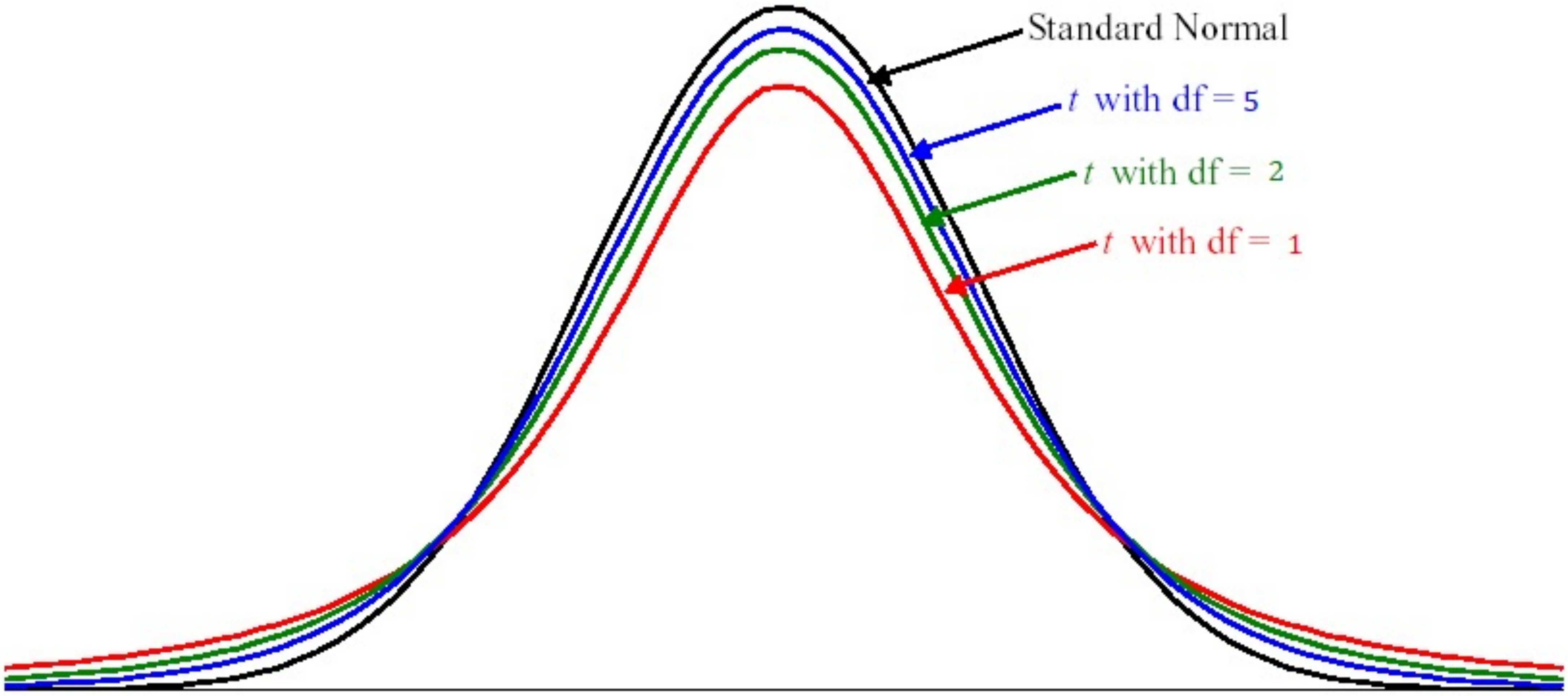
$X \sim N(15,9)$

c) Between 9 and 16

$$P(9 < X < 16) = P\left(\frac{9-15}{3} < Z < \frac{16-15}{3}\right) = P(-2 < Z < 0.33) = P(Z < 0.33) - P(Z \leq -2) = 0.6065$$



(Student's) t Distribution



Hypothesis Testing

- **Hypothesis**: an assumption that can be tested based on the evidence available
  - A novel drug is efficient in treating a certain disease
  - Regular smoking leads to lung cancer

  - Overweight individuals who (1) consume greasy food and (2) consume a low amount vegetables (1) have high levels of cholesterol and (2) have a higher risk of cardiovascular diseases

- **Hypothesis test**: investigation of the hypothesis using the sample
  - Assessing evidence provided by the data against the null claim (the claim which is to be assumed true unless enough evidence exists to reject it)

Null and Alternative Hypotheses

- H_0 – Null hypothesis
  - The mean of a variable is not different than *c*
  - There is no difference between the two groups' means
  - There is no difference compared to baseline
  - …
- H_a or H_1 – Alternative hypothesis
  - There is a difference between the two groups' means
  - The mean in group A is higher than group B
  - …

One- vs. Two-tailed Tests

- The coin is biased

Two-tailed

H_0: $p = 0.5$

H_a: $p \neq 0.5$

- The probability of heads is larger (or smaller) than 0.5

One-tailed

H_0: $p \leq 0.5$ (or $p \geq 0.5$)

H_a: $p > 0.5$ (or $p < 0.5$)

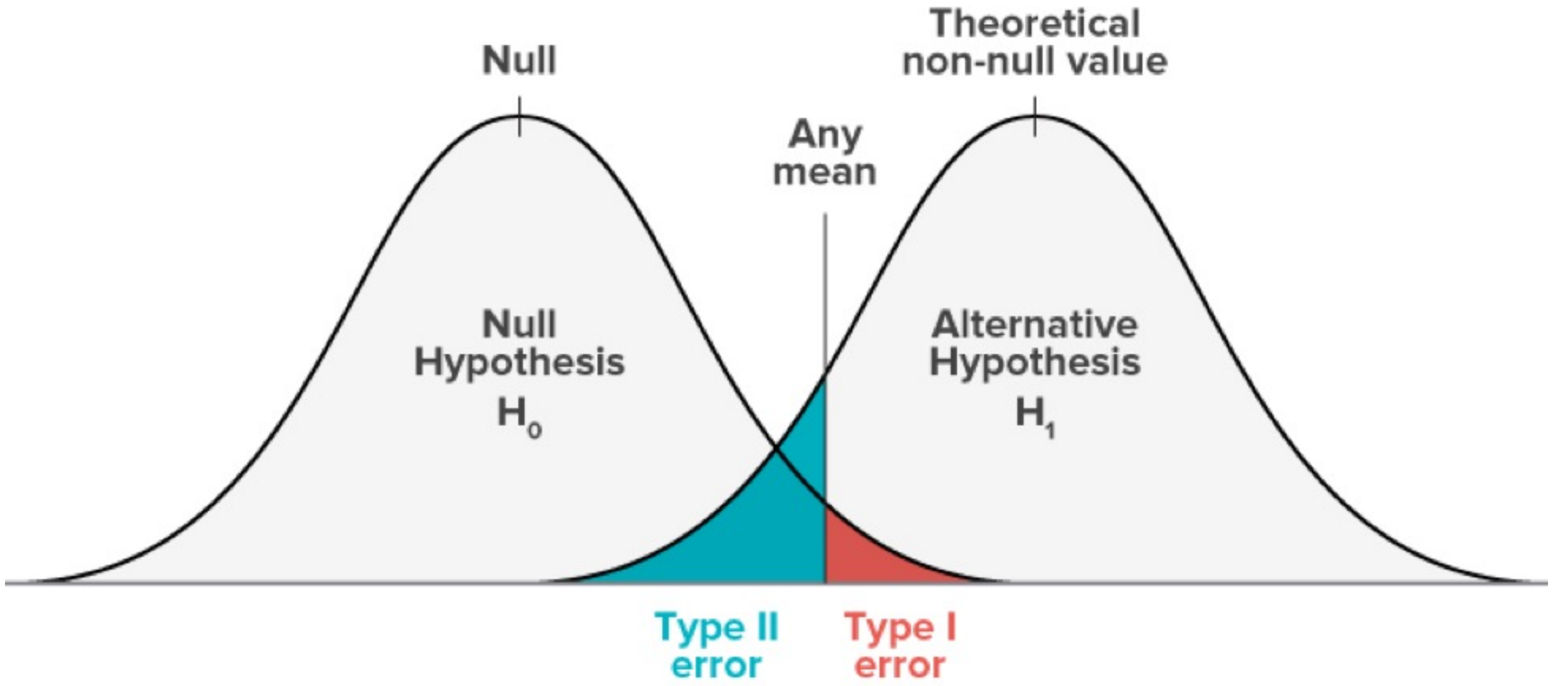
H_0	Decision	
	Fail to reject	Reject
True	Correct decision	**Type I Error** α
False	**Type II Error** **β**	Correct decision

Barai A. Hypothesis Testing | All a beginner needs to know [Internet]. Medium. 2020 [cited 2021 Oct 4]. Available from: https://medium.com/analytics-vidhya/hypothesis-testing-all-a-beginner-needs-to-know-dd1bdeebdbf1

Hypothesis Testing

- P(Type 1 error) = α = P(reject H_0| H_0 is true)
- P(Type 2 error) = β = P(fail to reject H_0| H_0 is false)

- As α gets larger β gets smaller, vice versa
- As *n* gets large, both α and β get smaller



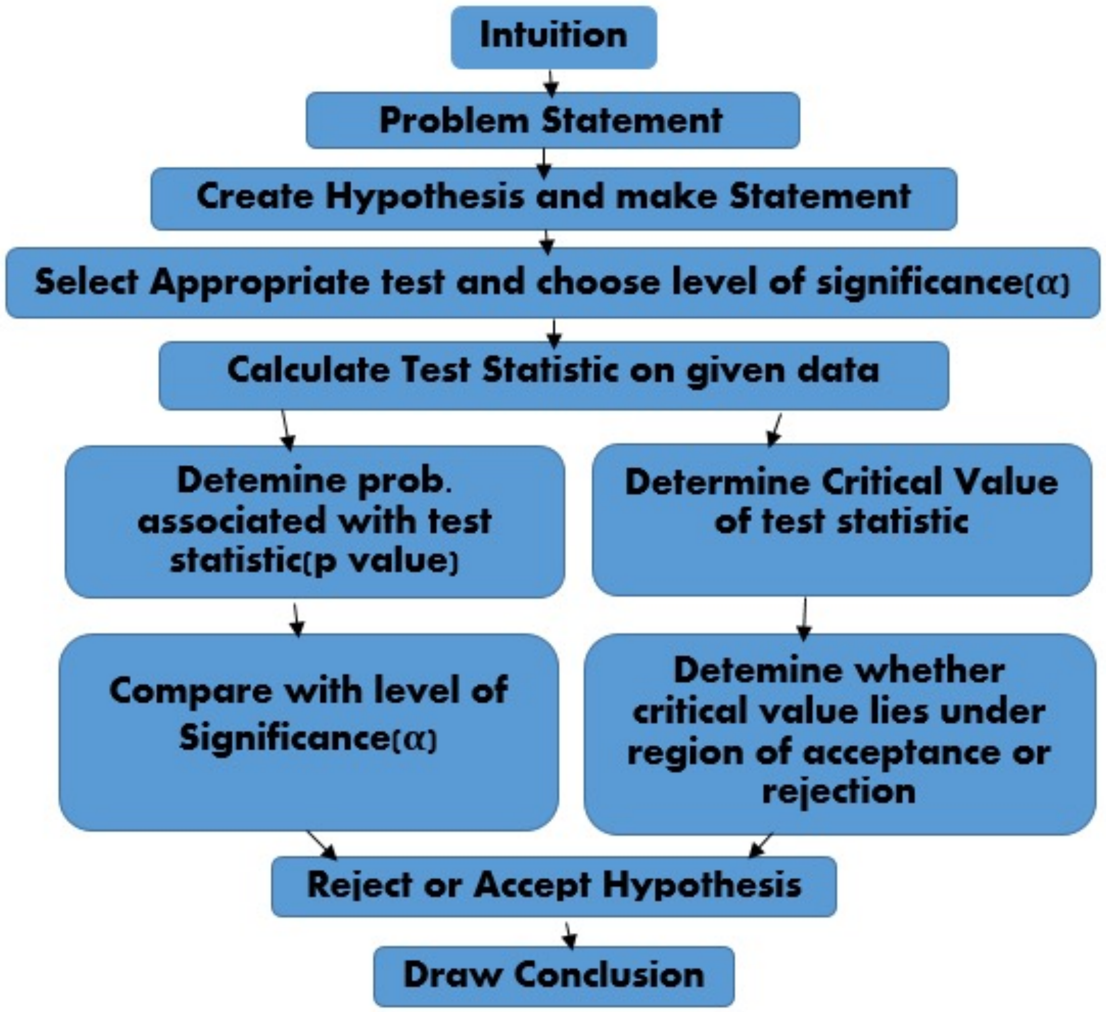
Hypothesis Testing

H_0	Decision	
	Fail to reject	Reject
True	Correct decision	**Type I Error** α
False	**Type II Error** **ß**	Correct decision

- **Confidence level** = $1 - \alpha$
  - P(fail to reject H_0\| H_0 is true)
- **Statistical power** = $1 - \beta$
  - P(reject H_0\| H_0 is false)

Barai A. Hypothesis Testing | All a beginner needs to know [Internet]. Medium. 2020 [cited 2021 Oct 4]. Available from: https://medium.com/analytics-vidhya/hypothesis-testing-all-a-beginner-needs-to-know-dd1bdeebdbf1

Hypothesis Testing - Steps



Hypothesis Testing - Steps

1. **Check assumptions, determine H_0 and H_a, choose α**
    - Assumptions differ based on the test
    - The null hypothesis always contains equality (=)
2. **Calculate the appropriate test statistic**
    - z, t, χ^2, …
3. **Calculate critical values/p value**
    - With the aid of precalculated tables/software
4. **Decide whether to reject/fail to reject H_0**
    - Reject if the statistic is within the critical region/$p \leq \alpha$

Test Statistic

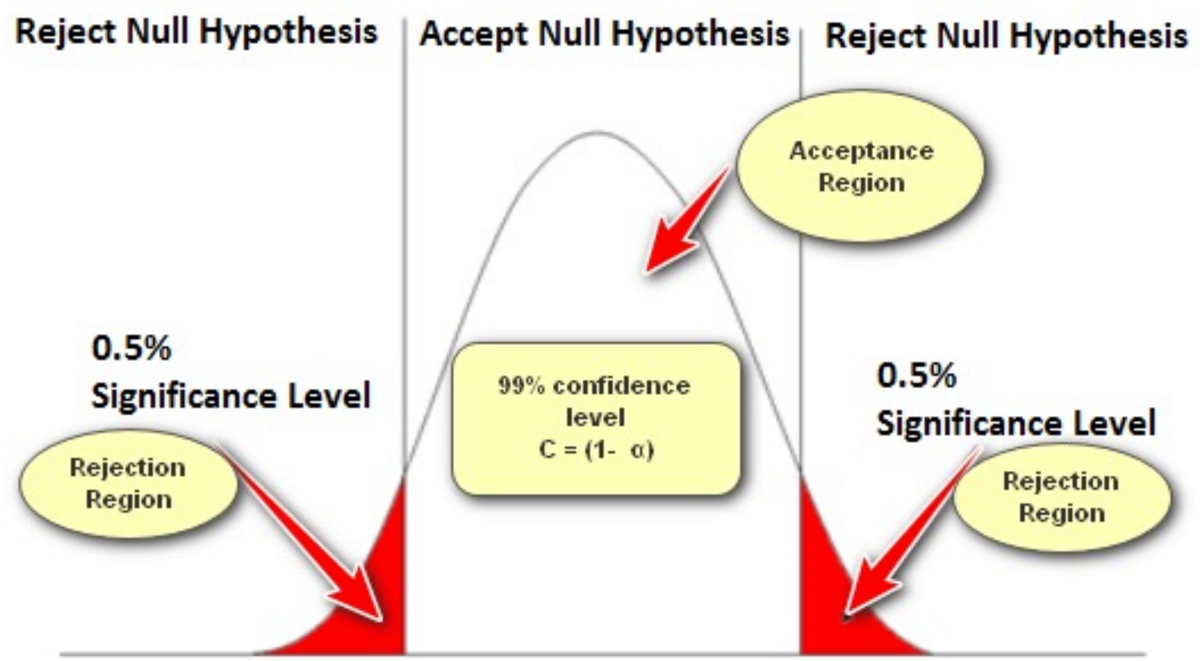
$$\text{test statistic} = \frac{\text{estimator} - \text{null value}}{\text{standard error of estimator}}$$

$$t = \frac{\boxed{\bar{X} - \mu}}{\boxed{s/\sqrt{n}}}$$

Barai A. Hypothesis Testing | All a beginner needs to know [Internet]. Medium. 2020 [cited 2021 Oct 4]. Available from: https://medium.com/analytics-vidhya/hypothesis-testing-all-a-beginner-needs-to-know-dd1bdeebdbf1

Critical Value/Rejection Region

- We select α (**significance level**) prior to performing a hypothesis test
  - Some common values for α are 0.01, **0.05** an 0.10
- Based on the selected α, the critical values are calculated, and the rejection region is determined
  - the region where the null hypothesis is rejected



$H_0: \mu = \mu_0$

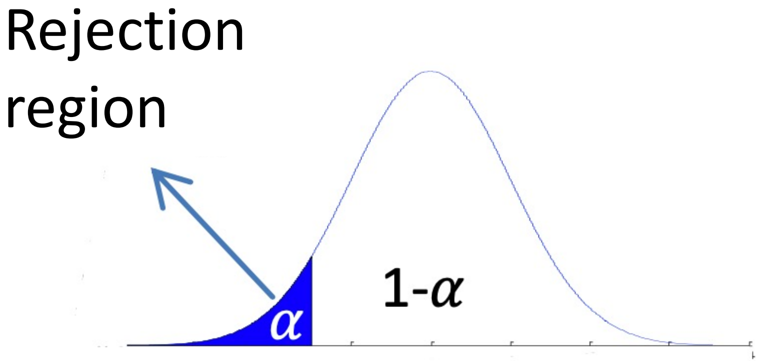
$H_1: \mu < \mu_0$

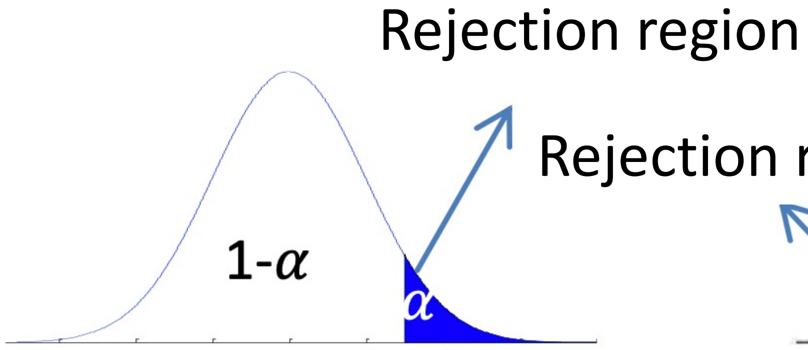
$H_0: \mu = \mu_0$

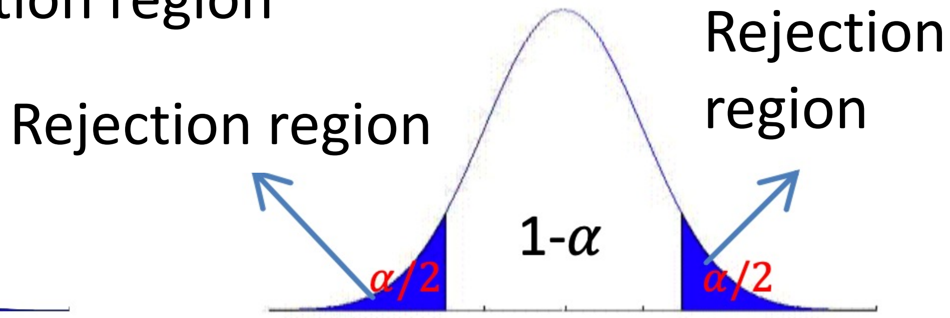
$H_1: \mu > \mu_0$

$H_0: \mu = \mu_0$

$H_1: \mu \neq \mu_0$

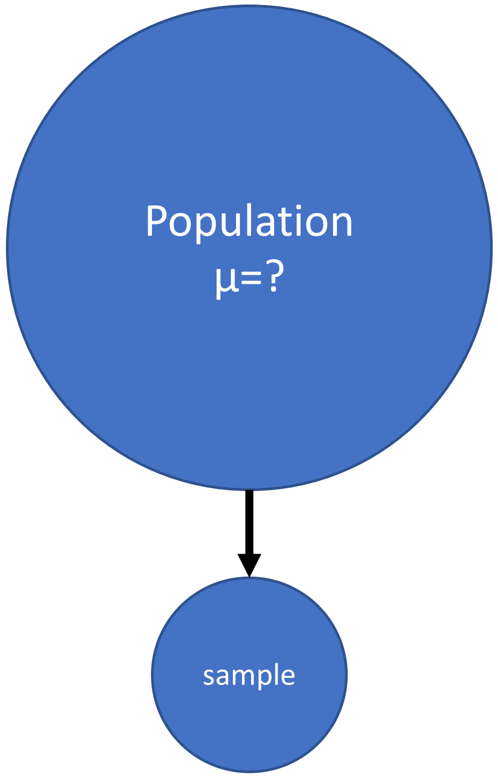




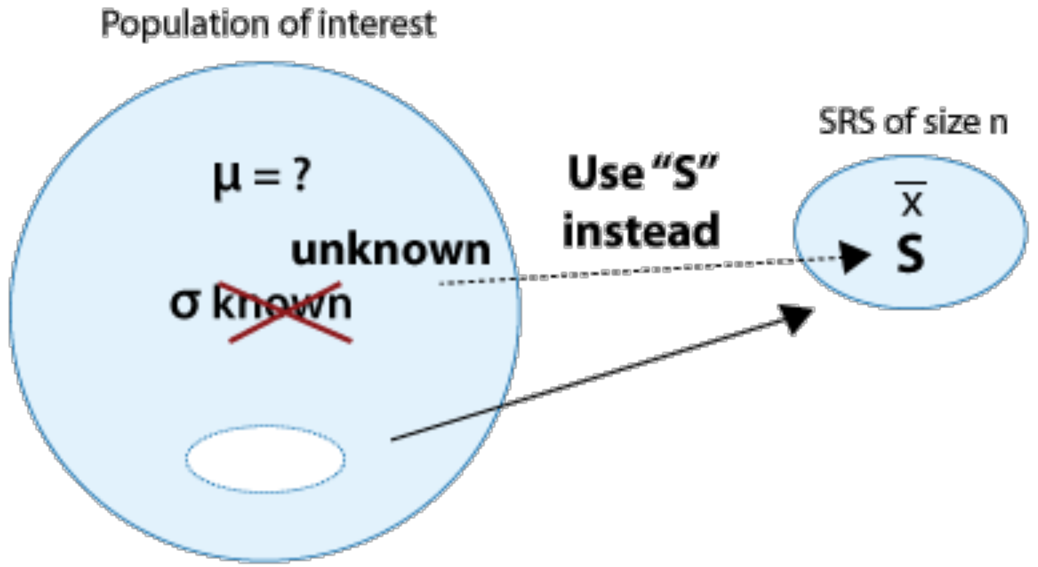


One-Sample t-Test

- a statistical hypothesis test used to determine whether an unknown population mean is different from a specific value



One-Sample t-Test



Means(All steps) [Internet]. [cited 2021 Oct 6]. Available from: https://bolt.mph.ufl.edu/6050-6052/unit-4/module-12/means-all-steps/

One-Sample t-Test – Example I

id	week_1	cd4_1	week_2	cd4_2	perc_benefit
361	0	26	7.43	3	-11.905994
1017	0	13	7.00	10	-3.296703
519	0	3	8.14	5	8.190008
1147	0	65	33.00	97	1.491841
1216	0	36	8.00	31	-1.736111
52	0	16	9.43	31	9.941676
660	0	34	8.43	32	-0.697788
1145	0	41	8.00	71	9.146341
697	0	33	8.00	45	4.545455
560	0	21	8.00	27	3.571429

- Mean percentage benefit is 1.925015
- Is it due to chance? Or does it indicate positive impact of the novel treatment?
  - What would be the value of mean percentage benefit what if you selected another set of 10 patients?

One-Sample t-Test – Example I (cont.)

1. Check assumptions, determine H_0 and H_a, choose α
   - Normality of the variable is checked (Quantile-quantile plot)
   - H_0: $\mu = 0$ H_a: $\mu \neq 0$
   - $\alpha = 0.05$

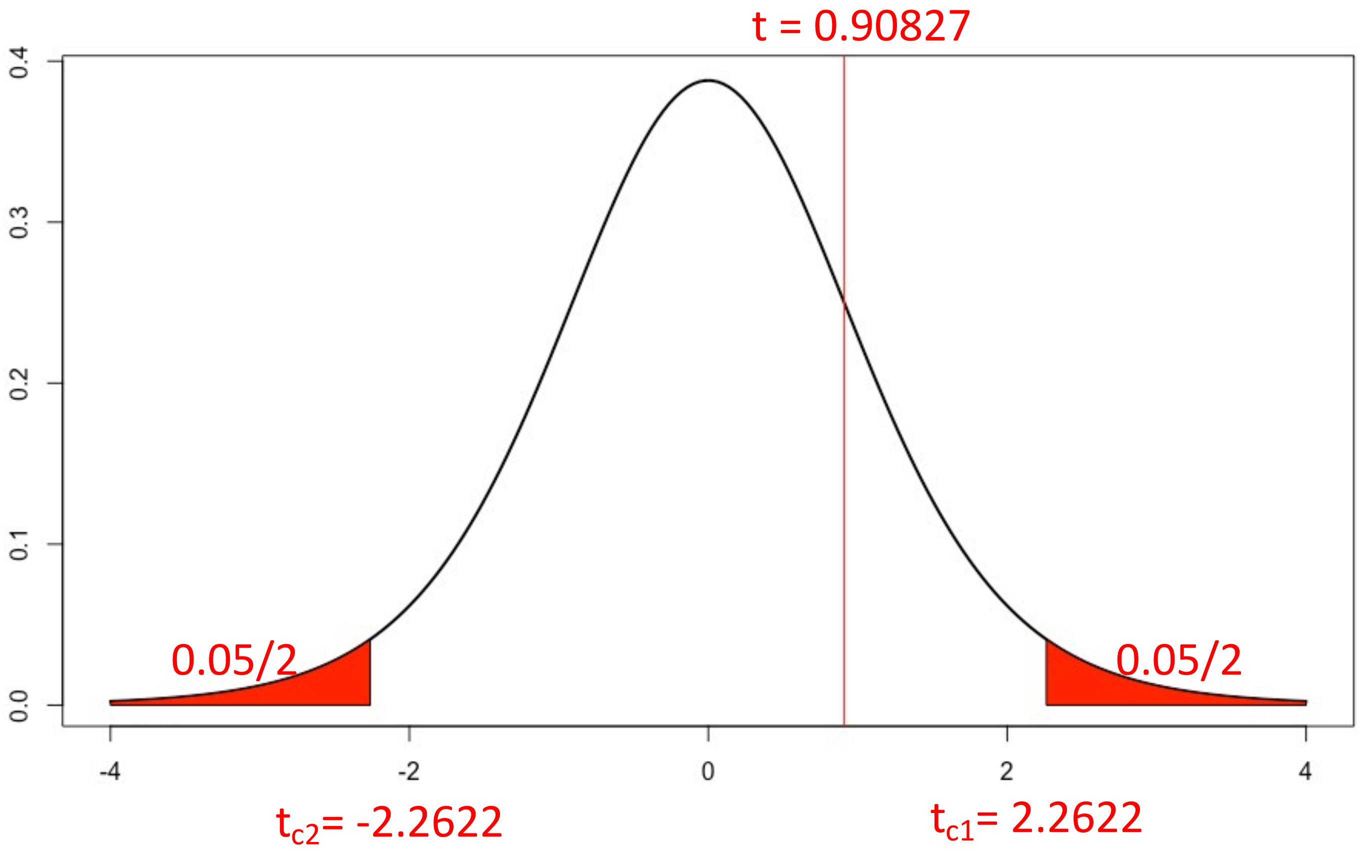
One-Sample t-Test – Example I (cont.)

2. Calculate the appropriate test statistic
    - Mean percentage benefit is 1.925015
    - Standard deviation is 6.702202
    - Sample size is 10

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{1.925015 - 0}{6.702202/\sqrt{10}} = 0.9082736 \quad (\sim t_{n-1} = t_9)$$

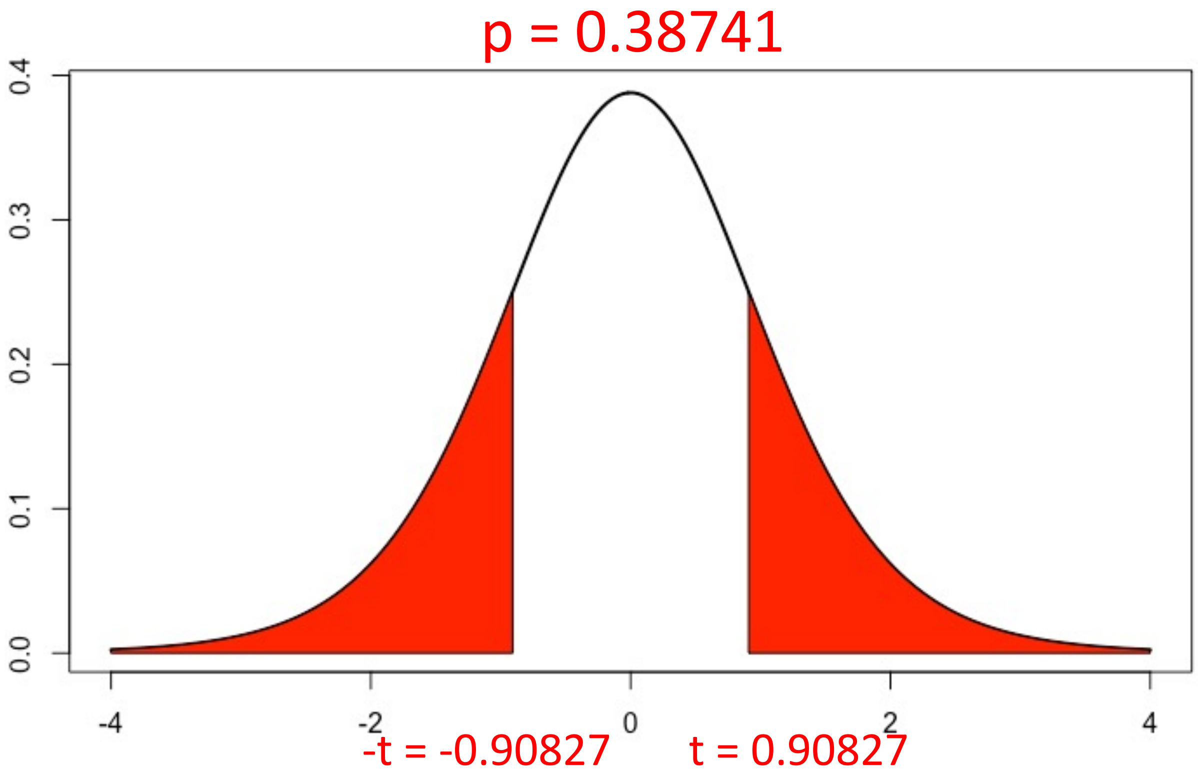
One-Sample t-Test – Example I (cont.)

3. Calculate **critical values**/p value
4. Decide whether to reject/fail to reject H_0



One-Sample t-Test – Example I (cont.)

3. Calculate critical values/**p value**
4. Decide whether to reject/fail to reject H_0



One-Sample t-Test – Example II

- It is claimed that the post-treatment tumor volume of glioblastoma patients subject to a novel treatment is different than 5 cm^3
- The mean tumor volume of 41 randomly-selected patients is 5.9 cm^3
- Sample standard deviation is 1.74

One-Sample t-Test – Example II (cont.)

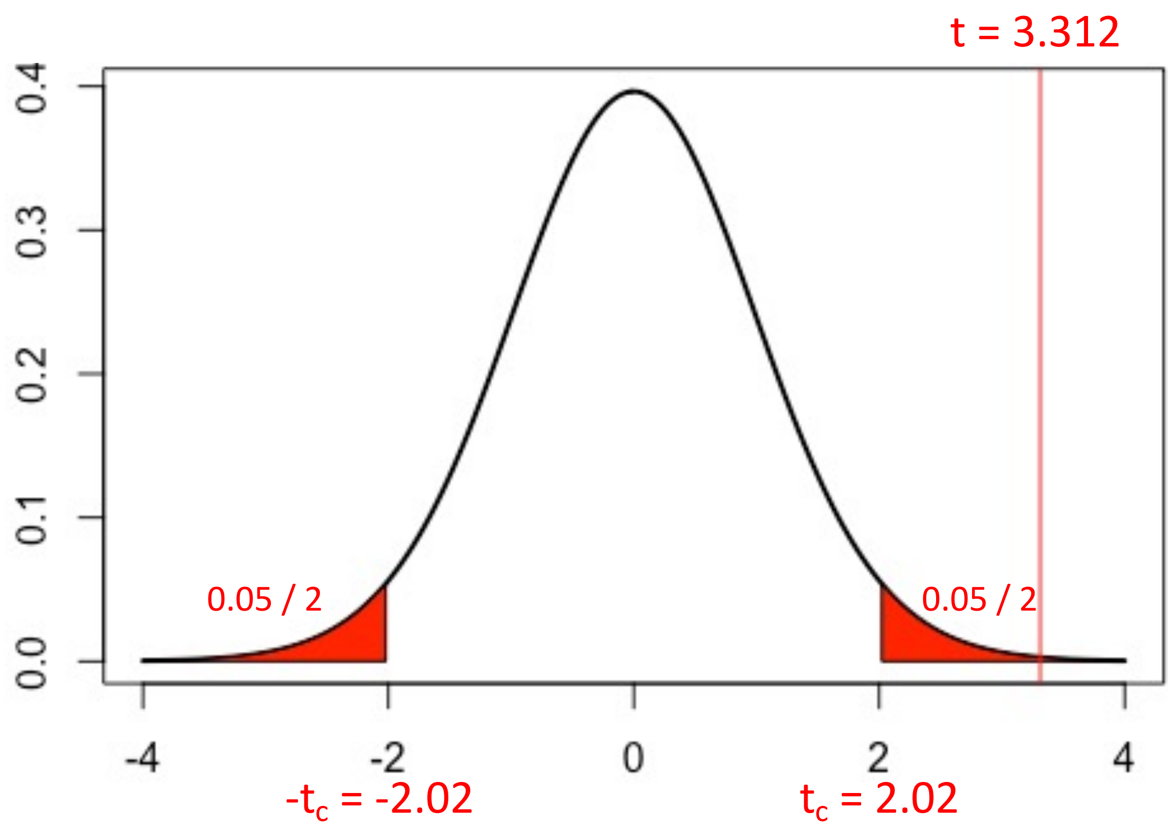
1. Check assumptions, determine H_0 and H_a, choose α
   - Normality of the variable is checked
   - H_0: $\mu = 5$ H_a: $\mu \neq 5$
   - $\alpha = 0.05$

2. Calculate the appropriate test statistic

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{5.9 - 5}{1.74/\sqrt{41}} = 3.312 \qquad (\sim t_{n-1} = t_{40})$$

One-Sample t-Test – Example II (cont.)

3. Calculate **critical values**/p value
4. Decide whether to reject/fail to reject H_0



One-Sample t-Test – Example II (cont.)

5. State a conclusion:

With 95% confidence, we can conclude that there is enough evidence to say that post-treatment tumor volume of glioblastoma patients subject to a novel treatment is different than 5 cm^3.

One-Sample t-Test – Example III

- It is claimed that:
- A novel drug reduces the recovery time of patients to less than 10 days
- Recovery time for 7 randomly-selected patients:

  2, 4, 11, 3, 4, 6, 8 ($\bar{X}$ = 5.43, s = 3.15)

- Test the hypothesis using $\alpha = 0.01$

One-Sample t-Test – Example III((cont.)

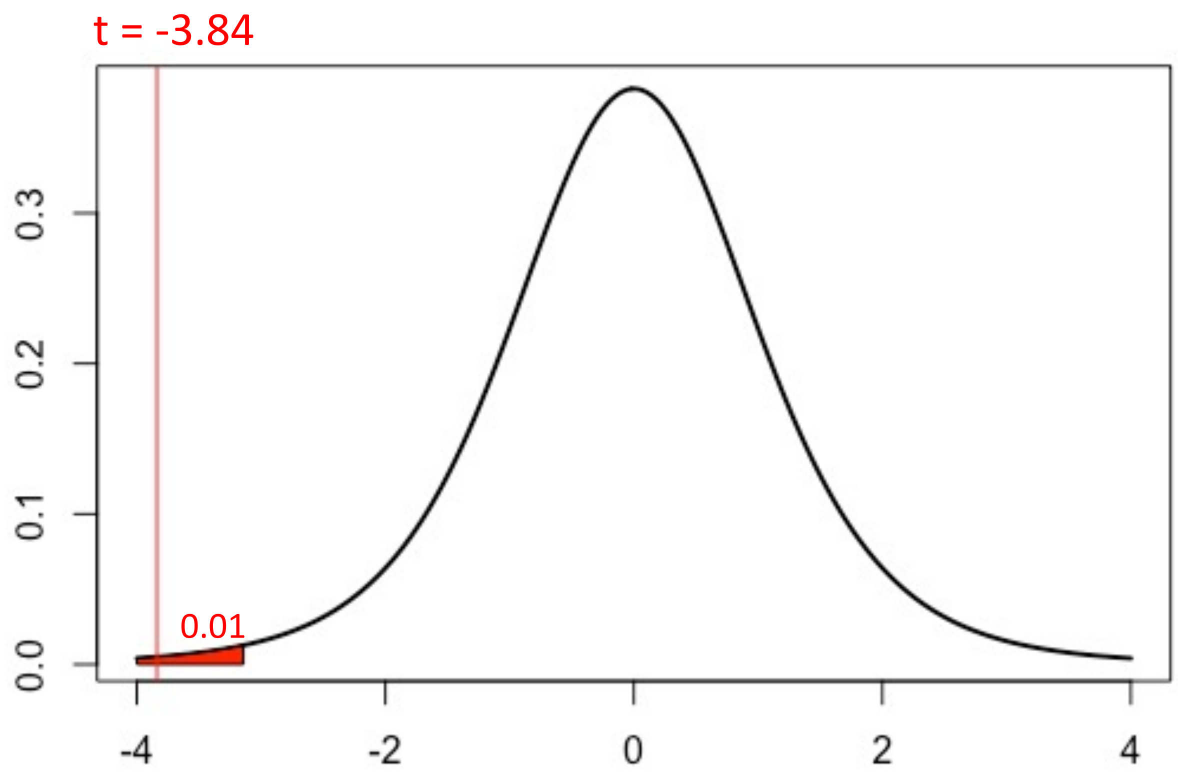
1. Check assumptions, determine H_0 and H_a, choose α
   - Normality of the variable is checked
   - H_0: $\mu \geq 10$ H_a: $\mu < 10$
   - $\alpha = 0.01$

2. Calculate the appropriate test statistic

$$t = \frac{\bar{X} - \mu}{s/\sqrt{n}} = \frac{5.43 - 10}{3.15/\sqrt{7}} = -3.84 \quad (\sim t_{n-1} = t_6)$$

One-Sample t-Test – Example III (cont.)

3. Calculate **critical values**/p value
4. Decide whether to reject/fail to reject H_0



Barai A. Hypothesis Testing | All a beginner needs to know [Internet]. Medium. 2020 [cited 2021 Oct 4]. Available from: https://medium.com/analytics-vidhya/hypothesis-testing-all-a-beginner-needs-to-know-dd1bdeebdbf1

Brief Summary