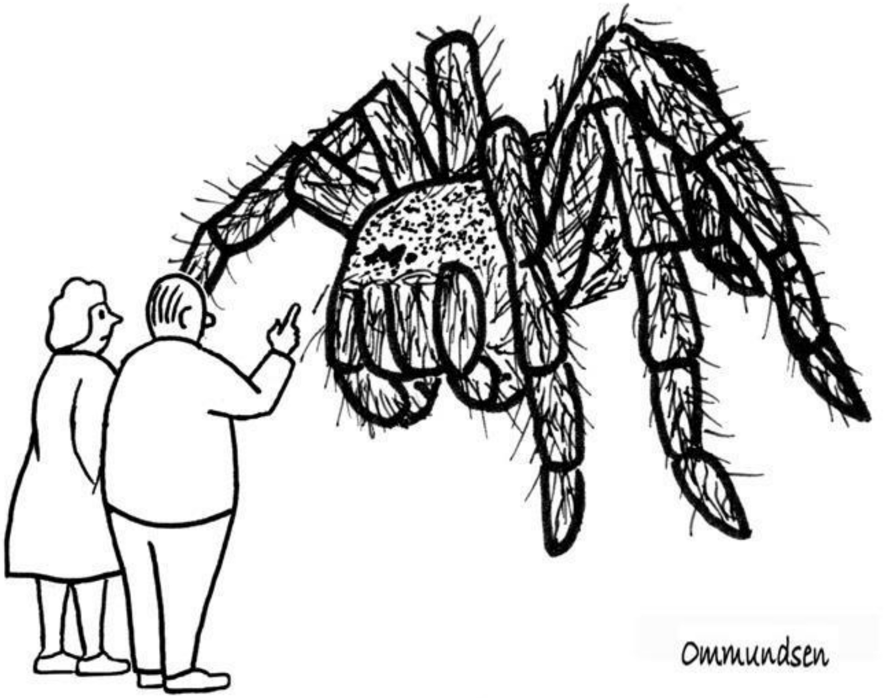


"I've narrowed it to two hypotheses: it grew or we shrunk."

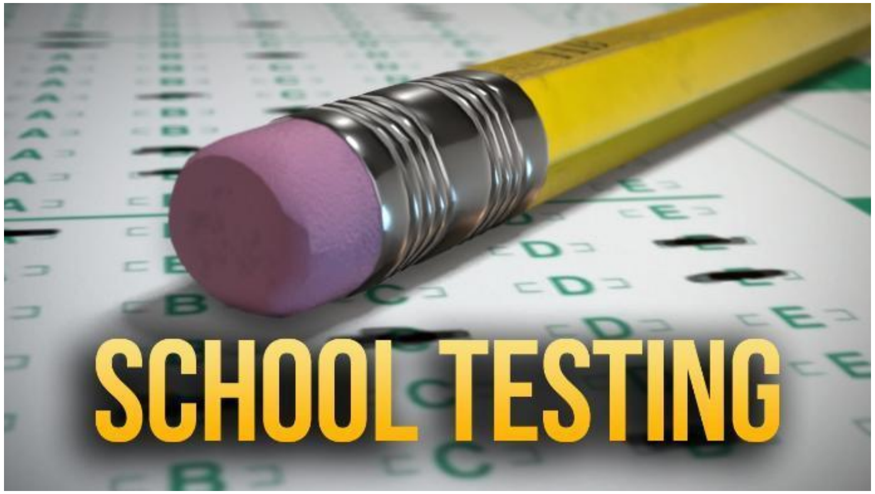
Hypothesis Testing

t-Distribution

Hypothesis tests give a way of using samples to test whether or not statistical claims are likely to be true or not.

Hypothesis testing

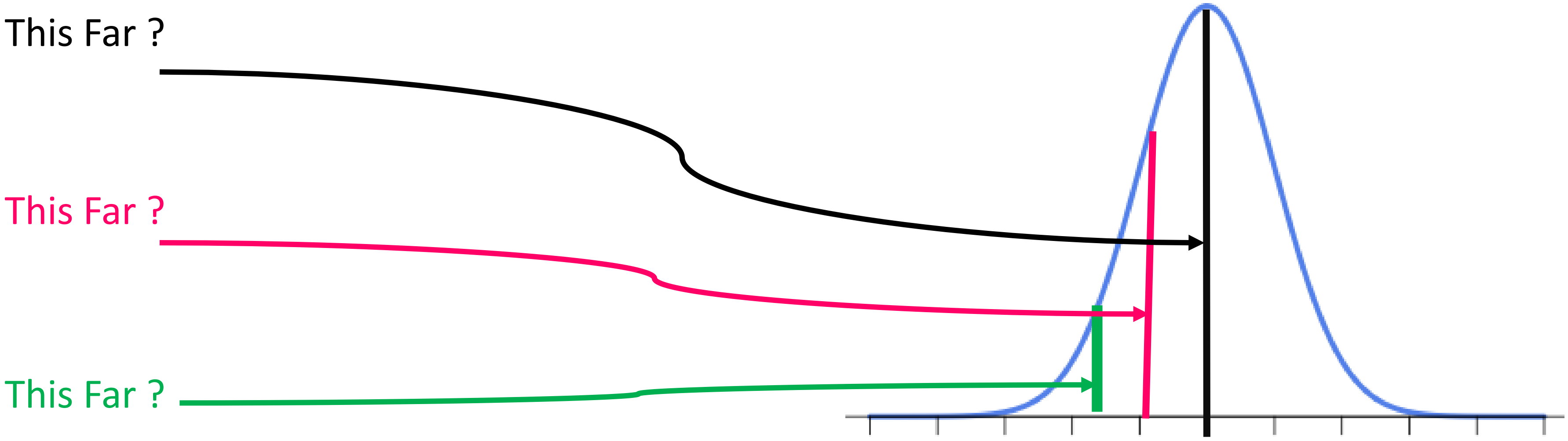
A school principal claims that the students from her school have an average score of 7/10 in a English Proficiency test.



You doubt that claim and take a random sample of 40 students and you find a mean score of 5.5/10, with a sample standard deviation of 1. Can you reject the principal's claim?

Hypothesis Testing Process

Considering variations in samples, how far away from 7/10 is acceptable to you as expected variation and when do you say “enough is enough; this is too far”?



This Far ?

This Far ?

This Far ?

Claim or Expectation, say, mean score = 7 /10

Step 1: Decide on the hypotheses

Average score on the test is 7/10.

This is called Null Hypothesis and is represented by H_0.

In this case, H0 : μ = 0.7

If Null Hypothesis is rejected based on evidence, an Alternate Hypothesis, H_1, needs to be accepted. **We always start with the assumption that Null Hypothesis is true.**

In this case, H1: μ < 0.7

Examples of Hypotheses

- Two hypotheses in competition:
  - H0 : The NULL hypothesis, usually the most conservative.
  - H1 or HA : The ALTERNATIVE hypothesis, the one we are actually interested in.
- Examples of NULL Hypothesis:
  - The coin is fair
  - The new drug is no better (or worse) than the placebo
- Examples of ALTERNATIVE hypothesis:
  - The coin is biased (either towards heads or tails)
  - The coin is biased towards heads
  - The coin has a probability 0.6 of landing on tails
  - The drug is better than the placebo

Step 2: Choose your statitics

Sample size = 40

Normal distribution is a good approximation

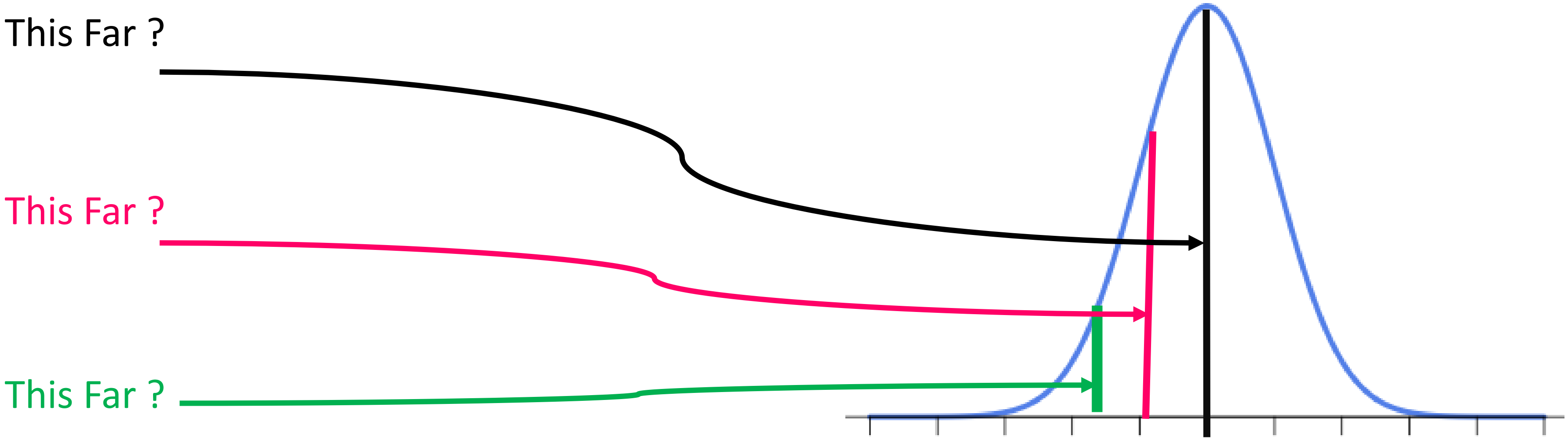
$$Std\ Err = \frac{s}{\sqrt{n}} = \frac{1.0}{\sqrt{40}} = 0.158$$

X ~ N(0.7, 0.158^2) = N(0.7, 0.025)

$$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{0.55 - 0.7}{0.158} = -0.9$$

Step 3: Specify the significance Level

First, we must decide on the Significance Level, α. It is a measure of how unlikely you want the results of the sample to be before you reject the null hypothesis, H_0.



This Far ?

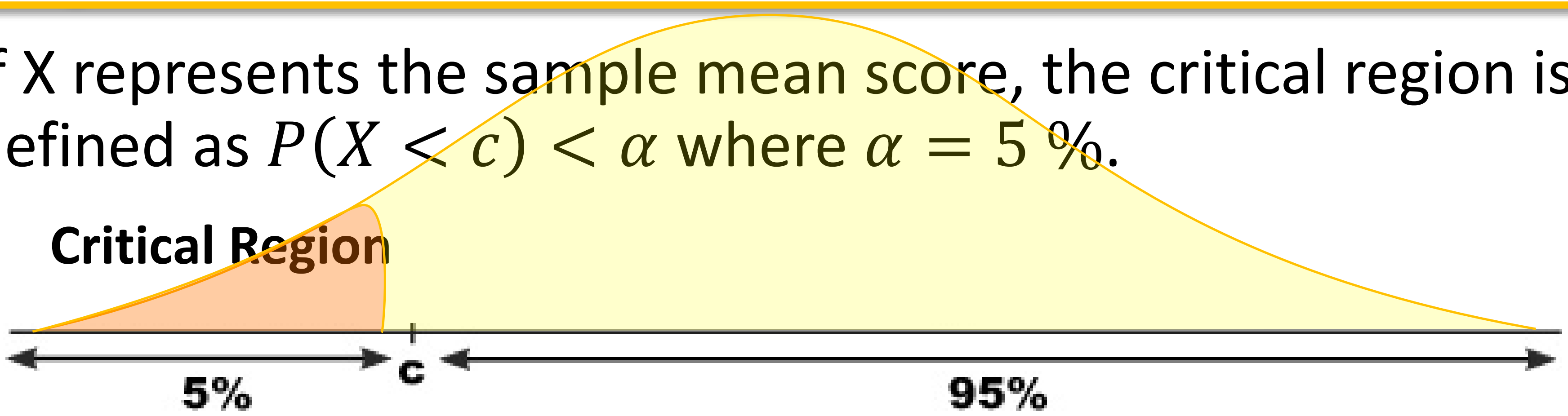
This Far ?

This Far ?

Claim or Expectation, say, mean score = 7 /10

Step 4: Determine the critical region

If X represents the sample mean score, the critical region is defined as $P(X < c) < \alpha$ where $\alpha = 5\ \%$.



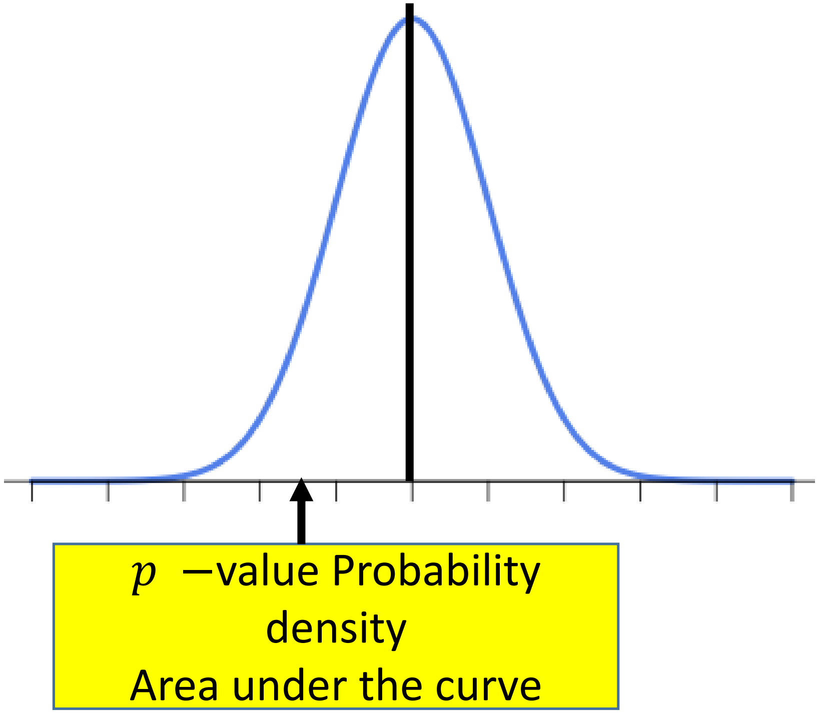
Recall that in a 95% CI, there is a 5% chance that the sample will not contain the population mean. Hence if the sample falls in the critical region, the null hypothesis that 0.7 is the mean score is rejected.

That is the reason 5% or 0.05 is called the Significance Level. In a 99% CI, 0.01 is the Significance Level.

Step 5: Find the p – value

p-value is the probability of getting a value up to and including the one in the sample in the direction of the critical region.

It is a way of taking the sample and working out whether the result falls within the critical region of the hypothesis test.



p –value Probability density
Area under the curve

Essentially, this is the value used to determine whether or not to reject the null hypothesis.

Step 5: Find the p −value

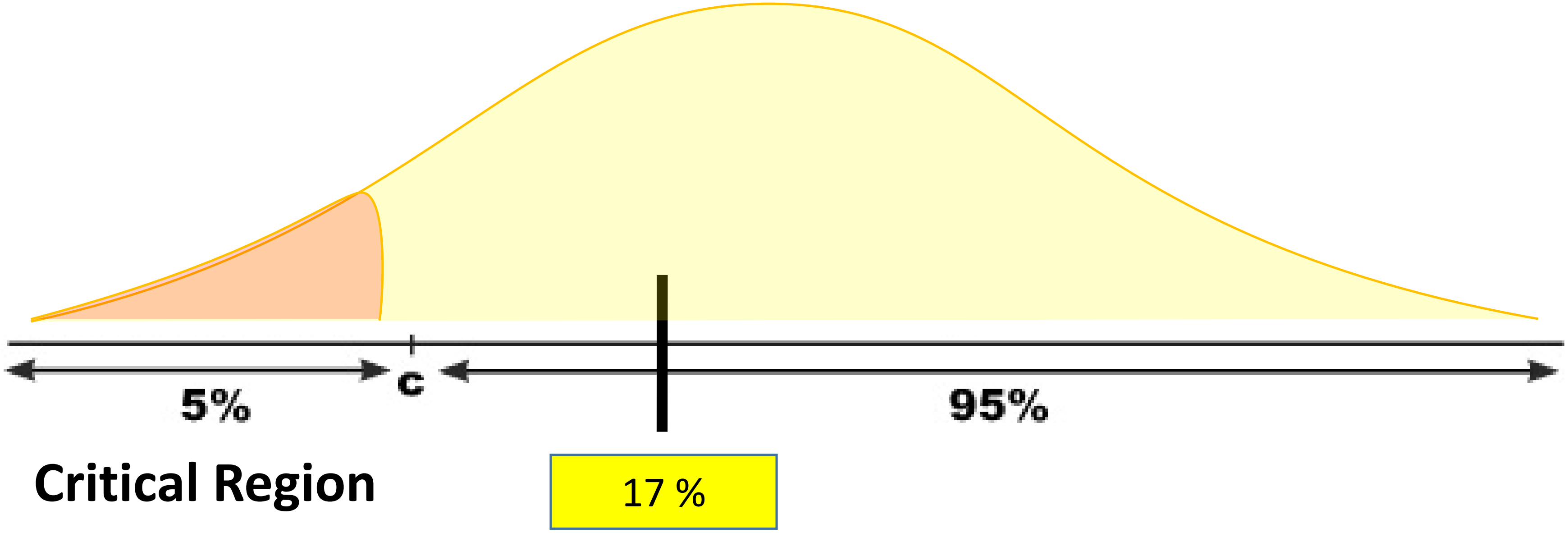
In our sample, we found a mean score of 5.5/10. This means our *p*-value is P(X = 0.55), where X is the distribution of the mean scores in the sample.

If P(X = 0.55) < 0.05 (Significance Level), it indicates that 0.55 is inside the critical region, and hence H_0 can be rejected.

Given that Z = -0.94 , P(X ≤ 0.55) = 0.171

So there is a 17% probability of find a mean score of 5.5/10 or less.

Step 6: Is the sample result in the critical region ?



Step 7: Make your decision

There isn’t sufficient evidence to reject the null hypothesis and so, the claims of the principal are accepted.

Would your conclusion be any different if the same average score of 5.5/10 was found from a sample of size 400?

What are the null and alternate hypotheses ?

$$H_0 : \mu = 0.7$$
$$H_1 : \mu = 0.7$$

What is the test statistics ?

$$Z = \frac{\bar{x} \ - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{0.55 \ - 0.7}{\frac{1}{\sqrt{400}}} = \ -3$$

p-value = P(Z < -3.0) = 0.00135

What is your decision ?

Since the *p*-value (0.00135) is less than the Significance Level of 0.05, the null hypothesis can be rejected.

Attention Check

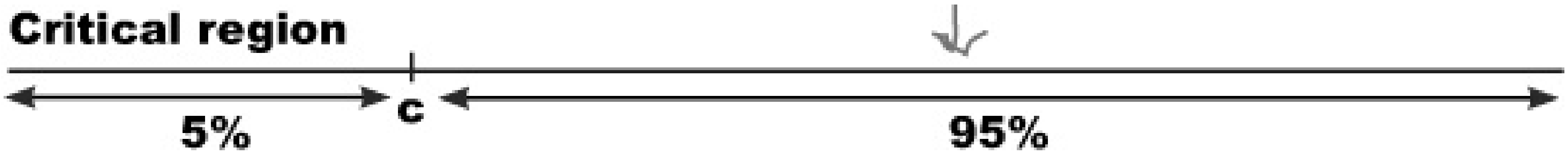
In hypothesis testing, do you assume the null hypothesis to be true or false?

True.

If there is sufficient evidence against the null hypothesis, do you accept it or reject it?

Reject it.

Attention Check



If the *p*-value is less than 0.05 for the above significance level, will you accept or reject the null hypothesis?

Reject it.

Do you need weaker evidence or stronger to reject the null hypothesis if you were testing at the 1% significance level instead of the 5% significance level?

Stronger.

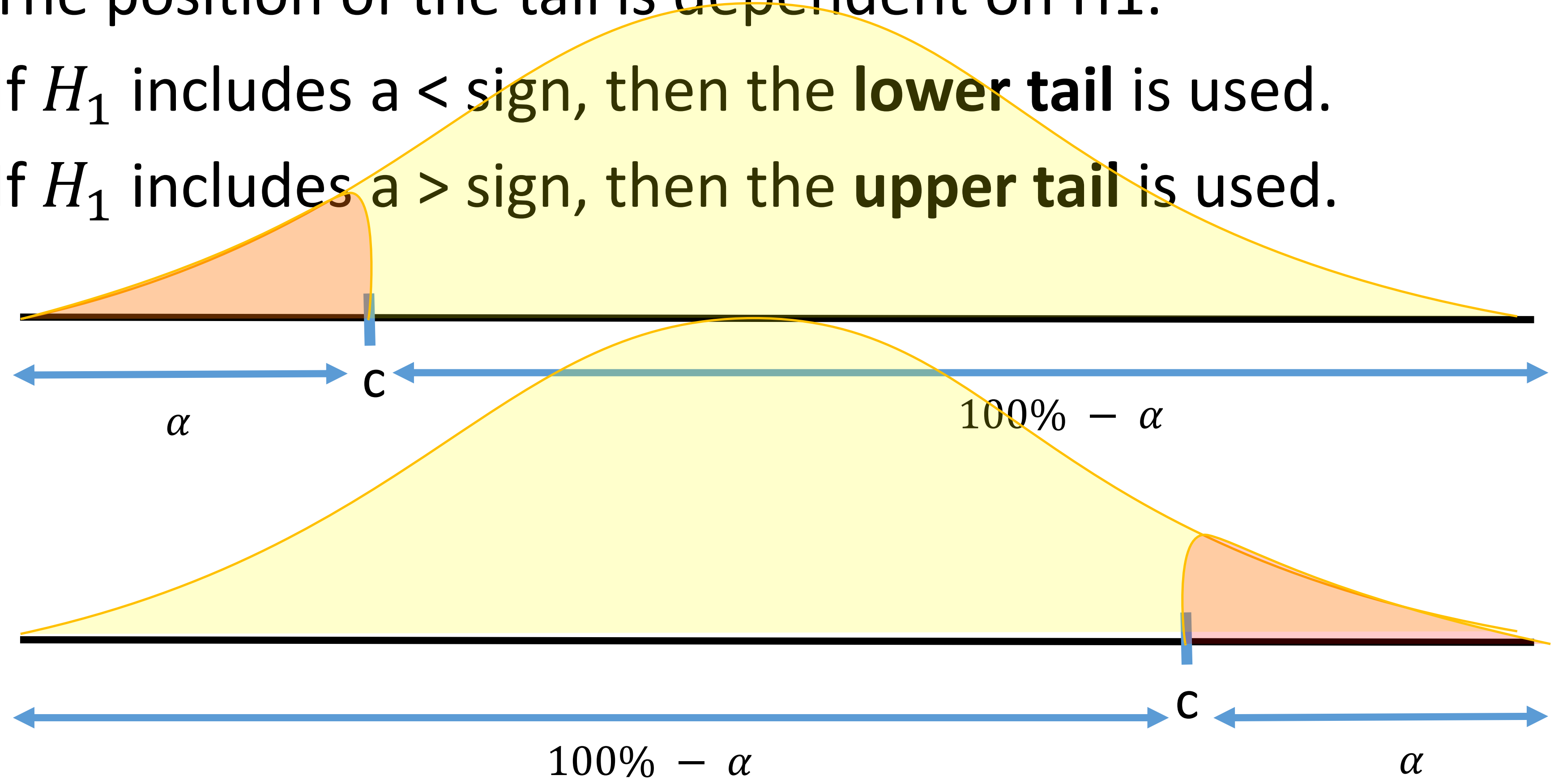
Critical Region Up Close

One-tailed tests

The position of the tail is dependent on H1.

If H_1 includes a < sign, then the **lower tail** is used.

If H_1 includes a > sign, then the **upper tail** is used.

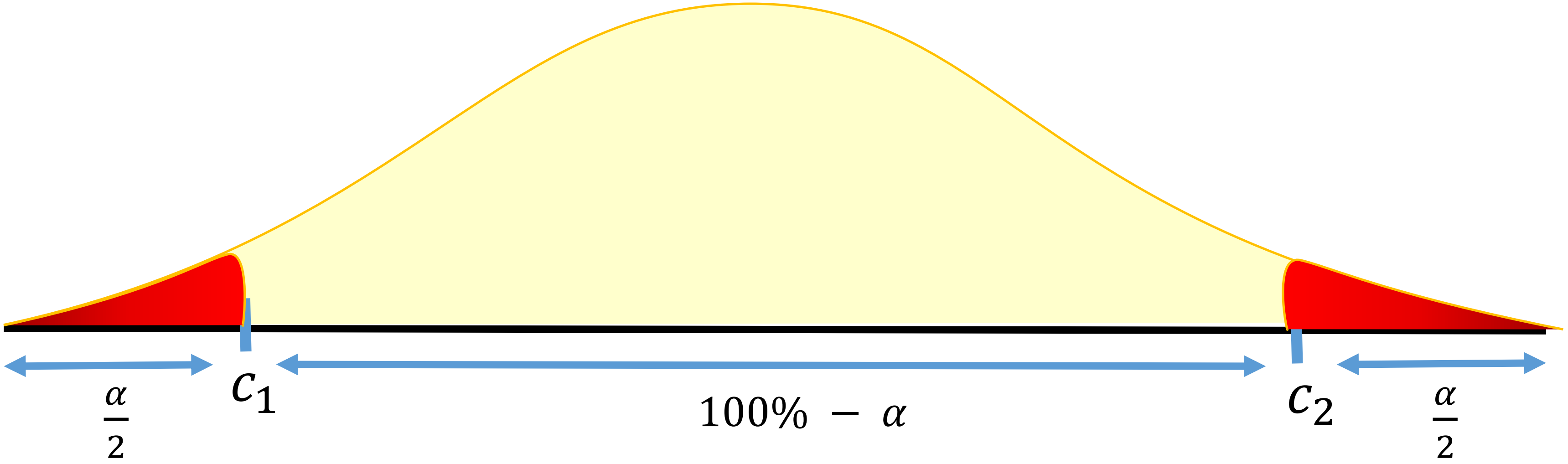


Critical Region Up Close

Two-tailed tests

Critical region is split over both ends. Both ends contain $\frac{\alpha}{2}$, making a total of α.

If H_1 includes a ≠ sign, then the two-tailed test is used as we then look for a change in parameter, rather than an increase or a decrease.



Critical Region Up Close

For each of the scenarios below, identify what type of test you would require.

- Average test score problem as discussed till now.

  One-tailed/Lower-tailed

- If we were checking whether the average is significantly different from 7/10, i.e., H1: $\mu \neq 0.7$.

  Two-tailed test

- The coin is biased.

  Two-tailed test

- The coin is biased towards heads with probability 0.8.

  One-tailed/Upper-tailed

The Missing Link in the Interview

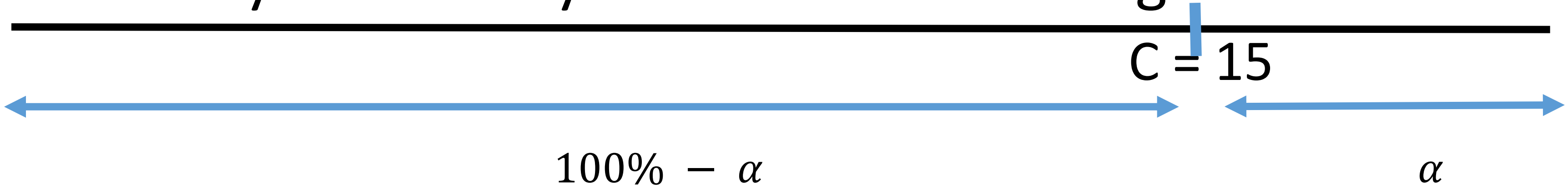
Q. What is the probability of getting 15 or more heads out of 20 coins?

A.

$$P(X \geq 15) = P(X = 15) + P(X = 16) + P(X = 17) + P(X = 18\ ) + P(X = 19) + P(X = 20)$$

$$= 0.021$$

What can you now say about the coin being biased or not?



C = 15

$100\% - \alpha$

α

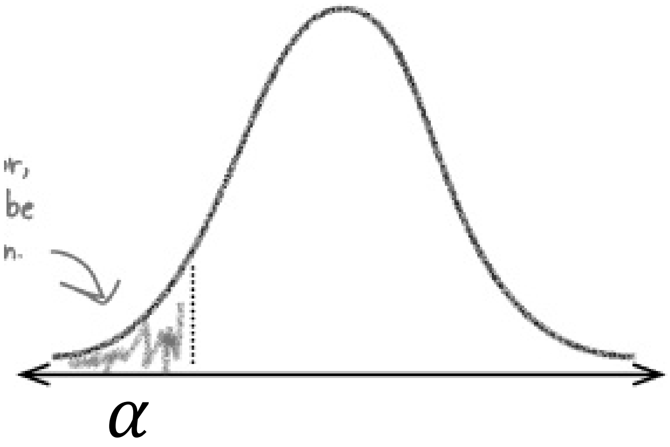
The hypothesis test doesn’t answer the question whether the coin is biased or not; it only states whether the evidence is enough to reject the null hypothesis or not ***at the chosen significance level*****.**

Errors

- Type I: We reject the NULL hypothesis incorrectly
- Type II: We “accept” it incorrectly

		State of Nature	
		Null True	Null False
Action	Fail to reject null (negative)	Correct decision True Negative $P(accept\ H_o \vert H_o True)$	Type II error(β) False Negative $P(Accept\ H_o \vert H_o\ False)$
	Reject null (positive)	Type II error(α) False Positive $P(Accept\ H_o \vert H_o\ True)$	Correct decision (Power) Sensitivity /Recall $P(accept\ H_o \vert H_o False)$

Probability of Getting Type I Error



$$P(Type\ I\ error) = \alpha$$

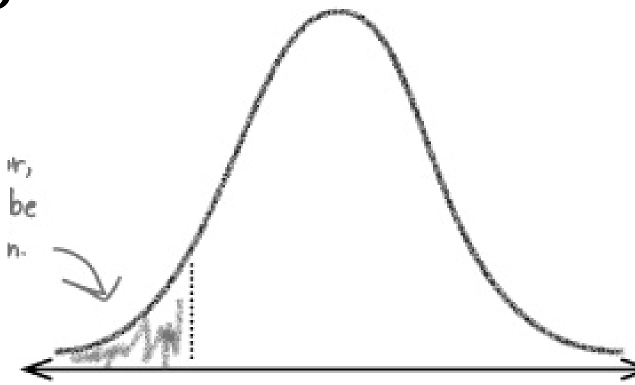
		State of Nature	
		Null True	Null False
Action	Fail to reject null (negative)	Correct decision True Negative $P(accept\ H_o \mid H_oTrue)$	Type II error(β) False Negative $P(Accept\ H_o \mid H_o\ False)$
	Reject null (positive)	Type II error(α) False Positive $P(Accept\ H_o \mid H_o\ True)$	Correct decision (Power) Sensitivity /Recall $P(accept\ H_o \mid H_oFalse)$

Probability of Getting Type II error

$$P(Type\ II\ error) = \beta$$

To find β

1. Check that you have a specific value for H1.



2. Find the range of values outside the critical region of the test. If the test statistic has been standardized, it needs to be de-standardized for the purpose.

3. Find the probability of getting this range of values, assuming H1 is true. In other words, find the probability of getting the range of values outside the critical region, but this time using the test statistic described by H1 and not H0.

A new miracle drug claims that it cures common cold and it has had a success rate of 90%. You conduct a random sample test with 100 patients and you find that 80 of them are cured. At 5% significant level, do you reject or accept the claim by the drug company?

At 5% significant level, do you reject or accept the claim by the drug company?

What are the null and alternate hypotheses ?

$$H_o: p = 0.9$$
$$H_1: p < 0.9$$

What is the test statistics ?

X ~ B(100,0.9)

Since np>5 and nq>5, Normal distribution can be used instead.

X ~ N(np, npq)

X ~ N(90, 9)

What is the probability of 80 or fewer getting cured?

$$Z = \frac{80.5\ - 90}{\sqrt{9}} = -3.17$$

p-value = P(Z < -3.17) = 0.0008

Probabilities of Errors in our Example

P(Type | error) = 0.05

To calculate P(Type II error)

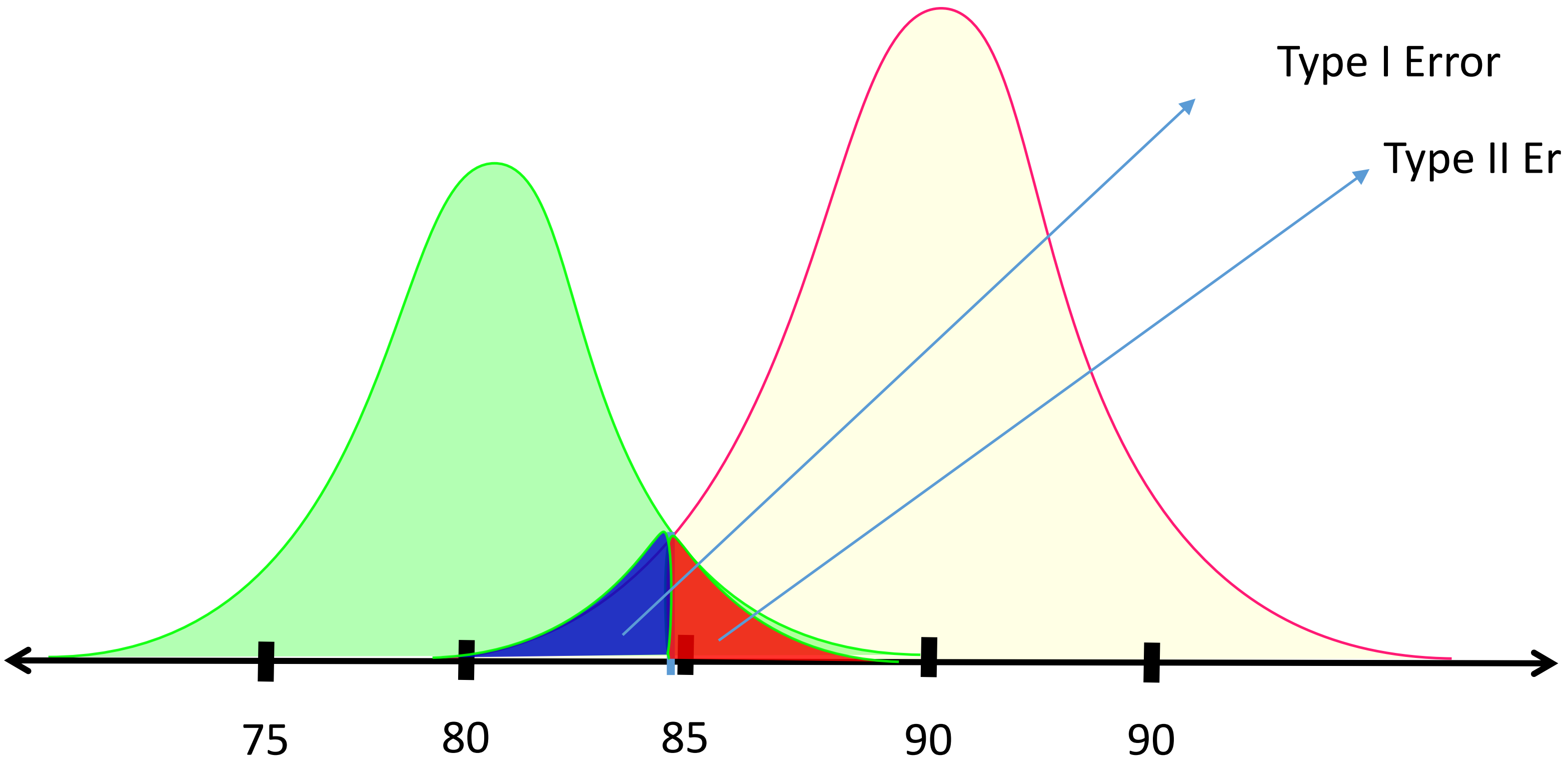
$H_o\colon p = 0.9$
$H_1\colon p = 0.8$

P(Z < c) = 0.05 for 5% Significance value. From probability tables, c = -1.64.

To de-standardize and find values outside the critical region, $\frac{X-90}{\sqrt{9}} \geq -1.64$;

X = 85.08, i.e., we would accept null hypothesis if 85.08 or more people had been cured.

Probabilities of Type I and Type II Errors



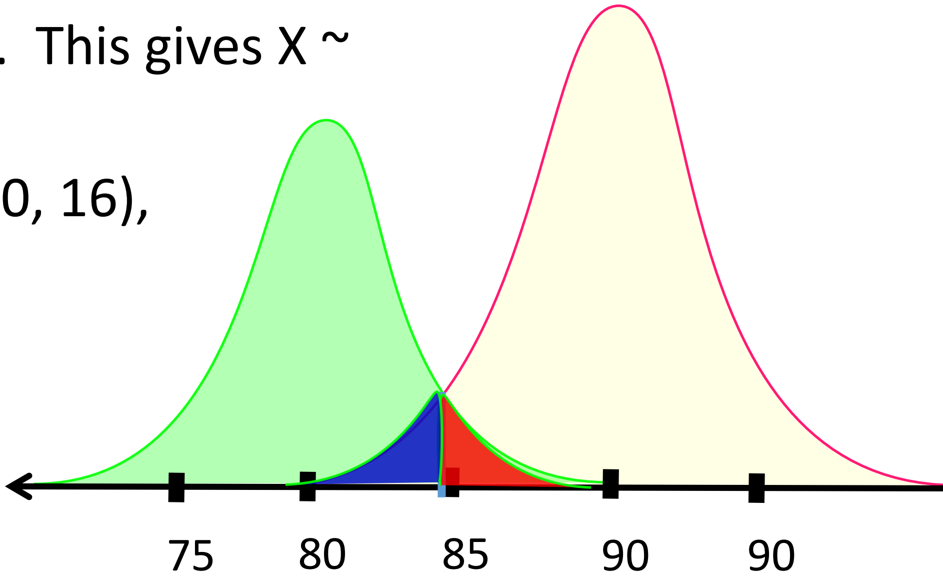
Probabilities of Errors in Our Example

Finally, we need to calculate P(X ≥ 85.08), assuming H1 is true.

X ~ N(np, npq) where n=100 and p=0.8. This gives X ~ N(80, 16).

To calculate P(X = 85.08) where X ~ N(80, 16),

We find



$$z = \frac{85.08\ - 80}{\sqrt{16}} = 1.27$$

P(Z = 1.27) = 1 – P(Z < 1.27) = 1-0.8980 = 0.102

P(Type II error) = 0.102

The probability of accepting the null hypothesis that 90% are cured when its actually 80% is 10.2%.

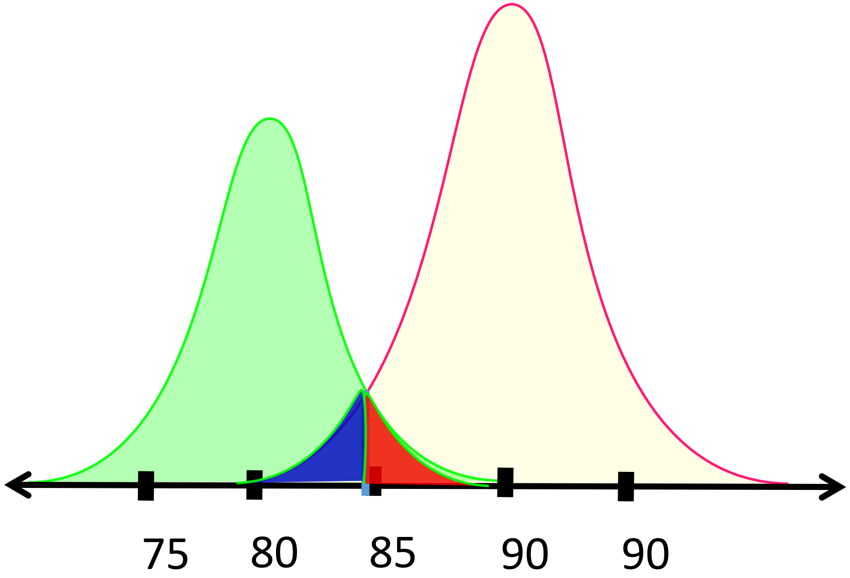
Power of Hypothesis Testing

		State of Nature	
		Null True	Null False
Action	Fail to reject null (negative)	Correct decision True Negative $P(accept\ H_o \vert H_o True)$	Type II error(β) False Negative $P(Accept\ H_o \vert H_o\ False)$
	Reject null (positive)	Type II error(α) False Positive $P(Accept\ H_o \vert H_o\ True)$	Correct decision (Power) Sensitivity /Recall $P(accept\ H_o \vert H_o False)$

Power of Hypothesis Test

We reject null hypothesis correctly when it is false.

It is actually the opposite of Type II error, and therefore,



Power = 1 – ß = 1-0.102 = 0.898, i.e., the probability that we will make the correct decision in rejecting the null hypothesis is 89.8%.

Hypothesis Testing

A prisoner is on trial and you are on the jury. The jury’s task is to assume that the accused is innocent, but if there is enough evidence, the jury needs to convict him.

- In the trial, what is the null hypothesis?

  The prisoner is innocent (or not guilty).

- What is the alternate hypothesis?

  The prisoner is guilty.

Hypothesis Testing

What are the possible ways of the jury coming to an incorrect verdict?

- If the prisoner is innocent, and the jury gives a ‘guilty’ verdict.
- If the prisoner is guilty, and the jury gives an ‘innocent’ verdict.

Which one is Type I and which one Type II?

First one is Type I because null hypothesis actually was correct but rejected incorrectly.

Second one is Type II because null hypothesis was false but was accepted incorrectly.

What is the Power of the test?

Since it is opposite of Type II, it will be finding the prisoner guilty when the prisoner is actually guilty, i.e., rejecting the null hypothesis correctly.

Common Test Statistics for Inferential Techniques

Inferential techniques (Confidence Intervals and Hypothesis Testing) most commonly use 4 test statistics:

- z
- t

Closely related to Sampling Distribution of **Means**

- $\chi 2$ (Chi-squared)
- F

- Closely related to Sampling Distribution of **Variances**
- Derived from Normal Distribution

TWO-SAMPLE *t*-TEST FOR MEANS

- Do two samples come from the same population?
- If they come from different populations, what is the difference in the means of the two populations?

- ✓ Does the average cost of a two-bedroom flat differ between Bengaluru and Hyderabad? What is the difference?
- ✓ What is the difference in the strength of steel produced under two different temperatures?
- ✓ Does the effectiveness of Head & Shoulders anti-dandruff shampoo differ from Pantene anti-dandruff shampoo?
- ✓ What is the difference in the productivity of men and women on an assembly line under certain conditions?
- ✓ Does an antibiotic affect the efficacy of another drug being taken by a patient?

Two-sample t-Test

- Paired Data
  - You have two sets of data, where there is a natural pairing in the elements. Eg: BloodPressure from 30 people – one from before a treatment and other from after treatment

- Unpaired Data
  - Comparing apartment costs from two cities
  - Two data sets of different length
  - No Natural pairing

Two-Sample t-Test for Paired Data

When the effects of two alternative treatments is to be compared, sometimes it is possible to make comparisons in pairs, where, e.g., the pair can be the same person at two different occasions or matched pairs where they are alike in all respects.

To study if their means are the same – we can create a new data set from the difference of the individual data points.

$$X_{new} = X_1 - X_2$$

We can then look at how far away from zero is the mean E(X_{new})

$$t = \frac{\bar{X}_{new} - 0}{SE\ (\bar{X}_{new})}$$

Two-Sample t-Test for Paired Data

A Yoga guru suggests that meditation increases concentration. To test this hypothesis, you get 12 volunteers and get them to complete a puzzle and you measure the time taken for completing the puzzle. The next day, you put them through a 30 minute meditation routine and have them complete another puzzle of similar difficulty. The time taken for completion is measured again.

You want to test at 5% Significance Level (or 95% Confidence Level) if the time taken is shorter after meditation.

Yoga Paired Data

Time to Solve the puzzle(min)			
Patient	After Yoga(A)	Before Yoga (B)	A-B
1	63	55	8
2	54	62	-8
3	79	108	-29
4	68	77	-9
5	87	83	4
6	84	78	6
7	92	79	13
8	57	94	-37
9	66	69	-3
10	53	66	-13
11	76	72	4
12	63	77	-14
TOTAL	**842**	**920**	**-78**
MEAN	**70.17**	**76.67**	**-6.5**

Yoga Paired Data

What are the null and alternate hypotheses?

H0: d = 0

H1: d < 0

One tail test or two tailed test?

One tail test

Significance level

α =0.05

Test statistic?

$t_{n-1,\alpha}$ (for two – tailed we would use $t_{n-1\,,\frac{\alpha}{2}}$

Mean of the difference, $\bar{d} = -6.5$

Standard Deviation of the difference, $s_d = 15.1$

Standard Error of the mean, SE ($\bar{d}$) = $\frac{s_d}{\sqrt{n}} = 4.37$

$$t = \frac{\bar{d}}{SE(\bar{d})} = -\frac{6.5}{4.37} = -1.487$$

Number of degrees of freedom = 12 – 1 = 11

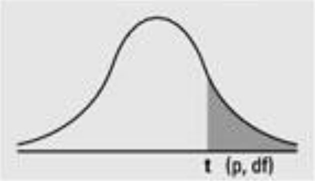
Two-Sample t-Test for Paired Data

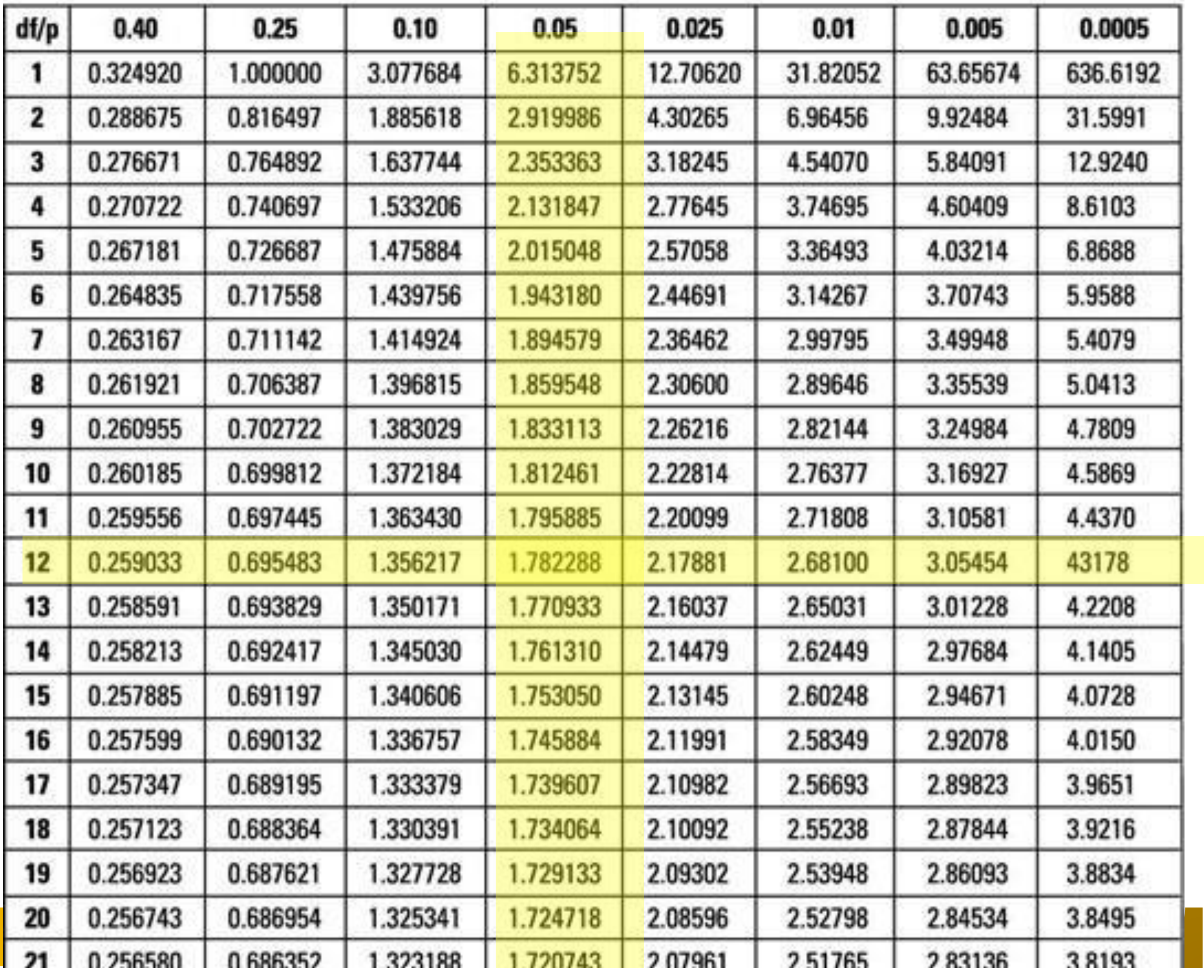
t = -1.487

$t_{11,0.05} = 1.795885$

Comparing the absolute t-value, we **cannot reject the Null** hypothesis that the mean completion time is the same.

Numbers in each row of the table are values on a *t*-distribution with (*df*) degrees of freedom for selected right-tail (greater-than) probabilities (*p*).





df/p	0.40	0.25	0.10	0.05	0.025	0.01	0.005	0.0005
1	0.324920	1.000000	3.077684	6.313752	12.70620	31.82052	63.65674	636.6192
2	0.288675	0.816497	1.885618	2.919986	4.30265	6.96456	9.92484	31.5991
3	0.276671	0.764892	1.637744	2.353363	3.18245	4.54070	5.84091	12.9240
4	0.270722	0.740697	1.533206	2.131847	2.77645	3.74695	4.60409	8.6103
5	0.267181	0.726687	1.475884	2.015048	2.57058	3.36493	4.03214	6.8688
6	0.264835	0.717558	1.439756	1.943180	2.44691	3.14267	3.70743	5.9588
7	0.263167	0.711142	1.414924	1.894579	2.36462	2.99795	3.49948	5.4079
8	0.261921	0.706387	1.396815	1.859548	2.30600	2.89646	3.35539	5.0413
9	0.260955	0.702722	1.383029	1.833113	2.26216	2.82144	3.24984	4.7809
10	0.260185	0.699812	1.372184	1.812461	2.22814	2.76377	3.16927	4.5869
11	0.259556	0.697445	1.363430	1.795885	2.20099	2.71808	3.10581	4.4370
12	0.259033	0.695483	1.356217	1.782288	2.17881	2.68100	3.05454	43178
13	0.258591	0.693829	1.350171	1.770933	2.16037	2.65031	3.01228	4.2208
14	0.258213	0.692417	1.345030	1.761310	2.14479	2.62449	2.97684	4.1405
15	0.257885	0.691197	1.340606	1.753050	2.13145	2.60248	2.94671	4.0728
16	0.257599	0.690132	1.336757	1.745884	2.11991	2.58349	2.92078	4.0150
17	0.257347	0.689195	1.333379	1.739607	2.10982	2.56693	2.89823	3.9651
18	0.257123	0.688364	1.330391	1.734064	2.10092	2.55238	2.87844	3.9216
19	0.256923	0.687621	1.327728	1.729133	2.09302	2.53948	2.86093	3.8834
20	0.256743	0.686954	1.325341	1.724718	2.08596	2.52798	2.84534	3.8495
21	0.256580	0.686352	1.323188	1.720743	2.07961	2.51765	2.83136	3.8193

Two-Sample t-Test for Paired Data

The 95% CI for mean difference is given by $\bar{d} \pm t_{n-1,\alpha} \ * SE(\bar{d})$

$$\Rightarrow -6.5 \ - 1.796 \ * 4.37 \leq D \leq -6.5 + 1.796 \ * 4.37$$

95% CI : (-14.35, 1.35)

As zero is included in the CI, we cannot reject the null hypothesis.

Business Decision (Yogic Decision?)

- Although zero is included in CI, the range is very wide, which should lead the us to conduct a larger study to be sure.

Two sample t-Test: unpaired data

The Central Limit Theorem states that the difference in two sample means, $\bar{x}_1 - \bar{x}_2$, is normally distributed for large sample sizes (both n_1 and $n_2 \geq 30$) whatever the population distribution

Also, $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$ *[Recall E(X-Y)=E(X)-E(Y)]*

And $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$ *[Recall E(X-Y)=E(X)-E(Y)]*

$$z = \frac{observed\ difference - expected\ difference}{SE\ of\ the\ difference} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

This is the test statistic for a 2-sample z-test.

Two-Sample t-Test for Unpaired Data

$H_0: \mu_1 = \mu_2 \, ; H_0: \mu_1 \neq \mu_2$

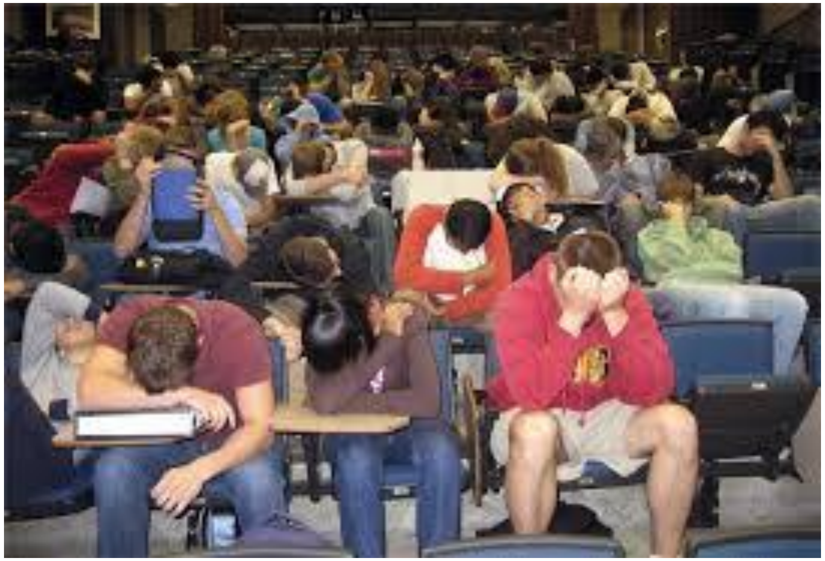
Test statistics, $t = \frac{\overline{x_1} - \overline{x_2}}{SE}$

Assuming the two samples come from populations with the same **standard deviation** (Rule of thumb: The ratio between the higher *s* and the lower s is less than 2), pooled variance can be used to calculate SE.

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}$$

$t = \frac{\overline{x_1} - \overline{x_2}}{SE}$ with $(n_1 + n_2 \; - 2)$ degrees of freedom

Insomnia Treatment



A statistics professor claims that his lectures can cure insomnia. You want to test the claim. You collect 30 patients with sleeping trouble and divide them into 2 groups of 15 each.

The control group were asked to follow their usual routine while the other group was exposed to 1-hour of his lecture on t-distribution shortly after dinner. The time taken to sleep was measured for each group.

Two sample t-test

Time taken to sleep (hr)					
Treated Subjects			Control Subjects		
0.81	0.56	0.46	1.15	1.15	0.92
1.06	0.45	0.43	1.28	0.72	0.67
0.43	0.88	0.37	1.00	0.79	0.76
0.54	0.73	0.73	0.95	0.67	0.82
0.68	0.43	0.93	1.06	1.21	0.82

$$n_2 = 15$$
$$\overline{x_2} = 0.633$$
$$s_2 = 0.216$$
$$s_2^2 = 0.0467$$

$$n_2 = 15$$
$$\overline{x_2} = 0.931$$
$$s_2 = 0.202$$
$$s_2^2 = 0.0408$$

Hypothesis Testing

What is the null hypothesis?

$H0$: μ_1- μ_2 = 0 (The lecture has no impact)

What is the alternative hypothesis?

$H1$: μ_1- $\mu_2 \neq 0$

Is it a one-tailed test or a two-tailed test?

Two-tailed

What could be a possible hypothesis for a one-tailed test?

The lecture helps people sleep better.

Two sample t-test

At α = 0.05, determine if there is a significant difference between the groups.

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}\,;\; t = \frac{\overline{x_1} - \overline{x_2}}{s_p\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \text{ with } (n_1 + n_2 - 2)df.$$

$$s_p^2 = \frac{(15 - 1) * 0.0408 + (15 - 1) * 0.0467}{(15 - 1) + (15 - 1)} = 0.04375; s_p = 0.209$$

$$t = \frac{(0.931 - 0.633)}{0.209 * \sqrt{\frac{1}{15} + \frac{1}{15}}} = 3.91$$

You can find the p-value for this t-score or knowing that the t-score is way more than the critical value for 28 df (~ 2) at this significance level, you see that it is in the critical region in the right tail.

Hypothesis Testing

Will you reject the null hypothesis or fail to do so?

Reject. That means lecture does affect the time-to-sleep.

Does it increase or decrease the time to sleep and by how much?

As the treated patients slept in shorter time (0.633 hr) compared to the control group (0.931 hr), the lecture reduces the time to sleep by 0.298 hr.

Reference

INSOFE: www.insofe.edu.in