Statistics

Part 3

Accuracy V/s Precision

- In order to get higher precision as well as higher accuracy, increase sample size. It reduces SE and hence ME. Therefore we can remain at a high confidence level while not needing to increase the Confidence Interval

Example

- A survey collected responses from 1154 US residents. Based on it, a 95% CI for the average no. of hours the Americans have to relax or pursue activities that they enjoy after an average work day was [3.53,3.83] hours.
- What is the point estimate of this sample?
- 3.68
- If we take 1000 such samples than in how many samples you will get true average number of hours Americans spend relaxing after a work day?
- 950

backtracking to n for a given ME

given a target margin of error, confidence level, and information on the variability of the sample (or the population), we can determine the required sample size to achieve the desired margin of error.

$$\boxed{ME = z^{\star} \frac{s}{\sqrt{n}} \quad \rightarrow n = \left(\frac{z^{\star} s}{ME} \right)^2}$$

Example

- A group of researchers want to test the possible effect of an epilepsy medication taken by pregnant mothers on the cognitive development of their children. As evidence, they want to estimate the IQ scores of 3-yr old children born to mothers who were on this medication during pregnancy. Previous studies suggest that the SD of IQ scores is 18 points.

- How many such children should the researchers sample in order to obtain a 90% confidence interval with a margin of error $\leq$ 4 points?

Example contd..

- ME = 4 points, For CL = 90%, Z* = 1.65, s = 18
- Hence n = 55.13, i.e. we need at least 56 such children in the sample to obtain a maximum ME of 4 points.

- If we further wish to decrease ME by 2 points, how would the required sample size change?
- Ans. To reduce ME to half, we need to quadruple our sample size.
- Increasing sample size takes resources

Example

- A given Confidence Interval is calculated, based on a random sample of ‘n’ observations. If we want to make this interval narrower, say 1/3rd of present size, how many observations should we sample?

a. (1/9n)

b. (1/3n)

c. 3n

d. 4n

e. 9n

Ans. ‘e’

- The General Social Survey (GSS) asks ‘for how many days during the past 30 days was your mental health not good?’. Based on the responses of 1151 US residents, the survey reported a 95% Confidence Interval of 3.40 to 4.24 days.
- Thus we are 95% confident that Americans on an average , have 3.40 to 4.24 bad mental health days per month.
- If a new survey was asking the same question to 500 Americans instead of 1151 Americans, would the standard error of the estimate be larger or smaller?
- Ans. Larger

Example

- A sample of 50 college students were asked, how many course projects they have done on their own so far. The students in the sample had an average of 3.2 projects with a SD of 1.74. Estimate the true average number of projects based on this sample, using a 95% Confidence Interval.
- SE=1.74/sqrt(50)=0.246
- 95%CI=3.2 ± 1.96(0.246)=[2.72,3.68]
- Meaning that we are 95% confident that college students on an average have done 2.72 to 3.68 course projects on their own.

Example

- Create a 99% confidence interval for the average days active per week of all VIT students using vit_csb sample. The point estimate is 3.75 and SE is 0.26.
- Solu: For 99% CI: point estimate ± 2.58*SE
- =3.75 ± 2.58(0.26)
- =(3.08,4.42)

Case Study:Promotion

Gender	Promoted	Not Promoted	Total
Male	21	3	24
Female	14	10	24
TOTAL	35	13	48
% Promotion	88%	58%	

simulation scheme

[use a deck of playing cards to simulate this experiment]

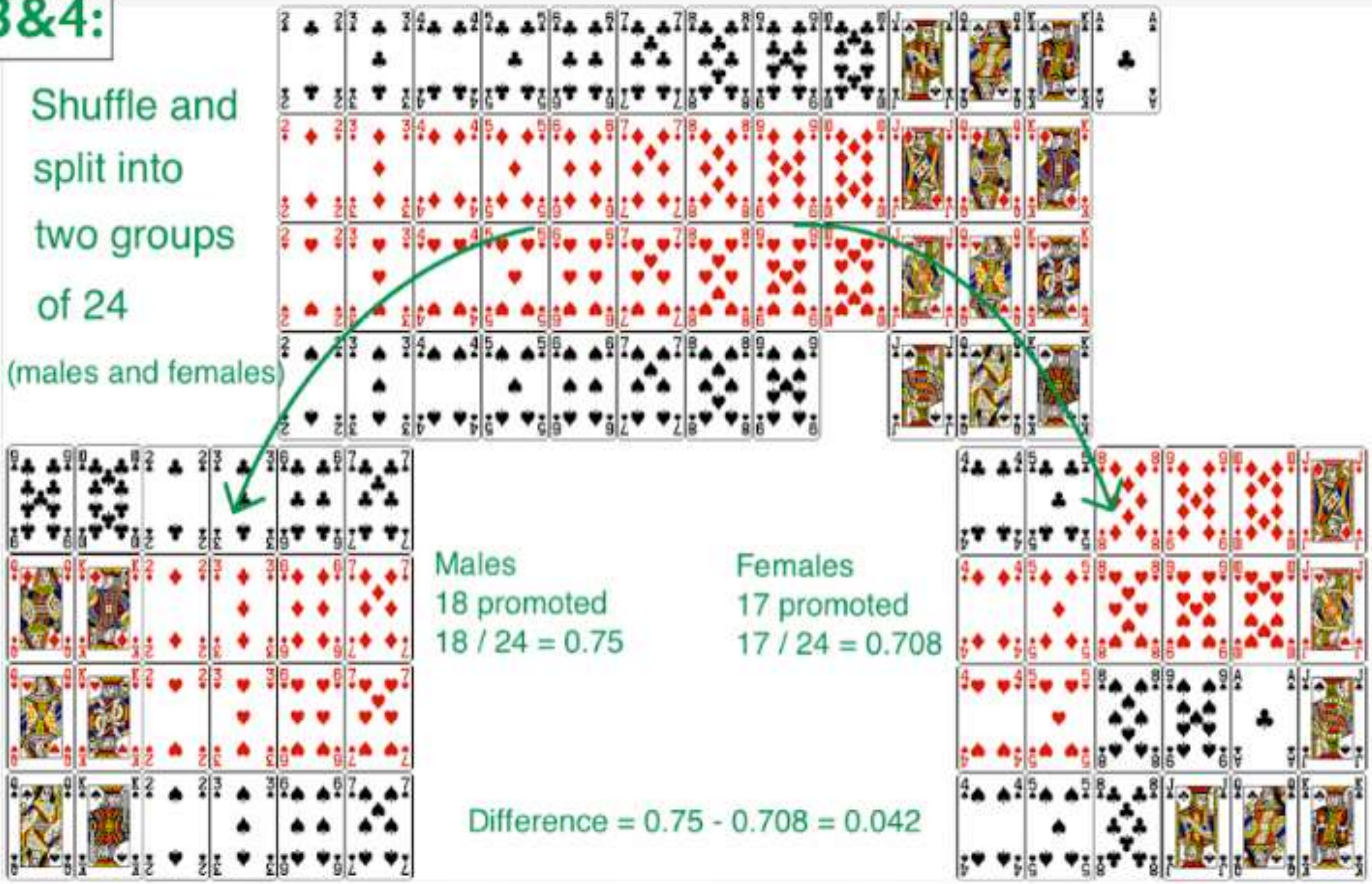
1. face card: not promoted, non-face card: promoted
   - set aside the jokers, consider aces as face cards
   - take out 3 aces → 13 face cards left in the deck (face cards: A, K, Q, J)
   - take out a number card → 35 number (non-face) cards left in the deck (number cards: 2-10)

Step 1:

Steps 3&4:

Shuffle and split into two groups of 24

(males and females)



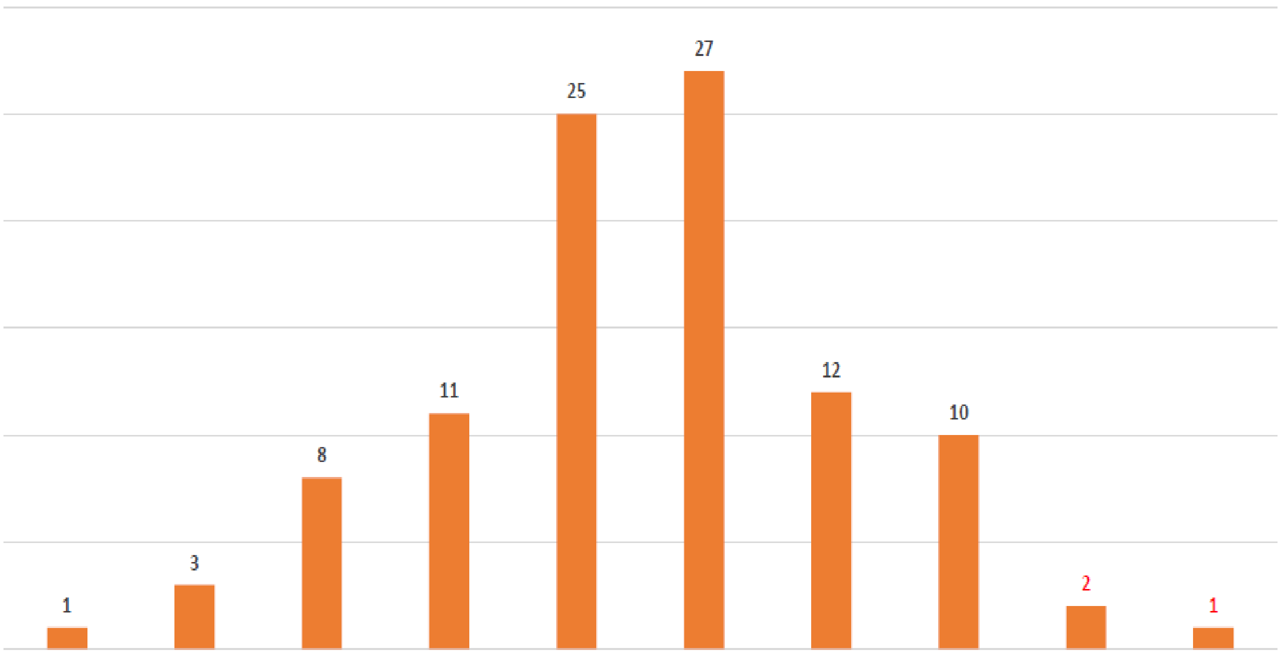
Males
18 promoted
18 / 24 = 0.75

Females
17 promoted
17 / 24 = 0.708

Difference = 0.75 - 0.708 = 0.042

Case study Contd..

- H0: There is no gender discrimination
- Ha: There is gender discrimination.
- Simulation was carried out and following facts were observed:



P-value

- P-value: The probability of observing data at least as extreme as the one observed in the original study, under the assumption that the null hypothesis is true , is called the p-value. It is the numerical criteria used for making decisions between competing hypotheses.
- In given case,21M–14F=7difference promotion cases, 3 instances are reported out of simulation. Hence the p-value is 0.03
- Keeping this value in mind, let us focus on following:
- Hypothesis testing is built around rejecting or failing to reject null hypothesis unless we have strong evidence.
- But what this strong evidence means? We define a term called as significance level(α). As a rule of thumb we consider α=0.05 or 5%.

P-value Contd…

Thus, If the p-value is lower, it means that it corresponds to sufficient evidence to reject Ho in favor of HA.

i.e

- P-value < α → Reject null hypothesis
- P-value > α →Fail to reject null hypothesis

Algorithm for Hypothesis Testing

- 1.Set the hypothesis,Ho:μ=null value;Ha:μ< or > or ≠nullvalue
- 2.Calculate the point estimate–sample mean ,(difference in mean)
- 3.Check for applicability conditions Similar to CLT i.e. independence (random sample for obs.Study/assignment for experiment) & sample size/skew. We consider this is met with each time.
- 4.Draw sampling distribution,shade p-value, calculate test statistic Z*
- 5.Make a decision and interpret it in the context of research question

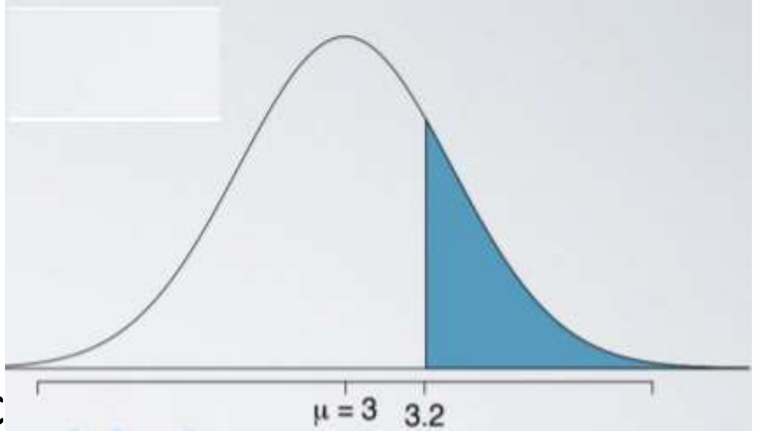
Example

- A sample of 50 college students were asked, how many course projects they have done on their own so far. The students in the sample had an average of 3.2 projects with a SD of 1.74. The project coordinator claims that the average mean of whole college is 3. Few faculty thinks that it is greater than 3.Verify the hypothesis. Consider 5% significance level.
- Solution: H0=3 , HA>3
- SE=1.74/sqrt(50)=0.246.

Test Statistics(Z score)= (3.2-3)/0.246

= 0.81

p-value=0.209



- Thus p value is greater than 0.05 hence we fail to rejec
hence the claim is correct.

Decision Based on p-value

- We used the test–statistic to calculate the p-value, which is the probability of observing data at least as favorable to the alternative hypothesis as our current dataset, if the null hypothesis was true.
- If the p-value is low–lower than a defined significance levelα, which is usually 5%, we say that it would be very unlikely to observe the data, if the null hypothesis were true and hence reject null hypothesis.
- If the p-value is high–(Higher than significance level α), we say that it is likely to observe the data even if then the null hypothesis were true, and hence do not reject the null hypothesis.
- Since p-value is 0.209,we do not reject the null hypothesis.

Example: College student school hours

- A poll by the National Sleep Foundation found that college students average about 7 hours of sleep per night. Researchers at a rural school are interested in showing that students at their school sleep longer than seven hours on average, and they would like to demonstrate this using a sample of students.
- The researchers at the rural school conducted a simple random sample of $n = 110$ students on campus. They found that these students averaged 7.42 hours of sleep and the standard deviation of the amount of sleep for the students was 1.75 hours.
- Consider significance value as 0.05, find whether the claim is true?

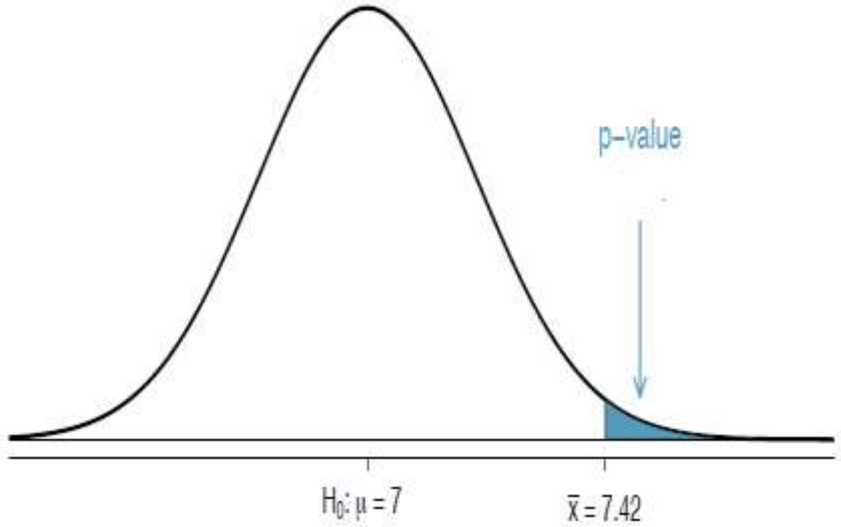
Solution Contd…

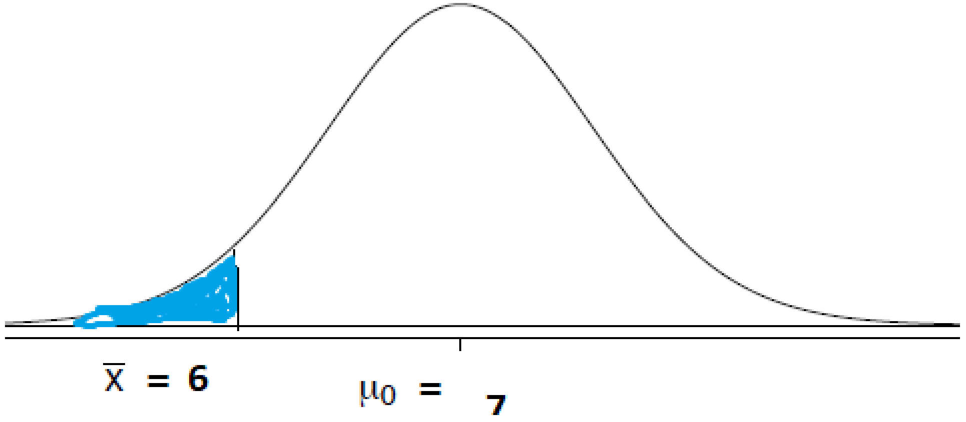
- Solu: H0: $\mu = 7$.   HA: $\mu > 7$.
- $SE=SD/\sqrt{n}= 1.75/ \sqrt{110} = 0.167$
- Test statistics, Z-score=$(7.42-7)/0.167=2.51$
- P-value=$1-0.9940=0.006$
- Since p-value<0.05: We reject the null hypothesis, i.e the average mean of sleep is not 7 hours.

One sided test

- Till now we have seen one sided test. Here we shade the single tail in the direction of the alternative hypothesis.
- Let H0: μ = 7.
- i. HA: μ > 7. p-value represented by upper tail.

  ii. HA: μ < 7 .p-value represented by lower tail.





Two-sided test

- In two sided test we shade two tails.
- E.g: Earlier we talked about a research group investigating whether the students at their school slept longer than 7 hours each night. Let's consider a second group of researchers who want to evaluate whether the students at their college differ from the norm of 7 hours. Write the null and alternative hypotheses for this investigation.
- H0: $\mu = 7$.   HA: $\mu \neq 7$. It means that $\mu > 7$ or $\mu < 7$
- This gives us two sided test. We shade the area in both the direction.

Example : The second college research group randomly samples 122 students and finds a mean of $\bar{x} = 6.83$ hours and a standard deviation of $s = 1.8$ hours. Does this provide strong evidence against H0? Use a significance level of $\alpha = 0.05$.

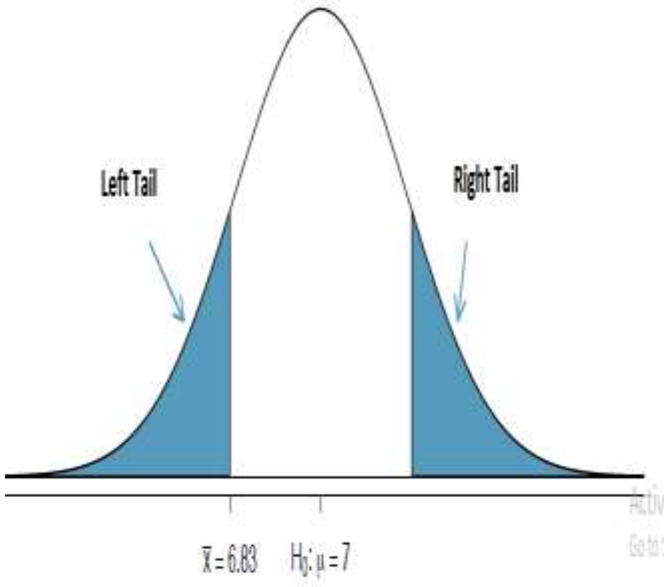
- $SE=SD/\sqrt{n}=1.8/\sqrt{122}=1.6$
- $Z =(6.83 - 7.00)/0.16= -1.06$
- left tail = 0.1446
- Because the normal model is symmetric, the right tail will have the same area as the left tail. The p-value is found as the sum of the two shaded tails:

p-value = left tail + right tail = 2x(left tail) = 0.2892

This p-value is relatively large (larger than

$\alpha = 0.05$), so we should not reject H0.

- Thus, we do not have sufficient evidence to conclude that the mean is different than 7 hours.



Example

- Colleges frequently provide estimates of student expenses such as housing. A consultant hired by a community college claimed that the average student housing expense was $650 per month. What are the null and alternative hypotheses to test whether this claim is accurate?

Solu: $H0: \mu=\$650$

$Ha: \mu\neq\$650$

- A study suggests that the average college student spends about 2 hours per week communicating with others online. You believe that this is an underestimate and decide to collect your own sample for the hypothesis test. You randomly sample 60 students from your Institute and find that on average, they spend 3.5 hours a week communicating with others online. What are the appropriate hypothesis ?
- a.Ho:X=2Hrs./wk;Ha:X>2Hrs./wk
- b.Ho:X=2Hrs./wk;Ha:X>3.5Hrs./wk
- c.Ho:μ=2Hrs./wk;Ha:μ>2Hrs./wk
- d.Ho:μ=2Hrs./wk
- Ans. ‘c’

- Chain restaurants have been required to display calorie count of each menu item. Prior to displaying calorie counts, the average calorie intake of diners at a restaurant was 1100 calories. After calorie counts started to be displayed on menus, a nutritionist collected data on the number of calories consumed at this restaurant from a random sample of diners. Which of the following would be used to test whether there was evidence of a difference in the average calorie intake of diners at this restaurant?
- A.Ho: $x = 1100$ calories; Ha: $x \neq 1100$ calories
- b.Ho: $x = 1100$ calories; Ha: $x < 1100$ calories
- c.Ho: $\mu = 1100$ calories; Ha: $\mu \neq 1100$ calories
- d.Ho: $\mu < 1100$ calories; Ha: $\mu > 1100$ calories
- Ans. ‘c’

Example

- P-value of approximately 0 provides –

a. Weak evidence against null hypothesis

b. Strong evidence against null hypothesis

c. Strong evidence against alternative hypothesis

d. Weak evidence against alternative hypothesis

e. No evidence against the null hypothesis

Ans: b

Decision Errors

		Decision	
		Fail to reject Ho	Reject Ho
Truth	Ho true	OK	Type 1 error
	Ha true	Type 2 error	OK