

PS06

Sejal Nimkar

2024-11-18

Q1) Assume that the hypotheses of a test are given by $H_0 : \mu \leq 14$ vs $H_1 : \mu > 14$ and we know $\sigma = 6$.

a. Assume that we obtain a sample of size $n = 6$ and $\bar{x} = 15.2$. Perform a test with $\alpha = 0.1$. What is your conclusion?

Ans: We shall compute the value of the test statistic and insert it in the code. Since the standard deviation is given, we shall use pnorm instead of pt.

The test statistic is given as follows:

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

```
# Inputs
mean <- 15.2 # Sample mean
n <- 36      # Sample size
std <- 6     # Population standard deviation
mu <- 14     # Null hypothesis mean

# Calculate the z-score
z <- (mean - mu) / (std / sqrt(n))

# Calculate the p-value for a one-tailed test
p_value <- 1 - pnorm(z)

# Print the p-value
p_value
```

```
## [1] 0.1150697
```

The p_value is approximately 0.12 which exceeds $\alpha = 0.1$. Thus we shall fail to reject the null hypothesis.

b. What is the sample mean that corresponds precisely to the boundary of the significance level? (i.e., the area under the PDF of \bar{X}_n to the right of this sample mean has to be 0.1).

Ans: To find the sample mean corresponding to the boundary of the significance level, we need to identify the critical value of the test statistic z where $\alpha = 0.1$

We shall find the value of z where $\alpha = 0.1$ such that $P(Z > z) = 0.1$. This can be found using `qnorm()` function.

```
z <- qnorm(1 - 0.1)
z
```

```
## [1] 1.281552
```

The z value where $P(Z > z) = 0.1$ is approximately 1.28

We shall back-solve for \bar{x} by inserting the value of z we found out earlier in the formula for test statistic:

$$z = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$\therefore \bar{x}_{\text{crit}} = z \times \frac{\sigma}{\sqrt{n}} + \mu_0$$

Substituting the values, we can find out the value of \bar{x}_{crit}

```
z<- 1.28
std<- 6
n<-36
mu<- 14

critical_mean <- z * (std / sqrt(n)) + mu
critical_mean
```

```
## [1] 15.28
```

The sample mean corresponding to the boundary of the significance level is approximately 15.28.

c. Now we learn that actually $\mu = 15$. Have you committed a Type I or Type II error or made a correct decision?

Ans: The hypothesis defined earlier were :

Null Hypothesis: $H_0 : \mu \leq 14$

Alternate Hypothesis: $H_1 : \mu > 14$

And the decision that was taken earlier was that we failed to reject the null hypothesis. Since the null hypothesis was actually false and we failed to reject it means that I committed a Type 2 error when the alternative hypothesis was true

d. Use the value obtained in part b, alongside the true mean $\mu = 15$ to get β , the probability of committing a type II error

Ans: The probability of committing a type 2 error can be denoted by the following:

$$\beta = P(\bar{x} < \bar{x}_{\text{crit}} | \mu = 15).$$

We have the following values :

$$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

Substitute the following values:

Critical value of mean: $\bar{x}_{\text{crit}} = 15.28$

True mean: $\mu = 15$

Sample size : $n = 36$

Standard deviation: $\sigma = 6$

```
# Inputs
mean <- 15.28 # Sample mean
n <- 36       # Sample size
std <- 6      # Population standard deviation
mu <- 15      # Null hypothesis mean

# Calculate the z-score
z <- (mean - mu) / (std / sqrt(n))
z
```

```
## [1] 0.28
```

The probability of $\bar{X} < \bar{x}_{\text{crit}}$ corresponds to the cumulative probability up to $Z = 0.28$:

$$\beta = P(Z < 0.28)$$

```
beta <- pnorm(0.28)
beta
```

```
## [1] 0.6102612
```

The probability of committing a Type II error β is approximately 0.61.

e. What is the power of the test?

Ans: The power of the test is defined as :

$$\text{Power} = 1 - \beta$$

where β is the probability of a Type 2 error.

$$\therefore \text{Power} = 1 - 0.61 = 0.39$$

Q2) Do the following

a. Using the sample obtained in PS05 question 4c, get a 94% confidence interval for the average arrival delay for NY flights.

Ans: Since we do not know the variance of the entire population, we shall use the sample variance. Thus we shall use t distribution. The confidence interval is given as :

$$I = \left(\bar{x} - q \frac{s_n}{\sqrt{n}}, \bar{x} + q \frac{s_n}{\sqrt{n}} \right)$$

For a 94% confidence interval, the α value is the complement of the confidence level:

$$\alpha = 1 - \text{Confidence Level}$$

$$\therefore \alpha = 1 - 0.94 = 0.06$$

Ans q is the critical value of the test statistic i.e. $1 - \frac{\alpha}{2}$ quantile of Normal(0,1)

$$\text{i.e. } 1 - \frac{0.06}{2} = 0.97^{\text{th}} \text{ quantile of Normal}(0,1)$$

```
# Load required library
library(nycflights13)

# Set seed for reproducibility
set.seed(100)

# Create a vector of arrival delays without missing values
arr_delay_vector <- na.omit(flights$arr_delay)

# Take a sample of 150 arrival delays
sample_arr_delay <- sample(arr_delay_vector, 150)

# Calculate sample mean, standard deviation, and sample size
sample_mean <- mean(sample_arr_delay)
sample_sd <- sd(sample_arr_delay)
n <- length(sample_arr_delay)

# Print the sample mean and standard deviation
cat("Sample mean:", sample_mean, "\n")
```

```
## Sample mean: 7.906667
```

```
cat("Sample standard deviation:", sample_sd, "\n")
```

```
## Sample standard deviation: 49.32352
```

```
# Calculate the critical value for a 94% confidence interval
alpha <- 0.06
q <- qt(1 - alpha / 2, df = n - 1)
cat("The critical value of test statistic:", q, "\n")
```

```
## The critical value of test statistic: 1.895221
```

```
# Calculate the confidence interval
lower_boundary <- sample_mean - q * sample_sd / sqrt(n)
upper_boundary <- sample_mean + q * sample_sd / sqrt(n)

# Display the confidence interval
cat("Confidence Interval I:", c(lower_boundary, upper_boundary), "\n")
```

```
## Confidence Interval I: 0.2741415 15.53919
```

The 94% confidence interval for the average arrival delay of NY flights, based on the sample, is between 0.27 minutes and 15.53 minutes.

b. What sample size is needed to get a 94% confidence interval that is only 5 minutes long? Assume that $\sigma = 58$

Ans: The formula to find the sample size for a 94% confidence interval is :

$$n = \left(\frac{2q\sigma}{L} \right)^2$$

Here, L is the length of the confidence interval which is % minutes long.

Standard deviation $\sigma = 58$

We will find the value of q using qnorm now that we know the standard deviation is 58

```
sigma<- 58
L<- 5

# Critical value
q <- qnorm(1 - alpha / 2)

# Calculate required sample size
n<- (2*q*sigma/L)^2

cat("Required sample size n:", ceiling(n), "\n")
```

```
## Required sample size n: 1904
```

c. Using the same sample, obtain a 96% confidence interval for the proportion of NY flights without arrival delays.

Ans: We shall use the following formula since we are to use a proportion of the sample:

$$I = \left(\hat{p} - q \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + q \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \right)$$

where:

\hat{p} represents the sample proportion.

q is the critical value from the standard normal distribution.

n is the sample size.

The square root term calculates the standard error for proportions.

```
# Calculate the sample proportion of flights without delays
p_hat <- mean(sample_arr_delay <= 0) # Proportion of flights without delays
n <- length(sample_arr_delay) # Sample size

# Calculate the critical value for a 96% confidence interval
alpha <- 0.04
q <- qnorm(1 - alpha / 2)

# Confidence interval
lower_boundary <- p_hat - q * sqrt((p_hat * (1 - p_hat)) / n)
upper_boundary <- p_hat + q * sqrt((p_hat * (1 - p_hat)) / n)

# Print results
cat("Sample proportion (p_hat):", p_hat, "\n")

## Sample proportion (p_hat): 0.6066667

cat("96% Confidence Interval for the proportion:", c(lower_boundary, upper_boundary), "\n")

## 96% Confidence Interval for the proportion: 0.5247529 0.6885805
```

Q3) Answer the following questions for the problems given:

- (1) What is the experimental unit?
- (2) From how many populations were the experimental units drawn? Identify the population(s). How many units were drawn from each population? Is this a 1- or a 2-sample problem?
- (3) How many measurements were taken on each experimental unit? Identify them.
- (4) Define the parameter(s) of interest for this problem. For 1- sample problems, this should be μ ; for 2-sample problems, this should be Δ .
- (5) State appropriate null and alternative hypotheses.

a. Irmina, a professional massage/physical therapist and ski instructor, decides to moonlight as an aerobics instructor. Her supervisor recommends that she begin each class with 10 minutes of static stretching, but Irmina believes that static stretching is detrimental to athletic performance. She devises an experiment, for which she recruits 20 aerobics students, that consists of two protocols. In protocol S, a participant walks for 5 minutes, then does 10 minutes of static stretches of the hamstring, quadricep, and calf muscles, then rides a stationary bike for 30 minutes. Protocol D replaces static stretches with dynamic stretches. Each bike is equipped with a heart monitor and the ability to measure watts of power expended. To equalize level of exertion, each participant is asked to maintain a constant training heart rate calculated using the Karvonen formula 5 with an intensity of 0.80. The study participants perform protocol D one week and protocol S the following week. Irmina records the number of watts expended during each 30-minute ride. How might she use the resulting data to persuade her supervisor that dynamic stretching is superior to static stretching?

Ans: 1) The experimental unit in the experiment is an aerobics student participating in the experiment.

- 2) The experimental units were drawn from a single population of aerobics students. There are 20 experimental units (20 participants). This is a 1-sample problem.
- 3) Number of measurements per unit: 2 (one for each protocol). Watts expended during protocol S. Watts expended during protocol D.
- 4) This is a paired problem, so the parameter of interest is the mean difference Δ in watts expended between protocol S and protocol D.

$\Delta = \mu_D - \mu_S$, where:

μ_D : Mean watts expended during protocol D.

μ_S : Mean watts expended during protocol S.

- 5) Null hypothesis (H_0): There is no difference in the mean watts expended between protocol D and protocol S. $H_0 : \Delta = 0$ or $\mu_D = \mu_S$

Alternative hypothesis (H_1): The mean watts expended during protocol D is greater than that during protocol S. $H_1 : \Delta > 0$ or $\mu_D > \mu_S$

b. A shoe company claims that wearing its racing flats will typically improve one's time in a 10K road race by more than 30 seconds. A running magazine sponsors an event to test this claim. It arranges for 120 runners to enter two road races, held two weeks apart on the same course. For the second race, each of these runners is supplied with the new racing flat. How might the race results be used to determine the validity of the shoe company's claim?

Ans: 1) The experimental unit is the individual runner, as the data is collected on each runner based on their performance in the two races.

2) The runners are drawn from the same population (participants in the two races). There are 120 experimental units (120 runners). This is a 1-sample paired problem, as each runner participates in both races.

3) Each experimental unit has 2 measurements:

Time to complete the first race (without the new racing flats).

Time to complete the second race (with the new racing flats).

4) This is a paired problem, so the parameter of interest is the mean difference (Δ) in race times between the first and second races.

$$\Delta = \mu_1 - \mu_2,$$

where:

μ_1 : Mean time to complete the first race (without racing flats).

μ_2 : Mean time to complete the second race (with racing flats).

5) Null hypothesis (H_0): The mean improvement in race times is at most 30 seconds. $H_0 : \Delta \leq 30$ or $\mu_1 - \mu_2 \leq 30$

Alternative hypothesis (H_1): The mean improvement in race times exceeds 30 seconds. $H_1 : \Delta > 30$ or $\mu_1 - \mu_2 > 30$

c. Susan theorizes that impregnating wood with an IGR (insect growth regulator) will reduce wood consumption by termites. To investigate this theory, she impregnates 60 wood blocks with a solvent containing the IGR and 60 wood blocks with just the solvent. Each block is weighed, then placed in a separate container with 100 ravenous termites. After two weeks, she removes the blocks and weighs them again to determine how much wood has been consumed. How might Susan use her results to determine if her theory is correct?

Ans: 1) The experimental unit is the individual wood block, as the data is collected for each block to measure the amount of wood consumed.

2) The experimental units were taken from 2 populations: Wood blocks impregnated with the IGR and solvent and Wood blocks impregnated with only the solvent. 60 wood blocks were drawn from each population, for a total of 120 blocks. This is a 2-sample problem.

3) Each experimental unit has 2 measurements:

Initial weight of the wood block (before termite exposure).

Final weight of the wood block (after termite exposure).

- 4) The parameter of interest is the difference in the mean amount of wood consumed (Δ) between the two groups. $\Delta = \mu_{\text{IGR}} - \mu_{\text{Solvent}}$, where:

μ_{IGR} : Mean wood consumption for blocks treated with IGR and solvent.

μ_{Solvent} : Mean wood consumption for blocks treated with solvent alone.

- 5) Null hypothesis (H_0): The IGR has no effect on wood consumption by termites. The mean wood consumption is the same for both groups. $H_0 : \Delta = 0$ or $\mu_{\text{IGR}} = \mu_{\text{Solvent}}$

Alternative hypothesis (H_1): The IGR reduces wood consumption by termites. The mean wood consumption is less for blocks treated with the IGR. $H_1 : \Delta < 0$ or $\mu_{\text{IGR}} < \mu_{\text{Solvent}}$

d. According to an article in Newsweek (May 10, 2004, page 89), recent “studies have shown consistently that women are better than men at reading and responding to subtle cues about mood and temperament.” Some psychologists believe that such differences can be explained in part by biological differences between male and female brains. One such psychologist conducts a study in which day-old babies are shown three human faces and three mechanical objects. The time that the baby stares at each face/object is recorded. Of interest is how much time the baby spends staring at faces versus how much time the baby spends staring at objects. The psychologist’s theory predicts that this comparison will differ by sex, with female babies preferring faces to objects to a greater extent than do male babies. How might the psychologist use his results to determine if his theory is correct?

Ans: 1) The experimental unit is the **individual baby**, as the data is collected for each baby to measure the time spent staring at faces and objects.

- 2) The experimental units were taken from 2 populations: Female babies and Male babies. The exact number of babies in each population is not provided but would consist of all male and female babies in the study. This is a 2-sample problem.
- 3) Each experimental unit has 2 measurements:

Time spent staring at human faces.

Time spent staring at mechanical objects.

- 4) The parameter of interest is the difference in the mean preference for faces over objects (Δ) between female and male babies. $\Delta = \mu_{\text{Female}} - \mu_{\text{Male}}$, where: μ_{Female} : Mean preference (time spent on faces minus time spent on objects) for female babies.

μ_{Male} : Mean preference (time spent on faces minus time spent on objects) for male babies.

- 5) Null hypothesis (H_0): There is no difference in the preference for faces over objects between female and male babies. $H_0 : \Delta = 0$ or $\mu_{\text{Female}} = \mu_{\text{Male}}$

Alternative hypothesis (H_1): Female babies prefer faces over objects to a greater extent than male babies. $H_1 : \Delta > 0$ or $\mu_{\text{Female}} > \mu_{\text{Male}}$

Q4) A researcher studies the use of color backgrounds to enhance online reading. A random sample of people was randomly divided into two groups: 235 individuals were asked to read a nonfiction novel online on a webpage with an orange background color, and another 197 individuals were given the same novel online on a blue background color. Readability was measured via the distance traveled by the mouse while scrolling the page in a fixed amount of time. They found that the average distance for the orange background group was 23.4 feet (sample standard deviation 5.7 feet) versus 21.9 feet (sample standard deviation 7.2 feet) for the blue background group.

a. Are the distance traveled by readers with orange background approximately Normal? How can we know? Does it matter whether it is normal or not? Explain

Ans: The sample size for the orange background group is 235, which is large enough for the Central Limit Theorem (CLT) to apply. The CLT states that, regardless of the underlying population distribution, the sampling distribution of the sample mean will be approximately normal for large samples. Therefore, normality of the individual data will not affect the validity of inference methods. The same reasoning applies to the blue background group.

b. Find a 98% confidence interval for the difference in average distance traveled using an orange instead of a blue background.

Ans: The formula for a confidence interval for the difference in two sample means is:

$$CI = (\bar{x}_1 - \bar{x}_2) \pm t \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

where:

- $\bar{x}_1 = 23.4$, $s_1 = 5.7$, $n_1 = 235$ (orange background),
- $\bar{x}_2 = 21.9$, $s_2 = 7.2$, $n_2 = 197$ (blue background),
- t is the critical value for 98% confidence with degrees of freedom calculated using the Welch-Satterthwaite formula.

First, we shall compute the standard error: $SE = \sqrt{\frac{5.7^2}{235} + \frac{7.2^2}{197}}$.

The degrees of freedom are given using: $df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$.

```
# Given data
x1 <- 23.4 # Mean for orange background
s1 <- 5.7  # Standard deviation for orange background
n1 <- 235  # Sample size for orange background

x2 <- 21.9 # Mean for blue background
s2 <- 7.2  # Standard deviation for blue background
n2 <- 197  # Sample size for blue background
```

```

# Calculate the standard error
SE <- sqrt((s1^2 / n1) + (s2^2 / n2))

# Calculate degrees of freedom using the Welch-Satterthwaite equation
df <- ((s1^2 / n1) + (s2^2 / n2))^2 /
      (((s1^2 / n1)^2 / (n1 - 1)) + ((s2^2 / n2)^2 / (n2 - 1)))

# Critical t-value for 98% confidence
alpha <- 0.02 # Two-tailed, so divide alpha by 2
t_critical <- qt(1 - alpha / 2, df)

# Calculate the confidence interval
mean_diff <- x1 - x2
lower_bound <- mean_diff - t_critical * SE
upper_bound <- mean_diff + t_critical * SE

# Display the results
cat("98% Confidence Interval for the Difference in Means:\n")

## 98% Confidence Interval for the Difference in Means:

cat("(", lower_bound, ", ", upper_bound, ")\n")

## ( 0.01970644 , 2.980294 )

```

c. Would you use Welch's or Student's two-sample T-test to perform a hypothesis test? Explain.

Ans: The sample standard deviations suggest the variances are not equal ($s_1 < 5.7$ and $s_2 < 7.2$) so it would be safer to use Welch's t-test. Welch's test accounts for differences in variances and is more appropriate when the assumption of equal variances is questionable.

d. Perform a hypothesis test using the method chosen in part c. State the null and alternative hypotheses, and conclude and interpret your result.

Ans: We are testing whether the orange background increases the distance traveled by the mouse compared to the blue background.

- Null Hypothesis (H_0): The mean distance traveled for the orange and blue backgrounds is equal. $H_0 : \mu_1 = \mu_2$
- Alternative Hypothesis (H_a): The mean distance traveled for the orange background is greater than for the blue background. $H_a : \mu_1 > \mu_2$

This is a one-tailed test.

The test statistic for Welch's t-test is given by:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Substituting the given values:

$$\bar{x}_1 = 23.4, \quad s_1 = 5.7, \quad n_1 = 235 \quad \bar{x}_2 = 21.9, \quad s_2 = 7.2, \quad n_2 = 197$$

The degrees of freedom df are calculated using the Welch-Satterthwaite equation:

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

```
# Given data
x1 <- 23.4 # Mean for orange background
s1 <- 5.7  # Standard deviation for orange background
n1 <- 235  # Sample size for orange background

x2 <- 21.9 # Mean for blue background
s2 <- 7.2  # Standard deviation for blue background
n2 <- 197  # Sample size for blue background

# Calculate the standard error
SE <- sqrt((s1^2 / n1) + (s2^2 / n2))

# Calculate degrees of freedom using the Welch-Satterthwaite equation
df <- ((s1^2 / n1) + (s2^2 / n2))^2 /
      (((s1^2 / n1)^2 / (n1 - 1)) + ((s2^2 / n2)^2 / (n2 - 1)))

# Calculate the t-statistic
t_stat <- (x1 - x2) / SE

# Calculate the p-value for a one-tailed test
p_value <- pt(t_stat, df, lower.tail = FALSE)

# Display the results
cat("t-statistic:", t_stat, "\n")
```

```
## t-statistic: 2.367561
```

```
cat("Degrees of Freedom:", df, "\n")
```

```
## Degrees of Freedom: 370.4134
```

```
cat("p-value:", p_value, "\n")
```

```
## p-value: 0.009209451
```

We reject the null hypothesis since the p-value (approximately 0.0092) is less than $\alpha = 0.05$ and conclude that there is statistically significant evidence that the orange background increases the distance traveled compared to the blue background.

Q5) Researchers obtained the following measurements of urinary -thromboglobulin excretion in 12 diabetic patients and 12 normal control subjects.

	4.1	6.3	7.8	8.5	8.9	10.4
Normal	11.5	12.0	13.8	17.6	24.3	37.2
	11.5	12.1	16.1	17.8	24.0	28.8
Diabetic	33.9	40.7	51.3	56.2	61.7	69.2

(a) Do these measurements appear to be samples from symmetric distributions? Why or why not?

Ans: No, these measurements do not appear to be samples from symmetric distributions. This is because the values show a clear skew towards higher values, especially in both groups (normal and diabetic). The data is not evenly distributed around the central values, with the upper range significantly larger than the lower range, indicating right-skewed distributions.

(b) Both samples of positive real numbers appear to be drawn from distributions that are skewed to the right; i.e., the upper tail of the distribution is longer than the lower tail of the distribution. Often, such distributions can be symmetrized by applying a suitable data transformation. Two popular candidates are:

(1) The natural logarithm: $u_i = \log(x_i)$ and $v_j = \log(y_j)$

(2) The square root: $u_i = \sqrt{x_i}$ and $v_j = \sqrt{y_j}$

Investigate the effect of each of these transformations on the above measurements. Do the transformed measurements appear to be samples from symmetric distributions? Which transformation do you prefer?

Ans: 1. Logarithmic Transformation ($\log(x)$): After applying the logarithmic transformation, the measurements appear more symmetric, as this transformation compresses the larger values more than the smaller ones, effectively reducing the right skew.

2. Square Root Transformation (\sqrt{x}): The square root transformation also reduces skewness but typically to a lesser extent compared to the logarithmic transformation. The data may still show mild skewness after applying the square root.

The logarithmic transformation is generally preferred in this case, as it better normalizes the data and reduces the skew more effectively than the square root transformation.

(c) Do the transformed measurements appear to be samples from normal distributions? Why or why not?

Ans: The transformed measurements (especially after the logarithmic transformation) appear closer to normal distributions, as the skewness is reduced, and the data becomes more symmetric. However, it is essential to formally test normality using statistical tests such as the Shapiro-Wilk test or by examining Q-Q plots.

The researchers claimed that diabetic patients have increased urinary β -thromboglobulin excretion. Assuming that the transformed measurements are samples from normal distributions, how convincing do you find the evidence for their claim?

Ans: Assuming the transformed measurements are normally distributed, the diabetic group consistently exhibits higher values of β -thromboglobulin excretion compared to the normal group. Statistical tests such as the two-sample t-test could be performed to assess the significance of this difference. If the p-value from the test is small (e.g., $p < 0.05$), this provides strong evidence supporting the researchers' claim that diabetic patients have significantly increased urinary β -thromboglobulin excretion.