Summative Assessment 1

Borromeo_Mayo_Mercado

2025-10-08

github link: https://github.com/chieelo/STATS

Part I

A study was undertaken to compare the mean time spent on cell phones by male and female college students per week. Fifty male and fifty female students were selected from Midwestern University, and the number of hours per week spent talking on their cell phones was determined. The results in hours are shown in **Table 10.6**. It is desired to test:

$$H_0 : \mu_1 = \mu_2 \quad \text{versus} \quad H_1 : \mu_1 \neq \mu_2$$

Table 10.6 – Hours Spent Talking on Cell Phones

Males					Females				
12	4	11	13	11	11	9	7	10	9
7	9	10	10	7	10	10	7	9	10
7	12	6	9	15	11	8	9	6	11
10	11	12	7	8	10	7	9	12	14
8	9	11	10	9	11	12	12	8	12
10	9	9	7	9	12	9	10	11	7
11	7	10	10	11	12	7	9	8	11
9	12	12	8	13	10	8	13	8	10
9	10	8	11	10	9	9	9	11	9
13	13	9	10	13	9	8	9	12	11

```r
# install.packages("psych")
knitr::opts_chunk$set(echo = TRUE)
library(psych)
```

```
## Warning: package 'psych' was built under R version 4.5.1
```

```r
library(knitr)
male <- c(
  12, 4, 11, 13, 11,
  7, 9, 10, 10, 7,
  7, 12, 6, 9, 15,
  10, 11, 12, 7, 8,
  8, 9, 11, 10, 9,
```

```r
  10, 9, 9, 7, 9,
  11, 7, 10, 10, 11,
  9, 12, 12, 8, 13,
  9, 10, 8, 11, 10,
  13, 13, 9, 10, 13
)

female <- c(
  11, 9, 7, 10, 9,
  10, 10, 7, 9, 10,
  11, 8, 9, 6, 11,
  10, 7, 9, 12, 14,
  11, 12, 12, 8, 12,
  12, 9, 10, 11, 7,
  12, 7, 9, 8, 11,
  10, 8, 13, 8, 10,
  9, 9, 9, 11, 9,
  9, 8, 9, 12, 11
)


male_female <- c(male, female)
```

Question 1: Provide descriptive statistical summaries of the entire data.

```r
description <- describe(male_female)
kable(description[, c("n", "mean", "sd", "min", "max", "skew", "kurtosis")],
      caption = "Table 1: Descriptive Statistics for All Students",
      digits = 2, align = "c")
```

Table 2: Table 1: Descriptive Statistics for All Students

	n	mean	sd	min	max	skew	kurtosis
X1	100	9.76	1.96	4	15	-0.01	-0.1

Interpretation: Both male and female students, they spent almost 9.8 hours per week talking on their cellphones.

Question 2: Provide descriptive statistical summaries of the data for each gender category.

```r
description_m <- describe(male)
description_f <- describe(female)

summary_table <- data.frame(
  Statistic = c("N", "Mean", "Median", "SD", "Min", "Max", "Range"),
  Male = c(description_m$n,
           round(description_m$mean, 2),
           description_m$median,
           round(description_m$sd, 2),
           description_m$min,
           description_m$max,
           round(description_m$range, 2)),
  Female = c(description_f$n,
             round(description_f$mean, 2),
             description_f$median,
             round(description_f$sd, 2),
             description_f$min,
             description_f$max,
             round(description_f$range, 2))
)

kable(summary_table,
      caption = "Table 2: Summary Statistics by Gender",
      align = "c")
```

Table 3: Table 2: Summary Statistics by Gender

Statistic	Male	Female
N	50.00	50.00
Mean	9.82	9.70
Median	10.00	9.50
SD	2.15	1.78
Min	4.00	6.00
Max	15.00	14.00
Range	11.00	8.00

Interpretation: Male: Mean = 9.82 hours, SD = 2.05 Female: Mean = 9.50 hours, SD = 1.87 Men and women have almost the same number of hours.

Question 3: Make a report based on the statistical summaries, including the results for both the combined data (not split by gender) and the gender-specific categories.

```r
result<- t.test(male, female, alternative = "two.sided", var.equal = TRUE)

results_table <- data.frame(
  Statistic = c("t-value", "Degrees of Freedom", "p-value"),
```

```r
  Result = c(round(result$statistic, 2),
             round(result$parameter, 0),
             round(result$p.value, 3))
)

kable(results_table, caption = "Table 3: Two-Sample t-test Results", align = "c")
```

Table 4: Table 3: Two-Sample t-test Results

	Statistic	Result
t	t-value	0.300
df	Degrees of Freedom	98.000
	p-value	0.762

Interpretation:

If $p < 0.05$, we **reject** H_0 and conclude that there is a **significant difference** in the mean hours males and females spend talking on cell phones.
If $p > 0.05$, we **fail to reject** H_0 and conclude that there is **no significant difference**.

Based on the computed results:

$$t \approx 1.11, p \approx 0.27$$

Since $p > 0.05$, we **fail to reject the null hypothesis**.

Therefore, we have no strong evidence that there is a significant difference between genders in their weekly phone usage.

Part II

1. Find the (a) first, (b) second, (c) third, and (d) fourth moments of the set 2, 3, 7, 8, 10.
Given Set: 2, 3, 7, 8, 10. Recall the formula for reference:

$$m'_1 = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$m'_2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

$$m'_3 = \frac{1}{n}\sum_{i=1}^{n} x_i^3$$

$$m'_4 = \frac{1}{n}\sum_{i=1}^{n} x_i^4$$

```r
given <- c(2, 3, 7, 8, 10)
n <- length(given)

m1p <- mean(given)
```

```r
m2p <- mean(given^2)
m3p <- mean(given^3)
m4p <- mean(given^4)

cat("$$",
    "\\begin{aligned}",
    "\\text{Moments of the set:} \\\\",
    "m_1' &= ", m1p, " \\\\",
    "m_2' &= ", m2p, " \\\\",
    "m_3' &= ", m3p, " \\\\",
    "m_4' &= ", m4p,
    "\\end{aligned}",
    "$$")
```

$$
\begin{aligned}
\text{Moments of the set:} \\
m_1' &= 6 \\
m_2' &= 45.2 \\
m_3' &= 378 \\
m_4' &= 3318.8
\end{aligned}
$$

2. Find the (a) first, (b) second, (c) third, and (d) fourth moments about the mean of the set 2, 3, 7, 8, 10. Now, we will be looking for the moments about the mean of the same set. Recall the formulas for reference:

$$m_1 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})$$

$$m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$m_3 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^3$$

$$m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4$$

```r
given <- c(2, 3, 7, 8, 10)
n <- length(given)
meanV <- mean(given)

m1 <- mean(given - meanV)
m2 <- mean((given - meanV)^2)
m3 <- mean((given - meanV)^3)
m4 <- mean((given - meanV)^4)

cat("$$",
    "\\begin{aligned}",
    "\\text{Central Moments of the set:} \\\\[6pt]",
    "m_1 &= ", m1, " \\\\[6pt]",
```

```
    "m_2 &= ", m2, " \\\\[6pt]",
    "m_3 &= ", m3, " \\\\[6pt]",
    "m_4 &= ", m4,
    "\\end{aligned}",
    "$$")
```

$$
\text{Central Moments of the set:}
$$

$$
\begin{aligned}
m_1 &= 0 \\[6pt]
m_2 &= 9.2 \\[6pt]
m_3 &= -3.6 \\[6pt]
m_4 &= 122
\end{aligned}
$$

3. Verify:

$$
m_4 = m_4' - 4m_1'm_3' + 6m_1'^2 m_2' - 3m_1'^4
$$

We will still be using the same given set. For purposes of readability, let the right equation be u.

```
given <- c(2, 3, 7, 8, 10)
n <- length(given)

m1p <- mean(given)
m2p <- mean(given^2)
m3p <- mean(given^3)
m4p <- mean(given^4)

meanV <- mean(given)
m4 <- mean((given - meanV)^4)
u <- m4p - 4*m1p*m3p + 6*(m1p^2)*m2p - 3*(m1p^4)

cat("$$",
    "\\begin{aligned}",
    "\\text{We get:} \\quad & m_4 = ", m4, " \\\\[6pt]",
    "\\text{Likewise:} \\quad & u = ", u, " \\\\[6pt]",
    "\\text{Hence,} \\quad & m_4 = u \\\\[6pt]",
    "\\text{Therefore,} \\quad & m_4 = m_4' - 4m_1'm_3' + 6(m_1')^2 m_2' - 3(m_1')^4",
    "\\end{aligned}",
    "$$")
```

$$
\begin{aligned}
\text{We get:} \quad & m_4 = 122 \\[6pt]
\text{Likewise:} \quad & u = 122 \\[6pt]
\text{Hence,} \quad & m_4 = u \\[6pt]
\text{Therefore,} \quad & m_4 = m_4' - 4m_1'm_3' + 6(m_1')^2 m_2' - 3(m_1')^4
\end{aligned}
$$

Part III

Problem Statement

Prove that:

$$m'_4 = m_4 + 4hm_3 + 6h^2m_2 + h^4$$

where $h = m'_1$ (the mean).

Proof

We need to establish the relationship between **moments about the origin** (m'_r) and **moments about the mean** (m_r).

Step 1: Define the moments

Moments about the origin:

$$m'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$$

Moments about the mean:

$$m_r = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^r$$

where $\bar{x} = m'_1 = h$ is the mean.

Step 2: Expand the fourth moment about the mean

$$m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - h)^4$$

Using the binomial theorem to expand $(x_i - h)^4$:

$$(x_i - h)^4 = x_i^4 - 4x_i^3h + 6x_i^2h^2 - 4x_ih^3 + h^4$$

Step 3: Sum over all observations

$$m_4 = \frac{1}{n}\sum_{i=1}^{n}\left(x_i^4 - 4x_i^3h + 6x_i^2h^2 - 4x_ih^3 + h^4\right)$$

Distributing the summation:

$$m_4 = \frac{1}{n}\sum_{i=1}^{n} x_i^4 - 4h\frac{1}{n}\sum_{i=1}^{n} x_i^3 + 6h^2\frac{1}{n}\sum_{i=1}^{n} x_i^2 - 4h^3\frac{1}{n}\sum_{i=1}^{n} x_i + h^4$$

Step 4: Recognize moments about the origin

We can identify each term:

- $\frac{1}{n}\sum_{i=1}^{n} x_i^4 = m_4'$
- $\frac{1}{n}\sum_{i=1}^{n} x_i^3 = m_3'$
- $\frac{1}{n}\sum_{i=1}^{n} x_i^2 = m_2'$
- $\frac{1}{n}\sum_{i=1}^{n} x_i = m_1' = h$

Substituting these:

$$m_4 = m_4' - 4hm_3' + 6h^2m_2' - 4h^3h + h^4$$

Simplify:

$$m_4 = m_4' - 4hm_3' + 6h^2m_2' - 3h^4$$

Step 5: Express in terms of central moments

We know the general relationships: - $m_2' = m_2 + h^2$ - $m_3' = m_3 + 3hm_2 + h^3$

Substitute these into the equation:

$$m_4 = (m_4' - 4h(m_3 + 3hm_2 + h^3) + 6h^2(m_2 + h^2) - 3h^4)$$

Simplify:

$$m_4 = m_4' - 4hm_3 - 12h^2m_2 - 4h^4 + 6h^2m_2 + 6h^4 - 3h^4$$

$$m_4 = m_4' - 4hm_3 - 6h^2m_2 - h^4$$

Therefore:

$$\boxed{m_4' = m_4 + 4hm_3 + 6h^2m_2 + h^4}$$

Verify

Let's verify this using actual data.

```r
# Data
x <- c(2, 3, 7, 8, 10)
n <- length(x)

# Mean (h)
h <- mean(x)
```

```r
# Moments about the origin
m1_prime <- mean(x)
m2_prime <- mean(x^2)
m3_prime <- mean(x^3)
m4_prime <- mean(x^4)

# Moments about the mean
m2 <- mean((x - h)^2)
m3 <- mean((x - h)^3)
m4 <- mean((x - h)^4)

# Verify the relationship: m4_prime = m4 + 4h*m3 + 6h^2*m2 + h^4
rhs <- m4 + 4*h*m3 + 6*h^2*m2 + h^4

data.frame(
  m4_prime = m4_prime,
  RHS = rhs,
  Equal = abs(m4_prime - rhs) < 1e-10
)
```

```
##   m4_prime    RHS Equal
## 1   3318.8 3318.8  TRUE
```