A fair six-sided die will be rolled fifteen times, and the numbers that land face up will be recorded. Let $\overline{x}_1$ represent the average of the numbers that land face up for the first five rolls, and let $\overline{x}_2$ represent the average of the numbers landing face up for the remaining ten rolls. The mean μ and variance σ^2 of a single roll are 3.5 and 2.92, respectively. What is the standard deviation $\sigma_{(\overline{x}_1 - \overline{x}_2)}$ of the sampling distribution of the difference in sample means $\overline{x}_1 - \overline{x}_2$?

(A) $2.92 + 2.92$

(B) $2.92 - 2.92$

(C) $\sqrt{\frac{2.92}{5} + \frac{2.92}{10}}$

(D) $\sqrt{\frac{2.92^2}{5} + \frac{2.92^2}{10}}$

(E) $\sqrt{\frac{2.92^2}{5} - \frac{2.92^2}{10}}$

C

Sample Distributions for Sample Means

1. **Mean – 均值 (jūn zhí)**
2. **Variance – 方差 (fāng chā)**
3. **Standard error – 标准误差 (biāo zhǔn wù chā)**
4. **Sample size – 样本容量 (yàng běn róng liàng)**
5. **Sampling distribution – 抽样分布 (chōu yàng fēn bù)**

Reminders
Mock: 40 MCQ and 6 FRQ. (Very mixed set of questions)
Jupiter Quiz due for Monday 16th Monday (3 quick questions)

Sampling Distribution of $\bar{x}$

- $\bar{x}$ is the sample mean used to estimate the population mean μ.
- The mean of $\bar{x}$ equals the population mean: $\mu_{\bar{x}} = \mu$.
- The standard deviation of $\bar{x}$ is $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$.
- The distribution of $\bar{x}$ is normal if the population is normal or approximately normal for large n (CLT).
- Independence is required (10% condition when sampling without replacement).

A company wants to estimate the average number of hours employees spend on training per month.

- A random sample of 36 employees is taken, and the sample mean is 12 hours with a standard deviation of 3 hours.
- Assume the distribution of hours per employee is approximately normal.

(a) Define the parameter of interest.

(b) Identify the sampling distribution of the sample mean. Include its mean and standard error.

(c) Calculate the standard error of the sample mean.

(d) Using the sampling distribution, what is the probability that the sample mean is greater than 13 hours?

(e) State the conditions necessary for using this sampling distribution.

(a) Parameter: μ = true mean hours of training per employee.

(b) Sampling distribution: $\bar{x} \sim N(\mu, SE = s/\sqrt{n})$

(c) SE: $SE = 3/\sqrt{36} = 0.5$

(d) Probability $\bar{x} > 13$: $z = (13 - 12)/0.5 = 2 \Rightarrow P \approx 0.023$

(e) Conditions: random sample, independent observations, population normal or $n \geq 30$.

Stretch and Challenge

A population has mean $\mu = 70$ and standard deviation $\sigma = 15$. The distribution of the population is **strongly right-skewed**.

A researcher repeatedly takes random samples of **size** $n = 100$ and records the **sample mean** $\bar{x}$.

Questions

a) Describe the **shape, center, and spread** of the **Sampling Distribution of the Sample Mean**.

b) Calculate the probability that a sample mean is **greater than 73**.

c) Suppose the researcher instead takes samples of **size** $n = 25$.
Explain how the **shape and spread** of the sampling distribution would change.

d) In one sample of $n = 100$, the researcher observes $\bar{x} = 76$.
Determine whether this result would be considered **unusual** and justify your answer.

e) Explain why it is valid to use a **normal model** to answer parts (b) and (d), even though the population distribution is strongly skewed. Your explanation should reference the **Central Limit Theorem**.

a) Shape, Center, Spread

For the **Sampling Distribution of the Sample Mean**:

- **Shape:** Approximately normal (large $n = 100$) by the **Central Limit Theorem**
- **Center:** $\mu_{\bar{x}} = 70$
- **Spread:** $\sigma_{\bar{x}} = \frac{15}{\sqrt{100}} = 1.5$

---

b) $P(\bar{x} > 73)$

$$z = \frac{73 - 70}{1.5} = 2$$

$$P(Z > 2) = 0.0228$$

---

c) If $n = 25$

- Shape: less normal (smaller sample)
- Spread: $\sigma_{\bar{x}} = \frac{15}{5} = 3$
- **More variability**

---

d) Is $\bar{x} = 76$ **unusual?**

$$z = \frac{76 - 70}{1.5} = 4$$

Probability ≈ **0.00003**

Yes — extremely unusual.

---

e) Why normal model works

Because the **Central Limit Theorem** states that for **large samples**, the distribution of the **Sample Mean** is approximately normal even if the population is skewed.

Distribution of Difference in Sample Means

1. Sampling distribution – 抽樣分佈
2. Difference of sample means ($\bar{X}_1 - \bar{X}_2$) – 樣本平均差
3. Standard error (SE) – 標準誤
4. Independent samples – 獨立樣本
5. Central Limit Theorem (CLT) – 中心極限定理

1. **Mean of the sampling distribution:**

$$\mu_{\bar{X}_1-\bar{X}_2} = \mu_1 - \mu_2$$

The expected value of the difference in sample means equals the difference of population means.

2. **Standard error (spread):**

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

Measures variability of the difference between sample means.

3. **Shape of distribution:**

- **Normal** if populations are normal.
- **Approximately normal** if n_1 and n_2 are large (CLT applies).

4. **Independence matters:**

   The formula assumes **samples are independent.** Paired samples require a different approach.

5. **Using z-scores for probability:**

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{X}_1-\bar{X}_2}}$$

Allows calculation of probabilities for differences in sample means.

A researcher wants to compare the **average test scores** of two different teaching methods.

- Population 1 (Method A): $\mu_1 = 78, \sigma_1 = 10$
- Population 2 (Method B): $\mu_2 = 72, \sigma_2 = 12$

Random samples are taken: $n_1 = 36$ students for Method A and $n_2 = 49$ students for Method B.

Questions

a) Find the **mean** and **standard error** of the sampling distribution of $\bar{X}_1 - \bar{X}_2$.

b) Calculate the probability that the **sample mean difference** is **greater than 10**.

c) Suppose a sample difference of **15** is observed. Determine if this is **unusual**, and justify your answer.

d) Explain why it is valid to use a **normal model** to calculate probabilities even if the populations are not perfectly normal.

7. Pooled vs. Unpooled Variance

When performing hypothesis tests or constructing confidence intervals, the approach to estimating variance depends on whether the population variances are assumed to be equal:

Pooled Variance: Assumes $\sigma_1^2 = \sigma_2^2$. The pooled variance (s_p^2) is calculated as:

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

This pooled estimate is used in the standard error calculation for the test statistic.

Unpooled Variance: Does not assume equal variances. The standard error is calculated separately for each sample, as shown earlier.

Pooled or Unpooled

- Two independent random samples are taken from two populations. Both populations are known to have **equal variances**, but the sample sizes are different.

Pooled

Pooled or Unpooled

- Two samples are taken from populations with **very different sample standard deviations**, and equality of variances is questionable.

Unpooled

Pooled or Unpooled

- A researcher compares the mean test scores of two teaching methods. Previous studies show the **population variances are approximately equal**.

Pooled

Pooled or Unpooled

THEIR RESULTS, COMPARED WITH CHAMELEON'S:

CHAMELEON

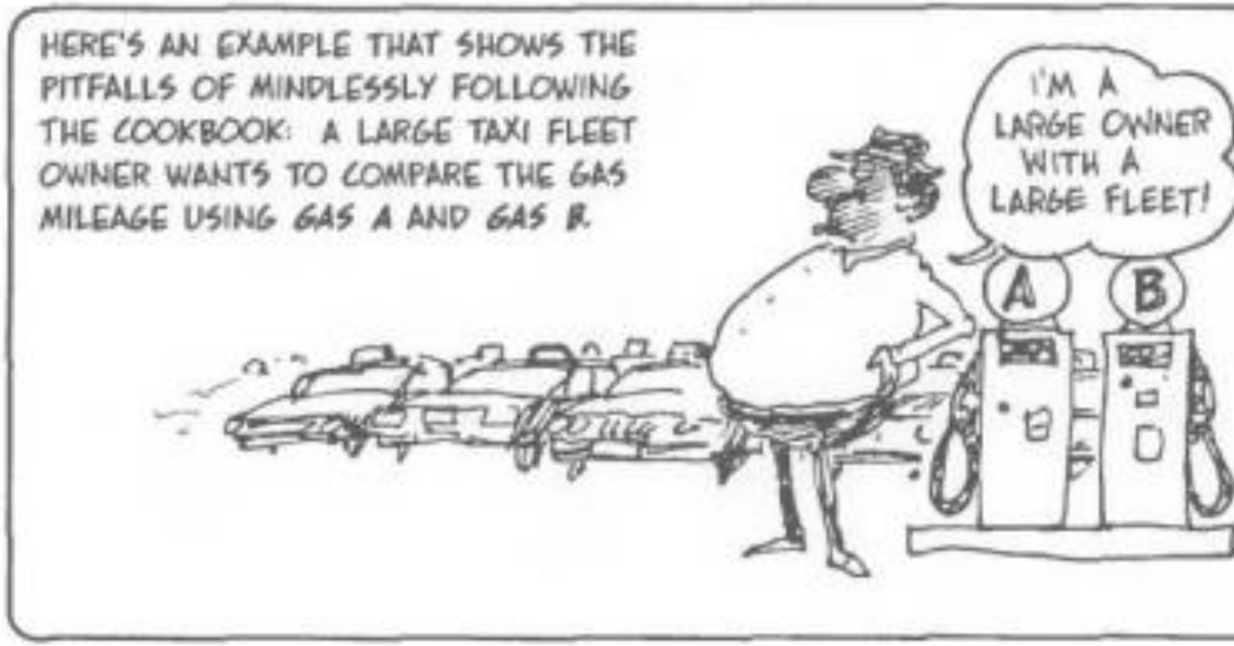
1	\$150
2	\$400
3	\$720
4	\$500
5	\$930
n_1	5
$\bar{x}_1$	\$540
s_1	\$299

IGUANA

1	\$50
2	\$200
3	\$150
4	\$400
5	\$750
6	\$400
7	\$150
n_2	7
$\bar{x}_2$	\$300
s_2	\$238

Pooled

Pooled or Unpooled



Unpooled

STARTING WITH 100 CABS, HE RANDOMLY ASSIGNS 50 TO EACH GASOLINE, AN
AFTER A DAY'S DRIVING, DETERMINES

	SAMPLE SIZE	MEAN MILEAGE	STANDARD DEVIATION
A	50	25	5.00
B	50	26	4.00

$\mu_1 = 78, \sigma_1 = 10, n_1 = 36;\ \mu_2 = 72, \sigma_2 = 12, n_2 = 49$

---

a) Mean & SE

$$\mu_{\bar{X}_1 - \bar{X}_2} = 78 - 72 = 6$$

$$\sigma_{\bar{X}_1 - \bar{X}_2} = \sqrt{100/36 + 144/49} \approx 2.39$$

---

b) $P(\bar{X}_1 - \bar{X}_2 > 10)$

$$z = \frac{10 - 6}{2.39} \approx 1.674 \quad \Rightarrow \quad P \approx 0.047$$

---

c) Is 15 unusual?

$$z = \frac{15 - 6}{2.39} \approx 3.77 \quad \Rightarrow \quad P \approx 0.00008$$

Yes, very unusual.

---

d) Why normal model valid

- Large samples ($n_1 = 36, n_2 = 49$)
- By **Central Limit Theorem**, sampling distribution ≈ normal even if populations are skewed

AP Exam Question – Pooled Variance

A researcher wants to compare the **mean test scores** of two classes that were taught using different methods.

Data:

Class	Sample Size	Sample Mean	Sample Standard Deviation
A	20	85	5
B	25	80	6

Assume **both populations have equal variances**.

Questions:

a) Calculate the **pooled variance** s_p^2.

b) Calculate the **standard error** of $\bar{X}_A - \bar{X}_B$.

c) Find the **t-statistic** to test whether Class A has a higher mean score than Class B.

d) Determine the **degrees of freedom** for this test.

e) Using a **0.05 significance level**, decide whether the difference is **statistically significant**.

Given

$$n_A = 20, \bar{X}_A = 85, s_A = 5; \quad n_B = 25, \bar{X}_B = 80, s_B = 6$$

---

a) Pooled variance

$$s_p^2 = \frac{(n_A - 1)s_A^2 + (n_B - 1)s_B^2}{n_A + n_B - 2} = \frac{19(25) + 24(36)}{43} = \frac{475 + 864}{43} = \frac{1339}{43} \approx 31.14$$

---

b) Standard error (SE)

$$SE = \sqrt{s_p^2 \left(\frac{1}{n_A} + \frac{1}{n_B}\right)} = \sqrt{31.14 \left(\frac{1}{20} + \frac{1}{25}\right)} = \sqrt{31.14(0.05 + 0.04)} = \sqrt{31.14 \cdot 0.09} = \sqrt{2.8026} \approx 1.673$$

---

c) t-statistic

$$t = \frac{\bar{X}_A - \bar{X}_B}{SE} = \frac{85 - 80}{1.673} \approx 2.99$$

---

d) Degrees of freedom

$$df = n_A + n_B - 2 = 20 + 25 - 2 = 43$$

---

e) Significance at $\alpha = 0.05$

- Critical $t_{0.05,43} \approx 1.681$ (one-tailed)
- $t = 2.99 > 1.681 \Rightarrow$ **Reject H_0**

Conclusion: Class A has a **significantly higher mean score** than Class B.

TRUE or FALSE

The formula for the standard error of $\bar{X}_1 - \bar{X}_2$ assumes the two samples are independent.

TRUE

TRUE or FALSE

If the populations are highly skewed, the sampling distribution of $\bar{X}_1 - \bar{X}_2$ can still be approximately normal if the sample sizes are large.

TRUE

TRUE or FALSE

Increasing one sample size while keeping the other fixed always decreases the standard error by the same amount.

FALSE