1 Mark for Review

A fair six-sided die, with sides numbered 1 through 6, will be rolled a total of 15 times. Let $\overline{x}_1$ represent the average of the *first* ten rolls, and let $\overline{x}_2$ represent the average of the *remaining* five rolls. What is the mean $\mu_{(\overline{x}_1-\overline{x}_2)}$ of the sampling distribution of the difference in sample means $\overline{x}_1 - \overline{x}_2$?

(A) $\frac{3.5}{10} - \frac{3.5}{5} = -0.35$

(B) $3.5 - 3.5 = 0$

(C) $10 - 5 = 5$

(D) $10(3.5) - 5(3.5) = 17.5$

(E) $6(10 - 5) = 30$

B

4 Mark for Review

The height requirement for a certain ride at an amusement park is 48 inches or taller. The distribution of height of six-year-old children is approximately normal with mean 45.2 inches and standard deviation 1.75 inches.

Which of the following is closest to the probability that a randomly selected six-year-old child will not meet the height requirement for the ride?

(A) 0.0401

(B) 0.0548

(C) 0.9452

(D) 0.9599

(E) 0.9974

C

5 Mark for Review

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

Sampling Distribution of $\hat{p}$

- $\hat{p}$ is the sample proportion used to estimate the population proportion p.
- The mean of $\hat{p}$ is the population proportion: $\mu_{\hat{p}} = p$.
- The standard deviation of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.
- The distribution of $\hat{p}$ is approximately normal if $np \geq 10$ **and** $n(1-p) \geq 10$.
- Independence is required (10% condition when sampling without replacement).

Sample Proportion	样本比例
Population Proportion	总体比例
Sampling Distribution	抽样分布
Standard Error	标准误
Central Limit Theorem	中心极限定理

A survey of 200 high school students finds that 120 of them prefer online learning over in-person classes

a) What is the **sample proportion** of students who prefer online learning?

b) Explain what the **sampling distribution** of the sample proportion represents.

c) If the true **population proportion** is 0.55, calculate the **standard error** of the sample proportion.

d) Why can we use the **Central Limit Theorem** to approximate the sampling distribution with a normal distribution in this case?

a) $\hat{p} = 120/200 = 0.6$

b) Sampling distribution = distribution of all possible sample proportions; shows variability around the population proportion.

c) $SE = \sqrt{0.55 \cdot 0.45/200} \approx 0.035$

d) CLT applies because $np \geq 10$ and $n(1-p) \geq 10$; distribution ≈ normal.

Question 1 (Concept + Conditions)

A population has proportion $p = 0.35$. A sample of size $n = 40$ is taken.

(a) Explain whether the sampling distribution of $\hat{p}$ can be approximated by a Normal distribution.
(b) If appropriate, calculate $P(\hat{p} > 0.50)$.
(c) Interpret your answer in context.

Question 2 (Reverse Engineering Probability)

A sample of size n is taken from a population with $p = 0.6$. It is found that

$$P(\hat{p} < 0.5) = 0.05$$

Assuming a Normal approximation is valid, **estimate the sample size** n.

Question 3 (Comparing Two Proportions)

Two independent samples are taken:

- Sample A: $n_1 = 100,\ \hat{p}_1 = 0.48$
- Sample B: $n_2 = 150,\ \hat{p}_2 = 0.55$

(a) Find the mean and standard deviation of $\hat{p}_1 - \hat{p}_2$.
(b) Calculate $P(\hat{p}_1 - \hat{p}_2 < -0.10)$.
(c) Comment on whether the observed difference is statistically significant.

Question 4 (Challenging – Validity of Model)

A survey reports that 12 out of 20 people prefer a new product.

(a) Explain why a Normal approximation may not be appropriate.
(b) Suggest an alternative approach.
(c) Estimate $P(\hat{p} > 0.7)$ using your suggested method.

Question 5 (Multi-step + Interpretation)

A factory claims that 80% of its products meet a quality standard. A sample of 120 items is tested.

(a) Find the probability that fewer than 70% meet the standard.
(b) If this occurs, discuss whether it provides evidence against the factory's claim.
(c) Explain the role of sampling variability in your conclusion.

Question 6 (Extreme Tail Reasoning)

A population proportion is $p = 0.25$, and $n = 200$.

(a) Calculate $P(\hat{p} < 0.20)$.
(b) A sample result gives $\hat{p} = 0.18$. Explain whether this is surprising.
(c) What does this suggest about the population proportion?

Q1

(a) Check conditions:
$np = 40(0.35) = 14$, $n(1-p) = 26$ → both ≥ 10 ✓ Normal OK

(b)
Mean $= 0.35$, SD $= \sqrt{\frac{0.35(0.65)}{40}} \approx 0.075$
$z = \frac{0.50-0.35}{0.075} \approx 2.00$
$P \approx 0.0228$

(c) About **2.3% chance** → quite unlikely

Q2

$$z = \frac{0.5 - 0.6}{\sqrt{0.6(0.4)/n}} = -1.645$$

Solve:

$$\frac{-0.1}{\sqrt{0.24/n}} = -1.645 \Rightarrow \sqrt{0.24/n} \approx 0.0608 \Rightarrow n \approx 65$$

Q3

(a)
Mean $= 0.48 - 0.55 = -0.07$
SD $= \sqrt{\frac{0.48(0.52)}{100} + \frac{0.55(0.45)}{150}} \approx 0.064$

(b)
$z = \frac{-0.10-(-0.07)}{0.064} \approx -0.47$
$P \approx 0.319$

(c) Not significant (probability large)

Q4

(a) $np = 12, n(1-p) = 8$ → fails (8 < 10) ❌

(b) Use **Binomial distribution**

(c)

$$P(X \geq 14),\ X \sim \text{Bin}(20, 0.6)$$

≈ **0.25 (calculator)**

Q5

(a)

Mean $= 0.8$, SD $= \sqrt{\frac{0.8(0.2)}{120}} \approx 0.0365$
$z = \frac{0.70-0.80}{0.0365} \approx -2.74$
$P \approx 0.0031$

(b) Very small → evidence **against claim**

(c) Could occur by chance, but unlikely

Q6

(a)

SD $= \sqrt{\frac{0.25(0.75)}{200}} \approx 0.0306$
$z = \frac{0.20-0.25}{0.0306} \approx -1.63$
$P \approx 0.051$

(b) Slightly unusual

(c) Weak evidence true $p < 0.25$

Statistical Inference

1. Statistical Inference — 统计推断
2. Parameter — 总体参数
3. Statistic — 样本统计量
4. Sampling Distribution — 抽样分布
5. Confidence Interval — 置信区间
6. Margin of Error — 误差范围
7. Null Hypothesis (H_0) — 原假设
8. Alternative Hypothesis (H_a) — 备择假设
9. p-value — p值
10. Significance Level (α) — 显著性水平

1. Inference uses sample data to draw conclusions about a population.
2. A parameter describes a population; a statistic describes a sample.
3. Sampling variability means sample results naturally change from sample to sample.
4. The sampling distribution shows how a statistic (like $\hat{p}$) varies across many samples.
5. For proportions, $\mu_{\hat{p}} = p$ and $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.
6. Normal approximation is valid only if $np \geq 10$ and $n(1-p) \geq 10$.
7. A confidence interval gives a range of plausible values for a population parameter.
8. The margin of error measures how far the estimate may be from the true value.
9. Hypothesis testing evaluates evidence against a null hypothesis using a p-value.
10. A small p-value suggests strong evidence against H_0, but does not prove it false.

FRQ: Statistical Inference – Sample Proportions

A university claims that **60% of its graduates find employment within 6 months** of graduating. A researcher believes this proportion is lower and takes a random sample of **200 graduates**, finding that **108 are employed within 6 months**.

(a) Conditions

Check whether a Normal approximation can be used for the sampling distribution of $\hat{p}$.

(b) Probability

Assuming the university's claim is correct, calculate the probability of observing a sample proportion **as low as or lower than 108/200**.

(c) Interpretation

Based on your result in (b), comment on whether the sample provides evidence against the university's claim.

(d) Inference conclusion

State a conclusion in context using a **5% significance level**.

(e) Extension (Challenge)

If the sample size were doubled to 400 but the sample proportion stayed the same (0.54), explain how this would affect:

- (i) the standard deviation of $\hat{p}$
- (ii) the strength of evidence against the claim

(a) Conditions

$np = 200(0.6) = 120 \geq 10, n(1-p) = 80 \geq 10$
✔ Normal approximation valid

(b) Probability

$$\hat{p} = \frac{108}{200} = 0.54$$

Mean $= 0.60$
SD $= \sqrt{\frac{0.6(0.4)}{200}} = \sqrt{0.0012} \approx 0.0346$

$$z = \frac{0.54 - 0.60}{0.0346} \approx -1.73$$

$P \approx 0.042$

(c) Interpretation

Probability ≈ 4.2% → **quite unlikely** if claim is true → evidence against claim

---

(c) Interpretation

Probability ≈ 4.2% → **quite unlikely** if claim is true → evidence against claim

---

(d) Conclusion (5% level)

Since $p \approx 0.042 < 0.05$:
✔ **Reject** H_0
Evidence suggests true proportion is **less than 0.60**

(e) Extension

(i) SD decreases:

$$\sigma_{\hat{p}} \propto \frac{1}{\sqrt{n}} \Rightarrow smaller$$

(ii) Evidence becomes **stronger** (same proportion but smaller variability → larger z-score magnitude)

True or False

- A sampling distribution of a sample proportion is based on repeated samples of the same size.

True

True or False

- Increasing sample size makes the sampling distribution narrower.

True

A software application (app) lets users enter questions to receive answers in the form of images, texts, or videos. Research indicates that 22 percent of high school students in Country W use the app to help them with their homework at least once per week. Karen is an AP Statistics student in Country W at a high school that has more than 2,000 students. She believes the proportion of all students at her school who use the app to help them with their homework at least once per week is greater than the proportion for her country. To investigate her belief, she took a simple random sample of 130 students from her school and found that 38 of the sampled students use the app to help them with their homework at least once per week.

Is there convincing statistical evidence, at a 0.05 significance level, to support Karen's belief? Justify your answer with the appropriate inference procedure.