

The AI Math Stress Test

Ten questions your AI will answer with total confidence. Some of those answers are wrong.

FACTORYOS BY COMPTRIO · YOUR DATA STAYS ON YOUR MACHINE. · [COMPTRIO.COM](https://comptrio.com)

A language model doesn't calculate — it generates what a right answer *looks like*. For prose, close-enough reads fine. For arithmetic, "looks right" and "is right" are different numbers, and the model can't tell you which one it handed you. These ten are picked from accounting, wealth management, and the kind of math HNW portfolios actually run on — each one a place where a confident wrong answer is easy and the real answer takes real work.

How to run it: paste each question into the AI you use today — ChatGPT, Copilot, Gemini, whatever. Write down its answer. Then turn to the answer key on the next pages: the exact figure, the worked steps, and the source. Where they don't match, you've found math you can't trust on your own screen.

THE TEN QUESTIONS

1. **ACCOUNTING** A single filer has **\$100,000** of taxable income. Using these federal brackets — 10% to \$11,600, 12% to \$47,150, 22% to \$100,525, 24% above — what is their **total federal income tax**?

2. **ACCOUNTING** What is the **monthly payment** on a **\$25,000** loan at **6% APR** over **5 years** (60 payments)?

3. **ACCOUNTING** A municipal bond yields **3.5%**, tax-free. For an investor in the **37%** bracket, what **taxable yield** would be equivalent?

4. **WEALTH MGMT** A fund's **12 monthly returns** are below. With a 0% risk-free rate, what is its **annualized Sharpe ratio**?

+2.1, -1.5, +3.4, +0.8, -2.7, +4.5, +1.2, -0.9, +3.0, -4.1, +1.8, +2.5 (percent)

5. **WEALTH MGMT** Same 12 monthly returns, 0% risk-free — what is the **annualized Sortino ratio**?

6. **WEALTH MGMT** An account's **month-end values** (indexed to 100) were below. What was the **maximum drawdown**?

100, 115, 108, 122, 119, 130, 116, 124, 112, 118, 121, 109

7. **WEALTH MGMT** Using those same 12 monthly returns, what is the **95% one-month historical Value at Risk**?

8. **HNW / ALTS** An investment costs **\$100,000** today and pays back **\$20,000, \$30,000, \$40,000, \$50,000** over the next four years. What is the **IRR**?

9. **HNW / CRYPTO** In a 50/50 liquidity pool, one asset's price **doubles** relative to the other. What is the **impermanent loss** versus simply holding the two?

10. **HNW / CRYPTO** A yield pool advertises **12% APR, compounded monthly**. What is the actual **APY**?

What you're really testing: not whether AI is useful — it is — but whether it should be the thing doing the arithmetic. The criterion for anything you'd put near your numbers: **ask where the math runs**. Inside the model, it's a guess in a suit. The next pages show the other way.

The Answer Key — computed, not guessed

Every figure below comes from a deterministic engine: the exact value, the work shown, the source cited.

This is how FactoryOS answers the same questions. It doesn't let the model do the math. The AI decides which data and which calculation; a deterministic engine computes every figure exactly — the same way every time — and shows the work with the source attached. Main answer up top, the steps underneath. Trust, but verify.

1 • Federal income tax on \$100,000

\$17,053

Where AI slips: the tempting shortcut is the top rate on the whole income — $22\% \times \$100,000 = \$22,000$. Wrong by ~\$5,000. Each bracket only taxes its own slice.

$$\begin{aligned}\text{tax} &= \Sigma (\text{income in bracket}) \times \text{rate} \\ &= 11,600 \times 10\% + 35,550 \times 12\% + 52,850 \times 22\% \\ &= 1,160 + 4,266 + 11,627 \\ &= \$17,053 \quad (\text{effective rate } 17.05\%, \text{ marginal rate } 22\%) \end{aligned}$$

Source: progressive marginal-bracket method, IRS Publication 17.

2 • Monthly payment, \$25,000 @ 6% / 60 mo

\$483.32

Where AI slips: it reverts to $\text{principal} \div \text{term}$ (\$416.67) or fumbles the amortization factor, landing a few dollars off — small per line, real over a portfolio of loans.

$$\begin{aligned}\text{pmt} &= \text{pv} \cdot r / (1 - (1 + r)^{-n}), \quad r = 6\%/12 = 0.005, \quad n = 60 \\ &= 25,000 \times 0.005 / (1 - 1.005^{-60}) \\ &= \$483.32 / \text{month} \quad (\text{total repaid} \approx \$28,999) \end{aligned}$$

Source: standard amortizing-loan / annuity payment formula (Excel PMT convention).

3 • Tax-equivalent yield, 3.5% muni @ 37%

5.56%

Where AI slips: the intuitive move is to *add* the rate — $3.5\% \times 1.37 = 4.80\%$. The correct inverse is to divide by $(1 - \text{rate})$.

$$\begin{aligned}\text{TEY} &= \text{tax-free yield} / (1 - \text{tax rate}) \\ &= 0.035 / (1 - 0.37) \\ &= 0.035 / 0.63 \\ &= 5.56\% \end{aligned}$$

Source: Investopedia — Tax-Equivalent Yield.

4 • Annualized Sharpe ratio

1.11

Where AI slips: a convention trap — using population stdev ($\div n$) instead of sample ($\div n-1$), or forgetting to annualize by $\sqrt{12}$. Miss either and the number drifts.

$$\begin{aligned}\text{Sharpe} &= \text{mean} / \text{stdev}(\text{sample}) \times \sqrt{12} \\ \text{mean} &= 0.101 / 12 = 0.008417 \quad (0.84\%/\text{mo}) \\ \text{stdev} &= 0.026190 \quad (\text{sample, } \div n-1) \\ &= 0.008417 / 0.026190 \times 3.4641 \\ &= 1.11 \end{aligned}$$

Source: Sharpe ratio (annualized), Sharpe 1966 / standard definition.

5 · Annualized Sortino ratio

1.94

Where AI slips: it reuses the full standard deviation instead of *downside* deviation, which counts only losing months and divides by *n*. Get that wrong and Sortino collapses toward Sharpe.

```
Sortino = mean / downside-deviation × √12
downside-dev = √( ∑ min(0, r)2 / n )
              = √( (0.0152 + 0.0272 + 0.0092 + 0.0412) / 12 )
              = √( 0.002716 / 12 ) = 0.015044
= 0.008417 / 0.015044 × 3.4641
= 1.94
```

Source: Sortino ratio (downside-deviation), standard definition.

6 · Maximum drawdown

-16.15%

Where AI slips: it reports the biggest *single-month* drop (130 → 116 = -10.8%). Max drawdown is the largest *peak-to-trough* fall — the peak of 130 all the way down to the later trough of 109.

```
max drawdown = (trough - prior peak) / prior peak
peak = 130 (month 6), subsequent trough = 109 (month 12)
= (109 - 130) / 130
= -16.15%
```

Source: drawdown (peak-to-trough), standard definition.

7 · 95% one-month historical VaR

3.33%

Where AI slips: it switches to *parametric* VaR (mean - 1.645 × σ) and presents that as the historical answer — a different method, a different number — or it just grabs the single worst month (-4.1%).

```
Historical VaR(95%) = -(5th-percentile return)
sorted returns, rank = 0.05 × (12 - 1) = 0.55
interpolate between -4.1% and -2.7%: -4.1% + 0.55 × 1.4% = -3.33%
VaR = 3.33% (a one-month loss exceeded ~5% of the time)
```

Source: historical Value at Risk (return quantile), standard definition.

8 · IRR of the four-year cash flow

12.83%

Where AI slips: IRR has no closed form — it requires iterative root-finding. A model can't iterate, so it guesses a plausible rate (often "~10%") or averages the cash back. Confident, and wrong.

```
IRR solves NPV = 0:
-100,000 + 20,000/(1+r) + 30,000/(1+r)2 + 40,000/(1+r)3 + 50,000/(1+r)4 = 0
solved numerically → r = 12.83%
```

Source: internal rate of return (NPV = 0), standard definition.

9 · Impermanent loss at a 2× price move

-5.72%

Where AI slips: an exotic formula it almost always misremembers — and the answer is counterintuitively small. Guesses run wild (some say ~25%, confusing it with the price move).

```
IL = 2√r / (1 + r) - 1,    r = price ratio = 2
   = 2√2 / 3 - 1
   = 2.8284 / 3 - 1
   = -5.72% (vs. simply holding the two assets)
```

Source: Uniswap v2 documentation — impermanent loss (matches their 1.25× → 0.6% ... 5× → 25.5% table).

10 · APY of 12% APR, compounded monthly

12.68%

Where AI slips: it ignores the compounding and answers 12% flat. The monthly compounding adds 68 basis points — small here, decisive across a yield comparison.

$$\begin{aligned} \text{APY} &= (1 + \text{APR}/n)^n - 1, \quad n = 12 \\ &= (1 + 0.12/12)^{12} - 1 \\ &= 1.01^{12} - 1 \\ &= 12.68\% \end{aligned}$$

Source: Investopedia — Annual Percentage Yield.

— Shawn Snarski · FactoryOS by Comptrio · your data stays on your machine. · comptrio.com

Every figure here was produced by FactoryOS's deterministic math engine, which keeps the arithmetic out of the language model: the AI chooses the data and the calculation, the engine computes it exactly and shows the work with the source attached. That's the difference between an answer that looks right and one that is. Tax brackets shown are illustrative — swap in the current schedule and the method holds.