Evaluating the Effectiveness of Human-Annotated Math Statements on Olympiad-Level Math Problems

Anonymous ACL submission

Abstract

Math statements, including definitions, theorems, axioms, lemmas, formulas, and so on, provide a clear and precise way to express mathematical concepts, which helps in constructing logical arguments and proofs for mathematical reasoning. Currently, there is a lack of systematic research to verify the role of math statements in solving math problems of Olympiad-level difficulty. In this paper, we conducted extensive experiments to evaluate the mathematical reasoning performance of multiple cutting-edge large language models (LLMs) with and without math statements as prompts. We found that problem-aligned math statements can substantially enhance the problem-solving capabilities of LLMs on complex Olympiad-level math problems. Notably, this enhancement is particularly pronounced in smaller-scale models such as Qwen2.5-Math-7B, where our curated math statements can achieve accuracy gains of over 10%. Even advanced deep reasoning models such as QwQ-32B still demonstrated a 3.5% accuracy improvement. Moreover, we construct the SA-Math dataset, which comprises 114 human-annotated Olympiad-level math problems, along with 130 domain-relevant math statements. We believe that our work can facilitate the math-problem-solving capabilities of LLMs.

1 Introduction

Mathematical reasoning, as a critical capability of LLMs, has garnered significant research attention following the emergence of advanced reasoning models such as DeepSeek-R1 (DeepSeek-AI, 2025) and OpenAI-o1 (Jaech et al., 2024). While these models exhibit strong performance on elementary mathematical benchmarks, many LLMs still struggle with knowledge-intensive problems in complex mathematical reasoning tasks, such as those found in Mathematical Olympiad competitions (He et al., 2024). Math statements like theorems, axioms, lemmas, and formulas, describe fundamental concepts in mathematics and logic, used to express relationships between mathematical objects or to make assertions about mathematical properties. While humans naturally use these as cognitive scaffolds for problem-solving, current LLM prompting techniques like Chain-of-Thought (COT) often neglect such domain-specific statements. This raises our key question: does explicitly encoding human-understandable math statements into prompts can guide LLMs to activate relevant knowledge during reasoning?

As shown in Table 1, recent studies have developed various datasets and benchmarks (Chen et al., 2023; Lucy et al., 2024; Wu et al., 2024; Zhao et al., 2024) to evaluate LLMs' ability to leverage mathematical knowledge in reasoning tasks. However, many existing work primarily focuses on K-12 level mathematics, where the limited complexity of elementary problems fails to adequately distinguish between knowledge-aware and knowledge-agnostic reasoning performance. Moreover, current practices mainly use math tags, educational curricula or math concepts generated by models or retrieved from websites (Li et al., 2025; Huang et al., 2025) as prompts. These methods may lack the precision, rigor, and completeness offered by manually curated mathematical statements.

In this paper, we focus on assessing whether math statements can improve LLM mathematical reasoning performance on Olympiad-level problems. We conducted experiments on multiple representative open-source and closed-source LLMs using manually verified question-statement-answer triples and prompts with and without human-annotated math statements. Experimental results demonstrate that relevant statements significantly enhance the math-problem-solving capabilities of LLMs, particularly on Olympiad-level problems. We also constructed a **S**tatement-**A**ugmented **Math** problem dataset (**SA-Math**) for evaluation. This dataset comprises 130 math statements and 114

Dataset	Domain	Source	Level	Including Statements?	Is Available?
TheoremQA (Chen et al., 2023)	STEM	Internet+Expert	College	✓	✓
MathFish (Lucy et al., 2024)	Math	Internet+Expert	K-12	✗	✓
ConceptMath (Wu et al., 2024)	Math	LLM+Expert	Grade 1-9	✗	✓
FinanceMath (Zhao et al., 2024)	Finance	Internet+Expert	College	✓	✗
SA-Math	Math	Expert	Olympiad	✓	✓

Table 1: Comparison of SA-Math dataset and other knowledge-intensive mathematical reasoning benchmarks.

curated problems spanning four core mathematical domains (*Algebra*, *Geometry*, *Number Theory*, and *Combinatorics*), where all problems are all tagged with one or more math statements through expert annotation. Furthermore, we propose to transfer the statements in SA-Math to existing mathematical benchmarks based on the embedding similarities of problems. This enables boosting LLM performance through statement integration in public datasets.

The contributions of our work are two-fold:

- We conduct experiments to assess the effectiveness of manually annotated math statements on multiple cutting-edge LLMs and Olympiad-Level math problems.
- We construct SA-Math dataset with 114 Olympiad-level problems annotated with human-verified math statements, which will be released at `https://anonymous.4open.science/r/SA-Math-FFCE`.

2 Related Work

Current methods for annotating math knowledge in mathematical problems primarily adopt two paradigms: direct generation via LLMs and knowledge base retrieval. The former leverages LLMs' intrinsic reasoning capabilities to extract math knowledge through zero-shot (Zhao et al., 2025) or few-shot prompting (Liu et al., 2022; Zhu et al., 2024). The latter retrieves relevant math knowledge through LLM-based relevance evaluation (Li et al., 2024b), agentic retrieval-augmented generation (RAG) frameworks (Li et al., 2025; Henkel et al., 2024), or embedding similarity metrics (Li et al., 2024a; Ding et al., 2025). The knowledge is subsequently employed to construct skill repositories (Didolkar et al., 2024) that facilitate either problem-solving assistance (Ozyurt et al., 2024) or problem generation (Huang et al., 2025).

Available datasets (Lucy et al., 2024; Wu et al., 2024) with math knowledge often lack detailed statements, while those few existing ones (Chen et al., 2023; Zhao et al., 2024) containing math statements are not specifically dedicated to Olympiad-level mathematical problems. In contrast, our proposed SA-Math dataset incorporates both Olympiad-level mathematical problems and corresponding human-annotated math statements, thereby addressing this critical gap in the field.

Math knowledge is typically leveraged to enhance LLM Reasoning through refined prompting strategies. These strategies encompass concatenating math knowledge with problem text (Liu et al., 2022), instructing explicit knowledge reference in outputs (Henkel et al., 2024), or embedding knowledge within reasoning paths (Li et al., 2025). However, due to the scarcity of math statements, current methodologies predominantly employ concise mathematical knowledge to facilitate reasoning in LLMs. This motivates us to systematically evaluate the impact of detailed statements on LLMs' math-problem-solving capabilities using our dataset.

3 SA-Math Dataset

The SA-Math dataset is built using expert-annotated Olympiad-level math problems with relevant statements, providing a verified benchmark to assess the enhancement of human-annotated math statements on LLM-based mathematical reasoning.

Source. We have carefully collected 114 Olympiad problems and 130 relevant math statements. Each problem is linked to one or more relevant math statements which are annotated with brief titles and detailed descriptions by experts. All the mathematical formulas within the content are preserved in their original LaTeX format.

Human verification. To mitigate unreasonable reasoning caused by inaccurate text and formulas, we conduct comprehensive content integrity and LaTeX format validation on SA-Math. Content in-

Statistics	Number
Total problems	114
Algebra	63(55%)
Geometry	12(11%)
Number Theory	14(12%)
Combinatorics	25(22%)
Total math statements	130
Algebra	77(59%)
Geometry	24(18%)
Number Theory	18(14%)
Combinatorics	11(9%)
Average problem tokens	129
Average math statement tokens	37

Table 2: Statistics of SA-Math

tegrity checks address omissions and maintain logical coherence, while LaTeX validation guarantees syntax correctness and symbol consistency for formula readability. Experts further verify the alignment between math statements and problems to eliminate mismatches.

Dataset Description. The details of the SA-Math dataset are presented in Table 2. The 114 problems in SA-Math dataset span 4 major mathematical domains including *Algebra*, *Geometry*, *Number Theory*, and *Combinatorics* (see Appendix A for examples). Math statements featuring a domain-title-description hierarchical structure can provide auxiliary information for LLM reasoning. Examples of domain-title structures in math statements include *Number Theory-Chinese Remainder Theorem*, *Geometry-Principle of Intersecting Chords*, *Combinatorics-The Pigeon-Hole Principle*, and *Algebra-Trigonometric Equations* (the detailed descriptions are provided in Appendix B). Additionally, because the SA-Math dataset is constructed from proprietary data sources, the potential training data leakage can be substantially mitigated.

4 Experimental Results

Evaluation Setting. We evaluated the following LLMs on SA-Math. The proprietary LLMs include GPT-4 Turbo, o1-preview (Jaech et al., 2024), and o3-mini, while open-source LLMs include DeepSeekMath-7B-Instruct (Zhihong Shao, 2024) QwQ-32B (Team, 2025b), Qwen2.5-32B-Instruct (Yang et al., 2024a), Qwen2.5-Math-7B (Yang et al., 2024b), Qwen3-8B, Qwen3-14B (Team, 2025a), and Llama-3.1-70B-Instruct (Grattafiori et al., 2024).

We evaluated the problem-solving capabilities of LLMs using Qwen2.5-Math scripts (Yang et al., 2024b) and reported the Pass@1 accuracy. For most LLMs, we employ greedy search decoding during inference by setting the temperature to 0 and the top-p to 1. For deep reasoning models such as QwQ-32B, we configure the temperature at 0.7 and the top-p at 0.95 following the recommendations in the Qwen2.5-Math evaluation scripts. All experiments in this paper are conducted on a compute node with 8 × H20 80GB GPUs.

Prompting without statements

System: Please reason step by step, and put your final answer within \boxed{}.

User: {Problem Description}

Prompting with statements

System: Please reason step by step, and put your final answer within \boxed{}.

User: Please answer the following question based on the konwledge points we have listed.

Knowledge Points:
Knowledge point-1. {Statement Title}
{Statement Description}

Knowledge point-2. {Statement Title}
{Statement Description}

...

Question: {Problem Description}

Figure 1: Template for the two prompting methods.

Methods. We employed two prompting methods respectively for these LLMs. The one is **prompting without statements** which prompting LLMs to generate step-by-step reasoning procedures in ordinary CoT manners. The other one is **prompting with statements** which extends CoT reasoning prompts with problem-specific statement integration, enabling LLMs to contextually leverage domain knowledge for articulated mathematical reasoning. The templates of the two prompting methods are shown in Figure 1.

4.1 Main Results

Table 3 shows a comprehensive performance comparison of various LLMs on our SA-Math dataset.

Proprietary LLMs. Regardless of whether math statements are embedded in the prompts, deep reasoning models like o3-mini outperform general-purpose chat models such as GPT-4 Turbo on the

Model	w/o Statements	w/ Statements
Proprietary LLMs		
GPT-4 Turbo	55.3	57.0(+1.7)
o1-preview	68.4	69.3(+0.9)
o3-mini	71.9	72.8(+0.9)
Open-source LLMs		
Qwen2.5-Math-7B	39.5	50.0(+10.5)
Qwen3-8B	59.6	64.0(+4.4)
Qwen3-14B	61.4	66.7(+5.3)
DeepSeekMath-7B-Instruct	31.6	34.2(+2.6)
QwQ-32B*	76.3*	79.8*(+3.5)
Qwen2.5-32B-Instruct	68.4	71.9(+3.5)
Llama-3.1-70B-Instruct	43.0	50.0(+7.0)

Table 3: Experimental results on SA-Math. Qwen2.5-Math-7B achieves the most notable accuracy improvement (up to 10.5%), while QwQ-32B maintains best performance both before and after the integration of math statements.

Model	w/o Statements	w/ Statements
Qwen2.5-Math-7B	19.5	40.2(+20.7)
Qwen3-8B	47.6	48.8(+1.2)
Qwen3-14B	52.4	57.3(+4.9)
DeepSeekMath-7B-Instruct	13.4	17.1(+3.7)
QwQ-32B*	75.6*	78.0*(+2.4)
Qwen2.5-32B-Instruct	46.3	50.0(+3.7)
Llama-3.1-70B-Instruct	29.3	30.4(+1.1)

Table 4: Experimental results on OlympiadBench-subset which contains 82 problems curated from OlympiadBench after matching.

SA-Math dataset (by up to 16.6%), highlighting their superior capabilities in math problem solving. For two prompting methods, our analysis shows that embedding domain-specific statements into prompts can improve the performance of LLMs on Olympiad-level math problems. However, prompting with statements yields less pronounced improvements on proprietary models compared to their open-source counterparts. This discrepancy arises because the frontier proprietary models have already internalized the domain knowledge for problem-solving, making explicit knowledge integration less effective for bridging reasoning gaps.

Open-source LLMs. Experimental results demonstrate that math statements significantly enhance the math-problem-solving capabilities of LLMs, with Qwen2.5-Math-7B achieving a 10.5% performance improvement and even deep-reasoning models such as QwQ-32B exhibiting a 3.5% performance gain. This substantiates that current open-source LLMs exhibit intrinsic limitations in problem-relevant knowledge. Consequently, the human-annotated statements in SA-Math that are well-aligned with problem exert substantial augmentation on the reasoning faculties of LLMs. Besides, we find that even the current best performing LLM (i.e., QwQ-32B) achieves an accuracy of less than 80% accuracy on SA-Math. This substantiates that the mathematical problems in SA-Math present enough difficulties to evaluate the math-problem-solving capabilities of existing LLMs.

4.2 Extension to Public Benchmark Augmentation

To demonstrate the potential of the SA-Math dataset, we perform embedding-based matching between its math statements and public benchmark problems. Specifically, we employ the E5-Mistral-7b-instruct (Wang et al., 2023, 2022) to compute embeddings for problems from both the public dataset and SA-Math. For each public problem, we retrieve SA-Math problems with cosine similarity scores exceeding a predetermined threshold of 0.85 as candidate problems. Finally, we collect the math statements of candidate problems to match target public problem and build statement-augmented prompts following Figure 1.

We adopt OlympiadBench (He et al., 2024) as our evaluation dataset due to its comparable problem difficulty to SA-Math. Following the alignment process, 82 OlympiadBench problems are matched with SA-Math statements, forming the OlympiadBench-subset for LLM evaluation. As evidenced by the experimental results in Table 4, the statement-augmented prompting demonstrates performance improvements when applied to problems from public datasets. Especially, Qwen2.5-Math-7B achieves a substantial accuracy improvement of 20.7%. This not only substantiates the validity of statement-augmented prompting but also demonstrates good compatibility between our math statements and the problems in public datasets.

5 Conclusion

This paper evaluates the effectiveness of manually annotated math statements on Olympiad-level mathematical problems with proposed SA-Math dataset. Our findings demonstrate that incorporating problem-relevant math statements into prompts significantly enhances the math-problem-solving capabilities of LLMs, which paves novel pathways in the field of mathematical reasoning.

Limitations

In this paper, we proposes to evaluate the enhancement of the domain-specific math statements in math-problem-solving capabilities of LLMs. There are still some limitations: (1) Our current benchmark comprises 114 carefully curated problems, which may exhibit potential coverage gaps in exhaustively representing the knowledge combinatorics inherent in mathematical reasoning tasks. (2) The statement-augmented prompting for public benchmarks relies on semantic similarities of problem embeddings, which may lead to mismatches in practice. In future work, we plan to annotate more problems to expand the SA-Math and explore fine-grained matching mechanisms between math statements and problems from public datasets.

References

Wenhu Chen, Ming Yin, Max Ku, Pan Lu, Yixin Wan, Xueguang Ma, Jianyu Xu, Xinyi Wang, and Tony Xia. 2023. TheoremQA: A theorem-driven question answering dataset. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 7889–7901, Singapore. Association for Computational Linguistics.

DeepSeek-AI. 2025. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *Preprint*, arXiv:2501.12948.

Aniket Didolkar, Anirudh Goyal, Nan Rosemary Ke, Siyuan Guo, Michal Valko, Timothy Lillicrap, Danilo Rezende, Yoshua Bengio, Michael Mozer, and Sanjeev Arora. 2024. Metacognitive capabilities of llms: An exploration in mathematical problem solving. In *Advances in Neural Information Processing Systems*, volume 37, pages 19783–19812. Curran Associates, Inc.

Ziqi Ding, Xiaolu Wang, Yuzhuo Wu, Guitao Cao, and Liangyu Chen. 2025. Tagging knowledge concepts for math problems based on multi-label text classification. *Expert Systems with Applications*, 267:126232.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Chaoqun He, Renjie Luo, Yuzhuo Bai, Shengding Hu, Zhen Thai, Junhao Shen, Jinyi Hu, Xu Han, Yujie Huang, Yuxiang Zhang, Jie Liu, Lei Qi, Zhiyuan Liu, and Maosong Sun. 2024. OlympiadBench: A challenging benchmark for promoting AGI with olympiad-level bilingual multimodal scientific problems. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3828–3850, Bangkok, Thailand. Association for Computational Linguistics.

Owen Henkel, Zach Levonian, Chenglu Li, and Millie Postle. 2024. Retrieval-augmented generation to improve math question-answering: Trade-offs between groundedness and human preference. In *Proceedings of the 17th International Conference on Educational Data Mining*, pages 315–320, Atlanta, Georgia, USA. International Educational Data Mining Society.

Yiming Huang, Xiao Liu, Yeyun Gong, Zhibin Gou, Yelong Shen, Nan Duan, and Weizhu Chen. 2025. Key-point-driven data synthesis with its enhancement on mathematical reasoning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 24176–24184.

Aaron Jaech, Adam Kalai, Adam Lerer, Adam Richardson, Ahmed El-Kishky, Aiden Low, Alec Helyar, Aleksander Madry, Alex Beutel, Alex Carney, and 1 others. 2024. Openai o1 system card. *arXiv preprint arXiv:2412.16720*.

Hang Li, Tianlong Xu, Jiliang Tang, and Qingsong Wen. 2024a. Automate knowledge concept tagging on math questions with llms. *arXiv preprint arXiv:2403.17281*.

Hang Li, Tianlong Xu, Jiliang Tang, and Qingsong Wen. 2024b. Knowledge tagging system on math questions via llms with flexible demonstration retriever. *arXiv preprint arXiv:2406.13885*.

Xiaoxi Li, Guanting Dong, Jiajie Jin, Yuyao Zhang, Yujia Zhou, Yutao Zhu, Peitian Zhang, and Zhicheng Dou. 2025. Search-o1: Agentic search-enhanced large reasoning models. *CoRR*, abs/2501.05366.

Jiacheng Liu, Alisa Liu, Ximing Lu, Sean Welleck, Peter West, Ronan Le Bras, Yejin Choi, and Hannaneh Hajishirzi. 2022. Generated knowledge prompting for commonsense reasoning. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3154–3169, Dublin, Ireland. Association for Computational Linguistics.

Li Lucy, Tal August, Rose E Wang, Luca Soldaini, Courtney Allison, and Kyle Lo. 2024. MathFish: Evaluating language model math reasoning via grounding in educational curricula. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 5644–5673, Miami, Florida, USA. Association for Computational Linguistics.

Yilmazcan Ozyurt, Stefan Feuerriegel, and Mrinmaya Sachan. 2024. Automated knowledge concept annotation and question representation learning for knowledge tracing. *arXiv preprint arXiv:2410.01727*.

Qwen Team. 2025a. Qwen3.

Qwen Team. 2025b. Qwq-32b: Embracing the power of reinforcement learning.

Liang Wang, Nan Yang, Xiaolong Huang, Binxing Jiao, Linjun Yang, Daxin Jiang, Rangan Majumder, and Furu Wei. 2022. Text embeddings by weakly-supervised contrastive pre-training. *arXiv preprint arXiv:2212.03533*.

Liang Wang, Nan Yang, Xiaolong Huang, Linjun Yang, Rangan Majumder, and Furu Wei. 2023. Improving text embeddings with large language models. *arXiv preprint arXiv:2401.00368*.

Yanan Wu, Jie Liu, Xingyuan Bu, Jiaheng Liu, Zhanhui Zhou, Yuanxing Zhang, Chenchen Zhang, ZhiqiBai ZhiqiBai, Haibin Chen, Tiezheng Ge, Wanli Ouyang, Wenbo Su, and Bo Zheng. 2024. ConceptMath: A bilingual concept-wise benchmark for measuring mathematical reasoning of large language models. In *Findings of the Association for Computational Linguistics: ACL 2024*, pages 6815–6839, Bangkok, Thailand. Association for Computational Linguistics.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, and 23 others. 2024a. Qwen2.5 technical report. *arXiv preprint arXiv:2412.15115*.

An Yang, Beichen Zhang, Binyuan Hui, Bofei Gao, Bowen Yu, Chengpeng Li, Dayiheng Liu, Jianhong Tu, Jingren Zhou, Junyang Lin, Keming Lu, Mingfeng Xue, Runji Lin, Tianyu Liu, Xingzhang Ren, and Zhenru Zhang. 2024b. Qwen2.5-math technical report: Toward mathematical expert model via self-improvement. *arXiv preprint arXiv:2409.12122*.

Xueliang Zhao, Wei Wu, Jian Guan, and Lingpeng Kong. 2025. Promptcot: Synthesizing olympiad-level problems for mathematical reasoning in large language models. *arXiv preprint arXiv:2503.02324*.

Yilun Zhao, Hongjun Liu, Yitao Long, Rui Zhang, Chen Zhao, and Arman Cohan. 2024. Financemath: Knowledge-intensive math reasoning in finance domains. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12841–12858, Bangkok, Thailand. Association for Computational Linguistics.

Qihao Zhu Runxin Xu Junxiao Song Mingchuan Zhang Y.K. Li Y. Wu Daya Guo Zhihong Shao, Peiyi Wang. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models.

Xunyu Zhu, Jian Li, Can Ma, and Weiping Wang. 2024. Key-point-driven mathematical reasoning distillation of large language model. *arXiv preprint arXiv:2407.10167*.

A Examples of problems in SA-Math

Each problem in the SA-Math dataset comprises four components: its affiliated domain, problem formulation, verified correct answer, and corresponding titles of math statements. Some exemplar problems is presented below.

Example of problem in SA-Math

Domain: Number Theory

Problem: A group of birds satisfy:
(i) Remainder 2 when counted by fives
(ii) Remainder 2 when counted by threes
(iii) Remainder 3 when counted by elevens
Find the smallest number of birds.

Answer:

$$\boxed{47}$$

Statements: ["The Chinese Remainder Theorem"]

Example of problem in SA-Math

Domain: Geometry

Problem: In a right-angled triangle with $a^2 + b^2 = c^2$, find r/R where r is inradius and R circumradius.

Answer:

$$\boxed{\frac{2ab}{c(a+b+c)}}$$

Statements: ["Triangle angle properties", "Pythagoras' theorem", "Circle geometry", "Tangent properties"]

Example of problem in SA-Math

Domain: Algebra

Problem: Find the sum: $\cos^2 0° + \cos^2 2° + \cos^2 4° + \cdots + \cos^2 358° + \cos^2 360°$.

Answer:

$$\boxed{91}$$

Statements: ["Inverse Trigonometric Functions", "Trigonometric Equations"]

B Examples of math statements in SA-Math

Each statement in the SA-Math contains three elements: its affiliated mathematical domain, the title of the statement, and its comprehensive description. We show several examples of math statements in following.

Example of math statement in SA-Math

Domain: Combinatorics

Title: The Pigeon-Hole Principle

Description: 1) Basic Pigeon-Hole Principle: If n objects are placed in fewer than n pigeon-holes, then at least two objects must occupy the same pigeon-hole. This principle explains results like having at least two people sharing a birth month in any group of 13 individuals.
2) General Form of the Pigeon-Hole Principle: If $mk + 1$ objects are distributed into m pigeon-holes, at least one pigeon-hole must contain at least $k + 1$ objects. This extends the basic principle to handle more complex distribution scenarios.
3) Formal Mathematical Statements: If a set of n elements is a union of $m < n$ subsets, at least one subset contains multiple elements. If a set of $mk + 1$ elements is a union of m subsets, at least one subset contains at least $k + 1$ elements.
4) Strategic Selection of Objects and Pigeon-Holes: Effective PHP application requires identifying suitable 'objects' (e.g., people) and 'pigeon-holes' (e.g., months). Correct pairing ensures conclusions like shared birthdays or overlapping spatial coordinates.

Example of math statement in SA-Math

Domain: Algebra

Title: Binomial Theorem

Description: The Binomial Theorem states that for any natural number n and real numbers x, y, and $\binom{n}{k} = \frac{n!}{(n-k)!k!}$:

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

The theorem is proven by induction. The base case $n = 1$ holds as $(x + y)^1 = x + y$. Assuming it holds for $n = t$, expanding $(x + y)^{t+1}$ and applying Pascal's rule confirms the inductive step. The theorem's shorthand form $(1 + x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k$ is useful for large powers, while the explicit expanded form

$$1 + nx + \frac{n(n-1)}{2!}x^2 + \frac{n(n-1)(n-2)}{3!}x^3 + \cdots + x^n$$

is practical for smaller powers or specific terms.

Example of math statement in SA-Math

Domain: Number Theory

Title: The Chinese Remainder Theorem

Description: The Chinese Remainder Theorem provides a systematic method for solving systems of congruences with pairwise coprime moduli. Let x be a number satisfying:

$$\begin{aligned} x &\equiv r_1 \pmod{d_1} \\ x &\equiv r_2 \pmod{d_2} \\ &\vdots \\ x &\equiv r_n \pmod{d_n} \end{aligned}$$

where $d_1, d_2, \ldots, d_n$ are pairwise coprime. Let $D = d_1 d_2 \cdots d_n$ and $y_i = \frac{D}{d_i}$. The theorem states that if we find integers a_i satisfying:

$$a_i y_i \equiv 1 \pmod{d_i} \quad \text{for each } i : 1 \leq i \leq n,$$

then a solution is:

$$x = \sum_{i=1}^{n} a_i y_i r_i.$$

Proof: For each modulus d_j, all terms in the sum except $a_j y_j r_j$ are multiples of d_j due to y_i containing d_j as a factor when $i \neq j$. The chosen a_j ensures $a_j y_j r_j \equiv r_j \pmod{d_j}$. Thus, x satisfies all congruences.
Remarks: Solutions are unique modulo D, with the smallest positive solution obtained by subtracting multiples of D from the initial solution.