

統計的仮説検定に関する推論と問題解決に対して事例の正誤に関する知識が与える影響¹⁾

堀内 正彦・永田 陽子

The influence of knowledge of true or false in cases on reasoning and problem solving about statistical hypothesis testing

Masahiko Horiuchi (Department of Psychology, Komazawa University, Japan)

Yoko Nagata (Department of Psychology, Komazawa University, Japan)

Abstract

One of the objectives of this study is to clarify the effect of showing the truth of a case on reasoning based on conditional and biconditional. The other purpose of this study is to examine whether the promotion of reasoning based on conditional and biconditional promotes understanding of logical systems, using statistical hypothesis testing as a subject. Based on the results, it was shown that understanding the truth of cases of conditional and biconditional sentences alone does not necessarily promote reasoning about them. As for the second purpose, it was difficult to examine from the results of this study. It was suggested that collaborative problem solving and procedural knowledge should be considered from multiple perspectives as a way to support understanding and reasoning about difficult events.

KEY WORDS: statistical hypothesis testing, knowledge, reasoning, problem solving

問題

心理学関連の学科の学生にとって心理統計学は難しい科目の1つであり、中でも統計的仮説検定は難解であり、そして重要な内容である。本研究は、統計的仮説検定の論理体系に論理学でいうところの条件文および双条件文が含まれることに着目して、条件文・双条件文に関する把握が心理統計学についての問題の解決に与える影響について検討する。

条件文と双条件文について

条件文とは、論理学における2項命題の1つであり、日常的な言葉では「もし p ならば q である」と表現できる。4枚カード問題は、この条件文に基づく推論を求める課題であり、長年にわたって研究されてきた認知課題である (Wason, 1968; Johnson-Liard, Legrenzi, & Legrenzi, 1972; Cheng & Holyoak, 1985)。Brain & O'Brien (1991) は、IF の理論 (theory of IF) において、条件文につい

での認知処理が語義の解釈 (lexical entry)、命題論理推論プログラム (propositional-logic reasoning program)、語用論的原理 (pragmatic principles) の3つの要素から構成されると主張し、条件文に関する認知処理が複雑であり、独特であることを示している。そのような条件文推論に関連して、堀内 (2000) は、4枚カード問題において一般的には条件文が用いられるところを連言、選言、双条件文に拡張して実験を行い、様々な2項命題に基づく推論について検討している。また、これらの命題において、選言と条件文は、論理的な構造に関して等価である。その実験の結果において、条件文と双条件文について論理的に推論することが特に困難であることを示している。これらの知見に基づく、Brain & O'Brien (1991) が述べているように、条件文によって表現される論理構造は、認知的負担がかかりやすく理解や問題解決を困難にさせると考えられる。そのような条件文の難しさの特徴は、双条件文についての推論は似ているところがあるが、他の2項命題につい

¹⁾ 本研究は、駒澤大学「人を対象とする研究」に関する倫理委員会の承認を2021年2月16日に得た上で実施された。また、本研究の結果の一部は、日本心理学会第85回大会において発表された。

では観察されず、条件文と双条件文に固有であると考えられる（堀内，2000）。

統計的仮説検定と条件文，双条件文の難しさについて

心理統計学が難しい科目であるという報告は複数見受けられる（Dunn, Carey, Richardson, & McDonald, 2016; Dolor & Noll, 2015）。心理統計学における主要な概念体系の1つに、統計的仮説検定があり、学生にとって難解な内容である。この統計的仮説的検定の論理体系には、条件文、双条件文の論理構造がいくつも含まれている。例えば、「帰無仮説に近い標本であれば、検定統計量は0に近い値になり、帰無仮説とかけ離れた標本であれば大きい値になる」あるいは「検定統計量の値が大きいならば、その確率は小さくなる」などである。条件文、双条件文の論理構造についての理解が獲得されることは、統計的仮説検定の論理体系を構成する個々の知識についての理解が進み、結果として統計的仮説検定についての理解も促進されると考えられる。

4枚カード問題の解法として図式適用の方法と仮説演繹的方法の2通りが高橋・服部（1996）によって示されており、どちらの解法であってもまず条件文の含意を正確に解釈することが必要になるはずである。しかし、それを正確に解釈すること自体が必ずしも容易ではなく、また、正確な解釈ではないとしても実験参加者自身の解釈に基づく推論はさらに困難であることが示されている（中垣，1989）。その一方で、条件文に関する認知処理に関して Brain & O'Brien（1991）も、条件文についての解釈と推論は別の過程としている。これらの知見に基づく、条件文あるいは双条件文の事例の正誤を明示することによる条件文の理解、推論の支援を検討する。

目的

本研究の目的の1つは、条件文および双条件文の事例正誤を把握することによって、直面する課題の論理構造を理解し、条件文・双条件文に基づく推論が促進されることを明らかにすることである。もう1つの目的は、条件文・双条件に基づく推論が促進されることによって、それらの構造を含む論理体系についての理解が促されることについて、統計的仮説検定を題材として検討すること

である。

方法

実験参加者

4年制大学心理学科の大学生22名（男性7名、女性15名、平均年齢20.6歳）。

実施期間・実施場所

2021年6月24日から7月15日までの間に、駒澤大学心理学実験室I-4において実施された。

実験計画

事例の正誤を示す正誤明示条件と示さない正誤非明示条件の2条件を設定した。

実験課題

課題は、統計的仮説検定に関する知識について合計6問であった。問1から問4は訓練試行であり、統計的仮説検定に関わる条件文もしくは双条件文に関する4枚カード問題と同じ形式の推論課題であった。正誤明示条件では、これらの課題に取り組む前に行う事例正誤についての判断を求める課題とその正答を提示すると提示し、正誤非明示条件では提示しなかった。問5と問6はテスト試行であり、問5は帰無仮説を棄却するか否かの判断について理由も含めて問う内容であった。問6は「検定統計量が臨界値を超えた」ということから言えることを問う内容であった。

問1から問4までの訓練試行の課題は、1) 規則（統計学の知識）の提示、2) 事例正誤の判断と明示、3) 4枚カード問題形式の推論、の3つで構成された。正誤明示条件に対して提示した問1を参考のために巻末資料に提示する。問1における冒頭から鍵括弧までの部分が規則の提示である。問2から問4において提示される規則（統計学の知識）は、Table 1に示す通りである。問1において鍵括弧の次に「以下の事例が正しいと思うか」どうかを尋ねている部分が、事例正誤の判断を求める部分である。この事例正誤の判断は、正誤明示条件にだけ提示され、正誤非明示条件においては提示されなかった。そして、その下の部分が4枚カード問題の形式による推論課題であった。

結果

事例正誤判断

事例正誤判断課題における主な反応パターンの1つは、論理学における条件文と同じ意味に解釈

Table 1 問2から問4において提示された規則

設問	規則
問2	検定統計量 F 値が大きい値ならば、その値になる確率は低い
問3	検定統計量 F 値が極端に大きい値ならば、帰無仮説を認めることには無理がある。反対に、 F 値が必ずしも極端に大きいとは言えない値ならば、帰無仮説を認めることに無理がない。
問4	帰無仮説は、比較する全ての母平均は等しいという意味です。反対に、対立仮説は、比較する母平均のどれか1つ以上が異なるという意味になります。

したと考えられる反応であり、これを条件文型判断と呼ぶことにする。このように規則を解釈することは、問2において正しい解釈である。また、主な反応パターンの2つ目は、論理学における双条件と同じ意味に解釈したと考えられる反応であり、これを双条件文型判断と呼ぶことにする。この解釈は、問1、問3、問4において正しい。問1から問4における事例正誤判断課題に対する主な反応パターンの人数を Table 2 に示す。

それぞれの設問における条件文型判断、双条件文型判断、その他の判断の3カテゴリーの人数について χ^2 検定（適合性の検定）を行ったところ、問2を除く3つの設問において有意差が認められた（問1： $\chi^2(2) = 16.55$, $p < .01$; 問3： $\chi^2(2) = 9.19$, $p < .05$; 問4： $\chi^2(2) = 16.55$, $p < .01$ ）。この結果は、双条件型の規則に関しては、事例の正誤が明示される前の段階で、ほとんどの実験参加者が規則を適切に解釈できていたことを示唆する。

訓練試行とテスト試行

訓練試行として行われた問1から問4の4枚

カード問題における正答者数を Table 3 に示す。ここで示される正答者数は、かなり低い値であり、相当に困難な課題であったことを表す。これらの各設問における正答者数・誤答者数について、正誤明示条件と非明示条件の間で χ^2 検定（独立性の検定）を行った。その結果、全ての設問において有意差が認められなかった。

また、この4枚カード問題は、「規則通りにカードが作られているかどうかを確認するために、いま見えていない裏側の面を確かめる必要のあるカードを選択」することが求められていたので、4枚すべてのカードを選択しないという反応も理論的にはありうるが、現実的には通常ほとんど起こらない反応である。しかしながら、本実験では Table 4 に示すように少数ながら、このような反応が現れた。

心理統計学に関するテスト試行として行われた問5と問6の正答者数を Table 5 に示す。各設問における正答者数・誤答者数について、正誤明示条件と非明示条件の間で χ^2 検定（独立性の検定）を行ったところ、全ての設問において有意差が認められなかった。Table 3 ならびに Table 5 に示

Table 2 事例正誤判断課題に対する主な反応パターンの人数

反応パターン	問1	問2	問3	問4
条件文型判断	0(0.00)	3(0.27)	2(0.18)	0(0.00)
双条件文型判断	10(0.91)	6(0.55)	9(0.82)	10(0.91)
その他の判断	1(0.09)	2(0.18)	0(0.00)	1(0.09)

(注：太枠は正しい解釈を表す。括弧内は比率（人数/ n ）を示す（ $n=11$ ）。)

Table 3 4枚カード問題における正答者数

	n	問1	問2	問3	問4
正誤非明示条件	11	2(0.18)	0(0.00)	0(0.00)	2(0.18)
正誤明示条件	11	0(0.00)	0(0.00)	2(0.18)	4(0.36)

(注：括弧内は比率（人数/ n ）を示す。)

Table 4 全てのカードを選択しないという反応の頻度

	<i>n</i>	問 1	問 2	問 3	問 4
正誤非明示条件	11	1(0.09)	3(0.27)	1(0.09)	4(0.36)
正誤明示条件	11	3(0.27)	1(0.09)	2(0.18)	3(0.27)

(注：括弧内は比率（人数/*n*）を示す。)

Table 5 テスト試行における正答者数

	<i>n</i>	問 5-(1)	問 5-(2)	問 5-(3)	問 5-(4)	問 6
正誤非明示条件	11	9(0.82)	8(0.72)	6(0.55)	6(0.55)	8(0.72)
正誤明示条件	11	8(0.72)	9(0.82)	9(0.82)	8(0.72)	8(0.72)

(注：括弧内は比率（人数/*n*）を示す。)

される正誤明示条件と正誤非明示条件の間で、全ての設問において有意差が認められなかったが、これは統計学の規則（知識）に関する事例の正誤を示すことは、その規則に関する4枚カード問題の遂行にも、また心理統計学に関する設問の解決にも顕著な影響を及ぼさないことを表す。

訓練試行として実施した問1から問4の4枚カード問題とテスト試行として実施した問5、問6に分類して、各条件における平均正答率をTable 6に示す。平均正答率について2要因2×2水準混合計画の分散分析を行ったところ、課題要因について主効果が認められ ($F(1, 20) = 39.947, p < .01$)、事例正誤要因の主効果と交互作用は認められなかった。この分散分析の結果は、事例の正誤明示要因はどちらの課題の遂行にも影響を与えることはないが、心理統計学の問題は4枚カード問題より容易であったことが示された。

考察

Table 2に示されるように、事例正誤判断課題においては双条件文については正答者が多く、主要な反応パターンの間における人数についての χ^2

検定で有意差が認められた一方で、条件文については正答者が主要な反応パターンの中でも最も多かったけれど、同様の χ^2 検定で有意差が認められなかった。この結果は、統計的仮説検定の論理体系に含まれる条件文の解釈は困難であるが、双条件文については適切に解釈されていたことを示す。

この事例正誤判断課題における結果に加えて、4枚カード問題において正答者数が少ないことと、正誤明示要因の主効果が有意ではないことの両面から、統計的仮説検定に関する知識（4枚カード問題における規則）について解釈することと、その知識について推論することは異なる認知過程であり、正しく解釈していても推論が困難になることがあることを示す。これは、中垣（1989）がアルファベットと数字を素材とする4枚カード問題において示した結果と整合する。このような結果は、規則として提示された条件文を適切に解釈することと、それに基づいて推論を行うことは難易度が異なることを意味するものであり、それらの認知的処理過程が異なることを示唆する。Brain & O'Brien（1991）が提唱する条件文に関する認知的処理のモデルは、語義の解釈、命題論理

Table 6 課題別の平均正答率

		<i>n</i>	課題種類要因		正誤明示要因における各条件の平均
			問 1～問 4 (4枚カード問題)	問 5, 問 6 (心理統計問題)	
正誤明示要因	正誤非明示条件	11	0.16(0.22)	0.67(0.35)	0.42(0.39)
	正誤明示条件	11	0.23(0.25)	0.76(0.34)	0.50(0.40)
課題種類要因における各条件の平均			0.19(0.24)	0.72(0.35)	0.46(0.40)

(注：括弧内は標準偏差を示す。)

推論プログラム、語用論の原理に分けており、本研究で示された結果と比較すると、語義の解釈に相当する解釈のプロセスと、その他の推論のプロセスには明確な相違が示されている点に関しては一致する。

また、Table 3 に示される 4 枚カード問題における正答者数についても、Table 5 に示される統計的仮説検定に関する設問（問 5、問 6）における正答者数についても、正誤明示条件と非明示条件の間で有意差が認められなかった。その一方で、Table 6 に示される課題種類要因の主効果が有意であった。これらの結果は、事例の正誤を明示することによる効果がなかったことを意味すると共に、4 枚カード問題において求められる推論を経験することによって、統計的仮説検定に関する知識が精緻化されるかどうか、あるいは訓練試行として位置づけられていた 4 枚カード問題で直接的には触れられていなかった知識が活性化するかどうかについても、確認できなかったことを意味する。ただし、Table 4 に示されるように、全てのカードを選択しないという通常は起こりにくい反応が起きたことも本実験の特徴的な結果であった。4 枚カード問題は、素材によっては難解であることが元々報告されている課題であり、心理統計学を苦手とする学生が多いことも指摘されている。学生にとって苦手であることが多い心理統計学の内容を、4 枚カード問題の素材にしたため、過剰に困難な課題となり、課題を遂行する動機が低減したとも考えられる。

本研究は、統計的仮説検定における論理体系に条件文あるいは双条件文が含まれていることと、4 枚カード問題が条件文に関する推論を求める課題であるという類似性に着目したが、訓練試行として位置づけられた 4 枚カード問題が困難であった。統計的仮説検定のように難解な論理体系を理解することや、4 枚カード問題のように困難な推論を遂行することを促進するためには、色々な誤認や見落としをしていることに気がつくことを支援すると共に、動機の低下を防ぐという効果の観点から、一部の統計学教育において試みられている（）のように、協働問題解決の影響を検討するべきである。また、論理的に推論するためにも、統計的仮説検定の論理体系全体を理解するためにも、そこに含まれる法則性や規則を正しく解釈すること必要なはずであるが、それらを事例に当てはめた

際の正誤を把握することだけでは、推論を効果的に支援するには至らないことも示された。事例正誤に関する知識だけでは十分でないとするならば、それと共に、何をすればよいのか、どのように考えるべきなのか、という手続きを支援することの必要性が考えられる。Yamamoto (2003) は、家電製品の取り付け作業を題材として、その作業手順の把握を支援することの効果を示している。ここでいう作業手順は、手続き的知識ともいえるものであり、1 つ 1 つの手続きの前後の関連性を理解することも促し、それによって個々の知識の間の論理手関係についての理解を促進することも期待されるものであり、今後検討されるべき視点の 1 つである。

本研究の目的の 1 つ目は、条件文および双条件文の事例正誤を把握することによって、課題の論理構造を理解し、条件文・双条件文に基づく推論が促進されることを明らかにすることであった。この目的に関して、条件文、双条件文の事例の正誤を把握するだけでは、それに関する推論が促進されるとは限らないことが示された。また、目的の 2 つ目は、条件文・双条件に基づく推論が促進されることによって、それらの構造を含む論理体系についての理解が促されることについて、統計的仮説検定を題材として検討することであった。この目的に関しては、本研究の結果からは検討困難であった。そのため、今後は、難解な事象についての理解と推論を支援する方策として、協働問題解決や手続き的知識の確立などを多角的に検討することが課題とされた。

引用文献

- Brain, M. D. S. & O'Brien, D.P. (1991). A theory of IF: A lexical entry, reasoning program and pragmatic principles. *Psychological Review*, 98, 182-203.
- Cheng, W. P. & Holyork, K. J. (1985) Pragmatic Reasoning Schemas. *Cognitive Psychology*, 17, 391-416.
- Dolor, J. & Noll, J. (2015). Using guided reinvention to develop teachers' understanding of hypothesis testing concepts. *Statistics Education Research Journal*, 14, 60-89.
- Dunn, P. K., Carey, M. D., Richardson, A. M., & McDonald, C. (2016). Learnig the language of statistics: Challenges and teaching approaches. *Statistics Education Research Journal*, 15, 8-27.

- 堀内正彦 (2000). Wason の選択課題の発展—試論的
実験— 基礎心理学研究, *18*, 156-164.
- 堀内正彦・永田陽子 (2020). 大学生にとって困難な
課題の解決に関わる認知構造—心理統計学を題材
として— 応用心理学研究, *46*, 130-138.
- Johnson-Laird, P.N., Legrenzi, P., & Legrenzi, M.S.
(1972). Reasoning and a sense of reality. *British
Journal of Psychology*, *63*, 395-400.
- 中垣啓 (1989). 抽象的 4 枚カード問題における課題
変質効果について 教育心理学研究, *37*, 36-45.
- 高橋和弘・服部雅史 (1996). 演繹的推論 市川伸一
(編) 認知心理学 4 思考 (pp.15-35) 東京大学出
版会
- Yamamoto, H. (2003). Signaling procedural instruc-
tions and supporting comprehension by older
adults. *Proceedings of XVth Triennial Congress of
the international Ergonomics Association*, *4*, 649-
651.
- Wason, P.C. (1968). Reasoning about a rule. *Quarterly
Journal of Experimental Psychology*, *20*, 273-281.

問1. ここでの検定統計量 F 値は、明るさ要因による分散と誤差変動の分散の比率であり、明るさ要因による分散が分子、誤差変動による分散を分母にします。

各条件の平均の大きさと検定統計量の値の間には、大まかな関数的な関係として次のような規則性があります。

「各条件の平均 $\bar{x}_a, \bar{x}_b, \bar{x}_c$ が近似した場合には、 F 値は 0 に近くなります。反対に、3つの平均 $\bar{x}_a, \bar{x}_b, \bar{x}_c$ がかけ離れた場合には、 F 値は大きな値になります。」

このかぎ括弧の記述に関して、以下の例は正しいか誤りかを判断して下さい。正しいと思う場合には○、誤りと思う場合には×を記して下さい。

事例	記入欄
各条件の平均 $\bar{x}_a, \bar{x}_b, \bar{x}_c$ が近い値で、 F 値は 0 に近い値。	
各条件の平均 $\bar{x}_a, \bar{x}_b, \bar{x}_c$ が近い値で、 F 値は大きい値。	
各条件の平均 $\bar{x}_a, \bar{x}_b, \bar{x}_c$ がかけ離れた値で、 F 値は 0 に近い。	
各条件の平均 $\bar{x}_a, \bar{x}_b, \bar{x}_c$ がかけ離れた値で、 F 値は大きい値。	

記入欄に回答したら、次のページの正答を確認して下さい。回答が間違っている場合でも、修正せずに、そのままにしておいて下さい。

以下の4枚のカードはある学生が、自習のために作成した暗記用のカードです。表面には各条件の平均が近似しているかどうか、裏面には F 値が 0 に近いかが書いてあります。上のかぎ括弧の記述に照らして、カードを正しく作ったかどうかを確認して欲しいと、この学生から頼まれました。今見えていない反対側の面を確認する必要があるカードの括弧内に○、確認する必要のないカードの括弧内に×を記して下さい。

各条件の平均： 近い値 (表面) ()	各条件の平均： かけ離れた値 (表面) ()	F 値： 0 に近い値 (裏面) ()	F 値： 大きい値 (裏面) ()
-----------------------------------	--------------------------------------	-------------------------------------	-----------------------------------