



New ESRI Working Paper No.40

欠測を伴う標本調査に基づく統計的仮説検定について

間 真実

June 2017



内閣府経済社会総合研究所
Economic and Social Research Institute
Cabinet Office
Tokyo, Japan

New ESRI Working Paper は、すべて研究者個人の責任で執筆されており、内閣府経済社会総合研究所の見解を示すものではありません（問い合わせ先：<https://form.cao.go.jp/esri/opinion-0002.html>）。

新ESRIワーキング・ペーパー・シリーズは、内閣府経済社会総合研究所の研究者および外部研究者によってとりまとめられた研究試論です。学界、研究機関等の関係する方々から幅広くコメントを頂き、今後の研究に役立てることを意図して発表しております。

論文は、すべて研究者個人の責任で執筆されており、内閣府経済社会総合研究所の見解を示すものではありません。

The views expressed in “New ESRI Working Paper” are those of the authors and not those of the Economic and Social Research Institute, the Cabinet Office, or the Government of Japan.

欠測を伴う標本調査に基づく統計的仮説検定について¹

間 真実（内閣府景気統計部）

平成 29（2017）年 6 月

要旨

データの欠測は、統計的推論にゆがみをもたらす。推定においては欠測バイアスが問題となる一方で、検定においては、分析者によって意図される検定と実行される検定との乖離が問題となる。意図される検定と実行される検定との乖離は、検定統計量に生じる欠測バイアスが原因であるが、この乖離には欠測バイアスだけでなく分析者の認知効果も作用する。

本稿は、意図される検定と実行される検定の乖離を測る尺度として「誤検定確率」を定義し、その定性・定量分析を行う。誤検定確率は、棄却すべきでないのに棄却する確率と棄却すべきなのに棄却しない確率との和である。定性分析により、誤検定確率の漸近収束値は、検定の有意水準、標本サイズ、欠測率の漸近収束値、回答・無回答群間の異質性指標の増加関数であることが示された。定量分析により、誤検定確率を有意水準や標本サイズによってコントロールすることは事実上できないことが明らかとなった。誤検定確率に対する標本サイズの効果が、非常に弾力的でありかつ他の効果を凌駕するためである。本稿ではまた、内閣府「機械受注統計調査」について、誤検定確率の決定要因のうちで観測されない変量である異質性指標の値の妥当な範囲を調べる方法を例示する。

¹ 本研究は、内閣府経済社会総合研究所（ESRI）景気統計部の平成 28 年度業務「欠測値補完に関する調査研究」の成果の一部である。研究に当たっては次の方々から有益な示唆をいただいた（肩書と所属は 2017 年 3 月時点）。星野崇宏教授（慶應義塾大学）、土屋隆裕教授（統計数理研究所）、元山斉准教授（青山学院大学）、杉原茂次長（ESRI）、坪内浩総括政策研究官（ESRI）、桑原進総務部長（ESRI）、西崎寿美景気統計部長（ESRI）、北島美雪景気統計部長補佐（ESRI）。分析に用いたデータセットの作成にあたっては、景気統計部機械受注班の方々の協力をいただいた。感謝申し上げます。もちろん本稿に関する責任は筆者にある。

1. はじめに

標本調査においては、通常、調査客体の無回答によりデータに欠測が生じる。欠測のあるデータによって統計的推論はゆがめられる²。推定においては、第1に、欠測バイアスの可能性、第2に、推定精度の低下という問題が伴う。検定においては、検定統計量に生じる欠測バイアスに起因して、分析者によって意図される検定と実行される検定との間の乖離が問題となる。

推定においては、精度低下の問題は欠測バイアス問題と切り離して考えることができ、調査で事前に予想される回答率を考慮に入れて標本サイズを設定することで対処できる。このため、データの欠測が推定にもたらす問題は欠測バイアスに尽きるといっても過言ではない。これに対して、検定においては、検定統計量の欠測バイアスに、完全ケース分析を行う分析者の認知効果が作用するため、欠測バイアスよりも複合的な問題が形成されている。

本稿は、欠測を伴うデータに基づく統計的推論における問題をいかに評価するかという課題の中で特に検定について考察する。欠測が推定にもたらす影響の大きさを間接的に測る尺度として最も基本的なものは、欠測率（あるいは回答率）である。実際に多くの標本調査では、回答率の値が統計データの信頼性に関する情報として調査結果に明示される。確かに、精度低下の問題に関しては、欠測率と標本誤差の間の単調な関係から、欠測率により推定精度の低下幅は正確に評価できる。しかし、欠測バイアスの問題に関しては、バイアスを欠測率により一般的に定量評価することは容易ではない。例えば、標本平均による母集団平均の推定においては、欠測率と欠測バイアスの絶対値の間には線形関係があるものの、欠測バイアスが正負のいずれであるかは、欠測率の情報からは分からない。また、より高い次数のモーメントの推定においては、欠測率と欠測バイアスの関係は非単調である。

図1は、標本平均について、欠測率と欠測バイアスの関係をグラフで示したものである。無作為抽出による標本調査において、欠測率の漸近収束値 π_{NR} 、回答標本の標本平均漸近収束値 μ_R 、無回答標本の標本平均漸近収束値 μ_{NR} に対して、回答標本の標本平均 $\hat{\mu}_{CC}$ のバイアスは、 $\text{Bias}(\hat{\mu}_{CC}) = \pi_{NR}(\mu_R - \mu_{NR})$ である。したがって、図1のグラフ(0)~(2)に示すとおり、標本平均の欠測バイアスと欠測率の関係には3通りの可能性がある。条件 $\mu_R = \mu_{NR}$ が成り立つならばグラフ(0)、条件 $\mu_R > \mu_{NR}$ が成り立つならばグラフ(1)、条件 $\mu_R < \mu_{NR}$ が成り立つならばグラフ(2)の関係となる。条件 $\mu_R = \mu_{NR}$ の下では欠測バイアスは生じない。条件 $\mu_R > \mu_{NR}$ の下では正の欠測バイアス、条件 $\mu_R < \mu_{NR}$ の下では負の欠測バイアスが生じ、いずれの場合も欠測バイアスの絶対値は欠測率に比例する。

【図1】

² 欠測を伴うデータの統計的処理に関する教科書としては、阿部（2016）、高井・星野・野間（2016）、星野（2009）、Little and Rubin（2002）、Molenberghs and Kenward（2007）、Molenberghs et al.（2014）などがある。

図2は、標本分散について、欠測率と欠測バイアスの関係をグラフで示したものである。無作為抽出による標本調査において、欠測率の漸近収束値 π_{NR} 、回答標本の標本平均漸近収束値 μ_R 及び標本分散漸近収束値 σ_R^2 、無回答標本の標本平均漸近収束値 μ_{NR} 及び標本分散漸近収束値 σ_{NR}^2 に対して、回答標本の標本分散 $\hat{\sigma}_{CC}^2$ のバイアスは、 $\text{Bias}(\hat{\mu}_{CC}) = \pi_{NR}(\mu_R - \mu_{NR})$ 及び $\text{Bias}(\hat{\sigma}_{CC}^2) = \pi_{NR}[(\mu_R - \mu_{NR})^2 \pi_{NR} - \{(\mu_R - \mu_{NR})^2 + (\sigma_{NR}^2 - \sigma_R^2)\}]$ である。したがって、図2のグラフ(0)~(3)に示すとおり、標本分散の欠測バイアスと欠測率の関係には4通りの可能性がある。条件 $\mu_R = \mu_{NR}$ 及び $\sigma_R^2 = \sigma_{NR}^2$ が成り立つならばグラフ(0)の関係であり、欠測バイアスは生じない。あとは、パラメータの大小関係に応じてグラフ(1)~(3)の関係が生じる。グラフ(1)は正の欠測バイアス、グラフ(2)は負の欠測バイアスが生じ、いずれの場合も欠測バイアスの絶対値は欠測率に対して単調に増加する。しかし、グラフ(3)の場合は、欠測バイアスと欠測率の関係に単調性はない。

【図2】

図1及び2の例では、それぞれの可能性のいずれが成立しているかをデータからは検証することができないという問題に直面している。標本平均の例(図1)では、欠測バイアスの符号を、標本分散の例(図2)では、バイアスの符号のみならず、バイアスの大きさが欠測率に対して単調・非単調のいずれであるかをデータから検証することはできない。これらの例からも、欠測を含む統計データの信頼性に関する情報としては、欠測率だけでは不十分であることが分かる。

欠測が統計的推論に対してもつ効果を一般的に計測することは困難である。このことは、図1及び2の例で、標本平均と標本分散で欠測の効果の現れ方が異なることから分かる。そこで、本稿では問題を限定し、特に母集団平均に関する検定について欠測の効果を評価することを試みる。

統計的推論における欠測問題は、推定においては欠測バイアスとして現れるのに対して、検定においては、分析者によって意図される検定と実行される検定との間の乖離として現れる。次節で説明する通り、意図される検定と実行される検定の乖離は、検定統計量に生じる欠測バイアスが原因であるが、分析者の認知効果も作用している。そこで本稿は、統計的仮説検定における欠測効果を、次節で定義する「誤検定確率」によって計測する。回答群と無回答群の異質性の上限について先験的な知見があれば、誤検定確率により欠測効果を評価できる。比較静学によると、直感的にも予想される結果であるが、誤検定確率は、欠測率、標本サイズ、検定の有意水準、及び適当に定義した回答群と無回答群の異質性の指標の増加関数である。定性分析の結果(第3.1節)からは、標本サイズと有意水準を適当に決めることで誤検定確率を抑制できる可能性が示唆されるものの、定量分析からは否定的な結果が得られた(第3.2節)。誤検定確率の決定要因のうち、回答群と無回答群の異質性は観測さ

れない変数である。第4節では、異質性指標の妥当な範囲をシミュレーションによって調べる方法を、内閣府「機械受注統計調査」のデータを用いて実演する。第5節は、結語である。

2. 統計的仮説検定における欠測効果の尺度

回答標本に基づいて統計的推論を行うことを「完全ケース分析」と呼ぶ。母集団平均 μ について、帰無仮説 $H_0: \mu = \mu_0$ 、対立仮説 $H_1: \mu \neq \mu_0$ 、有意水準 α の両側検定を、完全ケース分析により行うことを考える。完全ケース分析で検定を行うということは、分析者の意識では欠測バイアスの存在が無視されているということになる。このため、帰無仮説 $H_0: \mu = \mu_0$ と対立仮説 $H_1: \mu \neq \mu_0$ を検定しているつもりでも、実際には、回答標本の標本平均漸近収束値 μ_R についての帰無仮説 $H'_0: \mu_R = \mu_0$ と対立仮説 $H'_1: \mu_R \neq \mu_0$ の下での推定量の分布に基づいて検定を行っている。この分析者の意識と実際のずれによって、検定を誤る可能性が生じる。その図解を、図3に示す。

【図3】

図3で、 $F_{MNAR; H_0}$ 及び $f_{MNAR; H_0}$ はランダムでない欠測の下で帰無仮説 $H_0: \mu = \mu_0$ が正しい場合の推定量 $\hat{\mu}_{CC}$ の確率分布関数及び確率密度関数、 $F_{MCAR; H_0}$ 及び $f_{MCAR; H_0}$ は完全にランダムな欠測の下で帰無仮説 $H_0: \mu = \mu_0$ が正しい場合の推定量 $\hat{\mu}_{CC}$ の確率分布関数及び確率密度関数である。図3は、条件 $\mu_{NR} \geq \mu_0$ が成立する場合である。条件 $\mu_{NR} < \mu_0$ が成立する場合は、2つの確率密度関数 $f_{MNAR; H_0}$ と $f_{MCAR; H_0}$ のグラフの位置が逆転する。

完全ケース分析を行う分析者は、推定値が条件 $\hat{\mu}_{CC} \leq F_{MCAR; H_0}^{-1}(\alpha/2)$ 又は条件 $F_{MCAR; H_0}^{-1}(1 - \alpha/2) \leq \hat{\mu}_{CC}$ を満たすときに帰無仮説 $H_0: \mu = \mu_0$ を棄却する。しかし、帰無仮説 H_0 を棄却できる条件は、正しくは、 $\hat{\mu}_{CC} \leq F_{MNAR; H_0}^{-1}(\alpha/2)$ 又は $F_{MNAR; H_0}^{-1}(1 - \alpha/2) \leq \hat{\mu}_{CC}$ である。従って、図3の場合、推定値が $F_{MNAR; H_0}^{-1}(\alpha/2)$ よりも大きく $F_{MCAR; H_0}^{-1}(\alpha/2)$ よりも小さいときに、分析者は帰無仮説 H_0 を棄却すべきでないのに棄却し、推定値が $F_{MNAR; H_0}^{-1}(1 - \alpha/2)$ よりも大きく $F_{MCAR; H_0}^{-1}(1 - \alpha/2)$ よりも小さいときに、分析者は帰無仮説 H_0 を棄却すべきなのに棄却しないことになる。条件 $\mu_{NR} < \mu_0$ が成立する場合についても同様に考えると、棄却すべきでないのに棄却する確率 ξ_I と棄却すべきなのに棄却しない確率 ξ_{II} は、それぞれ次式で与えられる。

$$\xi_I = \begin{cases} 1 - \frac{\alpha}{2} - F_{MNAR; H_0} \left(F_{MCAR; H_0}^{-1} \left(1 - \frac{\alpha}{2} \right) \right) & (\mu_{NR} < \mu_0) \\ F_{MNAR; H_0} \left(F_{MCAR; H_0}^{-1} \left(\frac{\alpha}{2} \right) \right) - \frac{\alpha}{2} & (\mu_{NR} \geq \mu_0) \end{cases} \quad (1)$$

$$\xi_{II} = \begin{cases} \frac{\alpha}{2} - F_{MNAR; H_0} \left(F_{MCAR; H_0}^{-1} \left(\frac{\alpha}{2} \right) \right) & (\mu_{NR} < \mu_0) \\ F_{MNAR; H_0} \left(F_{MCAR; H_0}^{-1} \left(1 - \frac{\alpha}{2} \right) \right) - \left(1 - \frac{\alpha}{2} \right) & (\mu_{NR} \geq \mu_0) \end{cases} \quad (2)$$

棄却すべきでないのに棄却する確率 ξ_I と棄却すべきなのに棄却しない確率 ξ_{II} の和 ξ を「誤検定確率」と呼ぶことにする。誤検定確率 ξ は、次式で与えられる。

$$\xi \equiv \xi_I + \xi_{II} = \left| 1 - F_{MNAR; H_0} \left(F_{MCAR; H_0}^{-1} \left(\frac{\alpha}{2} \right) \right) - F_{MNAR; H_0} \left(F_{MCAR; H_0}^{-1} \left(1 - \frac{\alpha}{2} \right) \right) \right| \quad (3)$$

欠測データメカニズムが完全にランダムな欠測であれば、上式の関数 $F_{MNAR; H_0}$ は関数 $F_{MCAR; H_0}$ に置き換わるので、 $\xi = 0$ となる。また、欠測率が0であれば、上式の関数 $F_{MNAR; H_0}$ は関数 $F_{MCAR; H_0}$ と等しいので、 $\xi = 0$ となる。本稿では、統計的仮説検定における欠測効果の尺度として、誤検定確率 ξ に注目する。

3. 分析

(3)式で与えられる誤検定確率 ξ の性質を一般的に明らかにすることは容易ではない。以下では、決定論的欠測のもとで誤検定確率 ξ の漸近論的性質を解明する。標本サイズ n （回答標本だけでなく無回答標本も含めた標本のサイズ）の増大に対して、中心極限定理により、次が成り立つ³。

$$\frac{\hat{\mu}^{CC} - \frac{\mu_0 - \pi_{NR}\mu_{NR}}{1 - \pi_{NR}}}{\sigma_R / \sqrt{(1 - \pi_{NR})n}} \Bigg|_{MNAR, H_0, \pi_{NR}, \mu_{NR}, \sigma_R^2} \xrightarrow{d} N(0, 1)$$

$$\frac{\hat{\mu}^{CC} - \mu_0}{\sigma_R / \sqrt{(1 - \pi_{NR})n}} \Bigg|_{MCAR, H_0, \sigma_R^2} \xrightarrow{d} N(0, 1)$$

すなわち、欠測率の漸近収束値 π_{NR} 、回答標本の標本平均漸近収束値 μ_R 及び標本分散漸近収束値 σ_R^2 、無回答標本の標本平均漸近収束値 μ_{NR} 及び標本分散漸近収束値 σ_{NR}^2 を与件としたときに、完全ケース分析の推定量 $\hat{\mu}^{CC}$ が正規分布に分布収束する。従って、帰無仮説「 $H_0: \mu = \mu_0$ 」の下で、誤検定確率の漸近収束値 $\xi^* \equiv \lim_{n \rightarrow \infty} \xi$ について次式が成り立つ。

³ 決定論的欠測は、確率論的欠測において欠測データメカニズムを与件としたときの条件付き不完全データと等しい。決定論的欠測と確率論的欠測については、例えば土屋（2009）の11.5節参照。

$$\xi^* = \begin{cases} 1 - \Phi\left(\Phi^{-1}\left(\frac{\alpha}{2}\right) + \frac{\pi_{NR}\sqrt{nr}}{\sqrt{1-\pi_{NR}}}\right) - \Phi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \frac{\pi_{NR}\sqrt{nr}}{\sqrt{1-\pi_{NR}}}\right) & (\mu_{NR} < \mu_0) \\ \Phi\left(\Phi^{-1}\left(\frac{\alpha}{2}\right) + \frac{\pi_{NR}\sqrt{nr}}{\sqrt{1-\pi_{NR}}}\right) + \Phi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \frac{\pi_{NR}\sqrt{nr}}{\sqrt{1-\pi_{NR}}}\right) - 1 & (\mu_{NR} \geq \mu_0) \end{cases} \quad (4)$$

ただし、ここで新たに変数 $r \equiv (\mu_{NR} - \mu_0)/\sigma_R$ を定義した。定数 μ_0 は、完全ケース分析を行う分析者が帰無仮説 H_0 において想定する母集団平均 $\mu = (1 - \pi_{NR})\mu_R + \pi_{NR}\mu_{NR}$ の値であるから、変数 r は、回答者と無回答者の異質性の尺度である。例えば、異質性指標の絶対値 $|r| = 1$ は、無回答者の平均と母集団の平均（として分析者が帰無仮説で想定する値）との差が回答者の標準偏差と同じ大きさとなるほどの異質性を表している。変数 r の絶対値が大きいほど、欠測データメカニズムが「ランダムでない欠測」である度合いは強いといえる。

誤検定確率の漸近収束値 ξ^* の値を決める変数は、標本サイズ n 、検定の有意水準 α 、欠測率の漸近収束値 π_{NR} 、異質性の指標 r の4つである。このうち、標本サイズと有意水準の値は、分析者が決定する。欠測率の漸近収束値は、標本の欠測率により一致推定が可能である。異質性指標 r の値に関しては、データから情報を得ることはできず、先験的な知見からその上限が分かっているなければならない。異質性指標 r の値は、分析対象となる変数の分布と欠測データメカニズムによって決まる。第4節では、内閣府「機械受注統計調査」のデータを用いて、ロジットモデルに基づくシミュレーションにより異質性指標 r の妥当な範囲を調べる方法の一例を示す。

以下では、(4)式で与えられる誤検定確率 ξ の漸近収束値 ξ^* について、第3.1節で比較静学により定性的に、第3.2節で数値解析により定量的に、その性質を明らかにする。

3.1. 定性分析

(4)式に関する比較静学を行う。条件 $\mu_{NR} \neq \mu_0$ の下で次が成り立つ⁴。

$$\frac{\partial \xi^*}{\partial \pi_{NR}}, \frac{\partial \xi^*}{\partial |r|}, \frac{\partial \xi^*}{\partial \alpha}, \frac{\partial \xi^*}{\partial n} > 0$$

誤検定確率の漸近収束値 ξ^* は、欠測率の漸近収束値 π_{NR} 、回答者と無回答者の異質性の尺度 $|r|$ 、検定の有意水準 α 、標本サイズ n の増加関数である。

上記の結果に意外性はない。欠測率が大きいほど、あるいは回答者と無回答者の異質性が大きいほど、回答標本による検定は欠測バイアスによって歪められる。図3において、2つの確率密度関数の間の距離が大きくなることによって棄却すべきでないのに棄却する確率 ξ_I と棄却すべきなのに棄却しない確率 ξ_{II} の値が大きくなる効果である。この効果を「バイアス効果」と呼ぶことにする。また、検定の有意水準が高くなると、図3において、臨界点が

⁴ 導出は比較的容易であるが、念のため有意水準 α に関する結果についてのみ本稿末尾の補論に示す。

最頻値側へ寄せられることによって確率 ξ_I と確率 ξ_{II} の値が大きくなる効果が表れる。この効果を「有意水準効果」と呼ぶことにする。そして、標本サイズが増大すると、回答標本サイズも増大するので、欠測データメカニズムに関わりなく推定量の分散は低下する。これにより、図3においては、2つの確率密度関数がそれぞれの平均値を中心に先鋭化するため、確率 ξ_I と確率 ξ_{II} の値は大きくなる。この効果を「先鋭化効果」と呼ぶことにする。

ここで、比較静学の結果から、バイアス効果を有意水準効果あるいは先鋭化効果で相殺できる可能性が示唆される。すなわち、条件 $\mu_{NR} \neq \mu_0$ の下で次式が成り立つ。

$$\left. \frac{d\alpha}{d\pi_{NR}} \right|_{d\xi^*=0}, \left. \frac{d\alpha}{d|r|} \right|_{d\xi^*=0}, \left. \frac{dn}{d\pi_{NR}} \right|_{d\xi^*=0}, \left. \frac{dn}{d|r|} \right|_{d\xi^*=0} < 0$$

上式は、欠測率の上昇あるいは回答者と無回答者の異質性の増大に対して、有意水準の引き下げあるいは標本サイズの拡大によって、誤検定確率（の漸近収束値）を抑制できる可能性があることを示している。

最後に、条件 $\mu_{NR} \neq \mu_0$ の下で成り立つ次式の通り、先鋭化効果を有意水準効果で相殺できる可能性も注目に値する。

$$\left. \frac{d\alpha}{dn} \right|_{d\xi^*=0} < 0$$

一般的に標本サイズが大きいほど検定は有意な結果を出しやすいが、欠測バイアスを含む選択バイアスによって誤った検定が行われている可能性が先鋭化効果により大きくなっていることに注意が必要である。標本サイズが大きければ有意水準を引き下げて検定を厳格化することで、先鋭化効果を抑制することが考えられる。そのような処理が実際に可能かどうかを、次節の数値解析による定量分析で確認する。

3.2. 定量分析

欠測率の増大に対して誤検定確率の漸近収束値 ξ^* （以下では単に「誤検定確率」）がどのように増加するかを、図4のグラフに示す。有意水準 α 、標本サイズ n 、異質性指標 r の値の組合せを何通りか決めて、それぞれの値に対して欠測率と誤検定確率の関係を示す。図3からも予想される通り、誤検定確率がほぼ100%となるような欠測率の領域は比較的大きく、欠測を伴う標本調査に基づいて統計的仮説検定を行ってはならない場合が一般的に多くあることが分かる。図4で想定している条件 $(\alpha, n, r) \in \{0.001, 0.1\} \times \{1000, 10000\} \times \{0.5, 1.0, 1.5\}$ では、欠測率が30%を超えると、誤検定確率はほぼ100%となる。また図4のグラフからは、例えば、「サイズ1,000の標本で有意水準10%の検定を行う際に誤検定確率を10%以下に抑えたければ、異質性指標 r の値にもよるが、欠測率は高くても数%でなければならない」ということも分かる。

【図4】

次に、欠測率の上昇に対して検定の厳格化（有意水準の引き下げ）によって誤検定確率を抑制する可能性を確認するために、欠測率と有意水準に対する誤検定確率の等量線を図 5 に示す。図 5 では、標本サイズと異質性指標についての条件 $(n, r) \in \{500, 5000\} \times \{0.5, 1.0\}$ のそれぞれの下での等量線を示す。等量線は、右下がりであることは比較静学の結果から当然であるとして、原点に対して凸であることが図から分かる。誤検定確率を 10% 以下に抑えたいければ、欠測率と有意水準の値の組合せは、10% 等量線よりも左下方の領域になければならない。標本サイズが大きいほど、あるいは異質性指標の値が大きいほど、この領域は狭くなる。標本サイズ $n = 500$ 、異質性指標 $r = 0.5$ の場合でさえ、欠測率が十数%を超えると、誤検定確率 10% 以下の制約下で許容される有意水準の範囲は極端に狭くなる。数値解析の結果からは、欠測率の上昇に対して検定の厳格化によって誤検定確率を抑制する方法は、通常の標本調査においては実際的でないことが示された。

【図 5】

回答者と無回答者間の異質性の増大に対して検定の厳格化によって誤検定確率を抑制する可能性を確認するために、異質性指標と有意水準に対する誤検定確率の等量線を図 6 に示す。図 6 では、欠測率と標本サイズについての条件 $(\pi_{NR}, n) \in \{5\%, 10\%\} \times \{500, 5000\}$ のそれぞれの下での等量線を示す。ここでも、等量線は、右下がりであることは比較静学の結果から当然であるとして、原点に対して凸であることが図から分かる。欠測率が高いほど、あるいは標本サイズが大きいほど、誤検定確率が一定値以下に抑えられる領域は小さくなる。欠測率 $\pi_{NR} = 10\%$ 、標本サイズ $n = 5000$ の場合では、許容される異質性指標の範囲が低位側において非常に限られている様子が、図からみてとれる。図 6 の結果からは、例えば条件 $|r| = 1$ の場合では異質性の増大に対して検定の厳格化で誤検定確率を抑制する方法は、通常の標本調査においては実際的でないことが分かる。ただし、既述の通り、異質性の指標はデータの性質に大きく依存するため、統計調査ごとの個別性が大きいことには注意を要する。

【図 6】

欠測率の上昇に対して標本サイズの縮小によって誤検定確率を抑制する可能性を確認するために、欠測率と標本サイズ（の常用対数値）に対する誤検定確率の等量線を図 7 に示す。図 7 では、検定の有意水準と異質性指標についての条件 $(\alpha, r) \in \{1\%, 10\%\} \times \{0.5, 1.0\}$ のそれぞれの下での等量線を示す。ここでもまた、等量線は、右下がりであることは比較静学の結果から当然であるとして、原点に対して凸であることが図から分かる。有意水準が高いほど、あるいは異質性が大きいほど、誤検定確率が一定値以下に抑えられる領域は小さくな

る。有意水準 $\alpha = 10\%$ 、異質性 $r = 1$ の場合では、許容される欠測率の範囲が低位側において非常に限られている様子が、図からみてとれる。ここでもまた異質性の指標の取り得る範囲については注意を要するが、図7の結果から、例えば条件 $|r| = 1$ の場合では欠測率の上昇に対して標本サイズの拡大で誤検定確率を抑制する方法は、通常の標本調査においては実際的でないことが分かる。また、条件 $|r| = 0.5$ の場合でも、欠測率が10%程度であっても誤検定確率を10%以下に抑えるためには標本サイズは3桁程度でなければならず、かえって検定の検出力が低下するという問題に直面する。

【図7】

回答者と無回答者間の異質性の増大に対して標本サイズの縮小によって誤検定確率を抑制する可能性を確認するために、異質性指標と標本サイズ（の常用対数値）に対する誤検定確率の等量線を図8に示す。図8では、欠測率と有意水準についての条件 $(\pi_{NR}, \alpha) \in \{5\%, 10\% \} \times \{1\%, 10\% \}$ のそれぞれの下での等量線を示す。ここでもやはり、等量線は、右下がりであることは比較静学の結果から当然であるとして、原点に対して凸であることが図から分かる。欠測率が高いほど、あるいは有意水準が高いほど、誤検定確率が一定値以下に抑えられる領域は小さくなる。欠測率 $\pi_{NR} = 10\%$ 、有意水準 $\alpha = 10\%$ の場合では、許容される異質性指標の範囲が低位側において限定されている様子が、図からみてとれる。図8では標本サイズを常用対数変換していることから分かるように、異質性の増大に対して標本サイズの縮小によって誤検定確率の抑制を図る方法は、検出力低下の問題が制約となる。

【図8】

最後に、標本サイズが大きい場合に検定の厳格化によって誤検定確率を抑制する可能性を確認するために、標本サイズ（の常用対数値）と有意水準（の常用対数値）に対する誤検定確率の等量線を図9に示す。図9では、欠測率と異質性指標についての条件 $(\pi_{NR}, r) \in \{5\%, 10\% \} \times \{0.5, 1.0\}$ のそれぞれの下での等量線を示す。通常の標本調査における標本サイズの領域では、欠測率及び異質性に関する条件によっては有意水準1%で誤検定確率が10%以下に抑えられることもある。しかし、いわゆるビッグデータなどで考えられる標本サイズの領域では、先鋭化効果により有意水準をどれほど低くしても誤検定確率は100%に近い水準を維持する。一般的に、標本サイズの大きなデータによる検定は検出力が大きいものの、欠測ないし選択バイアスの先鋭化効果による誤検定を有意水準の引き下げによって抑制することは実際的ではないことが図9の結果から分かる。

【図9】

分析の結果をまとめると、次の通りである。第1に、回答者・無回答者間の異質性についての先験的な知見があれば、(4)式により誤検定確率の評価が可能である。定量分析の結果からは、通常の標本調査においても欠測率が数%程度の低水準でない限り、統計的仮説検定を行うべきではないことが分かる。第2に、欠測率や回答者・無回答者間の異質性の増加に対して検定の厳格化(有意水準の引き下げ)や標本サイズの縮小によって誤検定確率を抑制することは、実際的ではない。第3に、標本サイズの大きなデータでは検出力に余力があるため、誤検定確率を抑制するために有意水準を引き下げる余地が大きいと考えられるものの、定量分析によりこの方法は実際的でないことが分かった。

4. 「機械受注統計調査」への適用

誤検定確率の評価における最大の課題は、回答者・無回答者間の異質性の大きさが観測された情報からは検証できないことである。この節では、内閣府「機械受注統計調査」のデータを用いて、回答者・無回答者の異質性の尺度 r がとりうる値の妥当な範囲を調べる方法のひとつを実演する。

内閣府「機械受注統計調査」は、基本的に固定された作為抽出標本による月次調査である。国内の機械等を製造する企業のうち主要なものが選定され、現行では280社ベースの調査としている⁵。主要な調査項目は、発注部門別受注額と受注機種別受注額である。機械受注統計調査では、稀に無回答による欠測が生じているが、バランスパネルの利点を生かしてLOCF(直近値を欠測値に代入する方法。横置き補完法)により対処している。機械受注統計調査は、正確には標本調査ではないが、本節の分析では実演例を示すことを目的に、あたかも標本調査であるとして扱う。また、興味の対象となる変数は、企業ごと受注額の常用対数値とする。

異質性指標 r は、分析対象となる変数の母集団分布と欠測データメカニズムによって決まる。そこで、機械受注統計調査で欠測の出なかった調査月1時点分を適当に選び、下式に示すロジットモデルで定式化した欠測データメカニズムにより、欠測を発生させるシミュレーションを行うことで、異質性指標 r の取りうる妥当な範囲を調べる。

$$\ln \frac{Pr(R_i = 1|y_i)}{1 - Pr(R_i = 1|y_i)} = \psi_0 + \psi_1 y_i$$

ただし、変数 R_i は企業 i が変数 y_i の値と回答した場合に値1をとり回答しなかった場合に値0をとる2値変数、変数 y_i は企業 i の受注額の常用対数値である。第2パラメータ ψ_1 の値を妥当な範囲で動かし、それらの値ごとに第1パラメータ ψ_0 の値を調整して欠測率を一定値

⁵ 調査の目的は、「機械等製造業者の受注する設備用機械類の受注状況を調査することにより、設備投資動向を早期に把握し、経済動向分析の基礎資料を得る」(内閣府)こととされ、主要企業の作為抽出は、Gabaix(2011)のグラニューラ仮説(企業規模分布の裾野に位置する比較的少数の企業の動向がマクロ経済の変動の比較的大きな部分を説明する旨の仮説)と整合的である。

(10%)に固定する。パラメータ ψ_1 の絶対値が大きいほど、欠測データメカニズムが「ランダムでない欠測」である度合いが強く、異質性指標 r の絶対値は大きくなる。

表1は、欠測の生じなかった2011年5月分の機械受注統計調査データについて10,000回のシミュレーションを行った結果である。第2列から第5列までの値は各パラメータのシミュレーションにおける平均値である。第1列に示すとおり、パラメータ ψ_1 は、-3から3までの範囲を間隔0.5で動かした。 $\psi_1 = 0$ の場合は、「完全にランダムな欠測」であるから、 $r = 0$ となる(表1では省略している)。表の第6列には無回答者平均と回答者平均の差の平均値を示す。表の第7列に異質性指標 r の平均値を示すが、帰無仮説 $H_0: \mu = \mu_0$ における定数 μ_0 の値は、前期(2011年4月分)の標本平均とした。

例えば、 $\psi_1 = -2.0$ の場合、 $\mu_R = 3.29$ 及び $\mu_{NR} = 2.17$ すなわち $\mu_{NR} - \mu_R = -1.12$ となって、平均的に回答企業は無回答企業の10倍を超える受注額である。 $\psi_1 = 2.0$ の場合、 $\mu_R = 3.07$ 及び $\mu_{NR} = 4.08$ すなわち $\mu_{NR} - \mu_R = 1.01$ となって、平均的に回答企業は無回答企業の10分の1を下回る受注額である。表の第6列の値によって、回答企業と無回答企業の異質性について感覚的な評価が可能となる。すると、例えば「パラメータ ψ_1 の絶対値は高々2に過ぎない」すなわち「機械受注統計調査では異質性指標の絶対値は高々1.4に過ぎない」といった判断を下すことができる。

【表1】

「当該標本調査では異質性指標の絶対値は高々1.4に過ぎない」という先験的知見によれば、表10から、当該標本調査のデータに基づいて有意水準10%の統計的仮説検定を行う場合、誤検定確率を10%以下に抑えるためには欠測率は1.9%以下でなければならない、ということが分かる。

5. まとめ

本稿は、標本調査による母集団平均に関する両側検定について、欠測バイアスによって検定を誤る確率を定性的及び定量的に分析した。これは、標本調査に基づく統計的推論において欠測がもたらす問題の深刻さをいかに評価するかというより大きな課題の一部として位置付けることができる。

統計的仮説検定における欠測問題の深刻さの尺度として、「誤検定確率」を定義した。そのうえで、定性分析によって、誤検定確率と欠測率の単調関係を確認し、定量分析によって、欠測率、検定の有意水準、及び標本サイズに関する条件ごとに、感度分析を行った。

定性分析により、誤検定確率と欠測率の単調関係のほか、誤検定確率と有意水準、標本サイズ、回答者・無回答者間の異質性指標のそれぞれの間の単調関係も確認された。これらの関係は、図3から当然予想される結果である。定性分析の結果からは、誤検定確率の欠測率

及び異質性指標に対する単調増加関係で表される「バイアス効果」を、誤検定確率の有意水準に対する単調増加関係で表される「有意水準効果」、及び標本サイズに対する単調増加関係で表される「先鋭化効果」によって相殺する統計的処理の可能性が示された。しかしながら、定量分析の結果、そのような相殺処理は、通常の標本調査の条件に対しては、先鋭化効果が支配的であるため、実際的ではないことが明らかになった。また、標本サイズが非常に大きいデータに期待される検出力も、有意水準効果によって誤検定確率を抑制するに十分ではないことが分かった。今後ビッグデータを活用した統計的推論の普及が期待されるが、先鋭化効果には注意が必要である。

上記の分析の最大の問題は、回答者と無回答者の間の異質性の指標がみえないことである。本稿では、欠測データメカニズムをモデル化して、シミュレーションにより異質性指標の値の妥当な範囲を調べる方法を、内閣府「機械受注統計調査」のデータを用いて実演した。回答企業と無回答企業の平均受注額の比がさすがに10倍を超えるとは考えられないのであれば、異質性指標の絶対値は1.4を超えない。

文献

- 阿部貴行（2016）『欠測データの統計解析』朝倉書店
- 高井啓二・星野崇宏・野間久史（2016）『欠測データの統計科学』岩波書店
- 土屋隆裕（2009）『概説 標本調査法』朝倉書店
- 星野崇宏（2009）『観察データの統計科学—因果推論・選択バイアス・データ融合』岩波書店
- Gabaix, Xavier, 2011. “The Granular Origins of Aggregate Fluctuations.” *Econometrica*, 79(3), 733-772.
- Little, Roderick J. A., and Donald B. Rubin, 2002. *Statistical Analysis with Missing Data*. Second edition, John Wiley & Sons Ltd.
- Molenberghs, Geert, and Michael G. Kenward, 2007. *Missing Data in Clinical Studies*. John Wiley & Sons Ltd.
- Molenberghs, Geert, Garret Fitzmaurice, Michael G. Kenward, Anastasios Tsiatis, and Geert Verbeke (Editors), 2014. *Handbook of Missing Data Methodology*. Chapman & Hall CRC Press.

補論

第 3.1 節の比較静学

欠測率 π_{NR} 、回答者と無回答者の異質性の尺度 $|r|$ 、及び標本サイズ n については、「 $\mu_{NR} \leq \mu_0 \Leftrightarrow r \leq 0$ 」であることに注意すれば(4)式から直ちに導かれる。ここでは、検定の有意水準 α についての導出を示す。

(4)式から次式が成り立つ。

$$\frac{\partial \xi^*}{\partial \alpha} = \begin{cases} \frac{1}{2} \left\{ \frac{\varphi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \frac{\pi_{NR}\sqrt{nr}}{\sqrt{1 - \pi_{NR}}}\right)}{\varphi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)} - \frac{\varphi\left(\Phi^{-1}\left(\frac{\alpha}{2}\right) + \frac{\pi_{NR}\sqrt{nr}}{\sqrt{1 - \pi_{NR}}}\right)}{\varphi\left(\Phi^{-1}\left(\frac{\alpha}{2}\right)\right)} \right\} & (\mu_{NR} < \mu_0) \\ \frac{1}{2} \left\{ \frac{\varphi\left(\Phi^{-1}\left(\frac{\alpha}{2}\right) + \frac{\pi_{NR}\sqrt{nr}}{\sqrt{1 - \pi_{NR}}}\right)}{\varphi\left(\Phi^{-1}\left(\frac{\alpha}{2}\right)\right)} - \frac{\varphi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \frac{\pi_{NR}\sqrt{nr}}{\sqrt{1 - \pi_{NR}}}\right)}{\varphi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right)} \right\} & (\mu_{NR} \geq \mu_0) \end{cases}$$

ここで、

$$\varphi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right)\right) = \varphi\left(\Phi^{-1}\left(\frac{\alpha}{2}\right)\right)$$

であるから、上式右辺中括弧内の各項の分母は等しい。また、 $\mu_{NR} < \mu_0$ ならば、 $r < 0$ であるから、標準正規分布の確率密度関数の形状と臨界点の位置から、

$$\varphi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \frac{\pi_{NR}\sqrt{nr}}{\sqrt{1 - \pi_{NR}}}\right) > \varphi\left(\Phi^{-1}\left(\frac{\alpha}{2}\right) + \frac{\pi_{NR}\sqrt{nr}}{\sqrt{1 - \pi_{NR}}}\right)$$

が成り立つ。同様に、 $\mu_{NR} > \mu_0$ ならば、 $r > 0$ であるから、

$$\varphi\left(\Phi^{-1}\left(1 - \frac{\alpha}{2}\right) + \frac{\pi_{NR}\sqrt{nr}}{\sqrt{1 - \pi_{NR}}}\right) < \varphi\left(\Phi^{-1}\left(\frac{\alpha}{2}\right) + \frac{\pi_{NR}\sqrt{nr}}{\sqrt{1 - \pi_{NR}}}\right)$$

が成り立つ。従って、次が導かれる。

$$\mu_{NR} \neq \mu_0 \Rightarrow \frac{\partial \xi^*}{\partial \alpha} > 0$$

□

図表

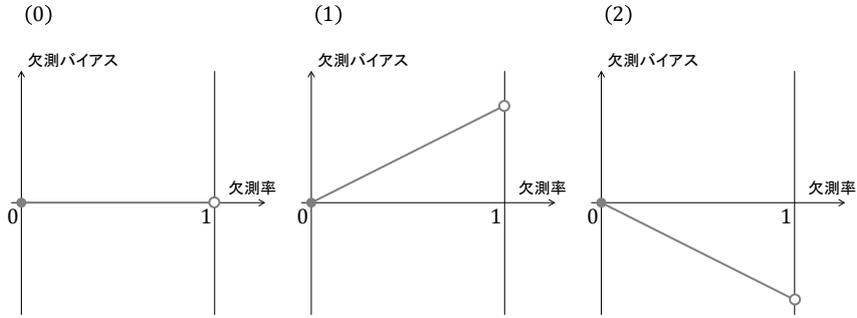


図1 標本平均における欠測バイアスと欠測率の関係：3通りの可能性

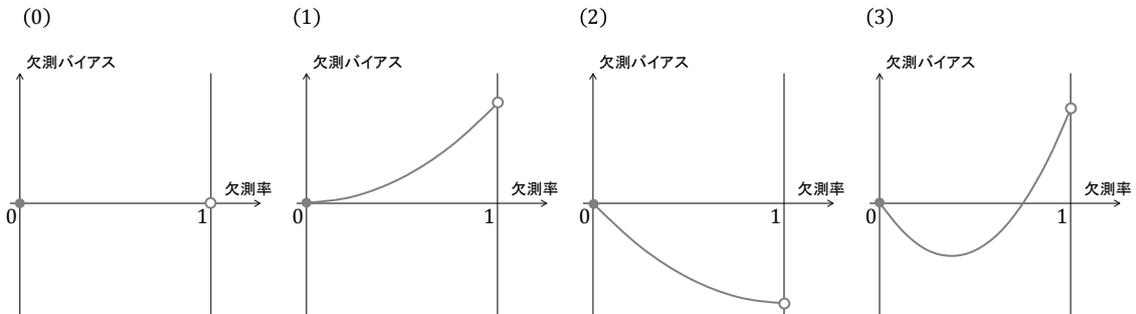


図2 標本分散における欠測バイアスと欠測率の関係：4通りの可能性

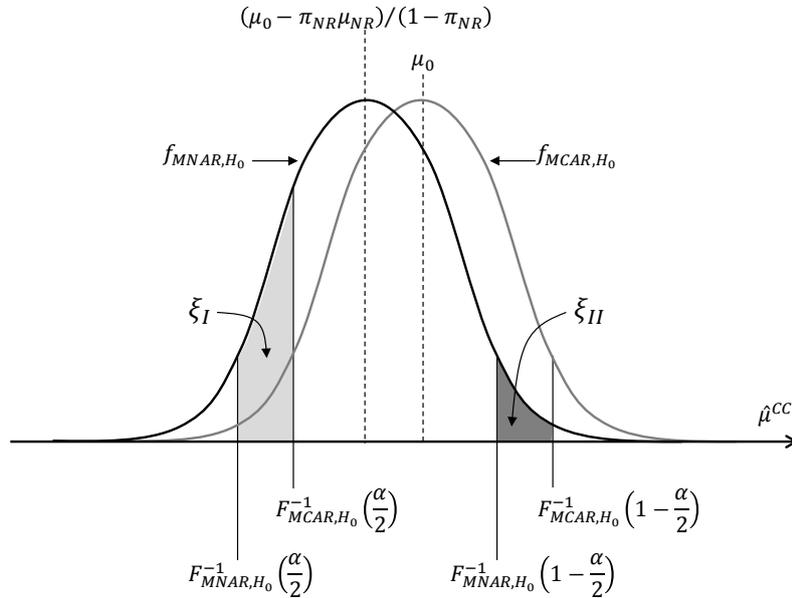


図3 母集団平均の検定において欠測により検定を誤る確率 ($\mu_{NR} \geq \mu_0$ の場合)

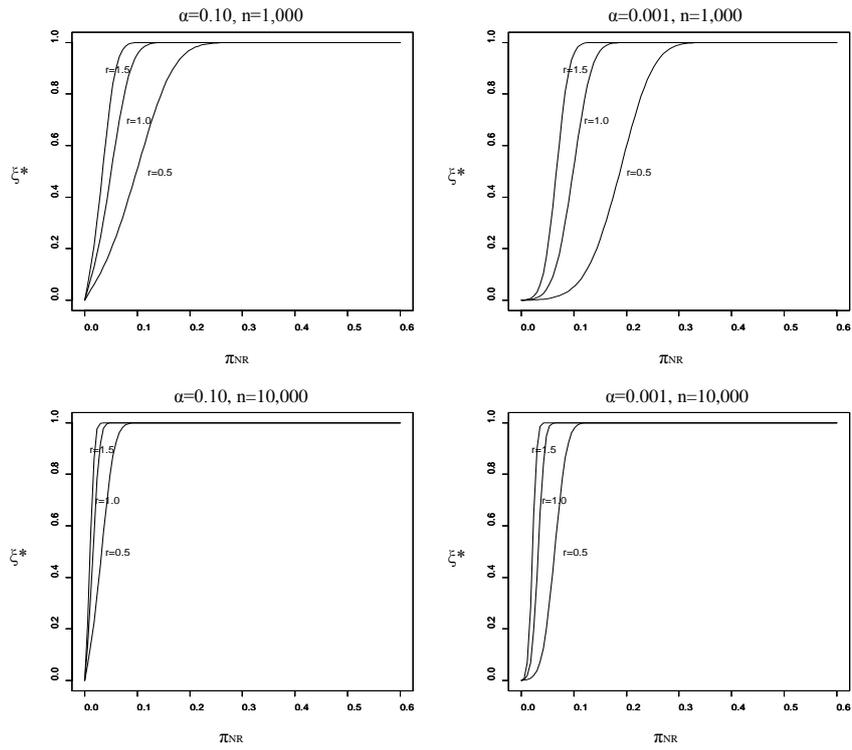


図4 欠測率と誤検定確率

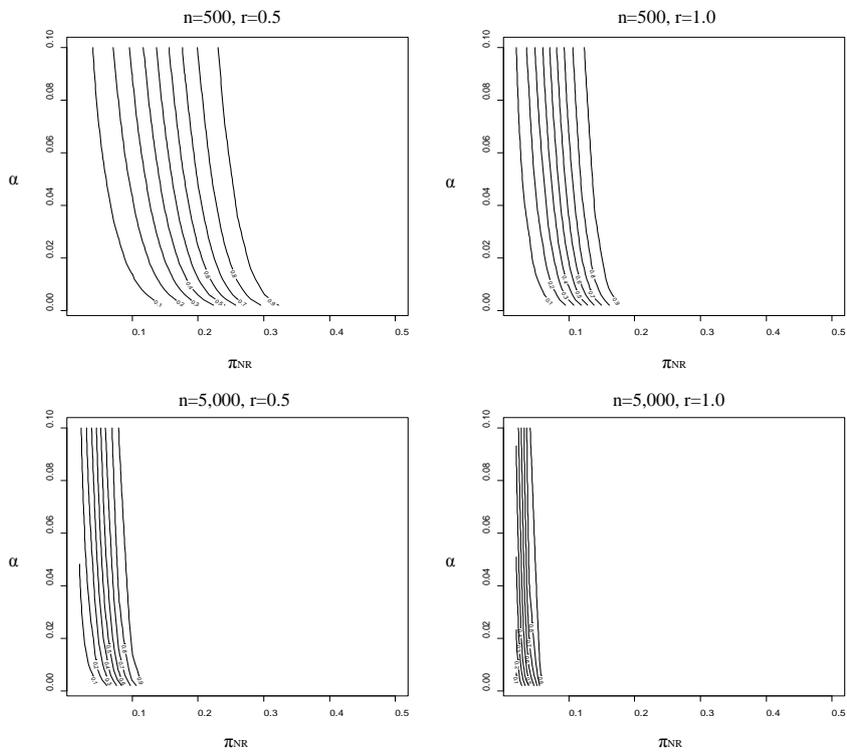


図5 π_{NR} - α 平面における誤検定確率の等量線

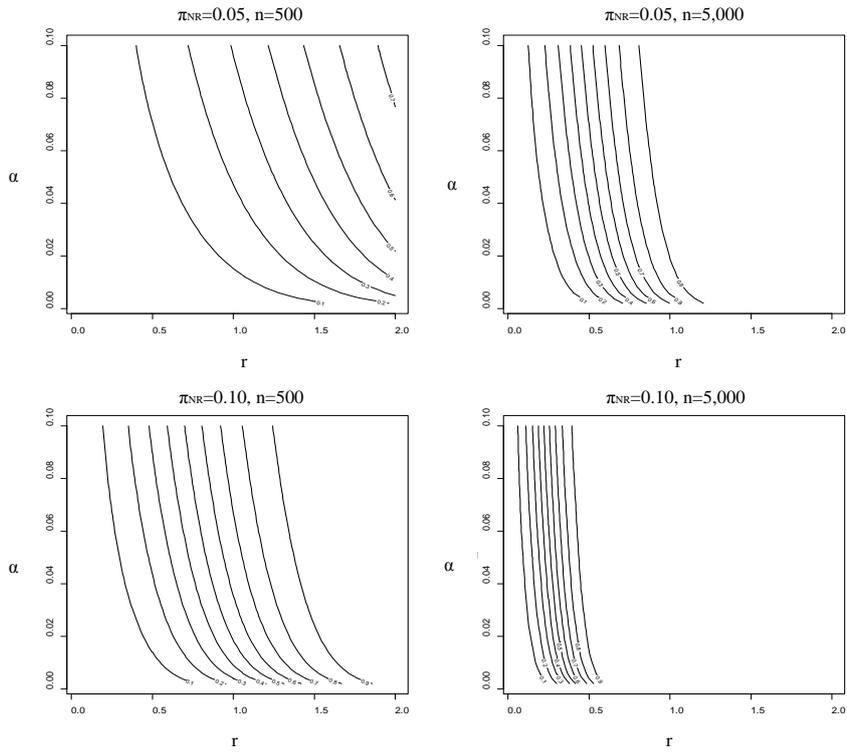


図6 $r - \alpha$ 平面における誤検定確率の等量線

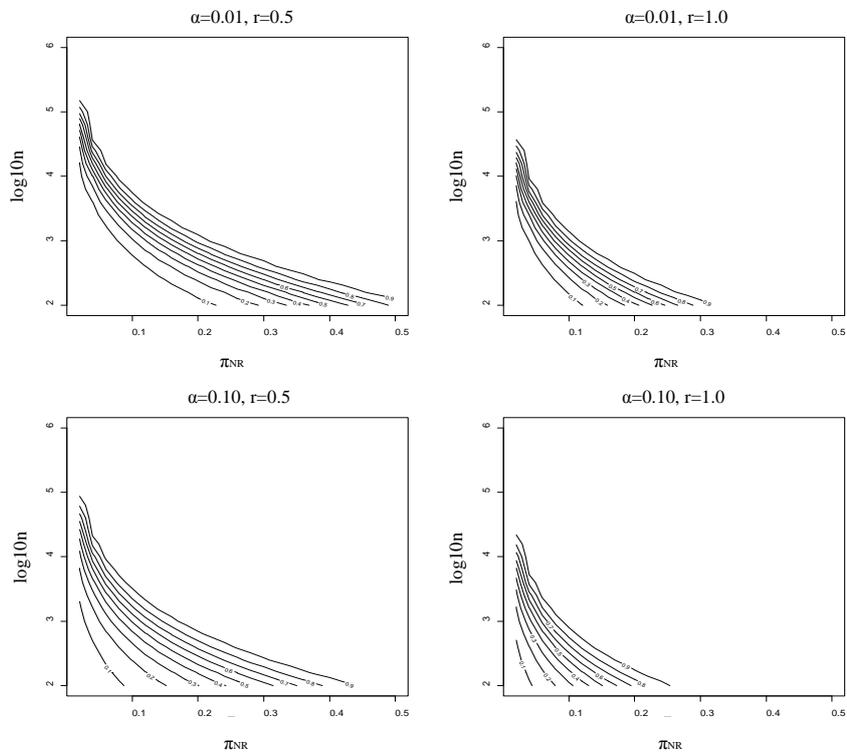


図7 $\pi_{NR} - \log_{10} n$ 平面における誤検定確率の等量線

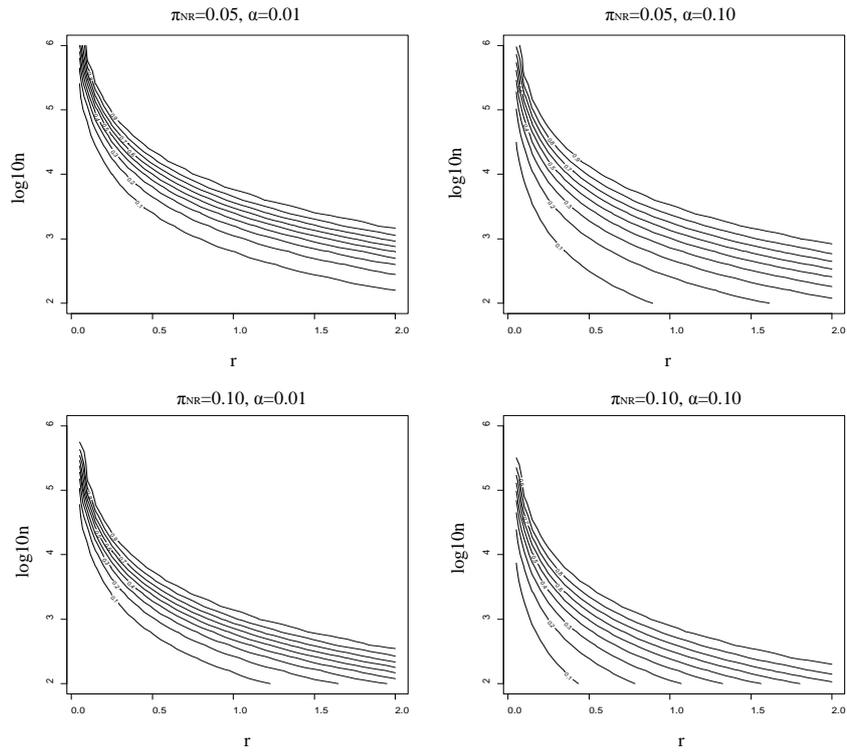


図8 $r - \log_{10} n$ 平面における誤検定確率の等量線

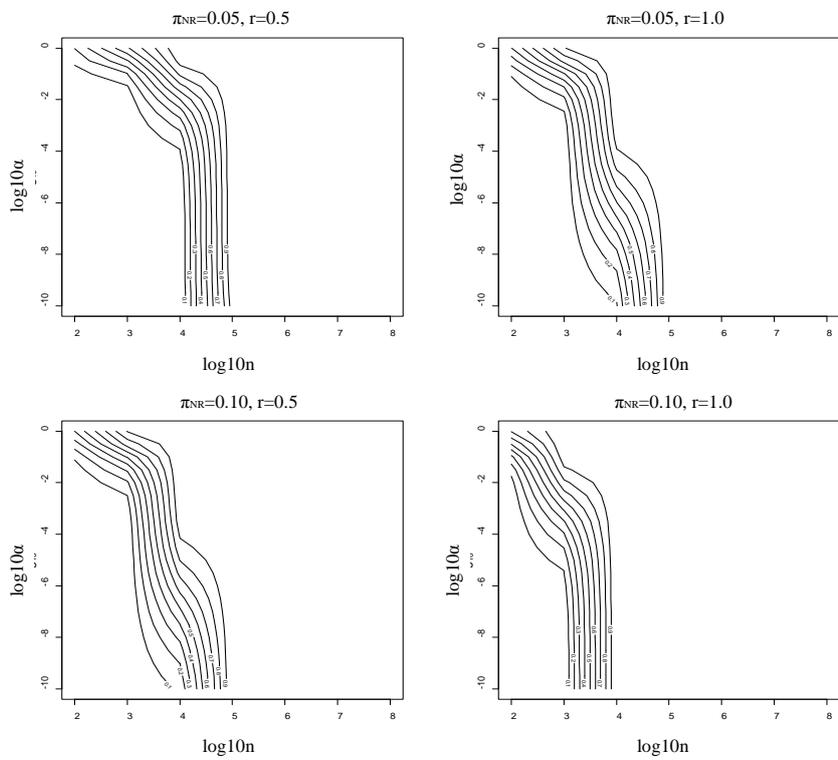


図9 $\log_{10} n - \log_{10} \alpha$ 平面における誤検定確率の等量線

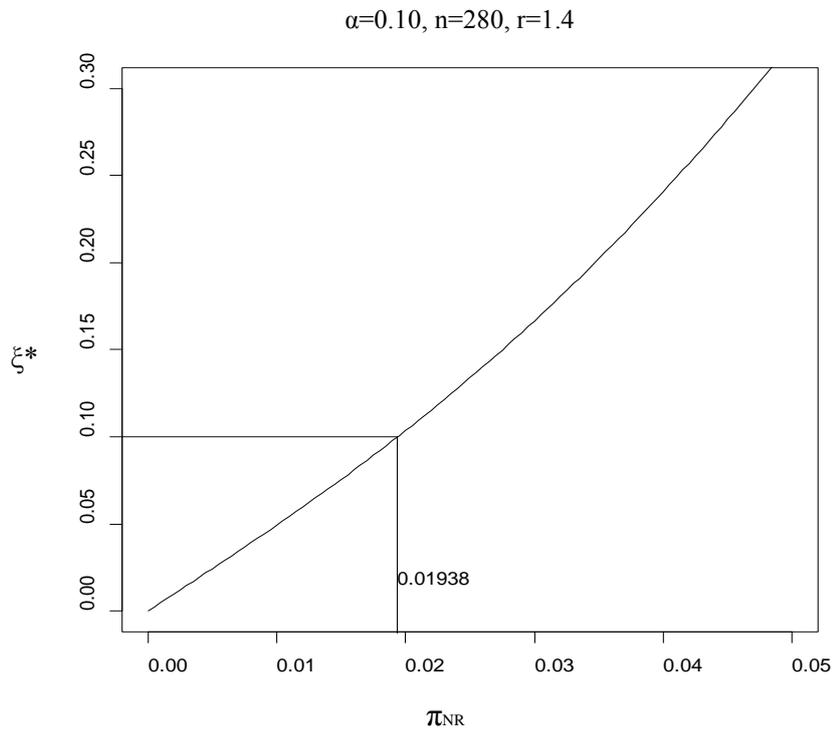


図 10 機械受注統計調査を想定した条件に対する誤検定確率曲線

表 1 機械受注統計調査に関するロジットモデルによるシミュレーション結果

ψ_1	μ_R	σ_R	μ_{NR}	σ_{NR}	$\mu_{NR} - \mu_R$	r
-3.0	3.31	0.71	1.97	0.75	-1.34	-1.73
-2.5	3.30	0.72	2.05	0.79	-1.25	-1.60
-2.0	3.29	0.74	2.17	0.84	-1.12	-1.39
-1.5	3.27	0.76	2.33	0.88	-0.94	-1.15
-1.0	3.24	0.78	2.56	0.90	-0.68	-0.82
-0.5	3.21	0.81	2.86	0.87	-0.35	-0.42
0.5	3.14	0.82	3.46	0.79	0.32	0.32
1.0	3.11	0.81	3.72	0.75	0.61	0.64
1.5	3.09	0.79	3.92	0.70	0.83	0.91
2.0	3.07	0.78	4.08	0.65	1.01	1.13
2.5	3.06	0.76	4.20	0.59	1.14	1.32
3.0	3.05	0.75	4.28	0.54	1.23	1.44