

RSS Higher Certificate in Statistics, 2010

Module 3: Basic Statistical Methods

1. 学年度のはじめに、ある学校で数学を勉強している最高学年の学生をグループ A (7人) とグループ B (11人) の 2つのクラスに分けた。それぞれのグループを受け持つ教師は異なっている。学生は全員同じ数学のテストを学年度の最後に受け、以下の表に示してあるような得点率をあげた。各グループの得点は昇順に並べている。

調査したい問題は 2つのグループ間において全体の得点レベルの差が有意かどうかということである。

Examination marks

	<i>Group A</i>	<i>Group B</i>
	12	36
	31	39
	34	46
	35	48
	55	58
	57	66
	61	69
		70
		77
		84
		85
Sample mean	40.71	61.64
Sample variance	312.90	303.85

- (i) 帰無仮説と対立仮説について注意して述べ、適切な t 検定を行い、結論を書け。
- (ii) t 検定を行う上での仮定を述べ、学生の得点がそれらを満たしているかどうか簡潔にコメントせよ。もし代わりにウィルコクソンの順位和検定を用いるのならば、その検定が妥当であるための仮定は何であるか述べよ。
- (iii) 得点データをウィルコクソンの順位和検定を用いて再分析せよ。同様に帰無仮説と対立仮説について述べ、結論を書け。

2. ある聖書の原書はタイプ A とタイプ B の 2 つの形式で作られた。エジプトのいくつかの遺跡からはこの聖書の複写が 10 点見つかり、そのうち 2 点がタイプ A のもの、8 点がタイプ B のものであった。これら 10 点の複写は、原書として作り出された聖書の数多くの複写の中から無作為に抽出されたものとみなしてよいとしたとき、原書として作り出されたものがタイプ A であった比率がタイプ B であった比率と異なるということが有意であるということ、出土された複写が示してくれるかを決めたい。

- (i) 写本のうちタイプ A の確率が θ であるという仮定の下で、無作為に 10 点抽出したときに含まれているタイプ A の写本の数の分布はどうなるか?
- (ii) (i) で述べられた分布のパラメーターによって検定される帰無仮説と対立仮説について明確に述べよ。
- (iii) 適切な統計表を用いて、データに対応する p -値を求め、結論を書け。

さらに遺跡を掘り進めた結果、全部で 100 点の聖書の複写が発見され、そのうち 39 点がタイプ A のもので、61 点がタイプ B のものであることが分かった。原書として作り出されたものがタイプ A のものである比率がタイプ B のものである比率と異なるということが有意であるかどうかを再び考えたい。

- (iv) 標本数 100 のより大きな標本において、タイプ A の写本が作り出された写本のうち θ であるとしたとき、標本に含まれるタイプ A の写本の数の分布の近似はどうなるか?
- (v) (ii) で記述された帰無仮説と対立仮説と適切な統計表を用いて、観測されたデータに対応する p -値を求め、結論を書け。

3. ある表計算ソフトの擬似乱数生成プログラムは数字の列を発生させるのに用いられる。もし乱数発生プログラムが適切なふるまいをしているのならば、その数字列の各値は 10 個の値 0, 1, ..., 9 のうちどれも同様に確からしくとりそうである。そのような乱数 100 個からなる数列を発生させた。そのときの 10 個の数字に対して観測された頻度の分布が以下の表に与えられている。

数字	0	1	2	3	4	5	6	7	8	9
頻度	10	7	8	12	10	13	7	4	15	14

- (i) 擬似乱数発生プログラムが適切にふるまう場合の各数字の期待頻度を計算し、そのように行った理由を述べよ。
- (ii) 観測された頻度が擬似乱数発生プログラムの想定された特性どおりであるかどうかを決定するための適切な検定統計量の値を計算せよ。

擬似乱数発生プログラムは標準正規偏差を発生するのにもまた用いることができる。正規分布の裾の部分にいかにあてはまっているか検定をするために、標準正規分布にしたがう 1000 個の乱数の標本をシミュレーションさせた。絶対値が 2 を超えたものは、これらのうち 54 個であることが分かった。

- (iii) 適切な統計表を用いて、標準正規分布にしたがう変量の絶対値が 2 を超える確率を求めよ。
- (iv) シミュレーションを行ったときに、値の絶対値が 2 を超えたと観測された数 54 は擬似乱数発生プログラムが適切なふるまいをしていないという有意な証拠を与えてくれるものかどうか、これを決定するための統計の検定を行え。

う

4.

- (a) 標本統計量の標本分布とは何を意味するか説明せよ。平均 μ 、分散 σ^2 の正規分布から無作為に n 個抽出したときの標本平均 \bar{X} と標本分散 S^2 の標本分布について述べよ。
- (b) ブランド A の小麦粉 100 袋の無作為標本の重さを量った。標本平均は 453.08 g で標本の標準偏差は 5.42 g であった。小麦粉の入っている袋の重さは未知の平均 μ と未知の分散 σ^2 の正規分布にしたがっているものとする。

(i) μ の 95% 信頼区間を求めよ。

(ii) σ^2 の 95% 信頼区間を求めよ。

(iii) ブランド B の袋に入っている小麦粉の重さはブランド A のものに比べて散らばりが大きいのではないかと思われる。ブランド B の小麦粉 20 袋の無作為標本の重さを量ると、標本の標準偏差は 11.15 g であった。

検定を行う上で基本となる分布理論について簡単に記述し、ブランド B の袋の重さの分散がブランド A のものよりも大きいということについて有意性があるかどうか決定するために検定を行え。