1. Здравствуйте, уважаемая комиссия! В прикладных исследованиях довольно часто возникает необходимость выяснить, имеют ли различия генеральные совокупности, из которых взяты две независимые выборки. В математической статистике данная задача формулируется как проверка гипотезы об однородности законов распределения вероятностей. Для решения данной задачи широко используются критерии однородности.  **(переключить)**
2. На практике нередко приходится иметь дело с данными ограниченной точности. Зачастую, это целые числа, или данные с одним, двумя знаками после запятой. При больших объемах выборок, количество повторений в выборках становится большим. Становится интересно, можно ли руководствоваться результатами исследований критериев однородности для таких выборок. Подчиняются ли статистики критериев предельным распределениям, и какие критерии являются более мощными на таких рода данных. Итак, целью моей работы было исследование критериев однородности по выборкам большого объема в случае ограниченной точности регистрации наблюдений. **(переключить)**
3. Итак, перейдем к критериям! На практике статистика критерия Лемана-Розенблатта используется в форме представленной на слайде. **(переключить)**

Данная статистика в пределе распределена как :



Вид предельной функции а1 представлен на слайде. **(переключить)**

1. Статистика критерия Смирнова измеряет расстояние между эмпирическими функциями распределения, построенными по выборкам



**(переключить)**

На практике значение статистики рекомен­дуется вычислять в соответствии с следующими соотношениями:

,

,

.

**(переключить)**

На практике используется статистика с поправкой Большова, с учетом которой распределение статистики в пределе подчиняется распределению Колмогорова с функцией распределения представленной на слайде. **(переключить)**

1. Для выборок непрерывных случайных величин выражение для статистики критерия Андерсона-Дарлинга принимает простой вид:



**(переключить)**

Предельным распределением данной статистики при справедливости проверяемой гипотезы  является распределение , представленное на экране:



. **(!переключить!)**

1. Исследования распределения статистик проводились с помощью метода компьютерного моделирования. Объем моделирования был задан 16600. Исследования проводились на данных ограниченных до 1,2 знаков и до целых чисел. Расстояние между эмпирической функцией распределения статистики и предельной функцией распределения критерия измерялось по метрике Колмогорова, формула для вычисления которого представлена на слайде. **(переключить)**
2. По данным представленным в таблице и в виде графика видно что, для критерия Лемана-Розенблатта распределения статистики остаются близкими к предельному закону при равных объемах выборок. **(переключить)**
3. Хотя для критерия Лемана-Розенблатта распределения статистики остаются близкими к предельному закону при равных объемах выборок, но при различных объемах выборок расстояние между эмпирической функцией распределения статистики и предельным увеличивается с ростом разницы между объемами выборок. **(переключить)**
4. При одинаковых размерностях обеих выборок для критерия Смирнова наблюдается медленная сходимость распределения статистики к предельному закону при увеличении объемов выборок. **(переключить)**
5. На данном слайде продемонстрированы результаты исследования распределения статистики для критерия Смирнова на выборках различных объемов. Не трудно заметить, что для выборок разного объема функция распределения статистики располагается несколько ближе к предельному. **(переключить)**
6. На данном слайде видно, что для критерия Андерсона-Дарлинга при одинаковых объемах выборок наблюдается увеличение расстояние между эмпирической функцией распределения и предельной функцией. **(переключить)**
7. С целью выяснить, зависит ли расстояние исключительно от объемов выборок или еще каких-то факторов, на следующем слайде представлены результаты исследований проведенных на всех типах округления. Изучив данные таблицы, можно сделать вывод, что для критерия Андерсона-Дарлинга расстояние между эмпирической функцией распределения статистики и предельным уменьшается с ростом отношения числа различных значений в объединенной выборке к объему объединенной выборки. Иначе говоря, близость функции распределения к предельной функции зависит от отношения числа различных значений в объединенной выборке к объему объединенной выборки. **(переключить)**
8. На данном слайде представлен сравнительный анализ результатов для случая одинаковых размеров выборок и различных для критерия Андерсона-Дарлинга. Из полученных результатов следует вывод, что для критерия Андерсона-Дарлинга функция распределения статистики располагается ближе к функции предельного распределения при выборках различного объема, чем при выборках одинакового объема. Также, были проведены дополнительные исследования, в ходе которых было выяснено, что чем разница в объеме выборок больше, тем расстояние до предельной функции меньше. (был зафиксирован объем для второй выборки(5000), менялся объем для первой(500, 1000, 2000) расстояние увеличивалось соответственно) **(переключить)**
9. Далее представлены результаты исследования мощности критериев. При всех конкурирующих гипотезах первая выборка всегда соответствовала стандартному нормальному закону, а вторая – некоторому другому. На слайде изображены виды распределений в альтернативах, относительно стандартного нормального распределения. **(переключить)**
10. На данных ограниченной точности наибольшую мощность среди рассмотренных критериев показали критерии Андерсона-Дарлинга и Лемана-Розенблатта. Однако в случае округления наблюдений в выборках до целых, первый график, критерий Андерсона-Дарлинга оказался смещенным относительно конкурирующих гипотез с пересечением функций распределения, то есть для гипотез H3, H4, H5. **(переключить)**
11. В заключении подведем итоги работы:

В результате исследования распределений статистик показано, что:

* для критерия Лемана-Розенблатта распределения статистики остаются близкими к предельному закону при равных объемах выборок, однако при  расстояние между эмпирической функцией распределения статистики и предельным увеличивается с ростом разницы в объемах выборок;
* для критерия Смирнова наблюдается довольно медленная сходимость распределения статистики к предельному закону при увеличении объемов выборок. Но при различных объемах выборок, расстояние между функцией распределения статистики и предельной функцией становится несколько меньше, чем в случае с одинаковыми объемами обеих выборок.
* для критерия Андерсона-Дарлинга расстояние между эмпирической функцией распределения статистики и предельным уменьшается с ростом отношения числа различных значений в объединенной выборке к объему объединенной выборки; А также, наблюдается сближение распределение статистики к предельному распределению при увеличении разности объемов выборок.

На данных ограниченной точности наибольшую мощность среди рассмотренных критериев показали критерии Андерсона-Дарлинга и Лемана-Розенблатта. Однако в случае округления наблюдений в выборках до целых критерий Андерсона-Дарлинга оказался смещенным относительно конкурирующих гипотез с пересечением функций распределения.

Обобщая полученные результаты, можно сделать вывод о предпочтительности использования критерия Лемана-Розенблатта при равных объемах выборок .

*На этом у меня все. Спасибо за внимание!*