Анализ данных по России

Данные, используемые для анализа, касаются доли людей в возрасте 70+ и приверженности к здоровому образу жизни (ЗОЖ) по регионам Российской Федерации. Анализ включает в себя вычисление коэффициента корреляции, построение графиков, а также оценку доверительных интервалов и сравнение средних значений для различных выборок.

# 1. Выводы из анализа данных

1.1 Средние значения выборок:  
Средние значения, полученные для генеральной выборки, случайной выборки и стратифицированной выборки, позволяют сделать несколько интересных выводов:  
 - Среднее значение для генеральной выборки составило 43.78. Это значение отражает общий тренд по России, что свидетельствует о том, что в среднем около 43-44% людей в возрасте 70+ привержены к здоровому образу жизни (ЗОЖ).  
 - Среднее значение случайной выборки составило 44.91, что немного выше среднего генеральной выборки. Это может означать, что в случайной выборке оказались регионы, где приверженность ЗОЖ выше.  
 - Среднее значение стратифицированной выборки составило 43.29, что чуть ниже среднего генеральной выборки. Это также подтверждает, что стратифицированный метод дает близкие результаты к генеральной выборке, но возможно с меньшими отклонениями.  
  
1.2 Доверительные интервалы:  
Для каждой из выборок были рассчитаны доверительные интервалы на уровнях доверия 90%, 95% и 99%. Доверительный интервал показывает, в каком диапазоне с определенной вероятностью находится среднее значение выборки. Чем уже интервал, тем точнее оценка.  
 - Для случайной выборки: 90% интервал составляет (43.06, 46.75), 95% интервал — (42.71, 47.11), 99% — (42.01, 47.81).  
 - Для стратифицированной выборки: 90% интервал составляет (41.45, 45.13), 95% интервал — (41.10, 45.48), 99% — (40.40, 46.18).  
Эти интервалы показывают, что стратифицированная выборка имеет более узкие интервалы, что может свидетельствовать о большей точности оценки среднего значения по сравнению с случайной выборкой.  
  
1.3 Сравнение средних значений выборок с генеральной выборкой:  
Средние значения случайной и стратифицированной выборок близки к среднему генеральной выборки, что говорит о том, что обе выборки хорошо отражают общее распределение данных по России. Сравнение средних значений показывает, что обе выборки имеют примерно одинаковое значение, что подтверждает их репрезентативность.  
  
1.4 Корреляция между переменными:  
Коэффициент корреляции между долей людей в возрасте 70+ и приверженностью к ЗОЖ составил 0.97, что говорит о сильной положительной связи между этими переменными. Это подтверждает, что с увеличением доли людей в возрасте 70+ в регионе, также увеличивается приверженность к здоровому образу жизни. Этот результат важен для планирования социальных и медицинских инициатив, направленных на улучшение здоровья пожилых людей в России.  
  
1.5 Коэффициент детерминации (R²):  
Коэффициент детерминации R² составляет 0.95, что означает, что 95% вариации данных можно объяснить с помощью линейной модели. Это высокий показатель, свидетельствующий о том, что линейная модель адекватно описывает связь между долей людей в возрасте 70+ и приверженностью к ЗОЖ.  
  
1.6 Заключение:  
В целом, результаты анализа показывают, что данные по регионам России имеют значимую положительную связь между возрастной группой 70+ и приверженностью к ЗОЖ. Выборки, как случайная, так и стратифицированная, хорошо отражают генеральную совокупность данных, что подтверждается близкими средними значениями и узкими доверительными интервалами. Для дальнейшего исследования рекомендуется использовать стратифицированный подход, чтобы обеспечить точность данных для каждого федерального округа.