

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ В ПАКЕТАХ ПРИКЛАДНЫХ ПРОГРАММ

Проверка статистических гипотез является одним из основных инструментов в научных исследованиях и бизнес-анализе. Она позволяет сделать выводы о генеральной совокупности на основе выборочных данных и определить, насколько вероятно, что наблюдаемые различия между выборками объясняются случайностью или они являются статистически значимыми. Настоящее исследование заключается в проверке статистических гипотез о значимости различий в размере заработной платы разных групп населения. Расчеты проведены по данным выборочного обследования членов домашних хозяйств Республики Беларусь по уровню жизни (15 000 наблюдений) в таких программах статистического анализа, как Excel, Statistica и SPSS.

При помощи пакета Statistica был проведен анализ распределения генеральной совокупности членов домашних хозяйств РБ и получены следующие результаты: гистограмма и критерий Колмогорова — Смирнова показали, что распределение не является нормальным ($p\text{-value} < 0.01$). Критерий Лиллиефорса также показал ненормальность распределения ($p\text{-value} < 0.01$). Из этих результатов можно сделать вывод, что генеральная совокупность членов домашних хозяйств РБ не имеет нормального распределения. Это важно учитывать при дальнейшем анализе и интерпретации результатов.

Далее проведено обследование, в ходе которого была выявлена статистически значимая разница в средней заработной плате между мужчинами и женщинами. Значение t -статистики составляет 19,46, т.е. является очень высоким. Эти результаты указывают на то, что различия в заработной плате между мужчинами и женщинами статистически значимы и не могут быть случайными. Средняя заработная плата у мужчин в выборке выше (723,85 руб.), чем у женщин (536,89 руб.).

Следующее исследование посвящено проверке гипотезы о том, что средняя заработная плата значительно превышает 500 руб. Расчеты проведены в программе Excel. Была получена выборка из 7265 наблюдений, среднее значение зарплаты составило 623,61 руб. Значение t -статистики равно 0,74, что меньше критического одностороннего значения t -статистики (1,64) при уровне значимости 0.05. Исходя из результатов, нельзя принять гипотезу о том, что заработная плата значительно превышает 500 рублей. То есть мы не можем отвергнуть нулевую гипотезу.

В программе SPSS было исследовано влияние места жительства (области) на уровень заработной платы. По результатам ANOVA значение F -статистики составляет 72,983, а p -значение (вероятность получить подобное или еще более экстремальное различие между группами случайно) равно 0,000. Это позволя-

ет сделать вывод о статистически значимой разнице в уровне заработной платы в зависимости от места жительства. Это значит, что наша статистическая гипотеза о том, что зарплата зависит от области проживания, подтверждается данными исследования. Различия в заработной плате между областями могут быть связаны с разными экономическими условиями, уровнем безработицы и т.д.

Нетрудно заметить, что все эти пакеты прикладных программ предоставляют широкий спектр инструментов и методов для проверки статистических гипотез. Однако, прежде чем проводить тесты, важно правильно сформулировать гипотезы и выбрать подходящий тест для вашего исследования. Кроме того, нельзя забывать о том, что статистические тесты не могут дать абсолютно точных результатов. Всегда нужно оценивать результаты тестов в контексте проводимого исследования и учитывать возможность ошибок первого и второго рода.

Источники

- Общая теория статистики : учебник / под ред. П. Р. Куликова. — М., 2002.
Джессен, Л. Статистические методы / Л. Джессен. — СПб., 2001.