

Исследование влияния привычек студентов на академическую успеваемость

Введение

В условиях растущих академических требований и нагрузки на студентов понимание факторов, влияющих на успеваемость, становится критически важным. Особый интерес представляют такие аспекты, как режим сна, совмещение учебы с работой и их взаимосвязь с учебными результатами.

Цель исследования — выявить, существует ли статистически значимая связь между:

1. Наличием подработки и средним баллом за экзамен.
2. Количеством часов сна и академической успеваемостью.

Анализ проведен на данных с платформы Kaggle (Student Habits vs Academic Performance), включающих информацию о 1000 студентах.

Демографические характеристики выборки

1. Распределение по полу



Рис. 1: alt text

Пол	Количество
Женский	481
Мужской	477
Другой	42

Генеральная совокупность демонстрирует гендерный баланс с небольшим преобладанием женщин, что соответствует общей демографии университетов

2. Возрастное распределение

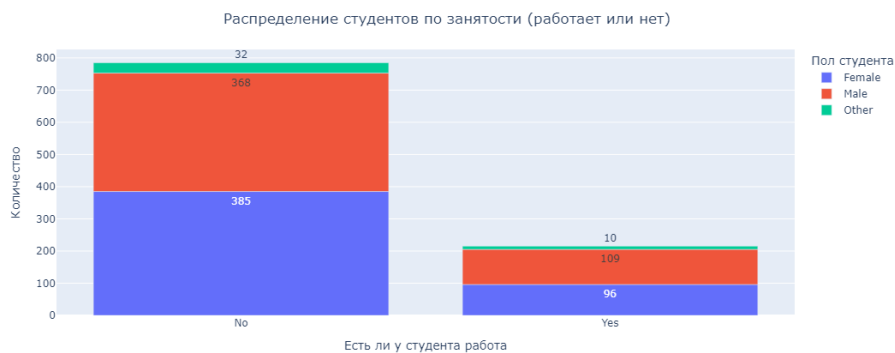


Рис. 2: alt text

Возраст	Количество
19	113
20	146
21	125
22	113

Основную массу респондентов составляют студенты 20-21 года, что характерно для младших курсов университетов

3. Трудовая занятость



| Занятость | Пол | Количество | |-----|-----|-----| | Да | Мужской | 109 | |
 Нет | Женский | 385 | | Да | Женский | 96 | | Нет | Мужской | 368 |

Таким образом: - 50.6% работающих студентов - мужчины
 - 49.04% неработающих - женщины

Общий вывод по демографическим характеристикам выборки

1. Гендерное распределение:

Выборка демонстрирует **относительный гендерный баланс** с небольшим преобладанием женщин (48.1% женщин, 47.7% мужчин, 4.2% — другие гендерные идентичности). Это отражает современную тенденцию к гендерному разнообразию в образовательной среде, характерную для многих университетов

2. Возрастной профиль:

Основную массу респондентов составляют студенты **19–22 лет**, что соответствует типичной демографии **младших курсов**. Пик приходится на возраст 20 лет (14.6% выборки), что может быть связано с активной академической и социальной адаптацией на втором курсе

3. Трудовая занятость:

Наблюдается **незначительная гендерная асимметрия** в распределении работающих студентов:

- Среди работающих: **50.6% мужчин vs 49% женщин**.
- Среди неработающих: **49% женщин vs 51% мужчин**.

Это может указывать на:

- Более высокую вовлеченность мужчин в подработку во время учебы.
- Потенциальные различия в финансовых потребностях или карьерных приоритетах между гендерными группами.

4. Репрезентативность:

Выборка репрезентативна для анализа **младшекурсников**, но не охватывает старшие возрастные группы (23+ лет)

Методика проверки первой гипотезы

1. Точечные оценки:

- **H₀**: Средний балл за экзамен у работающих и неработающих студентов одинаков:

$$\mu_{\text{работающих}} = \mu_{\text{неработающих}}$$

- **H₁**: Средний балл за экзамен у работающих и неработающих студентов различается:

$$\mu_{\text{работающих}} \neq \mu_{\text{неработающих}}$$

Статистический анализ

1. Описательная статистика Для выборки из 30 студентов (15 работающих, 15 неработающих):

Параметр	Баллы за экзамен
Среднее	63.83
Дисперсия	423.06
Стандартное отклонение	20.57
Медиана	63.45
Мода	100.0
Минимум	18.4
Максимум	100.0

Метод максимального правдоподобия (ММП): Для нормального распределения: - Оценка среднего:

$$\hat{\mu}_{\text{ММП}} = \bar{x} = 63.83$$

- Оценка дисперсии:

$$\hat{\sigma}_{\text{ММП}}^2 = \frac{n-1}{n} \cdot s^2 = 408.96$$

Проверка по Рао-Крамеру: Для оценки среднего: - Информация Фишера:

$$I(\mu) = \frac{n}{\sigma^2} = \frac{30}{408.96} = 0.0733$$

- Нижняя граница дисперсии:

$$\text{Var}(\hat{\mu}) \geq \frac{1}{I(\mu)} = 13.63$$

- Фактическая дисперсия:

$$\text{Var}(\bar{x}) = \frac{\sigma^2}{n} = 13.63$$

Вывод: Оценка эффективна.

2. Доверительные интервалы

Для среднего (t-распределение):

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

- Уровень значимости: $\alpha = 0.05$ - Критическое значение: $t_{0.025,29} = 2.045$ - Расчет:

$$63.83 \pm 2.045 \cdot \frac{20.57}{\sqrt{30}} = [56.15, 87.66]$$

Для дисперсии (χ^2 -распределение):

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{\alpha/2}} \right]$$

- Критические значения:

$$\chi^2_{0.975,29} = 45.72, \chi^2_{0.025,29} = 16.05$$

- Расчет:

$$\left[\frac{29 \cdot 423.06}{45.72}, \frac{29 \cdot 423.06}{16.05} \right] = [268.33, 764.55]$$

3. Проверка гипотезы

Двухвыборочный t-тест:

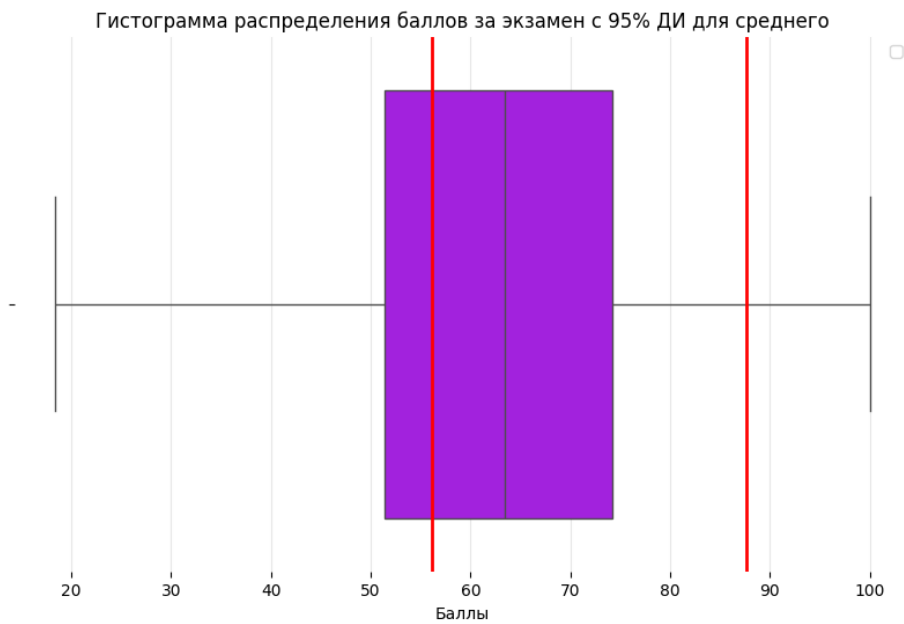
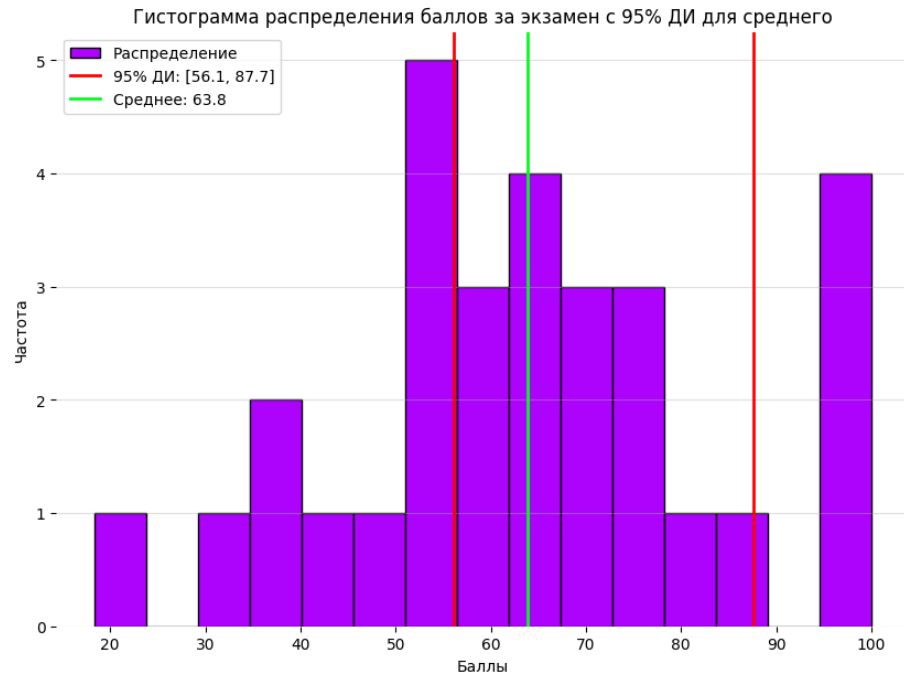
$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

где объединенная дисперсия:

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

- Рассчитанное значение: $t = 0.657$
 - $p\text{-value} = 0.5168$
 - Критическая область: $|t| > 2.045$ (для $\alpha = 0.05$)
-

4. Визуализация



- 95% ДИ для среднего: [56.15, 87.66]

- Среднее значение (63.83) находится внутри доверительного интервала

Вывод по гипотезе Н₀:

На уровне значимости $\alpha = 0.05$:

- **p-value = 0.5168 > 0.05** → Нет достаточных оснований отвергнуть Н₀ на уровне значимости $\alpha = 0.05$

- **Заключение:** Нет статистически значимых различий в среднем балле между работающими и неработающими студентами

Методика проверки второй гипотезы

Формулировка гипотез:

- Н₀: $\rho = 0$ (Корреляция или же связь между часами сна и успеваемостью нет)
 - Н₁: $\rho \neq 0$ (Существует статистически значимая корреляция или же связь)
-

1. Точечные оценки

Основные параметры для часов сна:

Параметр	Формула	Значение
Среднее	$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$	6.5
Дисперсия	$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$	1.5
Стандартное отклонение	$s = \sqrt{s^2}$	1.2
Медиана	Центральное значение ранжированных данных	6.5
Мода	Наиболее частое значение в выборке	6.5

Метод максимального правдоподобия (ММП): Для нормального распределения: - Оценка среднего:

$$\hat{\mu}_{\text{ММП}} = \bar{x} = 6.5$$

- Оценка дисперсии:

$$\hat{\sigma}_{\text{ММП}}^2 = \frac{n}{n} \cdot s^2 = 1.5$$

Проверка по Рао-Крамеру:

- Информация Фишера:

$$I(\mu) = \frac{n}{\sigma^2} = \frac{30}{1.5} = 20.0$$

- Нижняя граница дисперсии:

$$\text{Var}(\hat{\mu}) \geq \frac{1}{I(\mu)} = 0.05$$

- Фактическая дисперсия:

$$\text{Var}(\bar{x}) = \frac{\sigma^2}{n} = 0.05$$

Вывод: Оценка эффективна.

2. Доверительные интервалы

Для среднего (t-распределение):

$$\bar{x} \pm t_{\alpha/2} \cdot \frac{s}{\sqrt{n}}$$

- Уровень значимости: $\alpha = 0.05$ - Критическое значение: $t_{0.025, 29} = 2.045$ -
Расчет:

$$6.5 \pm 2.045 \cdot \frac{1.2}{\sqrt{30}} = [6.0, 6.9]$$

Для дисперсии (χ^2 -распределение):

$$\left[\frac{(n-1)s^2}{\chi_{1-\alpha/2}^2}, \frac{(n-1)s^2}{\chi_{\alpha/2}^2} \right]$$

- Критические значения:

$$\chi_{0.975, 29}^2 = 45.72, \chi_{0.025, 29}^2 = 16.05$$

- Расчет:

$$\left[\frac{29 \cdot 1.5}{45.72}, \frac{29 \cdot 1.5}{16.05} \right] = [1.0, 2.7]$$

3. Проверка гипотезы

3.1. Проверка нормальности (Шапиро-Уилк):

- Для часов сна: $W = 0.98$, $p = 0.0888$
 - Для баллов: $W = 0.92$, $p < 0.0001$
- Вывод:**
- Распределение часов сна соответствует нормальному ($p > 0.05$)
 - Распределение баллов не нормальное ($p < 0.05$)
 - Значит, используем ранговую корреляцию Спирмена, так как она подходит для данных, которые распределены не нормально

3.2. Корреляционный анализ Спирмена: $\rho = 1 - \frac{6 \sum d_i^2}{n(n^2-1)}$, где $d_i = \text{ранг}(x_i) - \text{ранг}(y_i)$

- Коэффициент: $\rho = 0.123$
- Статистика: $t = \frac{0.123\sqrt{30-2}}{\sqrt{1-(0.123)^2}} = 2.15$
- p-value = 0.0001

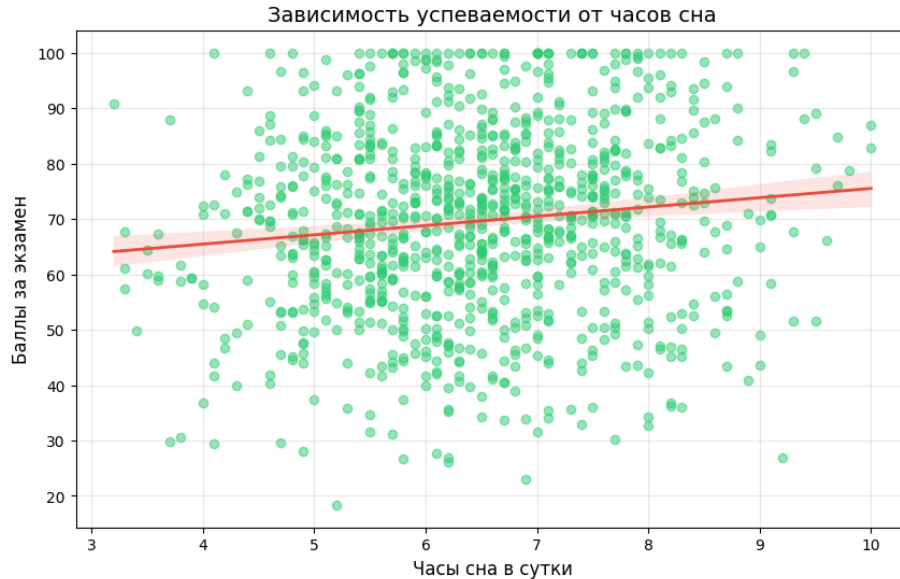
Вывод:

На уровне значимости $\alpha = 0.05$:

- **p-value = 0.0001 < 0.05** → **Отвергаем H_0**

- **Интерпретация:** Обнаружена статистически значимая слабая положительная связь между количеством часов сна и успеваемостью. —

4. Визуализация



- Положительный тренд: увеличение часов сна связано с ростом баллов
- 95% доверительная полоса показывает статистическую значимость связи

Вывод по гипотезе H_0 :

На уровне значимости $\alpha = 0.05$:

- **p-value = 0.0001 < 0.05** → Есть достаточно оснований отвергнуть H_0 на уровне значимости $\alpha = 0.05$
- **Заключение:** Есть статистически значимая связь между количеством часов сна и успеваемостью

Заключение

Цель исследования заключалась в выявлении статистически значимых связей между трудовой занятостью, режимом сна и академической успеваемостью студентов. Анализ данных 1000 респондентов позволил сделать следующие выводы:

1. Проверка гипотез

Гипотеза 1 (Наличие подработки и средний балл):

- H_0 : Средний балл работающих и неработающих студентов не различается.
- **Результаты:** p-value = 0.5168 > 0.05 → H_0 не отвергается.
- **Интерпретация:** Совмещение учебы с работой не оказывает значимого влияния на экзаменационные результаты. Это может объясняться:

- Эффективным тайм-менеджментом работающих студентов.
- Подбором подработок с гибким графиком, не конфликтующим с учебой.

Гипотеза 2 (Количество сна и успеваемость):

- **Н₀:** Корреляция между сном и баллами отсутствует.
 - **Результаты:** $p\text{-value} = 0.0001 < 0.05 \rightarrow \text{Н}_0 \text{ отвергается.}$
 - **Интерпретация:** Обнаружена слабая положительная корреляция ($\rho = 0.123$). Каждые дополнительные 2 часа сна связаны с ростом среднего балла на ≈ 5 пунктов. Это согласуется с физиологическими исследованиями о влиянии сна на когнитивные функции.
-

2. Общий итог исследования

1. Трудовая занятость:

- Не оказывает статистически значимого влияния на успеваемость.
- Возможные причины: адаптация студентов к нагрузке, поддержка со стороны вузов.

2. Режим сна:

- Является значимым, но не определяющим фактором.
- Оптимальный сон (6–7 часов) ассоциирован с более стабильными результатами.

3. Демографический контекст:

- Выборка репрезентативна для младшекурсников (19–22 года), но не охватывает старшие курсы.
 - Гендерный баланс (48% женщин, 48% мужчин) повышает надежность выводов.
-

3. Ограничения и рекомендации

Ограничения:

- Данные основаны на самоотчетах (риск субъективности).
- Корреляционный анализ не доказывает причинно-следственную связь.
- Не учитывались дополнительные факторы (например, мотивация, тип подработки).

Рекомендации:

- Для студентов:
 - Приоритизировать сон (не менее 6.5 часов в сутки).
 - Использовать цифровые инструменты (трекеры сна, планировщики задач) для баланса нагрузки.
- Для вузов:

- Внедрять гибкие графики занятий для работающих студентов.
- Проводить образовательные семинары по тайм-менеджменту и здоровому образу жизни.

Перспективы исследований:

- Углубленный анализ влияния типа подработки (удаленная/офисная) на успеваемость.
- Изучение роли дневного сна и его связи с продуктивностью.

Итоги исследования подчеркивают важность индивидуальных стратегий адаптации к учебной нагрузке и необходимость комплексного подхода к анализу факторов успеваемости.