

Воронеж 2024

Для данной лабораторной работы преподавателем была предоставлена таблица данных в формате электронной таблицы Excel. В данной таблице представлена информация о полетах за последний год. В таблице имеется 6 столбцов: кол-во полётов, пол пассажира, цель полёта, самая важная характеристика, годовой доход, возраст. Также представлены данные 201 пассажира.

Задание 1. Какие переменные в исследовании пассажиров являются количественными, а какие неколичественными – категориальными?

Количественные переменные:

- количество полетов за последний год;
- годовой доход;
- возраст.

Категориальные переменные:

- пол;
- цель полёта;
- самая важная характеристика.

Далее если мы хотим применять данные для обучения моделей, нужно произвести нормализацию данных. Поменять значение «пол» на бинарное, например, мужской = 1, женский = 0. Поменять значение «цель полёта» на бинарное, например, бизнес = 1, а личные цели = 0. Также поменять значение «самая важная характеристика», например, частота рейсов = 0, комфорт = 1, точность = 2.

Задание 2. Постройте три сводные таблицы.

1) Пол и цель полёта.

Таблица 1. Пол и цель полёта

Количество по полю цель полета	Названия столбцов		
Названия строк	бизнес	личные цели	Общий итог
женщина	30	20	50
мужчина	98	52	150
Общий итог	128	72	200

2) Пол и самая важная характеристика

Таблица 2. Пол и самая важная характеристика

Количество по полю самая важная характеристика	Названия столбцов		частота рейсов	Общий итог
	комфорт	точность		
Названия строк				
женщина	30	10	10	50
мужчина	46	36	68	150
Общий итог	76	46	78	200

3) Цель полёта и самая важная характеристика

Таблица 3. Цель полёта самая важная характеристика

Количество по полю самая важная характеристика	Названия столбцов		частота рейсов	Общий итог
	комфорт	точность		
Названия строк				
бизнес	48	26	54	128
личные цели	28	20	24	72
Общий итог	76	46	78	200

Из первой сводной таблицы видно:

- мужчин летало в 3 раза больше чем женщин;
- всего было 200 полётов;
- женщин, летавших для личных целей 20 (40%);
- женщин, летавших для бизнес целей 30 (60%);
- мужчин, летавших для личных целей = 52 (34,6%);
- мужчин, летавших для бизнес целей 98 (65,4%);
- женщины летали по личным целям на 5,4% больше, чем мужчины;
- женщины летали по бизнес делам меньше, чем мужчины на 5,4%.

Из второй сводной таблицы видно:

- для женщин комфорт на первом месте (60%);
- для женщин точность и частота на втором месте (по 20%);
- для мужчин на первом месте частота (45,4%);

- на втором месте для мужчин комфорт (30,6%);
- на третьем месте для мужчин точность (24%).

Из третьей таблицы видно:

- бизнес полётов было совершено 128 (64%);
- полётов в личных целях 72 (36%);
- для бизнес полетов на первом месте частота (54%);
- на втором месте для бизнеса комфорт (37,5);
- на третьем для бизнеса точность (20,3%);
- для личных целей на первом месте комфорт (38,8%);
- для личных целей на втором месте частота рейсов (33,3%);
- для личных целей на третьем месте точность (27,9).

Задание 3. Постройте две частотные таблицы: для распределения пассажиров по возрасту и по годовому доходу.

Таблица 4. Количество пассажиров по возрасту

возраст		
левая граница интервала	правая граница интервала	
11	15	0
16	20	2
21	25	14
26	30	32
31	35	32
36	40	22
41	45	24
46	50	22
51	55	18
56	60	20
61	65	8
66	70	2
71	75	4
всего		

Таблица 5. Количество пассажиров по доходу

годовой доход		
левая граница интервала	правая граница интервала	частота
1	10	0
11	20	0
21	30	0
31	40	4
41	50	0
51	60	34
61	70	18
71	80	2
81	90	24
91	100	12
101	110	32
111	120	4
всего		

Также для наглядности были построены две гистограммы, они продемонстрированы на рисунках 1-2.

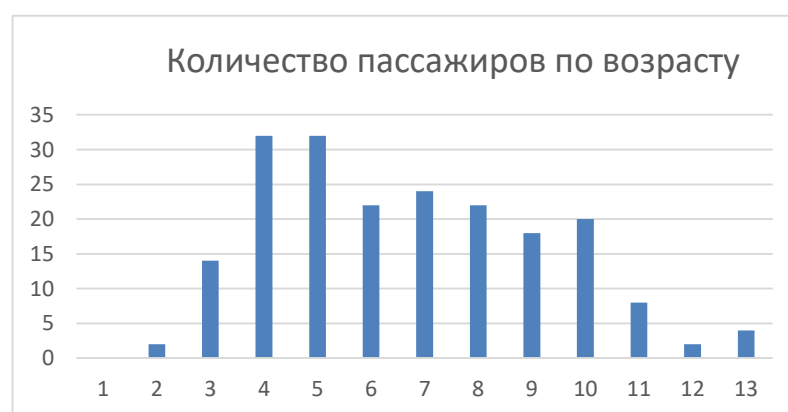


Рисунок 1 – Гистограмма кол-во пассажиров по возрасту

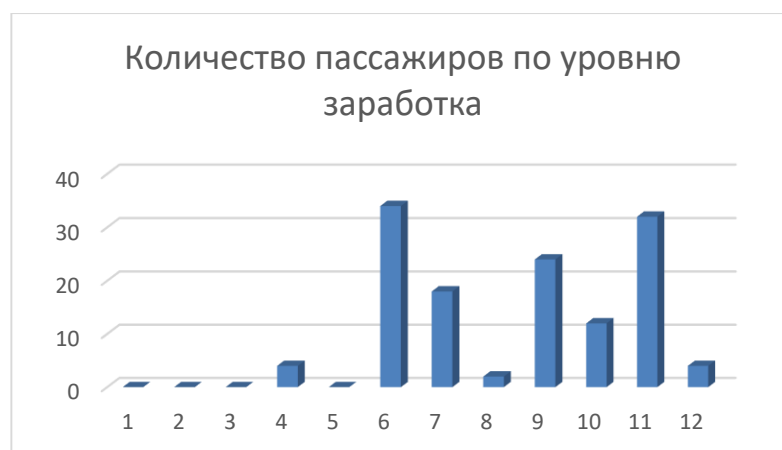


Рисунок 2 – гистограмма кол-ва пассажиров по уровню заработка

Из первой гистограммы видно, что больше всего пассажиров летало в возрасте от 26 до 35 лет. Также наглядно видно, что наибольший средний достаток от 51 до 61 тысячи рублей и от 101 до 110 тысячи рублей.

Задание 4. Представьте данные таблиц из п. 2 в графической форме. Выберите уместный тип диаграммы и поясните ее особенности. В таблице п. 3 добавьте столбики «средняя точка интервала» и «относительная частота» и постройте гистограммы.

Строю гистограмму из первой таблицы пункта 2 (рисунок 3).

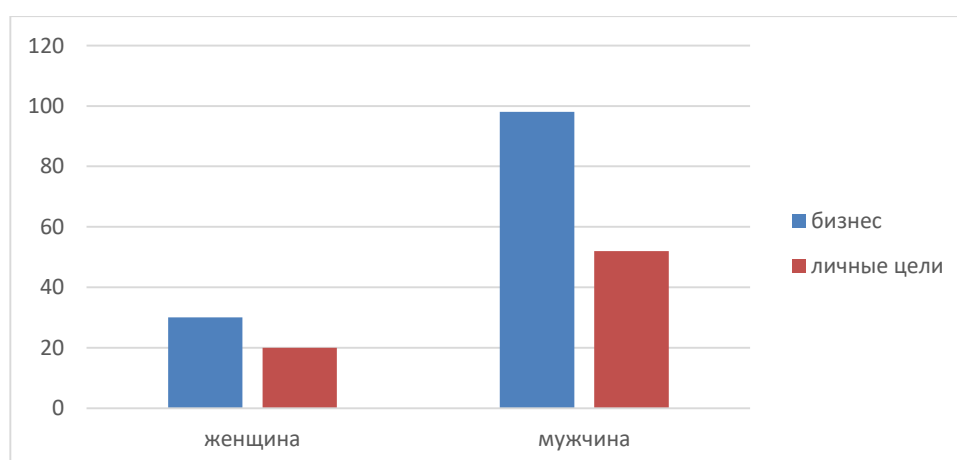


Рисунок 3 – Гистограмма отношение кол-ва полетов по целям

Из гистограммы видно, что в большей степени полёты были совершены по бизнес целям.

Строю гистограмму из второй таблицы пункта 2 (рисунок 4).

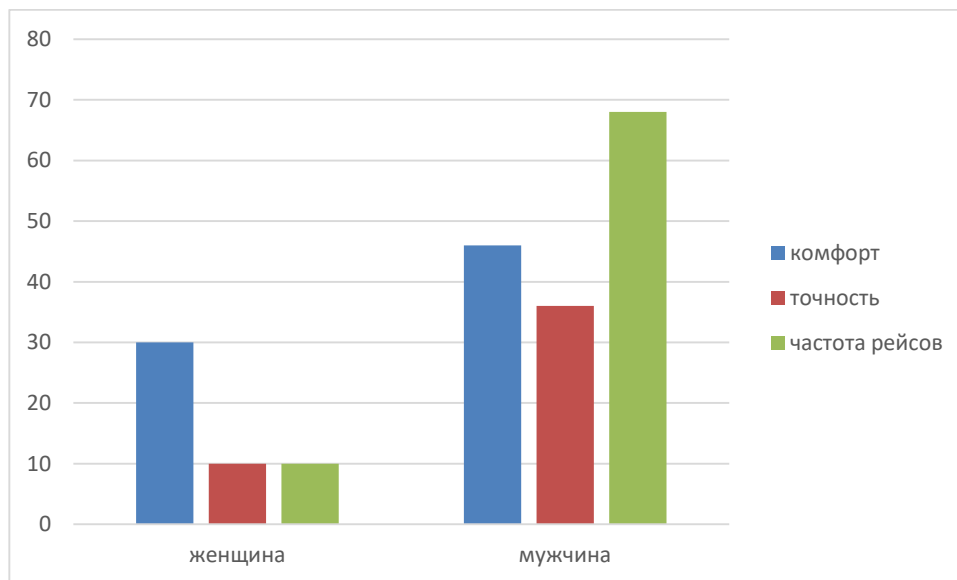


Рисунок 4 – Гистограмма отношения кол-ва пассажирам по главным критериям

Как видно из гистограммы мужчин в большей степени интересовала частота рейсов, а женщин комфорт.

Строю гистограмму из третьей таблицы второго пункта (рисунок 5).

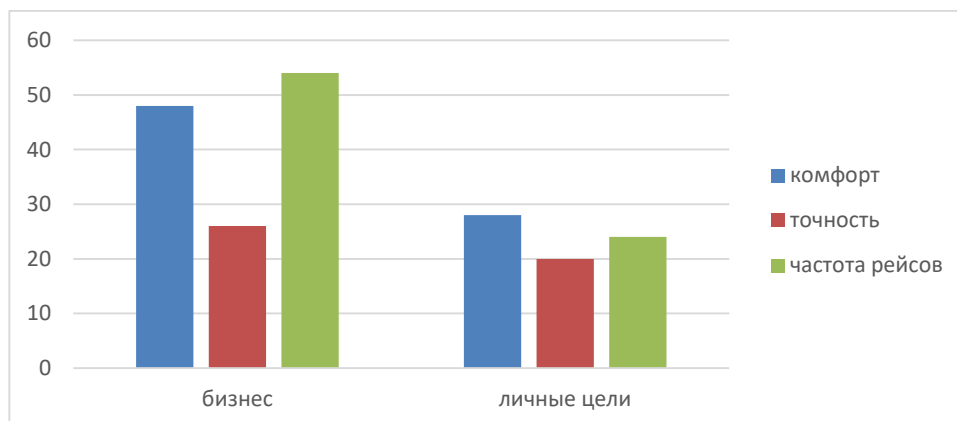


Рисунок 5 – Гистограмма важности факторов для различных целей

Из гистограммы видно, что для бизнес поездок главным критерием является частота рейсов, когда для поездок в личных целях главным является комфорт.

Новые таблицы представлены ниже.

Таблица 6. Частота полётов по возрастным интервалам

возраст				
левая граница интервала	правая граница интервала	Средняя точка интервала	Частота	Средняя частота
11	15	13	0	0
16	20	18	2	2
21	25	23	6	6
26	30	28	30	30
31	35	33	20	20
36	40	38	30	30
41	45	43	32	32
46	50	48	18	18
51	55	53	26	26
56	60	58	12	12
61	65	63	14	14
66	70	68	4	4
71	75	73	6	6
всего				

Таблица 7. Частота полётов по доходным границам

годовой доход				
левая граница интервала	правая граница интервала	Средняя точка интервала	Частота	Средняя частота
1	10	5,5	0	0
11	20	15,5	0	0
21	30	25,5	0	0
31	40	35,5	4	0
41	50	45,5	0	4
51	60	55,5	34	18
61	70	65,5	18	16
71	80	75,5	2	20
81	90	85,5	24	24
91	100	95,5	12	10
101	110	105,5	32	2
111	120	115,5	4	32
всего				

Задание 5. По данным таблицы из п.3 «Распределение пассажиров по годовому доходу» постройте кумулятивный частотный график («стрелку»).

График представлен на рисунке 6.



Рисунок 6 –Кумулятивный частотный график

Задание. 6 Оцените метод, которым была сделана выборка. Какие возможны ошибки? Предложите подходящую технику выборки респондентов для этого исследования. Что может стать источником ошибок в вашем методе?

Выборка была сделана методом опроса. Возможны ошибки только в том, что на большем количестве опрошенных могут быть другие отношения и статистические данные. Я бы делал выборку минимум от 2 тысяч людей. Если опрос заполняли от руки, могли возникнуть ошибки.

Задание 7. По форме гистограмм из п.4 оцените, симметричными или несимметричными являются распределения пассажиров по возрасту и доходу. Опишите данные по возрасту и доходу, используя уместные меры среднего и вариации.

Графики анализа симметрии представлены на рисунках 7-8.

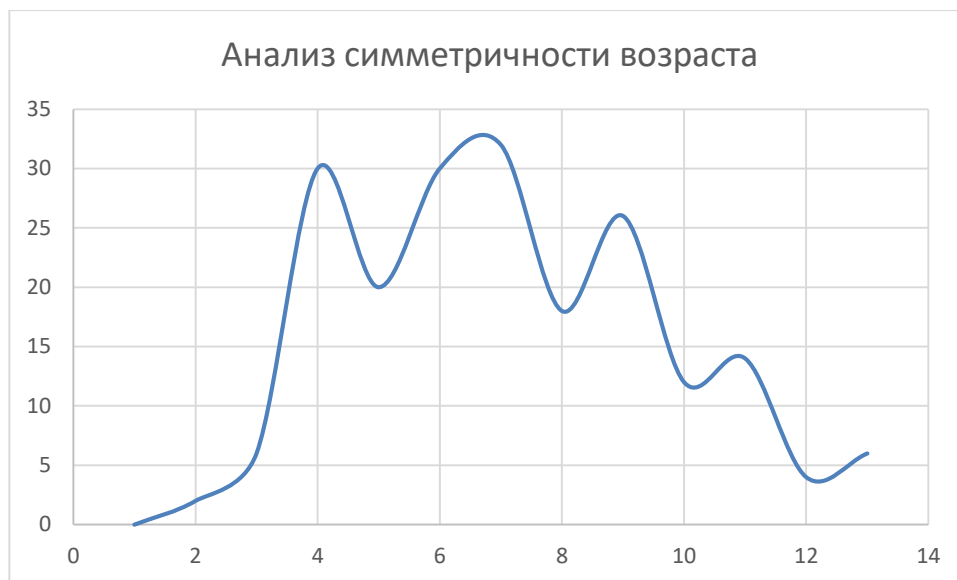


Рисунок 7 – График симметричности возраста

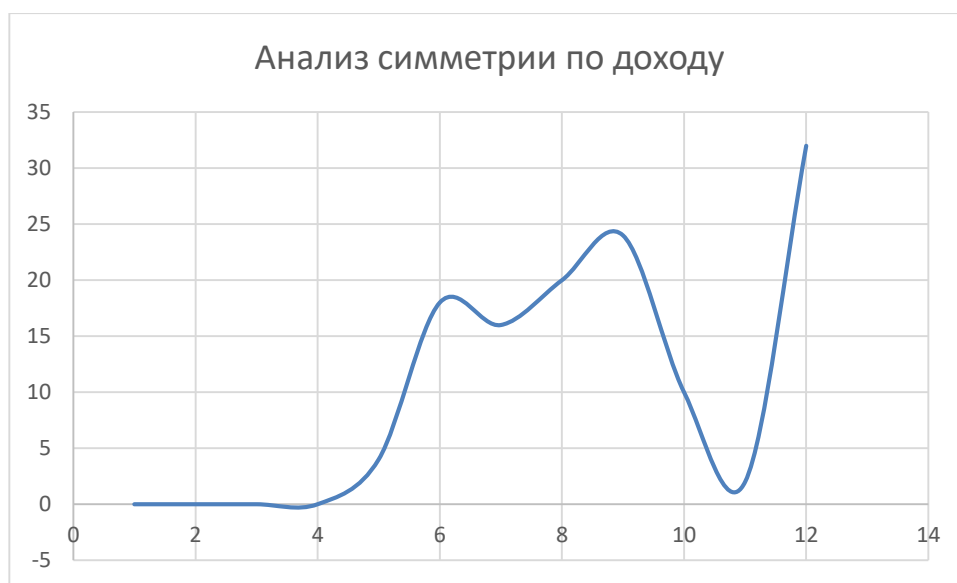


Рисунок 8 – График симметрии по доходу

Простейшим критерием симметрии является равенство среднего значения, моды и медианы. Найду среднее значение, моду и медиану для возрастов (при помощи функций в Excel):

Среднее значение (\bar{x}) = 41,54.

Мода (M_0) = 35.

Медиана ($med(x)$) = 40.

Данный набор данных является асимметричным и имеет правостороннюю асимметрию. \bar{x} смещена вправо, коэффициент асимметрии находится по формуле: $A_3 = \frac{\bar{x} - M_0}{\sigma}$, где \bar{x} - средняя, M_0 - мода, σ - среднее квадратическое отклонение:

$$\bar{x} = 41,54$$

$$M_0 = 35$$

$$\sigma = \sqrt{\frac{(x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}}$$

Для начала посчитаем отклонение возраста каждого человека от среднего для каждого пассажира и добавим в нашу таблицу:

$$x_i - \bar{x}$$

Теперь подсчитаем дисперсию:

$$\sigma^2 = \sqrt{\frac{(x_i - \bar{x})^2}{n}} = \sqrt{161,44}$$

Средне квадратическое отклонение равно:

$$\sigma = 12,7$$

Коэффициент асимметрии:

$$A_3 = \frac{42 - 35}{12,7} = 0,551$$

Найду среднее значение, моду и медиану для доходов (при помощи функций в Excel):

Среднее значение (\bar{x}) = 116.

Мода (M_0) = 108.

Медиана ($med(x)$) = 108.

Данный набор данных является асимметричным и имеет правостороннюю асимметрию. \bar{x} смещена вправо. Коэффициент асимметрии находится по формуле: $A_3 = \frac{\bar{x} - M_0}{\sigma}$, где \bar{x} - средняя, M_0 - мода, σ - среднее квадратическое отклонение:

$$\bar{x} = 116$$

$$M_0 = 108$$

$$\sigma = \sqrt{\frac{(x_1 - \bar{x}) + \dots + (x_n - \bar{x})}{n}}$$

Для начала посчитаем отклонение возраста каждого человека от среднего для каждого пассажира и добавим в нашу таблицу:

$$x_i - \bar{x}$$

Теперь подсчитаем дисперсию:

$$\sigma^2 = \sqrt{\frac{(x_i - \bar{x})^2}{n}} = \sqrt{3489,34}$$

Средне квадратическое отклонение равно:

$$\sigma = 59,07$$

Коэффициент асимметрии:

$$A_3 = \frac{116 - 108}{59,07} = 0,135$$

Также подсчитаем размах вариации для возраста и дохода:

$$R_{\text{возраста}} = X_{\max} - X_{\min}$$

$$R_{\text{возраста}} = 73 - 18 = 55$$

$$R_{\text{дохода}} = X_{\max} - X_{\min}$$

$$R_{\text{дохода}} = 321 - 39 = 282$$

Подсчитаю коэффициент вариации:

$$V = \frac{\sigma}{\bar{x}} \cdot 100\%$$

$$V_{\text{возраста}} = \frac{12,7}{41,54} \cdot 100\% = 30,57\%$$

$$V_{\text{дохода}} = \frac{59,07}{116} \cdot 100\% = 50,92\%$$

Для начала отсортирую данные по столбцу «цель полёта» (рисунок 9)

Рисунок 9 – Сортировка по столбцу цель полета

В качестве выборки возьмем 100 записей по цели полета - бизнес. Найдю доверительные интервалы. Формула доверительного интервала для среднего с известной дисперсией:

$$(\bar{x} - \frac{\sigma}{\sqrt{n}} Z_a; \bar{x} + \frac{\sigma}{\sqrt{n}} Z_a)$$

Найду $\bar{x}_{\text{бизнеса}} = 3$

Далее нужно найти дисперсию для этих 100 позиций:

$$\sigma = 1,06$$

Для 95% $Z_a = 1,96$

$$\text{Интервал равен } (3 - \frac{1,06}{\sqrt{100}} \cdot 1,96; 3 + \frac{1,06}{\sqrt{100}} \cdot 1,96) = (2,793; 3,207)$$

По критерию «личные цели» возьму выборку из 50 значений.

Найду $\bar{x}_{\text{личные цели}} = 2,88$

Далее нужно найти дисперсию для этих 50 позиций:

$$\sigma = 0,92$$

Для 95% $Z_a = 1,96$

$$\text{Интервал равен } (2,88 - \frac{0,92}{\sqrt{50}} \cdot 1,96; 2,88 + \frac{0,92}{\sqrt{50}} \cdot 1,96) = (2,625; 3,135)$$

Задание 9. Определите 95% доверительный интервал для доли женщин среди пассажиров авиакомпании.

Для удобства отсортирую таблицу по признаку «пол». В качестве выборки возьму 30 записей по признаку «пол» - женщина.

Найду $\bar{x}_{\text{бизнеса}} = 1,71$

Далее нужно найти дисперсию для этих 100 позиций:

$$\sigma = 0,46$$

Для 95% $Z_a = 1,96$

$$\text{Интервал равен } (1,71 - \frac{0,46}{\sqrt{50}} \cdot 1,96; 1,71 + \frac{0,46}{\sqrt{50}} \cdot 1,96) = (1,583; 1,837)$$

Задание 10. Для пассажиров-женщин определите 95% доверительный интервал доли тех, кто летает с личными целями.

Найду $\bar{x}_{\text{женщин-личные}} = 3,05$

Далее нужно найти дисперсию для этих 100 позиций:

$$\sigma = 2,94$$

Для 95% $Z_a = 1,96$

$$\text{Интервал равен } (3,05 - \frac{2,94}{\sqrt{19}} \cdot 1,96; 3,05 + \frac{2,94}{\sqrt{19}} \cdot 1,96) = (1,73; 4,37)$$

Задание 11. Есть ли основания полагать, что среднее число полетов у мужчин больше, чем у женщин?

Среднее кол-во полетов у мужчин = 2,76

Среднее кол-во полетов у женщин = 3,45

Следовательно, основания полагать, что среднее число полетов у мужчин больше, чем у женщин – есть.

Задание 12. Есть ли основания полагать, что среднее число полетов различно для тех пассажиров, кто летает с целью бизнеса, и тех, кто летает в личных целях?

Среднее кол-во полетов с целью бизнеса = 3,35

Среднее кол-во полетов в личных целях = 3,13

Следовательно, среднее число полётов с целью бизнеса и личных целей различны.

Задание 13. Существует ли связь между полом пассажира и целью полета?

Связь существует, но она не значительна:

- 65% мужчин летали в бизнес целях, а женщин в бизнес целях летало 60%.

35% мужчин летали в личных целях, а женщин в личных целях летало 40%.

Задание 14. Существует ли связь между полом пассажира и самой важной для него характеристикой полетов?

Отношения важных факторов показаны на рисунках 10-11

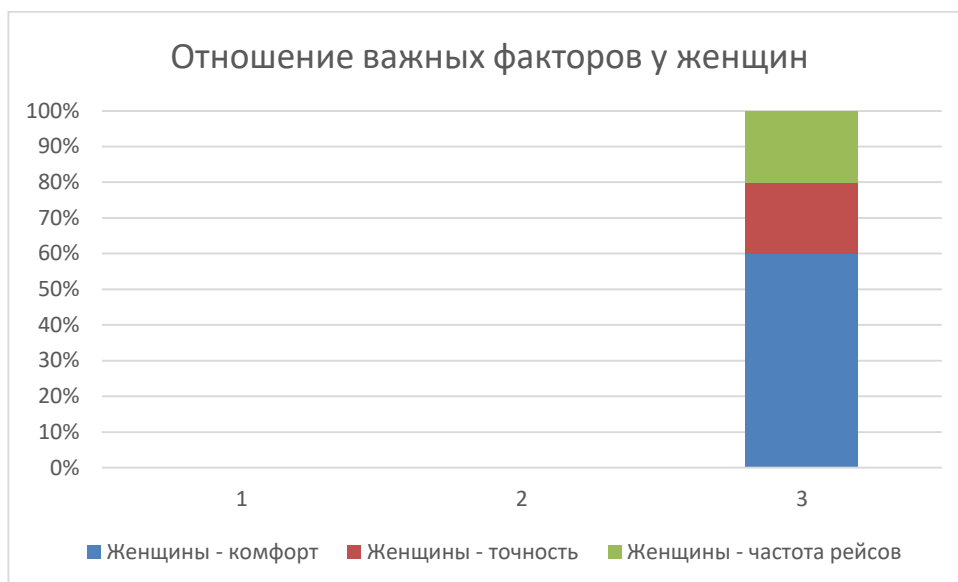


Рисунок 10 – Отношение важных факторов у женщин

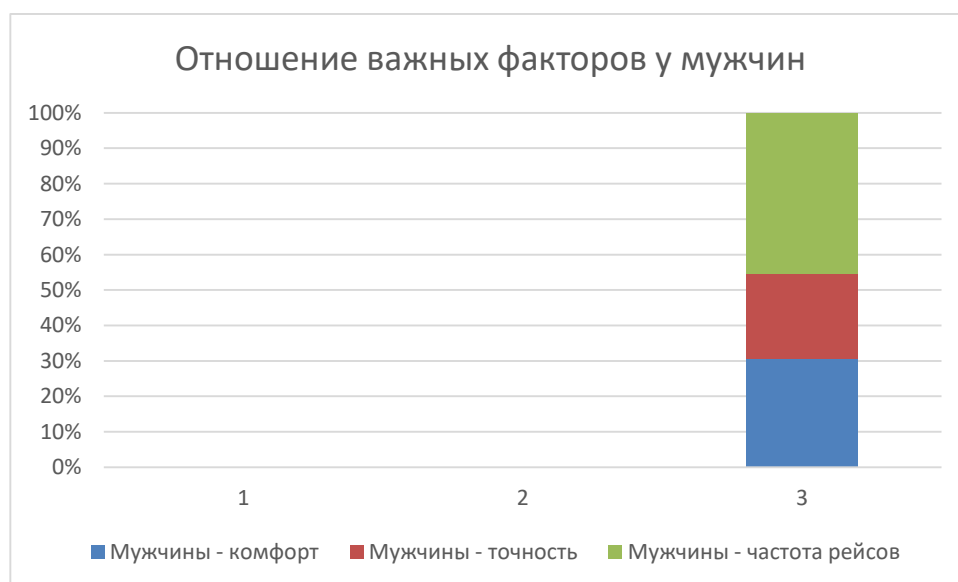


Рисунок 11 – Отношение важных факторов у мужчин

Из диаграмм отчетливо видно, что для женщин самый важный критерий – комфорт, а для мужчин – частота рейсов. Следовательно, данная связь существует.

Задание 15. Существует ли связь между целью полета пассажира и наиболее важной для него характеристикой полета?

Гистограммы важности характеристик для бизнес полётов и полётов в личных целях представлены на рисунках 12-13.



Рисунок 12 Гистограмма самая важна характеристика - бизнес

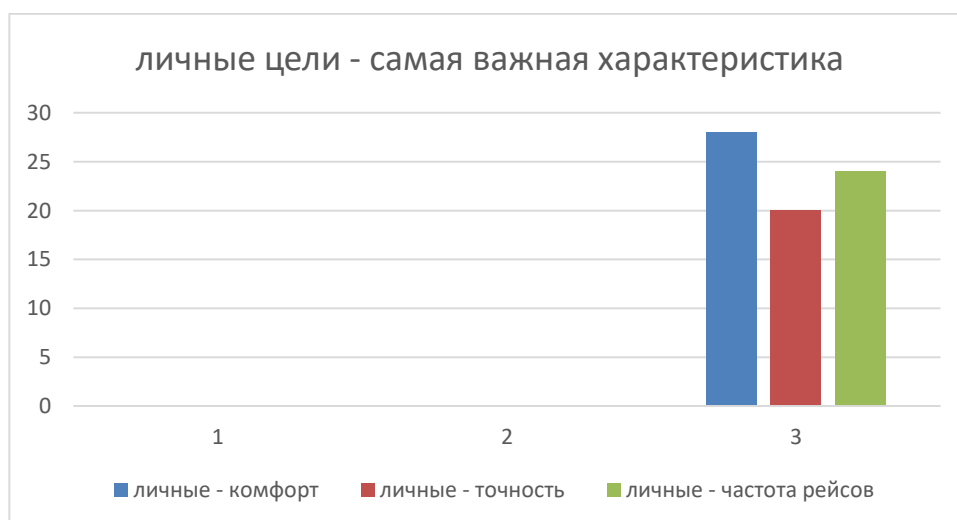


Рисунок 13 – Гистограмма самая важна характеристика – личные цели

Как видно из диаграмм для бизнеса самым важным критерием является частота рейсов, а для личных целей – комфорт. Следовательно, такая зависимость существует.

Задание 16. Есть ли основания полагать, что доля тех, кто ценит комфорт, больше для женщин?

Да, так как для 60% женщин важен комфорт, а для мужчин комфорт не стоит на первом месте.

Задание 17. Изложите смысл результатов, полученных в п.п.11-16.

Таким образом понятно, что мужчин в среднем летает больше, чем женщин. Также видно, что больше полётов совершено в бизнес целях, существует небольшая зависимость между полом и целью полетов. Видно, что существует четкая связь между полом и важностью факторов, целью полёта и важностью факторов. Также чётко видно, что женщины ценят комфорт, когда для мужчин важнее частота рейсов.

Задание 18. Постройте диаграммы рассеяния, чтобы показать связь между числом полетов и возрастом пассажира, и числом полетов и доходом пассажира. Прокомментируйте форму.

Диаграмма рассеивания показана на рисунке 14.

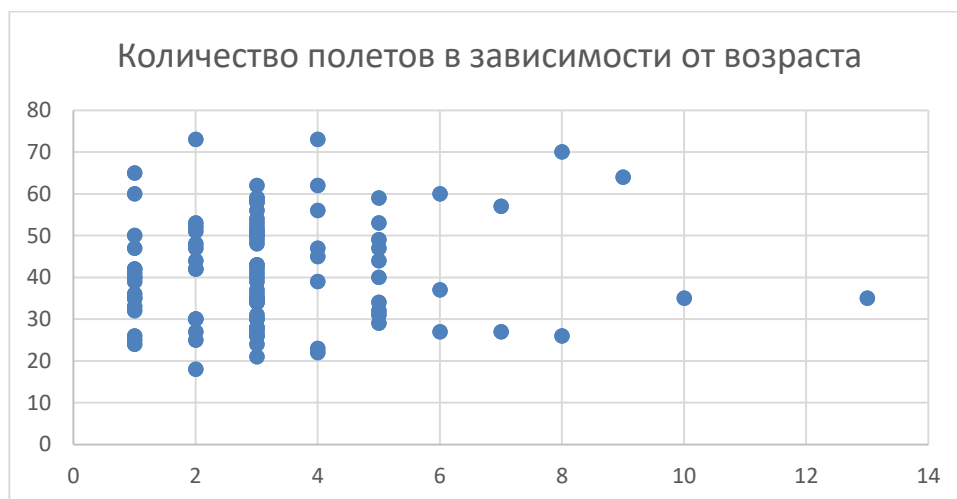


Рисунок 14 – Диаграмма рассеивания

Из диаграммы видно, что в наиболее часто встречаются люди летавшие 3, 2, 5 раз за год. Также наиболее часто летают люди в возрасте от 20 до 60 лет. Больше всего полётов у людей в возрасте 35 лет. Следующая диаграмма представлена на рисунке 15.

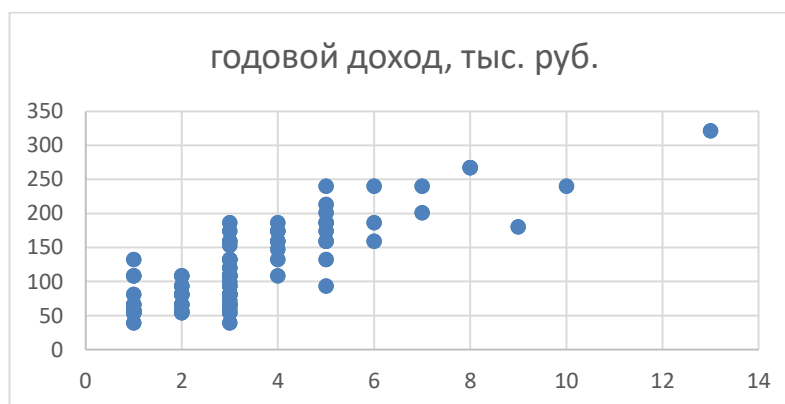


Рисунок 15 – Гистограмма рассеивания

Из диаграммы видно, что в основном от дохода зависит количество полётов.

Задание 19.

Вычислите коэффициент корреляции между

- числом полетов и возрастом пассажира
- числом полетов и доходом пассажира
- возрастом пассажира и доходом пассажира

Для значимых коэффициентов корреляции вычислите коэффициент детерминации и поясните его смысл.

Коэффициент корреляции между числом полетов и возрастом пассажира равна 0,077 – проглядывается маленькая положительная корреляция, следовательно, связи между этими переменными практически нет.

Коэффициент корреляции между числом полетов и доходом пассажира равен 0,835 – проглядывается сильная положительная корреляция, следовательно, количество полётов сильно зависит от дохода в год.

Коэффициент корреляции между возрастом пассажира и его доходом равен -0,031 – проглядывается очень маленькая отрицательная корреляция, следовательно, эти данные практически не зависят друг от друга.

Коэффициент детерминации между количеством полётов и доходом пассажира равен 0,70 (вычислен при помощи функции «КВПИРСОН»). Данный коэффициент говорит о том, что в данной ситуации количество полётов сильно зависит от дохода пассажира.

Задание 20. Выберите наиболее приемлемую независимую переменную и постройте простую линейную регрессию. Проверьте значимость коэффициентов.

Я хочу построить линейную регрессию для выявления зависимости между полом пассажира и его доходом. Для этого нужно нормализовать

данные в столбе пол. Пусть мужчина = 0, а женщина = 1. Результат нормализации представлен на рисунке 16.

1	Количество полетов за последний год	пол	цель полета	самая важная характеристика	годовой доход, тыс. руб.	возраст
2	1	1	бизнес	комфорт	66	35
3	1	1	бизнес	комфорт	54	60
4	1	1	бизнес	комфорт	66	35
5	1	1	бизнес	комфорт	54	33
6	1	1	бизнес	комфорт	54	33
7	1	1	бизнес	комфорт	54	60
8	1	0	бизнес	комфорт	57	65
9	1	0	бизнес	комфорт	57	36
10	1	0	бизнес	комфорт	57	65
11	1	0	бизнес	комфорт	57	36
12	1	1	бизнес	точность	108	26
13	1	1	бизнес	точность	66	39
14	1	1	бизнес	точность	66	39
15	1	1	бизнес	точность	108	26
16	1	1	бизнес	частота рейсов	54	41
17	1	1	бизнес	частота рейсов	54	41
18	1	0	бизнес	частота рейсов	57	50
19	1	0	бизнес	частота рейсов	57	32
20	1	0	бизнес	частота рейсов	57	32
21	1	0	бизнес	частота рейсов	39	35
22	1	0	бизнес	частота рейсов	39	35
23	1	0	бизнес	частота рейсов	57	50
24	2	0	бизнес	комфорт	81	30
25	2	0	бизнес	комфорт	54	52
26	2	0	бизнес	комфорт	54	52
27	2	0	бизнес	комфорт	81	30
28	2	1	бизнес	точность	66	51
29	2	1	бизнес	точность	66	51
30	2	0	бизнес	точность	93	44
31	2	0	бизнес	точность	93	44
32	2	0	бизнес	частота рейсов	81	27
33	2	0	бизнес	частота рейсов	54	42
34	2	0	бизнес	частота рейсов	108	42
35	2	0	бизнес	частота рейсов	108	42

Рисунок 16 - Нормализация данных «пол»

Строю регрессию (рисунок 17).

Вывод итогов								
Регрессионная статистика								
Множественный коэффициент корреляции	0,029909							
R-квадрат	0,000895							
Нормированный R-квадрат	-0,00415							
Стандартная ошибка	0,434999							
Наблюдения	200							
Дисперсионный анализ								
	df	SS	MS	F	Значимость F			
Регрессия	1	0,033545	0,033545	0,177275	0,674183			
Остаток	198	37,46646	0,189225					
Итого	199	37,5						
Коэффициенты регрессии								
У-пересечение	0,275507	0,067943	4,054992	7,21E-05	0,141523	0,409491	0,141523	0,409491
Переменная	-0,00022	0,000521	-0,42104	0,674183	-0,00125	0,000808	-0,00125	0,000808

Рисунок 17 - Таблица регрессии

Важными для нас являются значения R-квадрат и коэффициент Y-пересечения. R-квадрат показывает на сколько успешна наша модель линейной регрессии. В данном случае модель не является успешной. Успешной можно считать модель у которой R-квадрат находится в диапазоне от 0,5 до 1.0 (50-100%). Y-пересечение показывает какое значение будет Y, а коэффициент переменной X, показывает насколько X зависит от Y. У данных переменных наблюдается очень маленькая отрицательная корреляция, следовательно эти переменные не зависят друг от друга. Таким образом данная модель не является успешной и не имеет смысла.

Задание 21. Какие предположения сделаны при построении регрессии? Оцените эти предположения, используя график остатков. Можно ли использовать полученную модель как инструмент прогнозирования?

В данном случае модель не является успешной. Успешной можно считать модель у которой R-квадрат находится в диапазоне от 0,5 до 1.0 (50-100%). Y-пересечение показывает какое значение будет Y, а коэффициент переменной X, показывает насколько X зависит от Y. У данных переменных наблюдается очень маленькая отрицательная корреляция, следовательно эти переменные не зависят друг от друга. Таким образом данная модель не является успешной и не имеет смысла. Таблица рассеивания показана на рисунке 18.

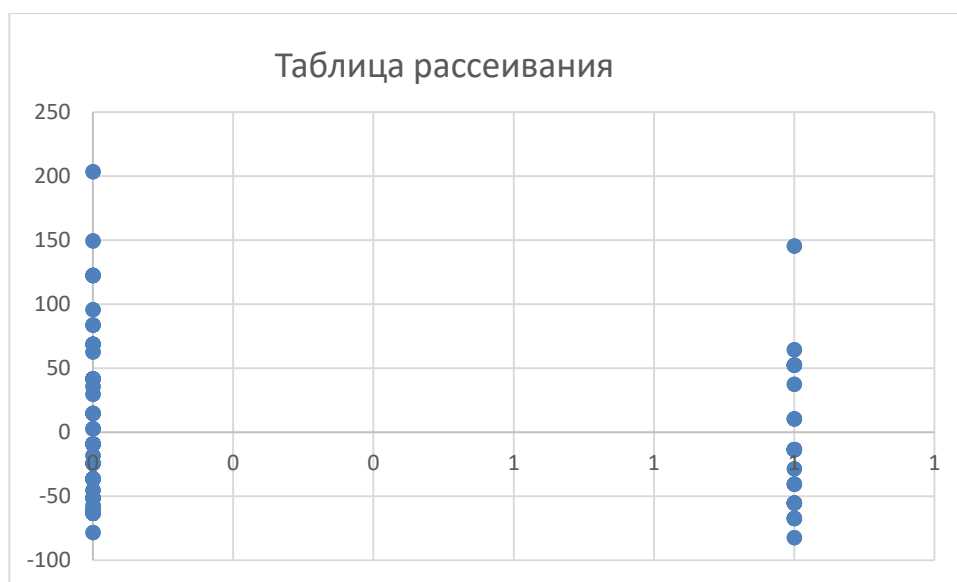


Рисунок 18 – Таблица рассеивания

Данную модель нельзя использовать как инструмент прогнозирования.

Задание 22. Спрогнозируйте число полетов пассажира со следующими характеристиками: женщина, летает с целью бизнеса, ценит точность, ей 45, она зарабатывает 32. Что влияет на вашу уверенность в прогнозе?

Для того, чтобы построить модель регрессии с такими характеристиками, как: пол, цель полёта, самая важная характеристика, возраст, годовой доход, нужно нормализовать данные по типу пол, цель полёта, самая важная характеристика.

Нормализация будет проходить путём заменой с текстовой информации на числовую. Для характеристики «пол» будет произведена следующая замена мужчина = 1, женщина = 0. Для характеристики «цель полёта» будет произведена следующая замена бизнес = 1, личные цели = 0. Для характеристики «самая важная характеристика» будет произведена следующая замена комфорт = 2, точность = 1, частота рейсов = 0.

Построю таблицу регрессии (рисунок 19).

	A	B	C	D	E	F	G	H	I	J
1	ВЫВОД ИТОГОВ									
2										
3	Дисперсионная статистика									
4	Множественный R	0,859929								
5	R-квадрат	0,739479								
6	Нормированный R-квадрат	0,732764								
7	Стандартная ошибка	1,06229								
8	Наблюдения	200								
9										
10	Дисперсионный анализ									
11		df	SS	MS	F	Значимость F				
12	Регрессия	5	621,3986	124,2797	110,1321	1,02E-54				
13	Остаток	194	218,9214	1,128461						
14	Итого	199	840,32							
15										
16	Коэффициентная статистика t-Значения нижние 95% верхние 95% нижние 95% верхние 95% нижние 95% верхние 95%									
17	Y-пересечение	-1,07947	0,339426	-3,18029	0,001713	-1,74891	-0,41003	-1,74891	-0,41003	
18	пол	0,383605	0,181282	2,116062	0,035615	0,026067	0,741142	0,026067	0,741142	
19	цель полёта	0,295548	0,157932	1,871369	0,062799	-0,01594	0,607031	-0,01594	0,607031	
20	самая важная характеристика	-0,26951	0,090575	-2,97553	0,003297	-0,44815	-0,09087	-0,44815	-0,09087	
21	годовой доход	0,029364	0,001278	22,98434	2,8E-57	0,026844	0,031883	0,026844	0,031883	
22	возраст	0,017652	0,00605	2,917939	0,00394	0,005721	0,029584	0,005721	0,029584	
23										
24										
25										
26										

Рисунок 1 – Таблица регрессии

R-квадрат равен 0,74 (74%), что является хорошим результатом для модели. Построю формулы по которой буду предсказывать кол-во полётов:

$$y = -1,079 + 0,383x_1 + 0,295x_2 - 0,269x_3 + 0,029x_4 + 0,0176x_5$$

где

x_1 – пол

x_2 – цель_полёта

x_3 – самая_важная_характеристика

x_4 – годовой_доход

x_5 – возраст

Подставляю данные и подсчитаю кол-во полётов:

$$y = -1,079 + 0,383 \bullet 0 + 0,295 \bullet 1 - 0,269 \bullet 1 + 0,029 \bullet 32 + 0,0176 \bullet 45$$

$$y = 0,680556$$

x1	0
x2	1
x3	1
x4	32
x5	45
y	0,680556

Моя модель предсказала, что пассажир с данными характеристиками совершит 1 полёт в год.

Задание 23. Учитывая результаты исследования, какие рекомендации Вы могли бы дать авиакомпании?

Я бы порекомендовал авиакомпании:

1. Для женщин предлагать перелёты опираясь на уровень комфорта.
2. Для мужчин предлагать перелёты опираясь на частоту рейсов.
3. Если люди летят для решения каких-то бизнес целей, в первую очередь для них учитывать частоту рейсов.
4. Если люди летят в каких-то личных целях (отдых, туризм, и тд.), в первую очередь учитывать уровень комфорта.
5. Также было бы хорошо смотреть статистику полётов клиентов и для тех, кто летает чаще предлагать более дорогие услуги, так как чем

больше полётов совершает пассажир за год, тем больше его уровень дохода.

6. Также можно комбинировать упомянутые выше пункты, например, если женщина летит в целях решения бизнес вопросов, то на ряду с частотой рейсов для неё важен комфорт.

Все файлы данной лабораторной работы были загружены на github, включая «.xlsx» файлы, в которых была проделана данная работа.