

1. КРАТКИЕ ТЕОРЕТИЧЕСКИЕ СВЕДЕНИЯ

Основные термины и понятия, используемые в контрольной работе.

Корреляция — это статистическая зависимость между случайными величинами, не имеющая строго функционального характера, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой. **Парная корреляция** — связь между двумя величинами (факторами) — результативным и факторным или двумя факторными.

Регрессионный метод заключается в определении аналитического выражения связи двух величин, в котором изменение одной (зависимой) величины обусловлено влиянием одной или нескольких независимых величин, а влияние множества всех прочих факторов, также оказывающих влияние на зависимую величину, принимается за постоянные средние значения. По форме зависимости различают **линейную регрессию** (выражается уравнением прямой линии) и **нелинейную регрессию** (описывается другими видами уравнений).

Модель парной линейной регрессии имеет вид:

$$y = \alpha + \beta x + u,$$

где y — зависимая переменная; x — независимая переменная; α, β — параметры модели; u — случайный остаточный член (случайная ошибка). Константу α называют также **свободным членом**, а угловой коэффициент β — **коэффициентом регрессии**.

Для оценки параметров модели используется уравнение:

$$\hat{y} = a + bx,$$

где \hat{y} — прогнозируемое значение объясняемой (зависимой) переменной; a — статистическая оценка параметра α ; b — статистическая оценка параметра β . Разность между фактическим значением зависимой переменной и значением, прогнозируемым по уравнению регрессии, называется **остатком**. Остатки e_i вычисляются по формуле:

$$e_i = y_i - \hat{y}_i.$$

Для определения существования и степени линейной зависимости между переменными сначала строится **диаграмма рассеивания** — графическое отображение точек $(x_1, y_1), \dots, (x_n, y_n)$ на плоскости, по которой делается вывод о допустимости предположения о линейной зависимости между x и y . Затем вычисляется коэффициент корреляции r по формуле:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}.$$

Линейный коэффициент корреляции характеризует тесноту и направление связи между двумя коррелируемыми признаками в случае наличия между ними линейной зависимости. Значения r находятся в пределах от -1 до 1 . Чем ближе абсолютное значение r к единице, тем сильнее связь между признаками.

Оценивание параметров уравнения регрессии с помощью метода наименьших квадратов (МНК) предполагает, что отсутствует систематическая связь между значениями случайной ошибки в любых двух наблюдениях:

$$\text{cov}(u_i, u_j) = 0, i \neq j.$$

Это условие часто нарушается в случае, когда данные являются временными рядами, и значения случайной ошибки в текущем и предыдущем наблюдениях взаимосвязаны. Если условие некоррелированности ошибок не выполняется, то говорят об *автокорреляции* ошибок.

Еще одним условием применения МНК является следующее — дисперсия ошибки должна быть постоянна для всех наблюдений:

$$D(u_i) = \sigma^2 = \text{const}, i = 1 \dots n.$$

Независимость дисперсии ошибки от номера наблюдения называется *гомоскедастичностью*. Случай непостоянства дисперсии ошибки для разных наблюдений называется *гетероскедастичностью*. Иными словами, если дисперсия остатков одинакова для всех наблюдений, то остатки гомоскедастичны, если не одинакова, то — гетероскедастичны.

Проверку гипотез рассматривают с двух точек зрения:

1. Сначала формулируется гипотеза, и цель эксперимента заключается в выяснении ее справедливости. Это приводит к проверке гипотезы о значимости.
2. Сначала проводится эксперимент и затем определяется, какие из теоретических гипотез соответствуют результатам эксперимента. Это приводит к построению доверительных интервалов.

Формулирование нулевой гипотезы выполняется следующим образом. Рассмотрим функцию $y = \alpha + \beta x + u$, где y — величина расходов населения, x — доходы. Предполагаем, что расходы зависят от доходов. Обычно строят нулевую гипотезу H_0 . Она проверяется с помощью альтернативной гипотезы H_1 , которая предполагается верной. В качестве нулевой гипотезы принимается утверждение о том, что расходы не зависят от доходов (y не зависит от x), т.е. $\beta = 0$. Альтернативная гипотеза — $\beta \neq 0$ (x влияет на y), т.е. доходы влияют на расходы.

В общем случае для нулевой гипотезы утверждают, что $\beta = \beta_0$, тогда альтернативная гипотеза — $\beta \neq \beta_0$. Если H_0 верна, то в МНК-оценки для β будут иметь распределение с математическим ожиданием и дисперсией:

$$E(b) = \beta_0, \quad D(b) = \frac{\sigma^2}{\sum (x_i - \bar{x})^2}.$$

Проверка гипотез заключается в следующем. Строится t -статистика для b :

$$t = \frac{b - \beta_0}{s_b}.$$

Если гипотеза H_0 верна, то t -статистика подчиняется распределению Стьюдента (t -распределению), заданному таблично. Параметром распределения Стьюдента является ν — *число степеней свободы*, которое определяется по правилу: оценивание каждого параметра поглощает одну степень свободы. Следовательно, $\nu = n - 1$, где n — количество наблюдений. В таблице распределения Стьюдента для различных ν заданы критические значения t -статистики, ко-

торые обозначаются $t_{кр}$. Гипотеза $t = 0$ эквивалентна H_0 . Надо проверить: $-t_{кр} < t < t_{кр}$, если выполнено, то не отказываемся от нулевой гипотезы. Если $t < -t_{кр}$ или $t > t_{кр}$, то H_0 отклоняем.

Процедура проверки гипотезы: вычислить t ; задать $\alpha = 0.05$; найти $t_{кр}$; проверить попало t в критическую область или нет; если попало, то H_0 отвергаем (есть влияние), если не попало, то H_0 не отвергаем.

Эквивалентная процедура проверки гипотез: вычислить t ; найти p -значение, равное $p_r(|t| > |t_{кр}|)$ — вероятности того, что при выполнении H_0 статистика критерия (t) принимает значение более экстремальное, чем $t_{кр}$; если p -значение меньше α , то H_0 отвергаем.

Построенные оценки параметров уравнения регрессии являются точечными оценками. Доверительные интервалы для параметров уравнения регрессии являются их интервальными оценками.

100 (1 - α) % -й доверительный интервал для b : $b \pm s_b \times t_{1-\frac{\alpha}{2}}(n-2)$,

100 (1 - α) % -й доверительный интервал для a : $a \pm s_a \times t_{1-\frac{\alpha}{2}}(n-2)$.

Оценивание надежности параметров парной регрессии может быть выполнено с помощью t -критерия Стьюдента. Для его вычисления сначала находится средняя ошибка оценки для b :

$$m_b = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (x_i - \bar{x}_i)^2 (n-2)}}$$

где $(n-2)$ — число степеней свободы (линейная парная регрессия имеет два параметра, поэтому теряются две степени свободы). Затем определяется t -критерий по формуле:

$$t = \frac{b}{m_b}$$

Табличное значение t -критерия Стьюдента при $(n-2)$ степенях свободы и заданном уровне значимости (например, 0.01) сравнивается с полученным. Если полученное значение больше табличного, то вероятность нулевого значения коэффициента регрессии менее 0.01. Следовательно, гипотезу о несущественности влияния независимой переменной на зависимую можно отклонить.

Дополнительно перед выполнением контрольной работы целесообразно изучить материал, приведенный в учебном пособии [5]. Особое внимание следует уделить темам: парный регрессионный анализ, коэффициенты регрессии, проверка гипотез, гетероскедастичность, свойства и автокоррелированность случайного члена (см. главы 2, 3, разделы 7.1 – 7.5 в учебнике [3]).

2. ИНДИВИДУАЛЬНЫЕ ЗАДАНИЯ

Провести эконометрический анализ данных о среднедушевых денежных доходах и расходах населения регионов РФ. Для этого:

1. Построить гистограмму, сделать содержательные выводы.
2. Построить диаграмму рассеивания, определить коэффициент корреляции между X и Y , сделать вывод.
3. Определить параметры линейной регрессионной модели. Записать уравнение регрессии.
4. Определить предсказанное значение Y , вычислить остатки.
5. Построить линию регрессии на диаграмме рассеивания.
6. Вынести на диаграмму уравнение и значение коэффициента детерминации.
7. Проверить адекватность модели и провести интерпретацию уравнения регрессии.
8. Определить остатки, построить график остатков, интерпретировать полученный результат.
9. С помощью модуля "Анализ данных" — "Регрессия" определить значение критерия Фишера, сделать вывод.
10. С помощью модуля "Анализ данных" — "Регрессия" проверить значимость оценки коэффициента регрессии, построить доверительный интервал для коэффициента регрессии.

Каждый студент выполняет индивидуальное задание. Исходные данные для выполнения контрольной работы приведены в разделе 2. Студент выбирает для анализа сведения о величине среднедушевых денежных доходов и расходов населения для того федерального округа и года, которые указаны в таблице 1 напротив порядкового номера студента в списке группы. Например, студенту, которому соответствует номер 1 в списке группы, надо анализировать данные для населения регионов Центрального федерального округа за 2004 год.

Таблица 1

Варианты индивидуальных заданий

Порядковый номер студента в списке группы	Федеральный округ	Год
1	Центральный	2004
2	Центральный	2005
3	Северо-Западный	2004
4	Южный	2004
5	Южный	2005
6	Приволжский	2004
7	Приволжский	2005
8	Сибирский	2004
9	Сибирский	2005
10	Дальневосточный	2004

22 - Л

23 - Юм Ш.

7 - Сете

16 - Юм Ш.

10 - Юм Ш.

Продолжение табл. 1

Порядковый номер студента в списке группы	Федеральный округ	Год
11	Дальневосточный	2005
12	Центральный	2002
13	Центральный	2003
14	Северо-Западный	2002
15	Северо-Западный	2003
16	Южный	2002
17	Южный	2003
18	Приволжский	2002
19	Приволжский	2003
20	Сибирский	2002
21	Сибирский	2003
22	Дальневосточный	2002
23	Дальневосточный	2003
24	Центральный	1999
25	Центральный	2000
26	Центральный	2001
27	Северо-Западный	1999
28	Северо-Западный	2000
29	Северо-Западный	2001
30	Южный	1999
31	Южный	2000
32	Южный	2001
33	Приволжский	1999
34	Приволжский	2000
35	Приволжский	2001
36	Сибирский	1999
37	Сибирский	2000
38	Сибирский	2001
39	Дальневосточный	1999
40	Дальневосточный	2000
41	Дальневосточный	2001

3. ИСХОДНЫЕ ДАННЫЕ ДЛЯ ВЫПОЛНЕНИЯ КОНТРОЛЬНОЙ РАБОТЫ

В таблицах 2, 3 представлены среднедушевые денежные доходы и расходы населения, характеризующие уровень жизни регионов Российской Федерации в 1999 – 2005 годах (модельные данные).

Таблица 2

Среднедушевые денежные доходы населения
(в месяц, рублей)

	1999	2000	2001	2002	2003	2004	2005
Центральный федеральный округ	698	1065	1313	1497	2481	3334	4446
Белгородская область	392	593	658	719	1190	1560	2131
Брянская область	340	524	595	586	910	1259	1685
Владимирская область	314	467	568	583	975	1152	1626
Воронежская область	343	465	672	782	1178	1476	2023
Ивановская область	306	454	546	512	740	1013	1264
Калужская область	435	567	639	690	1018	1298	1685
Костромская область	391	522	605	602	1004	1392	1838
Курская область	309	455	584	630	1054	1385	1863
Липецкая область	376	619	690	736	1180	1763	2247
Московская область	395	586	744	914	1488	1881	2684
Орловская область	380	577	651	707	1083	1387	1922
Рязанская область	335	506	608	607	973	1256	1836
Смоленская область	366	546	647	704	1214	1667	2256
Тамбовская область	307	457	562	616	1081	1461	2000
Тверская область	344	468	534	578	937	1206	1616
Тульская область	398	548	709	694	1104	1484	2013
Ярославская область	472	627	727	794	1297	1794	2503
г. Москва	1804	2846	3524	4083	6857	9285	12137
Северо-Западный федеральный округ	560	786	908	1007	1594	2244	3017
Республика Карелия	673	921	1023	1039	1684	2216	2798
Республика Коми	666	986	1260	1373	2081	2922	4501
Архангельская область	505	671	792	805	1369	2033	2879
Вологодская область	499	699	877	870	1371	1984	2638

Продолжение табл. 2

Северо-Западный федеральный округ	1999	2000	2001	2002	2003	2004	2005
Калининградская область	379	512	595	709	1333	1952	2139
Ленинградская область	371	544	601	653	1015	1366	1885
Мурманская область	740	1062	1300	1509	2367	3387	4377
Новгородская область	434	702	757	939	1358	1782	2371
Псковская область	340	476	558	597	963	1360	1742
г. Санкт-Петербург	672	923	1022	1179	1838	2583	3468
Южный федеральный округ	313	489	604	632	1053	1445	1930
Республика Адыгея	274	413	546	607	968	1387	1646
Республика Дагестан	193	271	373	405	682	984	1317
Кабардино-Балкарская Республика	266	396	479	526	898	1180	1654
Республика Калмыкия	232	338	476	470	740	957	1508
Карачаево-Черкесская Республика	232	325	438	455	718	1055	1366
Республика Северная Осетия - Алания	319	445	576	631	1104	1522	1699
Краснодарский край	358	577	676	695	1157	1583	2109
Ставропольский край	332	522	602	640	1060	1433	1807
Астраханская область	321	522	658	718	1209	1760	2300
Волгоградская область	331	518	672	631	1026	1410	1947
Ростовская область	333	524	655	720	1238	1692	2329
Приволжский федеральный округ	367	550	687	737	1233	1700	2285
Республика Башкортостан	325	499	626	695	1245	1738	2405
Республика Марий Эл	259	367	437	422	802	1046	1356
Республика Мордовия	284	409	513	513	830	1107	1538
Республика Татарстан	394	546	691	760	1248	1819	2482
Удмуртская Республика	319	581	707	658	1035	1478	1974
Чувашская Республика	301	431	480	469	817	1120	1525
Кировская область	381	544	619	610	984	1328	1785
Нижегородская область	382	532	654	721	1163	1703	2383

Продолжение табл. 2

Приволжский федеральный округ	1999	2000	2001	2002	2003	2004	2005
Оренбургская область	331	531	671	638	1078	1443	1916
Пензенская область	301	451	562	541	936	1234	1668
Пермская область	479	757	954	993	1624	2329	3151
Самарская область	476	731	928	1201	2028	2600	3256
Саратовская область	326	456	620	610	1064	1436	1930
Ульяновская область	312	505	625	661	973	1239	1662
Уральский федеральный округ	603	902	1131	1116	1827	2724	3790
Курганская область	275	411	521	528	917	1293	1714
Свердловская область	503	706	846	852	1460	2123	2969
Тюменская область	1085	1736	2209	2129	3264	4931	7102
Челябинская область	415	581	737	741	1303	1962	2492
Сибирский федеральный округ	489	708	867	844	1360	1888	2508
Республика Алтай	316	474	608	549	856	1168	1599
Республика Бурятия	372	497	738	643	1144	1590	2188
Республика Тыва	314	450	590	548	784	1147	1609
Республика Хакасия	465	737	831	756	1294	1823	2362
Алтайский край	349	436	506	533	907	1224	1691
Красноярский край	594	930	1139	1171	1938	2745	3526
Иркутская область	580	811	983	1046	1703	2281	2758
Кемеровская область	701	846	1063	972	1511	2257	3058
Новосибирская область	340	698	859	753	1240	1597	2126
Омская область	429	667	782	793	1133	1562	2309
Томская область	493	705	959	921	1492	2051	2823
Читинская область	421	549	570	509	791	1139	1574
Дальневосточный федеральный округ	639	941	1099	1149	1796	2365	3130
Республика Саха (Якутия)	957	1486	1741	1762	2844	3606	4617
Приморский край	538	730	843	910	1337	1770	2324

Продолжение табл. 2

Дальневосточный федеральный округ	1999	2000	2001	2002	2003	2004	2005
Хабаровский край	528	818	993	1135	1850	2419	3273
Амурская область	489	740	873	846	1294	1693	2135
Камчатская область	949	1438	1649	1597	2513	3349	4410
Магаданская область	961	1334	1516	1630	2466	3206	4293
Сахалинская область	672	989	1127	1146	1901	2653	3666
Еврейская автономная область	411	589	666	679	1022	1458	1983
Чукотский автономный округ	1072	1637	1872	1833	2623	3995	6581

Таблица 3

Среднедушевые денежные расходы населения
(в месяц, рублей)

	1999	2000	2001	2002	2003	2004	2005
Центральный федеральный округ	819	1297	1551	1679	2711	3626	4834
Белгородская область	332	502	559	631	1136	1471	1949
Брянская область	290	419	461	464	725	1086	1480
Владимирская область	272	398	461	496	816	1132	1564
Воронежская область	315	428	617	738	1125	1426	2005
Ивановская область	254	361	426	439	669	953	1271
Калужская область	382	472	511	588	911	1259	1675
Костромская область	333	413	463	481	810	1199	1690
Курская область	259	379	471	535	922	1225	1662
Липецкая область	309	494	541	638	1073	1536	1946
Московская область	461	534	680	851	1491	1952	2690
Орловская область	332	489	538	598	998	1309	1735
Рязанская область	281	391	465	501	855	1140	1669
Смоленская область	322	449	549	616	1089	1521	2082
Тамбовская область	246	358	440	548	946	1287	1697
Тверская область	300	390	434	505	840	1086	1428
Тульская область	310	427	531	544	906	1228	1693
Ярославская область	414	556	649	734	1202	1609	2227
г. Москва	2415	4125	4877	5154	8145	10826	14207

	1999	2000	2001	2002	2003	2004	2005
Северо-Западный федеральный округ	519	740	876	979	1536	2087	2836
Республика Карелия	536	693	745	792	1276	1708	2303
Республика Коми	446	632	822	1036	1566	2245	3521
Архангельская область	383	485	549	613	1014	1584	2365
Вологодская область	414	553	670	682	1090	1617	2232
Калининградская область	421	582	732	833	1513	2066	2201
Ленинградская область	355	472	499	575	923	1229	1732
Мурманская область	582	840	1047	1239	1956	2647	3523
Новгородская область	376	567	631	827	1199	1590	2128
Псковская область	282	359	450	499	836	1240	1624
г. Санкт-Петербург	735	1105	1311	1422	2169	2820	3775
Южный федеральный округ	261	421	526	571	987	1386	1894
Республика Адыгея	239	340	425	476	798	1103	1367
Республика Дагестан	94	162	230	289	467	675	923
Кабардино-Балкарская Республика	193	263	318	387	697	927	1349
Республика Калмыкия	141	202	269	303	442	604	820
Карачаево-Черкесская Республика	177	261	344	363	613	880	1088
Республика Северная Осетия - Алания	239	361	405	499	935	1302	1596
Краснодарский край	335	551	649	682	1188	1675	2286
Ставропольский край	302	507	605	640	1094	1569	2089
Астраханская область	253	403	504	582	1009	1490	2022
Волгоградская область	278	435	588	560	943	1284	1777
Ростовская область	270	437	575	663	1193	1687	2351
Приволжский федеральный округ	323	489	619	692	1173	1576	2138
Республика Башкортостан	286	440	549	642	1139	1511	2123
Республика Марий Эл	227	306	340	357	675	893	1206
Республика Мордовия	225	315	382	422	686	883	1183

Продолжение табл. 3

Приволжский федеральный округ	1999	2000	2001	2002	2003	2004	2005
Республика Татарстан	319	432	573	665	1107	1556	2169
Удмуртская Республика	270	506	592	586	873	1275	1723
Чувашская Республика	283	390	432	452	774	1074	1457
Кировская область	357	498	554	577	941	1300	1736
Нижегородская область	335	485	604	700	1206	1703	2344
Оренбургская область	237	382	470	507	837	1115	1487
Пензенская область	269	384	462	482	865	1179	1637
Пермская область	404	626	787	851	1387	1965	2771
Самарская область	516	869	1190	1376	2432	30Ю	3839
Саратовская область	260	345	471	522	916	1209	1624
Ульяновская область	289	459	580	643	920	1173	1551
Уральский федеральный округ	473	686	878	880	1493	2178	3086
Курганская область	207	286	370	418	754	1060	1432
Свердловская область	470	696	860	858	1459	2178	3145
Тюменская область	718	1044	1343	1353	2228	3153	4599
Челябинская область	348	486	651	633	1113	1654	2160
Сибирский федеральный округ	414	609	758	755	1221	1700	2298
Республика Алтай	190	285	346	334	536	767	1075
Республика Бурятия	323	408	623	558	981	1413	2012
Республика Тыва	193	262	295	341	454	636	906
Республика Хакасия	370	597	663	644	1116	1682	2259
Алтайский край	296	358	426	484	852	1170	1585
Красноярский край	456	747	907	944	1557	2234	2962
Иркутская область	472	656	812	899	1463	1946	2382
Кемеровская область	538	632	792	765	1142	1670	2370
Новосибирская область	443	900	1155	988	1657	2144	2840
Омская область	382	616	759	773	1130	1622	2378
Томская область	383	513	714	731	1211	1689	2319
Читинская область	341	398	413	394	620	920	1340

Продолжение табл. 3

	1999	2000	2001	2002	2003	2004	2005
Дальневосточный федеральный округ	509	763	941	981	1560	2116	2886
Республика Саха (Якутия)	626	952	1222	1229	1971	2555	3340
Приморский край	490	718	905	954	1428	1888	2561
Хабаровский край	463	721	905	1012	1716	2358	3275
Амурская область	416	636	771	745	1209	1585	2061
Камчатская область	698	1071	1279	1247	2066	2983	3898
Магаданская область	680	966	1105	1228	1760	2495	3464
Сахалинская область	522	749	846	852	1420	2112	3082
Еврейская автономная область	325	437	477	539	874	1248	1707
Чукотский автономный округ	493	845	992	1075	1357	1971	4419

4. ПРИМЕР ВЫПОЛНЕНИЯ КОНТРОЛЬНОЙ РАБОТЫ

В таблице 4 представлены среднедушевые денежные доходы и расходы населения Северо-Западного федерального округа (СЗФО) в 2005 г.

Таблица 4

Доходы и расходы населения СЗФО в 2005 г.

	X	Y	\hat{Y}	e
	среднедушевые денежные доходы населения, руб.	среднедушевые денежные расходы населения, руб.	предсказанные значения Y	остатки
Карелия	2896	2303	2614	-311
Коми	4092	3521	3564	-43
Архангельская	2757	2365	2504	-139
Вологодская	2471	2232	2276	-44
Калининградская	2003	2201	1905	296
Ленинградская	1824	1732	1762	-30
Мурманская	4250	3523	3690	-167
Новгородская	2332	2128	2166	-38
Псковская	1686	1624	1652	-28
Санкт-Петербург	3720	3775	3269	506

Коэффициент корреляции (r) между среднедушевыми денежными доходами населения (X) и среднедушевыми денежными расходами населения (Y) равен 0,95 и вычисляется с помощью функции КОРРЕЛ(...).

Уравнение регрессии имеет вид: $\hat{y} = a + bx$,
 где $a = 312,65$ — вычисляется с помощью функции ОТРЕЗОК(...);
 $b = 0,79$ — вычисляется с помощью функции НАКЛОН(...).

Предсказанные значения \hat{y} вычисляются в соответствии с уравнением регрессии, остатки — по формуле: $e_i = y_i - \hat{y}_i$.

Гистограмма доходов и расходов населения СЗФО приведена на рис. 1. Наибольшие среднедушевые доходы у населения Мурманской области, наименьшие — в Псковской. Наибольшие среднедушевые расходы — в Санкт-Петербурге, наименьшие — в Псковской области. В Калининградской области и в Санкт-Петербурге среднедушевые денежные расходы превышают доходы, что может свидетельствовать о существовании теневых доходов населения.



Рис. 1. Среднедушевые денежные доходы и расходы населения Северо-Западного федерального округа в 2005 г.

Диаграмма рассеивания представлена в виде точек на рис. 2. Визуально линейная зависимость между переменными существует. Коэффициент корреляции между X и Y $r = 0,95$, что свидетельствует о наличии линейной зависимости между X и Y . Зависимость — чем больше доходы (X), тем больше расходы (Y). Уравнение для оценки регрессионной модели имеет вид: $y = 312,65 + 0,79x$.

Уравнение регрессии строится с помощью опции "Добавить линию тренда", возникающей при выделении точек на диаграмме и нажатии правой кнопки мыши.

Отображение на графике уравнения и R^2 осуществляется установкой соответствующих отметок на вкладке "Формат линии тренда", "Параметры".

Предсказанные значения Y приведены в столбце 4, остатки — в столбце 5 таблицы 4. Коэффициент детерминации, близкий к 1 ($R^2 = 0.91$), свидетельствует об адекватности линейной модели.

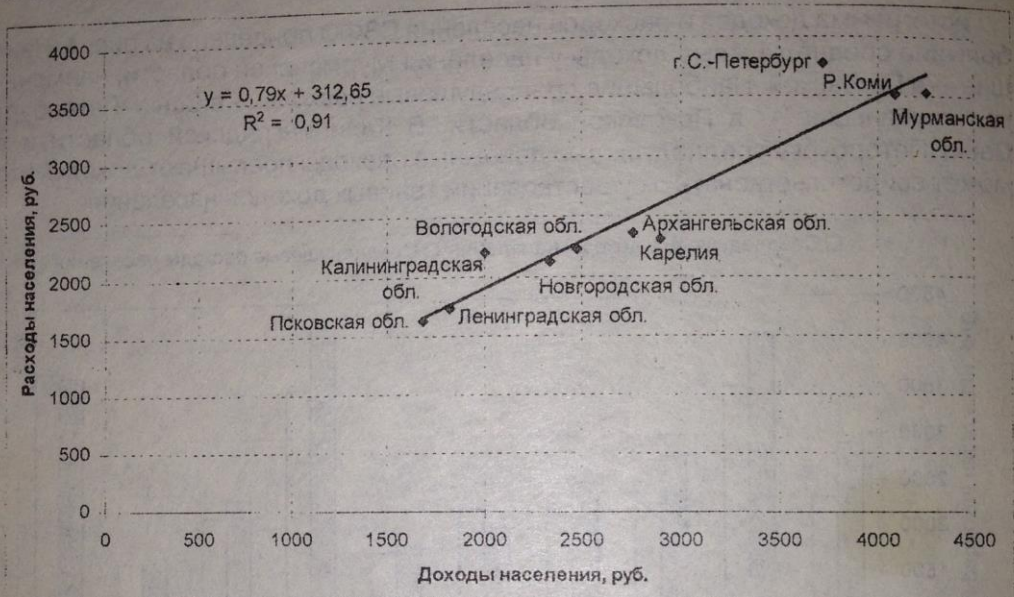


Рис. 2. Регрессионная зависимость между среднедушевыми денежными доходами и расходами населения Северо-Западного федерального округа в 2005 г.

Из рисунка 2 видно, что наибольшие отклонения от линии регрессии наблюдаются для Калининградской области и Санкт-Петербурга (превышение фактических значений среднедушевых расходов над прогнозируемыми значениями), а также для Республики Карелии и Мурманской области (значения среднедушевых денежных расходов ниже среднего уровня, определяемого регрессионной моделью). При приросте среднедушевых денежных доходов населения на 100 рублей можно ожидать возрастания среднедушевых расходов на 79 рублей.

По результатам анализа графика на рис. 3 можно сделать вывод, что остатки гомоскедастичны, автокорреляции нет.

На рис. 4 приведены результаты регрессионного анализа с помощью модуля "Анализ данных" – "Регрессия".

Нулевая гипотеза, H_0 : изменение среднедушевых доходов не влияет на среднедушевые расходы населения, т.е. коэффициент регрессии равен 0.

Альтернативная гипотеза, H_1 : изменение среднедушевых доходов влияет на среднедушевые расходы населения, т.е. коэффициент регрессии не равен 0.

См. справ
Excel

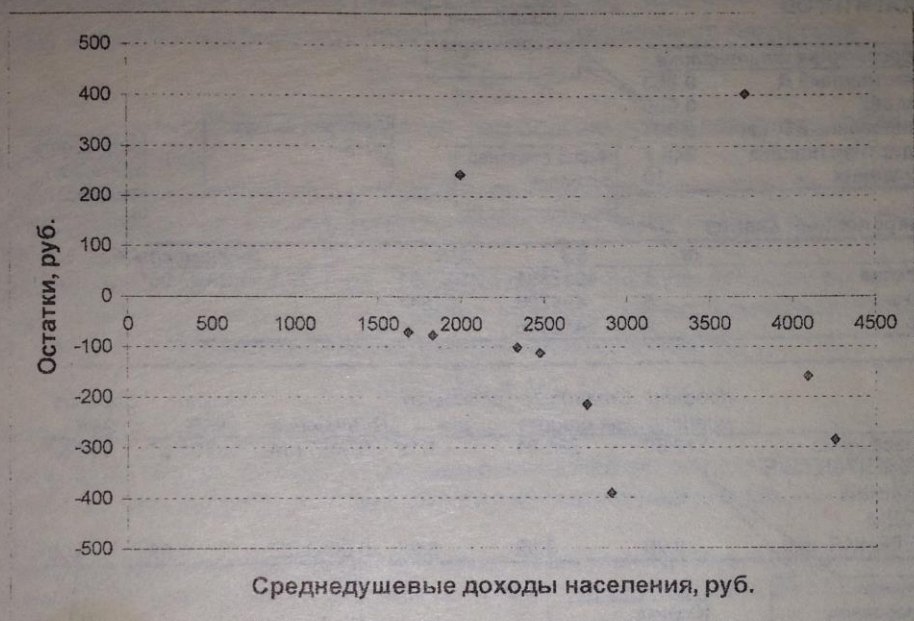


Рис. 3. График остатков

Критерий Фишера, $F = 79.5$, значимость критерия Фишера $0.00001983 < 0.05$, следовательно, в результате F-теста мы должны отклонить нулевую гипотезу и признать справедливость альтернативной гипотезы.

P -значение = $0.0000198 < 0.05$, следовательно, при уровне значимости 0.05 нулевая гипотеза может быть отвергнута. Справедлива гипотеза о регрессионной зависимости между переменными. Доверительный интервал: $0.59 < b < 1.00$.

Поясним содержимое таблицы дисперсионного анализа для простой линейной регрессии. Дисперсию оценки можно найти из таблицы дисперсионного анализа (см. табл. 5), которую можно получить стандартными средствами Excel (Сервис – Анализ данных – Регрессия).

Величина s^2 идентична среднему квадрату отклонения (остатка) от регрессии MS_R . Она вычисляется как отношение остаточной суммы квадратов SS_R к остаточному числу степеней свободы v_R .

Регрессионная статистика		Критерий Фишера (F-тест)		Значимость критерия Фишера	
Множественный R	0.953	79.5	1.983E-05		
R-квадрат	0.909				
Нормированный R-квэ	0.897				
Стандартная ошибка	248.7				
Наблюдения	10				

Дисперсионный анализ					
	df	SS	MS	F	Значимость F
Регрессия	1	4917391	4917391	79.5	1.983E-05
Остаток	8	494726	61841		
Итого	9	5412116			

	t-		P-Значение	Нижние 95%	Верхние 95%	
	Кoeffициенты	Стандартная ошибка				
У-пересечение	312.65	261.91	1.19	0.2667714	-291.31	916.62
СРЕДНЕДУШЕВЫЕ ДЕНЕЖНЫЕ ДОХОДЫ НАСЕЛЕНИЯ, руб.	0.79	0.09	8.92	0.0000198	0.59	1.00

Кoeffициент детерминации
 Число степеней свободы
 Оценка свободного члена "а"
 Оценка коэффициента регрессии "b"
 Значимость оценки коэффициента регрессии "b"
 Доверительный интервал для коэффициента регрессии

Отрезок 951,45
каждой 0,58

Рис. 4. Результаты регрессионного анализа

Таблица 5
Таблица дисперсионного анализа для простой линейной регрессии

Источник дисперсии	Сумма квадратов	Число степеней свободы	Средний квадрат	F-отношение
Регрессия	$SS_D = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$	$\nu_D = 1$	$MS_D = SS_D / \nu_D$	$F_0 = MS_D / MS_R$
Отклонение от регрессии	$SS_R = \sum_{i=1}^n (y_i - \hat{y}_i)^2$	$\nu_R = n - 2$	$MS_R = SS_R / \nu_R \equiv s^2$	-
Полная дисперсия	$SS_T = \sum_{i=1}^n (y_i - \bar{y})^2$	$\nu_T = n - 1$	-	-

Обусловленная регрессией сумма квадратов SS_D получила такое название потому, что ее можно выразить через оценку коэффициента регрессии:

$$SS_D = b^2 \sum_{i=1}^n (x_i - \bar{x})^2.$$

Чем больше коэффициент регрессии, тем больше сумма квадратов, "обусловленная регрессией".

Последняя колонка таблицы, называемая F -отношение, может быть использована для проверки гипотез, если ошибки предполагаются нормально распределенными.

Для проверки гипотезы о том, что простая линейная регрессия y по x отсутствует, то есть гипотезы $H_0: \beta = 0$ против альтернативной гипотезы $H_1: \beta \neq 0$, используется F -отношение из таблицы дисперсионного анализа:

$$F_0 = MS_D / MS_R \equiv MS_D / S^2.$$

Если верна нулевая гипотеза, то F_0 имеет F -распределение с $\nu_D = 1$ и $\nu_R = n - 2$ степенями свободы. Соответствующее критической области p -значение вычисляется стандартными средствами обработки статистических данных автоматически.

Мы отвергаем нулевую гипотезу, если p -значение меньше, чем уровень значимости α . В этом случае мы говорим о том, что регрессионная зависимость существует.