

**Контрольная работа по эконометрике
Вариант 29**

Задание:

1. Выбрать данные из файла Excel «Данные задание 4» по своему номеру в списке группы. Сформировать две таблицы: с данными по «развивающимся» странам и с данными по «развитым» странам.
2. *(Развернутый ответ)* В теории, какая связь ожидается между показателями? Где связь будет теснее: среди развитых или развивающихся стран? Почему?
3. Построить графики взаимного разброса Y и X в Excel (точечный график разброса). График должен быть оформлен корректно.
4. Рассчитать значения коэффициентов корреляции вручную.
5. Рассчитать коэффициенты корреляции с помощью пакета «Анализ данных». Убедиться, что значения коэффициентов, полученные разными способами, совпадают. Интерпретировать значения коэффициентов.
6. Оценить значимость коэффициента корреляции. Критическое значение соответствующей статистики определить с помощью встроенных функций MS Excel.
7. *(Развернутый ответ)* На основе проведенного анализа сделать вывод: на какой группе – развитых или развивающихся стран – следует строить модель парной линейной регрессии? Совпали ли результаты расчетов с предположениями, выдвинутыми в начале? Если нет, то почему? Влияют ли на результаты аномальные значения? Стоит ли очистить данные от выбросов? Если да, то необходимо исключить аномальные значения и дальше работать на сокращенной выборке.
Записать модель парной линейной регрессии в общем виде. Насколько высокое качество модели стоит ожидать?
8. Записать систему нормальных уравнений для нахождения коэффициентов модели парной линейной регрессии по выбранной группе. Решить систему.
9. По выбранной группе построить модель парной линейной регрессии с помощью пакета «Анализ данных». Убедиться в том, что значения коэффициентов равны значениям, полученным в предыдущем задании.
Записать модель парной линейной регрессии в явном виде.
10. Оценить значимость модели в целом, значимость отдельных коэффициентов. Записать проверяемые гипотезы. Критические значения соответствующих статистик определить с помощью встроенных функций MS Excel.
11. Рассчитать значение коэффициента эластичности.
12. *(Развернутый ответ)* Оценить качество построенной модели. Что показывает коэффициент детерминации? Интерпретировать значения коэффициентов полученной модели. Интерпретировать полученное значение коэффициента эластичности.

Решение

1. Выберем данные по варианту за 2014 г. и получим таблицы:

Развитые страны		Численность заболевших корью, чел.	Численность населения, чел.
№	Страна	Показатель Y	Показатель X
1	Australia	340	2346694
2	Austria	117	8541575
3	Belgium	70	112957
4	Cyprus	10	115239
5	Denmark	27	5643475
6	Germany	443	89825
7	Ireland	33	4617225
8	Japan	462	127276
9	Latvia	36	1993782
10	Lithuania	11	2932367
11	Netherlands	140	168658
12	Norway	3	5137232
13	Romania	59	1998979
14	Slovenia	52	26198
15	Spain	154	4648882
16	Sweden	26	969611
17	Switzerland	23	8188649
18	United Kingdom of Great Britain and Northern Ireland	133	65110000
19	United States of America	667	321 442 000

Развивающиеся страны		Численность заболевших корью, чел.	Численность населения, чел.
№	Страна	Показатель Y	Показатель X
1	Afghanistan	492	327582
2	Angola	11699	2692466
3	Argentina	1	42981515
4	Armenia	13	29622
5	Bangladesh	289	15945279
6	Belarus	64	9474511
7	Benin	786	1286712
8	Bosnia and Herzegovina	3000	35662
9	Botswana	1	2168573
10	Brazil	876	24213133
11	Burkina Faso	343	17585977
12	Cameroon	831	2223994
13	Canada	418	35544564
14	Central African Republic	210	4515392
15	Chad	1275	13569438

Контрольная работа по эконометрике выполнена в www.MatBuro.ru

©**МатБюро** – Консультации по математике, экономике, праву, другим предметам

Поможем с заданиями по эконометрике: https://www.matburo.ru/sub_subject.php?p=eckr

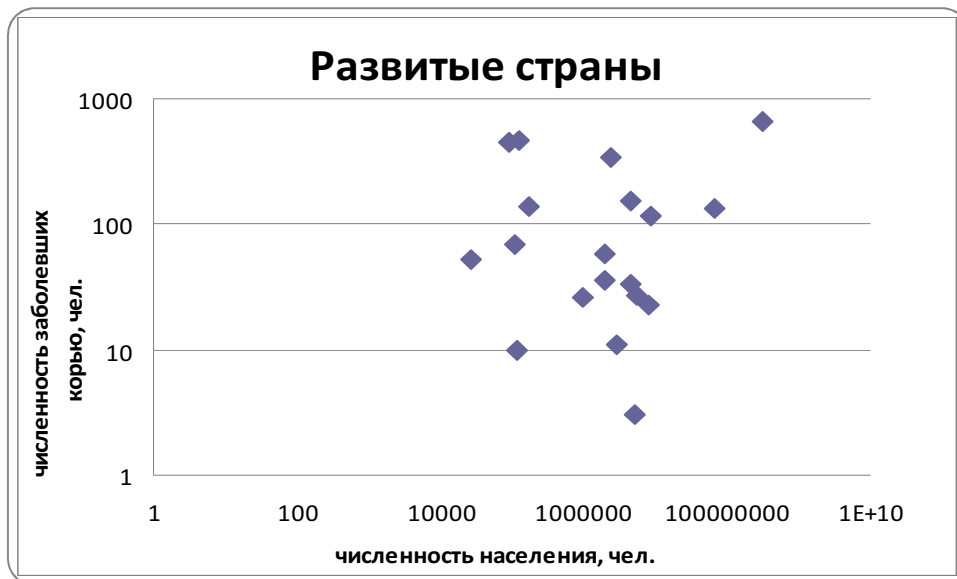
16	China	52628	136427
17	Costa Rica	1	4757575
18	Equatorial Guinea	13	1129424
19	Ethiopia	12739	97366774
20	France	267	66331957
21	Gabon	33	1875713
22	Georgia	3188	3727
23	Ghana	124	26962563
24	Greece	1	1892413
25	Guinea	175	118559
26	Guinea-Bissau	1	1725744
27	India	79563	1293859294
28	Indonesia	12943	255131116
29	Iraq	1317	3568
30	Israel	6	82157
31	Jordan	20	88936
32	Kazakhstan	321	17289224
33	Kenya	354	462425
34	Lebanon	112	563279
35	Liberia	34	439737
36	Madagascar	3	2358981
37	Malawi	3	1768838
38	Malaysia	221	322817
39	Mali	290	16962846
40	Mauritania	14	46392
41	Mexico	3	1242216
42	Morocco	10	3431882
43	Mozambique	9	27212382
44	Myanmar	122	51924182
45	Namibia	477	237992
46	Nepal	1279	28323241
47	Niger	1142	19148219
48	Nigeria	6855	1764652
49	Pakistan	1370	185546257
50	Papua New Guinea	2299	7755785
51	Philippines	58848	112249
52	Russian Federation	4711	143819666
53	Rwanda	10	11345357
54	Senegal	32	14546111
55	Serbia	37	713576
56	Sierra Leone	1006	779162
57	South Africa	66	54146734,7
58	South Sudan	441	1153971
59	Sri Lanka	1686	2771
60	Sudan	676	37737913
61	Syrian Arab Republic	594	19239

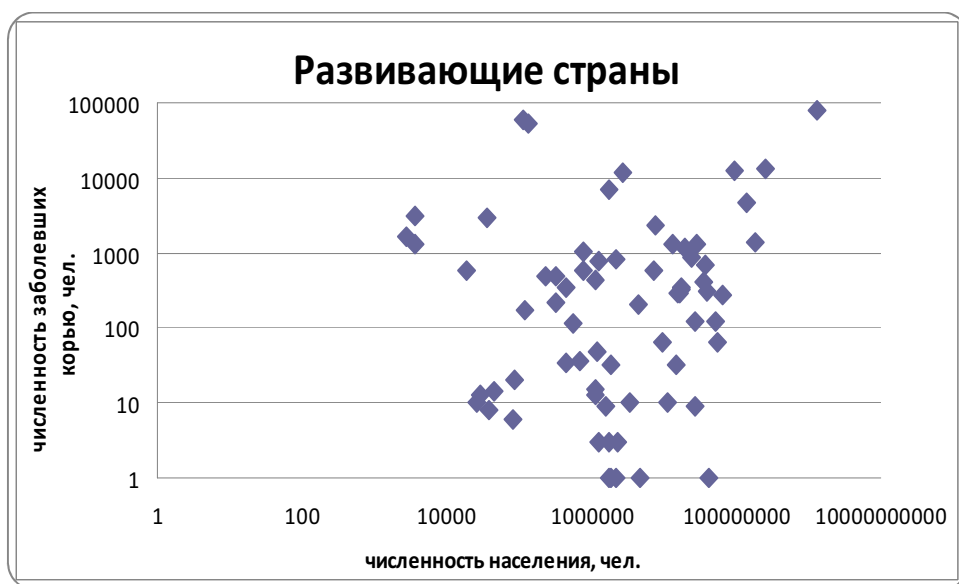
Поможем с заданиями по эконометрике: https://www.matburo.ru/sub_subject.php?p=eckr

62	Timor-Leste	47	1212814
63	Togo	577	7228915
64	Tunisia	15	1114398
65	Turkey	565	773628
66	Uganda	314	38833338
67	Uzbekistan	8	37577
68	Vanuatu	10	25885
69	Zambia	9	1562974

2. Теоретически ожидается, что чем больше численность населения, то тем больше будет число заболевших корью, то есть связь между показателями прямая. В развивающихся странах условия жизни населения хуже, чем в развитых странах, в том числе ниже уровень развития медицины и охват населения вакцинацией. Поэтому можно предположить, что заболеваемость корью в развивающихся странах будет носить систематический характер, а в развитых странах - случайный. Соответственно связь между численностью населения и количеством заболевших в развивающихся странах будет теснее.

3. Построим графики. Вследствие большого разброса показателей, целесообразно представить их в логарифмическом масштабе:





По виду графиков нельзя сделать однозначного вывода о положительной или отрицательной связи между численностью населения и количеством заболевших корью.

4. Коэффициенты корреляции определяются по формуле:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sqrt{\overline{x^2} - (\bar{x})^2} \cdot \sqrt{\overline{y^2} - (\bar{y})^2}}$$

Расчётные таблицы представлены в файле MS Excel. По расчётным данным рассчитаем вручную коэффициенты корреляции.

Для развитых стран:

$$r_{xy} = \frac{1,192E+10 - 22853190,74 \cdot 147,68}{\sqrt{5,675E+15 - (22853190,74)^2} \cdot \sqrt{55816,32 - (147,68)^2}} = 0,645$$

Для развивающихся стран:

$$r_{xy} = \frac{1,58E+12 - 37768521,63 \cdot 3882,42}{\sqrt{2,64E+16 - (37768521,63)^2} \cdot \sqrt{190444098,6 - (3882,42)^2}} = 0,682$$

5. Рассчитаем значения коэффициентов корреляции с помощью функции КОРРЕЛ, получим для развитых стран 0,645; для развивающихся стран 0,682. Значения коэффициентов, полученные разными способами, совпадают. Оба коэффициента показывают, что связь между численностью населения и числом заболевших корью прямая по направлению и заметная по тесноте. При этом, как и предполагалось ранее, связь между показателями теснее в развивающихся странах, чем в развитых.

6. Стандартная ошибка коэффициента корреляции рассчитывается по формуле:

$$m_r = \sqrt{\frac{1 - r^2}{n - p - 1}}, \text{ где}$$

p – число факторов в модели.

Поможем с заданиями по эконометрике: https://www.matburo.ru/sub_subject.php?p=eckr

Для развитых стран:
$$m_r = \sqrt{\frac{1 - 0,645^2}{19 - 1 - 1}} = 0,185$$

Для развивающихся стран:
$$m_r = \sqrt{\frac{1 - 0,682^2}{69 - 1 - 1}} = 0,089$$

Фактическое значение t-критерия рассчитывается по формуле:

$$t_r = \frac{r}{m_r}$$

Для развитых стран:
$$t_r = \frac{0,645}{0,185} = 3,48$$

Для развивающихся стран:
$$t_r = \frac{0,682}{0,089} = 7,63$$

Табличное значение t-критерия Стьюдента при $\alpha = 0,05$ и $k = 19 - 1 - 1 = 17$ будет равно $t_{\text{табл}} = 2,11$; при $k = 69 - 1 - 1 = 67$ табличное значение $t_{\text{табл}} = 1,996$ (по функции СТЬЮДРАСПОБР). Таким образом, оба коэффициента корреляции оказались значимы, так как расчётные значения t-критерия выше табличных значений.

7. Модель парной линейной регрессии следует строить по группе развивающихся стран, так как коэффициент корреляции и фактическое значение t-критерия в этой группе больше, чем в группе развитых стран. Результаты расчётов совпали с предположениями, выдвинутыми в начале – в развивающихся странах вследствие низкого уровня развития медицины заболеваемость кори выше, чем в развитых странах; она носит систематический характер и имеет более тесную связь с численностью населения. Аномальные значения не влияют на результат, что можно видеть на графиках, построенных в логарифмическом масштабе. Данные не имеет смысла очищать от выбросов.

Модель парной линейной регрессии имеет вид:

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

Качество модели ожидается среднее, так как коэффициент корреляции показывает среднюю тесноту связи.

8. Запишем систему нормальных уравнений:

$$\begin{cases} a \cdot n + b \cdot \sum_{i=1}^n x_i = \sum_{i=1}^n y_i; \\ a \cdot \sum_{i=1}^n x_i + b \cdot \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i \cdot y_i \end{cases}$$

Система нормальных уравнений с вычисленными коэффициентами:

$$\begin{cases} 69 \cdot a + 2606027993 \cdot b = 267887; \\ 2606027993 \cdot a + 1,825E+18 \cdot b = 1,087E+14 \end{cases}$$

Решим уравнение матричным способом. Найдём обратную матрицу X с помощью функции MS Excel МОБР:

Поможем с заданиями по эконометрике: https://www.matburo.ru/sub_subject.php?p=eckr

$$X^{-1} = \begin{pmatrix} 69 & 2606027993 \\ 2606027993 & 1,825E+18 \end{pmatrix}^{-1} = \begin{pmatrix} 0,0153 & -2,187E-11 \\ -2,187E-11 & 5,792E-19 \end{pmatrix}$$

Найдём вектор коэффициентов уравнения А:

$$A = X^{-1} \cdot Y = \begin{pmatrix} 0,0153 & -2,187E-11 \\ -2,187E-11 & 5,792E-19 \end{pmatrix} \cdot \begin{pmatrix} 267887 \\ 1,087E+14 \end{pmatrix} = \begin{pmatrix} 1725,85 \\ 5,71E-05 \end{pmatrix}$$

9. Построим модель регрессии с помощью надстройки Анализ данных:

ВЫВОД ИТОГОВ						
<i>Регрессионная статистика</i>						
Множественный R	0,682065702					
R-квадрат	0,465213622					
Нормированный R-квадрат	0,457231735					
Стандартная ошибка	9827,795065					
Наблюдения	69					
<i>Дисперсионный анализ</i>						
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>	
Регрессия	1	5629360636	5629360636	58,28366972	1,09782E-10	
Остаток	67	6471232241	96585555,83			
Итого	68	12100592877				
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
Y-пересечение	1725,85444	1216,382421	1,418841978	0,160580415	-702,055005	4153,763885
Переменная X 1	5,70996E-05	7,47927E-06	7,634374219	1,09782E-10	4,21709E-05	7,20282E-05

Очевидно, что значения коэффициентов совпадают с полученными ранее.

Уравнение регрессии имеет следующий вид:

$$\hat{y}_x = 1725,85 + 0,0000571x$$

10. Гипотеза H_0 : связь между y и x несущественна.

Фактическое значение F-критерия Фишера составляет 58,28. По таблице F-распределения Снедекора-Фишера при $\alpha = 0,05$ и $K_1 = 1$, $K_2 = 69 - 2 = 67$ величина $F_T = 3,98$ (по функции FРАСПОБР). Это означает, что гипотеза H_0 о несущественности связи между y и x с вероятностью ошибочности суждения $\alpha = 0,05$ отклоняется, то есть связь между этими переменными может быть признана существенной.

Гипотеза H_1 : коэффициент b незначимо отличен от 0.

Гипотеза H_2 : свободный член a незначимо отличен от 0.

Фактические значения t-критерия для параметров a и b составляют соответственно 1,42 и 7,63; табличное – 1,996. Это означает, что коэффициент уравнения b статистически значим, а свободный член a – незначим. Гипотеза H_1 отклоняется, гипотеза H_2 принимается.

11. Коэффициент эластичности определим по формуле:

Поможем с заданиями по эконометрике: https://www.matburo.ru/sub_subject.php?p=eckr

$$\mathcal{E}_{yx} = b \cdot \frac{\bar{x}}{y} = 0,0000571 \cdot \frac{37768521,63}{3882,42} = 0,56\%$$

12. Качество построенной модели среднее. Коэффициент детерминации показывает, что 46,5% вариации численности заболевших корью объясняется вариацией численности населения, а остальные 53,5% - вариацией прочих факторов.

В смысле интерпретации полученных результатов регрессионного моделирования можно отметить следующее:

- при увеличении численности населения на 1 чел. от среднего значения численность заболевших корью в стране может увеличиться на 0,0000571 чел.
- в относительном выражении при увеличении численности населения на 1% рост численности заболевших корью возможен на 0,56%.