Лабораторная работа №2

Полянцев И.С.

27 12 2020

# Раздел I.

Построить модели для Y, в т.ч. с фиктивными переменными по федеральным округам, довести их до состояния значимости всех параметров. Сравнить модели без поправки на p-значения и с поправкой на p-значения. В отчёт включить:

Таблицу ANOVA со сравнением значимых моделей. Таблицу с характеристиками качества моделей: F расчётное, средняя ошибка аппроксимации, скорректированный R-квадрат. Выбрать лучшую модель.

### Обозначения количественных показателей:

* *“Y.Wage”* – Среднедушевые денежные доходы населения
* *“X1.GRP”* – ВРП на душу населения
* *“X2.Retail”* – Оборот розничной торговли на душу населения
* *“X3.Social.Policy”* – Расходы консолидированных бюджетов субъектов Российской Федерации: на социальную политику на душу населения
* *“X4.Small.Businesses”* – Число малых предприятий на 10000 человек населения

##   
## Call:  
## lm(formula = Wage ~ GRP + Retail + Social.Policy + Small.Businesses,   
## data = reg.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8688.5 -1424.3 -206.1 1466.1 11733.2   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.572e+03 1.160e+03 1.355 0.179   
## GRP 4.748e-03 8.990e-04 5.281 1.08e-06 \*\*\*  
## Retail 7.894e-02 8.549e-03 9.234 3.05e-14 \*\*\*  
## Social.Policy 9.897e+02 8.737e+01 11.328 < 2e-16 \*\*\*  
## Small.Businesses 2.292e+00 6.120e+00 0.374 0.709   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2711 on 80 degrees of freedom  
## Multiple R-squared: 0.9278, Adjusted R-squared: 0.9242   
## F-statistic: 257.1 on 4 and 80 DF, p-value: < 2.2e-16

Модель значима (P-value < 0.05) и объясняет 92,42% разброса зависимой переменной с поправкой на степени свободы. Согласно процедуре последовательного исключения незначимых факторов, уберем из модели наименее значимый фактор. Это *Small.Businesses*

##   
## Call:  
## lm(formula = Wage ~ GRP + Retail + Social.Policy, data = reg.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -8520.5 -1407.8 -158.6 1379.5 11872.6   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.683e+03 1.116e+03 1.508 0.135   
## GRP 4.732e-03 8.932e-04 5.298 9.88e-07 \*\*\*  
## Retail 8.028e-02 7.716e-03 10.405 < 2e-16 \*\*\*  
## Social.Policy 9.904e+02 8.688e+01 11.399 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2696 on 81 degrees of freedom  
## Multiple R-squared: 0.9277, Adjusted R-squared: 0.925   
## F-statistic: 346.5 on 3 and 81 DF, p-value: < 2.2e-16

Теперь все параметры значимы и дальшейнее исключение связано с исключением значимых параметров, поэтому получим 3 варианта моделей только с одним параметром

##   
## Call:  
## lm(formula = Wage ~ GRP, data = reg.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12472.5 -2949.1 -964.5 1620.8 22485.4   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.764e+04 7.443e+02 23.70 <2e-16 \*\*\*  
## GRP 1.526e-02 1.114e-03 13.71 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 5484 on 83 degrees of freedom  
## Multiple R-squared: 0.6936, Adjusted R-squared: 0.6899   
## F-statistic: 187.9 on 1 and 83 DF, p-value: < 2.2e-16

##   
## Call:  
## lm(formula = Wage ~ Retail, data = reg.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9236 -3045 -1709 823 39586   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.019e+03 2.722e+03 0.742 0.46   
## Retail 1.560e-01 1.867e-02 8.354 1.28e-12 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 7301 on 83 degrees of freedom  
## Multiple R-squared: 0.4568, Adjusted R-squared: 0.4502   
## F-statistic: 69.79 on 1 and 83 DF, p-value: 1.279e-12

##   
## Call:  
## lm(formula = Wage ~ Social.Policy, data = reg.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11249 -2563 -289 2604 11428   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 9183.80 1002.56 9.16 3.12e-14 \*\*\*  
## Social.Policy 1606.79 95.06 16.90 < 2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4700 on 83 degrees of freedom  
## Multiple R-squared: 0.7749, Adjusted R-squared: 0.7722   
## F-statistic: 285.7 on 1 and 83 DF, p-value: < 2.2e-16

Добавим в каждую полученную модель фиктивные переменные, используя принадлежность каждого региона к одному из восьми федеральных округов.

##   
## Call:  
## lm(formula = Wage ~ GRP \* FO, data = reg.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -9607.0 -2015.7 -654.6 1591.0 11906.3   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.233e+04 2.304e+03 9.693 1.66e-14 \*\*\*  
## GRP 1.990e-02 3.487e-03 5.708 2.63e-07 \*\*\*  
## FOПФО -1.483e+04 4.237e+03 -3.500 0.000819 \*\*\*  
## FOСЗФО -2.298e+03 2.633e+03 -0.873 0.385754   
## FOСКФО -1.216e+04 7.279e+03 -1.670 0.099351 .   
## FOСФО -1.021e+04 3.994e+03 -2.557 0.012773 \*   
## FOУФО -4.242e+03 3.193e+03 -1.329 0.188356   
## FOЦФО -1.292e+04 2.859e+03 -4.519 2.50e-05 \*\*\*  
## FOЮФО -1.779e+04 6.736e+03 -2.641 0.010223 \*   
## GRP:FOПФО 2.841e-02 1.370e-02 2.075 0.041748 \*   
## GRP:FOСЗФО -8.151e-03 3.643e-03 -2.237 0.028491 \*   
## GRP:FOСКФО 3.283e-02 5.100e-02 0.644 0.521943   
## GRP:FOСФО 7.778e-03 1.244e-02 0.625 0.533776   
## GRP:FOУФО -5.561e-03 3.869e-03 -1.437 0.155156   
## GRP:FOЦФО 2.732e-02 6.139e-03 4.450 3.21e-05 \*\*\*  
## GRP:FOЮФО 4.556e-02 2.806e-02 1.624 0.108984   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3708 on 69 degrees of freedom  
## Multiple R-squared: 0.8835, Adjusted R-squared: 0.8582   
## F-statistic: 34.88 on 15 and 69 DF, p-value: < 2.2e-16

##   
## Call:  
## lm(formula = Wage ~ Retail \* FO, data = reg.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11356 -1840 -378 1008 35431   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.696e+04 9.510e+03 2.835 0.00601 \*\*  
## Retail 4.439e-02 6.357e-02 0.698 0.48741   
## FOПФО -2.320e+04 1.166e+04 -1.989 0.05062 .   
## FOСЗФО -4.247e+04 1.654e+04 -2.567 0.01242 \*   
## FOСКФО -1.582e+04 1.166e+04 -1.356 0.17939   
## FOСФО -1.932e+04 1.235e+04 -1.564 0.12242   
## FOУФО -3.524e+04 1.490e+04 -2.365 0.02083 \*   
## FOЦФО -2.907e+04 1.068e+04 -2.721 0.00822 \*\*  
## FOЮФО -2.196e+04 1.352e+04 -1.624 0.10886   
## Retail:FOПФО 7.933e-02 8.076e-02 0.982 0.32935   
## Retail:FOСЗФО 2.487e-01 1.109e-01 2.243 0.02810 \*   
## Retail:FOСКФО 1.439e-02 8.827e-02 0.163 0.87099   
## Retail:FOСФО 5.265e-02 9.016e-02 0.584 0.56119   
## Retail:FOУФО 1.719e-01 8.680e-02 1.980 0.05171 .   
## Retail:FOЦФО 1.257e-01 7.088e-02 1.774 0.08055 .   
## Retail:FOЮФО 5.999e-02 9.374e-02 0.640 0.52430   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6487 on 69 degrees of freedom  
## Multiple R-squared: 0.6435, Adjusted R-squared: 0.566   
## F-statistic: 8.303 on 15 and 69 DF, p-value: 2.283e-10

##   
## Call:  
## lm(formula = Wage ~ Social.Policy \* FO, data = reg.df)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -7199.2 -2434.0 -175.4 2419.1 10851.2   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 15358.6 3406.9 4.508 2.60e-05 \*\*\*  
## Social.Policy 1164.3 200.6 5.804 1.79e-07 \*\*\*  
## FOПФО -801.3 7066.5 -0.113 0.91004   
## FOСЗФО -10387.5 4293.8 -2.419 0.01820 \*   
## FOСКФО 4762.3 5841.0 0.815 0.41769   
## FOСФО -2009.3 7721.3 -0.260 0.79546   
## FOУФО -3304.0 4651.9 -0.710 0.47995   
## FOЦФО -8601.8 4073.1 -2.112 0.03832 \*   
## FOЮФО -19822.8 10932.1 -1.813 0.07414 .   
## Social.Policy:FOПФО -365.2 923.1 -0.396 0.69362   
## Social.Policy:FOСЗФО 871.8 288.6 3.020 0.00354 \*\*   
## Social.Policy:FOСКФО -1645.2 758.6 -2.169 0.03354 \*   
## Social.Policy:FOСФО -407.1 896.2 -0.454 0.65107   
## Social.Policy:FOУФО 310.9 280.8 1.107 0.27208   
## Social.Policy:FOЦФО 1058.9 340.5 3.109 0.00272 \*\*   
## Social.Policy:FOЮФО 2585.0 1650.2 1.566 0.12182   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4154 on 69 degrees of freedom  
## Multiple R-squared: 0.8538, Adjusted R-squared: 0.822   
## F-statistic: 26.86 on 15 and 69 DF, p-value: < 2.2e-16

Применяем функцию для исключения незначимых регрессоров к 1 модели

##   
## Call:  
## lm(formula = as.formula(paste(y.var.name, " ~ .", sep = "")),   
## data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -12618.5 -2364.5 -394.5 1669.5 10690.1   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.978e+04 9.986e+02 19.804 < 2e-16 \*\*\*  
## GRP 2.397e-02 2.213e-03 10.830 < 2e-16 \*\*\*  
## FOПФО -5.998e+03 1.306e+03 -4.591 1.74e-05 \*\*\*  
## FOСКФО -5.783e+03 1.667e+03 -3.469 0.000870 \*\*\*  
## FOСФО -6.699e+03 1.369e+03 -4.892 5.55e-06 \*\*\*  
## FOЦФО -1.037e+04 1.985e+03 -5.221 1.54e-06 \*\*\*  
## FOЮФО -6.074e+03 1.748e+03 -3.475 0.000853 \*\*\*  
## GRP.FOСЗФО -1.210e-02 2.195e-03 -5.514 4.75e-07 \*\*\*  
## GRP.FOУФО -1.062e-02 2.232e-03 -4.757 9.30e-06 \*\*\*  
## GRP.FOЦФО 2.325e-02 5.579e-03 4.168 8.14e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3759 on 75 degrees of freedom  
## Multiple R-squared: 0.8699, Adjusted R-squared: 0.8543   
## F-statistic: 55.71 on 9 and 75 DF, p-value: < 2.2e-16

Аналогично для 2 модели:

##   
## Call:  
## lm(formula = as.formula(paste(y.var.name, " ~ .", sep = "")),   
## data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -11281 -1964 -374 1022 35431   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.045e+04 4.403e+03 4.645 1.47e-05 \*\*\*  
## Retail 8.908e-02 2.648e-02 3.363 0.001229 \*\*   
## FOПФО -1.215e+04 2.748e+03 -4.423 3.34e-05 \*\*\*  
## FOСЗФО -3.596e+04 1.399e+04 -2.570 0.012216 \*   
## FOСКФО -1.242e+04 3.404e+03 -3.647 0.000494 \*\*\*  
## FOСФО -1.186e+04 2.889e+03 -4.104 0.000105 \*\*\*  
## FOУФО -2.873e+04 1.208e+04 -2.378 0.020043 \*   
## FOЦФО -2.256e+04 6.496e+03 -3.474 0.000867 \*\*\*  
## FOЮФО -1.339e+04 3.369e+03 -3.976 0.000163 \*\*\*  
## Retail.FOСЗФО 2.040e-01 9.299e-02 2.194 0.031409 \*   
## Retail.FOУФО 1.272e-01 6.376e-02 1.995 0.049815 \*   
## Retail.FOЦФО 8.102e-02 4.059e-02 1.996 0.049637 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 6365 on 73 degrees of freedom  
## Multiple R-squared: 0.6369, Adjusted R-squared: 0.5821   
## F-statistic: 11.64 on 11 and 73 DF, p-value: 3.498e-12

Аналогично для 3 модели:

##   
## Call:  
## lm(formula = as.formula(paste(y.var.name, " ~ .", sep = "")),   
## data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -10603.2 -2974.0 -587.2 2807.7 11489.4   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8525.55 986.19 8.645 3.66e-13 \*\*\*  
## Social.Policy 1608.03 91.03 17.664 < 2e-16 \*\*\*  
## Social.Policy.FOЦФО 419.16 143.65 2.918 0.00455 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 4501 on 82 degrees of freedom  
## Multiple R-squared: 0.7961, Adjusted R-squared: 0.7911   
## F-statistic: 160 on 2 and 82 DF, p-value: < 2.2e-16

Используем функцию anova(), чтобы проверить гипотезы об эквивалентности построенных моделей.

## Analysis of Variance Table  
##   
## Model 1: Wage ~ GRP  
## Model 2: Wage ~ GRP + FOПФО + FOСКФО + FOСФО + FOЦФО + FOЮФО + GRP.FOСЗФО +   
## GRP.FOУФО + GRP.FOЦФО  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 83 2495737256   
## 2 75 1059714184 8 1.436e+09 12.704 2.323e-11 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: Wage ~ Retail  
## Model 2: Wage ~ Retail + FOПФО + FOСЗФО + FOСКФО + FOСФО + FOУФО + FOЦФО +   
## FOЮФО + Retail.FOСЗФО + Retail.FOУФО + Retail.FOЦФО  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 83 4424317256   
## 2 73 2957695768 10 1466621488 3.6198 0.0006018 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: Wage ~ Social.Policy  
## Model 2: Wage ~ Social.Policy + Social.Policy.FOЦФО  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 83 1833547842   
## 2 82 1661076088 1 172471754 8.5142 0.004546 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Видно, что для двух рассматриваемых факторов добавление фиктивных переменных уменьшает остаточную дисперсию (RSS). Составим список моделей

## Модель R.2.скорр F.расч Станд.Ошибка  
## 1 fit.X1 0.690 187.86 5483.5  
## 2 fit.X1.fo 0.854 55.71 3758.9  
## 3 fit.X2 0.450 69.79 7301.0  
## 4 fit.X2.fo 0.582 11.64 6365.2  
## 5 fit.X3 0.772 285.69 4700.1  
## 6 fit.X3.fo 0.791 160.03 4500.8

Лучшей моделью является модель 2, т.к. она имеет наименьшую стандартную ошибку

# Раздел II.

##   
## Call:  
## lm(formula = Wage ~ GRP + Retail + Social.Policy + Small.Businesses,   
## data = DFF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.29192 -0.06439 -0.01727 0.05351 0.32789   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.41161 0.45965 5.247 1.24e-06 \*\*\*  
## GRP 0.19847 0.03633 5.463 5.14e-07 \*\*\*  
## Retail 0.41064 0.05371 7.645 3.99e-11 \*\*\*  
## Social.Policy 0.28506 0.04462 6.389 1.03e-08 \*\*\*  
## Small.Businesses -0.07055 0.03293 -2.143 0.0352 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1095 on 80 degrees of freedom  
## Multiple R-squared: 0.8969, Adjusted R-squared: 0.8917   
## F-statistic: 173.9 on 4 and 80 DF, p-value: < 2.2e-16

Модель значима (P-value < 0.05) и объясняет 89,17% разброса зависимой переменной с поправкой на степени свободы. Согласно процедуре последовательного исключения незначимых факторов, уберем из модели наименее значимый фактор. Это *Small.Businesses*

##   
## Call:  
## lm(formula = Wage ~ GRP + Retail + Social.Policy, data = DFF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.31890 -0.06896 -0.01010 0.05342 0.32355   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.61670 0.45944 5.695 1.91e-07 \*\*\*  
## GRP 0.18906 0.03685 5.130 1.94e-06 \*\*\*  
## Retail 0.37475 0.05215 7.186 2.95e-10 \*\*\*  
## Social.Policy 0.28480 0.04560 6.246 1.84e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1119 on 81 degrees of freedom  
## Multiple R-squared: 0.8909, Adjusted R-squared: 0.8869   
## F-statistic: 220.6 on 3 and 81 DF, p-value: < 2.2e-16

Теперь все параметры значимы и дальшейнее исключение связано с исключением значимых параметров, поэтому получим 3 варианта моделей только с одним параметром

##   
## Call:  
## lm(formula = Wage ~ GRP, data = DFF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.36560 -0.07529 -0.00951 0.08024 0.36103   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.34878 0.31488 13.81 <2e-16 \*\*\*  
## GRP 0.45011 0.02498 18.02 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.151 on 83 degrees of freedom  
## Multiple R-squared: 0.7964, Adjusted R-squared: 0.7939   
## F-statistic: 324.6 on 1 and 83 DF, p-value: < 2.2e-16

##   
## Call:  
## lm(formula = Wage ~ Retail, data = DFF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.28573 -0.13407 -0.06417 0.04864 1.01452   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.05530 0.90859 1.161 0.249   
## Retail 0.75936 0.07698 9.864 1.23e-15 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.227 on 83 degrees of freedom  
## Multiple R-squared: 0.5397, Adjusted R-squared: 0.5341   
## F-statistic: 97.3 on 1 and 83 DF, p-value: 1.227e-15

##   
## Call:  
## lm(formula = Wage ~ Social.Policy, data = DFF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.61057 -0.12048 0.00946 0.13166 0.41466   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.75457 0.11227 77.98 <2e-16 \*\*\*  
## Social.Policy 0.60130 0.05249 11.46 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.2083 on 83 degrees of freedom  
## Multiple R-squared: 0.6125, Adjusted R-squared: 0.6079   
## F-statistic: 131.2 on 1 and 83 DF, p-value: < 2.2e-16

Добавим в каждую полученную модель фиктивные переменные, используя принадлежность каждого региона к одному из восьми федеральных округов.

##   
## Call:  
## lm(formula = Wage ~ GRP \* FO, data = DFF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.26994 -0.09415 -0.01637 0.08700 0.25681   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.761488 1.002547 4.749 1.07e-05 \*\*\*  
## GRP 0.429969 0.076706 5.605 3.96e-07 \*\*\*  
## FOПФО -2.763547 1.825831 -1.514 0.1347   
## FOСЗФО -0.108302 1.199163 -0.090 0.9283   
## FOСКФО 0.809912 2.832447 0.286 0.7758   
## FOСФО 0.154394 1.672962 0.092 0.9267   
## FOУФО 0.660037 1.228946 0.537 0.5929   
## FOЦФО -2.747227 1.427624 -1.924 0.0584 .   
## FOЮФО -4.641421 2.749424 -1.688 0.0959 .   
## GRP:FOПФО 0.204524 0.144825 1.412 0.1624   
## GRP:FOСЗФО -0.003935 0.092000 -0.043 0.9660   
## GRP:FOСКФО -0.075930 0.237677 -0.319 0.7503   
## GRP:FOСФО -0.031913 0.132435 -0.241 0.8103   
## GRP:FOУФО -0.064839 0.093119 -0.696 0.4886   
## GRP:FOЦФО 0.210342 0.111892 1.880 0.0644 .   
## GRP:FOЮФО 0.360624 0.222217 1.623 0.1092   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1316 on 69 degrees of freedom  
## Multiple R-squared: 0.8714, Adjusted R-squared: 0.8434   
## F-statistic: 31.16 on 15 and 69 DF, p-value: < 2.2e-16

##   
## Call:  
## lm(formula = Wage ~ Retail \* FO, data = DFF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.37235 -0.06850 -0.00198 0.05053 0.82011   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 6.90720 3.25301 2.123 0.03732 \*   
## Retail 0.29234 0.27412 1.066 0.28993   
## FOПФО -6.37434 3.84980 -1.656 0.10231   
## FOСЗФО -14.68766 5.36405 -2.738 0.00785 \*\*  
## FOСКФО -0.33387 3.63909 -0.092 0.92717   
## FOСФО -3.24605 3.84494 -0.844 0.40145   
## FOУФО -10.56683 4.50338 -2.346 0.02183 \*   
## FOЦФО -9.21922 3.82516 -2.410 0.01862 \*   
## FOЮФО -4.46183 3.93807 -1.133 0.26114   
## Retail:FOПФО 0.50319 0.32534 1.547 0.12652   
## Retail:FOСЗФО 1.21642 0.45138 2.695 0.00884 \*\*  
## Retail:FOСКФО -0.01565 0.30891 -0.051 0.95975   
## Retail:FOСФО 0.23874 0.32564 0.733 0.46595   
## Retail:FOУФО 0.86129 0.37564 2.293 0.02491 \*   
## Retail:FOЦФО 0.74530 0.32233 2.312 0.02376 \*   
## Retail:FOЮФО 0.33573 0.33288 1.009 0.31672   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1754 on 69 degrees of freedom  
## Multiple R-squared: 0.7715, Adjusted R-squared: 0.7219   
## F-statistic: 15.53 on 15 and 69 DF, p-value: < 2.2e-16

##   
## Call:  
## lm(formula = Wage ~ Social.Policy \* FO, data = DFF)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.40171 -0.12740 0.00142 0.12828 0.34468   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.808368 0.423783 20.785 < 2e-16 \*\*\*  
## Social.Policy 0.589595 0.157662 3.740 0.000377 \*\*\*  
## FOПФО 0.648791 0.664763 0.976 0.332487   
## FOСЗФО -0.383836 0.538667 -0.713 0.478517   
## FOСКФО 1.242654 0.562220 2.210 0.030406 \*   
## FOСФО 0.520354 0.808656 0.643 0.522046   
## FOУФО 0.006473 0.543114 0.012 0.990525   
## FOЦФО -0.151622 0.504373 -0.301 0.764613   
## FOЮФО -1.435924 0.967464 -1.484 0.142306   
## Social.Policy:FOПФО -0.367625 0.311922 -1.179 0.242614   
## Social.Policy:FOСЗФО 0.157208 0.211515 0.743 0.459856   
## Social.Policy:FOСКФО -0.767251 0.260180 -2.949 0.004350 \*\*   
## Social.Policy:FOСФО -0.333381 0.371342 -0.898 0.372429   
## Social.Policy:FOУФО 0.016244 0.206434 0.079 0.937508   
## Social.Policy:FOЦФО 0.109420 0.211195 0.518 0.606046   
## Social.Policy:FOЮФО 0.756356 0.501320 1.509 0.135934   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1874 on 69 degrees of freedom  
## Multiple R-squared: 0.7392, Adjusted R-squared: 0.6825   
## F-statistic: 13.04 on 15 and 69 DF, p-value: 1.127e-14

Применяем функцию для исключения незначимых регрессоров к 1 модели

##   
## Call:  
## lm(formula = as.formula(paste(y.var.name, " ~ .", sep = "")),   
## data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.31657 -0.09055 -0.01196 0.10048 0.28787   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.510860 0.332720 13.558 < 2e-16 \*\*\*  
## GRP 0.447258 0.026567 16.835 < 2e-16 \*\*\*  
## FOПФО -0.187703 0.048480 -3.872 0.000227 \*\*\*  
## FOСЗФО -0.132124 0.052092 -2.536 0.013253 \*   
## FOЦФО -2.496599 1.076018 -2.320 0.023015 \*   
## FOЮФО -0.177174 0.063615 -2.785 0.006751 \*\*   
## GRP.FOСФО -0.016566 0.004078 -4.063 0.000117 \*\*\*  
## GRP.FOУФО -0.014652 0.004899 -2.991 0.003747 \*\*   
## GRP.FOЦФО 0.193052 0.086213 2.239 0.028064 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1325 on 76 degrees of freedom  
## Multiple R-squared: 0.8564, Adjusted R-squared: 0.8413   
## F-statistic: 56.64 on 8 and 76 DF, p-value: < 2.2e-16

Аналогично для 2 модели:

##   
## Call:  
## lm(formula = as.formula(paste(y.var.name, " ~ .", sep = "")),   
## data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.33329 -0.08461 -0.01138 0.04628 0.82011   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.40906 0.96839 4.553 2.07e-05 \*\*\*  
## Retail 0.50288 0.08146 6.173 3.41e-08 \*\*\*  
## FOПФО -0.43867 0.07676 -5.715 2.24e-07 \*\*\*  
## FOСЗФО -12.18952 4.44055 -2.745 0.00761 \*\*   
## FOСКФО -0.42427 0.09615 -4.413 3.47e-05 \*\*\*  
## FOСФО -0.41919 0.08040 -5.214 1.66e-06 \*\*\*  
## FOУФО -8.06869 3.30916 -2.438 0.01719 \*   
## FOЦФО -6.72107 2.26249 -2.971 0.00402 \*\*   
## FOЮФО -0.49329 0.09444 -5.223 1.60e-06 \*\*\*  
## Retail.FOСЗФО 1.00588 0.37338 2.694 0.00875 \*\*   
## Retail.FOУФО 0.65075 0.27339 2.380 0.01991 \*   
## Retail.FOЦФО 0.53475 0.19059 2.806 0.00643 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1782 on 73 degrees of freedom  
## Multiple R-squared: 0.7504, Adjusted R-squared: 0.7128   
## F-statistic: 19.96 on 11 and 73 DF, p-value: < 2.2e-16

Аналогично для 3 модели:

##   
## Call:  
## lm(formula = as.formula(paste(y.var.name, " ~ .", sep = "")),   
## data = data)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.48687 -0.11460 0.03292 0.11431 0.39384   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 8.73724 0.10916 80.040 < 2e-16 \*\*\*  
## Social.Policy 0.62417 0.04990 12.508 < 2e-16 \*\*\*  
## FOСКФО 1.31378 0.38333 3.427 0.000965 \*\*\*  
## Social.Policy.FOСКФО -0.80183 0.21181 -3.786 0.000295 \*\*\*  
## Social.Policy.FOСФО -0.08011 0.02861 -2.800 0.006411 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.1864 on 80 degrees of freedom  
## Multiple R-squared: 0.7008, Adjusted R-squared: 0.6859   
## F-statistic: 46.86 on 4 and 80 DF, p-value: < 2.2e-16

Используем функцию anova(), чтобы проверить гипотезы об эквивалентности построенных моделей.

## Analysis of Variance Table  
##   
## Model 1: Wage ~ GRP  
## Model 2: Wage ~ GRP + FOПФО + FOСЗФО + FOЦФО + FOЮФО + GRP.FOСФО + GRP.FOУФО +   
## GRP.FOЦФО  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 83 1.8922   
## 2 76 1.3347 7 0.5575 4.5349 0.0002864 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: Wage ~ Retail  
## Model 2: Wage ~ Retail + FOПФО + FOСЗФО + FOСКФО + FOСФО + FOУФО + FOЦФО +   
## FOЮФО + Retail.FOСЗФО + Retail.FOУФО + Retail.FOЦФО  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 83 4.2779   
## 2 73 2.3191 10 1.9588 6.166 9.482e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

## Analysis of Variance Table  
##   
## Model 1: Wage ~ Social.Policy  
## Model 2: Wage ~ Social.Policy + FOСКФО + Social.Policy.FOСКФО + Social.Policy.FOСФО  
## Res.Df RSS Df Sum of Sq F Pr(>F)   
## 1 83 3.6006   
## 2 80 2.7800 3 0.82061 7.8717 0.0001148 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Видно, что для двух рассматриваемых факторов добавление фиктивных переменных уменьшает остаточную дисперсию (RSS). Составим список моделей

## Модель R.2.скорр F.расч Станд.Ошибка  
## 1 fit.X1 0.794 324.62 0.2  
## 2 fit.X1.fo 0.841 56.64 0.1  
## 3 fit.X2 0.534 97.30 0.2  
## 4 fit.X2.fo 0.713 19.96 0.2  
## 5 fit.X3 0.608 131.21 0.2  
## 6 fit.X3.fo 0.686 46.86 0.2

Лучшей моделью является модель 2, т.к. она имеет наименьшую стандартную ошибку