УДК 51

СТАТИСТИЧЕСКИЙ АНАЛИЗ ЗАВИСИМОСТИ РАЗМЕРА ПЕНСИИ ОТ СТАЖА РАБОТЫ И ОТ СТРАХОВЫХ ВЫПЛАТ

Трепаков М.В., студент II курса, гр. ИТб-171
Кадочигова А.Н., студент II курса, гр. ИТб-171
Дягилева А.В., к.т.н., доцент
Кузбасский государственный технический университет
имени Т.Ф. Горбачева
г. Кемерово

Актуальность данной темы заключается в том, что пенсионная система страны – это особенно важный индикатор развитости и функционирования его социальных функций. В частности пенсионная система по праву занимает первые места в государствах, чей бюджет имеет социальную направленность. К таким государствам относится и Россия, являющаяся социальным государством. Однако в Российской Федерации пенсионная система на данный момент находится в стадии становления, укрепления и развития, и по этой причине, важно провести её статистический анализ для формирования общей картины пенсионной системы РФ.

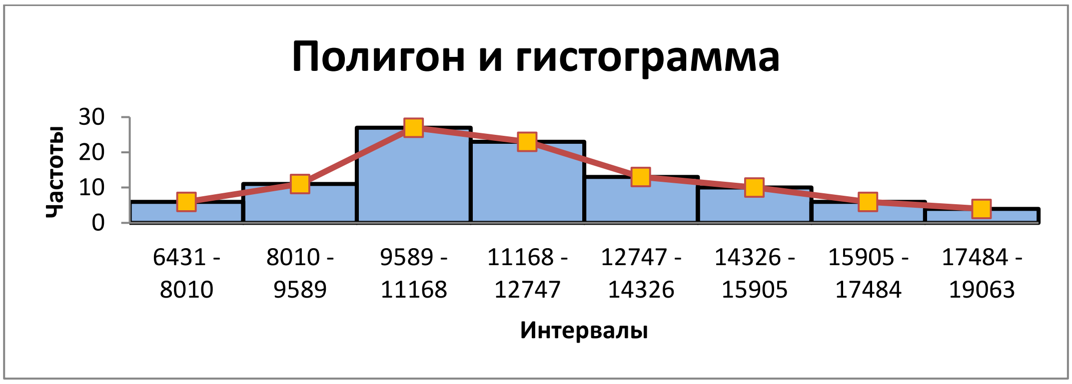
Применение в данной расчётной работе методов, используемых для статистической выборки данных, касающихся пенсий (в частности – пенсионных выплат, страховых взносов и стажа) является необходимым дополнением в работе общего анализа пенсий в Российской Федерации в нынешнее время и мало изучалось в научных статьях. Поэтому статистический анализ пенсий был взят как основа данной расчётной работы.

В данной работе, следует проанализировать массив полученных данных статистической выборки, выявить законы распределения для каждой величины и установить корреляционные законы между размером пенсии и стажем, а также между размером пенсии и страховыми взносами гражданина. Это позволит в дальнейшем предположить размер пенсии гражданина, зная его стаж и страховые выплаты.

Для достижения поставленной нами цели, необходимо сформировать определенные задачи:

1. Формирование статистической выборки данных за 2018 год;
2. Вывод закона распределения для каждой выборки одномерных данных;
3. Оценка адекватности выбранного закона при помощи различных критериев;
4. Составления корреляционных законов зависимости.

Анализируя параметр X, в первую очередь, был составлен полигон и гистограмма распределения, а также, составлена таблица интервального вариационного ряда, основываясь на известных формулах, используемых в статистике (В.Е. *ГМУРМАН*. Теория вероятностей и математическая статистика. Издание девятое, стереотипное)



Итак, для начала, была рассчитана выборочная средняя X_b, дисперсию $D_b(x)$ и среднее квадратичное отклонение $\sigma_b(x)$.

$\bar{x}_b$ =11957,5

$D_b(x)$ = 7379993,36

$\sigma_b\,(x)$ = 2716,6

По виду гистограммы выдвигается гипотеза о том, что это нормальное распределение. Рассматриваются числовые характеристики.

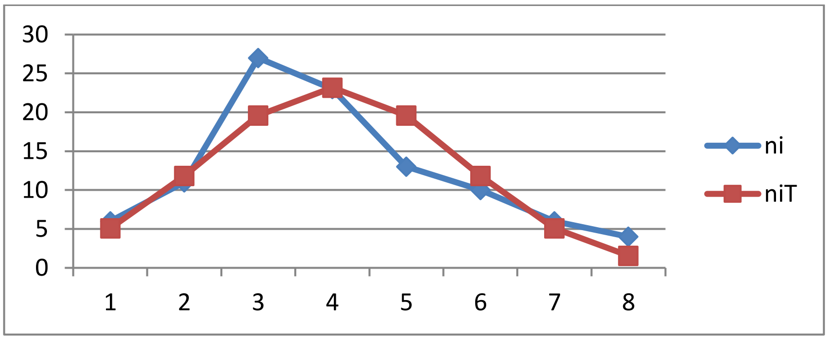
Для нормального распределения характерны формулы:

Используя формулы числовых характеристик, соответствующие нормальному распределению были получены приблизительно равные значения, а значит и предположение о том, что это нормальное распределение, также подтвердилось по числовым характеристикам.

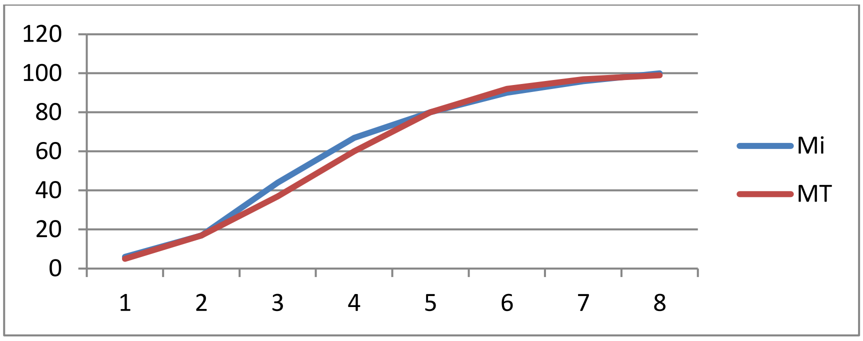
Для проверки нашей гипотезы, были использованы различные методы, такие как: критерий Пирсона, Критерий Колмогорова, Критерий Романовского, Числа Вестергарда и также использовалась проверка по структурным среним - моде и медиане, они также уовлетворят условиям нормального распрееления. По итогам проверки, согласно всем вышеперечисленным критериям, выборка данных **размера пенсии** (выборка X) полностью соответствует нормальному распределению.

Также в ходе проверки выборки X, для более точного понимания, были построены графики:

График зависимости n_i (эмпирической частоты) и n_i^T (теоритической частоты)



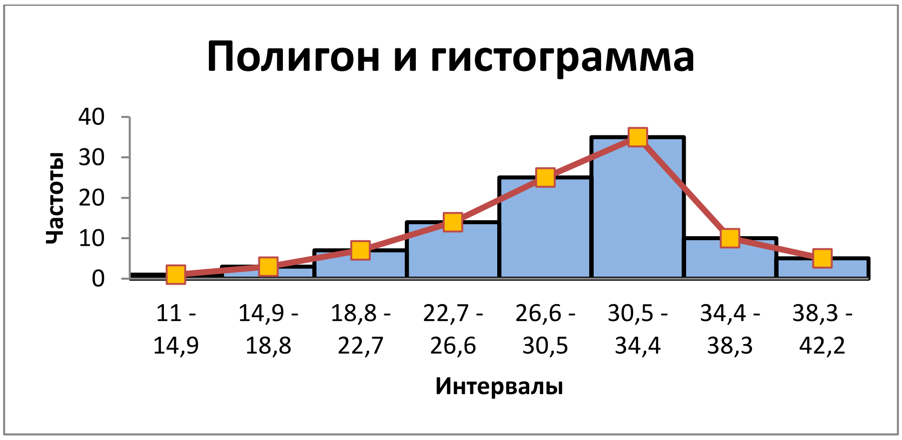
А также график накопленных частот, или кумулята



Из графика кумуляты видно, что линии накопленных частот и накопленных теоретических частот почти полностью совпадают, что еще раз свидетельствует о том, что выведенный закон справедлив.

Проверив случайные величины X (размер пенсии) по всем критериям, мы убедились, что они принадлежат нормальному распределению. Теперь необходимо проделать те же действия для случайных величин - стажа работы (Y) и страховых выплат (Z).

Анализируя параметр Y (стаж работы), так же как и для параметра X в первую очередь, был составлен полигон и гистограмма распределения и таблица интервального вариационного ряда, основываясь на известных формулах, используемых в статистике (В.Е. ГМУРМАН. Теория вероятностей и математическая статистика. Издание девятое, стереотипное)



Как и для выборки X, сначала, была рассчитана выборочная средняя $\overline{Y_b}$, дисперсию $D_b(y)$ и среднее квадратичное отклонение $\sigma_b(y)$.

$\overline{y_b} = 29{,}681$

$D_b(y) = 29{,}901$

$\sigma_b(y) = 5{,}468$

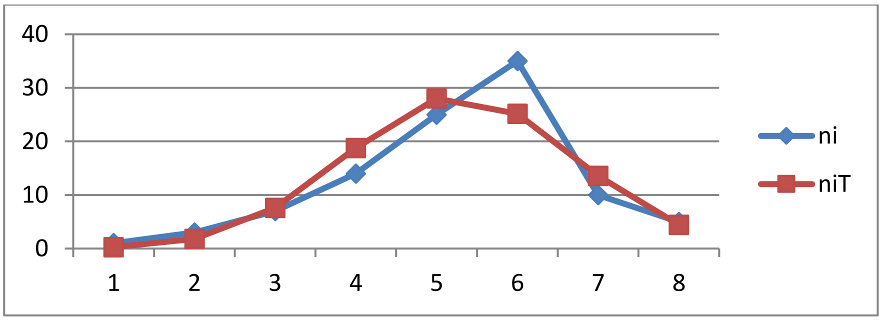
По виду гистограммы мы выдвигается гипотеза о том, что это нормальное распределение. Рассматриваются числовые характеристики.

Используя формулы числовых характеристик нормального распределения, были получены приблизительно равные значения, а значит и предположение о том, что это нормальное распределение, также подтвердилось по числовым характеристикам.

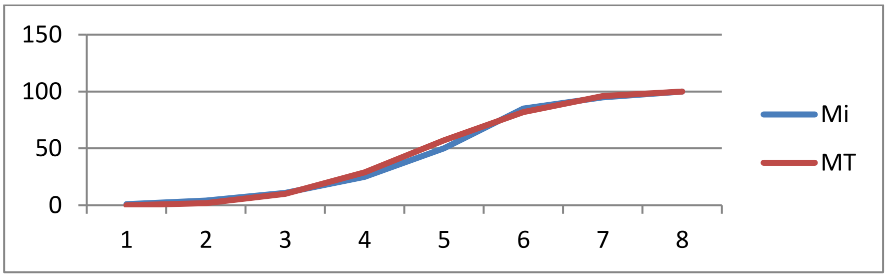
Для проверки теории о нормальном распределении выборки стажа работы (выборки Y) использовались те же критерии, что и для выборки размера пенсии (выборки X) и по итогу, все данные критерии подтвердили теорию о нормальном распределении выборки данных стажа работы.

Для более детального понимания, были построены несколько графиков:

График зависимости полигонов n_i (эмпирической частоты) и n_i^T (теоритической частоты):



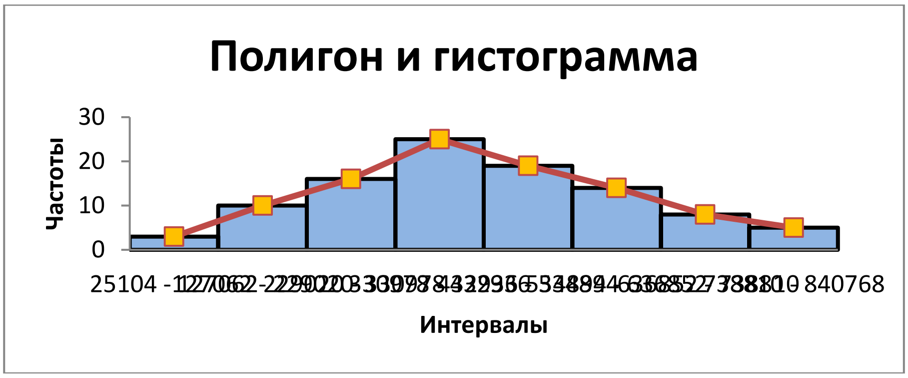
И график накопленных частот, или кумулята.



Из последнего графика видно, что линии накопленных частот и накопленных теоретических частот почти полностью совпадают, что еще раз свидетельствует о том, что выведенный закон справедлив.

Проверив случайные величины Y (стаж работы) по всем критериям, мы убедились, что они принадлежат нормальному распределению. Теперь необходимо проделать те же действия для случайных величин Z (страховых выплат).

Проанализировав параметр Z (страховые выплаты), так же как и предыдущие параметры, в первую очередь, был составлен полигон и гистограмма распределения т таблица интервального вариационного ряда, основываясь на известных формулах, используемых в статистике (В.Е. ГМУРМАН. Теория вероятностей и математическая статистика. Издание девятое, стереотипное)



Также, как и для предыдущих параметров, была рассчитана выборочная средняя $\overline{Z_b}$, дисперсию $D_b(z)$ и среднее квадратичное отклонение $\sigma_b(z)$.

$\overline{z_b} = 428858$

$D_b(z) = 30026170884$

$\sigma_b(z) = 173280{,}6$

По виду гистограммы мы выдвигается гипотеза о том, что это нормальное распределение. Используя формулы характерные нормальному распределению, рассматриваются числовые характеристики. При этом были получены приблизительно равные значения, а значит и предположение о том, что это нормальное распределение, также подтвердилось по числовым характеристикам. Используя все те же критерии, что и при анализе параметров X (размер пенсии) и Y (стаж работы), анализируется параметр Z или страховые отчисления. По итогу проверки по эти критериям, теория о нормальном распределении выборки страховых отчислений полностью подтвердилась.

Также, были составлены графики:

График зависимости полигонов n_i (эмпирической частоты) и n_i^T (теоритической частоты):

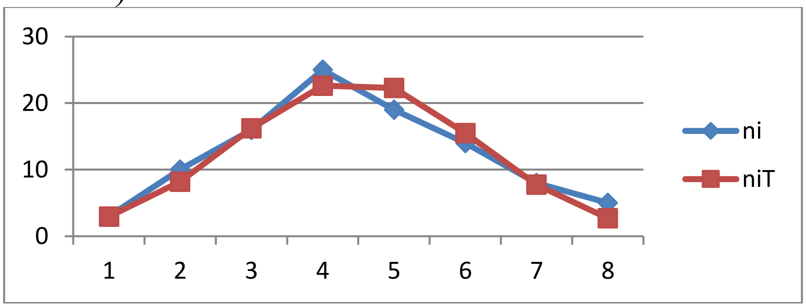
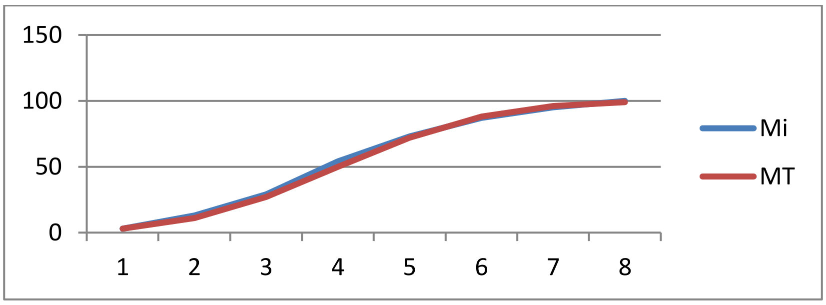


График накопленных частот, или кумулята.



Из графика накопленных частот видно, что линии накопленных частот и накопленных теоретических частот практически полностью совпадают, что еще раз свидетельствует о том, что выведенный закон справедлив.

И в соответствии с данным критерием, распределение также является нормальным.

Проверив случайные величины страховых выплат (Z) по всем критериям, мы убедились, что они принадлежат нормальному распределению. Далее необходимо рассчитать корреляцию величины пенсии от стажа (X - Y) и величины пенсии от страховых выплат (X - Z).

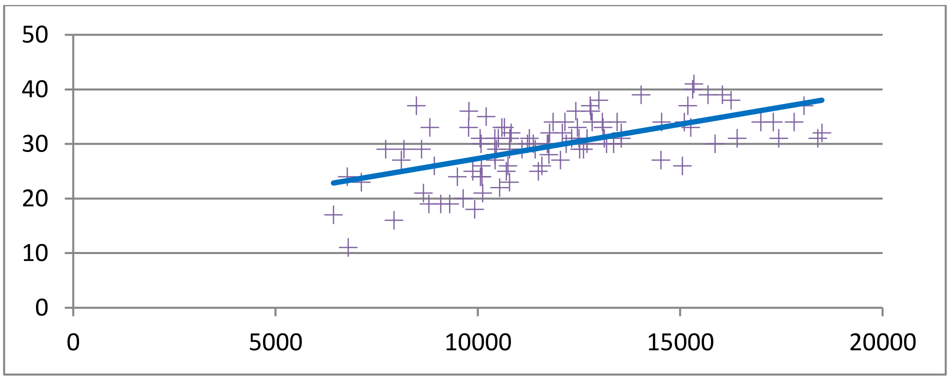
Рассчитаем корреляцию величины пенсии от стажа (X - Y). Для этого сначала определяется коэффициент корреляции по известным формулам из статистики (В.Е. *ГМУРМАН*. Теория вероятностей и математическая статистика. Издание девятое, стереотипное)

Для данных X и Y он будет равным 0,64. вычислили коэффициент корреляции, однако данный коэффициент является выборочным, поэтому проверим гипотезу H_0: $r_г=0$ о равенстве нулю генерального коэффициента корреляции. Если гипотеза принимается, то это будет означать, что взаимосвязи (корреляции) между нашими данными X и Y нет. Чтобы проверить гипотезу при уровне значимости α о равенстве нулю генерального коэффициента корреляции нормальной двумерной случайной величины при конкурирующей гипотезе $r_г \neq 0$, надо вычислить наблюдаемое значение критерия $T_{набл}$. В нашем случае, $T_{набл}=8{,}245$; k=98; α=0,05; $t_{кр}=1{,}66$. $|T_{набл}| > t_{кр}$, а значит мы отвергаем гипотезу H_0. Таким образом, мы принимаем обратную гипотезу о том, что наши величины X и Y имеют корреляцию. Необходимо также оценить надежность коэффициента корреляции. Для этого посчитаем среднее квадратическое отклонение коэффициента корреляции: Оно получилось равным 0,05904.

Далее составим уравнение регрессии размера пенсии от стажа. Для удобства расчётов была составлена корреляционная таблица X – Y. Используя эту таблицу были высчитаны коэффициенты для уравнения корреляции. Подставив их, мы получили следующее уравнение регрессии:

x = 318y + 2520

Так же, для наглядности, составим корреляционное поле и линию регрессии нём:



Подытожим, для того, чтобы найти зависимость (корреляцию) между величинами X и Y, были изучены распределения X и Y, определены виды распределений и выведены законы, также был определён – коэффициент корреляции, найдя который, было выяснено, что между X и Y всё таки существует корреляция. Это означает, что размер пенсии напрямую зависит от рабочего стажа определенного гражданина. И затем было составлено уравнение регрессии, для того, чтобы в дальнейшем можно было предположить размер пенсии гражданина, зная его рабочий стаж.

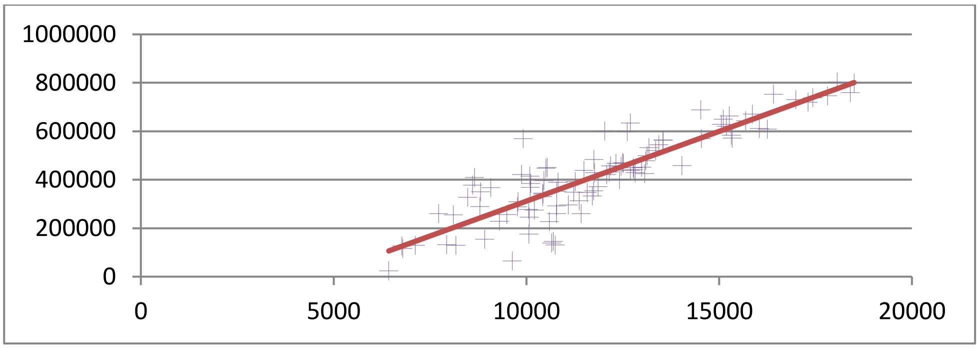
Рассчитаем корреляцию величины пенсии от стажа (X – Z). Для этого сначала определяется коэффициент корреляции по известным формулам из статистики (В.Е. *ГМУРМАН*. Теория вероятностей и математическая статистика. Издание девятое, стереотипное)

Для данных X и Y он будет равным 0,92. вычислили коэффициент корреляции, однако данный коэффициент является выборочным, поэтому проверим гипотезу H_0: $r_г=0$ о равенстве нулю генерального коэффициента корреляции. Если гипотеза принимается, то это будет означать, что взаимосвязи (корреляции) между нашими данными X и Y нет. Чтобы проверить гипотезу при уровне значимости α о равенстве нулю генерального коэффициента корреляции нормальной двумерной случайной величины при конкурирующей гипотезе $r_г \neq 0$, надо вычислить наблюдаемое значение критерия $T_{набл}$. В нашем случае, $T_{набл}$= 23,238; k = 98; α=0,05; $t_{кр}$=1,66. $|T_{набл}| > t_{кр}$, а значит мы отвергаем гипотезу H_0. Таким образом, мы принимаем обратную гипотезу о том, что наши величины X и Y имеют корреляцию. Необходимо также оценить <u>надежность</u> коэффициента корреляции. Для этого посчитаем среднее квадратическое отклонение коэффициента корреляции: Оно равно 0,01536.

Далее составим уравнение регрессии размера пенсии от стажа. Для удобства расчётов была составлена корреляционная таблица X – Y. Используя эту таблицу были высчитаны коэффициенты для уравнения корреляции. Подставив их, мы получили следующее уравнение регрессии:

$x = 0{,}0144z + 5772$

Так же, для наглядности, составим корреляционное поле и линию регрессии нём:



Итак, для того, чтобы найти зависимость (корреляцию) между величинами X и Z, были изучены распределения X и Z, определены виды распределений и выведены законы, также был определён – коэффициент корреляции, найдя который, было выяснено, что между X и Z всё таки существует корреляция. Это означает, что размер пенсии напрямую зависит также и от страховых взносов, определенного гражданина, которые он уплачивал. Затем было составлено уравнение регрессии, для того, чтобы в дальнейшем можно было предположить размер пенсии гражданина, зная величину его страховых взносов.

Вывод:

Подводя итоги наших расчётов по статистике пенсионных начислений, можно с уверенностью сказать, что тщательно обработав, полученный массив данных, включающий в себя непосредственно данные о размере пенсии, стаже и страховых выплатах, получилось выявить важную взаимозависимость, или корреляцию, между размером пенсии и стажем работы, а также формулу, по которой в дальнейшем можно предположить примерный размер пенсии. Так же была найдена корреляция между размером пенсии и страховыми взносами, которая, исходя из коэффициента корреляции близкого к 1 (0,92), оказалась практически полной. Для корреляции размера пенсии и страховых выплат также было составлено уравнение, позволяющее предположить приблизительный размер пенсии гражданина, зная сумму страховых взносов, которые он заплатил. В процессе обработки массивов данных был успешно применены: критерий Колмогорова, критерий Романовского и критерий Пирсона. Последний позволяет подтвердить или опровергнуть теорию об определенном законе распределения, которую мы выдвинули, основываясь на виде гистограммы и полигона распределения наших массивов данных. Для упрощения вычислений в ходе работы использовались такие программы как: Калькулятор Windows 10, набор программ Microsoft Office, в частности, Microsoft Office Excel для произведения расчётов и удобной структуризации полученных данных в таблицы, и Microsoft Office Word для упорядочивания полученной информации в текстовом формате.

Список литературы:

1. Гмурман, В.Е. Теория вероятностей и математическая статистика: Учебное пособие для бакалавров / В.Е. Гмурман. - М.: Юрайт, 2013. - 479 с.
2. Горлач, Б.А. Теория вероятностей и математическая статистика: Учебное пособие / Б.А. Горлач. - СПб.: Лань, 2013. - 320 с.
3. Калинина, В.Н. Теория вероятностей и математическая статистика: Учебник для бакалавров / В.Н. Калинина. - М.: Юрайт, 2013. - 472 с.
4. Федеральный закон от 28.12.2013 N 400-ФЗ (ред. от 27.12.2018) "О страховых пенсиях".
5. Федеральный закон "О государственном пенсионном обеспечении в Российской Федерации" от 15.12.2001 N 166-ФЗ (ред. от 27.12.2018)