Econometrics course work. Employee wellbeing.

Sofia Kyba

April 2021

1 Introduction. Problem area

Employee wellbeing is a question of health and safety at work. It is about feelings concerning human's job and his satisfaction. Nowadays the issue of employee wellbeing concerns all successful companies. It is not surprise that their main goal is better profit, that in fact directly depends on employees motivation and productivity.

Perfect employee wellbeing $\Longrightarrow$ High motivation $\Longrightarrow$ Better productivity $\Longrightarrow$ Increasing profit and more satisfied consumers.

Besides, there is also a problem related to migrations of younger specialists between companies, as they are more flexible than elder people and thus can easily change their workspace if their needs aren't satisfied. And another crucial moment nowadays is COVID-19 pandemics, that obviously has quite bad impact on employees feelings.

That's why this issue is very important for employers. It is essential for them to be aware of factors that have a significant impact on employees' wellbeing. Taking this information into account they will be able to control the situation by changing appropriate conditions for workers to make them feel more satisfied.

2 Literature review

I looked through some related works and articles connected with it. So I will provide some information I found just for better understanding of problem.

Migrations among young employees

Millennials are the most flexible and are most likely to switch jobs. Gallup report on the millennial generation revealed that 21% of millennials said they had changed jobs within the past year, which was more than three times the number of non-millennials who reported the same. Due to Gallup estimations millennial turnover costs the U.S. economy $30.5 billion annually.

There can be a lot of potential reasons, but one of them is certainly their low engagement in the workplace. Gallup also has found that only 29% of young people are engaged at work, emotionally and behaviorally connected to their job and company. Another 16% of millennials are actively disengaged, meaning they are more or less out to do damage to their company. And the majority of millennials (55%) are not engaged.

COVID-19 impact

"The coronavirus pandemic has made employees' mental health top-of-mind for employers, as many working adults are feeling a sense of uncertainty," said Nancy Reardon, chief strategy and product officer.

They mostly are concerned about the stability of their jobs, health of their relatives. They may feel anxious about the virus itself, uncertain of what the future prepares for them or worried about their financial situation. All of this can have an impact on their mental health and general wellbeing.

Considering these situations employers have to be more conscious of this problem and try to improve some factors that influence employees' wellbeing.

Factors that influence employee wellbeing most of all

Here I provide explored research by Banking Standards Board (BSB) relating to factors that may affect personal wellbeing in the workplace. 1.Work conditions (place, where employee works) 2.Workload (hours spent per day) 3.Freedom (how much employees are free to decide how to complete tasks related to their jobs) 4.Security and change 5.Relationships with colleagues 6.Organisational justice 7.Work-life balance 8.Meaning at work (how much employees feel appreciated)

Considering other researches, we can also state that employee's wage also influences his satisfaction.

3 Data description

I couldn't find appropriate data, as all the datasets were either quite old, or company specified, or sphere specified. That's why I decided to gather data by myself for this project. I made a poll in Google Forms and sent it to all my friends and relatives and posted it in several social networks. In result, in two weeks I received 199 responds from people from whole Ukraine, of different age and specialities.

Before creating poll I looked through several articles and datasets to define the factors I should take into account. In this survey I included some basic questions about person's age, gender, marital status and city they live in. Also there were few questions concerning their education and speciality. And finally I asked about their level of satisfaction concerning such factors as environment where their workplace is, their management, colleagues, wage, freedom level and ability for self-development and so on.

Some diagrams I built for better understanding of the data:

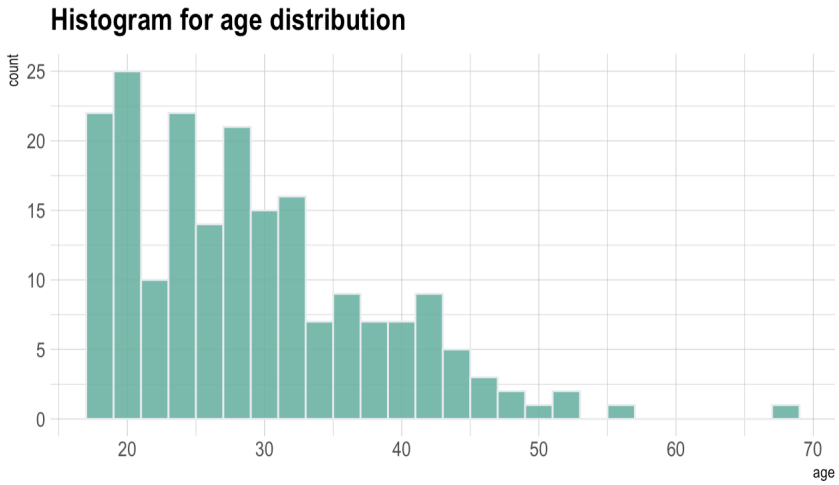

Figure 1. Age distribution

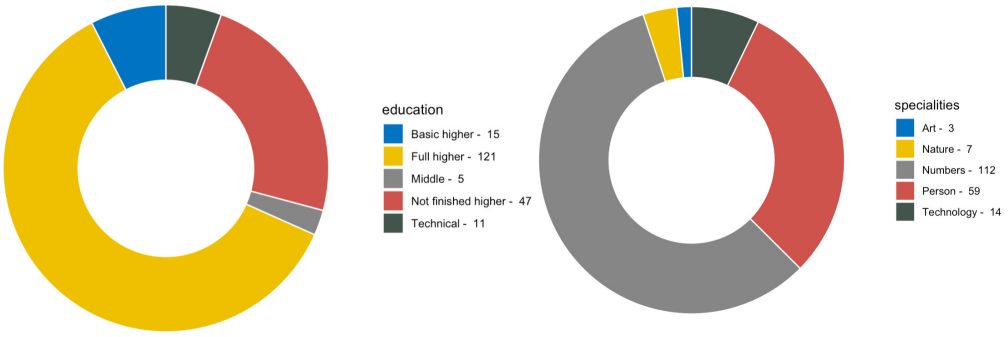

Figure 2. Education level and specialities

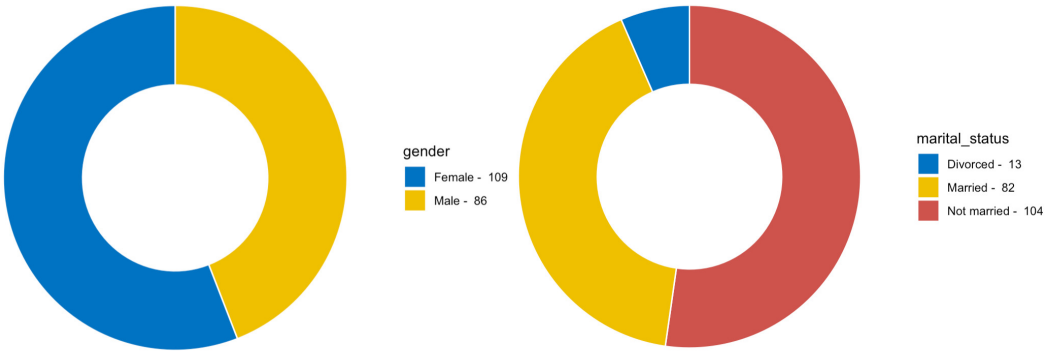

Figure 3. Gender and marital status

4 Methodology explanation

4.1 Correlation matrix

First of all, I constructed correlation matrix and plotted it to see the relationships between variables.

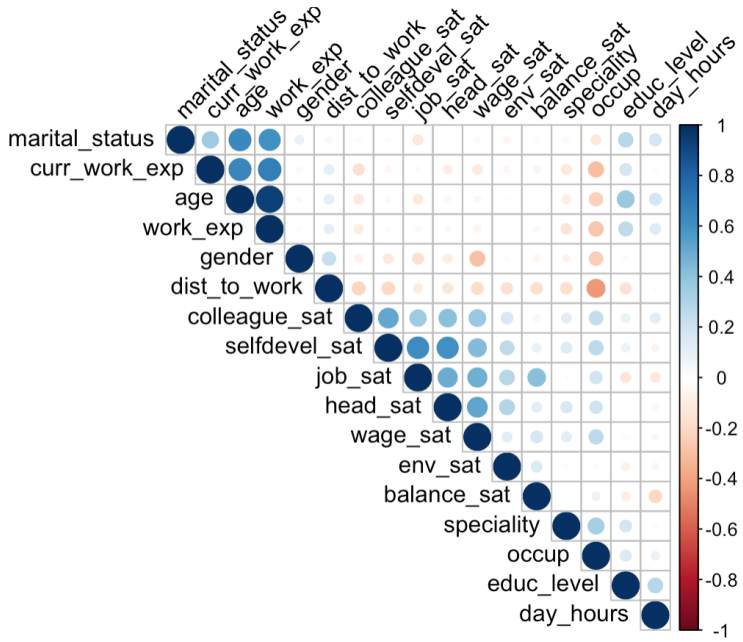

Figure 4. Correlation matrix

Here we can see positive correlation between age and work experience, that is in fact very obvious. Also we can observe positive relationships between management satisfaction and wage satisfaction together with satisfaction of self-development at work. And that is also clear.

But it is the most important for us to concentrate on job satisfaction. We see that it has the strongest correlation with self-development satisfaction, management satisfaction, wage satisfaction, balance satisfaction and little less with environment satisfaction and colleague satisfaction. I also did Pearson correlation tests that confirmed it.

4.2 Building linear models

Before building linear regression model I checked linearity among variables.

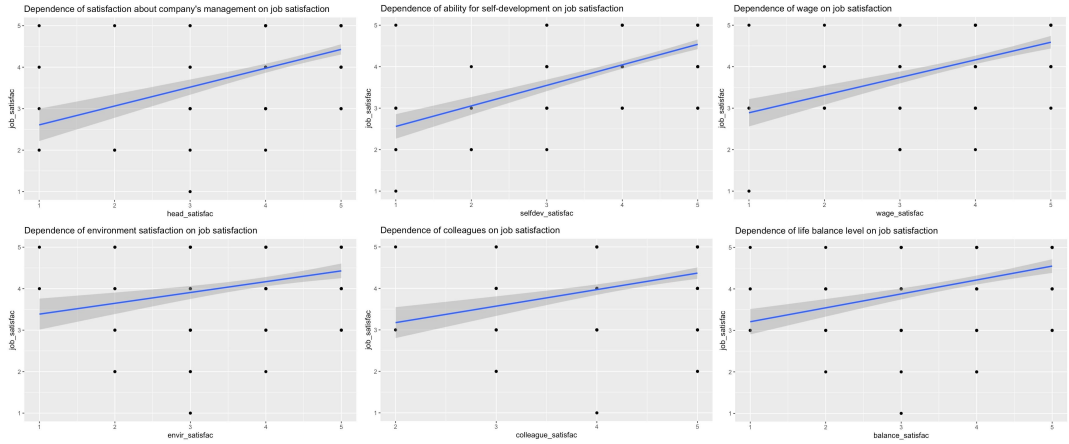

Figure 5. Linear dependence

Looking at these plots, we observe linearity in parameters.

Now I try to build linear model, having job satisfaction as dependent variable. I tried to add each factor that was strongly correlated with job satisfaction one by one and observed how the linear regression model changed.

	Self_development	Self_development and management	Self_development, management and wage	Self_development, management, wage and balance	Self_development, management, wage, balance and environment	Self_development, management, wage, balance, environment, education and distance to work
(Intercept)	4.14 ***	4.14 ***	4.14 ***	4.14 ***	4.14 ***	4.14 ***
	(0.05)	(0.05)	(0.04)	(0.04)	(0.04)	(0.04)
`poll_answers$selfdev_satisfac`	0.52 ***	0.42 ***	0.39 ***	0.38 ***	0.38 ***	0.40 ***
	(0.05)	(0.06)	(0.06)	(0.05)	(0.05)	(0.05)
`poll_answers$head_satisfac`		0.16 **	0.07	0.06	0.05	0.04
		(0.06)	(0.06)	(0.05)	(0.06)	(0.05)
`poll_answers$wage_satisfac`			0.20 ***	0.16 **	0.16 ***	0.18 ***
			(0.05)	(0.05)	(0.05)	(0.05)
`poll_answers$balance_satisfac`				0.27 ***	0.27 ***	0.27 ***
				(0.04)	(0.04)	(0.04)
`poll_answers$envir_satisfac`					0.07	0.07
					(0.04)	(0.04)
`poll_answers$education_level`						-0.10 *
						(0.04)
`poll_answers$dist_to_work`						0.07
						(0.04)
N	199	199	199	199	199	199
R2	0.39	0.41	0.45	0.55	0.56	0.58

All continuous predictors are mean-centered and scaled by 1 standard deviation. *** p < 0.001; ** p < 0.01; * p < 0.05.

Figure 6. Models' summary

This summary describes all built linear models. Now I will explain each step in more detail:

1. Firstly I built simple model, that included only self development satisfaction. As it is shown at the figure, the ability for self-development has a huge

influence on employees' wellbeing, that is in fact very obvious. If person doesn't see new opportunities for her at work, she loses enthusiasm and desire to work, her motivation and satisfaction reduces.

2. Having added one more factor - management satisfaction, our model became better. And we can make a conclusion, that employee's satisfaction of the management office in his company is really important.

3. Now I also added wage satisfaction. At this moment we can observe that our model again became better. We can see it by looking at the values of adjusted R-squared. We also can look at the p-value of F-statistics, that is lower than our significant level. However management satisfaction became less significant now and that can be explained by quite big correlation between wage and management satisfaction. I will return to this question further.

4. Having added balance satisfaction we observe significant improvement of model.

5. Now I add another factor - environment satisfaction, that has little correlation with job satisfaction and also can has impact.

6. I also decided to include the factors of education level and distance to workplace as they also have a little impact.

Now, when all variables are already included, I decided to check for correlation between management satisfaction and wage satisfaction and between management satisfaction and ability for self-development satisfaction. As I suggested earlier, there is quite considerable correlation between them. And this is quite natural observation. So, I decided to exclude head_satisfaction variable.

Here is summary for final model:

```
Coefficients:
                             Estimate Std. Error t value Pr(>|t|)
(Intercept)                  0.765523   0.321221   2.383  0.01814 *
poll_answers$selfdev_satisfac  0.398919   0.042675   9.348  < 2e-16 ***
poll_answers$wage_satisfac     0.194818   0.045643   4.268 3.09e-05 ***
poll_answers$balance_satisfac  0.256571   0.039110   6.560 4.86e-10 ***
poll_answers$envir_satisfac    0.078632   0.044938   1.750  0.08175 .
poll_answers$education_level  -0.093474   0.035566  -2.628  0.00928 **
poll_answers$dist_to_work      0.003348   0.001819   1.840  0.06730 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5471 on 192 degrees of freedom
Multiple R-squared:  0.5837,    Adjusted R-squared:  0.5707
F-statistic: 44.87 on 6 and 192 DF,  p-value: < 2.2e-16
```

Figure 7. Final model's summary

So, finally our linear model is constructed. Although we can observe that environment satisfaction and distance to workplace are not so significant, anyway there is quite significant relationship between them and job satisfaction

variable.

We can state that this model is quite good, as significant level is high and adjusted R-squared value is 0,57 that is also not bad.

4.3 Assumptions testing

Firstly I used R base function plot().

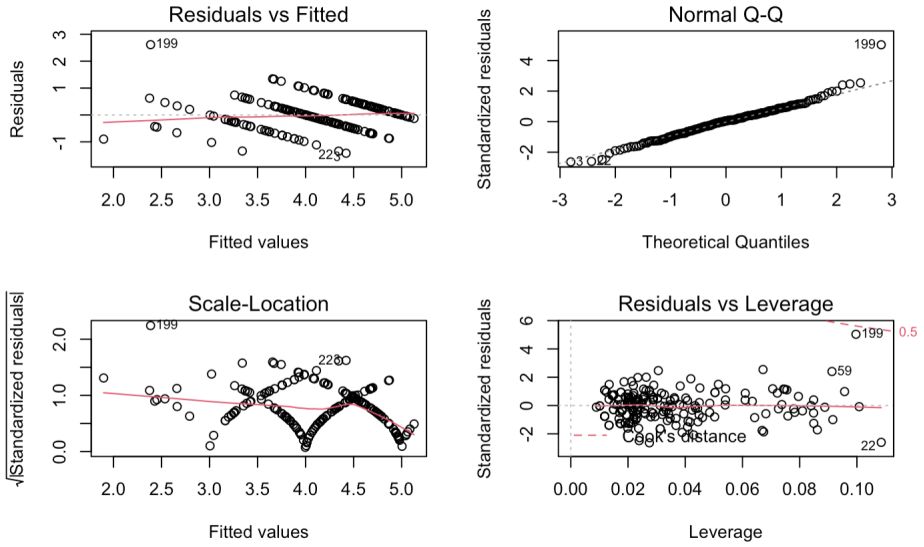

Figure 8. Residuals plot

First plot is a scatter plot of residuals on the y axis and fitted values (estimated responses) on the x axis. We see that the red line is approximately horizontal at zero, that proves assumption about linearity. Moreover, we can assume that there is no heteroscedasticity, as we observe no increase in variance, as the variables increase.

Looking at graph "Normal Q-Q" we check whether the residuals are normally distributed. Residuals points follow the straight dashed line and this proves second assumption concerning normal distribution of residuals.

Looking at graph "Residuals vs Leverage" we see that it highlights the top 3 most extreme points (#22, #59 and #199). However, there is no outliers that exceed 3 standard deviations, what is good.

We build a histogram for residuals to see whether the error term has zero population mean. And we can observe that it really has. So this assumption also holds.

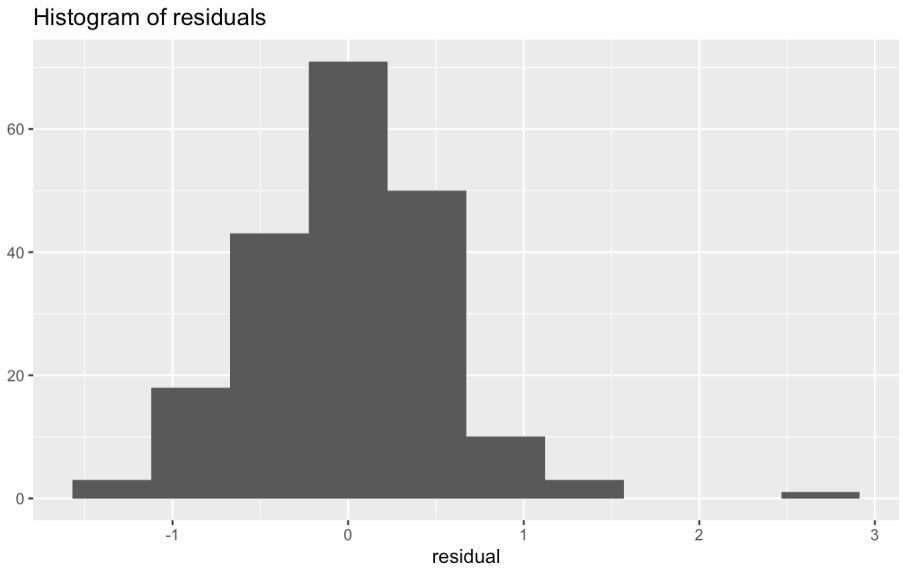

Figure 9. Histogram of residuals distribution

How we consider collinearity assumption. Collinearity is a phenomenon related to regression, in which some of the predictor variables are highly correlated among themselves and it is in fact a problem for regressions.

That's why I used function omcdiag() to find out whether there was really collinearity. Only one diagnostic (Farrar Chi-Square) showed that there was possible collinearity.

To be sure, I also tried individual collinearity diagnostic measure using imcdiag() function. I chose method VIF (Variance Inflation Factor) for this.
So I computed VIF values for each variable included in our model.

In result VIF Method failed to detect multicollinearity and each variable had VIF value less than 5, therefore no variables need to be removed and there is no multicolinearity.

Another assumption requires little or no autocorrelation in the data (residuals are independent). For checking this I used durbinWatsonTest().

In result we fail to reject H0 as p-value was quite small, and thus that means that there is no autocorrelation.

5 Conclusion. Results

The result of this course work is a linear regression model, that describes dependence of certain factors on employees' wellbeing. According to it we can conclude which factors influence it the most and which ones don't influence at all.

First of all, freedom level and ability for self-development at work have the

biggest impact on employees' satisfaction. I already mentioned that it is in fact very natural observation. It is really important for worker to feel no pressure from his management and also to have opportunities for improving himself at his work. The next important factor is life balance. That is actually the thing that company can't influence a lot, as it depends more on employee's ability to combine family, work and leisure. The 3rd factor is of course wage. It also isn't a surprise, as employee's monthly income really affects his wellbeing and motivation. Other factors that have pretty considerable impact are education level, satisfaction of environment where employee's workplace is situated and distance to work.

Almost all factors, that are included in model coincide with those ones, that were researched by Banking Standards Board.

We can also conclude that such factors as gender, marital status, age, speciality and working experience don't have considerable impact on employees' satisfaction.