



SZKOŁA GŁÓWNA HANDLOWA W WARSZAWIE
WARSAW SCHOOL OF ECONOMICS

Studium Magisteriskie

Kierunek: Metody ilościowe i systemy informacyjne

Yauheni Semianiuk

Nr albumu: 82591

Wpływ systemu pracy podczas COVID-19 na produktywność: dowody z Australii

Warszawa 2023

Spis treści

Wstęp	1
I. Przegląd literatury i hipotezy badawcze.....	2
II. Opis danych	4
2.1 Uczestnicy i procedura	4
2.2 Mierniki	4
III. Metoda badania.....	7
IV. Analiza.....	7
4.1 Istotność zmiennych i redukcja kategorii.	7
4.2 Istotność zmiennych numerycznych.....	9
4.3 Testowanie modelu.....	9
4.4 Jakość dopasowania.....	11
4.5 Interpretacja modelu. Efekty krańcowe.....	12
Dyskusja	13
Podsumowanie i zastosowanie	13
Ograniczenia.....	13
Kierunki dalszych badań	14
Bibliografia.....	14
Kod STATA	16

Wstęp

Pandemia COVID-19 zmieniła każdy aspekt pracy i życia. W odpowiedzi na krajową i lokalną politykę ograniczania rozprzestrzeniania się wirusa, przedsiębiorstwa prywatne, organizacje i instytucje państwowe i międzynarodowe zachęcały pracowników do pracy z domu. Zdalna praca jako zjawisko pojawiła się na początku XXI wieku, kiedy zaczęły się rozwijać technologie pracy zdalnej. Dzięki zdalnej pracy pracownicy mogą uniknąć dojazdów, zapewnić elastyczność harmonogramów i osiągnąć lepszą równowagę między życiem zawodowym a prywatnym. W czasie pandemii COVID-19 wiele pracowników odczuło na swoim przykładzie konsekwencje zdalnej pracy. Wcześniej ten rodzaj pracy miał charakter okazjonalny lub był stosowany tylko w szczególnych okolicznościach i w branżach produkcyjnych. Wiele przedsiębiorstw uważa, że praca hybrydowa, to znaczy pewne połączenie pracy z domu i z biura, stanie się bardziej powszechne ze względu na fakt, że pracodawcy ponieśli już stały koszt stworzenia systemów pracy zdalnej dla swoich pracowników. Dlatego wiele firm zatrudniających oraz instytucji badawczych skupia swoją uwagę na innych czynnikach związanych ze zdalnym pracowaniem.

Jednym z najważniejszych aspektów podlegających badaniu jest produktywność pracownika w nowych warunkach pracy. Tylko w latach 2020-2022 przeprowadzono kilka tysięcy badań (według wyników wyszukiwania na platformach Google Scholar i PubMed) na temat produktywności w trybie pracy zdalnej lub hybrydowej. Z tego powstaje naturalne pytanie co do konieczności niniejszego badania. Jednak będzie się ono charakteryzowało pewnymi cechami, których nie zauważono w poprzednich badaniach. Po pierwsze, zostanie sprawdzone podejście firmy zatrudniającej i jej poprzednie doświadczenie we wprowadzeniu pracy zdalnej, oraz wpływ tego czynnika na produktywność pracownika. Po drugie, użyte będą zmienne charakteryzujące firmę zatrudniającą – jej rozmiar ze względu na liczbę osób zatrudnionych, sektor oraz położenie geograficzne. Po trzecie będzie to zrobione na otwartym i ogólnodostępnym zbiorze danych, co jest częstym brakiem w dostępnych badaniach, zwłaszcza w przypadku raportów dużych firm zatrudniających, takich jak PWC i inne. Co więcej, liczne czynniki mają kontrowersyjny wpływ na produktywność pracowników, a mianowicie zmienne socjalno-demograficzne. Dlatego głównym celem badania będzie wykrycie wpływu wcześniej nieużywanych zmiennych na produktywność pracownika oraz ocena czynników, charakteryzujących się niejednoznacznym związkiem z produktywnością.

W niniejszym badaniu zbadano związek między preferencjami i odczuciami pracowników, czynnikami socjalno-demograficznymi oraz podejściem przedsiębiorstw a produktywnością pracownika w trybie pracy zdalnej i hybrydowej w czasie pandemii COVID-19 w Australii. Na podstawie przeglądu literatury zostaną skonstruowane hipotezy badawcze. Opisano również empiryczny komponent badania oraz wyniki analizy. Na końcu zostaną przedyskutowane najważniejsze wnioski, ograniczenia oraz sugestie i kierunki dalszych badań produktywności pracowników.

I. Przegląd literatury i hipotezy badawcze

Wzrost produktywności podczas pracy zdalnej jest nadal kwestią sporną. Badanie przeprowadzone jeszcze w 2016 roku we Francji ujawniło, że około dwudziestu procent respondentów wskazało na większą produktywność podczas pracy zdalnej niż w środowisku biurowym (Aguilera i in.). Ponadto, według badania autorstwa Wrigh z 2021 roku, osoby pracujące zdalnie były bardziej produktywne (około 80%), nawet gdy zachorowały. Z kolei raport PWC wykazał wzrost produktywności w przypadku dłuższej pracy zdalnej. Jednakże, w innych dostępnych badaniach dotyczących pracy zdalnej i hybrydowej, poziom produktywności nie uległ zmianie wraz z przejściem z biura do domu (Awada i in., 2021) lub nawet zmniejszył się (Toscano i in., 2020; Monteiro i in., 2021).

Satysfakcja z pracy zdalnej. Chęć pracowników do przejścia na tryb zdalny jest również niejasna, ponieważ niewielu z nich uważa skrócony czas dojazdu i większą równowagę między życiem zawodowym a prywatnym jako zachętę do preferowania pracy zdalnej (Mokhtarian i in. 2004; Bailey i Kurland, 2002). Niemniej jednak, pracownicy często uważają, że to ich kierownictwo nie pozwala na pracę zdalną (Aguilera). Pracownicy pracujący zdalnie mają tendencję do utraty satysfakcji w porównaniu z tradycyjnymi pracownikami przed pandemią COVID-19 (Parker et al., 2021). Zatarty czas i brak oddzielającej pracę i życie prywatne granicy zwiększają niepokój wśród

pracowników zdalnych, godziny pracy i skupienie się na zadaniach zawodowych (Fukumura i in., 2021). Na tej podstawie można sformułować pierwszą hipotezę badawczą:

H1: Brak satysfakcji z pracy zdalnej będzie częściej charakteryzował osób ze zmniejszoną produktywnością.

Izolacja socjalna jest innym istotnym wyzwaniem w pracy zdalnej. Autonomia podczas pracy hybrydowej i zdalnej prowadzi do zwiększonego poczucia osamotnienia i stresu (Wang i in., 2020). Związek między izolacją społeczną a produktywnością w warunkach pracy zdalnej został przeanalizowany w badaniu przeprowadzonym przez Goldeną i in., którzy zwrócili uwagę na negatywny wpływ izolacji zawodowej na produktywność, ocenianą przez przełożonych pracowników. Z tego postawiona druga hipoteza:

H2: Osoby o mniejszej produktywności będą częściej odczuwać samotność i izolację społeczną w pracy zdalnej lub hybrydowej.

Wiek. Starsze badania wykazywały pewną niezależność wieku od poziomu produktywności (Barrett i in., 1985), podczas gdy nowsze badania sugerują, że wydajność młodszych i starszych pracowników może zmieniać się w trakcie ich kariery (Alessandri i Borgogni, 2015). Kobiety i pracownicy w starszym wieku byli bardziej produktywni podczas pracy zdalnej również w badaniu Awady i in. Na podstawie powyższego, postawiona została trzecia hipoteza badawcza:

H3: Produktywność osób starszych wzrośnie w pracy zdalnej.

Doświadczenie. Jak sugerowano, poziom wydajności zazwyczaj ulega zmianie na etapie przejściowym, kiedy pracownik zaczyna na stanowisku lub w nowej firmie zatrudniającej, podczas gdy jest bardziej stabilny w fazie utrzymania, kiedy pracownik już dobrze nauczył się wymagań związanych z głównymi obowiązkami. (Murphy, 1989).

H4: Osoby mniej doświadczeni będą charakteryzowały się mniejszą produktywnością w pracy zdalnej.

Płeć, dochód, stanowisko. Jeśli chodzi o cechy społeczno-demograficzne, wcześniejsze badania wykazały, że płeć, dochód, wykształcenie i obecność dzieci miały wpływ na produktywność i inne zachowania podczas pracy zdalnej, oraz że osoby z wyższym wykształceniem, o wysokich dochodach i biali pracownicy byli bardziej skłonni do przejścia do pracy z domu (Bick i in., 2020). Ci sami autorzy stwierdzili również stosunkowo większą utratę pracy przez kobiety podczas pandemii, co wiąży się z utratą produktywności. Yassenov (2020) stwierdził, że pracownicy o niższych zarobkach mieli nawet trzykrotnie mniejsze szanse na możliwość pracy z domu niż pracownicy o wyższych zarobkach. Osoby z niższym poziomem wykształcenia, mniejszości etniczne i imigranci byli również skoncentrowani na zawodach, których wykonywanie z domu jest mniej prawdopodobne. Niemniej jednak, należy zachować ostrożność w uogólnieniach wpływu rys demograficznych, ponieważ sytuacja może się znacznie różnić, ze względu na różne konteksty społeczno-polityczne i kulturowe.

H5: produktywność kobiet podczas pracy zdalnej jest mniejsza.

Sektor produkcji. Kramer i Kramer (2020) zbadali status zawodowy w kontekście pracy z domu i doszli do wniosku, że wyzwania dla każdego sektora produkcyjnego mogą być różne, tzn. każdy sektor może doświadczać różnych skutków pandemii COVID-19. Na podstawie tego wniosku, postawiona została szósta i ostatnia hipoteza badawcza:

H6: Produktywność pracowników w pracy zdalnej zależy od sektora produkcyjnego i różni się w zależności od konkretnej branży lub sektora, w którym pracują.

Należy dodać, że zbadano wiele innych czynników, poprawiających produktywność, jak np. temperatura w pomieszczeniu, brak małych dzieci i wygodne miejsce do pracy (Awada i in.). Jednak analiza większego zbioru zmiennych objaśniających wymagałaby większej liczby obserwacji i wychodziłoby za możliwości oferowane przez otwarte zbiory danych. Tym niemniej, jak wspomniano we wstępie i jak wynika z przeglądu literatury, oceny wpływu poszczególnych czynników często różni się w zależności od badania, nawet dla dość jednoznacznych charakterystyk, jak wiek i płeć. Co więcej, przeanalizowany zostanie czynnik, związany z doświadczeniem firmy zatrudniającej we wprowadzeniu pracy hybrydowej i zdalnej, i jego wpływ na produktywność pracownika, co wcześniej nie było rozstrzygane w żadnym z przeglądanych badań. Z powyższymi hipotezami możliwe jest opracowanie problematyki badania, wstępnego opisu zebranych zmiennych objaśniających oraz metodologii badawczej.

II. Opis danych

2.1 Uczestnicy i procedura

Rozważany zbiór danych zawiera wyniki ankietowania przeprowadzone wśród 1500 pracowników hybrydowych i zdalnych na zlecenie stanu Nowa Południowa Walia (NSW) w Australii w sierpniu i wrześniu 2020 r. Ankietowanie miało na celu uchwycenie doświadczeń związanych z pracą zdalną i hybrydową w szczytowym okresie kwarantanny COVID-19. Z pierwotnej próby 1500 odpowiedzi, 250 nie spełniało kryterium włączenia; dotyczyło to odpowiedzi niepracujących zdalnie lub hybrydowo, respondentów o zadeklarowanym wieku powyżej 100 lat, oraz osób, których łączny czas spędzony na różne kategorie aktywności przekroczył 24 godz. Następnie, dla każdej jednostki badanej najczęstsze odpowiedzi na serię pytań: *From the following, please select the most significant barrier to doing your work remotely; Compare remote working to working at your employer's workplace. Select the best(worst) aspect of remote working for you*, zostały przedstawione jako największe wyzwania i benefity wynikające z pracy. Pełny zarys dotyczący wstępnego przygotowania danych został opisany na stronie internetowej github¹.

2.2 Mierniki

Zmienne osobiste. Ten zestaw zmiennych dotyczy wieku, płci, i typu gospodarstwa domowego jednostki badanej. Zbiór danych składa się z 551 kobiet i 699 mężczyzn. Charakterystyka typu gospodarstwa domowego obejmuje 4 poziomy: małżeństwo z dziećmi na utrzymaniu (453 os.),

¹ https://github.com/yauheni-se/RemoteWorkSurvey/blob/main/data_preparation.html, dostęp 05.05.2023

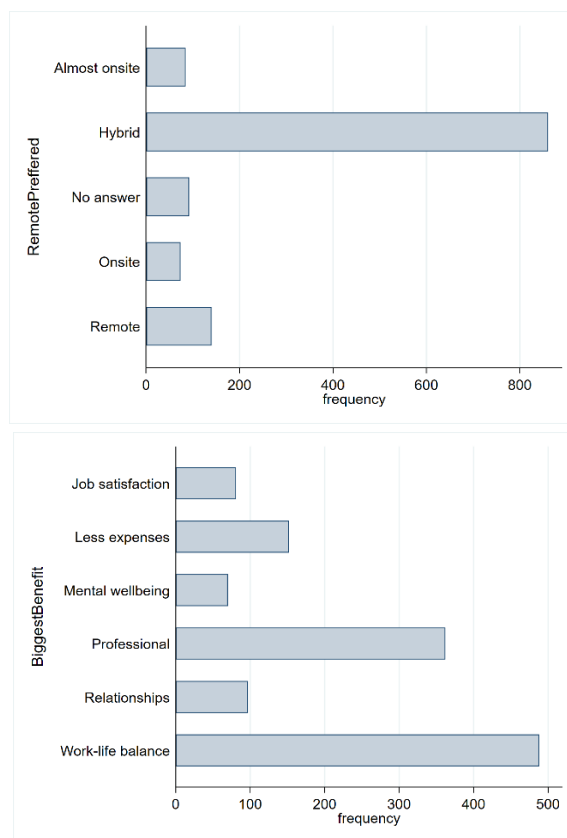
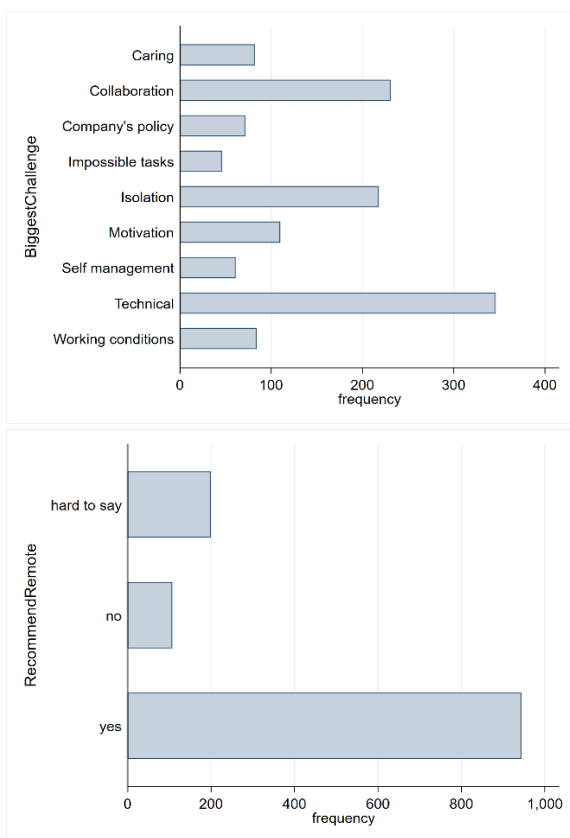
małżeństwo bez dzieci (435 os.), singli (214 os.) oraz pozostałe (148 os.): osoby wychowujące dziecko samodzielnie, rodziny nienuklearne i inne. Średni wiek respondentów wynosił 45 lat, z odchyleniem standardowym wynoszącym 11.6 lat. Najmłodsza osoba miała 19 lat, a najstarsza 83 lata. Warto jednak zauważyć, że rozkład kategorii wiekowych przedstawiony jest nierównomiernie: tylko 10% respondentów miało mniej niż 30 lat. Może to być powiązane z większym poziomem bezrobocia wśród osób w najmniejszym wieku produkcyjnym, które nie brały udziału w ankietowaniu.

Zmienne profesjonalne. Ten zestaw zmiennych dotyczy cech profesjonalnych jednostek badanych, takich jak stanowisko, doświadczenie i tryb pracy zdalnej respondenta. Rozkład zmiennej opisującej stanowisko składa się z czterech kategorii: urzędnik (184), menedżer (411), pracownik profesjonalny, , czyli osoba pracująca na stanowisku wymagającym umiejętności zawodowych (502) oraz inne (153). Zmienna doświadczenia charakteryzuje się trzema możliwymi odpowiedziami: 6-12 miesięcy (121 os.), od 1 roku do 5 lat (472 os.) oraz więcej niż 5 lat (657 os.). Należy jednak zauważyć, że doświadczenie może być silnie skorelowane z wiekiem respondenta. Większość badanych osób (657 odpowiedzi) pracowała w trybie hybrydowym, czyli spędzała 10 lub więcej procent swojego czasu pracy w domu. 470 osób spędzało 100% swojego czasu pracy w domu, natomiast pozostali (123 os.) pracowali z domu w mniej niż 10% przypadków.

Zmienne związane z firmą zatrudniającą, czyli: rozmiar przedsiębiorstwa, sektor produkcyjny, położenie (miasto (1016 odp.) lub regionalne (234 odp.)), dostępność pracy zdalnej, wsparcie firmy w kwestii pracy zdalnej oraz standardy związane z praktykami pracy zdalnej (czy praca zdalna zwykle się stosuje w firmie). Dwie trzecie respondentów pracowało w dużych firmach (615 os.), jedna trzecia w firmach liczących 19-199 pracowników (321 os.), a pozostała trzecia to firmy liczące mniej niż 19 pracowników (w tym osoby prowadzące działalność gospodarczą, 314 osoby). Sektor produkcyjny składał się z sektora pierwotnego (84 osoby), wtórnego (147 osoby) oraz trzeciorzędowego podzielonego na dwie podkategorie: sektor świadczeń usług bezpośrednio związanych z ludźmi, czyli quinary (552 osoby) (np. policjanci, lekarze, sprzedawcy, prawnicy) sektor świadczeń usług mających pośredni kontakt z ludźmi, czyli quaternary (467 osoby) (np. programiści i finansjści). 770 respondenci odpowiedziało, że ich firma zatrudniająca zezwala na pracę zdalną, 229 osób odpowiada, że nie, a 251 nie dało jednoznacznej odpowiedzi. Podobny jest rozkład opinii odnośnie wsparcia pracy zdalnej/hybrydowej: 721 odpowiedź tak, 229 odpowiedzi no oraz 300 opinii mieszanych. Już na etapie wstępnej analizy widać dość mocną korelację obu wymienionych powyżej zmiennych. Wśród osób badanych, 963 twierdzi, że praca zdalna lub hybrydowa jest rozpowszechniona; z kolei, 149 respondentów deklaruje, że nie jest to standardową praktyką, oraz 138 nie miało konkretnej opinii.

Zmienne opisujące preferencje: do tych zmiennych należy preferowany tryb pracy, polecenie bieżącego trybu pracy, największe wyzwanie w pracy zdalnej/hybrydowej oraz największa korzyść. Najlepiej te zmienne można opisać za pomocą wykresów:

Rysunek 2.2.1. Rozkłady zmiennych opisujących preferencje.

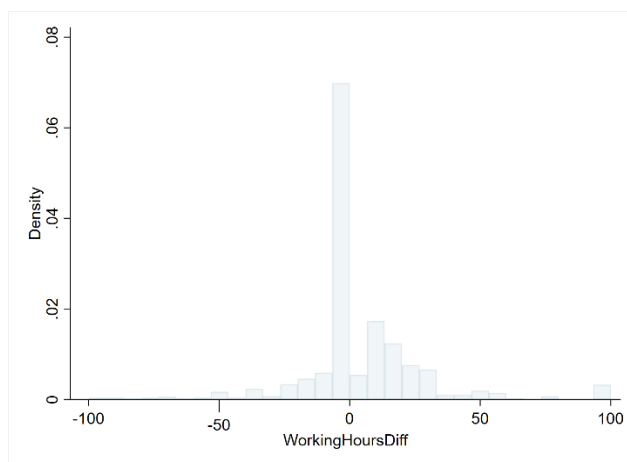


Zrodlo: opracowanie własne

Największym korzyścią dla pracowników, wynikającą ze stosowania pracy zdalnej lub hybrydowej, jest poprawa równowagi między pracą zawodową a życiem prywatnym. Przeważająca większość pracowników poleca również swój bieżący tryb pracy, czyli pracę hybrydową. Ostatecznie, największym wyzwaniem dla tych pracowników były problemy techniczne związane z oprogramowaniem, połączeniami internetowymi i innymi czynnikami.

Odrębnie znajduje się procentowa różnica ilości godzin pracy w domu w porównaniu do ilości godzin pracy w miejscu zatrudnienia. Średnia zmiana procentowa wynosiła 5.83% z odchyleniem standardowym wynoszącym 24.47%. Tylko 10% respondentów odnotowało zmniejszenie liczby przepracowanych godzin. W związku z tym, stworzony zostanie model oparty wyłącznie na zmiennej workhoursdiff, aby zbadać, czy zwiększona produktywność nie jest jedynie efektem zwiększenia liczby przepracowanych godzin.

Rysunek 2.2.2. Rozkład zmiennej workhoursdiff.



Zródło: opracowanie własne

Ten zbiór zmiennych odnosi się do sposobu pomiaru produktywności pracowników związanych z praktykami pracy zdalnej. Produktywność jest mierzona jako zwiększona, zmniejszona lub taka sama wielkość pracy i jakość pracy wykonanej na podstawie opinii samego respondentów. 221 osoba stwierdziła obniżoną produktywność, 369 osób stwierdziło tę samą wydajność, a 660 osób stwierdziło zwiększoną produktywność.

III. Metoda badania

Produktywność a jej związek z cechami respondentów w niniejszym badaniu modeluje się za pomocą modeli uporządkowanego modelu logitowego. Ten model jest rozszerzeniem standardowego binarnego modelu logitowego i jest specjalnie skonstruowany dla zmiennych, w których dane mają postać kategorii dyskretnych, ale z wyraźnym porządkiem. W przypadku produktywności można z pewnością stwierdzić, że ma ona jawny porządek, skoro poziom 'dropped' < 'same level' < 'increased'. Natomiast trudno było odzyskać przykłady badań, w których zastosowano uogólniony model logitowy do mierzenia produktywności; zwyczajnie, używano modeli binarnych. Model logitowy został użyty zamiast modelu probitowego z powodu łatwiejszej interpretacji efektów krańcowych modelu. Na podstawie dobranego modelu, przetestowanie zostanie założenie regresji równoległych, które jest fundamentem modelu uporządkowanego logitowego. W przypadku naruszenia założenia regresji równoległych porównywane zostaną model uogólniony logitowy uporządkowany oraz model częściowo proporcjonalnych szans.

IV. Analiza

4.1 Istotność zmiennych i redukcja kategorii.

Niniejsza analiza rozpoczyna się od przeglądu wstępnej istotności zmiennych kategoryalnych w kontekście zmiennej zależnej - produktywności. W tym celu wykorzystano test niezależności χ^2 do stwierdzenia, czy zmienne są istotne statystycznie (przy odrzuceniu hipotezy zerowej). Do potwierdzenia siły związku między zmiennymi nominalnymi zastosowano statystykę V-Kramera, która jest miarą siły związku między dwoma zmiennymi nominalnymi (im większa wartość V-

Kramera, tym silniejszy związek między zmiennymi). Poniższa tabela prezentuje wyniki testów niezależności zmiennych objaśniających ze zmienną objaśnianą (productivity):

Tabela 4.1.1. Istotność zmiennych kategoryalnych

Zmienna	Wartość p w χ^2	V-Kramera
gender	0.04	0.07
occupation	0.06	0.07
companysize	0.43	0.04
household	0.09	0.07
experience	0.17	0.05
transportation	0.78	0.02
remote	0.00	0.10
remoterecommend	0.00	0.20
remotepreferred	0.00	0.20
sector	0.02	0.08
companyallowchoice	0.07	0.06
companycommonremote	0.02	0.07
companysupportremote	0.02	0.08
biggestchallenge	0.04	0.11
biggestbenefit	0.02	0.10

Źródło: opracowanie własne

Jak wskazują wyniki zaprezentowane w tabeli, są wystarczająco dowodów, aby stwierdzić, że istnieje statystycznie istotny związek między zmiennymi: gender, remote, remoterecommend, remotepreferred, sector, companycommonremote, companysupportremote, biggestchallenge, biggestbenefit, a produktywnością pracowników. Dla tych zmiennych, których związek z produktywnością był uznawany za nieistotny, dokonano redukcji kategorii. Zmienne isnew, isfuturechoice i ismanager odnoszą się odpowiednio do czasu przepracowanego przez pracownika w firmie, możliwości pracy zdalnej w przyszłości i statusu menadżerskiego pracownika. Zmienna household przyjmuje wartości 1 gdy typ gospodarstwa domowego jest nietypowy (rodzina nienuklearna, 1 rodzic itp.), wartość 2 gdy jest to małżeństwo z dzieckiem, i 0 w pozostałych przypadkach. Rezultaty testu istotności i V-Kramera po przekodowaniu są podane poniżej:

Tabela 4.1.2. Istotność zmiennych kategoryalnych po redukcji kategorii

Zmienna	Wartość p w χ^2	V-Cramera
ismanager	0.01	0.09
household2	0.03	0.07
isfuturechoice	0.11	0.06
isnew	0.05	0.07

Okazało się, że trzy z czterech pierwotnie nieistotnych zmiennych mają związek z produktywnością po redukcji kategorii. Od tego momentu etap transformacji zmiennych mógłby być skończony. Jednak należy zwrócić uwagę na fakt, że liczba unikalnych kombinacji zmiennych kategoryalnych w rozważanym zbiorze danych jest nadal duża: $2 \cdot 3 \cdot 3 \cdot 5 \cdot 4 \cdot 3 \cdot 3 \cdot 9 \cdot 6 \cdot 2 \cdot 3 \cdot 2 \approx 2,1$ mln. Biorąc pod uwagę fakt, że zalecana liczba obserwacji w każdej podgrupie jest równa 30, potrzebne byłoby 60,2 mln obserwacji, żeby włączyć wszystkie wymienione zmienne do modelu. Z tego powodu zastosowano procedurę redukującą liczbę kategorii w zmiennych kategoryalnych, wyróżniając tylko te kategorie, które były statystycznie istotne w modelu uporządkowanym logitowym. Zmienne, które po redukcji wciąż były nieistotne, zostały wykluczone z finalnego modelu.

W wyniku powyższej procedury otrzymano model, składający ze zmiennych: **remoterecommend**, przyjmująca wartości tak lub nie; **remotepreferred**, przyjmująca wartości onsite, hybrid oraz remote; **islackmotivation**, która oznacza respondentów wyróżniających wśród największych wyzwań brak motywacji do pracy; **benefit**, która oznacza respondentów wyróżniających wśród największych benefitów mniej wydatków (1) korzyści profesjonalne lub poprawa równowagi między życiem zawodowym a prywatnym (2) lub inne (0); **isquaternary**, odpowiadająca za to czy pracownik działa w czwartorzędomy sektorze produkcyjnym. Daje to 72 unikalnych podgrupy, średnio po 17 obserwacji w każdej. Jest to bardziej zadowalający poziom niż przed redukcją poziomów, dlatego możliwe jest przejście do analizy istotności zmiennych numerycznych.

4.2 Istotność zmiennych numerycznych

Zaprezentowany zbiór danych zawiera dwie zmienne numeryczne: 'age' oraz 'workinghoursdiff'. Ze względu na to, że zmienne nie mają rozkładu normalnego, zamiast testu ANOVA wykorzystany został test Kruskal'a-Wallis'a. Wartość p dla zmiennej workinghoursdiff wyniosła 0.00, co oznacza, że średnia procentowa zmiana godzin pracy różni się statystycznie dla różnych poziomów produktywności. Innymi słowy, występuje statystycznie istotna relacja między tymi zmiennymi. Natomiast w przypadku zmiennej 'age', wartość p wyniosła 0.44, co wskazuje na to, że wiek pracownika nie ma istotnego wpływu na jego produktywność. Następnie, zmienna 'workinghoursdiff' została włączona do powyżej zdefiniowanego modelu uporządkowanego logitowego, ponieważ według testu istotności ta zmienna wpływa na kształtowanie produktywności. Dalsza część analizy dotyczyła testowania wybranego modelu z wymienionymi zmiennymi objaśniającymi.

4.3 Testowanie modelu

Przed interpretacją wyników modelu uporządkowanego logitowego, należy przetestować jego główne założenie, czyli założenie regresji równoległych/proporcjonalnych szans. W tym celu został zastosowany test Branta. Wyniki procedury są prezentowane poniżej:

Tabela 4.3.1. Wyniki testu Branta

Zmienna	χ^2	$p > \chi^2$	df
---------	----------	--------------	----

wszystkie	18.06	0.021	8
workinghoursdiff	0.04	0.128	1
1.rmtrecommend	2.32	0.251	1
1.rmtpref	1.32	0.077	1
2.rmtpref	3.13	0.265	1
1.islackmotivation	0.26	0.609	1
1.benefit	1.18	0.278	1
2.benefit	1.24	0.265	1
1.isquaternary	0.64	0.425	1

Źródło: opracowanie własne

Znacząca statystyka testowa dla wszystkich zmiennych ($\chi^2 = 18.06$) dostarcza dowodów na to, żeby stwierdzić, że założenie regresji równoległych dla przyjętego poziomu ufności $\alpha = 0.05$ zostało naruszone. Dlatego model logitowy uporządkowany (OL) zostanie porównany z modelem uogólnionym uporządkowanym (GTOM) oraz z modelem częściowo proporcjonalnych szans (POM). Finalna postać modelu zostanie wybrana za pomocą testów ilorazu wiarygodności (LR):

Tabela 4.3.2. Wyniki testów ilorazu szans

Założenie	χ^2	$p > \chi^2$	df
OL zagnieżdżony w GTOM	21.55	0.006	8
OL zagnieżdżony w POM	16.74	0.000	2
POM zagnieżdżony w GTOM	4.81	0.568	6

Źródło: opracowanie własne

Testy LR wykazały, że najbardziej oszczędny jest model z częściowymi proporcjonalnymi szansami. Ponadto, nieistotna statystyka testowa dla testu Walda ($p=0.537$) w modelu POM wskazuje, że ostateczny model nie narusza założenia proporcjonalnych szans/regresji równoległych. Dlatego ostatecznym modelem jest model częściowo proporcjonalnych szans o postaci:

$$\begin{cases} \log\left(\frac{p_1}{1-p_1}\right) = 0.41 + 0.01 \cdot whd + 0.71 \cdot rec + 1.04 \cdot pref_h + 2.52 \cdot pref_r - 0.59 \cdot motiv - 0.6 \cdot bf_{le} - 0.14 \cdot bf_{pb} - 0.3 \cdot quat \\ \log\left(\frac{p_2}{1-p_2}\right) = -0.83 + 0.01 \cdot whd + 0.71 \cdot rec + 1.04 \cdot pref_h + 1.33 \cdot pref_r - 0.59 \cdot motiv - 0.6 \cdot bf_{le} - 0.54 \cdot bf_{pb} - 0.3 \cdot quat, \end{cases} \quad (1)$$

gdzie:

- whd – procentowa różnica liczby godzin przepracowanych z domu do liczby godzin przepracowanych z biura,
- rec – zmienna zero-jedynkowa, oznaczająca czy pracownik poleca pracę zdalną lub hybrydową. Kategoria odniesienia – nie,
- pref – czy pracownik preferuje pracę hybrydową ($pref_h$), czy pracę zdalną ($pref_r$). Kategoria odniesienia – onsite, czyli praca z biura.

- *motiv* – zmienna zero-jedynkowa, mówiąca o tym, czy największym wyzwaniem dla pracownika podczas pracy zdalnej był brak motywacji. Kategoria odniesienia – nie,
- *bf* – czy największą korzyścią dla pracownika były mniejsze wydatki (bf_{le}), czy profesjonalne lub poprawa równowagi między życiem zawodowym a prywatnym (bf_{pb}). Kategoria odniesienia – inne korzyści,
- *quat* – czy pracownik pracuje w czwartorzędowym sektorze produkcyjnym. Kategoria odniesienia – nie,
- *y* – produktywność pracownika przyjmująca poziomy 1 (dropped), 2 (same level) oraz 3 (increased),
- p_1 – prawdopodobieństwo, że $y=2$ (lub $1 - P(y = 1)$),
- p_2 – prawdopodobieństwo, że $y=3$ (lub $1 - P(y = 2)$).

Należy zauważyć, że indywidualne testy istotności dla wszystkich zmiennych objaśniających wykazywały się prawdopodobieństwem nieodrzucenia hipotezy zerowej $p < 0.00$.

4.4 Jakość dopasowania

Następnie oceniana została tabela trafności, miary dokładności oraz wykres krzywej ROC.

Tabela 4.3.3. Tabela trafności

Predicted\Actual	Dropped	Same level	Increased
Dropped	23	11	4
Same level	1	1	1
Increased	197	357	655

Źródło: opracowanie własne

Jak wynika z tabeli, głównym wyzwaniem w przypadku zbudowanego modelu jest poprawne sklasyfikowanie obserwacji, dla których produktywność nie uległa zmianie. Tylko jedna spośród 369 obserwacji została poprawnie zaklasyfikowana przez wybrany model. Najlepsze wyniki osiągnięto w przypadku trzeciej klasy, czyli gdy produktywność wzrosła. Potwierdzeniem takiego wniosku są miary dopasowania, które zostały przedstawione w poniższej tabeli

Tabela 4.3.3. Miary dopasowania

Miara	Dropped	Same level	Increased	Ogółem
Trafność	0.545	0.500	0.53	0.543
Czułość	0.104	0.003	0.992	–
Swoistość	0.985	0.997	0.006	–
AUC	0.667	0.6263	0.784	0.692

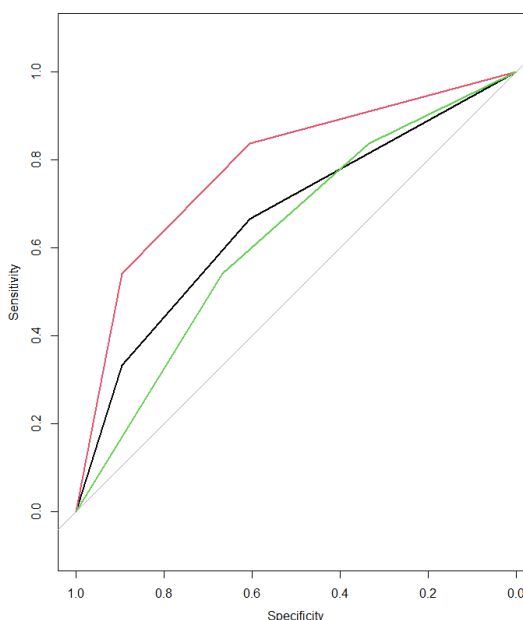
Źródło: opracowanie własne

Pod względem trafności dla kategorii "same level", wybrany model nie różni się zasadniczo od modelu losowego. Jednak, biorąc pod uwagę miarę AUC (ang. area under curve), zwłaszcza dla kategorii "increased", model wykazuje dość dobrą skuteczność w prognozowaniu. Warto zauważyć,

że gdyby celem badania było osiągnięcie jak najwyższej trafności, należałoby rozważyć inne modele, takie jak modele oparte na drzewach decyzyjnych. Jednak, ponieważ niniejsze badanie naśladuje cel ujawniania właściwości, mających statystycznie istotny związek z produktywnością, opartych na mocnej wiedzy teoretycznej, kwestie trafności można pominąć.

Ostateczną miarą dopasowania są wykresy krzywej ROC. Na czerwono zaznaczona jest krzywa ROC w przypadku, gdy występują tylko dwa poziomy zmiennej y: ‘dropped’ oraz ‘increased’; na zielono – gdy y przyjmuje poziomy ‘same level’ i ‘increased’; na czarno – gdy y przyjmuje poziomy ‘same level’ oraz ‘dropped’. Wykresy potwierdzają wnioski z badania miar dopasowania: najbardziej zbliżona do lewego górnego rogu jest krzywa dla kategorii trzeciej.

Rysunek 4.3.1. Krzywe charakterystyk operacyjnych odbiorników (ROC)



Źródło: opracowanie własne

4.5 Interpretacja modelu. Efekty krańcowe.

Ostatnim etapem niniejszej analizy jest interpretacja efektów krańcowych modelu. W tym celu należy zastosować odpowiednią transformację wykładniczą współczynników z równania (1). Przykładowo, wzrost różnicy liczby przepracowanych godzin zdalnie w stosunku do biurowych o 1% zwiększa szansę tego, że produktywność pracownika zostanie na tym samym poziomie, zamiast spaść, o $\exp(0.01) = 1\%$. To samo dotyczy szansy zwiększenia produktywności, w porównaniu z poziomem niezmiennym (wzrost procentowej różnicy godzin o 1% zwiększa tę szansę o 1%). Oznacza to, że zwiększenie godzin pracy ma dość mały wpływ na zwiększenie produktywności. Wśród osób, które polecają pracę zdalną, szansa na zwiększenie poziomu produktywności (w porównaniu do poziomu niezmiennego) jest większa o $\exp(0,71) = 103\%$, przy założeniu stałych pozostałych czynników.

Szansa na zwiększenie produktywności (w porównaniu do poziomu niezmiennego) jest również większa u osób, które preferują pracę zdalną, a nie z biura o $\exp(1.33) = 278\%$. Ten wpływ jest nieco mniejszy dla osób preferujących pracę hybrydową ($\exp(1.04)=182\%$), co świadczy o tym, że pracę zupełnie zdalną preferują najbardziej produktywni pracownicy. Warto zaznaczyć, że wpływ tego czynnika jest jeszcze większy w przypadku porównywania szansy produktywności na tym samym poziomie w porównaniu do produktywności zmniejszonej (wzrost o 1142%). Ponadto, szansa zwiększenia produktywności u osób, które odczuwają brak motywacji jako największe wyzwanie podczas pracy zdalnej, była o 55% mniejsza.

Wymienienie korzyści wynikających z pracy zdalnej, takich jak benefit profesjonalny i poprawa równowagi między życiem zawodowym a osobistym, skutkowało spadkiem szansy na zwiększenie lub utrzymanie poziomu produktywności (odpowiednio o $\exp(-0,54) = 58\%$ i $\exp(-0,14) = 87\%$). Oznacza to, że osoby, które były najbardziej produktywne, częściej wymieniały inne benefity niż te związane z pracą zawodową, np. wyższe poczucie satysfakcji. Należy także zwrócić uwagę na korzyść "less expenses": osoby, które wymieniły niższe wydatki jako główną korzyść z pracy zdalnej, miały mniejszą szansę na zwiększenie swojej produktywności (średnio o 54,8%). "Mniejsze wydatki" były jedynym korzyścią materialną, dlatego można to interpretować w taki sposób, że osoby, które nie odnosiły korzyści niematerialnych, miały większą szansę na spadek produktywności. Ostatnim czynnikiem jest sektor produkcyjny. Model wskazuje, że osoby pracujące w czwartorzędnym sektorze produkcyjnym (finansisci, pracownicy branży IT, R&D, media itp.) miały o 74% mniejsze szanse na zwiększenie lub utrzymanie swojej produktywności niż pracownicy z innych sektorów. Jest to dość zaskakujące, ponieważ naturalnie można oczekiwać spadku produktywności dla osób pracujących w sektorach takich jak rolnictwo czy produkcja przemysłowa, która wymaga fizycznej obecności w miejscu pracy. Jednak można zauważyć, że liczba osób pracujących w I i II sektorze była niewielka, w związku z tym osoby z IV i V sektora (human resources, urzędnicy, konsultanci, itp.) były bezpośrednio porównywane. W ten sposób okazuje się, że produktywność osób współpracujących z ludźmi, była większa podczas pracy zdalnej. Może to być spowodowane mniejszymi czynnikami negatywnymi po interakcji z ludźmi (np. mniejsza irytacja). Co do innych zmiennych, takich jak stosunek firmy do pracy zdalnej, wiek, płeć pracownika, stanowisko czy doświadczenie, nie miały one istotnego wpływu na poziom produktywności pracowników podczas pracy zdalnej w stanie Nowa Południowa Walia.

Dyskusja

Podsumowanie i zastosowanie

Po przeprowadzonej analizie okazało się, że tylko dwie z sześciu hipotez, zakładanych na początku, zostały potwierdzone (różny poziom produktywności w zależności od sektora oraz satysfakcja z pracy zdalnej). Częściowo wzrost produktywności był spowodowany większym czasem pracy, co nie zostało uwzględnione na etapie przeglądu literatury. Jednak najbardziej wpływowymi czynnikami okazały się preferencje pracownika w zakresie trybu pracy oraz sugestie lub jej brak co do pracy zdalnej. Zmienne socjalno-demograficzne nie miały statystycznie istotnego wpływu na produktywność, co tylko utwierdza niejednoznaczną rolę tych czynników. Ponadto, czynnik podejścia

firmy zatrudniającej do pracy zdalnej też nie okazał istotnego wpływu. Niniejsze badanie może pomóc w ułatwieniu wprowadzenia systemu hybrydowego (zdalnego) zarówno w przedsiębiorstwach prywatnych, jak i państwowych. Szczególną uwagę trzeba zwrócić na motywację pracowników, umożliwienie wyboru trybu pracy (hybrydowy czy całkowicie zdalny) oraz na korzyści wynikające dla pracowników w konkretnych przypadkach zastosowania.

Ograniczenia

Niniejsze badanie posiada kilka poważnych ograniczeń. Po pierwsze, zostały rozpatrzone jedynie obserwacje ze stanu Nowa Południowa Walia. Jest to tylko jeden z 16 stanów Australii i zawiera koło 1/3 całej populacji kraju. W tym przypadku próba danych może być niereprezentatywna ze względu na specyficzny profil ekonomiczny tego regionu. Po drugie, badanie nie uwzględnia osób, które nie pracowały w momencie badania, ale mogą mieć istotny wpływ na kształtowanie rynku pracy, np. studenci poszukujący pierwszej pracy. Po trzecie, ponieważ Australia i Polska różnią się pod względem ekonomicznym, socjalno-demograficznym i geograficznym, wyniki tego badania nie mogą być uogólniane na inne kraje, a w szczególności na Polskę. Charakterystyki te mogą być specyficzne tylko dla Australii lub regionu Oceanii, a ponadto Australia jest jednym z najbardziej rozwiniętych krajów na świecie, co stanowi znaczące odróżnienie w porównaniu z Polską. Z tego może wynikać, że największe korzyści i benefity mogą się różnić. W końcu, produktywność była mierzona przez respondentów, co implikuje pewien stopień obciążalności wyników. Pracownicy mogli zawyżać poziom swojej produktywności, aby uniknąć utraty pracy lub braku podwyżki.

Kierunki dalszych badań

Dalsze badania będą skupiać się głównie na poprawieniu ograniczeń, o których mowa w poprzednim rozdziale. Po pierwsze, warto przeprowadzić badania na poziomie całej Australii, aby uzyskać bardziej reprezentatywne wyniki, dotyczące wpływu pracy zdalnej na produktywność. Po drugie, warto rozważyć pomiar produktywności za pomocą średniej ważonej ocen zarówno pracowników,

jak i pracodawców, aby uzyskać szerszą perspektywę na temat wpływu pracy zdalnej na wyniki produktywności. Po trzecie, warto również przeprowadzić badania na danych po pandemii COVID-19, aby zaobserwować zachowania pracowników w normalnych warunkach i sprawdzić możliwe średnio- i długoterminowe skutki pracy zdalnej na produktywność pracowników.

Bibliografia

1. Aguilera A., Lethiais V., Rallet A., Proulhac L., *Home-based telework in France: characteristics, barriers and perspectives*, Transportation Research: Part A, Wyd. 92, 2016.
2. Alessandri, G., & Borgogni, L., *Stability and change of job performance across the career span*, Human Performance, Wyd. 28, 2015.
3. Alexander, R.A., Barrett, G.V., Caldwell, M.S., *The concept of dynamic criteria: A critical reanalysis*. Personnel Psychology, Wyd. 38, 1985.

4. Awada, M., Becerik-Gerber, B., Lucas, G.M., Roll, S., *Working from home during the COVID-19 pandemic: impact on office worker productivity and work experience*, Work, Wyd. 69, 2021.
5. Battisti E., Alfiero S., Leonidou E., *Remote working and digital transformation during the COVID-19 pandemic: Economic-financial impacts and psychological drivers for employees*, Journal of Business Research, Wyd. 150, 2022.
6. Bailey, D.E., Kurland, N.B., *A review of telework research: findings, new directions, and lessons for the study of modern work*, Behavior Wyd. 23, 2002.
7. Baudot L., Kelly K., *Work From Home During the COVID-19 Outbreak*, University of Central Florida, Florida, 2020.
8. Becerik-Gerber, B., Fukumura, Y.E., Lucas, G.M., Roll, S.C., Schott, J.M., *Negotiating time and space when working from home: experiences during COVID-19*, OTJR: Occupation, Participation and Health, Wyd. 41, 2021.
9. Bick, A., Blandin, A., Mertens, K., *Work from Home After the COVID-19 Outbreak*, CEPR Discussion Papers No. 15000, 2020.
10. Fan W., *Working More, Less or the Same During COVID-19? A Mixed Method*, University of Minnesota-Twin Cities, Minneapolis 2022.
11. Gavoiile N., Hazans M., *Personality traits, remote work and productivity*, Research Council of Lithuania, 2022 <https://docs.iza.org/dp15486.pdf> (dostęp 03.05.2023)
12. Golden, T.D., Dino, R.N., Veiga, J.F., *The Impact of Professional Isolation on Teleworker Job Performance and Turnover Intentions: Does Time Spent Teleworking, Interacting Face-to-Face, or Having Access to Communication-Enhancing Technology Matter?*, Journal of Applied Psychology, Wyd. 93, 2008.
13. Hunter P., *Remote working in research*, EMBO Reports, 2019.
14. Isaac B., Ayodeji O., Olanrewaju A., *COVID-19 Pandemic in South Africa: Application of Tobit and Probit Models*, University of Venda, HighTech and Innovation Journal, Wyd. 3, 2022.
15. Kramer, K.Z., Kramer, A., *The potential impact of the Covid-19 pandemic on occupational status, work-from-home, and occupational mobility*, Journal of Vocational Behavior, Wyd. 119, 2020.
16. Mokhtarian P.L., Collantes G.O., C. Telecommuting Gertz, *Residential location, and commute-distance traveled: evidence from state of California employees*, Transportation Research: Part A, Wyd. 36, 2004.
17. Monteiro N.P., Straume O.R., Valente M., *Does remote work improve or impair firm labour productivity? Longitudinal evidence from Portugal*, SSRN Electron, 2021.
18. PWC, *It's time to reimagine where and how work will get done*, <https://www.pwc.com/us/en/library/covid-19/us-remote-work-survey.html> (dostęp 03.05.2023).
19. Tleuken A. i in., *Effects of the residential built environment on remote work productivity and satisfaction during COVID-19 lockdowns: An analysis of workers' perceptions*, Building and Environment, Wyd. 219, 2022.

20. Toscano F., Zappal S., *Social isolation and stress as predictors of productivity perception and remote work satisfaction during the COVID-19 pandemic: the role of concern about the virus in a moderated double mediation*, Times Wyd. 12, 2020.
21. Murphy, K.R. *Is the relationship between cognitive ability and job performance stable over time?*, Human Performance, Wyd. 2, 1989.
22. Parker, K., Horowitz, J.M., Minkin, R., *How the coronavirus outbreak has – and hasn't – changed the way Americans work*, Pew Research Center (dostęp 03.05.2023).
23. Vafin A., *The Impact of Remote Work on Firm's Profitability: Optimizing Virtual Employee Productivity and Operational Costs*, Kazan Federal University, Kazan 2022.
24. Wang, B., Liu, Y., Parker, S.K., Qian, J., *Achieving effective remote working during the COVID-19 pandemic: a work design perspective*, Applied Psychology, Wyd. 69, 2020.
25. Wright, A.D., *Teleworkers more productive even when sick*, <https://www.shrm.org/resourcesandtools/hr-topics/technology/pages/teleworkers-more-productive-even-when-sick.aspx> (dostęp 03.05.2023).
26. Xiao Y., Becerik-Gerber B., Lucas G., Roll C.S., *Impacts of Working From Home During COVID-19 Pandemic on Physical and Mental Well-Being of Office Workstation Users*, Transportation Research Interdisciplinary Perspectives, Wyd. 11, 2021.
27. Yassenov, V., *Who can work-from-home?*, IZA Discussion Paper No. 13197, 2020.

Kod STATA

```
* PRESETS
pwd
cd C:\Projects\RemoteWorkSurvey
insheet using "data/remote_productivity.csv", comma clear
describe
codebook
* PLOTS
splitvallabels productivity
catplot gender productivity, percent(productivity) var1opts(label(labsize(small))) var2opts(label(labsize(small))) relabel("r(relabel)") ytitle("", size(small)) title("", span size(medium)) label(bar, format(%4.1f) size(vsmall)) intensity(25) asyvars plotregion(fcolor(white)) graphregion(fcolor(white))
graph export "C:\Projects\RemoteWorkSurvey\plots\distributions by prod\p1.png", as(png) name("Graph") replace
splitvallabels productivity
catplot occupation productivity, percent(productivity) var1opts(label(labsize(small))) var2opts(label(labsize(small))) relabel("r(relabel)") ytitle("", size(small)) title("", span size(medium)) label(bar, format(%4.1f) size(vsmall)) intensity(25) asyvars plotregion(fcolor(white)) graphregion(fcolor(white))
splitvallabels productivity
catplot companysize productivity, percent(productivity) var1opts(label(labsize(small))) var2opts(label(labsize(small))) relabel("r(relabel)") ytitle("", size(small)) title("", span size(medium)) label(bar, format(%4.1f) size(vsmall)) intensity(25) asyvars plotregion(fcolor(white)) graphregion(fcolor(white))
splitvallabels productivity
catplot household productivity, percent(productivity) var1opts(label(labsize(small))) var2opts(label(labsize(small))) relabel("r(relabel)") ytitle("", size(small)) title("", span size(medium)) label(bar, format(%4.1f) size(vsmall)) intensity(25) asyvars plotregion(fcolor(white)) graphregion(fcolor(white))
splitvallabels productivity
catplot experience productivity, percent(productivity) var1opts(label(labsize(small))) var2opts(label(labsize(small))) relabel("r(relabel)") ytitle("", size(small)) title("", span size(medium)) label(bar, format(%4.1f) size(vsmall)) intensity(25) asyvars plotregion(fcolor(white)) graphregion(fcolor(white))
splitvallabels productivity
catplot transportation productivity, percent(productivity) var1opts(label(labsize(small))) var2opts(label(labsize(small))) relabel("r(relabel)") ytitle("", size(small)) title("", span size(medium)) label(bar, format(%4.1f) size(vsmall)) intensity(25) asyvars plotregion(fcolor(white)) graphregion(fcolor(white))
splitvallabels productivity
catplot remote productivity, percent(productivity) var1opts(label(labsize(small))) var2opts(label(labsize(small))) relabel("r(relabel)") ytitle("", size(small)) title("", span size(medium)) label(bar, format(%4.1f) size(vsmall)) intensity(25) asyvars plotregion(fcolor(white)) graphregion(fcolor(white))
splitvallabels productivity
catplot recommendremote productivity, percent(productivity) var1opts(label(labsize(small))) var2opts(label(labsize(small))) relabel("r(relabel)") ytitle("", size(small)) title("", span size(medium)) label(bar, format(%4.1f) size(vsmall)) intensity(25) asyvars plotregion(fcolor(white)) graphregion(fcolor(white))
splitvallabels productivity
catplot remotepreferred productivity, percent(productivity) var1opts(label(labsize(small))) var2opts(label(labsize(small))) relabel("r(relabel)") ytitle("", size(small)) title("", span size(medium)) label(bar, format(%4.1f) size(vsmall)) intensity(25) asyvars plotregion(fcolor(white)) graphregion(fcolor(white))
splitvallabels productivity
catplot sector productivity, percent(productivity) var1opts(label(labsize(small))) var2opts(label(labsize(small))) relabel("r(relabel)") ytitle("", size(small)) title("", span size(medium)) label(bar, format(%4.1f) size(vsmall)) intensity(25) asyvars plotregion(fcolor(white)) graphregion(fcolor(white))
splitvallabels productivity
catplot companyallowchoice productivity, percent(productivity) var1opts(label(labsize(small))) var2opts(label(labsize(small))) relabel("r(relabel)") ytitle("", size(small)) title("", span size(medium)) label(bar, format(%4.1f) size(vsmall)) intensity(25) asyvars plotregion(fcolor(white)) graphregion(fcolor(white))
splitvallabels productivity
catplot companycommonremote productivity, percent(productivity) var1opts(label(labsize(small))) var2opts(label(labsize(small))) relabel("r(relabel)") ytitle("", size(small)) title("", span size(medium)) label(bar, format(%4.1f) size(vsmall)) intensity(25) asyvars plotregion(fcolor(white)) graphregion(fcolor(white))
splitvallabels productivity
catplot companysupportremote productivity, percent(productivity) var1opts(label(labsize(small))) var2opts(label(labsize(small))) relabel("r(relabel)") ytitle("", size(small)) title("", span size(medium)) label(bar, format(%4.1f) size(vsmall)) intensity(25) asyvars plotregion(fcolor(white)) graphregion(fcolor(white))
splitvallabels productivity
```

```

catplot biggestchallenge productivity, percent(productivity) var1opts(label(labsize(small))) var2opts(label(labsize(small)) relabel("r(relabel)")) ytitle("", size(small)) title("", span size(medium)) blabel(bar,
format(%4.1f) size(vsmall)) intensity(25) asyvars plotregion(fcolor(white)) graphregion(fcolor(white))
splitvallabels productivity
catplot biggestbenefit productivity, percent(productivity) var1opts(label(labsize(small))) var2opts(label(labsize(small)) relabel("r(relabel)")) ytitle("", size(small)) title("", span size(medium)) blabel(bar,
format(%4.1f) size(vsmall)) intensity(25) asyvars plotregion(fcolor(white)) graphregion(fcolor(white))
graph box age, over(productivity) plotregion(fcolor(white)) graphregion(fcolor(white))
graph box workinghoursdiff, over(productivity) intensity(25) plotregion(fcolor(white)) graphregion(fcolor(white))

* X-TESTS
tab productivity gender, chi2 V
tab productivity occupation, chi2 V
tab productivity companysize, chi2 V
tab productivity household, chi2 V
tab productivity experience, chi2 V
tab productivity transportation, chi2 V
tab productivity remote, chi2 V
tab productivity recommendremote, chi2 V
tab productivity remotepreferred, chi2 V
tab productivity sector, chi2 V
tab productivity companyallowchoice, chi2 V
tab productivity companycommonremote, chi2 V
tab productivity companysupportremote, chi2 V
tab productivity biggestchallenge, chi2 V
tab productivity biggestbenefit, chi2 V

* ENCODING
encode productivity, generate(y)
encode gender, generate(gendr)
encode occupation, generate(occ)
encode household, generate(hh)
encode experience, generate(exp)
encode remote, generate(rmt)
encode recommendremote, generate(rmtrecommend)
encode remotepreferred, generate(rmtpref)
encode sector, generate(sctr)
encode companycommonremote, generate(rmtcommon)
encode companysupportremote, generate(rmtsupport)
encode biggestchallenge, generate(challenge)
encode biggestbenefit, generate(benefit)
encode companyallowchoice, generate(rmtallowed)

* RECODING
recode y (1=1) (2=3) (3=2), generate(y2)
recode hh (1=0) (2=0) (3=1) (4=0), generate(hh2)
recode hh (1=0) (2=2) (3=1) (4=0), generate(hh3)
recode occ (1=0) (2=0) (3=0) (4=1), generate(occ2)
recode occ (1=0) (2=1) (3=0) (4=0), generate(occ3)
recode occ (1=0) (2=1) (3=0) (4=2), generate(occ4)
recode challenge (1=0) (2=1) (3=0) (4=0) (5=0) (6=2) (7=0) (8=0) (9=0), generate(challenge2)
recode challenge (1=0) (2=0) (3=0) (4=0) (5=0) (6=1) (7=0) (8=0) (9=0), generate(challenge3)
recode challenge (1=0) (2=0) (3=1) (4=0) (5=0) (6=2) (7=0) (8=0) (9=0), generate(challenge4)
recode challenge (1=0) (2=0) (3=0) (4=0) (5=0) (6=1) (7=0) (8=0) (9=0), generate(challenge5)
recode rmtcommon (1=0) (2=0) (3=1), generate(rmtcommon2)
recode rmtrecommend (1=0) (2=0) (3=1), generate(rmtrecommend2)
recode rmtsupport (1=0) (2=0) (3=1), generate(rmtsupport2)
recode rmtallowed (1=0) (2=0) (3=1), generate(rmtallowed2)
recode rmtpref (1=0) (2=1) (3=1) (4=0) (5=2), generate(rmtpref2)
recode rmt (1=0) (2=0) (3=1), generate(rmt2)
recode sctr (1=0) (2=0) (3=1) (4=0), generate(sctr2)
recode exp (1=0) (2=1) (3=0), generate(exp2)
recode benefit (1=0) (2=1) (3=0) (4=2) (5=0) (6=3), generate(benefit2)

* X-TESTS EXTENDED
tab productivity occ3, chi2 V
tab productivity occ4, chi2 V
tab productivity hh2, chi2 V
tab productivity hh3, chi2 V
tab productivity exp2, chi2 V
tab productivity rmtallowed2, chi2 V

* CLEANING
rename occ3 ismanager
drop gender occupation companysize household experience transportation remote recommendremote remotepreferred productivity sector companyallowchoice companycommonremote
companysupportremote personalhoursdiff domestichoursdiff commutehoursdiff biggestbenefit biggestchallenge y occ occ2 occ4 rmtallowed rmtallowed2 hh hh2
rename hh3 household2
rename exp2 isnew
cls

* NUM TESTS
histogram age
swilk age
histogram workinghoursdiff
kwallis age, by(y2)
kwallis workinghoursdiff, by(y2)
rename y2 y
drop age
cls

* MULTICOLLINEARITY
tab y rmtcommon, chi2 V
tab y rmtcommon2, chi2 V
tab y rmtsupport, chi2 V
tab y rmtsupport2, chi2 V
tab y rmtpref, chi2 V
tab y rmtpref2, chi2 V
tab y rmtrecommend, chi2 V
tab y rmtrecommend2, chi2 V
tab y rmt, chi2 V
tab y rmt2, chi2 V

* REGROUPING PREFERENCES
label define ynlabel 0 "no" 1 "yes"
label values rmtsupport2 ynlabel
label define rmtlabel 0 "onsite" 1 "hybrid" 2 "remote"
label values rmtpref2 rmtlabel

```

```

label values rmtrecommend2 ynlablel
label define rmt2label 0 "hybrid" 1 "remote"
label values rmt2 rmt2label
drop rmt rmtcommon2 rmtpref rmtrecommend rmtsupport
cls
* EXAMING THEIR CORRELATIONS
tab rmtsupport2 rmtcommon, chi2 V
tab rmtpref2 rmt2, chi2 V
rename rmt2 rmt
rename rmtpref2 rmtpref
rename rmtrecommend2 rmtrecommend
rename rmtsupport2 rmtsupport
rename household2 household
tab gendr ismanager, chi2 V
tab gendr household, chi2 V
tab gendr isnew, chi2 V
tab y sctr, chi2 V
tab y sctr2, chi2 V
tab y challenge, chi2 V
tab y challenge2, chi2 V
tab y challenge3, chi2 V
tab y challenge4, chi2 V
tab y challenge5, chi2 V
* FEATURE SELECTION
ologit y workinghoursdiff
ologit y workinghoursdiff i.gendr
ologit y workinghoursdiff i.gendr i.sctr
ologit y workinghoursdiff i.gendr i.rmtcommon
ologit y workinghoursdiff i.gendr i.challenge
ologit y workinghoursdiff i.gendr i.benefit
ologit y workinghoursdiff i.gendr i.household
ologit y workinghoursdiff i.gendr i.ismanager
ologit y workinghoursdiff i.gendr i.rmtrecommend
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtsupport
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtcommon
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.rmt
drop exp2
drop exp
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.sctr
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.sctr2
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.isnew
* FEATURE PREP AFTER FEATURE SEL
codebook sctr
rename sctr2 isquaternary
label values isquaternary ynlablel
drop sctr
drop rmtsupport isnew ismanager
codebook challenge
drop challenge2 challenge3 challenge4 challenge5
recode challenge (1=0) (2=0) (3=1) (4=0) (5=0) (6=2) (7=0) (8=0) (9=0), generate(challenge2)
recode challenge (1=0) (2=0) (3=1) (4=0) (5=0) (6=0) (7=0) (8=0) (9=0), generate(challenge3)
recode challenge (1=0) (2=0) (3=0) (4=0) (5=0) (6=1) (7=0) (8=0) (9=0), generate(challenge4)
codebook household
recode household (1=0) (2=1) (3=0), generate(ishhcomplex)
codebook rmtcommon
recode rmtcommon (1=0) (2=0) (3=1), generate(rmtcommon2)
drop benefit2
codebook benefit
recode benefit (1=0) (2=1) (3=0) (4=2) (5=0) (6=3), generate(benefit2)
recode benefit (1=0) (2=1) (3=0) (4=0) (5=0) (6=0), generate(benefit3)
recode benefit (1=0) (2=0) (3=0) (4=1) (5=0) (6=0), generate(benefit4)
recode benefit (1=0) (2=0) (3=0) (4=0) (5=0) (6=1), generate(benefit5)
recode benefit (1=0) (2=1) (3=0) (4=2) (5=0) (6=0), generate(benefit6)
recode benefit (1=0) (2=0) (3=0) (4=1) (5=0) (6=2), generate(benefit7)
recode benefit (1=0) (2=1) (3=0) (4=0) (5=0) (6=2), generate(benefit8)
* NEW FEATURE SELECTION FOR QUESTIONED VARS
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.isquaternary
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.challenge2
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.challenge3
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.challenge4
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.challenge4 i.ishhcomplex
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.challenge4 i.rmtcommon2
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.challenge4 i.benefit2
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.challenge4 i.benefit3
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.challenge4 i.benefit4
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.challenge4 i.benefit5
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.challenge4 i.benefit6
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.challenge4 i.benefit7
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.challenge4 i.benefit8
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.challenge4 i.benefit2
drop rmtcommon rmtcommon2 rmt household challenge challenge2 challenge3 ishhcomplex rmtcommon2 benefit benefit3 benefit4 benefit5 benefit6 benefit7 benefit8 id
rename benefit2 benefit
rename challenge4 islackmotivation
label values islackmotivation ynlablel
label define bnftlabel 0 "other" 1 "less expences" 2 "work-life balance"
label values benefit bnftlabel
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.islackmotivation i.benefit
label define bnftlabel 0 "other" 1 "less expences" 2 "professional" 3 "work-life balance"
label define bnftlabel2 0 "other" 1 "less expences" 2 "professional" 3 "work-life balance"
label values benefit bnftlabel2
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.islackmotivation i.benefit
recode benefit (0=0) (1=1) (2=2) (3=2), generate(benefit2)
ologit y workinghoursdiff i.gendr i.rmtrecommend i.rmtpref i.islackmotivation i.benefit2
codebook benefit
codebook benefit2
label define bnftlabel3 0 "other" 1 "less expences" 2 "professional or work-life balance"

```

```

label values benefit2 bnftlabel3
drop benefit
rename benefit2 benefit
tab y isquaternary, chi2 V
ologit y workinghoursdiff i.rmtrecommend i.rmtpref i.islackmotivation i.benefit i.isquaternary
drop genr
codebook
cls
* TESTING
ologit y workinghoursdiff i.rmtrecommend i.rmtpref i.islackmotivation i.benefit i.isquaternary
brant, detail
ologit y workinghoursdiff i.rmtrecommend i.rmtpref i.islackmotivation i.benefit i.isquaternary
est store logit_ordered
gologit2 y workinghoursdiff i.rmtrecommend i.rmtpref i.islackmotivation i.benefit i.isquaternary, npl lrf store(logit_generalized)
gologit2 y workinghoursdiff i.rmtrecommend i.rmtpref i.islackmotivation i.benefit i.isquaternary, auto lrf store(logit_partial)
lrtest logit_ordered logit_generalized, force
lrtest logit_ordered logit_partial, force
lrtest logit_generalized logit_partial, force
* MEASURING PERFORMANCE
label define ylbl 1 "dropped" 2 "same level" 3 "increased"
label values y ylbl
gologit2 y workinghoursdiff i.rmtrecommend i.rmtpref i.islackmotivation i.benefit i.isquaternary, auto lrf store(logit_partial)
gologit2 y workinghoursdiff i.rmtrecommend i.rmtpref i.islackmotivation i.benefit i.isquaternary, auto lrf
predict p1 p2 p3, pr
gen y_pred1 = (p1 > 0.5 & p2 <= 0.5 & p3 <= 0.5)
gen y_pred2 = (p2 > 0.5 & p3 <= 0.5)
gen y_pred3 = (p3 > 0.5)
gen y_actual1 = (y == 1)
gen y_actual2 = (y == 2)
gen y_actual3 = (y == 3)
tab y_actual1 y_pred1
tab y_actual1 y_pred2
tab y_actual1 y_pred3
tab y_actual2 y_pred1
tab y_actual2 y_pred2
tab y_actual2 y_pred3
tab y_actual3 y_pred1
tab y_actual3 y_pred2
tab y_actual3 y_pred3
codebook y
drop p1 p2 p3
gologit2 y workinghoursdiff i.rmtrecommend i.rmtpref i.islackmotivation i.benefit i.isquaternary, auto lrf
predict p1 p2 p3, pr
codebook p1
save "C:\Projects\RemoteWorkSurvey\ytyt.dta"
generate pred = .
replace pred = 1 if newvar == 1
replace pred = 2 if newvar2 == 1 & newvar != 1
replace pred = 3 if pred == .
tab pred y

```