



Customer Churn at Telco

Jack Bergstrom, Frances Genta Romero,
Steven Howley



Introduction

- Telco telecommunications dataset
- Churn rates for customers
- IBM ran analysis for Telco, we are going to run our own :)

Our Question:

How does contract type affect churn rate of customers differently for Telco?
Within these different contract types, how do families fair?



Background info

- Telco is a telecommunications agency
- Our dataset explores “fictional” Telco customers in California (anonymized)
- IBM was hired to run analysis on Telco customers and provide insight and support to help Telco compete in the market
- No date for the dataset but IBM ran their analysis in 2019



The data

- 7043 observations of 33 variables.
- Data was collected from Telco's customer accounts
- 73.5% of data is right-censored after 80 months of collection
- Churn was observed when a customer cancelled their account with Telco



Side Note

This dataset was huge so we explored a lot of options before figuring out what we wanted to do:

VARIABLES:	-Gender
-Location data	-Dependents
-Categorical internet plans	-Senior Citizen
-Monthly spend	-etc..

Covariates included different security options offered as well as location data options

We chose to explore the difference between contracts and how dependents affected the contract subgroups

Narrowing Focus

The largest z-scores
were for contract type
and dependents (all
significant)

Operating from these
features will yield the
best results

Call:

```
survreg(formula = theSurvObject ~ ds$`Senior Citizen` + ds$Dependents +  
        ds$Contract + ds$`Payment Method` + ds$Partner + ds$`Paperless Billing`)
```

	Value	Std. Error	z	p
(Intercept)	4.1197	0.0937	43.98	< 2e-16
ds\$`Senior Citizen`Yes	0.1116	0.0642	1.74	0.08190
ds\$DependentsYes	1.4384	0.1236	11.64	< 2e-16
ds\$ContractOne year	2.1806	0.1014	21.51	< 2e-16
ds\$ContractTwo year	3.8056	0.1830	20.80	< 2e-16
ds\$`Payment Method`Credit card (automatic)	0.0674	0.1059	0.64	0.52416
ds\$`Payment Method`Electronic check	-0.7600	0.0829	-9.17	< 2e-16
ds\$`Payment Method`Mailed check	-0.7041	0.1001	-7.03	2.0e-12
ds\$PartnerYes	0.4433	0.0586	7.56	3.9e-14
ds\$`Paperless Billing`Yes	-0.2178	0.0648	-3.36	0.00077
Log(scale)	0.1564	0.0191	8.20	2.5e-16

Scale= 1.17

Weibull distribution

Loglik(model)= -9077.4 Loglik(intercept only)= -10576.3

Chisq= 2997.85 on 9 degrees of freedom, p= 0

Number of Newton-Raphson Iterations: 8

n= 7043

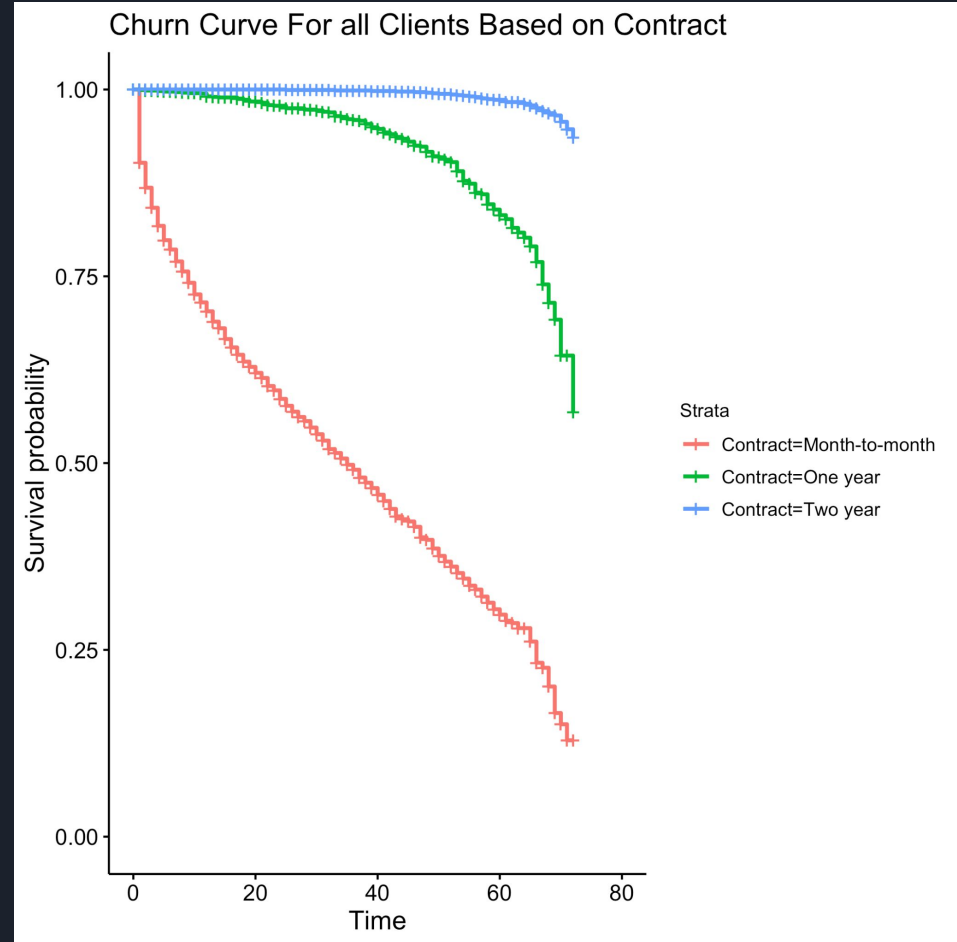
Churn Curves for Each Contract Plan

Time: months

Month to month contracts are churning much more rapidly than the yearly contracts

One year contracts are more stable but decline over time and take a large hit at around 5 years

Two year contracts are consistent and have barely any churning



Churn Curve for Month to Month Contracts with Dependents

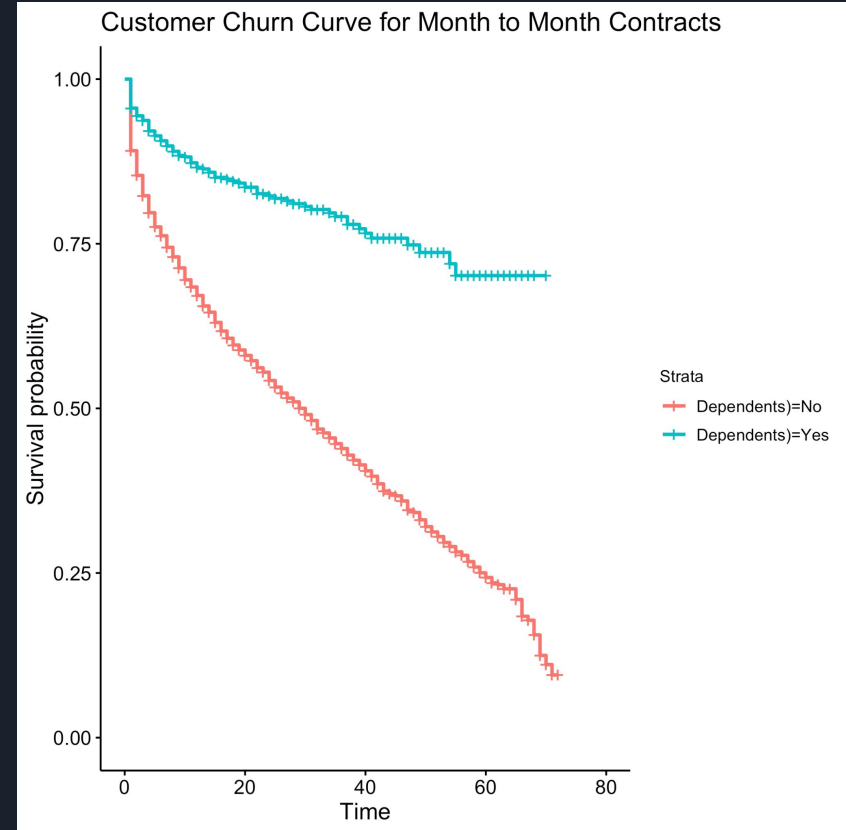
Within the month to month contracts, customers without dependents were much more likely to churn than those with dependents

`plnorm(24,meanlog = 3.146728,sdlog = 1.839436)`

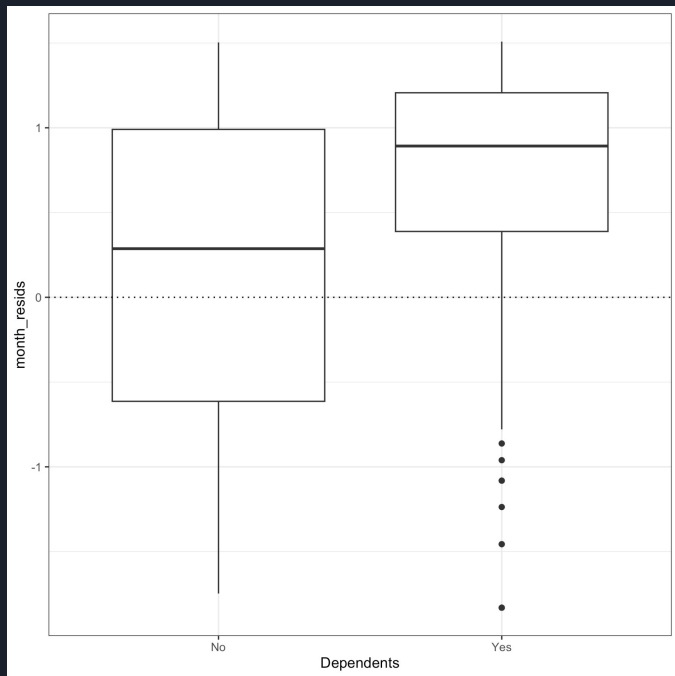
51% churn in 2 years w/out dependents

`plnorm(24,meanlog = 5.570644 ,sdlog = 2.720568)`

19% churn in 2 years w/ dependents



Log normal fit the month to month curve best but the residuals are a little wonky



w/ dependents mean = 5.57 sd = 2.72

w/out dependents mean = 3.15 sd = 1.84

only 636 people with dependents on month to month plan vs 3239 without - may be why we see weird residuals

Call:

```
survreg(formula = object ~ strata(ds_month$Dependents), dist = "lognorm")
```

	Value	Std. Error	z	p
(Intercept)	3.3866	0.0431	78.6	<2e-16
No	0.6618	0.0198	33.5	<2e-16
Yes	0.6150	0.0554	11.1	<2e-16

Scale:

No	Yes
1.94	1.85

Log Normal distribution

Loglik(model)= -7700.5 Loglik(intercept only)= -7700.5

Number of Newton-Raphson Iterations: 5

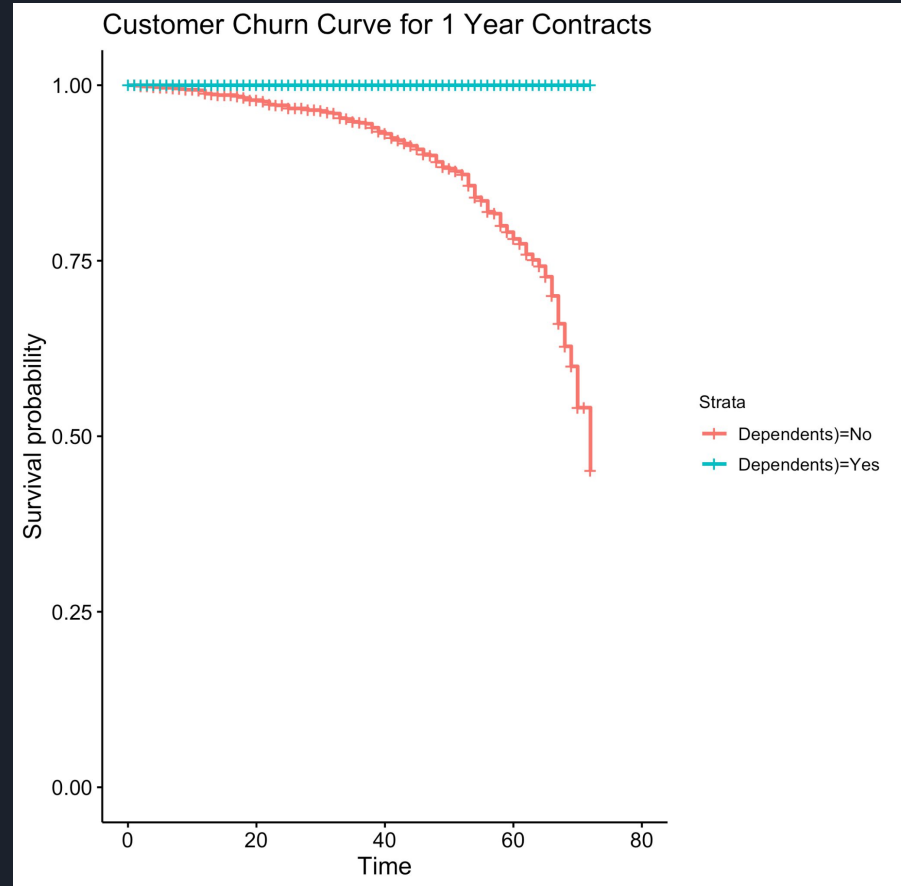
n= 3875

Churn Curves for Yearly Contracts With Dependents

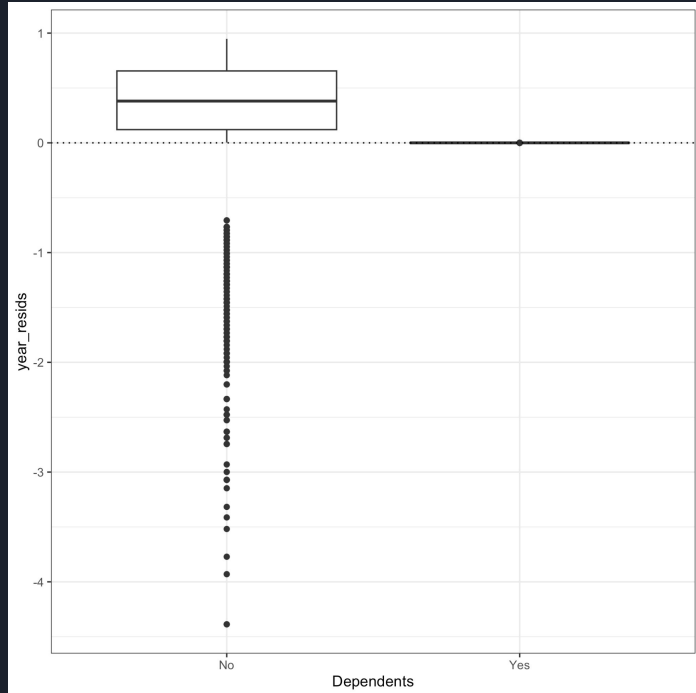
Within 1 year contracts, customers with dependants are consistent whereas customers without dependants are leaving rapidly

```
pweibull(24, shape = 2.741, scale = exp(4.568644))
```

2% churn in 2 years w/out dependents



Weibull fit the 1 year curve best and the residuals look good



Beta w/out dependents = 2.72 (wear out mode)
Alpha w/out dependents = 35.17

Call:

```
survreg(formula = object2 ~ strata(ds_1_year$Dependents), dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	4.5686	0.0450	101.46	<2e-16
No	-1.0084	0.0646	-15.62	<2e-16
Yes	-4.1399	155.7590	-0.03	0.98

Scale:

No	Yes
0.3648	0.0159

Weibull distribution

Loglik(model)= -1019.2 Loglik(intercept only)= -1019.2

Number of Newton-Raphson Iterations: 18

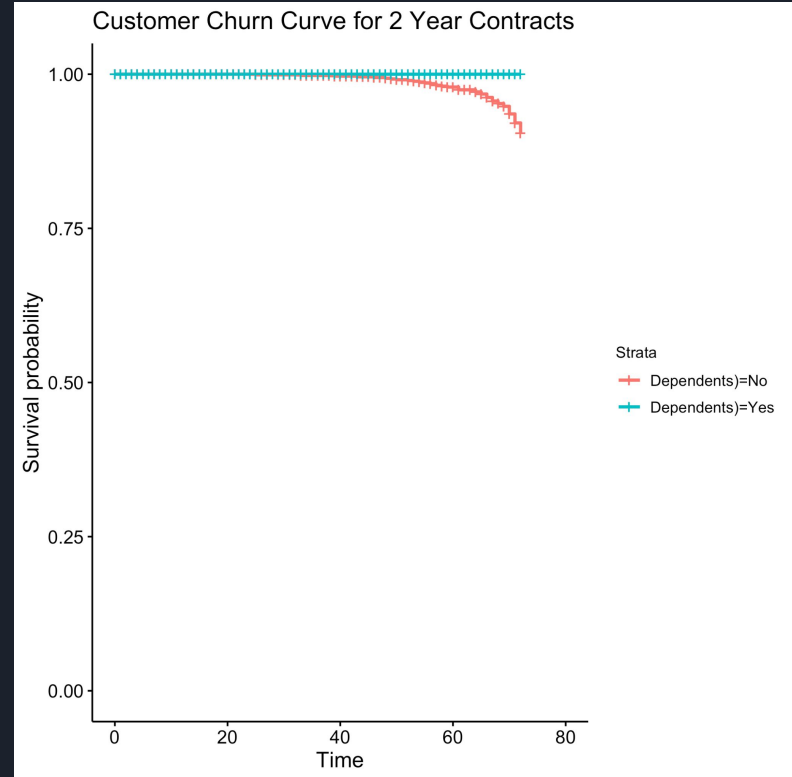
n= 1473

Churn Curve for Two Year Contract with Dependants

No churn observations in Dependents “yes” with 633 observations

This is unfortunate for modeling purposes but reinforces that a two year contract with dependents is very resistant to churn

We couldn't make models for 2 year data because it ran out of iterations and did not converge





Summary of Results

Original question: How does contract type affect churn rate of customers differently for Telco? Within these different contract types, how do families fair?

- Churn rate varied by contract type with month to month contracts having the highest churn rate and 2 year contracts the lowest
- Churn rate varied across all contract types by whether a customer had dependents, having dependents reduced churn rate
- The month to month contract showed infant mortality and very few of these customers had dependents
- The 1 year contract showed wearout mode and most customers with dependents had a 1 or 2 year contract



Actionable Insight

As data consultants from IBM we recommend offering competitive deals on yearly contracts to preserve existing customers and convert M2M customers into yearly ones.

Additionally, customers with dependents are much less likely to churn. Try running ads during back to school sales or changing the content of our ads to target families.

It may be worth it to buy out family contracts elsewhere as these people are locked in and will not switch over unless incentivized.

Future steps with this data:

Could explore if churn rate is higher in certain regions to target these areas with more offers

Questions?





Additional Slides



Month to Month AIC

```
parametric survival regression for month to month
```{r}
object <- Surv(ds_month$`Tenure Months`, ds_month$`Churn Value`)
mod_w <- survreg(object~strata(ds_month$Dependents), dist = "weibull")
mod_e <- survreg(object~(ds_month$Dependents), dist = "exponential")
mod_l <- survreg(object~strata(ds_month$Dependents), dist = "lognorm")

AIC(mod_w) #15470.49
AIC(mod_e) #15490.52
AIC(mod_l) #15406.98
#log normal fits best
summary(mod_l)
```

Fitting the  
correct model  
to the month  
to month ds  
and checking  
the AIC

# Yearly AIC

```
object2 <- Surv(ds_1_year$`Tenure Months`, ds_1_year$`Churn Value`)
mod_w_1 <- survreg(object2~strata(ds_1_year$Dependents), dist = "weibull")
mod_e_1 <- survreg(object2~(ds_1_year$Dependents), dist = "exponential")
mod_l_1 <- survreg(object2~strata(ds_1_year$Dependents), dist = "lognorm")

AIC(mod_w_1) #2044.389
AIC(mod_e_1) #2208.055
AIC(mod_l_1) #2107.419
#weibull fits best
summary(mod_w_1)
```

Fitting the correct model to the 1 year ds and checking the AIC

# Comparison

