

# 车险电销坐席通话与成交率的相关性研究

## 研究背景与问题



### 车险销售



#### 电销成为车险的重要销售渠道

江苏全省车险电销比例达到20%，个别地市已经超过30%



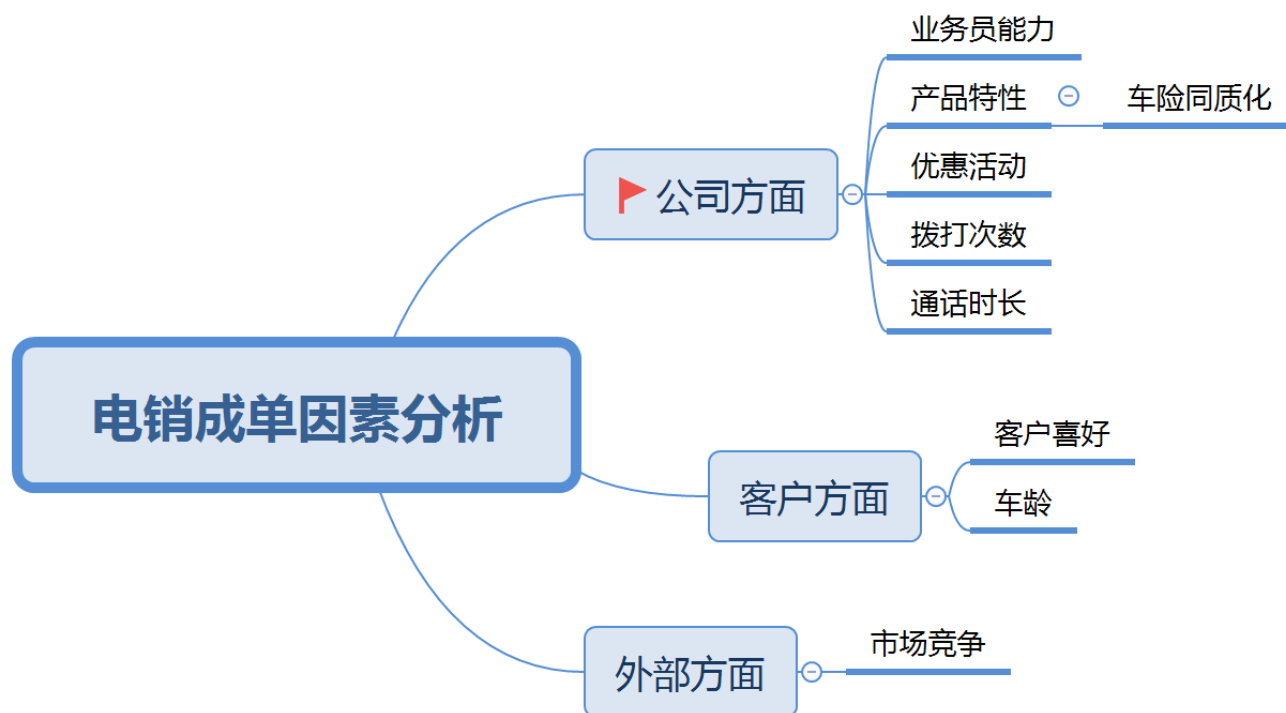
#### 如何提高电销坐席的成单业绩成为重要课题

所谓电销就是业务员通过电话与客户沟通，了解客户的需要，向客户推荐合适的产品实现车险的销售。

坐席重要的过程数据就是与客户的通话次数、通话时长等内容；

**车险电销坐席通话与保单成交有什么样的关系？  
是否可以通过控制通话次数、时长等提高销售业绩？**

## 课题分析



## 数据筛选与控制

### ✓ 关键数据筛选

- ◆ 筛选原则：可量化、可控制、有相关性
- ◆ 因变量：是否成单 --- 0, 1
- ◆ 自变量：业务员能力 --- 工龄  
拨打次数 --- 计数  
通话时长 --- 平均时长

### ✓ 干扰因素控制

- ◆ 数据分析要求具有普遍性：要求数据具有随机性和同质性；
- ◆ 客户数据选择：1. 一个地市的全部坐席续保拨打客户、全量统计；  
2. 客户选择相同的终保日期
- ◆ 产品与营销方案的一致性：1. 选择单一地市公司数据；  
2. 选择相同拨打区间的数据，同一时刻同一地区的营销方案是相同的。

## 原始数据

- 电销坐席人员入职时间清单
- 电销坐席人员与客户的联系记录清单
- 客户的成交清单

usercode	operatetimeforhis
A320601758	2019-04-15 15:00:23
83234574	2019-04-13 10:08:25
A320601010	2019-04-12 15:01:37

tasktargetid	tel_starttime	tel_endtime	intervaltime
PDAA201832060000045577	2019-02-20 14:06:16	2019-02-20 14:06:16	0
PDAA2018320600000105659	2019-03-04 17:37:04	2019-03-04 17:37:04	0
PDAA2018320600000078214	2019-02-15 15:44:59	2019-02-15 15:56:56	717
PDAA2018320600000099691	2019-02-26 14:57:47	2019-02-26 14:58:10	23



## 数据预处理

- 电销坐席能力处理：  
工作天数=样本截止日期-入职日期
- 联系记录清单：  
删除通话时长小于10秒的无效通话  
根据单一客户联系记录生成通话次数  
求取单一客户平均通话时长
- 客户的成单情况清单  
客户的成单转换：投保=1, 未投保=0

上年保单	坐席工号	工作天数	通话次数	平均时长	成功与否
PDAA201832060000134595	A320601639	143	2	215.5	0
PDAA201832060000048135	A320601637	143	5	356.6	0
PDAA201832060000118069	A320601380	391	3	64	0
PDAA201832060000043309	A320600628	1171	2	44.5	0
PDAA201832060000047116	A320601717	110	7	116.8571	1
PDAA201832060000095160	15187641	710	4	267.25	1

## Logistic模型

- 由于研究的因变量只有1和0，是离散的二值型变量，无法采用常规的线性模型。采用适用二值的Logistic模型进行建模分析研究。
- 自变量X为连续变量（天数、次数、分钟数）
- 线性模型的拟合形式如下：

Logistic回归适用于二值响应变量（0,1）。模型假设Y服从二项分布，线性模型的拟合形式为：

$$\ln\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \sum_{j=1}^p \beta_j X_j$$

其中  $\pi = \mu_Y$  是Y的条件均值（即给定一系列X的值时Y=1的概率）， $(\pi/1-\pi)$  为Y=1时的优势比， $\log(\pi/1-\pi)$  为对数优势比，或logit。本例中， $\log(\pi/1-\pi)$  为连接函数，概率分布为二项分布，可用如下代码拟合Logistic回归模型：

```
glm(Y~X1+X2+X3, family=binomial(link="logit"), data=mydata)
```

## 数据分析

### ➤ 采用R语言进行数据建模统计

#### ◆ glm函数介绍：

`glm(formula, family=family.generator, data, control = list(...))`

#### ◆ 常用的family：

- `binomial(link='logit')` ----响应变量服从二项分布，连接函数为logit，即logistic回归
- `binomial(link='probit')` ----响应变量服从二项分布，连接函数为probit
- `poisson(link='identity')` ----响应变量服从泊松分布，即泊松回归

#### ◆ control:控制算法误差和最大迭代次数

---



## 构建数据框

- 因变量Y : yon
- 自变量X1 : workdur
- 自变量X2 : calltimes
- 自变量X3 : avrdur

RStudio

File Edit Code View Plots Session Build Debug

+ + R + Go to file/function

dddd x

Filter

	workdur	calltimes	avrdur	yon
1	143	2	215.50000	0
2	143	5	356.60000	0
3	391	3	64.00000	0
4	1171	2	44.50000	0
5	110	7	116.85714	1
6	710	4	267.25000	1
7	539	1	239.00000	0
8	24	5	99.40000	0
9	520	2	106.50000	0
10	109	2	203.50000	0
11	996	1	264.00000	0

Showing 1 to 12 of 17,987 entries, 4 total columns

## R 语句与结果

dddd x				
	workdur	calltimes	avrdur	yon
1	143	2	215.50000	0
2	143	5	356.60000	0
3	391	3	64.00000	0

```
Console Terminal x Jobs x
d:/Program Files/RStudio/
> fit.full <- glm(yon~workdur+calltimes+avrdur, family = binomial(link="logit"), data =
  dddd)
> summary(fit.full)

call:
glm(formula = yon ~ workdur + calltimes + avrdur, family = binomial(link = "logit"),
  data = dddd)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.9441  -0.8620  -0.6679   1.2156   2.1302

Coefficients:
              Estimate Std. Error
(Intercept) -2.402e+00  4.917e-02
workdur      8.192e-04  4.722e-05
calltimes    1.907e-01  6.930e-03
avrdur       2.413e-03  8.868e-05
z value Pr(>|z|)
(Intercept) -48.85 <2e-16 ***
workdur      17.35 <2e-16 ***
calltimes    27.52 <2e-16 ***
avrdur       27.21 <2e-16 ***
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 22737  on 17986  degrees of freedom
Residual deviance: 20994  on 17983  degrees of freedom
AIC: 21002

Number of Fisher Scoring iterations: 4
```

## 结果初步多元分析

➤ 经过R语言glm建模获取模型如下：

$$Y = -2.4 + 0.00081 \times X_1 + 0.19 \times X_2 + 0.0024 \times X_3$$

$$Y = \text{Log} \left( \frac{P_i}{1 - P_i} \right)$$

```
Coefficients:
      Estimate Std. Error
(Intercept) -2.402e+00  4.917e-02
workdur      8.192e-04  4.722e-05
calltimes    1.907e-01  6.930e-03
avrdur       2.413e-03  8.868e-05
              z value Pr(>|z|)
(Intercept)  -48.85   <2e-16 ***
workdur       17.35   <2e-16 ***
calltimes     27.52   <2e-16 ***
avrdur        27.21   <2e-16 ***
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$P_i$ : i客户的成单概率,  $X_1$ :工作时间,  $X_2$ : 通话次数,  $X_3$ :通话时长

从以上分析结果来看,  $X_2$ 的属于大于 $X_3$ 两个数量级, 大于 $X_1$ 三个数量级, 在简化分析时, 可以省略处理, 所以下一步只对通话次数做一元建模分析。

## 二次一元分析

➤ 经过R语言glm建模获取模型如下：

$$P = \frac{1}{1 + e^{-(-1.31 + 0.16X)}}$$

P:成单概率, X: 通话次数

X 取值越大, P值越大  
也就是通话次数越多, 成  
单概率越高

```
> fit.single <- glm(yon~calltimes, family = binomial(link="logit"), data = dddd)
> summary(fit.single)
```

```
Call:
glm(formula = yon ~ calltimes, family = binomial(link = "logit"),
    data = dddd)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.7873	-0.9040	-0.7924	1.3381	1.6902

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-1.31114	0.02995	-43.78	<2e-16 ***
calltimes	0.15682	0.00657	23.87	<2e-16 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

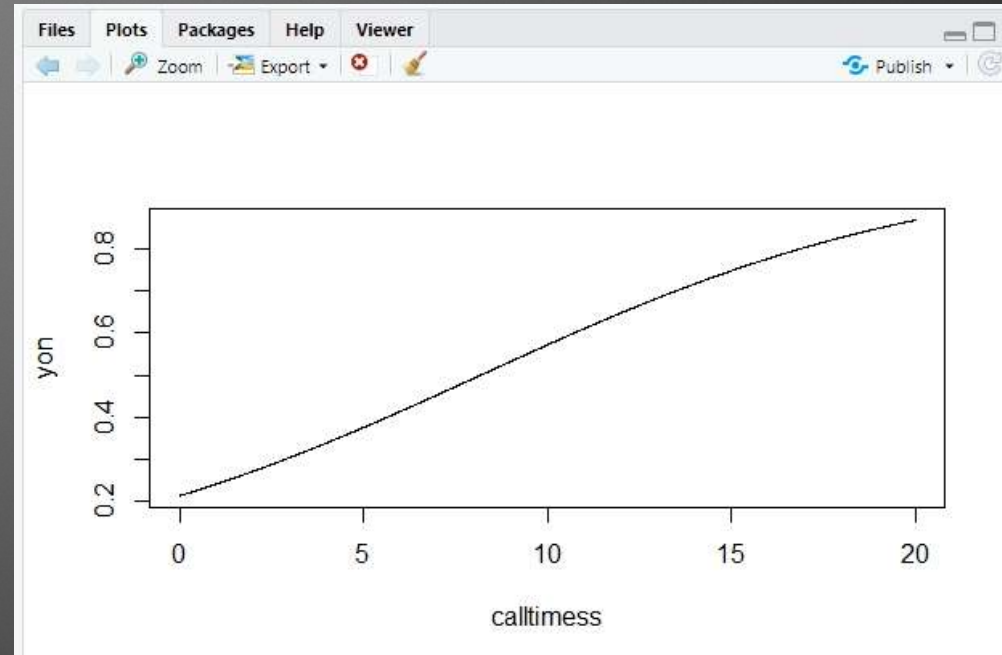
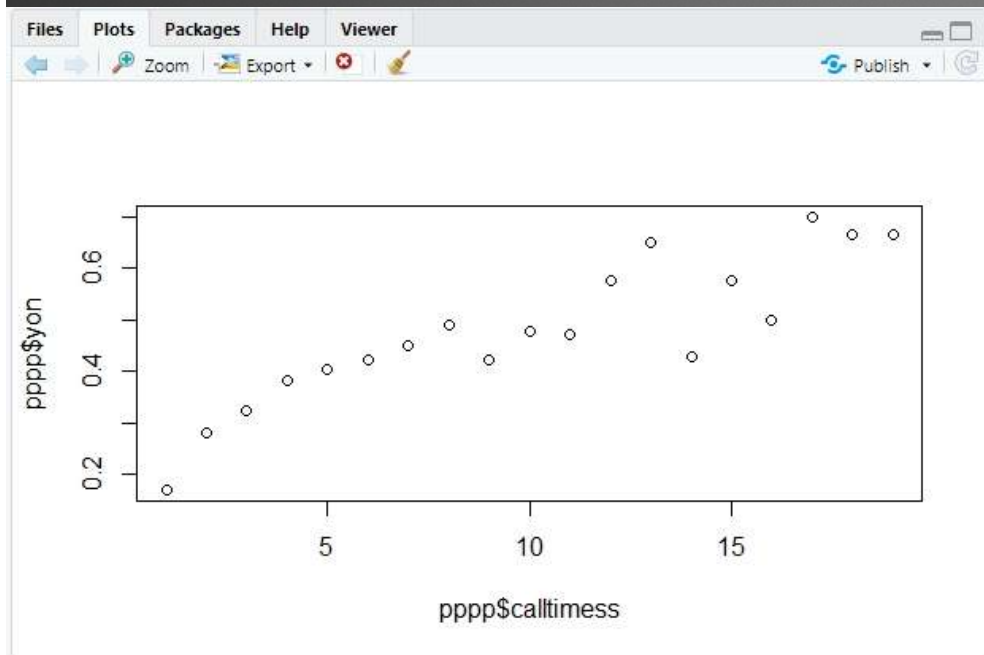
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 22737 on 17986 degrees of freedom  
Residual deviance: 22134 on 17985 degrees of freedom  
AIC: 22138

Number of Fisher scoring iterations: 4



## 二次一元分析





## 二次一元分析

➤ 经过R语言glm建模获取模型如下：

$$P = \frac{1}{1 + e^{-(-1.31 + 0.16X)}}$$

P:成单概率, X: 通话次数

**X 取值越大, P值越大  
也就是通话次数越多, 成  
单概率越高**

---

## 进一步对比验证

### ◆ 设计随机控制实验

通过在同一个地市的坐席中进行拨打次数的控制组与对照组试验进一步验证。

控制变量：与客户的拨打次数

其他变量：客户随机分配、坐席随机分配

#### ➤ 控制组

控制与客户电话沟通的最低次数



#### ➤ 对照组

保持原来的状态

### ◆ 结果分析

对两组数据的成单率进行分析，判断是否具有明显差异性

## 计划与预算

1. 历史数据分析  
各地市分别统计分析 ~ ~ 2019.05.11
2. 随机控制实验的详细设计讨论 ~ ~ 2019.05.25
3. 控制组与对照组的筛选 ~ ~ 2019.06.08
4. 随机控制试验执行 ~ ~ 2019.07.06
5. 试验结果分析 ~ ~ 2019.07.20

预算：2万

由于分析与处理对象全是内部员工，无需额外开始，但为了更好的开展随机控制试验，申请2万预算用于控制或对照组的奖励

---



**Thank You**