Applied Probability Models in Marketing Research: Extensions

Peter S. Fader
University of Pennsylvania
faderp@wharton.upenn.edu

Bruce G. S. Hardie
London Business School
bhardie@london.edu

14th Annual Advanced Research Techniques Forum
June 1–4, 2003

Probability Models 101

The Logic of Probability Models

- Many researchers attempt to describe/predict behavior using observed variables.
- However, they still use random components in recognition that not all factors are included in the model.
- We treat behavior as if it were "random" (probabilistic, stochastic).
- We propose a model of individual-level behavior which is "summed" across individuals (taking individual differences into account) to obtain a model of aggregate behavior.

Uses of Probability Models

- Understanding market-level behavior patterns
- Prediction
  - To settings (e.g., time periods) beyond the observation period
  - Conditional on past behavior
- Profiling behavioral propensities of individuals
- Benchmarks/norms

Building a Probability Model

(i) Determine the marketing decision problem/ information needed.

(ii) Identify the *observable* individual-level behavior of interest.

- We denote this by x.

(iii) Select a probability distribution that characterizes this individual-level behavior.

- This is denoted by $f(x|\theta)$.
- We view the parameters of this distribution as individual-level *latent traits*.

Building a Probability Model

(iv) Specify a distribution to characterize the distribution of the latent trait variable(s) across the population.

- We denote this by $g(\theta)$.
- This is often called the *mixing distribution*.

(v) Derive the corresponding *aggregate* or *observed* distribution for the behavior of interest:

$$f(x) = \int f(x|\theta)g(\theta)\,d\theta$$

Building a Probability Model

(vi) Estimate the parameters (of the mixing distribution) by fitting the aggregate distribution to the observed data.

(vii) Use the model to solve the marketing decision problem/provide the required information.

"Classes" of Models

- The first tutorial introduced simple models for three behavioral processes:
  - Timing ⟶ "when"
  - Counting ⟶ "how many"
  - "Choice" ⟶ "whether/which"
- Each of these simple models has multiple applications.
- More complex behavioral phenomena can be captured by combining models from each of these processes.

Outline

- Problem 4: Characterizing the Purchasing of Hard-Candy
  (Introduction to Finite Mixture Models)
- Problem 5: Who is Visiting khakichinos.com?
  (Incorporating Covariates in Count Models)
- Problem 6: Predicting New Product Trial (Again)
  (Extending Basic Models for Timing Data)

Problem 4: Characterizing the Purchasing of Hard-Candy

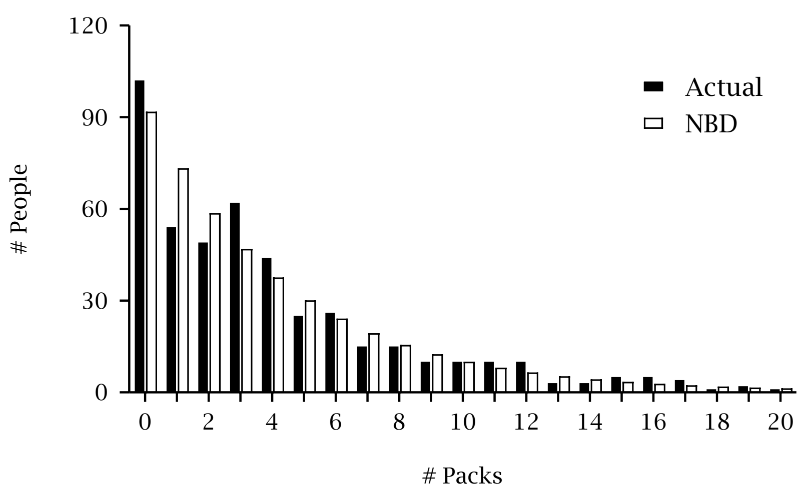
(Introduction to Finite Mixture Models)

Distribution of Hard-Candy Purchases

# Packs	# People	# Packs	# People
0	102	11	10
1	54	12	10
2	49	13	3
3	62	14	3
4	44	15	5
5	25	16	5
6	26	17	4
7	15	18	1
8	15	19	2
9	10	20	1
10	10		

Source: Dillon and Kumar (1994)

Fit of NBD



$\hat{r} = 0.998, \hat{\alpha} = 0.250$

Problem 4 -- NBD

	A	B	C	D	E	F	G	H	I	J
1	r	0.997657								
2	alpha	0.249962								
3	LL	=SUM(D6:D26)								
4										
5	# Packs	Observed	P(X=x)	LL	Expected		# Packs	Observed	Expected	(O-E)^2/E
6	0	102	=(B2/(B2+1))^B1	=B6*LN(C6)	=B$27*C6		=A6	=B6	=E6	=(H6-I6)^2/I6
7	1	54	=(B$1+A7-1)/(A7*(B$2+1))*C6	=B7*LN(C7)	=B$27*C7		=A7	=B7	=E7	=(H7-I7)^2/I7
8	2	49	=(B$1+A8-1)/(A8*(B$2+1))*C7	=B8*LN(C8)	=B$27*C8		=A8	=B8	=E8	=(H8-I8)^2/I8
9	3	62	=(B$1+A9-1)/(A9*(B$2+1))*C8	=B9*LN(C9)	=B$27*C9		=A9	=B9	=E9	=(H9-I9)^2/I9
10	4	44	=(B$1+A10-1)/(A10*(B$2+1))*C9	=B10*LN(C10)	=B$27*C10		=A10	=B10	=E10	=(H10-I10)^2/I10
11	5	25	=(B$1+A11-1)/(A11*(B$2+1))*C10	=B11*LN(C11)	=B$27*C11		=A11	=B11	=E11	=(H11-I11)^2/I11
12	6	26	=(B$1+A12-1)/(A12*(B$2+1))*C11	=B12*LN(C12)	=B$27*C12		=A12	=B12	=E12	=(H12-I12)^2/I12
13	7	15	=(B$1+A13-1)/(A13*(B$2+1))*C12	=B13*LN(C13)	=B$27*C13		=A13	=B13	=E13	=(H13-I13)^2/I13
14	8	15	=(B$1+A14-1)/(A14*(B$2+1))*C13	=B14*LN(C14)	=B$27*C14		=A14	=B14	=E14	=(H14-I14)^2/I14
15	9	10	=(B$1+A15-1)/(A15*(B$2+1))*C14	=B15*LN(C15)	=B$27*C15		=A15	=B15	=E15	=(H15-I15)^2/I15
16	10	10	=(B$1+A16-1)/(A16*(B$2+1))*C15	=B16*LN(C16)	=B$27*C16		=A16	=B16	=E16	=(H16-I16)^2/I16
17	11	10	=(B$1+A17-1)/(A17*(B$2+1))*C16	=B17*LN(C17)	=B$27*C17		=A17	=B17	=E17	=(H17-I17)^2/I17
18	12	10	=(B$1+A18-1)/(A18*(B$2+1))*C17	=B18*LN(C18)	=B$27*C18		=A18	=B18	=E18	=(H18-I18)^2/I18
19	13	3	=(B$1+A19-1)/(A19*(B$2+1))*C18	=B19*LN(C19)	=B$27*C19		=A19	=B19	=E19	=(H19-I19)^2/I19
20	14	3	=(B$1+A20-1)/(A20*(B$2+1))*C19	=B20*LN(C20)	=B$27*C20		=A20	=B20	=E20	=(H20-I20)^2/I20
21	15	5	=(B$1+A21-1)/(A21*(B$2+1))*C20	=B21*LN(C21)	=B$27*C21		15+	=SUM(B21:B26)	=SUM(E21:E26)	=(H21-I21)^2/I21
22	16	5	=(B$1+A22-1)/(A22*(B$2+1))*C21	=B22*LN(C22)	=B$27*C22					=SUM(J6:J21)
23	17	4	=(B$1+A23-1)/(A23*(B$2+1))*C22	=B23*LN(C23)	=B$27*C23					
24	18	1	=(B$1+A24-1)/(A24*(B$2+1))*C23	=B24*LN(C24)	=B$27*C24				# params	2
25	19	2	=(B$1+A25-1)/(A25*(B$2+1))*C24	=B25*LN(C25)	=B$27*C25				df	=16-J24-1
26	20	1	=(B$1+A26-1)/(A26*(B$2+1))*C25	=B26*LN(C26)	=B$27*C26					
27		=SUM(B6:B26)							p-value	=CHIDIST(J22,J25)

Problem 4 -- NBD

	A	B	C	D	E	F	G	H	I	J
1	r	0.998								
2	alpha	0.250								
3	LL	-1140.02								
4										
5	# Packs	Observed	P(X=x)	LL	Expected		# Packs	Observed	Expected	(O-E)^2/E
6	0	102	0.20073	-163.79	91.5		0	102	91.5	1.20
7	1	54	0.16021	-98.89	73.1		1	54	73.1	4.97
8	2	49	0.12802	-100.72	58.4		2	49	58.4	1.51
9	3	62	0.10234	-141.32	46.7		3	62	46.7	5.04
10	4	44	0.08183	-110.14	37.3		4	44	37.3	1.20
11	5	25	0.06543	-68.17	29.8		5	25	29.8	0.78
12	6	26	0.05233	-76.71	23.9		6	26	23.9	0.19
13	7	15	0.04185	-47.60	19.1		7	15	19.1	0.87
14	8	15	0.03347	-50.96	15.3		8	15	15.3	0.00
15	9	10	0.02677	-36.20	12.2		9	10	12.2	0.40
16	10	10	0.02141	-38.44	9.8		10	10	9.8	0.01
17	11	10	0.01713	-40.67	7.8		11	10	7.8	0.61
18	12	10	0.01370	-42.90	6.2		12	10	6.2	2.25
19	13	3	0.01096	-13.54	5.0		13	3	5.0	0.80
20	14	3	0.00876	-14.21	4.0		14	3	4.0	0.25
21	15	5	0.00701	-24.80	3.2		15+	18	11.8	3.27
22	16	5	0.00561	-25.92	2.6					23.35
23	17	4	0.00449	-21.63	2.0					
24	18	1	0.00359	-5.63	1.6				# params	2
25	19	2	0.00287	-11.71	1.3				df	13
26	20	1	0.00230	-6.08	1.0					
27		456							p-value	0.038

The Zero-Inflated NBD Model

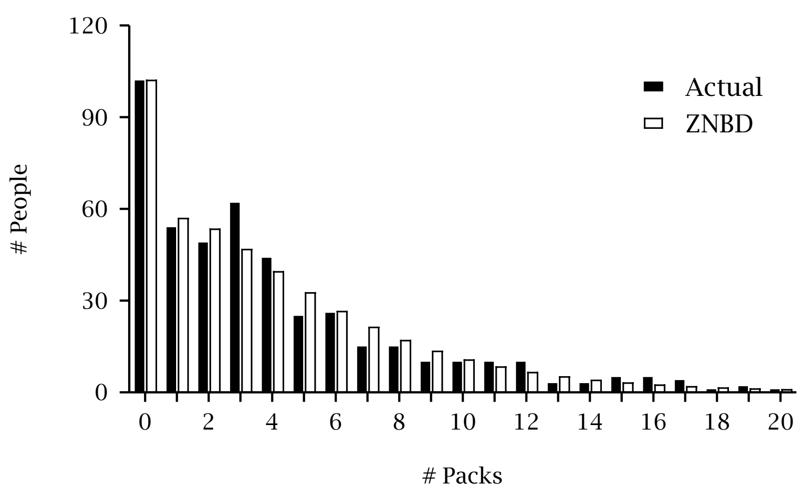
Because of the "excessive" number of zeros, let us consider the zero-inflated NBD (ZNBD) model:

- a proportion π of the population never buy hard-candy
- the visiting behavior of the "ever buyers" can be characterized by the NBD model

$$P(X = x) = \delta_{x=0}\pi + (1 - \pi) \times \frac{\Gamma(r + x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha + 1}\right)^r \left(\frac{1}{\alpha + 1}\right)^x$$

This is sometimes called the NBD with hard-core non-buyers model.

Fit of ZNBD



$$\hat{\pi} = 0.113, \hat{r} = 1.504, \hat{\alpha} = 0.334$$

Problem 4 -- ZNBD

	A	B	C	D	E	F	G	H	I	J	K
1	r	0.997657									
2	alpha	0.249962									
3	pi	0.113103									
4	LL	=SUM(E8:E28)									
5											
6			P(X=x)								
7	# Packs	Observed	NBD	ZNBD	LL	Expected		# Packs	Observed	Expected	(O-E)^2/E
8	0	102	=(B2/(B2+1))^B1	=(A8=0)*B$3+(1-B$3)*C8	=B8*LN(D8)	=B$29*D8		=A8	=B8	=F8	=(I8-J8)^2/J8
9	1	54	=(B$1+A9-1)/(A9*(B$2+1))*C8	=(A9=0)*B$3+(1-B$3)*C9	=B9*LN(D9)	=B$29*D9		=A9	=B9	=F9	=(I9-J9)^2/J9
10	2	49	=(B$1+A10-1)/(A10*(B$2+1))*C9	=(A10=0)*B$3+(1-B$3)*C10	=B10*LN(D10)	=B$29*D10		=A10	=B10	=F10	=(I10-J10)^2/J10
11	3	62	=(B$1+A11-1)/(A11*(B$2+1))*C10	=(A11=0)*B$3+(1-B$3)*C11	=B11*LN(D11)	=B$29*D11		=A11	=B11	=F11	=(I11-J11)^2/J11
12	4	44	=(B$1+A12-1)/(A12*(B$2+1))*C11	=(A12=0)*B$3+(1-B$3)*C12	=B12*LN(D12)	=B$29*D12		=A12	=B12	=F12	=(I12-J12)^2/J12
13	5	25	=(B$1+A13-1)/(A13*(B$2+1))*C12	=(A13=0)*B$3+(1-B$3)*C13	=B13*LN(D13)	=B$29*D13		=A13	=B13	=F13	=(I13-J13)^2/J13
14	6	26	=(B$1+A14-1)/(A14*(B$2+1))*C13	=(A14=0)*B$3+(1-B$3)*C14	=B14*LN(D14)	=B$29*D14		=A14	=B14	=F14	=(I14-J14)^2/J14
15	7	15	=(B$1+A15-1)/(A15*(B$2+1))*C14	=(A15=0)*B$3+(1-B$3)*C15	=B15*LN(D15)	=B$29*D15		=A15	=B15	=F15	=(I15-J15)^2/J15
16	8	15	=(B$1+A16-1)/(A16*(B$2+1))*C15	=(A16=0)*B$3+(1-B$3)*C16	=B16*LN(D16)	=B$29*D16		=A16	=B16	=F16	=(I16-J16)^2/J16
17	9	10	=(B$1+A17-1)/(A17*(B$2+1))*C16	=(A17=0)*B$3+(1-B$3)*C17	=B17*LN(D17)	=B$29*D17		=A17	=B17	=F17	=(I17-J17)^2/J17
18	10	10	=(B$1+A18-1)/(A18*(B$2+1))*C17	=(A18=0)*B$3+(1-B$3)*C18	=B18*LN(D18)	=B$29*D18		=A18	=B18	=F18	=(I18-J18)^2/J18
19	11	10	=(B$1+A19-1)/(A19*(B$2+1))*C18	=(A19=0)*B$3+(1-B$3)*C19	=B19*LN(D19)	=B$29*D19		=A19	=B19	=F19	=(I19-J19)^2/J19
20	12	10	=(B$1+A20-1)/(A20*(B$2+1))*C19	=(A20=0)*B$3+(1-B$3)*C20	=B20*LN(D20)	=B$29*D20		=A20	=B20	=F20	=(I20-J20)^2/J20
21	13	3	=(B$1+A21-1)/(A21*(B$2+1))*C20	=(A21=0)*B$3+(1-B$3)*C21	=B21*LN(D21)	=B$29*D21		=A21	=B21	=F21	=(I21-J21)^2/J21
22	14	3	=(B$1+A22-1)/(A22*(B$2+1))*C21	=(A22=0)*B$3+(1-B$3)*C22	=B22*LN(D22)	=B$29*D22		=A22	=B22	=F22	=(I22-J22)^2/J22
23	15	5	=(B$1+A23-1)/(A23*(B$2+1))*C22	=(A23=0)*B$3+(1-B$3)*C23	=B23*LN(D23)	=B$29*D23		15+	=SUM(B23:B28)	=SUM(F23:F28)	=(I23-J23)^2/J23
24	16	5	=(B$1+A24-1)/(A24*(B$2+1))*C23	=(A24=0)*B$3+(1-B$3)*C24	=B24*LN(D24)	=B$29*D24					=SUM(K8:K23)
25	17	4	=(B$1+A25-1)/(A25*(B$2+1))*C24	=(A25=0)*B$3+(1-B$3)*C25	=B25*LN(D25)	=B$29*D25					
26	18	1	=(B$1+A26-1)/(A26*(B$2+1))*C25	=(A26=0)*B$3+(1-B$3)*C26	=B26*LN(D26)	=B$29*D26				# params	3
27	19	2	=(B$1+A27-1)/(A27*(B$2+1))*C26	=(A27=0)*B$3+(1-B$3)*C27	=B27*LN(D27)	=B$29*D27				df	=16-K26-1
28	20	1	=(B$1+A28-1)/(A28*(B$2+1))*C27	=(A28=0)*B$3+(1-B$3)*C28	=B28*LN(D28)	=B$29*D28					
29		=SUM(B8:B28)								p-value	=CHIDIST(K24,K27)

Problem 4 -- ZNBD

	A	B	C	D	E	F	G	H	I	J	K
1	r	1.504									
2	alpha	0.334									
3	pi	0.113									
4	LL	-1136.17									
5											
6			P(X=x)								
7	# Packs	Observed	NBD	ZNBD	LL	Expected		# Packs	Observed	Expected	(O-E)^2/E
8	0	102	0.12468	0.22368	-152.75	102.0		0	102	102.0	0.00
9	1	54	0.14054	0.12465	-112.44	56.8		1	54	56.8	0.14
10	2	49	0.13188	0.11697	-105.15	53.3		2	49	53.3	0.35
11	3	62	0.11545	0.10239	-141.29	46.7		3	62	46.7	5.02
12	4	44	0.09743	0.08641	-107.74	39.4		4	44	39.4	0.54
13	5	25	0.08039	0.07130	-66.02	32.5		5	25	32.5	1.74
14	6	26	0.06531	0.05793	-74.06	26.4		6	26	26.4	0.01
15	7	15	0.05248	0.04654	-46.01	21.2		7	15	21.2	1.82
16	8	15	0.04181	0.03708	-49.42	16.9		8	15	16.9	0.22
17	9	10	0.03309	0.02935	-35.28	13.4		9	10	13.4	0.86
18	10	10	0.02605	0.02311	-37.68	10.5		10	10	10.5	0.03
19	11	10	0.02042	0.01811	-40.11	8.3		11	10	8.3	0.37
20	12	10	0.01595	0.01415	-42.58	6.5		12	10	6.5	1.95
21	13	3	0.01242	0.01101	-13.53	5.0		13	3	5.0	0.81
22	14	3	0.00964	0.00855	-14.28	3.9		14	3	3.9	0.21
23	15	5	0.00747	0.00663	-25.08	3.0		15+	18	10.4	5.48
24	16	5	0.00578	0.00512	-26.37	2.3					19.54
25	17	4	0.00446	0.00395	-22.13	1.8					
26	18	1	0.00343	0.00305	-5.79	1.4				# params	3
27	19	2	0.00264	0.00234	-12.11	1.1				df	12
28	20	1	0.00203	0.00180	-6.32	0.8					
29		456								p-value	0.076

What is Wrong With the NBD Model?

The assumptions underlying the model could be wrong on two accounts:

i. at the individual-level, the number of purchases is not Poisson distributed

ii. purchase rates (λ) are not gamma-distributed

Relaxing the Gamma Assumption

- Replace the continuous distribution with a discrete distribution by allowing for multiple (discrete) segments each with a different (latent) buying rate:

$$P(X = x) = \sum_{s=1}^{S} \pi_s P(X = x|\lambda_s), \quad \sum_{s=1}^{S} \pi_s = 1$$

- This is called a finite mixture model.
- We often reparameterize the mixing proportions for computational convenience:

$$\pi_s = \frac{\exp(\theta_s)}{\sum_{s'=1}^{S} \exp(\theta_{s'})}, \quad \theta_S = 0.$$

Problem 4 -- Poisson

	A	B	C	D	E	F	G	H	I	J
1	lambda	3.9912280686								
2	LL	=SUM(D5:D25)								
3										
4	# Packs	Observed	P(X=x)	LL	Expected		# Packs	Observed	Expected	(O-E)^2/E
5	0	102	=POISSON(A5,B$1,FALSE)	=B5*LN(C5)	=B$26*C5		=A5	=B5	=E5	=(H5-I5)^2/I5
6	1	54	=POISSON(A6,B$1,FALSE)	=B6*LN(C6)	=B$26*C6		=A6	=B6	=E6	=(H6-I6)^2/I6
7	2	49	=POISSON(A7,B$1,FALSE)	=B7*LN(C7)	=B$26*C7		=A7	=B7	=E7	=(H7-I7)^2/I7
8	3	62	=POISSON(A8,B$1,FALSE)	=B8*LN(C8)	=B$26*C8		=A8	=B8	=E8	=(H8-I8)^2/I8
9	4	44	=POISSON(A9,B$1,FALSE)	=B9*LN(C9)	=B$26*C9		=A9	=B9	=E9	=(H9-I9)^2/I9
10	5	25	=POISSON(A10,B$1,FALSE)	=B10*LN(C10)	=B$26*C10		=A10	=B10	=E10	=(H10-I10)^2/I10
11	6	26	=POISSON(A11,B$1,FALSE)	=B11*LN(C11)	=B$26*C11		=A11	=B11	=E11	=(H11-I11)^2/I11
12	7	15	=POISSON(A12,B$1,FALSE)	=B12*LN(C12)	=B$26*C12		=A12	=B12	=E12	=(H12-I12)^2/I12
13	8	15	=POISSON(A13,B$1,FALSE)	=B13*LN(C13)	=B$26*C13		=A13	=B13	=E13	=(H13-I13)^2/I13
14	9	10	=POISSON(A14,B$1,FALSE)	=B14*LN(C14)	=B$26*C14		=A14	=B14	=E14	=(H14-I14)^2/I14
15	10	10	=POISSON(A15,B$1,FALSE)	=B15*LN(C15)	=B$26*C15		=A15	=B15	=E15	=(H15-I15)^2/I15
16	11	10	=POISSON(A16,B$1,FALSE)	=B16*LN(C16)	=B$26*C16		=A16	=B16	=E16	=(H16-I16)^2/I16
17	12	10	=POISSON(A17,B$1,FALSE)	=B17*LN(C17)	=B$26*C17		=A17	=B17	=E17	=(H17-I17)^2/I17
18	13	3	=POISSON(A18,B$1,FALSE)	=B18*LN(C18)	=B$26*C18		=A18	=B18	=E18	=(H18-I18)^2/I18
19	14	3	=POISSON(A19,B$1,FALSE)	=B19*LN(C19)	=B$26*C19		=A19	=B19	=E19	=(H19-I19)^2/I19
20	15	5	=POISSON(A20,B$1,FALSE)	=B20*LN(C20)	=B$26*C20		15+	=SUM(B20:B25)	=SUM(E20:E25)	=(H20-I20)^2/I20
21	16	5	=POISSON(A21,B$1,FALSE)	=B21*LN(C21)	=B$26*C21					=SUM(J5:J20)
22	17	4	=POISSON(A22,B$1,FALSE)	=B22*LN(C22)	=B$26*C22					
23	18	1	=POISSON(A23,B$1,FALSE)	=B23*LN(C23)	=B$26*C23				# params	1
24	19	2	=POISSON(A24,B$1,FALSE)	=B24*LN(C24)	=B$26*C24				df	=16-J23-1
25	20	1	=POISSON(A25,B$1,FALSE)	=B25*LN(C25)	=B$26*C25					
26		=SUM(B5:B25)							p-value	=CHIDIST(J21,J24)

Problem 4 -- Poisson

	A	B	C	D	E	F	G	H	I	J
1	lambda	3.991								
2	LL	-1545.00								
3										
4	# Packs	Observed	P(X=x)	LL	Expected		# Packs	Observed	Expected	(O-E)^2/E
5	0	102	0.01848	-407.11	8.4		0	102	8.4	1039.25
6	1	54	0.07375	-140.78	33.6		1	54	33.6	12.34
7	2	49	0.14717	-93.89	67.1		2	49	67.1	4.89
8	3	62	0.19579	-101.10	89.3		3	62	89.3	8.34
9	4	44	0.19536	-71.85	89.1		4	44	89.1	22.82
10	5	25	0.15595	-46.46	71.1		5	25	71.1	29.90
11	6	26	0.10374	-58.91	47.3		6	26	47.3	9.59
12	7	15	0.05915	-42.42	27.0		7	15	27.0	5.31
13	8	15	0.02951	-52.85	13.5		8	15	13.5	0.18
14	9	10	0.01309	-43.36	6.0		9	10	6.0	2.72
15	10	10	0.00522	-52.55	2.4		10	10	2.4	24.37
16	11	10	0.00190	-62.68	0.9		11	10	0.9	96.58
17	12	10	0.00063	-73.69	0.3		12	10	0.3	328.19
18	13	3	0.00019	-25.65	0.1		13	3	0.1	96.07
19	14	3	0.00006	-29.42	0.0		14	3	0.0	351.76
20	15	5	0.00001	-55.65	0.0		15+	18	0.0	36512.67
21	16	5	0.00000	-62.59	0.0					38544.98
22	17	4	0.00000	-55.87	0.0					
23	18	1	0.00000	-15.47	0.0				# params	1
24	19	2	0.00000	-34.07	0.0				df	14
25	20	1	0.00000	-18.64	0.0					
26		456							p-value	0.000

Problem 4 -- 2seg Poisson

	A	B	C	D	E	F	G
1	lambda_1	1.8021538					
2	lambda_2	9.1206784					
3	pi	0.7008857					
4	LL	=SUM(F8:F28)					
5	BIC	=-2*B4+L26*LN(B29)					
6							
7	# Packs	Observed	Seg1	Seg2	P(X=x)	LL	Expected
8	0	102	=POISSON(A8,B$1,FALSE)	=POISSON(A8,B$2,FALSE)	=B$3*C8+(1-B$3)*D8	=B8*LN(E8)	=B$29*E8
9	1	54	=POISSON(A9,B$1,FALSE)	=POISSON(A9,B$2,FALSE)	=B$3*C9+(1-B$3)*D9	=B9*LN(E9)	=B$29*E9
10	2	49	=POISSON(A10,B$1,FALSE)	=POISSON(A10,B$2,FALSE)	=B$3*C10+(1-B$3)*D10	=B10*LN(E10)	=B$29*E10
11	3	62	=POISSON(A11,B$1,FALSE)	=POISSON(A11,B$2,FALSE)	=B$3*C11+(1-B$3)*D11	=B11*LN(E11)	=B$29*E11
12	4	44	=POISSON(A12,B$1,FALSE)	=POISSON(A12,B$2,FALSE)	=B$3*C12+(1-B$3)*D12	=B12*LN(E12)	=B$29*E12
13	5	25	=POISSON(A13,B$1,FALSE)	=POISSON(A13,B$2,FALSE)	=B$3*C13+(1-B$3)*D13	=B13*LN(E13)	=B$29*E13
14	6	26	=POISSON(A14,B$1,FALSE)	=POISSON(A14,B$2,FALSE)	=B$3*C14+(1-B$3)*D14	=B14*LN(E14)	=B$29*E14
15	7	15	=POISSON(A15,B$1,FALSE)	=POISSON(A15,B$2,FALSE)	=B$3*C15+(1-B$3)*D15	=B15*LN(E15)	=B$29*E15
16	8	15	=POISSON(A16,B$1,FALSE)	=POISSON(A16,B$2,FALSE)	=B$3*C16+(1-B$3)*D16	=B16*LN(E16)	=B$29*E16
17	9	10	=POISSON(A17,B$1,FALSE)	=POISSON(A17,B$2,FALSE)	=B$3*C17+(1-B$3)*D17	=B17*LN(E17)	=B$29*E17
18	10	10	=POISSON(A18,B$1,FALSE)	=POISSON(A18,B$2,FALSE)	=B$3*C18+(1-B$3)*D18	=B18*LN(E18)	=B$29*E18
19	11	10	=POISSON(A19,B$1,FALSE)	=POISSON(A19,B$2,FALSE)	=B$3*C19+(1-B$3)*D19	=B19*LN(E19)	=B$29*E19
20	12	10	=POISSON(A20,B$1,FALSE)	=POISSON(A20,B$2,FALSE)	=B$3*C20+(1-B$3)*D20	=B20*LN(E20)	=B$29*E20
21	13	3	=POISSON(A21,B$1,FALSE)	=POISSON(A21,B$2,FALSE)	=B$3*C21+(1-B$3)*D21	=B21*LN(E21)	=B$29*E21
22	14	3	=POISSON(A22,B$1,FALSE)	=POISSON(A22,B$2,FALSE)	=B$3*C22+(1-B$3)*D22	=B22*LN(E22)	=B$29*E22
23	15	5	=POISSON(A23,B$1,FALSE)	=POISSON(A23,B$2,FALSE)	=B$3*C23+(1-B$3)*D23	=B23*LN(E23)	=B$29*E23
24	16	5	=POISSON(A24,B$1,FALSE)	=POISSON(A24,B$2,FALSE)	=B$3*C24+(1-B$3)*D24	=B24*LN(E24)	=B$29*E24
25	17	4	=POISSON(A25,B$1,FALSE)	=POISSON(A25,B$2,FALSE)	=B$3*C25+(1-B$3)*D25	=B25*LN(E25)	=B$29*E25
26	18	1	=POISSON(A26,B$1,FALSE)	=POISSON(A26,B$2,FALSE)	=B$3*C26+(1-B$3)*D26	=B26*LN(E26)	=B$29*E26
27	19	2	=POISSON(A27,B$1,FALSE)	=POISSON(A27,B$2,FALSE)	=B$3*C27+(1-B$3)*D27	=B27*LN(E27)	=B$29*E27
28	20	1	=POISSON(A28,B$1,FALSE)	=POISSON(A28,B$2,FALSE)	=B$3*C28+(1-B$3)*D28	=B28*LN(E28)	=B$29*E28
29		=SUM(B8:B28)					

Problem 4 -- 2seg Poisson

	A	B	C	D	E	F	G	H	I	J	K	L
1	lambda_1	1.802										
2	lambda_2	9.121										
3	pi	0.701										
4	LL	-1188.83										
5	BIC	2396.03										
6												
7	# Packs	Observed	Seg1	Seg2	P(X=x)	LL	Expected		# Packs	Observed	Expected	(O-E)^2/E
8	0	102	0.16494	0.00011	0.11564	-220.04	52.7		0	102	52.7	46.03
9	1	54	0.29725	0.00100	0.20864	-84.63	95.1		1	54	95.1	17.79
10	2	49	0.26785	0.00455	0.18909	-81.61	86.2		2	49	86.2	16.07
11	3	62	0.16090	0.01383	0.11691	-133.07	53.3		3	62	53.3	1.42
12	4	44	0.07249	0.03154	0.06024	-123.61	27.5		4	44	27.5	9.95
13	5	25	0.02613	0.05753	0.03552	-83.44	16.2		5	25	16.2	4.78
14	6	26	0.00785	0.08745	0.03166	-89.77	14.4		6	26	14.4	9.26
15	7	15	0.00202	0.11395	0.03550	-50.07	16.2		7	15	16.2	0.09
16	8	15	0.00046	0.12991	0.03918	-48.60	17.9		8	15	17.9	0.46
17	9	10	0.00009	0.13165	0.03944	-32.33	18.0		9	10	18.0	3.55
18	10	10	0.00002	0.12007	0.03593	-33.26	16.4		10	10	16.4	2.49
19	11	10	0.00000	0.09956	0.02978	-35.14	13.6		11	10	13.6	0.94
20	12	10	0.00000	0.07567	0.02263	-37.88	10.3		12	10	10.3	0.01
21	13	3	0.00000	0.05309	0.01588	-12.43	7.2		13	3	7.2	2.48
22	14	3	0.00000	0.03459	0.01035	-13.71	4.7		14	3	4.7	0.63
23	15	5	0.00000	0.02103	0.00629	-25.34	2.9		15+	18	6.1	22.94
24	16	5	0.00000	0.01199	0.00359	-28.15	1.6					138.88
25	17	4	0.00000	0.00643	0.00192	-25.01	0.9					
26	18	1	0.00000	0.00326	0.00097	-6.93	0.4				# params	3
27	19	2	0.00000	0.00156	0.00047	-15.33	0.2				df	12
28	20	1	0.00000	0.00071	0.00021	-8.45	0.1					
29		456									p-value	0.000

	A	B	C	D	E	F	G	H
1	lambda_1	3.483317						
2	lambda_2	11.21581						
3	lambda_3	0.290543						
4	theta_1	0.674427	=EXP(B4)					
5	theta_2	-0.43042	=EXP(B5)					
6	theta_3	0	=EXP(B6)					
7	LL	=SUM(G11:G31)						
8	BIC	=-2*B7+M29*LN(B32)						
9			=C4/SUM(C4:C6)	=C5/SUM(C4:C6)	=C6/SUM(C4:C6)			
10	# Packs	Observed	Seg1	Seg2	Seg3	P(X=x)	LL	Expected
11	0	102	=POISSON(A11,B$1,FALSE)	=POISSON(A11,B$2,FALSE)	=POISSON(A11,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C11:E11)	=B11*LN(F11)	=B$32*F11
12	1	54	=POISSON(A12,B$1,FALSE)	=POISSON(A12,B$2,FALSE)	=POISSON(A12,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C12:E12)	=B12*LN(F12)	=B$32*F12
13	2	49	=POISSON(A13,B$1,FALSE)	=POISSON(A13,B$2,FALSE)	=POISSON(A13,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C13:E13)	=B13*LN(F13)	=B$32*F13
14	3	62	=POISSON(A14,B$1,FALSE)	=POISSON(A14,B$2,FALSE)	=POISSON(A14,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C14:E14)	=B14*LN(F14)	=B$32*F14
15	4	44	=POISSON(A15,B$1,FALSE)	=POISSON(A15,B$2,FALSE)	=POISSON(A15,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C15:E15)	=B15*LN(F15)	=B$32*F15
16	5	25	=POISSON(A16,B$1,FALSE)	=POISSON(A16,B$2,FALSE)	=POISSON(A16,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C16:E16)	=B16*LN(F16)	=B$32*F16
17	6	26	=POISSON(A17,B$1,FALSE)	=POISSON(A17,B$2,FALSE)	=POISSON(A17,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C17:E17)	=B17*LN(F17)	=B$32*F17
18	7	15	=POISSON(A18,B$1,FALSE)	=POISSON(A18,B$2,FALSE)	=POISSON(A18,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C18:E18)	=B18*LN(F18)	=B$32*F18
19	8	15	=POISSON(A19,B$1,FALSE)	=POISSON(A19,B$2,FALSE)	=POISSON(A19,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C19:E19)	=B19*LN(F19)	=B$32*F19
20	9	10	=POISSON(A20,B$1,FALSE)	=POISSON(A20,B$2,FALSE)	=POISSON(A20,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C20:E20)	=B20*LN(F20)	=B$32*F20
21	10	10	=POISSON(A21,B$1,FALSE)	=POISSON(A21,B$2,FALSE)	=POISSON(A21,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C21:E21)	=B21*LN(F21)	=B$32*F21
22	11	10	=POISSON(A22,B$1,FALSE)	=POISSON(A22,B$2,FALSE)	=POISSON(A22,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C22:E22)	=B22*LN(F22)	=B$32*F22
23	12	10	=POISSON(A23,B$1,FALSE)	=POISSON(A23,B$2,FALSE)	=POISSON(A23,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C23:E23)	=B23*LN(F23)	=B$32*F23
24	13	3	=POISSON(A24,B$1,FALSE)	=POISSON(A24,B$2,FALSE)	=POISSON(A24,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C24:E24)	=B24*LN(F24)	=B$32*F24
25	14	3	=POISSON(A25,B$1,FALSE)	=POISSON(A25,B$2,FALSE)	=POISSON(A25,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C25:E25)	=B25*LN(F25)	=B$32*F25
26	15	5	=POISSON(A26,B$1,FALSE)	=POISSON(A26,B$2,FALSE)	=POISSON(A26,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C26:E26)	=B26*LN(F26)	=B$32*F26
27	16	5	=POISSON(A27,B$1,FALSE)	=POISSON(A27,B$2,FALSE)	=POISSON(A27,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C27:E27)	=B27*LN(F27)	=B$32*F27
28	17	4	=POISSON(A28,B$1,FALSE)	=POISSON(A28,B$2,FALSE)	=POISSON(A28,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C28:E28)	=B28*LN(F28)	=B$32*F28
29	18	1	=POISSON(A29,B$1,FALSE)	=POISSON(A29,B$2,FALSE)	=POISSON(A29,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C29:E29)	=B29*LN(F29)	=B$32*F29
30	19	2	=POISSON(A30,B$1,FALSE)	=POISSON(A30,B$2,FALSE)	=POISSON(A30,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C30:E30)	=B30*LN(F30)	=B$32*F30
31	20	1	=POISSON(A31,B$1,FALSE)	=POISSON(A31,B$2,FALSE)	=POISSON(A31,B$3,FALSE)	=SUMPRODUCT(C$9:E$9,C31:E31)	=B31*LN(F31)	=B$32*F31
32		=SUM(B11:B31)						

Problem 4 -- 3seg Poisson

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	lambda_1	3.483											
2	lambda_2	11.216											
3	lambda_3	0.291											
4	theta_1	0.674	1.962908										
5	theta_2	-0.430	0.650233										
6	theta_3	0	1										
7	LL	-1132.04											
8	BIC	2294.70											
9			0.54327	0.17996	0.27677								
10	# Packs	Observed	Seg1	Seg2	Seg3	P(X=x)	LL	Expected		# Packs	Observed	Expected	(O-E)^2/E
11	0	102	0.03071	0.00001	0.74786	0.22367	-152.76	102.0		0	102	102.0	0.00
12	1	54	0.10696	0.00015	0.21728	0.11827	-115.28	53.9		1	54	53.9	0.00
13	2	49	0.18628	0.00085	0.03157	0.11009	-108.12	50.2		2	49	50.2	0.03
14	3	62	0.21629	0.00317	0.00306	0.11892	-132.02	54.2		3	62	54.2	1.11
15	4	44	0.18835	0.00887	0.00022	0.10399	-99.59	47.4		4	44	47.4	0.25
16	5	25	0.13122	0.01991	0.00001	0.07487	-64.80	34.1		5	25	34.1	2.45
17	6	26	0.07618	0.03721	0.00000	0.04808	-78.91	21.9		6	26	21.9	0.76
18	7	15	0.03791	0.05962	0.00000	0.03132	-51.95	14.3		7	15	14.3	0.04
19	8	15	0.01651	0.08359	0.00000	0.02401	-55.94	10.9		8	15	10.9	1.50
20	9	10	0.00639	0.10417	0.00000	0.02222	-38.07	10.1		9	10	10.1	0.00
21	10	10	0.00223	0.11684	0.00000	0.02224	-38.06	10.1		10	10	10.1	0.00
22	11	10	0.00070	0.11913	0.00000	0.02182	-38.25	10.0		11	10	10.0	0.00
23	12	10	0.00020	0.11134	0.00000	0.02015	-39.05	9.2		12	10	9.2	0.07
24	13	3	0.00005	0.09606	0.00000	0.01732	-12.17	7.9		13	3	7.9	3.04
25	14	3	0.00001	0.07696	0.00000	0.01386	-12.84	6.3		14	3	6.3	1.74
26	15	5	0.00000	0.05754	0.00000	0.01036	-22.85	4.7		15+	18	12.8	2.08
27	16	5	0.00000	0.04034	0.00000	0.00726	-24.63	3.3					13.07
28	17	4	0.00000	0.02661	0.00000	0.00479	-21.37	2.2					
29	18	1	0.00000	0.01658	0.00000	0.00298	-5.81	1.4				# params	5
30	19	2	0.00000	0.00979	0.00000	0.00176	-12.68	0.8				df	10
31	20	1	0.00000	0.00549	0.00000	0.00099	-6.92	0.5					
32		456										p-value	0.220

Parameter Estimates

	Seg 1	Seg 2	Seg 3	LL
λ	3.991			−1545.00
λ_s	1.802	9.121		−1188.83
π_s	0.701	0.299		
λ_s	0.291	3.483	11.216	−1132.04
π_s	0.277	0.543	0.180	

How Many Segments?

- Controlling for the extra parameters, is an $S + 1$ segment model better than an S segment model?
- We can't use the likelihood ratio test because its properties are violated
- It is standard practice to use "information-theoretic" model selection criteria
- A common measure is the Bayesian information criterion:

$$\text{BIC} = -2LL + p\ln(N)$$

  where p is the number of parameters and N is the sample size
- Rule: choose S to minimize BIC

Problem 4 -- 4seg Poisson

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	lambda_1	3.002												
2	lambda_2	0.205												
3	lambda_3	7.418												
4	lambda_4	12.873												
5	theta_1	1.598	4.943453											
6	theta_2	0.876	2.401291											
7	theta_3	0.398	1.488702											
8	theta_4	0	1											
9	LL	-1130.07												
10	BIC	2303.00												
11			0.50272	0.24420	0.15139	0.10169								
12	# Packs	Observed	Seg1	Seg2	Seg3	Seg4	P(X=x)	LL	Expected		# Packs	Observed	Expected	(O-E)^2/E
13	0	102	0.04969	0.81487	0.00060	0.00000	0.22406	-152.58	102.2		0	102	102.2	0.00
14	1	54	0.14917	0.16683	0.00445	0.00003	0.11641	-116.14	53.1		1	54	53.1	0.02
15	2	49	0.22390	0.01708	0.01652	0.00021	0.11925	-104.20	54.4		2	49	54.4	0.53
16	3	62	0.22404	0.00117	0.04084	0.00091	0.11919	-131.88	54.4		3	62	54.4	1.08
17	4	44	0.16814	0.00006	0.07574	0.00294	0.09631	-102.97	43.9		4	44	43.9	0.00
18	5	25	0.10095	0.00000	0.11237	0.00756	0.06853	-67.01	31.2		5	25	31.2	1.25
19	6	26	0.05051	0.00000	0.13892	0.01622	0.04807	-78.91	21.9		6	26	21.9	0.76
20	7	15	0.02166	0.00000	0.14722	0.02984	0.03621	-49.78	16.5		7	15	16.5	0.14
21	8	15	0.00813	0.00000	0.13652	0.04801	0.02964	-52.78	13.5		8	15	13.5	0.16
22	9	10	0.00271	0.00000	0.11252	0.06867	0.02538	-36.74	11.6		9	10	11.6	0.21
23	10	10	0.00081	0.00000	0.08347	0.08839	0.02203	-38.15	10.0		10	10	10.0	0.00
24	11	10	0.00022	0.00000	0.05629	0.10344	0.01915	-39.55	8.7		11	10	8.7	0.18
25	12	10	0.00006	0.00000	0.03480	0.11096	0.01658	-41.00	7.6		12	10	7.6	0.79
26	13	3	0.00001	0.00000	0.01986	0.10987	0.01419	-12.77	6.5		13	3	6.5	1.86
27	14	3	0.00000	0.00000	0.01052	0.10102	0.01187	-13.30	5.4		14	3	5.4	1.07
28	15	5	0.00000	0.00000	0.00520	0.08669	0.00960	-23.23	4.4		15+	18	14.1	1.11
29	16	5	0.00000	0.00000	0.00241	0.06975	0.00746	-24.49	3.4					9.17
30	17	4	0.00000	0.00000	0.00105	0.05281	0.00553	-20.79	2.5					
31	18	1	0.00000	0.00000	0.00043	0.03777	0.00391	-5.55	1.8				# params	7
32	19	2	0.00000	0.00000	0.00017	0.02559	0.00263	-11.88	1.2				df	8
33	20	1	0.00000	0.00000	0.00006	0.01647	0.00168	-6.39	0.8					
34		456											p-value	0.329

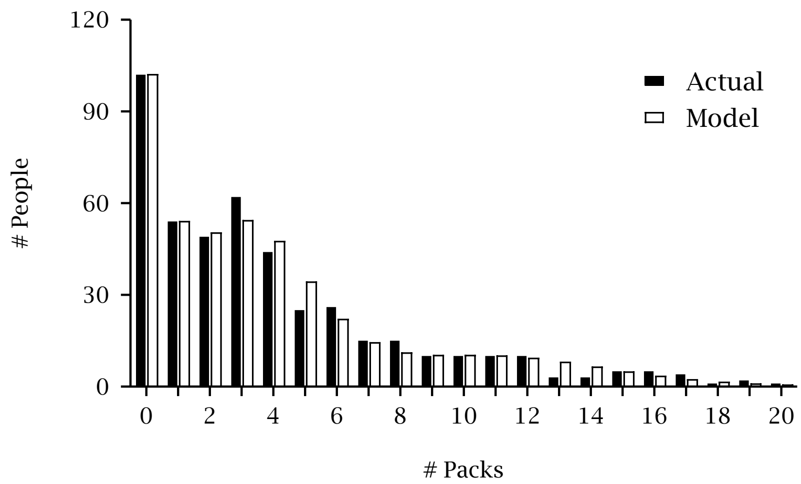
Summary of Model Fit

Model	LL	# params	BIC	χ^2 p-value
NBD	−1140.02	2	2292.29	0.04
ZNBD	−1136.17	3	2290.70	0.08
Poisson	−1545.00	1	3096.12	0.00
2 seg Poisson	−1188.83	3	2396.03	0.00
3 seg Poisson	−1132.04	5	2294.70	0.22
4 seg Poisson	−1130.07	7	2303.00	0.33

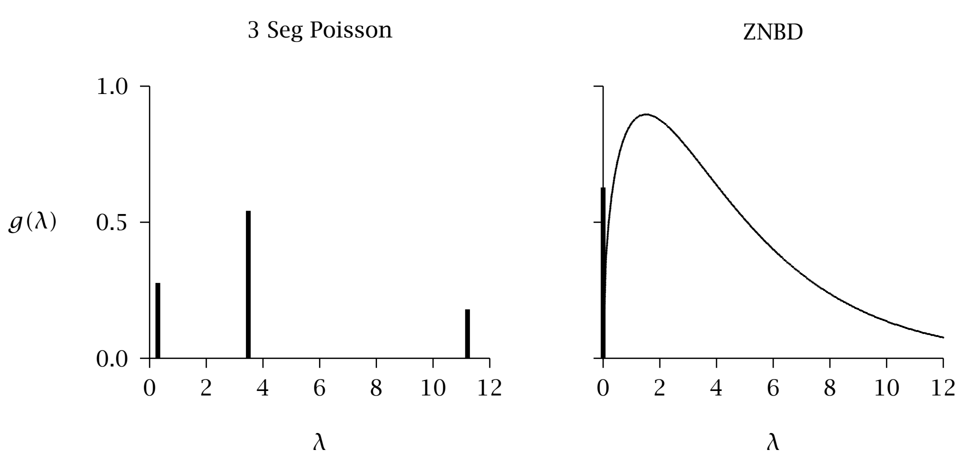
LatentGOLD Results

	Seg 1	Seg 2	Seg 3	Seg 4	LL
mean	3.991				−1545.00
class size	1.000				
mean	1.801	9.115			−1188.83
class size	0.700	0.300			
mean	3.483	0.291	11.210		−1132.04
class size	0.542	0.277	0.181		
mean	2.976	0.202	7.247	12.787	−1130.07
class size	0.500	0.243	0.156	0.106	

Fit of 3 Segment Poisson



Implied Heterogeneity Distribution



$\hat{\pi} = 0.113, \hat{r} = 1.504, \hat{\alpha} = 0.334$

Classification Using Bayes Theorem

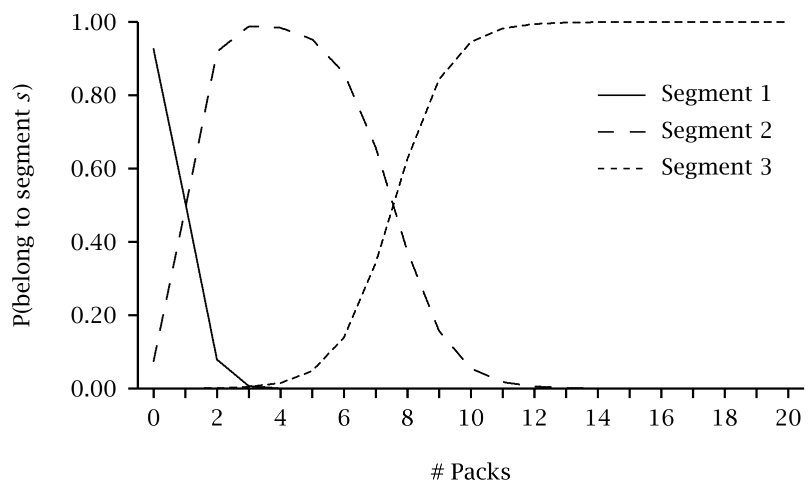
To which "segment" of the mixing distribution does each observation x belong?

- The π_s can be interpreted as the prior probability of λ_s
- By Bayes theorem,

$$P(s \mid X = x) = \frac{P(X = x \mid \lambda_s)\pi_s}{\sum_{s'=1}^{S} P(X = x \mid \lambda_{s'})\pi_{s'}},$$

  which can be interpreted as the posterior probability of λ_s

Posterior Probabilities



Conditional Expectations

What is the expected purchase quantity over the next month for a customer who purchased seven packs last week?

$$
\begin{aligned}
E[X(4)] &= E[X(4)|\text{seg }1]\,P(\text{seg }1|X=7) \\
&\quad + E[X(4)|\text{seg }2]\,P(\text{seg }2|X=7) \\
&\quad + E[X(4)|\text{seg }3]\,P(\text{seg }3|X=7) \\
&= (4\times 0.291)\times 0.0000 \\
&\quad + (4\times 3.483)\times 0.6575 \\
&\quad + (4\times 11.216)\times 0.3425 \\
&= 24.5
\end{aligned}
$$

… or 13.9 with “hard assignment” to segment 2.

Concepts and Tools Introduced

- Finite mixture models
- Discrete vs. continuous mixing distributions
- Probability models for classification

Further Reading

Dillon, Wiliam R. and Ajith Kumar (1994), "Latent Structure and Other Mixture Models in Marketing: An Integrative Survey and Overview," in Richard P. Bagozzi (ed.), *Advanced Methods of Marketing Research*, Oxford: Blackwell.

McLachlan, Geoffrey and David Peel (2000), *Finite Mixture Models*, New York: John Wiley & Sons.

Wedel, Michel and Wagner A. Kamakura (1999), *Market Segmentation: Conceptual and Methodological Foundations*, 2nd edn., Boston, MA: Kluwer Academic Publishers.

Problem 5:
Who is Visiting khakichinos.com?

(Incorporating Covariates in Count Models)

Background

Khaki Chinos, Inc. is an established clothing catalog company with an online presence at khakichinos.com. While the company is able to track the online *purchasing* behavior of its customers, it has no real idea as to the pattern of *visiting* behaviors by the broader Internet population.

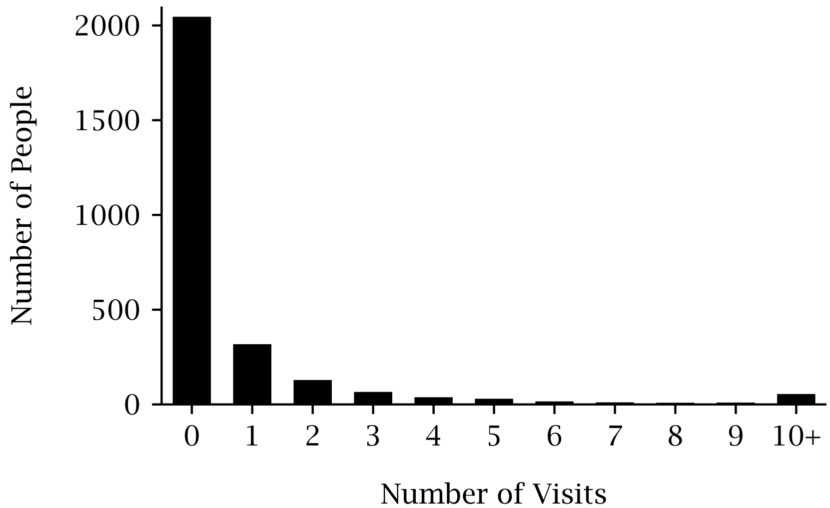
In order to gain an understanding of the aggregate visiting patterns, some Media Metrix panel data has been purchased. For a sample of 2728 people who visited an online apparel site at least once during the second-half of 2000, the dataset reports how many visits each person made to the khakichinos.com web site, along with some demographic information.

Management would like to know whether frequency of visiting the web site is related to demographic characteristics.

Raw Data

ID	# Visits	ln(Income)	Sex	ln(Age)	HH Size
1	0	11.38	1	3.87	2
2	5	9.77	1	4.04	1
3	0	11.08	0	3.33	2
4	0	10.92	1	3.95	3
5	0	10.92	1	2.83	3
6	0	10.92	0	2.94	3
7	0	11.19	0	3.66	2
8	1	11.74	0	4.08	2
9	0	10.02	0	4.25	1
	…				

Distribution of Visits



Modeling Count Data

Recall the NBD:

- At the individual-level, $Y \sim \text{Poisson}(\lambda)$
- λ is distributed across the population according to a gamma distribution with parameters r and α

$$P(Y = y) = \frac{\Gamma(r + y)}{\Gamma(r)y!} \left(\frac{\alpha}{\alpha + 1}\right)^r \left(\frac{1}{\alpha + 1}\right)^y$$

Observed vs. Unobserved Heterogeneity

Unobserved Heterogeneity:

- People differ in their mean (visiting) rate λ
- To account for heterogeneity in λ, we assume it is distributed across the population according to some (parametric) distribution
- But there is no attempt to *explain* how people differ in their mean rates

Observed Heterogeneity:

- We observe how people differ on a set of observable independent (explanatory) variables
- We explicitly link an individual's λ to her observable characteristics

The Poisson Regression Model

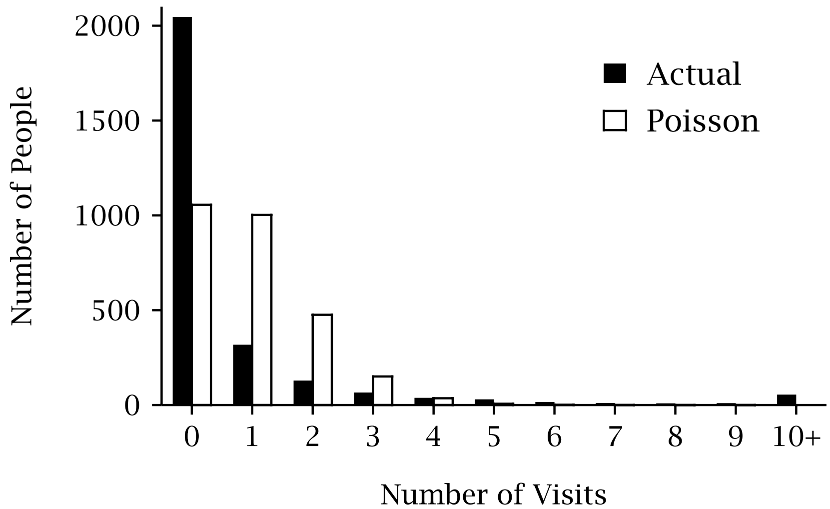
- Let the random variable Y_i denote the number of times individual i visits the site in a unit time period
- At the individual-level, Y_i is assumed to be distributed Poisson with mean λ_i:

$$P(Y_i = y|\lambda_i) = \frac{\lambda_i^y e^{-\lambda_i}}{y!}$$

- An individual's mean is related to her observable characteristics through the function

$$\lambda_i = \lambda_0 \exp(\boldsymbol{\beta}'\mathbf{x}_i)$$

Fit of Poisson



$\hat{\lambda} = 0.949, LL = -6378.6$

Poisson Regression Results

Variable	Coefficient
λ_0	0.0439
Income	0.0938
Sex	0.0043
Age	0.5882
HH Size	−0.0359
LL	−6291.5
LL_{Poiss}	−6378.6
LR (df = 4)	174.2

	A	B	C	D	E	F	G	H	I	J	K
1	\lambda_0	0.04387			LL =	=SUM(K9:K2736)					
2	B_inc	0.09385									
3	B_sex	0.00426									
4	B_age	0.58825									
5	B_size	-0.0359									
6				=TRANSPOSE(B2:B5)	=TRANSPOSE(B2:B5)	=TRANSPOSE(B2:B5)	=TRANSPOSE(B2:B5)				
7											
8	ID	Total		Income	Sex	Age	Size		lambda	P(Y=y)	ln[P(Y=y)]
9	1	0		11.3793940723457	1	3.87120101090789	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D9:G9))	=POISSON(B9,I9,FALSE)	=LN(J9)
10	2	5		9.76995615991161	1	4.04305126783455	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D10:G10))	=POISSON(B10,I10,FALSE)	=LN(J10)
11	3	0		11.0821425488778	0	3.3322045101752	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D11:G11))	=POISSON(B11,I11,FALSE)	=LN(J11)
12	4	0		10.9150884642146	1	3.95124371858143	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D12:G12))	=POISSON(B12,I12,FALSE)	=LN(J12)
13	5	0		10.9150884642146	1	2.83321334405622	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D13:G13))	=POISSON(B13,I13,FALSE)	=LN(J13)
14	6	0		10.9150884642146	0	2.94443897916644	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D14:G14))	=POISSON(B14,I14,FALSE)	=LN(J14)
15	7	0		11.1913418408428	0	3.66356164612965	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D15:G15))	=POISSON(B15,I15,FALSE)	=LN(J15)
16	8	1		11.7360690162844	0	4.07753744390572	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D16:G16))	=POISSON(B16,I16,FALSE)	=LN(J16)
17	9	0		10.0212705881925	0	4.24849524204936	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D17:G17))	=POISSON(B17,I17,FALSE)	=LN(J17)
18	10	0		10.9150884642146	0	3.85014760171006	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D18:G18))	=POISSON(B18,I18,FALSE)	=LN(J18)
19	11	1		10.7684849900227	0	3.93182563272433	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D19:G19))	=POISSON(B19,I19,FALSE)	=LN(J19)
20	12	0		10.9150884642146	0	3.98898404656427	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D20:G20))	=POISSON(B20,I20,FALSE)	=LN(J20)
21	13	3		10.5320962119585	0	3.63758615972639	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D21:G21))	=POISSON(B21,I21,FALSE)	=LN(J21)
22	14	0		10.9150884642146	0	3.61091791264422	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D22:G22))	=POISSON(B22,I22,FALSE)	=LN(J22)
23	15	0		10.2219412836547	1	3.58351893845611	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D23:G23))	=POISSON(B23,I23,FALSE)	=LN(J23)
24	16	1		10.7684849900227	1	3.25809653802148	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D24:G24))	=POISSON(B24,I24,FALSE)	=LN(J24)
25	17	2		12.2060726455302	0	3.66356164612965	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D25:G25))	=POISSON(B25,I25,FALSE)	=LN(J25)
26	18	0		10.7684849900227	0	3.95124371858143	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D26:G26))	=POISSON(B26,I26,FALSE)	=LN(J26)
27	19	6		11.1913418408428	1	3.3322045101752	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D27:G27))	=POISSON(B27,I27,FALSE)	=LN(J27)
28	20	0		10.3889953683178	1	3.58351893845611	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D28:G28))	=POISSON(B28,I28,FALSE)	=LN(J28)
29	21	2		10.7684849900227	1	3.3322045101752	4		=B$1*EXP(SUMPRODUCT(D$6:G$6,D29:G29))	=POISSON(B29,I29,FALSE)	=LN(J29)
30	22	0		11.1913418408428	1	3.46573590279973	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D30:G30))	=POISSON(B30,I30,FALSE)	=LN(J30)
31	23	0		11.1913418408428	1	3.43398720448515	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D31:G31))	=POISSON(B31,I31,FALSE)	=LN(J31)
32	24	2		11.7360690162844	1	3.80666248977032	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D32:G32))	=POISSON(B32,I32,FALSE)	=LN(J32)
33	25	0		11.3793940723457	0	4.27666611901606	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D33:G33))	=POISSON(B33,I33,FALSE)	=LN(J33)
34	26	0		10.3889953683178	0	4.21950770517611	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D34:G34))	=POISSON(B34,I34,FALSE)	=LN(J34)
35	27	0		10.6572593549125	1	3.49650756146648	4		=B$1*EXP(SUMPRODUCT(D$6:G$6,D35:G35))	=POISSON(B35,I35,FALSE)	=LN(J35)
36	28	0		12.0725412529057	0	3.95124371858143	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D36:G36))	=POISSON(B36,I36,FALSE)	=LN(J36)
37	29	0		10.9150884642146	1	3.80666248977032	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D37:G37))	=POISSON(B37,I37,FALSE)	=LN(J37)
38	30	0		10.9150884642146	0	3.52636052461616	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D38:G38))	=POISSON(B38,I38,FALSE)	=LN(J38)
39	31	0		11.1913418408428	1	3.36729582998647	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D39:G39))	=POISSON(B39,I39,FALSE)	=LN(J39)
40	32	0		10.2219412836547	1	3.13549421592915	4		=B$1*EXP(SUMPRODUCT(D$6:G$6,D40:G40))	=POISSON(B40,I40,FALSE)	=LN(J40)
41	33	0		11.3793940723457	0	3.3322045101752	4		=B$1*EXP(SUMPRODUCT(D$6:G$6,D41:G41))	=POISSON(B41,I41,FALSE)	=LN(J41)
42	34	0		9.07680897935166	1	3.40119738166216	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D42:G42))	=POISSON(B42,I42,FALSE)	=LN(J42)
43	35	0		10.0212705881925	1	3.52636052461616	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D43:G43))	=POISSON(B43,I43,FALSE)	=LN(J43)
44	36	0		11.0821425488778	0	4.06044301054642	4		=B$1*EXP(SUMPRODUCT(D$6:G$6,D44:G44))	=POISSON(B44,I44,FALSE)	=LN(J44)
45	37	2		10.2219412836547	1	3.68887945411394	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D45:G45))	=POISSON(B45,I45,FALSE)	=LN(J45)
46	38	2		12.0725412529057	1	3.68887945411394	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D46:G46))	=POISSON(B46,I46,FALSE)	=LN(J46)
47	39	1		11.0821425488778	0	4.17438726989564	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D47:G47))	=POISSON(B47,I47,FALSE)	=LN(J47)
48	40	0		9.52879410309472	1	2.70805020110221	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D48:G48))	=POISSON(B48,I48,FALSE)	=LN(J48)
49	41	0		11.0821425488778	1	3.80666248977032	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D49:G49))	=POISSON(B49,I49,FALSE)	=LN(J49)
50	42	0		11.3793940723457	1	4.12713438504509	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D50:G50))	=POISSON(B50,I50,FALSE)	=LN(J50)
51	43	0		11.3793940723457	0	4.17438726989564	3		=B$1*EXP(SUMPRODUCT(D$6:G$6,D51:G51))	=POISSON(B51,I51,FALSE)	=LN(J51)
52	44	0		10.5320962119585	1	3.55534806148941	6		=B$1*EXP(SUMPRODUCT(D$6:G$6,D52:G52))	=POISSON(B52,I52,FALSE)	=LN(J52)
53	45	0		10.7684849900227	0	3.2188758248682	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D53:G53))	=POISSON(B53,I53,FALSE)	=LN(J53)
54	46	0		11.3793940723457	1	3.36729582998647	2		=B$1*EXP(SUMPRODUCT(D$6:G$6,D54:G54))	=POISSON(B54,I54,FALSE)	=LN(J54)
55	47	0		11.7360690162844	0	3.04452243772342	4		=B$1*EXP(SUMPRODUCT(D$6:G$6,D55:G55))	=POISSON(B55,I55,FALSE)	=LN(J55)
56	48	0		10.7684849900227	0	3.52636052461616	1		=B$1*EXP(SUMPRODUCT(D$6:G$6,D56:G56))	=POISSON(B56,I56,FALSE)	=LN(J56)

Problem 5 -- Poisson reg

	A	B	C	D	E	F	G	H	I	J
1	\lambda_0	0.0439			LL =	-6291.497				
2	B_inc	0.0938								
3	B_sex	0.0043								
4	B_age	0.5882								
5	B_size	-0.0359								
6				0.0938	0.0043	0.5882	-0.0359			
7										
8	ID	Total		Income	Sex	Age	HH Size		lambda	P(Y=y)
9	1	0		11.38	1	3.87	2		1.16317	0.31249
10	2	5		9.77	1	4.04	1		1.14695	0.00525
11	3	0		11.08	0	3.33	2		0.82031	0.44029
12	4	0		10.92	1	3.95	3		1.12609	0.32430
13	5	0		10.92	1	2.83	3		0.58338	0.55801
14	6	0		10.92	0	2.94	3		0.62017	0.53785
15	7	0		11.19	0	3.66	2		1.00712	0.36527
16	8	1		11.74	0	4.08	2		1.35220	0.34977
17	9	0		10.02	0	4.25	1		1.31954	0.26726
18	10	0		10.92	0	3.85	3		1.05656	0.34765
19	11	1		10.77	0	3.93	2		1.13340	0.36488
20	12	0		10.92	0	3.99	2		1.18839	0.30471
21	13	3		10.53	0	3.64	2		0.93235	0.05317
22	14	0		10.92	0	3.61	1		0.98621	0.37299
23	15	0		10.22	1	3.58	3		0.84992	0.42745
24	16	1		10.77	1	3.26	3		0.73879	0.35291
25	17	2		12.21	0	3.66	2		1.10774	0.20266
26	18	0		10.77	0	3.95	2		1.14642	0.31777
27	19	6		11.19	1	3.33	2		0.83230	0.00020
28	20	0		10.39	1	3.58	2		0.89492	0.40864
29	21	2		10.77	1	3.33	4		0.74449	0.13163
30	22	0		11.19	1	3.47	2		0.90031	0.40644
31	23	0		11.19	1	3.43	2		0.88365	0.41327
32	24	2		11.74	1	3.81	2		1.15796	0.21060
33	25	0		11.38	0	4.28	2		1.47020	0.22988
34	26	0		10.39	0	4.22	2		1.29542	0.27378
35	27	0		10.66	1	3.50	4		0.81152	0.44418
36	28	0		12.07	0	3.95	2		1.29566	0.27372
37	29	0		10.92	1	3.81	3		1.03428	0.35548
38	30	0		10.92	0	3.53	3		0.87333	0.41756
39	31	0		11.19	1	3.37	2		0.84966	0.42756
40	32	0		10.22	1	3.14	4		0.62998	0.53260
41	33	0		11.38	0	3.33	4		0.78506	0.45609
42	34	0		9.08	1	3.40	1		0.73675	0.47867
43	35	0		10.02	1	3.53	1		0.86654	0.42040
44	36	0		11.08	0	4.06	4		1.17175	0.30982
45	37	2		10.22	1	3.69	2		0.93733	0.17206
46	38	2		12.07	1	3.69	2		1.11510	0.20385
47	39	1		11.08	0	4.17	1		1.39549	0.34568
48	40	0		9.53	1	2.71	3		0.47585	0.62136
49	41	0		11.08	1	3.81	3		1.05062	0.34972
50	42	0		11.38	1	4.13	3		1.30446	0.27132
51	43	0		11.38	0	4.17	3		1.33553	0.26302
52	44	0		10.53	1	3.56	6		0.77275	0.46174

Comparing Expected Visit Behavior

	Person A	Person B
Income	59,874	98,716
Sex	M	F
Age	55	33
HH Size	4	2

Who is less likely to have visited the web site?

$$\begin{aligned}\lambda_A &= 0.0439 \times \exp\big(0.0938 \times \ln(59,874) + 0.0043 \times 0 \\ &\quad + 0.5882 \times \ln(55) - 0.0359 \times 4\big) \\ &= 1.127 \\ \lambda_B &= 0.0439 \times \exp\big(0.0938 \times \ln(98,716) + 0.0043 \times 1 \\ &\quad + 0.5882 \times \ln(33) - 0.0359 \times 2\big) \\ &= 0.944\end{aligned}$$

Is $\boldsymbol{\beta}$ Different from 0?

Consider two models, A and B:

> If we can arrive at model B by placing k constraints on the parameters of model A, we say that model B is *nested* within model A.

The Poisson model is nested within the Poisson regression model by imposing the constraint $\boldsymbol{\beta} = \mathbf{0}$.

We use the *likelihood ratio test* to determine whether model A, which has more parameters, fits the data better than model B.

The Likelihood Ratio Test

- The null hypothesis is that model A is not different from model B
- Compute the test statistic

$$LR = -2(LL_B - LL_A)$$

- Reject null hypothesis if $LR > \chi^2_{.05,k}$

Computing Standard Errors

- Excel
  - indirectly via a series of likelihood ratio tests
- General modeling environments (e.g., MATLAB, Gauss)
  - easily computed from the Hessian matrix (computed directly or as a by-product of optimization)
- Advanced statistics packages (e.g., Limdep, S-Plus)
  - they come for free

S-Plus Poisson Regression Results

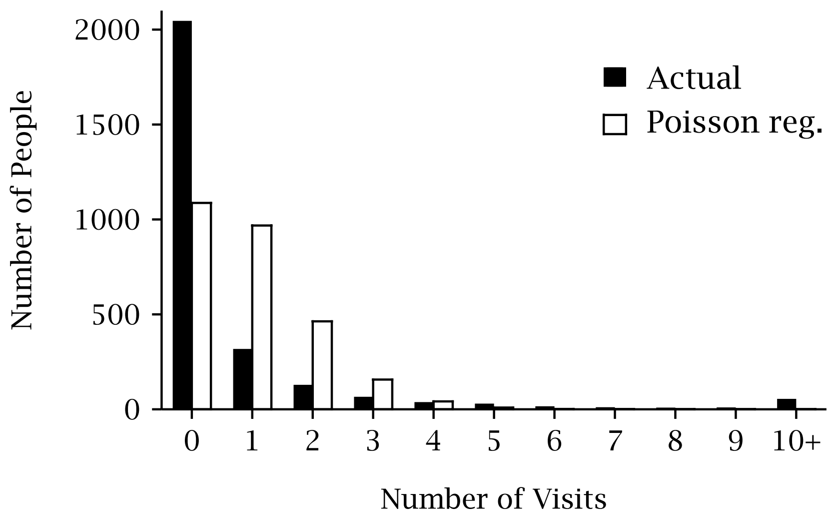
Coefficients:

	Value	Std. Error	t value
(Intercept)	-3.126238804	0.40578080	-7.7042552
Income	0.093828021	0.03436347	2.7304580
Sex	0.004259338	0.04089411	0.1041553
Age	0.588249213	0.05472896	10.7484079
HH Size	-0.035907406	0.01528397	-2.3493511

Limdep Poisson Regression Results

Variable	Coefficient	Standard Error	b/St.Er.
Constant	-3.122103284	.40565119	-7.697
INCOME	.9305546493E-01	.34332533E-01	2.710
SEX	.4312514407E-02	.40904265E-01	.105
AGE	.5893014445	.54790230E-01	10.756
HH SIZE	-.3577795361E-01	.15287122E-01	-2.340

Fit of Poisson Regression



The ZIP Regression Model

Because of the "excessive" number of zeros, let us consider the zero-inflated Poisson (ZIP) regression model:

- a proportion π of those people who go to online apparel sites will never visit khakichinos.com
- the visiting behavior of the "ever visitors" can be characterized by the Poisson regression model

$$P(Y_i = y) = \delta_{y=0}\pi + (1-\pi) \times \frac{[\lambda_0 \exp(\boldsymbol{\beta}'\mathbf{x}_i)]^y e^{-\lambda_0 \exp(\boldsymbol{\beta}'\mathbf{x}_i)}}{y!}$$

	A	B	C	D	E	F	G	H	I	J	K
1	\lambda_0	6.6231			LL =	=SUM(K10:K2737)					
2	pi	0.7433									
3	B_inc	-0.0891									
4	B_sex	-0.1327									
5	B_age	0.1141									
6	B_size	0.0196									
7				=TRANSPOSE(B3:B6)	=TRANSPOSE(B3:B6)	=TRANSPOSE(B3:B6)	=TRANSPOSE(B3:B6)				
8											
9	ID	Total		Income	Sex	Age	Size		lambda	P(Y=y)	ln[P(Y=y)]
10	1	0		11.3793940723457	1	3.87120101090789	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D10:G10))	=IF(B10=0,B$2,0)+(1-B$2)*POISSON(B10,I10,FALSE)	=LN(J10)
11	2	5		9.76995615991161	1	4.04305126783455	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D11:G11))	=IF(B11=0,B$2,0)+(1-B$2)*POISSON(B11,I11,FALSE)	=LN(J11)
12	3	0		11.0821425488778	0	3.3322045101752	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D12:G12))	=IF(B12=0,B$2,0)+(1-B$2)*POISSON(B12,I12,FALSE)	=LN(J12)
13	4	0		10.9150884642146	1	3.95124371858143	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D13:G13))	=IF(B13=0,B$2,0)+(1-B$2)*POISSON(B13,I13,FALSE)	=LN(J13)
14	5	0		10.9150884642146	1	2.83321334405622	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D14:G14))	=IF(B14=0,B$2,0)+(1-B$2)*POISSON(B14,I14,FALSE)	=LN(J14)
15	6	0		10.9150884642146	0	2.94443897916644	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D15:G15))	=IF(B15=0,B$2,0)+(1-B$2)*POISSON(B15,I15,FALSE)	=LN(J15)
16	7	0		11.1913418408428	0	3.66356164612965	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D16:G16))	=IF(B16=0,B$2,0)+(1-B$2)*POISSON(B16,I16,FALSE)	=LN(J16)
17	8	1		11.7360690162844	0	4.07753744390572	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D17:G17))	=IF(B17=0,B$2,0)+(1-B$2)*POISSON(B17,I17,FALSE)	=LN(J17)
18	9	0		10.0212705881925	0	4.24849524204936	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D18:G18))	=IF(B18=0,B$2,0)+(1-B$2)*POISSON(B18,I18,FALSE)	=LN(J18)
19	10	0		10.9150884642146	0	3.85014760171006	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D19:G19))	=IF(B19=0,B$2,0)+(1-B$2)*POISSON(B19,I19,FALSE)	=LN(J19)
20	11	1		10.7684849900227	0	3.93182563272433	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D20:G20))	=IF(B20=0,B$2,0)+(1-B$2)*POISSON(B20,I20,FALSE)	=LN(J20)
21	12	0		10.9150884642146	0	3.98898404656427	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D21:G21))	=IF(B21=0,B$2,0)+(1-B$2)*POISSON(B21,I21,FALSE)	=LN(J21)
22	13	3		10.5320962119585	0	3.63758615972639	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D22:G22))	=IF(B22=0,B$2,0)+(1-B$2)*POISSON(B22,I22,FALSE)	=LN(J22)
23	14	0		10.9150884642146	0	3.61091791264422	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D23:G23))	=IF(B23=0,B$2,0)+(1-B$2)*POISSON(B23,I23,FALSE)	=LN(J23)
24	15	0		10.2219412836547	1	3.58351893845611	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D24:G24))	=IF(B24=0,B$2,0)+(1-B$2)*POISSON(B24,I24,FALSE)	=LN(J24)
25	16	1		10.7684849900227	1	3.25809653802148	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D25:G25))	=IF(B25=0,B$2,0)+(1-B$2)*POISSON(B25,I25,FALSE)	=LN(J25)
26	17	2		12.2060726455302	0	3.66356164612965	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D26:G26))	=IF(B26=0,B$2,0)+(1-B$2)*POISSON(B26,I26,FALSE)	=LN(J26)
27	18	0		10.7684849900227	0	3.95124371858143	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D27:G27))	=IF(B27=0,B$2,0)+(1-B$2)*POISSON(B27,I27,FALSE)	=LN(J27)
28	19	6		11.1913418408428	1	3.3322045101752	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D28:G28))	=IF(B28=0,B$2,0)+(1-B$2)*POISSON(B28,I28,FALSE)	=LN(J28)
29	20	0		10.3889953683178	1	3.58351893845611	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D29:G29))	=IF(B29=0,B$2,0)+(1-B$2)*POISSON(B29,I29,FALSE)	=LN(J29)
30	21	2		10.7684849900227	1	3.3322045101752	4		=B$1*EXP(SUMPRODUCT(D$7:G$7,D30:G30))	=IF(B30=0,B$2,0)+(1-B$2)*POISSON(B30,I30,FALSE)	=LN(J30)
31	22	0		11.1913418408428	1	3.46573590279973	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D31:G31))	=IF(B31=0,B$2,0)+(1-B$2)*POISSON(B31,I31,FALSE)	=LN(J31)
32	23	0		11.1913418408428	1	3.43398720448515	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D32:G32))	=IF(B32=0,B$2,0)+(1-B$2)*POISSON(B32,I32,FALSE)	=LN(J32)
33	24	2		11.7360690162844	1	3.80666248977032	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D33:G33))	=IF(B33=0,B$2,0)+(1-B$2)*POISSON(B33,I33,FALSE)	=LN(J33)
34	25	0		11.3793940723457	0	4.27666611901606	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D34:G34))	=IF(B34=0,B$2,0)+(1-B$2)*POISSON(B34,I34,FALSE)	=LN(J34)
35	26	0		10.3889953683178	0	4.21950770517611	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D35:G35))	=IF(B35=0,B$2,0)+(1-B$2)*POISSON(B35,I35,FALSE)	=LN(J35)
36	27	0		10.6572593549125	1	3.49650756146648	4		=B$1*EXP(SUMPRODUCT(D$7:G$7,D36:G36))	=IF(B36=0,B$2,0)+(1-B$2)*POISSON(B36,I36,FALSE)	=LN(J36)
37	28	0		12.0725412529057	0	3.95124371858143	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D37:G37))	=IF(B37=0,B$2,0)+(1-B$2)*POISSON(B37,I37,FALSE)	=LN(J37)
38	29	0		10.9150884642146	1	3.80666248977032	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D38:G38))	=IF(B38=0,B$2,0)+(1-B$2)*POISSON(B38,I38,FALSE)	=LN(J38)
39	30	0		10.9150884642146	0	3.52636052461616	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D39:G39))	=IF(B39=0,B$2,0)+(1-B$2)*POISSON(B39,I39,FALSE)	=LN(J39)
40	31	0		11.1913418408428	1	3.36729582998647	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D40:G40))	=IF(B40=0,B$2,0)+(1-B$2)*POISSON(B40,I40,FALSE)	=LN(J40)
41	32	0		10.2219412836547	1	3.13549421592915	4		=B$1*EXP(SUMPRODUCT(D$7:G$7,D41:G41))	=IF(B41=0,B$2,0)+(1-B$2)*POISSON(B41,I41,FALSE)	=LN(J41)
42	33	0		11.3793940723457	0	3.3322045101752	4		=B$1*EXP(SUMPRODUCT(D$7:G$7,D42:G42))	=IF(B42=0,B$2,0)+(1-B$2)*POISSON(B42,I42,FALSE)	=LN(J42)
43	34	0		9.07680897935166	1	3.40119738166216	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D43:G43))	=IF(B43=0,B$2,0)+(1-B$2)*POISSON(B43,I43,FALSE)	=LN(J43)
44	35	0		10.0212705881925	1	3.52636052461616	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D44:G44))	=IF(B44=0,B$2,0)+(1-B$2)*POISSON(B44,I44,FALSE)	=LN(J44)
45	36	0		11.0821425488778	0	4.06044301054642	4		=B$1*EXP(SUMPRODUCT(D$7:G$7,D45:G45))	=IF(B45=0,B$2,0)+(1-B$2)*POISSON(B45,I45,FALSE)	=LN(J45)
46	37	2		10.2219412836547	1	3.68887945411394	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D46:G46))	=IF(B46=0,B$2,0)+(1-B$2)*POISSON(B46,I46,FALSE)	=LN(J46)
47	38	2		12.0725412529057	1	3.68887945411394	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D47:G47))	=IF(B47=0,B$2,0)+(1-B$2)*POISSON(B47,I47,FALSE)	=LN(J47)
48	39	1		11.0821425488778	0	4.17438726989564	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D48:G48))	=IF(B48=0,B$2,0)+(1-B$2)*POISSON(B48,I48,FALSE)	=LN(J48)
49	40	0		9.52879410309472	1	2.70805020110221	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D49:G49))	=IF(B49=0,B$2,0)+(1-B$2)*POISSON(B49,I49,FALSE)	=LN(J49)
50	41	0		11.0821425488778	1	3.80666248977032	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D50:G50))	=IF(B50=0,B$2,0)+(1-B$2)*POISSON(B50,I50,FALSE)	=LN(J50)
51	42	0		11.3793940723457	1	4.12713438504509	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D51:G51))	=IF(B51=0,B$2,0)+(1-B$2)*POISSON(B51,I51,FALSE)	=LN(J51)
52	43	0		11.3793940723457	0	4.17438726989564	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D52:G52))	=IF(B52=0,B$2,0)+(1-B$2)*POISSON(B52,I52,FALSE)	=LN(J52)
53	44	0		10.5320962119585	1	3.55534806148941	6		=B$1*EXP(SUMPRODUCT(D$7:G$7,D53:G53))	=IF(B53=0,B$2,0)+(1-B$2)*POISSON(B53,I53,FALSE)	=LN(J53)
54	45	0		10.7684849900227	0	3.2188758248682	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D54:G54))	=IF(B54=0,B$2,0)+(1-B$2)*POISSON(B54,I54,FALSE)	=LN(J54)
55	46	0		11.3793940723457	1	3.36729582998647	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D55:G55))	=IF(B55=0,B$2,0)+(1-B$2)*POISSON(B55,I55,FALSE)	=LN(J55)
56	47	0		11.7360690162844	0	3.04452243772342	4		=B$1*EXP(SUMPRODUCT(D$7:G$7,D56:G56))	=IF(B56=0,B$2,0)+(1-B$2)*POISSON(B56,I56,FALSE)	=LN(J56)
57	48	0		10.7684849900227	0	3.52636052461616	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D57:G57))	=IF(B57=0,B$2,0)+(1-B$2)*POISSON(B57,I57,FALSE)	=LN(J57)
58	49	0		10.3889953683178	1	2.83321334405622	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D58:G58))	=IF(B58=0,B$2,0)+(1-B$2)*POISSON(B58,I58,FALSE)	=LN(J58)
59	50	0		10.3889953683178	1	2.63905732961526	3		=B$1*EXP(SUMPRODUCT(D$7:G$7,D59:G59))	=IF(B59=0,B$2,0)+(1-B$2)*POISSON(B59,I59,FALSE)	=LN(J59)
60	51	0		11.0821425488778	0	3.73766961828337	5		=B$1*EXP(SUMPRODUCT(D$7:G$7,D60:G60))	=IF(B60=0,B$2,0)+(1-B$2)*POISSON(B60,I60,FALSE)	=LN(J60)
61	52	6		9.76995615991161	0	3.29583686600433	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D61:G61))	=IF(B61=0,B$2,0)+(1-B$2)*POISSON(B61,I61,FALSE)	=LN(J61)
62	53	16		9.76995615991161	0	3.13549421592915	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D62:G62))	=IF(B62=0,B$2,0)+(1-B$2)*POISSON(B62,I62,FALSE)	=LN(J62)
63	54	1		10.2219412836547	1	3.46573590279973	1		=B$1*EXP(SUMPRODUCT(D$7:G$7,D63:G63))	=IF(B63=0,B$2,0)+(1-B$2)*POISSON(B63,I63,FALSE)	=LN(J63)
64	55	0		11.3793940723457	1	3.55534806148941	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D64:G64))	=IF(B64=0,B$2,0)+(1-B$2)*POISSON(B64,I64,FALSE)	=LN(J64)
65	56	0		11.7360690162844	0	3.93182563272433	4		=B$1*EXP(SUMPRODUCT(D$7:G$7,D65:G65))	=IF(B65=0,B$2,0)+(1-B$2)*POISSON(B65,I65,FALSE)	=LN(J65)
66	57	0		10.3889953683178	1	3.73766961828337	2		=B$1*EXP(SUMPRODUCT(D$7:G$7,D66:G66))	=IF(B66=0,B$2,0)+(1-B$2)*POISSON(B66,I66,FALSE)	=LN(J66)

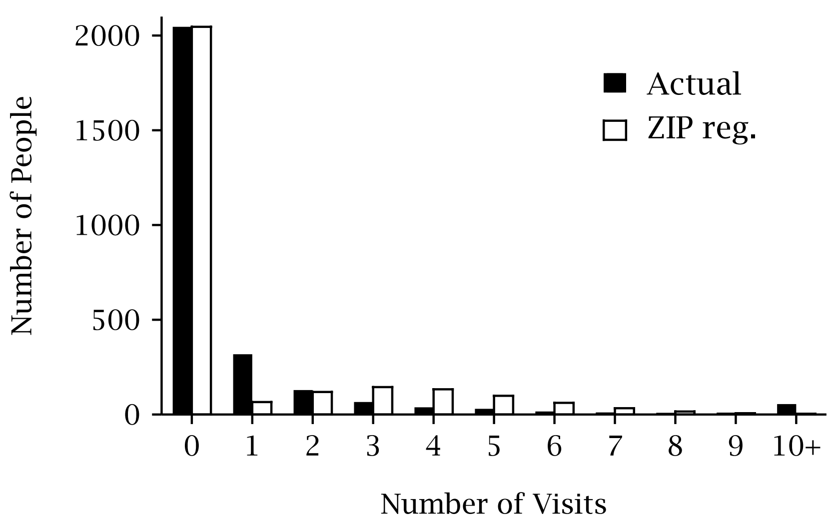
Problem 5 -- ZIP reg

	A	B	C	D	E	F	G	H	I	J
1	\lambda_0	6.6231			LL =	-4297.472				
2	pi	0.7433								
3	B_inc	-0.0891								
4	B_sex	-0.1327								
5	B_age	0.1141								
6	B_size	0.0196								
7				-0.0891	-0.1327	0.1141	0.0196			
8										
9	ID	Total		Income	Sex	Age	HH Size		lambda	P(Y=y)
10	1	0		11.38	1	3.87	2		3.40193	0.75184
11	2	5		9.77	1	4.04	1		3.92698	0.03936
12	3	0		11.08	0	3.33	2		3.75094	0.74932
13	4	0		10.92	1	3.95	3		3.64889	0.74996
14	5	0		10.92	1	2.83	3		3.21182	0.75363
15	6	0		10.92	0	2.94	3		3.71435	0.74954
16	7	0		11.19	0	3.66	2		3.85775	0.74871
17	8	1		11.74	0	4.08	2		3.85266	0.02099
18	9	0		10.02	0	4.25	1		4.48880	0.74617
19	10	0		10.92	0	3.85	3		4.11879	0.74746
20	11	1		10.77	0	3.93	2		4.13048	0.01705
21	12	0		10.92	0	3.99	2		4.10353	0.74752
22	13	3		10.53	0	3.64	2		4.07915	0.04914
23	14	0		10.92	0	3.61	1		3.85413	0.74872
24	15	0		10.22	1	3.58	3		3.72197	0.74949
25	16	1		10.77	1	3.26	3		3.41574	0.02881
26	17	2		12.21	0	3.66	2		3.52410	0.04699
27	18	0		10.77	0	3.95	2		4.13964	0.74737
28	19	6		11.19	1	3.33	2		3.25307	0.01633
29	20	0		10.39	1	3.58	2		3.59593	0.75033
30	21	2		10.77	1	3.33	4		3.51278	0.04722
31	22	0		11.19	1	3.47	2		3.30302	0.75272
32	23	0		11.19	1	3.43	2		3.29107	0.75284
33	24	2		11.74	1	3.81	2		3.27128	0.05214
34	25	0		11.38	0	4.28	2		4.06854	0.74767
35	26	0		10.39	0	4.22	2		4.41520	0.74639
36	27	0		10.66	1	3.50	4		3.61493	0.75019
37	28	0		12.07	0	3.95	2		3.68532	0.74973
38	29	0		10.92	1	3.81	3		3.58919	0.75038
39	30	0		10.92	0	3.53	3		3.96938	0.74813
40	31	0		11.19	1	3.37	2		3.26612	0.75308
41	32	0		10.22	1	3.14	4		3.60630	0.75025
42	33	0		11.38	0	3.33	4		3.79855	0.74904
43	34	0		9.08	1	3.40	1		3.88226	0.74857
44	35	0		10.02	1	3.53	1		3.62011	0.75016
45	36	0		11.08	0	4.06	4		4.23856	0.74699
46	37	2		10.22	1	3.69	2		3.69403	0.04356
47	38	2		12.07	1	3.69	2		3.13223	0.05493
48	39	1		11.08	0	4.17	1		4.04934	0.01812
49	40	0		9.53	1	2.71	3		3.58278	0.75042
50	41	0		11.08	1	3.81	3		3.53613	0.75076
51	42	0		11.38	1	4.13	3		3.57193	0.75050

ZIP Regression Results

Variable	Coefficient
λ_0	6.6231
Income	−0.0891
Sex	−0.1327
Age	0.1141
HH Size	0.0196
π	0.7433
LL	−4297.5
$LL_{\text{Poiss reg}}$	−6291.5
LR (df = 1)	3988.0

Fit of ZIP Regression



NBD Regression

The explanatory variables may not fully capture the differences among individuals

To capture the remaining (unobserved) component of differences among individuals, let λ_0 vary across the population according to a gamma distribution with parameters r and α:

$$P(Y_i = y) = \frac{\Gamma(r + y)}{\Gamma(r)y!} \left(\frac{\alpha}{\alpha + \exp(\boldsymbol{\beta}'\mathbf{x}_i)} \right)^r \left(\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{\alpha + \exp(\boldsymbol{\beta}'\mathbf{x}_i)} \right)^y$$

- Known as the “Negbin II” model in most textbooks
- Collapses to the NBD when $\boldsymbol{\beta} = \mathbf{0}$

NBD Regression Results

Variable	Coefficient
r	0.1388
α	8.1979
Income	0.0734
Sex	−0.0093
Age	0.9022
HH Size	−0.0243
LL	−2889.0

Problem 5 -- NBD reg

	A	B	C	D	E	F	G	H	I	J	K
1	r	0.1388			LL =	=SUM(K10:K2737)					
2	alpha	8.1979									
3	B_inc	0.0734									
4	B_sex	-0.0093									
5	B_age	0.9022									
6	B_size	-0.0243									
7				=TRANSPOSE(B3:B6)	=TRANSPOSE(B3:B6)	=TRANSPOSE(B3:B6)	=TRANSPOSE(B3:B6)				
8											
9	ID	Total		Income	Sex	Age	Size		exp(BX)	P(Y=y)	ln[P(Y=y)]
10	1	0		11.3793940723457	1	3.87120101090789	2		=EXP(SUMPRODUCT(D$7:G$7,D10:G10))	=EXP(GAMMALN(B$1+B10)-GAMMALN(B$1))/FACT(B10)*(B$2/(B$2+I10))^B$1*(I10/(B$2+I10))^B10	=LN(J10)
11	2	5		9.76995615991161	1	4.04305126783455	1		=EXP(SUMPRODUCT(D$7:G$7,D11:G11))	=EXP(GAMMALN(B$1+B11)-GAMMALN(B$1))/FACT(B11)*(B$2/(B$2+I11))^B$1*(I11/(B$2+I11))^B11	=LN(J11)
12	3	0		11.0821425488778	0	3.3322045101752	2		=EXP(SUMPRODUCT(D$7:G$7,D12:G12))	=EXP(GAMMALN(B$1+B12)-GAMMALN(B$1))/FACT(B12)*(B$2/(B$2+I12))^B$1*(I12/(B$2+I12))^B12	=LN(J12)
13	4	0		10.9150884642146	1	3.95124371858143	3		=EXP(SUMPRODUCT(D$7:G$7,D13:G13))	=EXP(GAMMALN(B$1+B13)-GAMMALN(B$1))/FACT(B13)*(B$2/(B$2+I13))^B$1*(I13/(B$2+I13))^B13	=LN(J13)
14	5	0		10.9150884642146	1	2.83321334405622	3		=EXP(SUMPRODUCT(D$7:G$7,D14:G14))	=EXP(GAMMALN(B$1+B14)-GAMMALN(B$1))/FACT(B14)*(B$2/(B$2+I14))^B$1*(I14/(B$2+I14))^B14	=LN(J14)
15	6	0		10.9150884642146	0	2.94443897916644	3		=EXP(SUMPRODUCT(D$7:G$7,D15:G15))	=EXP(GAMMALN(B$1+B15)-GAMMALN(B$1))/FACT(B15)*(B$2/(B$2+I15))^B$1*(I15/(B$2+I15))^B15	=LN(J15)
16	7	0		11.1913418408428	0	3.66356164612965	2		=EXP(SUMPRODUCT(D$7:G$7,D16:G16))	=EXP(GAMMALN(B$1+B16)-GAMMALN(B$1))/FACT(B16)*(B$2/(B$2+I16))^B$1*(I16/(B$2+I16))^B16	=LN(J16)
17	8	1		11.7360690162844	0	4.07753744390572	2		=EXP(SUMPRODUCT(D$7:G$7,D17:G17))	=EXP(GAMMALN(B$1+B17)-GAMMALN(B$1))/FACT(B17)*(B$2/(B$2+I17))^B$1*(I17/(B$2+I17))^B17	=LN(J17)
18	9	0		10.0212705881925	0	4.24849524204936	1		=EXP(SUMPRODUCT(D$7:G$7,D18:G18))	=EXP(GAMMALN(B$1+B18)-GAMMALN(B$1))/FACT(B18)*(B$2/(B$2+I18))^B$1*(I18/(B$2+I18))^B18	=LN(J18)
19	10	0		10.9150884642146	0	3.85014760171006	3		=EXP(SUMPRODUCT(D$7:G$7,D19:G19))	=EXP(GAMMALN(B$1+B19)-GAMMALN(B$1))/FACT(B19)*(B$2/(B$2+I19))^B$1*(I19/(B$2+I19))^B19	=LN(J19)
20	11	1		10.7684849900227	0	3.93182563272433	2		=EXP(SUMPRODUCT(D$7:G$7,D20:G20))	=EXP(GAMMALN(B$1+B20)-GAMMALN(B$1))/FACT(B20)*(B$2/(B$2+I20))^B$1*(I20/(B$2+I20))^B20	=LN(J20)
21	12	0		10.9150884642146	0	3.98898404656427	2		=EXP(SUMPRODUCT(D$7:G$7,D21:G21))	=EXP(GAMMALN(B$1+B21)-GAMMALN(B$1))/FACT(B21)*(B$2/(B$2+I21))^B$1*(I21/(B$2+I21))^B21	=LN(J21)
22	13	3		10.5320962119585	0	3.63758615972639	2		=EXP(SUMPRODUCT(D$7:G$7,D22:G22))	=EXP(GAMMALN(B$1+B22)-GAMMALN(B$1))/FACT(B22)*(B$2/(B$2+I22))^B$1*(I22/(B$2+I22))^B22	=LN(J22)
23	14	0		10.9150884642146	0	3.61091791264422	1		=EXP(SUMPRODUCT(D$7:G$7,D23:G23))	=EXP(GAMMALN(B$1+B23)-GAMMALN(B$1))/FACT(B23)*(B$2/(B$2+I23))^B$1*(I23/(B$2+I23))^B23	=LN(J23)
24	15	0		10.2219412836547	1	3.58351893845611	3		=EXP(SUMPRODUCT(D$7:G$7,D24:G24))	=EXP(GAMMALN(B$1+B24)-GAMMALN(B$1))/FACT(B24)*(B$2/(B$2+I24))^B$1*(I24/(B$2+I24))^B24	=LN(J24)
25	16	1		10.7684849900227	1	3.25809653802148	3		=EXP(SUMPRODUCT(D$7:G$7,D25:G25))	=EXP(GAMMALN(B$1+B25)-GAMMALN(B$1))/FACT(B25)*(B$2/(B$2+I25))^B$1*(I25/(B$2+I25))^B25	=LN(J25)
26	17	2		12.2060726455302	0	3.66356164612965	2		=EXP(SUMPRODUCT(D$7:G$7,D26:G26))	=EXP(GAMMALN(B$1+B26)-GAMMALN(B$1))/FACT(B26)*(B$2/(B$2+I26))^B$1*(I26/(B$2+I26))^B26	=LN(J26)
27	18	0		10.7684849900227	0	3.95124371858143	2		=EXP(SUMPRODUCT(D$7:G$7,D27:G27))	=EXP(GAMMALN(B$1+B27)-GAMMALN(B$1))/FACT(B27)*(B$2/(B$2+I27))^B$1*(I27/(B$2+I27))^B27	=LN(J27)
28	19	6		11.1913418408428	1	3.3322045101752	2		=EXP(SUMPRODUCT(D$7:G$7,D28:G28))	=EXP(GAMMALN(B$1+B28)-GAMMALN(B$1))/FACT(B28)*(B$2/(B$2+I28))^B$1*(I28/(B$2+I28))^B28	=LN(J28)
29	20	0		10.3889953683178	1	3.58351893845611	2		=EXP(SUMPRODUCT(D$7:G$7,D29:G29))	=EXP(GAMMALN(B$1+B29)-GAMMALN(B$1))/FACT(B29)*(B$2/(B$2+I29))^B$1*(I29/(B$2+I29))^B29	=LN(J29)
30	21	2		10.7684849900227	1	3.3322045101752	4		=EXP(SUMPRODUCT(D$7:G$7,D30:G30))	=EXP(GAMMALN(B$1+B30)-GAMMALN(B$1))/FACT(B30)*(B$2/(B$2+I30))^B$1*(I30/(B$2+I30))^B30	=LN(J30)
31	22	0		11.1913418408428	1	3.46573590279973	2		=EXP(SUMPRODUCT(D$7:G$7,D31:G31))	=EXP(GAMMALN(B$1+B31)-GAMMALN(B$1))/FACT(B31)*(B$2/(B$2+I31))^B$1*(I31/(B$2+I31))^B31	=LN(J31)
32	23	0		11.1913418408428	1	3.43398720448515	2		=EXP(SUMPRODUCT(D$7:G$7,D32:G32))	=EXP(GAMMALN(B$1+B32)-GAMMALN(B$1))/FACT(B32)*(B$2/(B$2+I32))^B$1*(I32/(B$2+I32))^B32	=LN(J32)
33	24	2		11.7360690162844	1	3.80666248977032	2		=EXP(SUMPRODUCT(D$7:G$7,D33:G33))	=EXP(GAMMALN(B$1+B33)-GAMMALN(B$1))/FACT(B33)*(B$2/(B$2+I33))^B$1*(I33/(B$2+I33))^B33	=LN(J33)
34	25	0		11.3793940723457	0	4.27666611901606	2		=EXP(SUMPRODUCT(D$7:G$7,D34:G34))	=EXP(GAMMALN(B$1+B34)-GAMMALN(B$1))/FACT(B34)*(B$2/(B$2+I34))^B$1*(I34/(B$2+I34))^B34	=LN(J34)
35	26	0		10.3889953683178	0	4.21950770517611	2		=EXP(SUMPRODUCT(D$7:G$7,D35:G35))	=EXP(GAMMALN(B$1+B35)-GAMMALN(B$1))/FACT(B35)*(B$2/(B$2+I35))^B$1*(I35/(B$2+I35))^B35	=LN(J35)
36	27	0		10.6572593549125	1	3.49650756146648	4		=EXP(SUMPRODUCT(D$7:G$7,D36:G36))	=EXP(GAMMALN(B$1+B36)-GAMMALN(B$1))/FACT(B36)*(B$2/(B$2+I36))^B$1*(I36/(B$2+I36))^B36	=LN(J36)
37	28	0		12.0725412529057	0	3.95124371858143	2		=EXP(SUMPRODUCT(D$7:G$7,D37:G37))	=EXP(GAMMALN(B$1+B37)-GAMMALN(B$1))/FACT(B37)*(B$2/(B$2+I37))^B$1*(I37/(B$2+I37))^B37	=LN(J37)
38	29	0		10.9150884642146	1	3.80666248977032	3		=EXP(SUMPRODUCT(D$7:G$7,D38:G38))	=EXP(GAMMALN(B$1+B38)-GAMMALN(B$1))/FACT(B38)*(B$2/(B$2+I38))^B$1*(I38/(B$2+I38))^B38	=LN(J38)
39	30	0		10.9150884642146	0	3.52636052461616	3		=EXP(SUMPRODUCT(D$7:G$7,D39:G39))	=EXP(GAMMALN(B$1+B39)-GAMMALN(B$1))/FACT(B39)*(B$2/(B$2+I39))^B$1*(I39/(B$2+I39))^B39	=LN(J39)
40	31	0		11.1913418408428	1	3.36729582998647	2		=EXP(SUMPRODUCT(D$7:G$7,D40:G40))	=EXP(GAMMALN(B$1+B40)-GAMMALN(B$1))/FACT(B40)*(B$2/(B$2+I40))^B$1*(I40/(B$2+I40))^B40	=LN(J40)
41	32	0		10.2219412836547	1	3.13549421592915	4		=EXP(SUMPRODUCT(D$7:G$7,D41:G41))	=EXP(GAMMALN(B$1+B41)-GAMMALN(B$1))/FACT(B41)*(B$2/(B$2+I41))^B$1*(I41/(B$2+I41))^B41	=LN(J41)
42	33	0		11.3793940723457	0	3.3322045101752	4		=EXP(SUMPRODUCT(D$7:G$7,D42:G42))	=EXP(GAMMALN(B$1+B42)-GAMMALN(B$1))/FACT(B42)*(B$2/(B$2+I42))^B$1*(I42/(B$2+I42))^B42	=LN(J42)
43	34	0		9.07680897935166	1	3.40119738166216	1		=EXP(SUMPRODUCT(D$7:G$7,D43:G43))	=EXP(GAMMALN(B$1+B43)-GAMMALN(B$1))/FACT(B43)*(B$2/(B$2+I43))^B$1*(I43/(B$2+I43))^B43	=LN(J43)
44	35	0		10.0212705881925	1	3.52636052461616	1		=EXP(SUMPRODUCT(D$7:G$7,D44:G44))	=EXP(GAMMALN(B$1+B44)-GAMMALN(B$1))/FACT(B44)*(B$2/(B$2+I44))^B$1*(I44/(B$2+I44))^B44	=LN(J44)
45	36	0		11.0821425488778	0	4.06044301054642	4		=EXP(SUMPRODUCT(D$7:G$7,D45:G45))	=EXP(GAMMALN(B$1+B45)-GAMMALN(B$1))/FACT(B45)*(B$2/(B$2+I45))^B$1*(I45/(B$2+I45))^B45	=LN(J45)
46	37	2		10.2219412836547	1	3.68887945411394	2		=EXP(SUMPRODUCT(D$7:G$7,D46:G46))	=EXP(GAMMALN(B$1+B46)-GAMMALN(B$1))/FACT(B46)*(B$2/(B$2+I46))^B$1*(I46/(B$2+I46))^B46	=LN(J46)
47	38	2		12.0725412529057	1	3.68887945411394	2		=EXP(SUMPRODUCT(D$7:G$7,D47:G47))	=EXP(GAMMALN(B$1+B47)-GAMMALN(B$1))/FACT(B47)*(B$2/(B$2+I47))^B$1*(I47/(B$2+I47))^B47	=LN(J47)
48	39	1		11.0821425488778	0	4.17438726989564	1		=EXP(SUMPRODUCT(D$7:G$7,D48:G48))	=EXP(GAMMALN(B$1+B48)-GAMMALN(B$1))/FACT(B48)*(B$2/(B$2+I48))^B$1*(I48/(B$2+I48))^B48	=LN(J48)
49	40	0		9.52879410309472	1	2.70805020110221	3		=EXP(SUMPRODUCT(D$7:G$7,D49:G49))	=EXP(GAMMALN(B$1+B49)-GAMMALN(B$1))/FACT(B49)*(B$2/(B$2+I49))^B$1*(I49/(B$2+I49))^B49	=LN(J49)
50	41	0		11.0821425488778	1	3.80666248977032	3		=EXP(SUMPRODUCT(D$7:G$7,D50:G50))	=EXP(GAMMALN(B$1+B50)-GAMMALN(B$1))/FACT(B50)*(B$2/(B$2+I50))^B$1*(I50/(B$2+I50))^B50	=LN(J50)
51	42	0		11.3793940723457	1	4.12713438504509	3		=EXP(SUMPRODUCT(D$7:G$7,D51:G51))	=EXP(GAMMALN(B$1+B51)-GAMMALN(B$1))/FACT(B51)*(B$2/(B$2+I51))^B$1*(I51/(B$2+I51))^B51	=LN(J51)
52	43	0		11.3793940723457	0	4.17438726989564	3		=EXP(SUMPRODUCT(D$7:G$7,D52:G52))	=EXP(GAMMALN(B$1+B52)-GAMMALN(B$1))/FACT(B52)*(B$2/(B$2+I52))^B$1*(I52/(B$2+I52))^B52	=LN(J52)
53	44	0		10.5320962119585	1	3.55534806148941	6		=EXP(SUMPRODUCT(D$7:G$7,D53:G53))	=EXP(GAMMALN(B$1+B53)-GAMMALN(B$1))/FACT(B53)*(B$2/(B$2+I53))^B$1*(I53/(B$2+I53))^B53	=LN(J53)
54	45	0		10.7684849900227	0	3.2188758248682	1		=EXP(SUMPRODUCT(D$7:G$7,D54:G54))	=EXP(GAMMALN(B$1+B54)-GAMMALN(B$1))/FACT(B54)*(B$2/(B$2+I54))^B$1*(I54/(B$2+I54))^B54	=LN(J54)
55	46	0		11.3793940723457	1	3.36729582998647	2		=EXP(SUMPRODUCT(D$7:G$7,D55:G55))	=EXP(GAMMALN(B$1+B55)-GAMMALN(B$1))/FACT(B55)*(B$2/(B$2+I55))^B$1*(I55/(B$2+I55))^B55	=LN(J55)
56	47	0		11.7360690162844	0	3.04452243772342	4		=EXP(SUMPRODUCT(D$7:G$7,D56:G56))	=EXP(GAMMALN(B$1+B56)-GAMMALN(B$1))/FACT(B56)*(B$2/(B$2+I56))^B$1*(I56/(B$2+I56))^B56	=LN(J56)
57	48	0		10.7684849900227	0	3.52636052461616	1		=EXP(SUMPRODUCT(D$7:G$7,D57:G57))	=EXP(GAMMALN(B$1+B57)-GAMMALN(B$1))/FACT(B57)*(B$2/(B$2+I57))^B$1*(I57/(B$2+I57))^B57	=LN(J57)
58	49	0		10.3889953683178	1	2.83321334405622	3		=EXP(SUMPRODUCT(D$7:G$7,D58:G58))	=EXP(GAMMALN(B$1+B58)-GAMMALN(B$1))/FACT(B58)*(B$2/(B$2+I58))^B$1*(I58/(B$2+I58))^B58	=LN(J58)
59	50	0		10.3889953683178	1	2.63905732961526	3		=EXP(SUMPRODUCT(D$7:G$7,D59:G59))	=EXP(GAMMALN(B$1+B59)-GAMMALN(B$1))/FACT(B59)*(B$2/(B$2+I59))^B$1*(I59/(B$2+I59))^B59	=LN(J59)
60	51	0		11.0821425488778	0	3.73766961828337	5		=EXP(SUMPRODUCT(D$7:G$7,D60:G60))	=EXP(GAMMALN(B$1+B60)-GAMMALN(B$1))/FACT(B60)*(B$2/(B$2+I60))^B$1*(I60/(B$2+I60))^B60	=LN(J60)
61	52	6		9.76995615991161	0	3.29583686600433	1		=EXP(SUMPRODUCT(D$7:G$7,D61:G61))	=EXP(GAMMALN(B$1+B61)-GAMMALN(B$1))/FACT(B61)*(B$2/(B$2+I61))^B$1*(I61/(B$2+I61))^B61	=LN(J61)
62	53	16		9.76995615991161	0	3.13549421592915	1		=EXP(SUMPRODUCT(D$7:G$7,D62:G62))	=EXP(GAMMALN(B$1+B62)-GAMMALN(B$1))/FACT(B62)*(B$2/(B$2+I62))^B$1*(I62/(B$2+I62))^B62	=LN(J62)
63	54	1		10.2219412836547	1	3.46573590279973	1		=EXP(SUMPRODUCT(D$7:G$7,D63:G63))	=EXP(GAMMALN(B$1+B63)-GAMMALN(B$1))/FACT(B63)*(B$2/(B$2+I63))^B$1*(I63/(B$2+I63))^B63	=LN(J63)
64	55	0		11.3793940723457	1	3.55534806148941	2		=EXP(SUMPRODUCT(D$7:G$7,D64:G64))	=EXP(GAMMALN(B$1+B64)-GAMMALN(B$1))/FACT(B64)*(B$2/(B$2+I64))^B$1*(I64/(B$2+I64))^B64	=LN(J64)
65	56	0		11.7360690162844	0	3.93182563272433	4		=EXP(SUMPRODUCT(D$7:G$7,D65:G65))	=EXP(GAMMALN(B$1+B65)-GAMMALN(B$1))/FACT(B65)*(B$2/(B$2+I65))^B$1*(I65/(B$2+I65))^B65	=LN(J65)
66	57	0		10.3889953683178	1	3.73766961828337	2		=EXP(SUMPRODUCT(D$7:G$7,D66:G66))	=EXP(GAMMALN(B$1+B66)-GAMMALN(B$1))/FACT(B66)*(B$2/(B$2+I66))^B$1*(I66/(B$2+I66))^B66	=LN(J66)
67	58	1		11.1913418408428	1	3.49650756146648	2		=EXP(SUMPRODUCT(D$7:G$7,D67:G67))	=EXP(GAMMALN(B$1+B67)-GAMMALN(B$1))/FACT(B67)*(B$2/(B$2+I67))^B$1*(I67/(B$2+I67))^B67	=LN(J67)
68	59	0		10.2219412836547	0	3.61091791264422	4		=EXP(SUMPRODUCT(D$7:G$7,D68:G68))	=EXP(GAMMALN(B$1+B68)-GAMMALN(B$1))/FACT(B68)*(B$2/(B$2+I68))^B$1*(I68/(B$2+I68))^B68	=LN(J68)
69	60	0		10.9150884642146	0	3.89182029811063	1		=EXP(SUMPRODUCT(D$7:G$7,D69:G69))	=EXP(GAMMALN(B$1+B69)-GAMMALN(B$1))/FACT(B69)*(B$2/(B$2+I69))^B$1*(I69/(B$2+I69))^B69	=LN(J69)
70	61	0		10.9150884642146	1	3.46573590279973	3		=EXP(SUMPRODUCT(D$7:G$7,D70:G70))	=EXP(GAMMALN(B$1+B70)-GAMMALN(B$1))/FACT(B70)*(B$2/(B$2+I70))^B$1*(I70/(B$2+I70))^B70	=LN(J70)
71	62	0		10.3889953683178	1	3.36729582998647	2		=EXP(SUMPRODUCT(D$7:G$7,D71:G71))	=EXP(GAMMALN(B$1+B71)-GAMMALN(B$1))/FACT(B71)*(B$2/(B$2+I71))^B$1*(I71/(B$2+I71))^B71	=LN(J71)
72	63	0		10.3889953683178	0	3.46573590279973	2		=EXP(SUMPRODUCT(D$7:G$7,D72:G72))	=EXP(GAMMALN(B$1+B72)-GAMMALN(B$1))/FACT(B72)*(B$2/(B$2+I72))^B$1*(I72/(B$2+I72))^B72	=LN(J72)
73	64	0		10.7684849900227	1	3.78418963391826	2		=EXP(SUMPRODUCT(D$7:G$7,D73:G73))	=EXP(GAMMALN(B$1+B73)-GAMMALN(B$1))/FACT(B73)*(B$2/(B$2+I73))^B$1*(I73/(B$2+I73))^B73	=LN(J73)
74	65	0		9.76995615991161	1	3.71357206670431	2		=EXP(SUMPRODUCT(D$7:G$7,D74:G74))	=EXP(GAMMALN(B$1+B74)-GAMMALN(B$1))/FACT(B74)*(B$2/(B$2+I74))^B$1*(I74/(B$2+I74))^B74	=LN(J74)
75	66	0		10.0212705881925	1	3.17805383034795	2		=EXP(SUMPRODUCT(D$7:G$7,D75:G75))	=EXP(GAMMALN(B$1+B75)-GAMMALN(B$1))/FACT(B75)*(B$2/(B$2+I75))^B$1*(I75/(B$2+I75))^B75	=LN(J75)
76	67	0		11.3793940723457	0	4.00733318523247	2		=EXP(SUMPRODUCT(D$7:G$7,D76:G76))	=EXP(GAMMALN(B$1+B76)-GAMMALN(B$1))/FACT(B76)*(B$2/(B$2+I76))^B$1*(I76/(B$2+I76))^B76	=LN(J76)

Problem 5 -- NBD reg

	A	B	C	D	E	F	G	H	I	J
1	r	0.1388			LL =	-2888.966				
2	alpha	8.1979								
3	B_inc	0.0734								
4	B_sex	-0.0093								
5	B_age	0.9022								
6	B_size	-0.0243								
7				0.0734	-0.0093	0.9022	-0.0243			
8										
9	ID	Total		Income	Sex	Age	HH Size		exp(BX)	P(Y=y)
10	1	0		11.38	1	3.87	2		71.51161	0.72936
11	2	5		9.77	1	4.04	1		76.02589	0.01587
12	3	0		11.08	0	3.33	2		43.42559	0.77467
13	4	0		10.92	1	3.95	3		72.50603	0.72810
14	5	0		10.92	1	2.83	3		26.44384	0.81876
15	6	0		10.92	0	2.94	3		29.50734	0.80919
16	7	0		11.19	0	3.66	2		59.02749	0.74680
17	8	1		11.74	0	4.08	2		89.25195	0.09014
18	9	0		10.02	0	4.25	1		94.07931	0.70456
19	10	0		10.92	0	3.85	3		66.80224	0.73555
20	11	1		10.77	0	3.93	2		72.89216	0.09075
21	12	0		10.92	0	3.99	2		77.57994	0.72197
22	13	3		10.53	0	3.64	2		54.93643	0.02795
23	14	0		10.92	0	3.61	1		56.51751	0.75075
24	15	0		10.22	1	3.58	3		49.45389	0.76289
25	16	1		10.77	1	3.26	3		38.38151	0.08984
26	17	2		12.21	0	3.66	2		63.59217	0.04587
27	18	0		10.77	0	3.95	2		74.18036	0.72603
28	19	6		11.19	1	3.33	2		43.37107	0.00859
29	20	0		10.39	1	3.58	2		51.29650	0.75957
30	21	2		10.77	1	3.33	4		40.04943	0.04257
31	22	0		11.19	1	3.47	2		48.92360	0.76387
32	23	0		11.19	1	3.43	2		47.54218	0.76647
33	24	2		11.74	1	3.81	2		69.25654	0.04625
34	25	0		11.38	0	4.28	2		104.05587	0.69552
35	26	0		10.39	0	4.22	2		91.89630	0.70667
36	27	0		10.66	1	3.50	4		46.07074	0.76932
37	28	0		12.07	0	3.95	2		81.63227	0.71736
38	29	0		10.92	1	3.81	3		63.63946	0.73996
39	30	0		10.92	0	3.53	3		49.88038	0.76211
40	31	0		11.19	1	3.37	2		44.76608	0.77192
41	32	0		10.22	1	3.14	4		32.21815	0.80143
42	33	0		11.38	0	3.33	4		42.27648	0.77710
43	34	0		9.08	1	3.40	1		40.49327	0.78098
44	35	0		10.02	1	3.53	1		48.58828	0.76449
45	36	0		11.08	0	4.06	4		79.79024	0.71943
46	37	2		10.22	1	3.69	2		55.72407	0.04515
47	38	2		12.07	1	3.69	2		63.83215	0.04589
48	39	1		11.08	0	4.17	1		95.12148	0.08988
49	40	0		9.53	1	2.71	3		21.33489	0.83709
50	41	0		11.08	1	3.81	3		64.42466	0.73884
51	42	0		11.38	1	4.13	3		87.92064	0.71066

S-Plus NBD Regression Results

Coefficients:

	Value	Std. Error	t value
(Intercept)	-4.047149702	1.10159557	-3.6738979
Income	0.074549233	0.09555222	0.7801936
Sex	-0.005240835	0.11592793	-0.0452077
Age	0.889862966	0.14072030	6.3236289
HH Size	-0.025094493	0.04187696	-0.5992435

Theta: 0.13878
Std. Err.: 0.00726

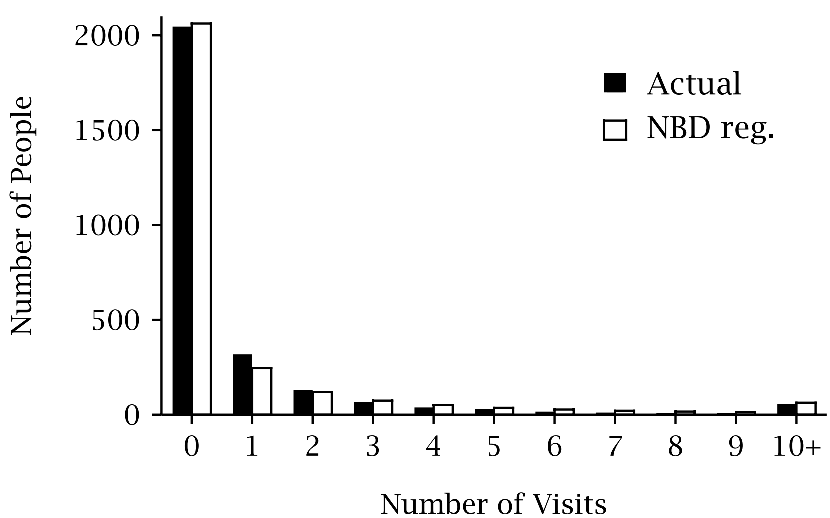
Limdep NBD Regression Results

Variable	Coefficient	Standard Error	b/St.Er.
Constant	-4.077239653	1.0451741	-3.901
INCOME	.7237686001E-01	.76663437E-01	.944
SEX	-.9009160129E-02	.11425700	-.079
AGE	.9045111135	.17741724	5.098
HH SIZE	-.2406546843E-01	.38695426E-01	-.622
	Overdispersion parameter		
Alpha	7.206708844	.33334006	21.620

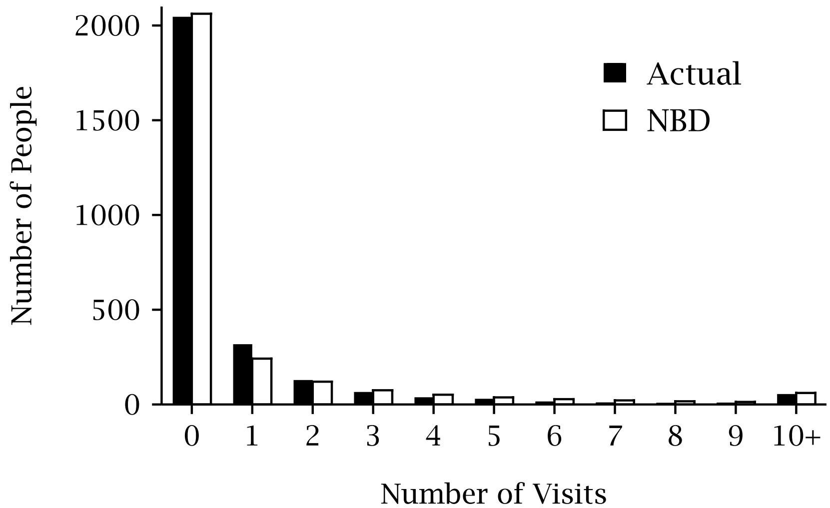
Summary of Regression Results

Variable	Poisson	ZIP	NBD
λ_0	0.0439	6.6231	
r			0.1388
α			8.1979
Income	0.0938	−0.0891	0.0734
Sex	0.0043	−0.1327	−0.0093
Age	0.5882	0.1141	0.9022
HH Size	−0.0359	0.0196	−0.0243
π		0.7433	
LL	−6291.5	−4297.5	−2889.0

Fit of NBD Regression



Fit of NBD



$\hat{r} = 0.134, \hat{\alpha} = 0.141, LL = -2905.6$

Concepts and Tools Introduced

- Incorporating covariate effects in count models
- Poisson (and NBD) regression models
- The value of covariates is frequently over-emphasized

Further Reading

Cameron, A. Colin and Pravin K. Trivedi (1998), *Regression Analysis of Count Data*, Cambridge: Cambridge University Press.

Wedel, Michel and Wagner A. Kamakura (1999), *Market Segmentation: Conceptual and Methodological Foundations*, 2nd edn., Boston, MA: Kluwer Academic Publishers.

Winkelmann, Rainer (2000), *Econometric Analysis of Count Data*, 3rd, revised edn., Berlin: Springer.

Problem 6:
Predicting New Product Trial (Again)

(Extending Basic Models for Timing Data)

Background

Ace Snackfoods, Inc. has developed a new snack product called Krunchy Bits. Before deciding whether or not to “go national” with the new product, the marketing manager for Krunchy Bits has decided to commission a year-long test market using IRI’s BehaviorScan service, with a view to getting a clearer picture of the product’s potential.

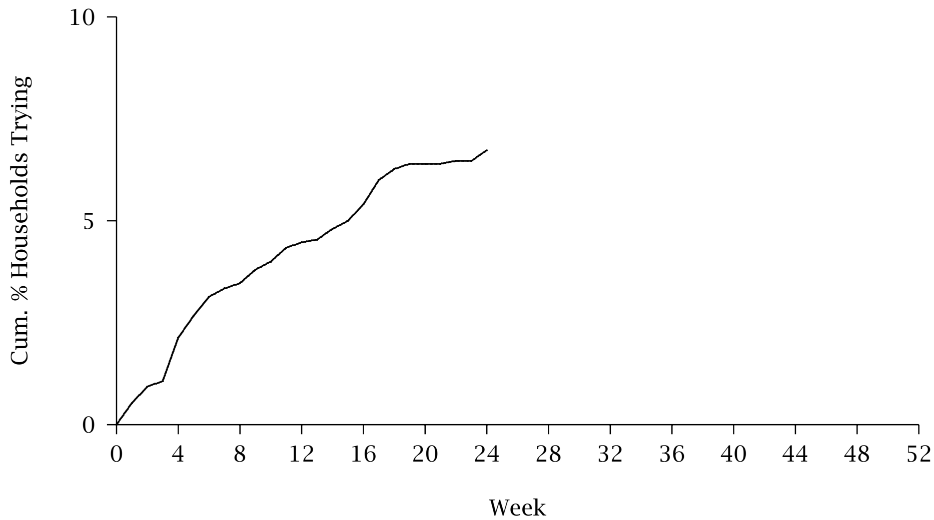
The product has now been under test for 24 weeks. On hand is a dataset documenting the number of households that have made a trial purchase by the end of each week. (The total size of the panel is 1499 households.)

The marketing manager for Krunchy Bits would like a forecast of the product’s year-end performance in the test market. First, she wants a forecast of the percentage of households that will have made a trial purchase by week 52.

Krunchy Bits Cumulative Trial

Week	# Households	Week	# Households
1	8	13	68
2	14	14	72
3	16	15	75
4	32	16	81
5	40	17	90
6	47	18	94
7	50	19	96
8	52	20	96
9	57	21	96
10	60	22	97
11	65	23	97
12	67	24	101

Krunchy Bits Cumulative Trial



Reflection

While the exponential-gamma (EG) fits the data and generates good forecasts, several questions arise:

- Is the exponential distribution the most appropriate model for characterizing the individual-level time-to-trial data?
- How can we incorporate the effects of (time-varying) marketing mix covariates?

Reflecting on the Exponential Assumption

The exponential distribution is often characterized as being "memoryless":

$$\begin{aligned} P(T > s + t | T > t) &= \frac{P(T > s + t, T > t)}{P(T > t)} \\ &= \frac{1 - \left(1 - e^{-\lambda(s+t)}\right)}{1 - \left(1 - e^{-\lambda t}\right)} \\ &= e^{-\lambda(s+t)} / e^{-\lambda t} \\ &= e^{-\lambda s} \\ &= P(T > s) \end{aligned}$$

The probability of "survival" to $s + t$, given survival to t, is the same as the initial probability of survival to s.

Reflecting on the Exponential Assumption

- This means that the probability that the event of interest occurs in the interval $(t, t + \Delta t]$ given that it has not occurred by t,

  $$P(t < T \leq t + \Delta t | T > t) = 1 - e^{-\lambda \Delta t}$$

  is also independent of t
- How can we make $P(t < T \leq t + \Delta t | T > t)$ depend on t?

The Hazard Function

The hazard function, $h(t)$, is defined by

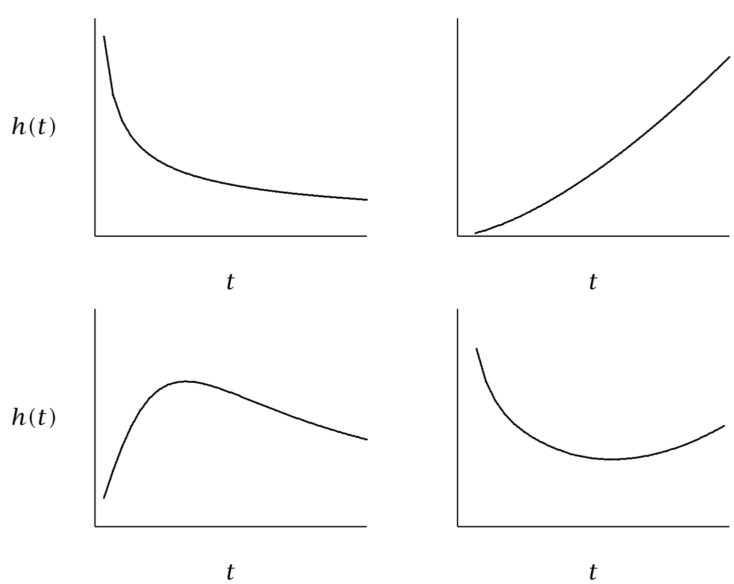
$$
\begin{aligned}
h(t) &= \lim_{\Delta t \to 0} \frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t} \\
&= \frac{f(t)}{1 - F(t)}
\end{aligned}
$$

and represents the instantaneous rate of "failure" at time t conditional upon "survival" to t.

The probability of "failing" in the next small interval of time, given "survival" to time t, is

$$P(t < T \leq t + \Delta t | T > t) \approx h(t) \times \Delta t$$

Shapes of the Hazard Function



The Hazard Function

The hazard function uniquely defines the distribution of a nonnegative random variable:

$$F(t) = 1 - \exp\left(-\int_0^t h(u)\,du\right)$$

Example:

the exponential distribution has a *constant* hazard function, λ

$$\begin{aligned} F(t) &= 1 - \exp\left(-\int_0^t \lambda\,du\right) \\ &= 1 - e^{-\lambda t} \end{aligned}$$

The Weibull Distribution

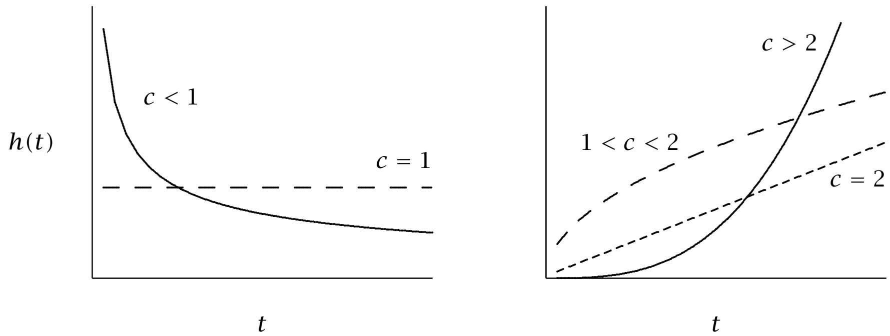
- A generalization of the exponential distribution that can represent decreasing or increasing hazard functions

  $$\begin{aligned} F(t) &= 1 - e^{-\lambda t^c}, \quad \lambda, c > 0 \\ h(t) &= c\lambda t^{c-1} \end{aligned}$$

  where c is the “shape” parameter and λ is the “scale” parameter

- Collapses to the exponential when $c = 1$
- $F(t)$ is S-shaped for $c > 1$

The Weibull Hazard Function



$$h(t) = c\lambda t^{c-1}$$

- Decreasing hazard function (negative duration dependence) when $c < 1$
- Increasing hazard function (positive duration dependence) when $c > 1$

The Weibull-Gamma Model

- Assuming λ is distributed across the population according to a gamma distribution, we have

$$\begin{aligned} P(T \le t) &= \int_0^\infty \left(1 - e^{-\lambda t^c}\right) \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}\, d\lambda \\ &= 1 - \left(\frac{\alpha}{\alpha + t^c}\right)^r \end{aligned}$$

- This collapses to the exponential-gamma model when $c = 1$
- Also know as the Burr Type XII distribution

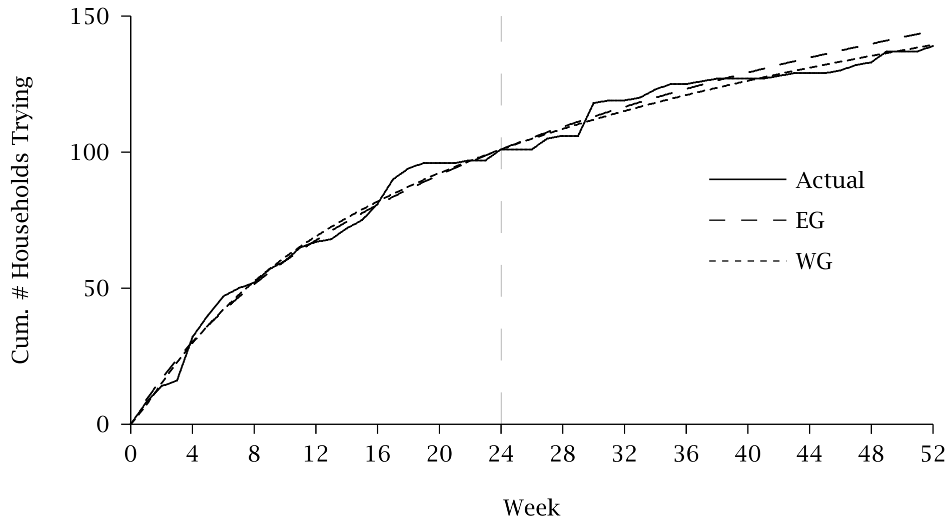
Problem 6 -- Weibull-gamma

	A	B	C	D	E	F	G	H	I
1	Product:	Krunchy Bits				r	0.031		
2	Panelists:	1499				\alpha	6.199		
3						c	1.241		
4						LL =	=SUM(G7:G31)		
5		Cum_Trl							
6	Week	# HHs	Incr_Trl		P(T <= t)	P(try week t)			E[T(t)]
7	1	8	=B7		=1-(G$2/(G$2+A7^G$3))^G$1	=E7	=C7*LN(F7)		=B$2*E7
8	2	14	=B8-B7		=1-(G$2/(G$2+A8^G$3))^G$1	=E8-E7	=C8*LN(F8)		=B$2*E8
9	3	16	=B9-B8		=1-(G$2/(G$2+A9^G$3))^G$1	=E9-E8	=C9*LN(F9)		=B$2*E9
10	4	32	=B10-B9		=1-(G$2/(G$2+A10^G$3))^G$1	=E10-E9	=C10*LN(F10)		=B$2*E10
11	5	40	=B11-B10		=1-(G$2/(G$2+A11^G$3))^G$1	=E11-E10	=C11*LN(F11)		=B$2*E11
12	6	47	=B12-B11		=1-(G$2/(G$2+A12^G$3))^G$1	=E12-E11	=C12*LN(F12)		=B$2*E12
13	7	50	=B13-B12		=1-(G$2/(G$2+A13^G$3))^G$1	=E13-E12	=C13*LN(F13)		=B$2*E13
14	8	52	=B14-B13		=1-(G$2/(G$2+A14^G$3))^G$1	=E14-E13	=C14*LN(F14)		=B$2*E14
15	9	57	=B15-B14		=1-(G$2/(G$2+A15^G$3))^G$1	=E15-E14	=C15*LN(F15)		=B$2*E15
16	10	60	=B16-B15		=1-(G$2/(G$2+A16^G$3))^G$1	=E16-E15	=C16*LN(F16)		=B$2*E16
17	11	65	=B17-B16		=1-(G$2/(G$2+A17^G$3))^G$1	=E17-E16	=C17*LN(F17)		=B$2*E17
18	12	67	=B18-B17		=1-(G$2/(G$2+A18^G$3))^G$1	=E18-E17	=C18*LN(F18)		=B$2*E18
19	13	68	=B19-B18		=1-(G$2/(G$2+A19^G$3))^G$1	=E19-E18	=C19*LN(F19)		=B$2*E19
20	14	72	=B20-B19		=1-(G$2/(G$2+A20^G$3))^G$1	=E20-E19	=C20*LN(F20)		=B$2*E20
21	15	75	=B21-B20		=1-(G$2/(G$2+A21^G$3))^G$1	=E21-E20	=C21*LN(F21)		=B$2*E21
22	16	81	=B22-B21		=1-(G$2/(G$2+A22^G$3))^G$1	=E22-E21	=C22*LN(F22)		=B$2*E22
23	17	90	=B23-B22		=1-(G$2/(G$2+A23^G$3))^G$1	=E23-E22	=C23*LN(F23)		=B$2*E23
24	18	94	=B24-B23		=1-(G$2/(G$2+A24^G$3))^G$1	=E24-E23	=C24*LN(F24)		=B$2*E24
25	19	96	=B25-B24		=1-(G$2/(G$2+A25^G$3))^G$1	=E25-E24	=C25*LN(F25)		=B$2*E25
26	20	96	=B26-B25		=1-(G$2/(G$2+A26^G$3))^G$1	=E26-E25	=C26*LN(F26)		=B$2*E26
27	21	96	=B27-B26		=1-(G$2/(G$2+A27^G$3))^G$1	=E27-E26	=C27*LN(F27)		=B$2*E27
28	22	97	=B28-B27		=1-(G$2/(G$2+A28^G$3))^G$1	=E28-E27	=C28*LN(F28)		=B$2*E28
29	23	97	=B29-B28		=1-(G$2/(G$2+A29^G$3))^G$1	=E29-E28	=C29*LN(F29)		=B$2*E29
30	24	101	=B30-B29		=1-(G$2/(G$2+A30^G$3))^G$1	=E30-E29	=C30*LN(F30)		=B$2*E30
31	25	101			=1-(G$2/(G$2+A31^G$3))^G$1	=E31-E30	=(B2-B30)*LN(1-E30)		=B$2*E31
32	26	101			=1-(G$2/(G$2+A32^G$3))^G$1	=E32-E31			=B$2*E32
33	27	105			=1-(G$2/(G$2+A33^G$3))^G$1	=E33-E32			=B$2*E33
34	28	106			=1-(G$2/(G$2+A34^G$3))^G$1	=E34-E33			=B$2*E34
35	29	106			=1-(G$2/(G$2+A35^G$3))^G$1	=E35-E34			=B$2*E35
36	30	118			=1-(G$2/(G$2+A36^G$3))^G$1	=E36-E35			=B$2*E36
37	31	119			=1-(G$2/(G$2+A37^G$3))^G$1	=E37-E36			=B$2*E37
38	32	119			=1-(G$2/(G$2+A38^G$3))^G$1	=E38-E37			=B$2*E38
39	33	120			=1-(G$2/(G$2+A39^G$3))^G$1	=E39-E38			=B$2*E39
40	34	123			=1-(G$2/(G$2+A40^G$3))^G$1	=E40-E39			=B$2*E40
41	35	125			=1-(G$2/(G$2+A41^G$3))^G$1	=E41-E40			=B$2*E41
42	36	125			=1-(G$2/(G$2+A42^G$3))^G$1	=E42-E41			=B$2*E42
43	37	126			=1-(G$2/(G$2+A43^G$3))^G$1	=E43-E42			=B$2*E43
44	38	127			=1-(G$2/(G$2+A44^G$3))^G$1	=E44-E43			=B$2*E44
45	39	127			=1-(G$2/(G$2+A45^G$3))^G$1	=E45-E44			=B$2*E45
46	40	127			=1-(G$2/(G$2+A46^G$3))^G$1	=E46-E45			=B$2*E46
47	41	127			=1-(G$2/(G$2+A47^G$3))^G$1	=E47-E46			=B$2*E47
48	42	128			=1-(G$2/(G$2+A48^G$3))^G$1	=E48-E47			=B$2*E48
49	43	129			=1-(G$2/(G$2+A49^G$3))^G$1	=E49-E48			=B$2*E49
50	44	129			=1-(G$2/(G$2+A50^G$3))^G$1	=E50-E49			=B$2*E50
51	45	129			=1-(G$2/(G$2+A51^G$3))^G$1	=E51-E50			=B$2*E51
52	46	130			=1-(G$2/(G$2+A52^G$3))^G$1	=E52-E51			=B$2*E52
53	47	132			=1-(G$2/(G$2+A53^G$3))^G$1	=E53-E52			=B$2*E53
54	48	133			=1-(G$2/(G$2+A54^G$3))^G$1	=E54-E53			=B$2*E54
55	49	137			=1-(G$2/(G$2+A55^G$3))^G$1	=E55-E54			=B$2*E55
56	50	137			=1-(G$2/(G$2+A56^G$3))^G$1	=E56-E55			=B$2*E56
57	51	137			=1-(G$2/(G$2+A57^G$3))^G$1	=E57-E56			=B$2*E57
58	52	139			=1-(G$2/(G$2+A58^G$3))^G$1	=E58-E57			=B$2*E58

Problem 6 -- Weibull-gamma

	A	B	C	D	E	F	G	H	I
1	Product:	Krunchy Bits				r	0.031		
2	Panelists:	1499				\alpha	6.199		
3						c	1.241		
4						LL =	-681.0		
5		Cum_Trl							
6	Week	# HHs	Incr_Trl		P(T <= t)	P(try week t)			E[T(t)]
7	1	8	8		0.00466	0.00466	-42.943		6.99
8	2	14	6		0.01004	0.00538	-31.350		15.06
9	3	16	2		0.01516	0.00512	-10.549		22.73
10	4	32	16		0.01988	0.00471	-85.722		29.79
11	5	40	8		0.02418	0.00430	-43.587		36.24
12	6	47	7		0.02811	0.00393	-38.773		42.14
13	7	50	3		0.03171	0.00360	-16.879		47.53
14	8	52	2		0.03502	0.00331	-11.419		52.50
15	9	57	5		0.03809	0.00306	-28.942		57.09
16	10	60	3		0.04093	0.00284	-17.589		61.35
17	11	65	5		0.04358	0.00265	-29.666		65.33
18	12	67	2		0.04606	0.00248	-11.999		69.04
19	13	68	1		0.04839	0.00233	-6.063		72.53
20	14	72	4		0.05058	0.00219	-24.490		75.82
21	15	75	3		0.05265	0.00207	-18.538		78.93
22	16	81	6		0.05461	0.00196	-37.401		81.87
23	17	90	9		0.05648	0.00186	-56.567		84.66
24	18	94	4		0.05825	0.00177	-25.339		87.32
25	19	96	2		0.05994	0.00169	-12.764		89.86
26	20	96	0		0.06156	0.00162	0.000		92.28
27	21	96	0		0.06311	0.00155	0.000		94.60
28	22	97	1		0.06459	0.00148	-6.513		96.82
29	23	97	0		0.06602	0.00143	0.000		98.96
30	24	101	4		0.06739	0.00137	-26.368		101.02
31	25	101			0.06871	0.00132	-97.536		103.00
32	26	101			0.06998	0.00127			104.91
33	27	105			0.07121	0.00123			106.75
34	28	106			0.07240	0.00119			108.53
35	29	106			0.07355	0.00115			110.25
36	30	118			0.07467	0.00111			111.92
37	31	119			0.07575	0.00108			113.54
38	32	119			0.07679	0.00105			115.11
39	33	120			0.07781	0.00102			116.64
40	34	123			0.07880	0.00099			118.12
41	35	125			0.07976	0.00096			119.56
42	36	125			0.08070	0.00094			120.96
43	37	126			0.08161	0.00091			122.33
44	38	127			0.08249	0.00089			123.66
45	39	127			0.08336	0.00087			124.96
46	40	127			0.08421	0.00084			126.22
47	41	127			0.08503	0.00082			127.46
48	42	128			0.08584	0.00081			128.67
49	43	129			0.08662	0.00079			129.85
50	44	129			0.08739	0.00077			131.00
51	45	129			0.08815	0.00075			132.13
52	46	130			0.08888	0.00074			133.24
53	47	132			0.08960	0.00072			134.32
54	48	133			0.09031	0.00071			135.38
55	49	137			0.09100	0.00069			136.41
56	50	137			0.09168	0.00068			137.43
57	51	137			0.09235	0.00067			138.43
58	52	139			0.09300	0.00065			139.41

Applying the WG Model



$$\hat{r} = 0.031,\ \hat{\alpha} = 6.199,\ \hat{c} = 1.241,\ LL = -681.0$$

Assessing Model Fit

Is the fit of the WG a significant improvement over that of the EG? (In other words, is c significantly different from 1.0?)

We compute the LR test statistic,

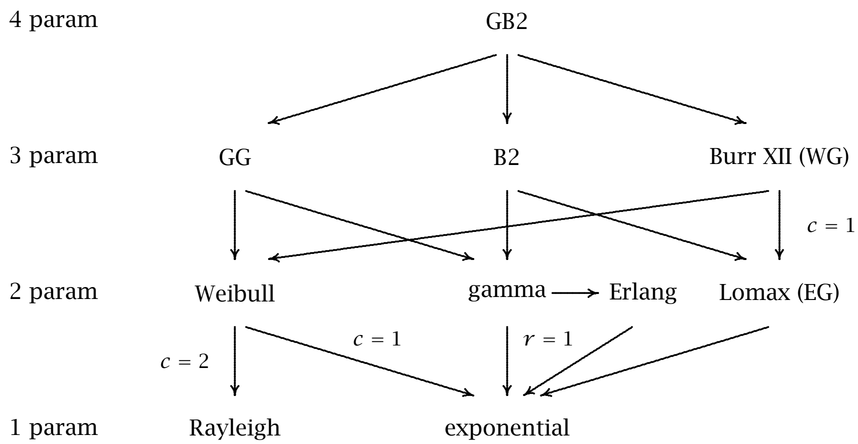
$$\begin{aligned} LR &= -2(LL_{\text{EG}} - LL_{\text{WG}}) \\ &= 0.75 \end{aligned}$$

and compare it again the critical value from the chi-squared distribution,

`=CHIINV(0.05,1)` = 3.84.

We fail to reject the null hypothesis that $c = 1$.

A Family Tree of Distributions



Adding Covariate Effects

In addition to the trial purchasing data, we also have information on the marketing activity associated with the new product while in the test market:

- *Coupon*: an aggregate measure of coupon activity, generated by IRI for modeling purposes, that reflects the face value and circulation of the coupon, along with standard decays in redemption rate.
- *AnyP:* %ACV with any merchandising (a standard market-level scanner data measure of promotional activity).

How do we incorporate these time-varying covariates into our timing model?

Adding Covariate Effects

- Intuitively, we would expect the probability of an individual buying in week t, given she has yet to make a trial purchase, to be a function of the marketing activity in week t.
- But the fact that she hasn't made a trial purchase by week t is, in part, a function of the marketing activity in the preceding weeks.
- This intuition is formally captured via the *proportional hazards* regression framework.

Proportional Hazards Regression

- Let $\mathbf{x}(t)$ denote the vector of covariates at time t and $\boldsymbol{\beta}$ the effects of these covariates.
- We assume that the covariates have a multiplicative effect on the baseline (underlying individual-level) hazard function

$$h(t|\theta, \mathbf{x}(t), \boldsymbol{\beta}) = h_0(t|\theta) \exp(\boldsymbol{\beta}'\mathbf{x}(t))$$

- For the exponential distribution, we have

$$h(t|\lambda, \mathbf{x}(t), \boldsymbol{\beta}) = \lambda \exp(\boldsymbol{\beta}'\mathbf{x}(t))$$

Proportional Hazards Regression

We derive the corresponding "with-covariates" distribution by recalling the fundamental relationship between the hazard function and cdf of a distribution:

$$F(t) = 1 - \exp\left(- \int_0^t h(u)\, du \right)$$

Proportional Hazards Regression

Assuming the covariates remain constant *within* each unit of time (e.g., week),

$$\begin{aligned}
\int_0^t h(u)\, du &= \int_0^1 \lambda \exp(\boldsymbol{\beta}'\mathbf{x}(u))\, du + \cdots + \int_{t-1}^t \lambda \exp(\boldsymbol{\beta}'\mathbf{x}(u))\, du \\
&= \lambda \exp(\boldsymbol{\beta}'\mathbf{x}(1)) + \cdots + \lambda \exp(\boldsymbol{\beta}'\mathbf{x}(t)) \\
&\equiv \lambda A(t)\,, \text{ where } A(t) = \sum_{i=1}^{t} \exp(\boldsymbol{\beta}'\mathbf{x}(i))
\end{aligned}$$

Therefore

$$F(t|\lambda, \mathbf{X}(t), \boldsymbol{\beta}) = 1 - \exp\left(- \lambda A(t)\right)$$

where $\mathbf{X}(t)$ denotes the covariate path up to time t.

Adding Covariate Effects

- Assuming λ is distributed across the population according to a gamma distribution, we have

$$\begin{aligned} P(T \le t) &= \int_0^\infty \Big(1 - \exp\big(-\lambda A(t)\big)\Big) \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}\, d\lambda \\ &= 1 - \left(\frac{\alpha}{\alpha + A(t)}\right)^r \end{aligned}$$

- We call this the EG+covariates model.
- This collapses to the EG model when $\boldsymbol{\beta} = \mathbf{0}$.

Fit of the EG+cov Model

Variable	Coefficient
r	0.103
α	55.008
Coupon	2.310
AnyP	0.015
LL	−674.0
LL_{EG}	−681.4
LR (df = 2)	14.7

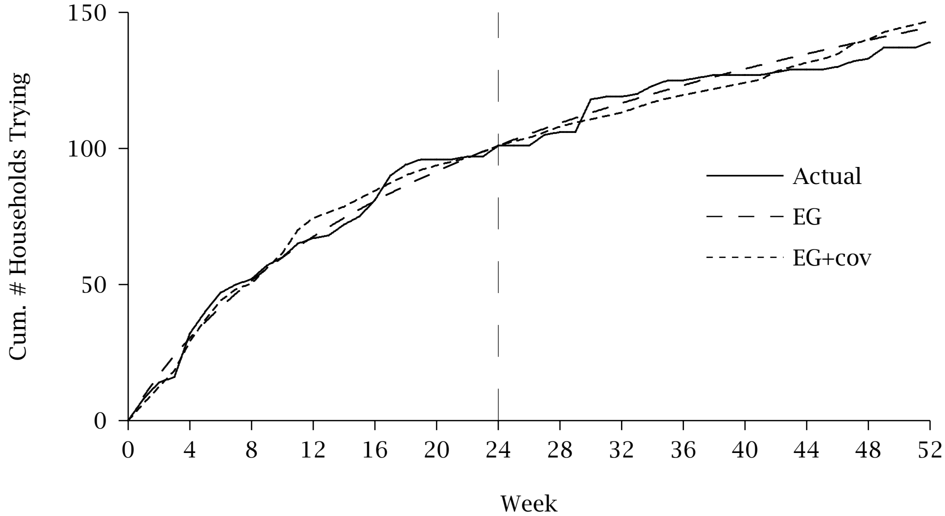
(Note: `=CHIINV(0.05,2)` = 5.99.)

Problem 6 -- EG+cov

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Product:	Krunchy Bits				r	0.103							
2	Panelists:	1499				\alpha	55.008							
3						B_coup	2.31							
4						B_AnyP	0.015							
5						LL =	=SUM(G8:G32)							
6		Cum_Trl											Covariates	
7	Week	# HHs	Incr_Trl		P(T <= t)	P(try week t)			E[T(t)]		A(t)	exp(BX)	Coupon	AnyP
8	1	8	=B8		=1-(G$2/(G$2+K8))^G$1	=E8	=C8*LN(F8)		=B$2*E8		=L8	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M8:N8))	0	55.61
9	2	14	=B9-B8		=1-(G$2/(G$2+K9))^G$1	=E9-E8	=C9*LN(F9)		=B$2*E9		=K8+L9	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M9:N9))	0	55.82
10	3	16	=B10-B9		=1-(G$2/(G$2+K10))^G$1	=E10-E9	=C10*LN(F10)		=B$2*E10		=K9+L10	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M10:N10))	0.0180370411	55.65
11	4	32	=B11-B10		=1-(G$2/(G$2+K11))^G$1	=E11-E10	=C11*LN(F11)		=B$2*E11		=K10+L11	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M11:N11))	0.1262592877	81.17
12	5	40	=B12-B11		=1-(G$2/(G$2+K12))^G$1	=E12-E11	=C12*LN(F12)		=B$2*E12		=K11+L12	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M12:N12))	0.078697419	74.45
13	6	47	=B13-B12		=1-(G$2/(G$2+K13))^G$1	=E13-E12	=C13*LN(F13)		=B$2*E13		=K12+L13	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M13:N13))	0.044160582	73.82
14	7	50	=B14-B13		=1-(G$2/(G$2+K14))^G$1	=E14-E13	=C14*LN(F14)		=B$2*E14		=K13+L14	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M14:N14))	0.0214475487	46.83
15	8	52	=B15-B14		=1-(G$2/(G$2+K15))^G$1	=E15-E14	=C15*LN(F15)		=B$2*E15		=K14+L15	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M15:N15))	0.0084743074	0
16	9	57	=B16-B15		=1-(G$2/(G$2+K16))^G$1	=E16-E15	=C16*LN(F16)		=B$2*E16		=K15+L16	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M16:N16))	0.0024570064	72.41
17	10	60	=B17-B16		=1-(G$2/(G$2+K17))^G$1	=E17-E16	=C17*LN(F17)		=B$2*E17		=K16+L17	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M17:N17))	0.015465007	71.52
18	11	65	=B18-B17		=1-(G$2/(G$2+K18))^G$1	=E18-E17	=C18*LN(F18)		=B$2*E18		=K17+L18	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M18:N18))	0.1052483182	89.32
19	12	67	=B19-B18		=1-(G$2/(G$2+K19))^G$1	=E19-E18	=C19*LN(F19)		=B$2*E19		=K18+L19	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M19:N19))	0.0655816056	56.05
20	13	68	=B20-B19		=1-(G$2/(G$2+K20))^G$1	=E20-E19	=C20*LN(F20)		=B$2*E20		=K19+L20	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M20:N20))	0.0368004851	12.58
21	14	72	=B21-B20		=1-(G$2/(G$2+K21))^G$1	=E21-E20	=C21*LN(F21)		=B$2*E21		=K20+L21	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M21:N21))	0.0178729572	13.33
22	15	75	=B22-B21		=1-(G$2/(G$2+K22))^G$1	=E22-E21	=C22*LN(F22)		=B$2*E22		=K21+L22	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M22:N22))	0.0070619229	43.31
23	16	81	=B23-B22		=1-(G$2/(G$2+K23))^G$1	=E23-E22	=C23*LN(F23)		=B$2*E23		=K22+L23	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M23:N23))	0.0020475053	41.46
24	17	90	=B24-B23		=1-(G$2/(G$2+K24))^G$1	=E24-E23	=C24*LN(F24)		=B$2*E24		=K23+L24	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M24:N24))	0.0003617828	41.95
25	18	94	=B25-B24		=1-(G$2/(G$2+K25))^G$1	=E25-E24	=C25*LN(F25)		=B$2*E25		=K24+L25	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M25:N25))	0.0000268709	44.47
26	19	96	=B26-B25		=1-(G$2/(G$2+K26))^G$1	=E26-E25	=C26*LN(F26)		=B$2*E26		=K25+L26	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M26:N26))	0.0000003526	22.9
27	20	96	=B27-B26		=1-(G$2/(G$2+K27))^G$1	=E27-E26	=C27*LN(F27)		=B$2*E27		=K26+L27	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M27:N27))	0.0000000001	8.82
28	21	96	=B28-B27		=1-(G$2/(G$2+K28))^G$1	=E28-E27	=C28*LN(F28)		=B$2*E28		=K27+L28	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M28:N28))	3.103112E-22	0
29	22	97	=B29-B28		=1-(G$2/(G$2+K29))^G$1	=E29-E28	=C29*LN(F29)		=B$2*E29		=K28+L29	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M29:N29))	0	0
30	23	97	=B30-B29		=1-(G$2/(G$2+K30))^G$1	=E30-E29	=C30*LN(F30)		=B$2*E30		=K29+L30	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M30:N30))	0	32.2
31	24	101	=B31-B30		=1-(G$2/(G$2+K31))^G$1	=E31-E30	=C31*LN(F31)		=B$2*E31		=K30+L31	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M31:N31))	0	33.4
32	25	101			=1-(G$2/(G$2+K32))^G$1	=E32-E31	=(B2-B31)*LN(1-E31)		=B$2*E32		=K31+L32	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M32:N32))	0	13.05
33	26	101			=1-(G$2/(G$2+K33))^G$1	=E33-E32			=B$2*E33		=K32+L33	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M33:N33))	0	0
34	27	105			=1-(G$2/(G$2+K34))^G$1	=E34-E33			=B$2*E34		=K33+L34	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M34:N34))	0	32.11
35	28	106			=1-(G$2/(G$2+K35))^G$1	=E35-E34			=B$2*E35		=K34+L35	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M35:N35))	0	31.26
36	29	106			=1-(G$2/(G$2+K36))^G$1	=E36-E35			=B$2*E36		=K35+L36	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M36:N36))	0	9.05
37	30	118			=1-(G$2/(G$2+K37))^G$1	=E37-E36			=B$2*E37		=K36+L37	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M37:N37))	0	0
38	31	119			=1-(G$2/(G$2+K38))^G$1	=E38-E37			=B$2*E38		=K37+L38	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M38:N38))	0	0
39	32	119			=1-(G$2/(G$2+K39))^G$1	=E39-E38			=B$2*E39		=K38+L39	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M39:N39))	0	0
40	33	120			=1-(G$2/(G$2+K40))^G$1	=E40-E39			=B$2*E40		=K39+L40	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M40:N40))	0	30.96
41	34	123			=1-(G$2/(G$2+K41))^G$1	=E41-E40			=B$2*E41		=K40+L41	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M41:N41))	0	31
42	35	125			=1-(G$2/(G$2+K42))^G$1	=E42-E41			=B$2*E42		=K41+L42	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M42:N42))	0	14.68
43	36	125			=1-(G$2/(G$2+K43))^G$1	=E43-E42			=B$2*E43		=K42+L43	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M43:N43))	0	0
44	37	126			=1-(G$2/(G$2+K44))^G$1	=E44-E43			=B$2*E44		=K43+L44	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M44:N44))	0	0
45	38	127			=1-(G$2/(G$2+K45))^G$1	=E45-E44			=B$2*E45		=K44+L45	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M45:N45))	0	0
46	39	127			=1-(G$2/(G$2+K46))^G$1	=E46-E45			=B$2*E46		=K45+L46	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M46:N46))	0	0
47	40	127			=1-(G$2/(G$2+K47))^G$1	=E47-E46			=B$2*E47		=K46+L47	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M47:N47))	0	0
48	41	127			=1-(G$2/(G$2+K48))^G$1	=E48-E47			=B$2*E48		=K47+L48	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M48:N48))	0	0
49	42	128			=1-(G$2/(G$2+K49))^G$1	=E49-E48			=B$2*E49		=K48+L49	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M49:N49))	0	68.12
50	43	129			=1-(G$2/(G$2+K50))^G$1	=E50-E49			=B$2*E50		=K49+L50	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M50:N50))	0	29.16
51	44	129			=1-(G$2/(G$2+K51))^G$1	=E51-E50			=B$2*E51		=K50+L51	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M51:N51))	0	28.13
52	45	129			=1-(G$2/(G$2+K52))^G$1	=E52-E51			=B$2*E52		=K51+L52	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M52:N52))	0	8.72
53	46	130			=1-(G$2/(G$2+K53))^G$1	=E53-E52			=B$2*E53		=K52+L53	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M53:N53))	0.0180651509	37.48
54	47	132			=1-(G$2/(G$2+K54))^G$1	=E54-E53			=B$2*E54		=K53+L54	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M54:N54))	0.1264560566	67.37
55	48	133			=1-(G$2/(G$2+K55))^G$1	=E55-E54			=B$2*E55		=K54+L55	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M55:N55))	0.078820065	21.53
56	49	137			=1-(G$2/(G$2+K56))^G$1	=E56-E55			=B$2*E56		=K55+L56	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M56:N56))	0.0442294041	61.61
57	50	137			=1-(G$2/(G$2+K57))^G$1	=E57-E56			=B$2*E57		=K56+L57	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M57:N57))	0.0214809736	24.84
58	51	137			=1-(G$2/(G$2+K58))^G$1	=E58-E57			=B$2*E58		=K57+L58	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M58:N58))	0.0084875142	20.95
59	52	139			=1-(G$2/(G$2+K59))^G$1	=E59-E58			=B$2*E59		=K58+L59	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M59:N59))	0.0024608355	21.62

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Product:	Krunchy Bits				r	0.103							
2	Panelists:	1499				\alpha	55.008							
3						B_coup	2.310							
4						B_AnyP	0.015							
5						LL =	-674.0							
6		Cum_Trl											Covariates	
7	Week	# HHs	Incr_Trl		P(T <= t)	P(try week t)			E[T(t)]		A(t)	exp(BX)	Coupon	AnyP
8	1	8	8		0.00421	0.00421	-43.769		6.31		2.29228	2.29228	0	55.61
9	2	14	6		0.00824	0.00404	-33.074		12.36		4.591751	2.299472	0	55.82
10	3	16	2		0.01226	0.00402	-11.033		18.38		6.982982	2.39123	0.018037	55.65
11	4	32	16		0.01937	0.00711	-79.141		29.04		11.47581	4.492827	0.126259	81.17
12	5	40	8		0.02476	0.00538	-41.795		37.11		15.11723	3.641419	0.078697	74.45
13	6	47	7		0.02942	0.00466	-37.579		44.10		18.44798	3.330751	0.044161	73.82
14	7	50	3		0.03226	0.00284	-17.594		48.35		20.56099	2.113016	0.021448	46.83
15	8	52	2		0.03359	0.00134	-13.233		50.36		21.58076	1.019768	0.008474	0
16	9	57	5		0.03737	0.00378	-27.892		56.02		24.54266	2.961902	0.002457	72.41
17	10	60	3		0.04106	0.00369	-16.809		61.55		27.55466	3.011996	0.015465	71.52
18	11	65	5		0.04668	0.00562	-25.911		69.97		32.38793	4.833271	0.105248	89.32
19	12	67	2		0.04965	0.00297	-11.636		74.42		35.07271	2.684783	0.065582	56.05
20	13	68	1		0.05107	0.00142	-6.558		76.55		36.38618	1.313461	0.0368	12.58
21	14	72	4		0.05242	0.00135	-26.423		78.58		37.65759	1.271417	0.017873	13.33
22	15	75	3		0.05444	0.00202	-18.608		81.61		39.59699	1.939398	0.007062	43.31
23	16	81	6		0.05635	0.00190	-37.584		84.47		41.46187	1.864882	0.002048	41.46
24	17	90	9		0.05822	0.00187	-56.537		87.27		43.33313	1.871262	0.000362	41.95
25	18	94	4		0.06012	0.00190	-25.066		90.12		45.27458	1.941442	2.69E-05	44.47
26	19	96	2		0.06147	0.00135	-13.213		92.14		46.68178	1.407205	3.53E-07	22.9
27	20	96	0		0.06255	0.00108	0.000		93.76		47.8224	1.140618	6.07E-11	8.82
28	21	96	0		0.06349	0.00094	0.000		95.16		48.8224	1	3.10E-22	0
29	22	97	1		0.06441	0.00093	-6.984		96.55		49.8224	1	0	0
30	23	97	0		0.06589	0.00148	0.000		98.77		51.43901	1.616615	0	32.2
31	24	101	4		0.06737	0.00148	-26.067		100.98		53.08483	1.645814	0	33.4
32	25	101			0.06844	0.00108	-97.501		102.60		54.29974	1.21491	0	13.05
33	26	101			0.06932	0.00088			103.91		55.29974	1	0	0
34	27	105			0.07071	0.00140			106.00		56.91418	1.614446	0	32.11
35	28	106			0.07207	0.00136			108.03		58.50829	1.594105	0	31.26
36	29	106			0.07303	0.00096			109.47		59.65283	1.144538	0	9.05
37	30	118			0.07386	0.00083			110.72		60.65283	1	0	0
38	31	119			0.07468	0.00082			111.95		61.65283	1	0	0
39	32	119			0.07550	0.00082			113.17		62.65283	1	0	0
40	33	120			0.07678	0.00128			115.09		64.23981	1.586987	0	30.96
41	34	123			0.07804	0.00126			116.98		65.82775	1.587934	0	31
42	35	125			0.07901	0.00098			118.44		67.07256	1.244812	0	14.68
43	36	125			0.07979	0.00078			119.60		68.07256	1	0	0
44	37	126			0.08056	0.00077			120.75		69.07256	1	0	0
45	38	127			0.08132	0.00076			121.89		70.07256	1	0	0
46	39	127			0.08207	0.00075			123.03		71.07256	1	0	0
47	40	127			0.08282	0.00075			124.15		72.07256	1	0	0
48	41	127			0.08356	0.00074			125.26		73.07256	1	0	0
49	42	128			0.08558	0.00202			128.28		75.83513	2.76257	0	68.12
50	43	129			0.08669	0.00111			129.94		77.38007	1.544942	0	29.16
51	44	129			0.08776	0.00108			131.56		78.90146	1.521386	0	28.13
52	45	129			0.08856	0.00080			132.75		80.04037	1.138917	0	8.72
53	46	130			0.08982	0.00126			134.64		81.864	1.823626	0.018065	37.48
54	47	132			0.09230	0.00248			138.35		85.5226	3.658594	0.126456	67.37
55	48	133			0.09339	0.00110			140.00		87.17666	1.654067	0.07882	21.53
56	49	137			0.09520	0.00181			142.71		89.95324	2.776573	0.044229	61.61
57	50	137			0.09618	0.00098			144.17		91.47545	1.52221	0.021481	24.84
58	51	137			0.09706	0.00088			145.49		92.86937	1.393922	0.008488	20.95
59	52	139			0.09793	0.00087			146.80		94.25783	1.388459	0.002461	21.62

Comparing EG with EG+cov



Adding Covariate Effects

For a general "baseline" distribution with cdf $F_0(t)$,

$$\int_0^t h(u|\theta, \mathbf{x}(u), \boldsymbol{\beta})\, du = \int_0^t h_0(u|\theta)\exp(\boldsymbol{\beta}'\mathbf{x}(u))\, du$$

$$= \sum_{i=1}^{t} \{\ln[1 - F_0(i-1|\theta)] - \ln[1 - F_0(i|\theta)]\} \exp(\boldsymbol{\beta}'\mathbf{x}(i))$$

For the Weibull distribution,

$$F(t|\lambda, c, \mathbf{X}(t), \boldsymbol{\beta}) = 1 - \exp\big(-\lambda B(t)\big)$$

where

$$B(t) = \sum_{i=1}^{t} [i^c - (i-1)^c] \exp(\boldsymbol{\beta}'\mathbf{x}(i))$$

The WG+cov Model

- Assuming λ is distributed across the population according to a gamma distribution, we have

$$P(T \le t) = \int_0^\infty \Big(1 - \exp\big(-\lambda B(t)\big)\Big) \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}\, d\lambda$$
$$= 1 - \left(\frac{\alpha}{\alpha + B(t)}\right)^r$$

- This collapses to the WG model when $\boldsymbol{\beta} = \mathbf{0}$.
- This collapses to the EG+cov model when $c = 1$.

Fit of the WG+cov Model

Variable	Coefficient
r	93.554
α	41760.598
c	0.810
Coupon	3.185
AnyP	0.015
LL	−673.6
$LL_{\text{EG+cov}}$	−674.0
LR (df = 1)	0.85

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Product:	Krunchy Bits				r	93.554					starting values:	1	
2	Panelists:	1499				\alpha	41760.6						1	
3						c	0.81						1	
4						B_coup	3.185						0	
5						B_AnyP	0.015						0	
6						LL =	=SUM(G9:G33)							
7		Cum_Trl											Covariates	
8	Week	# HHs	Incr_Trl		P(T <= t)	P(try week t)			E[T(t)]		B(t)	exp(BX)	Coupon	AnyP
9	1	8	=B9		=1-(G$2/(G$2+K9))^G$1	=E9	=C9*LN(F9)		=B$2*E9		=A9^G3*L9	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M9:N9))	0	55.61
10	2	14	=B10-B9		=1-(G$2/(G$2+K10))^G$1	=E10-E9	=C10*LN(F10)		=B$2*E10		=K9+(A10^G$3-A9^G$3)*L10	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M10:N10))	0	55.82
11	3	16	=B11-B10		=1-(G$2/(G$2+K11))^G$1	=E11-E10	=C11*LN(F11)		=B$2*E11		=K10+(A11^G$3-A10^G$3)*L11	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M11:N11))	0.0180370411	55.65
12	4	32	=B12-B11		=1-(G$2/(G$2+K12))^G$1	=E12-E11	=C12*LN(F12)		=B$2*E12		=K11+(A12^G$3-A11^G$3)*L12	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M12:N12))	0.1262592877	81.17
13	5	40	=B13-B12		=1-(G$2/(G$2+K13))^G$1	=E13-E12	=C13*LN(F13)		=B$2*E13		=K12+(A13^G$3-A12^G$3)*L13	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M13:N13))	0.078697419	74.45
14	6	47	=B14-B13		=1-(G$2/(G$2+K14))^G$1	=E14-E13	=C14*LN(F14)		=B$2*E14		=K13+(A14^G$3-A13^G$3)*L14	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M14:N14))	0.044160582	73.82
15	7	50	=B15-B14		=1-(G$2/(G$2+K15))^G$1	=E15-E14	=C15*LN(F15)		=B$2*E15		=K14+(A15^G$3-A14^G$3)*L15	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M15:N15))	0.0214475487	46.83
16	8	52	=B16-B15		=1-(G$2/(G$2+K16))^G$1	=E16-E15	=C16*LN(F16)		=B$2*E16		=K15+(A16^G$3-A15^G$3)*L16	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M16:N16))	0.0084743074	0
17	9	57	=B17-B16		=1-(G$2/(G$2+K17))^G$1	=E17-E16	=C17*LN(F17)		=B$2*E17		=K16+(A17^G$3-A16^G$3)*L17	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M17:N17))	0.0024570064	72.41
18	10	60	=B18-B17		=1-(G$2/(G$2+K18))^G$1	=E18-E17	=C18*LN(F18)		=B$2*E18		=K17+(A18^G$3-A17^G$3)*L18	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M18:N18))	0.015465007	71.52
19	11	65	=B19-B18		=1-(G$2/(G$2+K19))^G$1	=E19-E18	=C19*LN(F19)		=B$2*E19		=K18+(A19^G$3-A18^G$3)*L19	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M19:N19))	0.1052483182	89.32
20	12	67	=B20-B19		=1-(G$2/(G$2+K20))^G$1	=E20-E19	=C20*LN(F20)		=B$2*E20		=K19+(A20^G$3-A19^G$3)*L20	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M20:N20))	0.0655816056	56.05
21	13	68	=B21-B20		=1-(G$2/(G$2+K21))^G$1	=E21-E20	=C21*LN(F21)		=B$2*E21		=K20+(A21^G$3-A20^G$3)*L21	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M21:N21))	0.0368004851	12.58
22	14	72	=B22-B21		=1-(G$2/(G$2+K22))^G$1	=E22-E21	=C22*LN(F22)		=B$2*E22		=K21+(A22^G$3-A21^G$3)*L22	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M22:N22))	0.0178729572	13.33
23	15	75	=B23-B22		=1-(G$2/(G$2+K23))^G$1	=E23-E22	=C23*LN(F23)		=B$2*E23		=K22+(A23^G$3-A22^G$3)*L23	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M23:N23))	0.0070619229	43.31
24	16	81	=B24-B23		=1-(G$2/(G$2+K24))^G$1	=E24-E23	=C24*LN(F24)		=B$2*E24		=K23+(A24^G$3-A23^G$3)*L24	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M24:N24))	0.0020475053	41.46
25	17	90	=B25-B24		=1-(G$2/(G$2+K25))^G$1	=E25-E24	=C25*LN(F25)		=B$2*E25		=K24+(A25^G$3-A24^G$3)*L25	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M25:N25))	0.0003617828	41.95
26	18	94	=B26-B25		=1-(G$2/(G$2+K26))^G$1	=E26-E25	=C26*LN(F26)		=B$2*E26		=K25+(A26^G$3-A25^G$3)*L26	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M26:N26))	0.0000268709	44.47
27	19	96	=B27-B26		=1-(G$2/(G$2+K27))^G$1	=E27-E26	=C27*LN(F27)		=B$2*E27		=K26+(A27^G$3-A26^G$3)*L27	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M27:N27))	0.0000003526	22.9
28	20	96	=B28-B27		=1-(G$2/(G$2+K28))^G$1	=E28-E27	=C28*LN(F28)		=B$2*E28		=K27+(A28^G$3-A27^G$3)*L28	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M28:N28))	0.0000000001	8.82
29	21	96	=B29-B28		=1-(G$2/(G$2+K29))^G$1	=E29-E28	=C29*LN(F29)		=B$2*E29		=K28+(A29^G$3-A28^G$3)*L29	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M29:N29))	3.103112E-22	0
30	22	97	=B30-B29		=1-(G$2/(G$2+K30))^G$1	=E30-E29	=C30*LN(F30)		=B$2*E30		=K29+(A30^G$3-A29^G$3)*L30	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M30:N30))	0	0
31	23	97	=B31-B30		=1-(G$2/(G$2+K31))^G$1	=E31-E30	=C31*LN(F31)		=B$2*E31		=K30+(A31^G$3-A30^G$3)*L31	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M31:N31))	0	32.2
32	24	101	=B32-B31		=1-(G$2/(G$2+K32))^G$1	=E32-E31	=C32*LN(F32)		=B$2*E32		=K31+(A32^G$3-A31^G$3)*L32	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M32:N32))	0	33.4
33	25	101			=1-(G$2/(G$2+K33))^G$1	=E33-E32	=(B2-B32)*LN(1-E32)		=B$2*E33		=K32+(A33^G$3-A32^G$3)*L33	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M33:N33))	0	13.05
34	26	101			=1-(G$2/(G$2+K34))^G$1	=E34-E33			=B$2*E34		=K33+(A34^G$3-A33^G$3)*L34	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M34:N34))	0	0
35	27	105			=1-(G$2/(G$2+K35))^G$1	=E35-E34			=B$2*E35		=K34+(A35^G$3-A34^G$3)*L35	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M35:N35))	0	32.11
36	28	106			=1-(G$2/(G$2+K36))^G$1	=E36-E35			=B$2*E36		=K35+(A36^G$3-A35^G$3)*L36	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M36:N36))	0	31.26
37	29	106			=1-(G$2/(G$2+K37))^G$1	=E37-E36			=B$2*E37		=K36+(A37^G$3-A36^G$3)*L37	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M37:N37))	0	9.05
38	30	118			=1-(G$2/(G$2+K38))^G$1	=E38-E37			=B$2*E38		=K37+(A38^G$3-A37^G$3)*L38	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M38:N38))	0	0
39	31	119			=1-(G$2/(G$2+K39))^G$1	=E39-E38			=B$2*E39		=K38+(A39^G$3-A38^G$3)*L39	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M39:N39))	0	0
40	32	119			=1-(G$2/(G$2+K40))^G$1	=E40-E39			=B$2*E40		=K39+(A40^G$3-A39^G$3)*L40	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M40:N40))	0	0
41	33	120			=1-(G$2/(G$2+K41))^G$1	=E41-E40			=B$2*E41		=K40+(A41^G$3-A40^G$3)*L41	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M41:N41))	0	30.96
42	34	123			=1-(G$2/(G$2+K42))^G$1	=E42-E41			=B$2*E42		=K41+(A42^G$3-A41^G$3)*L42	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M42:N42))	0	31
43	35	125			=1-(G$2/(G$2+K43))^G$1	=E43-E42			=B$2*E43		=K42+(A43^G$3-A42^G$3)*L43	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M43:N43))	0	14.68
44	36	125			=1-(G$2/(G$2+K44))^G$1	=E44-E43			=B$2*E44		=K43+(A44^G$3-A43^G$3)*L44	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M44:N44))	0	0
45	37	126			=1-(G$2/(G$2+K45))^G$1	=E45-E44			=B$2*E45		=K44+(A45^G$3-A44^G$3)*L45	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M45:N45))	0	0
46	38	127			=1-(G$2/(G$2+K46))^G$1	=E46-E45			=B$2*E46		=K45+(A46^G$3-A45^G$3)*L46	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M46:N46))	0	0
47	39	127			=1-(G$2/(G$2+K47))^G$1	=E47-E46			=B$2*E47		=K46+(A47^G$3-A46^G$3)*L47	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M47:N47))	0	0
48	40	127			=1-(G$2/(G$2+K48))^G$1	=E48-E47			=B$2*E48		=K47+(A48^G$3-A47^G$3)*L48	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M48:N48))	0	0
49	41	127			=1-(G$2/(G$2+K49))^G$1	=E49-E48			=B$2*E49		=K48+(A49^G$3-A48^G$3)*L49	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M49:N49))	0	0
50	42	128			=1-(G$2/(G$2+K50))^G$1	=E50-E49			=B$2*E50		=K49+(A50^G$3-A49^G$3)*L50	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M50:N50))	0	68.12
51	43	129			=1-(G$2/(G$2+K51))^G$1	=E51-E50			=B$2*E51		=K50+(A51^G$3-A50^G$3)*L51	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M51:N51))	0	29.16
52	44	129			=1-(G$2/(G$2+K52))^G$1	=E52-E51			=B$2*E52		=K51+(A52^G$3-A51^G$3)*L52	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M52:N52))	0	28.13
53	45	129			=1-(G$2/(G$2+K53))^G$1	=E53-E52			=B$2*E53		=K52+(A53^G$3-A52^G$3)*L53	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M53:N53))	0	8.72
54	46	130			=1-(G$2/(G$2+K54))^G$1	=E54-E53			=B$2*E54		=K53+(A54^G$3-A53^G$3)*L54	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M54:N54))	0.0180651509	37.48
55	47	132			=1-(G$2/(G$2+K55))^G$1	=E55-E54			=B$2*E55		=K54+(A55^G$3-A54^G$3)*L55	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M55:N55))	0.1264560566	67.37
56	48	133			=1-(G$2/(G$2+K56))^G$1	=E56-E55			=B$2*E56		=K55+(A56^G$3-A55^G$3)*L56	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M56:N56))	0.078820065	21.53
57	49	137			=1-(G$2/(G$2+K57))^G$1	=E57-E56			=B$2*E57		=K56+(A57^G$3-A56^G$3)*L57	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M57:N57))	0.0442294041	61.61
58	50	137			=1-(G$2/(G$2+K58))^G$1	=E58-E57			=B$2*E58		=K57+(A58^G$3-A57^G$3)*L58	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M58:N58))	0.0214809736	24.84
59	51	137			=1-(G$2/(G$2+K59))^G$1	=E59-E58			=B$2*E59		=K58+(A59^G$3-A58^G$3)*L59	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M59:N59))	0.0084875142	20.95
60	52	139			=1-(G$2/(G$2+K60))^G$1	=E60-E59			=B$2*E60		=K59+(A60^G$3-A59^G$3)*L60	=EXP(SUMPRODUCT(TRANSPOSE(G$4:G$5),M60:N60))	0.0024608355	21.62

Problem 6 -- WG+cov

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Product:	Krunchy Bits				r	93.554				starting values:		1.000	
2	Panelists:	1499				\alpha	41760.6						1.000	
3						c	0.810						1.000	
4						B_coup	3.185						0.000	
5						B_AnyP	0.015						0.000	
6						LL =	-673.6							
7		Cum_Trl											Covariates	
8	Week	# HHs	Incr_Trl		P(T <= t)	P(try week t)			E[T(t)]		B(t)	exp(BX)	Coupon	AnyP
9	1	8	8		0.00514	0.00514	-42.160		7.71		2.302024	2.302024	0	55.61
10	2	14	6		0.00901	0.00387	-33.329		13.51		4.041492	2.309284	0	55.82
11	3	16	2		0.01270	0.00368	-11.207		19.03		5.70447	2.439589	0.018037	55.65
12	4	32	16		0.01981	0.00711	-79.144		29.69		8.930393	5.048604	0.126259	81.17
13	5	40	8		0.02504	0.00523	-42.025		37.53		11.31957	3.923125	0.078697	74.45
14	6	47	7		0.02948	0.00445	-37.911		44.19		13.36015	3.481473	0.044161	73.82
15	7	50	3		0.03214	0.00266	-17.785		48.18		14.58695	2.160735	0.021448	46.83
16	8	52	2		0.03337	0.00123	-13.402		50.03		15.15458	1.027355	0.008474	0
17	9	57	5		0.03685	0.00348	-28.304		55.24		16.76486	2.984715	0.002457	72.41
18	10	60	3		0.04034	0.00349	-16.973		60.48		18.38633	3.069718	0.015465	71.52
19	11	65	5		0.04627	0.00592	-25.644		69.36		21.15153	5.335575	0.105248	89.32
20	12	67	2		0.04937	0.00310	-11.552		74.00		22.60603	2.855477	0.065582	56.05
21	13	68	1		0.05082	0.00145	-6.538		76.17		23.28674	1.35773	0.0368	12.58
22	14	72	4		0.05217	0.00136	-26.412		78.21		23.92547	1.292763	0.017873	13.33
23	15	75	3		0.05420	0.00202	-18.610		81.24		24.87977	1.957872	0.007062	43.31
24	16	81	6		0.05610	0.00191	-37.570		84.10		25.78175	1.874142	0.002048	41.46
25	17	90	9		0.05799	0.00189	-56.462		86.93		26.67484	1.877854	0.000362	41.95
26	18	94	4		0.05992	0.00193	-25.000		89.82		27.59103	1.948085	2.69E-05	44.47
27	19	96	2		0.06130	0.00138	-13.172		91.89		28.24704	1.409665	3.53E-07	22.9
28	20	96	0		0.06240	0.00110	0.000		93.54		28.77291	1.141385	6.07E-11	8.82
29	21	96	0		0.06336	0.00096	0.000		94.98		29.22928	1	3.10E-22	0
30	22	97	1		0.06431	0.00095	-6.961		96.40		29.68154	1	0	0
31	23	97	0		0.06583	0.00152	0.000		98.68		30.40817	1.620591	0	32.2
32	24	101	4		0.06736	0.00153	-25.922		100.98		31.1419	1.650013	0	33.4
33	25	101			0.06848	0.00112	-97.496		102.66		31.67843	1.21612	0	13.05
34	26	101			0.06940	0.00091			104.02		32.11626	1	0	0
35	27	105			0.07086	0.00146			106.22		32.8197	1.618405	0	32.11
36	28	106			0.07229	0.00143			108.37		33.50935	1.597911	0	31.26
37	29	106			0.07331	0.00102			109.89		34.00033	1.145328	0	9.05
38	30	118			0.07419	0.00088			111.22		34.42621	1	0	0
39	31	119			0.07507	0.00088			112.53		34.8494	1	0	0
40	32	119			0.07594	0.00087			113.84		35.27001	1	0	0
41	33	120			0.07732	0.00137			115.90		35.93512	1.590739	0	30.96
42	34	123			0.07868	0.00137			117.94		36.59681	1.591694	0	31
43	35	125			0.07974	0.00106			119.54		37.11199	1.246207	0	14.68
44	36	125			0.08059	0.00085			120.80		37.52315	1	0	0
45	37	126			0.08143	0.00084			122.07		37.93215	1	0	0
46	38	127			0.08227	0.00084			123.32		38.33905	1	0	0
47	39	127			0.08310	0.00083			124.57		38.74392	1	0	0
48	40	127			0.08393	0.00083			125.80		39.14683	1	0	0
49	41	127			0.08475	0.00082			127.04		39.54782	1	0	0
50	42	128			0.08701	0.00227			130.43		40.65622	2.776962	0	68.12
51	43	129			0.08827	0.00126			132.32		41.27146	1.548382	0	29.16
52	44	129			0.08950	0.00123			134.16		41.87459	1.524654	0	28.13
53	45	129			0.09041	0.00091			135.53		42.32349	1.139675	0	8.72
54	46	130			0.09190	0.00148			137.75		43.05224	1.857975	0.018065	37.48
55	47	132			0.09515	0.00325			142.63		44.65668	4.107555	0.126456	67.37
56	48	133			0.09655	0.00140			144.73		45.34723	1.77504	0.07882	21.53
57	49	137			0.09882	0.00227			148.13		46.47085	2.899691	0.044229	61.61
58	50	137			0.10003	0.00121			149.94		47.07069	1.554024	0.021481	24.84
59	51	137			0.10112	0.00109			151.57		47.61156	1.406555	0.008488	20.95
60	52	139			0.10219	0.00107			153.18		48.14551	1.393748	0.002461	21.62

Fit of the W+cov Model

- Note that the fitted gamma distribution is tending to a "spike" at $r/\alpha = 0.00224$.
- We fit the plain Weibull+covariates model to the data.

Variable	Coefficient
λ	0.002
c	0.810
Coupon	3.184
AnyP	0.015
LL	−673.6

Concepts and Tools Introduced

- Hazard functions
- Alternative individual-level timing models (e.g., Weibull)
- Incorporating time-varying covariates into (single-event) timing models

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Product:	Krunchy Bits				\lambda	0.002					starting values:	0.1	
2	Panelists:	1499				c	0.81						1	
3						B_coup	3.184						0	
4						B_AnyP	0.015						0	
5						LL =	=SUM(G8:G32)							
6		Cum_Trl											Covariates	
7	Week	# HHs	Incr_Trl		P(T <= t)	P(try week t)			E[T(t)]		B(t)	exp(BX)	Coupon	AnyP
8	1	8	=B8		=1-EXP(-G$1*K8)	=E8	=C8*LN(F8)		=B$2*E8		=A8^G2*L8	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M8:N8))	0	55.61
9	2	14	=B9-B8		=1-EXP(-G$1*K9)	=E9-E8	=C9*LN(F9)		=B$2*E9		=K8+(A9^G$2-A8^G$2)*L9	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M9:N9))	0	55.82
10	3	16	=B10-B9		=1-EXP(-G$1*K10)	=E10-E9	=C10*LN(F10)		=B$2*E10		=K9+(A10^G$2-A9^G$2)*L10	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M10:N10))	0.0180370411	55.65
11	4	32	=B11-B10		=1-EXP(-G$1*K11)	=E11-E10	=C11*LN(F11)		=B$2*E11		=K10+(A11^G$2-A10^G$2)*L11	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M11:N11))	0.1262592877	81.17
12	5	40	=B12-B11		=1-EXP(-G$1*K12)	=E12-E11	=C12*LN(F12)		=B$2*E12		=K11+(A12^G$2-A11^G$2)*L12	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M12:N12))	0.078697419	74.45
13	6	47	=B13-B12		=1-EXP(-G$1*K13)	=E13-E12	=C13*LN(F13)		=B$2*E13		=K12+(A13^G$2-A12^G$2)*L13	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M13:N13))	0.044160582	73.82
14	7	50	=B14-B13		=1-EXP(-G$1*K14)	=E14-E13	=C14*LN(F14)		=B$2*E14		=K13+(A14^G$2-A13^G$2)*L14	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M14:N14))	0.0214475487	46.83
15	8	52	=B15-B14		=1-EXP(-G$1*K15)	=E15-E14	=C15*LN(F15)		=B$2*E15		=K14+(A15^G$2-A14^G$2)*L15	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M15:N15))	0.0084743074	0
16	9	57	=B16-B15		=1-EXP(-G$1*K16)	=E16-E15	=C16*LN(F16)		=B$2*E16		=K15+(A16^G$2-A15^G$2)*L16	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M16:N16))	0.0024570064	72.41
17	10	60	=B17-B16		=1-EXP(-G$1*K17)	=E17-E16	=C17*LN(F17)		=B$2*E17		=K16+(A17^G$2-A16^G$2)*L17	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M17:N17))	0.015465007	71.52
18	11	65	=B18-B17		=1-EXP(-G$1*K18)	=E18-E17	=C18*LN(F18)		=B$2*E18		=K17+(A18^G$2-A17^G$2)*L18	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M18:N18))	0.1052483182	89.32
19	12	67	=B19-B18		=1-EXP(-G$1*K19)	=E19-E18	=C19*LN(F19)		=B$2*E19		=K18+(A19^G$2-A18^G$2)*L19	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M19:N19))	0.0655816056	56.05
20	13	68	=B20-B19		=1-EXP(-G$1*K20)	=E20-E19	=C20*LN(F20)		=B$2*E20		=K19+(A20^G$2-A19^G$2)*L20	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M20:N20))	0.0368004851	12.58
21	14	72	=B21-B20		=1-EXP(-G$1*K21)	=E21-E20	=C21*LN(F21)		=B$2*E21		=K20+(A21^G$2-A20^G$2)*L21	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M21:N21))	0.0178729572	13.33
22	15	75	=B22-B21		=1-EXP(-G$1*K22)	=E22-E21	=C22*LN(F22)		=B$2*E22		=K21+(A22^G$2-A21^G$2)*L22	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M22:N22))	0.0070619229	43.31
23	16	81	=B23-B22		=1-EXP(-G$1*K23)	=E23-E22	=C23*LN(F23)		=B$2*E23		=K22+(A23^G$2-A22^G$2)*L23	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M23:N23))	0.0020475053	41.46
24	17	90	=B24-B23		=1-EXP(-G$1*K24)	=E24-E23	=C24*LN(F24)		=B$2*E24		=K23+(A24^G$2-A23^G$2)*L24	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M24:N24))	0.0003617828	41.95
25	18	94	=B25-B24		=1-EXP(-G$1*K25)	=E25-E24	=C25*LN(F25)		=B$2*E25		=K24+(A25^G$2-A24^G$2)*L25	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M25:N25))	0.0000268709	44.47
26	19	96	=B26-B25		=1-EXP(-G$1*K26)	=E26-E25	=C26*LN(F26)		=B$2*E26		=K25+(A26^G$2-A25^G$2)*L26	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M26:N26))	0.0000003526	22.9
27	20	96	=B27-B26		=1-EXP(-G$1*K27)	=E27-E26	=C27*LN(F27)		=B$2*E27		=K26+(A27^G$2-A26^G$2)*L27	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M27:N27))	0.0000000001	8.82
28	21	96	=B28-B27		=1-EXP(-G$1*K28)	=E28-E27	=C28*LN(F28)		=B$2*E28		=K27+(A28^G$2-A27^G$2)*L28	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M28:N28))	3.103112E-22	0
29	22	97	=B29-B28		=1-EXP(-G$1*K29)	=E29-E28	=C29*LN(F29)		=B$2*E29		=K28+(A29^G$2-A28^G$2)*L29	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M29:N29))	0	0
30	23	97	=B30-B29		=1-EXP(-G$1*K30)	=E30-E29	=C30*LN(F30)		=B$2*E30		=K29+(A30^G$2-A29^G$2)*L30	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M30:N30))	0	32.2
31	24	101	=B31-B30		=1-EXP(-G$1*K31)	=E31-E30	=C31*LN(F31)		=B$2*E31		=K30+(A31^G$2-A30^G$2)*L31	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M31:N31))	0	33.4
32	25	101			=1-EXP(-G$1*K32)	=E32-E31	=(B2-B31)*LN(1-E31)		=B$2*E32		=K31+(A32^G$2-A31^G$2)*L32	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M32:N32))	0	13.05
33	26	101			=1-EXP(-G$1*K33)	=E33-E32			=B$2*E33		=K32+(A33^G$2-A32^G$2)*L33	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M33:N33))	0	0
34	27	105			=1-EXP(-G$1*K34)	=E34-E33			=B$2*E34		=K33+(A34^G$2-A33^G$2)*L34	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M34:N34))	0	32.11
35	28	106			=1-EXP(-G$1*K35)	=E35-E34			=B$2*E35		=K34+(A35^G$2-A34^G$2)*L35	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M35:N35))	0	31.26
36	29	106			=1-EXP(-G$1*K36)	=E36-E35			=B$2*E36		=K35+(A36^G$2-A35^G$2)*L36	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M36:N36))	0	9.05
37	30	118			=1-EXP(-G$1*K37)	=E37-E36			=B$2*E37		=K36+(A37^G$2-A36^G$2)*L37	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M37:N37))	0	0
38	31	119			=1-EXP(-G$1*K38)	=E38-E37			=B$2*E38		=K37+(A38^G$2-A37^G$2)*L38	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M38:N38))	0	0
39	32	119			=1-EXP(-G$1*K39)	=E39-E38			=B$2*E39		=K38+(A39^G$2-A38^G$2)*L39	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M39:N39))	0	0
40	33	120			=1-EXP(-G$1*K40)	=E40-E39			=B$2*E40		=K39+(A40^G$2-A39^G$2)*L40	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M40:N40))	0	30.96
41	34	123			=1-EXP(-G$1*K41)	=E41-E40			=B$2*E41		=K40+(A41^G$2-A40^G$2)*L41	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M41:N41))	0	31
42	35	125			=1-EXP(-G$1*K42)	=E42-E41			=B$2*E42		=K41+(A42^G$2-A41^G$2)*L42	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M42:N42))	0	14.68
43	36	125			=1-EXP(-G$1*K43)	=E43-E42			=B$2*E43		=K42+(A43^G$2-A42^G$2)*L43	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M43:N43))	0	0
44	37	126			=1-EXP(-G$1*K44)	=E44-E43			=B$2*E44		=K43+(A44^G$2-A43^G$2)*L44	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M44:N44))	0	0
45	38	127			=1-EXP(-G$1*K45)	=E45-E44			=B$2*E45		=K44+(A45^G$2-A44^G$2)*L45	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M45:N45))	0	0
46	39	127			=1-EXP(-G$1*K46)	=E46-E45			=B$2*E46		=K45+(A46^G$2-A45^G$2)*L46	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M46:N46))	0	0
47	40	127			=1-EXP(-G$1*K47)	=E47-E46			=B$2*E47		=K46+(A47^G$2-A46^G$2)*L47	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M47:N47))	0	0
48	41	127			=1-EXP(-G$1*K48)	=E48-E47			=B$2*E48		=K47+(A48^G$2-A47^G$2)*L48	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M48:N48))	0	0
49	42	128			=1-EXP(-G$1*K49)	=E49-E48			=B$2*E49		=K48+(A49^G$2-A48^G$2)*L49	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M49:N49))	0	68.12
50	43	129			=1-EXP(-G$1*K50)	=E50-E49			=B$2*E50		=K49+(A50^G$2-A49^G$2)*L50	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M50:N50))	0	29.16
51	44	129			=1-EXP(-G$1*K51)	=E51-E50			=B$2*E51		=K50+(A51^G$2-A50^G$2)*L51	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M51:N51))	0	28.13
52	45	129			=1-EXP(-G$1*K52)	=E52-E51			=B$2*E52		=K51+(A52^G$2-A51^G$2)*L52	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M52:N52))	0	8.72
53	46	130			=1-EXP(-G$1*K53)	=E53-E52			=B$2*E53		=K52+(A53^G$2-A52^G$2)*L53	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M53:N53))	0.0180651509	37.48
54	47	132			=1-EXP(-G$1*K54)	=E54-E53			=B$2*E54		=K53+(A54^G$2-A53^G$2)*L54	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M54:N54))	0.1264560566	67.37
55	48	133			=1-EXP(-G$1*K55)	=E55-E54			=B$2*E55		=K54+(A55^G$2-A54^G$2)*L55	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M55:N55))	0.078820065	21.53
56	49	137			=1-EXP(-G$1*K56)	=E56-E55			=B$2*E56		=K55+(A56^G$2-A55^G$2)*L56	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M56:N56))	0.0442294041	61.61
57	50	137			=1-EXP(-G$1*K57)	=E57-E56			=B$2*E57		=K56+(A57^G$2-A56^G$2)*L57	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M57:N57))	0.0214809736	24.84
58	51	137			=1-EXP(-G$1*K58)	=E58-E57			=B$2*E58		=K57+(A58^G$2-A57^G$2)*L58	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M58:N58))	0.0084875142	20.95
59	52	139			=1-EXP(-G$1*K59)	=E59-E58			=B$2*E59		=K58+(A59^G$2-A58^G$2)*L59	=EXP(SUMPRODUCT(TRANSPOSE(G$3:G$4),M59:N59))	0.0024608355	21.62

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Product:	Krunchy Bits				\lambda	0.002					starting values:	0.100	
2	Panelists:	1499				c	0.810						1.000	
3						B_coup	3.184						0.000	
4						B_AnyP	0.015						0.000	
5						LL =	-673.6							
6		Cum_Trl											Covariates	
7	Week	# HHs	Incr_Trl		P(T <= t)	P(try week t)			E[T(t)]		B(t)	exp(BX)	Coupon	AnyP
8	1	8	8		0.00514	0.00514	-42.159		7.71		2.302504	2.302504	0	55.61
9	2	14	6		0.00901	0.00387	-33.329		13.51		4.041946	2.309767	0	55.82
10	3	16	2		0.01270	0.00368	-11.207		19.03		5.704751	2.440064	0.018037	55.65
11	4	32	16		0.01980	0.00711	-79.146		29.69		8.930218	5.049635	0.126259	81.17
12	5	40	8		0.02504	0.00523	-42.026		37.53		11.319	3.923976	0.078697	74.45
13	6	47	7		0.02948	0.00445	-37.911		44.19		13.35925	3.482315	0.044161	73.82
14	7	50	3		0.03214	0.00266	-17.785		48.18		14.58571	2.161078	0.021448	46.83
15	8	52	2		0.03337	0.00123	-13.403		50.03		15.15307	1.027348	0.008474	0
16	9	57	5		0.03685	0.00348	-28.304		55.24		16.76303	2.98552	0.002457	72.41
17	10	60	3		0.04034	0.00349	-16.973		60.47		18.38413	3.070504	0.015465	71.52
18	11	65	5		0.04627	0.00592	-25.644		69.35		21.14865	5.336917	0.105248	89.32
19	12	67	2		0.04937	0.00310	-11.552		74.00		22.60263	2.855929	0.065582	56.05
20	13	68	1		0.05082	0.00145	-6.538		76.17		23.283	1.357754	0.0368	12.58
21	14	72	4		0.05217	0.00136	-26.412		78.21		23.92141	1.29281	0.017873	13.33
22	15	75	3		0.05420	0.00202	-18.609		81.24		24.87534	1.958179	0.007062	43.31
23	16	81	6		0.05610	0.00191	-37.570		84.10		25.77696	1.874431	0.002048	41.46
24	17	90	9		0.05799	0.00189	-56.462		86.93		26.66969	1.878149	0.000362	41.95
25	18	94	4		0.05992	0.00193	-25.000		89.82		27.58551	1.94841	2.69E-05	44.47
26	19	96	2		0.06130	0.00138	-13.172		91.89		28.24119	1.409787	3.53E-07	22.9
27	20	96	0		0.06240	0.00110	0.000		93.54		28.76677	1.141423	6.07E-11	8.82
28	21	96	0		0.06336	0.00096	0.000		94.98		29.22288	1	3.10E-22	0
29	22	97	1		0.06431	0.00095	-6.961		96.40		29.67487	1	0	0
30	23	97	0		0.06583	0.00152	0.000		98.68		30.40114	1.620787	0	32.2
31	24	101	4		0.06736	0.00153	-25.921		100.98		31.13452	1.65022	0	33.4
32	25	101			0.06848	0.00112	-97.496		102.66		31.67074	1.216179	0	13.05
33	26	101			0.06940	0.00091			104.02		32.1083	1	0	0
34	27	105			0.07086	0.00146			106.22		32.81138	1.618601	0	32.11
35	28	106			0.07229	0.00143			108.37		33.50068	1.598098	0	31.26
36	29	106			0.07331	0.00102			109.89		33.99136	1.145367	0	9.05
37	30	118			0.07420	0.00088			111.22		34.41697	1	0	0
38	31	119			0.07507	0.00088			112.53		34.83988	1	0	0
39	32	119			0.07594	0.00087			113.84		35.26022	1	0	0
40	33	120			0.07732	0.00137			115.90		35.92497	1.590924	0	30.96
41	34	123			0.07868	0.00137			117.95		36.5863	1.591879	0	31
42	35	125			0.07974	0.00106			119.54		37.10117	1.246276	0	14.68
43	36	125			0.08059	0.00085			120.81		37.51205	1	0	0
44	37	126			0.08143	0.00084			122.07		37.92077	1	0	0
45	38	127			0.08227	0.00084			123.32		38.3274	1	0	0
46	39	127			0.08310	0.00083			124.57		38.732	1	0	0
47	40	127			0.08393	0.00083			125.81		39.13463	1	0	0
48	41	127			0.08475	0.00082			127.04		39.53535	1	0	0
49	42	128			0.08702	0.00227			130.44		40.64326	2.777672	0	68.12
50	43	129			0.08827	0.00126			132.32		41.25814	1.548552	0	29.16
51	44	129			0.08950	0.00123			134.17		41.86092	1.524815	0	28.13
52	45	129			0.09042	0.00091			135.54		42.30952	1.139713	0	8.72
53	46	130			0.09190	0.00148			137.76		43.03784	1.85821	0.018065	37.48
54	47	132			0.09516	0.00326			142.64		44.6414	4.108181	0.126456	67.37
55	48	133			0.09655	0.00140			144.74		45.33147	1.775072	0.07882	21.53
56	49	137			0.09882	0.00227			148.14		46.4545	2.900259	0.044229	61.61
57	50	137			0.10003	0.00121			149.95		47.05397	1.554142	0.021481	24.84
58	51	137			0.10112	0.00109			151.58		47.59448	1.406656	0.008488	20.95
59	52	139			0.10220	0.00107			153.19		48.12809	1.393858	0.002461	21.62

Further Reading

Evans, Merran, Nicholas Hastings, and Brian Peacock (2000), *Statistical Distributions*, 3rd edition, New York: Wiley.

Fader, Peter S., Bruce G. S. Hardie, and Robert Zeithammer (2003), "Forecasting New Product Trial in a Controlled Test Market Environment," *Journal of Forecasting*, forthcoming.

Kalbfleisch, John D. and Ross L. Prentice (2002), *The Statistical Analysis of Failure Time Data*, 2nd edn., New York: Wiley.

Lawless, J. F. (1982), *Statistical Models and Methods for Lifetime Data*, New York: Wiley.

Meeker, William Q. and Luis A. Escobar (1998), *Statistical Methods for Reliability Data*, New York: Wiley.

Recap

- Modeling timing data
- Modeling count data
- Modeling "choice" data

— — — — — — — — — — — — — — —

- Introduction to finite mixture models
- Incorporating covariates in count models
- Extending basic models for timing data

The Excel spreadsheets associated with this tutorial, along with electronic copies of the tutorial materials, can be found at:

`http://brucehardie.com/talks.html`