Graded Assignment: Data Analysis Project

Bayesian Statistics Specialization: Course 2, Techniques and Models

Tiago B. Lacerda

2025-05-11

Executive Summary

This report presents a comprehensive analysis of the Excellent Consistent Quality (ECQ) success rate across various network sites in Brazil's Northeast region. The primary objective is to identify sites with significantly lower ECQ success probabilities compared to the network average, enabling targeted interventions for network improvement. A three-level Bayesian hierarchical model is employed to account for data scarcity and local variability, providing stable estimates of each site's true ECQ success probability. The analysis reveals critical insights into site performance, guiding data-informed investment decisions.

Introduction

In contemporary 4G and 5G mobile networks, operators must efficiently allocate limited resources to ensure high service quality. A pivotal metric in this context is the Excellent Consistent Quality (ECQ) success rate at individual sites. ECQ assessments evaluate whether networks consistently support demanding applications such as video streaming, video calls, and gaming, ensuring a seamless user experience. These tests are typically conducted with embedded SDKs in applications. They measure KPIs including download speed, upload speed, latency, jitter, packet loss, and time to first byte, aligning with thresholds recommended for various demanding applications. However, the variability in the number of tests across sites—some reporting only a handful while others report hundreds due to natural user mobility—poses a significant challenge. Naïve "site-by-site" estimates can be misleading: small samples may produce extreme rates simply due to chance, and citywide averages can obscure localized underperformance. To address this, a three-level Bayesian hierarchical model is proposed, nesting individual sites within municipalities and municipalities within ANFs.

Problem Definition

Our network comprises multiple sites scattered across a city, each running a varying number of ECQ tests. Some sites may report as few as 5–20 tests in a given period, while others conduct several hundred. The core challenge is to identify which sites genuinely underperform in terms of ECQ success rate and thus prioritize network improvement investments, without being misled by the randomness inherent in small test counts.

Specific Question

> *Which sites have a true ECQ success probability significantly below the network average, and how can rank them for targeted interventions, accounting for both data scarcity and local variability?*

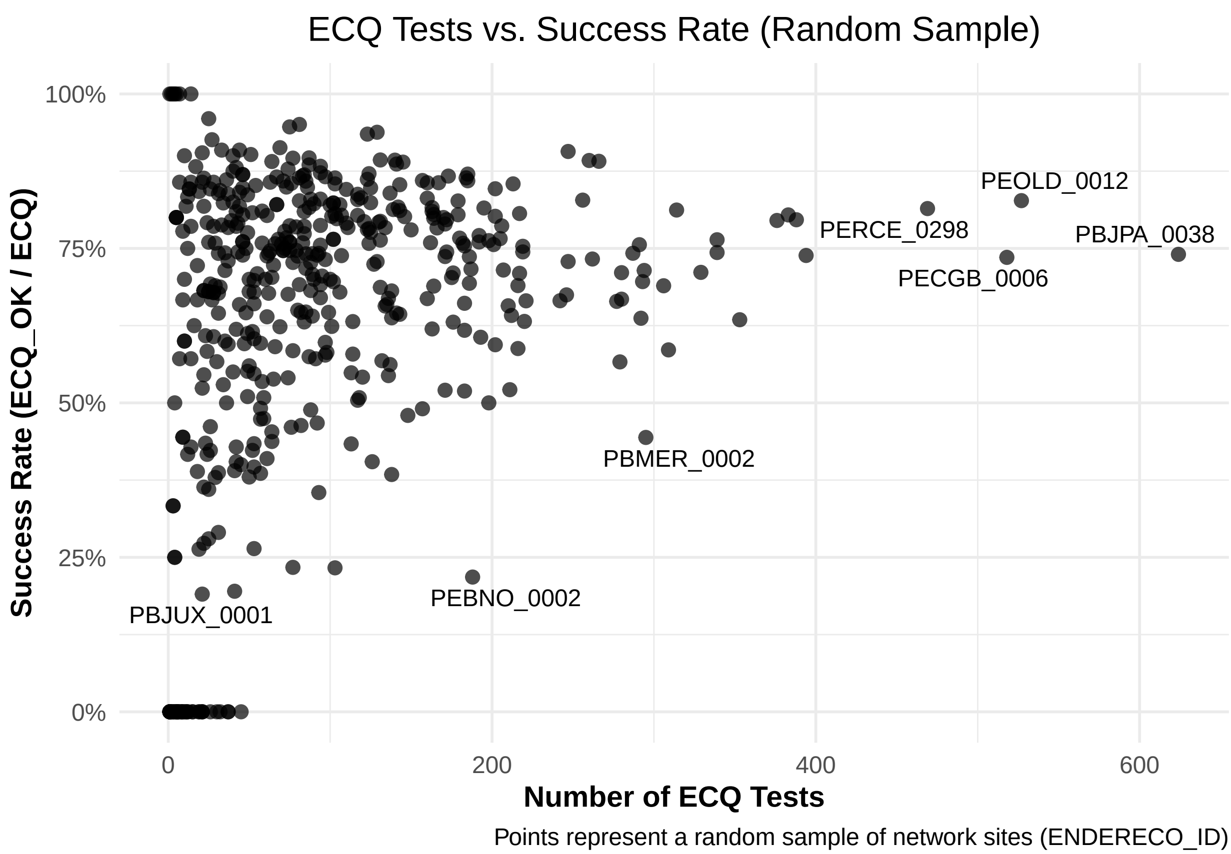
By formalizing this question, we set the stage for applying a hierarchical Bayes model that "borrows strength" across sites and municipalities, producing posterior distributions for each site's success probability. These posteriors underpin credible intervals and ranking metrics that guide robust, data-informed investment decisions.

Data

In this report, we analyze ECQ test data collected in October 2024 across Brazil's Northeast region, encompassing 8 ANFs. Each data point corresponds to a specific network site, identified by its unique ENDERECO_ID. For every site, we have recorded the total number of ECQ tests conducted and the number of successful tests (TESTES_ECQ_OK), indicating instances where the network met the stringent performance thresholds defined by the ECQ metric.

Below is a summary of the data we will be using in our analysis.

```
## tibble [958 x 6] (S3: tbl_df/tbl/data.frame)
##  $ group_id     : int [1:958] 101 103 93 106 104 107 95 96 94 97 ...
##  $ ANF          : chr [1:958] "83" "83" "83" "83" ...
##  $ MUNICIPIO    : chr [1:958] "ARACAGI" "ARARUNA" "AGUA BRANCA" "AREIAL" ...
##  $ ENDERECO_ID  : chr [1:958] "PBAAG_0001" "PBAAN_0001" "PBABW_0001" "PBAEA_0001" ...
##  $ TESTES_ECQ_OK: num [1:958] 9 27 12 10 80 0 19 5 26 0 ...
##  $ TESTES_ECQ   : num [1:958] 13 57 28 12 150 1 21 12 31 12 ...
```



Bayesian Model Structure

Model structure definition in JAGS.

```
model_string <- "
model {
  # Hiperparâmetros globais
  mu_global ~ dbeta(3, 3)
  sigma_global ~ dgamma(2, 0.5)  # Prior Gamma mais informativa

  # Parâmetros ANF com restrição
  alpha_anf <- mu_global * sigma_global
  beta_anf <- (1 - mu_global) * sigma_global
  mu_anf ~ dbeta(alpha_anf, beta_anf)

  # Priors para dispersão
  phi_municipio ~ dgamma(2, 0.9)  # Gamma mais suave
  phi_site ~ dgamma(2, 2)

  # Loop por municípios
  for(g in 1:N_group) {
    a_municipio[g] <- mu_anf * phi_municipio
    b_municipio[g] <- (1 - mu_anf) * phi_municipio
    mu_municipio[g] ~ dbeta(a_municipio[g], b_municipio[g])
  }

  # Loop por sites
  for(s in 1:N_sites) {
    logit_mu_site[s] <- logit(mu_municipio[group_per_site[s]])
    theta_site[s] <- ilogit(logit_mu_site[s] + epsilon[s])
    epsilon[s] ~ dnorm(0, 1/phi_site)

    n_success[s] ~ dbin(theta_site[s], n_tests[s])
  }
}
"
```

Conclusions

The application of the Bayesian hierarchical model enabled the identification of sites with performance significantly below average, highlighting priority areas for intervention. The consideration of credible intervals in the estimates reinforces the need for actions based on robust statistical evidence, aiming for the continuous improvement of network quality.

This approach allows for a nuanced understanding of performance variability across different sites and municipalities, ensuring that interventions are targeted where they are most needed. By quantifying uncertainty through credible intervals, decision-makers can assess the reliability of the estimates and prioritize actions with greater confidence.

In summary, the Bayesian hierarchical model provides a comprehensive framework for identifying underperforming areas and supports evidence-based decision-making to enhance overall network quality.

Table 1: Amostra de Sites Prioritários

	Identificação		Site — HDI			Município — HDI			Testes			Impacto / Pri.	
ANF	**City**	**Site**	**HDI Inf.**	**Avg.**	**HDI Sup.**	**HDI Inf**	**Avg**	**HDI Sup**	**Test**	**Succ**	**Fail**	**Impact**	**Prio**
81	BONITO	PEBNO_0002	0.16	0.21	0.27	0.05	0.16	0.28	188	41	147	0.61	1
83	MONTEIRO	PBMER_0002	0.39	0.44	0.50	0.20	0.42	0.65	295	131	164	0.48	2
83	ITAPORANGA	PBIRN_0005	0.43	0.49	0.55	0.26	0.45	0.67	252	124	128	0.34	3
83	CAMPINA GRANDE	PBCGE_0007	0.55	0.60	0.65	0.62	0.67	0.72	377	225	152	0.32	4
81	CUPIRA	PECUP_0003	0.16	0.24	0.32	0.09	0.29	0.51	103	24	79	0.32	5
81	JABOATAO DOS GUARARAPES	PEJBO_0007	0.65	0.69	0.72	0.70	0.74	0.77	607	417	190	0.31	6
83	JOAO PESSOA	PBJPA_0047	0.47	0.53	0.59	0.72	0.74	0.76	254	131	123	0.31	7
81	JABOATAO DOS GUARARAPES	PEJBO_0054	0.59	0.63	0.68	0.70	0.74	0.77	393	247	146	0.28	8
83	SUME	PBSUE_0002	0.35	0.42	0.50	0.28	0.47	0.66	160	67	93	0.28	9
81	SURUBIM	PESUU_0005	0.40	0.47	0.54	0.34	0.48	0.62	184	86	98	0.27	10
83	CATOLE DO ROCHA	PBCRH_0002	0.31	0.39	0.47	0.28	0.47	0.65	138	53	85	0.27	11
81	RECIFE	PERCE_0097	0.54	0.60	0.65	0.79	0.80	0.82	309	181	128	0.27	12
81	PAULISTA	PEPUI_0017	0.52	0.57	0.63	0.69	0.74	0.80	279	158	121	0.27	13
81	LIMOEIRO	PELIO_0004	0.41	0.48	0.55	0.28	0.42	0.57	188	91	97	0.27	14
81	RECIFE	PERCE_0150	0.45	0.52	0.59	0.79	0.80	0.82	198	99	99	0.25	15
81	PALMARES	PEPLS_0001	0.47	0.53	0.60	0.59	0.73	0.86	211	110	101	0.25	16
81	RECIFE	PERCE_0322	0.59	0.64	0.69	0.79	0.80	0.82	353	224	129	0.24	17
81	PRIMAVERA	PEPVE_0001	0.15	0.24	0.33	0.09	0.29	0.52	77	18	59	0.24	18
83	BAIA DA TRAICAO	PBBAI_0001	0.00	0.03	0.08	0.00	0.05	0.15	46	0	46	0.24	19
81	LIMOEIRO	PELIO_0003	0.32	0.41	0.49	0.28	0.42	0.57	126	51	75	0.23	20
83	JOAO PESSOA	PBJPA_0038	0.71	0.74	0.77	0.72	0.74	0.76	624	462	162	0.22	21
81	IPOJUCA	PEIPJ_0002	0.38	0.46	0.54	0.63	0.69	0.75	136	59	77	0.22	22
81	MACHADOS	PEMHO_0001	0.00	0.03	0.08	0.00	0.05	0.15	43	0	43	0.22	23
81	AGRESTINA	PEAGE_0002	0.03	0.08	0.15	0.09	0.23	0.39	45	0	45	0.22	24
81	OLINDA	PEOLD_0039	0.46	0.53	0.60	0.70	0.75	0.79	183	95	88	0.22	25
81	SURUBIM	PESUU_0003	0.40	0.48	0.55	0.34	0.48	0.62	148	71	77	0.21	26
83	JOAO PESSOA	PBJPA_0124	0.58	0.63	0.69	0.72	0.74	0.76	294	185	109	0.21	27
81	ABREU E LIMA	PEABU_0001	0.43	0.51	0.58	0.56	0.69	0.80	157	77	80	0.21	28
83	MARI	PBMAI_0001	0.45	0.52	0.60	0.28	0.52	0.75	171	89	82	0.21	29
81	RECIFE	PERCE_0073	0.62	0.67	0.72	0.79	0.80	0.82	352	234	118	0.20	30
83	JOAO PESSOA	PBJPA_0073	0.59	0.64	0.69	0.72	0.74	0.76	292	186	106	0.20	31
81	AGRESTINA	PEAGE_0001	0.32	0.41	0.49	0.09	0.23	0.39	113	49	64	0.20	32
81	OLINDA	PERCE_0018	0.54	0.60	0.66	0.70	0.75	0.79	230	136	94	0.20	33
81	VICENCIA	PEVCI_0002	0.23	0.33	0.43	0.15	0.37	0.61	83	27	56	0.20	34
81	JOAO ALFREDO	PEJFD_0003	0.02	0.09	0.16	0.09	0.23	0.40	41	0	41	0.20	35