THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

Likely Lies:
A Statistical Analysis of the Prevalence of Modern Forgeries

Katherine Jones
B.S. Candidate in Statistics

Introduction:

The authenticity of newly discovered artifacts is critically important. Scholars examine inscriptions for linguistic consistency, carbon dates, and intentions, among other factors. However, it can often be difficult to determine authenticity. This raises problems within the antiquities community. How ought media report on potentially forged artifacts? Are a few expert opinions enough to claim legitimacy? Which factors should be most heavily weighed when determining validity?

Objective:

This project seeks to establish an objective, statistical model for determining the authenticity of textual artifacts. A model is produced which analyzes qualities of an inscription, considers evidence against its legitimacy, and predicts the probability that it is forged.

Methods:

Data of sample size n=48 case studies was compiled, 25 forgeries and 23 non-forgeries. Eight criteria were considered:

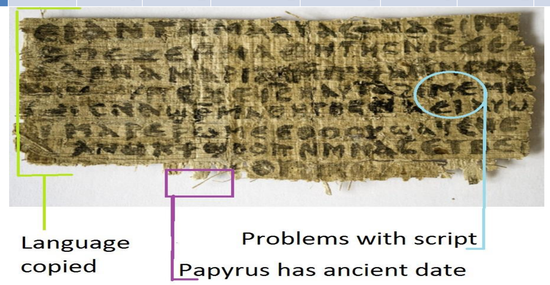
1. Date (C-14 or via textual analysis)
2. Linguistic inconsistency
3. Authenticity of script
4. Provenance
5. Location of purchase
6. Monetary motives for forgery
7. Social motives for forgery
8. Far-fetched or sensationalist content

Each criterion was represented as a binary variable, indicating evidence of forgery. A logit model was established, weighing each factor to produce the odds that an artifact was forged. The model output is the logarithm of the odds that an inscription is forged. The coefficients of the criteria are the factors by which they increased the odds.

Example of assigned binary values:

The Gospel of Jesus' Wife

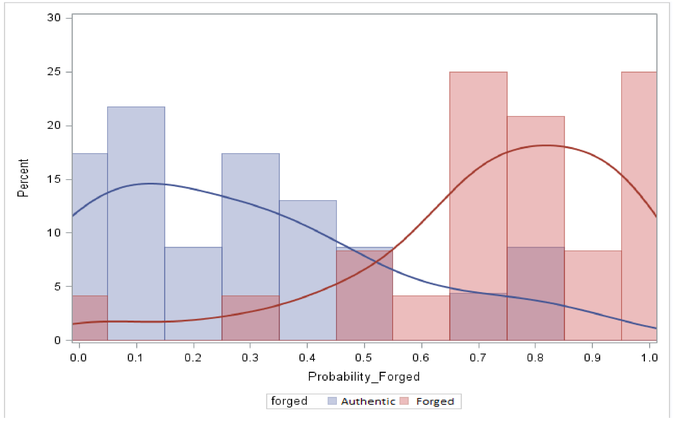
	Date	Lang-uage	Script	Prove-nance	Market	Motive ($)	Motive (social)	Sensa-tional
Flagged?	0	0	1	1	1	0	1	1



Increased odds from each variable

Variable:	Coefficient:	Factor of odds increase:	p- value
(intercept)	-5.27	n/a	.00204 **
Date	2.39	10.9 x	.02834 *
Script	2.20	9.03 x	.02183 *
Language	.26	1.30 x	.78412
Provenance	2.06	7.85 x	.03686 *
Motive ($)	1.08	2.94 x	.25639
Motive (social)	2.47	11.8 x	.02101 *
Sensational	1.80	6.05 x	.04602 *

Probabilities returned by model



The distribution of probabilities returned for authentic artifacts (blue) is centered below .5, and the distribution of probabilities returned for forged artifacts (red) is centered above .5.

Results:

- A simple logit model with dependent variable "log odds of forgery" and independent variables "date", "language", "script", "provenance", "monetary motive", "social motive", and "sensationalism" was found significant.
- Five coefficients ("date", "script", "provenance", "social motive", "sensationalism") and the intercept were statistically significant (α=.05).
- The variables which increased the odds most were "social motive", "date" and "script", with "provenance" and "sensationalism" also producing large impacts.
- A confusion matrix to test the model found an 81% success rate within the data set. Four forgeries and five non-forgeries were predicted incorrectly, while thirty-nine inscriptions were correctly classified.

Confusion Matrix

Model predicted status	True status: Forged	True status: Not
Forged	21	5
Not	4	18

Conclusions:

The model implies that poor script, incorrect date, and forger's motives are most heavily weighted in determining whether an artifact is legitimate. Poor provenance and sensational content also contribute. Each of these factors had a statistically significant impact on odds (α=.05).

This objective model could prevent media outlets from breaking incorrect information, or scholars from weighing some factors over more relevant evidence. It also demonstrates which aspects of an artifact are most significant and should therefore be examined preliminarily, potentially saving time and money for researchers and epigraphers.

Source for Photo :https://www.flickr.com/photos/centralasian/8008636912/in/photolist-dcGmJL