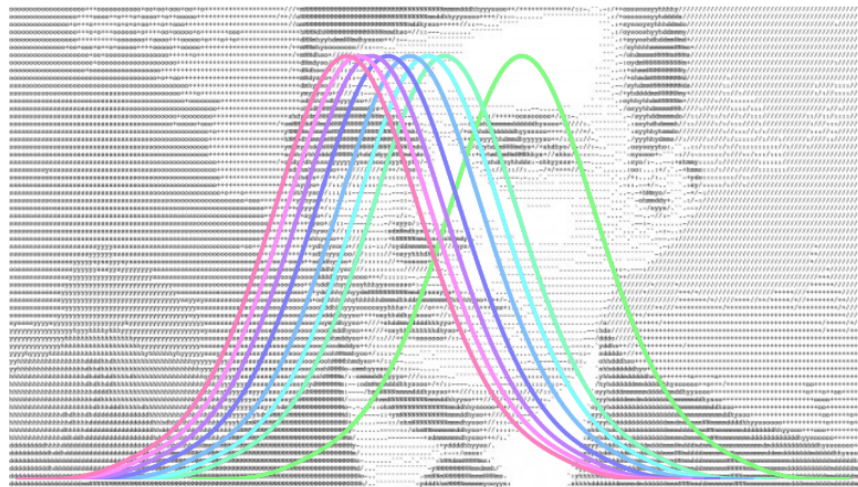# WOE, Log Odds, and Standard Errors

Quantifying uncertainty with WOE

Contents:

• WOE as centered log odds

• Logistic regression standard errors

• Uncertainty quantification with WOE

• FastWoe Python package

Weight of Evidence (WOE), Log Odds, and Standard Errors

Denis Burakov

June 2025

This document explains:

- How **WOE** is a centered version of **log odds**
- Why **WOE and log odds have the same standard error**
- That **subtracting the prior log odds (a constant) does not affect variance**
- Includes a Python simulation to demonstrate this as well as comparison with logistic regression

---

1. Definitions

Log odds for a binary feature group:

Let `n_11` = positives in group 1
Let `n_10` = negatives in group 1

$$\theta_1 = \log\left(\frac{n_{11}}{n_{10}}\right)$$

Global (prior) log odds:

Let `n_pos` = total positives, `n_neg` = total negatives:

$$\theta_{\text{prior}} = \log\left(\frac{n_{\text{pos}}}{n_{\text{neg}}}\right)$$

Weight of Evidence (WOE):

$$\text{WOE}_1 = \theta_1 - \theta_{\text{prior}}$$

---

2. Variance and Standard Error

> **Basic Property of Variance:** The variance of a random variable remains unchanged when a constant is subtracted from it.
>
> $$\mathrm{Var}(X - c) = \mathrm{Var}(X)$$
>
> This property is the cornerstone of our analysis. For more details on variance properties, see this discussion.

SE of log odds:

$$\mathrm{Var}(\theta_1) = \frac{1}{n_{11}} + \frac{1}{n_{10}} \Rightarrow \mathrm{SE}(\theta_1) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{10}}}$$

SE of WOE:

$$\mathrm{WOE}_1 = \theta_1 - \theta_{\mathrm{prior}} \Rightarrow \mathrm{Var}(\mathrm{WOE}_1) = \mathrm{Var}(\theta_1)$$

> **Why Standard Errors Are Identical**
>
> Because subtracting a constant does not change variance:
>
> $$\mathrm{Var}(X - c) = \mathrm{Var}(X)$$
>
> The prior log odds θ_{prior} is a **fixed constant** calculated from the entire dataset. When we subtract this constant from the log odds to obtain WOE, the variability (and hence standard error) remains exactly the same.

---

3. Difference of WOE = Coefficient in Logistic Regression

For two groups:

$$\beta = \theta_1 - \theta_0 = \mathrm{WOE}_1 - \mathrm{WOE}_0$$

$$\mathrm{Var}(\beta) = \mathrm{Var}(\mathrm{WOE}_1) + \mathrm{Var}(\mathrm{WOE}_0)$$

---

4. Python Simulation Example

> **Empirical Verification**
>
> Below we provide a simulation for the effect of constant in the variance calculation. We create 10,000 iterations by sampling from a binomial distribution and calculate the variance of log-odds and WOE.

```python
import numpy as np
import pandas as pd

np.random.seed(42)

# True probabilities
p1 = 30 / (30 + 20)  # x=1 group
p0 = 10 / (10 + 40)  # x=0 group
prior_p = (30 + 10) / (30 + 10 + 20 + 40)

# Sample sizes
n1_total = 50  # x=1 group total
n0_total = 50  # x=0 group total

# Number of simulations
n_sim = 10_000

# Store results
log_odds_1_samples = []
woe_1_samples = []

for _ in range(n_sim):
    # Simulate binary outcomes for x=1 group
    y1 = np.random.binomial(1, p1, size=n1_total)
    n11 = y1.sum()  # y=1, x=1
    n10 = n1_total - n11  # y=0, x=1

    # Compute log odds
    if n11 > 0 and n10 > 0:
        log_odds_1 = np.log(n11 / n10)
        log_odds_prior = np.log(prior_p / (1 - prior_p))
        woe_1 = log_odds_1 - log_odds_prior

        log_odds_1_samples.append(log_odds_1)
        woe_1_samples.append(woe_1)
```

```
# Convert to arrays
log_odds_1_samples = np.array(log_odds_1_samples)
woe_1_samples = np.array(woe_1_samples)

# Compare empirical variances
variance_log_odds = np.var(log_odds_1_samples, ddof=1)
variance_woe = np.var(woe_1_samples, ddof=1)

print(f"Empirical Variance (Log Odds): {variance_log_odds:.4f}")
print(f"Empirical Variance (WOE): {variance_woe:.4f}")
print(f"Difference: {abs(variance_log_odds - variance_woe):.4f}")
```

The results confirm that adding or subtracting a constant does not change the variance:

- Empirical Variance (Log Odds): 0.0890
- Empirical Variance (WOE): 0.0890
- Difference: 0.0000

This means that the centering of log-odds through WOE transformation does not affect the standard error.

5. Logistic Regression Example

Connecting WOE to Logistic Regression

In this section, we provide examples from logistic regression to demonstrate that the standard error (SE) of the Weight of Evidence (WOE) is linked to the variability of log odds per row and is not influenced by the centering effect.

Data

We will use the following table:

	y0	y1	Row Sum
x0	10	30	40
x1	20	10	30
Total	30	40	70

From the Table above, we can derive the log odds for $P(y = 1|x = 0)$ and $P(y = 1|x = 1)$. The log odds are defined as:

$$\log \left(\frac{P(y = 1|x)}{P(y = 0|x)} \right)$$

These are:

4

$$\log\left(\frac{P(y=1|x=0)}{P(y=0|x=0)}\right) = \log\left(\frac{30/40}{10/40}\right) = \log\left(\frac{30}{10}\right) = \log(3) \approx 1.099$$

$$\log\left(\frac{P(y=1|x=1)}{P(y=0|x=1)}\right) = \log\left(\frac{10/30}{20/30}\right) = \log\left(\frac{10}{20}\right) = \log(0.5) \approx -0.693$$

Logistic Regression

We can convert this table to a Bernoulli format to use this data with a logistic regression model. In the Bernoulli format, each observation is represented as a pair (x_i, y_i), where x_i is the predictor variable and y_i is the binary response variable taking values $\{0, 1\}$.

This transformation allows us to apply logistic regression, which models the log odds of the response variable y being 1 given the predictor x.

The Maximum Likelihood Estimation (MLE) solution if we assume $x = 0$ to be the intercept in a logistic regression model is:

Parameter	Estimate	Std. Error	Wald Statistic	P-value	Lower CI	Upper CI
intercept (x=0)	1.0986	0.3651	3.0087	0.0026	0.3829	1.8143
beta (x=1)	−1.7918	0.5323	−3.3661	0.0008	−2.8350	−0.7485

Notice that the sign of beta (x=1) flips depending on what we consider to be the intercept in our model. The coefficient represents the difference in the final log-odds between the two conditions and thus changes sign when we switch the reference base from $x = 0$ to $x = 1$.

Looking at the Std. Error column, we can see that the SE for the intercept is 0.3651, while for the coefficient it is 0.5323.

If we switch the intercept to represent a condition $x = 1$ then the model becomes:

Parameter	Estimate	Std. Error	Wald Statistic	P-value	Lower CI	Upper CI
intercept (x=1)	−0.6931	0.3873	−1.7897	0.0735	−1.4522	0.0659
beta (x=0)	1.7918	0.5323	3.3661	0.0008	0.7485	2.8350

Here we see that the standard error for the intercept is 0.3873.

We can get the standard errors reported for the intercepts from our 2×2 table:

$$SE(x = 0) = \sqrt{\frac{1}{10} + \frac{1}{30}} = 0.3651$$

$$SE(x = 1) = \sqrt{\frac{1}{20} + \frac{1}{10}} = 0.3873$$

The beta standard error is then the square root of the pooled variance of the two rows in our contingency table:

$$\text{SE}(\beta) = \sqrt{\text{SE}(x = 0)^2 + \text{SE}(x = 1)^2} = \sqrt{0.3651^2 + 0.3873^2} = 0.5323$$

Weight of Evidence (WOE)

To calculate WOE, we will use a conditional probability table. It should be noted that WOE can also be calculated through the odds form of the Bayes theorem, which would require knowing only the odds of y = 1 in a bin and the overall odds, an interested reader can find more details in [1].

$p(x\|y)$	$y = 0$	$y = 1$	WOE
$x = 0$	0.333	0.750	0.811
$x = 1$	0.667	0.250	-0.981
Column total	1.000	1.000	$-$

The Weight of Evidence (WOE) values for $x = 0$ and $x = 1$ are:

$$\text{WOE}(x = 0) = \ln\left(\frac{0.750}{0.333}\right) = \ln(2.25) \approx 0.811$$

$$\text{WOE}(x = 1) = \ln\left(\frac{0.250}{0.667}\right) = \ln(0.375) \approx -0.981$$

Base log odds are defined as follows:

$$\text{Base log odds} = \ln\left(\frac{40}{30}\right) = \ln\left(\frac{4}{3}\right) \approx 0.288$$

WOE Confidence Intervals

If we want to get an upper logit confidence interval using WOE:

$$\begin{aligned}
95\% \ \text{CI}_{\text{WOE}(x=0)} &= \text{Base log odds} + (\text{WOE}(x = 0) \times 1.96 \times \text{SE}(x = 0)) \\
&= 0.288 + (0.811 \times 1.96 \times 0.3651) \\
&\approx 0.288 + 0.576 = 1.8143
\end{aligned}$$

$$\begin{aligned}
95\% \ \text{CI}_{\text{WOE}(x=1)} &= \text{Base log odds} + (\text{WOE}(x = 1) \times 1.96 \times \text{SE}(x = 1)) \\
&= 0.288 + (-0.981 \times 1.96 \times 0.3873) \\
&\approx 0.288 - 0.731 = 0.0659
\end{aligned}$$

Notice that the resulting values correspond to the intercept CIs from the logistic regression fit summaries.

6. Implications

By understanding the properties of standard errors for WOE-transformed variables, we can derive valuable insights about the log of the likelihood ratios in predictive modeling. Applying standard errors to WOE allows us to identify bins containing the most uncertainty due to sampling variability. This, in turn, can indicate potential limitations in our inferences and highlight areas where additional data might be needed to improve model reliability.

Additionally, standard errors can help in assessing the true effect of the WOE values (since large absolute WOEs can be misleading due to low sample sizes) as well as in constructing confidence intervals, which are crucial for interpreting the model's predictions and making informed decisions.

> **Key Takeaway:** The centering operation in WOE transformation does not affect the uncertainty (standard error) of the estimates, making WOE a reliable transformation for feature engineering in machine learning and statistical modeling.

7. Python Implementation

You can find the Python implementation in the FastWoe package, which provides a fast and efficient Python implementation of WOE encoding and inference.

FastWoe
Fast Weight of Evidence (WOE) encoding and inference

Alan Turing's approach to explanation

- 'Naive Bayes' classifier:

$$\frac{p(H_0|s_1,...,s_p)}{p(H_1|s_1,...,s_p)} = \prod_i \frac{p(s_i|H_0)}{p(s_i|H_1)} \times \frac{p(H_0)}{p(H_1)}.$$

- So

$$\log \frac{p(H_0|s_1,...,s_p)}{p(H_1|s_1,...,s_p)} = \sum_i w_i + \log \frac{p(H_0)}{p(H_1)},$$

where w_i is the 'weight of evidence' $\log \frac{p(s_i|H_0)}{p(s_i|H_1)}$ (Turing and Good)
- Weights of evidence are positive/negative if evidence s_i is for/against H_0
- (Can use w_i's as predictors in a logistic regression to improve their calibration)
- Multiplying by 10 and rounding helps explanation

Image credit: David Spiegelhalter, Making Algorithms Trustworthy: What Can Statistical Science Contribute to Transparency, Explanation and Validation? NeurIPS 2018

Figure 1: FastWoe

7. References

[1] Good, I.J. Probability and the Weighing of Evidence. London: Griffin, 1950.
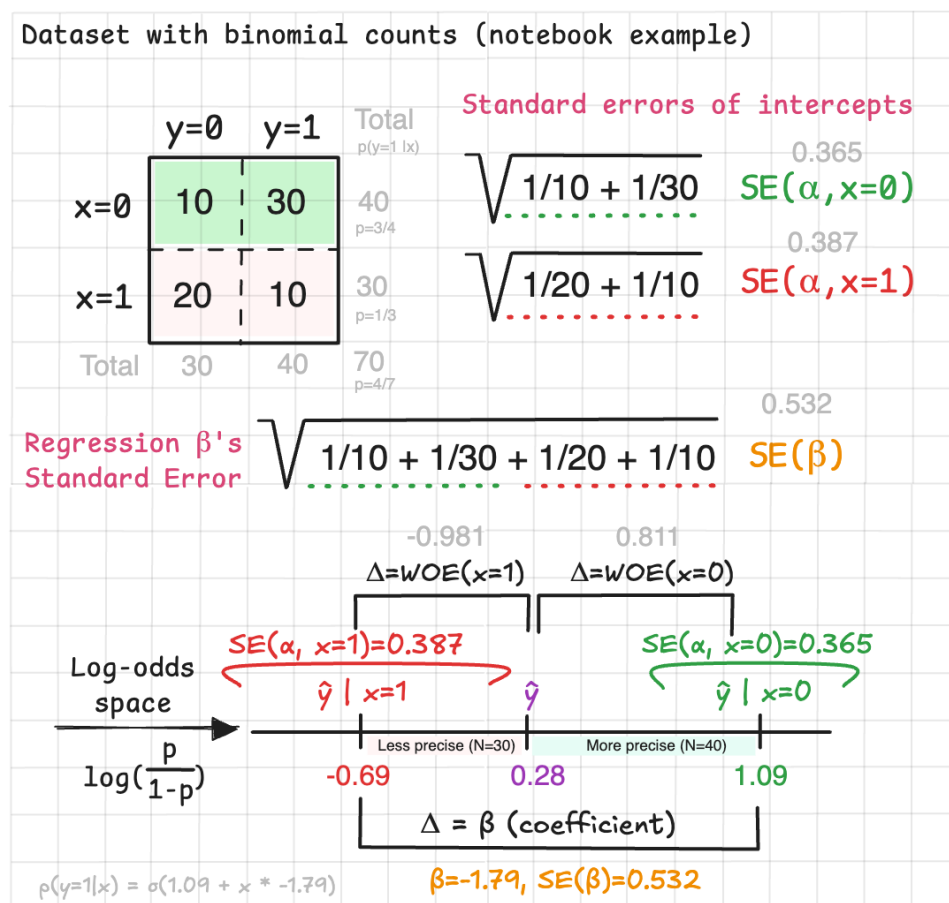
Appendix

Figure 2: A Notebook Sketch of WOE and Log Odds Standard Errors