

Replicating *The Bell Curve*

AFQT, SES, and Social Outcomes in NLSY79/97 and GSS

Troy Altus

2026-05-01

Table of contents

1 Abstract	1
2 Introduction	2
2.1 Research questions	2
3 Data	2
3.1 NLSY79	2
3.2 NLSY97	3
3.3 General Social Survey (GSS)	3
4 AFQT and SES Distribution	4
5 Poverty	5
6 Educational Attainment	5
7 Unemployment	6
8 Crime and Incarceration	7
9 GSS Cross-Sectional Extension (1972–2022)	8
10 Discussion	8
11 Conclusion	9
12 References	9

1 Abstract

Herrnstein and Murray’s 1994 *The Bell Curve* argued that cognitive ability, measured by AFQT scores from the National Longitudinal Survey of Youth 1979 (NLSY79), predicts a wide range of social outcomes — poverty, unemployment, illegitimacy, welfare dependency, crime — at least as strongly as socioeconomic background. This paper replicates their core analyses using the original NLSY79 cohort and extends the replication to NLSY97 and the General Social Survey

(GSS, 1972–2022). All analyses follow H&M’s logistic regression specification: binary outcome regressed on AFQT z-score and a composite SES index, with age and sex as covariates. Where H&M’s findings replicate, we note it; where they do not, we report the discrepancy and examine possible causes.

2 Introduction

Few social science books have provoked more sustained controversy than *The Bell Curve* (Herrnstein and Murray 1994). Its central empirical claim — that cognitive ability measured in early adulthood predicts life outcomes more robustly than family socioeconomic status — rested on a technically careful but highly contested analysis of NLSY79 data. Replication is straightforward in principle: the data are public, the method is described in the text and appendix, and the statistical tools are standard. In practice, several complications arise: variable coding has changed across NLSY data releases, the original extract is not publicly archived, and the political sensitivity of the topic has discouraged neutral replication attempts.

This paper proceeds without political agenda in either direction. The goal is to determine whether the key quantitative findings hold up under modern methods and newer data.

2.1 Research questions

1. Does AFQT predict poverty, unemployment, educational attainment, and crime after controlling for parental SES?
 2. Has the relative predictive power of cognitive ability vs. SES changed between the NLSY79 cohort (born 1957–64) and NLSY97 (born 1980–84)?
 3. Do results from WORDSUM (GSS vocabulary test) as an IQ proxy replicate the directional findings from AFQT?
-

3 Data

3.1 NLSY79

The Bureau of Labor Statistics National Longitudinal Survey of Youth, 1979 cohort, interviewed 12,686 respondents (born 1957–1964) annually from 1979 through 1994, then biennially through 2018. The AFQT (Armed Forces Qualification Test) was administered in 1980 to nearly the full sample.

H&M’s AFQT variable is the percentile score on the Armed Forces Qualification Test, sections 2 (arithmetic reasoning), 3 (word knowledge), 4 (paragraph comprehension), and half of section 5 (numerical operations). The BLS recomputed and revised the AFQT score in 1989; H&M used the revised score.

Data access: Custom extract via NLS Investigator (<https://www.nlsinfo.org/investigator>). Targeted variables listed in `src/acquire.py`. Targeted extract file size: ~5–15 MB CSV.

3.2 NLSY97

The NLSY97 cohort (born 1980–1984, $n \approx 8,984$) was interviewed annually from 1997 through 2011, then biennially. AFQT was administered in 1997–98. Used here as a robustness check: same design, 20-year-later cohort.

File size: ~10–25 MB CSV (targeted extract).

3.3 General Social Survey (GSS)

The GSS (NORC, University of Chicago) has fielded a nationally representative cross-sectional survey of U.S. adults nearly every year since 1972, with cumulative $n \approx 72,000$ through 2022. The WORDSUM variable — a 10-item vocabulary test — correlates ~0.7 with standard IQ measures and is the cognitive proxy H&M themselves validate against AFQT.

File size: Full cumulative Stata file download: ~200 MB (zip). After extracting to targeted variables: ~15–30 MB.

```
import sys
sys.path.insert(0, ".")

import pandas as pd
import numpy as np
import warnings
warnings.filterwarnings("ignore")

from pathlib import Path

DATA_READY = {
    "nlsy79": any(Path("data/raw/nlsy79").glob("*.csv")),
    "nlsy97": any(Path("data/raw/nlsy97").glob("*.csv")),
    "gss": any(Path("data/raw/gss").glob("*.dta")) or
           any(Path("data/raw/gss").glob("*.csv")),
}

print("Data availability:")
for k, v in DATA_READY.items():
    print(f" {k}: {' ' if v else ' NOT FOUND - see src/acquire.py'}")
```

Data availability:

```
nlsy79:
nlsy97:
gss:
```

```
from src.nlsy import load_nlsy79, load_nlsy97, cache as nlsy_cache
from src.gss import load_gss, cache as gss_cache

dfs = {}

if DATA_READY["nlsy79"]:
    dfs["nlsy79"] = load_nlsy79()
    nlsy_cache(dfs["nlsy79"], "nlsy79")
```

```

    print(f"NLSY79: {len(dfs['nlsy79']):,} respondents")

if DATA_READY["nlsy97"]:
    dfs["nlsy97"] = load_nlsy97()
    nlsy_cache(dfs["nlsy97"], "nlsy97")
    print(f"NLSY97: {len(dfs['nlsy97']):,} respondents")

if DATA_READY["gss"]:
    dfs["gss"] = load_gss()
    gss_cache(dfs["gss"])
    print(f"GSS: {len(dfs['gss']):,} respondent-years")

```

```

NLSY79: 12,686 respondents
NLSY97: 8,984 respondents
GSS: 75,699 respondent-years

```

4 AFQT and SES Distribution

Before replicating H&M's outcome models, we verify that the cognitive and SES measures are distributed as expected.

```

import plotly.express as px
import plotly.graph_objects as go

figs = []
for name, df in dfs.items():
    col = "afqt_z" if "afqt_z" in df.columns else "wordsum_z"
    if col not in df.columns:
        continue
    sub = df[col].dropna()
    figs.append(go.Histogram(x=sub, name=name, opacity=0.6, nbinsx=50))

if figs:
    fig = go.Figure(figs)
    fig.update_layout(
        barmode="overlay",
        title="Cognitive Measure Distribution",
        xaxis_title="Z-score",
        yaxis_title="Count",
        template="plotly_white",
    )
    fig.show()

```

Unable to display output for mime type(s): text/html

(a) AFQT z-score distribution by source

Unable to display output for mime type(s): text/html

(b)

Figure 1

5 Poverty

The Bell Curve, Chapter 5: Among whites ages 30–29 in the NLSY, the probability of being in poverty as a function of AFQT score, controlling for SES, shows a strong negative gradient. H&M report that moving from the 5th to the 95th AFQT percentile reduces poverty probability from ~26% to ~6%, holding SES at the median.

```
from src.models import quintile_table, logit_replicate
from src.plots import quintile_bar, prediction_curve

if "nlsy79" in dfs and "poor" in dfs["nlsy79"].columns:
    tbl = quintile_table(dfs["nlsy79"], "poor")
    display(tbl[["quintile", "rate_pct", "n"]].rename(
        columns={"rate_pct": "Poverty Rate (%)", "n": "N"}))
    fig = quintile_bar(tbl, "Poverty Rate by Cognitive Quintile (NLSY79)")
    fig.show()
```

Table 1: Poverty rate by cognitive quintile

	quintile	Poverty Rate (%)	N
0	Very Low	71.2	2089
1	Low	49.7	2123
2	Middle	38.1	2051
3	High	29.8	2039
4	Very High	21.9	2071

Unable to display output for mime type(s): text/html

```
if "nlsy79" in dfs and "poor" in dfs["nlsy79"].columns:
    result = logit_replicate(dfs["nlsy79"], "poor")
    print(result.summary_df.to_string())
    fig = prediction_curve(result, "Poverty Probability vs. AFQT (NLSY79)")
    fig.show()
```

	coef	OR	p	CI_lo	CI_hi
afqt_z	-0.776212	0.460146	4.207655e-229	0.438985	0.482326
ses_z	0.003967	1.003975	9.244570e-01	0.924941	1.089761

Unable to display output for mime type(s): text/html

Figure 2: Predicted poverty probability vs. AFQT, SES held at median

6 Educational Attainment 5

The Bell Curve, Chapter 6: H&M examine high school dropout as a binary outcome, arguing cognitive ability is a stronger predictor than SES.

```

if "nlsy79" in dfs:
    df79 = dfs["nlsy79"]
    if "educ_2018" in df79.columns:
        df79["dropout"] = (df79["educ_2018"] < 12).astype(float)
        df79.loc[df79["educ_2018"].isna(), "dropout"] = np.nan

if "nlsy79" in dfs and "dropout" in dfs["nlsy79"].columns:
    result = logit_replicate(dfs["nlsy79"], "dropout")
    print(result.summary_df.to_string())
    fig = prediction_curve(result, "High School Dropout Probability vs. AFQT (NLSY79)")
    fig.show()

```

	coef	OR	p	CI_lo	CI_hi
afqt_z	-1.605966	0.200696	1.223934e-279	0.183773	0.219177
ses_z	0.042729	1.043655	4.659312e-01	0.930408	1.170685

Unable to display output for mime type(s): text/html

Figure 3

7 Unemployment

```

for name in ["nlsy79", "gss"]:
    if name not in dfs:
        continue
    df = dfs[name]
    cog = "afqt_z" if "afqt_z" in df.columns else "wordsum_z"
    if "unemployed" not in df.columns or cog not in df.columns:
        continue
    result = logit_replicate(df, "unemployed", cognitive_col=cog)
    print(f"\n=== {name.upper()} ===")
    print(result.summary_df.to_string())
    fig = prediction_curve(result, f"Unemployment Probability vs. Cognitive Score ({name.upper()})")
    fig.show()

```

=== NLSY79 ===

	coef	OR	p	CI_lo	CI_hi
afqt_z	-0.265546	0.766787	2.891173e-33	0.734286	0.800726
ses_z	0.019808	1.020005	6.409227e-01	0.938541	1.108540

=== GSS ===

	coef	OR	p	CI_lo	CI_hi
wordsum_z	0.028249	1.028652	0.121331	0.992538	1.066079
ses_z	0.059330	1.061126	0.012468	1.012873	1.111678

Unable to display output for mime type(s): text/html

(a)

Unable to display output for mime type(s): text/html

(b)

Figure 4

8 Crime and Incarceration

The Bell Curve, Chapter 11: H&M report that AFQT strongly predicts self-reported incarceration history, with low-IQ males $\sim 5\times$ more likely to have been incarcerated than high-IQ males, holding SES constant.

```
for name in ["nlsy79", "gss"]:  
    if name not in dfs:  
        continue  
    df = dfs[name]  
    cog = "afqt_z" if "afqt_z" in df.columns else "wordsum_z"  
    out = "ever_arrested" if "ever_arrested" in df.columns else None  
    if not out or cog not in df.columns:  
        continue  
    result = logit_replicate(df, out, cognitive_col=cog)  
    print(f"\n=== {name.upper()} ===")  
    print(result.summary_df.to_string())  
    fig = prediction_curve(result, f"Arrest Probability vs. Cognitive Score ({name.upper()}")  
    fig.show()
```

=== NLSY79 ===

	coef	OR	p	CI_lo	CI_hi
afqt_z	-0.190954	0.826171	4.930885e-10	0.777935	0.877397
ses_z	-0.016574	0.983563	7.808833e-01	0.875158	1.105396

=== GSS ===

	coef	OR	p	CI_lo	CI_hi
wordsum_z	-0.093110	0.911093	0.008874	0.849722	0.976898
ses_z	-0.228664	0.795596	0.000003	0.722396	0.876214

Unable to display output for mime type(s): text/html

(a)

Unable to display output for mime type(s): text/html

(b)

Figure 5

9 GSS Cross-Sectional Extension (1972–2022)

The GSS allows a rough replication across five decades. We use WORDSUM as the cognitive proxy, replicate the poverty and unemployment models, and examine whether the AFQT/SES relative effect sizes have changed over time by estimating models on decade subsets.

```
if "gss" in dfs:
    gss = dfs["gss"]
    if "year" in gss.columns and "wordsum_z" in gss.columns:
        gss["decade"] = (gss["year"] // 10) * 10
        results = []
        for decade, sub in gss.groupby("decade"):
            for outcome in ["poor", "unemployed", "ever_arrested"]:
                if outcome not in sub.columns:
                    continue
                try:
                    r = logit_replicate(sub, outcome, cognitive_col="wordsum_z")
                    coef = r.summary_df.loc["wordsum_z", "OR"] if "wordsum_z" in r.summary_
                    results.append({"decade": decade, "outcome": outcome, "OR_wordsum": coef})
                except Exception:
                    pass
        if results:
            res_df = pd.DataFrame(results)
            import plotly.express as px
            fig = px.line(
                res_df, x="decade", y="OR_wordsum", color="outcome",
                title="WORDSUM Odds Ratio by Decade (GSS)",
                labels={"OR_wordsum": "Odds Ratio (cognitive)", "decade": "Decade"},
                template="plotly_white",
            )
            fig.add_hline(y=1, line_dash="dash", line_color="gray")
            fig.show()
```

Unable to display output for mime type(s): text/html

10 Discussion

[To be completed after data loaded and results reviewed.]

Key questions to address:

1. Do NLSY79 results replicate H&M's specific probability estimates within reasonable tolerance (± 5 percentage points)?
2. Is the AFQT coefficient larger in magnitude than the SES coefficient across all outcomes, as H&M claim?

3. Do NLSY97 and GSS results show attenuation or amplification of the cognitive-outcome gradient relative to NLSY79?
 4. What proportion of the racial gap in outcomes is explained by AFQT vs. SES?
-

11 Conclusion

[To be completed after analysis.]

12 References

Herrnstein, Richard J., and Charles Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. Free Press.