Sample Dissertation-Statistics Analysis Report

Effect of a Nurse-Led Educational Intervention on Medication-Adherence Scores in Adults With Type 2 Diabetes: A Pre/Post Controlled Design Analyzed by ANCOVA

Prepared by: Epsilon Solutions LLC — Dissertation & Research Statistics **Format:** Async, in writing, fixed-price · Columbia, South Carolina **Document type:** Portfolio sample (worked example) **Version:** Illustrative v1.0

> **⚠ Please read first — what this document is and is not**
>
> **This is a SAMPLE built on SYNTHETIC (computer-simulated) data.** No real patients, clinicians, sites, or datasets are involved. Every number below is produced by the reproducible simulation in the companion file `sample_analysis.R` (`set.seed(2024)`) and is included **for illustration only** — to show the structure, rigor, and reporting standard of a real Epsilon engagement.
>
> **Consulting is analysis and teaching — not ghostwriting.** In a real engagement, **you (the doctoral candidate) remain the sole author** of your dissertation. Epsilon designs the analysis, runs it reproducibly, defends the method in writing, builds your tables, and teaches you to explain and defend every result to your committee. We do not write your dissertation for you, and this report is a *template you learn from*, not text to paste in.
>
> **Numbers here are not findings about the real world.** They describe a made-up dataset. Do not cite them.

---

0. Engagement summary (what a client receives)

Deliverable	Included
Formal research questions & statistical hypotheses	Section 1
Analysis plan with the correct test *defended in writing*	Section 2
A priori power / sample-size justification	Section 3
Data screening & assumption checks (with fallbacks)	Section 4
Results with **APA 7th-edition tables** and correct statistical reporting	Section 5
Plain-language interpretation you can adapt for Chapter 4	Section 6
Honest limitations & threats to validity	Section 7
Fully reproducible analysis script (`sample_analysis.R`)	Companion file
Defense-prep: the questions your committee will actually ask	Section 8

Scenario. A DNP candidate evaluates whether a **nurse-led educational intervention** improves **medication adherence** in adults with **type 2 diabetes**, relative to usual care. Participants are assessed at **baseline (pre)** and at **12 weeks (post)**. The design is a **two-group pre/post (pretest–posttest control-group) design** with random assignment.

Primary outcome. Medication-adherence score on a **0–100 continuous scale** (higher = better adherence), adapted from a validated self-report adherence instrument (illustrative sample reliability, Cronbach's α = .84). *If your real instrument is the 8-item Morisky scale (MMAS-8, range 0–8) or a low/medium/high category, the appropriate analysis changes — see the sidebar in Section 2.*

Covariates. Baseline adherence (pretest), age (years), and highest education level (high school or less / some college / bachelor's or higher).

---

1. Research questions and hypotheses

Stated formally, with the **primary confirmatory** question pre-specified before data collection and separated from **secondary/estimation** aims. This separation is deliberate: testing many outcomes and reporting the winners manufactures significance from noise (the "garden of forking paths"). One primary test protects the Type I error rate.

RQ1 (primary, confirmatory). After controlling for baseline adherence, age, and education, do patients who receive the nurse-led intervention differ from usual-care patients in 12-week medication-adherence scores?

- **$H0_1$:** μ̂_intervention = μ̂_control (adjusted population means are equal)
- **$H1_1$:** μ̂_intervention ≠ μ̂_control (two-tailed; α = .05)

> A two-tailed test is used even though improvement is expected: it is the more conservative and more defensible choice, and it allows the analysis to detect an unintended *harmful* effect (e.g., education that overwhelms and reduces adherence).

RQ2 (secondary, estimation). *How large* is the adjusted intervention effect, and is it **clinically meaningful**? A minimal clinically important difference (**MCID**) of **5 points** on the 0–100 scale was defined a priori (≈ 0.4 SD of the baseline distribution, consistent with distribution-based MCID conventions and the clinical

team's judgment). RQ2 is answered with an **effect size and a confidence interval**, not a p-value — because a p-value tells you whether an effect is distinguishable from zero, not whether it is big enough to matter.

RQ3 (assumption/context). Is baseline adherence a strong enough predictor of follow-up adherence to justify its use as a covariate? (Answered in Sections 4–5.)

---

2. Analysis plan — the right test, defended

2.1 Recommended primary analysis: one-way ANCOVA

Model. A general linear model (ANCOVA) with:

- **Factor of interest:** Group (intervention vs. control)
- **Continuous covariates:** Baseline adherence, Age
- **Categorical covariate:** Education (3 levels)
- **Dependent variable:** 12-week adherence score

```
post_adherence ~ group + baseline_adherence + age + education
```

2.2 Why ANCOVA — and not the alternatives

This is the single most important methods decision in a pre/post controlled study, and committees ask about it. The recommendation rests on a direct comparison:

Candidate analysis	What it does	Verdict for this design
ANCOVA on posttest, adjusting for pretest	Compares groups on follow-up scores, holding baseline constant	✅ **Recommended.** Most powerful and least biased for a randomized pre/post design (Vickers & Altman, 2001). Adjusts for chance baseline imbalance and removes baseline variance from the error term, increasing power.
Independent-samples *t*-test on posttest only	Ignores baseline entirely	❌ Discards the pretest information; lower power; vulnerable to any baseline imbalance.
t-test on change scores (post − pre)	Compares gains	⚠ Unbiased under randomization but **less powerful** than ANCOVA and prone to regression-to-the-mean artifacts when baseline differs.
Repeated-measures / mixed ANOVA (time × group)	Tests the interaction	⚠ The time × group interaction answers a related question but is generally **less efficient** than ANCOVA for a single follow-up, and its sphericity/covariance assumptions add complexity. Reserve for ≥ 3 time points.

> **Committee-ready one-liner:** *"With random assignment and a single follow-up, ANCOVA on the posttest with the pretest as covariate is the recommended analysis (Vickers & Altman, 2001): it is more powerful than a change-score or posttest-only comparison and it corrects for any chance baseline difference between groups."*

2.3 Assumptions this model requires (checked in Section 4)

1. **Independence** of observations (design-based)
2. **Linearity** between each continuous covariate and the outcome
3. **Homogeneity of regression slopes** — the covariate–outcome relationship is the same in both groups (the defining ANCOVA assumption)
4. **Normality of residuals**
5. **Homogeneity of variance** (equal residual variance across groups)
6. **Covariate measured before/independently of the treatment and with good reliability** (baseline is measured pre-randomization, so this holds by design)

> **🔎 Sidebar — "What if my outcome isn't continuous?"**
>
> ANCOVA assumes an interval-level, roughly continuous outcome. If your adherence measure is: - **MMAS-8 (0–8) treated as ordinal** or **low/medium/high** → use **ordinal logistic regression** (proportional-odds) with baseline category and covariates, or a **Mann–Whitney U** on change if covariate adjustment isn't required. - **Binary "adherent / non-adherent"** → use **logistic regression** adjusting for baseline status and covariates; report adjusted odds ratios with 95% CIs. - **A count (e.g., missed doses)** → use **Poisson or negative-binomial regression** (negative binomial if overdispersed).
>
> The design logic (adjust for baseline + covariates, compare groups) is the same; only the link function changes. Tell us your exact instrument and we match the model to it.

---

3. A priori power analysis (sample-size justification)

Power was determined **before data collection** — the only time a power analysis is meaningful. (A "post hoc" power calculation computed from the observed effect is uninformative and is a known reviewer red flag; we do not report one.)

Parameter	Value	Rationale
Test	ANCOVA, fixed effects; group contrast (numerator *df* = 1)	Matches the primary analysis
Effect size	**Cohen's *f* = 0.25** (medium) ⇔ *d* = 0.50 ⇔ partial η² = .059	Smallest effect judged worth detecting; a medium standardized effect is a defensible, literature-consistent target for behavioral interventions (Cohen, 1988)
α (two-tailed)	**.05**	Conventional Type I error rate
Power (1 − β)	**.80**	Conventional; 20% false-negative risk
Covariates	3 (baseline, age, education)	As modeled

Result. Required **total N = 128 (n = 64 per group)** to achieve 80% power (computed with G*Power 3 / the `pwr` package; Faul et al., 2007). The companion script reproduces this.

Planned vs. analyzed. The illustrative study **enrolled 128** and **retained 120 (93.8%)** at 12 weeks; the primary analysis uses these **N = 120** complete cases (60 per group). Because retention exceeded 90% and loss was balanced across groups, complete-case analysis is reasonable here; a sensitivity analysis using multiple imputation would be added if attrition were higher or differential.

> **Note on the covariate bonus.** Baseline adherence correlates strongly with follow-up (here *r* = .77). The covariate explains about $r^2 \approx 60\%$ of the outcome variance, shrinking the ANCOVA error term to roughly $1 - r^2 \approx 40\%$ of the unadjusted variance, so the *realized* power at N = 120 comfortably exceeds .80 for the targeted effect. We sized to the conservative two-group requirement rather than banking on the covariate — under-powering is far costlier than a modest over-enrollment.

4. Data screening and assumption checks

All checks below are on the synthetic dataset and are reported exactly as they would be for a real client. Each check names the fallback if the assumption fails.

4.1 Preliminary — randomization / baseline equivalence

Confirming groups are comparable at baseline (expected under randomization; reassures the committee that any follow-up difference is attributable to the intervention, not pre-existing differences).

Baseline variable	Test	Result	Interpretation
Baseline adherence	Independent *t*	$t(118) = 0.46$, $p = .645$	Equivalent
Age	Independent *t*	$t(118) = 1.10$, $p = .272$	Equivalent
Education	χ^2	$\chi^2(2) = 0.57$, $p = .753$	Equivalent

No baseline imbalance. (Note: baseline balance tests are descriptive reassurance; ANCOVA adjusts for any residual imbalance regardless.)

4.2 The six ANCOVA assumptions

#	Assumption	How checked	Result (synthetic)	Verdict / fallback
1	Independence	Design: participants independent, one observation each	Met by design	—
2	Linearity (covariate ↔ outcome)	Scatterplot + partial-regression plots	Linear; baseline–post *r* = .77	Met. *Fallback:* transform or add polynomial term
3	**Homogeneity of regression slopes**	Test group × baseline interaction	$F(1, 113) = 1.08$, $p = .302$	**Met** (slopes parallel). *Fallback:* if $p < .05$, ANCOVA is invalid → report a moderation model (report the interaction) or use Johnson–Neyman regions
4	Normality of residuals	Shapiro–Wilk + Q–Q plot on model residuals	$W = 0.99$, $p = .609$	Met. *Fallback:* ANCOVA is robust at N = 120 (CLT); if severe, use bootstrapped/robust ANCOVA (`WRS2`)
5	Homogeneity of variance	Levene's test (median-centered)	$F(1, 118) = 0.46$, $p = .497$	Met. *Fallback:* heteroscedasticity-consistent (HC3) SEs or Welch-type robust ANCOVA
6	Reliable, pre-treatment covariate	Instrument reliability; timing	α = .84; baseline measured pre-randomization	Met (measurement error in the covariate biases ANCOVA; good reliability protects this)

Outliers / influence. One standardized residual marginally exceeded ±3 (max = 3.08), and no observation approached the conventional Cook's distance cutoff of 1.0 (max D = 0.12), so no case was unduly influential. (Eight cases exceeded the more liberal $4/N$ screening threshold — expected in a sample this size and not a concern by the $D > 1$ criterion.) **Missing data.** 8 of 128 (6.2%) lost to follow-up, balanced across arms (see §3).

> **Bottom line:** every assumption is satisfied, so the standard ANCOVA is appropriate and no fallback is triggered. In a real engagement we run and *show* each of these checks so the method is defensible under questioning — not assumed.

5. Results

> All values are **illustrative**, generated by `sample_analysis.R` (`set.seed(2024)`). Reported in APA 7th-edition style. Adherence is scored 0–100 (higher = better).

5.1 Descriptive statistics

Table 1 *Participant Characteristics and Adherence Scores by Group*

Characteristic	Control (n = 60)	Intervention (n = 60)	Baseline equivalence
Age (years), *M* (*SD*)	58.7 (8.5)	56.9 (9.5)	$t(118) = 1.10$, $p = .272$
Education, *n* (%)			$\chi^2(2) = 0.57$, $p = .753$
High school or less	20 (33.3)	22 (36.7)	
Some college	25 (41.7)	21 (35.0)	
Bachelor's or higher	15 (25.0)	17 (28.3)	
Baseline adherence, *M* (*SD*)	62.40 (12.52)	61.39 (11.41)	$t(118) = 0.46$, $p = .645$
12-week adherence, *M* (*SD*)	63.62 (11.07)	66.75 (11.72)	—
Change (post − pre), *M* (*SD*)	1.22 (7.58)	5.36 (7.74)	—

Note. Adherence measured on a 0–100 scale (higher = better adherence). Baseline equivalence tests confirm comparable groups at randomization. Over 12 weeks the control group was essentially unchanged (+1.2 points), whereas the intervention group improved by roughly 5.4 points on average.

5.2 Primary analysis — ANCOVA

Table 2 *Analysis of Covariance for 12-Week Medication-Adherence Score (Type III Sums of Squares)*

Source	*SS*	*df*	*MS*	*F*	*p*	partial η²
Baseline adherence (covariate)	9262.31	1	9262.31	186.06	< .001	.620
Age (covariate)	5.59	1	5.59	0.11	.738	.001
Education (covariate)	149.56	2	74.78	1.50	.227	.026
Group (intervention vs. control)	**433.41**	**1**	**433.41**	**8.71**	**.004**	**.071**
Error	5675.17	114	49.78			
Corrected total	15615.67	119				

Note. $R^2 = .637$ (adjusted $R^2 = .621$). Partial η² benchmarks: .01 = small, .06 = medium, .14 = large (Cohen, 1988). Type III sums of squares (correct for a design with covariates); categorical terms coded with sum-to-zero contrasts.

What the table says. After adjusting for baseline adherence, age, and education, group had a **statistically significant effect** on 12-week adherence, $F(1, 114) = 8.71$, $p = .004$, partial η² = .071 (a **medium** effect). As expected, **baseline adherence** was the dominant predictor (partial η² = .620) — which is exactly why adjusting for it sharpens the group comparison. Age and education were not significant covariates.

5.3 Adjusted (estimated marginal) means

Table 3 *Adjusted Means and the Group Contrast, Controlling for Baseline Adherence, Age, and Education*

Group	Adjusted *M*	*SE*	95% CI
Control	63.38	0.92	[61.55, 65.21]
Intervention	67.22	0.92	[65.40, 69.04]
Difference (Intervention − Control)	**3.84**	**1.30**	**[1.26, 6.41]**

Note. Means estimated at the sample mean of each covariate. Contrast: $t(114) = 2.95$, $p = .004$, **Cohen's *d* = 0.54, 95% CI [0.17, 0.92]**.

> **Teaching point (why the adjusted difference ≠ the raw difference).** The *raw* posttest gap is 66.75 − 63.62 ≈ 3.1 points, but the *adjusted* gap is **3.84** points. ANCOVA gives the intervention group "credit" for having started slightly *lower* at baseline (61.39 vs. 62.40): once you put both groups on an equal baseline footing, the estimated effect is a little larger than the raw comparison suggests. This is precisely the correction ANCOVA exists to make — and a question a sharp committee member may raise.

5.4 Effect size and clinical significance (RQ2)

- **Adjusted mean difference: 3.84 points (95% CI [1.26, 6.41]).**
- **Standardized effect: Cohen's *d* = 0.54** (medium), 95% CI [0.17, 0.92]; **partial η² = .071** (medium), 95% CI [.01, .18].
- **Against the pre-specified MCID of 5 points:** the **point estimate (3.84) is below the MCID**, but the **confidence interval [1.26, 6.41] spans it** — the data are compatible with effects from about 1 point (well below clinically important) up to about 6.4 points (above the MCID).

> **The honest read (and the reason a statistician earns their fee):** The intervention produced a **real, non-zero improvement** — the effect is distinguishable from chance (p = .004) and the CI excludes zero. But we **cannot claim with confidence that it reaches the pre-specified clinical threshold**, because the interval straddles it. *Statistical significance here is not the same as demonstrated clinical importance.* The correct conclusion is "promising and worth pursuing at larger scale," not "proven clinically important." A report that stopped at "$p < .05$, the intervention works" would over-claim.

---

6. Plain-language interpretation (adapt for Chapter 4)

> *Drop-in model paragraphs. Rewrite in your own voice — this is a scaffold to learn from, not text to submit verbatim. You are the author.*

Results paragraph (APA style). An analysis of covariance was conducted to compare 12-week medication-adherence scores between the nurse-led intervention and usual-care groups, controlling for baseline adherence, age, and education level. Preliminary checks supported the assumptions of the analysis, including homogeneity of regression slopes, $F(1, 113) = 1.08$, $p = .302$. After adjusting for the covariates, there was a statistically significant difference in adherence between groups, $F(1, 114) = 8.71$, $p = .004$, partial $\eta^2 = .071$, representing a medium effect. The adjusted mean adherence score was higher for the intervention group ($M = 67.22$, $SE = 0.92$) than for the control group ($M = 63.38$, $SE = 0.92$), a difference of 3.84 points, 95% CI [1.26, 6.41], $d = 0.54$. Baseline adherence was a strong predictor of follow-up adherence, partial $\eta^2 = .620$, whereas age and education were not significant covariates.

Interpretation paragraph. These results suggest that the nurse-led educational intervention improved medication adherence among adults with type 2 diabetes relative to usual care, over and above patients' starting adherence and demographic differences. The effect was of moderate magnitude and statistically reliable. However, because the confidence interval for the adjusted difference included values below the pre-specified 5-point threshold for clinical importance, the finding should be interpreted as encouraging evidence of benefit rather than definitive proof of a clinically decisive change. Replication in a larger, multi-site sample would tighten the estimate and clarify whether the average benefit reliably exceeds the clinical threshold.

One-sentence version (abstract / defense): *"Controlling for baseline adherence and demographics, the intervention produced a statistically significant, medium-sized improvement in adherence (adjusted difference ≈ 3.8 points, 95% CI [1.3, 6.4], d = 0.54, p = .004); the effect is real but its clinical magnitude is not yet certain."*

---

7. Limitations and threats to validity (state them before your committee does)

- **Single follow-up.** One post-measurement cannot speak to *durability*. Adherence gains often decay; a 6- or 12-month follow-up would address maintenance.
- **Self-reported outcome.** Self-report adherence is subject to social-desirability and recall bias; an objective anchor (pharmacy refill / proportion-of-days-covered, pill counts) would strengthen validity.
- **Clinical vs. statistical significance.** As shown, significance was achieved but the CI straddles the MCID — the practical size of the benefit remains uncertain.
- **Generalizability.** A single-site sample limits external validity; results may not transfer to other settings, languages, or comorbidity profiles.
- **No adjustment for clustering** (if patients were nested within nurses/clinics, a mixed-effects model with a random intercept would be required — tell us your recruitment structure).
- **Attrition.** Even at ~6%, loss to follow-up is a potential (here, low) source of bias; multiple-imputation sensitivity analysis is advisable if attrition is higher in your real data.

---

8. Defense-prep — questions your committee will likely ask

1. *"Why ANCOVA instead of a repeated-measures ANOVA or change scores?"* → See §2.2 (Vickers & Altman, 2001).
2. *"Did you check homogeneity of regression slopes?"* → Yes; the group × baseline interaction was non-significant, $F(1, 113) = 1.08$, $p = .302$ (§4.2).
3. *"Is your effect clinically meaningful, or just statistically significant?"* → Honest answer in §5.4 — significant, medium, but the CI straddles the MCID.
4. *"Why is your adjusted difference bigger than the raw difference?"* → Baseline imbalance correction (§5.3 teaching point).
5. *"Was the study adequately powered?"* → A priori N = 128 for a medium effect at 80% power; 120 retained (§3).
6. *"What did you do about the covariates that weren't significant?"* → They were pre-specified and retained for adjustment; covariate selection was not data-driven (avoids the forking-paths problem).

---

9. Reproducibility

Every number in this report is regenerated by the companion script:

- **`sample_analysis.R`** — simulates the dataset (`set.seed(2024)`, with the RNG pinned for cross-version reproducibility), runs the baseline-equivalence checks, tests all ANCOVA assumptions, fits the Type III ANCOVA, computes partial η² and Cohen's *d* with CIs, and prints Tables 1–3.

Run it top to bottom in R (≥ 4.1) after installing the listed packages. The figures above reproduce exactly at `set.seed(2024)` (only display rounding differs). The design, model, and conclusions are robust across seeds — the intervention effect is reliably a significant, medium-sized improvement.

References (methods)

American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.).

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191.

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science. *Frontiers in Psychology, 4*, 863.

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Vickers, A. J., & Altman, D. G. (2001). Analysing controlled trials with baseline and follow up measurements. *BMJ, 323*(7321), 1123–1124.

Epsilon Solutions LLC — dissertation & research statistics. We analyze, explain, and teach; you remain the author. This sample uses synthetic data for demonstration only.

Epsilon Solutions LLC · Columbia, SC · pdesai@epsilonsolutionsllc.com