Gov 2001 Final — Practice Exam (Extended)

Prof. Scott Cunningham, Spring 2026

About this practice

This packet contains **15 practice problems**, grouped into **five parts** of three. The real final will have only **4–5 problems**. So why are you seeing fifteen?

- Each part targets **one skill stack** you will need on the exam. The three problems in a part build up that stack: the first drills a single foundational move, the second combines two moves, the third puts the whole stack together on a full in-family problem.
- The exam will draw *one* problem from each family. You will not see the same numbers, the same distribution, or the same design matrix. But if you have practiced all three problems in a part, you will recognize the structure of the exam question and know which tools to pull out.
- The point of this packet is not to memorize answers. It is to invest in the skills — so that when you see a problem you have never seen before, you can recognize which family it belongs to and execute the steps.

Skills roadmap

Here is the full list of skills each part is designed to build:

Part	Skill stack	Practice problems
I	CLT for a sample mean; delta method for a smooth transformation; variance-stabilizing transformations that produce parameter-free CIs	1a, 1b, 1c
II	One-sided and two-sided CLT-based hypothesis tests; Chebyshev's inequality; comparing the width of a Chebyshev CI to the width of a CLT CI	2a, 2b, 2c
III	Distinguishing "exact" from "asymptotic"; distinguishing pre-data from post-data CI interpretation; Type I error vs power; the exact content of Gauss-Markov; robust variance estimators	3a, 3b, 3c
IV	Matrix OLS by hand: computing $X'X$, $(X'X)^{-1}$, $X'Y$, $\hat{\beta}$, $\hat{Y}$, residuals; applying FWL to isolate one coefficient via residualization	4a, 4b, 4c
V	Law of total variance and covariance; Bernoulli variance $p(1-p)$; using FWL to derive a variance-weighted average of within-group regression coefficients	5a, 5b, 5c

Each problem below carries a **Skill target** label naming the single skill you are practicing. Read the label before you start — it tells you what to invest in.

Mock final path

If you want to simulate the real exam, do the following **five problems closed-book and timed**, in this order, before consulting any solution:

Family	Mock problem	Why this one
I	Practice 1c	Variance-stabilizing transform; parameter-free CI
II	Practice 2c	Bernoulli, one-sided test, CLT *and* Chebyshev rejection regions
III	Practice 3a or 3b (pick one)	Conceptual T/F traps; the real exam draws from this category
IV	Practice 4c	No-intercept matrix OLS + FWL on a substantive control
V	Practice 5c	Variance-weighted decomposition with a binary control

Set a timer for 80 minutes. When you finish, then check the solutions. The other ten problems (a's and b's) are scaffolds — use them after the mock if you got something wrong, or before it if you need to warm up on a piece.

Triage if you have less than 8 hours. Do the c-problems first. The a's and b's exist to scaffold up to the c's; if you can already do a c-problem cleanly, skip its a and b. The single highest-leverage hour is Practice 5c (worth 25 points on the real exam and the densest skill stack).

Part I — Building up to an exam problem on CLT + Delta method

On the real exam, you may be asked to: given an iid sample from some distribution, find the asymptotic distribution of a transformation of the sample mean, and then build a confidence interval whose width does not depend on the unknown parameter. The three problems below walk you through that skill stack: CLT first, delta method second, variance-stabilizing transform third.

Practice 1a

Skill target. Apply the CLT to state the asymptotic distribution of a sample mean from a specific distribution, and report the standard error.

Let $X_1, \dots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$, with $\mathbb{E}[X_i] = \lambda$ and $\text{Var}(X_i) = \lambda$.

(a) State the asymptotic distribution of $\sqrt{n}(\bar{X}_n - \lambda)$ as $n \to \infty$. Which theorem are you invoking?

(b) Give the (asymptotic) standard error of $\bar{X}_n$ in terms of λ and n.

(c) Suppose $\lambda = 4$ and $n = 100$. Construct an approximate 95% confidence interval for λ using the plug-in standard error $\hat{\sigma}/\sqrt{n}$ with $\hat{\sigma}^2 = \bar{X}_n$.

Practice 1b

Skill target. Apply the delta method once to a smooth transformation of a sample mean.

Suppose $\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$, where $\mu \neq 0$ and σ^2 is known. Let $g(u) = u^2$.

(a) State the delta method. What regularity condition does g need to satisfy at μ?

(b) Derive the asymptotic distribution of $\sqrt{n}\big(g(\bar{X}_n) - g(\mu)\big)$.

(c) Suppose $\mu = 3$, $\sigma^2 = 4$, $n = 100$. Construct an approximate 95% CI for μ^2 using the delta-method variance (you may plug in $\mu = \bar{X}_n$ for the standard error).

Practice 1c

Skill target. Combine CLT and delta method with a variance-stabilizing transformation so that the asymptotic variance does not depend on the parameter. This is exactly the family of problem that appears on the exam.

Let $X_1, \dots, X_n \overset{iid}{\sim} \text{Poisson}(\lambda)$. Define the transformed statistic

$$\phi_n = 2\sqrt{\bar{X}_n}.$$

(a) State the asymptotic distribution of $\sqrt{n}(\bar{X}_n - \lambda)$.

(b) Consider the function $g(u) = 2\sqrt{u}$. Compute $g'(u)$ and evaluate it at $u = \lambda$.

(c) Apply the delta method to derive the asymptotic distribution of $\sqrt{n}(\phi_n - 2\sqrt{\lambda})$.

(d) Show that the asymptotic variance of $\sqrt{n}(\phi_n - 2\sqrt{\lambda})$ *does not depend on* λ.

(e) Construct an approximate 95% confidence interval for $2\sqrt{\lambda}$. Explain in one sentence why this CI is convenient: what does the fact that the variance is parameter-free buy you?

(f) Back out a 95% CI for λ itself by transforming the endpoints of your interval in (e).

Part II — Building up to an exam problem on CLT vs Chebyshev

On the real exam, you may be asked to conduct a hypothesis test and to construct confidence intervals using two different tools — the CLT (which assumes n is large enough for a normal approximation to be accurate) and Chebyshev's inequality (which works for any finite n but is looser). The three problems build to the comparison.

Practice 2a

Skill target. Conduct a one-sided, CLT-based hypothesis test for a Bernoulli proportion using the *plug-in* standard error $\sqrt{\hat{p}(1-\hat{p})/n}$. This is the SE convention the real exam expects — it is what makes CI–test duality go through.

A pollster surveys $n = 400$ voters. Let $X_i = 1$ if voter i supports candidate A, 0 otherwise. Treat $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$. Observed $\hat{p} = \bar{X}_n = 0.54$.

(a) State the null and alternative hypotheses for a one-sided test that support exceeds $1/2$.

(b) Compute the *plug-in* standard error of $\bar{X}_n$, using $\hat{p}(1-\hat{p})$ as the variance estimate. (Do *not* plug in $p_0 = 0.5$ here — that is a different test convention; see the note below.)

(c) Compute the z-statistic.

(d) At $\alpha = 0.05$, do you reject H_0? The one-sided critical value is $z_\alpha = 1.645$ (*not* $z_{\alpha/2} = 1.96$). Report the approximate p-value. (You may use $\Phi(1.605) \approx 0.946$.)

(e) **Note for your records.** Some textbooks use $\sqrt{p_0(1-p_0)/n}$ as the SE, justified by "compute the test statistic under the null." That is a valid Wald-style test, but it does *not* correspond to inverting the plug-in $(1-\alpha)$ CI to a rejection region. On this exam, use the plug-in SE so the CI and the rejection region are the same object.

Practice 2b

Skill target. Apply Chebyshev's inequality to a sample mean and compare the resulting CI half-width to the CLT half-width.

Let $X_1, \ldots, X_n \overset{iid}{\sim} F$ with $\mathbb{E}[X_i] = \mu$ and $\text{Var}(X_i) = \sigma^2$.

(a) State Chebyshev's inequality for the sample mean:

$$\mathbb{P}\big(|\bar{X}_n - \mu| \geq \epsilon\big) \leq ?$$

(b) Set the upper bound equal to $\alpha = 0.05$. Solve for ϵ in terms of σ, n, and α.

(c) This ϵ is the half-width of a (conservative) $(1-\alpha)$ Chebyshev CI for μ. Write down the CI.

(d) The CLT-based 95% half-width is $1.96 \cdot \sigma/\sqrt{n}$. Compute the ratio of the Chebyshev half-width to the CLT half-width. (Roughly, how much wider is Chebyshev?)

(e) In one sentence: why would anyone ever use Chebyshev when the CLT gives a tighter interval?

Practice 2c

Skill target. Full in-family problem mirroring the real exam. Bernoulli iid; *one-sided* test $H_0 : p \leq p_0$ vs. $H_1 : p > p_0$. Build a CLT CI *and* a Chebyshev CI for p, state both rejection regions, compare widths. Watch the z_α vs. $z_{\alpha/2}$ trap and the plug-in vs. Popoviciu trap.

Let $X_1, \ldots, X_n \overset{iid}{\sim} \text{Bernoulli}(p)$. Suppose $n = 200$ and observed $\hat{p} = \bar{X}_n = 0.58$. Test $H_0 : p \leq 0.5$ vs. $H_1 : p > 0.5$ at $\alpha = 0.05$.

(a) **CLT-based asymptotic** $(1-\alpha)$ **CI.** Construct the two-sided 95% CI for p using the plug-in SE $\sqrt{\hat{p}(1-\hat{p})/n}$. State the half-width.

(b) **CLT-based one-sided rejection region.** For the one-sided test $H_1 : p > 0.5$ at level α, the duality is with a one-sided $(1-\alpha)$ lower-confidence bound. Reject H_0 if $\hat{p} > p_0 + z_\alpha\sqrt{\hat{p}(1-\hat{p})/n}$, with $z_\alpha = z_{0.05} = 1.645$ (*not* $z_{\alpha/2} = 1.96$). At the observed $\hat{p} = 0.58$, do you reject?

(c) **Chebyshev-based conservative** $(1-\alpha)$ **CI.** Use Popoviciu's bound $\sigma^2 = p(1-p) \leq 1/4$. Set Chebyshev's bound equal to α to solve for the half-width ε, and write down the conservative two-sided 95% CI for p. (Do *not* plug in $\hat{p}(1-\hat{p})$ inside the Chebyshev bound — that defeats the distribution-free guarantee.)

(d) **Chebyshev-based rejection region.** Using the same ε, state the rejection region for the one-sided test: reject if $\hat{p} > p_0 + \varepsilon$. At $\hat{p} = 0.58$, do you reject? Compare your conclusion to part (b) and comment in one sentence.

(e) **Width comparison.** Compute the ratio of the Chebyshev half-width to the CLT half-width at $\alpha = 0.05$. Why is Chebyshev wider, in one sentence?

Part III — Building up to an exam problem on True/False

On the real exam, you will see a set of True/False items that probe whether you understand what the theorems and definitions in the course *actually* say — and do not say. The practice items below group the traps by topic. Each statement is a common misconception or a subtle wording issue you should recognize on sight. **Justify each answer in one sentence**; the exam will require justification.

Practice 3a — Probability, CLT, Chebyshev

Skill target. Distinguish "exact" from "asymptotic," distinguish pre-data from post-data CI interpretation, and recognize the scope of Chebyshev.

For each statement, circle True or False and write one sentence of justification.

1. If $X_1, \dots, X_n \overset{iid}{\sim} F$ with finite variance, then $\bar{X}_n$ is normally distributed for any finite n.
2. For any random variable X with $\mathrm{Var}(X) < \infty$, $\mathbb{P}(|X - \mathbb{E}[X]| \geq k\sqrt{\mathrm{Var}(X)}) \leq 1/k^2$ for all $k > 0$.
3. If $\sqrt{n}(\hat{\theta}_n - \theta) \xrightarrow{d} \mathcal{N}(0, V)$, then $\hat{\theta}_n \xrightarrow{p} \theta$ (i.e. consistency follows from asymptotic normality).
4. The delta method can be applied to any continuous transformation g of a consistent and asymptotically normal estimator.
5. If $[\hat{L}, \hat{U}]$ is a 95% CI for μ constructed from the data, then after the sample is observed, $\mathbb{P}(\hat{L} \leq \mu \leq \hat{U}) = 0.95$.

Practice 3b — OLS mechanics and Gauss-Markov

Skill target. Know what each OLS assumption buys you, what Gauss-Markov actually says, and what FWL claims.

1. The OLS estimator $\hat{\beta} = (X'X)^{-1}X'Y$ is unbiased for β provided $\mathbb{E}[\varepsilon \mid X] = 0$.
2. If errors are heteroskedastic but $\mathbb{E}[\varepsilon \mid X] = 0$ still holds, OLS is still unbiased but its usual (homoskedastic) variance formula is wrong.
3. The Frisch-Waugh-Lovell theorem says: the coefficient on D in the regression of Y on D and X equals the slope from regressing the *residualized outcome* $\tilde{Y}$ on the *residualized regressor* $\tilde{D}$, where both residuals are formed by partialling out X from Y and from D separately.
4. The Gauss-Markov theorem says OLS has the smallest variance among all unbiased estimators of β.

5. If $\sqrt{n}(\hat{\beta}_1 - \beta_1)/\widehat{\mathrm{SE}} \xrightarrow{d} \mathcal{N}(0,1)$ under H_0, then the test that rejects when $|\hat{\beta}_1/\widehat{\mathrm{SE}}| > 1.96$ has asymptotic size 0.05.

Practice 3c — Mixed traps (broader coverage)

Skill target. Extend your coverage of the trap categories in 3a and 3b. These items probe *related* but distinct misconceptions — so you are prepared for variations in wording on the exam.

1. If $\hat{\theta}_n$ is an unbiased estimator of θ, then $g(\hat{\theta}_n)$ is an unbiased estimator of $g(\theta)$ for any continuously differentiable g.
2. The delta method applied to $g(u) = u^2$ at $\mu = 0$ yields a nondegenerate normal limiting distribution for $\sqrt{n}(g(\bar{X}_n) - g(\mu))$.
3. A p-value of 0.03 means the null hypothesis is false with at least 97% confidence.
4. Including an irrelevant regressor (one whose true coefficient is zero) in an OLS regression leaves the other coefficient estimates unbiased — it only inflates their variance.
5. Under the Gauss-Markov assumptions, OLS is the maximum likelihood estimator of β.

Part IV — Building up to an exam problem on Matrix OLS + FWL

On the real exam, you may be asked to: given a small design matrix X and outcome vector Y, compute $\hat{\beta}$ by hand using $(X'X)^{-1}X'Y$, compute fitted values and residuals, and verify one coefficient using Frisch-Waugh-Lovell. The three problems build the mechanics.

Practice 4a

Skill target. Compute $X'X$, $(X'X)^{-1}$, $X'Y$, and $\hat{\beta}$ by hand for a two-column design.

Consider the data

$$X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix}, \qquad Y = \begin{pmatrix} 4 \\ 4 \\ 6 \\ 10 \end{pmatrix}.$$

(a) Compute $X'X$.

(b) Compute $\det(X'X)$ and $(X'X)^{-1}$.

(c) Compute $X'Y$.

(d) Compute $\hat{\beta} = (X'X)^{-1}X'Y$. Report $\hat{\beta}_0$ and $\hat{\beta}_1$.

(e) Compute the fitted values $\hat{Y} = X\hat{\beta}$ and the residuals $\hat{\varepsilon} = Y - \hat{Y}$.

(f) Verify that $\sum_i \hat{\varepsilon}_i = 0$ and $\sum_i X_{i1}\, \hat{\varepsilon}_i = 0$.

Practice 4b

Skill target. Apply FWL to the same data: partial the intercept out of Y and X_1, then regress. Verify the result matches $\hat{\beta}_1$.

Using the same data as in 4a:

(a) The “residual from regressing X_1 on the intercept” is just $\tilde{X}_{1,i} = X_{i1} - \bar{X}_1$ (within-demeaning). Compute $\tilde{X}_1$ for all four observations.

(b) Similarly compute $\tilde{Y}_i = Y_i - \bar{Y}$.

(c) Regress $\tilde{Y}$ on $\tilde{X}_1$ (no intercept, since both have sample mean zero). That is, compute

$$\hat{\beta}_1^{\text{FWL}} = \frac{\sum_i \tilde{X}_{1,i}\, \tilde{Y}_i}{\sum_i \tilde{X}_{1,i}^2}.$$

(d) Verify that $\hat{\beta}_1^{\text{FWL}}$ equals the $\hat{\beta}_1$ you computed in 4a.

(e) In one or two sentences: what is FWL saying about OLS when stated in this form?

Practice 4c

Skill target. Full in-family problem: *no intercept*, two demeaned substantive regressors. Compute $\hat{\beta}$ via the matrix formula. Verify $\hat{\beta}_1$ via FWL by residualizing X_1 *on* X_2 (not on an intercept) and residualizing Y on X_2. This is the move on the real exam.

Consider the data (already demeaned — there is no intercept column):

$$X_1 = \begin{pmatrix} 1 \\ 1 \\ -1 \\ -1 \end{pmatrix}, \qquad X_2 = \begin{pmatrix} 1 \\ 0 \\ -1 \\ 0 \end{pmatrix}, \qquad Y = \begin{pmatrix} 4 \\ 1 \\ -2 \\ -3 \end{pmatrix}, \qquad X = [\, X_1 \;\; X_2 \,].$$

(a) Compute $X'X$ (a 2×2 matrix), $\det(X'X)$, and $(X'X)^{-1}$.

(b) Compute $X'Y$ and then $\hat{\beta} = (X'X)^{-1}X'Y$. Report $\hat{\beta}_1$ and $\hat{\beta}_2$.

(c) Compute $\hat{Y} = X\hat{\beta}$ and residuals $\hat{\varepsilon} = Y - \hat{Y}$. Verify $X_1'\hat{\varepsilon} = 0$ and $X_2'\hat{\varepsilon} = 0$.

(d) **FWL Step 1 — residualize X_1 on X_2.** Compute $\hat{\gamma} = (X_2'X_2)^{-1}X_2'X_1$ and form $\tilde{X}_1 = X_1 - \hat{\gamma}\, X_2$.

(e) **FWL Step 2 — residualize Y on X_2.** Compute $\hat{\alpha} = (X_2'X_2)^{-1}X_2'Y$ and form $\tilde{Y} = Y - \hat{\alpha}\, X_2$.

(f) Compute $\hat{\beta}_1^{\text{FWL}} = (\tilde{X}_1'\tilde{Y})/(\tilde{X}_1'\tilde{X}_1)$. Confirm it matches $\hat{\beta}_1$ from (b).

Part V — Building up to an exam problem on FWL with a binary control: Angrist's variance-weighted decomposition

On the real exam, you may be asked: *when you regress Y on a binary treatment D and a binary stratifying control X, what does OLS compute?* The answer, derived from FWL alone, is that OLS returns a variance-weighted average of the within-stratum regression coefficients on D. The three problems build this answer step by step.

Practice 5a

Skill target. See FWL in action on a small dataset with a binary stratifying control. Learn that "residualizing D on X" is the same as "within-group demeaning."

Consider the following 8 observations:

i	X_i	D_i	Y_i	$\hat{D}_i$	$\tilde{D}_i$
1	0	0	2		
2	0	0	2		
3	0	1	4		
4	0	1	4		
5	1	0	3		
6	1	1	6		
7	1	1	6		
8	1	1	6		

(a) Regress D on X (intercept and X). Verify that the fitted value $\hat{D}_i$ is just the sample mean of D within the stratum $X = X_i$. Compute $\hat{D}_i$ for each row.

(b) Compute the residuals $\tilde{D}_i = D_i - \hat{D}_i$. Confirm that $\sum_i \tilde{D}_i = 0$ and $\sum_i X_i \tilde{D}_i = 0$.

(c) Apply FWL: compute

$$\hat{\beta}_D^{\text{FWL}} = \frac{\sum_i \tilde{D}_i Y_i}{\sum_i \tilde{D}_i^2}.$$

(d) You will be told (no need to verify by hand): the coefficient on D from the *long* regression of Y on D and X is $17/7 \approx 2.429$. Confirm your FWL answer matches.

(e) In one sentence: what does FWL say in words, applied to this problem?

Practice 5b

Skill target. Derive $\mathrm{Var}(\tilde{D})$ at the *population* level when X is a binary stratifying control. Use the law of total variance.

Let $X \sim \mathrm{Bernoulli}(\mu)$ and, conditional on $X = g$, let $D \mid X = g \sim \mathrm{Bernoulli}(p_g)$ for $g \in \{0, 1\}$. Define $\tilde{D} = D - \mathbb{E}[D \mid X]$.

(a) Show that $\mathbb{E}[\tilde{D} \mid X] = 0$.

(b) Show that $\mathrm{Var}(\tilde{D} \mid X = g) = p_g(1 - p_g)$. (Hint: conditional on $X = g$, D is $\mathrm{Bernoulli}(p_g)$, so $\tilde{D} = D - p_g$ has the same variance as D itself.)

(c) State the law of total variance for $\tilde{D}$:

$$\mathrm{Var}(\tilde{D}) = \mathbb{E}\big[\mathrm{Var}(\tilde{D} \mid X)\big] + \mathrm{Var}\big(\mathbb{E}[\tilde{D} \mid X]\big).$$

(d) Use (a) to show that the second term on the right-hand side is zero.

(e) Use (b) and (d) together to conclude:

$$\mathrm{Var}(\tilde{D}) = (1 - \mu)\, p_0(1 - p_0) + \mu\, p_1(1 - p_1).$$

(f) In one sentence: why are the weights $(1 - \mu)$ and μ in that final expression? (What is each one counting?)

Practice 5c

Skill target. Combine FWL with the result from 5b to derive Angrist's variance-weighted decomposition. Evaluate at specific numbers. State the asymptotic distribution of $\hat{\beta}_D$.

Keep the setup of Practice 5b: $X \sim \mathrm{Bernoulli}(\mu)$, $D \mid X = g \sim \mathrm{Bernoulli}(p_g)$. Define, for each stratum $g \in \{0, 1\}$, the within-stratum mean difference

$$\tau_g = \mathbb{E}[Y \mid D = 1, X = g] - \mathbb{E}[Y \mid D = 0, X = g].$$

This is just a difference in conditional means — a regression coefficient within the stratum $X = g$. (No potential-outcome language needed.)

Now derive the population coefficient β_D that OLS estimates when you regress Y on D and X.

(a) By FWL, $\beta_D = \mathrm{Cov}(\tilde{D}, Y)/\mathrm{Var}(\tilde{D})$. Quote this formula; the denominator was done in 5b.

(b) For the numerator, argue using the law of total covariance (or iterated expectations) that

$$\mathrm{Cov}(\tilde{D}, Y) = \mathbb{E}\big[\,\mathrm{Cov}(D, Y \mid X)\big].$$

(Hint: since $\mathbb{E}[\tilde{D} \mid X] = 0$, the "between" piece vanishes and only the "within" piece survives.)

(c) Within stratum $X = g$, expand

$$\operatorname{Cov}(D, Y \mid X = g) = \mathbb{E}[DY \mid X = g] - \mathbb{E}[D \mid X = g]\,\mathbb{E}[Y \mid X = g].$$

Use $\mathbb{E}[DY \mid X = g] = p_g\,\mathbb{E}[Y \mid D = 1, X = g]$ and $\mathbb{E}[Y \mid X = g] = p_g\,\mathbb{E}[Y \mid D = 1, X = g] + (1 - p_g)\,\mathbb{E}[Y \mid D = 0, X = g]$ to show that

$$\operatorname{Cov}(D, Y \mid X = g) = p_g(1 - p_g)\,\tau_g.$$

(d) Combine (b), (c), and the result of 5b to obtain

$$\beta_D = \frac{(1 - \mu)\,p_0(1 - p_0)\,\tau_0 + \mu\,p_1(1 - p_1)\,\tau_1}{(1 - \mu)\,p_0(1 - p_0) + \mu\,p_1(1 - p_1)}.$$

In one sentence: interpret this. What is OLS returning, mechanically?

(e) Evaluate β_D at the following numbers: $\mu = 0.4$, $p_0 = 0.3$, $p_1 = 0.7$, $\tau_0 = 2$, $\tau_1 = 5$.

(f) State the asymptotic distribution of $\sqrt{n}(\hat{\beta}_D - \beta_D)$. (You may cite the standard OLS asymptotic normality result; no new derivation required.)