Problem Set 02

Aaron Graybill
Collaborators: John Cho, Thomas Li, Blair Moore

2025-10-5

Contents

If you're having trouble running my code, see https://github.com/aarongraybill/MTGECON_603. Also the spacing on the tables text are really not ideal, Rmarkdown and LateX are not playing nicely, sorry about that!

1 Problem 1.

1.1 Problem 1.a.

Table 1 summarizes the results from the Neyman and OLS estimates for the post-treatment one and four year earnings. The estimates for Neyman versus OLS are quite similar given the large N. In every case, the 90% confidence intervals do not contain zero which indicates that the treatment had a significant average treatment effect on this population.

Table 1: Summary of Neyman and OLS point estimates, variances, and confidence intervals for one and four year earnings post-treatment.

method	outcome_var	n	point	variance	lb	ub
Neyman	earnings_1	5419	1.136210	0.0179932	0.9155717	1.356849
Neyman	earnings_4	5419	1.232332	0.0619079	0.8230711	1.641593
OLS	earnings_1	5419	1.136210	0.0290314	0.8559500	1.416470
OLS	earnings_4	5419	1.232332	0.0786663	0.7709917	1.693673

1.2 1.b.

When you stratify the data by high school completion and whether or not the subject has a young child and re-estimate the point estimates and variances, there are a couple things of note. Table 2 summarizes these results. Of course, each of the different strata have different point estimates and variances. Part of the difference in variance comes from the different population size in each stratum, but part of the difference in variance comes from the different groups empirically having different characteristics.

Table 2: Table of Neyman average treatment effects and variances for one and four year post-treatment earnings stratified by combinations of highs chool completion and the presence of children under 6.

young_child	high_school	method	outcome_var	n	point	variance	sd
0	0	Neyman	earnings_1	2216	0.7640112	0.0289347	0.1701021
0	1	Neyman	earnings_1	2316	1.5039758	0.0546637	0.2338027
1	0	Neyman	earnings_1	368	0.4951185	0.3487760	0.5905726
1	1	Neyman	earnings_1	519	1.4305537	0.1784069	0.4223824
0	0	Neyman	earnings_4	2216	0.9470774	0.0620227	0.2490436
0	1	Neyman	earnings_4	2316	1.1731442	0.2132351	0.4617739
1	0	Neyman	earnings_4	368	0.8861242	0.8233552	0.9073892
1	1	Neyman	earnings_4	519	2.8341227	0.8465675	0.9200911

1.3 1.c.

Finally table 3 aggregates the statistics in table 2 according to the appropriate aggregation for the respective summary statistics. Of course, the population size `n_tot` can simply be summed across the strata. The `point` column, the Neyman ATE point estimate, is the weighted average of each point estimates. Just to fix notation, index the strata $g \in \{1, \ldots, G\}$. Denote the size of each stratum N_g. I compute the aggregate point estimate with:

$$\sum_{g=1}^{G} \frac{N}{N_g} \hat{\tau}_g$$

Where $\hat{\tau}_g$ is the estimated ATE for each stratum.

To aggregate the variances, I use:

$$\sum_{g=1}^{G} \left(\frac{N}{N_g} \right)^2 \hat{\nu}_g$$

Where $\hat{\nu}_g$ is the estimated stratum level variance.

Comparing the variances in part 1.a. to part 1.c., we can see that the variance slightly increased when aggregating strata for 1 year earnings. Since we did not

Table 3: Table of Neyman average treatment effects and variances for one and four year post-treatment earnings combined across strata of combinations of highs chool completion and the presence of children under 6.

method	outcome_var	n_tot	point	var	sd
Neyman	earnings_1	5419	1.125838	0.0180683	0.1344183
Neyman	earnings_4	5419	1.220286	0.0608832	0.2467452

2 Problem 2

2.1 2.a

We showed in class and in the lecture notes that the estimator for the average treatment effect is also an unbiased estimator for the average treatment effect for the treated (ATT). Let's compute the variance of the ATE estimator around the (stochastic) ATT estimand.

That is:

$$\mathbb{V}_{\mathbf{W}}[\hat{\tau} - \tau_t]$$

Where $\mathbf{W}$ denote the random stochastic assignment vector and τ_t is the stochastic estimand.

Using the alternate variance formula, we can express the desired variance as:

$$\mathbb{V}_{\mathbf{W}}[\hat{\tau} - \tau_t] = \mathbb{E}_{\mathbf{W}}[(\hat{\tau} - \tau_t)^2] - \mathbb{E}_{\mathbf{W}}[(\hat{\tau} - \tau_t)]^2$$

Since the estimator $\hat{\tau}$ is unbiased for τ_t, that second term is $0^2 = 0$.

So computing the variance reduces to $\mathbb{E}[(\hat{\tau} - \tau_t)^2]$. Expanding relevant definitions, we have:

$$\mathbb{E}_{\mathbf{W}}\left[\left(\frac{1}{M}\sum_{i=1}^{N} W_i Y_i(T) - \frac{1}{N-M}\sum_{i=1}^{N}(1-W_i)Y_i(C) - \frac{1}{M}\sum_{i=1}^{N} W_i(Y_i(T) - Y_i(C))\right)^2\right]$$

We can do some cancellation

$$\mathbb{E}_{\mathbf{W}}\left[\left(\frac{1}{M}\sum_{i=1}^{N} W_i Y_i(T) - \frac{1}{N-M}\sum_{i=1}^{N}(1-W_i)Y_i(C) - \frac{1}{M}\sum_{i=1}^{N} W_i(Y_i(T) - Y_i(C))\right)^2\right]$$

$$\mathbb{E}_{\mathbf{W}}\left[\left(\frac{1}{M}\sum_{i=1}^{N} W_i Y_i(C) - \frac{1}{N-M}\sum_{i=1}^{N}(1-W_i)Y_i(C)\right)^2\right]$$

$$\mathbb{E}_{\mathbf{W}}\left[\left(\frac{1}{M}\sum_{i=1}^{N} W_i Y_i(C) + \frac{1}{N-M}\sum_{i=1}^{N} W_i Y_i(C) - \frac{1}{N-M}\sum_{i=1}^{N} Y_i(C)\right)^2\right]$$

$$\mathbb{E}_{\mathbf{W}}\left[\left(\frac{(N-M)+M}{M(N-M)}\sum_{i=1}^{N} W_i Y_i(C) - \frac{1}{N-M}\sum_{i=1}^{N} Y_i(C)\right)^2\right]$$

$$\mathbb{E}_{\mathbf{W}}\left[\left(\frac{N}{M(N-M)}\sum_{i=1}^{N} W_i Y_i(C) - \frac{1}{N-M}\sum_{i=1}^{N} Y_i(C)\right)^2\right]$$

$$\mathbb{E}_{\mathbf{W}}\left[\left(\frac{N}{M(N-M)}\sum_{i=1}^{N} W_i Y_i(C) - \frac{1}{N-M}\sum_{i=1}^{N} Y_i(C)\right)^2\right]$$

Using the alternate form of variance but now with $\mathbb{V}[X] + \mathbb{E}[X]^2 = \mathbb{E}^2[X]$, we can decompose the expression above into:

$$\mathbb{V}_{\mathbf{W}}\left[\frac{N}{M(N-M)}\sum_{i=1}^{N} W_i Y_i(C) - \frac{1}{N-M}\sum_{i=1}^{N} Y_i(C)\right] + \mathbb{E}_{\mathbf{W}}\left[\frac{N}{M(N-M)}\sum_{i=1}^{N} W_i Y_i(C) - \frac{1}{N-M}\sum_{i=1}^{N} Y_i(C)\right]^2$$

I will now argue that $\mathbb{E}_{\mathbf{W}}\left[\frac{N}{M(N-M)}\sum_{i=1}^{N} W_i Y_i(C) - \frac{1}{N-M}\sum_{i=1}^{N} Y_i(C)\right]^2 = 0$. Using the linearity of expectation, we have:

$$\left(\mathbb{E}_{\mathbf{W}}\left[\frac{N}{M(N-M)}\sum_{i=1}^{N} W_i Y_i(C)\right] - \mathbb{E}_{\mathbf{W}}\left[\frac{1}{N-M}\sum_{i=1}^{N} Y_i(C)\right]\right)^2$$

Pulling the constants out:

$$\left(\frac{N}{M(N-M)}\mathbb{E}_{\mathbf{W}}\left[\sum_{i=1}^{N} W_i Y_i(C)\right] - \frac{1}{N-M}\mathbb{E}_{\mathbf{W}}\left[\sum_{i=1}^{N} Y_i(C)\right]\right)^2$$

Noting that $W_i \perp Y_i(C)$, we can see that the second term is non-stochastic, so we must have:

$$\left(\frac{N}{M(N-M)}\mathbb{E}_{\mathbf{W}}\left[\sum_{i=1}^{N} W_i Y_i(C)\right] - \frac{1}{N-M}\sum_{i=1}^{N} Y_i(C)\right)^2$$

Similary since $W_i \perp Y_i(C)$, we can bring the expectation inside the summation to get:

$$\left(\frac{N}{M(N-M)}\sum_{i=1}^{N} \mathbb{E}_{\mathbf{W}}[W_i] Y_i(C) - \frac{1}{N-M}\sum_{i=1}^{N} Y_i(C)\right)^2$$

Across all of the realizations of $\mathbf{W}$, a given W_i is selected with probability M/N, so we have:

$$\left(\frac{N}{M(N-M)}\sum_{i=1}^{N} \frac{M}{N} Y_i(C) - \frac{1}{N-M}\sum_{i=1}^{N} Y_i(C)\right)^2 = \left(\frac{1}{N-M}\sum_{i=1}^{N} Y_i(C) - \frac{1}{N-M}\sum_{i=1}^{N} Y_i(C)\right)^2 = 0$$

So all that's left is to compute:

$$\mathbb{V}_{\mathbf{W}}\left[\frac{N}{M(N-M)}\sum_{i=1}^{N} W_i Y_i(C) - \frac{1}{N-M}\sum_{i=1}^{N} Y_i(C)\right]$$

Once again, assignments are orthogonal to outcomes, so the second term is non-varying in Y and doesn't contribute to variance. So we are left with:

$$\mathbb{V}_{\mathbf{W}}\left[\frac{N}{M(N-M)}\sum_{i=1}^{N} W_i Y_i(C)\right]$$

With the remaining sum, we need to to take seriously the covariances between the W_i's, so the sum reduces to:

$$\frac{N^2}{M^2(N-M)^2}\left(\sum_{i=1}^{N}\mathbb{V}_{\mathbf{W}}[W_i]Y_i(C)^2+\sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathbb{C}(W_i,W_j)\cdot Y_i(C)\cdot Y_j(C)\right)$$

Where i'm using the same orthogonality trick from before to pull the Ys out of the variances and covariances.

Over all the samples, the assignments have the same variance as a bernoulli assignment probabilities, so the variance of any individual $\mathbb{V}_{\mathbf{W}}[W_i]=\frac{M}{N}\cdot\left(1-\frac{M}{N}\right)$.

But for reasons that will be obvious in a moment it's more convenient to pull out a factor of $1/N$ from the second term giving: $\frac{M}{N^2}\cdot(N-M)$.

Plugging that back in we get:

$$\frac{N^2}{M^2(N-M)^2}\left(\sum_{i=1}^{N}\frac{M}{N^2}\cdot(N-M)\,Y_i(C)^2+\sum_{i=1}^{N}\sum_{j\neq i}^{N}\mathbb{C}(W_i,W_j)\cdot Y_i(C)\cdot Y_j(C)\right)$$

Thinking about the covariances between every distinct i and j, we can use the alternate formula for covariance that's:

$$\mathbb{E}[W_iW_j]-\mathbb{E}[W_i]\mathbb{E}[W_j]=\mathbb{E}[W_iW_j]-\frac{M}{N}\cdot\frac{M}{N}=\frac{M}{N}\cdot\frac{M-1}{N-1}-\frac{M}{N}\cdot\frac{M}{N}=\frac{M}{N}\left(\frac{M-1}{N-1}-\frac{M}{N}\right)$$

One final simplification gives:

$$\frac{M}{N}\left(\frac{M-1}{N-1}-\frac{M}{N}\right)=\frac{M}{N}\left(\frac{(MN-N)-(MN-M)}{N(N-1)}\right)=-\frac{M}{N^2}\cdot\frac{N-M}{N-1}$$

Plugging that back in we have:

$$\frac{N^2}{M^2(N-M)^2}\left(\sum_{i=1}^{N}\frac{M}{N^2}\cdot(N-M)\,Y_i(C)^2-\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{M}{N^2}\cdot\frac{N-M}{N-1}\cdot Y_i(C)\cdot Y_j(C)\right)$$

Then we have some nice cancellation giving:

$$\frac{1}{M(N-M)}\left(\sum_{i=1}^{N}Y_i(C)^2-\frac{1}{N-1}\sum_{i=1}^{N}\sum_{j\neq i}^{N}Y_i(C)\cdot Y_j(C)\right)$$

Pulling out a factor for $\frac{1}{N-1}$, we have:

$$\frac{1}{M(N-M)(N-1)}\left((N-1)\sum_{i=1}^{N}Y_i(C)^2-\sum_{i=1}^{N}\sum_{j\neq i}^{N}Y_i(C)\cdot Y_j(C)\right)$$

Now adding and subtracting a $\sum_{i=1}^{N}Y_i(C)^2$, we have:

$$\frac{1}{M(N-M)(N-1)}\left(N\sum_{i=1}^{N}Y_i(C)^2-\sum_{i=1}^{N}\sum_{j\neq i}^{N}Y_i(C)\cdot Y_j(C)-\sum_{i=1}^{N}Y_i(C)^2\right)$$

Note that $\left(\sum_i^N Y_i(C)\right)^2 = \sum_i^N Y_i(C)^2 + \sum_i^N \sum_{j\neq i} Y_i(C)Y_j(C)$ (just expanding out the quadratic form). Plugging that identity in, we have:

$$\frac{1}{M(N-M)(N-1)}\left(N\sum_{i=1}^N Y_i(C)^2 - \left(\sum_i^N Y_i(C)\right)^2\right)$$

Noting that $\sum_i^N Y_i(C) = N\overline{Y}^{obs}$, we have:

$$\frac{1}{M(N-M)(N-1)}\left(N\sum_{i=1}^N Y_i(C)^2 - N^2\overline{Y}^{obs\,2}\right) = \frac{N}{M(N-M)(N-1)}\left(\sum_{i=1}^N Y_i(C)^2 - N\overline{Y}^{obs\,2}\right)$$

Noting that summing something N times is the same as multiplying by N, we have:

$$\frac{N}{M(N-M)(N-1)}\left(\sum_{i=1}^N Y_i(C)^2 - \sum_{i=1}^N \overline{Y}^{obs\,2}\right)$$

Collecting into the sum:

$$\frac{N}{M(N-M)(N-1)}\left(\sum_{i=1}^N \left(Y_i(C)^2 - \overline{Y(C)}^2\right)\right)$$

Using the same algerba that makes, $\mathbb{E}[(X^2 - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, we know that:

$$\sum_{i=1}^N \left(Y_i(C)^2 - \overline{Y(C)}^2\right) = \sum_{i=1}^N \left(Y_i(C) - \overline{Y(C)}\right)^2$$

Leaving us with:

$$\frac{N}{M(N-M)}\frac{1}{N-1}\sum_{i=1}^N \left(Y_i(C) - \overline{Y(C)}\right)^2$$

and finally using the definition of S_C^2, we have that the final answer is:

$$\frac{N}{M(N-M)}S_C^2$$

Knowing from the lectures that (lowercase) s_C^2 is an unbiased estimator for (uppercase) S_C^2, then a multiple of s_C^2 will be an unbiased estimator for a multiple of S_C^2. Namely, our estimator will be:

$$\frac{N}{M(N-M)}s_C^2$$

2.2 2.b.

I'll quickly show that $\hat{\tau}$ is an unbiased estimator for τ_c, the ATU.

To show unbiasedness, we must show that $\mathbb{E}[\hat{\tau} - \tau_c] = 0$. Following the notes from section on 10/3, we proceed as follows:

$$\mathbb{E}_{\mathbf{W}}[\hat{\tau}-\tau_c] = \mathbb{E}_{\mathbf{W}}\left[\frac{1}{M}\sum_{i=1}^{N} W_i Y_i - \frac{1}{N-M}\sum_{i=1}^{N}(1-W_i)Y_i - \frac{1}{N-M}\sum_{i=1}^{N}(1-W_i)(Y_i(T)-Y_i(C))\right]$$

$$\begin{aligned}
&= \mathbb{E}_{\mathbf{W}}\left[\frac{1}{M}\sum_{i=1}^{N} W_i Y_i(T) - \frac{1}{N-M}\sum_{i=1}^{N}(1-W_i)Y_i(C) - \frac{1}{N-M}\sum_{i=1}^{N}(1-W_i)(Y_i(T)-Y_i(C))\right] \\
&= \mathbb{E}_{\mathbf{W}}\left[\frac{1}{M}\sum_{i=1}^{N} W_i Y_i(T) - \frac{1}{N-M}\sum_{i=1}^{N}(1-W_i)Y_i(T)\right] \\
&= \mathbb{E}_{\mathbf{W}}\left[\frac{1}{M}\sum_{i=1}^{N} W_i Y_i(T) - \frac{1}{N-M}\sum_{i=1}^{N}Y_i(T) + \frac{1}{N-M}\sum_{i=1}^{N} W_i Y_i(T)\right] \\
&= \mathbb{E}_{\mathbf{W}}\left[\left(\frac{1}{M}+\frac{1}{N-M}\right)\sum_{i=1}^{N} W_i Y_i(T) - \frac{1}{N-M}\sum_{i=1}^{N}Y_i(T)\right] \\
&= \mathbb{E}_{\mathbf{W}}\left[\left(\frac{N}{N-M}\right)\sum_{i=1}^{N} W_i Y_i(T) - \frac{1}{N-M}\sum_{i=1}^{N}Y_i(T)\right] \\
&= \left(\frac{N}{M(N-M)}\right)\sum_{i=1}^{N} \mathbb{E}_{\mathbf{W}}[W_i] Y_i(T) - \frac{1}{N-M}\sum_{i=1}^{N}Y_i(T) \\
&= \left(\frac{N}{M(N-M))}\right)\sum_{i=1}^{N} \frac{M}{N} Y_i(T) - \frac{1}{N-M}\sum_{i=1}^{N}Y_i(T) \\
&= \frac{1}{N-M}\sum_{i=1}^{N}Y_i(T) - \frac{1}{N-M}\sum_{i=1}^{N}Y_i(T) \\
&= 0
\end{aligned}$$

Then, a symmetric argument to the one in 2.a. shows that the true variance of the estimator, $\mathbb{V}_{\mathbf{W}}[\hat{\tau}-\tau_c]$, would be $\frac{N}{M(N-M)}S_T^2$ with unbiased estimator $\frac{N}{M(N-M)}s_T^2$

2.3 2.c.

I don't have a complete answer for this, and I'm not 100% sure about what it's asking. If it's asking for the variance of the true ATE, that would be zero (on the full sample, there is only one ATE), but if it's asking for the variance of the ATE estimator, $\mathbb{V}_{\mathbf{W}}[\hat{\tau}]$, I would note that we've already proven that $\mathbb{V}_{\mathbf{W}}[\hat{\tau}] = \frac{S_C^2}{N-M} + \frac{S_T^2}{M} - \frac{S_{CT}^2}{N}$, and given the variances from ATT and ATU, we still cannot estimate the S_{CT}^2 term, but if you're happy to make the conservative estimate that ignores S_{CT}^2, you could use:

$$\begin{aligned}
\mathbb{V}_{\mathbf{W}}[\hat{\tau}] &\leq \frac{\frac{M(N-M)}{N}\mathbb{V}_{\mathbf{W}}[\hat{\tau}-\tau_c]}{N-M} + \frac{\frac{M(N-M)}{N}\mathbb{V}_{\mathbf{W}}[\hat{\tau}-\tau_t]}{M} \\
&= \frac{M}{N}\mathbb{V}_{\mathbf{W}}[\hat{\tau}-\tau_c] + \frac{N-M}{N}\mathbb{V}_{\mathbf{W}}[\hat{\tau}-\tau_t]
\end{aligned}$$

3 2.a Graveyard

You don't have to read this, but this was my attempted derivation of for the variance of τ_t. I just have to leave it here for my sanity.

The average effect for the treated for a given realization of the assignment vector is given by:

$$\tau_t = \frac{1}{M}\sum_{i=1}^{N} \mathbf{1}_{W_i=1}(Y_i(T) - Y_i(C))$$

The true variance for those assignments would be:

$$\frac{1}{M}\sum_{i=1}^{N}(\mathbf{1}_{W_i=1}(Y_i(T) - Y_i(C)) - \tau_t)^2$$

Denote τ_t as the random variable corresponding to the ATT across the possible assignment vectors. Further let $\mathbf{W}$ be the random variable corresponding to the different possible assignment vectors. By the law of total variance, we can express $\mathbb{V}[\tau_t]$ as:

$$\mathbb{V}[\tau_t] = \mathbb{E}[\mathbb{V}[\tau_t|\mathbf{W}]] + \mathbb{V}[\mathbb{E}[\tau_t|\mathbf{W}]]$$

Consider the quantity $\mathbb{V}[\tau_t|\mathbf{W}]$. I will argue that this quantity must be zero, no matter the realization of $\mathbf{W}$. For any given realization of the assignment vector, there is only one possible value of τ_t (the average treatment effect for those treated under those assignments). Since there's only one possible value for τ_t, the conditional variance must be zero.

Therefore, the expected value of the variance across realizations of $\mathbf{W}$ must be zero ($E[0] = 0$).

That said, the second part of the above summation will not, in all likelihood, be zero. The quantity $\mathbb{E}[\tau_t|\mathbf{W}]$, while uniquely determined by a specific realization of $\mathbf{W}$, is a random variable that can take multiple values depending on the realizaton of the assignment vector. So to compute the variance over that expectation, we can use the definition of variance:

By the alternate definition of variance, we have that:

$$\mathbb{V}[\mathbb{E}[\tau_t|\mathbf{W}]] = \mathbb{E}[\mathbb{E}[\tau_t|\mathbf{W}]^2] - \mathbb{E}[\mathbb{E}[\tau_t|\mathbf{W}]]^2$$

$$\mathbb{V}[\mathbb{E}[\tau_t|\mathbf{W}]] = \mathbb{E}[(\mathbb{E}[\tau_t|\mathbf{W}] - \tau)^2]$$

for the second part of that difference, $\mathbb{E}[\mathbb{E}[\tau_t|\mathbf{W}]]^2$, it's relatively straightfoward to see that $\mathbb{E}[\mathbb{E}[\tau_t|\mathbf{W}]]^2 = \tau^2$ (where τ is the non-stochastic average treatment across across both treated and control units.) To see this, notice that the outer expectation is over assignments, which are all equally likely with probability $1/\binom{N}{M}$, so we can express the expectation as:

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[\tau_t|\mathbf{W}]] &= \sum_{W \in \Omega(\mathbf{W})} \frac{1}{\binom{N}{M}} \mathbb{E}[\tau_t|\mathbf{W} = W] \\
&= \frac{1}{\binom{N}{M}} \sum_{W \in \Omega(\mathbf{W})} \mathbb{E}[\tau_t|\mathbf{W} = W]
\end{aligned}$$

Those inner expectations are somewhat tricky to say anything about individually, but one thing we can say is that once we have conditioned on a particular W, there is only one possible value of τ_t and it occurs with probability 1. So it's not really an expectation, it's just whatever the value of τ_t is for that assignment vector. Each individual τ_t is just an average of the differences for treated groups, $\frac{1}{M}\sum_i W_i(Y_i(T) - Y_i(C))$.

Factoring out the shared $1/M$, we can rewrite $\mathbb{E}[\mathbb{E}[\tau_t|\mathbf{W}]]^2$ as:

$$\frac{1}{\binom{N}{M}} \cdot \frac{1}{M} \cdot \sum_{W \in \Omega(\mathbf{W})} \sum_i W_i(Y_i(T) - Y_i(C))$$

Another nice feature of the sum above is that every outcome, $Y_i(T) - Y_i(C)$, is treated exactly the same number of times across the possible values of $\mathbf{W}$. Since there are $\binom{N}{M}$ total realization vectors, and each vector has M total treatments, there are $M \cdot \binom{N}{M}$ total treatments across all assignment vectors. Since those treatments are occur evenly between all N participants, we must have that each i is treated $\frac{M \cdot \binom{N}{M}}{N}$ times. So instead of the nested summation, we can just say that each $Y_i(T) - Y_i(C)$ appears exactly $\frac{M \cdot \binom{N}{M}}{N}$. So that is:

$$\begin{aligned}
\mathbb{E}[\mathbb{E}[\tau_t|\mathbf{W}]]] &= \frac{1}{\binom{N}{M}} \cdot \frac{1}{M} \cdot \frac{M \cdot \binom{N}{M}}{N} \cdot \sum_i (Y_i(T) - Y_i(C)) \\
&= \frac{1}{N} \sum_i (Y_i(T) - Y_i(C)) \\
&= \tau
\end{aligned}$$

Beautiful, yet unsurprising. The expected average treatment effect for the treated is the true τ, so $\mathbb{E}[\mathbb{E}[\tau_t|\mathbf{W}]]]^2 = \tau^2$.

Now for $\mathbb{E}[\mathbb{E}[\tau_t|\mathbf{W}]^2]$. Consider an arbitrary $W \in \Omega\mathbf{W}$.

It's value for $\mathbb{E}[\tau_t|\mathbf{W} = W]^2$ would be something of the form:

$$\left(\frac{1}{M}\sum_i W_i \cdot (Y_i(T) - Y_i(C))\right)^2 = \frac{1}{M^2}\left(\sum_i W_i \cdot (Y_i(T) - Y_i(C))\right)^2$$

For a specific realization of W, that squared sum would be of the form:

$$\left(\sum_i W_i \cdot (Y_i(T) - Y_i(C))\right)^2 = \sum_{i:W_i=1} (Y_i(T) - Y_i(C))^2 + \sum_{i:W_i=1} \sum_{j>i:W_i=1} 2 \cdot (Y_i(T) - Y_i(C)) \cdot (Y_j(T) - Y_j(C))$$

There, all I have done is decompose the sum into its own terms and cross terms. The $j > i$ is accounting for the fact that each pair of items should only appear once in that sum. So we can have a term for $(i = 1, j = 2)$, but not simultaneously $(i = 2, j = 1)$ because we would collect those into the same term.

Now we try to think to think about what happens over the various realizations of $W \in \Omega(\mathbf{W})$. Similar to the logic from before, each realization of the assignment vector, W has equal probability with $1/\binom{N}{M}$. So the unconditional expected value is something of the form:

$$\frac{1}{\binom{N}{M}} \sum_{W \in \Omega(\mathbf{W})} \left[\sum_{i:W_i=1} (Y_i(T) - Y_i(C))^2 + \sum_{i:W_i=1} \sum_{j>i:W_i=1} 2 \cdot (Y_i(T) - Y_i(C)) \cdot (Y_j(T) - Y_j(C))\right]$$

We can separate that into separate sums for the own vs cross terms like follows:

$$\frac{1}{\binom{N}{M}} \sum_{W \in \Omega(\mathbf{W})} \left[\sum_{i:W_i=1} (Y_i(T) - Y_i(C))^2\right] + \frac{1}{\binom{N}{M}} \sum_{W \in \Omega(\mathbf{W})} \left[\sum_{i:W_i=1} \sum_{j\neq i:W_i=1} 2 \cdot (Y_i(T) - Y_i(C)) \cdot (Y_j(T) - Y_j(C))\right]$$

Starting with the first part of that sum, we're in a very similar case to before. Each inidividual i, must appear exactly $\frac{\binom{N}{M}M}{N}$ times across all of the inner summations and realizations of the assignment vectors. So we can replace that first sum with:

$$\frac{1}{\binom{N}{M}}\frac{\binom{N}{M}M}{N}\sum_{i=1}^{N}(Y_i(T)-Y_i(C))^2 = \frac{M}{N}\sum_{i=1}^{N}(Y_i(T)-Y_i(C))^2$$

Notice that the $\frac{\binom{N}{M}M}{N}\sum_{i=1}^{N}$ is not squared because it is simply counting the total number of times each outcome appears, the only thing squared is the actual difference in outcomes.

The other part of the sum, $\sum_{W\in\Omega(\mathbf{W})}\left[\sum_{i:W_i=1}\sum_{j\neq i:W_i=1} 2\cdot(Y_i(T)-Y_i(C))\cdot(Y_j(T)-Y_j(C))\right]$, is more combinatorially challenging. Here we need to enumerate each of the times that every *pair* of distinct $i \neq j$ appears in treatment group *together*. Fixing any given i and j, there are $M-2$ remaining treatments that must be distributed between the remaining $N-2$ units. So each $i \neq j$ grouping appears in exactly $\binom{N-2}{M-2}$ realizations of $\mathbf{W}$.

That portion of the sum reduces to:

$$\frac{1}{\binom{N}{M}}\cdot\binom{N-2}{M-2}\cdot 2\sum_{i=1}^{N}\sum_{j>i}\cdot(Y_i(T)-Y_i(C))\cdot(Y_j(T)-Y_j(C))$$

Noting that

$$\frac{\binom{N-2}{M-2}}{\binom{N}{M}} = \frac{\frac{(N-2)!}{(M-2)!((N-2)-(M-2))!}}{\frac{N!}{M!(N-M)!}} = \frac{\frac{1}{1}}{\frac{N\cdot(N-1)}{M\cdot(M-1)}} = \frac{M\cdot(M-1)}{N\cdot(N-1)}$$

We have:

$$\frac{M\cdot(M-1)}{N\cdot(N-1)}\cdot 2\sum_{i=1}^{N}\sum_{j>i}\cdot(Y_i(T)-Y_i(C))\cdot(Y_j(T)-Y_j(C))$$