Homework 2
Statistics 140

Due: February 18, 2026

Problem 1

(a) Why is the exact Fisher p-value preferable to an approximation?

The exact Fisher p-value is computed by enumerating *all* possible randomized allocations and calculating the exact fraction that produce a test statistic at least as extreme as the observed one. An approximation based on a smaller random sample of allocations introduces Monte Carlo noise: the estimated p-value varies across replications and can be above or below the true value. Specific drawbacks of the approximation include:

- **Variability.** The simulated p-value is itself a random variable, so two analysts running the same study with different seeds can reach different conclusions.

- **Discreteness near the boundary.** In a small experiment the exact distribution is highly discrete. A simulation may miss some tie-breaking allocations, distorting the p-value around decision thresholds (e.g. $\alpha = 0.05$).

- **Misleading precision.** Quoting a decimal p-value derived from B simulations with standard error $\approx \sqrt{p(1-p)/B}$ can convey false precision.

When the total number of allocations $\binom{N}{N_T}$ is feasible to enumerate, the exact computation is straightforward, free of simulation error, and fully reproducible.

(b) Minimum Fisher exact p-value: $N = 10$, $N_T = 4$, CRE

In a completely randomized experiment (CRE) with $N = 10$ units and $N_T = 4$ treated units, the reference distribution is uniform over all $\binom{10}{4}$ equally-likely allocations.

$$\binom{10}{4} = 210.$$

The minimum possible p-value is attained when exactly one allocation (the observed one) is as extreme as or more extreme than the observed statistic:

$$p_{\min} = \frac{1}{\binom{10}{4}} = \frac{1}{210} \approx 0.00476.$$

(c) Minimum Fisher exact p-value: `small_epi` dataset

Loading the `small_epi.RData` dataset reveals $N = 17$ participants with $N_T = 10$ exposed to ozone (O_3) and $N_C = 7$ exposed to clean air (CA).

$$\binom{17}{10} = 19{,}448.$$

$$p_{\min} = \frac{1}{19{,}448} \approx 5.14 \times 10^{-5}.$$

(d) Minimum Fisher exact p-value: $N = 20$, Bernoulli randomization

Under Bernoulli randomization each unit is independently assigned to treatment with probability $1/2$, so there are 2^{20} equally-likely allocation vectors (each of probability $(1/2)^{20}$).

$$2^{20} = 1{,}048{,}576.$$

$$p_{\min} = \frac{1}{2^{20}} = \frac{1}{1{,}048{,}576} \approx 9.54 \times 10^{-7}.$$

(e) Limitations of Figure 2 in Zhong et al. (2017) under Bernoulli assignment

Figure 2 in Zhong et al. reports the genome-wide results of an epigenome-wide association study (EWAS) for $N = 10$ volunteers under $PM_{2.5}$ exposure. Assuming a Bernoulli assignment mechanism, several statistical limitations arise.

1. **Unachievably small p-values.** Under Bernoulli randomization the reference distribution contains $2^{10} = 1{,}024$ equiprobable allocations. The minimum achievable Fisher exact p-value for any single test is therefore $1/1{,}024 \approx 0.001$. Figure 2D (Manhattan plot) and Figure 2B (QQ plot) display observed $-\log_{10} P$ values reaching 4–6, corresponding to p-values of 10^{-4}–10^{-6}—far below the achievable minimum. Such values cannot be validated by exact permutation testing; they rely entirely on asymptotic approximations from the linear mixed-effects model.

2. **Asymptotic, not exact, inference.** The reported p-values come from a parametric model (linear mixed effects with rank-normal transformation), not from the Fisher exact test. With only $N = 10$ participants, the normality assumptions underlying these asymptotic p-values are difficult to verify.

3. **Multiple testing with a limited reference distribution.** Approximately 450,000 CpG sites are tested. Because the Bernoulli reference distribution has only 1,024 distinct allocations, the permutation distribution is very coarse. Any Bonferroni- or FDR-corrected threshold that relies on very small p-values cannot be reached by exact methods with this sample size.

4. **Non-randomized treatment order.** The authors explicitly acknowledge that the order of placebo and B-vitamin supplementation was *not* randomized (to avoid long washout periods), a direct violation of Bernoulli randomization. Under a Bernoulli mechanism each volunteer should have had an equal probability of receiving placebo first or B-vitamin first. The fixed order introduces a potential temporal confound (learning effect, seasonal variation), which the authors attempt to address only through covariate adjustment.

5. **Small sample limits power.** With $N = 10$ and a minimum p-value of ≈ 0.001, the study is underpowered to distinguish true signals from noise at stringent genome-wide significance levels. The presentation of the "top 10 loci" selected by a combination of effect size and p-value (as in Maccani et al.) inflates the apparent significance via winner's curse.

Problem 2

A completely randomized experiment tests whether honey reduces nocturnal cough severity in 6 children ($N_1 = 3$ treated, $N_0 = 3$ control). Observed data:

Unit i	W_i^{obs}	Y_i^{obs}
1	1	3
2	1	2
3	1	0
4	0	4
5	0	6
6	0	1

(a) Observed data with potential outcomes

Under the sharp null hypothesis $H_0 : Y_i(1) = Y_i(0)$ for all i, both potential outcomes equal the observed outcome.

Unit i	W_i^{obs}	Y_i^{obs}	$Y_i(0)$	$Y_i(1)$
1	1	3	3	3
2	1	2	2	2
3	1	0	0	0
4	0	4	4	4
5	0	6	6	6
6	0	1	1	1

(b) Sharp null hypothesis

$$H_0 : Y_i(1) = Y_i(0) \quad \text{for all } i = 1, \ldots, 6.$$

That is, honey has *no effect* on any individual child's cough score.

(c) Matrix of all possible randomized allocations

With $N = 6$ and $N_1 = 3$, there are $\binom{6}{3} = 20$ possible allocations (no duplicates). Indexing by which units receive treatment ($W_i = 1$):

Treated units	$\mathbf{w}$	T	Treated units	$\mathbf{w}$	T
$\{1,2,3\}$	(1,1,1,0,0,0)	−2.000	$\{2,3,4\}$	(0,1,1,1,0,0)	−1.333
$\{1,2,4\}$	(1,1,0,1,0,0)	0.667	$\{2,3,5\}$	(0,1,1,0,1,0)	0.000
$\{1,2,5\}$	(1,1,0,0,1,0)	2.000	$\{2,3,6\}$	(0,1,1,0,0,1)	−3.333
$\{1,2,6\}$	(1,1,0,0,0,1)	−1.333	$\{2,4,5\}$	(0,1,0,1,1,0)	2.667
$\{1,3,4\}$	(1,0,1,1,0,0)	−0.667	$\{2,4,6\}$	(0,1,0,1,0,1)	−0.667
$\{1,3,5\}$	(1,0,1,0,1,0)	0.667	$\{2,5,6\}$	(0,1,0,0,1,1)	0.667
$\{1,3,6\}$	(1,0,1,0,0,1)	−2.667	$\{3,4,5\}$	(0,0,1,1,1,0)	1.333
$\{1,4,5\}$	(1,0,0,1,1,0)	3.333	$\{3,4,6\}$	(0,0,1,1,0,1)	−2.000
$\{1,4,6\}$	(1,0,0,1,0,1)	0.000	$\{3,5,6\}$	(0,0,1,0,1,1)	−0.667
$\{1,5,6\}$	(1,0,0,0,1,1)	1.333	$\{4,5,6\}$	(0,0,0,1,1,1)	2.000

(d) Null randomization distribution of T

Under the sharp null, potential outcomes are fixed, so for each allocation $\mathbf{w}$ we compute

$$T(\mathbf{w}) = \frac{1}{3} \sum_{i:\, w_i=1} Y_i^{\text{obs}} - \frac{1}{3} \sum_{i:\, w_i=0} Y_i^{\text{obs}}.$$
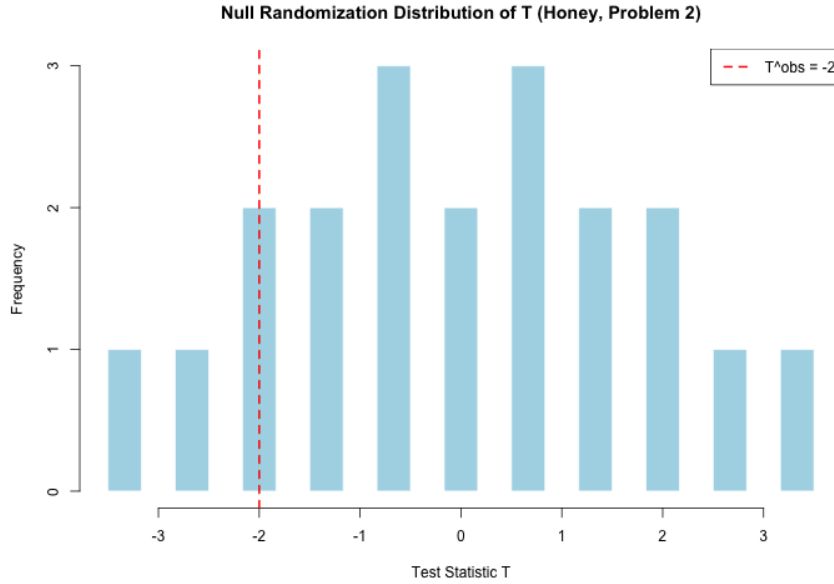
The histogram in Figure 1 shows all 20 values.

Figure 1: Null randomization distribution of T for the honey experiment (Problem 2). The red dashed line marks $T^{\text{obs}} = -2$.

(e) Observed test statistic

$$T^{\text{obs}} = \frac{3+2+0}{3} - \frac{4+6+1}{3} = \frac{5}{3} - \frac{11}{3} = -2.$$

This value is marked by the red dashed line in Figure 1.

(f) Fisher exact p-value

Because we expect honey to reduce cough (i.e. $T^{\text{obs}} < 0$), the one-sided p-value counts allocations with $T \le T^{\text{obs}} = -2$:

- $\{1, 2, 3\}$: $T = -2$ (the observed allocation)
- $\{1, 3, 6\}$: $T = -8/3 \approx -2.667$
- $\{2, 3, 6\}$: $T = -10/3 \approx -3.333$
- $\{3, 4, 6\}$: $T = -2$

$$p\text{-value} = \frac{4}{20} = 0.20.$$

There are 4 out of 20 allocations whose test statistic is at least as negative as the observed value of -2. We do not reject H_0 at the 5% level.

Problem 3

(a) Import `completelyrandomized.csv`

The dataset contains 17 participants and 6 CpG sites. Exposure is coded as `exp=2` (ozone) and `exp=0` (clean air). Note: the column is labelled `cg21036914` in the file (the homework states `cg21036194`); we use the actual column name throughout.

(b) Assignment vector `W.obs`

$$W_i^{\text{obs}} = \begin{cases} 1 & \text{if participant } i \text{ was exposed to ozone} \\ 0 & \text{if participant } i \text{ was exposed to clean air.} \end{cases}$$

There are $N_1 = 10$ ozone participants and $N_0 = 7$ clean-air participants.

(c) Matrix W of all possible allocations

The experiment is a CRE with $N = 17$, $N_T = 10$. The number of columns is

$$\binom{17}{10} = 19{,}448,$$

so the matrix W has dimensions $\mathbf{17 \times 19{,}448}$. Each column is a binary vector with exactly 10 ones, representing one possible assignment of 10 participants to ozone. This equals the total number of ways to choose which 10 of the 17 participants receive ozone, consistent with a completely randomized design that fixes $N_T = 10$.

(d) Outcome vector `Y.obs` for `cg00000029`

$\mathbf{Y}^{\text{obs}} = (0.116, 0.114, 0.140, 0.174, 0.064, 0.112, 0.123, 0.081, 0.100, 0.069, 0.182, 0.088, 0.194, 0.107, 0.122, 0.$

(e) Observed Welch statistic T^{obs}

$$T_{\text{Welch}} = \frac{\bar{Y}_1^{\text{obs}} - \bar{Y}_0^{\text{obs}}}{\sqrt{s_0^2/N_0 + s_1^2/N_1}},$$

where s_0^2 and s_1^2 are the sample variances of the clean-air and ozone groups, respectively. For `cg00000029`:

$$T^{\text{obs}} = 0.5724248.$$

(f) Verification with `t.test`

Running `t.test(Y.obs[W.obs==1], Y.obs[W.obs==0])$statistic` returns 0.5724248, confirming that the Welch two-sample t-statistic is identical.

(g) Null randomization distribution of T_{Welch}

For each of the 19,448 allocations, T_{Welch} is recomputed using the fixed $\mathbf{Y}^{\text{obs}}$ under the sharp null. The resulting distribution and the observed value $T^{\text{obs}} = 0.572$ are shown in Figure 2.
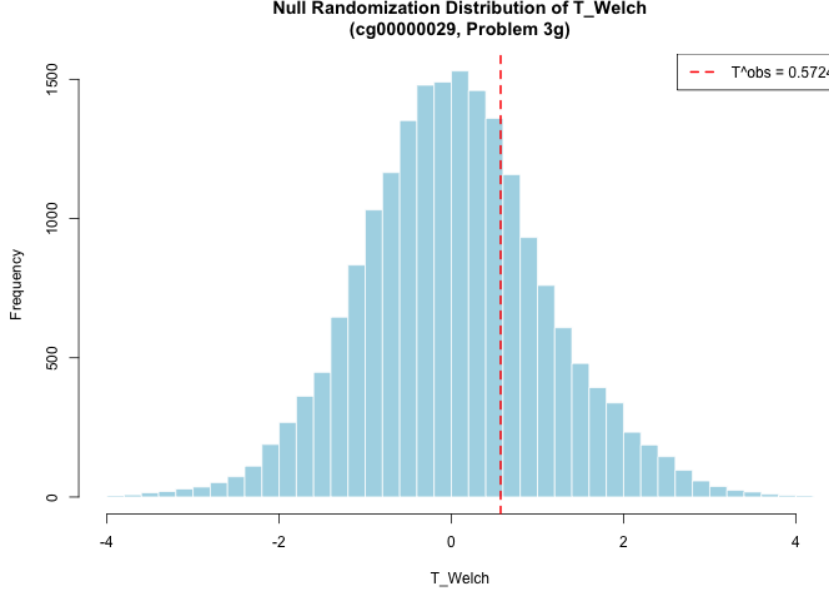
Figure 2: Null randomization distribution of T_{Welch} for `cg00000029` (Problem 3g). Red dashed line marks $T^{\text{obs}} = 0.572$.

(h) Fisher exact p-value ("extreme" = "greater than")

$$p\text{-value} = \frac{\#\{T_{\text{Welch}} \geq T^{\text{obs}}\}}{19{,}448} = \frac{5{,}679}{19{,}448} = 0.2920095.$$

(i) Comparison to `t.test` approximation

The `t.test` approximating p-value (assuming T_{Welch} follows a Student's t distribution under the Neymanian null) is 0.2888397. The two values are close but not identical because:

- The Fisher exact p-value is derived from the finite, discrete randomization distribution; the t-distribution approximation is continuous and asymptotic.

- With $N = 17$ the approximation is already fairly accurate, but exact and approximate values will always differ by a small amount due to the discreteness of the reference distribution.

(j) Fisher exact p-values for all six CpG sites

The table below is computed with two nested `for` loops: the outer loop iterates over the 6 CpG sites; the inner loop iterates over the 19,448 allocations in W. For the first three CpG sites, "extreme" means $T_{\text{Welch}} \geq T^{\text{obs}}$; for the last three, "extreme" means $T_{\text{Welch}} \leq T^{\text{obs}}$.

8

CpG site	T^{obs}	Fisher exact p-value	Approximating p-value
cg00000029	0.5724248	$5{,}679/19{,}448 = 0.2920095$	0.2888397
cg09008103	7.0847650	$1/19{,}448 = 0.0000514$	0.0000034
cg14354270	2.7647690	$2/19{,}448 = 0.0001028$	0.0154687
cg21036914	-1.8985240	$869/19{,}448 = 0.0446833$	0.0385190
cg00673208	-5.9502510	$1/19{,}448 = 0.0000514$	0.0000183
cg20976708	-8.1036900	$1/19{,}448 = 0.0000514$	0.0000004

Discussion in light of Problem 1. From Problem 1(c) we know the minimum achievable Fisher exact p-value in this CRE ($N = 17$, $N_T = 10$) is $1/19{,}448 \approx 5.14 \times 10^{-5}$. Four of the six CpG sites (cg09008103, cg14354270, cg00673208, cg20976708) yield very small exact p-values, with three of them attaining the minimum of $1/19{,}448$. This means no observed allocation could produce a more extreme statistic than the one actually observed—the observed allocation is *uniquely* the most extreme. Despite corresponding approximating p-values being even smaller (e.g. 4×10^{-7} for cg20976708), we cannot go below the floor $1/19{,}448$ with the exact approach. This illustrates the trade-off noted in Problem 1(a): the exact p-value is bounded below by $1/\binom{N}{N_T}$, whereas asymptotic approximations can in principle yield arbitrarily small values but may not be valid for small N.

(k) Six null randomization distributions

Figure 3 displays the null randomization distributions for all six CpG sites with the observed statistics marked.
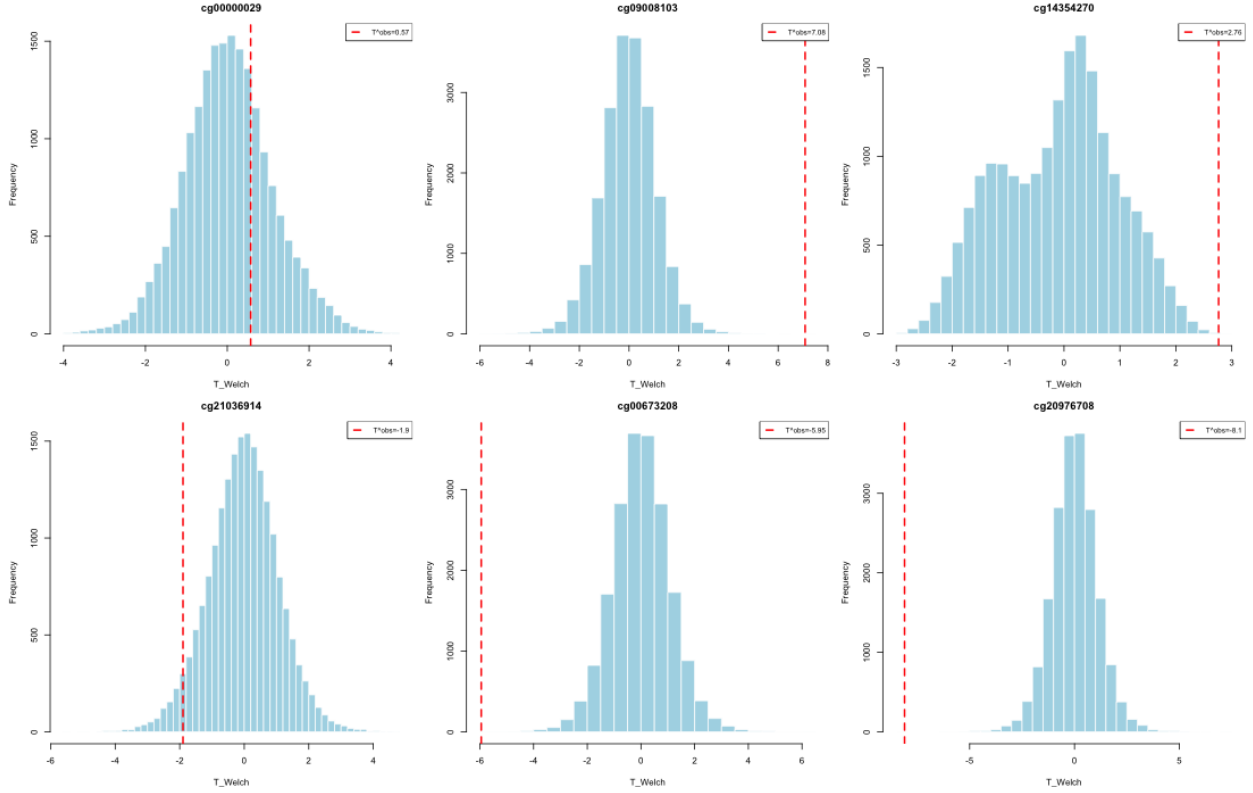
Figure 3: Null randomization distributions of T_{Welch} for the six CpG sites (Problem 3k). Red dashed lines mark the observed test statistics.

Comment on shape. All six null randomization distributions are approximately bell-shaped and symmetric about zero, closely resembling a normal (or Student's t) distribution. This is consistent with the theoretical result that the Welch t-statistic converges in distribution to a normal under the null as N grows. Even at $N = 17$, the continuous nature of the DNA methylation outcome produces a smooth bell curve. The centering at zero reflects the fact that under the sharp null there is no systematic difference between ozone and clean-air groups.

Problem 4

(a) Point estimate and 95% CI for `cg00000029`

The point estimate of the average treatment effect τ is the difference in sample means:

$$\hat{\tau} = \bar{Y}_1 - \bar{Y}_0.$$

Using the Neyman/Welch approach, a two-sided 95% CI is

$$\hat{\tau} \pm t^*_{\nu,0.025} \sqrt{s_0^2/N_0 + s_1^2/N_1},$$

where ν is the Welch–Satterthwaite degrees of freedom.

For `cg00000029`: $\hat{\tau} = 0.009$ and the 95% CI is $[-0.026,\ 0.044]$. The CI contains zero, so we cannot reject the null of no mean difference at the 5% level. This is consistent with the relatively large Fisher exact p-value of 0.292 obtained in Problem 3.

(b) Verification with `t.test`

Running `t.test(Y.obs[W.obs==1], Y.obs[W.obs==0], alternative="two.sided")` returns the same $\hat{\tau}$ and confidence interval.

(c) Results for all six CpG sites

CpG site	$\hat{\tau}$	Two-sided 95% CI
cg00000029	0.009	$[-0.026;\ 0.044]$
cg09008103	0.008	$[\ \ 0.006;\ 0.011]$
cg14354270	0.063	$[\ \ 0.008;\ 0.118]$
cg21036914	−0.006	$[-0.013;\ 0.001]$
cg00673208	−0.007	$[-0.010;\ -0.004]$
cg20976708	−0.030	$[-0.037;\ -0.022]$

Interpretation.

- `cg00000029`: CI contains 0; no significant effect of ozone.

- `cg09008103` and `cg14354270`: positive $\hat{\tau}$ with CIs entirely above zero, indicating ozone increases methylation at these sites.

- `cg21036914`: CI barely contains 0, marginal evidence for a decrease.

- `cg00673208` and `cg20976708`: negative $\hat{\tau}$ with CIs entirely below zero, indicating ozone *decreases* methylation at these sites.

These results are consistent with the p-values from Problem 3: the sites with the smallest Fisher exact p-values (1/19,448) also have CIs that exclude zero.