We will consider the regression model

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + \epsilon_i, \tag{1}$$

for $i = 1, ..., n$, where y_i is the outcome, $\mathbf{x}_i$ is a $1 \times p$ row-vector of covariates (including an intercept) for observation i, $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)'$ is a $p \times 1$ vector of regression coefficients, and ϵ_i is a mean-zero error term with variance σ_i^2. We shall assume that the errors are mutually independent. The model can be written in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2}$$

where $\mathbf{y} = (y_1, ..., y_n)'$ is an $n \times 1$ vector of outcomes, $\mathbf{X} = (\mathbf{x}_1', ..., \mathbf{x}_n')'$ is an $n \times p$ design matrix, and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of errors with $\mathrm{E}(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{\epsilon}) = \boldsymbol{\Sigma} = \mathrm{diag}\left(\sigma_1^2, ..., \sigma_n^2\right)$. The goal is to test the hypothesis that the q^{th} regression coefficient is equal to a constant c, i.e., $H_0 : \beta_q = c$ against the two-sided alternative $H_A : \beta_q \neq c$, with Type-I error rate α.

If the errors are homoskedastic, so that $\mathrm{Var}(\epsilon_i) = \sigma^2$ for $i = 1, ..., n$, then the hypothesis can be tested using a standard t-test. The regression coefficients are estimated using ordinary least squares, with

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}.$$

The residual variance of the regression is estimated as

$$\hat{\sigma}^2 = \frac{1}{n-p}\sum_{i=1}^{n} e_i^2,$$

where $e_i = y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}$ for $i = 1, ..., n$. The variance of $\boldsymbol{\beta}$ is then estimated by $\mathbf{V}^{hom} = \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$. Under H_0 and assuming that the errors are normally distributed and homoskedastic, the t-statistic $t_q^{hom} = \left(\hat{\beta}_q - c\right)/\sqrt{V_{qq}^{hom}}$ follows a t distribution with $n - p$ degrees of freedom. Thus, H_0 is rejected if $|t_q^{hom}| > F_t^{-1}\left(1 - \frac{\alpha}{2}; n - p\right)$, where $F_t^{-1}(x; \nu)$ is the quantile function for a t distribution with ν degrees of freedom. However, if the errors are instead heteroskedastic, the variance estimate V_{qq}^{hom} will be inconsistent and this t-test will generally have incorrect size (i.e., incorrect Type-I error).

1 Heteroskedasticity-consistent variance estimators

Heteroskedasticity-consistent (HC) variance estimators (a.k.a. "robust" variance estimators) provide a means to test hypotheses regarding the regression coefficients even if the errors are not homoskedastic or normally distributed. They are an attractive tool because violations of the homoskedasticity assumption are commonly observed and can be difficult to address through other methods of remediation. However, the guarantees that they provide are only asymptotic, they will provide correct estimates of the variance of $\hat{\boldsymbol{\beta}}$ and hypothesis tests of the correct size if the sample size is sufficiently large. In practice, it is not always clear whether a given sample is "sufficently large." Furthermore, some of the HC methods tend to be too liberal (producing variance estimates that are biased towards zero and hypothesis tests with size greater than nominal) when the sample size is small.

To see how the HC estimators work, we start by noting that under the general model (allowing for heteroskedasticity), the true variance of $\boldsymbol{\beta}$ is

$$\mathrm{Var}(\boldsymbol{\beta}) = \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\sigma_i^2\mathbf{x}_i\mathbf{x}'\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}'\right)^{-1} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{3}$$

The HC estimators all involve estimating $\mathrm{Var}(\boldsymbol{\beta})$ by replacing the σ_i^2 with crude estimates involving the squared residuals. Although taken one at a time, the squared residual e_i^2 is a very poor estimate of σ_i^2, they provide an adequate means of estimating the *average* of the σ_i^2 terms that appears in the middle of Equation (3). The HC estimators have the general form

$$\begin{aligned}\mathbf{V}^{HCx} &= \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}\omega_{xi}e_i^2\mathbf{x}_i\mathbf{x}'\right)\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}'\right)^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{diag}\left(\omega_{x1}e_1^2, ..., \omega_{xn}e_n^2\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},\end{aligned} \tag{4}$$

where $\omega_{x1}, ..., \omega_{xn}$ are correction terms that differ for the various HC estimators. Under weak assumptions, the weak law of large numbers ensures that the middle term in Equation (4) converges to the corresponding term in (3) as the sample size increases. Furthermore, the Wald statistic calculated as $t_q^{HCx} = \left(\hat{\beta}_q - c\right) / \sqrt{V_{qq}^{HCx}}$ follows a standard normal distribution if the sample size is sufficiently large (i.e., t_q^{HCx} converges in distribution to $N(0,1)$ as n increases to infinity). Thus, any asymptotically correct test can be constructed by rejecting H_0 when t_q^{HCx} is greater than the $1 - \alpha/2$ critical value from a standard normal critical value. Because this test often has inflated size in small samples, it is common to instead compare t_q^{HCx} to the critical value from a t distribution with $n - p$ degrees of freedom.

The various HC estimators use different correction terms that refine the behavior of the estimator in different ways. The original HC0 estimator did not use any correction, i.e., $\omega_{0i} = 1$ for $i = 1, ..., n$, producing an estimator with a downward bias. HC1 involves an ad hoc correction to the bias, with $\omega_{1i} = \frac{n}{n-p}$ for $i = 1, ..., n$. HC2 uses the correction term

$$\omega_{2i} = \frac{1}{1 - h_{ii}},$$

where $h_{ii} = \mathbf{x}_i \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{x}_i'$ is the i^{th} diagonal entry in the hat matrix. This correction has the property that the resulting variance estimator is exactly unbiased (at any sample size) when the errors are actually homoskedastic: $\mathrm{E}\left(\mathbf{V}^{HC2}\right) = \mathrm{Var}\left(\boldsymbol{\beta}\right) = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$. Intuitively, we would expect that HC2 would still be approximately unbiased if the degree of heteroskedasticity is small. However, using an unbiased variance estimator still does not guarantee that hypothesis tests constructed from it will have the correct size.

Several further HC variations have been proposed that aim to improve the accuracy of hypothesis tests. HC3 uses the correction term

$$\omega_{3i} = \frac{1}{(1 - h_{ii})^2},$$

which produces an estimator that closely approximates the omit-1 jackknife variance estimator. Note that these correction factors will always be larger than those for HC2 because $\frac{1}{n} \leq h_i i \leq 1$. HC4, proposed by Cribari-Neto (2004), uses a correction factor that is further inflated for observations with high leverage:

$$\omega_{4i} = \frac{1}{(1 - h_{ii})^{\delta_i^a}},$$

where $\delta_i^a = \min\left\{nh_{ii}/p, 4\right\}$. Cribari-Neto and da Silva (2011) later suggested an modified estimator, HC4m, that uses a correction factor of the same form as ω_{4i}, but where the exponent in the correction term is instead given by $\delta_i^b = \min\left\{nh_{ii}/p, 1\right\} + \min\left\{nh_{ii}/p, 1.5\right\}$. HC5, proposed by Cribari-Neto, Souza, and Vasconcellos (2007), also uses a correction factor that is tailored to account for the leverage of the observation:

$$\omega_{5i} = \frac{1}{(1 - h_{ii})^{\delta_i^c/2}},$$

where $\delta_i^c = \min\left\{nh_{ii}/p, \max\left\{4, knh_{max}/p\right\}\right\}$, $h_{max} = \max\left\{h_{11}, ..., h_{nn}\right\}$, and $0 < k < 1$ is a user-selected constant; based on simulation evidence, Cribari-Neto and colleagues suggest taking $k = 0.7$.

2 Distribution of HC estimators

The Wald statistic formed using an HC variance estimator is the ratio of a normally distributed estimator $\hat{\beta}_q - c$ to the corresponding variance estimator $V_q^{HCx} q$. The variance estimator can be written as a quadratic form in the residuals. Let the $n \times 1$ vector $\mathbf{g}_q = (g_{q1}, ..., g_{qn})'$ denote the q^{th} column of q of $\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}$; let $\mathbf{e} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ be the vector of residuals. It follows that

$$V_{qq}^{HCx} = \sum_{i=1}^{n} \omega_{xi} \left(g_{qi} e_i\right)^2 = \mathbf{e}'\mathbf{A}_{xq}\mathbf{e}, \tag{5}$$

where $\mathbf{A}_{xq} = \text{diag}\left(\omega_{x1} g_{q1}^2, ..., \omega_{xn} g_{qn}^2\right)$. Because the residuals are themselves a linear function of the outcome vector, V_{qq}^{HCx} can also be expressed as a quadratic form in $\mathbf{y}$:

$$V_{qq}^{HCx} = \mathbf{y}' \left(\mathbf{I} - \mathbf{H}\right) \mathbf{A}_{xq} \left(\mathbf{I} - \mathbf{H}\right) \mathbf{y},$$

where $\mathbf{H} = \mathbf{X} \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'$ is the full hat matrix. The numerator and denominator of the Wald statistic are not necessarily independent unless the errors are homoskedastic.

The distribution of V_{qq}^{HCx} can be obtained from the properties of quadratic forms. In general,

$$\text{E}\left(V_{qq}^{HCx}\right) = \text{tr}\left[\mathbf{A}_{xq} \left(\mathbf{I} - \mathbf{H}\right) \mathbf{\Sigma} \left(\mathbf{I} - \mathbf{H}\right)\right] = \sum_{i=1}^{n} \omega_{xi} g_{qi}^2 (1 - h_{ii}) \sigma_i^2.$$

If the errors are normally distributed, then

$$\text{Var}\left(V_{qq}^{HCx}\right) = 2\text{tr}\left[\mathbf{A}_{xq} \left(\mathbf{I} - \mathbf{H}\right) \mathbf{\Sigma} \left(\mathbf{I} - \mathbf{H}\right) \mathbf{A}_{xq} \left(\mathbf{I} - \mathbf{H}\right) \mathbf{\Sigma} \left(\mathbf{I} - \mathbf{H}\right)\right].$$

Furthermore (still assuming normality of the residuals), the HC variance estimator is distributed as a weighted sum of independent χ_1^2 random variates. Let $\lambda_1, ..., \lambda_n$ denote the eigenvalues of the matrix $\mathbf{A}_{xq} \left(\mathbf{I} - \mathbf{H}\right) \mathbf{\Sigma} \left(\mathbf{I} - \mathbf{H}\right)$ and let $z_1, ..., z_n$ be independent χ_1^2 random variables. Then

$$V_{qq}^{HCx} \sim \sum_{i=1}^{n} \lambda_i z_i. \tag{6}$$

These distributional properties provide a means to develop small-sample corrections for hypothesis tests based on the HC variance estimators. Several such corrections have been proposed, and are reviewed in the following sub-sections.

2.1 Satterthwaite approximation

Lipsitz, Ibrahim, and Parzen (1999) proposed a small-sample corrected hypothesis testing procedure that is based on a Satterthwaite approximation for the distribution of V_{qq}^{HC2}. The Satterthwaite approximation involves approximating the distribution of V_{qq}^{HC2} by a multiple of a χ^2 distribution with degrees of freedom $2\left[\text{E}\left(V_{qq}^{HC2}\right)\right]^2 / \text{Var}\left(V_{qq}^{HC2}\right)$. In practice, the mean and variance must be estimated because they involve the unknown quantity $\mathbf{\Sigma}$. Lipsitz and colleagues note that the variance of V_{qq}^{HCx} can be written as

$$\text{Var}\left(V_{qq}^{HCx}\right) = 2\text{tr}\left[\left(\mathbf{I} - \mathbf{H}\right) \mathbf{A}_{xq} \left(\mathbf{I} - \mathbf{H}\right) \left[\left(\left(\mathbf{I} - \mathbf{H}\right) \mathbf{A}_{xq} \left(\mathbf{I} - \mathbf{H}\right)\right) \circ \mathbf{S}\right]\right],$$

where $\circ$ denotes the element-wise (Hadamard) product and $\mathbf{S}$ has entries $S_{ij} = \sigma_i^2 \sigma_j^2$. They propose to estimate $\mathbf{S}$ using the matrix with entries

$$\hat{S}_{ii} = \frac{e_i^4}{3(1 - h_{ii})^2} \quad \text{for} i = 1, ..., n \quad \text{and} \quad \hat{S}_{ij} = \frac{e_i^2 e_j^2}{2h_{ij}^2 + (1 - h_{ii})(1 - h_{jj})} \quad \text{for} i \neq j.$$

They then construct estimated degrees of freedom ν_q by substituting V_{qq}^{HC2} in place of its expectation and taking

$$\nu_q = \frac{\left(V_{qq}^{HC2}\right)^2}{\text{tr}\left[\left(\mathbf{I} - \mathbf{H}\right) \mathbf{A}_{2q} \left(\mathbf{I} - \mathbf{H}\right) \left[\left(\left(\mathbf{I} - \mathbf{H}\right) \mathbf{A}_{2q} \left(\mathbf{I} - \mathbf{H}\right)\right) \circ \hat{\mathbf{S}}\right]\right]}. \tag{7}$$

The null hypothesis is tested by comparing t_q^{HC2} to a t-distribution with ν_q degrees of freedom, i.e., H_0 is rejected if $|t_q^{HC2}| > F_t^{-1}(1 - \alpha/2; \nu_q)$. Equivalently, the p-value corresponding to H_0 is $2\left[1 - F_t\left(|t_q^{HC2}|; \nu_q\right)\right]$, where $F_t(x; \nu)$ is the cumulative distribution function of a t_ν distribution.

2.2 Edgeworth approximation

Rothenberg (1988) developed an Edgeworth approximation for the distribution of Wald-type t-statistics based on the HC0 variance estimator. It is straight-forward to generalize the approach to any of the HC estimators. Let

$$\mathbf{Q} = (\mathbf{I} - \mathbf{H})\,\mathbf{\Sigma}\,(\mathbf{I} - \mathbf{H}), \qquad \mathbf{z}_q = (\mathbf{I} - \mathbf{H})\,\mathbf{\Sigma}\mathbf{g}, \qquad a = \frac{\sum_{i=1}^n g_{qi}^2 \omega_i z_{qi}^2}{\left(\sum_{i=1}^n g_{qi}^2 \sigma_i^2\right)^2}, \qquad b = \frac{\sum_{i=1}^n g_{qi}^2 \omega_i q_{ii}^2}{\sum_{i=1}^n g_{qi}^2 \sigma_i^2} - 1.$$

Let z_α denote the $1-\alpha/2$ quantile from a standard normal distribution. For an observed value of the test statistic t_q^{HC}, the corresponding p-value is calculated as

$$p = 2\left[1 - \Phi\left[\frac{\left|t_q^{HC}\right|}{2}\left(2 - \frac{1 + \left(t_q^{HC}\right)^2}{2\nu_q} - a\left(\left(t_q^{HC}\right)^2 - 1\right) - b\right)\right]\right],$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function. A further approximation provides a means for calculating a critical value for a specified α-level. Here, the hypothesis test is rejected if t_q^{HC} is greater than the critical value t_{crit} defined by

$$t_{crit} = \frac{z_\alpha}{2}\left[2 + \frac{z_\alpha^2 + 1}{2\nu_q} - a\left(z_\alpha^2 - 1\right) - b\right].$$

In practice, these testing procedures will need to be based on estimates of the quantities involved. Rothenberg proposes a simple estimate of the degrees of freedom:

$$\nu_q = \frac{\left(\sum_{i=1}^n g_{qi}^2 \omega_i e_i^2\right)^2}{\frac{1}{3}\sum_{i=1}^n g_{qi}^4 \omega_i^2 e_i^4}.$$

Alternately, one could use Equation (7). Rothenberg also proposes to calculate a, b, $\mathbf{z}_q$, and $\mathbf{Q}$ by simply replacing the values of σ_i^2 with $\omega_i e_i^2$.

2.3 Another edgeworth approximation

Kauermann and Carroll (2001) propose a method of constructing confidence intervals based on HC variance estimators that have close-to-nominal coverage rates. The hypothesis testing procedure corresponding to their proposed confidence intervals rejects H_0 if $\left|t_q^{HC}\right| > z_{\tilde{\alpha}}$, where $\tilde{\alpha}$ is implicitly defined as the solution to

$$\alpha = \tilde{\alpha} + \frac{\phi\left(z_{\tilde{\alpha}}\right)}{2\nu_q}\left(z_{\tilde{\alpha}}^3 + z_{\tilde{\alpha}}\right), \tag{8}$$

where $\phi(\cdot)$ is the density of the standard normal distribution and

$$\nu_q = \frac{2\left[\mathrm{Var}(\hat{\beta}_q)\right]^2}{\mathrm{Var}\left(V_{qq}^{HC}\right)}$$

is a degrees of freedom measure. Equivalently, the p-value for the test is given by

$$p = 2\left[1 - \Phi\left(\left|t_q^{HC}\right|\right)\right] + \frac{\phi\left(t_q^{HC}\right)}{2\nu_q}\left(\left|t_q^{HC}\right|^3 + \left|t_q^{HC}\right|\right),$$

These authors also offer a further approximation for the critical value $z_{\tilde{\alpha}}$, which saves the trouble of solving Equation (8):

$$z_{\tilde{\alpha}} = F_t^{-1}\left(1 - \frac{\alpha}{2}; n - p\right) + \frac{\left(z_\alpha^3 + z_\alpha\right)}{4}\left(\frac{1}{\nu_q} - \frac{\left(\sum_{i=1}^n g_{qi}^2\right)^2}{n}\right).$$

In contrast to the degrees of freedom estimator used by Lipsitz and colleagues (as given in Equation 7), Kauermann and Carroll calculate the degrees of freedom under the working assumption that the errors are actually homoskedastic. Under this working assumption, the HC2 variance estimator is unbiased, with degrees of freedom are given by

$$\nu_q = \left(\sum_{i=1}^{n} g_{qi}^2\right)^2 \left(\sum_{i=1}^{n} g_{qi}^4 + \sum_{i=1}^{n}\sum_{j\neq i} \frac{g_{qi}^2 g_{qj}^2 h_{ij}^2}{(1-h_{ii})(1-h_{jj})}\right)^{-1}$$

.

Differences between Kauermann and Carroll (2001) and Rothenberg (1988)?

2.4 Saddlepoint approximation

McCaffrey and Bell (2006) developed small-sample adjustments to test statistics based on cluster-robust variance estimators, of which HC variance estimators are a special case. They consider both a Satterthwaite approximation (similar to Lipsitz et al.) and a saddlepoint approximation for the distribution of the test statistic, finding that the latter produced tests with more accurate size. The saddlepoint approximation is obtained as follows. Let Observe that the cumulative distribution of t_q^{HC} can be expressed as

$$\Pr\left(t_q^{HC} \leq t\right) = \Pr\left(\frac{\left(\hat{\beta}_q - c\right)^2}{\mathrm{Var}(\hat{\beta}_q)} - t^2 \frac{V_{qq}^{HC}}{\mathrm{Var}(\hat{\beta}_q)} \leq 0\right).$$

Note that $\left(\hat{\beta}_q - c\right)^2 / \mathrm{Var}(\hat{\beta}_q)\ \chi_1^2$ and that V_{qq}^{HC} is distributed as a weighted sum of χ_1^2 random variables, with weights given by the eigen-values $\lambda_1, ..., \lambda_n$ of the matrix $\mathbf{A}_{xq}\left(\mathbf{I}-\mathbf{H}\right)\mathbf{\Sigma}\left(\mathbf{I}-\mathbf{H}\right)$. Assuming that V_{qq}^{HC} is unbiased, so that

$$\mathrm{E}\left(V_{qq}^{HC}\right) = \mathrm{tr}\left[\mathbf{A}_{xq}\left(\mathbf{I}-\mathbf{H}\right)\mathbf{\Sigma}\left(\mathbf{I}-\mathbf{H}\right)\right] = \sum_{j=1}^{n}\lambda_j,$$

and that $\hat{\beta}_q$ is independent of V_{qq}^{HC}, it follows that the $\Pr\left(t_q^{HC} \leq t\right)$ can be expressed as $\Pr(Z \leq 0)$, where $Z = \sum_{i=0}^{n}\gamma_i z_i$, $\gamma_0 = 1$, $\gamma_i = -t^2\lambda_i / \sum_{j=1}^{n}\lambda_j$, and $z_0, ..., z_n \overset{\text{iid}}{\sim} \chi_1^2$.

The saddlepoint technique is a means to approximate the distribution of Z. Let s be the saddlepoint, defined implicitly as the solution to

$$\sum_{i=0}^{n}\frac{\gamma_i}{1-2\gamma_i s} = 0.$$

Note that the solution to the saddlepoint equation will be in the range $\left(\left(2\min\left\{\gamma_0, ..., \gamma_n\right\}\right)^{-1}, 0\right)$ if $\sum_{i=0}^{n}\gamma_i > 0$ and in the range $\left(0, \left(2\max\left\{\gamma_0, ..., \gamma_n\right\}\right)^{-1}\right)$ if $\sum_{i=0}^{n}\gamma_i \leq 0$. Define the quantities r and q as

$$r = \mathrm{sign}(s)\sqrt{2sz + \sum_{i=0}^{n}\log\left(1-2\gamma_i s\right)}, \qquad q = s\sqrt{2\sum_{i=0}^{n}\frac{\gamma_i^2}{\left(1-2\gamma_i s\right)^2}}$$

for a constant z. The saddlepoint approximation is then

$$\Pr(Z \leq z) \approx \Phi(r) + \phi(r)\left[\frac{1}{r} - \frac{1}{q}\right]. \tag{9}$$

Given an observed value for the t-statistic t_q^{HC}, a p-value for H_0 can be calculated by taking $\gamma_i = -\left(t_q^{HC}\right)^2\lambda_i / \sum_{j=1}^{n}\lambda_j$ for $i = 1, ..., n$, finding s, r, and q, and evaluating $1 - \Pr(Z \leq 0)$ using Equation (9).

3 Summary

The various small-sample corrections described above differ in three distinct dimensions:

1. the form of the correction term used in constructing the variance estimator (i.e., HC0, HC1, HC2, HC3, HC4, HC4m, HC5);
2. how the variability of the variance estimator is approximated, with some authors proposing to calculate the degrees of freedom under a working model and others proposing to do so using adjustments to the residuals; and
3. how the p-values or critical values of the hypothesis test are calculated:
    - using a standard normal reference distribution
    - ad hoc approximation using a t distribution with $n - p$ degrees of freedom
    - Satterthwaite approximation
    - Rothenberg's Edgeworth approximation to the p-value
    - Rothenberg's Edgeworth approximation to the critical value
    - Kauermann & Carroll's approximation to the p-value
    - Kauermann & Carroll's approximation to the critical value
    - saddlepoint approximation

Approaches in each of these dimensions are interchangeable, and so we could in principle consider up to $7 \times 2 \times 8 = 112$ distinct test procedures.