1. (a) (Problem 1.1 of the notes)

1.1. (a)

$$\begin{aligned}
E(s-\sigma)^2 &= E(s-Es+Es-\sigma)^2 \\
&= E(s-Es)^2+2(Es-Es)(Es-\sigma)+(Es-\sigma)^2 \\
&= \mathrm{Var}(s)+(Es-\sigma)^2 \\
&= \sigma^2(1-c_4(n)^2)+(\sigma c_4(n)-\sigma)^2 \\
&= \sigma^2(2-2c_4(n)) \\
E(s/c_4-\sigma)^2 &= \mathrm{Var}(s/c_4(n)) \\
&= \sigma^2\left(\frac{1}{c_4^2(n)}-1\right)
\end{aligned}$$

where $c_4(n)$ is defined on page 4 of the notes. Consider $f(x) \equiv (2-2x) - (\frac{1}{x^2}-1)$. Note that $f(x) = -(1-x)^2(1+2x)/x^2 < 0$ for $0 < x < 1$. Thus $E(s-\sigma)^2 - E(s/c_4(n)-\sigma)^2 = \sigma^2 f(c_4(n)) < 0$, and we conclude that s has the smaller MSE.

1.1. (b) For simplicity, let $c_4 = c_4(n)$. Then

$$\begin{aligned}
MSE(ks) &= \mathrm{Var}(ks)+(E(ks)-\sigma)^2 \\
&= k^2\sigma^2(1-c_4^2)+(kc_4\sigma-\sigma)^2 \\
&= (k^2-2kc_4+1)\sigma^2 \\
&= \{(k-c_4)^2+1-c_4^2\}\sigma^2,
\end{aligned}$$

which is minimized when $k = c_4(n)$.

(b) (Problem 1.2 of the notes)

Note that $MSE[R/d_2(n)] = \mathrm{Var}[R/d_2(n)] = \sigma^2\left(\frac{d_3(n)}{d_2(n)}\right)^2$ and $MSE[s/c_4(n)] = \sigma^2\left(\frac{1}{c_4^2(n)}-1\right)$. It can be shown numerically, as shown in the following table, that the MSE of $s/c_4(n)$ is smaller than that of $R/d_2(n)$.

n	d_2	d_3	c_4	$MSE[R/d_2(n)]$	$MSE[s/c_4(n)]$
2	1.128	0.853	0.7979	$0.5718\sigma^2$	$0.5707\sigma^2$
3	1.693	0.888	0.8862	$0.2751\sigma^2$	$0.2733\sigma^2$
4	2.059	0.880	0.9213	$0.1827\sigma^2$	$0.1781\sigma^2$
5	2.326	0.864	0.9400	$0.1380\sigma^2$	$0.1317\sigma^2$
6	2.534	0.848	0.9515	$0.1120\sigma^2$	$0.1045\sigma^2$
7	2.704	0.833	0.9594	$0.0949\sigma^2$	$0.0864\sigma^2$
8	2.847	0.820	0.9650	$0.0830\sigma^2$	$0.0739\sigma^2$
9	2.970	0.808	0.9693	$0.0740\sigma^2$	$0.0643\sigma^2$
10	3.078	0.797	0.9727	$0.0670\sigma^2$	$0.0569\sigma^2$

(c) (Problem 1.5 of the notes)

$$\mathrm{Var}[\bar{R}/d_2(n)] = \frac{\mathrm{Var}(\bar{R})}{d_2^2(n)} = \frac{\mathrm{Var}(R)}{d_2^2(n)r} = \frac{\sigma^2 d_3^2(n)}{d_2^2(n)r}$$

Estimating σ^2 by $\tilde{\sigma}^2 = \left(\bar{R}/d_2(n)\right)^2$, we have

$$\widehat{\mathrm{Var}}[\bar{R}/d_2(n)] = \left(\frac{d_3(n)}{d_2(n)}\right)^2 \frac{1}{r}\left(\frac{\bar{R}}{d_2(n)}\right)^2.$$

Thus, a reasonable standard error for $\bar{R}/d_2(n)$ is $\bar{R}d_3(n)/(\sqrt{r}d_2^2(n))$.

2. Let $U \equiv g(X, Y, \cdots, Z)$, where $X, Y \cdots, Z$ are independent random variables with means $\mu_X, \mu_Y, \cdots, \mu_Z$ and variance-covariance matrix $\mathrm{diag}(\sigma_X^2, \sigma_Y^2, \cdots, \sigma_Z^2)$. By the delta method, it follows that

$$\mu_U = EU \approx g(\mu_X, \mu_Y, \cdots, \mu_Z)$$

and

$$\begin{aligned}\sigma_U^2 = \mathrm{Var}U &\approx \begin{pmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} & \cdots & \frac{\partial g}{\partial z}\end{pmatrix} \begin{pmatrix} \sigma_X^2 & 0 & \cdots & 0 \\ 0 & \sigma_Y^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_Z^2 \end{pmatrix} \begin{pmatrix} \frac{\partial g}{\partial x} \\ \frac{\partial g}{\partial y} \\ \vdots \\ \frac{\partial g}{\partial z}\end{pmatrix} \\ &= \left(\frac{\partial g}{\partial x}\right)^2 \sigma_X^2 + \left(\frac{\partial g}{\partial y}\right)^2 \sigma_Y^2 + \cdots + \left(\frac{\partial g}{\partial z}\right)^2 \sigma_Z^2,\end{aligned}$$

where the partial derivatives are evaluated at the point $(\mu_X, \mu_Y, \cdots, \mu_Z)$.

3. (a) Note that the variance-covariance matrix for $(s_y^2, s^2)'$ is

$$\Sigma = \begin{bmatrix} \frac{2\sigma_y^4}{n-1} & 0 \\ 0 & \frac{2\sigma_{meas}^4}{m-1} \end{bmatrix}.$$

Note also that $\hat{\sigma}_x = g(s_y^2, s^2)$, where $g(t_1, t_2) = \sqrt{\max(0, t_1 - t_2)}$, and $s = f(s_y^2, s^2)$, where $f(t_1, t_2) = \sqrt{t_2}$. Then,

$$D = \begin{bmatrix} \frac{\partial g}{\partial t_1} & \frac{\partial g}{\partial t_2} \\ \frac{\partial f}{\partial t_1} & \frac{\partial f}{\partial t_2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2\sqrt{t_1-t_2}} & -\frac{1}{2\sqrt{t_1-t_2}} \\ 0 & \frac{1}{2\sqrt{t_2}} \end{bmatrix}\Bigg|_{\sigma_y^2, \sigma_{meas}^2} = \begin{bmatrix} \frac{1}{2\sqrt{\sigma_y^2-\sigma_{meas}^2}} & -\frac{1}{2\sqrt{\sigma_y^2-\sigma_{meas}^2}} \\ 0 & \frac{1}{2\sqrt{\sigma_{meas}^2}} \end{bmatrix}.$$

By the delta method, we obtain an approximate variance-covariance matrix for $[\hat{\sigma}_x \; s]'$ as follows:

$$\begin{aligned} D\Sigma D' &= \begin{bmatrix} \frac{1}{2\sigma_x} & -\frac{1}{2\sigma_x} \\ 0 & \frac{1}{2\sigma_{meas}} \end{bmatrix} \begin{bmatrix} \frac{2\sigma_y^4}{n-1} & 0 \\ 0 & \frac{2\sigma_{meas}^4}{m-1} \end{bmatrix} \begin{bmatrix} \frac{1}{2\sigma_x} & 0 \\ -\frac{1}{2\sigma_x} & \frac{1}{2\sigma_{meas}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{\sigma_y^4}{2(n-1)\sigma_x^2} + \frac{\sigma_{meas}^4}{2(m-1)\sigma_x^2} & -\frac{\sigma_{meas}^3}{2(m-1)\sigma_x} \\ -\frac{\sigma_{meas}^3}{2(m-1)\sigma_x} & \frac{\sigma_{meas}^2}{2(m-1)} \end{bmatrix}. \end{aligned}$$

Note that $\sigma_x = 1, \sigma_{meas} = .5$ and thus $\sigma_y = \sqrt{1.25}$. Then

$$D\Sigma D' = \begin{bmatrix} \frac{1.25^2}{2(n-1)} + \frac{.25^2}{2(m-1)} & -\frac{.5^3}{2(m-1)} \\ -\frac{.5^3}{2(m-1)} & \frac{.25}{2(m-1)} \end{bmatrix}.$$

(b) Note that the determinant is a one-number summary that roughly measures how well the estimator is. The determinants of $D\Sigma D'$ are identical at the value of 0.002712674 for both $(n, m) = (5, 10)$ and $(n, m) = (10, 5)$. In fact, it can be shown that

$$\det D\Sigma D' = \frac{\sigma_y^4 \sigma_{meas}^2}{4(n-1)(m-1)\sigma_x^2}.$$

This implies that an exchange of n and m does not change the value of the determinant.

4. (a) Let $\tilde{\sigma}_x = g(R, R')$, where $g(t_1, t_2) = \sqrt{\max(0, t_1^2/d_2^2(n) - t_2^2/d_2^2(m))}$. Note that

$$E \begin{bmatrix} R \\ R' \end{bmatrix} = \begin{bmatrix} \sigma_y d_2(n) \\ \sigma_{meas} d_2(m) \end{bmatrix}$$

$$\text{Var} \begin{bmatrix} R \\ R' \end{bmatrix} = \begin{bmatrix} \sigma_y^2 d_3^2(n) & 0 \\ 0 & \sigma_{meas}^2 d_3^2(m) \end{bmatrix} \equiv \Sigma.$$

Then by the delta method, we have

$$\begin{aligned} \text{Var}\tilde{\sigma}_x &\approx \left(\frac{\partial g}{\partial t_1}\bigg|_{ER,ER'} \right)^2 \text{Var}R + \left(\frac{\partial g}{\partial t_2}\bigg|_{ER,ER'} \right)^2 \text{Var}R' \\ &= \frac{\sigma_y^4 \frac{d_3^2(n)}{d_2^2(n)} + \sigma_{meas}^4 \frac{d_3^2(m)}{d_2^2(m)}}{\sigma_y^2 - \sigma_{meas}^2}. \end{aligned}$$

Estimating σ_y and σ_{meas} by $R/d_2(n)$ and $R'/d_2(m)$, respectively, we have a standard error for $\tilde{\sigma}_x$ as

$$\sqrt{\frac{\frac{R^4 d_3^2(n)}{d_2^6(n)} + \frac{R'^4 d_3^2(m)}{d_2^6(m)}}{\frac{R^2}{d_2^2(n)} - \frac{R'^2}{d_2^2(m)}}}.$$

(b) Let $\breve{\sigma}_x = g(MSTr, MSE)$, where $g(t_1, t_2) = \sqrt{t_1/m - t_2/m}$. Note that

$$E \begin{bmatrix} MSTr \\ MSE \end{bmatrix} = \begin{bmatrix} m\sigma_x^2 + \sigma \\ \sigma \end{bmatrix}$$

$$\text{Var} \begin{bmatrix} MSTr \\ MSE \end{bmatrix} = \begin{bmatrix} \frac{2(m\sigma_x^2 + \sigma^2)^2}{r-1} & 0 \\ 0 & \frac{2\sigma^4}{r(m-1)} \end{bmatrix}.$$

Then by the delta method, we have

$$\begin{aligned}
\mathrm{Var}\check{\sigma}_x &\approx \left(\frac{\partial g}{\partial t_1}\bigg|_{EMSTr,EMSE}\right)^2 \mathrm{Var}MSTr + \left(\frac{\partial g}{\partial t_2}\bigg|_{EMSTr,EMSE}\right)^2 \mathrm{Var}MSE \\
&= \frac{(\sigma_x^2+\sigma^2/m)^2}{2(r-1)\sigma_x^2} + \frac{\sigma^4}{2r(m-1)m^2\sigma_x^2}.
\end{aligned}$$

As in (a), we obtain a standard error for $\check{\sigma}_x$ as

$$\sqrt{\frac{1}{2m(MSTr-MSE)}\left[\frac{(MSTr)^2}{r-1}+\frac{(MSE)^2}{r(m-1)}\right]}.$$

5. Fitting a one-way ANOVA, we have $MSTr = 2.533 \times 10^{-8}$ and $MSE = 1.000 \times 10^{-9}$. With $m = 2$, we then have

$$\check{\sigma}_x = \sqrt{\max(0, MSTr/m - MSE/m)} = 0.0001103$$

Using the result in problem 4(b), we obtain a standard error of the estimator as 2.7086×10^{-5}. An approximate 90% confidence interval is

$$\check{\sigma}_x \pm 1.645 s.e. = (6.5744 \times 10^{-5}, 1.5486 \times 10^{-4}).$$

The Brandon Paris program computes the confidence interval as $(7.899 \times 10^{-5}, 1.838 \times 10^{-4})$.

In this problem, specification limits of the diameter have been set with a spread of .002 inches, but our estimates of standard deviations are smaller than that. This fact makes us sceptical about the calculations using those estimates.

6. Fitting a two-way ANOVA, we have the following ANOVA table.

	Df	Sum of Sq	Mean Sq	F Value	Pr(F)
parts	9	1.303500e-06	1.448330e-07	22.2821	0.000000000
operators	2	3.782333e-06	1.891167e-06	290.9487	0.000000000
parts:operators	18	3.210000e-07	1.783300e-08	2.7436	0.007039827
Residuals	30	1.950000e-07	6.500000e-09		

$$\hat{\sigma} = \sqrt{MSE} = 8.062 \times 10^{-5}$$

$$\hat{\sigma}_{reproducibility} = \sqrt{\max\left(0, \frac{1}{mI}MSB + \frac{1}{m}\left(1-\frac{1}{I}\right)MSAB - \frac{1}{m}MSE\right)} = 0.000315$$

$$\hat{\sigma}_{parts} = \sqrt{\hat{\sigma}_\alpha^2 + \hat{\sigma}_{\alpha\beta}^2} = \sqrt{\frac{MSA-MSAB}{mJ} + \frac{MSAB-MSE}{m}} = 0.000164.$$

Andy Chiang's program yielded the following results.

			90.00 % CONFIDENCE INTERVAL	
PARAMETER	ESTIMATE	STD ERROR	LOWER LIMIT	UPP LIMIT
---------	--------	---------	-----------	---------
SIG_REPEA	0.00008062	0.00001041	0.00006674	0.00010269
SIG_REPRO	0.00031517	0.00015008	0.00018794	0.00135944
SIG_OVERAL	0.00032532	0.00014539	0.00020678	0.00136189
SIG_PARTS	0.00016381	0.00003535	0.00012085	0.00026137

The reproducibility variation is about twice as large as the part (process) variation and about four times as large as the repeatability variation.

7. (Problem 1.7 of the notes)

   (a) We assume $Y_{new} - Y_{old} \sim N(\mu_{new} - \mu_{old}, 2\sigma^2_{meas})$, where $\sigma_{meas} = .002$. We want

   $$P(\text{False positive}) = P(Y_{new} - Y_{old} > L_c|\mu_{new} = \mu_{old}) = .05$$

   $$\begin{aligned}\Rightarrow\ & P((Y_{new} - Y_{old})/(\sqrt{2}\sigma_{meas}) > L_c/(\sqrt{2}\sigma_{meas})|\mu_{new} = \mu_{old}) = .05\\ \Rightarrow\ & P(Z > L_c/(\sqrt{2}\sigma_{meas})) = .05,\ \text{where}\ \ Z \sim N(0,1)\\ \Rightarrow\ & L_c = \sqrt{2}\sigma_{meas}z_{.95} = \sqrt{2}(.002)(1.645) = 0.00465\ \text{(instrument units)}.\end{aligned}$$

   (b) We want $P(\text{Correct Detection}) = P(Y_{new} - Y_{old} > L_c|\mu_{new} - \mu_{old} = L_d) = .9$.

   $$\begin{aligned}\Rightarrow\ & P((Y_{new} - Y_{old} - L_d)/(\sqrt{2}\sigma_{meas}) > (L_c - L_d)/(\sqrt{2}\sigma_{meas})|\mu_{new} - \mu_{old} = L_d) = .9\\ \Rightarrow\ & P(Z > (L_c - L_d)/(\sqrt{2}\sigma_{meas})) = .9,\ \text{where}\ \ Z \sim N(0,1)\\ \Rightarrow\ & (L_c - L_d)/(\sqrt{2}\sigma_{meas}) = z_{.10} = -1.28\end{aligned}$$

   Hence,

   $$L_d = L_c + (\sqrt{2}\sigma_{meas})(1.28) = 0.008273\ \text{(instrument units)}.$$

   This is, $0.008273/58.2 = 0.000142$ (g/l).

8. (Problem 1.12 of the notes)
   We have $R' = .0003$ with $m = 15$ and $R = .0008$ with $n = 12$. Using the formula from the notes, we have

   $$\tilde{\sigma}_x = \sqrt{\max(0, (R/d_2(n))^2 - (R'/d_2(m))^2)} = 2.298 \times 10^{-4}.$$

9. (a) We may use $\sqrt{MSE} = 23.5811$ as $\hat{\sigma}_{meas}$. A recommended conversion formula for translating "local lab measurements" to estimated "gold standard measurements" would be obtained by inverting the regression function:

   $$\hat{y} = 41.43693 + 0.88274x \qquad \Rightarrow \qquad \hat{x} = \frac{y - 41.43693}{0.88274}.$$

(b) Note that for $y = 2000, \hat{x}_{n+1} = (2000 - 41.43693)/0.88274 = 2218.732$.

(i) The 95% prediction limits for y are given by

$$b_0 + b_1 x \pm t_{.975,12} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

To obtain a confidence interval for x, invert the prediction limits for y by solving

$$y_{n+1} = 2000 = b_0 + b_1 x \pm t_{.975,12} \cdot s \cdot \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

for x, or solving $ax^2 + bx + c = 0$, where

$$\begin{aligned} a &= \frac{b_1^2}{t^2 s^2} - \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\ b &= -\frac{2b_1(y_{n+1} - b_0)}{t^2 s^2} + \frac{2\bar{x}}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \\ c &= \frac{(y_{n+1} - b_0)^2}{t^2 s^2} - \left(1 + \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right), \end{aligned}$$

which yields a confidence interval of $(2156.085, 2282.445)$. (Note that $s = 23.5811$, $b_0 = 41.43693$, $b_1 = 0.88274$, and $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 4274776.929$.)

(ii) Using the delta method formula on page 11 of the notes, we have an approximate 95% confidence interval

$$\hat{x}_{n+1} \pm t\frac{\sqrt{MSE}}{|b_1|}\sqrt{1 + \frac{1}{n} + \frac{(\hat{x}_{n+1} - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} = (2155.579, 2281.884).$$