PI a) Note that, no matter how you assign callers to each group (std. or new procedure) the total call duration $m = x_1 + x_2 + y_1 + y_2 + y_3$ will always be the same. Therefore

$$
\begin{aligned}
T &= \frac{1}{2}(x_1 + x_2) - \frac{1}{3}(y_1 + y_2 + y_3) \\
&= \frac{1}{2}(x_1 + x_2) - \frac{1}{3}(m - (x_1 + x_2)) \\
&= \frac{1}{2}(x_1 + x_2) + \frac{1}{3}(x_1 + x_2) - \frac{1}{3}m \\
&= \frac{5}{6}(x_1 + x_2) - \frac{1}{3}m \\
&= \frac{5}{6}S - \underbrace{\frac{1}{3}m}_{\text{constant under random assignments.}}
\end{aligned}
$$

Since there is a strickly monotonic relation between T and S the two statistics are equivalent.

b)

Standard Procedure call times		s
95	145	240
95	28	123
95	47	142
95	135	230
145	28	173
145	47	192
145	135	280
28	47	75
28	135	163
47	135	182

Therefore, the dist. of S under H_0 is uniform over

$$\mathcal{S} := \{75, 123, 142, 163, 173, 182, 192, 230, 240, 280\}$$

c) We should reject the null hypothesis if s_0 is large (meaning the std. procedure displays a large wall duration). The p-value of the test is simply

$$\text{p-value} = \frac{1}{\binom{5}{2}} \sum_{s \in \mathcal{S}} 1\{s \geqslant s_0\} = \frac{2}{10} = \frac{1}{5} = 0.2$$

Therefore, there is not enough evidence to reject H_0 with 95% confidence (since $0.2 > 0.05$).

PII a) Note that $U_i = F(X_i)$ are iid r.v. with uniform dist. over $[0,1]$. Now

$$S_n = \sup_{t\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^n 1\{X_i \le t\} - \frac{1}{n}\sum_{i=1}^n 1\{-X_i \le t\} \right|$$

F is strictly monotone increasing
$$= \sup_{t\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^n 1\{F(X_i) \le F(t)\} - \frac{1}{n}\sum_{i=1}^n 1\{F(-X_i) \le F(t)\} \right|$$

(under H_0)
$$= \sup_{t\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^n 1\{U_i \le F(t)\} - \frac{1}{n}\sum_{i=1}^n 1\{1-F(X_i) \le F(t)\} \right|$$

$$= \sup_{t\in\mathbb{R}} \left| \frac{1}{n}\sum_{i=1}^n 1\{U_i \le F(t)\} - \frac{1}{n}\sum_{i=1}^n 1\{U_i \ge 1-F(t)\} \right|$$

(1)
$$= \sup_{u\in[0,1]} \left| \frac{1}{n}\sum 1\{U_i \le u\} - \frac{1}{n}\sum_{i=1}^n 1\{U_i \ge 1-u\} \right| ,$$

where the last step follows from the fact that $t \mapsto F(t)$ has range $[0,1]$ (since F is continuous). As the expression (1) doesn't involve F we conclude that S_n is distribution free.

b) Note that the GC thm. tells us that:

$$\sup_t \left| \hat{F}_n(t) - F(t) \right| \xrightarrow{P} 0 \qquad \text{as } n\to\infty$$

and
$$\sup_t \left| \hat{G}_n(t) - G(t) \right| \xrightarrow{P} 0 \qquad \text{as } n\to\infty, \text{ where}$$

$$\hat{F}_n(t) = \frac{1}{n}\sum_i 1\{X_i \le t\} \quad , \quad \hat{G}_n(t) = \frac{1}{n}\sum_i 1\{-X_i \le t\}$$

and $G(t) = 1 - F(-t)$ (the dist of $-X_i$)

X_i is a cont. r.v.

$$G(t) = P(-X_i \le t) = P(X_i \geq -t) = 1 - P(X_i < -t) \overset{\uparrow}{=} 1 - P(X_i \le -t) = 1 - F(-t).$$

Now, under H_0, we have $G(t) = F(t)$, so

$$S_n = \sup_t \left| \hat{F}_n(t) - F(t) - \hat{G}_n(t) + G(t) + \underbrace{F(t) - G(t)}_{=0} \right|$$

$$\le \sup_t \left| \hat{F}_n(t) - F(t) \right| + \sup_t \left| \hat{G}_n(t) - G(t) \right|$$

$$\xrightarrow{P} 0 \quad \text{as } n \to \infty.$$

Since $S_n \geq 0$ this means that

$$\forall_{\varepsilon > 0} : P(S_n > \varepsilon) \to 0 \quad \text{as } n \to \infty$$

However, under H_1 $\exists_{a \in \mathbb{R}} : F(a) \neq 1 - F(-a) = G(a)$.

Therefore $S_n = \sup_t \left| \hat{F}_n(t) - F(t) - \hat{G}_n(t) + G(t) + F(t) - G(t) \right|$

$$\geq \left| \hat{F}_n(a) - F(a) - \left(\hat{G}_n(a) - G(a) \right) + F(a) + F(-a) - 1 \right|$$

$$\geq \left| F(a) + F(-a) - 1 \right| - \left| \hat{F}_n(a) - F(a) - \left(\hat{G}_n(a) - G(a) \right) \right|$$

$$\xrightarrow{P} \left| F(a) + F(-a) - 1 \right| \quad \text{as } n \to \infty$$

In conclusion, if $\varepsilon = \frac{|F(a) + F(-a) - 1|}{2}$

$$P(S_n > \varepsilon) \to P\left(|F(a) + F(-a) - 1| > \varepsilon \right) = 1 \quad \text{as } n \to \infty,$$

as we wanted to show.

c) Since we showed in a) that S_n is distribution-free we can generate samples under the null hyp. very easily:

- Generate B iid samples of size n from a symmetric distribution (e.g. $\mathrm{Unif}([-1,1])$):

$$\{X^*_{1,j}, \dots, X^*_{n,j}\} \quad , \quad j \in \{1, \dots, B\}$$

- Compute $$S^*_{n,j} = \sup_t \left| \frac{1}{n} \sum_{i=1}^{n} 1\{X^*_{i,j} \leq t\} - \frac{1}{n} \sum_{i=1}^{n} 1\{-X^*_{i,j} \leq t\} \right|$$

for $j \in \{1, \dots, B\}$

- Calculate the p-value as $= \frac{1}{B} \sum_{b=1}^{B} 1\{S^*_{n,b} \geq s_n\}$,

where s_n is the observed test statistic.

Remarks: By taking B large this computation of the p-value approaches the true p-value with arbitrary accuracy.

The parametric bootstrap approach discussed in the notes cannot be used here, as the null hyp. is not parametric. Keep in mind that, to compute the p-value, you need to generate samples under H_0. A sensible bootstrap approach would be to generate samples from

$$\frac{\hat{F}_n(t) - \hat{F}_n(-t) + 1}{2},$$

the symmetrized ECDF.

d) If we knew t_0, then a natural modification of the test statistic would be

$$\sup_t \left| \frac{1}{n} \sum_i 1\{X_i - t_0 \leq t\} - \frac{1}{n} \sum_i 1\{-X_i + t_0 \leq t\} \right|$$

Since we don't know it, we must plug-in an estimate. For instance, the sample median $\hat{t}_0 = X_{(\frac{n}{2})}$. The proposed test statistic is therefore

$$\tilde{S}_n = \sup_t \left| \frac{1}{n} \sum_i 1\{X_i \leq t + \hat{t}_0\} - \frac{1}{n} \sum_i 1\{-X_i \leq t - \hat{t}_0\} \right|.$$

Because of the plugged-in parameter est. it is highly unlikely that $\hat{S}_n$ will be distribution free.

PIII a) This was given both in class and in the book of Wasserman:

$$l_j(x) = \frac{K\left(\frac{x - x_j}{h}\right)}{\sum_{k=1}^{n} K\left(\frac{x - x_k}{h}\right)}$$

b)

$L_{ij} = l_j(x_i)$. Note that

$$K\left(\frac{x_i - x_j}{h}\right) = K\left(\frac{i/n - j/n}{2/n}\right) = K\left(\frac{i-j}{2}\right)$$

$$= c\,\frac{\left|\frac{i-j}{2}\right|\left(1 - \left|\frac{i-j}{2}\right|\right)}{1 + \left(\frac{i-j}{2}\right)}\; 1\left\{\left|\frac{i-j}{2}\right| \leq 1\right\}$$

Since $i, j \in \{1, \dots, n\}$ we have

$$K\left(\frac{x_i - x_j}{h}\right) = c \begin{cases} 0 & i = j \\ c\,\frac{\frac{1}{2}\frac{1}{2}}{1+\frac{1}{4}} & |i-j| = 1 \\ 0 & |i-j| > 1 \end{cases}$$

$$= c\,\frac{4}{5}\, 1\{|i-j| = 1\}$$

Therefore, the matrix A: $A_{ij} = K\left(\frac{x_i - x_j}{2/n}\right) =$

$$A = \begin{bmatrix} 0 & c\frac{4}{5} & 0 & \cdots & & 0 \\ c\frac{4}{5} & 0 & c\frac{4}{5} & 0 & \cdots & 0 \\ 0 & c\frac{4}{5} & 0 & c\frac{4}{5} & \cdots & 0 \\ & & & \vdots & & \\ 0 & & \cdots & & c\frac{4}{5} & 0 \end{bmatrix}$$

Finally, the matrix $L_{ij} = l_j(x_i) = \frac{A_{ij}}{\sum_{k=1}^{n} A_{ik}}$, so we just have to normalize the rows to sum to one:

$$L = \begin{bmatrix} 0 & 1 & 0 & \cdots & & & 0 \\ 1/2 & 0 & 1/2 & 0 & \cdots & & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & \cdots & 0 \\ & & & \vdots & & & \\ 0 & & & & & 1 & 0 \end{bmatrix}$$

c) $\nu = \text{trace}(L) = 0$

d) $\tilde{\nu} = \sum_{i,j} L_{ij}^2 = 1 + 1 + \frac{1}{4} \times 2 \times (n-2) = \frac{n-2}{2} + 2 = \frac{n}{2} + 1$

this is very sensible, as this estimator essentially averages two points

e) $\mathbb{E}\left[R(h)\right] = \mathbb{E}\left[\frac{1}{n}\sum\left(Y_i - r(x_i) + r(x_i) - \hat{r}_n^{(h)}(x_i)\right)^2\right]$

$$= \mathbb{E}\left[\frac{1}{n}\sum_i \left(Y_i - r(x_i)\right)^2 + \frac{1}{n}\sum_i\left(\hat{r}_n^{(h)}(x_i) - r(x_i)\right)^2 + \frac{2}{n}\sum_{i=1}^{n}\left(Y_i - r(x_i)\right)\left(r(x_i) - \hat{r}_n^{(h)}(x_i)\right)\right]$$

$$= \frac{1}{n}\sum_i \mathbb{E}\left[\left(Y_i - r(x_i)\right)^2\right] + \frac{1}{n}\sum_i \mathbb{E}\left[\left(\hat{r}_n^{(h)}(x_i) - r(x_i)\right)^2\right] + \frac{2}{n}\sum_i \mathbb{E}\left[\underbrace{\left(Y_i - r(x_i)\right)}_{A}\underbrace{\left(r(x_i) - \hat{r}_n^{(h)}(x_i)\right)}_{B}\right]$$

not a function of Y_i, since $L_{ii} = 0$, therefore A and B are independent!

$$= \sigma^2 + \frac{1}{n}\sum_i \mathbb{E}\left[\left(r(x_i) - \hat{r}_n^{(h)}(x_i)\right)^2\right]$$

$$+ \underbrace{\frac{2}{n}\sum_{i=1}^{n} \underbrace{\mathbb{E}\left[Y_i - r(x_i)\right]}_{=0} \mathbb{E}\left[\left(r(x_i) - \hat{r}_n^{(h)}(x_i)\right)\right]}_{=0}.$$

f) This is just the CV approach, because indeed $R(h)$ is the cross-validation score since $L_{ii} = 0$ (see Thm. 5.34 of W). So this is quite reasonable.

PIV. a) The "parameter" of interest is S. The log-likelihood is simply given by

$$\ell(S) = \sum_{i=1}^{n} \log f_{Y_i;S}(Y_i;S) =$$

$$= \sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(Y_i - \mu 1\{i\in S\})^2}{2\sigma^2}}$$

$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(Y_i - \mu 1\{i\in S\}\right)^2$$

Therefore $\hat{S} = \arg\max_{S\in \mathcal{C}_s} \ell(S)$

$$= \arg\max_{S\in\mathcal{C}_s} -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_i \left(Y_i - \mu 1\{i\in S\}\right)^2$$

$$= \arg\min_{S\in\mathcal{C}_s} \sum_{i=1}^{n}\left(Y_i - \mu 1\{i\in S\}\right)^2$$

b) $\frac{1}{n}\sum_{i=1}^{n}\left(\mu 1\{i\in\hat{S}\} - \mu 1\{i\in S^*\}\right)^2$

$$= \frac{\mu^2}{2}\sum_{i=1}^{n}\left(1\{i\in\hat{S}\} - 1\{i\in S^*\}\right)^2$$

$$= \frac{\mu^2}{2}\sum_i \begin{cases} (0-0)^2 & i\notin\hat{S},\ i\notin S^* \\ (0-1)^2 & i\notin\hat{S},\ i\in S^* \\ (1-0)^2 & i\in\hat{S},\ i\notin S^* \\ (1-\phi)^2 & i\in\hat{S},\ i\in S^* \end{cases} = \frac{\mu^2}{2}\sum_i 1\{i\in S^*\Delta\hat{S}\}$$

$$= \frac{\mu^2}{2} d(\hat{S}, S^*)$$

c) Note that we are precisely in a regression setting, where $r^*(x_i) = \mu 1\{i\in S^*\}$. As we seen, the estimator in the theorem corresponds to the maximum likelihood, provided $c(r) = \text{const.}$

So we can take $c(S) = \log_2 |\mathcal{C}_s| = \log_2 \binom{n}{s}$, since

$$\sum_{S \in \mathcal{C}_s} 2^{-c(S)} = \sum_{S \in \mathcal{C}_s} \frac{1}{|\mathcal{C}_s|} = 1.$$

Now $$\hat{S}_n = \arg\min_{S \in \mathcal{C}_s} \Big\{ \frac{1}{n} \sum_{i=1}^n (Y_i - \mu 1\{i \in S\})^2 + \underbrace{\frac{4\sigma^2 c(S) \log 2}{n}}_{\text{not a function of } S} \Big\}$$

$$= \arg\min_{S \in \mathcal{C}_s} \sum_{i=1}^n (Y_i - \mu 1\{i \in S\})^2.$$

The theorem tells us that

$$\frac{1}{n} \sum_i \mathbb{E}\left[(\mu 1\{i \in \hat{S}\} - \mu \{i \in S^*\})^2 \right]$$
$$\leq 2 \min_{S \in \mathcal{C}_s} \Big\{ \frac{1}{n} \sum (\mu 1\{i \in S\} - \mu 1\{i \in S^*\})^2 + \frac{4\sigma^2 c(S) \log 2}{n} \Big\}$$

$$(=) \frac{\mu^2}{n} \mathbb{E}\left[d(\hat{S}, S^*) \right] \leq 2 \min_{S \in \mathcal{C}_s} \Big\{ \frac{\mu^2}{n} d(S, S^*) + \frac{4\sigma^2 c(S) \log 2}{n} \Big\}$$

$$(=) \frac{\mu^2}{n} \mathbb{E}\left[d(\hat{S}, S^*) \right] \leq \frac{8\sigma^2}{n} \log_2 |\mathcal{C}_s| \log 2$$

$$(=) \mathbb{E}\left[d(\hat{S}, S^*) \right] \leq \frac{8\sigma^2}{\mu^2} \log |\mathcal{C}_s| = \frac{8\sigma^2}{\mu^2} \log \binom{n}{s}$$
$$\leq \frac{8\sigma^2}{\mu^2} \log n^s$$
$$\leq \frac{8\sigma^2}{\mu^2} s \log n$$

d) As done in class, we want to find the "best" set S from the class

$$\mathcal{C} = \bigcup_{s=1}^{n} \mathcal{C}_s .$$

Note that the sets $\mathcal{C}_s$ are disjoint, so it is quite easy to proceed. As done in the class and lecture notes take

$$c(S) = |S| + \log_2 |\mathcal{C}_{|S|}| \text{ , where } |S| \text{ denotes the}$$

cardinality of S. Clearly

$$\sum_{S\in\mathcal{C}} 2^{-c(S)} = \sum_{s=1}^{n} \sum_{S\in\mathcal{C}_s} 2^{-s-\log_2|\mathcal{C}_s|} = \sum_{s=1}^{n} 2^{-s} \sum_{S\in\mathcal{C}_s} \frac{1}{|\mathcal{C}_s|}$$

$$= \sum_{s=1}^{n} 2^{-s} = 1 .$$

Therefore, we can use the estimator

$$\hat{S} = \arg\min_{S\in\mathcal{C}} \left\{ \frac{1}{n} \sum_i \left(Y_i - \mu 1\{i\in S\} \right)^2 + \frac{4\sigma^2 (|S| + \log_2 |\mathcal{C}_{|S|}|) / \log 2}{n} \right\}$$

From the theorem

$$\frac{\mu^2}{n} \mathbb{E}\left[d(\hat{S}, S^*) \right] \leq 2 \min_{S\in\mathcal{C}} \left\{ \frac{\mu^2}{n} d(S,S^*) + \frac{4\sigma^2 (|S| + \log_2 |\mathcal{C}_{|S|}|) / \log 2}{n} \right\}$$

$$\leq \frac{2\mu^2}{n} d(S^*,S^*) + \frac{8\sigma^2 \left(|S^*| + \log_2 \binom{n}{|S^*|}\right) / \log 2}{n}$$

so $$\mathbb{E}\left[d(\hat{S}, S^*) \right] \leq \frac{8\sigma^2}{\mu^2} \left(|S^*| \log 2 + |S^*| \log n \right) \leq \frac{8\sigma^2}{\mu^2} |S^*| (\log 2 + \log n)$$