All tests always have the hypothesis that we want to test. In this case, the null hypothesis is $H_0: \mu_{NB} = \mu_{BR}$, given a channel and a time. Also we want to calculate the 95% confidence intervals for population mean $\mu_{NB}$ and the 95% confidence intervals for the mean difference of $\mu_{NB}$ and $\mu_{BR}$.

1. **t-test**

Suppose we have two samples: $NB_1, \ldots, NB_{n_{NB}}$ and $BR_1, \ldots, BR_{n_{BR}}$ $n_{NB}$ and $n_{BR}$. $NB_i \sim N(\mu_{NB}, \sigma_{NB})$ and $BR_j \sim N(\mu_{BR}, \sigma_{BR})$. for $i = 1, \ldots, n_{NB}$ and $j = 1, \ldots, n_{BR}$. Then we can calculate the sample means and standard deviations such that

$$\hat{\mu}_{NB} = \overline{NB}, \quad \hat{\sigma}_{NB} = s_{NB},$$

$$\hat{\mu}_{BR} = \overline{BR}, \quad \hat{\sigma}_{BR} = s_{BR},$$

$$\hat{\mu}_p = \frac{n_{NB}\overline{NB} + n_{BR}\overline{BR}}{n_{NB} + n_{BR}}, \quad \hat{\sigma}_p^2 = \frac{SS_{NB} + SS_{BR}}{v_{NB} + v_{BR}}, \quad v = v_{NB} + v_{BR},$$

where $SS_{NB} = \sum_{i=1}^{n_{NB}} (NB_i - \overline{NB})^2$, $SS_{BR} = \sum_{j=1}^{n_{BR}} (BR_j - \overline{BR})^2$, $v_{NB} = n_{NB} - 1$, and $v_{BR} = n_{BR} - 1$.

Then 95% CI for population mean $\mu_{NB}$ is

$$\overline{NB} \pm t_{(0.975,v)} \sqrt{\frac{s_p^2}{n_{NB}}}$$

where $t_{(0.975,v)} \approx 2$.

The 95% CI for $\mu_{NB} - \mu_{BR}$ is

$$\overline{NB} - \overline{BR} \pm t_{(0.975,v)} s_{NB - BR}$$

If $H_0$ is not rejected then the 95% CI for pooled mean $\mu$ is

$$\hat{\mu}_p \pm t_{(0.975,v)} \sqrt{\frac{s_p^2}{n_1 + n_2}}$$

The above statistics can be obtained by t test functions in MATLAB or R.

2. **R code**

```R
NB <- rnorm(304) ; BR <- rnorm(53) ;
pooled <- c(NB, BR)
mean_p <- (304 * mean(NB)+53*mean(BR))/(304+53)
s_p <- sqrt((304 * sd(NB)^2 + 53*sd(BR)^2)/(304+53-2))
mean(pooled) ; mean_p ; sd(pooled) ; s_p
df_p <- (var(NB)^2/304 + var(BR)/53)^2/((var(NB)/304)^2/303 + (var(BR)/53)^2/52)
t.test(NB, BR) ;
```
wilcox.test(NB, BR, conf.int = TRUE, conf.level = 0.95);

# if H_0 is not rejected
# 95% CI for mu_NB is
# c(mean(NB) - qt(0.975, df_p) *sd(NB)/sqrt(304), mean(NB) + qt(0.975, df_p) *sd(NB)/sqrt(304))
# 95% CI for mu_BR is
# c(mean(BR) - qt(0.975, df_p) *sd(BR)/sqrt(53), mean(BR) + qt(0.975, df_p) *sd(BR)/sqrt(53))
# 95% CI for the mean difference in mu_NB and mu_BR is
# c(mean(NB)-mean(BR) - qt(.975, df_p) *s_p*sqrt(1/304+1/53), 0,
# mean(NB)-mean(BR) + qt(.975, df_p) *s_p*sqrt(1/304+1/53))

• Permutation test

1. Suppose we have \(NB_1, \ldots NB_{304}\) and \(BR_1, \ldots BR_{53}\), given a time and a channel. Let’s \(n_{NB} = 304\) and \(n_{BR} = 53\). Suppose \(NB_k \sim (\mu_{NB}, \sigma_{NB})\) and \(BR_l \sim (\mu_{BR}, \sigma_{BR})\), for \(k = 1, \ldots, n_{NB}\) and \(l = 1, \ldots, n_{BR}\).

(a) Test \(H_0: \mu_{NB} = \mu_{BR}\).

(b) Instead perform the all permutation case, just random sample, say 1000 samples: \(sample^{(i)}, \ldots, sample^{(1000)}\) and each \(sample^{(i)}\) contains 53 numbers from NBs. Hence \(sample^{(i)} = \{NB_1^{(i)}, \ldots, NB_{53}^{(i)}\}\) for \(i = 1, \ldots, 1000\).

(d) \(mean(sample^{(i)}) = \frac{1}{53} \sum_{j=1}^{53} NB_j^{(i)}\). Let’s say \(mSam_i = mean(sample^{(i)})\), for \(i = 1, \ldots, 1000\). Then \(mSam_i \approx N(\mu_{mSam}, \sigma_{mSam}) \approx N(\mu_{NB}, \frac{\sigma_{NB}}{\sqrt{53}})\) for \(i = 1, \ldots, 1000\).

(e) Hence \(\hat{\sigma}_{mSam} = \frac{\hat{\sigma}_{NB}}{\sqrt{53}}\) and \(\hat{\mu}_{mSam} = \hat{\mu}_{NB} = \bar{NB}\).

(f) Hence 95% Confidence Interval for population mean, \(\mu_{BR}\) is either \(\bar{NB} \pm 2\hat{\sigma}_{NB} \sqrt{\frac{1}{53}}\) or \(\bar{mSam} \pm 2\hat{\sigma}_{mSam}\).

(g) # By permutations,
\[
\text{mSam} <- \text{rep}(0, 1000)
\]
\[
\text{for}(i \text{ in } 1:1000)\{
\text{mysample} <- \text{sample}(NB, 53)
\text{mSam}[i] <- \text{mean(mysample)}
\}
\]
\[
c(\text{mean(mSam)} - 2 * \text{sd(mSam)}, \text{mean(BR)}, \text{mean(mSam)} + 2 * \text{sd(mSam)}))
\]
\[
c(\text{mean(NB)} - 2 * \text{sd(NB)}/\text{sqrt}(53), \text{mean(BR)}, \text{mean(NB)} + 2 * \text{sd(NB)}/\text{sqrt}(53))
\]

2. Bootstrap

(a) Test \(H_0: \mu_{NB} = \mu_{BR}\).

(b) Pool all observed data together: Let \(Pooled = \{NB_1, \ldots, NB_{n_{NB}}, BR_1, \ldots, BR_{n_{BR}}\}\)

(c) Sample \(n_{BR}\) numbers from \(n_{NB} + n_{BR}\) Pool’s. In this case, we have \(357\choose 304\) = \(357\choose 53\) = 7.858354e + 63 samples, which is a lot of cases to permute.
(d) Instead perform the all permutation case, just random sample, say 1000 samples: \(sample^{(1)}, \ldots, sample^{(1000)}\) and each \(sample^{(i)}\) contains two sets that one contains 304 numbers from Pool’s and the other 53 numbers. Hence \(sample^{(i)} = \{sample^{(i)}_{NB}, sample^{(i)}_{BR}\}\), \(sample^{(i)}_{NB} = \{NB_1^{(i)}, \ldots, NB_{304}^{(i)}\}\) and \(sample^{(i)}_{BR} = \{BR_1^{(i)}, \ldots, BR_{53}^{(i)}\}\) for \(i = 1, \ldots, 1000\).

(e) \(mdiff_i = \text{mean}(sample^{(i)}_{NB}) - \text{mean}(sample^{(i)}_{BR})\), for \(i = 1, \ldots, 1000\). Then we can calculate get the density for \(mdiff\) and the p-value. Our observed \(mdiff\) is \(NB - BR\). The p-value is

\[
\frac{\text{number of } (mdiff \geq \text{observed } mdiff)}{\text{number of simulations}}.
\]

(f) Hence 95% Confidence Interval for population mean difference \(\overline{mdiff} \pm 2SE\).

(g) # By Bootstrap,

\[
\text{Pool} \leftarrow \text{c(NB, BR)}
\]

\[
n_{NB} \leftarrow \text{length(NB)} ; n_{BR} \leftarrow \text{length(BR)} ; n_{total} \leftarrow n_{NB} + n_{BR};
\]

\[
\text{obsdiff} \leftarrow \text{mean(NB)} - \text{mean(BR)} ;
\]

\[
\text{mdiff} \leftarrow \text{rep}(0, 1000)
\]

\[
\text{for}(i \text{ in } 1:1000)\{
\]

\[
\text{NB_index} \leftarrow \text{sample(n_total, n_NB)};
\]

\[
\text{sam}_{NB} \leftarrow \text{Pool[NB_index]} ;
\]

\[
\text{sam}_{BR} \leftarrow \text{Pool[-NB_index]} ;
\]

\[
\text{mdiff}[i] \leftarrow \text{mean(sam}_{NB}) - \text{mean(sam}_{BR})
\]

\[
}\)

# p-value

\[
\text{length(mdiff[ abs(mdiff) > abs(obsdiff) ] )/1000}
\]

#confidence interval

\[
\text{c(mean(mdiff) - 2*sd(mdiff), mean(mdiff) + 2*sd(mdiff))}
\]