Repeated Significance Testing

Werner Brannath

VO ”Sequential and Adaptive Designs”,
University Bremen

Fixed size sample z-test (reminder)

Two-sided z-test with fixed sample size

- $X_1, \ldots, X_n$ stochastically independent with $X_i \sim N(\mu, \sigma^2)$.
- For some fixed μ_0 (e.g. $\mu_0 = 0$) and $0 < \alpha < 1$ ($\alpha = 0.05$):

  we test $H_0 : \mu = \mu_0$ versus $H_0 : \mu \neq \mu_0$ at significance level α.
- Given the sample mean $\overline{X} := \sum_{i=1}^{n} X_i/n$ we use the test statistics

$$Z = \sqrt{n}\left(\overline{X} - \mu_0\right)\Big/\sigma \qquad \textit{(z-score)}$$

- Under H_0 we have $Z \sim N(0,1)$ and so the rejection region

$$\mathcal{R} := \left\{|Z| \geq z_{\alpha/2}\right\} \qquad \text{with} \qquad z_{\alpha/2} := \Phi^{-1}(1 - \alpha/2)$$

  has probability α under H_0 , i.e. the type I error rate is α.

Power and sample size for the z-test

- For $\Delta := \mu - \mu_0 \neq 0$ we have that

$$Z \sim N(\sqrt{n}\,\delta\,, 1) \quad \text{with} \quad \delta := \Delta/\sigma \text{ (relative effect)}$$

- We fix $0 < \beta < 1$ (type II error) and $\Delta_1 > 0$ and choose n such that:

$$\mathbf{P}_{\sqrt{n}\delta_1}\left(Z \geq z_{\alpha/2}\right) = 1 - \beta \quad \text{with} \quad \delta_1 := \Delta_1/\sigma$$

- This leads to the sample size formula:

$$n = (z_\beta + z_{\alpha/2})^2/\delta_1^2.$$

- n guarantees power $1 - \beta$ (type II error β) for all $\mu \geq \mu_1$.

Remarks on power and sample size

- The non-centrality parameter (ncp) $\vartheta := \sqrt{n}\delta$ uniquely determines the distribution of $Z \sim N(\vartheta, 1)$, and the ncp-value

  $$\vartheta_\beta = z_\beta + z_{\alpha/2}$$

  gives power

  $$\mathbf{P}_{\vartheta_\beta}\left(Z \geq z_{\alpha/2}\right) = 1 - \beta.$$

  Hence, the sample size for power $1 - \beta$ at μ_1 is $n = (\vartheta_\beta/\delta_1)^2$

- We have focused here on the "one-sided power" to reject H_0 by $Z \geq z_{\alpha/2}$ in favour of $\mu > \mu_0$ when $\mu \geq \mu_0 + \Delta_1$.

  Similar arguments lead to the same n for the "one-sided power" to reject H_0 by $Z \leq -z_{\alpha/2}$ in favour of $\mu < \mu_0$ when $\mu \leq \mu_0 - \Delta_1$.

General z-test

- We test for some parameter $\theta \in \mathbb{R}$ and parameter value $\theta_0 \in \mathbb{R}$:

  $$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0$$

  with a test statistic $Z \sim N(\sqrt{I}(\theta - \theta_0), 1)$ where I is the "information" on θ in our sample of the given size.

- **Examples:**
  - One-sample z-test: $I = n/\sigma^2$
  - Balanced two-sample z-test: $I = (1/2)(n/\sigma^2)$, n the sample size per group and σ^2 the common variance.
  - General two-sample z-test: $I = 1\big/\big(\sigma_1^2/n_1 + \sigma_2^2/n_2\big)$, n_i sample size and σ_i^2 variance of group $i = 1, 2$.
  - Binomial endpoint with probability p: $I = n/\big[p(1-p)\big]$
  - Wald test (general): $\hat{\theta}$ MLE for θ, I Fisher information, $Z = \sqrt{I}(\hat{\theta} - \theta_0)$ (MLE = Maximum Likelihood Estimate)

Basic theory for group sequential designs

Cumulative means and z-scores

Analysis scheme (IA = Interim Analysis):

$$\underbrace{\underbrace{\underbrace{\underbrace{X_{11},\ldots,X_{1n_1}}_{\text{1. IA with } N_1 := n_1 \text{ obs.}}\;,\underbrace{X_{21},\ldots,X_{2n_2}}_{n_1 \text{ obs.}}}_{\substack{\text{2. IA with } N_2 := n_1+n_2 \text{ observations}\\ \ldots\quad\ldots\quad\ldots}}\;,\ldots,\underbrace{X_{k1},\ldots,X_{kn_k}}_{n_k \text{ obs.}}}_{\substack{\text{k. IA with } N_k := n_1+n_2+\cdots+n_k \text{ observations}\\ \ldots\quad\ldots\quad\ldots}}\ldots,\underbrace{X_{K1},\ldots,X_{Kn_K}}_{n_K \text{ obs.}}}_{\text{Final analysis with } N_K := n_1+\ldots+n_K \text{ observations}}$$

Cumulative z-score:

Let $\overline{X}^{(k)} := \left(X_{11}+\cdots+X_{kn_k}\right)/N_k$ be the overall mean up to stage k and use at interim (final) analysis $k = 1,\ldots,K$ the test statistics:

$$Z_k^* := \sqrt{N_k}\left(\overline{X}^{(k)} - \mu_0\right)/\sigma \sim N(\sqrt{N_k}\delta, 1)$$

Stage-wise means and z-scores

Sampling scheme

$$\underbrace{X_{11},\ldots,X_{1n_1}}_{n_1 \text{ observations}},\underbrace{X_{21},\ldots,X_{2n_2}}_{n_2 \text{ observations}},\ \ldots,\ \underbrace{X_{k1},\ldots,X_{kn_k}}_{n_k \text{ obs.}},\ \ldots,\ \underbrace{X_{K1},\ldots,X_{Kn_K}}_{n_K \text{ observations}}.$$

Stage-wise z-score: $Z_k := \sqrt{n_k}\,\left(\overline{X}_k - \mu_0\right)/\sigma$

where $\overline{X}_k := \left(X_{k1} + \ldots + X_{kn_k}\right)/n_k$ is stage-wise mean of stage k.

Cumulative z-score: One can easily calculate that

$$Z_k^* = \left(\sqrt{n_1}Z_1 + \cdots + \sqrt{n_k}Z_k\right)\Big/\sqrt{N_k} \;=\; w_{11}Z_1 + \cdots + w_{1k}Z_k \tag{1}$$

where $w_{ik} := \sqrt{n_i/N_k}$ for $i = 1,\ldots,k$.

Distribution of stage-wise z-scores

- Due to the independence of the observations, the stage-wise z-scores $Z_1, \ldots, Z_k$ are stochastically independent.
- For all $1 \leq i < j \leq K$: $\quad \mathrm{Cov}(Z_i, Z_j) = 0$

$$\text{and by (1):} \quad \mathrm{Cov}(Z_i^*, Z_j^*) = \mathrm{Cor}(Z_i^*, Z_j^*) = \sqrt{N_i/N_j}$$

- In summary, we have that $Z_1^*, \ldots, Z_k^*$ is *multivariate normal* with means $\sqrt{N_1}\delta, \ldots, \sqrt{N_k}\delta$ and covariance matrix

$$\Sigma := \begin{pmatrix} 1 & \sqrt{N_1/N_2} & \cdots & \sqrt{N_1/N_{K-1}} & \sqrt{N_1/N_K} \\ \sqrt{N_1/N_2} & 1 & \cdots & \sqrt{N_2/N_{K-1}} & \sqrt{N_2/N_K} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sqrt{N_1/N_K} & \sqrt{N_1/N_K} & \cdots & \sqrt{N_{K-1}/N_K} & 1 \end{pmatrix}$$

Properties of the multivariate normal distribution

Let Y be *multivariate normal* with mean vector $\vartheta = (\theta_1, \ldots, \theta_K)^t$ and covariance matrix $\Sigma = (\Sigma_{ij})_{i,j=1,\ldots,K}$, i.e.

$$Y \sim MNV(\theta, \Sigma)$$

Then:

- For all $k = 1, \ldots, K$: $\quad Y_k \sim N(\theta_k, \Sigma_{kk})$
- For every $\mathbf{a} \in \mathbb{R}^K$ and $(K \times K)$-matrix $\mathbf{A}$:

$$\mathbf{a} + \mathbf{A}Y \sim MNV(\mathbf{a} + \mathbf{A}\theta, \mathbf{A}\Sigma\mathbf{A}^t)$$

- If the inverse Σ^{-1} exists, then Y has the joint density

$$f(y) := (2\pi)^{-K/2} \det(\Sigma)^{-1/2} e^{-(y-\theta)^t \Sigma^{-1} (y-\theta)/2}$$

Calculations for multivariate normal distribution

- Calculation of density and joint cumulative distribution in R:

  R-packages `mvtnorm` and `mnormt`

- Standard bivariate normal distribution function in SAS:

  `PROBNRM(x,y,r)`

Type I error rate calculation with two stages

Calculation for un-adjusted repeated tested

We assume iid $X_{lj} \sim N(\mu, \sigma^2)$, $j = 1, \ldots, n_l$, $l = 1, 2$, and test

$$H_0 : \mu = \mu_0 \qquad \text{against} \qquad H_1 : \mu \neq \mu_0$$

We test H_0 at each stage $1 \leq k \leq 2$ with the un-adjusted rejection rule ($Z_k^* =$ cumulative z-score):

$$|Z_k^*| \geq z_{\alpha/2}.$$

How to calculate the type I error?

Un-adjusted repeated testing for $K = 2$ (Slide I)

With two stages (only) the type I error rate is

$$\mathbf{P}\Big(|Z_1^*| \geq z_{\alpha/2} \text{ or } |Z_2^*| \geq z_{\alpha/2}\Big)$$

$$=$$

$$\underbrace{\mathbf{P}\Big(|Z_1^*| \geq z_{\alpha/2}\Big)}_{=\alpha} + \underbrace{\mathbf{P}\Big(|Z_1^*| < z_{\alpha/2},\ |Z_2^*| \geq z_{\alpha/2}\Big)}_{=:B \text{ (type I error inflation)}} > \alpha$$

By symmetry under H_0, i.e. $(Z_1^*, Z_2^*) \sim (Z_1^*, -Z_2^*)$, we obtain

$$B = 2\,\mathbf{P}\Big(|Z_1^*| < z_{\alpha/2},\ Z_2^* \leq -z_{\alpha/2}\Big)$$

Un-adjusted repeated testing for $K = 2$ (Slide II)

We can continue, calculating

$$\mathbf{P}\Big(|Z_1^*| < z_{\alpha/2},\ Z_2^* \le -z_{\alpha/2}\Big) =$$

$$\mathbf{P}\Big(Z_1^* < z_{\alpha/2},\ Z_2^* \le -z_{\alpha/2}\Big) - \mathbf{P}\Big(Z_1^* < -z_{\alpha/2},\ Z_2^* \le -z_{\alpha/2}\Big) =$$

$$F(z_{\alpha/2}, -z_{\alpha/2}) - F(-z_{\alpha/2}, -z_{\alpha/2})$$

where F is the joint distribution function of (Z_1^*, Z_2^*). Recall that

$$(Z_1^*, Z_2^*) \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \sqrt{N_1/N_2} \\ \sqrt{N_1/N_2} & 1 \end{pmatrix}\right).$$

As mentioned, F is implemented in `R` and `SAS`

Exercise: Calculate the type I error for $\alpha = 0.05$.

Adjusted repeated significance testing ($K = 2$)

- Obviously, we need to adjust the critical values.
- We could determine $u > z_{\alpha/2}$ numerically, such that

$$\mathbf{P}\Big(|Z_1^*| \geq u \text{ or } |Z_2^*| \geq u\Big) = \alpha$$

- **Example 1:** For $\alpha = 0.05$ we obtain $u = 2.178$.
- **Exercise:** Verify this u and calculate the u for $\alpha = 0.01$
- We could also choose different $u_1, u_2 > z_{\alpha/2}$ such that

$$\mathbf{P}\Big(|Z_1^*| \geq u_1 \text{ or } |Z_2^*| \geq u_2\Big) = \alpha \qquad (2)$$

- **Problem:** There is **no** unique way for choosing u_1 and u_2 !
- We can fix $\gamma > 0$, define $u_2 = \gamma u_1$ and determine u_1 to meet (2).

Group Sequential Designs

with K stages

General notation for GSD

GSD with $K-1$ interim analyses (I A) and one final analysis.

- As before, $Z_k^* =$ cumulative z-score at stage $k = 1, \ldots, K$.
- For each stage $k = 1, \ldots, K$ we pre-define:
  - a *continuation region* $\mathcal{C}_k^*$, whereby $\mathcal{C}_K^* = \emptyset$;
  - a *rejection region* $\mathcal{R}_k^* \subseteq \mathbb{R} \setminus \mathcal{C}_k^*$;
  - and the *acceptance region* $\mathcal{A}_k^* = \mathbb{R} \setminus (\mathcal{C}_k^* \cup \mathcal{R}_k^*)$.
- Note that $\mathcal{A}_K^* = \mathbb{R} \setminus \mathcal{R}_K^*$.
- At every IA $k \leq K-1$, we continue to the next stage, if $Z_k^* \in \mathcal{C}_k^*$.
- At all stages $k = 1, \ldots, K$, we stop if $Z_k^* \notin \mathcal{C}_k^*$ and
  - reject H_0 if $Z_k^* \in \mathcal{R}_k^*$,
  - retain H_0 if $Z_k^* \in \mathcal{A}_k^*$.

Examples

- **Example 1:** $K = 2$ equally sized stages ($n_1 = n_2$) with

$$\mathcal{A}_1^* = \emptyset, \qquad \mathcal{C}_1^* = \mathcal{A}_2^* = (-2.178, 2.178)$$

and

$$\mathcal{R}_1^* = \mathcal{R}_2^* = (-\infty, -2.178] \cup [2.178, \infty)$$

- **Example 2:** $K = 2$ equally sized stages ($n_1 = n_2$) with

$$\mathcal{A}_1^* = (-1, 1), \qquad \mathcal{A}_2^* = (-2.178, 2.178),$$

$$\mathcal{C}_1^* = (-2.178, -1] \cup [1, 2.178)$$

and

$$\mathcal{R}_1^* = \mathcal{R}_2^* = (-\infty, -2.178] \cup [2.178, \infty)$$

Rejection probability of a GSD

We have

$$\mathbf{P}_\mu\Big(\text{reject } H_0\Big) = \mathbf{P}_\mu\big(Z_1^* \in \mathcal{R}_1^*\big) +$$

$$+ \sum_{k=2}^{K} \mathbf{P}_\mu\big(\underbrace{Z_1^* \in \mathcal{C}_1^*, \ldots, Z_{k-1}^* \in \mathcal{C}_{k-1}^*}_{\text{continue to stage } k}, Z_k^* \in \mathcal{R}_k^*\big) \tag{3}$$

- For $\mu = \mu_0$ expression (3) gives the type I error rate.
- For $\mu \neq \mu_0$ expression (3) gives the (possibly two-sided) power.

- I generally prefer the one-sided power and will discuss this later.

Type I error in our first examples

- **Example 1:** We already know that

$$\mathbf{P}_{\mu_0}\Big(\text{reject } H_0\Big) = \mathbf{P}_{\mu_0}\Big(\bigcup_{k=1}^{2}\{|Z_k^*| \geq 2.178\}\Big) = 0.05$$

- **Example 2:** We can calculate numerically that

$$\mathbf{P}_{\mu_0}\Big(\text{reject } H_0\Big) = \mathbf{P}_{\mu_0}(|Z_1^*| \geq 2.178) +$$

$$+\mathbf{P}_{\mu_0}(1 \leq |Z_1^*| < 2.178, |Z_2^*| \geq 2.178) = 0.0458$$

Type I error in a new example

- **Example 3:**

  Like example 2 but with 2.178 replaced by 2.14:

  $$\mathcal{A}_1^* = (-1, 1), \qquad \mathcal{C}_1^* = (-2.14, -1] \cup [1, 2.14)$$

  and $$\mathcal{R}_1^* = \mathcal{R}_2^* = (-\infty, -2.14] \cup [2.14, \infty).$$

  The type I error $\alpha = 0.05$ is now exhausted:

  $$\mathbf{P}_{\mu_0}\Big(\text{reject } H_0\Big) = \mathbf{P}_{\mu_0}\big(|Z_1^*| \geq 2.14\big) +$$

  $$+\,\mathbf{P}_{\mu_0}\big(1 \leq |Z_1^*| < 2.14, |Z_2^*| \geq 2.14\big) = 0.05$$

Calculating power by shifting the regions

We know that $\mathbf{E}(Z_k^*) = \vartheta_k$ for the non-centrality parameter

$$\vartheta_k = \sqrt{N_k}(\mu - \mu_0)/\sigma, \qquad k = 1, \ldots, K.$$

Therefore $(Z_1^* - \vartheta_1, \ldots, Z_K^* - \vartheta_k) \sim N(\mathbf{0}, \Sigma)$ and

$$\mathbf{P}_\mu\Big(\text{reject } H_0\Big) = \mathbf{P}_{\mu_0}(Z_1^* \in \mathcal{R}_1^* - \vartheta_1) +$$

$$+ \sum_{k=2}^{K} \mathbf{P}_{\mu_0}(Z_1^* \in \mathcal{C}_1^* - \vartheta_1, \ldots, Z_{k-1}^* \in \mathcal{C}_{k-1}^* - \vartheta_{k-1}, Z_k^* \in \mathcal{R}_k^* - \vartheta_k)$$

Examples: See solid lines in Figure 1.4 (Example 1) and Figure 1.3 (Example 3) of Wassmer & Brannath (2016, WaBr16) for power in dependence of relative effect $\delta = (\mu - \mu_0)/\sigma$ when $n_1 = n_2 = 20$.

Exercise: Calculate the power for Example 1 with $\mu_0 = 0$, $\mu_1 = 0.4$, $\sigma = 2$ and $n_1 = n_2 = 100$.

Sample Size Calculation for a GSD (Slide I)

- We fix all ratios $r_k = n_k/n_1$ for $k = 2, \ldots, K$.
- A common choice is $n_1 = n_2 = \cdots = n_K$ (i.e. $r_k = 1$ for all $k \geq 2$).
- Obviously, $N_k = n_1 \underbrace{(r_1 + \cdots + r_k)}_{=:R_k}$ and $\vartheta_k = \vartheta_1 \sqrt{R_k}$.
- Given r_k for $k \geq 2$, the power is

$$\text{Power}(\vartheta_1) = \mathbf{P}_{\mu_0}\left(\bigcup_{k=1}^{K} \{Z_k^* \in \mathcal{R}_k^* - \vartheta_1 \sqrt{R_k}\} \right)$$

  and depends only on $\vartheta_1 := \sqrt{n_1}(\mu - \mu_0)/\sigma$.
- Instead of fixing r_k we will later fix the *information times*:

$$t_k := N_k/N_K = R_k/R_K, \qquad k = 1, \ldots, K-1 \qquad (t_K = 1)$$

Sample Size Calculation for a GSD (Slide II)

- For the anticipated type II error $0 < \beta < 1$ there exist a unique $\vartheta_{1,\beta}$ such that
$$\text{Power}(\vartheta_{1,\beta}) = 1 - \beta.$$

- For given R_k ($k \geq 2$) and $\delta_1 = (\mu_1 - \mu_0)/\sigma$, the sample sizes required for power $1 - \beta$ are
$$n_1 = \vartheta_{1,\beta}^2/\delta_1^2 \qquad \text{and} \qquad n_k = r_k n_1, \; k = 2, \ldots, K.$$

- **Example:** GSD in Example 1 ($K = 2$, $u = 2.178$). If $\delta_1 = 0.4$:
$$n_1 = n_2 = 27 \;\rightarrow\; \vartheta_1 = 0.4\sqrt{27} = 2.0785 \;\rightarrow\; \text{Power} = 0.797$$
$$n_1 = n_2 = 28 \;\rightarrow\; \vartheta_1 = 0.4\sqrt{28} = 2.1166 \;\rightarrow\; \text{Power} = 0.81$$

- **Exercise:** Calculate the exact and universal $\vartheta_{1,0.2}$ and $\vartheta_{1,0.1}$.

Sample size *inflation factor*

- Recall the sample size formula for fixed sample size z-test:

$$n_f := (z_{\alpha/2} + z_\beta)^2/\delta_1^2$$

- For equally sized stages $n_1 = \cdots = n_K$ we get for the maximum sample size of the GSD

$$N_k = Kn_1 = K\vartheta_{1,\beta}^2/\delta_1^2$$

- This leads to the sample size inflation factor

$$N_k/n_f = K\vartheta_{1,\beta}^2/(z_{\alpha/2} + z_\beta)^2$$

  that depends only on β, K and the GSD boundaries (and thereby α), but is independent from δ_1, i.e. from μ_0, μ_1 and σ.

- N_K/n_f is tabulated for the common GSD e.g. in WaBr16.

The random sample size

- The sample size of a GSD depends on the stage the trial stops which depends on the data.
- Hence, the sample size is random (and not a fixed number).
- The random sample size is

$$N^* = n_1 + \sum_{k=2}^{K} n_k \mathbf{1}_{\{Z_1^* \in \mathcal{C}_1^*, \ldots, Z_{k-1}^* \in \mathcal{C}_{k-1}^*\}}$$

  where $\mathbf{1}_{\{Z_1^* \in \mathcal{C}_1^*, \ldots, Z_{k-1}^* \in \mathcal{C}_{k-1}^*\}} = 1$ if we continue to stage k and 0 otherwise.
- The sample size is always between $N_1 = n_1$ and N_K.
- We are usually interested in distribution or mean of N^*.

Average Sample Size (ASN)

- The average (expected) sample size is

$$\mathrm{ASN}(\mu) := \mathbf{E}_\mu(N^*) = n_1 + \sum_{k=2}^{K} n_k \, \mathbf{P}_\mu\big(Z_1^* \in \mathcal{C}_1^*, \ldots, Z_{k-1}^* \in \mathcal{C}_{k-1}^*\big)$$

- For its calculation we need to calculate the probabilities

$$\mathbf{P}_\mu\big(Z_1^* \in \mathcal{C}_1^*, \ldots, Z_{k-1}^* \in \mathcal{C}_{k-1}^*\big)$$

  for all $k = 1, \ldots, K-1$.

- The ASN depends on μ.
- The ASN is always between N_1 and N_K.
- **Examples:** See slashed lines in Figure 1.4 (Example 1) and Figure 1.3 (Example 3) of WaBr16.