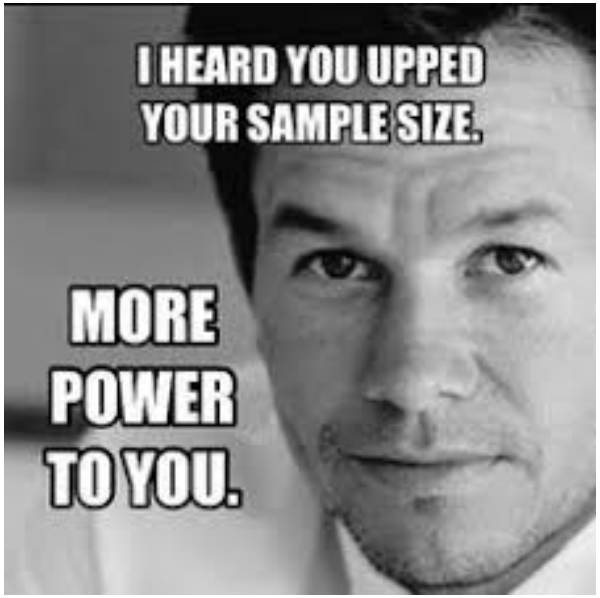
BST 210
Applied Regression Analysis



Lecture 25
Plan for Today

Happy snow days! ☺

- Welcome to sample size considerations and calculation!
- Review of
  - Hypothesis testing
  - Type I and Type II error
- Sample Size for one and two sample problems
- Examples galore: binomial and survival data
- General patterns and relationship between sample size, power, and related constituents

Sample Size Justification

- Thus far in BST 210 we have focused on the *data analysis* of observational, clinical trial, and case control studies—all which have been already planned and conducted.

- But, the *design of studies* is also a very important issue!
  - How many patients we recruit
  - How long we follow them
  - What covariate information we collect
  - What analyses we conduct

  -- all have profound impact on what *effects* we can detect, and what *conclusions* we may draw.

- Developing the study hypotheses, choosing a sampling plan, and defining the population of interest, primary outcome(s), important exposure or treatment variable(s), and possible adjustment factors -- are all absolutely essential to good study design.

Sample Size Justification

- Another important part of study design is that of *determining the sample size necessary to obtain sufficiently high power* to detect an effect of clinical or epidemiological interest

- It is generally not appropriate (or ethical) to run a study that likely will not be able to detect an effect of interest (except perhaps a pilot study)

- Determination of the appropriate sample size helps provide the scope of study feasibility, time frame, necessary resources, and budget

Sample Size Justification

- In other words:

  Q: Why do we conduct sample size calculations?

  A: In order to determine how large a study is needed to guarantee a certain level of power for detecting a pre-specified difference in effect between groups (or other measures of effect)!

Sample Size Justification

- In a perfect world:

  Q: What scenario would we want regarding sample size and power?

  A: We would want the smallest possible sample for detecting the largest effect size!

Sample Size Justification

- In considering approach:

  Q: What can go wrong if we aim *solely* on making the sample as small and cost-effective as possible?

  A: If study has too small a sample size → could have insufficient power to detect true effect → may fail to reject null hypothesis at end of study, even if true and meaningful effect does exist. That would be a big waste of time and resources for 'getting close but not close enough'!

Sample Size Justification

- In considering approach:

  Q: Given the above, why then not just use as big a sample as we can possibly recruit/collect?

  A: Ethical, practical, and cost considerations make this approach untenable. We may only need 250 subjects to power a meaningful effect size, but we collect 500. That's a waste of 250 subjects time, and twice the money—not to mention ethical issues.

Sample Size Justification

- In other words:

  - We want to be not only thoughtful, but calculated in how many subjects we plan to enroll, in order to utilize a given sample for our desired power and our effect of interest

  - Power calculations thus impose a necessary and delicate balance between

    (1) efficiency/'trim-ness' of design

    versus

    (2) having enough power to detect the effect we hypothesize exists in the population

Sample Size Justification

- In conducting power or sample size calculations we typically are interested in:

  - Type I error rate

  - Desired power (for sample size calculations, or Desired sample size (for power calculations)

  - Effect size under the alternative hypothesis that we'd like to detect

Sample Size Justification

- The precise method of calculating sample size and power will change, depending on

  (1) type of data we are working with

  (2) analysis we plan to conduct at end of study

Sample Size Justification

- Generally speaking:

(1) sample size N = function(1-β, α, prevalence of outcome (across groups), clinically meaningful effect size, 1 or 2-sided test, drop-out/missingness rate**)**

and

(2) Power 1-β = function(N, α, prevalence of outcome (across groups), clinically meaningful effect size, 1 or 2-sided test, drop-out/missingness rate**)**

Sample Size Justification

- No matter the type of data or research question(s) at hand, the motivation behind power and sample size calculations is always the same:

We just want to make sure that—if there really is a clinically meaningful association between our outcome and our predictor of interest—that we are able to design a study that is actually powered to be able to detect this association!

Computing Caveats around Sample Size

- SAS is excellent and well documented (regulatory/audit-driven)
- Stata is also great and quite well documented
- R…less so

Recall: Hypothesis Testing

- Hypotheses being tested:

  H_0 : Null hypothesis

  H_1 : Alternative hypothesis

- Decisions to be made:

  Reject H_0 in favor of H_1

  Fail to reject H_0

- Two types of errors:

  α = P(reject H_0 | H_0 true) (ie false positive)

  = P(Type I error)

  = significance level of test

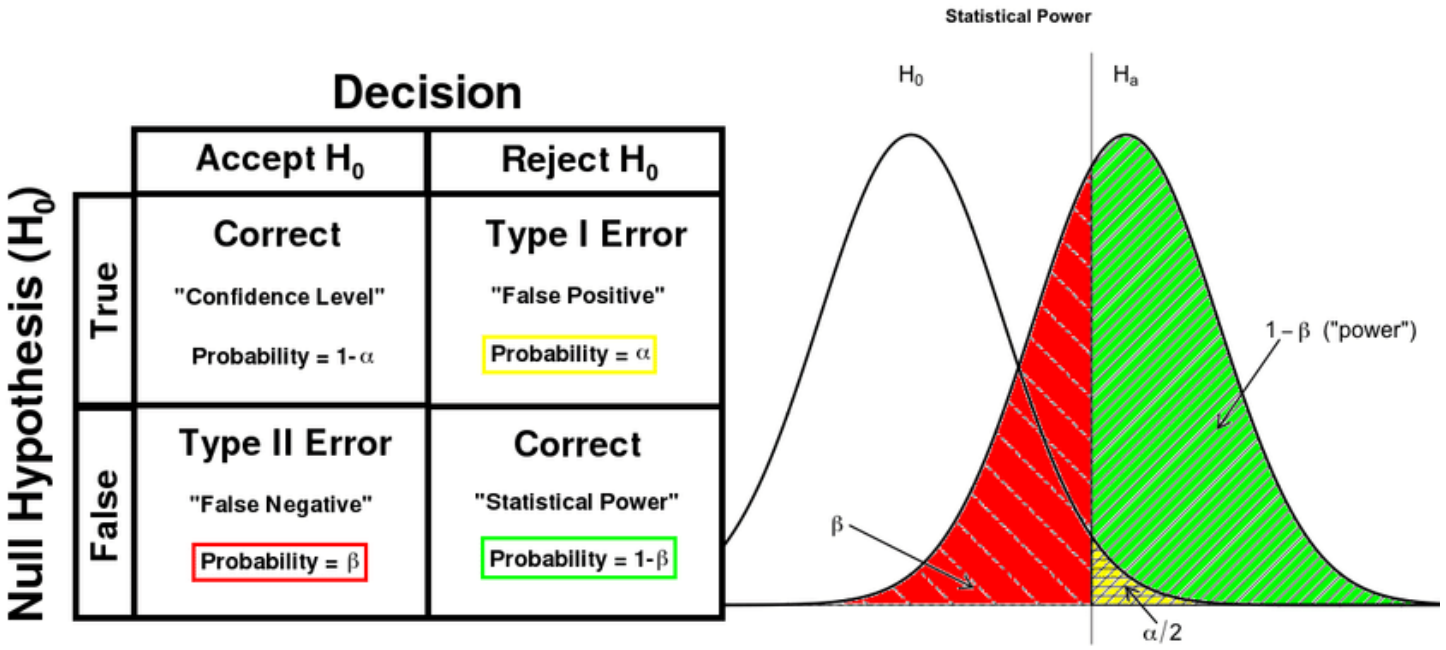
  β = P(fail to reject H_0 | H_0 false) (ie false negative)

  = P(Type II error)

Recall: Hypothesis Testing

- *Main event*:

Power = P(reject H_0 | H_0 false) = 1 – β



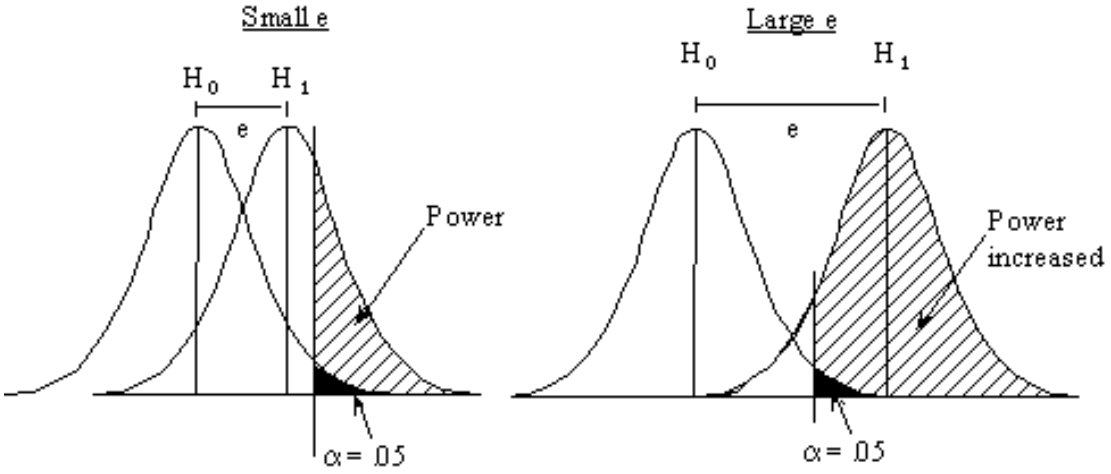
General Patterns in the Sample Size – Power Relationship

In general, what do you think will happen to the necessary sample size if we:

(1) Increase our alternative effect size of interest?

Decrease our alternative effect size of interest?

*As the alternative effect size that we are interested in detecting **increases**, the sample size needed to detect that effect with a fixed power **decreases**. Conversely, as our alternative effect **decreases**, the necessary sample size **increases**.*



General Patterns in the Sample Size – Power Relationship

In general, what do you think will happen to the necessary sample size if we:

(2) Increase our desired power? Decrease our desired power?

*As the power with which we wish to detect a particular effect **increases**, the sample size needed to achieve that power also **increases**. As our desired power **decreases**, our sample size also **decreases**.*

(3) Incorporate potential loss to follow up or censoring?

Here we must assume that dropout or censoring is not related to the outcome of interest (sometimes this is not realistic). In doing so, we can then simply multiply our calculated sample size N by (1-rate of missingness).

General Patterns in the Sample Size – Power Relationship

In general, what do you think will happen to the power of our study if we:

(1) Increase our alternative effect size of interest?

Decrease our alternative effect size of interest?

For a fixed sample size, **increasing** the effect we are interested in detecting will **increase** our power; **decreasing** it will **decrease** our power.

(2) Increase our sample size? Decrease our sample size?

*For a fixed effect size of interest, **increasing** the sample size will **increase** our power to detect that effect; **decreasing** the sample size will **decrease** our power.*

General Patterns in the Sample Size – Power Relationship

In general, what do you think will happen to the power of our study if we:

(3) Use a 1-sided test? A 2-sided test? (or alter the α-level)

Since the one-sided test only considers alternatives that are in one particular direction, we will have more power to detect effects in that direction, versus had we 'split' the alpha level between 2 tails of a distribution for a 2-sided test.

Example: Power for One-Sample Binomial Test

- Let's get to some examples! Suppose we wish to test the null hypothesis

  H_0: $p = p_0$ versus the alternative H_1: $p \neq p_0$

- H_0 would be rejected if the sample proportion $\hat{p}$ is either too large or too small, relative to p_0

- The decision cut-points chosen are determined by α

- For the moment, assume we are interested in the one-sided alternative H_1: $p > p_0$

- If H_1 is true, then $p = p_1$ where $p_1 > p_0$

Example: Power for One-Sample Binomial Test

- H_0 would be rejected only if the sample proportion $\hat{p}$ is too large (too much higher than p_0)

- Remember that $\alpha = P(\text{reject } H_0 \mid H_0 \text{ true})$

$$= P(\text{reject } H_0 \mid p = p_0)$$
$$= P(\hat{p} > c \mid p = p_0)$$
$$= P\left((\hat{p} - p_0) / \sqrt{p_0 q_0 / n} > (c - p_0) / \sqrt{p_0 q_0 / n} \right)$$
$$= P\left(Z > (c - p_0) / \sqrt{p_0 q_0 / n} \right)$$
$$= P(Z > z_{1-\alpha})$$

- Therefore $c = p_0 + z_{1-\alpha} \sqrt{p_0 q_0 / n}$. Here $Z \sim N(0, 1)$.

Example: Power for One-Sample Binomial Test

- Also, β = P(fail to reject H_0 | H_0 false)

  = P(fail to reject H_0 | $p = p_1$)

  = P($\hat{p} \leq c \mid p = p_1$)

  = P($(\hat{p} - p_1) / \sqrt{p_1 q_1 / n} \leq (c - p_1) / \sqrt{p_1 q_1 / n}$)

  = P($Z \leq (c - p_1) / \sqrt{p_1 q_1 / n}$)

  = P($Z \leq z_\beta$)

- Therefore $c = p_1 + z_\beta \sqrt{p_1 q_1 / n}$ and thus,

$$p_0 + z_{1-\alpha} \sqrt{p_0 q_0 / n} = p_1 + z_\beta \sqrt{p_1 q_1 / n}$$

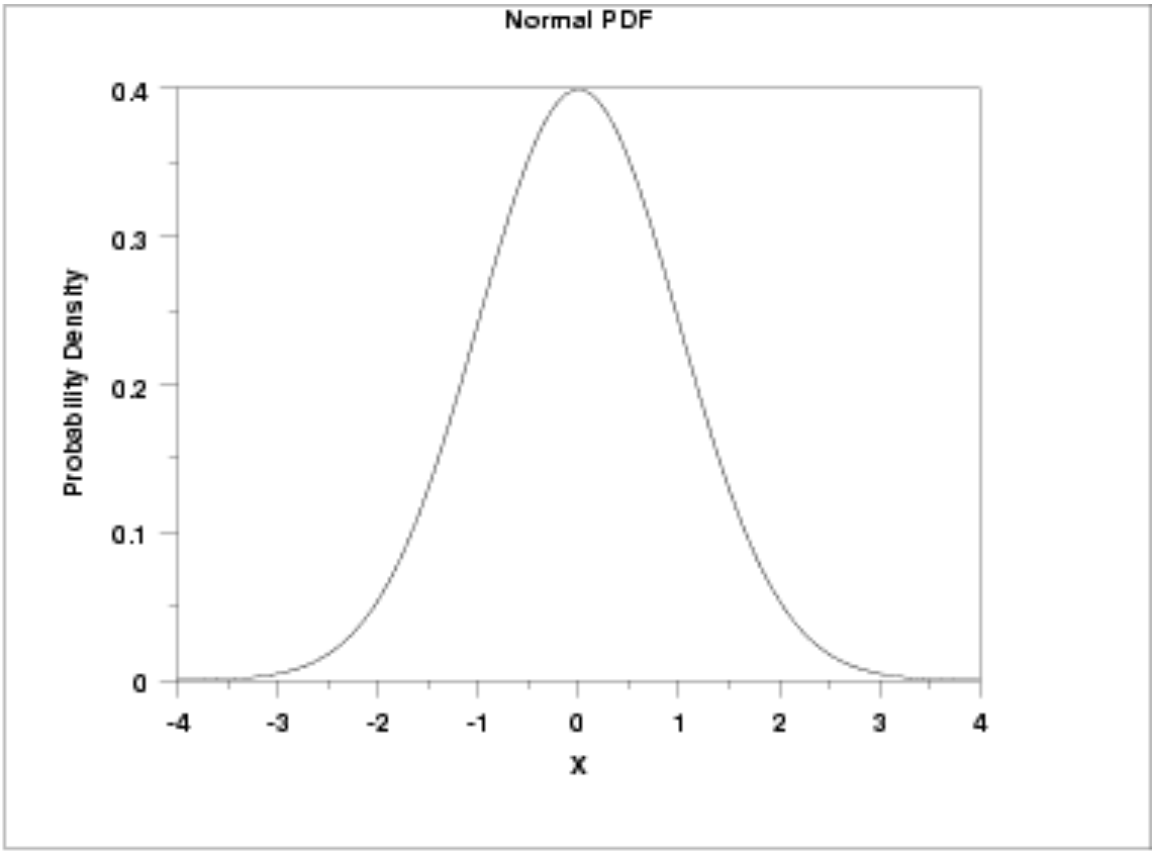
Example: Power for One-Sample Binomial Test

- This equation can be solved for z_β, and then for the power $1 - \beta$

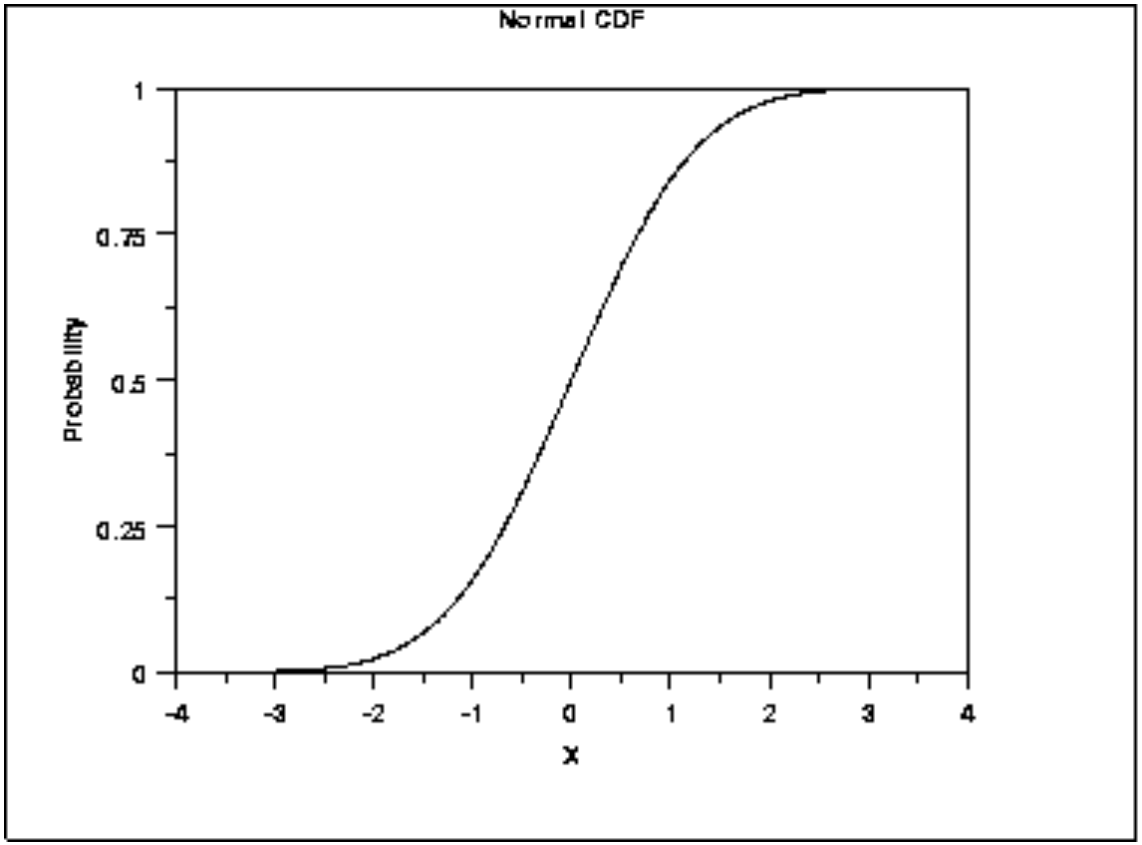
- **Power 1 – β** = $\phi[\sqrt{p_0 q_0 / p_1 q_1}(z_\alpha + [|p_0 - p_1|\sqrt{n} / \sqrt{p_0 q_0}])]$

where ϕ is the standard normal cumulative distribution function. This method assumes that the normal approximation to the binomial distribution is valid, say for $np_0q_0 \geq 5$.

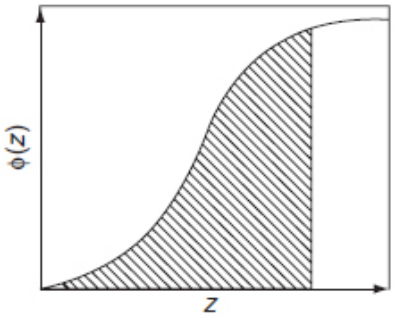
Recall: N(0, 1) Density Function



Recall: N(0, 1) Cumulative Distribution Function



Recall: N(0, 1) Cumulative Distribution Function



$$\Phi(z) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{z} \exp\left(\frac{-x^2}{2}\right) dx$$

$$\text{for } 0.00 \le z \le 4.00$$

$$1 - \Phi(z) = \Phi(-z)$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6985	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319

Recall: N(0, 1) Cumulative Distribution Function

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9430	0.9440
1.6	0.9452	0.9463	0.9474	0.9485	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9700	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9762	0.9767
2.0	0.9773	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9865	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9^2010	0.9^2061	0.9^2035	0.9^2086	0.9^2111	0.9^2134	0.9^2158
2.4	0.9^2180	0.9^2202	0.9^2224	0.9^2245	0.9^2266	0.9^2286	0.9^2305	0.9^2324	0.9^2343	0.9^2361
2.5	0.9^2379	0.9^2396	0.9^2413	0.9^2430	0.9^2446	0.9^2461	0.9^2477	0.9^2492	0.9^2506	0.9^2520
2.6	0.9^2534	0.9^2547	0.9^2560	0.9^2573	0.9^2586	0.9^2598	0.9^2609	0.9^2621	0.9^2632	0.9^2643
2.7	0.9^2653	0.9^2664	0.9^2674	0.9^2683	0.9^2693	0.9^2702	0.9^2711	0.9^2720	0.9^2728	0.9^2737
2.8	0.9^2745	0.9^2752	0.9^2760	0.9^2767	0.9^2774	0.9^2781	0.9^2788	0.9^2795	0.9^2801	0.9^2807
2.9	0.9^2813	0.9^2819	0.9^2825	0.9^2831	0.9^2836	0.9^2841	0.9^2846	0.9^2851	0.9^2856	0.9^2861
3.0	0.9^2865	0.9^2869	0.9^2874	0.9^2878	0.9^2882	0.9^2886	0.9^2889	0.9^2893	0.9^2897	0.9^2900
3.1	0.9^3032	0.9^3065	0.9^3096	0.9^3126	0.9^3155	0.9^3184	0.9^3211	0.9^3238	0.9^3264	0.9^3289
3.2	0.9^3313	0.9^3336	0.9^3359	0.9^3381	0.9^3402	0.9^3423	0.9^3443	0.9^3462	0.9^3481	0.9^3499
3.3	0.9^3517	0.9^3534	0.9^3550	0.9^3566	0.9^3581	0.9^3596	0.9^3610	0.9^3624	0.9^3638	0.9^3651
3.4	0.9^3663	0.9^3675	0.9^3687	0.9^3698	0.9^3709	0.9^3720	0.9^3730	0.9^3740	0.9^3749	0.9^3759
3.5	0.9^3767	0.9^3776	0.9^3784	0.9^3792	0.9^3800	0.9^3807	0.9^3815	0.9^3822	0.9^3822	0.9^3835
3.6	0.9^3841	0.9^3847	0.9^3853	0.9^3858	0.9^3864	0.9^3869	0.9^3874	0.9^3879	0.9^3883	0.9^3888
3.7	0.9^3892	0.9^3896	0.9^4004	0.9^4043	0.9^4116	0.9^4116	0.9^4150	0.9^4184	0.9^4216	0.9^4257
3.8	0.9^4277	0.9^4305	0.9^4333	0.9^4359	0.9^4385	0.9^4409	0.9^4433	0.9^4456	0.9^4478	0.9^4499
3.9	0.9^4519	0.9^4539	0.9^4557	0.9^4575	0.9^4593	0.9^4609	0.9^4625	0.9^4641	0.9^4655	0.9^4670

Example: Power for One-Sample Binomial Test

- For a two-sided test, we replace α by α/2
- **Power 1 – β** = $\phi[\sqrt{p_0 q_0 / p_1 q_1}(z_{\alpha/2} + [|p_0 - p_1|\sqrt{n} / \sqrt{p_0 q_0}])]$
- The power of the hypothesis test depends on α, n, p_0, and p_1
- If the power of a test is low, then there is little chance of detecting a difference in proportions even if one truly exists

Example: Power for One-Sample Binomial Test

The power of a test can be increased by:

(1) Raising the significance level α (though usually α is set at 0.05, two-sided)

(2) Considering larger deviations from p_0, i.e., increasing $|p_0 - p_1|$ (we must make an assumption about the value of p_1)

(3) Increasing the sample size n

Now: Sample Size for One-Sample Binomial Test

- For a two-sided test, solve the previous power equation for the sample size *n* to get

$$\textbf{sample size } \boldsymbol{n} = \frac{p_0 q_0 \left(z_{\alpha/2} + z_{1-\beta} \sqrt{\frac{p_1 q_1}{p_0 q_0}} \right)^2}{(p_1 - p_0)^2}$$

- The sample size needed depends on α, β, p_0, and p_1

Example: Otitis Media- 1 sample

- We are interested in studying children with otitis media (OTM), and want to know whether the proportion of children with language deficits by the age of 3 years is the same in this group as it is in the general population

- If 15% of normal children have language deficits by age 3, how many OTM children would need to be studied to have 90% power to test whether the proportion with language deficits is *different* for OTM cases and normal children?

Example: Otitis Media- 1 sample

- This will be a one-sample test conducted at the $\alpha = 0.05$ level of significance, with power = 0.90
- H_0: $p = 0.15$ versus H_1: $p \neq 0.15$
- Based on a small pilot study with 20 children with otitis media (OTM), 25% have language deficits at age 3
- We can use this as the "alternative" proportion p_1

Example: Otitis Media – 1 sample

- $p_0 = 0.15$, $p_1 = 0.25$, $z_{1-\alpha/2} = 1.96$, and $z_{1-\beta} = 1.28$

- $$n = .15(.85)\frac{\left[1.96 + 1.28\sqrt{\frac{.25(.75)}{.15(.85)}}\right]^2}{(.25-.15)^2}$$

  $=$ 157.3 or 158 subjects need to be studied to have 90% power.

Example: Otitis Media – 1 sample

- Suppose that only 80 subjects are available. What is the power?

- $1 - \beta = \phi[\sqrt{p_0 q_0 / p_1 q_1}(z_{\alpha/2} + [|p_0 - p_1|\sqrt{n} / \sqrt{p_0 q_0}])]$

$$= \phi[\sqrt{(.15)(.85)/(.25)(.75)}\,(1.96 + [|.15 - .25|\,\sqrt{80} / \sqrt{(.15)(.85)}])]$$

$$= \phi[0.45]$$

- Power = $\phi(0.45) = 0.6735$ (found in CDF table)

- In general, this would not be considered sufficient

Example: Breast Cancer – 2 sample

- A 5-year study is planned to look at the effect of a certain gene on diagnosis of breast cancer
- Among women without the gene, 1% of women will be diagnosed with breast cancer over a 5-year period
- It is anticipated that among women with the gene, that probability will double
- How many women should be enrolled in the study if it is expected that 25% of women are gene-positive and 75% are gene-negative, and we require 80% power? This is a two-sample hypothesis test.

Example: Breast Cancer – 2 sample

- For the two-sample test for binomial proportions, we test the hypothesis

$$H_0: p_1 = p_2 \text{ versus } H_1: p_1 \neq p_2$$

- The test statistic is:

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}(1/n_1 + 1/n_2)}}$$

- We reject H_0 if $z > z_{1-\alpha/2}$ or $z < z_{\alpha/2}$

Example: Breast Cancer – 2 sample

The power is approximately:

$$Power = \Phi\left[\frac{\Delta}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} - z_{1-\alpha/2}\sqrt{\frac{\bar{p}\bar{q}(1/n_1 + 1/n_2)}{p_1 q_1 / n_1 + p_2 q_2 / n_2}}\right],$$

$$\text{where } \bar{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}, \bar{q} = 1 - \bar{p}.$$

and $\Delta = |p_2 - p_1|$

Example: Breast Cancer – 2 sample

- We must specify n_1 and n_2; they may or may not be equal
- α is the significance level of the test
- p_1 and p_2 are the proposed values of the population proportions if the alternative hypothesis is true
- $\Delta = |p_2 - p_1|$ should be clinically important

Example: Breast Cancer – 2 sample

- To determine sample size, we assume that $n_2 = k\, n_1$, substitute this into the equation for power, and solve for n_1

$$n_1 = \frac{\left[z_{1-\alpha/2}\sqrt{\bar{p}\bar{q}(1+1/k)} + z_{1-\beta}\sqrt{p_1q_1 + p_2q_2/k} \right]^2}{\Delta^2},$$

$$n_2 = kn_1.$$

Example: Breast Cancer – 2 sample

- Suppose group 1 = gene-negative women and group 2 = gene-positive women
- $p_1 = 0.01$, $p_2 = 0.02$, $k = 0.25/0.75 = 1/3$
- $z_{1-\alpha/2} = z_{0.975} = 1.96$, $z_{1-\beta} = z_{0.80} = 0.84$
- $\bar{p} = (p_1 + k\, p_2) / (1 + k)$

then

$$\bar{p} = \frac{.01 + (1/3)(.02)}{4/3} = .0125, \quad \bar{q} = .9875.$$

Example: Breast Cancer – 2 sample

Therefore,

- $n_1 = [1.96 \sqrt{(0.0125)(0.9875)(1+3)} + 0.84 \sqrt{(0.01)(0.99) + (0.02)(0.98)(3)}]^2 / (0.01)^2$
  $= 4299.3$

- We would require 4300 gene-negative women, and $n_2 = k\, n_1 = (1/3)\ (4300) = 1434$ gene-positive women

- The total sample size is 5734

- Note that this is roughly in keeping with what we expected from general population, ie 25:75

Example: Breast Cancer – 2 sample

- Suppose next that a total of 4000 women are recruited for the breast cancer study, of whom 3000 are gene-negative and 1000 gene-positive

- What would be the power of the study? Let's write out what we know:

- $p_1 = 0.01$, $p_2 = 0.02$

- $z_{1-\alpha/2} = z_{0.975} = 1.96$, $n_1 = 3000$, $n_2 = 1000$

Example: Breast Cancer – 2 sample

- $\bar{p} = \dfrac{.01(3000) + .02(1000)}{4000} = .0125,$

  $\bar{q} = 1 - .0125 = .9875.$

  Hence, the power is given by $\Phi(a - b)$, where

  $$a = \frac{.01}{\sqrt{.01(.99)/3000 + .02(.98)/1000}} = 2.090,$$

  $$b = 1.96\frac{\sqrt{.0125(.9875)(1/3000 + 1/1000)}}{\sqrt{.01(.99)/3000 + .02(.98)/1000}} = 1.662$$

- Power = $\phi(0.428) = 0.67 = 67\%$

But what if there are Missing Outcomes?

- All the calculations above assume that all observations are *observed*.

- In cases where there are missing outcomes, say with λ as the missingness proportion, and the missingness is unrelated to the outcome (*strong assumption; not always realistic!*), one can *increase the sample size(s) by the* ***fraction 1 / (1 – λ)*** to achieve the same power.

- E.g., if λ = 20% = 0.20, multiply the sample size(s) by the fraction 1 / 0.80 = 1.25 to achieve sufficient power.

Example: Weight Loss study

- Suppose we enroll subjects with type II diabetes and BMI ≥ 30 kg/m^2 into a weight loss study

- The outcome variable for the study is a decrease in BMI of ≥ 2 kg/m^2 after 1 year

- Subjects are randomized to

  (1) a control intervention (just receiving health-related pamphlets) – group 1,

  or

  (2) an active intervention (diet/exercise counseling) – group 2

Example: Weight Loss study

- It is projected that 10% of the subjects in the control group will have a positive outcome
- It is also thought that 30% of the subjects in the active group *who perform the intervention* will have a positive outcome
- All subjects who do not return for a 1 year follow-up visit are assumed to be failures

Example: Weight Loss study

- It is anticipated that 40% of subjects in the active group will not complete the intervention (i.e., 40% drop-outs or non-compliers), and will have success rates similar to the control group

- How many subjects should be enrolled in the study to have 80% power if there are an equal number of active and control subjects and a two-sided test at the $\alpha = .05$ level is planned?

Sample Size in a Clinical Trial Setting

- The observed success probability in the active group (p_2^*) is given by

$$p_2^* = (1 - \lambda)\, p_2 + \lambda\, p_1$$

where λ is the *drop-out proportion*

- Here p_2^* is a weighted average of p_1 and p_2 – we are assuming that a proportion λ drop out of the active group and look like the control group, while a proportion 1 – λ look like the intervention group.

Sample Size in a Clinical Trial Setting

- The observed success probability in the control group (p_1^*) is given by

$$p_1^* = (1 - \theta)\, p_1 + \theta\, p_2$$

where θ is the *drop-in rate* (= 0 for this study)

- The drop-in rate is the proportion of control subjects who perform the active intervention outside the study protocol. In our example, it is not possible for control subjects to get the intervention, but sometimes scenario is possible.

Example: Weight Loss

- Assuming 100% compliance to the assigned treatment regimen (hence no drop-ins or drop-outs, or cross-overs), to get 80% power for a two-sided 0.05 level test, we would compare proportions 0.10 and 0.30.

- In this simple scenario, calculations show that 124 subjects are required (62 per group).

Example: Weight Loss

- If we allow for non-compliance, we have $p_1 = p_1^* = 0.10$, $p_2 = 0.30$, $\lambda = 0.40$, and $\theta = 0$
- $p_2^* = 0.60(0.30) + 0.40(0.10) = 0.22$
- Comparing 0.10 to 0.22 with 80% power and a two-sided 0.05 level test requires 292 subjects (146 per group)
- Wow! Clearly a substantially larger sample size! Missingness is something to take note of (more upcoming in this course)

Next - Sample Size Estimation: Survival Outcomes

- We wish to compare two survival curves -- maybe an experimental group (E, e.g., the chemotherapy maintenance group in the leukemia study) and a control group (C)

- We will use a **two-sided log-rank test** with probability of making a type I error equal to α, and power $1 - \beta$

- We will *assume proportional hazards* between groups

Sample Size Estimation: Survival Outcomes

- *k* is the ratio of the sample size in E versus C, often set to 1 (if expecting equal sample sizes per group)

- When comparing hazard rates (or survival functions) between groups, the key determinant of the

  sample size n

  is the

  total number of observed (non-censored) events

  represented by *d*.

Sample Size Estimation: Survival Outcomes

- To detect a hazard ratio of HR, the total number of events required (using Freedman's method) is:

$$d = \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 (k\,\mathrm{HR} + 1)^2}{k\,(\mathrm{HR} - 1)^2}$$

- This is for a two-sided α level test with power $1 - \beta$. (When $k = 1$, you get the same d whether you use HR or 1/HR.)

Example: Sample Size Estimation - Survival

- We want to design a clinical trial to detect a 25% reduction in the hazard of a major cardiovascular event with 90% power; therefore, the hazard ratio HR would be 0.75

- We will use a log-rank test conducted at the two-sided 0.05 level of significance, and wish to have equal sample sizes in each group ($k = 1$)

Example: Sample Size Estimation - Survival

- The total number of CV events required is

$$d = \frac{(z_{0.975} + z_{0.90})^2 (HR + 1)^2}{(HR - 1)^2}$$

$$= \frac{(1.96 + 1.28)^2 (0.75 + 1)^2}{(0.75 - 1)^2}$$

$$= 514.4 \text{ (or round up to 516)}$$

- So, we need 516 events (an even number), assuming all subjects in the trial 'fail', or have an event (ie *no observations censored!*)

Example: Sample Size Estimation - Survival

- If censoring will occur, the number of *events* required remains the same, but the number of *subjects* to be enrolled in the study increases

- Censoring is often assumed to occur equally in the two groups, i.e., that we have a common proportion of subjects in both groups that withdraw or are censored

Other Assumptions re Missingness

- Depending on your problem, you may have to adjust the sample size/power calculations to fit the scenario the best
- For example, in a clinical trial with dropouts, you might assume that all dropouts have the worst outcome (be it a binary or survival outcome)
- In an observational study with missing covariates or outcomes, you might study different assumptions about the missingness with a sensitivity analysis

Still Other Issues and Caveats

- Formulas and results can vary depending on assumptions
  - Two-sample binary comparisons based on normal approximations as we have done, vs. using a continuity correction or performing exact calculations
  - Two-sample survival comparisons based on assumptions about accrual of subjects into the study, e.g., constant accrual over time with earlier subjects having a higher chance of having the event
- Different software may give slightly different answers

Bottom Line when working with sample size/power

- Often, the sample sizes you estimate from simplistic assumptions are just a starting point – you probably need to adjust for dropouts or drop-ins, noncompliance, censored observations, missing values, or related factors – and you may also have covariates or confounding factors to control for

- Not only that, investigators often anticipate effect sizes larger than what may be realistic

- You'll need to work closely with collaborators with expertise regarding the clinical reasoning

Coming Up

- Computing sample sizes; some stat packages
- Basics of sample size/power estimation for regression problems
- Introduction to the analysis of missing data!
- Preparation for Final Exam