MEDSTA 2: Regression models in medical research

18 January 2014

Øystein Ariansen Haaland, PHD

Department of Global Public Health and Primary Care, University of Bergen

Statistical power

- What is (statistical) power?

		Result of test	
		Keep H_0	Reject H_0
The truth	H_0 true	True negative	False positive
	H_0 false	False negative	True positive

Statistical power

- What is (statistical) power?

		Result of test	
		Keep H_0	Reject H_0
The truth	H_0 true	True negative	Type I error
	H_0 false	Type II error	True positive

Statistical power

- What is (statistical) power?

		Result of test: Keep H_0	Result of test: Reject H_0
The truth	H_0 true	True negative 1-α	Type I error α
	H_0 false	Type II error β	True positive $1-\beta$

Statistical power

- What is (statistical) power?
  - The probablity of rejecting a false H_0

Result of test

		Keep H_0	Reject H_0
The truth	H_0 true	True negative 1-α	Type I error α
	H_0 false	Type II error β	True positive $1 - \beta$

Statistical power

- What is (statistical) power?
  - The probablity of rejecting a false H_0

		Result of test	
		Keep H_0	Reject H_0
The truth	H_0 true	True negative 1-α	Type I error α
	H_0 false	Type II error β	True positive $1-\beta$

Statistical power

- What is (statistical) power?
  - The probablity of rejecting a false H_0
  - P(Reject $H_0 | H_0$ false)
  - P(p<0.05|H_0 false)
- If the power is low, we do not reject H_0 even if we should have.

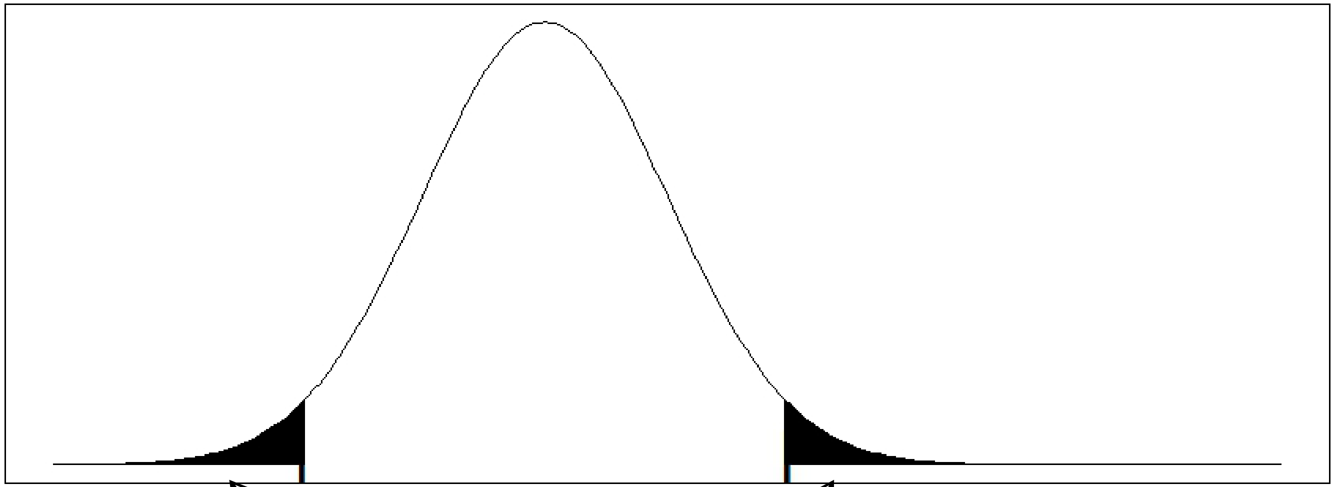
Statistical power

- Consequences of low power
  - More difficult to get financing.
  - More difficult to get aproval from ethical commitees.
  - We risk wasting our time. Have to rely on luck.
  - Power affects how we have to interpret our results (will come back to this).

Statistical power

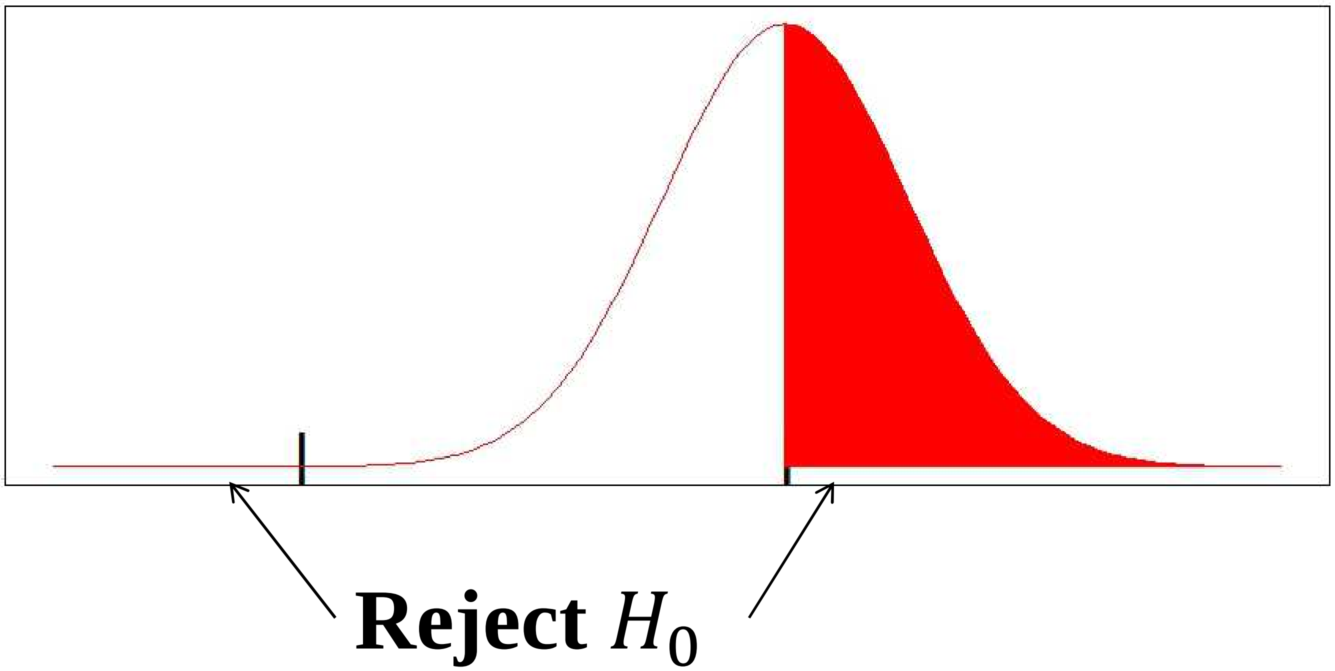
- Power depends on
  - Sample size, n
  - Effect size, Δ
  - Significance level, α

Statistical power

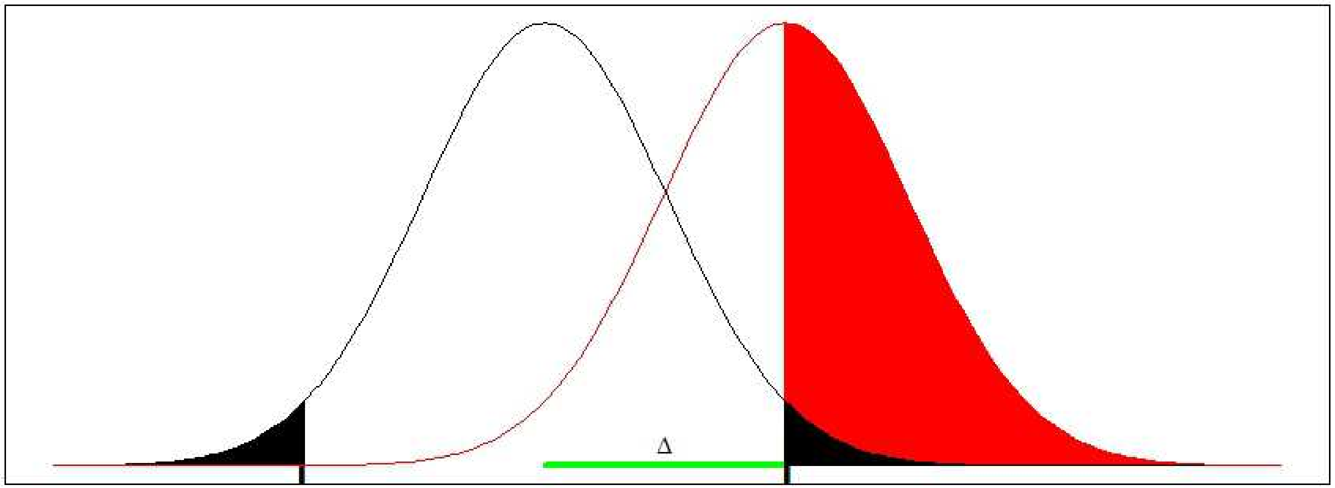


Reject H_0

Statistical power



Statistical power



Reject H_0

Statistical power

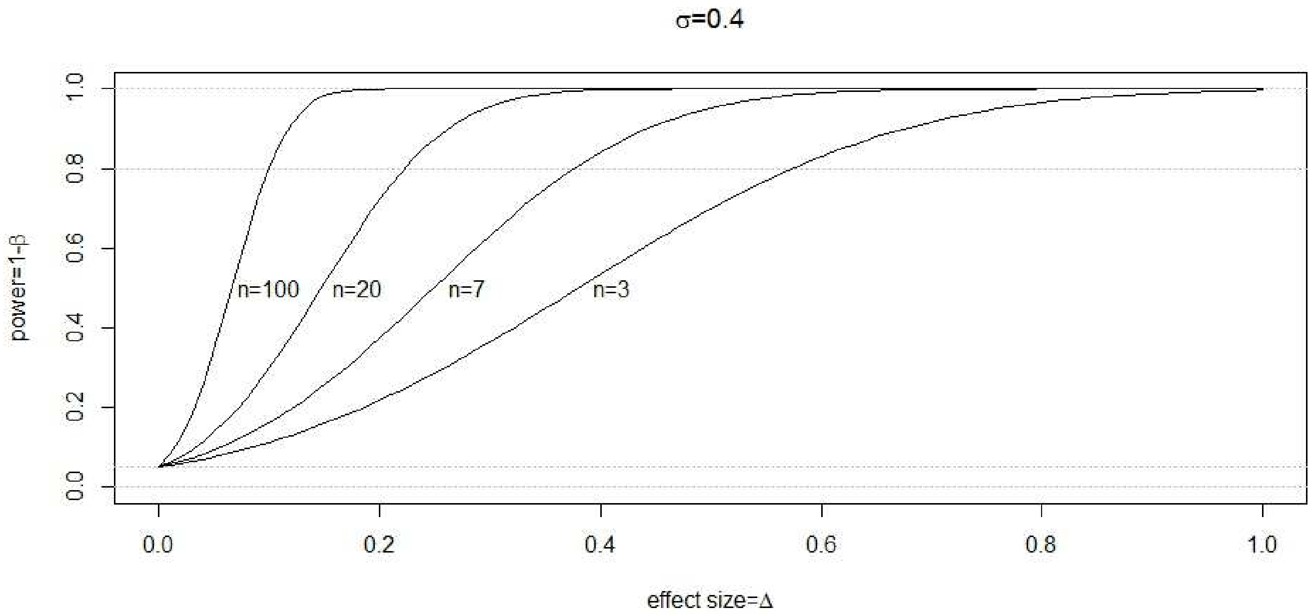
- Power calculations z-test (σ known)
  - $H_0: \mu = \mu_0$, $H_A: \mu < \mu_0$
  - Power (unknown effect size Δ, n known)
    - $1 - \beta$
    - $P(\text{Reject } H_0 | \mu = \mu_0 - \Delta)$
    - $P\left(Z < z_\alpha + \frac{\Delta\sqrt{n}}{\sigma}\right)$

Statistical power

- Power calculations z-test (σ known)
  - Consider a drug to reduce fever
  - Give drug to a group of patients and measure how the fever develops
  - $H_0: \mu = 0$ (no decrease)
  - $H_A: \mu < 0$ (fever drops)
  - Assume $\sigma = 0.4$

Statistical power

- Power calculations z-test (σ known)
  - $H_0: \mu = 0,$ $H_A: \mu < 0$
  - Power (unknown effect size Δ, n known)



Statistical power

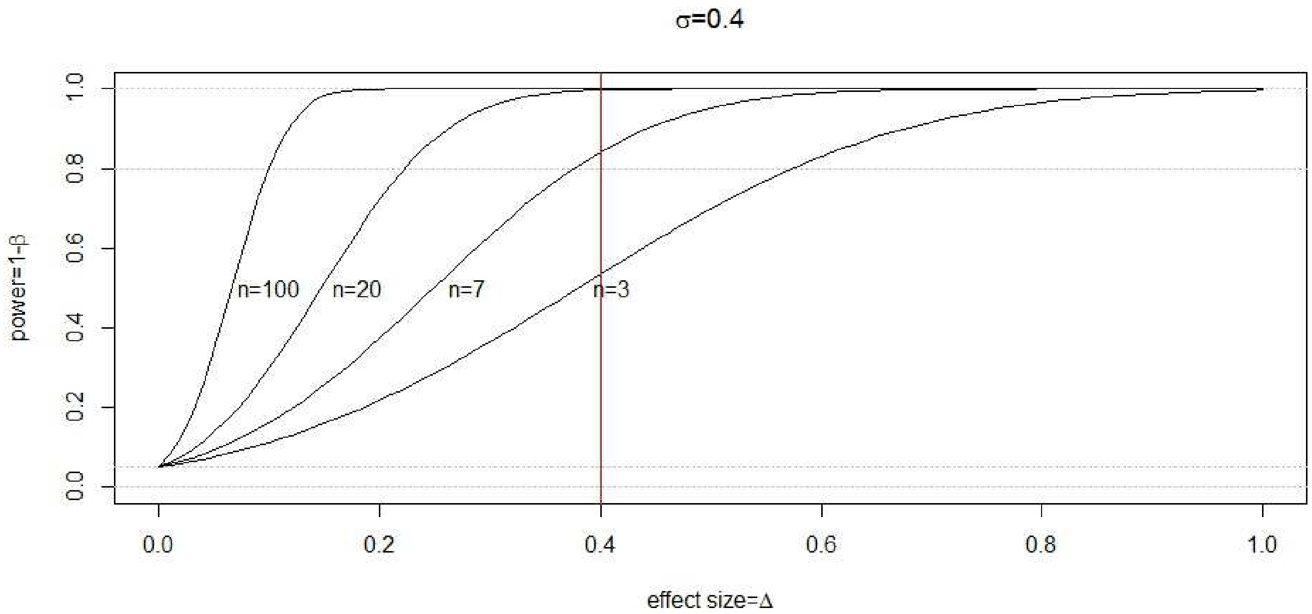
- Power calculations: z-test (σ known)
  - One sample
  - Consoder power =$1 - \beta$
  - Consider effect size = Δ
  - One-sided test ($H_0: \mu \leq \mu_0$ vs. $H_A: \mu > \mu_0$)
    - $n = \left(z_\alpha + z_\beta\right)^2 \times \left(\frac{\sigma}{\Delta}\right)^2$

Statistical power

- Power calculations: z-test ($\sigma = 0.4$)
  - One sample
  - Consoder power =$1 - \beta = 0.8$
  - Consider effect size = $\Delta = 0.4$
  - One-sided test ($H_0 : \mu \leq 0$ vs. $H_A : \mu > 0$)
    - $z_{0.05} = -1.65$, $z_{0.20} = -0.84$
    - $n = (-1.65 - 0.84)^2 \times \left(\frac{0.4}{0.4}\right)^2 \approx 6.1 \rightarrow 7$

Statistical power

- Power calculations z-test ($\sigma = 0.4$)
  - $\Delta = 0.4$



Statistical power

- Mostly we do not know σ or Δ
  - Use published data
    - Caution! Published effect estimates are often too high because of publication bias.
  - Guess: («$\sigma = 2$ is too high and $\sigma = 0.1$ is too low.»)
  - Consider what is clinically relevant («We don't care about $\Delta < 0.4$.»

Statistical power

- Will not go through the mathematics for
  - T-tests
  - More than one sample
  - Paired observations
  - χ^2-tests
  - Regressions
  - Etc.

Statistical power

- T-tests are similar to z-tests
  - Do not assume a known σ
  - Power is lower than for z-tests if *n* and Δ are equal
  - Unless we feel very certain about our σ estimates, we should base the power on T-tests

Statistical power

- Stata 13 syntax
  - power *method* μ_0 $(\mu_0 + \Delta)$ [,options]
- One-sample T-test
  - power onemean 0 0.4, n(7) onesided sd(0.4)
    - 0.7544
  - Default:
    - Two-sided test
    - sd(1), power(0.8) and alpha(0.05)

Statistical power

- Stata 13 syntax
  - power *method* μ_0 $(\mu_0 + \Delta)$ [,options]
- One-sample T-test
  - Omitting n() propts sample size calculations
  - power onemean 0 0.4, onesided sd(0.4)
    - N=8 (for $1 - \beta = 0.8$)
  - power onemean 0 0.4, onesided sd(0.4) power(0.95)
    - N=13 (for $1 - \beta = 0.95$)

Statistical power

- Stata 13 syntax
  - power *method* μ_0 $(\mu_0 + \Delta)$ [,options]
- One-sample T-test
  - Adding n() AND power() AND removing $(\mu_0 + \Delta)$ propts effect size calculations
  - power onemean 0, onesided sd(0.4) n(7) power(0.8)
    - delta=1.0667 ($\text{delta} = \delta = \Delta/\sigma$)
    - ma=0.4267 ($\text{ma} = \Delta + 0$)

Statistical power

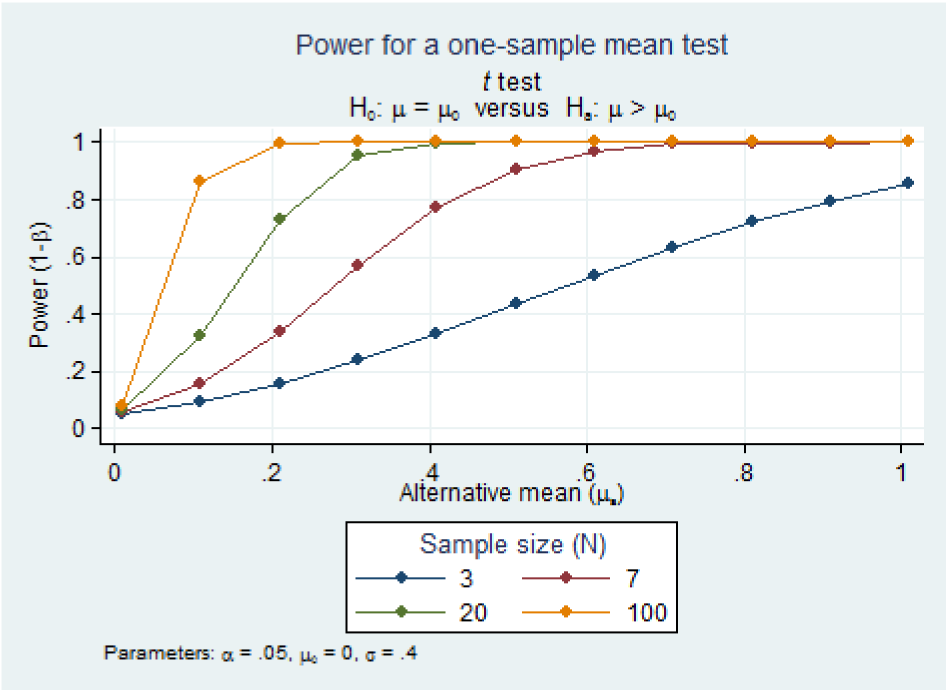
- Stata 13 syntax
  - power *method* μ_0 $(\mu_0 + \Delta)$ [,options]
- One-sample T-test
  - Plots power as function of *n*
    - power onemean 0 0.4, onesided sd(0.4) n(1(1)20) graph
  - Plots power as function of Δ
    - power onemean 0 (0.01(0.1)1.01), onesided sd(0.4) n(7) graph

Statistical power

- Stata 13 syntax
  - power *method* μ_0 $(\mu_0 + \Delta)$ [,options]
- One-sample T-test
  - Plots power as function of *n*
    - power onemean 0 0.4, onesided sd(0.4) n(1(1)20) graph
  - Plots power as function of Δ for several *n*
    - power onemean 0 (0.01(0.1)1.01), sd(0.4) n(3 7 20 100) graph(y(power) x(ma)) onesided

Statistical power

- Stata 13 syntax
  - power *method* μ_0 $(\mu_0 + \Delta)$ [,options]



Statistical power

- Two-sample T-test
- Controls vs. Cases
  - $H_0: \mu_1 = \mu_2$, $H_A: \mu_1 \neq \mu_2$
  - $\sigma_1 = \sigma_2 = 0.4$
  - $\Delta = \mu_2 - \mu_1$
- power twomeans 0 0.4, sd(0.4)
  - N=34
  - N per group=17

Statistical power

- Two-sample T-test
- Controls vs. Cases
  - Assume $\sigma_1 = 0.1$ and $\sigma_2 = 0.4$
- power twomeans 0 0.4, sd1(0.1) sd2(0.4)
  - N=22
  - N per group=11

Statistical power

- Two-sample T-test
- Controls vs. Cases
  - Assume $\sigma_1 = 0.1$ and $\sigma_2 = 0.4$
  - We already had 17 in each group
- power twomeans 0 0.4, sd1(0.1) sd2(0.4) n(34)
  - power=0.9655

Statistical power

- Two-sample T-test
- Controls vs. Cases
  - Assume $\sigma_1 = 0.1$ and $\sigma_2 = 0.4$
  - 9 cases and 7 controls were ineligeble
- power twomeans 0 0.4, sd1(0.1) sd2(0.4) n1(10) n2(8)
  - power=0.6729

Statistical power

- Two-sample T-test
- Controls vs. Cases
  - Assume $\sigma_1 = 0.1$ and $\sigma_2 = 0.4$
  - Which Δ can we detect at 80% power?
- power twomeans 0, sd1(0.1) sd2(0.4) n1(10) n2(8) power(0.80)
  - delta=3.2192 (?)
  - m2=0.4665 (m2=Δ)

Statistical power

- Testing proportions (T-test)
- Test if a drug works
  - H_0: p=30% get well within two days
  - H_A: p>30% get well within two days
  - How many patients do we need if we assume p=50% (so that $\Delta = p - 0.3 = 0.2$)?

Statistical power

- Testing proportions (T-test)
- Test if a drug works
  - H_0: p=30% get well within two days
  - H_A: p>30% get well within two days
  - How many patients do we need if we assume p=50% (so that $\Delta = p - 0.3 = 0.2$)?
- power oneproportion 0.3 0.5
  - N=44

Statistical power

- Testing proportions (T-test)
- Test if a drug works
  - H_0: p=30% get well within two days
  - H_A: p>30% get well within two days
  - For which Δ do we have a power of 0.80 if we only get 35 patients?

Statistical power

- Testing proportions (T-test)
- Test if a drug works
  - H_0: p=30% get well within two days
  - H_A: p>30% get well within two days
  - For which Δ do we have a power of 0.80 if we only get 35 patients?
- power oneproportion 0.3, n(35) power(.8)
  - delta=0.2229 (delta=Δ)
  - pa=0.5229 (pa=$p_0 + \Delta$)

Statistical power

- Testing proportions (χ^2-test)
- Test which of two drugs work the best
  - $\mathrm{H_0}$: $p_1 = p_2$, $\mathrm{H_A}$: $p_1 \neq p_2$
  - $\Delta = p_2 - p_1$
  - Assume $p_1 = 0.75$ and $p_2 = 0.60$
  - Which n do we need to obtain a power of 0.8
- power twoproportions 0.75 0.60

Statistical power

- Testing proportions (χ^2-test)
- Test which of two drugs work the best
  - $\mathrm{H_0}$: $p_1 = p_2$, $\mathrm{H_A}$: $p_1 \neq p_2$
  - $\Delta = p_2 - p_1$
  - Assume $p_1 = 0.75$ and $p_2 = 0.60$
  - Which n do we need to obtain a power of 0.8
- power twoproportions 0.75 0.60
  - N=304
  - N per group=152

Statistical power

- Testing proportions (χ^2-test)
- Test which of two drugs work the best
  - H_0: $p_1 = p_2$, H_A: $p_1 \neq p_2$
  - $\Delta = \mathrm{OR}(p_2, p_1)$
  - Assume $p_1 = 0.75$ and $p_2 = 0.60$
  - Which n do we need to obtain a power of 0.8
- power twoproportions 0.75 0.60, effect(oratio)
  - delta=0.5 (odds-ratio)

Statistical power

- Testing proportions (χ^2-test)
- Test which of two drugs work the best
  - $\mathrm{H_0}$: $p_1 = p_2$, $\mathrm{H_A}$: $p_1 \neq p_2$
  - $\Delta = \mathrm{OR}(p_2, p_1) = 0.5$
  - Assume $p_1 = 0.75$, $n_1 = 30$, $n_2 = 31$
  - What is the power?
- power twoproportions 0.75, n1(30) n2(31) oratio(0.5)
  - power=0.2359

Statistical power

- For linear regression
  - T-test
  - Simulations
- For logistic regression
  - χ^2-test
  - Simulations
- Very difficult to predict how adjustment variables affect power

Statistical power

- Consequences of low power
- The positive predictive (PPV) value is the probability that a positive result is due to a true association
- $\mathrm{PPV} = P(\mathrm{H}_0 \text{ false} | p < 0.05)$
- $\mathrm{PPV} = \frac{(1-\beta)R}{(1-\beta)R+0.05}$
  - R is the odds of H_0 being false
  - $\mathrm{R} = \mathrm{P}(\mathrm{H}_0 \text{ false})/\mathrm{P}(\mathrm{H}_0 \text{ true})$

Statistical power

- $\mathrm{PPV} = \frac{(1-\beta)R}{(1-\beta)R+0.05}$
- Test if gene is associated with disease
  - 1000 candidate genes
  - 3 are associated (we don't know which)
- $R = \frac{0.003}{1-0.003} \approx 0.003$
- $\mathrm{PPV} = \frac{0.8R}{0.8R+0.05}$

Statistical power

- $\text{PPV} = \frac{(1-\beta)R}{(1-\beta)R+0.05}$
- Test if gene is associated with disease
  - 1000 candidate genes
  - 3 are associated (we don't know which)
- $R = \frac{0.003}{1-0.003} \approx 0.003$
- $\text{PPV} = \frac{0.8R}{0.8R+0.05} \approx 0.046$

Statistical power

- $\text{PPV} = \frac{(1-\beta)R}{(1-\beta)R+0.05}$
- Test if medicatin works for disease
  - Worked for similar disease
  - About 75% certain that it works this time too
- $R = \frac{0.75}{1-0.75}$ =3
- $\text{PPV} = \frac{0.8R}{0.8R+0.05}$

Statistical power

- $\text{PPV} = \frac{(1-\beta)R}{(1-\beta)R+0.05}$
- Test if medicatin works for disease
  - Worked for similar disease
  - About 75% certain that it works this time too
- $R = \frac{0.75}{1-0.75}$ =3
- $\text{PPV} = \frac{0.8R}{0.8R+0.05} \approx 0.98$

Statistical power

- $\text{PPV} = \frac{(1-\beta)R}{(1-\beta)R+0.05}$
- Test if medicatin works for disease
  - What if it's either way?
  - 50% sure
- $R = \frac{0.50}{1-0.50}$ =1
- $\text{PPV} = \frac{0.8R}{0.8R+0.05}$

Statistical power

- $\text{PPV} = \frac{(1-\beta)R}{(1-\beta)R+0.05}$
- Test if medicatin works for disease
  - What if it's either way?
  - 50% sure
- $R = \frac{0.50}{1-0.50} = 1$
- $\text{PPV} = \frac{0.8R}{0.8R+0.05} \approx 0.94$

Statistical power

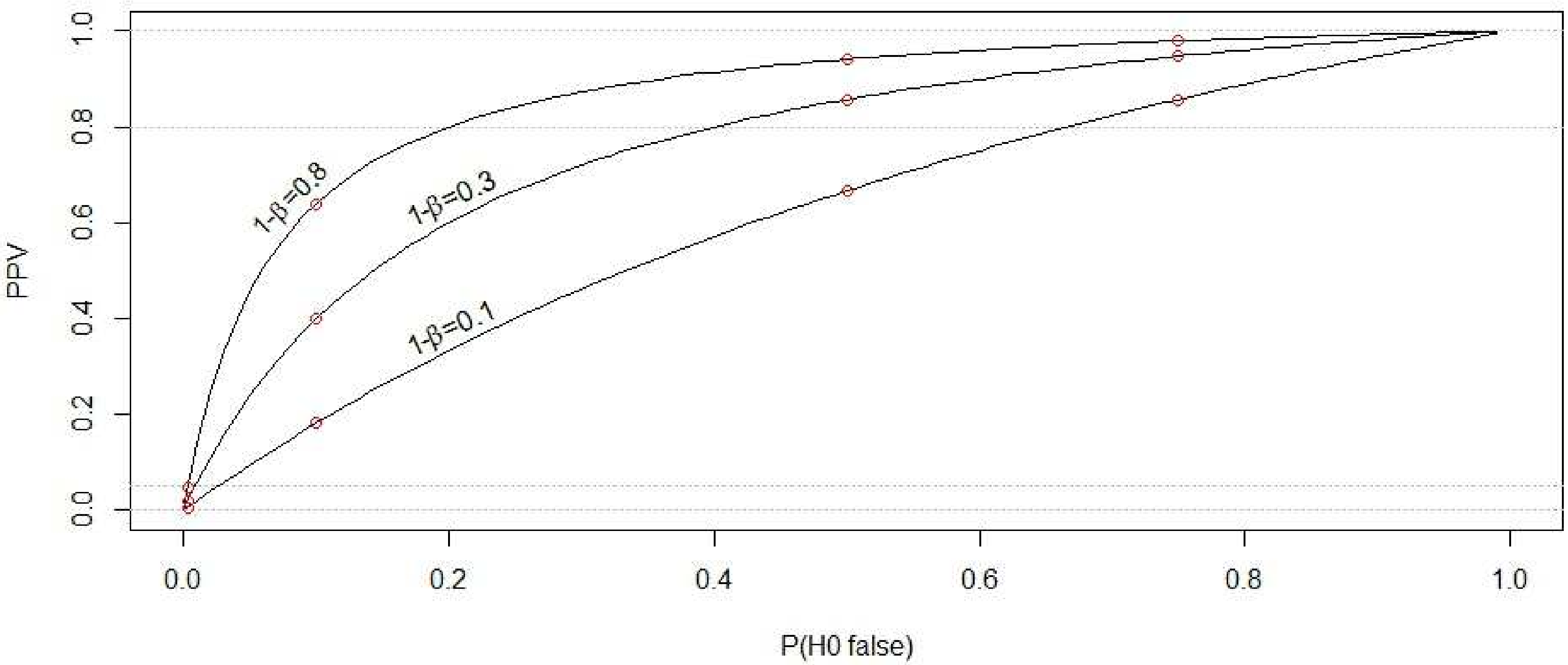
- $\text{PPV} = \frac{(1-\beta)R}{(1-\beta)R+0.05}$
- Test if medicatin works for disease
  - What if the power is low ($1 - \beta = 0.2$)?
  - 50% sure
- $R = \frac{0.50}{1-0.50} = 1$
- $\text{PPV} = \frac{0.2R}{0.2R+0.05}$

Statistical power

- $\text{PPV} = \frac{(1-\beta)R}{(1-\beta)R+0.05}$
- Test if medicatin works for disease
  - What if the power is low $(1 - \beta = 0.2)$?
  - 50% sure
- $R = \frac{0.50}{1-0.50} = 1$
- $\text{PPV} = \frac{0.2R}{0.2R+0.05} = 0.80$

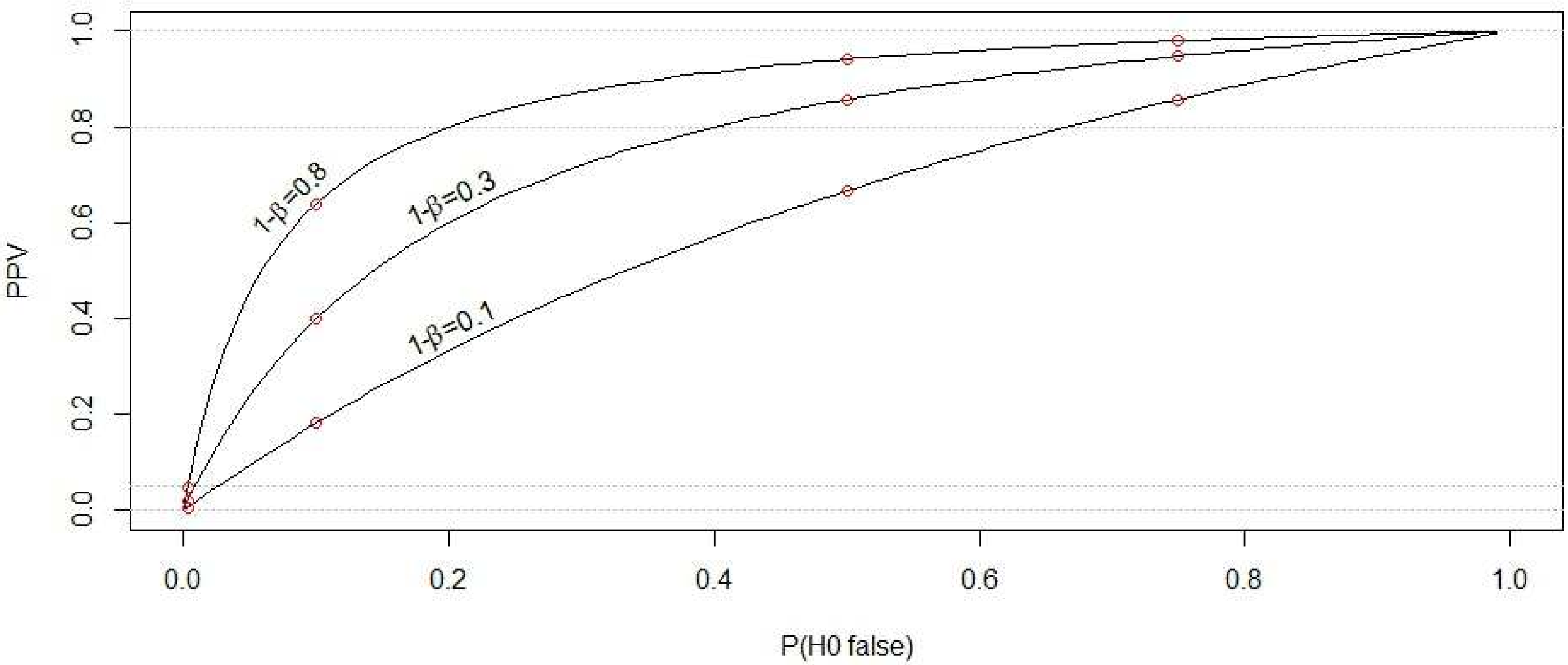
Statistical power

- Different values of power and P($\mathrm{H_0}$ false)



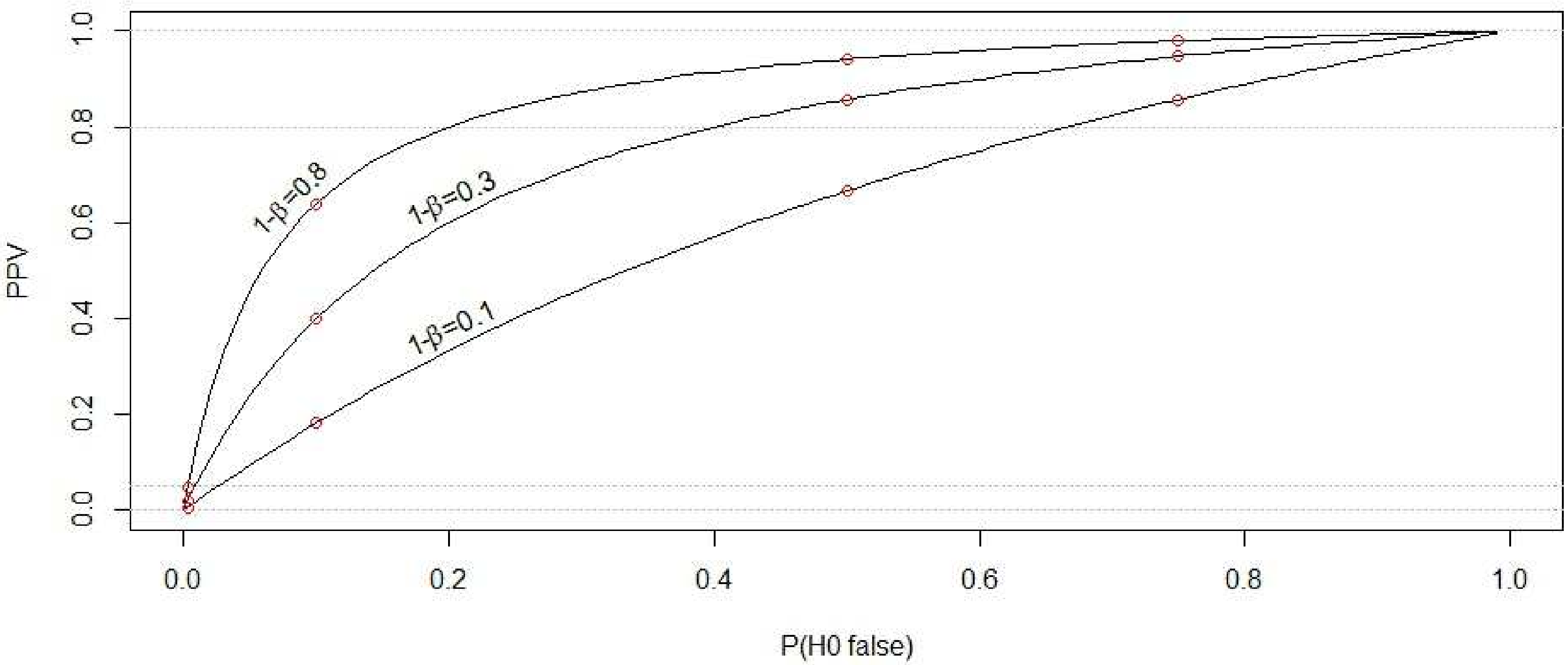
Statistical power

- Surprising results are not likely to be true!



Statistical power

- Low powered results are not likely to be true!



Statistical power

- Recommended reading
  - Ioannidis (2005) Why most published research findings are false. PLoS Medicine.
  - Button et al. (2013) Power failure: Why small sample size undermines the reliability of neuroscience. Nature reviews, neuroscience.
  - http://simplystatistics.org/2013/12/16/a-summary-of-the-evidence-that-most-published-research-is-false/