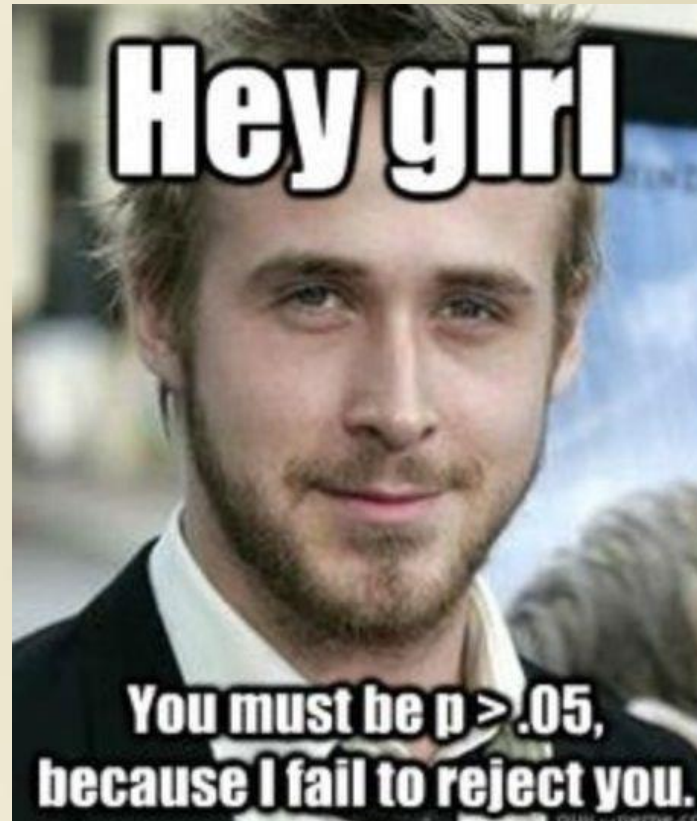


Power, Effect Size & Sample Size Planning

Rizqy Amelia Zein
Departemen Psikologi
Kepribadian dan Sosial



FAKULTAS
PSIKOLOGI
UNIVERSITAS AIRLANGGA

*Imagining
Learning
& Creating
for life*

psikologi.unair.ac.id

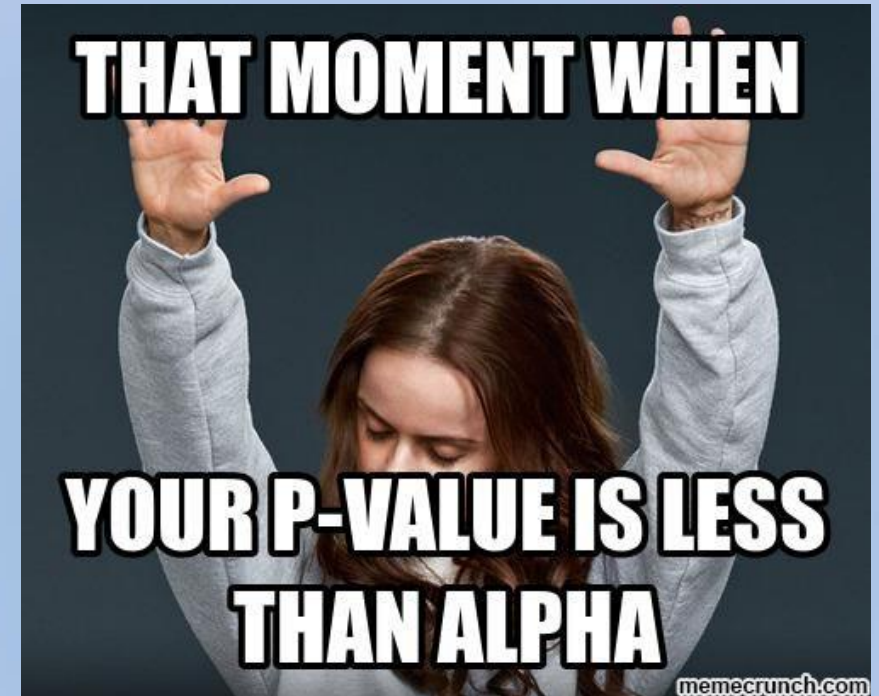
Taraf signifikansi: masih signifikan? (1)

- Hampir semua artikel ilmiah di Psikologi menyandarkan kesimpulan inferensialnya dengan menggunakan informasi mengenai **taraf signifikansi** (*p-value - statistical significance*) sebagai **patokan/bukti** untuk **menerima/menolak hipotesis**, padahal informasi ini **bermasalah dalam tataran praktis**.
 - Signifikan = seberapa beda?
 - Signifikan = seberapa kuat/besar?
 - Taraf signifikansi **tidak mengandung informasi apapun** mengenai jarak antara dua kelompok yang diteliti (pada kasus penelitian komparasi antara dua kelompok; *substantive significance* – Kelley & Preacher 2012), dan taraf signifikansi tidak dapat dibandingkan antar studi.
 - Akibatnya, nilai informasi suatu artikel ilmiah menjadi amat kecil.



Taraf signifikansi: masih signifikan? (2)

- Untuk mengatasi persoalan ini, American Psychological Association (APA) merekomendasikan bahwa seluruh laporan analisis statistik dalam semua naskah terpublikasi **harus mencakup informasi mengenai *effect size*** (lihat APA 5th edition manual, 1.10: Results section).



Taraf signifikansi: masih signifikan? (3)

- Apabila peneliti **tidak berintensi** melakukan generalisasi (hanya tertarik pada temuan di level sampel), **taraf signifikansi tidak diperlukan**. Dalam situasi ini, informasi mengenai **effect size** saja sudah memadai untuk penarikan kesimpulan.
- Ketika menginvestigasi *effect size* dengan **sample size yang kecil**, **taraf signifikansi** akan memberikan **informasi yang menyesatkan**. Berbeda dengan yang umum diketahui, taraf signifikansi **bukan indikator besarnya efek**, melainkan fungsi dari *sample size*, *power* dan *p-values*.
- Ketika menginvestigasi *effect size* dengan **sample size yang besar**, taraf signifikansi dapat memberikan **informasi yang menyesatkan**, karena **efek yang amat kecil dan trivial** dapat sama-sama **menghasilkan kesimpulan signifikan** secara statistik.

(Neill, 2008)

<http://www.wilderdom.com/research/effectsizes.html>



Taraf signifikansi: masih signifikan? (4)

Kasus 1

- Joko mengambil data dari 2 kelompok sampel yang independen, dengan sampel masing-masing 10 tiap kelompok ($N=20$).
- $M_1=6$, $SD_1=3.16$ dan $M_2=7$, $SD_2=3.16$ menghasilkan $t=-0.5$, $p=0.63$ ($p>0.05$) --- **H0 gagal ditolak!**

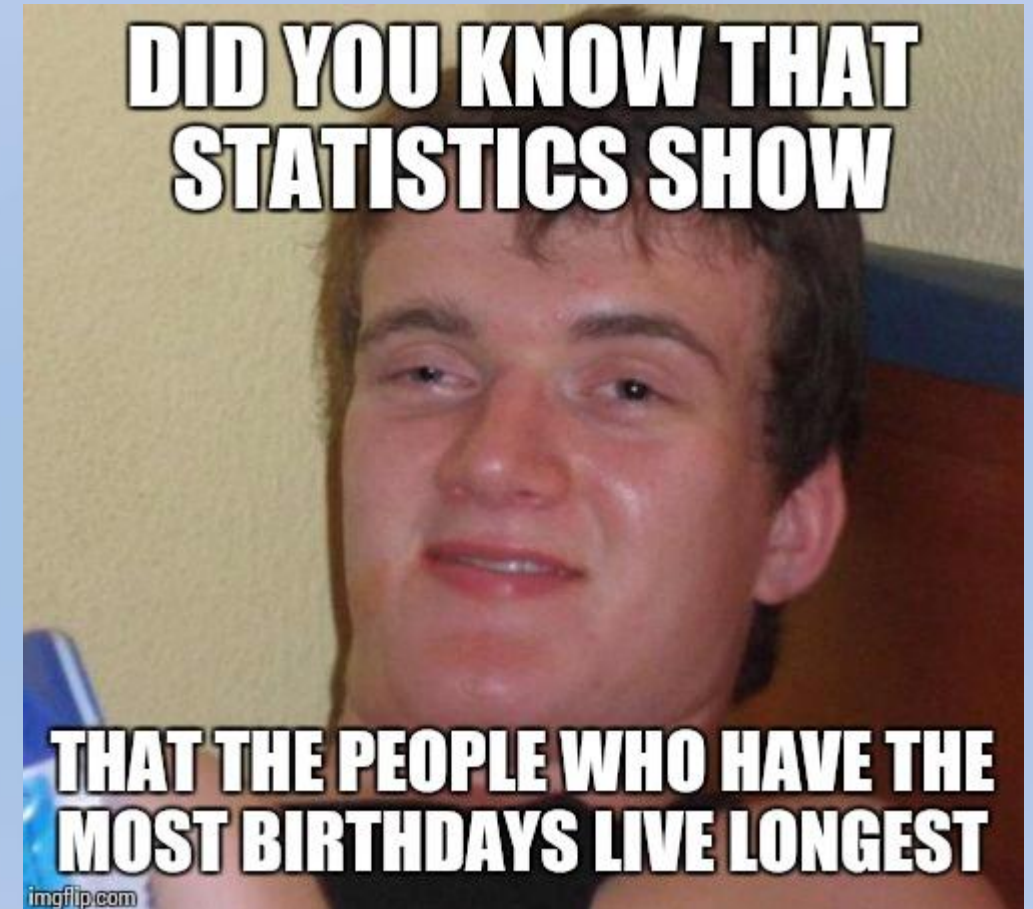
Kasus 2

- Bowo mengambil data dari 2 kelompok sampel yang independen, dengan sampel masing-masing 100 tiap kelompok ($N=200$).
- $M_1=6$, $SD_1=3.16$ dan $M_2=7$, $SD_2=3.16$ menghasilkan $t=-2.46$, $p=0.02$ ($p<0.05$) --- **H0 ditolak!**

Ini menunjukkan p-value sangat sensitif dengan sample size!!

Effect size (1)

- Definisi yang paling lugas dari *effect size* adalah “...***the magnitude, or size, of an effect..***” (Cohen 1992)
- Kadang-kadang didefinisikan sebagai **nilai relatif** dibandingkan dengan ***null effect*** (Kelly & Preacher 2012)
- *Effect size* dapat dibandingkan antar studi
- Salah satu bentuknya adalah Cohen’s *d*, atau hanya “*d*.” Digunakan dalam uji perbedaan antara 2 kelompok independen (*independent sample t-test*)



Effect size (2)

- Interpretasi d (Cohen 1988) (**disclaimer**: Cohen memohon patokan ini diabaikan)
 - .8 = besar (8/10 dari unit standar deviasi)
 - .5 = sedang (1/2 dari unit standar deviasi)
 - .2 = kecil (1/5 dari unit standar deviasi)
- Uji hipotesis yang berbeda menggunakan parameter *effect size* yang berbeda (Lalongo 2012)
- <http://imaging.mrc-cbu.cam.ac.uk/statswiki/FAQ/effectSize>



Bagaimana cara menggunakan *p-value* dan ES?

P-value/ES	Besar	Kecil
Signifikan	Efek IV terhadap DV besar dan cenderung konsisten	Efek IV terhadap DV sebenarnya kecil , tetapi mengalami overestimasi karena <i>sample size</i> besar
Tidak signifikan	Efek IV terhadap DV besar, tetapi <u>power terlalu rendah</u> untuk mendeteksi efek tsb atau Efek IV terhadap DV besar, tapi reliabilitasnya tidak diketahui.	Efek IV terhadap DV <u>kecil</u>

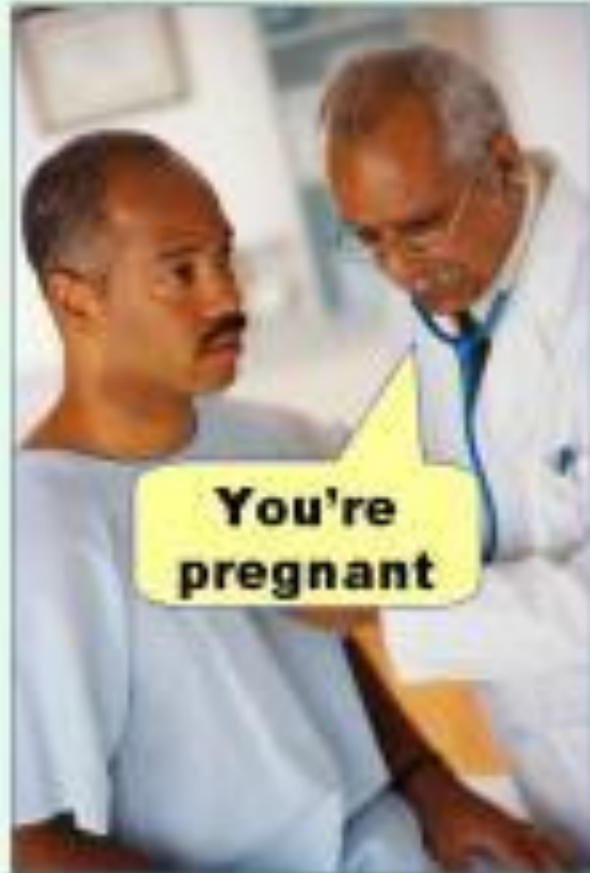


Effect size dan power

- Kalau pada sampel kecil *nothing is significant*, sedangkan pada sampel besar *anything is significant...*
- Lalu bagaimana cara kita menentukan jumlah sampel yang tidak terlalu kecil dan tidak terlalu besar?
- Maksimalkan *statistical power*
- Apa itu *statistical power*?



Type I error
(false positive)



Type II error
(false negative)



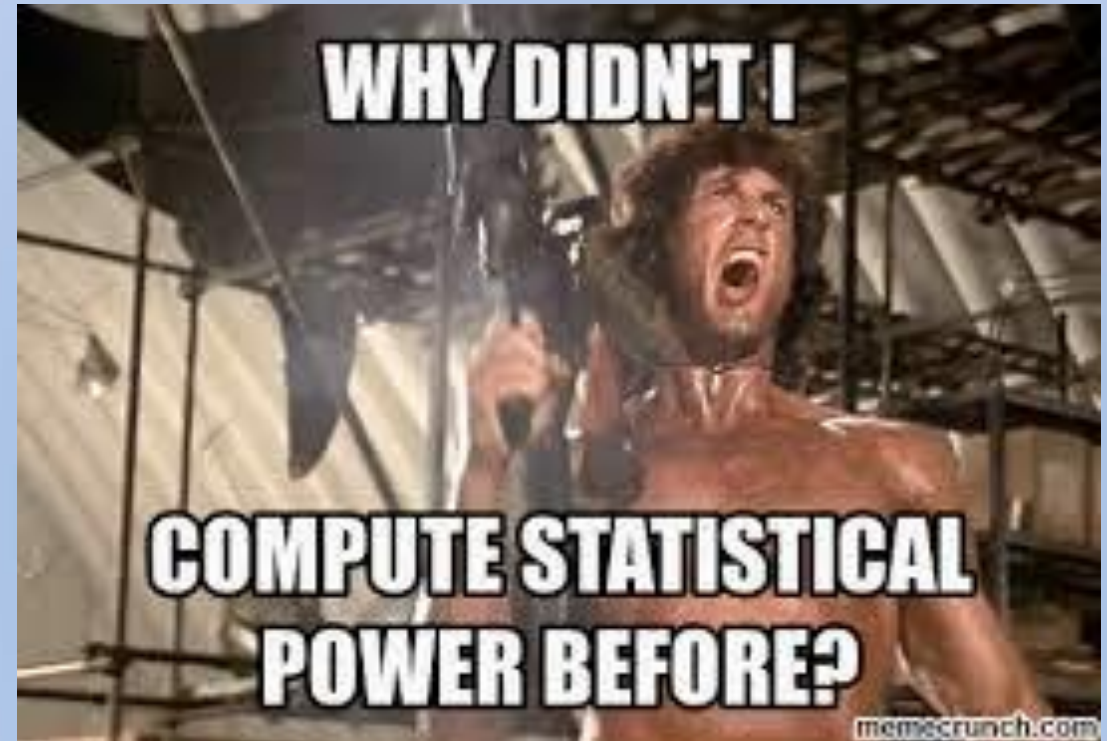
Luruskan definisi... (1)

- ***P-value***: probabilitas peneliti mendapatkan nilai effect size yang ekstrim, dengan asumsi H0 benar (no effect) → berlaku dalam jangka panjang (pada N studi = ∞)
 - Amel berasumsi bahwa semua dosen Fakultas Psikologi sama cantiknya
 - Setelah Amel bandingkan kecantikannya dengan dosen yang lain, lho... ternyata Amel adalah yang tercantik di Fakultas Psikologi
 - *P-value* adalah “...peluang dosen lain bisa secantik Amel, kalo seandainya benar bahwa semua dosen sama cantiknya”
- **Kesalahan tipe 1**: menyimpulkan ada efek/hasil signifikan ($p < .05$), ketika H0 yang benar (no effect)
- **α** : ***probabilitas*** terjadinya kesalahan tipe 1



Luruskan definisi... (2)

- **Kesalahan tipe 2**: menyimpulkan **tidak ada efek**/hasil tidak signifikan ($p > .05$), ketika **H1 yang benar**
- **β : *probabilitas*** terjadinya kesalahan tipe 2
- **$1-\beta$ (power): *probabilitas*** mendeteksi **adanya efek**/hasil signifikan ($p < .05$), ketika **H1 yang benar**.
- **α dan β** seperti jungkat-jungkit.
 - Yang satunya naik, pasti mengakibatkan satunya turun.



**P-value/Kondisi
riil**

Signifikan

**Tidak
signifikan**

H0 benar

H1 benar

False Positive
(α)

True Positive
($1-\beta$)

True Negative
($1-\alpha$)

False Negative
(β)



Misalnya...

- Esmeralda ingin menguji hipotesis:

H1: Makan Samyang dapat menyebabkan diare

H0: Makan Samyang tidak berdampak apapun pada frekuensi pup perharinya

- Probabilitas H1:H0 adalah 50:50, kita tidak tahu mana hipotesis yang paling mungkin
- Kita gunakan asumsi *default*
 - α : 0.05/5% sehingga $1 - \alpha = 0.95 / 95\%$
 - β : 0.2/20%, sehingga $1 - \beta = 0.8 / 80\%$



P-value/Kondisi riil	H0 benar (50%)	H1 benar (50%)
Signifikan $\alpha: 5\%; 1-\beta: 80\%$	False Positive $5\% * 50\% = 2.5\%$	True Positive $80\% * 50\% = 40\%$
Tidak signifikan $1-\alpha: 95\%; \beta: 20\%$	True Negative $95\% * 50\% = 47.5\%$	False Negative $20\% * 50\% = 10\%$



Tanpa melakukan apapun, bahkan
sebelum kita mengambil data,
True Negative merupakan
outcome yang paling mungkin

Lalu apa yang bisa kita lakukan?

P-value/Kondisi riil	H0 benar (50%)	H1 benar (50%)
Signifikan α : 5%; $1-\beta$: 99%	False Positive $5\% * 50\% = 2.5\%$	True Positive $99\% * 50\% = 49.5\%$
Tidak signifikan $1-\alpha$: 95%; β : 1%	True Negative $95\% * 50\% = 47.5\%$	False Negative $1\% * 50\% = 0.5\%$



P-value/Kondisi riil	H0 benar (50%)	H1 benar (50%)
Signifikan α : 1%; $1-\beta$: 80%	False Positive $1\% * 50\% = 0.5\%$	True Positive $80\% * 50\% = 40\%$
Tidak signifikan $1-\alpha$: 99%; β : 20%	True Negative $99\% * 50\% = 49.5\%$	False Negative $1\% * 50\% = 0.5\%$

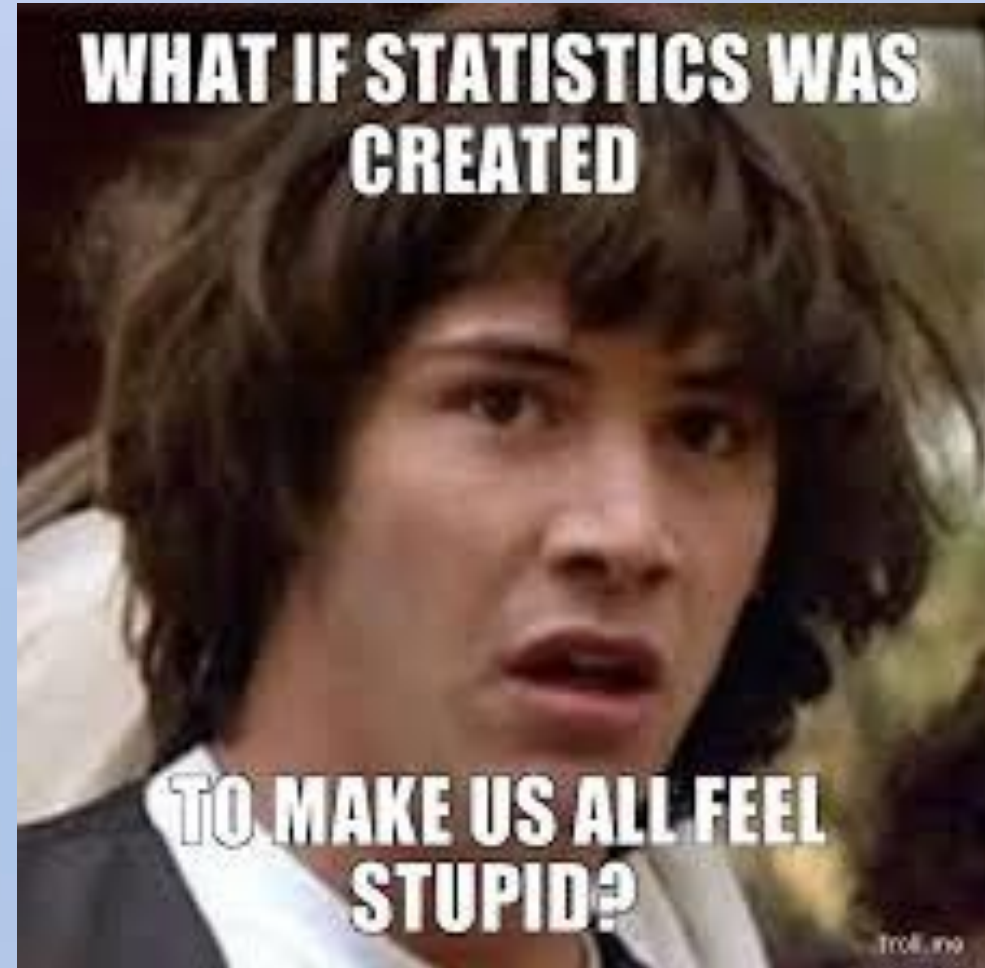
P-value/Kondisi riil	H0 benar (10%)	H1 benar (90%)
Signifikan $\alpha: 5\%; 1-\beta: 80\%$	False Positive $5\% * 10\% = 0.5\%$	True Positive $80\% * 90\% = 72\%$
Tidak signifikan $1-\alpha: 95\%; \beta: 20\%$	True Negative $95\% * 10\% = 9.5\%$	False Negative $20\% * 90\% = 18\%$

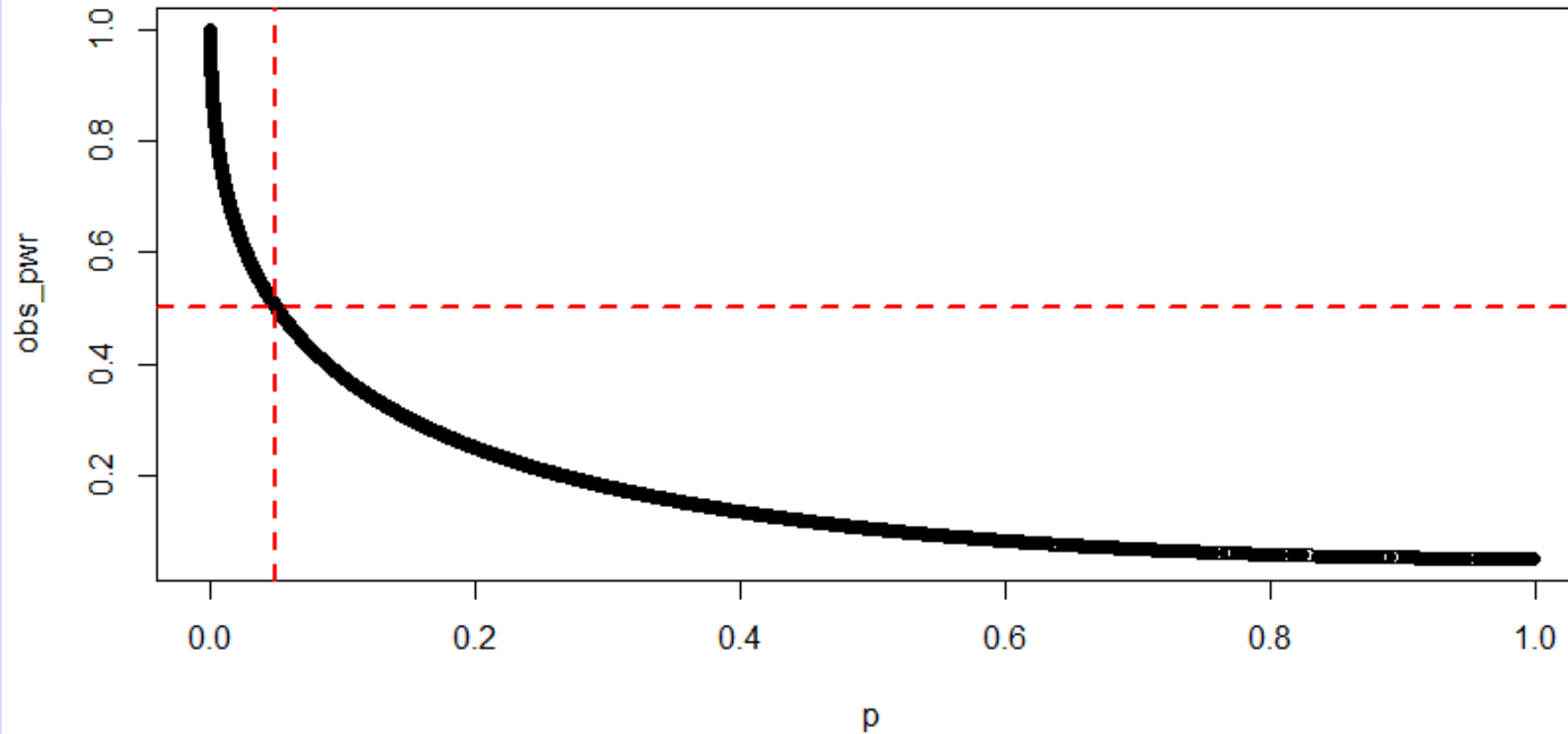


Seandainya $p=0.05$, berapa *power*
yang dimiliki?

Seandainya...

- Kita dapat membuat simulasi 100.000 studi dengan rata-rata *power* sebesar 90%.
- Berikut adalah *scatterplot* yang menggambarkan korelasi antara *p-value* dengan *power*



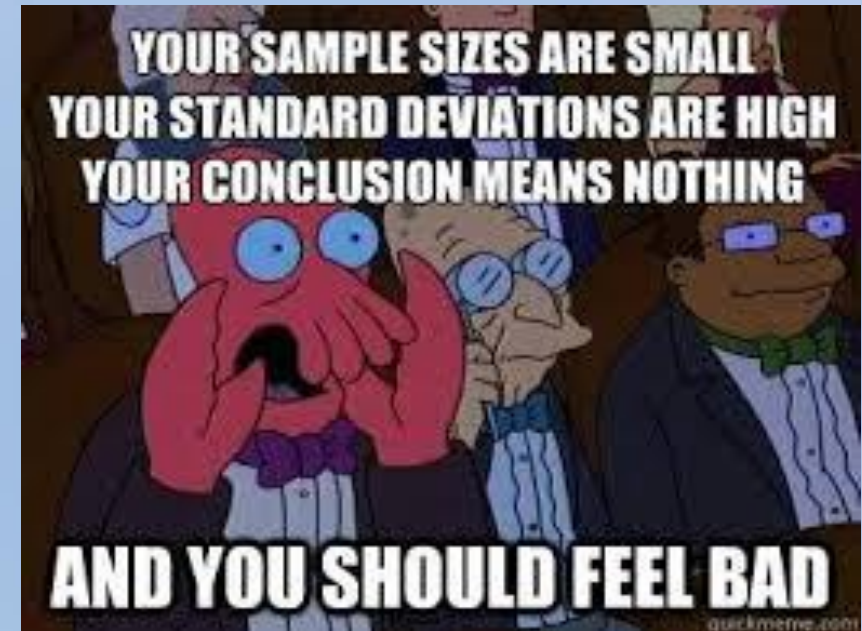


Syntax credit: Lakens, 2014



Artinya...

- Dalam jangka panjang ($N \sim$), studi dengan ***p-value*=0.05** kemungkinan memiliki *power* sebesar **hanya 50% saja!**
- Bahkan Cohen (1962) menyatakan bahwa rata-rata *power* penelitian Psikologi adalah 50%
 - Berarti **sebagian besar penelitian** yang dikutip oleh buku teks acuan kita kemungkinan besar adalah *false positive*
- Ada banyak sekali penelitian Psikologi (yang dipublikasikan di jurnal *high-impact*) yang melaporkan ***p-value* yang mendekati 0.05** (contoh: 0.049, 0.04, ...) tapi **dilaporkan** sebagai **temuan yang signifikan** 😞
- Yang lebih gawat... ada beberapa peneliti yang melaporkan *p-value*=0.051, 0.055, dst. sebagai *marginally significant!*



Open access, freely available online

Essay

Why Most Published Research Findings Are False

John P. A. Ioannidis



Berbagai mitos...

- Dalam paradigma ***null hypothesis significance testing (NHST)***, ketika kita melakukan uji hipotesis, kita **hanya** melakukan **kontrol** pada **kesalahan tipe 1**, sehingga kita hanya bisa menghasilkan **bukti yang mendukung/menolak H_0** → Fisher's NHST!
- Kita **cenderung 'mengabaikan' kesalahan tipe 2** (Power Primer, Cohen) --- tapi anehnya, kalau H_0 ditolak, kenapa H_a/H_1 yang diterima? Padahal kan yang diuji H_0 ?
- Ketika H_0 gagal ditolak ($p > 0.05$), ada dua kemungkinan:
 - **Mungkin** H_0 benar (sehingga harus diterima)
 - **Mungkin** *sample size*nya tidak cukup besar untuk menolak/menerima H_0 , sehingga *power*nya rendah!
- Namun *power* sebaiknya diestimasi sebelum studi (*a priori*), dengan merencanakan jumlah sampel sebelum pengambilan data.



Bagaimana mengoptimisasi *power*?

- Meningkatkan $\alpha \rightarrow$ misal dari umumnya 0.05 menjadi 0.1
 - tentu saja ini tidak diinginkan, kita mau α dan β sama-sama rendah.
- Naikkan *effect size*
 - Pilih prediktor dengan justifikasi yang jelas
 - Dalam eksperimen, sebaiknya bandingkan *treatment* dengan *no treatment*
- Tes statistik yang *one-tailed* memiliki *power* yang lebih tinggi daripada *two-tailed test*
- Naikkan *n* (*sample size*) atau turunkan *s* (standar deviasi)
 - Dalam menghitung *test statistics*, kita membagi standar deviasi dengan akar *n* (s/\sqrt{n}) - penurunan *s* membuat *t* menjadi lebih kecil, seperti halnya meningkatkan *n* - membagi dengan jumlah yang lebih kecil memberi kita nilai yang lebih besar dari *z* atau *t*, yang menghasilkan peningkatan peluang menolak H_0



Berapa besar *power* yang optimal?

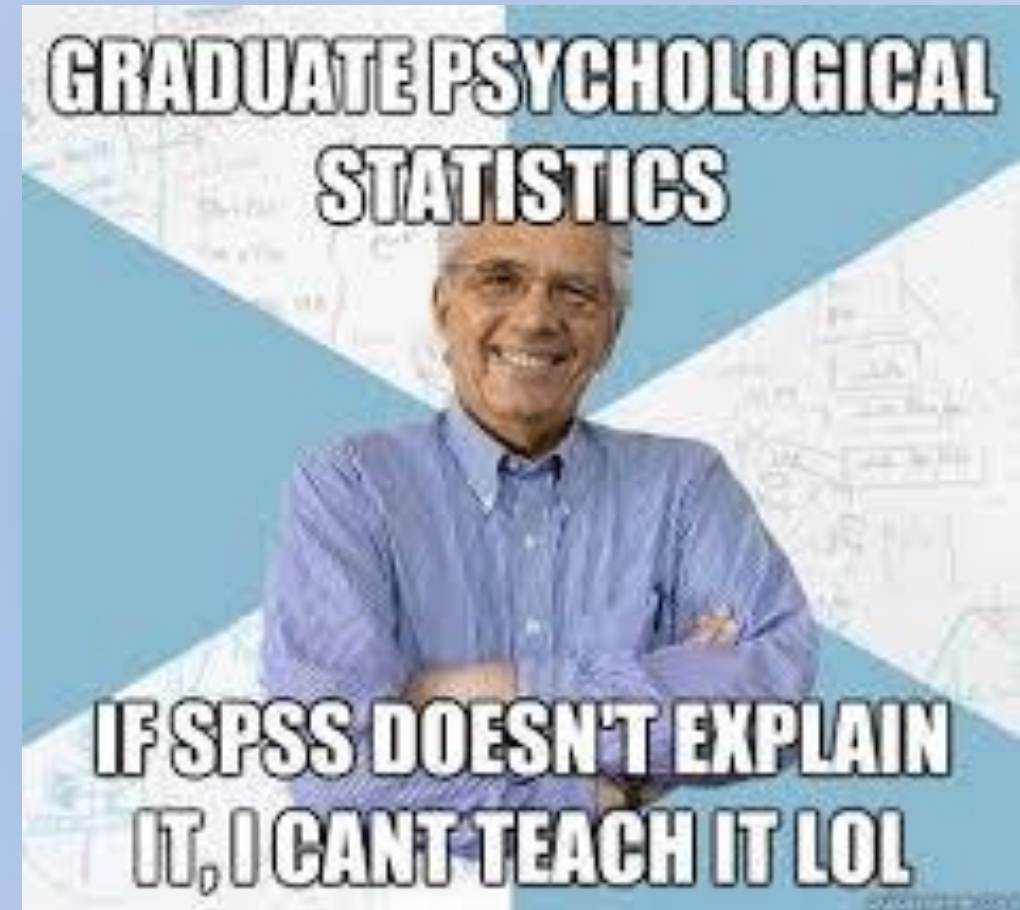
- Umumnya, konsensus menyatakan 0.8 (80%) termasuk memadai untuk mendeteksi adanya efek.
 - Power=0.8 → 80% peluang mendeteksi adanya efek, 20% peluang melakukan kesalahan tipe 2
- Sebelum melakukan pengambilan data, kita dapat melakukan *power analysis* (secara *a priori*) yang memberikan informasi kepada kita **berapa banyak jumlah sampel yang dibutuhkan untuk mendeteksi ES sebesar x dengan *power* sebesar 80%**
- *Software* yang dapat digunakan:
 - R (*package* pwr, PowerR, psych, dll)
 - jamovi (*module* jpower)
 - G*Power (<http://www.gpower.hhu.de/>)



**Cohen (1992) merekomendasikan
1- β (power) setidaknya = 0.8 / 80%
(tapi ia berharap rekomendasi ini diabaikan) 😊**

Mengapa *power* diabaikan ☹️

- Membingungkan ☹️
- Pemahaman yang keliru soal teknik statistik
 - Statistik (pendekatan *frequentist* utamanya) adalah *the science of changing our mind (from default action)*
 - H_0 adalah *default action*
 - Sering disalah artikan sebagai cara untuk *confirming/predicting* fenomena, padahal dengan statistik kita hanya mendapat informasi tentang derajat ketidakpastian (*uncertainty*), bukan menegaskan/mendapatkan/melegitimasi kepastian-kepastian.
- SPSS tidak pernah mengeluarkan *power analysis*
 - Untuk versi 25 disediakan, tapi harus *install add-on* tambahan



Imagining
Learning
& Creating
for life



Sampling error & sample size planning

Sampling error

- ..merupakan besaran **variasi** proporsi/*mean* **sampel** dibandingkan dengan **variasi** proporsi/*mean* **populasi**..
- Dikenal juga dengan *margin error*
- Dasarnya dari dua konsep yang saling berkaitan: *standard error* dan *confidence level*
- *Standard error* dihitung berdasarkan....
- ...derajat variasi (*standard deviation*) yang dibandingkan dengan jumlah sampel



Standard error (of proportion)

- p = proporsi yang akan diukur SE-nya
- n = jumlah sampel

Dalam sebuah survei elektabilitas calon Presiden, diketahui pemilih Alfonso = 33%, sedangkan Marimar = 31% ($n=1000$), berapa SEnya?

$$\text{Standard error} = \sqrt{\frac{p(1-p)}{n}}$$

$$\begin{aligned} \text{SE} &= \sqrt{(p \cdot [1-p] / n)} \\ &= \sqrt{(0.33 \cdot [1-0.33] / 1000)} \\ &= 0.015 \text{ atau sekitar } 1.5\% \end{aligned}$$



Menghitung *sampling error*

- Untuk menghitung *sampling error* dengan tingkat kepercayaan (*confidence level*) sebesar 95%, maka kita akan mengalikan SE dengan nilai Z (distribusi normal/*Gaussian*)
 - 95% [1.96]
 - 98% [2.32]
 - 99% [2.58]
- Oke, kita mau *defaultnya* aja (95%) berarti $SE \times 1.96$ ---- $0.015 \times 1.96 = 0.29 = \pm 2.9\%$
- Berarti di populasi, Alfonso kemungkinan mendapatkan 30.1% - 39.9% dan Marimar 28.1% - 33.9% → *rentang kepercayaan/confidence interval (CI)*
- **Tetapi!** Kita hanya **yakin 95% proporsi di populasi jatuh di rentang ini**, masih ada **peluang 5%** proporsi sebenarnya di populasi **jatuh diluar rentang tsb.**
- SE kecil = CI makin sempit = estimasi makin akurat

Confidence interval

- Semua parameter statistik (koefisien korelasi, ES, *path coefficient*, *mean*, bahkan koefisien reliabilitas) dapat dihitung sekalian SEnya.
 - Padahal parameter tsb kan hanya mencerminkan sampel
 - Kalo estimasi parameter di populasi? Gunakan *confidence interval*!
 - Jadi, *confidence interval* merupakan estimasi kita atas parameter di populasi yang sifatnya tidak diketahui (*unknown*)
- Oleh karena itu....
- Laporkan ES dan berikan *confidence intervalnya*.

Contoh:

- Setelah dilakukan analisis Spearman's Rho, korelasi antara dukungan sosial dan depresi diketahui positif dan amat kuat (0.89, 95% CI [0.81 – 0.92])
- Koefisien reliabilitas skala OCD adalah $\alpha = 0.97$ (95% CI 0.91 – 0.99)

1. Fernando Jose ingin mengukur intelegensi mahasiswa Psikologi yang N populasinya = 1400, dengan mengambil sampel sebanyak 500 orang
2. Maria Mercedes ingin mengukur intelegensi remaja di Jawa Timur dengan N populasinya = 15000, dengan mengambil sampel sebanyak 500 orang

**Mana yang lebih akurat &
representatif?**

Jawabannya?
Kekuatan estimasinya sama!

Meminimalisasi *sampling error*

- Tidak ada kaitannya dengan besar populasi (yang biasanya *unknown*).
 - Mau di populasi 1000, 100 juta dan 10 milyar, *sampling error* pada kelompok sampel $n=100$ akan sama saja.
- Kalau mau mengurangi *sampling error*, maka gedeinlah jumlah sampel!

25 partisipan +/- 20%

50 partisipan +/- 14%

100 partisipan +/- 10%

200 partisipan +/- 7%

400 partisipan +/- 5%

1,000 partisipan +/- 3%

2,000 partisipan +/- 2%

10,000 partisipan +/- 1%



Menentukan *sample size*

...agar *sample size* tidak terlalu sedikit, dan tidak terlalu besar...

- *A priori* power analysis
- *Sensitivity analysis*
- *Accuracy of interval parameter estimation* (AIPE)
- Monte carlo simulation (MC)
 - Digunakan juga untuk mengimpute data yang *missing* (menggunakan Markov Chain Monte Carlo (MCMC))

Kita akan coba teknik yang pertama dan kedua.



A priori power (1)

- Prinsipnya digunakan untuk mengestimasi jumlah sampel yang dibutuhkan untuk mendeteksi ES sebesar n , dengan diketahui:
 - $\alpha = 0.05$; dan
 - $1-\beta = 0.8/80\%$
 - Teknik statistik tertentu
- Sudah banyak dianjurkan oleh jurnal-jurnal *high impact* untuk menentukan jumlah sampel sekaligus *rule of data collection stopping*.
- Mencegah peneliti curang, yaitu sengaja menambah responden agar hasilnya signifikan \rightarrow *p-value fishing*



A priori power (2)

- Amel ingin menguji obat diare baru ciptaannya. Ia memberikannya pada **sekelompok pasien diare**, lalu **mencatat frekuensi pup** mereka dalam seminggu dan dibandingkan dengan **frekuensi pup sekelompok pasien lain** yang tidak diberikan apa-apa.
- Amel tidak yakin obatnya sangat ampuh. Ia berasumsi mungkin efektif, tapi kecil efeknya ($d=0.2$).
- Namun tentu saja ada kemungkinan obat tsb tidak ampuh atau malah menimbulkan efek samping.
- Berapa banyak responden yang harus diambil untuk mendeteksi ES sebesar $d=0.2$ dengan $\alpha = 0.05$ dan $1-\beta = 0.8$?



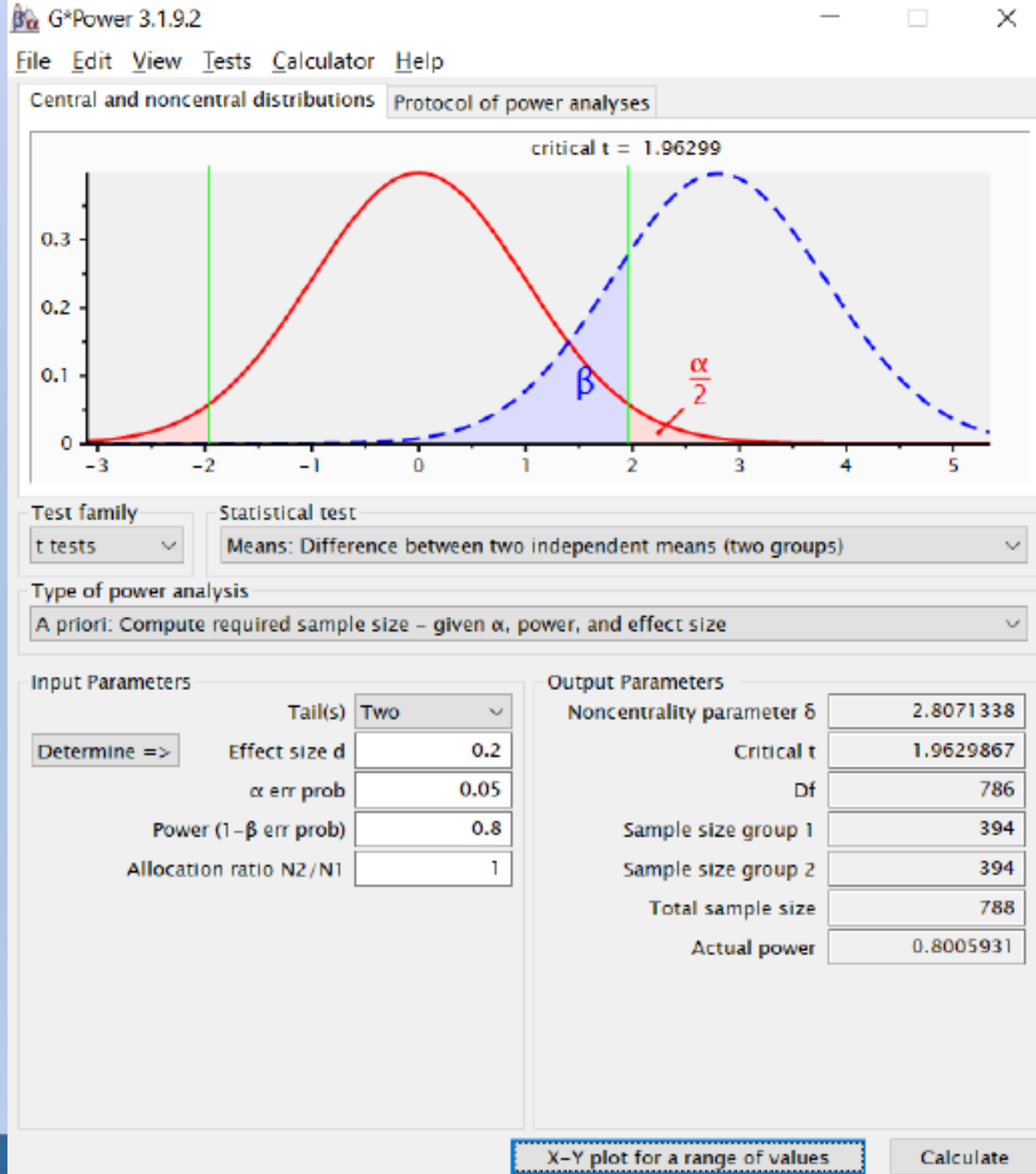
Menggunakan G*Power (*t*-test)

- Dalam opsi test family, pilih ***t-test***.
- Dalam opsi statistical test, pilih ***Means: difference between two independent means (two group)***.
- Untuk opsi type of power analysis, pilih ***a priori: computed required sample size – given α , power and effect size***.
- Pilih **two-tailed test**, **ES = 0.2**, **$\alpha=0.05$** , **power=0.8**, dan **allocation ratio N1/N2** (rasio antara kelompok *treatment* dan kontrol) = **1**



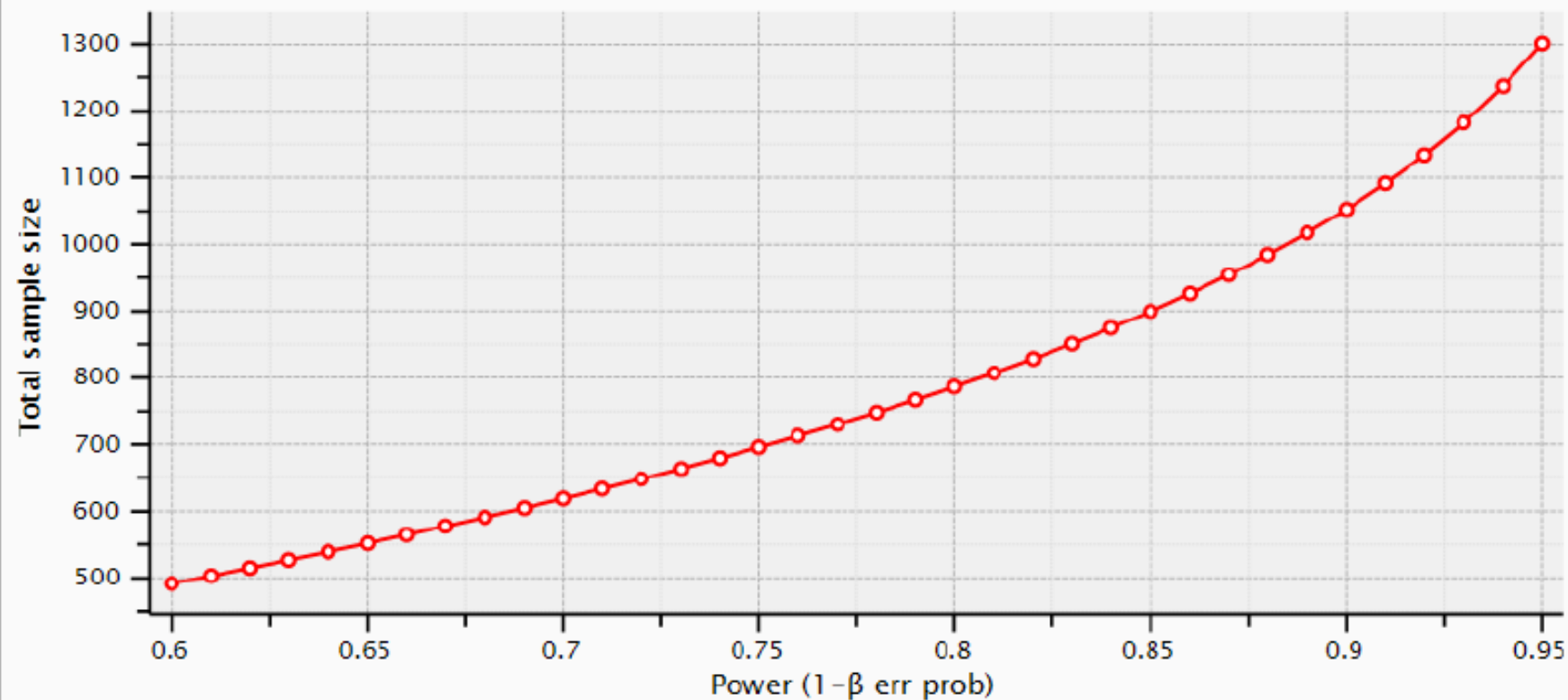
Interpretasi

- Untuk mendeteksi $d=0.2$, maka Amel harus mendapatkan setidaknya 394 responden utk masing2 kelompok (total $N=788$)
- Coba klik *X-Y plot for range of values*



t tests – Means: Difference between two independent means (two groups)

Tail(s) = Two, Allocation ratio $N2/N1 = 1$, α err prob = 0.05, Effect size $d = 0.2$



Plot Parameters

Plot (on y axis) Total sample size ☒ with markers ☐ and displaying the values in the plot
as a function of Power ($1 - \beta$ err prob) ☐ from 0.6 in steps of 0.01 through to 0.95
Plot 1 ☐ graph(s) interpolating points ☐
with Effect size d ☐ at 0.2
and α err prob ☐ at 0.05

Draw plot



A priori power (3)

- Amel ingin melakukan penelitian untuk menyelidiki hubungan antara jumlah setoran suami tiap bulan dengan berat badan.
 - Menggunakan korelasi bivariat.
- Beberapa literatur menyebutkan ada hubungan yang **positif** dan **cenderung moderate** antara kedua variabel ini.
 - Katakanlah $r=0.4$
- Kedua variabel diharapkan berdistribusi normal.
- Berapa banyak responden yang harus diambil untuk mendeteksi ES sebesar $r=0.4$ dengan $\alpha = 0.05$ dan $1-\beta = 0.8$?



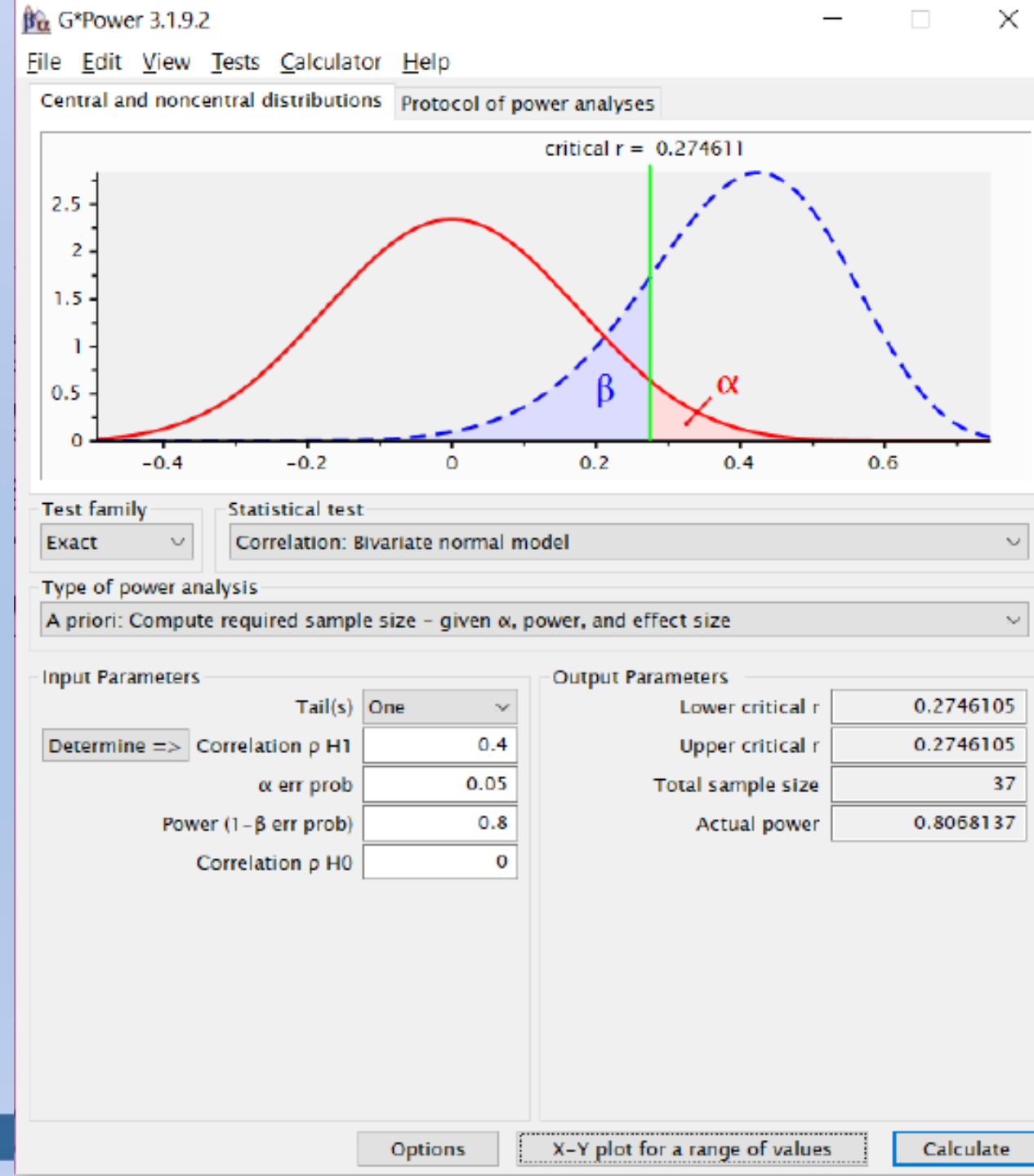
Menggunakan G*Power (korelasi bivariat)

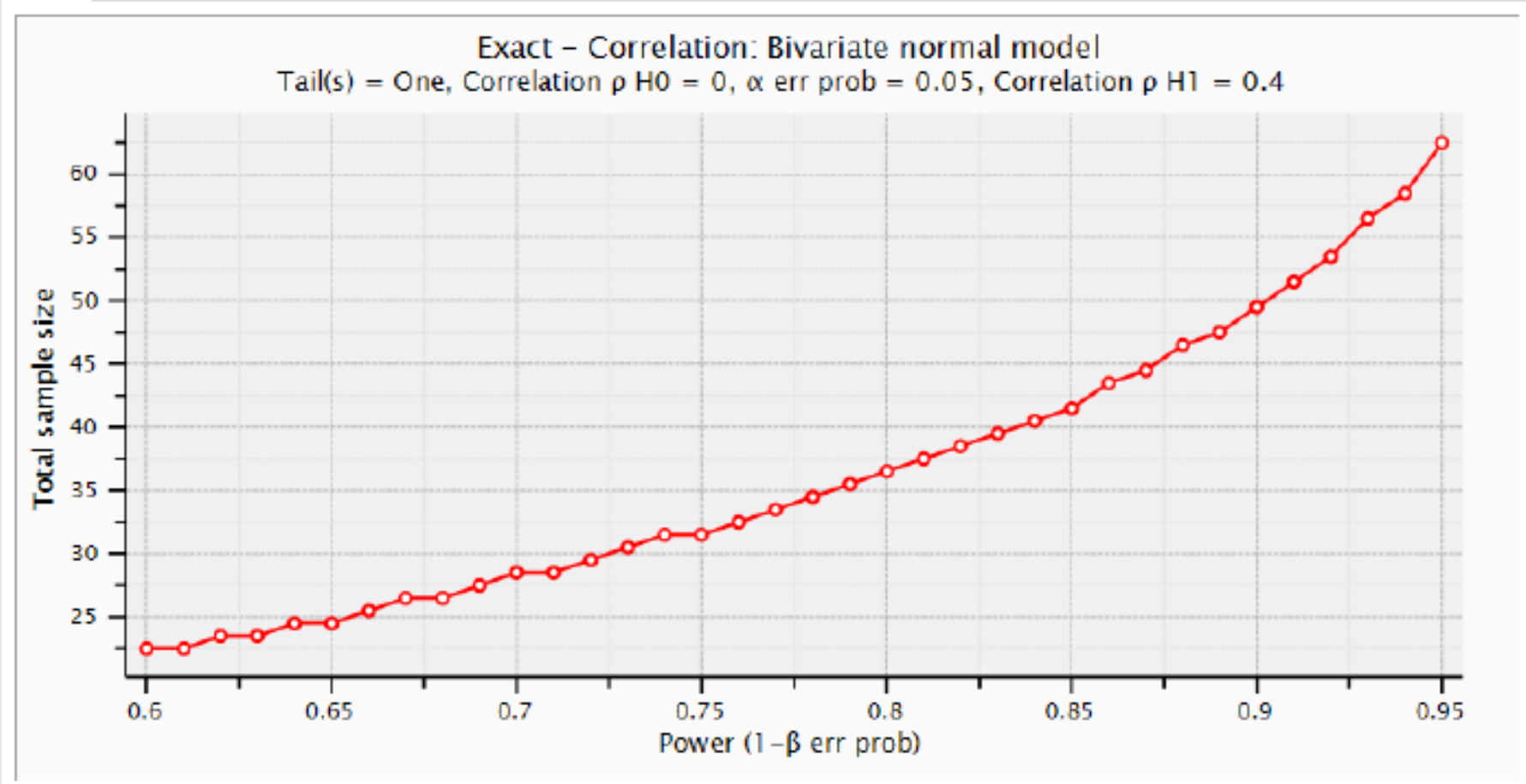
- Dalam opsi test family, pilih **exact**.
- Dalam opsi statistical test, pilih **Correlations: bivariate normal model**
- Untuk opsi type of power analysis, pilih **a priori: computed required sample size – given α , power and effect size**.
- Pilih one-tailed test, correlation ρ H1 = 0.4, $\alpha=0.05$, power=0.8, dan correlation ρ H0 = 0



Interpretasi

- Untuk mendeteksi $r=0.4$, maka Amel harus mendapatkan setidaknya 37 responden
- Coba klik *X-Y plot for range of values*





Plot Parameters

Plot (on y axis) Total sample size ☒ with markers ☐ and displaying the values in the plot

is a function of Power ($1 - \beta$ err prob) from 0.6 in steps of 0.01 through to 0.95

Plot 1 graph(s) interpolating points

with Correlation ρH_1 at 0.4

and α err prob at 0.05

Draw plot

A priori power (4)

- Amel ingin melakukan penelitian untuk menyelidiki hubungan antara **jumlah setoran suami tiap bulan** dengan **berat badan**.
- **Setelah** dilakukan analisis **korelasi bivariat**, dan **mendapatkan bukti** yang cukup bahwa **kedua variabel tsb berkorelasi**, Amel memutuskan untuk membuat **model regresi linier**.
- Berapa banyak responden yang harus diambil untuk mendeteksi **ES sebesar $R^2=0.16$** dengan $\alpha = 0.05$ dan $1-\beta = 0.8$?



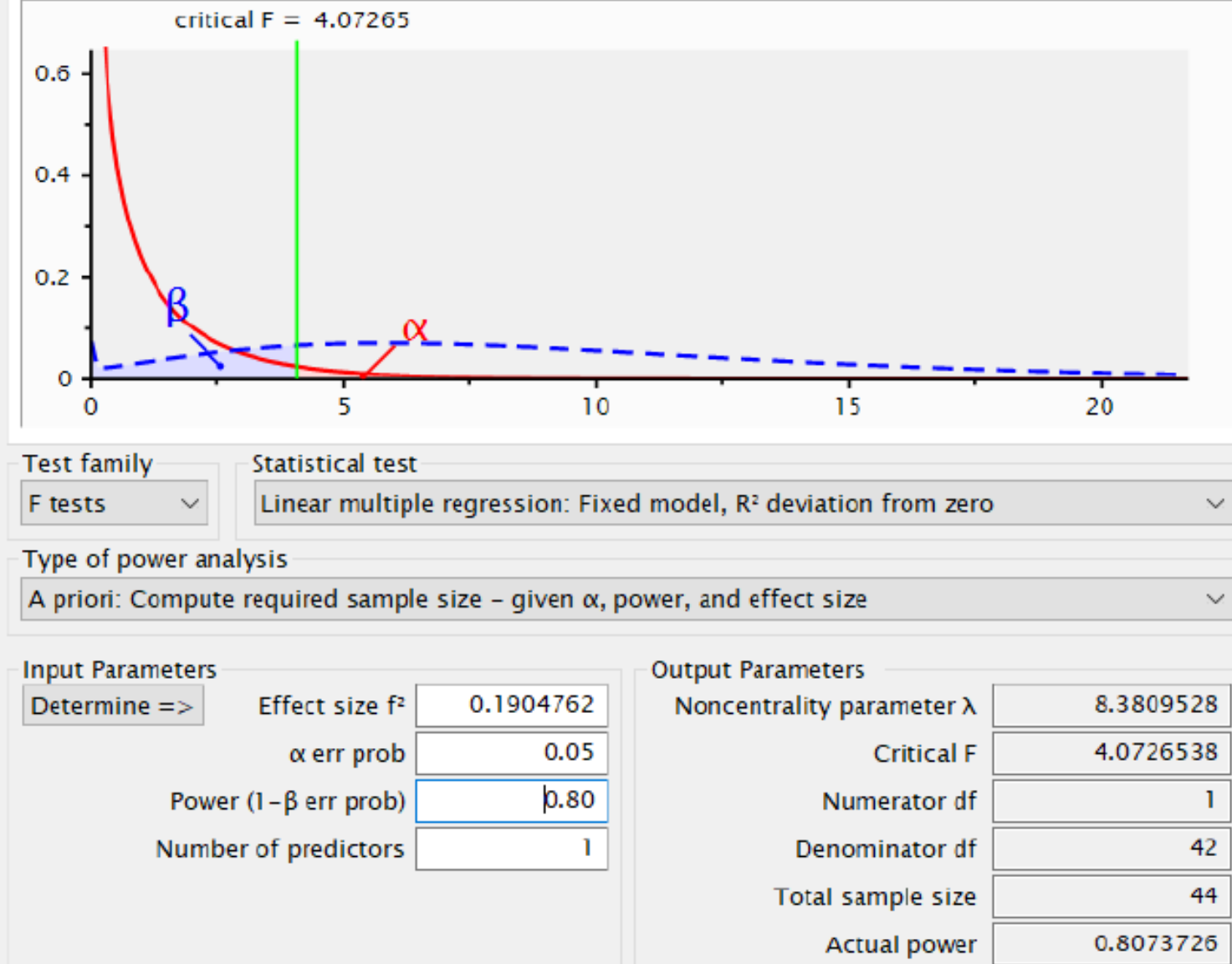
Menggunakan G*Power (regresi linear)

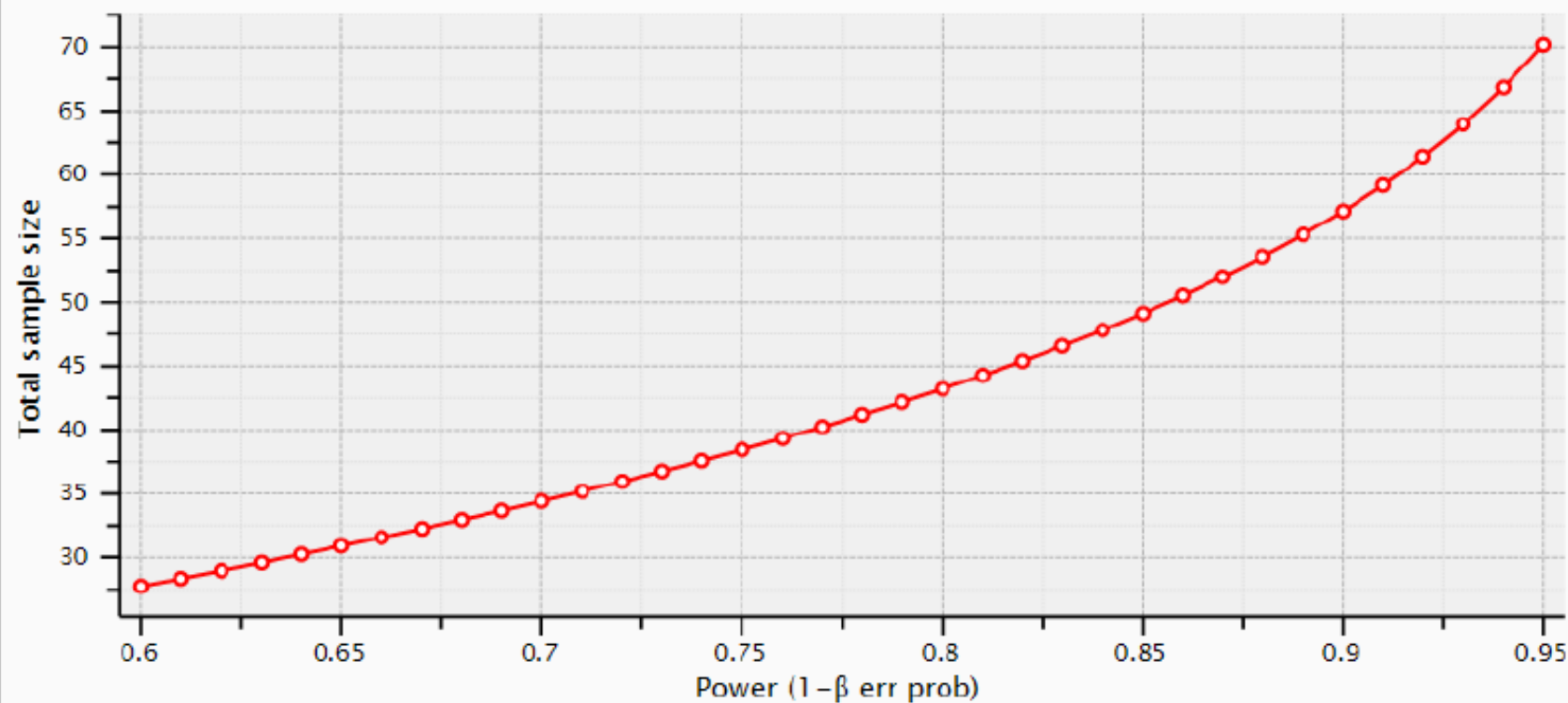
- Dalam opsi test family, pilih ***F-tests***.
- Dalam opsi statistical test, pilih ***Linear multiple regression: Fixed model, R^2 deviation from zero***
- Untuk opsi type of power analysis, pilih ***a priori: computed required sample size – given α , power and effect size***.
- Klik *determine* pada kolom effect size f^2 , kemudian pilih from correlation coefficient (karena cuma ada 1 prediktor). Isikan squared multiple correlation $\rho^2 = 0.16$, lalu klik calculate and transfer to main window.
- Pilih $\alpha=0.05$, power=0.8, dan number of predictors = 1



Interpretasi

- Untuk mendeteksi $R^2=0.16$, maka Amel harus mendapatkan setidaknya 44 responden
- Coba klik *X-Y plot for range of values*



F tests – Linear multiple regression: Fixed model, R^2 deviation from zeroNumber of predictors = 1, α err prob = 0.05, Effect size $f^2 = 0.190476$ 

Plot Parameters

Plot (on y axis) Total sample size ☒ with markers ☐ and displaying the values in the plot

as a function of Power (1 - β err prob) from 0.6 in steps of 0.01 through to 0.95

Plot 1 graph(s) interpolating points

with Effect size f^2 at 0.1904762

and α err prob at 0.05

Draw plot



Sensitivity analysis

- *A priori* power analysis memiliki 1 kelemahan, yaitu mengasumsikan bahwa peneliti punya sumberdaya tak terbatas untuk mengambil n berapapun.
- *Padahal....*
 - Dana terbatas
 - Akses terbatas
 - Belum lagi ada risiko *systematic bias* dan *sampling error*
- Alternatifnya, kita bisa menggunakan *sensitivity analysis* untuk mendeteksi:
Berapa besar ES yang dapat kita deteksi, dengan diketahui n , α dan *power*.
- Caranya, tinggal mengganti kolom *type of power analysis* menjadi *sensitivity* (bukan *a priori* lagi)



Perform an a priori power calculation

Use the existing literature to estimate the size of effect you are looking for and design your study accordingly. If time or financial constraints mean your study is underpowered, make this clear and acknowledge this limitation (or limitations) in the interpretation of your results.

Disclose methods and findings transparently

If the intended analyses produce null findings and you move on to explore your data in other ways, say so. Null findings locked in file drawers bias the literature, whereas exploratory analyses are only useful and valid if you acknowledge the caveats and limitations.

Pre-register your study protocol and analysis plan

Pre-registration clarifies whether analyses are confirmatory or exploratory, encourages well-powered studies and reduces opportunities for non-transparent data mining and selective reporting. Various mechanisms for this exist (for example, the [Open Science Framework](#)).

Make study materials and data available

Making research materials available will improve the quality of studies aimed at replicating and extending research findings. Making raw data available will enhance opportunities for data aggregation and meta-analysis, and allow external checking of analyses and results.

Work collaboratively to increase power and replicate findings

Combining data increases the total sample size (and therefore power) while minimizing the labour and resource impact on any one contributor. Large-scale collaborative consortia in fields such as human genetic epidemiology have transformed the reliability of findings in these fields.



"Thank you. You've been a great audience."