



Modul 2

FPM 226-Methodologi Penelitian Fisioterapi II

Materi 3

Statistik Inferensial

Disusun Oleh

Wahyuddin

UNIVERSITAS ESA UNGGUL

2019

## Pendahuluan

Dalam materi sebelumnya telah dijelaskan tentang statistik deskriptif yang digunakan untuk meringkas dan menggambarkan data. Namun demikian terdapat keterbatasan dalam penggunaannya sehingga diperlukan statistik inferensial. Statistik inferens meliputi pembuatan keputusan berdasarkan kriteria objektif dengan estimasi karakteristik populasi dari data sampel. Keberhasilan proses ini mensyaratkan bahwa kita membuat asumsi tertentu tentang seberapa baik sampel merepresentasikan populasi yang lebih besar. Asumsi ini didasarkan pada dua konsep penting berdasarkan pertimbangan statistik yaitu probability dan sampling error.

## Probability

Probability adalah sebuah konsep yang kompleks tapi penting untuk memahami statistik inferensial. Kita semua memiliki beberapa gagasan tentang probabilitas apa artinya, sebagaimana dibuktikan dengan penggunaan istilah-istilah seperti "mungkin" atau "kesempatan baik." Kita menggunakan probabilitas sebagai alat prediksi misalnya terdapat peluang 50% hujan akan terjadi besok atau tingkat keberhasilan suatu tindakan operasi ini sekitar 75%. Probabilitas adalah kemungkinan bahwa setiap salah satu akan terjadi. Misalnya kemungkinan untuk mendapatkan sisi angka pada pelemparan koin sebesar  $\frac{1}{2}$  atau 50%.

## Aplikasi Probabilitas pada Distribusi Angka

Konsep probabilitas dapat diterapkan pada distribusi angka. Misalnya kita mendapatkan data usia rata-rata pria pada suatu populasi sekitar 69 tahun dengan standard deviation 3. Bagaimana jika kita memilih salah seorang dari populasi tersebut secara acak? Kemungkinan usia yang terpilih berada diantara 66 sampai 72 tahun. Hal ini penting untuk dipahami bahwa probabilitas adalah prediktif yang akan terjadi. Kita menggunakan probabilitas dalam penelitian sebagai pedoman untuk membuat keputusan tentang seberapa baik estimasi data sampel menggambarkan karakteristik populasi. Kita juga menggunakan probabilitas untuk menentukan jika diamati perbedaan efek intervensi atau karena terjadi secara kebetulan.

## Sampling Error

Estimasi karakteristik populasi dari data sampel didasarkan pada asumsi bahwa sampel yang diacak secara valid merepresentasikan populasi. Pilihan secara acak menyiratkan bahwa semua anggota populasi memiliki kesempatan yang sama untuk dipilih, tetapi jelas, ini tidak menjamin representasi proporsional seluruh penduduk.

## Standard Error of the Mean

Karena suatu distribusi sampel digambarkan dengan kurva normal, kita dapat menentukan variabilitas. Standard deviation dari suatu distribusi sampel dinamakan standard error of the mean. Nilai ini dapat mengestimasi standard deviation pada suatu populasi. Terkait dengan jumlah sampel, jika jumlah sampel meningkat, maka akan semakin merepresentasikan populasi. Karena itu, standard deviation dari distribusi

sampel merupakan suatu indikator tingkat kesalahan pengukuran sampel yang merefleksikan seberapa akurasi variasi estimasi rata-rata sampel terhadap populasi.

### Confidence Intervals (CI)

Dalam banyak aplikasi penelitian, data sampel digunakan untuk memperkirakan parameter populasi yang tidak diketahui. Sebagai contoh, kita dapat menggunakan catatan medis untuk menentukan lama tinggal di rumah sakit untuk pasien dengan diagnosis tertentu atau kita bisa mempelajari nilai-nilai yang normatif untuk tes fungsi motorik. Tujuan dari jenis analisis adalah untuk mengestimasi perilaku populasi dan menggunakan informasi ini untuk pengambilan keputusan atau sebagai dasar untuk penelitian lebih lanjut.

CI adalah serangkaian nilai atau keyakinan dengan batas-batas tertentu, yang harus merupakan nilai mean dari populasi. Batas-batas CI berdasarkan mean sampel dan standard error. Semakin lebar interval, semakin tinggi keyakinan bahwa nilai-nilai populasi berada pada batas tersebut. CI ini dinyatakan sebagai persentase probabilitas, misalnya 95% CI.

### Pengujian hipotesis

Estimasi parameter populasi adalah salah satu bagian dari statistik inferensial. Kesimpulan lebih sering digunakan untuk menjawab pertanyaan mengenai perbandingan atau hubungan, seperti apakah salah satu intervensi yang lebih efektif dibanding yang lain? atau apakah ada hubungan antara jangka waktu treatment dengan derajat peningkatan pada suatu pengamatan? Jenis pertanyaan ini biasanya melibatkan perbandingan berarti, proporsi, korelasi atau beberapa statistik lainnya.

Pertimbangkan sebuah studi di mana sebuah hipotesis diajukan bahwa sesi intervensi mobilisasi jaringan lunak akan efektif untuk meningkatkan eksternal rotasi pada pasien dengan problems otot-otot bahu. Kedua kelompok baik kelompok intervensi dan kelompok kontrol diseleksi secara acak. Pengukuran rentang gerak (ROM) dilakukan sebelum dan sesudah intervensi pada setiap subjek. Para peneliti menyimpulkan bahwa terdapat peningkatan eksternal rotasi 16.4 derajat untuk kelompok pengobatan intervensi dan 1 derajat pada kelompok kontrol. Berdasarkan perbedaan antara mean, apakah peneliti harus membuat konklusi bahwa hal tersebut mendukung hipotesis penelitian?

Berdasarkan konsep sampling error, kita berharap untuk melihat beberapa perbedaan antara kelompok bahkan ketika intervensi sama sekali tidak efektif karena perbedaan karakteristik subjek. Oleh karena itu, kita perlu beberapa mekanisme untuk memutuskan jika efek diamati akan mencerminkan kesempatan. Kita melakukan ini melalui proses pengujian hipotesis.

### Statistical Hypotheses

Sebelum menginterpretasikan suatu perbedaan, kita harus mempertimbangkan penjelasan terkait observasi outcome. Penjelasan yang mungkin adalah bahwa perbedaan antara kelompok terjadi secara kebetulan, sebagai akibat dari sampling

error. Ini disebut sebagai null hypothesis null ( $H_0$ ), yang menyatakan bahwa tidak ada perbedaan antara kelompok. Tidak peduli bagaimana hipotesis penelitian dibuat, tujuan peneliti akan selalu menguji null hipotesis. Penjelasan kedua adalah bahwa ada perbedaan antara kelompok, dan treatment tersebut efektif. Ini dinyatakan sebagai hipotesis alternatif ( $H_1$ ).

### Null hypothesis

Secara formal hipotesis statistik merupakan terminologi dalam parameter populasi, meskipun uji statistik yang sebenarnya akan didasarkan pada data sampel. Oleh karena itu, hipotesis null dapat dinyatakan untuk memprediksi bahwa rata-rata populasi penduduk A tidak berbeda dari rata-rata populasi B.

Misalnya pada studi terkait mobilisasi sendi bahu yang menyimpulkan bahwa tidak ada perbedaan antara mereka yang menerima mobilisasi dengan kelompok yang tidak menerima mobilisasi. Karena ada kemungkinan perbedaan karena faktor kebetulan, kita harus menyadari bahwa meskipun tidak ada efek treatment, kedua grup secara probabilitas tidak mempunyai mean yang sama.

### The Alternative Hypothesis

Penjelasan kedua terkait dengan alternative hypothesis. Hal ini menjelaskan tentang temuan bahwa suatu treatment efektif. Karena itu, perbedaan nilai observasi terlalu besar untuk dipertimbangkan sebagai hasil karena faktor kebetulan. Pernyataan dari hipotesis tersebut bahwa ada perbedaan antara mean populasi yang bukan karena faktor kebetulan atau peluang karena faktor tersebut sangat kecil. Ketika kita menolak null hypothesis, kita dapat menerima alternative hypothesis. Ini berarti bahwa alternative hypothesis konsisten dengan temuan kita.

Dalam kebanyakan kasus, peneliti berharap bahwa data akan mendukung alternative hypothesis dan menolak null hypothesis. Ini mungkin dalam situasi bahwa peneliti tidak mengharapkan adanya perbedaan antar kelompok, seperti mencoba untuk menunjukkan bahwa suatu treatment baru pada kelompok eksperimental seefektif dengan treatment standar. Dalam kasus tersebut, hipotesis penelitian akan sama dengan null hypothesis, dan peneliti berharap untuk menolak alternative hypothesis.

### Directional and Nondirectional Hypotheses

Alternative hypotheses sebelumnya menyatakan bahwa akan ada perbedaan antara populasi A dan B. Ini disebut sebagai nondirectional hipotesis, karena mereka tidak menspesifikasi kelompok ini diharapkan akan lebih besar. Alternative hypotheses juga dapat dinyatakan dalam bentuk terarah, menunjukkan arah yang diharapkan dari perbedaan antara mean sampel. Hipotesis ini memperkirakan bahwa rata-rata populasi penduduk A itu lebih besar atau lebih kecil daripada rata-rata populasi B. Sebagai contoh, kita mungkin memprediksi bahwa kelompok eksperimental yang menerima mobilisasi akan menunjukkan peningkatan yang lebih besar daripada kelompok kontrol.

## Errors in Hypothesis Testing

Pengujian hipotesis akan selalu menghasilkan salah satu dari dua keputusan apakah menolak atau menerima null hypothesis. Dengan menolak null hypothesis, peneliti menyimpulkan bahwa tidak mungkin ada faktor kebetulan terkait adanya perbedaan. Ini disebut sebagai significant effect, dimana salah satu kemungkinan bukan karena faktor kebetulan.

Ketika tidak menolak null hypothesis, peneliti menyimpulkan bahwa perbedaan yang terjadi mungkin karena faktor kebetulan dan tidak signifikan. Keputusan untuk menolak atau menerima null hypothesis didasarkan pada hasil prosedur statistik yang objektif; Namun, objektivitas ini tidak menjamin bahwa keputusan yang benar akan dibuat. Karena keputusan tersebut didasarkan hanya pada data sampel, mungkin bahwa benar hubungan antara populasi eksperimental tidak secara akurat tercermin dalam hasil statistik.

Setiap keputusan dapat benar atau salah. Oleh karena itu, kita dapat mengelompokkan empat hasil mungkin keputusan, Dalam analisis statistik apapun kita dapat menarik kesimpulan yang benar, atau kita mungkin melakukan salah satu dari dua jenis kesalahan. Jika kita menolak null hypothesis, dan memutuskan bahwa pengobatan kelompok berbeda, kita mungkin benar atau kita mungkin membuat sejenis saya kesalahan (tapi bukan tipe II). Jika kita menerima null hypothesis dan memutuskan bahwa ada tidak ada perbedaan, kita mungkin benar atau kita mungkin membuat kesalahan tipe II (tapi bukan tipe I).

Type I error: level of significance

Ketika melihat hasil suatu eksperimen, kita tahu bahwa perbedaan pengamatan bisa karena dampak atau akibat treatment yang diberikan atau mungkin karena faktor kebetulan. Pada level berikutnya, bagaimana kita bisa membuat keputusan mengenai null hypothesis, jika kita tidak yakin jika itu benar atau salah?

Karena itu, jika kita ingin menentukan probabilitas terkait type I error, kita harus menetapkan standar untuk menolak null hypothesis. Standar ini disebut sebagai tingkat kemaknaan, dilambangkan sebagai alpha ( $\alpha$ ). Tingkat kemaknaan mewakili kriteria untuk menilai jika terdapat perbedaan dapat dianggap sebagai akibat sampling error atau nyata. Semakin besar perbedaan pengamatan, semakin kecil kemungkinan itu terjadi secara kebetulan.

Kemungkinan bahwa perbedaan yang diamati terjadi secara kebetulan ditentukan oleh uji statistik. Probabilitas ini dilambangkan sebagai p. Sebagai contoh, kita menemukan bahwa suatu analisis membandingkan dua cara menghasilkan nilai  $p = .18$ . Ini berarti bahwa ada 18% kemungkinan terjadi perbedaan antara mean terjadi karena faktor kebetulan. Oleh karena itu, jika kita memutuskan untuk menolak null hypothesis dan menyimpulkan bahwa kelompok yang diuji berbeda antara satu dengan yang lain, kita memiliki peluang kesalahan sebanyak 18%.

Pertanyaan yang sering dihadapi peneliti adalah bagaimana memutuskan apakah probabilitas dapat diterima. Kita tahu ada beberapa kesempatan bahwa

perbedaan merupakan hasil dari sampling error. Tapi seberapa banyak peluang kesalahan yang dapat diterima?

Misalnya laporan cuaca mengatakan bahwa terdapat peluang hujan hari ini sebesar 75%. Apakah kita akan mengambil payung? Jika kita memutuskan untuk tidak mengambil dan kemudian hujan, berarti kita salah dalam mengambil keputusan. Sekarang laporan mengatakan bahwa kemungkinan hujan hari ini sebesar 5%. Apakah kita akan meninggalkan rumah tanpa membawa payung? Jika kita melakukan itu maka kecenderungan kesalahan dalam mengambil keputusan lebih kecil.

Untuk tujuan penelitian, pemilihan level alpha mendefinisikan risiko maksimal yang diterima untuk membuat suatu type I error jika kita menolak null hypothesis. Biasanya, para peneliti menetapkan standar ini pada 5%, yang dianggap sebagai risiko kecil. Ini berarti bahwa kita akan bersedia menerima kemungkinan sampai 5%, tapi tidak lebih. Oleh karena itu, untuk analisis tertentu, jika  $p$  adalah sama dengan atau kurang dari .05, kita akan bersedia untuk menolak null hypothesis. Karena itu perbedaan akan dipertimbangkan sebagai hal yang signifikan. Sebaliknya jika nilai  $p$  lebih besar dari .05, kita tidak akan menolak null hypothesis. Untuk contoh sebelumnya dengan  $p = .18$ , jika kita telah menentukan  $\alpha = .05$ , kita tidak akan menolak null hypothesis.

#### Memilih tingkat signifikansi

Bagaimana seorang peneliti menentukan tingkat signifikansi sebagai kriteria untuk pengujian statistik?. Seorang peneliti dapat memilih kriteria lain tingkat tergantung pada pentingnya mengkritisi. Misalnya, suatu studi terkait penelitian obat untuk mengurangi spastisitas dengan membandingkan kelompok kontrol dan kelompok-eksperimental. Obat ini bermanfaat bagi pasien dengan kondisi upper motor neuron; Namun, obat tersebut memiliki efek samping yang berpotensi serius dan produksinya sangat mahal. Dalam situasi seperti ini kita ingin menjadi sangat yakin bahwa mengamati hasil yang nyata, dan bukan karena kebetulan. Jika kita menolak null hypothesis dan merekomendasikan obat tersebut, kita ingin probabilitas melakukan kesalahan (type I error) sangat kecil.

Peneliti harus menentukan tingkat signifikansi minimal yang diperlukan untuk menolak null hypothesis sebelum pengumpulan data. Hal ini tidak tepat untuk menentukan tingkat kriteria setelah probabilitas statistik telah ditentukan. Karena data dipengaruhi oleh sampling error, penting bahwa penentuan tingkat kemaknaan merupakan suatu proses yang tidak bias.

Kita juga harus berhati-hati untuk menghindari penggunaan besarnya  $p$  sebagai indikasi tingkat validitas hipotesis penelitian. Selain itu tidak disarankan menggunakan istilah seperti "sangat signifikan" atau "lebih signifikan" karena nilai  $p$  sendiri tidak menyiratkan sebagai efek treatment. Makna dari tingkat ini dapat dianggap sebagai titik kontinum yang menandakan jalur antara faktor kebetulan dan realitas. Setelah tingkat kemaknaan dipilih, ini merepresentasikan suatu aturan keputusan. Keputusan ini bersifat dichotomous misalnya ya atau tidak, signifikan atau tidak signifikan. Setelah keputusan dibuat, besarnya  $p$  mencerminkan tingkat kepercayaan yang dapat ditempatkan dalam keputusan itu.

## Type II error: : Statistical power

Penjelasan-penjelasan sebelumnya telah membentuk logika tentang statistik inferensi, al berdasarkan kemungkinan menolak sebuah null hypothesis yang benar, atau type I error. Apa yang terjadi ketika kita menemukan ada perbedaan yang signifikan antara kelompok dan kita tidak menolak null hypothesis? Apakah ini berarti bahwa tidak ada perbedaan nyata? Jika kita tidak menolak null hypothesis ketika itu memang palsu, kita telah melakukan kesalahan tipe II (type II error); dimana kita menemukan tidak ada perbedaan yang signifikan ketika perbedaan benar-benar ada.

Sayangnya ketika hasil tidak signifikan, peneliti sering berasumsi bahwa treatment kelompok eksperimental itu tidak efektif. Probabilitas membuat type II error dilambangkan dengan beta ( $\beta$ ). Penentuan statistical power berdasarkan empat faktor: kriteria signifikansi, varians dalam data ( $s^2$ ), ukuran sampel ( $n$ ) dan faktor yang mencerminkan besarnya perbedaan diamati yang disebut efek size (ES).

## Power Analysis

Kita dapat menganalisis kekuatan (power) untuk dua tujuan. Salah satu adalah untuk memperkirakan ukuran sampel dalam merekrut sampel selama tahap perencanaan studi. Tujuan kedua adalah untuk menentukan probabilitas bahwa kesalahan tipe II sesuai ketika hasil studi menunjukkan tidak signifikan.

## Penentuan ukuran sampel: a priori analysis

Pertanyaan yang sering diajukan oleh peneliti pada tahap perencanaan studi adalah "berapa banyak subjek yang diperlukan?" Jawaban yang mudah dengan mengatakan sebanyak mungkin jumlah pasien akan semakin baik. Namun, hal ini sangat dipengaruhi oleh batasan waktu dan sumber daya dalam proses pengumpulan data. Peneliti menyarankan bahwa ukuran sampel 30 atau 50 adalah logis dan masuk akal.

Sayangnya, perkiraan ini mungkin tidak memadai untuk banyak desain penelitian atau untuk studi dengan ukuran kecil efek. Dengan menspesifikasi tingkat kemaknaan dan power yang diinginkan dalam tahap perencanaan studi, seorang peneliti dapat memperkirakan berapa banyak subjek yang diperlukan untuk mendeteksi perbedaan yang signifikan untuk ukuran efek yang diharapkan. Ini disebut a priori power analysis.

## Pengukuran type II error: post hoc analysis

Ketika sebuah studi telah selesai dan hasil menunjukkan tidak signifikan, peneliti akan ingin menentukan kemungkinan akibat kesalahan tipe II error. Dengan mengetahui ukuran effect size, level signifikansi yang digunakan dan ukuran sampel, peneliti dapat menentukan tingkat power yang dicapai dalam analisis. Jika power rendah, peneliti mungkin menarik hanya kesimpulan tentatif tentang kurangnya efek treatment yang signifikan, dan mempertimbangkan mereplikasi studi dengan sampel lebih besar untuk meningkatkan power tes.

## Konsep pengujian statistik

Prosedur statistik yang digunakan adalah untuk menguji hipotesis melalui perhitungan tes statistik, atau tes rasio. Statistik yang berbeda yang digunakan untuk menguji perbedaan mean, korelasi dan proporsi. Tes statistik yang digunakan untuk menentukan jika efek yang signifikan dicapai, dengan menentukan probabilitas bahwa perbedaan tersebut tidak akan terjadi jika null hypothesis benar.

## Referensi:

Leiyu Shi. Health Services Research Methods 2nd ed, Thomson Delmar Learning United States 2009

Leslie G. Portney, Mary P. Watkins. Foundations of Clinical Research. Applications to Practice 3<sup>rd</sup> ed, F.A Davis Company Philadelphia