

Experimentálna ekonómia

Cvičenie VI - Experiment zo štatistickej perspektívy

Matej Lorko

matej.lorko@euba.sk

Materiály: www.lorko.sk

Referencie:

- Weimann, J., & Brosig-Koch, J. (2019). Methods in experimental economics. Springer International Publishing. Chicago

Experiment zo štatistickej perspektívy

- Ak má byť výskumná otázka zodpovedaná experimentom a s využitím štatistických metód, potom musí byť experiment navrhnutý tak, aby na túto otázku odpovedal čo najlepšie. „Čo najlepšie“ znamená, že formálna metóda analýzy musí byť kompatibilná so štatistickou povahou vygenerovaných dát.
- Inými slovami, výskumná otázka, návrh experimentu, výsledné vygenerované dáta a štatistická analýza musia byť navzájom zladené.
- Zle navrhnutý experiment vedie k slabému vedeckému výsledku – ani tá najsofistikovanejšia metóda analýzy to nemôže zmeniť. Na druhej strane, dobre vybraná analytická metóda môže z dobre navrhnutého experimentu vyvodiť ešte zaujímavejšie zistenia.

Experiment zo štatistickej perspektívy

- Zo štatistického pohľadu je vhodné priebeh experimentálnej štúdie rozdeliť na fázu návrhu a fázu realizácie. Fáza návrhu, ktorá sa realizuje ako prvá, pozostáva z nasledujúcich úloh a typických problémov:
- Operacionalizácia výskumnej otázky: Aké dáta potrebujeme v experimente zbierať, aby bolo možné odpovedať na výskumnú otázku? Dajú sa tieto dáta jasne definovať ako premenné? Ako by sa mali tieto premenné merať? Ktorá z nich je závislá premenná? Ktoré z nich sú nezávislé premenné?
- Štruktúrovanie štatistického dizajnu: Ktoré premenné má experimentátor manipulovať a akým spôsobom (výber treatmentu)? Ktoré premenné môžeme ovládať a ako možno minimalizovať nežiaduce variácie závislej premennej? Ako by sa mala vybrať vzorka subjektov? Koľko subjektov potrebujeme, aby sme dokázali určiť, či hľadaný efekt „existuje“ s dostatočnou pravdepodobnosťou? Na akých úrovniach sa budú sa premenné merať (within vs. between-subject)? Ako často a kedy by sme mali tieto premenné merať? Ktoré premenné sú kvalitatívne a ktoré kvantitatívne?
- Prevod výskumnej otázky do štatistickej hypotézy alebo štatistického modelu: Aký formálny vzťah by mohol existovať medzi pozorovanou variáciou závislej premennej a variáciou nezávislých premenných? Ktoré premenné majú fixné efekty a ktoré náhodné efekty?

Výber vhodnej štatistickej metódy analýzy

- Aký je účel štatistickej analýzy: poskytnúť popisnú reprezentáciu dát a efektov treatmentu? Urobiť štatistický záver týkajúci sa populácie, z ktorej sa vzorka čerpá (inferencia)? Robiť predpoveď na základe odhadovaného modelu? Aké sú hlavné štatistické charakteristiky experimentálneho dizajnu alebo výsledných dát (odpovede z predchádzajúcich otázok)? Aké analytické metódy možno použiť vzhľadom na hlavné štatistické charakteristiky?
- Počítačové spracovanie dát: Chýbajú v dátach niektoré hodnoty? Existujú odľahlé pozorovania (outliers)? Existujú subjekty, ktoré zjavne urobili náhodné rozhodnutia (nepochopili experiment)?
- Vytváranie nových premenných z (kombinácie) už zhromaždených premenných (napr. skupinové priemery); Vytvorenie zoznamu premenných s popisom.
- Analýza dát: Popis dát pomocou kľúčových ukazovateľov; Grafické znázornenie dát; prispôbenie štatistického modelu dátam a odhadom parametrov modelu; Diagnostika modelu; Vyvodzovanie záverov; Predpovede pomocou odhadovaného modelu.
- Závery: Dajú sa účinky treatmentu štatisticky overiť? Dokáže model dobre vysvetliť pozorované dáta? Sú potrebné ďalšie experimentálne treatmenty?

Premenné

- Aby bolo možné experimentálne testovať výskumné otázky, je potrebné vygenerovať rôzne typy premenných. Predpokladajme napríklad, že výskumná hypotéza znie, že „sumy ponúkané v ultimátnej hre sú nižšie, ak prvý hráč hrá proti počítaču namiesto človeka (a vie o tom)“. V tomto prípade je závislou premennou čiastka, ktorú ponúka prvý hráč.
- Experimentátor očakáva, že nezávislá premenná bude mať vplyv na závislú premennú, ale nie naopak. V súlade s našou výskumnou hypotézou očakávame, že binárna premenná „počítačový protivník“ (áno/nie) bude mať vplyv na ponúkané sumy. V kontrolovanom experimente sú hodnoty týchto nezávislých premenných nastavené experimentátorom. Vo vyššie uvedenom príklade experimentátor meria závislú premennú „ponúknuté sumy“ raz pod hodnotou „áno“ a raz pod hodnotou „nie“, aby bolo možné porovnanie oboch podmienok a testovanie výskumnej hypotézy. V tomto prípade sa nezávislá premenná nazýva aj intervenčná premenná, pretože jej hodnoty predstavujú „intervencie“ (treatmenty), za ktorých sa pozoruje závislá premenná.
- Ak má štúdia vyvodiť kauzálny záver o závislých a nezávislých premenných (a to je hlavným účelom kontrolovaných experimentov), je potrebné zvážiť aj iné premenné, ktoré môžu mať na výsledkok vplyv. Musíme byť schopní vylúčiť, že rozdiel medzi treatmentami, ktorý nameráme, bol spôsobený inými vplyvmi, ako iba výlučne nezávislou (treatment) premennou, ktorú v experimente manipulujeme. Rôzne rušivé premenné môžu rozmazať kauzalitu medzi závislými a nezávislými premennými, ak majú „skrytý“ vplyv na závislú premennú, ktorá nie je explicitne súčasťou experimentu.
- Bohužiaľ, existujú aj rušivé premenné, ktoré nemožno kontrolovať. Sú to predovšetkým individuálne charakteristiky subjektov, napr. IQ. Samozrejme, nie všetky možné nekontrolovateľné premenné sú pre experiment relevantné, pretože mnohé nemajú žiadnu súvislosť so závislou premennou. Napriek tomu by sme mali dôkladne zvážiť, čo s vysokou pravdepodobnosťou ovplyvní našu závislú premennú a zároveň sa môže líšiť od subjektu k subjektu, pričom zostáva mimo našej kontroly.

Kontrola, randomizácia a veľkosť vzorky

- Bez ohľadu na to, či je nekontrolovaná rušivá premenná merateľná alebo nie, jej vplyv na našu závislú premennú by sa mal z experimentu čo najviac odstrániť; inak už nie je možné urobiť jednoznačný kauzálny záver o vplyve intervenčnej premennej. 100% kontrola takýchto premenných je sotva možná, keďže mnohé z nich sú nielen nemerateľné, ale dokonca neznáme a ich vplyv je teda „skrytý“.
- Napriek tomu existuje jednoduchý štatistický trik, ktorý dokáže zmierniť ich dopad. Základnou myšlienkou je vytvoriť skupiny subjektov, v ktorých sú možné rušivé premenné rozdelené čo najrovnomernejšie. Robí sa to náhodným zaradením každého subjektu do jednej zo skupín (randomizácia). V tomto procese by sa malo zabezpečiť, aby skupiny pozostávali z dostatočne veľkého počtu nezávislých subjektov.
- Celkovo vzaté, v laboratórnom experimente je ústrednou premennou závislá premenná. Zmeny v tejto premennej sú spôsobené vplyvom vysvetľujúcich premenných a rôznych rušivých premenných. Ak sa má pozorovaná zmena v závislej premennej pripísať zmene vo vysvetľujúcej premennej vyvolanej experimentátorom, treba zvážiť tri najdôležitejšie pojmy:
 - Kontrola (všetky nežiaduce vplyvy, ktoré možno udržiavať konštantné, by sa mali udržiavať konštantné);
 - Randomizácia (vytváranie skupín, ktoré sú v priemere homogénne - ponechaním na náhodu, ktorý subjekt je zaradený do ktorej skupiny);
 - Veľkosť vzorky (alebo replikácia) - zabezpečiť dostatočný počet nezávislých pozorovaní pre každý treatment, t.j. dostatočne veľké skupiny subjektov, ktoré systematicky nevykazujú rovnaké správanie.

Náhodné premenné a ich distribúcia

- Pri štatistickom modelovaní vzťahu medzi premennými sa závislá premenná interpretuje ako náhodná premenná. Hodnoty náhodnej premennej ktoré sú najpravdepodobnejšie a tie, ktoré sú menej pravdepodobné, určuje jej rozdelenie. Funkcia hustoty diskkrétnej náhodnej premennej udáva pravdepodobnosť, s akou nastane určitá hodnota. Napríklad výsledky hodu kockou sú rovnomerne rozdelené, každý s príslušnou pravdepodobnosťou $1/6$.
- V prípade spojitej náhodnej premennej, ako je dĺžka času rozhodnutia, nie je možné určiť pravdepodobnosť individuálnej hodnoty. Ak existuje nekonečný počet hodnôt, pravdepodobnosť jednej hodnoty musí byť nekonečne blízka nule. Z tohto dôvodu je pri spojitých premenných možné indikovať špecifické pravdepodobnosti len pre rozsahy hodnôt, pričom celková plocha pod funkciou hustoty je vždy 1. Kumulatívna (spojitá) distribučná funkcia je, matematicky povedané, integrálom spojitej funkcie hustoty. Hodnota funkcie v bode x teda udáva pravdepodobnosť, s ktorou náhodná premenná nadobudne hodnotu menšiu alebo rovnú x .
- Väčšina štatistických rozdelení má určité parametre, ktoré určujú tvar funkcie hustoty. Tri najdôležitejšie parametre sú očakávaná hodnota, rozptyl a stupne vol'nosti. Očakávaná hodnota je priemerom všetkých hodnôt, ak náhodnú vzorku (teoreticky) vytiahneme nekonečne veľa krát. Napríklad, keďže existuje rovnaká pravdepodobnosť hodu každého čísla na (normálnej) kocke, očakávaná hodnota je $1/6 \cdot (1 + 2 + 3 + 4 + 5 + 6) = 3,5$. Rozptyl je stredná kvadratická odchýlka všetkých realizácií očakávanej hodnoty a teda predstavuje informáciu o "rozhodení" náhodnej premennej. Čím je rozptyl väčší, tým je funkcia hustoty širšia a viac plochá.
- Matkou všetkých distribúcií je normálne rozdelenie. Jeho parametrami sú očakávaná hodnota μ a rozptyl σ^2 . Hustota pravdepodobnosti je zvonovitá a symetrická okolo μ , kde má najvyššiu funkčnú hodnotu hustoty. Iné dôležité rozdelenia nie sú parametrizované priamo pomocou očakávanej hodnoty a rozptylu, ale aj nepriamo pomocou takzvaných stupňov vol'nosti, ktoré ovplyvňujú očakávanú hodnotu a/alebo rozptyl. Napríklad (Studentovo) t-rozdelenie má takéto stupne vol'nosti, pričom tvar funkcie hustoty sa stále viac a viac približuje funkcii hustoty štandardného normálneho rozdelenia so zvyšujúcimi sa stupňami vol'nosti.

Tvorba štatistického designu

- Výber určitého počtu subjektov z celkovej populácie sa v štatistike označuje ako vzorkovanie. Je potrebné sa zamyslieť nad veľkosťou vzorky, t. j. otázkou „Koľko subjektov vyberiem zo špecifikovanej populácie?“ Žiaľ, v experimentálnej praxi sa na túto otázku často odpovedá len na základe rozpočtu, podľa hesla: „Jednoducho berieme toľko subjektov, koľko dokážeme zaplatiť, bez ohľadu na to, či je toto číslo dosť veľké alebo malé“. Napríklad v neurovedách sú laboratórne časy extrémne drahé, takže veľkosti vzoriek sú (často musia byť) v jednocifernom rozsahu.
- Takéto malé vzorky sú však problematické najmä z pohľadu inferenčnej štatistiky. Pravdepodobnosť, že test štatistickej hypotézy správne identifikuje skutočný efekt ako prítomný (toto sa nazýva sila testu), drasticky klesá s menšími vzorkami. Inými slovami, aj keď v skutočnosti existuje v populácii relatívne silný a vedecky relevantný efekt, bude prinajlepšom rozpoznatelný ako „náhodný artefakt“ a nie ako štatisticky významný efekt. Na druhej strane existuje aj „príliš veľká“ vzorka, pretože veľké vzorky môžu spôsobiť, že testy štatistických hypotéz budú príliš citlivé. To znamená, že aj tie najmenšie, možno vedecky nevýznamné účinky sa stávajú štatisticky významnými.
- Štatistická významnosť by sa preto nemala zamieňať s vedeckou významnosťou. V závislosti od veľkosti vzorky môžu byť obe úplne odlišné. Je to preto, že štatistická významnosť je silne ovplyvnená veľkosťou vzorky, zatiaľ čo skutočný účinok, ktorý sa má zistiť v populácii, nie je.
- Ak je jasné, že (dostatočne veľká) náhodná vzorka nie je dostupná ale stále je potrebná reprezentatívna vzorka, potom je vhodnou možnosťou stratifikovaný výber. Populácia sa najskôr rozdelí na subpopulácie (vrstvy), pričom subjekty v rámci každej subpopulácie majú aspoň jednu spoločnú vlastnosť, ktorá ich odlišuje od subjektov ostatných subpopulácií. Z každej vrstvy sa potom vyberie náhodná vzorka. Každá z týchto vzoriek musí tvoriť rovnaký podiel z celkového počtu všetkých vzoriek ako každá vrstva v celkovej populácii.

Tvorba štatistického designu

- Ako sa líšia experimentálne treatmenty?
- Experimentálne treatmenty je možné klasifikovať podľa počtu faktorových premenných a ich typu, ako aj podľa počtu možných hodnôt. V najjednoduchšom faktoriálnom dizajne sa mení iba jedna premenná. Ak ide o binárnu premennú len s dvoma hodnotami alebo úrovňami, hovoríme o 1×2 faktoriálnom dizajne. 1×2 faktoriálny design sa dá vyhodnotiť obzvlášť jednoducho, pretože zvyčajne stačí porovnať iba stredné hodnoty závislých premenných. V ideálnom prípade je tento rozdiel spôsobený samotným treatmentom, a preto sa nazýva (jednoduchý) efekt treatmentu. Kvantitatívny rozdiel medzi týmito dvoma hodnotami sa nazýva veľkosť efektu treatmentu. Na druhej strane, ak má faktorová premenná viac ako dve úrovne, stredné hodnoty závislej premennej môžu byť porovnávané párovo pre každé dve úrovne alebo súčasne pre všetky úrovne.
- Návrh s dvoma faktormi je podstatne zložitejší ako návrh s jedným faktorom. Napríklad, ak chceme experimentálne preskúmať, ako ovplyvňujú správanie faktory „hra proti počítaču“ (nie/áno alebo 0/1) a „experimentátor vie, kto som“ (nie/áno alebo 0/1), potom ide o 2×2 faktoriálny dizajn.
- V dizajne opakovaných meraní každý subjekt absolvuje niekoľko meraní, buď v jednom a tom istom treatmente v rôznych časoch (pozdlžny dizajn) alebo v rôznych treatmentoch, samozrejme aj v rôznych časoch (cross-over design). Sekvencia treatmentov, ktorými subjekt prechádza, je opäť náhodná.
- Hlavným štatistickým problémom pri viacnásobných meraniach je vzájomná závislosť pozorovaní. Vo faktoriálnom dizajne 1×2 s viacerými meraniami dostaneme kontrolnú skupinu (meraní na úrovni 1) a intervenčnú skupinu (meraní na úrovni 2), ktoré spolu súvisia. Účinok meraní pomocou závislej premennej teda už nemožno jednoznačne pripísať treatmentu, pretože by to mohol byť aj časový alebo sekvenčný účinok (napr. učenie, oboznámenie sa, únava). Často nám v tomto prípade pomáha vyvažovanie (balancovanie), t.j. vytvorí sa dve homogénne skupiny a jedna skupina sa meria v poradí 1, potom 2 a druhá v poradí 2, potom 1.
- Výhodou opakovaných meraní sú nižšie náklady v dôsledku menšieho počtu subjektov, nižšie rozpätie chýb, čo vedie k vyššej štatistickej sile ako pri porovnateľných between-subject dizajnoch, a možnosť merania efektov v čase (dynamika). Nevýhody takéhoto návrhu spočívajú v tom, že zahŕňa podstatne zložitejšie metódy analýzy v dôsledku závislosti pozorovaní a slabších príčinných súvislostí v dôsledku sekvencie, času a efektov prenosu.

Štatistické testy

- V každodennom živote sa až príliš často stretávame s tým, že vyvodzujeme úplne nevedecké a neplatné závery, ako napríklad „Môj priateľ bol raz okradnutý v meste A, takže je to zločinecké mesto“ alebo „Bezpečnostný pás nie je potrebný. Koniec koncov, nikdy som nemal nehodu.“
- Aj bez formálnej analýzy si môžeme byť celkom istí, že tieto závery príliš zovšeobecňujú, keďže sú založené len na jednom pozorovaní. Ako však možno urobiť konkrétne tvrdenia o kvalite záveru? Ako si môže byť experimentátor istý, že pozorovaný efekt nie je úplne náhodný?
- V takýchto situáciách nám pomáhajú nástroje z inferenčnej štatistiky. Dôraz je kladený na testovanie štatistických hypotéz. Takýmito testami možno určiť, nakoľko je všeobecné tvrdenie o charakteristikách populácie v súlade s pozorovanými laboratórnymi údajmi alebo so vzorkou.

Formulácia testovateľných hypotéz

- Východiskovým bodom štatistického testovania je výskumná hypotéza. Zvyčajne postuluje obsah výskumnej otázky. V podstate to, čo považujeme za pravdivé, je formulované ako nulová hypotéza H_0 a jej opak alebo doplnok ako alternatívna hypotéza H_1 .
- Tento princíp štatistického testovania je porovnateľný s prezumpciou nevinu v súdnom spore. Počiatočná alebo nulová hypotéza je: "Obžalovaný je nevinný." Namiesto priameho preukázania, že obžalovaný je vinný, prokurátor predkladá viac či menej silné dôkazy, ktoré nie sú v súlade s nevinou obžalovaného. Ak sú tieto dôkazy dostatočne silné, prezumpcia nevinu už neplatí a obžalovaný je uznaný vinným. Ak však nie je možné predložiť dostatočne silné dôkazy proti prezumpcii nevinu, obžalovaný nebude uznaný vinným, pretože jeho predtým predpokladanú nevinu nebolo možné spochybniť bez akýchkoľvek pochybností.
- Predpokladá sa, že nulová hypotéza je pravdivá, až pokiaľ zhromaždené dáta nebudú dostatočne silné proti nej, a vtedy musí byť zamietnutá. Akonáhle sa tak stane, alternatívna hypotéza je nepriamo prijatá. Ak však údaje nedokážu vyvrátiť nulovú hypotézu, treba stále predpokladať, že je pravdivá a výskumná hypotéza nie je akceptovaná. Keďže v teste hypotéz sa testuje iba nulová hypotéza a hľadajú sa dôkazy proti nej, nulová hypotéza môže byť iba zamietnutá alebo nezamietnutá, ale nikdy nie akceptovaná.

Inferenčná štatistika

- Ak je výskumná otázka formulovaná vo forme štatistickej hypotézy, môžeme ju testovať. Vždy však existuje určitá pravdepodobnosť chýb.
- Vždy treba mať na pamäti, že žiadny štatistický test nedokáže určiť, či je hypotéza skutočne pravdivá alebo nepravdivá. Aj keď je testovacia štatistika vzorky v kritickej oblasti a dospejeme k záveru, že nulovú hypotézu treba zamietnuť, stále môže byť pravdivá.
- Čím väčšiu kritickú oblasť alebo hladinu významnosti zvolíme, tým pravdepodobnejšia je táto takzvaná chyba I. typu. Avšak, ak je nulová hypotéza v skutočnosti nepravdivá, môžeme urobiť chybu v prípade, že ju neodmietneme (chyba typu II).

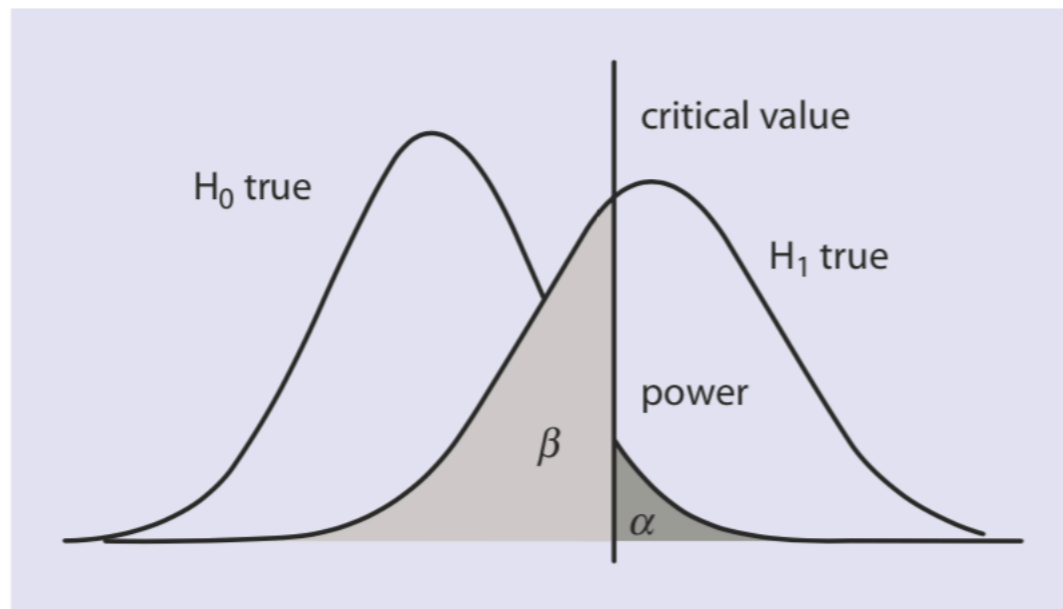
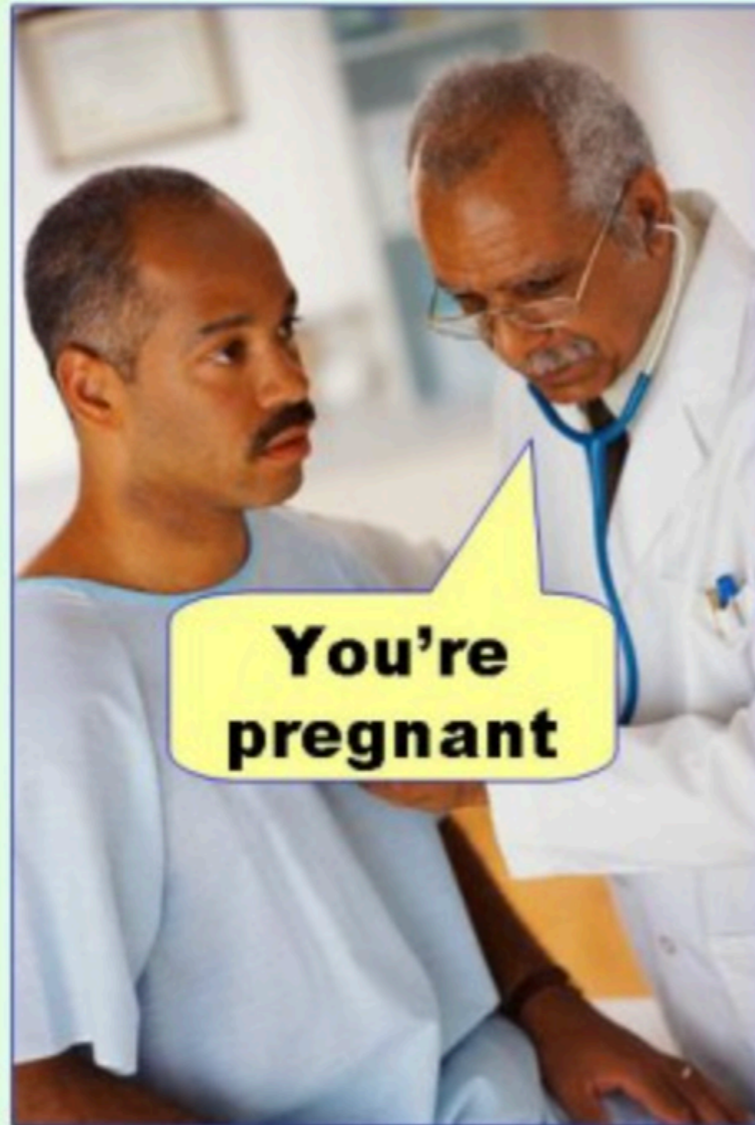


Table 4.1 Summary of error probabilities

	Truth	
	H ₀ true	H ₀ not true
Rejection H ₀	Type I error (Prob. α)	correct
Non-rejection H ₀	correct (Prob. $1 - \alpha$)	Type II error (Prob. $1 - \beta$)

Type I error
(false positive)



Type II error
(false negative)

