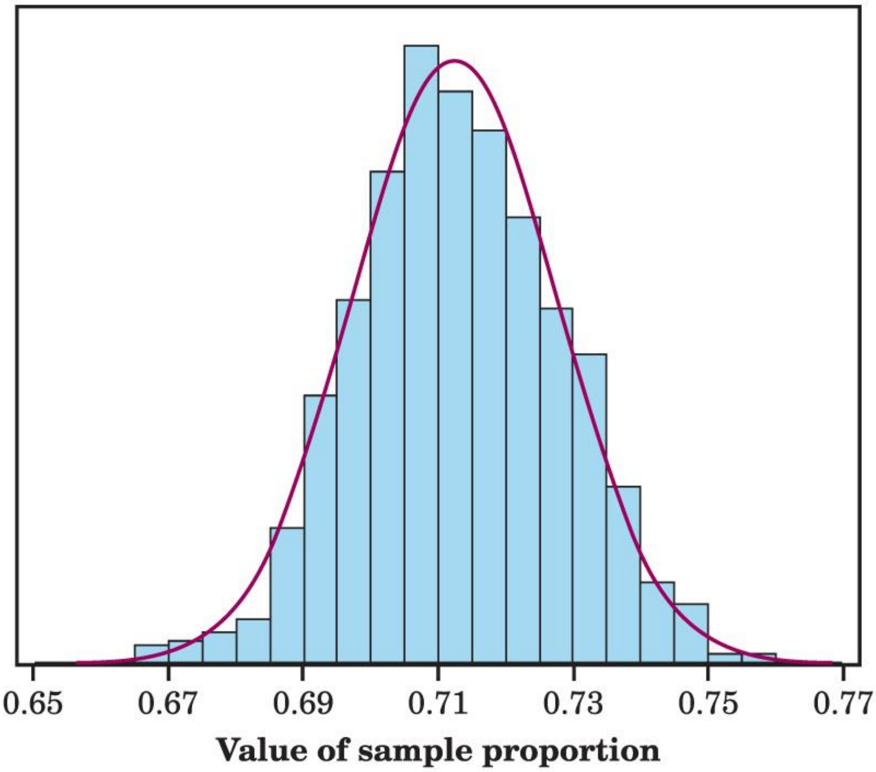
Vjerovatnoće i Uzorkovanje

1. Uvod u Distribuciju Uzorkovanja

Distribucija uzorkovanja nam pokazuje koje vrijednosti statistika može uzeti u ponovljenim uzorcima iz iste populacije i koliko često se te vrijednosti javljaju. Ova distribucija se često opisuje distribucijom frekvencija vjerovatnoće (PDF), kao što je normalna distribucija.

2. Primjer Distribucije Uzorkovanja

Uzmimo jednostavan slučajni uzorak od 1004 odrasle osobe i pitamo ih da li smatraju da su vakcinacije u djetinjstvu veoma važne. Proporcija onih koji kažu "Da" se označava kao p^. Ako ponovimo ovo istraživanje 1000 puta, dobit ćemo 1000 proporcija uzorka *p*^. Histogram ovih proporcija pokazuje distribuciju uzorka kada je stvarna proporcija u populaciji 71%.



Moore/Notz, *Statistics: Concepts and Controversies*, 10e,
© 2020 W. H. Freeman and Company

Horizontalna osa na grafu predstavlja vrijednosti proporcije uzorka, koje se kreću u rasponu od 0,65 do 0,77, sa intervalima od 0,02. Zvonasta kriva, koja je karakteristična za normalnu distribuciju, počinje od vrijednosti 0,65, raste do vrha na 0,71, i postepeno opada do vrijednosti 0,77. Ova kriva ilustrira kako se proporcije uzorka raspoređuju kada se ponavljaju slučajni uzorci iz iste populacije.

Ponavljanje slučajnih uzoraka i posmatranje njihovih proporcija pokazuje jedan od ključnih principa statistike: redovan obrazac koji se pojavljuje u mnogim uzorcima. Ovaj obrazac se može dobro aproksimirati normalnom krivom, koja daje vjerovatnoće za različite proporcije uzoraka.

3. Normalna Distribucija i Pravilo 68-95-99.7

Normalna kriva je dobra aproksimacija histogramu distribucije uzorka. Ova kriva ima srednju vrijednost 0,710 i standardnu devijaciju oko 0,014. Prema pravilu 68-95-99.7, 95% svih uzoraka će dati $\hat{p}$ unutar 2 standardne devijacije od srednje vrijednosti, tj. između 0,682 i 0,738.

4. Krive Gustine i Vjerovatnoće

Krive gustine imaju ukupnu površinu od 1 ispod sebe, što odgovara ukupnoj vjerovatnoći od 1. Kada normalna kriva dodjeljuje vjerovatnoće, možemo koristiti pravilo 68-95-99.7 ili tabele percentila normalnih distribucija za izračunavanje vjerovatnoća.

5. Primjer: Odobravanje Kockanja

Istraživanje javnog mnjenja pita slučajni uzorak od 501 tinejdžera da li odobravaju legalno kockanje. Pretpostavimo da bi tačno 50% svih tinejdžera reklo "Odobravam". Proporcija uzorka koji kaže "Odobravam" će se razlikovati u ponovljenim uzorcima prema normalnoj distribuciji sa srednjom vrijednošću 0,5 i standardnom devijacijom oko 0,022.

6. Sažetak

Model vjerovatnoće opisuje slučajni fenomen navodeći moguće ishode i vjerovatnoće koje im se dodjeljuju. Postoje dva načina za modeliranje vjerovatnoće: dodjeljivanje vjerovatnoće svakom pojedinačnom ishodu i dodjeljivanje vjerovatnoća kao površina ispod krive, kao što je normalna kriva.

Očekivana Vrijednost

Matematičko Očekivanje

Matematičko očekivanje je koncept koji se koristi za predviđanje prosječnog ishoda slučajnog fenomena s numeričkim ishodima. Na primjer, u lutrijskoj igri "Straight" iz DC-3 igre DC lutrije, plaćate 1,00 dolar i birate trocifreni broj. Ako vaš broj bude izabran, dobijate 500 dolara. Pošto postoji 1000 trocifrenih brojeva, vjerovatnoća da dobijete je 1 u 1000.

Model Vjerovatnoće za Dobitak

- **Ishod:** $0 ili $500
- **Vjerovatnoća:** 0.999 za $0 i 0.001 za $500

Prosječni Dobici

Uobičajeni prosjek dva moguća ishoda ($0 i $500) bio bi $250, ali to nema smisla jer je $500 mnogo manje vjerovatno od $0. Dugoročno gledano, dobijate $500 jednom na svakih 1000 opklada i $0 na preostalih 999 od 1000 opklada. Vaši dugoročni prosječni dobici od tiketa su: [500 * 0.001 + 0 * 0.999 = 0.50]

Očekivana Vrijednost

Očekivana vrijednost slučajnog fenomena s numeričkim ishodima se nalazi množenjem svakog ishoda njegovom vjerovatnoćom, a zatim sabiranjem svih proizvoda. U simbolima, ako su mogući ishodi $a_1, a_2, \ldots, a_k$ i njihove vjerovatnoće $p_1, p_2, \ldots, p_k$, očekivana vrijednost je:

$$E(X) = a_1p_1 + a_2p_2 + \ldots + a_kp_k$$

Očekivana vrijednost je prosjek mogućih ishoda, ali nije običan prosjek gdje svi ishodi imaju istu težinu. Umjesto toga, svaki ishod je ponderisan njegovom vjerovatnoćom, tako da ishodi koji se javljaju češće dobijaju veće pondere.

Primjer: Broj Motornih Vozila u Američkim Domaćinstvima

Američka uprava za energetske informacije (2017) daje distribuciju vozila po domaćinstvu:

- **Broj vozila:** 0, 1, 2, 3, 4, 5, 6
- **Udio domaćinstava:** 0.09, 0.34, 0.33, 0.15, 0.06, 0.02, 0.01

Očekivana vrijednost broja vozila po domaćinstvu je:

E(X) =1.85

Ovaj rezultat pokazuje da prosječno američko domaćinstvo ima 1.85 motornih vozila.

Zakon Velikih Brojeva

Osnovni Princip

Prema zakonu velikih brojeva, ako se slučajna pojava s numeričkim ishodima ponavlja mnogo puta nezavisno, srednja vrijednost stvarno posmatranih ishoda približava se očekivanoj vrijednosti. Očekivana vrijednost je prosjek mogućih ishoda, ali ponderisan vjerovatnoćama tih ishoda.

Dugoročni Prosjek

Očekivana vrijednost predstavlja dugoročni prosjek koji će se zapravo vidjeti ako ponavljamo opkladu mnogo puta ili nasumično biramo mnogo domaćinstava. Matematičari su dokazali da je očekivana vrijednost izračunata iz modela vjerovatnoće zaista dugoročni prosjek. Ova poznata činjenica se zove zakon velikih brojeva.

Primjena u Kockanju

Zakon velikih brojeva objašnjava zašto je kockanje posao za kazino, dok je za pojedince rekreacija ili zavisnost. Kazino ne kocka, već unaprijed izračunava očekivanu vrijednost i zna šta će biti dugoročno. Prosečan dobitak velikog broja kupaca biće prilično blizu očekivanoj vrijednosti. Kazina se fokusiraju na jeftinu zabavu i mala ugađanja kupcima kako bi održali priliv kupaca. Ako se stavi dovoljno opklada, zakon velikih brojeva garantuje kazinu profit.

Primjena u Osiguranju

Kompanije za životno osiguranje rade slično kao kazina. One se klade da ljudi koji kupuju osiguranje neće umrijeti. Neki, naravno, umiru, ali osiguravajuće društvo zna vjerovatnoće i oslanja se na zakon velikih brojeva da bi predvidjelo prosječan iznos koji će morati da isplati.