ACH3584 - Estatística II

Aula 5: Distribuição da proporção amostral e Tamanho de amostra

Alexandre Ribeiro Leichsenring
alexandre.leichsenring@usp.br

Organização

1. Distribuição Amostral da Proporção

2. Determinação do Tamanho de uma Amostra

Organização

1. Distribuição Amostral da Proporção

2. Determinação do Tamanho de uma Amostra

Distribuição Amostral da Proporção

Vamos recordar o que já sabemos sobre a distribuição amostral da **média**.

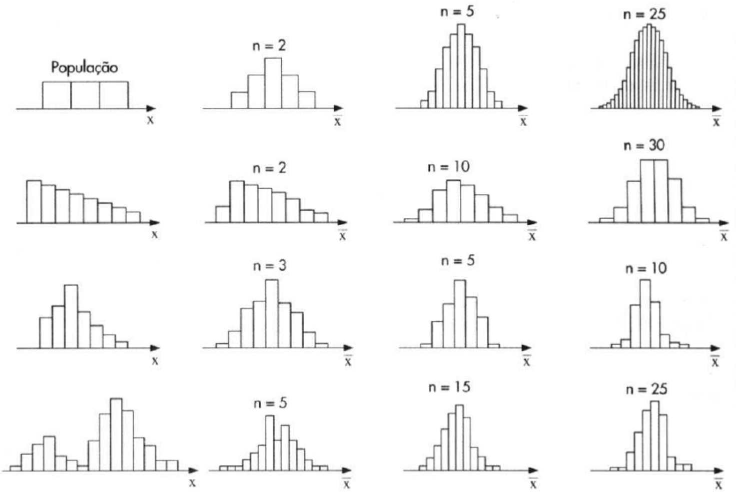
Se X é uma v.a. com média μ e variância σ^2, e seja $(X_1, X_2, \ldots, X_n)$ uma amostra aleatória simples de X, então:

$$\begin{aligned} \mathbf{E}(\bar{X}) &= \mu \\ Var(\bar{X}) &= \frac{\sigma^2}{n} \end{aligned}$$

Além disso, o TLC afirma que:

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Distribuição de $\bar{X}$ para amostras extraídas de algumas populações



Distribuição amostral da proporção

Vamos considerar uma população em que a proporção de elementos portadores de uma característica é p. Vamos definir as v.a.'s $X_1, X_2, \ldots$ tais que:

$$X_i = \begin{cases} 1, & \text{se o } i\text{-ésimo indivíduo é portador da característica} \\ 0, & \text{se o } i\text{-ésimo indivíduo } \textbf{não} \text{ é portador da característica} \end{cases}$$

Logo, $X_i \sim Bernoulli(p)$, com:

- $\mathbf{E}(X) = p$
- $Var(X) = p(1-p)$

Retirada uma amostra de tamanho n da população e indicando por Y_n o total de indivíduos portadores da característica na amostra, já vimos que:

$$Y_n = X_1 + X_2 + \ldots + X_n$$

$$Y_n \sim \text{Binomial}(n, p)$$

Definindo $\hat{p}$ a proporção de indivíduos portadores da característica na amostra, temos que:

$$\hat{p} = \frac{Y_n}{n} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

Ou seja, $\hat{p}$ é uma média amostral, e logo vale o TLC:

$$\hat{p} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \text{aproximadamente.}$$

No caso da Bernoulli:

- $\mu = \mathbf{E}(X) = p$ (proporção de indivíduos na população com a característica);
- $\sigma^2 = Var(X) = p(1-p)$

Logo,

$$\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right), \quad \text{aproximadamente.}$$

Exemplo

Um sistema de produção opera de tal maneira que 10% das peças produzidas são defeituosas. Suponha que os itens sejam vendidos em caixas com 100 unidades. Qual a probabilidade de que uma caixa:

a) tenha mais do que 10% de defeituosas?

b) tenha menos do que 15% de defeituosas?

Solução

a) Seja $\hat{p}$ a proporção de peças defeituosas na amostra. Queremos $\mathbf{P}(\hat{p} > 0,1)$.

$$\hat{p} \sim N\left(p; \frac{p(1-p)}{n}\right), \quad \text{aproximadamente.}$$

Do enunciado: $\begin{cases} p = 0,1 \\ n = 100 \end{cases}$

Logo,

$$\mu_{\hat{p}} = p = 0,1$$

$$\sigma^2_{\hat{p}} = \frac{p(1-p)}{n} = \frac{0,1(1-0,1)}{100} = 0,0009$$

$$\hat{p} \sim N(0,1; 0,0009)$$

Exemplo

Um sistema de produção opera de tal maneira que 10% das peças produzidas são defeituosas. Suponha que os itens sejam vendidos em caixas com 100 unidades. Qual a probabilidade de que uma caixa:

a) tenha mais do que 10% de defeituosas?

b) tenha menos do que 15% de defeituosas?

Solução

a) Seja $\hat{p}$ a proporção de peças defeituosas na amostra. Queremos $\mathbf{P}(\hat{p} > 0,1)$.

$$\hat{p} \sim N\left(p; \frac{p(1-p)}{n}\right), \quad \text{aproximadamente.}$$

Do enunciado: $\begin{cases} p = 0,1 \\ n = 100 \end{cases}$

Logo,

$$\begin{aligned} \mu_{\hat{p}} &= p = 0,1 \\ \sigma^2_{\hat{p}} &= \frac{p(1-p)}{n} = \frac{0,1(1-0,1)}{100} = 0,0009 \end{aligned}$$

$$\hat{p} \sim N(0,1; 0,0009)$$

O exercício agora se resume a fazer cálculos com a distribuição normal de média 0,1 e variância 0,0009.

Seja X uma v.a. com distribuição Normal com

- média $\mu_X = 0,1$; e
- variância $\sigma_X^2 = 0,0009 \Rightarrow \sigma_X = \sqrt{0,0009} = 0,03$

Então,

$$
\begin{aligned}
\mathbf{P}(\hat{p} > 0,1) &= \mathbf{P}(X > 0,1) \\
&= \mathbf{P}\left(\frac{X - \mu_X}{\sigma_X} > \frac{0,1 - \mu_X}{\sigma_X}\right) \\
&= \mathbf{P}\left(Z > \frac{0,1 - \mu_X}{\sigma_X}\right) \\
&= \mathbf{P}\left(Z > \frac{0,1 - 0,1}{0,03}\right) \\
&= \mathbf{P}(Z > 0) \\
&= 0,5
\end{aligned}
$$

b) De maneira análoga:

$$
\begin{aligned}
\mathbf{P}\left(\hat{p} < 0,15\right) &= \mathbf{P}\left(\frac{X - \mu_X}{\sigma_X} < \frac{0,15 - \mu_X}{\sigma_X}\right) \\
&= \mathbf{P}\left(Z < \frac{0,15 - \mu_X}{\sigma_X}\right) \\
&= \mathbf{P}\left(Z < \frac{0,15 - 0,1}{0,03}\right) \\
&= \mathbf{P}\left(Z < 1,67\right) = 0,9525
\end{aligned}
$$

Organização

1 Distribuição Amostral da Proporção

2 Determinação do Tamanho de uma Amostra

Determinação do Tamanho de uma Amostra

- Em alguns casos, queremos justamente determinar o tamanho de uma amostra a ser extraída de uma população
- Queremos controlar o erro amostral (de natureza aleatória) envolvido, estipulando uma margem de erro tolerável

Suponha que queremos estimar a média populacional μ e para tanto usaremos a média amostral, $\bar{X}$, baseada numa amostra de tamanho n.

Suponha que queremos estabelecer a *priori* uma determinada confiança (γ) em que a estimativa não ultrapasse uma determinada margem de erro ϵ.

Em outras palavras, queremos determinar o valor de n de tal modo que

$$\mathbf{P}\left(|\bar{X} - \mu| \leq \epsilon\right) \geq \gamma$$

Antes de prosseguir, vamos apresentar outra formulação do TLC:

TLC

- Se $X_1, X_2, \ldots, X_n$ é uma amostra aleatória simples de uma população X;
- X tem média μ e variância σ^2;
- $\bar{X} = \frac{X_1+X_2+\ldots+X_n}{n}$ é a média amostral.

Então:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$$

Prosseguindo... Como $\frac{\bar{X}-\mu}{\sigma\sqrt{n}} \sim N(0,1)$, temos que

$$\begin{aligned}
& \mathbf{P}\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| \leq \frac{\epsilon}{\sigma/\sqrt{n}}\right) \geq \gamma \\
\Leftrightarrow \quad & \mathbf{P}\left(|Z| \leq \frac{\epsilon}{\sigma/\sqrt{n}}\right) \approx \gamma \\
\Leftrightarrow \quad & \mathbf{P}\left(-\frac{\epsilon}{\sigma/\sqrt{n}} \leq Z \leq \frac{\epsilon}{\sigma/\sqrt{n}}\right) \approx \gamma
\end{aligned}$$

Dado γ, podemos obter z_γ da $N(0,1)$ tal que

$$\mathbf{P}(-z_\gamma \leq Z \leq z_\gamma) = \gamma$$

Dessa forma,

$$\frac{\epsilon}{\sigma/\sqrt{n}} = z_\gamma,$$

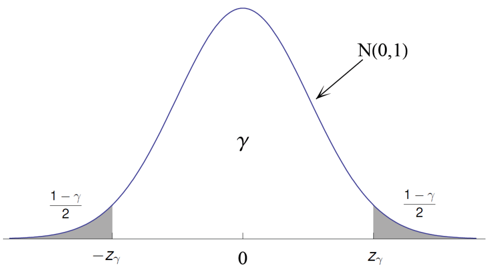
do que obtemos finalmente

$$n = \frac{\sigma^2 z_\gamma^2}{\epsilon^2} \tag{1}$$

Obs 1. Na fórmula (1) conhecemos z_γ e ϵ, mas σ^2 é a variância desconhecida da população. Para podermos ter uma ideia sobre n devemos ter alguma informação prévia sobre σ^2 ou, então, usar uma pequena amostra piloto para estimar σ^2.

Obs 2. z_γ é o quantil da Normal Padrão Z tal que

$$\mathbf{P}\left(-z_\gamma \leq Z \leq z_\gamma\right) = \gamma$$



Para $\gamma = 0,95$, por exemplo, $z_\gamma = 1,96$.

Exemplo

Suponha que uma determinada variável apresenta variância populacional $\sigma^2 = 16$. Calcule o tamanho da amostra *n* para obter uma estimativa para a média μ dessa população com margem de erro $\epsilon = 0,5$ e confiança $\gamma = 0,95$.

Solução Basta aplicar a fórmula (1).

$$n = \frac{\sigma^2 z_\gamma^2}{\epsilon^2} = \frac{16 \times (1,96)^2}{(0,5)^2} \approx 246$$

Exemplo

Suponha que uma determinada variável apresenta variância populacional $\sigma^2 = 16$. Calcule o tamanho da amostra *n* para obter uma estimativa para a média μ dessa população com margem de erro $\epsilon = 0,5$ e confiança $\gamma = 0,95$.

Solução Basta aplicar a fórmula (1).

$$n = \frac{\sigma^2 z_\gamma^2}{\epsilon^2} = \frac{16 \times (1,96)^2}{(0,5)^2} \approx 246$$

Tamanho de amostra para estimar proporção

No caso de proporções, a fórmula equivalente fica:

$$n = \frac{z_\gamma^2 p(1-p)}{\epsilon^2}$$

No entanto, como não conhecemos p, a verdadeira proporção populacional, podemos usar o fato de que $p(1-p) \leq \frac{1}{4}$. E a fórmula para determinação do tamanho da amostra fica

$$n \leq \frac{z_\gamma^2}{4\epsilon^2}.$$

Comentário.

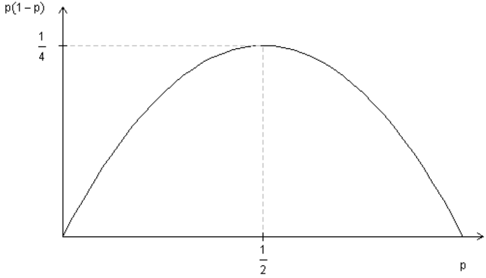
Lembramos que a fórmula original para dimensionamento da amostra é dada por:

$$n = \frac{z_\gamma^2 \sigma^2}{\epsilon^2}.$$

A fórmula da amostra inclui a variância, que não é dada. Acontece que a variância da Bernoulli(p) é igual a $p(1-p)$, e nesse caso, a fórmula para o tamanho da amostra fica:

$$n = \frac{z_\gamma^2 p(1-p)}{\epsilon^2}.$$

Se é verdade que desconhecemos o valor de p, que é justamente o que queremos estimar (um aparente paradoxo), por outro lado sabemos que $p(1-p)$ vale no máximo 1/4, que ocorre quando $p = 0,5$:



Assim, substituindo $p(1-p)$ por $\frac{1}{4}$ na fórmula anterior, chegamos a:

$$n \leq \frac{z_{\gamma}^{2}}{4\epsilon^{2}},$$

que é a fórmula apresentada no slide 17.

Determinando n desse jeito, estamos na verdade superestimando o valor de n que atende os requisitos de margem de erro e confiança solicitados. Para qualquer valor de p diferente de $p = 0,5$, precisaremos na verdade de uma amostra de tamanho menor do que a calculada assim, mas como não sabemos o valor de p, calculando assim saberemos que o nível de confiança e a margem de erro serão pelo menos tão exigentes quanto os solicitados.

Exemplo

Suponha uma pesquisa de intenção de votos. Calcule o tamanho da amostra para que o erro amostral de $\hat{p}$ seja menor do que $\epsilon = 0,03$, com probabilidade $\gamma = 0,95$.

Solução Considerando que:

- $\epsilon = 0,03$
- $\gamma = 0,95 \Rightarrow z_{0,95} = 1,96$

$$n \approx \frac{z_\gamma^2}{4\epsilon^2} = \frac{(1,96)^2}{4 \times (0,03)^2} = 1067$$

Exemplo

Suponha uma pesquisa de intenção de votos. Calcule o tamanho da amostra para que o erro amostral de $\hat{p}$ seja menor do que $\epsilon = 0,03$, com probabilidade $\gamma = 0,95$.

Solução Considerando que:

- $\epsilon = 0,03$
- $\gamma = 0,95 \Rightarrow z_{0,95} = 1,96$

$$n \approx \frac{z_\gamma^2}{4\epsilon^2} = \frac{(1,96)^2}{4 \times (0,03)^2} = 1067$$

Exercício

Suponha que você quer fazer uma pesquisa para avaliar a proporção da população de uma cidade favorável a uma determinada proposta de política pública. Determine o tamanho da amostra para obter estimativas da proporção populacional favorável à proposta com margem de erro não superior a 5% com confiança 95%.

B.1 Normal Padrão

Valores de p tais que $\mathbf{P}(Z < z_c) = p$.

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,10	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,20	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,30	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,40	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,50	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,60	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,70	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,80	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,90	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,00	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,10	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,20	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,30	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,40	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,50	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,60	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,70	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,80	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,90	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,00	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,10	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,20	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,30	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,40	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,50	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,60	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,70	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,80	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,90	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,00	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,10	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,20	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,30	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,40	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998