Statistical Analysis

Probability distributions

Ana Cristina Costa

ccosta@novaims.unl.pt

Topics

- **LU2 – Probability distributions**
  - Binomial distribution
  - Poisson distribution
  - Normal distribution
  - Approximation of the Binomial to the Normal distribution
  - t, Chi-squared and F distributions

Objectives

- **At the end of this learning unit students should be able to**
  - Describe the Binomial distribution and its applications
  - Describe the Poisson distribution and its applications
  - Describe the Normal distribution and its applications
  - Understand the parameters of the Binomial, Poisson and Normal distributions
  - Calculate probabilities based on Binomial, Poisson and Normal distributions
  - Explain the use of the Normal distribution as an approximation of the Binomial distribution
  - Describe the main characteristics of the Student's t, Chi-squared and Fisher-Snedecor's F distributions

Suggested reading

- Newbold, P., Carlson, W. L., Thorne, B. (2013). Statistics for Business and Economics. 8th Edition, Boston: Pearson, pages 159-173 (ch. 4), 206-225 (ch. 5).

- Mariappan, P. (2019). Statistics for Business. New York: Chapman and Hall/CRC, chapters 10 & 11.

Resources on the Internet

- The Pennsylvania State University (2021) "Section 2: Discrete Distributions". In STAT 414 Introduction to Probability Theory. (accessed: March 2021)

- The Pennsylvania State University (2021) "Section 3: Continuous Distributions". In STAT 414 Introduction to Probability Theory. (accessed: March 2021)

- Lane, D. M. (2015) "Normal Distributions". In *Online Statistics Education: An Interactive Multimedia Course of Study*, http://onlinestatbook.com.

- **Online tools and calculators**
  - Daniel Soper, Statistics Calculators, DanielSoper.com. (accessed: March 2021)
  - EssyCode, Plot Distributions. (accessed: March 2021)
  - Stat Trek, Distributions Calculators, StatTrek.com (accessed: March 2021)

Binomial distribution

- **Notation**
  - X~B(n, p) means that the random variable (r.v.) X follows a Binomial distribution with parameters *n* and *p*
    - The symbol '~' is used with the meaning "has the probability distribution"
    - The name of the probability distribution is abbreviated by its first letter
    - The probabilistic model is completely specified with its parameters

Binomial distribution

- **Bernoulli distribution**
  - **X~B(p)**

    $P(X = 1) = p$

    $P(X = 0) = 1 – p, \qquad 0 \le p \le 1$

  - Main application
    - Bernoulli trial: random experiment with exactly two possible outcomes, "success“ (X=1) and "failure“ (X=0), in which the probability of success (*p)* is the same every time the experiment is conducted

  - Mean and variance
    - $E(X) = p$
    - $V(X) = p(1 – p) = pq$

Binomial distribution

- **Example of a Bernoulli distribution**
  - Suppose 80% of villagers should be vaccinated. What is the probability of randomly choosing a vaccinated villager?
    - Success → 1 (vaccinated)
    - Failure → 0 (not vaccinated)
  - This experience corresponds to randomly selecting 1 individual
    - $P(X = 1) = 0.80$
    - $P(X = 0) = 1 – 0.80 = 0.20$
  - $X \sim B(p=0.80)$

Binomial distribution

- **Examples of Bernoulli distributions**
  - Bernoulli trials are quite ubiquitous in real-life situations
    - A team will win a championship or not
    - A student will pass or fail an exam
    - A statement is true or false
    - People agree or do not agree with a statement from a survey
    - A product is defective or not
    - …

Binomial distribution

- **Binomial distribution**
  - **X~B(n, p)**

$$P(X = k) = C_k^n p^k (1 - p)^{n-k}, \ k = 0, 1, 2, \dots, n, \quad 0 \le p \le 1$$

  - Main application
    - X represents the number of successes in *n* independent Bernoulli trials, with the probability of success on a single trial equal to *p*
  - Mean and variance
    - E(X) = np
    - V(X) = np(1 – p) = npq

Binomial distribution

- **Examples**
  - Sara usually wins 60% of the chess games against Rita. X represents the number of games that Sara wins Rita in 10 chess matches.
    - X~B(n=10, p=0.60)
  - A goalkeeper defends 95% of penalty shoot-outs. When games end in a tie it is necessary a penalty shoot-out with 5 shots on the goal from the penalty mark to find the winner. X is the number of shots the goalkeeper defends in a penalty shoot-out.
    - X~B(n=5, p=0.95)
  - A college administrator randomly samples students until he finds four that have volunteered to work for a local organization. Let X equal the number of students sampled. Is X Binomial?
    - No, because the number of trials *n* was not fixed in advance. Moreover, X counts the number of trials, not the number of successes.

Binomial distribution

- **Examples**
  - A company makes sports bikes. 90% pass final inspection (and 10% fail and need to be fixed). Let X equal the number of passes from ten inspections. How many bikes (out of ten) would you expect to pass?
    - $X \sim B(n=10, p=0.90)$. Hence, $E(X) = 10 \times 0.90 = 9$
  - In a store, out of all the customers who came there 25% bought a shirt. Let X equal the number of customers that by a shirt from the next nine customers entering the store. How many of them would you expect to buy a shirt?
    - $X \sim B(n=9, p=0.25)$. Hence, $E(X) = 9 \times 0.25 = 2.25 \approx 2$

Binomial distribution

- **Sampling with and without replacement**
  - In a set of 10 cell phones, 3 are defective. In a random sample of size n=2, consider the event P(B to occur, given that A has occurred) = P(B | A)
    - A = the 1st phone is defective
    - $\overline{\mathrm{A}}$ = the 1st phone is not defective
    - B = the 2nd phone is defective
  - With replacement sample
    - P(B | A) = P(B | $\overline{\mathrm{A}}$) = 3/10
    - ✓ Independent trials!
  - Without replacement sample
    - P(B | A) = 2/9 $\neq$ P(B | $\overline{A}$) = 3/9

Binomial distribution

- **Sampling with and without replacement**
  - Formally, the Binomial distribution should only apply to sampling with replacement, which guarantees independent trials
  - In practice, when the sample size *n* is small in relation to the population size *N*, we assume a random variable X, whose value is determined by sampling without replacement, follows (approximately) a Binomial distribution
  - On the other hand, if the sample size *n* is close to the population size N, then we assume the random variable X follows a Hypergeometric distribution

Binomial distribution

- **Examples**
  - Suppose individuals with a certain gene have a 0.70 probability of eventually contracting a certain disease. Let X equal the number of individuals who will contract the disease in a sample of 100 individuals with the gene participating in a lifetime study. Is X Binomial?
    - X follows (approximately) a B(n=100, p=0.7)
  - According to recent estimates, approximately 55 million people living in households in the USA own a “Sport/utility/special purpose” (SUV) vehicle. A poll of 100 random people living in households is conducted. Let X equal the number of people in the sample who own a SUV. Is X Binomial?
    - X follows (approximately) a Binomial distribution

Statista (2017) Statistics and Facts about Vehicles & Traffic. https://www.statista.com/statistics/227851/people-living-in-households-that-own-a-domestic-vehicle-of-the-type-sport-utility-special-purpose-usa/. (accessed Nov 2017)

Binomial distribution

- **Examples**
  - The Binomial distribution has shown to be an excellent model for a wide array of phenomena such as
    - Vote counts for two different candidates in an election
    - The number of male/female employees in a company
    - The number of accounts that are in compliance or not in compliance with an accounting procedure
    - The number of successful sales calls
    - The number of defective products in a production run
    - The number of days in a month your company’s computer network experiences a problem

Binomial distribution

- **Computing Binomial probabilities in Excel**
  - $F(x) = P(X \leq x) =$ BINOM.DIST(x; n; p; true)
    - Cumulative distribution function = true
  - $P(X = x) =$ BINOM.DIST(x; n; p; false)
    - Cumulative distribution function = false
  - $P(a \leq X \leq b) =$ BINOM.DIST.RANGE(n; p; a; b)
    - If *b* is omitted, it computes $P(X = a)$

Binomial distribution

- **Example 1**
  - Suppose you ask 100 people if they prefer Coke or Pepsi. Assume half the people prefer Coke to Pepsi or Pepsi to Coke. Let X be the number of people that prefer Coke to Pepsi in your sample.
    - $X \sim B(n=100, p=0.5)$

a) What is the probability that exactly 60 people will prefer Coke to Pepsi?

b) What's the chance that less than or equal to 60 people prefer Coke to Pepsi?

c) What's the chance that more than 60 people prefer Coke to Pepsi?

d) What's the chance that less than 60 people prefer Coke to Pepsi?

e) What's the chance that between 40 and 60 people, inclusive, prefer Coke to Pepsi?

Solved in the **Example1** sheet of the LU2_Examples Excel file

Poisson distribution

- **Poisson distribution**
  - **X~P(λ)**

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad k = 0, 1, 2, 3, \ldots; \ \lambda > 0 \qquad E(X) = V(X) = \lambda$$

  - X represents the number of events occurring in a fixed time interval (or spatial area, volume, ...), where the average rate is λ per unit time interval and

1. The probability of having more than one random event occurring in a very short time interval is essentially zero.
2. For a very short subinterval of length 1/n where *n* is a sufficiently large integer, the probability of a random event occurring in this subinterval is λ/n. In other words, the probability of a random event occurring in a given subinterval is proportional to the length of that subinterval.
3. The numbers of random events occurring in non-overlapping time intervals are independent.

Poisson distribution

- **Examples**
  - On average, three telephone calls arrive every 30 minutes at NOVA IMS's reception. X represents the number of calls arriving at the reception between 11:00 and 11:30.
    - $X \sim P(\lambda=3)$
  - The daily average requests for an ambulance at a certain health facility is two. X is the daily number of ambulance requests at the health facility.
    - $X \sim P(\lambda=2)$
  - A football team averages 1.7 goals per game. X is the number of goals per game.
    - $X \sim P(\lambda=1.7)$

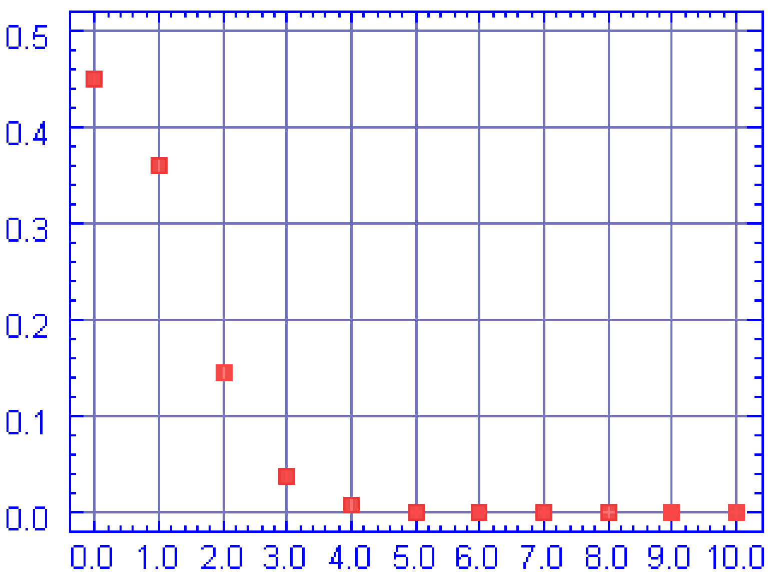
Benjamin Cronin (2017) "Poisson Distribution: Predict the score in soccer betting", Apr 27, Pinnacle.

Poisson distribution

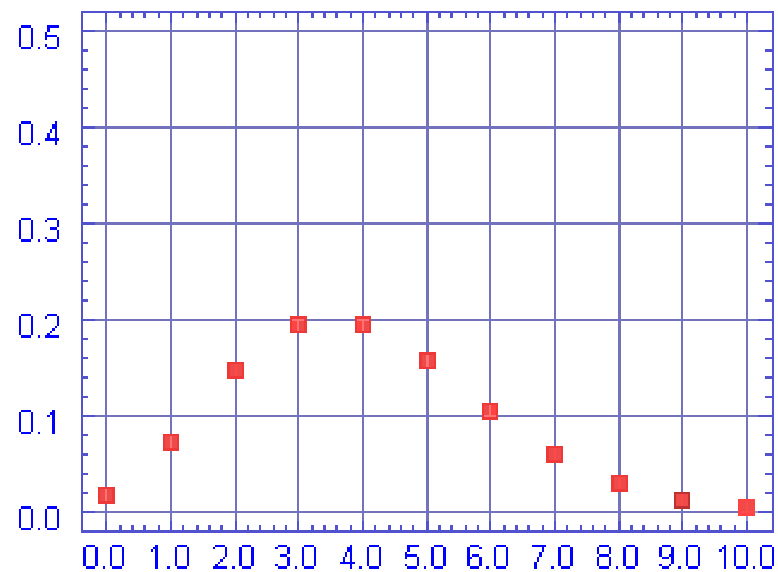
- **Examples**
  - The Poisson distribution has shown to be an excellent model for a wide array of phenomena such as
    - The number of traffic accidents in a stretch of highway in a given period of time
    - The number of accesses to a particular web server in a given period of time
    - The number of insurance losses/claims in a given period of time
    - The hourly number of customers arriving at a bank
    - The number of earthquakes occurring in a fixed period of time
    - The number of hurricanes in a year that originate in the Atlantic ocean
    - The number of typos in a book

Poisson distribution

- **Mode of the Poisson distribution**
  - If λ is an integer, the mode is bimodal in λ–1 an λ
  - Otherwise, the mode is the integer part of λ



X~P(λ=0.8)



X~P(λ=4)

Poisson distribution

- **Computing Poisson probabilities in Excel**
  - $F(x) = P(X \leq x) =$ POISSON.DIST(x; λ; true)
    - Cumulative distribution function = true
  - $P(X = x) =$ POISSON.DIST(x; λ; false)
    - Cumulative distribution function = false

Poisson distribution

- **Example 2**
  - Suppose a technical support centre receives on average 30 calls per hour.

a) What's the chance they'll get exactly 60 calls in 2 hours?

✓ $X \sim P(\lambda = 2 \times 30)$

b) What is the probability they'll get less than or equal to 60 calls in 2 hours?

c) What's the chance they'll get more than 60 calls in 2 hours?

d) What's the chance they'll get less than 60 calls in 2 hours?

e) What's the chance they'll get between 50 and 100 people calls, inclusive, in 2 hours?

f) What's the chance they'll get between 50 (exclusive) and 100 (inclusive) people calls in 2 hours?

Normal distribution

- **Normal or Gaussian distribution**
  - **X~N(μ, σ)**

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, -\infty < x < +\infty, \ -\infty < \mu < +\infty, \ \sigma > 0$$

  - Interminable real-life applications, and important statistical tool
    - Height and weight of people
    - Blood pressure
    - Measurement errors
    - Exam grades
    - IQ scores
    - Stock market returns

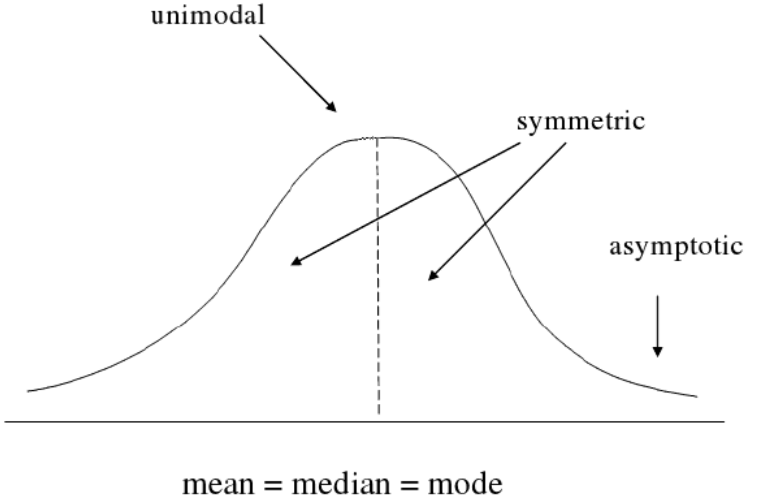
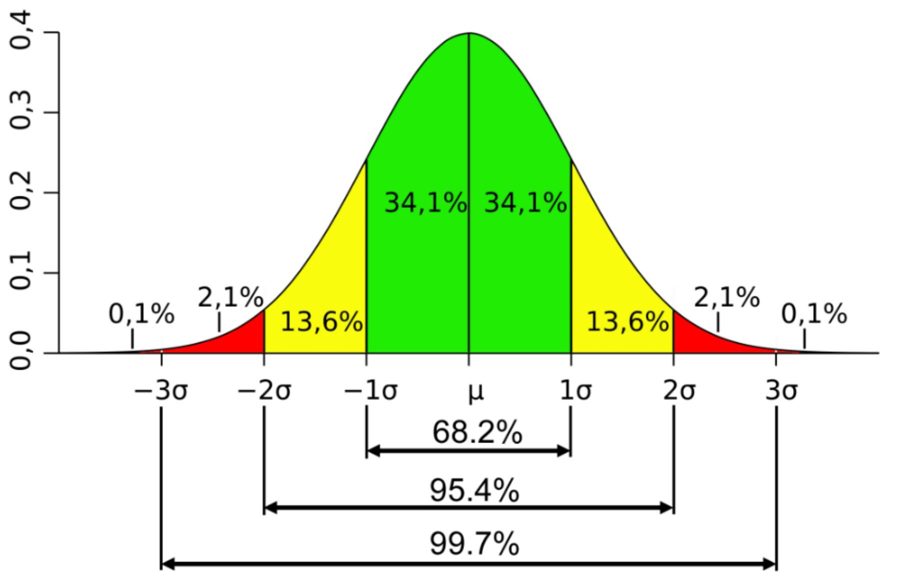


Fig. source: Yolanda Williams (2017) "Normal Distribution of Data: Examples, Definition & Characteristics", Study.com.

Normal distribution

- **Properties**
  - Approximately,
    - 68.2% of the area falls within 1 standard deviation of the mean
    - 95.4% of the area falls within 2 standard deviations of the mean
    - 99.7% of the area falls within 3 standard deviations of the mean



Normal distribution

- **Computing Normal probabilities in Excel**
  - $F(x) = P(X \leq x)$ = NORM.DIST(x; μ; σ; true)
    - Cumulative distribution function = true
  - NORM.INV(p; μ; σ; true) = x, so that $P(X \leq x) = p$
    - Cumulative distribution function = true

Normal distribution

- **Example 3**
  - A study showed that the weight of the “Smile“ candies followed a Normal distribution with mean 830 mg and standard deviation 24 mg.
    a) What is the probability of randomly selecting a candy weighting less than 860 mg?
    b) What is the probability of randomly selecting a candy weighting between 840 mg and 860 mg?

Normal distribution

- **Example 4**
  - Suppose you are selling Prozac, and you have an estimate that during the next year the average number of units sold will be 60 000 with a standard deviation of 5000.
  - Assuming the demand will be normally distributed, if you want to have a 1% chance of running out, how much Prozac should you make?
    - X~N(60 000, 5 000)
      - ✓ What is the value of *x* so that P(X > x) = 0.01?
      - ✓ The 99th percentile for the demand for Prozac is 71631.74. Hence, you should make at least 71632 units to guarantee that the chance of running out is less than 1%.

Normal distribution

- **Example 5**
  - The white blood cell (WBC) count may be indicative of infection if the count is high. WBC counts are approximately normally distributed in healthy people with a mean of 7550 WBC per mm^3 (i.e., per microliter) and a standard deviation of 1085.

    a) What proportion of subjects have WBC counts exceeding 9000?

    b) What proportion of patients have WBC counts between 5000 and 7000?

    c) If the top 10% of WBC counts are considered abnormal, what is the upper limit of normal?

Solved in the **Example5** sheet of the LU2_Examples Excel file

Normal distribution

- **Standard Normal distribution**
  - **Z~N(0, 1)**

$$X \sim N(\mu, \sigma) \quad \Rightarrow \quad Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

✓ $P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$

Normal distribution

- **Computing Standard Normal probabilities**
  - $\phi(z) = P(Z \leq z)$ = NORM.S.DIST(z; true)
    - Cumulative distribution function = true
  - Areas under the N(0, 1) can be found using a Standard Normal Table
    - $P(Z \leq 0.34) = \phi(0.34) = 0.63307$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793

Normal distribution

- **Example 6**
  - A study showed that the weight of the “Smile“ candies followed a Normal distribution with mean 830 mg and standard deviation 24 mg.
  - Compute the following probabilities <u>using the Standard Normal Table</u>

a) What is the probability of randomly selecting a candy weighting less than 860 mg?

b) What is the probability of randomly selecting a candy weighting less than 800 mg?

c) What is the probability of randomly selecting a candy weighting between 840 mg and 860 mg?

Normal distribution

- **Sums of independent Normal random variables**
  - Any linear combination of independent Normal random variables $X_1, X_2, \ldots, X_n$ has a Normal distribution
    - **Theorem**

If $X_1, X_2, \ldots, X_n$ are mutually independent random variables with distributions $N(\mu_i, \sigma_i)$ and a_i are constants ($i=1,\ldots,n$), then $\boldsymbol{Y = \sum a_i X_i \sim N(\mu, \sigma)}$ where

$$\mu = a_1\mu_1 + a_2\mu_2 + \ldots + a_n\mu_n$$

$$\sigma^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \ldots + a_n^2\sigma_n^2$$

Approximation of the Binomial to the Normal

- **Binomial distribution**
  - X~B(n, p) represents the number of successes in *n* independent Bernoulli trials, with the probability of success on a single trial equal to *p*
    - $x = 0, 1, 2, \ldots, n$; $n = 1, 2, 3, \ldots$; $0 \leq p \leq 1$
    - Why use the approximation? Because it might be easy to run into computational difficulties with the binomial formula. Moreover, the determination of a probability that X falls within a range of values is tedious to calculate!
  - Abraham de Moivre (1667-1754) found that for large values of *n* and for values of *p* not too close to zero or one, the Binomial distribution approaches the Normal distribution
    - This is the first example of the application of a fundamental theorem for the practice of statistics - the Central Limit Theorem

Approximation of the Binomial to the Normal

- **Examples**
  - Philip B. Stark (2016) “Java applet for the normal approximation to the binomial probability histogram”. SticiGui: https://www.stat.berkeley.edu/~stark/Java/Html/BinHist.htm
  - Micky Bullock (2011) “GeoGebra: Binomial Distribution with Normal and Poisson Approximation”: http://www.mickybullock.com/blog/2011/06/interactive-dynamic-binomial-distribution-applet/

Approximation of the Binomial to the Normal

- **Use the normal approximation to solve a binomial problem**
  - X~B(n, p)
    - Rule of thumb: **$np \geq 5$ and $n(1-p) \geq 5$**
    - Better rule of thumb: $np \geq 10$ and $n(1-p) \geq 10$

$$Z = \frac{X - np}{\sqrt{npq}} \text{ approx. } N(0, 1)$$

  - **Continuity correction** to improve the approximation
    - $P(X = a) = P(a - \mathbf{0.5} \leq X \leq a + \mathbf{0.5})$
    - $P(X \leq a) = P(X \leq a + \mathbf{0.5})$
    - $P(X \geq a) = P(X \geq a - \mathbf{0.5})$
    - $P(a \leq X \leq b) = P(a - \mathbf{0.5} \leq X \leq b + \mathbf{0.5})$

Approximation of the Binomial to the Normal

- **Example 7**
  - Suppose that, in a village, the vaccination coverage rate is 80% and 25 villagers are selected at random. Use the Normal distribution to find the following probabilities.

    a) What is the probability that a maximum of 7 vaccinated villagers were selected?

    b) What is the probability that more than 17 vaccinated villagers have been selected?

Approximation of the Binomial to the Normal

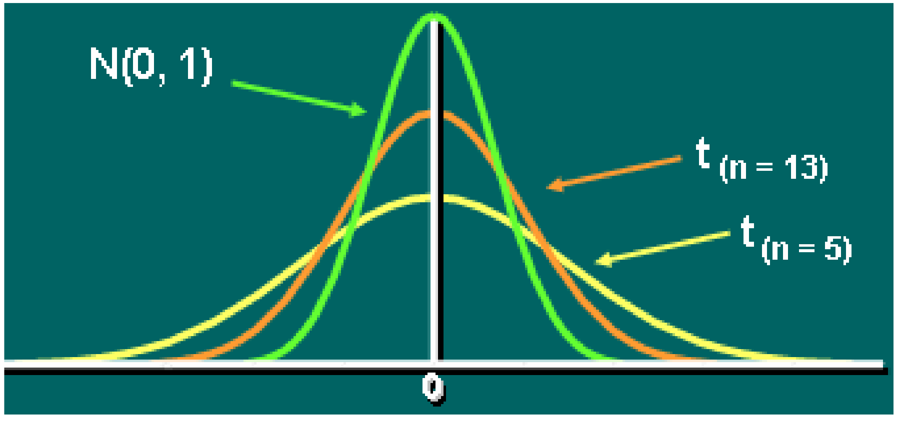
- **Example 7** (continued)
  - X~B(n=25, p=0.8)
    - ✓ $np = 25 \times 0.8 = 20 \geq 5$ and $nq = 25 \times 0.2 = 5 \geq 5$
    - • $\mu = np = 25 \times 0.8 = 20$ and $\sigma = \sqrt{npq} = \sqrt{25 \times 0.8 \times 0.2} = 2$

a) $P(X \leq 7 + 0.5) \approx P\left(Z \leq \frac{7.5-20}{2}\right) = P(Z \leq -6.25) = 0$

b) $P(X > 17) = P(X \geq 18) = P(X \geq 18 - 0.5) \approx P\left(Z \geq \frac{17.5-20}{2}\right) = P(Z \geq -1.25) = P(Z \leq 1.25) = 0.8944$

t, Chi-squared and F distributions

- **Student's t distribution** with n degrees of freedom
  - $X \sim t_{(n)}$
    - The pdf is bell shaped, symmetric in zero, unimodal and asymptotic
      - Similar to N(0,1) but with heavier tails
    - When the degrees of freedom (*df*) increase the Student's t distribution becomes similar to the Standard Normal distribution



t, Chi-squared and F distributions

Computing Student's t probabilities in Excel

- $F(x) = P(t_{(n)} \leq x) =$ T.DIST(x; n; true)
  - Cumulative distribution function = true

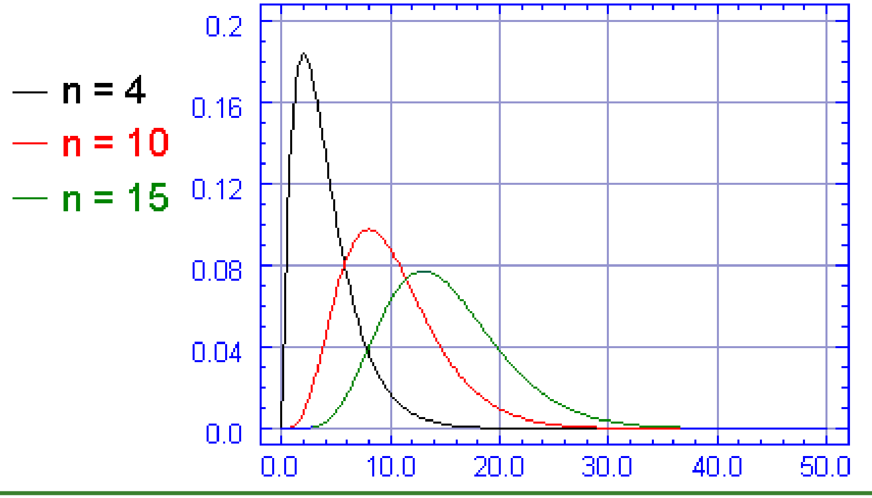
- T.INV(p; n; true) = x, so that $P(t_{(n)} \leq x) = p$
  - Cumulative distribution function = true

- Areas under the $t_{(n)}$ can be found using a Student's t Table
  - $P(t_{(n=3)} \leq 1.638) = 0.9$

$n \backslash q$	0.6	0.7	0.8	0.9	0.95	0.975	0.99	0.995	0.999	0.9995
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.656	318.289	636.578
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925	22.328	31.600
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841	10.214	12.924
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032	5.894	6.869
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707	5.208	5.959

t, Chi-squared and F distributions

- **Chi-Squared distribution** with n degrees of freedom
  - $X \sim \chi^2_{(n)}$
    - X only takes positive values
    - The pdf is positively skewed, unimodal and right-sided asymptotic
    - When the degrees of freedom (*n*) increase the Chi-Squared distribution becomes similar to the Standard Normal distribution



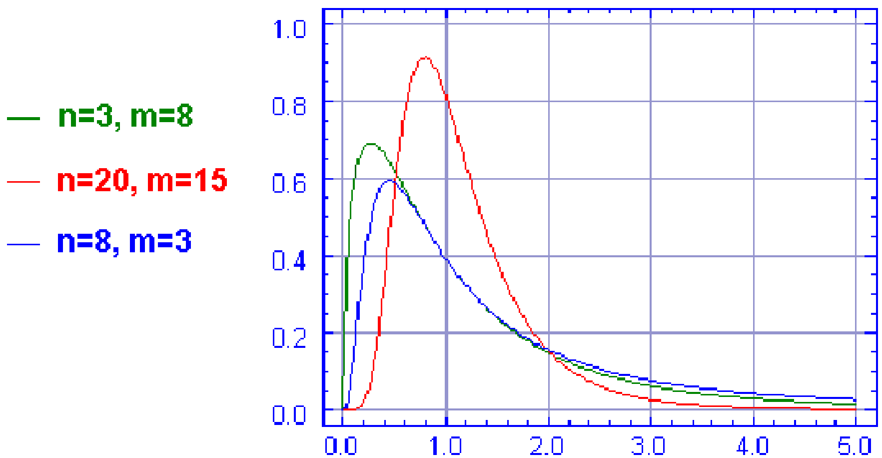
t, Chi-squared and F distributions

- **Computing Chi-Squared probabilities in Excel**
  - $F(x) = P(\chi^2_{(n)} \leq x) =$ CHISQ.DIST(x; n; true)
    - Cumulative distribution function = true
  - CHISQ.INV(p; n; true) = x, so that $P(\chi^2_{(n)} \leq x) = p$
    - Cumulative distribution function = true
  - Areas under the $\chi^2_{(n)}$ can be found using a Chi-Squared Table
    - $P(\chi^2_{(n=5)} \leq 11.070) = 0.95$

$n \backslash q$	0.005	0.01	0.025	0.05	0.2	0.5	0.8	0.95	0.975	0.99	0.995
1	0.000039	0.00016	0.001	0.004	0.064	0.455	1.642	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.446	1.386	3.219	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	1.005	2.366	4.642	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.649	3.357	5.989	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	2.343	4.351	7.289	11.070	12.832	15.086	16.750
6	0.676	0.872	1.237	1.635	3.070	5.348	8.558	12.592	14.449	16.812	18.548

t, Chi-squared and F distributions

- **Fisher-Snedecor's F distribution** with *n* and *m* degrees of freedom
  - $X \sim F_{(n, m)}$
    - X only takes positive values
    - The pdf is positively skewed, unimodal and right-sided asymptotic
    - When the degrees of freedom (*n* and *m*) increase the mode approximates 1



t, Chi-squared and F distributions

- **Computing F probabilities in Excel**
  - $F(x) = P(F_{(n,\,m)} \le x)$ = F.DIST(x; n; m; true)
    - Cumulative distribution function = true
  - F.INV(p; n; m; true) = x, so that $P(F_{(n,\,m)} \le x) = p$
    - Cumulative distribution function = true
  - Percentiles can be found using $F_{(n,\,m)}$ Tables for p=0.95, p=0.975, p=0.99
    - Table for p=0.95: $P(F_{(n=9,\,m=4)} \le 6) = 0.95$

$m \backslash n$	1	2	3	4	5	6	7	8	9	10	20	40	120	∞
1	161	199	216	225	230	234	237	239	241	242	248	251	253	254
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.45	19.47	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.66	8.59	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.80	5.72	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.56	4.46	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.87	3.77	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.44	3.34	3.27	3.23

t, Chi-squared and F distributions

Theorems

- If $Z_1, Z_2, \ldots, Z_n$ are iid N(0, 1) then

$$\sum_{i=1}^{n} {Z_i}^2 \sim \chi^2_{(n)}$$

- If Z~N(0, 1) and $Y \sim \chi^2_{(n)}$ are independent random variables then

$$T = \frac{Z}{\sqrt{Y/n}} \sim t_{(n)}$$

- If $X \sim \chi^2_{(n)}$ and $Y \sim \chi^2_{(m)}$ are independent random variables then

$$T = \frac{X/n}{Y/m} \sim F_{(n,m)}$$

Probability distributions

Do the homework*!*

Ana Cristina Costa, ccosta@novaims.unl.pt