Probability Theory

Genetics

Prof. Dr. Akin Pala

http://akin.houseofpala.com

akin@comu.edu.tr

Probability, Genetics

- Probability concepts are used in heredity calculations.
- Probability is used in calculations of
  - Inbreeding
  - Relationship coefficient (rel. matrix)
- P value in statistical genetics is a probability.

Genetics & Probability

- Mendel's (1865 Austrian monk) laws:
  - segregation
  - independent assortment

  reflect the same laws of probability that apply to tossing coins or rolling dice

Mendel's Laws of Inheritance

Law	Definition
Law of Segregation	During gamete formation, the alleles for each gene segregate from each other so that each gamete carries only one allele for each gene.
Law of Independent Assortment	Genes for different traits can segregate independently during the formation of gametes.
Law of Dominance	Some alleles are dominant while others are recessive; an organism with at least one dominant allele will display the effect of the dominant allele.

Probability, introduction

- Events occur based on chance:
  - Considering numbers from 1 to 10, 7 occurs
  - Genotype Ff for any individual
  - An individual picked randomly may be 1.75 m tall. His height is determined by
    - Genotype (has a chance factor)
    - Environment (has a chance factor)
    - Genotype x Environment (has a chance factor)

Chance

- Anything we cannot calculate/estimate
  - Dice
  - Within error
  - Tossing coin:
- Is called **chance**
- If we could calculate everything, chance would not exist.

Probability, introduction

- Lets call the probability of an event X. It is the ratio of the event where X occurs to the total number of trials, or the fraction of outcomes in which event X occurs
  - P(A)=the # of cases A occurs / the # of total cases
  - You have 6 red, 3 yellow and 1 green marbles in a jar.
    - The probability of drawing a red marble: 6/10
- The probability X occurs: Pr(X) or Prob (X)
- Probability of heads or tails when tossing one Lira coin

Heads or tails

- If the probability of heads is 0.90,
- Pr (heads)=0.90, meaning it is expected that 9 out of 10 times it will be heads, or the probability is %90 heads
  - Pr(tails)=1-0.90=0.10

Heads or tails

- Probability of heads is 0.50 and you approach that probability as the # of observations increase
  - It is 0.50 in infinity

Independent, exclusive

- Back to back events, such as tossing a coin 12 times, may be independent events,
- Those that may happen at the same time, such as tossing the coin and getting heads or tails, one instead of the other are mutually exclusive.
- Homopolymer sequence: AAAAAA

Independent events

- One event has no influence on the outcome of another event
- If events A & B are independent
  - then P(A&B) = P(A)*P(B)
- If P(A&B) = P(A)*P(B)
  - then events A & B are independent

Independent events

- Coin flipping

  if P(H) = P(T) = .5 then

  P(HTHTH) = P(HHHHH) =

  $.5*.5*.5*.5*.5 = .5^5 = .03$

- If you are flipping a coin and it has already come up heads 6 times in a row, what are the odds of an 7th head?

- 0.5

Mutually exclusive Events

- Mutually exclusive events are not independent
- rather, the most dependent kinds of events
  - if not heads, then tails
  - joint probability of 2 mutually exclusive events is 0
    - P(A&B)=0

Probability Rules

- Probabilities are between 0 (never) and 1 (always).
- All exclusive (independent, cannot happen at the same time) probabilities add up to 1.
- For example, if there are n possible outcomes,
- $Pr(1) + Pr(2) + .. + Pr(n) = 1$
- So, $Pr(1) = 1 - (Pr(2) + .. + Pr(n))$

AND/OR Rules

- **AND rule**: If X and Y are independent (knowing one does not give us a clue about the other event), probability that both X and Y occurs:
- Pr(X and Y) = Pr(X)*Pr(Y)
- **AND= multiply the probabilities, rule of multiplication**
- **OR rule: If X and Y** are separate (cannot happen at the same time) events, probability of X OR Y
- Pr(X or Y) = Pr(X) + Pr(Y)
- OR **= add the probabilities**, rule of addition

Example 1

- Assume rolling a dice and tossing 1 coin Lira,
- Probability that the dice will show an even number
- Probability that the dice will show an odd number?
- Possible outcomes of an honest dice 1, 2, 3, 4, 5, 6 with the same probability: 1/6. Getting an even number means 2 OR 4 OR 6.
  - These three events (2, 4, 6) are mutually exclusive, cannot happen at the same time (there is only one dice) OR = adding rule:
  - Pr(even) = Pr(2) + Pr(4) + Pr(6) = 3/6 = 1/2
- Probability of 5 on a dice AND head on a coin?
- Rolling dice and tossing coin are independent events, result of one does not affect the other.
- Pr(Head AND dice 5) = Pr(Head) * Pr(5) = 1/2*1/6 = 1/12

Sock Drawer

- Red, yellow, green and blue socks. Pick one without looking: red sock. Put it back.
  - Probability of picking another red sock?
- Because the sock is put back to the drawer, these are independent events. Picking red in the first event has nothing to do with the second event.
- P(Red)= ¼
- ¼ * ¼ = 1/16

Conditional Probability

- Conditional probability measures the probability of an event given that (by assumption, presumption, assertion or evidence) another event has occurred.
- For example, the probability that any given person has a cough on any given day may be only 5%. But if we know or assume that the person has a cold, then they are much more likely to be coughing. The conditional probability of coughing given that you have a cold might be a much higher 75%

Conditional Probability

- If Pr(X | Y) = Pr(X), then X and Y are independent.
  - Knowing Y occurs does not inform us about the result of X
  - Then the normal multiplication rule: Pr(X,Y) = Pr(X) * Pr(Y)
- When X and Y are not independent
  - X occurring gives us a clue about Y
- X and Y both occurring conditional probability Pr(X, Y),
- Pr(X,Y) = Pr(X | Y) * Pr(Y)
- Pr (X | Y): probability of X given Y
  - **Pr (X | Y)= Pr(X,Y) / Pr(Y)**

Conditional Probability, breeds

- Event **A** may depend on 3 different events, lets call them **B**, **C**, and **D**.
  - **Pr(A) = Pr(A|B)*Pr(B) + Pr(A|C)*Pr(C) + Pr(A|D)*Pr(D)**
- For example, 3 breeds may have different disease resistance. If event X is the disease occurring and Angus (A), Holstein (H) and Jersey (J) are the different genotypes: **Pr(X|H)**, is the risk of the H genotype for this disease.

Conditional Probability, class

- Genetics class midterm
  - %42 of the class over 50
  - %25 of the class both midterm & final over 50
  - Percentage of those that passed the midterm passing the final?
  - The condition here is that they passed the first exam.
- $Pr(X,Y) = Pr(X \mid Y) * Pr(Y)$ or,
- $P(X, Y)=P(X)\ P(Y \mid X)$

Conditional Probability, class

- P(X and Y)=P(X) P(Y | X)
- Both sides are divided by P(X),
- P(Y | X)= P(X and Y)/P(X)
- P(Y | X)= 0.25/0.42=0.60=%60 percent of the class that passed the midterm passed the final. 25 is 60 percent of 42.
  - 42 to 25
  - 100 to X

Conditional Probability, Aa

- Aa x Aa
  - AA and Aa Black Holstein, aa Red Holstein (red patches instead of black)
- Pr(AA) = Pr(aa) = 1/4
- Pr(Aa) = 1/2
- Pr(Holstein) = Pr(AA) + Pr(Aa) = 3/4.
- Probability that Black Holstein progeny has Aa genotype?

Conditional Probability, Aa

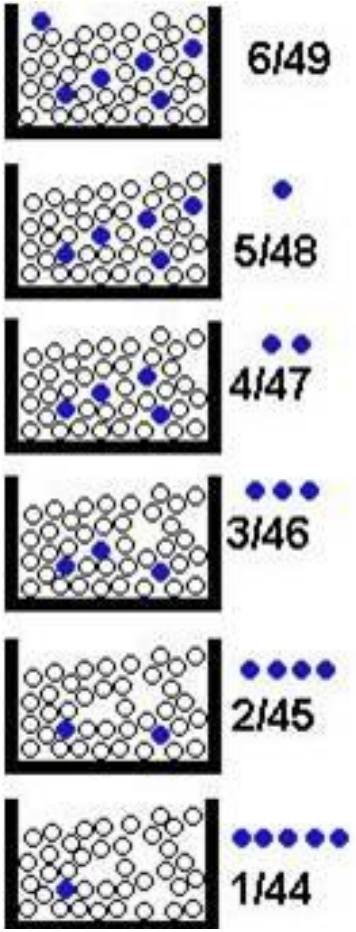
$$P(AA \mid blackHolstein) = \frac{1/4}{3/4} = \frac{1}{3}$$

Numbers Lottery (Sayısal Loto)

- 49 numbers and 6 balls are drawn at random.
- If all 6 of these balls match what you have, jackpot.
- Probability of matching only one of the balls is 6/49

Numbers Lottery (Sayısal Loto)

- Pr of matching one ball is 6/49
- Pr 1st ball and 2nd ball match:
- P(2|1)=5/48
- P(3|1and2)=4/47
- P(4|1,2,3)=3/46
- P(5|1,2,3,4)=2/45
- P(6|1,2,3,4,5)=1/44



Summary of Numbers Lottery

- Pr(Jackpot)

$=(6/49)*(5/48)*(4/47)*(3/46)*(2/45)*(1/44)$

$= 720/10{,}068{,}347{,}520$

$=1/13{,}983{,}816$

How long for %50 percent chance at winning?

- If every time 100 tickets are purchased, how long for %50 chance at winning?
- We are picking 100 sets of 6 numbers out of 13,983,816 sets of 6 numbers, so the chance of winning= 100/13,983,816=0.00000715
- Chance of losing=1-0.00000715

  = 0.99999285

How long for %50 percent chance at winning?

- Chance of losing= 0.99999285
- Chance of losing in k number of trials

$=(0.99999285)^k$

- We have to find such a k number of trials that chance of losing is 0.5 or smaller:
- $(0.99999285)^k \leq 0.5$

How long for %50 percent chance at winning?

- $(0.99999285)^k \leq 0.5$
- Taking the log of both sides,
- $k*\ln(0.99999285) = \ln(0.5)$

$$k = \frac{\ln(0.5)}{\ln(0.99999285)} \approx 96943$$

- The number of trials k is 96943

How long for %50 percent chance at winning?

- If we play every week, 52 times a year and 96943/52=1864 years to play 96 thousand times.
- If every play is 1 TL and if we win on 96943rd time, because every time we played 100 tickets,
  - 100 * 96943=9,694,300,
  - You need 9 million 694 lira and 1864 years to have a 50 percent chance of winning.

Spend 9 million TL & 1864 years

Milli Piyango Ana sayfa

SAYISAL LOTO 6/49

Sayısal Loto Çekiliş Sonuçları

Çekiliş Tarihi: 17-01-2015

Hafta: 949

02 19 30 39 42 49

Büyük İkramiye Kazanan İl/ İlçe: ADANA / SEYHAN

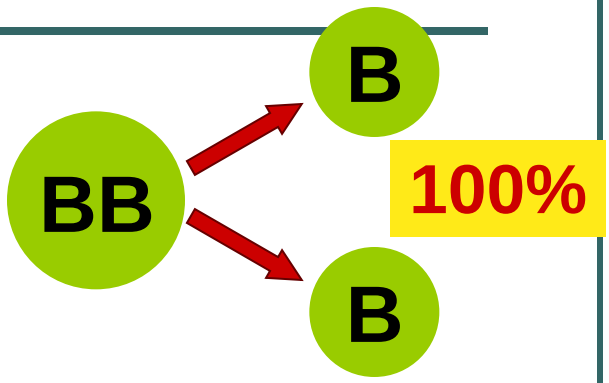
6 bilen kişi sayısı :	1	Kişi başına düşen ikramiye tutarı:	1.551.122,15 TL
5 bilen kişi sayısı :	181	Kişi başına düşen ikramiye tutarı:	3.608,35 TL
4 bilen kişi sayısı :	8.930	Kişi başına düşen ikramiye tutarı:	54,90 TL
3 bilen kişi sayısı :	173.425	Kişi başına düşen ikramiye tutarı:	8,00 TL

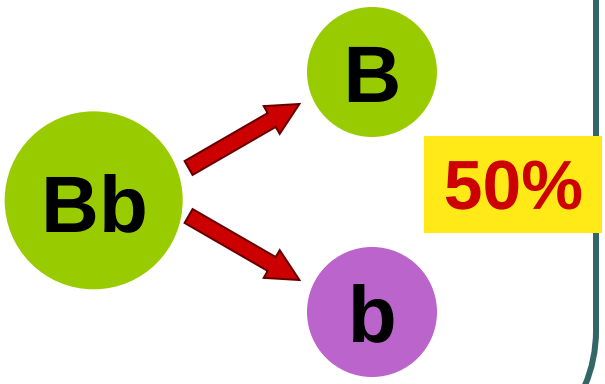
Milli Piyango Topu | Sayısal Loto | Süper Loto | On numara | Şans

PDF Büyük İkramiye İsabet Eden Bayiler

Probability & genetics

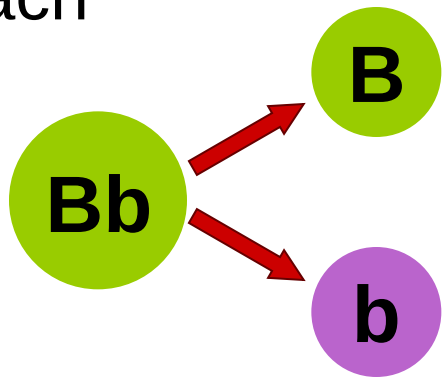
- Calculating probability of making a specific gamete is just like calculating the probability in flipping a coin
  - probability of tossing heads?
  - probability making a B gamete?





Probability & genetics

- Outcome of 1 toss has no impact on the outcome of the next toss
  - probability of tossing heads each time? **50%**
  - probability making a B gamete each time? **50%**



Calculating probability

P*p* x P*p*

Punnett square

female / eggs	male / sperm: **P**	male / sperm: ***p***
P	**PP**	**P*p***
p	**P*p***	***pp***

sperm		egg		offspring
P		**P**		**PP**
1/2	x	1/2	=	1/4
P		***p***		**P*p***
1/2	x	1/2	=	1/4
p		**P**		+
1/2	x	1/2	=	1/4
				1/2
p		***p***		***pp***
1/2	x	1/2	=	1/4

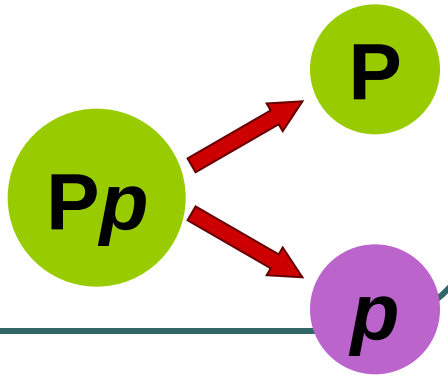
Rule of multiplication

- Chance that 2 or more independent events will occur together
  - probability that 2 coins tossed at the same time will land heads up

1/2 x 1/2 = 1/4

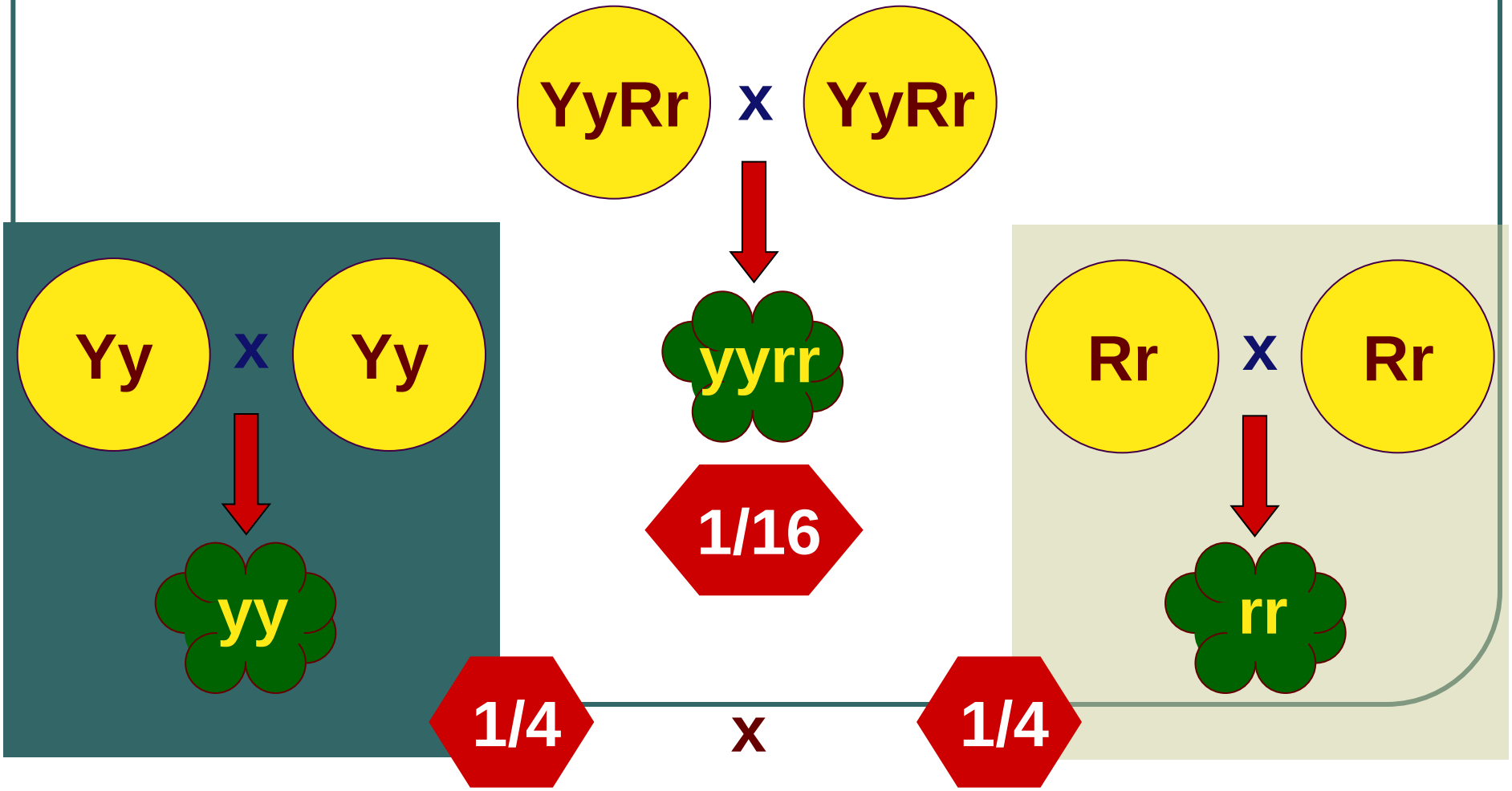
  - probability of P*p* x P*p* → *pp*

1/2 x 1/2 = 1/4



Calculating probability in crosses

Use rule of multiplication to predict crosses



Apply the Rule of Multiplication

AABbccDdEEFf x AaBbccDdeeFf

Hybrid Chicken Genes?

AabbccDdEeFF

AA x Aa → Aa	1/2
Bb x Bb → bb	1/4
cc x cc → cc	1
Dd x Dd → Dd	1/2
EE x ee → Ee	1
Ff x Ff → FF	1/4

1/64

Rule of addition

- Chance that an event can occur 2 or more different ways
  - sum of the separate probabilities
  - probability of Bb x Bb → Bb

sperm		egg		offspring
B		b		Bb
1/2	x	1/2	=	1/4
b		B		Bb
1/2	x	1/2	=	1/4

$$\begin{array}{r} 1/4 \\ +\ 1/4 \\ \hline 1/2 \end{array}$$

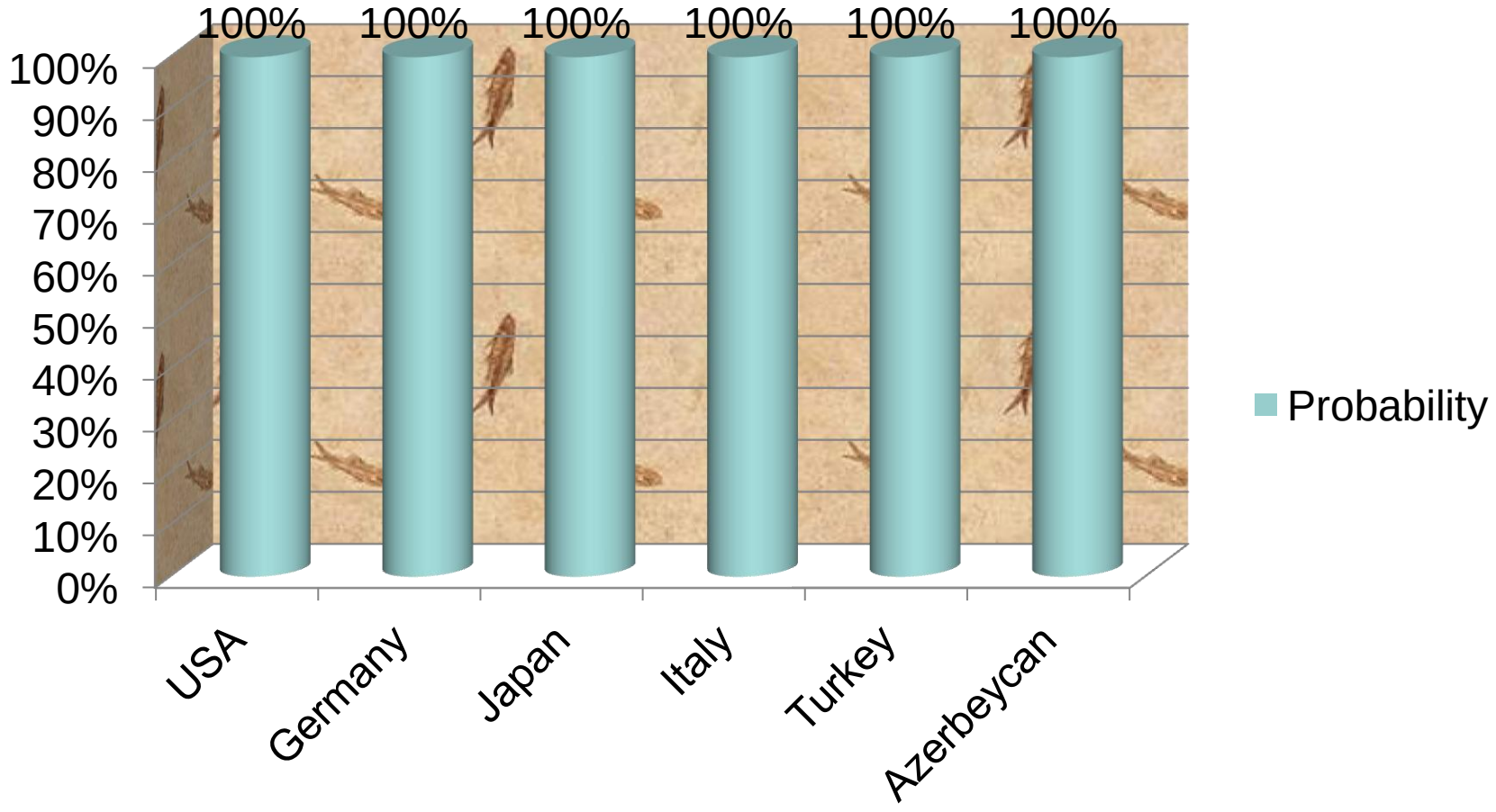
Chi-square test

- Test to see if your data supports your hypothesis
- Compare “observed” vs. “expected” data
  - is variance from expected due to “random chance”?
  - or is there another factor influencing data?
    - null hypothesis
    - degrees of freedom
    - statistical significance

CHI-SQUARE DISTRIBUTION TABLE

Degrees of freedom	Probability (p) value									
	← ACCEPT NULL HYPOTHESIS								REJECT →	
	0.99	0.95	0.80	0.70	0.50	0.30	0.20	0.10	**0.05**	0.01
1	0.001	0.004	0.06	0.15	0.46	1.07	1.64	2.71	**3.84**	6.64
2	0.02	0.10	0.45	0.71	1.30	2.41	3.22	4.60	**5.99**	9.21
3	0.12	0.35	1.00	1.42	2.37	3.67	4.64	6.25	**7.82**	11.34
4	0.30	0.71	1.65	2.20	3.36	4.88	5.99	7.78	**9.49**	13.28
5	0.55	1.14	2.34	3.00	4.35	6.06	7.29	9.24	**11.07**	15.09
6	0.87	1.64	3.07	3.38	5.35	7.23	8.56	10.65	**12.59**	16.81
7	1.24	2.17	3.84	4.67	6.35	8.38	9.80	12.02	**14.07**	18.48

Homework

- What is the probability that you will get an A this semester?

Probability of Death by Country