Binomial vs. Proportion

- Similarities and differences and how to decide which to use .

Example: We are told that historically only 55% of the students at this school vote in student elections.

A. BINOMIAL –

parameters

- p - probability of success
- n - # trials
- x - # of successes - must be discrete – here it would be the # of students who voted
- use *binompdf* for a single value
- use *binomcdf* for an interval of values

Sample questions:

1) Find the probability that no more than 300 of a group of 500 students will vote in the next election.

Solution: $n = 500$ $p = 0.55$ $x \leq 300$

$P(x \leq 300)$ = *binomcdf* (500, .55, 300) = 0.9893

2) Find the probability that at least 300 of the 500 will vote.

Solution: $n = 500$ $p = 0.55$ $x \geq 300$

$P(x \geq 300) = 1 - P(x \leq 299)$ = 1 – *binomcdf* (500, .55, 299) = 0.0136

3) Find the probability that between 150 and 250 of the 500 will not vote.

Solution: $n = 500$ $\mathbf{p = 0.45}$ $150 \leq x \leq 250$

$P(150 \leq x \leq 250)$ = *binomcdf* (500, .45, 250) – *binomcdf* (500,.45,149) = 0.9889

B. PROPORTION –

parameters

- p - population proportion with the given characteristic – i.e. here students who vote
- n - sample size
- $\hat{p}$ - sample proportion with the given characteristic
- use *normalcdf* (lower value, upper value, $\mu_{\hat{p}}$, $\sigma_{\hat{p}}$)

Sample questions:

1) Find the probability that no more than 60% of the 500 students will vote in the next election.

Solution: $n = 500$ $p = 0.55$ $\hat{p} \leq .60$ $\mu_{\hat{p}} = p = .55$ $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(.55)(.45)}{500}}$

$P(\hat{p} \leq .60)$ = *normalcdf* $\left(0, .60, .55, \sqrt{\frac{(.55)(.45)}{500}}\right)$ = .9877

2) Find the probability that at least 60% of the 500 students will vote in the next election.

Solution: $n = 500$ $p = 0.55$ $\hat{p} \geq .60$ $\mu_{\hat{p}} = p = .55$ $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(.55)(.45)}{500}}$

$P(\hat{p} \geq .60) = normalcdf\ (.60, 1, .55, \sqrt{\frac{(.55)(.45)}{500}}) = .0123$

3) Find the probability that between 30% and 50% of the 500 students will *not* vote in the next election.

Solution: $n = 500$ $\mathbf{p = 0.45}$ $.30 \leq \hat{p} \leq .50$ $\mu_{\hat{p}} = p = \mathbf{.45}$ $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(.45)(.55)}{500}}$

$P(.30 \leq \hat{p} \leq .50) = normalcdf\ (.30, .50, .45, \sqrt{\frac{(.45)(.55)}{500}}) = .9877$

M101 **COURSE REVIEW - Major Topics- may not be exhaustive**

1. Terminology - - classification of data - qualitative vs quantitative, discrete, continuous, level of measurement (NOIR), types of sampling etc.

2. Graphing of Data - Histograms - frequency, relative frequency, cumulative and relative cumulative dist'ns - class width, boundaries, class marks etc.

3. Measures of Central Tendency - Mean, median, mode - for group and ungrouped data

4. Measures of Spread - Range, Mean Deviation, Variance, Standard Deviation - know defining formula and how to use your calculator to find 's' and $\bar{x}$ using the Lists

5. Measures of Position - z - scores, Median, Quartiles, Percentiles (especially for the Normal)

- Boxplots, Chebyshev's Theorem, Empirical Rule

6. Probability - know rules, mutually exclusive, independent, conditional, Tree and VENN diagrams, expected value, "AND" and "OR" implications, Probability Dist'ns $\mu = \Sigma x \cdot P(x)$ $\sigma = \sqrt{\Sigma[x^2 \cdot P(x)] - \mu^2}$

7. Distribution of Random Variables -

Binomial - when to use and defining characteristics - discrete data

- Use of calculator binompdf(n,p,x) - for single value

binomcdf (n,p,x) - for values from 0 → X

- For Normal Approximation use $\mu = np \pm E.V$ $\sigma = \sqrt{npq}$

Normal - calculation of z - scores : $Z = \frac{x - \mu}{\sigma}$ OR $\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$ OR $\frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$

- Finding values given a probability: Z = invNorm (area to left)

X = invNorm (area, μ , σ)

t = invT (area to left, DF)

- Use of calculator to find probability $P(z_1 < z < z_2)$ = normalcdf(z_1, z_2)

OR $P(a < x < b)$ = normalcdf (a, b, μ, σ)

OR $P(a < \overline{x} < b)$ = normalcdf (a, b, $\mu, \frac{\sigma}{\sqrt{n}}$)

OR $P(a < \hat{p} < b)$ = normalcdf (a, b, $p, \sqrt{\frac{pq}{n}}$)

Distribution of Sample Means ($\bar{X}$) - Central Limit Theorem - Importance - restrictions

$$\mu_{\bar{x}} = \mu \qquad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

T-value - when to use - when σ is not known (S.D. is from the sample)

8. Estimation - Important for Final Exam

- Confidence Intervals - what they mean, common z - values, how to find other z - values or t - values

- Interval estimates for i) μ : $\bar{X} - E < \mu < \bar{X} + E$ where $E = z \cdot \frac{\sigma}{\sqrt{n}}$ OR $t \cdot \frac{s}{\sqrt{n}}$

ii) p : $\hat{p} - E < p < \hat{p} + E$ where $E = z \cdot \sqrt{\frac{pq}{n}}$ OR $z \cdot \sqrt{\frac{\hat{p}\hat{q}}{n}}$

- Sample size problems for i) μ where $n = \left(\frac{z \cdot \sigma}{E}\right)^2$ OR $\left(\frac{z \cdot s}{E}\right)^2$

ii) p where $n = \frac{z^2 pq}{E^2}$

- Know when to use the Z or t - value

- Margin of Error - for confidence intervals, sample size calculations or z-values

- If p and $\hat{p}$ are both unknown - use $\hat{p} = \hat{q} = 0.5$ for maximum sample size (ie. safest bet)

- If not possible to calculate s.d. then you can approximate s.d using Range Rule

9. Hypothesis Testing - 5 steps

(1) Identify the parameter and be able to (2) set up Null and alternative hypotheses

(3) Upper or Lower or Two - Tailed test - draw the sketch

(4) Find critical values - z^* or t^* and know when to use each

(5) Know proper format for setting up and making conclusions - claim is H_0 - Reject or FTR

- claim is H_1 - Support or Fail to Support

- Types of Error I, II

YOU ARE ALLOWED AN 8.5 " BY 11 " FORMULA SHEET (2 SIDES IN YOUR OWN HANDWRITING - NOT PHOTOCOPIED.

FINAL EXAM - TUESDAY, DEC. 15 - 9 - 12 NOON BR 205

M101 CUMULATIVE REVIEW I

1. An insurance company has determined that 65 % of new drivers have taken formal driving lessons. Of those that have taken lessons, the probability of having an accident in the first 6 months after getting their licence is .09 while the probability for those who didn't take lessons is .15. One experienced driver is selected at random :

a) what is the probability the person took lessons and was accident - free for the first 6 months?
b) what is the probability the person had an accident in the first 6 months?
c) If the person was accident - free, what is the probability they did not take lessons ?
d) if 2 people are selected at random,
   i) what is the probability that they both had accidents within 6 months ?
   ii) what is the probability that at least one was accident - free for the first 6 months ?

2. Assume that the maximum amount that a Grey Cup spectator spends on food and drink at the Grey Cup is $ 20 How large a random sample of spectators would you need to estimate the average amount that the Grey cup spectators spend on food and drink at the game accurate to within $ 1.00 with 95 % confidence?

3. How large a random sample is needed to estimate the percentage of trucks on the road with faulty brakes accurate to within 2 percentage points 24 times out of 25 :
a) If no previous information about the percentage is available ?
b) If it is known that the percentage is no more than 20 %

4. A survey of 700 North Shore adult residents , indicated that 72 had travelled to Europe within the last year. what is the margin of error in a 95 % C. I. for the proportion of all North Shore adult residents that had travelled to Europe last year ?

5. Telus claims that fewer than 10 % of all cell phone customers use prepaid cards. If we accept this claim, how large a random sample of Telus customers is needed to estimate the percentage of prepaid card customers if the estimate is to be accurate within 3 % at the 90 % confidence level ?

6. How large a random sample is needed to estimate the percentage of Vancouver adults who own a cell phone if the estimate is to be accurate within 4 percentage points " 19 times in 20 "

7. Classify each quantity as *continuous* or *discrete:*
a) Annual snowfall (in centimetres) at Whistler. _____
b) The number of season pass holders at Whistler . _____

8. Classify the data as *qualitative* or *quantitative* a) Age _____ b) Marital status _____

9. Determine which level of measurement is most appropriate - nominal, ordinal, interval or ratio
a) Olympic medals (Gold, silver, bronze) b) Countries that have hosted the Olympic Games

10. Identify the type of sampling used: random, systematic, stratified or cluster
a) One student is selected from each high school home room to sit on the student council.

11. On a class test, mean = 62 and s = 6
a) Chebyshev says that at least 3/4 of the data values on this test lie between what 2 values ?
b) If the minimum is 53, use the Range Rule of Thumb to approximate the maximum.

12. Determine the following for the data set : 7, 9, 10, 4, 15, 10, 13, 7, 6

a) mean ___ median ____ mode_____ b) Q_1 = _____ (show your work)

c) What percentile is represented by a score of 9 ? d) Draw the box-plot

13.

CLASS	FREQUENCY
10 - 14	2
15 - 19	1
20 -24	3
25 - 29	5

Use the given table to answer the following:

a) Class Width _____

b) Mean ≈

14. Determine the sample standard deviation: Enter data into your calculator to evaluate to 1 decimal:

X	Frequency
1	4
4	2
6	3

15. A Box contains 4 Red and 6 Blue Balls. A coin is tossed and if the result is Heads - a red ball is added to the box - if the result is Tails - a blue ball is added to the box. The coin is tossed, a ball is added and then one ball is drawn at random.

a) Draw the tree diagram to represent the act of tossing a coin and drawing a ball.

b) What is the probability that the ball drawn is Blue ? Red ?

16. Given P(A) = .40 , P(B) = .25 and P(A and B) = .10

a) Use a Venn diagram to answer the following questions:

i) P (A or B) = ii) P (B | A) = iii) are A & B independent ? Justify your answer.

iv) are A & B mutually exclusive ? Justify your answer.

17. The amount of sleep college students get is normally distributed with a mean of 8 hours and s.d. of 1. 5 hours . Find the probability that a randomly selected student will get:

a) At least 7 hours sleep tonight. b) Between 6 and 8. 5 hours sleep tonight.

c) Find the 65 th percentile sleeping time

18. A tele - marketing firm has determined that between 6:00 and 6:30 p.m. , 40 % of the telephones in the Vancouver area will be answered when the phone rings.

a) If the company makes 12 calls at this time, what is the probability that at least 4 will be answered ?

b) If you were to call people at 6 :00 p.m. until you received an answer, what is the probability that you would have your first answered call on your 3 rd call ?

19. In Canada 47 % of households with annual incomes under $ 20,000 own a microwave oven. If 25 households are chosen at random, find the probability that:

a) exactly 15 own a microwave. b) At least 10 but less than 15 own a microwave.

20. The diameters of the base of all trees harvested by a logging company was found to be normally distributed with a mean of 62 cm. and a standard deviation of 8 cm.

a) If you sample a log at random, what is the probability that its diameter will be within 4 cm. of the average diameter of all logs ?

b) A log is considered to be superior if the base diameter is over 75 cm. What percent of logs harvested will be considered superior ?

c) To be in the top 10 % largest logs, a tree would have to have a diameter greater than what value ?

M101 REVIEW 11

1. During the recent Olympics, it was noted among a group of friends that a total of 55% of the group went to a athletic event **(A)** and 85% went to a street celebration **(S)** with some of the group doing both. It was further noted that 10% of the group did neither.

- Answer the following using a Venn diagram or the probability formulas - answer as a %.

a) What per cent did both?

b) For this example what is P($A \cup S$) ?

c) Determine $P(A \cap \overline{S})$ and explain what this means *in the context of this example.*

d) Given that a particular person went to an athletic event, what is the probability that they also attended a street celebration ?

e) Are the two events independent? Justify your answer mathematically.

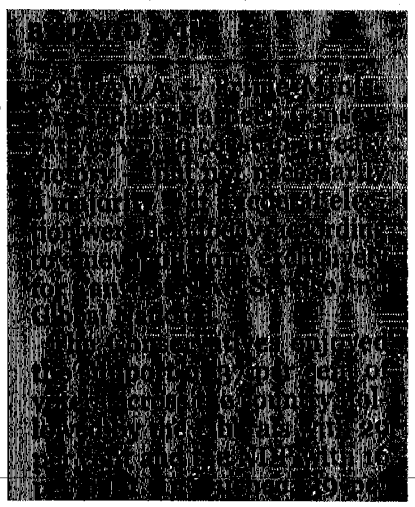
2. . A patient with kidney failure is awaiting a kidney transplant. His doctor informs him that for patients in his condition, 90% survive the transplant operation and 10% do not. For those that survive, the transplant is successful 60% of the time and the other 40% must return to dialysis. The proportions who survive for at least 5 years after surgery is 50% for those who returned to dialysis and 70% for those who did not have to return to dialysis.

a) Set up the tree diagram to illustrate this information. Identify variables used.

b) What is the probability that the patient will survive for at least 5 years? Answer as a per cent.

. . . . . .

3. "**Tories Not in Majority Territory**" use the information contained in the attached article from the "Vancouver Sun" on March 2, 2010 to answer the questions - be sure to **define variables**, **include the probability statement** and the **calculator programmes used** in your answers . (answer to 4 decimals)



A random sample of 200 Canadian voters is taken, what is the probability that:

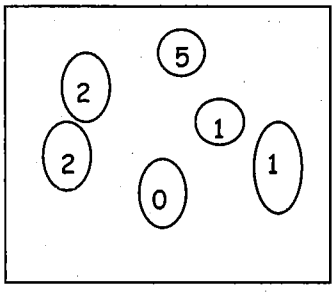
a) there are 60 Liberal supporters in the group?

b) At least 60 will support the Conservatives?

c) there are between 20 and 40 NDP supporters in the group?

4. **Two tags** are chosen at random ***without replacement*** from the box having the six numbered tags as shown

Construct the probability distribution for the *product of the numbers on the tags* drawn.



M101 CUMULATIVE REVIEW I - ANSWER KEY

Note: Normal dist'n solutions were done on the calculator so answer found with the table may vary slightly

1. a) .5915 b) .111 c) .3346 d) i) $(.111)^2 = .012$ ii) . 988

2. Approximate s using Range rule with max \$ 20 and min \$ 0 $n \geq 97$

3. a) use $p = q = .5$ $n \geq 2627$ b) use $p = .19$ $n \geq 1617$

4. E = 2 % 5. $n \geq 247$ 6. Use $p = .5$ $n \geq 601$ 7. c) Continuous d) Discrete

8. a) Quantitative b) Qualitative 9. a) ordinal b) nominal 10. Stratified

11. a) k = 2 the min is $\overline{X} - 2s = 62 - 2(6) = 50$ and max is 74
b) Range $\approx 4s = 24$ therefore 62 ± 12 ie. 50 to 74

12. a) Mean = 9 Median = 9 Mode = 7 and 10 (bimodal) b) $Q_1 = 6.5$ c) P_{44}

13. a) width = 5 b) Mean = 22 (use class mark)

14. Enter data into List 1, frequencies into List 2 then s = 2.3

15. b) P (B) = P (H B or T B) = $\frac{1}{2} \cdot \frac{6}{11} + \frac{1}{2} \cdot \frac{7}{11} = \frac{13}{22}$ P (R) = 1 - P (B)

16. a) .40 + .25 − .10 = .55 ii) .10 / .40 = .25 iii) Independent since P(B) = P(B / A)
iv) not m.e. since P (A and B) ≠ 0

17. a) P ($x \geq 7$) = . 7475 b) P ($6 \leq x \leq 8.5$) = .5393 c) invNorm(.65, 8, 1.5) = 8 .58 hours

18. a) P($x \geq 4$) = 1 - P($x \leq 3$) = 1 - binomcdf (12, .4, 3) = .7747 b) P(3rd call) = (.6)(.6)(.4) = . 144

19. Best ⇒ lowest time then x = invNorm(.25, 462.1, 76.3) = 410 . 6 minutes

20. a) P (x = 15) = binompdf (25, .47, 15) = . 0690

b) P ($10 \leq x < 15$) = P ($10 \leq x \leq 14$) = binomcdf (25, .47, 14) − binomcdf (25, .47, 9) = .6807

21. a) P ($E \leq 4$) = normalcdf (-4/8 , 4/8) = . 3829

b) P ($x > 75$) = normalcdf (75, 1 E 99, 62, 8) = .0521 i.e. 5 . 21 %

c) x = invNorm (.90 , 62, 8) = 72.25 cm for top 10 %

TESTING HYPOTHESES PROBLEMS - SD NOT KNOWN

1. Reject $H_0: \mu = 400$ for $\alpha = .01$ - there is enough evidence to reject claim that coverage is 400 sq.ft per gallon

2. Reject $H_0: \mu \geq 400$ for $\alpha = .01$ - there is enough evidence to support the claim that the painter is applying paint at less than 400 sq. feet per gallon

3. FTR $H_0: \mu \geq \$\ 18\,000$ at $\alpha = .05$ - there is not enough evidence to reject the claim that average income is at least $ 18,000

4. FTR $H_0: \mu \geq \$\ 12.50$ at $\alpha = .05$ - there is not enough evidence to reject the claim that the average cab fare is $ 12.50

5. $ 10.56 < μ < $ 14.94

6. Mean = 59 and s = 4.4 hours

a) Reject $H_0: \mu \leq 55$ hours at $\alpha = .05$ - we reject claim that the average # hours per week is 55 hours or less

b) We are 90 % confident that the mean lies in the interval 55.8 < μ < 62.2 hours

TESTING HYPOTHESES ABOUT P

1. Reject $H_0: p = .22$ at $\alpha = .05$ - claim is not reasonable

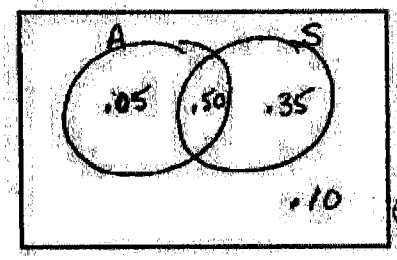
2. Reject H_0 if $z^* < -2.33$ and FTR H_0 if $z^* \geq -2.33$

3. . Reject H_0 if $z^* > 2.33$ and FTR H_0 if $z^* \leq 2.33$

4. FTR $H_0: p \leq .25$ at $\alpha = .05$ - insufficient evidence to reject claim

6. During the recent Olympics, it was noted among a group of friends that a total of 55% of the group went to a athletic event (A) and 85% went to a street celebration (S) with some of the group doing both. It was further noted that 10% of the group did neither.

- Answer the following using a Venn diagram or the probability formulas - answer as a %.



(1)

a) What per cent did both?

10% + 85% + 55% = 150%

∴ 50% did both.

(1)

b) For this example what is $P(A \cup S)$?

$P(A \cup S) = .05 + .50 + .35 = \boxed{.90}$ or 90%

1.5

c) Determine $P(A \cap \overline{S})$ and explain what this means *in the context of this example.*

$P(A \cap \overline{S}) = .05$ or 5%

1.5

- 5% of the people went to an athletic events but did not attend the street celebrations

d) Given that a particular person went to an athletic event, what is the probability that they also attended a street celebration?

$$P(S|A) = \frac{P(S \cap A)}{P(A)} = \frac{.50}{.55} \approx .9090.. \approx \boxed{91\%}$$

1.5

8

e) Are the two events independent? Justify your answer mathematically.

1.5

check $P(S)$ and $P(S|A)$

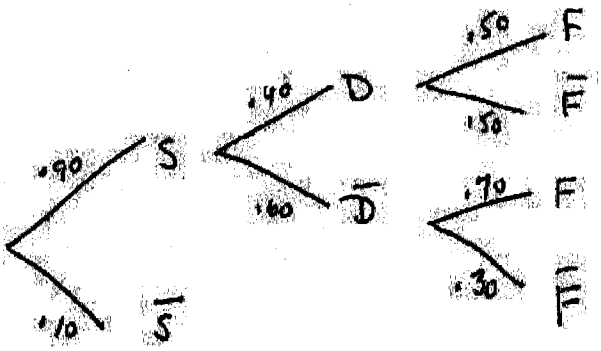
$P(S) = .85$ > Not the same ∴ not independent

$P(S|A) = .91$

9. A patient with kidney failure is awaiting a kidney transplant. His doctor informs him that for patients in his condition, 90% survive the transplant operation and 10% do not. For those that survive, the transplant is successful 60% of the time and the other 40% must return to dialysis. The proportions who survive for at least 5 years after surgery is 50% for those who returned to dialysis and 70% for those who did not have to return to dialysis.

a) Set up the tree diagram to illustrate this information. Identify variables used.

Let S = survive Surgery D = return to dialysis F = survive 5 years



.6 1 1.5

3

b) What is the probability that the patient will survive for at least 5 years? Answer as a per cent.

$P(\text{survive at least 5 years}) = P(SDF \text{ or } S\bar{D}F)$

$= .90(.40)(.50) + .90(.60)(.70)$

$= .558 \approx \boxed{55.8\%}$ survive at least 5 years.

2

10. "Tories Not in Majority Territory" use the information contained in the attached article from the "Vancouver Sun" on March 2, 2010 to answer the questions - be sure to define variables, include the probability statement and the calculator programmes used in your answers. (answer to 4 decimals)

BY DAVID AKIN

OTTAWA — Prime Minister Stephen Harper's Conservatives would coast to an easy victory — but not necessarily a majority — if a federal election were held today, according to a new poll done exclusively for Canwest News Service and Global National.

The Conservatives enjoyed the support of 37 per cent of voters across the country, followed by the Liberals with 29 per cent and the NDP with 16 [illegible]

A random sample of 100 Canadian voters is taken, what is the probability that:

let x = # supporters

a) there are 40 Conservative supporters in the group?

$n=100$, $p=.37$, $x=40$

$P(x=40) = \text{binompdf}(100, .37, 40)$

$= \boxed{.0671}$ 1.5

b) the majority of the group will support the Conservatives?

$x \geq 51$

$n=100$, $p=.37$, $x \geq 51$

$P(x \geq 51) = 1 - P(x \leq 50)$

$= 1 - \text{binomcdf}(100, .37, 50)$

$= \boxed{.0030}$ 1.5

c) there are between 20 and 40 Liberal supporters in the group?

$n=100$, $p=.29$

$P(20 \leq x \leq 40) =$

$= \text{binomcdf}(100, .29, 40) - \text{binomcdf}(100, .29, 19)$

$= \boxed{.9778}$ 2

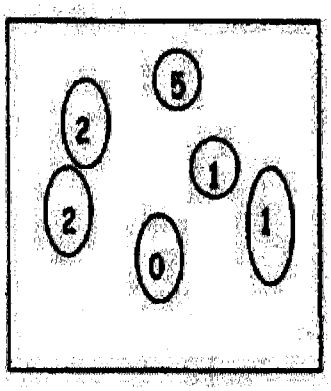
5

10

11. Two tags are chosen at random *without replacement* from the box having the six numbered tags as shown

Construct the probability distribution for the *sum of the numbers on the tags* drawn.

3



Let X = sum of 2 tags

$n = 6 \times 5 = 30$ outcomes.

2nd \ 1st	0	1	1	2	2	5
0	X	1	1	2	2	5
1	1	X	2	3	3	6
1	1	2	X	3	3	6
2	2	3	3	X	4	7
2	2	3	3	4	X	7
5	5	6	6	7	7	X

Sum

X	P(X)
1	4/30
2	4/30
3	8/30
4	2/30
5	2/30
6	4/30
7	4/30
①	1

②