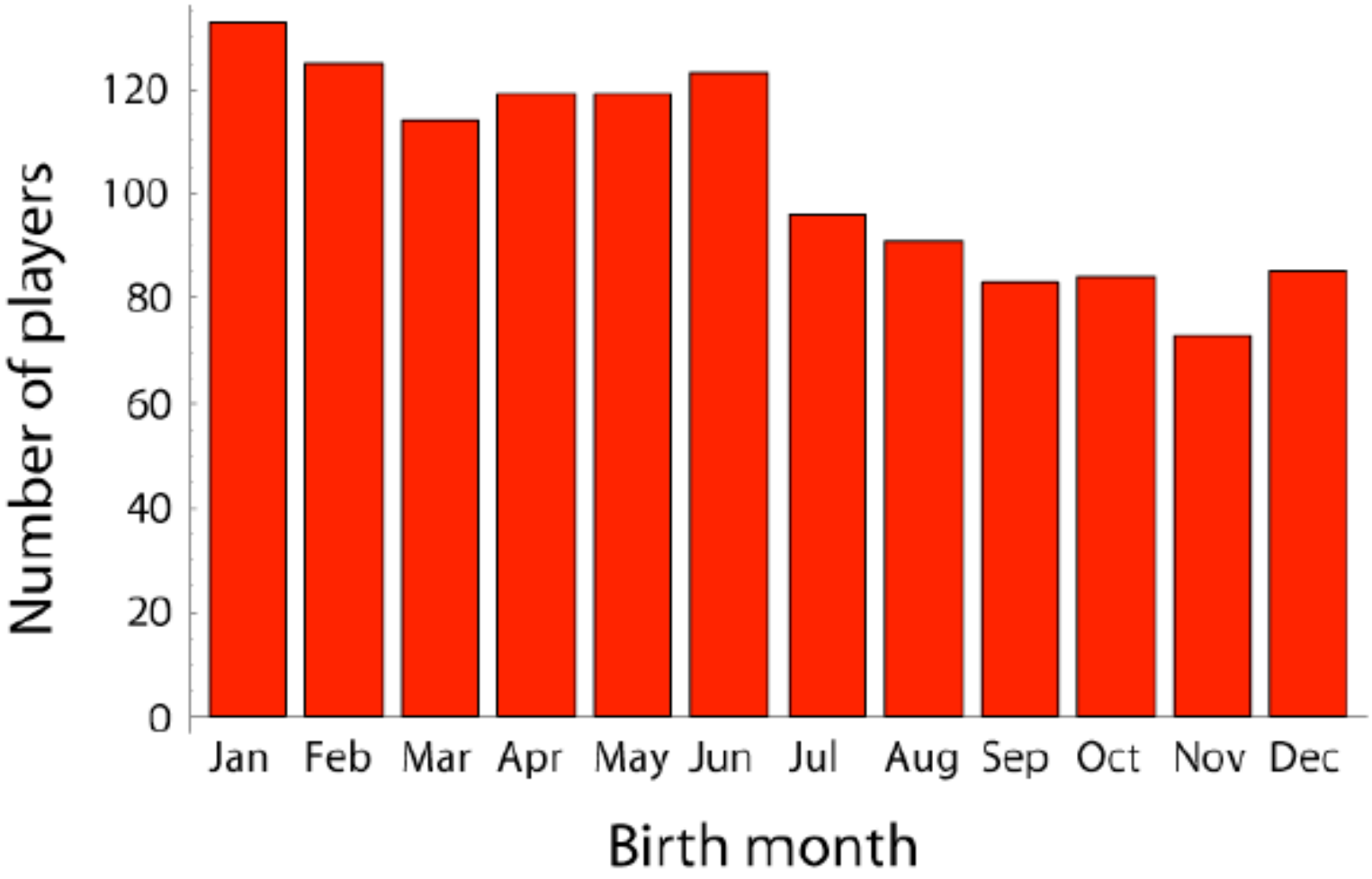
The birth month of Canadian hockey players

Month	Number of Players
January	133
February	126
March	114
April	119
May	119
June	123
July	96
August	91
September	83
October	84
November	73
December	85

Goodness of Fit



Goodness of Fit

H_0:

The probability of a NHL birth occurring in any given month is equal to national birth proportions

H_A:

*The probability of a NHL birth occurring in any given month is **NOT** equal to national birth proportions*

Goodness of Fit

Month	# of Players	Expected %
January	133	7.94
February	126	7.63
March	114	8.72
April	119	8.63
May	119	8.95
June	123	8.57
July	96	8.76
August	91	8.5
September	83	8.54
October	84	8.19
November	73	7.70
December	85	7.85
TOTAL	1245	100.0

Goodness of Fit

Month	# of Players	Expected (%)	Expected (of 1245)
January	133	7.94	99
February	126	7.63	95
March	114	8.72	109
April	119	8.63	107
May	119	8.95	111
June	123	8.57	107
July	96	8.76	109
August	91	8.5	106
September	83	8.54	106
October	84	8.19	102
November	73	7.70	96
December	85	7.85	98
TOTAL	1245	100.0	1245

We'll go through the calculation for January:

$$\frac{(Observed_i - Expected_i)^2}{Expected_i} = \frac{(133-99)^2}{99} = \frac{1156}{99}$$

Calculating χ^2:

$$\chi^2 = \sum_i \frac{(Observed_i - Expected_i)^2}{Expected_i}$$

$$= \frac{1156}{99} + \frac{900}{95} + \frac{25}{109} + \frac{144}{107} + \frac{64}{111} + \frac{256}{107} + \frac{169}{109} + \frac{225}{106} + \frac{529}{106} + \frac{324}{102} + \frac{529}{96} + \frac{169}{98}$$

= 44.77

$$\text{dof} = 12 - 1 - 0 = 11$$

What is true about p-value in terms of x^2 goodness-of-fit test?

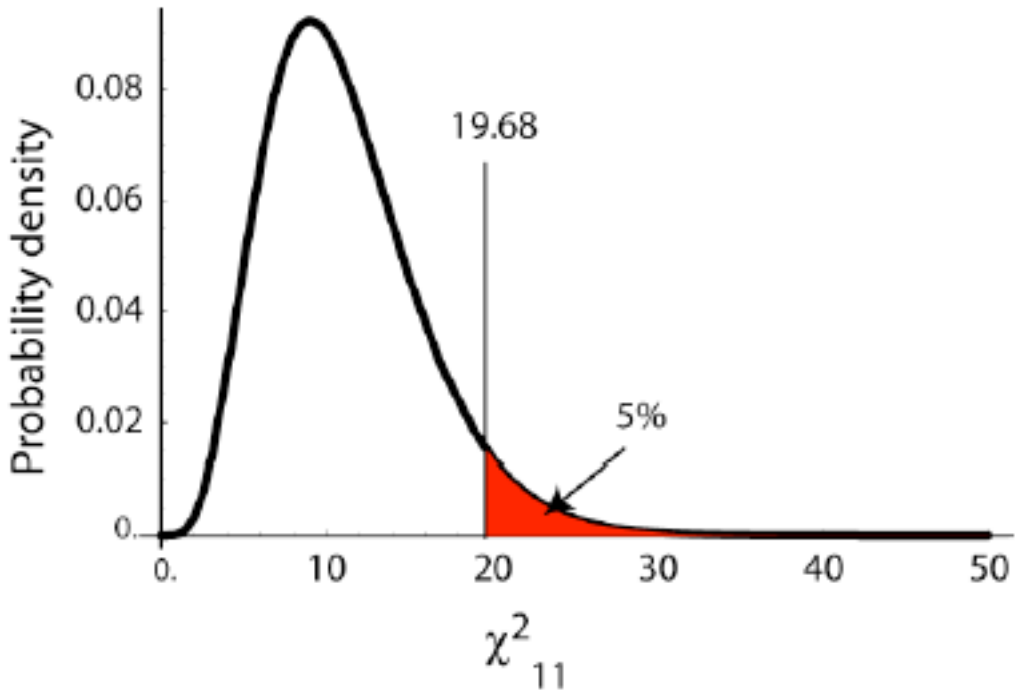
---

a. The p-value is the probability of getting a x^2 value less than the observed x^2 value calculated from the data
b. The p-value is the probability of getting a x^2 value equal to the observed x^2 value calculated from the data
c. The p-value is the probability of getting a x^2 value greater than or equal to the observed x^2 value calculated from the data
d. The p-value does not make any significant impact

Find Critical Value using Table A:

df	0.999	0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005	0.001
1	0.0000016	0.000039	0.00016	0.00098	0.00393	3.84	5.02	6.63	7.88	10.83
2	0.002	0.01	0.02	0.05	0.10	5.99	7.38	9.21	10.60	13.82
3	0.02	0.07	0.11	0.22	0.35	7.81	9.35	11.34	12.84	16.27
4	0.09	0.21	0.30	0.48	0.71	9.49	11.14	13.28	14.86	18.47
5	0.21	0.41	0.55	0.83	1.15	11.07	12.83	15.09	16.75	20.52
6	0.38	0.68	0.87	1.24	1.64	12.59	14.45	16.81	18.55	22.46
7	0.60	0.99	1.24	1.69	2.17	14.07	16.01	18.48	20.28	24.32
8	0.86	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.95	26.12
9	1.15	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59	27.88
10	1.48	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19	29.59
11	1.83	2.60	3.05	3.82	4.57	19.68	21.92	24.72	26.76	31.26
12	2.21	3.07	3.57	4.40	5.23	21.03	23.34	26.22	28.30	32.91

Find Critical Value using Table A:



ALWAYS CONCLUDE:

We can reject the null hypothesis: NHL players are *not* born in the same proportions per month as the population at large with P-value $\leq \alpha$ (=0.05).

Was Mendel's data 'too good'?

- RA Fisher accused Mendel of having data that fit too well (approximately 3/100,000 experiments should have data that fit as well as Mendel's)

- A raging debate ever since:
  - http://www.istics.net/stat/
    - Go to "Not random enough" in the sidebar
  - http://arxiv.org/pdf/1104.2975.pdf

Example: the results of a **Monohybrid cross** between a (heterozygous) yellow pea plant **(Yy)** and a green pea **(yy)** plant are as follows: **14 yellow** and **6 green**. Are these results consistent with Mendel's first law (**segregation)** which should a 1:1 ratio in this case? (alpha = 0.05). Punnet square shown:

Yellow (Hetero)\Green (Homo)	y	y
Y	Yy	Yy
y	yy	yy

A. Yes – the results FTR the null hypothesis
B. No – the results Reject the null hypothesis
C. Yes – the results reject the null hypothesis
D. No – the results FTR the null hypothesis

Goodness of Fit

Example: the results of a **Monohybrid cross** between a (heterozygous) yellow pea plant (Yy) and a green pea (yy) plant are as follows: **14 yellow** and **6 green**. Are these results consistent with Mendel's first law (**segregation)** which should a 1:1 ratio in this case?

step 1

H_0:

If the hypothesis (Yy is yellow and yy is green) is true then we would expect a 1:1 ratio in progeny (or, to put it into counts: 10 **Yellow** and 10 green)

H_A:

The genotypes that result in **Yellow** and green *are not simply Yy and yy (there may be something else going on)*

Goodness of Fit

Example: the results of a **Monohybrid cross** between a (heterozygous) yellow pea plant (Yy) and a green pea (yy) plant are as follows: **14 yellow** and **6** green. Are these results consistent with Mendel's first law (**segregation)** which should a 1:1 ratio in this case?

Step 2 Test Statistic:

assumptions: no category with expected freq < 1; no more than 20% categories have expected freq < 5

$$X^2 = (14-10)^2/10 + (6-10)^2/10 = 1.6 + 1.6 = 3.2$$

step 3 dof = # categories – 1 = 1

critical value that corresponds to 0.05 = 3.84

step 4 conclusion: since the critical value > value calculated from the X^2 test statistic, the ratio of progeny obtained is **consistent** with a 1:1 ratio