Outline

1. The chi-square test for proportions
   - Can humans generate random numbers?
   - The chi-squared statistic
   - Case of 2 categories
   - How to do it in R

Testing more than 1 proportion more than success/fail

We saw 2 ways to test a proportion with H_0: $p_{\text{success}} = p_0$:

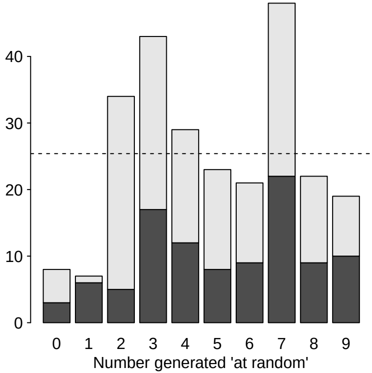
binomial test,

or the z-test (if np_0 and $n(1-p_0) > 5$).

What if more than 2 categories (success/fail)?

Example: online survey in Springs 2005 and 2006. Last question: “Select a number between 0 and 9 at random.”

Are humans able to pick numbers at random in an unbiased way, i.e. with same probability for each digit?



Example with one sample, 10 proportions

	0	1	2	3	4	5	6	7	8	9	total
2005	3	6	5	17	12	8	9	22	9	10	101
2006	5	1	29	26	17	15	12	26	13	9	153
combined	**8**	**7**	**34**	**43**	**29**	**23**	**21**	**48**	**22**	**19**	254
sample proportions	.031	.028	.134	.169	.114	.091	.083	.189	.087	.075	

Do humans have a bias, i.e. do not select all digits with the same probability? Do we have evidence against

H_0: $p_0 = p_1 = p_2 = p_3 = \cdots = p_9 = 1/10 = 0.1$
versus H_A: at least one $p_i \neq 0.1$?

The binomial or z-test can help us with a single question, like $p_0 = 0.1$ vs. $p_0 \neq 0.1$, but cannot handle more complex questions from > 2 categories. The chi-square test can!

Chi-square test, goodness of fit

	0	1	2	3	4	5	6	7	8	9	total
observed	8	7	34	43	29	23	21	48	22	19	254
expected under H_0	25.4	25.4	25.4	25.4	25.4	25.4	25.4	25.4	25.4	25.4	254
$\frac{(O_i - E_i)^2}{E_i}$	11.9	13.3	2.9	12.2	.5	.2	.8	20.1	.5	1.6	64.0

Expected counts: $E_i =$ Row total $* p_i = np_i$

Test statistic: measure of similarity

$$X^2 = \sum_{\text{all cells}} \frac{(\text{obs} - \text{exp})^2}{\text{exp}}$$

(use counts, not proportions)

Here: $x^2 =$

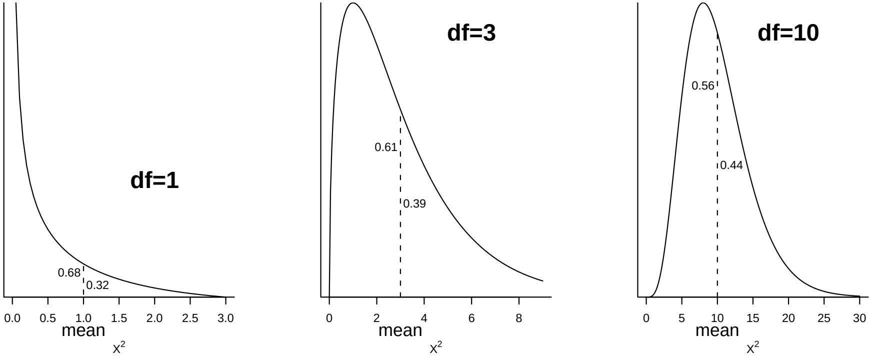
$= 11.9 + 13.3 + \cdots + 1.6 = 64.0$

The χ^2 distribution

$X^2 = 0$ when data and H_0 in perfect agreement ($O_i = E_i$)

If H_0 is true, X^2 close to 0: it has a χ^2 distribution on $10 - 1 = 9$ degree of freedom here.

If H_A is true, X^2 will be bigger. More extreme = larger



benchmark: $X^2 \leq$ df supports H_0.

Null distribution, p-value and conclusion

Why 9 df? Once 9 of the probabilities are fixed, no freedom for the 10th probability.

degree of freedom for chi-square test, goodness-of-fit

with c categories, $df = c - 1$.

p-value $= \mathbb{P}\{\chi^2_9 \geq 64.0\} = 2.225e - 10 < 0.001$
Use the chi-square table A (p.669-671):

df	...	0.05	.025	0.01	.005	.001
1	...	3.84	5.02	6.63	7.88	10.83
2	...	5.99	7.38	9.21	10.6	13.82
...						
9	...	16.92	19.02	21.67	23.59	27.88

Conclusion: very strong evidence that humans have a bias when picking a digit at random.

Chi-square test: facts and assumptions

For 2 categories (success/fail) and one proportion:

the chi-square test is equivalent to the z-test

The chi-square distribution is only an approximation to the true null distribution, i.e. the p-value obtained is an approximation to the true p-value.

just like the z-test uses the normal to approximate the binomial.

Assumption, for the validity of the χ^2 test

n is large enough, for all expected counts np_i to be > 5.

Always double-check.

here: the smallest E_i's was 25.4. ✓

chisq.test() in R gives a warning if some E_i's are < 5.

other software of calculators may not...

Chi-square test, goodness of fit, with 1 proportion

Recall missing sons,
H_0: $p = 0.51$.

	male	female	total
O_i	30	57	87
E_i	44.4	42.6	87

Expected counts:
$E_i =$ Row total $* p_i = np_i$
Test statistic:

$$X^2 = \sum_{\text{all cells}} \frac{(\text{obs} - \text{exp})^2}{\text{exp}}$$

Here: $x^2 =$ $= 4.65 + 4.84 = 9.50$

More extreme is so p-value $=$

Chi-square test, goodness of fit, with 1 proportion

Use the chi-square table (p.669-671) for $p = \mathbb{P}\{X^2 \geq a\}$:

df	...	0.05	.025	0.01	.005	.001
1	...	3.84	5.02	6.63	7.88	10.83
2	...	5.99	7.38	9.21	10.6	13.82
...						
9	...	16.92	19.02	21.67	23.59	27.88

We get: $<$ p-value $<$ at df=1.

Or use R:

```
> pchisq(9.50, df=1, lower.tail=FALSE)
[1] 0.002057
```

Notice that $x^2 = 9.50 = (-3.082)^2$ where $z = -3.082$ is what we had from the z-test. Equivalent tests: same p-value.

One-sided alternative for χ^2 test

Appropriate only with 2 categories.

Here: H_A: $p_{\text{son}} < 0.51$.

First check that the data goes in is the same direction as H_A.

if not: p-value > 0.50, accept H_0.

if so: p-value $= 1/2$ that found from the table.

Here: $\hat{p}_{\text{son}} = 30/87 = 0.34$ goes in the direction of H_A.
p-value $= 1/2 * \mathbb{P}\{X^2 > 9.498\} = .001$.
Strong evidence that $p_{\text{son}} < 0.51$ for male radiologists.

χ^2 test in R: chisq.test()

```
> dat = c(8,7,34,43,29,23,21,48,22,19)
> x2 = sum((dat-25.4)^2/25.4)
> x2
[1] 64.0315
> 1 - pchisq( 64.0315, df=9)
2.225407e-10

> chisq.test(dat)

        Chi-squared test for given probabilities

data:  dat
X-squared = 64.0315, df = 9, p-value = 2.225e-10

> chisq.test( c(30,87-30), p=c(.51,.49) )

        Chi-squared test for given probabilities
data:  c(30, 87 - 30)
X-squared = 9.4979, df = 1, p-value = 0.002057
```

χ^2 test in R: prop.test() for success/fail

```
> prop.test(30,87,0.51)

        1-sample proportions test with continuity correction

data:  30 out of 87, null probability 0.51
X-squared = 8.8485, df = 1, p-value = 0.002933
alternative hypothesis: true p is not equal to 0.51
95 percent confidence interval:
 0.2482987 0.4552176
sample estimates:
        p
0.3448276

> prop.test(30,87,0.51, correct=F)

        1-sample proportions test without continuity correction

data:  30 out of 87, null probability 0.51
X-squared = 9.4979, df = 1, p-value = 0.002057
alternative hypothesis: true p is not equal to 0.51
95 percent confidence interval:
 0.2534266 0.4493523
sample estimates:
        p
0.3448276
```

key R commands for proportions: recap

```
binom.test(y, n, p0)
binom.test( c(y,n-y),  p0)

prop.test(y, n, p0, alternative="greater")

chisq.test(c(n1,n2,...,nk), p=c(p1,p2,...,pk))
                                       # for k categories
```

default probability for H_0, if not specified, is

$p_0 = 0.50$ with 2 categories, or

$(1/k, \dots, 1/k)$ with k categories.

default alternative is `"two-sided"`