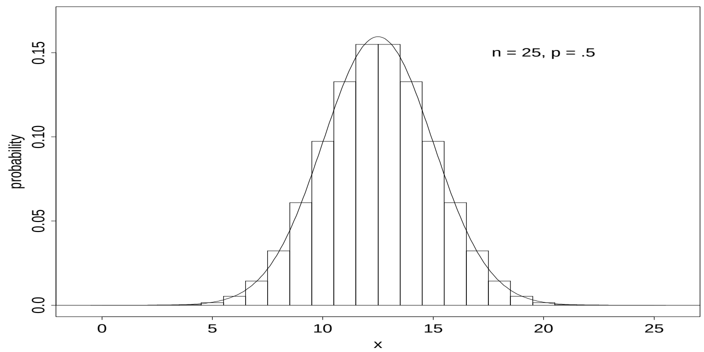
Inferences about proportions: the binomial distribution

Readings: Ch 22
Review: Ch 17, 19, 20, 21

- So far we have talked about continuous variables, which can be modeled using a density function.
- The t distributions was used to make inferences about one or more means.
- In many cases the data consists of counts.
- When there are only two possible outcomes and the interest is in the number of times one of these outcome occurs, the binomial distribution is the appropriate model.
- Suppose there are n independent trials, and call the two possible outcomes "success" and "failure", denoted S and F.
- Assume that the probability of success $P(S) = p$, is the same for each trial, and let X be the number of successes in the n trials.
- Then the binomial probability function is

$$p(x) = \binom{n}{x} p^x (1-p)^{(n-x)}$$

- The term $p^x(1-p)^{(n-x)}$ is the probability of x successes and $n-x$ failures in any particular order.
- The binomial coefficient $\binom{n}{x}$ is the number of possible orders.
- This distribution has mean $\mu = np$ and variance $\sigma^2 = np(1-p)$.
- If both np and $n(1-p)$ are greater than 10, the binomial can be approximated by the normal distribution with the same mean and variance.
- The figure shows the good agreement for $n = 25$ and $p = .5$



Testing for a single population proportion

- Inferences about p are based on the observed proportion $\hat{p} = X/n$, which has an approximate normal distribution with mean p and variance $\sigma^2 = p(1-p)/n$.

- Using the normal approximation, a $(1-\alpha)100\%$ confidence interval for p is given by
$$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$
where $z_{\alpha/2}$ is the value of the standard normal which has probability $\alpha/2$ to the right.

- To test $H_0 : p = p_0$ we use the test statistic
$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} = \frac{X - np_0}{\sqrt{np_0(1-p_0)}}$$

- The alternative hypothesis can be one-sided or two-sided, and this will affect the way the P value is calculated.

- Note that the null value for p, p_0, is used in the denominator of the test statistic because in testing we use the distribution of the test statistic when the null hypothesis is true.

Comparing two proportions

- We are often interested in comparing two binomial probabilities.

Example: To test the effectiveness of a new pain-relieving drug, 80 patients at a clinic were given a pill containing a drug and 80 others were given a placebo. In the first group, 56 of the patients showed improvement. In the second group, 38 of the patients showed improvement. How effective is the drug compared to the placebo?

Group	n	X	$\hat{p}$
Drug	80	56	.70
Placebo	80	38	.475

1. construct a confidence interval for $p_1 - p_2$
2. test the hypothesis $H_0 : p_1 = p_2$, or $p_1 - p_2 = 0$

- We require confidence intervals and tests for $p_1 - p_2$, the difference in two proportions.
- We assume independent random samples from two populations.
- The following notation is used

Population		Sample		
no.	probability	size	count	proportion
1	p_1	n_1	X_1	$\hat{p}_1 = X_1/n_1$
2	p_2	n_2	X_2	$\hat{p}_2 = X_2/n_2$

- The difference between the sample proportions

$$\hat{p}_1 - \hat{p}_2$$

is the natural estimate of $p_1 - p_2$.

- The variance of $\hat{p}_1 - \hat{p}_2$ is

$$\sigma^2_{\hat{p}_1 - \hat{p}_2} = \sigma^2_{\hat{p}_1} + \sigma^2_{\hat{p}_2}$$

or

$$\sigma^2_{\hat{p}_1 - \hat{p}_2} = \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}.$$

- $\hat{p}_1 - \hat{p}_2$ is approximately normally distributed if n_1 and n_2 are large and p_1 and p_2 are not too close to 0 or 1.
- For a confidence interval, substitute observed proportions in the standard error

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

- The confidence interval is

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} SE(\hat{p}_1 - \hat{p}_2)$$

or

$$\hat{p}_1 - \hat{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}.$$

Example: Pain clinic. Recall the data

Group	n	X	$\hat{p}$
Drug	80	56	.70
Placebo	80	38	.475

- The estimates are $\hat{p}_1 = 56/80 = .7$ and $\hat{p}_2 = 38/80 = 0.475$.

- The standard error for confidence intervals is

$$SE(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{.7(1-.7)}{80} + \frac{.475(1-.475)}{80}} = .0758$$

- The approximate 90% confidence interval for $p_1 - p_2$ is

$$.7 - .475 \pm 1.645(.0758)$$

  or

$$.225 \pm .1247$$

  or

$$(.1003, .3497).$$

- This corresponds to a two-sided test at the $\alpha = .10$ level. Since the 90% CI does not contain 0, so p-value < 0.1.

- We can also estimate the 95% or 99% confidence interval for $p_1 - p_2$, which correspond to $\alpha = 0.05$ and $\alpha = .01$ levels, respectively, for a two-sided test.

- For instance, for a 99% confidence interval:

$$.7 - .475 \pm 2.576(.0758)$$

  or

$$.225 \pm .1953$$

  or

$$(.0297, .4203).$$

- Since the 99% CI still does not contain 0, so p-value < 0.01.

Hypothesis test for two proportions

- The test for

$$H_0 : p_1 = p_2 \quad (p_1 - p_2 = 0)$$

versus

$$H_a : p_1 > p_2, \qquad H_a : p_1 < p_2$$

or

$$H_a : p_1 \neq p_2 \quad (p_1 - p_2 \neq 0)$$

uses the test statistic

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{SE_p(\hat{p}_1 - \hat{p}_2)}.$$

- The standard error in the denominator is calculated using the *pooled* estimate of the population proportion p

$$\hat{p}_p = \frac{X_1 + X_2}{n_1 + n_2}$$

  - This is appropriate because the proportions are assumed equal under H_0.

- Assuming that $p_1 = p_2 = p$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})},$$

so the standard error for testing is

$$SE_p(\hat{p}_1 - \hat{p}_2) = \sqrt{\hat{p}_p(1-\hat{p}_p)(\frac{1}{n_1} + \frac{1}{n_2})}.$$

- The normal approximation is used to calculate P-values

- The number of successes and failures in each group should be at least 5.

- The test statistic is

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}_p(1-\hat{p}_p)(\frac{1}{n_1} + \frac{1}{n_2})}}$$

Example: The drug minoxidil has been approved for the treatment of male pattern baldness. A clinical trial was carried out, in which some patients were given the drug, and others were given an identical looking placebo. The treatment group had $n_1 = 310$ subjects, of whom 99 demonstrated new hair growth. The control group had $n_2 = 309$ subjects and 62 of them had new hair growth. Does this study suggest that minoxidil is effective in promoting new hair growth?

- The hypotheses are $H_0 : p_1 = p_2$ versus $H_a : p_1 > p_2$.
- The pooled estimate is

$$\hat{p_p} = \frac{99 + 62}{310 + 309} = .2601.$$

- The estimated standard error of the difference is

$$SE_p(\hat{p}_1 - \hat{p}_2) = \sqrt{\hat{p_p}(1 - \hat{p_p})(\frac{1}{n_1} + \frac{1}{n_2})}$$

$$\begin{aligned} &= \sqrt{.2601(1 - .2601)(\frac{1}{310} + \frac{1}{309})} \\ &= \sqrt{.1924(0.0065)} \\ &= .0353. \end{aligned}$$

- The test statistic is

$$Z = \frac{99/310 - 62/309}{.0353} = 3.3662.$$

- The P value is $P(Z > 3.3662) = .0004$ using the computer.
- Using the standard normal tables we find the value 3.3 in the left column, and go across the row to the column labelled .07.
- There we find the probability .9996.
- We subtract this from 1 to get $P = .0004$.
- We therefore have very strong evidence against the null hypothesis of no difference due to the drug.
- The drug effect is statistically significant using $\alpha = .05$ and with $\alpha = .01$.

Paired binary samples
McNemar's test is appropriate in the case of paired binary data.
You will not be responsible for McNemar's test.

- In some sitiuations samples are not independent, and the methods described above cannot be used.
- Pairing arises with matched controls, siblings, and repeated trials on the same subjects, for example.

Example: A study was carried out to determine whether marijuana users had the same difficulty sleeping as matched controls, with the following results:

		Marijuana group		
		+	-	Total
Control	+	4	9	13
group	-	3	16	19
Total		7	25	32

- Note that there are 64 subjects in 32 pairs in this study.
- Each pair appears once in the table, so for example there were 3 pairs where the marijuana user slept well but the matched control did not.
- The hypotheses are the same as before

$$H_0 : p_1 = p_2 \quad (p_1 - p_2 = 0)$$

versus

$$H_a : p_1 > p_2, \qquad H_a : p_1 < p_2$$

or

$$H_a : p_1 \neq p_2 \quad (p_1 - p_2 \neq 0)$$

- The analysis uses only the data for pairs which give a different result.
- These are the off-diagonal entries in the table. (3 and 9 in the example)
- If there is no difference in the probability of "Success" in the two groups, the two off-diagonal entries should be close together.

- Denote the table entries as follows

		Group 1		
		S	F	Total
Group	S	a	b	a+b
2	F	c	d	c+d
Total		a+c	b+d	N

- The test amounts to a test of $H_0 : p = .5$ versus $H_a : p \neq .5$ (or a one-sided alternative) for the $n = b + c$ trials with response $X = b$.
- From page 2 of these notes, the test statistic is

$$\begin{aligned} Z &= \frac{b - n/2}{\sqrt{n/4}} = \frac{2b - n}{\sqrt{n}} \\ &= \frac{b - c}{\sqrt{b + c}} \end{aligned}$$

- The P value is calculated from the standard normal tables.
- In the literature, one often sees the squared statistic

$$X^2 = \frac{(b - c)^2}{b + c}$$

which is compared to a chi-squared (χ^2) distribution with 1 degree of freedom.

- This version is not as good for one-sided alternatives.
- To use the normal or χ^2 distributions here, one should have $b + c > 10$.

Example: Marijuana

- Here $Z = (9 - 3)/\sqrt{12} = 1.73$.
- Using the normal table, we find the row 1.7 and column .03, which gives probability .9582.
- This is the probability to the left of 1.73.
- We subtract from 1 to get the right tail probability, and multiply by 2 because we are using a two-sided alternative, to get $P = 2(1 - .9582) = .0836$.

- We conclude there is only weak evidence against the null hypothesis of no difference between sleeping difficulty in the marijuana and matched control groups.
- If we used the squared statistic $X^2 = 3.000$, we compare to the first row of the χ^2 tables.
- Our value is greater than 2.706 and less than 3.841, so we find $.05 < P < .10$.
- The P value is not as precisely determined, but the conclusion is the same.
- Our results are statistically significant at the $\alpha = .10$ level but not at the $\alpha = .05$ level.