Chapter 10: Analysis of Categorical Data

1. Binomial Random Variables
2. Inference for a Single Proportion π (Large and Small Samples)
3. Power and sample size for a Single Proportion π
4. Inference for Comparing Two Proportions (Large Samples)
5. Power and sample size calculations for comparing Proportions
6. McNemar's test for comparing Paired Proportions
7. Chisquare Goodness of Fit Test
8. Chisquare test for Contingency Tables and Fisher's Exact Test
9. Odds Ratio
10. Three-way tables and "Simpson's Paradox"
11. Breslow-Day and Mantel-Haenszel Tests
12. The Poisson distribution

Chapter 10 Examples

1. Maize Example: Chisquare goodness of fit test
2. Chisquare Tests for Contingency Tables and FET
3. Odds Ratio Examples
4. Gender Discrimination at Berkeley
5. Drug clinic three-way table
6. Mulekick Example: Poisson GOF

0. Some Perspective

In CH5-9, we focused on inference for means and standard deviations. **Mean and standard deviation** are parameters for center and spread for a **numerical variables** (ex: hormone level, lead consumption).

In CH10, we focus on **categorical variables** (ex: infested or not infested, cold or no cold). In this case, the parameter of interest is **proportion**.

CH5-9: Normal, t, $\chi 2$, F distributions

- All continuous distributions

CH10: Binomial and Poisson distributions

- Both discrete distributions
- BUT sometimes these discrete distributions are approximated by continuous distributions!

1. Binomial Random Variables

Consider an experiment that has only two outcomes (a binary response).
Examples: 0 or 1, Heads or Tails, Event or No Event, Success or Failure

Let π = P(observe a 1) ; then $1-\pi$ = P(observe a 0).
Note: π represents a probability between 0 and 1. ($\pi \neq 3.14159$ here.)

Repeat the experiment **n** times **independently**
Let Y = number of 1's observed out of the n trials
= sum of the outcomes of the n trials

An experiment involving n repeats is called a **binomial experiment**, and Y is called a **binomial random variable.** For $Y \sim \text{bin}(n, \pi)$

$$P(Y = y) = \frac{n!}{y!(n-y)!}(\pi)^y(1-\pi)^{(n-y)}$$

For y = 0, 1, …, n
mean(Y) = $\mu_Y = n\pi$ and var(Y) = $\sigma_Y^2 = n\pi(1-\pi)$

Note n! = n·(n-1)·(n-2)…1, and 0! defined as 1.

Example :

n = 4 (4 trials)

π = 0.6 (probability of event)

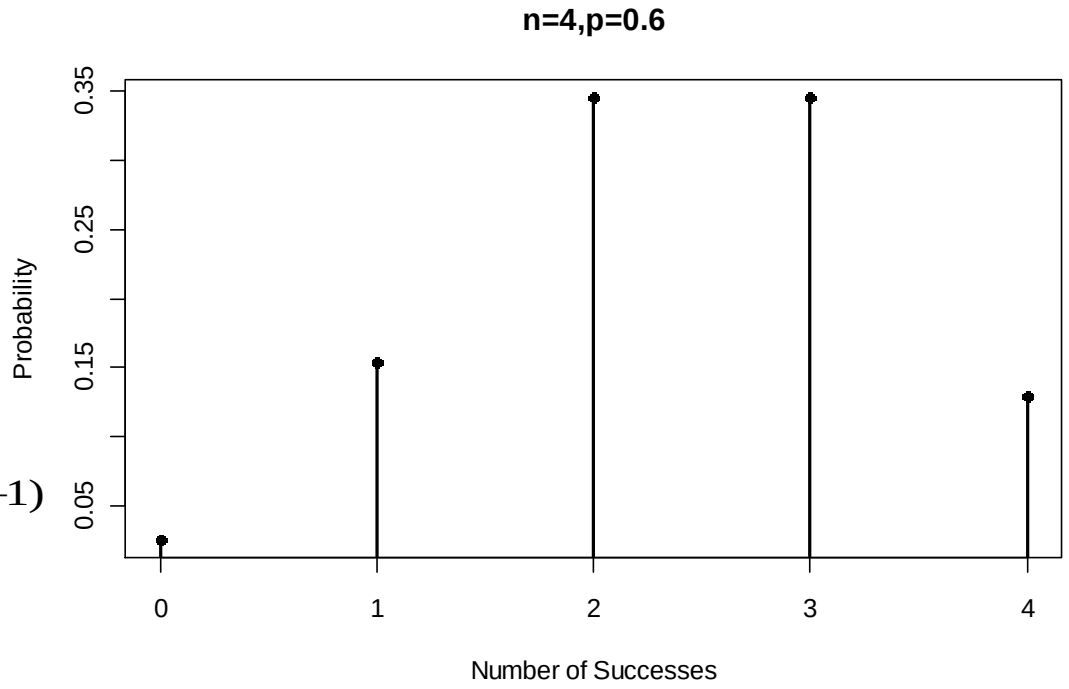
y = 1 (total # of events)

$$P(Y=1)$$

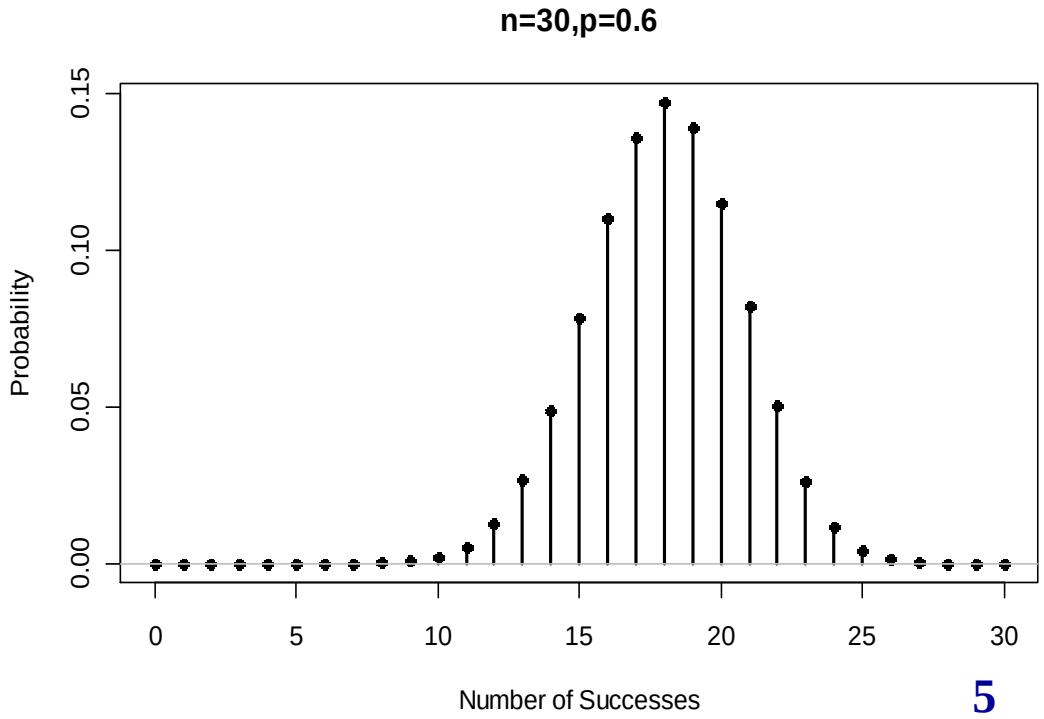
$$=\frac{4!}{1!(4-1)!}(0.6)^1(1-0.6)^{(4-1)}$$

$$=0.1536$$

$$\text{Note}: 0!=1 \quad \text{and } y^0=1$$



With n=30 and π=0.6 the distribution looks close to normal.



Cumulative distribution functions gives the probability of being less than or equal to a value.

Example: For the Binomial distribution

$$P(Y \leq y) = \sum_{k=0}^{y} \frac{n!}{k!(n-k)!} (\pi)^k (1-\pi)^{(n-k)}$$

NOTE: Be careful about inequalities!

P (Y < y) ≠ P (Y ≤ y) !

P(Y ≤ 2) = P(Y=0) + P(Y=1) + P(Y=2)

P(Y < 2) = P(Y=0) + P(Y=1)

= P(Y ≤ 1)

Binomial Probabilities in R and Rcmdr

- In R, `dbinom()` gives $P(Y=y)$ or `pbinom()` gives $P(Y \leq y)$.
- In Rcmdr, use Distributions > Discrete Distributions > Binomial Distribution.

  Choose Binomial Probabilities (ex: $P(Y=5)$) or Binomial Tail Probabilities (ex: $P(Y \leq 5)$).

Example: Suppose $n=15$ and $\pi=0.5$.

```
> dbinom(5, size = 15, prob = 0.5)
[1] 0.09164429
> pbinom(5, size = 15, prob = 0.5)
[1] 0.1508789
> pbinom(5, 15, 0.5) - pbinom(4, 15, 0.5)
[1] 0.09164429
```

Mean and standard deviation of a Binomial RV

The mean and standard deviation of a population of outcomes from imaginary replications of the experiment.

The mean and standard deviation of a binomial RV

If Y~bin(n,π) then

$$\mu_Y = n\pi \quad \text{and} \quad \sigma_Y = \sqrt{n\pi(1-\pi)}$$

Example: n=4 π=0.6

$$\mu_Y = n\pi = (4)(0.6) = 2.4$$

$$\sigma_Y = \sqrt{n\pi(1-\pi)} = \sqrt{(4)(0.6)(1-.6)} = 0.98$$

Normal Approximation to the Binomial

(an example of the Central Limit Theorem)

Let X be the result of a single trial with probability of success π.
Let X = 1 in case of a success and 0 in case of failure.

Observe n independent trials $X_1, X_2, \ldots, X_n$.

By the central limit theorem, the sample mean $\overline{X}$ has a distribution that is approximately normal, for large n.

Let: $Y = \sum X_i = \#successes$

$$= n\left(\sum X_i / n\right) = n\overline{X}$$

Y is a binomial RV, but Y is just a multiple of $\overline{X}$ and $\overline{X}$ is approximately normal, so Y must also be approximately normal.

Example: See binomial distribution with n=30 and π=0.6 (slide 5).

Normal Approximation to the Binomial *(Example)*

Example: Give a dose of insecticide to 30 flies. Assuming that the probability that an individual fly dies is 0.7, what is the probability that 25 or more die? $P(Y \geq 25) = 1 - P(Y < 25) = 1 - P(Y \leq 24)$

Let Y represent the number of dead flies. Then Y ~ bin (n=30, π=0.7). (Here an "event" is a dead fly.)

EXACT binomial probability using R:
1 - pbinom(24, size = 30, prob = 0.7) = 0.0766

Normal Approximation (without Continuity Correction):

$$\mu_Y = n\pi = 30(0.7) = 21$$
$$\sigma_Y = \sqrt{n\pi(1-\pi)} = \sqrt{30(0.7)(1-0.7)} = 2.51$$

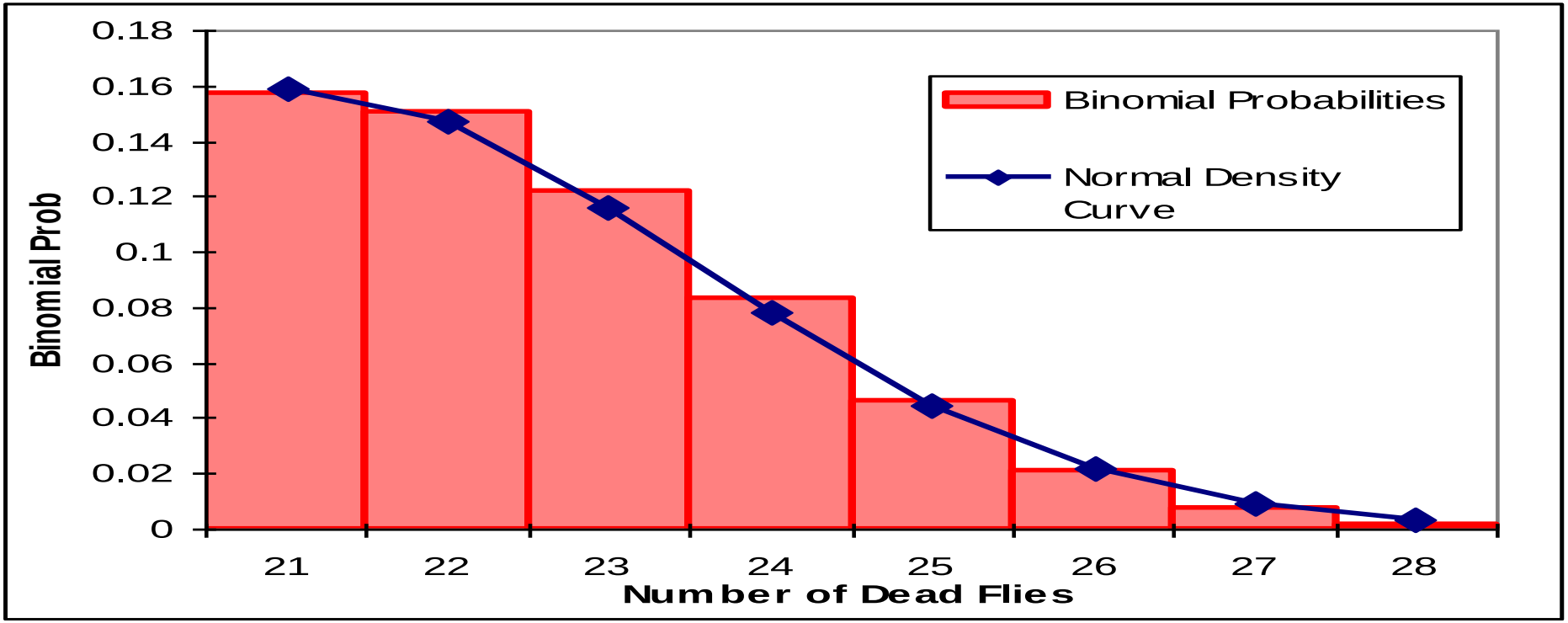
$$P(Y \leq 24) = P\left(\frac{Y-21}{2.51} \leq \frac{24-21}{2.51}\right)$$

$$\cong P(Z \leq 1.20) = 0.8849$$

$$\text{Therefore: } P(Y \geq 25) \cong 1 - .8849 = 0.1151$$

Normal Approximation to the Binomial *(continuity correction)*

Continuity Correction: Use 24.5 in place of 24 to include whole 24 box. The idea extends to other situations.



$$P(Y \le 24.5) = P\left(\frac{Y-21}{2.51} \le \frac{24.5-21}{2.51}\right)$$

$$\cong P(Z \le 1.39) = 0.9177$$

Therefore: $P(Y \ge 25) \cong 1-.9177 = 0.0823$

If we can calculate exact probabilities, **why are we talking about the normal approximation and continuity correction?**

Because we will see that many standard approaches for inference for proportions are based on a normal approximation.

Examples include large sample CI and Z-test for a single proportion and chi-square test for contingency tables.

In R, we will see that these methods will usually include a "continuity correction" by default.

We will not be too concerned about the exact details of the continuity correction because it is most important when sample sizes are small. And in those cases, we prefer (exact) small sample methods. For example, Fisher's Exact Test.

2. Inference about a Single Proportion π

Aphid Example #1: We observe 53 plots infested with Russian Wheat Aphids in 100 randomly selected plots in Larimer Co.

Let: Y = # of plots having aphids (event/success)
n = # trials
π = Prob (a randomly selected plot will have aphids)
= proportion of plots in the population with aphids.

The random variable Y is binomial, which is approx. normal (for large n).
Recall that:

$$\mu_Y = n\pi \quad \text{and} \quad \sigma_Y = \sqrt{n\pi(1-\pi)}$$

Objectives for inference:

1. Estimate π.
2. Large sample confidence interval for π (normal approximation).
3. Large sample hypothesis test about π (normal approximation).
4. Small sample hypothesis test about π (exact binomial).
5. Small sample confidence interval for π (exact binomial).

We use the sample proportion to estimate the population proportion:

$$\hat{\pi} = \frac{y}{n} = \frac{\#\text{ events}}{\#\text{ trials}}$$

The mean and the std. dev. of the distribution of $\hat{\pi}$ are given by:

$$\mu_{\hat{\pi}} = \pi \quad \text{and} \quad \sigma_{\hat{\pi}} = \sqrt{\frac{\pi(1-\pi)}{n}}$$

The mean of the estimator is the quantity that we are trying to estimate. (This property in any estimator is called "unbiasedness").

The standard deviation of $\hat{\pi}$ is

$$\sigma_{\hat{\pi}} = \sqrt{\frac{\pi(1-\pi)}{n}}$$

which is estimated by

$$\hat{\sigma}_{\hat{\pi}} = SE(\hat{\pi}) = \sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

Large sample confidence interval for π (Normal Approximation)

An approximate (1-α)100% confidence interval for π is:

$$\hat{\pi} \pm Z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

where the table value $Z_{\alpha/2}$ is determined from the Normal distribution

Assumptions: Random sample, independent observations, large sample size. (See discussion later in notes.)

95% CI for the Aphid Example:

$$\hat{\pi} = \frac{53}{100} = 0.53 \qquad 0.53 \pm 1.96\sqrt{\frac{(0.53)(1-0.53)}{100}}$$

$$0.53 \pm 0.098, \quad \text{i.e.,} \quad (0.432, 0.628)$$

Common Table Values from the Normal Distribution

Other values can be found using Table 1 (Normal) or the qnorm() function in R.
Another possibility: use Table 2 (t) with df=infinity!

α	CI	R Code	Z_α
0.005	99%	qnorm(0.995)	2.576
0.01	98%	qnorm(0.99)	2.326
0.025	95%	qnorm(0.975)	1.960
0.05	90%	qnorm(0.95)	1.645
0.10	80%	qnorm(0.90)	1.282

Large sample confidence interval for π
Formulas used by Ott & Longnecker and R

R and the book use (different) modifications of what is given in these notes. The differences will be most noticeable when sample size is small (but then we will want to use the exact binomial CI and test).

Book: $\tilde{y} = 0.5z_{\alpha/2}^2 \quad \tilde{n} = n + z_{\alpha/2}^2 \quad \tilde{\pi} = {}^{\tilde{y}}/_{\tilde{n}}$

$$\tilde{\pi} \pm z_{\alpha/2}\sqrt{\frac{\tilde{\pi}(1-\tilde{\pi})}{\tilde{n}}}$$

R prop.test(): $(\hat{\pi} + z^*) \pm z_{\frac{\alpha}{2}}\left(\sqrt{\frac{\hat{\pi}\,(1-\hat{\pi})}{n} + \frac{z^*}{2n}}\right)/(1 + 2z^*)$

where $z^* = (z_{\frac{\alpha}{2}}^2)/2\text{n}$

Question: How can we tell what formula R is using?
Answer: Help is not helpful, so we need to use the source!

Large sample hypothesis tests for π (Normal Approximation)

Assumptions: Random sample, independent observations**,** large sample size. (See discussion later in notes.)

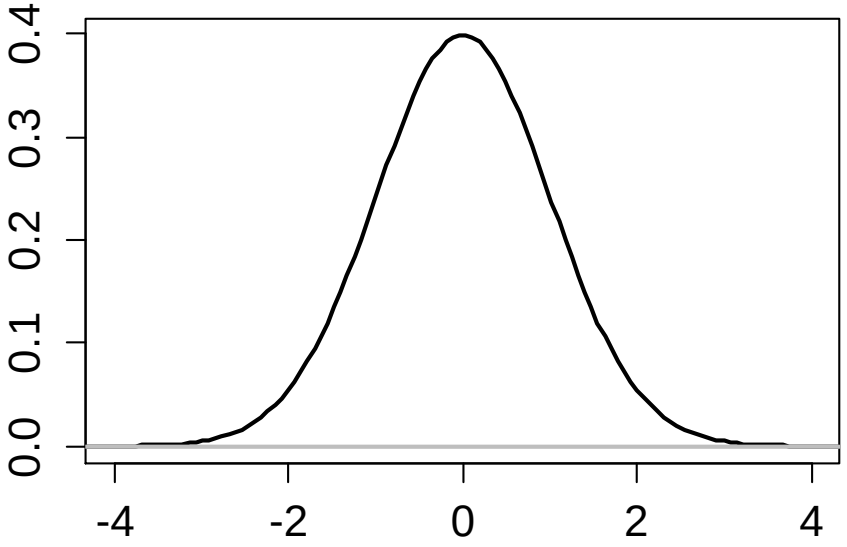
Test Statistic: $$z = \frac{\hat{\pi} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

Alternative Hypothesis:	**Rejection Region:**
(1) $H_A: \pi > \pi_0$	$z \geq z_\alpha$
(2) $H_A: \pi < \pi_0$	$z \leq -z_\alpha = z_{1-\alpha}$
(3) $H_A: \pi \neq \pi_0$	$\lvert z \rvert \geq z_{\alpha/2}$

P-values: (1) area to the right of z (test statistic), (2) area to the left, (3) double the area to the right of |z|.

For the Aphid Example:

$H_0: \pi \leq 0.5$ versus $H_A: \pi > 0.5$



Test Statistic:

$$z = \frac{0.53 - 0.50}{\sqrt{\frac{0.50(1-0.50)}{100}}} = \frac{0.03}{0.05} = 0.6$$

Rejection Rule:

Reject H_0 if $z > z_{0.05} = 1.645$

Decision: Since z= 0.6 is NOT > $z_{0.05}$ = 1.645, we Fail to Reject H_0. We cannot conclude that the population proportion of infested fields is greater than 0.5.

One-sided p-value: P(Z>0.6)=0.27

Large Sample Test & CI for Single Proportion in R

- In R, use `prop.test()`

For the Aphids Example #1 (n=100, y=53):

```
> prop.test(53, 100, p=0.5, alternative =
  "greater", correct = FALSE)

        1-sample proportions test without
  continuity correction

data:  53 out of 100, null probability 0.5

X-squared = 0.36, df = 1, p-value = 0.2743

alternative hypothesis: true p is greater than
  0.5

95 percent confidence interval:

 0.4481999 1.0000000

sample estimates:

   p

0.53
```

Notes on the Large Sample Normal Approximation

1. When forming the large sample CI use $\hat{\pi}$ to compute the standard error, but when testing use π_0 to compute the estimated standard error.
2. Due to #1, the set of π that would not be rejected by the large sample hypothesis test is slightly different from the set of π that are in the CI.
3. The X-squared statistic is equal to Z^2 where the formula for Z is given on slide 18. To get the sign of Z look at $\hat{\pi} - \pi_0$ (numerator of Z test statistic).
4. The formulas presented assume a large population. A finite population correction (FPC) is possible, but not covered here.

More detail about the Continuity Correction

1. Yates continuity correction is possible (default in R).
2. The idea of the continuity correction is to better approximate binomial distribution (discrete) with normal (continuous). However the effect of the continuity correction is most noticeable when the sample size is small (in which case we will use the exact binomial approach for testing and CI).
3. I am fine with the continuity correction, but I will not ask for (or show) a continuity correction when doing hand calculations. Hence, I use correct = FALSE here to more closely match the hand calculations.

How Large does the sample need to be in order to use the Large Sample Approximation?

Depends on both n and π. Need BOTH:

$$\hat{\pi} > 0 + 3\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

$$\hat{\pi} < 1 - 3\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

In other words, $\hat{\pi}$ should be more than 3SE away from 0 and 1.

Checking Sample Size for Aphid Example #1:

$$\hat{\pi} = \frac{53}{100} = 0.53 \qquad 3\sqrt{\frac{0.53(1-0.53)}{100}} = 0.15$$

$\hat{\pi} > 0.15$ AND $\hat{\pi} < 1 - 0.15 = 0.85$

Both conditions are satisfied, so the large sample normal approximation is adequate.

ME Reported for Polls in the News

Polls reported in the news should give the sampling dates, number of participants and margin of error.

Note that the formula for ME depends on the estimated proportion:

$$ME = Z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

However, often a single ME will be reported even if several different questions (with different estimates) were asked.

$$95\%ME = 1.96\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}} \leq 2\sqrt{\frac{0.5(1-0.5)}{n}} = \frac{1}{\sqrt{n}}$$

Hence, the news will often report the ME as 1/sqrt(n)!

Small sample hypothesis test for π (Exact Binomial)

Aphids Example #2: Suppose we want to test
$H_0: \pi \leq 0.25$ versus $H_A: \pi > 0.25$
Say that n=8 plots are sampled and of these y = 4 are found to be infested.

Test statistic is Y, which has a binomial distribution.
Under H_0, Y~bin(n=8, π=0.25)

P-value=$P(Y \geq y_{obs})$ In our example, $y_{obs}=4$.

p-value = $P(Y \geq 4) = 1 - P(Y \leq 3)$
= 1 - pbinom(3, size = 8, prob = 0.25)
= 0.114
Since p-value > α =0.05, we Fail to Reject H_0.

Exact binomial test can be run using `binom.test()`.
One and two-sided sided alternatives can be done.

Small sample Confidence Interval for π (Exact binomial)

An exact confidence interval for π can be obtained by using `binom.test()`.

The exact confidence interval is constructed as the set of values of π that are not rejected in the exact hypothesis test on the previous page.

This cannot be done easily by hand, because you have to test all possible π values. Use R.

If n = 8 and the observed response y = 4 then, a 95% CI is (0.157,0.843).

Small Sample Test & CI for Single Proportion in R

- In R, use `binom.test()`

For the Aphids Example #2 (n=8, y=4):

```
> binom.test(4,8,p=0.25,alternative="greater")

        Exact binomial test

data:  4 and 8
number of successes = 4, number of trials = 8,
  p-value = 0.1138
alternative hypothesis: true probability of
  success is greater than 0.25
95 percent confidence interval:
 0.1929029 1.0000000
sample estimates:
probability of success
                   0.5
```

Return to the Sign Test for the Median

In chapter 5, we used the sign test to make inference about a single median. The sign test is based on the binomial test of a single proportion.

Suppose we are testing H_0: $M \leq M_0$ vs H_A: $M > M_0$.
If the null hypothesis was true (M_0 was the population median) then we would expect 50% of the differences ($y_i - M_0$) to be positive.

To calculate the test statistic, we look at the sign of each difference ($y_i - M_0$). The test statistic (s) is the number of positive differences.

To calculate the p-value we compare the test statistic (s) to the binomial distribution with n = # observations, $\pi = 0.5$.

Return to the Sign Test for the Median

Example: Suppose we are testing H_0: $M \leq 2$ vs H_A: $M > 2$. Based on a sample of size n = 8, we find that s = 5 observations are greater than 2. Then the p-value is $P(Y \geq 5) = 1 - P(Y \leq 4)$.

```
> InData <- c(0, 1, 1, 3, 5, 7, 9, 10)
> 1 - pbinom(4, size = 8, prob = 0.5)
[1] 0.3632813
> library(BSDA)
> SIGN.test(InData, md=2, alternative =
"greater")

        One-sample Sign-Test
data:  InData
s = 5, p-value = 0.3633
alternative hypothesis: true median is greater
than 2
```

3. Power and sample size for a single proportion π

In this section we consider a variety of objectives for planning experiments to study a single proportion. We consider large sample methods based on the normal approximation and the Z-test and Z-intervals, as well as small sample methods based on the exact binomial distribution:

Objectives:

A. Planning to achieve a desired CI width (large sample only)

B. Large sample power for a hypothesis test about π.

C. Small sample power for a hypothesis test about π.

3A. Sample Size based on CI width (Large Sample)

Recall that the (large sample) confidence interval for π has the form

$$\hat{\pi} \pm z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

Let E represent the desired ME of the confidence interval is:

$$E = z_{\alpha/2}\sqrt{\frac{\hat{\pi}(1-\hat{\pi})}{n}}$$

Solving this for n we get

$$\boxed{n = \frac{z^2_{\alpha/2}\hat{\pi}(1-\hat{\pi})}{E^2}}$$

We need a conjecture for $\hat{\pi}$ or we can substitute $\hat{\pi} = 0.5$ corresponding to a "worst case" scenario. The choice $\hat{\pi} = 0.5$ will give the largest possible sample size that may be needed, a conservative answer for the required sample size.

Example: Suppose we want to find the sample size required to achieve a 95% ME for π to be ≤ 0.10. We do not have a specific conjecture for π, so we use the "worst case" conjecture $\pi = 0.50$.

95%ME -> $Z_{\alpha/2} = 1.96$

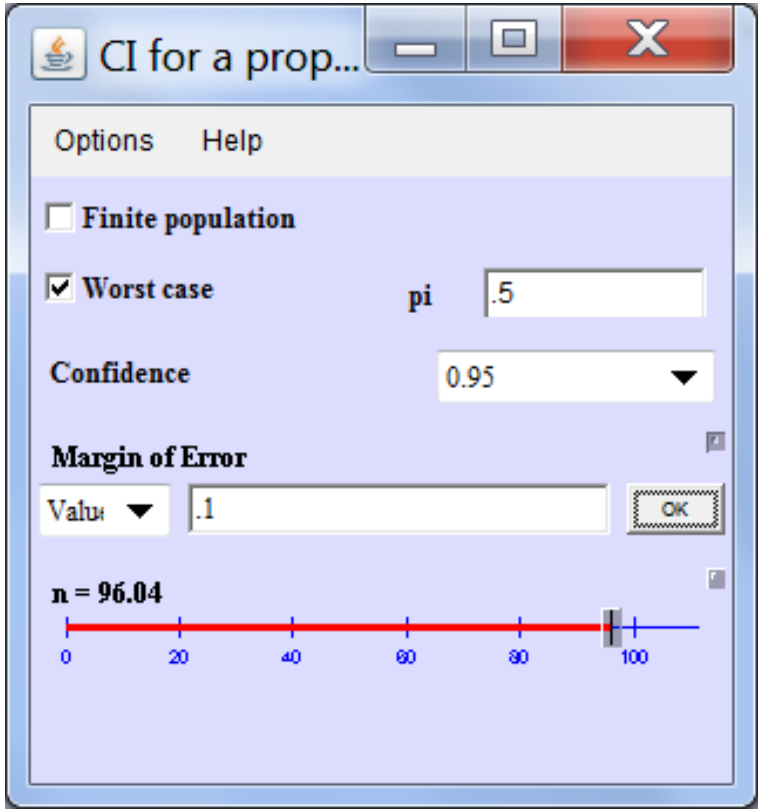
$$n = \frac{(1.96)^2(0.5)(1-0.5)}{(0.1)^2} = 96.04$$

Round up to n=97!

3A. Sample Size based on CI width using Lenth

http://homepage.stat.uiowa.edu/~rlenth/Power/

- Choose CI for one Proportion.
- Here we calculate the required sample size to achieve a 95% ME of 0.1
- We use the "worst case scenario" conjecturing that π=0.5



3B. Large Sample power for a hypothesis test about π

$H_0: \pi \le \pi_0$ versus $H_A: \pi > \pi_0$ (one-sided)

Test Statistic: $z = \dfrac{\hat{\pi} - \pi_0}{\sigma_\pi}$ where $\sigma_\pi = \sqrt{\dfrac{\pi_0(1-\pi_0)}{n}}$

Reject H_0 if $z > z_\alpha$

For a specific conjectured value of π, the distribution of Z is approximately normal with variance approximately 1.0 and mean:

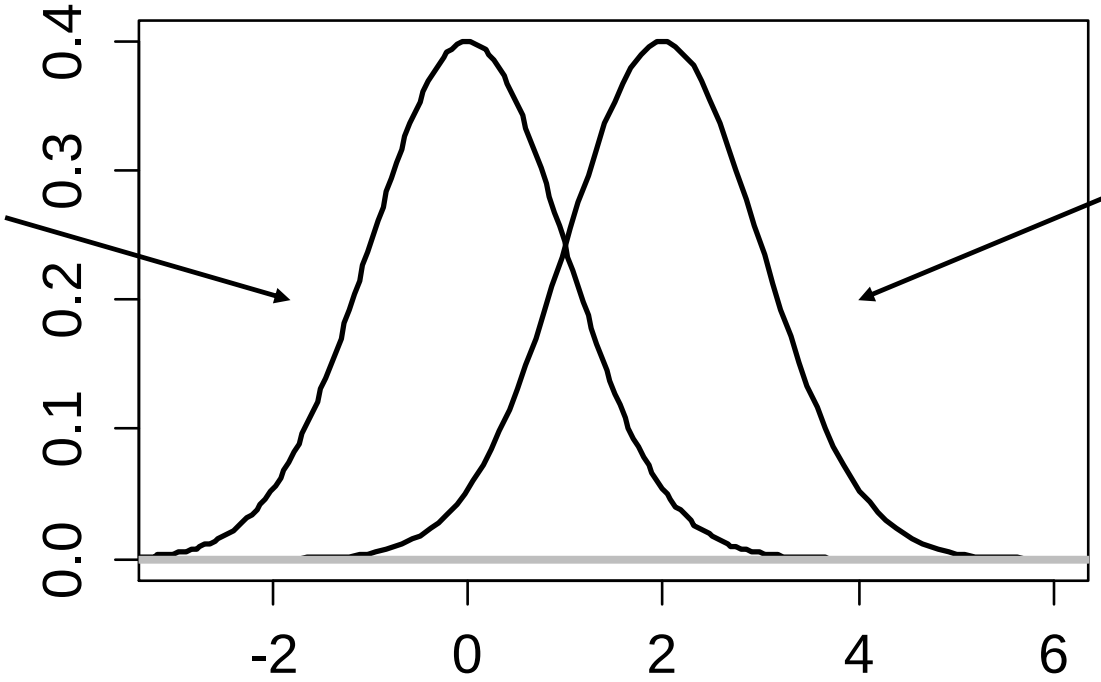
$$\lambda = \frac{\pi - \pi_0}{\sqrt{(\pi_0(1-\pi_0))/n}}$$

We can use this information to calculate power.

Example: Suppose we want to calculate power for testing H_A: $\pi > 0.5$ using n = 100 and $\alpha = 0.05$. (1) Hence we Reject H0 if $Z > Z_\alpha = 1.645$. (2) We conjecture that π=0.6.

$$\lambda = \frac{0.6 - 0.5}{\sqrt{(0.5(1-0.5))/100}} = 2$$

Distbn of Z when H_0 is true



Distbn of Z when H_A is true

Calculate power "by hand" using R:

1-pnorm(1.645, mean = 2, sd = 1) = 0.6388

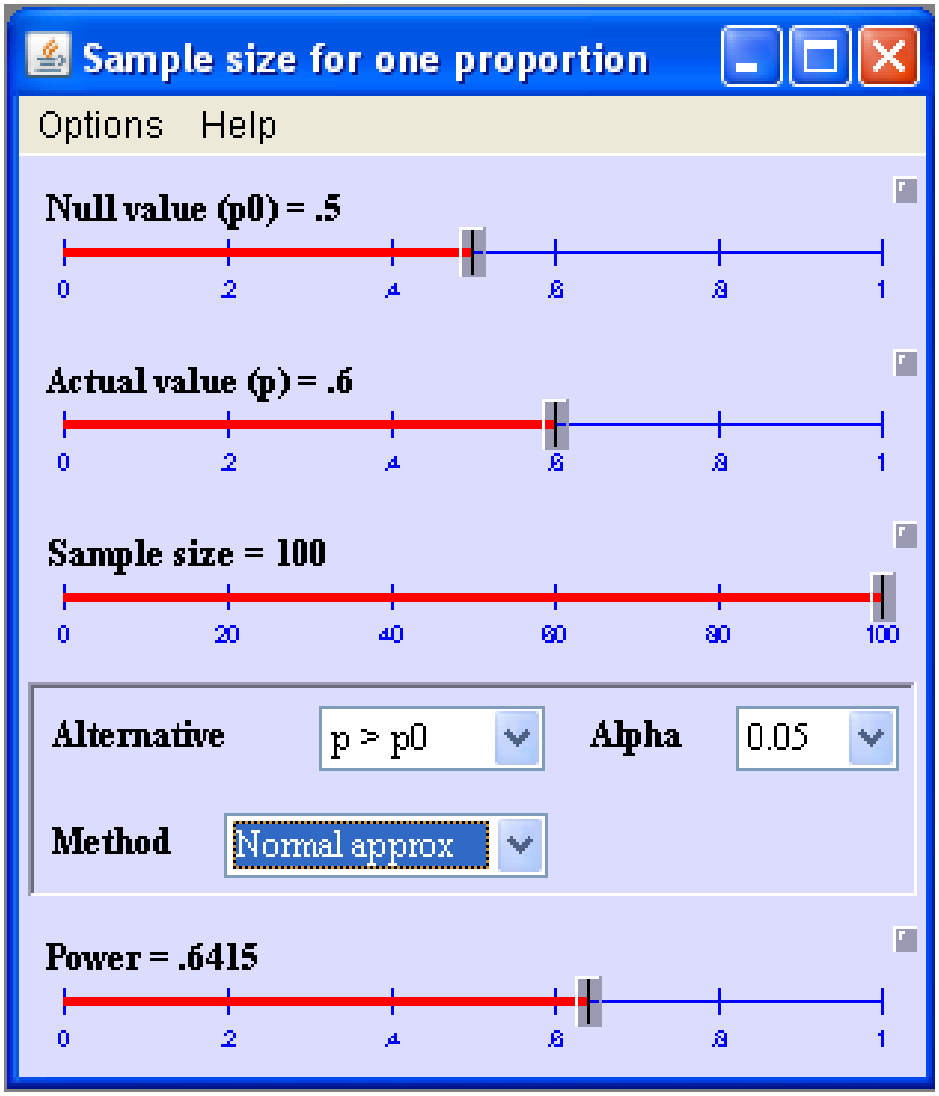
3B. Power for Large Sample Test about π using Lenth

http://homepage.stat.uiowa.edu/~rlenth/Power/

- Choose Test of one Proportion.

- Here we calculate power for HA: $\pi > 0.5$ using n=100 based on a conjectured value of $\pi = 0.6$.

- Since n=100 the large sample normal approximation is appropriate.



3C. Small Sample power for test about π

Example: Compute power for an exact binomial test with n=20:

$H_0: \pi \leq 0.25$ versus $H_A: \pi > 0.25$ (one-sided) ($\alpha = 0.05$)

Step 1: Define the Reject Region.

Reject H_0 when $Y \geq Y_{crit}$

Find Y_{crit} such that under H_0 (π_0=0.25), $P(Y \geq Y_{crit}) \leq \alpha = 0.05$.

Using pbinom() with n=20 and π=0.25:

P(Y≥9) = P(Y>8) = 0.0409

Therefore, we Reject H0 if Y≥9.

The test is slightly conservative, meaning α < 0.05.

k	P(Y≤k)	P(Y>k)
0	0.0032	0.9968
1	0.0243	0.9757
2	0.0913	0.9087
3	0.2252	0.7748
4	0.4148	0.5852
5	0.6172	0.3828
6	0.7858	0.2142
7	0.8982	0.1018
8	0.9591	0.0409
9	0.9861	0.0139
10	0.9961	0.0039
11	0.9991	0.0009
….		

Step 2: Use a specific conjectured value to calculate power.

For a conjectured value of π=0.5, we calculate the power as:

Power = $P(Y \geq Y_{crit}) = P(Y \geq 9)$

= 1- $P(Y \leq 8)$

= 1- pbinom(8, 20, 0.5)

= 0.7483

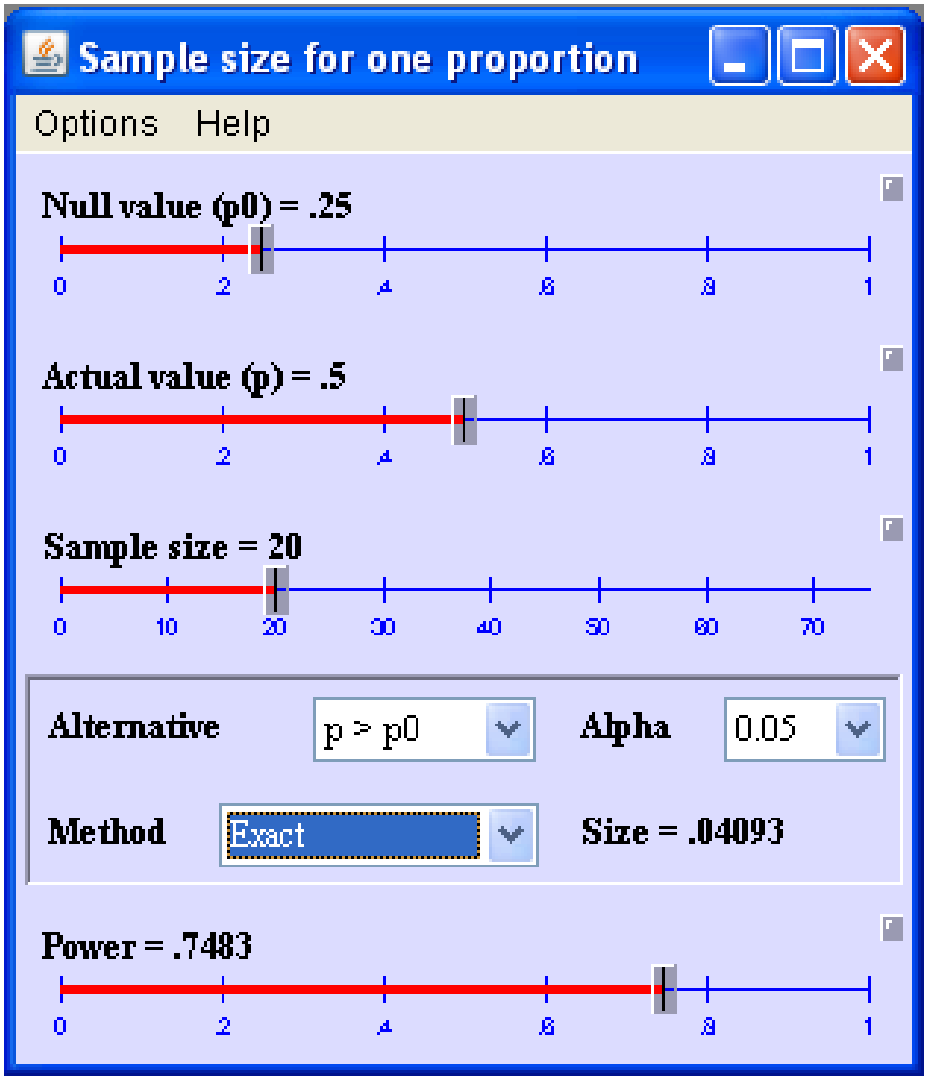
3C. Power for Small Sample Test about π using Lenth

http://homepage.stat.uiowa.edu/~rlenth/Power/

•Choose Test of one Proportion

•Here we calculate power for H_A: $\pi > 0.25$ using n=20 based on a conjectured value of $\pi = 0.5$.

•Since n=20 the small sample exact test is appropriate.



4. Large Sample Z test and CI Comparing Two Proportions π_1 and π_2

Example: Researchers were interested in examining the effect of Vitamin C for prevention of colds. 279 French skiers randomly divided into two groups. 139 are given Vitamin C, of these 17 developed colds. 140 given Placebo, of these 31 developed colds

	Cold	No Cold	Total
Vitamin C	17	122	139
Placebo	31	109	140

$\widehat{\pi_{VitC}} = 17/139$

$\widehat{\pi_P} = 31/140$

Objectives:

1. Estimate $\pi_1 - \pi_2$
2. A confidence interval for $\pi_1 - \pi_2$ (large samples).
3. A hypothesis test about $\pi_1 - \pi_2$ (large samples).
4. Sample size and power for $\pi_1 - \pi_2$ (large samples).

Estimation of $\pi_1 - \pi_2$

$$\hat{\pi}_1 = \frac{y_1}{n_1} \quad \text{and} \quad \hat{\pi}_2 = \frac{y_2}{n_2}$$

So we estimate $\pi_1 - \pi_2$ by $\hat{\pi}_1 - \hat{\pi}_2 = \frac{y_1}{n_1} - \frac{y_2}{n_2}$

Std. dev$(\hat{\pi}_1 - \hat{\pi}_2)$ is

$$\sigma_{\hat{\pi}_1 - \hat{\pi}_2} = \sqrt{\frac{(\pi_1)(1-\pi_1)}{n_1} + \frac{(\pi_2)(1-\pi_2)}{n_2}}$$

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

Large Sample Confidence Interval for $\pi_1 - \pi_2$

(Normal Approximation)

An approximate (1-α)100% confidence interval for $\pi_1 - \pi_2$ is:

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

where the table value $z_{\alpha/2}$ is determined from the Normal distribution

Assumptions: Independent random samples, large sample sizes.

In the French skier example:

Vit C group	Placebo group
$n_1=139$	$n_2=140$
$y_1=17$	$y_2=31$
π_1=Prob(cold)	π_2=Prob(cold)

$$\hat{\pi}_1 - \hat{\pi}_2 = \frac{17}{139} - \frac{31}{140} = 0.122 - 0.221 = -0.099$$

95% Confidence Interval:

$$0.122 - 0.221 \pm 1.96\sqrt{\frac{(0.122)(1-0.122)}{139} + \frac{(0.221)(1-0.221)}{140}}$$

$$-0.099 \pm 0.088$$

$$(-0.187, -0.011)$$

Large Sample Hypothesis Test for $\pi_1 - \pi_2$

(Normal Approximation)

Assumptions: Independent random samples with large sample sizes.

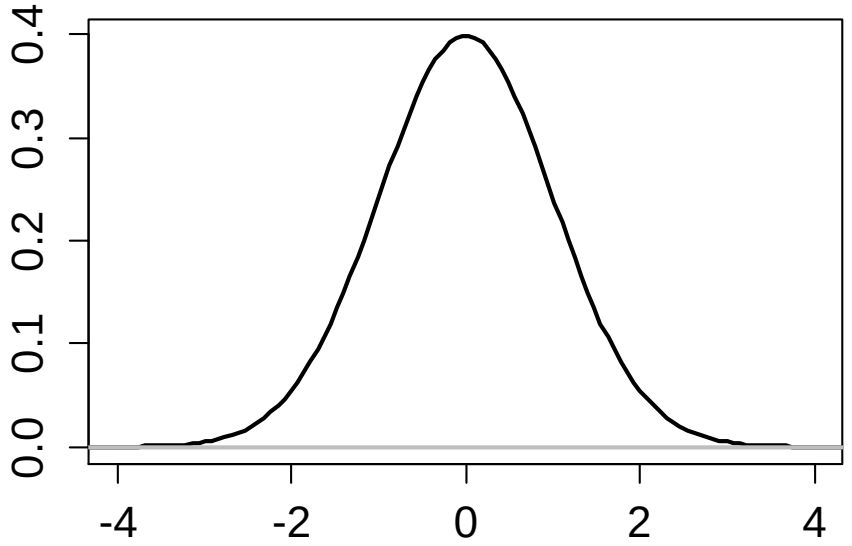
Test Statistic:
$$z = \frac{\hat{\pi}_1 - \hat{\pi}_2}{\sqrt{\hat{\pi}(1-\hat{\pi})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

where
$$\hat{\pi} = \frac{y_1 + y_2}{n_1 + n_2}$$

Alternative Hypothesis:	**Rejection Region:**
(1) $H_A: \pi_1 - \pi_2 > 0$	$z \geq z_\alpha$
(2) $H_A: \pi_1 - \pi_2 < 0$	$z \leq -z_\alpha = z_{1-\alpha}$
(3) $H_A: \pi_1 - \pi_2 \neq 0$	$\lvert z \rvert \geq z_{\alpha/2}$

P-values: (1) area to the right of z (test statistic), (2) area to the left, (3) area to the right of |z|.

In the French Skier example:



$H_0: \pi_1 - \pi_2 = 0$ vs $H_A: \pi_1 - \pi_2 \neq 0$

Recall: $\widehat{\pi_1} = 0.122, \widehat{\pi_2} = 0.221$

$$\hat{\pi} = \frac{y_1 + y_2}{n_1 + n_2} = \frac{17 + 31}{139 + 140} = 0.172$$

$$SE(\hat{\pi}_1 - \hat{\pi}_2) = \sqrt{0.172(1 - 0.172)\left(\frac{1}{139} + \frac{1}{140}\right)} = 0.0452$$

Test Statistic $z = \dfrac{0.122 - 0.221}{0.0452} = -2.19$

Reject H_0 because $|z| > z_{\alpha/2} = 1.96$

p - value = 0.028

<u>Conclusion:</u> Vitamin-C reduces incidence of colds in the population from which the sample of French skiers was taken.

Large Sample Test & CI for Two Proportions in R

- In R, use `prop.test()`

For the Skiers Example:

```
> prop.test(c(17,31), c(139,140), alternative =
  "two.sided", conf.level = 0.95, correct =
  FALSE)

 2-sample test for equality of proportions
  without continuity correction

data:  c(17, 31) out of c(139, 140)

X-squared = 4.8114, df = 1, p-value = 0.02827

alternative hypothesis: two.sided

95 percent confidence interval:

 -0.18685917 -0.01139366

sample estimates:

   prop 1    prop 2

0.1223022 0.2214286
```

Notes on the Large Sample test and CI for $\pi_1 - \pi_2$

1. When forming the CI, we used a standard error expression that involved separate estimates of $\pi_1$and π_2, but when testing used an estimate that involved a common estimate of $\pi_1$and π_2, which we called $\hat{\pi}$.
2. Due to #1, rejecting H_0: $\pi_1=\pi_2$ is not <u>exactly</u> equivalent to a CI for $\pi_1-\pi_2$ **not** containing zero.
3. **The (two-sided) Z-test is equivalent to the commonly used chi-squared (χ^2) test for 2x2 tables which we will discuss later in these notes.** The χ^2 statistic is equal to Z^2 where the formula for Z is given on slide 43.
4. Above test and CI require large enough samples, so that the normal approximation to the Binomial distribution is adequate.
5. A continuity correction (called the Yates correction) is possible. This is used by default in R. I am fine with the continuity correction.
6. **When sample sizes are small, tests of H_0: $\pi_1=\pi_2$ can be done using "Fisher's Exact Test" to be discussed later in these notes.**

5. Power and sample size for comparing proportions

All of the methods considered here are based on the large sample normal approximation:

A. Planning to achieve a desired CI width
B. Power calculation using R
C. Power calculation using Lenth

Note: Small sample power will be based on Fisher's Exact Test (FET). As far as I know, a power calculation for FET is not available in R or Lenth.

5A. Sample Size for a Desired Confidence Interval Width

Say we want a 95% C.I. for π_1-π_2 to be

$\hat{\pi}_1 - \hat{\pi}_2 \pm \mathrm{E}$ (width=2E)

$$E = z_{0.025} \times SE(\hat{\pi}_1 - \hat{\pi}_2) = z_{0.025}\sqrt{\frac{\hat{\pi}_1(1-\hat{\pi}_1)}{n_1} + \frac{\hat{\pi}_2(1-\hat{\pi}_2)}{n_2}}$$

Assume n_1 and n_2 are the same (=n), put in conjectures for $\hat{\pi}_1$ and $\hat{\pi}_2$, and solve for n:

$$n = \frac{z^2_{\alpha/2}\left(\hat{\pi}_1(1-\hat{\pi}_1) + \hat{\pi}_2(1-\hat{\pi}_2)\right)}{E^2}$$

What to do for $\hat{\pi}_1$ and $\hat{\pi}_2$ in this formula?

Conjecture the ranges of values for π_1 and π_2 and follow the guidelines given for the single sample case. Or take the worst case: $\pi_1 = 0.5$ and $\pi_2 = 0.5$.

Sample Size Example: How large must n (per group) be to get a 95% CI for π_1-π_2 of width 0.10.(i.e. E=0.05)

Say π_1 is thought to be between 0 and 0.6

and $\pi_2$is thought to be between 0.2 and 0.4.

$$n = \frac{(1.96)^2\left[(0.5)(1-0.5)+(0.4)(1-0.4)\right]}{(0.05)^2} = 753$$

Note This sample size estimate can be inaccurate when π_1 or π_2 is near (or thought to be near) 0 or 1 or sample sizes are small.

5B. Power calculation for comparing proportions in R

- In R, use power.prop.test()
- **Example:** Design an experiment to test H_0: $\pi_1 \geq \pi_2$ vs H_A: $\pi_1 < \pi_2$.
  Use conjectured proportions: $\pi_1 = 0.15$ $\pi_2 = 0.35$.
  Calculate power when $n_1 = n_2 = 50$.

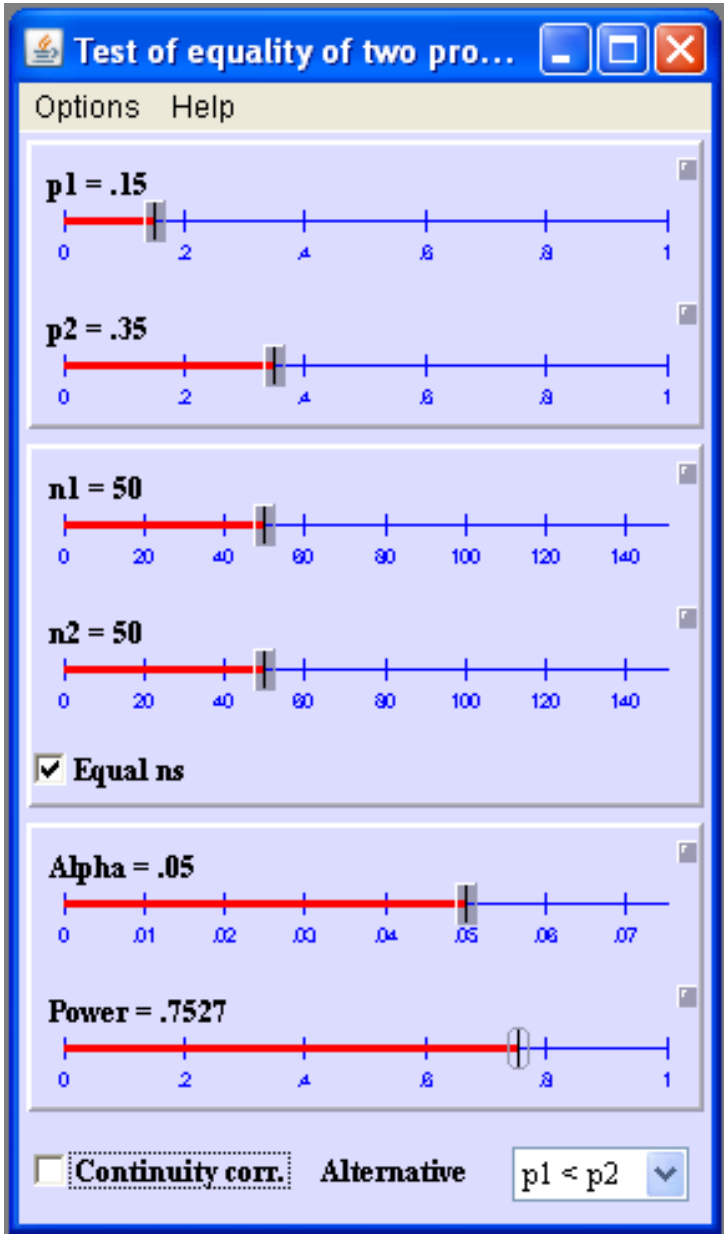
```
> power.prop.test(n = 50, p1 = 0.15, p2 = 0.35,
  sig.level = 0.05, alternative = "one.sided")

 Two-sample comparison of proportions power calc

              n = 50
             p1 = 0.15
             p2 = 0.35
      sig.level = 0.05
          power = 0.7526999
    alternative = one.sided

 NOTE: n is number in *each* group
```

5C. Power calculations for comparing proportions using Lenth

http://homepage.stat.uiowa.edu/~rlenth/Power/

- Choose test comparing two proportions

- Here we calculate power for HA: $\pi_1 < \pi_2$ using conjectured proportions $\pi_1 = 0.15$ and $\pi_2 = 0.35$, with n=50 per group.



6. McNemar's test for comparing paired proportions

Example (from Samuels & Witmer): A study was done to determine a woman's risk of transmitting HIV to her unborn children. A total of 114 HIV-infected women who gave birth to two children were included in the study.

	Older Sibling	Younger Sibling
HIV Yes	19	20
HIV No	95	94
Total	114	114

Summarized this way, the Z-test (or χ^2 test) might seem appropriate but the samples (older and younger siblings) are NOT independent.

We start by reformatting the data retaining the information about pairs.

	Younger HIV Yes	Younger HIV No	Total
Older HIV Yes	2	17	19
Older HIV No	18	77	95
Total	20	94	114

Now we have each sibling pair grouped by HIV status.
Note that there are **79 concordant** pairs **and 35 discordant** pairs.

H_0: $\pi_{Older}=\pi_{Younger}$

or equivalently

H_0: Among discordant pairs, Pr(yes/no) = Pr (no/yes) = 0.5

We will use `mcnemar.test()` to run the test.

McNemar's test of paired proportions in R

- In R, use `mcnemar.test()`

```
> Siblings<-matrix(c(2,17,18,77), byrow = TRUE,
nrow = 2)
> Siblings
     [,1] [,2]
[1,]    2   17
[2,]   18   77
> mcnemar.test(Siblings)
        McNemar's Chi-squared test with
  continuity correction
data:  Siblings
McNemar's chi-squared = 0, df = 1, p-value = 1
```

So for this example, we Fail to Reject H_0. We cannot conclude that there is a difference in infection rates for older and younger siblings.

7. Chisquare Goodness of Fit (GOF) Test

Maize Example (Snedecor and Cochran): Two types of maize were crossed. A sample of n=1301 plants is taken and 4 types of maize are observed. Hence k = 4 categories.

	Green (1)	Gold (2)	Green Striped (3)	Green/Gold Striped (4)	Total
Observed	n1=773	n2=231	n3=238	n4=59	1301
H0	π1=9/16	π2=3/16	π3=3/16	π4=1/16	1

Question: Are these data consistent with the Mendelian laws of inheritance? These laws would imply that in the long run the four types would occur with the following proportions:
π1=9/16, π2=3/16, π3=3/16, π4=1/16.

(Pearson's) Chisquare Goodness of Fit Test

Assumptions: Independent observations, large sample size.
Rule of thumb for sample size: No E_i can be less than 1, and no more than 20% of E_i's can be less than 5. (See discussion later in notes.)

H_0: $\pi_i = \pi_{i0}$ for categories i=1,…,k. (π_{i0} are specified probabilities or proportions.)

H_A: At least one of the cell probabilities differs from the hypothesized value.

Test statistic: $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - E_i)^2}{E_i}$

where $E_i = n\pi_{i0}$ (the expected count under H_0)

Rejection Region: Reject H_0 if $\chi^2 > \chi^2_\alpha$ with df=k-1

In R, use `chisq.test()` see "**Chisquare GOF test**" Example.

GOF Test for the Maize Example

$H_0 : \pi_1 = 9/16, \pi_2 = 3/16, \pi_3 = 3/16, \pi_4 = 1/16$

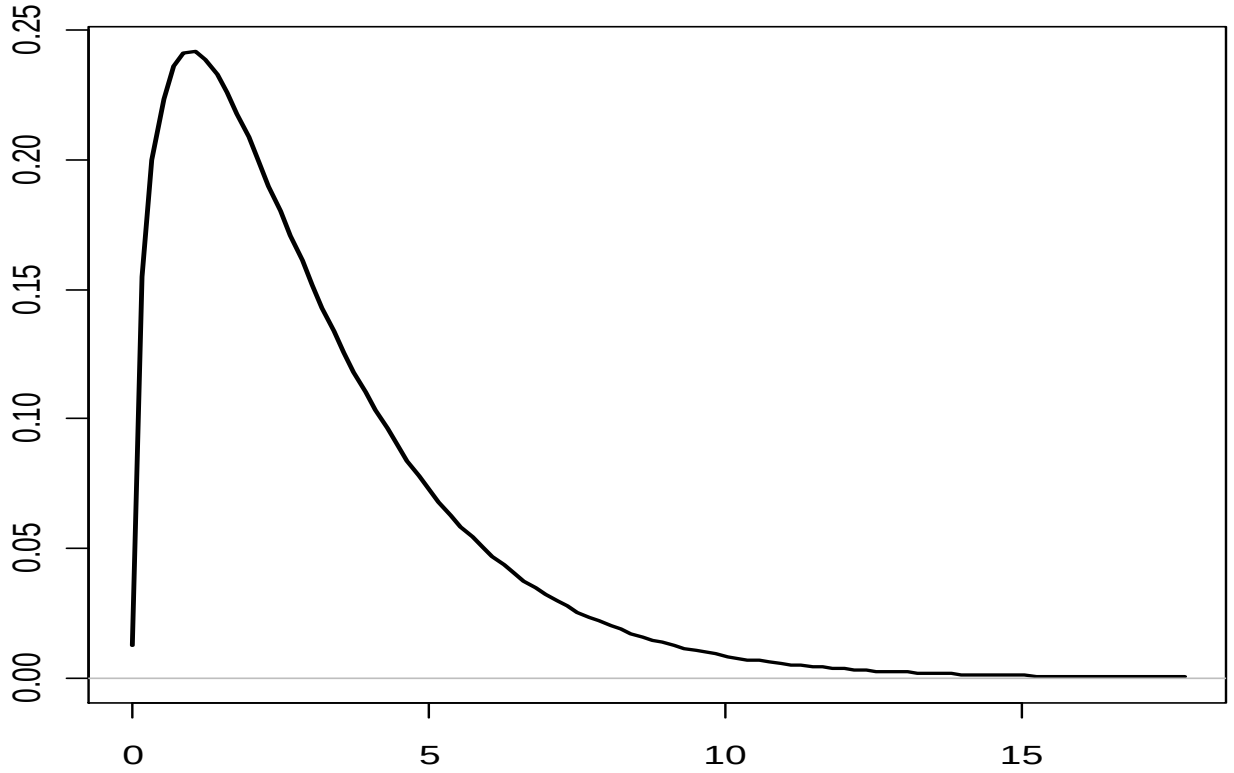
$H_A : \text{not } H_0$

	Green (1)	Gold (2)	Green Striped (3)	Green/Gold Striped (4)
Observed	n1=773	n2=231	n3=238	n4=59
	1301(9/16)	1301(3/16)	1301(3/16)	1301(1/16)
Expected	E1=731.8	E2=243.9	E3=243.9	E4=81.3

$$\chi^2 = \frac{(773-731.8)^2}{731.8} + \frac{(231-243.9)^2}{243.9} + \frac{(238-243.9)^2}{243.9} + \frac{(59-81.3)^2}{81.3} = 9.27$$

$\chi^2_{0.05} = \texttt{qchisq(0.95,df=3)} = 7.81$ (or Table 7 with df = k – 1 =3).
Since $9.27 = \chi^2 > \chi^2_{0.05} = 7.81$, Reject H_0.

GOF Test for the Maize Example (continued)



In R: pvalue=1-pchisq(9.27, df = 3) = 0.026
critval=qchisq(0.95, df = 3) = 7.81

Conclusion: Reject H_0. Evidence against Mendel's law.
But which category (color) is different than expected? Look at Pearson residuals

Pearson residuals $r_i = \dfrac{n_i - E_i}{\sqrt{n(\pi_i)(1-\pi_i)}}$

For large E_i these are approximately standard normal:
If the null hypothesis is true, there is only a 5% chance of a Pearson residual to take a value outside the interval between -1.96 and 1.96.
(For small E_i the residuals distribution is skewed)

For the Maize Example, the residuals are:
2.30 (green), -0.92 (gold), -0.42 (gr striped) and -2.56 (green/gold striped)

Conclusion: Data consistent with Mendel's law except for the last category and perhaps the first (a "chlorophyl abnormality").

Follow-up: Omit the last category, renormalize the probabilities to add up to 1, then re-run the Pearson's Chisquare test: $\chi^2 = 2.7$, df=2, p=0.25.

Follow-up conclusion: Data consistent with Mendel's law in the first three categories.

Notes about the χ^2 Goodness of Fit Test

1. One possible strategy: After doing the overall χ^2 test, look at the residuals. Sometimes, cells can be combined or omitted to demonstrate a point.

2. The distribution of the χ^2 –statistic is approximately chisquare for "large n":

   a) A very conservative rule of thumb: all E_i must be at least 5.

   b) A <u>better</u> rule of thumb (Cochran): no E_i can be less than 1, and no more than 20% of the E_i can be less than 5.

3. The more cells, the less strict you have to be about the above rules of thumb.

4. You can combine cells to increase the E_i values.

5. Generally, hypothesis tests have the research hypothesis as the <u>alternative</u>. Most χ^2 goodness of fit tests have the research hypothesis as the <u>null</u>. This is risky because the type II error rate is harder to compute and often left uncontrolled.

Notes about the χ^2 Goodness of Fit Test *(continued)*

6. The rejection region is almost always to the right, but the test should be thought of as "non-directional" with respect to the π_ι 's.

A (rare) exception (rejection region to the left): Fisher's analysis of a group of experiments reported by Mendel:

H_0: Mendel's law holds.
H_A: Data were "edited" so that they are closer to the theoretically expected E_i's than they would be if "editing" was not done.
(He combined several experiments into one test.)
$\chi^2 = 42$, d.f = 84, p-value = 0.00004

Conclusion: Reject H_0. Data closer to E_i's than random data would be; i.e., the data were "edited".

Fisher: "I have no doubt that Mendel was deceived by a gardening assistant, who knew only too well what his principal expected from each trial made."

8. Chi-square Test for Contingency Tables

Assumptions: Independent observations, large sample size.
Rule of thumb for sample size: No E_i can be less than 1, and no more than 20% of E_i's can be less than 5. (See discussion for GOF test.)

H_0: The row and column variables are independent.
H_A: The row and column variables are dependent (associated).
(Can be other statements of hypotheses.)

Test statistic:
$$\chi^2 = \sum_{i,j} \frac{(n_{ij} - E_{ij})^2}{E_{ij}}$$

E_{ij} = what you would expect if H_0 were true

$$= \frac{(\text{i}^{\text{th}} \text{ row total})(\text{j}^{\text{th}} \text{ column total})}{\text{grand total}}$$

Rejection Region: Reject H_0 if $\chi^2 > \chi^2_\alpha$ with df=(#rows-1)(#cols-1)

Test of Equality of proportions (2x2Table): French Skier Example

	Cold	No Cold	Total
Vitamin C	17	122	139
Placebo	31	109	140
Total	48	231	279

$\widehat{\pi_{VitC}} = 17/139$

$\widehat{\pi_P} = 31/140$

$H_0: \pi_{VitC} = \pi_P$ vs $H_A: \pi_{VitC} \neq \pi_P$

$\chi^2 = 4.811$

$\chi^2_{0.05} = 3.84$ (df = (2-1)(2-1) = 1)

p-value = 0.028

So we Reject H_0

Note that $\chi^2 = 4.811 = 2.19^2 = Z^2$

p-values are the same.

E_{ij} = what you would expect if H_0 were true

$$= \frac{(i^{th} \text{ row total})(j^{th} \text{ column total})}{\text{grand total}}$$

e.g. $E_{11} = \frac{(139)(48)}{279} = 23.91$

(note: 48/279 is $\hat{\pi}$ from previous analysis.)

$E_{12} = 115.09$ $E_{21} = 24.09$ $E_{22} = 115.91$

See: "**Chisquare Tests for Contingency Tables**" Example

Chi Square Test for Contingency Tables in R/Rcmdr

- In R, use `chisq.test()`
- In Rcmdr, choose Statistics > Contingency Tables > Enter and analyze two-way table.

For the Skiers Example:

```
> Skiers<-
  matrix(c(17,122,31,109),byrow=TRUE,nrow=2)
> Skiers
     [,1] [,2]
[1,]   17  122
[2,]   31  109
> chisq.test(Skiers,correct=FALSE)
        Pearson's Chi-squared test
data:  Skiers
X-squared = 4.8114, df = 1, p-value = 0.02827
```

Test of Equality of proportions (3x2Table): Carcinogenicity Study
Three groups of 100 rats were given different doses of a drug scheduled for testing in humans.

	Yes Tumors	No Tumors	Total
Control	10	90	100
Low Dose	14	86	100
High Dose	19	81	100
Total	43	257	300

$\widehat{\pi_C}$= 10/100 =0.10

$\widehat{\pi_L}$= 14/100 =0.14

$\widehat{\pi_H}$= 19/100 =0.19

H_0: $\pi_{Ctrl} = \pi_{Low} = \pi_{High}$
H_0: Not all the proportions are the same

$\chi^2 = 3.31$, df=(3-1)(2-1)=2
p-value = 0.191
Fail to Reject H_0. No evidence that probability of tumors depends on dose.

NOTE: This test does not use information about what appears to be trend. Could use logistic regression to account for this.

Test of Independence (2x2 Table): Opinion Example

1397 people were surveyed and asked two questions:

1. Do you favor hand gun registration (GR)?
2. Do you favor the death penalty (DP)?

	DP Yes	DP No	Total
GR Yes	784	236	1020
GR No	311	66	377
Total	1095	302	1397

$\widehat{\pi_1}$= 784/1020

$\widehat{\pi_2}$= 311/377

H_0: Opinion about GR and DP are independent ("No Relationship")

H_A: Opinion about GR and DP are associated ("Some Relationship")

$\chi^2 = 5.15$, df=(2-1)(2-1)=1

p-value = 0.023

Reject H_0. The data support the conclusion that the responses are dependent. There is a relationship between GR and DP opinions!

A Chisquare Test of Independence - Pearson Residuals

$$\text{Pearson Residual for cell in row } i, \text{ column } j = \frac{\left(\begin{matrix}\text{Observed}\\ \text{Cell Count}\end{matrix}\right) - \left(\begin{matrix}\text{Expected}\\ \text{Cell count}\end{matrix}\right)}{\sqrt{n\hat{\pi}_{ij}(1-\hat{\pi}_{ij})}}$$

$\hat{\pi}_{ij} = (\text{expected count for cell i, j}) / (\text{grand total})$

$n = (\text{total number of observations})$

	DP Yes	DP No	Total
GR Yes	784	236	1020
GR No	311	66	377
Total	1095	302	1397

$$\text{Pearson Residual for the (No,No) cell} = \frac{(66-81.5)}{\sqrt{1397\times(81.5/1397)\times(1-81.5/1397)}} = -1.77$$

This residual is the largest of the 4 residuals (verify).

Notes about the Chi-Squared Test for Contingency Tables

1. The (two-sided) two-sample Z-test comparing two proportions is equivalent to the chi-squared (χ^2) test for 2x2 tables.
2. Note the difference in how the sampling was done for the Skiers data (2x2 Table) versus the Opinions data (also 2x2 Table). The hypotheses may be stated differently (equality of proportions versus independence), but the test is done the same way.
3. Reordering rows or columns (or transposing rows and columns) will give the same result.
4. The chi-squared test can be used for tables with any number of rows and columns. However, the expected values for each cell must be reasonably large. Sometimes it helps to combine categories to achieve this.
5. When sample size is small, use Fisher's Exact Test (next slide).

Fisher's Exact Test (FET) for Small Samples

Example: Song Discrimination in Warblers (2x2 Table)

Blue-winged and Golden-winged warblers of Southeastern Michigan were tested using tape recordings. If they responded to recorded songs from only their own species they were termed "discriminators". If they responded to songs from both species, they were "non-discriminators". We are interested in whether the proportion of discriminators differ by species.

	Disc Yes	Disc No	Total
Blue	4	6	10
Gold	3	9	12
Total	7	15	22

$\widehat{\pi_B}$ = 4/10 =0.40

$\widehat{\pi_G}$ = 3/12 = 0.25

In this example, the sample size is so small that we doubt whether the Chi-square test is valid.

Fisher's exact test (FET): A randomization test, valid for any sample size. Designed for sampling in which both row and column marginal totals are set by the experimenter; however, Fisher and others argue that it should be used whenever the sample size is small.

FET for Contingency Tables in R/Rcmdr

- In R, use `fisher.test()`
- In Rcmdr, choose Statistics > Contingency Tables > Enter and analyze two-way table, select Fisher's Exact Test.

- For the Birds Example:

```
> Birds<-matrix(c(4,6,3,9),nrow=2,byrow=TRUE)
> Birds
     [,1] [,2]
[1,]    4    6
[2,]    3    9
> fisher.test(Birds)

        Fisher's Exact Test for Count Data

p-value = 0.6517
alternative hypothesis: true odds ratio is not
  equal to 1
```

An Example in which both row and column marginal totals are fixed by the experiment: Draft Lottery Data Example

Example:

In 1970, Congress instituted a “random” selection process for the military draft. All 366 possible birth dates were placed in plastic capsules in a rotating drum and were selected one by one. The first date drawn from the drum received draft number one and eligible men born on that date were drafted first. In a truly random lottery there should be no relationship between the date and the draft number. However, this dataset suggests that men born later in the year were more likely to be drafted. An investigation of the lottery revealed that the birth dates were placed in the drum by month and were not thoroughly mixed.

Draft Lottery Data (3x3 Table)

	Jan-Apr	May-Aug	Sep-Dec	Total
1-122	29	45	48	122
123-244	42	28	52	122
245-366	50	50	22	122
Total	121	123	122	366

Question: Is there evidence against the claim that the numbers were drawn in random order?
Since both margins are fixed, Fisher's Exact Test is appropriate, but for such large sample sizes the χ^2 test is appropriate also.
$\chi^2 = 25.107$ df=4 p-value < 0.001

Conclusion: Reject H_0. Conclude drawing not random. Looking at the column proportions (Jan-Apr: 29/121 = 0.24, May-Aug: 45/123 = 0.37, Sep-Dec: 48/122 = 0.39), we conclude that people born later in the year had a higher probability of getting a low draft number.

Some Perspective

When we have a **binary response** variable and **two (or more) groups**, we have been focusing on methods to compare **proportions**.

- Large sample Z-test for $\pi_1 - \pi_2$
  - 2 groups
  - In R: prop.test()
- X^2 test for contingency tables
  - With 2 groups, equivalent to Z-test above.
  - Can be used with more than two groups ("larger" tables)
  - In R: chisq.test()
- Fisher's Exact Test (FET)
  - Alternative to X^2 test when sample size is small.
  - In R: fisher.test()
- But instead of summarizing as proportions, we can also summarize as **odds ratio(s).**
- We could also consider logistic regression with a categorical predictor, but that approach is not covered in STAT511.

9. Odds Ratios

Given that proportions are not homogeneous, or two variables are not independent, how do we describe the **strength** of the relationship? We'll use the Vitamin C study for the discussion.

Note: When calculating odds ratios in R, it helps to have (1) reference/control group in first row and (2) "event" in last column.

	No Cold	Cold	Total
Placebo	109	31	140
Vitamin C	122	17	139

1. Difference of proportions

$$\hat{\pi}_1 - \hat{\pi}_2 = \frac{17}{139} - \frac{31}{140} = -0.101$$

Advantage: simple

Disadvantage: not good over a wide range of cold prevalence. A difference of 0.1 is relatively large when π's are small, but not when π's large.

2. Ratios of proportions ("Risk Ratio")

$$\frac{\hat{\pi}_1}{\hat{\pi}_2} = \frac{17/139}{31/140} = 0.552$$

Advantage: simple

Disadvantage: relatively hard to work with mathematically, can be misleading.

3. Ratio of Odds ("Odds Ratios")

$$\text{Definition : Odds} = \frac{P(yes)}{P(no)} = \frac{P(yes)}{1 - P(yes)}$$

$$\text{If } P(yes) = 0.8, \text{ then Odds} = \frac{0.8}{0.2} = 4$$

Usually expressed as 4:1

Definition :

$$\text{Odds ratio for situation 1 relative to situation 2} = \frac{\text{Odds for situation 1}}{\text{Odds for situation 2}} = \lambda$$

Odds Ratio for the French Skiers Data:

	No Cold	Cold	Total
Placebo	109	31	140
Vitamin C	122	17	139

$$\text{Odds ratio} = \lambda = \frac{\text{odds of cold in VitC group}}{\text{odds of cold in Placebo group}} = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}$$

$$\text{Estimate by } \hat{\lambda} \cong \frac{\left(17/139\right)/\left(122/139\right)}{\left(31/140\right)/\left(109/140\right)} = \frac{17/122}{31/109} = \frac{17\times 109}{31\times 122}$$

$$= 0.49$$

<u>Conclusion:</u> Odds of getting a cold in the Vitamin-C group are estimated to be <u>about half</u> the odds of getting a cold in the Placebo group.

NOTES about odds ratio λ:

1. $\lambda=1$ indicates no difference between the groups. So, if a CI for λ includes 1, then we cannot conclude a difference between the odds for the two groups. See the Skiers example.
2. The odds ratio is not affected if an entire row or column is multiplied by a constant. This is important for analyzing case-control studies. See the Birth Control example.
3. If the rows (or columns) are interchanged, then $\lambda_{\text{new table}} = 1/\lambda_{\text{old table}}$.
4. Odds ratio can be used for larger tables by combining rows and/or columns down to 2 by 2, or by looking at a series of 2 by 2 subtables. See the Tumors example.
5. The odds ratio and corresponding confidence interval can be computed in R using the `oddsratio()` function in the `epitools` package. See the "**Odds Ratio Examples**".

A Confidence Interval for the odds ratio (λ)

It is known that ln(λ) (called "**logodds**") has an approximately normal distribution when the cell counts are large. So, our strategy is to first construct a confidence interval for ln(λ). The point estimate of ln(λ) is given by $\ln(\hat{\lambda})$. An approximate large sample standard error for $\ln(\hat{\lambda})$

$$SE(\ln(\hat{\lambda})) = \sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}}$$

$$\text{A 95\% C.I. for } \ln(\lambda) \text{ is: } \ln(\hat{\lambda}) \pm (1.96)\, SE(\ln(\hat{\lambda}))$$

To obtain a confidence interval for the odds-ratio λ, compute the interval for ln(λ), then **exponentiate both bounds to get an interval for λ**.
*NOTE: ln = natural log!

Odds Ratio CI for the French Skiers Data:

$$\ln(\hat{\lambda}) = \ln(0.49) = -0.71$$

$$SE(\ln(\hat{\lambda})) = \sqrt{\frac{1}{17} + \frac{1}{122} + \frac{1}{31} + \frac{1}{109}} = 0.32$$

$$95\%\text{CI for } \ln(\lambda): -0.71 \pm 1.96 * 0.32 \rightarrow (-1.34, -0.08)$$

$$95\%\text{CI for } \lambda: (e^{-1.34}, e^{-0.08}) \rightarrow (0.26, 0.93)$$

NOTES:

- CI includes 0.49, but not symmetric about the estimate.
- The resulting CI does not include 1.0. This is consistent with the rejection of the null hypothesis that the probability of a cold is the same for both groups.

Odds Ratios in 2x2 Sub Tables *(Tumor Data Example)*

	No Tumors	Yes Tumors	Total
Control	90	10	100
Low Dose	86	14	100
High Dose	81	19	100

Odds ratio in favor of tumors for low dose relative to control :

$$\hat{\lambda}_1 = \frac{14/86}{10/90} = 1.46$$

95% C.I. for λ_1 is $(0.62, 3.47)$

Odds ratio in favor of tumors for high dose relative to control :

$$\hat{\lambda}_2 = \frac{19/81}{10/90} = 2.11$$

95% C.I. for λ_2 is $(0.93, 4.8)$

Conclusion: Odds of tumors are estimated to increase by a factor of about 1.5 for low dose, and about 2.1 for high dose, relative to control. (These estimates are not significantly different from 1.0)
Note: our analysis ignores an apparent trend. (See: Logistic Regression examples later for analysis of trends in categorical data.)

Odds Ratios in Retrospective Studies *(Case-Control Studies)*

Birth Control Case-Control Study: In a 1975 study, 58 married women under 45 being treated for myocardial infarction ("heart attack") were each matched with three women (similar age, weight, etc), and all classified on whether they had used oral contraceptives.

From Wikipedia (10/13/17):
"A case-control study is a type of observational study in which two groups differing the outcome are identified and compared based on some supposed causal attribute. They require fewer resources but provide less evidence for causal inference as compared to randomized experiment."

This type of study is often used when the outcome (typically a disease or particular cause of death) is relatively rare.

Important note: We **cannot** estimate the probability of having a heart attack from this data! We certainly do NOT have a random sample, since we identified women with heart attacks and matched them to controls. The odds ratio is an appropriate analysis for case-control studies.

Odds Ratio for the Birth Control Data:

Myocardial infarction

Contraceptive practice (p=0.004)	No	Yes	Total
Never Used	132	35	167
Used	34	23	57
Total	166	58	224

$$\hat{\lambda} = \frac{(23/34)}{(35/132)} = 2.55 \qquad 95\%\ \text{C.I.}\ (1.34, 4.87)$$

Conclusion: The odds of myocardial infarction are estimated to be 2.55 times higher among those that have used oral contraceptives. This value is ***significantly*** higher than 1.0.

10. Three-way tables and "Simpson's Paradox"

Example: Death Penalty and Race in Florida.

Defendants convicted of first degree murder were classified by whether they were sentenced to death and their race.

		Death Penalty	
		Yes	**No**
Defendant's Race	**White**	**19**	**141**
	Black	**17**	**149**

$$\text{Odds ratio} = \frac{(19/141)}{(17/149)} = 1.18 \qquad 95\%\ \text{C.I. } (0.59, 2.36)$$

Conclusion:

Whites are estimated to have odds of getting the death penalty 1.18 times greater than odds for blacks. This is not significantly different from 1.0, but we are looking at trends here.

Death Penalty Example continued

Since whites getting the death penalty more often was surprising, we look at the data more closely, dividing the data according to **race of victim**.

Victim's race = WHITE

		Death Penalty	
		Yes	**No**
Defendant's Race	**White**	**19**	**132**
	Black	**11**	**52**

$$\hat{\lambda} = \frac{(19/132)}{(11/52)} = 0.68$$

Victim's race = BLACK

		Death Penalty	
		Yes	**No**
Defendant's Race	**White**	**0**	**9**
	Black	**6**	**97**

$$\hat{\lambda} = \frac{(0/9)}{(6/97)} = 0$$

Death Penalty Example continued

Question: How is it possible that the odds ratios are **less than 1.0** for both the victim-race sub-tables, but **greater than 1.0** in the combined table?

Answer: It is possible when **both** the variables in the combined table are related to the third variable over which data was combined.
This is called **Simpson's Paradox (sometimes called Yule's Paradox).**

		Death Penalty	
		Yes	**No**
Victim's Race	**White**	**30**	**184**
	Black	**6**	**106**

$$\hat{\lambda} = \frac{(30/184)}{(6/106)} = 2.88$$

		Defendant's Race	
		White	**Black**
Victim's Race	**White**	**151**	**63**
	Black	**9**	**103**

$$\hat{\lambda} = \frac{(151/63)}{(9/103)} = 27.4$$

Death Penalty Example continued

Conclusion: The relationship between Defendant's Race and Death Penalty in the combined ("collapsed") table is different from the relationship between Defendant's Race and Death Penalty in the two individual Victim Race sub-tables. This appears to have occurred because Race of Victim is related to **both** of the other two variables.

Of course, there are many other variables that we are not accounting for: prior record, victim gender, jury composition, etc!

In an observational study (a study in which you observe, but do not manipulate the experimental conditions for your subjects) it is very difficult to identify one variable as "causing" another because the relationship you are observing could be a result of both variables being related to other unobserved variables.
Association is NOT causation!

Berkeley Gender Discrimination Example (1975)

Looking at all graduate records (combining departments), the estimated odds of admission for males vs females is found to be $\hat{\lambda} = 1.84$ (95% CI = (1.62, 2.09), p-value < 0.001). This supports that men have higher odds of admission as compared to women.

Logical next step is to look at individual departments. When we do this we find department odds ratios of 0.35, 0.81, 0.83, 0.92, 1.13, and 1.22. Most departments have higher odds of admission for women!

Based on Breslow-Day test (next slide), we find significant differences between the odds ratios for the departments. Hence it is not appropriate to combine departments!

Alternative Analysis: Logistic Regression.

11. The Breslow-Day Test

Breslow – Day Tests whether the odds ratios are the same for all sub-groups:

H_0: λ_h's all equal

H_A: some λ_h different from others

The Breslow-Day test can be done in R using the `rma.mh()` function in the `metafor` package.

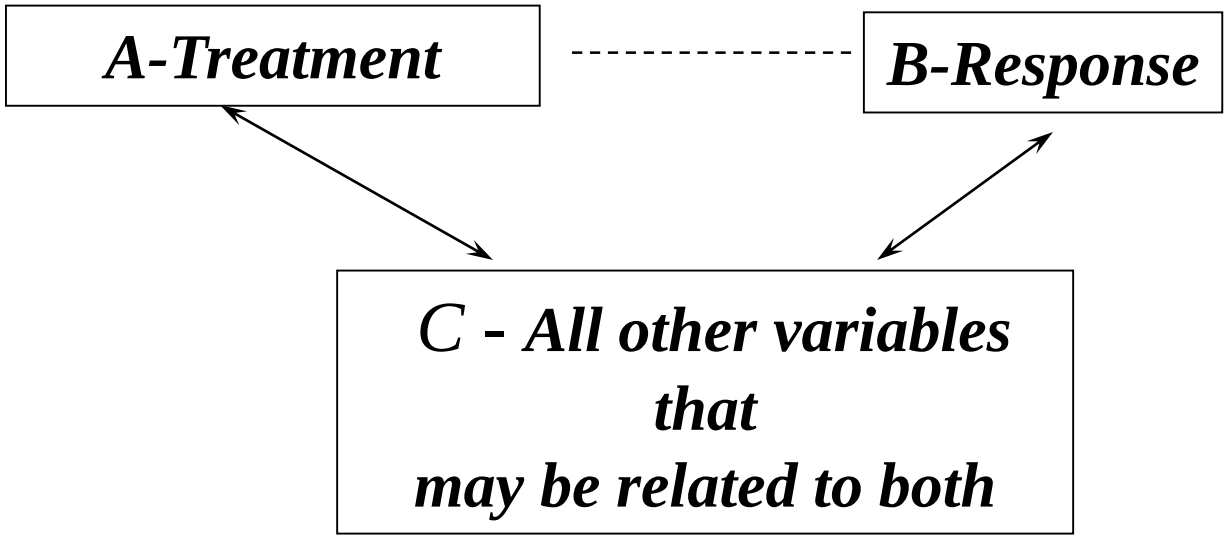
For the Berkeley Data, see the "**Discrimination Example**".

The BD p-value = 0.0021, so we Reject H0 and conclude that the odds ratios are not the same for all departments.

Conclusion: We should not combine information across departments! Women were applying in higher numbers to Departments that were harder to get into (see bar plots).

Designed Experiments versus Observational Studies

Question: Why can you conclude more in a **designed** study (with treatments randomly assigned) than in an **observational** study?



Randomization of subjects (experimental units) to treatment groups helps achieve (approximate) balance with respect to the "confounding variables". It "breaks the circuit" between A and C.

A Designed Experiment – the Drug Clinic Example

Example: A drug is tested at three clinics:

Clinic	Drug group	Improved	Not Improved	Total
1	Drug	40 (80%)	10	50
	Placebo	15 (30%)	35	50
	Total	***55***	***45***	***100***
2	Drug	35 (70%)	15	50
	Placebo	20 (40%)	30	50
	Total	***55***	***45***	***100***
3	Drug	43 (86%)	7	50
	Placebo	31 (62%)	19	50
	Total	***74***	***26***	***100***

See: "**Drug Clinic Three-way Table**" Example

Drug Clinic Example

1. Use the Breslow-Day test to evaluate whether we want to combine data from three clinics:

   λ_1=9.33, λ_2=3.50, λ_3 = 3.77

   BD p-value = 0.245

   Conclusion: No evidence that odds ratios depend on clinic.

2. It makes sense to combine information across the clinics, controlling for clinic. We do this using the (Cochran) Mantel-Haenszel Test (see next slide for details).

   Controlling for clinic (estimating λ_h separately for each clinic, then combining) we find:

   λ_{CMH} = 4.90, 95% CI (2.92, 8.21)

   Based on the CI or the CMH test (p-value < 0.001) we conclude that the odds ratio is different from 1. The odds of improvement are higher for the active treatment as compared to placebo.

(Cochran) Mantel-Haenszel Test

Assume an experiment involving a 2 by 2 table is replicated at k locations. Let n_{hij} be the count for the i^{th} row, the j^{th} column, of the h^{th} table.

$$\chi^2_C = \frac{\left\{\sum_h \left(n_{h11} - \frac{n_{h1+} n_{h+1}}{n_{h++}} \right)\right\}^2}{\sum_h \frac{n_{h1+} n_{h2+} n_{h+1} n_{h+2}}{n^2_{h++}(n_{h++} - 1)}} \qquad \text{df} = 1$$

Let the odds ratio at the h^{th} location be λ_h. The CMH statistic tests:

$$H_0 : \text{"average" of } \lambda_h\text{'s equals } 1.$$

$$H_A : \text{"average" of } \lambda_h\text{'s does not equal } 1.$$

In R, use (1) `mantelhaen.test()`, (2) `cmh.test()` in the `lawstat` package or (3) `rma.mh()` in the `metafor` package.

12. **The Poisson distribution** (for count data)

The Poisson distribution is an example of a discrete distribution. It is a common choice for modeling count data. Specifically used to model the number of events over a particular unit of time or space.

Examples:
Number of calls to a telephone operator per hour
Number of insects per leaf

Assumptions:

1. Events occur one at a time; two or more events do not occur at precisely the same time or space.
2. The occurrence of an event in a given period of time (or region of space) is independent of the occurrence of the event in a non-overlapping time period (or region of space).
3. The expected number of events during one period (or region) μ, is the same as the expected number of events that occur in any other period (or region).

Poisson Distribution

Example: Observe the number of insects per leaf.

1. Imagine that the leaf is divided into a large number of "grids", say n grids.
2. Each grid may have an insect in it (success) or not have an insect in it (failure)
3. There can only be 0 or 1 insect per grid. The probability of a grid having an insect in it is π.
4. The average number of insects per leaf is $\mu = n\pi$.
5. Whether there is an insect or not in one grid doesn't change the probability of there being an insect in any other grid (independence).

Poisson Probabilities

If Y follows the poisson distribution with parameter μ we say Y~Poisson(μ).
Then the following formula gives the probability that
Y is equal to a specified value y (which may be 0, 1, 2, …..)

$$P(Y = y) = \frac{\mu^y e^{-\mu}}{y!}, \text{ for } y = 0,1,2,...$$

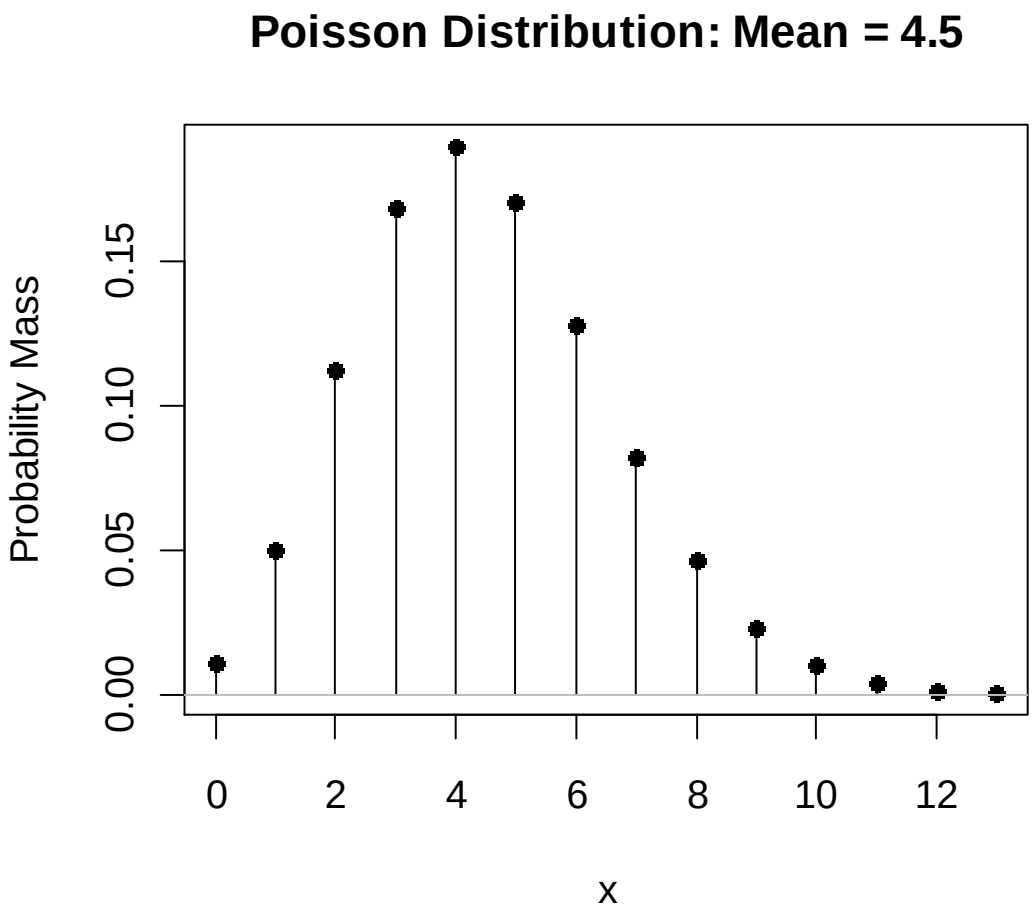
where $e = 2.7182...$(base of natural logarithms)

The mean of Y is μ and the variance of Y is also μ.

In R, use `dpois()` which gives P(Y=k) or `ppois()` which gives P(Y≤k).

Poisson Distribution (Example)

Example: Let Y denote the number of clams captured in a trap during a given time period. Suppose that Y has a Poisson distribution with $\mu=4.5$, so on average traps will contain 4.5 clams.



Poisson Probabilities in R and Rcmdr

- In R, use `dpois()` which gives P(Y=y) or `ppois()` which gives P(Y≤y)
- In Rcmdr, use Distributions > Discrete Distributions > Poisson Distribution.

  Choose Poisson Probabilities (ex: P(Y=5)) or Poisson Tail Probabilities (ex: P(Y≤5)).

Example: Suppose μ=4.5.

```
> dpois(4,4.5)
[1] 0.1898076
> ppois(4,4.5)
[1] 0.5321036
> ppois(4,4.5)-ppois(3,4.5)
[1] 0.1898076
```

Poisson χ^2 Goodness of Fit Test

H0: Data are from Poisson distribution
HA: Data are NOT from Poisson distribution.

1. Estimate $\hat{\mu}$ = #events/#units = average count per unit.
2. Calculate $P(y_i)=P(Y=y_i)$ for each cell: y= 0,1,2,3,4, .. based on the Poisson distribution with mean = $\hat{\mu}$.
3. For each cell, calculate $E_i = n \cdot P(y_i)$ where n = total # units.
   Group cells if necessary, so that all $E_i > 1$)
4. Do a χ^2 goodness of fit test with df = k – 2, where k = #categories/cells.

Mulekick Example: Poisson GOF test

The number of soldiers in the Prussian army that were kicked to death by mules in the twenty years from 1875 to 1894 (von Bortkewitsch, 1898) were recorded. There were 10 army corps and 20 years (n = 200 corp-years). Is the number of deaths per corps per year Poisson distributed?

# Events (# Deaths)	# Units (# Corps)
0	109
1	65
2	22
3	3
4	1

1. Calculate $\hat{\mu}$ = #events/#units.

   # events = 109(0) + 65(1) + 22(2) + 3(3) + 1(4)

   = 122 deaths by mule kick

   # units = 10 corps x 20 years = 200

   $\hat{\mu}$ = 122/200 = 0.61

Mulekick Example continued

2. Calculate Poisson probabilities (Prob) based on $\hat{\mu} = 0.61$.
3. Calculate expected counts: Exp = n x Prob
4. Do a χ^2 goodness of fit test with df = k – 2, where k = #categories/cells.

# Events (Y)	# Units (Obs)	Prob (dpois(Y, 0.61))	Exp (200*Prob)	X2
0	109	0.5434	108.7	0.001
1	65	0.3314	66.3	0.025
2	22	0.1011	20.2	0.157
3	3	0.0206	4.1	0.300
4	1	0.0035	0.7	0.117

Test Statistic $\chi^2 = 0.5999$, df = 5–2 = 3.

P-value = 1-pchisq(0.599, df = 3) = 0.8965

Conclusion: Fail to Reject H0. No evidence against Poisson.

See **Mulekick Example** in R.

Another Example from O&L

A sample of lake water was taken, and n=150 microscope slide fields (units) were inspected for algae cell clumps (events).

Question: Is the distribution of cell clumps on a slide Poisson distributed?

Let Y_i = Number of clumps in the i^{th} field.

A frequency table for Y is given below:

y_i	0	1	2	3	4	5	6	7	8	9	10	11
n_i	6	23	29	31	27	13	8	9	2	0	1	1

$$\hat{\mu} = \left(\sum n_i y_i\right) / \left(\sum n_i\right) = (495/150) = 3.3$$

Calculate the probabilities for each cell in the table using $\hat{\mu}$

Checking to see if data are Poisson – Second Example

y_i	$P(y_i)$	E_i	n_i
0	0.036883	5.5325	6
1	0.121714	18.2572	23
2	0.200829	30.1243	29
3	0.220912	33.1368	31
4	0.182252	27.3378	27
5	0.120286	18.0430	13
6	0.066158	9.9236	8
7	0.031189	4.6783	9
8	0.012865	1.9298	2
9	0.004717	0.7076	0
10	0.001557	0.2335	1
11	0.000467	0.0701	1
12	0.000128	0.0193	0
13	0.000033	0.0049	0
14	0.000008	0.0012	0
15	0.000002	0.0003	0
16	0.000000	0.0001	0

y_i	$P(y_i)$	E_i	n_i
0	**0.036883**	**5.53**	**6**
1	**0.121714**	**18.26**	**23**
2	**0.200829**	**30.12**	**29**
3	**0.220912**	**33.14**	**31**
4	**0.182252**	**27.34**	**27**
5	**0.120286**	**18.04**	**13**
6	**0.066158**	**9.92**	**8**
≥7	**0.050966**	**7.65**	**13**

$\chi^2 = 6.975$ (df = 8–1–1 = 6)

p-value = 0.3231

Conclusion: Fail to Reject H0. No evidence against the Poisson Model.

Some Properties of the Poisson Distribution

Recall that if Y~Poisson(μ), the mean (Y) = μ and variance(Y) = μ.

1. As μ gets larger (>50) the Poisson distribution becomes approximately normal.

2. If Y has a Poisson distribution with mean μ, then the standard deviation of Y is $\sqrt{\mu}$ which can be estimated by $\sqrt{Y}$.

3. If Y_1 is Poisson with mean μ_1, and Y_2 is Poisson with mean μ_2 , then $Y_1 + Y_2$ is also Poisson with mean $\mu_1 + \mu_2$.

Normal Approximation CI for Poisson μ (Single Count)

Properties #1 and 2 can be combined to get an approximate confidence interval for μ from a single observed Y (provided Y is large – say, 50 or more)

$$y \pm z_{\alpha/2}\sqrt{y}$$

Example: You count y = 82 particles of radioactive decay in one hour. Assuming that radioactive decay is distributed according to the Poisson distribution, (a very good assumption) a 95% C. I. for the true mean rate of decay (μ) (per hour) is:

$$y \pm z_{0.025}\sqrt{y} \qquad \text{i.e.,} \quad 82 \pm 1.96\sqrt{82}$$

82 ± 17.7. The required CI is $(63.3, 99.7)$

Normal Approximation CI for Poisson μ (Multiple Counts)

Mulekick Example: The number of soldiers kicked to death by mules per corps per year was seen to be approximately Poisson distributed. The total number of deaths for the 20 years and 10 corps (n=200) was 122. If the death rates for individual corp-year's are Poisson with mean μ, then the death rate for the entire army must be Poisson with mean 200μ. (The sum of independent Poisson's is a Poisson - property 3 from previous slide).

We start with a 95% C.I. for 200μ, based on the total for the entire army, based solely on the total number of events (y=122):

$$y \pm z_{0.025}\sqrt{y}$$

$$122 \pm 1.96\sqrt{122} \qquad 122 \pm 21.65 \quad = \quad (100,144)$$

Divide by 200 to get a 95% CI for μ, the rate for a single corps-year.

$$(0.50,\ 0.72)$$

Mulekick Example (continued)

The above interval for μ depends heavily on the assumption that the data are Poisson (which seems supported in this example). However, if you are in doubt about the adherence to the Poisson distribution (a crazy mule killing several people, i.e. lack of independence), you might consider some alternatives:

Alternative #1: (when n is large, or μ not too small)
Recall that the raw data are actual counts from each army-corp for each of 10 years. There will be a total of 200 data values. Since n = 200 here, we might compute the usual t-interval based (despite the lack of normality) on the sample mean (mean of the 200 counts) and the sample standard deviation s (standard deviation of the 200 counts). In this example we get $\bar{y}$ = 0.61 and s = 0.78. So the t-interval is (df = 199)

$$0.61 \pm 1.972 \frac{0.78}{\sqrt{200}}; \qquad (0.501,\ 0.719),\ \text{same as above!}$$

Alternative #2: A bootstrap interval, like the one from CH 5.