Notation:

Par = population parameter of interest
Est = sample estimate of the parameter
SE(Est) = standard error of the estimate; a measure of the variability of the estimate
Dist^n = a probability distribution for finding confidence level cutoffs and p-values

Inferences for the population proportion:

1. One sample

Par = p
Est = $\hat{p}$

SE(Est) = $\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$ for confidence intervals SE(Est) = $\sqrt{\dfrac{p_0(1-p_0)}{n}}$ for a test of Ho: $p = p_o$

Dist^n = N(0,1)

Assumptions:

- SRS
- the population is at least 10 times as large as the sample
- for a test of H_0: $p = p_0$, the sample size n is so large that both np_0 and $n(1\text{-}p_0)$ are 10 or more
- for a confidence interval, n is so large that both $n\hat{p}$ and $n(1-\hat{p})$ are 15 or more.

a. Confidence Intervals:

$$\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

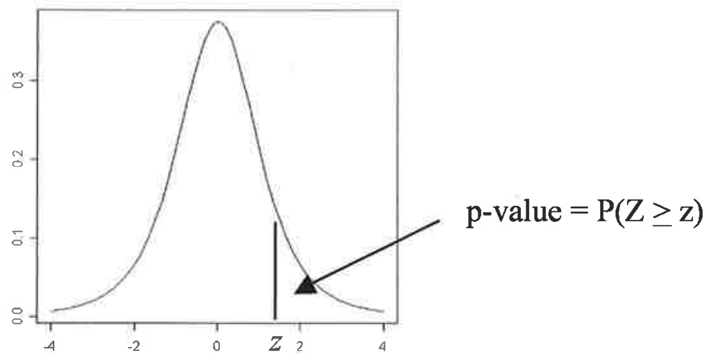
where z^* is the cutoff from an N(0,1) distribution corresponding to the confidence level of the confidence interval. Common values for z* are:

C	90%	95%	99%
z^*	1.645	1.960	2.576

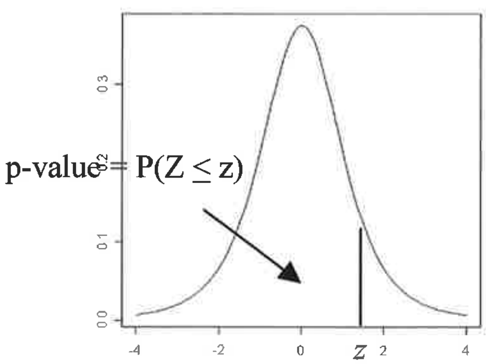
b. Hypothesis Tests:

The test statistic is $z = \dfrac{\hat{p} - p_0}{\sqrt{p_0(1 - p_0)/n}}$

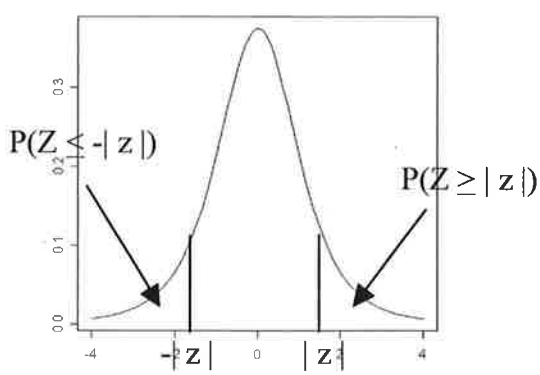
- Ho: $p = p_o$ vs. Ha: $p > p_o$ has p-value = P(Z ≥ z)



- Ho: $p = p_o$ vs. Ha: $p < p_o$ has p-value = P(Z ≤ z)



- Ho: $p = p_o$ vs. Ha: $p \neq p_o$ has p-value = P(Z ≤ -| z |) + P(Z ≥ | z |) = 2P(Z ≥ | z |) where | z | = absolute value of z



2. Two sample

Par = $p_1 - p_2$
Est = $\hat{p}_1 - \hat{p}_2$
SE(Est) = $\sqrt{\dfrac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \dfrac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$ for confidence intervals
SE(Est) = $\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}$ for a test of Ho: $p_1 - p_2 = 0$ where $\hat{p} = \dfrac{X_1 + X_2}{n_1 + n_2}$ is the overall proportion of successes (X_1 = number of successes in population 1 and X_2 = number of successes in population 2)

Dist^n = N(0,1)

Assumptions:
- Each samples is an SRS
- The samples are independent of each other
- the populations should be at least 10 times as large as the sample
- counts of successes (np) and failures ($n(1-p)$) are 10 or more in both samples for confidence intervals and 5 or more for hypothesis tests

a. Confidence Intervals:

$$\hat{p}_1 - \hat{p}_2 \pm z^* \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

where z^* is the cutoff from a N(0,1) distribution corresponding to the confidence level of the confidence interval.

b. Hypothesis Tests:

The test statistic is $z = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$

- Ho: $p_1 - p_2 = 0$ vs. Ha: $p_1 - p_2 > 0$ has p-value = $P(Z \geq z)$
- Ho: $p_1 - p_2 = 0$ vs. Ha: $p_1 - p_2 < 0$ has p-value = $P(Z \leq z)$
- Ho: $p_1 - p_2 = 0$ vs. Ha: $p_1 - p_2 \neq 0$ has p-value = $2P(Z \geq |z|)$

where the probabilities are calculated from a N(0,1) distribution.

Inference for Proportions
One Sample

Example 1

The National AIDS Behavioral Surveys interviewed a sample of adults in cities where AIDS is most common. This sample included 803 heterosexuals who reported having more than one sexual partner in the past year. We can consider this an SRS of size 803 from the population of all heterosexuals in high-risk cities who have multiple partners. These people risk infection with the AIDS virus. Yet 304 of the respondents said they never use condoms. Find a 95% confidence interval for the population proportion that never use condoms.

Assumptions for inference about a proportion:

- data are an SRS from the population of interest
- the population is at least 10 times as large as the sample
- for a test of H_0: $p = p_0$, the sample size n is so large that both np_0 and $n(1 - p_0)$ are 10 or more.
- for a confidence interval, n is so large that both $n\hat{p}$ and $n(1-\hat{p})$ are 15 or more.

Example 2

Can dogs detect cancer by smell?

A recent study investigated whether dogs can be trained to distinguish a patient with bladder cancer by smelling certain compounds released in the patient's urine. Ten dogs of varying breeds were trained to discriminate between urine from patients with bladder cancer and urine from control patients without it. The dogs were taught to indicate which among several specimens was from the bladder cancer patient by lying beside it.

An experiment was conducted to analyze how the dog's ability to detect the correct urine specimen compared to what would be expected with random guessing. Each of the ten dogs was tested with nine trials. In each trial, one urine sample from a bladder cancer patient was randomly placed among six control urine samples. In the total of 90 trials with the ten dogs, the dogs made the correct selection 37 times.

Did this study provide strong evidence that the dogs' predictions were better or worse than with random guessing? Set-up and test the appropriate hypotheses. Is a confidence interval consistent with the results of the hypothesis test?

Inference for Proportions
Two Sample

1) Are urban students more successful?

North Carolina State University looked at the factors that affect the success of students in a required chemical engineering course. Students must get a C or better in the course in order to continue as the chemical engineering majors. There were 65 students from urban or suburban backgrounds, and 52 of these students succeeded. Another 55 were from rural or small-town backgrounds with 30 of these students succeeding in the course.

Is there good evidence that the proportion of students who succeed is different from urban/suburban versus rural/small-town backgrounds?

Give a 90% confidence interval for the size of the difference.

Assumptions:

- 2 independent SRSs
- the populations should be at least 10 times as large as the sample
- counts of successes (np) and failures ($n(1 - p)$) are 10 or more in both samples for confidence intervals and 5 or more for hypothesis tests

2) Effect of Aspirin on Heart Attacks

A 1988 study on the effects of aspirin on the incidence of heart attacks gives the data in the table below. We are interested in whether taking aspirin reduces the incidence of heart attacks. To test this, compare the proportion of patients who had a heart attack under the two treatments.

	Heart Attack	No Heart Attack	Total
Aspirin	104	10,933	11,037
Placebo	189	10,845	11,034
Total	293	21,778	22,071