Comparing Proportions

Many questions take the form of comparing two proportions:

- Is support for gun registration higher in Quebec than in Ontario?
- Which of two surgical procedures has the higher survival rate?
- How much can the fraction of defective items manufactured be reduced by adjusting the production machinery more often?

These questions involve two proportions, π_1 and π_2. We may wish to test $H_0 : \pi_1 = \pi_2$, or we may want a confidence interval for $\pi_1 - \pi_2$.

We will try to answer these questions based on independent samples from the two populations.

The Difference of Sample Proportions

The observed proportions in the two samples, of sizes n_1 and n_2, will be called p_1 and p_2.

An obvious estimate for $\pi_1 - \pi_2$ is $D = p_1 - p_2$. What is its sampling distribution?

$$\begin{aligned} \mu_D &= \mu_{p_1} - \mu_{p_2} = \pi_1 - \pi_2 \\ \sigma_D^2 &= \sigma_{p_1}^2 + \sigma_{p_2}^2 \\ &= \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2} \end{aligned}$$

When n_1 and n_2 are large, p_1 and p_2 will have approximately normal distributions, as will D.

We will use the distribution of D to find confidence intervals and P-values.

Hypothesis Test for $H_0 : \ \pi_1 = \pi_2$

We might wish to test $H_0 : \ \pi_1 = \pi_2$ versus $H_a : \ \pi_1 \neq \pi_2$, using $D = p_1 - p_2$ as the test statistic.

However, even assuming H_0, we don't know the distribution of D, since we don't know what π_1 and π_2 are (just that they are equal).

We will have to use an estimate for π_1 and π_2, which if H_0 is true can be based on the proportion from both samples:

$$p = \frac{\text{total successes}}{\text{total sample size}} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

From this, we estimate

$$\sigma_D = \sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}} = \sqrt{p(1-p)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

We can now do a z test of whether D is big enough for us to reject H_0.

Confidence Interval for $\pi_1 - \pi_2$

We might find a confidence interval for $\pi_1 - \pi_2$ instead of doing a hypothesis test.

Since we then aren't using a null hypothesis in which $\pi_1 = \pi_2$, we will estimate π_1 and π_2 separately, by the sample proportions, p_1 and p_2. This gives the following estimate for the standard error of $p_1 - p_2$:

$$SE = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

The level C confidence interval is then computed as

$$(p_1 - p_2 - z^* SE, \ p_1 - p_2 + z^* SE)$$

where z^* is the point such that the area under the standard normal curve between $-z^*$ and z^* is C.

Comparing Two Means

We have looked so far at "one sample" hypothesis tests, for a *single* mean, such as:

$$H_0: \quad \mu = 0, \quad H_a: \quad \mu \neq 0$$

Similarly, the confidence intervals we have found are for a single mean parameter.

What if we are interested in comparing two means?

- Do Catholics and Protestants in Ontario contribute different amounts to charity, on average?
- Does taking calcium reduce blood pressure?
- Is some of the Vitamin C in potatoes destroyed by cooking them?

Here, we have hypotheses such as

$$H_0: \quad \mu_1 = \mu_2, \quad H_a: \quad \mu_1 \neq \mu_2$$

or maybe $H_a : \mu_1 > \mu_2$ or $H_a : \mu_1 < \mu_2$.

Comparing Means with Matched Pairs

Sometimes we can compare means using *matched pairs*. For example, to test if cooking potatoes destroys vitamin C, we could

1. Obtain a sample of n potatoes.
2. Cut each potato into two similar pieces.
3. Randomly choose one of the two pieces to cook, one not to cook.
4. Measure the vitamin C content of each piece.
5. For each potato, find the difference of the amount of vitamin C in the cooked piece minus the amount in the uncooked piece.

We then test whether the mean difference is zero, versus it being negative. We can do this with the "one sample" procedures. The light speed example was actually of this sort.

Comparing Means with Two Samples

We can't always use matched pairs. It would be hard to match up Catholic and Protestant charity contributors, for example. We may not even have the same number of each.

Instead, we compute sample means, $\bar{y}_1$ and $\bar{y}_2$, separately from two samples, of size n_1 and n_2. We can then estimate $\mu_1 - \mu_2$ by $\bar{y}_1 - \bar{y}_2$.

We would like to find the sampling distribution of $\bar{y}_1 - \bar{y}_2$.

We know that its mean is $\mu_1 - \mu_2$.

If the two samples are independent, we know that its variance is

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

If we know σ_1 and σ_2, we can now do a hypothesis test, or compute a confidence interval for $\mu_1 - \mu_2$.

What if We Don't Know σ_1 and σ_2?

Usually we don't know σ_1 and σ_2. Instead, we use the sample standard deviations, s_1 and s_2.

To test $H_0 : \mu_1 = \mu_2$ we use the two-sample t statistic:

$$t = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$$

Unfortunately, the sampling distribution of this is hard to figure out, but it is approximately a t distribution with some degrees of freedom that can be calculated from s_1, s_2, n_1, and n_2. (The formula is on page 561 of the book.)

A rough (and conservative) approximation can be obtained by using the smaller of $n_1 - 1$ and $n_2 - 1$ as the degrees of freedom. But it's easiest to just use Minitab.

Example: Calcium and Blood Pressure

Recall the data from a randomized experiment we looked at regarding the effect of calcium on systolic B.P. in 21 Black men:

B.P. before	Treatment received	B.P. after	Change in B.P.
107	Calcium	100	-7
110	Calcium	114	+4
123	Calcium	105	-18
129	Calcium	112	-17
112	Calcium	115	+3
111	Calcium	116	+5
107	Calcium	106	-1
112	Calcium	102	-10
136	Calcium	125	-11
102	Calcium	104	+2
123	Placebo	124	+1
109	Placebo	97	-12
112	Placebo	113	+1
102	Placebo	105	+3
98	Placebo	95	-3
114	Placebo	119	+5
119	Placebo	114	-5
112	Placebo	114	+2
110	Placebo	121	+11
117	Placebo	118	+1
130	Placebo	133	+3

Two-Sample t Test for B.P. After

Do men who receive calcium have a lower blood pressure after treatment than men who receive a placebo?

We can test this directly, with a t test comparing the mean B.P. after for the two groups. Here is the Minitab output for a two-sided test:

```
Two sample T for after

treatmen     N      Mean     StDev   SE Mean
Calcium     10    109.90      7.80       2.5
Placebo     11     113.9      11.3       3.4

95% CI for mu (Calcium) - mu (Placebo): ( -12.9,  4.9)
T-Test mu (Calcium) = mu (Placebo) (vs not =):
   T = -0.95  P = 0.35  DF = 17
```

From this test, there seems to be no evidence of a difference.

Two-Sample t Test for Change in B.P.

A more powerful way to test the null hypothesis of no effect is to look at the *change* in B.P. in the two groups. This eliminates some of the random variation from one subject to the next.

Here's the Minitab output for a two-sided test of a difference in the mean change:

```
Two sample T for change

treatmen     N      Mean     StDev   SE Mean
Calcium     10     -5.00      8.74       2.8
Placebo     11      0.64      5.87       1.8

95% CI for mu (Calcium) - mu (Placebo): ( -12.6,  1.4)
T-Test mu (Calcium) = mu (Placebo) (vs not =):
   T = -1.72  P = 0.11  DF = 15
```

The P-value of 0.11 still gives us almost no reason to doubt the null hypothesis.

Should the Test be One-Sided?

These P-values are for a two-sided test. If we do a one-sided test of

H_0 : mean change is the same for both groups

H_a : mean change is less for the calcium group

the P-value will be half as large:

```
Two sample T for change

treatmen     N      Mean     StDev   SE Mean
Calcium     10     -5.00      8.74       2.8
Placebo     11      0.64      5.87       1.8

95% CI for mu (Calcium) - mu (Placebo): ( -12.6,  1.4)
T-Test mu (Calcium) = mu (Placebo) (vs <):
   T = -1.72  P = 0.053  DF = 15
```

Is a one-sided test appropriate? Do the researchers know that calcium can't increase blood pressure? Would they ignore a result that seemed to show that it does?