# POL 345: Quantitative Analysis and Politics

Precept Handout 9

Week 11 (Verzani Chapter 10: 10.1–10.2)

Remember to complete the entire handout and submit the precept questions to the Blackboard 24 hours before precept. In this handout, we cover the following new materials:

- Using `prop.test()` to conduct hypothesis testing for two-sample proportions
- Using `t.test()` to conduct two-sample hypothesis testing under the Student's $t$-Distribution
- Using `power.prop.test()` to calculate the power of a two-sample hypothesis test for proportions
- Using `power.t.test()` to calculate the power of a two-sample hypothesis test under the Student's $t$-Distribution
- Using `var()` to calculate the variance

# 1 Two-Sample Hypothesis Testing for Proportions

- In many cases, it is of interest to investigate the difference in the means of two populations. We can use the difference in sample means as our estimate of the difference in population means.

$$D = \overline{X}_1 - \overline{X}_2$$

where $\overline{X}_j$ is the mean of sample $j$ for $j = 1, 2$. The standard error is given by the square root of the sum of two variances

$$\text{s.e.} = \sqrt{\frac{\text{Var}(X_1)}{n_1} + \frac{\text{Var}(X_2)}{n_2}}$$

where $\text{Var}(X_j)$ is the variance of $X$ in population $j$ and $n_j$ is the size of sample $j$ for $j = 1, 2$. For example, if we are testing the difference in two population proportions,

$$\text{s.e.} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}$$

Using these standard errors, we can easily compute the $(1-\alpha) \times 100\%$ confidence interval as

$$[D - \text{s.e.} \times z_{\alpha/2},\ D + \text{s.e.} \times z_{\alpha/2}]$$

where the critical value $z_{\alpha/2}$ can be calculated via `qnorm`$(1 - \alpha/2)$ as before.

- Here, the null hypothesis is that two populations have the identical mean: $H_0 : p_1 = p_2$. The alternative hypothesis could be two-sided, i.e., $H_1 : p_1 \neq p_2$, or one-sided, i.e., $H_1 : p_1 > p_2$ or $H_1 : p_1 < p_2$, depending on prior knowledge.

- Hypothesis testing will proceed in the same way by deriving the sampling distribution under the null hypothesis. For example, when testing the equality of the two population proportions, we have the following approximate sampling distribution (for a sufficiently large sample size),

$$Z = \frac{D - 0}{\text{s.e.}} = \frac{D}{\sqrt{\frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}}} \sim \mathcal{N}(0, 1)$$

where $p = p_1 = p_2$ is the population proportion under the null hypothesis.

- **Example: Public Opinion and the Iraq War II.** We return to Gardner's data on public opinion and the Iraq war used in precept handout 8. Recall that the data set contains information on each respondent's gender as well as whether s/he felt that the war in Iraq was a mistake. We calculate the 95% confidence interval of the difference in the average attitude among women versus men. Additionally, we calculate a two-sided hypothesis test where the null hypothesis is that the proportion of women who felt the war was a mistake is the same as the proportion of men who felt in the same way. The alternative is that those two proportions are different.

```
> load("iraq.RData")
> ## Calculate the mean of attitude toward the war among women
> mean.women <- mean(iraq$mistake[iraq$female == 1])
> n.women <- nrow(iraq[iraq$female == 1,])
> ## Calculate mean and sample size for men
```

```
> mean.men <- mean(iraq$mistake[iraq$female == 0])
> n.men <- nrow(iraq[iraq$female == 0,])
> ## Calculate Difference in Means and Standard Error
> D <- mean.women - mean.men
> se <- sqrt(mean.women * (1 - mean.women) / n.women
+            + mean.men * (1 - mean.men) / n.men)
> ## Compute 95% confidence interval
> lower <- D - se * qnorm(0.975)
> upper <- D + se * qnorm(0.975)
> lower; upper

[1] -0.114693

[1] 0.0257888

> ## Estimated proportion under the null
> p.null <- mean(iraq$mistake) ## Equal support under the null
> var.null <- p.null * (1-p.null) ## Variance follows from equal support assumption
> se.null <- sqrt(var.null / n.men + var.null / n.women)
> z <- D / se.null
> z ## z-value

[1] -1.25650

> 2*pnorm(abs(z), lower.tail = FALSE) ## Two-sided p-value

[1] 0.208934
```

- As with the one-sample test, we may use `prop.test()` to test for difference in proportions for two samples. For a two-sample test, `x` and `n` will be entered as vectors. The command will automatically calculate the proportion from the sum of observations meeting the condition (i.e. taking a value of 1) for each group and the size of the group.

```
> nw.mistake <- sum(iraq$mistake[iraq$female == 1])
> nm.mistake <- sum(iraq$mistake[iraq$female == 0])
> prop.test(c(nw.mistake, nm.mistake), c(n.women, n.men),
+           conf.level = 0.95, alternative = "two.sided", correct = FALSE)

        2-sample test for equality of proportions without continuity
        correction

data:  c(nw.mistake, nm.mistake) out of c(n.women, n.men)
X-squared = 1.5788, df = 1, p-value = 0.2089
alternative hypothesis: two.sided
95 percent confidence interval:
 -0.1146933  0.0257888
sample estimates:
  prop 1   prop 2
0.149425 0.193878
```

While we cannot reject the null hypothesis that mean support rates are equal, our findings are still somewhat informative. The results contrast with Gardner's hypothesis. If anything, women likely express less antagonism towards the war; a lower percentage think that the invasion into Iraq was a mistake.

# 2 The Power of a Two-Sample Test for Proportions

- The command `power.prop.text(n, p1, p2, sig.level, power, alternative, strict)` may be used to calculate the power of a two-sample test for proportions. Note that this command may only be used to calculate power for a two-sample test. Further, the command assumes equal sample size for the two groups. The arguments of the command are as follows:
  - `n` - number of observations (per group); assumed to be equal
  - `p1` - probability in group 1
  - `p2` - probability in group 2
  - `sig.level` - significance level (Type I error probability)
  - `power` - power of test (1 - Type II error probability)
  - `alternative` - one- or two-sided test
  - `strict` - includes the probability of rejection in the opposite direction of the true effect
- We revisit Gardner's data and calculate the power of the statistical test by assuming that the true proportion of women (men) who believe the war was a mistake equals the sample proportion (of course, you can assume the true values that are different from the sample values).

```
> p.women <- mean(iraq$mistake[iraq$female == 1])
> p.men <- mean(iraq$mistake[iraq$female == 0])
> ## calculate the power
> power.prop.test(n = n.women, p1 = p.women, p2 = p.men, sig.level = 0.05,
+                  power = NULL, alternative = "two.sided", strict = TRUE)

     Two-sample comparison of proportions power calculation

              n = 261
             p1 = 0.149425
             p2 = 0.193878
      sig.level = 0.05
          power = 0.269958
    alternative = two.sided

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

- We may also conduct the same power analysis by hand as you may need to do in the quiz.

```
> p.null <- (p.women + p.men) / 2
> se.null <- sqrt(2 * p.null * (1 - p.null)/ n.women)
> thld.low <- -qnorm(0.975) * se.null
```

```
> thld.high <- qnorm(0.975) * se.null
> ## Step 2: calculate the power
> se.true <- sqrt(p.women * (1 - p.women) / n.women +
+                  p.men * (1 - p.men) / n.women)
> power.low <- pnorm(thld.low, p.women - p.men, se.true)
> power.high <- pnorm(thld.high, p.women - p.men, se.true, lower.tail = FALSE)
> total.power <- power.low + power.high
> total.power

[1] 0.269958
```

We find that the power of the test is very low - about 27%.

- As with a one-sample test, we may calculate the number needed in each group to achieve a given level of power by specifying `power` rather than `n`. For instance, we may calculate the sample size needed in each group to reach a power of 90%.

```
> power.prop.test(p1 = p.women, p2 = p.men, sig.level = 0.05, power = 0.90,
+                  alternative = "one.sided")

     Two-sample comparison of proportions power calculation

              n = 1230.58
             p1 = 0.149425
             p2 = 0.193878
      sig.level = 0.05
          power = 0.9
    alternative = one.sided

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

We find that we need at least 1,231 in each group to reach a power of 90%.

# 3 Two-Sample Hypothesis Testing for Continuous Variables

- If the distributions of both populations are assumed to be normal, one can conduct a **two-sample *t*-test**. This can be done easily using the `t.test()` command we learned earlier. In particular, we use the command in the following manner, `t.test(x, y, alternative, conf.level)` where `y` indicates the vector of the second sample.

- **Example: Civil War and GDP growth.** We turn next to the relationship between civil war and GDP. A broad range of literature in International Relations and Comparative Politics explores the relationship between civil war and economic growth. The literature has argued that the incidence of civil war is often associated with low levels of development and economic growth. We rely on the `growth.RData` data set. The data set contains the following variables:
  - `year` - year of observation
  - `country` - country

- `war` - binary indicator of civil war (1 if the country experienced civil war in the observed year, 0 else)
- `gdppc` - GDP per capita for the year of observation
- `gdppc.lag` - GDP per capita from the previous year
- `growth.rate` - GDP per capita growth rate, in percentage points

To begin, we wish to consider the association between civil war and levels of economic development. As such, we wish to compare the GDP per capita between countries that experienced a war versus those that did not. We would expect those countries that experience war to have levels of economic development (i.e.: measured as GDP per capita) that are significantly lower than countries that do not experience war. Due to the skew in the data, we must log GDP per capita.

```
> load("growth.RData") ## Load data
> ## Two-sided test using log of the variables to ensure
> ## sample distribution is closer to normal distribution
> ## two-sided test
> t.test(log(growth$gdppc[growth$war == 1]), log(growth$gdppc[growth$war == 0]),
+        alternative = "two.sided", conf.level = 0.95)

        Welch Two Sample t-test

data:  log(growth$gdppc[growth$war == 1]) and log(growth$gdppc[growth$war == 0])
t = -19.2408, df = 1184.07, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.745573 -0.607591
sample estimates:
mean of x mean of y
 0.264058  0.940640

> ## one-sided test
> t.test(log(growth$gdppc[growth$war == 1]), log(growth$gdppc[growth$war == 0]),
+        alternative = "less", conf.level = 0.95)

        Welch Two Sample t-test

data:  log(growth$gdppc[growth$war == 1]) and log(growth$gdppc[growth$war == 0])
t = -19.2408, df = 1184.07, p-value < 2.2e-16
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf -0.618697
sample estimates:
mean of x mean of y
 0.264058  0.940640
```

We find that levels of development are indeed significantly different (and specifically, lower) among countries that experience war versus those that do not. However, it is important to note that we cannot draw conclusions about causality due to the existence of potential confounders.

More recent literature has emphasized that civil war is more closely associated with a low growth rate. We next compare GDP growth per capita between countries engaged in civil war versus those that were not. We conduct a **two-sample t-test** with the null hypothesis that there was no difference in GDP growth per capita between countries engaged in civil war versus those that were not. The alternative hypothesis is that the economic growth rate is less among countries that experience war, as compared to those that do not.

```
> t.test(growth$growth.rate[growth$war == 1],
+        growth$growth.rate[growth$war == 0],
+        alternative = "less", conf.level = 0.95)

        Welch Two Sample t-test

data:  growth$growth.rate[growth$war == 1] and growth$growth.rate[growth$war == 0]
t = -3.9555, df = 889.488, p-value = 4.122e-05
alternative hypothesis: true difference in means is less than 0
95 percent confidence interval:
      -Inf -0.860111
sample estimates:
mean of x mean of y
 0.596103  2.069583
```

We can approximate this result using the normal distribution. We may use the command `var()` to find the variance of a given variable:

```
> D <- mean(growth$growth.rate[growth$war == 1]) -
+   mean(growth$growth.rate[growth$war == 0])
> ## Calculate Standard Error
> n.war <- length(growth$growth.rate[growth$war == 1])
> n.nowar <- length(growth$growth.rate[growth$war == 0])
> se <- sqrt(var(growth$growth.rate[growth$war == 1])/n.war +
+            var(growth$growth.rate[growth$war == 0])/n.nowar)
> ## Calculate z-score
> z <- D/se
> pnorm(z)

[1] 3.81861e-05
```

As expected, we are able to reject the null hypothesis of no difference in GDP per capita between countries engaged in civil war in the previous year versus those that were not.

# 4 The Power of a Two-Sample Test for Continuous Variables

- As with a one-sample test, we can calculate the power of a two-sample test using the command `power.t.test(n, delta, sd, sig.level, power, type, alternative, strict)`. The arguments are as follows:
  - `n` is the number of observations;

- `delta` is the true difference in means;
- `sd` is the standard deviation within the population;
- `sig.level` is the test's level of significance (Type I error probability);
- `type` is the type of $t$-test (`"two.sample"`, `"one.sample"` or `"paired"`);
- `alternative` specifies a direction of the test (`"two.sided"` or `"one.sided"`);
- `strict = TRUE` allows for null hypothesis to be rejected by data in the opposite direction of the truth. The default is `strict = FALSE`

We calculate the power of the hypothesis test conducted above by assuming that the standard deviation is equal to the sample standard deviation and that the difference in growth rate between the war and non-war periods equals 0.5. Note that both the sample size and the standard deviation are assumed to be identical between the two groups.

```
> n <- length(growth$growth.rate[growth$war == 1])
> sd <- sd(growth$growth.rate)
> power.t.test(n = n, delta = 0.5, sd = sd, type = "two.sample",
+              alternative = "one.sided", sig.level = 0.05)

     Two-sample t test power calculation

              n = 757
          delta = 0.5
             sd = 7.39197
      sig.level = 0.05
          power = 0.370896
    alternative = one.sided

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

We may again approximate this with the normal approximation.

```
> se <- sqrt(2 * sd^2 / n)
> thld.null <-  qnorm(0.95) * se
> ## Calculate power
> power.norm <- pnorm(thld.null, 0.5, se, lower.tail = FALSE)
> power.norm

[1] 0.371118
```

We find that the power of the test is just over 35%. Again, we may calculate the sample size needed in each group to reach a specified level of power.

```
> power.t.test(power = 0.90, delta = 1, sd = sd(growth$growth.rate),
+              type = "two.sample", alternative = "one.sided", sig.level = 0.05)
```

```
     Two-sample t test power calculation

              n = 936.555
          delta = 1
             sd = 7.39197
      sig.level = 0.05
          power = 0.9
    alternative = one.sided

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

We find that we need 937 states in each group to reach a power of 90%.

# 5 Precept Questions

We again return to the `chechen.RData` data set from Jason Lyall's article "Does Indiscriminate Violence Incite Insurgent Attacks? Evidence from Chechnya" (*Journal of Conflict Resolution, 2009*). Recall that for the first problem set, we conducted an initial analysis of the difference in differences. That is, calculated the difference in the difference of pre-shelling and post-shelling insurgency attacks among shelled versus non-shelled villages.

1. Begin by calculating the difference in differences, as done in the first problem set. Next, conduct a one-sided, two-sample hypothesis test where the null hypothesis is that the average effect of shelling on the frequency of insurgency attacks is zero (i.e. the difference in attacks among both shelled and non-shelled villages is identical) and the alternative hypothesis is that the difference in insurgency attacks is greater among shelled versus non-shelled villages. Note that the direction of your alternative hypothesis must be determined according to your calculation of the `diffattack` variable.

2. Calculate the sample size needed for the power of the above test to reach 0.9 by assuming that the true average effect is 0.25 and the standard deviation is equal to the sample standard deviation.