PO11Q: Introduction to Quantitative Political Analysis I

Worksheet Week 9 - Solutions

Dr Florian Linke

F.Linke@warwick.ac.uk

The t-distribution

1. How df changes the shape and critical values
   Using two-tailed, $\alpha = 0.05$ (so $t_{\text{crit}} = t_{0.025,df}$ with $df = n - 1$):

   - $n = 3$ ($df = 2$): $t_{\text{crit}} \approx 4.303$
   - $n = 5$ ($df = 4$): $t_{\text{crit}} \approx 2.776$
   - $n = 10$ ($df = 9$): $t_{\text{crit}} \approx 2.262$
   - $n = 20$ ($df = 19$): $t_{\text{crit}} \approx 2.093$
   - $n = 50$ ($df = 49$): $t_{\text{crit}} \approx 2.009$

   The normal reference is $z_{0.025} \approx 1.960$.
   Within 0.05 of 1.96 occurs by $n \approx 50$ ($df = 49$, 2.009 is 0.049 above 1.96).

   Explanation: as $df \to \infty$, the t distribution $\to N(0, 1)$, so tails thin and t_{crit} decreases toward 1.96.

2. One-tailed vs two-tailed critical regions
   With $df = 14$ and $\alpha = 0.05$:

   - Two-tailed: $t_{0.025,14} \approx 2.145$, cut points at ± 2.145
   - One-tailed (upper): $t_{0.05,14} \approx 1.761$, single cut point at $+1.761$

   For $H_0 : \mu = \mu_0$ vs $H_1 : \mu > \mu_0$, the one-tailed threshold $t_{0.05,14}$ is relevant because only the upper tail provides evidence against H_0.

   In both cases, the shaded area equals α:

   - two-tailed shades two symmetric regions each of area $\alpha/2$;
   - one-tailed shades a single upper tail of area α.

Effect Size, sample size, and power

1. Same effect size, different n (mapping t, n, and $\hat{d}$)
   a. With $t = 1.6$ and $n = 16$, $\hat{d} = t/\sqrt{n} = 1.6/4 = 0.40$. The power at $n = 16$ for $\hat{d} \approx 0.40$ and $\alpha = 0.05$ will be modest (typically well below 0.80).
   b. Keeping $\hat{d}$ fixed at 0.40 and raising n to 64 requires $t = \hat{d}\sqrt{n} = 0.40 \times 8 = 3.2$. The app will show the same $\hat{d}$ but a higher power at $n = 64$.
   c. As n increases while the underlying effect stays fixed, the standard error shrinks: $\text{SE} = s/\sqrt{n}$. The test statistic is $t = \dfrac{\bar{x} - \mu_0}{\text{SE}}$, so a smaller SE makes $|t|$ larger on average. For a fixed α, the critical cutoff (e.g., $t_{\alpha/2,df}$ for a two-sided test) is essentially fixed, so larger typical $|t|$ increases the chance that $|t| > t_{\alpha/2,df}$. Therefore, bigger $n \Rightarrow$ less sampling noise $\Rightarrow$ tighter estimates $\Rightarrow$ higher power.

2. Planning with a SESOI and α sensitivity
   a. For $t = 2.0$ and $n = 30$, $\hat{d} = t/\sqrt{n} \approx 2.0/\sqrt{30} \approx 0.37$. The app will report the power at $n = 30$ using this $\hat{d}$ (typically moderate at $\alpha = 0.05$).
   b. Turning on the SESOI with $d = 0.5$ switches the curve to a fixed target effect. The orange marker shows the n giving 80% power at $\alpha = 0.05$; for $d = 0.5$ this is typically in the few–dozen range for a one-sample t-test (on the order of the 30s).
   c. Increasing α to 0.10 lowers the critical threshold and reduces the required n for a given power; decreasing α to 0.01 raises the threshold and increases the required n. Formally, stricter α increases the critical t value, so a larger n is needed for the same probability of exceeding it under the alternative.

3. Was the study well powered? Post-hoc check and replication planning
   a. With $t = 2.1$ and $n = 25$, $\hat{d} = 2.1/\sqrt{25} = 2.1/5 = 0.42$. The app will show the power at $n = 25$ for $\hat{d} \approx 0.42$ and $\alpha = 0.05$; this is typically only moderate, not comfortably high.
   b. For replication planning, turn on SESOI and choose the closest option to $\hat{d}$ (e.g., $d = 0.5$). The orange marker will give the n needed for 80% power at the chosen α; expect a sample size in the few–dozen range for $d = 0.5$ at $\alpha = 0.05$.
   c. A significant result with low power can be fragile because small shifts in sampling may miss the effect and estimates are noisier. Picking a SESOI (a target d tied to practical importance) and planning for 80% power helps ensure a replication has adequate sensitivity to detect a meaningfully sized effect.

Summarising Conclusions

- Power is the probability your test will detect a real effect (reject H_0 when the effect truly exists).
- The curve fixes an effect size d and shows how power increases with n.
- If you base d on your observed test ($\hat{d} = t/\sqrt{n}$), the curve answers: "If the effect really was the size we observed, how likely would the test be to detect that effect at other sample sizes n?"
- If you base d on a meaningful minimum (SESOI), the curve answers: "How large should n be to reliably detect an effect that actually matters?"
- The 0.80 (80%) line is a convention (not a law). Many fields use it as a planning minimum; higher power (e.g., 90%) is desirable when stakes are high.
- Rule of thumb: because $t = d\sqrt{n}$, halving d requires about 4× the sample size to keep power roughly the same.

Why do many plots show an 80% power line?

- 80% power means choosing $\beta = 0.20$ (so power $= 1 - \beta = 0.80$). This is a planning convention, not a mathematical law.
- The convention became common through methodological guidance (e.g., Cohen's power analysis texts) and is echoed in many applied fields.
- In higher-stakes settings (e.g., confirmatory clinical trials), targets of 80–90% power are typical; 90