Statistical Thinking for Forensic Practitioners

Quiz on Part 6: Inference

1 Sample size calculations

Collecting data can be expensive and time-intensive. As such, we may want to know how large a sample needs to be to accomplish the goals of an analysis. Consider the following situations and determine which sample size calculation is needed to accomplish the goals of the analysis. (Note: the 4 sample size calculation methods discussed in lecture can be found on slides 37-45 and slide 93. As some of these exercises involve applying formulas, it may be useful to perform your calculations in Excel. In particular, the `NORM.S.INV` function will be useful. Remember that for a confidence level of $(1-\alpha)$, we need to consider the $\left(1-\frac{\alpha}{2}\right)$ standard normal quantile in calculating critical values.)

1. A bloodstain pattern researcher is interested in the average identifying the average diameter of a droplet in a stain caused by a firearm. Based on a previous study, they initially estimate the average diameter to be 1.43 mm with standard deviation .36 mm. They would like to estimate the true average diameter to within 5% with 90% confidence. (Note: remember for a $(1-\alpha)$% confidence level to consider a critical value of $z_{1-\alpha/2}$.)

   Relative margin of error for a population mean: $n = \left(\frac{\sigma}{\mu}\right)^2 \left(\frac{z_\alpha}{d}\right)^2 = 68.588$ **rounding to 3 decimal places. To be safe, the researcher should consider 69 bloodstains patterns.**

2. An examiner receives 40 individual fibers recovered from a crime scene to analyze using a time-intensive X-ray fluorescence microanalysis procedure. Rather than considering all of these fibers, the examiner would like to know how many fibers they need to analyze to conclude with 95% certainty that at least 70% of the 40 fibers match, assuming 2 of the fibers sampled turn out to be non-matches. (Source)

   Considering the table on slide 45, we have $N = 40$, $k = .7$, 2 negative (non-matches), and 95% confidence. Then $n = 15$.

3. A statistician working for the U.S. Department of Justice is interested in estimating the proportion of 2020 criminal cases for which the defendant is sentenced to prison that involve a firearm. In particular, they want to estimate this proportion with a 99% confidence level. Using an estimate from 2016 that 21% of all state and federal prisoners reported that they possessed or carried a firearm when they committed the offense for which they are serving time, the statistician wants their margin of error to be .07. How many cases should the statistician consider to achieve the desired goals? (Source)

   Margin of error for a population proportion π: $n = \left(\frac{z_\alpha}{ME}\right)\hat{p}(1-\hat{p}) = \left(\frac{2.576}{.07}\right)^2 (.21)(1-.21) = 224.639$ **rounding to 3 decimal places. To be safe, the statistician should consider 225 cases.**

4. Glass analysts are interested in performing a hypothesis tests to determine whether the average concentration of Calcium in building float glass is different than in vehicle float glass. They want to detect an effect size of .5 (ppm) with 85% power and 95% confidence. How many samples of glass of each type should they consider?

   Sample size based on effect size: $n_i = \frac{2(Z_{\alpha/2}+Z_{1-\beta})}{d^2} = \frac{2(1.96+1.036)^2}{(.5)^2} = 71.827$ **rounding to 3 decimal places. To be safe, the analysts should consider 72 samples of each type.**

2 Comparing proportions

We will practice hypothesis testing procedures to answer questions about population proportions. As some of these exercises involve applying formulas, it may be useful to perform your calculations in Excel. You may even want to work through the Part 6 lab prior to completing these exercises to practice using Excel to perform hypothesis tests. Slides 108-120 will be useful to reference.

Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) is a risk assessment tool that predicts how likely a defendant is to recitivate (reoffend). In particular, a numerical score quantifying the "risk of recidivism" is assigned to a defendant. It has been used by some jurisdictions to inform parole decisions. A 2016 ProPublica article criticized the tool for categorically assigning a higher risk to Black defendants compared to White defendants (in fact, they go even further and argue that the tool *mis*classifies Black defendants as being higher risk based on follow-up recidivism data). We will analyze the findings in this article using hypothesis testing procedures. We will use a significance level of $\alpha = .05$.

The article authors aggregated the scores into "low risk " and "high risk" classifications. We will first consider whether the proportion of Black defendants classified as high risk is different from the proportion of all defendants classified as high risk (note: we are considering "different" from rather specifically "greater than" to be careful about the fact that we didn't formulate our research question until after collecting/observing the data). In a sample of 3,696 Black defendants, 2,174 were classified as high risk. Assume that the sample is representative of all Black defendants for whom COMPAS assigned a score. Also assume that 46% of all defendants were classified as high risk. Let π_1 be the proportion of all Black defendants that would be classified as high risk.

1. Formulate the null and alternative hypotheses H_0 and H_a using similar notation as used on slide 109 of the lecture slides.

   $H_0 : \pi_1 = 0.46$ **vs.** $H_a : \pi_1 \neq 0.46$

2. Calculate $\hat{p}_1$, the estimate of the π_1. Why can't we claim $\hat{p}_1 = \pi_1$?

   $\hat{p}_1 = \frac{2174}{3696} = .588$ **rounding to 3 decimal places. Because $\hat{p}_1$ is based on only a *sample* of Black defendants for whom COMPAS assigned a score, not the *population* of all Black defendants for whom COMPAS assigned a score, we cannot claim that it is equal to the population proportion π_1.**

3. Calculate the standard error $SE_{\hat{p}_1}$ of $\hat{p}_1$ calculated above.

   $SE_{\hat{p}_1} = \sqrt{\frac{\pi_0(1-\pi_0)}{n}} = \sqrt{\frac{.46(1-.46)}{3696}} = .008$ **rounding to 3 decimal places.**

4. Calculate the z statistic for this hypothesis test. Explain why a z statistic is used here instead of a t statistic.

   $z = \frac{\hat{p}_1 - \pi_0}{SE_{\hat{p}_1}} = \frac{.588-.46}{.008} = 15.663$ **rounding to 3 decimal places. We use a z statistic here because we assume the standard error (with respect the population proportion $\pi_0 = .46$) is known. If the standard error instead had to be estimated, then we would use a t statistic.**

5. Calculate the critical value for this z-test assuming $\alpha = .05$.

   $z^* = z_{.05/2} = 1.96$

6. Do you reject or fail to reject H_0? Explain what your decision means in the context of the problem.

   We reject H_0 because the z statistic is far outside of the "fail-to-reject" region of $[-1.96, 1.96]$. We conclude that there is significant evidence that the proportion of Black defendants classified as high risk is different from the proportion of all defendants classified as high risk.

7. Calculate the p-value associated with the z statistic calculated above. (Note: the `NORM.S.DIST` function with `cumulative` set to `TRUE` should be useful. This Excel function returns specifically the area under the normal curve on the interval $(-\infty, z]$ where z is the z-statistic. We desire either the area under the normal curve on $[z, \infty)$ or, by symmetry of the normal distribution, on $(-\infty, -z]$ for the value of z calculated in question 4. Keep this in mind when performing your calculations).

   Since we are performing a two-sided test, we will want to double the p-value obtained from plugging -15.663 into the `NORM.S.DIST` function. However, this statistic is so far from the center

of the normal distribution (0) that the resulting p-value Excel returns is 0. In reality, the p-value isn't 0 yet is so small that Excel can't represent it (this is referred to as machine precision).

8. Explain what a Type I and Type II error would be in the context of the problem.

   A Type I error would mean incorrectly rejecting the notion that proportion of Black defendants classified as high risk is equal to the proportion of all defendants classified as high risk. A Type II error would mean failing to reject the notion that the proportion of Black defendants classified as high risk is equal to the proportion of all defendants classified as high risk, when in reality they are different.

9. Construct a 95% confidence interval for the proportion of Black defendants classified as high risk.

   The 95% confidence interval is given by $\hat{p}_1 \pm 1.96 SE_{\hat{p}_1} = .5888 \pm 1.96(.008198) = [.572, .604]$ **rounding to 3 decimal places.**

Now consider that of the 2,454 White defendants sampled, 854 were classified as high risk.

10. Perform a hypothesis test to determine whether the proportion of Black defendants classified as high risk, π_1 still, is different from the proportion of White defendants classified as high risk, π_2 say. In your answer, include null and alternative hypotheses, a test statistic, a critical value, a p-value, and a conclusion of rejecting/failing to reject the null and what this means in the context of the problem. Use $\alpha = .05$ as a critical value. (Note: We shouldn't z test here as we don't assume knowledge of a population proportion. Refer to slides 113-117.)

    $H_0 : \pi_1 = \pi_2$ **vs.** $H_a : \pi_1 \neq \pi_2$ **(or, equivalently,** $H_0 : \pi_1 - \pi_2 = 0$ **vs.** $H_a : \pi_1 - \pi_2 \neq 0$**). The common population proportion estimate is** $\hat{p} = \frac{2174+854}{3696+2454} = .492$ **rounding to 3 decimal places. The** t **statistic is** $t = \frac{\hat{p}_1 - \hat{p}_2}{SE_{\hat{p}}} = \frac{\frac{2174}{3696} - \frac{854}{2454}}{\sqrt{.492(1-.492)}\sqrt{\frac{1}{3696}+\frac{1}{2454}}} = 18.451$**. The degrees of freedom for this test are** $df = n_1 + n_2 - 2 = 3696 + 2454 - 2 = 6148$**. Thus, the** t^* **critical value assuming a confidence level of 95% is 1.960 rounding to 3 decimal places (note that this is close to the** z **test critical value due to the high degrees of freedom – the** t **distribution "converges" to the standard normal distribution as the degrees of freedom grow). Very clearly, we would reject the null hypothesis and conclude that there is significant evidence of a difference between the proportion of Black and White defendants classified as high risk.**

Now suppose data were also collected on Hispanic defendants. We would be interested in determining whether the proportions among the three groups of defendants differ.

11. How many hypothesis tests would we need to perform to fully compare the three groups?

    3 tests – compare Black defendants to White defendants, Black defendants to Latino defendants, and White defendants to Latino defendants.

12. If we were to naively assume a Type I error rate of $\alpha = .05$ for each hypothesis tests, what would be the *family-wise* confidence level of all 3 hypothesis tests? (Hint: refer to slide 118)

    The family-wise confidence level would be $(1 - .05)^3 = .857$ **rounding to 3 decimal places.**

13. If we were to apply a Bonferroni correction to control the family-wise confidence level of all 3 hypothesis tests, what should be the corrected Type I Error rate *per* test? (Hint: refer to slide 119)

    For $m = 3$ **tests, we set the individual test level to be** $\alpha_B = \alpha/m = .05/3 = .017$ **rounding to 3 decimal places.**