Statistical Inference - Solutions

Exercise A - (5 min)

Two researchers carried out independent studies to answer the same research question. The first reports an effect estimate and standard error of 25 ± 10. The second reports 10 ± 10.

1. Are the results of the first study statistically significant at traditional significance thresholds?
2. What about the results of the second study?
3. Is there a statistically significant difference between the results of the studies?

Solution

1. The test statistic for a two-sided test of the null hypothesis of no effect is 25/10 = 2.5. This gives a p-value of `2 * (1 - pnorm(2.5)` $\approx$ 0.01, so the results are statistically significant at all "traditional" thresholds: 10%, 5%, and 1%.
2. Here the test statistic is 1 so the p-value is approximately 0.32. The results are not significant at any of the traditional thresholds.
3. The difference of effects is 25 - 10 = 15. Because the studies are independent, the variance of the difference is the sum of the variances. Thus, the standard error of the difference is $\sqrt{10^2 + 10^2} = 10\sqrt{2}$. Thus, the test statistic for a two-sided test of no difference between the studies is $15/(10\sqrt{2}) \approx 1.06$, yielding a p-value of around 0.29. The results of the first study are *highly* statistically significant, the results of the second study are *nowhere close* to significant, and yet there is *no statistically significant difference between the studies*.

Exercise B - (4 min)

True or False. If false, explain.

1. A small p-value indicates the presence of a large effect.
2. I tested the null that my treatment has no effect against the one-sided alternative that it is effective. My p-value was 0.01. Hence there is a 99% chance that the treatment is effective and a 1% chance that it is ineffective or harmful.

Solutions

1. False. A small p-value indicates that the estimated effect is large *relative to the standard error*. This can occur even if the effect size is minuscule, provided that the standard error is smaller still. On its own, a p-value tells us *nothing* about the size of an effect.
2. False. A p-value is the probability of observing a test statistic at least as extreme as the one we actually observed *assuming that the null is true*. It is not the probability that the null is false.

Exercise C - (6 min)

1. Suppose that $Z \sim \text{N}(0,1)$ and κ and c are constants. Write a line of R code to compute each of the following:
   a. $\mathbb{P}(Z + \kappa < -c)$
   b. $\mathbb{P}(Z + \kappa > c)$
   c. $\mathbb{P}(|Z + \kappa| > c)$
2. Suppose $\hat{\theta} \sim \text{N}(\theta, \text{SE}^2)$ and consider a test of $H_0\colon \theta = \theta_0$. If the null is *false* what is the distribution of the test statistic $T \equiv (\hat{\theta} - \theta_0)/\text{SE}$?

Solution

- **Part 1**:
  a. `pnorm(-c - kappa)`
  b. `1 - pnorm(c - kappa)`
  c. `pnorm(-c - kappa) + 1 - pnorm(c - kappa)`
- **Part 2**: Since $\left(\hat{\theta} - \theta\right)/\text{SE} \equiv Z \sim \text{N}(0,1)$

$$T = \frac{\hat{\theta} - \theta_0}{\text{SE}} = \frac{\hat{\theta} - \theta}{\text{SE}} + \frac{\theta - \theta_0}{\text{SE}} = Z + \left(\frac{\theta - \theta_0}{\text{SE}}\right).$$

Therefore $T \sim \text{N}(\kappa, 1)$ where $\kappa = (\theta_0 - \theta)/\text{SE}$.

Exercise D - (10 min)

$X_1, \ldots, X_n \sim \text{N}(\mu, \sigma^2)$; estimate μ using $\bar{X}_n$

1. How does $\text{SE}(\bar{X}_n)$ depend on n and σ?
2. Let $H_0\colon \mu = 0$. What is κ this example?
3. Continuing from 2, plot the power of a one-sided test with $\alpha = 0.05$, $n = 100$ and $\sigma^2 = 25$ as a function of μ.
4. Suppose that $\mu = \sigma/5$. Plot the power of a one-sided test with $\alpha = 0.05$ as a function of n.

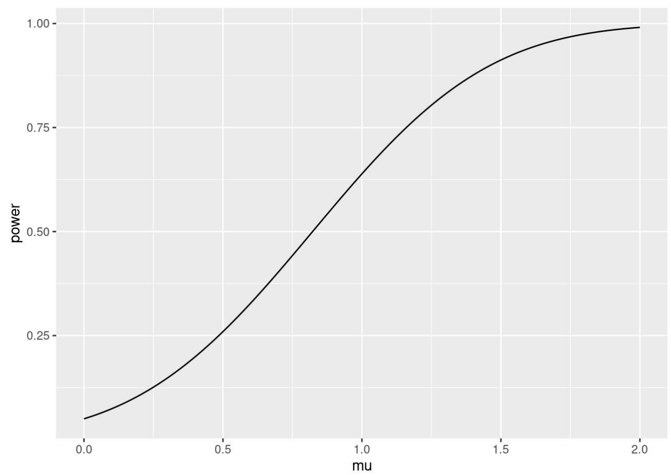
Solution

Parts 1-2

1. $\text{SE}(\bar{X}_n) = \sigma/\sqrt{n}$
2. $\kappa = \sqrt{n}\mu/\sigma$

Part 3

```r
library(tidyverse)
n <- 100
s <- sqrt(25)
alpha <- 0.05
crit <- qnorm(1 - alpha)
tibble(mu = seq(0, 2, 0.01)) |>
  mutate(kappa = sqrt(n) * mu / s,
         power = 1 - pnorm(crit - kappa)) |>
  ggplot(aes(x = mu, y = power)) +
  geom_line()
```



Part 4

```r
mu_over_sigma <- 0.2
tibble(n = 1:500) |>
  mutate(kappa = sqrt(n) * mu_over_sigma,
         power = 1 - pnorm(crit - kappa)) |>
  ggplot(aes(x = n, y = power)) +
  geom_line()
```