1 Hypothesis testing for a single mean

1. The null, or status quo, hypothesis is labeled H_0, the alternative H_a or H_1 or H_2 ...

2. A **type I error** occurs when we falsely reject the null hypothesis. The probability of a type I error is usually labeled α.

3. A **type II error** occurs when we falsely fail to reject the null hypothesis. A type II error is usually labeled β.

4. A **Power** is the probability that we correctly reject the null hypothesis, $1 - \beta$.

5. The Z test for $H_0 : \mu = \mu_0$ versus $H_1 : \mu < \mu_0$ or $H_2 : \mu \neq \mu_0$ or $H_3 : \mu > \mu_0$ constructs a test statistic $TS = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$ and rejects the null hypothesis when

   H_1 $TS \leq -Z_{1-\alpha}$

   H_2 $|TS| \geq Z_{1-\alpha/2}$

   H_3 $TS \geq Z_{1-\alpha}$

   respectively.

6. The Z test requires the assumptions of the CLT and for n to be large enough for it to apply.

7. If n is small, then a Student's T test is performed exactly in the same way, with the normal quantiles replaced by the appropriate Student's T quantiles and $n - 1$ df.

8. Tests define confidence intervals by considering the collection of values of μ_0 for which you fail to reject a two sided test. This yields exactly the T and Z confidence intervals respectively.

9. Conversely, confidence intervals define tests by the rule where one rejects H_0 if μ_0 is *not in* the confidence interval.

10. A **P-value** is the probability of getting evidence as extreme or more extreme than we actually got under the null hypothesis. For H_3 above, the P-value is calculated as $P(Z \geq TS_{obs}|\mu = \mu_0)$ where TS_{obs} is the observed value of our test statistic. To get the P-value for H_2, calculate a one sided P-value and double it.

11. The P-value is equal to the **attained significance level**. That is, the smallest α value for which we would have rejected the null hypothesis. Therefore, rejecting the null hypothesis if a P-value is less than α is the same as performing the rejection region test.

12. The power of a Z test for H_3 is given by the formula (know how this is obtained)

$$P(TS > Z_{1-\alpha}|\mu = \mu_1) = P\left(Z \geq \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + Z_{1-\alpha}\right).$$

Notice that power required a value for μ_1, the value under the null hypothesis. Correspondingly for H_1 we have

$$P\left(Z \leq \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - Z_{1-\alpha}\right).$$

For H_2, the power is approximately the appropriate one sided power using $\alpha/2$.

13. Some facts about power.

   a. Power goes up as α goes up.

   b. Power of a one sided test is greater than the power of the associated two sided test.

   c. Power goes up as μ_1 gets further away from μ_0.

   d. Power goes up as n goes up.

14. The prior formula can be used to calculate the sample size. For example, using the power formula for H_1, setting $Z_{1-\beta} = \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} - Z_{1-\alpha}$ yields

$$n = \frac{(Z_{1-\beta} + Z_{1-\alpha})^2 \sigma^2}{(\mu_0 - \mu_1)^2},$$

which gives the sample size to have power $= 1 - \beta$. This formula applies for H_3 also. For the two sided test, H_2, replace α by $\alpha/2$.

15. Determinants of sample size.

   a. n gets larger as α gets smaller.

   b. n gets larger as the power you want gets larger.

   c. n gets lager the closer μ_1 is to μ_0.

2 Binomial confidence intervals and tests

1. Binomial distributions are used to model proportions. If $X \sim \text{Binomial}(n, p)$ then $\hat{p} = X/n$ is a sample proportion.

2. $\hat{p}$ has the following properties.

   a. It is a sample mean of Bernoulli random variables.

   b. It has expected value p.

   c. It has variance $p(1-p)/n$. Note that the largest value that $p(1-p)$ can take is $1/4$ at $p = 1/2$.

d. $Z = \frac{\hat{p}-p}{\sqrt{p(1-p)/n}}$ follows a standard normal distribution for large n by the CLT. The convergence to normality is fastest when $p = .5$.

3. The **Wald test** for $H_0 : p = p_0$ versus one of $H_1 : p < p_0$, $H_2 : p = p_0$, and $H_3 : p > p_0$ uses the test statistic

$$TS = \frac{\hat{p} - p}{\sqrt{\hat{p}(1 - \hat{p})/n}}$$

which is compared to standard normal quantiles.

4. The **Wald confidence interval** for a binomial proportion is

$$\hat{p} \pm Z_{1-\alpha/2}\sqrt{\hat{p}(1 - \hat{p})/n}.$$

The Wald interval is the interval obtained by inverting the Wald test (and vice versa).

5. The **Score test** for a binomial proportion is

$$TS = \frac{\hat{p} - p}{\sqrt{p_0(1 - p_0)/n}}.$$

The score test has better finite sample performance than the Wald test.

6. The **Score interval** is obtained by inverting the score test (and vice versa)

$$\hat{p}\left(\frac{n}{n+Z_{1-\alpha/2}^2}\right) + \frac{1}{2}\left(\frac{Z_{1-\alpha/2}^2}{n+Z_{1-\alpha/2}^2}\right)$$
$$\pm Z_{1-\alpha/2}\sqrt{\frac{1}{n+Z_{1-\alpha/2}^2}\left[\hat{p}(1-\hat{p})\left(\frac{n}{n+Z_{1-\alpha/2}^2}\right) + \frac{1}{4}\left(\frac{Z_{1-\alpha/2}^2}{n+Z_{1-\alpha/2}^2}\right)\right]}.$$

7. An approximate score interval for $\alpha = .05$ can be obtained by taking $\tilde{p} = \frac{X+2}{n+4}$ and calculating the Wald interval using $\tilde{p}$ instead of $\hat{p}$.

8. An exact binomial test for H_3 can be performed by calculating the exact P-value

$$P(X \geq x_{obs}|p = p_0) = \sum_{k=x_{obs}}^{n}\binom{n}{k}p_0^k(1 - p_0)^{n-k}.$$

where x_{obs} is the observed success count. For H_1 the corresponding exact P-value is

$$P(X \leq x_{obs}|p = p_0) = \sum_{k=0}^{x_{obs}}\binom{n}{k}p_0^k(1 - p_0)^{n-k}.$$

These confidence intervals are **exact**, which means that the actual type one error rate is *no larger than* α. (The actual type one error rate is generally smaller than α.) Therefore these tests are **conservative**. For H_2, calculate the appropriate one sided P-value and double it.

9. Occasionally, someone will try to convince you to obtain an exact Type I error rate using supplemental randomization. Ignore them.

10. Inverting the exact test, choosing those value of p_0 for which we fail to reject H_0, yields an exact confidence interval. This interval has to be calculated numerically. The coverage of the exact binomial interval is no lower than $100(1 - \alpha)\%$.

3   Group comparisons

1. For group comparisons, make sure to differentiate whether or not the observations are paired (or matched) versus independent.

2. For paired comparisons for continuous data, one strategy is to calculate the **differences** and use the methods for testing and performing hypotheses regarding a single mean. The resulting tests and confidence intervals are called **paired Student's** T tests and intervals respectively.

3. For independent groups of iid variables, say X_i and Y_i, *with a constant variance σ^2* across groups

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$$

limits to a standard normal random variable as both n_x and n_y get large. Here

$$S_p^2 = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$$

is the **pooled estimate** of the variance. Obviously, $\bar{X}$, S_x, n_x are the sample mean, sample standard deviation and sample size for the X_i and $\bar{Y}$, S_y and n_y are defined analogously.

4. If the X_i and Y_i happen to be normal, then Z follows the Student's T distribution with $n_x + n_y - 2$ degrees of freedom.

5. The test statistic $TS = \frac{\bar{X} - \bar{Y}}{S_p\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}}$ can be used to test the hypothesis that $H_0 : \mu_x = \mu_y$ versus the alternatives $H_1 : \mu_x < \mu_y$, $H_2 : \mu_x \neq \mu_y$ and $H_3 : \mu_x > \mu_y$. The test statistic should be compared to Student's T quantiles with $n_x + n_y - 2$ df.

6. $\frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2}$ follows what is called the F distribution with $n_x - 1$ **numerator degrees of freedom** and $n_y - 1$ denominator degrees of freedom.

7. To test the hypothesis $H_0 : \sigma_x^2 = \sigma_y^2$ versus th hypotheses $H_1 : \sigma_x^2 < \sigma_y^2$, $H_2 : \sigma_x^2 \neq \sigma_y^2$ and $H_3 : \sigma_x^2 > \sigma_y^2$ compare the statistic $TS = S_1^2/S_2^2$ to the F distribution. We reject H_0 if:

H_1 if $TS < F_{n_x-1, n_y-1, \alpha}$,

H_2 if $TS < F_{n_x-1, n_y-1, \alpha/2}$ or $TS > F_{n_x-1, n_y-1, 1-\alpha/2}$,

4

H_3 if $TS > F_{n_x-1,n_y-1,1-\alpha}$.

8. The F distribution satisfies the property that $F_{n_x-1,n_y-1,\alpha} = F_{n_y-1,n_x-1,1-\alpha}$. So that, it turns out, that our results are consistent whether we put S_x^2 on the top or bottom.

9. Using the fact that

$$1 - \alpha = P\left(F_{n_x-1,n_y-1,\alpha/2} \leq \frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2} \leq F_{n_x-1,n_y-1,1-\alpha/2}\right)$$

we can calculate a confidence interval for $\frac{\sigma_y^2}{\sigma_x^2}$ as $\left[F_{n_x-1,n_y-1,\alpha}\frac{S_x^2}{S_y^2}, F_{n_x-1,n_y-1,1-\alpha/2}\frac{S_x^2}{S_y^2}\right]$. Of course, the confidence interval for $\frac{\sigma_x^2}{\sigma_y^2}$ is $\left[F_{n_y-1,n_x-1,\alpha}\frac{S_y^2}{S_x^2}, F_{n_y-1,n_x-1,1-\alpha/2}\frac{S_y^2}{S_x^2}\right]$.

10. F tests are not robust to the normality assumption.

11. The statistic

$$\frac{\bar{X} - \bar{Y} - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}}$$

follows a standard normal distribution for large n_x and n_y. It follows an approximate Students T distribution if the X_i and Y_i are normally distributed. The degrees of freedom are given below.

12. For testing $H_0 : \mu_x = \mu_y$ in the event where there is evidence to suggest that $\sigma_x \neq \sigma_y$, the test statistic $TS = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_x^2}{n_x}+\frac{S_y^2}{n_y}}}$ follows an approximate Student's T distribution under the null hypothesis when X_i and Y_i are normally distributed. The degrees of freedom are approximated with

$$\frac{(S_x^2/n_x + S_y^2/n_y)^2}{(S_x^2/n_x)^2/(n_x - 1) + (S_y^2/n_y)^2/(n_y - 1)}.$$

13. The power for a Z test of $H_0 : \mu_x = \mu_y$ versus $H_3 : \mu_x > \mu_y$ is given by

$$P\left(Z \geq Z_{1-\alpha} - \frac{\mu_x - \mu_y}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}\right)$$

while for $H_1 : \mu_x < \mu_y$ it is

$$P\left(Z \leq -Z_{1-\alpha} - \frac{\mu_x - \mu_y}{\sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}}\right).$$

14. Sample size calculation assuming $n_x = n_y = n$

$$n = \frac{(Z_{1-\alpha} + Z_{1-\beta})^2(\sigma_x^2 + \sigma_y^2)}{(\mu_x - \mu_y)^2}.$$