Course summary

Stephanie Ranft S2459825

September 25, 2019

Statistics 3
PSBE2-12

1 Refresher

1.1 More on prioris

Prior to commencing a study, a psychologist/statistician needs to be aware that sample data can be misleading in that it is not always representative of the population. You should be familiar with Table 1 but you may not understand what it means graphically. It is important to remember that the probabilities α and β are conditional probabilities, in that they are conditional on which reality is actually true.

		Reality	
		H_0 is true	H_A is true
Decision	Reject H_0	**Type I error** α "false alarm"	True positive **Power** $1-\beta$ Sensitivity
	Accept H_0	True negative $1-\alpha$ Specificity	**Type II error** β "miss rate"

Table 1

In hypothesis testing, you assume that H_0 is the true reality, but you accept that you might be wrong so H_A is a possible reality. Under the null hypothesis, you assume (if you are taking about means) that $\mu = \mu_0$ where μ_0 is some value, say zero ($\mu_0 = 0$). You accept that there is an $100 \times \alpha\%$ possibility that you will have data which is not representative of the population, for whatever reason (usually attributed to human error in the data collection itself). This says that there is a $100 \times (1-\alpha)\%$ possibility that the sample you collected is representative of your population. If H_0 is the true reality it means that (in terms of $\mu = \mu_0$ and $\alpha = 0.05$) the probability that the sample mean is within 1.96 standard deviations (or lies 1.645 standard deviations away, if one-tailed) from the population mean, given that the population mean is equal to μ_0, is 0.95.
In the case that you were wrong, so H_0 is not the true reality and H_A is, you accept that this is probable with $\alpha\%$, and in fact can calculate **when** you would discover this via your sample.

Two-sided H_A: $\mu \neq \mu_0$, with $\alpha = 0.05$

$$\frac{|\bar{x}-\mu_0|}{\sigma/\sqrt{n}} \geq 1.96 \tag{1.1.1}$$

$$|\bar{x}-\mu_0| \geq 1.96 \times \frac{\sigma}{\sqrt{n}} \tag{1.1.2}$$

$$\bar{x} \geq \mu_0 + 1.96 \times \frac{\sigma}{\sqrt{n}} \tag{1.1.3}$$

$$\bar{x} \leq \mu_0 - 1.96 \times \frac{\sigma}{\sqrt{n}} \tag{1.1.4}$$

If we have that the sample mean is more than some value equal to $\mu_0 + 1.96 \times \sigma/\sqrt{n}$, or less than some value equal to $\mu_0 - 1.96 \times \sigma/\sqrt{n}$, then our p-value is less than α (refer to Figure 1a).

$$\mathbb{P}\left(\bar{X} \geq \mu_0 + 1.96 \times \frac{\sigma}{\sqrt{n}}\right) \approx 0.025 \tag{1.1.5}$$

$$\mathbb{P}\left(\bar{X} \leq \mu_0 - 1.96 \times \frac{\sigma}{\sqrt{n}}\right) \approx 0.025 \tag{1.1.6}$$

One-sided H_A: $\mu > \mu_0$ (right-tailed), with $\alpha = 0.05$

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \geq 1.645 \tag{1.1.7}$$

$$\bar{x} - \mu_0 \geq 1.645 \times \frac{\sigma}{\sqrt{n}} \tag{1.1.8}$$

$$\bar{x} \geq \mu_0 + 1.645 \times \frac{\sigma}{\sqrt{n}} \tag{1.1.9}$$

If we have that the sample mean is more than some value equal to $\mu_0 + 1.645 \times \sigma/\sqrt{n}$ then our p-value is less than α (refer to Figure 1b).

$$\mathbb{P}\left(\bar{X} \geq \mu_0 + 1.645 \times \frac{\sigma}{\sqrt{n}}\right) \approx 0.05 \tag{1.1.10}$$

One-sided H_A: $\mu < \mu_0$ (left-tailed), with $\alpha = 0.05$

$$\frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \leq -1.645 \tag{1.1.11}$$

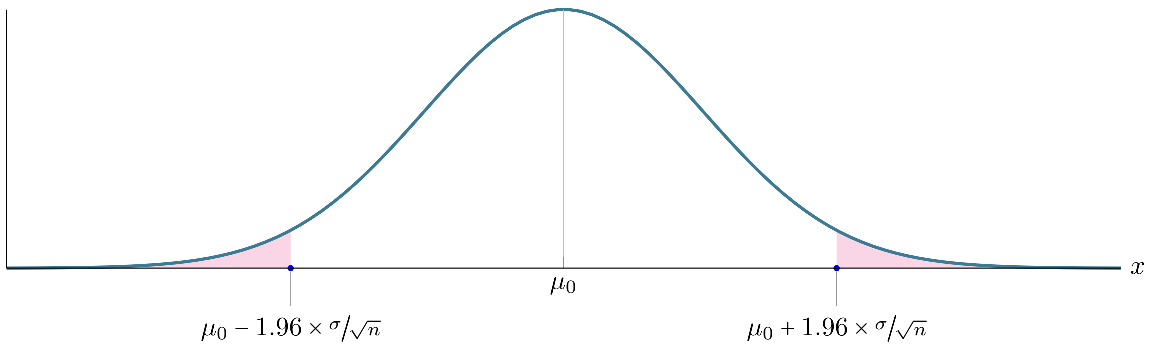
$$\bar{x} - \mu_0 \leq -1.645 \times \frac{\sigma}{\sqrt{n}} \tag{1.1.12}$$

$$\bar{x} \leq \mu_0 - 1.645 \times \frac{\sigma}{\sqrt{n}} \tag{1.1.13}$$

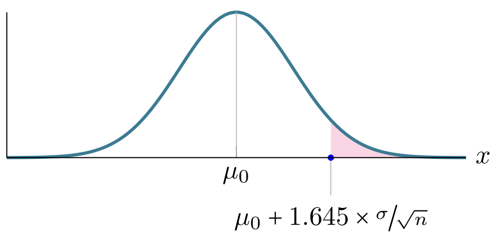
If we have that the sample mean is less than some value equal to $\mu_0 - 1.645 \times \sigma/\sqrt{n}$ then our p-value is less than α (refer to Figure 1c).

$$\mathbb{P}\left(\bar{X} \leq \mu_0 - 1.645 \times \frac{\sigma}{\sqrt{n}}\right) \approx 0.05 \tag{1.1.14}$$

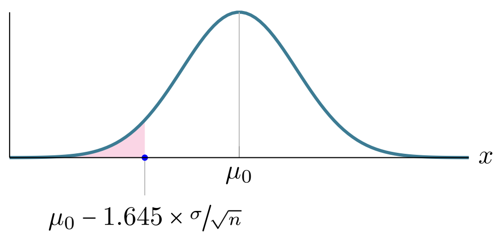
What happens if we modify our acceptance level, α? If we *decrease* α, the shaded areas in Figure 1 will become smaller, which means that our "cut-off values" will change to values which are further from μ_0. A decrease in α will result in an increase of β in samples of unchanged size (Figure 3) - *reduction of power*. If we *increase* our α, the shaded areas in Figure 1 will become larger, which means that our "cut-off values" will change to values which are closer to μ_0. An increase in α will result in an decrease of β in samples of unchanged size (Figure 3) - *increase of power*.



(a) Approximately 95% of sample means lie within 1.96 standard deviations from the mean.



(b) Approximately 95% of sample means lie below 1.645 standard deviations above the mean.



(c) Approximately 95% of sample means lie above 1.645 standard deviations below the mean.

Figure 1: For a given random variable X which has mean μ_0 and standard deviation σ, the sampling distribution of the mean of size n is Normal with mean μ_0 and standard deviation $\sigma/\sqrt{n}$ (the Standard Error) which is denoted $\bar{X} \sim \mathcal{N}\left(\mu_0, \sigma^2/n\right)$.

Now that we have determined the "cut-off values" (when we reject H_0, based on μ_0, n and α), if we were to learn the **true value of μ in the population**, let's call it μ_A, it is possible to compute the power of our test, i.e. the probability of rejection given that H_0 **should** be rejected.

Two-sided H_A: $\mu \neq \mu_0$, with $\alpha = 0.05$

$$\text{power} = \mathbb{P}\left(\text{reject } H_0 \,|H_0 \text{ is not the true reality}\right) \tag{1.1.15}$$

We define two scenarios, where the true mean μ_A is greater than the hypothesised mean μ_0 and when it is less than.

$$= \begin{cases} \mathbb{P}\left(\bar{X} \geq \mu_0 + 1.96 \times \dfrac{\sigma}{\sqrt{n}} \middle| \mu = \mu_A \right), & \mu_A > \mu_0 \\ \mathbb{P}\left(\bar{X} \leq \mu_0 - 1.96 \times \dfrac{\sigma}{\sqrt{n}} \middle| \mu = \mu_A \right), & \mu_A < \mu_0 \end{cases} \tag{1.1.16}$$

Now convert to the standard normal distribution and get your z-scores:

$$= \begin{cases} \mathbb{P}\left(\dfrac{\bar{X} - \mu_A}{\sigma/\sqrt{n}} \geq \dfrac{\mu_0 - \mu_A}{\sigma/\sqrt{n}} + 1.96 \right), & \mu_A > \mu_0 \\ \mathbb{P}\left(\dfrac{\bar{X} - \mu_A}{\sigma/\sqrt{n}} \leq \dfrac{\mu_0 - \mu_A}{\sigma/\sqrt{n}} - 1.96 \right), & \mu_A < \mu_0 \end{cases} \tag{1.1.17}$$

$$= \begin{cases} \mathbb{P}\left(Z \geq \dfrac{\mu_0 - \mu_A}{\sigma/\sqrt{n}} + 1.96 \right), & \mu_A > \mu_0 \\ \mathbb{P}\left(Z \leq \dfrac{\mu_0 - \mu_A}{\sigma/\sqrt{n}} - 1.96 \right), & \mu_A < \mu_0 \end{cases} \tag{1.1.18}$$

In two-sided hypothesis testing, we can use the "cut-off values" to calculate power using the z-test. So, if $\mu_A > \mu_0$, we transform the "cut-off value" $\bar{x} = \mu_0 + 1.96 \times {}^{\sigma}/_{\sqrt{n}}$ into a z-score using μ_A **which should be negative**. The reason it **should be negative** is that will imply that our power is more than 50%; if you achieve a positive z-score with the $\geq$ sign, then you will certainly have low statistical power. We can transform this negative z-score with the $\geq$ sign into a positive z-score with the $\leq$ sign as follows:

$$\mathbb{P}\left(Z \geq \frac{\mu_0 - \mu_A}{\sigma/\sqrt{n}} + 1.96 \right) = \mathbb{P}\left(Z \leq \frac{\mu_A - \mu_0}{\sigma/\sqrt{n}} - 1.96 \right), \qquad \mu_A > \mu_0. \tag{1.1.19}$$

We are able to do this because the standard Normal distribution is symmetric about the mean of zero. It is now possible to read this value from the "regular" z-table, which usually shows $\leq$ and positive z-values. If $\mu_A < \mu_0$, then the transformed "cut-off" value that we calculated is already positive and we already have the $\leq$ sign, so it is easy to read from the table.

One-sided H_A: $\mu > \mu_0$ (right-tailed), with $\alpha = 0.05$

$$\text{power} = \mathbb{P}\left(\text{reject } H_0 \,|H_0 \text{ is not the true reality}\right) \tag{1.1.20}$$

$$= \mathbb{P}\left(\bar{X} \geq \mu_0 + 1.645 \times \frac{\sigma}{\sqrt{n}} \middle| \mu = \mu_A \right) \tag{1.1.21}$$

$$= \mathbb{P}\left(\frac{\bar{X} - \mu_A}{\sigma/\sqrt{n}} \geq \frac{\mu_0 - \mu_A}{\sigma/\sqrt{n}} + 1.645 \right) \tag{1.1.22}$$

$$= \mathbb{P}\left(Z \geq \frac{\mu_0 - \mu_A}{\sigma/\sqrt{n}} + 1.645 \right) \tag{1.1.23}$$

Again, we can transform this negative z-score (**should be negative** because $\mu_0 < \mu_A$ and **power should be more than 50%**) with $\geq$ sign to a positive z-score with $\leq$ sign:

$$= \mathbb{P}\left(Z \leq \frac{\mu_A - \mu_0}{\sigma/\sqrt{n}} - 1.645 \right) \tag{1.1.24}$$

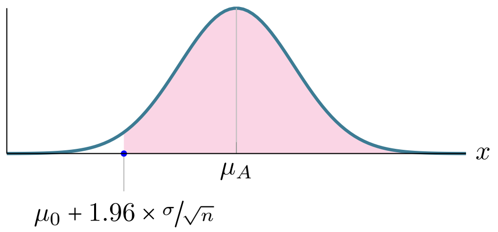
One-sided H_A: $\mu < \mu_0$ (left-tailed), with $\alpha = 0.05$

$$\text{power} = \mathbb{P}\left(\text{reject } H_0 \,|H_0 \text{ is not the true reality}\right) \tag{1.1.25}$$

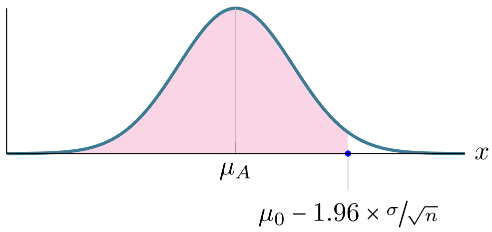
$$= \mathbb{P}\left(\bar{X} \leq \mu_0 - 1.645 \times \frac{\sigma}{\sqrt{n}} \middle| \mu = \mu_A \right) \tag{1.1.26}$$

$$= \mathbb{P}\left(\frac{\bar{X} - \mu_A}{\sigma/\sqrt{n}} \leq \frac{\mu_0 - \mu_A}{\sigma/\sqrt{n}} - 1.645 \right) \tag{1.1.27}$$

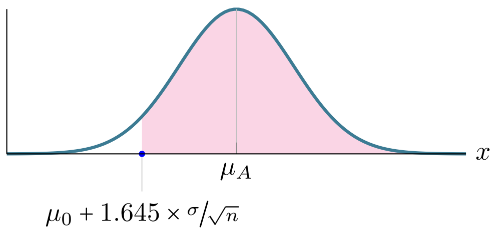
$$= \mathbb{P}\left(Z \leq \frac{\mu_0 - \mu_A}{\sigma/\sqrt{n}} - 1.645 \right) \tag{1.1.28}$$



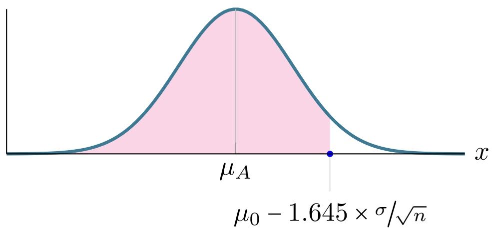
(a) If the true mean μ_A is greater than μ_0, we calculate power as the probability of observing a sample mean greater than our "cut-off value" $\mu_0 + 1.96 \times \sigma/\sqrt{n}$, where $\alpha = 0.05$.



(b) If the true mean μ_A is less than μ_0, we calculate power as the probability of observing a sample mean less than our "cut-off value" $\mu_0 - 1.96 \times \sigma/\sqrt{n}$, where $\alpha = 0.05$.



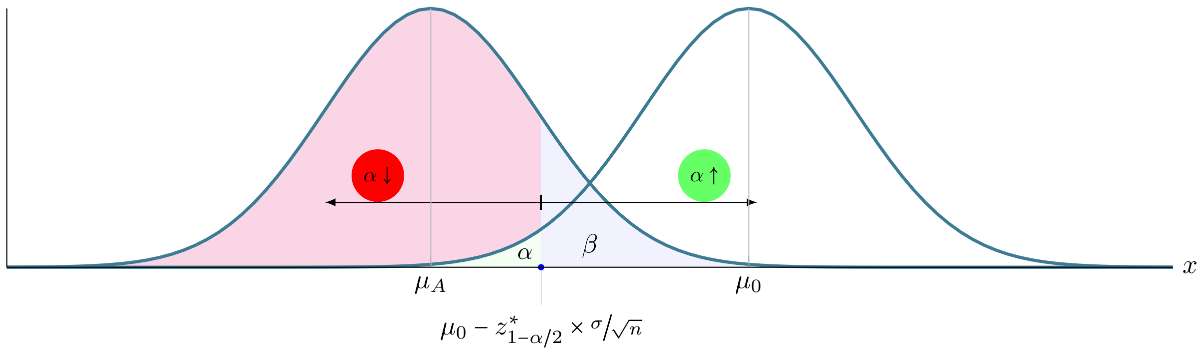
(c) For a right-tailed hypothesis test, we calculate statistical power as the probability of observing a sample mean greater than our "cut-off value" $\mu_0 + 1.645 \times \sigma/\sqrt{n}$, where $\alpha = 0.05$.



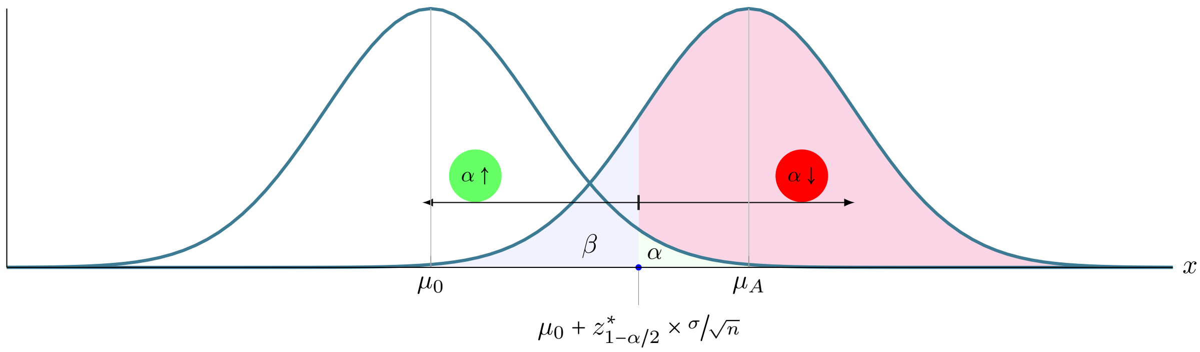
(d) For a left-tailed hypothesis test, we calculate statistical power as the probability of observing a sample mean less than our "cut-off value" $\mu_0 - 1.645 \times \sigma/\sqrt{n}$, where $\alpha = 0.05$.

Figure 2: When the true mean of the population is $\mu = \mu_A$, unequal to μ_0, we can calculate the statistical power based on the "cut-off values" proposed by μ_0 given in Figure 1.

So, if we are to finalise all of these ideas together (visually):



(a) Left-tailed power calculation ($\mu_A < \mu_0$). The red shaded area is the probability of observing a sample mean less than $\mu_0 - z^*_{1-\alpha/2} \times \sigma/\sqrt{n}$ conditional on $\mu = \mu_A$. The blue shaded area is β and the green shaded area is α. The use of the critical value $z^*_{1-\alpha/2}$ v.s. $z^*_{1-\alpha}$ depends on whether the null hypothesis is two- or one-sided.



(b) Right-tailed power calculation ($\mu_A > \mu_0$). The red shaded area is the probability of observing a sample mean more than $\mu_0 + z^*_{1-\alpha/2} \times \sigma/\sqrt{n}$ conditional on $\mu = \mu_A$. The blue shaded area is β and the green shaded area is α. The use of the critical value $z^*_{1-\alpha/2}$ v.s. $z^*_{1-\alpha}$ depends on whether the null hypothesis is two- or one-sided.

Figure 3: Power calculation for $\mu_A < \mu_0$ and $\mu_A > \mu_0$ when the population variance is known (z-test). If α is decreased, then the critical value $z^*_{1-\alpha/2}$ (or $z^*_{1-\alpha}$, if H_0 is one-sided) becomes larger which is visually shown by the red shaded area becoming smaller (smaller power). If α is increased, then the critical value $z^*_{1-\alpha/2}$ (or $z^*_{1-\alpha}$, if H_0 is one-sided) becomes smaller which is visually shown by the red shaded area becoming larger (larger power).

2 Homework

2.1 Homework 3

1. In Exercise 3 of Section 2.1 we learned that the time it takes for adult men to fall asleep is distributed as a normal with a mean of $\mu = 335$ seconds and a standard deviation of $\sigma = 15$ seconds. Twenty-five adult men undergo hypnotherapy to decrease the time required to fall asleep. We wish to test the following hypotheses:

$$\begin{aligned} H_0 &: \mu = 335 \\ H_A &: \mu < 335 \end{aligned}$$

(a) The mean observed time for the 25 men to fall asleep is 330.1 seconds. Will H_0 be rejected at a significance level of $\alpha = .05$?

Solution We reject H_0 when our sample statistic is more extreme than 95% of the population (note that the defined hypotheses imply a one-sided test):

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z^*_{1-\alpha} \tag{2.1.1}$$

$$\implies \frac{330.1 - 335}{15/\sqrt{25}} < -1.645 \tag{2.1.2}$$

$$\implies -1.6\bar{3} < -1.645. \tag{2.1.3}$$

As the above inequality does not hold (obviously $-1.633 > -1.645$), we can conclude that our statistic is not extreme enough to warrant us rejecting the null hypothesis. If we wish to find the p-value:

$$p = \mathbb{P}\left(\bar{X} < 330.1 \,|\mu_0 = 335\right) \tag{2.1.4}$$

$$= \mathbb{P}\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < \frac{330.1 - 335}{15/\sqrt{25}}\right) \tag{2.1.5}$$

$$= \mathbb{P}\left(Z < -1.6\bar{3}\right) \tag{2.1.6}$$

$$= 0.0516. \tag{2.1.7}$$

Again, we do not reject H_0 as we have that $p > \alpha$.

(b) The mean observed time for the 25 men to fall asleep is 330.0 seconds. Will H_0 be rejected at a significance level of $\alpha = .05$?

Solution In the same manner as before, we compare our statistic to the critical z-value:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -z^*_{1-\alpha} \tag{2.1.8}$$

$$\implies \frac{330.0 - 335}{15/\sqrt{25}} < -1.645 \tag{2.1.9}$$

$$\implies -1.\bar{6} < -1.645. \tag{2.1.10}$$

As the above inequality does hold (obviously $-1.667 < -1.645$), we can conclude that our statistic is extreme enough to warrant us rejecting the null hypothesis. If we wish to find the p-value:

$$p = \mathbb{P}\left(\bar{X} < 330.0 \,|\mu_0 = 335\right) \tag{2.1.11}$$

$$= \mathbb{P}\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < \frac{330.0 - 335}{15/\sqrt{25}}\right) \tag{2.1.12}$$

$$= \mathbb{P}\left(Z < -1.\bar{6}\right) \tag{2.1.13}$$

$$= 0.0475. \tag{2.1.14}$$

Again, we reject H_0 as we have that $p < \alpha$, but realise that we only reject at a 5% significance level and not at 1%.

(c) State to what extent you believe it is useful to think about results in terms of rigid, accept/reject decisions based on a particular significance level.

Solution How powerful is our test? Firstly, for what value $\bar{x}$ do we certainly reject H_0 (it's somewhere between 330.0 and 330.1 seconds):

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < -1.645 = -z^*_{1-\alpha} \tag{2.1.15}$$

$$\implies \bar{x} < \mu_0 - 1.645 \times \frac{\sigma}{\sqrt{n}} \tag{2.1.16}$$

$$= 335 - 1.645 \times 3 \tag{2.1.17}$$

$$= 330.065 \tag{2.1.18}$$

$$\implies \text{power} = \text{probability of correctly rejecting } H_0 \tag{2.1.19}$$

Recall that we reject H_0 when $\bar{x} < 330.065$ and we are correct in doing so when the alternative hypothesis is true.

$$= \mathbb{P}\left(\bar{X} < 330.065 \,|\mu = \mu_A\right), \tag{2.1.20}$$

where μ_A is the true population mean. Suppose $\mu_A = 330.0$

$$\implies \text{power} = \mathbb{P}\left(\frac{\bar{X} - \mu_A}{\sigma/\sqrt{n}} < \frac{330.065 - 330.0}{15/5}\right) \tag{2.1.21}$$

$$= \mathbb{P}\left(Z < 0.021\bar{6}\right) \tag{2.1.22}$$

$$= 0.5080 \tag{2.1.23}$$

So, the probably of a Type II error is around 50%, which is relatively high so it does not seem useful to base our decision on a significance test in this case[1]. In order to increase the power of our test, we need to increase the sample size and in doing so we aim for power of at least 80%:

$$0.80 \approx \mathbb{P}\left(Z < 0.85\right) \tag{2.1.24}$$

You can find the above z critical value from Table A.

$$= \mathbb{P}\left(\frac{\bar{X} - \mu_A}{\sigma/\sqrt{n}} < \frac{330.065 - 330.0}{15/\sqrt{n}}\right). \tag{2.1.25}$$

So, we are looking for a n which satisfies

$$\implies 0.85 = \frac{330.065 - 330.0}{15/\sqrt{n}} \tag{2.1.26}$$

$$\implies n = \left(\frac{0.85 \times 15}{330.065 - 330.0}\right)^2 \approx 38476.33. \tag{2.1.27}$$

This tells us that, if the researchers want to be 80% certain that they are correctly rejecting the null hypothesis, they would require at least 38,477 participants for this study (assuming that the true mean is 330.0 seconds). For a hypothetical μ_A, we require an n that satisfies:

$$n = \left(\frac{0.85 \times 15}{330.065 - \mu_A}\right)^2. \tag{2.1.28}$$

The size of the required n is extremely large if the distance between μ_A and 330.065 is less then 1, i.e. $329.065 < \mu_A < 331.065$, but is decreasing if the distance is more than 1. This also gives you a bit of an idea about power: obviously μ_A is **fixed in the population** but a researcher might have a hypothesised mean μ_0 is stupidly too far away from μ_A in order to overflate the power of their test.
We can also look at this from the perspective of confidence intervals, which we learnt in the previous homework is more informative than a p-value.

$$90\% \text{ CI for } \mu: \qquad \bar{x} \pm z^*_{1-\alpha/2} \times \frac{\sigma}{\sqrt{n}} = 330.1 \pm 1.645 \times \frac{15}{5} \tag{2.1.29}$$

$$= \left(325.165, 335.035\right). \tag{2.1.30}$$

If we have a sample mean of 330.1 seconds, then the 90% CI (90%, because the hypothesis test is one-tailed) for μ contains $\mu_0 = 335$.

$$90\% \text{ CI for } \mu: \qquad \bar{x} \pm z^*_{1-\alpha/2} \times \frac{\sigma}{\sqrt{n}} = 330.0 \pm 1.645 \times \frac{15}{5} \tag{2.1.31}$$

$$= (324.065, 334.935)\,. \tag{2.1.32}$$

Here we see that $\mu_0 = 335$ is not in the CI (only just), but if we were to round up to the nearest second then it would be included. If we were to increase our n, then $\mu_0 = 335$ might not be in the 90% CI for either part a or b. This question is just trying to show you the silly-side of significance testing.

2. In Section 2.2.1, we discussed a neuropsychologist studying about information processing in parents with closed head injury (CHI). The neuropsychologist used a number of reaction time tasks that are assumed to require different forms of information processing. His hypothesis is that CHI patients process information more slowly than healthy people.
The test scores for healthy adults in the mirror-reading task are normally distributed with mean of 3000ms and standard deviation 400ms. The neurologist administers the task to 16 randomly selected CHI patients, at least 1 year after their accident. The researcher wonders how easy it would be to determine that the CHI patients were slower if they were much slower than the healthy participants. He chooses a 5% significance level, and "much slower" to him is 3300ms.

[Extract important information:]
$X \sim \mathcal{N}\left(3000, 400^2\right)$: test scores for healthy adults in the mirror-reading task.
$n = 16$ of CHI patients.

$$\begin{aligned} H_0 &: \mu = 3000 \\ H_A &: \mu > 3000 \end{aligned}$$

$\alpha = 0.05$ significance level.
$\bar{X} \sim \mathcal{N}\left(3000, 400^2/16\right)$: sampling distribution of the mean of X.

(a) Find the power of the test under the alternative hypothesis that μ for the CHI patients is 3300ms.

Solution The neuropsychologist considers rejecting H_0 if the sample statistic is more extreme than 3164.5ms:

$$\bar{x} > \mu_0 + z^*_{1-\alpha} \times \frac{\sigma}{\sqrt{n}} \tag{2.1.33}$$

$$= 3000 + 1.645 \times \frac{400}{\sqrt{16}} \tag{2.1.34}$$

$$= 3164.5. \tag{2.1.35}$$

Given that we are told the true population mean is 3300ms, how powerful is our test?

$$\text{power} = \mathbb{P}\left(\frac{\bar{X} - \mu_A}{\sigma/\sqrt{n}} > \frac{3164.5 - 3300}{400/\sqrt{16}}\right) \tag{2.1.36}$$

$$= \mathbb{P}\left(Z > -1.355\right) \tag{2.1.37}$$

$$= 0.9123. \tag{2.1.38}$$

This tells us that, given a sample size of 16 we will correctly reject the null hypothesis 91.23% of the time.

(b) Does the experiment have a good chance of finding a significant effect if the mean score for the CHI patients is truly 3300ms?

Solution You will find a significant effect is the p-value is less that α. This question is asking you for the probably of rejecting H_0 given that H_0 is false, i.e. what is the power. This question is hoping to trick you, but it's the same as part a and only hopes that you explore the theory a little deeper.
In general, researchers like to ensure that the power of their test is around 80% and in our case we have 91%.

[1] This paper might be of interest `https://www.jstor.org/stable/449153`, or this one `https://academic.oup.com/ptj/article/79/2/186/2837119`.

(c) Explain why the experiment will give much clearer results if the researcher were to take a larger sample.

Solution We reject the null hypothesis at 5% significance level when

$$\bar{x} > \mu_0 + z^*_{1-\alpha} \times \frac{\sigma}{\sqrt{n}} \tag{2.1.39}$$

$$= 3000 + \frac{658}{\sqrt{n}}. \tag{2.1.40}$$

Assuming that the true population mean is 3300ms, the power of our test is

$$= \mathbb{P}\left(\frac{\bar{X} - \mu_A}{\sigma/\sqrt{n}} > \frac{\left(3000 + \frac{658}{\sqrt{n}}\right) - 3300}{400/\sqrt{n}} \right) \tag{2.1.41}$$

$$\text{power} = \mathbb{P}\left(Z > 1.645 - 0.75\sqrt{n}\right) \tag{2.1.42}$$

$$= \begin{cases} \mathbb{P}\left(Z > -1.355\right) = 0.9123, & n = 16; \\ \mathbb{P}\left(Z > -2.105\right) = 0.9821, & n = 25; \\ \mathbb{P}\left(Z > -2.855\right) = 0.99785, & n = 36; \\ \mathbb{P}\left(Z > -3.605\right) \approx 1, & n = 49. \end{cases} \tag{2.1.43}$$

If n increases, then the standard error of the estimate of μ decreases: $\bar{X}$ estimates μ such that $\bar{X} \sim \mathcal{N}\left(\mu, \sigma^2/n\right)$, where $\sigma/\sqrt{n}$ is the standard error of the estimate. In our case, $\bar{X} \sim \mathcal{N}\left(3000, 400^2/n\right)$ and in Figure 4 you can observe how increasing n causes the sampling distribution of the mean to become more centralised around μ, whether it be 3000ms or 3300ms.

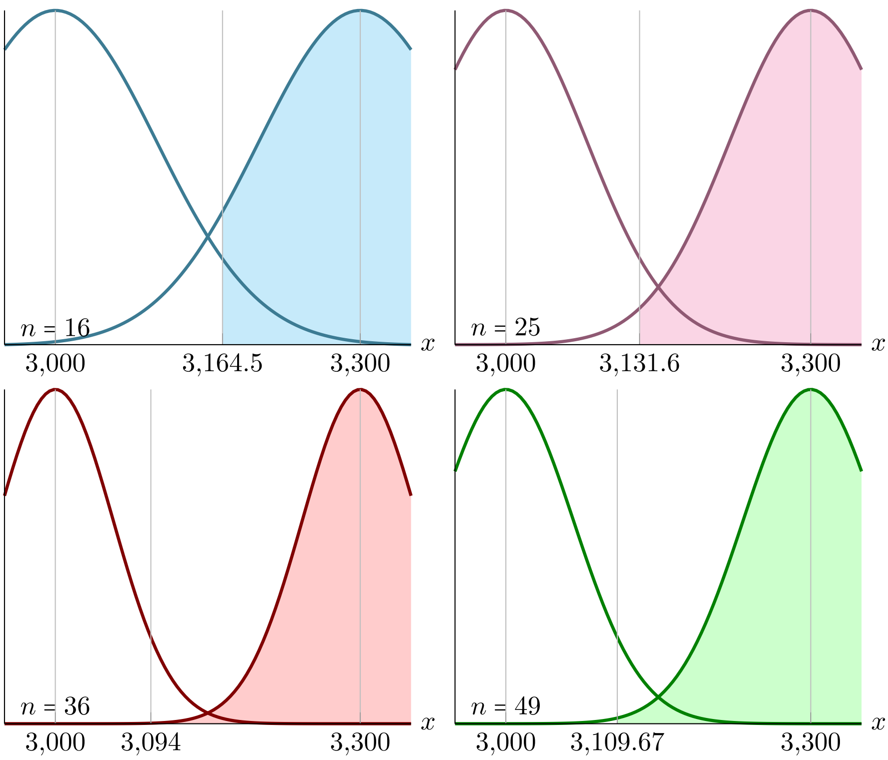


Figure 4: The shaded area represents the power of the test for increasing values of n, where the random variable $\bar{X}$ is distributed normally with standard error $400/\sqrt{n}$ ($\alpha = 0.05$; $H_0 : \mu = 3000$ms; $H_a : \mu > 3000$ms; true mean $\mu = 3300$ms). As n increases, the "cut-off value" decreases resulting in an increase of statistical power (caeteris paribus).

(d) Suppose that the alternative hypothesis is farther away from H_0; say, for instance, μ is truly 3500 for the CHI patients. Will the power than be higher or lower than the value you found in Question 2a?

Solution In the same way that we calculated power in (2.1.41):

$$\text{power} = \mathbb{P}\left(\frac{\bar{X} - \mu_A}{\sigma/\sqrt{n}} > \frac{\left(3000 + \frac{658}{\sqrt{n}}\right) - 3500}{400/\sqrt{n}}\right) \tag{2.1.44}$$

$$= \mathbb{P}\left(Z > 1.645 - 1.25\sqrt{n}\right) \tag{2.1.45}$$

$$= \begin{cases} \mathbb{P}(Z > -3.355) = 0.9996, & n = 16; \\ \mathbb{P}(Z > -3.5) \approx 1, & n > 16. \end{cases} \tag{2.1.46}$$

Fixing n and α, the power of the test increases dramatically if the true mean is much farther away from the hypothesised mean than if they are close.

(e) Another way to determine whether the sample size is large enough to yield worthwhile results is to determine whether the 90% confidence interval will be narrow enough. (90%, because the hypothesis test is one-tailed). The researcher would like the confidence interval to be more narrow than 60ms. Is his sample size large enough?

Solution If the researcher would like the confidence interval to be more narrow than 60ms, then they would like to ensure that the margin of error is no more than 30ms:

$$\text{margin of error} = z^*_{1-\alpha/2} \times \frac{\sigma}{\sqrt{n}} < 30 \tag{2.1.47}$$

$$\implies 1.645 \times \frac{400}{\sqrt{n}} < 30 \tag{2.1.48}$$

$$\implies n > \left(\frac{1.645 \times 400}{30}\right)^2 \approx 481.071. \tag{2.1.49}$$

The sample size of 16 is too small to ensure that the confidence interval be narrower than 60ms; a sample size of at least 482 is needed.

(f) Explain the relationship between the width of the confidence interval and how worthwhile the experiment is.

Solution If the researcher wants to ensure that their experiment is worthwhile then they need to ensure that the probabilities of Type I or II errors are maintained at an acceptable level, which are typically $\alpha = 0.05$ and $\beta = 0.20$. The value of α will dictate which critical value is used for the width of the interval and β will dictate the appropriate sample size.

$$\text{width of the CI} = 2 \times z^*_{1-\alpha/2} \times \frac{\sigma}{\sqrt{n}}. \tag{2.1.50}$$

$$1 - \beta = \mathbb{P}\left(Z < \left|z^*_{1-\alpha} - \frac{|\mu_0 - \mu_A|}{\sigma/\sqrt{n}}\right|\right). \tag{2.1.51}$$

The researcher prescribes $\alpha = 0.05$ and $\beta = 0.20$.

$$\implies \text{width of the CI} = 2 \times 1.96 \times \frac{\sigma}{\sqrt{n}}. \tag{2.1.52}$$

The researcher can increase or decrease their sample size to ensure an appropriate width, or alternatively they can ensure that their requirements for power are met:

$$\implies 0.80 = \mathbb{P}\left(Z < \left|1.645 - \frac{|\mu_0 - \mu_A|}{\sigma/\sqrt{n}}\right|\right) \tag{2.1.53}$$

Checking table A we note that $\mathbb{P}(Z < 2.05) \approx 0.8$.

$$\implies \left|1.645 - \frac{|\mu_0 - \mu_A|}{\sigma/\sqrt{n}}\right| = 2.05 \tag{2.1.54}$$

$$\implies n = \left(\frac{|\mu_0 - \mu_A|}{\sigma \times (2.05 - 1.645)}\right)^2 \tag{2.1.55}$$

To summarise, a researcher ensures that their experiment is worthwhile (reducing probability of Type I and II errors) by setting the sample size to restrict the width of the confidence interval.

3. The scores of men on the Chapin Test for Social Insight (CTSI; [Chapin, 1942]) are normally distributed with mean 25 and standard deviation 5. Assume a sample size of $N = 16$. A study is run to determine whether women have a lower average score on the CTSI than men. We let μ be the mean of the population of women, and we wish to test the following hypotheses:

$$\begin{aligned} H_0 &: \mu = 25 \\ H_a &: \mu < 25 \end{aligned}$$

We will assume that the standard deviation of the population of women is also 5. For $\alpha = 0.05$, H_0 can be rejected when $\bar{x} < 22.95$.

(a) Compute the probability of a Type I error. In other words, compute the probability that one will reject H_0 if the mean of the population of women is exactly $\mu = 25$.

Solution Under H_0, the probability of a Type I error is equal to α.

$$\begin{aligned} \alpha &= \mathbb{P}\left(\bar{X} < 22.95 \,|\mu = 25\right) && (2.1.56) \\ &= \mathbb{P}\left(\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} < \frac{22.95 - 25}{5/\sqrt{16}}\right) && (2.1.57) \\ &= \mathbb{P}\left(Z < -1.64\right) && (2.1.58) \\ &= 0.0516 && (2.1.59) \end{aligned}$$

(b) Compute the probability of a Type II error when $\mu = 22$. In other words, compute the probability that one will not reject H_0 if the mean of the population of women truly is $\mu = 22$.

Solution We don't reject H_0 if $\bar{x} \geq 22.95$, so

$$\begin{aligned} \beta &= \mathbb{P}\left(\bar{X} > 22.95 \,|\mu = 22\right) && (2.1.60) \\ &= \mathbb{P}\left(\frac{\bar{X} - \mu_A}{\sigma/\sqrt{n}} > \frac{22.95 - 22}{5/\sqrt{16}}\right) && (2.1.61) \\ &= \mathbb{P}\left(Z > 0.76\right) && (2.1.62) \\ &= 0.2236. && (2.1.63) \\ \implies \text{power} &= 1 - \beta = 0.7764. && (2.1.64) \end{aligned}$$

(c) Compute the probability of a Type II error if $\mu = 20$.

Solution

$$\begin{aligned} \beta &= \mathbb{P}\left(\bar{X} > 22.95 \,|\mu = 20\right) && (2.1.65) \\ &= \mathbb{P}\left(\frac{\bar{X} - \mu_A}{\sigma/\sqrt{n}} > \frac{22.95 - 20}{5/\sqrt{16}}\right) && (2.1.66) \\ &= \mathbb{P}\left(Z > 2.36\right) && (2.1.67) \\ &= 0.0091. && (2.1.68) \\ \implies \text{power} &= 1 - \beta = 0.9909 && (2.1.69) \end{aligned}$$

(d) Using your computation of the Type II error rate of the test, find the power if $\mu = 22$.

Solution Power $= 1 - \beta = 0.7764$.

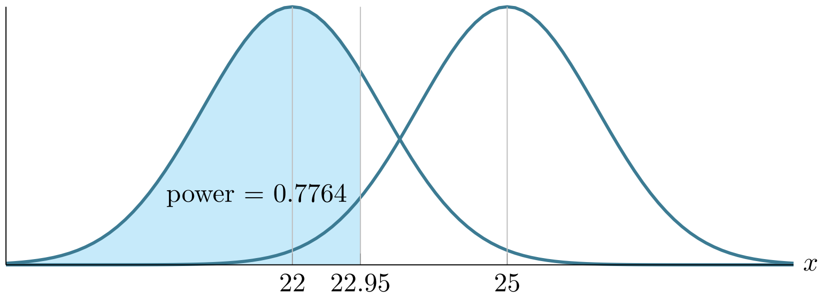


Figure 5

(e) Using your computation of the Type II error rate of the test, find the power if $\mu = 20$.

Solution Power $= 1 - \beta = 0.9909$.

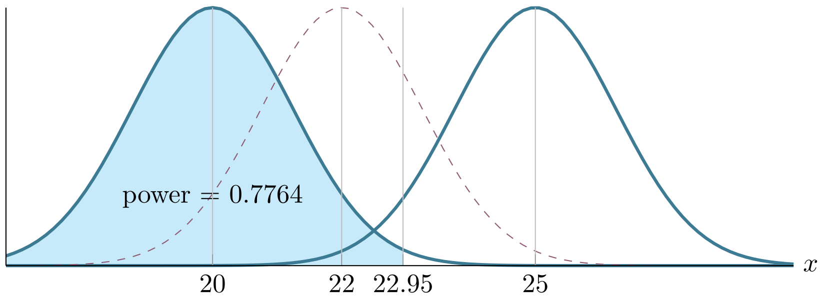


Figure 6

4. If σ is unknown and we must estimate it, we generally use the t statistic. Suppose the t statistic for the test

$$\begin{aligned} H_0 &: \mu = 25 \\ H_a &: \mu < 25 \end{aligned}$$

based on $N = 10$ observations, has a value of $t = -4.45$.

(a) What are the appropriate degrees of freedom for this t test?

Solution Degrees of freedom is just allocation in a loose sense: in this case, we have 10 observations and once we have allocated 9 of these observations with a place number, there is no freedom for the 10th allocation. Hence, we have 9 degrees of freedom.

(b) Using the t-distribution Critical Values Table, find the p value resulting from the t test.

Solution Check table B at the back of the Agresti book, and you will see that $\mathbb{P}\left(|t|_{\text{df=9}} < 4.297\right) = 0.9980$, so this means that our p-value must be less than $(1 - 0.9980)/2 = 0.001$ as our statistic is more extreme than the critical value. This can be seen in Figure 7.

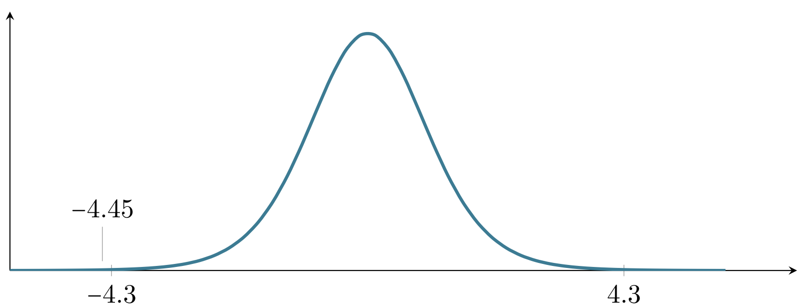


Figure 7: For the t-distribution with 9 degrees of freedom, a test statistic of −4.45 is significant at every level as the (two-sided) critical value for $\alpha = 0.002$ is −4.297.