STAT 513 fa 2019 Lec 05

Some large-sample tests and sample size calculations

Karl B. Gregory

Large-sample tests for means and proportions

- We consider some tests concerning means and proportions which can be used when sample sizes are large. Each relies upon the Central Limit Theorem.

- **Central Limit Theorem:** Let $X_1, \dots, X_n$ be a random sample from a distribution with mean μ and variance $\sigma^2 < \infty$. Then

$$\sqrt{n}(\bar{X}_n - \mu)/\sigma \to \text{Normal}(0,1) \text{ in distribution}$$

  as $n \to \infty$.

- **Corollary via Slutzky's Theorem**. Under the same settings, if S_n is a consistent estimator of σ, then

$$\sqrt{n}(\bar{X}_n - \mu)/S_n \to \text{Normal}(0,1) \text{ in distribution}$$

  as $n \to \infty$.

- Theses results allow us to easily construct tests of hypotheses about the mean which, when n is large, do not require that the population distribution be Normal.

- **Formulas for large-sample tests about the mean:** Let $X_1, \dots, X_n$ be a random sample from a distribution with mean μ and variance $\sigma^2 < \infty$, and for some μ_0, let $T_n = \sqrt{n}(\bar{X}_n - \mu_0)/S_n$. We tabulate below some tests of hypotheses about the mean along with their large-n power functions and p-value formulas:

H_0	H_1	Reject H_0 at α iff	Approx. power function $\gamma(\mu)$	Approx. p-value
$\mu \leq \mu_0$	$\mu > \mu_0$	$T_n > z_\alpha$	$1 - \Phi(z_\alpha - \sqrt{n}(\mu - \mu_0)/\sigma)$	$1 - \Phi(T_n)$
$\mu \geq \mu_0$	$\mu < \mu_0$	$T_n < -z_\alpha$	$\Phi(-z_\alpha - \sqrt{n}(\mu - \mu_0)/\sigma)$	$\Phi(T_n)$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\lvert T_n \rvert > z_{\alpha/2}$	$1 - [\Phi(z_{\alpha/2} - \sqrt{n}(\mu - \mu_0)/\sigma)$ $-\Phi(-z_{\alpha/2} - \sqrt{n}(\mu - \mu_0)/\sigma)]$	$2[1 - \Phi(\lvert T_n \rvert)]$

- We now consider a special case of sampling from a non-Normal population: when $X_1, \dots, X_n$ is a random sample from the $\text{Bernoulli}(p)$ distribution.

- **Formulas for large-sample tests about the proportion:** Let $X_1, \dots, X_n$ be a random sample from the $\text{Bernoulli}(p)$ distribution and for some $p_0 \in (0,1)$ let $Z_n = \sqrt{n}(\hat{p}_n - p_0)/\sqrt{p_0(1-p_0)}$, where $\hat{p}_n = \bar{X}_n$. We tabulate below some tests of hypotheses about p along with their large-n power functions and p-value formulas, using the notation $\sigma_0^2 = p_0(1-p_0)$ and $\sigma^2 = p(1-p)$.

H_0	H_1	Reject H_0 at α iff	Approx. power function $\gamma(p)$	Approx. p-value
$p \le p_0$	$p > p_0$	$Z_n > z_\alpha$	$1 - \Phi((\sigma_0/\sigma)(z_\alpha - \sqrt{n}(p-p_0)/\sigma_0))$	$1 - \Phi(Z_n)$
$p \ge p_0$	$p < p_0$	$Z_n < -z_\alpha$	$\Phi((\sigma_0/\sigma)(-z_\alpha - \sqrt{n}(p-p_0)/\sigma_0))$	$\Phi(Z_n)$
$p = p_0$	$p \ne p_0$	$\lvert Z_n \rvert > z_{\alpha/2}$	$1 - [\Phi((\sigma_0/\sigma)(z_{\alpha/2} - \sqrt{n}(p-p_0)/\sigma))$ $-\Phi((\sigma_0/\sigma)(-z_{\alpha/2} - \sqrt{n}(p-p_0)/\sigma_0))]$	$2[1 - \Phi(\lvert Z_n \rvert)]$

  A rule of thumb is to use these tests only when $\min\{np_0, n(1-p_0)\} \ge 15$. Then the sample size is "large enough" for Z_n to be approximately Normal.

- **Exercise:** *Verify the asymptotic power function in the above table for the test of H_0: $p \le p_0$ versus H_1: $p > p_0$.*

  Answer: If $X_1, \dots, X_n$ is a random sample from the $\text{Bernoulli}(p)$ distribution, then the Central Limit Theorem gives

$$\sqrt{n}(\hat{p} - p)/\sqrt{p(1-p)} \to \text{Normal}(0,1) \text{ in distribution}$$

  as $n \to \infty$. Therefore, if n is large, the power is given approximately by

$$\begin{aligned}
\gamma(p) &= P_p(\sqrt{n}(\hat{p} - p_0)/\sqrt{p_0(1-p_0)} > z_\alpha) \\
&= P_p(\sqrt{n}(\hat{p} - p_0)/\sqrt{p(1-p)} > \sqrt{p_0(1-p_0)}/\sqrt{p(1-p)}z_\alpha) \\
&= P_p(\sqrt{n}(\hat{p} - p)/\sqrt{p(1-p)} + \sqrt{n}(p - p_0)/\sqrt{p(1-p)}) > \sqrt{p_0(1-p_0)}/\sqrt{p(1-p)}z_\alpha) \\
&= P_p(\sqrt{n}(\hat{p} - p)/\sqrt{p(1-p)} > \sqrt{p_0(1-p_0)}/\sqrt{p(1-p)}z_\alpha - \sqrt{n}(p - p_0)/\sqrt{p(1-p)}) \\
&= P(Z > \sqrt{p_0(1-p_0)}/\sqrt{p(1-p)}z_\alpha - \sqrt{n}(p - p_0)/\sqrt{p(1-p)}), \quad Z \sim \text{Normal}(0,1) \\
&= 1 - \Phi(\sqrt{p_0(1-p_0)}/\sqrt{p(1-p)}z_\alpha - \sqrt{n}(p - p_0)/\sqrt{p(1-p)}) \\
&= 1 - \Phi((\sigma_0/\sigma)(z_\alpha - \sqrt{n}(p-p_0)/\sigma_0)),
\end{aligned}$$

  letting $\sigma = \sqrt{p(1-p)}$ and $\sigma_0 = \sqrt{p_0(1-p_0)}$.

- We can also use the Central Limit Theorem to construct tests for comparing the means of two non-Normal populations.

- **Formulas for large-sample tests comparing two means:** Let

$$X_{11},\dots,X_{1n_1} \text{ be a random sample from a distribution with mean } \mu_1 \text{ and variance } \sigma_1^2<\infty$$
$$\text{and } X_{21},\dots,X_{2n_1} \text{ be a random sample from a distribution with mean } \mu_2 \text{ and variance } \sigma_2^2<\infty,$$

and for some δ_0, define

$$T = \frac{\bar{X}_1 - \bar{X}_2 - \delta_0}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}.$$

The following table give some tests concerning $\mu_1 - \mu_2$ and gives power functions and p-value formulas which are approximately correct if n_1 and n_2 are large (say $\min\{n_1,n_2\} \geq 30$):

H_1	Rej. H_0 at α iff	Approx. power function $\gamma(\delta)$, $\delta = \mu_1 - \mu_2$	Approx. p-value
$\mu_1 - \mu_2 > \delta_0$	$T > z_\alpha$	$1 - \Phi(z_\alpha - (\delta - \delta_0)/\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2})$	$1 - \Phi(T)$
$\mu_1 - \mu_2 < \delta_0$	$T < -z_\alpha$	$\Phi(-z_\alpha - (\delta - \delta_0)/\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2})$	$\Phi(T)$
$\mu_1 - \mu_2 \neq \delta_0$	$\lvert T\rvert > z_{\alpha/2}$	$1 - [\Phi(z_{\alpha/2} - (\delta - \delta_0)/\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2})$ $-\Phi(-z_{\alpha/2} - (\delta - \delta_0)/\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2})]$	$2[1 - \Phi(\lvert T\rvert)]$

Sample size calculations

- Sample size calculations center on two questions:
  (i) How small a deviation from the null is it of interest to detect?
  (ii) With what probability do we wish to detect it?

- **Exercise:** *Suppose you wish to test the hypotheses H_0: $\mu = 500\,\text{mL}$ versus H_1: $\mu \neq 500\,\text{mL}$, where μ is the mean amount of a drink in bottles labeled as containing $500\,\text{mL}$, and suppose you assume that the standard deviation is $\sigma = 2\,\text{mL}$. Based on the volumes $X_1,\dots,X_n$ of n randomly selected bottles, you plan to test the hypotheses using the test*

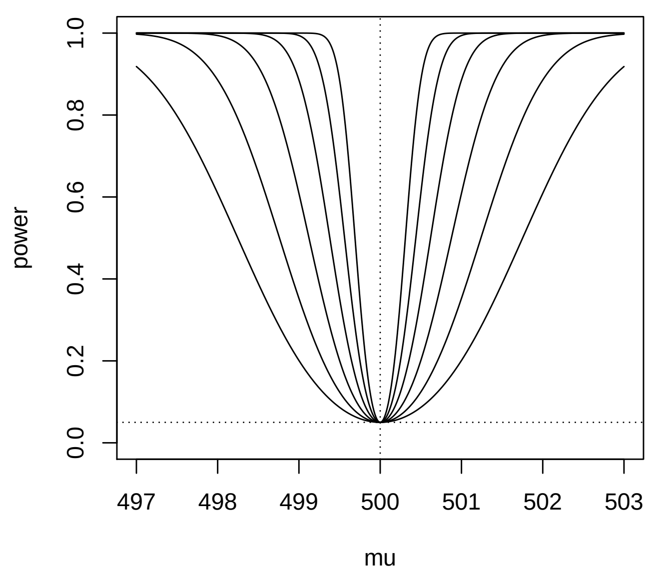
$$\textit{Reject } H_0 \textit{ iff } |\sqrt{n}(\bar{X}_n - 500)/2| > z_{0.025}.$$

  (i) Plot power curves for the sample sizes $n = 5, 10, 20, 40, 80, 160$ across $\mu \in (497, 503)$, assuming that the volumes follow a Normal distribution.
  (ii) Suppose you wish to detect a deviation as small as $1\,\text{mL}$ of the mean from $500\,\text{mL}$ with probability at least 0.80 (deviations of μ from $500\,\text{mL}$ of less than $1\,\text{mL}$ are nothing to worry about, say). Based on the plot, which of the sample sizes $n = 5, 10, 20, 40, 80, 160$ ensure this?
  (iii) What is the smallest of the sample sizes $n = 5, 10, 20, 40, 80, 160$ under which the test has power at least 0.80 for values of μ at least $1\,\text{mL}$ away from $500\,\text{mL}$?

Answers:

(i) The following R code produces the plot:

```
mu.0 <- 500
sigma <- 2
mu.seq <- seq(497,503,length=500)
n.seq <- c(5,10,20,40,80,160)
power <- matrix(NA,length(mu.seq),length(n.seq))

for(j in 1:length(n.seq))
{
    power[,j] <- 1-(pnorm(qnorm(.975)-sqrt(n.seq[j])*(mu.seq - mu.0)/sigma) -
                    pnorm(-qnorm(.975)-sqrt(n.seq[j])*(mu.seq - mu.0)/sigma))
}

plot(mu.seq,power[,1],type="l",ylim=c(0,1),xlab="mu",ylab="power")
for(j in 2:length(n.seq))
{
    lines(mu.seq,power[,j])
}

abline(v=mu.0,lty=3) # vert line at null value
abline(h=0.05,lty=3) # horiz line at size
```



(ii) The sample sizes $n = 40, 80, 160$ ensure this.

(iii) The sample size $n = 40$ is the smallest of these which ensures this.

- In the above exercise, the power curves of the test under several sample sizes were plotted, and these were used to determine a sample size which would give high enough power. We are interested in getting formulas for the smallest sample size under which a test will have at least a certain power under some magnitude of deviation from the null.

- To make sample size calculations, the researcher must specify two values in answer to the two central questions of sample size calculations stated above. Let

  (i) δ^* be the smallest deviation from the null that we wish to detect.

  (ii) γ^* be the smallest probability with which we wish to detect it.

  We want to find the smallest sample size n under which the test has power γ^* to detect a deviation from the null as small as δ^*. We will denote this sample size by n^*.

- Given a parameter $\theta \in \Theta \subset \mathbb{R}$ and given null and alternate hypotheses H_0: $\theta \in \Theta_0$ versus H_1: $\theta \in \Theta_1$, we define the "deviation of θ from the null" as the distance

$$d(\theta; \Theta_0) = \inf_{\tilde{\theta} \in \Theta_0} |\theta - \tilde{\theta}|.$$

  If $\theta \in \Theta_0$, then the deviation $d(\theta; \Theta_0)$ of θ from is equal to zero. If $\theta \notin \Theta_0$, then $d(\theta; \Theta_0)$ is equal to the distance between θ and the value in Θ_0 which is closest to θ.

- If the parameter θ has dimension $d \geq 1$, such that $\theta \in \Theta \subset \mathbb{R}^d$, we define for any set $\Theta_0 \subset \mathbb{R}^d$ the distance

$$d(\theta; \Theta_0) = \inf_{\tilde{\theta} \in \Theta_0} \|\theta - \tilde{\theta}\|_2,$$

  where for any $x \in \mathbb{R}^d$, $\|x\|_2 = (x_1^2 + \cdots + x_d^2)^{1/2}$ is the length of the d-dimensional vector x in d-dimensional Euclidean space. Thus for any θ and $\tilde{\theta}$ in $\mathbb{R}^d$, $\|\theta - \tilde{\theta}\|_2$ is the Euclidean distance in $\mathbb{R}^d$ between θ and $\tilde{\theta}$.

  Examples:

  - For $p \in (0, 1)$, consider H_0: $p = 1/2$ and H_1: $p \neq 1/2$. Then the null set consists of a single point $1/2$, so that the deviation of any p from the null is

$$d(p; \{1/2\}) = \inf_{\tilde{p} \in \{1/2\}} |p - \tilde{p}| = |p - 1/2|,$$

    which is simply the distance between p and $1/2$.

  - For $\mu \in (-\infty, \infty)$, consider H_0: $\mu \leq 10$ and H_1: $\mu > 10$. Then the null set is the set $(-\infty, 10]$, so that the deviation of any μ from the null is

$$d(\mu; (-\infty, 10]) = \inf_{\tilde{\mu} \in (-\infty, 10]} |\mu - \tilde{\mu}| = \begin{cases} 0 & \text{if } \mu \leq 10 \\ \mu - 10 & \text{if } \mu > 10, \end{cases}$$

    so that when H_0 is true the distance is zero and when H_1 is true the distance is the amount by which μ exceeds 10.

– For sets of hypotheses in which the alternate hypothesis takes any of the forms, $\theta > \theta_0$, $\theta < \theta_0$, or $\theta \neq \theta_0$, for some θ_0, we have

$$d(\theta; \Theta_0) = \begin{cases} 0 & \text{if } \theta \in \Theta_0 \\ |\theta - \theta_0| & \text{if } \theta \in \Theta_1. \end{cases}$$

– Suppose θ consists of two parameters so that $\theta = (\theta_1, \theta_2)^T$, and consider the hypotheses H_0: $\theta_1 = \theta_2$ versus H_1: $\theta_1 \neq \theta_2$. Then the null space is $\Theta_0 = \{(x, y) : x = y\}$. In this case for any (θ_1, θ_2) pair, we have

$$d((\theta_1, \theta_2)^T; \Theta_0) = \inf_{\{(x,y):x=y\}} \|(\theta_1, \theta_2)^T - (x, y)^T\|_2 = |\theta_1 - \theta_2|/\sqrt{2},$$

which is the shortest distance of any line segment connecting the point (θ_1, θ_2) to the line given by $y = x$.

- We may now give a general expression for n^* in terms of γ^* and δ^*. Given a parameter $\theta \in \Theta$, null and alternate hypotheses H_0: $\theta \in \Theta_0$ versus H_1: $\theta \in \Theta_1$, and a test which has, under a sample of size n, the power function $\gamma_n(\theta) : \Theta \to [0, 1]$, for $n = 1, 2, \dots$, the smallest sample size n under which the test has power greater than or equal to γ^* to detect deviations from the null as small as δ^* is given by

$$n^* = \inf \left\{ n \in \mathbb{N} : \inf_{\{\theta \in \Theta : d(\theta; \Theta_0) \geq \delta^*\}} \gamma_n(\theta) \geq \gamma^* \right\},$$

where $\mathbb{N} = \{1, 2, 3, \dots\}$ is the set of natural numbers.

- Is the above expression needlessly complicated? Well, we will come to some simpler formulas for tests which we often encounter, but the above expression works for any situation. To get a sample size "recipe" which works for any situation, it has to be a bit abstract. To break it down, read each piece separately: The first infimum over $n \in \mathbb{N}$ indicates that we are looking for a small sample size. The second infimum asks, "what is the lowest power of the test over values of the parameter for which we would like to reject the null?". We want this lowest power to be no lower than γ^*.

- **Exercise (cont):** *Suppose you now wish to test the one-sided set of hypotheses H_0: $\mu \geq 500\,\text{mL}$ versus H_1: $\mu < 500\,\text{mL}$ (it will be easier to begin with a one-sided example), where μ is the mean amount of a drink in bottles labeled as containing $500\,\text{mL}$, and suppose you assume that the standard deviation is $\sigma = 2\,\text{mL}$. Based on the volumes $X_1, \dots, X_n$ of n randomly selected bottles, you plan to test the hypotheses using the test*

$$\textit{Reject } H_0 \textit{ iff } \sqrt{n}(\bar{X}_n - 500)/2 < -z_{0.05}.$$

(i) Make a sketch showing what the power curve should look like.

(ii) Find the smallest sample size under which the test will have a power of at least 0.80 to detect a deviation as small as $1\,\text{mL}$ from the null (suppose it does not matter if the average volume is less than 500 by less than $1\,\text{mL}$).

(iii) *Plot the power curve of the test under this sample size. Also include power curves for the test under a sample with* 10 *fewer observations and a sample with* 10 *more observations.*

Answers:

(ii) We have $\gamma^* = 0.80$ and $\delta^* = 1$. The power function of the test for any sample size n is

$$\gamma_n(\mu) = \Phi(-z_{0.05} - \sqrt{n}(\mu - 500)/2).$$

We are interested the power whenever μ is at least $\delta^* = 1\,\text{mL}$ below 500, that is, when μ is in the set

$$\{\mu \in (-\infty, \infty) : d(\mu; [500, \infty)) \geq 1\} = (-\infty, 499].$$

Since the power function $\gamma_n(\mu)$ is decreasing in μ, its smallest value over $\mu \in (-\infty, 499]$ occurs at $\mu = 499$, so we have

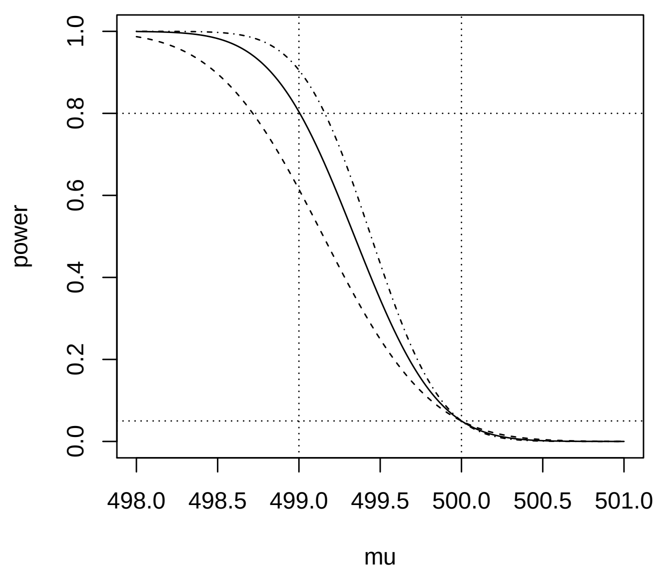
$$\inf_{\{\mu\in(-\infty,\infty):d(\mu;[500,\infty))\geq 1\}} \gamma_n(\mu) = \gamma_n(499) = \Phi(-z_{0.05} - \sqrt{n}(499 - 500)/2).$$

Now we must find the smallest n such that this power is at least $\gamma^* = 0.80$. That is, we need

$$\begin{aligned}
n^* &= \inf\left\{n \in \mathbb{N} : \Phi(-z_{0.05} - \sqrt{n}(499 - 500)/2) \geq 0.80\right\}\\
&= \inf\left\{n \in \mathbb{N} : -z_{0.05} - \sqrt{n}(499 - 500)/2 \geq \Phi^{-1}(0.80)\right\}\\
&= \inf\left\{n \in \mathbb{N} : -\sqrt{n} \geq 2(z_{0.20} + z_{0.05})/(499 - 500)\right\}\\
&= \inf\left\{n \in \mathbb{N} : n \geq 2^2(z_{0.20} + z_{0.05})^2/(499 - 500)^2\right\}\\
&= \lfloor 2^2(z_{0.20} + z_{0.05})^2/(1)^2\rfloor + 1\\
&= \texttt{floor(2**2*(qnorm(.8)+qnorm(.95))**2/(1)**2)+1}\\
&= 25.
\end{aligned}$$

(iii) The following R code makes the plot:

```
mu.0 <- 500
sigma <- 2
mu.seq <- seq(498,501,length=500)
n.seq <- c(15,25,35)
power <- matrix(NA,length(mu.seq),length(n.seq))

for(j in 1:length(n.seq))
{
    power[,j] <-  pnorm(-qnorm(.95) - sqrt(n.seq[j])*(mu.seq - mu.0)/2)
}

plot(mu.seq,power[,1],type="l",ylim=c(0,1),xlab="mu",ylab="power",lty=2)
lines(mu.seq,power[,2],lty=1)
lines(mu.seq,power[,3],lty=4)
abline(v=mu.0,lty=3) # vert line at null value
abline(h=0.05,lty=3) # horiz line at size
abline(v = mu.0 - 1,lty=3)  # vert line at detect boundary
abline(h = 0.8,lty=3)       # horiz line at desired power at detect boundary
```



- If we want a test to have power greater than or equal to γ^*, it is the same as wanting the probability of a Type II error to be less than or equal to $1-\gamma^*$. Let $\beta^* = 1-\gamma^*$ denote from now on a desired upper bound on the Type II error probability.

- **Sample size formulas for tests about a mean:** Suppose you are to draw a random sample $X_1, \dots, X_n$ from a population with mean μ and variance $\sigma^2 < \infty$, and for some μ_0 define $T_n = \sqrt{n}(\bar{X}_n - \mu_0)/S_n$. The following table gives the sample sizes required in order that the power of the test to detect a deviation from the null of size $\delta^* > 0$ be at least γ^*. In the table $\beta^* = 1-\gamma^*$, so that β^* is the desired upper bound on the probability of Type II error. Thus n^* is the smallest sample size required such that for deviations from the null as small as δ^*, the Type II error probability will not exceed β^*.

H_0	H_1	Rej. at α iff	choose n^* as smallest integer greater than
$\mu \leq \mu_0$	$\mu > \mu_0$	$T_n > z_\alpha$	$\sigma^2(z_\alpha + z_{\beta^*})^2/(\delta^*)^2$
$\mu \geq \mu_0$	$\mu < \mu_0$	$T_n < -z_\alpha$	$\sigma^2(z_\alpha + z_{\beta^*})^2/(\delta^*)^2$
$\mu = \mu_0$	$\mu \neq \mu_0$	$\|T_n\| > z_{\alpha/2}$	$\sigma^2(z_{\alpha/2} + z_{\beta^*})^2/(\delta^*)^2$

- **Exercise:** *Verify the sample size formula in the above table for the two-sided test.*

  Answer: For each n, the power function for the test is given by

$$\Phi(-z_{\alpha/2} - \sqrt{n}(\mu - \mu_0)/\sigma) + 1 - \Phi(z_{\alpha/2} - \sqrt{n}(\mu - \mu_0)/\sigma).$$

If we consider the case $\mu > \mu_0$, then as n increases, most of the contribution to the power comes from the right tail

$$1 - \Phi(z_{\alpha/2} - \sqrt{n}(\mu - \mu_0)/\sigma),$$

whereas in the case $\mu < \mu_0$, more of the power comes from the left tail

$$\Phi(-z_{\alpha/2} - \sqrt{n}(\mu - \mu_0)/\sigma).$$

In the case $\mu > \mu_0$, ignoring the contribution from the left tail, we write

$$\begin{aligned} & 1 - \Phi(z_{\alpha/2} - \sqrt{n}(\mu - \mu_0)/\sigma) \geq \gamma^* \\ & \iff 1 - \gamma^* \geq \Phi(z_{\alpha/2} - \sqrt{n}(\mu - \mu_0)/\sigma) \\ \text{(draw a picture!)} & \iff z_{\alpha/2} - \sqrt{n}(\mu - \mu_0)/\sigma \leq z_{\gamma^*} = -z_{\beta^*} \quad (z_{\gamma^*} = -z_{1-\gamma^*} = -z_{\beta^*}) \\ & \iff z_{\alpha/2} + z_{\beta^*} \leq \sqrt{n}(\mu - \mu_0)/\sigma \\ & \iff n \geq \sigma^2(z_{\alpha/2} + z_{\beta^*})^2/|\mu - \mu_0|^2. \end{aligned}$$

In the case $\mu < \mu_0$, ignoring the contribution from the right tail, we write

$$\begin{aligned} & \Phi(-z_{\alpha/2} - \sqrt{n}(\mu - \mu_0)/\sigma) \geq \gamma^* \\ \text{(draw a picture!)} & \iff -z_{\alpha/2} - \sqrt{n}(\mu - \mu_0)/\sigma \geq z_{1-\gamma^*} = z_{\beta^*} \\ & \iff \sqrt{n} \geq -\sigma(z_{\alpha/2} + z_{\beta^*})/(\mu - \mu_0) \\ & \iff n \geq \sigma^2(z_{\alpha/2} + z_{\beta^*})^2/|\mu - \mu_0|^2. \end{aligned}$$

Now if we replace $|\mu - \mu_0|$ with δ^* we have the formula.

Our ignoring one or the other tail based on whether $\mu > \mu_0$ or $\mu < \mu_0$ corresponds to using the approximation

$$\begin{aligned}\gamma_n(\mu) &= \Phi(-z_{\alpha/2} - \sqrt{n}(\mu-\mu_0)/\sigma) + 1 - \Phi(z_{\alpha/2} - \sqrt{n}(\mu-\mu_0)/\sigma) \\ &\approx 1 - \Phi(z_{\alpha/2} - \sqrt{n}|\mu-\mu_0|/\sigma),\end{aligned}$$

which becomes more accurate for larger $|\mu - \mu_0|$.

- **Remark:** Note that the sample size formulas do *not* assume that the sample will be drawn from a Normal population; they assume only that the sample size resulting from the calculation will be large enough for the Central Limit Theorem to have taken effect. If we *knew* that the population was Normal, we could do sample size calculations using the exact distribution of the test statistic $\sqrt{n}(\bar{X}_n - \mu)/S_n$, which would then be the t_{n-1} distribution. However, since the distribution of $\sqrt{n}(\bar{X}_n - \mu)/S_n$ is different for every sample size, we cannot get simple sample size formulas as we have done by using the asymptotic distribution. In practice, even when the population is believed to be Normal, the formulas in the table above are used because they are simple and give answers which are very close to the sample sizes one would obtain using exact calculations.

- **Sample size formulas for tests about a proportion:** Suppose you are to draw a random sample $X_1, \dots, X_n$ from the Bernoulli(p) distribution, and for some p_0 define $Z_n = \sqrt{n}(\hat{p} - p_0)/\sqrt{p_0(1-p_0)}$. The following table gives the sample sizes required in order that the power of the test to detect a deviation from the null of size $\delta^* > 0$ be at least γ^*. In the table $\beta^* = 1 - \gamma^*$.

H_0	H_1	Rej. at α iff	choose n^* as smallest integer greater than
$p \le p_0$	$p > p_0$	$Z_n > z_\alpha$	$\left[\sqrt{(p_0+\delta^*)(1-(p_0+\delta^*))}z_{\beta^*} + \sqrt{p_0(1-p_0)}z_\alpha\right]^2/(\delta^*)^2$
$p \ge p_0$	$p < p_0$	$Z_n < -z_\alpha$	$\left[\sqrt{(p_0-\delta^*)(1-(p_0-\delta^*))}z_{\beta^*} + \sqrt{p_0(1-p_0)}z_\alpha\right]^2/(\delta^*)^2$
$p = p_0$	$p \ne p_0$	$\lvert Z_n\rvert > z_{\alpha/2}$	$\max\left\{\left[\sqrt{(p_0+\delta^*)(1-(p_0+\delta^*))}z_{\beta^*} + \sqrt{p_0(1-p_0)}z_{\alpha/2}\right]^2, \left[\sqrt{(p_0-\delta^*)(1-(p_0-\delta^*))}z_{\beta^*} + \sqrt{p_0(1-p_0)}z_{\alpha/2}\right]^2\right\}/(\delta^*)^2$

- **Exercise:** *Verify the first sample size formula in the above table.*

  Answers: The asymptotic power function for any sample size n is

$$\gamma_n(p) = 1 - \Phi((\sigma_0/\sigma)(z_\alpha - \sqrt{n}(p-p_0)/\sigma_0)).$$

  We are interested in the probability of rejecting H_0 when $p > p_0 + \delta^*$ for some $\delta^* > 0$. The smallest power of the test over all $p > p_0 + \delta^*$ is given by

$$\inf_{p \ge p_0+\delta^*} \gamma_n(p) = \gamma_n(p_0+\delta^*) = 1 - \Phi(\sigma_0/\sqrt{(p_0+\delta^*)(1-(p_0+\delta^*))}(z_\alpha - \sqrt{n}(p_0+\delta^*-p_0)/\sigma_0)).$$

We now find the smallest n such that the above is greater than or equal to γ^*. We have

$$
\begin{aligned}
1 - \Phi(\sigma_0/\sqrt{(p_0 + \delta^*)(1 - (p_0 + \delta^*))}(z_\alpha - \sqrt{n}(p_0 + \delta^* - p_0)/\sigma_0)) \geq \gamma^* \\
\iff \Phi(\sigma_0/\sqrt{(p_0 + \delta^*)(1 - (p_0 + \delta^*))}(z_\alpha - \sqrt{n}\delta^*/\sigma_0)) \leq 1 - \gamma^* \\
\iff \sigma_0/\sqrt{(p_0 + \delta^*)(1 - (p_0 + \delta^*))}(z_\alpha - \sqrt{n}\delta^*/\sigma_0) \leq \Phi^{-1}(1 - \gamma^*).
\end{aligned}
$$

Letting $\beta^* = 1 - \gamma^*$, we have $\Phi(\gamma^*) = z_{\beta^*}$ so that we can write $\Phi^{-1}(1 - \gamma^*) = -z_{\beta^*}$. Then we have

$$
\begin{aligned}
z_\alpha - \sqrt{n}\delta^*/\sigma_0 \leq -\sqrt{(p_0 + \delta^*)(1 - (p_0 + \delta^*))}/\sigma_0 z_{\beta^*} \\
\iff \sqrt{n}\delta^* \geq \sqrt{(p_0 + \delta^*)(1 - (p_0 + \delta^*))}z_{\beta^*} + \sigma_0 z_\alpha \\
\iff n \geq [\sqrt{(p_0 + \delta^*)(1 - (p_0 + \delta^*))}z_{\beta^*} + \sqrt{p_0(1 - p_0)}z_\alpha]^2/(\delta^*)^2.
\end{aligned}
$$

This verifies the formula.

- Quite often sample sizes are determined based on the desired width of a confidence interval for a parameter. One specifies the desired width as well as the confidence level and works backwards to find the minimum sample size required. This strategy does not consider the desired power under certain deviations from the null. For details about the confidence interval approach, see the STAT 512 notes.

- **Exercise:** *Let* $X_1, \dots, X_n$ *represent a random sample from the* $\text{Bernoulli}(p)$ *distribution, where* p *is unknown and suppose there are three researchers:*
  - *Fausto to test* H_0: $p \leq 1/4$ *vs* H_1: $p > 1/4$ *with Rej.* H_0 *iff* $\frac{\hat{p}_n - 1/4}{\sqrt{(1/4)(1-1/4)/n}} > z_\alpha$
  - *Inés to test* H_0: $p \geq 1/4$ *vs* H_1: $p < 1/4$ *with Rej.* H_0 *iff* $\frac{\hat{p}_n - 1/4}{\sqrt{(1/4)(1-1/4)/n}} < -z_\alpha$
  - *Germán to test* H_0: $p = 1/4$ *vs* H_1: $p \neq 1/4$ *with Rej.* H_0 *iff* $\left|\frac{\hat{p}_n - 1/4}{\sqrt{(1/4)(1-1/4)/n}}\right| > z_{\alpha/2}$

  (i) If each researcher wishes to detect a deviation from the null as small as 0.10 *with probability at least* 0.80*, what sample size should each use?*

  (ii) Using these sample sizes, plot the power curves for the three researchers' tests.

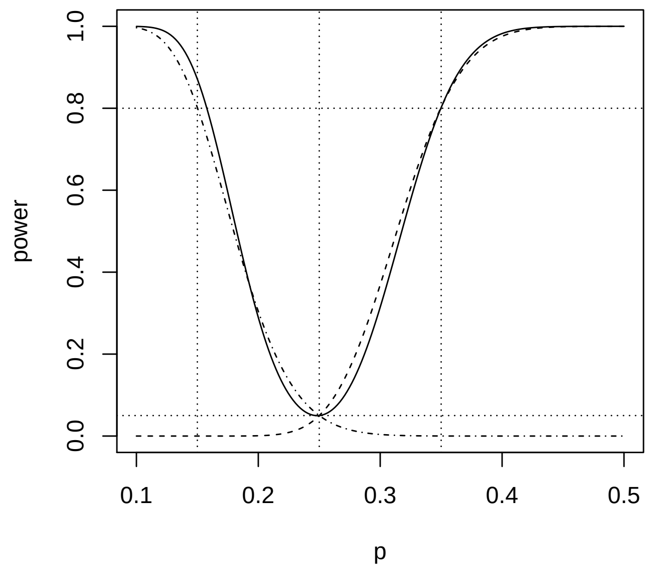
Answers:

(i) Using the sample size formulas, we get

- for Fausto, $n^* = 125$.
- for Inés, $n^* = 103$.
- for Germán, $n^* = 157$.

(ii) The following R code makes the plot:

```
p.seq <- seq(0.1,.5,length=200)
p.0 <- 1/4
sig.0 <- sqrt(p.0*(1-p.0))
sig <- sqrt(p.seq*(1-p.seq))

pwr.gt <- 1-pnorm(sig.0/sig*(qnorm(.95)-sqrt(125)*(p.seq-p.0)/sig.0))
pwr.lt <- pnorm(sig.0/sig*(-qnorm(.95)-sqrt(103)*(p.seq-p.0)/sig.0))
pwr.neq <- pnorm(sig.0/sig*(-qnorm(.975)-sqrt(157)*(p.seq-p.0)/sig.0))
            +1-pnorm(sig.0/sig*(qnorm(.975)-sqrt(157)*(p.seq-p.0)/sig.0))

plot(p.seq, pwr.neq,type="l",ylim=c(0,1),xlab="p",ylab="power")
lines(p.seq, pwr.gt,lty=2)
lines(p.seq, pwr.lt,lty=4)

abline(v=p.0,lty=3)  # vert line at null value
abline(h=0.05,lty=3) # horiz line at size

abline(v=p.0-.1,lty=3)
abline(v=p.0+.1,lty=3)
abline(h=.8, lty=3)
```



- **Sample size formulas for tests comparing two means:** Suppose you are to draw random samples

$$X_{11}, \ldots, X_{1n_1} \text{ from a distribution with mean } \mu_1 \text{ and variance } \sigma_1^2 < \infty$$
$$\text{and } X_{21}, \ldots, X_{2n_1} \text{ from a distribution with mean } \mu_2 \text{ and variance } \sigma_2^2 < \infty,$$

and define the quantity

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (0)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}},$$

on the basis of which you plan to test hypotheses about the difference $\mu_1 - \mu_2$. The following table gives the smallest values of n such that for all n_1 and n_2 satisfying

$$n_1 \geq \left(\frac{\sigma_2}{\sigma_1 + \sigma_2}\right) n \quad \text{and} \quad n_2 \geq \left(\frac{\sigma_2}{\sigma_1 + \sigma_2}\right) n$$

the power of the test to detect a difference in means in the direction of the alternative of size $\delta^* > 0$ will be at least γ^*. In the table $\beta^* = 1 - \gamma^*$.

H_0	H_1	Rej. at α iff	choose n greater than or equal to
$\mu_1 - \mu_2 \leq 0$	$\mu_1 - \mu_2 > 0$	$T > z_\alpha$	$[(z_{\beta^*} + z_\alpha)(\sigma_1 + \sigma_2)]^2/(\delta^*)^2$
$\mu_1 - \mu_2 \geq 0$	$\mu_1 - \mu_2 < 0$	$T < -z_\alpha$	$[(z_{\beta^*} + z_\alpha)(\sigma_1 + \sigma_2)]^2/(\delta^*)^2$
$\mu_1 - \mu_2 = 0$	$\mu_1 - \mu_2 \neq 0$	$\|T\| > z_{\alpha/2}$	$[(z_{\beta^*} + z_{\alpha/2})(\sigma_1 + \sigma_2)]^2/(\delta^*)^2$

- **Exercise:** *Verify the first sample size formula.*

  Answers: Firstly, for any $n > 0$, let

$$n_1 = \left(\frac{\sigma_1}{\sigma_1 + \sigma_2}\right) n \quad \text{and} \quad n_2 = \left(\frac{\sigma_2}{\sigma_1 + \sigma_2}\right) n,$$

  ignoring the fact that these may not be whole numbers. Then for any $n > 0$, we have

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{1}{n}\left[\frac{\sigma_1^2}{\sigma_1/(\sigma_1+\sigma_2)} + \frac{\sigma_2^2}{\sigma_2/(\sigma_1+\sigma_2)}\right] = \frac{1}{n}[\sigma_1(\sigma_1+\sigma_2) + \sigma_2(\sigma_1+\sigma_2)] = \frac{1}{n}(\sigma_1+\sigma_2)^2,$$

  which is the minimum variance of $\bar{X}_1 - \bar{X}_2$ under a fixed total sample size n. Under these choices of n_1 and n_2, the power function of the test for large n is approximately

$$\gamma_n(\delta) = 1 - \Phi(z_\alpha - \sqrt{n}\delta/(\sigma_1+\sigma_2)).$$

  The smallest power of the test over all $\delta > \delta^*$ is given by

$$\inf_{\delta > \delta^*} \gamma_n(\delta) = \gamma_n(\delta^*) = 1 - \Phi(z_\alpha - \sqrt{n}\delta^*/(\sigma_1+\sigma_2)).$$

  We now find the smallest value of n such that the above is greater than or equal to γ^*. We have

$$\begin{aligned} & 1 - \Phi(z_\alpha - \sqrt{n}\delta^*/(\sigma_1+\sigma_2)) \geq \gamma^* \\ \iff & z_\alpha - \sqrt{n}\delta^*/(\sigma_1+\sigma_2) \leq \Phi^{-1}(1-\gamma^*). \end{aligned}$$

  Letting $\beta^* = 1 - \gamma^*$, we may write $\Phi^{-1}(1-\gamma^*) = -z_{\beta^*}$, so we have

$$\begin{aligned} & z_\alpha - \sqrt{n}\delta^*/(\sigma_1+\sigma_2) \leq -z_{\beta^*} \\ \iff & -\sqrt{n}\delta^*/(\sigma_1+\sigma_2) \leq -z_{\beta^*} - z_\alpha \\ \iff & \sqrt{n}\delta^* \geq (z_{\beta^*} + z_\alpha)(\sigma_1+\sigma_2) \\ \iff & n \geq [(z_{\beta^*} + z_\alpha)(\sigma_1+\sigma_2)]^2/(\delta^*)^2. \end{aligned}$$

- **Exercise:** *A researcher wishes to compare the means μ_1 and μ_2 of two populations by testing the hypotheses H_0: $\mu_1 - \mu_2 = 0$ versus H_1: $\mu_1 - \mu_2 \neq 0$. A pilot study resulted in estimates of the variances of the two populations equal to $\hat{\sigma}_1^2 = 1.22$ and $\hat{\sigma}_2^2 = 0.26$. Suppose that a difference between μ_1 and μ_2 less than 0.50 units is not practically meaningful, but the researcher wishes to find a difference if it is greater than 0.50 units with probability at least 0.90. The researcher will reject H_0 if*

$$\left|\frac{\bar{X}_1 - \bar{X}_2 - (0)}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}\right| > 2.575829.$$

  (i) Recommend sample sizes n_1 and n_2 to the researcher.

(ii) *Plot the power curve of the test under these sample sizes, assuming* $\sigma_1^2 = 1.22$ *and* $\sigma_2^2 = 0.26$. *Add to the plot the power curve resulting from using the same total sample size as in part (i), but when samples of equal size are drawn from the two populations. Explain how the power curves are different and why.*

Answers:

(i) We have $2.575829 = z_{\alpha/2}$ for $\alpha = 0.01$. With $\gamma^* = 0.90$ and $\delta^* = 0.20$, the formula gives

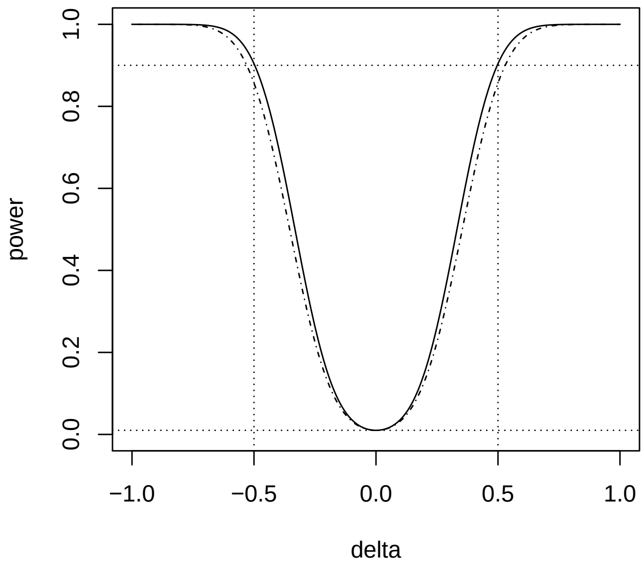
$$
\begin{aligned}
&[(z_{0.10}+z_{0.005})(\sqrt{1.22}+\sqrt{0.26})]^2/(0.20)^2\\
&\qquad = \texttt{((qnorm(.9)+qnorm(.995))*(sqrt(1.22)+sqrt(0.26)))**2/(0.5)**2}\\
&\qquad = 155.1271.
\end{aligned}
$$

So we recommend

$$
n_1 = \left\lfloor \left(\frac{\sqrt{1.22}}{\sqrt{1.22}+\sqrt{0.26}} \right) 156 \right\rfloor + 1 = 107 \text{ and } n_2 = \left\lfloor \left(\frac{\sqrt{0.26}}{\sqrt{1.22}+\sqrt{0.26}} \right) 156 \right\rfloor + 1 = 50.
$$

(ii) The following R code produces the plot:

```
sig1 <- sqrt(1.22)
sig2 <- sqrt(0.26)
d.star <- 0.5
alpha <- 0.01
beta.star <- 0.1

n.pre <- ceiling(((qnorm(1-beta.star)+qnorm(1-alpha/2))*(sig1+sig2))^2/d.star^2)
n1 <- ceiling(sig1/(sig1 + sig2) * n.pre)
n2 <- ceiling(sig2/(sig1 + sig2) * n.pre)
n <- n1 + n2

d.seq <- seq(-1,1,length=500)

power.n1n2opt <- 1-(pnorm(qnorm(1-alpha/2)-d.seq/sqrt(sig1^2/n1+sig2^2/n2))
                -pnorm(-qnorm(1-alpha/2)-d.seq/sqrt(sig1^2/n1+sig2^2/n2)))
power.n1n2eq <- 1-(pnorm(qnorm(1-alpha/2)-d.seq/sqrt(sig1^2/(n/2)+sig2^2/(n/2)))
                -pnorm(-qnorm(1-alpha/2)-d.seq/sqrt(sig1^2/(n/2)+sig2^2/(n/2))))

plot(d.seq, power.n1n2opt,type="l",ylim=c(0,1),xlab="delta",ylab="power",lty=1)
lines(d.seq, power.n1n2eq,lty=4)

abline(v = d.star,lty=3)
abline(v = -d.star,lty=3)
abline(h = alpha,lty=3)
abline(h = 1 - beta.star,lty=3)
```



The power curve under $n_1 = n_2$ is lower for all $\delta \neq 0$ than the power curve under the choices of n_1 and n_2 which take into account the different variances of the two populations.