Lab11: Hypothesis Testing Concepts & One-Sample Tests

Handout date: Wednesday, November 11, 2020
Due date: Friday, November 20, 2020 by midnight into the eLearning link SubmitLab11
This lab counts 4 % toward your total grade

Note: Study for this lab Chapter 10 "Power Analysis" in Kabacoff "R in Action. Data Analysis and Graphics with R". Available online at UTD's Library.

Task 1: Hypothesis Testing from a Binomial Distribution (2points)

Assume that a developer takes 225 random and independent soil samples in a 150 acres plot, which is zoned for a new residential development. From previous studies in the broader area around the building site it is known that 24% of the homesteads need subsequent foundation repairs because they are built on shifting clay soil, which would require to lay substantially more expensive foundations to avoid foundation repairs (welcome to the DFW real estate market).

The builder is interested whether it worthwhile to immediately build the more expensive foundations for all homes in the new development, or just to set some money aside for warrantee work to fix approximately 24% of the foundations build by her company.

Out of the 225 surface samples taken 95 indicate potential clay soil problems. The developer is willing to take a 5% risk of making a wrong decision, which is spending more money on reinforced all foundations even though it is not necessary. On the other hand, if a substantial number of foundations fail, he will make a loss and by exhausting the warrantee fund.

[a] Verbally explain whether the given scenario leads to a *one*- or a *two*-sided hypothesis test and what the *null* and the *alternative* hypothesis are. (0.3 points)

Comment: The given scenario leads to a one-side hypothesis test because the question requires to test whether it is worthwhile to build the more expensive foundations for all homes in the new development. The expenditure is decided by the proportion of clay soil in the study area because it costs more to build foundations on clay soil.

$$H_0: \text{The proportion of clay soil is smaller or equal to 24\%}$$
$$H_1: \text{The proportion of clay soil is larger than 24\%}$$

[b] Formulate H_0 and H_1 in a statistical notation. (0.1 points)

$$H_0: \pi \leq 0.24$$
$$H_1: \pi > 0.24$$

[c] Based on the sample observations perform an exact test of the null hypothesis using the binomial distribution. Try to exhaust the error probability α as closely as possible. Properly interpret the test outcome in terms of its *prob*-value. Execute the test using R-code and show your code. (0.5 points)

```
pi.H0 <- 0.24
n <- 225
alpha <- 0.05
( x.hi <- qbinom(1-alpha,size=n, prob=pi.H0) ) # upper critical value
[1] 65
( pi.hi <- pbinom(x.hi,size=n, prob=pi.H0,lower.tail=FALSE) ) # exact
upper alpha
[1] 0.03849099
( pi.hi <- pbinom(x.hi-1,size=n, prob=pi.H0,lower.tail=FALSE) ) #
exact upper alpha for pi.hi-1
[1] 0.0527559
( pi.hi <- pbinom(95-1,size=n, prob=pi.H0,lower.tail=FALSE) ) # p-
value
[1] 1.378579e-09
```

Comment: For the count 65 and the corresponding tail probability is 0.03849099.

However, this error probability doesn't cover the probability at the critical value since the definition in pbinom() is $\Pr[X > x]$ when the options lower.tail=FALSE is set. In other words, the probability of $\Pr[X > 100] = 0.03849099$ is a cumulative probability from 66 to 225. Therefore, the critical value is 64= 65 - 1. Because the binomial distribution is a discrete distribution, we can only obtain an alpha error approximately equal to $\alpha = 0.05$.

The *p*-value for 95 is 1.378579e-09. Therefore, we can reject the null hypothesis.

To simply the test the function **binom.test( )** can be used:

```
binom.test(95, 225, p=0.24, alternative = "greater")
     Exact binomial test
data:  95 and 225
number of successes = 95, number of trials = 225, p-value = 1.379e-09
alternative hypothesis: true probability of success is greater than
0.24
95 percent confidence interval:
 0.3668522 1.0000000
sample estimates:
probability of success
            0.4222222
```

[d] Perform the same test using the normal approximation for the binomial distribution. Properly interpret the test outcome in terms of its *prob*-value. Execute the test using R-code and show your code. (0.5 points)

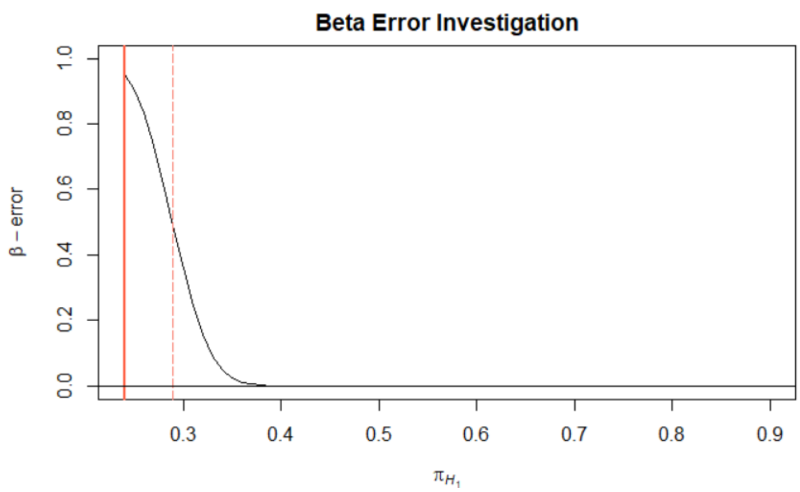
```
(p <- 95/225)
```

```
[1] 0.4222222
( z <- (p-pi.H0)/sqrt((pi.H0*(1-pi.H0))/n) )
[1] 6.400003
pnorm(z, mean=0, sd=1, lower.tail = F)
[1] 7.768693e-11
```

Comment: The test statistic is $z = \frac{p-\pi_0}{\sqrt{\pi_0*(1-\pi_0)/n)}}$ (see BBR page 346) The one-sided *p*-value of the normal test is$\Pr(Z > z) = 7.768693e - 11$which is smaller than 0.05, so we can reject the null hypothesis.

[e] Plot the β-error, i.e., $\beta = 1 - power$, of the exact test assuming the true population probability ranges from $\pi \in [0.24, 0.90]$, i.e., **`pi.H1 <- seq(0.25, 0.70, by=0.01)`**. Show your code and interpret the graph. (0.4 points)

```
library(binom) # installation required
pi.H1 <- seq(0.24, 0.90, by=0.01) # value range of pi under H1
beta <- 1 - binom.power(pi.H1, n=n, p=pi.H0, alpha=alpha,alternative =
"greater", method="exact")
plot(pi.H1, beta, type="l", ylim=c(0,1),
xlab=expression(italic(pi[H[1]])),                ylab=expression(beta-
error),main="Beta Error Investigation" )
abline(v=pi.H0, col="tomato",lwd=2) # value pi under H0
abline(v=(x.hi)/n, col="salmon",lty=5) # critical values of pi
abline(h=0)
```



Comment: As the hypothetical population probability π increases with regards to $\pi_0 = 0.24$, the β-error decreases. The β-error is approximately $\beta \approx 0.4$ at the critical probability $\pi_{crit} = \frac{64}{225} = 0.285$ The increasing difference between the tested population and the reference population causes the declining β-error.

[f] Why is the β-error at $\pi_0 = 0.24$ equal to $\beta_{\pi=\pi_0} = 1 - \alpha$? At which hypothetical value of the estimated sample proportion do you think the beta error becomes acceptable. Justify. (0.2 points)

Comment: At $\pi_0 = 0.24$ the distribution of the test statistic under the null hypothesis is exactly identical to the distribution assuming a hypothetical population with the parameter $\pi = 0.24$. For this hypothetical population distribution, the power at the critical value of $\pi_{critical} = 0.285$ is equal to the α-error. Therefore, the beta error becomes β = 1 - α = 0.95.

Hints: The R-functions **`qbinom( )`** and **`pbinom( )`** allow you to perform the exact test (see Chapter 8 script **`ExactBinomialTest.r`**) and the function **`pnorm( )`** will give you the error probability of the approximate test. The function **`binom::binom.power( )`** calculates the power.

Task 2: Testing for Correlation (1 point)

In Lab 5 Task 8 you visualized the correlation pattern of 3 datasets. Continue with the three data-frames **`cc1`**, **`cc2`** and **`cc3`** in workspace **`Part1Data.Rdata`** to test whether pairs of variables are significantly correlated. See the function **`cor.test()`** for the significance test. For all tests assume that an error probability of $\alpha = 0.05$ and use independence, i.e., $\rho_0 = 0.0$, as benchmark. Interpret the *prob*-values and the calculated confidence intervals. Draw your conclusions about the stated hypotheses. Make sure that you properly account for the one- and two-sided test scenarios.

[a] Test the hypotheses $H_0: \rho \leq \rho_0$ against $H_1: \rho > \rho_0$ for the variable pair in **`cc1`**. Interpret the output. (0.2 points)

```
load("Part1Data.RData")
cor.test(cc1$X1, cc1$X2, alternative="greater", method="pearson")
     Pearson's product-moment correlation

data:  cc1$X1 and cc1$X2
t = -0.57885, df = 998, p-value = 0.7186
alternative hypothesis: true correlation is greater than 0
95 percent confidence interval:
 -0.07029893  1.00000000
sample estimates:
        cor
-0.01831999
```

<u>Comment</u>: The *p*-value is larger than 0.05 and the confidence interval covers both negative and positive correlation levels. Therefore, we fail to reject the null hypothesis.

[b] Test the hypotheses $H_0: \rho = \rho_0$ against $H_1: \rho \neq \rho_0$ for the variable pair in **`cc2`**. Interpret the output. (0.2 points)

```
cor.test(cc2$X1, cc2$X2, alternative="two.sided", method="pearson")
     Pearson's product-moment correlation
data:  cc2$X1 and cc2$X2
t = -21.71, df = 998, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.6070443 -0.5227251
sample estimates:
```

```
        cor
-0.5663649
```

Comment: The *p*-value is much smaller than 0.05 and the confidence interval does not cover zero, so we reject the null hypothesis.

[c] Test the hypotheses $H_0: \rho \geq \rho_0$ against $H_1: \rho \leq \rho_0$ for the variable pair in **cc3**. Interpret the output. (0.2 points)

```
cor.test(cc3$X1, cc3$X2, alternative="less", method="pearson")
      Pearson's product-moment correlation

data:  cc3$X1 and cc3$X2
t = 3.1209, df = 998, p-value = 0.9991
alternative hypothesis: true correlation is less than 0
95 percent confidence interval:
 -1.0000000  0.1495937
sample estimates:
       cor
0.09831319
```

Comment: The *p*-value is larger than 0.05 and the confidence interval covers both negative and positive correlation levels. Therefore, we fail to reject the null hypothesis.

[d] Use the function **pwr::pwr.r.test( )** to evaluate for $\alpha = 0.05$ and $\rho = 0.5$ the required sample sizes to achieve a power $1 - \beta$ of 0.7, 0.8 and 0.9. Explain how the power and sample size relate to each other. You may want to think in terms of the test statistic distributions assuming the null and the alternative hypotheses are true. (0.2 points)

```
pwr::pwr.r.test(r=0.5, sig.level=0.05, power=0.7)
     approximate correlation power calculation (arctangh transformation)
              n = 22.71375
              r = 0.5
      sig.level = 0.05
          power = 0.7
    alternative = two.sided
pwr::pwr.r.test(r=0.5, sig.level=0.05, power=0.8)
     approximate correlation power calculation (arctangh transformation)

              n = 28.24841
              r = 0.5
      sig.level = 0.05
          power = 0.8
    alternative = two.sided
pwr::pwr.r.test(r=0.5, sig.level=0.05, power=0.9)
```

```
     approximate correlation power calculation (arctangh transformation)

              n = 37.03547
              r = 0.5
      sig.level = 0.05
          power = 0.9
    alternative = two.sided
```

Comment: We observe that a larger sample is needed to obtain a higher power of a test. When the sample size is increases, the standard error of the test statistic shrinks, which decreases the overlapped areas of the distribution of the test statistic under H_0 and under H_1. Due to power being $power = 1 - \beta$, a shrinking β-error causes the power to grow.

[e] Use the function **pwr::pwr.r.test( )** to evaluate for a sample size of $n = 35$ and $\rho = 0.5$ the power for given error probabilities α of 0.1, 0.05 and 0.01. Explain how the power and the error probability relate to each other. You may want to think in terms of the test statistic distributions assuming the null and the alternative hypotheses are true. (0.2 points)

```
pwr::pwr.r.test(r=0.5, sig.level=0.10, n=35)
     approximate correlation power calculation (arctangh transformation)

              n = 35
              r = 0.5
      sig.level = 0.1
          power = 0.9339196
    alternative = two.sided
pwr::pwr.r.test(r=0.5, sig.level=0.05, n=35)
     approximate correlation power calculation (arctangh transformation)

              n = 35
              r = 0.5
      sig.level = 0.05
          power = 0.8820248
    alternative = two.sided
pwr::pwr.r.test(r=0.5, sig.level=0.01, n=35)
     approximate correlation power calculation (arctangh transformation)

              n = 35
              r = 0.5
      sig.level = 0.01
          power = 0.7087765
    alternative = two.sided
```

Comment: The power tends to decrease with shrinking an error probability. The critical value moves toward to the upper tail when the significance level decreases. The area of β-error on the left side of the critical value increases, so the power declines.

Task 3: One-Sample Testing for a Given Expected Value (1 point)

A stream has been monitored weekly for several years. Its total dissolved solids in the stream is 60 parts per million. Overall, the distribution of the dissolved solids appears to be normal distributed without systematic variation.

Following recent changes in the land-use within the catchment area new weekly samples were taken for 36 weeks. These samples indicate that the dissolved solids have increased on average to 68 parts per million with a week to week standard deviation of 16 parts per million.

[a] Perform the six steps of a classical hypothesis testing using a one-sided hypothesis $H_0: \mu \leq 60$ against $H_1: \mu > 60$ with a given error probability of $\alpha = 0.1$. (0.3 points)

Note: Since the sample size is larger than 30, the normal approximation rather than the exact *t*-distribution can be used.

(1) Formulating hypothesis: $H_0: \mu \leq 60$ against $H_1: \mu > 60$

(2) The distribution of the test statistic $z = \frac{\bar{X}-60}{16/\sqrt{36}}$ under the null hypothesis is $z \sim N(0,1)$ based on a sample size of n = 36.

(3) Selection of a level of significance: $\alpha = 0.1$

(4) Critical value:

```
qnorm(0.1,0,1,lower.tail = F)
[1] 1.281552
```

(5) Given the observed sample data the test statistic becomes $z = \frac{\bar{X}-\mu_0}{s/\sqrt{n}} = \frac{68-60}{16/\sqrt{36}} = 3$

(6) Because the test statistic is larger than the critical value, we can reject the null hypothesis.

[b] Calculate the prob-value of the observed sample. (0.1 points)

```
pnorm(3, 0, 1, lower.tail=F)

[1] 0.001349898
```

[c] Use the **pwr::pwr.t.test( )** function to fill in the missing values into the table below. (0.2 points)

```
pwr::pwr.t.test(d = 0.25, sig.level = 0.05, n=36,type = 'one.sample',alternative = "two.sided")
pwr::pwr.t.test(d = 0.25, sig.level=0.01, n=36,type = 'one.sample',alternative = "two.sided")
pwr::pwr.t.test(power = 0.5, sig.level=0.01, n=36,type = 'one.sample',alternative = "two.sided")
```

```
pwr::pwr.t.test(power = 0.5, sig.level=0.01, n=144,type = 'one.sample',alternative = "two.sided")
pwr::pwr.t.test(power = 0.99, sig.level=0.01,d = 0.50,type = 'one.sample', alternative = "two.sided")
```

Scenario:	Effect Size:	Sample Size:	α-error:	Power:
1	0.25	36	0.05	**0.31**
2	0.25	36	0.01	**0.13**
3	**0.45**	36	0.01	0.5
4	**0.22**	144	0.01	0.5
5	0.50	**100**	0.01	0.99

[d] Address the following questions in a sentence or two (0.4 points)

Why is the power shrinking when the α-error shrinks?

Comment: The power tends to decrease with shrinking error probability. The critical value moves toward to the upper tail when the significance level decreases. The area of the β-error on the left side of the critical value increases, so the power declines.

Why is the power shrinking for a smaller effect size?

Comment: When the effect size is smaller, the difference between μ_{1970} and μ_{1980} decreases. Therefore, the distribution of the test statistic under the alternative hypothesis moves closer to the distribution under the null hypothesis while the critical value remains the same. Consequently, the β-error increases and the power = 1 - β shrinks.

Why is the power increasing of a larger sample size?

Comment: We observe that a larger sample is needed to obtain a higher power of a test. When the sample size is increases, the standard error of the test statistic shrinks, which decreases the overlapped areas of the distribution of the test statistic under H_0 and under H_1. Therefore, the β-error shrinks causing the power to grow.

Can power analysis be used to determine a required sample size?

Comment: Yes. When the desired effect size, α-error, and power are fixed, sample size is determined