Pair_Programming_Module_06_simple_hypothesis_testir

HDS

2024-08-30

Pair Programming Exercise 6 DSE5001

Hypothesis tests based on Confidence intervals

HD Sheets, August 30, 2024 checked 01/03/2025

#Student Info

Your name:Ryan Waterman Your team-mates names: Trinity Tobin

Counts or proportions

Diez et al problem 5.21

5.21 Minimum wage, Part I.

Do a majority of US adults believe raising the minimum wage will help the economy, or is there a majority who do not believe this? A Rasmussen Reports survey of a random sample of 1,000 US adults found that 42% believe it will help the economy. Conduct an appropriate hypothesis test to help answer the research question.

We can do a basic hypothesis test using the confidence interval on the proportion

p=0.42 and n=1000 for this survey

the standard error in the proportion p is (p(1-p)/n)^0.5

```
(.42*(1-.42)/1000)^0.5
```

```
## [1] 0.01560769
```

so the Standard error in p is 0.0156,

the 95% confidence interval is .42+-1.96*0.0156

```
1.96*.0156
```

```
## [1] 0.030576
```

So this survey indicates that the confidence interval for the belief that raising the minimum wage is held by.42 =-.0306, or roughly 39% to 45%

For a two tailed test, we simply note that the 95% confidence excludes 50%, and reject the hypothesis that the support is 50%. This is using the symmetric confidence interval, from the lower 2.5% bound to the upper 97.5 bound

But the question asked “Is there a major that do not belief this”, which implies a one tailed test, the hypothesis indicates the result not just excludes 50%, but is above it. What we want is the 95% upper bound.

We the mean p at 0.42 with sigma=0.0156, we can use the normal model to find the 95% upper quantile under the null

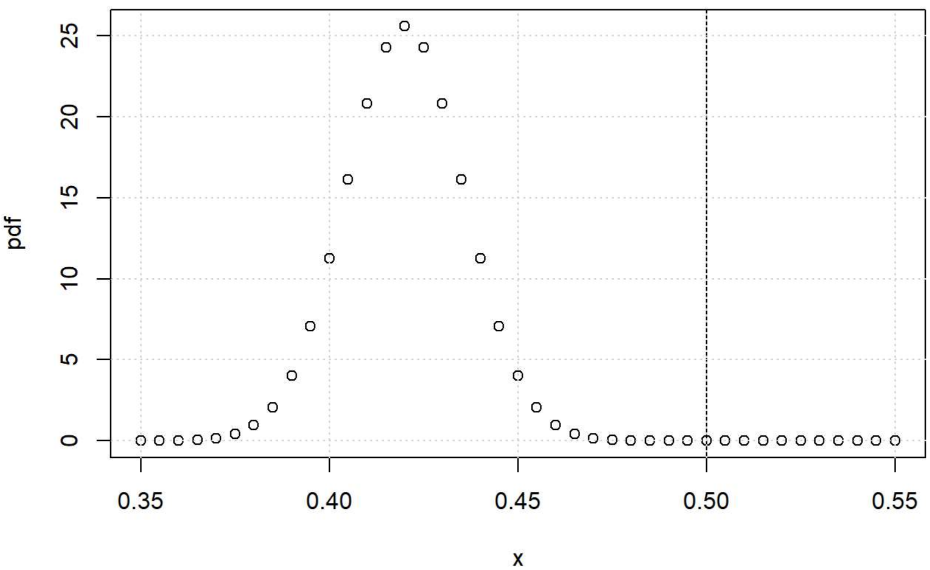
```
qnorm(.95,.42,.0156)
```

```
## [1] 0.4456597
```

That tells use the 95% upper bound for this survey is 44.5%, so there is only a 5% chance of reaching 44.5% on a random survey.

Here is the visual situation

```
x=seq(0.35, 0.55, 0.005)
pdf=dnorm(x,0.42,0.0156)
plot(x,pdf)
abline(v=0.50)
grid()
```



This is the probability of the p value taking on each x value, the probability of the value being 0.5 or higher is very low.

We could use the cummulative distribution to find this, the chance that p is 0.5 or greater, as 1-p(x<.5)

```
1-pnorm(0.5, 0.42, 0.0156)
```

```
## [1] 1.462588e-07
```

Question/action

Complete this on your own

Question 5.23

Suppose we gave the drug MyroShoe to 100 people and found that 54% of them reported that the drug made their feet feel better. What is the proportion p and the 95% confidence interval on p? What do you think this means about MyroShoe?

(p(1-p)/n)^0.5

```
p=0.54
n=100

SE=(p*(1-p)/n)^0.5

ci=1.96*SE

SE
```

```
## [1] 0.04983974
```

```
ci
```

```
## [1] 0.0976859
```

Based on the standard error, this drug only seems to be effective about half of the time.

If you give people medical attention, they tend to feel better. If you give them a harmless pill, it often appears to work, this is of course the placebo effect. In the same trial, 120 people got the placebo and 64 of those patients said their feet felt better. What was the proportion for this group? What was the confidence interval?

```
p=64/120
n=120

SE=(p*(1-p)/n)^0.5

ci=1.96*SE

p
```

```
## [1] 0.5333333
```

```
SE
```

```
## [1] 0.045542
```

```
ci
```

```
## [1] 0.08926233
```

In this group, p was 53.3% and the standard error/confidence interval were the same, indicating that the placebo was as effective as the drug.

Continuous Variables

Confidence interval for the mean

1.) Suppose our typical monthly sales over the last 36 months have been $15,431 per month with a standard deviati0n of $1200

What is the 95% confidence interval on our sales?

What is the chance the real average is under $15,000 over the long term?

---

mean =15,431 sigma=1200 n=36

standard error in the mean is sigma/n^0.5

```
1200/(36)^0.5
```

```
## [1] 200
```

The standard error in the mean is 200, so our 95% confidence interval is 15,431 +- 1.96*200

```
se=200
upper_bound=15431+se*1.96
lower_bound=15431-se*1.96
print(se)
```

```
## [1] 200
```

```
print(upper_bound)
```

```
## [1] 15823
```

```
print(lower_bound)
```

```
## [1] 15039
```

So the 95% bound excludes 15000,

we could find the probability of 15000 or less using the cdf p(x<15000)

```
pnorm(15000,15431,200)
```

```
## [1] 0.01558092
```

So a 1.55% chance our true or underlying mean sales rate is 15000 or less

Question/Action

We want our average profit per sales order to be over 100 dollars. In a one month interval, we have 55 sales with an average profit of 98 dollars per sale and a standard deviation of $27.5.

Do we have meaningful evidence the the average profit is below 100 dollars (so we need to make changes in the business) or is this just a random fluctuation.

```
sd=27.5
n=55

SE=sd/(n)^0.5

upper=98+(se*1.96)
lower=98-(se*1.96)

SE
```

```
## [1] 3.708099
```

```
upper
```

```
## [1] 490
```

```
lower
```

```
## [1] -294
```

No, the standard deviation is very large compared to the mean, and the sample size is small, which leads to meaningless data.