STAT 2507

Lab 4

To access these slides...

GitHub: www.github.com/melissavanbussel

Today’s Tutorial

- Common mistakes from T2: I’ll email you once I know! (Same with A4)
- Quick review: CLT and hypothesis test for a population mean
- Using SPSS to compare sampling distribution of means, and perform a hypothesis test for a population mean

Review: CLT

- The sampling distribution of the mean of a simple random sample of size n from <u>any population</u> with mean μ and variance σ^2 is approximately normal when n is “large enough” (at least 30).

That is, $\bar{X}$ is approximately normal with mean μ and variance σ^2/n .

$$\bar{X} \overset{\text{approx}}{\sim} N(\mu, \sigma^2/n).$$

Hypothesis test for a population mean

- We want to test if the average cholesterol level drop for some medication is greater than 4 units.
  - This is a hypothesis test for a **population mean**
- Which of the following cases apply to this question?

- **Case (i)-Two-tailed test:** $H_0 : \mu = \mu_0$ versus $H_a : \mu \neq \mu_0$
- **Case (ii)-Right-tailed test:** $H_0 : \mu = \mu_0$ versus $H_a : \mu > \mu_0$
- **Case (iii)-Left-tailed test:** $H_0 : \mu = \mu_0$ versus $H_a : \mu < \mu_0$

Hypothesis test for a population mean

- Our null hypothesis is that the average cholesterol drop is equal to 4
- Our alternative hypothesis is that the average cholesterol drop is greater than 4
- Therefore, we have a **one-tailed test**

- **Case (ii)-Right-tailed test:** $H_0 : \mu = \mu_0$ versus $H_a : \mu > \mu_0$

H_0: $\mu = 4$ (or $\mu \leq 4$) $\quad\quad$ H_a: $\mu > 4$

Hypothesis test for a population mean

- Which of the following cases apply for the data we have?
  - We are given the Excel file shown on the right

1. **Case 1:** σ is known
2. **Case 2:** σ is **not** known and sample size is large $n \geq 30$
3. **Case 3:** σ is **not** known and sample size is small $n < 30$

1	Drop
2	6.2
3	5.2
4	8.2
5	4.2
6	3.1
7	5.3
8	3.9
9	6.5
10	7.1
11	6.2
12	9.1
13	8.2
14	5.3
15	1.2
16	2.2
17	3.4
18	6.3
19	5.2
20	3.1
21	1.9
22	1.8
23	2
24	3.1
25	2.2
26	1.6
27	1.3
28	2.1
29	

Hypothesis test for a population mean

- We have case 3, so our test statistic is given by: $$t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$
- Our p-value is given by: $P(T_{n-1} > t_0)$

This is what we would do if we were solving the question by hand. We will use SPSS to solve it in the lab, but you should try this question by hand to make sure you can get the same answer as we get with SPSS.

3 **Case 3:** σ is **not** known and sample size is small $n < 30$

Results from SPSS

- SPSS will only give us the p-value for the two-tailed test (and we’re doing one-tailed!)

One-Sample Test

	Test Value = 4					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Drop	.648	26	.523	.29259	-.6356	1.2208

Results from SPSS

- No need to worry, though! We can compute the p-value for our right-tailed test using this value. All we need to do is divide it by 2.
- 0.523 / 2 = 0.2615

Sig. (2-tailed)
.523

NOTE: It's not always as simple as "dividing by two", if you're trying to go from a two-tailed p-value to a one-tailed p-value. The only reason we can do this is because the sample mean we had agreed with our alternative hypothesis (i.e., our sample mean was indeed greater than 4).

Mean
4.2926

Results from SPSS

- If the p-value > alpha = 0.05, we fail to reject the null hypothesis
    - And conclude that there is not statistically significant evidence that the average cholesterol level drop is greater than 4 units
- If the p-value <= alpha = 0.05, we reject the null hypothesis
    - And conclude that there is statistically significant evidence that the average cholesterol level drop is greater than 4 units
- Our p-value was 0.2615
    - Fail to reject the null hypothesis