8.3 The One-Sample t Test:

If σ is unknown so we have to use s

(if σ is known and normal population then we use Z)

	$n \le 40$	$n > 40$
population approx. normal	t	t/Z
not approx. normal	??	Z

case-by-case basis

In this section, we look at hypothesis tests for a population mean where the population is approximately normal. (We do not need to assume $n > 40$ now.)

Like when we did CIs, $T = \dfrac{\bar{X} - \mu}{S/\sqrt{n}}$ has a t distribution with $n-1$ degrees of freedom.

The t-Test Structure:

1) Hypotheses $H_0: \mu = \mu_0$

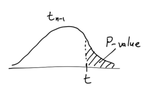
$H_a: \mu > \mu_0$, $H_a: \mu < \mu_0$, or $H_a: \mu \ne \mu_0$

2) Test Statistic If H_0 is true, $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}}$ has a t_{n-1} distribution.

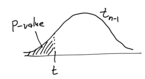
Test Statistic Value is $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$.

3) P-Value

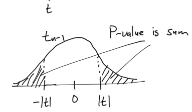
If $H_a: \mu > \mu_0$, then P-value is the area under the t_{n-1} curve to the right of t.



If $H_a: \mu < \mu_0$, then P-value is the area under the t_{n-1} curve to the left of t.



If $H_a: \mu \ne \mu_0$, then P-value is $2 \cdot$(area under t_{n-1} curve to the right of $|t|$).



Can use pt(t, df) in R.

4) Conclusion

If P-value $\le \alpha$, then reject H_0.

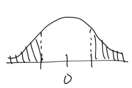
If P-value $> \alpha$, then fail to reject H_0.

Example 1: The true average diameter of ball bearings of a certain type is supposed to be 0.5 in. A random sample of $n=15$ ball bearings is gathered to test if the mean actually differs from 0.5 in. The sample yields $\bar{x} = 0.55$ in and $s = 0.1$ in. Perform the hypothesis test with $\alpha = .05$.

Hypotheses: $H_0: \mu = 0.5$ vs $H_a: \mu \ne 0.5$

Test Statistic: $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}} = \dfrac{0.55 - 0.5}{0.1/\sqrt{15}} \approx 1.936$

P-value:



P-value $= 2 \cdot$ pt$(-1.936,$ df$=14) \approx 0.0698$

Conclusion: P-value is larger than $\alpha = 0.05$, so we fail to reject H_0. We have insufficient evidence to suggest the ball bearings average size is not 0.5 in.

8.4 Tests for a Population Proportion:

In this section, we look at hypothesis tests for a population proportion p.

Large Sample Test:

When $np \geq 10$ and $n(1-p) \geq 10$, the number of successes in the sample, X, is approximately normal by the CLT. So the sample proportion $\hat{p} = \frac{X}{n}$ is also approximately normal.

In 7.2, we showed $\hat{p} \approx N\left(p, \frac{p(1-p)}{n}\right)$

Test Structure:

1) Hypotheses $H_0: p = p_0$

$H_a: p > p_0$, $H_a: p < p_0$, or $H_a: p \neq p_0$

2) Test Statistic If H_0 is true, $Z = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \approx N(0,1)$.

Test Statistic Value is $z = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$

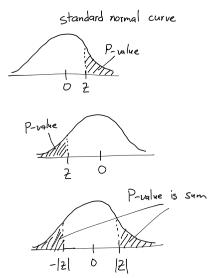
3) P-Value

If $H_a: p > p_0$, then P-value is the area under the standard normal curve to the right of z.

If $H_a: p < p_0$, then P-value is the area under the standard normal curve to the left of z.

If $H_a: p \neq p_0$, then P-value is

$2 \cdot$ (area under standard normal curve to the right of $|z|$.



4) Conclusion

If P-value $\leq \alpha$, then reject H_0.

If P-value $> \alpha$, then fail to reject H_0.

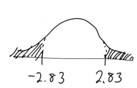
Example 1: We want to test if a coin is fair, meaning when it is flipped, it comes up Heads with probability $p = 0.5$. We flipped the coin 200 times and it came up Heads 80 times. Perform a hypothesis test with $\alpha = 0.05$ to determine if the coin is fair.

Hypotheses: $H_0: p = 0.5$

$H_a: p \neq 0.5$

Test Statistic: $z = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \dfrac{0.4 - 0.5}{\sqrt{\frac{0.5(1-0.5)}{200}}} \approx -2.83$ $\left(\hat{p} = \frac{80}{200} = 0.4\right)$

P-value: P-value $= 2 \cdot \text{pnorm}(-2.83) \approx 0.00465$



Conclusion: P-value is less than $\alpha = 0.05$, so we reject H_0. We have strong evidence that this coin is not fair.

Small Sample Test:

If $np < 10$ or $n(1-p) < 10$, then we cannot use the CLT.

In this case, we can use X, the number of successes in the sample, as our test statistic. $X \sim Bin(n, p)$.

See page 2 of the 8.1 notes for an example of this type of test.

8.5 Further Aspects of Hypothesis Testing:

Statistical Significance vs Practical Significance:

Statistical significance means H_0 was rejected at the chosen significance level α. But it is possible that the departure from H_0 is minor and has little practical significance.

Ex: Let μ is average daily intake of zinc for some cohort.

$H_0: \mu = 15$ mg

Not statistically significant: $\bar{x}$ is close to 15, and we fail to H_0.

Statistically significant but not practically: $\bar{x}$ is close to 15, and we reject H_0

Statistically and practically significant: $\bar{x}$ is far from to 15, and we reject H_0.

The Relationship between Confidence Intervals and Hypothesis Tests:

Consider doing a z-test with $\alpha = .05$ for $H_0: \mu = \mu_0$, $H_a: \mu \neq \mu_0$.

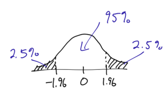
Let's compare with a 95% CI for μ.

CI

$$\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

z-Test

$$Z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$



reject H_0 if $Z \geq 1.96$ or $Z \leq -1.96$

fail to reject H_0 if $-1.96 < Z < 1.96$

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

Let's solve for μ_0 in center.

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

$$-\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \leq -\mu_0 \leq -\bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

$$\bar{x} + 1.96\frac{\sigma}{\sqrt{n}} \geq \mu_0 \geq \bar{x} - 1.96\frac{\sigma}{\sqrt{n}}$$

$$\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu_0 \leq \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}$$

This is when we fail to reject H_0

Prop: Let (θ_L, θ_U) be a $100(1-\alpha)\%$ CI for θ. A hypothesis test of $H_0: \theta = \theta_0$ vs $H_a: \theta \neq \theta_0$ with significance level α will

- reject H_0 if θ_0 is not in the CI (θ_L, θ_U), and
- fail to reject H_0 if θ_0 is in the CI (θ_L, θ_U).

What is the purpose of hypothesis tests then?

Hypothesis tests will give us the p-value.

Simultaneously Testing Multiple Hypotheses:

The probability we make a Type I error increases for each additional test we conduct.

ex: We have list of characteristics. Want to see which characteristic is most predictive of car accidents (brakes, road condition, ...), or pairs of characteristics, etc.

If we want to perform k hypothesis tests, with an overall probability of a Type I error of at most α, then we can use a significance level of $\frac{\alpha}{k}$ for each test. An inequality from probability then tells us

$$P(\text{Type I error}) = P(A_1 \cup A_2 \cup \cdots \cup A_k) \underbrace{\leq}_{\text{union bound}} P(A_1) + P(A_2) + \ldots + P(A_k) = k \cdot \frac{\alpha}{k} = \alpha$$