Testing hypotheses

Testing hypotheses

Our focus in these last weeks of class will be almost exclusively on confidence intervals and hypothesis testing.

Testing hypotheses

Our focus in these last weeks of class will be almost exclusively on confidence intervals and hypothesis testing. Last time we slowly developed the language and methods of basic hypothesis testing.

Testing hypotheses

Our focus in these last weeks of class will be almost exclusively on confidence intervals and hypothesis testing. Last time we slowly developed the language and methods of basic hypothesis testing. We will review these now but be more direct.

General setup: Suppose you are studying some random variable of some population. Suppose someone else makes a guess that the mean

$$\mu = \mu_0$$

for some number μ_0. (And suppose that you *know* the standard deviation, σ for the population. We'll remove this assumption when we study §16).

General setup: Suppose you are studying some random variable of some population. Suppose someone else makes a guess that the mean

$$\mu = \mu_0$$

for some number μ_0. (And suppose that you *know* the standard deviation, σ for the population. We'll remove this assumption when we study §16).

You suspect that μ_0 is incorrect; that is that

$$\mu \neq \mu_0.$$

Your audience will be convinced, if you can show that

$$P(\mu = \mu_0) < \alpha.$$

Your audience will be convinced, if you can show that

$$P(\mu = \mu_0) < \alpha.$$

(α is typically .05 or .1 or .01).

Your audience will be convinced, if you can show that

$$P(\mu = \mu_0) < \alpha.$$

(α is typically .05 or .1 or .01). Our setup is now:

- $H_0 : \mu = \mu_0$. Null hypothesis.
- $H_a : \mu \neq \mu_0$. Alternative hypothesis.
- Significance level α.

Your audience will be convinced, if you can show that

$$P(\mu = \mu_0) < \alpha.$$

(α is typically .05 or .1 or .01). Our setup is now:

- $H_0 : \mu = \mu_0$. Null hypothesis.
- $H_a : \mu \neq \mu_0$. Alternative hypothesis.
- Significance level α.

To try to convince an audience that H_0 is wrong (to *reject the null hypothesis*) we do the following:

a. Take a SRS from population of size n.

To try to convince an audience that H_0 is wrong (to *reject the null hypothesis*) we do the following:

a. Take a SRS from population of size n.

b. Calculate $\overline{x}$ (our sample mean) and the z-statistic

$$z = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$$

c. Calculate the "P-value:" $=$ the probability (assuming H_0 is true) that $\overline{x}$ is at least this far from μ_0

c. Calculate the "P-value:" $=$ the probability (assuming H_0 is true) that $\overline{x}$ is at least this far from μ_0

$$= P(\text{sample mean} \geq \mu_0 + |\overline{x} - \mu_0|) + P(\text{sample mean} \leq \mu_0 - |\overline{x} - \mu_0|)$$

c. Calculate the "P-value:" $=$ the probability (assuming H_0 is true) that $\overline{x}$ is at least this far from μ_0

$$
\begin{aligned}
&= P(\text{sample mean} \geq \mu_0 + |\overline{x} - \mu_0|) + \\
&\qquad P(\text{sample mean} \leq \mu_0 - |\overline{x} - \mu_0|) \\
&\qquad\qquad = P(Z \geq |z|) + P(Z \leq -|z|).
\end{aligned}
$$

d. If $P \leq \alpha$ this is *statistically significant at level* α and leads you to reject the null hypothesis. If $P > \alpha$ this is not statistically significant at that significance level, which means that you can neither reject or accept the null hypothesis.

NOTE: We are concentrating on the situation where H_0 is of form $\mu = \mu_0$. H_a can be serveral things. Either

- $H_a : \mu \neq \mu_0$ (if we don't know which way μ might be off from μ_0). This is called the *two-sided alternative.*

NOTE: We are concentrating on the situation where H_0 is of form $\mu = \mu_0$. H_a can be serveral things. Either

- $H_a : \mu \neq \mu_0$ (if we don't know which way μ might be off from μ_0). This is called the *two-sided alternative.*
- $H_a : \mu > \mu_0$ *(if we suspect μ might really be larger than μ_0). This is called a* one-sided alternative.

NOTE: We are concentrating on the situation where H_0 is of form $\mu = \mu_0$. H_a can be serveral things. Either

- $H_a : \mu \neq \mu_0$ (if we don't know which way μ might be off from μ_0). This is called the *two-sided alternative.*

- $H_a : \mu > \mu_0$ *(if we suspect μ might really be larger than μ_0). This is called a* one-sided alternative.

- $H_a : \mu < \mu_0$ (if we suspect μ might really be smaller than μ_0). This is another one-sided alternative.

Example 1. *Height of men in their 20s in the US is* $N(69.3, 2.8)$.

Example 1. *Height of men in their 20s in the US is* $N(69.3, 2.8)$. *We suspect that the mean height of UO undergrad men is larger than this.*

Example 1. *Height of men in their 20s in the US is* $N(69.3, 2.8)$*. We suspect that the mean height of UO undergrad men is larger than this.*

- *Hypotheses:* $H_0 : \mu = 69.3$. $H_a : \mu > 69.3$.

Example 1. *Height of men in their 20s in the US is* $N(69.3, 2.8)$*. We suspect that the mean height of UO undergrad men is larger than this.*

- *Hypotheses:* $H_0 : \mu = 69.3$. $H_a : \mu > 69.3$.
- *Calculate* z*-statistic: We take a SRS of 32 undergrad men from the UO. We discover that* $\overline{x} = 70.4$ *and then that*

$$z = \frac{70.4 - 69.3}{2.8/\sqrt{32}} = 2.22.$$

Example 1. *Height of men in their 20s in the US is* $N(69.3, 2.8)$. *We suspect that the mean height of UO undergrad men is larger than this.*

- *Hypotheses:* $H_0 : \mu = 69.3$. $H_a : \mu > 69.3$.
- *Calculate* z*-statistic: We take a SRS of 32 undergrad men from the UO. We discover that* $\overline{x} = 70.4$ *and then that*
$$z = \frac{70.4 - 69.3}{2.8/\sqrt{32}} = 2.22.$$
- *Calculate* P*-value:* $P(Z \geq 2.22) = .0132$ *from table A.*

- *Conclusion: If* H_0 *is* true, *the probability that* $\overline{x} \geq 70.4$ *is only .0132. This is evidence against* H_0

- *Conclusion: If* H_0 *is* true, *the probability that* $\overline{x} \geq 70.4$ *is only .0132. This is evidence against* H_0 *This* P*-value gets smaller if we assume that* μ *is even smaller than* 69.3*, so in fact it gives evidence for* H_a*.*

- *Conclusion: If* H_0 *is* true, *the probability that* $\overline{x} \geq 70.4$ *is only .0132. This is evidence against* H_0 *This* P*-value gets smaller if we assume that* μ *is even smaller than* 69.3*, so in fact it gives evidence for* H_a*.*

  At significance level $\alpha = .05$ *this is statistically significant so we reject the null hypotheses and believe* H_a *is probably true.*

- *Conclusion: If* H_0 *is* true, *the probability that* $\overline{x} \geq 70.4$ *is only .0132. This is evidence against* H_0 *This* P*-value gets smaller if we assume that* μ *is even smaller than* 69.3*, so in fact it gives evidence for* H_a*.*

  At significance level $\alpha = .05$ *this is statistically significant so we reject the null hypotheses and believe* H_a *is probably true.*

  At significance level $\alpha = .01$ *this is not statistically significant so we don't reject the null hypotheses.*

(Although it may still be false.)

Example 2. *SAT scores are roughly normally distributed with a mean of 1682 and a deviation of 220.*

Example 2. *SAT scores are roughly normally distributed with a mean of 1682 and a deviation of 220. You want to know if TestQuick test preparation helps improve scores. A random sample of 625 students who have used TestQuick are chosen, and they score an average of 1695.*

Example 2. *SAT scores are roughly normally distributed with a mean of 1682 and a deviation of 220. You want to know if TestQuick test preparation helps improve scores. A random sample of 625 students who have used TestQuick are chosen, and they score an average of 1695. Is this convincing evidence that TestQuick improves scores? If so, at what level is it convincing?*

Example 3. *Yearly household incomes in the U.S have a mean of 38 thousand and a standard deviation which is less than forty thousand dollars.*

Example 3. *Yearly household incomes in the U.S have a mean of 38 thousand and a standard deviation which is less than forty thousand dollars. You do a random survey of 1600 housholds in Oregon and you find an average income of 35 thousand. Is this convincing evidence that household incomes in Oregon are on average less than the national average?*

Example 3. *Yearly household incomes in the U.S have a mean of 38 thousand and a standard deviation which is less than forty thousand dollars. You do a random survey of 1600 housholds in Oregon and you find an average income of 35 thousand. Is this convincing evidence that household incomes in Oregon are on average less than the national average?*

Relating hypothesis testing to confidence intervals

Relating hypothesis testing to confidence intervals

The basic calculations for hypothesis testing and confidence intervals are the same:

Relating hypothesis testing to confidence intervals

The basic calculations for hypothesis testing and confidence intervals are the same: take a simple random sample, compute its mean, and then calculate its z-statistics using the Central Limit Theorem.

Relating hypothesis testing to confidence intervals

The basic calculations for hypothesis testing and confidence intervals are the same: take a simple random sample, compute its mean, and then calculate its z-statistics using the Central Limit Theorem.

In fact, these two procedures are logically related

Relating hypothesis testing to confidence intervals

The basic calculations for hypothesis testing and confidence intervals are the same: take a simple random sample, compute its mean, and then calculate its z-statistics using the Central Limit Theorem.

In fact, these two procedures are logically related (seeing how helps us better understand both concepts).

Fact 4. *If some guess for a mean* μ_0 *is not within a* $1-\alpha$ *confidence interval about an observed mean* $\bar{x}$ *then we may reject the null hypothesis* $\mu = \mu_0$ *at significance level* α.

Fact 4. *If some guess for a mean μ_0 is not within a $1-\alpha$ confidence interval about an observed mean $\bar{x}$ then we may reject the null hypothesis $\mu = \mu_0$ at significance level α.*

It helps to sketch a picture of what this is saying and to choose some concrete numbers.

Fact 4. *If some guess for a mean* μ_0 *is not within a* $1-\alpha$ *confidence interval about an observed mean* $\bar{x}$ *then we may reject the null hypothesis* $\mu = \mu_0$ *at significance level* α.

It helps to sketch a picture of what this is saying and to choose some concrete numbers.

Example 5. *Suppose a variable has a standard deviation of* 5 *and a measured mean of* 172.1 *from a sample size of* 80.

Fact 4. *If some guess for a mean* μ_0 *is not within a* $1-\alpha$ *confidence interval about an observed mean* $\bar{x}$ *then we may reject the null hypothesis* $\mu = \mu_0$ *at significance level* α.

It helps to sketch a picture of what this is saying and to choose some concrete numbers.

Example 5. *Suppose a variable has a standard deviation of* 5 *and a measured mean of* 172.1 *from a sample size of* 80.

- *Find a confidence interval with* $C = 95\%$ *for this variable.*

- *Find a confidence interval with* $C = 95\%$ *for this variable.*

- *Show that the null-hypothesis of a mean equal to* 175 *can be rejected at level* 5%.

Caution: garbage in, garbage out

Caution: garbage in, garbage out

The most common mistakes one sees in applications of statistics are not in the mathematics of means and deviations

Caution: garbage in, garbage out

The most common mistakes one sees in applications of statistics are not in the mathematics of means and deviations – with practice one can master these methods, and there are sophisticated computer packages to help out

Caution: garbage in, garbage out

The most common mistakes one sees in applications of statistics are not in the mathematics of means and deviations – with practice one can master these methods, and there are sophisticated computer packages to help out – but in starting with biased data, whose error is comparable to the confidence intervals in question.

Caution: garbage in, garbage out

The most common mistakes one sees in applications of statistics are not in the mathematics of means and deviations – with practice one can master these methods, and there are sophisticated computer packages to help out – but in starting with biased data, whose error is comparable to the confidence intervals in question.

Example 6. *What if in our UO height example, two of the sample taken were members of the basketball team over 80 inches tall?*

Caution: garbage in, garbage out

The most common mistakes one sees in applications of statistics are not in the mathematics of means and deviations – with practice one can master these methods, and there are sophisticated computer packages to help out – but in starting with biased data, whose error is comparable to the confidence intervals in question.

Example 6. *What if in our UO height example, two of the sample taken were members of the basketball team over 80 inches tall? If these outliers are thrown out, what*

is our new confidence level for the hypothesis that UO men are above average height?

is our new confidence level for the hypothesis that UO men are above average height?

One of the topics of Chapter 15 is to elaborate on problems which can plague samples and surveys, in light of their effects on confidence intervals and hypothesis tests.

is our new confidence level for the hypothesis that UO men are above average height?

One of the topics of Chapter 15 is to elaborate on problems which can plague samples and surveys, in light of their effects on confidence intervals and hypothesis tests. This material mostly applies common sense, not mathematics, as we did when we looked at designing surveys and experiemens.

is our new confidence level for the hypothesis that UO men are above average height?

One of the topics of Chapter 15 is to elaborate on problems which can plague samples and surveys, in light of their effects on confidence intervals and hypothesis tests. This material mostly applies common sense, not mathematics, as we did when we looked at designing surveys and experiemens.

We will not at all cover the material from Chapter 15 on type I and II errors and power of tests.

is our new confidence level for the hypothesis that UO men are above average height?

One of the topics of Chapter 15 is to elaborate on problems which can plague samples and surveys, in light of their effects on confidence intervals and hypothesis tests. This material mostly applies common sense, not mathematics, as we did when we looked at designing surveys and experiemens.

We will not at all cover the material from Chapter 15 on type I and II errors and power of tests. This development of language is straightforward enough so that you could learn it on your own if you encountered these terms outside of class.