Lecture 27/Chapters 22 & 23

Hypothesis Tests for Means

- Four Steps
- Null and Alternative Hypotheses
- Standardizing Sample Mean
- *P*-value, Conclusions
- Examples

Probability then Inference, Proportions then Means

Probability theory dictates behavior of sample proportions (categorical variable of interest) and sample means (quantitative variable) in random samples from a population with known values. Now

perform **inference** with confidence intervals

- for proportions (Chapter 20)
- for means (Chapter 21)

or with hypothesis testing

- for proportions (Chapters 22&23)
- for means (Chapters 22&23)

Two Forms of Inference

Confidence interval: Set up a range of plausible values for the unknown population proportion (if variable of interest is categorical) or mean (if variable of interest is quantitative).

Hypothesis test: Decide if a particular proposed value is plausible for the unknown population proportion (if variable of interest is categorical) or mean (if variable of interest is quantitative).

Last time, we tested about an unknown population proportion when the variable of interest was categorical (for or against gay civil unions, M&M color). Now we test about a mean when the variable of interest is quantitative (wt, IQ, cost).

Testing Hypotheses About Pop. Value

1. Formulate hypotheses.
2. Summarize/standardize data.
3. Determine the *P*-value.
4. Make a decision about the unknown population value (proportion or mean).

Null and Alternative Hypotheses

For a test about a single mean,

- **Null hypothesis:** claim that the population mean equals a proposed value.
- **Alternative hypothesis:** claim that the population mean is greater, less, or not equal to a proposed value.

  An alternative formulated with ≠ is **two-sided**; with > or < is **one-sided**.

Standardizing Normal Values *(Review)*

Put a value of a normal distribution into perspective by standardizing to its *z*-score:

$$z = \frac{\text{observed value - mean}}{\text{standard deviation}}$$

The observed value that we need to standardize in this context is the **sample mean**. We've established Rules for its mean and standard deviation, and for when the shape is approximately normal, so that a probability (the *P*-value) can be assessed with the normal table.

Conditions for Rule of Sample Means

- Randomness [affects center]
- Independence [affects spread]
  - If sampling without replacement, sample should be less than 1/10 population size
- Large enough sample size [affects shape]
  - If population shape is normal, any sample size is OK
  - If population if not normal, a larger sample is needed.

If 1st two conditions don't hold, the mean and sd in *z* are wrong; if 3rd doesn't hold, *P*-value is wrong.

Rule for Sample Means (if conditions hold)

- **Center:** The mean of sample means equals the true population mean.
- **Spread:** The standard deviation of sample means is standard error =

  $$\frac{\text{population standard deviation}}{\sqrt{\text{sample size}}}$$

  Substitute sample standard deviation if population standard deviation is unknown.
- **Shape:** (Central Limit Theorem) The frequency curve will be approximately normal, depending on how well 3rd condition is met.

Standardized Sample Mean

- To test a hypothesis about an unknown population mean, find sample mean (and standard deviation) and standardize to

$$z = \frac{\text{sample mean - population mean}}{\frac{\text{standard deviation}}{\sqrt{\text{sample size}}}}$$

- z is called the **test statistic**.

Note that "sample mean" is what we've observed, "population mean" is the value proposed in the null hypothesis, and "standard deviation" is from population (preferred) or sample (OK if sample size ≥ 30).

P-value in Hypothesis Test about Mean

The *P*-value is the probability, assuming the null hypothesis is true, of a sample mean at least as low/high/different as the one we observed. In particular, it depends on whether the alternative hypothesis is formulated with a less than, greater than, or not-equal sign.

Making a Decision Based on a *P*-value

If the *P*-value in our hypothesis test is small, our sample mean is improbably low/high/different, assuming the null hypothesis to be true. We conclude it is **not** true: we reject the null hypothesis and believe the alternative.

If the *P*-value is not small, our sample mean is believable, assuming the null hypothesis to be true. We are willing to believe the null hypothesis.

P-value small ⟶ reject null hypothesis

P-value not small ⟶ don't reject null hypothesis

Hypothesis Test for Means: Details

1. null hypothesis: pop mean = proposed value
   alt hyp: pop mean < or > or ≠ proposed value
2. Find sample mean (and sd) and standardize to z.
3. Find the *P*-value= probability of sample mean as low/high/different as the one observed; same as probability of z this far below/above/away from 0.
4. If the *P*-value is small, conclude alternative is true. In this case, we say the data are **statistically significant** (too extreme to attribute to chance). Otherwise, continue to believe the null hypothesis.

Example: *Testing a Hypothesis about a Mean*

- **Background**: Wts (in g) in a large colony of lab mice have mean 30, sd 5. Grad students pick 25 "at random" and find mean wt is 32.6.
- **Question:** Was their sample actually biased?
- **Response:**

1. Null: __________________ Alt: ____________________
2. Sample mean=____, sd=___, z =
3. *P*-value=prob of z this far away from 0: ____________
4. Because the *P*-value is ___________________________
   Conclude ___________________________

Example: *Hypothesis Test about Smoking & IQ*

- **Background**: IQs of children of a sample of 36 women who smoked while pregnant had mean 91.
- **Question:** Could this have been chance (null) or is it significantly lower than pop mean IQ 100 (with sd 16)?
- **Response:**

1. Null:__________________Alt:____________________
2. Sample mean=___, pop sd=___, z =
3. *P*-value=prob of z this far below 0: ____________
4. *P*-value is small, so reject null hypothesis. Conclude
   ____

Choosing the Right Display *(Review)*

Display type depends on variable types:

- 1 measurement variable (students' heights): stemplot, boxplot, histogram, freq. curve (chs 7&8)
- 1 categorical + 1 measurement var. (sex + ht): multiple boxplots (ch 7, see p. 136)
- 2 measurement variables:
  - Time is expl (yr + cremation): time series (ch 15)
  - in general (age + wt): scatterplot (ch 10)
- 1 categorical var: (radio show type): piechart (ch 9)
- 2 or more cat vars (sex,smoke,on/off):barchart (ch 9) (for 2 cat vars, use two-way table to organize data)

Choosing the Right Test

Type of test depends on variable types:

- 1 categorical: z test about population proportion (done)
- 1 measurement (quan) [pop sd known or sample large]: z test about mean (done)
- 1 measurement (quan) [pop sd unknown & sample small]: t test about mean (to do)
- 1 categorical (2 groups)+ 1 quan: two-sample z or t (to do)
- 2 categorical variables: chi-square test (done in Chapter 13)
- 2 quan variables: regression test (not done in this course)

Note: The t curve, like z, is bell-shaped and symmetric about 0. Because t has a bit more spread than z, our reaction to a t statistic is similar to what it would be for a z statistic but it takes a larger value of t to impress us, especially if the sample is small.

Example: *t Test*

- **Background**: Cost (in $1000s) of coronary bypass surgery at a sample of 9 hospitals had mean 24, sample sd 8.
- **Question:** Are we convinced that the overall mean is >20?
- **Response:**

1. Null: ____________________ Alt: ______________________
2. Sample mean=___, sample sd=___, *t* =
3. *P*-value = prob of *t* this far above 0 = ?
   Note: the *t* curve is similar to *z* but more spread out: *t* values must be more extreme to achieve significance.
4. Since +1.5 is not large for *z*, ___________________________.
   P-value is ____
   ____________________ the population mean cost is more than 20 thousand dollars.

Two-Sample *z* or *t* Test

1. Null: mean for 1st population=mean for 2nd population
2. Two-sample $t = \dfrac{\text{1st sample mean} - \text{2nd sample mean}}{\sqrt{\dfrac{(\text{1st sd})^2}{\text{1st sample size}} + \dfrac{(\text{2nd sd})^2}{\text{2nd sample size}}}}$
3. Obtain *P*-value based on *z* or *t* distribution
   (*z* for large samples, *t* for small samples).
4. Reject null hypothesis if *P*-value is small.

Example: *Two-Sample Test*

- **Background**: Wait times (in seconds) at 7 banks on the west coast had mean 231.6, sd 27.8, while 18 banks on the east coast had mean 272.7, sd 72.5. The two-sample *t* statistic was 2.05 and the *P*-value for a two-sided test was 0.052.
- **Question:** Do mean wait times differ in general, east vs. west?
- **Response:** _______________________ Note that if *z*=2.05 (instead of *t*), the *P*-value for a two-sided *z* test would be __________________, and the results would be somewhat more convincing.

Chi-Square Test

We learned to use chi-square to test for a relationship between two categorical variables.

1. Null hypothesis: the two variables are not related
   Alternative hypothesis: the two variables are related
2. Test statistic=chi-square=$\dfrac{(\text{observed count} - \text{expected count})^2}{\text{expected count}}$
3. *P*-value= probability of chi-square this large, assuming the two variables are not related. For a 2-by-2 table, chi-square > 3.84 ⟷ *P*-value < 0.05.
4. If the *P*-value is small, conclude the variables are related. Otherwise, we have no convincing evidence of a relationship.

In general, a large test statistic is accompanied by a small *P*-value.

Note: Next lecture we'll do another example of a chi-square test.