Monday PM

- Presentation of AM results
- Hypothesis testing review
- Testing means
  - One sample t test
  - Two sample t test
  - One way ANOVA
  - Paired sample t- test

Hypothesis testing: a review

- In hypothesis testing statistics, we set up a null hypothesis about the data, and then proceed to try to reject this hypothesis.
- The null hypothesis usually represents “no effect”, “no difference”, or “no relationship”, though it may represent other possibilities as well.

Errors in hypothesis testing

	H_o is false (effect)	H_0 is true (no effect)
Reject H_0	Sig. effect (TP) 1-β (power)	Type I error (FP) α
Fail to reject H_0	Type II error (FN) β	No effect (TN) 1-α (confidence)

One- and two-tailed tests

- Tests can be one-tailed or two-tailed
  - A two tailed test looks for any difference from the null hypothesis, no matter what direction.
  - A one tailed test looks for a specified *directional* difference from the null hypothesis, and does not test for differences in the other direction.
- One-tailed tests are more powerful for a given α, but two-tailed tests can find effects in either direction.

Degrees of freedom

- Inferential (sample) statistics essentially involve fitting a model of the null hypothesis to the data, and finding that the model is a poor fit.
- The more data you have and the simpler the model, the less constraint there is upon how the data can be distributed, and the more ways the data might not be fitted.
- Formally, the number of unconstrained data points that the model is free to fit or not are the statistic's *degrees of freedom*, or *df*.

Degrees of freedom, example

- A line is a model defined by two parameters: a slope and an intercept.
- If I give you two points, you can always fit a perfect line. There are no degrees of freedom left to determine if the line fits well or not.
- If I give you three points, you need two of them to fit a line, and one is left to test whether a line is a good fit to the data. You have 1 df. This is a weak test.
- If I give you 100 points, you need 2 to fit the line, and you have 98 df. This is a powerful test.

One-sample t-test

- Goal: Given a sample, test to see if it comes from a population with a given mean value of a variable
- Example: Is the mean GPA of medical students different from 3.0?
- H_0: $\mu = k$
- H_1 $\mu \neq k$ (2-tailed) or $\mu > k$ (1-tailed)

One-sample t-test in SPSS

- Analyze…Compare Means…One-sample t-test
- Enter variable and test value

```
                                           N  Mean    SD      SE mean
Highest Year of School Completed       1510    12.88   2.98
         7.68E-02

One-Sample Test
Test Value = 8
         t      df      Sig.   Mean Diff   95% CI of diff

                                           Lower      Upper
HYSC    63.602  1509    .000     4.88        4.73      5.03
```

- On average, Americans sampled had more than 8 years of education ($t(1509) = 63.6$, $p < .05$).

Two-sample t-test

- Goal: Given 2 samples, test to see if they come from populations with different mean values of a variable.
- Example: Is the mean GPA of male medical students greater than that of female?
- H_0: $\mu_m = \mu_f$
- H_1 $\mu_m \neq \mu_f$ (2-tailed) or $\mu_m < \mu_f$ (1-tailed)

Two-sample t-test in SPSS

- Analyze…Compare means…Independent samples t-test
- Enter test (dependent) variable, and grouping (independent) variable, and define the two groups by their value on the grouping variable.

Group Statistics

	Sex	N	Mean	SD	SE Mean
HYSC	Male	633	13.23	3.14	.12
	Female	877	12.63	2.84	9.59E-02

Independent Samples Test

Levene's Test for Equality of Variances: F=11.226, p < .001

t-test, equal variances not assumed:

t	df	Sig.	Mean Diff	SE Diff	95% CI Diff
3.824	1276.454	.000	.60	.16	[.29,.91]

Two-sample t-test, reporting

- Men had a significant higher mean number of years of education than women (unequal-variance t(1276)=3.83, p<.05).
- On average, men had 0.6 more years of school than women (95% CI: [.29,.91]).

One-way ANOVA (analysis of variance)

- Goal: Given many samples, test to see if they come from populations with different mean values of a variable.
- Example: Do the mean GPAs of medical students from Sâo Paulo, Marilia, and Botucatu differ?
- H_0: $\mu_{sp} = \mu_m = \mu_b$
- H_1: at least one mean differs

One-way ANOVA in SPSS

- Analyze…Compare means…One-way ANOVA
- Enter test and group vars, optionally set contrasts, post-hoc tests, and options.

Descriptives - Highest Year of School Completed

	N	Mean	SD	Std. Error
White	1262	13.06	2.95	8.32E-02
Black	199	11.89	2.68	.19
Other	49	12.47	4.00	.57

ANOVA - Highest Year of School Completed

	Sum Squares	df	Mean Square	F	Sig.
Betw Grps	240.725	2	120.362	13.746	.000
W/in Grps	13195.994	1507	8.756		
Total	13436.719	1509			

One-way ANOVA, reporting

- An ANOVA was conducted to examine the effect of race (white, black, other) on highest year of education completed.
- There was a significant difference in education between races ($F(2,1507) = 13.7$, $p < .05$).

Contrasts in a one-way ANOVA

- If you reject the overall null hypothesis, you still haven't show which particular means are different from each other.
- You can analyze specific contrasts or comparisons of means to do this.
- You can either do *t* tests on pairs, or (better) program the contrasts into the ANOVA analysis itself. This is more powerful, because it will use the ANOVA's error estimate, which is based on the full sample and usually more accurate

Contrasts in SPSS

- To specify a contrast in a one way ANOVA, use the Contrasts button.
- Contrasts are specified as weights for each group in the grouping variable. For example, if the groups are white, black, other, some contrasts are:

1,-1,0	Compare white and black
-1,1,0	Compare black and white
1,0,-1	Compare white and other
1,-0.5, -0.5	Compare white and mean nonwhite

Contrast output and reporting

Contrast Coefficients

Race of Respondent

White	Black	Other
1	-1	0

Contrast Tests - Highest Year of School Completed

	Value of Contrast	SE	t	df	Sig. (2-tailed)
Equal var	1.16	.23	5.147	1507	.000
Unequal var	1.16	.21	5.607	279.767	.000

- A planned contrast between white and black respondents found that white respondents had significantly more schooling than black ($t(1507)=5.15$, $p<.05$).

Post-hoc tests

- Post-hoc tests are *unplanned* comparisons, performed after looking at the data pattern.
- As such, they capitalize on chance in the data, and should not be accorded as much weight as planned comparisons.
- Moreover, if you perform a large number of statistical tests (common in post-hoc analyses), you must consider the *familywise* (Type I) error rate, which is much larger than the per-test rate.

Example: Bonferroni test

Bonferroni: Highest Year of School Completed

(I) Race	(J) Race	Mean Diff	SE	Sig.
White	Black	1.16*	.23	.000
	Other	.59	.43	.520
Black	White	-1.16*	.23	.000
	Other	-.57	.47	.670
Other	White	-.59	.43	.520
	Black	.57	.47	.670

* The mean difference is significant at the .05 level.

- Corrects for familywise error by setting each test's α to 0.05/(number of tests)

Paired-sample t-test

- Goal: Given 2 measurements of a variable from the same sample, test to see if their means differ between measurements.
- Example: For graduating medical students, are mean GPAs higher at the end of year 2 or the end of year 4?
- H_0: $\mu_{GPA,2} = \mu_{GPA,4}$
- H_1: $\mu_{GPA,2} \neq \mu_{GPA,4}$ (2-tailed)
  or $\mu_{GPA,2} < \mu_{GPA,4}$ (1-tailed)

Paired-sample t-test in SPSS

- Analyze…Compare means…Paired-sample t-test
- Enter pairs of variables

Paired Samples Statistics: LE in 109 countries

	Mean	N	SD	SE Mean
Female LE	70.16	109	10.57	1.01
Male LE 64.92	109	9.27	.89	

Paired Samples Test

Paired Differences							
Mean	SD	SE	95% CI		t	df	sig (2-tail)
5.24	2.27	.22	4.81	5.67	24.11	108	.000

- Average life expectancy for women is higher than for men in the same country (t(108)=24.11, p<.05)

Assumptions of t-tests

- t tests and ANOVA are *parametric* tests: they make assumptions about distribution of scores in the populations from which the means are taken:
  - Distributions are assumed to be normal
  - If two or more population means are being compared, populations are assumed to have equal variances
- These are fairly strong assumptions, but the tests are often ok even if they're violated moderately.
- We'll see *nonparametric* tests later that don't make these assumptions

Monday PM assignment

- Using the hyp data set, test these hypotheses:
  1. The mean spatial perception score is 50
  2. The mean midterm score is different for case-based and lecture formats
  3. The mean final score is higher for case-based than lecture formats
  4. Mean final scores are higher than mean midterm scores
  5. Create a new variable with 3 categories: new (<5 years post-MD), medium (5-15 years post-MD) and old (>15 years post-MD). Do mean satisfaction scores differ by this category?