Final exam

- 2.5 hours allotted
- Chapters 1-17, 19, 20
- Excluding section 9.3, 9.6, 14.7, 15.3, 15.6, confidence interval for $r$
- Bring: Calculator (not programmable), pen or pencil, UBC ID
- You will be given: formula sheet & stats tables

One variable: Which test?

<table>
<thead>
<tr>
<th>Data type</th>
<th>Goal</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical</td>
<td>Comparing a proportion to a hypothesized value</td>
<td>Binomial test (7)</td>
</tr>
<tr>
<td></td>
<td>Comparing frequency data to a probability distribution</td>
<td>$\chi^2$ Goodness-of-fit test (8)</td>
</tr>
<tr>
<td>Numerical</td>
<td>Comparing mean to a hypothesized value when data are approximately normal (possibly only after a transformation)</td>
<td>One-sample $t$-test (11)</td>
</tr>
<tr>
<td></td>
<td>Comparing median to a hypothesized value when data are not normal (even after transformation)</td>
<td>Sign test (13)</td>
</tr>
<tr>
<td></td>
<td>Comparing frequency data to a discrete probability distribution</td>
<td>$\chi^2$ Goodness-of-fit test (8)</td>
</tr>
<tr>
<td></td>
<td>Comparing data to the normal distribution</td>
<td>Shapiro–Wilk test (13)</td>
</tr>
</tbody>
</table>

Two variables: Which test?

<table>
<thead>
<tr>
<th>Type of explanatory variable</th>
<th>Type of response variable</th>
<th>Categorical</th>
<th>Numerical</th>
</tr>
</thead>
<tbody>
<tr>
<td>Categorical</td>
<td>Contingency analysis (9)</td>
<td>Logistic regression (17)</td>
<td></td>
</tr>
<tr>
<td>Numerical</td>
<td>$t$-tests, ANOVA, Mann–Whitney $U$-test, etc. [See Table 3 for more details.]</td>
<td>Linear and nonlinear regression (17)</td>
<td></td>
</tr>
</tbody>
</table>

Tests to know how to do...

- Binomial test
  - Directly
  - Using normal approximation
- $\chi^2$ Goodness-of-fit test
- $\chi^2$ Contingency analysis
- One-sample $t$-test
- Paired $t$-test
- Two-sample $t$-test
- F test
- Sign test
- Mann–Whitney U test
- Single factor ANOVA
- Correlation
- Linear regression
- Likelihood ratio test
Tests to know about...

- Fisher's exact test
- Welch's t-test
- Shapiro-Wilk test
- Mann-Whitney U test
- Tukey-Kramer test
- Kruskal-Wallis test
- Multifactor ANOVA
- Nested ANOVA
- Spearman's correlation

- ANCOVA
- Polynomial regression
- F-tests
- Logistic regression
- Randomization
- Simulation
- Bootstrapping

Confidence intervals to know how to do...

- Proportion
- Mean of a normally distributed variable
- Variance of a normally distributed variable
- Regression slope

Know tests to...

- Compare proportion to constant
- Compare proportion from two groups
- Test independence of two categorical variables
- Compare frequency data to a model

- Compare mean to a constant
- Compare means of two groups
- Compare mean difference of two groups (paired data)
- Compare means of more than two groups
- All of the above, without assuming normal distribution

Know tests to...

- Test for independence of two numerical variables - with assumptions
- Test for independence of two numerical variables - with fewer assumptions
- Predict one numerical variable from another

- Compare two slopes
- Test the interaction of two categorical factors effects on a numerical variable factor
- Compare to a normal distribution
- Compare the variances of two groups
- Predict a binary variable from a numerical variable
Know tests to...

- Deal with new situations, not dealt with by traditional tests

Practice - which test?

- 1. In the World Series, the teams have won at home 286 times and lost at home 243 times. Is there a "home-field advantage"?

- 2. Are the mean numbers of parasites per host different in multiply-infected vs. singly-infected hosts? (Answer if number of parasites per host is normally distributed, and if it is not.)

Practice - which test?

- 3. Are longevity and fecundity related?

- 4. The number of eggs per worm seems to decline in a curved relationship with worm density. How can you test this relationship?

- 5. Longevity (number of days until death) in response to cold temperature may be a function of temperature or of genotype. How can you test whether these two effects are independent?

- 6. The developmental pathway leading to the formation of spots on butterfly wings has been studied by surgical excision of a small amount of tissue on the left wings of a set of butterflies, with the right wings left untouched. The size of the spots on these wings was subsequently measured. How would you test whether the manipulation had an effect on spot size?

Practice - which test?

- 7. Does human height follow a normal distribution?

- 8. Is the variance of male weight equal to the variance in female weight (if weight is not perfectly normal)?
9. Five different kinds of dog food were each tested on 25 young dogs, to determine which made them grow most quickly (as measured by g/day/kg). Which test(s) might you use?

10. Five different kinds of dog food were each tested on 25 dogs, to determine whether they had an effect on whether or not the dogs developed diabetes. Which test might you use?

11. 25 dogs were grown in each of five different temperatures. Is there an effect on growth rate?

12. 25 dogs were each grown in five different temperatures. Data were collected on whether or not the dogs survived to an age of 5 years. Is there an effect of temperature on mortality?

13. An analysis of variance has successfully shown an effect of diet on the amount of bacon a pig produces. How do you determine whether a particular diet is significantly better than the others?

14. You discover that a data set you have addresses a very important question, but the analysis required to address that question does not fit any known existing statistical test. What are your options?