Stat 111 Spring 2025 Week 8: k-sample Normal

1. **The unpooled 2-Sample t Test**
Suppose $Y_{11}, \ldots, Y_{1n_1} \stackrel{\text{iid}}{\sim} N(\mu_1, \sigma_1^2)$ are independent of $Y_{21}, \ldots, Y_{2n_2} \stackrel{\text{iid}}{\sim} N(\mu_2, \sigma_2^2)$. Let $\bar{Y}_1$ and $\bar{Y}_2$ be the two averages and s_1^2 and s_2^2 the two sample variances. Consider a test of $H_o : \mu_1 = \mu_2$ vs. $H_a : \mu_1 \neq \mu_2$.

   a) The GLR test rejects for large values of T^2, where
$$T = \frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$
Set up the GLR test and describe what it would take to show this (the algebra gets a bit dense, so don't go through it all). Explain why T does not follow a t distribution, even under H_o.

   b) Use the context of the matched pairs t test to explain why $m_o = \min(n_1 - 1, n_2 - 1)$ is a conservative degree of freedom value in that, if t^* is the $1-\alpha/2$ quantile of the $t_{m_o)}$ distribution, then the test that rejects H_o for $|T| > t^*$ has significance level at most α, and the interval $\bar{Y} \pm t^* \frac{s}{\sqrt{n}}$ contains μ for at least 95% of samples of size n.

   c) There is a degrees of freedom value m such that the $t_{(m)}$ distribution is a close approximation to the null distribution of T. Most statistical software implements Welch's method, which uses degrees of freedom $\hat{m}$:
$$\hat{m} = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{(s_1^2/n_1)^2/(n_1 - 1) + (s_2^2/n_2)^2/(n_2 - 1)}.$$
Derive Welch's formula for the "estimated degrees of freedom" by matching the unpooled 2-sample t statistic to the definition of a $t_{(m)}$ random variable. Define $Z \sim N(0, 1)$ to be the standardized value of $\bar{Y}_1 - \bar{Y}_2$ (using the unknown values σ_1 and σ_2) and define U to be the variable playing the role of the independent $\chi^2_{(m)}$ random variable. Now find m to match the mean and variance of U to that of a $\chi^2_{(m)}$ random variable, approximating σ_1 and σ_2 by s_1 and s_2. Show that m is always between the conservative df value m_o =min$(n_1 - 1, n_2 - 1)$ and the pooled df value $n_1 + n_2 - 2$.

   d) Use data on NSF grants as an example, testing for differences in grants for male and female principal investigators. In 1998, a Swarthmore student took a random sample of NSF grants funded that year and recorded the sex of the principal investigator (PI) and the dollar value of the grant (grantsSub.txt). Often taking logarithms of dollar values makes the data more appropriate for a t test. Explain how the test for the dollar values compares the mean value while the test for the log dollars essentially compares the median values.

2. **The pooled 2-Sample t Test** Suppose $Y_{11}, \ldots, Y_{1n_1} \stackrel{\text{iid}}{\sim} N(\mu_1, \sigma_1^2)$ are independent of $Y_{21}, \ldots, Y_{2n_2} \stackrel{\text{iid}}{\sim} N(\mu_2, \sigma_2^2)$. Let $\bar{Y}_1$ and $\bar{Y}_2$ be the two averages and s_1^2 and s_2^2 the two sample variances. Consider a test of $H_o : \mu_1 = \mu_2$ vs. $H_a : \mu_1 \neq \mu_2$ and assume $\sigma_1 = \sigma_2 = \sigma$.

   a) The MLE for σ^2 is $\hat{\sigma}^2 = \frac{1}{n_1+n_2}(\sum(Y_{1i} - \bar{Y}_1)^2 + \sum(Y_{2i} - \bar{Y}_2)^2)$. Find the bias of this estimate as a function of n_1 and n_2. The *restricted* maximum likelihood (REML) estimate is calculated by first integrating the joint likelihood function with respect to the two mean parameters μ_1 and μ_2, and then maximizing the resulting function over σ^2. Show that this leads to the unbiased pooled sample variance $s_p^2 = \frac{\sum(Y_{1i} - \bar{Y}_1)^2 + \sum(Y_{2i} - \bar{Y}_2)^2}{n_1+n_2-2} = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$. Note that the only justification for this procedure from a frequentist perspective is that it leads to an improved estimate of σ that reflects the degrees of freedom lost to estimating the two mean parameters (it makes much more sense as an objective Bayes procedure).

b) Show, using results we have already proved, that s_p^2 is a Gamma random variable, independent of $\bar{Y}_1$ and $\bar{Y}_2$. Show that the pivot $W = \frac{(n_1+n_2-2)s_p^2}{\sigma^2} \sim \chi^2_{(n_1+n_2-2)}$, when conditioning on μ_1, μ_2 and σ^2.

c) Show that the pooled two sample t statistic T satisfies the definition of a $t_{(n_1+n_2-2)}$ random variable for any hypothesized value of $\mu_1 - \mu_2$ (e.g. $\mu_1 - \mu_2 = 0$).

$$T = \frac{\bar{Y}_1 - \bar{Y}_2 - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

What values of T would lead you to reject H_o at level α? What values of T^2 would lead you to reject? What is the null sampling distribution of T^2? What is a CI for $\mu_1 - \mu_2$? Make the distinction between the pooled standard deviation estimate (root mean square error) and the standard error.

d) As an example, imagine dividing $N = 200$ subjects into two equal-sized treatment groups and administering a treatment to one group and a placebo to the other. What values of T would lead you to reject the null hypothesis? What values would lead you to conclude there is a positive difference in means? Explain why it is justifiable to claim to have shown a positive difference when the alternative hypothesis does not specify a direction. How would the test change if you assumed the target null mean ($\mu_o = 0$) and variances ($\sigma_1 = \sigma_2 = 1$) were correct?

e) Find an expression for the power of the test if $\mu_1 - \mu_2 = c\sigma$. For example, suppose $c = 0.25$ would be on the lower boundary of being an important (practically significant) difference in means. Find the smallest n to have power at least 0.99 of detecting such a difference at $\alpha = 0.1$. Explain what having such high power allows you to say if you fail to reject H_o (as compared to having, say, 50% or 10% power).

3. **Oneway ANOVA**

Suppose $Y_{i1}, \ldots, Y_{in_i} \overset{\text{iid}}{\sim} N(\mu_i, \sigma^2)$ for $i = 1, \ldots, k$ groups. Let Y_i be the group i average and let $\bar{Y}$ be the overall average. Consider a test of $H_o : \mu_1 = \ldots = \mu_k$ vs. H_a : "not H_o".

a) Explain how the F statistic generalizes the pooled 2-sample t statistic.

$$F - \text{stat} = \frac{\text{SSM}/(k-1)}{\text{SSE}/(n-k)} = \frac{\sum_{i=1}^k n_i(Y_i - \bar{Y})^2/(k-1)}{\sum_{i=1}^k (n_i - 1)s_i^2/(n-k)}$$

For example, show how rejecting for large values of the this is the same as rejecting for large T^2 when $k = 2$ and $N = n_1 + n_2$ (what would unusually small values of T^2 suggest?).

b) To simplify, suppose $n_i = n$, for $i = 1, \ldots, k$, so $N = nk$ and $\bar{Y} = \frac{1}{k}\sum_{i=1} Y_i$. Show, using facts we have already proved, that $\frac{\text{SSE}}{\sigma^2} = \sum_{i=1}^k \sum_{j=1}^n \frac{(Y_{ij} - Y_i)^2}{\sigma^2} \sim \chi^2_{(N-k)}$ and independent of SSM, and that $\frac{\text{SST}}{\sigma^2} = \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{(Y_{ij} - \bar{Y})^2}{\sigma^2} \sim \chi^2_{(N-1)}$ if H_o is true. Also show SST = SSM + SSE, and infer that $\frac{\text{SSM}}{\sigma^2} \sim \chi^2_{(k-1)}$ if H_o is true. Conclude that the null sampling distribution of the F-statistic is $F_{(k-1,N-k)}$. Note that MSE = SSE$/(N-k)$ is an unbiased estimate for σ^2. If H_o is true, then MSM=SSM$/(k-1)$ is also unbiased for σ^2, but otherwise is biased high.

c) Define $R^2 = \frac{\text{SSM}}{\text{SST}} = 1 - \frac{\text{SSE}}{\text{SST}}$ as the proportion of variability explained by the groups. Show that the null sampling distribution of R^2 is Beta$(k-1, N-k)$. With $k = 5$ and $n = 10$, what values of the F-statistic would lead you to reject H_o? What values of R^2 would lead you to reject? What conclusion could you make?

d) Use my Wordle data to demonstrate, with $k = 5$ different start words and $n = 10$ games played with each start word. The responses are the numbers of attempts to guess the word (assume these are approximately Normal with a constant variance but possibly different means). Show a graph of the data, fill in the ANOVA table and compute the F-statistic and R^2.

4. **Multiple Comparisons**

The significance level α for a single test indicates the probability we will reject H_o when H_o is true. With multiple tests, the probability of at least one false rejection grows with the number of tests.

a) With m independent tests, we can compute the exact 'family-wise' error rate (FWER). For example, imagine administering drug tests to $m = 10$ members of a sports team, none of whom have used drugs. If each test has false positive rate α, what is the probability of at least one false positive? What level α should be used to make the family-wise rate equal to 0.05? Show this is greater than $0.05/m$.

b) For multiple testing without independence, we can use the Bonferroni adjustment. Prove the finite version of Boole's inequality:

$$P(\cup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$$

Explain how this leads to the Bonferroni adjustment for multiple tests and interval estimates that may or may not be independent.

c) For k independent samples, there are $\binom{k}{2}$ possible pairwise tests of equal means. Explain why these do not represent independent tests. Use my Wordle data as an example, with $k = 5$ start words.

d) For an ANOVA F test, describe the estimates made for the k unknown means when we reject and fail to reject H_o. In 1962 Charles Stein showed that, for estimating $k \geq 3$ mean parameters, using the k sample averages is *inadmissible*. Explain what this means and why it is called "Stein's Paradox in Statistics". Describe the improved James-Stein "shrinkage" estimates.

5. **The Normal Hierarchical Model**

The null hypothesis for a one-way ANOVA model is well defined, but the alternative hypothesis is not. If the means are not all equal, how do they vary? A Normal hierarchical model assumes they follow a Normal distribution. The 2024 WNBA scoring data gives the points scored in the each of the first $n = 5$ games for $k = 12$ women's professional basketball teams.

a) Write out the 2-level Normal hierarchical model and explain how it parameterizes the alternative hypothesis in a one-way ANOVA model, while allowing for the null hypothesis. Describe the parameters in the context of the WNBA data.

$$\text{level-1:} \qquad Y_i \,|\, \theta_i, \sigma^2 \overset{\text{indep}}{\sim} N(\theta_i, \sigma^2/n), \quad i = 1, \ldots, k$$

$$\text{level-2:} \qquad \theta_i \,|\, \mu, A \overset{\text{i.i.d.}}{\sim} N(\mu, A)$$

b) Explain the 1-way ANOVA F test in the context of this model. Say how the test changes if σ is assumed known (and why this doesn't make much difference when n and k are large).

c) Give the conditional distribution of $\theta_i|Y_i, \sigma^2, \mu, A$ and show how the James-Stein estimate anticipates regression towards the mean.

d) Use the WNBA data, assuming the mean square error is the true value of σ^2. Write out the marginal joint likelihood function for the hyperparameters μ and A. Assuming $p(\mu, A) \propto c$ (for $A > 0$) this becomes the joint posterior density. Transform to μ and B, for $B = \frac{V}{V+A}$ and show this is the joint pdf for a Gamma variable constrained to be between 0 and 1, as B is, and a conditional Normal variable μ. Time permitting, describe how Bayesian inference would proceed, and show the connection to the James-Stein estimate.