STAT 245 Spring 2021: Solutions Midterm

May 8, 2021

Intro

Let $X_1, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ and $Y_1, \ldots, Y_m \overset{i.i.d.}{\sim} N(\theta, \sigma^2)$, all independent. Want to make inference on $\Delta = \mu - \theta$.

Part 1

Let $\hat{\Delta} = \bar{X} - \bar{Y}$. Then

1. $\mathbb{E}(\hat{\Delta}) = \mathbb{E}(\bar{X}) - \mathbb{E}(\bar{Y}) = \mu - \theta = \Delta\,, \quad$ by linearity of $\mathbb{E}$.
2. $Var(\hat{\Delta}) = Var(\bar{X}) + Var(\bar{Y}) = \frac{\sigma^2}{n} + \frac{\sigma^2}{m}\,, \quad$ because of independence of $X = (X_1, \ldots, X_n)$ and $Y = (Y_1, \ldots, Y_m)$.
3. $\mathbb{E}((\hat{\Delta} - \Delta)^2) = Var(\hat{\Delta}) + (\mathbb{E}(\hat{\Delta}) - \Delta)^2 = Var(\hat{\Delta}) = \frac{\sigma^2}{n} + \frac{\sigma^2}{m}\,, \;$ by decomposition of MSE proven in HW 2 - Problem 4.

Part 2

Let $S_X = \sum_{i=1}^n (X_i - \bar{X})^2$ and $S_Y = \sum_{j=1}^m (Y_j - \bar{Y})^2$ and $S = \frac{S_X + S_Y}{\sigma^2}$. From class we know that

$$\frac{S_X}{\sigma^2} \sim \chi^2_{n-1} \quad \text{and} \quad \frac{S_Y}{\sigma^2} \sim \chi^2_{m-1}\,,$$

and since S_X and S_Y are functions of independent R.V.s, they are also independent. So S is the sum of two independent Chi-squared R.V.s. We can use the fact that a sum of independent $\chi^2_{r_i}$ is again χ^2_r with degrees of freedom given by $r = \sum_i r_i$ and conclude that

$$S = \frac{S_X}{\sigma^2} + \frac{S_Y}{\sigma^2} \sim \chi^2_{n+m-2}\,.$$

Alternatively, you could also say that $\frac{S_X}{\sigma^2} \sim \sum_{i=1}^{n-1} Z_i^2$ and $\frac{S_Y}{\sigma^2} \sim \sum_{j=1}^{m-1} Z_j^2$ all independent and identically distributed $N(0,1)$ so that $S = \sum_{i=1}^{n-1} Z_i^2 + \sum_{j=1}^{m-1} Z_j^2 \sim \chi^2_{n+m-2}$.

Part 3

From HW 2 - Problem 3, we know that $\mathbb{E}(\chi^2_p) = p$ so that $\mathbb{E}(S) = n + m - 2$ so that the statistics

$$S^\star = \frac{S_X + S_Y}{n + m - 2} = \frac{\sigma^2}{n + m - 2}\, S$$

is unbiased for σ^2 as $\mathbb{E}(S^\star) = \frac{\sigma^2}{n+m-2} \mathbb{E}(S) = \sigma^2$.
Any other estimator like $\frac{1}{n-1} S_X$ or $\frac{1}{m-1} S_Y$ is also a fine answer.

Part 4

From HW 3 - Problem 4, we know that $\bar{X} \perp S_X$ and $\bar{Y} \perp S_Y$, so that also $(\bar{X}, \bar{Y}) \perp (S_X, S_Y)$. Then $g(\bar{X}, \bar{Y}) \perp h(S_X, S_Y)$ for any functions g, h. Take $g(x, y) = x - y$ and $h(x, y) = x + y$ and conclude that $\bar{X} - \bar{Y} \perp S_X + S_Y$.

Part 5

To construct an $(1-\alpha)$ C.I. for Δ using the t distribution, first observe that we know

$$\bar{X} \sim N(\mu, \sigma^2/n) \quad \text{and} \quad \bar{Y} \sim N(\theta, \sigma^2/m)\,.$$

Thus, taking the difference $\bar{X} - \bar{Y}$ (which is a linear combination with weights +1 and -1) we still have a Normal distribution. So that

$$Z = \frac{\bar{X} - \bar{Y} - (\mu - \theta)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} = \frac{\hat{\Delta} - \Delta}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0, 1)\,.$$

Moreover, from Part 2,

$$S = \frac{S_X}{\sigma^2} + \frac{S_Y}{\sigma^2} \sim \chi^2_{n+m-2}\,,$$

which is independent of Z, by Part 4. Therefore, by definition of a Student's t distributed R.V., we have that

$$T = \frac{Z}{\sqrt{S/(n+m-2)}} \sim t_{n+m-2}\,,$$

so that

$$T = \frac{\hat{\Delta} - \Delta}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \cdot \frac{\sqrt{n+m-2}}{\sqrt{\frac{S_X}{\sigma^2} + \frac{S_Y}{\sigma^2}}} = \frac{\hat{\Delta} - \Delta}{\sqrt{S_X + S_Y}} \cdot \frac{\sqrt{n+m-2}}{\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim t_{n+m-2}\,.$$

Therefore letting $t^\star = t_{n+m-2, 1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of a t_{n+m-2}

$$\mathbb{P}(-t^\star \leq T \leq t^\star) = \mathbb{P}\left(-t^\star \leq \frac{\hat{\Delta} - \Delta}{\sqrt{S_X + S_Y}} \cdot \frac{\sqrt{n+m-2}}{\sqrt{\frac{1}{n} + \frac{1}{m}}} \leq t^\star\right)\,,$$

so that the $(1-\alpha)$ C.I. for Δ can be written as

$$\hat{\Delta} \pm t^\star\sqrt{S_X + S_Y} \cdot \frac{\sqrt{\frac{1}{n} + \frac{1}{m}}}{\sqrt{n+m-2}}\,.$$