Math 373  Lecture 34  

Comparing ranked populations  

How would you determine if male dogs are heavier on average than female dogs? Pick a sample of $n_m$ male dogs and a sample of $n_f$ female dogs, calculate the average weights $\mu_m, \mu_f$ of each sample and then determine if the data show (this is the alternative hypothesis) that $\mu_m > \mu_f$.  

Suppose we don’t have a scale but do have a balance. We can’t get the average weights of males and females, the one dog is lighter, equal to, or heavier than another. We can’t measure a dog’s weight but we can determine if a dog’s weight is lighter, equal to, or heavier than another.

Here $\alpha$ is the number of dogs, $\alpha = 4, 5, 6$. In the example, there are fewer male dogs, hence $P_1$ is the population of female dogs and $n_f = 6$. To distinguish the data for the two populations, we underline the data for $P_1$ as in the example.

Here $T_{1-} = 1.5 + 4.5 + 8.5 + 8.5 = 23$.

If all members of $P_1$ are smaller than $P_2$, then $T_1 = 1 + 2 + 3 + ... + n_1 = n_1(n_1 + 1)/2$. This is $T_{\min}$, the smallest $T_1$. If all members of $P_1$ are larger than $P_2$, then $T_1 = n_1(n_1 + 2n_2 + 1)/2$. This is $T_{\max}$, the largest possible $T_1$.

Theorem. $T_1 \in [T_{\min}, T_{\max}] = [n_1(n_1 + 1)/2, n_1(n_1 + 2n_2 + 1)/2]$. This is the range of $T_1$. The distribution is symmetric about the midpoint $T_{\mid \text{mid}} = (T_{\min} + T_{\max})/2$.

In the example above, $n_1 = 4, n_2 = 6$. Hence we will always have $T_1 \in [T_{\min}, T_{\max}] = [4(4+1)/2, 4(4+2\cdot6+1)/2] = [10, 34]$.

Suppose $x$ is in the lower half of $[a, b]$. Let $x^*$ be the symmetrical point in the upper half of $[a, b]$, $x^*$ is the same distance below $b$ as $x$ is above $a$. In $[3, 7]$, $3^* = 7$, $4^* = 6$, $5^* = 5$, $6^* = 4$ and $7^* = 3$.

Theorem. For $x \in [a, b]$, $x^* = a + b - x$.

Proof. The distance $x$ above $a$ equals the distance $x^*$ below $b$. Hence $x - a = b - x^*$. Hence $x^* = a + b - x$.

When the interval is $[T_{\min}, T_{\max}]$, $T^* = T_{\min} + T_{\max} - T$.

In the dog example, $[T_{\min}, T_{\max}] = [10, 34]$. Hence if $T = 13$, then $T^* = T_{\min} + T_{\max} - T = 10 + 34 - 13 = 31$.

For normal curves, $z_{1-\alpha} = -z_\alpha$. For the $F$ distribution we have $F_{1-} = 1/F_\alpha$. Hence tables are only given for one of the two half regions. For $T$ distribution $T_{1-\alpha} = T^*$. So

Theorem. $T_{1-\alpha} = T_{\min} + T_{\max} - T_\alpha$.

If we knew the weights, we would average them and “males usually weigh less than males” would be formalized as “$\mu_1 < \mu_2$”. Since we know only the ranks, we average ranks instead of weights.

Definition. $P_1 \leq_{\text{avg}} P_2$ iff the average of $P_1$’s ranks is $\leq$ the average of $P_2$’s ranks.

If $P_1$ is, on average, smaller than $P_2$, then its elements will have low ranks in the combined list and $T_1$ will be near the left end of its interval. Thus we will have $T_1 \in [T_{\min}, T_{\mid \text{mid}}]$ where $T_{\mid \text{mid}}$ is the midpoint of $[T_{\min}, T_{\max}]$.

Theorem. On average: $P_1 \leq_{\text{avg}} P_2$ right-tailed acceptance region $[T_{\min}, T_{\mid \text{mid}}]$.

Hypotheses

H$_0$: $P_1 \leq_{\text{avg}} P_2$, H$_a$: $P_1 >_{\text{avg}} P_2$, $[T_{\min}, T_{\mid \text{mid}}]$. $T_{1-}$ right-tailed

H$_0$: $P_1 \leq_{\text{avg}} P_2$, H$_a$: $P_1 >_{\text{avg}} P_2$, $[T_\alpha, T_{\max}]$, $[T_{\min}, T_\alpha]$ left-tailed

H$_0$: $P_1 \leq_{\text{avg}} P_2$, H$_a$: $P_1 \neq_{\text{avg}} P_2$, $[T_\alpha/2, T_{1-\alpha/2}]$, two-sided, $[T_{\min}, T_\alpha/2) \cup (T_{1-\alpha/2}, T_{\max}]$.

As before, “left-tailed” describes the direction of the rejection region $[T_{\min}, T_\alpha]$ which has area $\alpha$.

Does the data show that male dogs are heavier?

In the dog example $T_{1-} = 23 \in [10, 34]$. Since this is close to the center $T_{\mid \text{mid}} = 22$ of the interval, this probably is not significant evidence that males are heavier. Precisely —

H$_0$: males $>_{\text{avg}}$ females, H$_a$: males $\leq_{\text{avg}}$ females.

iff H$_a$: $P_1 >_{\text{avg}} P_2$, H$_0$: $P_1 \leq_{\text{avg}} P_2$ since $P_1$ = males,

iff H$_a$: $T_{1-} \in [T_\alpha, T_{\max}]$, H$_0$: $T_{1-} \in [T_{\min}, T_{\mid \text{mid}}]$ = null region.

$n_1 = 4$, $n_2 = 6$, $T_{\mid \text{mid}} = 13$, $T_{1-} = 13* = 10 + 34 - 13 = 31$.

Acceptance region $[T_{\min}, T_{1-\alpha}] = [10, 31]$.

$T_{1-} = 23 \in [10, 31]$.

No significant evidence that males are heavier.