STAT 135, 2. Midterm exam, Spring 2017, H. Pitters

NAME (IN CAPS):____
SID number and SECTION:____________________________________

Please write your answers on the exam sheets. Show your work or provide a brief explanation for all answers. This quiz is closed books. You are allowed to use a calculator. Answers should be simplified as much as possible. Good luck!

1	2	3	4	5	Σ

Question 1. [8] We have observations

$$X_1, \ldots, X_{n_1} \qquad Y_1, \ldots, Y_{n_2}$$

drawn independently from a continuous population of treatments and controls, respectively. We have no further information about the distribution of the populations, and we employ Wilcoxon's test based on rank sums. Let μ_X/μ_Y denote the population mean in the population of treatments/controls.

(1) What is the advantage of Wilcoxon's test based on rank sums over a parametric test for comparing two populations? [2]
A:
(a) No assumptions have to be made about the distribution underlying the observations (which comes at the prize of a somewhat technical alternative hypothesis).
(b) Wilcoxon's rank sum test is not sensitive to outliers.

(2) What is the null hypothesis H_0 in Wilcoxon's rank sum test? [2]
A:

$$H_0\colon \text{"treatment and control population are identical"}.$$

More formally: $H_0\colon X_1 =_d Y_1$.

(3) Suppose Wilcoxon's rank sum test does not reject H_0 based on the data. Does this imply that Wilcoxon's rank sum test does not reject $\mu_X \neq \mu_Y$? Explain. [2]
A: If Wilcoxon's test does not reject H_0 this implies that the test does not reject $\mu_X = \mu_Y$ (as the latter statement is weaker than the null hypothesis). Put differently, in light of the data, $\mu_X = \mu_Y$ cannot be rejected against the alternative H_A which implies $\mu_X \neq \mu_Y$.

(4) What is the minimal/maximal value of the sum of ranks $R_1 := \sum_{i=1}^{n_1} \operatorname{rank}(X_i)$ of the observations from the treatment population? [2]

A: Minimal value is $1 + \cdots + n_1 = \frac{1}{2}n_1(n_1+1)$, maximal value is $n_2 + 1 + \cdots + n_2 + n_1 = \frac{1}{2}n_1(n_1+1) + n_1 n_2$.

Question 2. [9] Car manufacturers are interested in mileage performance of two models of automobiles. The table shows the miles per gallon rating for each of twelve randomly selected vehicles of each model.

model 1	rank	model 2	rank
20.5	19.5	20.1	17
19.2	12	18.0	6
19.8	14.5	16.9	1
19.8	14.5	17.8	5
20.5	19.5	18.8	9.5
21.0	22	17.3	2
20.2	18	21.1	23
18.8	9.5	19.7	13
18.9	11	18.5	7
18.6	8	17.4	3
19.9	16	17.6	4
20.6	21	21.3	24

(1) Compute the sum of ranks R_1 for model 1. [2]

A: First, rank the pooled observations (see table). Summing ranks for model 1 one obtains 185.5.

(2) According to a rule of thumb, the null distribution of R_1 is approximately normal for sample sizes n_1, n_2 greater than or equal to 10. We don't trust this rule, and we want to convince ourselves that it can be reasonably employed in our specific example. To this end we bootstrap the distribution of R_1. Write down the corresponding algorithm in pseudocode. (Be concise!) [4]

A: We carry out a nonparametric bootstrap:

(a) Let F'_{n_1} and F'_{n_2} denote the empirical cdf of the control and treatment population, respectively.

(b) Draw (simulated) independent samples

$$(x_1^*, \ldots, x_{n_1}^*) \sim F'_{n_1} \qquad (y_1^*, \ldots, y_{n_1}^*) \sim F'_{n_2},$$

and compute the corresponding sum of ranks $R_1^* = \sum_{i=1}^{n_1} \text{rank}(x_i^*)$.

(c) Repeat this B times to obtain simulated values $R_{1,1}^*, \ldots, R_{1,B}^*$.

(d) Check whether $R_{1,1}^*, \ldots, R_{1,B}^*$ are independent draws from a normal distribution with mean $n_1(n_1+n_2+1)/2$ and variance $n_1 n_2(n_1+n_2+1)/12$. (E.g. could be checked informally by looking at the histogram of the $R_{1,i}^*$s.)

(3) Compute the p-value of Wilcoxon's rank-sum-test using a normal approximation to the null distribution of R_1. (Hint: $\mathbb{E}[R_1] = n_1(n_1 + n_2 + 1)/2$, $\mathrm{Var}(R_1) = n_1 n_2(n_1 + n_2 + 1)/12$.) [3]

A: Wilcoxon's test rejects the null for large values of $|R_1 - \mathbb{E}[R_1]|$. In our example we have $\mathbb{E}[R_1] = n_1(n_1+n_2+1)/2 = 6 \cdot 25 = 150$, $\mathrm{Var}(R_1) = n_1 n_2(n_1+n_2+1)/12 = 300$. The p-value is the probability to see values as extreme or even more extreme than $|185.5 - 150| = 35.5$, hence given by

$$\mathbb{P}\left\{|R_1 - \mathbb{E}[R_1]| > 35.5\right\} = \mathbb{P}\left\{\frac{|R_1 - \mathbb{E}[R_1]|}{\sqrt{300}} > \frac{35.5}{\sqrt{300}}\right\} \approx 2\Phi(\frac{35.5}{\sqrt{300}}) \approx 2\Phi(2.05) = 0.0404.$$

Question 3. [6] To measure the effect of a tranquilizer on patients their pulse rates were recorded before and after taking the tranquilizer. Model the observations as draws from a normal distribution, and denote by μ_b, μ_a the mean (population) pulse rates before/after taking the tranquilizer.

before (X_i)	after (Y_i)	difference ($X_i - Y_i$)
80	74	6
84	78	6
79	80	-1
82	75	7
80	79	1
81	77	4

(1) Conduct a test at significance level $\alpha = 0.05$ for the hypothesis that the tranquilizer has no effect on patients' pulse rates versus the alternative that the tranquilizer changes their pulse rates. [4]

A: We consider a matched samples test with observations the differences $d_1, \ldots, d_6$ in pulse rates before and after taking the tranquilizer (see table). We test the null hypothesis that there is no effect,

$$H_0\colon \mu_b = \mu_a$$

versus the alternative that the tranquilizer has an effect,

$$H_A\colon \mu_b \neq \mu_a.$$

We reject H_0 for large absolute values of the test statistic

$$\frac{\bar{D}_n - (\mu_b - \mu_a)}{s_{\bar{D}_n}/\sqrt{n}},$$

where $\bar{D}_n = \frac{1}{n}\sum_{i=1}^n D_i$, $D_i = X_i - Y_i$, and $s_{\bar{D}_n} = \frac{1}{n-1}\sum_{i=1}^n (d_i - \bar{d}_n)^2 = 3.19$. Recall that the null distribution is a t distribution with $n - 1$ degrees of freedom. Thus, the critical value is $t_{\text{crit}} = |t_5(0.025)| = t_5(0.975) = 2.571$, the $\alpha/2$-percentile of the t distribution with $n-1=5$ df (from distribution table). The test statistic applied to our data yields $2.94 > t_{\text{crit}}$ (since $\bar{D}_n = 3.83$), thus the test rejects H_0, and the data contradicts the hypothesis that the tranquilizer has no effect on patients' pulse rates.

(2) Find a 95% confidence interval for the average decrease $\mu_b - \mu_a$ in pulse rate. [2]

A: Use $\bar{D}_n$ as an estimator for $\mu_b - \mu_a$. Notice that the question allows to provide any 95% confidence interval. Want to find interval $C(X) := [l(X), r(X)]$ such that $\mathbb{P}\{\mu_b - \mu_a \in C(X)\} = \alpha = 95\%$. From part (1) we obtain

$$\begin{aligned}\alpha &= \mathbb{P}\left\{t_{n-1}(\alpha/2) \le \frac{\bar{D}_n - (\mu_b - \mu_a)}{s_{\bar{D}_n}/\sqrt{n}} \le t_{n-1}(1-\alpha/2)\right\}\\ &= \mathbb{P}\left\{\bar{D}_n - \frac{s_{\bar{D}_n}}{\sqrt{n}} t_{n-1}(1-\alpha/2) \le \mu_b - \mu_a \le \bar{D}_n - \frac{s_{\bar{D}_n}}{\sqrt{n}} t_{n-1}(\alpha/2)\right\},\end{aligned}$$

hence $[l(X), r(X)] = [0.48, 7.18]$ is a 95% confidence interval for $\mu_b - \mu_a$.

Question 4. [8] True or false? Explain.

(1) The significance level of a statistical test is equal to the probability that the null hypothesis is true. [1]

A: False. The significance level is the probability of a type I error, that is the probability of rejecting H_0 given that H_0 is true.

(2) If a significance level of a test is decreased, the power would be expected to increase. [1]

A: False. Usually, one can decrease the significance level α at the expense of increasing β. Since the power is defined as $1-\beta$, the power is expected to decrease.

(3) If a test is rejected at the significance level α, the probability that the null hypothesis is true equals α. [1]

A: False. If H_0 is rejected at significance level α this means that the data contradict H_0 and if a large number of independent data sets would be collected, only in $(100\alpha)\%$ of the cases would one falsely reject H_0.

(4) The probability that the null hypothesis is falsely rejected is equal to the power of the test. [1]

A: False. The former is the probability for a type I error, while the latter is $1-\beta$, where β is the probability of a type II error.

(5) A type I error occurs when the test statistic falls in the rejection region of the test. [1]

A: False. A type I error occurs when the test statistic falls in the rejection region given that the null hypothesis is true.

(6) A type II error is more serious than a type I error. [1]

A: False. By definition the type I error is the more serious error.

(7) The power of a test is determined by the null distribution of the test statistic. [1]

A: False. The power of a test is determined by the distribution of the test statistic under the alternative hypothesis.

(8) The likelihood ratio is a random variable. [1]

A: True. The likelihood ratio $\Lambda = \Lambda(X_1, \ldots, X_n)$ is a statistic that depends on the random variables $X_1, \ldots, X_n$ modeling the observations. Once one evaluates the likelihood ratio (i.e. one plugs in the data $x_1, \ldots, x_n$ into the map Λ), one obtains a realization of Λ which is, of course, deterministic.