EXAM SOLUTION

1 Landing the Mars rover [17 pt]

a) **[1 pt]** $p^6 = 0.5314$

b) **[1 pt]** $p^6 + 6 * p^5(1 - p) = 0.886$

c) **[3 pt]**

[2 pt] $p_2^6 + 6 * p_2^5(1 - p_2) + 15 * p_2^4(1 - p_2)^2 = 0.901$

[1 pt] The alternative solution has a higher probability to operate on Mars.

d) **[2 pt]**

[1 pt] χ^2 test.

[1 pt] $H_0 : p = 0.99$; $H_1 : p \neq 0.99$.

e) **[10 pt]**

All terms in E with 0.01^2 and lower are ignored. That leaves:

[1 pt] $E(5) = 100 * 6 * p_3^5(1 - p_3) = 5.71$ $E(6) = 100 * p_3^6 = 94.15$

[3 pt] We compute the χ^2 Statistic as $\chi^2 = \sum_i (O_i - E_i)^2/E_i = (29 - 5.71)^2/5.71 + (71 - 94.15)^2/94.15 = 100.69$.

[3 pt] For 7 possible outcomes, we have $df = 6$ degrees of freedom. We look up in the table $q\chi_6^2(0.95) = 12.59$

[3 pt] As $\chi^2 >> q\chi_6^2(0.95)$, we reject the null hypothesis. This data is incompatible with the claimed $p_3 = 0.99$. As fewer complete surviving rovers were observed, its clear the real probability must be lower. The suspensions are not as good as claimed.

2 Urban architecture - in wood [23 pt]

a) **[3 pt]** $\sum_i (X_i - 10)^2 = \sum_i X_i^2 - 20 \sum_i X_i + \sum_i 100$

$\bar{X} = 1/400 * (\sum_i X_i^2 - \sum_i (X_i - 10)^2 + 2000) = 361.4$

b) **[3 pt]**

[1 pt] We have $n = 20$ datapoints: $n * 0.18 = 3.6$ non integer. $q_{0.18} = X(4) = 346$

All other points are integer:

[0.66 pt] $q_{0.25} = 1/2(X(5) + X(6)) = 348$

[0.66 pt] $q_{0.5} = 1/2(X(10) + X(11)) = 359$

[0.66 pt] $q_{0.75} = 1/2(X(15) + X(16)) = 369$

c) **[5 pt]**

[2 pt] boxplot well drawn to proportions.

[0.5 pt] IQR=$q_{0.75} - q_{0.25}$=21

[0.5 pt] box between $q_{0.75}$ and $q_{0.25}$

[0.5 pt] candlesticks to $q_{0.75} + 1.5IQR = 400.5$ and $q_{0.25} - 1.5 * IQR = 316.5$

[0.5 pt] mark 425 as outlier

[1 pt] notice: quite symmetric distribution but one clear outlier

d) **[5 pt]**

[1 pt] recompute mean without datapoint: $\bar{X}_{new} = 1/19(20\bar{X} - X_{20}) = 358.05$

[1 pt] n-1 in estimator

[3 pt] $\sigma^2 = 1/18(\sum_i^{20} X_i^2 - 425^2 - 19\bar{X}_{new}^2) = 131.386$, giving $\sigma = \sqrt{131.386} = 11.462$

[-1 pt] for confusing σ and σ^2

e) **[5 pt]**

[4 pt] $P(X > 350) = 1 - P(X < 350) = 1 - P(Z < (350 - 358.05)/11.462) = 1 - P(Z < -0.702) = P(Z < 0.702) = 0.758$

[1 pt] This exceeds the requried acceptance rate of 0.70, hence we are good to go with the building!

f) **[2 pt]** $P = 15/20 = 0.75$. Not surprisingly, this value is very close to the model.

3 Transport in quantum materials [19 pt]

a) **[6 pt]**

[1 pt] (just writing down) $SSE = \sum_i(\hat{U}(X_i) - U_i)^2 = \sum_i(RI_i - U_i)^2$

[1 pt] (seeing concept) We find the coefficient R by minimizing the SSE $d/dRSSE = 0$

[4 pt] $\sum_i(RI_i - U_i)I_i = 0$ -> $R = \sum_i I_iU_i/\sum_i I_i^2 = 2.45$

b) **[5 pt]**

[1 pt] Exactly the same as in a), just replace I_i by I_i^2.

[4 pt] $\beta = \sum_i I_i^2U_i/\sum_i I_i^4 = 0.985$

c) **[1 pt]** no.

d) **[7 pt]**

[3 pt] $SSE^{(1)} = \sum_i(RI_i - U_i)^2 = 14.23$

[3 pt] $SSE^{(2)} = \sum_i(\beta I_i^2 - U_i)^2 = 3.26$

[1 pt] The second model explains much more of the variance at the same number of parameters (1). It is the better model to the data.

4 VegiSTEAK [27 pt]

a) **[2 pt]**

[1 pt] $\bar{X}_1 = 7.83$

[1 pt] $s_1 = 1.47$

b) **[2 pt]**

[1 pt] H_0: $\mu_1 \leq \mu_2$ $H_1 : \mu_1 > \mu_2$.

[1 pt] We do a one sided test as we want to know if our product beats the steak.

c) **[8 pt]**

[1 pt] We are not given a known variance, hence we have to perform a Welch test.

[2 pt] Welch's T-statistic is computed as $T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_1^2 + s_2^2}}\sqrt{6} = 1.872$.

[3 pt] degrees of freedom: $a = (\frac{s_1^2}{n} + \frac{s_2^2}{n})^2 = 1.79$ and $b = \frac{1}{2n-1}((\frac{s_1^2}{n})^2 + (\frac{s_2^2}{n})^2) = 0.217$. Then $df = round(a/b) = round(8.24) \approx 8$.

[1.5 pt] In this one-sided test, large values of T speak against H_0. $qt(p = 0.95, df = 8) = 1.86$.

[0.2 pt] As $T > qt(p = 0.95, df = 8)$, we have (barely) statistically significant evidence that VegiSTEAK beats the T-bone.

d) **[7 pt]**

	df [1 pt]	SS [4 pt]	MS [1 pt]	F [1 pt]
Model	2	$6 * \sum_{i=1}^{3}(\bar{X}_i - \bar{X})^2 = 25$	12.5	F=MST/MSE=8.72
Error	15	$5 * \sum_{i=1}^{3} s_i^2 = 21.5$	1.43	
Total	17	46.5		

e) **[4 pt]**

The F-statistic of this model is 8.72.

[1 pt] We have $\nu_1 = 2$, $\nu_2 = 15$ degrees of freedom.

[1 pt] Table: $qF_{2,15}(p = 0.95) = 3.682$.

[2 pt] Our F falls well inside the critical region. We reject the null Hypothesis.

f) **[4 pt]**

[2 pt] ANOVA tells us there is a difference between the groups. The VegiSTEAK has the highest mean, hence there is significant evidence that it is the best meat-replacement product.

[2 pt] VegiSTEAK tasted better than the T-bone steak, yet not enough to claim statistical significance. We have to accept that they may be the same or the T-bone even tastes better.