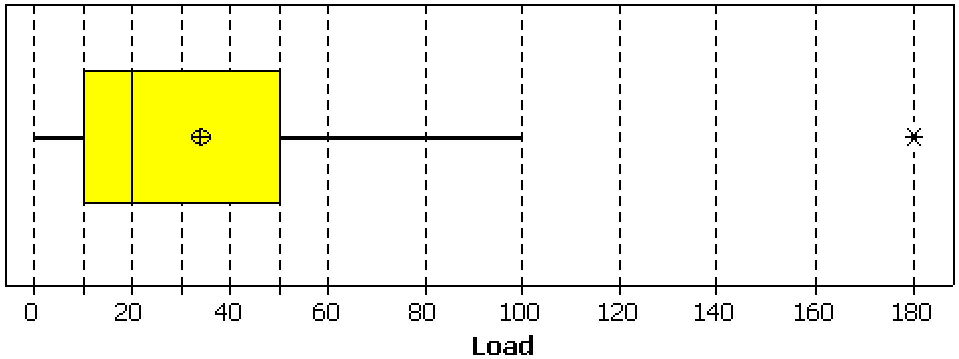
1. A sample of 43 measurements of breaking load is summarized in this boxplot.



The location of the sample mean is indicated by the symbol ⊕.

(a) Determine whether the outlier is mild or extreme. Show your working. [5]

(b) Excluding the outlier, state any two other reasons to deduce positive skew. [2]

(c) Using the summary statistics [3]

$$n = 43\,, \quad \sum x = 1462\,, \quad \sum x^2 = 107\,212$$

find the value of the sample standard deviation s for these data.

---

2. The load W (in g) needed to break a new type of cotton thread is known to be a random quantity that follows a Normal distribution. It is believed that the population mean breaking load is 20 g and the strength of that belief is represented by the standard deviation $\sigma_o = 1\text{g}$. A random sample of 25 such threads has a mean breaking load of 18.72 g with a sample standard deviation of 2.50 g.

(a) Construct a Bayesian 95% confidence interval estimate for the true mean breaking load μ. [5]

(b) Construct a classical 95% confidence interval estimate for the true mean breaking load μ. [5]

(c) Is there sufficient evidence to conclude that the true mean breaking load μ is not 20 kg? [2]

(d) Provide a *brief* reason for the different widths of your two confidence intervals. [2]

---

3. A continuous random quantity X has a probability density function defined by

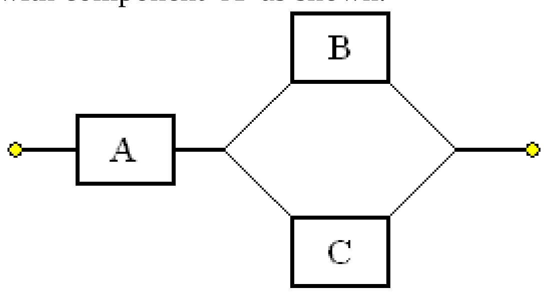
$$f(x) = \begin{cases} \dfrac{\pi}{4}\cos\left(\dfrac{\pi x}{2}\right) & (-1 \le x \le +1) \\ 0 & (\text{otherwise}) \end{cases}$$

(a) Find the exact value of $\mathrm{P}\left[\,|X| < \dfrac{1}{2}\right]$. [5]

(b) Find the exact value of the population variance σ^2. [5]

---

4. In a network, components B and C are connected in parallel and that subsystem is connected in series with component A as shown.



The subsystem works if at least one of the components B or C functions properly. The complete system works only if component A and the subsystem both function properly.
Let A = the event that component A functions properly;
B = the event that component B functions properly; and
C = the event that component C functions properly.
It is known that P[A] = . 80, independently of the other two components, and that P[B] = .50, independently of component A . However, the event C is dependent on the event B . A failure in component B increases the load on component C and therefore reduces the probability that component C will work.
It is known that $\mathrm{P}[C \mid B] = .90$ and $\mathrm{P}[C \mid \widetilde{B}] = .50$.

(a) Find $\mathrm{P}[B \cup C]$. [6]

(b) Find the probability that the complete system works. [3]

(c) Find $\mathrm{P}[C]$. [5]

---

5. The performances of eight machines of varying capacities in a standard test are measured before (x) and after (y) the application of a proposed new cleaning device. It is known that both populations are normally distributed with a common variance. The manager will invest in the new cleaning device only if there is evidence, at a level of significance of 5%, that the device has improved the average performance score by more than 1.0.

(a) Which of the two sample t-tests (paired or unpaired) should be conducted? State the reason for your selection. [3]

(b) Write down the appropriate null and alternative hypotheses. [3]

(c) Use the appropriate set of Minitab output below, to conduct the appropriate hypothesis test, at a level of significance of 5%. [4]

(d) What should the manager's decision be? [2]

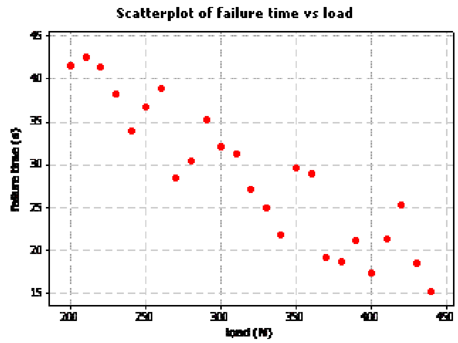
Two-Sample T-Test and CI: After, Before

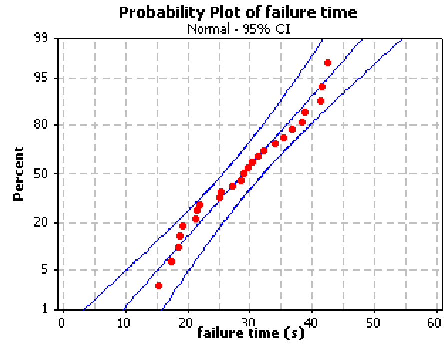
```
        N   Mean  StDev  SE Mean
After   8  154.0   49.9       18
Before  8  150.0   49.0       17

Difference = mu (After) - mu (Before)
Estimate for difference:  4.0
95% lower bound for difference:  -39.5
T-Test of difference = 1 (vs >): T-Value = 0.121  P-Value = 0.453  DF = 14
Both use Pooled StDev = 49.4397
```

Paired T-Test and CI: After, Before

```
            N   Mean  StDev  SE Mean
After       8  154.0   49.9     17.6
Before      8  150.0   49.0     17.3
Difference  8   4.00   4.47     1.58

95% lower bound for mean difference: 1.004
T-Test of mean difference = 1 (vs > 1): T-Value = 1.897  P-Value = 0.0498
```

6. The time (y) to failure of a device is measured for each of 25 values of a load (x). A plot of the observed values of (x, y) and a normal probability plot of the residuals from simple linear regression are shown here.





(a) State two reasons why the simple linear regression model for Y as a function of x is appropriate. [2]

Use the following summary statistics to answer the remaining parts of this question.

$$n = 25 \qquad \sum x = 8\,000 \qquad \sum y = 719.81$$
$$\sum x^2 = 2\,690\,000 \qquad \sum xy = 216\,777.00 \qquad \sum y^2 = 22\,391.88$$
$$n\,S_{xx} = 3\,250\,000 \qquad n\,S_{xy} = -339\,055.00 \qquad n\,S_{yy} = 41670.5639$$

(b) Find the equation of the regression line $\left(\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x\right)$. [3]

(c) Construct the ANOVA table for this simple linear regression model. [5]

(d) To the nearest 1%, how much of the total variation in Y is explained by the simple linear regression model? [3]

(e) Conduct an appropriate hypothesis test to determine whether or not there is a significant linear association between Y and x. [3]

(f) Construct a 95% prediction interval for a future value of Y when $x = 320$. [4]

7. A partially completed decision tree is shown here.

G = the item is good
D = the item is defective
$D = \widetilde{G}$

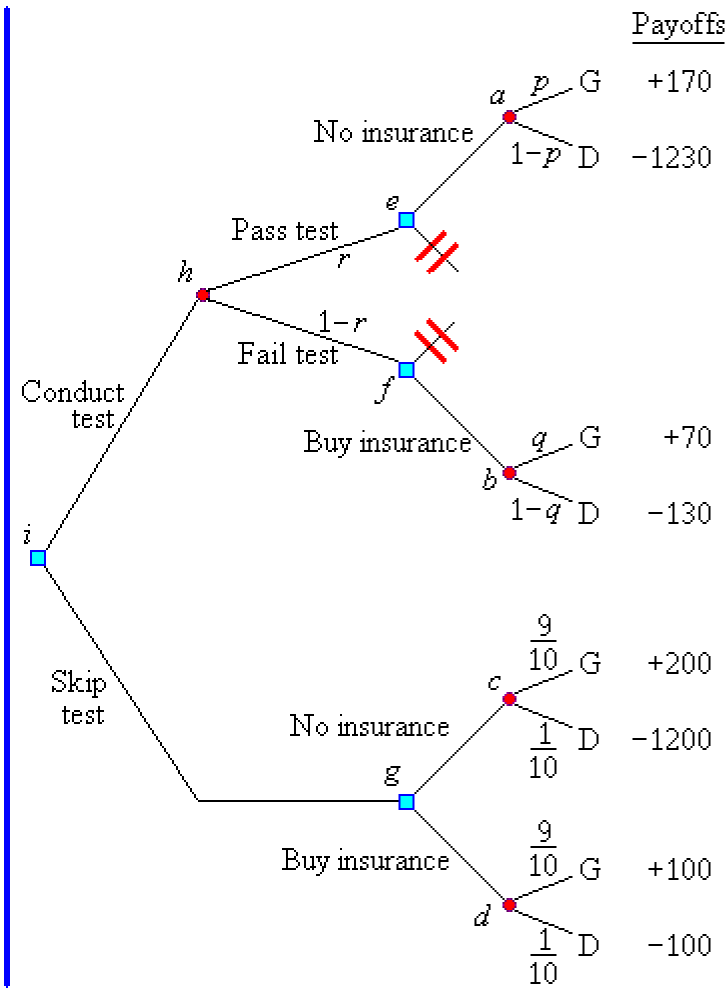
The payoffs for items are as shown at the right edge of the diagram (in units of dollars). Insurance protects against the loss due to a defective item.

The test costs \$30.
Insurance, costing \$100, is purchased if the item fails the test and is not purchased if the item passes the test.

It is known that 90% of all items are good.

If an item is good then it will pass the test 80% of the time.

If an item is defective, then it will pass the test 20% of the time.



(a) Verify that the probability labelled ***(p)*** is exactly $\frac{36}{37}$. [3]

(b) Verify that the probability labelled ***(q)*** is exactly $\frac{9}{13}$. [2]

(c) Find the probability labelled ***(r)***. [2]

(d) Verify that the expected value labelled ***(a)*** is exactly $\frac{4890}{37}$. [1]

(e) Find the other expected values labelled ***(b)***, ***(c)***, ***(d)***, ***(e)***, ***(f)***, ***(g)***, ***(h)***, ***(i)*** and hence determine the optimum strategy that maximizes the expected payoff. [10]

(f) By how much can the test cost be increased before the optimum strategy changes? [2]

BONUS QUESTION

8. A two-sample hypothesis test is designed to determine whether or not the true mean efficiency rating μ_{new} of a prototype machine is higher than the true mean efficiency rating μ_{old} of an existing machine. Random samples of equal size n are taken from each type of machine and their efficiencies are recorded. It is known that the efficiencies are normally distributed with the same known population standard deviation $\sigma = 10$ in both cases. [+5]

Find the minimum sample size $n_{\min}$ needed to ensure that the probabilities α (of type I error) and $\beta(5)$ (type II error when $\mu_{\text{new}} - \mu_{\text{old}} = 5$) are both 5%.
[If you quote a formula for $n_{\min}$, then derive that formula.]

[Also provided with this examination paper were tables of the standard normal c.d.f. (the z tables) and of the critical values of the t distribution (the *t* tables).]

☜ Back to the index of questions On to the solutions of this examination ☞