Statistics 1 - Reviewed Notes

Dom Hutchinson

April 28, 2019

Contents

1 General

Definition 1.1 - *Order Statistic*
An *Order Statistic* is a data set where the data has been placed in increasing order of value, not time.

1.1 Exploratory Data Analysis

Definition 1.2 - *Exploratory Data Analysis*
Exploratory Data Analysis is an approach to data analysis which focuses on summarising the main characteristics of the set
Definition 1.3 - *Median*
The *Median* is the central value of a data set. For an odd-sized ($n = 2m + 1$) data set the median is $x_{(m+1)}$. For an even-sized ($n = 2m$) data set the median is $\frac{1}{2}(x_m + x_{m+1})$.

Definition 1.4 - *Sample Mean*
The *Sample Mean* is the average value of all data points within a sample. For a sample $\{x_1, \dots, x_n\}$

$$\bar{x} := \frac{1}{n}\sum_{i=1}^{n} x_i$$

Definition 1.5 - *Trimmed Sample Mean*
The *Trimmed Sample Mean* is the average value of a subset of data points within a sample. The subset is defined to ignore the $\frac{\Delta}{2}\%$ largest & smallest values of the sample. For a $\Delta\%$ trimmed mean we define

$$\bar{x}_\Delta := \frac{1}{n-2k}\sum_{i=k+1}^{n-k-1} x_i \text{ with } k = \left\lfloor \frac{n\Delta}{100} \right\rfloor$$

Definition 1.6 - *Sample Variance*
Sample Variance is a measure of spread of data in a sample around the sample mean. For a sample $\{x_1, \dots, x_n\}$

$$s^2 := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n-1}\left(\left(\sum_{i=1}^{n} x_i^2\right) - n\bar{x}\right)$$

Definition 1.7 - *Hinges*
Hinges describe the spread of data in a sample, while trying to ignore extreme data. The *Lower Hinge*, H_1, is the median of the set containing the median & values with rank <u>less</u> than the sample median . The *Upper Hinge*, H_3, is the median of the set containing the median & values with rank <u>greater</u> than the sample median.

Definition 1.8 - *Quartiles*
Quartiles describe the spread of data in a sample. The *Lower Quartile*, Q_1, is the median of the set of values with rank <u>less</u> than the sample median . The *Upper Quartile*, Q_3, is the median of the set of values with rank <u>greater</u> than the sample median.
<u>N.B.</u> - These sets do <u>not</u> contain the median.

Definition 1.9 - *Five-Number Summary*
The *Five-Number Summary* of a sample contains the sample's: median; lower hinge; upper hinge; minimum value; & maximum value.

Definition 1.10 - *Skew*
Skew describes the spread of values in a sample which are less than the median, relative to the

spread of values greater than the median. A sample is *Left-Skewed* if $H_3 - H_2 > H_1 - H_2$. A sample is *Right-Skewed* if $H_3 - H_2 < H_1 - H_2$.

1.2 Graphical Plots

Definition 1.11 - *Histogram*
A *Histogram* is a plot used to visualise the shape & distribution of a sample. A *Histogram* can be produced by the following process

i) Divide the range of data values into K intervals (bins) of equal width.

ii) Counter the frequency of observations falling into each interval.

iii) Display a plot of joined columns above each interval, with the columns height proportional to the count for that interval.

Definition 1.12 - *Stem-and-Leaf Plot*
A *Stem-and-Leaf Plot* is a plot used to visualise the shape & distribution of a sample. A *Stem-and-Leaf Plot* gives more information about the data in a sample than a *Histogram*, since it displays the value of each element. A *Stem-and-Leaf Plot* can be produced by the following process

i) Truncate or round the data values so that all the variation is the in the last two, or three, significant figures;

ii) Separate each data value into a stem (consisting of all digits except the last) and a leaf (last digit);

iii) Write the stems in a vertical column, smallest to biggest, and draw a vertical line to separate from the right column;

iv) Write each leaf in the row to the right of its corresponding stem, in increasing order;

v) Record any strikingly low or high values separately from the main stem, displaying the individual values in a group above the main stem or below.

1.3 χ^2 Distribution

Definition 1.13 - *χ^2 Distribution*
Let W be a random variable whose *Moment Generating Function* $\mathcal{M}_W(t) = (1-2t)^{-r/2}$, with $r \in \mathbb{N}$ & $t < 1/2$, then $W \sim \chi^2_r$. Here W is said to be distributed by the χ^2 distribution with r degrees of freedom.

Theorem 1.1 - *Samples from Normal Distribution are χ^2 Distributed*
Let $X_1, \dots, X_n$ be a simple random sample from $N(\mu, \sigma^2)$. Then

$$\begin{aligned} \sum_{i=1}^n \frac{(X_n - \mu)^2}{\sigma^2} &\sim \chi^2_n \\ \sum_{i=1}^n \frac{(X_n - \overline{X})^2}{\sigma^2} &\sim \chi^2_{n-1} \end{aligned}$$

1.4 t-Distribution

Definition 1.14 - *t-Distribution*
Let X, Y be irv with $X \sim N(0,1)$ & $Y \sim \chi^2_r$. Define random variable $Z := \dfrac{X}{\sqrt{Y/r}} \implies Z \sim t_r$.
The t-distribution is shaped similarly to $N(0,1)$, but with heavier tails.

Theorem 1.2 - *Distance between Sample Mean & Population Mean*
Let $X_1, \dots, X_n$ be a simple random sample from $N(\mu, \sigma^2)$. Then

$$\frac{\sqrt{n}}{s}(\overline{X} - \mu) \sim t_{n-1}$$

This allows us to estimate how far apart μ & $\overline{X}$ are without knowing σ^2.

2 Parametric Models

Definition 2.1 - *Parametric Models*
Parametric Models are the class of statistical distributions whose probability mass/density function take parameters. These parameters represent values of interest in the population, such as mean or variance. We generally do not know these values so we estimate them from samples.

Definition 2.2 - *Quantity of Interest*
When analysing distributions it often helps to define *Quantities of Interest* about the distributions (*e.g.* Mean). These are defined as functions in terms of the parameters $\tau(\theta)$. We estimate *Quantities of Interest* by passing estimated values of the parameters $\hat{\tau} = \tau(\hat{\theta})$.

Definition 2.3 - *Joint Probability Density of Simple Random Sample*
Let $X_1, \dots, X_n$ be iid random variables representing the values of a simple random sample. The probability of obtaining $x_1, \dots, x_n$ as the values obtained by a simple random sample is

$$f_{X_1,\dots,X_n}(x_1, \dots, x_n) = \prod_{i=1}^{n} f_X(x_i; \theta)$$

3 Estimating Parameters

Definition 3.1 - *Likelihood Function*
The *Likelihood Function* provides an curve which estimates the parameters of a *Parametric Distribution*, for an observed value of the distribution.

$$\begin{array}{rcl} L(\theta; x) & := & \begin{cases} p(x;\theta) & \text{Discrete} \\ f(x;\theta) & \text{Continuous} \end{cases} \\ L(\theta; x_1, \dots, x_n) & := & \begin{cases} \displaystyle\prod_{i=1}^{n} p_{X_i}(x_i;\theta) & \text{Discrete} \\ \displaystyle\prod_{i=1}^{n} f_{X_i}(x_i;\theta) & \text{Continuous} \end{cases} \end{array}$$

Definition 3.2 - *Log-Likelihood Function*
The *Log-Likelihood Function* is the natural logarithm of the *Likelihood Function*. Since the natural logarithm is an increasing function, the *Log-Likelihood Function* has the same maximum

as the *Likelihood Function.* The *Log-Likelihood Function* is useful when dealing with multiple observations since it requires the sum of the probability functions, rather than the product.

$$\begin{aligned}\ell(\theta;x) &:= \ln L(\theta;x)\\ \ell(\theta;x_1,\dots,x_n) &:= \sum_{i=1}^n \ln L(\theta;x_i)\end{aligned}$$

Definition 3.3 - *Maximum Likelihood Estimate*
Maximum Likelihood Estimate is a technique for estimating the parameters of a *Parametric Distribution* using the *Likelihood Function. Maximum Likelihood Estimate* takes in a series of observations & returns the parameters which are the most likely to cause these observations. Generally differentiation will be required to find this value.

$$\hat{\theta}_{mle} := \text{argmax}_\theta L(\theta;x)$$

For multiple observations $\hat{\theta}_{mle}$ can generally be found as the solution to

$$0=\sum_{i=1}^n \frac{\partial}{\partial\theta}\ell(x_i;\hat{\theta}_{mle})$$

To find the *Maximum Likelihood Estimate* for *Parametric Distributions* with multiple parameters we use the previous equation, taking the partial derivative wrt each parameter in turn & then solve as a set of simultaneous equations. For a model with parameters α & β we have that $\hat{\alpha}_{mle}$ & $\hat{\beta}_{mle}$ are the solutions to the following pair of simultaneous equations

$$0=\sum_{i=1}^n \frac{\partial}{\partial\alpha}\ell(x_i;\hat{\alpha}_{mle},\hat{\beta}_{mle}) \quad \& \quad 0=\sum_{i=1}^n \frac{\partial}{\partial\beta}\ell(x_i;\hat{\alpha}_{mle},\hat{\beta}_{mle})$$

N.B. - When given a distribution which is non-regular (*e.g.* Uniform Distribution) it is best to work with $L(\theta)$ directly (by inspection).

Definition 3.4 - *Method of Moments*
Method of Moments is a method for estimating the parameters of a *Parametric Model.* The *Method of Moments* states that we should equate $\mathbb{E}(X^k;\hat{\boldsymbol{\theta}}_{mom})=m_k$ and solve these for $k\in[0,m]$ (where m is the number of parameters) as simultaneous equations.

4 Assessing Fit

Definition 4.1 - *Probability Plot*
A *Probability Plot* compares a sample against a theoretical model (using estimators). A *Probability Plot* plots $F_X(y;\hat{\theta})$ against $\hat{F}(y)$ for $y\in\{x_1,\dots,x_n\}$. The closer a plot is to $y=x$ the better the estimator is.

Definition 4.2 - $(Q-Q)$ *Plot*
A *Quartile Plot* compares a sample against a theoretical model (using estimators). A $(Q-Q)$ *Plot* plots $F_X^{-1}(y;\hat{\theta})$ against $\hat{F}^{-1}(y)$ for $y\in\{x_1,\dots,x_n\}$. It is has to derive $\hat{F}^{-1}$ so we notice that $\hat{F}^{-1}\left(\frac{k}{n}\right)=x_{(k)}$ & now plot $F_X^{-1}\left(\frac{k}{n};\hat{\theta}\right)$ against $x_{(k)}$. The closer a plot is to $y=x$ the better the estimator is.

Definition 4.3 - *Sample Distribution*
A *Sampling Distribution* is the distribution of an estimator. This distribution depends on the sample & not the population distribution.

Definition 4.4 - *Bias*
Bias is the mean distance of the estimated value for a parameter, from its true value.

$$\text{Bias}(\hat{\theta};\theta) := \mathbb{E}(\hat{\theta}-\theta;\theta) = \mathbb{E}(\hat{\theta};\theta) - \theta$$

N.B. - An estimator is unbiased if $\forall\ \theta,\ \text{Bias}(\hat{\theta};\theta) = 0$.

Definition 4.5 - *Mean-Square Error*
Mean-Square Error is the mean distance of the estimated value for a parameter, from its true value.

$$\text{mse}(\hat{\theta};\theta) := \mathbb{E}\left[(\hat{\theta}-\theta)^2;\theta\right] = \text{Var}(\hat{\theta};\theta) + \text{bias}(\hat{\theta};\theta)^2$$

Definition 4.6 - *Simulation*
Simulation is the process of generating data sets from a given distribution. This provides an empirical approach to assessing estimators, rather than theoretical. Simulations only provide information about one scenario, only. We cannot deduce anything for other sample sizes or true values of the parameters.

5 Linear Regression

Definition 5.1 - *Linear Regression, Predictor Variables & Response Variables*
Consider a data set $\{x_1,\dots,x_n\}$ & a definition for another variable $y_i = \alpha x_i + \beta$. Here x_i is the *Predictor Variable* & y_i is the *Response Variable. Linear Regression* is a technique for estimating the values of α & β from a sample $\{(x_1,y_1),\dots,(x_n,y_n)\}$.

Definition 5.2 - *Linear Regression Model*
The *Linear Regression Model* states that $\mathbb{E}(y_i|x_i) = \alpha + \beta x_i$ for some $\alpha,\beta \in \mathbb{R}$. The *Linear Regression Model* assumes that the relationship between x_i & y_i is of the form $y_i = \alpha + \beta x_i + e_i$ where e_i is an error measure with $\mathbb{E}(e_i) = 0,\ \text{Var}(e_i) = \sigma^2$ & $\text{Cov}(e_i,e_j) = 0\ \forall\ i \neq j$.

Theorem 5.1 - *Finding $\hat{\alpha}$ & $\hat{\beta}$*
Using the *Linear Regression Model*, the *Least-Squares Estimates* for α & β are the solutions

$$\hat{\alpha},\hat{\beta} := \text{argmin}_{\alpha,\beta}\sum_{i=1}^{n}(y_i - (\alpha+\beta x_i))^2$$

Which can be found explicitly as

$$\hat{\beta} = \frac{ss_{xy}}{ss_{xx}} \quad \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

Definition 5.3 - *Fitted values*
Fitted values are the estimated values for y_i using the estimated values for α & β

$$\hat{y}_i := \hat{\alpha} + \hat{\beta}x_i$$

Definition 5.4 - *Residual Values*
Residual values are the difference between observed values & fitted values

$$\hat{e}_i := y_i - \hat{y}_i$$

We define an error measure, the *Residual Sum of Squares*

$$\text{RSS} := \sum_{i=1}^{n}\hat{e}_i^2 \equiv ss_{yy} - \frac{ss_{xy}^2}{ss_{xx}}$$

We can produce a *Residual Plot* by producing a bar chart with the *Predictor Variable* on the x-axis & residuals on y-axis.

Definition 5.5 - *Best Predictor*
We define the *Best Predictor* to be an explicit equation for the *Response Variable* in terms of the *Predictor Variable*

$$\hat{y} := \hat{\alpha} + \hat{\beta}x$$

We can plot this as a line on a graph & plot the observed values as scatter points on the graph. This plot can be used to assess the quality of fit. If the fit seems poor we may want to increase the complexity of the model *e.g.* $y = \alpha + \beta x + \gamma x^2$ etc.

6 Confidence

Definition 6.1 - *Percentage Points*
Percentage Points, x_α are values within a distribution which have a pre-determined likelihood, α, of values in that distribution being greater than them.

$$\mathbb{P}(X \geq x_\alpha) = \alpha$$

Definition 6.2 - *Confidence Interval*
Consider a simple random sample $X_1, \dots, X_n$. We define a $100(1-\alpha)\%$ *Confidence Interval* (for $\alpha \in (0,1)$ for a parameter θ to be the interval $[c_l, c_u]$ where the probability of the true value of the parameter being in this interval is $1-\alpha$.

$$\mathbb{P}(c_l \leq \theta^* \leq c_u; \theta) \geq 1 - \alpha$$

N.B. - We define $|c_u - c_l|$ to be the *Length* of the confidence interval.

Proposition 6.1 - *Finding Confidence Intervals - Normal Distribution*
Consider finding a confidence interval for μ using a sample from $N(\mu, \sigma_0^2)$, with σ_0^2 known. We know $\overline{X} \sim N(\mu, \frac{1}{n}\sigma_0^2)$ and $\frac{\overline{X}-\mu}{\sigma_0/\sqrt{n}} \sim N(0,1)$. Thus we find the confidence interval by rearranging the following to be in terms of μ.

$$\mathbb{P}\left(-z_{\alpha/2} \leq \frac{\overline{X}-\mu}{\sigma_0/\sqrt{n}} \leq z_{\alpha/2}; \mu, \sigma_0^2\right) = 1 - \alpha$$

Consider finding a confidence interval for μ using a sample from $N(\mu, \sigma^2)$, with σ^2 unknown. We know that $\frac{\overline{X}-\mu_0}{s/\sqrt{n}} \sim t_{n-1}$ thus we find the confidence interval by rearranging the following to be in terms of μ.

$$\mathbb{P}\left(-t_{n-1;\alpha/2} \leq \frac{\overline{X}-\mu}{\sigma_0/\sqrt{n}} \leq t_{n-1;\alpha/2}; \mu, \sigma^2\right) = 1 - \alpha$$

Consider finding a confidence interval for σ^2 using a sample from $N(\mu, \sigma^2)$, with μ unknown. We know that $\frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \overline{X})^2 \sim \chi^2_{n-1}$ then we find the confidence interval by rearranging the following to be in terms of σ^2.

$$\mathbb{P}\left(\chi^2_{n-1;1-\frac{\alpha}{2}} \leq \frac{1}{\sigma^2}\sum_{i=1}^n (X_i - \overline{X})^2 \leq \chi^2_{n-1;\frac{\alpha}{2}}; \mu, \sigma^2\right) = 1 - \alpha$$

Proposition 6.2 - *Affecting Confidence Interval Length*

i) Increasing sample size decrease confidence interval length;

ii) Increasing population variance increases confidence interval length; and,

iii) Increasing confidence level, α, increases confidence interval length.

7 Hypothesis Testing

Definition 7.1 - *Hypothesis Testing*
Hypothesis Testing is the process of evaluating whether a sample is consistent with one of two contrasting statements about the population parameters. The statement which is deemed true, unless proved otherwise, is called the *Null Hypothesis*, H_0. The other statement is called the *Alternative Hypothesis*, H_1. There are two types of *Hypothesis Tests*: two-tailed ($\theta \neq \alpha$); or, one-tailed ($\theta > \alpha$ or $\theta < \alpha$). It is important to pay attention to which sort of test we are using as it affects the confidence level.

Proposition 7.1 - *Comparing Two Groups*
Some *Hypothesis Tests* require the comparison of two groups. There are two possible relationships between the provided examples

i) *Independent Samples* - The two groups are independent of one another & thus can be modelled by different population distributions.

ii) *Paired Samples* - The samples represent two observations from each member of the population (each member appears in both samples).
e.g. Patients given drug A & then drug B, some time later.

We can often combine two samples into a single distribution by comparing sample means & then proceeding as if a single sample test.

Proposition 7.2 - *Pooled Variance*
If we can assume that two *Independent Samples* (X of size m & Y of size n) have the <u>same</u> variance we can define a *Pooled Variance*

$$s_p^2 := \frac{\left(\sum_{i=1}^m (X_i - \overline{X})^2\right) + \left(\sum_{i=1}^n (Y_i - \overline{Y})^2\right)}{m+n-2}$$

Then we can define test statistic

$$T = \frac{\overline{X} - \overline{Y}}{s_p\sqrt{(1/m)+(1/n)}} \sim t_{m+n-2}$$

Definition 7.2 - *Types of Error*

i) A *Type I Error* is when the *Null-Hypothesis* is rejected, when in fact it is true. *i.e.* False Negative.

ii) A *Type II Error* is when the *Null-Hypothesis* is accepted, when in fact the *Alternative Hypothesis* is true. *i.e.* False Positive.

Definition 7.3 - *Power*
The *Power* of a *Hypothesis test* is the probability of correctly identifying the *Alternative Hypothesis* as true

$$\text{Power} := 1 - \mathbb{P}(\text{Type II Error})$$

Proposition 7.3 - *Hypothesis Testing Procedure*

1) State *Model Assumptions* for distribution & independence of variables.

2) State *Null Hypothesis* & *Alternative Hypothesis*;

3) Choose an appropriate *Test Statistic*, T, & calculate its value, t_{obs}, using observations from sample.

4) Compute the p-value for the test statistic value;
   i.e. Probability of a sample having the calculated observed value, assuming H_0 is true.
   e.g. If $H_1 : \mu > \mu_0 \implies p = \mathbb{P}(T \geq t_{obs} | H_0 \text{ true})$.

5) Make conclusions about whether there is sufficient evidence to reject the *Null Hypothesis*.

Proposition 7.4 - *Which Test Statistic to use when comparing means*

Test Name	Samples	Variance?	Test Statistic, T
One-Sample t-Test	Single Sample	Known	$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$
One-Sample t-Test	Single Sample	**Un**known	$\frac{\overline{X}-\mu}{s/\sqrt{n}} \sim t_{n-1}$
Pooled Two-Sample t-Test	Independent Samples	Equal	$\frac{\overline{X}-\overline{Y}}{s_p\sqrt{(1/m)+(1/n)}} \sim t_{m+n-2}$
Welch Two-Sample t-Test	Independent Samples	**Un**equal	$\frac{\overline{X}-\overline{Y}}{\sqrt{(\widehat{\sigma_X^2}/m)+(\widehat{\sigma_Y^2}/n)}} \sim t_\nu$
Paired Two-Sample t-test	Paired Samples		$W_i := X_i - Y_i \sim N(\mu_W, \sigma_W^2) \implies \frac{\overline{W}-\mu_W}{s_W/\sqrt{n}} \sim t_{n-1}$

N.B. - $v := \dfrac{((s_X^2/m)+(s_Y^2/n))^2}{\frac{1}{m-1}(s_X^2/m)^2 + \frac{1}{n-1}(s_Y^2/n)^2}$.

Definition 7.4 - *Critical Region*
Defining a *Critical Region* is an alternative approach to testing two hypotheses about population parameters. Here we calculate the value(s), c^*, such that if a certain observed value falls on the wrong side of this value(s) then H_0 is rejected. These value(s) are called *Critical Value(s)*, c^*.

Proposition 7.5 - *Linear Regression & Hypothesis Tests*
In order to perform *Hypothesis Test* on the results of *Linear Regression* we define $y_i = \alpha + \beta x_i + e_i$ where $e_1, \dots, e_n$ are iid $N(0, \sigma^2)$. Here α, β & σ^2 are unknown. Since e_i are normal distributed we have

$$\hat{\alpha} \sim N\left(\alpha, \frac{\sigma^2}{nss_{xx}} \sum_{i=1}^n x_i^2\right) \quad \hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{ss_{xx}}\right)$$

By defining estimates for the standard deviation of $\hat{\alpha}$ & $\hat{\beta}$ as $s_{\hat{\alpha}} := \sqrt{\frac{\widehat{\sigma^2}}{nss_{xx}} \sum_{i=1}^n x_i^2}$ & $s_{\hat{\beta}} := \sqrt{\widehat{\sigma^2}/ss_{xx}}$. Then

$$\frac{\hat{\alpha} - \alpha^*}{s_{\hat{\alpha}}} \sim t_{n-2} \quad \frac{\hat{\beta} - \beta^*}{s_{\hat{\beta}}} \sim t_{n-2}$$

0 Reference

0.1 Definitions

Definition 0.1 - *Empirical Distribution Function*
For a simple random sample $(X_1, \dots, X_n)$ the *Empirical Distribution Function* is an estimate of the cumulative frequency function. *Empirical Distribution Function* is defined as

$$\hat{F}(y) := \frac{1}{n} |\{X_i \leq y\}|$$

Definition 0.2 - *Moment Generating Function*
The *Moment Generating Function* is a real-valued function that uniquely describes a probability distribution. Two distributions with the same *Moment Generating Function* are equivalent.

$$\begin{array}{rrcccl} \text{Population} & \mathcal{M}_X(s) & := & \mathbb{E}\left(e^{kX}\right) & = & \int_{-\infty}^{\infty} e^{kx} f_X(x)dx \\ \text{JointPopulation} & \mathcal{M}_{X,Y}(s,t) & := & \mathbb{E}\left(e^{sX+tY}\right) & = & \mathcal{M}_X(s)\mathcal{M}_Y(t) \\ \text{Sample} & m_k & := & \frac{1}{n}\sum_{i=1}^{n} x_i^k & & \end{array}$$

Definition 0.3 - *Simple Random Sample*
A *Simple Random Sample* is an discrete, unbiased sample from a population. The likelihood of a particular element being in the sample depends upon the distribution of the population.

0.2 Notation

Definition 0.4 - *Estimation*
$\hat{\theta}$ denotes an estimation. θ^* denotes the true value. $\hat{\tau} = \tau(\hat{\theta})$ denotes applying an estimation of θ to a function τ of θ.

Notation 0.1 - *Order Statistic*
For a data set $D = \{x_1, \dots, x_n\}$ we denote the i^{th} smallest value as $x_{(i)}$.
Notation 0.2 - *Percentage Points*
For *Percentage Points* different notation is used for different distributions

RV	Notation
$Z \sim N(0,1)$	$\mathbb{P}(Z \geq z_\alpha) = \alpha$
$T \sim t_r$	$\mathbb{P}(T \geq t_{r,\alpha}) = \alpha$
$W \sim \chi^2_r$	$\mathbb{P}(W \geq \chi^2_{r,\alpha}) = \alpha$

Notation 0.3 - *Summary Statistics*
For samples of equal size $X = \{x_1, \dots, x_n\}$ & $Y = \{y_1, \dots, y_n\}$ we define

$$\begin{array}{rcl|rcl} \bar{x} & = & \frac{1}{n}\sum_{i=1}^n x_i & \bar{y} & = & \frac{1}{n}\sum_{i=1}^n y_i \\ ss_{xx} & = & \left(\sum_{i=1}^n x_i^2\right) - n\bar{x}^2 & ss_{yy} & = & \left(\sum_{i=1}^n y_i^2\right) - n\bar{y}^2 \\ ss_{xy} & = & \left(\sum_{i=1}^n x_i y_i\right) - n\bar{x}\bar{y} & & & \end{array}$$

0.3 Theorems

Theorem 0.1 - *Central Limit Theory*
Let $X_1, \dots, X_n$ be iid rv from a random sample of a population with mean $\mathbb{E}(X)$ & variance $\sigma^2 = \text{Var}(X)$. Then, for sufficiently large n

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \simeq N(0,1)$$

Since the distribution of X may only allow discrete values but the normal distribution is continuous we perform *Continuity Correction*

$$\begin{array}{rcl}\mathbb{P}(X=x) & \simeq & \mathbb{P}\left(x-\frac{1}{2}\leq X\leq x+\frac{1}{2}\right)\\ \mathbb{P}(X\leq x) & \simeq & \mathbb{P}\left(X\leq x+\frac{1}{2}\right)\end{array}$$

Theorem 0.2 - *Independence of Sample Mean & Mean-Square Distance*
Let $X_1,\dots,X_n$ be iid rv distributed with $N(\mu,\sigma^2)$ then

$$\overline{X}\ \&\ \sum_{i=1}^{n}\left(X_j-\bar{X}\right)^2 \text{ are independent.}$$

Theorem 0.3 - *Moment Generating Function Identities*

$$\begin{array}{rcl}\mathcal{M}_{aX+b}(t) & = & \mathbb{E}\left(e^{atX+bt}\right)=e^{tb}\mathcal{M}_X(ta)\\ \mathcal{M}_X(s) & = & \mathcal{M}_{X,Y}(s,0)\\ \mathcal{M}_Y(t) & = & \mathcal{M}_{X,Y}(0,t)\\ \mathcal{M}_{X_1+\dots+X_n}(s) & = & \mathcal{M}_{X_1}(s)\dots\mathcal{M}_{X_n}(s)\end{array}$$

Theorem 0.4 - *Transformations of Normal Distribution*
Let $X\sim N(\mu,\sigma^2)$. Then

$$\begin{array}{rcl}aX+b & \sim & N(a\mu+b,a^2\sigma^2)\\ \sum_i a_iX_i & \sim & N\left(\sum_i a_i\mu_i,\sum_i a_i^2\sigma^2\right)\end{array}$$

Theorem 0.5 - χ^2 *Distribution Identities*
Let $W\sim\chi_r^2$ then

$$\begin{array}{rcll}X & \sim & \Gamma\left(\frac{r}{2},\frac{1}{2}\right) & \\ N(0,1)^2 & \simeq & \chi_1^2 & \\ X+Y & \sim & \chi_{r+s}^2 & \text{for } X\sim\chi_r^2,\ Y\sim\chi_s^2\end{array}$$

Theorem 0.6 - t *Distribution Identities*
Let $X\sim t_r$ then

$$\text{As } r\to\infty \quad \text{then} \quad X\sim N(0,1)$$

0.4 Probability Distributions

Definition 0.5 - *Binomial Distribution*
Let X be a discrete random variable modelled by a *Binomial Distribution* with n events and rate of success p.

$$\begin{array}{rcl} p_X(k) & = & \binom{n}{k}p^k(1-p)^{n-k} \\ \mathbb{E}(X) & = & np \\ \text{Var}(X) & = & np(1-p) \end{array}$$

Definition 0.6 - *Gamma Distribution*
Let T be a continuous randmo variable modelled by a *Gamma Distribution* with shape parameter α & scale parameter λ. Then

$$\begin{array}{rcl} f_T(x) & = & \dfrac{\lambda^\alpha x^{\alpha-1}e^{-\lambda x}}{\Gamma(\alpha)} \quad \text{for } x > 0 \\ \mathbb{E}(T) & = & \dfrac{\alpha}{\lambda} \\ \text{Var}(T) & = & \dfrac{\alpha}{\lambda^2} \end{array}$$

N.B. - $\alpha, \lambda > 0$.

Definition 0.7 - *Exponential Distribution*
Let T be a continuous random variable modelled by a *Exponential Distribution* with parameter λ. Then

$$\begin{array}{rcll} f_T(t) & = & \lambda e^{-\lambda t} & \text{for } t > 0 \\ F_T(t) & = & 1 - e^{-\lambda t} & \text{for } t > 0 \\ \mathbb{E}(X) & = & \frac{1}{\lambda} & \\ \text{Var}(X) & = & \frac{1}{\lambda^2} & \end{array}$$

Definition 0.8 - *Normal Distribution*
Let X be a continuous random variable modelled by a *Normal Distribution* with mean μ & variance σ^2.
Then

$$\begin{array}{rcl} f_X(x) & = & \dfrac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ F_X(x) & = & \dfrac{1}{\sqrt{2\pi\sigma^2}}\displaystyle\int_{-\infty}^{x} e^{-\frac{(y-\mu)^2}{2\sigma^2}}\,dy \\ \mathcal{M}_X(\theta) & = & e^{\mu\theta+\sigma^2\theta^2(1/2)} \\ \mathbb{E}(X) & = & \mu \\ \text{Var}(X) & = & \sigma^2 \end{array}$$

Definition 0.9 - *Poisson Distribution*
Let X be a discrete random variable modelled by a *Poisson Distribution* with parameter λ. Then

$$\begin{array}{rcl} p_X(k) & = & \frac{e^{-\lambda}\lambda^k}{k!} \quad \text{for } k \in \mathbb{N}_0 \\ \mathbb{E}(X) & = & \lambda \\ \text{Var}(X) & = & \lambda \end{array}$$

Definition 0.10 - χ^2 *Distribution*
Let X be a random variable modelled by the χ^2 *Distribution* with r degrees for freedom. Then

$$\begin{array}{rcl} \mathcal{M}_X(t) & = & (1-2t)^{-r/2} \\ \mathbb{E}(X) & = & r \\ \text{Var}(X) & = & 2r \end{array}$$

Definition 0.11 - *$t2$ Distribution*
Let X be a random variable modelled by the *t-Distribution* with r degrees for freedom. Then

$$\begin{array}{rcl}\mathbb{E}(X) & = & 0 \\ \text{Var}(X) & = & \dfrac{r}{r-2}\end{array}$$